BMC Plant Biology

RESEARCH Open Access

Evolutionary dynamics of Orchid *DL* paralogs: gene duplication, functional divergence, and expression patterns across Orchid subfamilies

Francesca Lucibelli[1*], Angela Carfora[1], Annette Becker[2], Katrin Ehlers[2] and Serena Aceto[1*]

Abstract

Background Orchids are known for their extraordinarily diversified floral structures and evolutionary adaptations. The study of transcription factor genes, such as the *YABBY* gene *DROOPING LEAF*, is crucial for understanding the molecular mechanisms underlying orchid development and evolution. This study aims to elucidate the evolutionary dynamics and expression patterns of *DL* genes across orchid subfamilies.

Results Through genomic and transcriptomic analyses, we identified 25 full-length *DL* genes in orchids, with two paralogs (*DL1* and *DL2*-like genes) observed in Epidendroideae, Orchidoideae, Cypripedioideae, and Vanilloideae, while the most ancestral Apostasioideae retained a single-copy gene. In addition to the functional *DL*, genomic features reveal the presence of a *DL* pseudogene within Apostasioideae. Evolutionary analyses revealed relaxed selection pressures acting on orchid *DL2* paralogs. Sequence comparison and expression analyses uncovered potential pseudogenization events affecting *DL2* paralogs of Vanilloideae and Cypripedioideae, while in the most recent subfamily Epidendroideae, differential expression of *DL2* in inner perianth tissues suggests the possible acquisition of a new function in the development of the lip callus.

Conclusions Our study provides insights into the evolutionary trajectory of *DL* genes in orchids. The relaxed selection on *DL2* paralogs might be related to pseudogenization or functional divergence. Pseudogenization of *DL2* in most ancestral orchids and possible neofunctionalization in Epidendroideae indicate a dynamic evolutionary process shaping the functional repertoire of *DL* genes. These findings contribute to our understanding of the genetic basis of orchid diversity and evolution, with implications for future studies on the role of transcription factors in plant development and adaptation.

Keywords Orchidaceae, *DROOPING LEAF*, Gene duplication, Evolutionary dynamics

*Correspondence:
Francesca Lucibelli
francesca.lucibelli@unina.it
Serena Aceto
serena.aceto@unina.it
[1]Department of Biology, University of Naples Federico II, Napoli, Italy
[2]Fachbereich 08 Biologie und Chemie, Institute of Botany, Justus-Liebig-University, Giessen, Germany

Background

Despite being non-model organisms, Orchidaceae play a crucial role in the study of biodiversity and evolution due to their remarkable diversity in flower architecture, specialized developmental programs, pollination syndromes, and colonization of diverse habitats. In particular, the extreme diversification of orchid flowers has long fascinated researchers, prompting investigations into the genetic mechanisms that regulate the orchid flower development and the establishment of floral symmetry [1–3]. Among the more than 30,000 species of Orchidaceae (WFO Plant List, www.worldfloraonline.org) there exists considerable phenotypic variation in their flowers. However, most orchid flowers share a typical organization of perianth structures, consisting of three outer tepals (one dorsal and two lateral), two lateral inner tepals, and a diversified median inner tepal known as the labellum or lip (Fig. 1A-D). The column, which comprises fused male and female reproductive tissues with the ovary at the base and pollen grains at the top, is another distinctive feature of orchid flowers [4, 5].

The lip, a crucial structure of the orchid perianth, contributes to the zygomorphy of the flower in the subfamilies Vanilloideae, Cypripedioideae, Orchidoideae, and Epidendroideae [5] and plays a pivotal role in pollination by attracting insects through its shape and pigmentation [6]. This specialized structure often exhibits various adaptations to facilitate successful reproduction, such as the presence of calli (Fig. 1F), steleidia, and mentum, which aid in insect adhesion, proper positioning of insects for pollination, and effective pollen transfer [7].

While the traditional genetic model ABC(DE) has provided a genetic framework for understanding floral organ identity [8, 9], monocots exhibit deviations from this model, particularly the expression domain of class B genes expands in the first whorl [10]. Consequently, orchids present a petaloid perianth with similar sepals and petals known as tepals. Specific models such as the "orchid code," HOT, and P-code have been proposed to explain the unique floral organization in orchids, emphasizing molecular pathways underlying the differentiation of outer and inner tepals, with particular focus on lip development [11–15].

Advancements in sequencing technology have resulted in the availability of an increasing number of orchid genomes and transcriptomes, allowing for a more comprehensive understanding of the molecular networks governing orchid flower development. For instance, MYB, TCP, and NAC transcription factors were proposed as relevant in establishing the orchid flower bilateral symmetry [16–22].

Moreover, investigations into YABBY transcription factors discovered duplication of *DL* genes in orchids, suggesting their involvement in perianth development through possible acquisition of new functions [23, 24]. In *Phalaenopsis* (Epidendroideae), the two *DL* paralogs do not exhibit the same expression profile: they show a conserved expression pattern in the column and ovary; however, *DL2* displays higher expression in the lip compared to the lateral inner tepals. This observation is intriguing, given that *DL* genes are typically not expressed in perianth organs of other angiosperms. For example, in the model plant *Arabidopsis thaliana*, the YABBY *CRC* gene, ortholog of *DL*, is involved in carpel development, floral meristem termination, and nectary formation [25]. In grasses, the *DL* genes maintain their role in the differentiation of reproductive structures and have also acquired a function in forming the leaf midrib [26].

Despite previous studies have hypothesized a possible neofunctionalization of *DL* genes in orchids, several questions remain unanswered regarding the evolutionary dynamics and expression patterns of the *DL* paralogs outside the *Phalaenopsis* genus and across the different orchid subfamilies. In this study, we aim to address these gaps by investigating the evolutionary trajectory of *DL* genes within the Orchidaceae family, identifying orthologs and paralogs across subfamilies, and exploring their expression patterns during early flower development both computationally and experimentally.

Methods

Plant material

The orchids *Dendrobium* Sonia (Fig. 1C) (an intrageneric hybrid between *Dendrobium* Caesar and *Dendrobium* Tomie Drake) and *Phalaenopsis aphrodite* (Fig. 1D) (Epidendroideae) used in this study were cultivated in the greenhouse of the Department of Biology (University of Naples Federico II, Napoli, Italy). Outer tepals, lateral inner tepals, lip, column, and ovary were collected from three floral buds of *P. aphrodite* (approximately 1 cm in size, Fig. 1F, left) and *Dendrobium* Sonia (approximately 1.5 cm, Fig. 1E). These sizes correspond to early developmental stages, as reported by previous works [24, 27]. The *P. aphrodite* lip at the 1 cm bud size was dissected into callus, lateral, and central lobes (Fig. 1F, right). The specific tissues were pooled and stored in RNAlater (Ambion, Austin, Texas) at −80 °C until RNA extraction. Total RNA was extracted from pooled tissues using Trizol (Ambion, Austin, Texas), followed by DNase treatment. Lip calli of *P. aphrodite* were fixed in 4% (v/v) paraformaldehyde, 0.5% (v/v) glutaraldehyde, 0.1% Triton X-100, and 4% dimethyl sulfoxide in phosphate–saline buffer 1X for 16 h at 4 °C. Subsequently, they were dehydrated through an ethanol series, paraffin-embedded, and stored at 4 °C until in situ hybridization experiments [28].

Total RNA of *Vanilla planifolia* (Vanilloideae, Fig. 1A), *Phragmipedium longifolium* (Cypripedioideae, Fig. 1B),

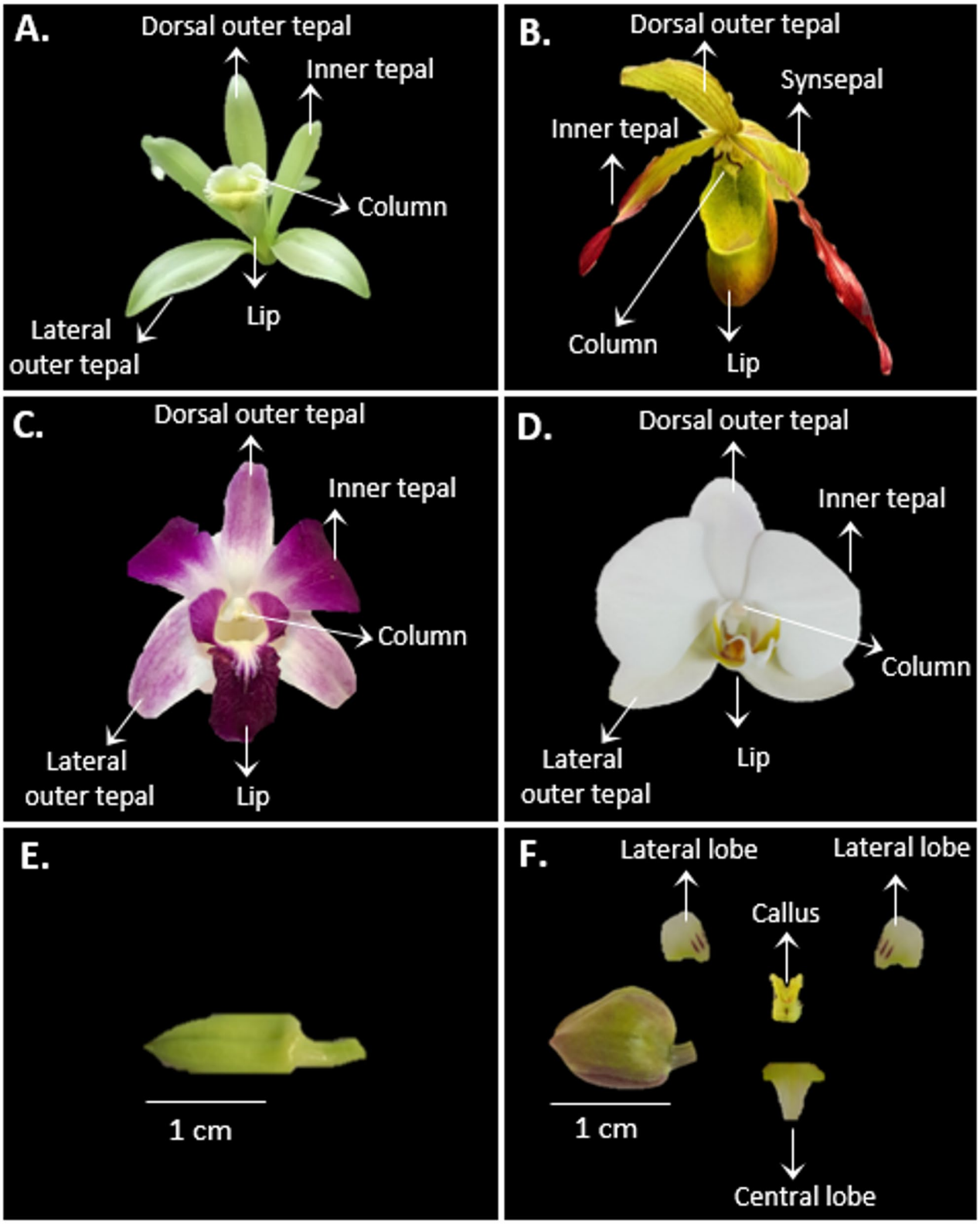

Fig. 1 Flowers of orchid species used in this work. **A** *Vanilla planifolia* (Vanilloideae); **B** *Phragmipedium longifolium* (Cypripedioideae), **C** *Dendrobium* Sonia (Epidendroideae); **D** *Phalaenopsis aphrodite* (Epidendroideae). In *Phragmipedium* lateral outer tepals are fused to form a unique concave synsepal. **E** floral bud of *Dendrobium* Sonia; **F** floral bud of *P. aphrodite* (on the left) and its lip dissected in callus, lateral lobes and central lobes (on the right)

and *Phalaenopsis* hyb. "Athens" was extracted from floral tissues dissected from floral buds of approximately 1.5 cm (*V. planifolia*), 1 cm (*P. longifolium*), and 1.5 cm (*Phalaenopsis* hyb. "Athens"). This material was available from previous studies [12, 23].

The *Nicotiana benthamiana* plants used for subcellular localization experiments were grown in the greenhouse of the Institute of Botany (Justus Liebig University, Giessen, Germany).

Orchid DL sequence retrieval

The *DL* nucleotide and amino acid sequences of *Apostasia shenzhenica* (AshDL) (Apostasoideae) and *Phalaenopsis equestris* (PeqDL1, PeqDL2) (Epidendroideae) were downloaded from the NCBI database. Sequences of *Apostasia wallichii* (AwaDL) and *Neuwiedia zollingeri* (NzoDL) (Apostasoideae), *Orchis italica* (OitDL1, OitDL2) (Orchidoideae), and *Cymbidium ensifolium* (CenDL1, CenDL2) (Cypripedioideae) were obtained from the Orchidstra2 database [29]. Sequences of *Haemaria discolor* (HdiDL1, HdiDL*2*) (Orchidoideae), and *Bletilla striata* (BstDL2) (Epidendroideae) were downloaded from the CNGB database.

The sequence of *Apostasia ramifera* (AraDL) (Apostasoideae) was reconstructed by performing a tBLASTn search on the reference genome ASM1929769v1 using AshDL as query. The same approach was used to obtain the DL sequences of *Dendrobium officinale* (DofDL1) and *Bletilla sinensis* (BsiDL1) (Epidendroideae) using PeqDL1 as query on the reference genomes ASM1951458v1 and ASM2249587v1, respectively.

The RNA-seq reads of *Vanilla planifolia* (PRJNA237672) (Vanilloideae), *Paphiopedilum malipoense* (PRJNA310678) (Cypripedioideae), and *Erycina pusilla* (PRJNA278061) (Epidendroideae) were in silico assembled using Trinity v2.12.0 software [30]. *DL* sequences were identified through tBLASTn searches against each assembled transcriptome using PeqDL1 and PeqDL2 as queries and the default e-value setting (10.0).

All orchid DL sequences retrieved are listed in the Additional file 1.

Evolutionary analyses

Amino acid sequences of orchid and other angiosperms DL proteins (Additional file 1) were aligned using T-Coffee (https://tcoffee.crg.eu/) [31], and the resulting alignm ent was manually adjusted. The corresponding orchid *DL* nucleotide sequence alignment was obtained from the multiple protein alignment using Pal2Nal [32].

The best amino acid substitution model was determined using MEGA X software [33]. The amino acid Neighbor-Joining tree was constructed by applying the JTT + G + I model (shape parameter = 1.16) with 500 bootstrap replicates using MEGA X software [33]. The analysis involved 62 sequences, and all ambiguous positions were removed (pairwise deletion option), resulting in a final dataset of 364 positions.

Pairwise orchid nucleotide and amino acid identity were calculated using the Ident and Sim online tool (https://www.bioinformatics.org/sms2/ident_sim.html).

The amino acid conserved motifs of the orchid DL sequences were searched using the MEME v5.5 online tool by setting to 15 the maximum number of identified motifs with motif width between 6 and 50, and allowing the zero or one occurrence per sequence motif distribution [34]. The identified motifs were scanned for matches against the InterPro protein signature databases, using InterProScan tool (https://www.ebi.ac.uk/interpro/search/sequence/).

GARD analysis [35] implemented in the Datamonkey Adaptive Evolution Server (https://www.datamonkey.org/) [36] was conducted to detect recombination breakpoints within the aligned nucleotide *DL* sequences.

Evolutionary analysis was performed using the CODEML application of PAML v4.9 software [37]. Site models, branch models, and branch-site models, coupled with LRT calculations, were used to analyze selective pressures, evolutionary rates, and signals of positive selection across the orchid *DL* gene phylogeny. The analyses were conducted on the aligned coding sequences and the two partitions detected by GARD analysis.

Site models M0 (one-ratio), M1a/M2a (nearly neutral/positive selection), M7 (beta), and M8 (beta,&ω > 1) were employed to assess variation in selection pressure among sites within the orchid *DL* nucleotide sequences. LRTs were performed to compare the fit of different models with the null hypothesis, assuming uniform selective pressure across all sites.

To assess variation in evolutionary rates among different branches of the *DL* phylogenetic tree, branch model 2 was implemented. The LRT was utilized to compare the fit of the null model (uniform rates across all branches) against the alternative model, allowing for distinct rates on specified branches (DL2; Epidendroideae and Orchidoideae DL2; Epidendroideae DL2).

Branch-site models were employed to detect positive selection acting on specific lineages. This involved designating foreground and background branches, with the null hypothesis assuming neutrality on the foreground branches. LRTs were conducted to assess the significance of positive selection, providing insights into adaptive evolution along specific lineages (DL2; Epidendroideae and Orchidoideae DL2; Epidendroideae DL2).

The LRT statistic, calculated as twice the difference in log-likelihood between nested models, follows a chi-squared distribution. Significance levels were determined based on the degrees of freedom equal to the difference

in the number of parameters between the two models compared.

In addition, the RELAX tool [38] in Datamonkey was used to identify relaxed selection in the orchid *DL2* coding sequences, setting as test branches the same used as foreground in the CODEML analysis. The analysis was conducted also setting *DL1* coding sequences as test branch and *DL2* as reference.

In silico expression analysis

In silico expression analysis of *DL* genes was conducted in *Apostasia odorata*, *Dendrobium* Suriya Gold, *Rhyncholaeliocattleya* Beauty Girl "KOVA", and *Paphiopedilum micranthum*. *Dendrobium* Suriya Gold is an intrageneric hybrid between *Dendrobium* Siam's Gift and *Dendrobium* Chaisri Gold; *Rhyncholaeliocattleya* Beauty Girl "KOVA" is an intergeneric hybrid between *Rhyncholaelia* Schltr. and *Cattleya* Lindl.

Illumina raw reads were downloaded from the SRA database and reference transcriptomes were assembled using Trinity v2.12.0 software [30]. BUSCO analysis was conducted to assess the completeness of the assembled transcriptomes [39].

The *A. odorata* floral bud transcriptome was assembled using reads from the PRJNA310678 Bioproject, derived from a single biological sample of floral bud tissue (SRR5722152).

The *Dendrobium Suriya Gold* perianth organ transcriptome was assembled using all reads from the PRJNA880415 Bioproject, which were obtained from outer tepals, lateral inner tepals, and lips of floral buds. For each tissue type, reads from three distinct biological samples were analyzed.

The *Rhyncholaeliocattleya* Beauty Girl "KOVA" transcriptome was assembled using all reads from the PRJNA559603 Bioproject. Differential expression analysis was focused on outer tepals, lateral inner tepals, and lips, with the latter dissected into epichile (distal portion) and hypochile (basal portion) at four developmental stages (S1, S2, S3, and S4 reported by previous work [40]). For each tissue and stage, reads from three independent biological samples were analyzed.

The *P. micranthum* inner perianth transcriptome was assembled using all reads from the PRJNA918555 Bioproject, obtained from early and late developmental stages of inner tepals and lips. For each tissue type and stage, reads from three different biological samples were analyzed.

In the Additional File 2 are reported the main statistics of the assembled transcriptomes and used genomes available on NCBI.

Subsequently, the differential gene expression analysis was performed using edgeR v3.13 software, which includes normalization based on the trimmed mean of M-values (TMM) method [41]. This analysis excluded *Apostasia odorata*, where only one biological sample was available and no differential expression analysis was performed.

Quantitative expression analysis

Reverse-transcription reactions were performed on 500 ng of pooled RNAs (see Plant material) using the LunaScript® RT SuperMix kit (NEB, Ipswich, Massachusetts). Relative expression of *DL* genes was evaluated in *V. planifolia*, *P. longifolium*, *Dendrobium* Sonia, and *P. aphrodite* by real-time qPCR using the PowerUp™ SYBR™ Green Master Mix (Applied Biosystems™, Foster City, California), with ribosomal 18 S as the reporter gene [21]. Primer pairs used in qPCR experiments are listed in Additional file 3. qPCR experiments were conducted in technical triplicates. In order to carry out the relative quantification of the transcript levels of the genes of interest compared to the housekeeping gene 18 S we applied the Normalized Relative Quantity method (NRQ) [42]. For each gene, NRQ ± SEM was calculated, and ANOVA analysis followed by the Holm–Sidak *post-hoc* test was carried out to estimate the statistical significance of the differences in NRQ among the tissues, with the significance threshold set at $p < 0.05$.

In situ hybridization

Phalaenopsis aphrodite paraffin-embedded calli (see Plant material) were sectioned at 5 µm using a Leica RM2235 microtome. Probe synthesis involved PCR amplification of a *DL2* fragment of *P. aphrodite* (DreamTaq, Thermo Scientific™, Waltham, Massachusetts) using a specific primer pair (Additional file 3). The PCR amplification product was cloned into pGEM®-T Easy Vector (Promega, Madison, Wisconsin). T7 and SP6 RNA polymerases and the DIG RNA Labeling kit (Roche, Basel, Switzerland) were used to transcribe DIG-labeled sense and antisense RNA probes. The DIG Nucleic Acid Detection kit (Roche, Basel, Switzerland) was used to detect the hybridization signal following the manufacturer's instructions. The signal was observed using a Zeiss Axioskop microscope, and the images were acquired using Axiovision 4.7 software (Zeiss, Oberkochen, Germany).

PeqDL2 protein subcellular localization

The *PeqDL2* coding sequence was divided into regions encoding the specific protein domains Zinc Finger (ZF, aa 14–42), Intermediate (INT, aa 43–98), and YABBY (YAB, aa 99–148) based on the alignment with the *Arabidopis* CRC protein [43] (Additional file 4, Figure A1). The full-length coding sequence and the regions encoding the specific domains were PCR amplified (DreamTaq, Thermo Scientific™, Waltham, Massachusetts) using primers listed in the Additional file 2 and cloned into the

N-terminal GFP fusion vector pEGAD under the control of the CaMV 35 S promoter [43, 44].

The obtained constructs were transformed into *Agrobacterium tumefaciens* GV3101 used to infiltrate leaves of 4-week-old *Nicotiana benthamiana* [43]. Three days after infiltration, a disk of the infiltrated leaf was isolated and stained with 1 ng/ml DAPI. Microscopy analysis was conducted with the Leica fluorescence microscope DCM5500 (Leica Microsystems GmbH, Wetzlar, Germany) using an A4 filter for DAPI fluorescence and an L5 filter for GFP fluorescence.

Results

Structure and evolution of the orchid *DL* genes

To explore the evolutionary trajectory of *DL* genes within the Orchidaceae family, a search for *DL*-like orthologs and paralogs across various subfamilies was conducted using various approaches, including sequence analysis and processing RNA-seq data from publicly available databases. A total of 25 full-length open reading frames encoding DL proteins were identified (Additional file 1). Among the subfamilies Epidendroideae, Orchidoideae, Cypripedioideae and Vanilloideae, two *DL* paralogs were observed, while within the basal subfamily Apostasioideae, *DL* exists as a single-copy gene. Nevertheless, the tBLASTn analysis conducted on the genomes of *A. shenzhenica* and *A. ramifera* [45, 46] using the AshDL and AraDL sequences as queries to identify additional *DL* genes, uncovered two scaffolds (KZ451899.1 for *A. shenzhenica* and JAEMOT010001828.1 for *A. ramifera*) containing coding sequences of *DL* with premature stop codons (Additional file 5). In vivo and/or in silico analyses demonstrate the expression of the single-copy gene of Apostasioideae and both orchid *DL* paralogs of the other orchid subfamilies.

Figure 2 shows heatmaps illustrating the percentage similarity and identity among the orchid DLs. The amino acid similarity is high among the DL proteins of Apostasioideae (91–99%) and within DL1, ranging from ~85% to ~97%. Excluding *Vanilla* and *Phragmipedium*, amino acid similarity spans from ~85% to ~96% for DL2. The DL proteins of Apostasioideae exhibit greater similarity to orchid DL1 (from ~75% to ~80%) than DL2 (from ~57% to ~67%). The similarity between DL1 and DL2 sequences spans from ~59% to ~74%. Notably, VplDL2 of *V. planifolia* displays greater similarity to DL1 (and DL proteins of Apostasioideae) than to DL2 proteins. In addition, the PmaDL2 and PmiDL2 of *Paphiopedilum* show lower similarity than other DL2s when compared to all the other DL proteins. Additional file 6 reports the pairwise amino acid similarity and identity and pairwise nucleotide identity of the orchid DL sequences.

The Neighbor-Joining tree of the *DL* amino acid sequences clusters the orchid *DL1* and *DL2* paralogs in two groups and includes the Apostasioideae *DL*s within the DL1 group (Fig. 3). The topology of both clusters is in general agreement with the orchid phylogeny, with Epidendroideae and Orchidoideae representing the most recent subfamilies. The position of the *VplDL2* sequence is outside the group that includes the *DL*s of Apostasioideae and the *DL1* sequences.

The MEME tool was used to search for conserved and novel domains among orchid DL-like proteins that may be associated with the acquisition of new functions. The MEME analysis identified conserved and distinctive amino acid motifs of orchid DL proteins (Fig. 4; Additional file 4, Figure A2) [34]. Motifs 1 and 2 match with YABBY proteins of the InterPro database (IPR006780) and are shared by all sequences, as well as the unclassified motifs 3 and 4. Motifs 5 and 6 are classified as

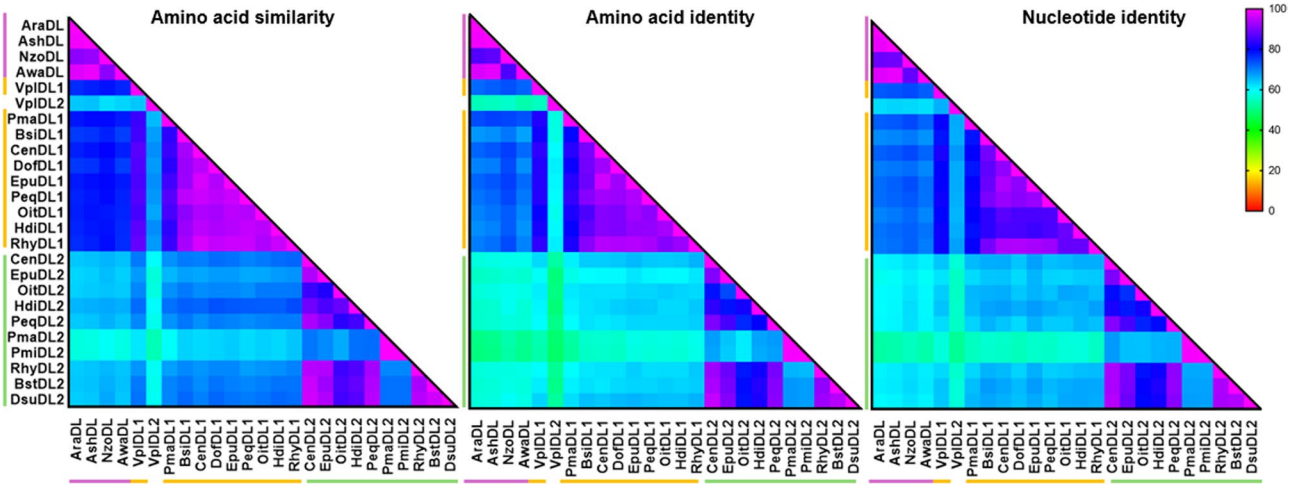

Fig. 2 Graphical representation of pairwise amino acid and nucleotide comparisons among orchid DLs. The heatmaps display similarity and identity percentages at both amino acid and nucleotide levels, using a color map ranging from red (0%) to pink (100%). Colored bars indicate the Apostasioideae DL (pink), DL1 (yellow), and DL2 (green) sequences

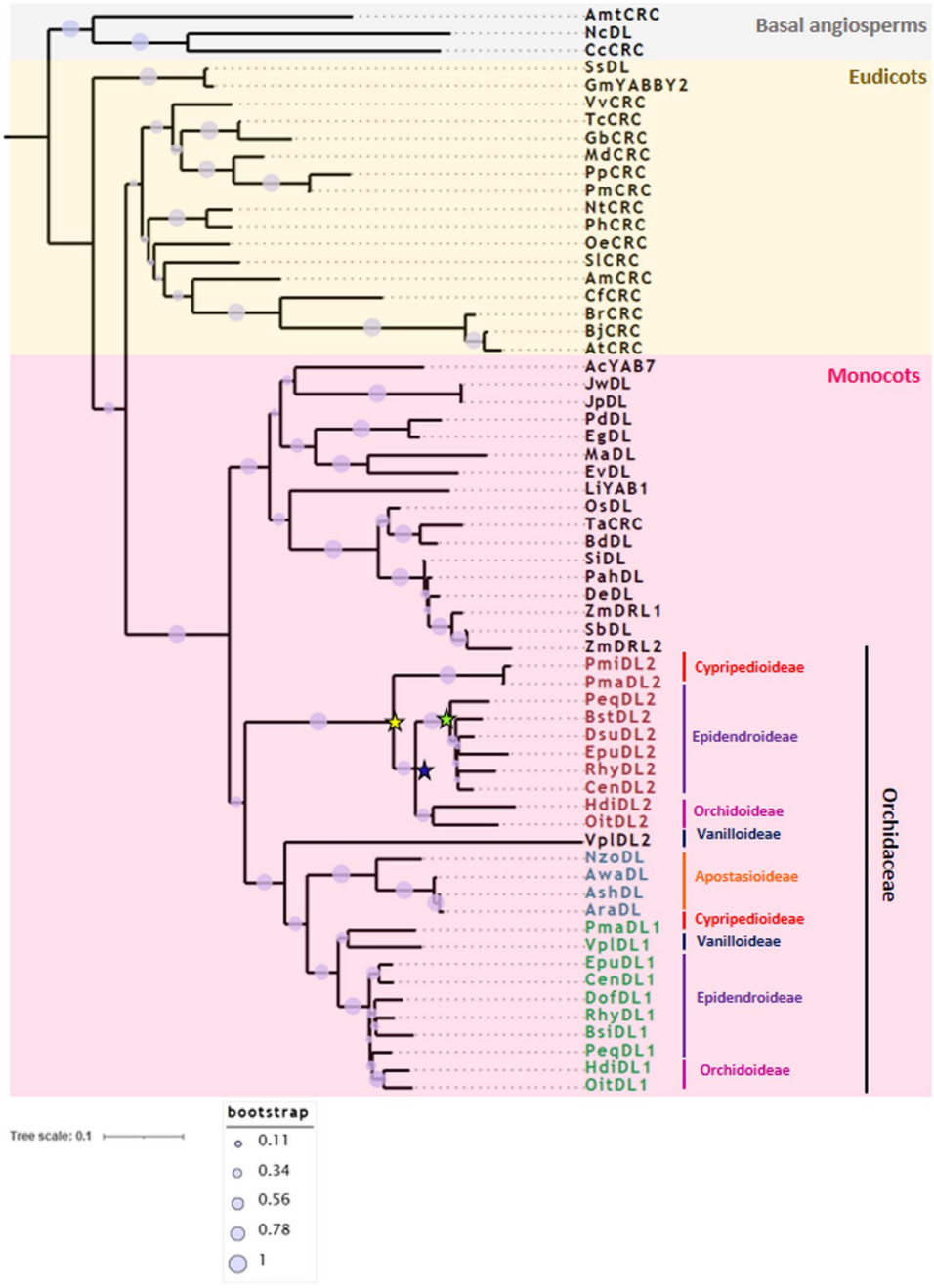

Fig. 3 Neighbor-Joining tree of the DL proteins. Neighbor-Joining tree of the amino acid DL sequences of orchids and other angiosperms. Colored stars indicate the foreground branches used in the evolutionary analyses

intrinsically disordered regions (MobiDBLite database), both missing in VplDL2, and motif 6 is missing in PmaDL2 and PmiDL2. All the other motifs do not correspond to any protein deposited in the InterPro database; however, some are characteristic of specific groups. Specifically, motif 10 is specific to the Apostasioideae DLs, motif 8 is shared by DL1s and DLs of Apostasioideae, and motif 11 is specific to DL2 proteins. Motif 13 is found only in Cypripedioideae DL2.

The analysis of recombination sites inferred a single breakpoint at nucleotide position 455 (with support 0.83). Therefore, evolutionary analyses were conducted on the entire codon alignment and its two partitions: positions 1-456 and 457–660. The position of the predicted recombination breakpoint is indicated on the amino acid sequences in the Additional file 4, Figure A3.

The results of PAML analysis running CODEML under various site models gave significant results when comparing the M0 model (one-ratio) and the nested model

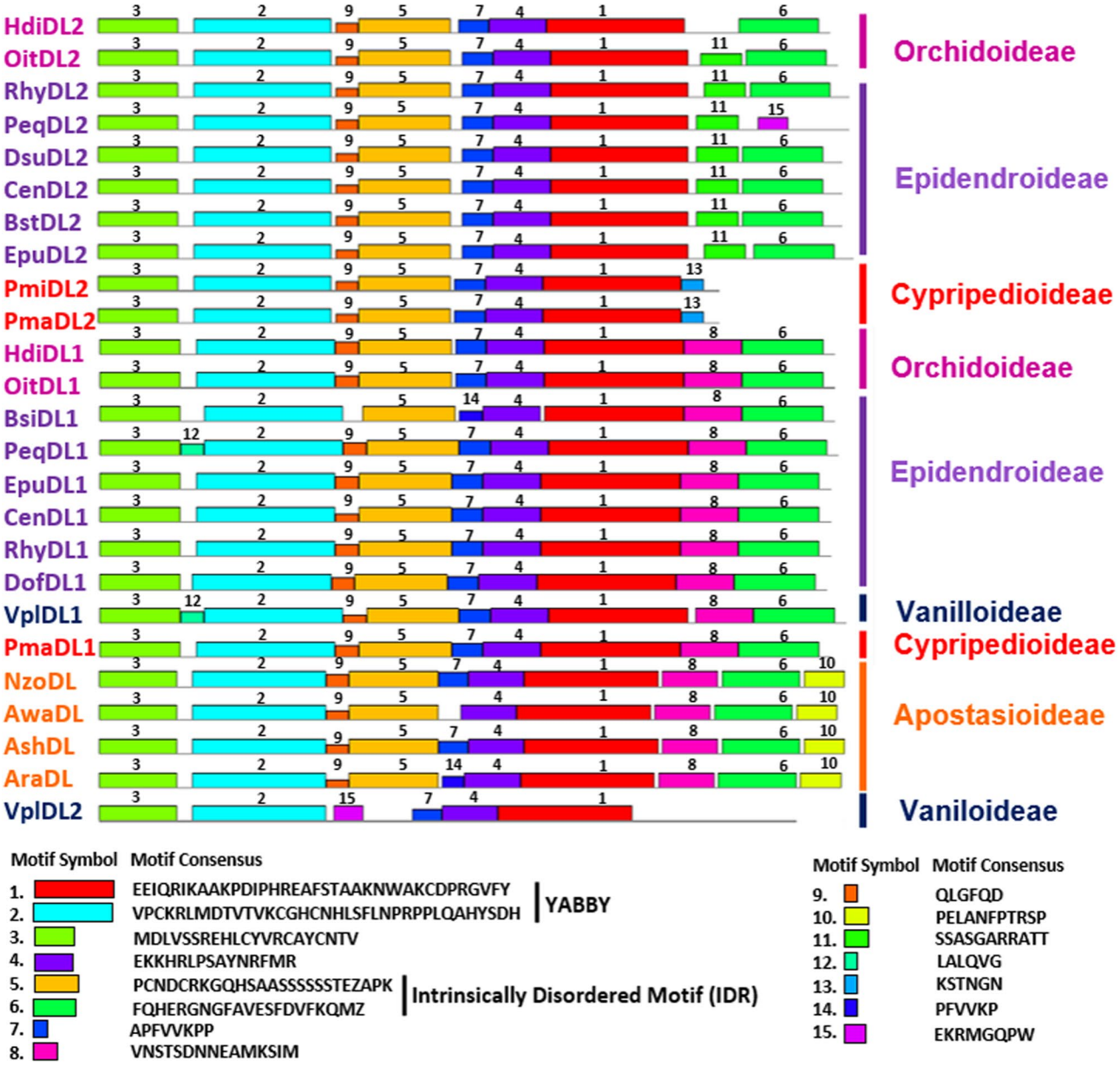

Fig. 4 Conserved motifs shared by DL orchid proteins. Amino acid motifs on DL orchid protein sequences identified by MEME analysis. Motifs 1 and 2, characteristics of YABBY transcription factors, are present in all sequences. Motifs 5 and 6 represent intrinsically disordered regions (IDR) absent in VplDL2 and both Cypripedioideae DL2, respectively

M1a (nearly neutral), supporting variability of selective pressure among amino acid sites when considering the entire codon alignment as well as its partitions (Additional file 7). Branch models assume that ω varies across the branches of the tree and are compared to the null site model M0, which assumes the same ω for all branches. Three different tests were conducted, setting as foreground branches the DL2 clade, the DL2 of Epidendroideae and Orchidoideae, and the DL2 of Epidendroideae (see Fig. 4). Even though the LRT test was significant considering the DL2 clade as the foreground branch in the partition 1-456 (Additional file 7), the estimate of ω values does not clearly suggest positive selection. Branch-site models assume heterogeneous pressure across sites and branches of a gene phylogeny. These models were used to detect positive selection affecting specific amino acid sites along the three foreground branches specified for the branch models. Different positively selected sites were detected both considering the entire coding sequence and its partitions, in particular when the DL2 of Epidendroideae were set as the foreground branch (Additional file 7); however, LRT tests against the respective null models were not significant, even though with

chi-square very close to the critical value for 1 degree of freedom.

Tests for selection relaxation on *DL2* were significant considering both the whole coding region and its partitions (Table 1). The selection intensity parameter K indicates the stringency of selection acting on the test branch. When significant results show $K < 1$, selection strength has been relaxed along the test branches, as observed for the orchid *DL2* genes. In contrast, the RELAX analysis considering *DL1* as test branch gave significant results for selection intensification, with $K > 1$ (Table 1).

The *DL* orchid genes exhibit a genomic organization with seven exons and six introns. This gene structure is conserved in comparison to *A. thaliana* (*AthCRC)* and grasses such as *Triticum aestivum* (*TaeCRC*) and *Oryza sativa* (*OsaDL*) [47–49]. The exon size is quite uniform, while intron size is variable and increases probably due to the presence of many transposable elements characteristic of the orchid genomes (Additional file 4, Figure A4).

The comparison of orchid DL nucleotide and amino acid alignments highlighted the presence of a single-nucleotide indel shared by all orchid *DL1* and *DL2* coding sequences compared to the *DL*s of Apostasioideae (Fig. 5). This frameshift mutation alters the sequence composition and length of the encoded C-terminal region. Additionally, a single-nucleotide indel is shared by DL2 sequences of Orchidoideae and Epidendroideae, occurring just before the last codon. Vanilloideae and Cypripedioideae *DL2s* appear to deviate from these mutation-induced sequence changes, resulting in an evident sequence divergence. In particular, *VplDL2* has accumulated numerous substitutions as well as some in-frame and single indels (Fig. 5). *PmaDL2* and *PmiDL2* share two large in-frame deletions and a premature stop codon spanning exon 6 and 7, resulting in a shorter (162 aa) encoded amino acid sequence (Fig. 5, Additional file 4, Figure A5).

Table 1 Results of the RELAX analysis on the entire CDS of the Orchid *DL* genes and on its partitions into 1-456 and 457–660

Partition	Test branch	K	*p*-value
1-660	DL2	0.18	**0.002**
	DL2 Epidendroideae + Orchidoideae	0.75	**0.006**
	DL2 Epidendroideae	0.64	**0.001**
	DL1	1.31	**0.004**
1-456	DL2	0.75	**0.015**
	DL2 Epidendroideae + Orchidoideae	0.83	0.077
	DL2 Epidendroideae	0.73	0.050
	DL1	1.27	**0.027**
457–660	DL2	0.28	**0.043**
	DL2 Epidendroideae + Orchidoideae	0.09	0.677
	DL2 Epidendroideae	0.20	**0.015**
	DL1	0.00	0.959

Cellular localization of the *Phalaenopsis* DL2 protein and its domains

To support the hypothesis that PeqDL2 functions as a transcription factor, its subcellular localization was verified by transiently expressing recombinant fluorescent proteins in *N. benthamiana* leaf epidermal cells. The analysis of subcellular localization of the full-length *Phalaenopsis* PeqDL2 protein and its domains by fluorescence microscopy demonstrates that PeqDL2 is exclusively confined in the nucleus (Fig. 6B). This result is consistent with the transcription factor role of this protein. The fused proteins constituted by only the zinc-finger or intermediate domain show a nuclear and cytoplasmatic localization (Fig. 6C, D), as well as the native GFP adopted as control (Fig. 6A). In contrast, the constructs composed by only the YABBY domain localize in the nucleus, implying that this domain contains the nuclear localization signal (Fig. 6E). These results are consistent with previous work on the CRC protein of *Arabidopsis* [43].

Expression analyses

Expression analyses were conducted to examine the expression pattern of orchid *DL* genes beyond the *Phalaenopsis* genus. These analyses were performed on representatives of the Epidendroideae subfamily, to which *Phalaenopsis* belongs, as well as on species belonging to other orchid subfamilies, namely Vanilloideae, Cypripedioideae, and Apostasioideae, using both in vivo and in silico approaches. Limited data on Apostasioideae and Orchidoideae result from challenges in collecting/dissecting plant material at early flower developmental stages and the absence of tissue-specific RNA-seq data in public databases for these orchid subfamilies.

Quantitative real-time PCR was conducted to detect the expression pattern of *DL1* and *DL2* genes in the floral organs of orchids collected at the early stage of development. In *V. planifolia* (Vanilloideae), *VplDL1* shows higher expression in the column and ovary than *VplDL2* (Fig. 7A-B) and is expressed at very low levels in the perianth (Fig. 7A). The levels of *VplDL2* are uniform in the reproductive tissues as well as in the different tissues of the perianth (Fig. 7B).

In *P. longifolium* (Cypripedioideae), *PloDL1* is more expressed than *PloDL2* in the reproductive tissues and poorly expressed in the perianth (Fig. 7C-D). As in *Vanilla*, *PloDL2* is detected in the perianth and reproductive tissues of *Phragmipedium* with similar expression levels (Fig. 7D).

In the Epidendroideae *P. aphrodite* and *Dendrobium* Sonia, the *DL* genes are expressed in the column and ovary, and *DL1* is more expressed than *DL2* (Fig. 7E-H). The *DL2* genes are also expressed in perianth, with higher levels in the lip than other organs (Fig. 7F, H).

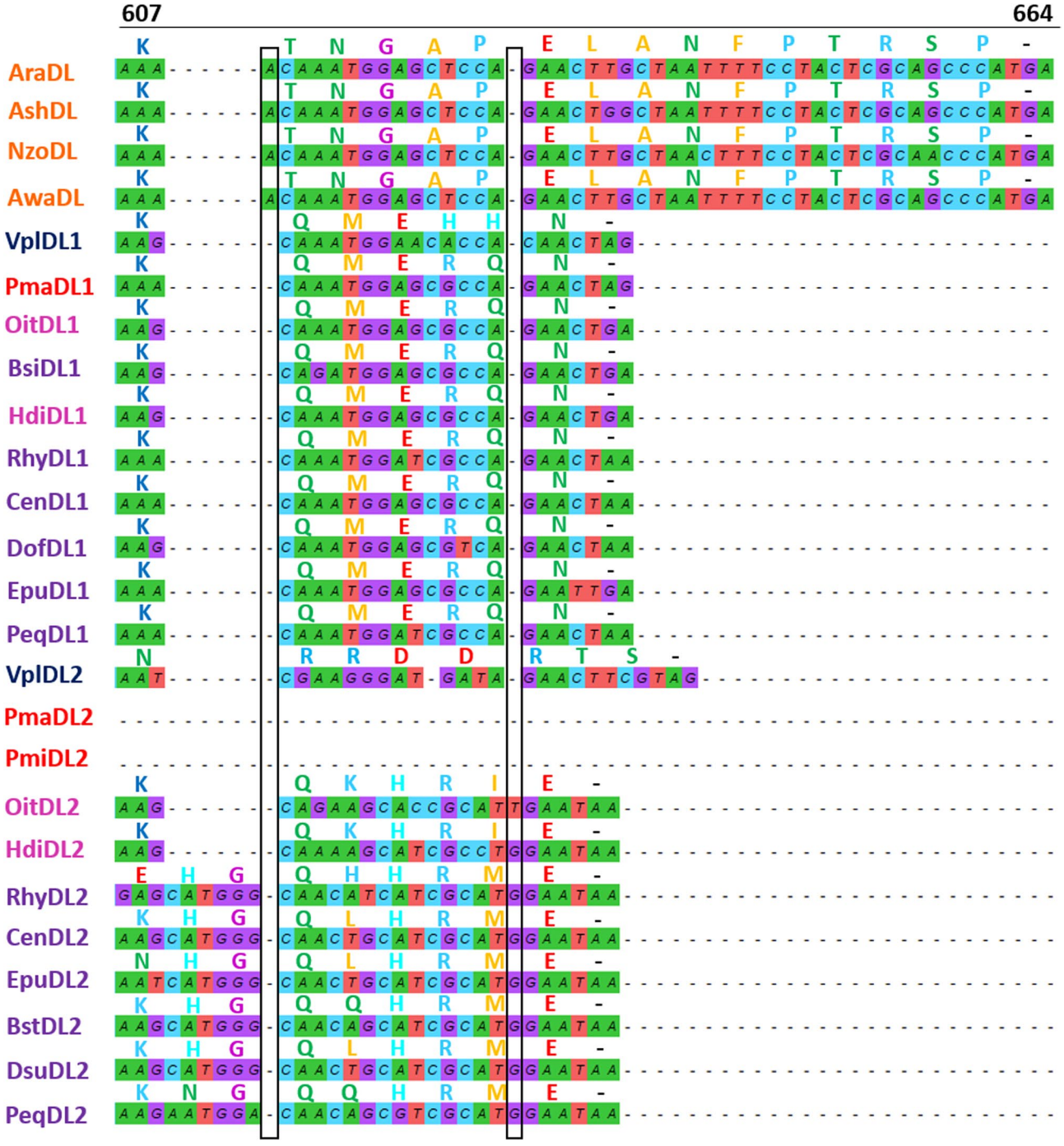

Fig. 5 Frameshift mutations in orchid *DL* genes. Virtual translation of the *DL* region spanning from nucleotide position 607 to 664. The vertical box highlights the indel position. The color of each species abbreviation indicates its subfamily: orange for Apostasioideae, blue for Vanilloideae, red for Cypripedioideae, pink for Orchidoideae, and purple for Epidendroideae

These results match with the differential expression analysis conducted in previous work [23]. In addition, the expression profile of *PapDL2* in the lip of *P. aphrodite* shows higher levels in the callus, located in the basal part of the lip (hypochile), than in the lateral and central lobes (Fig. 8A). To determine whether *PapDL2* expression in the callus is localized to specific regions or distributed uniformly, we performed in situ RNA hybridization. The results indicate that *PapDL2* is expressed homogeneously throughout the callus (Fig. 8B).

To verify the expression domains of the *DL* genes in the different organs of the perianth in other orchid species we conducted in silico differential expression analysis. Firstly, the perianth organs of *Dendrobium* Suriya

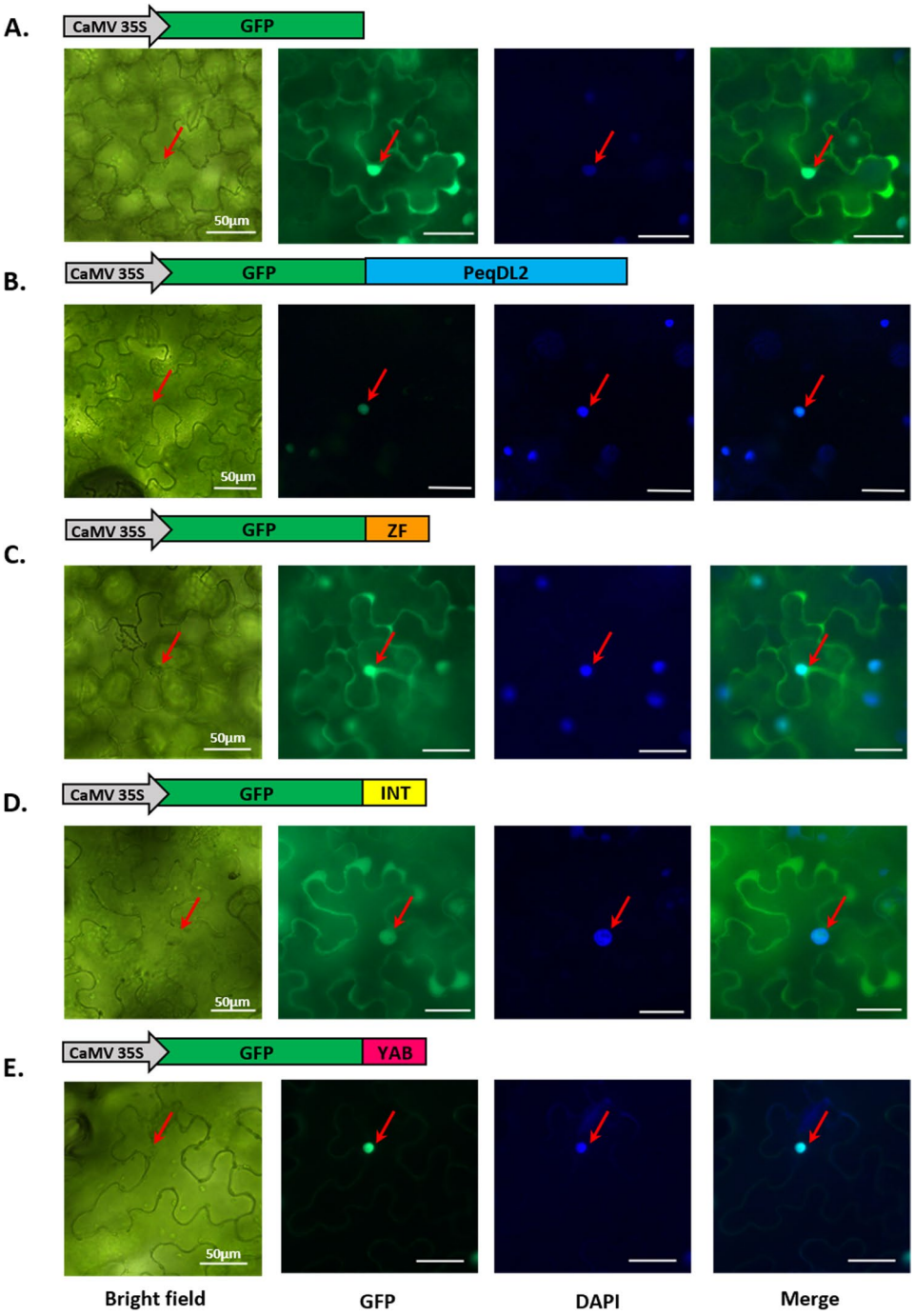

Fig. 6 Cellular localization of PeqDL2 and its domains. Fluorescence microscopy images of *Nicotiana benthamiana* leaf sections agroinfiltrated with recombinant constructs containing GFP native protein (**A**); GFP-PeqDL2 (**B**); GFP-ZF (**C**); GFP-INT (**D**); GFP-YAB (**E**). ZF, zinc-finger domain; INT, intermediate domain; YAB, YABBY domain. The red arrows show the main nucleus in focus

Gold (Epidendroideae) transcriptome was assembled from the public RNA-seq reads. In the assembled transcriptome, tBLASTn analysis revealed the presence of a *DL2* transcript (*DsuDL2*), while *DL1* was absent. This suggests very low or absent expression of *DL1* in the perianth organs of *Dendrobium*. Differential expression analysis indicated higher expression of *DsuDL2* in the lip compared to the lateral inner tepals and outer tepals (Additional file 4, Table A1), consistent with the expression pattern observed in other analysed Epidendroideae [23].

In silico differential expression analysis was also performed by assembling RNA-seq reads of *Rhyncholaeliocattleya* Beauty Girl "KOVA" (Epidendroideae) [40]. As

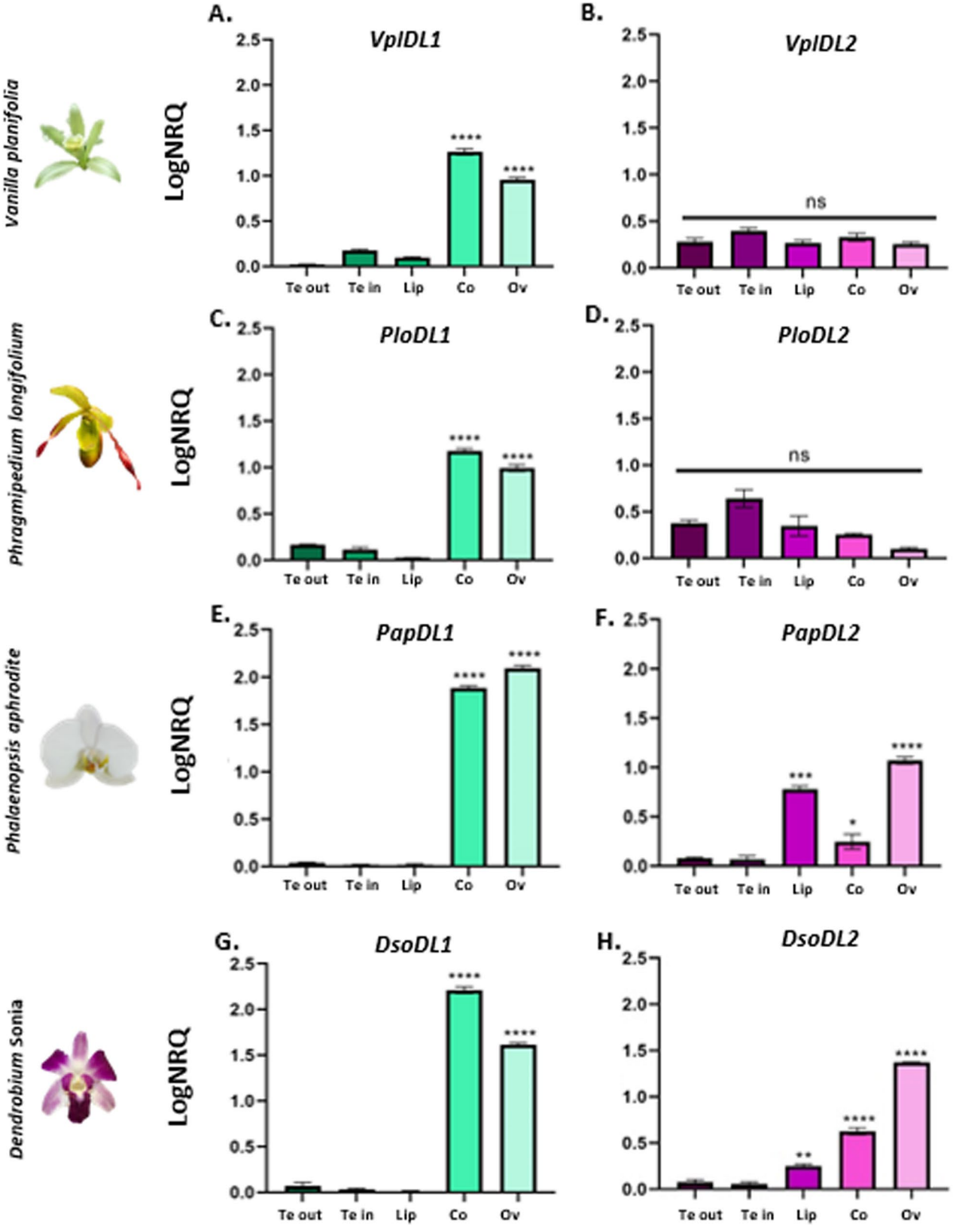

Fig. 7 (See legend on next page.)

(See figure on previous page.)

Fig. 7 Relative expression of the *DL* genes in early flower tissues of orchids belonging to different subfamilies. Relative expression of *DL1* and *DL2* genes in different floral tissues of *Vanilla planifolia* (Vanilloideae) collected from 1.5 cm floral buds (**A**-**B**); *Phragmipedium longifolium* (Cypripedioideae) collected from 1 cm floral buds (**C**-**D**); *Phalaenopsis aphrodite* (Epidendroideae) collected from 1 cm floral buds (**E**-**F**); *Dendrobium* Sonia (Epidendroideae) collected from 1.5 cm floral buds (**G**-**H**). The expression is reported as logarithm of 1 + Normalized Relative Quantity (LogNRQ). The bars represent the SEM of the biological and technical replicates. The asterisks indicate the statistically significant difference of the expression compared to outer tepals, with $p < 0.05$ considered the threshold for significance. *p*-values *<0.05; **<0.01; ***<0.001; ****<0.0001; ns, not significant. Te out, outer tepals; Te inn, inner tepals; Co, column; Ov, ovary

expected, *RhyDL1* has low expression in the perianth organs, whereas *RhyDL2* is expressed only in the lip at the S1 early stage of development and is not expressed during later developmental stages (S2, S3 and S4), as in *Phalaenopsis* [23]. In particular, *RhyDL2* is expressed only in the hypochile (basal part) of the early lip (S1 stage) (Additional file 4, Figure A6, Table A1). These results are congruent with the higher expression of *PapDL2* transcripts of *P. aphrodite* in lip callus (Fig. 8B).

Transcriptome assembly and differential expression analysis were also conducted on *Paphiopedilum micranthum* (Cypripedioideae), using RNA-seq reads from lateral inner tepals and lip. As for *Dendrobium* and *Rhyncholaeliocattleya*, *PmiDL2* is expressed in the *Paphiopedilum* inner perianth organs, while the *PmiDL1* transcript is not detectable. The differential expression of *PmiDL2* between the early lateral inner tepals and lip is not statistically significant (FDR > 0.1; Additional file 4, Table A1).

The transcriptome of *Apostasia odorata* was assembled using RNA-seq reads from flower bud to verify the expression of the single-copy *DL* gene. One full-length *DL* transcript is present in this transcriptome, indicating that the only transcribed *DL* in *Apostasia* is the single-copy gene.

Discussion

The exploration of orchid *DL* genes reveals intriguing insights into their structure, evolution, expression patterns, and functional characteristics. Across various orchid subfamilies, distinctive patterns emerge regarding *DL* gene duplication, expression profiles, sequence conservation and evolutionary dynamics, providing clues to possible functional specialization. These findings deepen our understanding of orchid *DL* genes evolution, structure, and expression dynamics, opening the way for further exploration of their biological significance in orchid development and adaptation.

Orchid *DL* genes duplication: pseudogenization or paralog retention

Gene duplication is a frequent event in genomes across diverse taxa, contributing to the expansion of gene families [50–52]. Paralogs, resulting from gene duplications, often undergo divergent evolutionary trajectories, with one copy retaining the original function. At the same time, the other can experience different fates, including pseudogenization, sub-functionalization, or neofunctionalization with the acquisition of new functions [52].

DL proteins are transcription factors belonging to the YABBY family, generally present as a single-copy gene in most monocots, as well as in basal and core eudicots, including *A. thaliana*. However, in certain angiosperm families, such as Solanaceae and Brassicaceae, multiple copies of *DL/CRC* genes have been identified, typically resulting from whole-genome duplication, whole-genome triplication, or polyploidization events [53, 54]. In the Poaceae family, the number of *DL/CRC* genes varies among species. In *Oryza sativa*, only a single copy of the *DL* gene is present [26], whereas in *Zea mays*, the *drl1* and *drl2* loci represent a case of recent gene duplication [55].

Genome- and transcriptome-wide analyses of *DL* genes across various orchid species reveal that the most ancestral orchids Apostasioideae have a single-copy *DL* gene, while Vanilloideae, Cypripedioideae, Orchidoideae, and Epidendroideae have duplicated *DL* genes, consistent with previous findings [23, 24]. Earlier studies proposed the acquisition of a new function of the duplicated *DL2* paralog in *Phalaenopsis* orchid flower development [23], and our work aimed to explore the evolution of the duplicated *DL*s across the orchid subfamilies.

Gene duplication is a major driver of functional innovation in regulatory networks, providing genetic material that can diverge over time. Following duplication, one or both gene copies may accumulate changes that alter the properties of their encoded proteins, such as DNA-binding activity, protein-protein interactions, or expression patterns. These functional divergences can lead to key evolutionary adaptations in plants, influencing flower development and contributing to variation in blooming time, floral scent, coloration, and morphology [1, 48, 50, 51].

The topology of the Neighbor-Joining tree indicates that the duplication event responsible for the emergence of two *DL* paralogs occurred in the common ancestor of orchids rather than after the divergence between Apostasioideae and Vanilloideae. This gene duplication could be related to the whole-genome duplication event occurred in the ancestor of all orchid species before the diversification in different subfamilies [45]. During evolution, pseudogenization affected the duplicated *DL* gene of Apostasioideae, resulting in the retention of a single functional *DL* gene. This hypothesis is supported by the

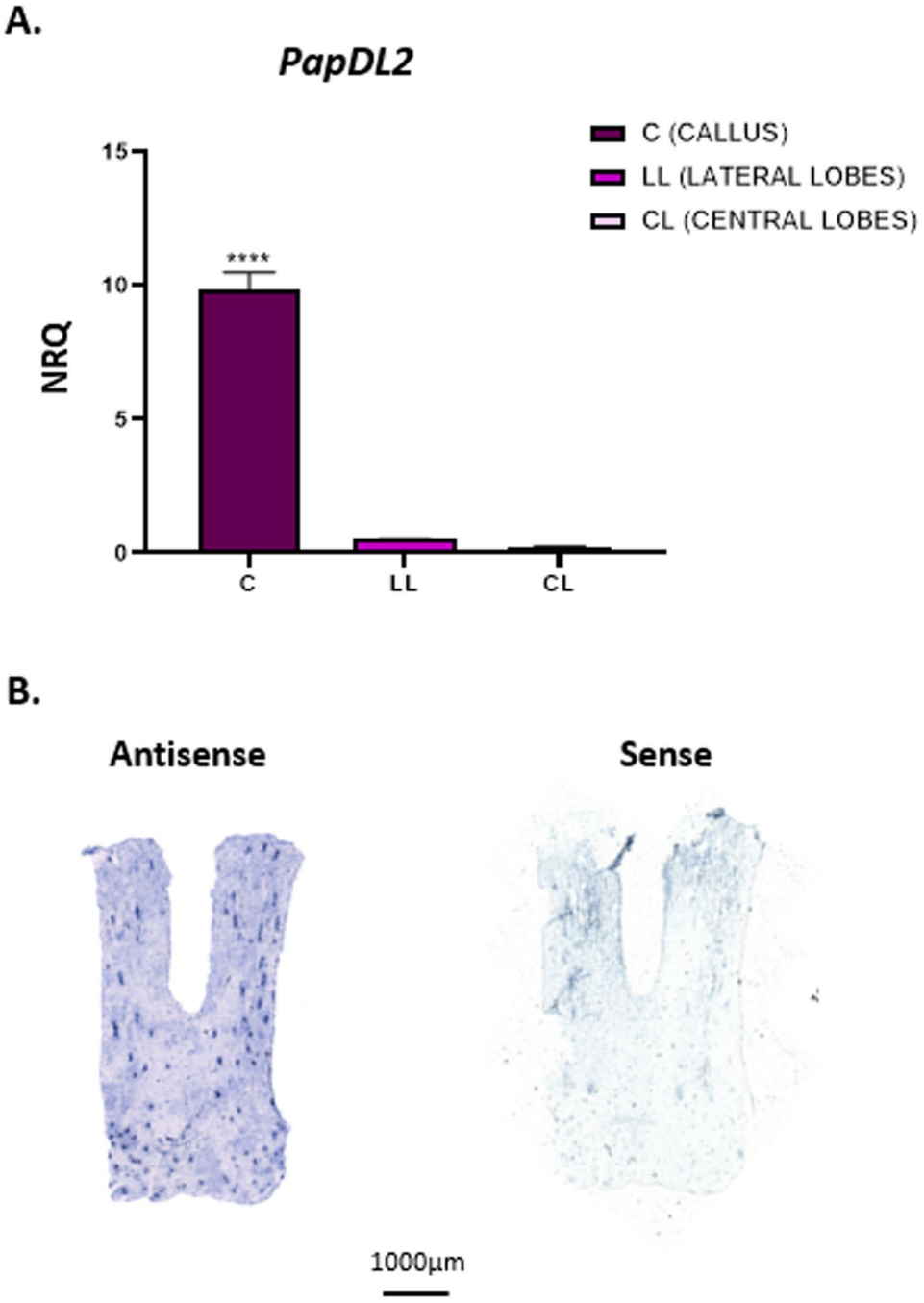

Fig. 8 Expression of *PapDL2* gene in the lip of *P. aphrodite*. Relative expression of *PapDL2* in callus, lateral, and central lobes (**A**); in situ RNA hybridization of the antisense (left) and sense (right) *PapDL2* probes on callus. In (**A**), the expression is reported as NRQ and the bars represent the SEM of the biological and technical replicates. The asterisks indicate the statistically significant difference of the expression compared to the central lobes, with $p < 0.05$ considered the threshold for significance. *p*-values ****<0.0001

presence of regions that encode amino acid stretches of DL in the genomes of *A. shenzhenica* and *A. ramifera*. These regions are located on scaffolds different from the full-length *DL*, and both start with a *DL* open reading frame containing a premature stop codon. Sequences downstream of the stop codon encode retrotransposon protein fragments, followed by additional fragments encoding amino acid stretches similar to DL proteins. The presence of a single full-length *DL* transcript in the floral bud transcriptome of *Apostasia odorata*, assembled in this study, along with the transcriptomes of Apostasioideae *A. wallichi* and *N. zollingeri* available on Orchidstra2.0 [29], further supports this hypothesis.

The accumulation of numerous nucleotide substitutions and indels in the C-terminal region of the duplicated gene *VplDL2* of *Vanilla planifolia* suggests an in-progress pseudogenization process. Indeed, although *VplDL2* encodes a protein of 187 amino acids, with an N-terminal region more similar to DL1 than DL2 proteins, its C-terminal region presents a significant divergence of nucleotides and amino acid sequence from all other DLs. Similarly, in *DL2* of analysed Cypripedioideae, deletions and a premature stop codon interrupt the open reading frame that is shorter than that of other DLs (162 aa), suggesting a possible pseudogenization process acting on these sequences (Additional file 3, Figure A4). The pseudogenization hypothesis gains further support from the absence of the intrinsically disordered motifs (either one or both) in the sequences of VplDL2, PmaDL2, and PmiDL2, a feature shared by all other orchid DL proteins. Increasing evidence suggests the involvement of such motifs in protein functionality [56–58], and their absence in the DL2 sequences of Vanilloideae and Cypripedioideae may be linked to the loss-of-function of these proteins.

The N-terminal region of orchid DL1 and DL2 proteins is well conserved, while the C-terminal (downstream of the YABBY domain) includes distinctive amino acid motifs. The nucleotide alignment suggests that the evolution of the orchid DL proteins has occurred through nucleotide substitutions and frameshift mutations responsible for the diversification of the C-terminal regions between DL1 and DL2. Frameshift mutations have been described as responsible for the evolution of various plant transcription factor families, e.g., MADS-box, MYB [59–61], and seems to be the molecular mechanism leading to the sequence diversification between DL1 and DL2 in orchids.

Molecular evolution analyses do not highlight positive selection acting on specific sites and/or branches of the orchid *DL* genes but rather show relaxed selection acting on the *DL2* branches. Relaxed selection implies reduced selective pressures acting on orchid *DL2*, leading to increased mutation tolerance. In the context of *DL1* and *DL2* paralogs, selection acts asymmetrically, with *DL1* experiencing stronger selective constraints, while *DL2* undergoes relaxed selection, potentially enabling functional divergence or pseudogenization. Both these outcomes seem to have happened during the evolution of the orchid *DL* genes.

Evolution of the expression patterns of orchid *DL1* and *DL2* genes

The nuclear localization of PeqDL2 supports the hypothesis that the duplicated *DL2* gene in *Phalaenopsis* still functions as a transcription factor, as also indicated by the presence of a Cys2Cys2 type zinc-finger domain capable of binding DNA and RNA [62].

Thus, PeqDL2 likely participates in transcriptional regulation, potentially controlling gene expression and contributing to the evolution of new molecular pathways. This hypothesis is consistent with the evolution of *DL/CRC* homologs in angiosperms, which often involves the acquisition of new functions, likely driven by changes in regulatory sequences and coding regions [63, 64]. While maintaining their ancestral role in carpel development across various flowering plants, including basal angiosperms, these genes have diversified functions in different lineages [65]. For example, *CRC* genes contribute to nectary formation in certain rosids and asterids [47, 66]. In *Eschscholzia californica*, *DL* genes have conserved roles in the gynoecium and are involved in placenta development and ovule initiation [67]. Grass *DL* genes influence carpel identity and midrib formation [26]. Additionally, the expression of *DL* genes in leaves may have emerged in monocots, as seen in species like *Asparagus asparagoides* and *Lilium longiflorum* [49, 68, 69]. In *Phalaenopsis*, *DL2* exhibits differential expression between the lip and lateral inner tepal during early flower development, suggesting a role in perianth development [23].

Our expression analyses aim to verify if the *Phalaenopsis DL* expression pattern is conserved across the orchid subfamilies. *DL* paralogs exhibit distinct expression profiles in the examined orchids during early flower development. *DL1* expression is always prominent in the column and ovary, with minimal presence in the perianth, consistent with its conserved role in reproductive organ development. Conversely, *DL2* expression domains vary, showing expression in the column and ovary, albeit less than *DL1*, and expanding to perianth tissues with variable levels and distribution among the different subfamilies.

The absence of a transcribed *DL2* paralog in Apostasioideae, the basal orchid subfamily with actinomorphic flowers, strongly supports the loss-of-function of the duplicated gene, whose presence is still detectable in the genome of *Apostasia*. Similarly, the expression patterns of the duplicated *DLs* of Vanilloideae and Cypripedioideae, basal orchid subfamilies with zygomorphic flowers, agree with the hypothesis of pseudogenization based on the presence of indels and stop codons within their sequences. More in details, the low expression levels of *VplDL2* detected in reproductive organs of Vanilloideae, and its uniform expression in the perianth tissues are consistent with the hypothesis that *VplDL2* could have lost (at least partially) its original redundant role in reproductive structures and have not acquired a specific role in perianth development. Analogously, the low expression of *PloDL2* in column and ovary and the absence of *DL2* differential expression among the perianth tissues of

Cypripedioideae is coherent with loss of redundancy and absence of neofunctionalization.

Contrary to what has been observed in more basal subfamilies, the transcriptional profile of *DL2* in Epidendroideae, one of the most recent subfamilies of orchids, does not indicate pseudogenization processes taking place. As previously documented by studies conducted by our research group, in wild-type *Phalaenopsis*, *DL2* is expressed both in the column and ovary, as well as in the perianth, where it is significantly more expressed in the lip compared to the other tepals and this differential expression is not observed in peloric mutants [23]. Here we demonstrate that this expression profile is conserved in other Epidendroideae, and the expression in lip is localized in the callus-hypochile. This expression pattern is consistent with the conservation of a redundant role of *DL2* in column and ovary, and with the acquisition of a new function in the lip, specifically in the callus. This specialized lip structure originates from a primitive and sterile stamen-like complex, located near fertile organs and exhibits a mixed petaloid-staminodial origin [7, 70]. The ontogenetic origin of the callus in Epidendroideae is reflected in the expression of class A, B, and E MADS-box genes involved in stamen and petal development [7, 70]. Hence, the expression of *DL2* in the lip hypochile, proximal to reproductive organs, is related to the ontogenetic origin of this tissue and to the ancestral role of *DL* genes in angiosperm reproductive organs. Relaxed selective constraints on *DL2* might have allowed sequence divergence and acquisition of new specific role in callus formation of Epidendroideae. Although specialized structures similar to calli are present in other orchid subfamilies, these are not morphologically homologous across different groups [71, 72]. Additionally, the uniform expression of *DL2* in *Phalaenopsis* callus suggests that this transcript is not involved in the polarization of this organ, in contrast to the *CRC* gene in *Arabidopsis*, which is essential for establishing the abaxial tissue identity [73].

The expression of the duplicated *DL2* gene in the callus of more evolutionarily recent orchids is particularly intriguing, given the important role this structure plays in pollination efficiency. Gene duplication events have frequently been linked to the evolution of plant-insect interactions. A well-documented example involves desaturase genes, which encode enzymes responsible for synthesizing key volatile compounds that attract pollinators in the sexual mimicry systems of *Ophrys* orchids [74]. Structural variations or changes in the expression levels of such genes can represent rapid evolutionary adaptations that enhance insect interactions. Furthermore, as illustrated by the "orchid code," the duplication of class B *DEFICIENS*-like *MADS*-box genes has played a critical role in reshaping the floral morphology of orchids, leading to the evolution of zygomorphic symmetry that promotes specialized pollination strategies [75]. In this context, the duplication of *DL*-like genes in orchids may have ecological implications, potentially influencing both callus development and interactions with pollinators. However, functional studies are required to validate this hypothesis.

Conclusions

In conclusion, this study offers a comprehensive analysis of the evolutionary trajectory and expression patterns of *DL* genes in orchids, shedding light on significant genomic and functional dynamics within this plant family. Our findings suggest that gene duplication of *DL* occurred in the common ancestor of orchids, resulting in the retention of two paralogs in Vanilloideae, Cypripedioideae, Orchidoideae, and Epidendroideae, while Apostasioideae possess a single functional gene. Comparative analyses of nucleotide and amino acid sequences indicate that pseudogenization has also influenced the duplicated gene in basal subfamilies Vanilloideae and Cypripedioideae. Moreover, gene expression profiling reveals distinct patterns between *DL1* and *DL2* across various subfamilies. While *DL1* likely regulates the development of reproductive organs, as observed in other angiosperms, our data suggest a potential new role for *DL2* in the development of the lip callus in Epidendroideae, possibly through the activation or repression of specific genes involved in its differentiation. Overall, this research advances our understanding of *DL* gene evolution in orchids and provides a basis for further studies on the function and evolution of transcription factors in plant development and adaptation processes.

Abbreviations

ANOVA	Analysis of Variance
CaMV	Cauliflower Mosaic Virus
CRC	CRABS CLAW
DAPI	4',6-Diamidino-2-Phenylindole
DIG	Digoxigenin
DL	DROOPING LEAF
FDR	False Discovery Rate
GARD	Genetic Algorithm for Recombination Detection
GFP	Green Fluorescent Protein
HOT	Homeotic orchid tepal
LRT	Likelihood Ratio Test
NRQ	Normalized Relative Quantity
PAML	Phylogenetic Analysis by Maximum Likelihood
P-code	Perianth code
qPCR	quantitative Polymerase Chain Reaction
SEM	Standard Error of the Mean
SRA	Sequence Read Archive

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06936-6.

Additional file 1. List of the sequences, sources, and accession numbers of the DL examined.

Additional file 2. Main statistics of the assembled transcriptomes and used genomes available on NCBI.

Additional file 3. List of the primers used.

Additional file 4. Figure A1. Amino acid alignment of PeqDL2 of ***Phalaenopsis equestris*** and AthCRC of ***Arabidopsis thaliana*** showing the position of the zinc-finger (ZF), intermediate (INT), and YABBY (YAB) domains. Figure A2. Logos of the conserved motifs of the orchid DL proteins identified by MEME analysis. Figure A3. The orchid DL amino acid alignments. The black arrow indicates the position of the predicted recombination breakpoint. Figure A4. Genomic organization of the ***DL/CRC*** genes of ***Arabidopsis thaliana*** (***AthCRC***)**,** ***Triticum aestivum*** (***TaeCRC***)**,** ***Oryza sativa*** (***OsaDL***)**,** ***Apostasia ramifera*** (***AraDL***)**,** ***A. shenzhenica*** (***AshDL***)**,** ***Vanilla planifolia*** (***VplDL1***, ***VplDL2***)**,** and ***Phalaenopsis equestris*** (***PeqDL1***, ***PeqDL2***)**.** Figure A5. Alignment of the ***DL*** sequences spanning from nucleotide position 481 to 586. The gray boxes represent the deletions of the ***DL2*** sequences of the Cypripedioideae analyzed. The stars indicate premature stop codons. Figure A6. Expression of ***DL*** genes in ***Rhyncholaeliocattleya*** perianth. Schematic representation of the ***Rhyncholaeliocattleya*** Beauty Girl "KOVA" flower (A); heatmap of the expression pattern of ***RhyDL1 and RhyDL2*** in different perianth organs of ***Rhyncholaeliocattleya*** at developmental stage S1 (bud length < 2 cm) (B). The heatmap and the relative color gradient are obtained from the TMM (Trimmed Mean of M-values) read counts of the DL transcripts. Table A1. Results of the ***in silico*** differential expression analyses of the ***DL2*** gene.

Additional file 5. Regions of genomic scaffolds of ***Apostasia shenzhenica*** and ***A. ramifera*** containing truncated open reading frames of ***DL***.

Additional file 6. Pairwise percentage of nucleotide identity and amino acid similarity and identity of the orchid DLs.

Additional file 7. Results of the evolutionary PAML analysis on the orchid ***DL*** coding sequences.

Acknowledgements
We are deeply grateful to Dr. Mariana Mondragon-Palomino for inspiring us to further investigate the role of DL genes in orchids and for generously providing us with RNA from Vanilla, Phragmipedium, and Phalaenopsis produced in her previous studies. Furthermore, we acknowledge the invaluable support of former Professor Giovanni Iazzetti in the fixation of orchid tissues and preparation of slides for our in situ hybridization experiments.

Authors' contributions
F.L., A.B. and S.A. designed the experiments; F.L., A.C. and K.E. performed the experiments; F.L., A.B. and S.A. analyzed the data; F.L. and S.A. wrote the article. All the authors have read and agreed to the submitted version of the manuscript.

Funding
Work funded by National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union– NextGenerationEU; Award Number: Project code CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP H43C22000530001, Project title "National Biodiversity Future Center - NBFC".

Data availability
All RNA-seq data in this study were downloaded from the NCBI (https://www.ncbi.nlm.nih.gov/), with biological projects PRJNA310678, PRJNA237672, PRJNA918555, PRJNA559603, PRJNA278061, PRJNA880415, PRJNA635894, PRJNA192198, PRJNA662181, PRJNA777357. Database consulted: Orchidstra2 (https://orchidstra2.abrc.sinica.edu.tw/orchidstra2/index.php), CNGB Sequence Archive (https://db.cngb.org/cnsa/), MaizeGDB (https://www.maizegdb.org).

Declarations

Ethics approval and consent to participate
The study was carried out in compliance with relevant institutional, national, and international guidelines and legislation for plant ethics.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 30 April 2024 / Accepted: 20 June 2025
Published online: 08 July 2025

References

1. Zhang D, Zhao XW, Li YY, Ke SJ, Yin WL, Lan S, Liu ZJ. Advances and prospects of Orchid research and industrialization. Hortic Res. 2022;9:uhac220.
2. Darwin C. On the Various Contrivances by Which British and Foreign Orchids are Fertilised by Insects. Cambridge, UK: Cambridge University Press; 2011.
3. Cozzolino S, Widmer A. Orchid diversity: an evolutionary consequence of deception? Trends Ecol Evol. 2005;20:487–94.
4. Aceto S, Gaudio L. The MADS and the beauty: genes involved in the development of Orchid flowers. Curr Genomics. 2011;12:342–56.
5. Rudall PJ, Bateman RM. Roles of synorganisation, zygomorphy and heterotopy in floral evolution: the gynostemium and labellum of orchids and other lilioid monocots. Biol Rev Camb Philos Soc. 2002;77:403–41.
6. Bateman RM, Rudall PJ. The good, the bad, the ugly: using naturally occurring terata to distinguish the possible from the impossible in Orchid floral evolution. Aliso. 2006;22:481–96.
7. Pramanik D, Dorst N, Meesters N, Spaans M, Smets E, Welten M, Gravendeel B. Evolution and development of three highly specialized floral structures of bee-pollinated Phalaenopsis species. Evodevo. 2020;11:16.
8. Coen ES, Meyerowitz EM. The war of the whorls: genetic interactions controlling flower development. Nature. 1991;353:31–7.
9. Theissen G, Saedler H. Plant biology. Floral quartets. Nature. 2001;409:469–71.
10. Kanno A, Nakada M, Akita Y, Hirai M. Class B gene expression and the modified ABC model in Nongrass monocots. Sci World J. 2007;7:268–79.
11. Mondragon-Palomino M, Theissen G. Why are Orchid flowers so diverse? Reduction of evolutionary constraints by paralogues of class B floral homeotic genes. Ann Bot. 2009;104:583–94.
12. Mondragon-Palomino M, Theissen G. Conserved differential expression of paralogous DEFICIENS- and GLOBOSA-like MADS-box genes in the flowers of orchidaceae: refining the 'orchid code'. Plant J. 2011;66:1008–19.
13. Mondragon-Palomino M, Theissen G. MADS about the evolution of Orchid flowers. Trends Plant Sci. 2008;13:51–9.
14. Pan Z, Cheng C, Tsai W, Chung M, Chen W, Hu J. The duplicated B-class MADS-Box genes display dualistic characters in Orchid floral organ identity and growth. Plant Cell Physiol. 2011;52:1515–31.
15. Hsu H, Hsu W, Lee Y, Mao W, Yang J, Li J, Yang C. Model for perianth formation in orchids. Nat Plants. 2015;1:15046.
16. Valoroso MC, De Paolo S, Iazzetti G, Aceto S. Transcriptome-Wide identification and expression analysis of DIVARICATA- and RADIALIS-Like genes of the mediterranean Orchid Orchis Italica. Genome Biol Evol. 2017;9:6.
17. Valoroso MC, Sobral R, Saccone G, Salvemini M, Costa MMR, Aceto S. Evolutionary conservation of the Orchid MYB transcription factors DIV, RAD, and DRIF. Front Plant Sci. 2019;10:1359.
18. De Paolo S, Gaudio L, Aceto S. Analysis of the TCP genes expressed in the inflorescence of the Orchid Orchis Italica. Sci Rep. 2015;5:16265.
19. Lucibelli F, Valoroso MC, Aceto S. Radial or bilateral?? The molecular basis of floral symmetry. Genes (Basel). 2020;11:4.
20. Lin YF, Chen YY, Hsiao YY, Shen CY, Hsu JL, Yeh CM, Mitsuda N, Ohme-Takagi M, Liu ZJ, Tsai WC. Genome-wide identification and characterization of TCP genes involved in ovule development of Phalaenopsis equestris. J Exp Bot. 2016;67:5051–66.
21. Valoroso MC, Lucibelli F, Aceto S. Orchid NAC Transcription Factors: A Focused Analysis of CUPULIFORMIS Genes. Genes (Basel). 2022;13(12):2293.
22. De Paolo S, Salvemini M, Gaudio L, Aceto S. De Novo transcriptome assembly from inflorescence of Orchis italica: analysis of coding and non-coding transcripts. PLoS ONE. 2014;9:e102155.

23. Lucibelli F, Valoroso MC, Theissen G, Nolden S, Mondragon-Palomino M, Aceto S. Extending the toolkit for beauty: differential Co-Expression of DROOPING LEAF-Like and class B MADS-Box genes during Phalaenopsis flower development. Int J Mol Sci. 2021;22(13):7025.
24. Chen YY, Hsiao YY, Chang SB, Zhang D, Lan SR, Liu ZJ, Tsai WC. Genome-Wide identification of YABBY genes in Orchidaceae and their expression patterns in Phalaenopsis Orchid. Genes (Basel). 2020;11(9):955.
25. Gross T, Becker A. Transcription factor action orchestrates the complex expression pattern of CRABS CLAW in Arabidopsis. Genes (Basel). 2021;12(11):1663.
26. Ishikawa M, Ohmori Y, Tanaka W, Hirabayashi C, Murai K, Ogihara Y, Yamaguchi T, Hirano HY. The Spatial expression patterns of DROOPING LEAF orthologs suggest a conserved function in grasses. Genes Genet Syst. 2009;84:137–46.
27. Piluk P, Ratanasut K. Expression profiles of the dihydroflavonol 4-reductase (DFR) gene inthe sepals and petals of Dendrobium Sonia Earsakul. In: Proceedings of the 1st MaeFah Luang University International Conference, Chiang Rai,Thailand. 29 Nov.–1 Dec. 2012. Mae Fah Luang Univ., Chi-ang Rai. p. 1–8. https://conf.mfu.ac.th/archive/mfuic2012/electronic_proceeding/Documents/00_PDF/O-SC-B/O-SC-B-1%20Pornnapa.pdf.
28. Chen TK, Yang HT, Fang SC, Lien YC, Yang TT, Ko SS. Hybrid-Cut: an improved sectioning method for recalcitrant plant tissue samples. J Vis Exp. 2016;23(117):54754.
29. Chao YT, Yen SH, Yeh JH, Chen WC, Shih MC. Orchidstra 2.0-A transcriptomics resource for the Orchid family. Plant Cell Physiol. 2017;58:e9.
30. Grabherr MGHB, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.
31. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol. 2000;302:205–17.
32. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34:W609–612.
33. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.
34. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37:W202–208.
35. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol. 2006;23:1891–901.
36. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol. 2018;35:773–7.
37. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.
38. Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol. 2015;32:820–32.
39. Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38:4647–54.
40. Li BJZB, Wang JY, Tsai WC, Lu HC, Zou LH, Wan X, Zhang DY, Qiao HJ, Liu ZJ, Wang Y. New insight into the molecular mechanism of colour differentiation among floral segments in orchids. Commun Biol. 2020;3:89.
41. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
42. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. 2007;8:R19.
43. Gross T, Broholm S, Becker A. CRABS CLAW acts as a bifunctional transcription factor in flower development. Front Plant Sci. 2018;9:835.
44. Cutler SR, Ehrhardt DW, Griffitts JS, Somerville CR. Random gfp::cdna fusions enable visualization of subcellular structures in cells of Arabidopsis at a high frequency. Proc Natl Acad Sci U S A. 2000;97:3718–23.
45. Zhang GQ, Liu KW, Li Z, Lohaus R, Hsiao YY, Niu SC, Wang JY, Lin YC, Xu Q, Chen LJ, et al. The apostasia genome and the evolution of orchids. Nature. 2017;549:379–83.
46. Zhang W, Zhang G, Zeng P, Zhang Y, Hu H, Liu Z, Cai J. Genome sequence of apostasia ramifera provides insights into the adaptive evolution in orchids. BMC Genomics. 2021;22:536.
47. Bowman JL, Smyth DR. CRABS CLAW, a gene that regulates carpel and nectary development in arabidopsis, encodes a novel protein with zinc finger and helix-loop-helix domains. Development. 1999;126:2387–96.
48. Hao L, Zhang J, Shi S, Li P, Li D, Zhang T, Guo H. Identification and expression profiles of the YABBY transcription factors in wheat. PeerJ. 2022;10:e12855.
49. Yamaguchi T, Nagasawa N, Kawasaki S, Matsuoka M, Nagato Y, Hirano HY. The YABBY gene DROOPING LEAF regulates carpel specification and midrib development in Oryza sativa. Plant Cell. 2004;16:500–9.
50. Panchy N, Lehti-Shiu M, Shiu SH. Evolution of gene duplication in plants. Plant Physiol. 2016;171:2294–316.
51. Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. 2017;45:D1040–5.
52. Airoldi CA, Davies B. Gene duplication and the evolution of plant MADS-box transcription factors. J Genet Genomics. 2012;39:157–65.
53. Lu YH, Alam I, Yang YQ, Yu YC, Chi WC, Chen SB, Chalhoub B, Jiang LX. Evolutionary Analysis of the YABBY Gene Family in Brassicaceae. Plants (Basel). 2021;10:2700.
54. Phukela BGR, Das S, Tandon R. Ancestral segmental duplication in Solanaceae is responsible for the origin of CRCa-CRCb paralogues in the family. Mol Genet Genomics. 2020;295:563–77.
55. Strable J, Vollbrecht E. Maize YABBY genes drooping leaf1 and drooping leaf2 regulate floret development and floral meristem determinacy. Development. 2019;146:dev171181.
56. Martins IC, Santos NC. Intrinsically disordered protein domains in flavivirus infection. Arch Biochem Biophys. 2020;683:108298.
57. Brodsky S, Jana T, Mittelman K, Chapal M, Kumar DK, Carmi M, Barkai N. Intrinsically disordered regions direct transcription factor in vivo binding specificity. Mol Cell. 2020;79:459–e471454.
58. Sun X, Jones WT, Rikkerink EH. GRAS proteins: the versatile roles of intrinsically disordered proteins in plant signalling. Biochem J. 2012;442:1–12.
59. Vandenbussche MTG, Van de Peer Y, Gerats T. Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. Nucleic Acids Res. 2003;31(15):4401–9.
60. Kramer EM, Su HJ, Wu CC, Hu JM. A simplified explanation for the frameshift mutation that created a novel C-terminal motif in the APETALA3 gene lineage. BMC Evol Biol. 2006;6:30.
61. Rodrigues JA, Espley RV, Allan AC. Genomic analysis uncovers functional variation in the C-terminus of anthocyanin-activating MYB transcription factors. Hortic Res. 2021;8:77.
62. Burdach J, O'Connell MR, Mackay JP, Crossley M. Two-timing zinc finger transcription factors liaising with RNA. Trends Biochem Sci. 2012;37:199–205.
63. Lee JY, Baum SF, Alvarez J, Patel A, Chitwood DH, Bowman JL. Activation of CRABS CLAW in the nectaries and carpels of Arabidopsis. Plant Cell. 2005;17:25–36.
64. Fourquin C, Vinauger-Douard M, Chambrier P, Berne-Dedieu A, Scutt CP. Functional conservation between CRABS CLAW orthologues from widely diverged angiosperms. Ann Bot. 2007;100:651–7.
65. Fourquin C, Vinauger-Douard M, Fogliani B, Dumas C, Scutt CP. Evidence that CRABS CLAW and TOUSLED have conserved their roles in carpel development since the ancestor of the extant angiosperms. Proc Natl Acad Sci U S A. 2005;102:4649–54.
66. Lee JY, Baum SF, Oh SH, Jiang CZ, Chen JC, Bowman JL. Recruitment of CRABS CLAW to promote nectary development within the eudicot clade. Development. 2005;132:5021–32.
67. Orashakova S, Lange M, Lange S, Wege S, Becker A. The CRABS CLAW ortholog from California poppy (Eschscholzia californica, Papaveraceae), EcCRC, is involved in floral meristem termination, gynoecium differentiation and ovule initiation. Plant J. 2009;58:682–93.
68. Nakayama H, Yamaguchi T, Tsukaya H. Expression patterns of aadl, a CRABS CLAW ortholog in Asparagus asparagoides (Asparagaceae), demonstrate a Stepwise evolution of CRC/DL subfamily of YABBY genes. Am J Bot. 2010;97:591–600.
69. Wang A, Tang J, Li D, Chen C, Zhao X, Zhu L. Isolation and functional analysis of LiYAB1, a YABBY family gene, from Lily (Lilium longiflorum). J Plant Physiol. 2009;166:988–95.
70. Dirks-Mulder A, Butot R, van Schaik P, Wijnands JW, van den Berg R, Krol L, Doebar S, van Kooperen K, de Boer H, Kramer EM, et al. Exploring the

evolutionary origin of floral organs of Erycina pusilla, an emerging Orchid model system. BMC Evol Biol. 2017;17:89.
71. Swamy BGL. Vascular anatomy of Orchid flowers. Bot Museum Leaflets. 1948;13:61–95.
72. Deressler RL. Phylogeny and classification of the Orchid family. Cambridge, UK: Cambridge University Press; 1993. p. 30.
73. Eshed Y, Baum SF, Bowman JL. Distinct mechanisms promote Polarity establishment in carpels of Arabidopsis. Cell. 1999;99:199–209.
74. Clare EL, Schiestl FP, Leitch AR, Chittka L. The promise of genomics in the study of plant-pollinator interactions. Genome Biol. 2013;14:207.
75. Mondragon-Palomino M, Hiese L, Harter A, Koch MA, Theissen G. Positive selection and ancient duplications in the evolution of class B floral homeotic genes of orchids and grasses. BMC Evol Biol. 2009;9:81.

Publisher's note