UNIVERSITY OF NAPLES FEDERICO II

DEPARTMENT OF BIOLOGY

Ph.D. IN ADVANCED BIOLOGY

SECTION: MOLECULAR SYSTEMATIC

XXV CYCLE

Genes involved in flower development of *Orchis italica* (Orchidaceae)

Tutor

Prof. Serena Aceto

Coordinator

Prof. Luciano Gaudio

Ph.D. Student

Dr. Marinella Salemme

INDEX

ABSTRACT **3**

INTRODUCTION **4**

Flower development 4

The MADS-box gene family 7

The APETALA2/ERF gene family 9

The Orchidaceae 10

The orchid class B MADS-box genes and the "orchid code" 12

The orchid class C and D genes 14

AIMS OF THE STUDY **15**

MATERIALS AND METHODS **17**

Plant material 17

Nucleic acids extraction 17

Isolation of the OitaAP2 cDNA 17

Rapid Amplification of cDNA Ends (5' RACE) 18

Mir172 cleavage assay on the OitaAP2 cDNA 20

Expression analysis 20

PCR amplification of genomic DNA 22

Sequence analysis 23

Construction of microRNA libraries from floral bud and leaf tissue 25

RESULTS AND DISCUSSION **26**

Expression pattern of the two class B paralogs OrcPI and OrcPI2 26

The class C OitaAG and class D OitaSTK cDNAs 30

Expression pattern of the OitaAG and OitaSTK genes 33

Genomic structure of the OitaAG and OitaSTK loci 36

The class A OitaAP2 cDNA 41

Genomic structure of the OitaAP2 locus …… 45
Expression pattern of the OitaAP2 gene …… 47
Comparative expression analysis of the OitaAP2 isoforms, mir172 and OitaAG … 50
High throughput sequencing of microRNA libraries from floral bud and leaf ……. 53

REFERENCES ……. 56

APPENDIX …….. 64
Poster abstracts and scientific publications produced during the PhD course …… 64

ABSTRACT

Il presente lavoro è incentrato sullo studio di cinque geni appartenenti a quattro classi funzionali differenti, coinvolti nello sviluppo del fiore dell'orchidea mediterranea *Orchis italica*. Tali geni codificano fattori trascrizionali, quattro appartenenti alla famiglia MADS-box (*OrcPI*, *OrcPI2*, *OitaAG* e *OitaSTK*) e uno alla famiglia AP2/ERF (*OitaAP2*).

I geni *OrcPI* e *OrcPI2* sono due paraloghi di classe B e, secondo il modello ABCDE dello sviluppo del fiore nelle orchidee, sono coinvolti nella formazione dei tepali e della colonna. L'analisi di espressione di *OrcPI* e *OrcPI2* in differenti tessuti e in vari stadi dello sviluppo lascia ipotizzare una loro possibile sub-funzionalizzazione, con ripartizione delle funzioni tra le due copie paraloghe.

Il gene *OitaAG* appartiene alla classe C ed è coinvolto principalmente nella formazione della colonna, pur essendo espresso anche nell'ovario in maturazione, mentre il gene di classe D *OitaSTK* regola principalmente la formazione degli ovuli, pur essendo espresso anche nella colonna. La caratterizzazione genomica di *OitaAG* e *OitaSTK* ha evidenziato la presenza di introni di dimensioni piuttosto elevate, all'interno dei quali sono presenti numerose tracce di elementi trasponibili. L'introne 2 del gene *OitaAG* contiene numerose sequenze di elementi *cis*-regolativi conservati tra monocotiledoni e dicotiledoni.

Il gene *OitaAP2* appartiene alla classe A ed è coinvolto nella formazione dei tepali e del labello. Un evento di *splicing* alternativo produce due diverse isoforme, *OitaAP2* e *OitaAP2_ISO*. Entrambe le isoforme conservano il sito bersaglio del microRNA miR172 che regola negativamente l'attività del gene *OitaAP2* promuovendo il taglio del trascritto. Nei tessuti del fiore i geni *OitaAP2* e *OitaAP2_ISO* sono espressi nei tepali, nel labello e nell'ovario prima dell'impollinazione. *OitAP2_ISO* viene espresso anche nelle radici e nel fusto, facendo ipotizzare che ciascuna isoforma sia specializzata a svolgere ruoli funzionali differenti nello sviluppo di *O. italica*. Nelle fasi che precedono la fioritura, il profilo di espressione di entrambe le isoforme è sempre complementare a quello di *OitaAG* e a quello di mir172, confermando l'attività inibitoria di OitaAP2 su *OitaAG* e di mir172 su *OitaAP2*. Dopo l'antesi tale equilibrio si rompe e miR172 non viene più espresso nella colonna, lasciando ipotizzare l'esistenza anche di un meccanismo di regolazione negativa di *OitaAP2* alternativo a mir172.

INTRODUCTION

Flower development

The flower is the reproductive structure of the flowering plants (angiosperms). Despite the variety of sizes, shapes and colors, the general flower architecture is quite conserved and consists of four concentric organ types: sepals, petals, stamens and carpels. The development of flower is a multistep process including the formation of a floral meristem and the establishment of unique organs identity, followed by the differentiation of floral structures.

Flower development starts from populations of actively dividing cells named shoot apical meristems and floral meristems. In response to floral inductive cues, the shoot apical meristems that are producing leaves and branches convert to a reproductive fate. In the model plant *Arabidopsis thaliana* two genes, *LEAFY* (*LFY*) and *APETALA1* (*AP1*) are required to specify the floral meristem identity. One of the most dramatic differences between the floral and vegetative shoot apical meristem is that the floral meristem is determinate: it produces a finite set of organs and then stops the proliferation. In *Arabidopsis*, the specification of the floral meristem depends on the early expression of the gene *AGAMOUS* (*AG*) in the apical meristem (*Drews et al., 1991*). This expression pattern depends on an intricate feedback loop among *AG* and two other transcription factors, *LFY* and the homeodomain containing *WUSCHEL* (*WUS*) gene product. The LFY protein is required for the transcription of *AG* in the floral meristem, as *lfy* mutants show both a delayed and reduced domain of *AG* expression (*Weigel and Meyerowitz, 1993*). Together, the WUS and LFY proteins activate the expression of *AG* and both have been shown to bind directly to sequences within the *AG* gene (*Busch et al. 1999; Lenhard et al., 2001; Lohmann et al., 2001*).

In the early 1990s, analyses of homeotic mutants in both *Arabidopsis* and *Antirrhinum* have led to the ABC model of floral organ identity specification (*Bowman et al., 1991; Coen and Meyerowitz 1991; Weigel and Meyerowitz 1994*). More recently, two new functional classes (D and E) were added, resulting in the revised ABCDE model (*Pelaz et al., 2000, 2001; Pinyopich et al; 2003*). Based on this model, the five functional classes A, B, C, D and E, each consisting of one or more genes, determine the floral

organ identities. Function A alone controls the development of sepals in the first (outer) whorl of a flower. Functions A and B together specify petal development in the second whorl; B and C functions together control stamen development in the third whorl; fuction C alone determines carpel development in the fourth whorl. Functions A and C are mutual antagonists and the expression of one represses the other (*Bowman et al., 1991; Drews et al., 1991; Gustafson-Brown et al, 1994*). In the revised model, the function D determines ovule development, while function E is required for development of all floral organs (**Fig. 1A**). The above mentioned mechanism of floral patterning is applicable to many plant species; however, the existence of unique flower structures in some species indicates the ABCDE model is not fully conserved (*Irish and Litt, 2005; Zahn et al; 2005*) (**Fig. 1B**).

In *Arabidopsis*, class A genes include *APETALA1* (*AP1*) and *APETALA2* (*AP2*), class B genes are represented by *APETALA3* (*AP3*) and *PISTILLATA* (*PI*). The *AGAMOUS* (*AG*) and *SEEDSTICK* (*STK*) genes are members of the class C and D, respectively. Class E genes consist of four members, *SEPALLATA1-4* (*SEP1-4*), which show partially redundant functions (*Pelaz et al., 2000; Honma and Goto, 2001; Ditta et al., 2004*). All the known homeotic genes of the ABCDE model, with the exception of the function A gene *AP2*, belong to the MADS-box family of transcription factors, regulators of developmental programs from vegetative growth to flowering process and seed development (*Theissen, 2001*). The molecular interaction of floral organ identity genes is well depicted in the "floral quartet" model (*Honma and Goto, 2001*). The proteins encoded by the ABCDE genes are assembled to form different quaternary complexes, one for each type of floral organ. For example, the complexes AP1/AP1/SEP/SEP, AP1/SEP/AP3/PI, AG/SEP/AP3/PI and AG/AG/SEP/SEP are present within whorls 1, 2, 3 and 4, respectively, to induce the formation of the specific floral organs (**Fig. 2**).

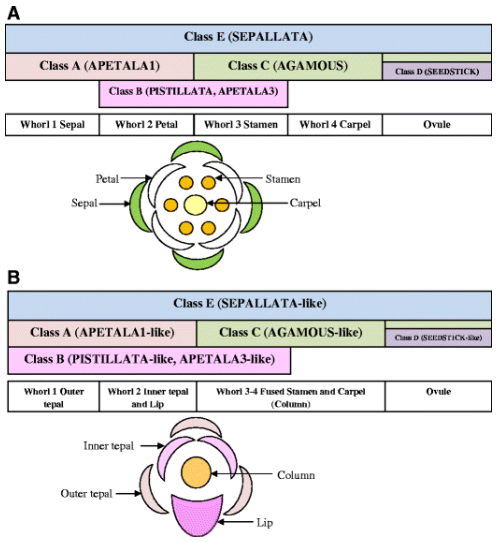


Figure 1: The classical (**A**) and modified (**B**) ABCDE model of floral development (*Salemme et al., 2011*).

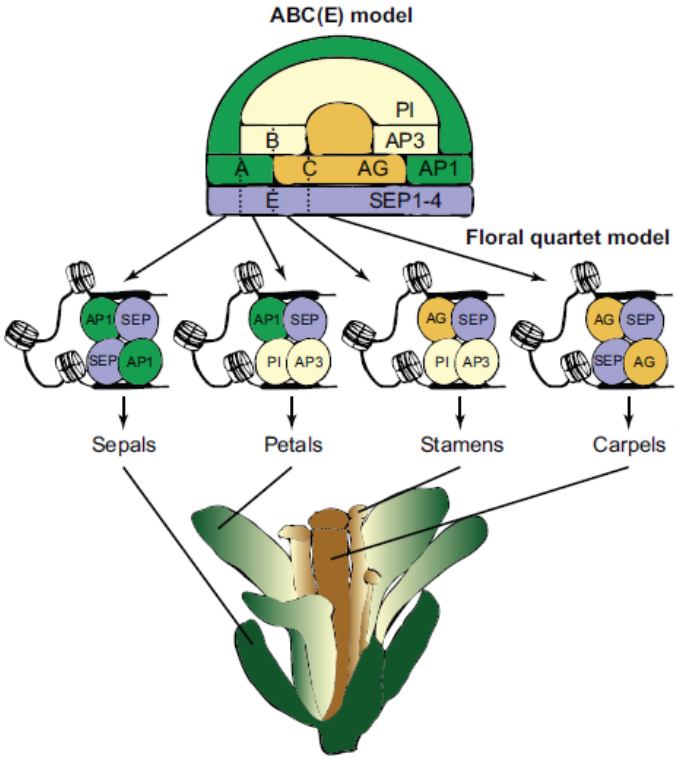


Figure 2: The floral quartet model of floral organ specification (*Smaczniak et al., 2012*).

These protein complexes exert their function by binding to the promoters of target genes which they either activate or repress, as appropriate for the development of the identities of the different floral organs. According to this quartet model, two protein dimers of each tetramer recognize two different DNA sites, the CArG-boxes (consensus sequence 5'-CC(A/T)$_6$ GG-3'), which are brought into close vicinity by DNA bending (*Theissen, 2001; Theissen and Saedler, 2001)* (**Fig. 3**).

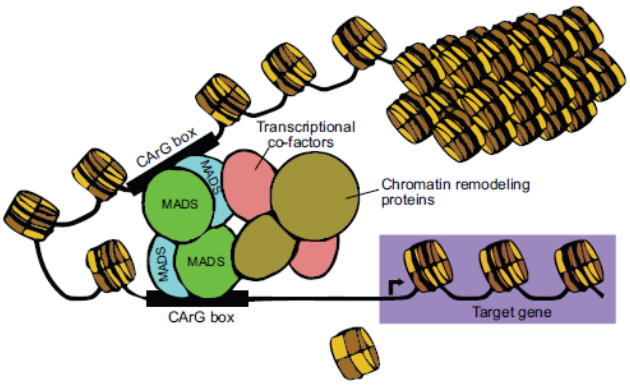


Figure 3: Interaction between the DNA and the MADS-domain (*Smaczniak et al., 2012*).

The MADS-box gene family

The first MADS-box genes discovered as regulators of floral organ identity were *AGAMOUS* (*AG*) from *Arabidopsis thaliana* (*Yanofsky et al., 1990*) and *DEFICENS* (*DEF*) from *Antirrhinum majus* (*Schwarz-Sommer et al., 1990*). The AG and DEF proteins contain a DNA-binding domain of ~60 amino acids that exhibits a striking similarity with the previously characterized proteins serum response factor (SRF) of *Homo sapiens* (*Norman et al., 1988*) and Minichromosome maintenance 1 (Mcm1) of *Saccharomyces cerevisiae* (*Passmore et al., 1988*). This shared and conserved region was named MADS-box (for MCM1, AG, DEF and SRF) and is present in all the MADS-domain transcription factor family members (*Schwarz-Sommer et al., 1990*).

The sequence similarity between the DNA-binding domain of the MADS-box transcription factors and the subunit A of the topoisomerase IIA (TOPOIIA-A) supports the origin of the MADS-box family from the TOPOIIA. A duplication of the TOPOIIA-A might have occurred in the common ancestor of all eukaryotes. After duplication, one

of the copies retained the topoisomerase activity, necessary for processes such as DNA replication, and the other one acquired the DNA-binding capacity, becoming the progenitor of the MADS-box transcription factors (*Gramzow and Theissen, 2010*). Based on the protein domain structure, the MADS-box gene family can be divided into the Type I and Type II lineages, arisen from duplication events occurred in a more recent common ancestor of plants and animals. Type I lineage can be further classified into three subclasses, Mα, Mβ and Mγ, discovered after the completion of the *Arabidopsis* genome sequence. These groups of genes to which no function has been assigned yet are an heterogeneous group sharing only ~180 bp of DNA sequence encoding the MADS-domain (*De Bodt et al., 2003; Kofuji et al., 2003; Parenicovà et al., 2003*).

The floral homeotic genes belong to the MADS-box type II lineage. They have a modular domain organization, the MIKC structure, and contain the N-terminally located DNA-binding MADS-domain, followed by the I- (intervening) and K- (keratin-like) regions, essential for dimerization and higher-order complex formation, and finally the highly variable C-terminal domain, which may have roles in protein-complex formation and transcriptional regulation (*Kaufmann et al., 2005*). Based on the different exon/intron structure, MIKC-type MADS-box genes can be further divided into MIKC^{C} and MIKC* (*Henshel et al., 2002*) (**Fig. 4**). Functional studies have suggested a major specialization of the MIKC* genes in the development of the male gametophyte (*Zobell et al., 2010*), whereas the MIKC^{C} genes, the best characterized group of MADS-box genes (that are often referred to simply as the MIKC genes), are involved in many functions related to plant growth and development and are closely linked to the origin of the floral organs and fruits of angiosperms. The genomic organization of the MIKC^{C} genes is generally conserved, with the presence of seven introns and eight exons (*De Bodt et al., 2003; Gramzow and Theissen, 2010*).

Figure 4: Structure-based classification of the MADS-box proteins (*Smaczniak et al., 2012*).

The MADS-box gene family evolved through genetic duplication events followed by

functional and structural divergences. In particular, after duplication, the evolution of the MADS-box family has occurred through two main different processes, sub-functionalization and neo-functionalization, consisting in a partition of the existing functions or in an the acquisition new ones, respectively (*De Bodt et al., 2006*). There is a strict temporal correlation between the duplication events of the MADS-box genes and relevant events of the angiosperm evolution (*Nei, 2005; Nei and Rooney, 2005; Freeling and Thomas, 2006; Soltis et al., 2007*). In addition, the type II MADS-box transcription factors can be classified in several distinctive subfamilies in which the proteins are often characterized by distinct sequence motifs in their C-terminal domain originated from frame-shift mutations verified during their evolution, some of them necessary for specific protein functions (*Piwarzyk et al., 2007; Benlloch et al., 2009*).

The APETALA2/ERF gene family

The *AP2/ERF* (*APETALA2/Ethylene-Responsive element-binding Factor*) genes define a large family encoding for three different classes of proteins that share the AP2 DNA-binding domain. The AP2-like proteins present two AP2 domains, the ERF-like proteins have only one AP2 domain and the RAV1-2 proteins have two different DNA-binding domains: AP2 and B3 (*Giraudat et al., 1992*). As reported for the MADS-box gene family, also the evolution of the *AP2/ERF* gene family has occurred through duplications followed by sub- and neo-functionalization (*Ohno 1970; Levin 1983; Stebbins 1966; Lynch and Conery 2000*).

According to the ABCDE model, *AP2* is a class A gene. In *Arabidopsis* it is required for perianth identity and repression of the C class gene function. In addition, *AP2* plays several different roles during the plant life cycle, such as the establishment of floral meristem identity (*Irish and Sussex, 1990; Bowman et al., 1993*), the determination of floral organ identity (*Komaki et al., 1988; Jofuku et al., 1994*) and the temporal and spatial regulation of flower homeotic gene expression (*Drews et al., 1991*). As mentioned above, the *AP2* gene belongs to a subfamily of transcription factors characterized by the presence of two DNA-binding motifs (AP2 domains) of 60 and 61 conserved amino-acids, respectively. The structure of the *AP2* gene includes 10 exons

and 9 introns and a highly conserved sequence of 21 nucleotide in the exon 10 that represents the target of the microRNA miR172 (**Fig. 5**). In addition to the two AP2 domains strongly conserved among species, all the AP2-like proteins present two conserved motifs at the N-terminus (motif 1 and 2) followed by a nuclear localizing signal, upstream the AP2 domains, and a motif 3 downstream the AP2 domains (*Gil-Humanes et al., 2009).*

In *Arabidopsis*, the *AP2* mRNA accumulates in the perianth, even though its expression is visible also in reproductive organ primordia. The *AP2* activity is regulated by the microRNA 172 that acts regulating the *AP2* gene expression at post-transcriptional level, predominantly through mRNA cleavage and translational inhibition (*Rhoades et al., 2002; Aukerman and Sakai, 2003; Kasschau et al., 2003; Chen, 2004*). In addition, some evidences show the existence of a negative feedback loop in which *AP2* represses its own transcription (*Aukerman and Sakai, 2003; Chen, 2004; Kasschau et al., 2003; Schwab et al., 2005; Mlotshwa et al., 2006*).

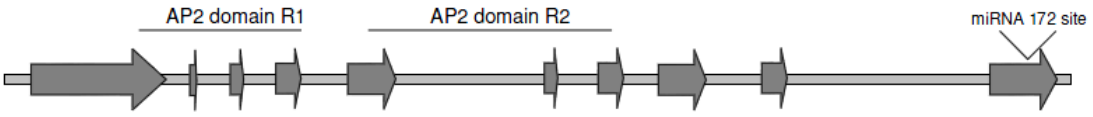


Figure 5: Structure of an *AP2*-like gene (*Gil-Humanes et al., 2009*).

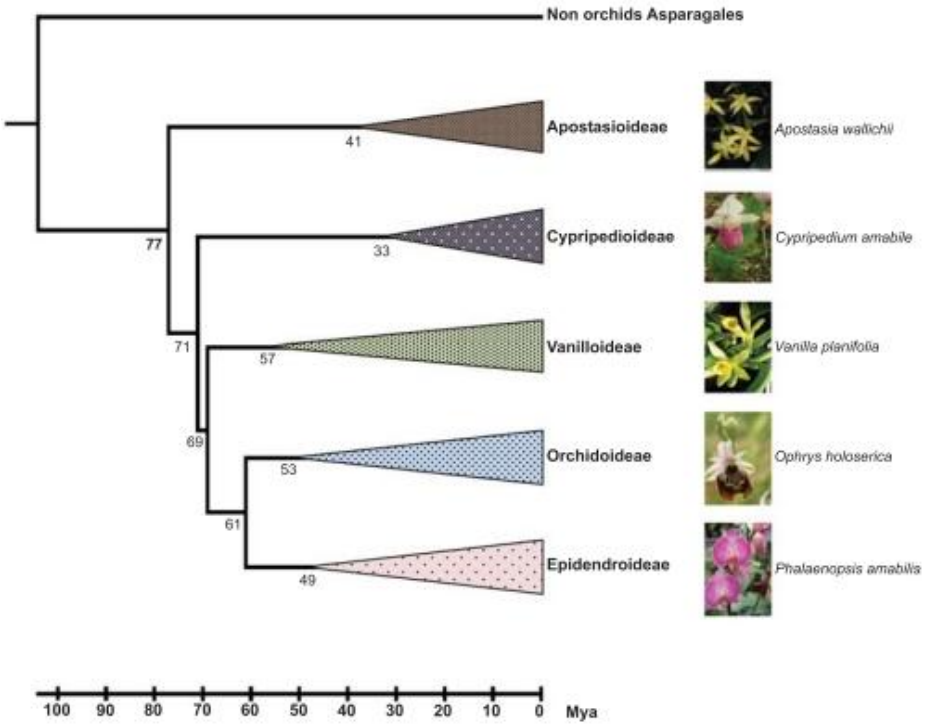
The Orchidaceae

The Orchidaceae family is one of the largest among the angiosperms, including five subfamilies (Apostasioideae, Cypripedioideae, Epidendroideae, Orchidoideae and Vanilloideae) (**Fig. 6A**), each comprising a large number of tribes and subtribes (*Cameron et al., 1999; Cameron et al., 2007*). Orchid species have colonized almost all habitats, showing great morphological and ecological diversification. Epiphytism, highly diversified pollination strategies, natural selection and genetic drift represent the most relevant causes determining the wide species diversity in orchids (*Cozzolino and Widmer, 2005; Tremblay et al., 2005*).

The orchid flower is zygomorphic (bilateral symmetry): it includes three outer tepals (sepals) and three petals distinct into two lateral inner tepals and a highly modified

median tepal (lip or labellum). The orchid male and female reproductive organs are fused in a peculiar structure termed column. At the top of the column is positioned the mature pollen grains (pollinia), while at the base is positioned the ovary (**Fig. 6B**). Another peculiar aspect of the orchids is the ovary development, triggered by fertilization (*Yu and Goh, 2001; Rudall and Bateman, 2002*).

A



B

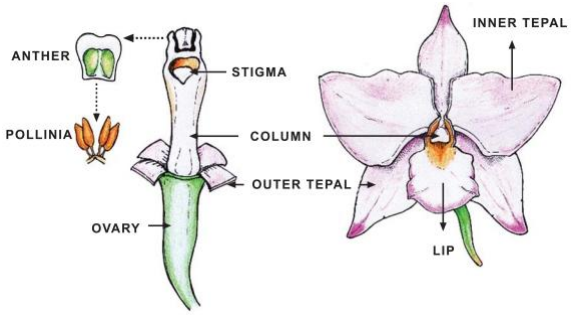


Figure 6: Phylogenetic relationships of the five sub-families of Orchidaceae (**A**) and scheme of an orchid flower (**B**) (*Aceto and Gaudio, 2011*).

The orchid class B MADS-box genes and the "orchid code"

The class B MADS-box genes includes two major gene lineages: the *AP3/DEF*-like (from *APETALA3* of *A. thaliana* and *DEFICIENS* of *A. majus*) and the *PI/GLO*-like (from *PISTILLATA* of *A. thaliana* and *GLOBOSA* of *A. majus*). A duplication in an ancestral gene containing a paleoAP3 motif gave rise to the *PI/GLO*-like lineage (*Kramer et al., 1998; Zhan et al., 2005*), while a second duplication event resulted in the *AP3/DEF*-like lineage that includes the paleoAP3 clade and two further clades, TM6 and euAP3 (*Kramer et al., 1998*).

In orchids, the class B *AP3/DEF*-like genes evolved through a series of duplication events in the coding and regulatory regions followed by sub- and neo-functionalization processes that greatly contributed to the morphological evolution of this plant family (*Tsai et al., 2005; Xu et al., 2006; Mondragòn-Palomino et al., 2009*). These duplications produced four different orchid *AP3/DEF*-like clades, each exhibiting a specific expression pattern. An hetero-dimerization process is required to permit the AP3/DEF-like proteins to interact with class A, C and E MADS-box proteins (reviewed in *Theissen and Melzer, 2007*). According to the classical ABCDE model, these genes are necessary to ensure the correct petals and stamens development (*Kramer et al., 1998; Gramzow and Theissen, 2010*). As described previously, in contrast with the flower morphology showed by *Arabidopsis* and other eudicots (normally represented by sepals in the first whorl and petals in the second whorl), the flower of orchids presents in the first and in the second whorl a series of morphologically similar organs termed tepals (**Fig. 6B**). The modified ABCDE model attributes the presence of sepaloid petals (tepals) to the expansion of the class B genes expression into the first whorl, in addition to the second and the third one (*Yu et al., 1999; Chang et al., 2010*). However, the ABCDE model does not suggest a convincing explanation of the presence in orchids of the greatly specialized median inner tepal, the lip or labellum. In this regard, the recent theory on "the orchid code" suggests that the class B *AP3/DEF*-like genes have an important role into determining tepals and lip identity (*Mondragòn-Palomino et al., 2009*). The experimental data show that in the inner lateral tepals are required high levels of expression of the clade-1 and clade-2 *AP3/DEF*-like genes and low levels of

the clade-3 and clade-4 genes; on the contrary, lip development requires high expression levels of the clade-3 and clade-4 *AP3/DEF*-like genes and low expression levels of the clade-1 and clade-2 genes (**Fig. 7A**). According to these observations, the model proposes that a first duplication event generated the clade-1 and clade-2 genes and the ancestor of the clade-3 and clade-4 genes. In this "intermediate" phase the expression of the ancestor gene of the clade-3 and 4 has been excluded from the outer tepals, generating distinctive outer and inner tepals. A second duplication event gave rise to the differentiated clade-3 and clade-4 genes; at this stage, the restriction of the expression of the clade-4 genes led to the development of the modified median inner tepal (lip).

The first proposal of the "orchid code" model was referred to species belonging to Epidendroideae and Orchidoideae. More recently, the applicability of the "orchid code" has been evaluated also in other species such as *Vanilla planifolia* (Vanilloideae) and *Phragmipedium longifolium* (Cypripedioideae). The results obtained lead to a revised model in which the main difference with "the old" one is represented by the combined differential expression of distinct levels of mRNAs from the four *AP3/DEF*-like genes rather than the presence/absence of a transcript that specifies the organ identity (**Fig. 7B**) (*Mondragòn-Palomino and Theissen,2011*).

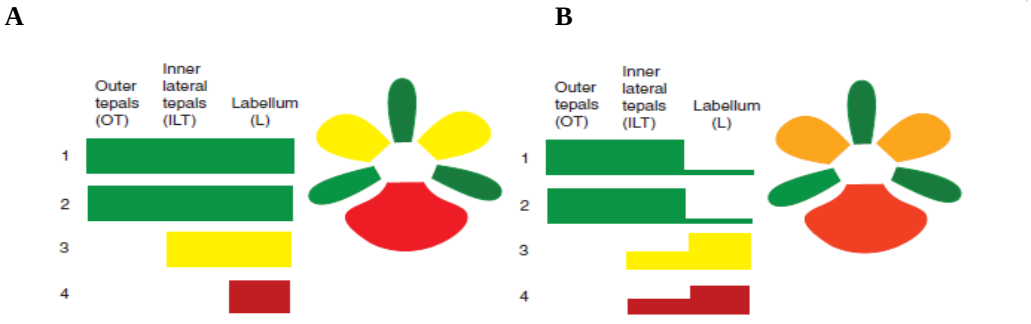


Figure 7: The first proposal (**A**) and the modified version (**B**) of the "orchid code" (*Mondragòn-Palomino and Theissen, 2011*).

In contrast to the *AP3/DEF*-like genes, the orchid class B *PI/GLO*-like genes are documented as single-copy loci in almost all the orchid subfamilies, with the only exception of the tribe Orchidinae where two *PI/GLO*-like paralogs have been described

(*Kim et al., 2007; Cantone et al., 2011*). Evolutionary analyses showed that the two paralogs evolved through different selective constraints acting on synonymous sites, suggesting a possible sub-functionalization of these *PI/GLO*-like genes within Orchidinae (*Aceto et al., 2007; Cantone et al., 2009; Mondragòn-Palomino et al., 2009; Cantone et al., 2011).*

The orchid class C and D genes

According to the ABCDE model, the class C genes are known to regulate, alone, the development of carpels and, together with the class B genes, the development of the stamens. The class D genes are involved in the development of ovules. The class C and D genes are sister clades evolved after an early duplication event during angiosperm evolution (*Kramer et al., 2004*). They encode for transcriptional factors characterized by two motifs, AG motifs I and II that are common to the class C and D genes.

Different class C and D genes have been identified in orchids, in particular in the subfamily Epidendroideae, where they are known to be involved in different aspects of flower and ovule development with just a few differences among the species examined (*Hsu et al., 2010; Chen et al., 2012*). In some species, such as *Phalaenopsis* and *Oncidium*, the class C and D genes show redundant functions (*Hsu et al., 2010; Song et al., 2006; Chen et al., 2012*), while in others, such as *Dendrobium thyrsiflorum*, they present different expression patterns due to their differential expression during the ovule development (*Skipper et al., 2006*). Further evidences have shown that in *Cymbidium ensifolium* a class C gene duplication followed by sub-functionalization gave rise to different functional paralogs (*Wang et al., 2011*).

In *Arabidopsis*, the class C and D genes exhibit similar regulatory mechanisms that involve the presence of intragenic *cis*-acting elements. The expression of the *AG* gene is regulated by *cis*-acting elements present within the large intron 2 (*Deyholos and Sieburth, 2000; Hong et al., 2003*) and conserved among species (*Causier et al., 2009*). The *cis*-regulatory elements of the *STK* gene are present within the promoter region and the intron 1, where there are GA-rich elements able to bind the BASIC PENTACYSTEINE 1 (BPC 1) protein, with consequent DNA conformational changes (*Kooiker et al., 2005*).

AIMS OF THE STUDY

My PhD project is part of a wide study that includes different aspects of the Mediterranean orchid biology, such as molecular characterization of naturally occurring hybrids, molecular phylogeny, population genetic variability and characterization of genes involved in flower development (e.g.: *Aceto et al., 1999a; Aceto et al., 1999b; Montieri et al., 2004; Aceto et al., 2007; Cantone et al., 2009; Aceto and Gaudio, 2011*). Since the time of Darwin, biologists have addressed their interest on the study of the origin and evolution of the Orchidaceae, with particular emphasis to questions regarding the extreme diversity and specialization of floral morphology and, more recently, the absence of relationships between morphological and molecular evolution.

The MADS-box genes represent one of the most studied plant gene family. Despite the absence of a published genome for orchids, comparative genetic analyses are clarifying the functional role and the evolutionary patterns of the orchid MADS-box genes. These studies have revealed that different evolutionary forces, such as purifying selection and relaxation of selective constraints, act on orchid flower development genes, sometimes highlighting an heterogeneous evolutionary pattern of the coding and non-coding regions (*Aceto et al., 2007, Cantone et al., 2009, Mondragòn-Palomino et al., 2009; Cantone et al., 2011*).

My PhD project is focused on the study of some genes involved in flower development of *Orchis italica* (Orchidoideae, Orchidinae), a Mediterranean orchid species. Firstly, I have addressed my attention on the analysis of the expression pattern of two class B *PI/GLO*-like paralog genes. As described in the previous sections, orchid class B *PI/GLO*-like genes are single-copy loci; however, in the tribe Orchidinae (subfamily Orchidoideae) two *PI/GLO*-like paralogs have been reported (*Kim et al., 2007; Cantone et al., 2011*). These two paralogs seem to evolve through different selective constraints acting on synonymous sites, suggesting their possible sub-functionalization (*Aceto et al., 2007; Cantone et al., 2009; Cantone et al., 2011*). The available data on the expression pattern of the two orchid *PI/GLO*-like paralogs were not complete and were obtained by non-quantitative techniques. To integrate the data obtained previously, in the present study the expression pattern of the *PI/GLO*-like paralogs *OrcPI* and *OrcPI2*

was analyzed in different floral tissues of *O. italic*a at different developmental stages, using the Real Time RT-PCR technique. The results obtained were object of a scientific publication (*Salemme et al., 2011*).

The second part of my PhD project is focused on the class C and D MADS-box genes, involved in the formation of male and female reproductive organs (in orchids fused into the column) and ovule development, respectively. In this work, the *OitaAG* and *OitaSTK* genes, homologs of the class C *AG* and the class D *STK* genes of *Arabidopsis*, respectively, were identified in *O. italica* and their expression pattern was examined in different tissues and at different developmental stages. In addition, for the first time in an orchid species, the *OitaAG* and *OitaSTK* genomic organization was analyzed and the comparative analysis of their putative conserved *cis*-acting regulatory elements was conducted. The results obtained were object of a scientific publication (*Salemme et al., 2013*).

The third part of my PhD project is focused on the class A, non-MADS-box gene *AP2*. In *Arabidopsis* the *AP2* gene is involved in the formation of perianth (sepals and petals), is a negative regulator of the *AG* gene and is targeted by the microRNAs miR172 that acts on the *AP2* transcript by RNA cleavage (*Rhoades et al., 2002; Aukerman and Sakai, 2003; Chen, 2004; Kasschau et al., 2003*). In the present study the *OitaAP2* gene, homolog of the *AP2* gene of *Arabidopsis*, was identified in *O. italica*. Two alternatively spliced isoforms were identified, named *OitaAP2* and *OitaAP2_ISO*. The genomic organization of the *OitaAP2* gene and its expression profile in different tissues of *O. italica* were analyzed. In addition, the expression levels of miR172 were evaluated and it was verified the cleavage of the *OitaAP2* mRNA at the miR172 target site. The results obtained are object of a scientific publication in preparation.

One of the possible regulatory mechanisms of gene expression involves the microRNAs (miRNAs), small non coding RNA molecules that act at post-transcriptional level. These molecules are implied in many processes such as response to stress, plant organ development, signal transduction, etc. In different plant species, miRNAs have been identified and validated; however, little is known about this class of small RNAs in orchids (*An et al., 2011; An and Chan, 2012*). In order to characterize miRNAs from *O.*

italica, in the present work two miRNA libraries obtained from floral bud and leaf tissue were sequenced using the 454 High-Throughput technology and the obtained reads were subjected to *in silico* analyses to identify known and novel miRNAs.

MATERIALS AND METHODS

Plant material

Florets of *O. italica* were collected from early (10 days before anthesis) and late (after anthesis) inflorescence. Some florets were dissected to collect outer and inner tepals, lip, column and ovary not pollinated. In addition, *O. italica* was manually pollinated and the ovaries were collected at 3, 7 and 10 days after pollination (*dap*). Leaf, root and stem tissue was also collected. All the tissues were quickly stored in RNA Later (*Ambion*) at -80 °C until processing.

Nucleic acids extraction

Total RNA was extracted from the collected tissues of *O. italica* using the TRIzol reagent (*Ambion*) and treated with DNase I (*Ambion*). Genomic DNA was extracted from leaves of *O. italica* following the Doyle & Doyle method (*Doyle and Doyle, 1987*). After extraction, nucleic acids quantification was determined by spectrophotometric measurement (Nanodrop 2000c, *ThermoScientific*) and their integrity was checked by agarose gel electrophoresis.

Total RNA from all tissues was reverse transcribed using the Advantage RT PCR Kit (*Clontech*) with oligo dT primer and the first strand cDNA was quantified using the spectrophotometer Nanodrop 2000c.

Isolation of the OitaAP2 cDNA

Based on the amino acid and nucleotide alignments of the conserved domains of the *AP2* genes present in GenBank, two degenerated forward primers were designed, AP2F1 (5'-GATGGGARTCKCAYATYTGGGA-3') and AP2F2 (5'-TGGGARGCTMGNATGGGNCARTT-3'), and used in combination with a poly-T primer to obtain double strand cDNA through two PCR amplification rounds. The obtained amplicons were cloned into the pGEM-T Easy vector (*Promega*) and positives

clones were sequenced using the T7 and SP6 universal primers. The sequence reactions were run on the 310 Automated Sequencer (*Applied Biosystems*). BLAST analysis was performed to verify the nature of the sequenced fragments and a set of specific forward and reverse primers was designed for subsequent analyses (**Tab. 1**).

Table 1: Nucleotide sequence of the *OitaAP2* primers. F, forward; R, reverse.

Name	Direction	Sequence (5'-3')
AP2F8	F	atggtgctagatctcaacgtgtcat
AP2F4 new	F	catcaaactcctctgttctcaatg
AP2F5	R	cttcacaaatgtgggcggtg
AP2R5	F	caccgcccacatttgtgaag
AP2F6	R	ggaagaatttgtgcatattcttcgg
AP2F4 old	F	gagcatcctcatgtttggggcag
AP2R4 old	R	ctgccccaaacatgaggatgctc
APF3	F	tgcagcatcatcaggattc
AP2R4	R	cctctggcttcatttgatattgag

Rapid Amplification of cDNA Ends (5' RACE)

5' RACE experiments were performed to obtain the nucleotide sequence of 5'-untranslated region (UTR) of the *OitaAP2* cDNA and of two previously isolated cDNAs, *OitaAG* and *OitaSTK*, encoding for the class C AG-like and class D STK-like MADS-box protein, respectively. Based on the partial nucleotide sequence of the *OitaAG* and *OitaSTK* cDNAs, specific primers were designed and used for subsequent analyses (**Tab. 2** and **3**, respectively). The reactions were conducted using the 5' RACE System (*Invitrogen*) and specific reverse primers for each cDNA (**Tab. 1-3**). Amplification products were cloned and sequenced as described above. The full length nucleotide sequence of the *OitaAG* and *OitaSTK* cDNAs were deposited in GenBank with the accession number JX205496 and JX205497, respectively. The full length nucleotide sequence of the *OitaAP2* cDNA is not yet deposited in GenBank.

Table 2: Nucleotide sequence of the *OitaAG* primers. F, forward; R, reverse.

Name	Direction	Sequence (5'-3')
AG1-UTRF	F	tctctgccctccctctctctcccc
AG1-217R	R	agaagatgacgagggcgacc
AG1-UTRF_{long}	F	tctctgccctccctctctcccccccaccatcttctctc
AG1-540R_{long}	R	gttgcagatattgatccagaactggagttatca
AG1-217F	F	ggtcgccctcgtcatcttct
AG1-121R	R	tgcttgctttcttgtaacgctcga
AG1-int2F3	F	tacacataagatgtgaggctaata
AG1-840R	R	ggagtcaaagggaggcatcat
AG1-749R_{long}	R	gcatgtactctatctcagcatatagcaactca
AG1-439F	F	caaaacttcgacagcaaatcaca
AG1-1043R	R	ctgtagtggagaaaaacctgct
AG1-19R	R	gaggacagagagctcgatgg
AG1-308R	R	ttcttgtaacgctcgatagtgcc
AG1-10R	R	gcggcgcttacaaaaggtgac

Table 3: Nucleotide sequence of the *OitaSTK* primers. F, forward; R, reverse.

Name	Direction	Sequence (5'-3')
UTRF0	F	ttctggccggccatggatgct
63R	R	attcatagaggcggccgcgg
123F	F	gccaacagttctacttcagctgtgg
123R	R	ccacagctgaagtagaactgttggc
420R	R	tgctgcttcctgctgatagta
181F	F	aggaagcagcaaagttgcggca
181R	R	tgccgcaactttgctgcttcct
282F	F	agagaccagacttgaacgaggcct
282R	R	aggcctcgttcaaagtctggtctct
508F	F	aagcaactagagaccagacttgaacgaggcct
739R	R	tgtgataagctgctgctgcctcgaacatgttgaca
577F	F	gaggaaatagagttcatgcaaaag
362F	F	gcggagctacacgatgaaagtatgt
362R	R	acatactttcatcgtgtagctccgc
854R	R	taagctgctgctgcctcgaacatg

Mir172 cleavage assay on the OitaAP2 cDNA

Using the RLM-RACE GeneRacer™ kit (*Invitrogen*), a modified 5' RACE experiment (*Llave et al., 2002*) was performed to verify the cleavage of the *OitaAP2* cDNA induced by mir172. Total RNA (2 µg) extracted from the early column tissue of *O. italica* was directly ligated to the RLM-RACE 5'RACE RNA Oligo Adaptor (45 nucleotides) without enzymatic treatment with Calf Intestine Alkaline Phosphatase (CIP) to remove the 5' CAP. After the RNA ligase reaction, RNA was reverse transcribed with the oligo dT primer. A first PCR amplification was performed using the resulting first strand cDNA as template, the forward 5' RACE Outer Primer and the *OitaAP2* specific reverse primer AP2R4 (**Tab. 1**) that anneals downstream the putative mir172 target site. The 5'RACE Inner Primer and the AP2R4 primer were used to perform a nested PCR and the amplification product was cloned and sequenced as described above.

Expression analysis

Real Time RT-PCR experiments were conducted on the collected tissues of *O. italica* to examine the expression profile of the *OitaAP2*, *OitaAG*, *OitaSTK* genes and of the two class B MADS-box paralog genes *OrcPI* and *OrcPI2*, previously isolated (*Cantone et al., 2011;* GenBank accession number AB094985 and AB537504, respectively). Relative abundance of the transcripts was assessed applying the $\Delta\Delta C_T$ method, using the actin encoding gene *OitaAct* as endogenous gene (GeneBank accession number AB630020) and leaf as reference tissue. The nucleotide sequences of the primers used are listed in **Table 4**. For all the genes examined, the reactions were conducted in triplicates on two independent biological samples, using 30 ng of cDNA from each tissue. Negative controls were performed without cDNA in the reaction mixture. The reactions were performed using a 7500 Real Time PCR System (*Applied Biosystems*) in the presence of ×1 Power Sybr® Green PCR Master mix (*Applied Biosystems*) and 0.1 µM of each primer. The thermal protocol was as follows: 2 min at 50°C, 10 min at 95°C, followed by 40 cycles of 15 s at 95°C and 1 min at 60°C. A melting curve of PCR products (95–60°C) was performed to ensure the absence of artifacts.

The raw amplification data were examined using the PCR Miner online tool (*Zhao and*

Fernald 2005). This software is able to calculate the PCR efficiency (E) and the optimal threshold cycle (C_T) for each well through a non-linear regression algorithm, without the need of a standard curve. The mean PCR efficiency for each gene was used in the subsequent analysis. The relative expression ratios (rER) of each target gene relative to the expression of the endogenous *OitaAct* gene was calculated applying the following formula (*Schefe et al., 2006*):

$$rER = (1 + E_{target})^{-\Delta C_T target}/(1 + E_{actin})^{-\Delta C_T actin}$$

where, E represents the mean PCR efficiency of the gene, $\Delta C_{T\ target}$ is the difference between the C_T value of the target gene in the tissue of interest and the C_T value of the target gene in the reference tissue (leaf), ΔC_T actin is the difference between the C_T value of the *OitaAct* gene (endogenous control) in tissue of interest and the C_T value of the *OitaAct* gene in leaf tissue. The mean rERs and the corresponding standard deviations were calculated for each triplicate and for each biological sample. Differences in the relative expression levels of the target genes between and/or among different samples were assessed by the two-tailed t test and ANOVA followed by the Tukey HSD post hoc test, respectively.

Table 4: Nucleotide sequence of the primers used in the Real Time RT-PCR experiments. F, forward; R, reverse.

Name	Direction	Sequence (5'-3')
AP2RealF	F	tgtgtaccccggattatttcct
AP2R3	R	gaatcctgatgatgctgca
AP2RealF_iso	F	caagaaattgaaggaaagggccatgg
OrcPIRealF	F	cccagaatatgcggaccagatgcc
OrcPIRealR	R	tgggctggaaaggctgcacg
OrcPI2RealF	F	gagagtacgcaccgccaccg
OrcPI2RealR	R	gctggatgggctgcacacga
OitaAG1realF	F	tctgcaacaaatgcgcagtat
OitaAG1realR	R	aagcttgtgatttgctgtcgaa
OitaAG2realF	F	cggagctacacgatgaaagtatgt
OitaAG2realR	R	ccgcgccctctcgtttt
ActRealF	F	tcgcgacctcaccaatgtac
ActRealR	R	ccgctgtagttgtgaatgaatagc
Phal_172_1	F	cgaatcttgatgatgctgcat
Poly T adaptor	R	gcgagcacagaattaatacgac

For the expression analysis of the micro RNA miR172, the Poly(T) Adaptor RT-PCR method was used (*Shi et al., 2012*). In brief, starting from 350 ng of total RNA from each tissue, a poly-T adaptor was ligated to the 3'-terminus of the miRNAs and subsequently the reverse transcription reaction was performed. The forward primer specific for miR172 was designed based on the nucleotide sequence of miR172 of the orchid *Phalaenopsis equestris* (*An et al., 2011*) and used in combination with the poly-T adaptor reverse primer during the Real Time PCR experiments. Also in this case *OitaAct* was used as endogenous gene. The raw data were analyzed using the same approach described above.

PCR amplification of genomic DNA

Based on the putative intron positions of the examined genes, PCR amplifications were conducted using 100 ng of DNA of *O. italica* to amplify the genomic region spanning from 5'- to 3'-UTRs of the *OitaAG*, *OitaSTK* and *OitaAP2* genes. The reactions were

conducted through two PCR amplification rounds using the Long Amp Taq PCR kit (*New England Biolabs*), following the manufacturer instructions. The amplification of large size introns (intron 2 and 3 of the *OitaAG* locus, intron 2 and 5 of the *OitaSTK* locus, intron 9 of the *OitaAP2* locus) was performed using the thermal cycle described in *Keeney, 2011* (**Tab. 5**). Based on the size of the obtained amplicons, the optimal cloning vector was chosen: the small size fragments were cloned into pGEM-T Easy (*Promega*), while the large ones were cloned using the CopyControl cDNA, Gene and PCR Cloning Kit (*Epicentre*). Positives clones were sequenced using pGEM-T Easy or pCC1 plasmid specific primers and gene-specific nested primers.

Sequence analysis

The nucleotide sequences of the *OitaAP2, OitaAG* and *OitaSTK* cDNAs were virtually translated and aligned with their respective AP2-like, AG-like and STK-like sequences of monocots present in public databases using the MUSCLE software (*Edgar, 2004*). For each alignment, the maximum likelihood (ML) tree with 1000 bootstrap replicates was constructed under the best amino acid evolutionary model using the MEGA software (*Tamura et al., 2011*).

The nucleotide sequences of the genomic fragments of the *OitaAG, OitaSTK* and *OitaAP2* loci were overlapped and the resulting contigs were aligned with the corresponding cDNA sequences using Bioedit (*Hall, 1999*). The whole genomic sequences of the *OitaAG* and *OitaSTK* loci were deposited in GenBank with the accession number JX205498 and JX205499, respectively. The whole genomic sequence of the *OitaAP2* locus is not yet deposited in GenBank.

The nucleotide sequence of the introns of the *OitaAG, OitaSTK* and *OitaAP2* genes were scanned using the CENSOR online tool (*Kohany et al., 2006*) to search for the presence of repetitive elements and traces of plant mobile elements. The analysis was performed using the default settings and selecting the Viridiplantae section of REPBASE, a reference database of eukaryotic repetitive/mobile DNA.

Sequences of intron 2 of the *OitaAG* gene, of intron 1 of the *OitaSTK* gene and of corresponding introns of *AG*- and *STK*-like genes, respectively, present in GenBank

(**Tab. 6**) were screened to search for shared elements using the MEME online tool (*Bailey et al., 2009*). These sequences were also analyzed using the PLANTPAN database of plant transcription factor binding sites and *cis*-regulatory elements (*Chang et al., 2008*).

Table 5: Thermal cycle of the long template PCR.

Step	Temperature	Time
Initial denaturation	94°C	2 min
10 cycles consisting of:		
Denaturation	94°C	12 sec
4 subcycles per main cycle consisting of:		
Annealing	60°C	2 min
Extension	65°C	2 min
16 cycles consisting of:		
Denaturation	94°C	12 sec
4 subcycles per main cycle consisting of:		
Annealing	60°C	2 min + 3 sec per cycle
Extension	65°C	2 min + 3 sec per cycle

Table 6: Species name, abbreviation, accession number and sequence size of the intron 2 (*AG*-like gene) and intron 1 (*STK*-like gene) examined

Locus	Intron	Species	Abbreviation	Accession number	Length (bp)
AG-like	2	*Antirrhinum majus*	PLE	AY935269	6,666
		Oryza sativa	OsMADS3	AP008207	5,351
		Populus trichocarpa	PTAG1	AF052570	4,865
		Petunia hybrida	PMADS3	AB076051	4,010
		Ipomea nil	DP	AB281192	3,432
		Solanum lycopersicum	TAG1	AY254705	3,190
		Arabidopsis thaliana	AG	At4g18960	2,999
		Antirrhinum majus	FAR	AJ239057	2,965
		Cucumis sativus	CUM1	AY254704	1,903
STK-like	1	*Oryza sativa*	OsMADS13	Os12g10540	1,362
		Arabidopsis thaliana	STK	AT4G09960	1,328
		Zea mays	ZAG2	NM_001111908	1,261

Construction of microRNA libraries from floral bud and leaf tissue

Small non coding RNAs, including miRNAs, were isolated starting from 20 mg of floral bud and leaf tissue of *O. italica* using the mirVANA miRNA Isolation Kit (*Ambion*), following the manufacturer's instructions. The small RNAs were used to construct two miRNA libraries, floral bud and leaf, through a multistep process (*Ambros, 2001; Lau et al., 2001; Klevebring et al., 2009*). Size selection of small RNAs (~15-30 nucleotides) was performed through denaturing polyacrylamide gel electrophoresis. An end-modified linker (5'-Adenylate, 3'-Dideossi) was added to the 3'-terminus of the selected RNAs using the T4 RNA Ligase 1 (*New England Biolabs*) according to the instructions of the manufacturer. After gel electrophoresis, elution and purification, an hybrid DNA/RNA adaptor (5'-atcgtAGGCACCUGAUA-3') was added to 5'-terminus of the previously linked product, using the T4 RNA ligase. The obtained small RNA fragments were further size selected, eluted, purified and reverse transcribed using the SuperScript III Reverse transcriptase (*Invitrogen*). PCR amplification using the Phusion High-Fidelity Taq DNA Polymerase (*New England Biolabs*) and 5'-terminus mono phosphorylated primers was performed to produce concatamers that were ligated using the T4 DNA Ligase (*Invitrogen*) and sequenced using the 454 high-throughput sequencing technology (**Fig. 8**).

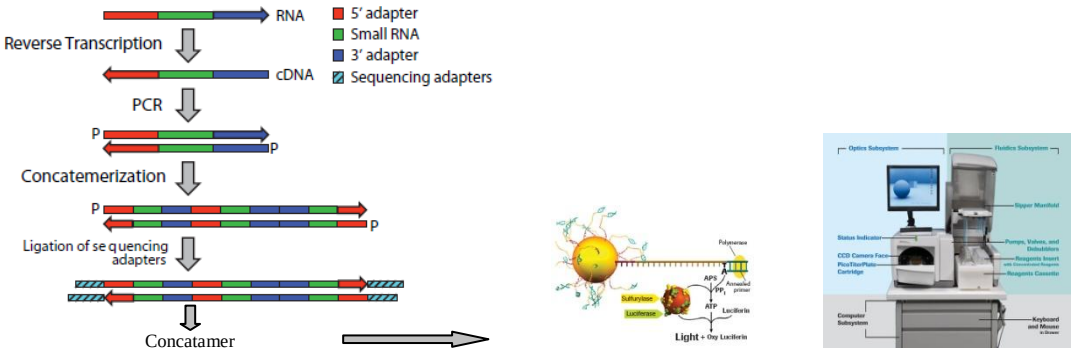


Figure 8: Workflow of the miRNA library construction and sequencing.

RESULTS AND DISCUSSION

Expression pattern of the two class B paralogs OrcPI and OrcPI2

In contrast to other orchids, Orchidinae have two class B MADS-box *PI/GLO*-like paralogs that evolve under selective constrains different from those acting on the single copy *PI/GLO*-like gene of the other orchid species (**Fig. 9**). In *O. italica* these two paralogs are *OrcPI* and *OrcPI2*, both exhibiting relaxation of purifying selection when compared to the single copy lineages. In addition, a high percentage of sites between *OrcPI* and *OrcPI2* show different rate of synonymous substitutions, suggesting their possible functional divergence (*Cantone et al., 2011*).

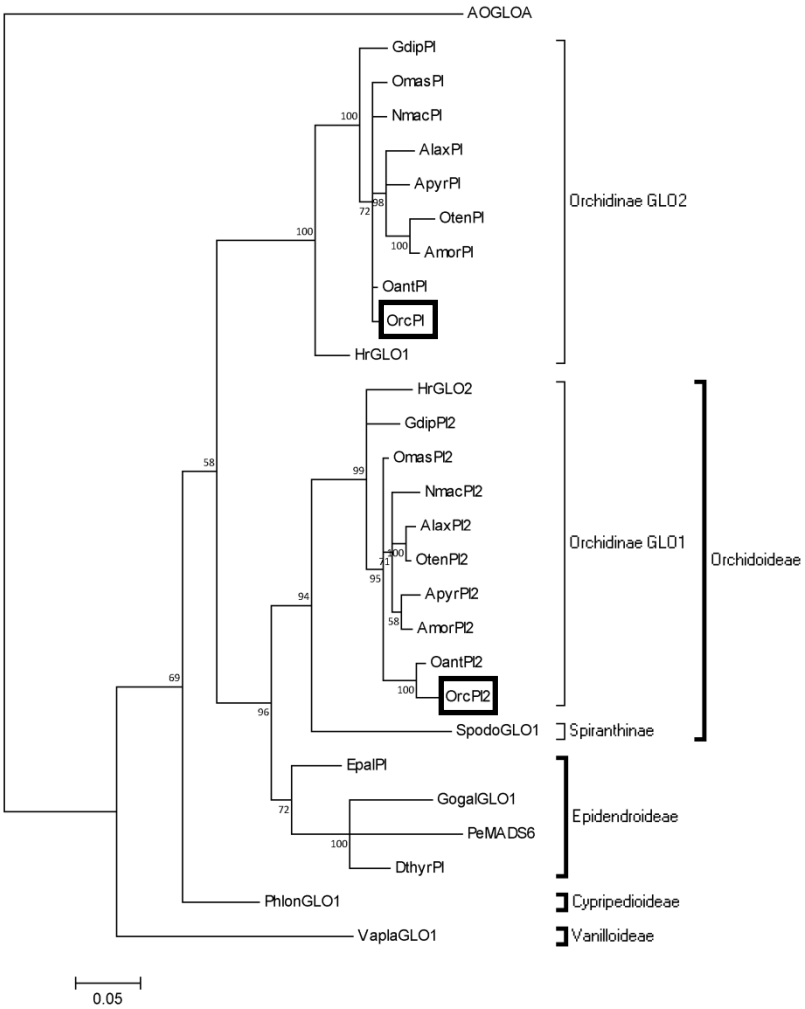


Figure 9: Phylogeny of the orchid *PI*/*GLO*-like genes (*Cantone et al., 2011*).

In the present study, Real Time PCR analysis has been conducted in *O. italica* in order to clarify the expression pattern of the two paralogs *OrcPI* and *OrcPI2* and to compare it with the expression pattern of the single copy lineages. The expression pattern of the two *PI/GLO*-like paralog genes has been examined in different floral tissue and in different developmental stages.

The melting curve plots demonstrated the effectiveness of the primers used. The mean PCR efficiency for each gene (the target genes *OrcPI* and *OrcPI2* and the endogenous control *OitaAct*) showed comparable values, ranging from 0.92±0.02 for the *OrcPI* gene to 1.04±0.02 for the *OitaAct* gene. **Figure 10** shows the relative expression ratio of the *OrcPI* and *OrcPI2* genes in different floral tissues of immature (**Fig. 10A, B**) and mature (**Fig. 10C, D**) inflorescence. In the floral tissues of *O. italica* the transcripts of *OrcPI* and *OrcPI2* present different levels of expression, with the *OrcPI2* gene showing a significantly higher expression when compared to *OrcPI* in all the examined tissues (*t* test $p \leq 0.030$).

In the immature inflorescence (**Fig. 10B**), the column shows the lowest expression of both genes, with an expression ratio *OrcPI2/OrcPI* of 1.6. The expression level of *OrcPI* in the outer tepal and lip is comparable, such as the expression level of *OrcPI2*. The expression ratio *OrcPI2/OrcPI* is 7.05 in the outer tepal and 3.69 in the lip. Both genes have the highest expression level in the lateral inner tepals, with the expression ratio *OrcPI2/OrcPI* of 5.65.

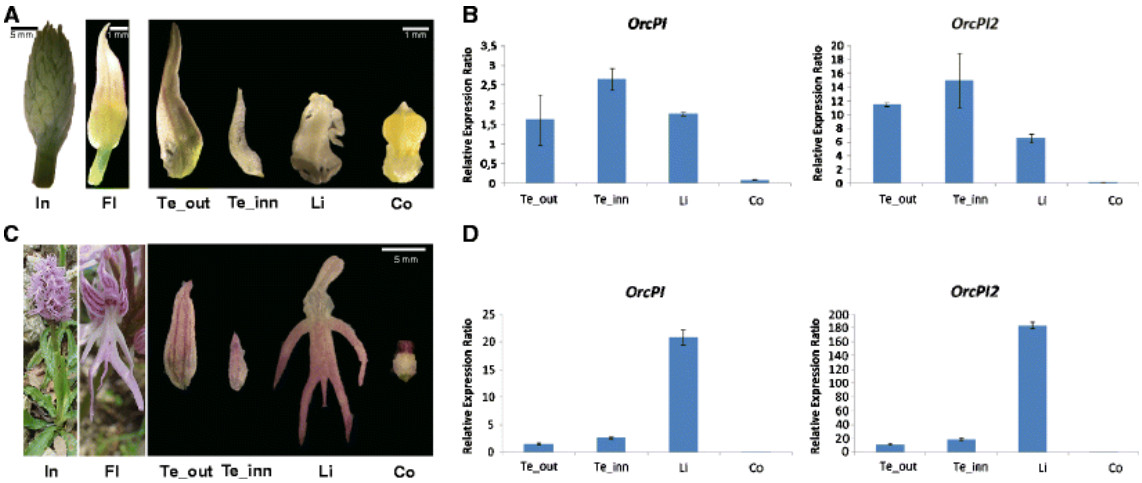


Figure 10: Floral tissues of *O. italica* and relative expression pattern of the *OrcPI* and *OrcPI2* genes. **A** and **C**: Immature and mature inflorescence of *O. italica*; **B** and **D**: Relative expression ratio of *OrcPI* and *OrcPI2* in floral tissues from immature and mature inflorescence, respectively. In, Inflorescence; Fl; Floret; Te_out, Outer tepal; Te_inn, Inner tepal; Li, Lip; Co, Column. (*Salemme et al., 2011*).

In the mature inflorescence (**Fig. 10D**), the expression pattern of the *OrcPI* and *OrcPI2* paralogs is different from that observed in the immature inflorescence. For both genes the column continues to be the tissue where the lowest expression is detectable (*t* test $p \leq 0.016$), with an expression ratio *OrcPI2/OrcPI* of 2.6. The lip shows the significantly highest expression level of both *OrcPI* and *OrcPI2*, with an expression ratio *OrcPI2/OrcPI* of 8.82. The amount of *OrcPI* and *OrcPI2* mRNA in the outer and lateral inner tepal is higher than that detected in the column. The expression ratio of *OrcPI2/OrcPI* is ~7 in both outer and lateral inner tepal tissues.

The pairwise comparison of the *OrcPI* and *OrcPI2* tissue expression level between immature and mature inflorescence produced significant results in the lip and column tissues ($p \leq 0.002$). In particular, both genes show an increasing expression level in the lip from immature to mature inflorescence (expression ratio mature/immature inflorescence: 11.77 for *OrcPI* and 28.1 for *OrcPI2*) and an opposite trend in the column (expression ratio mature/immature inflorescence 0.16 for *OrcPI* and 0.27 for *OrcPI2*). In the outer and lateral inner tepals the expression level of both genes remains approximately the same, with an expression ratio mature/immature inflorescence of ~1.

Thus, the *OrcPI* and *OrcPI2* paralog genes display a common, co-expression pattern in the different floral tissues and in the different developmental stages examined, which is in general agreement with that observed in some *PI/GLO*-like single-copy orchid lineages (*Mondragòn-Palomino and Theissen, 2011*).

Based on the hypothesis that the *PI/GLO*-like gene *PeMADS6* of the orchid *Phalaenopsis equestris* (Epidendroideae) might be a negative regulator of the ovary development (*Tsai et al., 2005*), expression analysis of both *OrcPI* and *OrcPI2* was conducted on ovary tissue of *O. italica* before and after pollination. In *O. italica*, such as in all orchids, ovary development is triggered by pollination. In this Mediterranean species the ovary reaches the complete maturation approximately 10 days after pollination (*Cozzolino*, personal communication). **Figure 11** shows the expression pattern of the *OrcPI* and *OrcPI2* genes in ovary tissue collected from immature and mature inflorescences before pollination and 3, 7, and 10 days after manual pollination. In the ovary of the immature inflorescence both genes show a weak expression level, whereas in the ovary of the mature inflorescence before pollination they are expressed at a significantly higher level. After pollination the expression of *OrcPI* and *OrcPI2* rapidly decrease, according to the expression profile of *PeMADS6* in *P. equestris*. This result supports the previously hypothesized role of the *PI/GLO*-like genes as negative regulators of ovary development in orchids. Also in the ovary tissue the expression of the *OrcPI2* gene is higher than that of the *OrcPI* gene, with an expression ratio *OrcPI2/OrcPI* of 3.58 in the ovary collected from mature inflorescence.

In conclusion, the expression analysis of the *OrcPI* and *OrcPI2* genes in *O. italica* revealed that these *PI/GLO*-like paralogs are co-expressed, even though the expression level of the *OrcPI2* gene, ortholog of the single-copy *PI/GLO*-like gene from other orchid lineages, is higher than that the *OrcPI* gene in all the examined tissues. The observed transcriptional pattern confirms the hypothesis of a recent lineage-specific duplication event within Orchidinae, with a functional role of the duplicated gene *OrcPI* fully or partially redundant relative to *OrcPI2*. The evolutionary and expression pattern of the two paralogs suggest that the duplicated locus *OrcPI* might evolve through sub-functionalization resulting in a possible partition between the OrcPI and OrcPI2

protein's ability to interact with different class-B AP3/DEF-like proteins during the formation of the specific floral quartets necessary for correct floral tissue development.

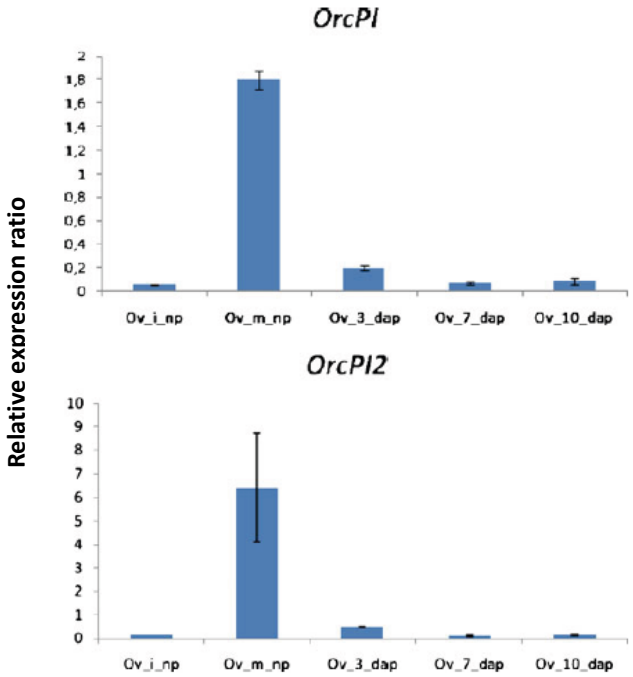


Figure 11: The expression pattern of *OrcPI* and *OrcPI2* genes in ovary tissue of *O. italica*. Ov_i_np, Immature ovary not pollinated; Ov_m_np, Mature ovary not pollinated; Ov_3_dap, Ovary 3 days after pollination; Ov_7_dap, Ovary 7 days after pollination; Ov_10_dap, Ovary 10 days after pollination (*Salemme et al., 2011*).

The class C OitaAG and class D OitaSTK cDNAs

The whole cDNA sequence of *OitaAG* (1.149 bp) includes a 5'- and 3'-UTR of 208 and 235 bp, respectively. The virtual translation of this sequence results in a protein of 234 residues. BLAST analysis reveals the highest similarity of the *OitaAG* sequence with the *AG1* gene of the orchid *Dendrobium crumenatum* (88% nucleotide and 87% amino acid identity). The whole cDNA sequence of *OitaSTK* (903 bp) includes a 5'- and 3'-UTR of 105 and 113 bp, respectively. The virtual translation of this sequence produces a protein of 227 residues. BLAST analysis shows the highest similarity of the *OitaSTK* sequence with the *AG2* sequence of *D. crumenatum* (82% nucleotide and 84% amino acid identity). The amino acid alignment of the OitaAG and OitaSTK sequences with other AG-like and STK-like sequences of monocots present in GenBank (**Fig. 12**)

reveals the presence of two conserved regions at the C-terminus known as the AG motifs I and II. Furthermore, the OitaAG sequence presents an additional stretch of 7 residues localized at the N-terminus and the OitaSTK sequence has an additional stretch of 6 residues at the C terminus, known as the MD motif. These stretches are common for the monocots AG-like and STK-like proteins, respectively (*Yun et al; 2004*).

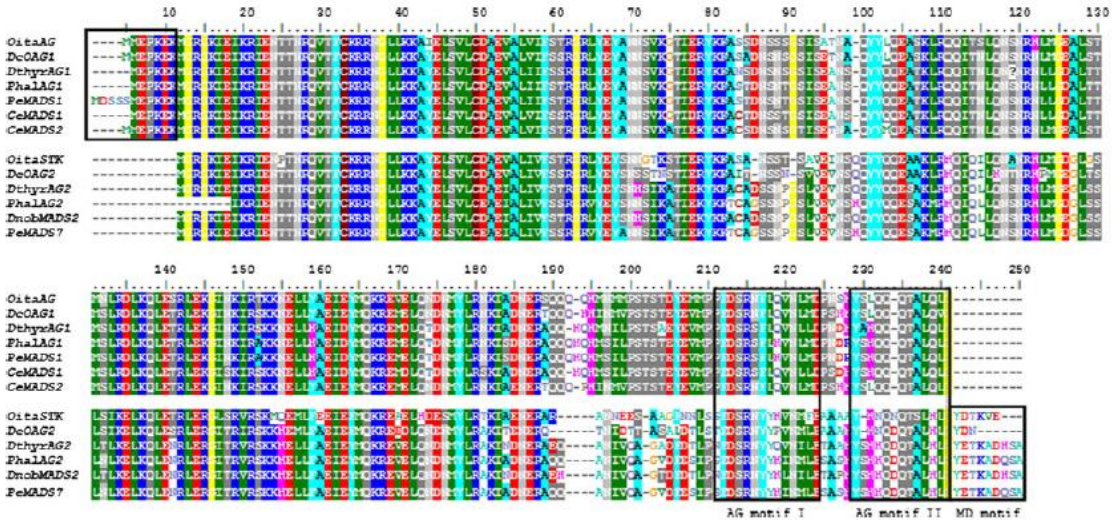


Figure 12: Amino acid alignment of the OitaAG and OitaSTK sequences with other AG-like and STK-like proteins of other orchid species. Boxes refer to N-terminal extension, AG motif I and II and MD motif (*Salemme et al., 2013*).

The bootstrap consensus ML tree was constructed under the best amino acid evolutionary model JTT+G +I (**Fig. 13**). The tree topology shows two main different branches corresponding to the class C and D MADS-box functional classes: OitaAG belongs to the orchid class C clade (bootstrap value 86 %) and OitaSTK belongs to the class D clade.

Sequence comparison revealed that the *OitaAG* and *OitaSTK* genes of *O. italica* belong to the MADS-box class C and D genes, respectively. Phylogenetic analysis showed that these genes are closely related to the *AG*-like and *STK*-like genes isolated in other orchid species.

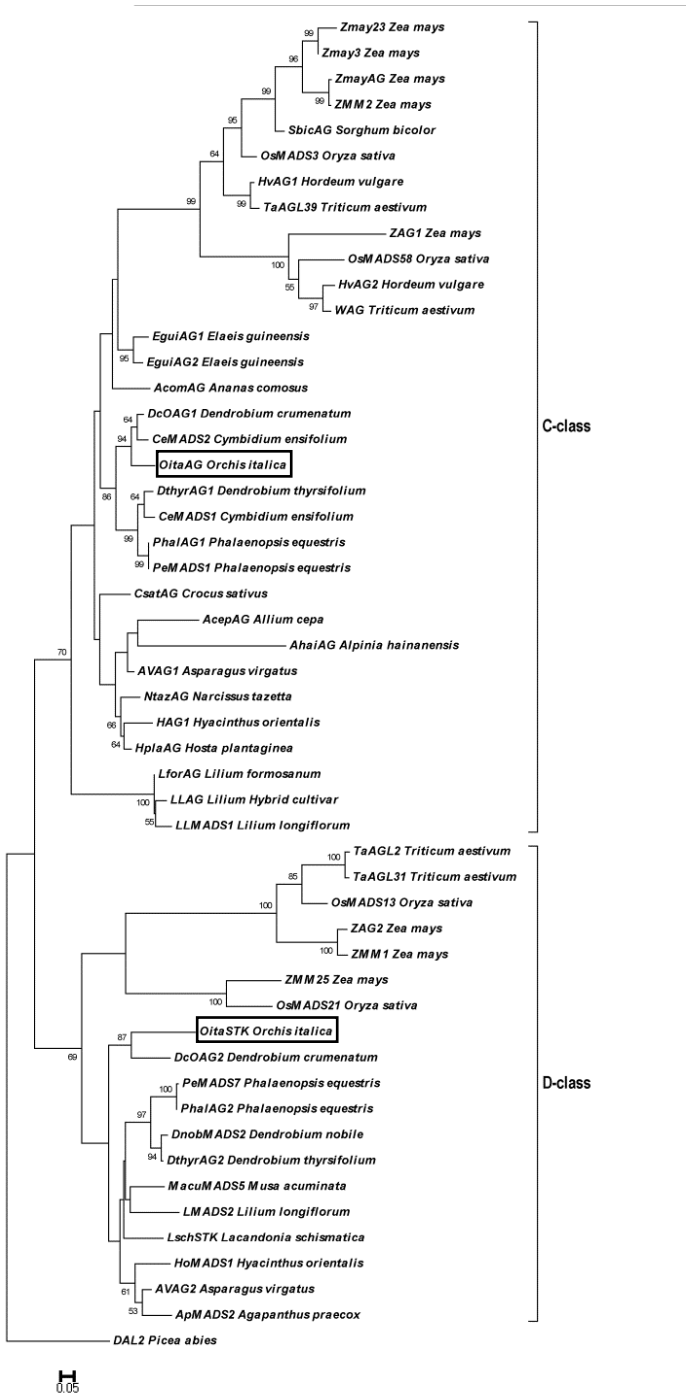


Figure 13: Maximum linkelihood (ML) tree based on the amino acid alignment of the AG-like and STK-like proteins of some monocot species. Numbers indicate the bootsrap percentage (values lower than 50% are not shown) (*Salemme et al., 2013*).

Expression pattern of the OitaAG and OitaSTK genes

The relative expression ratio of the *OitaAG* and *OitaSTK* genes has been evaluated in floral tissues from early and late inflorescence (**Fig. 14A, B**). The melting curve plots demonstrated the effectiveness of the primers used. The mean PCR efficiency for each gene (the target genes *OitaAG* and *OitaSTK* and the endogenous control *OitaAct*) showed comparable values. Both genes are expressed mainly in the column, with the expression of *OitaSTK* significantly different between early and late stage tissues. In late inner tepals the expression of *OitaAG* is weak and in late lip tissue a low level of *OitaSTK* mRNA is present (**Fig 14C, D**).

In the ovary before pollination both the *OitaAG* and *OitaSTK* genes are expressed, with the level of *OitaSTK* transcripts much higher than that of *OitaAG* (**Fig. 15A**). Three days after pollination the expression of both genes begins to increase, reaching the highest levels at 7 and 10 days after pollination. In both pollinated and unpollinated ovaries the expression of the *OitaSTK* gene is much higher than that of *OitaAG* (**Fig. 15A**).

In root and stem tissue the expression level of the *OitaAG* gene is comparable to that in leaf (reference tissue), whereas in root tissue a slightly higher expression level of the *OitaSTK* gene is detectable (**Fig. 15B**).

Taken together, these results allow to compare the expression pattern of the *OitaAG* and *OitaSTK* genes with that of other *AG*-like and *STK*-like genes in other species. *OitaAG* appears to be specifically expressed in floral tissues, like the other *AG*-like genes *ZAG1* of maize (*Schmidt et al., 1993*) *TAG1* of tomato (*Pnueli et al., 1994*), *AVAG1* of *Asparagus* (*Yun et al., 2004*), etc. Also *OitaSTK* is expressed in floral tissues like other *STK*-like genes, e.g. *FBP11* of *Petunia* (*Colombo et al., 1995*), *LMADS2* of lily (*Tzeng et al., 2002*), *AVAG2* of *Asparagus* (*Yun et al., 2004*). Surprisingly, the *OitaSTK* gene is expressed also in roots, even though at low levels. To date, the available studies on the C and D class MADS-box gene do not report the expression of these genes in root tissue; instead, some evidences in *Arabidopsis* (*Rounsley et al., 1995; Tapia-Lopez et al., 2008*) and rice (*Lee et al., 2008*) attribute to the sister C/D class gene *XAL1* (*AGL12* group) a root specific expression. Further studies are needed to understand if the

OitaSTK gene is involved in root development of *O. italica* or its weak expression represents the background level in this tissue.

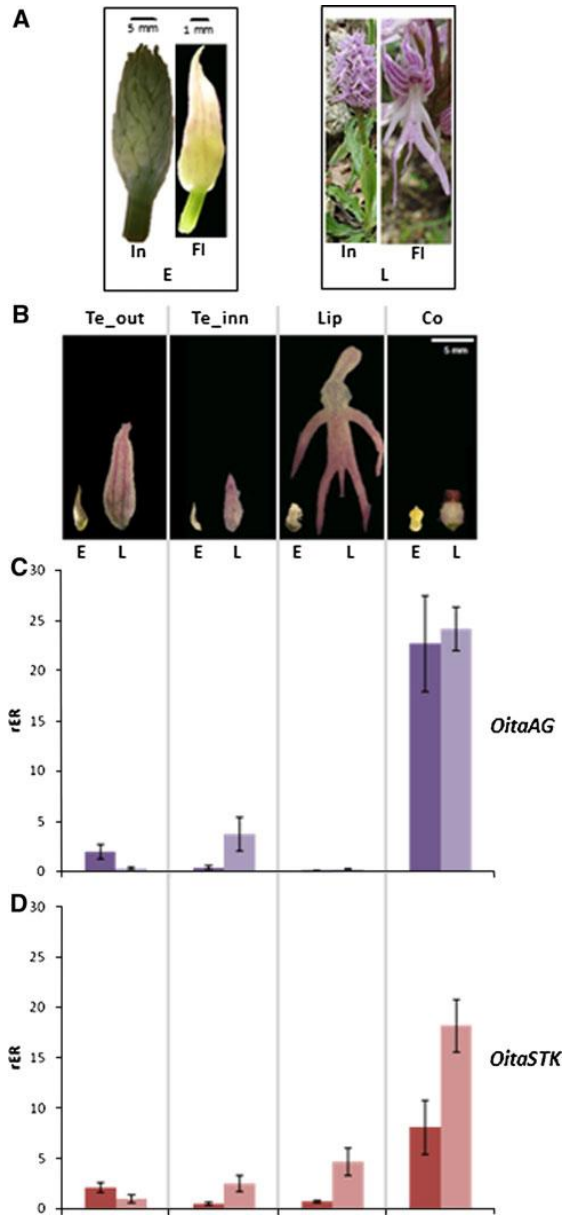


Figure 14: Expression pattern of the *OitaAG* and *OitaSTK* genes in tissues from early and late inflorescence of *O. italica*. **A:** On the left, early inflorescence and floret; on the right, late inflorescence and floret. **B**: On the left, early floral tissues; on the right late floral tissue. **C**: *OitaAG* relative expression ratio. **D**: *OitaSTK* relative expression ratio. E, Early; L, Late; In, Inflorescence; Fl, Floret; Te-out, Outer tepal; Te_in, Inner tepal; Co, Column; rER, Relative expression ratio. *(Salemme et al.,2013).*

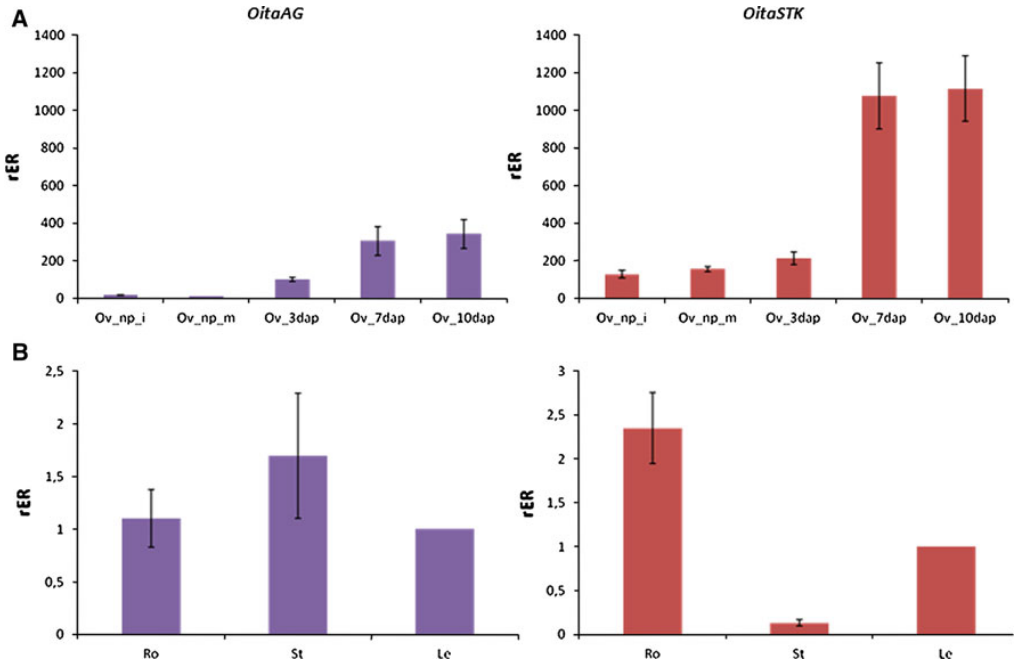


Figure 15: Expression pattern of the *OitaAG* and *OitaSTK* genes in ovary (**A**) and in vegetative tissues (**B**) of *O. italica.* Ov_np_i, Immature ovary not pollinated; Ov_np_m, Mature ovary not pollinated; Ov_3dap, Ovary 3 days after pollination; Ov_7dap, Ovary 7 days after pollination; Ov_10dap, Ovary 10 days after pollination; Ro, Root; St, Stem; Le, Leaf; rER, Relative expression ratio (*Salemme et al., 2013*).

As expected, both the *OitaAG* and *OitaSTK* genes display a similar expression pattern that confirms their role in the development and maintenance of reproductive structures. The general similarity in the expression pattern of *OitaAG* and *OitaSTK* suggests a possible redundant function; however, some differences are detectable. In particular, in early and late column tissue there is a similar amount of the *OitaAG* mRNA, whereas the expression level of the *OitaSTK* gene increases from early to late stages (**Fig. 14 C, D**). This difference might reflect a not completely redundant role for these two genes during the cell distension of gynostemium in *O. italica*, with *OitaAG* implicated both in the formation and maintenance of the column and *OitaSTK* involved mainly in the maintenance of this structure. In the ovary, before fertilization *OitaAG* is expressed at very low levels and the amount of its mRNA starts to increase at 3 days after pollination, reaching the highest level at 7 and 10 days after pollination (**Fig. 15A**). In

the unpollinated ovary *OitaSTK* is expressed at high levels and fertilization progressively enhances its accumulation from 3 to 10 days after pollination (**Fig. 15A**). This difference in expression patterns suggests a relevant role of the class D *OitaSTK* gene in ovary formation and maturation, while the class C *OitaAG* gene seems to be involved only in events subsequent to pollination. This expression profile overlaps only partially with that described in *Phalaenopsis*, where the class C *PeMADS1* and the class D *PeMADS7* genes are both expressed at very low levels in the immature ovary, with expression increasing only after pollination (*Chen et al., 2012*). Further differences are detectable comparing the expression profile of the *OitaAG* and *OitaSTK* genes of *O. italica* with that of their homologs in other orchid species, especially within tepal, lip and vegetative tissue. The *OitaAG* gene is weakly expressed in late inner tepal and *OitaSTK* shows a low expression in late stage lip and root tissues. Previous studies report the expression of the class C *DcOAG1* gene of *D. crumenatum* and of the class C *PhalAG1* and class D *PhalAG2* genes of *Phalaenopsis* in tepals and lip tissue (*Song et al., 2006; Chen et al., 2012*). However, recently it has been reported the expression of the *PeMADS1* and *PeMADS2* genes of *Phalaenopsis* only in the column and in post pollinated ovary (*Chen et al., 2012*). These contrasting results might reflect different evolutionary histories of the class C and D genes in orchids, a hypothesis also supported by the finding of two class C genes in *C. ensifolium* with distinct roles in gynostemium development (*Wang et al., 2011*).

Genomic structure of the OitaAG and OitaSTK loci

In dicots species, the presence of *cis*-regulatory elements within intron 2 of *AG*-like genes and intron 1 of *STK*-like genes has been widely documented, whereas in monocots analyses have been restricted to rice. For this reason, in the present study it was determined the position and sequence of introns in both the *OitaAG* and *OitaSTK* genes of *O. italica*.

The *OitaAG* gene has eight introns and *OitaSTK* has seven (**Fig. 16**); all the sequenced introns display the canonical 5'-GT and the 3'-AG termini. Intron 1 of both genes is included into the 5'-UTR; intron 8 of the *OitaAG* gene is located within the codon

preceding the TAG stop codon. In both loci the intron position results conserved, maintaining a similar exon size; on the contrary, introns are not conserved in size, that results greatly variable. The large intron 3 of the *OitaAG* gene has been amplified but not sequenced. All of the attempts to amplify intron 2 of the *OitaSTK* gene gave negative results.

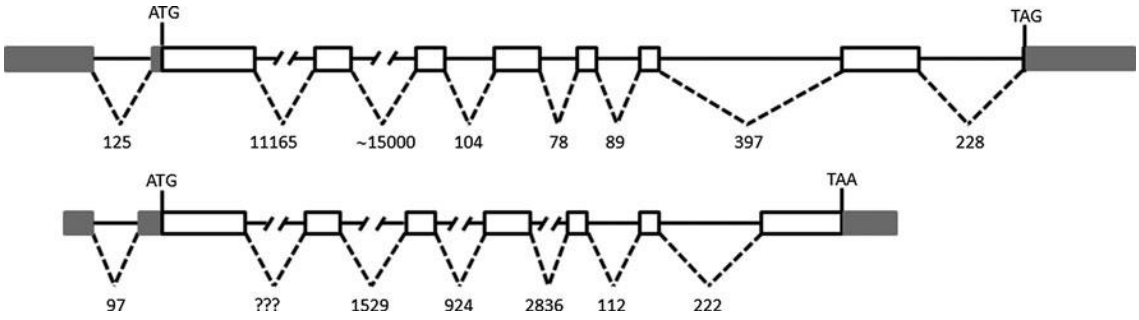


Figure 16: Genomic organization of the *OitaAG* (upper) and *OitaSTK* (lower) loci. Gray boxes, white boxes, and lines represent UTR, exons and introns, respectively. The numbers indicate the lengths of introns in bp, question marks refer to the unknow size of the intron *(Salemme et al., 2013).*

The CENSOR analysis of the intron sequences has shown traces of transposable elements mainly within the largest introns (**Tab. 7**), although some short sequences similar to transposable elements are present also within the small introns 4 and 8 of the *OitaAG* gene. Among the 25 traces of repetitive/mobile elements identified, 17 shared homology to Class I retrotransposons (9 LTR Copia or Gypsy and 8 NonLTR), 6 to Class II DNA transposons (4 En/Spm, 1 Helitron and 1 Mutator) and 2 to Interspersed repeats. The type and abundance of these repetitive/mobile elements are in agreement with those observed in a partial genome analysis conducted on the orchid *Phalaenopsis* (*Hsu et al., 2011*). In particular, Class I retrotransposons are highly represented (68% in *O. italica* and 75.7% in *Phalaenopsis*) and the most abundant are copia- and gypsy-like elements (36% in *O. italica* and 54.7% in *Phalaenopsis*). Class II transposons are less frequent than Class I retrotransposons (24% in *O. italica* and 10.8% in *Phalaenopsis*) and the most abundant are En/Spm elements (16% in *O. italica* and 5.3% in *Phalaenopsis*). The interspersed repeats, centromere and unclassified elements represent 8% in *O. italica* and 13.5% in *Phalaenopsis*.

Table 7: Results of the CENSOR analysis conducted on the introns of *OitaAG* and *OitaSTK*. Sim, nucleotide similarity; Length is expressed in bp.

Gene	Intron	Length	Element	Class	Sim
OitaAG	2	11165	ENSPM-6_ZM	DNA/EnSpm	0.69
			DNA9-3C_Mad	Interspersed_Repeat	0.78
			MuDR1_HV	DNA/MuDR	0.72
			ATCOPIA70_I	LTR/Copia	0.69
			Copia-2_PD-I	LTR/Copia	0.70
			ENSPM-6_ZM	DNA/EnSpm	0.70
			RTE-1_Mad	NonLTR/RTE	0.69
			RTE-1_Mad	NonLTR/RTE	0.67
			RTE-1B_Mad	NonLTR/RTE	0.64
			EnSpm-12_OS	DNA/EnSpm	0.76
	4	104	ATHILA-4_SBi-I	LTR/Gypsy	0.84
	8	228	EnSpm-5_STu	DNA/EnSpm	0.78
OitaSTK	3	1529	RTE-1_Mad	NonLTR/RTE	0.78
			Copia-55_BRa-I	LTR/Copia	0.75
			Copia-100_Mad-I	LTR/Copia	0.81
	4	924	SINE2-2_STu	NonLTR/SINE/SINE2	0.78
			SINE2-2_STu	NonLTR/SINE/SINE2	0.78
			SINE2-2_STu	NonLTR/SINE/SINE2	0.82
			SINE2-2_STu	NonLTR/SINE/SINE2	0.80
	5	2836	HELITRON7_OS	DNA/Helitron	0.85
			Gypsy-76_SB-I	LTR/Gypsy	0.90
			DNA9-3C_Mad	Interspersed_Repeat	0.83
			Copia1A-VV_I	LTR/Copia	0.73
			Copia-8_CP-I	LTR/Copia	0.66
			Copia2-VV_I	LTR/Copia	0.70

According to the general small size of introns in plant MIKC MADS-box genes, the size of some of the isolated introns is quite small. In contrast, the *OitaAG* gene has very large intron 2 and 3, to date, the largest ever reported (11,165 bp and ~15,000 bp, respectively). Also the size of intron 3, 4 and 5 of the *OitaSTK* gene is quite large, ranging from 924 to 2,836 bp. The size expansion of these noncoding regions might be a consequence of the insertion of mobile elements. For example, within intron 5 of the *OitaSTK* gene is still detectable a quite large (~850 bp) trace of an LTR/copia element. Furthermore, BLAST analyses revealed the presence in intron 2 of the *OitaAG* gene of two regions (spanning from the position 2,900 to 3,099 and from 4,986 to 5,065) that

show identity (~76%) with two regions (the first inverted and the second direct) in intron 1 of the *LFY* gene of the orchid *Ophrys*. These results indicate high and possibly recent activity of mobile elements that could also have captured genomic segments surrounding their original position and, after the insertion within the introns of the *OitaAG* and *OitaSTK* genes, undergone mutations and fragmentation leading to their inactivation.

Intron 1 of the *OitaSTK* gene is much shorter than those of the other *STK*-like genes here examined (**Tab. 6**, Materials and Methods) and their nucleotide sequences are not alignable. MEME analysis highlighted the presence of a single GA-rich motif (GAAGAAA) within intron 1 of the *OitaSTK* gene, whereas in the corresponding region of the other *STK*-like genes there are 5 (within *STK* and *OsMADS13*) and 7 (within *ZAG2*) GA-rich motifs whose sequence does not exactly matches to the consensus motif RGARAGRRA that, present in multiple copies within intron 1 of the *STK* gene of *Arabidopsis*, *in vitro* binds the BPC1 protein (*Kooiker et al., 2005*). The absence of conservation of these *cis*-regulatory motifs within intron 1 of the *OitaSTK*, *OsMADS13* and *ZAG2* genes might suggest the existence of regulators different from BPC1 involved in the transcriptional modulation of the class D genes in monocots. Alternatively, in monocots all the BPC1 binding sites might have a different consensus sequence or might be located in different sites, e.g. within the promoter.

As already discussed, intron 2 and 3 of the *OitaAG* gene are the largest among intron 2 and 3 of the known *AG*-like gene. Due to their global sequence divergence, the nucleotide sequence of intron 2 of the *OitaAG* gene is not alignable to those of the other *AG*-like genes here examined. However, MEME analysis revealed the presence within intron 2 of the *OitaAG* gene of conserved boxes corresponding to known *cis*-acting *AG* regulatory elements (*Hong et al., 2003; Causier et al., 2009*): the aAGAAT-box, the LEAFY binding site (LBS) and the 70 bp element known as CCAATCA-box. These motifs are present in conserved order and conserved distance relative to the other examined sequences (**Fig. 17**). The direct interaction of the LFY transcription factor with intron 2 of the *AG* gene is important for the early activation of the *AG* expression in *Arabidopsis* (*Weigel and Meyerowitz, 1993*), the 70bp element is involved in the late

stage activity of *AG* (*Hong et al., 2003*) while the function of the aAGAAT-box is still not clear. The conserved order and distance of these elements within intron 2 of the *AG*-like genes here examined highlights the importance of the nucleotide sequence and spatial distribution conservation of these three regulatory motifs in monocots as well in dicots.

In addition to these three conserved motifs, PLANTPAN analysis showed in all the sequences examined the presence of different known transcription factor binding sites (TFBSs). Among them, it has been observed the presence of the binding sites of three important *AG* regulators WUSCHEL (WUS), BELLRINGER (BLR) and APETALA2 (AP2) shared by intron 2 of the *OitaAG* gene and its homologs (**Fig. 17**).

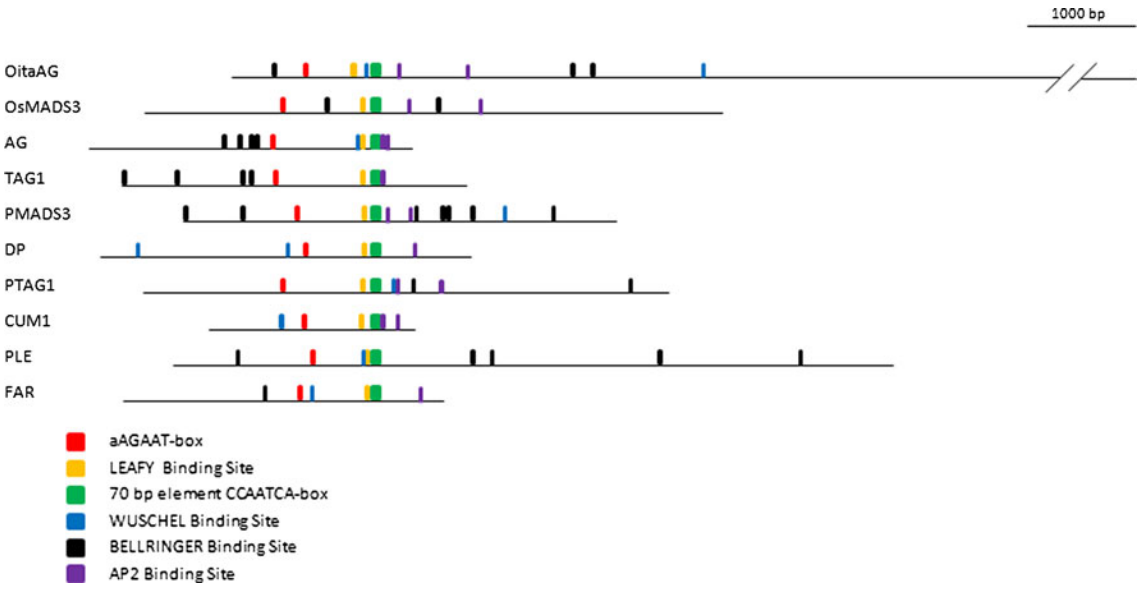


Figure 17: Scheme of the distribution of the conserved boxes and TFBSs within the intron 2 of *OitaAG* and its homologs (*Salemme et al., 2013*).

WUS and BLR are known as positive and negative regulators of *AG*, respectively (*Bao et al., 2004; Ikeda et al., 2009).* The WUS and BLR TFBSs are both present in multiple copies within intron 2 of the *OitaAG* gene and its homologs, even though their spatial location is not strictly conserved, especially in monocots (orchid and rice). The absence of canonical WUS binding sites within intron 2 of the *OsMADS3* gene might be due to

the partition of the class C functions between *OsMADS3* and its paralogous gene *OsMADS58* in rice (sub-functionalization) (*Yamaguchi et al., 2006; Hu et al., 2011*) and/or to the expression pattern of the *WUS* gene of rice (*OsWUS*), different from that of the *WUS* gene of *Arabidopsis* (*Nardmann and Werr, 2006*).

Within intron 2 of the *OitaAG* gene and its homologs the recently described AP2 binding site is located downstream the CCAATCA-box (**Fig. 17**). In *Arabidopsis*, the class A AP2 protein negatively regulates the expression of the *AG* gene recognizing a not-canonical AT-rich sequence within intron 2 (*Dinh et al., 2012*). In some of the examined genes, the position of the AP2 binding sites in intron 2 is just downstream of the 70 bp element, whereas in *O. italica* and rice it is farther downstream. The presence of conserved regions and TFBSs within intron 2 of the *OitaAG* gene of *O. italica* suggests that the general regulatory mechanisms of this gene are strictly conserved; however, differences in the number and distribution of the TFBSs might be related to peculiar aspects of the expression profile of this gene in orchids.

The class A OitaAP2 cDNA

The isolation of the *OitaAP2* cDNA was performed using the first strand cDNA from inflorescence of *O. italica* as template in PCR reactions conducted in presence of a degenerated forward primer that anneals within the region encoding for the AP2 domain 1 and the poly-T reverse primer. This reaction produced two different amplification fragments of ~1200 and 1100 bp that were cloned and sequenced. Alignment of the two sequenced fragments revealed their full nucleotide identity, excluding a 105 bp region absent in the shortest fragment. BLAST analysis showed both these fragment encode for a partial AP2-like protein. Based on these preliminary results, specific reverse primers were designed to obtain the 5'-terminus of these cDNAs and, subsequently, a number of forward and reverse specific primers were designed (**Tab. 1**, Materials and Methods) to verify if the observed fragments could be product of an alternative splicing event. PCR amplifications of the cDNA from inflorescence *O. italica* were conducted with different primer pairs. When the forward primer AP2F4new (that anneals 88 bp downstream of the ATG start codon) is used in combination with the AP2R4 primer (that anneals within

the 3'-UTR), two fragments are produced (1480 and 1380 bp, respectively) (**Fig. 18A, lane 1**). Two amplification fragments (1183 and 1083 bp, respectively) are produced also when the same forward primer is used in combination with the AP2RealR primer (that anneals 191 bp upstream of the TGA stop codon) (**Fig. 18A, lane 2**), whereas a single fragment (697 bp) is produced when the AP2F4new primer is used in combination with the AP2R5 reverse primer (that anneals within the region encoding for the AP2 domain 2) (**Fig. 18A, lane 3**). Two amplification fragments are produced when the forward primer AP2F5 (reverse and complement of AP2R5) is used in combination with the AP2RealR (439 and 339 bp) (**Fig. 18A, lane 5**) or the APR3 reverse primer (that anneals ~60 bp downstream of AP2RealR) (546 and 446 bp) (**Fig. 18A, lane 6**). A single amplification fragment (360 bp) is produced when the primer AP2RealF_iso, that anneals 10 bp upstream and 10 bp downstream of the region missing in the shortest isoform (**Tab. 4**), is used in combination with AP2R4 (**Fig. 18B, lane 2**).

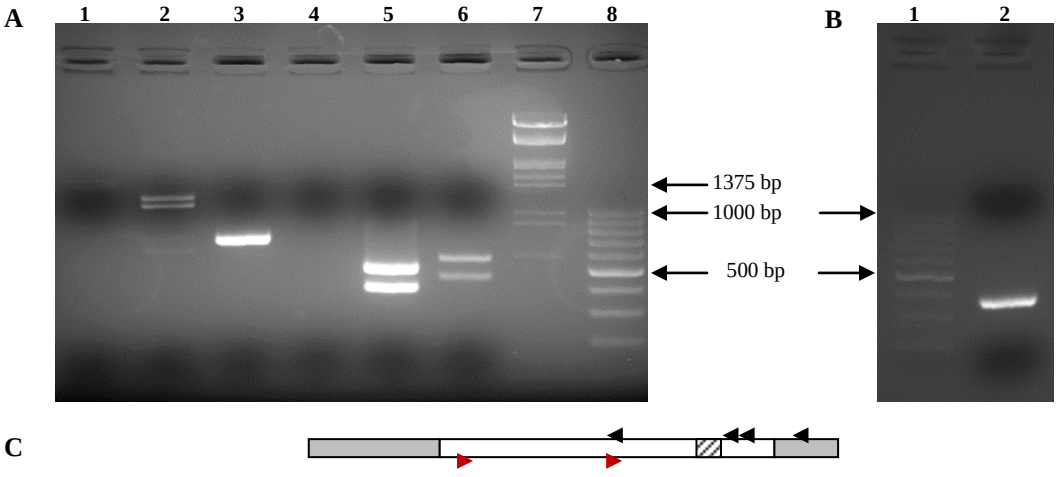


Figure 18: Gel electrophoresis of the PCR amplifications of the *OitaAP2* cDNA with different primer pairs. **A.** 1, AP2F4new/AP2R4; 2, AP2F4new/AP2RealR; 3, AP2F4new/AP2R5; 4, AP2F4new/AP2RealR without first strand cDNA template (negative control); 5, AP2F5/AP2RealR; 6, AP2F5/APR3; 7, DNA Marker III; 8, 100 bp ladder. **B.** 1, 100 bp ladder; 2, AP2RealF_iso/AP2R4. **C.** Diagram of the *OitaAP2* cDNA. Gray boxes represent UTRs; box with transverse lines represents the alternatively spliced fragment; arrows indicate the primer position and direction (red, forward; black, reverse).

These results seem to confirm the existence of an alternative splicing resulting in two isoforms of different size named *OitaAP2* and *OitaAP2_ISO*. The deletion of 105 bp that distinguishes between the two isoforms is in frame with the main ORF. The whole cDNA sequence of the *OitaAP2* cDNA of *O. italica* is 2264 bp, whereas the size of the *OitaAP2_ISO* cDNA is 2164 bp. Both *OitaAP2* and *OitaAP2_ISO* include identical 5'- and 3'-UTR of 560 and 273 bp, respectively.

BLAST analysis revealed the highest similarity with the *AP2*-like locus of the orchid *Dendrobium crumenatum* (72% nucleotide and 60% amino acid identity). The amino acid alignment of the virtually translated sequence of the *OitaAP2* (476 residues) and *OitaAP2_ISO* (441 residues) cDNA with the AP2-like sequences of different species retrieved from GenBank shows the presence of the two conserved AP2 domains (AP2R1 and AP2R2). In addition, upstream of the AP2 domains there are the canonical Motifs 1 and 2 and a nuclear localizing signal. A short conserved amino acid stretch, located at the C-terminus, corresponds to the target site of the microRNA mir172 on the *AP2* mRNA (**Fig. 19A**). All these features and the topology of the ML tree (**Fig. 19B**) indicate that the *OitaAP2* locus of *O. italica* is member of the *AP2*-like gene subfamily.

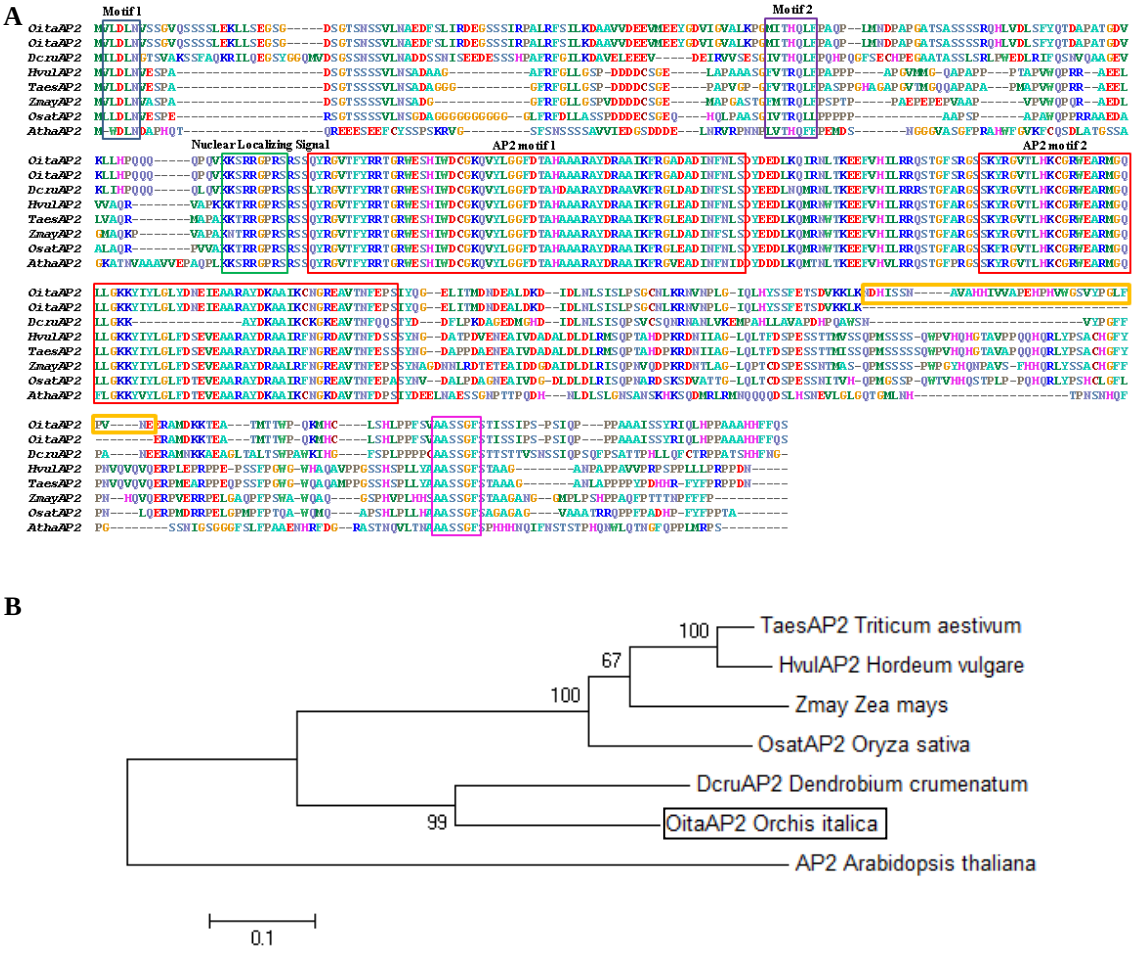


Figure 19: Amino acid alignment of the OitaAP2 and OitaAP2_ISO sequences with AP2-like sequences of other species (**A**) and the resulting maximum likelihood (ML) tree (**B**).

Both the *OitaAP2* and *OitaAP2_ISO* cDNAs present the conserved sequence of 21 bp that represents the target site of mir172 on the mRNA. In *Arabidopsis*, mir172 represents a negative post-transcriptional regulator of *AP2*: it cleaves the *AP2* mRNA and acts predominantly determining translational inhibition (*Aukerman and Sakai, 2003; Chen, 2004; Schwab et al., 2005*). This regulatory mechanism seems to be conserved among the plant species (*Lauter et al., 2005; Chuck et al., 2007; Varkonyi-Gasic et al., 2012*) . Using a modified 5'RACE reaction (*Llave et al., 2002*) the mir172 5'-end cleavage product of the *OitaAP2/OitaAP2_ISO* mRNA was successfully amplified (**Fig. 20A**). Cloning and sequencing of this fragment (~290 bp) revealed the position of the cleavage site within the mir172 target site (**Fig. 20B**). This result demonstrates that in orchids the regulatory mechanism that determines the translational repression of *OitaAP2* through mir172 is conserved.

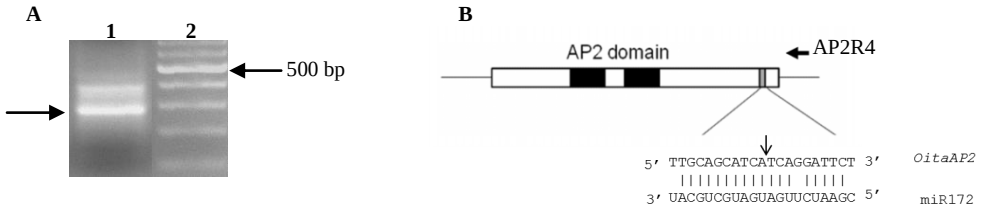


Figure 20: (**A**) 1, PCR amplification of the 5'-end mir172 cleavage product of the *OitaAP2/OitaAP2_ISO* mRNA; 2, 100 bp ladder. (**B**) (Upper) Diagram of the *OitaAP2* cDNA. The gray box indicates the mir172 target site. The position of the reverse primer AP2R4 is also indicated. (Lower) Sequence of the cleaved *OitaAP2/OitaAP2_ISO* mRNA aligned with mir172. The arrow indicates the cleavage site.

Genomic structure of the OitaAP2 locus

Based on the differential splicing observed for the *OitaAP2* mRNA, it was of crucial relevance to characterize the genomic organization of the *OitaAP2* locus of *O. italica*. In order to evaluate and to compare the structure of the *OitaAP2* gene with that of known *AP2*-like genes, the *OitaAP2* locus was amplified from genomic DNA using different primer pairs. Sequence comparison of the *OitaAP2* and *OitaAP2_ISO* cDNA with the genomic sequence of the *OitaAP2* locus revealed the presence of 10 exons and 9 introns (**Fig. 21**).

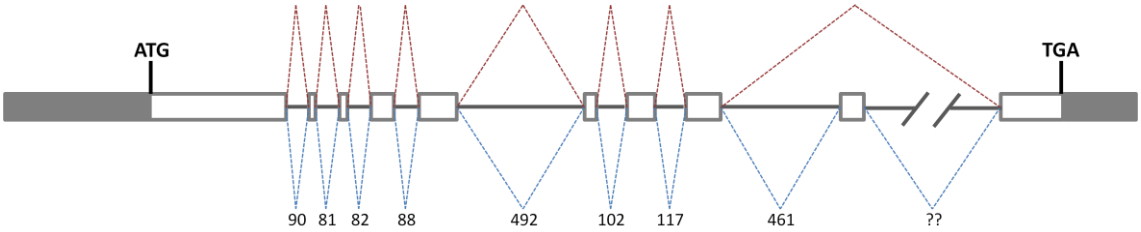


Figure 21: Genomic organization of the locus *OitaAP2*. Gray boxes, white boxes, and lines show UTRs, exons and introns, respectively. The numbers indicate the lengths of introns in bp. Dotted lines indicate the two alternative splicing events that produce the *OitaAP2* (lower) and *OitaAP2_ISO* (upper) mRNA.

This gene structure appears conserved for the *AP2* genes of *Arabidopsis*, grapevine (*Velasco et al., 2007*) and apple (*Velasco et al., 2010*), all constituted by 10 exons and 9 introns. The size of the exons and the position of the introns is broadly conserved among the *AP2*-like genes. All the introns of the *OitaAP2* gene have a quite small size (spanning from 81 to 492 bp), the only exception represented by the intron 9 that is unusually larger than the others (>10,000 bp). Currently, intron 9 of the *OitaAP2* locus is under sequencing. All the introns identified present the canonical 5'GT and 3'AT, with the only exception of the intron 4 that presents a not canonical 5'GC donor splicing site.

CENSOR analysis of the intron sequences showed traces of mobile elements belonging to class I and II transposable elements within the longest introns 5 and 8 of *OitaAP2* (**Tab. 8**). This result further confirms the abundance of mobile elements within the orchid genomes, as detected for the *OitaAG* and *OitaSTK* genes.

Table 8: Results of the CENSOR analysis conducted on the introns of the *OitaAP2* locus. Sim, nucleotide similarity; Length is expressed in bp.

Gene	Intron	Length	Element	Class	Sim
OitaAP2	5	492	Chapaev-16_ HM	Chapaev	0.7586
	8	461	Copia-10_TC-I	Copia	0.8043

Nucleotide comparison of the genomic sequence of the *OitaAP2* locus with the *OitaAP2* and *OitaAP2_ISO* cDNA reveals that the isoform *OitaAP2_ISO* is produced through the differential splicing of the intron 8, exon 9 and intron 9 that preserves the correct reading frame (**Fig. 21**).

To date, alternative splicing of the *AP2* locus has been documented only recently in *Actinidia deliciosa* (kiwifruit) (*Varkonyi-Gasic et al., 2012*). Similarly to *O. italica*, the region involved in the alternative splicing of the *AdAP2* locus of kiwifruit spans from exon 8 to 10 and includes an unusually large intron. However, the alternative kiwifruit isoform *AdAP2Δ*, whose functional significance is not clear, retains an additional fragment (named exon 9a) that alters the main reading frame and introduces a stop codon and a polyadenilation site, excluding the mir172 target site from the transcript.

Expression pattern of the OitaAP2 gene

Real-time RT-PCR experiments have been performed to gain insight into the role of the *OitaAP2* gene in *O. italica*. Isoform-specific reverse primers were used to distinguish between the expression profile of *OitaAP2* and *OitaAP2_ISO* in floral tissues from early and late inflorescence. The melting curve plots demonstrated the effectiveness of the primers used. The mean PCR efficiency for each gene (the target isoforms *OitaAP2* and *OitaAP2_ISO* and the endogenous control *OitaAct*) showed comparable values. Both the isoforms result ubiquitously expressed in the organ of perianth in early and late inflorescence and are absent in the column (**Fig. 22**). More in detail, in the early inflorescence *OitaAP2* results more expressed in the outer and inner tepals than in the lip, whereas in the same tissues *OitaAP2_ISO* is expressed at comparable levels, always significantly lower than those detected for *OitaAP2*. In the late inflorescence the expression level of *OitaAP2* decreases in outer tepals and increases in inner tepals and lip, whereas the amount of *OitaAP2_ISO* mRNA increases in outer and inner tepals and, more strongly, in lip (**Fig. 22**).

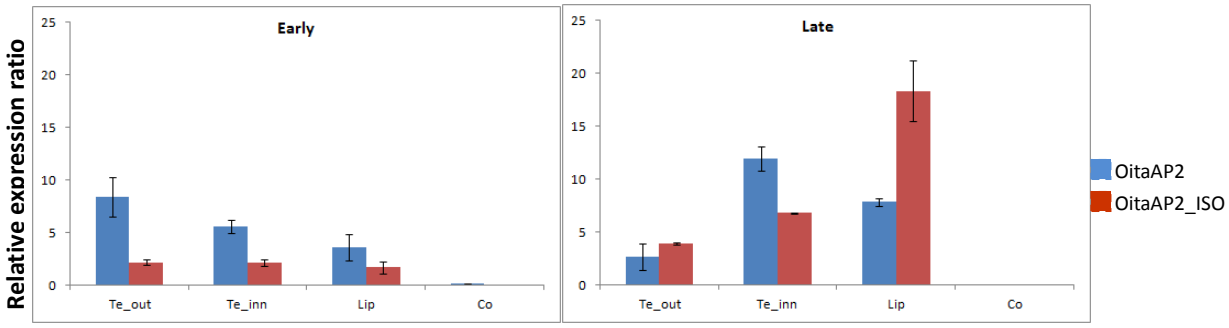


Figure 22: Relative expression ratio of the *OitaAP2* and *OitaAP2_ISO* isoforms in floral tissues from early (before anthesis) and late (after anthesis) inflorescence. Te_out, Outer tepals; Te_inn, Inner tepals; Co, Column.

The relative expression ratio of *OitaAP2* and *OitaAP2_ISO* has been evaluated also in ovary tissue of immature and mature inflorescence before and after pollination (3, 7 and 10 days after pollination) (**Fig. 23**). Both the isoforms are expressed only in the mature ovary before pollination, where the *OitaAP2_ISO* is significantly more abundant than that of *OitaAP2*.

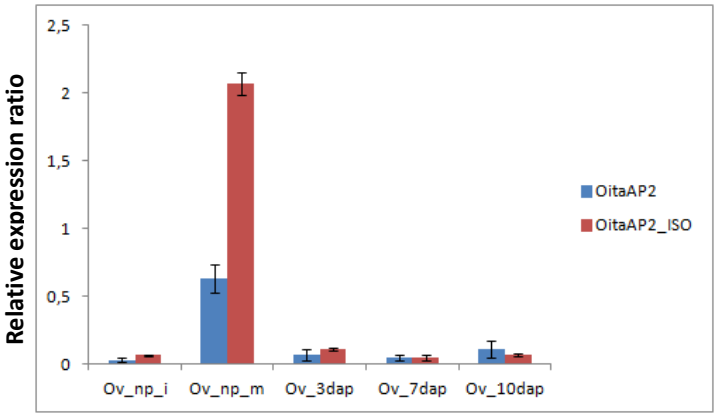


Figure 23: Relative expression ratio of *OitaAP2* and *OitaAP2_ISO* in the ovary tissue. Ov_i, Immature ovary; Ov_m, Mature ovary; Ov_3dap, Ovary 3 days after pollination; Ov_7dap, Ovary 7 days after pollination; Ov_10dap, Ovary 10 days after pollination.

The relative expression analysis of *OitaAP2* and *OitaAP2_ISO* in vegetative tissues reveals very low amounts of the *OitaAP2* mRNA, with a slight increase in stem tissue. Contrastingly, the expression of the *OitaAP2_ISO* isoform is significantly higher in root and stem tissue (**Fig. 24**).

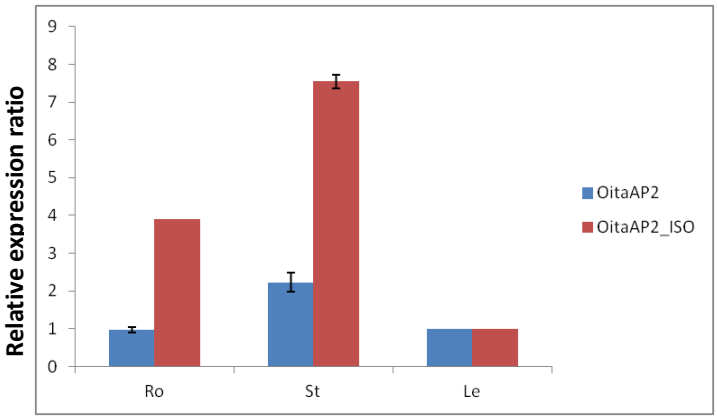


Figure 24: Relative expression ratio of *OitaAP2* and *OitaAP2_ISO* in the vegetative tissues. Ro, Root; St, Stem; Le, Leaf.

Previous studies conducted in different species, e.g. *Arabidopsis* (*Jofuku et al., 1994; Würschum et al., 2006; Zhao et al., 2007*), *Petunia* (*Maes et al., 2001*), *Nymphaea* (*Luo et al., 2012*) reported the expression of the *AP2* gene not only in the perianth, but also in the reproductive organs, even though with different levels of mRNA accumulation (*Alvarez-Venegas et al., 2003*). In *Antirrhinum* and maize the expression of the *AP2* homolog genes is restricted to specific domains (*Chuck et al., 1998; Keck et al., 2003*). A more recent study demonstrated that during the early stage of flower development of *Arabidopsis AP2* is expressed specifically within the future perianth and stamen primordia as well as in vegetative tissue. Later, *AP2* is expressed in petals, stamens and developing ovules (*Wollmann et al., 2010*).

In contrast with the pattern described in the orchid *Dendrobium crumenatum*, where the *DcAP2* gene seems to be expressed in all floral organs and without differential splicing (*Xu et al., 2006*), in *O. italica* the expression pattern of both the *AP2* isoforms *OitaAP2* and *OitaAP2_ISO* is restricted to outer and inner tepals and lip, being absent in the fused reproductive organs (column), confirming the role of the *OitaAP2* gene into the development and maintenance of the perianth organs, as expected for a class A gene. The different expression level of the two isoforms, in particular in the lip from the late inflorescence, where it is detectable the highest expression of the *OitaAP2_ISO* isoform, might reflect the possible functional partition of the two isoforms in development and maintenance of the perianth organs. The expression profile detected in mature ovary and

in vegetative tissues seems to confirm a not redundant function of *OitaAP2* and *OitaAP2_ISO*, with the latter specifically involved in ovary formation (but not in post-pollination processes) and in root and stem specific functions. However, further studies are needed to demonstrate this hypothesis and to verify the presence of two *AP2* isoforms in other orchid species.

Comparative expression analysis of the OitaAP2 isoforms, mir172 and OitaAG

Based on the ABCDE model, in *Arabidopsis* the class A *AP2* and the class C *AG* genes drive the formation of perianth and reproductive organs, respectively, acting in an antagonistic manner. *AP2* is a negative regulator of *AG* and its function is regulated by *miR172* through the RNA cleavage. In order to evaluate the relationship existing among *OitaAP2*, *OitaAG* and mir172 in *O. italica*, their relative expression pattern in different floral tissues was compared (**Fig. 25**).

In the tissues from early inflorescence, the expression profile of mir172 is similar to the pattern observed for *OitaAG* and clearly complementary to that observed for *OitaAP2*.

In particular, miR172 appears to be expressed mainly in the column from the early inflorescence and almost absent in the other tissues. This pattern fully agrees with the repressive action of mir172 on *OitaAP2* that, in turns, represses *OitaAG*. In the floral tissues from the late inflorescence, the expression pattern of *OitaAG* is still opposite to that observed for *OitaAP2*; however, miR172 is expressed at very low levels in all tissues, including the column. This behavior suggests that in *O. italica* the inhibitory role of miR172 on the *OitaAP2* transcripts is exerted until the stages preceding the anthesis, whereas the absence of *OitaAP2* mRNA in the column after anthesis could be related to different regulatory mechanisms. In both stages (before and after anthesis) the absence of *OitaAP2* within column permits the expression of the *OitaAG* gene.

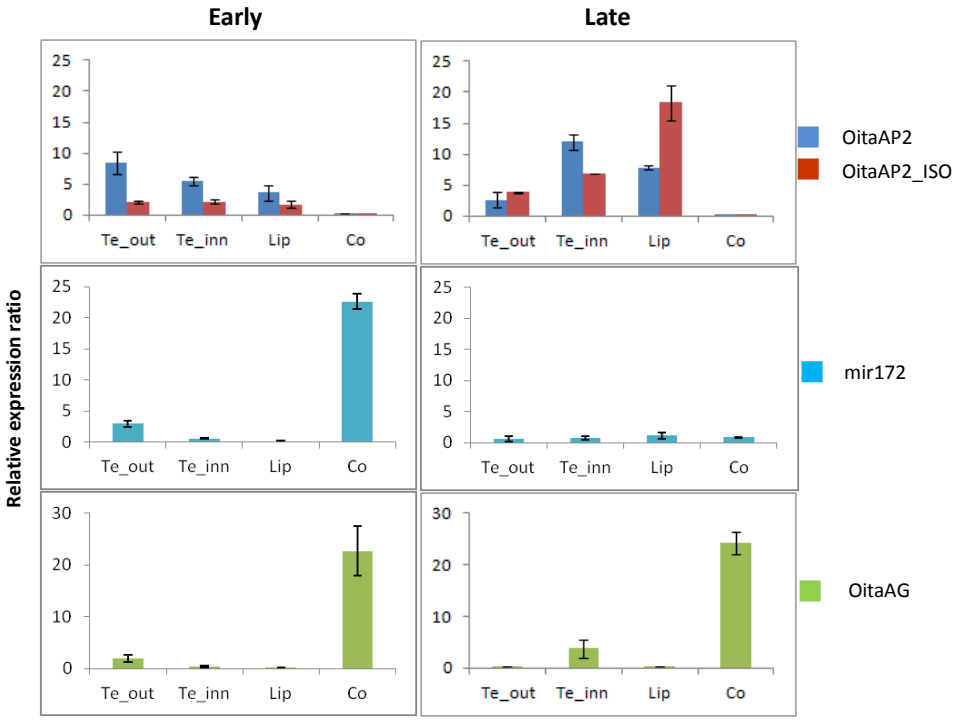


Figure 25: Comparison among the relative expression ratio of *OitaAP2*, *OitaAP2_ISO*, miR172 and *OitaAG* in floral tissues from early (before anthesis) and late (after anthesis) inflorescence. Te_out, Outer tepals; Te_inn, Inner tepals; Co, Column.

In the ovary tissue, mir172 is weakly detectable in the immature ovary, whereas it is expressed from pre-pollinated mature ovary to 10 days after pollination, as observed in *Arabidopsis* (*Wollmann et al., 2010*), showing its possible involvement in ovary development. However, in the ovary tissue of *O. italica* the mir172 expression pattern is not fully overlapping with that observed for *OitaAG* and it is not fully complementary to that observed for *OitaAP2* (**Fig. 26**). These results suggest that also during the ovary maturation the inhibition of *OitaAP2* might not be directly realized through the mir172 mediated cleavage.

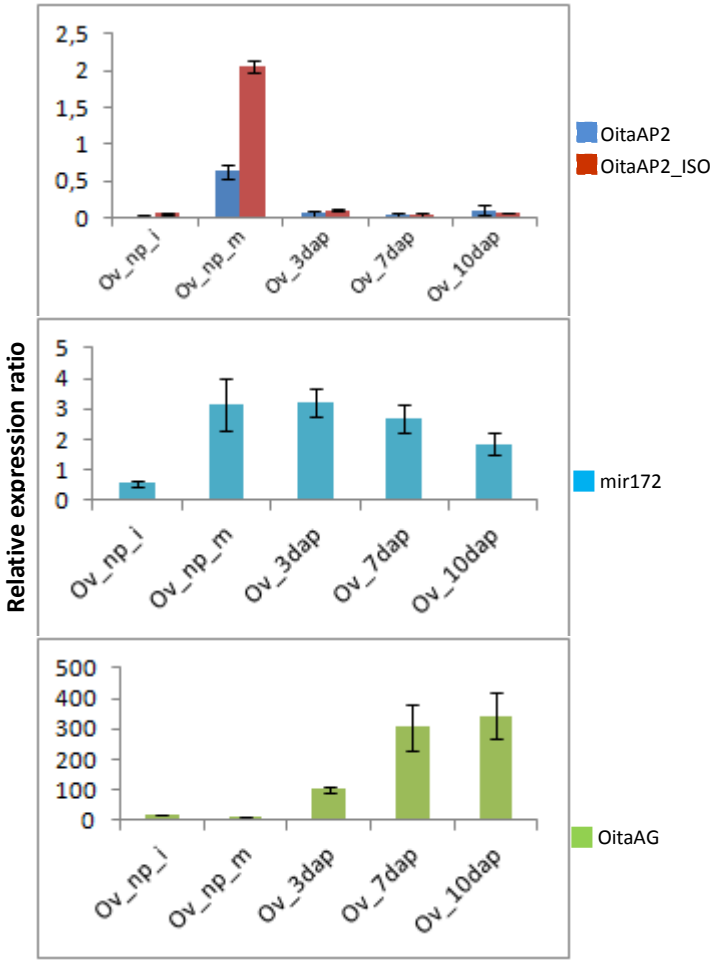


Figure 26: Comparison among the relative expression ratio of *OitaAP2*, *OitaAP2_ISO*, mir172 and *OitaAG* in ovary tissue. Ov_i, Immature ovary; Ov_m, Mature ovary; Ov_3dap, Ovary 3 days after pollination; Ov_7dap, Ovary 7 days after pollination; Ov_10dap, Ovary 10 days after pollination.

High throughput sequencing of microRNA libraries from floral bud and leaf

In the last few years, increasing evidences have highlighted the role of plant microRNAs in the regulation of different pathways, such as signal transduction, hormone biosynthesis, response to stress and flower development (e.g.: *Chen, 2009; Axtell, 2013*). Although many miRNA families have been identified in different plant species, they remain largely unknown in orchids, with the only exception of *Phalaenopsis equestris* (*An et al., 2011; An and Chan, 2012*). In order to characterize miRNAs from *O. italica*, miRNA libraries obtained from floral bud and leaf tissue were constructed and sequenced using the 454 high-throughput sequencing technology. The establishment of an *O. italica* miRNA database could be the foundation for comparative and functional studies aimed at elucidating the role of these molecules in non-model plant species.

After sequencing, the first *in silico* processing consisted in adaptor trimming and filtering of the raw reads using the online platforms GALAXY and UEA sRNA, producing a total of 608,141 and 415,040 reads from floral bud and leaf, respectively (**Tab. 9**). Standalone BLAST (E-value 0.0001) revealed that the two libraries share 361,853 reads. The reads were then clustered based on their perfect sequence identity (collapsing) producing 67,242 and 47,608 unique reads from floral bud and leaf, respectively. However, standalone BLAST revealed the persistence of contaminating adaptor sequences within the reads, mainly due to the concatameric nature of the two libraries (see **Fig. 8**, Materials and Methods). In facts, the GALAXY and UEA sRNA platforms eliminate the adaptor sequences from the 5'- and 3'-terminus of the reads, but not from the central regions. In order to completely remove the adaptor sequences from the reads, a manual cleaning procedure was applied. As result, the number of effective reads was enormously reduced (1394 and 1010 from floral bud and leaf, respectively). The reads distribution length ranged between 16 and 32 nucleotides, with a peak at 16 in both tissues, an unusually short size given the mean plant miRNA length of 24 nucleotides (**Fig. 27**).

Table 9: Reads obtained from floral bud and leaf tissue of *O.italica* with 454 high-throughput sequencing.

	Leaf	**Floral bud**
Reads after adaptor trimming and filtering (min=18)	415.040	608.141
Total nucleotides	13.805.641	22.583.520
Average length	33	37
Total reads highly represented (>100)	306.999	455.258
Different reads highly represented (>100)	320	313
Total reads averagely represented (50<r<99)	14.771	19.271
Different reads averagely represented (50<r<99)	215	277
Total basal reads (18<r<49)	19.492	26.222
Different basal reads (18<r<49)	686	894
Reads after manual cleaning	1349	1010

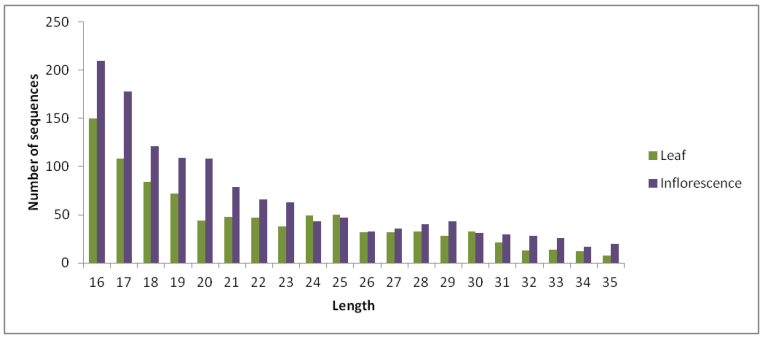


Figure 27: Analysis of the reads distribution length in leaf tissue and floral bud of *O. italica***.**

Mature and stem and loop plant miRNA sequences were downloaded from miRBase (v. 18), the main microRNA database, and standalone BLAST was performed to search for sequence similarity between the known plant miRNA sequences and the reads of the floral bud and leaf libraries of *O. italica*. The analysis of the floral bud library revealed 315 positive matches against the known miRNA mature sequence database and 813 against the stem and loop database, corresponding to 17 different families of known miRNAs (**Fig. 28**). The analysis of the leaf library showed 103 positive matches against the known miRNA mature sequence database and 763 against the stem and loop database, corresponding to 10 different families of known miRNAs (**Fig. 29**). All the positive matches have 100% sequence identity; however, they are no longer than 11 nucleotides.

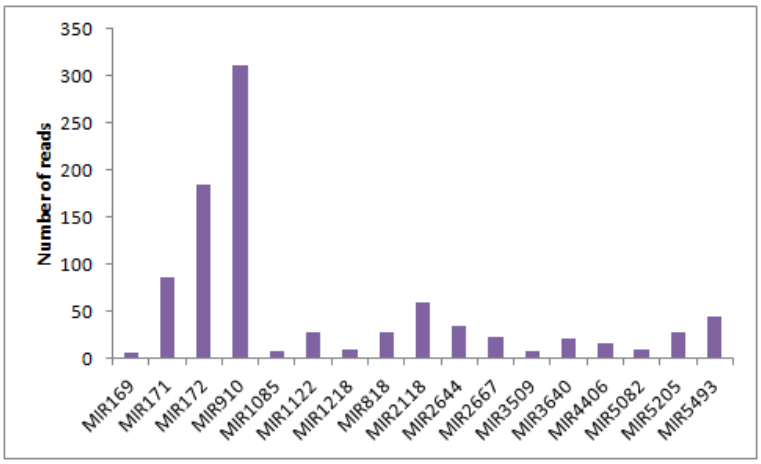


Figure 28: Conserved miRNA families detected by BLAST search of the floral bud reads of *O. italica* against the plant known miRNA mature database.

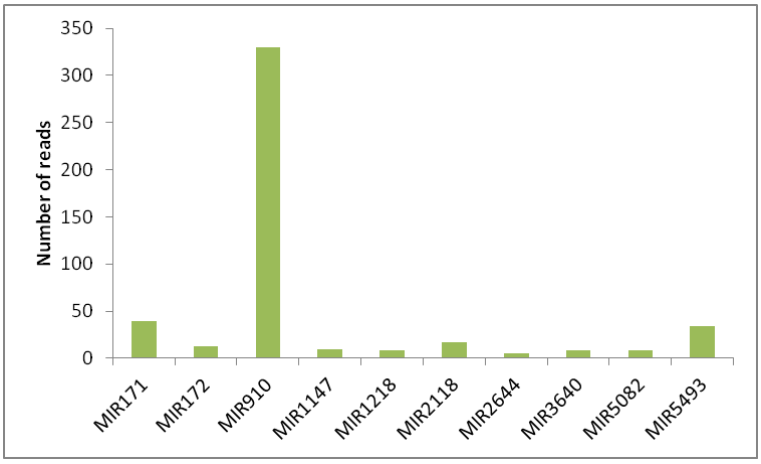


Figure 29: Conserved miRNA families detected by BLAST search of the leaf reads of *O. italica* against the plant known miRNA mature database.

Taken together, the small mean size of the reads, the low number of conserved miRNA families detected and the small sequence length of the positive matches indicate that both the libraries are partial and not fully representative of the real microRNA population of *O. italica*, possibly due to the concatamerization step during the library construction. Currently, a new library generation and sequencing approach is in course on floral bud and leaf of *O. italica* based on the Illumina technology that bypass the concatamerization step and produces reads by direct sequencing of the monomeric small RNAs.

REFERENCES

Aceto S., Caputo P., Cozzolino S., Gaudio L., Moretti A. (1999a). Phylogeny and evolution of *Orchis* and allied genera based on ITS DNA variation: morphological gaps and molecular continuity. *Mol Phylogenet Evol* 13(1): 67-76.

Aceto S., Cozzolino S., Gaudio L., Nazzaro R., De Luca P. (1999b). Pollination flow in hybrid formation between *Orchis morio* and *Orchis papilionacea* (Orchidaceae) in two different habitats. *Int J Plant Sci* 160(6): 1153-56.

Aceto S., Gaudio L. (2011). The MADS and the beauty: genes involved in the development of or wers. *Curr Genomics* 12(5): 342-56.

Aceto S., Montieri S., Sica M., Gaudio L. (2007). Molecular evolution of the *OrcPI* locus in natural populations of Mediterranean orchids. *Gene* 392(1-2): 299-305.

Alvarez-Venegas R., Pien S., Sadder M., Witmer X., Grossniklaus U., Avramova Z. (2003). *ATX-1*, an *Arabidopsis* homolog of *trithorax*, activates flower homeotic genes. *Curr Biol* 13: 627-37.

Ambros V. (2001). microRNAs: tiny regulators with great potential. *Cell* 107(7): 823-26.

An F.M., Chan M.T. (2012). Transcriptome-wide characterization of miRNA-directed and non-miRNA-directed endonucleolytic cleavage using degradome analysis under low ambient temperature in *Phalaenopsis aphrodite* subsp. *formosana*. *Plant Cell Physiol* 53(10):1737-50.

An F.M., Hsiao S.R., Chan M.T. (2011). Sequencing-based approaches reveal low ambient temperature-responsive and tissue-specific microRNAs in *Phalaenopsis* orchid. *PLoS One* 6(5): e18937.

Aukerman M.J., Sakai H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. *Plant Cell* 15(11): 2730-41.

Axtell M.J. (2013). Classification and comparison of small RNAs from plants. *Annu Rev Plant Biol* DOI 10.1146/annurev-arplant-050312-120043.

Bailey T.L., Boden M., Buske F.A., Frith M., Grant C.E., Clementi L., Ren J., Li W.W., Noble W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37: W202-W208.

Bao Z., Franks R.G., Levin J.Z., Liu Z. (2004). Repression of *AGAMOUS* by *BELLRINGER* in floral and inflorescence meristems. *Plant Cell* 16(6): 1478-89.

Benlloch R., Roque E., Ferrándiz C., Cosson V., Caballero T., Penmetsa R.V., Beltrán J.P., Cañas L.A., Ratet P., Madueño F. (2009). Analysis of B function in legumes: PISTILLATA proteins do not require the PI motif for floral organ development in *Medicago truncatula*. *Plant J* 60(1): 102-11.

Bowman J.L., Alvarez J., Weigel D., Meyerowitz E.M., Smyth D.R. (1993). Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development* 119: 721-43.

Bowman J.L., Smyth D.R., Meyerowitz E.M. (1991). Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* 112: 1-20.

Busch M.A., Bomblies K., Weigel D. (1999). Activation of a floral homeotic gene in *Arabidopsis*. *Science* 285: 585-87.

Cameron D.D., Johnson I., Leake J.R., Read D.J. (2007). Mycorrhizal acquisition of inorganic phosphorus by the green-leaved terrestrial orchid *Goodyera repens*. *Ann Bot* 99: 831–34.

Cameron K.M., Chase M.W., Whitten W.M., Kores P.J., Jarrell D.C., Albert V.A., Yukawa T., Hills H.G., Goldman. D.H. (1999). A phylogenetic analysis of the Orchidaceae: evidence from *rbcL* nucleotide sequences *Am J Bot* 86: 208-24.

Cantone C., Gaudio L., Aceto S. (2011). The *PI/GLO*-like locus in orchids: duplication and purifying selection at synonymous sites within Orchidinae (Orchidaceae). *Gene* 481(1): 48-55.

Cantone C., Sica M., Gaudio L., Aceto S. (2009). The *OrcPI* locus: genomic organization, expression pattern, and noncoding regions variability in *Orchis italica* (Orchidaceae) and related species. *Gene* 434(1-2): 9-15.

Causier B., Bradley D., Cook H., Davies B. (2009). Conserved intragenic elements were critical for the evolution of the floral C-function. *Plant J* 58(1): 41-52.

Chang W.C., Lee T.Y., Huang H.D., Huang H.Y., Pan R.L. (2008). PlantPAN: Plant promoter analysis navigator, for identifying combinatorial *cis*-regulatory elements with distance constraint in plant gene groups. *BMC Genomics* 9: 561.

Chang Y.Y., Kao N.H., Li J.Y., Hsu W.H.,. Liang Y.L., Wu J.W., Yang C.H. (2010). Characterization of the possible roles for B class MADS box genes in regulation of perianth formation in orchid. *Plant Physiol* 152: 837-53.

Chen X. (2004). A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. *Science* 303: 2022-25.

Chen X. (2009). Small RNAs and their roles in plant development. *Annu Rev Cell Dev Biol* 25: 21-44.

Chen Y.Y., Lee P.F., Hsiao Y.Y., Wu W.L., Pan Z.J., Lee Y.I., Liu K.W., Chen L.J., Liu Z.J., Tsai W.C. (2012). C- and D-class MADS-box genes from *Phalaenopsis equestris* (Orchidaceae) display functions in gynostemium and ovule development. *Plant Cell Physiol* l53(6): 1053-67.

Chuck G., Meeley R., Irish E., Sakai H., Hake S. (2007). The maize *tasselseed4* microRNA controls sex determination and meristem cell fate by targeting *Tasselseed6/indeterminate spikelet1*. *Nat Genet* 39(12): 1517-21.

Chuck G., Meeley R.B., Hake S. (1998). The control of maize spikelet meristem fate by the *APETALA2*-like gene *indeterminate spikelet1*. *Genes Dev* 12: 1145-54.

Coen E.S., Meyerowitz E.M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353(6339): 31-37.

Colombo L., Franken J., Koetje E., Went J.V., Dons H.J.M., Angenent G.C., Tunenai A.J.V. (1995). The petunia MADS box gene *FBPI7* determines ovule identity. *Plant Cell* 7: 1859-68.

Cozzolino S., Widmer A. (2005). Orchid diversity: an evolutionary consequence of deception?. *Trends Ecol Evol* 20(9): 487-94.

De Bodt S., Raes J., Florquin K., Rombauts S., Rouzé P., Theissen G., Van de Peer Y. (2003). Genome wide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. *J Mol Evol* 56 (5): 573-86.

De Bodt S., Theissen G., Van de Peer Y. (2006). Promoter analysis of MADS-box genes in eudicots through phylogenetic footprinting. *Mol Biol Evol* 23(6): 1293-303.

Deyholos M.K., Sieburth L.E. (2000). Separable whorl-specific expression and negative regulation by enhancer elements within the *AGAMOUS* second intron. *Plant Cell* 12(10): 1799-810.

Dinh T.T., Girke T., Liu X., Yant L., Schmid M., Chen X. (2012).The floral homeotic protein APETALA2 recognizes and acts through an AT-rich sequence element. *Development* 139(11): 1978-86.

Ditta G., Pinyopich A., Robles P., Pelaz S., Yanofsky M.F. (2004). The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr Biol* 14(21): 1935-40.

Doyle J.J., Doyle J.L. (1987). A rapid DNA isolation procedure for small amounts of leaf tissue. *Phytochem Bull* 19: 11-15.

Drews G.N., Bowman J.L., Meyerowitz E.M. (1991). Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the APETALA2 product. *Cell* 65: 991-1002.

Edgar R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.

Freeling M., Thomas B.C. (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res* 16(7): 805-14.

Gil-Humanes J., Pistón F., Martín A., Barro F. (2009). Comparative genomic analysis and expression of the *APETALA2*-like genes from barley, wheat, and barley-wheat amphiploids. *BMC Plant Biology* 9: 66.

Giraudat J., Hauge B.M., Valon C., Smalle J., Parcy F., Goodman H.M. (1992). Isolation of the *Arabidopsis ABI3* gene by positional cloning. *Plant Cell* 4(10): 1251-61.

Gramzow L., Theissen G. (2010). A hitchhiker's guide to the MADS world of plants. *Genome Biol* 11(6): 21.

Gustafson-Brown C., Savidge B., Yanofsky M.F. (1994). Regulation of the *Arabidopsis* floral homeotic gene *APETALA1*. *Cell* 76(1): 131-43.

Hall T.A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acid Symp Ser* 41: 95-98.

Henschel K., Kofuji K., Hasebe M., Saedler H., Munster T, Theissen G. (2002). Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. *Mol Biol Evol* 19(6): 801-14.

Hong R.L., Hamaguchi L., Busch M.A., Weigel D. (2003). Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing. *Plant Cell* 15(6): 1296-1309.

Honma T., Goto K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409(6819): 525-29.

Hsu C.C., Chung Y.L., Chen T.C., Lee Y.L., Kuo Y.T., Tsai W.C., Hsiao Y.Y., Chen Y.W., Wu W.L., Chen H.H. (2011). An overwiew of the *Phalaenopsis* orchid genome through BAC end sequence analysis. *BMC Plant Biol* 11: 3.

Hsu H.F., Hsieh W.P., Chen M.K., Chang Y.Y., Yang C.H. (2010). 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 Cell Physiol* 51(6): 1029-45.

Hu L., Liang W., Yin C., Cui X., Zong J., Wang X., Hu J., Zhang D. (2011). Rice MADS3 regulates ROS homeostasis during late anther development. *Plant Cell* 23: 515-33.

Ikeda M., Mitsuda N., Ohme-Takagi M. (2009). *Arabidopsis* WUSCHEL is a bifunctional transcription factor that acts as a repressor in stem cell regulation and as an activator in floral patterning. *Plant Cell* 21(11): 3493-505.

Irish V.F., Litt A. (2005). Flower development and evolution: gene duplication, diversification and redeployment. *Curr Opin Genet Dev* 15: 454-60.

Irish V.F., Sussex I.M. (1990). Function of the *apetala-1* gene during *Arabidopsis* floral development. *Plant Cell* 2: 741-53.

Jofuku K.D., den Boer B.G.W., Van Montagu M., Okamuro J.K. (1994). Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. *Plant Cell* 6: 1211-25.

Kasschau K.D., Xie Z., Llave E.A.C., Chapman E.J., Krizan K.A., Carrington J.C. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function extent. *Dev Cell* 4: 205-17.

Kaufmann K., Melzer R., Theissen G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein

interactions and network evolution in land plants. *Gene* 347: 183-98.

Keck E., McSteen P., Carpenter R., Coen E. (2003). Separation of genetic functions controlling organ identity in flowers. *EMBO J* 22: 1058-66.

Keeney S. (2011). Long-PCR amplification of human genomic DNA. *Methods Mol Biol* 688: 67-74.

Kim S.Y., Yun P.Y., Fukuda T., Ochiai T., Yokoyama J., Kameya T., Kanno A. (2007). Expression of a *DEFICIENS*-like gene correlates with the differentiation between sepal and petal in the orchid *Habenaria radiata* (Orchidaceae). *Plant Sci* 172(2): 319-26.

Klevebring D., Street N.R., Fahlgren N., Kasschau K.D., Carrington J.C, Lundeberg J., Jansson S. (2009). Genome-wide profiling of *Populus* small RNAs. *BMC Genomics* 10: 620.

Kofuji R., Sumikawa N., Yamasaki M., Kondo K., Ueda K., Ito M., Hasebe M. (2003). Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Mol Biol Evol* 20(12): 1963-77.

Kohany O., Gentles A.J., Hankus L., Jurka J. (2006). Annotation, submission and screening of repetitive elements in Repbase: Repbase Submitter and Censor. *BMC Bioinformatics* 7: 474.

Komaki M.K., Okada K., Nishino E., Shimura Y. (1988). Isolation and characterization of novel mutants of *Arabidopsis thaliana* defective in flower development. *Development* 104: 195-203.

Kooiker M., Airoldi C.A., Losa A., Manzotti P.S., Finzi L., Kater M.M, Colombo L.. (2005). BASIC PENTACYSTEINE1, a GA binding protein that induces conformational changes in the regulatory region of the homeotic *Arabidopsis* gene *SEEDSTICK*. *Plant Cell* 17: 722-29.

Kramer E.M., Dorit R.L., Irish V.F. (1998). Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. *Genetics* 149: 765-83.

Kramer E.M., Jaramillo M.A., Di Stilio V.S. (2004). Patterns of gene duplication and functional evolution during the diversification of the *AGAMOUS* subfamily of MADS box genes in angiosperms. *Genetics* 166: 1011-23.

Lau N.C., Lim L.P., Weinstein E.G., Bartel D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294(5543): 858-62.

Lauter N., Kampani A., Carlson S., Goebel M., Moose S.P. (2005). microRNA172 down-regulates *glossy15* to promote vegetative phase change in maize. *Proc Natl Acad Sci USA* 102(26): 9412-17.

Lee S., Woo Y.M., Ryu S.I., Shin Y.D., Kim W.T., Park K.Y., Lee I.J., An G. (2008). Further characterization of a rice *AGL12* group MADS-box gene, *OsMADS26*. *Plant Physiol* 147: 156-68.

Lenhard M., Bohnert A., Jurgens G., Laux T. (2001). Termination of stem cell maintenance in *Arabidopsis* floral meristems by interactions between WUSCHEL and AGAMOUS. *Cell* 105: 805-14.

Levin D.A. (1983). Polyploidy and novelty in flowering plants. *Am Nat* 122: 1-25.

Llave C., Xie Z., Kasschau K.D., Carrington J.C. (2002). Cleavage of *Scarecrow*-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* 297: 2053-56.

Lohmann J.U., Hong R.L., Hobe M., Busch M.A., Parcy F., Simon R., Weigel D. (2001). A molecular link between stem cell regulation and floral patterning in *Arabidopsis*. *Cell* 105: 793-803.

Luo H., Chen S., Jiang J., Teng N., Chen Y., Chen F. (2012). The *AP2*-like gene *NsAP2* from water lily is involved in floral organogenesis and plant height. *J Plant Physiol* 169(10): 992-98.

Lynch M., Conery J.S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151-55.

Maes T., Van de Steene N., Zethof J., Karimi M., D'Hauw M., Mares G., Van Montagu M., Gerats T. (2001). Petunia

Ap2-like genes and their role in flower and seed development. *Plant Cell* 13: 229-44.

Mlotshwa S., Yang Z., Kim Y., Chen X. (2006). Floral patterning defects induced by *Arabidopsis APETALA2* and microRNA172 expression in *Nicotiana benthamiana*. *Plant Mol Biol* 61: 781-93.

Mondragón-Palomino M., Hiese L., Härter A., Koch M.A., Theissen G. (2009). Positive selection and ancient duplications in the evolution of class B floral homeotic genes of orchids and grasses. *BMC Evolutionary Biology* 9: 81.

Mondragón-Palomino M., Theissen G. (2011). High conservation in the pattern of expression of class B MADS-box genes in the Orchidaceae: testing and revising the 'orchid code'. *Plant J* 66: 1008.

Montieri S., Gaudio L., Aceto S. (2004). Isolation of the *LFY/FLO* homologue in *Orchis italica* and evolutionary analysis in some European orchids. *Gene* 333: 101-09.

Nardmann J., Werr W. (2006). The shoot stem cell niche in angiosperms: expression patterns of *WUS* orthologues in rice and maize imply major modifications in the course of mono- and dicot evolution. *Mol Biol Evol* 23(12):2492-504.

Nei M. (2005). Selectionism and neutralism in molecular evolution. *Mol Biol Evol* 22: 2318-42.

Nei M., Rooney A.P. (2005). Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39: 121-52.

Norman C., Runswick M., Pollock R., Treisman R. (1988). Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* 55: 989-1003.

Ohno S. (1970). Evolution by gene duplication. Springer-Verlag, Heidelberg,Germany.

Parenicová L., de Folter S., Kieffer M., Horner D.S., Favalli C., Busscher J., Cook H.E, Ingram R.M., Kater M.M., Gerco B.D., Angenent C., Colombo L. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* 15: 1538-51.

Passmore S., Maine G.T., Elble R., Christ C., Tye B.K. (1988). *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MAT alpha cells. *J Mol Biol* 204: 593-606.

Pelaz S., Ditta G.S., Baumann E., Wisman E., Yanofsky M.F. (2000). B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* 405: 200-03.

Pelaz S., Gustafson-Brown C., Kohalmi S.E., Crosby W.L., Yanofsky M.F. (2001). APETALA1 and SEPALLATA3 interact to promote flower development. *Plant J* 26(4): 385-94.

Pinyopich A., Ditta G.S., Savidge B., Liljegren S.J., Baumann E., Wisman E., Yanofsky M. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424(6944): 85-88.

Piwarzyk E., Yang Y., Jack T. (2007). Conserved C-terminal motifs of the *Arabidopsis* proteins APETALA3 and PISTILLATA are dispensable for floral organ identity function. *Plant Physiol* 145(4): 1495-505.

Pnueli L., Hareven D., Broday L., Hurwitz C., Lifschitz E. (1994). The *TM5* MADS box gene mediates organ differentiation in the three inner whorls of tomato flowers. *Plant Cell* 6(2): 175-86.

Rhoades M.W., Reinhart B.J., Lim L.P., Burge C.B., Bartel B., Bartel D.P. (2002). Prediction of plant microRNA targets. *Cell* 110: 513-20.

Rounsley S.D., Ditta G.S., Yanofsky M.F. (1995). Diverse roles for MADS-box genes *Arabidopsis* development. *Plant Cell* 7: 1259-69.

Rudall P.J., Bateman R.M. (2002). Roles of synorganisation, zygomorphy and heterotopy in floral evolution: the gynostemium and labellum of orchids and other lilioid monocots. *Biol Rev Camb Philos Soc* 3: 403-41.

Salemme M., Sica M., Gaudio L., Aceto S. (2011). Expression pattern of two paralogs of the *PI/GLO*-like locus during *Orchis italica* (Orchidaceae, Orchidinae) flower development. *Dev Genes Evol* 221(4): 241-46.

Salemme M., Sica M., Gaudio L., Aceto S. (2013). The *OitaAG* and *OitaSTK* genes of the orchid *Orchis italica*: a comparative analysis with other C- and D-class MADS-box genes. *Mol Biol Rep* DOI 10.1007/s11033-012-2426-x.

Schefe J.H., Lehmann K.E., Buschmann I.R., Unger T., Funke-Kaiser H. (2006). Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's CT difference" formula. *J Mol Med* (*Berl*) 84(11): 901-10.

Schmidt R.J., Veit B., Mandel M.A., Mena M., Hake S., Yanofsky M.F. (1993). Identification and molecular characterization of *ZAG1*, the maize homolog of the *Arabidopsis* floral homeotic gene *AGAMOUS*. *Plant Cell* 5(7): 729-37.

Schwab R., Palatnik J.F., Riester M., Schommer C., Schmid M., Weigel D. (2005). Specific effects of microRNAs on the plant transcriptome. *Dev Cell* 8(4): 517-27.

Schwarz-Sommer Z., Huijser P., Nacken W., Saedler H., Sommer H. (1990).Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250(4983): 931-36.

Shi R., Sun Y.H., Zhang X.H., Chiang V.L. (2012). Poly(T) adaptor RT-PCR. *Methods Mol Biol* 822: 53-66.

Skipper M., Johansen L.B., Pedersen K.B., Frederiksen S., Johansen B.B. (2006). Cloning and transcription analysis of an *AGAMOUS* and *SEEDSTICK* ortholog in the orchid *Dendrobium thyrsiflorum* (Reichb. f.). *Gene* 366(2): 266-74.

Smaczniak C., Immink R.G., Angenent G.C., Kaufmann K. (2012). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* 139(17): 3081-98.

Soltis D.E., Ma H., Frohlich M.W., Soltis P.S., Albert V.A., Oppenheimer D.G., Altman N.S., de Pamphilis C., Leebens-Mack J. (2007). The floral genome: an evolutionary history of gene duplication and shifting patterns of gene expression. *Trends Plant Sci* 12(8): 358-67.

Song I.J., Nakamura T., Fukuda T., Yokoyama J., Ito T., Ichikawa H., Horikawa Y., Kameya T., Kanno A. (2006). Spatiotemporal expression of duplicate *AGAMOUS* orthologues during floral development in *Phalaenopsis*. *Dev Genes Evol* 216(6): 301-13.

Stebbins G.L. (1966). Chromosomal variation and evolution. *Science* 152: 1463-69.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10): 2731-39.

Tapia-López R., García-Ponce B., Dubrovsky J.G., Garay-Arroyo A., Pérez-Ruíz R.V., Kim S.H., Acevedo F., Pelaz S., Alvarez-Buylla E.R. (2008). An *AGAMOUS*-related MADS-box gene, *XAL1* (*AGL12*), regulates root meristem cell proliferation and flowering transition in *Arabidopsis*. *Plant Physiol* 146(3): 1182-92.

Theissen G. (2001). Development of floral organ identity: stories from the MADS house. *Curr Opin Plant Biol* 4(1): 75-85.

Theissen G., Melzer R. (2007). Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann Bot* 100(3): 603-19.

Theissen G., Saedler H. (2001). Plant biology. Floral quartets. *Nature* 409(6819): 469-71.

Tremblay R.L., Ackerman J.D., Zimmerman J.K., Calvo R.N. (2005). Variation in sexual reproduction in orchids and

its evolutionary consequences: a spasmodic journey to diversification. *Biol J Linn Soc Lond* 84: 1-54.

Tsai W.C., Lee .PF., Chen H.I., Hsiao Y.Y., Wei W.J., Pan Z.J., Chuang M.H., Kuoh C.S., Chen W.H., Chen H.H. (2005). *PeMADS6*, a *GLOBOSA/PISTILLATA*-like gene in *Phalaenopsis equestris* involved in petaloid formation, and correlated with flower longevity and ovary development. *Plant Cell Physiol* 46(7): 1125-39.

Tzeng T.Y., Chen H.Y., Yang C.H. (2002). Ectopic expression of carpel-specific MADS box genes from lily and lisianthus causes similar homeotic conversion of sepal and petal in *Arabidopsis*. *Plant Physiol* 130(4): 1827-36.

Varkonyi-Gasic E., Lough R.H., Moss S.M., Wu R., Hellens R.P. (2012). Kiwifruit floral gene *APETALA2* is alternatively spliced and accumulates in aberrant indeterminate flowers in the absence of miR172. *Plant Mol Biol* 78(4-5): 417-29.

Velasco R., Zharkikh A., Affourtit J., Dhingra A., Cestaro A., Kalyanaraman A., Fontana P., Bhatnagar S.K., Troggio M., Pruss D., Salvi S., Pindo M., Baldi P., Castelletti S., Cavaiuolo M., Coppola G., Costa F., Cova V., Dal Ri A., Goremykin V., Komjanc M., Longhi S., Magnago P., Malacarne G., Malnoy M., Micheletti D., Moretto M., Perazzolli M., Si-Ammour A., Vezzulli S., Zini E., Eldredge G., Fitzgerald L.M., Gutin N., Lanchbury J., Macalma T., Mitchell J.T., Reid J., Wardell B., Kodira C., Chen Z., Desany B., Niazi F., Palmer M., Koepke T., Jiwan D., Schaeffer S., Krishnan V., Wu C., Chu VT., King ST., Vick J., Tao Q., Mraz A., Stormo A., Stormo K., Bogden R., Ederle D., Stella A., Vecchietti A., Kater MM., Masiero S., Lasserre P., Lespinasse Y., Allan AC., Bus V., Chagne D., Crowhurst R.N., Gleave A.P., Lavezzo E., Fawcett J.A., Proost S., Rouze P., Sterck L., Toppo S., Lazzari B., Hellens R.P., Durel C.E., Gutin A., Bumgarner R.E., Gardiner S.E., Skolnick M., Egholm M., Van de Peer Y., Salamini F., Viola R. (2010). The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat Genet* 42: 833-39.

Velasco R., Zharkikh A., Troggio M., Cartwright D.A., Cestaro A., Pruss D., Pindo M., FitzGerald L.M., Vezzulli S., Reid J., Malacarne G., Iliev D., Coppola G., Wardell B., Micheletti D., Macalma T., Facci M., Mitchell J.T., Perazzolli M., Eldredge G., Gatto P., Oyzerski R., Moretto M., Gutin N., Stefanini M., Chen Y., Segala C., Davenport C., Dematte` L., Mraz A., Battilana J., Stormo K., Costa F., Tao Q., Si-Ammour A., Harkins T., Lackey A., Perbost C., Taillon B., Stella A., Solovyev V., Fawcett JA., Sterck L., Vandepoele K., Grando S.M., Toppo S., Moser C., Lanchbury J., Bogden R., Skolnick M., Sgaramella V., Bhatnagar S.K., Fontana P., Gutin A., Van de Peer Y., Salamini F., Viola R. (2007). A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLOS One* 2: e1326.

Wang S.Y., Lee P.F., Lee Y.I., Hsiao Y.Y., Chen Y.Y., Pan Z.J., Liu Z.J., Tsai W.C. (2011). Duplicated C-class MADS-box genes reveal distinct roles in gynostemium development in *Cymbidium ensifolium* (Orchidaceae). *Plant Cell Physiol* 52(3): 563-77.

Weigel D., Meyerowitz E.M. (1993). Activation of floral homeotic genes in *Arabidopsis*. *Science* 261(5129): 1723-26.

Weigel D., Meyerowitz E.M. (1994). The ABCs of floral homeotic genes. *Cell* 78(2): 203-09.

Wollmann H., Mica E., Todesco M., Long J.A., Weigel D. (2010). On reconciling the interactions between *APETALA2*, miR172 and *AGAMOUS* with the ABC model of flower development. *Development* 137(21): 3633-42.

Würschum T., Gross-Hardt R., Laux T. (2006). *APETALA2* regulates the stem cell niche in the *Arabidopsis* shoot meristem. *Plant Cell* 18: 295-307.

Xu Y., Teo L.L., Zhou J., Kumar P.P., Yu H. (2006). Floral organ identity genes in the orchid *Dendrobium*

crumenatum. *Plant J* 46(1): 54-68.

Yamaguchi T., Lee D.Y., Miyao A., Hirochika H., An G., Hirano H.Y. (2006). Functional diversification of the two C-class MADS box genes *OSMADS3* and *OSMADS58* in *Oryza sativa*. *Plant Cell* 18(1): 15-28.

Yanofsky M.F., Ma H., Bowman J.L., Drews G.N., Feldmann K.A., Meyerowitz E.M. (1990). The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature* 346(6279): 35-39.

Yu D., Kotilainen M., Pöllänen E., Mehto M., Elomaa P., Helariutta Y., Albert V.A., Teeri T.H. (1999). Organ identity genes and modified patterns of flower development in *Gerbera hybrida* (Asteraceae). *Plant J* 17(1): 51-62.

Yu H., Goh C.J. (2001). Molecular genetics of reproductive biology in orchids. *Plant Physiol* 127(4): 1390-93.

Yun P.Y., Ito T., Kim S.Y., Kanno A., Kameya T. (2004). The *AVAG1* gene is involved in development of reproductive organs in the ornamental asparagus *Asparagus virgatus*. *Sex Plant Reprod* 17(1): 1-8.

Zahn L.M., Kong H., Leebens-Mack J.H., Kim S., Soltis P.S., Landherr L.L., Soltis D.E., de Pamphilis C.W., Ma H. (2005). The evolution of the *SEPALLATA* subfamily of MADS-box genes: a pre angiosperm origin with multiple duplications throughout angiosperm history. *Genetics* 169: 2209-23.

Zhao L., Kim Y., Dinh T.T., Chen X. (2007). miR172 regulates stem cell fate and defines the inner boundary of *APETALA3* and *PISTILLATA* expression domain in *Arabidopsis* floral meristems. *Plant J* 51: 840-49.

Zhao S., Fernald R.D. (2005). Comprehensive algorithm for quantitative real-time polymerase chain reaction. *J Comput Biol* 12: 1047-64.

Zobell O., Faigl W., Saedler H., Münster T. (2010). MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. *Mol Biol Evol* 27(5): 1201-11.

APPENDIX

Poster abstracts and scientific publications produced during the PhD course

- **Salemme M.**, Sica M., Gaudio L., Aceto S. (2011) Identification and expression analysis of class C and D MADS-box genes in *Orchis italica* (Orchidaceae). *Joint Meeting AGI – SIBV – SIGA*, Cittadella di Assisi (PG), 19-22 September 2011, 2A.57.
- Sica M., **Salemme M.**, D'Argenio V., Casaburi G., Cantiello P., Gaudio L., Salvatore F., Aceto S. (2012) Isolation of microRNAs from floral bud and leaf tissue of Orchis italica (Orchidaceae). *EMBO Workshop Evolution in the time of Genomics*, Venice (Italy), May 7-9 2012, pag 31.
- **Salemme M.**, Sica M., Gaudio L., Aceto S. (2012) Expression profile of *APETALA2*, mir172 and *AGAMOUS* in the floral tissue of the orchid *Orchis italica*. XII FISV Congress, September 24-27, 2012 Rome, Italy, pag. 115-116.

- **Salemme M.**, Sica M., Gaudio L., Aceto S. (2011) Expression profile of two paralogs of the *PI*/*GLO*-like locus during *Orchis italica* (Orchidaceae, Orchidinae) flower development. *Development Genes and Evolution*, 221(4): 241-246.
- **Salemme M.**, Sica M., Gaudio L., Aceto S. (2013) 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 (in press).* DOI: 10.1007/s11033-012-2426-x.