UNIVERSITY OF NAPLES "FEDERICO II"

PhD in Advanced Biology (XXVIII Cycle)

The inflorescence transcriptome of *Orchis italica*: analysis of coding and non-coding transcripts

PhD Student:
Sofia De Paolo

Supervisor:
Prof. Serena Aceto

PhD Coordinator:
Prof. Luciano Gaudio

Academic year 2015/2016

Alla mia famiglia

Index

Abstract

Orchis italica è un'orchidea selvatica mediterranea appartiene alla famiglia delle Orchidaceae, sottofamiglia Orchidoideae. Sebbene di recente siano stati condotti numerosi studi sui fattori di trascrizione e trascritti non codificanti che regolano lo sviluppo del fiore, il meccanismo molecolare alla base della fioritura nelle orchidee rimane ancora poco chiaro. In questo lavoro è stato utilizzato un approccio di *Next Generation Sequencing* (NGS) per studiare i trascritti corti e lunghi espressi nei tessuti fiorali di *O. italica.* Per caratterizzare il miRNoma fiorale di *O.italica* è stata costruita e sequenziata una libreria di small RNA di infiorescenza. Tra le 37.818 *reads* uniche, il sottogruppo di 24 nt è risultato il più abbondante. L'analisi dei micro RNA (miRNA) ha consentito l'identificazione in *O. italica* di 23 famiglie conservate di miRNA e di nuovi possibili miRNA specifici di orchidea. Il profilo di espressione dei miRNA nei tessuti fiorali di *O. italica* e la predizione dei loro possibili bersagli ne indicano una funzione conservata rispetto a quella delle specie modello. Per analizzare il trascrittoma di *O.italica*, il cui genoma non è ancora disponibile, è stato utilizzato un approccio *de novo.* Partendo da più di 100 milioni di *reads*, sono stati assemblati 132.565 trascritti, raggruppati in 86.079 *unigenes*. L'annotazione funzionale ha assegnato il 45,3% degli *unigenes* alle sequenze presenti nella banca dati NCBI, il 37,4% alla GO terms, il 18,3% alla KOG e l'8,3% alla KEGG. L'analisi di espressione *in silico* è stata validata con la Real-Time PCR su dieci *unigenes* selezionati, indicando una correlazione positiva statisticamente significativa. Oltre ai trascritti codificanti, sono stati analizzati anche quelli non codificanti (lncRNA), la cui analisi del profilo di espressione ne ha lasciato ipotizzare un ruolo funzionale nei tessuti fiorali di *O. italica*. L'analisi del trascrittoma di *O. italica* ha consentito di identificare 12 trascritti appartenenti alla famiglia di fattori di trascrizione TCP, probabilmente coinvolti nella determinazione della simmetria fiorale (nelle orchidee bilaterale). L'analisi filogenetica ha rivelato che essi appartengono a differenti classi (I e II) e gruppi (PCF, CIN and CYC/TB1) e che mostrano numerosi motivi conservati comparti con quelli di *Arabidopsis* e *Oryza.* Inoltre è stata dimostrata la presenza in un trascritto TCP di *O.italica* di uno specifico sito di taglio per miR319 ed è stata condotta l'analisi di espressione di tutti di trascritti TCP identificati in differenti e tessuti fiorali e in foglia in due stadi di sviluppo. I risultati suggeriscono che alcuni trascritti TCP svolgono

funzioni ridondanti nei vari tessuti e stadi di sviluppo esaminati, mentre altri trascritti esercitano una funzione specifica e solo in determinati tessuti.

1. Introduction

1.1 The Orchidaceae family

> "In my examination of orchids, hardly any fact has so much struck me as the endless diversity of structure, the prodigality of resources, for gaining the same end, namely, fertilisation of one flower by the pollen of another. The fact to a certain extent is intelligible on the principle of natural selection".

Charles Darwin, *On the various contrivances by which British and foreign orchids are fertilised by insects*, 1862

Since ancient times, researchers have studied the origin and the evolution of the Orchidaceae. Even Darwin, after the Origin of Species, published a volume on orchids to present their adaptations as the result of innumerable natural forces and to expand the theory of adaptation through natural selection.

The Orchidaceae family is one of the largest among the flowering plants and includes species that have invaded every habitat with highly specialized reproductive strategies and extremely diversified flowers. Despite their morphological diversity, orchids share bilaterally symmetrical (zygomorphic) flowers with three outer tepals (sepals), two lateral inner tepals (petals) and a highly modified median inner tepal (lip or labellum) (fig. 1) [1].

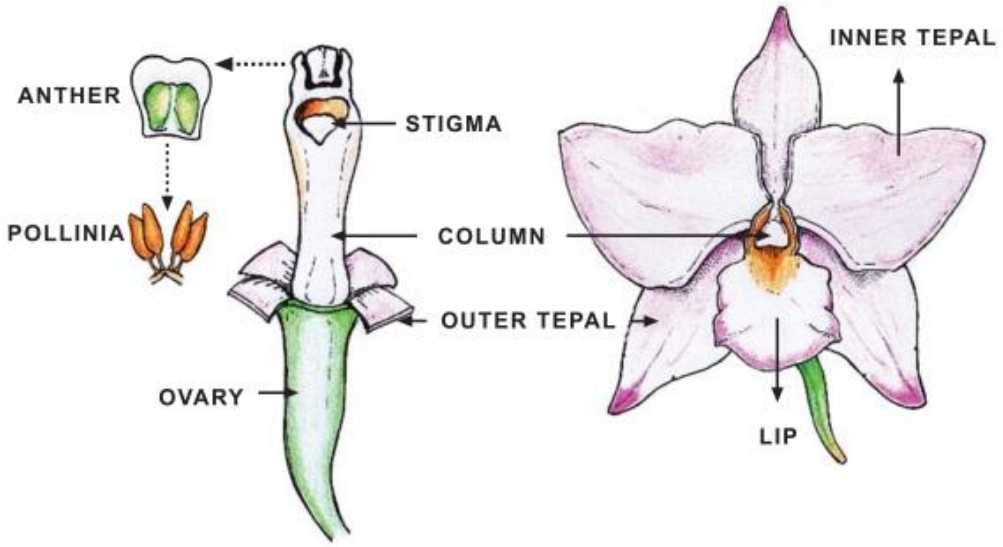

Figure 1 Reprinted from Aceto *et al.* (2012) [1] by permission of Eureka Science Ltd. Model of an orchid flower.

The lip exhibits a distinctive shape and color pattern different from that of the other tepals. In some orchids, the upper surface of the lip may be adorned with a callus, spurs or glands. The male (stamen/anther) and female (pistil/stigma) tissues are fused together and form the column, the orchid's reproductive structure. At the top of the column are the pollinia and at the base of the column is the ovary, which develops when triggered by pollination [2-4]. An interesting feature of the development of most orchid flowers is the phenomenon of resupination during which the lip undergoes a 180° rotation in floral orientation that moves the lip to the lowest tepal position and opposite to the fertile anther. Its collocation suggests that its highly diversified shape and pigmentation are the result of adaptations to specific pollinators [5]. The lip is an important adaptation of the orchid to facilitate cross pollination. It can be imagined as a coloured flag to attract potential and specific pollinators and can be supposed to act as a landing platform.

In the past years, the phylogeny of this family was based on some morphological characteristics, though many phenotypic characters often have an adaptive nature. This approach induced confusion and contradictory phylogenetic reconstructions of the Orchidaceae family [6], but the advent of molecular systematics has revolutionized our understanding of the phylogeny and evolution of this plant family. The data from molecular markers have been progressively added to the morphological ones and the current classification system designates five subfamilies within the Orchidaceae: Apostasioideae, Cypripedioideae, Epidendroideae, Orchidoideae and Vanilloideae [1]. Each subfamily includes numerous tribes and subtribes [7, 8] (fig 2).

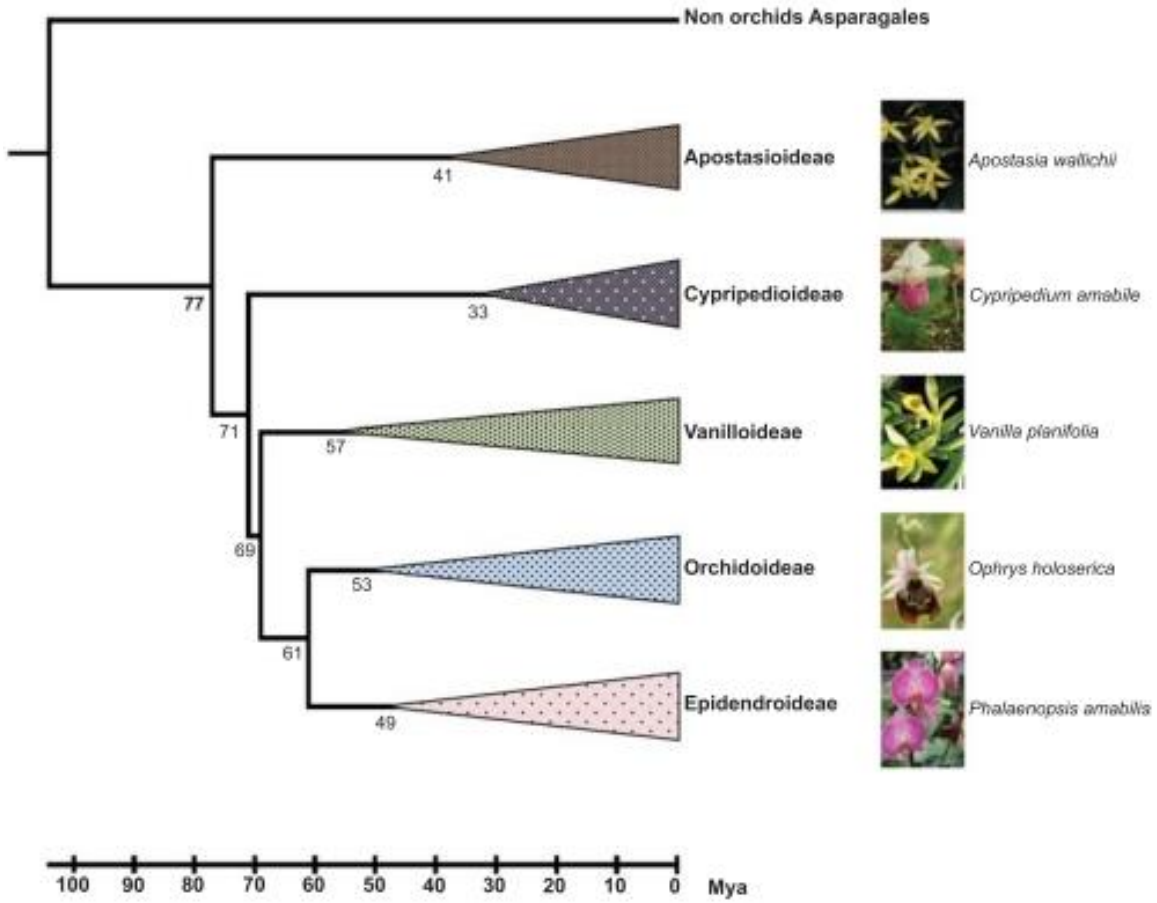

Figure 2 Reprinted from Aceto *et al.* (2012) [1] by permission of Eureka Science Ltd. Phylogeny of the five sub-families of Orchidaceae. The numbers indicate the divergence time in millions of years (Mya). The images of orchid species on the right of each subfamily indicate the most representative species.

One of the still open questions about orchids is related to the great diversification of their flowers: why are orchid flowers so different and which are the changes in the developmental processes that have caused their morphological diversity?

The selection acts upon the phenotype and the knowledge of the link between the evolution of genes and the morphological innovations becomes essential to understand the evolution of new forms and functions in orchids. However, in this family this connection remains poorly understood, as well as the causes determining the wide species diversity. For the latter question, relevant roles have been attributed to epiphytism, highly diversified pollination strategies [9], natural selection and genetic drift [10].

Evolutionary developmental biology (evo-devo) uses comparative approaches to explain how changes in developmental pathways occur during the evolution. Numerous studies indicated that gene regulatory networks controlling developmental processes are

conserved and changes in the regulation and/or in the expression pattern of these genes induce phenotypic changes. So, if the evolution of novel morphological characters is related to changes in the expression pattern of "candidate genes" the study of these genes and of their transcriptional regulation is probably the key to understand the evolution of a given lineage.

Changes in temporal and spatial expression patterns may be caused by different events. Among them, cis-regulatory changes by accumulation of mutations and sub- or neo-functionalization following gene or whole-genome duplication events are the most frequent. In the first case, the gene acquires a new transcription factor binding site (TFBS) that can be recognized by a novel transcription factor (TF) which could extend or reduce the expression domain of the gene under certain conditions. The second case is very common in plants where multiple gene and/or whole lineage-specific duplication events have occurred. It is speculated that these events generate the raw genetic material that drives the evolution of new forms and functions. Indeed, in addition to redundant functions, duplicate genes can undergo functional diversification (sub- or neo-functionalization) that provides a substrate for morphological evolution.

The evolution of the morphological complexity in plants has been linked with the expansion of genes encoding transcriptional regulators [11].

The main transcriptional regulators in plant are the TFs, which control all the major processes of life such as the flower development. They do so by a DNA-binding domain, which is almost always the most highly conserved part of the TF and is used to classify the TFs in different families. During plant evolution, whole or segmental duplication events have lead to the expansion in the number of TFs families.

In orchids, the difference between the rate of morphological and molecular evolution suggests that the cause of such a wide morphological diversification might be due to the evolution of a small number of genes. This hypothesis has led many researchers to study the genes involved in the regulation of flower development in this plant family. Among the TFs involved in the complex molecular network of the flower development, the most studied are the MADS-box genes, involved in specifying meristem and organ identity, and the TCP genes that control floral zygomorphy. Furthermore, the study of developmental

processes in plants highlights the role of the non-coding RNAs (ncRNAs) in the regulation of the TFs involved in flower development.

1.2 The MADS-box genes

The MADS-box genes family is present in almost all the major eukaryotic groups, larger in higher plants than in animals or fungi [12, 13]. The MADS-box genes encode for transcription factors showing a typical DNA binding domain called MADS-box domain. These TFs represent one of the most studied gene families in plants for their essential roles in almost every developmental process.

The acronym MADS was established after the discovery of the first MADS-box genes *AGAMOUS* (*AG*) from *Arabidopsis thaliana* [14] and *DEFICIENS* (*DEF*) from *Antirrhinum majus* [15]. Their products are proteins showing a ~60 amino acid DNA-binding domain with similarities to the serum response factor (SRF) in *Homo sapiens* [16] and Minichromosome maintenance 1 (Mcm1) in *Saccharomyces cerevisiae* [17]. This conserved domain is present in all the MADS-box transcription factors [18].

The origin of this transcription factor family is associated with a series of gene duplications followed by gene loss, neo- or sub-functionalization [19]. Indeed, it was discovered that the MADS-box family has evolved from a region of the topoisomerase II subunit A [20] and that a second gene duplication occurred before the divergence of plants and animals, giving rise to two main groups of MADS-box genes: type I and type II [21]. These two classes of MADS-box genes have distinct functions and evolutionary histories. The type I originated mainly by recent duplication of single genes. In addition, they are so heterogeneous that can be further classified into three subclasses: M-alpha, M-beta and M-gamma. All of type I MADS-box genes share a 180 bp DNA sequence encoding the MADS domain [22] and are involved predominantly in development of seed, embryo and female gametophyte [23]. The Type II MADS-box genes are mainly the product of whole genome duplications and have a modular domain structure called MIKC. It contains the highly conserved DNA-binding MADS domain (M) at the amino terminus, followed by the a poorly conserved I (intervening) domain and a moderately conserved K (keratin-like) domain, both essential for protein–protein interactions and the formation of high-order protein complexes. Finally, a variable

carboxyl-terminal (C) region has roles in the formation of protein complexes and may function as a trans-activation domain [24, 25].

Based on differences inside the domain structure, MIKC-type MADS-box genes can be further divided into $MIKC^C$ and MIKC* genes [26]. The latter are involved in the development of the male gametophyte [27], whereas the $MIKC^C$ genes, the most studied group of MADS-box genes, are involved in many functions related to plant growth and development, including the flower formation, and can be divided into several distinct subfamilies.

An interesting aspect of the MADS-box genes evolution is that the number and the functional diversity of this family of TFs increased considerably during land plant evolution. Moreover, the functions and expression patterns of the $MIKC^C$-type genes suggest their involvement in the origin and evolution of seed plant reproductive structures [28,29].

The identy of floral organs depends on the expression and interaction of floral homeotic genes. The spatial and functional activity of these genes is described by the ABCDE model of flower development [30, 31]. This model classifies the homeotic genes into five classes (from A to E) based on the mutant analyses of the model species *Arabidopsis thaliana.* All but one (*AP2*) the floral homeotic genes encode MADS-box TFs.

The flower of *Arabidopsis thaliana* is structured into four concentric whorls of floral organs and, according to the ABCDE model, the A-class genes *APETALA1* (*AP1*) and *AP2* alone specify sepal identity in whorl 1. The A-class and B-class genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) together determine petal identity in whorl 2. The B- and C-class (*AGAMOUS, AG*) genes together specify stamen identity in whorl 3. The C-class genes alone in whorl 4 determine the formation of carpel. Finally, the class D genes *SEEDSTICK* (*STK)* and *SHATTERPROOF 1 and 2* (*SHP1, 2*) specify ovule identity within the carpel and the class E genes *SEPALLATA* (*SEP1-4)*, expressed in all the whorls, act in a redundant manner for the correct formation of all of the floral organs (fig. 3A).

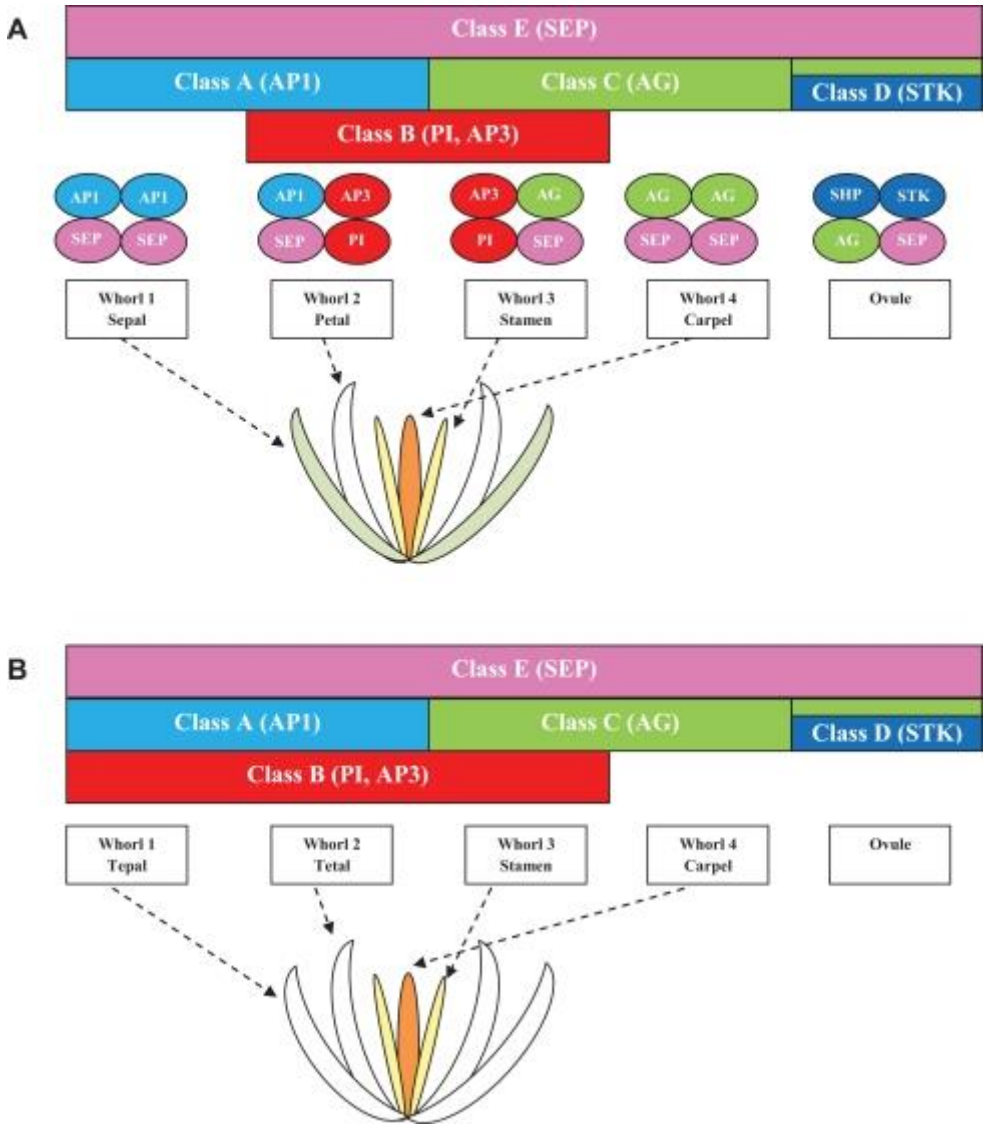

Figure 3 Reprinted from Aceto *et al.* (2012) [1] by permission of Eureka Science Ltd. (A) Schematic representation of the ABCDE and the quartet models of floral development. (B) The expanded ABCDE model.

The MADS-box TFs act forming homo- and heterodimers to recognize the conserved nucleotide CC(A/T)$_6$GG DNA sequences (known as the CArG boxes) [32]. According to the "floral quartet model", after the formation of dimers, the MADS-box proteins interact to form tetrameric protein complexes consisting of two dimers that bind the target DNA sequence containing two CArG boxes and thereby the complexes activate floral organ-specific expression programs [33, 34]. The interaction with other homeotic proteins takes place through the K domain that in some cases contributes to heterodimerization [35]. An important role as mediators in the higher-order complex formation is played by the members of the SEP subfamily [30,36].

The ABCDE model is generally conserved among plants [37-41]; however, changes in homeotic gene expression in non-model species have been observed. In particular, recent studies conducted on monocots, including *Orchis italica,* indicated that there is an expansion of the expression profile of the class B MADS-box genes into the first floral whorl, while the expression profile of the other MADS-box genes is generally in agreement to the canonical model [42-45] (fig 3B).

Among orchids, *Orchis italica* is one of the most widespread Mediterranean species, belonging to the Orchidoideae subfamily. Commonly known as the "naked man" for the shape of the lip that seems to imitate the body of a man, the inflorescence of *O. italica* is dense with light pink flowers. The sepals are rosy with evident purple streaks, the slightly darker petals. The lip is three-lobed, white-pinkish, speckled with purple (fig. 4).

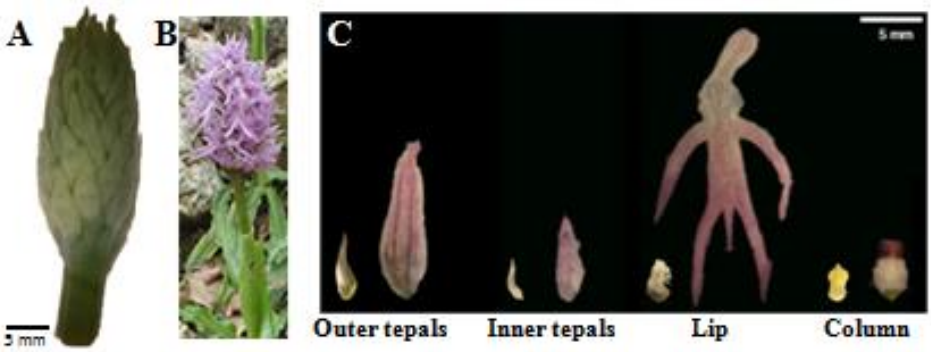

Figure 4 Reprinted from De Paolo *et al.* [46]. (A) An inflorescence of *Orchis italica* before anthesis (early stage) and (B) after anthesis (late stage). (C) Single tissues (outer tepals, inner tepals, lip and column) collected from a floret of *O.italica* at the two developmental stages: early (left) and late (right).

The orchid flower does not show difference in the morphology of the sepals and petals, rather it exhibits similar organs in the outer whorls 1 and 2 called tepals. This phenotypic character seems to be related to the expression of the class B genes into the whorl 1 as well as in the whorl 2 and 3 [47, 48]. However, although the expansion of the ABCDE model provides a good explanation of the development of the tepals, it does not explain the the wide morphological diversification of the lip.

Recent research focused on of the class B MADS-box gene family have improved the understanding of the mechanisms involved in the perianth differentiation in orchids. The evolutionary analysis of the class B MADS-box genes indicated the presence of two major

lineages, the *AP3/DEF*-like and the *PI/GLO*-like genes, arising from a duplication event that took place before the origin of angiosperms [49, 50]. Subsequently, a second duplication event occurred in the paleo-*AP3/DEF* lineage producing two other distinct clades: TM6 and euAP3 [49]. More recently, the studies conducted on the *AP3/DEF*-like genes in orchids indicated the presence of four clades: *AP3/DEF*-like clade 1 (*PeMADS2*-like), clade 2 (*OMADS3*-like), clade 3 (*PeMADS3*-like) and clade 4 (*PeMADS4*-like) [49, 50].

The orchid code is a developmental-genetic code according to which, in the orchids, the key to understand the evolution and morphological diversification of the perianth has to be found in the duplication events followed by sub- and neo- functionalization in the regulatory region of the class B *AP3/DEF*-like genes [47, 51-53]. This theory speculates that the interaction of one PI/GLO-like protein with the different four AP3/DEF-like gene products determined the identity of the tepals and lip [51, 52]. According to the orchid code theory, the identity of the organs of the perianth depends on the different expression levels of the members of the distinct clades of B-class proteins. The interaction between the clade 1 and 2 products is involved in the development of the three outer tepals, while the formation of the two lateral inner tepals is regulated by the interaction of high expression levels of the clade 1 and 2 and low levels of the clade 3 and 4 gene products. High expression levels of the clade 3 and 4 gene products and low expression levels of the clade 1 and 2 mediate the development of the lip [5, 51].

1.3 Genes involved in the flower symmetry

Variation in floral symmetry is one of the most fascinating aspect in the study of the evolution and diversification of flowering plants. Flower development in higher plants gives rise to an enormous variation of flower morphologies and symmetries. Phylogenetic analyses have shown that during the diversification of flowering plants, numerous evolutionary transitions have occurred between radial flower symmetry (polysymmetry, actinomorphy; fig. 5a), with multiple planes of mirror image symmetry, and bilateral flower symmetry (monosymmetry, zygomorphy; fig. 5d), with just a single plane of mirror image symmetry. In addition, flowers may show disymmetry, with two planes of mirror image symmetry, or asymmetry with zero planes of mirror image symmetry (fig. 5b, c). Transitions

from radial to bilateral symmetry are probably associated to the evolution of specialized flower-pollinator interactions [54-57].

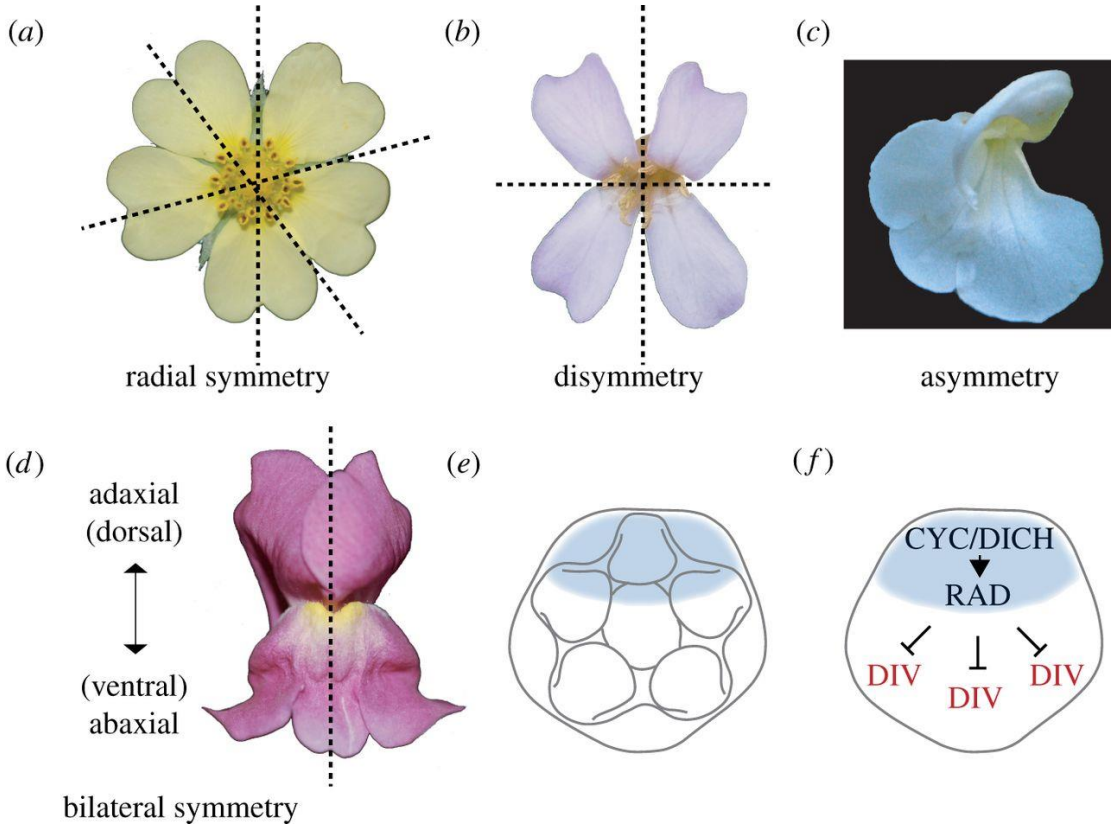

Figure 5 Reprinted from Hileman (2014) [58] by permission of the Royal Society. The different kinds of flower symmetry: radial symmetry (a, *Potentilla* sp.), disymmetry (b, *Cardaminopsis arenosa*), asymmetry (c, *Pedicularis racemosa*) and bilateral symmetry (d, *Antirrhinum majus*). One or more genetic signals that differentiate the dorsal (adaxial) from the ventral (abaxial) domains of the developing flower are shown in dorsal shading of the early developing flower (*e*). In the model species *A. majus*, the genetic program that establishes dorso-ventral flower identity from early stages of development (*f*).

The flower-pollinator interaction is particularly interesting in orchids, where some species have flowers that are pollinated by a single pollinator insect. A classic example of coevolution between flower and pollinator is the Darwin's orchids (*Angraecum sesquipedale)*. This orchid is particularly fascinating because the length of the flowers' nectar spurs is up to 35 centimetres long. Charles Darwin in his book "*On the various contrivances by which British and foreign orchids are fertilised by insects"*, speculated that this orchid must be pollinated by a gigantic moth, with an enormous proboscis capable of accessing the nectar collected in the bottom of the long spur. However, only many years later it was discovered a moth (*Xanthopan morganii praedicta*) with a proboscis so long to

get to the nectar. Given this relationship (symmetry-pollination specificity), it was hypothesized that the transition to flowers with bilateral symmetry represents the key innovation contributing to the diversification of flowering plants in species-rich flower lineages as Orchidaceae [58].

The genetic machinery for the occurrence of zygomorphic flowers (monosimmetry) was first identified in *Antirrhinum majus*, where it was found that the *cycloidea (cyc) and dichotoma (dich)* mutants returned to radially symmetrical flowers, eliminating individual organ asymmetry [59]. The *DICH* and *CYC* genes, belonging to the TCP family, are the key regulators that establish dorso-ventral (DV) symmetry and are expressed in the dorsal region of the floral meristem [60]. These two regulators are also responsible for the elaboration of organ symmetry and interact with two distinct MYB proteins DIVARICATA (DIV) and RADIALIS (RAD), respectively, to determine lateral and ventral identities [61, 62].

The TCP genes are found only in plants and encode transcription factors that share a 60-residue homologous region called TCP domain [60], common to all the members. This domain was initially identified in four proteins, from which the name 'TCP' was derived: teosinte branched1 (tb1) from maize (*Zea mays*) [63], CYCLOIDEA (CYC) from snapdragon (*A. majus*) [64], and the PROLIFERATING CELL FACTORS 1 and *2* (PCF1 and PCF2) from rice (*Oryza sativa*) [65]. The TCP domain is predicted to adopt a basic helix–loop–helix (bHLH) motif that allows DNA binding and protein–protein interactions [60].

Phylogenetic analyses based on the TCP domain have identified two subfamilies: class I, also known as PCF class or TCP-P class, and class II, also known as TCP-C class [66-68]. Class II is subdivided into two clades: the CYC/tb1 clade, or angiosperm-specific ECE clade, and the more ancient CINCINNATA (CIN) clade [69]. The members of the ECE clade have a 18-20 residues arginine-rich motif (the R domain) that might be involved in protein–protein interactions [60], and a relatively conserved glutamic acid–cysteine–glutamic acid (ECE) motif between the TCP and R domain; some members of the CIN clade independently acquired the R domain [67].

Class I and class II factors have different consensus binding sites. The consensus for class I is GGNCCCAC, while the consensus for class II binding site is distinct but overlapping with that of class I sites: G(T/C)GGNCCC [66, 70, 71].

The TCP genes encode for transcription factors that directly modulate the transcriptional status of genes involved in many processes of the plant growth and development [72]. Specific TCP proteins forming homo- and heterodimers can act as transcriptional activators or repressors and some may have both functions. The molecular mechanisms by which the TCP proteins regulate the transcription are still poorly understood, but it seems that some TCP proteins require interaction with other proteins to bind target site on DNA. Probably these proteins act as part of a multimeric complex of TFs to control the transcription of target genes [69]. These evidences indicate that there are regulatory mechanisms that act at different levels.

In general, class I TCPs have been associated with the promotion of the cell cycle machinery whereas class II TCPs have been suggested to promote the arrest of the cell cycle [66, 73]. By contrast, the class II CIN-type genes limit cell proliferation at the margins of the developing leaf primordia in *Antirrhinum* [74], *Arabidopsis thaliana* [75] and *Solanum lycopersicum* [76], while their function in monocots is unknown.

Phylogenetic analyses on ECE genes in core eudicot species revealed that this group evolved through a series of duplication events that gave rise to the three subgroups CYC1, CYC2 and CYC3 [77]. The CYC1 clade probably resulted from the first ECE split and is thus sister to CYC2 and CYC3, while CYC3 genes exhibit a duplication pattern similar to CYC2 [78].

During the evolution of monosymmetry (zygomorphic flowers), the duplications and subsequent accumulation of mutations of the *CYC2* genes, were an important source of new gene functions (sub-functionalization) that have facilitated the evolution of variable angiosperm symmetry forms [79].

The TCP genes *CYC* and *DICH*, members of CYC/TB1-like clade, are involved in the estabilishment of flower and petal symmetry through the interaction with DIVARICATA and RADIALIS TFs belonging to the MYB family [80].

The MYB family is very important in the transcriptional regulation of a large number of genes involved in many plant-specific processes. MYB proteins are characterized by the presence of a specific DNA-binding domain composed of one, two or three repeats of about

52 residues, folded to form a "helix-turn-helix" [81]. The DIV protein is formed by two MYB domains and contains both an N-terminal protein interaction and a C-terminal DNA binding domain, while RAD presents one MYB domain which constitutes two thirds of the protein. Interestingly, the RAD protein is closely related to the N-terminal MYB domain of DIV, suggesting that the RAD could derive from an ancestral DIV-like protein after the deletion of its C-terminal domain [62].
In *A. majus*, the *CYC* and *DICH* genes are expressed and act redundantly in the dorsal region of the floral meristem during the early stages of differentiation of the petals and stamens. Instead, the ventral identity of the floral meristem is specified by the MYB gene *DIV*, expressed in the ventral region of the flower [62, 82]. During early wild type flower development, *DIV* is expressed in both the dorsal and ventral domains; in the later stages, its expression is restricted to the ventral petals [61]. In the *div* mutants the ventral petal acquires lateral identity.

The effects of *CYC* and *DICH* on the dorsal domains and of *DIV* on the ventral domains of the flower are in part mediated by RAD. The RAD protein, whose expression is positively regulated by *CYC* and *DICH* and restricted to the dorsal domain, antagonizes the DIV protein with a post-translational mechanism [62] (fig. 5 e, f). However, probably there are some *CYC* or *DICH* functions that are independent from *RAD* activation because *rad* mutants in snapdragon do not show a completely radial flower [71].

The DIV and RAD proteins form heterodimers with another MYB protein, DRIF (DIV-and-RAD-interacting factor): the heterodimers DIV/DRIF bind DNA on sites with DIV consensus sequence, regulating presumably the activity of target genes necessary for the development of the ventral region of the flower. However, in the dorsal region the RAD protein rivals with DIV for the interaction with the DRIF proteins, limiting the action of DIV only in the ventral region of the floral organs where RAD is not expressed [83].

The involvement of the CYC/tb1-like genes in the evolution and maintenance of bilateral symmetry has been extensively analyzed in dicotyledonous species [64, 84, 87-98], while there are only few studies on monocot species [85, 86]. In core eudicots, the involvement of CYC-like genes in the control of floral symmetry has been shown in both asterids [64, 87-92], and rosids [93-98]. A shared hypothesis is that outside the order Lamiales, to which

the snapdragon belongs, the developmental genetic program of bilateral symmetry (and back to radial symmetry) has evolved several times by a parallel or independent recruitment of the CYC/RAD/DIV system. Across the rosids and asterids (eudicot flowering plants), the *CYC* orthologs expression in the dorsal petal tissue is a key factor in the dorsal identity [92-99].

In *Oryza sativa*, a monocot species, floral zygomorphy along the lemma-palea axis is partially or indirectly determined by the CYC-like homolog *RETARDED* PALEA1 (*REP1*), which regulates palea identity and development [100]. Researches on the possible involvement of the *CYC* genes on floral symmetry in other monocot plants as *Z. mays* [61], Zingiberales (Costaceae and Heliconiaceae) [101], *Commelina* and *Tradescantia* (Commelinaceae) [86] and Alstroemeria (Alstroemeriaceae) [102] have also been conducted. In the flowers of the monocots *Costus* and *Heliconia* (Zingerberales), as well as in *Commelina* (Commelinales), all bilaterally symmetrical, the expression of the *CYC*-like genes is not uniform along the dorso-ventral flower axis, but in contrast to the general pattern of a CYC-dependent program, the asymmetric *CYC*-like expression is limited to the ventral side of the flower [86-101]. However, it is still not clear whether this emerging pattern of dorsal flower expression in eudicots and ventral flower expression in monocots of the *CYC* homologs is general or not. Further comparative studies might give an answer to this question, to better understand how the *CYC* homologue expression is regulated during flower development in both monocot and euicot plants.

1.4 Plant non-coding RNAs (ncRNAs)

Nowadays, thanks to the development and application of high-throughput deep sequencing, we know that although ~90% of the eukaryotic genome is transcribed, only the 1%–2% of total RNAs is mRNAs. This suggests that a large number of RNA molecules are non-coding RNAs (ncRNAs). NcRNAs can be classified as “housekeeping” ncRNAs, as ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA), and small nucleolar RNA (snoRNA), and “regulatory” ncRNAs which include small ncRNA (sncRNA) and long non-coding RNA (lncRNA).

1.4.1 Small non coding RNAs

Small ncRNAs are approximately 21-24 nucleotides RNAs that act as a critical component of plant gene regulation at the transcriptional and post-transcriptional level. SncRNAs include a wide family of molecules that are different in the biogenesis, length and function [103].

Following the classification proposed by Axtell (2013) [103], small RNAs include two main categories, which are distinguished by their biogenesis (fig. 6). The first difference is the precursor of the small RNA. There are two different types of precursor molecules: those formed by a single-stranded RNA (ssRNA) that has the ability to fold and form an imperfectly double-stranded RNA called “hairpin” structure (hpRNA); those formed by a double-stranded RNA (dsRNA) with intermolecular perfect hybridization of two complementary RNA strands (siRNA) that derive from indipendent transcription of inverted repeat sequences, convergent transcription of sense-antisense gene pairs or synthesis by RNA-dependent RNA polymerases (RDRs) [104].

The hpRNAs can be divided into two subclasses: miRNAs (which include lineage-specific miRNAs and long miRNAs), and other hpRNAs. The siRNA, can be divided into three subclasses: heterochromatic siRNA, secondary siRNA (which include phased siRNAs and trans-acting-siRNAs), NAT-siRNAs (which include cis-NAT-siRNA and trans-NAT-siRNA).

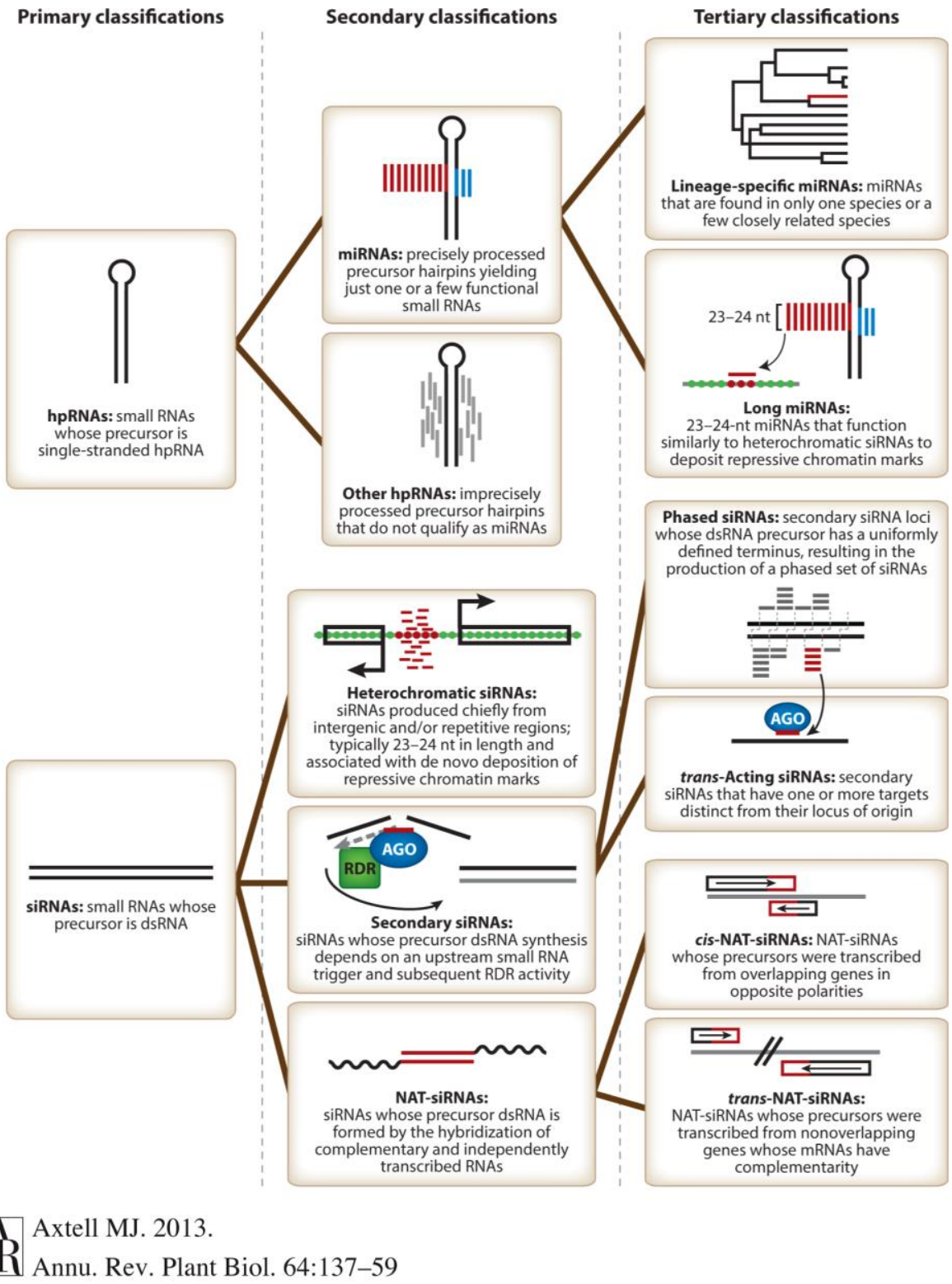

Figure 6 Reprinted from Axtell (2013) [103]. Classification of the endogenous plant small RNAs.

MiRNAs are non-coding RNAs of 21-24 nucleotides that regulate gene expression at the post-transcriptional level. They were discovered in *Caenorhabditis elegans* [105] and later were found widely distributed throughout the animal [106] and plant [107] kingdom. Plant

miRNAs are involved in the regulation of many aspects of plant biology, such as metabolism, hormonal response [108], biotic [109] and abiotic [110] stress and plant development [111-115].

MiRNAs are transcribed by RNA Pol II to generate long hpRNA precursors (pri-miRNAs) that are subsequently processed by the RNAse III-like enzyme DICER-like 1 (DCL1) together with two other enzymes (the dsRNA binding protein HYPONASTIC LEAVES1 (HYL1) [116] and the C2H2 zinc finger protein SERRATE (SE) [117, 118]) to generate a hairpin structure containing the miRNAs of 21-24 nucleotides called pre-miRNA. The pre-miRNA is processed a second time by DCL1 and the S-adenosylmethionine-dependent HUA ENHANCER 1 (HEN1), and the duplex miRNA/miRNA* is carried into the cytoplasm from the HASTY protein, that is homologous to exportin 5 protein [119]. In the cytoplasm, the mature miRNA is incorporated in the 'Induced Silencing Complex' (RISC RNA), in which a key component is the protein AGO that contains domains similar to those of the endonuclease RNase H. Once incorporated into the RISC complex, the miRNA recognizes sequences complementary or partially complementary called miRNA-binding sites that are located on the mRNA target; generally, a miRNA recognizes multiple targets, up to several dozens [120, 121]. The binding of the miRNA to its binding site involves the degradation or the translational repression of the target. Between these two mechanisms, the best characterized in plant is the first that is realized thanks to the action of the endonucleolytic AGO1 protein that catalyzes the cut of the target mRNA inside the duplex miRNA-mRNA. Moreover, plant miRNAs are almost perfectly complementary to their mRNA targets, unlike the animal miRNAs and their targets. The degree of complementarity between miRNAs and their targets is responsable, at least in part, of the regulatory mechanism (cleavage in plant versus translation repression in animals).

Until a few years ago, it was believed that the most widely used mechanism of post-transcriptional regulation of the plant miRNAs was the degradation of the target mRNA and the translational repression was uncommon. However, this idea is changing as it seems that many of the plant miRNAs act in both ways [122, 123]. In a developmental context, these two mechanisms may offer different potential benefits to the organism. The cut of the target mRNA directed by miRNAs may represent an irreversible way to remove a transcript accumulated, while the translational repression is reversible and could be used to adjust

the levels of transcripts expressed in the cell. When the miRNA and its target have an expression profile that does not overlap, the miRNA acts to define the spatial limits of the target mRNA; instead, when the miRNA and its target have a coincident expression profile, the function of miRNA could be to regulate the expression levels of the transcript or the protein accumulation.

MiRNAs that have a defined set of target mRNAs are grouped into families that are often conserved among plants [124]. The conserved miRNAs have homolog targets in many species, probably because they are involved in fundamental processes of development and therefore the miRNA/target relationship also have been conserved during the evolution of plants. Nevertheless, the conserved miRNAs can have new functional interactions with different targets, in addition to the canonical ones [125]. Not all plant miRNAs are conserved. There are some miRNAs found only in a few species of plants and therefore called lineage-specific miRNAs that are distinguished in many features from the conserved miRNAs [108, 126-128]. The long miRNAs are 24 nucleotides long and have a function similar to heterochromatic siRNAs that direct repressive chromatin modifications [129-132].

Interesting, miRNAs are involved both in the regulation of TFs described by the ABCDE model and in modulating the expression of the TCP genes. Several studies conducted on orchid family [133-146], including *O. italica* [147], have demostred the conserved role of miR172 in the regulation of the *AP2* gene, the only non MADS box belonging to the A class in the ABCDE model. In addition, five class II TCP genes (*TCP2*, *TCP3*, *TCP4*, *TCP10*, and *TCP24*) and their homologs in different species are regulated by the conserved microRNA miR319 [148]. High levels of miR319 and/or inactivation of miR319-regulated TCPs cause important changes in *Arabidopsis* leaf morphogenesis and the generation of crinkled leaves [149]. In addition, miR319-regulated TCPs have been described to control also the size and shape of the floral organs [150].

The secondary siRNAs include the phased siRNAs and the trans-acting-siRNAs, although many of the trans-acting siRNAs are also phased-siRNAs. Phased siRNAs result from precursors on which occur coordinated subsequent cutting events, operated by a DICER-LIKE (DCL) enzyme. The end and the specific site from which DCL starts generating

secondary siRNA are defined by a cut on the precursor transcribed from a small RNA. Some of these secondary siRNAs are able to act in trans to regulate the repression of a target mRNA and for this reason they are called trans-acting siRNAs (ta-siRNAs).

The ta-siRNAs were discovered by the analysis of the mutants in *A. thaliana*. It has been demonstrated that the mutant phenotypes that showed abnormalities in the transition to adulthood phase derived by an excess of the expression of genes encoding transcription factors of response to auxin, regulated by ta-siRNAs derived from the *TAS3* locus [151-155]. The production of the ta-siRNAs is triggered by the cut on the *TAS* primary transcript (pri-TAS) by a specific miRNA. The pri-TASs are transcribed by a DNA-dependent RNA polymerase II (RNA pol II) from *TAS* loci and they have a cap at the 5 ' and a poly A tail to the 3'. In *Arabidopis,* 8 *TAS* loci have been identified (*TAS 1 a-c*, *TAS2*, *TAS3 a-c*, *Tas4*) and each pri-TAS is around 1 Kb [151, 156-158].

Based on their biogenesis, there are two different types of pri-TAS: "one-hit", that contains a single binding site of miRNAs, and "two-hits", that contains two binding sites to the miRNA [159]. The pri-TAS1 a-c and pri-TAS2 have a binding site for miR173, while pri-TAS4 has a binding site for miR828; all of them are one-hit. Otherwise, the pri-TAS3 a-c is two-hits, containing two binding sites for miR390 and both the complementary sites of miR390 are highly conserved in higher plants and are essential for the production of the ta-siRNA [159]. Furthermore, Montgomery et al. (2008) [160] have shown that only the binding of miR390 at the 3' end of the pri-TAS3 induces the cut. A further difference regards the ARGONAUTE (AGO) protein associated with miRNAs: miR390 was only found associated with AGO7 [160], while both miRNA173 and miRNA828 are incorporated into a RISC complex with the AGO1 protein [161-163].

Unlike other siRNAs, which are dependent on RDR for the synthesis of the precursor, the dsRNA precursor of a NAT-siRNA is formed by hybridization between two complementary RNAs transcribed from opposite strands of the same locus (cis-NAT-siRNAs) or from non overlapping genes (trans-NAT-siRNA). Cis-NAT-siRNAs were identified only in plants such as in *Arabidopsis* [164, 165].

1.4.2 Long non-coding RNAs

Initially considered “transcriptional noise”, lncRNAs are RNA molecules longer than 200 nucleotides, belonging to a group of ncRNAs that are always capped and polyadenylated [166]. Despite lncRNA investigations have begun only recently, many studies have demonstrated that lncRNAs can interact with DNA, RNA and transcription factors to regulate target gene expression through DNA methylation, histone modification and chromatin remodelling [167]. All lncRNAs place their functions through four main ways: as signals, decoys, guides, or scaffolds (fig. 7) [168].

Although a few lncRNAs have been characterized in plants [169], compared with lncRNA discovered in animals, they are involved in many development processes, as plant reproductive development [170-172]. Recent studies showed that two lncRNAs, *COOLAIR* (Cold Induced Long Antisense Intergenic noncoding RNA) and *COLDAIR* (Cold Assisted Intronic noncoding RNA), could regulate *A. thaliana* flowering time through *Flowering locus C* (*FLC*) repression [173]. FLC is known as a regulator of flowering transition in plants, that is a central event for plant reproductive development. *COOLAIR* seems to promote *FLC* transcriptional repression without epigenetic silencing [174], otherwise the studies of knockdown of *COLDAIR* by RNA interference (RNAi) indicated an alteration of the vernalization response and its role in *FLC* epigenetic silencing [175].

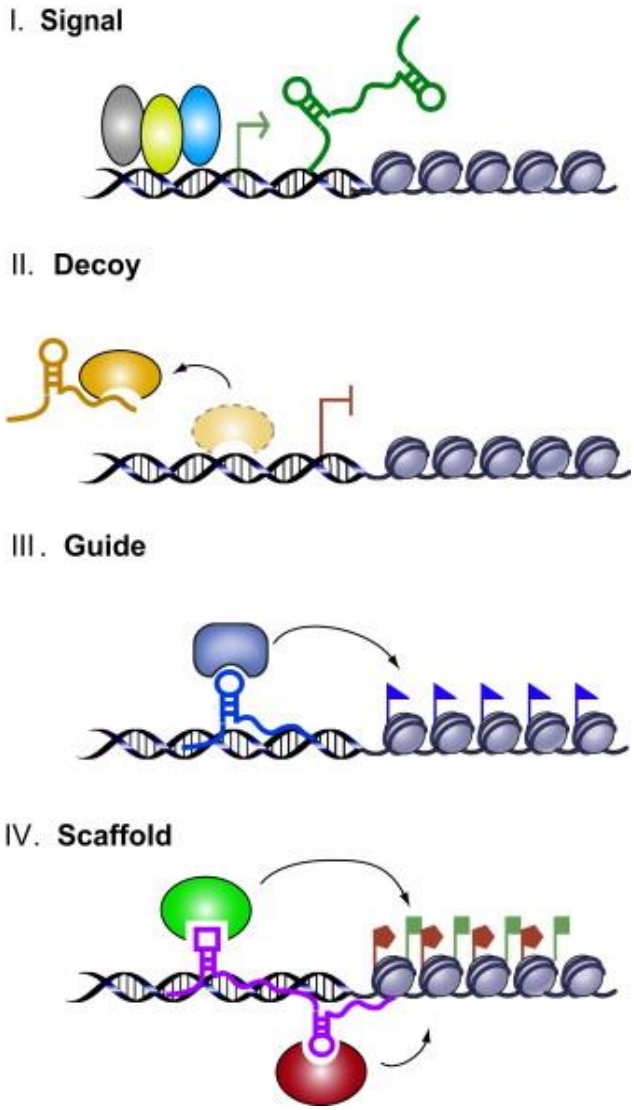

Figure 7 Reprinted from Wang *et al*. 2011 [168] (Copyright 2016) by the permission of Elsevier. Four ways to act of lncRNAs. Example I: lncRNA can regulate gene expression acting as signal and reflecting the combinatorial actions of transcription factors (colored ovals). Example II: lncRNAs can act as decoy for transcription factors and other proteins bringing them on the chromatin. Example III: lncRNAs can act as guide recruiting chromatin-modifying enzymes to target genes. Example IV: lncRNAs can act as scaffolds to form ribonucleoprotein complexes (lncRNA-RNP) that may regulate the accessibility of the chromatin affecting histone modifications.

2. The PhD research project

My PhD project concerns the study of the evolution and diversification of the orchid flower. During my PhD activity I focused on the study of the coding and non coding genes involved in the flower development of the orchid species *Orchis italica*. In particular, the aim of my PhD project was:

- To characterize the population of small RNAs expressed in the inflorescence of *O. italica* through a next generation sequencing (NGS) approach;
- To sequence, assemble and annotate the inflorescence transcriptome of *O.italica*;
- To analyze both the coding and long non-coding transcripts expressed in the inflorescence of *O. italica* that are potentially involved in flower development;
- To analyze the *TCP* genes and other gene families of interest involved in the development of the flower of *O. italica.*

3. Results and Discussion

3.1 Analysis of the inflorescence miRNome of *O. italica*

3.1.1 Conserved and novel miRNAs

The Illumina sequencing of the small RNA library of inflorescence tissue of *O. italica* produced 4,718,127 total and 2,100,557 distinct reads. After adaptor trimming, length filtering (18–35 nucleotides), removal of sequences representing less than 5 reads and tRNA and rRNA contaminant sequences, we obtained 1,064,237 total and 37,818 distinct reads (Table 1).

	Total	Distinct
Raw reads	4,718,127	2,100,557
Remaining after adaptor removal	4,412,197	1,953,969
Remaining after length range filtering (18-35)	3,019,126	1,672,887
Remaining after low-complexity filtering	3,018,960	1,672,730
Remaining after minimum abundance filtering (5)	1,161,336	40,054
Remaining after invalid sequence filtering	1,160,925	40,027
Remaining after tRNA/rRNA filtering	1,064,237	37,818

Table 1 Summary statistics of small RNA sequencing in inflorescence of *O. italica.*

The analysis of the length distribution of the total and distinct short reads in the inflorescence of *O. italica* indicated that the number of reads was highest between 21–24 nucleotides, with a peak at 24 nt (fig. 8), in accordance with the studies of the distribution pattern of small RNAs in other plant species [176-181, 144].

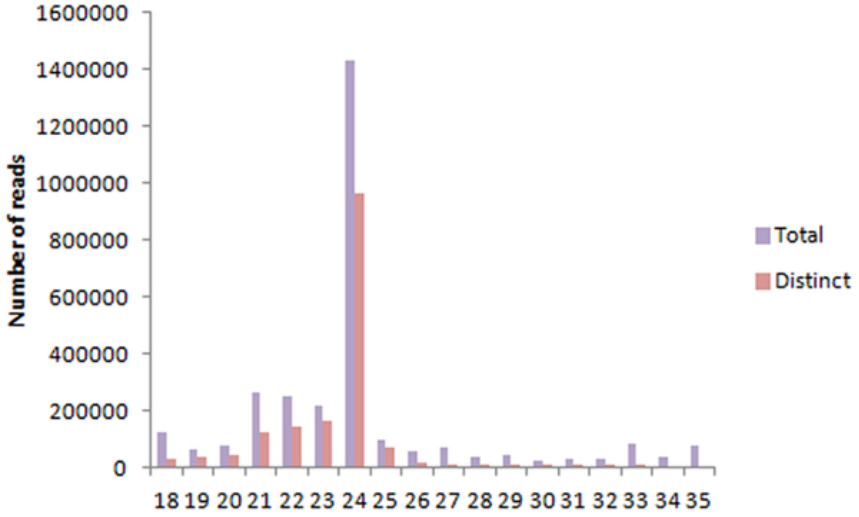

Figure 8 Reprinted from Aceto *et al.* (2014) [176] Length distribution of the short reads in the inflorescence of *O. italica*. In violet the total and in red the distinct short reads.

The analysis of the conserved miRNAs revealed a total of 175 putative miRNAs that correspond to 23 known plant miRNAs (fig. 9A). Counting the number of members from each family, the most abundant families (miR6300, miR166, miR319, miR396, miR393 and miR168) are also the most heterogeneous (fig. 9B).

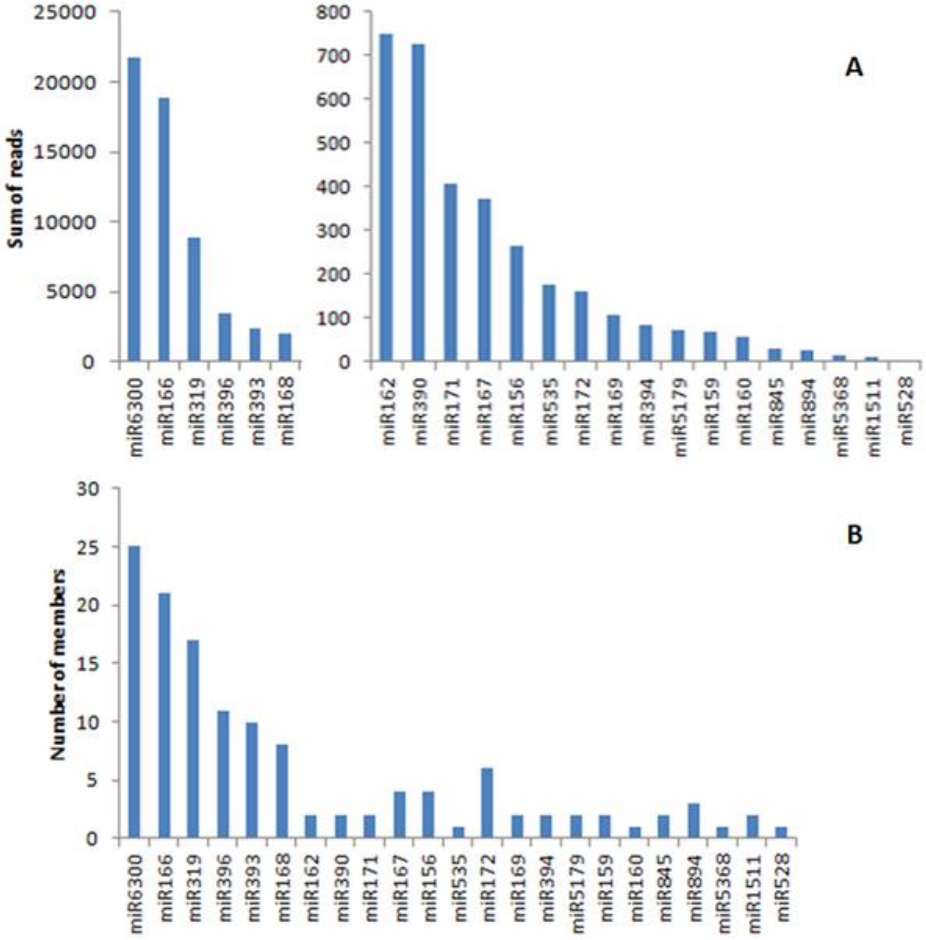

Figure 9 Reprinted from Aceto *et al.* (2014) [176]. *In silico* analysis of the expression level of conserved miRNAs in the inflorescence of *O. italica.* A) Sum of reads; B) number of members.

The comparison between the evolutionary conserved miRNAs described in Cuperus *et al.* (2011) [124] and the known miRNAs identified indicated that in the inflorescence of *O. italica* are present 16 of the 37 conserved miRNA families (fig. 10).

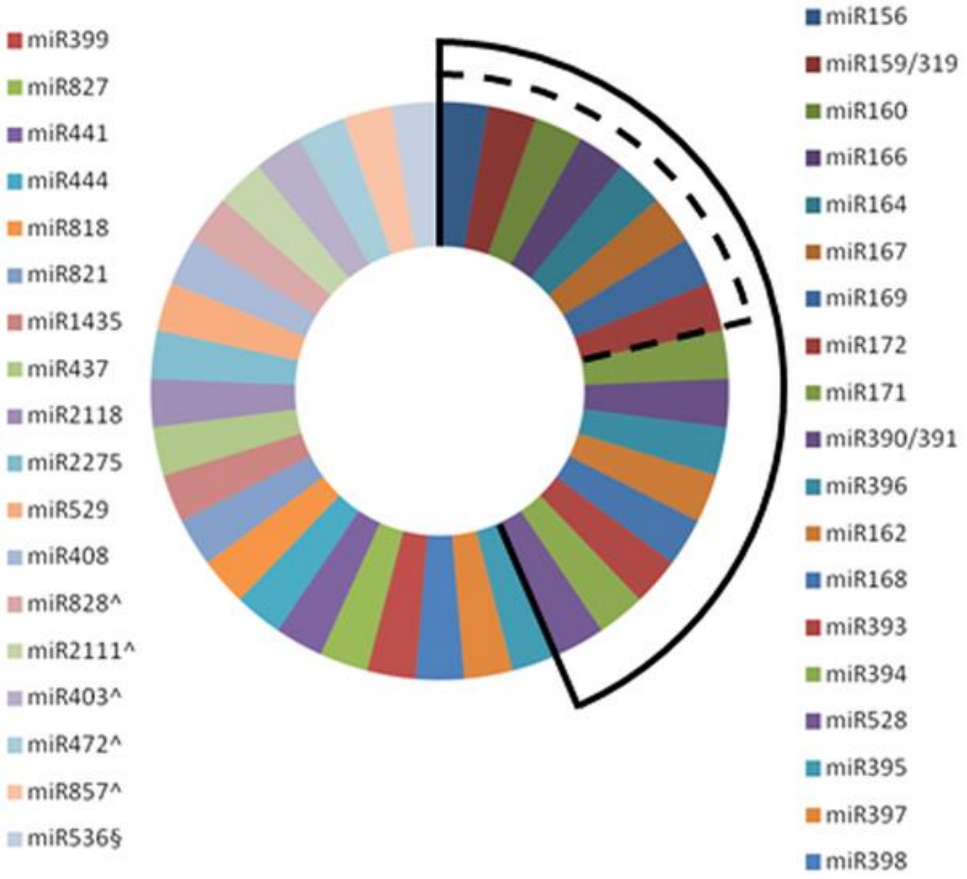

Figure 10 Reprinted from Aceto *et al.* (2014) [176]. Evolutionary conserved miRNAs identified in the inflorescence of *O. italica*. The evolutionarily conserved miRNAs family from Cuperus *et al*. (2011) [124] are indicate with different colors along the circle. Dotted lines group indicates the evolutionary conserved miRNAs involved in flower development [182]. Continuous lines group indicates the evolutionary conserved miRNAs detected in the inflorescence of *O. italica.* ^ indicates miRNA specific of dicots; § indicates miRNA specific of Bryophyta and Lycopodiophyta.

In addition, in the inflorescence of *O.italica* we have identified all of the 8 conserved miRNA families involved in flower development described in Luo *et al*. 2013 [182]. The presence of a such high degree of conservation suggests that these miRNAs play a key role in the flower development processes.

The analysis of the putative novel miRNAs was performed using the miRDeep-P software. The results revealed 478 distinct sequences that reflect the criteria of plant miRNA structures. To exclude the conserved miRNAs, these 487 distinct sequences were used as queries in a BLAST search against the miRBase database. The results showed that 109 sequences correspond to known plant miRNA families and 100 to miRNAs with limited similarity (11–15 nucleotides) with known miRNAs present in miRBase. The remaining 296 sequences were clusterized using the CD-HIT suite (see methods), resulting in 161 clusters

with a sequence identity cut-off of 80%. These distinct reads should be considered putative novel orchid-specific miRNAs.

3.1.2 *In silico* analysis of putative targets and cleavage validation

Using the psRNATarget online tool, we conducted an *in silico* analysis to predict the putative targets of all the miRNAs identified in the inflorescence of *O.italica.* When we used the reads against the transcriptome of the orchid *Phalaenopsis aphrodite* (maximum expectation 3.0), a total of 5349 putative targets corresponding to 1456 distinct Gene Ontology (GO) terms were detected. This result suggests that the role of miRNAs is conserved among orchids. When the putative target analysis was conducted against the coding transcripts of *O. italica* (maximum expectation 0.0), we used transcripts involved in flower development of *O. italica* that have been previously isolated (*OrcLFY* GenBank accession number AB088851, *OitaAP2* KF152921, *OrcPI* AB094985, *OrcPI2* AB537504, *OitaAG* JX205496, *OitaSTK* JX205497).

To increase the number of class B MADS-box genes of *O. italica* belonging to the DEF-like lineage and to verify if they are targets of miRNAs, we isolated four different *DEF*-like cDNA sequences using a MADS-box and a poly-T primer (see Methods, Table 6). The ORFs of these cDNAs (called *Oita*DEF1, *Oita*DEF2, *Oita*DEF3, *Oita*DEF4) is 681 bp, 603 bp, 672 bp and 675 bp, respectively. All of them have the typical highly conserved MADS-box domain, the moderately conserved intervening I-domain, the keratin-like K-domain and the variable C-terminal domain [25] and are deposited in GenBank with the accessions AB857726, AB857727, AB857728, AB857729.

The results of the psRNATarget analysis against the coding transcripts of *O. italica* revealed that two *O. italica* miRNAs, INF_28502 and INF_26041, that are both homologs of miR5179 of *Oryza sativa* and *Brachypodium distachyon*, target all the four *DEF*-like transcripts with better expectation value for *OitaDEF2* and *OitaDEF4* transcripts. Although in the orchids *Ericina pusilla* and *P. aphrodite* it was identified a miRNA that putatively cleaves the transcripts of class B MADS-box genes, more detailed studies are lacking. In order to validate the *in silico* analysis, we performed a modified 5'RACE experiment for all the four *DEF*-like transcripts. We obtained positive results only for one transcript, with a fragment of the expected size (129 bp) derived from the cleavage product of miR5179 on the

OitaDEF2 transcript of *O. italica* (fig. 11). We did not detect any cleavage product of the homolog of miR5179 in *O. italica* for the other *OitaDEF*-like genes, including *Oita*DEF4, although the *in silico* analysis predicted the same expectation value of *OitaDEF2*. This result could be related to different transcriptional regulation mechanisms that do not involve cutting as well as to the difference in the nucleotide sequence in the regions surrounding the predicted target site on the four *DEF*-like transcripts, resulting in decreased accessibility of the RISC complex to the *OitaDEF4*, *OitaDEF1* and *OitaDEF3* transcripts.

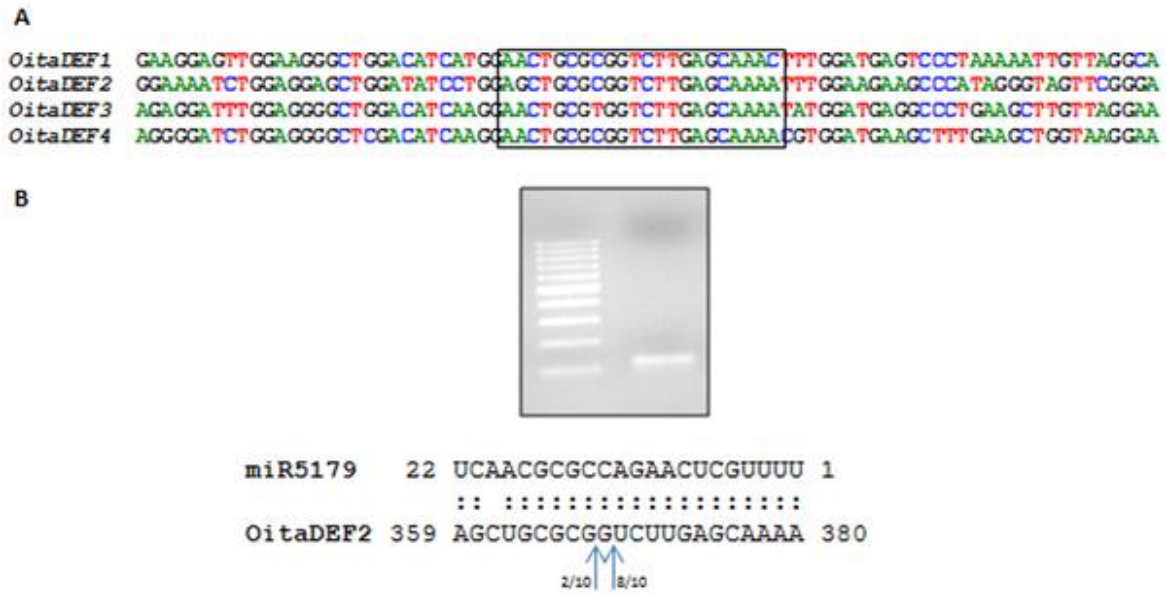

Figure 11 Reprinted from Aceto *et al.* (2014) [176] Cleavage of the homolog of miR5179 on *Oita*DEF2 in *O. italica*. A). In black box the nucleotide alignment of the putative target site of miR5179 on *Oita*DEF-like transcripts and outsite the box its surrounding region. B) Agarose gel electrophoresis of the amplified product from the modified 5' RACE experiment on *Oita*DEF2 and 100 bp ladder (Fermentas). It is also reported the alignment of the miR5179 and its target site on *Oita*DEF4 and the arrows indicate the position of the cleavage site and the number of clones corresponding to each site as deduced by the cloning and sequencing of the obtained fragment.

3.1.3 Expression analysis

The expression analysis was performed in different inflorescence tissues of *O. italica* on 10 selected miRNAs chosen based on their *in silico* expression level or their putative targets (see Methods, Table 6). Among the selected miRNAs, eight correspond to known plant miRNAs and two correspond to putative novel miRNAs. In order to amplify the miRNAs we used the Poly(T) Adaptor RT-PCR method (see methods) and the 5.8S RNA transcript as endogenous control gene.

The fig. 12A shows the expression pattern of the homolog of miR390 in *O. italica* that reveals a variable expression profile in the different tissues examined. Interestingly, this miRNA is implicated in the biogenesis of the ta-siRNAs cutting the *TAS3* mRNA [183]. The ta-siRNAs inhibit the auxine response factors mRNAs (ARFs) that control the responses to the auxin, a phytohormone with a conserved function in plant development [183].

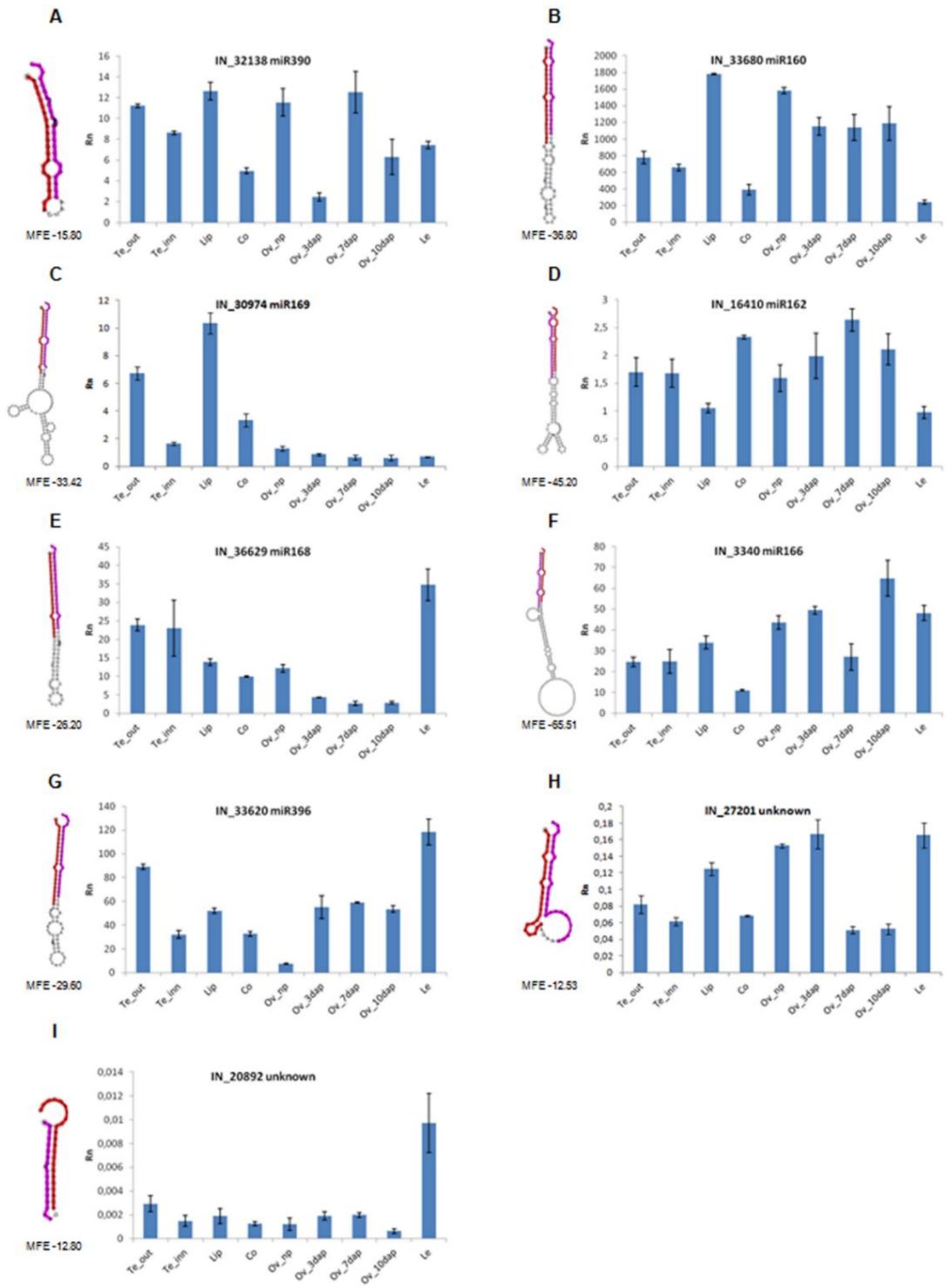

Figure 12 Reprinted from Aceto *et al.* (2014) [176]. Relative expression pattern of selected conserved and putative novel miRNAs in different tissues of *O. italica*. On the left of each figures is reported the predicted structure of the pre-miRNA and the miRNA and miRNA* sequences are shown in red and pink, respectively. MFE, minimum free energy; Rn, relative expression ratio; Te_out, outer tepal; Te_inn, inner tepal; Co, column; Ov_np, not pollinated ovary; Ov_3dap, Ov_7dap, Ov_10dap, ovary 3, 7 and 10 days after pollination, respectively; Le, leaf. Bars indicate the standard deviation.

In addition, the miR160 homolog is expressed in all the examinated tessues of *O.italica* but with a higher level in the lip and ovary (fig. 12B). This results support the role of miR160 in the shaping of the inflorescence and of the floral organs and in fertility [184, 185]. The targets of miR169 are NF-YA transcription factors. These TFs act to positively regulate the expression of the class C MADS-box gene *AGAMOUS*. miR169 regulates negatively the expression of NF-YAs and act an indirect manner to limit the expression domain of the *AG* gene mainly in outer tepals and lip, where it is expressed at levels higher than in the column, ovary or leaf. In addition, the comparison between the expression pattern of miR169 (fig. 12C) and *OitaAG* reported in Salemme et al. [147] in the same floral tissues shows a complementary profile, supporting a conserved role of this miRNA in the flower development.

The *O. italica* miR162 is expressed in all the examined tissues, with a relatively low level (fig. 12D), and targets *Dicer1-like* transcripts. Because Dicer1 protein is involved in the biogenesis of the miRNAs [186]. the low expression level of miR162 is probably related to the high processing of the microRNAs in these tissues. Also the homolog of miR168 of *O. italica* is involved in the miRNAs pathway because regulates the transcript *AGO1* that is a core component of the RISC complex [187]. However, the expression pattern among the floral organs of *O. italica* is higher than that of miR162 in all the tissues except in the ovary after pollination (fig. 12E). These results could be related to the additional role of miR168 in the stress responses and signal transduction [188].

The homolog of miR166 of *O. italica* reveals a not uniform expression profile in the examined tissues, with the lowest level detected in the column and the in ovary 7 days after pollination (fig. 12F). miR166 cleaves the transcripts of the *HD-ZIP III* gene that encodes transcription factors involved in shoot apical and lateral meristem formation, organ polarity and vascular development [189, 190]. Finally, the miRNA homolog of miR396 in *O. italica* regulates the expression of growth factors involved in flower and leaf development [125]. miR396 shows a variable expression level in the tissues examined in *O.italica* with a lower level in the ovary before pollination and higher in outer tepals and leaf (fig. 12G).

The expression profile of the two putative novel orchid-specific miRNAs IN_27201 and IN_20892 is low in the floral and leaf tissues examined and lower than that of the selected conserved miRNAs (fig. 12H and I). However, this result is not surprising because different studies showed similar expression differences between novel miRNAs compared with the known ones [178, 180]. Figure 13 A shows the expression analysis of the *DEF*-like genes of *O. italica* conducted using actin *OitaAct* gene as endogenous control and the comparison between the expression level of *O. italica* miR5179 and its putative target *OitaDEF2*.

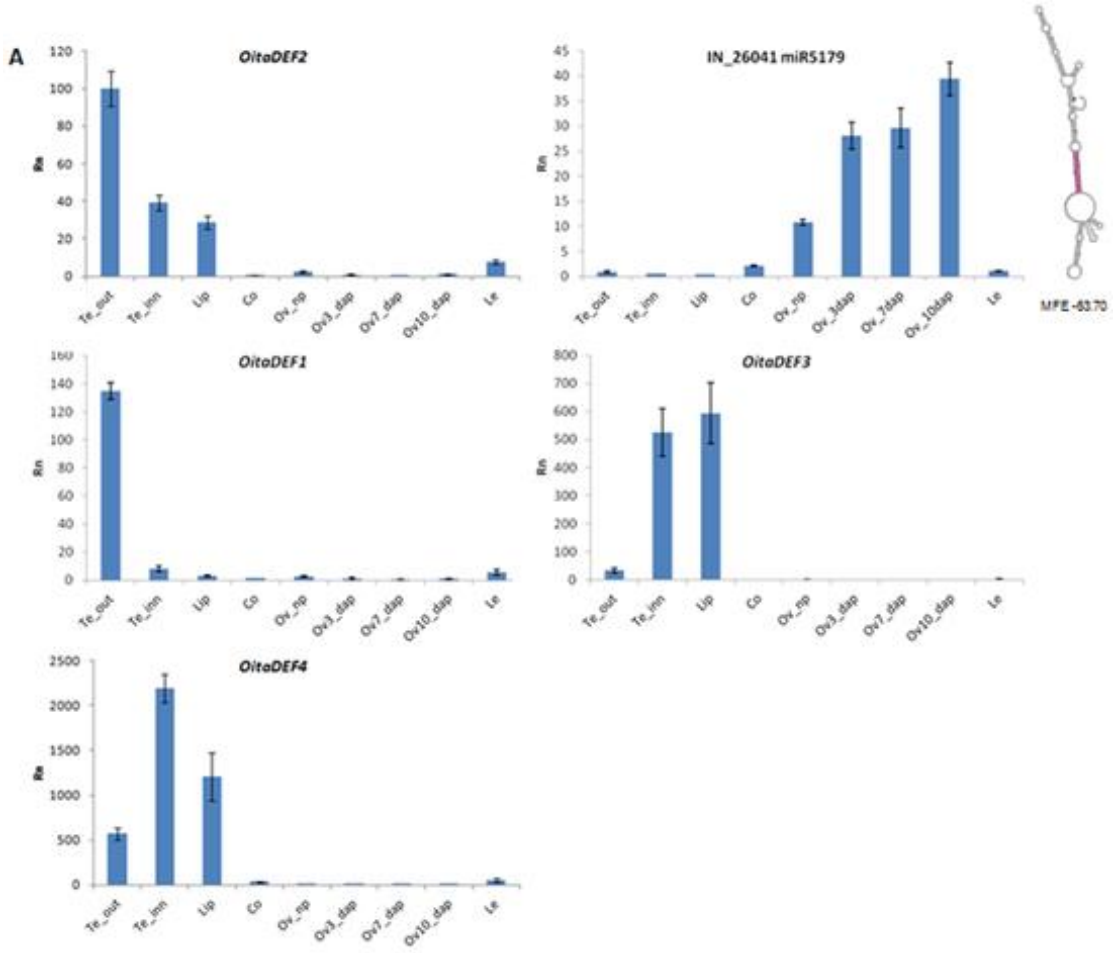

Figure 13 Reprinted From Aceto *et al.* (2014) [176]. Relative expression pattern of the *Oita*DEF-like genes and the homolog of miR5179 of *O. italica*. A) Expression profile of the *Oita*DEF1-4 genes and of miR5179 in different tissues of *O. italica*. On the rigth of the expression graph of the miR5179 is reported the predicted structure of the pre-miRNA of *O. italica*, where the miRNA and miRNA* sequences are shown in red and pink, respectively. MFE, minimum free energy; Rn, relative expression ratio; Te_out, outer tepal; Te_inn, inner tepal; Co, column; Ov_np, not pollinated ovary; Ov_3dap, Ov_7dap, Ov_10dap, ovary 3, 7 and 10 days after pollination, respectively; Le, leaf. Bars indicate the standard deviation.

The expression profile of miR5179 gradually increases from the column to the ovary 10 days after pollination, while in the tepals, labellum and leaf the expression is almost absent. The expression level of *OitaDEF2* is higher in the outer floral organs that form the perianth than in the other tissues. This complementary expression profile further supports the repressive role of miR5179 on *Oita*DEF2.

OitaDEF1 is expressed almost exclusively in outer tepals, while *OitaDEF3* and *OitaDEF4* are expressed mainly in inner tepals and lip than in the other tissues (fig. 13). The "orchid code" theory speculates that the identity of the perianth organs (outer and inner tepals and lip) depends on the fine regulation of the relative expression levels of the four *DEF*-like gene products. According to the this theory, high expression levels of the clade 1 and the clade 2 gene products are mainly involved in the development of the tepals, while high expression levels of the clade 3 and 4 gene products mediate the development of the lip [5, 51,52]. The expression pattern of the *DEF*-like genes detected in *O. italica* not only is in agreement with the ''orchid code'' theory but it allows us to hypothesize that also a miRNA is involved in the regulation of the orchid perianth development by targeting the clade 2 *DEF*-like gene.

The results obtained have been published in Aceto *et al*. 2014 [176].

3.2 *De novo* transcriptome assembly from inflorescence of *Orchis italica*

3.2.1 Illumina sequencing and *de novo* assembly

The cDNA library of the inflorescence of *O. italica*, obtained from high quality total RNA (RIN= 9.0), was sequenced with the Illumina technology. After filtering reads of good quality (Phred quality score ≥33) and without contaminants, we obtained ≈ 94 million of paired-end (PE) 100-bp reads (86.2% of the original reads). The cleaned reads were processed using the *de novo* assembler Trinity (see methods), that generated 132,565 assembled transcripts. These were clustered into 86,079 not redundant transcripts (unigenes) based on their sequence identity (set to 85%). Table 2 shows that the N50 value is 956 and that the mean size of the unigenes is 606 bp, indicating a good quality of the assembled transcriptome.

	Number	N_{50}	Mean	Min	Max	>1000	>2000	>3000	>5000	>10000
Starting reads	108,911,910									
After contaminant cleaning	108,738,395									
After quality checking/adaptor trimming	93,926,808									
Assembled transcripts	132,565	786	564	201	12,047	18,004	4,357	1,210	143	3
Unigenes	86,079	956	606	201	12,047	13,996	3,928	1,185	140	3

Table 2 Summary statistics of sequence assembly from inflorescence of *O. italica*. Mean, Min and Max indicates the average, minimum and maximum length expressed in base pairs.

The analysis of the size distribution of the assembled transcripts and unigenes (fig. 14 A, B) indicated that the size ranging between 200 and 300 bp is the most abundant. 13.6% of the transcripts were more than 1,000 bp in length and 32.3% were more than 500 bp. Among the unigenes, 16.3% were more than 1,000 bp , and 33.2% were more than 500 bp.

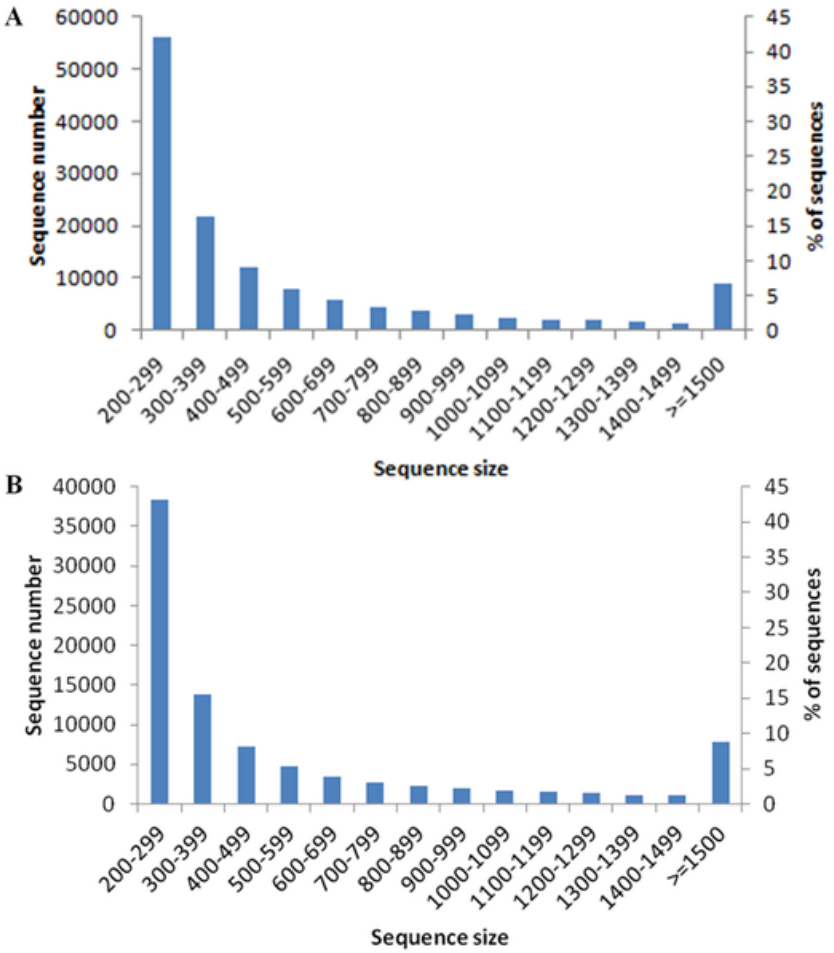

Figure 14 Reprinted from De Paolo *et al*. (2014) [191]. Sequence size distribution of the assembled transcripts (A) and unigenes (B) of the inflorescence of *O. italica.* The lengths are indicated in base pairs.

The comparison of the assembled transcriptomes of other orchids obtained with the same sequencing approach and that of *O. italica* revealed that the number of transcripts and unigenes assembled for the inflorescence of *O. italica* is higher than that of *Cymbidium ensifolium* [192] and similar or slightly lower than those assembled for mixed vegetative and reproductive tissues of *Cymbidium sinense* [193] and *Erycina pusilla* [194]. Among the orchid transcriptomes currently available [195-197], that of *Ophrys* (Orchidoideae) is the closest to *O.italica* and for this reason it was used for more specific comparative analyses. However, the trascriptome of *Ophrys* was obtained applying combined approaches of various NGS techniques and from a mixture *of O. exaltata, O. garganica* and *O. sphegodes.* The trascriptome of *Ophrys* includes 51,795 contigs (Illumina data) and 70,122 singletons (454 and Sanger data) [197]. The difference in the number of assembled transcripts could be related to the different sequencing approaches and to their great diversity in genome size. Indeed, Orchidaceae are the angiosperm family with the most variable genome size.

In particular, the genome size estimated for the genus *Orchis* and *Ophrys* (subfamily Orchidoideae) is 8.6 Gb and 10 Gb respectively, while for *Cymbidium* and *Erycina* (subfamily Epidendroideae) 4 Gb and 1.7 Gb respectively [198].

3.2.2 Functional annotation

The transcriptome of *O. italica* was annotated using the web platform FastAnnotator (see methods). The 45,3% (38,984) of all the unigenes matched at least one significant hit against the NCBI nr protein database (Table 3).

	All	NCBI-nr	GO	Enzyme	Pfam	KOG	KEGG
Number of unigenes	86,079	38,984	32,161	3,085	32,011	15,775	7,143
% of unigenes	100	45.3	37.4	3.6	37.2	18.3	8.3

Table 3 Statistics of the annotation results for the *O. italica* unigenes.

Figure 15 shows that the percentage of annotated unigenes of *O. italica* was positively correlated with the sequence length (Pearson correlation coefficient r = 0.57, p=0.001).

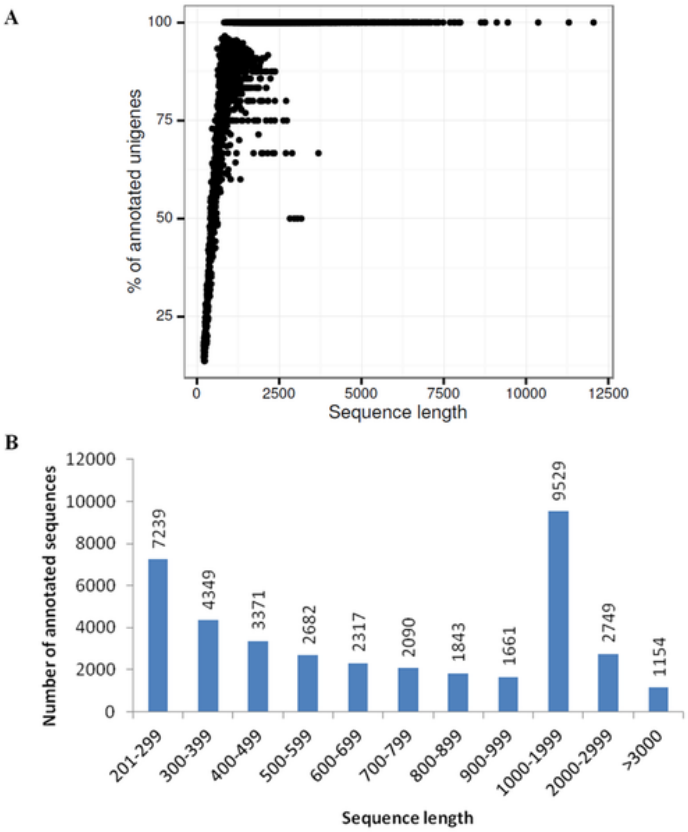

Figure 15 From De Paolo *et al.* (2014) [191]. Size distribution of the annotated transcripts. (A) The relationship between the sequence length of the assembled unigenes and the percentage of annotations in the NCBI nr protein database. (B) Sequences size distribution of the annotated unigenes. The lengths are indicated in base pairs.

BLASTN analysis between the unannotated unigenes of *O. italica* and *Ophrys* revealed that only the 1,3% of the unannotated unigenes of *O. italica* is best reciprocal hit of *Ophrys*. So, probably most of the unannotated unigenes of *O. italica* are novel transcripts**.** The size distribution analysis of the annotated transcripts revealed that the most abundant class had a sequence length between 1,000 and 2,000 bp (fig. 15 B).

The annotated unigenes are divided in functional categories belonging to three main classes of GO terms. Among them, the most abundant class was biological process (28,558 unigenes), followed by molecular function (27,378) and cellular component (24,304) (fig. 16 A).

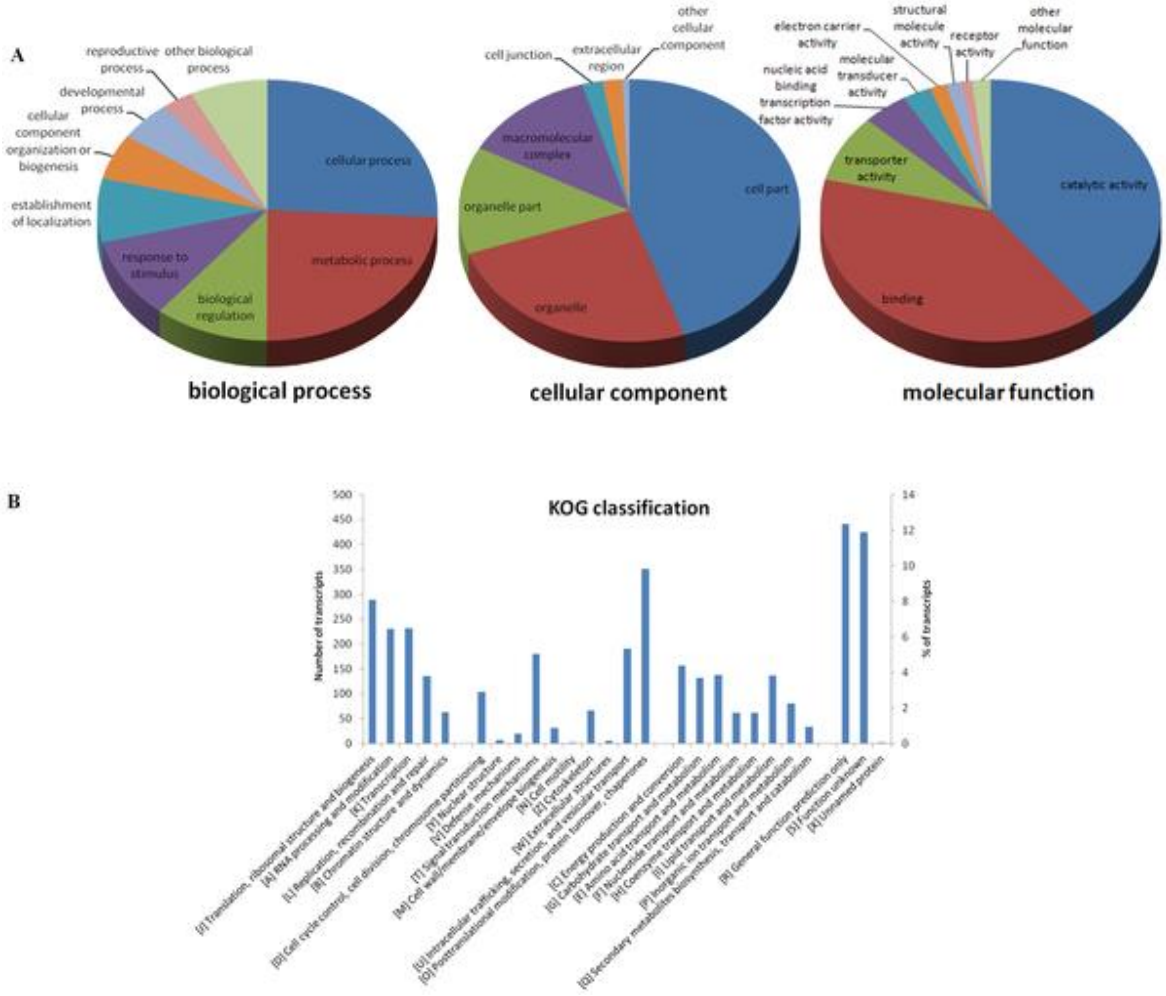

Figure 16 Reprinted from De Paolo *et al.* (2014) [191]. Functional annotations of the unigenes of *O. italica.* (A) Level 2 GO term distribution. (B) KOG annotation.

The comparison of the GO terms between *O. italica* and *Ophrys* revealed difference in the total number of the uniges as well as in the relative level in each class. In *Ophrys* the most abundant class is molecular function, followed by biological process and cellular component and the total numbers are lower than those reported for *O. italica.* In the biological process terms, the most abundant classes were cellular and metabolic process (25.8% and 24.2%, respectively). Among the molecular function terms, most of the unigenes were assigned to binding (39.1%) and catalytic activity (39.7%), while in the cellular function category the classes with the highest number of assigned unigenes were cell part (44.8%) and organelle (24.8%). The level 2 GO classification of the transcriptome of *O. italica* is in agreement with that reported for *Ophrys* [197].

To identify the putative orthologs and paralogs in the unigenes of *O. italica,* we conducted a search within the KOG database (see methods). Among the total unigenes of *O. italica*,

15,775 (18.3%) were assigned to 26 eukaryotic orthologous groups (Table 3, fig. 16 B). The figure 16B shows that the most represented were the general functions (R, 12.3%), unknown functions (S, 11.9%) and post-translational modifications, protein turnover and chaperones (O, 9.8%). With limited differences only in some groups, the KOG classification is in general agreement with those reported for *Ophrys* [197].

An additional functional annaotation was conducted to identify transcription factors within the assembled transcripts of *O. italica*. A search against the Plant Transcription Factor Database was performed using *A. thaliana* and *O. sativa* as reference dicot and monocot species, respectively. The results of this analysis revelaled that a total of 4,095 unigenes (4.8%) matched with 57 plant transcription factor families (fig. 17).

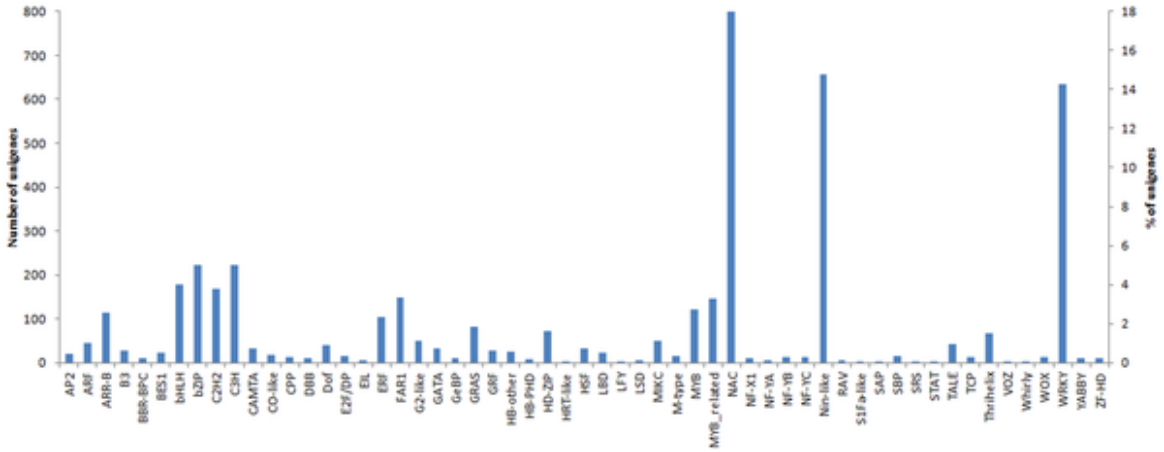

Figure 17 Reprinted from De Paolo *et al*. (2014) [191]. Annotations of the unigenes of *O. italica* obtained from the plant transcription factor database (TFDB).

Among them, the most abundant TF families were NAC (18.1%), Nin-like (14.7%) and WRKY (14.3%); however, also other families were well represented, such as those involved in flower development (MYB, AP2, LFY, MIKC, TCP). These results are partially in agreement with those reported for *Ophrys*, where the number and percentage of total unigenes is lower than those reported for *O. italica* (2,7% versus 4,8%). The most abundant TF families in *Ophrys* were comparable to those reported for *O. italica*: WRKY (21.4%), NAC (7.8%) and NF-YA (12.5%) while the Nin-like family was at 5.8% [197]. The search conducted in the in the Pfam database (with coverage greater than 50%) indicated that 32,011 (37.2%) of the assembled transcripts of *O. italica* matched with 7,208 protein domains (Table 3). Among

them, the most highly represented was the PPR domain followed by RVT_2 and Pkinase (Table 4).

Short name	Accession	Description	Occurrence
PPR_2	PF13041	Pentatricopeptide repeat family	825
RVT_2	PF07727	Reverse transcriptase (RNA-dependent DNA polymerase)	670
Pkinase	PF00069	Protein kinase domain	527
rve	PF00665	Integrase core domain	372
ABC_tran	PF00005	ATP-binding domain of ABC transporters	252
MFS_1	PF07690	Major facilitator superfamily	249
LysR_substrate	PF03466	LysR substrate binding domain	197
Pkinase_Tyr	PF07714	Tyrosine kinase	191
RVT_1	PF00078	Reverse transcriptase	183
UBN2_3	PF14244	gag-polypeptide of LTR copia-type	173
AMP-binding	PF00501	AMP-binding enzyme	167
RRM_1	PF00076	RNA recognition motif	167
gag_pre-integrs	PF13976	gag-pre-integrase domain	156
WD40	PF00400	WD40 repeat	144
Tymo_45kd_70kd	PF03251	Tymovirus 45/70Kd protein	140
Retrotrans_gag	PF03732	Retrotransposon gag protein	138
BPD_transp_1	PF00528	Binding-protein-dependent transport system inner membrane	131
LRR_8	PF13855	Leucine-rich repeat	127
ACR_tran	PF00873	AcrB/AcrD/AcrF family integral membrane proteins	124
adh_short	PF00106	Short-chain dehydrogenase	124
Response_reg	PF00072	Response regulator receiver domain	122
Aldedh	PF00171	Aldehyde dehydrogenase family	121
DYW_deaminase	PF14432	DYW family of nucleic acid deaminases	120
Myb_DNA-binding	PF00249	Myb-like DNA-binding domain	118
zf-RING_2	PF13639	RING finger domain	118
p450	PF00067	Cytochrome P450	116
Abhydrolase_6	PF12697	Alpha/beta hydrolase fold	114
TonB_dep_Rec	PF00593	TonB-dependent receptors	102
Other domains			26,022

Table 4 Summary of the Pfam domain annotations with occurrence > 100.

The PPR domain proteins are involved in various aspects of plant physiology and development by RNA editing and/or regulation of the mRNA turnover and translation [199]. The RVT proteins are reverse transcriptases and the presence of a high number of members in the transcriptome of *O. italica*, together with other abundant protein classes such as rve, gag_pre-integrs, Retrotrans_gag, is in agreement with the high number of mobile elements reported in the orchid genomes [147, 200]. The protein kinases (Pkinase), involved in cell proliferation, differentiation and death [201], are also highly represented in the transcriptome of *Ophrys*, and are also related to the presence of mobile elements [197]. An additional search was conducted in the Enzyme database (Table 3) which revealed that 3,085 transcripts (3.5%) had at least one enzyme hit (Table 3).

To analyze the involvement of the assembled transcripts in biochemical pathways, the unigenes of *O. italica* were used for a functional annotation in the KEGG database. The results reveals that among all the transctipts, 7,143 (8.3%) matched with 2,651 enzymes involved in essential biochemical pathways (Table 3 and 5). Table 6 shows that the enzymes involved in metabolism are the most represented, followed by genetic information processing, cellular processes, environmental information processing and organismal systems.

KEGG pathway	N unigenes	N enzymes
Metabolism		
Global and overview maps	3032	1116
Carbohydrate metabolism	977	271
Amino acid metabolism	654	458
Lipid metabolism	433	165
Energy metabolism	278	61
Biosynthesis of other secondary metabolites	241	53
Metabolism of other amino acids	227	56
Metabolism of cofactors and vitamins	206	120
Nucleotide metabolism	192	73
Metabolism of terpenoids and polyketides	139	78
Glycan biosynthesis and metabolism	79	43
Genetic Information Processing		
Folding, sorting and degradation	178	24
Translation	109	37
Replication and repair	81	25
Transcription	48	8
Cellular Processes		
Transport and catabolism	112	28
Environmental Information Processing		
Signal transduction	87	25
Organismal Systems		
Environmental adaptation	58	10

Table 5 Summary of the KEGG pathways analysis indicating the number (N) of unigenes and the number of corresponding enzyme matches.

3.2.3 Expression analysis

The RSEM software was used to evaluate the expression level of the unigenes of *O. italica* (see methods). The results revealed that the 36.4% of the unigenes presented a FPKM value lower than 1 and were considered as unexpressed; the 50.1% of the unigenes presented a FPKM value between 1 and 10 and were considered as moderately expressed, while the

12.2% were considered moderately expressed with a FPKM values between 10 and 100. The unigenes with FPKM values higher than 100 (1.3%) were considered highly expressed.

To validate the *in silico* analysis, the expression level of ten selected unigenes was measured by real-time RT-PCR. Among them, one is a housekeeping gene and the other nine unigenes encode transcriptional factors involved in the ABCDE model of flower development (Table 7). Both measures were normalized relative to the actin levels.

The R0 values of each gene was divided by the R0 value of the actin to obtain the mean Rn value, while the FPKM value was normalized (FPKMn) by dividing the FPKM value of each unigene by the FPKM value of the actin. The resulting values were compared and the Pearson correlation coefficient showed a strong positive correlation between the two datasets (r= 0.87, p= 0.002) (fig. 18, Table 7). These results demonstrated that the *in silico* analysis indicates with a good approximation the real expression level of the transcripts in the inflorescence tissue.

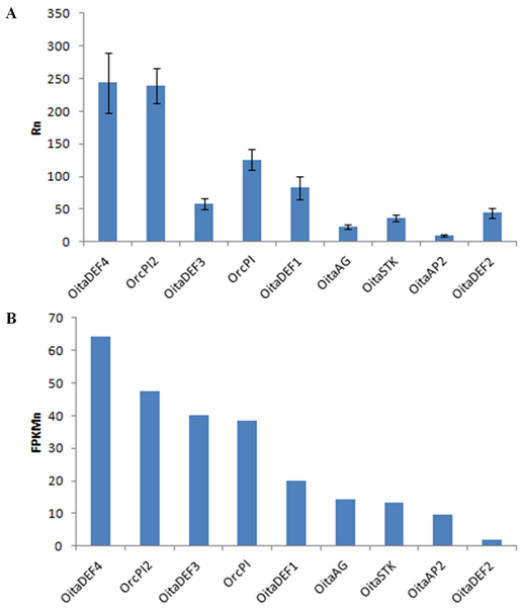

Figure 18 Reprinted from De Paolo *et al*. (2014) [191]. Relative expression levels of selected unigenes of *O. italica* detected by real-time PCR in inflorescence tissue (A) and by normalized FPKM counts (B). The bars indicate the standard deviation.

3.2.4 Non-coding transcripts

Despite the increasing interest in the study of the plant lncRNAs for their involvement in a wide range of regulatory processes, to date lncRNAs in orchids are completely unknown. The development of *ad hoc in silico* analysis tools has facilitated the ability to predict potential lncRNAs. To understand if a unannotated assembled transcript is a lncRNA or a misassembled sequence, it would preferable to have the assembled genome but unfortunately it is absent for *O. italica* (and at that time also for other orchid species). To identify the putative lncRNAs assembled in the inflorescence transcriptome of *O. italica,* a preliminary analysis was conducted using separately two different software packages: Coding Potential Calculator (CPC) and Portrait (see methods). The 47,097 unannotated unigenes were analyzed and the results reveled 45,266 (CPC) and 7,888 (Portrait) potential non-coding transcripts, with 7,779 transcripts matching both thresholds. Within the group of the unannotated unigenes, 10 were selected to verify their existence in *O. italica* and exclude they were assembly artifacts (Table 8).

Among them, 7 matched both the CPC and Portrait threshold values, two were chosen for their length (~1,000 bp) but matched only the CPC threshold and one did not match any threshold but was chosen for its high FPKM value (20,357).

A first analysis was conducted by RT-PCR amplification on total RNA extracted from inflorescence of *O. italica*. The figure 19A shows that single amplification product of the expected size was obtained for 7 of the 10 analyzed transcripts. Among the remaining three transcripts resulting in multiple fragments (fig. 19 A, lane 7–9), two matched only the CPC threshold. Six out of seven amplification products of the expected size were confirmed by cloning and sequencing, while one resulted a contaminant sequence (fig. 19 A, lane 4).

The selected six non-coding transcripts were used to performed a Real-time PCR experiment to analyze their expression pattern in different floral tissues and leaf of *O. italica* (fig. 19 B–G).

All the transcripts showed a variable expression profile in the examined tissues. They were absent in the ovary and absent or weakly expressed in the leaf.

The expression analysis of the comp0_c0_seq1 and comp3328_c0_seq1 transcripts showed that they are mainly expressed in the column (fig. 19 B and C, respectively), suggesting a possible role in regulation the development of the reproductive tissues. The comp1231_c0_seq1, comp48038_c0_seq1 and comp6669_c0_seq1 transcripts (fig. 19 D–F, respectively) are expressed almost exclusively in the perianth organs. The comp134696_c0_seq1 transcript (fig. 19 G) is expressed almost exclusively in inner tepals and absent or weakly expressed in the other tissues. The results of this analysis revealed the presence of specific putative lncRNAs in the perianth of *O. italica* suggesting their possible role in flower development processes.

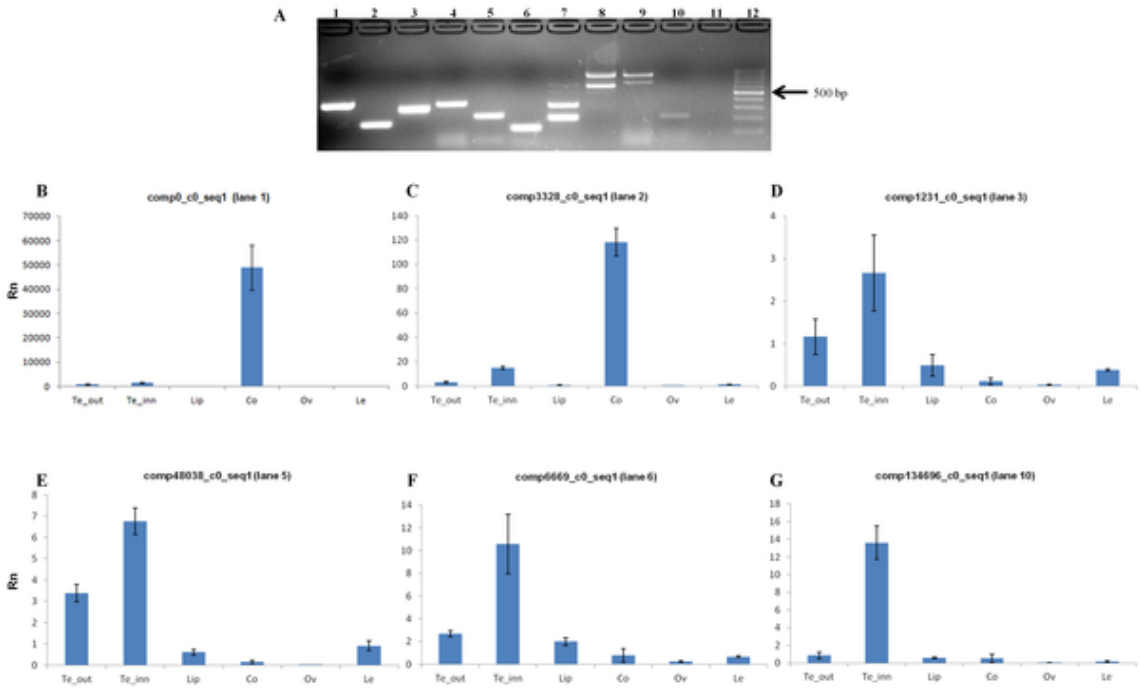

Figure 19 Reprinted from De Paolo *et al*. (2014) [191]. Selected putative long non-coding RNAs expressed in the inflorescence of *O. italica*. (A) Agarose gel electrophoresis of the RT-PCR of the selected transcripts (Lane 1, comp0_c0_seq1; lane 2, comp3328_c0_seq1; lane 3, comp1231_c0_seq1; lane 4, comp3311_c0_seq1; lane 5, comp48038_c0_seq1; lane 6, comp6669_c0_seq1; lane 7, comp4129_c0_seq1; lane 8, comp1308_c0_seq1; lane 9, comp15481_c0_seq1; lane 10, comp134696_c0_seq1; lane 11, empty; lane 12, 100 bp ladder). (B–G) Relative expression profile of the transcripts comp0_c0_seq1 (B), comp3328_c0_seq1 (C), comp1231_c0_seq1(D), comp48038_c0_seq1 (E), comp6669_c0_seq1, (F), and comp134696_c0_seq1 (G) in the outer tepals (Te_out), inner tepals (Te_inn), labellum (Lip), column (Co), ovary (Ov) and leaf (Le). Rn, relative expression ratio. The bars indicate the standard deviation.

The BLASTN search performed on the comp134696_c0_seq1 revealed that it is a homolog of the *TAS3* long non-coding transcript (Fig. 20) from which it is known that derive tasiRNAs involved in the regulation of the the auxin response factor genes [108, 125].

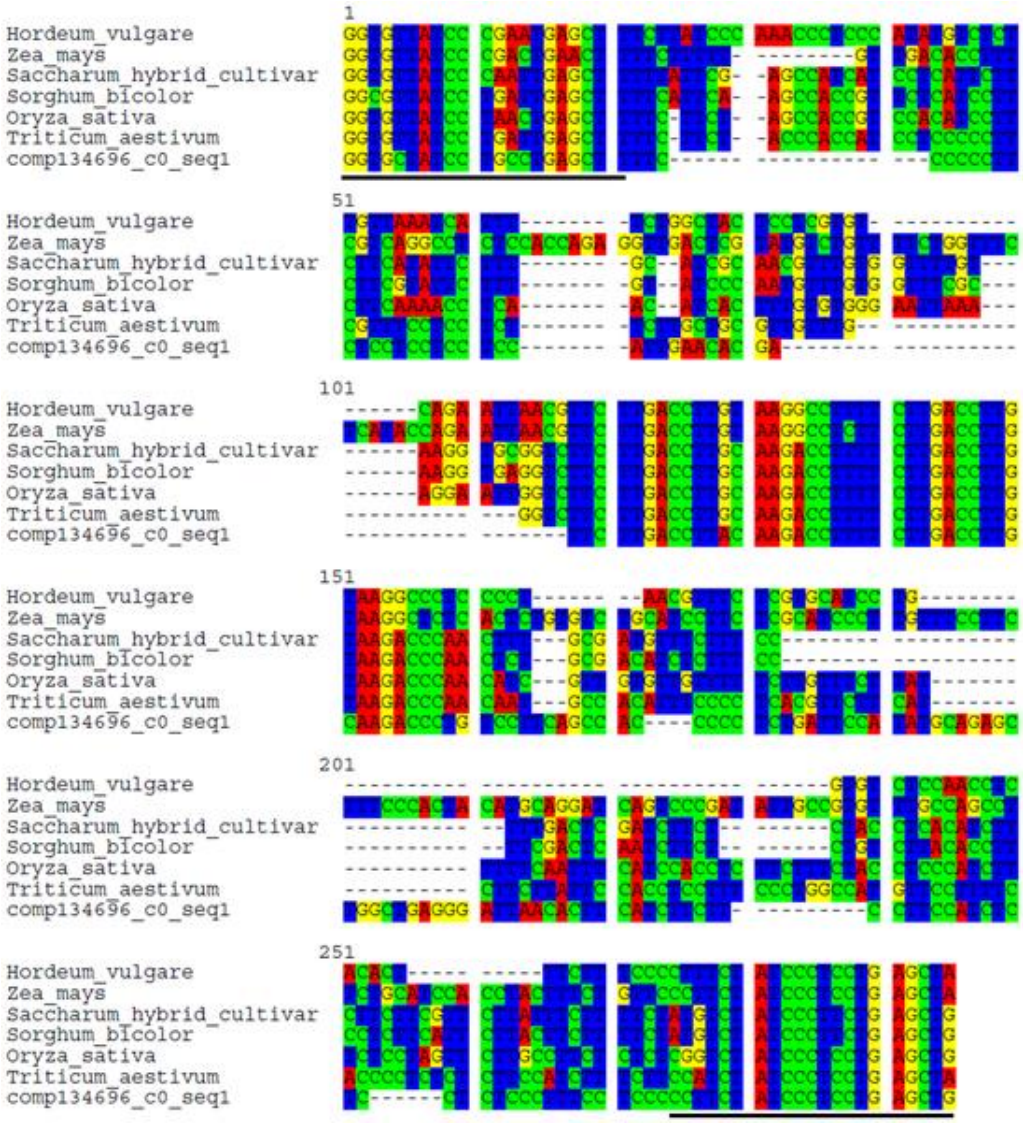

Figure 20 Reprinted from De Paolo *et al.* (2014) [191]. Alignment of nucleotide sequence of the comp134696_c0_seq1 identified in *O. italica* and the TAS3 sequences of *Hordeum vulgare* (accession number BF264964), *Zea mays* (BE519095), *Saccharum* hybrid cultivar (CA145655), *Sorghum bicolor* (CD464142), *Oryza sativa* (AU100890) and *Triticum aestivum* (CN010916).

The biogenesis of tasiRNAs is regulateted by miR-390 whose homolog in *O. italica* is differentially expressed in the various tissues of the inflorescence of *O. italica* [176].

In addition, the discovery of other transcripts involved in the biogenesis of tasiRNAs, RDR6 (comp44794_c0_seq1), DCL4 (comp3192_c0_seq1) and 11 transcripts matching different AGO proteins in the annotated transcriptome of *O. italica,* suggests the presence of a

conserved pathway for the *TAS3* ta-siRNA biogenesis in plant and their possible role in the floral organs. Interestingly, the hypothesis of a possible functional role of tasiRNAs in flower development is a novelty because previous studies conducted on the processes regulated by tasiRNAs only concern response to pathogens, lateral roots development and leaf morphology or transitions to juvenile/adult stage.

The results obtained have been published in De Paolo *et al*. [191].

3.3 Analysis of the *TCP* genes of *Orchis italica*

3.3.1 The TCP genes in the inflorescence transcriptome of *O. italica*

To identify transcripts encoding TCP proteins in the inflorescence transcriptome of *O. italica* a TBLASTN search was performed using as query the sequences of the TCP DNA-binding domain of *A. thaliana* and *O. sativa*. The search reveled 11 different transcripts containing a region encoding the TCP domain (Table 9). This result increased the number of known *TCP* genes of orchids, where only five *TCP* genes have been reported [85].

After the virtual translation of the 11 *TCP* sequences of *O. italica*, 24 of *Arabidopsis thaliana* and 22 of *Oryza sativa*, the amino acid alignment was used to construct a Neighbor-Joining (NJ) tree. However, this first phylogenetic analysis revealed that none of the *TCP* transcripts expressed in the inflorescence of *O. italica* belong to the class II CYC/TB1-like group, whose members are involved in the establishment of bilateral symmetry in numerous plant species [82, 92, 98, 99, 202, 203].

To verify the presence of CYC/TB1-like genes within orchids, standalone BLASTN and BLASTX searches were performed using the *TCP* transcripts of *O. italica* as queries on the recently released genome of *P. equestris*. The results revealed that in the genome of *P. equestris* there are 17 genes encoding TCP proteins and among them three belong to the CYC/TB1-like group. Based on the CYC/TB1-like sequences of *P. equestris*, degenerate primers were designed spanning from the TCP to the R domain, in order to amplify the genomic DNA of *O. italica* and to check for the presence of CYC/TB1-like sequences also in *O. italica.* The resulting amplicon (*OitaTB1,* 378 bp) has homology with *CYC/TB1*-like genes and was deposited in GenBank with the accession number KR858306. To verify if the absence of the *OitaTB1* transcript in the transcriptome of *O. italica* was due to a transcriptome mis-assembly, RT-PCR amplification was conducted on RNA extracted from various tissues of the inflorescence and from leaf tissue of *O. italica.* However, every attempts failed. This result suggests that *OitaTB1* is not expressed in the inflorescence of *O. italica* at the stage of development examined.

3.3.2 Phylogeny and analysis of conserved motifs

A phylogenetic analysis was performed on the multiple amino acid alignment of all the TCP domains encoded by the selected transcripts of *O. italica, A. thaliana* and *O. sativa* (fig. 21).

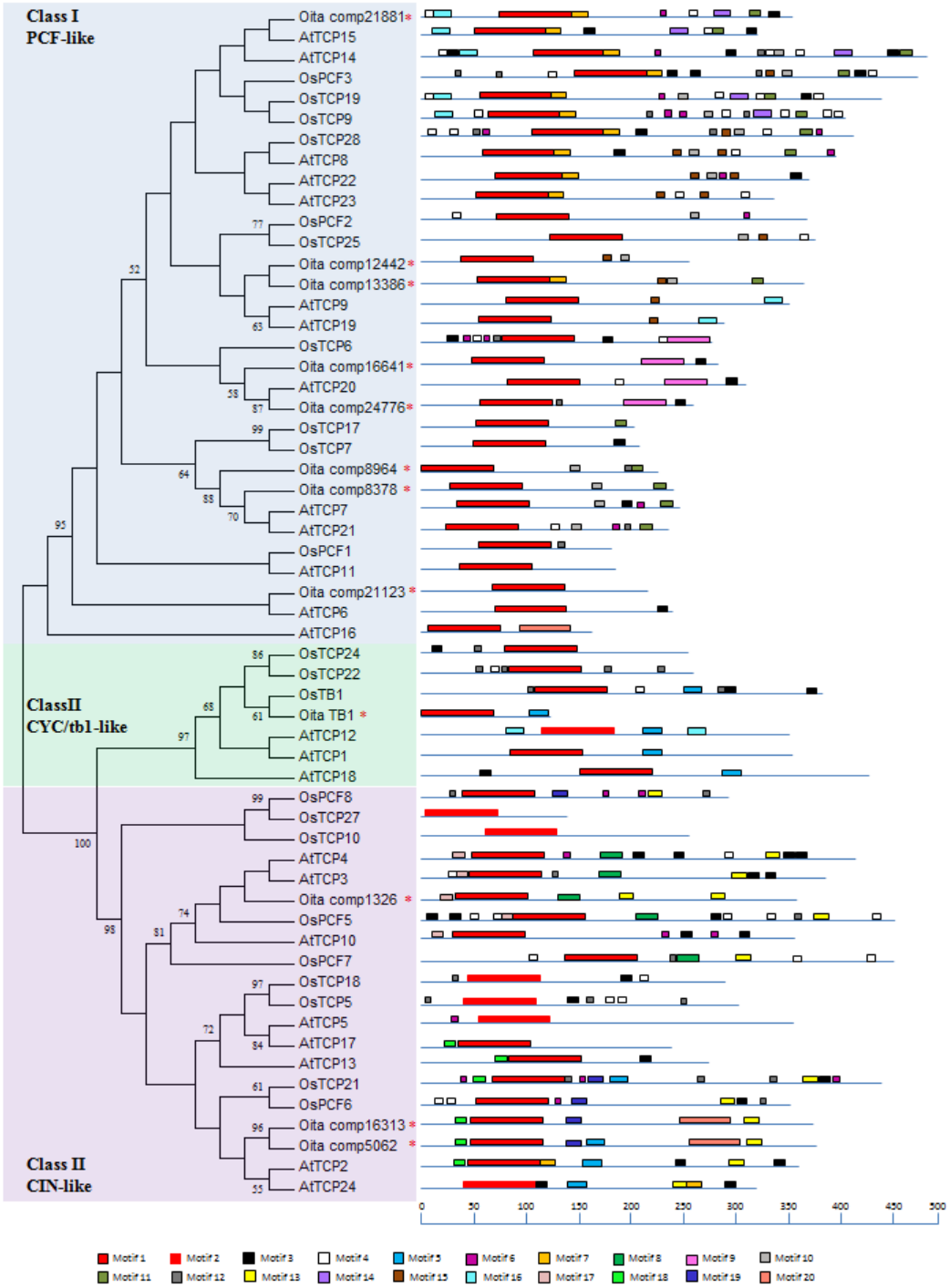

Fig. 21 Reprinded from De Paolo *et al.* (2015) [46]. The NJ tree of the TCP proteins examined and graphic representation of the conserved domains identified. On the left, the NJ tree was obtained from the amino acid alignment of the TCP domain of *Orchis italica*, *Arabidopsis thaliana* and *Oryza sativa.* The bootstrap percentages >50% are shown and the red asterisks indicate the sequences of *O. italica*. On the right, the graphical representation of the conserved domains and the relative legend obtained from the MEME search using the full length of the TCP proteins. The amino acid position is indicated from the scale below the conserved domains.

The NJ tree obtained is divided into two main branches, both statistically well supported (bootstrap value 95% and 98%, respectively). Among them, the first includes the class I PCF-like group, with eight *TCP* transcripts of *O. italica*; the second is divided into the CIN-like group (bootstrap value 98%) that includes three *TCP* transcripts of *O. italica* and the CYC/TB1-like group (bootstrap value 97%) that includes *OitaTB1*. The topology of the tree suggests that the expansion of the TCP family occurred before the divergence of the examined lineages and is in agreement with previous studies conducted in some dicot species [204–206].

To search for conserved shared domains, the amino acid sequences of the TCP proteins of *O. italica*, *A. thaliana* and *O. sativa* were scanned using the motif-based sequence analysis tool MEME (fig. 21). The pattern distribution of the conserved motifs is in general agreement with the NJ tree. The Motif 1 and 2 include the TCP domain and are present in all the sequences. The Motif 5 corresponds to the R domain and is shared by 9 sequences belonging to the class II TCP family among which two are sequences of *O. italica*, one CIN-like (comp5062) and one CYC/TB1-like (*OitaTB1*). The Motif 13 corresponds to the amino acid stretch encoded by the target site of the microRNA miR319 [76,165,207,208] and is shared by 12 sequences of the CIN-like group, among which three are sequences of *O. italica* (comp5062, comp1326 and comp16313). Other motifs are restricted to specific sub-groups of the tree and could play specific roles.

To understand the function of the identified conserved motifs, a scanning was performed against the database of protein domains PROSITE (see methods). Excluding the TCP domain, the other motifs have unknown function.

3.3.3 microRNA target sites and expression analysis

To verify whether the selected transcripts of *O. italica* encoding TCP-like proteins are target of specific miRNAs, they were scanned with the psRNATarget online tool using the inflorescence miRNAs of *O. italica* as queries (see methods). As expected from the conserved motifs analysis, the results predicted on three different transcripts of *O. italica* (comp5062, comp1326 and comp16313) a putative cleavage site for miR319 (fig. 22 A). Phylogenetic analysis shows that these three transcripts belong to the class II CIN-like group and that comp5062 and comp16313 are related to AtTCP2 and AtTCP24 of *A. thaliana* and comp1326 is related to AtTCP3 and AtTCP4 of *A. thaliana* and OsPCF5 of *O. sativa.* These

transcripts of *Arabidopsis* and *Oryza* have a target site for miR319 and are involved in different developmental processes [76, 150, 207-209]. In order to validate the *in silico* miRNA target analysis, a modified 5'RACE experiment was performed. The results confirmed the cleavage only for the transcript comp5062 (Fig. 22 B, C). Despite the *in silico* analysis predicted a putative target site also for the transcripts comp1326 and comp16316, no cleavage fragment of the expected size was detected.This results could be related to differences in the sequence of the upstream and downstream regions surrounding the predicted cleavage site.

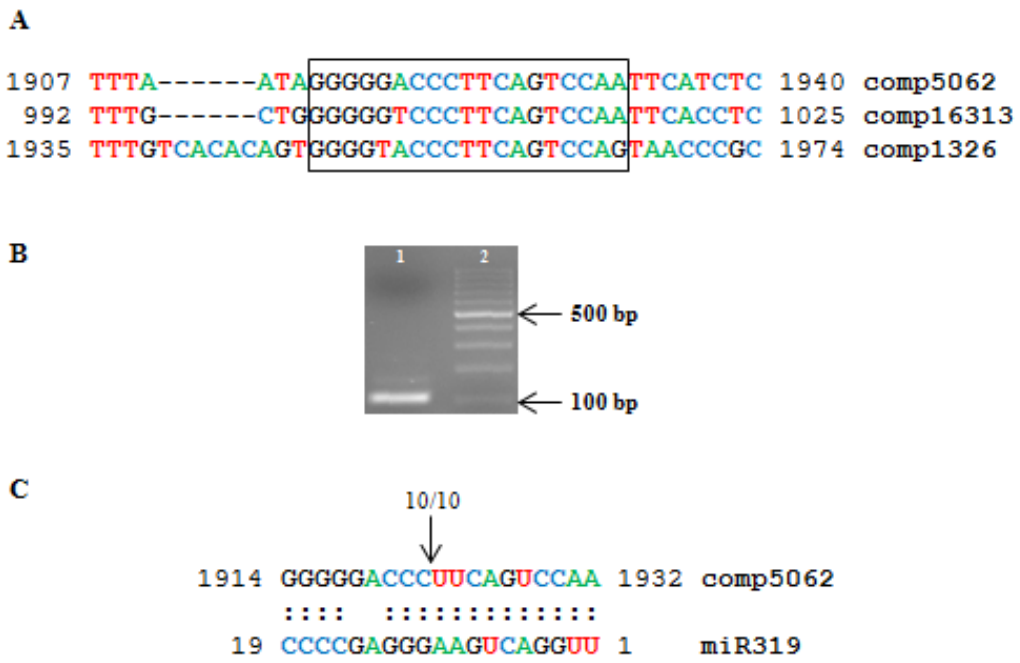

Figure 22 Reprinted from De Paolo *et al*. (2015) [46]. The cleavage site of miR319 on the TCP transcript of *O. italica* (A). In the box the nucleotide alignment of the miR319 target site on three TCP transcripts of *O. italica* predicted *in silico*. Numbers at the sides of the sequences indicate the nucleotide positions on the transcript. (B) Agarose gel electrophoresis of the modified 5'RACE experiment; in the lane 1 the TCP transcript comp5062 and in the lane 2 the 100 bp Ladder (Fermentas). (C) The alignment of the miR319 and its target site on the comp5062. The arrow indicates the position of the cleavage site and the numbers of sequenced clones that revealed the cleavage in that position.

Real-Time PCR experiments were performed to analyze the expression pattern of miR319 and of its putative target transcripts in different tissues of the inflorescence of *O. italica* at two developmental stages (fig. 23).

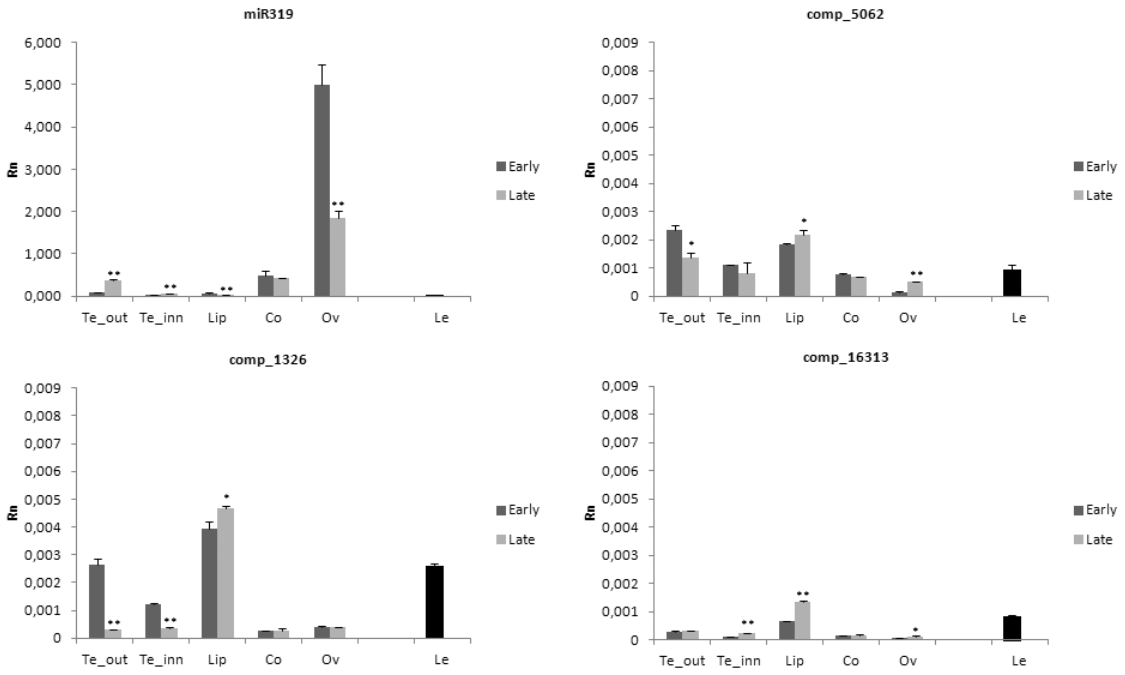

Figure 23 Reprinted from De Paolo *et al.* (2015) [46]. The expression profile of selected *TCP* putative target of miR319 (comp5062, comp1326 and comp16313) and of miR319 in different floral tissues and in leaves of *O. italica*. The bars indicate standard deviation. Statistically significant differences between the relative expression of the early and the late stages is indicated by asterisks (*p < 0.05, **p < 0.01). Te_out, outer tepals; Te_inn, inner tepals; Lip, labellum; Co, column; Ov, ovary; Le, leaf. Rn, relative normalized expression.

The three examined transcripts (comp5062, comp1326 and comp16313) show a similar expression pattern in the tissues of *O. italica* at the two different stages, with difference in the expression levels. In the inflorescence, the transcripts are mainly expressed in the tepals and lip, even if some differences between the early and late stage were detectable. In the column and ovary the three transcripts were expressed at levels lower than in the other tissues, including leaves. The microRNA miR319 shows a complementary expression profile, with the highest expression levels observed in the column and ovary. Although the expression of miR319 shows an opposite trend when compared to that of the three transcripts, the only statistically significant Pearson correlation coefficient ($r = -0.61$, $p < 0.05$) is relative to the transcript comp5062. This result is in agreement with that of the modified 5'RACE experiment, demonstrating for the first time the activity of miR319 on a *TCP* target in the floral tissues of a monocot species. In *Arabidopsis* miR319 is involved in the development of petal and stamens through the cleavage of a CIN-like *TCP* target [148]. This data suggests a possible conserved function of miR319 in flower development of dicots

and monocots. In addition, the expression profile of the transcripts comp1326 and comp16316, together with the failure to detect specific fragments cleaved by miR319, suggests the existence of an alternative transcriptional regulatory mechanism that does not involve cutting. The expression pattern of the other identified *TCP* transcripts was examined in different tissues of *O. italica* (fig. 24).

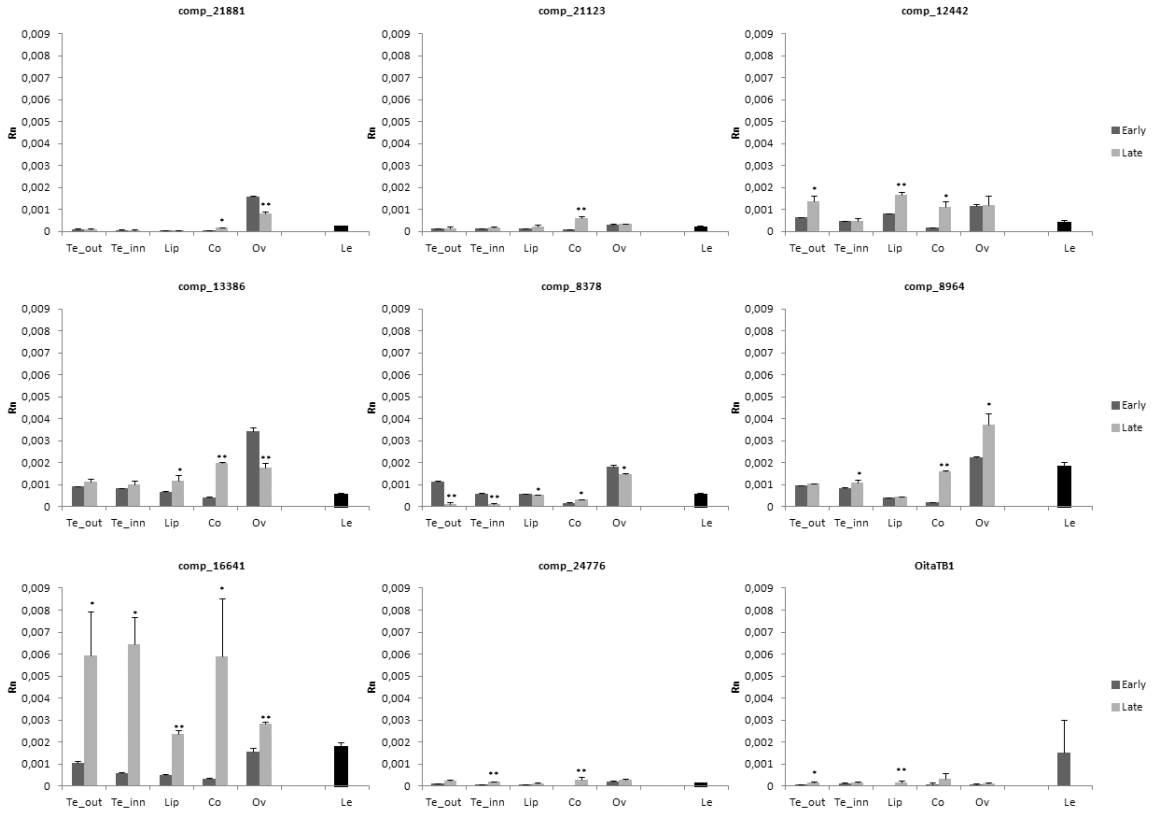

Figure 24 Reprinted from De Paolo *et al*. (2015) [46]. The expression profile of nine *TCP* transcripts in different floral tissues and in leaves of *O. italica*. The bars indicate standard deviation. Statistically significant differences between the relative expression of the early and the late stages is indicated by asterisks (*p < 0.05, **p < 0.01). Te_out, outer tepals; Te_inn, inner tepals; Lip, labellum; Co, column; Ov, ovary; Le, leaf. Rn, relative normalized expression.

The results revealed similar patterns for some transcripts and different for others that probably have distinct function in the tissues and stages examined. For example, the transcript comp21881 is expressed mainly in the ovary tissue and is related to the gene *AtTCP15* of *Arabidopsis* (fig. 21) involved in the developmental pathways of the gynoecium [210, 211]. So, the expression pattern observed in *O. italica* revealed a possible functional conservation in the development of the female reproductive structures. The transcript

comp21123 is expressed almost exclusively in the column at the late stage, suggesting that it might have a specific function in this tissue.

Among the transcripts belonging to the same branch of the NJ tree, some showed a similar expression pattern while others revealed a distinct profile. In particular, the couples comp12442-comp13386, comp8378-comp8964 and comp16641 and comp24776 are phylogenetically close.

The comp12442 and comp13386 display generally overlapping profiles, with some differences at the two developmental stages in the ovary tissue. These results suggest a possible pleiotropic and redundant function of these two transcripts in the tissues examined of *O. italica.* The transcripts comp8378 and comp8964 show an overlapping pattern at the early stage while at the late stage the comp8964 shows an overall level of expression higher than that of the comp8378. These two transcripts are related to the gene *AtTCP21*, involved in leaf development [212] and *AtTCP7* that is a component of the circadian clock [213] (fig. 17). The expression profile of these two related transcripts revealed a possible functional diversification also in *O. italica.* The transcripts comp16641 and comp24776 show expression profiles very different. The comp16641 is highly expressed in all the floral tissues at the late stage and in leaves, the comp 24776 shows a very low level of expression in all the tissues, indicating that probably it acts in different organs and/or developmental stages. The transcript comp16641 shows the strongest variation of the expression profile between the two stages examined. This result suggests a possible sub- or neo-functionalization of these transcripts in *O.italica* that could have a role in the development and maintenance of floral structures in orchids.

Finally, *OitaTB1* is weakly expressed in all the examined tissues, with a slightly higher level in leaves than in the other tissues. *OitaTB1* is related to *OsTB1* of *Oryza sativa* that is involved in the development of lateral branching [214]. However, the understanding of the role played by this transcript in the development of *O. italica* and in orchids in general still remains an open question.

The results of this study were published in De Paolo *et al.* [46].

4. Conclusions

The development of this PhD project was based on the application of the NGS approach to study small and long transcripts expressed in the floral tissues of the orchid *O. italica*, increasing the RNA-seq data currently available for orchids.

The analysis of the miRNome revealed the presence of evolutionary and taxonomically conserved miRNA families involved in flower development and the presence of novel miRNAs that might be considered orchid-specific. For the first time, putative lncRNAs were also identified in the floral organs of an orchid species using a new *in silico* approach that could be used to extend similar investigations in the analysis of other non-model species. The *in silico* analyses conducted to identify transcripts containing putative target sites for miRNAs showed for the first time that a miRNA (miR5179) acts specifically on a class B MADS-box mRNA (*OitaDEF2*). This result highlights the involvement of a miRNA in the diversification of the organs of the perianth in orchids.

The presence of flower-specific long non-coding transcripts, differentially expressed in the various tissues of the perianth of *O. italica*, suggests they might have a relevant role in flower development, expanding the ‘‘orchid code’’ theory. This innovative hypothesis requires further investigations to be confirmed also in other species, in order to clarify the possible role of the lncRNAs during the flower development.

The transcriptome-wide analysis of one of the gene families involved in the establishment of floral symmetry, the TCP transcription factors, showed the presence of 12 TCP transcripts in the inflorescence of *O. italica*. This number is lower than that reported in the model species *A. thaliana* (24) and *O. sativa* (22). Even though it is possible that other *TCP* genes are present in the genome of *O. italica*, probably expressed in other tissues and/or developmental stages, the number of identified TCP transcripts of *O. italica* is similar to that we found in the genome of the orchid *P. equestris* (17), suggesting that in orchids there are fewer *TCP* genes than in *Arabidopsis* and *Oryza*. The analysis of the expression profiles of the *TCP* transcripts of *O. italica* indicated that some of them could have pleiotropic and/or redundant effects, being expressed in all the tissues and developmental stages examined, while others seem to have specific functions, showing an expression profile restricted to specific tissues/stages. As for the class B MASD-box genes, in *O. italica* also

the expression of some members of the TCP family is regulated by a specific miRNA (miR319), supporting the existence of an evolutionary conserved mechanism that regulates the TCP gene expression through small RNAs.

5. Material and Methods

5.1 Material

5.1.1 *Orchis italica*

The tissues used in this work were dissected from the inflorescence of *O. italica* before anthesis (defined as the early stage) and after anthesis (defined as late stage). The early stage corresponds to floral buds with a diameter of approximately 9 mm and the late stage to completely open flowers after anthesis (fig. 4). Although in the early stage cell division has been completed and flower organs formed, cell elongation is still occurring. Outer tepals (Te_out), inner tepals (Te_inn), labellum (Lip) and column (Co) were collected from both develpmental stages of the inflorescence of *O. italica*. Ovary was collected before (Ov) and 3 (Ov_3dap), 7 (Ov_7dap) and 10 (Ov_10dap) days after pollination. Before the pollination the ovules of *O.italica* contain megaspore mother cells that are in the first meiotic division. Their maturation occurs at 3 days after pollination and 7 days after pollination the seeds are in the early developmental stage. At 10 days after pollination the seeds are mature (Barone Lumaga, personal communication). Single florets were collected from the bottom of a single inflorescence of *O. italica* before anthesis. Although not synchronous, the selected florets displayed approximately the same size and could be considered in the same developmental stage. Leaf tissue was also collected.

5.2 Methods

5.2.1 RNA extraction

Using the Trizol Reagent (Ambion), total RNA was extracted from ten pooled florets and from all the different tissues at both the developmental stages. After DNase treatment, RNA was quantified using the Nanodrop 2000c spectrophotometer (ThermoScientific). The integrity of the extracted RNA was assessed by measuring the RNA integrity number (RIN) using the Agilent 2100 BioAnalyzer (Agilent).

5.2.2 RNA library construction and sequencing

5.2.2.1 Small RNA library

Using the total RNA extracted from 10 pooled florets, small RNA library preparation and sequencing were carried out according to the manufacturer instruction (Illumina). The library obtained was sequenced using the MiSeq instrument (Illumina).

5.2.2.2 Long RNA library

The Long RNA library preparation was performed using the total RNA extracted from 10 pooled florets. The Illumina sequencing was performed at Genomix4Life S.r.l. (Salerno, Italy) following the Illumina TruSeq Stranded sample preparation protocol. Paired-end (PE) strand-specific sequencing was performed on an Illumina HiSeq 1500 instrument following the supplier-provided protocols and generating 100 nt long reads.

5.2.3 *In silico* analysis of the small RNA reads

The plant version of the UEA sRNA workbench was used to process the raw reads of inflorescence tissue of *O. italica* [215,216] in order to remove the adaptor sequences, the low quality reads and the reads with abundance lower than 5. Filtering options were set to include in the analysis only the sequences with a length ranging from 18 to 35 nucleotides. After the removal of tRNA and rRNA sequences that was carried out using the Bowtie aligner v 1.0 [217], the reads were collapsed to estimate the number of different sequences and for each the read count was summed. In order to identify the conserved miRNAs in the inflorescence of *O. italica,* the pre-filtered reads were used as query in a standalone BLAST search against the known plant mature and hairpin miRNA sequences downloaded from mirBase 20 [218]. Reads matching at least 18 nt and with less than 3 mismatches were considered positive. In order to identify the new small RNAs of the inflorescence of *O. italica* we used the miRDeep-P software [219]. This software is specific to identify plant miRNAs. To perform this analysis we needed a reference transcriptome to conduct a bowtie alignment with the collapsed short reads. Hovewer, at that time the *O. italica* transcriptome was not available and we used that of the orchid *Phalaenopsis aphrodite* [144]. Using the deep sequencing data deposited in the Sequence Read Archive under the accession code SRA030409, we assembled the vegetative, seed and inflorescence

transcriptome of *P. aphrodite* with the Trinity software [220]. From these three tissue-specific transcriptomes we obtained a not-redundant collection of unigenes to use as reference transcriptome. The annotation of the assembled transcriptomes were performed using the FastAnnotator online tool [221]. The bowtie alignment was performed setting the parameters with a maximum of three mismatches and reads mapping to multiple positions (maximum 15) were retained. The potential miRNA precursors were then selected by setting the maximum length to 250 nt and their secondary structure was predicted using RNAfold.

5.2.4 *In silico* analysis of the long RNA reads

5.2.4.1 Pre-processing, assembly and clustering

Using Trimmomatic, we conducted the quality control by sliding window analysis and adapter trimming of the raw reads [222]. In order to remove the contaminating sequences matching with rRNAs, tRNAs, *Cymbidium mosaic virus* (accession number NC_001812), *Odontoglossum ringspot virus* (NC_001728) and *E. coli,* we used the Bowtie aligner v 1.0 [3] allowing for 2 mismatches (-v 2). Using Trinity 2013.11.10 [220, 223], we assembled the filtred reads obteined with the fixed default k-mer size of 25, minimum contig length of 200, maximum length expected between fragment pairs of 500 and a butterfly HeapSpace of 20 Gb. The similarity clustering of the assembled transcripts was performed using CDHIT EST [224] with an identity cut-off of 85%.

5.2.4.2 Functional annotation

Using FastAnnotator [221] with the default search parameters, we annotated the assembled transcripts. FastAnnotator assigns the Gene Ontology terms (GO) using the Blast2Go software [225]. In this way we identified the Pfam protein domains and the Enzyme Commission (EC) numbers. Executing a RPSTBLASTN search [226] against the NCBI KOG database (cut-off Evalue of e^{-5}), we obtained the KOG (Eukaryotic Orthologous Groups) [227] annotations. Using *Arabidopsis thaliana* and *Oryza sativ*a as reference (cut-off E-value e^{-5}) of dicot and monocot model species, respectively, we obtained the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [228, 229]. In addition, the Transcription Factor

(TF) databases of *A. thaliana* and *O. sativa* (downloaded from PlantTFDB v3.0 [230]), were used to conduct a BLASTX search (cut-off E-value e^{-5}).

5.2.4.3 Analysis of the coding and non coding transcripts

To evaluate the *in silico* expression level of the assembled transcripts we used RSEM [231]. This software calculates the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values for each assembled transcript normalizing the counts of the PE reads and the total number of mapped reads in the sample [232]. Among the assembled protein coding transcripts, 10 were selected to compare their abundance estimated in silico (FPKM) with that measured by quantitative RT-PCR as described below, using the actin *Oita*Act gene [GenBank: AB630020] as the endogenous control.

Coding Potential Calculator (CPC) [233] and Portrait [234] software packages were used to perform the analysis of the potential non-coding transcripts. The CPC software extimates the coding potential of a transcript evaluating the extent and quality of the ORF and then performing a BLASTX search against the UniProt Reference Clusters. So, a CPC positive value indicates that the transcript probably encodes for a protein; vice versa, negative values predicts a potential non-coding transcript. The Portrait software predicts putative proteins by a support vector machine and no homology information is required. To extract potential non-coding transcripts from the assembled transcriptome we applied the arbitrary threshold values ≥0.8 for the CPC coding potential score and ≤95% for the Portrait non-coding probability. We selected ten unannotated transcripts for the experimental validation. Specific primer pairs (Table 8) were designed and used to amplify the cDNA obtained from the total RNA of inflorescence of *O. italica.* The specific primer pairs were used to amplify 30 ng of first strand cDNA using the LongAmp Taq PCR Kit (New England Biolabs). After the cloning of the amplification product into the pGEM-T Easy vector (Promega), they were sequenced using the plasmid primers T7 and SP6 and were run on a 310 Genetic Analyzer (Applied Biosystems). To exclude artifacts, we aligned the obtained nucleotide sequences with those resulting from the *in silico* analysis of the transcriptome of *O. italica.* Quantitative RT-PCR experiments were performed in order to verify the expression pattern of the selected putative long non-coding transcripts as described below.

5.2.5 *In silico* search for miRNA target sites and cleavage analysis

The putative miRNA target analysis was conducted using the psRNATarget online tool [235]. When we decided to carry out this search using the cleaned small RNA reads of the inflorescence of *O.italica* against the newly assembled transcriptome of *P. aphrodite,* the parameters of the search were set with the default values (maximum expectation 3.0).When the search was conducted using the small RNA library of *O. italica* against the transcripts expressed in the inflorescence of the same species, the search was conducted using more stringent parameters (maximum expectation 0.0).

We verified the presence of the cleavage product induced by the cut of the miRNA on the transcript using a modified 5'-RACE experiment. This method was applied using the RLM-RACE GeneRace kit (Invitrogen) and consists in the binding of a the 5' adaptor to the 5'-terminus of the RNA extracted from inflorescence tissue of *O. italica* (500 ng) without any enzymatic treatment to remove the 5' cap [236]. The RNA was reverse transcribed and the cDNA was amplified using a transcript-specific reverse primer designed downstream of the predicted putative miRNA cleavage site and a GeneRace 5' Primer. A second PCR reaction was conducted on 1 µl of the first reaction using the nested adaptor forward primer and the nested specific reverse primers. The amplification products were cloned and sequenced as described above.

5.2.6 Real-Time PCR

Total RNA was reverse transcribed using the Advantage RT-PCR kit (Clontech) and an oligo dT primer. The Real Time experiments were conducted on 30 ng of the first strand cDNA from each tissue in technical and biological triplicates and using the SYBR Green PCR Master Mix (Life Technologies). To calculate the PCR efficiency (E) and the threshold cycle (CT) we used the Real Time PCR Miner online tool [237]. The mean relative expression ratio (Rn) and standard deviation of the target transcripts in the examinated tissues was calculated following formula $Rn = (1+E\ target)^{-CT\ target}/(1+E\ reference)^{-CT\ reference}$ and using 5,8 S or Actin as reference. ANOVA and Tukey HSD post hoc test were used to detect differences in the expression levels of the analized RNAs in the various tissues. Real Time PCR product of several samples was cloned and sequenced to exclude the presence of amplification artifacts.

5.2.6.1 Poly(T) Adaptor Real-Time PCR

We used the Poly(T) Adaptor RT-PCR method [238] in order to amplify the miRNAs. An amount of 350 ng of RNA from each tissue was used to conduct a reaction of poly-T adaptor ligation to the 3'-terminus of total RNA that was subsequently reverse transcribed with oligo dT primers. The Real Time PCR amplification was performed using a forward primer specific for each selected miRNA and a poly-T adaptor reverse primer. The procedures to calculate the mean relative expression ratio (Rn), as well as the differences in relative expression levels of the microRNAs between the different tissues are described above.

5.2.6.2 Stem and Loop Real-Time PCR

Stem-loop real time PCR experiments were conducted to evaluate the expression pattern of the microRNA miR319 in the examinated tissues [239]. In bref, we used a microRNA and a reference stem-loop primers separately (Table 9) to reverse transcribe 150 ng of RNA from each tissue. Then the real-time PCR was performed in technical triplicates and biological duplicates using 5 ng of first strand cDNA and the specific microRNA or reference primers and the stem-loop universal primer (Table 9). The procedures to calculate the mean relative expression ratio (Rn), as well as the differences in relative expression levels of the microRNA between the different tissues are described above.

5.2.7 Isolation of the class B *DEF-like* transcripts

First strand cDNA of inflorescence was amplified with the MADS-box degenerate primer MADS_F (see methods, Table 6) and a poly-T primer using the LongAmp Taq PCR Kit (New England Biolabs). The amplification products were cloned into the pGEM-T Easy vector (Promega) and about 50 clones were sequenced using the plasmid primers T7 and SP6. To verify the sequences obtained we conducted a BLAST search that revealed they correspond to four different *DEF*-like cDNAs.

5.2.8 Isolation of the *TCP* genes expressed in the inflorescence of *O. italica*

To isolate the transcripts of the TCP genes present in the inflorescence of *O.italica*, we used the sequences of the TCP DNA-binding domain of *Arabidopsis* and *Oryza* (Pfam PF0363) as query to perform a standalone TBLASTN (e-value 1 e^{-003}) against the inflorescence

transcriptome of *O. italica* [191]. The transcripts of *O. italica* with significant hits from the TBLASTN search and those previously annotated as *TCP* genes were selected to perform the analyses.

5.2.9 Isolation of the class II CYC/TB1-like genomic sequences

Based on three CYC/TB1-like nucleotide sequences of *P. equestris,* we designed degenerated primers to amplify the region spanning from the TCP domain to the R domain on the genomic DNA of *O. italica* extracted from leaf tissue [240]. The amplification product was cloned into the pGEM-T Easy vector (Promega) and the positive clones were sequenced as described above.

5.2.10 Phylogeny and analysis of conserved motifs

To conduct the phylogenetic analisys and the identification of the conserved motifs, the virtual translation of the selected transcripts was performed and the alignment was constructed using MUSCLE [241]. The Neighbor-Joining (NJ) trees was constructed using MEGA 6.06 [242] with 1000 bootstrap replicates. To identify shared conserved motifs in the TCP proteins of *O. italica*, *A. thaliana* and *O. sativa* we used the online tool MEME [243]. The parameters were set to any number of repetitions, the optimum width from 4 to 70 and the maximum number of motifs to 20.

Name	Sequence (5'-3')	Locus	Putative miRNA target
MADS_F	AAGATAGAGAATCCDACDAACD	MADS-box	
OitaDEF1F	CCTTCGCAGGGAGATAAGGCAAAGGA	*OitaDEF1*	
OitaDEF2F	CCTTCGGAAGGAGATAAGGCAGAGGA	*OitaDEF2*	
OitaDEF3F	CCTGAGGAGGGAGATAAGGCAGAGAA	*OitaDEF3*	
OitaDEF4F	TCTGAGGAGGGATGTAAGACAGAGGA	*OitaDEF4*	
OitaDEF1R1	TCATGCATAAGGGCCCTGTATACTTC	*OitaDEF1*	
OitaDEF2R1	TCATGCACTAGGGCCATGCACATTTC	*OitaDEF2*	
OitaDEF3R1	TCACGGAGTAAGCTCTTGTGGGTTTC	*OitaDEF3*	
OitaDEF4R1	TCACGCAGCAAATTATGGTGTGTCTC	*OitaDEF4*	
OitaDEF1R2	GTAAGTGTCTGTTTGCGTGGCGATCA	*OitaDEF1*	
OitaDEF2R2	GTAAGTGTCAGTTTGGGTAGCGATCA	*OitaDEF2*	
OitaDEF3R2	GTATGTATCAGTCTGGGTGCTAATGC	*OitaDEF3*	
OitaDEF4R2	ATAGGTGTCTGTCTGCGTACTGATTA	*OitaDEF4*	
OitaActF	TCGCGACCTCACCAATGTAC	*OitaAct*	
OitaActR	CCGCTGTAGTTGTGAATGAATAGC	*OitaAct*	
IN_3340	TCTCGGACCAGGCTTCATTCC	miR166	Leucine-zipper transcription factor
IN_36629	TCGCTTGGTGCAGGTCGGGA	miR168	AGO1
IN_33620	TTCCACAGCTTTCTTGAACTG	miR396	Growth regulating factor 4
IN_16410	TCGATAAACCTCTGCATCCGG	miR162	Dicer1-like
IN_32138	AAGCTCAGGAGGGATAGCGCC	miR390	TAS3
IN_30974	CAGCCAAGGATGACTTGCCGA	miR169	NF-YA
IN_33680	TGCCTGGCTCCCTGTATGCCA	miR160	Auxine response factor
IN_26041	TTTTGCTCAAGACCGCGCAAC	miR5179	*DEF*-like genes
IN_20892	ATATGAGCTCAAATCTAAGCTTG	Unknown miRNA	Leucine-rich repeat protein kinase
IN_27201	AAACTCTCTGAAATCACCCGAGAGG	Unknown miRNA	Transmembrane kinase
5.8S	ACGTCCGCCTGGGCGTCAAGC	Ribosomal 5.8S	

Table 6 Sequence of the primers used to isolate the *DEF*-like cDNAs and to analyze their tissue expression. Also, the sequence of the primers used to analyse the expression of the selected conserved and novel miRNAs of *O. italica* and their putative targets (best score).

Encoded gene	GenBank id	Unigene name	Primer (5'-3')	FPKM	FPKMn	Rn
OitaDEF4	AB857729	comp900_c0_seq1	TCTGAGGAGGGATGTAAGACAGAGGA	181.29	64.29	243.76
			ATAGGTGTCTGTCTGCGTACTGATTA			
OrcPI2	AB537504	comp1173_c0_seq1	GAGAGTACGCACCGCCACCG	134.3	47.62	239.04
			GCTGGATGGGCTGCACACGA			
OitaDEF3	AB857728	comp7668_c0_seq1	CCTGAGGAGGGAGATAAGGCAGAGAA	112.74	39.98	58.50
			GTATGTATCAGTCTGGGTGCTAATGC			
OrcPI	AB094985	comp1989_c0_seq1	CCCAGAATATGCGGACCAGATGCC	108.63	38.52	126.00
			TGGGCTGGAAAGGCTGCACG			
OitaDEF1	AB857726	comp3831_c0_seq1	CCTTCGCAGGGAGATAAGGCAAAGGA	56.84	20.16	82.81
			GTAAGTGTCTGTTTGCGTGGCGATCA			
OitaAG	JX205496	comp7958_c0_seq1	TCTGCAACAAATGCGCAGTAT	40.55	14.38	23.23
			AAGCTTGTGATTTGCTGTCGAA			
OitaSTK	JX205497	comp3859_c0_seq1	CGGAGCTACACGATGAAAGTATGT	37.75	13.39	36.35
			CCGCGCCCTCTCGTTTT			
OitaAP2	KF152921	comp8045_c0_seq1	TGTGTACCCCGGATTATTTCCT	26.78	9.50	9.60
			TTTCTGGGGCCAAGTGGTCATGGT			
OitaDEF2	AB857727	comp22604_c0_seq1	CCTTCGGAAGGAGATAAGGCAGAGGA	5.2	1.84	44.06
			GTAAGTGTCAGTTTGGGTAGCGATCA			
OitaAct	AB630020	comp44267_c0_seq1	TCGCGACCTCACCAATGTAC	2.82	1.00	1.00
			CCGCTGTAGTTGTGAATGAATAGC			

Table 7 Protein coding unigenes selected for the expression analysis validation. The table show the sequences of the primer pairs used in the Real Time PCR experiments, the FPKM counts for each assembled unigene and their respective normalized value (FPKMn) relative to the actin counts. Rn indicates the relative expression value obtained in the Real Time PCR experiments.

Unigene name	Length	Primer (5'-3')	Amplicon length	FPKM	CPC	Portrait
comp48038_c0_seq1	300	ACACCTTAATACAACCCTAAACCCT	224	2.67	-1.62	96.26
		TAACACCGGGGCAATGTCTT				
comp1308_c0_seq1	1246	ATCTGCAACGGGGGCATAAA	917	435.18	-1.03	32.33
		TGTTTCGCGGTCAGATCCAA				
comp0_c0_seq1	597	AAGCCTGCTGCCTTCGTTAT	386	20357.49	-0,31	4.87
		CAACACAGACTGGCTGGCTA				
comp3328_c0_seq1	214	CGTTCTGGTGGAGTTTGTCC	173	87.04	-1.13	95.64
		AATTGGCATGCATCAAGAAA				
comp1231_c0_seq1	772	AACGAATCCTGACCGCAGTT	308	61.91	-1.03	96.08
		ACTCATTTGCGGTCCTCCTG				
comp3311_c0_seq1	894	CCTCGGCCTAAAGAGGTAGC	360	52.42	-1.10	96.22
		ACAGTTGACCATCGCTCTCC				
comp6669_c0_seq1	217	ACACAGCAGCAAGTTGGTCTT	126	51.02	-1.32	95.00
		TGACCCCCAACACACAACAG				
comp4129_c0_seq1	611	CAGACATGGCAGAACGAAGA	202	46.77	-1.19	96.38
		AGCCGGAAGATAAGCTGACA				
comp15481_c0_seq1	2888	GAAGAAGCAATGAGCCCCCT	924	9.90	-1.33	84.89
		CAACCTACCAGTTCCGGTCC				
comp134696_c0_seq1	203	GGCGTTATCCTGATTGAGCTTTTC	203	0.64	-0.92	96.89
		CAGCTCAGGAGGGATAGAAGGGGG				

Table 8 Putative long non-coding unigenes selected for the expression analysis. Nucleotide sequence of the primer pairs used in the amplification experiments and amplificon length are shown. The CPC and Portrait columns indicate the coding potential score and the percentage of the non-coding probability, respectively.

Name	Target	Forward	Reverse	AP
TB1_orchid_TCP	CYC/TB1-like genes	AGRAARGAYMKRCAYARHAAGAT	YTTCTTCTCCAARGTYCTYTCYCT	slRT-PCR
comp8378_c0_seq1 c	comp_8378	CCAGCTATCTCAGGACGGTT	ATAGCGGCTAGCAGGTTGAG	RT-PCR
comp8964_c0_seq1	comp_8964	CAGTCTCCGGGAGGTACG	ACAAAGAGGCGAGGAGGTT	RT-PCR
comp21881_c0_seq1	comp_21881	GGATTCTTGCAGCCCTTAAC	GAGTCACTCGTGCTCATCGT	RT-PCR
comp16641_c0_seq1	comp_16641	CATATGGGACAGGGAAGAGG	GAGTCGACCCATCAGAATCA	RT-PCR
comp24776_c0_seq1	comp_24776	GGCTCGAGCTAGGACTTTCA	GTGCTGTTGATGGTGATGCT	RT-PCR
comp13386_c0_seq1	comp_13386	CCCTTGCAGTTTATGTCGAG	AAGGTTCGAATCAGCAATCC	RT-PCR
comp12442_c0_seq1	comp_12442	CCGGAACTATACCTGCCATC	GAAGTGGATGATCGGAACCT	RT-PCR
comp16313_c0_seq1	comp_16313	TTTCCATCATGCAAGACCAT	AGGTGAATTGGACTGAAGGG	RT-PCR
comp5062_c0_seq1	comp_5062	GTCAGCTCCAGGGATTGACT	AGCAGTGCCAAAGAAGAAGG	RT-PCR
comp1326_c0_seq1	comp_1326	TGGTCCAGAACCAGTTTGTC	GCTGTTAGGTGCATCTGGTG	RT-PCR
comp21123_c0_seq1	comp_21123	CCCACAAGCTTCTTCCAGTT	AACAGCATGGCCGTGTAATA	RT-PCR
TB1_Oita_TCP	OitaTB1	CAAGTTCTTCGATCTCCAGGAT	GACGAGCTCTTTAATGGCTGATT	RT-PCR
REAL_5.8S_ITA	5.8S	GGATATCTTGGCTCTCGCAT	GATGGTTCACGGGATTCTG	RT-PCR
Stem-Loop_miR319	miR319		GTCGTATCCAGTGCAGGGTCCG AGGTATTCGCACTGGATACGAC AGGGAG	slRT-PCR
miR319F	miR319	GCGGCGGTTGGACTGAAGGGAG		slRT-PCR
Stem-Loop_5.8S	5.8S		GTCGTATCCAGTGCAGGGTCCG AGGTATTCGCACTGGATACGAC GATTCA	slRT-PCR
REAL_5.8S_ITA_F	5.8S	GGATATCTTGGCTCTCGCAT		slRT-PCR
Stem-Loop_Univ_Rev	Universal		GTGCAGGGTCCGAGGT	slRT-PCR
1326_Rout	comp_1326		GAAGCCAACCACATCGCCGGCG GCG	m5'RACE
1326_Rinn	comp_1326		GCTGTTAGGTGCATCTGGTG	m5'RACE
5062_Rout	comp_5062		CCTGCAGACGGCCATTGAAACC GCC	m5'RACE
5062_Rinn	comp_5062		AGCAGTGCCAAAGAAGAAGG	m5'RACE
16313_Rout	comp_16313		CCGGCTGCATGTATGCCTCATTA CAG	m5'RACE
16313_Rinn	comp_16313		CTGGTTCTCCAGCGAGCTTCCCG CTG	m5'RACE

Table 9 The nucleotide sequence of the primers used. In the last column, the application (AP) in which they are used: PCR (PCR); Real Time PCR (RT-PCR); Stem-Loop Real Time PCR (slRT-PCR); modified 5' RACE (m5'RACE).

References

[1] Aceto S, Gaudio L (2011) The MADS and the Beauty: Genes Involved in the Development of Orchid Flowers. Current Genomics 12, 342-356
[2] Yu H, Goh CJ, (2001) Molecular genetics of reproductive biology in orchids. Plant Physiol, 127, 1390-1393.
[3] Tsai WC, Chen HH (2006) The orchid MADS-box genes controlling floral morphogenesis. Scientific World Journal, 6, 1933-1944.
[4] Rudall PJ, Bateman RM (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, 77, 403-441.
[5] Mondragon-Palomino M, Theissen G (2009) Why are orchid flowers so diverse? Reduction of evolutionary constraints by paralogues of class B floral homeotic genes. Ann Bot, 104, 583-594.
[6] Cameron KM (2007) Molecular phylogenetics of Orchidaceae: the first decade of DNA sequencing. Orchid Biology Reviews and Perspectives, 9, 163-200.
[7] Cameron KM, Chase MW, Whitten WM, Kores PJ, Jarrell DC, Albert VA, Yukawa T, Hills HG, Goldman DH (1999) A phylogenetic analysis of the Orchidaceae: evidence from rbcL nucleotide sequences. Am J Bot , 86, 208-224.
[8] Gorniak M, Paun O, Chase MW (2010) Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, Xdh: Congruence with organellar and nuclear ribosomal DNA results. Mol Phylogenet Evol, 56, 784-795.
[9] Cozzolino S, Widmer A (2005) Orchid diversity: an evolutionary consequence of deception? Trends Ecol Evol, 20, 487-494.
[10] Tremblay RL, Ackerman JD, Zimmerman JK, Calvo RN (2005) Variation in sexual reproduction in orchids and its evolutionary consequences: a spasmodic journey to diversification. Biol J Linn Soc, 84, 1-54.
[11] Lang D, Weiche B, Timmerhaus G, Richardt S, Riaño-Pachón DM, Corrêa LGG, Reski R, Mueller-Roeber B, Rensing SA (2010) Genome-Wide Phylogenetic Comparative Analysis of Plant Transcriptional Regulation: A Timeline of Loss, Gain, Expansion, and Correlation with Complexity. Genome Biol Evol, 2, 488–503.
[12] De Bodt S, Maere S, Van de Peer Y (2005) Genome duplication and the origin of angiosperms. Trends Ecol Evol, 20, 591-597.
[13] Yang C, Liu H, Li G, Liu M, Yun Y, Wang C, Ma Z, Xu JR (2015) The MADS-box transcription factor FgMcm1 regulates cell identity and fungal development in Fusarium graminearum Environ Microbiol, 8, 2762-76.
[14] Yanofsky MF, Ma H, Bowman J L, Drews GN, Feldmann KA, Meyerowitz EM (1990) The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature, 346, 35-39
[15] 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, 931-936.
[16] 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.

[17] Passmore S, Maine GT, Elble R, Christ C, Tye BK (1988) *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MAT alpha cells. J Mol Biol, 204, 593-606.
[18] 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, 931-936.
[19] Airoldi CA, Davies B (2012) Gene duplication and the evolution of plant MADS-box transcription factor. Journal of Genetics and genomics, 39, 157-165
[20] Gramzow L, Ritz MS, Theissen G (2010) On the origin of MADSdomain transcription factors. Trends Genet, 26, 149-153.
[21] Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana L, Martinez-Castilla L, Yanofsky MF (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc Natl Acad Sci U S A, 97, 5328-5333.
[22] Parenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, 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, 1538-1551.
[23] Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM (2011) The emerging importance of type I MADS box transcription factors for plant reproduction. Plant Cell, 2011, 23, 865-872.
[24] Riechmann JL, Meyerowitz EM (1997) MADS domain proteins in plant development. Biol Chem, 378, 1079-1101.
[25] Kaufmann K, Melzer R, Theissen G (2005) MIKC-type MADS domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene, 347, 183-198.
[26] Henschel K, Kofuji R, 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, 801-814.
[27] Zobell O, Faigl W, Saedler H, Munster T (2010) MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. Mol Biol Evol, 2010, 27, 1201-1211.
[28] Becker A, Theissen G (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol Phylogenet Evol, 29, 464-489.
[29] Kaufmann K, Melzer R, Theissen G (2005) MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene 347, 183-198.
[30] Coen ES, Meyerowit EM (1991) The war of the whorls–genetic interactions controlling flower development. Nature, *353*, 31-37.
[31] Krizek BA, Fletcher JC (2005) Molecular mechanisms of flower development: An armchair guide. Nat Rev Genet, 6, 688-698.
[32] Riechmann JL, Meyerowitz EM (1997) MADS domain proteins in plant development. Biol Chem, *378*, 1079-1101.
[33] Theissen G, Saedler H (2001) Plant biology. Floral quartets. Nature, 409, 469-471
[34] Smaczniak C, Immink RG, Angenent GC, Kaufmann K (2012) Developmental and evolutionary diversity of plant MADSdomain factors: insights from recent studies. Development 139, 3081-3098.
[35] Yang Y, Fanning L, Jack T (2003) The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. Plant J, 33, 47-59.

[36] Liu ZC, Mara C (2010) Regulatory mechanisms for floral homeotic gene expression. Semin Cell Dev Biol, 21, 80-86.
[37] Ambrose BA, Lerner DR, Ciceri P, Padilla CM, Yanofsky MF, Schmidt RJ (2000) Molecular and genetic analyses of the *silky1* gene reveal conservation in floral organ specification between eudicots and monocots. Mol Cell, 2000, 5, 569-579.
[38] Whipple CJ, Ciceri P, Padilla CM, Ambrose BA, Bandong SL, Schmidt RJ (2004) Conservation of B-class floral homeotic gene function between maize and *Arabidopsis.* Development, 131, 6083-6091.
[39] Whipple CJ, Zanis MJ, Kellogg EA, Schmidt RJ (2007) Conservation of B class gene expression in the second whorl of a basal grass and outgroups links the origin of lodicules and petals. Proc Natl Acad Sci U S A, 104, 1081-1086.
[40] Kim S, Koh J, Yoo MJ, Kong H, Hu Y, Ma H, Soltis PS, Soltis DE (2005) Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. Plant J, 43, 724-744.
[41] Ferrario S, Immink RG, Angenent GC (2004) Conservation and diversity in flower land. Curr Opin Plant Biol, 2004, 7, 84-91.
[42] Kanno A, Nakada M, Akita Y, Hirai M (2007) Class B gene expression and the modified ABC model in non grass monocots. Scientific World Journal, 7, 268–279.
[43] 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, 241–246.
[44] 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, 9–15.
[45] 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, 40, 3523–3535.
[46] De Paolo S, Gaudio L, Aceto S (2015) Analysis of the *TCP* genes expressed in the inflorescence of the orchid *Orchis italica*. Sci Rep, 5, 16265.
[47] Kanno A, Saeki H, Kameya T, Saedler H, Theissen G (2003) Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). Plant Mol Biol, *52*, 831-841.
[48] Nakamura T, Fukuda T, Nakano M, Hasebe M, Kameya T, Kanno A (2005) The modified ABC model explains the development of the petaloid perianth of *Agapanthus praecox ssp. Orientalis* (Agapanthaceae) flowers. Plant Mol Biol, 58, 435-445.
[49] Kramer EM, Dorit RL, Irish VF (1998) Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. Genetics, 149, 765-783.
[50] Zahn LM, Leebens-Mack J, DePamphilis CW, Ma H, Theissen G (2005) To B or Not to B a flower: the role of *DEFICIENS* and *GLOBOSA* orthologs in the evolution of the angiosperms. J Hered, *96*, 225-240.
[51] Mondragon-Palomino M, Theissen G (2008) MADS about the evolution of orchid flowers. Trends Plant Sci, 13, 51-59.
[52] Mondragon-Palomino M, Theissen G (2011) Conserved differential expression of paralogous *DEFICIENS*- and *GLOBOSA*-like MADS-box genes in the flowers of Orchidaceae: refining the 'orchid code'. Plant J, 66, 1008-1019.

[53] Cantone C, Gaudio L, Aceto S (2011) The *GLO*-like locus in orchids: duplication and purifying selection at synonymous sites within *Orchidinae* (Orchidaceae). Gene, 481, 48-55.
[54] Kuriya S, Hattori M, Nagano Y, Itino T (2015). Altitudinal flower size variation correlates with local pollinator size in a bumblebee-pollinated herb, *Prunella vulgaris L*. (Lamiaceae). J Evol Biol, (10):1761-9.
[55] Shimizu A, *et al.* (2014) Fine-tuned bee-flower coevolutionary state hidden within multiple pollination interactions. Sci Rep, 4:3988.
[56]Fenster CB, Armbruster WS, Dudash MR (2009) Specialization of flowers: is floral orientation an overlooked first step? New Phytol, 183, 502–506.
[57] Ushimaru A, Dohzono I, Takami Y, Hyodo F (2009) Flower orientation enhances pollen transfer in bilaterally symmetrical flowers. Oecologia, 160, 667–674.
[58] Hileman LC (2014) Trends in flower symmetry evolution revealed through phylogenetic and developmental genetic advances. Philos Trans R Soc Lond B Biol Sci 369, 20130348
[59] Luo D, Carpenter R, Copsey L, Vincent C, Clark J, Coen E (1999) Control of organ asymmetry in flowers of *Antirrhinum.* Cell, 99, 367–376.
[60] Cubas P, Lauter N, Doebley J, Coen E (1999). The TCP domain: A motif found in proteins regulating plant growth and development. Plant J, 18, 215–222.
[61] Almeida J, Rocheta M, Galego L (1997) Genetic control of flower shape in *Antirrhinum majus*. Development, 124, 1387–1392.
[62] Corley SB, Carpenter R, Copsey L, Coen E (2005) Floral asymmetry involves an interplay between TCP and MYB transcription factors in *Antirrhinum.* Proc Natl Acad Sci USA, 102, 5068–5073.
[63] Doebley J, Stec A, Hubbard L. (1997) The evolution of apical dominance in maize. Nature, 386, 485–488.
[64] Luo D, Carpenter R, Vincent C, Copsey L, Coen E (1996) Origin of floral asymmetry in *Antirrhinum*. Nature, 383, 794–799.
[65] Kosugi S, Ohashi Y (1997) PCF1 and PCF2 specifically bind to cis elements in the rice proliferating cell nuclear antigen gene. The Plant Cell, 9, 1607–1619.
[66] Kosugi S, Ohashi Y (2002) DNA binding and dimerization specificity and potential targets for the TCP protein family. The Plant Journal, 30, 337–348.
[67] Cubas P (2002) Role of TCP genes in the evolution of morphological characters in angiosperms. In QCB Cronk, RM Bateman, JA Hawkins, eds, Developmental Genetics and Plant Evolution, Special Volume Series 65. The Systematics Association, London, pp 247–266.
[68] Navaud O, Dabos P, Carnus E, Tremousaygue D, Herve C (2007) TCP transcription factors predate the emergence of land plants. Journal of Molecular Evolution, 65, 23–33.
[69] Martın-Trillo M, Cubas P (2010) TCP genes: a family snapshot ten years later. Trends in Plant Science, 15, 31–39.
[70] Masuda HP, Cabral LM, De Veylder L, Tanurdzic M, de Almeida Engler J, Geelen D, Inzé D, Martienssen RA, Ferreira PC, Hemerly AS (2008) ABAP1 is a novel plant Armadillo BTB protein involved in DNA replication and transcription. EMBO J, 27(20), 2746-56.
[71] Costa MM, Fox S, Hanna AI, Baxter C, Coen E (2005) Evolution of regulatory interactions controlling floral asymmetry. Development 132(22), 5093-101.
[72] Aggarwala P, Das Gupta M, Joseph AP, Chatterjee N, Srinivasan N and Nath U (2010) Identification of Specific DNA Binding Residues in the TCP Family of Transcription Factors in *Arabidopsis*. The Plant Cell, 22 (4), 1174-1189.

[73] Li C, Potuschak T, Colo ´n-Carmona A, Gutierrez RA, Doerner P (2005). *Arabidopsis* TCP20 links regulation of growth and cell division control pathways. Proc Natl Acad Sci USA, 102, 12978–12983.
[74] Nath U, Crawford BC, Carpenter R, Coen E (2003) Genetic control of surface curvature. Science, 299, 1404–1407.
[75] Koyama T, Furutani M, Tasaka M, Ohme-Takagi M (2007) TCP transcription factors control the morphology of shoot lateral organs via negative regulation of the expression of boundary-specific genes in *Arabidopsis*. The Plant Cell, 19, 473–484.
[76] Ori N, Cohen AR, Etzioni A, et al. (2007) Regulation of *LANCEOLATE* by miR319 is required for compound-leaf development in tomato. Nature Genetics, 39, 787–791.
[77] Howarth DG, Donoghue MJ (2006) Phylogenetic analysis of the ''ECE'' (CYC/TB1) clade reveals duplications predating the emergence of the core eudicots. Proc Natl Acad Sci USA, 103, 9101–9106.
[78] Busch A and Zachgo S (2009) Flower symmetry evolution: towards understanding the abominable mystery of angiosperm radiation. BioEssays, 31, 1181–1190.
[79] Preston JC and Hileman LC (2009) Developmental genetics of floral symmetry evolution. Trends in Plant Science, 14 (3), 147-154.
[80] Hileman LC, Cubas P (2009) An expanded evolutionary role for flower symmetry genes. J Biol, 8 (10), 90.
[81] Martin C, Paz-Ares J (1997) MYB transcription factors in plants. Trends genet. Cell, 13 (2), 67–73.
[82] Almerida J, Galego L (2005) Flower symmetry and shape in *Antirrhinum Int*. J Dev Biol, 49, 527-537.
[83] Raimundo J,Sobral R, Bailey P, Azevedo H, Galego L, Almeida J, Coen E, Costa MMR (2013) A subcellular tug of war involving three MYB-like proteins underlies a molecular antagonism in *Antirrhinum* flower asymmetry. The Plant Journal, 75, 527–538.
[84] Wang J, Wang Y, Luo D (2010) *Lj*CYC genes constitute floral dorsoventral asymmetry in *Lotus japonicus.* J Integr Plant Biol, 52, 959–970.
[85] Mondragón-Palomino M, Trontin C (2011) High time for a roll call: gene duplication and phylogenetic relationships of TCP-like genes in monocots. Ann Bot, 107(9), 1533-44.
[86] Preston JC, Hileman LC (2012) Parallel evolution of TCP and B-class genes in commelinaceae flower bilateral symmetry. Evodevo, 3, 6.
[87] Preston JC, Martinez CC, Hileman LC (2011) Gradual disintegration of the floral symmetry gene network is implicated in the evolution of a wind-pollination syndrome. Proc Natl Acad Sci U S A, 108, 2343-2348.
[88] Zhou XR, Wang YZ, Smith JF, Chen R (2008) Altered expression patterns of TCP and MYB genes relating to the floral developmental transition from initial zygomorphy to actinomorphy in *Bournea* (Gesneriaceae). New Phytol, 178, 532-543.
[89] Song CF, Lin QB, Liang RH, Wang YZ (2009) Expressions of ECE-CYC2 clade genes relating to abortion of both dorsal and ventral stamens in *Opithandra* (Gesneriaceae). BMC Evol Biol, 9, 244.
[90] Broholm SK, Tähtiharju S, Laitinen RA, Albert VA, Teeri TH et al. (2008) A TCP domain transcription factor controls flower type specification along the radial axis of the *Gerbera* (Asteraceae) inflorescence. Proc Natl Acad Sci U S A, 105, 9117-9122.
[91] Tähtiharju S, Rijpkema AS, Vetterli A, Albert VA, Teeri TH et al. (2012) Evolution and diversification of the CYC/TB1 gene family in Asteraceae - a comparative study in *Gerbera* (Mutisieae) and sunflower (Heliantheae). Mol Biol Evol, 29, 1155-1166.

[92] Howarth DG, Martins T, Chimney E, Donoghue MJ (2011) Diversification of *CYCLOIDEA* expression in the evolution of bilateral flower symmetry in *Caprifoliaceae* and *Lonicera* (Dipsacales). Ann Bot Lond, 107, 1521-1532.
[93] Citerne HL, Pennington RT, Cronk QC (2006) An apparent reversal in floral symmetry in the legume *Cadia* is a homeotic transformation. Proc Natl Acad Sci U S A, 103, 12017-12020.
[94] Xu S, Luo Y, Cai Z, Cao X, Hu X et al. (2013) Functional diversity of *CYCLOIDEA*-like TCP genes in the control of zygomorphic flower development in *Lotus japonicus*. J Integr Plant Biol, 55, 221-231.
[95] Busch A, Zachgo S (2007) Control of corolla monosymmetry in the *Brassicaceae Iberis amara*. Proc Natl Acad Sci U S A, 104, 16714-16719.
[96] Busch A, Horn S, Mühlhausen A, Mummenhoff K, Zachgo S (2012) Corolla monosymmetry: evolution of a morphological novelty in the Brassicaceae family. Mol Biol, Evol, 29, 1241-1254.
[97] Zhang W, Kramer EM, Davis CC (2010) Floral symmetry genes and the origin and maintenance of zygomorphy in a plant-pollinator mutualism. Proc Natl Acad Sci U S A, 107, 6388-6393.
[98] Wang Z et al. (2008) Genetic control of floral zygomorphy in pea *(Pisum sativum L.)*. Proc Natl Acad Sci USA, 105, 10414–10419.
[99] Feng XZ, ZhaoZ, Tian ZX et al. (2006) Control of petal shape and floral zygomorphy in *Lotus japonicus.* Proc Natl Acad Sci USA, 103, 4970-4975
[100] Yuan Z, Gao S, Xue DW, Luo D, Li LT et al. (2009) *RETARDED PALEA1* controls palea development and floral zygomorphy in rice. Plant Physiol, 149, 235-244.
[101] Bartlett ME, Specht CD (2011) Changes in expression pattern of the *teosinte branched1*-like genes in the Zingiberales provide a mechanism for evolutionary shifts in symmetry across the order. Am J Bot, 98, 227-24.
[102] Hoshino Y, Igarashi T, Ohshima M, Shinoda K, Murata N, Kanno A, Nakano M (2014) Characterization of *CYCLOIDEA*-like genes in controlling floral zygomorphy in the monocotyledon *Alstroemeria*. Scientia Horticulturae, 169, 6–13.
[103] Axtell MJ (2013) Classification and Comparison of Small RNAs from Plants. Annu Rev Plant Biol, 64, 137-59.
[104] Mallory AC, Vaucheret H (2006) Functions of microRNAs and related small RNAs in plants. Nature Genetics, 38, S31 - S36.
[105] Lee RC, Feinbaum LR, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell, 75, 843-854.
[106]Lagos-Quintana M, Rauhut R, Lendeckel W, Tusch T (2001) Identification of novel genes coding for small expressed RNAs. Science, 294, 853-858.
[107] Llave C, Kasschau KD, Rector MA, Carrington JC (2002) Endogenous and silencing-associated small RNAs in plants. Plant Cell, 14, 1605-1619.
[108] Liu Q, Chen YQ (2009) Insights into the mechanism of plant developmental: Interactions of miRNA pathway with phytohormone response. Biochem Biophys Res Commun, 384, 1-5.
[109] Voinnet O (2008) Post-transciptional RNA silencing in plant-microbe interactions: a touch of robustness and versatility. Curr Opin Plant Biol, 11, 464-470.
[110] Sunkar R, Chinnusamy V, Zhu J, Zhu JK (2007) Small RNAs as a big players in plant abiotic stress responses and nutrient deprivation. Trends Plant Sci, 12, 301-309.

[111] Jacobsen SE, Running MP, Meyerowitz EM (1999) Disruption of an *RNA helicase/RNAse III* gene in *Arabidopsis* causes unregulated cell division in floral meristems. Development, 126, 5231–5243.
[112] Lu C, Fedoroff N (2000) A mutation in the *Arabidopsis HYL1* gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin and cytokinin. Plant Cell, 12, 2351–2366.
[113] Telfer A, Poethig RS (1998) *HASTY*: a gene that regulates the timing of shoot maturation in *Arabidopsis thaliana.* Development 125, 1889–1898.
[114] Bohmert K et al. AGO1 defines a novel locus of *Arabidopsis* controlling leaf development. EMBO J, 17, 170–180.
[115] Chen X, Liu J, Cheng Y, Jia D (2002) *HEN1* functions pleiotropically in *Arabidopsis* development and acts in C function in the flower. Development, 129, 1085–1094.
[116] Han MH, Goud S, Song L, Fedoroff N (2004) The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. Proc Natl Acad Sci USA, 101, 1093-1098.
[117] Grigg SP, Canales C, Hay A, Tsiantis M (2005) *SERRATE* coordinates shoot meristem function and axial leaf pattering in *Arabidopsis.* Nature, 437, 1022-1026.
[118] Kurihara Y, Takashi Y, Watanabe Y (2006) The interaction between DLC1 and HYL1 is important for efficient and precise processing of pri-mRNA in plant micro-RNA biogenesis. RNA, 12, 206-212.
[119] Park MY, Wu G, Gonzalez-Suiser A, Vaucheret H, Poethig RS (2005) Nuclear processing and export of microRNAs in *Arabidopsis.* Proc Natl Acad Sci USA, 102, 3691-3696.
[120] Jones-Rhoades MW, Bartel D (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell, 14, 787-799.
[121] Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatory roles in plants. Ann Rev Plant Biol, 57, 19-53.
[122] Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, et al. (2008) Widespread translational inhibition by plant miRNAs and siRNAs. Science, 320, 1185–90.
[123] Dugas DV, Bartel B (2008) Sucrose induction of *Arabidopsis* miR398 represses two Cu/Zn superoxide dismutases. Plant Mol Biol, 67, 403–17.
[124] Cuperus JT, Fahlgren N, Carrington JC (2011) Evolution and functional diversification of MIRNAgenes. Plant Cell, 23, 431–42.
[125] Debernardi JM, Rodriguez RE, Mecchia MA, Palatnik JF (2012) Functional specialization of the plant miR396 regulatory network through distinct microRNA-target interactions. PLoS Genet, 8, e1002419.
[126] Axtel MJ (2008) Evolution of microRNAs and their targets: Are all microRNAs biologically relevant? Biochimica et Biophysica Acta, 1779, 725–734
[127] Fahlgren N, Jogdeo S, Kasschau KD, Sullivan CM, Chapman EJ, et al. 2010. MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. Plant Cell, 22, 1074–89.
[128] Ma Z, Coruh C, Axtell MJ. 2010. *Arabidopsis lyrata* small RNAs: transient MIRNA and small interfering RNA loci within the *Arabidopsis* genus. Plant Cell, 22, 1090–103.
[129] Qi Y, He X, Wang XJ, Kohany O, Jurka J, Hannon GJ. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature, 443, 1008–12.
[130] Chellappan P, Xia J, Zhou X, Gao S, Zhang X, et al. (2010) siRNAs from miRNA sites mediate DNA methylation of target genes. Nucleic Acids Res, 38, 6883–94.

[131] Vazquez F, Blevins T, Ailhas J, Boller T, Meins F Jr (2008) Evolution of ArabidopsisMIR genes generates novel microRNA classes. Nucleic Acids Res, 36, 6429–38.
[132] Wu L, Zhou H, Zhang Q, Zhang J, Ni F, et al. (2010) DNA methylation mediated by a microRNA pathway. Mol Cell, 38, 465–75.
[133] Hsu HF, Yang CH (2002) An orchid (*Oncidium Gower Ramsey*) AP3-like MADS gene regulates floral formation and initiation. Plant Cell Physiol, 43, 1198–1209.
[134] Tsai WC, Kuoh CS, Chuang MH, Chen WH, Chen HH (2004) Four DEF-like MADS box genes displayed distinct floral morphogenetic roles in *Phalaenopsis* orchid. Plant Cell Physiol, 45, 831–844.
[135] Skipper M, Johansen LB, Pedersen KB, Frederiksen S, Johansen BB (2006) Cloning and transcription analysis of an *AGAMOUS*- and *SEEDSTICK* ortholog in the orchid *Dendrobium thyrsiflorum* (Reichb. f.). Gene, 366, 266–274.
[136]Song IJ, Nakamura T, Fukuda T, Yokoyama J, Ito T, et al. (2006) Spatiotemporal expression of duplicate *AGAMOUS* orthologues during floral development in *Phalaenopsis*. Dev Genes Evol, 216, 301–313.
[137] Xu Y, Teo LL, Zhou J, Kumar PP, Yu H (2006) Floral organ identity genes in the orchid *Dendrobium crumenatum*. Plant J, 46, 54–68.
[138] Tsai WC, Pan ZJ, Hsiao YY, Jeng MF, Wu TF, et al. (2008) Interactions of B- class complex proteins involved in tepal development in *Phalaenopsis* orchid. Plant Cell Physiol, 49, 814–824.
[139] Mondragon-Palomino M, Hiese L, Harter A, Koch MA, Theissen G (2009) Positive selection and ancient duplications in the evolution of class B floral homeotic genes of orchids and grasses. BMC Evol Biol, 9, 81.
[140]Chang YY, Kao NH, Li JY, Hsu WH, Liang YL, et al. (2010) Characterization of the possible roles for B class MADS box genes in regulation of perianth formation in orchid. Plant Physiol, 152, 837–853.
[141] Wang SY, Lee PF, Lee YI, Hsiao YY, Chen YY, et al. (2011) Duplicated C-class MADS-box genes reveal distinct roles in gynostemium development in *Cymbidium ensifolium* (Orchidaceae). Plant Cell Physiol, 52, 563–577.
[142] Su CL, Chen WC, Lee AY, Chen CY, Chang YC, et al. (2013) A modified ABCDE model of flowering in orchids based on gene expression profiling studies of the moth orchid *Phalaenopsis aphrodite*. PLoS One, 8, e80462.
[143] Acri-Nunes-Miranda R, Mondragon-Palomino M (2014) Expression of paralogous *SEP*-, *FUL*-, *AG*- and *STK*-like MADS-box genes in wild-type and peloric *Phalaenopsis* flowers. Front Plant Sci, 5, 76.
[144] An FM, Hsiao SR, Chan MT (2011) Sequencing-based approaches reveal low ambient temperature-responsive and tissue-specific microRNAs in *Phalaenopsis* orchid. PLoS One, 6, e18937.
[145] Lin CS, Chen JJ, Huang YT, Hsu CT, Lu HC, et al. (2013) Catalog of *Erycina pusilla* miRNA and categorization of reproductive phase-related miRNAs and their target gene families. Plant Mol Biol, 82, 193–204.
[146]Chao YT, Su CL, Jean WH, Chen WC, Chang YC, et al. (2014) Identification and characterization of the microRNA transcriptome of a moth orchid *Phalaenopsis aphrodite*. Plant Mol Biol, 84, 529–548.
[147] Salemme M, Sica M, Iazzetti G, Gaudio L, Aceto S (2013) The *AP2*-like gene *OitaAP2* is alternatively spliced and differentially expressed in inflorescence and vegetative tissues of the orchid *Orchis italica*. PLoS One, 8, e77454.

[148] Schommer C, Debernardi JM, Bresso EG, Rodriguez RE, Palatnik JF (2014) Repression of cell proliferation by miR319-regulated TCP4. Mol Plant, 7, 1533-1544.
[149] Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D (2003) Control of leaf morphogenesis by microRNAs. Nature, 425, 257–263.
[150] Nag A, King S, Jack T (2009) miR319a targeting of *TCP4* is critical for petal growth and development in *Arabidopsis*. Proc Natl Acad Sci USA, 106, 22534-22539.
[151] Peragine A, Yoshikawa M, Wu G, Albrecht H L, Poethig R S (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabid*opsis. Genes Dev, 18, 2368–2379.
[152] Adenot X, Elmayan T, Lauressergues D, Boutet S, Bouche N, Gasciolli V, Vaucheret H (2006) DRB4-dependent TAS3 trans-acting siRNAs control leaf morphology through AGO7. Curr Biol, 16, 927–932.
[153] Fahlgren N, Montgomery TA, Howell MD, Allen E, Dvorak SK, Alexander AL, Carrington JC (2006) Regulation of AUXIN RESPONSE FACTOR3 by TAS3 tasiRNAs affects developmental timing and pattering in *Arabidopsis*. Curr Biol, 16, 939-944.
[154] Garcia D, Collier SA, Byrne ME, Martienssen RA (2006) Specification of leaf polarity in *Arabidopsis* via the trans-acting siRNA pathway. Curr Biol, 16, 933–938.
[155] Hunter C, Willmann MR, Wu G, Yoshikawa M, de la Luz Gutierrez-Nava M, Poethig RS (2006) Trans-acting siRNA-mediated repression of *ETTIN* and *ARF4* regulates heteroblasty in *Arabidopsis*. Development, 133, 2973–2981.
[156] Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, Mallory AC, Hilbert JL, Bartel DP, Crete P (2004) Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. Moll Cell, 16, 69-79.
[157] Allen E, Xie Z, Gustafson AM, Carrington JC (2005) MicroRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell, 121, 207–221.
[158] Yoshikawa M, Peragine A, Park MY, Poethig RS (2005) A pathway for the biogenesis of trans-acting siRNAs in *Arabidopsis*. Genes Dev, 19, 2164–2175.
[159] Axtell M J, Jan C, Rajagopalan R, Bartel DP (2006) A two-hit trigger for siRNA biogenesis in plants. Cell, 127, 565–577.
[160] Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, Chapman EJ, Fahlgren N, Allen E, Carrington JC (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNAs formations. Cell, 133, 128-141.
[161] Baumberger N, Baulcombe DC (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci USA, 102, 11928–11933.
[162] Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C, al. (2008) Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5' terminal nucleotide. Cell, 133, 116–127.
[163] Allen E, Howell MD (2010) miRNAs in the biogenesis of trans-acting siRNAs in higher plant. Semin Cell Dev Biol, 21, 798-804.
[164] Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell, 123, 1279–91.
[165] Ron M, Alandete Saez M, EshedWilliams L, Fletcher JC, McCormick S (2010) Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in *Arabidopsis.* Genes Dev, 24, 1010–21.

[166] Yang Y, Wen L, Zhu H (2015) Unveiling the hidden function of long non-coding RNA by identifying its major partner-protein. Cell Biosci, 5, 59.
[167] Chen LL, Carmichael GG (2010) Decoding the function of nuclear long non-coding RNAs. Curr Opin Cell Biol, 22, 357–364.
[168] Wang KC, Chang HY (2011) Molecular mechanisms of long noncoding RNAs. Mol Cell, 43, 904–914.
[169] Liu J, Wang H, Chua NH (2015) Long noncoding RNA transcriptome of plants. Plant Biotechnol J, 13, 319–328.
[170] Li X, Wu Z, Fu X, Han W (2014) LncRNAs: Insights into their function and mechanics in underlying disorders. Mutat Res Rev Mutat Res, 762, 1–21.
[171] Wierzbicki AT (2012) The role of long non-coding RNA in transcriptional gene silencing. Curr Opin Plant Biol, 15, 517–522.
[172] Zhang YC, Chen YQ (2013) Long noncoding RNAs: New regulators in plant development. Biochem Biophys Res Commun, 436, 111–114.
[173] Csorba T, Questa JI, Sun Q, Dean C (2014) Antisense coolair mediates the coordinated switching of chromatin states at *FLC* during vernalization. Proc Natl Acad Sci USA, 111, 16160–16165.
[174] Wang ZW, Wu Z, Raitskin O, Sun Q, Dean C (2014) Antisense-mediated *FLC* transcriptional repression requires the P-TEFb transcription elongation factor. Proc Natl Acad Sci USA, 111, 7468–7473.
[175] Kim ED, Sung S (2012) Long noncoding RNA: Unveiling hidden layer of gene regulatory networks. Trends Plant Sci, 17, 16–21.
[176] Aceto S, Sica M, De Paolo S, D'Argenio V, Cantiello P, et al. (2014) The Analysis of the Inflorescence miRNome of the Orchid *Orchis italica* Reveals a *DEF* Like MADS-Box Gene as a New miRNA Target. PLoS ONE, 9(5), e97839.
[177] Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, et al. (2007) High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS ONE, 2, e219.
[178] Zhu J, Li W, Yang W, Qi L, Han S (2013) Identification of microRNAs in *Caragana intermedia* by high-throughput sequencing and expression analysis of 12 microRNAs and their targets under salt stress. Plant Cell Rep, 32 (9), 1339, 1349.
[179] Zhang R, Marshall D, Bryan GJ, Hornyik C (2013) Identification and characterization of miRNA transcriptome in potato by high-throughput sequencing. PLoS One, 8, e57233.
[180] Khraiwesh B, Pugalenthi G, Fedoroff NV (2013) Identification and analysis of red sea mangrove (*Avicennia marina*) microRNAs by high-throughput sequencing and their association with stress responses. PLoS One, 8, e60774
[181] Morin RD, Aksay G, Dolgosheina E, Ebhardt HA, Magrini V, et al. (2008) Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. Genome Res, 18, 571–584.
[182] Luo Y, Guo Z, Li L (2013) Evolutionary conservation of microRNA regulatory programs in plant flower development. Dev Biol, 380, 133–144.
[183] Jouannet V, Moreno AB, Elmayan T, Vaucheret H, Crespi MD, et al. (2012) Cytoplasmic Arabidopsis *AGO7* accumulates in membrane-associated siRNA bodies and is required for ta-siRNA biogenesis. EMBO J, 31, 1704–1713.
[184] Liu PP, Montgomery TA, Fahlgren N, Kasschau KD, Nonogaki H, et al. (2007) Repression of *AUXIN RESPONSE FACTOR10* by microRNA160 is critical for seed germination and post-germination stages. Plant J, 52, 133–146.

[185] Mallory AC, Bartel DP, Bartel B (2005) MicroRNA-directed regulation of *Arabidopsis AUXIN RESPONSE FACTOR17* is essential for proper development and modulates expression of early auxin response genes. Plant Cell, 17, 1360–1375.
[186] Hirsch J, Lefort V, Vankersschaver M, Boualem A, Lucas A, et al. (2006) Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived transcripts. Plant Physiol, 140, 1192–1204.
[187] Vaucheret H (2009) *AGO1* homeostasis involves differential production of 21-nt and 22-nt miR168 species by MIR168a and MIR168b. PLoS One, 4, e6442.
[188] Li W, Cui X, Meng Z, Huang X, Xie Q, et al. (2012) Transcriptional regulation of *Arabidopsis* MIR168a and argonaute1 homeostasis in abscisic acid and abiotic stress responses. Plant Physiol, 158, 1279–1292.
[189] Miyashima S, Honda M, Hashimoto K, Tatematsu K, Hashimoto T, et al. (2013) A comprehensive expression analysis of the *Arabidopsis* MICRORNA165/6 gene family during embryogenesis reveals a conserved role in meristem specification and a non-cell-autonomous function. Plant Cell Physiol, 54, 375–384.
[190] Husbands AY, Chitwood DH, Plavskin Y, Timmermans MC (2009) Signals and prepatterns: new insights into organ polarity in plants. Genes Dev, 23, 1986– 1997.
[191] De Paolo S, Salvemini M, Gaudio L, Aceto S (2014) *De novo* transcriptome assembly from inflorescence of *Orchis italica*: analysis of coding and non-coding transcripts. PLoS One, 9, e102155.
[192] Li X, Luo J, Yan T, Xiang L, Jin F, et al. (2013) Deep sequencing-based analysis of the *Cymbidium ensifolium* floral transcriptome. PLoS One, 8, e85480.
[193] Zhang J, Wu K, Zeng S, Teixeira da Silva JA, Zhao X, et al. (2013) Transcriptome analysis of *Cymbidium sinense* and its application to the identification of genes associated with floral development. BMC Genomics, 14, 279.
[194] Chou ML, Shih MC, Chan MT, Liao SY, Hsu CT, *et al.* (2013) Global transcriptome analysis and identification of a *CONSTANS*-like gene family in the orchid *Erycina pusilla*. Planta, 237, 1425–1441.
[195] Su CL, Chao YT, Alex Chang YC, Chen WC, Chen CY, *et al.* (2011) *De novo* assembly of expressed transcripts and global analysis of the *Phalaenopsis aphrodite* transcriptome. Plant Cell Physiol, 52, 1501–1514.
[196] Chang YY, Chu YW, Chen CW, Leu WM, Hsu HF, *et al.* (2011) Characterization of *Oncidium 'Gower Ramsey'* transcriptomes using 454 GSFLX pyrosequencing and their application to the identification of genes associated with flowering time. Plant Cell Physiol, 52, 1532–1545.
[197] Sedeek KE, Qi W, Schauer MA, Gupta AK, Poveda L, *et al*. (2013) Transcriptome and proteome data reveal candidate genes for pollinator attraction in sexually deceptive orchids. PLoS One, 8, e64621.
[198] Leitch IJ, Kahandawala I, Suda J, Hanson L, Ingrouille MJ, *et al.* (2009) Genome size diversity in orchids: consequences and evolution. Ann Bot, 104, 469–481.
[199] Barkan A, Small I (2014) Pentatricopeptide Repeat Proteins in Plants. Annu Rev Plant Biol 65: 415–442.
[200] Hsu CC, Chung YL, Chen TC, Lee YL, Kuo YT, *et al*. (2011) An overview of the *Phalaenopsis* orchid genome through BAC end sequence analysis. BMC Plant Biol, 11, 3.
[201] Manning G, Plowman GD, Hunter T, Sudarsanam S (2002) Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci, 27, 514–520.

[202] Chapman, M. A. *et al.* Genetic analysis of floral symmetry in Van Gogh's sunflowers reveals independent recruitment of *CYCLOIDEA* genes in the *Asteraceae*. PLoS Genet, 8, e1002628.
[203] Jabbour F. *et al.* (2014) Specific duplication and dorsoventrally asymmetric expression patterns of Cycloidea-like genes in zygomorphic species of Ranunculaceae. PLoS One 9, e95727.
[204]Xu R, Sun P, Jia F, Lu L, Li Y, Zhang S, Huang J (2014) Genome wide analysis of TCP transcription factor gene family in *Malus domestica.* J Genet, 93, 733–746 (2014).
[205] Ma, J. Wang Q, Sun R, Xie F, Jones DC, Zhang B (2014) Genome-wide identification and expression analysis of TCP transcription factors in *Gossypium raimondii*. Sci Rep, 4, 6645.
[206] Parapunova, V. *et al.* (2014) Identification, cloning and characterization of the tomato TCP transcription factor family. BMC Plant Biol, 14, 157.
[207] Palatnik JF *et al.* (2007) Sequence and expression differences underlie functional specialization of arabidopsis microRNAs miR159 and miR319. Dev Cell, 13, 115–125.
[208] Schommer, C. *et al*. (2008) Control of jasmonate biosynthesis and senescence by miR319 targets. PLoS Biol, 6, e230.
[209] Zhou M, Luo H (2014) Role of microRNA319 in creeping bentgrass salinity and drought stress response. Plant Signal Behav, 9, e28700.
[210] Uberti-Manassero NG, Lucero LE, Viola IL, Vegetti AC, Gonzalez DH (2012) The class I protein AtTCP15 modulates plant development through a pathway that overlaps with the one affected by CIN-like TCP proteins. J Exp Bot 63, 809–823.
[211] Lucero LE *et al.* (2015) *TCP15* modulates cytokinin and auxin responses during gynoecium development in *Arabidopsis*. Plant J, 84(2), 267-282.
[212] Pruneda-Paz JL, Breton G, Para A, Kay SA (2009) A functional genomics approach reveals CHE as a component of the *Arabidopsis circadian* clock. Science, 323, 1481–1485.
[213] Aguilar-Martinez JA, Sinha N (2013) Analysis of the role of *Arabidopsis* class I TCP genes *AtTCP7, AtTCP8, AtTCP22*, and *AtTCP23* in leaf development. Front Plant Sci, 4, 406.
[214] Takeda T *et al.* (2003) The *OsTB1* gene negatively regulates lateral branching in rice. Plant J, 33, 513–520.
[215] Moxon S, Schwach F, Dalmay T, Maclean D, Studholme DJ, et al. (2008) A toolkit for analysing large-scale plant small RNA datasets. Bioinformatics, 24, 2252–2253.
[216] Stocks MB, Moxon S, Mapleson D, Woolfenden HC, Mohorianu I, *et al*. (2012) The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. Bioinformatics, 28, 2059–2061.
[217] Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory efficient alignment of short DNA sequences to the human genome. Genome Biol, 10, R25.
[218] Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res, 39, D152–157.
[219] Yang X, Li L (2011) miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics, 27, 2614–2615.
[220] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, *et al.* (2011) Full length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol, 29, 644–652.
[221] Chen TW, Gan RC, Wu TH, Huang PJ, Lee CY, *et al.* (2012) FastAnnotator an efficient transcript annotation web tool. BMC Genomics, 13 Suppl 7, S9.

[222] Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, *et al.* (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res, 40, W622–627.
[223] Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, *et al.* (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc, 8, 1494–1512.
[224] Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics, 26, 680–682.
[225] Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, *et al*. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 21, 3674–3676.
[226] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, *et al.* (1997) Gapped BLAST and PSIBLAST: a new generation of protein database search programs. Nucleic Acids Res, 25,3389–3402.
[227] Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, et al. (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol, 5, R7.
[228] Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res, 28, 27–30.
[229] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res, 32, D277–280.
[230] Jin J, Zhang H, Kong L, Gao G, Luo J (2014) PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res, 42, D1182–1187.
[231] Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNASeq data with or without a reference genome. BMC Bioinformatics, 12, 323.
[232] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods, 5, 621–628.
[233] Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, *et al.* (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res, 35, W345–349.
[234] Arrial RT, Togawa RC, Brigido Mde M (2009) Screening non-coding RNAs in transcriptomes from neglected species using PORTRAIT: case study of the pathogenic fungus Paracoccidioides brasiliensis. BMC Bioinformatics, 10, 239.
[235] Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res, 39, W155–159.
[236] Llave C, Xie Z, Kasschau KD, Carrington JC (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. Science, 297, 2053–2056.
[237] Zhao S, Fernald RD (2005) Comprehensive algorithm for quantitative real-time polymerase chain reaction. J Comput Biol, 12, 1047–1064.
[238] Shi R, Sun YH, Zhang XH, Chiang VL (2012) Poly(T) adaptor RT-PCR. Methods Mol Biol, 822, 53–66.
[239] Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP (2007) Protocol: a highly sensitive RT PCR method for detection and quantification of microRNAs. Plant Methods, 3, 12.
[240] Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small amounts of leaf tissue. Phytochem Bull, 19, 11–15.

[241] Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res, 32, 1792–1797.
[242] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol, 30, 2725–2729.
[243] Bailey TL *et al.* (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res, 37, W202–208.

Personal bibliography

Communications to congress

S. De Paolo, M. Sica, V. D'Argenio, P. Cantiello, F. Salvatore, L. Gaudio, and S. Aceto. A deep sequencing approach to uncover the inflorescence miRNome of the orchid *Orchis italica.* Conference AGI (Italian Genetics Association), Cortona 25 to 27 September 2013.

S. De Paolo. A deep sequencing of the inflorescence miRNome of the orchid Orchis italic reveals a DEF-like gene as new miRNA target , Young Researchers in life sciences (YRLS), Paris 26 to 28 May 2014

S. De Paolo, M. Salvemini, L. Gaudio, S. Aceto. The inflorescence transcriptome of *Orchis italica*, a wild Mediterranean orchid species. Italian Federation of Life Sciences (FISV), Pisa 24 to 27 September 2014.

S. De Paolo, L.Gaudio, and S. Aceto (2015*) TCP* genes genes of the Mediterranean Orchid *Orchis Italica.* AGI, Cortona 28 to 30 september 2015.

Seminars

Sofia De Paolo, University of Naples Federico II "The inflorescence transcriptome of Orchis Italica: analysis of coding and non coding transcripts" National Pingtung University of Science and Technology, Taiwan 27/10/2014.

Papers

Aceto S, Sica M, **De Paolo S,** D'Argenio V, Cantiello P, *et al.* (2014) The Analysis of the Inflorescence miRNome of the Orchid *Orchis italica* Reveals a DEF-Like MADS-Box Gene as a New miRNA Target. PLoS ONE, 9(5), e97839.

De Paolo S, Salvemini M, Gaudio L, Aceto S (2014) De Novo Transcriptome Assembly from Inflorescence of *Orchis italica*: Analysis of Coding and Non-Coding Transcripts. PLoS ONE, 9(7), e102155.

De Paolo S, Gaudio L, Aceto S (2015) Analysis of the TCP genes expressed in the inflorescence of the orchid *Orchis italica*. Sci Rep, 5, 16265.

Borrelli L, Aceto S, Agnisola C, **De Paolo S**, Dipineto L, Stilling RM, Dinan TG, Cryan JF, Menna LF and Fioretti A (2016) Probiotic modulation of the microbiota-gut-brain axis and behaviour in zebrafish. Sci Rep 6, 30046.