ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Adaptive sequence evolution is driven by biotic stress in a pair of orchid species (*Dactylorhiza*) with distinct ecological optima

Francisco Balao[1,2] (iD) | Emiliano Trucchi[1,3] | Thomas M. Wolfe[1,4] | Bao-Hai Hao[1] | Maria Teresa Lorenzo[1,2] | Juliane Baar[1] | Laura Sedman[5] | Carolin Kosiol[6,7] | Fabian Amman[8] | Mark W. Chase[9,10] | Mikael Hedrén[11] | Ovidiu Paun[1] (iD)

[1]Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria

[2]Departamento de Biología Vegetal y Ecología, University of Seville, Sevilla, Spain

[3]Department of Life Sciences and Biotechnologies, University of Ferrara, Ferrara, Italy

[4]Vienna Graduate School of Population Genetics, Vienna, Austria

[5]Gregor Mendel Institute for Plant Molecular Biology, Vienna, Austria

[6]Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria

[7]Centre of Biological Diversity, School of Biology, University of St Andrews, St Andrews, UK

[8]Department of Chromosome Biology, University of Vienna, Vienna, Austria

[9]Royal Botanic Gardens Kew, Richmond, UK

[10]School of Plant Biology, University of Western Australia, Crawley, Perth, WA, Australia

[11]Department of Biology, Lund University, Lund, Sweden

Correspondence
Ovidiu Paun, Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria.
Email: ovidiu.paun@univie.ac.at

Abstract

The orchid family is the largest in the angiosperms, but little is known about the molecular basis of the significant variation they exhibit. We investigate here the transcriptomic divergence between two European terrestrial orchids, *Dactylorhiza incarnata* and *Dactylorhiza fuchsii*, and integrate these results in the context of their distinct ecologies that we also document. Clear signals of lineage-specific adaptive evolution of protein-coding sequences are identified, notably targeting elements of biotic defence, including both physical and chemical adaptations in the context of divergent pools of pathogens and herbivores. In turn, a substantial regulatory divergence between the two species appears linked to adaptation/acclimation to abiotic conditions. Several of the pathways affected by differential expression are also targeted by deviating post-transcriptional regulation via sRNAs. Finally, *D. incarnata* appears to suffer from insufficient sRNA control over the activity of RNA-dependent DNA polymerase, resulting in increased activity of class I transposable elements and, over time, in larger genome size than that of *D. fuchsii*. The extensive molecular divergence between the two species suggests significant genomic and transcriptomic shock in their hybrids and offers insights into the difficulty of coexistence at the homoploid level. Altogether, biological response to selection, accumulated during the history of these orchids, appears governed by their microenvironmental context, in which biotic and abiotic pressures act synergistically to shape transcriptome structure, expression and regulation.

KEYWORDS
abiotic stress, defence, ecological divergence, positive selection, small RNAs, transcriptomics

---

1 | INTRODUCTION

The orchid family represents an extraordinary biological diversification, containing over 8% of all flowering plant species, making Orchidaceae the largest family of angiosperms (Chase et al., 2015; Givnish et al., 2015). They are distributed from equatorial lowlands to arctic and alpine locations, through a remarkable variability, affecting their morphology, ecology, genomes and life history strategies (Benzing, 1986). Orchids are particularly renowned for their specialized and diverse pollination systems (Darwin, 1877), and pollinator-driven isolation is often discussed as a main factor driving their diversification (Benzing, 1986; Cozzolino & Widmer, 2005). However, a third of all orchids are estimated to be pollinated on a deceptive basis (i.e., they offer no reward, Tremblay, Ackerman, Zimmerman, & Calvo, 2005), often with the help of unspecific pollinators. Indeed, pollinator-mediated reproductive isolation has been shown to play little or no role in food-deceptive orchids (Moccia, Widmer, & Cozzolino, 2007).

Other factors put forward to explain orchid diversity range from epiphytism (Gravendeel, Smithson, Slik, & Schuiteman, 2004; Silvera, Santiago, Cushman, & Winter, 2009), low fruiting success setting the stage for a dynamic interplay between drift and natural selection (Tremblay et al., 2005), obligate orchid–mycorrhiza interactions (Tupac Otero & Flanagan, 2006), CAM photosynthesis and tropical distribution (Silvera et al., 2009) or a combination of these factors (Givnish et al., 2015). In general, orchid species diversity has been reported to correlate to area size and, even more strongly, with habitat heterogeneity (e.g., with maximum island elevation as a surrogate for habitat diversity in the Caribbean, Ackerman, Trejo-Torres, & Crespo-Chuy, 2007), indicating that ecological conditions may be a major driver of diversity in the family. However, only little is known about the molecular factors underlying divergence between orchid species, mainly due to disproportionately limited genomic data in public databases despite the recent availability of a reference genome for *Phalaenopsis equestris* (Cai et al., 2015).

The molecular basis of biodiversity can reside in both amino acid sequence variation (Nielsen, 2005) and regulatory divergence that triggers protein abundance shifts (Wray, 2007), and these molecular components may evolve at different rates. For example, accumulating evidence suggests a role for divergent small RNAs (sRNAs) and other regulatory elements in postzygotic reproductive isolation, and the hybrid incompatibility they trigger is predicted to evolve relatively early in the speciation process (Ha, Pang, Agarwal, & Chen, 2008; Hollister et al., 2011; Kitano, Yoshida, & Suzuki, 2013; McManus et al., 2010; Michalak, 2008). Small RNAs, including plant microRNAs (miRNAs), secondary small interfering RNAs (siRNAs) and heterochromatic small interfering RNAs (hetsiRNAs, Borges & Martienssen, 2015), are involved in (post)transcriptional gene regulation, RNA-directed DNA methylation and chromatin remodelling, playing major roles in the maintenance of genome stability and function (Zhang, 2008). Importantly, regulatory divergence is generally dynamic, potentially exposing phenotypes to natural selection

differentially across time and environmental conditions. Biological responses to selection depend heavily on the environmental context, with regard to biotic and abiotic conditions that can act individually or synergistically to either accelerate or constrain trait evolution (Anderson, Wagner, Rushworth, Prasad, & Mitchell-Olds, 2014). Understanding the extent and timescale of the interactions between different molecular levels in response to the plethora of characteristics of complex environments is a primary goal of evolutionary biology.

In this study, we aimed to investigate the molecular basis of adaptation to specific microenvironments in *Dactylorhiza fuchsii* and *D. incarnata*, two terrestrial, food-deceptive orchid species (Sletvold, Grindeland, & Ågren, 2010) that provide a suitable model to trace effects of natural selection. The two orchid species broadly share geographical distribution (northern and central Europe and Western Asia, Tutin et al., 1980) and often grow in proximity. However, they have clearly distinct morphologies and also differ in their ecological preference (Paun et al., 2011). *Dactylorhiza fuchsii* inhabits grasslands, open woods and disturbed habitats such as road margins, whereas *D. incarnata* prefers marshes, fens and bog habitats. The genetic structure of the two species has most likely been shaped by different histories as, in contrast to *D. fuchsii*, *D. incarnata* appears to have experienced a dramatic reduction in population size that homogenized its genetic variation at neutral markers (Balao, Tannhäuser, Lorenzo, Hedren, & Paun, 2016; Hedrén, 1996; Pillon et al., 2007), despite its current wide distribution and its significant morphological heterogeneity. The two species have hybridized on multiple occasions, producing an array of ecologically divergent allotetraploid sibling species (Devos, Raspe, Oh, Tyteca, & Jacquemart, 2006; Hedrén, 1996; Hedrén, Fay, & Chase, 2001; Pillon et al., 2007), several of which are also broadly distributed in Europe.

With RNAseq across several replicates originating from different European regions, we searched for genes with lineage-specific evolutionary patterns shaped by positive selection, and tested for differential expression (DE) that correlate to phenotypic divergence between the two species (De Wit, Pespeni, & Palumbi, 2015). We further aimed to time the divergence between the two species and tested whether the coding single nucleotide polymorphism (cSNP) patterns obtained follow the low neutral genetic diversity in *D. incarnata*. Finally, we sequenced sRNAs to track divergent post-transcriptional gene regulation (Ehrenreich & Purugganan, 2008). One specific aim within the latter analysis was to search for a potential deviating negative control of transposable elements (TEs) that may have resulted in the difference in genome size between the two species (1C values: 3.55 pg for *D. incarnata* vs. 2.98 pg for *D. fuchsii*, Aagaard, Sastad, Greilhuber, & Moen, 2005). Besides understanding in detail the molecular divergence between the diploid orchid species, this work is also setting a foundation for further studies of the molecular principles underlying widespread allopolyploidy in this orchid group and its impact on evolutionary novelty and ecological diversification.

2 | MATERIALS AND METHODS

2.1 | RNAseq and de novo transcriptome assemblies

We sequenced leaf RNAseq libraries (see Appendix S1) of five wild individuals each of *D. fuchsii* and *D. incarnata* from different European regions (Table 1, Fig. S1) after they had been transplanted and grown in a common garden in Vienna, Austria, for at least one growth cycle in order to remove temporary environmental effects. Every year, *Dactylorhiza* plants produce a new tuber that carries them through to the next year. We focused on leaves as an ecologically relevant tissue, which also avoids potential contamination by mycorrhizal fungi associated with orchid roots. Across the individuals analysed, the directional RNA sequencing generated almost 1.5 billion pairs of 100-bp reads (Table 1).

As the closest assembled genome (*P. equestris*, Cai et al., 2015) is highly divergent from *Dactylorhiza* (their most recent common ancestor lived 55–70 Ma, Givnish et al., 2015), we have assembled de novo with *Trinity* (Haas et al., 2013) a combined reference transcriptome for both *Dactylorhiza* species (for details see Appendix S1 and Fig. S2). The individual transcriptomes are available from GenBank Transcriptome Shotgun Assembly (TSA) Database (Accession nos.: GFHX00000000 and GFJV00000000 for *D. fuchsii* and *D. incarnata*, respectively). The combined final transcriptome is available from https://plantgenomics.univie.ac.at/downloads/. Due to its low number of assembled contigs pointing to a technical depletion of complexity, and due to its outlier position in ordination analyses, the individual fA6 was excluded from all further RNAseq analyses.

2.2 | Variant discovery and assessment of genetic diversity

In order to conduct evolutionary analyses, the filtered RNAseq reads were mapped to an available, closely related outgroup reference, the transcriptome of *Orchis italica* (De Paolo, Salvemini, Gaudio, & Aceto, 2014), which is a member of the same orchid subtribe (Chase et al., 2015). The mapping has been performed with the second-pass approach (Engström et al., 2013) of STAR v2.4.1d (Dobin et al., 2013), by lowering the maximum allowed ratio of mismatches to read length to 0.11. The best practices recommendations (DePristo et al., 2011; Van Auwera et al., 2013) for GATK version 3 (McKenna et al., 2010) have been followed, but with a hybrid approach between RNA and DNA sequencing as the analyses have been performed on a reference transcriptome, not a full genome. After processing the BAM files by adding read groups and removing duplicates with PICARD TOOLS (v.1.119, http://broadinstitute.github.io/picard/), we split the reads into exon segments and reassigned STAR mapping qualities. The INDELREALIGNER module from GATK v3.5-0-g36282e4 was used to improve local alignments around indels, downsampling to a maximum of 100× coverage. Variants were further called for each sample in the GVCF mode of the GATK HAPLOTYPECALLER to generate an intermediate gVCF, by ignoring any soft clipped bases in the reads to minimize false positives and negative calls. Next, we processed all samples in the cohort in a joint genotyping analysis with GENOTYPEGVCFS with the minimum Phred-scaled confidence threshold of 20 at which variants should be called and emitted, following GATK best practices recommendations for RNAseq. After retaining only SNPs with the SELECTVARIANTS module of GATK to avoid later uncertainties in alignments, variants have been further filtered out if any of the following criteria were fulfilled: the quality normalized by the coverage (QD) was <2.0, the Phred-scaled *p*-value for Fisher's exact test to detect strand bias (FS) was >60.0, or the root mean square of mapping quality across all samples (MQ) was <40. Diversity measures were calculated with VCFTOOLS v.0.1.14 (Danecek et al., 2011) with the options –het (for the per-individual heterozygosity and inbreeding coefficient F), –site-pi (for per site nucleotide divergence π) and –singletons (for private alleles).

TABLE 1 Details of the leaf samples analysed in this study

Region	No.	Acronym	Collectors	Latitude/Longitude	Raw pairs of RNAseq reads (M)	Raw sRNAseq reads (M)
D. fuchsii						
Pannonian	1566	fA6	GS, OP	47°53′41″N 19°52′01″E	226.8	16.7
Pyrenees	1001	fP1	MC, MF, OP	43°12′43″N 0°49′50″E	79.4	23.8
Pyrenees	1707	fP7	OP	42°53′40″N 1°55′46″E	166.0[a]	21.8
Britain	1804	fB4	FB, OP	54°15′10″N 0°41′06″W	107.1	18.6
Britain	1855	fB5	FB, OP	57°40′03″N 7°14′32″W	224.3	10.4
D. incarnata						
Alps	1586	iA6	OP	47°16′54″N 11°11′04″E	199.8	5.3
Scandinavia	1904	iS4	MH	57°20′26″N 18°19′16″E	82.0	13.1
Scandinavia	1908	iS8	MH	57°49′01″N 18°53′43″E	201.1	8.9
Britain	1176	iB6	RB	57°40′19″N 7°14′17″W	101.0	-
Britain	1870	iB0	FB, OP	57°40′14″N 7°14′22″W	75.4	7.7

Collector abbreviations are FB, Francisco Balao; GS, Gabor Szramko; MC, Mark Chase; MF, Michael Fay; MH, Mikael Hedrén; OP, Ovidiu Paun and RB, Richard Bateman.
[a]Summed over two half lanes.

For visualizing genetic relationships between individuals, we filtered out with VCFTOOLS the variants with data in less than six (of the nine) *Dactylorhiza* individuals. The 449,518 retained SNPs were further concatenated to a Phylip file with PGDSPIDER v.2.0.9.2 (Lischer & Excoffier, 2012), summarizing heterozygosities as IUPAC ambiguities. A maximum-likelihood phylogenetic tree with 1,000 rapid bootstrap inferences and a GTRCAT substitution model were inferred with RAXML v.7.2.8 (Stamatakis, 2014) and the results were graphically visualized with FIGTREE v.1.4.2 (available from http://tree.bio.ed.ac.uk/software/figtree/).

The genetic similarity among individuals of the two *Dactylorhiza* species was also visualized with a principal component analysis (PCA). The data were first filtered with VCFTOOLS by retaining only SNPs with the minor allele present in at least two individuals. The retained 129,511 biallelic SNPs were imported in the BIOCONDUCTOR (Huber et al., 2015) with GDSFMT, and the PCA was finally drawn with the R package SNPRELATE (Zheng et al., 2012).

2.3 | Rates of sequence evolution and targets of selection

Synthetic transcripts of each sample have been generated from the GATK-filtered vcf file with the GATK FASTAALTERNATEREFERENCEMAKER, using IUPAC encodings for heterozygote states. We further used for each sample the GATK CALLABLELOCI with default settings to filter only the callable transcriptomic intervals. These have been intersected with BEDTOOLS v.2.25.0 (Quinlan & Hall, 2010) with exonic regions of the *O. italica* reference predicted with TRANSDECODER (v.2.0.1, https://transdecoder.github.io/). BEDTOOLS GETFASTA has been used to retain from each individual only the target intervals. Finally, indel-free coding DNA sequence (CDS) alignments of at least 150-bp coding regions have been retained for positive selection analyses.

Nonsynonymous (K_a) and synonymous (K_s) substitution rates for CDS alignments of pairwise combinations of sequences of *D. fuchsii*, *D. incarnata* and *O. italica* were first calculated using an approximate method (MYN) modified after Yang and Nielsen (2000) and implemented in the software package KAKS_CALCULATOR v.2.0 (Zhang, Li, & Yu, 2006). The level of significance of Fisher's exact tests was adjusted with the false discovery rate (FDR) approach with the P.ADJUST function in R. The contigs shown to be under positive selection (i.e., $K_a/K_s > 1$) between *D. incarnata* and *D. fuchsii* within an FDR level <0.1 have been confirmed by estimating pairwise dN/dS (ω) with a maximum-likelihood method implemented in CODEML of the PAML package v.4.8a (Danecek et al., 2011). To test for statistical significance ($p < .05$), a likelihood ratio test was applied against a model with fixed ω = 1 (Wong, Yang, Goldman, & Nielsen, 2004). We further aimed to detect the particular branch on which positive selection acted by performing branch-site model tests M1a versus M2a (Zhang, Nielsen, & Yang, 2005) in CODEML, comparing models with estimated ω versus the corresponding null models with fixed ω = 1 along the *D. fuchsii* branch, and the *D. incarnata* branch, respectively. A simple phylogenetic tree has been used for the inference as ((*D. fuchsii*, *D. incarnata*), *O. italica*)). Using the initial list of tested

contigs as reference set and the genes under positive selection on each branch as the test sets, GO term enrichments were determined with Fisher's exact tests and a *p*-value <.01 in BLAST2GO to place the genes in a meaningful context. The GO terms were summarized with REVIGO as explained above.

2.4 | Abundance estimation

For analyses of DE, the filtered RNAseq reads were mapped to the *Dactylorhiza* reference using STAR v.2.4.1b (Dobin et al., 2013) with default settings allowing for up to 20 multiple mapping reads. A unigene-level table of counts was obtained with CORSET v.1.04 (Davidson & Oshlack, 2014), by filtering out any transcripts with fewer than 18 reads aligning over the nine samples. CORSET uses multimapping reads to hierarchically cluster similar transcripts based on their expression patterns and proportion of shared reads. Three different methods within BIOCONDUCTOR were used for DE analyses, DESEQ2 v. 1.6.3 (Anders & Huber, 2010), BAYSEQ v. 2.0.50 (Hardcastle & Kelly, 2010) and EDGER v. 3.8.6 (Robinson, McCarthy, & Smyth, 2010). Clusters were considered DE if they were detected in each of the three methods used at an FDR <0.05, in order to reduce the rate of false positives. Using a custom *python* script, the longest transcript from each CORSET cluster was extracted and used for further GO enrichment analyses in BLAST2GO v.3.2.7 (Conesa & Götz, 2008). Fisher's exact tests were implemented at a threshold *p*-value of .01, using as the reference set the list of the longest contigs for each cluster in the data. Overrepresented GO terms were then summarized using REVIGO (http://revigo.irb.hr, Supek, Bošnjak, Škunca, & Šmuc, 2011), applying thinning based on semantic similarity with the *SimRel* algorithm (Schlicker, Domingues, Rahnenführer, & Lengauer, 2006).

2.5 | Small RNA analysis

For the same *Dactylorhiza* individuals used in the RNA sequencing experiment (i.e., including individual fA6, but excluding iB6, for which not enough leaf material was still available at this stage), we also sequenced leaf sRNAseq libraries as multiplexed 50-bp single-end Illumina libraries (see Appendix S1 and Table 1). After demultiplexing and adapter removal, we selected for each accession the reads with lengths from 20 to 22 nucleotides (hereafter "mi/tasiRNA") and, separately, of 24 nucleotides (hereafter "siRNA") and analysed them as two classes. Length-selected sRNA reads were mapped to the *Dactylorhiza* reference transcriptome using STAR v2.5.1b (Dobin et al., 2013) not allowing for any mismatch (–outFilterMismatchNmax 0) and discarding any read mapping to more than one contig (–outFilterMultimapNmax 0). These settings for mapping on the *Dactylorhiza* reference transcriptome were chosen to include only unambiguous signals in further analyses. A replicated analysis allowing for one mismatch produced consistent results. Read counts per contig were further summarized in CORSET with default settings and the resulting tables of counts were then imported into R. Those contigs showing less than 10 reads in at least three individuals in one of the two species were filtered out. The function *betweenLaneNormalization* in the

R package EDASEQ (BIOCONDUCTOR) was used for sequencing depth normalization among samples applying a nonlinear full quantile method. A further normalization step based on a factor analysis of putative nondifferentially regulated genes was performed. First, a set of non-differentiated empirical control genes was produced in silico by running a preliminary analysis in EDGER (Robinson et al., 2010) and selecting all genes scored as undifferentiated (FDR > 0.95). This list of empirical genes was then used as negative control in the *RUVg* normalizing function in the R package RUVSEQ (Risso, Ngai, Speed, & Dudoit, 2014) removing $k = 2$ factors of unwanted variation ($k = 1$ or 3 were also tested). Results of the applied normalization were checked by plotting the relative log expression (*plotRLE* function) and the PCA (*plotPCA* function) across samples using the R package EDASEQ. The model resulting from the *RUVg* analysis was then used in EDGER, together with the group assignation, to implement the design of the differential regulation analysis between *D. fuchsii* and *D. incarnata*. The *glmLRT* function was used in EDGER and a FDR <0.05 was applied as a threshold for identifying differentially regulated transcripts. Differentially overregulated contigs were used in GO enrichment analyses in BLAST2GO. All transcriptome contigs were set as a reference in Fisher's exact tests and a *p*-value of .01 was applied as a threshold. Overrepresented GOs were then visualized in *REVIGO* as explained above.

2.6 | Assessment of ecological divergence

We finally aimed to characterize interspecific ecological divergence and place our transcriptomic results in this context. Locality information was collected from *Global Biodiversity Information Facility* (GBIF) using the R package RGBIF (available from https://cran.r-project.org/web/packages/rgbif). For robustness, we restricted the data set to localities with known herbarium vouchers. A few additional localities from our personal observations were also added. The geographical information was converted into decimal degrees and cleaned by removing exact duplicate localities. Locally dense sampling was finally reduced by thinning the records to one per 10-km^2 grid cell. Based on occurrence information at 298 localities for *D. fuchsii* and 393 for *D. incarnata*, spatial environmental data were extracted with RASTER (available from https://cran.r-project.org/web/packages/raster) from Bioclim (www.worldclim.org/bioclim) and Landsat Tree Cover (http://glcf.umd.edu/data/landsatTreecover/). We removed highly correlated bioclimatic variables in the obtained data set using a threshold of 0.80 using the package CARET (available from https://cran.r-project.org/web/packages/caret). This retained only seven of the 19 Bioclim variables for further analyses: annual mean temperature (BIO1), mean diurnal temperature range (BIO2), isothermality (BIO3), mean temperature of wettest quarter (BIO8), annual precipitation (BIO12), precipitation seasonality (BIO15) and precipitation of warmest quarter (BIO18). Ecological niche overlap between the two species was tested with ECOSPAT v.1.1 (available from https://cran.r-project.org/web/packages/ecospat). This method calculates the kernel smoothing densities of species occurrences and climatic variables along environmental axes from a PCA. Then, niche overlap along these axes was assessed

with three statistical tests: niche similarity/equivalency (Broennimann et al., 2012), Schoener's D metric (Schoener, 1970) and a niche divergence test (McCormack, Zellmer, & Knowles, 2010) based on the differences in PCA scores between species in comparison with the differences in scores for distinct "background regions" for each species. For these tests, we extracted climate data from 2,000 randomly selected points (background region). These points were randomly selected from polygons around the projected coordinate occurrences with a 10-km buffer. For the null distribution of D statistics, we used 1,000 replicates and a grid environmental value of 100. For the differences in PCA scores in the niche divergence test, to calculate the 95% confidence interval, we used a resampling approach based on 1,000 replicates of 500 samples for each species' background region.

Finally, soil pH at ca. 7 cm depth near 45 *D. incarnata* and 19 *D. fuchsii* plants has been measured with a Direct Soil pH Meter HI99121 (Hanna Instruments) across nine and, respectively, five populations from Britain, Scandinavia, Alps and Pyrenees between 2010 and 2015. The data have been summarized with box plots, and significance of the differences between the pH preferences of the two species has been tested with a *t* test in R.

3 | RESULTS

3.1 | Variation analyses and rates of sequence evolution

Mapping efficiency on the *Orchis* reference transcriptome ranged from 77% (fP1) to 87% (fB5) in *D. fuchsii* and from 82% (iB0) to 87% (iA6) for *D. incarnata* individuals. Initially, the GATK (McKenna et al., 2010) variant calling pipeline identified 23,185 indels and 727,350 SNPs in the data. After retaining only SNPs and implementing a set of variant filters in GATK, we retained 682,118 high-quality SNPs, of which 11,583 were multiallelic across the *Dactylorhiza–Orchis* group. The average transition/transversion rate found was 1.61. Heterozygosity was higher in *D. fuchsii* individuals than in *D. incarnata* (Figure 1), resulting in increased inbreeding coefficient F in *D. incarnata* (Table 2).
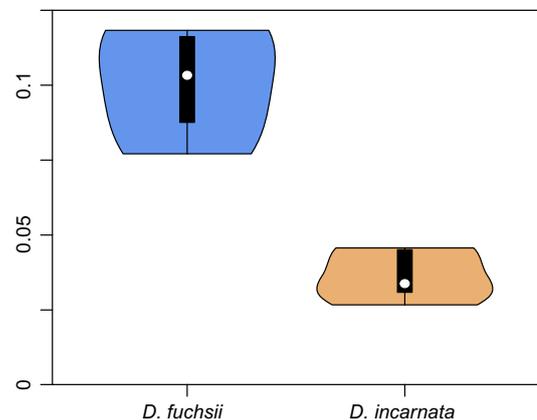

FIGURE 1 Violin plot of the proportions of heterozygous SNPs per individual among the 682,118 high-quality, filtered variant positions

TABLE 2 Summary of cSNP diversity measures in the *Dactylorhiza* transcriptomes

	H_O	F	P	π
fB4	0.12	0.73	0.40	
fB5	0.09	0.79	0.24	
fP1	0.08	0.81	0.16	
fP7	0.12	0.72	0.25	
D. fuchsii	**0.1**	**0.76**	**0.26**	**.077**
iA6	0.03	0.92	0.06	
iB6	0.05	0.89	0.19	
iB0	0.03	0.92	0.14	
iS4	0.03	0.94	0.11	
iS8	0.05	0.89	0.19	
D. incarnata	**0.04**	**0.91**	**0.14**	**.037**

H_O, observed proportion of heterozygous SNPs; F, per-individual inbreeding coefficient; P, ratio of private variants (i.e., singletons or doubletons) to total number of variant sites; π, nucleotide diversity per species averaged over all biallelic SNP positions with more than two genotypes present per group. Species values are given in bold.

Between individuals, diversity measures and the number of private alleles were also higher in *D. fuchsii* individuals than those in *D. incarnata* (Table 2). A maximum-likelihood phylogenetic tree (Fig. S3) and a PCA based on cSNPs (Fig. S4a) separated clearly the two species. Less significant, but clear geographical patterns were also visible.

The positive selection analyses included 5,082, 3,290 and 2,923 indel-free pairwise CDS alignments longer than 150 bp for *D. fuchsii–O. italica*, *D. incarnata–O. italica* and *D. fuchsii–D. incarnata*, respectively (Table 3). The average values of K_s and K_a, as indicated by KAKS_CALCULATOR (Zhang et al., 2006) analyses, were similar for both *Dactylorhiza–Orchis* comparisons (i.e., $K_s = 0.1$ and, respectively, $K_a = 0.02$), whereas for the *D. fuchsii–D. incarnata* comparison lower values have been obtained as expected because of their more recent divergence ($K_s = 0.062$, $K_a = 0.014$; Fig. S5). Based on the K_s values and the divergence time of ca. 16.75 Ma previously estimated for the split *Dactylorhiza–Orchis* (Givnish et al., 2015; Inda, Pimentel, & Chase, 2012), the ancestral lineages of the two *Dactylorhiza* species split around 10.4 Ma. Between 61% and 67% of transcripts with alignments showed signs of purifying selection (i.e., $K_a/K_s < 1$,

TABLE 3 Summary of the KAKS_CALCULATOR analysis of rates of sequence evolution in pairwise alignments over 150 bp of the *Dactylorhiza* species and *O. italica*

	Alignments	AvKs	AvKa	Purifying	Positive
D. fuchsii–O. italica	5,082	0.096	0.022	3,416	79
D. incarnata–O. italica	3,290	0.096	0.021	2,059	66
D. fuchsii–D. incarnata	2,923	0.062	0.014	1,790	41

AvKs, average synonymous substitution rate, AvKa, average nonsynonymous substitution rate, purifying and positive – the number of putative CDS showing K_a/K_s <1 and, respectively, >1 at an FDR < 0.1.

FDR < 0.1; Table 3). For the pair *D. fuchsii–D. incarnata*, 41 putative CDS were inferred to be under positive selection by the approximate MYN method implemented in KAKS_CALCULATOR (i.e., $K_a/K_s > 1$, FDR < 0.1), of which 35 (Appendix S2) were confirmed (i.e., ω > 1, $p < .05$) with the maximum-likelihood approach implemented in PAML (Yang, 2007). Among them, CODEML branch-site model tests have indicated 18 CDS to be under positive selection on the *D. fuchsii* branch and 14 on the *D. incarnata* branch. Enrichment tests found for the *D. fuchsii* and *D. incarnata* branches 27 and 56 enriched GO terms, respectively (Appendix S2), related mostly to response to stress, primarily biotic, but also abiotic (Figure 2, Fig. S6).

3.2 | Abundance estimation and DE analyses

The table of counts included 41,851 retained clusters (i.e., unigenes). A PCA (Fig. S4b) drawn with EDGER confirmed that gene expression variation between the two *Dactylorhiza* species explained the largest principal component (70.5% of the total variation), but some geographical structure was also evident, in particular in *D. incarnata*. Reflecting the substantial interspecific expression variation, DE tests indicated a significant difference in expression between the two *Dactylorhiza* species (Fig. S7), ranging from 14,165 clusters (EDGER) to 14,866 clusters (BAYSEQ). A total of 13,157 DE clusters were present in each of the three tests at an FDR < 0.05; of these, 47% were found to be overexpressed in *D. fuchsii*. Some of these genes have not yet been characterized, but Fisher's exact tests on annotated genes identified 184 enriched GO terms with overexpressed elements in *D. fuchsii* and 133 in *D. incarnata* (Appendix S3). The GO terms that had elements affected by DE were of a broad spectrum (Figure 3, Fig. S8): photosynthesis, chlorophyll and pigment binding, ribosome biogenesis, generation of precursor metabolites and energy, protein–chromophore linkage, structural molecule activity and tetrapyrrole binding were some of the significantly enriched GO terms in upregulation of both *D. fuchsii* and *D. incarnata*, but representing different transcripts. Enriched GO terms specifically overexpressed in *D. fuchsii* included, for example, response to temperature, carbohydrate storage, systemic acquired resistance, organonitrogen metabolism and cysteine biosynthetic process (Figure 3a, Fig. S8a). For *D. incarnata*, specific enriched overexpressed GO terms included a variety of fluid transmembrane transport-related GO terms, light harvesting in photosystem I, carbohydrate catabolic processes, DNA integration, RNA-dependent DNA polymerase activity, but also, perhaps importantly, the induced systemic resistance, detection of stimulus and nitric oxide signal transduction (Figure 3b, Fig. S4b).

3.3 | Small RNA analysis

After selection by length, we retained an average of 3.1 million 20- to 22-nt-long reads per sample and an average of 1.5 million 24-nt-long reads per sample, respectively. Average mapping efficiency of these two categories of reads to the *Dactylorhiza* reference transcriptome was 30.1% (SD = 2.8%) and 21.3% (SD = 6.7%), respectively. This is in line with the expectation that an important part of
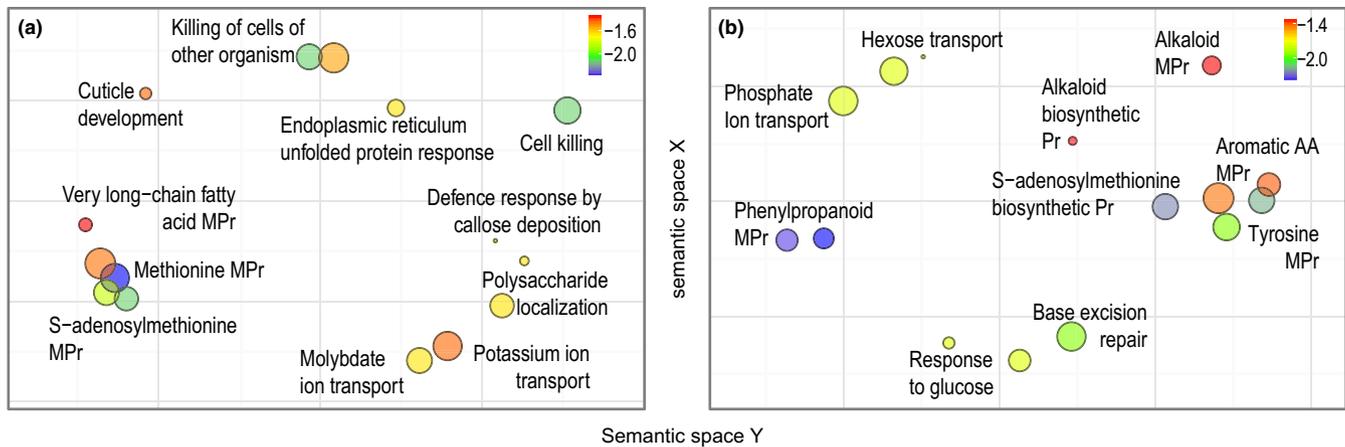
FIGURE 2 Enriched biological processes ($p < .01$) with elements targeted by positive selection in *D. fuchsii* (a) and in *D. incarnata* (b). Bubble size is proportional to the frequency of the respective term in the public GO database. The colour represents the log10 value of the significance of Fisher's exact tests of enrichment, corresponding to the indicated scale. Pr, process; MPr, metabolic process

sRNAs target genomic regions outside of the transcriptome (e.g., introns, repetitive elements and promoters), albeit some of the active repetitive elements and parts of introns can be represented in the reference transcriptome constructed from a total RNA library.

Clustering of the samples (Fig. S4c,d), based on the siRNAs and, respectively, the mi/tasiRNAs confirmed a significant divergence in regulation between the two species. The differential regulation tests resulted in 1,232 contigs overregulated by mi/tasiRNAs in *D. fuchsii* and 1,449 contigs in *D. incarnata* over a total of 9,848 transcripts passing initial filters. Enrichment tests identified 48 enriched GO terms with mi/tasiRNA overregulated elements in *D. fuchsii* and 38 in *D. incarnata*. DNA integration, DNA metabolism, nitrogen metabolism, RNA-directed DNA polymerase (RdDp) activity, nucleic acid binding, heterocyclic and organic cyclic compound binding are some of the significantly enriched GO terms in the set of contigs overregulated in *D. fuchsii* (Fig. S9). In *D. incarnata*, several of the enriched GO terms were related to organelle function (e.g., light harvesting in photosystem I, protein–chromophore linkage, chlorophyll and pigment binding) and some to iron transmembrane transport. Concerning the siRNA, 6,001 contigs passed initial filters and 855 and 1,045 contigs were overregulated in *D. fuchsii* and *D. incarnata*, respectively. Fisher's exact tests gave evidence for 35 GO terms (Fig. S10) with elements overregulated by siRNAs in *D. fuchsii* and 30 in *D. incarnata*. In this case, the most significant GO terms enriched in both species appeared related to DNA integration, DNA metabolism, nucleic acid metabolism, binding, nucleic acid binding, zinc ion binding, cation binding and heterocyclic and organic cyclic compound binding (Fig. S10). Several GO terms specifically overregulated in *D. fuchsii* included metabolism and modifications of small nuclear (snRNA), small nucleolar RNAs (snoRNA) and pseudouridine synthesis.

3.4 | Ecological differentiation

As Bioclim parameters have a resolution of 30 arc seconds (ca. 1 km^2), they are considered representative for the macrohabitat preference of each species. A test of climatic niche overlap

(Broennimann et al., 2012) indicated that overall, the (macro)habitats of the two species were similar ($D = 0.75$; similarity tests in both directions with $p < .01$), but an equivalency test ($p = .002$) indicated that the climatic niches of the two species were not completely identical. Accordingly, a divergence test of the climatic niches (McCormack et al., 2010) showed a significant differentiation on the PC2 niche axis (mean divergence = -0.36, null hypothesis = -0.21 to 0.13; paired samples Student's $t = -3.650$, $df = 624.5$, $p < .001$), summarizing annual mean temperature (BIO1), isothermality (BIO3), mean temperature of wettest quarter (BIO8), annual precipitation (BIO12) and precipitation seasonality (BIO15), but not on PC1 axis (mean divergence = 0.23, null hypothesis = 0.15–0.55; paired samples Student's $t = 2.437$, $df = 632.84$, $p = .01$) including mean diurnal temperature range (BIO2) and precipitation of warmest quarter (BIO18).

The tree cover data of 30-m resolution are considered here a microhabitat descriptor. Significant differences in tree cover were apparent between the localities occupied by the two species (Student's $t = 6.776$, $df = 641.11$, $p < .001$; Figure 4). Another microhabitat descriptor soil pH directly measured in the immediate vicinity of plants indicated a significantly divergent preference for soil acidity (Student's $t = -3.535$, $df = 62$, $p < .001$, Figure 4).

4 | DISCUSSION

Orchids make up the largest angiosperm family, but the molecular variation shaping this important biodiversity is still little understood (Givnish et al., 2015). Here, we have identified thousands of transcripts with distinct alleles, expression patterns and/or post-transcriptional regulation between two food-deceptive, terrestrial orchid species that are currently parapatric across Europe. We show that despite their generally shared distribution area, the two *Dactylorhiza* species inhabit distinct niches, as the microenvironmental conditions they prefer differ significantly with regard to soil acidity and tree cover (Figure 4), but also with respect to temperature evenness over
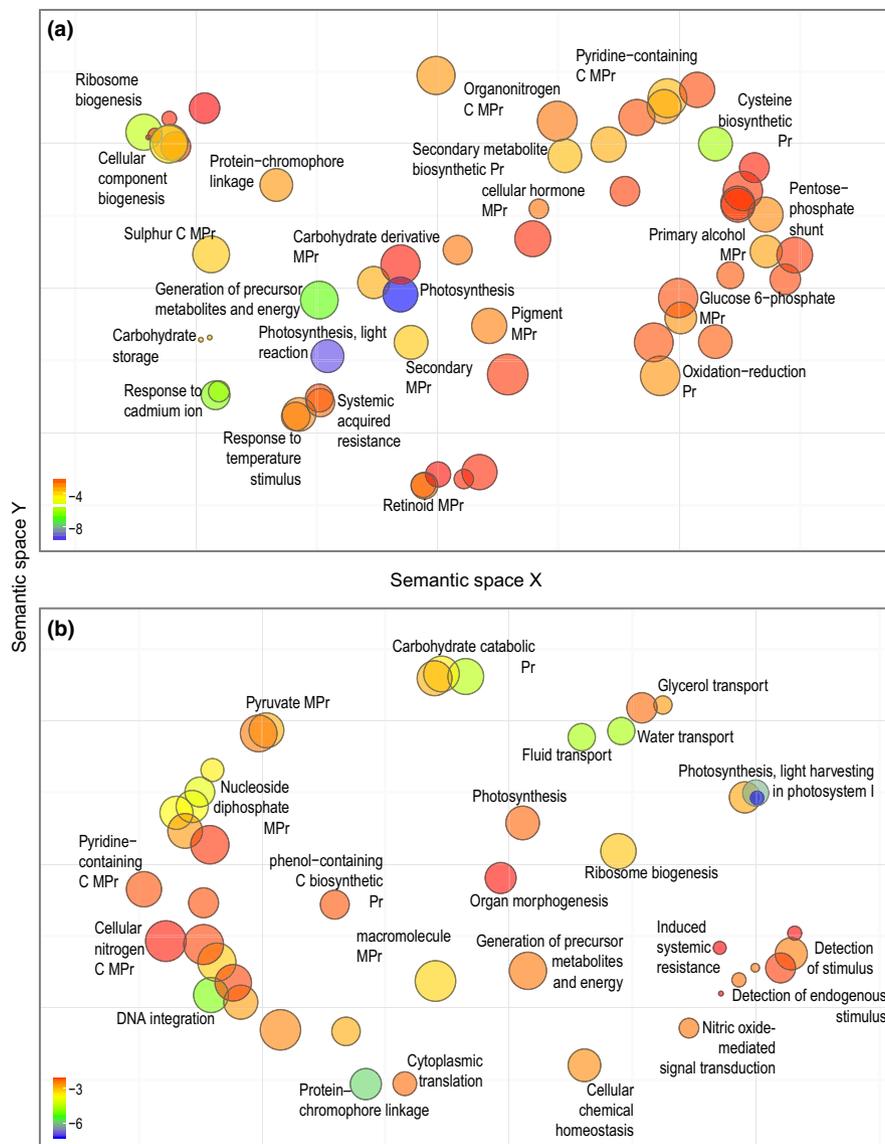
FIGURE 3 REVIGO visualization of the enriched biological processes ($p < .01$) that are affected by differential expression (FDR < 0.05) between *D. fuchsii* and *D. incarnata*. (a) Enriched GO terms of genes overexpressed in *D. fuchsii*. (b) Enriched GO terms of genes overexpressed in *D. incarnata*. Bubble size is proportional to the frequency of the respective term in the public GO database. The colour represents the log10 value of the significance of Fisher's exact tests of enrichment, corresponding to the indicated scale. Pr, process; C, compound; MPr, metabolic process

the year and annual precipitation. If maintained over generations, such deviating ecological preferences likely triggered distinct selection that in combination to specific demographic histories have moulded different genomic landscapes between the two species (Slotte, Foxe, Hazzouri, & Wright, 2010).

The current prevalence of *D. incarnata* within small, localized populations (i.e., subhabitat specialization) over a highly fragmented distribution resulting in a Wahlund effect (Hedrén & Nordström, 2009) is corroborated here with considerable levels of inbreeding of this taxon. In turn, *D. fuchsii*, which currently grows in larger, diffuse populations (Kull & Hutchings, 2006), exhibits higher levels of heterozygosity and, in general, higher genetic diversity (Table 2). We estimate the lineages leading to *D. fuchsii* and *D. incarnata* to have diverged in the second half of Miocene, most probably within the first million years after

Dactylorhiza split from its sister clade (Givnish et al., 2015). An early split of the clades of the two species has previously been demonstrated in phylogenetic (Bateman et al., 2003) and population genetic studies (Balao et al., 2016; Hedrén, 1996) that described significant genetic divergence between them. The two *Dactylorhiza* species studied here can therefore be considered to be representative for the depth of molecular variation in the genus, retaining effects of natural selection throughout their evolutionary histories.

4.1 | Protein sequence evolution, positive selection and biotic pressure

Nonsynonymous mutations may trigger novel functions and new phenotypes and are hence expected to be under strong selective
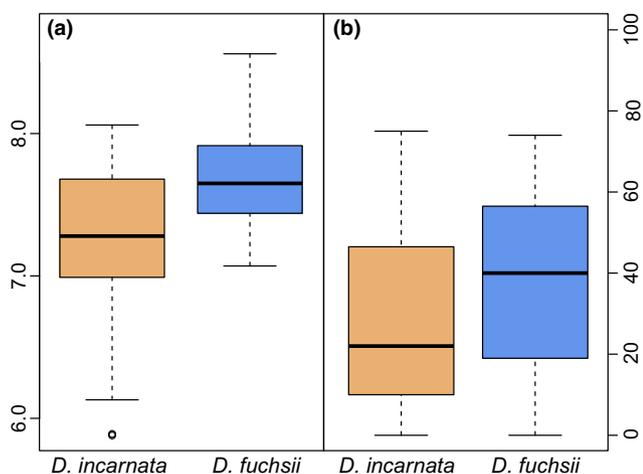
FIGURE 4 Ecological divergence between *D. fuchsii* and *D. incarnata*. (a) Preference for soil pH measured at 14 populations in Britain, Scandinavia, the Alps and Pyrenees. (b) Preference for tree cover of the two species. The difference in distribution in both cases is significant at $p < .001$

pressure, either purifying or positive. Along the deep histories of the two *Dactylorhiza* species, we observe significant effects of purifying selection, targeting close to 70% of genes. However, the estimated overall ratio of nonsynonymous to synonymous substitutions of 1:5 appears in line to those recorded for example between *Arabidopsis thaliana* and *A. lyrata* (diverged 5–10 Ma, Yang & Gaut, 2011) or *Gossypium arboreum* and *G. raimondii* (diverged 7–11 Ma, Senchina et al., 2003).

Our results further bring evidence that positive selection elevated the interspecific K_a/K_s ratios across at least 1.2% of orthologous genes. This estimate is, however, limited to a cohort of protein-coding sequences that were expressed in both *Dactylorhiza* species at the time of tissue sampling in the common garden and also had an ortholog in the *O. italica* reference. Hence, our inferences may be mainly representative for constitutively expressed genes/exons, rather than those transcribed only as direct response to stimuli or stressors. In addition, given the long history of the two species and therefore the substantial amount of synonymous substitutions that could have accumulated, our test for positive selection requires a fairly large number of codons per gene to be affected by positive selection in order for K_a/K_s to significantly exceed one. Therefore, the stringent approach we used likely uncovers rather extreme cases where positive selection is operating (Roth & Liberles, 2006).

Consistent with a dynamic evolution of protein sequences shaped by exposure to habitat-specific pathogens and herbivores (Brown & Tellier, 2011), the most frequent and overrepresented functions among the selected genes along the phylogenetic branches of the two *Dactylorhiza* species are linked to biotic responses. The type and strength of biotic stress imposed on plants depends heavily on both geography and environmental context, but also on host genotype and ability of plants to resist particular stressors (Tiffin & Moeller, 2006). For example, given the open versus more forested habitats specific for *D. incarnata* and *D. fuchsii*, respectively, each of these species had most probably to adapt to distinct herbivores. Plant responses to herbivores and pathogens (i.e., host–pathogens interactions, leading to pathogen recognition and host defence systems; Chisholm, Coaker, Day, & Staskawicz, 2006) are generally context dependent, moulded by the surrounding plant communities, local nutrient conditions and the plant metapopulation dynamics and structure, likely distinct in the long-term between the two species investigated here (Hedrén & Nordström, 2009; Pillon et al., 2007). Aiming to minimize frequency and impact of attack, the diverse defences that plants specifically evolved appear to include both physical and chemical adaptations. The specific genes that have been found to be under lineage-specific positive selection with our approach (see Section "2") and their GO term enriched in the respective test are discussed below.

On the *D. fuchsii* branch (Figure 2a, Fig. S6a), evidence for adaptive evolution is found for a homolog of *DEFENSIN J1-2*, known to inhibit the growth of pathogenic fungi by modifying their morphology or physiology, disrupting or killing exogenous cells (Thomma, Cammue, & Thevissen, 2002). Another candidate gene for adaptive evolution on the *D. fuchsii* branch is a homolog of *HAI1 PHOSPHATASE 2C 78*, which is involved in defence response mechanisms by deposition of callose, a polysaccharide produced as a response to wounding, but also in response to drought and salt stress (Zhang et al., 2013). *HAI1* is further involved in the pathway of the endoplasmic reticulum response to an accumulation of unfolded or misfolded proteins on stress (Wahyu Indra Duwi, Lee, & Lee, 2013). A *D. fuchsii* homolog of *TETRAKITIDE α-PYRONE REDUCTASE 2* also shows a significantly elevated K_a/K_s ratio. This is part of the *DIHYDROKAEMPFEROL 4-REDUCTASE* activity with a role in flavonoid biosynthesis, again potentially linked to plant defence. Finally, a homolog of *3-KETOACYL-SYNTHASE 10*, with a role in developing cuticular wax, is also evidenced as a target of positive selection on the *D. fuchsii* branch, potentially in response to viral pressure (Lolle et al., 1997).

On the other hand, along the history of *D. incarnata* (Figure 2b, Fig. S6b), positive selection acted on a *POLYPHENOL OXIDASE* homolog with a putative role in the activity of *CATECHOL OXIDASE* as a response to plant tissue damage, producing ortho-benzoquinones that further polymerize to form melanins, protecting plant wounds against further microbial attack (Queiroz, Mendes Lopes, Fialho, & Valente-Mesquita, 2008). Adaptive evolution also appears to have affected a *PRIMARY AMINE OXIDASE* homolog potentially involved in cell wall maturation and lignification during development, as well as wound healing and cell wall reinforcement during pathogen invasion (Cona, Rea, Angelini, Federico, & Tavladoraki, 2006). The *S-ADENOSYLMETHIONINE SYNTHETASE* orthologs appear to have undergone specific adaptive evolution, independently during both *D. fuchsii* and *D. incarnata* histories. S-adenosylmethionine acts as a universal methyl group donor in transmethylation reactions, but also serves as a precursor to the phytohormone ethylene, implicated in the control of numerous developmental processes, and regulating resistance to bacterial and fungal pathogens, nematodes and lepidopteran herbivores (Wang, Li, & Ecker, 2002).

Further adaptive evolution specific to the *D. incarnata* history has affected elements of phenylpropanoid biosynthesis, namely a *LACCASE* homolog with a putative role in formation of lignin and a *PHENYLALANINE AMMONIA-LYASE (PAL)* homolog potentially involved in alkaloid biosynthesis. Apart their role in defence mechanisms against herbivores and pathogens, phenylpropanoids have been shown, for example in the orchid *Phalaenopsis*, to enhance protection to UV light (Ali, Khatun, Hahn, & Paek, 2006), potentially relevant for *D. incarnata* given its preference for more open habitats compared to *D. fuchsii*. In this context, we also detected adaptive evolution along the *D. incarnata* branch of a homolog of *TRANS-MEMBRANE 9 SUPERFAMILY MEMBER 4* with a putative role in base excision repair, a cellular DNA repair mechanism.

Elevated evolution of further protein-coding sequences potentially related to abiotic-imposed selection in *D. incarnata* includes a *SUGAR CARRIER C* homolog, potentially involved in hexose transmembrane transport likely linked to hypoxia (Gharbi, Ricard, Smiti, Bizid, & Brouquisse, 2009) that the roots of *D. incarnata* may suffer, given the high soil moisture to which this species adapted (Paun et al., 2011). Preferences for distinct soil chemistry (i.e., organic vs. mineral soil typically preferred by *D. incarnata* and *D. fuchsii*, respectively) and soil pH may be linked to elevated rates of evolution of an *INORGANIC PHOSPHATE TRANSPORTER 1-8* homolog in *D. incarnata* and of *POTASSIUM TRANSPORTER 9* in *D. fuchsii*. Finally, a *MOLYBDATE TRANSPORTER 2* homolog underwent adaptive evolution along the *D. fuchsii* history. Molybdate is an essential plant micronutrient that regulates, for example, nitrate reduction to nitrite and plays a role in redox reactions, and its availability and efficiency of its uptake from soil are highly dependent on soil pH (Tomatsu et al., 2007). Altogether, these candidates open up the ground for further research to identify the specific functional impact of codon evolution, its physiological implications and potential transferability of defence gene variants to cultivated orchids.

4.2 | Divergent regulation of expression and abiotic preference

We further explored differentiation in gene expression rates between the two *Dactylorhiza* species in a common garden setting. Our results estimate around a third of the identified orthologous genes show significant DE. This extensive expression divergence could be the result of both adaptation and drift and appears mid-range compared to other pairs of plant species that diverged within a similar time frame (*Brassica rapa* and *B. oleracea* diverged ca. 13 Ma and show ca. 22% DE, Jiang et al., 2013; e.g., *Gossypium* species diverged 5–10 Ma and show up to 53% DE, Rapp, Udall, & Wendel, 2009).

The uncovered DE between the two orchid species affected genes involved in various biosynthetic and metabolic processes—with several GO term enrichments that may be consistent with adaptation (i.e., long-term constitutive expression divergence) and/or acclimation to multiple abiotic (and biotic) stresses, some of them likely specific for the conditions of the common garden. We further exemplify gene ontology terms of a potential ecological relevance,

which have been found to be enriched within the group of DE genes.

The shaded common garden mimicked more closely the light preferences of *D. fuchsii*. Light is one of the most important environmental factors for plants, providing an energy source, but also governing their developmental programme (Neff, Fankhauser, & Chory, 2000). Our tree cover analysis shows a significant difference in preference for the openness of the habitats for *D. incarnata* and *D. fuchsii* (Figure 4b). The need to enhance efficiency of light absorption corroborates an observed overexpression in *D. incarnata* of 14 chlorophyll binding transcripts, part of the light harvesting pathway in photosystem I, which is the most significantly enriched biological process ($p = 6.2 \times 10^{-8}$) for upregulation in this species (Figure 3b). This further correlates to a significant increase in mi/tasiRNAs targeting of this pathway in *D. incarnata* (i.e., affecting five chlorophyll binding transcripts; enrichment $p = 3.1 \times 10^{-4}$). This may indicate a dynamic regulation of acclimation to shaded conditions in *D. incarnata*. Further gene expression differences potentially increasing photosynthetic efficiency, this time in the way light is used to reduce molecules in light reaction, are visible in 63 transcripts overexpressed in *D. fuchsii* (or with decreased expression in *D. incarnata*, corresponding also to increased targeting by mi/tasiRNAs of 15 transcripts in this pathway, enriched at $p = 1.1 \times 10^{-4}$). Broadly, transcripts with a putative role in photosynthesis appear to have higher expression (Figure 3a; enriched at $p = 1.6 \times 10^{-10}$) and to be less targeted by mi/tasiRNAs (Fig. S9b; enriched at $p = 6.1 \times 10^{-7}$) in *D. fuchsii* in comparison with *D. incarnata*—a pattern that may mirror the long-term preference of the former for more shaded environments.

In turn, the location of the common garden has likely featured rather alien temperature conditions for *D. fuchsii* (i.e., higher temperature in summer, limited snow cover in winter), given that no population of this species grows within at least 50 km radius around the city, whereas some *D. incarnata* stands are present even within the city boundary (e.g., compare distributions on www.gbif.at). Indeed, for *D. fuchsii* we identified a significant overexpression of 55 annotated transcripts putatively involved in a response to temperature (enriched at $p = .0011$). This abiotic stress can trigger misfolding of newly synthesized proteins and denaturation of existing proteins (Barnabas, Jäger, & Feher, 2008). Genes known to be upregulated as response to high temperature stress (e.g., Altenbach, 2012) and having a homolog found to be overexpressed in *D. fuchsii* include, for example, heat-shock proteins (i.e., *HSP70 kDa17* and *DNAJ HSP40* homologs), a *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE* (*GADPH A*), a *GALACTINOL SYNTHASE* and a *TRIOSEPHOSPHATE ISOMERASE* homolog involved in glycolysis. The uncovered ecological divergence in bioclimatic variables such as annual mean temperature (BIO1) is congruent with a significant divergence in expression of genes related to a temperature response.

4.3 | Genome stability and small RNA regulation

Gene expression is generally governed by a complex combination of regulators, including among others transcription factors, splicing

factors and miRNAs (Warnefors & Eyre-Walker, 2011). The comparative analysis of sRNA profiles confirms that the extensive differences in gene expression between the *Dactylorhiza* species are complemented by divergent post-transcriptional regulations. miRNAs function as negative regulators of gene expression via translational repression or mRNA cleavage (Borges & Martienssen, 2015) and are known to mediate response to genomic or environmental stresses (Ding, Tao, & Zhu, 2013; Ivashuta et al., 2011). From the transcripts that were targeted by miRNAs and/or tasiRNAs, 27.2% were found to be differentially regulated between the two orchid species, with deviating post-transcriptional regulation of metabolism (e.g., nucleic acids, aromatic compounds, tetrapyrrole and alkaloid metabolism targeted in *D. fuchsii*, and DNA and tetrapyrrole metabolism targeted in *D. incarnata*) and binding (Fig. S9). As discussed above, we observed an increased targeting by mi/tasiRNA of photosynthetic pathways in *D. incarnata*, potentially as an acclimation response to the shaded conditions in the common garden. Likely in connection to its photosynthetic activity, the starch transitory breakdown including maltose metabolism is also overregulated in *D. incarnata*. Starch represents an overflow of photosynthetically fixed carbon that is exported from chloroplasts at night, predominantly as maltose (Lu & Sharkey, 2006). Inhibition of maltose metabolism slows down starch degradation, which has been reported to be a response to changes in photoperiod length, circadian clock and temperature (Lu & Sharkey, 2006).

Importantly, our results point to distinct strategies and efficiency in the maintenance of genome stability between the two orchid species. In the RNAseq data, RdDp activity (enriched at $p = .0016$; with 60 transcripts overexpressed, representing in particular Ty3-gypsy subclass), mainly associated with retrotransposons and retroviruses, together with the process of DNA integration (193 transcripts, enriched at $p = 9.1 \times 10^{-6}$) appear to be overexpressed in *D. incarnata* (Figure 3, Fig. S8). Moreover, the RdDp activity shows increased control by mi/tasiRNAs in *D. fuchsii* (20 transcripts, enriched at $p = .003$; Fig. S9c). These suggest a significant increase in activity of class I TEs in *D. incarnata* compared to *D. fuchsii*, which over time may have triggered their difference in genome size. In this context, it is notable that the two species exhibit different effective population sizes (Balao et al., 2016), reflected in higher inbreeding in *D. incarnata* that may correlate also with less efficient purging of active insertions of TEs.

On the other hand, we also detect divergent regulation of 31.7% of the transcripts that are targeted by siRNAs in the two *Dactylorhiza* species. GO enrichment analyses of overregulated siRNAs for each species revealed that the most meaningful biological processes appear related to DNA integration, DNA metabolism and other processes related to nucleic acid binding (Fig. S10). These biological processes are congruent with the presumed defence function of siRNAs in plants against the proliferation of endogenous TEs or invading viruses through transcriptional or post-transcriptional gene silencing (Carrington & Ambros, 2003). However, siRNAs could also indirectly affect gene expression of adjacent genes in the proximity of TEs (Hollister et al., 2011). Thus, changes in expression in sRNAs could

ultimately be adaptive and evolve quickly under natural selection (Chapman & Carrington, 2007).

4.4 | Implications

Altogether, the distinct responses to selection accumulated during the history of the two parapatric *Dactylorhiza* species investigated here appear governed by their divergent microenvironmental context, in which biotic and abiotic pressures act synergistically to shape transcriptome structure, expression and regulation. Apart from adaptation to divergent abiotic conditions, distinct habitat affiliation of nascent species requires divergent adaptation to abiotic conditions, if they exceed possibilities granted by phenotypic plasticity. However, distinct microhabitat preferences readily require divergent defence strategies in order to detect and respond to specific pathogens and herbivores (e.g., Turcotte, Davies, Thomsen, & Johnson, 2014). The divergent mechanisms of biotic responses apparent between the two *Dactylorhiza* species both at the level of coding sequence evolution, but also in the rates of expression across the systemic acquired resistance (upregulated in *D. fuchsii*, enriched at $p < .005$) and the induced systemic resistance (overexpressed in *D. incarnata*, enriched at $p < .01$), are remarkable given their obligate orchid–mycorrhiza interactions (Tupac Otero & Flanagan, 2006). Modulation of plant defence responses could have detrimental secondary effects on their symbiotic relationship. On the other hand, the plant mechanism of mycorrhiza recognition may constitutively activate systemic plant immune responses (Pozo, Verhage, García-Andrade, García, & Azcón-Aguilar, 2009). Such activation can lead to a primed state of the plant that allows for a more efficient defence mechanisms in response to attack by potential enemies. Although *Dactylorhiza* species associate with a wide range of common and generalist mycorrhizal fungi, these two species show an incipient differentiation in mycorrhizal spectra (Jacquemyn, Deja, De hert, Cachapa Bailarote, & Lievens, 2012).

The detailed transcriptome knowledge for *Dactylorhiza* gathered in this study provides critical resources to understand evolution after hybridization and whole-genome duplication, phenomena especially frequent in this genus. In particular, the hybridization of the two species analysed here iteratively formed a myriad of morphologically and ecologically distinct sibling allopolyploids (Pillon et al., 2007). We demonstrate here an extensive molecular divergence between the two diploids, which suggests an extensive genomic and transcriptomic shock in their hybrids (McClintock, 1984) and offers potential insights into the difficulties of coexistence for these two genomes at the homoploid level (Paun, Forest, Fay, & Chase, 2009). In turn, whole-genome duplication was proposed to alleviate the transcriptomic shock of hybridization between divergent progenitors (Hegarty et al., 2006). The present study hence contributes useful resources and background information to further explore allelic interactions between parental genomes when combined in a single duplicated nucleus. This notably leads to altered reprogramming of gene expression, including

dominance and transgressive patterns that may promote metabolic and developmental changes, stress tolerance and increased fitness in allopolyploids (Chen, 2013).

DATA ACCESSIBILITY

The raw RNAseq data have been deposited into the Short Read Archive (SRA) of the NCBI under Accession nos. SRR3330392–SRR3330400 and SRR3330403–SRR3330404. The raw sRNA data files have been deposited into the NCBI SRA under Accession nos. SRR3331132–SRR3331133, SRR3331136, SRR3331143, SRR3331149–SRR3331151, SRR3331195 and SRR5365231. The Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the Accession nos. GFHX00000000 and GFJV00000000.

AUTHOR CONTRIBUTIONS

OP designed the research. F.B., M.H. and O.P. collected the sample material. M.T.L. and J.B. carried out the wet laboratory work with advice from O.P. and L.S. B.H.H. assembled the reference genome. F.B. and T.W. carried out the D.E. analyses, with input from F.A. O.P. performed the cSNP analyses, with input from C.K. E.T. performed the sRNA analyses. F.B. carried out the ecological analyses. M.W.C. and M.H. contributed to data interpretation. O.P., F.B. and E.T. drafted the manuscript. All authors revised and approved the final manuscript.

REFERENCES

Aagaard, S. M., Sastad, S. M., Greilhuber, J., & Moen, A. (2005). A secondary hybrid zone between diploid *Dactylorhiza incarnata* ssp. *cruenta* and allotetraploid *D. lapponica* (Orchidaceae). *Heredity*, *94*, 488–496.

Ackerman, J. D., Trejo-Torres, J. C., & Crespo-Chuy, Y. (2007). Orchids of the West Indies: Predictability of diversity and endemism. *Journal of Biogeography*, *34*, 779–786.

Ali, M. B., Khatun, S., Hahn, E.-J., & Paek, K.-Y. (2006). Enhancement of phenylpropanoid enzymes and lignin in Phalaenopsis orchid and their influence on plant acclimatisation at different levels of photosynthetic photon flux. *Plant Growth Regulation*, *49*, 137–146.

Altenbach, S. B. (2012). New insights into the effects of high temperature, drought and post-anthesis fertilizer on wheat grain development. *Journal of Cereal Science*, *56*, 39–50.

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*, R106.

Anderson, J. T., Wagner, M. R., Rushworth, C. A., Prasad, K. V. S. K., & Mitchell-Olds, T. (2014). The evolution of quantitative traits in complex environments. *Heredity*, *112*, 4–12.

Balao, F., Tannhäuser, M., Lorenzo, M. T., Hedren, M., & Paun, O. (2016). Genetic differentiation and admixture between sibling allopolyploids in the *Dactylorhiza majalis* complex. *Heredity*, *116*, 351–361.

Barnabas, B., Jäger, K., & Feher, A. (2008). The effect of drought and heat stress on reproductive processes in cereals. *Plant, Cell & Environment*, *31*, 11–38.

Bateman, R. M., Hollingsworth, P. M., Preston, J., Luo, Y. B., Pridgeon, A. M., & Chase, M. W. (2003). Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (Orchidaceae). *Botanical Journal of the Linnean Society*, *142*, 1–40.

Benzing, D. H. (1986). The genesis of orchid diversity: Emphasis on floral biology leads to misconceptions. *Linleyana*, *1*, 73–89.

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

Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., & Yoccoz, N. G., . . . Guisan, A. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, *21*, 481–497.

Brown, J. K. M., & Tellier, A. (2011). Plant-parasite coevolution: Bridging the gap between genetics and ecology. *Annual Review of Phytopathology*, *49*, 345–367.

Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W. -C., & Liu, K. -W., . . . Liu, Z. -J. (2015). The genome sequence of the orchid *Phalaenopsis equestris*. *Nature Genetics*, *47*, 65–72.

Carrington, J. C., & Ambros, V. (2003). Role of MicroRNAs in plant and animal development. *Science*, *301*, 336–338.

Chapman, E. J., & Carrington, J. C. (2007). Specialization and evolution of endogenous small RNA pathways. *Nature Reviews Genetics*, *8*, 884–896.

Chase, M. W., Cameron, K. M., Freudenstein, J. V., Pridgeon, A. M., Salazar, G., vanden Berg, C., & Schuiteman, A. (2015). An updated classification of Orchidaceae. *Botanical Journal of the Linnean Society*, *177*, 151–174.

Chen, Z. J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. *Nature Reviews Genetics*, *14*, 471–482.

Chisholm, S. T., Coaker, G., Day, B., & Staskawicz, B. J. (2006). Host-microbe interactions: Shaping the evolution of the plant immune response. *Cell*, *124*, 803–814.

Cona, A., Rea, G., Angelini, R., Federico, R., & Tavladoraki, P. (2006). Functions of amine oxidases in plant development and defence. *Trends in Plant Science*, *11*, 80–88.

Conesa, A., & Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*, *2008*, 12.

Cozzolino, S., & Widmer, A. (2005). Orchid diversity: An evolutionary consequence of deception? *Trends in Ecology & Evolution*, *20*, 487–494.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., & DePristo, M. A., . . . Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

Darwin, C. (1877). *The various contrivances by which orchids are fertilized by insects*. London: John Murray.

Davidson, N. M., & Oshlack, A. (2014). Corset: Enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biology*, *15*, 410.

De Paolo, S., Salvemini, M., Gaudio, L., & Aceto, S. (2014). *De novo* transcriptome assembly from inflorescence of *Orchis italica*: Analysis of coding and non-coding transcripts. *PLoS One*, *9*, e102155.

De Wit, P., Pespeni, M. H., & Palumbi, S. R. (2015). SNP genotyping and population genomics from expressed sequences—Current advances and future possibilities. *Molecular Ecology*, *24*, 2310–2323.

DePristo, M. A., Banks, E., Poplin, R. E., Garimella, K. V., Maguire, J. R., & Hartl, C., . . . Daly, M. J. (2011). A framework for variation discovery

and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43, 491–498.

Devos, N., Raspe, O., Oh, S. H., Tyteca, D., & Jacquemart, A. L. (2006). The evolution of *Dactylorhiza* (Orchidaceae) allotetraploid complex: Insights from nrDNA sequences and cpDNA PCR-RFLP data. *Molecular Phylogenetics and Evolution*, 38, 767–778.

Ding, Y., Tao, Y., & Zhu, C. (2013). Emerging roles of microRNAs in the mediation of drought stress response in plants. *Journal of Experimental Botany*, 64, 3077–3086.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., & Jha, S., . . . Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15–21.

Ehrenreich, I. M., & Purugganan, M. D. (2008). Sequence variation of MicroRNAs and their binding sites in *Arabidopsis*. *Plant Physiology*, 146, 1974–1982.

Engström, P. G., Steijger, T., Sipos, B., Grant, G. R., Kahles, A., & Rätsch, G., . . . Bertone, P. (2013). Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature Methods*, 10, 1185–1191.

Gharbi, I., Ricard, B., Smiti, S., Bizid, E., & Brouquisse, R. (2009). Increased hexose transport in the roots of tomato plants submitted to prolonged hypoxia. *Planta*, 230, 441–448.

Givnish, T. J., Spalink, D., Ames, M., Lyon, S. P., Hunter, S. J., & Zuluaga, A., . . . Cameron, K. M. (2015). Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society of London B: Biological Sciences*, 282, 20151553.

Gravendeel, B., Smithson, A., Slik, F. J. W., & Schuiteman, A. (2004). Epiphytism and pollinator specialization: Drivers for orchid diversity?. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 359, 1523–1535.

Ha, M., Pang, M., Agarwal, V., & Chen, Z. J. (2008). Interspecies regulation of microRNAs and their targets. *Biochimica et Biophysica Acta*, 1779, 735–742.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., & Bowden, J., . . . Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8, 1494–1512.

Hardcastle, T. J., & Kelly, K. A. (2010). baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics*, 11, 422.

Hedrén, M. (1996). Genetic differentiation, polyploidization and hybridization in northern European *Dactylorhiza* (Orchidaceae): Evidence from allozymes markers. *Plant Systematics and Evolution*, 201, 31–55.

Hedrén, M., Fay, M. F., & Chase, M. W. (2001). Amplified fragment length polymorphisms (AFLP) reveal details of polyploid evolution in *Dactylorhiza* (Orchidaceae). *American Journal of Botany*, 88, 1868–1880.

Hedrén, M., & Nordström, S. (2009). Polymorphic populations of *Dactylorhiza incarnata* s.l. (Orchidaceae) on the Baltic island of Gotland: Morphology, habitat preference and genetic differentiation. *Annals of Botany*, 104, 527–542.

Hegarty, M. J., Barker, G. L., Wilson, I. D., Abbott, R. J., Edwards, K. J., & Hiscock, S. J. (2006). Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Current Biology*, 16, 1652–1659.

Hollister, J. D., Smith, L. M., Guo, Y.-L., Ott, F., Weigel, D., & Gaut, B. S. (2011). Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 2322–2327.

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., . . . Morgan, M. (2015). Orchestrating high-throughput genomic analysis with bioconductor. *Nature Methods*, 12, 115–121.

Inda, L. A., Pimentel, M., & Chase, M. W. (2012). Phylogenetics of tribe Orchideae (Orchidaceae: Orchidoideae) based on combined DNA

matrices: Inferences regarding timing of diversification and evolution of pollination syndromes. *Annals of Botany*, 110, 71–90.

Ivashuta, S., Banks, I. R., Wiggins, B. E., Zhang, Y., Ziegler, T. E., Roberts, J. K., & Heck, G. R. (2011). Regulation of gene expression in plants through miRNA inactivation. *PLoS One*, 6, e21330.

Jacquemyn, H., Deja, A., De hert, K., Cachapa Bailarote, B., & Lievens, B. (2012). Variation in mycorrhizal associations with tulasnelloid fungi among populations of five *Dactylorhiza* species. *PLoS One*, 7, e42212.

Jiang, J., Shao, Y., Du, K., Ran, L., Fang, X., & Wang, Y. (2013). Use of digital gene expression to discriminate gene expression differences in early generations of resynthesized *Brassica napus* and its diploid progenitors. *BMC Genomics*, 14, 72.

Kitano, J., Yoshida, K., & Suzuki, Y. (2013). RNA sequencing reveals small RNAs differentially expressed between incipient Japanese threespine sticklebacks. *BMC Genomics*, 14, 1–13.

Kull, T., & Hutchings, M. J. (2006). A comparative analysis of decline in the distribution ranges of orchid species in Estonia and the United Kingdom. *Biological Conservation*, 129, 31–39.

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28, 298–299.

Lolle, S. J., Berlyn, G. P., Engstrom, E. M., Krolikowski, K. A., Reiter, W. D., & Pruitt, R. E. (1997). Developmental regulation of cell interactions in the Arabidopsis fiddlehead-1 mutant: A role for the epidermal cell wall and cuticle. *Developmental Biology*, 189, 311–321.

Lu, Y. A. N., & Sharkey, T. D. (2006). The importance of maltose in transitory starch breakdown. *Plant, Cell & Environment*, 29, 353–366.

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

McCormack, J. E., Zellmer, A. J., & Knowles, L. L. (2010). Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation? Insights from tests with niche models. *Evolution*, 64, 1231–1244.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20, 1297–1303.

McManus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R., & Wittkopp, P. J. (2010). Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Research*, 20, 816–825.

Michalak, P. (2008). Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. *Heredity*, 102, 45–50.

Moccia, M. D., Widmer, A., & Cozzolino, S. (2007). The strength of reproductive isolation in two hybridizing food-deceptive orchid species. *Molecular Ecology*, 16, 2855–2866.

Neff, M. M., Fankhauser, C., & Chory, J. (2000). Light: An indicator of time and place. *Genes & Development*, 14, 257–271.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, 39, 197–218.

Paun, O., Bateman, R. M., Fay, M. F., Luna, J. A., Moat, J., Hedrén, M., & Chase, M. W. (2011). Altered gene expression and ecological divergence in sibling allopolyploids of *Dactylorhiza* (Orchidaceae). *BMC Evolutionary Biology*, 11, 113.

Paun, O., Forest, F., Fay, M. F., & Chase, M. W. (2009). Hybrid speciation in angiosperms: Parental divergence drives ploidy. *New Phytologist*, 182, 507–518.

Pillon, Y., Fay, M. F., Hedrén, M., Bateman, R. M., Devey, D. S., Shipunov, A. B., . . . Chase, M. W. (2007). Evolution and temporal diversification of western European polyploid species complexes in *Dactylorhiza* (Orchidaceae). *Taxon*, 56, 1185–1208.

Pozo, M., Verhage, A., García-Andrade, J., García, J., & Azcón-Aguilar, C. (2009). Priming plant defence against pathogens by arbuscular mycorrhizal fungi. In C. Azcón-Aguilar, J. Barea, S. Gianinazzi, & V.

Gianinazzi-Pearson (Eds.), *Mycorrhizas-functional processes and ecological impact* (pp. 123–135). Berlin Heidelberg: Springer.

Queiroz, C., Mendes Lopes, M. L., Fialho, E., & Valente-Mesquita, V. L. (2008). Polyphenol oxidase: Characteristics and mechanisms of browning control. *Food Reviews International*, 24, 361–375.

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

Rapp, R. A., Udall, J. A., & Wendel, J. F. (2009). Genomic expression dominance in allopolyploids. *BMC Biology*, 7, 18.

Risso, D., Ngai, J., Speed, T. P., & Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, 32, 896–902.

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

Roth, C., & Liberles, D. A. (2006). A systematic search for positive selection in higher plants (Embryophytes). *BMC Plant Biology*, 6, 12.

Schlicker, A., Domingues, F. S., Rahnenführer, J., & Lengauer, T. (2006). A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics*, 7, 302.

Schoener, T. W. (1970). Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology*, 51, 408–418.

Senchina, D. S., Alvarez, I., Cronn, R. C., Liu, B., Rong, J., Noyes, R. D., ... Wendel, J. F. (2003). Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Molecular Biology and Evolution*, 20, 633–643.

Silvera, K., Santiago, L. S., Cushman, J. C., & Winter, K. (2009). Crassulacean acid metabolism and epiphytism linked to adaptive radiations in the Orchidaceae. *Plant Physiology*, 149, 1838–1847.

Sletvold, N., Grindeland, J. M., & Ågren, J. (2010). Pollinator-mediated selection on floral display, spur length and flowering phenology in the deceptive orchid *Dactylorhiza lapponica*. *New Phytologist*, 188, 385–392.

Slotte, T., Foxe, J. P., Hazzouri, K. M., & Wright, S. I. (2010). Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. *Molecular Biology and Evolution*, 27, 1813–1821.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, 6, e21800.

Thomma, B. P., Cammue, B. P., & Thevissen, K. (2002). Plant defensins. *Planta*, 216, 193–202.

Tiffin, P., & Moeller, D. A. (2006). Molecular evolution of plant immune system genes. *Trends in Genetics*, 22, 662–670.

Tomatsu, H., Takano, J., Takahashi, H., Watanabe-Takahashi, A., Shibagaki, N., & Fujiwara, T. (2007). An *Arabidopsis thaliana* high-affinity molybdate transporter required for efficient uptake of molybdate from soil. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 18807–18812.

Tremblay, R. L., Ackerman, J. D., Zimmerman, J. K., & Calvo, R. N. (2005). Variation in sexual reproduction in orchids and its evolutionary consequences: A spasmodic journey to diversification. *Biological Journal of the Linnean Society*, 84, 1–54.

Tupac Otero, J., & Flanagan, N. S. (2006). Orchid diversity—beyond deception. *Trends in Ecology & Evolution*, 21, 64–65.

Turcotte, M. M., Davies, T. J., Thomsen, C. J. M., & Johnson, M. T. J. (2014). Macroecological and macroevolutionary patterns of leaf herbivory across vascular plants. *Proceedings of the Royal Society of London B: Biological Sciences*, 281, 20140555.

Tutin, T., Heywood, V., Moore, D. M., Valentine, D. H., Walters, S. M., & Webb, D. A. (1980). *Flora Europaea, volume 5: Alismataceae to Orchidaceaea (Monocotyleones)*. Cambridge: Cambridge University Press.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C. Poplin, R., Levy-moonshine, A., Jordan, T., ... Depristo, M. A. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics 43* 11.10.1–11.10.33.

Wahyu Indra Duwi, F., Lee, S. Y., & Lee, K. O. (2013). The unfolded protein response in plants: A fundamental adaptive cellular response to internal and external stresses. *Journal of Proteomics*, 93, 356–368.

Wang, K. L.-C., Li, H., & Ecker, J. R. (2002). Ethylene biosynthesis and signaling networks. *The Plant Cell*, 14, S131–S151.

Warnefors, M., & Eyre-Walker, A. (2011). The accumulation of gene regulation through time. *Genome Biology and Evolution*, 3, 667–673.

Wong, W. S. W., Yang, Z., Goldman, N., & Nielsen, R. (2004). Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics*, 168, 1041–1051.

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, 8, 206–216.

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24, 1586–1591.

Yang, L., & Gaut, B. S. (2011). Factors that contribute to variation in evolutionary rate among *Arabidopsis* genes. *Molecular Biology and Evolution*, 28, 2359–2369.

Yang, Z., & Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, 17, 32–43.

Zhang, X. (2008). The epigenetic landscape of plants. *Science*, 320, 489–492.

Zhang, J., Li, X., He, Z., Zhao, X., Wang, Q., Zhou, B., ... Liu, X. (2013). Molecular character of a phosphatase 2C (PP2C) gene relation to stress tolerance in *Arabidopsis thaliana*. *Molecular Biology Reports*, 40, 2633–2644.

Zhang, Z., Li, J., & Yu, J. (2006). Computing Ka and Ks with a consideration of unequal transitional substitutions. *BMC Evolutionary Biology*, 6, 44.

Zhang, J., Nielsen, R., & Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, 22, 2472–2479.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28, 3326–3328.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.