University of Naples “Federico II”

SCUOLA POLITECNICA E DELLE SCIENZE DI BASE
AREA DIDATTICA DI SCIENZE MATEMATICHE, FISICHE E NATURALI

PhD in Biology (XXXI Cycle)

PhD Coordinator: Salvatore Cozzolino

Analysis of the genes involved in floral symmetry of the orchid *Orchis italica*

PhD Student:
Maria Carmen Valoroso

PhD Supervisor:
Prof. Serena Aceto

Anno accademico: 2017-2018

Summary

Abstract

Orchis italica è un'orchidea selvatica mediterranea appartenente alla sottofamiglia Orchidoideae delle Orchidaceae, una delle più vaste famiglie di piante a fiore. Lo scopo di questo lavoro è lo studio dei geni coinvolti nella determinazione della simmetria bilaterale del fiore di *O. italica*. Studi sulla determinazione della simmetria fiorale effettuati sulla specie modello *Antirrhinum majus* hanno evidenziato che cinque geni, due appartenenti alla famiglia TCP (*CYC* e *DICH*) e tre appartenenti alla famiglia MYB di fattori di trascrizione (*DIV, RAD* e *DRIF*), svolgono un ruolo cruciale nella formazione del fiore zigomorfo. In *A. majus* queste proteine interagiscono tra di loro attraverso un meccanismo antagonistico che permette la formazione nel fiore di una regione dorsale e una ventrale. In *O. italica*, l'analisi del trascrittoma dell'infiorescenza matura ha evidenziato la presenza di 8 trascritti *DIV*-like, 4 trascritti *RAD*-like e due trascritti *DRIF1/2*. L'organizzazione genomica dei geni *DIV*-like e *RAD*-like presenta un singolo introne, fatta eccezione per un unico gene *RAD*-like che risulta essere privo di introni. L'analisi evolutiva ha evidenziato che sulle regioni codificanti dei geni *DIV*-like e *RAD*-like agisce selezione purificante. I geni *DIV*, *RAD* e *DRIF* di *O. italica* hanno un'espressione conservata rispetto ad *A. majus.* L'analisi delle interazioni proteiche tra i fattori di trascrizione MYB di *O. italica* ha dimostrato che il modello alla base della determinazione della simmetria bilaterale sembra essere conservato anche in orchidea.

I geni MADS-box di tipo II MIKC^{C}, classificati in cinque classi funzionali ABCDE, hanno un ruolo cruciale della determinazione della struttura e dell'organizzazione fiorale. In particolare, in orchidea, secondo il modello "*orchid code*", i geni MADS-box appartenenti alla classe B hanno un ruolo fondamentale nella formazione della struttura zigomorfa del fiore di orchidea. Nell'infiorescenza matura di *O. italica* sono espressi 29 trascritti MADS-box appartenenti al tipo I e al tipo II. I geni MADS-box di *O. italica* hanno un profilo di espressione in linea con il modello di sviluppo fiorale "*fading borders*".

Poiché entrambe le famiglie geniche, MADS-box e MYB, sono alla base della determinazione fiorale è stato effettuato uno studio sull'interazione proteica tra i fattori di trascrizione MYB (DIV, RAD e DRIF) e le proteine MADS-box di classe B di *O. italica*. L'analisi condotta ha evidenziato che le singole proteine MADS-box appartenenti alla classe B non sono in grado di interagire con i fattori di trascrizione MYB.

CHAPTER 1 – The molecular basis of flower symmetry

1. Introduction

1.1 The Orchidaceae family: a focus on *Orchis italica*

The origin and diversification of angiosperms, defined by Charles Darwin as “an abominable mystery”, has been subject of many studies during the last 100 years [1].The emergence of molecular techniques has improved the understanding of angiosperm phylogeny and evolution [2-5]. Among the flowering plants, the monocot family Orchidaceae is one of the most species-rich and widespread in the world, adapted to different habitats (epiphytic and terrestrial) and with highly specialized reproductive strategies. This family includes more than 28,000 species divided in five subfamilies: Apostasioideae, Cypripedioideae, Vanilloideae, Epidendroideae and Orchidoideae [6]. Each subfamily includes numerous tribes and subtribes [7, 8] (Fig. 1). The origin of Orchidaceae has been placed ~112 million years ago (Mya). The different subfamilies started their diversification ~90 Mya and the separation between Orchidoideae and Epidendroideae happened ~64 Mya [9].

Despite the high morphological difference among the flowers of the orchid species, they share common features. The orchid flower is generally characterised by bilateral symmetry (zygomorphy) and it is composed by three outer tepals, two lateral inner tepals and an inner medial tepal, called **labellum** (or lip), highly diversified [10]. The orchid’s reproductive structure, called **gynostemium** or **column**, is composed by the fusion of male (stamen/anther) and female (pistil/stigma) tissue. The pollen grains (pollinia) are located at the top of the column, while at its base is located the ovary, whose maturation is triggered by pollination [10, 11] (Fig. 2A). In addition, many orchid species show a resupinate flower, where the pedicel and ovary undergo a 180° degree twisting process resulting in the ventral position of the lip in the mature flower [12] (Fig. 2B).

The shape and pigmentation of the lip and its abaxial orientation in the opposite part of the fertile anther after resupination suggest that these characters are adaptations to specific pollinators [6, 13].

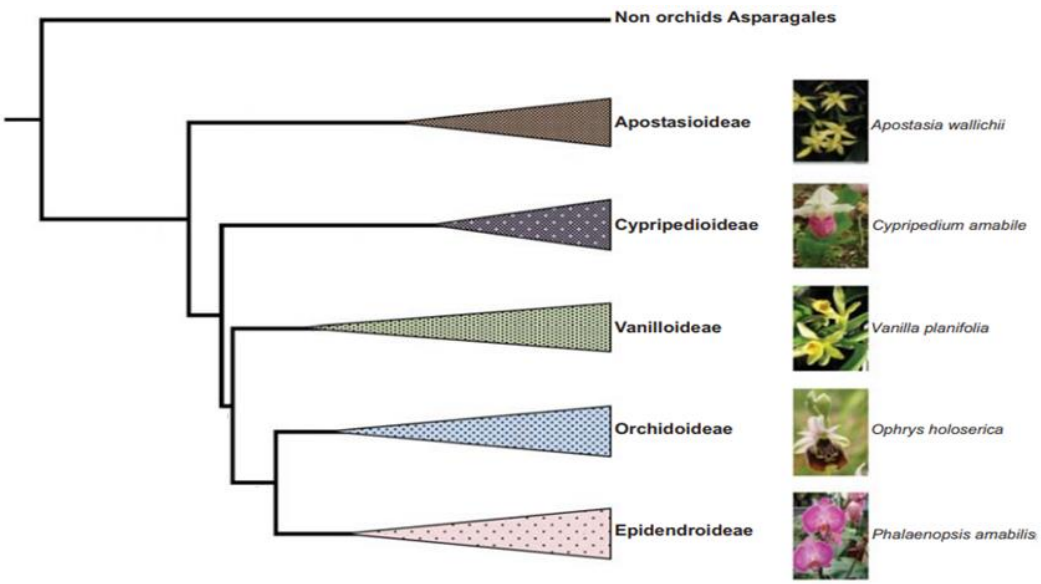

Fig. 1 Phylogenetic tree of the Orchidaceae family. Modified from Aceto *et al.* (2011) [6]. On the right there are images of representative orchid species of the different subfamilies.

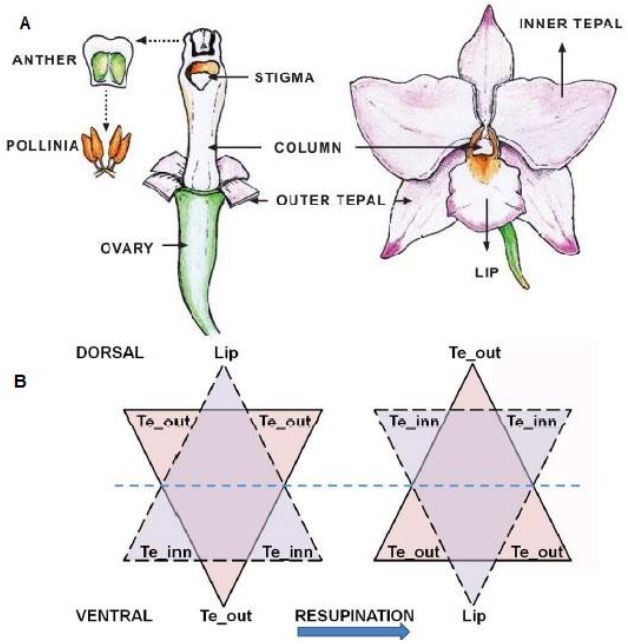

Fig. 2 The orchid flower structure and the resupination of the flower. (A) Reprinted from *Aceto et al.* (2011) [6]. Structure of an orchid flower with a detail of the reproductive structures. **(B)** Reprinted from *Valoroso et. al* (2017) [14]. The diagram shows the position of the organs of the first (pink triangle) and second (violet triangle) whorls of the perianth before (left) and after (right) the resupination.

Orchis italica is an orchid species belonging to the subfamily Orchidoideae. It is one of most widespread Mediterranean orchids, characterized by a white-purple cluster inflorescence (Fig. 3A) and a lip with white-purple flaps that assume an anthropomorphic form (Fig. 3B).

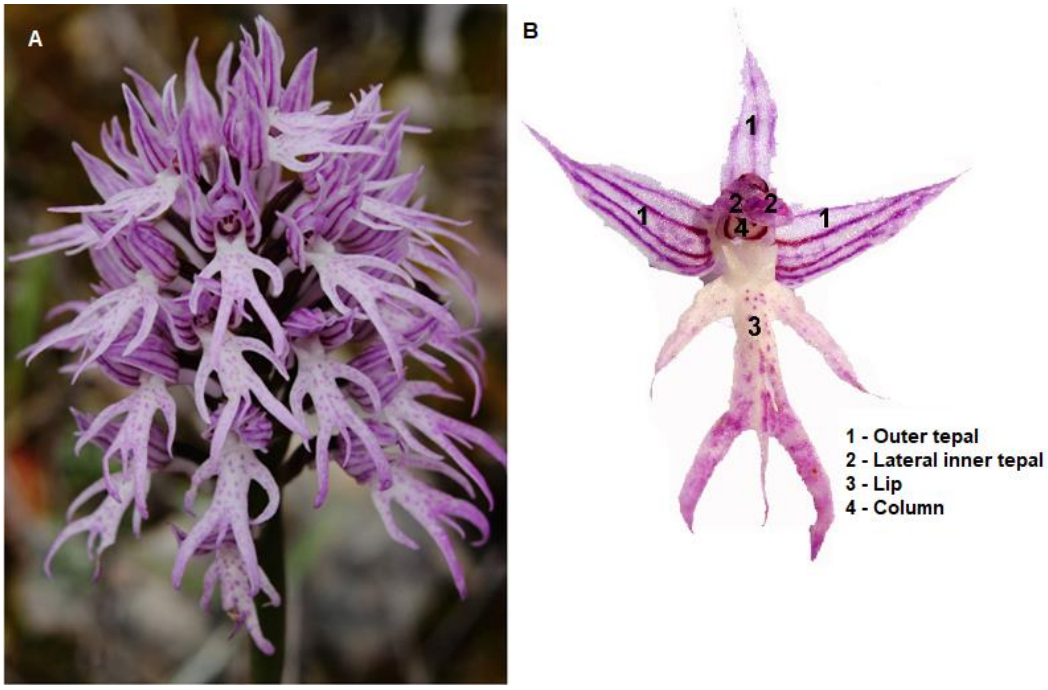

Fig. 3 *Orchis italica*. Reprinted from *Valoroso et al.* (2017) [14]. **(A)** The inflorescence and **(B)** the flower of *Orchis italica*.

1.2 The molecular basis of bilateral symmetry of the flower

During the evolution of flowering plants, actinomorphy (radial symmetry) (Fig. 4) represents the ancestral state of flower symmetry. The first changes from radial symmetry have evolved during the first angiosperm radiation (Turonian age, late Cretaceous), with the appearance of asymmetric flowers that can be considered “precursors” of zygomorphic flowers [15, 16]. The analysis of fossil records indicates that zygomorphy evolved in several plant lineages during the same period, as well as the rise of some bee families, supporting the idea that there was a coevolution between flower structure, symmetry and specific pollinating insects [17].

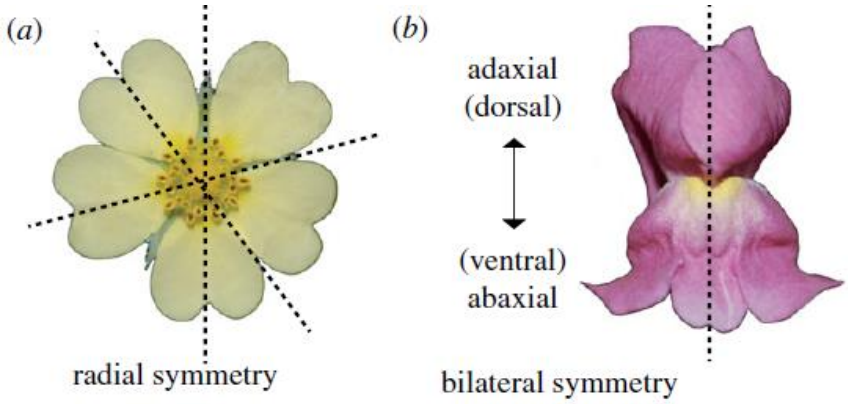

Fig. 4 Difference between radial and bilateral symmetry of the flower. Reprinted from *Hileman (2014)* [18]. (a) Flower with radial symmetry characterized by multiple planes of mirror image symmetry; (b) flower with bilateral symmetry characterized by a single plane of mirror image symmetry.

The molecular basis underlying the establishment of bilateral symmetry of the flower has been dissected in the model plant *Antirrhinum majus* (snapdragon), where a key role is played by the genes *CYCLOIDEA (CYC), DICHOTOMA (DICH), RADIALIS (RAD), DIVARICATA (DIV)* and *DIV and RAD INTERACTING FACTOR 1 (DRIF1)*.

1.2.1The TCP family: CYCLOIDEA and DICHOTOMA

TCPs are plant-specific transcription factors whose name derives from four proteins that share the TCP domain: **T**EOSINTE BRANCHED 1 (Tb1) from *Zea mays* [19], **C**YCLOIDEA (CYC) from *Antirrhinum majus* [20] and **P**ROLIFERATING CELL FACTORS 1 and 2 (PCF1 and PCF2) from *Oryza sativa* [21]. The TCP factors contain two conserved regions: the **TCP** and the **R** domain. The predicted structure of the TCP domain is a basic helix–loop–helix (bHLH) composed of 21 amino acids with DNA binding function [22]. The R domain is rich of polar residues (arginine, lysine and glutamic acid) and, together with the TCP domain, forms coiled-coil structures (leucine zippers) [22-24].

Based on the sequence of the TCP domain, the members of this family are divided in two different groups: TCP-P (class I) and TCP-C (class II) [22, 25, 26]. These two classes have different but overlapping consensus binding sites that are GGNCCCAC for class I and G(T/C)GGNCCC for class II [27-29]. Both these consensus motifs are over-represented in the promoters of cell cycle and protein synthesis related genes [28, 30, 31]. The TCP-C factors are further divided into two clades: ECE (CYC/TB1-like), characterized by both the TCP and R domain, and CINCINNATA (CIN), generally showing only the TCP domain [25,

26]. Phylogenetic analysis of the ECE clade shows that it is divided in three subgroups (CYC1, CYC2 and CYC3), raised after duplication events from the ancestral ECE sequence [32].

The transcription factors encoded by the TCP genes are involved many processes of plant growth and development [18, 22, 33]. An important role in the establishment of bilateral symmetry is played by the ECE clade, in particular by the *CYC* and *DICH* genes [18, 26, 32]. During the evolution of zygomorphy, duplications and mutations within the subgroup CYC2 have probably facilitated the transition from radial to bilateral flower symmetry [34].

In *A. majus* the *CYC* and *DICH* genes are involved in the establishment of the lateral and dorsal part of the flower. The snapdragon wild-type flower (bilateral symmetry) has one ventral, two dorsal and two lateral petals (Fig. 5A). In the wild type flower, *CYC* is expressed in the dorsal and lateral domains and its expression is stable until the late stage of development [20]. The loss of function *cyc* mutant (Fig. 5B), named semipeloric, has five or six petals, four with ventral identity and two with combination of lateral and dorsal identity. In this mutant, the bilateral plan of symmetry is misaligned respect to the dorsal-ventral axis. The expression pattern of *CYC* and the phenotype of the semipeloric *cyc* mutant have suggested the involvement of this gene in the establishment of the dorsal and lateral identity and in the organ placement of the snapdragon flower [20]. The *DICH* gene is expressed in the dorsal domain, but differently from *CYC,* it is restricted to the most dorsal part of the flower. The *dich* mutants show partial loss of the dorsal petals asymmetry [35]. These evidences have suggested that also the *DICH* gene is related to the establishment of the identity and asymmetry of the dorsal petals [35].The role of *CYC* and *DICH* in the dorsal and lateral asymmetry of *A. majus* has been further confirmed by the phenotype of the peloric double mutants *cyc:dich* (Fig. 5C)*,* radially symmetric and with ventral identity of all petals [20].

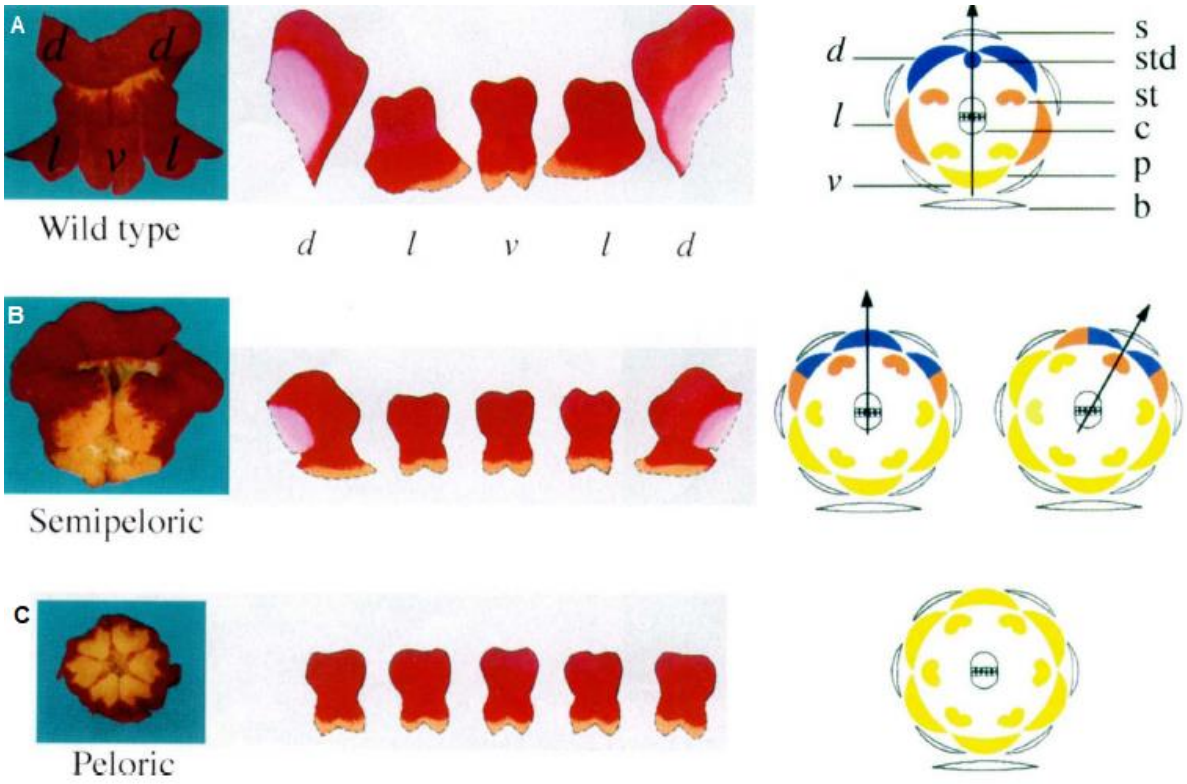

Fig. 5 Wild type and mutant flower of *Antirrhinum majus*. Reprinted from *Luo et al. (1996)* [20]. **(A)** Wild-type, **(B)** semipeloric *cyc* mutant and **(C)** peloric *cyc:dich* double mutant flower of *A. majus*, their different petals and diagram of their symmetry.

1.2.2 The MYB family: RADIALIS, DIVARICATA and their interacting factor

The MYB transcription factors have been described in all the eukaryotic organisms [36, 37]. The first MYB gene isolated in plants, more than 30 years ago, was *COLORED1* from *Zea mays*, involved in the anthocyanin synthesis [38]. The MYB transcription factors are characterized by the presence of MYB repeats (R), from one to four or more. Each repeat is long 52 amino acids and contains three regularly spaced residues of tryptophan, or other aliphatic amino acids, that form a hydrophobic core [39, 40]. Each repeat adopts a helix-turn-helix conformation with DNA binding and protein-protein interaction function [39, 41].

Based on the number of R repeats, the MYB proteins are divided into three groups [39, 42]:

1. **MYB3R**: showing three MYB repeats (R1, R2, R3), involved in the control of cyclins during the late G_2 and M phase of the cell cycle.
2. **R2R3-MYB:** with two similar MYB repeats (R2-R3) and very variable C-terminus. Frequently, the C-terminus contains a transcriptional activation or repression domain composed by residues of serine and threonine. These proteins represent the largest

part of the plant MYB factors and are implicated in many functions, from metabolism regulation to biosynthesis and floral organ determination.

3. **Other MYB-type:** this group includes all the other forms of MYB proteins with a variable number of R repeats. Based on their structure, they can be divided into four sub-groups:
    - **R-MYB,** with a single R repeat, involved in the chromatin and histone metabolism;
    - The second sub-group includes proteins related to the evolutionary old R1/R2 class, involved in the circadian clock control;
    - **GARP,** involved in the control of organ polarity;
    - **4R-MYB** that contains four repeats of R1/R2 domains.

In *A. majus*, three genes related to the establishment of bilateral symmetry belong to the MYB family: *DIV*, *RAD* and *DRIF1*.

DIV belongs the R2R3-class of MYB proteins and contains two similar MYB domains [14, 43]. The DIV transcription factor is implicated in the ventralization of the snapdragon flower and it is expressed in all the whorls of the flower. In the loss of function *div* mutant (radial symmetry) the ventral petals adopt the shape of dorsal and lateral petals (Fig. 6) [43, 44]. The *div:cyc* mutant has radial symmetry and the ventral petals have lateral identity, as in the single *cyc* mutant. The *div:dich* mutant has the ventral petals similar to the lateral petals; however, the dorsal petals resemble the phenotype of the *dich* single mutant. Finally, in the triple mutant *div:cyc:dich* (Fig. 6) all the petals have lateral identity, suggesting that the flower domain influenced by *DIV* is extended to all the flower and that both genes, *CYC* and *DICH*, are necessary to inhibit the DIV function in the dorsal part of the flower [44].

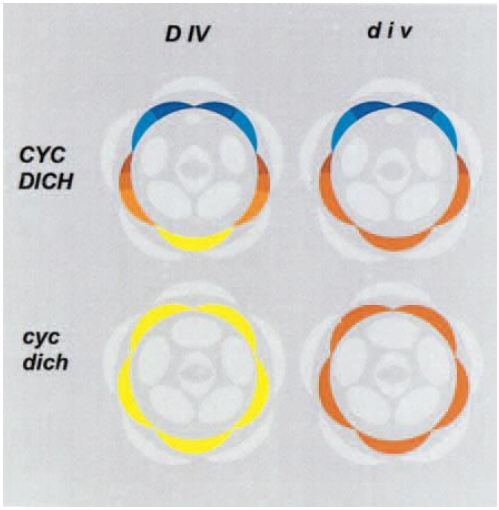

Fig. 6 *Antirrhinum majus div* mutants. Reprinted from *Galego and Almeida (2002) [43]*. Floral diagram of the wild-type and *div* mutant (first line) and *cyc:dich* double and *cyc:dich:div* triple mutant (second line) of *A. majus*.

RAD belongs to the third group of MYB proteins. It is a small protein that contains only one MYB domain and is expressed in the dorsal part of the snapdragon flower [14, 45]. The phenotype of different snapdragon single and double mutants reveals the involvement of the *RAD* gene in the dorsalization of the flower and its connection with the activity of the *CYC* gene. Indeed, all the *rad* mutants have petals that resemble the wild type ventral ones, with only the dorsal half of each petal showing dorsal lateral identity (Fig. 7). In addition, the double mutant *cyc:rad* shows a fully ventralized phenotype, while the double mutant *rad:div* shows the features of both the single mutants [45].

The DRIF proteins are MYB-like transcription factors that display a MYB-like domain at the N-terminus and a domain with unknown function at the C-terminus [46]. Phylogenetic analysis has shown that DRIFs can be divided into two sub-groups, where the group 1 contains both DRIF1 and DRIF2 of *A. majus*, involved in the establishment of bilateral symmetry.

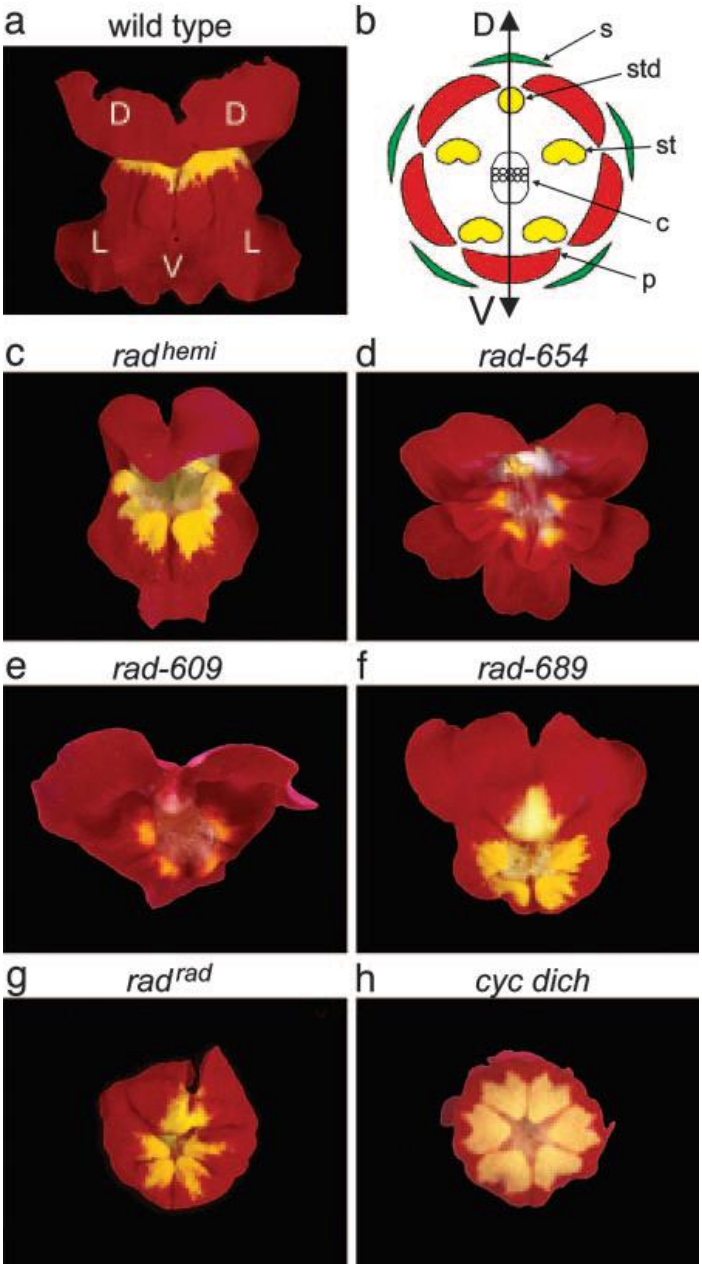

Fig. 7 The different snapdragon *rad* mutants. Reprinted from Corley *et al.* (2005) [45]. (a) *A. majus* wild-type flower; (b-g) different *rad* mutants, from the weaker to the strongest; (h) the peloric *cyc:dich* mutant.

Previous studies conducted in tomato have shown that the DRIF proteins can interact with DIV and RAD [47]. More recently, these interactions have been highlighted also in *A. majus*. In particular, both DIV and RAD can interact with DRIF1 and DRIF2 and the cellular localization of these latter proteins depends on the expression of the *RAD* gene. If the *RAD* gene is expressed, the DRIF proteins are present in the cytoplasm and form a complex with the RAD protein. If *RAD* is not expressed, the DRIF proteins are localized in the nucleus and form a complex with the DIV protein [46].

1.2.3 The molecular network underlying the establishment of bilateral symmetry in *A. majus*

At the basis of the establishment of bilateral symmetry of *A. majus* there is an antagonistic mechanism due to the presence of the DRIF proteins, shown in Fig. 8 [27]. The *CYC* gene is expressed in all the dorsal parts of the flower, as well as the *RAD* gene, whereas *DICH* is expressed only in the more dorsal part [20, 27, 45]. The *DIV* and *DRIF1* genes are both expressed in all the whorls of the flower [43, 46].

In the dorsal part of the flower, the CYC and DICH proteins are able to induce the transcription of the *RAD* gene, binding its promoter and intron regions [27]. Once activated, the RAD protein goes in the cytoplasm of the cell and here binds the DRIF1 protein, a co-activator of DIV. In this way, in the dorsal part of the flower DIV is expressed but it is not active because DRIF1 is sequestered by RAD in the cytoplasm. In absence of the active DIV function, the proteins CYC, DICH and RAD can induce the dorsalization of the flower. On the contrary, in the ventral part of the flower *CYC* and *DICH* are not expressed. In this way, the RAD protein is not expressed and DRIF1 is able to move inside the nucleus, where it binds and activates the DIV protein that can induce the ventralization of the flower.

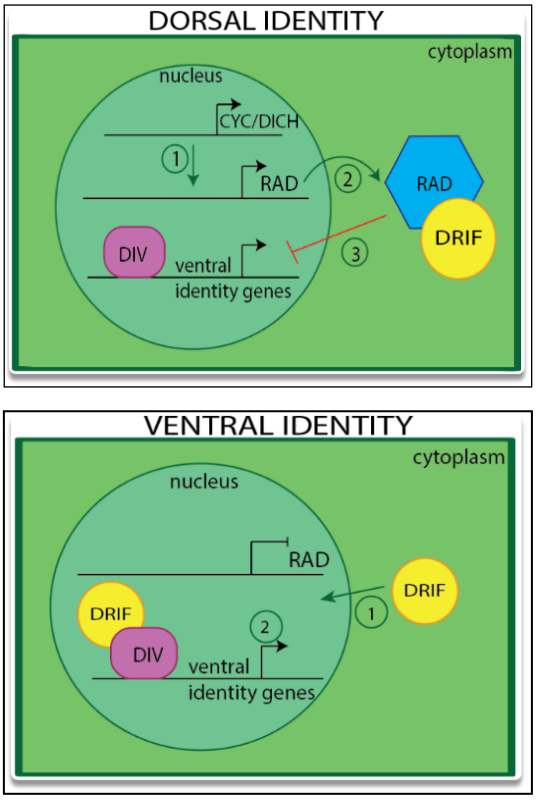

Fig. 8 Molecular network underlying the establishment of bilateral symmetry in the flower of *A. majus*.

1.3 Aim of the work

The aim of my PhD project was the study of the molecular basis of the orchid flower development, paying particular attention to the genes involved in floral symmetry of a non-model orchid species, the Mediterranean orchid *Orchis italica*. Starting from the knowledge of the molecular network of interactions described in the model plant *A. majus*, the goal of my PhD research project was to verify the existence of evolutionary conserved modules of interaction underlying the flower symmetry in distantly related flowering plants (*O. italica* and *A. majus*).

At the beginning of my PhD research project, the orchid TCP genes had just been described in *O. italica* [48] and subsequently in *Phalaenopsis equestris* [49]. Unfortunately, in *P. equestris* the role in flower symmetry of the three *CYC*-like genes identified was only marginally investigated, whereas in *O. italica* only a small, conserved fragment of the *TB1* (*CYC*-like) gene was reported. In addition, the orchid MYB genes *DIV*, *RAD* and *DRIF* had never been studied. For these reasons, I decided to undertake the analysis starting from the search of the *DIV*, *RAD* and *DRIF* genes expressed in the inflorescence of *O. italica*. More in details, the first part of my work can be divided in different points:

- Identification of the *DIV*-, *RAD*- and *DRIF*-like transcripts expressed in the available inflorescence transcriptome of *O. italica* [48] by *in silico* analysis;
- Analysis of the gene structure of the *DIV*- and *RAD*-like genes of *O. italica* and *P. equestris*;
- Phylogenetic and evolutionary analysis of the orchid *DIV*- and *RAD*-like genes;
- Expression analysis of the *DIV*, *RAD* and *DRIF1* genes of *O. italica* by real time PCR and *in situ* hybridization;
- Study of the protein interactions among DIV, RAD and DRIF1 of *O. italica* by yeast two-hybrid analysis.

2. Results and Discussion

2.1 Identification and structure of *DIV*- and *RAD*-like genes in orchids

- **Identification of *DIV*- and *RAD-like* transcripts**

The analysis of the inflorescence transcriptome of *O. italica* reveals the presence of 8 *DIV*- and 4 *RAD*-like transcripts.

The 8 *DIV*-like transcripts contain two conserved MYB domains. Their length ranges from 737 to 1901 bp and they encode for proteins ranging from 175 to 297 amino acids (Tab. 1 - Appendix). Two *DIV*-like transcripts, *OITA_23026* and *OITA_13252*, are subjected to alternative splicing (Fig. 9). In particular, the alternative transcript of *OITA_23026* includes a 111 bp fragment (whose ends are 5'-GT and AG-3') that generates a premature stop codon within the coding sequence (CDS), leading to the formation of a 117 amino acids protein with a single MYB domain. The two transcripts of *OITA_13252* differ in the 5'-UTR, 31 nucleotides upstream the translation start codon, for the presence/absence of a fragment of 88 bp (with 5'-GT and AG-3' ends) whose presence does not change the protein product.

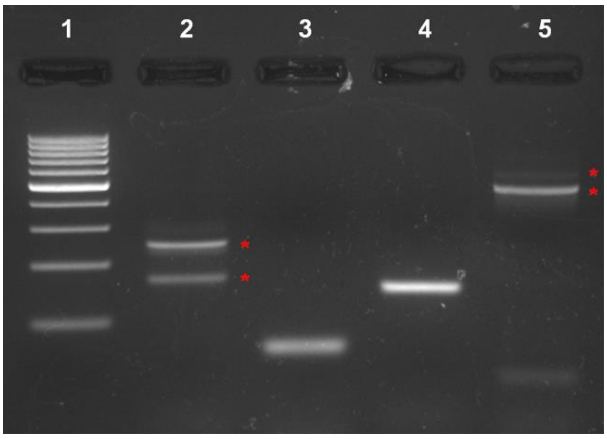

Fig. 9 Alternatively spliced isoforms of *DIV*-like transcripts of *O. italica*. Reprinted from *Valoroso et al. (2017)* [14]. The agarose gel electrophoresis shows different PCR amplified fragments of the alternative isoforms of two *DIV*-like transcripts. 1) GeneRuler 100 bp DNA ladder (Thermo Scientific); 2) the two different isoforms *OITA_13252* and *OITA_13252_AS*; 3) fragment of *OITA_13252*; 4) fragment of *OITA_23026_AS*; 5) the two different isoforms *OITA_23026* and *OITA_23026_AS.* The red asterisks indicate the alternatively spliced isoforms. Forward primers that specifically amplify the intron-retaining isoforms were designed in the region covering the exon–intron junction [14].

The *in silico* analysis of the transcriptome of the orchid *Ophrys sphegodes* revealed the presence of four *DIV*-like transcripts. Even though they are not full length, these transcripts were included in the further analyses because *Ophrys* belongs to the same subfamily of *O. italica* (Orchidoideae). It is possible that the low number of *DIV*-like transcripts found in *Ophrys* is due to the sequencing and assembly strategies used to obtain the transcriptome of this species [50]. In addition, the recently released orchid genomes of *P. equestris* [51] and *D. catenatum* [52] were analysed, showing the presence of 7 and 8 *DIV*-like transcripts, respectively (Tab. 1 - Appendix). This result shows that the number of *DIV*-like genes identified in the present study reflects the real number of *DIV*-like copies in orchids and that this number is similar to that found in other species (e.g., 8 *DIV*-like genes in *Dipsacales* [53]).

The four *O. italica RAD*-like transcripts present a single MYB domain, a length ranging from 543 to 420 bp (Tab. 1 - Appendix) and encode for proteins ranging from 82 to 94 amino acids. The *in silico* analysis of the *Ophrys* transcriptome showed the presence of a single *RAD*-like transcript, whereas the search conducted on the genomes of *P. equestris* and *D. catenatum* revealed the presence of 5 *RAD*-like genes (Tab. 1 - Appendix). As for the *DIV*-like genes, the number of *RAD*-like genes seems to be quite conserved within orchids and similar to that of other species (e.g. six genes in *A. majus* and *Arabidopsis thaliana* [54]). The difference relative to the lower number of *RAD*-like transcripts found in *O. italica* compared to the other orchid species could be due to the very low expression level of the fifth *RAD*-like gene, resulting in the absence of its transcript in the assembled transcriptome of *O. italica*. In alternative, it is possible that in *O. italica* there are only four *RAD*-like genes. As currently the assembled genome of *O. italica* is not available, it is impossible to discriminate between the two hypotheses.

- **Genome organization of the orchid *DIV*- and *RAD*-like genes**

The orchid *DIV*-like genes of *O. italica, P. equestris* and *D. catenatum* include a single intron (Fig. 10C). In *O. italica* intron position is conserved relative to *P. equestris* and *D. catenatum*, with canonical donor and acceptor splicing sites. Intron length of the different *DIV*-like genes of *O. italica* varies in a range from 76 to ~10,000 bp, as in *P. equestris* and *D. catenatum* (Tab. 1 - Appendix). Due to their length, introns of the genes *OITA_13233* (~5,000 bp), *OITA_8681* (~6,000 bp) and *OITA_12910* (~10,000 bp) were only partially sequenced (Tab. 1 - Appendix).

Three out of the four *RAD-like* genes of *O. italica* have a single intron (Fig. 10C). In addition, a premature stop codon in *OITA_32153* shifts the intron position to the 3' UTR. Among the examined orchids, this feature, where intron is located within the 3'UTR, with a premature stop codon, is peculiar of *O. italica.* Although it seems to be an intermediate condition between the intron presence and absence, the orthologs of *OITA_32153* in *P. equestris* and *D. catenatum* present a canonical intron within the coding region, suggesting that this intron shift is a derived character of *O. italica*. The presence of *RAD*-like genes with a single intron within the CDS, within the 3'UTR or intronless has been described also in other species, as *A. majus* and *A. thaliana* [54]. These evidences support the possible evolution of the RAD-like subfamily from an ancestral gene organization with two exons to a more recent structure without introns. However, it cannot be excluded the independent origin of these two gene structures in the different lineages.

The analysis of the three orchid genomes currently released (*P. equestris*, *D. catenatum* and the basal species *Apostasia shenzenica*) has shown that they are very rich in transposable elements, generally localized within the large intron sequences [51, 55]. CENSOR analysis revealed traces of transposable/repetitive elements within the introns of the *DIV*-like genes *OITA_13252* and *OITA_13233*, of their orthologs of *P. equestris* and *D. catenatum* and in all introns of the *RAD*-like genes. The large number of repetitive/transposable elements within the orchid genomes might have a functional significance in the regulation of gene expression. It is possible that the presence of these elements can drive the antisense transcription and/or promote heterochromatin formation, reducing the transcriptional levels [56].

2.2 Phylogenetic and evolutionary analysis of the orchid *DIV-* and *RAD-*like genes

- **Phylogenetic analysis**

The Maximum Likelihood (ML) tree shown in Fig. 10A is obtained from the DIV- and RAD-like amino acid alignment of the orchid sequences identified by *in silico* analysis. Both the DIV-like and RAD-like groups present a high bootstrap support value (87%).

Within the DIV-like lineage, the ML tree shows the presence of three subgroups (1-3), all including sequences of *O. italica, P. equestris* and *D. catenatum*, whereas the sequences of *Ophrys* are included only in the subgroups 1 and 2. Two well-supported subgroups are detected within the RAD-like group: the largest subgroup includes the sequences of all the

species considered in this study, the smaller one includes only two sequences, one of *P. equestris* and one of *D. catenatum.*

The presence in almost all the subgroups of the ML tree of the DIV- and RAD-like sequences of all the orchid species considered in this study demonstrates that the duplication that has originated the different *DIV*- and *RAD*-like genes occurred before the divergence of Orchidoideae and Epidendroideae (~64 Mya). In addition, the ML tree topology is in agreement with the results previously reported in *Dipsacales* [53], suggesting that the origin of these ortholog groups predates the monocot/dicot divergence (140-150 Mya [2]).

The analysis of the conserved domains (Fig. 10B) identified nine amino acid motifs whose distribution reflects the partition of the orchid *DIV*- and *RAD*-like genes into three and two subgroups, respectively. The motifs 1, 2 and 3 are part of the MYB DNA binding domain, while the others have unknown function.

Evolutionary analysis was conducted on the coding region of the orchid *DIV*- and *RAD*-like genes. The ratio between the mean nonsynonymous and synonymous substitution rates (ω) shows that on these genes is acting purifying selection ($\omega<1$). Different evolutionary models were compared to verify the existence of variations of the ω ratios among the different branches of the ML tree of the *DIV*- and *RAD*-like genes. The results obtained from evolutionary analysis are summarized in Tab. 2-5 (Appendix).

The one-ratio model assumes an equal ω for all the branches of the tree, whereas the two- and three-ratio models consider two and three different ω values, respectively. Within the *DIV*-like genes, the three-ratio model fits the data better than the two-ratio models, whereas within the *RAD*-like genes the two-ratio model is statistically more supported than the one-ratio model.

The clade model infers different selective pressures on a proportion of sites within a specific branch of the tree. The clade model 2 is statistically more supported than its null model, showing that ~34% of sites of the DIV-like subgroup 2 has a different ω (0.01754) from that of the other subgroups (0.03487). On the contrary, within the RAD-like group, the presence of sites with different selective pressures is not statistically supported.

The sites and branch-sites models assume positive selection in specific sites and branches of the tree. Signals of positive selection are not detected either within the *DIV*- or *RAD*-like coding sequences.

Although evolutionary analysis highlights strong purifying selection acting on the orchid *DIV*- and *RAD*-like coding sequences, the three ortholog groups of the *DIV*-like genes have different evolutionary rates, whereas the selective constraints of the *RAD*-like genes are more uniform.

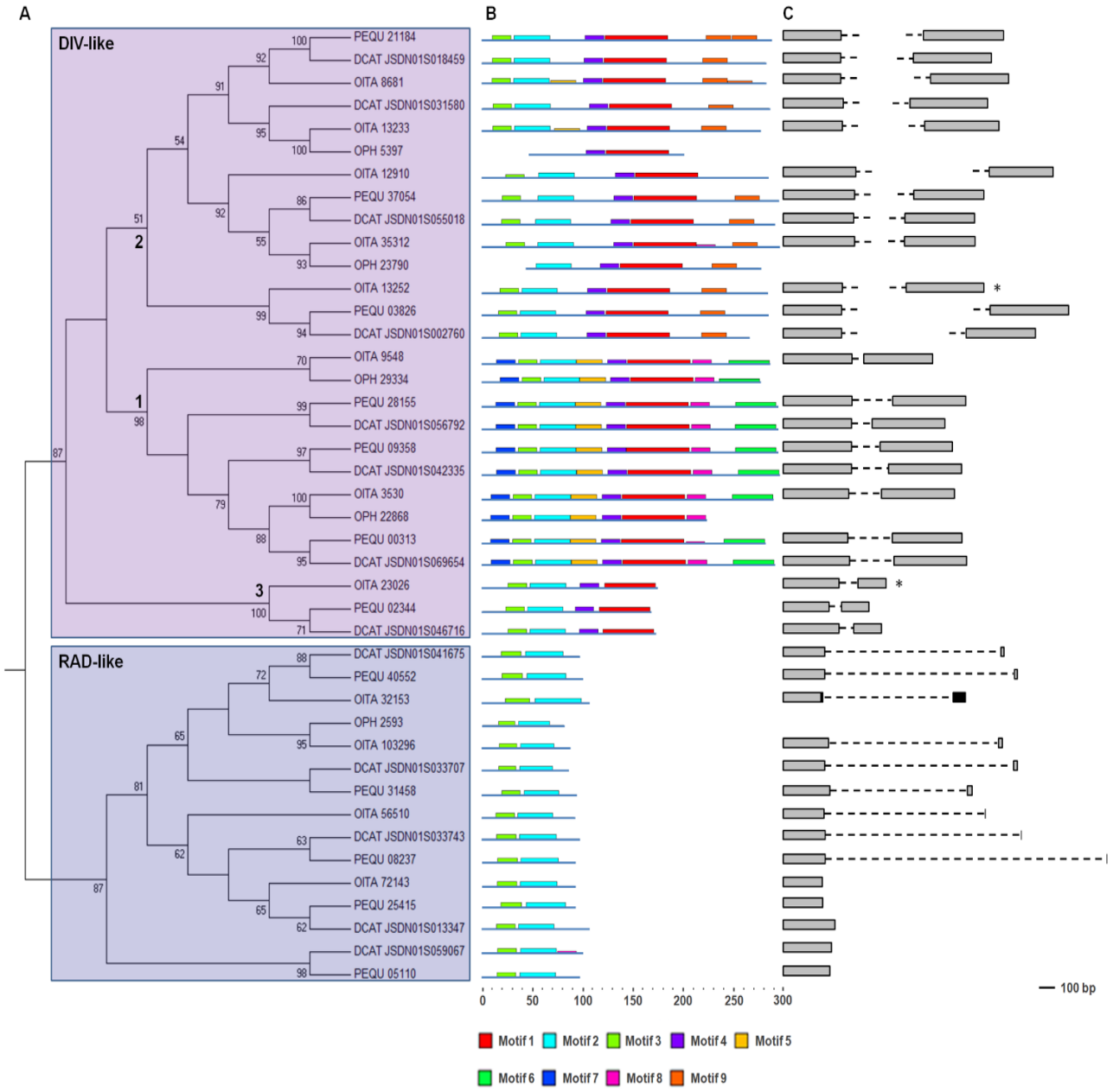

Fig. 10 Phylogeny and genomic organization of the orchid *DIV*- and *RAD*-like genes. Reprinted from *Valoroso et al. (2017)* [14]. **(A)** ML phylogenetic tree; **(B)** diagram of the conserved domains; **(C)** genomic organization and intron size/position.

2.3 Identification of the DRIF-like genes of *O. italica*

Within the inflorescence transcriptome of *O. italica* there are two transcripts, *OITA_10599* and *OITA_6376*, encoding for DRIF-like proteins. The transcript *OITA_10599* (1,250 bp) encodes for a protein product of 258 amino acids that shows 56% identity with the DRIF1 protein of *A. majus.* The transcript *OITA_6376* (2,083 bp) encodes for a protein product of 305 amino acids showing 39% identity with the DRIF2 protein of *A. majus.* The BLASTP analysis of both proteins reveals a higher identity score (65% for OITA_10599 and 56% for OITA_6376) with uncharacterized proteins; however, due to the poorly conserved sequences of the DRIF proteins, we have assumed *OITA_10599* as the putative ortholog of *DRIF1* of *A. majus*. A recent study has revealed the presence of a larger number of *DRIF*-like genes in dicots and monocots. For example, there are 6 *DRIF*-like genes in *A. majus* and in *Oryza sativa* [57]. Currently, *in silico* genome and transcriptome analyses have been undertaken to verify the copy number of the *DRIF*-like genes in orchids.

2.4 Expression analysis of the *DIV*-, *RAD*- and *DRIF*-like genes of *O. italica*

O. italica is a wild species very difficult to propagate *in vitro*. In addition, it is not possible to perform functional studies based on knock-out techniques. For these reasons, the analysis of gene expression is an important tool to infer gene function in this non-model species. In *O. italica*, the orthologs of the *DIV*, *RAD* and *DRIF1* genes of *A. majus* are *OITA_9548*, *OITA_56510* and *OITA_10599*, respectively. Figure 11 shows the RNA *in situ* hybridization of these transcripts in the early floral tissues of *O. italica*. *OITA_9548* (*DIV*) and *OITA_10599* (*DRIF1*) are localized in all the floral tissues, whereas *OITA_56510* (*RAD*) is expressed in the lip and outer tepals. This pattern is similar to that described in *A. majus*, where *DIV* and *DRIF1* are expressed in all the parts and *RAD* is expressed in the dorsal domain of the snapdragon flower [43, 44]. In analogy with *Antirrhinum*, the observed expression pattern suggests that in the lip of *O. italica* OITA_10599 (DRIF1) could bind OITA_56510 (RAD), thus inhibiting the formation of the DIV-DRIF complex and the subsequent activation of ventralization genes. In *O. italica*, RAD and DRIF1 could also interact in outer tepals (where both are expressed), thus inhibiting ventralization as in the lip. The expression of *RAD* (expressed in the dorsal domains of the snapdragon) in the ventral part of the flower of *O. italica* is due to resupination, the 180° twist of the median inner tepal (lip - dorsal structure) that becomes a ventral structure (Fig. 1). The expression pattern of *DIV*, *RAD* and *DRIF1* in *O. italica* is maintained in the perianth tissues after anthesis, as revealed by real time PCR

experiments (Fig. 12-14). The expression data suggest a model of interaction among DIV, RAD and DRIF1 to determine the bilateral symmetry of the flower generally conserved between *Antirrhinum* and *O. italica*, with some differences due to the peculiarities of the orchid flower. In particular, this model should be integrated with the expression pattern of some MADS-box genes (e.g. *AP3/DEF* and *AGL6*) involved in the formation of the orchid lip [58-60]. The role of the MADS-box genes in the formation of the orchid perianth will be analysed in the Chapter 2.

After anthesis, *OITA_10599* (*DRIF1*) is expressed in all floral tissues and has a significantly higher expression in the lip, whereas the *OITA_6376* (*DRIF2*) is expressed at similar levels in all tissues except the ovary and leaf, where it is very weakly expressed (Fig. 14).

In the mature flower, some *DIV*-like transcripts of *O. italica* show overlapping profiles and others have very different expression patterns (Fig. 12). In addition to its expression in the perianth, *OITA_9548* is detectable in the column, ovary and leaf. Although at lower levels, also *OITA_3530* is expressed in all floral organs, as well as *OITA_8681*, *OITA_35312* and *OITA_13233*, whereas *OITA_12910* is mostly expressed in the column. The two isoforms of OITA_23026 differ in their level of expression, significantly lower in all tissues (excluding leaf) for *OITA_23026_AS*, the isoform that retains the intron. Also the two isoforms of *OITA_13252* show different expression profiles. In this case *OITA_13252*_AS, the isoform that retains the intron in the 5' UTR, is expressed at higher levels in all tissues (excluding leaf) than *OITA_13252* (Fig 12). The expression levels of the *DIV*-like genes in *O. italica* seem to be inversely correlated with their intron size, in line with previous results obtained from genome analysis and gene expression of other orchid species [51]. This feature could be due to the presence of repetitive/transposable elements within the large introns of the *DIV*-like genes that can negatively regulate their expression levels [56].

Although with different levels, the expression patterns of the *RAD*-like transcripts in the perianth tissues after anthesis are similar (Fig 13), being detectable mainly in the outer tepals and lip and with lower expression in the ovary. *OITA_32153* is also highly expressed in the column and leaf, where also *OITA_56510* is expressed. The similar expression pattern of the *RAD*-like genes of *O. italica* suggests their redundant function in the tissues of the perianth and more specialized roles in reproductive and vegetative tissues.

All the *DIV*-like genes of *O. italica* are mainly expressed in outer and inner tepals and lip, suggesting a possible redundant function in these tissues. However, a possible pleiotropic role of some transcripts, as previously reported in other species [53], is supported by their expression also in the column, ovary and leaves. The expression profile of the two

OITA_13252 isoforms are similar, even though *OITA_13252_AS*, that retains the intron in the 5' UTR, is expressed at higher levels. This difference could be due to a transcriptional regulatory role of the small intron in the 5' UTR, whose presence might increase the transcription efficiency. Interestingly, the expression pattern of the *DIV*-like isoform *OITA_23026_AS* (that encodes for a protein with a single MYB domain) is similar to that of the *RAD*-like gene *OITA_103296*, leading to hypothesize that this isoform might be an evolutionary intermediate step towards the *RAD*-like genes, possibly evolved from a *DIV*-like gene after the loss of exon 2 through alternative splicing.

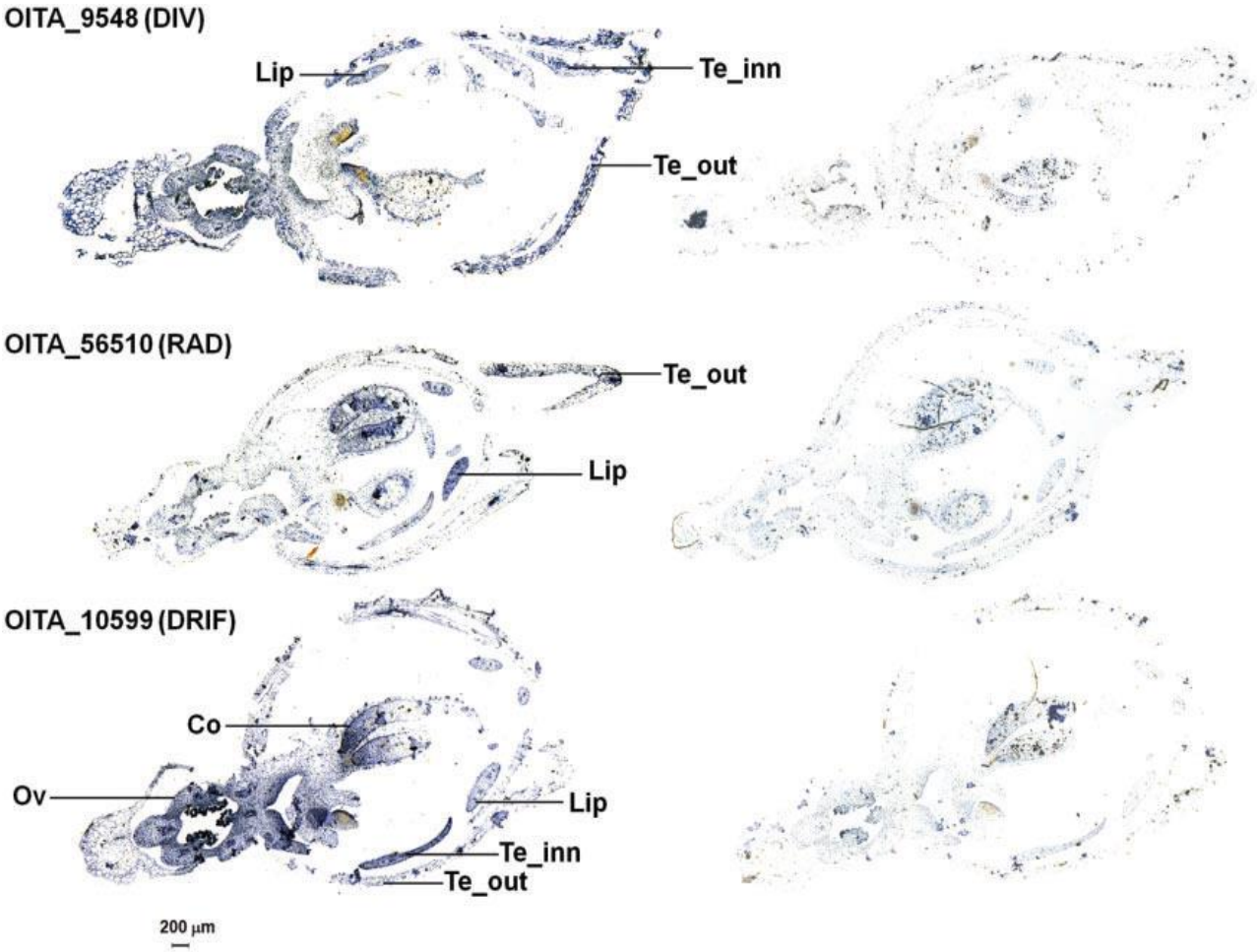

Fig. 11 RNA *in situ* hybridization of the *DIV-, RAD-* and *DRIF1* transcripts on early floral tissues of *O. italica*. Reprinted from *Valoroso et al. (2017)* [14]. RNA *in situ* hybridization with the antisense (left) and sense (right) probes of the transcripts *OITA_9548* (*DIV*), *OITA_56510* (*RAD*) and *OITA_10599* (*DRIF1*). Te_out, outer tepal; Te_inn, inner tepal; Co, column; Ov, ovary.

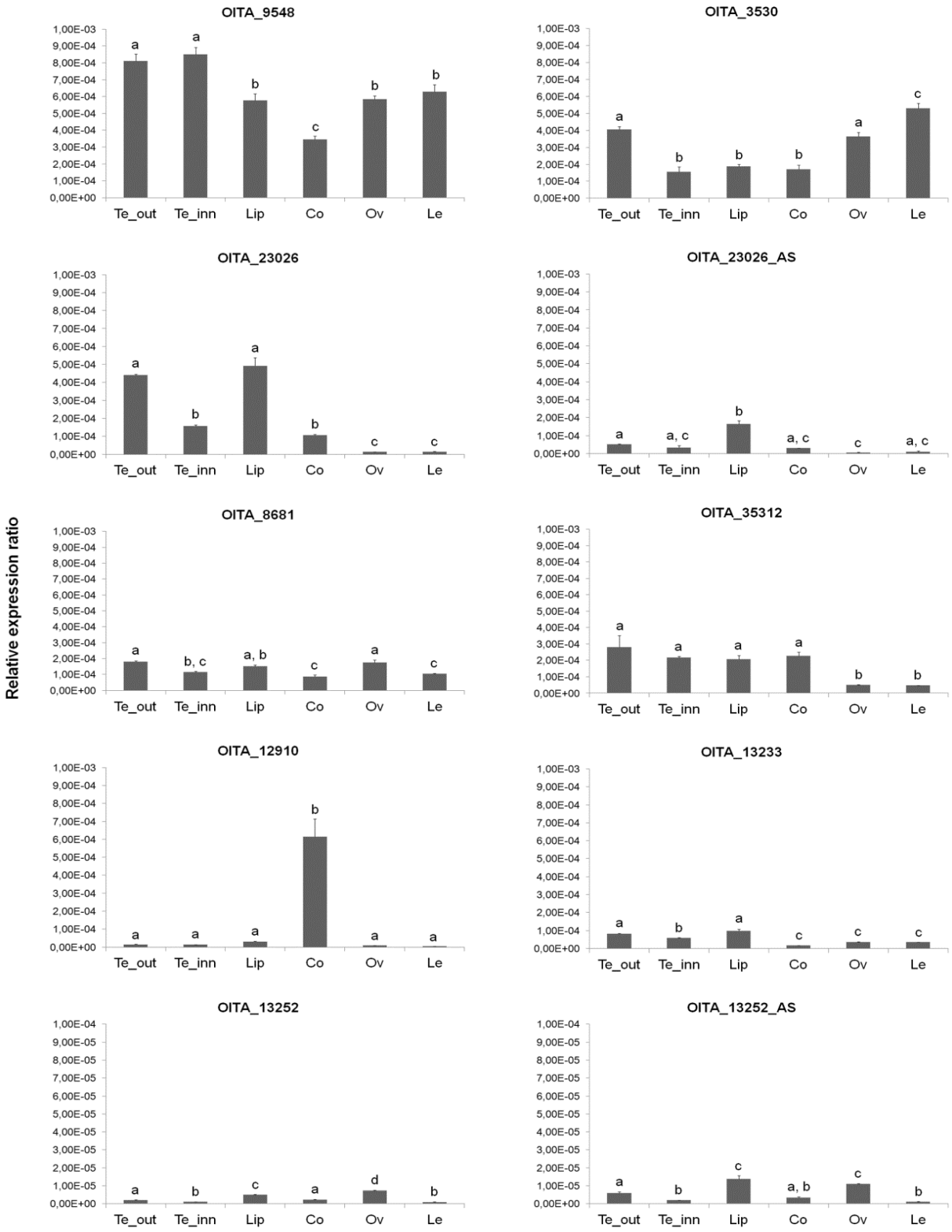

Fig. 12 Expression pattern of the *DIV*-like genes of *O. italica.* Reprinted from *Valoroso et al. (2017)* [14]. Relative expression pattern of the *DIV*-like transcripts in the different tissues of *O. italica.* Te_out, outer tepal; Te_inn, inner tepal; Co, column; Le, leaf. The letters above the bars indicate statistically significant groups, as assessed by the Tukey HSD post-hoc test.

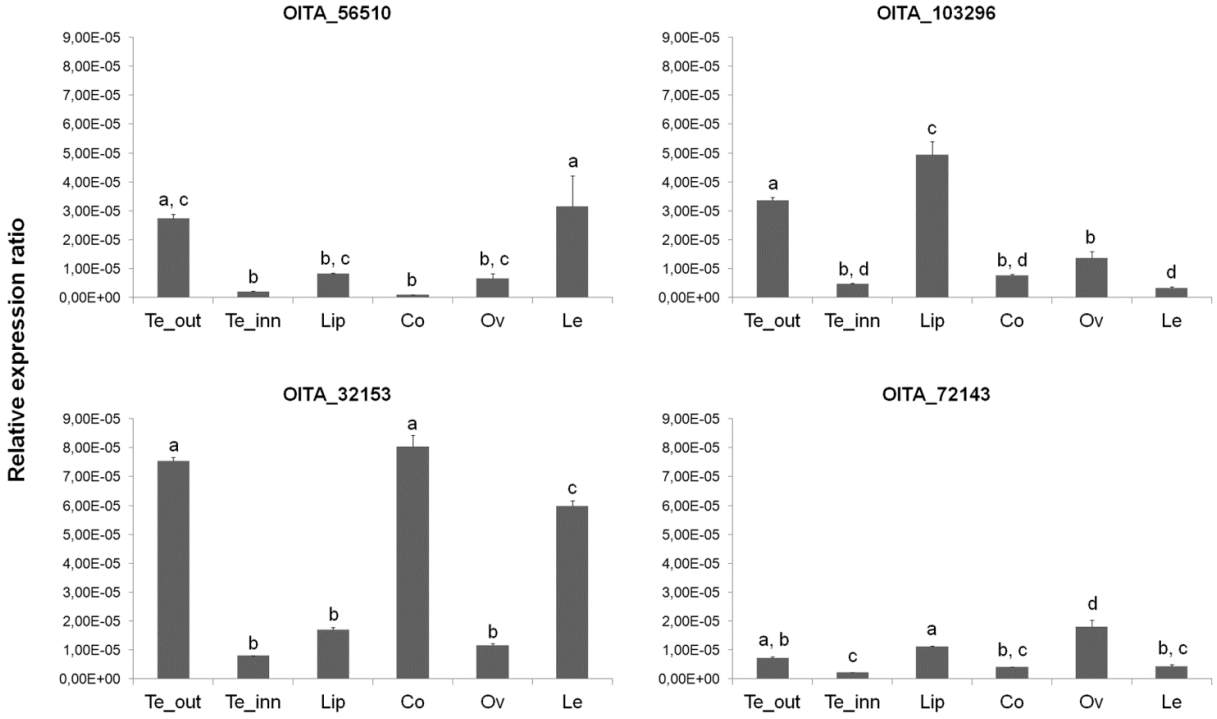

Fig. 13 Expression pattern of the *RAD*-like genes of *O. italica.* Reprinted from *Valoroso et al. (2017)* [14]. Relative expression pattern of the *RAD*-like transcripts in the different tissues of *O. italica.* Te_out, outer tepal; Te_inn, inner tepal; Co, column; Le, leaf. The letters above the bars indicate statistically significant groups, as assessed by the Tukey HSD post-hoc test.

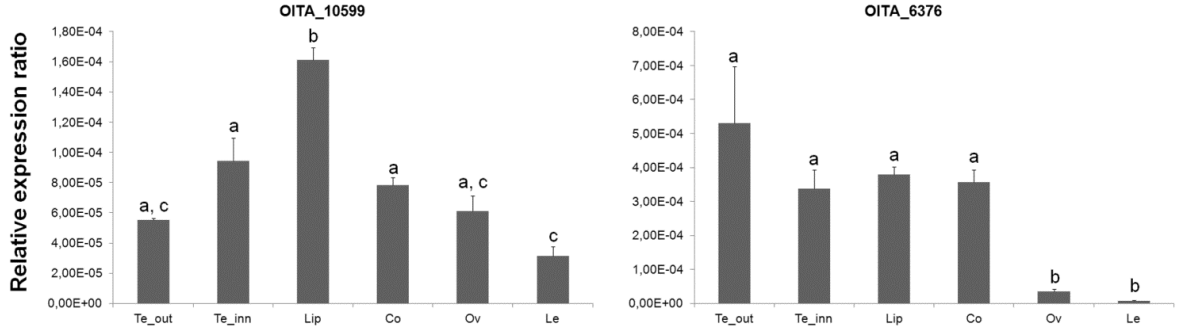

Fig. 14 Expression pattern of the *DRIF1/2*-like genes of *O. italica.* Reprinted from *Valoroso et al. (2017)* [14]. Relative expression pattern of the *DIV*-like transcripts in the different tissues of *O. italica.* Te_out, outer tepal; Te_inn, inner tepal; Co, column; Le, leaf. The letters above the bars indicate statistically significant groups, as assessed by the Tukey HSD post-hoc test.

The results discussed in the previous paragraphs (Chapter 1, from 2.1 to 2.4) were published in:

Valoroso MC, De Paolo S, Iazzetti G, Aceto S (2017). Transcriptome-Wide Identification and Expression Analysis of *DIVARICATA*- and *RADIALIS*-Like Genes of the Mediterranean Orchid *Orchis italica*. *Genome Biology and Evolution*, Volume 9, Issue 6, 1 June 2017, Pages 1418–1431, https://doi.org/10.1093/gbe/evx101.

2.5 Protein interaction among DIV, RAD and DRIF1 in *O. italica*

In order to confirm if the model of interaction of DIV, RAD and DRIF1 of *A. majus* is conserved in *O. italica*, a "yeast two hybrid system assay" (Y2H) was conducted in collaboration with Prof. Maria Manuela Ribeiro Costa, at the Plant Functional Biology Center, University of Minho (Braga, Portugal).

The open reading frame (ORF) of the *DIV (OITA_9548), RAD (OITA_56510)* and *DRIF1 (OITA_10599)* genes were cloned into two vectors containing the GAL4 DNA-binding or activation domain, respectively. The interactions were assayed in both the protein fusion forms. When the protein (in this case DIV) is able to promote the reporter gene transcription, it was assayed only the protein form fused to GAL4 activation domain.

The results of the Y2H screening and the quantitative analysis of the interaction levels are shown in Fig. 15. In *O. italica*, DRIF1 establishes a strong interaction both with the DIV and RAD proteins, whereas the RAD and DIV proteins do not interact each other. The ability of DRIF to interact with DIV and RAD has been demonstrated in *Antirrhinum* and tomato. This interaction has a key role in the establishment of different developmental pathways (snapdragon flower symmetry and tomato fruit development) [46, 47]. Our Y2H results demonstrate that these protein interactions are conserved also within monocot species, as orchids, where their role in the establishment of flower symmetry is supported also by the *DIV*, *RAD* and *DRIF* expression pattern.

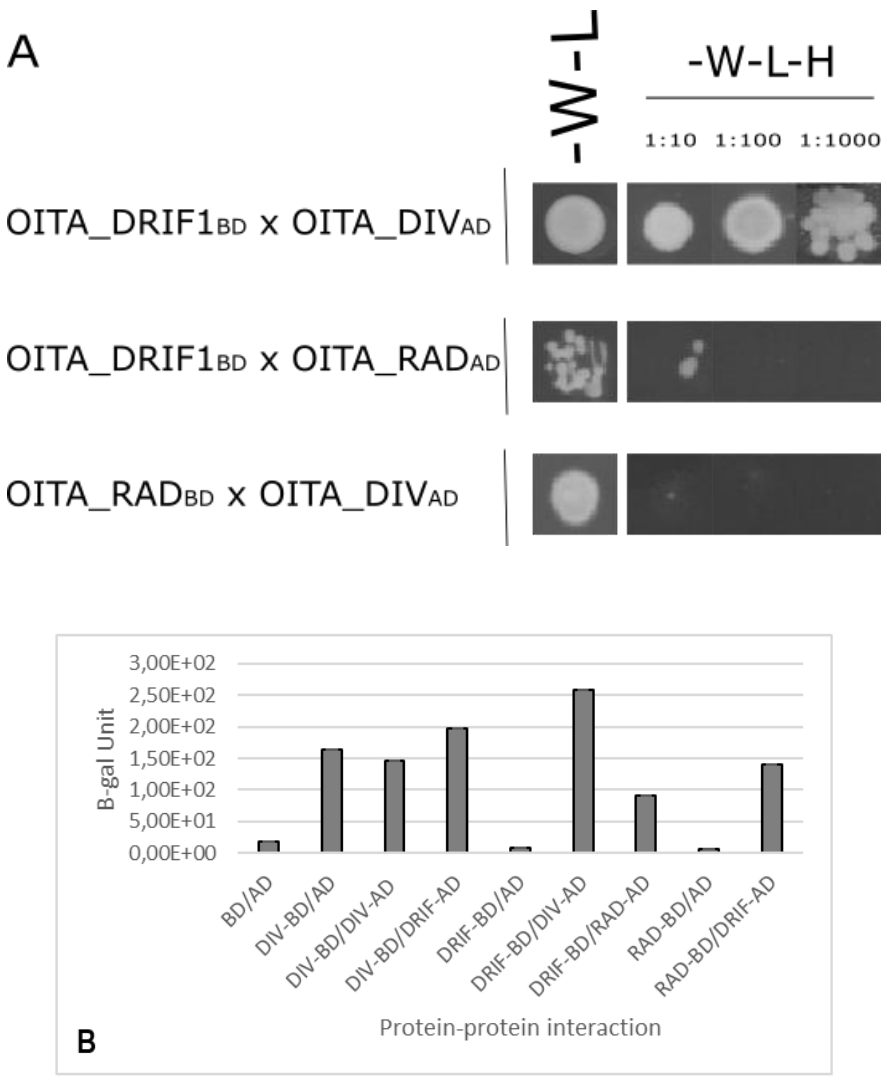

Fig. 15 The DIV, RAD and DRIF1 protein-protein interactions in *O. italica*. **(A)** Protein interaction between DRIF1$_{BD}$ - DIV$_{AD}$, DRIF1 $_{BD}$ - RAD$_{AD}$, RAD $_{BD}$ - DIV$_{AD}$. (-W-L, medium lacking the tryptophan and leucine; -W-L-H, medium lacking tryptophan, leucine and histidine. The dilution factor applied for the yeast inoculate is indicated with 1:10, 1:100 and 1:1000). **(B)** Quantitative analysis of the interaction levels by β-galactosidase assay. BD, binding domain; AD, activation domain.

CHAPTER 2 – The MADS-box genes

1. Introduction

1.1 The MADS-box genes: from flower development to bilateral symmetry

A crucial role in the evolution of flower architecture and in the control of flowering time is played by the MADS-box genes family. The MADS-box genes, known also as plant homeotic genes, are found in almost all eukaryotes and encode for transcription factors that contain a conserved DNA-binding domain called MADS domain [61]. This family takes its name from the first four MADS-box loci found in different species: ***M****INICHROMOSOME MAINTENANCE 1 (MCMI)* of *Saccharomyces cerevisiae*, ***A****GAMOUS (AG)* of *Arabidopsis thaliana*, ***D****EFICIENS (DEF)* of *Antirrhinum majus* and ***S****ERUM RESPONSE FACTOR (SRF)* of *Homo sapiens* [62].

The MADS-box transcription factors originated from the topoisomerase II subunit A [6, 63, 64] and are divided into two lineages, type-I and type-II, that differ in genomic organization, developmental role, evolutionary rate and level of functional redundancy. The type-I MADS-box proteins (Fig. 16A) contain the MADS domain and are divided in M-alpha (Mα), M-beta (Mβ) and M-gamma (Mγ) due to the sequence divergence of their C-terminus [65]. They are involved in the development of seed, embryo and female gametophyte [66]. The type-II is the most studied class of MADS-box genes due to its involvement in different plant developmental processes, including flower formation. These transcription factors are characterized by three conserved domains and a variable domain that form the so-called MIKC structure (Fig. 16B-C) [6]:

- The MADS-box domain (**M**) at the N-terminus;
- The intervening domain (**I**);
- The keratin domain (**K**);
- The variable **C**-terminus.

The MADS-box is the DNA-binding domain that forms an α-helix structure capable, together with a partner protein, to recognize the CArG-box motif ($CC[A/T]_6GG$) in the promoter of the target genes [67-69]; the I and K domains are involved in the protein-protein interaction and in the formation of protein complexes [70, 71]; the variable C-domain has a role in the formation of protein complexes functioning as trans-activator domains [62, 71].

A duplication event involving the 5' region of the exon encoding for the K domain, followed by neofunctionalization, gave rise to the two classes of MIKC-type MADS-box genes: MIKC^C and MIKC* [71, 72]. The MIKC* genes (sometimes called Mδ) are involved in the male gametophyte development [73], while the MIKC^C, the most studied, play different roles in various processes of plant growth and in the establishment and maintenance of floral organs [6].

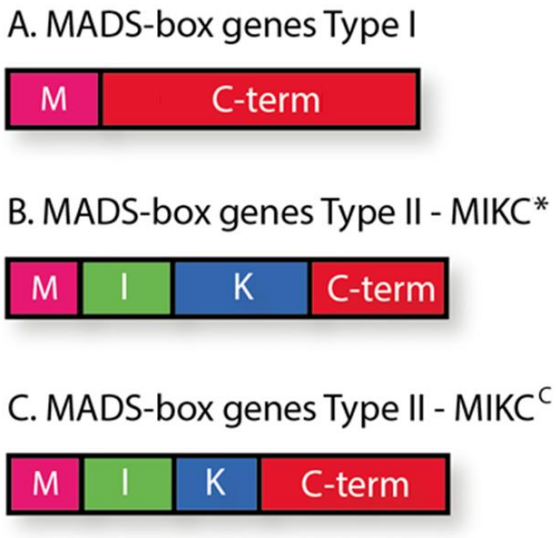

Fig. 16 The MADS-box protein structure. (A) Structure of a type-I **(B)** type II MIKC* and **(C)** type II MIKC^C protein.

The MIKC^C MADS-box genes involved in flower formation are divided in five functional classes (from A to E) and their activity is described by the ABCDE model of flower development [74, 75].

1.2 The ABCDE model of flower development

The first molecular model proposed to explain the identity and development of floral organs of the model plant *A. thaliana* dates back to 1991: the ABC model of flower development [76]. Since then, the model was improved, modified and extended to many plant species and currently it is known as the ABCDE model [75, 77-80]. The identity of floral organs depends on the expression and interaction of floral homeotic genes described by the ABCDE model [81, 82].

The wild-type flower of *A. thaliana* is composed by four concentric whorls where the floral organs develop: sepals in the outermost whorl 1, petals in the whorl 2, stamens in the whorl

3 and carpels in the innermost whorl 4 [83]. Excluding *APETALA2* (*AP2*), all the genes involved in the ABCDE model belong to the MADS-box family. According to this model (Fig. 16A), in *A. thaliana,* the A-class genes *APETALA1 (AP1)* and *AP2* specify the sepal identity. Together, the A- and B-class genes *APETALA3 (AP3)* and *PISTILLATA (PI)* specify the petal identity. The stamens in the third whorl are specified by the combination of the B-class genes and *AGAMOUS (AG)* of class C. The C-class alone determines formation of the carpels in the fourth whorl. The D-class genes *SEEDSTICK* (*STK*) and *SHATTERPROOF1/2* (*SHP1/2*) are involved in the determination of ovule and carpel identity. Finally, the E-class genes *SEPALLATA 1-4* (*SEP1-4*) are expressed in all the whorls and act redundantly to specify all the floral organs together with the other gene classes.

The MADS-box genes involved in the ABCDE model of flower development work forming homo- and heterodimers that regulate specific expression programmes in the different floral whorls [62]. This model of interactions is called **floral quartet.** According to this model, the complexes AP1/AP1/SEP/SEP, AP1/SEP/AP3/PI, AG/SEP/AP3/PI and AG/AG/SEP/SEP regulate the formation of the floral structures within whorls 1, 2, 3 and 4, respectively [75, 84].

Although very useful, the plant models *A. thaliana* and *A. majus* are highly derived species, whose developmental pathways could be very different from those of other angiosperm species. The “classic” ABCDE floral quartet model is applicable to different plant species (mainly eudicots) where it is well conserved [85-88]; however, analyses of non-model species (mainly monocots, as orchids) have revealed differences probably due to the different structure of the flower [89]. For example, in orchids the class B MADS-box genes show an expression profile expanded to the first floral whorl, explaining the presence of petaloid sepals (Fig. 16B) [90-93].

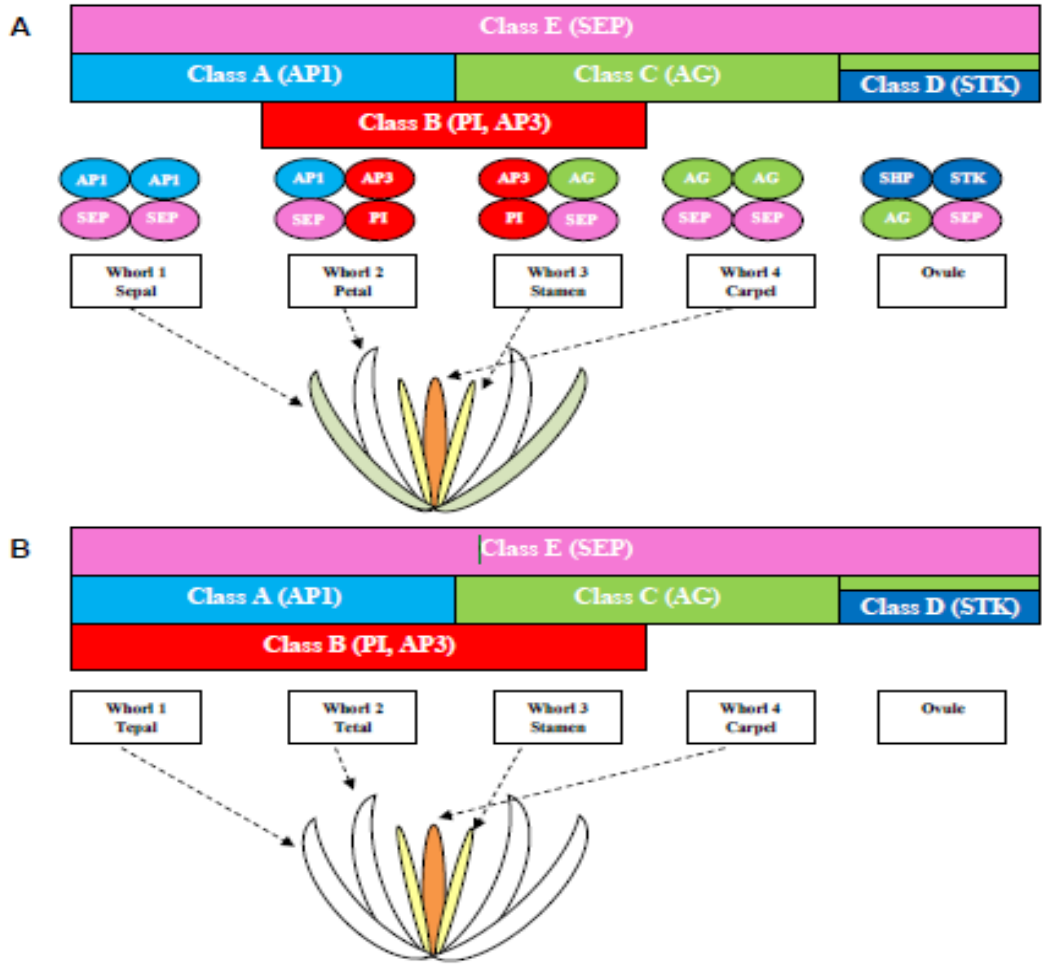

Fig. 16 The ABCDE model of flower development. Reprinted from *Aceto et al. (2011)* [6]. **(A)** The ABCDE model and floral quartet in *A. thaliana*; **(B)** modified ABCDE model in orchids.

Current advances in transcriptome and genome sequencing are highlighting the molecular programs that underly floral evolution and development of non-model species. For many species belonging to basal angiosperms, magnoliids and basal eudicots, the “fading borders model” proposes a gradient of the expression levels of floral homeotic genes to explain the flower morphology. In particular, low expression levels at the boundary of the domain of a given gene belonging to the class A, B and C overlap with the expression of a different homeotic gene in the adjacent domain (Fig. 17).

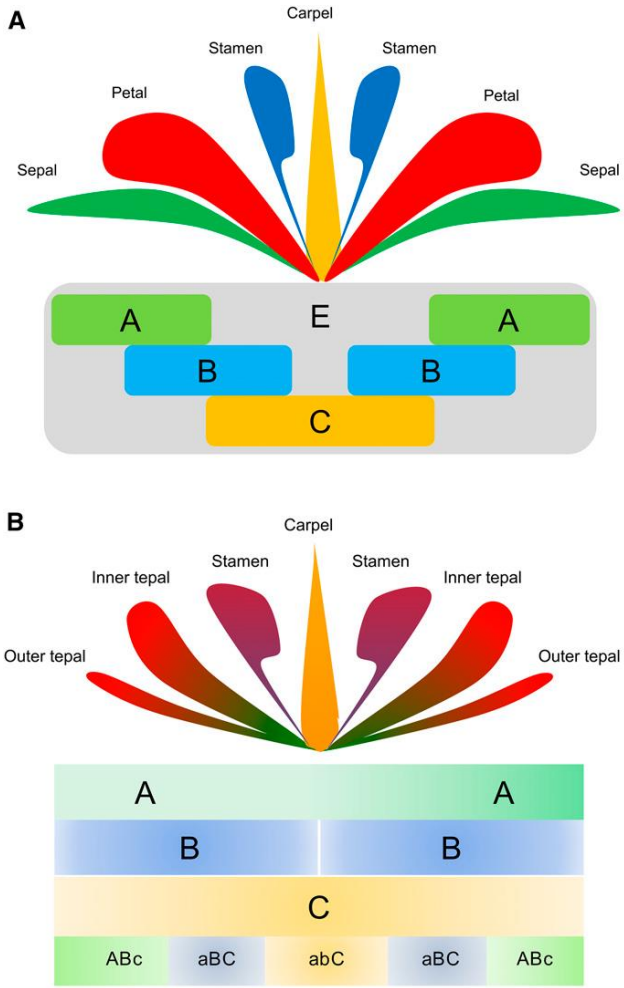

Fig. 17 The differences between the ABCE model and the fading borders model. Reprinted from *Chanderbali et al. (2015*) [94]. **(A)** Classic and **(B)** fading borders model of flower development.

- **The orchid code and the Homeotic Orchid Tepal (HOT) model**

In orchids, the MADS-box involved in the formation of the perianth are well characterized, in particular the B-class MADS-box genes [6, 59, 91-93, 95-98]. This class of genes originated after the duplication from an ancestral gene containing a paleoAP3 motif, giving rise to two lineages: the *AP3/DEF* lineage (from *APETALA3* of *A. thaliana* and *DEFICIENS* of *A. majus*) and the *PI/GLO* lineage (from *PISTILLATA* of *A. thaliana* and *GLOBOSA* of *A. majus*) [99, 100]. Two subsequent duplication events within the orchid AP3/*DEF* genes have played an important role in the evolutionary origin of the current structure of the orchid flower explained by the so-called "**orchid code**" theory (Fig. 18) [101, 102]. The orchid code theory proposes that the identity of orchid perianth is established through the formation of protein complexes among the four different orchid AP3/DEF proteins and one PI/GLO-like protein [6]. The orchid *AP3/DEF* genes are divided in 4 clades with different expression patterns in the

organs of the perianth (Fig. 18). Genes of clade 1 and 2 are expressed in all tepals and drive the formation of the outer tepals; high levels of clade 1 and 2 proteins and low levels of clade 3 and 4 proteins determine the formation of inner lateral tepals; high levels of clade 3 and 4 proteins and low levels of clade 1 and 3 proteins induce the formation of the lip (Fig. 18) [103-105].The orchid code theory proposes an evolutionary reconstruction of the perianth origin from an ancestral orchid flower with radial symmetry to the current flower with bilateral symmetry. The ancestral orchid flower had undifferentiated tepals; after the first duplication of the *AP3*/*DEF* gene, two gene clades were formed: 1/2 and 3/4, with the origin of a perianth composed of outer and inner tepals; after the second duplication, the four gene clades gave origin to the zygomorphic orchid flower with outer and inner tepals and a differentiated lip [102].

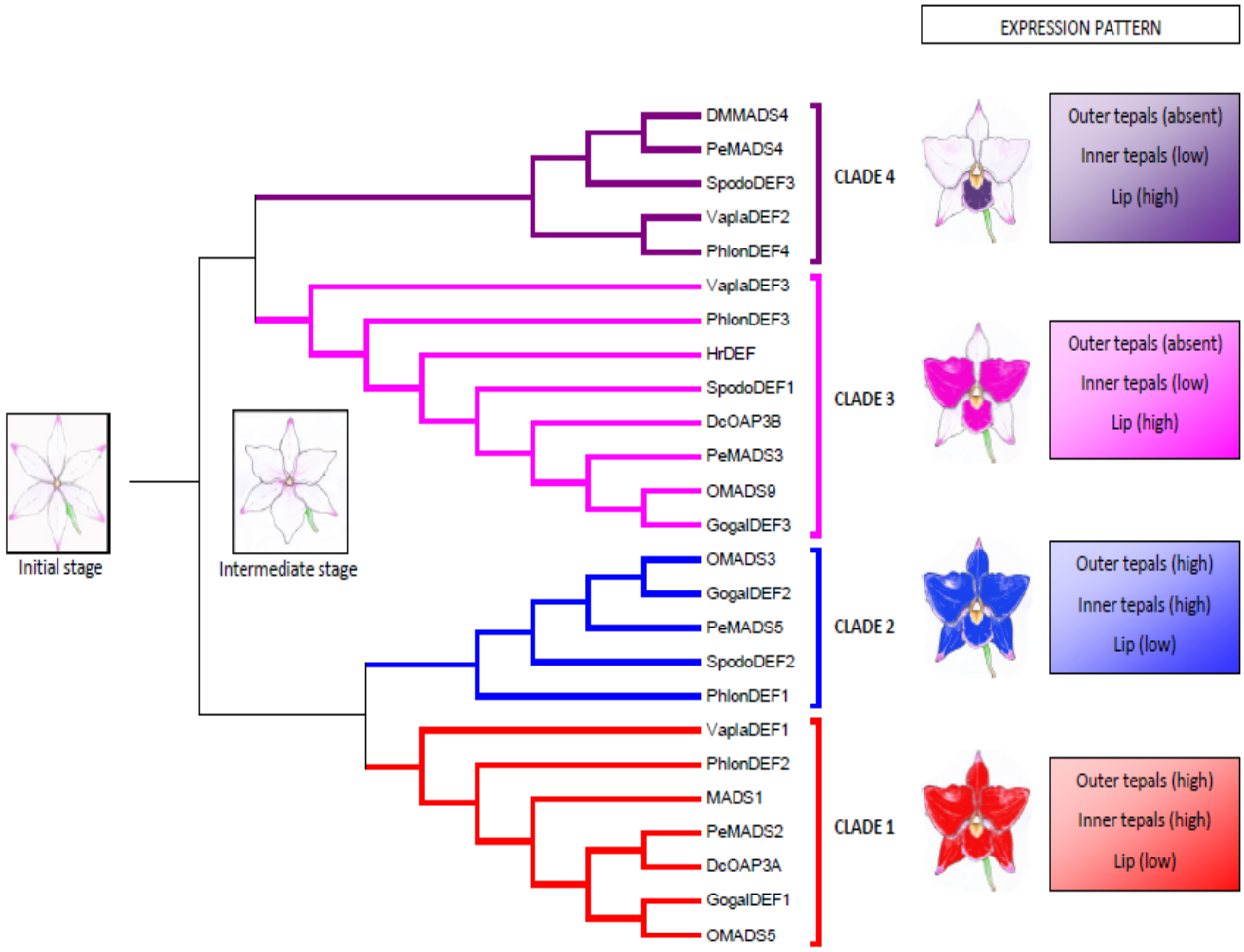

Fig. 18 Duplications of the orchid B-class MADS-box genes, their expression and evolution of flower symmetry. Reprinted from *Aceto et al. (2011)* [6].

The regulation of perianth morphogenesis in orchids is explained more in details by the recently proposed "**Homeotic Orchid Tepal (HOT)**" model (Fig. 19) [98]. In this model, the *PI/GLO* and *AP3/DEF* clades 1 and 2 determine the formation of the outer tepals, while the combination of *PI/GLO* and *AP3/DEF* clades 1-2-3 induces the formation of the lateral inner tepals. The *PI/GLO* and *AP3/DEF* clades 3 and 4, together with the MADS-box *AGL6*-like genes, contribute to the lip morphogenesis [98].

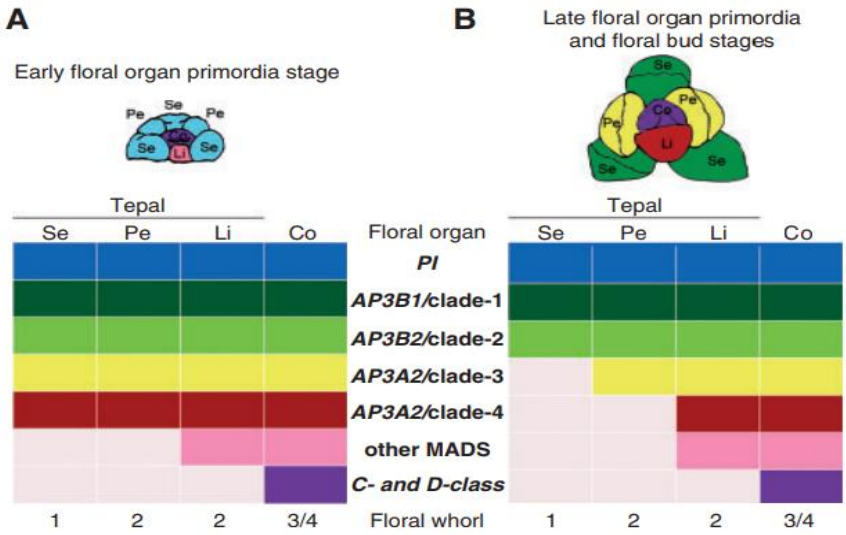

Fig. 19 The Homeotic Orchid Tepal (HOT) model. Reprinted from *Pan* et al. (2011). Expression of the MADS-box genes at early **(A)** and late **(B)** stage of orchid development.

1.3 Aim of work

When I started my PhD project, previous studies had analysed the structure, expression and evolution of some MADS-box genes of *O. italica*, in particular the B-class genes *PI/GLO* and *AP3/DEF* and the C- and D-class genes *AG* and *STK* [6, 59, 91, 92, 95, 106-108].

As evidences were accumulating supporting an involvement of the MADS-box genes in the establishment of bilateral symmetry during the development of orchid flower (orchid code and HOT model), I decided to perform a transcriptome-wide characterization of the MADS-box genes expressed in the inflorescence of *O. italica* and to verify if some of them could interact with the MYB transcription factors DIV, RAD and DRIF1, potentially involved in the zygomorphy of the orchid flower.

More in details, the second part of my work can be divided in different points:

- Identification of all the MADS-box genes expressed in the inflorescence transcriptome of *O. italica* by *in silico* analysis;
- Phylogenetic analysis of the orchid MADS-box genes;
- Expression analysis of the MADS-box genes in the floral tissues of *O. italica* by real time PCR;
- Evolutionary analysis of the MADS-box genes most expressed in the floral tissues of *O. italica;*
- Study of the protein interaction between the B-class MADS-box proteins and the MYB factors DIV, RAD and DRIF1 of *O. italica* by yeast two hybrid analysis.

2. Results and Discussion

2.1 Identification of the MADS-box genes of *O. italica*

The analysis of the inflorescence transcriptome of *O. italica* reveals the presence of twenty-nine transcripts encoding for MADS-box proteins. Respect to the number of MADS-box genes reported in the orchid species with a released genome assembly (*P. equestris, D. catenatum* and *A. shenzenica*) the number of MADS-box isolated in this study is lower because were considered only the transcripts expressed in the inflorescence of *O. italica* [51, 52, 109]. BLAST and phylogenetic analysis show that in *O. italica* are expressed both class I and class II MADS-box genes (Fig. 20).

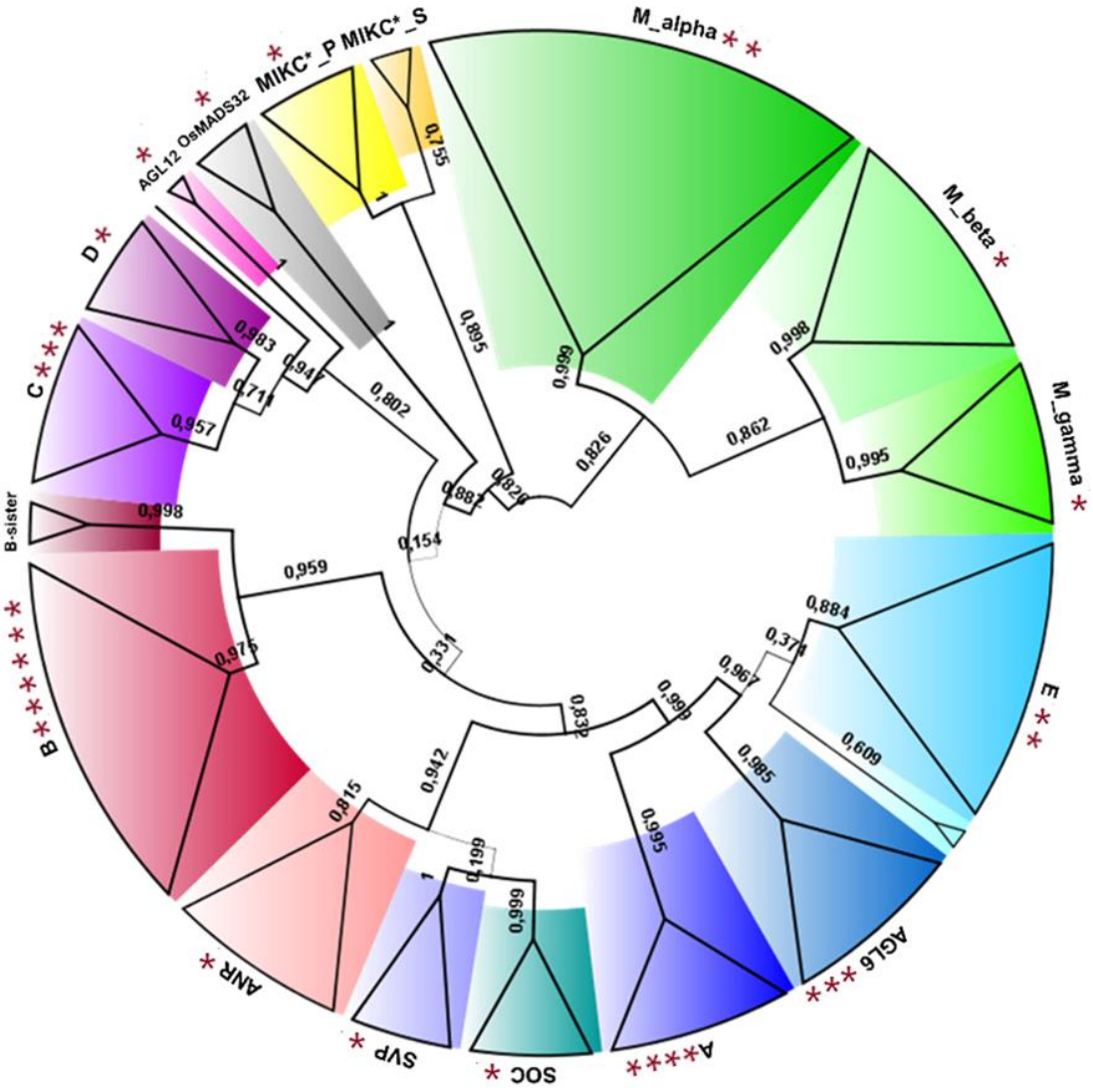

Fig. 20 Maximum likelihood tree of the orchid MADS-box proteins. The collapsed branches include MADS-box proteins of *O. italica*, *P. equestris* and *A. shenzenica*. The red asterisks indicate the number of sequences identified in *O. italica*. The numbers indicate the statistical support of the branches.

2.1.1 Class I MADS-box genes

Previous studies proposed that the orchid genomes lack the class I Mβ genes [109]. In the transcriptome of *O. italica* are expressed four class I transcripts, two belonging to the Mα type, one to the Mγ and one to the Mβ (Fig. 21). This result demonstrates the presence of Mβ genes in orchids, as recently reported also in *P. aphrodite* [110]. The class I MADS-box transcripts of *O. italica* are poorly expressed in all the floral tissues, as expected based on their role in embryo and endosperm maturation [66] and in agreement with their expression levels in the orchid *A. shenzenica* [109]. Although at low levels, they are expressed in all the floral tissues, also in the column (Fig. 22), at whose top are located the pollen grains. The expression in this tissue is in line with the pattern observed the orchid *P. aphrodite* [110] and suggests the involvement of the class I MADS-box genes in the development of the orchid pollinia.

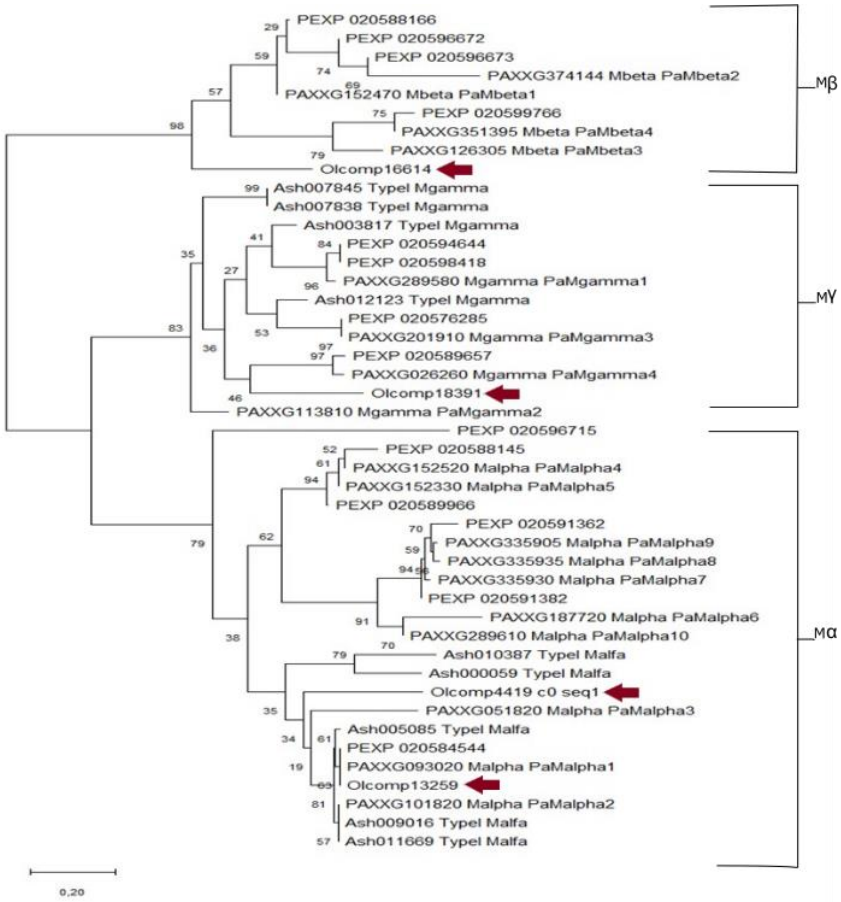

Fig. 21 Maximum likelihood tree of the orchid MADS-box class I proteins. The numbers indicate the statistic support of the branches. The arrows indicate the *O. italica* MADS-box class I proteins.

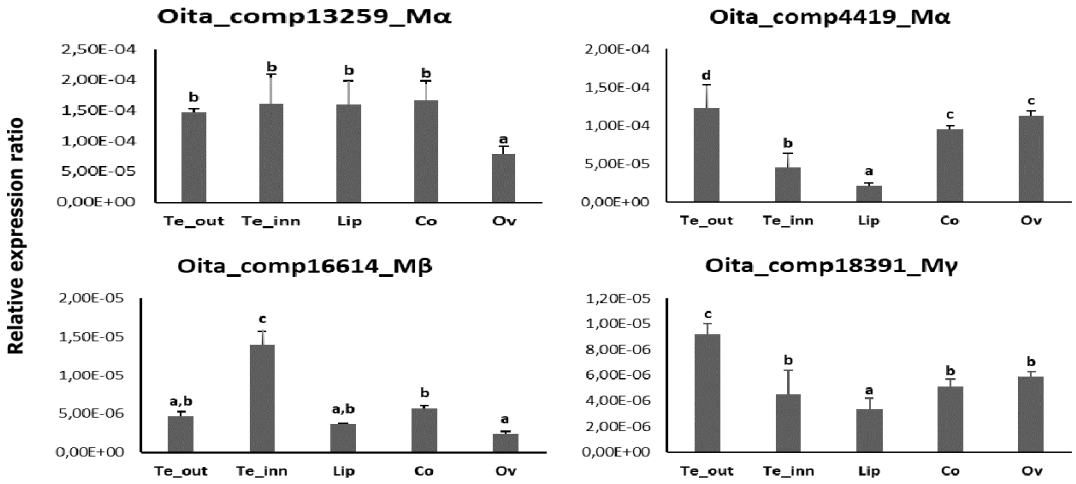

Fig. 22 Expression pattern of the class I MADS-box genes of *O. italica*. Relative expression of the class I transcripts of *O. italica* in different floral tissues. Te_out, outer tepal; Te_inn, inner tepal; Co, column; Ov, ovary. The letters above the bars indicate statistically significant groups, as assessed by the Tukey HSD post-hoc test.

2.1.2 Class II MADS-box genes

The class II MADS-box genes are divided into two sub-classes: MIKC* and MIKC^C [71]. The analysis of the inflorescence transcriptome of *O. italica* reveals the presence of one MIKC* and twenty-four MIKC^C transcripts.

- **The MIKC* genes**

The MIKC* transcript expressed in the inflorescence of *O. italica* belongs to the P-subclade. The MIKC* genes are involved in the development of gametophyte [111, 112] and the MIKC* transcript of *O. italica* is highly expressed in the column (Fig. 23), as in the orchids *E. pusilla* [60] and *P. aphrodite* [110], supporting its function in the development of the orchid gametophyte.

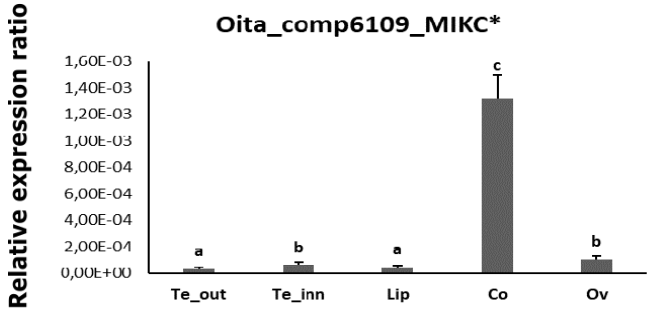

Fig. 23 Expression pattern of the MIKC* gene of *O. italica*. Relative expression of the class II MIKC* transcript of *O. italica* in different floral tissues. Te_out, outer tepal; Te_inn, inner tepal; Co, column; Ov, ovary. The letters above the bars indicate statistically significant groups, as assessed by the Tukey HSD post-hoc test.

- **The *SOC, SVP, ANR1*, *AGL12* and *OsMADS32* genes**

Some MIKC^C genes are involved in the regulation of flowering (as *SOC1* and *SVP*) and root development (as *AGL12* and *ANR1*) [113, 114]. In the inflorescence of *O. italica* are expressed two genes involved in the regulation of flowering (*SOC* and *SVP*), two related to the root development *AGL12* and *ANR1* and one involved in the seed development *(OsMADS32*) [115, 116]. As expected, all these transcripts have a low expression in all the floral tissues (Fig. 24), as reported also in other orchids [60, 117, 118].

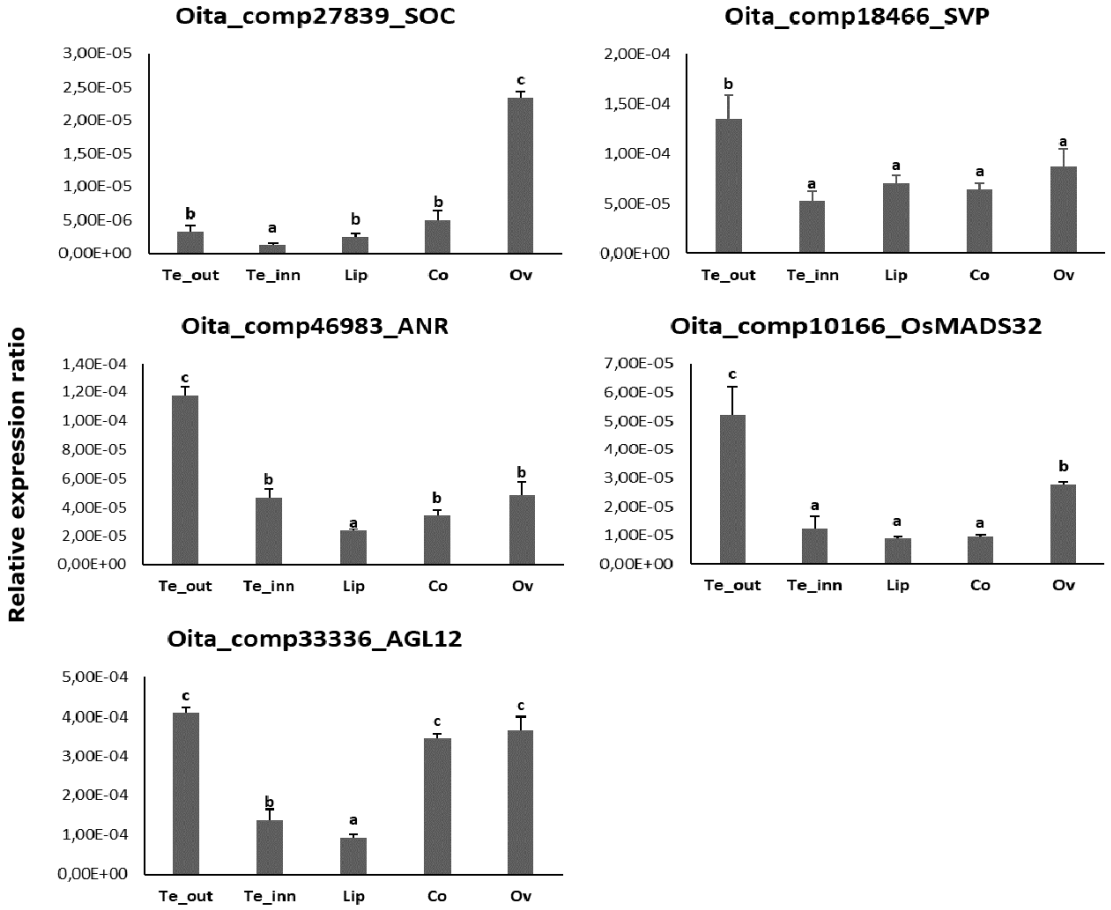

Fig. 24 Expression pattern of the MIKC^C genes of *O. italica.* Relative expression of the *SOC, SVP, ANR, OsMADS32, AGL12* transcripts of *O. italica* in different floral tissues. Te_out, outer tepal; Te_inn, inner tepal; Co, column; Ov, ovary. The letters above the bars indicate statistically significant groups, as assessed by the Tukey HSD post-hoc test.

- **The AGL6-class genes**

Recent studies in *Oncidium* and *P. equestris* [101] have highlighted the role of the MADS-box gene *AGL6* in the formation of the orchid lip. In many orchids there are three *AGL6* transcripts divided in two clades, one that includes also other monocot species and the other

one that includes only orchid species [110, 119]. As in other orchid species, in *O. italica* there are three *AGL6*-like transcripts. Phylogenetic analysis of the *AGL6* transcripts of *O. italica* shows that they belong to the two clades previously reported in orchids. The topology of tree reveals the formation of two different paralog groups probably due to a duplication within the orchid-specific AGL6 clade (Fig. 25). Both groups include transcripts of basal orchid species (Apostasioideae, Cypripedioideae or Vanilloideae*)* in addition to Epidendroideae and Orchidoideae.

Analysis of the expression pattern of these transcripts (Fig. 26) reveals an overlapping profile of expression of *AGL6_OITA_8204* and *AGL6_OITA_1386*. Both transcripts are expressed at high level in lip and lower level in outer tepals. This result suggests that the paralogs *AGL6_OITA_8204* and *AGL6_OITA_1386* probably have a redundant functional role in the formation of lip in *O. italica*. In addition, *AGL6_OITA_4335* presents a high expression in outer tepals and ovary, suggesting a role in the reproductive structures.

The evolutionary analysis of the orchid *AGL6* coding sequences reveals no evidence of positive selection or relaxation of selective constrains (Tab. 6-7 - Appendix).

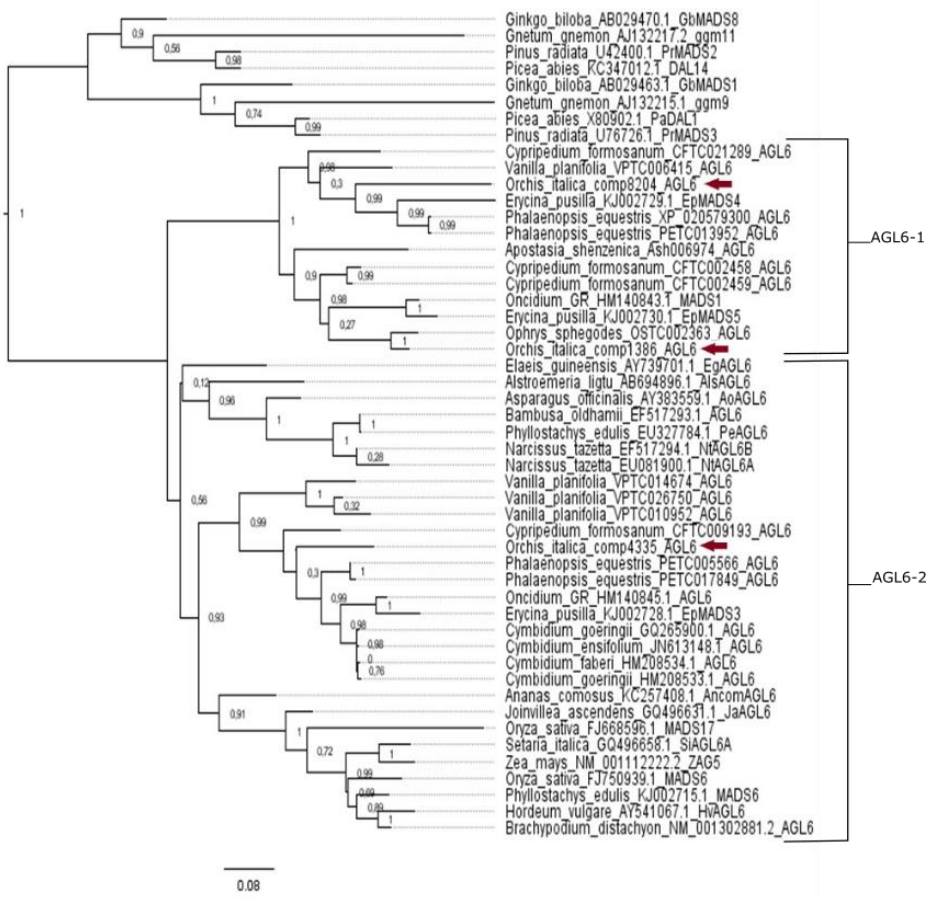

Fig. 25 Maximum likelihood tree of the AGL6 proteins. The numbers indicate the statistic support of the branches. The arrows indicate the *O. italica* AGL6 proteins.

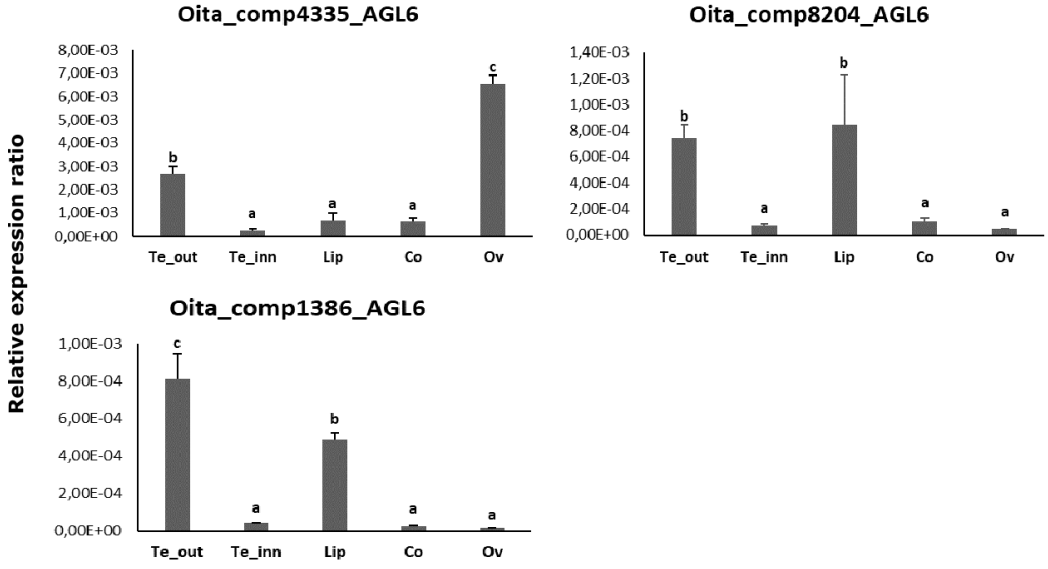

Fig. 26 Expression pattern of the *AGL6* genes of *O. italica*. Relative expression of the AGL6 class transcripts of *O. italica* in different floral tissues. Te_out, outer tepal; Te_inn, inner tepal; Co, column; Ov, ovary. The letters above the bars indicate statistically significant groups, as assessed by the Tukey HSD post-hoc test.

- **The A-Class genes**

In the inflorescence transcriptome of *O. italica* there are four transcripts belonging to the two monocot *FUL*-like groups previously identified in other species [120]. An orchid-specific duplication expanded the number of the orchid *AP1/FUL* genes producing four, well supported AP1/FUL-like clades (Fig. 27). Three *AP1/FUL*-like transcripts of *O. italica* (*OITA_9283_AP1, OITA_11046_AP1, OITA_2508_AP1)* encode for proteins that present the complete *FUL*-like motif LPPWML at the C-terminus, while in *OITA_3679_AP1* it is absents. This is not a peculiar feature of *O. italica* and it is reported also in other species belonging to Epidendroideae [96] and Apostasioideae. As the divergence of the C-terminus is shared by members of the AP1/FUL proteins belonging to the basal Apostasioideae subfamily, [121] it is possible that it is an ancient condition, raised early during the orchid evolution.

The expression analysis of the *AP1/FUL* genes *OITA_3679_AP1*, *OITA_9283_AP1* and *OITA_2508_AP1* of *O. italica* (Fig. 28) reveals a pattern very similar to *EpMADS10*, *11* and *12* of *E. pusilla* [60, 122]. They show a high expression in column and ovary and are expressed also in outer and inner tepals. Contrastingly, *OITA_11046_AP1* has a high expression in outer and inner tepals and a weak expression in column and ovary (Fig. 28).

The evolutionary analysis reveals absence of positive selection within the *AP1/FUL* coding sequences of orchids; however, relaxation of purifying selection is evident when the orchid *AP1/FUL* genes (ω=0.23) are compared with the orchid MADS-box genes of the related functional classes *SEP* (ω=0.11) and *AGL6* (ω=0.14) (Tab. 6-7 - Appendix). This result is in agreement with the previous report of diversifying selection in the orchid *AP1/FUL* genes [122] and suggests a possible functional diversification after duplication of these orchid genes, supported also by the divergence of the C-terminus sequence of some members of this clade and by the peculiar expression profile of *OITA_11046_AP1*.

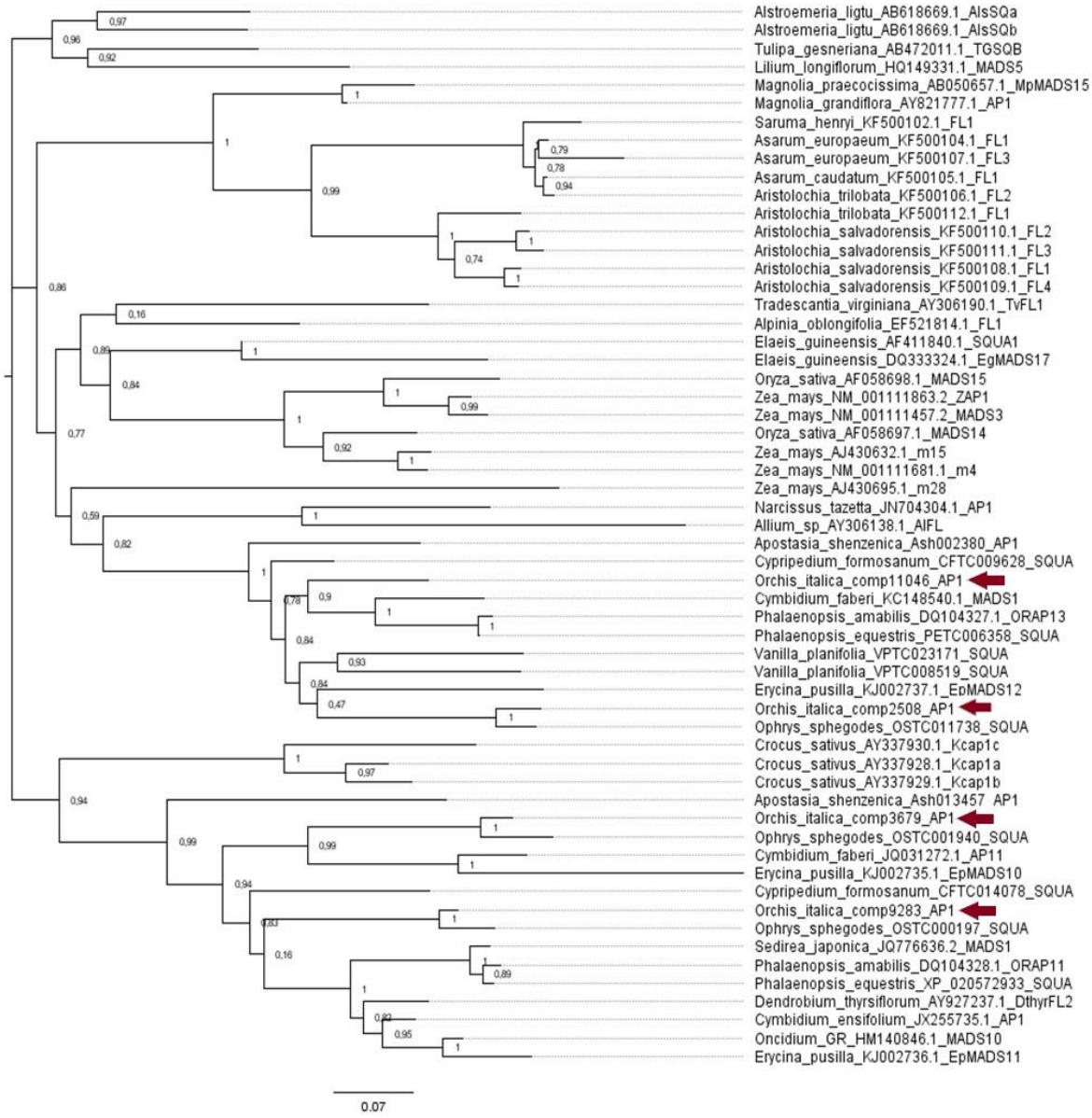

Fig. 27 Maximum likelihood tree of the AP1/FUL proteins. The numbers indicate the statistic support of the branches. The arrows indicate the *O. italica* MADS-box class A proteins.

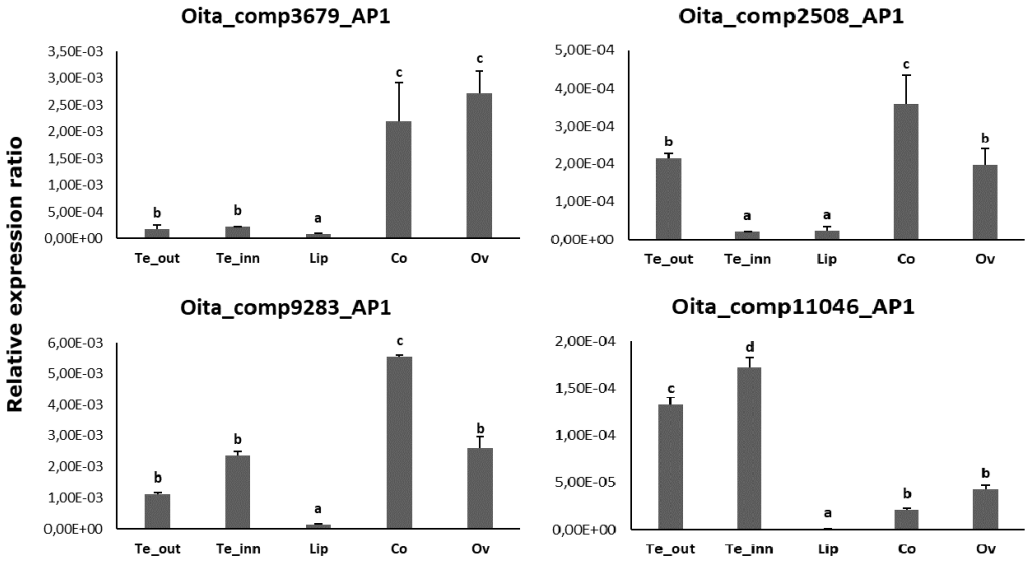

Fig. 28 Expression pattern of the class A genes of *O. italica.* Relative expression of the A class transcripts of *O. italica* in different floral tissues. Te_out, outer tepal; Te_inn, inner tepal; Co, column; Ov, ovary. The letters above the bars indicate statistically significant groups, as assessed by the Tukey HSD post-hoc test.

- **The C- and D- class genes**

Previous studies have identified one C- (*AG*) and one D-class (*STK*) gene in *O. italica* [108]. In addition to these, in the present study two additional *AG* transcripts have been identified: *OITA_1784_AG* and *OITA_16614_AG.* Three C-class and one D-class genes are present also in *E. pusilla* [60, 122], suggesting that this is the number of *AG/STK* genes also in *O. italica*.

Both the transcripts encode for proteins containing the AG-motif I and II at the C-terminus. The OITA_1784_AG shows 69% amino acid identity with the C-class OitaAG and 62% with the D-class OitaSTK, whereas OITA_16614_AG shows 71% identity with OitaAG and 52% with OitaSTK. Also in the transcriptome of *Ophrys sphegodes* (Orchidoideae) there are transcripts encoding for AG proteins similar (84% identity) to those of *O. italica.*

The expression analysis reveals that the newly identified *AG* transcripts of *O. italica* have a high level of expression in ovary and column (Fig. 29), in agreement with the canonical expression pattern of the C-class genes in other orchid species and with their role in the development of female reproductive structures [58, 96, 123].

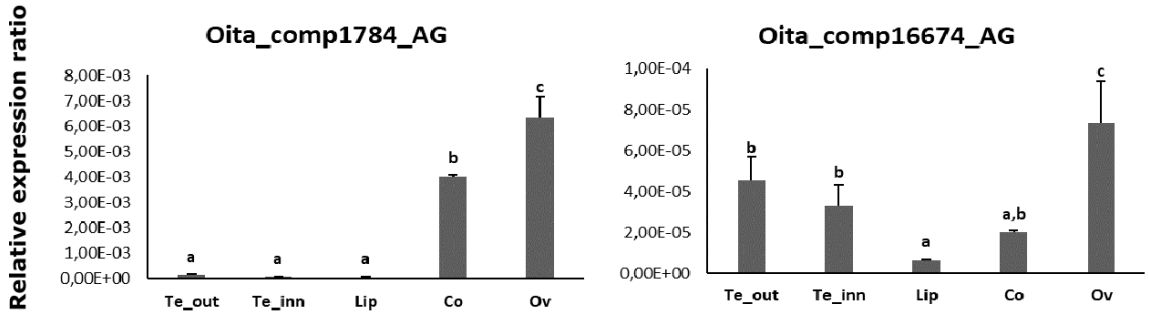

Fig. 29 Expression pattern of the class C genes of *O. italica*. Relative expression of the C class transcripts of *O. italica* in different floral tissues. Te_out, outer tepal; Te_inn, inner tepal; Co, column; Ov, ovary. The letters above the bars indicate statistically significant groups, as assessed by the Tukey HSD post-hoc test.

- **The E-class genes**

The E-class of MADS-box genes are involved in the establishment of all the floral organs in *Arabidopsis thaliana*. In orchids there are four clades of *SEP*-like genes raised from orchid-specific duplications within the two SEP monocot clades [60, 96, 121]. In the inflorescence transcriptome of *O. italica* there are only two transcripts (*OITA_7010_SEP* and *OITA_1006_SEP)* that encode for SEP proteins, both containing the SEP I and II motifs.

Phylogenetic analysis (Fig. 30) reveals that one transcript of *O. italica* belongs to the SEP clade 3 (also including *EpMADS9* and *PeSEP3*) while the other belongs to the SEP clade 4 (also including *EpMADS7* and *PeSEP4*). BLAST search in the transcriptome of *Ophrys sphegodes* [50] shows that also in this species only two SEP-like transcripts are detectable, with significant hits and sequence identity to the SEP transcripts isolated in *O. italica*. In addition, in *Habenaria radiata* (Orchidoideae) a recent study reports the presence of only two SEP-like transcripts in the floral buds [124]. Taken together, all these evidences suggest that in the Orchidoideae subfamily there are only two SEP genes or that only two SEP genes are expressed in the inflorescence. Due to the absence of an available assembled genome of Orchidoideae, it is not possible to discriminate between these two hypotheses.

The expression pattern of the *SEP* transcripts of *O. italica* (Fig. 31) shows that both are expressed in all the organs of the perianth and that *OITA_7010_SEP* is expressed also in column and ovary, as reported in other orchids as *H. radiata* [124], *P. equestris* [96] and *E. pusilla* [60, 122].

The evolutionary analysis reveals a strong and homogeneous purifying selection acting on the *SEP* orchid genes (Tab. 6-7 - Appendix).

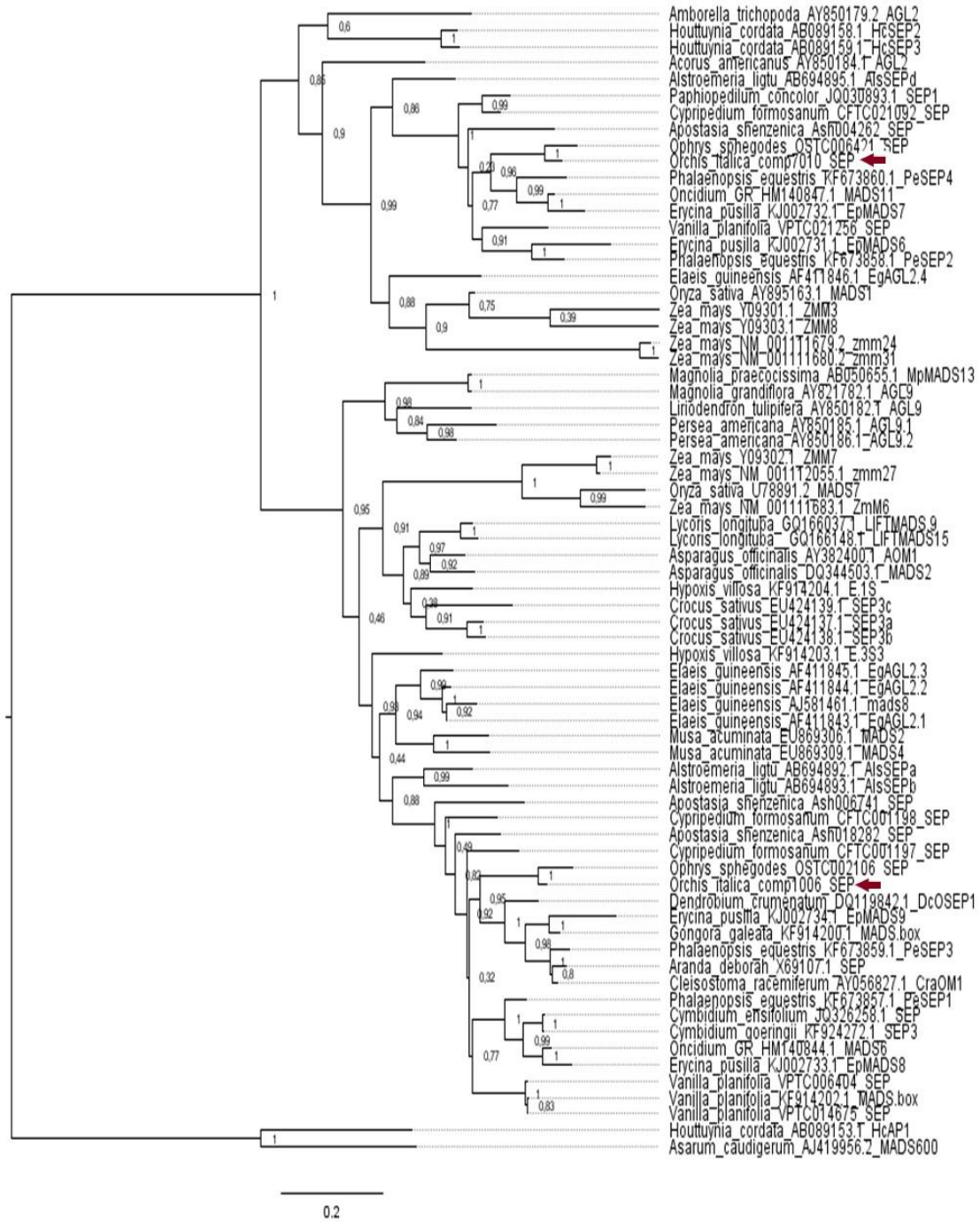

Fig. 30 Maximum likelihood tree of the SEP proteins. The numbers indicate the statistic support of the branches. The arrows indicate the *O. italica* MADS-box class E proteins.

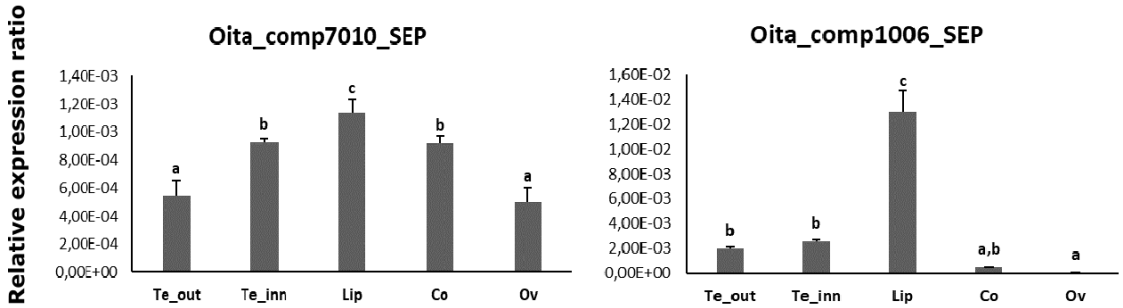

Fig. 31 Expression pattern of the E class genes of *O. italica.* Relative expression of the E class transcripts of *O. italica* in different floral tissues. Te_out, outer tepal; Te_inn, inner tepal; Co, column; Ov, ovary. The letters above the bars indicate statistically significant groups, as assessed by the Tukey HSD post-hoc test.

2.2 Protein-protein interaction between B-class MADS-box and MYB transcription factors

Studies on the evolution of the class-B genes (specifically on the *DEF*-like genes) have shown that two rounds of duplications of these genes have played an important role in the evolutionary origin of the current structure of the orchid flower [13, 58]. As both the class-B MADS-box and the MYB genes are implicated in the establishment of flower symmetry, we have hypothesized a possible interaction between MYB and class-B MADS-box factors during the development of the orchid flower. In collaboration with Prof. Maria Manuela Ribeiro Costa at the Plant Functional Biology Center, University of Minho (Braga, Portugal), a “yeast two hybrid system assay” (Y2H) was conducted to verify this hypothesis. The open reading frame (ORF) of the orthologs of the B-class MADS-box genes (*Oita_DEF1-4* and *OITA_PI-PI2*) and of the MYB genes *DIV (OITA_9548*), *RAD* (*OITA_56510*) and *DRIF1 (OITA_10599)* (see Chapter 1) were cloned into two vectors containing the GAL4 DNA-binding and activation domain, respectively.

The results reveal that there is not interaction between the B-class MADS-box and MYB transcription factors (Fig. 32). This result suggests two hypotheses: i) the proteins are involved in the same process (floral zygomorphy) but with two distinct functions and they do not directly interact; ii) using the Y2H screening, their eventual interaction is not detectable because, based on the ABCDE floral quartet model of flower development [75, 84], the MADS-box proteins work after the formation of heteromultimeric complexes (protein quartets). For this reason, to detect the eventual interaction between MADS-box and MYB transcription factors, it will be necessary to perform yeast three or four hybrid screenings.

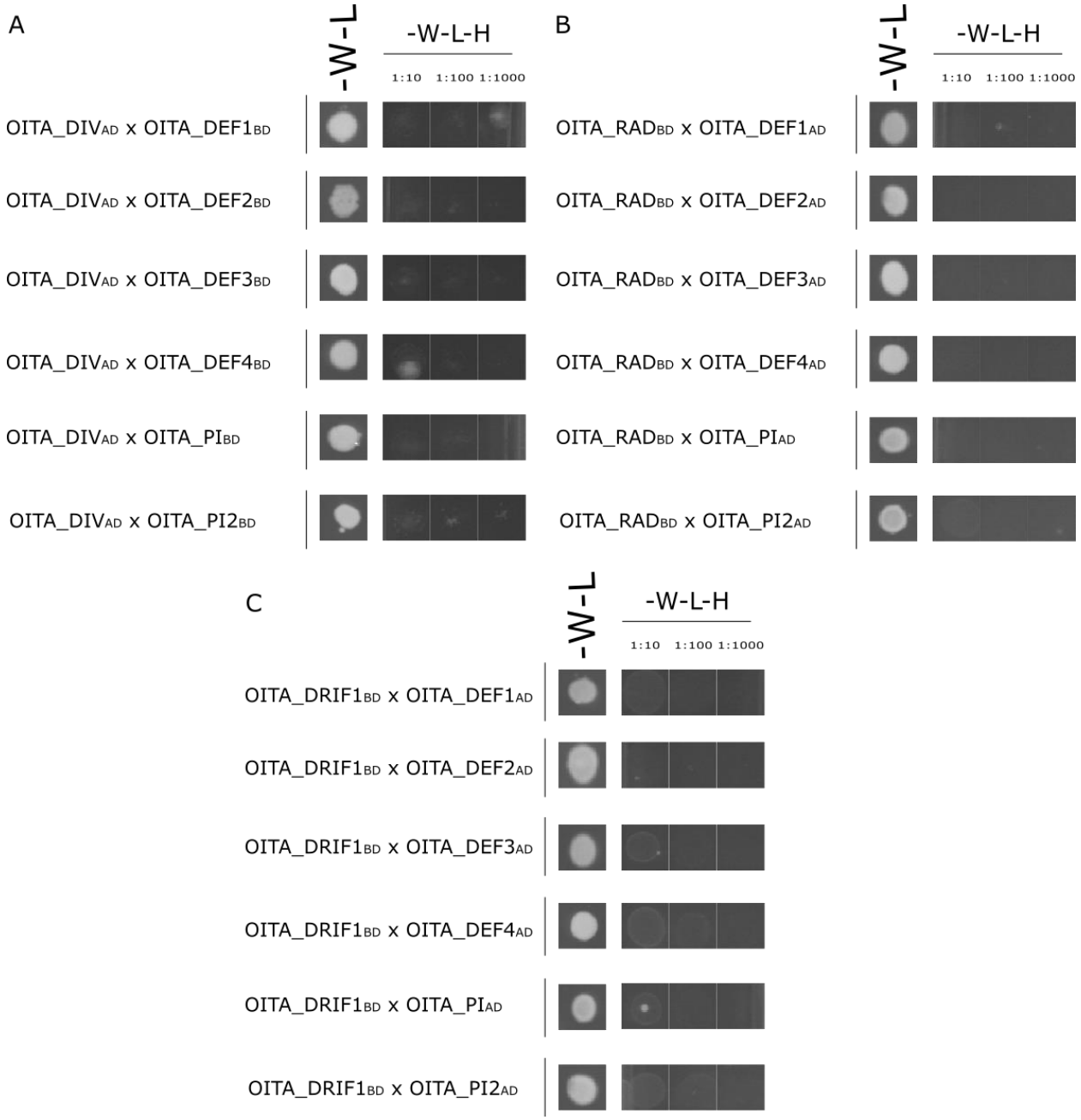

Fig. 32 Protein-protein interaction between B-class MADS-box and MYB transcription factors. Interaction between DIV_{AD} **(A)**, RAD_{BD} **(B)** and DRIF1 **(C)** proteins with all the B-class MADS box genes of *O. italica*. (-W-L, medium lacking the tryptophan and leucine; -W-L-H, medium lacking tryptophan, leucine and histidine. The dilution factor applied for the yeast inoculate is indicated as 1:10, 1:100 and 1:1000.) BD, binding domain; AD, activation domain.

Conclusions

The results obtained during the development of my PhD project demonstrate the existence of a pattern of gene expression and protein interaction of the transcription factors DIV, RAD and DRIF 1 conserved between two distantly relates species: the model eudicot *Antirrhinum majus* and the non-model monocot *Orchis italica*. In *A. majus* these MYB genes are key regulators of the floral dorso-ventral identity and the evidences obtained let us hypothesize that in *O. italica* they could be involved in the establishment of floral zygomorphy. The idea that they could also interact with MADS-box transcription factors that play a relevant role in the formation of the asymmetrical structures of the orchid perianth (mainly the lip) has not been confirmed from our results, even though it cannot be excluded. Yeast three and four hybrid experiments could clarify the existence of a direct MYB/MADS interaction.

The study of non-model organisms is always very exciting and interesting. However, often many problems have to be solved to obtain a clear picture of the molecular pathways underlying the different developmental mechanisms. In particular, one of the main problems of *O. italica* is the extreme rarity of mutants (e.g. peloric or semi-peloric) and the impossibility to obtain them through genetic manipulations. One possible approach to overcome this problem is to analyse the expression pattern of the *DIV*, *RAD* and *DRIF* homologs in peloric mutants of other orchid species (e.g. *Phalaenopsis*).

Other questions about the molecular mechanisms of orchid flower symmetry are still open. First, the role in orchid zygomorphy of the *CYC* genes (TCP transcription factors), whose identification in *O. italica* based on whole sequence similarity with other species gave only partial results (identification of a region of the *TB1* gene). Currently, we are performing both *in silico* analyses using small conserved regions of the CYC proteins as molecular bait and PCR with degenerated primers to identify all the *CYC* genes of *O. italica*.

In addition, other transcription factors might be involved in orchid floral zygomorphy. A recent study demonstrated the role of the *CUP* genes (NAC transcription factors) in floral shaping of *A. majus*, in coordination with MYB and TCP genes [125]. We have just started studying these genes in *O. italica* to clarify their role in orchid flower symmetry.

Material and Methods

1. Plant material

The *Orchis italica* plants used in this work were grown in the greenhouse of the Department of Biology of the University of Naples Federico II (Napoli, Italy). I warmly thank Prof. Salvatore Cozzolino and Prof. Giovanni Scopece for plant material and support during the tissue dissection.

Tissues were collected from two developmental stages: before anthesis at bud stage (early stage of development) and mature inflorescence after anthesis (late stage of development) for different analyses. Single flowers at the two different developmental stages and flower tissues (outer tepals, inner lateral tepals, lip, column, and ovary before pollination) from the late developmental stage were stored in RNA-later (Ambion) until nucleic acids extraction. Single florets of *O. italica* at early stage of development were collected and 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 [126]. They were then dehydrated through ethanol series, paraffin embedded and stored at 4°C until *in situ* hybridization experiments.

2. Identification of the orchid MYB and MADS-box transcripts

- **Identification of the orchid *DIV* and *RAD*-like genes**

A TBLASTN analysis was conducted using the amino acid sequence of DIV and DVL1 (AAL78741, AAL78742) and RAD and RAD-like (AAX48042, ABI14752, ABI14753, ABI14755, AJ791699, AJ793240) of *Antirrhinum majus* on the available transcriptomes of two Orchidoideae species: *Orchis italica* [48] and *Ophrys sphegodes* [50]. In addition, a TBLASTN analysis on the *Phalaenopsis equestris* CDSs (downloaded from ftp://ftp.genomics.org.cn/from_BGISZ/20130120/) was conducted. Genome scaffolds of *P. equestris* and *Dendrobium catenatum* were scanned [51, 55] to identify the exon/intron junctions of the *RAD*-like and *DIV*-like genes.

The selected CDSs of *O. italica, Ophrys, P. equestris* and *D. catenatum* were virtually translated to verify the absence of indels and/or stop codons.

To identify the orchid DIV- and RAD-like ortholog groups, all-versus-all BLASTP searches were conducted between the predicted amino acid sequences of *O. italica*, *Ophrys*, *P. equestris*, *D. catenatum* and *A. majus* to select the best reciprocal hits. In addition, BLASTP

search was performed against the nonredundant protein database and an OrthoMCL analysis was conducted [127]. Paralogs were identified through a BLASTCLUST search within each of the orchid species cited above using the following parameters: e-value ≤ 10^{-10} and 30% minimum similarity (–S 30) over 75% of the protein (–L 0.75) [128, 129].

- **Identification of the *DRIF* and *AGL6* transcripts of *O. italica***

The *DRIF* and *AGL6* transcripts were identified through a TBLASTN analysis against the inflorescence transcriptome of *O. italica* using as query the DRIF proteins of *A. majus* (DRIF1, AGL11918; DRIF2, AGL11919) [46] and the *P. aphrodite* AGL6 proteins (AGL6-1, PATC154379; AGL6-2 PATC138772) [130].

- **Identification of the MADS-box transcripts of *O. italica***

To identify the transcripts encoding for the MADS-box protein of *O. italica*, a TBLASTN analysis was conducted on the *O. italica* inflorescence transcriptome using as query the conserved amino acid sequence of the known MADS-box domains of *O. italica* [6, 92, 93, 95, 108]. The selected transcripts were virtually translated to check the presence of stop codons and to exclude those missing the regions downstream the MADS domain. The annotation of the selected transcripts was performed through a BLASTP search using their virtual translation as query against the Viridiplantae protein database.

The MADS-box nucleotide and protein sequences of the orchid *P. equestris* and *Apostasia shenzenica* [109] [51] and other plant species were downloaded from GenBank and their accession numbers are included in the name used in the subsequent analyses.

3. Analysis of the *Orchis italica* transcripts

Total RNA was extracted from pool of 10 single florets of *O. italica* after anthesis using the extraction RNA kit for Plants PureLink™ RNA Mini Kit (Invitrogen™) followed by DNase treatment. After spectrophotometric quantification using the Nanodrop 2000 (Thermo Fisher Scientific), RNA (500 ng) was reverse transcribed using the Advantage RT-PCR kit (Euroclone) and oligo dT primer and hexamer random primers. Using the nucleotide sequences obtained from *in silico* analysis, specific primer pairs were designed to amplify the whole selected transcripts of *O. italica* [14] (Tab. 8 - Appendix). For the transcripts where the sequences were not complete, specific primers [14] were designed to perform the 5’ and 3’ RACE using the FirstChoice RLM-RACE kit (Thermo Fisher Scientific). The PCR amplifications were conducted using two different *Taq* polymerases. The *DIV-*, *RAD* and

DRIF1-like transcripts were amplified using 50 ng of first strand cDNA, 0.20 mM dNTPs, 0.5 µM of each primer, 1x buffer and 2.5 U HotMaster *Taq* DNA polymerase (5 Prime) with the following thermal cycle: 94°C for 3 min, 35 cycles of 94°C for 30 s, 60°C for 20 s, 65°C for a time dependent upon the amplicon size (from 30 s to 3 min), followed by a final extension of 10 min at 65°C. The MADS-box transcripts were amplified using 50 ng of first strand cDNA 0.20 mM of dNTPs, 0.5 µM of each primer and 2.5 U of Wonder *Taq* DNA polymerase (Euroclone) with the following thermal cycle: 95°C for 1 min, 35 cycles of 35°C for 15 s, 60°C for 15 s and 72°C for a time dependent from the amplicon size (from 15 s to 5 min). All the amplicons were cloned into the pSC-A-amp/kan vector (Agilent), sequenced using the T3 and T7 plasmid primers and analysed using the ABI 310 Automated Sequencer (Applied Biosystems).

4. Gene structure analysis

- **Analysis of *DIV*- and *RAD*-like genes**

Total DNA was extracted from leaf tissue of *O. italica* using the ISOLATE II Plant DNA kit (Bioline) followed by RNase treatment and quantification. Based on *in silico* analysis and intron position on the *DIV*- and *RAD*-like genes of *P. equestris* and *D. catenatum,* specific primers pairs were designed to amplify the introns of these genes in *O. italica* [14]. PCR amplifications were conducted using two different High fidelity *Taq* polymerases, Herculase II Fusion DNA polymerase (Agilent) and RANGER DNA Polymerase (Bioline), following the manufacturer guidelines.

Before cloning the genomic amplicons, a dATP tailing reaction was performed by adding 0.5 U of *DreamTaq* DNA polymerase (Thermo Fisher Scientific) and 100 µM dATP to the PCR reaction and incubating at 72°C for 15 min. The amplicons were then cloned into the pSC-A-amp/kan vector (Agilent), and sequenced using vector- and intron-specific primers.

The intron nucleotide sequences of the *DIV*- and *RAD*-like genes of *O. italica*, *P. equestris*, and *D. catenatum* were scanned for the presence of transposable/repetitive elements using the CENSOR software [131].

5. Phylogenetic and evolutionary analysis

- **Analysis of the *DIV*- and *RAD*-like genes**

The predicted amino acid sequences of the DIV- and RAD-like proteins of *O. italica, Ophrys, P. equestris* and *D. catenatum* were aligned using MUSCLE [132] and the corresponding nucleotide alignments of the *DIV*- and *RAD*- like CDSs were obtained using PAL2NAL [133]. Based on the amino acid alignment of the DIV- and RAD-like sequences, the Maximum Likelihood (ML) and Neighbor-Joining (NJ) trees were constructed using the MEGA7 software [134] with the JTT+G amino acid substitution model with 1,000 bootstrap replicates.

The presence of shared conserved motifs among the predicted amino acid sequences of the DIV- and RAD-like proteins *of O. italica, Ophrys, P. equestris, and D. catenatum* was verified using the MEME online tool [135].

The *DIV*- and *RAD*-like CDSs of all the orchid species analysed were used to test the variation of evolutionary rates at specific codons and among the tree branches using the CODEML program implemented in PAML v.4.8 [136]. The ω value (ratio between nonsynonymous and synonymous substitution rates) of the *DIV*- and *RAD*-like sequences was evaluated under four different evolutionary models: branch, sites, branch-sites and clades. The branch models assume one or different ω values among the branches of the tree, while the sites models verify the presence of positively selected codons among the sequences. The branch-site models test the existence of positive selection on individual codons in specific branches of the tree. Finally, the clade model verifies the presence of different selective constraints between clades after gene duplication If ω is 1, neutral selection is acting on the examined sequences, whereas if ω is lower or higher than 1, there is purifying or positive selection, respectively [137]. A likelihood ratio test was applied to establish which model best fits the data.

- **Analysis of the MADS-box genes**

The predicted amino acid sequences of *O. italica, Ophrys, P. equestris* and *A. shenzenica* were aligned using MAFFT [138] and the corresponding nucleotide sequences were aligned by PAL2NAL [133]. Using the amino acid alignment of the MADS-box sequences of *O. italica, Ophrys, P. equestris* and *A. shenzenica*, the ML tree was constructed by PhyML [139]. In addition, the amino acid alignments of the different classes of MADS-box proteins of different plant species were separately constructed and used to obtain class-specific ML trees by PhyML.

CODEML program from PAML v.4.9 [136] was used to analyse the evolutionary rates of the *AP1/FUL, SEP* and *AGL6* CDSs of *O. italica* and other orchids. The ratio between nonsynonymous and synonymous substitution rates was calculated (ω) and the branch and branch-sites evolutionary models were tested. A likelihood ratio test was applied to establish which model best fits the data.

6. Expression Analysis

- ***In situ* hybridization of the *DIV RAD* and *DRIF1* transcripts**

The paraffin embedded samples (see Plant material) were sectioned at 9 µm.
Primer pairs specific for the transcripts *OITA_9548 (DIV), OITA_56510 (RAD)* and *OITA_10599 (DRIF1)* were designed to amplify a specific region for the probe synthesis [14]. The amplicons were cloned into pGEM®-T Easy Vector (Promega) and the antisense/sense RNA probes synthesis was carried out using the T7 and SP6 RNA polymerases and the DIG RNA Labeling kit (Roche). The detection of the signals with alkaline phosphatase was performed using the DIG Nucleic Acid Detection kit (Roche).

- **Real Time PCR**

Total RNA was extracted from different tissues of *O. italica* after anthesis: outer tepals (Te_out), inner lateral tepals (Te_inn), labellum (Lip), column (Co), ovary not pollinated (Ov) and leaf (Le). After the reverse transcription of RNA from each tissue, 100 ng of first stand cDNA were amplified using the SYBR Green PCR Master Mix (Life Technologies) in technical triplicates and biological duplicates. Specific primers pairs were designed for all genes *O. italica* [14] and Tab. 9 – Appendix, using the 5.8S RNA as the endogenous control gene.

The relative expression ratio (Rn) was calculated using the following calculation:

$$Rn = \frac{(1 + E_{target})^{-CT_{target}}}{(1 + E_{5.8S})^{-CT_{5.8S}}}$$

where E is the PCR efficiency and C_T the threshold cycle. The mean Rn and standard error (SE) were calculated for each tissue and the statistical significance of the mean Rn differences among and between tissues was evaluated using the ANOVA test followed by the Tukey post hoc test, and two-tailed t-tests, respectively.

7. Yeast two-hybrid analysis

Specific primer pairs (Tab. 10 – Appendix) were designed to amplify the open reading frame of the transcripts *OITA_9548 (DIV), OITA_56510 (RAD), OITA_10599 (DRIF1)* and the B-class of MADS box genes *OITA_PI, OITA_PI2, OITA_DEF1, OITA_DEF2, OITA_DEF3, OITA_DEF4* (AB094985, AB537504, AB857726, AB857727, AB857728, AB857729, respectively). The amplified sequences were cloned into pGBT9 (bait vector; Clontech) and pGAD424 (prey vector; Clontech).

Protein-protein interactions were analysed using a GAL-4 based yeast two hybrid system (Matchmaker two-hybrid system; Clontech) and the proteins fused to the binding domain of GAL4 were tested for self-activation monitoring the transformed cells in SD medium with 5 mM 3-amino triazole and without histidine.

The different vectors, in all the combinations, were transformed in the *Saccharomyces cerevisiae* strain AH109 using the LiAc/DNA/PEG transformation method [140]. The selection of positive interactions and β-galactosidase assay was performed according to Causier and Davies (2002) [141].

Appendix

Subfamily	Species	Transcript_Name	cDNA_Size	ORF_Size	Intron_Size	Accession/Scaffold
DIV-like	*O. italica*	OITA_9548	1,361	864	76	KY089088
		OITA_23026[a]	737	528	111	KY089095
		OITA_3530	1,901	873	201	KY089091
		OITA_35312	1,182	894	1,615	KY089094
		OITA_13252[b]	1,694	861	3,591	KY089093
		OITA_13233	1,427	840	~5,000	KY089092
		OITA_8681	1,472	855	~6,000	KY089089
		OITA_12910	1,101	861	~10,000	KY089090
	Ophrys	OPH_5397	465	462	Unknown	-
		OPH_23790	809	708	Unknown	-
		OPH_29334	1,384	840	Unknown	-
		OPH_22868	1,145	675	Unknown	-
	P. equestris	PEQU_28155	-	891	246	Scaffold000684_20
		PEQU_00313	-	846	276	Scaffold000002_542
		PEQU_37054	-	891	2,090	Scaffold000306_1
		PEQU_03826	-	861	12,810	Scaffold000002_5659
		PEQU_21184	-	870	5,008	Scaffold000065_25
		PEQU_09358	-	885	174	Scaffold000006_57
	D. catenatum	DCAT_JSDN01S056792	-	888	124	JSDN01S056792
		DCAT_JSDN01S046716	-	525	87	JSDN01S046716
		DCAT_JSDN01S069654	-	876	275	JSDN01S069654
		DCAT_JSDN01S055018	-	882	1,327	JSDN01S055018
		DCAT_JSDN01S002760	-	804	8,197	JSDN01S002760
		DCAT_JSDN01S031580	-	867	2,705	JSDN01S031580
		DCAT_JSDN01S018459	-	855	3,275	JSDN01S018459
		DCAT_JSDN01S042335	-	891	231	JSDN01S042335
RAD-like	*O. italica*	OITA_56510	420	282	1,012	KY089097
		OITA_103296	406	285	1,056	KY089098
		OITA_32153	369	249	1,135	KY089096
		OITA_72143	543	267	No	KY089099
	Ophrys	OPH_2593	660	282	Unknown	-
	P. equestris	PEQU_08237	-	288	1,775	Scaffold000054_59
		PEQU_31458	-	294	864	Scaffold000460_4
		PEQU_40552	-	279	1,191	Scaffold198270_12
		PEQU_25415	-	249	No	Scaffold000067_1
		PEQU_05110	-	294	No	Scaffold000840_202
	D. catenatum	DCAT_JSDN01S033743	-	288	1,240	JSDN01S033743
		DCAT_JSDN01S033707		285	1,178	JSDN01S033707
		DCAT_JSDN01S041675	-	279	1,108	JSDN01S041675
		DCAT_JSDN01S013347	-	324	No	JSDN01S013347
		DCAT_JSDN01S059067	-	303	No	JSDN01S059067

Tab. 1 The *DIV*- and *RAD*-like genes in the orchid species analysed. The size of the cDNA, ORF and intron are expressed in bp. [a] Alternative splicing form with intron retention and premature stop codon. [b] Alternative splicing of an additional intron within the 5'UTR (88 bp) that is conserved in PEQU (86 bp) and DCAT (90 bp).

Model	Foreground_branch	Statistics	lnL
One-ratio		ω=0.1860	-15089,66289
Two-ratios1	1	ω_1=0.0435, ω_2=ω_3=0.0843	-4355,82460
Two-ratios2	2	ω_2=0.0779, ω_1=ω_3=0.0591	-4360,53299
Two-ratios3	3	ω_3=0.1265, ω_1=ω_2=0.0638	-4358,40636
Three-ratios		ω_1=0.0435, ω_2=0.0777, ω_3=0.1272	-4354,19245
Sites_Alternative		p0=0.99937 p=0.60319 q=1.92439 (p1=0.00063) ω=6.73361	-14626,28686
Sites_Null		p=0.60240 q=1.91735	-14626,29774
Branch-site_Alternative1	1	site class 0 1 2a 2b proportion 0.59004 0.30228 0.07120 0.03648 background ω 0.11995 1.00000 0.11995 1.00000 foreground ω 0.11995 1.00000 1.00000 1.00000	-14895,86638
Branch-site_Alternative2	2	site class 0 1 2a 2b proportion 0.48474 0.12768 0.30678 0.08081 background ω 0.09740 1.00000 0.09740 1.00000 foreground ω 0.09740 1.00000 1.00000 1.00000	-14832,92932
Branch-site_Alternative3	3	site class 0 1 2a 2b proportion 0.53170 0.26709 0.13393 0.06728 background ω 0.12852 1.00000 0.12852 1.00000 foreground ω 0.12852 1.00000 1.32697 1.32697	-14899,87767
Branch-site_Null1	1	site class 0 1 2a 2b proportion 0.59004 0.30228 0.07120 0.03648 background ω 0.11995 1.00000 0.11995 1.00000 foreground ω 0.11995 1.00000 1.00000 1.00000	-14895,86638
Branch-site_Null2	2	site class 0 1 2a 2b proportion 0.48474 0.12768 0.30678 0.08081 background ω 0.09740 1.00000 0.09740 1.00000 foreground ω 0.09740 1.00000 1.00000 1.00000	-14832,92932
Branch-site_Null3	3	site class 0 1 2a 2b proportion 0.51988 0.26425 0.14312 0.07275 background ω 0.12707 1.00000 0.12707 1.00000 foreground ω 0.12707 1.00000 1.00000 1.00000	-14900,61465
Clades_Alternative1	1	site class 0 1 2 proportion 0.64839 0.35161 0.00000 branch type 0: 0.13376 1.00000 28.57514 branch type 1: 0.13376 1.00000 25.94735 (note that p[2] is zero)	-14922,19598
Clades_Alternative2	2	site class 0 1 2 proportion 0.60412 0.05844 0.33744 branch type 0: 0.29715 1.00000 0.01754 branch type 1: 0.29715 1.00000 0.03487	-14662,13922
Clades_Alternative3	3	site class 0 1 2 proportion 0.64839 0.35161 0.00000 branch type 0: 0.13376 1.00000 27.74734 branch type 1: 0.13376 1.00000 11.80042 (note that p[2] is zero)	-14922,19598
Clade_Null		p: 0.64839 0.30028 0.05134 ω: 0.13376 1.00000 1.00000	-14922,19598

Tab. 2 Models used, and parameters estimated under different conditions for the evolutionary analysis of the orchid *DIV*-like coding regions.

Model Comparison	LRT	df	p	
One-ratio_vs_Two-ratios1	21467,677	1	<0.00001	
One-ratio_vs_Two-ratios2	21458,26	1	<0.00001	
One-ratio_vs_Two-ratios3	21462,513	1	<0.00001	
One-ratio_vs_Three-ratios	21470,941	2	<0.00001	
Two-ratios1_vs_Three-ratios	3,264292	1	0.070803	ns
Two-ratios2_vs_Three-ratios	12,681068	1	0.000369	
Two-ratios3_vs_Three-ratios	8,427824	1	0.003695	
Sites_Null_vs_Sites_Alternative	0,021774	2	0.88262	ns
Null1_vs_Branch-site_Alternative1	0	1	1	ns
Null2_vs_Branch-site_Alternative2	0	1	1	ns
Null3_vs_Branch-site_Alternative3	1,47395	1	0.224716	ns
Null_vs_Clades_Alternative1	0	2	1	ns
Null_vs_Clades_Alternative2	520,11351	2	< 0.00001	
Null_vs_Clades_Alternative3	0	2	1	ns

Tab. 3 Likelihood ratio statistics for the comparison of the evolutionary models of the orchid *DIV*-like coding regions. df, degrees of freedom; ns, not significant.

Model	Foreground branch	Statistics	lnL
One-ratio		ω=0.1111	-2951,16452
Two-ratios	1	ω_1=0.0899, ω_2=0.0518	-2063,51814
Sites_Alternative		p_0= 0.99999 p= 0.61306 q= 3.56596 (p_1= 0.00001) ω= 1.00000	-2879,02731
Sites_Null		p= 0.61305 q= 3.56584	-2879,02703
Branch-site_Alternative	1	site class 0 1 2a 2b proportion 0.89050 0.03270 0.07408 0.00272 background ω 0.10229 1.00000 0.10229 1.00000 foreground ω 0.10229 1.00000 1.00000 1.00000	-2930,65466
Branch-site_Null		site class 0 1 2a 2b proportion 0.89050 0.03270 0.07408 0.00272 background ω 0.10229 1.00000 0.10229 1.00000 foreground ω 0.10229 1.00000 1.00000 1.00000	-2930,65466
Clades_Alternative	1	site class 0 1 2 proportion 0.89289 0.08277 0.02434 branch type 0: 0.10418 1.00000 0.00000 branch type 1: 0.10418 1.00000 26.29186	-2931,57908
Clades_Null		p: 0.89431 0.03905 0.06664 ω: 0.10341 1.00000 1.00000	-2932,22443

Tab. 4 Models used, and parameters estimated under different conditions for the evolutionary analysis of the orchid *RAD*-like coding regions.

Model Comparison	LRT	df	p	
One_vs_Two_ratios	1775,29277	1	< 0.00001	
Sites_Null_vs_Sites_Alternative	-0,0006	2	0.9997	ns
Null_vs_Branch-site_Alternative	0	1	1	ns
Null_vs_Clades_Alternative	1,2907	2	0.524479	ns

Tab. 5 Likelihood ratio statistics for the comparison of the evolutionary models of the orchid *RAD*-like coding regions. df, degrees of freedom; ns, not significant.

Model	**Foreground branch**	**ω_0**	**ω_1**	**ω_2**	**lnL**	**# parameters**	
one_ratio	//	0,15407	//	//	-25962,02	165	
two-ratios	A	0,11796	0,23284	//	-25899,96	166	
two-ratios	AGL6	0,16162	0,13553	//	-25958,58	166	
two-ratios	E	0,18707	0,10409	//	-25919,45	166	
three-ratios	A *(1)* AGL6 *(2)*	**0,1052**	**0,2327**	**0,1353**	-25894,84	167	
Site class	**0**	**1**	**2a**	**2b**	**lnL**	**# parameters**	**Positively selected sites**
%	60,245	6,789	29,628	3,339			
alternative_A_back_ω	0,1093	1	0,1093	1	-25506,04	168	yes
alternative_A_fore_ω	0,1093	1	1	1			
null_A_back_ω	0,1093	1	0,1093	1	-25506,04	167	
null_A_fore_ω	0,1093	1	1	1			
% site	73,021	19,147	6,205	1,627			
alternative_AGL6_back_ω	0,12294	1	0,12294	1	-25646,90	168	yes
alternative_AGL6_fore_ω	0,12294	1	1	1			
null_AGL6_back_ω	0,12294	1	0,12294	1	-25646,90	167	
null_AGL6_fore_ω	0,12294	1	1	1			
% site	75,533	18,282	4,98	1,205			
alternative_E_back_ω	0,12467	1	0,1246	1	-25643,72	168	yes
alternative_E_fore_ω	0,12467	1	1	1			
null_E_back_ω	0,12467	1	0,1246	1	-25643,72	167	
null_E_fore_ω	0,12467	1	1	1			

Tab. 6 Models used, and parameters estimated under different conditions for the evolutionary analysis of the orchid *AGl6*, A-class and E-class MADS-box coding regions.

Model Comparison	LRT	df	p-value
one_ratio_vs_foreground_A	124,112	1	**< 0,00001**
one_ratio_vs_foreground_AGL6	6,88317	1	**0,008702**
one_ratio_vs_foreground_E	42,56812	1	**< 0,00001**
one_ratio_vs_three_ratios	134,362	2	**< 0,00001**
foreground_A_vs_three_ratios	10,2507	1	**0,001367**
foreground_AGL6_vs_three_ratios	127,479	1	**< 0,00001**
foreground_E_vs_three_ratios	49,2262	1	**< 0,00001**

Tab. 7 Likelihood ratio statistics for the comparison of the evolutionary models of the orchid *AGL6*, A-class and E-class coding regions. df, degrees of freedom; ns, not significant.

Sub-family	Transcript_name	Forward	Reverse
AGL6-like	*Oita_comp1386_AGL6*	ACTCTCCGTTCTCTGCGACGC	TCCGATCAAACAGCCCATCCCG
	Oita_comp4335_AGL6	GCTTGGAGAGGATCTTGGACCCC	ACAACAGCAAGCTCAGTCCAGA
	Oita_comp8204_AGL6	CACCTTCTCCAAGCGCAGGA	CATCAGGGGCCCAGGATCTG
MADS-box	*Oita_comp6109_MIKC**	TGATTGCGCCGCGTGAGTCT	CGTCGGCTCGTACTATTCCGTCG
	Oita_comp27839_SOC	CGCCGCTTCGCCGGAAATAG	GCCCGGTCTCCACGTCCTTC
	Oita_comp18466_SVP	TCGCAGCAAAGCCATCACTTGA	CCGCTTCCCTCCATCCCCAG
	Oita_comp33336_AGL12	GCAGGTTGGTCAAGTCGCAG	AGACATGTTAACCCGAACCGC
	Oita_comp46983_ANR	CTGAGGTACATGCTGGGACA	TTGCCTTCAGTACTCCCTCCT
	Oita_comp10166_OsMADS32	GCTGGTGATTGTTTTGAAGGCT	CGCCCTCGTGAATAGAAGAGCCT
	Oita_comp3679_AP1	TCGCCAGCCTCTTCGTTGAT	AAACGAGAGAAGCCTGCGGT
	Oita_comp9283_AP1	ACCCCTCCCTGCTCCATCGA	AAGCGTGGAAAGAAGGAGAAAA
	Oita_comp11046_AP1	CGACGGAGAATGGGGAGGGG	ACATGCACGGGATAATAAGAAAC
	Oita_comp2508_AP1	TCACAGGTTGATTGGCGTCT	TCATATGCGTCTTTCTCGACCTT
	Oita_comp1784_AG	ATTCTCACCAGCAGCAGACT	AGCAAGCCTACACAATTAAGCCT
	Oita_comp16674_AG	TGCCTTCGTTTGATTCCCGA	AGACATGTTAACCCGAACCGC
	Oita_comp7010_SEP	TGGGGATCGGATCGGAGGAGG	ATGTGGGGTGCATGTTGACCA
	Oita_comp1006_SEP	AGGGCAATGAGGGGCAAGGG	ACAAACATGTCAACGCGTGA
	Oita_comp16614_Mbeta	GGCCGTGCGAGGGTTGAGAT	TCCCCTCCAGAGTCATTAGGCCA
	Oita_comp4419_Malpha	CGAAGGGTTTTGCAGCGAAA	CCGAATCAATCCTCCACCGG
	Oita_comp13259_Malpha	TGGGCGTGGTTCAGATCGATGA	AGCGATGAACAGCTGCCCCA
	Oita_comp18391_Mgamma	CTCTCCTGTTGACACCGTCC	CCGCCGGGTCAAAATCATCC

Tab. 8 Sequences of the primers (5'-3') used to amplify the MADS-box transcripts expressed in the inflorescence of *O. italica*.

Sub-family	Transcript_name	Forward	Reverse
AGL6-like	*Oita_comp1386_AGL6*	CGGGATGGGCTGTTTGATCGGA	CCACCCAATTAATCGAGCCCCA
	Oita_comp4335_AGL6	ACAGCAATTCCCAGGACTCCT	ACAACAGCAAGCTCAGTCCAGA
	Oita_comp8204_AGL6	GCTACAGCGATTCCAGAGGC	CATCAGGGGCCCAGGATCTG
MADS-box	*Oita_comp6109_MIKC**	GCCCAACCTTACACTCAGTCA	CGTCGGCTCGTACTATTCCGTCG
	Oita_comp27839_SOC	CGAACAGAGCTTAGCAAAAGTCA	GCCCGGTCTCCACGTCCTTC
	Oita_comp18466_SVP	TGGAAACCCACTCAAGAACGA	CCGCTTCCCTCCATCCCCAG
	Oita_comp33336_AGL12	TGCCTTCGTTTGATTCCCGA	AGACATGTTAACCCGAACCGC
	Oita_comp46983_ANR	CTGAGGTACATGCTGGGACA	TTGCCTTCAGTACTCCCTCCT
	Oita_comp10166_OsMADS32	GCTGGTGATTGTTTTGAAGGCT	GCTGGTGATTGTTTTGAAGGCT
	Oita_comp3679_AP1	CCTTGTCACCCACCTTCCAA	AAACGAGAGAAGCCTGCGGT
	Oita_comp9283_AP1	GATGCTTCACTGCTCCACCT	AAGCGTGGAAAGAAGGAGAAAA
	Oita_comp11046_AP1	TCTTCCTACTCCCACCAGCA	ACATGCACGGGATAATAAGAAAC
	Oita_comp2508_AP1	TCACAGGTTGATTGGCGTCT	TCATATGCGTCTTTCTCGACCTT
	Oita_comp1784_AG	ATTCTCACCAGCAGCAGACT	AGCAAGCCTACACAATTAAGCCT
	Oita_comp16674_AG	TGCCTTCGTTTGATTCCCGA	AGACATGTTAACCCGAACCGC
	Oita_comp7010_SEP	CGGGATGGATGTGATGATTTCA	ACATACACAGCTGCAGAGGG
	Oita_comp1006_SEP	GATCGGACCAGTAGCTGCTC	TCTGAGGGACATTTCTAGTTGGT
	Oita_comp16614_Mbeta	TGCAGGAACAACGCTCAGAT	TGAAGAATATGCCGAGCCACT
	Oita_comp4419_Malpha	CGAAGGGTTTTGCAGCGAAA	CCGAATCAATCCTCCACCGG
	Oita_comp13259_Malpha	CTCTCCTGTTGACACCGTCC	CCGCCGGGTCAAAATCATCC
	Oita_comp18391_Mgamma	GAAGCCGATAAGAGCCCCTC	AGCGATGAACAGCTGCCCCA

Tab. 9 Sequences of the primers (5'-3') used to perform the Real Time PCR amplifications of the MADS-box genes of *O. italica*.

Transcript_name	Forward	Reverse
OITA_PI_Y2H	CCGGAATTCCGGATGGGGCGGGGAAATACG	AGGATCCAATCCCACAAAACGCCGTAAAA
OITA_PI2_Y2H	CCGGAATTCCGGATGGGGCGGGGAAAGATAGA	AGGATCCAATTTTCCACAGCCCAGCAGT
OITA_DEF1_Y2H	CCGGAATTCCGGATGGGGAGGGGGAAGATAGAGA	AGGATCCAAAAGGAAAAGTGTTCTCTGGCAC
OITA_DEF2_Y2H	CCGGAATTCCGGATGGGGAGGGGGAAGATAGAGA	AGGATCCAAGCACTGCATCTATCAACTGACCA
OITA_DEF3_Y2H	CCGGAATTCCGGATGGGGAGGGGGAAGATAGAGA	AGGATCCAATGAAGGAAATGTGTGGTTGATCA
OITA_DEF4_Y2H	CCGGAATTCCGGATGGGGAGGGGGAAGATAGAGA	AGGATCCAAAGAACTTACAAATTATTAGGGCA
OITA_DIV9548_Y2H	GGATCCAAATGATGACAAACTCGTGGATG	AACTGCAGCTCACACTCCGACACCAACA
OITA_RAD56510_Y2H	AGGATCCAAATGGCGTCAAGGAACTGTG	AACTGCAGCCGCCGCCTTCTTTAGTT
OITA_DRIF1_Y2H	AGGATCCAAATGGGAACGAATGTGGACCT	AACTGCAGTCCAAAACAGATTGTGGAGGTG

Tab. 10 Sequences of the primers (5'-3') used to amplify the ORFs of the MADS-box and MYB transcripts of *O. italica* for the "Two hybrid system assay".

Bibliography

1. Darwin C: **The Different Forms of Flowers on Plants of the Same Species.** *D Appleton, New York* 1903.
2. Chaw SM, Chang CC, Chen HL, Li WH: **Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes**. *Journal of molecular evolution* 2004, **58**(4):424-441.
3. Crane PR, Friis, EM, and Pedersen, KR: **The origin and early diversification of angiosperm.** *Nature* 1995, **353**:31–37.
4. Davies TJ, Barraclough TG, Chase MW, Soltis PS, Soltis DE, Savolainen V: **Darwin's abominable mystery: Insights from a supertree of the angiosperms**. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(7):1904-1909.
5. Soltis PS, Soltis DE, Savolainen V, Crane PR, Barraclough TG: **Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(7):4430-4435.
6. Aceto S, Gaudio L: **The MADS and the Beauty: Genes Involved in the Development of Orchid Flowers**. *Current genomics* 2011, **12**(5):342-356.
7. Cameron KM, Chase MW, Whitten WM, Kores PJ, Jarrell DC, Albert VA, Yukawa T, Hills HG, Goldman DH: **A phylogenetic analysis of the Orchidaceae: evidence from rbcL nucleotide**. *American journal of botany* 1999, **86**(2):208-224.
8. Gorniak M, Paun O, Chase MW: **Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, Xdh: Congruence with organellar and nuclear ribosomal DNA results**. *Molecular phylogenetics and evolution* 2010, **56**(2):784-795.
9. Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, Iles WJ, Clements MA, Arroyo MT, Leebens-Mack J *et al*: **Orchid phylogenomics and multiple drivers of their extraordinary diversification**. *Proceedings Biological sciences* 2015, **282**(1814).
10. Rudall PJ, Bateman RM: **Roles of synorganisation, zygomorphy and heterotopy in floral evolution: the gynostemium and labellum of orchids and other lilioid monocots**. *Biological reviews of the Cambridge Philosophical Society* 2002, **77**(3):403-441.
11. Dressler R: **Phylogeny and classification of the orchid family.** *Portland, OR: Discorides Press* 1993.
12. Arditti J: **Resupination.** *Natural History Publications,* 2002.:111 –121.
13. Mondragon-Palomino M, Theissen G: **Why are orchid flowers so diverse? Reduction of evolutionary constraints by paralogues of class B floral homeotic genes**. *Annals of botany* 2009, **104**(3):583-594.
14. 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 biology and evolution* 2017, **9**(6).
15. Crepet WL: **Timing in the evolution of derived floral characters: Upper Cretaceous (Turonian) taxa with tricolpate and tricolpate-derived pollen.** *Rev Palaeobot Palynol* 1996, **90**:339–359.
16. Crepet WL, Niklas KJ: **Darwin's second 'abominable mystery': Why are there so many angiosperm species?** *American journal of botany* 2009, **96**(1):366-381.
17. Neal PR, Dafni, A, Giurfa, M: **Floral symmetry and its role in plant-pollinator systems: terminology, distribution, and hypotheses.** *Annu Rev Ecol Syst* 1998, **29**:345–373.
18. Hileman LC: **Trends in flower symmetry evolution revealed through phylogenetic and developmental genetic advances**. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2014, **369**(1648).
19. Doebley J, Stec A, Hubbard L: **The evolution of apical dominance in maize**. *Nature* 1997, **386**(6624):485-488.
20. Luo D, Carpenter R, Vincent C, Copsey L, Coen E: **Origin of floral asymmetry in Antirrhinum**. *Nature* 1996, **383**(6603):794-799.
21. Kosugi S, Ohashi Y: **PCF1 and PCF2 specifically bind to cis elements in the rice proliferating cell nuclear antigen gene**. *The Plant cell* 1997, **9**(9):1607-1619.

22. Cubas P LN, Doebley J, Coen E: **The TCP domain: a motif found in proteins regulating plantgrowth and development**. *The Plant Journal* 1999, **18**(2):215–222.
23. Lupas A: **Coiled coils: new structures and new functions**. *Trends in biochemical sciences* 1996, **21**(10):375-382.
24. Lupas A, Van Dyke M, Stock J: **Predicting coiled coils from protein sequences**. *Science* 1991, **252**(5009):1162-1164.
25. Navaud O, Dabos P, Carnus E, Tremousaygue D, Herve C: **TCP transcription factors predate the emergence of land plants**. *Journal of molecular evolution* 2007, **65**(1):23-33.
26. Cubas P: **Role of TCP genes in the evolution of morphological characters in angiosperms**. *In QCB Cronk, RM Bateman, JA Hawkins, eds, Developmental Genetics and Plant Evolution, Special Volume Series 65 The Systematics Association, London,* 2009:247–266.
27. Costa MM, Fox S, Hanna AI, Baxter C, Coen E: **Evolution of regulatory interactions controlling floral asymmetry**. *Development* 2005, **132**(22):5093-5101.
28. Kosugi S, Ohashi Y: **DNA binding and dimerization specificity and potential targets for the TCP protein family**. *The Plant journal : for cell and molecular biology* 2002, **30**(3):337-348.
29. Masuda HP CL, De Veylder L, Tanurdzic M, de Almeida Engler J, Geelen D, Inzé D, Martienssen RA, Ferreira PC, Hemerly AS: **ABAP1 is a novel plant Armadillo BTB protein involved in DNA replication and transcription.** *EMBO J* 2008, **27**(20):2746-2756.
30. Tatematsu K, Nakabayashi K, Kamiya Y, Nambara E: **Transcription factor AtTCP14 regulates embryonic growth potential during seed germination in Arabidopsis thaliana**. *The Plant journal : for cell and molecular biology* 2008, **53**(1):42-52.
31. Tremousaygue D, Garnier L, Bardet C, Dabos P, Herve C, Lescure B: **Internal telomeric repeats and 'TCP domain' protein-binding sites co-operate to regulate gene expression in Arabidopsis thaliana cycling cells**. *The Plant journal : for cell and molecular biology* 2003, **33**(6):957-966.
32. Howarth DG, Donoghue MJ: **Phylogenetic analysis of the "ECE" (CYC/TB1) clade reveals duplications predating the core eudicots**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(24):9101-9106.
33. Aggarwal P, Das Gupta M, Joseph AP, Chatterjee N, Srinivasan N, Nath U: **Identification of specific DNA binding residues in the TCP family of transcription factors in Arabidopsis**. *The Plant cell* 2010, **22**(4):1174-1189.
34. Preston JC, Hileman LC: **Developmental genetics of floral symmetry evolution**. *Trends in plant science* 2009, **14**(3):147-154.
35. Luo D, Carpenter R, Copsey L, Vincent C, Clark J, Coen E: **Control of organ asymmetry in flowers of Antirrhinum**. *Cell* 1999, **99**(4):367-376.
36. Kranz H, Scholz K, Weisshaar B: **c-MYB oncogene-like genes encoding three MYB repeats occur in all major plant lineages**. *The Plant journal : for cell and molecular biology* 2000, **21**(2):231-235.
37. Lipsick JS: **One billion years of Myb**. *Oncogene* 1996, **13**(2):223-235.
38. Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H: **The regulatory c1 locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators**. *EMBO J* 1987, **6**(12):3553-3558.
39. Feller A, Machemer K, Braun EL, Grotewold E: **Evolutionary and comparative analysis of MYB and bHLH plant transcription factors**. *The Plant journal : for cell and molecular biology* 2011, **66**(1):94-116.
40. Kanei-Ishii C, Sarai A, Sawazaki T, Nakagoshi H, He DN, Ogata K, Nishimura Y, Ishii S: **The tryptophan cluster: a hypothetical structure of the DNA-binding domain of the myb protooncogene product**. *The Journal of biological chemistry* 1990, **265**(32):19990-19995.
41. Ogata K, Kanei-Ishii C, Sasaki M, Hatanaka H, Nagadoi A, Enari M, Nakamura H, Nishimura Y, Ishii S, Sarai A: **The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation**. *Nature structural biology* 1996, **3**(2):178-187.
42. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L: **MYB transcription factors in Arabidopsis**. *Trends in plant science* 2010, **15**(10):573-581.

43. Galego L, Almeida J: **Role of DIVARICATA in the control of dorsoventral asymmetry in Antirrhinum flowers**. *Genes & development* 2002, **16**(7):880-891.
44. Almeida J, Rocheta M, Galego L: **Genetic control of flower shape in Antirrhinum majus**. *Development* 1997, **124**(7):1387-1392.
45. Corley SB, Carpenter R, Copsey L, Coen E: **Floral asymmetry involves an interplay between TCP and MYB transcription factors in Antirrhinum**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(14):5068-5073.
46. Raimundo J, Sobral R, Bailey P, Azevedo H, Galego L, Almeida J, Coen E, Costa MM: **A subcellular tug of war involving three MYB-like proteins underlies a molecular antagonism in Antirrhinum flower asymmetry**. *The Plant journal : for cell and molecular biology* 2013, **75**(4):527-538.
47. Machemer K, Shaiman O, Salts Y, Shabtai S, Sobolev I, Belausov E, Grotewold E, Barg R: **Interplay of MYB factors in differential cell expansion, and consequences for tomato fruit development**. *The Plant journal : for cell and molecular biology* 2011, **68**(2):337-350.
48. 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**(7):e102155.
49. Zhou Y, Xu Z, Zhao K, Yang W, Cheng T, Wang J, Zhang Q: **Genome-Wide Identification, Characterization and Expression Analysis of the TCP Gene Family in Prunus mume**. *Frontiers in plant science* 2016, **7**:1301.
50. Sedeek KEM, Qi WH, Schauer MA, Gupta AK, Poveda L, Xu SQ, Liu ZJ, Grossniklaus U, Schiestl FP, Schluter PM: **Transcriptome and Proteome Data Reveal Candidate Genes for Pollinator Attraction in Sexually Deceptive Orchids**. *PloS one* 2013, **8**(5).
51. Cai J, Liu X, Vanneste K, Proost S, Tsai WC, Liu KW, Chen LJ, He Y, Xu Q, Bian C *et al*: **The genome sequence of the orchid Phalaenopsis equestris**. *Nature genetics* 2015, **47**(1):65-72.
52. Zhang GQ, Xu Q, Bian C, Tsai WC, Yeh CM, Liu KW, Yoshida K, Zhang LS, Chang SB, Chen F *et al*: **The Dendrobium catenatum Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution**. *Sci Rep* 2016, **6**:19029.
53. Howarth DG, Donoghue MJ: **Duplications and expression of DIVARICATA-like genes in dipsacales**. *Molecular biology and evolution* 2009, **26**(6):1245-1258.
54. Baxter CE, Costa MM, Coen ES: **Diversification and co-option of RAD-like genes in the evolution of floral asymmetry**. *The Plant journal : for cell and molecular biology* 2007, **52**(1):105-113.
55. Guo-Qiang Zhang QX, Chao Bian, Wen-Chieh Tsai, Chuan-Ming Yeh, Ke-Wei Liu, Kouki Yoshida, Liang-Sheng Zhang, Song-Bin Chang, Fei Chen, Yu Shi, Yong-Yu Su, Yong-Qiang Zhang, Li-Jun Chen, Yayi Yin, Min Lin, Huixia Huang, Hua Deng, Zhi-Wen Wang, Shi-Lin Zhu, Xiang Zhao, Cao Deng, Shan-Ce Niu, Jie Huang, Meina Wang, Guo-Hui Liu, Hai-Jun Yang, Xin-Ju Xiao, Yu-Yun Hsiao, Wan-Lin Wu, You-Yi Chen, Nobutaka Mitsuda, Masaru Ohme-Takagi, Yi-Bo Luo, Yves Van de Peer & Zhong-Jian Liu: **The Dendrobium catenatum Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution**. *Scientific Reports* 2015, **6**.
56. Feschotte C: **Transposable elements and the evolution of regulatory networks**. *Nature reviews Genetics* 2008, **9**(5):397-405.
57. Raimundo J, Sobral R, Laranjeira S, Costa MMR: **Successive domain rearrangements underlie the evolution of a regulatory module controlled by a small interfering peptide**. *Molecular biology and evolution* 2018.
58. Hsu HF, Hsieh WP, Chen MK, Chang YY, Yang CH: **C/D Class MADS Box Genes from Two Monocots, Orchid (Oncidium Gower Ramsey) and Lily (Lilium longiflorum), Exhibit Different Effects on Floral Transition and Formation in Arabidopsis thaliana**. *Plant and Cell Physiology* 2010, **51**(6):1029-1045.
59. 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'**. *The Plant journal : for cell and molecular biology* 2011, **66**(6):1008-1019.
60. Lin CS, Hsu CT, Liao DC, Chang WJ, Chou ML, Huang YT, Chen JJ, Ko SS, Chan MT, Shih MC: **Transcriptome-wide analysis of the MADS-box gene family in the orchid Erycina pusilla**. *Plant biotechnology journal* 2016, **14**(1):284-298.

61. Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H: **Genetic Control of Flower Development by Homeotic Genes in Antirrhinum majus**. *Science* 1990, **250**(4983):931-936.
62. Riechmann JL, Meyerowitz EM: **MADS domain proteins in plant development**. *Biological chemistry* 1997, **378**(10):1079-1101.
63. Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, Vergara-Silva F, Yanofsky MF: **MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes**. *The Plant journal : for cell and molecular biology* 2000, **24**(4):457-466.
64. Gramzow L, Ritz MS, Theissen G: **On the origin of MADS-domain transcription factors**. *Trends in genetics : TIG* 2010, **26**(4):149-153.
65. Parenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B *et al*: **Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world**. *The Plant cell* 2003, **15**(7):1538-1551.
66. Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM: **The emerging importance of type I MADS box transcription factors for plant reproduction**. *The Plant cell* 2011, **23**(3):865-872.
67. Dolan JW, Fields S: **Cell-type-specific transcription in yeast**. *Biochimica et biophysica acta* 1991, **1088**(2):155-169.
68. Treisman R: **The serum response element**. *Trends in biochemical sciences* 1992, **17**(10):423-426.
69. Tilly JJ, Allen DW, Jack T: **The CArG boxes in the promoter of the Arabidopsis floral organ identity gene APETALA3 mediate diverse regulatory effects**. *Development* 1998, **125**(9):1647-1657.
70. Riechmann JL, Krizek BA, Meyerowitz EM: **Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS**. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(10):4793-4798.
71. Kaufmann K, Melzer R, Theissen G: **MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants**. *Gene* 2005, **347**(2):183-198.
72. Henschel K, Kofuji R, Hasebe M, Saedler H, Munster T, Theissen G: **Two ancient classes of MIKC-type MADS-box genes are present in the moss Physcomitrella patens**. *Molecular biology and evolution* 2002, **19**(6):801-814.
73. Zobell O, Faigl W, Saedler H, Munster T: **MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants**. *Molecular biology and evolution* 2010, **27**(5):1201-1211.
74. Theissen G: **Development of floral organ identity: stories from the MADS house**. *Current opinion in plant biology* 2001, **4**(1):75-85.
75. Theissen G, Saedler H: **Plant biology. Floral quartets**. *Nature* 2001, **409**(6819):469-471.
76. Bowman JL, Smyth DR, Meyerowitz EM: **Genetic interactions among floral homeotic genes of Arabidopsis**. *Development* 1991, **112**(1):1-20.
77. Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF: **B and C floral organ identity functions require SEPALLATA MADS-box genes**. *Nature* 2000, **405**(6783):200-203.
78. Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF: **The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity**. *Current biology : CB* 2004, **14**(21):1935-1940.
79. Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, Yanofsky MF: **Assessing the redundancy of MADS-box genes during carpel and ovule development**. *Nature* 2003, **424**(6944):85-88.
80. Zahn LM, Kong H, Leebens-Mack JH, Kim S, Soltis PS, Landherr LL, Soltis DE, Depamphilis CW, Ma H: **The evolution of the SEPALLATA subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history**. *Genetics* 2005, **169**(4):2209-2223.
81. Coen ES, Meyerowitz EM: **The war of the whorls: genetic interactions controlling flower development**. *Nature* 1991, **353**(6339):31-37.
82. Krizek BA, Fletcher JC: **Molecular mechanisms of flower development: an armchair guide**. *Nature reviews Genetics* 2005, **6**(9):688-698.
83. Ng. M, Yanofsky MF: **Function and evolution of the plant MADS-box gene family**. *Nature reviews Genetics* 2001, **2**(3):186-195.
84. Liu Z, Mara C: **Regulatory mechanisms for floral homeotic gene expression**. *Seminars in cell & developmental biology* 2010, **21**(1):80-86.

85. Ambrose BA, Lerner DR, Ciceri P, Padilla CM, Yanofsky MF, Schmidt RJ: **Molecular and genetic analyses of the silky1 gene reveal conservation in floral organ specification between eudicots and monocots**. *Molecular cell* 2000, **5**(3):569-579.
86. Whipple CJ, Ciceri P, Padilla CM, Ambrose BA, Bandong SL, Schmidt RJ: **Conservation of B-class floral homeotic gene function between maize and Arabidopsis**. *Development* 2004, **131**(24):6083-6091.
87. Whipple CJ, Zanis MJ, Kellogg EA, Schmidt RJ: **Conservation of B class gene expression in the second whorl of a basal grass and outgroups links the origin of lodicules and petals**. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(3):1081-1086.
88. Ferrario S, Immink RG, Angenent GC: **Conservation and diversity in flower land**. *Current opinion in plant biology* 2004, **7**(1):84-91.
89. Irish VF, Litt A: **Flower development and evolution: gene duplication, diversification and redeployment**. *Current opinion in genetics & development* 2005, **15**(4):454-460.
90. Kanno A, Nakada M, Akita Y, Hirai M: **Class B gene expression and the modified ABC model in nongrass monocots**. *TheScientificWorldJournal* 2007, **7**:268-279.
91. Cantone C, Sica M, Gaudio L, Aceto S: **The OrcPI locus: genomic organization, expression pattern, and noncoding regions variability in Orchis italica (Orchidaceae) and related species**. *Gene* 2009, **434**(1-2):9-15.
92. Salemme M, Sica M, Gaudio L, Aceto S: **Expression pattern of two paralogs of the PI/GLO-like locus during Orchis italica (Orchidaceae, Orchidinae) flower development**. *Development genes and evolution* 2011, **221**(4):241-246.
93. Salemme M, Sica M, Iazzetti G, Gaudio L, Aceto S: **The AP2-like gene OitaAP2 is alternatively spliced and differentially expressed in inflorescence and vegetative tissues of the orchid Orchis italica**. *PloS one* 2013, **8**(10):e77454.
94. Chanderbali AS, Berger BA, Howarth DG, Soltis PS, Soltis DE: **Evolving Ideas on the Origin and Evolution of Flowers: New Perspectives in the Genomic Era**. *Genetics* 2016, **202**(4):1255-1265.
95. Aceto S, Sica M, De Paolo S, D'Argenio V, Cantiello P, Salvatore F, Gaudio L: **The analysis of the inflorescence miRNome of the orchid Orchis italica reveals a DEF-like MADS-box gene as a new miRNA target**. *PloS one* 2014, **9**(5):e97839.
96. Acri-Nunes-Miranda R, Mondragon-Palomino M: **Expression of paralogous SEP-, FUL-, AG- and STK-like MADS-box genes in wild-type and peloric Phalaenopsis flowers**. *Frontiers in plant science* 2014, **5**:76.
97. 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 evolutionary biology* 2009, **9**:81.
98. Pan ZJ, Cheng CC, Tsai WC, Chung MC, Chen WH, Hu JM, Chen HH: **The duplicated B-class MADS-box genes display dualistic characters in orchid floral organ identity and growth**. *Plant & cell physiology* 2011, **52**(9):1515-1531.
99. Kramer EM, Jaramillo MA, Di Stilio VS: **Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms**. *Genetics* 2004, **166**(2):1011-1023.
100. Zahn LM, Leebens-Mack J, DePamphilis CW, Ma H, Theissen G: **To B or Not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms**. *The Journal of heredity* 2005, **96**(3):225-240.
101. Hsu H HW, Lee Y, Mao W, Yang J, Li J, Yang C: **Model for perianth formation in orchids**. *Nat Plants* 2015, **1**.
102. Mondragon-Palomino M, Theissen G: **MADS about the evolution of orchid flowers**. *Trends in plant science* 2008, **13**(2):51-59.
103. Tsai WC, Kuoh CS, Chuang MH, Chen WH, Chen HH: **Four DEF-like MADS box genes displayed distinct floral morphogenetic roles in Phalaenopsis orchid**. *Plant & cell physiology* 2004, **45**(7):831-844.
104. Xu Y, Teo LL, Zhou J, Kumar PP, Yu H: **Floral organ identity genes in the orchid Dendrobium crumenatum**. *The Plant journal : for cell and molecular biology* 2006, **46**(1):54-68.

105. Kim SY YP, Fukuda T, Ochiai T, Yokoyama J, Kameya T, Kanno A: **Expression of a DEFICIENS-like gene correlates with the differentiation between sepal and petal in the orchid, Habenaria radiata (Orchidaceae)**. *Plant Science* 2007, **172**:319-326.
106. Kanno A, Saeki H, Kameya T, Saedler H, Theissen G: **Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (Tulipa gesneriana)**. *Plant Mol Biol* 2003, **52**(4):831-841.
107. Nakamura T, Fukuda T, Nakano M, Hasebe M, Kameya T, Kanno A: **The modified ABC model explains the development of the petaloid perianth of Agapanthus praecox ssp. orientalis (Agapanthaceae) flowers**. *Plant Mol Biol* 2005, **58**(3):435-445.
108. Salemme M, Sica M, Gaudio L, Aceto S: **The OitaAG and OitaSTK genes of the orchid Orchis italica: a comparative analysis with other C- and D-class MADS-box genes**. *Molecular biology reports* 2013, **40**(5):3523-3535.
109. 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**(7672):379-383.
110. Chao YT, Chen WC, Chen CY, Ho HY, Yeh CH, Kuo YT, Su CL, Yen SH, Hsueh HY, Yeh JH *et al*: **Chromosome-level assembly, genetic and physical mapping of Phalaenopsis aphrodite genome provides new insights into species adaptation and resources for orchid breeding**. *Plant biotechnology journal* 2018.
111. Kwantes M, Liebsch D, Verelst W: **How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes**. *Molecular biology and evolution* 2012, **29**(1):293-302.
112. Liu Y, Cui S, Wu F, Yan S, Lin X, Du X, Chong K, Schilling S, Theissen G, Meng Z: **Functional conservation of MIKC*-Type MADS box genes in Arabidopsis and rice pollen maturation**. *The Plant cell* 2013, **25**(4):1288-1303.
113. Tapia-Lopez R, Garcia-Ponce B, Dubrovsky JG, Garay-Arroyo A, Perez-Ruiz RV, Kim SH, Acevedo F, Pelaz S, Alvarez-Buylla ER: **An AGAMOUS-related MADS-box gene, XAL1 (AGL12), regulates root meristem cell proliferation and flowering transition in Arabidopsis**. *Plant physiology* 2008, **146**(3):1182-1192.
114. Zhang H, Forde BG: **An Arabidopsis MADS box gene that controls nutrient-induced changes in root architecture**. *Science* 1998, **279**(5349):407-409.
115. Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, Kapoor S: **MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress**. *BMC genomics* 2007, **8**:242.
116. Zhao T, Ni Z, Dai Y, Yao Y, Nie X, Sun Q: **Characterization and expression of 42 MADS-box genes in wheat (Triticum aestivum L.)**. *Molecular genetics and genomics : MGG* 2006, **276**(4):334-350.
117. Ding L, Wang Y, Yu H: **Overexpression of DOSOC1, an ortholog of Arabidopsis SOC1, promotes flowering in the orchid Dendrobium Chao Parya Smile**. *Plant & cell physiology* 2013, **54**(4):595-608.
118. Liu XR, Pan T, Liang WQ, Gao L, Wang XJ, Li HQ, Liang S: **Overexpression of an Orchid (Dendrobium nobile) SOC1/TM3-Like Ortholog, DnAGL19, in Arabidopsis Regulates HOS1-FT Expression**. *Frontiers in plant science* 2016, **7**:99.
119. Hsing-Fun Hsu W-HH, Yung-I Lee, Wan-Ting Mao, Jun-Yi Yang, Jen-Ying Li, Chang-Hsien Yang: **Model for perianth formation in orchids**. *Nature Plants* 2015.
120. Litt A, Irish VF: **Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: Implications for the evolution of floral development**. *Genetics* 2003, **165**(2):821-833.
121. Kocyan A. QY-L, Endress P. K. , Conti E.: **A phylogenetic analysis of Apostasioideae (Orchidaceae) based on ITS, trnL-F and matK sequences**. *Plant Syst Evol* 2004.
122. Dirks-Mulder A, Butot R, van Schaik P, Wijnands JWPM, 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**.
123. Skipper M, Johansen LB, Pedersen KB, Frederiksen S, Johansen BB: **Cloning and transcription analysis of an AGAMOUS- and SEEDSTICK ortholog in the orchid Dendrobium thyrsiflorum (Reichb. f.)**. *Gene* 2006, **366**(2):266-274.

124. Mitoma M, Kanno A: **The Greenish Flower Phenotype of Habenaria radiata (Orchidaceae) Is Caused by a Mutation in the SEPALLATA-Like MADS-Box Gene HrSEP-1**. *Frontiers in plant science* 2018, **9**.
125. Rebocho AB, Kennaway JR, Bangham JA, Coen E: **Formation and Shaping of the Antirrhinum Flower through Modulation of the CUP Boundary Gene**. *Current biology : CB* 2017, **27**(17):2610-2622 e2613.
126. Javelle M, Marco CF, Timmermans M: **In situ hybridization for the precise localization of transcripts in plants**. *Journal of visualized experiments : JoVE* 2011(57):e3328.
127. Li L, Stoeckert CJ, Jr, Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes**. *Genome research* 2003, **13**(9):2178-2189.
128. Devos N, Szovenyi P, Weston DJ, Rothfels CJ, Johnson MG, Shaw AJ: **Analyses of transcriptome sequences reveal multiple ancient large-scale duplication events in the ancestor of Sphagnopsida (Bryophyta)**. *The New phytologist* 2016, **211**(1):300-318.
129. Ponting CP, Russell RR: **The natural history of protein domains**. *Annual review of biophysics and biomolecular structure* 2002, **31**:45-71.
130. Su CL, Chen WC, Lee AY, Chen CY, Chang YC, Chao YT, Shih MC: **A modified ABCDE model of flowering in orchids based on gene expression profiling studies of the moth orchid Phalaenopsis aphrodite**. *PloS one* 2013, **8**(11):e80462.
131. Kohany O, Gentles AJ, Hankus L, Jurka J: **Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor**. *BMC bioinformatics* 2006, **7**:474.
132. Edgar RC: **MUSCLE: a multiple sequence alignment method with reduced time and space complexity**. *BMC bioinformatics* 2004, **5**:113.
133. Suyama M, Torrents D, Bork P: **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments**. *Nucleic acids research* 2006, **34**(Web Server issue):W609-612.
134. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: **MEGA6: Molecular Evolutionary Genetics Analysis version 6.0**. *Molecular biology and evolution* 2013, **30**(12):2725-2729.
135. 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 research* 2009, **37**(Web Server issue):W202-208.
136. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood**. *Computer applications in the biosciences : CABIOS* 1997, **13**(5):555-556.
137. Yang Z, Bielawski JP: **Statistical methods for detecting molecular adaptation**. *Trends in ecology & evolution* 2000, **15**(12):496-503.
138. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability**. *Molecular biology and evolution* 2013, **30**(4):772-780.
139. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0**. *Systematic biology* 2010, **59**(3):307-321.
140. Gietz RD, Schiestl RH, Willems AR, Woods RA: **Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure**. *Yeast* 1995, **11**(4):355-360.
141. Causier B, Davies B: **Analysing protein-protein interactions with the yeast two-hybrid system**. *Plant molecular biology* 2002, **50**(6):855-870.