Mutations in orthologous *PETALOSA TOE*-type genes cause dominant double-flower phenotype in phylogenetically distant eudicots

Stefano Gattolin¹,²*, Marco Cirilli³, Stefania Chessa¹, Alessandra Stella¹, Daniele Bassi³, Laura Rossini³*

¹ CNR-IBBA, Via Bassini 15, 20133 Milano, Italy
² PTP Science Park, Via Einstein, Loc. Cascina Codazza, 26900 Lodi, Italy
³ Università degli Studi di Milano – DiSAA, Via Celoria 2, 20133 Milano, Italy

Corresponding authors:
Stefano Gattolin stefano.gattolin@ibba.cnr.it
Laura Rossini laura.rossini@unimi.it

Highlight
A conserved mechanism regulates the dominantly inherited double flower trait, involving natural allelic variation at microRNA172 binding site in the novel PETALOSA subclade of TARGET OF EAT (TOE)-type genes.
Abstract
The double-flower phenotype has been selected by humans for its attractiveness in various plant species and it is of great commercial value for the ornamental market. In this paper we investigated the genetic determinant of the dominant double-flower trait in carnation, petunia and Rosa rugosa, identifying mutant alleles of TARGET OF EAT (TOE)-type genes characterized by a disruption of the miR172 target sequence and of the C-terminal portion of the encoded protein. Despite the phylogenetic distance between these eudicots, which diverged in the early Cretaceous, the orthologous genes carrying such mutations all belong to a single TOE-type subgroup, herein referred to as PETALOSA (PET). Homology searches allowed us to identify PET sequences in various other species. To confirm the results on naturally occurring mutations, we used CrispR-Cas9 to induce lesions within the miR172 target site of Nicotiana tabacum PET genes, and this resulted in the development of supernumerary petaloid structures. This study describes pet alleles in economically important ornamentals and provides evidence about the possibility of identifying and engineering PET genes to obtain the desirable double-flower trait in different plants.

Keywords
Floral development, petal number, Dianthus, Rosa, petunia, tobacco
Introduction

The variety of flower shapes and colours has fascinated humans for millennia as witnessed by artistic depictions since ancient times. Among features of ornamental value, Double Flower (DF) is a desirable trait, increasing the number of petals and overall flower size (Crane and Lawrence, 1952). Investigations of the genetic control of this trait have uncovered different loci with recessive and dominant inheritance in several plant species (Vainstein 2013). Recently, candidate mutations for the dominant DF trait were described in orthologous genes of the euAP2 transcription factor lineage in peach (Prunus persica) and Rosa genus (Gattolin et al., 2018, Hibrand Saint-Oyant et al., 2018, François et al., 2018). euAP2 transcription factors are encoded by genes of the APETALA2/ETHYLENE RESPONSIVE ELEMENT-BINDING FACTOR (AP2/ERF) family and are divided into two groups, AP2- and TOE-type, both characterized by the presence of a target site for the binding of microRNA172 (miR172), which negatively regulates their activity (Jofuku et al., 1994; Riechmann et al., 2000; Kim et al., 2006). This class of transcription factors play a conserved role in the regulation of flower patterning and development transitions (Huijser and Schmid, 2011), with AP2 playing a major role in the ABCDE model proposed for Arabidopsis thaliana (Rijpkema et al., 2010). During flower development, euAP2 repression by miR172 is crucial for maintaining organ determinacy and defining the boundary between the outer sterile perianth and the reproductive inner flower parts (Zhao et al., 2007; Wollmann et al., 2010). Wild-type Rosaceae flowers typically bear five petals. In both peach and rose, mutated alleles in orthologous TOE-type genes were dominantly associated with the DF phenotype, suggesting they might act as upstream regulators of AGAMOUS (AG) and restrict its expression to the inner floral whorls (Dubois et al., 2010); such variants cause the expression of transcripts encoding truncated proteins carrying all key functional motifs but lacking the miR172 target site. In particular, a deletion in the 3' region of peach gene Prupe.6G242400 results in the expression of a truncated mRNA, and ultimately in a potentially functional TOE-type transcription factor not post-transcriptionally regulated by miR172 (Gattolin et al., 2018). In rose, an insertion in the 8th intron of Rchm3g0468481, encoding TOE-type factor XP_024182693, causes a premature stop that also leads to the transcription of a truncated mRNA lacking the miR172 binding site (Gattolin et al., 2018). The insertion in rose seems to consist in a Gypsy LTR transposon and this allele can be found in the Old Blush rose genome, suggesting that this variant could have originated in Rosa chinensis, a major contributor of the genetic makeup of modern roses (Martin et al., 2001, Bendahmane et al., 2013, François et al., 2018). However, previous research suggests that different mutations leading to the DF phenotype arose in the genus Rosa and were likely selected independently in R. chinensis, R. gallica and R. rugosa (Dubois et al., 2010).
Dominant flower doubleness is widespread amongst angiosperms, and it was selected as leading ornamental trait in commercially important plants such as carnation (D85 locus, Yagi et al., 2014a) and petunia (Do1 locus, Sink 1984, de Vlaming et al., 1984; Vanderkrol and Chua, 1993). However, the genes controlling this phenotype remain unknown in these species. In this paper the genetic determinant of the dominant DF trait has been investigated in carnation, petunia and R. rugosa, discovering mutant alleles of TOE-type genes - hereafter referred to as PETALOSA (PET) genes - characterized by a disruption of the miR172 target sequence. To further support a conserved role of allelic variability at PET genes, a genome editing approach was used to induce mutations within the miR172 target site of Nicotiana tabacum PET genes, obtaining tobacco lines characterized by additional petals and petaloid stamens.

Materials and methods

Plant Material

Dianthus varieties were kindly provided by ‘Floricoltura Billo’ (https://www.floricolturabillo.it/) and Hybrida srl. R. rugosa material was kindly provided by ‘Le Rose di Firenze’ (https://www.lerosedifirenze.com). Petunia plants were obtained from local garden centres and identified with their commercial names.

RT-PCR and 3’ RACE

Total RNA was extracted from floral buds using a Quick-RNA Miniprep Kit (Zymo) following the manufacturer’s protocol, with the modification of adding 2% PVP and 4% beta-mercaptoethanol (Sigma-aldrich) to the tissue lysis buffer just before use. 1 μg total RNA was treated with DNaseI (Invitrogen) and first strand cDNA obtained with Goscript Reverse Transcriptase (Promega), using either a standard oligo-dT primer or the B26 primer containing an adaptor sequence (Frohman et al., 1988). Reactions were diluted 1:10 and 1 μl used as template for RT-PCR analysis or 3’-RACE using GOTAq (Promega). RT-PCR analysis was carried out in a 25 μl reaction with GOTAq, using specific primers CA-9f/CA-1r (Dianthus pet allele), CA-9f/CA-4r (Dianthus wild type allele), PH-9F/PH-M-3R (petunia pet allele) or PH-9F/PH-10R (petunia wild type allele). RACE analysis was carried out in a 25 μl reaction using GOTAq from DF petunia cDNA using PH-9F and B25 or from DF Dianthus cDNA using CA-9f and B25.

Genome-walking in R. rugosa

2 μg of single flower and DF R. rugosa genomic DNA were digested with 2 μl TaqIa (NEB) in Buffer Cutsmart in a 20 μl final reaction (65°C, 2h30’). The digestion reaction was purified with PCR cleanup kit (Promega) and the concentration adjusted to 50 ng/μl. 20 μl of B25 and B25_TaqI_adapter primers (100 pmol/μl) were mixed, heated for 5 minutes at 75°C and let cool
down to room temperature to anneal into 5'-CG-3' overhang adapters. 200 ng of digested DNA and 100 pmol of B25 adapter were ligated using T4 Ligase (Invitrogen). The ligation was used as template for a two-step nested PCR. A 13 cycles pre-PCR cycle was carried out using RO-8F/B25 and this was used as template for nested PCR, using RO-8F2/B25. Amplified bands were gel purified and sequenced.

Genotyping
Genomic DNA was extracted from 200 μg of leaf tissue using DNeasy 96 Plant Kit (Qiagen); 10 ng of genomic DNA were used in PCR reactions using GoTaq and the appropriate primer combinations (Supplementary Table S1): CA-9f/CA-4r or CA-9f/CA-1r, (Dianthus wild type and pet alleles respectively); PH-9F/PH-10R or PH-9F/PH-M-3R (Petunia hybrida wild type and pet alleles respectively); RO-8f/RR-3UTRr for both R. rugosa wild type and pet alleles.

Full length genomic DNA sequencing
Full length genomic amplicons were obtained with GoTaq Long (Promega) using genomic DNA obtained from D. barbatus ‘Sweet William’, D. superbus ‘Primadonna’, D. caryophyllus ‘Widecombe fair’, P. hybrida ‘Viva Double Purple Vein’ (Florensis), R. rugosa ‘Hansa’ and primers CA-F1/CA-4R (Dianthus), PH-F1/PH-UTR-R (petunia wild type allele), PH-F1/PH-M-3R (petunia pet allele) and RO-1F/RR-UTR-R (R. rugosa). PCR bands were extracted from agarose gels using a Wizard SV Gel and PCR Clean-Up System (Promega) and paired-end sequenced on an Illumina MiSeq instrument (Illumina, San Diego, CA, USA), following the manufacturer’s instructions, using a transposome-based Nextera XT kit (Illumina) to generate the libraries. Fastq files were mapped with BWA-MEM against reference sequences on the Galaxy Platform (Afgan et al., 2018) and visualized using IGV (Robinson et al., 2011).

Molecular Phylogenetic analysis by Maximum Likelihood method
P. persica, R. chinensis, P. hybrida, P. axillaris, A. thaliana, Vitis vinifera peptide sequences used for phylogenetic analysis were obtained from previously published work (Morel et al., 2017; Gattolin et al., 2018), Nicotiana tomentosiformis, Camellia sinensis, Spinacia oleracea and Carica papaya sequences were obtained from NCBI (https://www.ncbi.nlm.nih.gov/), while D. caryophyllus sequences were obtained from the reference genome website (http://carnation.kazusa.or.jp/). For Dianthus Dca21030 the wild-type allele (Dca21030.2) was used (see results). Phylogenetic relationships were estimated in MEGAX (Kumar et al., 2018). Peptide sequences (Supplementary Fig. S1A) were aligned by MUSCLE with default settings. Evolutionary relationships among TOE-type members were inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The rate variation model allowed for some sites to be evolutionarily invariable and a discrete Gamma distribution was used to model
evolutionary rate differences among sites. The reliability of the phylogenetic tree was estimated by setting 200 bootstrap replicates.

**CrispR-Cas9 editing of tobacco**

For genome editing of tobacco plants, we used *Agrobacterium*-mediated T-DNA transformation with binary vector pHAtC (Kim et al., 2016), obtained from Addgene (https://www.addgene.org). To ensure the transcription of specific guide RNA, oligonucleotides TOB-CRISPR_for and TOB-CRISPR_rev were annealed and ligated into *AarI* digested pHAtC. Tobacco cv. TI 527 ‘Kentucky’ plants were transformed with a c58 Agrobacterium suspension following Sparkes et al., 2006 and transformants selected in media containing Hygromycin 30 mg/l. T₀ plants were grown under standard greenhouse conditions until flowering. Mutations in *NtBEN* miR172 binding sites were assessed by Sanger sequencing of PCR fragments obtained using either forward primer NtBEN_016482517_F (XP_016482517) or NtBEN_016499635_F (XP_016499635), and common reverse primer NtBEN_SEQR. T₁ seedlings were PCR screened for the presence of the transgene using primers LBfor/LBrev and RBfor/RBrev, specific for the T-DNA sequence. The presence of off-targets in transgene-free T₁ plants was assessed through an HRMA-based approach. Primers (Supplementary Table S1) were designed to amplify fragments flanking the PAM recognition sequence within miR172 target site of euAP2 target genes and HRM analyses were carried out in a Corbett Rotor-Gene 6000 series (Qiagen) using Type-it HRM PCR Kit (Qiagen). The reactions were carried out with the following program: 2 min at 94°C, 40 cycles of 30 s at 94°C, 30 s annealing at 58°C and 30 s at 72°C, followed by a melting step over a 70–95°C gradient with 0.1°C/s ramp rate. Data were analysed using Rotor-Gene software 1.7 and visualized using both a derivative and difference plot, according to software instructions.

**Results**

A mutation disrupting the miR172 target site of a *PET* gene is associated with DF phenotype in *Dianthus*

The *Dianthus* genus comprises species of horticultural interest (carnations and pinks), including both single- and double-flower varieties. Homology searches on the carnation (*Dianthus caryophyllus* L.) genome (Yagi et al., 2014b) allowed for the identification of a *TOE*-type gene orthologous to *Prupe.6G242400*, annotated as *Dca21030.1*. Notably, *CES0212*, an SSR markers tightly associated with the DF *D85* locus (Yagi 2014a), maps less than 5 Kb from the 5’ region of *Dca21030.1*, making this gene a prime candidate for the DF phenotype. Analysis of the reference genome sequence of DF cultivar ‘Francesco’ revealed that this gene consists of 10 exons. Sequence comparison with peach *Prupe.6G242400* suggested the presence of a 1 Kb insertion.
within the 10th exon (Supplementary Figure S2A) also affecting the miR172 target site. The sequence of the insertion shows similarity with a putative mobile element present in multiple copies within the D. caryophyllus genome (Supplementary Figure S2B), and we reasoned that this annotated gene could be the mutated PET allele (pet) leading to flower doubleness in ‘Francesco’. Allelic comparison of the 3’ region revealed the presence of only the wild-type allele in single flower accessions D. superbus ‘Primadonna’ and D. deltoides ‘Flashing lights’ (Figure 1A, B, Supplementary Figure S2C) and both wild-type and pet alleles in DF variety ‘Widecombe fair’ (Figure 1A). Supporting the role of these sequence variants in flower development, both alleles are expressed in “Widecombe fair” floral buds (Figure 1C). Combined evidence from cDNA and genomic resequencing confirmed the presence of a 1 Kb insertion, causing a CC to AG substitution at the 3’ end of the miRNA target site and introducing a stop codon 11 bp downstream (Supplementary Figure S2C, D), consistent with the annotated genome sequence. Therefore, Dca21030.1 could represent a de facto pet allele encoding a transcript escaping miR172 regulation and conferring the DF phenotype. Conversely, the identified wild-type allele (Dca21030.2) harbours a complete miR172 target site and further 83 bp of coding sequence (Supplementary Figure S3A, B). Co-segregation of the pet allele with the dominant DF phenotype was confirmed in 25 commercial varieties (Figure 1B, Supplementary Table S2A).

A mutation disrupting the miR172 target site of a PET gene is associated with DF phenotype in petunia

BLIND ENHANCER (BEN) and BROTHER OF BEN (BOB), two Petunia hybrida TOE-type transcription factors orthologous to Prupe.6G242400, have been finely characterized, suggesting they redundantly regulate the development of the second and third floral whorls (Morel et al., 2017). The sequence of genomic marker SSR7 associated with the DF phenotype in petunia (Liu et al., 2016) was searched against the genomes of petunia wild parents, P. axillaris and P. inflata (Bombarely et al., 2016): this analysis positioned this marker within the P. inflata scaffold Peinf101Scf00457, at 1,310 Kb from the BOB genomic sequence. The possible involvement of this gene in P. hybrida DF phenotype was investigated by 3’ RACE on bud cDNA from DF variety “Double Purple Vein”. Of the two expressed alleles obtained, one was nearly identical to the BOB transcript (KU096996) and P. axillaris Peaxi162Scf00472g00069.1, while the second showed an insertion positioned at the level of the 10th exon, 69 bp upstream of the miR172 target site (Figure 2A and Supplementary Figure S4A), as confirmed by targeted resequencing. The insertion - a probable LTR mobile element present in multiple copies in the genomes of P. axillaris (Supplementary Figure S4B, C) and P. inflata - results in a shorter transcript predicted to escape
miR172 post-transcriptional regulation, but still encoding euAP2 functional domains (Supplementary Figure S3C, D). Genotype-phenotype association was validated in different commercial petunia varieties (Figure 2B). Interestingly, in the four DF varieties tested, both the wild-type and the pet allele were found, in agreement with previous reports on the presence of both alleles in DF petunia varieties (Sink, 1984). Expression of both alleles was further confirmed in “Double Purple Vein” using specific primers (Figure 2C).

**Identification of a previously uncharacterized pet allele in R. rugosa**

The screening of different rose varieties with the previously developed *R. hybrida* pet marker (Gattolin et al., 2018) only detected the wild-type allele in the DF *R. rugosa* variety ‘Hansa’. Considering the complex breeding history of modern roses (Dubois et al., 2010), we hypothesized the existence of a different but functionally similar pet allele in *R. rugosa*. A genome-walking approach revealed a 500 bp deletion in the *R. rugosa* gene orthologous to *RcHm3g0468481*, spanning part of the last exon and the adjacent 3’ UTR (Figure 3A). The deletion includes the miR172 binding sequence, originating a pet allele similar to those previously identified in the other DF species, not found in single flower *R. rugosa* accessions (Figure 3B). Targeted resequencing of the entire gene regions of both ‘Hansa’ alleles using the wild-type *R. chinensis* allele as reference confirmed the presence of the deletion in this pet allele (Figure 3A, Supplementary Figure S3E, F, G). The presence of this pet allele was further confirmed in 4 commercial DF varieties (Supplementary Table S2B), supporting that in the *Rosa* genus the DF phenotype can be caused by at least two independent pet mutations.

**Phylogenetic analysis of TOE-type transcription factors**

To gain insight into the phylogenetic relationships among TOE-type genes harbouring pet mutations, we analysed protein sequences from different plant species (listed in Supplementary Figure S1A). As the Rosaceae family was represented by sequences from both rose and peach, sequences from a second Solanaceae species, diploid *N. tomentosiformis*, were included to complement those of petunia. Phylogenetic analysis suggests the existence of three subgroups within TOE-type genes and all genes associated to DF mutations belong to a single orthologous PET subgroup, which includes a gene/genomic duplication in *N. tomentosiformis* as well as petunia, where this is consistent with functional redundancy (Morel et al., 2017) (Figure 4). Hence, independent selection of distinct mutations in orthologous PET genes gave rise to the DF trait in peach, rose, carnation and petunia. The lack of a PET gene in Arabidopsis could be due to a recent loss following polyploidization and genome rearrangements in the Brassicaceae lineage.
(Blanc et al., 2003; Town et al., 2006; Ren et al., 2018) and a survey of TOE-type genes (Supplementary Figure S1B) confirmed the absence of PET genes in Brassica napus. Interestingly, a PET orthologue (XM_022035806) was found in C. papaya, which belongs to the order Brassicales, suggesting a PET was indeed originally present in the lineage. A survey of the reference allotetraploid N. tabacum genome (Murad et al., 2002) revealed the existence of at least 13 putative euAP2 proteins, including 3 putative PET sequences: two closely related homoeologous NtBENa and NtBENb (XP_016482517 and XP_016499635 derived from N. tomentosiformis and N. sylvestris, respectively), and a NtBOB (XP_016502850) derived from N. tomentosiformis (XP_018630941) (Supplementary Figure S1C, D). Noteworthy, in both N. tabacum and N. tomentosiformis BOB transcripts, a miR172 binding site is present in the 3’ UTR, while a stop codon positioned 14 nucleotides upstream of the site itself (Supplementary Figure S5A-C) originates a predicted protein lacking 54 amino acids at C-terminus, compared to petunia BOB (Supplementary Figure S5D).

Validation of the effect of lesions in the miR172 binding site in PET genes via CrispR-Cas9 mediated editing of tobacco plants

Our results demonstrate a strict association between a range of naturally occurring mutations in orthologous PET genes and the dominant DF trait in different plant species. We thus investigated whether artificially induced mutations in the miR172 target site of a PET gene are sufficient to induce the DF phenotype in a different plant species without naturally occurring DF variants. To this end, we used CrispR-Cas9 to specifically create mutations at the PET-miR172 target sequence in tobacco (Supplementary Figure S5E). Of seven T₀ tobacco lines carrying the CRISPR-Cas9 construct, three were characterized by a range of floral phenotypes including conversion of stamens into petaloid structures and double flowers (Supplementary Figure S6). Molecular analysis on leaf and petal tissues confirmed an array of mutations, ranging from 1 to 3 bp insertions and/or deletions in the miR172 binding site of target NtBENs and NtBOB, while the other euAP2 were not affected by editing (Supplementary Figure S7). Thus, preliminary T₀ analysis confirmed that plants displaying DF phenotype also carried lesions in PET miR172 binding regions. From selfing T₀ line 7, four T₁ plants lacking the CRISPR-Cas9 construct and carrying mutations in NtBENs were selected, while no plant with mutation in NtBOB was obtained (Supplementary Figure S7). Sequencing of genomic fragments spanning the miR172 target site of both NtBENs revealed the presence of differently edited alleles, associated in heterozygosity to various degree of flower doubleness (Figure 5). A single nucleotide insertion within the miR172 core recognition sequence of either NtBEN gene is sufficient to induce the development of petaloid
structures within the corolla (Figure 5 lines II and III). A 1bp deletion in one of the two NtBEN genes resulted in flowers indistinguishable from the wild-type (line I), while a 1bp insertion in both NtBEN genes resulted in the strongest DF phenotype (line IV).

Discussion

The understanding of the mechanisms governing flower development has been a major goal in plant biology, although knowledge gathered on model species does not always apply to other plants. First insights into the molecular basis of dominant mutations conferring the DF trait were recently obtained when structural variants disrupting miR172 target sites within a subclass of TOE-type genes were pinpointed as prime candidates in Rosaceae (Gattolin et al., 2018, François et al., 2018), although functional evidence was still lacking. Also, whether the mechanism behind the dominant DF trait is conserved in phylogenetically distant genera of angiosperms remained unknown and of great interest for both plant scientists and breeders, particularly considering the economic value of the ornamental market.

In the present work, we show that similar variations may be responsible for the dominant DF character in different plants, broadening the spectrum of species bearing this type of mutation beyond the Rosaceae family. Through allele mining approaches, pet mutations in orthologous genes encoding PET euAP2 transcription factors were also identified in carnation, petunia and R. rugosa. Insertions in carnation and petunia and a deletion in R. rugosa were shown to result in the transcription of mRNAs encoding truncated proteins lacking the C-terminus portion and disrupting the miR172 target site within. By editing the genome of a single flower species (tobacco), we clearly demonstrated the crucial role of PET genes in the DF phenotype: single nucleotide insertions within the miR172 sequence core induced the development of supernumerary petaloid stamens in whorl 3. In contrast to the strong reduction in petal development caused by double knockout of petunia PET genes BEN and BOB (Morel et al., 2017), pet alleles are dominantly associated with supernumerary petals indicating they are gain-of-function mutations. In agreement with the well-characterized mechanism of miR172-regulated euAP2 expression, all these pet alleles are likely to escape post-transcriptional regulation. This might lead to the DF phenotype as a consequence of altered AG repression, similar to AG regulation by AP2 in Arabidopsis (Krogan et al., 2012, Morel et al., 2017).

As pet mutations also impact the C-terminal portion of the encoded transcription factors, the precise mechanism by which naturally occurring or artificially induced mutations modify the functionality of PET genes and give rise to DF remains to be more finely validated. Nevertheless, different lines of evidence support a crucial role for the miR172 binding site rather than
modification of protein C-ter functionality. First, all identified pet alleles still encode all the highly conserved functional euAP2 domains (Wang et al., 2016), while the C-ter portions of PET proteins show little conservation beyond the motif corresponding to the miR172 binding site (Supplementary Figure S8A). Indeed, BOB genes in wild-type single flower N. tabacum and N. tomentosiformis encode shorter PET proteins (Supplementary Figure S8B) and harbour the miR172 binding sequence in the 3’ UTR. Second, in cases where the miR172 target site is affected but not completely absent, the severity of the phenotype depends on the position of the mutation within the site: this can be interpreted as the result of a different stability of the resulting mRNA:miRNA duplexes and a consequent moderate loss of PET regulation, lowering but not abolishing AG expression. In Dianthus pet we observed two consecutive mismatches corresponding to the miRNA 5’ end, which has been reported to fully abolish target site efficacy (Liu et al., 2014). Furthermore, mRNA regions flanking the miR172 binding site have also been shown to affect silencing efficiency (Li et al., 2014; Wong et al., 2018) and the different downstream sequence may well contribute to the phenotype severity in Dianthus. Third, the tobacco T1 plant with the strongest phenotype (line IV) harbours a 1 bp heterozygous insertion in both NtBEN genes, suggesting a dose-dependent effect. This is reminiscent of previous observations on the AP2-like Q gene in polyploid wheat: a single nucleotide mutation in the miR172 target site was shown to alter the balance between miR172 and Q gene expression and correlate with dose-dependent phenotypes of varying intensity (Debernardi et al., 2017).

In the present work we identified the PET subgroup of TOE-type euAP2 transcription factors comprising sequences from Eudicots species belonging to the Pentapetalae clade, a group of plants characterized by pentameric flowers with whorled phyllotaxis. Pentapetalae are divided in two groups (Zeng et al., 2017) and PET sequences were found both in Rosids (peach, rose, papaya) and Vitales (V. vinifera), belonging to Group I, as well as in Asterids (petunia, Nicotiana, camellia) and Caryophyllales (Dianthus, Spinacia) belonging to Group II. These two groups are estimated to have diverged around 120 Mya in agreement with the radiation of the Pentapetalae lineage, that recent works set in the early Cretaceous (Kumar et al., 2017; Zeng et al., 2017). Given single flower is the ancestral state of peach, rose, carnation and petunia, and independent mutations were found in orthologous PET genes, the dominant DF trait likely represents an example of convergence under strong human selection in phylogenetically distant Eudicots. Arabidopsis TOE-type genes (TOE1 and TOE2) were reported to redundantly act in various plant developmental processes, such as seedling innate immunity (Zou et al., 2018), epidermal leaf identity (Wu et al., 2009) and flowering repression (Aukerman and Sakai, 2003; Jung et al., 2007, Zhang et al., 2015, Zhai et al., 2015). While our phylogenetic reconstruction places these proteins
close to one of the non-PET TOE-type subgroups (Figure 4), this positioning is not strongly supported (bootstrap value 47): in conjunction with the long branches, this suggests an uncertain phylogenetic placement. These considerations and the lack of a PET orthologue indicate that Arabidopsis may not be representative of the functions of TOE-type genes in other species. An outstanding question regards the role of genes from the other two TOE-type subgroups. Further studies in a range of plants are therefore important to gain a better understanding of the biological roles of different TOE-type genes.

In summary, we used information on the causal mutation of the dominant DF phenotype in peach and rose to investigate the causal mutations in other economically important ornamental plants: carnation, petunia and R. rugosa. In all cases we identified strong candidate causal mutations in orthologous TOE-type genes, resembling those previously described. These findings are confirmed by phenotypic alterations of gene-edited tobacco plants, providing a proof-of-concept of the possibility to manipulate flower morphology in different plants through PET engineering.

Supplementary data

Figure S1. TOE-type peptides from different species.

Figure S2. Analysis of the Dianthus pet allele.

Figure S3. Sequence alignment of PET alleles from different species.

Figure S4: Analysis of the petunia pet allele.

Figure S5. Analysis of the tobacco PET alleles

Figure S6. T₀ tobacco lines transformed with the CrispR-Cas9 construct.

Figure S7. High-Resolution Melting Analysis for detection of potential off-target edited alleles of tobacco TOE-type genes.

Figure S8. Degree of conservation of PET amino acid sequences and effect of CrispR-Cas9-induced mutations on protein sequences.
Table S1. List of primers used in this study.

Table S2. PCR analysis of *Dianthus* and *R. rugosa* varieties.

**Acknowledgements**

The authors thank Federico Billo and Flavio Sapia for providing *Dianthus* plant material and precious varietal knowledge, and Stefano Magi, Adriana Balzi, Alipio Cortopassi for *R. rugosa* samples. We are grateful to Stefania Prati, Remo Chiozzotto, Andrea Giupponi (DISAA, University of Milan, Italy), Paola Cremonesi (IBBA CNR, Italy) and Valeria Rizzi (PTP Science Park) for technical assistance.

**Author contributions**

S.G. and L.R. conceived the project. S.G. designed and performed most of the experiments and analysed the data. M.C. performed HRM analysis and S.C. conducted amplicon resequencing. A.S. and D.B. provided financial support and critically revised the manuscript. S.G., L.R. and M.C. wrote the article.
References


Figure Legends

Figure 1. Molecular analysis of sequence variants of the Dianthus PETALOSA gene. (A) TOP: Dca21030.1 gene model (pet Dianthus allele) where CDS exons (solid boxes), the position of primers CA-1f, CA-9f, CA-1r, CA-4r (arrow heads) and of the miR172 target region (asterisk) are indicated. The coding sequences for AP2-R1/AP2-R2 conserved domains and the EAR motif are marked in red and green, respectively. BOTTOM: Amplicon sequencing coverage plot of the two alleles amplified with primers CA-1f and CA-4r from DF D. caryophyllus ‘Widecombe Fair’ and of the single amplicon obtained from both D. barbatus and D. superbus. (B) PCR analysis of four single-flower (top) and 4 double-flower (bottom) Dianthus accessions. Expected band sizes for pet (CA-9f/CA-1r, 411bp) or wild type allele (CA-9f/CA-4r, 441bp) are indicated. (C) Amplification with primers CA-9f/CA-1r and CA-9f/CA-4r from flower bud cDNA (pet: 315bp, wild-type: 345bp) or genomic DNA of D. caryophyllus ‘Widecombe Fair’.

Figure 2. Molecular analysis of sequence variants of the petunia PETALOSA gene. (A) TOP: Gene model for petunia BOB pet allele deduced from the genomic sequence of Peaxi162Scf00472g00069 and the available sequence information for the insertion; CDS exons (solid boxes), the position of primers PH-1f, PH-9f, PH-10r, PH-M3r, PH-UTRr (arrow heads) and of the miR172 binding site (asterisk) are indicated. The coding sequences for AP2-R1/AP2-R2 conserved domains and the EAR motif are marked in red and green, respectively. BOTTOM: Amplicon sequencing coverage plot of the two alleles amplified from DF petunia ‘Double Purple Vein’ with primers PH-1F and PH-M3r (pet) or PH-UTRr (wild type). (B) PCR analysis of four single-flower (top) and four double-flower (bottom) petunia varieties. Expected band sizes for pet (PH-9F/PH-M-3R, 724bp) or wild-type allele (PH-9F/PH-10R, 606bp) are indicated. (C) PCR products obtained with the primers PH-9F/PH-M-3R and PH-9F/PH-10R using flower bud cDNA.
(pet: 333bp, wild-type: 215bp) or genomic DNA of petunia variety ‘Double Purple Vein’. (Photograph of Double Pirouette by 阿橋 HQ, own photostream on https://www.flickr.com, CC BY-SA 4.0 - https://creativecommons.org/licenses/by-sa/4.0)

Figure 3. Molecular analysis of sequence variants of the *R. rugosa* PETALOSA gene. (A) TOP: Gene model for *R. chinensis* wild-type PET (XP_024186592) where CDS exons (solid boxes), the position of primers RO-1f, RO-8f, RR-UTRr (arrow heads) and of the miR172 binding site (asterisk) are indicated. The coding sequences for AP2-R1/AP2-R2 conserved domains and the EAR motifs are marked in red and green, respectively. BOTTOM: Amplicon sequencing coverage plot of the two alleles amplified with primers RO-1f and RR-UTRr from DF *R. rugosa* ‘Hansa’. (B) PCR analysis of two single-flower accessions and DF ‘Hansa’ using primers RO-8f e RR-3UTRr, the expected band sizes for *pet* (853bp) or wild-type allele (1338bp) are indicated.

Figure 4. Evolutionary relationships among TOE-type proteins in different plant species. The tree was obtained using peptide sequences from: *P. persica* (Peach), *R. chinensis* (Rosa), *A. thaliana* (Arabidopsis), *P. hybrida/axillaris* (Petunia), *D. caryophyllus* (Dianthus), *S. oleracea* (Spinacia), *N. tomentosiformis* (Ntomentosiformis), *V. vinifera* (Vitis), *C. papaya* (Papaya) and *C. sinensis* (Camellia). Blue dots indicate characterized PET genes. A square indicates *N. tomentosiformis* BEN. *R. chinensis* sequence XP_024186592 corresponds to RAG04722 described previously (Gattolin *et al.*, 2018). An AP2-type sequence from petunia (PhROB1) was included as outgroup. Only bootstrap values above 50 are shown.

Figure 5. CrispR-Cas9 mediated editing of tobacco plants (A) Heterozygous mutations at *NtBEN* miR172 target sites in four T1 genome edited tobacco lines: I, II, III and IV. (B) Flower phenotype of a wild-type and the four T1 plants. (C) Dissection of a wild-type flower showing sepals, fused petals, stamens and pistil. (D) Dissection of a T1 line IV flower showing sepals, fused petals, petaloid stamens and pistil. Scale bar: 1 cm. XP_016482517 is *NtBENa* and XP_016499635 is *NtBENb* (Supplementary Figure S1D).
Figure 1

A  
Dca21030.1 - pet

B  
D. barbatus  
Sweet William  Primadonna  India Star
D. superbus  Olivia
D. caryophyllus
D. caryophyllus
Pink Kisses  Starburst  Widecombe Fair  Francesco

C  
Bud cDNA  Genomic DNA

441 bp
413 bp
345 bp
315 bp
Figure 2

A

Peaxi162Scf00472g00069 (BOB) - pet

'Double Purple Vein' - pet

'Double Purple Vein' - wild type

B

Picobella F1  Pacta Parade  Giant Blue  Corona Amethyst

Double Pirouette  Double Purple Vein  Double Liliac  Sweet Sunshine Red

C

Bud cDNA  Genomic DNA

724 bp  606 bp

333 bp  215 bp
**Figure 3**

**A**  
*R. chinensis* genomic for XP_024186592 wild-type

**B**  
*R. rugosa typica* 'Nyveldt's White'  
'Hansa'
Figure 4
Figure 5

A

<table>
<thead>
<tr>
<th>T1</th>
<th>Gene</th>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>XP_016499635</td>
<td>TGCAGCATC ATCAGGATTCG</td>
<td>Ref. sequence</td>
</tr>
<tr>
<td>II</td>
<td>XP_016499635</td>
<td>TGCAGCATC ATCAGGATTCG</td>
<td>1nt deletion</td>
</tr>
<tr>
<td>III</td>
<td>XP_016482517</td>
<td>TGCAGCATC ATCAGGATTCG</td>
<td>1nt insertion</td>
</tr>
<tr>
<td>IV</td>
<td>XP_016499635</td>
<td>TGCAGCATTCAATCAGGATTCG</td>
<td>1nt insertion</td>
</tr>
<tr>
<td></td>
<td>XP_016482517</td>
<td>TGCAGCATTCAATCAGGATTCG</td>
<td>1nt insertion</td>
</tr>
</tbody>
</table>

B

wt  I   II  III  IV

C

wt

D

IV