the plant journal

SEB Society for Experimental Biology

The Plant Journal (2011) **68**, 11–27 doi: 10.1111/j.1365-313X.2011.04661.x

Revealing impaired pathways in the *an11* mutant by high-throughput characterization of *Petunia axillaris* and *Petunia inflata* transcriptomes

Sara Zenoni[1,†], Nunzio D'Agostino[2,†,‡], Giovanni B. Tornielli[1,†], Francesca Quattrocchio[3], Maria L. Chiusano[2], Ronald Koes[3], Jan Zethof[4], Flavia Guzzo[1], Massimo Delledonne[1], Luigi Frusciante[2], Tom Gerats[4] and Mario Pezzotti[1,*]

[1]*Department of Biotechnology, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy,*
[2]*Department of Soil, Plant, Environmental, and Animal Production Sciences, University of Naples ''Federico II,'' Via Università 100, 80055 Portici, Italy,*
[3]*Department of Molecular Cell Biology, VU-University, de Boelelaan 1075, 1081 HV Amsterdam, The Netherlands, and*
[4]*Institute for Water and Wetland Research, Radboud University, Heyendaalseweg 135, 6525AC Nijmegen, The Netherlands*

Received 25 March 2011; revised 26 May 2011; accepted 27 May 2011; published online 14 July 2011.
*For correspondence (fax +39 045 8027929; e-mail mario.pezzotti@univr.it).
†These authors contributed equally to this work.
‡Present address: CRA-ORT, Research Centre for Vegetable Crops, via Cavalleggeri 25, 84098 Pontecagnano (Salerno), Italy.

SUMMARY

Petunia is an excellent model system, especially for genetic, physiological and molecular studies. Thus far, however, genome-wide expression analysis has been applied rarely because of the lack of sequence information. We applied next-generation sequencing to generate, through *de novo* read assembly, a large catalogue of transcripts for *Petunia axillaris* and *Petunia inflata.* On the basis of both transcriptomes, comprehensive microarray chips for gene expression analysis were established and used for the analysis of global- and organ-specific gene expression in *Petunia axillaris* and *Petunia inflata* and to explore the molecular basis of the seed coat defects in a *Petunia hybrida* mutant, *anthocyanin 11* (*an11*), lacking a WD40-repeat (WDR) transcription regulator. Among the transcripts differentially expressed in *an11* seeds compared with wild type, many expected targets of AN11 were found but also several interesting new candidates that might play a role in morphogenesis of the seed coat. Our results validate the combination of next-generation sequencing with microarray analyses strategies to identify the transcriptome of two petunia species without previous knowledge of their genome, and to develop comprehensive chips as useful tools for the analysis of gene expression in *P. axillaris, P. inflata and P. hybrida.*

Keywords: ***Petunia axillaris, Petunia inflata*, next-generation sequencing, transcriptome, AN11.**

INTRODUCTION

Petunia hybrida, ('garden petunia') is a popular ornamental all around the world and a useful model system for plant research. Due to its small dimensions, short generation time, abundant seed production, petunia has been widely used for genetic and physiological studies (for reviews see Gerats and Vandenbussche, 2005 and Gerats and Strommer, 2009). Moreover, petunia lines in which *dTPH1* elements transpose at high frequency provide excellent tools for forward and reverse genetic analyses (van Houwelingen *et al.*, 1998; Vandenbussche *et al.*, 2008). In addition, petunia is easy to transform and susceptible to virus-induced gene silencing (VIGS) (Chen *et al.*, 2005). Another important point is that petunia functions in many ways different from Arabidopsis, as it, for example, interacts with mycorrhiza, has a different architecture, is pollinated by insects, and has fused petals (Gerats and Strommer, 2009). Thus it is not only an excellent system for comparative analyses but also to study processes that can or cannot be easily analyzed in Arabidopsis.

Genome-wide expression analysis has so far rarely been applied in petunia, simply due to the lack of sequence information. Recently, an expressed sequence tag (EST) collection was established to identify genes involved in mycorrhiza interactions (Breuillin *et al.*, 2010). However, the

existing EST catalogue is still rather limited for an in-depth characterization of petunia biology and genome activity. Therefore, the almost exhaustive list of transcribed sequences provided from a set of representative organs is an important achievement. RNA-seq, is a new approach based on 454 next-generation sequencing technology and aiming to analyze a transcriptome in a very short time, even when a reference genome is lacking (Bellin *et al.*, 2009). This situation suggests that the approach can be exploited to characterize the transcriptome of petunia and to develop high density microarray chips for transcriptomic analyses. Some of the valuable insights obtained from petunia are related to the genetic control of flower pigmentation (Quattrocchio *et al.*, 1993; Koes *et al.*, 2005; Tornielli *et al.*, 2009) and plant development (Angenent *et al.*, 2005; Cartolano *et al.*, 2007). For example *ANTHOCYANIN1* (*AN1*) and *AN11* encode transcription regulators with respectively HLH and an WDR domain that in petals interact with the MYB protein AN2 to activate genes involved in anthocyanin synthesis. Interaction with a distinct MYB protein, PH4, activates another pathway that acidifies the vacuole where anthocyanins are stored (Quattrocchio *et al.*, 2006). In seeds, *AN1* and *AN11* control the pigmentation and morphogenesis of the seed coat epidermis, via yet unknown mechanisms. Homologous genes have been shown to control anthocyanin synthesis in variety of other species, including maize, *Antirrhinum* and Arabidopsis. In Arabidopsis, the HLH proteins GLABRA3 (GL3) and ENHANCER OF GL3 (EGL3), and the AN11-homologue TTG1 control, besides anthocyanin synthesis, also the formation of trichomes on aerial tissues and non-hair cells in the root epidermis, in part via activation of the downstream gene *GL2* (Ishida *et al.*, 2008). Interestingly, such a role in (non)-hair development is unique for Arabidopsis and is not seen in other species, suggesting that this function of AN11 homologues evolved recently (Spelt *et al.*, 2002; Carey *et al.*, 2004; Pang *et al.*, 2009). Similarly, the role of AN1 in promoting cell growth and preventing cell division in the developing seed coat epidermis is seen in petunia, but not in other species (Spelt *et al.*, 2002).

To further unravel how the conserved HLH and WDR proteins acquired these novel functions, we set out to characterize the transcriptome of wild-type and *an11* mutant seeds. We used 454 sequencing to generate, through *de novo* read assembly, a 'complete' catalogue of unique transcripts (Tentative Consensus sequences plus singletons) for two parental species, *Petunia axillaris* and *Petunia inflata*, from which the large array of *Petunia hybrida* cultivars were generated by conventional breeding nearly two centuries ago. The transcriptome of each parental species was analyzed, to collect the whole set of genetic information which potentially contributes to the *Petunia hybrida* genotypes, to reveal species-specific and common features, and to establish comprehensive microarray chips for large scale gene expression analysis. In addition, RNA from wild-type and *an11 P. hybrida* mutant seeds was hybridized to both *P. axillaris* and *P. inflata* arrays to gain information about the molecular mechanisms by which AN11 controls the morphogenesis of the seed coat.

RESULTS AND DISCUSSION

Transcriptome sequencing and gene discovery in *Petunia axillaris* and *Petunia inflata*

We aimed to prepare a complete collection of transcript fragments from two petunia species, *P. axillaris* and *P. inflata*, generally considered the ancestors of *P. hybrida* (garden petunia). *P. axillaris* has white, salverform flowers with petals that emit scent to attract night moths for pollination, whereas, *P. inflata*, exhibits purple flowers with blue pollen to attract bees for pollination (Figure 1 and Stehmann *et al.*, 2009).

To obtain an exhaustive set of expressed sequences, we pooled total RNA from fertilized ovary, floral bud, callus, inflorescence and seedling of each of the two petunia species and prepared normalized cDNA libraries enriched for 3′ cDNA ends (Bellin *et al.*, 2009). 454 sequencing resulted in 578 107 raw reads from *P. axillaris* and 602 753

Figure 1. Comparison between *P. axillaris* and *P. inflata* flowers and whole plants.

from *P. inflata*, which were reduced to respectively 513 135 and 535 446 high quality reads via pre-processing. The two collections were processed independently using the ParPEST pipeline (D'Agostino *et al.*, 2005) and assembled into 27 731 tentative consensus sequences (TCs) of 555 bp average length for *P. axillaris* and 29 329 TCs of 556 bp average length for *P. inflata*, using 77 and 78% of the total reads, respectively (Table 1). All sequence reads that did not meet the criteria to be clustered/assembled with any other sequence in the collection were defined as singletons (sESTs).

The TCs and sESTs represent a collection of unique transcript fragments providing a survey of the *P. axillaris* and *P. inflata* transcriptomes (data accessible at http://biosrv.cab.unina.it/454petuniadb/) which comprises 45.63 Mbp and 47.93 Mbp of expressed sequences, respectively.

We compared our collection of transcript fragments with the total of EST/mRNA sequences from petunia available in GenBank on November 2010 (Table S1) and the non-redundant set of unique sequences recently described by Breuillin *et al.* (2010). We aligned all 454-based unique transcript fragments (TCs and sESTs) from both *P. axillaris* and *P. inflata* to the entire known petunia transcriptome using BLASTn and accepted only the best hits with an E-value < $1e^{-10}$. This showed that only 50% of the TCs and sESTs have significant similarity with previously published expressed sequences, implying that the remaining represents novel transcripts (Table S2).

To explore the functional information content of these petunia TC collections we followed the annotation workflow depicted in Figure S1. BLASTx searches (cut off E-value = 10^{-6}) of petunia TCs were performed against the UniProtKB/TrEMBL database (Dataset S1). Association to Gene Ontology terms (GO) were inferred electronically by exploiting TrEMBL entries. In addition, the *Arabidopsis thaliana* protein complement (TAIR 9 release) was selected as a reference proteome for carrying out BLAST comparisons, resulting in 16 418 (59.2%) TCs from *P. axillaris* showing significant matches to 10 396 distinct Arabidopsis proteins, and 16 900 (57.6%) TCs from *P. inflata* to 10 574 distinct Arabidopsis proteins.

Sequences that did not find a match in the UniProtKB/TrEMBL database were used for BLASTn searches against the GenBank non-redundant nucleotide database. Finally, the sequences, for which no homologues were found so far, were compared with the dbEST database in order to check their expression in independent expression data collections (i.e. those obtained by Sanger EST sequencing). As expected, most EST sequences identified by BLAST searches originate from either petunia (*P. axillaris* or *P. hybrida*) or from the closely related species *Nicotiana tabacum* and *Solanum lycopersicum* (data not shown). Sequences for which no matches were found are referred to as 'orphan sequences'. Overall, significant similarity was found for more than 80% of the TCs (Figure 2).

The sESTs were annotated according to a similar workflow resulting in 67.9% (*P. axillaris*) and 65.4% (*P. inflata*) of the total sESTs showing sequence similarity with known sequences in public databases. The informative potential of sESTs is restrained by the fact that they are represented by one single sequence which make their quality hard to assess. For this reason we decided to exclude them from further consideration in this work.

Table 1 Summary of *P. axillaris* and *P. inflata* unique transcript collections after 454-reads assembly

	P. axillaris	*P. inflata*
TCs		
Number of sequences	27 731	29 329
Average length	555.77 ± 145.15	556.70 ± 143.60
Min seq length	95	89
Max seq length	1922	1545
Reads in TCs		
Number of sequences	396 454	417 921
Average length	357.89 ± 67.53	363.61 ± 66.14
Min seq length	87	88
Max seq length	693	677
sESTs		
Number of sequences	116 681	117 525
Average length	261.20 ± 133.69	271.28 ± 134.72
Min seq length	40	40
Max seq length	648	635

Tentative consensus (TCs), high quality reads (reads) and singletons (sESTs) lengths are expressed as bp.

Development of microarray chips and their application in transcriptional analysis

From the collections of *P. axillaris* and *P. inflata* TC sequences, we designed comprehensive microarrays for expression analysis. To create specific oligonucleotide probes we adopted the OLIGOARRAY 2.1 software (Rouillard *et al.*, 2003). This probe design produced 27 594 oligonucleotide sequences for *P. axillaris*, (among which 859 probes matching more than one TC) and 29 163 oligonucleotides sequences for *P. inflata* (875 matching more than one TC).

In a first experiment, we verified that the *P. axillaris* and *P. inflata* arrays could identify sequences transcribed in different petunia organs, and we compared results obtained by hybridization of the two different arrays with three biological replicates of RNA from fertilized ovary, bud, callus, seedling and inflorescence (the same RNA samples as used for the cDNA library preparation).

The principal component analysis (PCA) system efficiently separated seedling, bud, fertilized ovary and inflorescence samples in both species, confirming also the uniformity of the three biological replicates (Figure 3a, b). The two principal components explained 44 and 41% of the overall variance of transcription profiles, in *P. axillaris* and in

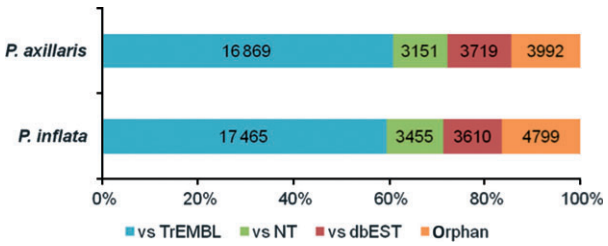


Figure 2. More than 80% of the *P. axillaris* and *P. inflata* tentative consensus sequences (TCs) find significant similarity.
In the 100% stacked bar chart is represented the percentage of genes that found sequence similarity in each (database) step of the annotation workflow.

P. inflata respectively. The greatest variance in gene expression was found, for both species, between fertilized ovary and seedling, as they were separated along the first component, while it was not possible to clearly differentiate callus from the other organs, supporting the idea that callus is not a tissue on its own, but rather a mixture of cells with different identities, or dedifferentiated cells.

The transcriptome of the different samples identified 18 807 *P. axillaris* probes (out of 27 594, amounting to 68.15%) and 19 668 *P. inflata* probes (out of 29 163, or 67.44%) that detected expression in at least one of the five samples tested. The lack of any hybridization signal for approximately 30% of the transcripts that were detected by 454-sequencing procedure, demonstrated the higher sensitivity of deep sequencing of normalized cDNA libraries as compared with the microarray hybridization analysis.

In particular 11 309 (60.13% of all expressed genes) and 11 707 (59.52%) genes were expressed in all five samples, while 2502 (13.3% of all expressed genes) and 2660 (13.52%) genes were expressed only in a single sample, in *P. axillaris* and *P. inflata* species, respectively (Figure 3c).

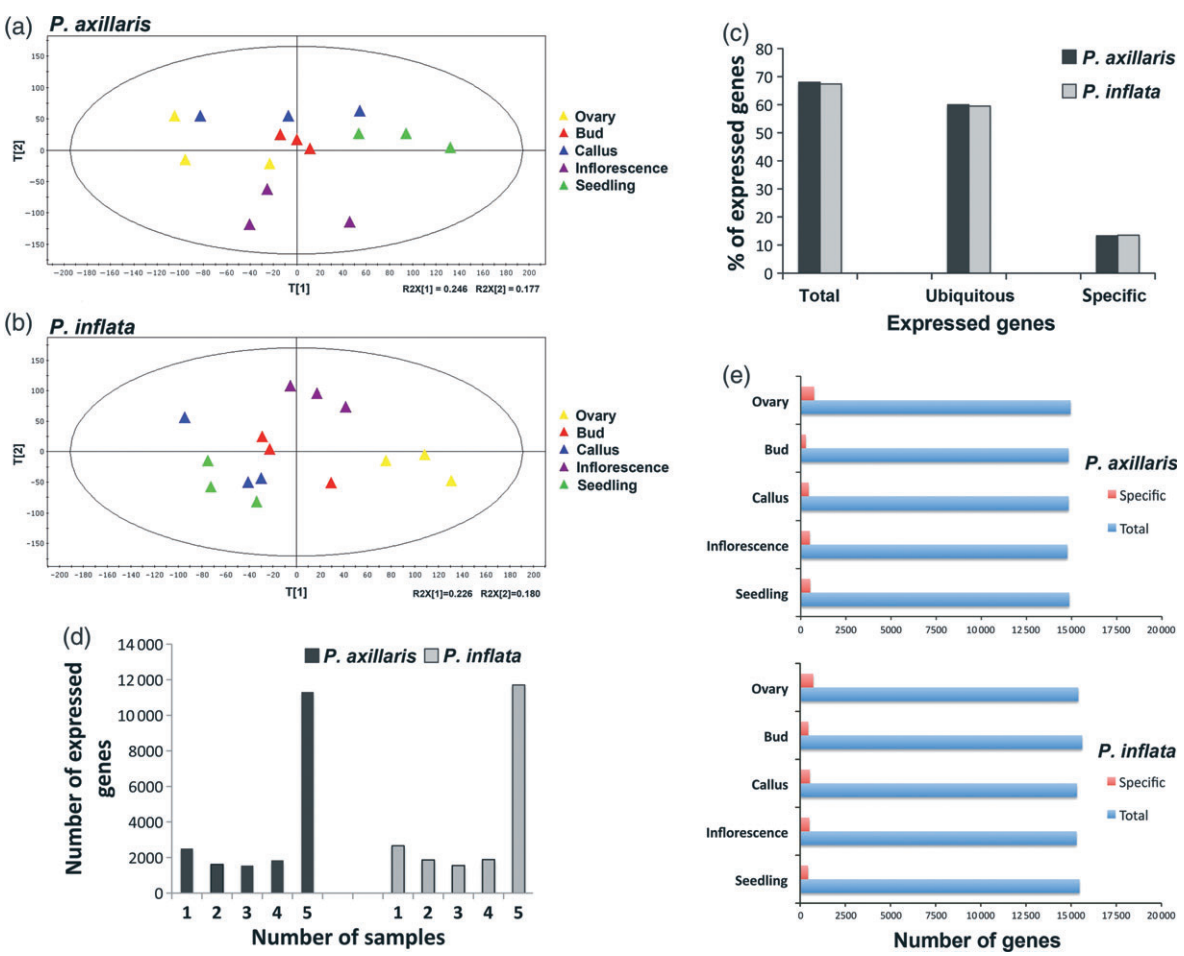


Figure 3. Comparison between *P. axillaris* and *P. inflata* organ expression analysis.
Transcriptomic dataset structure description by principal component analysis (PCA): scatter plot showing the three biological replicates of the *P. axillaris* (a) and *P. inflata* (b) samples. (c) Percentages of genes expressed in at least one sample over the total of the genes represented into the chip (Total), genes expressed in all samples over the total of expressed genes (Ubiquitous) and genes expressed in only one sample over the total of expressed genes (Specific), in *P. axillaris* and

The distribution of the number of genes expressed in multiple samples (Figure 3d), the total number of transcripts represented in each sample and the number of sample-specific transcripts (Figure 3e) are consistently similar in the two petunia species.

To determine significant differences in gene expression in the five petunia samples, all expressed genes were analyzed by significance analysis of microarrays (SAM) (T-MeV 4.3). Using a false discovery rate (FDR) of 5%, we identified 8547 transcripts for *P. axillaris* and 4070 for *P. inflata* that are expressed differentially in two or more samples. We have selected a group of probes showing at least a two-fold change in transcript abundance in at least one sample comparison (5407 probes for *P. axillaris* and 4069 for *P. inflata*); we clustered those into 15 groups (Figure 4) according to specificity in expression by k-mean method (T-Mev 4.3). A detailed analysis of the sample-specific transcript detection and the cluster assignment of each gene are contained in Dataset S2. The expression profile of a selection of modulated genes was further verified by real-time reverse transcriptase polymerase chain reaction (RT-PCR) (Figure S2).

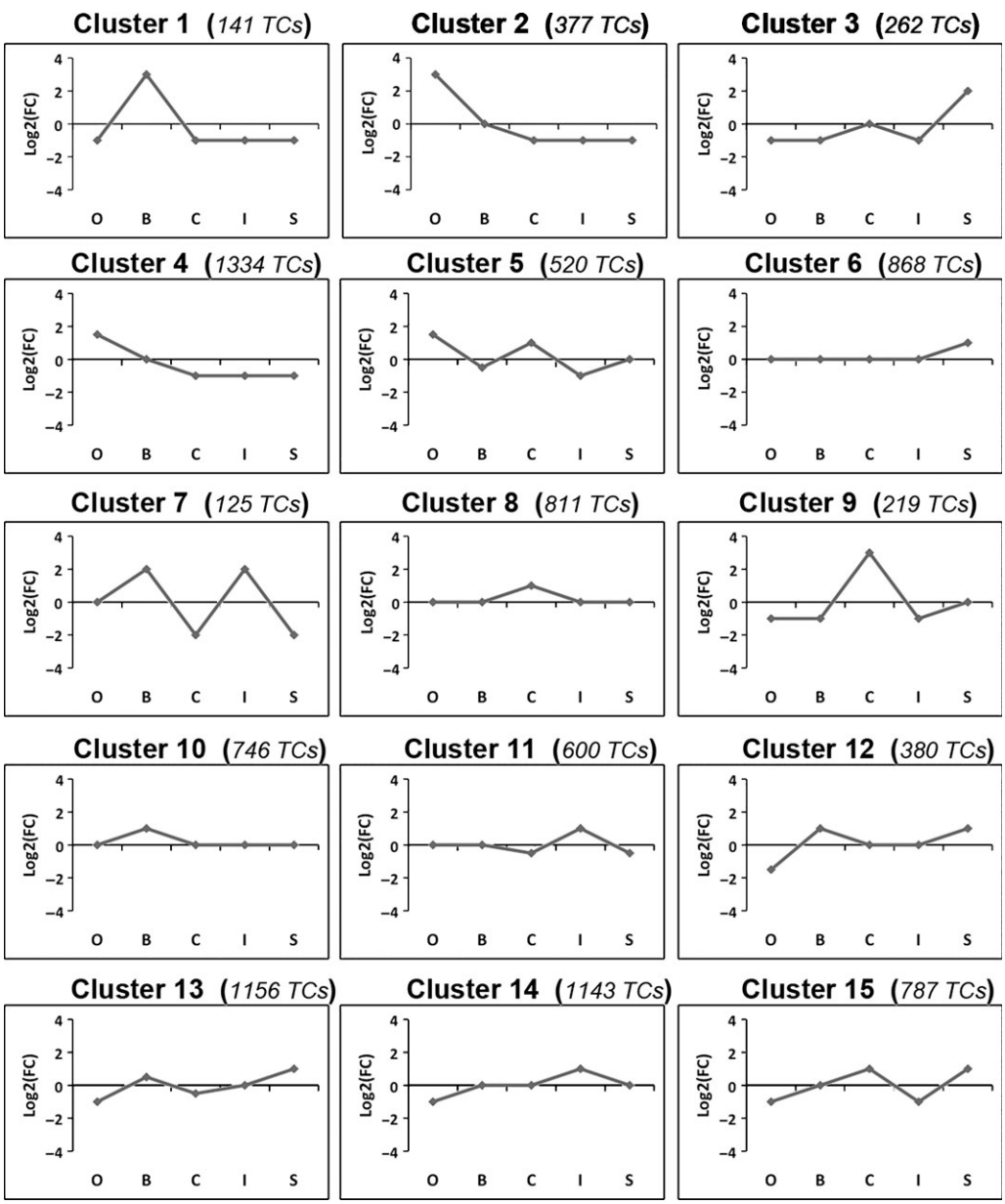


Figure 4. Distribution of *P. axillaris* and *P. inflata* differentially expressed tentative consensus sequences (TCs) into 15 expression clusters. The number of TCs belonging to the cluster is showed in brackets. O, fertilized ovary; B, floral bud; C, callus; I, inflorescence; S, seedling.

A list of genes that can be reliable used in normalization of relative transcript abundance has recently been reported for *Petunia hybrida* (Mallona *et al.*, 2010) and this list was confirmed by the chip hybridization results. *Actin 11* (*ACT*), *Elongation factor 1-alpha* (*EF1alfa*), *Ribosomal protein S13* (*RPS13*), *GTP-binding nuclear protein* (*RAN1*), and *SAND family protein* (*SAND*) are expressed ubiquitously and their expression levels are not significantly different in the five analyzed samples for both petunia species (Table S3). *Tubulin beta-6 chain* (*TUB*) has been found only in *P. inflata* where it is constitutively expressed. In constrast, *Glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) shows significant differences in expression level in both species, confirming that this gene is not a good one to be used in normalization (Mallona *et al.*, 2010). The candidates *Cyclophilin-2* (*CYP*) and *Ubiquitin* (*UBQ*) genes are not represented in the two petunia chips.

Together these data show that the petunia microarray chips are suitable for transcriptomic studies.

Comparison of *P. inflata* and *P. axillaris* transcriptomes

We investigated sequence conservation between the transcriptomes of *P. axillaris* and *P. inflata* by three different bioinformatics approaches.

First, we assessed the extent to which *P. inflata* and *P. axillaris* TC sequences overlap, organizing them into 'mixed super-assemblies' using the CAP3 algorithm-based approach. We found that 11 786 TCs from *P. axillaris* and 11 804 from *P. inflata* collapse into 10 371 mixed super-assemblies (8055 mixed super-assemblies are made up of two sequences, one from each species and the remaining 2316 are made up of more than two sequences. Dataset S3).

Second, a reciprocal BLASTn comparison (E-value cut off $1*e^{-10}$) between *P. inflata* and *P. axillaris* TCs was used to evaluate sequence similarity. By this approach, we identified 8064 sequence pairs, with an identity cut off ≥90% and a minimum match length of 75% of the subject length (Dataset S3).

Third, we estimated how many homologues of proteins in the UniProtKB/TrEMBL protein dataset are present in both *P. axillaris* and *P. inflata* TC collections (Dataset S3). 8342 TrEMBL entries turned out to find correspondence to both *P. axillaris* and *P. inflata* TCs (11 326 and 11 525, respectively); 5118 proteins in TrEMBL were solely associated to *P. axillaris* TCs (in total they matched 5543 TCs), while 5475 TrEMBL entries matched exclusively to *P. inflata* TCs (in total they matched 5940 TCs). The association of more than one TC of one petunia species to the same TrEMBL entry could be due to transcripts that are represented by two or more TCs, or to multigene families. The fact that a remarkable number of TCs from either species did not find a common match is not surprising, as results returned by BLAST are certainly affected by the partial nature and the length of the query sequences.

Using these three different approaches almost two-thirds of the 27 731 *P. axillaris* and of the 29 329 *P. inflata* TCs found matches (Figure 5). The remaining one-third of TCs (9937 for *P. axillaris* and 11 302 for *P. inflata*) did not match by any of these analyses.

We also evaluated the frequency of matching TCs showing the same expression profile in microarray hybridization, to assess the reliability of the matches. This analysis supports a positive correlation between TCs association and their expression pattern, as the percentage of TCs having a counterpart with the same expression is on average 34% for *P. axillaris* and 37% for *P. inflata* (Table S4) of the total TCs belonging to a mixed super-assembly, with peaks of 68% (cluster 7 – *P. axillaris*) and 58% (cluster 2 – *P. inflata*). Similar results were obtained considering TCs matching by BLASTn: on average, 33% of *P. axillaris* TCs and 38% of *P. inflata* TCs of the total belonging to a pair by BLASTn in the same expression cluster, found a counterpart in the same cluster (Table S4).

Overall the evaluation of the similarity between the two petunia transcript collections and the consistent expression patterns of matching TCs, provide a good description of the potential overlapping information between the two petunia transcriptomes. This similarity is likely to be underestimated, due to the lack of a reference genome for the mapping of the assembled reads, and to the fact that the two TC collections mainly represent the 3′ untranslated region of transcripts.

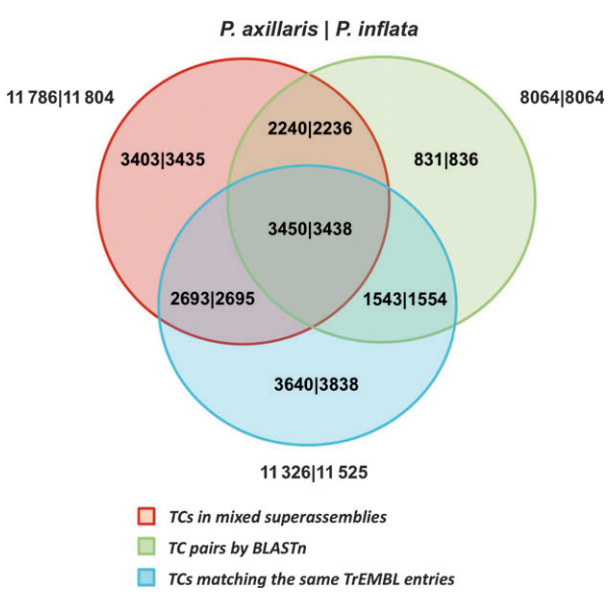
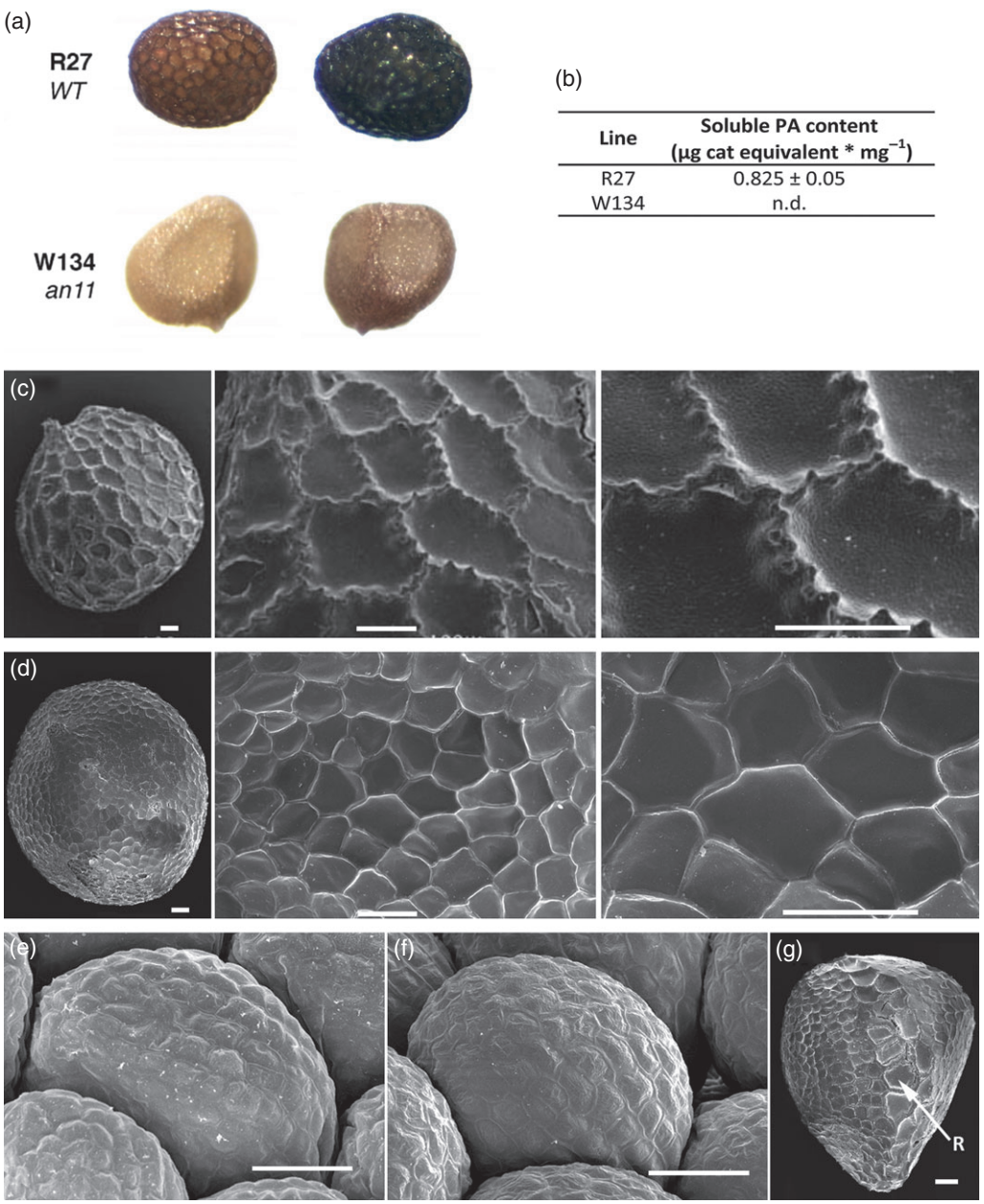


Figure 5. Almost two-thirds of *P. axillaris* and *P. inflata* tentative consensus sequences (TCs) are overlapping.

Phenotypic characterization of *an11* mutant seeds

The seed coat of wild-type petunia lines has a dark brown colour and consists of large cells, whereas the seed coat of *an1* mutants has a light yellowish colour and consist of many more small cells (Spelt *et al.*, 2002). Seeds of *an11* lines display the same light colour as *an1* seeds and also harbour an increased number of cells (Figure 6a). Scanning electron microscopy showed that the seed coat epidermis of mature wild-type seeds is made up of relatively large cells (approximately 100 μm) that are joined by thick wavy crests to the neighbouring cells. The *an11* seed coat epidermis instead consists of smaller cells that are joined by thinner and smoother crests (Figure 6c, d, g). Moreover, in wild type the



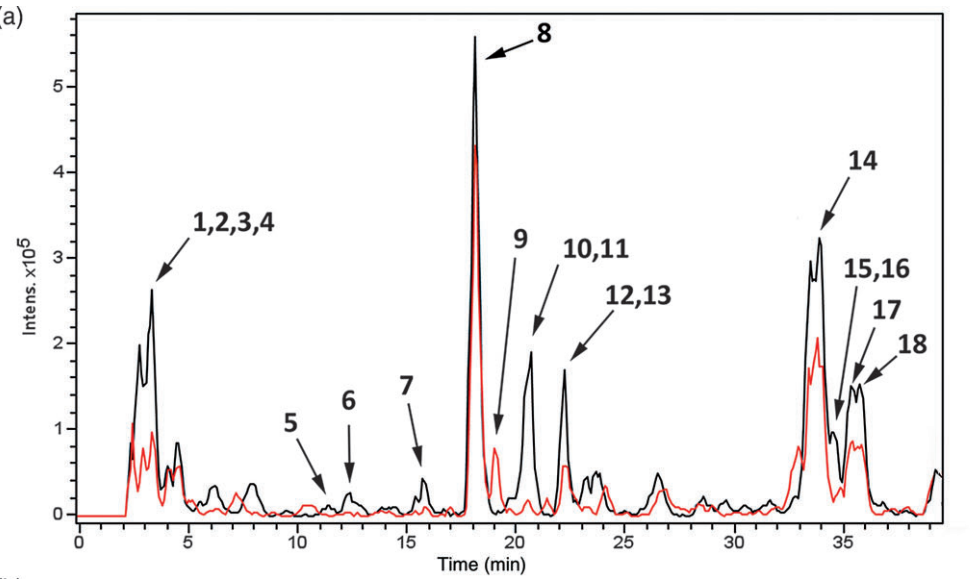
Line	Soluble PA content (μg cat equivalent * mg^{-1})
R27	0.825 ± 0.05
W134	n.d.

Figure 6. Effect of *an11* mutation on the PA content and on the morphology of seed coat cells.
(a) Mature seeds of the wild type line R27 (top) and *an11* W134 line (bottom), unstained (left) and stained with dimethylaminocinnamaldehyde (DMACA) reagent to detect proanthocyanidins (PAs) (right). (b) Soluble PAs levels from wild type line R27 and *an11* W134 line; mean values ± SD of three replicates for each line are reported. (c–g) Scanning electron micrographs of the wild type line R27 seed coats (c), *an11* line W134 seed coats (d), ovule of wild-type line R27 (e), ovule of *an11* line W134 (f), seed coat of W137 line (g) homozygous for the *dTph1* transposon unstable *an11-W137* allele from which the stable recessive *an11-W134* allele originated: note a sector of revertant cells marked with R. Bars = 50 μm.

number of cells in the seed epidermis appears similar to that of the unfertilized ovule, indicating that during seed development the epidermal seed coat cells only grow in size but do not divide. In *an11* the epidermal seed coat cells appear to divide during seed development, as the number of cells in the *an11* ovule epidermis is similar to wild type, whereas the cell number in the epidermis of the mature seed is much larger (Figure 6e, f). Because the increase in cell number is counterbalanced by a reduction in cell size, the coat of *an11* seeds reaches the same final size as in wild type seeds.

The brown pigments in the seed coat yield an anthocyanin (delphinidin) after acid hydrolysis and are eliminated by

(a)



(b)

ID	RT (min)	m/z(–)	Putative identification	R27	W134
1	2.9	341	Sucrose	50.9 ± 12.2	23.5 ± 5.8
2	2.9	377	Sucrose, chloride adduct	513.2 ± 4.2	28.5 ± 1.7
3	2.9	191	Citric acid	897 ± 4.9	268.4 ± 12.9
4	2.9	421	Citric acid derivative	257.4 ± 18.6	36.6 ± 4.2
5	11.4	517	Ferulic acid derivative	46.1 ± 7.9	12.2 ± 3.5
6	13.4	278	Coumaric acid + aspartic acid	43.7 ± 1.9	4.2 ± 1.5
7	15.8	449	Tetrahydroxyflavanone hexoside	46.1 ± 10.2	2.7 ± 1.2
8	18.3	625	Quercetin di-hexoside	1170 ± 212.1	1107 ± 188
9	19.4	463	ui	0	165.5 ± 5.4
10	20.7	609	Kaempferol O-di-hexoside	368.2 ± 22.5	31 ± 0.8
11	20.8	579	Catechin/epicatechin 2M-1 adduct (fragment of procyanidin)	481.8 ± 8.9	0
12	22.3	639	Petunidin di-hexoside	363.7 ± 10.7	149.2 ± 4.2
13	22.3	675	Petunidin di-hexoside, chloride adduct	83.4 ± 5.5	45.8 ± 4.8
14	33.7	1097	Tetrahydroxy flavanone O-penta hexoside	1650.3 ± 92.7	969.7 ± 39.4
15	34.7	655	ui	403.3 ± 34.6	21.8 ± 0.2
16	34.7	691	Compound 15, chloride adduct	99.2 ± 3.1	15 ± 5.2
17	35.2	935	Tetrahydroxy flavanone O-tetra hexoside	878 ± 49.6	444.7 ± 18.2
18	35.8	655	ui	147.2 ± 23.3	7.1 ± 1.2

Figure 7. Changes in mature seeds metabolite levels in *an11* line W134 compared with wild type line R27.
(a) HPLC-MS chromatograms of wild type line R27 (black) and *an11* line W134 (red) seed methanolic extracts. (b) Table showing the retention time (RT), mass to charge ratio value in negative mode (*m/z*(–)), the putative identification and the relative amount of the identified molecule, represented as peak area means/ 10 000 ± SD, from two biological replicates. The identification numbers (ID) correspond to those indicated in the chromatogram. ui, unidentified compound.

mutations in the genes encoding FLAVONOID 3-HYDROXYLASE (*an3*) or DIHYDROFLAVONOL 4-REDUCTASE (*an6*) suggesting that they are proanthocyanidins (PAs) (Spelt *et al.*, 2002). Indeed, wild type seeds stain blue with dimethylaminocinnamaldehyde (DMACA), which reacts with PAs and its flavan 3-ol precursors, whereas *an11* seeds do not stain. The PA phenotype was further confirmed by a staining assay on total soluble PA in methanolic extracts of ground seeds. The PA content of R27 seeds was 0.825 μg mg^{-1} (±0.05), while in *an11* mutant seeds no PAs were detectable with this method (Figure 6a, b). Thus, inactivation of *AN11* abolishes PA accumulation, similar to the effect of mutation in the homologues *AtTTG1* of Arabidopsis (Baudry *et al.*, 2004) and *MtWD40-1* of *Medicago truncatula* (Pang *et al.*, 2009).

Next, we analyzed the seed extracts by high-performance liquid chromatography coupled to ion trap mass spectroscopy (HPLC-MS). Several compounds appeared reduced or not detectable in *an11* seeds at 25 DAP (days after pollination) compared with R27 seeds at the same developmental stage (Figure 7a, b), in particular intermediates of the phenylpropanoid pathway, whereas the amount of quercetin was constant in both genotypes. Sugar content was also dramatically reduced in the mutant, which might indicate a delayed (or reduced) development of the embryo resulting in lower accumulation of storage compounds in the seed.

In *M. truncatula MtWD40-1* mutant multiple phenylpropanoid compounds such as benzoic acids, flavonols and flavan 3-ols are also reduced (Pang *et al.*, 2009). However this is partially in contrast with the role described for the *PhAN11* in petunia petals (de Vetten *et al.*, 1997) and *AtTTG1* in Arabidopsis seeds (Shirley *et al.*, 1995; Lepiniec *et al.*, 2006) which were shown to control only the late steps of the phenylpropanoid/flavonoid pathway that do not affect flavonols and other uncolored products of this biosynthetic pathway.

Analysis of the effect of the *an11* mutation on seed transcriptomes

Contrary to the HLH and MYB gene partners *AN1* and *AN2*, *AN11* mRNA is expressed in all plant organs (de Vetten *et al.*, 1997), which is confirmed by our microarray data (Dataset S2: PETAX014888|Contig1 and PETIN001385|Contig 1).

As *AN11* promotes flower pigmentation indirectly by activating the transcription of downstream structural genes encoding enzymes of the anthocyanin and vacuolar acidification pathway, we presumed that AN11 promotes PA synthesis and seed coat morphogenesis via yet unknown target genes. To identify these target genes, we hybridized RNA from seeds of line W134 (*an11^{W134}* mutant) and the progenitor line R27 (*AN11*) to the *P. axillaris* and *P. inflata* chips. *an11^{W134}* is a stable recessive allele with a frame shift mutation in the region encoding the first (of five) WD40 repeats that severely truncates the protein (de Vetten *et al.*, 1997). In *an11^{W134}* flowers anthocyanin synthesis is completely blocked indicating that *an11^{W134}* is a null allele, and *an11^{W134}* seeds display the same phenotype as *an11^{W137}* (de Vetten *et al.*, 1997; Spelt *et al.*, 2002).

By SAM analysis of microarray hybridization data (FDR = 0.125%) we identified 648 genes and 1279 genes that were differentially expressed in *AN11* and *an11^{W134}* seeds using the *P. axillaris* and the *P. inflata* chip, respectively. Among these, a subset of 521 genes from the *P. axillaris* chip and of 1049 genes from the *P. inflata* chip displayed a two-fold or greater change in transcript abundance (Dataset S4). That the two distinct microarrays detect such different numbers of expressed differentially genes may be due to larger genetic contribution of *P. inflata* to the genome of *P. hybrida* lines R27 and W134. Although generally only few mismatches are found between alleles from different *P. hybrida* lines or ancestral species, these can cause poor hybridization to the probes on the two chips as these are only 40 nt long. Moreover, because most probes are complementary to 3′ untranslated mRNA regions, the presence of mismatches in the probes could be significantly higher than on average along the whole transcript.

For further analysis we have chosen genes showing a fold change (FC) ≥4, resulting in 63 differentially expressed genes detected on the *P. axillaris* chip and 149 on the *P. inflata* chip. We have searched for a putative function for all these sequences by BLAST analyses (Table 2). Nine *P. axillaris* and 28 *P. inflata* differential transcripts remained without any putative function either because of lack of functional information for the most similar sequences in public databases (unknown proteins) or because little or no similarity was detected at all with known sequences (no hits). Notably, 24 TCs of *P. axillaris* that share identical function with 35 TCs of *P. inflata* showed a differential expression consistent with their counterpart(s), in the *an11* mutant seeds compared to wild type (Table 2).

The *an11^{W134}* mutation destabilizes the transcript in petals (de Vetten *et al.*, 1997) possibly because the frameshift and premature stop codons target this RNA for nonsense-mediated mRNA decay (NMD) (Bhuvanagiri *et al.*, 2010). Interestingly, this transcript (recognized by PETAX014888|Contig1 and PETIN001385|Contig 1 probes) is not down-regulated in *an11^{W134}* seeds as compared with *AN11*, which we confirmed by real-time RT-PCR (Figure S3). The observation that *an1^{W134}* transcript are present in the same amount as wild type *AN11* RNA might imply that the NMD mechanism is not active in seeds. This is an interesting possibility which needs however more experimental evidence to be confirmed.

Expected targets of AN11 in seeds are identified by microarray hybridization

Chip hybridizations showed, as expected, that several genes of the phenylpropanoid/flavonoid pathway are

Table 2 Differentially expressed TCs (≥4-fold) in mutant W134 seeds compared to the wild type line R27

P. axillaris TC ID	FC	Gene description	FC of *P. inflata* counterpart(s)
PETAX038429|Contig4	−58.58	Tetrapyrrole-binding protein, chloroplast precursor	
PETAX099154|Contig1	−35.18	CLAVATA3/ESR-RELATED 13 (CLE13) [*Arabidopsis thaliana*]	−67.25
PETAX003779|Contig1	−20.42	Phenylalanine ammonia-lyase (PAL) [Petunia hybrida]	−17.96
PETAX035006|Contig1	−19.02	Unknown protein	
PETAX003691|Contig1	−16.74	Skp1-like protein 2 (PSK2)	−11.14
PETAX040465|Contig1	−15.41	S-adenosylmethionine decarboxylase (SAMD)	−19.53; −18.40
PETAX006022|Contig1	−12.36	Metallothionein-like protein	
PETAX010809|Contig1	−12.33	Serine carboxypeptidase III	−18.84
PETAX089413|Contig1	−11.43	Zinc binding dehydrogenase	
PETAX054602|Contig1	−10.26	Flavonoid 3′,5′-hydroxylase (Hf2) [Petunia integrifolia]	
PETAX046078|Contig1	−9.57	NAC domain protein	−7.83; −4.56
PETAX093891|Contig1	−9.36	Homeobox protein GLABRA 2 [*Arabidopsis thaliana*]	
PETAX051907|Contig1	−9.09	Cytochrome P450 (CYP92B2v1) [Nicotiana tabacum]	−8.50
PETAX031301|Contig1	−8.13	Periaxin like	
PETAX003776|Contig3	−7.69	Aspartic proteinase 4	
PETAX099818|Contig2	−6.88	Auxin induced extracellular matrix structural constituent	
PETAX031287|Contig1	−6.85	Myo-inositol oxygenase	
PETAX033257|Contig1	−6.84	Late embryogenesis abundant protein	−4.16; −4.12
PETAX006472|Contig1	−6.79	BHLH 68 Transcription factor [*Arabidopsis thaliana*]	−7.25
PETAX003118|Contig2	−6.22	Beta (1,4)-xylosidase	−5.10
PETAX004857|Contig1	−6.03	ACC oxidase 4 (ACO4) [Petunia hybrida]	−5.45
PETAX023177|Contig2	−5.48	Mitochondrial dicarboxylate carrier protein, putative	
PETAX089424|Contig1	−5.44	no hit	
PETAX087639|Contig1	−5.41	Unknown protein	
PETAX042833|Contig1	−5.34	ACC oxidase 3 (ACO3) [Petunia hybrida]	−5.45
PETAX041066|Contig2	−5.33	Zing finger (C3HC4-type RING finger) family protein	
PETAX090363|Contig1	−5.19	Dihydroflavonol-4-reductase (DFR-A) [Petunia hybrida]	
PETAX093240|Contig1	−5.05	Carbon-monoxyde oxygenase (ACYB-2)	−7.44; −5.60
PETAX006013|Contig3	−4.99	Glycine-rich protein	
PETAX003440|Contig1	−4.69	ATP citrate synthase (ACLA-1) [*Arabidopsis thaliana*]	−4.07
PETAX013056|Contig1	−4.56	Expansin-like protein (EXLA1) [Solanum lycopersicum]	
PETAX003927|Contig2	−4.50	Unknown protein	
PETAX043384|Contig1	−4.45	NAC domain protein	−7.83; −4.56
PETAX074350|Contig1	−4.43	Nodulin	−9.38; −7.69; −5.40; −4.02
PETAX087874|Contig1	−4.27	Unknown protein	
PETAX005496|Contig1	−4.17	Cyclotide 3d	
PETAX064549|Contig1	−4.16	Adenosine 3′-phospho 5′-phosphosulfate transporter	
PETAX017204|Contig3	−4.15	Anthocyanidin reductase (ANR)	
PETAX000482|Contig2	−4.13	Chloroplast ribosomal protein	
PETAX002380|Contig1	−4.13	GDSL-motif lipase/hydrolase family protein	−6.72; 4.04; 5.62
PETAX094955|Contig1	−4.13	Quercetin3-O-glucosyl transferase	
PETAX001794|Contig1	−4.12	S-adenosylmethionine synthetase	
PETAX001281|Contig2	−4.12	CP12-2, protein binding	
PETAX011304|Contig1	−4.10	Unknown protein	
PETAX003927|Contig1	−4.06	Glucosyltransferase	
PETAX017296|Contig1	−4.02	Hydroxymethylglutaryl-CoA reductase	
PETAX010438|Contig1	4.04	No hit	
PETAX008201|Contig1	4.11	SOUL heme-binding protein	
PETAX005642|Contig4	4.21	Dehydrin	
PETAX082641|Contig1	4.28	Alpha-galactosidase	
PETAX097967|Contig1	4.46	Unknown protein	
PETAX029664|Contig1	4.50	GDSL-motif lipase/hydrolase family protein	−6.72; 4.04; 5.62
PETAX024227|Contig1	4.55	Unknown protein	
PETAX089506|Contig1	5.04	Similar to pathogenesis-related protein	
PETAX013600|Contig2	5.11	GDSL-motif lipase/hydrolase family protein	−6.72; 4.04; 5.62
PETAX015583|Contig1	5.27	Glutathione S-transferase GST 23 [Glycine max]	
PETAX003488|Contig1	6.01	Vesicle transport v-SNARE 11 [*Arabidopsis thaliana*]	−5.93
PETAX082833|Contig1	6.06	Pyruvate kinase isozyme A, chloroplastic	
PETAX014273|Contig1	6.28	3-ketoacyl-CoA synthase	

Table 2 (*Continued*)

P. axillaris TC ID	FC	Gene description	FC of *P. inflata* counterpart(s)
PETAX001216|Contig2	6.81	40S ribosomal protein	−4.49; −5.31; −4.53
PETAX006800|Contig1	6.98	P18 protein	6.98
PETAX029664|Contig2	7.92	GDSL-motif lipase/hydrolase family protein	−6.72; 4.04; 5.62
PETAX093977|Contig1	28.25	Multidrug-efflux transporter	23.44
P. inflata TC ID	F.C.	Gene description	FC of *P. axillaris* counterpart(s)
PETIN013810|Contig1	−89.63	No hit	
PETIN077377|Contig1	−67.25	CLAVATA3/ESR-RELATED 13 (CLE13) [*Arabidopsis thaliana*]	−35.18
PETIN049299|Contig2	−46.55	(1–4)-beta-mannan endohydrolase, putative	
PETIN049986|Contig1	−36.79	No hit	
PETIN012345|Contig1	−33.91	Chalcone synthase (CHSj) [Petunia hybrida]	
PETIN031950|Contig1	−31.49	Abscisic stress-ripening protein (ASR)	
PETIN014529|Contig1	−31.37	Unknown protein	
PETIN027852|Contig2	−26.05	Anthocyanidin synthase [Petunia hybrida]	
PETIN012786|Contig1	−24.52	No hit	
PETIN079771|Contig1	−19.85	Zing-binding dehydrogenase	−11.43
PETIN033913|Contig1	−19.58	Leucoanthocyanidin reductase (LAR)	
PETIN004567|Contig1	−19.53	S-adenosylmethionine decarboxylase (SAMD)	−15.41
PETIN020617|Contig1	−18.98	Cytochrome b5 (DIFf) [Petunia hybrida]	
PETIN004704|Contig2	−18.84	Serine carboxypeptidase III	−12.33
PETIN027391|Contig1	−18.40	S-adenosylmethionine decarboxylase (SAMD)	−15.41
PETIN027852|Contig1	−18.28	Anthocyanidin synthase [Petunia hybrida]	
PETIN001564|Contig1	−17.96	Phenylalanine ammonia-lyase (PAL) [Petunia hybrida]	−20.42
PETIN007466|Contig3	−14.18	No hit	
PETIN033012|Contig1	−13.92	Ribonuclease	
PETIN007408|Contig4	−13.40	Anthocyanidin reductase (ANR)	
PETIN007466|Contig2	−11.14	Skp1-like protein 2 (PSK2)	−16.74
PETIN035820|Contig1	−9.97	No hit	
PETIN053028|Contig1	−9.74	No hit	
PETIN065677|Contig1	−9.44	Glycine-rich protein-interacting protein	−4.99
PETIN083072|Contig1	−9.38	Nodulin	−4.43
PETIN059287|Contig1	−9.28	Putative phosphatase	
PETIN003226|Contig1	−9.12	Fladoxin-like quinone reductase	
PETIN003122|Contig2	−8.95	Unknown protein	
PETIN000292|Contig1	−8.54	60S ribosomal protein	
PETIN053843|Contig1	−8.50	Cytochrome P450 (CYP92B2v1) [Nicotiana tabacum]	−9.09
PETIN003878|Contig1	−8.47	Protein kinase	
PETIN017724|Contig1	−8.23	Aldo-keto reductase, putative	
PETIN001037|Contig2	−7.90	Flavanone 3beta-hydroxylase (F3H) [Petunia hybrida]	
PETIN000852|Contig1	−7.83	Dehydration-responsive family protein	
PETIN039188|Contig1	−7.83	NAC domain protein	−9.57; −4.45
PETIN060596|Contig1	−7.70	UDP-glucose 6-dehydrogenase (UGDH)	
PETIN002725|Contig1	−7.69	Nodulin	−4.43
PETIN014912|Contig1	−7.52	Chalcone synthase (CHS-A) [Petunia hybrida]	
PETIN008477|Contig1	−7.44	Carbon-monoxide oxygenase (ACYB-1)	−5.05
PETIN005362|Contig3	−7.36	Thiamine biosynthesis protein ThiC	
PETIN000373|Contig1	−7.25	BHLH 68 Transcription factor [*Arabidopsis thaliana*]	−6.79
PETIN021589|Contig1	−7.08	Unknown protein	
PETIN063105|Contig1	−7.00	Auxin-responsive protein	
PETIN045934|Contig1	−6.96	Unknown protein	
PETIN043039|Contig1	−6.74	Unknown protein	
PETIN093053|Contig1	−6.72	Unknown protein	
PETIN004586|Contig1	−6.72	GDSL-motif lipase	−4.13; 4.50; 5.11; 7.92
PETIN043866|Contig1	−6.63	Adenine phosphoribosyltransferase 5 (APT5)	
PETIN011957|Contig1	−6.29	Unknown protein	
PETIN070127|Contig1	−6.10	Trypsin and protease inhibitor family protein	
PETIN014914|Contig1	−5.97	Reticulon family protein	
PETIN001337|Contig1	−5.93	Vesicle trasport v-SNARE 11 [*Arabidopsis thaliana*]	6.01
PETIN012203|Contig2	−5.90	No hit	
PETIN000852|Contig2	−5.77	Dehydration-responsive family protein	

Table 2 (*Continued*)

P. inflata TC ID	FC	Gene description	FC of *P. axillaris* counterpart(s)
PETIN004397|Contig3	−5.72	Sulfotransferase	
PETIN008477|Contig2	−5.60	Carbon-monoxide oxygenase (ACYB-1)	−5.05
PETIN002380|Contig1	−5.59	Glucosyltransferase, putative	
PETIN021332|Contig1	−5.58	Unknown protein	
PETIN029252|Contig1	−5.50	2-oxoglutarate-dependent dioxygenase	
PETIN074066|Contig2	−5.47	stearoyl acyl carrier protein desaturase	
PETIN008620|Contig1	−5.46	CTP synthase	
PETIN010064|Contig3	−5.45	ACC oxidase 1 (ACO1) [Petunia hybrida]	−6.03; −5.34
PETIN000286|Contig1	−5.42	Elongation factor 1-alpha	
PETIN002725|Contig3	−5.40	Nodulin	−4.43
PETIN008232|Contig1	−5.38	ULTRAPETALA1 (ULT1) DNA binding [*Arabidopsis thaliana*]	
PETIN032202|Contig1	−5.37	Unknown protein	
PETIN003197|Contig3	−5.31	40S ribosomal protein	6.81
PETIN010330|Contig1	−5.29	Epoxide hydrolase	
PETIN052227|Contig1	−5.27	No hit	
PETIN007378|Contig1	−5.12	Cysteine protease	
PETIN044467|Contig1	−5.10	Xylem serine proteinase	
PETIN045318|Contig1	−5.10	Beta-xylosidase	−6.22
PETIN090716|Contig1	−5.09	No hit	
PETIN012702|Contig1	−5.08	2-oxoglutarate-dependent dioxygenase	
PETIN026119|Contig1	−5.08	IRE1A; endoribonuclease [*Arabidopsis thaliana*]	
PETIN034762|Contig1	−5.08	Wall-associated receptor kinase-like 21 [*Arabidopsis thaliana*]	
PETIN068822|Contig1	−5.07	Heat-shock protein	
PETIN076947|Contig1	−5.05	UDP-glucuronic acid/UDP-N-acetylgalactosamine transporter-related	
PETIN004665|Contig3	−5.04	Cytochrome P450 monooxygenase	
PETIN001306|Contig1	−4.91	Heat-shock protein binding	
PETIN035463|Contig2	−4.90	CBL-interacting protein kinase	
PETIN000884|Contig2	−4.82	ATF2; enzyme activator	
PETIN021642|Contig1	−4.79	Unknown protein	
PETIN011646|Contig6	−4.78	Unknown protein	
PETIN058188|Contig1	−4.77	2-oxoglutarate-dependent dioxygenase	
PETIN078827|Contig1	−4.76	Homogentisate phytylprenyltransferase	
PETIN008706|Contig1	−4.72	Gamma-tocopherol methyltransferase	
PETIN004815|Contig1	−4.71	No hit	
PETIN023182|Contig3	−4.68	60S ribosomal protein	
PETIN022406|Contig1	−4.65	3-dehydroquinate dehydratase	
PETIN005538|Contig1	−4.64	Putative 4-coumarate-CoA ligase 2 (4CL)	
PETIN025093|Contig1	−4.57	Unknown protein	
PETIN040578|Contig1	−4.57	Unknown protein	
PETIN098061|Contig1	−4.57	Annexin, putative	
PETIN086640|Contig1	−4.56	NAC domain protein	−9.57; −4.45
PETIN000480|Contig2	−4.53	40S ribosomal protein	6.81
PETIN073715|Contig2	−4.52	Unknown protein	
PETIN034019|Contig1	−4.52	WRKY protein TTG2 [*Arabidopsis thaliana*]	
PETIN002683|Contig1	−4.49	40S ribosomal protein	6.81
PETIN007216|Contig1	−4.46	HISTONE H2A PROTEIN 11 (HTA11) [*Arabidopsis thaliana*]	
PETIN000698|Contig3	−4.45	PSBX (photosystem II subunit X)	
PETIN001023|Contig1	−4.43	Lipid transfer protein	
PETIN006747|Contig1	−4.42	Heat-shock protein	
PETIN061177|Contig1	−4.42	Pectinesterase inhibitor	
PETIN015427|Contig1	−4.41	Glutamyl-tRNA(Gln) amidotransferase	
PETIN022898|Contig1	−4.38	Hydroxycinnamoyl transferase	
PETIN012942|Contig1	−4.36	Putative sterol desaturase	
PETIN019711|Contig1	−4.34	Vacuole-associated annexin	
PETIN020635|Contig1	−4.34	Alpha/beta fold hydrolase (ABE)	
PETIN006621|Contig2	−4.29	Glycine-rich protein	−4.99
PETIN008174|Contig2	−4.28	Polyphosphoinositide binding protein	
PETIN029931|Contig1	−4.26	No hit	
PETIN025757|Contig1	−4.25	Calcium-dependent protein kinase	
PETIN076682|Contig1	−4.25	Polyadenylate-binding protein, putative	

Table 2 (*Continued*)

P. inflata TC ID	FC	Gene description	FC of *P. axillaris* counterpart(s)
PETIN081195|Contig1	−4.23	Response regulator histidine kinase	
PETIN017284|Contig1	−4.16	Late embryogenesis abundant protein	−8.64
PETIN011877|Contig1	−4.16	Unknown protein	
PETIN002706|Contig1	−4.15	Subtilisin-like protease	
PETIN010548|Contig1	−4.13	Auxin response transcription factor	
PETIN003249|Contig1	−4.13	Oligosaccharide transmembrane transporter OST3/OST6 [*Arabidopsis thaliana*]	
PETIN026196|Contig1	−4.12	Aminotransferase class I and II	
PETIN006326|Contig2	−4.12	Late embryogenesis abundant protein	−8.64
PETIN008644|Contig1	−4.07	ATP citrate synthase (ACLA-2) [*Arabidopsis thaliana*]	−4.69
PETIN056876|Contig1	−4.07	Unknown protein	
PETIN014719|Contig1	−4.06	Cell wall protein	
PETIN007867|Contig1	−4.06	AP2 domain-containing transcription factor	
PETIN008271|Contig1	−4.06	Alpha-expansin precursor	
PETIN004319|Contig1	−4.05	Mitochondrial 2-oxoglutarate/malate carrier protein	
PETIN046603|Contig1	−4.04	Vacuolar-type H^+-translocating inorganic pyrophosphatase	
PETIN004986|Contig3	−4.03	Obtusifoliol-14-demethylase [Petunia hybrida]	
PETIN013586|Contig1	−4.03	Enoyl-[acyl carrie -protein] reductase [NADH]	
PETIN002725|Contig2	−4.02	Nodulin	−4.43
PETIN002414|Contig1	4.04	GDSL-motif lipase/hydrolase family protein	−4.13; 4.50; 5.11; 7.92
PETIN062500|Contig1	4.13	Unknwon protein	
PETIN044748|Contig1	4.51	Glucose-methanol-choline (GMC) oxidoreductase family protein	
PETIN058662|Contig1	4.66	Transketolase	
PETIN006670|Contig1	4.73	MLP-like protein 28 [*Arabidopsis thaliana*]	
PETIN037855|Contig1	5.12	Inositol oxygenase 5 [*Arabidopsis thaliana*]	
PETIN006313|Contig1	5.12	Pectinesterase	
PETIN074226|Contig1	5.24	Lipid transfer protein	
PETIN033346|Contig1	5.38	Cytochrome P450 monooxygenase	
PETIN011298|Contig1	5.62	GDSL-motif lipase/hydrolase family protein	−4.13; 4.50; 5.11; 7.92
PETIN000556|Contig1	6.31	Glycine-rich protein	−4.99
PETIN016822|Contig5	6.98	P18 protein [Nicotiana alata]	6.98
PETIN000556|Contig2	7.44	Glycine-rich protein	−4.99
PETIN038182|Contig1	7.83	No hit	
PETIN008679|Contig2	11.02	Proline-rich extensin-like	
PETIN003438|Contig1	15.20	Phytanoyl-CoA dioxygenase	
PETIN024413|Contig1	23.44	Multidrug-efflux transporter	28.25

down-regulated in *an11* seeds. Among these are genes involved in the anthocyanin/proanthocyanidin pathway, such as *CHSj*, *F3′5′H*, *DIFf*, *DFR*, and *ANS*, which are known to be under the control of AN11 in petals (Koes *et al.*, 2005).

Among the down-regulated transcripts are also sequences encoding enzymes specific for the biosynthesis of PA, such as a putative leucoanthocyanidin reductase (LAR) and anthocyanidin reductase (ANR), which were not identified before in petunia (Table 2). Furthermore, we found an RNA encoding a glucosyltransferase gene (detected by the PETAX003927|Contig1 probe in the *P. axillaris* chip and by the PETIN002380|Contig1 probe in the chip of *P. inflata*) that is down-regulated in *an11* seeds. In Arabidopsis and *M. truncatula* seeds epicatechins are thought to be converted by a glycosyltransferase into epicatechin hexosides, prior to their transport into the vacuolar lumen via the MATE protein TT12 (Pang *et al.*, 2008; Kitamura *et al.*, 2010). However, the similarity of this petunia glucosyltransferase with the epicatechin-specific glucosyltransferase UGT72L1 of *M. truncatula* is very low (data not shown) and therefore its role in PA synthesis is unclear. The down-regulation of the abovementioned PA-specific genes, is in line with the tannin-deficient phenotype of *an11* seeds and identifies the PA branch of the pathway in petunia which was not yet described.

Similar to what reported for seeds of *M. truncatula MtWD40-1* mutant (Pang *et al.*, 2009), many early phenylpropanoid pathway genes (*PAL*, *4CL*, *CHSA*, *F3H* and *F3′H*) were down-regulated in *an11* seeds compared with wild type. The strong down-regulation of *PAL* was confirmed by real-time RT-PCR (Figure S3). These alterations in gene expression can explain the significant decrease in benzoic acid and kaempferol derivatives in *an11* seeds compared to wild type. The down-regulation of these early genes, which was not observed in *an11* petals, could be explained by: (i) direct control of their transcription by AN11 in the seed, (ii) a

consequence of a metabolic feedback, or (iii) an indirect effect due to the impairment of other biochemical pathways and/or developmental processes.

In the petal epidermis AN11 activates, together with AN1 and the MYB protein PH4, genes involved in the acidification of the vacuoles (Quattrocchio *et al.*, 2006). Among the downstream genes in this pathway are *PH3*, encoding a WRKY transcription factor, and *PH5*, which encodes a P-ATPase proton pump residing in the tonoplast (Verweij, 2007; Verweij *et al.*, 2008). It is known that the disruption of this pH-regulating pathway impairs the accumulation of PA in the seed coat (Baxter *et al.*, 2005 and Verweij *et al.*, 2008). Among the genes that are down-regulated in *an11* seeds, we found *PH3* (down-regulated 4.5-fold and 3.2-fold on the *P. inflata* and *P. axillaris* chips respectively), but not *PH5*. *PH5* was not represented on the microarrays because it was present as a singleton in the 454 sequencing. However, by real-time RT-PCR we could assess that *PH5* transcripts are nearly 5-fold less abundant in $an11^{W134}$ seeds compared with wild type (Figure S3). Interestingly, we found that a distinct pyrophosphatase H^+ pump (PETIN046603|Contig1) was also down-regulated in *an11* seeds, suggesting that AN11 controls, at least in seeds, multiple vacuolar acidification mechanisms.

The *AN11*-regulated genes described above are all expected to be involved in the pigmentation of the seed, consistent with the *an11* phenotype.

New AN11 target genes potentially involved in seed coat morphogenesis

The altered morphology of the *an11* seed coat cells cannot be attributed to defects in the synthesis of anthocyanins or PAs, or vacuolar acidification, as mutations in structural genes of these pathways do not affect seed coat morphology (Figure 6; Spelt *et al.*, 2002; Verweij *et al.*, 2008). Among the transcripts that are down-regulated in *an11* seeds were several interesting candidates that might play a role in morphogenesis of the seed coat. In *an11* seeds a homologue of the Arabidopsis *GLABRA2* gene (*GL2*) is down-regulated 9.36-fold (on the *P. axillaris* chip and confirmed by real-time RT-PCR; Figure S3). GL2 encodes a homeo-domain transcription factor that acts downstream of GL3/EGL3 and TTG1 in the specification of trichome fate in leaves and non-hair fate in roots. Because the specification of trichome fate causes trichome initials to exit from the normal mitotic cell cycle and to enter an endoreduplication cycle (Ishida *et al.*, 2008), this suggests that the role of AN11 in seed coat development – inhibition of cell divisions – may be mechanistically and evolutionary linked to certain aspects of trichome development in a previously unsuspected way.

A direct target of GL2 in Arabidopsis roots, a phospholipase D (PLD) is required for normal development of root hairs by regulation of membrane traffic (Ohashi *et al.*, 2003). We found a PLD encoding transcript that was down-regulated in *an11* seeds, albeit only two-fold (Dataset S4). In *an11* seeds RNAs encoding a different type of phospholipase (GDSL-motif lipase/hydrolase family protein) and a Qb-SNARE protein similar to the vacuolar v-SNARE11 of Arabidopsis (Mortimer *et al.*, 2008) are more than four-fold down- or up-regulated, suggesting that a similar pathway of membrane remodelling through lipid signalling and vesicle transport could control cell shape definition in seed coat development.

We also detected strong down regulation of a transcript with high similarity to *CLAVATA3/ESR-RELATED 13* (*CLE13*) of Arabidopsis (35.18-fold in *P. axillaris* and 67.25-fold in *P. inflata*) by chip hybridization, which we confirmed by real-time RT-PCR (Figure S3). CLEs are small plant peptides represented in Arabidopsis by a family of at least 31 members, some of which work as short-range signals for the maintenance of cell proliferation (Fiers *et al.*, 2007). As the peptides are rather small and the conserved motifs cover a large part of the protein, it is difficult to identify the real orthologous gene in Arabidopsis. To our knowledge there are no reports about a role of AtCLE13 or other CLE peptides in seed development. However, in legumes CLE homologues were shown to be involved in long distance signalling between shoots and roots to prevent hypernodulation, which is coupled with a high rate of cell proliferation (Okamoto *et al.*, 2009).

The similarities with the nodulation extend to the differential expression of several TC probes similar to the late nodulin *MtN21* from *M. truncatula*, (a member of the plant drug/metabolite exporter family *P-DME*). It is possible that the nodulin encoding TCs derive from the same transcript, but they could also represent different members of the same family. The differential expression of the *P. inflata* PETIN083072|Contig1 nodulin encoding TCs in the wild type and mutant seeds was confirmed by real-time RT-PCR (Figure S3). An Arabidopsis homologue, *WAT1* (Ranocha *et al.*, 2010), was shown to play a role in secondary cell wall formation. Remarkably, a transcript with similarity to this class of nodulins was described as down-regulated in *M. truncatula* seeds mutant for the *AN11* homologue (Pang *et al.*, 2009).

It is tempting to propose that homologues of CLEs and nodulins in seeds are involved in the proliferation of epidermal seed coat cells and could therefore play a role in defining the number and dimension of the cells in the seed coat epidermis. However, the isolation of gain- or loss of function mutants for these genes is necessary to assess their involvement in the phenotype of *an11* (as well as *an1*) mutant seeds.

It has been proposed that ethylene and polyamines control the activity of cell wall modifying enzymes during the last phase of seed development (Matilla and Matilla-Vázquez, 2008). The finding that the *an11* mutation down-regulates genes involved in the biosynthesis of these

hormones, such as *S*-adenosylmethionine decarboxylase (*SAMD*) and 1-aminocyclopropane-1-carboxylate oxidase (*ACO*), suggests that *AN11* may promote seed development in part via these hormones.

Several genes putatively involved in cell wall metabolism show differential expression in *an11* mutant seeds, including a (1–4)-beta-mannan endohydrolase (46.5-fold down regulation in *P. inflata*), a beta-xylosidase, an expansin and an expansin-like protein, a pectin esterase inhibitor and a cell wall related protein. The bifunctional beta-D-xylosidase/alpha-L-arabinofuranosidase *BXL1* gene of Arabidopsis was shown to play a role in the modification of the pectin structure in mucilage secretion from seed coat cells (Arsovski *et al.*, 2009). Petunia does not produce mucilage, but a similar enzyme could be involved in modifications of the cell wall structure that give rise to the peculiar crests of epidermal seed coat cells. A specific modification of pectin metabolism is also suggested by the opposite modulation of a pectinesterase inhibitor (down-regulated) and a pectinesterase (up-regulated).

The reduced seed coat epidermis cell size, characteristic of the *an11* mutant could be directly related to a reduced loosening of the cell walls. The down-regulation of two members of the expansin family and the up regulation of an extensin-like protein are consistent with this view.

General considerations

Taken together, our data show that the newly developed petunia chips, based on the transcriptomes of *P. axillaris* and *P. inflata*, are a useful tool for the analysis of gene expression in *Petunia hybrida*, which we applied here to define a collection of genes that are regulated by AN11 to obtain insight into functional diversification of the well known WDR–bHLH–MYB complex.

EXPERIMENTAL PROCEDURES

Plant material

Petunia axillaris and *Petunia inflata* plants were grown in soil under normal greenhouse conditions in Nijmegen. Callus was made from leaf discs, on MS medium, containing macro- and microsalts with Gamborg B5 vitamins, 3% sucrose, 2 mg L^{-1} 6-benzylaminopurine (BAP) and 0.1 mg L^{-1} 1-naphthalene acetic acid (NAA). Seedlings were grown for 3 weeks on half-strength standard MS medium.

Petunia hybrida (lines R27 and W134) were grown under normal greenhouse conditions in Verona. Seeds were collected at 15, 20 and 25 days after self-pollination (dap), and proportionally pooled.

RNA extraction and normalized library preparation

Total RNA from *P. axillaris* and *P. inflata* plant organs was isolated using Trizol reagent (Life Technologies, http://www.lifetechnologies.com). RNA from fertilized ovaries (17 dap), flower buds, callus, inflorescences and seedlings was mixed in a 1:2:2:2:1 ratio for both *P. axillaris* and *P. inflata* and used to prepare cDNA libraries (Eurofins MWG Operon, http://www.eurofinsdna.com) as described by Bellin *et al.* (2009). The resulting non-cloned libraries were normalized and subsequently sequenced on a half plate each, by GS FLX technology using Titanium series chemistry.

For hybridization analysis RNA was extracted from pooled seeds of R27 and W134 lines by using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, http://www.sigmaaldrich.com).

Bioinformatics tools for TC assembly and annotation

In total, 454 reads were processed in order to clip key, adapters, bad quality and primer tags used during the generation of cDNA libraries. Reads with percentage of nucleotides with a quality score <20, >50% per read, and, sequences shorter than 40 nucleotides in size were removed. Files containing 454 reads and their quality score are available from the Short Reads Archive division of the NCBI (National Center for Biotechnology Information) repository [Accession Number: SRA027293.1]. The two datasets were processed by the ParPEST pipeline (D'Agostino *et al.*, 2005). The masking of simple sequence repeats, low complexity subsequences and repetitive elements was performed as described in Alagna *et al.* (2009). Reads were grouped into clusters by the *wcd* cluster program (Hazelhurst *et al.*, 2008) and clusters were then assembled with CAP3 (85% sequence identity and 60 nucleotide overlap; Huang and Madan, 1999). Automated annotation was performed by BLASTx (E-value < 10^{-6}) searches against the UniProtKB/TrEMBL database (release 14.0) and the *Arabidopsis thaliana* protein complement (version TAIR 9 pep). Association to GO terms (GO) was electronically inferred by exploiting TrEMBL entries and a local MySQL copy of the full GO database (relase January 2010). In addition, BLASTn (E-value < 10^{-6}) searches against the nucleotide non-redundant database of GenBank (January 2010) and dbEST (est_others.gz, January 2010) were carried out. Files containing TCs sequences are available from the Transcriptome Shotgun Assembly division of the NCBI repository [Accession Numbers: JI332784-JI35915; JI359155-JI387099]. The complete collection is available at http://biosrv.cab.unina.it/454petuniadb/.

Microarray chips construction, hybridization and data analysis

The oligonucleotide probes were designed using the OLIGOARRAY 2.1 software (Rouillard *et al.*, 2003), and Custom 90K CombiMatrix arrays were prepared as described by Bellin *et al.* (2009).

One microgram of the total RNA from each of *Petunia axillaris* and *inflata* organs, and of RNA from the R27 and W134 seeds, was used for each array hybridization. RNA labeling, hybridization reaction, data extraction and quantile data normalization were performed as described in Bellin *et al.* (2009). Expression data are available from the NCBI [Accession Number: GSE28120].

Pearson correlation of three biological replicates for each sample ranged from 0.96 to 0.99. A gene was considered expressed when the probe signal was higher than threshold, calculated as the 97th percentile of the intensities from the negative controls, in at least two replicates.

Principal Component Analysis was performed by using SIMCA P+ (Umetrics, http://www.umetrics.com). Differentially expressed genes were defined by multiclass comparison method of Significance Analysis of Microarray (Tusher *et al.*, 2001), with FDR = 5%. To define 15 expression clusters, *P. axillaris* and *P. inflata* differentially expressed genes were analysed together; clusters were obtained by the k-means method, using Pearson correlation (T-Mev, 4.3). Genes expressed differentially between R27 and W134 pooled seeds were determined by SAM unpaired method (T-Mev, 4.3), with a FDR = 0.125%, for both *P. axillaris* and *P. inflata* chip arrays.

Transcriptome comparison

P. inflata and P. axillaris TCs were combined and organized into 'super-assemblies' by using the CAP3 assembler (85% sequence identity and 60 nucleotide overlap; Huang and Madan, 1999). Pairs of TCs were determined by reciprocal BLASTn analysis at E-value < 10^{-10}. The alignments were filtered by fixing an identity cut off equal to 90% and a match length cutoff to 75% of the subject length.

Scanning electron microscopy

Mature seeds were fixed, critical point dried, and imaged with a scanning electron microscope as described by Souer *et al.* (1996).

Seed staining and PA analysis

To determine the presence of PAs in the seed coat, mature seeds were stained with dimethylaminocinnamaldehyde (DMACA) as described by Pang *et al.* (2009). For PA analysis, powdered seeds were extracted with 0.2 ml of 1% HCl/methanol by vortexing, sonication for 30 min and standing overnight at room temperature. Following centrifugation at 3000 ***g*** for 10 min, total soluble PA content in the supernatant was spectrophotometrically determined after reaction with DMACA reagent (0.1% [w/v] DMACA in methanol – 3 N HCl) at 640 nm, with (+)-catechin as standard.

LC–MS analysis

Powdered mature seeds of line R27 and W134 were extracted with 10 volumes of methanol containing 0.1% HCl. The samples were sonicated for 20 min and then placed at room temperature with constant agitation O/N. After a centrifugation at 2900 ***g*** for 30 min, the supernatants were filtered and transferred to liquid chromatography vials.

The HPLC (high pressure liquid chromotography) system and analysis and the mass spectrometer parameters are described in Toffali *et al.* (2011).

Metabolites were identified by comparing the retention times, *m/z* values and fragmentation patterns (MS/MS and MS3) with an in house library obtained analyzing commercial standards. When no commercial standards were available, fragmentation patterns were compared with those reported in the literature.

Real-time RT-PCR

cDNA synthesis and real-time PCR analysis were performed as described by Zenoni *et al.* (2010). Each expression value, relative to actin was determined in triplicate. The list of gene-specific primers, designed on the 3′ untranslated gene region is reported in Table S5. The PCR involved a 95°C hold for 10 min followed by 40 cycles at 95°C for 30 sec, 55°C for 30 sec, and 72°C for 20 sec. The expression *ratio* value was calculated for organ samples relative to the fertilized ovary (O) while for R27 and W134 seeds relative to the R27, according to the Pfaffl (2001) and Pfaffl *et al.* (2002).

ACKNOWLEDGEMENTS

This work was supported by: Fondazione Cariverona (Completamento e attività del Centro di genomica Funzionale Vegetale), Verona, Italy; Dave Clark laboratory (Department of Environmental Horticulture, University of Florida, USA) and Florida Agricultural Experiment Station (USA). Nunzio D'Agostino was supported by the Agronanotech Project, MIPAF. Special thanks to Alberto Ferrarini for his support in the chip design.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. Annotation workflow followed for TCs annotation.

Figure S2. Real-time RT-PCR validation of the organ expression profiles of four genes.

Figure S3. Real-time RT-PCR validation of six differentially expressed genes between R27 and W134 seeds.

Table S1. Number of EST/mRNA sequences from petunia species available in GenBank.

Table S2. TCs and sESTs that match the known petunia transcriptome.

Table S3. Microarray hybridization validates the expression pattern of candidate genes for use in normalization expression analysis.

Table S4. Count of TCs belonging to the 15 expression clusters.

Table S5. List of primers used for real-time RT-PCR analyses.

Dataset S1. All TC IDs with the best hits from diverse BLAST searches.

Dataset S2. All Probes (PROBE ID) with the expression pattern description.

Dataset S3. All TC IDs with matching counterpart(s).

Dataset S4. Differentially expressed TCs in *an11* mutant seeds displaying a two-fold or greater change in transcript abundance compared to the wild type, in *P. axillaris* and *P. inflata.*

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

REFERENCES

Alagna, F., D'Agostino, N., Torchia, L., Servili, M., Rao, R., Pietrella, M., Giuliano, G., Chiusano, M.L., Baldoni, L. and Perrotta, G. (2009) Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics*, **10**, 399.

Angenent, G.C., Stuurman, J., Snowden, K.C. and Koes, R. (2005) Use of *Petunia* to unravel plant meristem functioning. *Trends Plant Sci.* **10**, 243–250.

Arsovski, A.A., Popma, T.M., Haughn, G.W., Carpita, N.C., McCann, M.C. and Western, T.L. (2009) AtBXL1 encodes a bifunctional beta-D-xylosidase/alpha-L-arabinofuranosidase required for pectic arabinan modification in Arabidopsis mucilage secretory cells. *Plant Physiol.* **150**, 1219–1234.

Baudry, A., Heim, M.A., Dubreucq, B., Caboche, M., Weisshaar, B. and Lepiniec, L. (2004) TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J.* **39**, 366–380.

Baxter, I.R., Young, J.C., Armstrong, G., Foster, N., Bogenschutz, N., Cordova, T., Peer, W.A., Hazen, S.P., Murphy, A.S. and Harper, J.F. (2005) A plasma membrane H+-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **102**, 2649–2654.

Bellin, D., Ferrarini, A., Chimento, A., Kaiser, O., Levenkova, N., Bouffard, P. and Delledonne, M. (2009) Combining next-generation pyrosequencing with microarray for large scale expression analysis in non-model species. *BMC Genomics*, **10**, 555.

Bhuvanagiri, M., Schlitter, A.M., Hentze, M.W. and Kulozik, A.E. (2010) NMD: RNA biology meets human genetic medicine. *Biochem. J.* **430**, 365–377.

Breuillin, F., Schramm, J., Hajirezaei, M. *et al.* (2010) Phosphate systemically inhibits development of arbuscular mycorrhiza in *Petunia hybrida* and represses genes involved in mycorrhizal functioning. *Plant J.* **64**, 1002–1017.

Carey, C.C., Strahle, J.T., Selinger, D.A. and Chandler, V.L. (2004) Mutations in the pale aleurone color1 regulatory gene of the *Zea mays* anthocyanin pathway have distinct phenotypes relative to the functionally similar *TRANSPARENT TESTA GLABRA1* gene in *Arabidopsis thaliana*. *Plant Cell*, **16**, 450–464.

Cartolano, M., Castillo, R., Efremova, N., Kuckenberg, M., Zethof, J., Gerats, T., Schwarz-Sommer, Z. and Vandenbussche, M. (2007) A conserved microRNA module exerts homeotic control over *Petunia hybrida* and *Antirrhinum majus* floral organ identity. *Nat. Genet.* **39**, 901–905.

Chen, J.C., Jiang, C.Z. and Reid, M.S. (2005) Silencing a prohibitin alters plant development and senescence. *Plant J.* **44**, 16–24.

D'Agostino, N., Aversano, M. and Chiusano, M.L. (2005) ParPEST: a pipeline for EST data analysis based on parallel computing. *BMC Bioinformatics*, **6**(Suppl 4), S9.

Fiers, M., Ku, K.L. and Liu, C.M. (2007) CLE peptide ligands and their roles in establishing meristems. *Curr. Opin. Plant Biol.* **10**, 39–43.

Gerats and Vandenbussche, M. (2005) A model system for comparative research: Petunia. *Trends Plant Sci.*. **10**, 251–256.

Gerats, T. and Strommer, J. (2009) *Petunia, Evolutionary, Developmental and Physiological Genetis*, 2nd edn. Berlin: Springer-Verlag.

Hazelhurst, S., Hide, W., Liptak, Z., Nogueira, R. and Starfield, R. (2008) An overview of the wcd EST clustering tool. *Bioinformatics*, **24**, 1542–1546.

van Houwelingen, A., Souer, E., Spelt, K., Kloos, D., Mol, J. and Koes, R. (1998) Analysis of flower pigmentation mutants generated by random transposon mutagenesis in *Petunia hybrida*. *Plant J.* **13**, 39–50.

Huang, X. and Madan, A. (1999) CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–877.

Ishida, T., Kurata, T., Okada, K. and Wada, T. (2008) A genetic regulatory network in the development of trichomes and root hairs. *Annu. Rev. Plant Biol.* **59**, 365–386.

Kitamura, S., Matsuda, F., Tohge, T., Yonekura-Sakakibara, K., Yamazaki, M., Saito, K. and Narumi, I. (2010) Metabolic profiling and cytological analysis of proanthocyanidins in immature seeds of *Arabidopsis thaliana* flavonoid accumulation mutants. *Plant J.* **62**, 549–559.

Koes, R., Verweij, W. and Quattrocchio, F. (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* **10**, 236–242.

Lepiniec, L., Debeaujon, I., Routaboul, J.M., Baudry, A., Pourcel, L., Nesi, N. and Caboche, M. (2006) Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* **57**, 405–430.

Mallona, I., Lischewski, S., Weiss, J., Hause, B. and Egea-Cortines, M. (2010) Validation of reference genes for quantitative real-time PCR during leaf and flower development in *Petunia hybrida*. *BMC Plant Biol.* **10**, 4.

Matilla, A.J. and Matilla-Vázquez, M.A. (2008) Involvement of ethylene in seed physiology. *Plant Sci.* **175**, 87–97.

Mortimer, J.C., Laohavisit, A., Macpherson, N., Webb, A., Brownlee, C., Battey, N.H. and Davies, J.M. (2008) Annexins: multifunctional components of growth and adaptation. *J. Exp. Bot.* **59**, 533–544.

Ohashi, Y., Oka, A., Rodrigues-Pousada, R., Possenti, M., Ruberti, I., Morelli, G. and Aoyama, T. (2003) Modulation of phospholipid signaling by GLABRA2 in root-hair pattern formation. *Science*, **300**, 1427–1430.

Okamoto, S., Ohnishi, E., Sato, S., Takahashi, H., Nakazono, M., Tabata, S. and Kawaguchi, M. (2009) Nod factor/nitrate-induced CLE genes that drive HAR1-mediated systemic regulation of nodulation. *Plant Cell Physiol.* **50**, 67–77.

Pang, Y., Peel, G.J., Sharma, S.B., Tang, Y. and Dixon, R.A. (2008) A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of *Medicago truncatula*. *Proc. Natl Acad. Sci. USA*, **105**, 14210–14215.

Pang, Y., Wenger, J.P., Saathoff, K. *et al.* (2009) A WD40 repeat protein from Medicago truncatula is necessary for tissue-specific anthocyanin and proanthocyanidin biosynthesis but not for trichome development. *Plant Physiol.* **151**, 1114–1129.

Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45.

Pfaffl, M.W., Horgan, G.W. and Dempfle, L. (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30**, e36.

Quattrocchio, F., Wing, J.F., Leppen, H., Mol, J. and Koes, R.E. (1993) Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. *Plant Cell*, **5**, 1497–1512.

Quattrocchio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J. and Koes, R. (2006) PH4 of Petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell*, **18**, 1274–1291.

Ranocha, P., Denance, N., Vanholme, R. *et al.* (2010) Walls are thin 1 (WAT1), an Arabidopsis homolog of *Medicago truncatula* NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *Plant J.* **63**, 469–483.

Rouillard, J.M., Zuker, M. and Gulari, E. (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res.* **31**, 3057–3062.

Shirley, B.W., Kubasek, W.L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F.M. and Goodman, H.M. (1995) Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. *Plant J.* **8**, 659–671.

Souer, E., van Houwelingen, A., Kloos, D., Mol, J. and Koes, R. (1996) The no apical meristem gene of Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell*, **85**, 159–170.

Spelt, C., Quattrocchio, F., Mol, J. and Koes, R. (2002) ANTHOCYANIN1 of petunia controls pigment synthesis, vacuolar pH, and seed coat development by genetically distinct mechanisms. *Plant Cell*, **14**, 2121–2135.

Stehmann, J.R., Lorenz-Lemke, A.P., Freitas, L.B. and Semir, J. (2009) The Genus *Petunia*. *In Petunia*, 2nd edn (Gerats, T. and Strommer, J., eds). Berlin: Springer-Verlag, pp. 1–28.

Toffali, K., Zamboni, A., Anesi, A., Stocchero, M., Pezzotti, M., Levi, M. and Guzzo, F. (2011) Novel aspects of grape berry ripening and post-harvest withering. *Metabolomics*, in press (DOI 10.1007/s11306–010–0259-y).

Tornielli, G., Koes, R. and Quattrocchio, F. (2009) The genetics of flower color. *In Petunia*, 2nd edn (Gerats, T. and Strommer, J., eds). Berlin: Springer-Verlag, pp. 269–299.

Tusher, V.G., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA*, **98**, 5116–5121.

Vandenbussche, M., Janssen, A., Zethof, J. *et al.* (2008) Generation of a 3D indexed *Petunia* insertion database for reverse genetics. *Plant J.* **54**, 1105–1114.

Verweij, C.W. (2007) *Vacuolar Acidification: Mechanism, Regulation and Function in Petunia flowers*. Amsterdam: VU-University, pp. 136.

Verweij, W., Spelt, C., Di Sansebastiano, G.P., Vermeer, J., Reale, L., Ferranti, F., Koes, R. and Quattrocchio, F. (2008) An H^+ P-ATPase on the tonoplast determines vacuolar pH and flower colour. *Nat. Cell Biol.* **10**, 1456–1462.

de Vetten, N., Quattrocchio, F., Mol, J. and Koes, R. (1997) The *an11* locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. *Genes Dev.* **11**, 1422–1434.

Zenoni, S., Ferrarini, A., Giacomelli, E., Xumerle, L., Fasoli, M., Malerba, G., Bellin, D., Pezzotti, M. and Delledonne, M. (2010) Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. *Plant Physiol.* **152**, 1787–1795.