universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Ecological and evolutionary consequences of smRNA activity after allopolyploidization in marsh orchids (*Dactylorhiza majalis* s.l.)"

verfasst von / submitted by

Matthew Ryan Thornton

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 833
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Ecology and Ecosystems
Betreut von / Supervisor:	Assoz. Prof. Dipl.-Ing. Dr. Ovidiu Paun

Table of Contents

Abstract

Polyploidy and hybridization have a major impact on the adaptation and speciation of angiosperms. Together, they provide evolutionary avenues by which established allopolyploids can diversify and expand their ecological niche outside of parental ranges. There is a complex suite of molecular processes behind the diversification of allopolyploids, some known and some requiring attention, and uncovering the underlying mechanisms yields insights into their formation, evolution, and speciation. We investigate a system of two ecologically distinct sibling allotetraploids (*Dactylorhiza majalis, D. traunsteineri*), rising from the unidirectional hybridization of the same diploid progenitors (*D. incarnata, D. fuchsii)*, to uncover the role of differential small RNA (smRNA) regulation in their observed eco-physiological divergence. smRNAs have known roles in stabilizing the genomes of nascent polyploids, and are heavily involved in plant development, metabolism, and stress response making them interesting candidates for studying polyploid diversification. We find that in a common garden, the sibling allotetraploids largely share smRNA targeting characteristics, and the inherited targeting patterns from diploid progenitors are highly similar and partially conserved between the two. However, several genes that were differentially targeted by smRNAs were found to play key roles in environmentally linked signaling pathways. Reciprocal transplant experiments revealed a strong environmental effect on smRNA targeting with the polyploids showing differential plasticity in smRNA activity. We find that smRNA targeting is influenced by gene flow between the polyploids, and heavily dependent on the tissue type under study. This work is part of a larger goal of identifying the driving factors behind the rich allopolyploid diversity in the *Dactylorhiza* genus.

Kurzfassung

Polyploidie und Hybridisierung haben einen großen Einfluss auf die Anpassung und Artbildung von Angiospermen. Zusammen bieten sie evolutionäre Möglichkeiten, durch die etablierte Allopolyploide diversifizieren und ihre ökologische Nische außerhalb der elterlichen Verbreitungsgebiete erweitern können. Hinter der Diversifizierung von Allopolyploiden steht eine komplexe Reihe von molekularen Prozessen, von denen einige bekannt sind und andere noch der Aufmerksamkeit bedürfen, und die Aufdeckung der zugrundeliegenden Mechanismen ermöglicht Einblicke in ihre Entstehung, Evolution und Artbildung. Wir untersuchen ein System von zwei ökologisch unterschiedlichen allotetraploiden Geschwistern (*Dactylorhiza majalis, D. traunsteineri*), die aus der unidirektionalen Hybridisierung derselben diploiden Vorläufer (*D. incarnata, D. fuchsii*) hervorgegangen sind, um die Rolle der unterschiedlichen Regulierung kleiner RNAs (smRNAs) bei der beobachteten ökophysiologischen Divergenz aufzudecken. Es ist bekannt, dass smRNAs eine Rolle bei der Stabilisierung der Genome von entstehenden Polyploiden spielen und stark in die Pflanzenentwicklung, den Stoffwechsel und die Stressreaktion involviert sind, was sie zu interessanten Kandidaten für die Untersuchung der polyploiden Diversifizierung macht. Wir stellen fest, dass die allotetraploiden Geschwister in einem gemeinsamen Garten weitgehend die gleichen smRNA-Targeting-Merkmale aufweisen, und dass die von den diploiden Vorfahren vererbten Targeting-Muster sehr ähnlich und teilweise zwischen den beiden konserviert sind. Es wurde jedoch festgestellt, dass mehrere Gene, auf die smRNAs unterschiedlich abzielen, eine Schlüsselrolle in umweltbezogenen Signalwegen spielen. In Experimenten zur reziproken Transplantation wurde ein starker Umwelteinfluss auf das smRNA-Targeting festgestellt, wobei die Polyploiden eine unterschiedliche Plastizität der smRNA-Aktivität aufwiesen. Wir stellen fest, dass das smRNA-Targeting durch den Genfluss zwischen den Polyploiden beeinflusst wird und stark von dem untersuchten Gewebetyp abhängt. Diese Arbeit ist Teil eines

größeren Ziels, die treibenden Faktoren hinter der reichen allopolyploiden Vielfalt in der Gattung *Dactylorhiza* zu identifizieren.

1. Introduction

Polyploidy is a driving force of evolution and plays a major role in the adaptation and speciation of angiosperm taxa (Soltis & Soltis 1999; Soltis et al., 2009; Van de Peer et al., 2009). Polyploids arise through whole genome duplication (WGD), and according to their formation they can be characterized as autopolyploids, resulting from the duplication of a single genome, or allopolyploids, resulting from the hybridization of genomes from different species. It is known that the angiosperms have undergone several rounds of WGD in their history and these events were critical for the abundance of diversity observed today (Soltis et al., 2009). An estimated 47-70% of angiosperm species are currently polyploids (Ramsey & Schemske, 1998) and polyploidy is estimated to be directly implicated in 15% of angiosperm speciation events (Wood et al., 2009). Thus, WGD events are recurrent, and ubiquitous in angiosperms, suggesting a conferred advantage under a diversity of selective pressures. The proven implications of polyploidy in evolution and adaption, coupled with an incomplete bank of knowledge concerning the mechanisms by which polyploidy events can affect the evolutionary and ecological trajectories of the taxa involved, provide many opportunities for the advancement of research in these fields.

If a neopolyploid can surpass the obstacles hindering its establishment, a variety of short- and long-term evolutionary benefits may come into play. WGD can provide the genomic material required for adaption due to the presence of duplicated genes. With multiple copies of each gene, polyploids have a “safety net” for evolutionary neo-functionalization experiments that could otherwise be deleterious or lethal at the diploid level. This can lead to a diversification of gene function and allow polyploids to expand their phenotypic variation and tolerance to new biotic or abiotic conditions (Moore

& Purugganan, 2005; Adams & Wendel, 2005; Comai, 2005). Another advantage conferred by polyploidy is the fixation of heterosis and hybrid vigor effects. It has been well documented that selfing or inbreeding in a diploid population leads to a reduction in fitness and overall vigor of the affected individuals (Crnokrak & Roff, 1999; Wang et al., 1999; Keller & Waller 2002). Polyploids can escape this evolutionary depression by masking deleterious allele accumulation (dominance model of Crow, 1948), or altering gene expression landscapes (Birchler et al., 2003); and the degree of hybrid vigor in viable hybrids was found to be proportional to the genic distance of the parents, supporting a stronger benefit in allopolyploid lineages (East, 1936). Polyploidy can also interfere with self-incompatibility systems and allow for self-fertilization in progeny. The mechanisms are not fully elucidated but there is evidence for polyploid induced self-fertilization capabilities, conferring an advantageous mechanism of propagation (Miller & Venable, 2000).

However, while potentially advantageous, WGD is not trivial for an organism to tolerate. There is a range of challenges that nascent polyploids must overcome to propagate and establish as a population. A neopolyploid must tolerate meiotic disruption (Santos et al., 2003; Paun et al., 2007), changes in gene expression levels (Guo et al., 1996; Adams & Wendell, 2005), epigenetic instability (Scheid et al., 1996; Wang et al., 2004), and physiological changes including increased cell volume (Kondorosi et al., 2000). Additionally, the polyploids must overcome population level challenges including minority cytotype and mating disadvantages (Levin, 1975; Burton & Husband, 2000), and a sharp initial reduction in genetic diversity creating an evolutionary bottleneck event.

On a molecular level, there are regulatory mechanisms that can help neopolyploids adapt to the genomic shock resulting from drastic shifts in genomic composition, gene expression and regulation, and meiotic patterns. One class of molecules that have a role in buffering against the genomic shock of WGD is small RNA (smRNA; Ha et al., 2009). smRNAs are a subset of RNA molecules with a base length ranging in plants from 20-24 nucleotides. Due to differences in smRNAs among organisms this paper

will consider only smRNA characteristics described in plants. In short, smRNAs are synthesized either by the action of Dicer Like proteins (DLCs) on single stranded RNA precursors with a hairpin structure (hpRNA), or RNA-dependent RNA polymerases (RDRs) forming double stranded RNA (dsRNA) precursors. The former mechanism leads to the formation of micro-RNA (miRNA) and other hairpin derived non-miRNA, while the latter results in the formation of different types of small interfering RNA (siRNA) including phased (pha) siRNA, trans-acting (ta) siRNA, and epigenetically activated (ea) siRNA. Once a double stranded smRNA duplex is formed, it complexes with a variety of Argonaute (AGO) proteins which effect binding to either coding or non-coding RNA molecules based on sequence complementarity. The RNA target and the type of smRNA-AGO complex determines the final effect which can range from direct degradation of mRNA, repressive chromatin modification, or the recruitment of co-factors that repress gene expression (Axtell, 2013; Borges et al., 2015).

smRNAs are involved in a wide range of regulatory processes and act on the transcriptional and post-transcriptional level. A significant function of smRNA is repression of gene expression through various mechanisms including mRNA cleavage, repressive epigenetic modifications, and translation repression (Borges et al., 2015). Historically, smRNAs are thought to have evolved as a defense against RNA viruses and were later adapted to regulate endogenous gene expression and mobile elements. They have documented roles in embryonic development (Nodine & Bartel, 2010), and in response to a wide range of biotic and abiotic stress factors. For example, research in several plant models detected miRNAs involved in responding to drought, salinity, and temperature stress with different miRNAs effecting either positive or negative regulation on the target genes (Li et al., 2008; Zhou et al., 2009; Wang et al., 2012; Kumar et al., 2018). Work done in *Arabidopsis thaliana* demonstrated the role of trans-acting siRNA (ta-siRNA) in modulating heat tolerance through targeting of components of the Heat Shock Protein complex (Li et al., 2014). miRNAs have also been implicated in many studies involving nutrient stress response including copper (Cu^{2+}), nitrogen (N), and phosphate (P_i) deficiencies in

different plant models (Fujii et al., 2005; Sunkar et al., 2006; Gifford et al., 2008; Zhao et al., 2011). In polyploids, smRNAs have been shown to reduce the initial shock of WGD in part by controlling aberrant transposable element (TE) activation and chromatin remodeling following hybridization (Ha et al., 2009). Furthermore, previous work has demonstrated the implications of differential smRNA action accounting for morphological and phenotypic divergence between allopolyploids (Ng et al., 2012; Guan et al., 2014; Wei et al., 2019). Thus, smRNA involvement stabilizing the genomes in nascent polyploids, regulating development, metabolism, and stress response makes them an interesting candidate when studying the formation, and diversification of polyploids sharing similar genetic material.

In this study we utilize a system of phenotypically diverse sibling polyploids and their diploid parents. The setup investigates a group of four marsh orchids (*Dactylorhiza*) containing two diploids, the maternal *D. fuchsii* and paternal *D. incarnata* which hybridized multiple times to produce allopolyploid sibling offspring, *D. majalis* and *D. traunsteineri* (among other polyploid offspring not included in this study). The diploids have distinct morphologies and ecological preferences with an estimated divergence time estimated around 5.5MYA (Brandrud et al., 2020). *Dactylorhiza incarnata* has a larger genome though it is genetically less heterogeneous than *D. fuchsii* and is hypothesized to have gone through a bottleneck in its history (Brandrud et al., 2020). The allotetraploids are both phenotypically and ecologically distinct but can grow near each other in locations in the Alps and have reported gene flow (Balao et al., 2016). They are thought to have separate origins with *D. majalis* as the older of the two, thought to have formed roughly 100K generations ago, also exhibiting a broader ecological preference compared to *D. traunsteineri*, which likely formed 74K generations ago (Hawranek, 2021). They are known to prefer distinct soil chemistries with *D. traunsteineri* inhabiting nitrogen, phosphate, and potassium deficient soil (Wolfe et al., 2021). This system of established polyploids is excellent to investigate the underlying mechanisms which allow differences to establish and endure in the face of common genetic material and gene flow between sibling allopolyploids. It was

hypothesized that the recurrent formation of *D. majalis* and *D. traunsteineri* drove the diversification of their eco-physiological characteristics and allowed them to remain distinct despite intense gene flow and common genetic backgrounds (Wolfe et al., 2021).

This study aims to investigate the role that smRNA regulation plays in the ecological divergence, and adaptation following the unidirectional formation of two sibling allopolyploids by utilizing smRNAseq data from common garden, and reciprocal transplantations experiments. We seek to uncover the role of smRNA in driving the biodiversity of the *Dactylorhiza* genus and, more broadly, the mechanisms facilitating and maintaining divergent evolution and adaptation of polyploids. A common garden setup was prepared to investigate encoded regulatory differences in ecologically relevant genes between the polyploid siblings, as well as the inherited regulatory patterns from each progenitor, by reducing sources of environmental variation. In addition, reciprocal transplantation experiments were set up for *D. majalis* and *D. traunsteineri* to understand the plasticity of smRNA targeting in response to environmental stressors. By using populations from localities with different degrees of gene flow between the siblings, we aimed to further investigate the effects of such regulatory differences on maintaining diversity.

2. Methods

2.1 Study Design: Common Garden

The common garden experiment included a system of two diploid marsh orchids, *Dactylorhiza fuchsii* (N=4) and *D. incarnata* (N=4); and two polyploid offspring, *D. majalis* (N=9) and *D. traunsteineri* (N=12). In short, samples included in this study originated from populations in the Pyrenees (*D. fuchsii, D. majalis*), Scandinavia (*D. incarnata, D. majalis*, *D. traunsteineri*), Britain (*D. incarnata, D. fuchsii, D. traunsteineri*), and the Alps (*D. incarnata, D. majalis*, *D. traunsteineri*). These samples were transplanted and grown in a common garden in Vienna, Austria for two years prior to sampling to dilute the residual

environmental effects of native environments. smRNA libraries were prepared following methods and parameters described in Balao et al., 2017.

2.2 Study Design: Reciprocal Transplants

We used control and transplanted individuals of *D. majalis* and *D. traunsteineri* at two localities in Austria (St. Ulrich and Kitzbuhel). Five samples from each *D. traunsteineri* and *D. majalis* were transplanted and grown in the environment of the other, and five other samples were transplanted within their own environment to control for any residual transplantation effects. Leaf and root tissue were collected after two growing seasons and libraries were prepared and sequenced following the workflow described in Balao et al., 2017.

2. 3 Read Mapping and Feature Counts

CLC Genomics Workbench v.8.0 (QIAGEN) software was used to filter for small RNA reads with a length of 20-24 nucleotides. This resulted in smRNAseq datasets for diploid and polyploid individuals from the common garden experiment (N=31), leaf tissue from transplanted polyploid individuals (N=38), and root tissue from transplanted polyploid individuals (N=39). The three datasets were analyzed independently but following the same procedure outlined in Table S3.

The filtered reads were mapped to a *Dactylorhiza incarnata* reference genome v.1.0 (Wolfe et al., 2021) using STAR v.2.7.3a (--AlignIntronMax 1, --outFilterMismatchNoverLmax 0.05, --outFilterMultimapNmax 100). The options here allow STAR to function as an ungapped aligner, which is recommended when working with smRNA sequences, while allowing multimapping reads under a threshold of 100 loci to include repeat associated sequences. Read coverage was assessed per individual along 100bp windows using bamCoverage from the deeptools2 v.3.5.1 package (Ramirez et al., 2016). Windows with more than ten mapped reads were isolated and became the functional unit of the study, to be used subsequently as features (peaks) in the featureCounts analysis. Please note that the

terms peak and region will be used interchangeable throughout this thesis. Consecutive (adjacent) windows were combined per accession, and a final file including merged peaks from all accessions was created using bedtools merge v.2.29.2 (Quinlan & Hall, 2010) to create the input feature file for featureCounts analysis. The 10 reads per 100bp threshold was chosen to reduce the noise, and therefore increase sensitivity, in subsequent differential targeting analyses by removing peaks unlikely to be biologically interesting for this study. Since the final peak file was in bed format, conversion to gff was necessary prior to featureCounts using R package Rtracklayer (Lawrence et al., 2009). A table of counts was generated using the featureCounts function of the subread v2.0.0 package (Liao et al., 2014), with flags accounting fractionally for multiple mappers (-M, --fraction). The resulting tables of counts were imported into R v.3.6.3 for differential expression (targeting) analyses. Technical replicates for select samples (mA10_1573, tA9_1553, trL_mTK3, trL_tMK4, trL_tTS1, trR_mTK1, trR_mTS2, trR_tMS2, trR_tMS3) were combined in the featureCounts table before any further analyses as samples will be normalized according to library size in subsequent steps. Sample tA11_1578 was removed due to suspected triploidy and samples trL_mMK2r, trL_tTK1, and trR_tTK1 were removed due to potential technical issues during smRNA isolation.

2.4 Differential Targeting and Dominance Patterns

To assess which regions of the genome were differentially targeted by smRNAs the edgeR v.3.34.0 differential expression pipeline (McCarthy et al., 2012; workflow outline in Chen et al., 2016) was followed. For common garden individuals, the model matrix included both diploid (*D. fuchsii* N=4, *D. incarnata* N=4) and polyploid data (*D. majalis* N=9, *D. traunsteineri* N=12). The model matrices for the transplanted individuals were separated into leaf tissue data (*D. majalis* N=18, *D. traunsteineri* N=17) and root tissue data (*D. majalis* N=18, *D. traunsteineri* N=17) for both localities. Group structures are shown in Table S1. All functions referenced below are included in the edgeR package unless otherwise stated.

The table of counts was filtered for low targeted peaks and the threshold filtered for peaks with more than two cpm, in at least the minimum sample group size; for common garden samples (minimum polyploid group size, N=9), transplanted leaf tissue samples (N=4), and transplanted root tissue samples (N=4). The counts tables were normalized using the calcNormFactors function which utilizes the trimmed mean of M-values (TMM) method to account for library size (Robinson & Oshlack, 2010). Following filtering and normalization, samples were grouped on an MDS plot using the plotMDS function from the limma package v.3.9 (Ritchie et al., 2015) to assess clustering and identify potential outlier samples. The edgeR pipeline uses a negative binomial distribution to model the peak counts and the dispersion parameter was calculated using estimateDisp (Chen et al., 2016). This parameter estimates the amount of variation between true targeting levels of biological replicates. A quasi-likelihood negative binomial model was fit using glmQLFit to prepare for differential targeting analysis which was conducted with function glmTREAT (lfc = 1.2) which is a statistically robust method to introduce a log fold change requirement to call significance. This analysis makes use of contrasts within the design matrix and these contrasts were created to compare each group of interest via makeContrasts. Differentially targeted peaks were called using decideTestsDGE with a Benjamini-Hochberg correction and an FDR threshold of 0.05.

To assess smRNA targeting dominance patterns the following definitions were applied. A peak was considered additive if it was significantly down-targeted in the polyploid relative to one parental diploid (parent A) and significantly up-targeted relative to the other parental diploid (parent B), indicating that the targeting was intermediate in value between the parental levels. A peak was considered transgressive if it was significantly DT relative to both parental species (targeting level exceeds the parental range in either direction). Lastly, a peak was considered dominant when there was no significant difference in targeting level relative to parent A, but a significant up or down targeting

level relative to parent B, indicating a larger inherited influence of one parental targeting pattern over the other.

2.5 Functional Annotation of Differentially Targeted Regions

The *D. incarnata* reference genome used in this study was previously annotated and included mRNA, 3'UTR, 5'UTR, and gene annotations. These were expanded to include exons, annotated by combining CDS, 3'UTR and 5'UTR regions using bedtools merge; introns, annotated by subtracting the exon annotations from the gene annotations with bedtools subtract; promoters, defined ad hoc by 1,000bp upstream from a gene start site; and intergenic, annotated by subtracting all other annotated regions from each full scaffold region. Transposable element annotations, including miniature inverted-repeat transposable elements (MITE), were also included, and annotations were provided by Mimmi Eriksson. For each accession, the distribution of read lengths was mapped to each genomic feature to provide an ad hoc overview of how potential smRNA species are targeting each genomic feature.

2.6 Gene Ontology Enrichments

Significantly DT regions were mapped onto genomic features associated with gene expression (exons, introns, promoters). Gene names for annotation features containing a DT region were extracted and compiled for each comparison and processed with Blast2GO v.5.2.5 (Götz et al., 2008) using Fisher's exact test (biological processes, reduced to most specific terms) to identify significantly enriched gene ontology (GO) terms in genes under differential smRNA regulation. For promoter regions, the gene adjacent to a promoter containing a DT region was used. Significant GO terms were visualized using REVIGO (Supek et al., 2011) and ggplot2 (Wickham, 2016).

2.7 Read Length Distributions

To provide a preliminary look at the distributions of different classes of smRNA and how their activity changes between different genomic features, read lengths totals were extracted for each genomic feature annotation using a combination of bioawk and sed following a script from the lab rotation report of Malte Mederacke (personal communication). Distributions were normalized by using each read length as a proportion of total reads and significance differences were calculated using Wilcoxon signed-rank test.

3. Results

3.1 Common Garden

Reads were size selected (20-24nt) and mapped to the *D. incarnata* reference genome. The polyploids had a higher average mapping efficiency (*D. majalis* and *D. traunsteineri* 84%; *D. fuchsii* 82%*; D. incarnata* 80%) though, interestingly, total reads mapped were highest in *D. fuchsii* and lowest in *D. incarnata* (Appendix A). Mismatch rates were highest in *D. fuchsii* (0.86) and lowest in *D. incarnata* (0.31) with *D. majalis* and *D. traunsteineri* intermediate (0.56, 0.59 respectively) which was expected given alignment towards the *D. incarnata* reference genome. Peak designation and subsequent merging produced a total of 2,454,438 peaks across all samples. Following featurecounts analysis, peaks were filtered by targeting level retaining 1.7% of the total peaks (40,832) for differential targeting analysis.

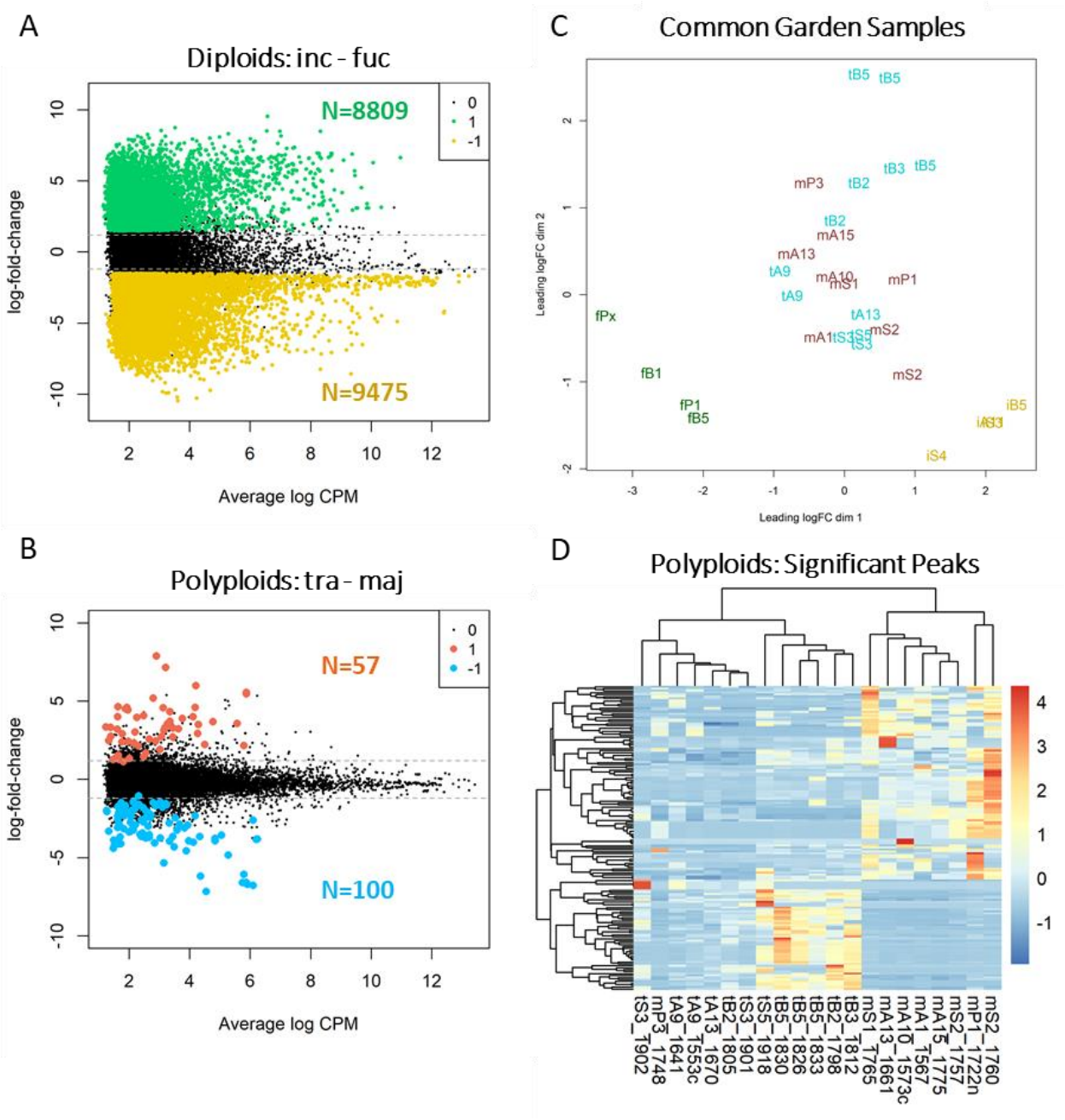


Figure 1. A-B) Mean difference (MA) plot for common garden diploids and polyploids which plots the log fold change (y-axis) against the average targeting level in counts per million (x-axis). Dashed lines indicate a log fold change of 1.2 which was used as a cutoff to call significance. **A)** Green points are significantly over-targeted regions in *D. incarnata* compared to *D. fuchsii,* and yellow points are significantly under-targeted. **B)** Red points are significantly over targeted regions in *D. traunsteineri* compared to *D. majalis*, and blue are significantly under targeted. **C)** MDS plot for common garden samples showing D. *fuchsii* (green), *D. incarnata* (yellow), *D. majalis* (red), and *D. traunsteineri* (blue). Samples are clustered based on the leading log fold change in the count data matrix. **D)** Heatmap of z-transformed smRNA targeting levels in DT regions (N=157) among common garden polyploids. The regions are DT in *D. traunsteineri* compared to *D. majalis* with red colors indicating higher targeting levels and blue indicating lower targeting levels.

Following filtering and normalization steps the sample similarity was visualized with a multidimensional scaling (MDS) plot of the count matrix (Fig. 1C). MDS plots provide a valuable first insight on how large, multivariate datasets are behaving and facilitate the identification of outliers in a given group. The clear clustering by diploid sample groups, and intermediate clustering of the polyploids in respect to each diploid provided confidence in the filtering and integrity of the data included in the subsequent analyses.

Differential targeting analysis yielded 157 regions (0.39%) that were differentially regulated by smRNAs between *D. traunsteineri* and *D. majalis*. Of these, almost two thirds (N=100) were under-targeted regions in *D. traunsteineri* (Fig. 1B). Diploid comparisons were included in the DT analyses as a sanity check to validate the parameters of the workflow and its output. There is a high degree of differential targeting between the diploid species (44.8% of included peaks), which is expected given their estimated 5.5MYA divergence (Brandrud et al., 2020), and subsequent genetic and transcriptomic differences (Fig. 1A). The heatmap in Fig. 1D show how targeting levels of significantly DT regions behave between polyploid samples, and cluster samples by similarity in these levels on the x-axis. Accessions cluster largely by sibling type with exception to one *D. majalis* sample originating from the Pyrenees (mP3_1748) which diverges from the rest of the *D. majalis* accessions.

To investigate biological implications of the differentially regulated regions, all significant DT peaks were intersected with annotated genomic features associated with gene expression. As the reference annotation file only contained annotations for gene, mRNA, CDS, and 3'- and 5'UTR, the following regions were defined manually (see methods section). DT regions were mapped to exons to provide a proxy for post-transcriptional regulation via transcript degradation, and a combination of intron and promoter regions to give insight on pre-transcriptional regulation via gene silencing. The gene names of all regions containing a significantly DT region were extracted and used as input for the

GO enrichment analysis. Gene names lists were analyzed separately for exons (post-transcriptional regulation) and introns + promoters (transcriptional regulation; Fig. 2).

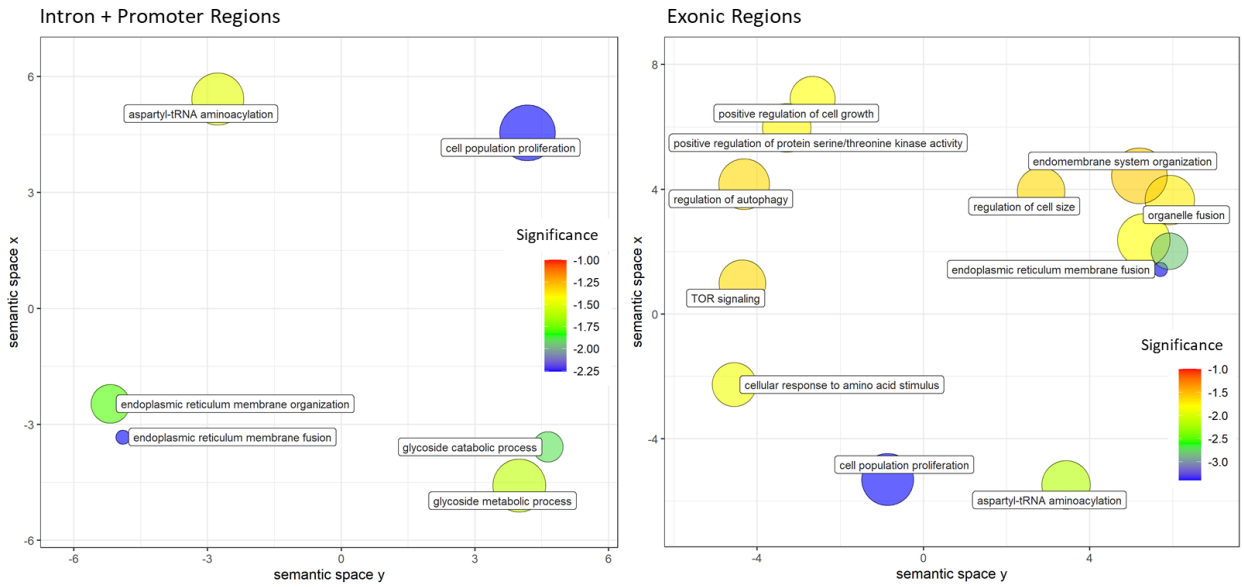


Figure 2. Left) Significantly enriched GO terms in DT intronic and promoter regions between the polyploids in a common garden. The gene names were derived, and combined, from DT regions mapping to introns and promoters to infer a set of pre-transcriptionally regulated genes. **Right)** Significantly enriched terms in exon regions of common garden polyploids. The gene names were derived from mapping DT regions to exon annotations to provide a list of post-transcriptionally regulated genes. GO enrichment was performed with Blast2GO (Fisher's exact test, FDR<0.05) and visualized with REVIGO. Circle size is proportional to the frequency of the GO term in the database (smaller circles are more specific terms) and color represents significance level (blue colors are more significant and red colors less significant).

This study did not endeavor to classify the different types of smRNA. Instead, read length was used as an ad hoc classifier of smRNA type to infer how the different classes were targeting different genomic features. The distribution of read lengths were extracted for each genomic feature as well as the overall pattern across the full reference genome (Fig. 3A-B). Significantly DT regions were also mapped and summarized for each annotation to understand which regions are most differentially regulated between all common garden comparison (Fig. 3C). Exons show the highest proportions of 21nt reads in both polyploids with approximately 40% of all reads, while transposable elements contain the highest proportions of 24nt reads ranging from 30-45% of all reads. This pattern fits nicely with the expected smRNA classes thought to be acting on these features.

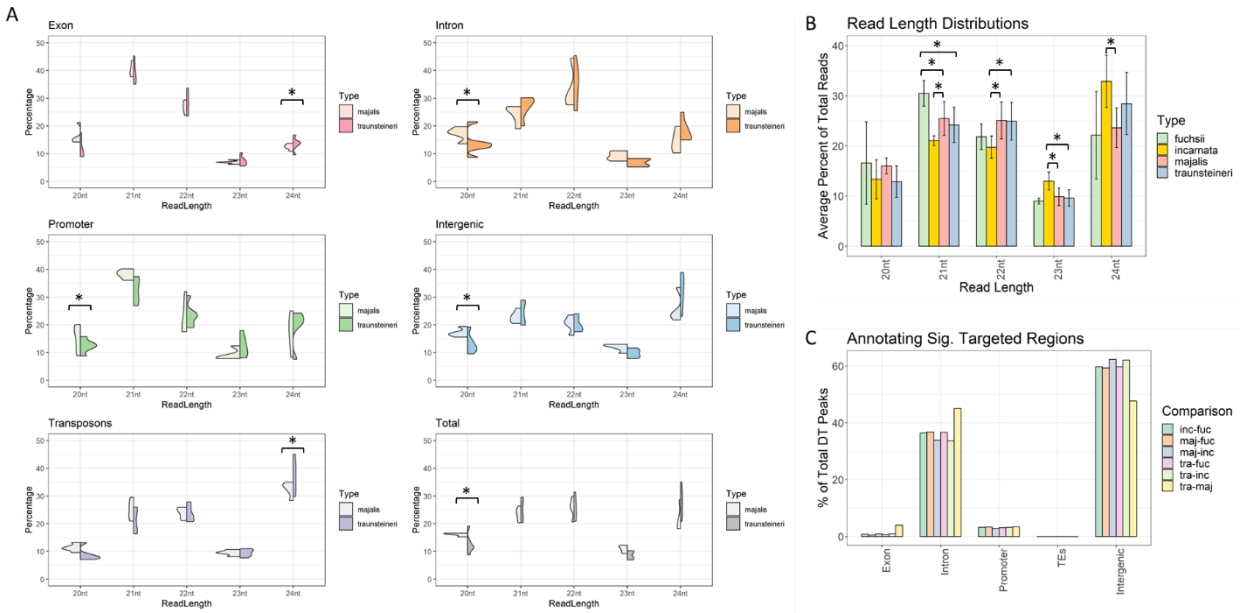
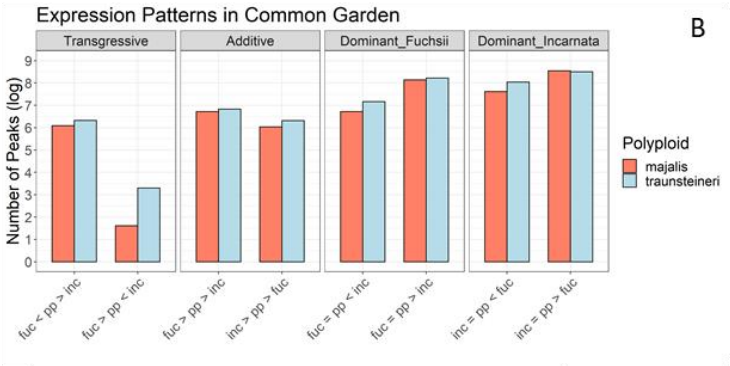


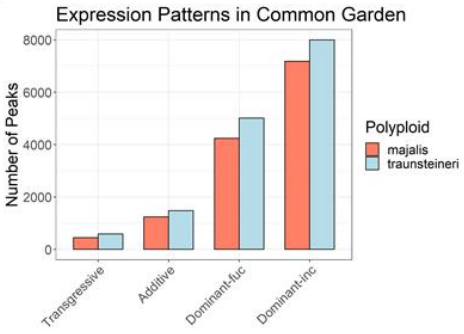
Figure 3. A) Split bean plots showing read length distributions across genomic features in the common garden polyploids. The y-axis depicts the percentage of total reads mapping to the feature annotation and the x-axis is the read length. Density distributions are shown for each polyploid at each read length to visualize how closely the biological replicates share mapping patterns and significance values were calculated with the Wilcoxon signed-rank test (*p<0.05). **B)** Read length mapping patterns across the whole reference genome for polyploids and diploids with error bars representing standard deviation. **C)** Genome annotations for each significantly DT region between common garden polyploids, diploids, and each polyploid-diploid comparison. Significantly over- and under-targeted regions were combined and mapped to each reference annotation. The peak numbers per feature then divided by the total number of DT regions for each comparison to normalize by showing percentages of total DT regions per feature.

In addition to comparisons among diploids and polyploids, targeting patterns between the polyploids and diploid progenitors were analyzed to investigate dominance patterns. These patterns were defined by the following: transgressive regions are defined as regions which are significantly over- or under-targeted in the polyploid compared to both diploid progenitors. Additive regions are regions that are intermediate between the diploids with an over-targeted pattern in reference to one diploid, and under-targeted pattern to the other. Dominant regions are defined by regions which are conserved (non-significant differences) in reference to one diploid, while over- or under-targeted in reference to the other.

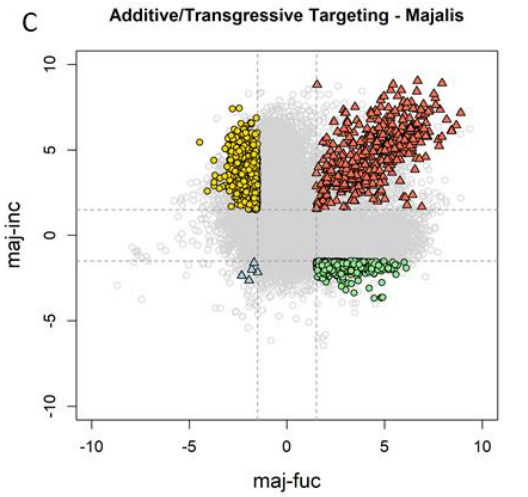
A



B

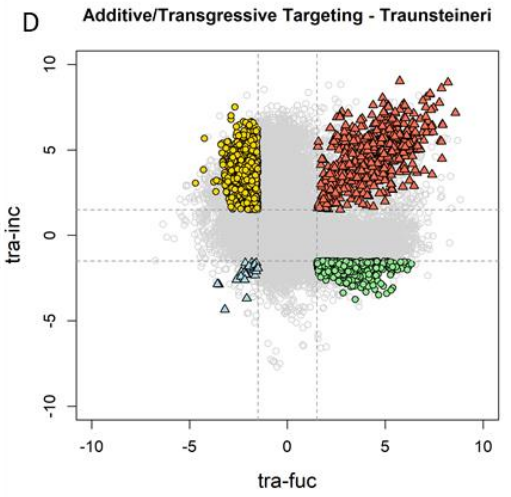


C



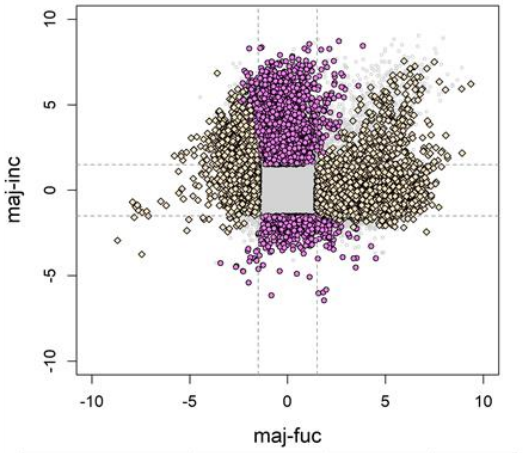
Type	Pattern	Color	Peak #
Additive	fuc > maj > inc	Yellow	827
Additive	inc > maj > fuc	Green	417
Transgressive Down	fuc > maj < inc	Blue	5
Transgressive Up	fuc < maj > inc	Red	442

D



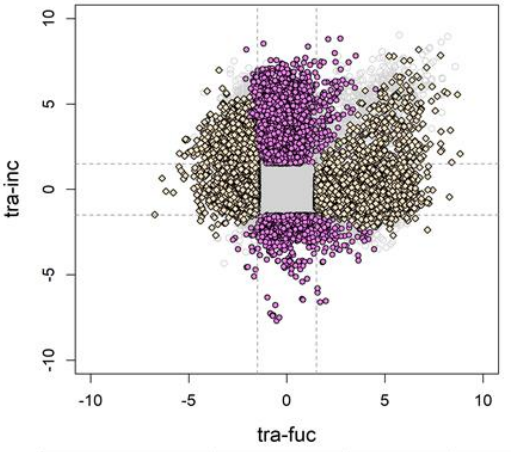
Type	Pattern	Color	Peak #
Additive	fuc > tra > inc	Yellow	925
Additive	inc > tra > fuc	Green	551
Transgressive Down	fuc > tra < inc	Blue	27
Transgressive Up	fuc < tra > inc	Red	560

E



Type	Pattern	Color	Peak #
Dominant Fuchsii	fuc = maj > inc	Pink	3415
Dominant Fuchsii	fuc = maj < inc	Pink	828
Dominant Incarnata	inc = maj > fuc	Tan	5143
Dominant Incarnata	inc = maj < fuc	Tan	2037

F



Type	Pattern	Color	Peak #
Dominant Fuchsii	fuc = tra > inc	Pink	3727
Dominant Fuchsii	fuc = tra < inc	Pink	1288
Dominant Incarnata	inc = tra > fuc	Tan	4909
Dominant Incarnata	inc =traj < fuc	Tan	3088

Figure 4. A) Targeting patterns for *D. majalis* (red) and *D. traunsteineri* (blue) in reference to the diploid progenitors. The x-axis represents the comparison being made (pp = polyploid; represented by color) and the number of regions (log transformed) that are defined by this comparison are shown on the y-axis. **B)** Total peaks for each inheritance pattern category. **C-D)** Targeting patterns between *D. majalis* and *D. traunsteineri* and their diploid progenitors. The y-axis represents the logFC of regions in the polyploid in reference to *D. incarnata* while the x-axis is the logFC of the same regions in reference to *D. fuchsii.* Red and blue triangles are regions that are transgressive up and down respectively. Yellow circles represent regions that are under targeted compared to *D. fuchsii* but over targeted compared to *D. incarnata,* and the green circles represent the opposite (additive targeting). **E-F)** Dominant targeting patterns between *D. majalis* and *D. traunsteineri* and the diploid progenitors. Pink circles are regions that are *D. fuchsii* dominant. Tan diamonds represent dominant *D. incarnata* targeting patterns.

Figure 4 summarizes the number of regions in each inheritance category. Dominant targeting patterns were most prevalent from *D. incarnata* with 10.9% of all peaks retaining a *D. incarnata* dominant pattern in both polyploids, and 5.7% qualifying as *D. fuchsii* dominant. Additive peaks shared between polyploids accounted for 1.9% of the total with the majority under targeted in respect to *D. fuchsii* and over-targeted with respect to *D. incarnata* (1.3%, 0.6% respectively). The polyploids shared more conserved (non-significant) regions with *D. incarnata* than with *D. fuchsii* (69.3%, 57.5% respectively) and 42.3% of all peaks included in the analysis were conserved among the total diploid-polyploid system. Of the peaks in each category, an average of 39% (±3.1% SD) were conserved between the polyploids. Transgressive peaks accounted for a small fraction of analyzed regions (0.7%) but were consistently characterized by increased smRNA regulation (transgressive up) in polyploids compared to either diploid progenitor. Of the peaks that were transgressive in both *D. majalis* and *D. traunsteineri* a near majority were under increased regulation from smRNAs in the polyploids (295 up, 2 down). These transgressive peaks were run through the same GO enrichment pipeline as previously mentioned. Intron and promoter gene sets did not yield any significantly enriched GO terms, but genes targeted across exons were significantly enriched for terms dealing with auxin response and signaling pathways (Fig. 5).

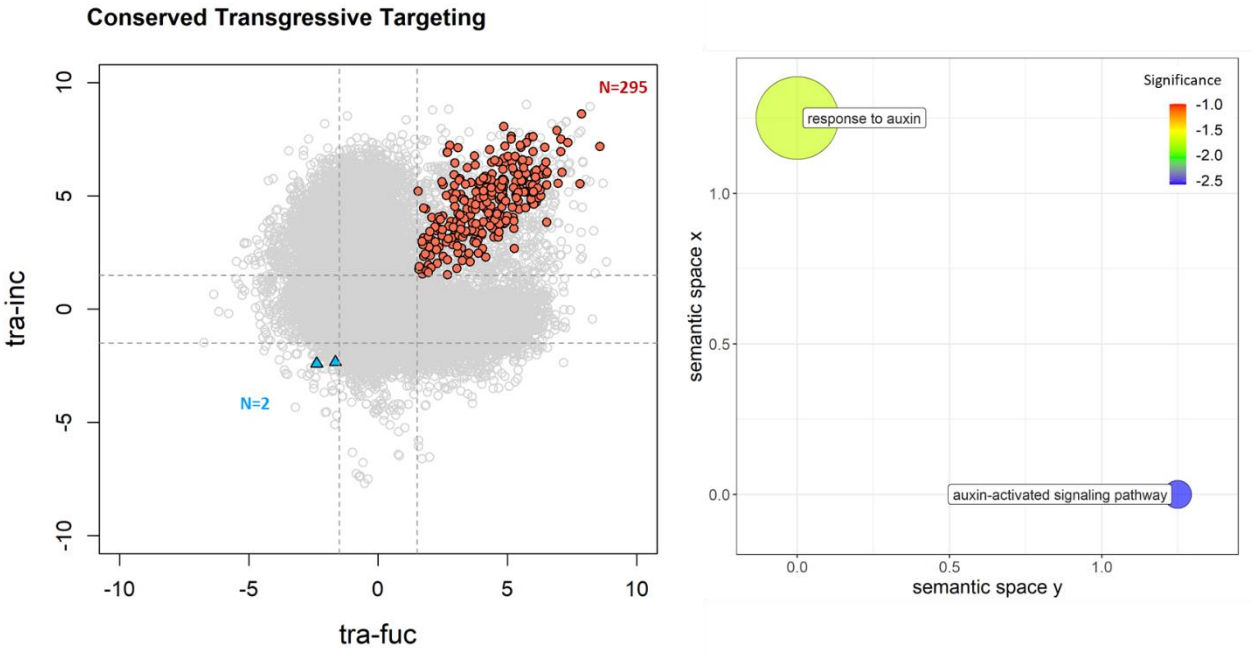


Figure 5. Left) Conserved transgressive (down - blue, up - red) regions in both *D. majalis* and *D. traunsteineri.* Shown are logFC values for *D. traunsteineri.* **Right)** Enriched GO terms in the transgressively targeted exon regions. GO enrichment were performed with Blast2GO (Fisher's exact test, FDR < 0.05), and results visualized with REVIGO.

3.2 Reciprocal Transplants – Leaf Tissue

Following size selection an average of 6,485,140 reads were retained (range: 2,499,121 - 10,178,526) of which 85.3% (range: 76.7% – 88.8%) were successfully mapped to the reference. Mismatch rates did not vary significantly (0.63, ±0.8 standard deviation, Fig. S2). Mapped reads predominantly aligned to multiple loci at 60.7% on average, and 24.5% mapped uniquely. Following feature counts, samples were filtered for low targeting regions retaining 4.6% (102,094) of peaks, and visualized on an MDS plot (Fig. 6)

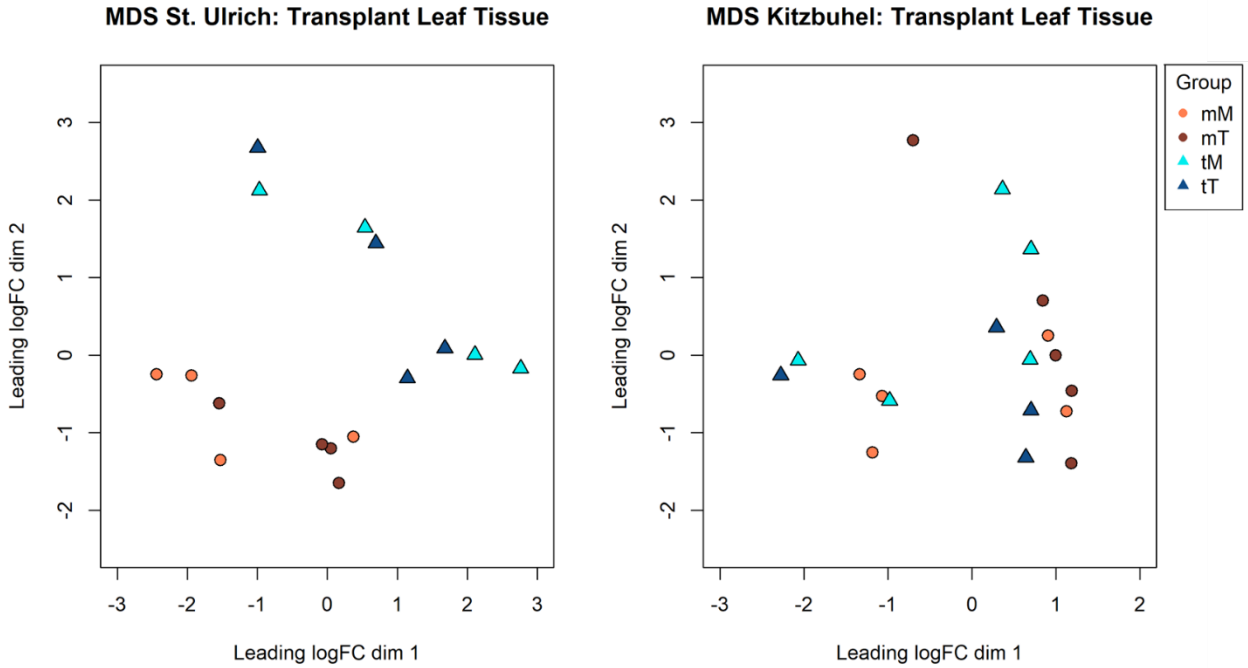


Figure 6. MDS plots for leaf tissure in transplanted samples shown separately for both tissue types and localities. Color shade represents environment with the lighter colors (coral, cyan) depicting samples grown in *D. majalis* environment (M) and darker colors (red, blue) for samples grown in *D. traunsteineri* environment (T). Shape represents the polyploid with triangles for *D. traunsteineri* (t) and circles for *D. majalis* (m).

Following the differential targeting analysis, significantly DT regions were plotted (Fig. 7A & C). A barplot shows the overall trends between over- and under-targeted regions allowing visual comparsion between the localities. The MA plots look further accounting for the overall log fold change distributions in reference to the average targeting level of all regions in the analysis, while highlighting the significantly DT regions. Generally, highly targeted regions had lower variation between polyploids agreeing with trends noted in the leaf tissue of common garden samples. The variation between biological replicates was similar in transplanted leaf samples in respect to the common garden (BCV=0.26, 0.27 in transplants and common garden repectively).

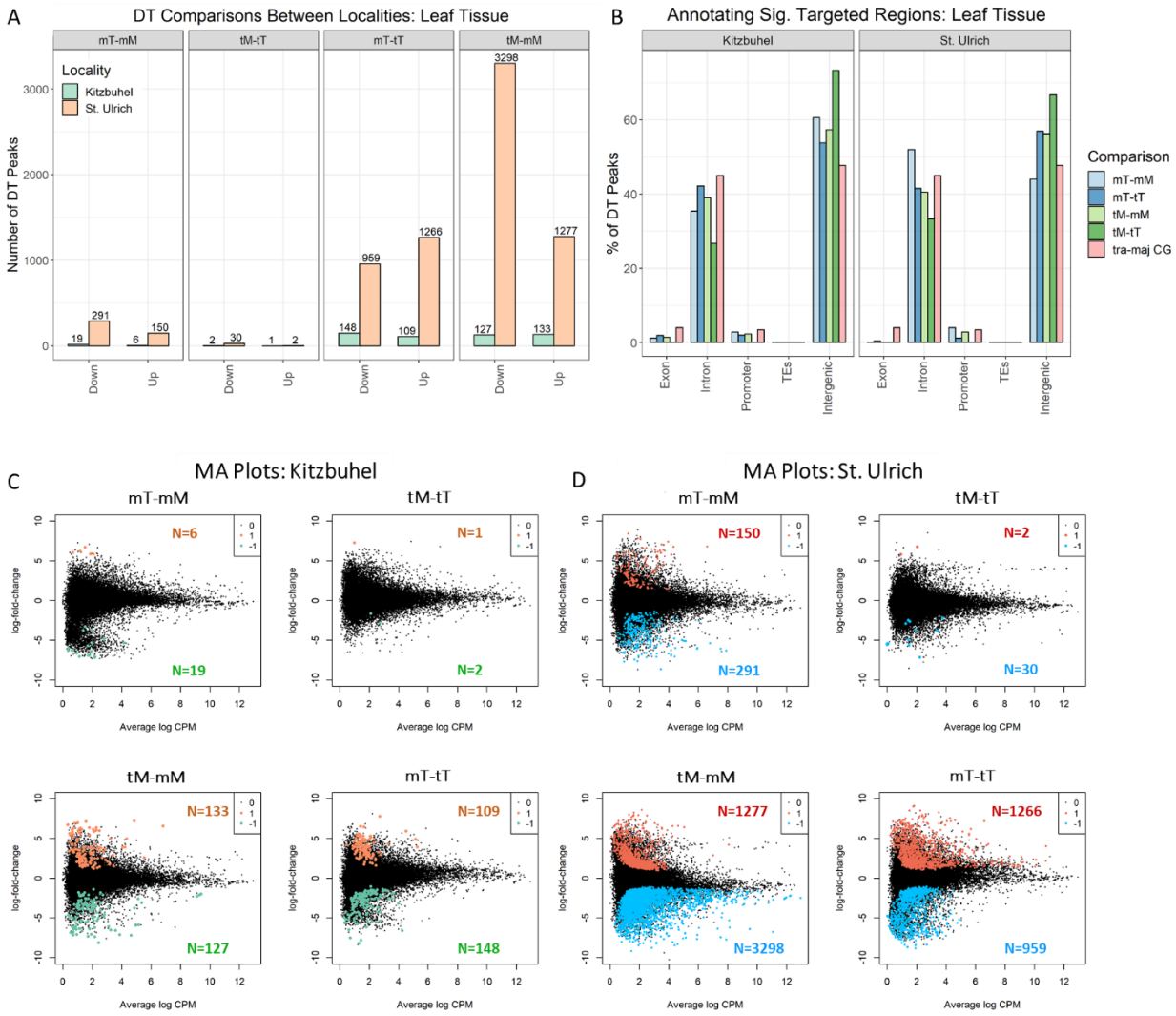
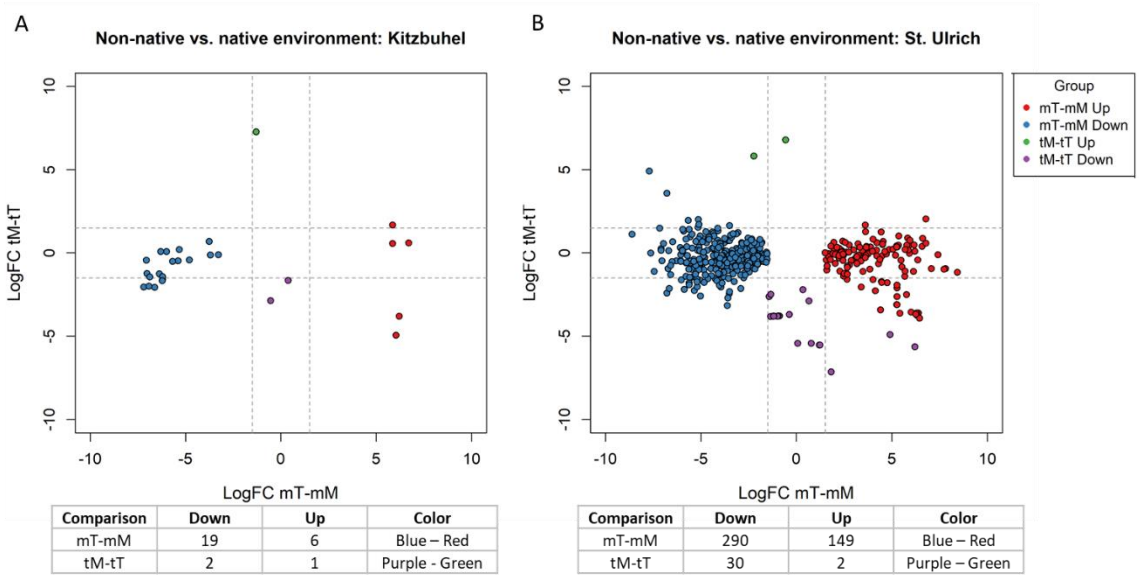


Figure 7. A) DT peak counts in leaf tissue samples. Color indicates sample origin locality, and the plot is faceted by comparison. Peak numbers are always shown for the polyploid in the non-native environment with the naming scheme; m = *D. majalis* organism, M = *D. majalis* environment, t = *D. traunsteineri* organism, T = *D. traunsteineri* environment**. B)** Genome annotations for each significantly DT region in leaf tissue transplanted accessions (blue, green) with the sample comparison from the common garden for reference (salmon). The y-axis represents the percentage of sig DT regions for each genomic feature shown on the x-axis. **C-D)** MA plots showing the log fold change on the y-axis against the average targeting level (abundance, normalized by log cpm) on the x-axis. In Kitzbuhel, orange points are significantly over-targeted regions and green points are significantly under-targeted. In St. Ulrich, red points represent significantly over-targeted regions and blue represent significantly under-targeted. Significance was called at FDR<0.05, logFC ±1.2.

Significantly DT regions were intersected between comparisons to identify any conserved regulatory patterns in response to non-native environmental cues. For example, significantly DT regions between *D. majalis* accessions grown in its native environment and those grown in *D. traunsteineri*'s

environment were interesected with significantly DT regions in *D. traunsteineri* in native and non-native environments. Regions that were shared in both comparisons were deemed a conserved plastic reponse. In leaf tissue such conservation was not observed in polyploids in either locality. To further visualize patterns in the targeting levels between the transplants, logFC values from significantly DT regions were plotted for each polyploid-envirnoment comparsion (Fig. 8). Regions were interesected to search for a signal of conservation between the polyploids but all regions were targeted uniquely for each polyploid.



Comparison	Down	Up	Color
mT-mM	19	6	Blue – Red
tM-tT	2	1	Purple - Green

Comparison	Down	Up	Color
mT-mM	290	149	Blue – Red
tM-tT	30	2	Purple – Green

Comparison	Down	Up	Color
tM-mM	126	127	Blue – Red
mT-tT	142	109	Purple – Green

Comparison	Down	Up	Color
tM-mM	3183	1085	Blue – Red
mT-tT	890	1167	Purple – Green

Figure 8. Targeting patterns in leaf tissue among transplanted accessions. Dashed lines represent a logFC of ±1.5 and plots are faceted by comparison and sample origin locality. **A-B)** Comparing the same polyploid in different environments. The y-axis shows logFC of significant DT regions between *D. traunsteineri* grown in *D. majalis* environment (tM), and its native environment (tT); while the x-axis shows logFC values for *D. majalis* grown in *D. traunsteineri* environment (mT) and its native environment (mM). **C-D)** Comparing different polyploids in the same environment. The y-axis shows logFC of significant DT regions between *D. majalis* grown in *D. traunsteineri* environment (mT), and *D. traunsteineri* in its native environment (tT); while the x-axis shows logFC values for D. traunsteineri grown in *D. majalis* environment (tM) and *its D. majalis* in its native environment (mM).

3.3 Reciprocal Transplants – Root Tissue

After size selection an average of 5,497,691 reads were retained (range 2,322,556 – 8,408,971) of which 77.8% (range: 65.7% – 84.6%) were successfully mapped to the reference, which was lower than leaf tissue samples, though this could be due to the potential increased complexity of root tissue samples containing, for example, symbiont mycorrhiza. The average mapping mismatch rate was 0.60 (0.08 standard deviation, Fig. S2). Like with leaf tissue, reads mapped predominantly to multiple loci at 56.7% on average (SD=4.4), and 27.2% mapped uniquely (SD=2.6). Filtering for low targeting regions retained 5.5% (133,692) of peaks, and samples were subsequently visualized on an MDS plot (Fig. 9).

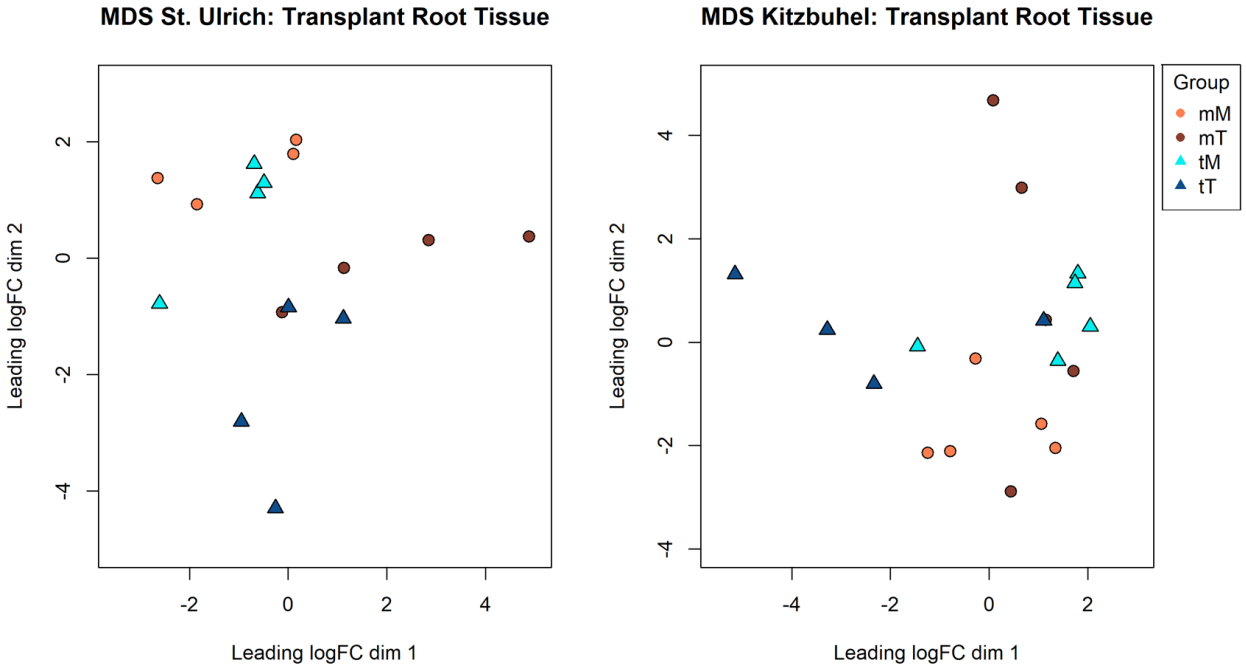


Figure 9. MDS plots for root tissue of transplanted samples shown separately for both sample origin localities. Color shade represents environment with the lighter colors (coral, cyan) depicting samples grown in *D. majalis* environment (M) and darker colors (red, blue) for samples grown in *D. traunsteineri* environment (T). Shape represents the polyploid with triangles for *D. traunsteineri* (t) and circles for *D. majalis* (m).

Further analyses followed the workflow described previously for leaf tissue. A reversal of the trend noticed in leaf tissue was evident in roots with Kitzbuhel accessions having greater regulatory divergences than St. Ulrich accessions in all comparisons (Fig. 10A). DT regions again predominantly targeted intron and intergenic regions in all transplant comparisons accounting for >95% of all significantly DT regions which followed trends noted in leaf tissue and common garden samples. Variation between biological replicates was high in roots (BCV=0.40) and targeting level variation in significant regions increased in more highly targeted regions as can be seen in the MA plots in figure 10C-D. Despite this, much higher numbers of DT regions were still detected in all but one comparisons in reference to the leaf tissue (tM-mM - St. Ulrich: 4.2% DT in leaf tissue, 1.4% DT in root tissue).

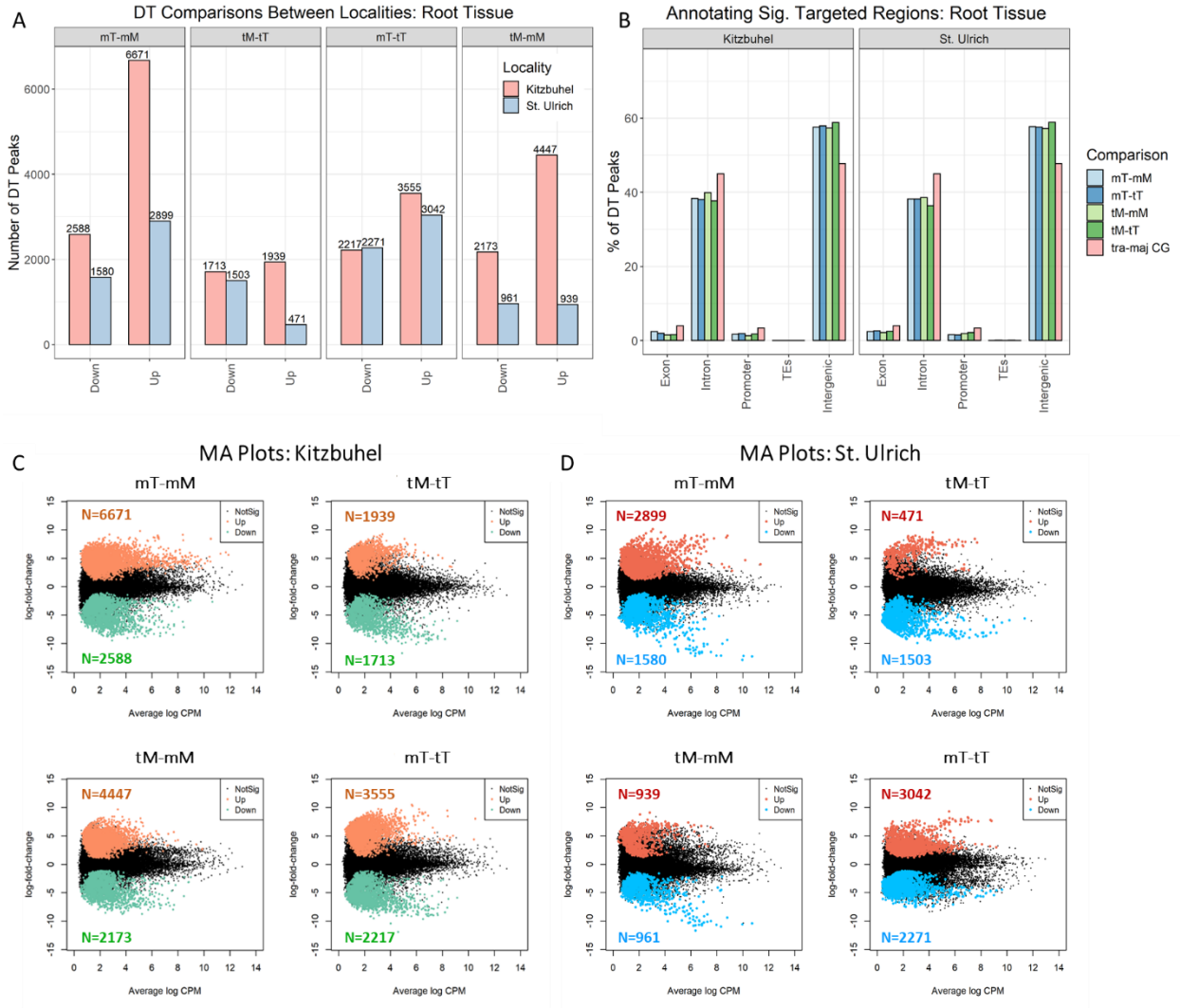
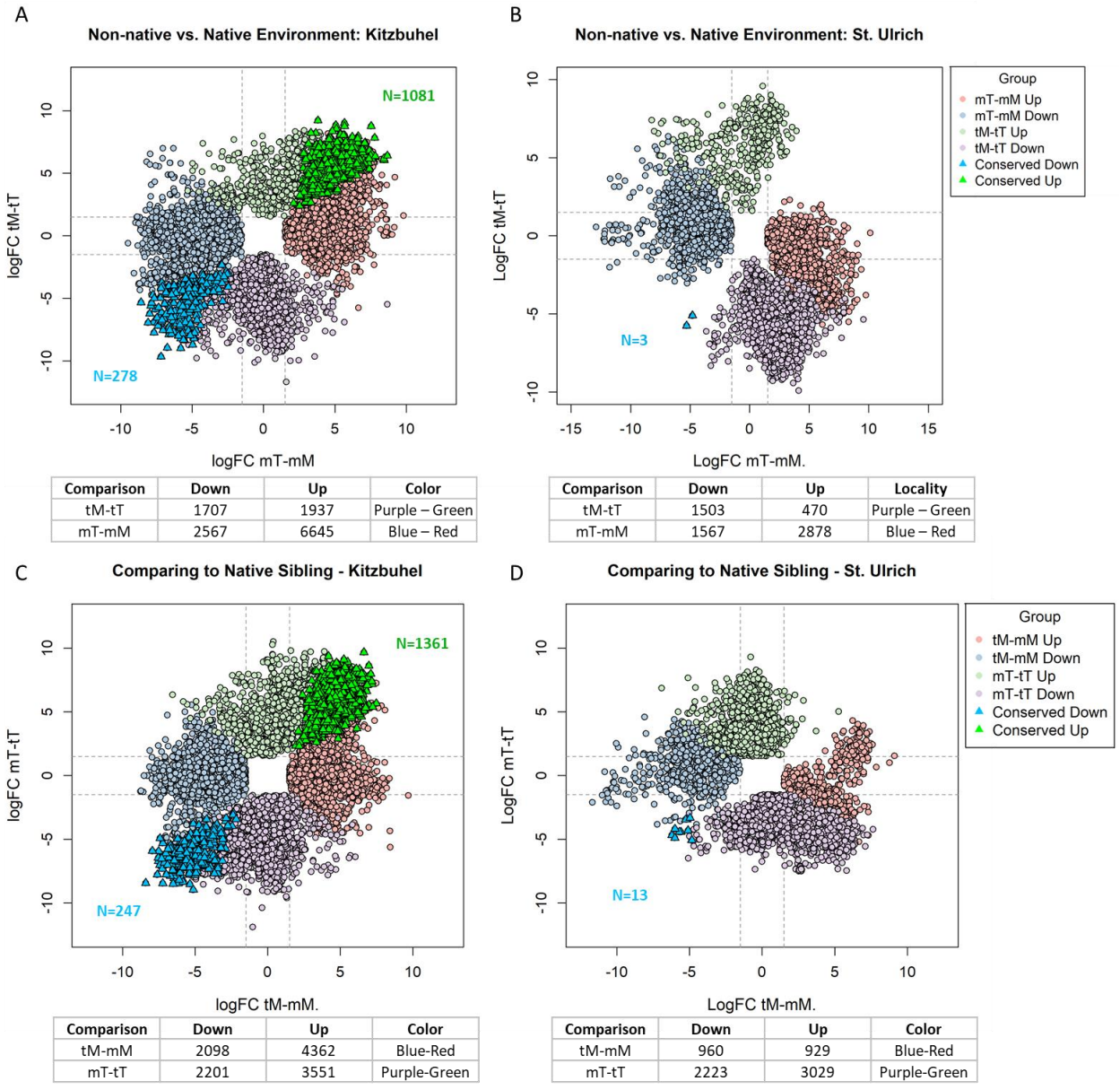


Figure 10. A) DT regions in root tissue samples. Color indicates sample origin locality, and the plot is faceted by comparison. Peak numbers are always shown for the polyploid in the non-native environment with the naming scheme; m = *D. majalis* organism, M = *D. majalis* environment, t = *D. traunsteineri* organism, T = *D. traunsteineri* environment. **B)** Genome annotations for each significantly DT region in root tissue transplanted accessions (blue, green) with the sample comparison from the common garden for reference (salmon). The y-axis represents the percentage of signifantly DT regions for each genomic feature shown on the x-axis. Plot is faceted by sample origin locality. **C-D)** MA plots showing the log fold change on the y-axis against the average targeting level (abundance, normalized by log cpm) on the x-axis. Plots are separated for each comparison in each locality. In Kitzbuhel, orange points are significantly over-targeted regions and green points are significantly under-targeted. In St. Ulrich, red points represent significantly over-targeted regions and blue represent significantly under-targeted. Significance was called at FDR<0.05, logFC ±1.2.

Signals of conserved targeting patterns were prevelant in Kiztbuhel. Interestingly, a greater regulatory conservation was noted between different polyploids in the same environment (15.2% of all

DT peaks) than the same polyploid in different environments (11.8%) in Kitzbuhel accessions. St. Ulrich accessions had very little conservation in either comparison (Fig. 11).



Comparison	Down	Up	Color
tM-tT	1707	1937	Purple – Green
mT-mM	2567	6645	Blue – Red

Comparison	Down	Up	Locality
tM-tT	1503	470	Purple – Green
mT-mM	1567	2878	Blue – Red

Comparison	Down	Up	Color
tM-mM	2098	4362	Blue-Red
mT-tT	2201	3551	Purple-Green

Comparison	Down	Up	Color
tM-mM	960	929	Blue-Red
mT-tT	2223	3029	Purple-Green

Figure 11. Targeting patterns in root tissue. Dashed lines represent a logFC of ±1.5 and plots are faceted by sample origin locality. Pastel circles represent the significant regions for each comparison listed while the brighter triangles (blue, green) represent significant regions which are conserved between both polyploids indicating conserved adaptive targeting patterns. **A-B)** The y-axis represents logFC of significant DT regions between *D. traunsteineri* grown in *D. majalis* environment (tM), and its native environment (tT); while the x-axis represents logFC values for *D. majalis* grown in *D. traunsteineri* environment (mT) and its native environment (mM). **C-D)** The y-axis represents logFC of significant DT regions between *D. majalis* grown in *D. traunsteineri* environment (mT), and *D. traunsteineri* in its native environment (tT); while the x-axis represents logFC values for *D. traunsteineri* grown in *D. majalis* environment (tM) and its *D. majalis* in its native environment (mM).

Read lengths distributions were averaged for each group of leaf and root tissue samples and summarized in figure 12. In leaves, *D. majalis* accessions had higher average proportions 20-22nt reads while *D. traunsteineri* had higher average proportions on 24nt reads (significance was not tested). In roots, distributions are similar across all comparisons thought the overall profile differs from leaves with higher proportions of 23nt reads and less 22nt and 24nt reads.

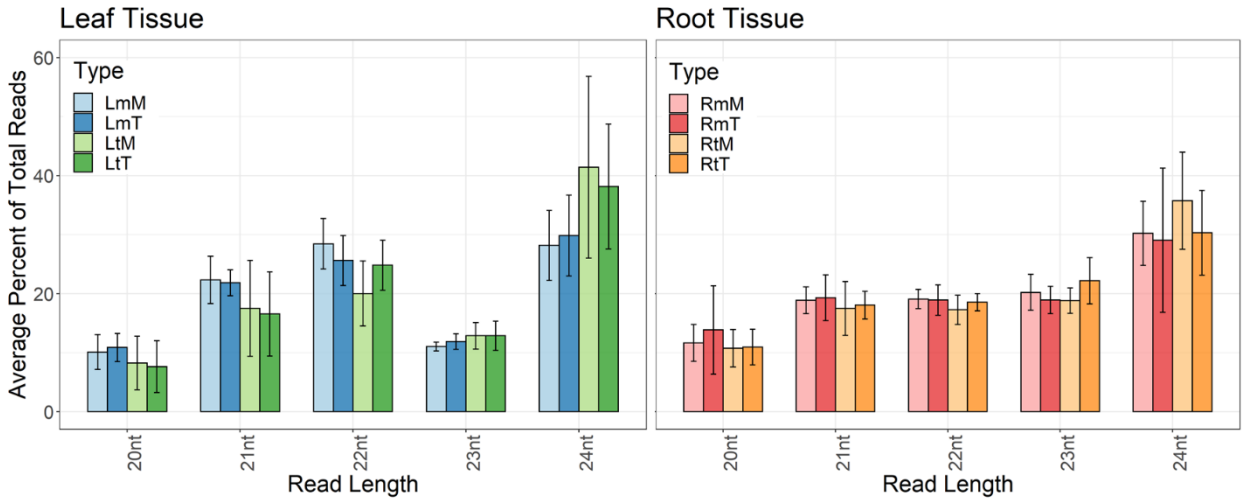


Figure 12. Read length distribution plots for transplanted accessions. Plots are faceted by tissue type, and color indicates polyploid-environment setup with darker colors representing the polyploids in the non-native environment. Error bars represent standard deviation.

4. Discussion

The *Dactylorhiza* polyploid-diploid system used in this study provides an excellent model to study how sibling allopolyploids, arising from recurrent and directional hybridization of genomes with distinct demographics histories, diversify and maintain unique morphologies and ecological niches. There are many complex molecular mechanisms underlying the formation and ecological differentiation of allopolyploids and this study endeavored to provide a comprehensive overview of how smRNA-mediated differential regulation plays a role in this system.

4.1 Polyploids share a common smRNA regulatory system

Common garden experiments, in combination with advances in genomic analyses, strive to uncover the genetic components of divergence between organisms. The goal is to remove sources of environmental variation and study what remains; in this case to see if any regulatory differences can still be detected. It is evident from our data that the sibling allopolyploids largely share an underlying smRNA regulatory system. Evidence of this can be found in the count-based dimension reduction analysis where no clear clustering can be observed between the polyploids suggesting minimal variation in smRNA targeting patterns (Fig. 1C), and in the results of the differential targeting analysis showing merely 0.4% of peaks under differential regulation (Fig. 1B). The lack of regulatory differences may stem in part from the formation of the respective polyploids via a consistent inheritance of parental regulatory patterns, with both siblings displaying similar patterns in all categories. We find that 40.4% of additive peaks were conserved between the siblings while an average of 38.5% of non-additive peaks were conserved. Both polyploids showed more regulatory similarity with the larger, paternal *D. incarnata* genome reflected in the higher levels of dominant-incarnata targeting patterns, which are conserved between the polyploids and *D. incarnata* while diverging from *D. fuchsii.* Higher levels of *D. incarnata* dominance is found also in gene expression patterns from RNAseq data of the same system (Wolfe, 2019). *D. incarnata* is shown to have a higher TE load than *D. fuchsii* (Balao et al., 2017, Eriksson, 2022) which is in line with the notion that a submissive subgenome in a hybrid more likely contains a higher percentage of TEs (Edgar, 2017). This result could be biased as the only available reference genome in this study was of *D. incarnata* making it difficult to deconvolute true progenitor-specific signals, though mapping rates were higher on average in *D. fuchsii* suggesting that mapping bias is not driving the observed trends. The observed similarity could also be reinforced by gene flow between the siblings, especially in accessions from the Alps, where the siblings grow in sympatry and have demonstrated gene flow (Balao et al., 2016, Brandrud et al., 2020). Epigenetic divergences have

been found to largely explain environmental differences between the polyploids (Paun et al., 2010), which, in combination with our results from the common garden, indicate a dominant role in environmental selective pressure, rather than inherited genetic differences, in the observed differentiation. Perhaps supplementary analyses excluding highly sympatric regions could be more sensitive in identifying regulatory regions associated with the observed differences in the polyploids by investigating more ecologically distinct populations.

Profiles of smRNA classes are also largely conserved between the siblings with no significant differences in the proportions of 21-23nt smRNA classes mapping to genomic feature annotations. 20nt reads, likely consisting of miRNAs differed significantly in intron, promoter, and intergenic regions while 24nt reads differed in exon and MITE-like transposon regions (Fig. 2A). In comparison to the diploid parents there is evidence for differential abundances of post-transcriptionally regulating smRNAs (namely miRNAs; 20-22nt). *D. majalis* differed in 21nt read distributions with a larger proportion than the paternal *D. incarnata* and a smaller proportion than maternal *D. fuchsii* and the same trend was observed in 21nt and 22nt reads in *D. traunsteineri*. Mapping of the significantly DT regions to genome annotations revealed a pattern dominated by intron and intergenic targeting in all comparisons from the common garden system (Fig. 2C, >92%). Exons and promoter regions had similar significant targeting differences but across all comparisons no significant DT regions mapped to TEs. It must be noted that only MITE-like TE annotations were included in this study and therefore this is not representative of the full population of TEs in the genomes. A more thorough analysis of TE regulation by smRNA showed high levels of TE targeting, though these patterns were largely conserved for 24nt targeting between *D. majalis* and *D. traunsteineri.* Comparisons between the polyploids had higher levels of gene targeting than diploid or polyploid-diploid comparisons suggesting an increase in gene silencing potentially via post-transcriptional miRNA-mediated transcript degradation.

Despite the overarching similarities, there is evidence of differential regulation of developmental and environmental gene targets. In *D. traunsteineri* a geographic signal can be observed in the British accessions in both the overall count data (Fig. 1C) and in the significantly DT regions (Fig. 1D). Populations in the British Isles are found to have slightly different genetic signatures and morphologies which is also reflected in the regulatory patterns shown here. Most GO terms enriched in the genes under differential regulation concerned cell proliferation and development in both pre- and post-transcriptional groups (Fig. 2). These results agree with numerous reports of the role of smRNA in regulating development and metabolic activity in plants (Bonnet et al., 2006; Guan et al., 2014; Li et al., 2017). Of note are the GO term enrichments in the serine/threonine kinase target of rapamycin (TOR) signaling, response to amino acids, and autophagy regulation in DT genes under post-transcriptional regulation. The TOR signaling pathway is an important link between environmental stimulus and developmental consequences in plants (Xiong & Sheen, 2015; Dobrenel et al., 2016; reviewed in Shi et al., 2018). TOR signaling in plants has known involvement in modulating metabolism, growth and development, energy production, and nutrient recycling via autophagy (Pu et al., 2017) in response to nutrient profiles (Deprost et al., 2007; Dong et al., 2017), amino acids (Xiong & Sheen, 2015), and photosynthetic input (Xiong et al., 2013). Interestingly, photosynthesis-derived glucose was shown to drive TOR-signaling cascades which can reprogram transcriptional profiles of genes involved in a wide range of metabolic, regulatory, and developmental processes (Xiong et al., 2013). A key difference in the gene expression profiles of *D. majalis* and *D. traunsteineri* involved genes concerning photosynthesis pathways and hardware, and significant differences were also found in chlorophyll amounts and the excited-electron input to photosystem II that can be used for glucose production, among other photosynthetic related variables (Wolfe et al., 2021). One possibility is that changes in photosynthetic energy production between the polyploids is linked to the observed divergent regulation of TOR signaling cascades. Furthermore, since the polyploids are also known to differ in soil chemistry

preference, we hypothesize that metabolic regulation would diverge in response to the nutrient profiles of the local soils. Together, it seems possible, though currently highly speculative, that a divergence in the TOR-mediated regulatory networks that control metabolism and growth could play a role in maintaining different environmental preferences of the polyploids. TOR signaling is not the only pathway involved in metabolic and developmental regulation, but it has very important roles in linking these processes to environmental stimuli and is an interesting candidate to keep in mind in subsequent analyses. It is also important to note that the number of DT regions, and therefore the reporting of genes affected by DT is dependent on the data filtering steps performed before DT analysis. By setting less stringent thresholds, more peaks are allowed in the analysis and thus there is room for more significantly DT peaks to be reported, affecting such downstream analyses as GO enrichments. This should be taken into consideration concerning the reproducibility of results with different data filtering methods or thresholds.

4.2 Auxin related pathways are transgressively regulated in the polyploids

Given the potential for neofunctionalization in polyploids, a divergent regulation of developmentally or ecologically relevant genes could facilitate the expansion of polyploids outside of their parental range through transgressive gene regulation or expression. For example, in *Arabidopsis* it was shown that non-additive expression of miRNA and tasi-RNA lead to morphological variation between polyploid and progenitor (Ng et al., 2012). The majority of transgressively DT regions in this study were over-targeted (transgressive up) in the polyploids compared to the diploids. An RNAseq study with the same system found that in transgressively expressed genes, a consistent trend is observed with more genes under expressed (transgressive down) with respect to the diploids (Wolfe, 2021), suggesting that smRNA-mediated gene silencing could be playing a role here, but definitive conclusions cannot be drawn until direct comparisons of the DE genes and DT regions are made. In our system transgressive regulation was exclusively associated with auxin signaling and response (Fig. 5).

Auxin is a major player in plant development and works in concert with other regulatory hormones to control cell differentiation and division (Aloni et al., 2006; Zhang et al., 2013). It regulates response to environmental cues such as light and gravity, developmental structuring of roots and shoots, and is under significant pre- and post-transcriptional regulation (reviewed in Schepetilnikov & Ryabova, 2017). Studies have demonstrated the role of auxin response factor (ARF) regulation via tasi- and miRNAs in response to various environmental stressors in both leaf and root tissue (Tang et al., 2012; Sun et al., 2017) suggesting that modulation of auxin responses could provide an avenue for ecological differentiation in the polyploids. The smRNA data shown in this study could be supplemented by identifying if any auxin-associated genes under differential smRNA regulation are also differentially expressed to provide more conclusive support that the DT regions are leading to true expression level.

4.3 The polyploids exhibit differential regulatory plasticity in response to environment

As previously described, one of the major roles of smRNAs is mounting responses to various biotic and abiotic stressors (environmental stress). The ability of a neopolyploid to respond plastically (ie. by adapting its regulatory patterns) to new environments will increase its chances of establishing itself in novel ranges, perhaps one outside of the parental range. As *D. majalis* occupies a more competitive and wider ecological niche (Balao et al., 2017), we hypothesize that there must be an increased capacity to adapt smRNA regulatory patterns in response to environmental stress (ie. a more plastic smRNA response) compared to the more environmentally restricted *D. traunsteineri.* To investigate the effects of environment on smRNA activity, the reciprocal transplants provided an excellent system in which we investigated smRNA targeting patterns between accessions grown in their native environment and those transplanted to a non-native environment. The data presented here indicates that *D. majalis* consistently exhibits more plasticity in its regulatory patterns and that these patterns are influenced by the location of origin of the population studied, and tissue type. In leaf tissue

samples, the difference between localities is clear. Samples from St. Ulrich cluster by polyploid, though clear environmental signals are not evident, whereas in Kitzbuhel there is little to no clustering of samples by either variable (Fig. 6). Furthermore, when referencing the number of DT peaks between the same polyploid in different environments there are 17.6X more DT peaks in *D. majalis* from St. Ulrich than Kitzbuhel and 10.6X more for *D. traunsteineri* (Fig. 7 & Fig. 8A-B) showing a clear effect of sample origin on the regulatory patterns of the polyploids. *D. majalis* had 8.3X and 13.7X more significantly DT peaks between native and non-native accessions compared to *D. traunsteineri* (in Kitzbuhel and St. Ulrich respectively), and, in the case of Kitzbuhel accessions, *D. traunsteineri* had only three total DT regions suggesting a very limited ability to modulate regulatory responses in new environments. Non-native accessions were predominately under-targeted by smRNAs in all cases indicating a relaxation of smRNA-mediated gene silencing following exposure to new environments. GO enrichments of genes with DT regions between native and non-native *D. majalis* accessions were predominately associated with the regulation of metabolic processes, specifically with the catabolism and metabolism of mRNA; and gene regulation (including RNA stability, post-transcription gene regulation, and translation; Fig. S3). Due to the small number of DT regions in *D. traunsteineri* there were no enriched GO terms in genes under DT.

Additionally, the two polyploids were compared within a native environment to understand if the non-native polyploid could adopt similar targeting profiles as the native, and to compare with common garden results to infer the environmental effects. Both comparisons (mT-tT and tM-mM) had much higher DT levels than observed in the common garden, suggesting the native environmental conditions are indeed playing a large role in the observed regulatory divergences (Fig. 7 & Fig. 8C-D). Again, in Kitzbuhel accessions this effect is less pronounced with both comparisons (tM-mM and mT-tT) having only 1.6X more DT peaks than in the common garden, but in St. Ulrich this jumps to 27X and 13X more respectively. *Dactylorhiza majalis* accessions from St. Ulrich grown in *D. traunsteineri*

environment (mT-tT) had less than half the number of significant differences with the native *D. traunsteineri* than the non-native *D. traunsteineri* with native *D. majalis* (tM-mM). This suggests the ability of *D. majalis* to better adopt its regulatory patterns with that of the native polyploid, though differences are still far greater than when environmental variables are controlled for. However, this effect was not observed in root tissue for St. Ulrich accessions where the non-native *D. traunsteineri* showed more similarity with the native *D. majalis* (Fig. 11C-D). In many comparisons there emerges a clear signal of tissue dependent results requiring caution when drawing conclusions based on single tissue types.

4.4 smRNA activity is tissue dependent

smRNA activity has been shown to differ in different tissues depending on environmental conditions (Chuck & O'Connor 2010; Candar-Cakir et al., 2016). Our data reveal an effect of tissue type on smRNA activity, differential targeting, and class distributions across genomic features. Interestingly, when observing MDS plots for leaf and root tissue samples in St. Ulrich accessions, there is a change in the clustering patterns from polyploid type in leaf tissue (Fig. 6), to environmental setup in root tissue (Fig. 9) indicating a shift in regulatory patterns in roots corresponding to the environmental stimuli, rather than genetic identity. We also observe a large-scale difference in the amount of DT regions in the same comparisons with higher DT levels in root compared to leaf tissue (Fig. 7,9; Fig. 10,11). The most drastic differences were in *D. traunsteineri* in native and non-native environments (tM-tT) with root tissue having 1,217X and 62X more DT regions than leaf tissue samples in Kitzbuhel and St. Ulrich respectively. *D. majalis* also had much higher DT levels in roots with 370X (Kitzbuhel) and 10X (St. Ulrich) increases compared to leaf tissue samples. Despite the different effect size, the non-native polyploids were consistently more under-targeted in each comparison, as seen with leaf data. As roots provide the interface for nutrient absorption their structure is modulated by nutrient deficiency-induced

signaling pathways which may alter expression patterns of various transporter and metabolic genes. Small RNAs are involved in such responses to deficiencies in nitrogen (Yang et al., 2019), phosphorous and sulfur (reviewed in Kumar et al., 2017), and various micronutrients (Abdel-Ghany & Pilon, 2008; Kong & Yang, 2010). The observed increase in differential smRNA activity in roots is not surprising when considering a transplantation of polyploids with different soil nutrient preferences.

An interesting observation was found when looking at the conserved signals between polyploids with different tissue types. As previously stated, there was no such conservation of smRNA regulation in leaf tissue comparisons, but in root tissue, in Kitzbuhel accessions, there is substantial conservation of targeting patterns (11.8% of all significantly DT regions, Fig. 12C-D). As Kitzbuhel accessions have substantial gene flow between the two, this could explain the conservation in regulatory patterns when polyploids are adapting to new environments.

5. Conclusions

This study aimed to describe a potential mechanism explaining the diversification of a pair of sibling allopolyploids following recurrent, unidirectional formation. Our findings illustrate that despite significant morphological and ecological differences, there is little evidence for an extensive deviation of encoded regulatory patterns of smRNA between the polyploids. Coupled with the similarity and conservation of inherited smRNA targeting patterns from the progenitor when controlling for environmental variation, we show that despite independent formation, these regulatory elements are likely conserved from the unidirectional formation of the polyploids. When testing environmental effects, the divergent regulation from smRNAs increases drastically and the older and more broadly distributed *D. majalis* exhibits more regulatory plasticity. We see that tissue type, and gene flow, effect both the variation and conservation levels of smRNA regulation. Taking gene expression data and smRNA data together, we see that the environment has a dominant role in shaping the divergent

patterns we observe. It is possible that the increased standing variation of polyploid genomes provides different avenues for environmental selective pressures to act leading to larger diversification over time, though effects of drift, especially early in the establishment of the polyploids, cannot be discounted.

This study was meant to provide a descriptive overview of smRNA activity in this allopolyploid system. Bioinformatic analyses have limitations but provide powerful tools to help guide further research and experimental design to validate what is seen *in silico.* We found that filtering threshold can greatly affect the number of DT regions reported, though biological relevance (ex. GO enrichment analyses) remain largely unaffected. It would be interesting to implement a more thorough classification of the smRNA species and how their behavior is changing between the polyploids. Read length is an acceptable ad hoc solution for smRNA type but there are many tools available for more rigorous classifications which would allow potential de novo identification, and functional comparison with previously annotated smRNA species. Overall, our results provide an additional piece in the puzzle of macro and microevolutionary processes that fuel the diversification of allopolyploids, and *Dactylorhiza* genus.

Acknowledgements

Thanks first and foremost to Ovidiu Paun for agreeing to take me on as a master's student, and all the help and support he provided throughout my thesis. I would like to thank Christian Lexer for his lectures and discussions, which initially inspired me to think about a project in evolutionary research. Many thanks to Mimmi, Clara, Aglaia, Dana, Thelma, and the group of plant ecological genomics who all have directly or indirectly helped me throughout my time; as well as Thomas, Malte, and Slava whose previous work was extremely helpful in guiding me in the early stages. Lastly, I would like to thank and acknowledge my parents and my wife Irene whose care and support were invaluable through the long and often complicated process of completing a masters in Austria. I truly could not have managed

without them. This work was financially supported by the Austrian Science Fund (FWF) through grant Y661-B16 to Ovidiu Paun. Francisco Balao, Mark W. Chase, Michael F. Fay and Mikael Hedrén are acknowledged for help during collecting, and Juliane Baar 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.

Supplementary Figures

Locality	Group	N	Meaning
Common Garden	*D. fuchsii*	4	*D. fuchsii* grown in common garden
	D. incarnata	4	*D. incarnata* grown in common garden
	D. majalis	9	*D. majalis* grown in common garden
	D. traunsteineri	12	*D. traunsteineri* grown in common garden
Kitzbühel	mM	5	*D. majalis* individuals (m)grown in *majalis* environment (M)
	mT	5	*D. majalis* individuals (m) grown in *traunsteineri* environment (T)
	tM	5	*D. traunsteineri* individuals (t) grown in *D. majalis* environment (M)
	tT	4	*D. traunsteineri* individuals (t) grown in *D. traunsteineri* environment (T)
St. Ulrich	mM	4	*D. majalis* individuals (m)grown in *majalis* environment (M)
	mT	4	*D. majalis* individuals (m) grown in *traunsteineri* environment (T)
	tM	4	*D. traunsteineri* individuals (t) grown in *D. majalis* environment (M)
	tT	4	*D. traunsteineri* individuals (t) grown in *D. traunsteineri* environment (T)

Table S1. Study design and glossary for reciprocal transplant accessions.

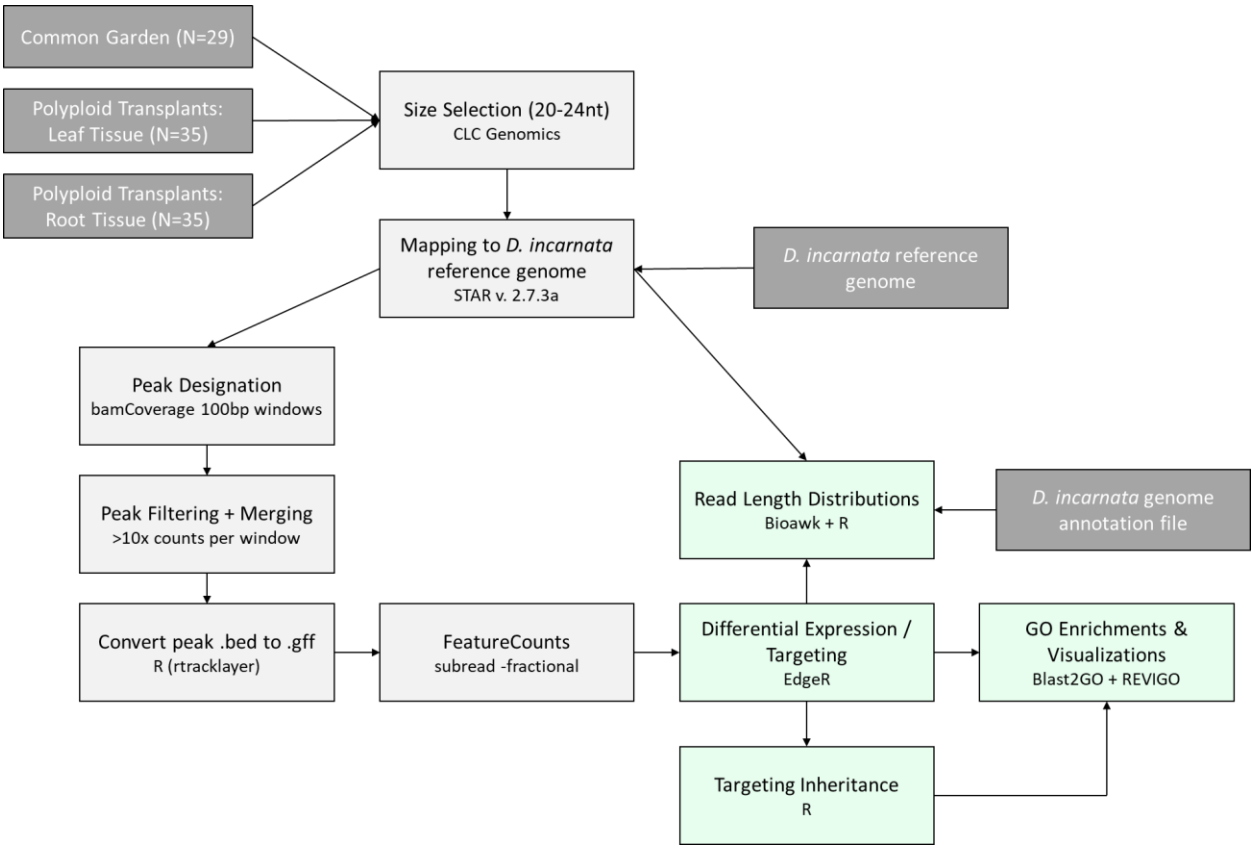
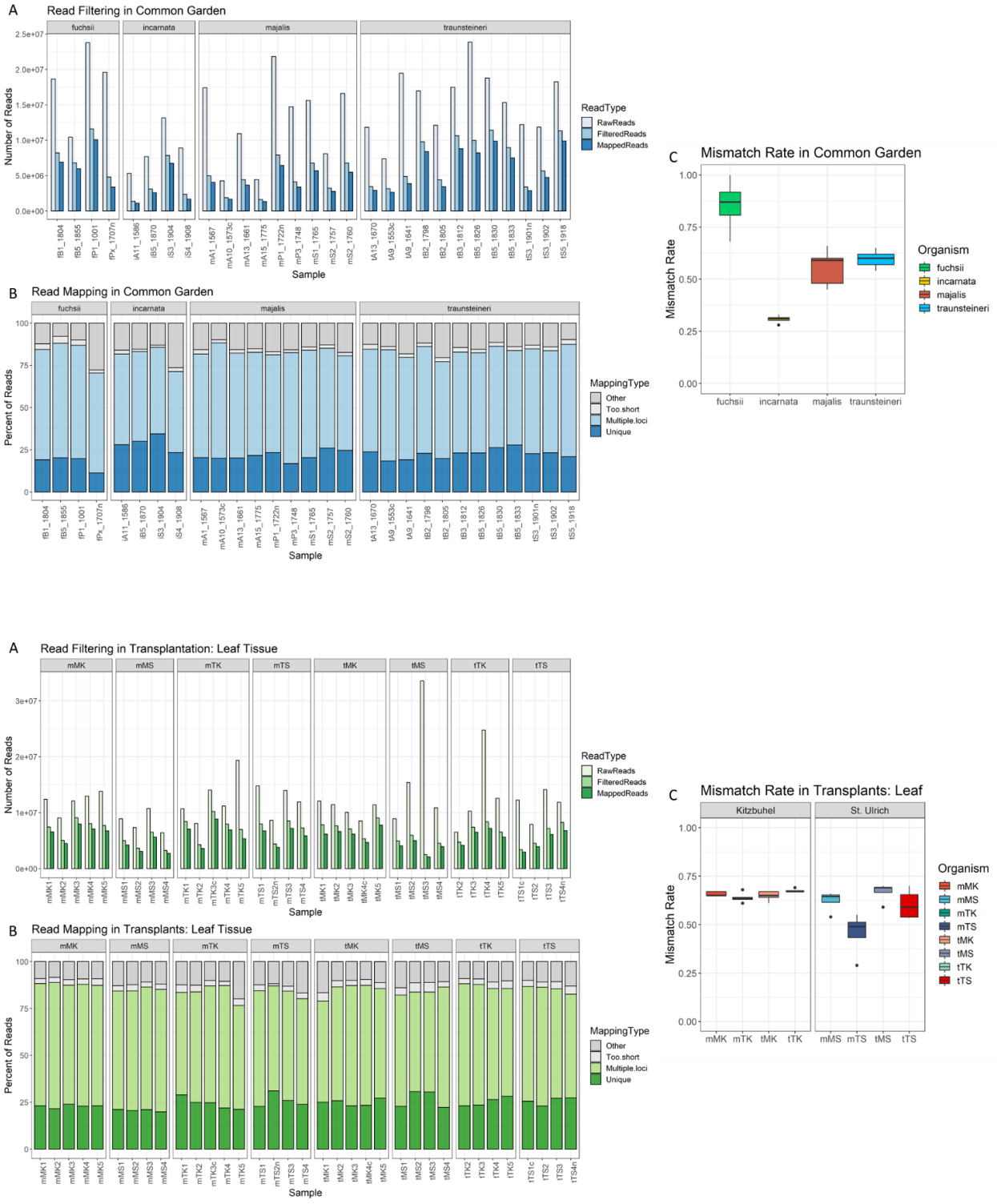


Figure S1. Pipeline overview. Dark gray represents input files, light gray represents data structuring and manipulation, and light green represents analyses and visualization steps.


A
Read Filtering in Common Garden
fuchsii
incarnata
majalis
traunsteineri
Number of Reads
Sample
ReadType
RawReads
FilteredReads
MappedReads
B
Read Mapping in Common Garden
Percent of Reads
MappingType
Other
Too.short
Multiple.loci
Unique
C
Mismatch Rate in Common Garden
Mismatch Rate
Organism
A
Read Filtering in Transplantation: Leaf Tissue
mMK
mMS
mTK
mTS
tMK
tMS
tTK
tTS
B
Read Mapping in Transplants: Leaf Tissue
C
Mismatch Rate in Transplants: Leaf
Kitzbuhel
St. Ulrich

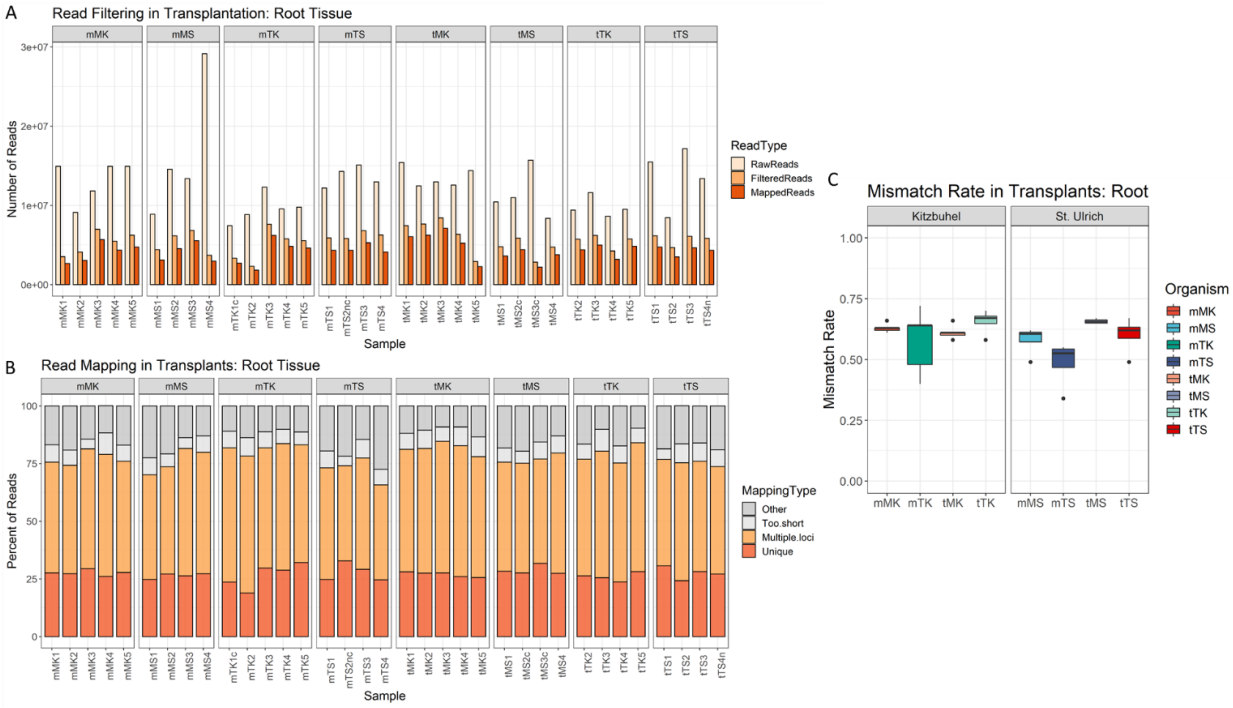


Figure S2. A) Mapping statistics for common garden (top, blue), transplant leaf tissue (middle, green) and transplant root tissue (bottom, orange). RawReads indicate non-filtered reads obtained from smRNAseq library preparation and sequencing. FilteredReads refer only to 20-24nt reads filtered with CLC Genomic Workbench, and MappedReads are reads aligned with STAR (multimapper and unique). **B)** Mapping outcomes for the listed samples. "Other" and "Too short" refer to unmapped reads while "Multiple loci" and "Unique" refer to mapped reads. **C)** Mismatch rate for each sample group.

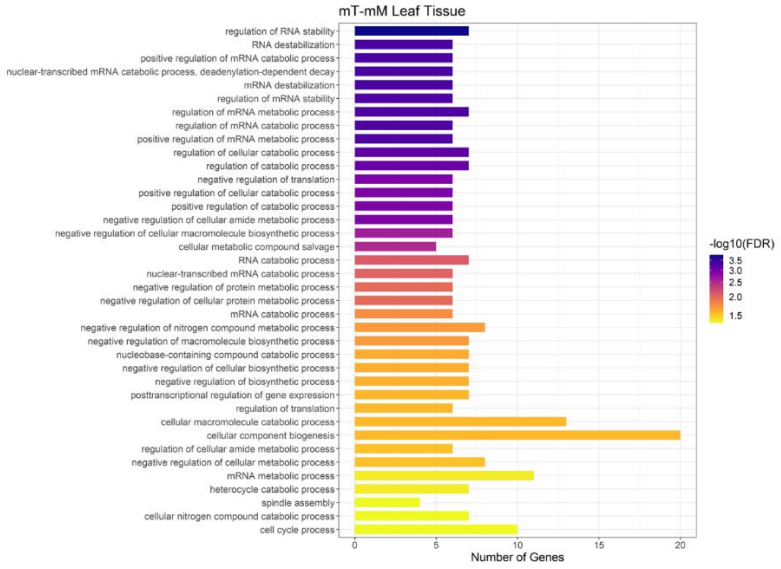


Figure S3. GO terms significantly over-enriched (FDR<0.01) in DT genes from leaf tissue samples between *D. majalis* grown in native vs. *D. traunsteineri* environment. Color indicates the significance gradient (dark = more significant) and bar length indicates specificity of the term in the public database (more general terms have larger bars).

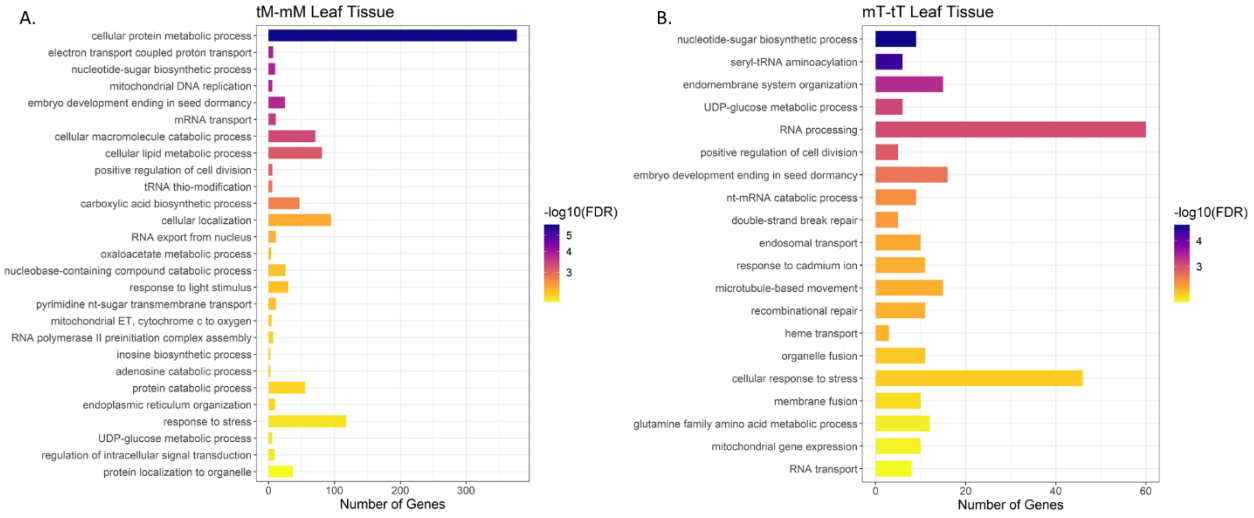


Figure S4. GO terms significantly over-enriched (FDR<0.01) between DT genes in leaf tissue samples of **A)** *D. traunsteineri* and *D. majalis* grown in *D. majalis'* native environment and **B)** *D. majalis* and *D. traunsteineri* grown in *D. traunsteineri's* native environments and color indicates the -log10 significance gradient (dark = more significant) and bar length indicates specificity of the term in the public database (more general terms have larger bars).

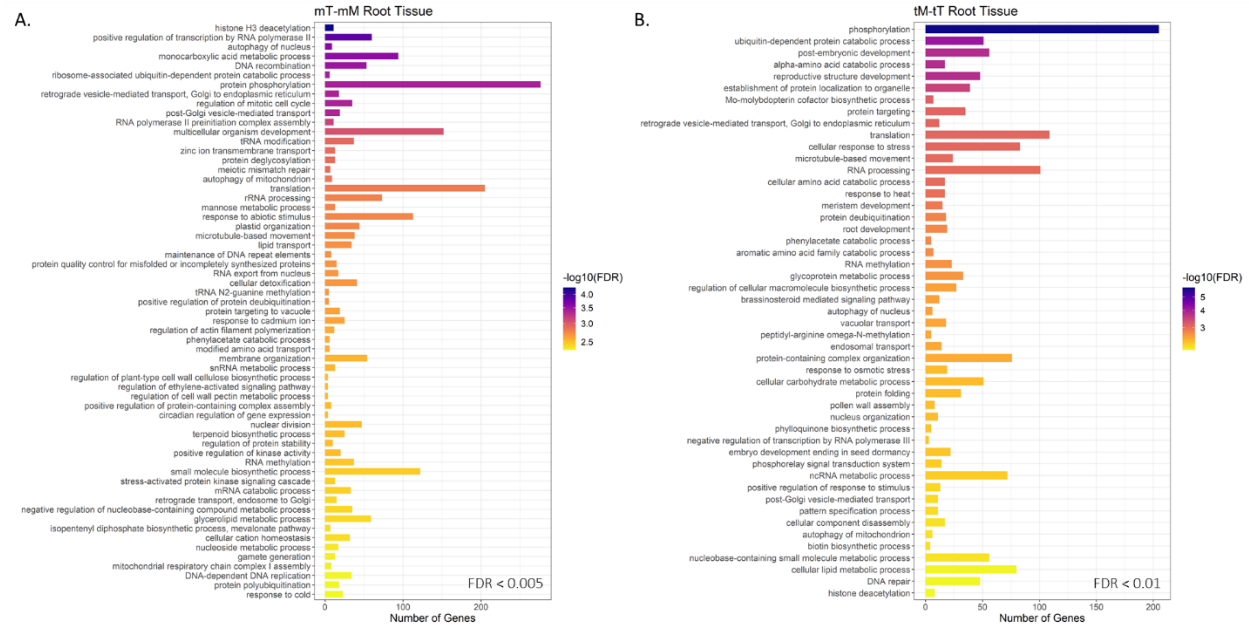


Figure S5. GO terms significantly over-enriched in DT genes from root tissue samples between **A)** *D. majalis* grown in native vs. *D. traunsteineri* environment and **B)** *D. traunsteineri* grown in native vs. *D. majalis* environment. Color indicates the -log10 significance gradient (dark = more significant) and bar length indicates specificity of the term in the public database (more general terms have larger bars).

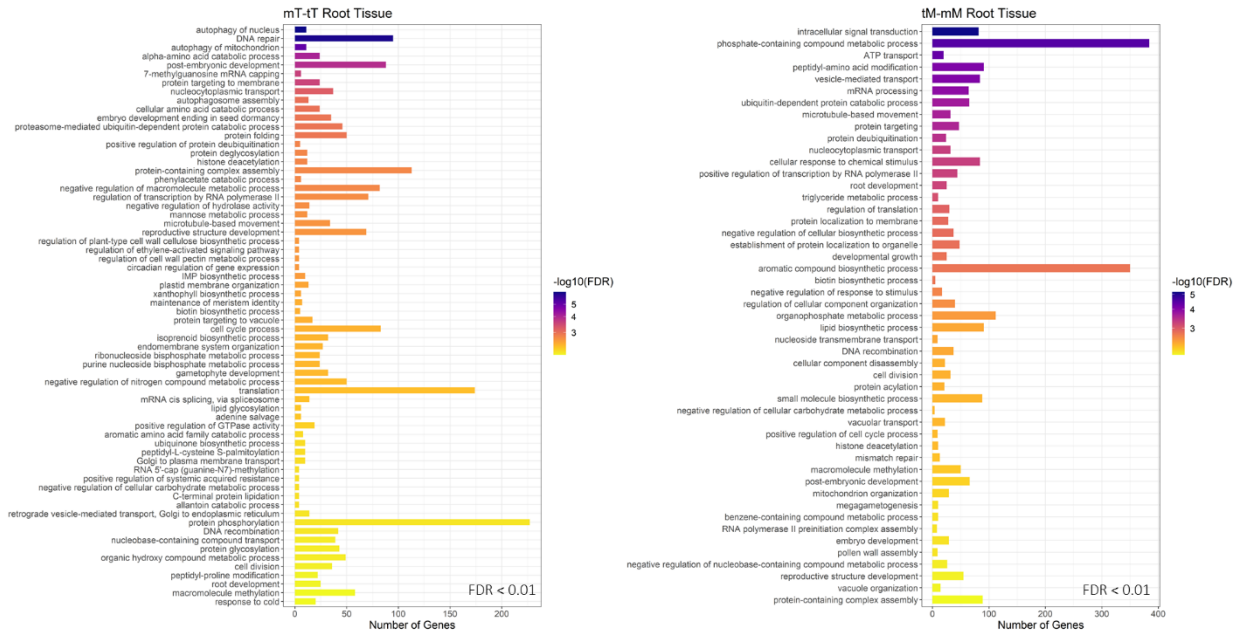


Figure S6. GO terms significantly over-enriched (FDR<0.01) in DT genes between root tissue samples of **A)** *D. traunsteineri* and *D. majalis* grown in *D. majalis'* native environment and **B)** *D. majalis* and *D. traunsteineri* grown in *D. traunsteineri's* native environments and color indicates the -log10 significance gradient (dark = more significant) and bar length indicates specificity of the term in the public database (more general terms have larger bars).

References

Abdel-Ghany, S. E., & Pilon, M. (2008). MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in Arabidopsis. *Journal of Biological Chemistry, 283* (23). https://doi.org/10.1074/jbc.M801406200

Adams, K. L., & Wendel, J. F. (2005). Novel patterns of gene expression in polyploid plants. In *Trends in Genetics* (Vol. 21, Issue 10). https://doi.org/10.1016/j.tig.2005.07.009

Alix, K., Gérard, P. R., Schwarzacher, T., & Heslop-Harrison, J. S. P. (2017). Polyploidy and interspecific hybridization: Partners for adaptation, speciation and evolution in plants. In *Annals of Botany* (Vol. 120, Issue 2, pp. 183–194). Oxford University Press. https://doi.org/10.1093/aob/mcx079

Aloni, R., Aloni, E., Langhans, M., & Ullrich, C. I. (2006). Role of cytokinin and auxin in shaping root architecture: Regulating vascular differentiation, lateral root initiation, root apical dominance and root gravitropism. Annals of Botany, 97 (5). https://doi.org/10.1093/aob/mcl027

Axtell, M. J. (2013). Classification and comparison of small RNAs from plants. In Annual Review of Plant Biology (Vol. 64, pp. 137–159). https://doi.org/10.1146/annurev-arplant-050312-120043

Birchler, J. A., Auger, D. L., & Riddle, N. C. (2003). In Search of the Molecular Basis of Heterosis. *Plant Cell* (Vol. 15, Issue 10). https://doi.org/10.1105/tpc.151030

Balao, F., Tannhäuser, M., Lorenzo, M. T., Hedrén, M., & Paun, O. (2016). Genetic differentiation and admixture between sibling allopolyploids in the Dactylorhiza majalis complex. *Heredity*, *116* (4), 351–361. https://doi.org/10.1038/hdy.2015.98

Balao, F., Trucchi, E., Wolfe, T. M., Hao, B. H., Lorenzo, M. T., Baar, J., Sedman, L., Kosiol, C., Amman, F., Chase, M. W., Hedrén, M., & 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. https://doi.org/10.1111/mec.14123

Bonnet, E., van de Peer, Y., & Rouzé, P. (2006). The small RNA world of plants. *New Phytologist* (Vol. 171, Issue 3, pp. 451–468). https://doi.org/10.1111/j.1469-8137.2006.01806.x

Borges, F., & Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nature Reviews Molecular Cell Biology* (Vol. 16, Issue 12, pp. 727–741). Nature Publishing Group. https://doi.org/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 (Orchidaceae). *Systematic Biology, 69* (1). https://doi.org/10.1093/sysbio/syz035

Burton, T. L., & Husband, B. C. (2000). FITNESS DIFFERENCES AMONG DIPLOIDS, TETRAPLOIDS, AND THEIR TRIPLOID PROGENY IN CHAMERION ANGUSTIFOLIUM: MECHANISMS OF INVIABILITY AND IMPLICATIONS FOR POLYPLOID EVOLUTION. *Evolution*, *54* (4). https://doi.org/10.1554/0014-3820(2000)054[1182:fdadta]2.0.co;2

Candar-Cakir, B., Arican, E., & Zhang, B. (2016). Small RNA and degradome deep sequencing reveals drought-and tissue-specific micrornas and their important roles in drought-sensitive and drought-tolerant tomato genotypes. *Plant Biotechnology Journal, 14* (8). https://doi.org/10.1111/pbi.12533

Chen, Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology* (Vol. 58). https://doi.org/10.1146/annurev.arplant.58.032806.103835

Chen, Y., Lun, A. T. L., & Smyth, G. K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research, 5*, 1438. https://doi.org/10.12688/f1000research.8987.1

Chuck, G., & O'Connor, D. (2010). Small RNAs going the distance during plant development. *Current Opinion in Plant Biology* (Vol. 13, Issue 1). https://doi.org/10.1016/j.pbi.2009.08.006

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* (Vol. 6, Issue 11, pp. 836–846). https://doi.org/10.1038/nrg1711

Crnokrak, P., & Roff, D. A. (1999). Inbreeding depression in the wild. Heredity, 83 (3). https://doi.org/10.1038/sj.hdy.6885530

Deprost, D., Yao, L., Sormani, R., Moreau, M., Leterreux, G., Bedu, M., Robaglia, C., & Meyer, C. (2007). The Arabidopsis TOR kinase links plant growth, yield, stress resistance and mRNA translation. *EMBO Reports*, 8 (9). https://doi.org/10.1038/sj.embor.7401043

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics, 29* (1). https://doi.org/10.1093/bioinformatics/bts635

Dobrenel, T., Caldana, C., Hanson, J., Robaglia, C., Vincentz, M., Veit, B., & Meyer, C. (2016). TOR Signaling and Nutrient Sensing. *Annual Review of Plant Biology* (Vol. 67). https://doi.org/10.1146/annurev-arplant-043014-114648

Dong, Y., Silbermann, M., Speiser, A., Forieri, I., Linster, E., Poschet, G., Allboje Samami, A., Wanatabe, M., Sticht, C., Teleman, A. A., Deragon, J. M., Saito, K., Hell, R., & Wirtz, M. (2017). Sulfur availability regulates plant growth via glucose-TOR signaling. *Nature Communications, 8* (1). https://doi.org/10.1038/s41467-017-01224-w

East EM. Heterosis. Genetics 1936; 21:375–397. [PubMed: 17246801]

Fujii, H., Chiou, T. J., Lin, S. I., Aung, K., & Zhu, J. K. (2005). A miRNA involved in phosphate-starvation response in Arabidopsis. Current Biology, 15 (22). https://doi.org/10.1016/j.cub.2005.10.016

Gifford, M. L., Dean, A., Gutierrez, R. A., Coruzzi, G. M., & Birnbaum, K. D. (2008). Cell-specific nitrogen responses mediate developmental plasticity. Proceedings of the National Academy of Sciences of the United States of America, 105 (2). https://doi.org/10.1073/pnas.0709559105

Götz S., García-Gómez J. M., Terol J., Williams T. D., Nagaraj S. H., Nueda M. J., Robles M., Talón M., Dopazo J., Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420-3435 (2008).

Guan, X., Song, Q., & Chen, Z. J. (2014). Polyploidy and small RNA regulation of cotton fiber development. *Trends in Plant Science* (Vol. 19, Issue 8). https://doi.org/10.1016/j.tplants.2014.04.007

Guo, M., Davis, D., & Birchler, J. A. (1996). Dosage effects on gene expression in a maize ploidy series. *Genetics*, *142*(4). https://doi.org/10.1093/genetics/142.4.1349

Ha, M., Lu, J., Tian, L., Ramachandran, V., Kasschau, K. D., Chapman, E. J., Carrington, J. C., Chen, X., Wang, X. 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). https://doi.org/10.1073/pnas.0907003106

Hawranek AS. 2021. The impact of recurrent origins and gene flow on the genetic structure of allopolyploid marsh orchids (Dactylorhiza, Orchidaceae). MSc Thesis, University of Vienna. https://doi.org/10.25365/thesis.70360, https://utheses.univie.ac.at/detail/60380/

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. Trends in Ecology and Evolution (Vol. 17, Issue 5). https://doi.org/10.1016/S0169-5347(02)02489-8

Kondorosi, E., Roudier, F., Gendreau, E. (2000). Plant cell-size control: Growing by ploidy? *Current Opinion in Plant Biology* (Vol. 3, Issue 6). https://doi.org/10.1016/S1369-5266(00)00118-7

Kong, W. W., & Yang, Z. M. (2010). Identification of iron-deficiency responsive microRNA genes and cis-elements in Arabidopsis. *Plant Physiology and Biochemistry, 48* (2–3). https://doi.org/10.1016/j.plaphy.2009.12.008

Kumar, S., Verma, S., & Trivedi, P. K. (2017). Involvement of small RNAs in phosphorus and sulfur sensing, signaling and stress: Current update. *Frontiers in Plant Science* (Vol. 8). https://doi.org/10.3389/fpls.2017.00285

Lawrence M, Gentleman R, Carey V (2009). “rtracklayer: an R package for interfacing with genome browsers.” *Bioinformatics*, **25**, 1841-1842. doi: 10.1093/bioinformatics/btp328

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078-9. [PMID: 19505943]

Li, S., Liu, J., Liu, Z., Li, X., Wu, F., & He, Y. (2014). HEAT-INDUCED TAS1 TARGET1 mediates thermotolerance via heat stress transcription factor A1a-directed pathways in arabidopsis. Plant Cell, 26 (4). https://doi.org/10.1105/tpc.114.124883

Liao Y, Smyth GK and Shi W. featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923-30, 2014

Liu, H. H., Tian, X., Li, Y. J., Wu, C. A., & Zheng, C. C. (2008). Microarray-based analysis of stress-regulated microRNAs in Arabidopsis thaliana. RNA, 14 (5). https://doi.org/10.1261/rna.895308

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). https://doi.org/10.1093/nar/gks042

Miller, J. S., & Venable, D. L. (2000). Polyploidy and the evolution of gender dimorphism in plants. *Science*, *289*(5488). https://doi.org/10.1126/science.289.5488.2335

Moore, R. C., & Purugganan, M. D. (2005). The evolutionary dynamics of plant duplicate genes. In *Current Opinion in Plant Biology* (Vol. 8, Issue 2). https://doi.org/10.1016/j.pbi.2004.12.001

Nodine, M. D., & Bartel, D. P. (2010). MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes and Development, 24 (23). https://doi.org/10.1101/gad.1986710

Ng, D. W. K., Lu, J., & Chen, Z. J. (2012). Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Current Opinion in Plant Biology* (Vol. 15, Issue 2, pp. 154–161). https://doi.org/10.1016/j.pbi.2012.01.007

Paun, O., Fay, M. F., Soltis, D. E., & Chase, M. W. (2007). Genetic and epigenetic alterations after hybridization and genome doubling. *Taxon* (Vol. 56, Issue 3). https://doi.org/10.2307/25065850

Paun, O., Bateman, R. M., Fay, M. F., Hedrén, M., Civeyrel, L., & Chase, M. W. (2010). Stable epigenetic effects impact adaptation in allopolyploid orchids (Dactylorhiza: Orchidaceae). *Molecular Biology and Evolution*, *27* (11). https://doi.org/10.1093/molbev/msq150

Pu, Y., Luo, X., & Bassham, D. C. (2017). Tor-dependent and -independent pathways regulate autophagy in arabidopsis thaliana. *Frontiers in Plant Science, 8.* https://doi.org/10.3389/fpls.2017.01204

QIAGEN CLC Genomics Workbench 20.0 (https://digitalinsights.qiagen.com/)

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

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Ramírez, Fidel, Devon P. Ryan, Björn Grüning, Vivek Bhardwaj, Fabian Kilpert, Andreas S. Richter, Steffen Heyne, Friederike Dündar, and Thomas Manke. deepTools2: A next Generation Web Server for Deep-Sequencing Data Analysis. *Nucleic Acids Research* (2016). doi:10.1093/nar/gkw257

Robinson, M.D., Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol, 11*, R25 (2010). https://doi.org/10.1186/gb-2010-11-3-r25

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research, 43* (7). https://doi.org/10.1093/nar/gkv007

Santos, J. L., Alfaro, D., Sanchez-Moran, E., Armstrong, S. J., Franklin, F. C. H., & Jones, G. H. (2003). *Partial Diploidization of Meiosis in Autotetraploid Arabidopsis thaliana. Genetics*, 165 (3). https://doi.org/10.1093/genetics/165.3.1533

Scheid, O. M., Jakovleva, L., Afsar, K., Maluszynska, J., & Paszkowski, J. (1996). A change of ploidy can modify epigenetic silencing. *Proceedings of the National Academy of Sciences of the United States of America*, *93* (14). https://doi.org/10.1073/pnas.93.14.7114

Shi, L., Wu, Y., & Sheen, J. (2018). TOR signaling in plants: conservation and innovation. *Development* (Cambridge, England) (Vol. 145, Issue 13). https://doi.org/10.1242/dev.160887

Soltis Douglas E., & Soltis Pamela S. (1999). *Polyploid: recurrent formation and genome evolution* (Vol. 14, Issue 9).

Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., & Soltis, P. S. (2009). Polyploidy and angiosperm diversification. *American Journal of Botany*, *96* (1), 336–348. https://doi.org/10.3732/ajb.0800079

Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual Review of Plant Biology*, *60*. https://doi.org/10.1146/annurev.arplant.043008.092039

Sunkar, R., Kapoor, A., & Zhu, J. K. (2006). Erratum: Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in Arabidopsis is mediated by downregulation of miR398 and important for oxidative stress tolerance (Plant Cell (2006) 18 (2051-2065)). In Plant Cell (Vol. 18, Issue 9). https://doi.org/10.1105/tpc.106.180960

Supek F, Bošnjak M, Škunca N, Šmuc T. "*REVIGO summarizes and visualizes long lists of Gene Ontology terms*" PLoS ONE 2011. doi:10.1371/journal.pone.0021800

Van de Peer, Y., Fawcett, J. A., Proost, S., Sterck, L., & Vandepoele, K. (2009, December). The flowering world: a tale of duplications. *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2009.09.001

Wang, J., Hill, W. G., Charlesworth, D., & Charlesworth, B. (1999). Dynamics of inbreeding depression due to deleterious mutations in small populations: Mutation parameters and inbreeding rate. Genetical Research, 74 (2). https://doi.org/10.1017/S0016672399003900

Wang, J., Tian, L., Madlung, A., Lee, H. S., Chen, M., Lee, J. J., Watson, B., Kagochi, T., Comai, L., & Chen, Z. J. (2004). Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. *Genetics*, *167* (4). https://doi.org/10.1534/genetics.104.027896

Wei, Y., Li, F., Zhang, S., Zhang, S., Zhang, H., & Sun, R. (2019). Analysis of small RNA changes in different Brassica napus synthetic allopolyploids. *PeerJ*, *2019* (9). https://doi.org/10.7717/peerj.7621

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org

Wolfe, T. M., Balao, F., Trucchi, E., Bachmann, G., Gu, W., Baar, J., Paun, O. (2021). Recurrent allopolyploidization events diversify eco-physiological traits in marsh orchids. *bioRxiv*, 2021.08.28.458039.

Wolfe T.M. (2019) Genomic interactions and their eco-physiological implications for the evolution of allopolyploid marsh orchids. PhD Dissertation, University of Vienna. Chapter 3. Available at https://doi.org/10.25365/thesis.58381

Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., & Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America*, *106* (33). https://doi.org/10.1073/pnas.0811575106

Xiong, Y., & Sheen, J. (2015). Novel links in the plant TOR kinase signaling network. *Current Opinion in Plant Biology* (Vol. 28). https://doi.org/10.1016/j.pbi.2015.09.006

Yang, Z., Wang, Z., Yang, C., Yang, Z., Li, H., & Wu, Y. (2019). Physiological responses and small RNAs changes in maize under nitrogen deficiency and resupply. *Genes and Genomics*, *41* (10). https://doi.org/10.1007/s13258-019-00848-0

Zhang, W., Swarup, R., Bennett, M., Schaller, G. E., & Kieber, J. J. (2013). Cytokinin induces cell division in the quiescent center of the arabidopsis root apical meristem. *Current Biology,* 23 (20). https://doi.org/10.1016/j.cub.2013.08.008

Zhao, B., Ge, L., Liang, R., Li, W., Ruan, K., Lin, H., & Jin, Y. (2009). Members of miR-169 family are induced by high salinity and transiently inhibit the NF-YA transcription factor. BMC Molecular Biology, 10. https://doi.org/10.1186/1471-2199-10-29

Zhao, M., Ding, H., Zhu, J. K., Zhang, F., & Li, W. X. (2011). Involvement of miR169 in the nitrogen-starvation responses in Arabidopsis. New Phytologist, 190 (4). https://doi.org/10.1111/j.1469-8137.2011.03647