Supplementary Note 1

Assembly and annotation of the *Petunia* genomes.

Aureliano Bombarely [1], Michel Moser [2], Noe Fernandez-Pozo [3], Lukas A. Mueller [3], Cris Kuhlemeier [2] and Thomas L. Sims[4]

Affiliations

1- Department of Horticulture, Virginia Tech, Blacksburg, VA 24061 USA.
2- University of Bern, Altenbergrain 21, CH-3013 Bern, Switzerland.
3- Boyce Thompson Institute for Plant Research, 533 Tower Rd, Ithaca, NY 14853-1801, USA.
4- Department of Biological Sciences, Northern Illinois University, DeKalb, IL 60115-2861, USA.

EXTENDED MATERIAL AND METHODS FOR GENOME ASSEMBLY AND ANNOTATION

Plant Material and DNA preparation

Petunia inflata S6 was obtained from the University of Florida in 1974 and maintained by inbreeding (for >15 generations at Free University of Amsterdam and University of Amsterdam). *Petunia axillaris* N was provided by the Botanical Garden of Rostock to the Botanical Garden of Bern and deposited in the Amsterdam collection under the designation of *P. axillaris* S26. Plants used for DNA extraction were grown axenically in tissue culture containers. Mature plants (leaves and stems) were harvested, flash-frozen in liquid nitrogen and stored at -80C until used for DNA extractions. Plant material (approximately 15 g) was extracted using a modification of methods designed to isolate high molecular weight DNA from nuclei (Fischer and Goldberg, 1982; Carrier et al., 2011). Briefly, frozen plant material was homogenized in a blender with liquid nitrogen until a fine powder was obtained. Powdered material was thawed in 1X SEB plus mercaptoethanol (10 mM Tris pH 8.0, 100 mM KCl, 10 mM Na_2EDTA, 0.5 M sucrose, 4 mM spermidine, 1 mM spermine, 0.13% carbamic acid, 0.25% PVP-40, 0.2% β-mercaptoethanol), then filtered through Nitex mesh. Triton X-100 was added to 0.5% and nuclei isolated and washed by repeated low speed centrifugation and washing with SEB. Nuclei were lysed by adding an equal volume of NLB (2% Sodium N-lauryl sarcosine, 40 mM Na_2EDTA, 0.1 M Tris-HCl pH 8.0 and 1mg/ml proteinase K) followed by incubation at 55 C for 1 hour. Cesium chloride was added to 50% w/w along with ethidium bromide to a final concentration of 0.4%. DNA gradients were centrifuged in a 70.1 Ti rotor at 40,000 rpm for 36 hours followed by re-banding in a Vti65.2 rotor at 60,000 rpm for 6 hours. Ethidium bromide was removed by extraction with SSC-saturated isopropanol and the remaining solution dialyzed against TNE (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.1 mM EDTA) for 24 hours. DNA was precipitated, washed with 70% ethanol, dried and resuspended in EB (10 mM Tris pH 8.0) to a final concentration of 150 µg/ml.

For the PacBio sequencing the DNA was prepared in a different manner. Briefly, genomic DNA from *Petunia axillaris* subsp. *axillaris* N was extracted from sterile plant cultures. The plants were grown at 16 h photoperiod and constant 20 °C. Young folded leaves from 3 plants were collected and ground in liquid nitrogen. Genomic DNA was extracted using a CTAB protocol (Clarke, 2009) and DNA was treated with RNase digestion and an extra ethanol washing step before library preparation.

For the BAC PacBio sequencing, the BAC DNA from a *Petunia axillaris* subsp. *axillaris* N library (in house, unpublished) was isolated with columns from the QIAGEN Large-Construct kit (QIAGEN Redwood City, CA, USA) following the manual supplied by the QIAGEN Large-Construct kit.

DNA Sequencing

Illumina library preparation and sequencing

Illumina reads were produced in two phases. In the first one, BGI-Zhenzhen prepared four pair-ends libraries with insert sizes of 170, 350, 500 and 800 bp and two mate-pair libraries with insert sizes of 2 and 5 Kb with TruSeq library preparation kit (Illumina, San Diego, CA, USA) and sequenced in a HiSeq1000 system (2x100). In the second one, three more libraries were prepared with the Nextera DNA library prep. kit (Illumina, San Diego, CA, USA) and sequenced at University of Illinois Roy J. Carver Biotechnology Center, one pair-end library with an insert size of 1 Kb, sequenced as 2x150 in a HiSeq2000 Illumina system and two mate-pair libraries with insert sizes of 8 and 15 Kb sequenced as 2x100 in a HiSeq2500 Illumina system **(Supplementary table 1)**.

Petunia axillaris PacBio library preparation and sequencing

High molecular weight genomic DNA was sheared in a Covaris g-TUBE (Covaris, Woburn, MA, USA) to obtain 18 Kb fragments. After shearing the DNA size distribution was checked on a DNA Fragment Analyzer (Advanced Analytical Technologies, Ames, IA, USA). The sheared DNA was used to prepare a SMRTbell library with the PacBio DNA Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA,

USA) according to the manufacturer's recommendations. The recovered library was sequenced with P4/C2 chemistry and MagBeads on a PacBio RSII system (Pacific Biosciences, Menlo Park, CA, USA) at 120 min movie length. The library was sequenced with a total of 62 SMRT Cells. Sequencing produced 7,501,054 reads longer than 500 bp and a total of 21x coverage (29 .1 Gb). The average read length was 3,888 bp and maximal read length reached 24,597 bp.

Petunia axillaris BAC Library Preparation and Sequencing
BAC DNA from 7 BAC clones and a total of 40 microgram genomic DNA was sent on ice to the Lausanne Genomic Technologies Facility (LGTF) for library preparation and sequencing. Per BAC clone, a single library was prepared. High molecular weight DNA from BACs was sheared in a Covaris g-TUBE (Covaris, Woburn, MA, USA) to obtain 10-20 Kb fragments. After shearing the DNA size distribution was checked on a DNA Fragment Analyzer (Advanced Analytical Technologies, Ames, IA, USA). The sheared DNA was used to prepare a SMRTbell library with the PacBio DNA Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA) according to the manufacturer's recommendations. The recovered library was sequenced with XL/C2 chemistry and MagBeads on a PacBio RSII system (Pacific Biosciences, Menlo Park, CA, USA) at 120 min movie length using one SMRT Cell per BAC clone.

Read processing, sequence assembly and draft genome evaluation
Read processing
Illumina reads were processed in three different steps: 1) Quality filtering and adapter removal using Fastq-mcf (Aronesty, 2013) with a minimum quality score of 30 (-q 30) and a minimum length of 50 bp (-l 50). 2) Read pair duplications filtering using Prinseq (Schmieder and Edwards, 2011). 3) Read error correction using Musket with the default parameters (Liu et al., 2013). Pacbio reads were processed using the SMRT Analysis pipeline (v.2.0.1). PacBio reads were filtered for a minimum length of 500 bp and minimum quality score of 0.70.

Genome assembly
Illumina reads were assembled using SOAPdenovo2 (Luo et al., 2012) with different k-mer sizes (23, 31, 39, 47, 55, 63, 71, 79, 87). Different assembly results were compared using its corresponding total assembly size, N50/L50 and N90/L90, selecting the assembly with the best stats (kmer=79 for both species, *Petunia axillaris* N and *P. inflata* S6). Gaps between contigs were completed using GapCloser, from the SOAPdenovo2 package with the default parameters (Luo et al., 2012). For the *P. axillaris* assembly, where Pacbio reads were also available, the assembly was broken in its contigs using a Perl script, BreakScaffolds (https://github.com/aubombarely/GenoToolBox/blob/master/SeqTools/BreakScaffolds) and then a new assembly was performed using AHA (Bashir et al., 2012) to combine Pacbio reads and the contigs produced by the Illumina read assembly (4 iterations using only reads longer than 3000 bp). Gaps produced by the hybrid assembly were filled using PBJelly (English et al., 2012b), and then new scaffolding using the Illumina pair ends and mate pairs was performed using SSPACE (Boetzer et al., 2011). Finally a last round of gap filling was performed using PBJelly (English et al., 2012a).

Genome size estimation
Genome size estimation was performed through the k-mers abundance distribution (Li et al., 2010) with the Illumina 1000 paired end libraries. Jellyfish was used to count k-mers (Marçais and Kingsford, 2011) with a kmer of 31. K-mer distribution was analyzed with R.

Organelle genomes
Chloroplastic and mitochondrial PacBio sequences from *P. axillaris* were filtered out using the chloroplast genomes of S*olanum lycopersicum* LA3032 (GI:84371962) and *Nicotiana tabacum* (GI:76559634) and mitochondrial DNA of *Nicotiana tabacum* (GI:56806513) with the alignment tool BLASR (Chaisson and Tesler, 2012). Filtered reads were assembled separately using HGAP (Chin et al., 2013). The chloroplast genome of *P. axillaris* contains 158,794 bp and was assembled in a single

contig with a GC content of 37.7 %. The mitochondrial genome of *P. axillaris* has a total length of 440,114 bp, fragmented over 10 contigs.

Petunia axillaris BAC Clone Assembly
De novo assembly of the P. *axillaris* BAC inserts was performed using the SMRT Analysis (v.2.0.1) pipeline. The complete BAC clones were assembled with HGAP followed by consensus calling with Quiver (Chin et al., 2013). The resulting contig was blasted against the vector sequence (pCC1BAC) and the identified vector sequence was cut out to retain the BAC insert. PacBio reads were mapped back to the BAC insert using BLASR (Chaisson and Tesler, 2012). Alignments were reviewed for possible miss-assemblies by visualizing read depth in IGV (Robinson et al., 2011). The 7 BAC clones could be assembled into single contigs each of sizes between 90 Kb and 155 Kb. As 2 pairs of BAC clones were extracted from the BAC library using the same marker, their sequence could be overlaid and merged. The final 5 BAC clone sequences had an average length of 170,478 bp and a total length of 852,393 bp.

Genome quality assessment
Gene space completeness of the assembly was evaluated running CEGMA v2.5 (Parra et al., 2007) with the default parameters. We mapped the 248 Core Eukaryotic Genes (CEGs) to assess the completeness of both assemblies and found 239 (94%) and 243 (98 %) in the assembly of *P. axillaris* and *P. inflata*, respectively.

BAC sequences were used to evaluate genome accuracy of P*. axillaris* assembly. Similarity searches were conducted using BLASTN (McGinnis and Madden, 2004) in ungapped mode with an E-value cut-off of 1E-10. The percent nucleotide identity was calculated by matcher from EMBOSS tools version 2.0u4 (www.ebi.ac.uk/Tools/psa/emboss_matcher/nucleotide.html). The five BAC sequences mapped against the *P. axillaris* scaffolds showed an average similarity of 95 % (Supplementary figure 1).

Heterozygosity Evaluation
The heterozygosity of the sequenced genomes was estimated mapping the Illumina reads to the assembly using Bowtie2 (Langmead and Salzberg, 2012), calling SNPs using FreeBayes (Garrison and Marth, 2012) with a minimum mapping score of 20 and a minimum read depth of 5. SNPs were annotated using SnpEff (Cingolani et al., 2014). Heterozygosity was calculated following two methodologies: A- Number of heterozygous variants divided by the whole genome sequence with a minimum read depth of 5 (total heterozygosity). The total heterozygosity was 0.27 and 0.34 variants/Kb (0.03%) for *P. axillaris* N and *P. inflata* S6 genomes respectively; B- Using Vcftools (Danecek et al. 2011) to calculate the variants/Kb in 10 Kb genomic bins (119,652 *P. axillaris* and 124,348 *P. inflata* bins). In this case, regions with read mapping depth below 5 and over 2 times the read mapping depth mean (56.7 for *P. axillaris* and 52.7 for *P. inflata*) were removed from analysis to avoid possible paralog regions collapsing problems yielding 117,797 and 122,481 bins respectively. The estimated heterozygosity calculated by 10 Kb bins was 0.25 and 0.37 variants/Kb (0.03 and 0.04 %) respectively. The distribution was represented in the Supplementary figure 2.

Genome annotation
Structural annotation
The genome structural annotation was performed using Maker (Cantarel et al., 2007). This program creates the gene model annotation using *ab-initio* gene predictions programs such as Augustus (Stanke et al., 2006) and SNAPP (Korf, 2004) integrating this results with experimental data such as ESTs, RNAseq and protein alignments. Augustus was trained with 400 *P. axillaris* N gene models curated manually using Web Apollo (Lee et al., 2013), experimental data from 454 (Zenoni et al., 2011) and Illumina RNASeq reads from several tissues and developmental stages (Supplementary table 2), and compared with protein alignments from the tomato genome project, ITAG2.4 (Tomato Genome Consortium, 2012) and SwissProt Solanaceae proteins set (Magrane and Consortium, 2011).

RNAseq Illumina data was mapped using Tophat2 and assembled with Cufflinks (Trapnell et al., 2009). The same datasets were also used as experimental supporting data for Maker. Additionally Maker annotates the repeat content. A previous repeat analysis was done using RepeatModeler (http://www.repeatmasker.org/RepeatModeler.html). tRNA were annotated using tRNAscan (Lowe and Eddy, 1997). After the automatic annotation, more than 500 gene models were structurally curated using Web Apollo (Lee et al., 2013) including the genes described in the different sections of this publication. The resulting annotation contained a total of 35,812 and 39,408 genes, thereof 32,928 and 36,697 being protein-coding genes for *P. axillaris* and *P. inflata*, respectively (Supplementary table 3). As non-coding RNAs, 546 rRNAs and 2884 tRNAs and 558 rRNAs and 2711 tRNAs were predicted for *P. axillaris* and *P. inflata*, respectively.

Functional annotation

The gene functional annotation was performed in two steps: 1) Functional description association by sequence homology search from different sources such as protein datasets from SwissProt (Magrane and Consortium, 2011), arabidopsis genome annotation version TAIR10 (Lamesch et al., 2012) and tomato genome annotation version ITAG2.3 (Tomato Genome Consortium, 2012) and GenBank (Benson et al., 2013) using BlastX (McGinnis and Madden, 2004) and protein domains using InterProScan (Mulder and Apweiler, 2007). 2) Functional annotation filtering, scoring and integration using AHRD (https://github.com/groupschoof/AHRD). Swissprot, TAIR10, ITAG2.3 and GenBank were scored 100, 50, 50 and 30 respectively. 58, 73 and 88% of the proteins showed at least one hit with a minimum e-value of 1e-20 with Swissprot, TAIR10 and GenBank respectively. 83% presented at least one protein domain from the InterPro database. Gene Ontology (GO) annotations were inferred from the protein domain hits.

Supplementary table 1

Sequencing machine	Insert-size [bp]	Read length / type	P. axillaris Processed reads coverage (X)*	P. inflata Processed reads coverage (X)*
Illumina HiSeq1000	170	2x100 bp / paired-end	17	22
Illumina HiSeq1000	350	2x100 bp / paired-end	12	6
Illumina HiSeq1000	500	2x100 bp / paired-end	10	25
Illumina HiSeq1000	800	2x100 bp / paired-end	12	15
Illumina HiSeq1000	2000	2x100 bp / mate-pair	11	15
Illumina HiSeq1000	5000	2x100 bp / mate-pair	8	20
Illumina HiSeq2500	1000	2x150 bp / paired-end	48	NA
Illumina HiSeq2500	8000	2x150 bp / mate-pair	10	17
Illumina HiSeq2500	15000	2x150 bp / mate-pair	9	15
Total short reads:	-	-	137	135
PacBio RS 2	-	-	21	NA

*see section Read processing for details about read processing

Supplementary table 2

Species	Tissue/ Develop. Stage	Treatment	Reads	SRA Accession	Publication
P. axillaris[1]	Mixed tissues	NA	578,107	SRX036998	Zenoni et al. 2011
P. axillaris	Floral buds	NA	34,150,939	SRX1402727	Sheehan et al. 2015
P. axillaris	Floral buds	NA	44,731,386	SRX1402609	Sheehan et al. 2015
P. axillaris	Floral buds	NA	52,405,543	SRX1402587	Sheehan et al. 2015
P. axillaris	Trichomes	NA	23,620,517	SRX710299	Guo et al. 2015
P. axillaris	Seedling	NA	32,715,526	SRX710174	Guo et al. 2015
P. axillaris	Apical shoots	NA	35,221,375	SRX709906	Guo et al. 2015
P. axillaris	Mature flower	NA	27,178,746	SRX709957	Guo et al. 2015
P. axillaris	Callus	NA	28,505,785	SRX710298	Guo et al. 2015
P. integrifolia[1]	Mixed tissues	NA	602,753	SRX036999	Zenoni et al. 2011
P. integrifolia	Trichomes	NA	23,957,225	SRX711430	Guo et al. 2015
P. integrifolia	Seedling	NA	25,099,736	SRX711429	Guo et al. 2015
P. integrifolia	Apical shoots	NA	30,675,565	SRX711427	Guo et al. 2015
P. integrifolia	Mature flower	NA	32,220,783	SRX711426	Guo et al. 2015
P. integrifolia	Callus	NA	28,961,576	SRX711385	Guo et al. 2015
P. inflata	Mixed tissues[5]	NA	83,739,290	In preparation	No published
P. inflata	Mixed tissues[5]	NA	41,077,858	In preparation	No published
P. inflata	Mixed tissues[5]	NA	71,909,258	In preparation	No published
P. inflata	Mixed tissues[5]	NA	71,334,072	In preparation	No published
P. inflata	Mixed tissues[5]	NA	92,583,881	In preparation	No published
P. hybrida[2]	Mature leaves	Control NaCl 0h	14,299,136	SRX1530795	Villarino et al. 2014
P. hybrida[2]	Mature leaves	Control NaCl 0h	23,084,764	SRX1530797	Villarino et al. 2014
P. hybrida[2]	Mature leaves	Control NaCl 0h	18,140,264	SRX1530799	Villarino et al. 2014
P. hybrida[2]	Mature leaves	Control NaCl 6h	21,904,108	SRX1530801	Villarino et al. 2014
P. hybrida[2]	Mature leaves	Control NaCl 6h	13,362,088	SRX1530803	Villarino et al. 2014
P. hybrida[2]	Mature leaves	Control NaCl 6h	14,459,178	SRX1530805	Villarino et al. 2014
P. hybrida[2]	Mature leaves	Control NaCl 24h	12,733,646	SRX1530807	Villarino et al. 2014
P. hybrida[2]	Mature leaves	Control NaCl 24h	14,364,874	SRX1530809	Villarino et al. 2014
P. hybrida[2]	Mature leaves	Control NaCl 24h	16,181,450	SRX1530811	Villarino et al. 2014
P. hybrida[2]	Mature leaves	150 mM NaCl 6h	16,771,012	SRX1530813	Villarino et al. 2014
P. hybrida[2]	Mature leaves	150 mM NaCl 6h	16,458,846	SRX1530814	Villarino et al. 2014
P. hybrida[2]	Mature leaves	150 mM NaCl 6h	18,340,684	SRX1530815	Villarino et al. 2014
P. hybrida[2]	Mature leaves	150 mM NaCl 24h	14,368,884	SRX1530822	Villarino et al. 2014
P. hybrida[2]	Mature leaves	150 mM NaCl 24h	16,404,210	SRX1530823	Villarino et al. 2014
P. hybrida[2]	Mature leaves	150 mM NaCl 24h	16,508,928	SRX1530825	Villarino et al. 2014
P. hybrida[3]	Mature petals	NA	15,259,555	In preparation	In preparation
P. hybrida[3]	Mature petals	NA	9,829,522	In preparation	In preparation
P. hybrida[3]	Mature petals	NA	11,170,745	In preparation	In preparation
P. hybrida[3]	Mature petals	NA	11,523,112	In preparation	In preparation
P. hybrida[3]	Mature petals	NA	13,949,735	In preparation	In preparation
P. hybrida[3]	Mature petals	NA	8,617,423	In preparation	In preparation
P. hybrida[3]	Mature petals	NA	11,569,546	In preparation	In preparation
P. hybrida[3]	Mature petals	NA	15,817,998	In preparation	In preparation
P. hybrida[4]	Mature petals	NA	13,060,169	In preparation	In preparation
P. hybrida[4]	Mature petals	NA	11,883,830	In preparation	In preparation

Notes: All the sequenced libraries were Illumina HiSeq2000 Pair Ends except for (1) that were 454 libraries publically available at NCBI SRA. The Petunia hybrida sequenced cultivars were Mitchell (2), R27 (3) and R143 (4). For more details about the plant growth, RNA extraction, library preparation and sequencing, consult the corresponding publication. For the libraries were no publication was specified, the plants were growth in standard conditions of light and temperature (long day, 22C). RNA was extracted with RNeasy Qiagen kit. (5) For library produced with mixed tissues, RNA for the most representative tissues (leaves, roots, flowers, stems…) were mixed to maximize the gene space sequenced.

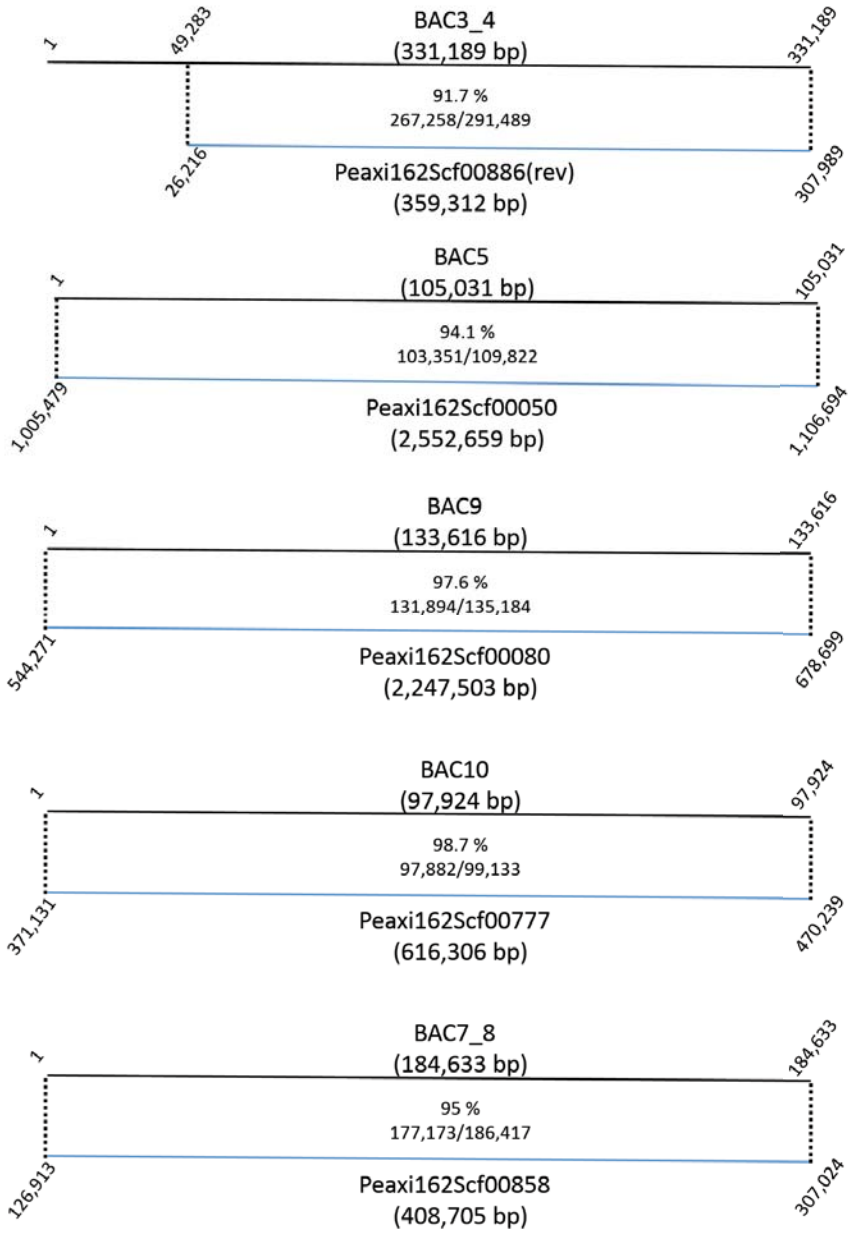
Supplementary table 3

Protein-coding gene statistics	*P. axillaris*	*P. inflate*	*S. lycopersicum*
Gene number	32,928	36,697	34,725
Average gene length [bp]	4252	4152	3162
Exon number	173,712	188,372	160,001
Average exon number per gene	5.27	5.13	4.61
Average exon length [bp]	238.2	238.0	261.4
Total exon length [bp]	41,387,658	44,837,367	41,821,567
Total CDS length [bp]	38,480,559	42,303,603	35,813,852
Average exonic length per gene [bp]	1256.9	1221.8	1204.4
Intron number	138,743	150,011	125,276
Total intron length [bp]	98,287,394	106,892,756	67,748,290
Average intron length [bp]	708.4	712.6	540.8

Protein coding gene statistics of the two Petunia assemblies in comparison with S. lycopersicum (ITAG 2.4)

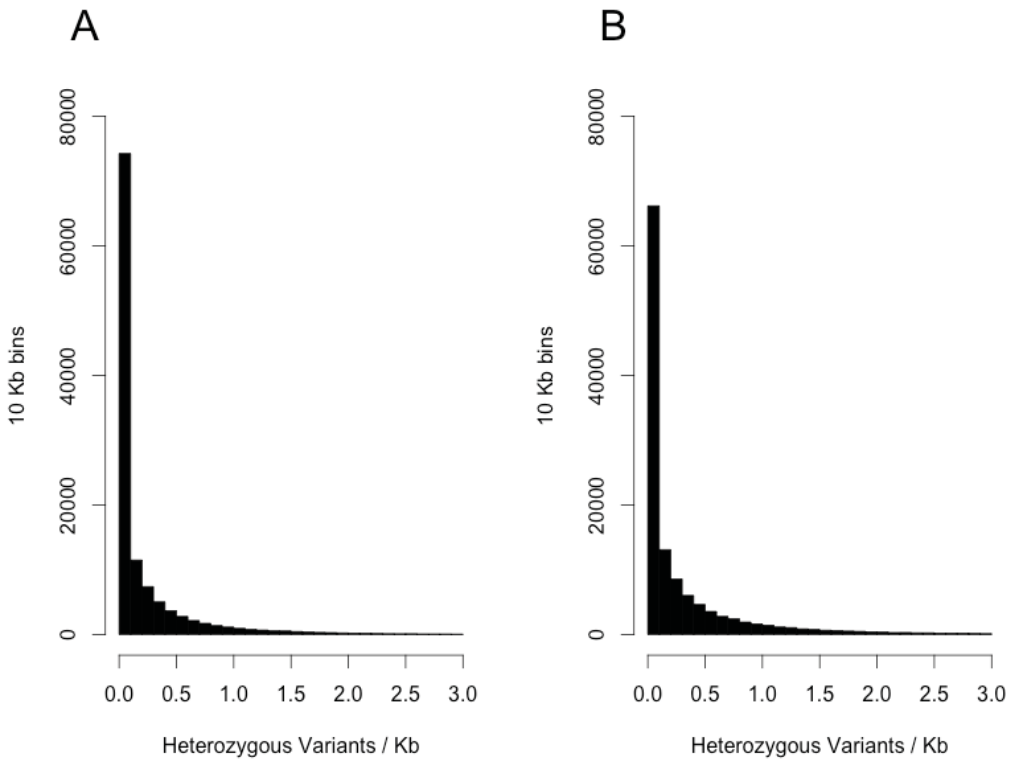
Supplementary Figure 1

BAC sequences were mapped against the scaffolds of the *P. axillaris* assembly using BLAST v. Black lines and blue lines indicate the BAC inserts and the *P. axillaris* scaffolds, respectively. Numbers along the lines indicate the start and end of sequences and their overlapping regions which are connected by dotted lines. Percent nucleotide identity for each homologous region is indicated along with the length of the matching basepairs out of the total alignment length.



Supplementary Figure 2

Heterozygosity distribution for 10 Kb genomic bins for: A- 117,797 bins for *P. axillaris N* genome assembly v1.6.2 (Variants/Kb mean = 0.25); B- 122,481 bins for *P. inflata* S6 genome assembly v1.0.1 (Variants/Kb mean = 0.37).



REFERENCES

Aronesty, E. (2013). Comparison of sequencing utility programs. Open Bioinform J. 7:1-8

Bashir, A. et al. (2012). A hybrid approach for the automated finishing of bacterial genomes. Nat Biotechnol **30**: 701–707.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013). GenBank. Nucleic Acids Res **41**: D36–42.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27**: 578–579.

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sanchez Alvarado, A., and Yandell, M. (2007). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res **18**: 188–196.

Carrier, G., Santoni, S., Rodier-Goud, M., Canaguier, A., Kochko, A. de, Dubreuil-Tranchant, C., This, P., Boursiquot, J.-M., and Le Cunff, L. (2011). An efficient and rapid protocol for plant nuclear DNA preparation suitable for next generation sequencing methods. Am J Bot **98**: e13–5.

Chaisson, M.J. and Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. Bmc Bioinformatics **13**: 238.

Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W., and Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Meth **10**: 563–569.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2014). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) **6**: 80–92.

Clarke, J.D. (2009). Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. Cold Spring Harb Protoc **2009**: pdb.prot5177–pdb.prot5177.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R., Lunter G., Marth G., Sherry S.T., McVean G., Durbin R. and 1000 Genomes Project Analysis Group (2011). The Variant Call Format and VCFtools. Bioinformatics **27**:2156-2158.

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., and Gibbs, R.A. (2012a). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS ONE **7**: e47768.

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., and Gibbs, R.A. (2012b). Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. PLoS ONE **7**: e47768.

Fischer, R.L. and Goldberg, R.B. (1982). Structure and flanking regions of soybean seed protein genes. Cell **29**: 651–660.

Garrison, E. and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv:**1207**.3907

Guo, Y., Wiegert-Rininger, K.E., Vallejo, V.A., Barry, C.S. and Warner, R.M., 2015. Transcriptome-enabled marker discovery and mapping of plastochron-related genes in Petunia spp. *BMC genomics*, *16*(1), p.726.

Korf, I. (2004). Gene finding in novel genomes. Bmc Bioinformatics **5**: 59.

Lamesch, P. et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res **40**: D1202–10.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Meth **9**: 357–359.

Lee, E., Helt, G.A., Reese, J.T., and Munoz-Torres, M.C. (2013). Web Apollo: a web-based genomic annotation editing platform. Genome Biology. **14**:R93 doi:10.1186/gb-2013-14-8-r93

Li, R. et al. (2010). The sequence and de novo assembly of the giant panda genome. Nature **463**: 311–317.

Liu, Y., Schroder, J., and Schmidt, B. (2013). Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. Bioinformatics **29**: 308–315.

Lowe, T.M. and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25**: 955–964.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., and He, G. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience **1**: 18.

Magrane, M. and Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford) **2011**: bar009–bar009.

Marçais, G. and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics **27**: 764–770.

McGinnis, S. and Madden, T.L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. **32** (Web Server issue):W20-5

Mulder, N. and Apweiler, R. (2007). InterPro and InterProScan. In Comparative Genomics, Methods in Molecular Biology. (Humana Press: Totowa, NJ), pp. 59–70.

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics **23**: 1061–1067.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol **29**: 24–26.

Schmieder, R. and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics **27**: 863–864.

Sheehan, H., Moser, M., Klahre, U., Esfeld, K., Dell'Olivo, A., Mandel, T., Metzger, S., Vandenbussche, M., Freitas, L. and Kuhlemeier, C., (2015). MYB-FL controls gain and loss of floral UV absorbance, a key trait affecting pollinator preference and reproductive isolation. *Nature genetics*. doi:10.1038/ng.3462

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res **34**: W435–W439.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635–641.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25**: 1105–1111.

Villarino, G.H., Bombarely, A., Giovannoni, J.J., Scanlon, M.J. and Mattson, N.S., (2014). Transcriptomic analysis of Petunia hybrida in response to salt stress using high throughput RNA sequencing. *PloS one*, *9*(4), p.e94651.

Zenoni, S., D'Agostino, N., and Tornielli, G.B. (2011). Revealing impaired pathways in the an11 mutant by high-throughput characterization of Petunia axillaris and Petunia inflata transcriptomes. The Plant Journal 68(1):11-27

Supplementary Note 2

Analysis of *Petunia vein clearing virus* (PVCV) sequences, retroelements and tandem repeats in *Petunia axillaris* N and *P. inflata* S6

Trude Schwarzacher[1*], J.S. (Pat) Heslop-Harrison[1] and Katja R. Richert-Pöggeler[2]

1) University of Leicester, Department of Genetics, Leicester LE1 7RH, UK.
2) Institute for Epidemiology and Pathogen Diagnostics, Julius Kühn-Institut, (JKI) – Federal Research Centre for Cultivated Plants, 38104 Braunschweig, Germany

*) Address correspondence to TS32@le.ac.uk

Short Title: PVCV, retroelements and tandem repeats in *Petunia*

ABSTRACT

Within the genome sequence assemblies of *P. axillaris* (*PaxiN*) and *P. inflata* (*PinfS6*) and unassembled reads, we analysed the occurrence of endogenous *Petunia vein clearing virus* (PVCV) sequences, other endogenous pararetrovirus (EPRV) sequences, LTR-retroelements, and tandem repeats. Petunia genomes show substantial diversity in their pararetroviral sequences as revealed in searches using the polymerase motif. Homologies to two genera of *Caulimoviridae*, *Petu*- and *Florendoviruses,* with more than 60% amino acid identity, were present in both genomes. Almost complete PVCV copies, fragments, and degenerate copies, sometimes in tandem arrays, were found. PVCV motifs were more frequent in *P. axillaris*, with the results seen in the assemblies confirmed by *in situ* hybridization of PVCV fragments to metaphase chromosomes indicating that *P. axillaris* is likely a more permissive host for EPRVs. LTR-retroelements are localised near centromeres; about 6500 full length elements were found in the *PinfS6* assembly while 4500 were in *PaxiN*. Apart from rDNA, microsatellites and telomeric sequences, no highly abundant tandem repeats were identified in the assembly or raw reads. Repeat cluster analysis indicates that LTR-retroelements are associated with simple sequence repeats and low complexity DNA families and that repeats within *Petunia* are very diverse, with none having amplified to form a major proportion of the genome. The repeat landscape of *Petunia* is different from other species of *Solanaceae*, in particular the x=12 crown group including *Solanum* and *Nicotiana*, with a relative low proportion (60-65%) of total repeats for a genome size of 1.4Gb, x=7, and a high degree of genome plasticity.

KEYWORDS

Petunia vein clearing virus, pararetrovirus, retroelement, *Copia* superfamily, *Gypsy* superfamily, telomere, tandem repeat, repeat explorer, K-mer analysis, genome organisation, chromosome size, fluorescent *in situ* hybridization

INTRODUCTION

The genome sizes reported for 20 species within the genus *Petunia* are similar with a range of 1.30 to 1.57 pg 1C for diploid species with a chromosome number of 2n=14 (Mishiba et al., 2000) corresponding to on average 1.4 Gb. The *Petunia* genome therefore is considerably larger than tomato and potato (900Mb and 844Mb respectively, Tomato Genome Consortium, 2012), but not as large as the hot pepper genome (3,480Mb; Kim et al., 2014) which contains a large proportion of repetitive sequences, in particular long terminal repeat (LTR) retroelements. DNA transposons have been described in *Petunia* (Gerats, 2009) and are analysed in Supplementary Note 3. In contrast, there are few reports of LTR retroelements including the *Gypsy* and *Copia* superfamilies. Matsubara et al. (2005) described the full length rTph1 element that shares features with the C*opia*-superfamily, and recently Kriedt et al. (2013) studied *Petunia* species relationships with the RNase H – 3'LTR region of eight families of *Ty1/Copia* –Tork clade elements that are related to the tobacco Tnt1. Both *Ty3*/*Gypsy* and *Ty1*/*Copia* reverse transcriptase domains were described from several petunia species by Richert-Pöggeler and Schwarzacher (2009). Tandemly repeated satellite DNA families in *Petunia* have not been widely reported.

Petunia (Richert-Pöggeler et al., 2003), similar to other members of *Solanaceae* (Hansen et al., 2005; Geering et al., 2014), has integrated pararetrovirus sequences. A spontaneous outbreak of vein-clearing disease in the hybrid species *P. hybrida* could be traced back to activation of *Petunia vein clearing virus* (PVCV) genomes inserted into the host chromosomes in a tandem array (Richert-Pöggeler et al., 2003). PVCV belongs to the genus *Petuvirus* within the family of *Caulimoviridae* (King et al., 2012). This virus family is also referred to as plant pararetroviruses since it uses reverse transcription for genome amplification. In contrast to retroviruses, their genome consists of a circular double stranded DNA with single strand gaps (Hohn et al., 2008). Furthermore, integration into the host genome is not obligatory for the replication cycle of parartroviruses. Comparing the genome sequences of *P. axillaris* N (*PaxiN*) and *P. inflata* S6 (*PinfS6*) allows the study of the genomic context to determine any effect on diversity and evolution of *Caulimoviridae*.

RESULTS

Petunia vein clearing virus (PVCV) insertions

Integrated sequences of PVCV were found in the genomic scaffolds and on chromosomes of both petunia species analyzed. In both genome assemblies, several arrays of almost complete genome length and degenerated PVCV sequences were found (Tables 1 and 2, Figure 1 and details below). Data obtained from BlastN searches using the whole PVCV genome of 7206 bp were filtered for alignments larger than 500 nt in length. For *PaxiN,* 30 sequences were selected that ranged in size from 542 to 2848 nt in length and 80-99% identity, whereas *PinfS6,* 9 sequences with sizes of 563-635 nt and 78-80% identity were found. Fluorescent *in situ* hybridization with PVCV-specific probes comprising the complete viral genome created a stronger signal in *P. axillaris* than in *P. inflata*, both at centromeric regions of chromosomes III and VI for *P. axillaris* and chromosome IV for *P. inflata* (Figures 2 and 3). The bioinformatic and cytogenetic data indicate that PVCV abundance and perseverance is markedly higher in *P. axillaris* compared to *P. inflata.*

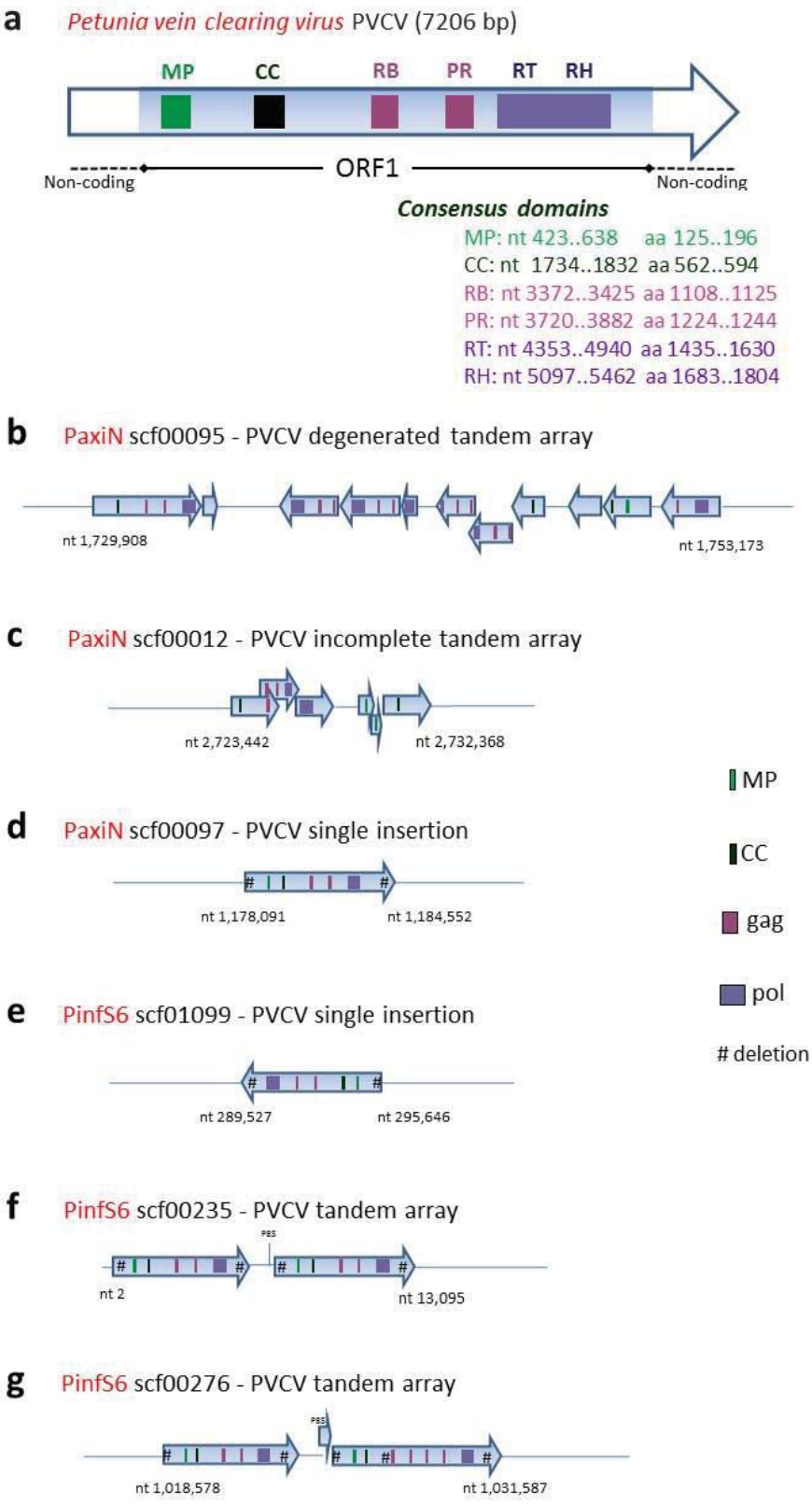


Figure 1: Full length PVCV genomes and PVCV integrations in *P. axillaris* N and *P. inflata* S6 genome assembly scaffolds. Integrations show head-to-tail and inverted orientations, deletions, duplications and overlapping protein motifs. For protein and regulatory DNA element boundaries see Tables 1 and 2.

MP: movement protein, CC: coiled coil domain, gag: group-specific antigen with RB (RNA binding Zn-finger motif of the capsid protein) and PR (protease), pol: polymerase with RT (reverse transcriptase) and RH (RNaseH); PBS: primer binding site comprising first 14nt of PVCV genome. Not shown are TSS (transcription start site) and Pro (Promoter) within the 3'-non coding region. Note that the single insertion depicted in d) for *PaxiN* shows 74% amino acid identity with PVCV compared to the single insertion depicted in e) for *PinfS6* displaying 56% amino acid identity.

The behavior and mode of integration seem to be similar for both species: the PVCV elements are present as single insertions as well as small tandem arrays. The latter display deletions as well as rearrangements and thus are not able to generate full length infectious viral RNA molecules via direct transcription as reported for *P. hybrida* W138 (Richert-Pöggeler et al., 2003). The investigated petunia species differ in preservation of integrated viral sequences. Those in *PinfS6* are subjected to a higher degree of degradation. At a stringent level of more than 2000aa and an id>70% we found only one PVCV copy in *P. inflata* compared to four copies in *PaxiN*.

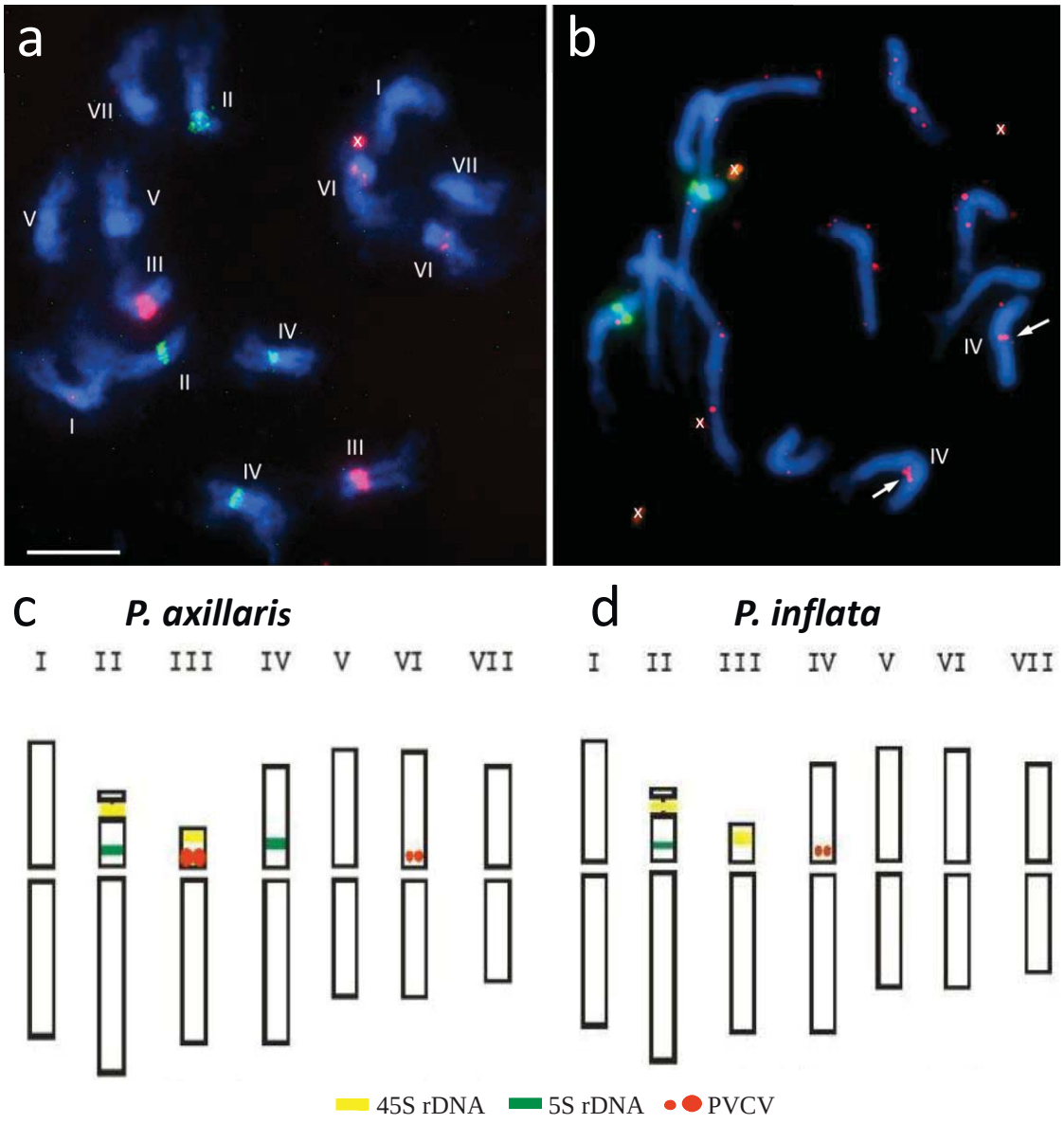


Figure 2: PVCV distribution on metaphase chromosomes *P. axillaris* and *P. inflata*
a,b: Fluorescent *in situ* hybridization using PVCV probe (Richert-Pöggeler et al., 2003; depicted in red) and 5S-rDNA (depicted in green) counterstained with DAPI (blue) **a:** *P. axillaris* chromosome identifications are indicated; PVCV signal is found at the centromeres of chromosomes III and VI (this spread was reprobed with *gypsy*-related retroelement junction fragment 4-18, see Figure 3). **b:** *P. inflata* chromosome IV is identified and the location of PVCV is indicated by arrows while remaining dots are background signal visible due to enhancement of weak signal. x denotes stain precipitates. Bar = 10 µm
c, d: Idiograms showing PVCV, 5S and 45S rDNA.

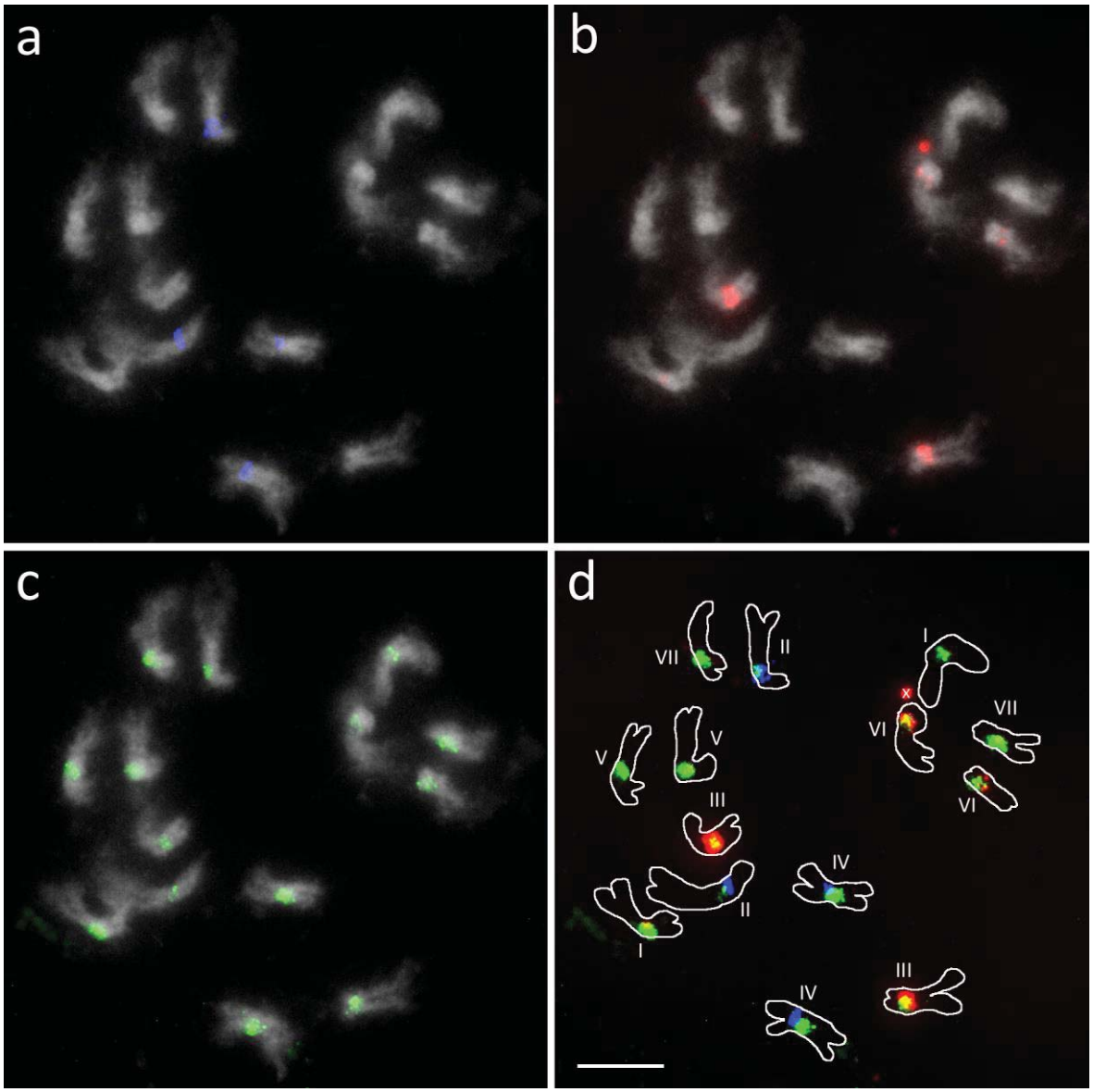


Figure 3: **Fluorescent *in situ* hybridization to metaphase chromosomes of *P. axillaris.***
a, b, c: Chromosomes were counterstained with DAPI (shown in grey), and hybridization sites are shown in blue for the 5S-rDNA (**a**) and red for PVCV (**b**) from the first hybridization (see Figure 2a) and in green for the *Gypsy* super family retroelement junction fragment 4-18 used to reprobe the spread (**c**).
d:The overlay is a drawing of the chromosomes with overlayed and enhanced FISH signal (overlapping signal of red and green appears yellow; original see main text Figure 2b). Chromosomes are identified by their size and 5S rDNA signal. Two large and two small PVCV sites close to the centromeres of chromosomes III and VI are present while all centromeres show dispersed signal with the *Gypsy*-related retroelement probe. X denotes stain precipitate. Bar = 10 µm.

PVCV sequence insertions in *PaxiN*

Highly preserved stretches of the N-terminal half of PVCV ORF1 with 99% aa identity (Table 1) comprising the MP consensus domain were identified in scf03256 and in scf01628. The latter scaffold contained at its 5'-end, nt 7239 -7835, highly preserved (96% identity) regulatory sequences out of the PVCV untranslated region (nt 6591 -7193) including the promoter, transcription start site as well as a polyA signal (Noreen et al., 2007). Both scaffolds contained also other parts of ORF1 that were less conserved (45 to 87% aa identity).

Table 1: Identified PVCV sequences within *PaxiN* scaffolds annotated in the assembly Peaxi162.

Scaffold (% identity)	Scaffold length (bp)	PVCV location in scf	PVCV-ORF1 (2179aa)	Orientation, frame	FISH detection by PVCV probes[1)]
***PaxiN* scaffolds with single PVCV sequences, filter: alignment length > 1500aa, id > 70%**					
scf00097 (74)	2.428.612	1.178.091 - 1.184.552	11-2176	forward	PVCV-L, M, R
scf00254 (74)	1.519.427	751.256 - 744.795	11-2176	reverse	PVCV-L, M, R
scf00447 (72)	1.217.748	1.146.839 - 1.152.364	318-2176	forward	PVCV-L, M
scf00560 (74)	869.919	238.998 - 245.459	11-2176	forward	PVCV-L, M, R
scf00674 (74)	473.906	265.554 - 271.907	11-2141	reverse	PVCV-L, M, R
scf00911 (74)	507.819	34.116 – 29.446	598-2176	reverse	PVCV-M
***PaxiN* scaffolds with highly preserved PVCV sequences, filter: alignment length > 380aa, id = 99%**					
scf03256 (99)	12295	552 - 1964	1-470	forward	PVCV-R
scf01628 (99)	9910	313-1460	50-432	forward	PVCV-R
***PaxiN* scf00012 with multiple PVCV sequences, filter: alignment length > 100aa, id > 40%**					
scf00012 (43%)	3.935.541	2.723.442 - 2.725.583	417-1139	forward, +3	?
scf00012 (64%)	3.935.541	2.725.433 - 2.727.046	1088-1631	forward, +2	PVCV-M
scf00012 (50%)	3.935.541	2.727.046 - 2.728.506	1636-2134	forward, +1	?
scf00012 (65%)	3.935.541	2.729.342 -2.729.914	4-198	forward, +2	?
scf00012 (67%)	3.935.541	2.729.863 - 2.730.255	179-309	forward, +1	?
scf00012 (43%)	3.935.541	2.730.362 - 2.732.368	345-1007	forward, +2	PVCV-L, M
***PaxiN* scf00095 with multiple PVCV sequences, filter: alignment length > 100aa, id > 40%**					
scf00095 (53%)	1.774.960	1.729.908 - 1.734.470	287-1811	forward, +3	PVCV-M
scf 00095 (52%)	1.774.960	1.734.563 - 1.734.898	1854-1965	forward, +2	?
scf 00095 (63%)	1.774.960	1.739.779 - 1.737.299	1126-1956	reverse, -1	PVCV-M
scf 00095 (64%)	1.774.960	1.742.275- 1.739.783	1059-1905	reverse, -1	PVCV-M
scf 00095 (71%)	1.774.960	1.742.929 -1.742.255	1571-1795	reverse, -1	PVCV-M
scf 00095 (58%)	1.774.960	1.744.585 - 1.742.936	945-1503	reverse, -1	PVCV-M
scf 00095 (60%)	1.774.960	1.746.506 - 1.744.659	945-1569	reverse, -3	PVCV-M
scf 00095 (44%)	1.774.960	1.747.798 - 1.746.485	497-951	reverse, -1	PVCV-M
scf 00095 (50%)	1.774.960	1.749.980 - 1.748.940	679-1046	reverse, -3	PVCV-M
scf 00095 (51%)	1.774.960	1.752.079 - 1.749.992	4-665	reverse, -1	PVCV-L
scf 00095 (63%)	1.774.960	1.755.647 - 1.753.173	1221-2059	reverse, -3	PVCV-M

1) For FISH stringent washing conditions were applied, therefore sequences with < 63 % of identity probably did not contribute to the observed signal as letters of PVCV-probes in grey indicate. "?": identified sequences are only covered partly (< 180 aa) by probes used.

In total, 49 scaffolds carrying 1 to 16 fragments of PVCV-like sequences, with a majority (73%) containing 1 to 5 PVCV fragments of various sizes were detected in searches using a cut off >100aa aligned, >40% aa identity. In six scaffolds (see Table 1), large continuous stretches of integrated sequences homologous to PVCV (cut off >1500aa aligned, >70% aa identity) comprising ORF1 in forward or reverse orientation were revealed. Besides single blocks, integrated PVCV sequences were found as tandem array-like structures (see scf00012 and scf00095, in Table 1 and Figure 1b and c). Thus the following integration patterns can be distinguished for the annotated sequences within the *P. axillaris N* genome: 1) the insertion spots (scf00095, scf00012) show rearrangement and fragmentation or 2) the integrated PVCV sequences (scf00097, scf00254, scf00447, scf00560, scf00674, scf00911) cover almost the complete ORF1 region. Interestingly, four of the latter copies

comprise the same region of aa 11-2141 in PVCV ORF1 while the other two copies show larger deletion on the N-terminal end of ORF1.

Filtered results obtained for scf00095 (id% > 40 and alignment length >100 aa) identified 11 copies of various ORF1 fragments which can be assembled into a "tail to tail - tail to head" array. The first part of the array contains deletions at the N- and C- terminal ends of ORF1 as well as near the C-terminal end. The latter is accompanied by a frameshift from +3 to +2. The "tail to head" part contains repetitive blocks of PVCV sequences and frameshifts (Table 1). The same filter applied for scf00012 revealed an almost complete ORF1 of PVCV comprising amino acids 4 to 2134 containing rearranged fragments, smaller deletions as well as frameshifts. For detection of PVCV sequences petunia chromosomes were hybridized with a mixture of three virus specific probes (Richert-Pöggeler et al., 2003) named PVCV-L (nt 657-1793), PVCV-M (nt 2235-5321) and PVCV-R (nt 5445-7206 and nt 1-671) using fluorescent *in situ* hybridization (FISH). In Table 1 we indicated detection only for those scaffold sequences that either are completely covered or contain at least 180 aa in common with probes PVCV-L, –M or -R. We suggest that mostly the single insertions detected by PVCV-L, –M and -R probes are responsible for the PVCV FISH signal identified within the pericentromeric region on chromosomes III and VI of *P. axillaris* (Figures 2 and 3, Table 1).

PVCV sequence insertions in *PinfS6*:

There are two near-complete copies of PVCV ORF1 at the 5'-end of scf00235 (56% identity), separated by 1000nt (nt 5903-6975). Within the region separating the two PVCV copies, a domain identical to the PVCV primer binding site (PBS) was located at nt 6907-6920 of the scaffold (Figure 1f). Besides the PBS no homologies to known sequences were identified, but as shown in Table 3 this junction sequence was frequently adjacent to PVCV insertions.

The strongest similarity to PVCV was found for sequences in scf00276 (Figure 1g). It contained a tandem array of PVCV sequences (head to tail) that were separated by 1 kb of sequences with no known homology (nt 1.024.645-1.025.740 in scf00276). The first copy in frame +3 covered most of PVCV ORF1 with a major (144 aa) and minor (3 aa) deletion at the N-terminal and carboxy terminal end respectively and a stop codon towards the carboxy terminal end. The 2nd copy showed deletions at both ends as well as in the centre of ORF1. The sequence with homology to PVCV aa 1007-1247 was inserted leading to a duplication of the gag region. The tandem array separating sequence (nt 1.024.645-1.025.740) was a hybrid of PVCV- and non-PVCV sequences. The latter showed no homology to sequences deposited in GenBank, but stretches of it could be found in scf09521 and scf16590 (see Table 3).

In scf00059, two blocks of fragmented PVCV sequences in forward orientation followed by clustered PVCV sequences comprising aa 4 to 2047 from PVCV ORF1 in reverse orientation were located at its 3'-end. The latter contained frameshifts. In the first quarter of scf00753, a cluster of repetitive PVCV fragments is present, mostly in the same orientation. They are not continuous but separated by non-PVCV sequences. The existence of only one continuous PVCV-sequence with >70% aa identity found in PinS6scf00276 (Table 2, Figure 1g) compared to numerous of those identified in *P. axillaris N* (Table 1) may explain the weak detection of PVCV in the *P. inflata S6* genome (Figure 2b).

Table 2: Identified PVCV sequences within *PinfS6* scaffolds annotated in the assembly Peinf101.

Scaffold (identity)	Scaffold length scf (bp)	PVCV location in scf	PVCV-ORF1 (2179aa)	Orientation, frame	Comments
PinfS6 scaffolds with single PVCV sequences, filter: alignment length > 1500aa, id > 50%, *alignment length > 800aa, id > 65%*					
scf00251 (55%)	2.742.664	2.734.378 - 2.729.603	4-1591	reverse	
scf 00844 (56%)	223.102	199.961 - 193.842	4-2059	reverse	
scf 01099 (56%)	1.012.215	295.646 - 289.527	4-2059	reverse	
scf 01671 (56%)	432.778	241.933 - 235.814	4-2059	reverse	
***PinfS6* scf00235 with multiple PVCV sequences, filter: alignment length > 1500aa, id > 50%**					
scf 00235 (56%)	1.810.835	6.976 – 13.095	4-2059	forward	tandem array head to tail with filler (s. text)
scf 00235 (56%)	1.810.835	2 – 5.902	78-2059	forward	
scf 00235	1.810.835	5903 - 6975	unknown sequence (s. Table 3) carrying PVCV-PBS nt 6907-69		
***PinfS6* scf00276 with multiple PVCV sequences, filter: alignment length > 800aa, id > 65%**					
scf 00276 (73%)	1.557.018	1.018.578 - 1.024.634	145-2176	forward, +3	tandem array head to tail with filler sequences (s. text)
scf 00276	1.557.018	1.024.645 - 1.025.264	unknown sequence (s. Table 3)		
scf 00276 (75%)	1.557.018	1.025.265 - 1.025.740	PVCV nt 1-476 (PBS, degenerated N terminal part of ORF1		
scf 00276 (68%)	1.557.018	1.025.741 - 1.028.308	145-1003	forward, +2	
scf 00276 (76%)	1.557.018	1.028.309 - 1.029.025	1007-1247	forward, +2	
scf 00276 (78%)	1.557.018	1.029.026 - 1.031.587	1248-2102	forward, +2	
***PinfS6* scf00059 with multiple PVCV sequences, filter: id>50%, alignment length >100aa**					
scf 00059 (61%)	1.578.608	64.320 - 65.480	4-390	forward, +3	
scf 00059 (55%)	1.578.608	66.366 - 67.226	1088-1379	forward, +3	
scf 00059 (51%)	1.578.608	76.324 - 74.510	1433-2047	reverse, -2	
scf 00059 (51%)	1.578.608	77.739 - 76.684	970-1333	reverse, -3	
scf 00059 (46%)	1.578.608	80.153 - 77.736	170-970	reverse, -1	
00059 (68%)	1.578.608	80.641 - 80.156	4-166	reverse, -2	
***PinfS6* scf00753 with multiple PVCV sequences, filter: id>60%, alignment length >100aa**					
scf 00753 (70%)	2.084.733	418.612 - 419.550	4-317	forward, +1	
scf 00753 (72%)	2.084.733	444.046 - 443.360	23-251	reverse, -3	
scf 00753 (73%)	2.084.733	447.339 – 446.515	1316-1587	reverse, -1	
scf 00753 (62%)	2.084.733	450.381 – 450.016	1196-1317	reverse, -1	
scf 00753 (67%)	2.084.733	452.331 – 453.290	4-324	forward, +3	
scf 00753 (69%)	2.084.733	458.072 – 459.013	7-317	forward, +2	
scf 00753 (70%)	2.084.733	472.568 – 473.221	1357-1574	forward, +2	
scf 00753 (61%)	2.084.733	475.644 – 475.988	4-119	forward, +3	
scf 00753 (74%)	2.084.733	475.958 – 476.356	110-242	forward, +2	
scf 00753 (70%)	2.084.733	477.668 – 478.750	1357-1717	forward, +2	
scf 00753 (71%)	2.084.733	480.792 – 481.346	4-189	forward, +3	
scf 00753 (65%)	2.084.733	483.902 – 485.188	1130 - 1561	forward, +2	
scf 00753 (68%)	2.084.733	486.922 – 487.746	4-279	forward, +1	
scf 00753 (71%)	2.084.733	502.531 – 502.031	113-279	reverse, -3	
scf 00753 (61%)	2.084.733	502.856 – 502.530	4-113	reverse, -2	
scf 00753 (63%)	2.084.733	517.270 – 516.731	4-184	reverse, -3	

Table 3: Scaffolds with sequences associated with PVCVs

scf number (identity)	scf Length (bp)	location in scf	orientation, association with PVCV sequences (Ps)
PVCV-sequences containing scaffolds with similar sequences to *PinfS6* scf00235_nt5903-6975 including PVCV-PBS at nt 6907-6920 [1)]			
scf00059 (92%)	1.578.608	68.896 – 67.916	reverse, between Ps
scf00235 (100%)	1.810.835	5.903 – 6.975	forward, between Ps
scf00235 (99%)	1.810.835	13.095 – 13.871	forward, no association with Ps
scf00753 (92%)	2.084.733	456.955 – 457.923	forward, upstream Ps with gap
scf00753 (94%)	2.084.733	479.737 – 480.702	forward, upstream Ps
scf00753 (94%)	2.084.733	492.613 – 493.385	forward, between repetitive Ps with gaps
scf00753 (93%)	2.084.733	438.095 – 437.356	reverse, between Ps with gaps
scf00753 (95%)	2.084.733	444.927 – 444.189	reverse, downstream with gap Ps
scf00753 (92%)	2.084.733	503.928 – 503.327	reverse, downstream with gap Ps
scf00844 (99%)	223.102	193.842 – 192.768	reverse, upstream Ps
scf01099 (100%)	1.012.215	289.527 - 288.522	reverse, upstream Ps
scf01671 (99%)	432.778	235.814 - 234.820	reverse, upstream Ps
Scaffolds with fragments similar to parts of sequence *PinfS6* scf 00276, nt 1.024.645 - 1.025.264			
scf09521 (92%)	17538	8.085 – 8.285	forward
scf09521 (87%)	17538	8.318 – 8.510	forward
scf16590 (95%)	13872	11.193 – 11.038	reverse

1) This sequence shows no homology to known sequences and was found in all PVCV containing scaffolds mentioned in Table 2, with the exception of scf00276.

Similarity analysis of selected integrated PVCV sequences within *PaxiN* and *PinfS6* genomes

Similarity analysis of aligned amino acid sequences from the nearly full-length PVCV insertions identified within the petunia genomes revealed two major clusters for chromosomal PVCV sequences (Figure 4). The tree topology supports the hypothesis that after speciation there had been at least two separate invasion events of petunia genomes by episomal PVCV sequences (Staginnus and Richert-Pöggeler, 2006). At an earlier event only *P. inflata* has been affected which is also illustrated in a lower value of sequence identity (Table 2). A second round of invasion included both *P. inflata* and *P. axillaris* genomes. This clade also includes episomal PVCV, and it probably has contributed to the tandem array structure of inducible endogenous PVCV existing after genome hybridization in *P. hybrida* (Richert-Pöggeler et al., 2003). Here the almost full-length single copies of PVCV show a slightly higher degree of preservation for the *P. axillaris N* genome. Evolution within the genome resulted in deletions, amino acid exchanges and ORF disruption by a stop codon in case of *PaxiN* scf00447 and scf00674. In the *P. inflata* genome context, detrimental forces seem to act more strongly on foreign DNA. The PVCV continuous copy *PinfS6* scf00276 carries not only a stop codon but also an amino acid exchange (*aspartic acid D* to aspartate N) within the reverse transcriptase consensus domain "VYID*D*VLL" common to a broad range of retroelements (Richert-Pöggeler and Shepherd, 1997). Adjacent to the full-length copy fragmented and rearranged PVCV sequences are located in *PinfS6* scf00276. Most likely they originate from the same integration event since the fragments show similar values for amino acid identity and PVCV sequences within the tandem array structure were most conserved.

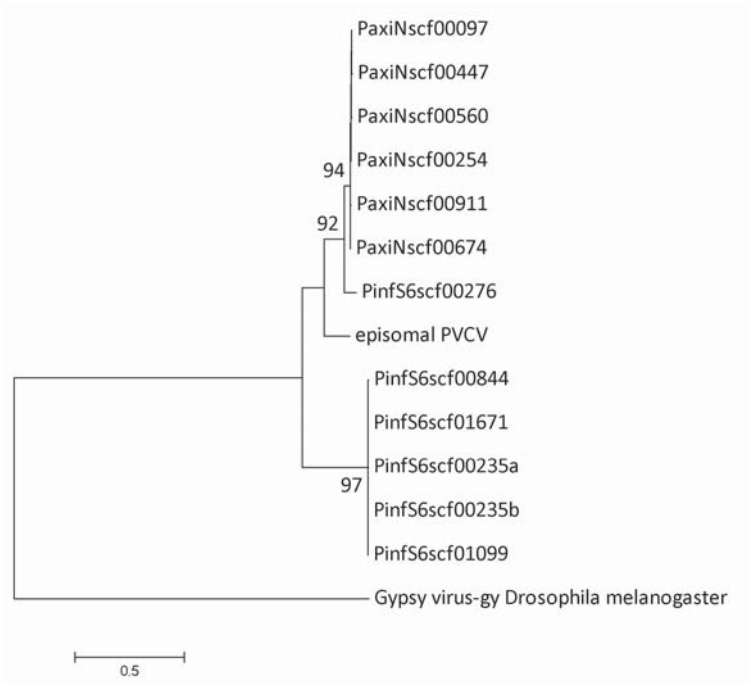
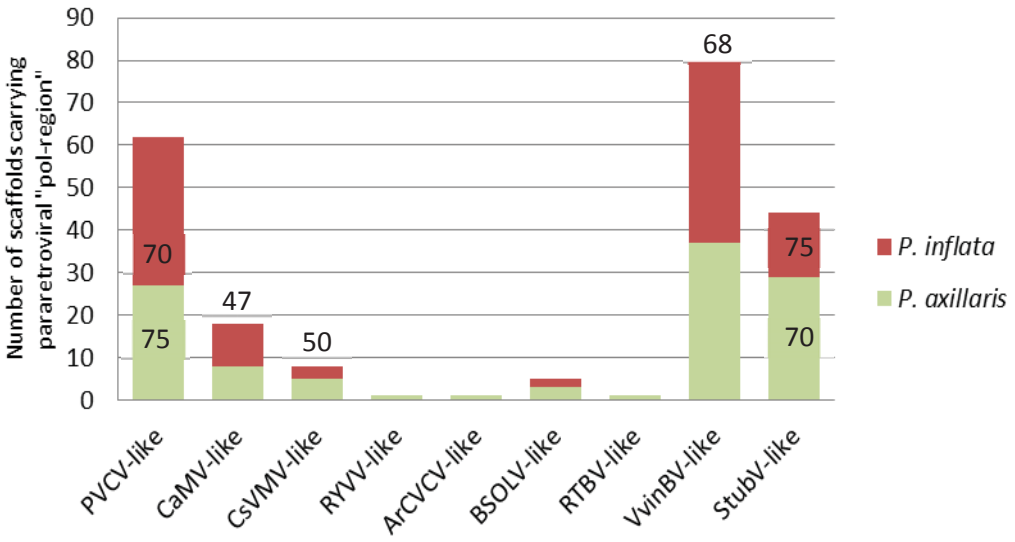


Figure 4: **Similarity analysis of PVCV ORF1 insertions within *P. axillaris N* and *P. inflata S6* genomes.**
The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992). The tree with the highest log likelihood (-7930.8490) is shown. The percentage of trees, cut off > 70, in which the associated taxa clustered together, is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically: when the number of common sites was < 100, or less than a quarter of the total number of sites, the maximum parsimony method was used; otherwise BIONJ method with MCL distance matrix was used. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 14 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 923 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011). The *Gypsy virus –gy* of *Drosophila melanogaster* (DmeGypV-gy) was used as the outgroup.

Comparison of EPRV diversity in *PaxiN* and *PinfS6*

The pol region comprising the reverse transcriptase and RNase H is the most conserved motif of EPRVs. To identify additional PVCV insertions and EPRV diversity, scaffolds were searched for *Caulimoviridae* pol-like domains using the PVCV pol-region comprising aa 1425 to 1804 and *Caulimoviridae* sequences with homology to the PVCV pol-region as identified in BLASTp searches. These include pol-like regions with homology to bacilliform pararetroviruses illustrated by *Banana streak OL virus* and *Rice tungro bacilliform virus* and sequences from the genera *Caulimo*/*Cavemovirus* both displaying isometric particle morphology and a putative pararetrovirus with unknown morphology, *Aristotelia chilensis vein clearing virus*. Sequences were investigated further if the alignment length was >200 amino acids, or if the identity was >60% in case of PVCV and the two selected florendoviruses (Geering et al., 2014) or >45% for all other *Caulimoviridae.*

a



b

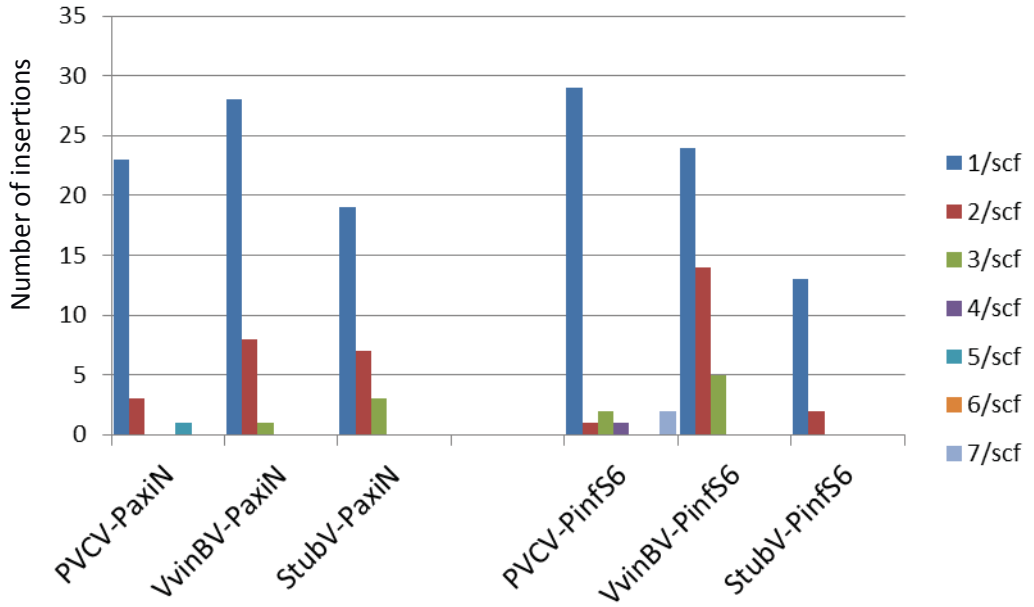


Figure 5: Distribution of EPRV pol-like sequences in *PaxiN* and *PinfS6* assemblies.

a. Numbers connected with columns indicate average amino acid identity in % of the identified pol region compared to the corresponding EPRV.

b. Insertion numbers per scaffold (scf) of pol-like petu- and florendovirus sequences PVCV: *Petunia vein clearing (Petu-) virus;* CaMV: *Cauliflower mosaic (Caulimo-) virus;* CsVMV: *Cassava vein mosaic (Cavemo-) virus;* RYVV: *Rose yellow vein virus* (unassigned genus)*;* ArCVCV: *Aristotelia chilensis vein clearing virus (unassigned genus).* BSOLV: *Banana streak OL (Badna-) virus*; RTBV: *Rice tungro bacilliform (Tungro-) virus;* VvinBV: *Vitis vinifera B (Florendo-) virus isolate -compAsc*1 and StubV: *Solanum tuberosum (Florendo-) virus isolate* -scSt1. The corresponding known episomal viruses displaying different particle morphologies: isometric shape for PVCV, CaMV, CsVMV and RYVV, bacilliform shape for BSOLV and RTBV. For ArCVCV, VvinBV and StubV no information on episomal virus or particle morphology is available.

The two petunia genomes differ slightly in content and diversity of pol-like regions homologous to endogenous pararetroviruses (EPRVs, Figure 5a). In most cases, only one conserved motif was found on a scaffold (Figure 5b). The most abundant pol-like consensus sequences in both genomes, belonged to Petu- and Florendoviruses. Those showed also a high degree of conservation illustrated by 68-75% of sequence identity (Figure 5a). Multiple insertions of three different EPRVs with homology to *Petu-*, *Badna-* and *Florendoviruses* were only found in *PinfS6* scf00909. PVCV-like sequences with 4 separated insertions within the first quarter of *PinfS6* scf00909 were accompanied by 2 StubV-like and 1 BSOLV-like sequences closer to the 3' end of the scaffold. In *PinfS6* scf00073 the two *Florendoviruses* were found at opposite ends of the scaffold with one insertion each. In *PinfS6* scf07983 a StubV-like pol-region was found closer to the middle of the scaffold whereas two VvinBV-like insertions were positioned at the 5' end and 3' end respectively. Only *PaxiN*_scf00380 harbored a combination of the two investigated *Florendoviruses* at its 3'end consisting of two StubV-like pol-region and one VvinBV-like element. The observed higher variability regarding numbers and conservation among *Florendoviruses* in *PinfS6* compared to *PaxiN* (Figure 5) might indicate various time points of invasion.

Other EPRV-like sequences were more degenerate with respect to the full-length virus sequence, showing an average of 49% identity. Multiple occurrences of the same or distinct elements on single scaffolds occurred less frequently, and the motifs were separated by several kb, and did occur in tandem repeats. Examples of linked elements include scf00027 of *PaxiN* with both PVCV-like and CaMV-like domains; scf00654 with both PVCV- and RTBV-like sequences; scf00514 of *Pinf6* with both CaMV-like domains and CsVMV-like pol domains.

Retrotransposons

LTR-STRUCT (McCarthy et al. 2003) was used for *de novo* retroelement searches in the assembled scaffolds. In *P. inflata* S6, a total of 595 RT (reverse transcriptase) active site types, 914 PBS (primer binding site) types, and 996 PPT (polypurine tract) 5'-end types were found from a total of 7354 LTR retrotransposons of which 4147 had RT domains. In *PaxiN*, 4573 LTR retrotransposons were found of which 1,850 included an RT region. These ranged between 1,183bp and 24,737 bp and had LTRs of 76bp to 5344bp. They were classified into *Ty3/Gypsy* (*Metaviridae*) superfamily and *Ty1/Copia* (*Pseudoviridae*)-superfamily elements by their gene order; *Gypsy* elements having the order RT-RH-INT (integrase) while in *Copia* elements the order is INT-RT-RH (see Hansen and Heslop-Harrison, 2004). 52% of the retroelements were categorized as *Gypsy* superfamily, and most of the remaining were *Copia s*uperfamily elements. Twelve selected elements were annotated in Geneious with a local database of motifs taken from the LTR-STRUCT analysis and from Hansen and Heslop-Harrison (2004). The most common retroelement families including *Athila* and *Cyclops Gypsy*-like elements and *BARE Copia*-like elements were found, but some showed various insertions, deletions and inversions (see examples in Figures 6 and 7).

Figure 6: Annotation of the *Copia BARE*-like LTR retroelement 86. It was identified by LTR_STRUC analysis of *PaxiN* assembly Peaxi162 with Ps20000_RT19_B1_L41_86, is 7418 bp long and contains LTR 41 (red) at each end (447bp and 451bp with 96.9% homology), protein binding site (PBS, cyan), the aspartic protein gene (BARE-1_APR, yellow), integrase (BARE-1_INT; magenta), a functional RT region (BARE-1_RT, blue) with the putative active sites SYDDVLF and YVDDILM, RNaseH (BARE-1_RNaseH, green), and a polypurine tract (orange).

Figure 7: Annotation of the *Cyclops Gypsy*-like element 414. It was identified by LTR_STRUC analysis of *PaxiN* assembly Peaxi162 with Ps15485_RT49_B65_L129_414; it is 13252 bp long and contains LTR 129 (red) at each end (1307bp and 1306bp with 96.9% homology), primer binding site (PBS, cyan), integrase (Athila_int and overlapping split Cyclops_INT, magenta), RNaseH (Cyclops_RNaseH, green), a functional RT region (Cyclops_NT, blue) with the putative active sites FLDDLLF and WLDDGII, that is inverted between nt5,708 and nt 11,150, an aspartic protease motif (Cyclops_APR, yellow) and a polypurine tract (orange)

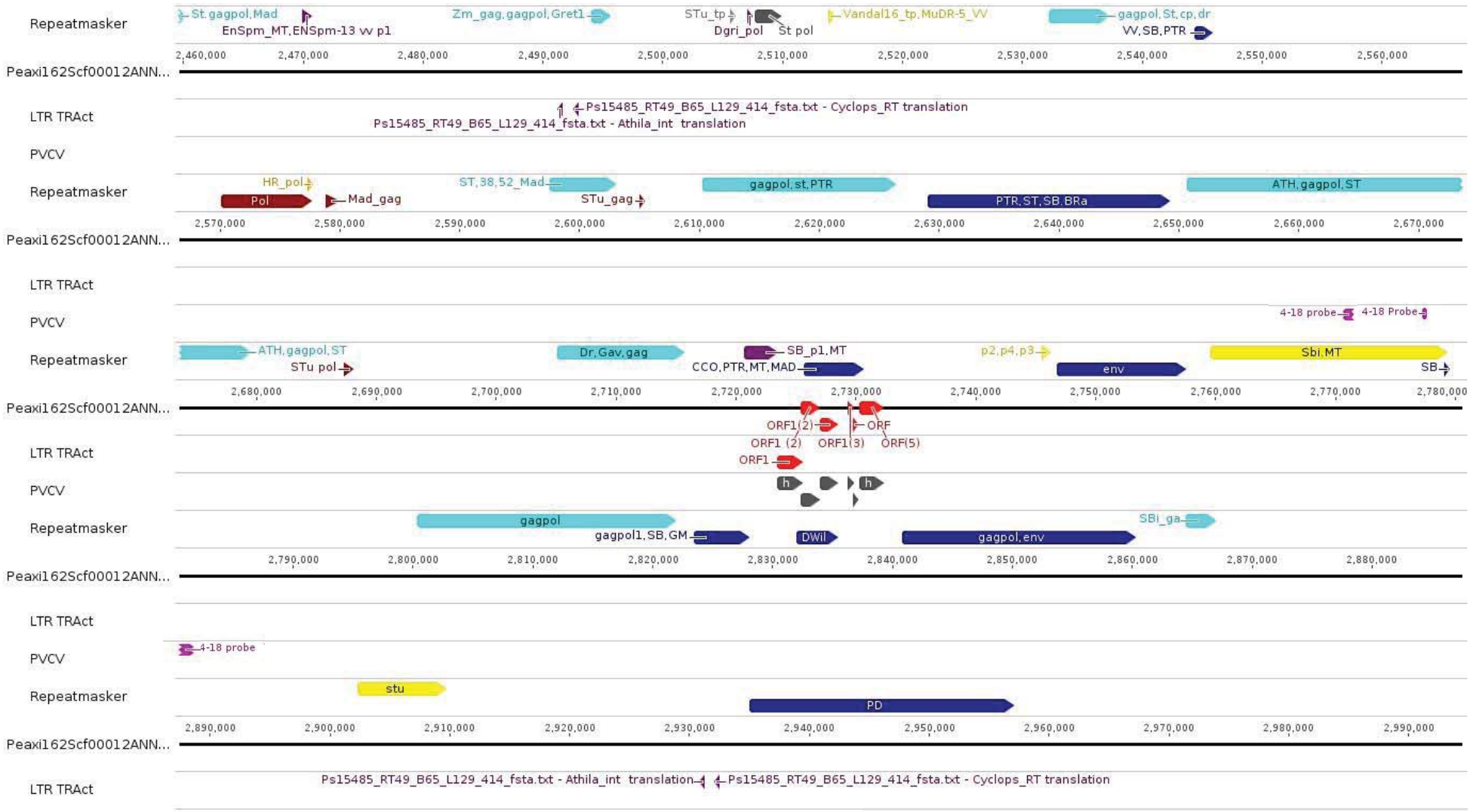


Figure 8
Annotation of *PaxiN* scf00012 from nt 2,460,000 to 2,990,000 using Geneious software with LTR-STRUC analysis (LTR TRAct, PVCV ORFs in red), Repeatmasker, manual alignment with reference transposon sequences from Hansen and Heslop-Harrison (2004), PVCV (grey) and the *Gypsy* super family retroelement junction fragment 4-18 (purple), identification from Tables 1 and 4

Retroelement and endogenous PVCV relation

During the analysis of lambda clones obtained from screening a genomic DNA library of *P. hybrida* (Richert-Pöggeler et al., 2003) it was noted that *Metaviridae* (LTR-*Gypsy* superfamily) sequences were adjacent to integrated PVCV sequences. One such sequence, the 1.2 kb *Gypsy* superfamily retroelement junction fragment 4-18 (1,233bp;nt 4483-5715 of GenBank AY333912, *P. hybrida* lambda clone 4, Richert-Pöggeler et al., 2003) was used for FISH experiments to *P. axillaris* chromosomes (Figure 3c and d). A strong signal is visible at the centromeres of all chromosomes and next to or interspersed with both the strong PVCV signal on chromosome III and the weaker PVCV signal on chromosome VI. To further analyse the surrounding sequences of PVCV, scf00012 of *PaxiN* was annotated (Figure 8). Many gag-pol regions indicative of LTR retroelements were found including some in the immediate vicinity of the PVCV tandem array.

Because of the distribution of fragment 4-18 (Figure 3c and d), scaffolds of *PaxiN* assembly at least 500kb in length were searched by BlastN to identify sequences related to fragment 4-18 that were at least 150 bp long. Within scaffolds scfs0000-scfs0999, 635 scaffolds contained at least one copy, but only 7% of scaffolds had more than a total of 1% of sequences homologous to 4-18 and only 9 clones contained 3-6% (see Figure 9A). Longer scaffolds on the whole contained less 4-18 related sequences, while not all but some shorter scaffolds have more (see Figure 9B). Possibly, this is an under-representation and unassembled shorter reads need to be checked. Results for scaffolds scf00012, scf00095 and scf00097 that also contain PVCV are given in Table 4. Interestingly only in scf00012, a larger number (58, 0.9%) of 4-18 related sequences were found, three copies in the vicinity of PVCV.

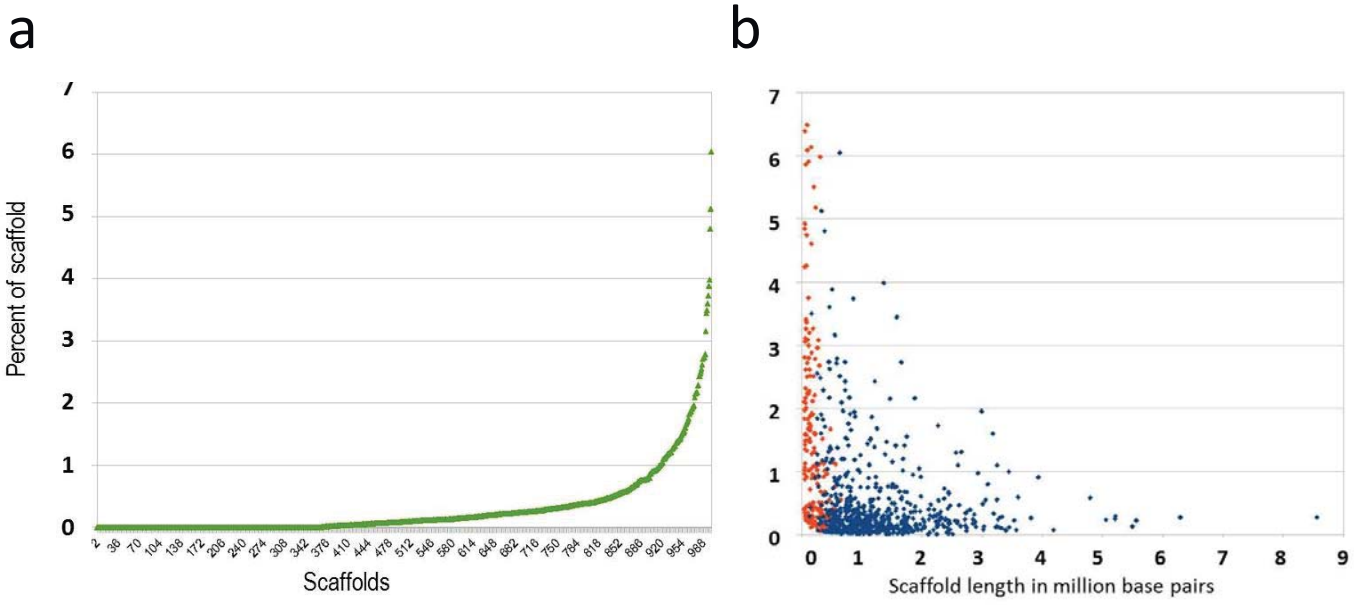


Figure 9: Percentage of sequences longer than 150bp and homologous to *Gypsy* gag-pol 1.2kb fragment 4-18 in *P. axillaris* N scaffolds.

a. Arranged by increasing frequency in scfs 0-999

b. Plotted against scaffold length of all scaffolds with a length greater than 500kb; scaffolds without 4-18 sequences are not plotted; scaffolds 0-999 blue, remaining red

Table 4: *Gypsy* superfamily retroelement junction fragement 4-18 in scf00012, scf00095 and scf00097 of *PaxiN* assembly Peaxi162. These also contain PVCV sequences. Hits longer than 150 bp are listed with start and end for hit (scf) and query (4-18 fragment) as well as total number of hits, length and proportion in scaffold highlighted in yellow. Those sequences in the vicinity of PVCV (±200kb) are highlighted in orange.

Scaffold	caffold length	Sequence length	% scf	Hit end (nt)	Hit start (nt)	Query nd (nt)	Query art (nt)	Pairwise identity	E value
scf00012		187		51,828	51,642	1219	1405	84.8%	1.73E-48
scf00012		522		88,675	88,154	465	937	67.5%	8.97E-46
scf00012		670		154,346	153,677	3	663	73.2%	1.16E-107
scf00012		649		159,814	160,462	6	612	66.0%	1.02E-38
scf00012		455		174,278	174,716	1324	1778	78.0%	3.56E-95
scf00012		409		180,259	180,650	1370	1778	78.9%	4.06E-88
scf00012		1113		230,971	232,083	3	1105	85.0%	0
scf00012		938		389,698	390,635	6	889	65.2%	1.97E-60
scf00012		700		399,287	399,986	6	663	66.5%	2.57E-46
scf00012		205		424,250	424,046	1410	1614	83.3%	7.36E-47
scf00012		353		430,032	429,681	1426	1778	76.0%	1.17E-50
scf00012		342		430,554	430,222	1324	1665	76.1%	2.74E-52
scf00012		1120		502,183	503,298	1	1120	85.4%	0
scf00012		992		535,999	535,008	6	907	63.9%	4.96E-49
scf00012		700		596,817	596,118	6	663	65.9%	3.57E-38
scf00012		1110		665,804	666,913	1	1105	85.2%	0
scf00012		695		682,086	681,392	6	663	66.9%	4.96E-49
scf00012		385		716,677	717,061	6	372	71.2%	2.11E-47
scf00012		491		724,892	725,382	495	936	65.6%	4.96E-30
scf00012		1149		738,774	737,626	1	1120	79.2%	0
scf00012		977		760,183	761,159	6	923	63.1%	4.96E-30
scf00012		415		774,848	774,464	1364	1778	74.6%	1.97E-60
scf00012		564		791,730	791,230	1215	1778	69.2%	5.29E-55
scf00012		1105		869,720	870,819	1	1105	86.3%	0
scf00012		984		915,597	916,580	6	936	67.3%	4.64E-81
scf00012		237		954,226	953,991	1191	1427	84.8%	3.81E-63
scf00012		913		957,558	956,646	21	897	70.6%	2.57E-65
scf00012		434		961,128	960,709	1219	1652	80.5%	6.44E-92
scf00012		698		962,392	961,695	6	663	65.7%	1.02E-38
scf00012		954		985,715	984,762	6	907	63.5%	1.42E-30
scf00012		608		990,652	991,224	1171	1778	72.3%	6.88E-79
scf00012		199		1,010,491	1,010,688	465	663	75.1%	2.57E-27
scf00012		976		1,022,877	1,021,902	6	935	67.9%	1.33E-100
scf00012		589		1,082,067	1,082,655	37	616	75.4%	6.43E-111
scf00012		266		1,083,013	1,083,278	674	937	71.9%	4.65E-24
scf00012		345		1,157,369	1,157,025	1435	1778	75.7%	4.07E-50
scf00012		265		1,157,894	1,158,158	674	937	72.1%	1.09E-25
scf00012		1036		1,177,630	1,176,595	6	937	64.1%	4.96E-49
scf00012		265		1,179,759	1,179,495	674	937	73.1%	4.96E-30
scf00012		234		1,180,756	1,180,523	6	230	71.8%	2.57E-27
scf00012		388		1,183,735	1,183,348	6	375	68.4%	1.42E-30
scf00012		229		1,206,353	1,206,126	3	231	72.5%	3.13E-26
scf00012		229		1,218,306	1,218,079	3	231	73.9%	6.05E-29
scf00012		487		1,245,074	1,245,539	1169	1655	68.5%	2.75E-33
scf00012		458		1,257,141	1,257,598	1	449	83.4%	2.4E-135

Table 4 cont

Scaffold	Scaffold length	Sequence length	% scf	Hit end	Hit start	Query end	Query start	Pairwise identity	E value
scf00012		296		1,281,496	1,281,783	1483	1778	75.7%	5.66E-42
scf00012		408		1,288,213	1,288,613	1371	1778	72.7%	4.96E-49
scf00012		1332		1,298,230	1,299,509	442	1773	73.2%	0
scf00012		810		1,327,235	1,326,430	311	1120	85.1%	0
scf00012		288		1,328,193	1,327,906	1	288	89.6%	2.1E-104
scf00012		1134		1,372,898	1,371,765	1	1120	84.8%	0
scf00012		648		2,665,696	2,666,283	1100	1747	85.8%	0
scf00012		179		2,671,061	2,671,239	1597	1775	99.4%	2.56E-84
scf00012		976		2,881,365	2,882,340	34	937	64.9%	1.42E-49
scf00012		949		3,006,423	3,005,475	11	886	65.1%	5.29E-55
scf00012		691		3,401,387	3,402,077	6	663	67.1%	7.86E-34
scf00012		695		3,479,170	3,478,476	430	1120	85.8%	0
scf00012		347		3,480,607	3,480,261	1	347	89.1%	4.94E-125
scf00012 SUM	3,935,541	35793	0.9095%						
number of hits		58							
scf00095		234		475,877	476,110	231	6	72.3%	7.37E-28
scf00095		195		484,575	484,381	231	37	73.8%	3.82E-25
scf00095		492		1,094,204	1,094,695	937	493	65.9%	1.42E-30
scf00095		226		1,094,981	1,095,206	231	6	74.8%	7.86E-34
scf00095		700		1,483,284	1,483,983	663	6	67.4%	2.74E-52
scf00095		1192		1,491,641	1,492,832	1191	1	82.3%	0
scf00095		231		1,541,528	1,541,758	936	706	73.4%	8.97E-27
scf00095		691		1,542,092	1,542,782	663	6	65.0%	3.82E-25
scf00095		668		1,551,339	1,550,672	663	1	75.4%	4.06E-126
scf00095		265		1,558,575	1,558,839	935	674	70.6%	1.33E-24
scf00095 SUM	1,774,960	4894	0.2757%						
number of hits		10							
scf00097		990		193,085	192,096	923	6	62.7%	1.33E-24
scf00097		640		197,909	198,548	663	39	69.8%	1.73E-67
scf00097		700		258,127	257,428	663	6	66.8%	4.07E-50
scf00097		233		323,849	324,081	906	674	75.1%	1.52E-36
scf00097		473		324,118	324,590	889	465	65.3%	3.13E-26
scf00097		235		324,848	325,082	231	6	71.9%	7.37E-28
scf00097		697		508,542	507,846	663	6	66.7%	2.74E-52
scf00097		265		509,117	508,853	937	674	75.0%	4.35E-37
scf00097		1200		576,557	575,358	1199	1	84.2%	0
scf00097		1199		856,964	858,155	1199	1	79.5%	0
scf00097		179		1,304,062	1,303,891	1773	1595	80.1%	4.96E-30
scf00097		1195		1,308,810	1,307,626	1195	1	86.5%	0
scf00097		179		1,309,276	1,309,099	1773	1595	82.5%	1.25E-37
scf00097		422		1,416,974	1,416,553	1778	1368	75.3%	4.64E-62
scf00097		280		1,489,752	1,490,031	937	674	70.9%	6.05E-29
scf00097		199		1,490,342	1,490,540	663	465	74.8%	3.82E-25
scf00097		195		1,850,949	1,851,143	231	37	74.4%	8.97E-27
scf00097		863		2,225,747	2,226,609	1105	249	85.8%	0
scf00097	2,428,612	10144	0.4177%						
number of hits		19							

K-mer analysis in *P. axillaris* N

K-mer frequencies (measurements of the occurrence of each sequence motif k bases long in raw read data) are independent of assembly algorithm and thus an unbiased method to access the repetitive portion of a genome. Genome repetivity was assessed via 16- and 32-mer frequencies (Figure 10a and b) as in the tomato genome (Tomato Genome Consortium, 2012, Supplementary Figure 42) and were overlaid onto the tomato, potato and sorghum data (Figure 10c). The *P. axillaris* genome is larger than tomato and potato, so it should be expected that the slope indicates larger repetivity, and 16-mers occurring ≥10 times account for 50% of the genome, approximately twice the frequency in the smaller solanaceous genomes. However, interestingly for 16-mers occurring ≥ 30 times, the *Petunia* slope follows sorghum, even though the sorghum genome is only half the size of *Petunia*. It is notable that the top 10-20 most frequent k-mers are composites of AT, AAT, AG, A, C and AAG microsatellites. Larger abundant k-mers were also analysed, and for example some 54-mers and 64-mers have a few thousand repeats in the genomes, but only short tandem arrays with 10-20 copies were found in scaffolds suggesting that the assembly had collapsed some arrays of near-identical repeats.

In an attempt to identify larger tandem repeats in *P. axillaris*, 128-mers (Table 5) that were repeated more than 1001 times (in total 2347 of them) were subjected to a *de novo* assembly. They assembled into contigs with the largest of 1943 bp, 531 bp, 185 bp and 175 bp. The contig of 1943 bp has a total of 8,217 hits in *PaxiN* assembly and was found to be part of a *Gypsy* superfamily LTR-retroelement. There are some small duplications/tandem repeats within this large repeat and it is found as single repeat unit in most of the large scaffolds of *PaxiN* and also in the GenBank accession AY136628 of *P. hybrida*. The second 128-mer contig contains more of a repeat structure and hits to three large *Petunia* sequences in GenBank (AY136628, AB472856 and EF517793) and is present in 530 of the *PaxiN* contigs (0.8%) with high homology of near 100%. Dotplots of pairs of these scaffolds indicated that there is a larger repeated unit, up to 8 kb – with homology to a *Gypsy* superfamily retroelement with RNaseH, RT, INT, LTRs and other domains.

Table 5. Assembly using Geneious assembler of 128-mer identified in *PaxiN* raw reads.

	Unused reads	**Contigs >=128 bp**	**Contigs >=1000 bp**
Number	2	10	1
Minimum length (bp)	128	129	1,943
Median length (bp)		143	
Mean length (bp)	128	363	1,943
Max length (bp)	128	1,943	1,943
N50 length (bp)		1,943	
Number of contigs >= N50		1	1
Length sum (bp)	256	3,638	1,943

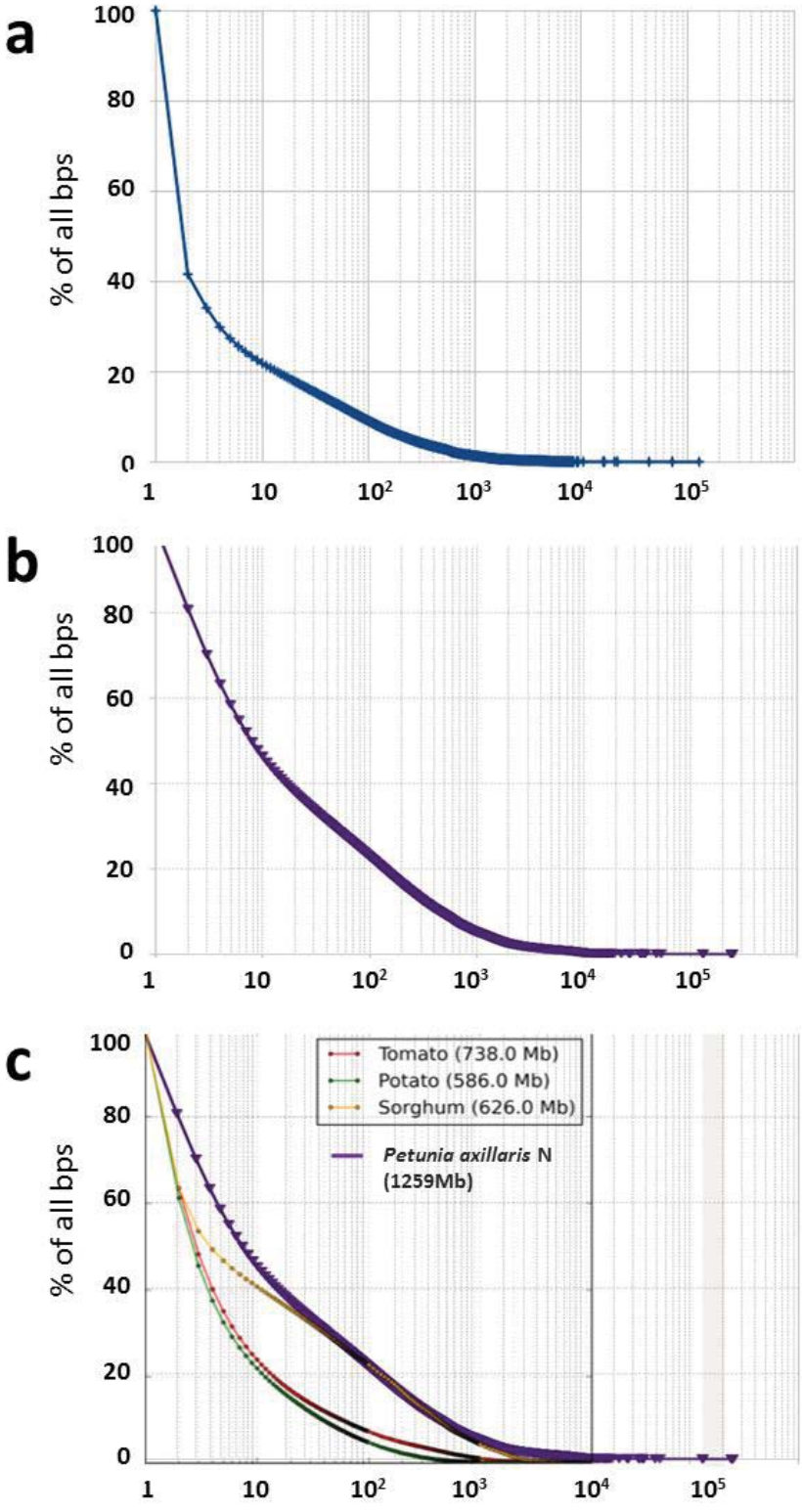


Figure 10: ***P. axillaris N*** **genome repetivity analyzed by k-mer frequency in raw reads**

a. 32-mer frequency.

b. 16-mer frequency

c. 16-mer data overlaid onto graph from tomato (Tomato Genome Consortium, 2012, Supplementary Figure 42). Genome size is given in Mb of assembled genomes.

Telomeres, tandem repeats and larger repeats

A few scaffolds included multiple copies of the plant telomere sequence TTTAGGG; some occurrences appear in intercalary positions. For example, *PaxiN* scf02211 was 12 kb long with telomere motifs at both ends totalling 2.5 kb, while scf47951 with a length of 2.5 kb was mostly composed of degenerate telomere arrays. In scf01230, telomere tracts up to 3 kb long were part of a longer tandem repeat unit of 11,246 bp, while scf03941 had a 180bp tandem repeat next to the telomere tract.

Another tandem repeat was extracted from *PaxiN* scf02038 and is 169 bp long. Many hundreds of smaller scaffolds were composed largely of a 169bp repeat (eg scf36935 is 16.5 copies over 2795 bp; scf01515 has 22 copies over 3778 bp; scf1920 has c.64 copies). Of larger scaffolds, scf01294 (409 kb) ends with 12 copies, scf00744 (498 kb) has 21 internal copies, scf00700 (826 kb) has 41 copies at the end over 7 kb (both orientations). Further, scf00420 has copies at the end, scf00451 an internal array, scf00286 multiple dispersed short arrays, and scf00207, scf00160, scf00128, scf00074 and scf00003 all have multiple copies. However, no scaffolds were found that could be related clearly to centromere structures, nor were any scaffolds candidates for giving the distribution of the fragment 4-18 on chromosomes (Figure 8 and above): the FISH results show a strong signal around the centromeres suggesting large numbers of a repetitive DNA motif. Two scaffolds could be candidates to locate around centromeres (e.g. Figure 11 for *PaxiN* scf00160), but further analysis including FISH and targeted cloning are needed to establish the position and distribution of these sequences, and identify them as centromere related.

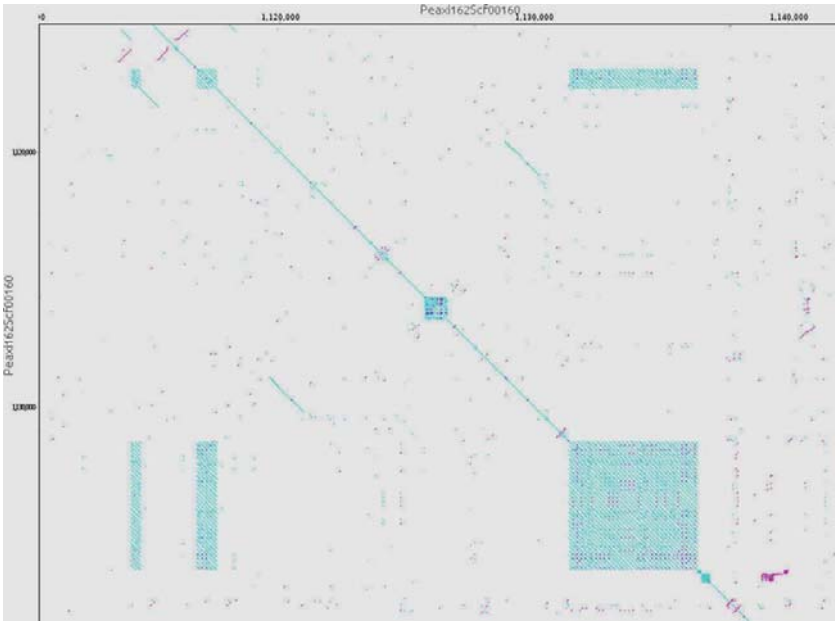


Figure 11: Tandem repeat structure.
Dotplot of *PaxiN* scf00160 nt1,110,000 to 1,140,000 with (c. 160 bp monomer) showing longer and shorter units.

Repeat analysis

As well as the k-mer analysis, genome-wide characterization of repetitive elements can use graph based clustering of DNA sequences from raw reads (Novak et al., 2010; 2013). The program RepeatExplorer was run using unassembled reads. For *P. axillaris N* 3,833,689 reads were selected

from an Illumina run, and for *P. inflata S6* 3,229,900 reads were examined. They resulted in 195,174 clusters using 2,488,367 reads for *P. axillaris,* about 65% of the analysed data pool (Figure 12a) and for *P. inflata*, 198,865 clusters using 2,511,217 reads, about 78% of the analysed data pool (Figure 12B); the remaining were not clustered. No cluster in either genome dominated the analysis, with the top 5 clusters representing 3.6% in *PinfS6* and 3.3% in *PaxiN* of the total and then declining in frequency gradually over several hundred motifs (Figure 12), contrasting with other genomes where a high proportion of repeats are represented by a small number of clusters (e.g. cacao, Sveinsson et al., 2013). RepeatExplorer depicts clusters graphically as connected dots; protein domains of transposable elements are colour coded. The total number of base pairs, reads and genome proportion are calculated; in addition the hits to known repeats present in the repeat masker database are identified. In Figure 13 the most frequent clusters of *PaxiN* and two most frequent clusters of *PinfS6* are shown. The most striking feature is that clusters are made up of a varied composition of sequence types, including LTR and low complexity repeats. The programme identified many further complex clusters, enriched in degenerate LTR-*Gypsy* and LTR *Copia* elements. Tandem repeats, rDNA and simple repeats were included in a few clusters.

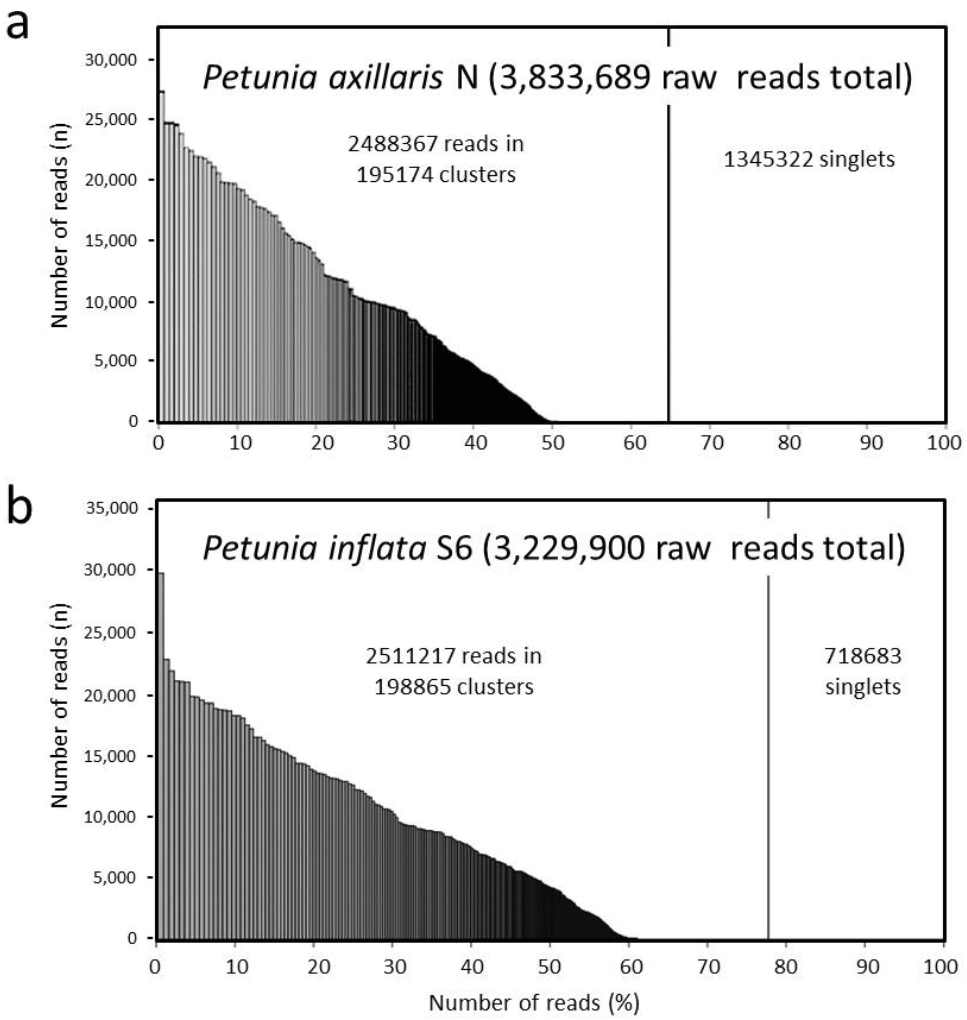


Figure 12: **Summary of Repeat Explorer (Novak et al., 2010, 2013) analysis using 3 million randomly selected raw reads a.** from *P. axillaris* N *and* ***b.*** *P. inflata* S6.

The most common clusters are not greatly more abundant than subsequent clusters. The top 350 clusters (all >0.01% proportion of the genome) represented 65 to 78% of all repeat clusters and include 50% (*P. axillaris* N) and 60% (*P. inflata* S6) of the genome.

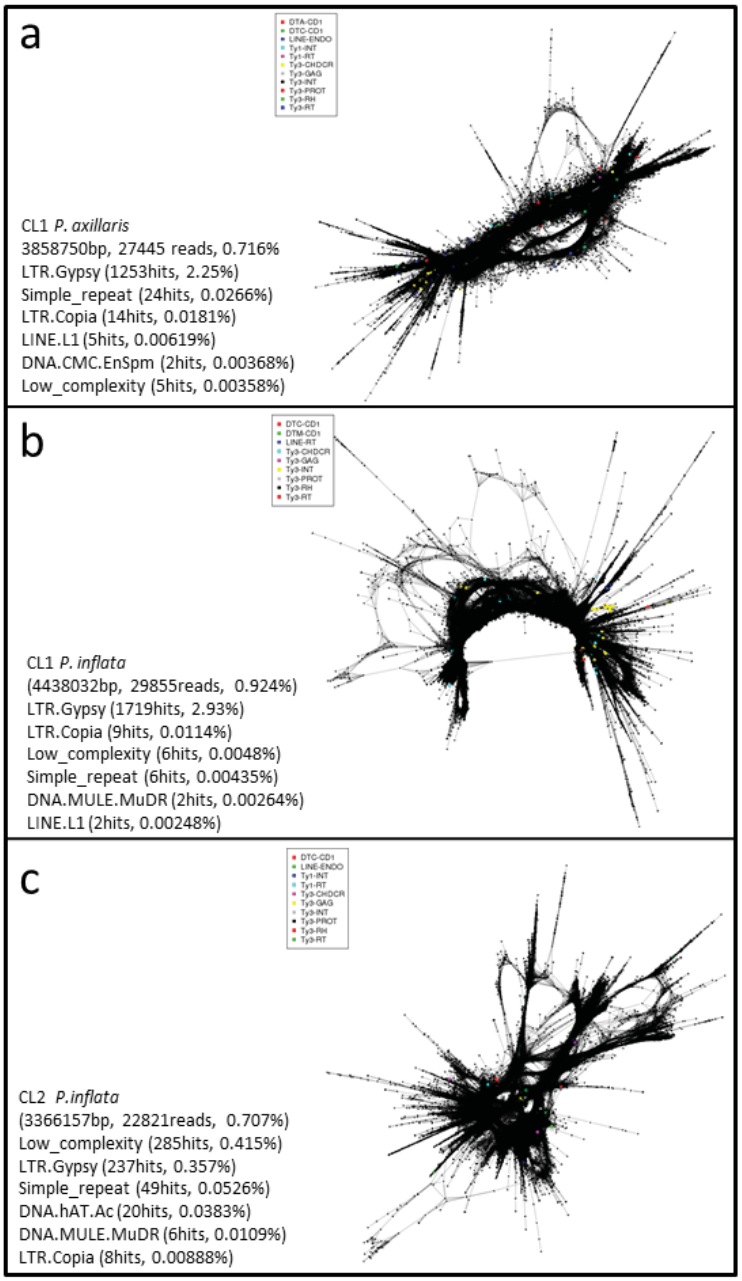


Figure 13: Most frequent repeat clusters CL1 of *P. axillaris* N (A) and CL1 and CL2 of *P. inflate* S6 (B, C). Output from Repeat Explorer (Novak et al., 2010, 2013). The clusters are displayed as composites containing low complexity and degenerate LTR-transposable elements with no abundant retroelement protein motifs identified (shown by the few coloured dots). Each cluster is less than 1% of the genome.

DISCUSSION

Members of Solanaceae are known to be suitable hosts for a number of plant viruses. More than 150 plant viruses have been shown to infect *P. hybrida* (Engelmann and Hamacher, 2008). Plant genome sequencing indicates that several *Caulimoviridae* invaded the genome of their host in the course of viral infection (Tomato Genome Consortium, 2012; Kim et al., 2014; Geering et al., 2014). The identified single insertion sites within the *P. axillaris* N and *P. inflata* S6 genome respectively are suggestive of provirus stages typical for retroviruses and endogenous retroviruses. Integration of retroviruses requires an integrase and is enhanced in actively dividing cells (Young et al., 2013). The high regeneration potential of Solanaceae species like *Petunia* might promote capture of plant pararetroviruses of the family of *Caulimoviridae* that lack an integrase from the *Petunia* genome. De-differentiation of pararetrovirus infected somatic cells as been happening during callus proliferation and plant regeneration might offer gateways for invasion followed by vertical transmission (Hohn et al., 2008). Our analysis provides insight into the elements which have been manifested in the male

and female germline lineages (Schmidt et al., 2012) and we expect that the frequency of pararetroviral invasion in chromosomal DNA of somatic cells is much higher. We have speculated in the past (Richert-Pöggeler and Schwarzacher, 2009; Staginnus and Richert-Pöggeler, 2006) that the close proximity of endogenous pararetroviruses and LTR-retroelements points to a co-evolution of these two similar elements as well as to their integration and silencing mechanisms. The available sequence data of two petunia species prove *Petunia* as an ideal model system to study endogenous pararetrovirus co-evolution within solanaceous hosts as well as potential functions besides being an infectious entity. Furthermore EPRVs can be suitable markers for monitoring dynamic processes during genome hybridisation as happened during generation of *P. hybrida*. Both genomes contain insertions of regulatory sequences (*PaxiN* scf01628, *PinfS6* scf00235 and *PinfS6* scf00276) from the untranslated region of the PVCV genome that await further analysis of their functionality in gene expression and reverse transcription respectively. Recent studies by Mushegian and Elana (2015) discuss molecular functions for the host provided by integration of pararetroviral movement protein (MP)-like sequences. Based on their phylogenetic analyses and known MP functions in macromolecule trafficking they propose a possible role of integrated MP in control of tissue differentiation. Indeed, in *P. axillaris N* integrated MP sequences of scf01628 and scf03256 showed the highest degree of conservation and thus may be active in the host rather than viral context.

The pericentromeric localization and array structure of integrated PVCV sequences is similar in both petunia species. However, the preservation and copy number of these endogenous viral sequences is higher in *P. axillaris* compared to *P. inflata*. That was also true with regard to diversity of EPRVs. Thus *P. axillaris* seems to be a more permissive host for EPRV invasion, so we suggest that there are several factors controlling EPRV invasion and preservation that differ even between related species, and probably even cultivars, including the organization in the genome, presence of miRNA, DNA methylation and histone modifications; such factors may be under evolutionary selection depending on disease pressure and consequences for the different species.

Our analysis indicates that *Petunia* genomes are rich in repetitive DNA and the K-mer analysis of *P. axillaris* N indicates more repeats are present than in tomato and potato (Tomato Genome Consortium, 2012). However, when comparing total genome size, the amount of repeats in both the Pa*xi*N and *Pinf*S6 assemblies with about 60-65% (Table 6) is low for genomes of 1.4Gb. In particular LTR-retroelements are unusually low both in total DNA bps and number full length elements that we were able to identify. We found about 5,000-6,000 full elements with roughly equal numbers of *Gypsy* and *Copia* superfamily retroelements. This number is similar to tomato with a genome assembly of 740Mbp, but a relatively large LTR-retrolement component. It is in contrast to hot pepper (*Capsicum annuum*) with the largest solanaceous genome (more than 3Gb) so far sequenced where large numbers of LTR-retroelements in particular *Gypsy* superfamily elements make up 70% of the repetitive DNA fraction, are responsible for the genome expansion and conversion of euchromatin into heterochromatin (Kim et al., 2014). *Nicotiana* species also have relatively few identified LTR-retroelements (Table 6), and here *Copia* superfamily retroelementss make the difference between *N. tomentosiformis* and *N. sylvestris* (Sierro et al., 2013). Interestingly, *N. tomentosiformis* and *N. sylvestris* can also be distinguished by their EPRV composition (Gregor et al., 2004). The difference to the smaller tomato genome however, is mainly attributed to shorter repeats (listed under 'others' in Table 6) that represent almost 30% of the repetitive DNA in the about 2.4Gb *Nicotiana* genomes. This situation is mirrored in the much small cucumber genome (244Mb) where transposable elements are relatively low in abundance, but satellite tandem repeats make up almost half of the repetitive DNA and are concentrated at centromeres and telomeres as evidenced by FISH (Huang et al., 2009). A

similar situation is found in *Brassica rapa* (Brassica rapa Genome Consortium, 2011) and *Beta vulgaris* (Dohm et al., 2014) where satellites repeats have been found (with only two families abundant, locating at centromeres, in *Brassica*, Harrison and Heslop-Harrison, 1994); however, the larger genome sizes of 500-600 Mbp can be attributed to the LTR-retroelement fractions.

The relative low proportion of repeats, make in turn the gene and low copy sequence space relative larger and would correspond to the lower fragmentation found in *P. axillaris* N after the last triplication event (paleohexaploidisation) in comparison to tomato and potato (see Supplementary Note 5).

In *Petunia*, FISH has indicated a concentration of retro-elements like sequences around the centromeres (this study) as well as their dispersion throughout the chromosomes (Richert-Pöggeler and Schwarzacher, 2009), but this is not reflected in the assembly, as is normal with shotgun sequence data; even the relatively high proportion of PacBio and mate-pair reads apparently spanning the length of retroelements. However, DNA transposons are common and were found at a much higher frequency than in *Nicotiana* and *Solanum* (Table 6 and Supplementary Note 3). Interestingly, the literature does not report tandem satellite repeats in *Petunia* and our repeat searches within the assembly have not found typical 180bp or 340bp repeats that wrap around nucleosomes in a specific manner (see Heslop-Harrison and Schwarzacher, 2013). However, shorter repeats of about 60bp have been found, as well as some longer repeats of 500-1000bp. In addition, many mixed repeat family clusters incorporating retroelements, simple sequence repeats and low complexity repeats were identified by the RepeatExplorer algorithm, but none present a substantial percentage in the genome in contrast to the RepeatExplorer data in cacao (Sveinsson et al., 2013).

It is therefore apparent that the repeat structure of *Petunia* differs from other species of Solanaceae so far analysed in detail and indicates a high degree of genome plasticity. Genome size alone might however not dictate the distribution, type and amount of repetitive elements. It is notable that *Petunia* chromosomes (Table 6), with an average of 200Mb per chromosome (three times that of tomato or potato), are relatively large for the overall genome size that is distributed over 7 rather than the more common 12 pairs of chromosomes in the family that form the related x=12 clade (Saerkinen et al., 2013). This has consequences for chromosomal organisation, recombination and homogenisation events and together with DNA transposon frequency and the presence of EPRVs might have an effect on the overall genome organisation.

The genomic sequence data will be seminal for investigating interactions of the identified reverse transcribing elements of the family of *Caulimoviridae, Metaviridae* and others both *in planta* as well as at the single cell level and in culture. The identified diversity and abundance of the polymerase motif of viral retroelements raises questions about possible functions of reverse transcription in genome maintenance and/or speciation. Genome sequence data reveal existence of petuvirus-like sequences not only in the solanaceous plant family but also in woody plants, for example in the family of *Rutaceae* (Yang et al., 2003; Roy et al., 2014). The combination of horizontal and vertical transmission among multiple members of *Caulimoviridae* probably contributed to the abundancy of EPRVs within angiosperms (reviewed by Teycheney and Geering, 2010, Geering et al., 2014). Whereas information about the contributing genomes is available, participating vectors mediating transfer of the episomal forms still need to be elucidated. The effects of genome hybridization during generation of *P. hybrida* with *P. axillaris* and *P. inflata* as parental crossing partners, on repeat and in particular EPRV evolution, activation and function can now be studied in greater detail. Thus will also contribute to the general understanding of mechanisms involved in lateral DNA transfer.

Table 6: Comparison of repeat content, total genome and chromosome sizes in *Solanaceae* and selected eudicot species.

Species	Repeats	DNA transposons		LTR retroelements and retrotransposons		non-LTR retroelements (SINES, LINES)		others (satellites, unknown, low complexity)		sequence assembled	1C DNA content	Chromo -some number	DNA/chr
	% assembled genome	bp	%	bp	%	bp	%	bp	%	bp	Mbp	n=	Mbp
Cucumis sativus 1)	24.01	2,808,075	1.24	23,622,636	10.43	3,961,988	1.75	25,762,300	10.58	243,500,000	367	7	52
Brassica rapa 2)	44.79	15,518,826	3.2	131,612,046	27.14	15,925,293	3.28	54,174,500	11.17	485,000,000	560	10	56
Beta vulgaris 3)	42.3	19,820,000	3.33	122,670,000	20.59	32,190,000	5.40	77,320,000	12.98	595,744,681	730	9	81
Solanum lycopersicum 4)	68	6,050,581	0.86	459,739,604	61.77	4,089,807	0.55	31,421,760	4.26	737,600,000	900	12	75
Solanum tuberosum 4)	62.20	6,543,927	1.2	311,628,974	54.35	5,796,327	1.16	32,394,740	5.53	585,800,000	844	12	70
Petunia axillaris N 5)	63.08	65,589,038	5.21	508,788,466	40.41	29,284,495	2.33	190,486,700	15.13	1,259,000,000	1380	7	197
Petunia inflata S6 5)	59.22	59,714,447	4.64	475,871,680	36.98	38,215,801	2.97	188,141,800	14.63	1,286,000,000	1430	7	204
Nicotiana tomentosiformis 6)	74.84	22,593,004	1.34	882,169,158[8)]	52.21[8)]	8,078,343	0.48	571,894,844	20.33	1,689,000,000	2360	12	197
Nicotiana sylvestris 6)	71.95	33,621,895	1.51	1,082197,020[9)]	48.65[9)]	9,869,117	0.44	703,763,729	21.34	2,222,000,000	2680	12	223
Capsicum annuum 7)	76.36	165,894,072	5.41	1,780,527,144	58.11	47497259	1.55	345,248,200	11.29	3,058,000,000	3480	12	290
Capsicum chinense 7)	79.55	197,445,015	6.69	1,649,035,494	55.84	69182970	2.34	433,647,200	14.68	2,954,000,000	3140	12	262

1) Huang et al. (2009);
2) Brassica Genome Consortium (2011);
3) Dohm et al. (2014);
4) Tomato Genome Consortium (2012);
5) this study numbers taken from the repeatmasker analysis used for the assembly (see supplementary Note 1);
6) Sierro et al. (2013);
7) Kim et al. (2013)
8) this number contains 666,441,913 bp (39.13%) LTR-retroelements and 220,727,245 bp (13.08%) non-identified retrotransposons,
9) this number contains 851,543,954 bp (38.32%) LTR-retroelements and 230,653,066 bp (10.33%) non-identified retrotransposons

METHODS

Identification of PVCV-like sequences in the Petunia genomes

Nucleotide sequences of PVCV accession U95208.2 as well as amino acid (aa) sequences of PVCV ORF 1 (accession NP_127504.1) were compared with scaffolds of *PaxiN* and *PinfS6*respectively deposited on the Blast Server: http://petuniasp.sgn.cornell.edu/blast/blast.html using Blast settings without filter and tBlastN respectively. Due to various degrees of sequence degradation distinct thresholds were set as indicated in Tables 1 and 2.

Determination of EPRV diversity

Pol regions of selected *Caulimoviridae* were compared with scaffolds of *PaxiN* and *PinfS6* respectively using the Blast server as above. The thresholds were set at > 200 amino acid (aa) alignment length for all elements and >60% aa identity for *Petu*- and *Florendoviruses*, as well as >45% for all other *Caulimoviridae*. The following accessions and pol regions were incorporated in the search:
NP_569141.1 for PVCV, *Petuvirus*, ORF1, aa 1425..1804,
BAO53400.1 for *Cauliflower mosaic virus* (CaMV), *Caulimovirus*, isolate JPNS2, ORF 5, aa 285..671,
Q89703.1 for *Cassava vein mosaic virus* (CsVMV), *Cavemovirus*, ORF 3, aa 237..637,
YP_007761644.1 for *Rose yellow vein virus* (RYVV), unassigned genus, ORF 3, aa 446..837,
AHN13810.1 for *Aristotelia chilensis vein clearing virus* (ArCVCV), unassigned genus, putative ORF, aa 33..412,
AHA62452.1 for *Banana streak OL virus* (BSOLV), *Badnavirus*, ORF 3 partial, aa 1..413,
NC_001914 for *Rice tungro bacilliform virus* (RTBV), ORF3, aa 1225..1612,
Vitis vinifera B virus (VvinBV_compAsc1), *Florendovirus*, Geering et al. 2014, ORF 1, aa 965..1346,
and *Solanum tuberosum virus* (StubV_scSt1), *Florendovirus*, Geering et al. 2014, ORF 1, aa 1433..1816.
Alignment of sequences has been done using ClustalW within the *MEGA* version 5 software package (Tamura et al., 2011). Identified hits were manually edited to remove overlapping hits and the sequence with the higher score was selected.

*DNA similarity analyse*s

The following PVCV sequences were used for alignment using Clustal W in the *MEGA* version 5 (Tamura et al., 2011) with default settings applied: scaffolds of *P. axillaris* with single insertion of PVCV coding sequences (*PaxiN*_00097, *PaxiN*_00254, *PaxiN*_00447, *PaxiN*_00560, *PaxiN*_00674 and *PaxiN*_00911), scaffolds of *P. inflata* with single insertion of PVCV coding sequences (*PinfS6*_00844, *PinfS6*_01099, *P.inf6S*_01671 and *P.inf6S*_00276) and with double insertions (*P.inf6S*_00235a and *PinfS6*_00235b). Episomal PVCV sequences, isolated from *N. glutinosa* (infectious virus; AAK68664) have been also included in the analysis; LTR-retrotransposon from *Drosophila melanogaster* (Accession AAA70219) has been used as outgroup.

Tandem repeat and retroelement analysis

Basic analysis of the assemblies were performed on Ubuntu Linux 13.10, with Geneious version 7.1.4 (and earlier) by Biomatters (Kearse et al., 2012; available from http://www.geneious.com/). K-mer analyses were performed using Jellyfish version 2.1.3 (Marcais and Kingsford, 2011). Other programmes to search for repeats included LTR-STRUC (McCarthy et al., 2003), LTR finder (http://tlife.fudan.edu.cn/ltr_finder) and RepeatExplorer (Novak et al., 2010; 2013).

Fluorescent in situ *hybridization*

Probe labelling, chromosome preparation and *in situ* hybridization followed the procedure of Schwarzacher and Heslop-Harrison (2000). The 5S rDNA probe was a 410 bp fragment from the clone pTa794 (Gerlach and Dyer, 1980) containing the 5S rDNA repeat unit of *Triticum aestivum*. Three viral probes that, in combination, cover most of the sequence of an infectious chromosomal PVCV copy were produced using the SacI subclone l5-7 from lambda clone 5 (Richert-Pöggeler et al., 2003) while fragment 4-18 (1,233bp long) originates from lambda clone 4 (GenBank AY333912) and contains part of a *Gypsy* gag-pol region. Probes were labelled with biotin-11-dUTP (Roche) or digoxigenin-11-dUTP (Roche) by PCR using M13 forward and reverse sequencing primers for cloned sequences or template specific primers for virus (see Richert-Pöggeler et al., 2003) and 4-18 fragement (forward primer #34935, TGG TAG CGA CTT GTA TCG AGC, reverse primer #34936, TCA ACA AGT AAG CCA CGC AGG, nt 4483-5715 .

Root tips from young plants, *P.axillaris* and *P. inflata* (both from the University of Nottingham collection received in 2001) were fixed with 96% ethanol:glacial acetic acid (3:1) after treatment with 0.2 M 8-hydroxyquinoline for 3-4 h. Chromosome preparations were made following proteolytic digestion with cellulase and pectinase, treated with RNase and fixed in 4% paraformaldehyde (see Schwarzacher and Heslop-Harrison, 2000)

The probe mixtures contained 100-200ng labelled probes, 50% (v/v) formamide, 20% (w/v) dextran sulphate, 2x SSC, 0.025µg of salmon sperm DNA and 0.125% SDS (sodium dodecyl sulphate) and 0.125mM EDTA (ethylenediamine-tetraacetic acid). Chromosomes and 40µl of probe were denatured together and allowed to hybridize overnight at 37°C. Post-hybridization washes were at 42°C in 20% formamide and 0.1xSSC, giving a stringency of 80±85%. Detection of hybridization sites was carried out with 4µg/ml Fluorescein-conjugated anti-digoxigenin (Roche) and 2µg/ml Alexa 495-conjugated streptavidin (Molecular Probes). Chromosomes were counterstained with 4µg/ml DAPI (4´,6-diamidino-2-phenylindole) and mounted in CitifluorAF. Slides were analysed with a Zeiss Axioplan2. Fluorescent microscope and images captured with an Optronix S97790 cooled CCD camera. Overlays of hybridization signal and DAPI images were prepared with Adobe Photoshop CS4 using only cropping and functions that treat all pixels equally. For some slides, after photographing and noting down the coordinates of the metaphase, the first probing was washed away during a repeat denaturation step and a second probing was carried out.

ACKNOWLEDGEMENTS

We thank and the Botanic Garden University of Leicester for growing petunia plants for root tip collection and George Heslop-Harrison for help with the retroelement annotation. We are grateful for critical reading and helpful comments by J. Schoelz, University of Missouri, USA. Part of the work was supported by EU-Framework V Paradigm.

AUTHOR CONTRIBUTIONS

KRP analysed the PVCV and EPRV in the genome assemblies; PHH and TS analysed the tandem repeats and retroelements; KRP and TS performed and analysed the FISH experiments; all authors contributed to the MS.

REFERENCES

Brassica rapa genome consortium. (2011). The genome of the mesopolyploid crop species *Brassica rapa.* Nature **43**: 1035-1039. doi: 10.1038/ng.919

Dohm, J.C., Minoche, A.E., Holtgrawe, D., Capella-Gutie´rrez, S., Zakrzewski, F., Tafer, H., Rupp, O., Sorensen, T.R., Stracke, R., Reinhardt, R., Goesmann, A., Kraft, T., Schulz, B., Stadler, B.F., Schmidt, T., Gabaldo, T., Lehrach, H., Weisshaar, B. and Himmelbauer, H. (2014). The genome of the recently domesticated crop plant sugar beet (Beta vulgaris). Nature **505**: 546-549. doi: 10.1038/nature12817.

Engelmann, J. and Hamacher, J. (2008). Plant Virus Diseases: Ornamental Plants. In: Mahy, B. W. and van Regenmortel, M.H. V. (Eds.) Desk Encyclopedia of Plant and Fungal Virology Vol. 4, Acad. Press. Elsevier pp. 436-458.

Geering, A.D.W., Maumus, F., Copetti, D., Choisne, N., Zwickl, D.J., Zytnicki, M., McTaggart, A.R., Scalabrin, S., Vezzulli, S., Wing, R.A., Quesneville, H., Teycheney, P.-Y. (2014). Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. Nature Communications 5:5269. doi: 10.1038/ncommms6269.

Gerats, T. (2009). Identification and exploitation of Petunia transposable elements: a brief history. In Petunia: Evolutionary, Developmental and Physiological Genetics, 2nd edition. Eds. Gerats, T. and Strommer, J. Springer, New York. doi: 10.1007/978-0-387-84796-2.

Gregor, W., Mette, M.F., Staginnus, C., Matzke, M.A., Matzke, A.J. (2004). A distinct endogenous pararetrovirus family in Nicotiana tomentosiformis, a diploid progenitor of polyploid tobacco. Plant Physiol. 13: 1191-**119**9.

Hansen, C.N., Harper, G. and Heslop-Harrison, J.S. (2005). Characterization of pararetrovirus-like sequences in the genome of potato (*Solanum tuberosum*). Cytogenetic and Genome Research **10**: 559-565. doi: 10.1159/000084989.

Hansen, C.N. and Heslop-Harrison, J.S. (2004). Sequences and phylogenies of plant pararetroviruses, viruses and transposable elements. Adv. Bot. Res. **41**: 165-193.

Heslop-Harrison, J.S. and Schwarzacher, T. (2013). Nucleosomes and centromeric DNA packaging. Proc. Nat. Acad. Sci. USA.http://www.pnas.org/content/110/50/19974.full.pdf+html http://dx.doi.org/10.1073/pnas.1319945110 -

Hohn, T., Richert-Pöggeler, K.R., Staginnus, C., Harper, G., Schwarzacher, T., Teo, C.H., Teycheney, P.-Y., Iskra-Caruana, M.-L. and Hull, R. (2008). Evolution of integrated plant viruses. In: Roossinck, M.J. (ed.) Plant Virus Evolution. Springer Berlin Heidelberg pp. 53-81.

Huang, S. et al. (2009). The genome of the cucumber, *Cucumis sativus* L. Nature Genetics **41**: 1275-1281. doi: 10.1038/ng.475.

Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences **8**: 275-282.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649. doi: 10.1093/bioinformatics/b199.

Kim et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nature Genetics **46**: 270-278. doi: 10.1038/ng.2877.

King, A.M.Q., Adams, M.J., Carstens, E.B. and Lefkowitz, E.J. (2012). Virus taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses, Elsevier Academic Press, San Diego.

Kriedt, R.A., Cruz, G.M.Q., Bonatto, S.A. and Freitas, L.B. (2014). Novel transposable elements in Solanaceae: Evolutionary relationships among Tnt1-related sequences in wild petunia species. Plant Mol. Biol. Rep. **32**: 142-152. doi: 10.1007/s11105-013-0626-8.

Marcais, G. And Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics **27**: 764-770. doi: 10.1093/bioinformatics/btr011.

Matsubara, K., Kodama, H., Kokubun, H., Watanabe, H. and Toshio, A. (2005). Two novel transposable elements in a cytochrome P450 gene govern anthocyanin biosynthesis of commercial petunias. Gene **358**: 121-126.

McCarthy, E.M. and McDonald, J.F. (2003). LTR_STRUC: a novel search and identification program for LTR retrotransposons. Bioinformatics **19**: 362–367.

Mishiba, K.I., Ando, T., Mii, M., Watanabe, H., Kokubun, H., Hashimoto, G. and Marchesi, E. (2000). Nuclear DNA content as an index character discriminating taxa in the genus etunia sensu Jussieu (Solanaceae). Annals of Botany **85**: 665-673. doi: 10.1006/anbo.2000.112.

Mushegian, A.R. and Elena, S.F. (2015). Evolution of plant virus movement proteins from the 30K superfamily and their homolgs integrated in plant genomes. Virology **476**: 304-315.

Noreen, F., Akbergenov, R., Hohn, T., Richert-Pöggeler, K.R. (2007). Distinct expression of endogenous *Petunia vein clearing virus* and the DNA transposon *dTph1* in two *Petunia hybrida* lines is correlated with differences in histone modification and siRNA production. Plant J. **50**: 219-229.

Novak, P., Neumann, P. and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics **11**: 378. doi:10.1186/1471-2105-11-378.

Novak, P., Neumann, P., Pech, J., Steinhaisl, J. and Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. Bioinformatics **29**: 792-793.

Richert-Pöggeler, K.R., Noreen, F., Schwarzacher, T., Harper, G. and Hohn, T. (2003). Induction of infectious petunia vein clearing (pararetro) virus from endogenous provirus in petunia. EMBO J. **22**: 4836–4845.

Richert-Pöggeler, K.R. and Schwarzacher, T. (2009). Impact of etroelements in shaping the petunia genome in petunia: Evolutionary, Developmental and Physiological Genetics, 2nd edition. Eds. Gerats, T. and Strommer, J. Springer, New York. doi: 10.1007/978-0-387-84796-2.

Richert-Pöggeler KR and Shepherd, R.J. (1997). Petunia vein clearing virus: a plant pararetrovirus with the core sequences for an integrase function. Virology **236**: 137-146.

Roy, A., Shao, J., Schneider, W.L., Hartung, J.S. and Brlansky, R.H. (2014). Population of endogenous pararetrovirus genomes in Carrizo citrange. Genome Announc. **2**: e01063-13. doi:10.1128/genomeA.01063-13.

Saerkinen, T., Bohs, L., Olmstead, R.G. & Knapp, S. (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evolutionary Biology **13:** 214.

Schmidt, A., Schmid, M.W. and Grossniklaus, U. (2012). Analysis of plant germline development by high-throughput RNA profiling: technical advances and new insights. Plant J. **70**: 18-29.

Schwarzacher, T. and Heslop-Harrison, J.S. (2000). Practical in situ hybridization. BIOS (Oxford).

Sierro, N., Battey, J.N.D., Ouadi, S., Bovet, L., Goepfert, S., Bakaher, N., Peitsch, M.C. and Ivanov, N.V. (2013). Reference genomes and transcriptomes of Nicotiana sylvestris and Nicotiana tomentosiformis. Genome Biology **14**: R60.

Staginnus, C. and Richert-Pöggeler, K.R. (2006). Endogenous pararetroviruses: Two-faced travelers in the plant genome. Trends Plant Sci. **11**: 485–491.

Sveinsson, S., Gill, N., Kane, N.C. and Cronk, Q. (2013). Transposon fingerprinting using low coverage whole genome shotgun sequencing in Cacao (Theobroma cacao L.) and related species. BMC Genomics **14**: 502.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution **28**: 2731-2739. doi:10.1093/molbev/msr121.

Teycheney, P.-Y. and Geering, A. (2011). Endogenous viral sequences in plant genomes. In: Caranta, C., Aranda, M.A., Tepfer, M. and Lopez-Moya, J.J. (eds.) Recent Advances in Plant Virology. Caister Academic Press, Norfolk, UK pp. 343-362.

Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635-641. doi: 10.1038/nature11119.

Yang, Z.N., Ye, X.R., Molina, J., Roose, M.L. and Mirkov, T.E. (2003). Sequence analysis of a 282-kilobase region surrounding the citrus tristeza virus resistance gene (*Ctv*) locus in *Poncirus trifoliata* L. Raf. Plant Physiol. **131**: 482–492.

Young, G.R., Stoye, J.P., Kassiotis, G. (2013). Are human endogenous retroviruses pathogenic? An approach to testing the hypothesis. Bioessays **35**: 794–803.

Supplementary Note 3

Genome wide analysis of Petunia *dTPH* transposable elements in *Petunia axillaris* and *Petunia inflata*

Michiel Vandenbussche[1]*, Jan Zethof[2], Mattijs Bliek[3], Tom Gerats[2], Patrice Morel[1] and Ronald Koes[3]

[1]Laboratoire de Reproduction et Développement des Plantes, CNRS, INRA, ENS Lyon, UCBL, Université de Lyon, 69364 Lyon, France.
[2]Plant Genetics, Institute for Water and Wetland Research, Radboud University Nijmegen, 6525AJ Nijmegen, The Netherlands
[3]Department of Plant Development and (Epi)Genetics, Swammerdam Institute for Life Sciences, University of Amsterdam, 1098 XH Amsterdam, The Netherlands.
*for correspondance: Email: michiel.vandenbussche@ens-lyon.fr

Running title: *dTPH* transposable elements in Petunia

Abstract

Since its discovery in 1990, the small endogenous non-autonomous *hAT*-like *dTPH1* transposon that transposes at high frequency in the *Petunia* W138 line has been developed into an extremely powerful tool for forward and reverse genetics. Yet, the origin of this highly efficient mutagenesis system remains poorly understood, mainly because no *Petunia* genome sequence was available. Here we have mined the *PaxiN* and *PinfS6* genomes to better understand the origin of *dTPH1* and other related elements. We found that both species contain only a small number of *dTPH1* copies compared to W138, in agreement with the idea that *dTPH1* copy number has increased only recently, especially in the line leading to W138. In contrast, we found that the *dTPH7* element has vastly expanded reaching around ~200 copies in both species, while the other *dTPH*-like elements are present at much lower copy numbers. This suggests that different *dTPH* elements in *Petunia* may require different transacting factors for their mobility. Interestingly, comparison of *dTPH1* insertion loci found in *PaxiN* and *PinfS6* genomes with W138 provides independent evidence that both species indeed contributed to the hybrid nature of W138. Furthermore, we analyzed *dTPH1* distribution patterns in wild *Petunia axillaris* accessions collected all over Uruguay, showing that *dTPH1* transposition activity is very low in these natural populations. This confirms previous genetic studies, and further supports the hypothesis that a recessive *activator1* (*act1*) epiallele, the locus required for *dTPH1* transposition, got reactivated after the interspecific crosses that gave rise to *P. hybrida*. Finally, we report the discovery of a new member of the *dTPH* family, named *dTPH12*.

Introduction

Transposable elements or remains thereof constitute a major part of eukaryotic genomes. Even though they should probably be seen as molecular parasites living in the host genomes, they have an important role in evolution of host species as generators of genetic diversity (Feschotte and Pritham, 2007; Lisch, 2013). For the same reason they are being used in certain model species as tools to identify genes by forward genetics (transposon-tagging) or to establish the function of genes by reverse genetics. Because of the genetic heterogeneity of *P. hybrida* cultivars, detailed genetic maps are lacking, which makes map-based approaches for the isolation of (mutated) genes cumbersome. Hence, the isolation of *Petunia* genes that are defined by mutations relied to a large extent on the use of transposons and specific *Petunia* lines, like W138, in which transposons are highly active.

Genetically unstable petunia mutants, which are typical for transposon insertion alleles, were first documented in 1970s (Bianchi et al., 1978), but presumably occurred many times before during breeding of *P. hybrida* varieties, which added to the broad pallet of flower colors of *P. hybrida* varieties (Kroon et al., 1994; Faraco et al., 2014). A particularly important event was the appearance of a novel mutant, having white flowers with numerous red revertant spots due to a transposon insertion in *ANTHOCYANIN1* (*AN1*), among progeny of an inbred line R27, with red colored flowers (Doodeman et al., 1984; Spelt et al., 2000). The new mutant was maintained by inbreeding as line W138. Whereas spontaneous mutations occur very rarely in R27, new (unstable) mutations are found at high incidence among W138 progeny (Doodeman et al., 1984; van Houwelingen et al., 1998). This suggests that in W138, or the R27 plant(s) from which W138 originates, transposition frequencies were strongly upregulated for reasons that remain unknown. The unstable $an1^{W138}$ mutation that founded the line W138, and the large majority of new mutations that subsequently arose in W138 progeny are due to insertions of a small DNA-type (class II) transposon of only ~284 bp (van Houwelingen et al., 1998; Spelt et al., 2000). As this element is too small to encode a transposase and to transpose autonomously, it was named *defective Transposon Petunia hybrida 1* (*dTPH1*) (Gerats et al., 1990). *dTPH1* elements belong to the *hobo, Ac*, *Tam* (*hAT*) family of transposons, because (i) *dTPH1* generates, like other *hAT* elements, a target site duplication of 8 bp and because (ii) the *dTPH1* ends are 12-bp inverted repeats with similarity to the terminal inverted repeats (TIR) of *ACTIVATOR* (*AC*) from maize and other *hAT* elements (Gerats et al., 1990).

Genetic analyses indicated that the excisions of the *dTPH1* copy in the $an1^{W138}$ and the $dfrC^{W138}$ allele required the presence of a dominant allele at the *ACTIVATOR1* (*ACT1*) locus. Surprisingly, *ACT1* mapped to chromosome 1 in all *P. hybrida* lines that were analyzed (Huits et al., 1995). As these lines derived from distantly related cultivars, this suggested that *ACT1* is immobile. The molecular nature of *ACT1* is unknown. One possibility is that *ACT1* is a transposase gene that resides in an (previously) autonomous *dTPH1* element with a defect that rendered it immobile. Alternatively, *ACT1* may be a host gene that is required for transposition. Most of the tested *P. hybrida* lines, which derived from distantly related cultivars, contain a dominant *ACT1* allele, suggesting that it originates from the parental species, whereas only a small group of related lines contain a recessive *act1* allele, suggesting that this allele results from a loss of function mutation during breeding of some *P. hybrida* cultivars. Although the genomes of *P. axillaris* and *P. inflata* contain sequences that hybridize to *dTPH1* (Huits et al., 1995), none of the accessions that were tested appeared to contain a dominant *ACT1* allele, which raises questions about the origin of *ACT1* in *P. hybrida*. One possibility that was put forward is that the parental species contain a recessive *act1* epiallele, which got reactivated after the interspecific crosses that gave rise to *P. hybrida*.

Despite the fact that the identity of the *ACT1* locus has remained a mystery, the *dTPH1* system in the W138 line has a number of characteristics that makes it particularly well suited for use in forward and reverse genetics approaches: First, *dTPH1* inserts preferentially in genic regions and some 20-40 novel insertions arise per individual/per generation (Koes et al., 1995; Vandenbussche et al., 2003). Consequently, relatively small mutant populations suffice to saturate the genome with genic insertions, although it should be kept in mind that *dTPH1* does not insert randomly and that the *P. hybrida* genome contains both "hot spots" and "cold spots" for *dTPH1* insertions. Second, *dTPH1* is the major source of novel mutations in W138 (van Houwelingen et al., 1998), despite the activity of other types of transposable elements (see further below), which facilitates the isolation of (mutated) genes by transposon tagging. Third, the small size of the *dTPH1* element allows for a very straightforward PCR-based genotyping assay, simply using a gene specific primer pair flanking the insertion site.

Since the isolation of the *dTPH1* element, different techniques have been developed and further improved to use *dTPH1* in forward (Van den Broeck et al., 1998) and reverse genetics (Koes et al., 1995; Vandenbussche et al., 2003; Vandenbussche et al., 2008). The exploitation of these *dTPH1* mutagenesis based forward and reverse genetics tools has particularly boosted the understanding of flower pigmentation, inflorescence architecture and flower development in *Petunia*. More recently, a new approach was developed that provides a highly accurate, efficient and cost-effective way of amplifying and sequencing *dTPH1* flanking sequences simultaneously from all individuals of a population of several thousands of plants (unpublished). This will make the creation of a semi-saturated sequence indexed mutant collection for reverse genetics screens very feasible. Such a collection would revolutionize functional genomics in *Petunia*, in the same way as the SALK collection has revolutionized *Arabidopsis* research.

Following the isolation of the *dTPH1* element, several other small related *hAT*-like elements have been isolated, often in genetic backgrounds other than W138. These have been termed *dTPH* followed by a number, which reflects their order of isolation. *dTPH2* and *dTPH4* were cloned via deliberate transposon trapping strategies in a collection of corolla pigmentation genes (van Houwelingen et al., 1998), or the nitrate reductase gene (Renckens et al., 1996). *dTPH3* was isolated as an insertion sequence in the *RT* gene, a structural gene in the anthocyanin pathway (Kroon et al., 1994). *dTPH5* was first encountered as an insertion in the *AN1* gene (Spelt et al., 2002). *dTPH7* was found as an insertion element in *PH4*, an R2R3 MYB transcription factor that activates vacuolar acidification (Quattrocchio et al., 2006). *dTPH8* was identified in a mutant allele of the *Petunia* B-class floral MADS-box gene *PhGLO1* through a gene candidate cloning approach (Vandenbussche et al., 2003). *dTPH10* was found in a *PH5* mutant allele, encoding an H+ P-ATPase on the tonoplast that determines vacuolar pH and flower colour (Verweij et al., 2008). Like *dTPH1*, all these *dTPH* elements are all too small to internally encode a transposase; therefore they are all considered to be non-autonomous. More recently, a different Petunia *hAT*-like element called *PhAT1* has been reported (Hamiaux et al., 2012) that would be large enough to encode a putative transposase (4897bp), but the element is clearly distinct from the classical *dTPH* elements (see further Figure 2).

Although most attention focused on *hAT* elements, in particular *dTPH1*, also active transposons from distinct families were identified in *P. hybrida*. Three nearly identical *dTPH6* elements of about 4 kb were found as insertions in (weakly) unstable alleles of *AN1* and *PH4* that arose in the R27/W138 background (Spelt et al., 2002; Quattrocchio et al., 2006), whereas the 9.9-kb element *PETUNIA SPM-LIKE* (*PSL*) was identified as an insertion in an unstable *hydroxylation at five 1*

allele that arose in a distinct background (Snowden and Napoli, 1998). *dTPH6* and *PSL* have TIRs of about 30 bp, which start with the sequence CACTA*,* and create 3-bp target site duplications upon insertion, which are typical features that define transposons of the *CACTA* family (Feschotte and Pritham, 2007). *MUTATOR,* which is an element that is highly active in maize, defines a large family of *MU-LIKE ELEMENTS* (MULES) that are wide spread in the plant kingdom (Feschotte and Pritham, 2007). However, in *P. hybrida* MULEs seem a rather infrequent cause of mutations compared to *hAT* and *CACTA* elements, because only few mutant alleles with MULE insertions have been found until today. The first *P. hybrida* MULE, *dTPH9*, was found as an insertion in the *HF1* gene of a red flowering commercial *P. hybrida* variety (Matsubara et al., 2005). However, because the *HF1* parent was unknown, it is unclear whether this 2.3 kb element is active in *P. hybrida*, or is an old immobile insertion that was introgressed from a parental *Petunia* species. A larger 8.5-kb *dTPH9* element was found in a mutant allele of *VEGGIE*, which is the *Petunia* homolog of *FLOWERING LOCUS T* (*FT*) from *Arabidopsis*, that arose in W138 (Roobeek, 2011). Because this *dTPH9* copy lacks some nucleotides in one of the TIRs and no reversions of the *veggie* allele were observed among ~ 4000 progeny, it appears that this *dTPH9* copy got immobilized upon, or immediately after, the insertion in *VEGGIE* (Roobeek, 2011). A third MULE, *dTPH11,* was found as an (old) insertion in the flower color gene *METHYLATION AT THREE.* This insertion occurred in *P. axillaris* and was subsequently introgressed in *P. hybrida,* suggesting that also this MULE excises very infrequently, if at all (Provenzano et al., 2014); see also the section of this paper on flower pigmentation).

Here we have mined the parental genomes (*PaxiN* and *PinfS6*) of *Petunia x hybrida* to better understand the biology and origin of the various DNA transposon families in the modern *Petunia* cultivars, with a focus on *dTPH* like elements.

Results

Presence of the small non-autonomous *dTPH* elements (*dTPH1*,2,3,4,5,7,8,10) in the *PaxiN* and *PinfS6* genomes

To create a global overview, we have for each element identified and annotated the respective copies in each genome (Table 1 and genome annotation). Overall, for all different *dTPH* elements isolated from various *Petunia* varieties (see introduction), we found copies in both genomes, showing that their origin predates the *P. axillaris/P. inflata* split. While there are important differences in copy-number between the different elements (Table 1), copy numbers per element are in the same range in the two species, except perhaps for *dTPH3* and *dTPH4* elements. Even though *dTPH1* is the major source of mutations in the *Petunia* W138 line, *dTPH7* is, surprisingly, the most abundant element with ~200 copies in each *PaxiN* and *PinfS6* genomes, and thus outnumbers the *dTPH1* elements by far (13 and 18 copies in *PaxiN* and *PinfS6* respectively).

Table 1: Distribution analysis of the small *dTPH* elements in the *PaxiN* and *PinfS6* genomes.

Name	Reference	Genbank Locus	Length[1]	TSD[2]	# unique copies *PaxiN*	# unique copies *PinfS6*	# common insertion loci
dTPH1	Gerats et al., 1990	S44589	284 bp	8bp	13	18	3
dTPH2	van Houwelingen et al., 1998	AFO22143	177 bp	8bp	8	12	1
dTPH3	Kroon et al., 1994	AF260904	443 bp	8bp	21	4	1
dTPH4	Renckens et al., 1996	U27320	787 bp	8bp	2	8	2
dTPH5	Spelt et al., 2002	AF260920	873 bp	8bp	11	16	4
dTPH7	Souer et al., 2008	AY187282	177 bp	8bp	211	188	37
dTPH8	Vandenbussche et al., 2003	AY283799	189 bp	8bp	23	26	7
dTPH10	Verweij et al., 2008	DQ515972	1176 bp	8bp	10	6	0
dTPH12	This study	Peaxi162Scf00017	702 bp	8bp	13	15	2

[1]Length of firstly described copy; [2]TSD: size of target size duplication

dTPH12@Peinf101Scf02332 AAGATGGTAG-*dTPH12*-GATGGTAGCT
dTPH12@Peaxi162Scf00900 ATGTGTCACC-*dTPH12*-GTGTCACCTC
dTPH12@Peinf101Scf00039 CCGCTGAATC-*dTPH12*-GCTGAATCCT
dTPH12@Peaxi162Scf01114 ATCCTTTGAT-*dTPH12*-CCTTTGATGA
dTPH12@Peaxi162Scf00017 ATGAAATACC-*dTPH12*-GAAATACCTC
dTPH12@Peinf101Scf01051 GATGTAGCGG-*dTPH12*-TGTAGCGGGT
dTPH12@Peaxi162Scf00638 TGTTGCATAG-*dTPH12*-TTGCATAGTA
dTPH12@Peinf101Scf00633 TCCTAACTAT-*dTPH12*-CTAACTA**C**CC
dTPH12@Peaxi162Scf00164 TGCCTGTAGG-*dTPH12*-CCTGTAGGAA
dTPH12@Peinf101Scf00035 TATTTACATC-*dTPH12*-TTTACATCAA
dTPH12@Peinf101Scf00613 *ATGTCATAGA-dTPH12-ATGAAATCAC*
dTPH12@Peinf101Scf00068 GTAGTTGGAT-*dTPH12*-AGTTGGATAG
dTPH12@Peinf101Scf00694 ATTATATCAA-*dTPH12*-TATATCAACA
dTPH12@Peinf101Scf02235 GCTTTACTTT-*dTPH12*-TTTACTTTCC
dTPH12@Peinf101Scf00102 GGGCTCTAAC-*dTPH12*-GCTCTAACCA
dTPH12@Peinf101Scf01064 TCAAGATATG-*dTPH12*-AAGATATGTG
dTPH12@Peinf101Scf00835 GCCACATGAC-*dTPH12*-CACATGA**T**CT
dTPH12@Peinf101Scf01078 *TATAAACTGT-dTPH12-AAGATATGTG*

Figure 1: Analysis of Target site duplications (TSDs) induced by *dTPH12* insertions. For each insertion, 10 bp immediately flanking the *dTPH12* element are shown on both sides. 8 bp target site duplications are underlined. Nucleotides in bold indicate mismatches. Nucleotide sequences in italics show absence of TSD.

While searching for *dTPH1* copies, we discovered a hitherto unknown *dTPH* element that we named *dTPH12* and that is closely related to *dTPH1*, but of larger size (~700 bp compared to ~284 bp). *dTPH12* elements have a comparable copy-number as the *dTPH1* element (Table 1). Comparison of the Terminal Inverted Repeat (TIR) sequences and adjacent regions show that the different *dTPH* elements can be grouped in two distinct families, with one group containing *dTPH2*, *dTPH3* and *dTPH5*, while all other *hAT* elements, including the newly identified element *dTPH12*, are more similar to *dTPH1* (Figure 2). Furthermore, the *dTPH2,3,5* group seems to possess 14-bp TIR sequences, while the others have 12-bp TIRs. Note that the TIR sequences of the large *PhAT1* element are clearly distinct from the two former groups, suggesting that this element is the first representative of yet another group of *hAT like* transposons in *Petunia* (Figure 2).

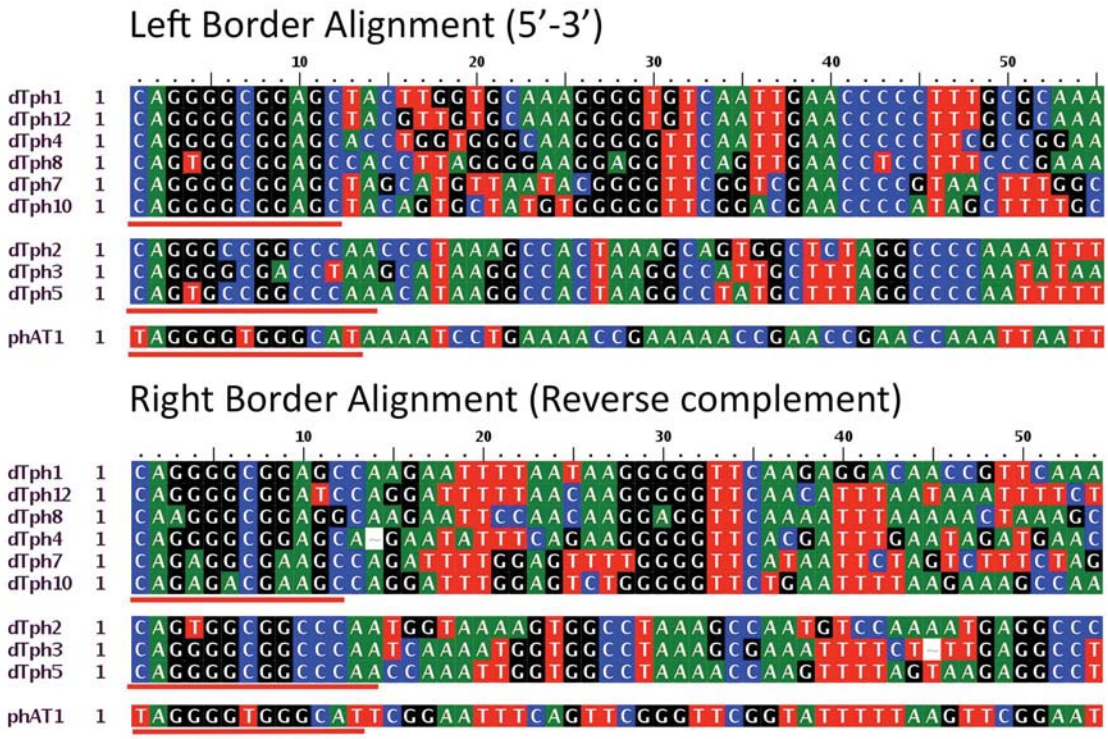


Figure 2: Comparison of Left and Right Border regions of *dTPH* elements in *Petunia*. For each border region, the first 55 bp are displayed. Terminal Inverted Repeat (TIR) sequences are underligned in red.

Comparison of *dTPH* insertion loci between *PaxiN* and *PinfS6* genomes

To be able to compare insert positions, we have retrieved on both sides of each detected *dTPH* copy an additional 500 bp of flanking sequence and have used these extended fragments for multiple alignment analysis. Common insertion loci element were then simply identified by visual inspection, based on the presence of homologous flanking sequences on both sides of the insertion element. For all elements, we found that most insertion loci (Table 1) are unique to either species, suggesting that the majority of the detected insertions in each species have transposed after the *P. axillaris/P. inflata* split. However, insertion loci that are common to both species are also identified, suggesting that these insertions might have been immobilized. In line with that, we found in at least some cases degraded or deleted TIR sequences that could explain their immobility. For example, among the three common *dTPH1* insertion loci, only one locus displays fully conserved TIR sequences on both sides of the *dTPH1* element. Alternatively, these loci might be hypermethylated and thus cannot be detected by transposases.

Transposition activity of *dTPH1* elements in wild *P. axillaris* accessions.

Because the analysis of *dTPH1* elements for each species is based on the genome sequence of one chosen individual, this provides no information on the variation of insertion sites between individuals within a species and their current transposition activity. To analyze this, we have performed a Transposon Display (Van den Broeck et al., 1998) experiment to visualize *dTPH1* elements in the progeny of *Petunia axillaris axillaris* wild accessions that we collected from local populations in Uruguay. Populations were usually found in recently disturbed soil and distances between different populations ranged from 5 to 250 km (Figure 3). The results (Figure 4) reveal around 10 insertion loci that are common to all or most individuals (indicated with black arrows). Furthermore, a number of insertion loci (green arrows, not all indicated) are shared between individuals even from

geographically distant populations but do appear only in a subset of the plants. Finally, very few insertion loci are visible that are either present in only one plant (e.g. red rectangle), only present in direct siblings (e.g. blue rectangle) or only present in plants from a same population (e.g. yellow rectangle). Overall, these results indicate that there is some diversity in *dTPH1* transposon loci within the wild *P. axillaris* population in Uruguay, but that transposition activity currently is extremely low to absent. For comparison, in W138 the total number of *dTPH1* copies per plant is >200, with 25-40 new insertions per individual/generation.

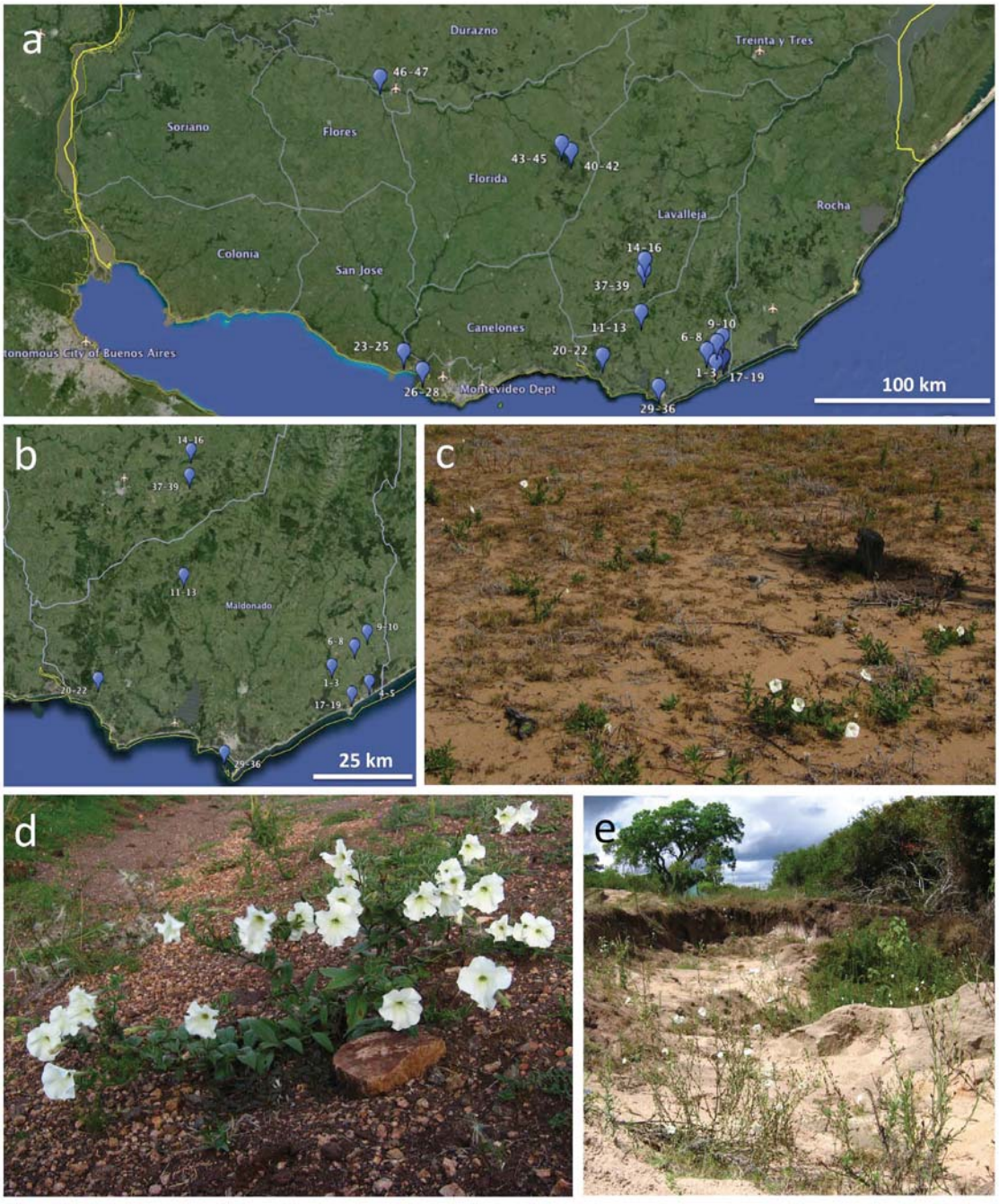


Figure 3: Geographical distribution of wild *Petunia axillaris* populations used for Transposon Display analysis and growth habitat examples. a, Uruguay map locations of *P. axillaris* populations (Google Earth). **b,** close-up of (**a**). The 15 populations used in this study are marked with the numbers of the individuals used in the Transposon display experiment. **c-e** Examples of local *P. axillaris* populations, usually encountered in recently disturbed soils.

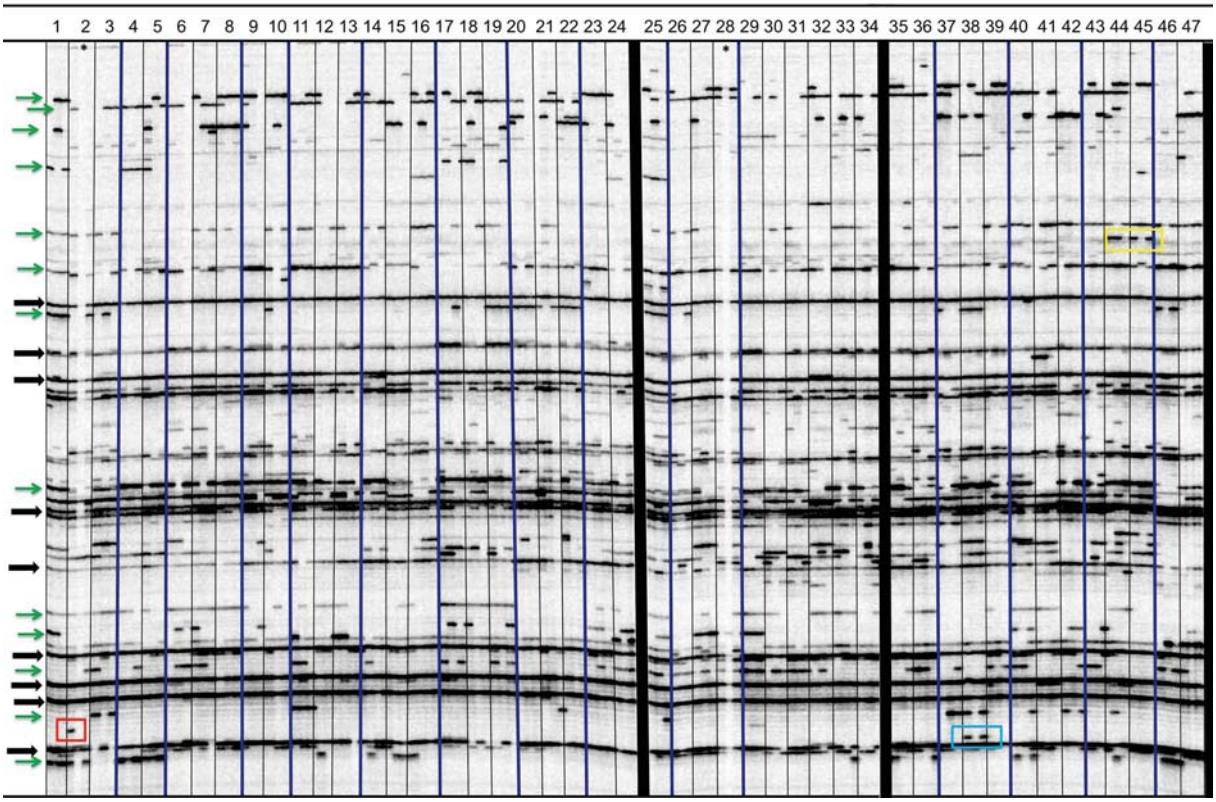


Figure 4: *dTPH1* Transposon Display analysis on progeny of *Petunia axillaris axillaris* accessions collected from distinct populations in Uruguay. All populations were encountered on the mainland with the exception of population 29-36; which was found on Gorriti Island, a small island approximately 1.5km of the coast of Maldonado (see Figure 3 for global position of the populations). Of 47 collected seed capsules, each time 3 siblings were analyzed (141 individuals in total; siblings are grouped between black vertical lines). From each local population (separated with blue vertical lines) progeny of 3 individual plants was used for analysis (with the exception of populations 9-10; 46-47; and 29-36: progeny of respectively 2;2 and 8 individuals). Black arrows: insertion loci common to all or most individuals; Green arrows: insertion loci found in individuals even from geographically distant populations, but appearing only in a subset of the plants (not all indicated). Rectangles highlight examples (not all indicated) of insertion loci that are either present in only one plant (red), only present in direct siblings (blue) or only present in plants from a same population (yellow). The asterisks indicate dropout lanes.

Comparison of W138 *dTPH1* insertion loci with those of the *P. axillaris* and *P. inflata* genomes

To further analyze the origin of the *dTPH1* elements in the W138 line, we compared in the insertion loci identified in *PaxiN* and *PinfS6* genomes with our *dTPH1* Transposon Flanking Sequence (TFS) collection (Vandenbussche et al., 2008 and unpublished data) amplified from W138 populations. Of the three common *dTPH1* insertion loci found in *PaxiN*/*PinfS6*, we found that the element displaying intact TIR sequences is still present in W138 (Table 2). The two other common insertion loci could not be detected, simply because the requirements for successful amplification are not fulfilled (sequence to divergent). In addition, from the insertion loci unique either to *PaxiN* or *PinfS6*, we could detect the presence of another 12 insertions in the W138 line (Table 2). In total, we found that four insertion loci have been inherited from *P. inflata*, and nine from *P. axillaris*, all displaying identical insertion sites in W138. Moreover, all W138-derived flanking sequences display 100% identity with the species in which the insertion was identified (with the exception of one sequence). These data strongly confirm the hybrid nature of the W138 line and identify *P. axillaris* and *P. inflata* as contributing parental species. To provide further independent proof, we took advantage of the existing sequence polymorphisms found between *PaxiN* and *PinfS6* in the regions flanking the

transposon insertion. In each case sufficient polymorphisms could be found to assign the corresponding W138 flanking sequence unambiguously to either *PaxiN* or *PinfS6*. In each case, the match corresponded to the species in which also the transposon insertion was found, independently confirming the species origin of the corresponding genomic region. Likewise, for the insertion common to *PaxiN*/*PinfS6* and W138 (Table2); the insertion in W138 clearly has been inherited from *P. inflata*, based on the polymorphisms in the transposon flanking region.

Table 2: Origin of W138 *dTPH1* insertion loci introgressed from the parental *PaxiN* and *PinfS6* genomes.

Scaffold	homology with W138	Origin of the W138 insertion based on polymorphisms between *PaxiN* and *PinfS6*
Peinf101Scf01482*	100%	*PinfS6*
Peaxi162Scf00516*	94%	
Peinf101Scf00255	100%	*PinfS6*
Peaxi162Scf00294	100%	*PaxiN*
Peaxi162Scf00258	100%	*PaxiN*
Peaxi162Scf00883	100%	*PaxiN*
Peaxi162Scf00263	3 mismatches/68bp	*PaxiN*
Peaxi162Scf00160	100%	*PaxiN*
Peinf101Scf00652	100%	*PinfS6*
Peaxi162Scf00534	100%	*PaxiN*
Peaxi162Scf00700	100%	*PaxiN*
Peinf101Scf01234	100%	*PinfS6*
Peaxi162Scf00296	100%	*PaxiN*
Peaxi162Scf00111	100%	*PaxiN*

*: Insertion site common between *PaxiN* and *PinfS6* genomes

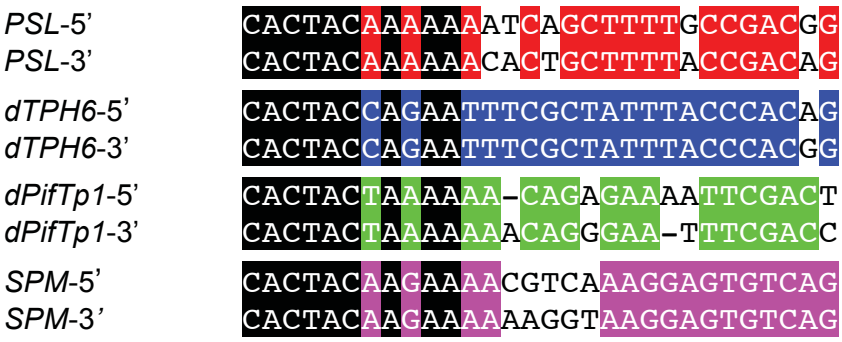


Figure 5. Comparison of Terminal Inverted Repeats of *PSL*, *dTPH6* from *P. hybrida*, *Tp1* from *P. integrifolia* and *SPM* from maize. Black shading indicates nucleotides conserved in all element. Red, blue, green and magenta shading indicates similarities between the TIRS at the 5' and 3' end of the same element

CACTA elements: *dTPh6* and *PSL*

We used sequences of the *P. hybrida dTPH6* and *PSL* elements to identify homologs in the *PaxiN* and *PinfS6* genomes. The TIRs of *PSL* and *dTPH6* share a few conserved nucleotides, which are also conserved in *SPM* from maize, but otherwise the TIRs (and the internal sequences) are different (Figure 5), indicating that (*d*)*TPH6* and *PSL* represent distinct families of CACTA elements.

A BLASTn search with the entire ~6-kb *dTPH6* sequence previously identified in *P. hybrida* yielded >1000 sequences with similarity in both genomes. However, the large majority of these regions had similarity only to sequences in the center of *dTPH6* and no similarity at all to the termini of *dTPH6*. As such (moderately repetitive) sequences are unlikely to correspond to transposition competent *dTPH6* elements they were not further analyzed. However, we found in the *PaxiN* genome three elements, ranging in size from 5705 to 10455 bp, and five in *PinfS6*, ranging from 2303-12184 bp, that may represent mobile *(d)TPH6* copies, as (i) they contain complete TIRs, (ii) display in their internal sequence significant similarity to the active *dTPH6* copy from *P. hybrida*, and (iii) were flanked by 3-bp target site duplications (Figure 6). Interestingly, all eight insertions were unique for either *PaxiN* or *PinfS6*, indicating that these elements still transposed after the *P. axillaris* /*P. inflata* split and may be transposition competent until today. The *dTPH6* elements in the 10-12 kb range share similarity with transposase encoding regions from other CACTA elements, such as *PSL* and *SPM*, while the copies, including the three *dTPH6* copies identified in *P. hybrida*, do not. This indicates that one or more of the larger elements may be autonomous, whereas the smaller elements are non-autonomous elements.

dTPH6* insertions in *P. axillaris N

```
Paxi162Scf01326   10455 bp   GAAGACTGGGAA CACTACCAGAATTTCGCTATTTACCCAC-dTPH6-GTGGGTAAATAGCGAAATTCTGGTAGTG GAAGATGTGTCAAAG
PinfS6                       GAAGACTGGGAA                                                                 GATGTGTCAAAG

Paxi162Scf00177    5705 bp   CCTGCTCACTCC CACTACCAGAATTTCGCTATTTACCCAC-dTPH6-GTGGGTAAATAGCGAAATTCTGGTAGTG TCCAATCTTATTTCA
PinfS6                       CCTGCTCACTCC                                                                 AATCTTATTTCA

Paxi162Scf01016   10148 bp   CATCATCAACAA CACTACCAGAATTTCGCTATTTACCCAC-dTPH6-GTGGGTAAATAGCGAAATTCTGGTAGTG CAAGTGGAAGCTTAA
PinfS6                       CATCATCAACAA                                                                 GTGGAAGCTTAA

PaxiScf00745a     12786 bp   TCCGAGTTGCCC CATTACCAGAATTTTGCTATTTACCCAT-dTPH6-GTGGGTAAATGGCGAAATTCTGGTAGTG CCCGAAACCAAAGAC
PaxiScf00745b     19559 bp   TCCGAGTTGCCC CATTACCAGAATTTTGCTATTTACCCAC-dTPH6-GTGGGTAAATAGCGAAATTCTAGTAGTG CCCGAAACCAAAGAC
PinfS6                       TCCGAGTTGCCC                                                                 GAAACCGAAGAC
```

dTPH6* insertions in *P. inflata S6

```
Pinf101Scf00236   10688 bp   CTAGAAAAAAAA CACTACCAGAATTTCGCTATTTACCCAC-dTPH6-GTGGGTAAATAGCGAAATTCTGGTAGTG AAAATAATATCGAA
PeaxiN                       CTAGAAAAAAAA                                                                 --AATATCGAA

Pinf101Scf00177   11173 bp   AGAGGGAACTCC CACTACCAGAATTTCGCTATTTACCCAC-dTPH6-GTGGGTAAATAGCGAAATTCTGGTAGTG CTTCTCTCTCCATC
PeaxiN                       no homologous (repetitive) sequence found

Pinf101Scf01047   12184 bp   GCTTGGCTCTAC CACTACCAGAATTTCGCTATTTACCCAC-dTPH6-GTGGGTAAATAGCGAAATTCTGGTAGTG TACACACTAAATCA
PeaxiN                       GCTTGGCTCTAC                                                                 ACACTAAATCA

Pinf101Scf00592    2303 bp   TCGATCCCAAAG CACTACCAGAATTTCGCTATTTACCCAC-dTPH6-GTGGGTAAATAGCGAAATTCTGGTAGTG AAGGTGGCACAGGT
PeaxiN                       AAG sequence

Pinf101Scf00660    9745 bp   ACTGCTGAAGGA CACTACCAGAATTTCGCTATTTACCCAC-dTPH6-GTGGGTAAATAGCGAAATTCTGGTAGTG GGACTACCTTGACA
PeaxiN                       ACTGCTGAAGGA                                                                 CTACCTTGACA
```

Figure 6 (*d*)*TPH6* elements identified in the *PaxiN* and *PinfS6* genomes. For each (*d*)*TPH6* copy the scaffold in which it resides, and size of the element is shown on the left. (*d*)*Tph6* sequences (TIRs only) are depicted in blue and flanking sequences in black. Underlined sequences denote the target site duplication. For each insertion the sequence of the homologous locus in the other species is shown below. Alterations in the TIRs and mismatches between the flanking sequence and the homologous region in the other species are highlighted in red. For the insertion site of one element in *PinfS6* the homologous sequence in *PaxiN* could not be identified.

DNA gel blot data suggested that *P. axillaris*, *P. parodii*, (a white flowering species related to *P. axillaris*) and *P. integrifolia* (a violet flowering species related to *P. inflata*) harbor about 3 fragments that strongly hybridize to the right end of *PSL* and 3-8 additional fragments that hybridize weakly (Snowden and Napoli, 1998). BLASTn searches with the *P. hybrida PSL* sequence identified about a dozen sites in the *PaxiN* and *PinfS6* genomes with similarity to central sequences of *PSL*. Only two of these regions in *PaxiN* and three in *PinfS6* are bracketed by TIRs with similarity to those of *PSL* and a 3-bp target site duplication and may represent (putative) competent transposons. In addition *PinfS6* contains a fourth *PSL* element of 17.3 kb, the largest of the *PSL* elements that we found, in which the conserved terminal CACTA sequence is changed into GAGTA and which is not flanked by an obvious target site duplication. Hence, this *PSL* copy is probably an element that got immobilized. As the sequences that flank this element in *P. inflata* are not present in the *PaxiN* genome, we could not assess whether the original TSD was lost after the insertion by a secondary deletion event (Figure 7).

P. axillaris N

Paxi162Scf00034	AGCTCCACTTCCAAGCCTA	CACTA-PSL-TAGTG	CTAGAACCTCCTCGAACTAT	9931 bp	TSD=3
PinfS6	AGCTCCACCTCCAAGCCTA		TCACCTCCTCGAGCTAT		
Peaxi162Scf00695	TCAACGATTGTCCCAGTTC	CACTA-PSL-TAGTG	GAGGTAGGGCCAAAAGAGAG	15552 bp	TSD=????
PinfS6	Homologous sequence not found				

P. inflata S6

Pinf101Scf00546	TTTTTATATGTCTCTTAAT	CACTA-PSL-TAGTG	AATCTTGTATTTGTGTTAA	4301 bp	TSD=3
PaxiN	TTTGTATATGTCTCTTAAT		CTTGTATTTGCATTAA		
Pinf101Scf01289	CATTCAAGTTGATAATATC	CACTA-PSl-TAGTG	ATCCTACATGTCGTAGTGG	4608 bp	TSD=3
PaxiN	CATTCAAGTTGATAATACT		--ACATGTCGTAGTGG		
Pinf101Scf01887	AAAATATACTACATGTTAT	CACTA-PSl-TAGTG	ATATGCTAATGAAAAAAAT	12352 bp	TSD=3
PaxiN	AAAATATACAACATGTTAT		GCTAATGTAAAAAAT		
Pinf101Scf00044	AAAGAGGTGGTGGGTGGAA	GAGTA-PSL-TAGTG	ATATACATAGATGCCAAAC	17295 bp	TSD=??
PaxiN	Homologous sequence not found				

Figure 7. *PSL* elements in the *PaxiN* and *PinfS6* genomes**.** For each *PSL* copy the scaffold in which it resides is shown on the left and the size of the element and the target site duplication (TSD) on the right. *PSL* sequences (only the terminal 5 nt are shown) are depicted in blue and flanking sequences in black. Underlined sequences denote the target site duplication. For each insertion the sequence of the homologous locus in the other species is shown below. Alterations in the TIRs and mismatches between the flanking sequence and the homologous region in the other species are shown in red lettering.

MULES: *dTPH9* and *dTPH11*

BlastX searches showed that the *PaxiN* and *PinfS6* genomes contain >5000 regions that potentially encode an amino acid motif that is conserved among transposases (usually defined a MurA) of a variety of MULEs from different species. This suggests that the *PaxiN* and *PinfS* genome may contain an enormous number of MULEs. We analyzed two MULE families, *dTPH9* and *dTPH11*, in more detail, as previous findings suggested that these elements were active, at least until recently, and created mutant alleles in *P. hybrida*.

In a screen of *P. hybrida* W138 progeny we identified a new locus, *VEGGIE*, that controls flowering time and inflorescence architecture through a new mutant allele containing an insertion of the MULE *dTPH9*. To identify potential transposition competent *dTPH9* elements we performed BLASTn searches of the *PaxiN* and *PinfS6* genomes using the *dTPH9* insertion in the *P. hybrida veggie*^{C3382} allele as a query, and manual/visual inspection to select *PaxiN* and *PinfS6* sequences that have similarity to *dTPH9* and are, like *dTPH9*, bracketed by ~120-bp TIRs. This revealed that the *PaxiN* genome contains at least 24 *dTPH9* elements, ranging in size from 901 to 14707 bp (Figure 8). Fourteen of these elements represented insertions unique for *PaxiN*, as the homologous *PinfS6* locus contain no insertion, and possess complete TIRs and are flanked by a 9 or 10 bp TSD, similar to the TSDs created by other MULEs, indicating that these element may still be mobile. Five of the *PaxiN dTPH9* insertions apparently predated the *P. axillaris* / *P. inflata* split, as the *PinfS6* genome contained the same *dTPH9* insertion. One of these elements lost 8 bp in one TIR, which may account for the apparent immobility of this element. However, the four other common *dTPH9* insertions that are in *PinfS6 and PaxiN* have no obvious defects in their TIRs and are flanked by 9 or 10-bp TSDs, suggesting that their apparent immobility is due to other defects that are unknown. The *PinfS6* genome contains, besides the five insertions shared with *PaxiN*, another 22 *dTPh9* insertions that are unique for *PinfS6*. Eighteen of these concern *dTPH9* elements, ranging in size from 1269 to 20272 bp, with apparently complete TIRs and are flanked by a 9 or 10-bp TSD, suggesting that these may be mobile *dTPH9* copies.

Whereas most *dTPH9* insertions in *PaxiN* and *PinfS6* are flanked by 9 or 10-bp insertions, we found in about 20% (10 out of 47) of the cases no obvious TSD. In four of these cases the TSD seems to have been lost by a secondary deletion that also removed a few nucleotides of the TIR. However the other six cases, in which no nucleotide seems deleted from the *dTPH9* element and the flanking sequence, the absence of a TSD cannot be easily explained by secondary deletions, unless one assumes that small deletion removed each time only the TSD, which seems improbable if happening by chance only. An alternative explanation may be that in substantial fraction of the cases (about 20%) dTPH9 inserts by an alternative mechanism that does not cause the duplication of target sequences.

The MULE *dTPH11* was identified as an inactivating insertion in *METHYLATION AT THREE* (*MT*), which encodes an anthocyanin methyltransferase, that occurred in the *P. axillaris* lineage after the split from the *P. inflata* group (Provenzano et al., 2014). The *PaxiN* and *PinfS6* genomes contain at least 23 and 35 *dTPH1*1 copies respectively (Figure 9). Only 4 of those insertions are common in *PaxiN* and *PinfS6* indicating that these predate the *P. axillaris* / *P. inflata* split. Most of these elements are terminated by intact TIRs and bracketed by 9-bp target site duplications, suggesting that they are still transposition competent. Interestingly all *dTPH11* insertions caused target site duplications of 9 bp, but unlike *dTPH9*, no duplication of 10 bp or 0 bp were observed, except for one insertion in *PaxiN* which seems to have undergone some secondary rearrangements that make it impossible to discern whether this insertion duplicated 9, 10 or 11 bp.

a, Unique insertions in *PinfS6* (18)

```
Pinf101Scf00001   AACTAATTATATAAC GAAG-dTPH9-CTTC TTATATAACTCTCGA   1269 bp  TSD=9
Paxi              AACTAATTATATAAC                 TCTCCA

Pinf101Scf00889   TCACTTTCTTTCTTT GAAG-dTPH9-CTTC TTCCTTCTTTTGTAT   2102 bp  TSD=10
Paxi              TCCCTTTCTTTCTTT                 TGTAT

Pinf101Scf00125   TATGTCATGTGTTTG GAAG-dTPH9-CTTC CATGTGTTTGTGGTG   2325 bp  TSD=10
Paxi              No blast hits

Paxi00341         GAAAGCGAAAACACG GAAG-dTPH9-CTTC CGAAAACACGCATGA   10422 bp  TSD=10
Pinf              homogous sequence not found

Pinf101Scf00906   AACAAGCTTTTCAGC GAAG-dTPH9-CTTC CTTTTCAGCCCCGCC   8302 bp  TSD=9
Paxi              AACAAGCTTTTCAGC                 CTCGCC

Pinf101Scf00118   TGGTGCAAGTCAATC GAAG-dTPH9-CTTC CAAGTCAATCCCACG   20272 bp  TSD=10
Paxi              -GGTGCAAGTCAATC                 CCATG

Pinf101Scf03436   TAAGCTAATTTTTTC GAAG-dTPH9-CTTC TAATTTTCTCAATAA   4837 bp  TSD=10
Paxi              TAAGCTAATTTTCTC                 AATAA

Pinf101Scf03119   TTGTTTGAGAAAATTA GAAG-dTPH9-CTTC GAGAAAATTAGCTTA   7088 bp  TSD=10
Paxi              TTATTGAGAAAATTA                 GCTTA

Pinf101Scf00013   GACCCTCTCATCTAG GAAG-dTPH9-CTTC TTCATCTATCCTGGA   2115 bp  TSD=9
Paxi              GACCCTCTCAT-TAT                 CCTGGA

Pinf101Scf01189   CATCTTCTCGTTCAC AAAG-dTPH9-CTTC TCTCGTTCACCTGTC   2038 bp  TSD=10
Paxi              No clear hit

Pinf101Scf01179   TGCCAAAAGGCCTTG GAAG-dTPH9-CTTC AAAGGCCTTGCCACA   2097 bp  TSD=10
Paxi              GCCTAAAGCCCTTG                  TTACA

Pinf101Scf00019   TAGTATAACTACAAT GAAG-dTPH9-CTTC TAACTACAACTCAAC   2079 bp  TSD=10
Paxi              TAGTATAACTACAAT                 TCAAC

Pinf101Scf00191   AGTTGAGTTATGGCG GAAG-dTPH9-CTTC AGTTATGGCGTTCTT   6315 bp  TSD=10
Paxi              AGTTGAGTTATGGCG                 T--TT

                        Δ3bp       Δ3bp
Pinf101Scf00736   CTAATTAACCAAA-T ---G-dTPH9-C--- CAAAATAATACGTTT   10414 bp  TSD=6
Paxi              CTAATTAACCAAAAT                 AATACATTT

Pinf101Scf00060   GTAAAATGATAAATG GAAG-dTPH9-TATC TGATAAATGTGGTGT   7096 bp  TSD=9
Paxi              GTAAAATGATAAATA                 TGGTGT

Pinf101Scf01199   ACAATGGATACAACC GAAG-dTPH9-CTTC GATACAACCCGAAAA   6290 bp  TSD=9
Paxi              No clear hit

Pinf101Scf00301   ATTTAAAATCACATG GAAG-dTPH9-CTTC AATCACATGCAAGCA   14289 bp  TSD=9
Paxi              ATCCAAAGTCACATG                 CAAACA

Pinf101Scf00258   GACTTACAGTTTTTG GAAG-dTPH9-CTTC CCGTTTTTGTTTCCT   2881 bp  TSD=9
Paxi              GACTTACATTTTTTG                 TTTCCT
```

b, Unique insertions in *PaxiN* (14)

```
Paxi162Scf00428   ACTAGTTTTTCACAC GAAG-dTPH9-CTTC TTTTTCACACTGTGA   901 bp  TSD=10
Pinf101Scf        ACTAGTTTTTCACAC                 TGTGA

Paxi162Scf00057   CTAACTCATATTTGA GAAG-dTPH9-CTTC TCATATTTAAGGCTA   11937 bp  TSD=10
Pinf              CTAACTCATTTTTGA                 GGCTA

Paxi162Scf00456   CCATGTATTAGTAAT GAAG-dTPH9-CTTC TATTAGTAATACATG   02084 bp  TSD=10
Pinf              CTATGTATTACTAAT                 ACATG

Paxi162Scf01311   TGGCAAAATTATATA GAAG-dTPH9-CTTC ATTTATATAAGGAGT   5721 bp  TSD=9
Pinf              TGGCAAAATTATATA                 AGGAGT

Paxi162Scf00677   CCTGTCTATTATGTG GAAG-dTPH9-CTTC TATTATGTGTGCTCA   4874 bp  TSD=9
Pinf              CCTGTCTATTATGTG                 TGCTCA

Paxi162Scf00376   AGTTGTGATCCAAAG GAAG-dTPH9-CTTC TGATCCAAAGTCCCT   5152 bp  TSD=10
Pinf              AGTTGTGAGCCAAAG                 TCCCT

Paxi162Scf00003   TTTCCAATCTAAATT GAAG-dTPH9-CTTC ATCTAAATTAGTTGG   9068 bp  TSD=9
Pinf              TTTCCAATCTAAATT                 AGTTGG

Paxi162Scf00961   GGATGCAGTTTAGAT GAAG-dTPH9-CTTC CAGTTTAGATTCACG   7769 bp  TSD=9
Pinf              GGATGCAGTTTAAAT                 TTCACG

Paxi162Scf00273   TCAAAATTAGATTAT GAAG-dTPH9-CTTC TTAGATTCTCAATAA   4993 bp  TSD=9
Pinf              TCAAAATTAGATTCT                 CAACAA

                                                  +1bp
Paxi162Scf00755   GTACATTGTTAAAAT GAAG-dTPH9-CTTC TCTTTTAAAATGTTA   04905 bp  TSD=10
Pinf              GTACATCTTTAAAAT                 GTTA

Paxi162Scf00564   GTTATGTCTTCTTTT GAAG-dTPH9-CTTC TCTTCTTTTGATTCC   9857 bp  TSD=9
Pinf              GTTATGTCTTCTTTT                 GATTCC

Paxi162Scf00052   CCATGTATTAGTAAT GAAG-dTPH9-CTTC TATTAGTAATACATG   2843 bp  TSD=10
Pinf              CTATGTATTACTAAT                 ACATG

Paxi162Scf00038   CAACATAAAGGATAA GAAG-dTPH9-CTTC AAAAGGAAAAGGAGG   2658 bp  TSD=10
Pinf              CAACATAAAGGAAAA                 GGAGG

Paxi162Scf00012   TTAGCTTCAAGGTTT GAAG-dTPH9-CTTC TCAAGGTTTCCACTA   14707 bp  TSD=9
Pinf              TTAGCTTCAAGGTTT                 CTAGTATTAATTT
```

d, Unique insertions in *PaxiN* and *PinfS6* without clear TSD(5+5)

```
                        Δ9bp      Δ10bp
Paxi162Scf00145   ATAAAG--------- -----dTPH9-CTTC AACTTTAAACCTTGC   902 bp  no TSD
Pinf              ATAAAGATTCCGGAA                 AACTTTAAACCTTGC

Paxi162Scf00132   TTTATTGATCCGTTC --AG-dTPH9-CT-- CCATTTTTCACTTAT   3885 bp  no TSD
Pinf              TTTATTGATCCGTTT                 -CATTTTTCACTTAT

Paxi162Scf00474   TAGAGAACCGTGTAA -----dTPH9----- TATTATTTTGGTCAA   3269 bp  no TSD
Pinf              TAGAGAACCGTGTAA                 -CTAATGATAGTCTA

Paxi162Scf00241   TATGAATTCTAACCC GAAG-dTPH9-CTTC CTAGCGAGTGTGAAT   12464 bp  no TSD
Pinf              TATGAATTCTAACCC                  TAACGAGTGTGAAT

Paxi162Scf00618   ATACAGGTTGAATCA GAAG-dTPH9-CTTC TTAAATGGAGCATAT   9680 bp  no TSD
Pinf              GTACAGGTTGAACTA                 TGGAGCATAT

Pinf101Scf02187   AATTGGTAGTAATAT GAAG-dTPH9-CTTC CNTAATGATACATTT   7074 bp  NO TSD
Pax               AATTGGTAGTAAT   <   289 bp   >   TAATGATACATTT

Pinf101Scf00755   GGTCAAATCGCTTCC -----dTPH9-CTTC TCTTCTCACCCATAG   10104 bp  no TSD
Paxi              GGTCAAATCGCCTCC                 TCTTCTCACCCATAG

Pinf101Scf01292   CCATCAGGGTTAGCA GAAG-dTPH9-CTTC AATAGCTTAGTACGG   3585 bp  no TSD
Paxi              Absent!

                        Δ24 bp
Pinf101Scf01996   ACTTATCTACGCTAC -----dTPH9-CTTC AATTTGTTGTTCATT   4264 bp  no TSD
Paxi              Absent

Pinf101Scf02339   GATGGATAAGGCAAT GAAG-dTPH9-CTTC TGAAATACTGTAAGTA   9963 bp  no TSD
Paxi              Absent
```

c, Common insertions in *PaxiN* and *PinfS6* (5)

```
                        Δ8bp
Paxi162Scf00337   TGATGCTTTTGTTAA -----dTPH9-CTTC TAATGTAGATTGCTA   13547 bp  No TSD
Pinf101Scf00734   TGATGCTTTTGTTAA -----dTPH9-CTTC TAATGTAGATTGCTA   27124 bp
                        Δ8bp

Paxi162Scf00993   CAGACTCAAAGAAGA GAAG-dTPH9-CTTC CAAAGAAAGAAGTAA   3741 bp  TSD=10
Pinf101Scf00713   CAGACTCAAAGAAGA GAAG-dTPH9-CTTC CAAAGAAAGAAGTAA   3579 bp

Paxi162Scf00450   TGGAATATATATAAG GAAG-dTPH9-CTTC TATATATTATAAGGG   25422 bp  TSD=10
Pinf101Scf00713   TGGAATATATATAAG GAAG-dTPH9-CTTC TATATATTATAAGGG   30576 bp
Pinf101Scf04980                   dTPH9-CTTC TATATATTATAAGGG   ~2200 bp

Paxi162Scf00535   CAACAATGTTATTGC GAAG-dTPH9-CTTC AGTTATTGCAATAAC   3795 bp  TSD=9
Pinf101Scf01659   CAACAATGTTATTGC GAAG-dTPH9-CTTC AGTTATTGCAATAAC   3698 bp  TSD=9

Paxi162Scf00345   TGTAAAAGAACTCCT GAAG-dTPH9-CTTC AGAACTCCTTGACTC   3464 bp  TSD=9
Pinf101Scf12821   TGTAAAAGAACTCCT GAAG-dTPH9-CTTC AGAACTCCTTGACTC   3262 bp  TSD=9
```

Figure 8. *dTPH9* elements in *PaxiN* and *PinfS6*. a, Unique insertions in *PaxiN*. **b,** Unique insertions in *PinfS6* **c**, Common insertions present in both *PaxiN* and *PinfS6*. **d**, Unique insertions in *PaxiN* and *PinfS6* not flanked by a target site duplication. For each *dTPH9* copy the scaffold in which it resides, is shown on the left and the size of the element and the target site duplication (TSD) on the right. *dTPH9* sequences (only the terminal 4 nt are shown) are depicted in blue and flanking sequences in black. Underlined sequences denote the target site duplication. For each insertion the sequence of the homologous locus in the other species is shown below. Alterations in the TIRs and mismatches between the flanking sequence and the homologous region in the other species are highlighted in red.

Figure 9. *dTPH11* elements in *PaxiN* and *PinfS6*. a, Unique insertions in *PaxiN*. **b,** Unique insertions in *PinfS6.* **c,** common insertions in *PaxiN* and *PinfS6*. For each *dTPH11* copy the scaffold in which it resides is shown on the left and the size of the element and the target site duplication (TSD) on the right. *dTPH11* sequences (only the terminal 10 nt are shown) are depicted in blue and flanking sequences in black. Underlined sequences denote the target site duplication. For each insertion the sequence of the homologous locus in the other species is shown below. Alterations in the TIRs and mismatches between the flanking sequence and the homologous region in the other species are highlighted in red.

DISCUSSION

Virtually all mutant alleles that have been isolated in Petunia until today arose from "spontaneous" mutations that were isolated over decades in breeding programs, or more recently in large scale screens of progeny from the *P. hybrida* line W138, whereas chemical of radiation mutagenesis was employed only rarely. Molecular analysis indicated that the large majority of the spontaneous mutations were due to insertions from transposable elements or large rearrangements (deletions) with breakpoints in an transposon, indicating that transposons were a major source of the genetic variation that drove the divergence of the ancestral *P. axillaris* and *P. inflata* species and the appearance of novel phenotypes, including for example wide variety off lower colors, during 2 centuries of breeding.

Since the cloning of the *dTPH1* element from *P. hybrida*, now 25 years ago, *dTPH1* transposon mutagenesis has become an extremely powerful tool in forward and reverse genetics approaches in *Petunia*, but knowledge on its origin and biology has remained very limited, mainly because no *Petunia* genome sequence was available. The currently sequenced genomes of the *P. axillaris* N and *P. inflata* S6 lines, both considered as parental species of the modern *Petunia x hybrida* varieties, have opened the way to a more detailed analysis of the origin and biology of *dTPH1* and other transposable elements in *Petunia*. Confirming earlier estimates based on Southern blot hybridizations (Huits et al. 1995), both species appear to contain only a small number of *dTPH1* copies (Table 1) compared to the W138 line, in line with the idea that *dTPH1* copy number has vastly increased only recently and specifically in the line leading to W138. Furthermore, it was shown by genetic analysis (Huits et al. 1995) that both parental species do not exhibit *ACT1* activity, the genetic locus required for *dTPH1* transposition. In line with that, our *dTPH1* transposon display analysis on progeny of *Petunia axillaris* accessions collected in the wild all over Uruguay revealed few new transposition events.

When we compared the identified *PaxiN* and *PinfS6* loci harbouring a *dTPH1* insertion with *dTPH1* loci in W138 populations, we found back 9/16 of the *PaxiN* loci and 5/21 of the *PinfS6* loci in W138, providing independent evidence that both species indeed contributed to the hybrid genome of the W138 cultivar. Note that these numbers are minimal estimates, since some *dTPH1* copies in the parental genomes might be too divergent to be amplified by our *dTPH1* flanking sequence amplification protocol (Vandenbussche et al., 2008). While *dTPH1* copy numbers are low in the *PaxiN* and *PinfS6* genomes, we found that the *dTph7* element has vastly expanded reaching around ~200 copies in both species, while all other *dTPH*-like elements are present at much lower copy numbers. Together with the observation that in W138 *dTPH1* seems to be the major cause of mutations, this further supports the idea that at least some of the different *dTPH* elements in *Petunia* may require different transacting factors for their mobility (Kroon et al., 1994). Finally, several other small related *h*AT-like elements, and active transposons from the CACTA and MULE families have been isolated over the years, often, but not always, in genetic backgrounds other than W138. We found copies of all of these elements in both parental genomes, indicating that these transposable elements already existed before the *P. axillaris/P. inflata* split and have been transposing ever since, albeit at very low frequencies compared to the activity of *dTPH1* in the W138 line

Materials & Methods

Identification of *dTPH*-like copies and annotation

Copies of *dTPH*-like elements were identified in the genomes of *P. axillaris* and *P. inflata* by repetitive BLASTN search, and together with the element, each time 500 bp of flanking sequence on both sides was retrieved to allow insertion locus comparison between the two species. Sequences were aligned using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Insertion loci common between the two species were identified by clustering, and confirmed by visual inspection (Bioedit, http://www.mbio.ncsu.edu/bioedit/bioedit.html) of the homology of left and right border flanking sequences between *P. inflata* and *P. axillaris*. Each individual element copy was annotated in the genome sequence by defining the start of left and end of right TIRS. In case one (or two of the TIRs) were missing, the copy was marked as 'fragment'. *dTPH* insertion loci that are shared by the two species are indicated by 'Common locus *PaxiN* /*PinfS6* '. Per element, the TSD (Target Site Duplication) size was verified by comparing 5' and 3' border sequences and comparison with the corresponding WT locus in either *P. axillaris* or *P. inflata* as shown for the newly identified *dTPH1*2 element in Figure 2.

Comparison with W138 *dTPH1* insertion loci

For the detection of W138 *dTPH1* insertion loci inherited from the *P. axillaris* and *P. inflata* parental genomes, we BLAST-searched the *PaxiN*/*PinfS6 dTPH1* elements including the 500-bp flanking sequence against our TFS collection (Vandenbussche et al., 2008 and unpublished data). Because the W138 flanking sequences allow determining the exact insertion site (the first nucleotide of the flanking sequence directly flanks the *dTPH1* transposon), inherited insertions can be identified by alignments starting immediately at the beginning of the genomic sequences flanking the *dTPH1* insertion loci found in the two genomes. We used the *PaxiN*/*PinfS6* polymorphisms found in the genomic flanking regions to independently identify the species origin of the genomic region in W138.

Transposon display on *Petunia axillaris* accessions.

During a two-week trip in Uruguay, we crossed the country in search of local *Petunia axillaris* populations, and collected ripe seed capsules from individual plants, usually from three different plants per local population (Figure 3). Progenies of these lines were grown in the lab, and per line, three siblings were subjected to a transposon display analysis. *dTPH1* insertion loci derived fragments were visualized as described earlier (Van den Broeck et al., 1998), with the exception that we used adapter and TIR based primers without selective nucleotides during the 'hot PCR', to anticipate on the much lower *dTPH1* copy numbers found in wild accessions compared to W138. The image shown in Figure 4 is composed of three different gel(part)s (separated by solid black lines). For the alignment of the images, the two parts on the right have been slightly adjusted in height to compensate for small runtime differences with the gel on the left.

REFERENCES

Bianchi, F., Cornelissen, P.T., Gerats, A.G., and Hogervorst, J.M. (1978). Regulation of gene action in Petunia hybrida: Unstable alleles of a gene for flower colour. Theor Appl Genet **53,** 157-167.

Doodeman, M., Boersma, E.A., Koomen, W., and Bianchi, F. (1984). Genetic analysis of instability in Petunia hybrida : 1. A highly unstable mutation induced by a transposable element inserted at the An1 locus for flower colour. Theor Appl Genet **67,** 345-355.

Faraco, M., Spelt, C., Bliek, M., Verweij, W., Hoshino, A., Espen, L., Prinsi, B., Jaarsma, R., Tarhan, E., de Boer, A.H., Di Sansebastiano, G.P., Koes, R., and Quattrocchio, F.M. (2014). Hyperacidification of vacuoles by the combined action of two different P-ATPases in the tonoplast determines flower color. Cell Rep **6,** 32-43.

Feschotte, C., and Pritham, E.J. (2007). DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet **41,** 331-368.

Gerats, A.G., Huits, H., Vrijlandt, E., Marana, C., Souer, E., and Beld, M. (1990). Molecular characterization of a nonautonomous transposable element (*dTPH1*) of petunia. Plant Cell **2,** 1121-1128.

Hamiaux, C., Drummond, R.S., Janssen, B.J., Ledger, S.E., Cooney, J.M., Newcomb, R.D., and Snowden, K.C. (2012). DAD2 is an alpha/beta hydrolase likely to be involved in the perception of the plant branching hormone, strigolactone. Curr Biol **22,** 2032-2036.

Huang, J., Zhang, K., Shen, Y., Huang, Z., Li, M., Tang, D., Gu, M., and Cheng, Z. (2009). Identification of a high frequency transposon induced by tissue culture, nDaiZ, a member of the hAT family in rice. Genomics **93,** 274-281.

Huits, H.S., Wijsman, H.J., Koes, R.E., and Gerats, A.G. (1995). Genetic characterisation of Act1, the activator of a non-autonomous transposable element from Petunia hybrida. Theor Appl Genet **91,** 110-117.

Koes, R., Souer, E., van Houwelingen, A., Mur, L., Spelt, C., Quattrocchio, F., Wing, J., Oppedijk, B., Ahmed, S., Maes, T., and et al. (1995). Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants. Proc Natl Acad Sci U S A **92,** 8149-8153.

Kroon, J., Souer, E., de Graaff, A., Xue, Y., Mol, J., and Koes, R. (1994). Cloning and structural analysis of the anthocyanin pigmentation locus Rt of Petunia hybrida: characterization of insertion sequences in two mutant alleles. Plant J **5,** 69-80.

Lisch, D. (2013). How important are transposons for plant evolution? Nat Rev Genet **14,** 49-61.

Matsubara, K., Kodama, H., Kokubun, H., Watanabe, H., and Ando, T. (2005). Two novel transposable elements in a cytochrome P450 gene govern anthocyanin biosynthesis of commercial petunias. Gene **358,** 121-126.

Provenzano, S., Spelt, C., Hosokawa, S., Nakamura, N., Brugliera, F., Demelis, L., Geerke, D.P., Schubert, A., Tanaka, Y., Quattrocchio, F., and Koes, R. (2014). Genetic Control and Evolution of Anthocyanin Methylation. Plant Physiol **165,** 962-977.

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.

Renckens, S., De Greve, H., Beltran-Herrera, J., Toong, L.T., Deboeck, F., De Rycke, R., Van Montagu, M., and Hernalsteens, J.P. (1996). Insertion mutagenesis and study of transposable elements using a new unstable virescent seedling allele for isolation of haploid petunia lines. Plant J **10,** 533-544.

Roobeek, I. (2011). Molecular analysis of the inflorescence architecture in Petunia hybrida. PhD. thesis (Vrije Universiteit Amsterdam, The Netherlands).

Snowden, K.C., and Napoli, C.A. (1998). Psl: a novel Spm-like transposable element from Petunia hybrida. Plant J **14,** 43-54.

Spelt, C., Quattrocchio, F., Mol, J.N., and Koes, R. (2000). anthocyanin1 of petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. Plant Cell **12,** 1619-1632.

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.

Van den Broeck, D., Maes, T., Sauer, M., Zethof, J., De Keukeleire, P., D'Hauw, M., Van Montagu, M., and Gerats, T. (1998). Transposon Display identifies individual transposable elements in high copy number lines. Plant J **13,** 121-129.

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.

Vandenbussche, M., Zethof, J., Souer, E., Koes, R., Tornielli, G.B., Pezzotti, M., Ferrario, S., Angenent, G.C., and Gerats, T. (2003). Toward the analysis of the petunia MADS box gene family by reverse and forward transposon insertion mutagenesis approaches: B, C, and D floral organ identity functions require SEPALLATA-like MADS box genes in petunia. Plant Cell **15,** 2680-2693.

Vandenbussche, M., Janssen, A., Zethof, J., van Orsouw, N., Peters, J., van Eijk, M.J., Rijpkema, A.S., Schneiders, H., Santhanam, P., de Been, M., van Tunen, A., and Gerats, T. (2008). Generation of a 3D indexed Petunia insertion database for reverse genetics. Plant J **54,** 1105-1114.

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.

Supplementary Note 4

Analysis of Tandem Duplications and Gene Families in *Petunia* Species as Compared to Other Solanaceae and *Arabidopsis*

Mitrick A. Johns*, Jennifer Hintzsche

Dept. of Biological Sciences, Northern Illinois University, DeKalb IL 60115, USA
*Corresponding author: e-mail: rjohns@niu.edu

Short title: *Petunia* gene families

Abstract

Polypeptide sequences from *Petunia axillaris* (*PaxiN*) and *P. inflata* (*PinfS6*) as well as tomato, potato, *Nicotiana benthamiana*, and *Arabidopsis thaliana* were subjected to an all-vs-all blast search, then clustered into gene families with OrthoMCL. Of the resulting 27,600 families, 9430 families (39.2% of all genes) contained genes from all six species, and 24.6% of the genes were not clustered into any family. Most multispecies gene families followed the accepted evolutionary lineage, with the *Petunia*, *Solanum*, and Solanaceae clades sharing gene families far more often than other species groupings. Most of the genes unique to a single species or small clade were transposable element-related or had unknown functions, while most genes shared widely among species were involved with cellular and metabolic functions. Gene families with known functions that were almost unique to *Petunia* included several families of F-box proteins, which may function to ubiquitylate specific proteins destined for degradation, and several families of pentatricopeptide repeat genes, which are probably involved with RNA modification in the organelles. *Petunia* also contained several unique families of genes coding for cytochrome P450 and disease resistance genes. Among the families of genes coding for metabolic enzymes that differentially amplified in *Petunia* were genes for 1-aminocyclopropane-1-carboxylate synthase and oxidase, cupredoxin, HXXXD-type acyl-transferase, caffeate O-methyltransferase, and replication protein A. Tandemly duplicated genes fell into a similar set of functions, except that they did not include transposon genes or genes with unknown functions.

INTRODUCTION

Gene duplication is a major way of increasing an organism's diversity of response to specific environmental conditions (Ohno, 1970; Flagel and Wendel, 2009). Genes that have proliferated into multiple copy families can provide important clues to the specialized life habit of a species. For this reason we have studied gene families in the two *Petunia* species, grouping similar genes into families, determining which families have expanded, and analyzing their annotations. We concentrated on three types of comparison: *PinfS6* and *PaxiN* with each other, the two *Petunia* species with other Solanaceae with sequenced genomes, and all Solanaceae with *Arabidopsis thaliana*.

RESULTS

Gene Families Generated by OrthoMCL

For OrthoMCL analysis, genes from the six input species were filtered by requiring them to be at least 20 amino acids long and contain no internal stop codons. Only one representative peptide was used from each gene. A total of 251,612 genes were analyzed, with 35,233 from *PaxiN*, 38,826 from *PinfS6*, 34,727 from *S. lycopersicum*, 39,031 from *S. tuberosum*, 76,379 from *N. benthamiana*, and 27,416 from *A. thaliana*. Table 1 summarizes the number of genes and the number of gene families found in different possible combinations of species. A Venn diagram showing the percentage of genes in families found in all combinations of the five Solanaceae species can be found as Figure 2e in the main text.

Table 1. **Number of genes clustered into families by OrthoMCL.**

Species	Total genes	Genes in families with members of these species only	Families containing members of these species only	Singletons
P. axillaris (PaxiN)	35233	829	314	6615
P. inflata (PinfS6)	38826	1084	428	7782
S. lycopersicum (Sl)	34727	1414	445	8433
S. tuberosum (St)	39031	4243	752	7659
N. benthamiana (Nb)	76379	10848	2430	26425
A. thaliana (At)	27416	3820	1024	4890
Petunia (PaxiN PinfS6)	74059	5948	2402	
Solanum (Sl St)	73758	4469	1175	
All other groups of 2 (mean)	85403	398	94	
All other groups of 2 (range)	62113-115410	27-974	8-268	
Solanaceae (PaxiN PinfS6 Sl St Nb)	224196	23238	3076	
All other groups of 5 (mean)	206773	2041	300	
All other groups of 5 (range)	175233-216885	558-6343	89-946	
All species	251612	98683	9430	61804

OrthoMCL grouped these genes into 27,600 gene families, of which 9430 families contained genes from all six species. These common families had 98,683 genes, or 39.2% of all genes. In contrast, 24.6% of the genes (61,804 genes) were not placed in any gene family (singletons).

The number of gene families unique to specific groups of species was strongly influenced by the evolutionary relatedness of the two *Petunia* species and, to a lesser extent, the two *Solanum* species. There were 314 and 428 families containing only members of *PaxiN* and *PinfS6*, respectively, but 2402 families containing members of both species. Similarly, tomato and potato had 445 and 752 families unique to each species, but 1175 families had genes from both species. As a comparison, all other combinations of two species were found to have between 8 and 268 families (average 94 families) containing members of both species. When extended to all of the Solanaceae (i.e. all species except *Arabidopsis*), 3076 families were found to have genes from all five species. All other combinations of 5 species had between 89 and 946 families in common, with an average of 300.

Of the 27,600 gene families, 2347 families had at least five members from at least one species. These families were used for further analysis. The largest family, consisting of retrotransposon polyproteins found almost exclusively in *S. tuberosum*, contained 1026 members.

Gene Families Found in All Species

"Balanced" gene families were defined as having at least 30 members distributed across the six species, with fewer than 80% in any single species or genus. This category included 277 families containing 15,337 genes (data not shown). Most of the gene families in this category contained genes involved in normal cellular metabolism. Only three of these families, with a total of 99 genes, had no known function, and none of the balanced gene families seemed to be associated with transposable elements.

The largest balanced gene family contained subtilase family proteins, with 45-85 genes in each species. The largest 20 families also included six families of protein kinases, four families of membrane transporters, three families of carbohydrate-active enzymes, and two families of pentatricopeptide repeat proteins. Proteins in the latter group are involved with processing organellar RNAs (Barkan, 2011). Other types of family prominent on the list of balanced gene families include transcription factors, disease resistance genes, cytochrome P450 genes, and lipid-active genes.

Gene Families Found Primarily in a Single Species

Gene families that are almost unique to a single species were defined as having at least 30 family members with at least 80% of the members from a single species. There were 108 gene families containing 10,736 genes in this category (Table 2). The number of families varied widely among species, from a low of three families with 155 genes in *PaxiN* to a high of 62 families with 5434 genes in *N. benthamiana*. Families unique to *A. thaliana* were not included. In contrast to the balanced gene families, genes in the almost unique families were overwhelmingly either transposable element genes or genes with no known function. Only four families containing 394 genes are not clearly categorized in one of these two categories: a FAR1-related transcription factor, a 60S ribosomal protein, an F-box protein, and a chromo domain-containing protein.

One group of genes found frequently in species-specific groups is Ulp1 proteases, which regulate the removal of small ubiquitin-like modifier (SUMO) peptides in a cell cycle-dependent manner (Elmore et al., 2011). Ulp1 protease genes are found in several large families that are nearly species-specific in all of the Solanaceae species. Most of these genes are found adjacent to Mutator-like transposase genes. These finding suggest that the Solanaceae contain large numbers of CaMULE or Kaonashi elements (Hoen et al., 2006; van Leeuwen et al., 2007), which consist of an active Ulp1 protease gene coupled to a transposase gene, as seen in cucumber and Arabidopsis.

Table 2. **Near-unique gene families.**

A. *PaxiN* near-unique families: 3 families with 155 genes.[a]

group	total	paxi[b]	pinf	soly	sotu	nben	ath	Family Annotation
ppssna_223	48	43	1	0	2	2	0	RNase H
2 families	107	98	2	1	1	5	0	Unknown protein

B. *PinfS6* near-unique families: 6 families with 349 genes.

group	total	paxi	pinf	soly	sotu	nben	ath	Family Annotation
ppssna_119	67	3	59	0	3	2	0	gag-pol polyprotein
2 families	137	6	128	0	2	1	0	Ulp1 protease
3 families	145	5	137	2	0	1	0	Unknown protein

C. *Solanum lycopersicum* near-unique families: 7 families with 394 genes.

group	total	paxi	pinf	soly	sotu	nben	ath	Family Annotation
ppssna_84	85	2	6	70	4	0	3	Mutator-like transposase
ppssna_91	82	1	1	69	7	2	2	gag-pol polyprotein
ppssna_129	65	0	0	58	7	0	0	Ulp1 protease
ppssna_198	52	1	0	49	2	0	0	retrotransposon zinc finger CCHC-type protein
ppssna_301	40	0	0	40	0	0	0	Unknown protein
ppssna_325	38	0	0	37	0	0	1	transposase-related hAT dimerisation domain
ppssna_427	32	1	0	28	3	0	0	En/Spm-like transposon protein

D. *Solanum tuberosum* near-unique families: 33 families with 4496 genes.

group	total	paxi	pinf	soly	sotu	nben	ath	Family Annotation
ppssna_30	141	0	0	1	140	0	0	'chromo' domain containing protein

ppssna_259	44	0	0	4	40	0	0	RRNA intron-encoded homing endonuclease
15 families	2168	2	2	7	2155	2	0	gag-pol polyprotein
ppssna_11	260	0	0	0	260	0	0	Retrotransposon gag protein
10 families	1480	0	10	4	1466	0	0	Integrase core domain
6 families	403	1	4	5	386	7	0	Unknown Protein

E. *Nicotiana benthamiana* near-unique families: 59 families with 5342 genes.

group	total	paxi	pinf	soly	sotu	nben	ath	Family Annotation
ppssna_95	80	2	3	3	4	67	1	ribosomal protein L37a; Protein synthesis initiation factor
ppssna_186	53	0	0	0	0	53	0	F box protein
ppssna_50	120	3	2	2	0	113	0	FAR1 related sequence
4 families	241	0	0	1	4	236	0	gag-pol polyprotein
ppssna_48	121	0	0	0	0	121	0	Retrotransposon gag protein
ppssna_195	52	2	1	0	1	48	0	Endonuclease/exonuclease/phosphatase
ppssna_169	56	0	0	0	0	56	0	Ribonuclease H
ppssna_151	60	0	0	0	0	60	0	Transposon MuDR mudrA protein
ppssna_422	32	0	2	0	0	30	0	Mutator-like transposase
6 families	730	2	7	34	4	682	1	Ulp1 protease
44 families	3889	26	8	5	7	3843	0	Unknown protein

[a]The families listed here contained at least 30 members with least 80% of the genes in a single species.

[b]Species abbreviations as in Table 1.

Gene Families Found in a Single Genus

This study includes two pairs of species from the same genus: the two *Petunia* species, and the two *Solanum* species*, S. lycopersicum* and *S. tuberosum*. To examine genus-level differences, we searched for gene families with at least 30 members that had less than 80% of members from one species, but more than 80% in any two species (Table 3).

Of the eight families that were mainly found in the two *Petunia* species, six families had similar numbers of genes in *PinfS6* and *PaxiN*. The other two families were found at least four times more frequently in one family than the other, but did not reach the cutoff of 80% from one species to be classified as single-species families. The *Petunia*-specific gene families included three categories that were not unknown proteins or related to transposable elements: two families of cytochrome P450 genes, a family of HXXXD-type acyl-transferase family proteins, and a family of replication protein A 70 kDa DNA-binding subunit B genes. The latter two families are quite interesting as they represent potentially significant differences between *Petunia* and other Solanaceae; they are discussed below.

All of the three *Solanum*-specific gene families were primarily found in one species: two from *S. lycopersicum*, which were both transposable element genes, and one from *S. tuberosum*, a family of

genes with unknown function. This finding is consistent with the idea that the two *Solanum* species are more evolutionarily diverged than the two *Petunia* species.

We also found two gene families that reached the criterion of 80% of members from two different species that were not members of the same genus. One of these was a family of disease resistance genes that has many members in the two *Solanum* species as well as in *Arabidopsis*. The other was a family of self-incompatibility proteins that surprisingly was found primarily in *Arabidopsis*.

Table 3. Gene families in which 80% of members came from two different species.

A. Genus *Petunia*.

Group	Total	paxi[a]	pinf	soly	sotu	nben	ath	Family Annotation
ppssna_3	506	218	228	0	4	56	0	Mutator transposase
ppssna_85	85	33	42	0	2	8	0	Unknown protein
ppssna_100	77	6	60	4	1	6	0	Mutator transposase
ppssna_111	70	41	28	0	1	0	0	replication protein A 70 kDa DNA-binding subunit B
ppssna_128	65	48	11	1	5	0	0	RNase H family protein
ppssna_145	61	28	23	3	7	0	0	HXXXD-type acyl-transferase family protein
ppssna_197	52	26	26	0	0	0	0	Cytochrome P450
ppssna_276	42	16	19	2	2	3	0	Cytochrome P450

B. Genus *Solanum*.

Group	Total	paxi	pinf	soly	sotu	nben	ath	Family Annotation
ppssna_203	51	3	3	38	3	1	3	Helitron helicase-like protein
ppssna_230	47	2	2	36	5	2	0	Mutator transposase
ppssna_451	31	2	3	4	21	1	0	Unknown protein
ppssna_477	30	2	0	1	24	3	0	Unknown protein

C. Other Groups.

Group	Total	paxi	pinf	soly	sotu	nben	ath	Family Annotation
ppssna_33	136	2	4	16	36	4	74	Disease resistance protein (TIR-NBS-LRR class)
ppssna_456	30	2	0	3	4	0	21	Plant self-incompatibility protein S1 family

[a]Species abbreviations as in Table 1.

Gene families in Solanacae vs. *Arabidopsis*

We attempted to locate and analyze gene families that are well-distributed among the five Solanaceae species but rare in *Arabidopsis*, or which are almost unique to *Arabidopsis*. Table 4A shows a summary of gene families with at least 30 members that had fewer than 80% of genes in any of the Solanaceae species or in the *Petunia* or the *Solanum* genera, but which had fewer than 2.5% of their genes in *Arabidopsis*. These included 46 gene families containing 3285 genes. The largest annotation categories were disease resistance (749 genes), transposable elements (736 genes), cytochrome P450 (307 genes), and protein degradation (210 genes). In addition, there were 486 genes with no annotation other than

"unknown". At least some of these unknown genes are probably transposon-related: blastp searches using these genes with the NCBI nr database show some weak hits to both DNA transposons and retrotransposons. There were also 797 genes with functions that did not any of the other categories; these included several families of receptor protein kinases, genes involved in the biosynthesis of primary and secondary metabolites, and transporters.

Table 4. Numbers of genes and gene families found in functional categories: Solanaceae vs. *Arabidopsis*.

	families	paxi genes	pinf genes	soly genes	sotu genes	nben genes	ath genes
A. Solanaceae Gene Families.							
Total number	59	538	794	515	663	775	8
DNA transposon	9	55	172	183	27	111	0
Retrotransposon	3	38	7	22	47	44	0
F-box protein	0	0	0	0	0	0	0
cytochrome P450	4	58	86	40	83	40	0
DNA-interacting	1	6	6	5	8	7	0
Disease resistance	11	139	193	105	204	108	1
Metabolic enzymes	5	38	42	38	59	33	2
Protein degradation	6	68	112	20	50	22	1
Protein kinase	4	24	27	26	49	31	1
Carbohydrate active	4	39	54	33	46	40	0
Lipid active	2	10	13	18	29	13	2
Other	1	12	12	6	4	7	1
Unknown protein	9	51	70	19	57	319	0
B. *Arabidopsis* Gene Families							
Total number	15	11	9	7	8	5	913
DNA transposon	0	0	0	0	0	0	0
Retrotransposon	0	0	0	0	0	0	0
F-box protein	6	6	4	3	5	2	469
cytochrome P450	1	1	2	1	1	1	33
DNA-interacting	0	0	0	0	0	0	0
Disease resistance	0	0	0	0	0	0	0
Metabolic enzymes	0	0	0	0	0	0	0
Protein degradation	0	0	0	0	0	0	0
Protein kinase	2	1	2	2	1	1	80
Carbohydrate active	2	0	0	0	0	0	79
Lipid active	0	0	0	0	0	0	0
Other	3	2	1	1	1	0	216
Unknown protein	1	1	0	0	0	1	36

Table 4B shows 15 gene families (913 genes) that had at least 80% of their members from *Arabidopsis*. The largest single annotation category (469 *Arabidopsis* genes) is proteins containing an F-box. There were also 36 genes with unknown functions and 408 genes in other categories. The unexpected lack of transposable element genes may be attributable to the removal of such genes in the original *Arabidopsis* annotation process.

Functional Categories of Gene families Expanded in *Petunia*

In a further attempt to find interesting gene families that were expanded in *Petunia*, we identified all gene families with at least five members from one of the *Petunia* species and no more than one member from each of the non-*Petunia* species. The families were divided into *PaxiN*-specific (30 families with 261 genes), *PinfS6*-specific (24 families with 198 genes), and *Petunia*-specific (88 families with 1292 genes) categories. The families were then categorized (Table 5).

Table 5. **Gene families with at least 5 members from *Petunia* and no more than 1 member from each of the non-*Petunia* species.**

Family	Total	paxi	pinf	soly	sotu	nben	ath	Family annotation
A. *PaxiN*								
ppssna_3008	11	1	10	0	0	0	0	Blue copper protein; cupredoxin
ppssna_4370	9	9	0	0	0	0	0	AT hook motif DNA-binding family protein
ppssna_5915	8	7	1	0	0	0	0	Cyclin-D-binding Myb-like transcription factor
ppssna_5912	8	8	0	0	0	0	0	1-aminocyclopropane-1-carboxylate synthase
ppssna_5909	8	8	0	0	0	0	0	Cytochrome P450
ppssna_14546	5	5	0	0	0	0	0	Ribonuclease H-like domain
6 families	58	55	3	0	0	0	0	Mutator-like Transposase
3 families	28	27	1	0	0	0	0	Endonuclease/exonuclease/phosphatase
13 families	115	103	12	0	0	0	0	Unknown protein
B. *PinfS6*								
ppssna_3024	11	0	11	0	0	0	0	CCR4-NOT transcription complex subunit
2 families	11	5	6	0	0	0	0	Ribonuclease H
ppssna_5832	8	0	8	0	0	0	0	mitogen-activated protein kinase 12
ppssna_14492	5	0	5	0	0	0	0	protein disulfide isomerase
19 families	0	0	0	0	0	0	0	Unknown protein
C. *Petunia* genus								
11 families	129	52	62	6	5	3	1	F-box family protein
ppssna_111	70	41	28	0	1	0	0	Replication protein A
ppssna_197	52	26	26	0	0	0	0	Cytochrome P450
ppssna_699	24	7	17	0	0	0	0	FAR1-related protein
ppssna_759	23	5	18	0	0	0	0	Disease resistance protein (CC-NBS-LRR class)

ppssna_886	21	7	13	1	0	0	0	FAR1-related protein
ppssna_1900	14	6	7	1	0	0	0	ubiquitin-protein ligase
ppssna_2246	13	5	4	1	1	1	1	cysteine-type endopeptidase inhibitor
ppssna_2248	13	4	6	1	1	1	0	Interferon related developmental regulator
ppssna_2593	12	4	8	0	0	0	0	Zinc finger, RING/FYVE/PHD-type
ppssna_2595	12	6	6	0	0	0	0	Histone superfamily protein
ppssna_2594	12	4	6	1	1	0	0	HXXXD-type acyl-transferase
ppssna_2591	12	10	2	0	0	0	0	1-aminocyclopropane-1-carboxylate oxidase
ppssna_3010	11	3	8	0	0	0	0	Nse4 component of Smc5/6 DNA repair complex
ppssna_3036	11	7	4	0	0	0	0	HXXXD-type acyl-transferase
ppssna_3019	11	7	4	0	0	0	0	GDSL esterase/lipase
ppssna_3020	11	7	4	0	0	0	0	ATP synthase subunit-like protein
ppssna_3027	11	5	3	1	1	1	0	UDP-Glycosyltransferase
ppssna_3531	10	5	4	0	0	1	0	GDSL esterase/lipase
ppssna_3551	10	2	6	0	1	1	0	RNase Phy3 [*Petunia* x hybrida]
ppssna_3544	10	3	7	0	0	0	0	Disease resistance protein (CC-NBS-LRR class)
ppssna_4347	9	5	4	0	0	0	0	HXXXD-type acyl-transferase
ppssna_5824	8	1	5	1	0	1	0	FAR1-related protein
ppssna_5835	8	3	5	0	0	0	0	serpin serine protease inhibitor
ppssna_5846	8	1	5	1	1	0	0	Caffeate O-methyltransferase (COMT) family)
ppssna_5878	8	2	5	0	1	0	0	Cytochrome P450
ppssna_11782	6	6	0	0	0	0	0	Zinc finger, CCHC-type; transcription regulation
13 families	172	103	61	1	2	5	0	Mutator-like Transposase
2 families	93	11	80	1	0	1	0	Ulp1 protease
ppssna_3011	11	0	9	0	1	1	0	Helitron helicase protein
ppssna_1177	18	17	0	0	1	0	0	gag non-LTR retrotransposase
4 families	84	66	10	3	3	2	0	Ribonuclease H
2 families	63	10	53	0	0	0	0	Aminotransferase-like, plant mobile domain
26 families	275	130	124	7	6	8	0	Unknown protein

Most of the gene families that are expanded in one *Petunia* species relative to the other, or in both *Petunia* species relative to the other Solanaceae, are transposable element genes (35 families with 543 genes) or genes with no known function (57 families with 553 genes).

Both cytochrome P450 genes (three families with 60 genes) and disease resistance genes (two families with 33 genes) of the CC-NBS-LRR type (Moffett et al, 2002; McHale et al., 2006) are found in several families that are shared widely among all species, and in other families are almost species-specific (Tables 3 and 7). Both of these categories contain many genes, some of which are needed for common disease and metabolic issues, while others are needed for species-specific problems.

The DNA-interacting proteins include several families of transcription factor as well as some genes that may be involved in chromatin remodeling, DNA repair, or other higher level regulatory functions. Similarly, there are four families of genes that can be categorized as affecting protein degradation, including 11 families of F-box proteins, which are involved with ubiquitylation of proteins targeted for degradation.

The most interesting families in this analysis have relatively specific metabolic functions that may be useful clues to *Petunia* phenotypes; several of these genes are discussed more fully below.

Tandem duplications

We performed a search for tandem duplications that were amplified in *Petunia* relative to tomato. The size distribution of tandem arrays in the three genomes is quite similar, with *PaxiN* having 7732 genes in 2865 tandem arrays, *PinfS6* having 7883 genes in 2967 tandem arrays, and tomato having 8715 genes in 3018 arrays (Figure 1). Despite similar size distributions, the *Petunia* tandem arrays differ strikingly from the tomato arrays in that the average distance between array members in both *Petunia* species is more than twice that of tomato (Table 6).

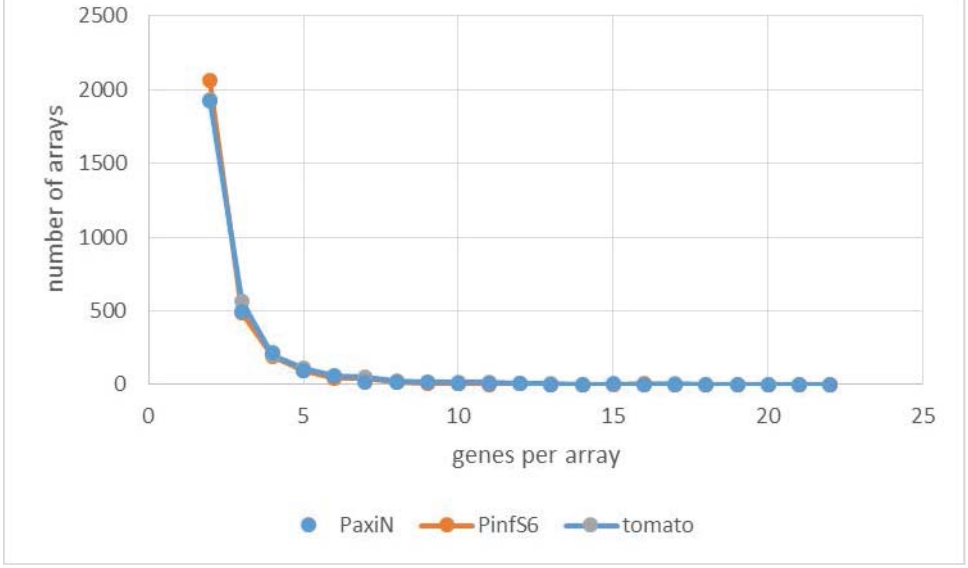


Fig 1. Number of tandem arrays vs. number of genes per array.

Individual tandem arrays were compared between species by examining the best blastp hits and determining which arrays they belonged to. Arrays were considered to be of significantly different sizes if the sum of the array sizes of best blastp hit genes in the target species was less than half the array size in the original species.

Of the 26 tandem arrays with 10 or more members in *PaxiN*, ten were unique to this species, while six of the 20 large arrays were unique to *PinfS6*. Out of 46 total large arrays in the two species combined, 19 were found in *Petunia* but not tomato (Table 7). Prominent among these were three arrays of pentatricopeptide genes unique to *PaxiN*; these genes are involved in RNA editing in the organelles (Nakamura et al., 2012). Three arrays of genes encoding 2-oxoglutarate and Fe(II)-dependent dioxygenase superfamily proteins, two in *PaxiN* only and one in both *Petunia* species, were found; these enzymes play a variety of metabolic roles including hormone and pigment synthesis as well as DNA repair and histone demethylation (Farrow and Facchini, 2014). Each *Petunia* species also has a unique array of cytochrome P450 genes. Many of the *Petunia*-specific tandem arrays encode F box protein

genes, many of which are involved in ubiquitylation of proteins destined to be degraded as well as hormone signaling pathways and self-incompatibility (Smalle and Vierstra, 2004). The arrays that are shared by all three species, including glutaredoxin, protease inhibitors, and genes involved with carbohydrate metabolism.

Table 6. Average distance between tandem array members.

	PaxiN		*PinfS6*		tomato	
Distance	number of arrays	cumulative percentage	number of arrays	cumulative percentage	number of arrays	cumulative percentage
< 10,000 bp	958	33.4	967	32.6	1678	55.6
10,000 - 49,999 bp	1273	77.9	1327	77.3	998	88.7
50,000 - 99,999 bp	327	89.3	393	90.6	167	94.2
100,000 -199,999 bp	215	96.8	201	97.3	88	97.1
200,000 - 499,999 bp	77	99.5	68	99.6	51	98.8
500,000 - 999,999 bp	15	100	8	99.9	21	99.5
> 1,000,000 bp	0	100	3	100	15	100
total number of arrays	2865		2967		3018	
median average dist	19331		20504		8384	

Table 7. Large tandem arrays that are expanded in *Petunia* species.

Array number	size	annotation
***PaxiN* only.**		
paxi_1645	22	Pentatricopeptide repeat-containing protein
paxi_291	15	2-oxoglutarate and Fe(II)-dependent oxygenase superfamily protein
paxi_887	14	Pentatricopeptide repeat-containing protein
paxi_1870	12	LRR receptor-like serine/threonine-protein kinase
paxi_1311	11	F-box protein
paxi_331	11	cytochrome P450
paxi_582	10	Protein phosphatase 2C family protein (signal transduction)
paxi_1014	10	Pentatricopeptide repeat-containing protein
paxi_1881	10	Chaperone DnaJ-domain superfamily protein
paxi_2296	10	2-oxoglutarate and Fe(II)-dependent oxygenase superfamily protein
***PinfS6* only.**		
pinf_1629	16	Unknown protein GO:0003677 (DNA binding)
pinf_2365	16	Disease resistance protein (CC-NBS-LRR class)
pinf_941	13	Cytochrome P450
pinf_848	10	Actin
pinf_891	10	F-box family protein
pinf_970	10	unknown protein, chloroplast related

***PaxiN* and *PinfS6*.**

paxi_2069	17	Polyadenylate-binding protein 2 (=pinf_1516)
paxi_1140	15	F-box/FBD/LRR-repeat protein
paxi_2715	13	Cc-nbs-lrr, resistance protein with an R1 specific domain (=pinf_1603)
paxi_24	12	Protein of Unknown Function (DUF239)
paxi_65	12	2-oxoglutarate and Fe(II)-dependent oxygenase superfamily protein (=pinf_2470)
paxi_558	12	F-box/RNI-like superfamily protein (=pinf_936)
paxi_848	11	F-box/RNI-like superfamily protein
paxi_853	11	F-box/FBD/LRR-repeat protein
paxi_1101	11	Cc-nbs-lrr, resistance protein
paxi_1890	11	F-box/FBD/LRR-repeat protein (=pinf_188)
paxi_2438	10	Unknown protein
pinf_1603	18	Cc-nbs-lrr, resistance protein with an R1 specific domain (=paxi_2715)
pinf_2470	16	2-oxoglutarate and Fe(II)-dependent oxygenase superfamily protein (=paxi_65)
pinf_2209	14	F-box/RNI-like superfamily protein
pinf_1516	12	Polyadenylate-binding protein 2 (=paxi_2069)
pinf_188	11	F-box/FBD/LRR-repeat protein (=paxi_1890)
pinf_323	10	F-box family protein
pinf_936	10	F-box/RNI-like superfamily protein (=paxi_558)
pinf_1451	10	wall-associated kinase

***PaxiN* and *S. lycopersicum*.**

paxi_292	18	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
paxi_1683	15	SAUR-like auxin-responsive protein family
paxi_38	10	cytochrome P450

***PinfS6* and *S. lycopersicum*.**

pinf_967	13	alpha/beta-Hydrolases superfamily protein

All three species.

paxi_2165	12	subtilisin-like serine protease (=pinf_1903)
paxi_913	10	Glutaredoxin (=pinf_1755)
paxi_1778	10	UDP-glucosyltransferase
pinf_468	12	Serpin (Serine protease inhibitor)
pinf_902	12	cytochrome P450
pinf_1903	11	subtilisin-like serine protease (=paxi_2165)
pinf_526	10	xyloglucan specific endoglucanase inhibitor
pinf_1755	10	Glutaredoxin (=paxi_913)

DISCUSSION

The gene families used in this study were generated by OrthoMCL from an all-vs-all blastp search. OrthoMCL is a clustering method and as such the groups it forms are partly based on unknown internal parameters and vagaries of the input data. However, several intermediate sets of gene families were generated from various combinations of the species, and we found that gene families were relatively stable. On the average, the effect of adding the genes from one additional species was that 96.9% of the genes in the largest 500 gene families remained clustered together in the same family. Also, an average of 434 of these families had all members clustered into the same family when an additional species was added, and an average of 469.5 families had at least 90% of their members clustered into the same family. In general, gene family membership remained intact when new sets of genes were added. For this reason we believe that when genes of the same functional annotation are placed in different gene families, the two families will be detectably different when subjected to phylogenetic analysis and probably functionally differentiated.

The evolution of transposons, as well as that of the Ulp1 protease sequences that are associated with many Mutator-like transposons, is apparently much less constrained than other gene families, as these genes are the largest category of species- and genus-specific genes. The large number of genes with unknown functions in the species-specific categories might also include novel transposons. However, genes with unknown function may also include some with functions that are truly novel to *Petunia*, and they may also include some sequences that are incorrectly labeled as genes.

The most interesting gene families that have been differentially amplified in *Petunia* are those with known metabolic functions that might point to *Petunia*-specific activities. Genes coding for 1-aminocyclopropane-1-carboxylate synthase (ACS) were found in all the species studied, with between 9 and 17 members per genome. However, *PaxiN* has an additional 18 ACS genes in two families, with only two *PinfS6* genes in one family and none in the other. The enzyme coded by these genes is the rate-limiting step in ethylene biosynthesis. It is encoded as a multigene family in all sequenced plant genomes, with different isozymes that have different patterns of gene expression (Yamagami et al., 2003). Similarly, both *Petunia* species share a family of 1-aminocyclopropane-1-carboxylate oxidase genes, not found in the other species. These genes code for the final step in ethylene production (Kende, 1989).

The HXXXD-type acyl-transferase genes are found in both *Petunia* species in relatively equal numbers, but there are very few members from any of the other species. An important sub-category of these enzymes are the hydroxycinnamoyl CoA quinate transferases, which are involved in the production of chlorogenic acids (Sonnante et al., 2010). These secondary metabolites have multiple functions in plants, including lignin biosynthesis, UV light protection and pest resistance.

Caffeate O-methyltransferase (COMT) family enzymes are involved with the biosynthesis of many phenylpropanoids, including lignin, anthocyanin, and a number of defensive and stress-induced compounds (Joshi and Chiang, 1998). There is one member of this family in each of the Solanaceae except *PinfS6*, which has five members.

PaxiN has a family of genes that contain a cupredoxin fold, including blue copper protein and laccase. Genes of this general type are found in all the species studied, with several different families found in each species. In *Petunia*, one family has ten copies in *PinfS6* but only one in *PaxiN*, with no

members in any of the other species. These proteins perform oxidation-reduction reactions in photosynthesis, oxygenations and they have an intense blue color (Nersissian et al., 1998).

FAR1-related sequences are transcription factors that mediate the far-red light responses in higher plants. They have a domain structure similar to Mutator-like transposases. It is possible that the gene families annotated as FAR1-related are transposon-related (Hudson et al., 2003).

Replication protein A is a heterotrimer composed of 70 kDa, 32 kDa, and 14 kDa subunits that is involved with several aspects of DNA metabolism, including replication and repair. Animals and yeast have only a single set of RPA genes, but plants have at least two distinct types of RPA proteins, which may have separated replication functions from repair functions, and which may be specific to organelles or the nucleus (Sakaguchi et al., 2009). Gene family ppssna_111 contains 70 members, 69 of which are from *Petunia*.

METHODS

Peptide sequences from *S. lycopersicon*, *S. tuberosum*, and *N. benthamiana* were downloaded from the Sol Genomics Network web site (http://solgenomics.net). For tomato, the ITAG Release 2.3 genomic annotations were used (Tomato Gene Consortium, 2012). The PGSC version 3.4 was used for *S. tuberosum* (Xu et al., 2011). *N. benthamiana* draft genome sequences from release 0.4.4 (Bombarely et al., 2012). The *Arabidopsis thaliana* peptides and annotations were downloaded from http://www.arabidopsis.org , using version TAIR10.

Blastp from blast+ version 2.2.27 was used to do an all-versus-all comparison of the peptides from the six species (Altschul et al., 1997). The blastp results were used to group genes into families with OrthoMCL version 2.0.8 (Li et al., 2003). The annotations from the genes grouped into families were analyzed using custom software written in Perl. Functional annotations for groups were created by attempting to find a consensus among annotations for individual genes, with a bias against leaving the function as “unknown”.

Tandem duplications in the two *Petunia* species and in tomato were detected using the SynMap function of CoGe (Lyons and Freeling, 2008; Lyons et al., 2008), and were defined as two or more genes with significant sequence homology lying within ten genes of each other on the same contig. Individual tandem arrays were compared between species by examining the best blastp hits and determining which arrays they belonged to. Arrays were considered to be of significantly different sizes if the sum of the array sizes of best blastp hit genes in the target species was less than half the array size in the original species.

ACKNOWLEDGEMENTS

The authors thank the NIU Department of Computer Science for use of its Gaea computing cluster.

AUTHOR CONTRIBUTIONS

Both authors performed the research and analyzed the data. M.A.J. wrote the article.

REFERENCES

Altschul, S.F., Madden,T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller,W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**: 3389-3402.

Barkan, A. (2011). Expression of plastid genes: Organelle-specific elaborations on a prokaryotic scaffold. Plant Physiol. **155**: 1520–1532.

Bombarely, A., Rosli, H.G., Vrebalov, J. , Moffett, P. , Mueller, L.A. , and Martin, G.B. (2012). A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. Molecular Plant-Microbe Interactions **25**: 1523-1530.

Elmore, Z.C., Donaher, M., **Matson, B.C.**, **Murphy, H.**, **Westerbeck, J.W.,** and **Kerscher, O**. (2011). Sumo-dependent substrate targeting of the SUMO protease Ulp1. BMC Biology **9**: 74.

Farrow, S.C. and Facchini, P.J. (2014). Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. Front. Plant Sci. **5**: 524.

Flagel, L.E. and Wendel, J.F. (2009). Gene duplication and evolutionary novelty in plants. New Phytol. **183**: 557-564.

Hillwig, M.S., Liu, X., Liu, G., Thornburg, R.W., and MacIntosh, G.C. (2010). Petunia nectar proteins have ribonuclease activity. J Exp. Bot. **61**: 2951–2965.

Hoen, D.R., Park, K.C., Elrouby, N., Yu, Z., Mohabir, N., Cowan, R.K., and Bureau, T.E. (2006). Transposon-mediated expansion and diversification of a family of ULP-like genes. Mol. Biol. Evol. **23**: 1254-1268.

Hudson, M.E., Lisch, D.R., and Quail, P.H. (2003). The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. Plant J. **34**: 453–471.

Joshi, C.P. and Chiang, V.L. (1998). Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. Plant Mol. Biol. **37**: 663–674.

Kende, H. (1989). Enzymes of ethylene biosynthesis. Plant Physiol. **91**: 1-4.

Li, L, Stoekert, C. J., and Roos, D.S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. **13**: 2178-2189.

Lyons, E., and Freeling M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J. **53**: 661-673.

Lyons E, Pedersen B, Kane J, and Freeling M (2008). The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates rosids. Tropical Plant Biol. **1**: 181-190**.**

McHale, L., Tan, X., Koehl, P., and Michelmore, R.W. (2006). Plant NBS-LRR proteins: adaptable guards. Genome Biology **7**: 212.

Moffett, P., Farnham, G., Peart, J., and Baulcombe, D.C. (2002). Interaction between domains of a plant NBS–LRR protein in disease resistance-related cell death. EMBO J. **21**: 4511–4519.

Nakamura, T., Yagi, Y., and Kobayashi, K. (2012). Mechanistic insights into pentatricopeptide repeat proteins as sequence-specific RNA-binding proteins for organellar RNAs in plants. Plant Cell. Physiol. **53**: 1171-1179.

Nersissian, A.M., Immoos, C., Hill, M.G., Hart, P.J., Williams, G., Herrmann, R.G., and Valentine, J.S. (1998). Uclacyanins, stellacyanins, and plantacyanins are distinct subfamilies of phytocyanins: Plant-specific mononuclear blue copper proteins. J. Biol. Chem. **278**: 49102-49112.

Ohno, S. (1970). *Evolution by Gene Duplication.* Springer-Verlag, New York.

Sakaguchi, K., Ishibashi, T., Uchiyama, Y., and Iwabata, K. (2009). The multi-replication protein A (RPA) system – a new perspective. FEBS J. **276**: 943–963.

Smalle, J. and Vierstra, R.D. (2004). The ubiquitin 26S proteasome proteolytic pathway. Annu. Rev. Plant Biol. **55**: 555-590.

Sonnante, G., D'Amore,R., Blanco, E., Pierri, C.L., De Palma, M., Luo, J., Tucci, M., and Martin, C. (2010). Novel hydroxycinnamoyl-coenzyme A quinate transferase genes from artichoke are involved in the synthesis of chlorogenic acid. Plant Physiol. **153**: 1224-1238.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635-641.

Van Leeuwen, H., Monfort, A., and Puigdomenech, P. (2007). Mutator-like elements identified in melon, Arabidopsis and rice contain ULP1 protease domains. Mol. Genet. Genomics **277**: 357-364.

Xu, X. et al. (2011). Genome sequence and analysis of the tuber crop potato. Nature **475**: 189-195.

Yamagami, T., Tsuchisaka, A., Yamada,K., Haddon, W.F., Harden, L.A. and Theologis, A. (2003). Biochemical diversity among the 1-amino-cyclopropane-1-carboxylate synthase isozymes encoded by the Arabidopsis gene family. J. Biol. Chem **278**: 49102-49112.

Supplementary Note 5

Incomplete gene fractionation after paleopolyploidy: the first study case in flowering plants revealed by comparison of the *Petunia axillaris N*. and *Solanum lycopersicum* genomes

Laurie Grandont[1 +], Haibao Tang[2, 3 +], Mitrick Johns[4], Eric Lyons[3]*, M. Eric Schranz[1]*

[1]Biosystematics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.
[2] Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian Province, China.
[3] School of Plant Sciences, iPlant Collaborative, University of Arizona, Tucson, AZ 85721, USA.
[4]Department of Biological Sciences, Northern Illinois University, DeKalb IL 60115, USA.

[+] These authors contributed equally to this work

*for correspondence: eric.schranz@wur.nl; ericlyons@email.arizona.edu

Running title: Comparative genomic analysis of *Petunia* and tomato

ABSTRACT

Polyploidy events, or whole genome duplications, have had a strong impact on land plant diversification, adaptation and speciation. Genomic investigations have found that polyploidy is ubiquitous among angiosperms and have identified independent lineage-specific ancient polyploidizations. Traces of these ancient polyploidy events, or paleopolyploidy events, can still be identified although duplication events are followed by massive gene loss (fractionation) and chromosomal structural rearrangements. The sequencing of the genome of *Petunia axillaris N* offers an ideal opportunity to study paleopolyploidy and gene fractionation in the evolutionary context of radiation due to the unique phylogenetic location of *Petunia* in the Solanaceae family. Our study confirms the previously inferred Solanaceae paleohexaploidy event. We also demonstrate that the *Petunia* lineage has experienced at least two rounds of paleohexapolyploidization, the older gamma hexaploidy event, which is shared with other Eudicots, and the more recent Solanaceae paleohexaploidy event that is shared with tomato and other Solanaceae species. Despite the shared paleohexaploidy event, we found that the process of gene fractionation is less profound in Petunia compared to tomato. This indicates that fractionation of gene content was not complete when these lineages diverged and independent gene loss events may have contributed to the speciation of the lineages, similar to what has been observed in *Saccharomyces* yeasts but so far not shown in flowering plants.

INTRODUCTION

Genomic analysis of numerous plant species has found that polyploidy is ubiquitous among angiosperms, with shared and independent lineage-specific ancient polyploidy (or paleopolyploidy) events (Soltis et al., 2008; Jiao et al., 2011). While most paleopolyploidy events are ancient genome doublings (paleotetraploidies), there are a few important examples of ancient triplications (paleohexaploidies). For example, the gamma polyploidy event near the origin of Eudicots (e.g. Rosids and Asterids) is a genome triplication most clearly seen by the analysis of the grape genome (Jaillon et al., 2007). Ancient genome triplication events have also been detected by analysis of Brassica and Cleomaceae species (Barker et al., 2009; Wang et al., 2011; Cheng et al., 2013). Additionally, the genome analysis of tomato suggested that there was an ancient polyploidy event somewhere during the evolution of the plant family Solanaceae (Tomato Genome Consortium, 2012), however the exact timing and nature of this triplication is not yet established.

Sub-genomes created by paleo-polyploidization can differentiate in terms of gene density due to uneven or biased gene fractionation (Thomas et al., 2006), and levels of gene expression due to genome dominance (Schnable et al., 2011). Biased gene fractionation and genome dominance have been found between subgenomes/duplicated regions in some plants and yeast (Schnable et al., 2011; Wang et al., 2011). In other plants, a second pattern of gene fractionation involving no bias either in gene fractionation or in genome dominance has been observed (Garsmeur et al., 2013; Chalhoub et al., 2014). Interestingly, patterns of genome fractionation from a shared polyploidy event seem to be similar within plant clades (for example, across grasses and across crucifers). This contrasts with observations from *Saccharomyces* yeasts where early-branching clades show an independent pattern of genome fractionation than later branching species (Scannell et al., 2006). Paleopolyploidy events may contribute to the radiation between crown-groups (large phylogenetic clades) and sister-groups (smaller early-branching clades) by providing genetic material for the evolution of novel trait(s) (reviewed in Schranz et al., 2012; Tank et al., 2015). However, it is not yet clear why there appears to be a lag between the timing of the polyploidy event and the eventual radiation of species (referred to as the WGD radiation lag-time). One possible cause and/or mechanism could be due to the time to establish differential genome fractionation between crown-group species and early-branching sister-species. However, to date very few sister-lineages have been sequenced (*Aethionema* of the Brassicaceae being a rare example of this) to address this question.

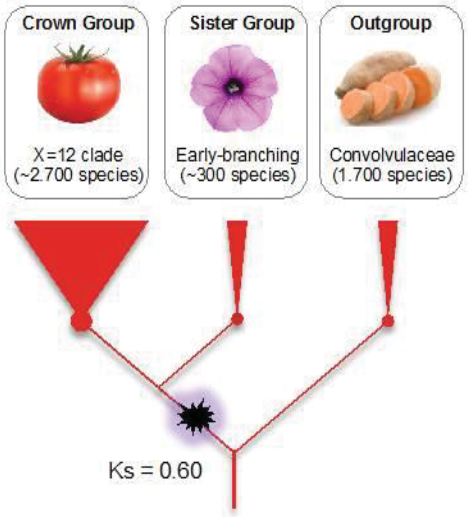


Figure 1: Simplified phylogeny of crown-group Solanaceae (x=12 clade) which includes most important crop species in relation to the early-branching sister-group including Petunia and the out-group family Convovulaceae including sweet potato. The Crown- and Sister-groups of the Solanaceae share an ancient genome triplication (paleohexaploidy) (shown by star-burst) with duplicate gene pairs having an average Ks divergence of 0.60.

In this study we focus on the paleopolyploid history and gene fractionation in two Solanaceae species, tomato (a member of the crown-group) and *Petunia axillaris N* (a member of the smaller-sister group). The Solanaceae plant family contains more than 3000 species, with most of the species found in the crown-group defined as the large "x=12" clade (Särkinen et al., 2013) and the Convolulaceae, including sweet potato being the outgroup family (Figure 1). The Solanaceae crown-group contains a variety of important crops such as tomato, potato, tobacco, and eggplant. Genome collinearity analysis between tomato and grape has been used in establishing a genome triplication in the Solanaceae (Tomato Genome Consortium, 2012). Genomic dot plots revealed that much of the tomato genome was covered in regions either duplicated or triplicated in comparison to grape (Tomato Genome Consortium, 2012). While the existence of triplicated segments suggested that a genome triplication was highly probable in the Solanaceae lineage, they could also have arisen through genome doubling followed by segmental duplications. These ambiguous genomic patterns have led to alternative interpretations. Analysis of the *Mimulus guttatus* genome, which is another related lineage to the Solanaceae in the family Phrymaceae, does not share the Solanaceae paleohexaploidy event, but has an independent paleotetraploidy event (Ibarra-Laclette et al., 2013). Analysis of *Mimulus* and tomato did not yield additional insight into the structure of the Solanaceae paleohexplaoidy event due to the independence of their respective WGD events.

Here we validate previous findings and investigate the role of whole genome triplication and gene fractionation in the evolutionary context of radiation by taking advantage of the phylogenetic location of *Petunia*, which is part of the Solanaceae "first-branching" or "sister-group". We showed that *Petunia* has experienced a paleohexaploidy event subsequent to its divergence from the grape lineage and that this event is dated to have occurred before tomato-*Petunia* split (and thus predated the other x=12 crown-group species as well). Additionally, we showed that following this shared genome triplication; the tomato genome has retained fewer genes than the *Petunia* genome. The high fractionation level in the tomato lineage may have contributed to the initial difficulties in clearly identifying the three triplicated regions within the tomato genome. We finally show that gene fractionation occurred in "two steps" with a first shared fractionation process in the *Petunia* and tomato lineages consecutive/subsequent to their common Solanaceae paleohexaploidy event. Gene fractionation then continued independently following the divergence of these two lineages. This is evidence that fractionation of gene content was not complete when these lineages diverged and may have contributed to the diversification of the lineages, similar to what has been observed in *Saccharomyces* yeasts but until now not yet described in flowering plants.

RESULTS

Petunia shares the Eudicot paleohexaploidy event

We first addressed if the *Petunia* genome shares the gamma genome triplication with other Eudicots. We performed whole genome collinearity analysis of grape and *Petunia axillaris N* (Figure 2). The genome hexaploidy event, or "gamma", could be clearly highlighted in a self-self grape comparison, as the grape genome has not undergone a more recent polyploidy event since gamma (Jaillon et al., 2007). Indeed, by analyzing a given genomic region by traversing the grape/grape dot plot horizontally we identified the three syntenic regions generated by the Eudicot paleohexaploidy gamma event (Figure 2a; see red dashed lines for example). Similarly, in the grape vs. *Petunia* dot plot we observed three syntenic regions (in green), called out-paralogs, which also originated from the gamma event (Figure 2b; see red dashed lines). In the grape-*Petunia* dot plot, the syntenic blocks on the diagonal correspond to the orthologous regions between grape and *Petunia*. The orthologous regions show a smaller Ks value (~1.58) than the "gamma" regions (~2.69), confirming that the gamma triplication predates the divergence of grape and *Petunia* (Figure 2c).

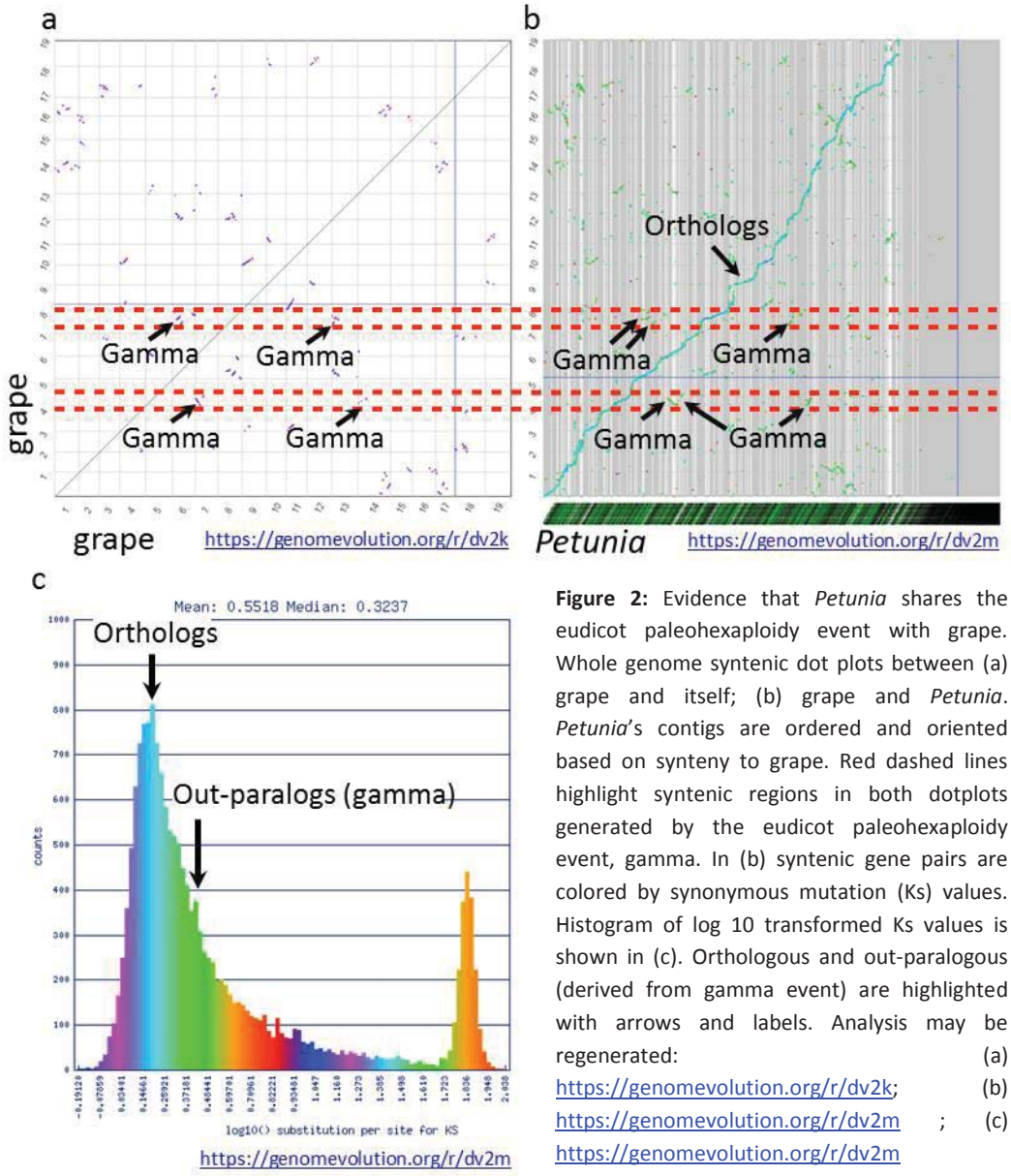


Figure 2: Evidence that *Petunia* shares the eudicot paleohexaploidy event with grape. Whole genome syntenic dot plots between (a) grape and itself; (b) grape and *Petunia*. *Petunia*'s contigs are ordered and oriented based on synteny to grape. Red dashed lines highlight syntenic regions in both dotplots generated by the eudicot paleohexaploidy event, gamma. In (b) syntenic gene pairs are colored by synonymous mutation (Ks) values. Histogram of log 10 transformed Ks values is shown in (c). Orthologous and out-paralogous (derived from gamma event) are highlighted with arrows and labels. Analysis may be regenerated: (a) https://genomevolution.org/r/dv2k; (b) https://genomevolution.org/r/dv2m ; (c) https://genomevolution.org/r/dv2m

***Petunia* shares the more recent hexaploidy event with crown-group Solanaceae**

Refined analyses of orthologous regions from our grape-*Petunia* comparison also reveal that the *Petunia* lineage underwent one additional independent hexaploidy event after its divergence from the grape lineage (Figure 3). By zooming in the grape-*Petunia* dot plot along the orthologous regions, we observed that cluster of orthologous regions appear to be either duplicated or triplicated in *Petunia* (Figure 3b), providing evidence for a subsequent polyploidy event (either tetraploidy or hexaploidy) in the *Petunia* lineage. These duplicated or triplicated segments were isolated as "boxed" areas along the diagonal, as a result of the Syntenic Path Assembly (SPA) algorithm in CoGe SynMap (Lyons et al., 2011). We then scrutinized manually each boxed region for microsynteny and duplicated regions were subsequently searched in order to capture additional syntenic regions, which can be missed or misplaced by the automated syntenic path assembly algorithm used to order *Petunia*'s contigs. In all cases of syntenic regions that appeared duplicated in the grape-*Petunia* dot plot, an additional syntenic region was identified in an exhaustive search. More precisely, we carefully analyzed the microsynteny for each box regions identified in the Figure 3b (labeled A-I) as shown in Figures 4-6. Each figure identifies three syntenic orthologous regions of *Petunia* to one region of grape, consistent with a paleohexaploidy event in *Petunia*.

The *Petunia* regions show consistent patterns of fractionation following polyploidy when compared to an unduplicated outgroup region from grape. Of the nine regions from grape analyzed, all but one had three syntenic regions in *Petunia*, thus confirming that the triplication is genome-wide although with varying degrees of clarity. The genomic regions A, C, F, G, H and I showed similar patterns of retention/loss (fractionation) across each of the triplicated regions. Note that in region F, it is possible to see a genomic insertion in grape. By contrast, the third (small) syntenic region found by SynFind for the B region is highly fractionated with relatively fewer retained duplicated genes. The D and E regions appear to be more complicated and are impacted by other genomic rearrangement events. For the D region, a number of tandem duplications/inversions are present. In addition, the third region shows weak synteny with the grape genome indicating a higher degree of fractionation or poor assembly. For the E region, the first and the third *Petunia* regions do not have overlapping synteny with grape . These results provide unequivocal evidence that *Petunia* lineage has an independent hexaploidy event after its divergence with grape lineage.

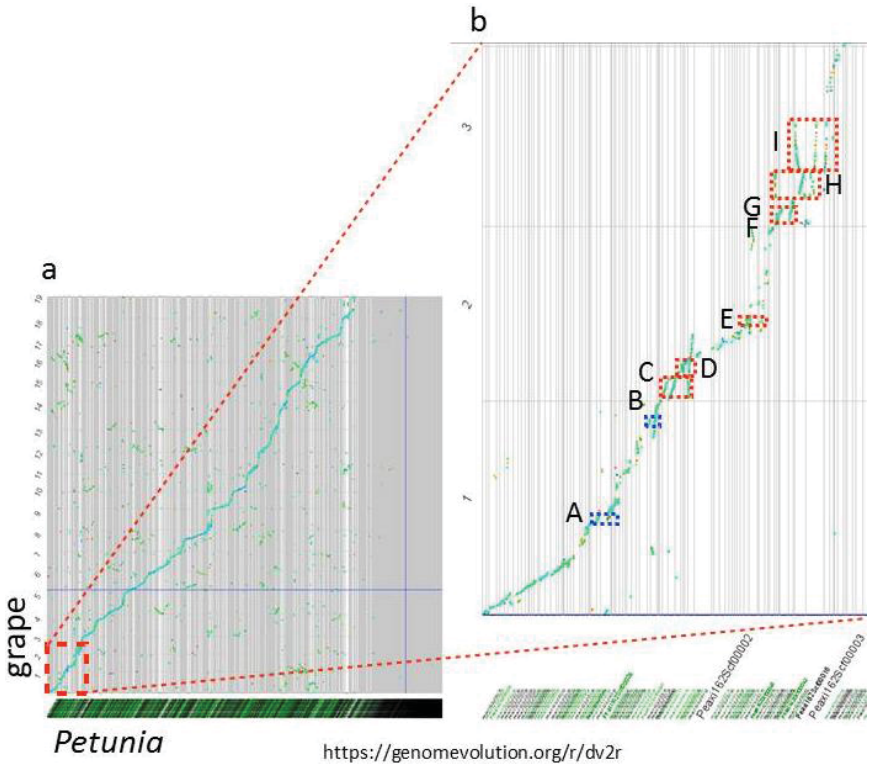


Figure 3: Evidence that the *Petunia* lineage had a subsequent hexaploidy event following its divergence with the grape lineage. Whole genome syntenic dot plot between grape and *Petunia* showing their entire genomes (a) or a zoomed in region of the dot plot (b). In the zoomed in region (b) multiple syntenic orthologous regions of *Petunia* are seen. These were given boxes to identify regions that appear to be triplicated (red) or duplicated (blue). Analysis may be regenerated: https://genomevolution.org/r/dv2r

Figure 4

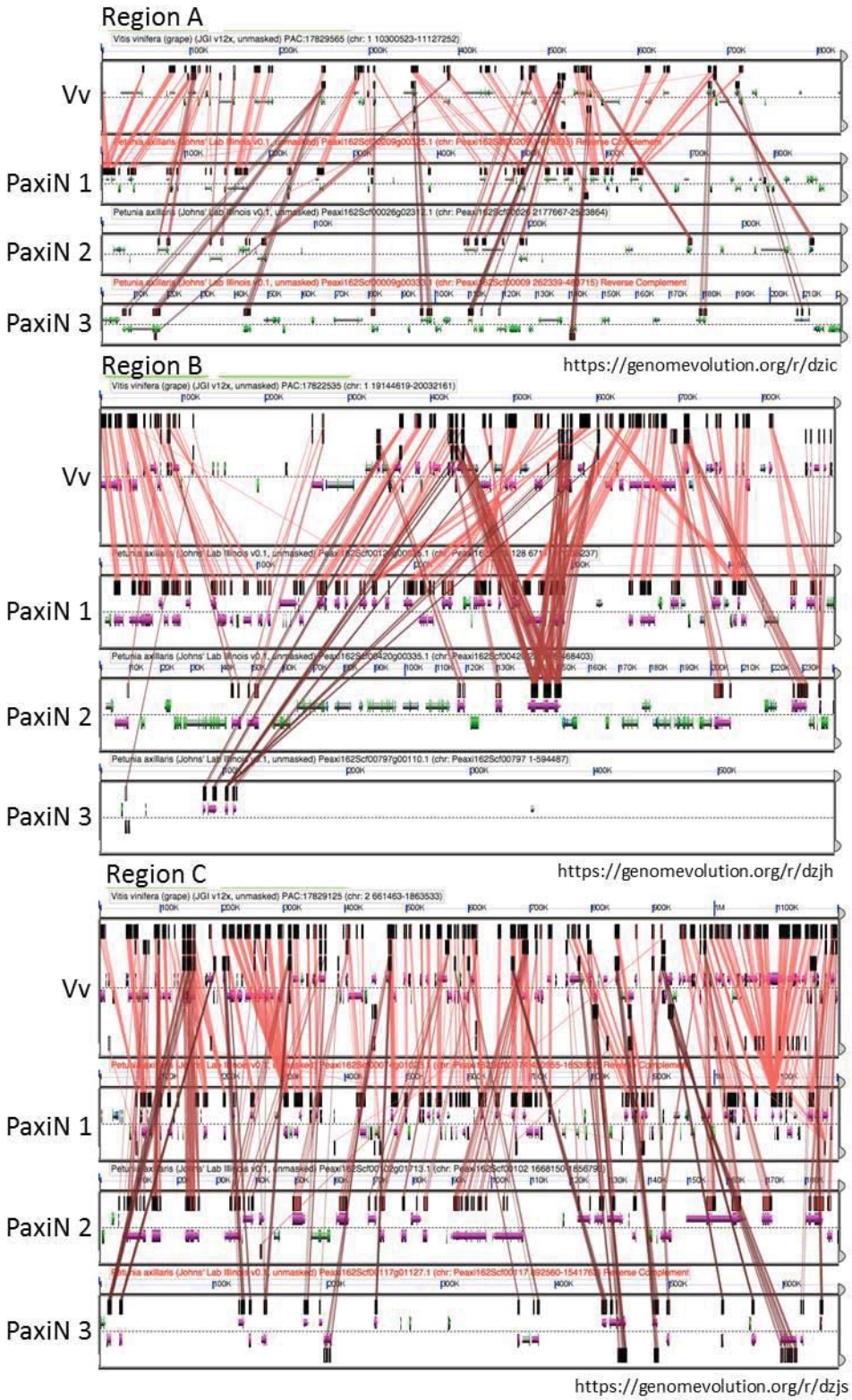
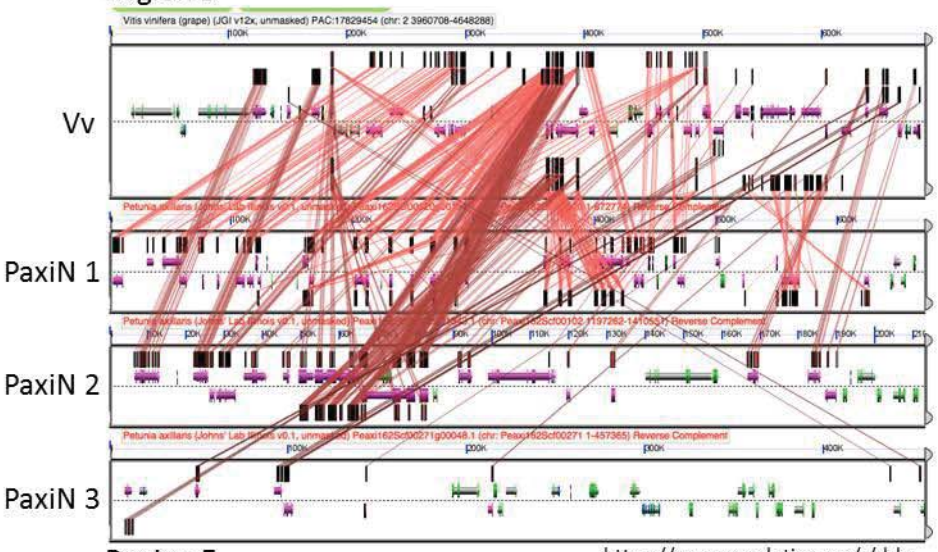

Region A
Vv
PaxiN 1
PaxiN 2
PaxiN 3
https://genomevolution.org/r/dzic
Region B
Vv
PaxiN 1
PaxiN 2
PaxiN 3
https://genomevolution.org/r/dzjh
Region C
Vv
PaxiN 1
PaxiN 2
PaxiN 3
https://genomevolution.org/r/dzjs

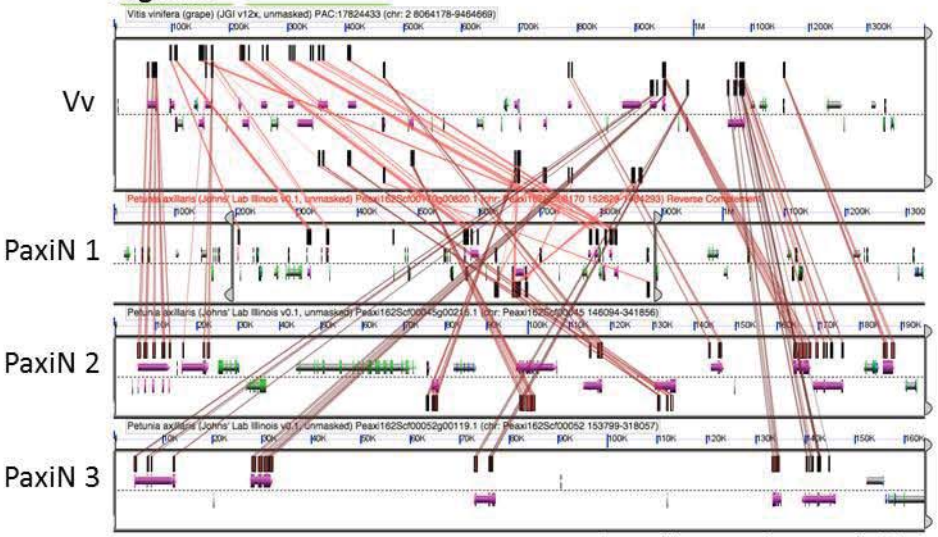
Figure 5

Region D



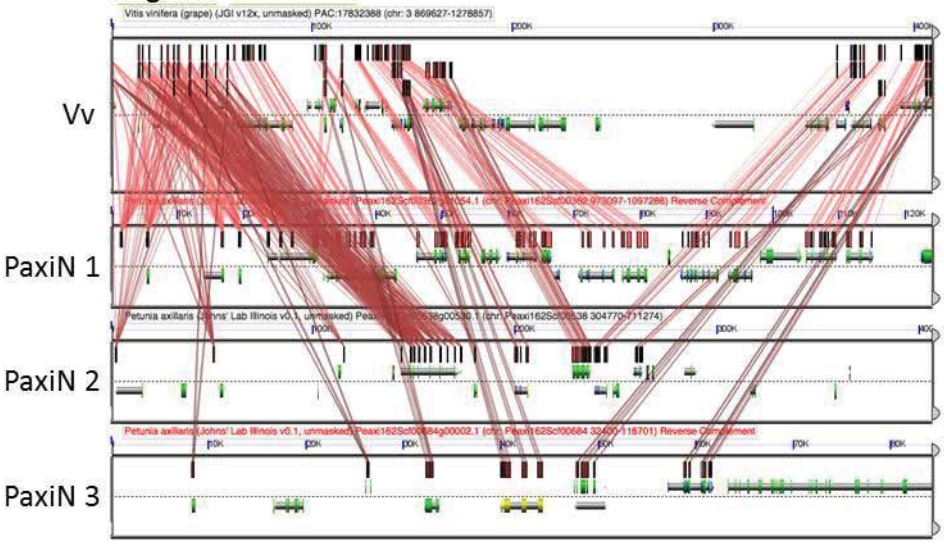
https://genomevolution.org/r/dzkg

Region E



https://genomevolution.org/r/dzkn

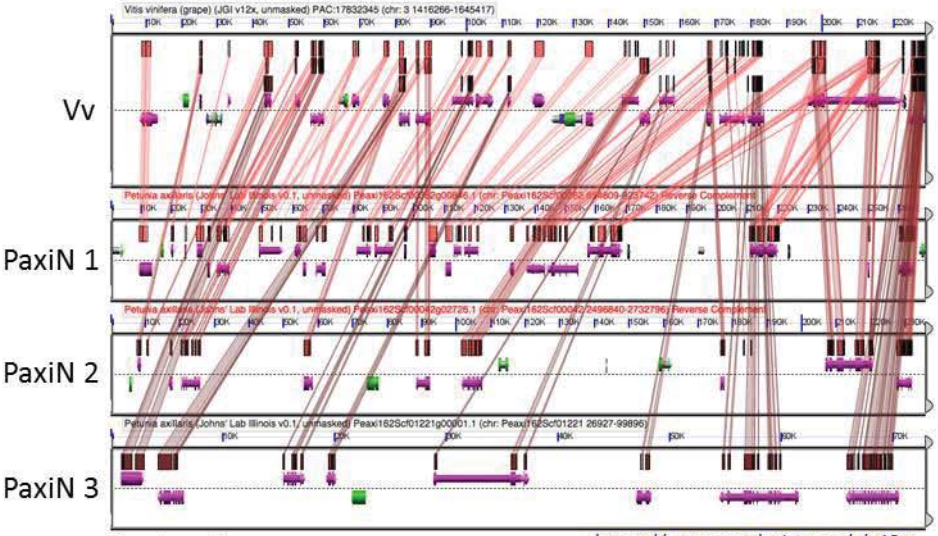
Region F



https://genomevolution.org/r/e19j

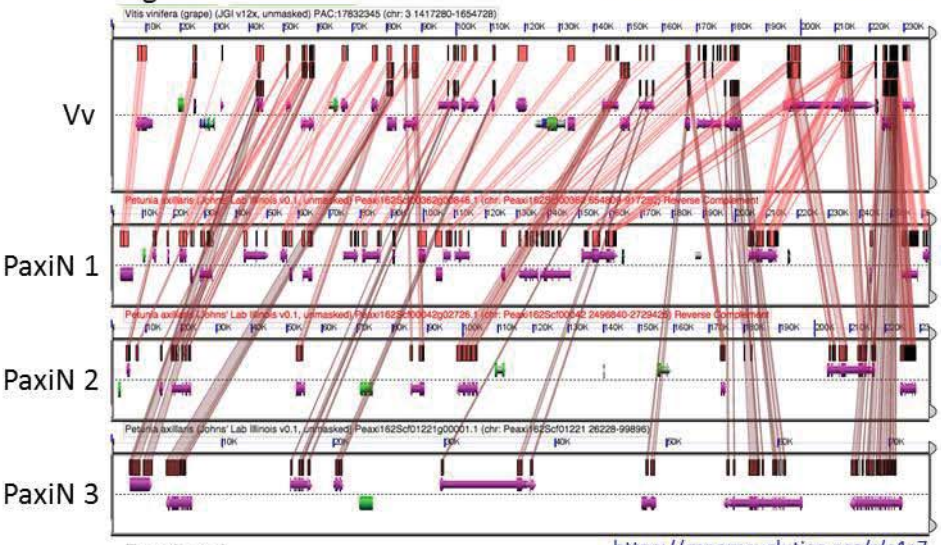
Figure 6

Region G



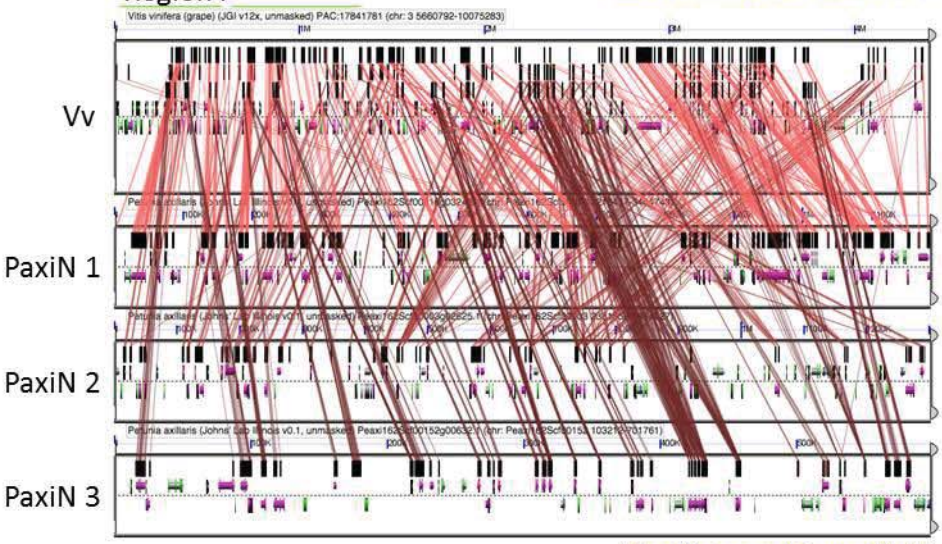
https://genomevolution.org/r/e19m

Region H



https://genomevolution.org/r/e1a7

Region I



https://genomevolution.org/r/e1bk

Figures 4-6: Evidence that *Petunia* is triplicated compared to grape. Microsynteny analysis of selected regions identified in Figure 3. Each figure identifies three syntenic orthologous regions of *Petunia axillaris N* (PaxiN) to one region of grape (*Vitis vinifera*, Vv). Each figure has a link to generate the analysis along with notes about the quality of the pattern of synteny. Each genomic region is represented by a horizontal panel with a dashed line separating the top and bottom strands of DNA. Gene models are represented as colored arrows located immediately above and below the dashed line. Regions of sequence similarity are shown as colored blocks with lines connecting them between genomic panels. Synteny is inferred as a colinear arrangement of homologous genes forming parallel connecting lines.

We then examined if the subsequent hexaploidy event we found in the *Petunia* genome corresponds to the *Solanum* hexaploidy event previously found by analyzing the tomato, potato and tobacco genomes (Tomato Genome Consortium, 2012).

We analyzed the synteny between tomato and *Petunia* genomes (Figure 7). The corresponding dot plot highlights three types of genomic regions colored according to the *Ks* values; the orthologous regions (in purple) and two different out-paralogous regions (in blue and green). Much of the two genomes were covered in one-to-one matching orthologous regions, which suggests that the paleopolyploidy level is similar between the two genomes. The out-paralogous (II), in green, represent regions that are originated by the old Eurosid hexaploidy gamma event as indicated by the "weak" peak in the *Ks* value histogram (Figure 7b). While the out-paralogs (I), in blue, are originated from the youngest *Solanum* hexaploidy event. This result provides strong evidence that *Petunia* shares the *Solanum* hexaploidy event. This is further confirmed by the similar syntenic patterns seen between both *Solanum* and *Petunia* to grape (Figure 8).

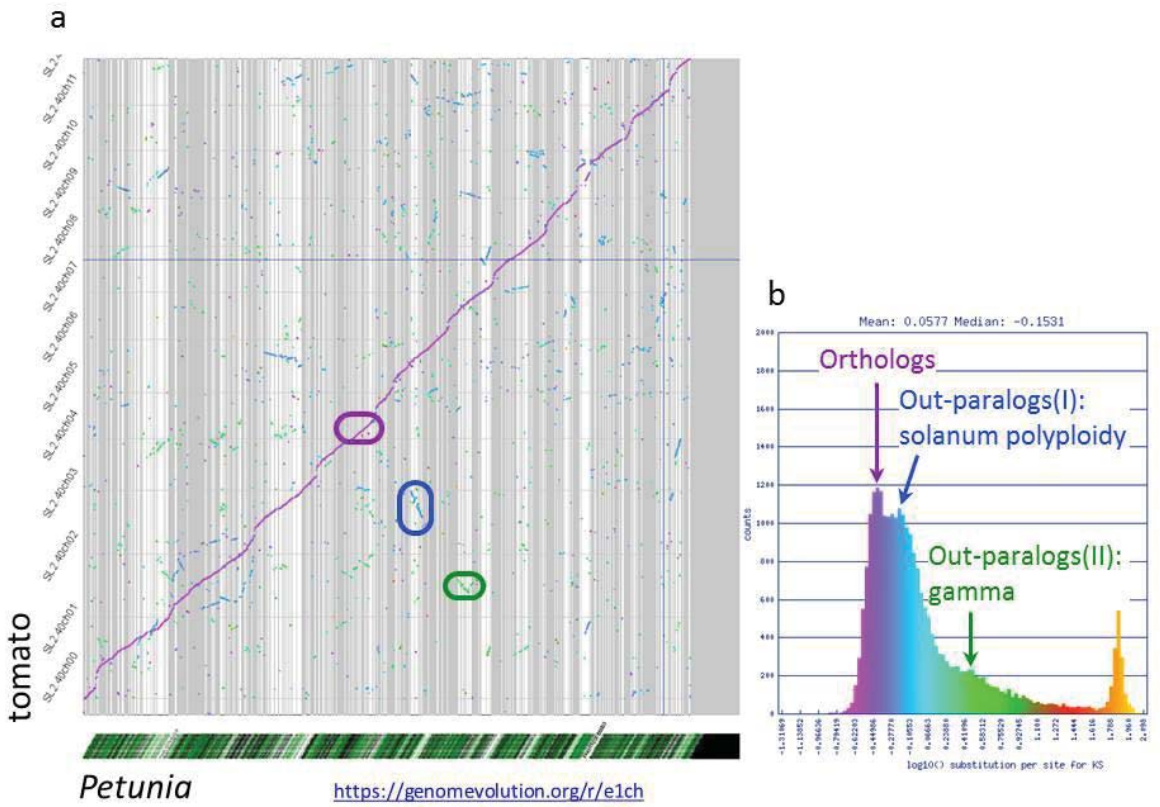


Figure 7: Evidence that *Petunia* and *Solanum* share a paleohexaploidy event. (a) Whole genome syntenic dot plot between tomato and *Petunia* with (b) histogram of log 10 transformed *Ks* values for syntenic gene pairs. Orthologous regions are colored purple; those derived from their shared *Solanum*-specific polyploidy event are colored blue; and those derived from the Eurosid/Eudicot paleohexaploidy (gamma) event are colored in green. Analysis may be regenerated: https://genomevolution.org/r/e1ch

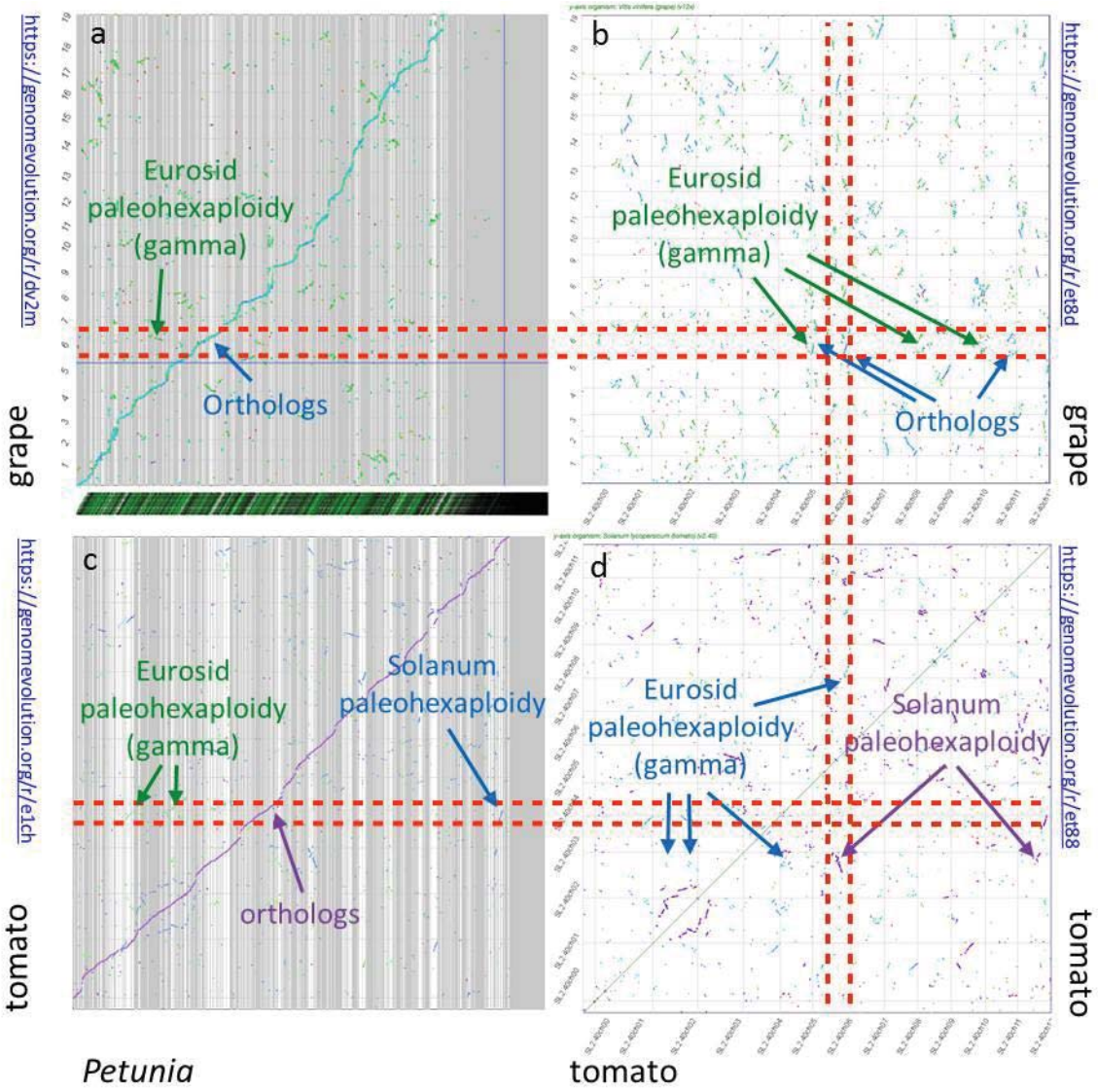


Figure 8: Evidence that *Petunia* shares the *Solanum* paleohexaploidy event. Whole genome syntenic dot plots between (a) grape and *Petunia*; (b) grape and tomato; (c) tomato and *Petunia*; (d) tomato and tomato. Syntenic regions are labeled in each dot plot as to their evolutionary origins and correlated patterns are highlighted with dashed red lines. Note, dashed red lines are not drawn vertically for *Petunia* across tomato and grape because the syntenic path ordering of *Petunia*'s contigs is not necessarily the same for those two genomes. Analyses may be regenerated: (a) https://genomevolution.org/r/dv2m; (b) https://genomevolution.org/r/et8d; (c) https://genomevolution.org/r/e1ch; (d) https://genomevolution.org/r/et88

An in-depth depiction of a set of syntenic regions

To go a step further we performed a detailed examination of one grape syntenic region, the region between bases 1,452,300 and 1,645,417 on chromosome 3 of grape (see Figure 6, region G, Table 1). This region provides evidence for the ancient triplication of the Solanaceae, followed by a later separation of the *Petunia* lineage from the *Solanum* lineage, and then by numerous independent gene loss and tandem duplication events. Here, a single grape region matches three distinct regions in both *Petunia axillaris* and *Solanum lycopersicum*. The region contains 22 grape genes, of which 20 have at least one match in a syntenic region of *Petunia* or tomato. All genes in all 7 syntenic segments are found in the same order and on the same strand relative to each other, allowing for

reverse complementation of the entire region as necessary. One region in each species is clearly dominant, which is the best match to the grape region. The dominant syntenic regions in *Petunia* (on scaffold Scf00362) and tomato (on chromosome 1) share 19 of the 20 matching genes, respectively; the one shared gene found in grape but not in either of the dominant regions is found on one of the other petunia syntenic regions.

The other four secondary syntenic regions share 6-9 genes with the grape region. The two secondary regions within each species show little similarity with each other. However, the syntenic regions on *Petunia* Scf00042 and tomato chromosome 10 appear to be derived from a common ancestor after the triplication event. They share 2 genes at the left end of the region not found in the other regions, as well as 7 other genes also found in the dominant regions. The only differences between the *Petunia* Scf00042 region and the tomato chromosome 10 region are a tandem duplication of a tomato gene, and one syntenic gene found in *Petunia* but not tomato. This similarity demonstrates that the triplication event predates the *Petunia-Solanum* split. The third pair of syntenic regions, found on *Petunia* Scf01221 and tomato chromosome 2, shows little evidence of shared common ancestry more recent than the triplication; at most one gene is shared by these regions that is not found in the other syntenic regions. These regions appear to either have been derived independently from the dominant region or subjected to enough fractionation to completely obscure their common ancestry.

The syntenic regions show evidence of many independent gene loss and tandem duplication events presumably since divergence from a common ancestor. Each of the six regions contains 1-14 genes with no homologues in the other regions. There are also 4 tandem duplications of shared genes (1 in *Petunia*, 3 in tomato) that are unique to a single syntenic region. In contrast, the tandem duplication of grape genes 17832342 and 17832343 must pre-date split between these species, as it is shared by one Petunia and two tomato syntenic regions. The length of the syntenic regions are also quite variable, ranging from 74,000 to 334,000 bp, highlighting different evolutionary trajectory among those regions following the shared event.

Another important feature of this region is the variation concerning which *Petunia* or tomato region contains the closest match to the syntenic grape gene in a tblastn search. For 10 of the 20 genes, the dominant syntenic region contains the best hit. However, for five *Petunia* genes and four tomato genes, the closest match is a gene not found in any syntenic region. The secondary regions have the best match for the remaining genes. The evolutionary forces that affect the subfunctionalization of duplicated genes have apparently affected individual genes in the *Petunia* and tomato genomes differently, and independently of the fractionation process.

Table 1. Description of grape syntenic region G, compared to *P. axillaris* and tomato.

		Matching genes						best tblastn hit to grape	
		petunia			tomato				
Grape gene	position on chr 3	Scf 00362	Scf 00042	Scf 01221	ch 01	ch 10	ch 02	petunia	tomato
--		--	g00274.1	--	--	g055230	--		

--		--	g02718.1	--	--	g055240	--		
PAC:17832347	1,459,600-1,461,579	g00940.1	--	g00001.1	g111270	--	g083820	other	other
PAC:17832346	1,468,355-1,469,527	g00091.1	g00273.1	--	g111280, g111300	g055250	g083800	Scf 00042	Chr 10
PAC:17832345	1,472,790-1,476,160	g00846.1	g02726.1	--	g111310	g055260	--	other	other
PAC:17832344	1,482,091-1,483,122	--	--	--	--	--	--		
PAC:17832343	1,484,329-1,485,994	g00852.1	--	--	g111320	--	g083790	Scf 00362	Chr 01
PAC:17832342	1,493,153-1,494,510	g00841.1	--	--	g111330	--	g083760	Scf 00362	Chr 01
PAC:17832341	1,497,503-1,500,305	g00845.1	g02623.1	g00031.1	g111340	g055340, g055370	--	Scf 01221	other
PAC:17832340	1,502,460-1,506,275	g00832.1	g02616.1	--	g111350	g055390	--	Scf 00362	Chr 10
PAC:17832339	1,512,511-1,518,020	g00840.1	--	g00026.1	g111360	--	--	Scf 00362	other
PAC:17832338	1,519,145-1,523,441	g00839.1	--	--	g111370	--	--	Scf 00362	Chr 01
PAC:17832337	1,527,977-1,528,708	g00838.1	--	g00027.1	g111380	--	--	Scf 01221	Chr 01
PAC:17832336	1,534,987-1,537,909	g00088.1	--	--	g111400	--	--	Scf 00362	Chr 01
PAC:17832335	1,543,989-1,546,820	--	--	--	--	--	--		Chr 01
PAC:17832334	1,549,919-1,551,865	g00082.1	--	--	g111430	--	--	Scf 00362	Chr 01
PAC:17832333	1,560,187-1,564,616	g00848.1	--	--	g111440	--	g083690	other	Chr 01
PAC:17832332	1,565,251-1,567,724	--	g02619.1	--	--	--	--	other	--
PAC:17832331	1,570,696-1,574,690	g00836.1	--	g00028.1	g111450, g111460	--	--	Scf 00362	Chr 01
PAC:17832330	1,584,232-1,585,073	g00831.1	g02519.1	g00042.1	g111500	g055410	--	Scf 00362	Chr 10
PAC:17832329	1,587,828-1,594,273	g00830.1	--	--	g111510	--	g083620, g083630	other	Chr 02
PAC:17832328	1,595,399-1,600,077	g00720.1	--	g00043.1	g111520	--	--	Scf 01221	Chr 01
PAC:17832327	1,613,566-1,639,214	g00725.1, g00716.1	g02521.1	--	g111530	g055450	--	Scf 00362	Chr 01
PAC:17832325	1,641,239-1,645,417	g00645.1	g02513.1	g00034.1	g111540	g055470	--	Scf 00042	Chr 10

The tomato genome is more fractionated than the *Petunia* genome.

We finally investigated gene fractionation in *Petunia* in comparison with tomato. Gene fractionation corresponds to the loss of duplicate genes after whole genome duplication (Langham et al., 2004; Thomas et al., 2006). We compared gene fractionation in Region H in detail (see Figure 6) by analyzing microsynteny between *Petunia* and tomato genomes with grape as reference (Figure 9 and Figure 10). While we selected H region for illustratrative purposes, patterns of other regions offer similar trends of fractionation.

As expected, due to the *Solanum*-hexaploidy event, we can see that both the *Petunia* and tomato genomes are triplicated in comparison with the grape genome (Figure 9). It appears that PaxiN 3 and Sl 3 are under-fractionated or dominant (i.e., retain more genes than other regions) while PaxiN 1, PaxiN 2, Sl 1 and Sl 2 are over-fractionated. Note that for the under-fractionated regions PaxiN 3 and Sl 3, gene fractionation is identical. To better identify regions that are differentially fractionated in the over-fractionated regions, synteny lines between genomic regions of *Petunia* and tomato vs. grape in PaxiN 3 and Sl 3 are not shown. Gene fractionation is significantly less complete in *Petunia* in comparison with tomato. In addition to shared fractionation events, we also observed independent gene fractionation in the two lineages with more fractionation events in tomato. Of the 13 regions differently fractionated between *Petunia* and tomato, 9 regions are independently retained in *Petunia* while only 3 regions are independently retained in tomato. In addition, *Petunia* retains more genes in over-fractionated regions than tomato (compare PaxiN 1 and PaxiN 2 with Sl 1 and Sl 2). These results are also validated by comparing the same genomic regions but with a different order (Figure 10). This representation highlights synteny between orthologs and allows us to observe independent fractionation in *Petunia* and tomato as well as shared fractionation events between these two lineages.

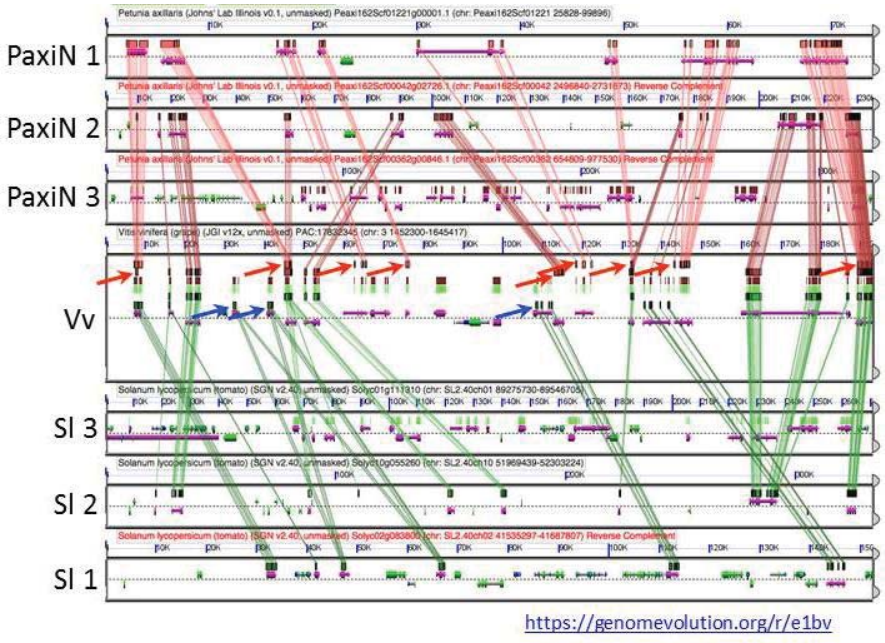


Figure 9: Evidence that divergence of tomato and *Petunia* lineages happened prior to the completion of fractionation. Microsynteny analysis of one region of grape (Vv) to three syntenic orthologous regions of *Petunia* (PaxiN) and tomato (Sl). Description of how to read this figure is described in summary for figures 4-6. Red arrows indicate genes retained in *Petunia* and fractionated in tomato. Blue arrows indicate genes retained in tomato and fractionated in *Petunia*. Analysis may be regenerated: https://genomevolution.org/r/e1bv

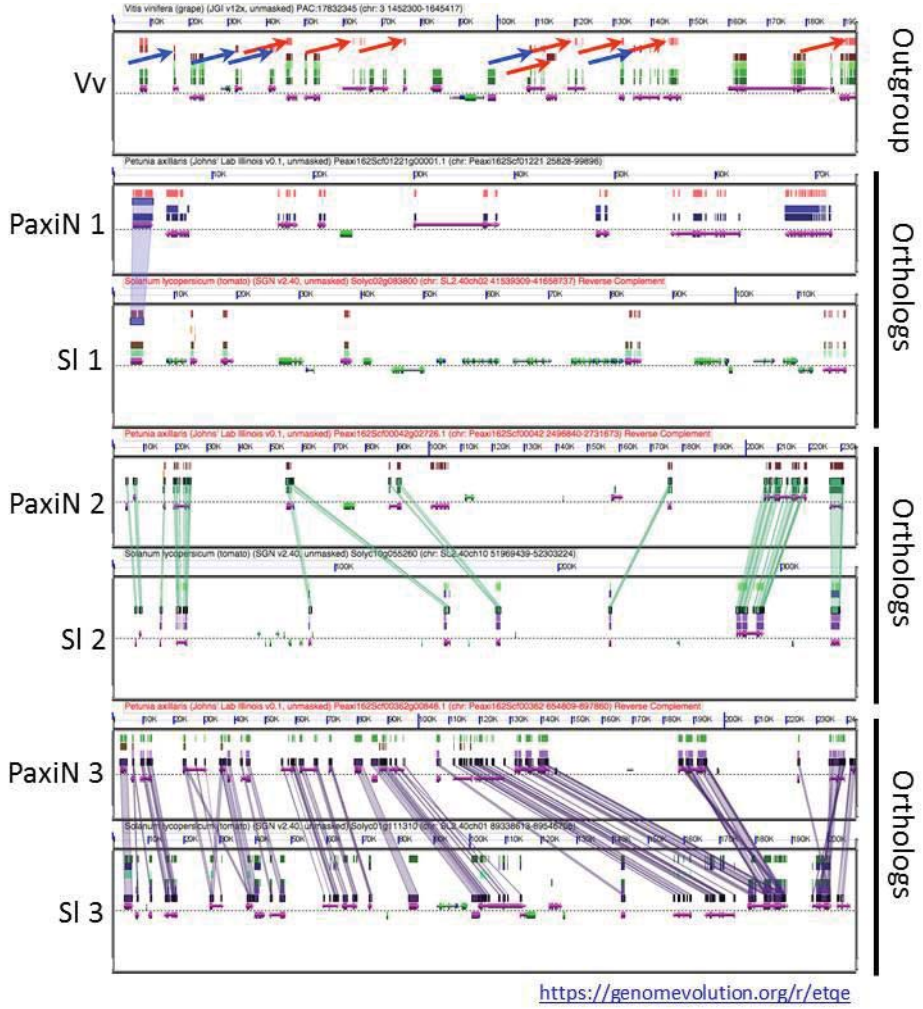


Figure 10: Fractionation partially occurred independently in each lineage, with more fractionation events in the tomato lineage. Microsynteny analysis description is similar to figure 9, but with different order to the regions. Red arrows indicate genes retained in *Petunia* and fractionated in tomato. Blue arrows indicate genes retained in tomato and fractionated in *Petunia*. Note that there is only a single syntenic gene set shared between orthologous regions PaxiN 1 and Sl 1. Analysis may be regenerated: https://genomevolution.org/r/etge

We finally included *Mimulus guttatus* in the analysis of microsynteny between *Petunia* and grape (Figure 11). *Mimulus guttatus* was shown to lack the *Solanum* paleohexaploidy event and to have an independent paleotetraploidy event (Ibarra-Laclette et al., 2013). Our analysis is consistent with this previous study as we could observe an independent fractionation in *Mimulus* in comparison with *Petunia* and tomato.

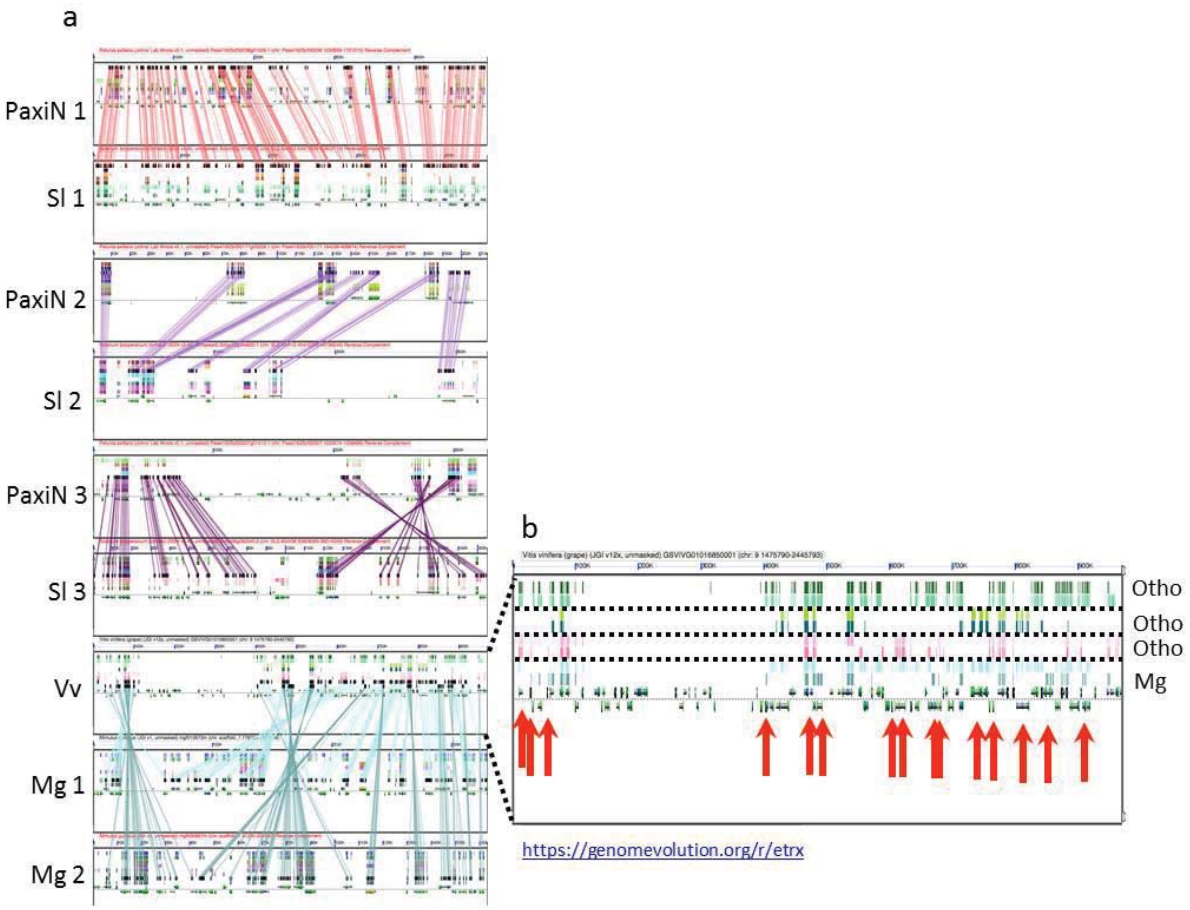


Figure 11: Independent gene fractionation in *Mimulus* in comparison with *Petunia* and tomato.
(a) Microsynteny analysis of *Petunia* (PaxiN) and tomato (Sl), *Mimulus guttatus* (Mg) and grape (Vv). (b) Close up of microsynteny panel for grape with red arrows marking independent fractionation of the *Mimulus* lineage and tomato/*Petunia* lineages. Analysis may be regenerated: https://genomevolution.org/r/etrx

DISCUSSION

Here we present the first analysis of paleopolyploidy for *Petunia*, which represents the sister-group to the larger and more diverse x=12 crown-group clade of the Solanaceae family (Särkinen et al., 2013) (Figure 1). Our study confirms the until now ambiguous Solanaceae paleohexaploidy event initially inferred in the analysis of the tomato genome. Further, we demonstrate that the *Petunia* lineage has experienced (at least) two rounds of paleopolyploidization, the older gamma hexaploidy event (Figure 2), which is shared with other Eudicots (Jaillon et al., 2007), and the more recent Solanaceae paleohexaploidy event which is shared with tomato and other x=12 species (Figures 3-8) (Tomato Genome Consortium, 2012). We have shown that the process of gene fractionation that facilitates the return to a diploid state (also known as diploidization), occurred, in part, independently in *Petunia* and tomato despite the shared Solanaceae paleohexaploidy event, similar to what has been observed in *Saccharomyces* yeasts but until now not yet described in flowering plants.

Validation of the Solanaceae paleohexaploidy event in the genome of tomato.

Genome collinearity analysis between tomato and grape is ambiguous in establishing a complete independent genome triplication in the Solanaceae or, potentially evidence for segmental duplication (i.e. some regions duplicated and others triplicated) (Ibarra-Laclette et al., 2013; Figure 8). Our study notably showed that the *Petunia* genome is less fractionated than the tomato genome (Figure 9). This would explain the difficulties in identifying a clear 3:1 orthologous syntenic relationship between tomato and grape by performing genome colinearity analysis (Ibarra-Laclette et al., 2013; Figure 8). In view of our results, we could hypothesize that an independent genome triplication occurred in the Solanaceae family but the high degree of gene fractionation in tomato makes it difficult to identify the three orthologous regions. A similarly high degree of gene fractionation could also be explain the interpretation of paleopolyploidy in the potato genome. Both tomato and potato genomes are relatively complete, suggesting that the lack of triplicated regions is not due to incomplete genome assembly. Indeed, genomic analysis showed that one genomic region of grape is (only) syntenic to two regions in the potato genome, leading to an inaccurate inference of only paleotetraploidy (The Potato Genome Consortium, 2011).

Gene fractionation occurred independently in *Petunia* and tomato genomes following their divergence despite the shared paleopolyploidy history

Our analysis is the first detailed description of an independent gene fractionation from a common polyploidy event in plants. While some fractionation of gene content probably occurred prior to the divergence of these lineages, as indicated by the shared gene fractionation events in *Petunia* and tomato genomes, our results demonstrate that gene fractionation also continued independently in these two lineages (Figure 9 and Figure 10) following their divergence, even though they shared a paleohexaploidy event (Figure 2 and Figure 3).

First, we observed in *Petunia* and tomato the presence of two classes of genomic regions with distinct levels of gene fractionation, the over-fractionated (i.e., more genes are lost) and the under-fractionated (i.e. fewer genes are lost) regions (Figure 9, Figure 10 and Figure 11), supporting a biased fractionation process following polyploidy. In each microsynteny analyses, one genomic region is still under-fractionated (PaxiN 3 and Sl 3 in Figures 9 and 10, PaxiN 1 and Sl 1 in Figure 11) while the two other are over-fractionated. This indicates that 3 sub-genomes, with contrasting gene contents, coexist in *Petunia* and tomato genomes. The same situation is found in other

paleopolyploids such as in *Brassica rapa* (Wang et al., 2011; Tang et al., 2012), a diploid species with three subgenomes originating from a whole genome triplication (Wang et al., 2011). A two-step theory implicating a differential subgenome evolution was proposed to explain the genome triplication event in *B. rapa* and also more broadly to explain fractionation after a paleohexaploidy event (Lyons et al., 2008). The first step of this model involves the formation of a tetraploid with two subgenomes. The two new subgenomes then experience loss of duplicated genes resulting in two fractionated subgenomes. The second step consists of the formation of the new tetraploid between the previous fractionated diploid genome and a new diploid genome, which experienced another round of gene fractionation. The final diploid genome will thus contain three subgenomes, two subgenomes that experienced two rounds of fractionation (and are thus over-fractionated plus one subgenome with only one round of fractionation (under-fractionated).

Second, *Petunia* genome is less fractionated than the tomato genome (Figure 9). Indeed, Petunia retains more genes in over-fractionated regions than tomato while difference in the dominant region is less profound. Analysis of gene fractionation between tomato and potato as well as between *Petunia* and potato would be helpful to decipher whether gene fractionation is similar in tomato and in potato and to confirm if fractionation in Petunia and potato is also independent. This observation is similar to what happens in yeast where early-branching clades, in our study *Petunia*, show a different pattern of fractionation than latter branching species, represented here by tomato and eventually by potato (Scannell et al., 2006). It would be interesting to compare gene fractionation between *Petunia* and other “later branching” Solanaceae species such as tobacco in order to confirm this similarity. *Petunia* is well situated to be a better genome comparator for Solanaceae genomes than the tomato genome, due to its low degree of gene fractionation. Nevertheless, some aspects of the *Petunia* genome structure and evolution still need to be elucidated in order to represent an ideal reference for genomics, specifically, a higher quality assembly into pseudomolecules.

METHODS

Whole genome collinearity analysis

Whole genome dot plots were performed using SynMap in the comparative genomics platform, CoGe (Lyons et al., 2008). Each analysis generates a tiny-url that links back to SynMap configured to regenerated the analysis exactly as shown. URLs include embedded options for configuring SynMap and for these analyses, the syntenic path assembly and coloration of syntenic gene pairs by synonymous mutation rate (Ks; CodeML; http://abacus.gene.ucl.ac.uk/software/paml.html) were extensively used. For details on how to use SynMap, please see Tang and Lyons (2012).

Microsynteny analysis

Microsynteny analysis was performed using GEvo in the comparative genomics platform, CoGe (Lyons and Freeling, 2008). As with SynMap, each GEvo analysis generates a tiny-url to regenerate the analysis exactly as configured; these links are provided in the figure legends.

ACKNOWLEDGEMENTS

LG is supported by a H2020 Marie Curie fellowship (H2020-MSCA-IF-2014). EL and HT are supported by USDA 2013-00984 and NSF IOS – 1339156. HT is supported by the "100 Talent Plan" award by the Fujian provincial government. CoGe is part of the Powered by iPlant program (NSF DBI – 1265383). MES was funded by a NWO-VIDI grant.

REFERENCES

Barker, M.S., Vogel, H., and Schranz, M.E. (2009). Paleopolyploidy in the Brassicales: analyses of the Cleome transcriptome elucidate the history of genome duplications in Arabidopsis and other Brassicales. Genome biology and evolution **1,** 391-399.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Correa, M., Da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.C., Fan, G., Renault, V., Bayer, P.E., Golicz, A.A., Manoli, S., Lee, T.H., Thi, V.H., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C.H., Wang, X., Canaguier, A., Chauveau, A., Berard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J.M., Adams, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C., and Wincker, P. (2014). Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science **345,** 950-953.

Cheng, S., van den Bergh, E., Zeng, P., Zhong, X., Xu, J., Liu, X., Hofberger, J., de Bruijn, S., Bhide, A.S., Kuelahoglu, C., Bian, C., Chen, J., Fan, G., Kaufmann, K., Hall, J.C., Becker, A., Brautigam, A., Weber, A.P., Shi, C., Zheng, Z., Li, W., Lv, M., Tao, Y., Wang, J., Zou, H., Quan, Z., Hibberd, J.M., Zhang, G., Zhu, X.G., Xu, X., and Schranz, M.E. (2013). The Tarenaya hassleriana genome provides insight into reproductive trait and genome evolution of crucifers. The Plant cell **25,** 2813-2830.

Garsmeur, O., Schnable, J.C., Almeida, A., Jourda, C., D'Hont, A., and Freeling, M. (2014). Two evolutionarily distinct classes of paleopolyploidy. Molecular biology and evolution **31,** 448-454.

Ibarra-Laclette, E., Lyons, E., Hernandez-Guzman, G., Perez-Torres, C.A., Carretero-Paulet, L., Chang, T.H., Lan, T., Welch, A.J., Juarez, M.J., Simpson, J., Fernandez-Cortes, A., Arteaga-Vazquez, M., Gongora-Castillo, E., Acevedo-Hernandez, G., Schuster, S.C., Himmelbauer, H., Minoche, A.E., Xu, S., Lynch, M., Oropeza-Aburto, A., Cervantes-Perez, S.A., de Jesus Ortega-Estrada, M., Cervantes-Luevano, J.I., Michael, T.P., Mockler, T., Bryant, D., Herrera-Estrella, A., Albert, V.A., and Herrera-Estrella, L. (2013). Architecture and evolution of a minute plant genome. Nature **498,** 94-98.

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pe, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.F., Weissenbach, J., Quetier, F., Wincker, P., and French-Italian Public Consortium for Grapevine Genome, C. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449,** 463-467.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J., and dePamphilis, C.W. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature **473,** 97-100.

Langham, R.J., Walsh, J., Dunn, M., Ko, C., Goff, S.A., and Freeling, M. (2004). Genomic duplication, fractionation and the origin of regulatory novelty. Genetics **166,** 935-945.

Lyons, E., and Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. The Plant journal : for cell and molecular biology **53,** 661-673.

Lyons, E., Freeling, M., Kustu, S., and Inwood, W. (2011). Using genomic sequencing for classical genetics in E. coli K12. PloS one **6,** e16717.

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., Wang, X., Bowers, J., Paterson, A., Lisch, D., and Freeling, M. (2008). Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant physiology **148,** 1772-1781.

Sarkinen, T., Bohs, L., Olmstead, R.G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC evolutionary biology **13,** 214.

Scannell, D.R., Byrne, K.P., Gordon, J.L., Wong, S., and Wolfe, K.H. (2006). Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature **440,** 341-345.

Schnable, J.C., Springer, N.M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences of the United States of America **108,** 4069-4074.

Schranz, M.E., Mohammadin, S., and Edger, P.P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Current opinion in plant biology **15,** 147-153.

Soltis, D.E., Bell, C.D., Kim, S., and Soltis, P.S. (2008). Origin and early evolution of angiosperms. Annals of the New York Academy of Sciences **1133,** 3-25.

Tang, H., and Lyons, E. (2012). Unleashing the genome of brassica rapa. Frontiers in plant science **3,** 172.

Tang, H., Woodhouse, M.R., Cheng, F., Schnable, J.C., Pedersen, B.S., Conant, G., Wang, X., Freeling, M., and Pires, J.C. (2012). Altered patterns of fractionation and exon deletions in Brassica rapa support a two-step model of paleohexaploidy. Genetics **190,** 1563-1574.

Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B., and Harmon, L.J. (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. The New phytologist **207,** 454-467.

Thomas, B.C., Pedersen, B., and Freeling, M. (2006). Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome research **16,** 934-946.

Tomato Genome, C. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485,** 635-641.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F., Huang, S., Li, X., Hua, W., Wang, J., Wang, X., Freeling, M., Pires, J.C., Paterson, A.H., Chalhoub, B., Wang, B., Hayward, A., Sharpe, A.G., Park, B.S., Weisshaar, B., Liu, B., Li, B.,

Liu, B., Tong, C., Song, C., Duran, C., Peng, C., Geng, C., Koh, C., Lin, C., Edwards, D., Mu, D., Shen, D., Soumpourou, E., Li, F., Fraser, F., Conant, G., Lassalle, G., King, G.J., Bonnema, G., Tang, H., Wang, H., Belcram, H., Zhou, H., Hirakawa, H., Abe, H., Guo, H., Wang, H., Jin, H., Parkin, I.A., Batley, J., Kim, J.S., Just, J., Li, J., Xu, J., Deng, J., Kim, J.A., Li, J., Yu, J., Meng, J., Wang, J., Min, J., Poulain, J., Wang, J., Hatakeyama, K., Wu, K., Wang, L., Fang, L., Trick, M., Links, M.G., Zhao, M., Jin, M., Ramchiary, N., Drou, N., Berkman, P.J., Cai, Q., Huang, Q., Li, R., Tabata, S., Cheng, S., Zhang, S., Zhang, S., Huang, S., Sato, S., Sun, S., Kwon, S.J., Choi, S.R., Lee, T.H., Fan, W., Zhao, X., Tan, X., Xu, X., Wang, Y., Qiu, Y., Yin, Y., Li, Y., Du, Y., Liao, Y., Lim, Y., Narusaka, Y., Wang, Y., Wang, Z., Li, Z., Wang, Z., Xiong, Z., Zhang, Z., and Brassica rapa Genome Sequencing Project, C. (2011). The genome of the mesopolyploid crop species Brassica rapa. Nature genetics **43,** 1035-1039.

Supplementary Note 6

Analysis of the genomic origin of *Petunia hybrida*

Aureliano Bombarely[1*], Gonzalo H. Villarino[2], Mattijs Bliek[3], Neil S. Mattson[2], Francesca M. Quattrocchio[3], Ronald Koes[3]

[1] Department of Horticulture, Virginia Tech, Blacksburg, VA, 24061, USA.
[2] Horticulture Section, School of Integrative Plant Sciences, Cornell University, Ithaca, NY, 14850
[3] Department of Plant Development and (Epi)Genetics, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, Amsterdam, The Netherlands

[*]Corresponding author, email: aurebg@vt.edu

ABSTRACT
Petunia hybrida, commonly named as garden petunia has a complex origin. It was originated by the hybridization of white flowers and purple flowers petunias, probably *P. axillaris* and one of the taxa of the *P. integrifolia* complex. Its history is probably populated by uncountable events of backcrossing with the same species that originally produced the first hybrids. In this analysis, we have used transcriptomic data from three different accessions (Mitchell/W115, R27 and R143) to elucidate the gene contribution of *P. axillaris* N and *P. inflata* S6 to the garden petunias. Our results show that these garden petunia accessions are highly homozygous with major contributions of *P. axillaris*, with minor contributions of *P. inflata*. We have also found events of partial gene conversions for more than one thousand genes between both ancestral genomes. For these genes, a part has a *P. axillaris* type origin and another part *P. inflata* type origin.

INTRODUCTION
The garden petunia (*Petunia hybrida*) has a complex origin. It was originated by the cross of two native species, one with white flowers, *P. axillaris*, and another one with purple flowers. Atkins reported in 1834 a cross between *P. axillaris* and *P. integrifolia* that it could be the origin of the cultivated petunia. Nevertheless the popularity of this species and its quickly appearance in the English gardens probably encouraged the production of other hybrids using different parents and the frequent backcrosses making difficult to describe the complexity of this species (Griesbach, 2006). *P. integrifolia* is a complex group divided in five taxa with the same flower color and similar flower morphologies with differences in habitats and vegetative structures. Based in the historical records and phylogenetic analysis, three possible taxa could be the origin of the garden petunia: *P. integrifolia ssp. integrifolia, P. inflata* and *P. interior* (Ando et al., 2005; Griesbach, 2006; Segatto et al., 2014). Although the domestication history of the modern garden petunia has not been completely elucidated is clear that involve at least one event of hybridization between these species. Hybridization is a common process in the plant domestication. The cross between different species is the fastest way to develop new phenotypes and to improve some of the traits of the cultivated species through heterosis (DeHaan and Van Tassel, 2014), but also to avoid the *P. integrifolia* self incompatibility mechanism (Gerats and Strommer, 2009). Nevertheless, although the hybrid origin of the garden petunia, its history probably is full of backcrossing with other species making it more similar to different introgression lines rather than a hybrid per-se. An example is the petunia hybrida line Mitchell. It was originated by the somatic hybridization of *P. axillaris* and *P. hybrida* cultivar “Rose of Heaven” and then inbreeded for several generations (Gerats and Vandenbussche, 2005).
The goal of this study is to elucidate the gene contribution of *P. axillaris* and *P. inflata* in three *P. hybrida* accessions (Mitchell/W115, R27 and R143) using transcriptomic data (Figure 1).

Figure 1: Flowers for the petunia accessions used in this analysis. A- *P. axillaris* N; B- *P. inflata* S6; C- *P. hybrida* R27; D- *P. hybrida* R143 and E- *P. hybrida* Mitchell/W115. White rule is scaled to 1 cm per mark (Figure supplied by R. Koes).

RESULTS

Origin of the Petunia hybrida Gene Space

Petunia hybrida, or garden petunia, is a hybrid between the white flower petunia, *P. axillaris* and the purple flower petunia, probably *P. inflata* (Gerats and Strommer, 2009). RNASeq reads from three different *petunia hybrida* accessions (Mitchell/W115, R27, R43) were mapped to the *P. axillaris* N draft genome and compared with the *P. inflata* S6 polymorphisms to elucidate the contribution of each of the progenitors to the *P. hybrida* transcribed gene space. Genome coverage for the RNASeq reads and SNP contributions are summarized in the Table 1.

The percentage of contributions ranges from 49.8% (R27) to 62.1% (Mitchell) for *P. axillaris* N, from 4.7% (R27) to 5.4 (Mitchell) for *P. inflata* S6 and from 15.4% (Mitchell) to 16.0% (R143) for an unidentified petunia not recognized as *P. axillaris* N or *P. inflata* S6 that could indicate a contribution of different accessions from these species, or even from different species. As it was expected most of the SNPs are synonymous polymorphisms, although non-synonymous polymorphisms and frame shifts were detected indicating a possible source functional variation between accessions and its ancestors (Table 2).

Species / Accession	*P. hybrida* var. Mitchell	*P. hybrida* var. R27	*P. hybrida* var. R143
Genome coverage (Mb)[1]	77.52	70.41	36.90
Total SNPs[2]	1,464,375	827,349	1,721,775
% P. axillaris SNPs	62.1	49.8	58.9
% P. inflata SNPs	5.4	4.7	4.8
% No *P. axillaris* or *P. inflata* SNPs	15.4	15.9	16.0
% Heterozygous SNPs (Pa/Pi)	0.6	0.6	0.5
% Heterozygous SNPs (Pa=Pi/X)[3]	16.5	29.0	19.7
INDELS	34,378	92,466	43,242

Table 1: Summary of the polymorphism analysis of the *P. hybrida* compared with *P. axillaris* and *P. inflata*. [1]Genome coverage is estimated with a 5 minimum read depth per position. [2]Single Nucleotide Polymorphism (SNPs) using *P. axillaris* as reference and with a minimum read depth of 5 and a minimum mapping quality of 20. [3] Percentage of SNPs where *P. axillaris* and *P. inflata* share the same nucleotide and one of the *P. hybrida* alleles is the same as these two species (Pa=Pi) and the other one is different (X).

Species / Accession	*P. hybrida* var. Mitchell	*P. hybrida* var. R27	*P. hybrida* var. R143
Codon INDELs	876	1,856	1,074
Frame shift	3,866	7,768	4,235
Non Synonymous Coding	43,231	48,732	31,687
Start/Stop Changes	1,364	1,936	1,095
Synonymous Coding	53,133	57,670	39,242
3' and 5' UTR	22,863	21,542	10,504

Table 2: Summary of the polymorphism effects for different *P. hybrida* accessions compared with *P. axillaris.*

To have a better perspective of the origin of each gene, the different gene models were assigned to 1- *P. axillaris*, 2- *P. inflata*, 3- *P. axillaris x P. inflata* (hybrid), or 4- *P. axillaris + P. inflata* (gene conversions where a part of the gene is *P. axillaris* and another part is *P. inflata*) groups based in the polymorphisms when they were compared with *P. axillaris* and *P. inflata.* The polymorphisms not identified as *P. axillaris* N or *P. inflata* S6 were considered as neutral data for purposes of origin assignment. Results are summarized in the Table 3.

Species / Accession	*P. hybrida* var. Mitchell	*P. hybrida* var. R27	*P. hybrida* var. R143
P. axillaris	15,514	15,327	15,740
P. inflata	587	683	469

P. axillaris x P. inflata	12	27	16
P. axillaris + P. inflata	1,722	1,802	1,636
Unclear	1,926	1,924	1,902

Table 3: *Petunia hybrida* gene assignment to *P. axillaris* (all the polymorphism are the same than *P. axillaris*), *P. inflata* (all the polymorphism are the same than *P. inflata*), *P. axillaris x P. inflata* (heterozygous polymorphisms, two alleles: *P. axillaris* and *P. inflata*) and *P. axillaris + P. inflata* (some part of the gene can be assigned to *P. axillaris* and the other part of the gene can be assigned to *P. inflata*).

These results indicate that the origin of most of these petunia accessions is the same: *P. axillaris*, with probably an undetermined low number of hybridization events with *P. inflata* and probably other petunia species not included in this analysis. The number of *P. inflata* identified genes in these cultivars range from 469 (R143) to 687 (R143), sharing 416 of those. The three lines used, Mitchel (also known as W115), R27 and R143 have no traceable direct relations. The Mitchell cultivar was originated by the somatic hybridization of *P. axillaris* and *P. hybrida* cultivar "Rose of Heaven", and then developed as a double haploid line for more than 20 generations (Mitchell et al. 1980, Gerats and Vandenbussche, 2005) so it is expected that a high percentage of the genes should have *P. axillaris* origin. R27 was established around 1960 by inbreeding (> 40 generations) from a cultivar "Roter Vogel" of a German breeder (R.E. Koes personal communication). R143 originates by inbreeding (>10 generations) of lines that had been inbred around 1960 from cultivars of American "Fire chief" and "Admiral" and Dutch breeders "Peach satin" and "Gypsy" (R.E. Koes personal communication).

Additionally we have found evidences that could indicate gene conversions between the two ancestor genes, *P. axillaris* and *P. inflata* where different parts of the *P. hybrida* gene come from different origins (Figure 2). Gene conversions and homoeologous exchanges have been documented in polyploid genomes analysis such as cotton and oilseed rape (Salmon et al., 2010; Wang et al., 2011; Paterson et al., 2012) but, to our knowledge, this is the first time that this type of genomic event has been reported in a species with a hybrid origin. Other explanations can be possible such as contributions of unknown parents or random repair of heteroduplexes. Sequencing of different populations of wild petunias species, as well as *P. hybrida* accessions could be necessary to conclude the origin of these genes with multiple origins.

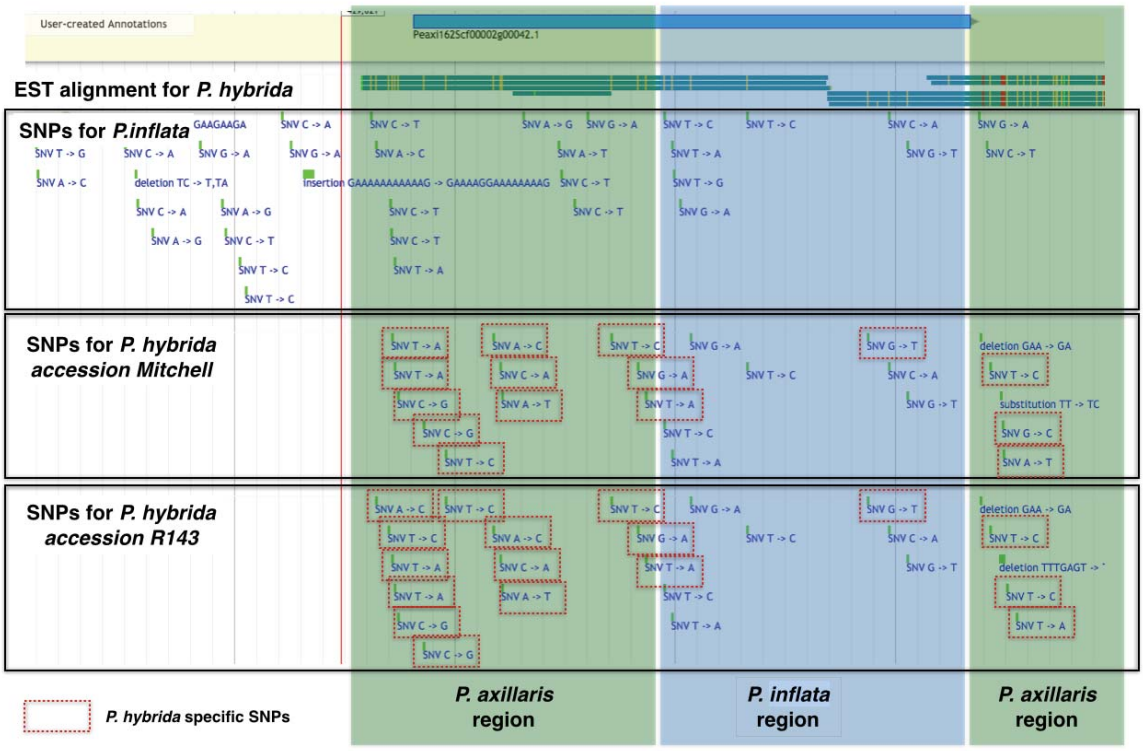


Figure 2: Example of a *Petunia hybrida* gene (Peaxi162Scf00002g00042.1 as *P. axillaris* reference) showing partial gene conversion between both ancestral genomes, *P. axillaris* and *P. inflata*. Four tracks beside the reference are shown: 1- Sanger EST from *P. hybrida* Mitchell with SNP marked in yellow, insertions in green and deletions in red; 2- *P. inflata* S6 polymorphisms; 3- *P. hybrida* Mitchell polymorphisms; 4- *P. hybrida* R143 accessions. For tracks 3 and 4 polymorphisms not present in *P. inflata* S6 are highlighted with a red box. Identified regions for the *P. hybrida* accessions are marked in green (*P. axillaris*) and blue (*P. inflata*) (Deletions and insertions are not marked).

To elucidate if there is a relation between the genes with non synonymous polymorphisms and possible functions, a Gene Set Enrichment Analysis (GSEA) was performed. The GSEA for the Biological Process categories are summarized in the supplementary tables 4 and 5. The gene sets with *P. inflata* origin show enrichment in categories such as “Formaldehyde catabolic process” (one gene present of the two annotated), “DNA replication, synthesis of RNA primer” (one gene present of the two annotated), “mRNA cleavage” (one gene present of the two annotated) and “histone lysine methylation” (two or three genes present of the sixteen annotated). The case of the genes with gene conversion events from *P. axillaris* and *P. inflata* is different. There are strong enrichments several terms: “Regulation of the cell shape” with three of the four annotated genes present (see Tables 4 and 5 for more details).

GO ID	Term	Mitchell GSEA p-value	R27 GSEA p-value	R143 GSEA p-value
GO:0015976	Carbon utilization	0.0039	0.0050	N.S.
GO:0015977	Carbon fixation	0.0164	0.0208	N.S.
GO:0009987	**Cellular process**	**0.0193**	**0.0251**	**0.012**
GO:0005991	Trehalose metabolic process	0.0201	0.0228	N.S.
GO:0018343	Protein farnesylation	0.0202	0.0229	N.S.
GO:0019288	Isopentenyl diphosphate biosynthetic process	0.0202	0.0229	N.S.
GO:0042127	Regulation of cell proliferation	0.0202	0.0229	N.S.
GO:0000079	Regulation of cyclin-dependent protein serine/threonine kinase activity	0.0307	0.0420	N.S.
GO:0006269	**DNA replication, synthesis of RNA primer**	**0.0400**	**0.0452**	**0.029**
GO:0006379	**mRNA cleavage**	**0.0400**	**0.0452**	**0.029**
GO:0006656	Phosphatidylcholine biosynthetic process	0.0400	0.0452	N.S.
GO:0045040	**Protein import into mitochondrial outer layer**	**0.0400**	**0.0452**	**0.029**
GO:0046294	**Formaldehyde catabolic process**	**0.0400**	**0.0452**	**0.029**
GO:0006298	Mismatch repair	0.0405	N.S.	N.S.
GO:0034968	**Histone lysine methylation**	**0.0405**	**0.0053**	**0.023**
GO:0034477	U6 snRNA 3'-end processing	N.S.	N.S.	N.S.
GO:0009772	Photosynthetic electron transport in photosystem II	N.S.	N.S.	N.S.
GO:0006378	mRNA polyadenylation	N.S.	N.S.	0.044
GO:0009306	Protein secretion	N.S.	N.S.	0.044
GO:0042823	Pyrodoxal phosphate biosynthetic process	N.S.	N.S.	0.044

Table 4: Results for GSEA for Biological Process, for the *P. inflata* origin homozygous genes from *P. hybrida*. In bold the categories statically significant in all the accession.

GO ID	Term	Mitchell GSEA p-value	R27 GSEA p-value	R143 GSEA p-value
GO:0008360	**Regulation of cell shape**	**0.00078**	**0.00092**	**0.00074**
GO:0006499	**N-terminal protein myristoylation**	**0.00348**	**0.00389**	**0.00333**
GO:0051301	Cell division	0.00357	N.S.	N.S.
GO:0015937	**Coenzyme A biosynthetic process**	**0.01003**	**0.01119**	**0.00962**
GO:0006544	Glycine metabolic process	0.01313	0.01531	N.S.
GO:0006810	**Transport**	**0.01411**	**0.01304**	**0.01104**
GO:0046907	**Intracellular transport**	**0.01900**	**0.02118**	**0.01842**
GO:0071805	**Potassium ion transmembrane transport**	**0.02948**	**0.03636**	**0.02719**
GO:0000902	Cell morphogenesis	0.03065	0.03407	N.S.
GO:0006563	L-serine metabolic process	0.03752	0.00697	N.S.
GO:0015995	**Chlorophyll biosynthetic process**	**0.04452**	**0.04935**	**0.04280**
GO:0006879	Cellular iron ion homeostasis	N.S.	0.01119	N.S.
GO:0006108	Malate metabolic process	N.S.	0.02754	N.S.
GO:0006825	Copper ion transport	N.S.	0.04332	N.S.
GO:0006564	L-serine biosynthetic process	N.S.	0.04935	0.04280
GO:0006826	Iron ion transport	N.S.	004935	N.S.
GO:0006750	Glutathione biosynthetic process	N.S.	N.S.	0.01842
GO:0015693	Magnesium ion transport	N.S.	N.S.	0.02552
GO:0055085	Transmembrane transport	N.S.	N.S.	0.03104

Table 5: Results for GSEA for Biological Process, for the partial gene conversions between *P. axillaris* and *P. inflata* origins from *P. hybrida*. In bold the categories statistically significant in all the accession.

DISCUSSION

The analysis of the reads produced by the transcriptome sequencing of three *Petunia hybrida* accessions (Mitchell/W115, R27, R143) supports the previous data about the hybrid origin of the garden petunia. Nevertheless the high contribution of *P. axillaris* shared between the three accessions could indicate multiple early backcrosses of the hybrid with *P. axillaris* accessions. The high number of polymorphisms not shared with the *P. axillaris* reference (PaxiN) also may indicate that this accession probably was very different from the *P. axillaris* N, but further research should be done to identify the most probable population that was used in the original crosses. Previous research has reported that the self incompatibility mechanism "it does not appears to have arisen an artifact of the hybrid crosses between *P.inflata*/*P. integrifolia* and *P. axillaris*" (Gerats and Strommer, 2009) so it is likely that early crosses were driven for other causes such the development of scented purple flowers (being the scent a *P. axillaris* contribution and flower color a *P. inflata/P. integrifolia* contribution) or simply plant access (early *P. hybrida* were wrongly identified as *P. violaceae*)(Griesbach, 2006). Still, more accessions should be analyzed before get any conclusion. As it was described before, *P. hybrida* was

probably originated by different hybridization events and it would be possible to find accessions originated in backcrosses with purple flower petunias (R.E. Koes personal communication). Other result produced for this analysis is the evidence for gene conversions in diploid hybrid origin species. Gene conversions are common in allopolyploids. They have been reported in several allopolyploid species such as cotton (Salmon et al., 2009; Paterson et al., 2012), tobacco (Lim et al., 2000) and oilseed rape (Chalhoub et al., 2014). Allopolyploids are considered fixed hybrids, so it is not surprising to find these results in a hybrid origin diploid species. It has been reported that homoeologous exchanges in allopolyploids, as a source of gene conversions, has been selected during the domestication of oilseed rape (Chalhoub et al., 2014). Some authors propose gene conversion as a mechanism for neofuncionalization for duplicated genes (Teshima and Innan, 2008), although it is not clear the role that this mechanism could have during the plant domestication. Plant domestication often involves gain of new traits by gene function loss (nonfunctionalization) such as *SH1* in *Sorghum bicolor* and *tb1* and *Sh1-5.1* in *Zea mays* (Meyer and Purugganan, 2013). In this context, we hypothesize that gene conversion could be a mechanism for gene non-functionalization during the early stages of plant domestication with a hybrid origin, but dozens wild and domesticated accessions should be analyzed to support this hypothesis.

MATERIAL AND METHODS

Reads from *Petunia hybrida* accessions Mitchell (Villarino et al., 2014), R27, R143 and W115 (also considered Mitchell) (supplied by the Prof. Francesca Quattrocchio) were processed with Fastq-Mcf with a minimum quality score of 30 and a minimum length of 50 bp (Aronesty, 2013) and mapped to the *P. axillaris* genome draft (v1.6.2) using Bowtie2 (Langmead and Salzberg, 2012). SNPs were called using FreeBayes (Garrison and Marth, 2012) with a minimum read depth of 5 and a minimum mapping quality of 20. Polymorphisms were annotated using Snpeff (Cingolani et al., 2012). To assign exons and genes to *P. axillaris* or *P. inflata*, a Perl script was used (RegionSNPAnalyzer). A minimum of 5 SNPs where used with a min. of two SNPs per Petunia progenitor. Five categories were used: 1- Homozygous *P. axillaris* gene/exon; 2- Homozygous *P. inflata* gene/exon; 3- Heterozygous *P. axillaris/P. inflata* gene/exon; 4- Homozygous *P. axillaris* and *P. inflata* gene/exon and 5- unclear assignment. Some of the homozygous SNPs for the genes with exons from both species were confirmed aligning *P. hybrida* EST supplied by the Prof. Reinhard (Breuillin et al., 2010). Read alignments and SNP were visualized using Web Apollo (Lee et al., 2013).

Gene Set Enrichment Analysis (GSEA) was performed using the Bioconductor package TopGO (Alexa and Rahnenfuhrer, 2010) for the three different types “Biological Process”, “Cellular Component” and “Molecular Function”. Because the gene sets were associated to the present/absent in the categories homozygous *P. inflata* gene/exon and homozygous *P. axillaris* and *P. inflata* gene/exon, a Fisher test was used. Three different algorithms were tested, “classic”, “elim” and “weight01”. The results presented are based in the most conservative analysis Fisher with the “weight01” algorithm.

REFERENCES

Alexa, A. and Rahnenfuhrer, J. (2010). Bioconductor - topGO. R package version 28.

Ando, T., Kokubun, H., Watanabe, H., Tanaka, N., Yukawa, T., Hashimoto, G., Marchesi, E., Suarez, E., and Basualdo, I.L. (2005). Phylogenetic analysis of Petunia sensu Jussieu (Solanaceae) using chloroplast DNA RFLP. Ann Bot-London **96**: 289–297.

Aronesty, E. (2013). Comparison of sequencing utility programs. Open Bioinform J.

Breuillin, F. 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.

Chalhoub, B. et al. (2014). Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science **345**: 950–953.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) **6**: 80–92.

DeHaan, L.R. and Van Tassel, D.L. (2014). Useful insights from evolutionary biology for developing perennial grain crops. Am J Bot **101**: 1801–1819.

Garrison, E. and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv.

Gerats, T. and Strommer, J. eds (2009). Petunia (Springer New York: New York, NY).

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

Griesbach, R.J. (2006). Flower Breeding and Genetics (Springer Netherlands: Dordrecht) Chapter 11: Petunia.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Meth **9**: 357–359.

Lee, E., Helt, G.A., Reese, J.T., and Munoz-Torres, M.C. (2013). Web Apollo: a web-based genomic annotation editing platform. Genome.

Lim, K., Kovarik, A., Matyasek, R., Bezdek, M., Lichtenstein, C., and Leitch, A.R. (2000). Gene conversion of ribosomal DNA in Nicotiana tabacum is associated with undermethylated, decondensed and probably active gene units. Chromosoma **109**: 161–172.

Meyer, R.S. and Purugganan, M.D. (2013). Evolution of crop species: genetics of domestication and diversification. Nat Rev Genet **14**: 840–852.

Mitchell, A.Z., M.R. Hanson, R.C. Skvirsky, and F.M. Ausubel (1980) Anther culture of Petunia: Genotypes with high frequency of callus, root, or plantlet formation. Z. Pflanzenphysiol. 100:131-146

Paterson, A.H. et al. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature **492**: 423–427.

Salmon, A., Flagel, L., Ying, B., Udall, J.A., and Wendel, J.F. (2010). Homoeologous nonreciprocal recombination in polyploid cotton. New Phytol **186**: 123–134.

Salmon, A., Flagel, L., Ying, B., Udall, J.A., and Wendel, J.F. (2009). Homoeologous nonreciprocal recombination in polyploid cotton. New Phytol **186**: 123–134.

Segatto, A.L.A., Ramos-Fregonezi, A.M.C., Bonatto, S.L., and Freitas, L.B. (2014). Molecular insights into the purple-flowered ancestor of garden petunias. Am J Bot **101**: 119–127.

Teshima, K.M. and Innan, H. (2008). Neofunctionalization of duplicated genes under the pressure of gene conversion. Genetics **178**: 1385–1398.

Villarino, G.H., Bombarely, A., Giovannoni, J.J., Scanlon, M.J., and Mattson, N.S. (2014). Transcriptomic analysis of Petunia hybrida in response to salt stress using high throughput RNA sequencing. PLoS ONE **9**: e94651.

Wang, X. et al. (2011). The genome of the mesopolyploid crop species Brassica rapa. Nat Genet **43**: 1035–1039.

Supplementary Note 7

The genes behind the different colors of *P. axillaris* and *P. inflata* flowers

Contributors: Mattijs Bliek, Kees Spelt, Valentina Passeri, Susan L. Urbanus, Ronald Koes and Francesca M. Quattrocchio

Department of Plant Development and (Epi)Genetics, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, Amsterdam, The Netherlands.

*For correspondence: f.quattrocchio@uva.nl

Abstract

Petunia wild accessions differ in pigmentation patterns and this is suspected to have played a role in species separation. Comparison of the structural anthocyanin genes in *P. axillaris*N and *P.inflata*S6 genomes has not shown obvious differences in coding sequences which could account for the divergence in pigmentation displayed by these species. A promoter rearrangement in the *CHIa* gene of *PaxiN* abolishes expression in anthers and is responsible for the accumulation of yellow chalcon. Mutations in MYB anthocyanin regulators are responsible for further pigmentation differences between *Pinf*S6 and *Paxi*N. The *AN2* gene is inactivated by a small insertion in the coding sequence in *Paxi*N, while in the same species, *AN4* is duplicated and both copies present rearrangements in the promoter. The comparison of the genomic regions of *Paxi*N and *Pinf*S6 containing these genes and other two MYBs known to regulate anthocyanin accumulation in green tissues reveals multiple insertions/deletions and loss of synteny indicating that vast rearrangements have occurred. The analysis of corresponding genomic fragments in tomato and the less closely related monkey flower, suggest events of proliferation of this genes occurring often during evolution, leading to species specific sets of MYB anthocyanin regulators. This is supported by phylogenetic analysis. Each set of MYB regulators displays different expression patterns as result of gene duplication followed by heavy rearrangements that have interested mainly the regulatory regions of each copy. We propose that the extraordinary plasticity of the DNA sequences in which the MYB coding sequences are embedded is responsible for the large variety of pigmentation patterns that we observe in nature.

Introduction

One of the most obvious differences between *P. inflata* and *P. axillaris* plants is the color of their flowers. This eye-catching character is a main factor in the different pollination strategies of both species and may have strongly contributed to their genetic separation.

Color and patterns, along with other features, such as scent and flower shape, are a major cue to attract specific pollinating animals (Hoballah et al., 2007; Klahre et al., 2011; Willmer, 2011). The evolution of pigmentation patterns is consequently an important player in speciation as mutations that result in attraction of a different pollinator contribute to reproductive isolation (Hoballah et al., 2007; Klahre et al., 2011). Comparison of genes involved in the distinct coloration of the *P. inflata* and *P. axillaris* flowers provides therefore insight into speciation processes.

Among the pigments that color foliage, flowers, fruits, and seeds, anthocyanins and the related proanthocyanins are the most diffused in the plant kingdom. Because pigmentation is a major factor determining the ornamental value of flowers and because it is a convenient visual marker for gene activity, pigmentation has been an all-time favorite of breeders and geneticists alike. Hence, over the decades, if not centuries, a plethora of pigmentation mutants has been isolated in a wide variety of species with defects in the synthesis of anthocyanin pigments and/or their sequestrations in the vacuole.

Our understanding of the genetic and biochemical processes that contribute to the final color of pigmented cells is perhaps most advanced in *Petunia hybrida*. Several studies describe the genetic analysis of in *P. hybrida* identified regulatory genes encoding transcription factors and downstream "structural" genes encoding enzymes involved in the synthesis of anthocyanins, their accumulation and stabilization in the vacuole, the synthesis of flavonol co-pigments, and the acidification of the vacuolar lumen where anthocyanins are stored (de Vlaming et al., 1984; Tornielli et al., 2009), as well as genes that control the shape and optical characteristics of the colored epidermal cells (van Houwelingen et al., 1998; Baumann et al., 2007).

Over the years most of these loci have been isolated from *P. hybrida* and their roles in the specification of flower color and pigmentation patterns analyzed. Further analysis of these genes in the *P. inflata S6* and *P. axillaris N* genomes can provide insight in the installation and reinforcement of two different pollination syndromes in purple versus white flowering petunias in order to reconstruct the sequence of events that caused the appearance of new species.

Anthocyanin biosynthetic genes

Anthocyanins belong to the larger secondary metabolite family of flavonoids and all share the same basic structure consisting of three carbon rings. The pathway that synthesizes anthocyanins starts from phenylalanine and can be divided in 3 major parts (Figure 1). The conversion of phenylalanine into 4-coumaroyl-CoenzymeA by the enzymes Phenylalanine Ammonialyase (PAL), Cinnamate 4-Hydroxylase (C4H) and 4-Coumaroyl-CoA ligase (4CL) is known as the general phenylpropanoid pathway and delivers precursors for the synthesis of various compounds, including lignin, volatiles and flavonoids.

The conversion of 4-coumaroyl-CoA into a dihydroflavonol constitutes the general part of flavonoid metabolism and generates the precursors for the synthesis of all flavonoid classes, including anthocyanins and flavonols. This involves the condensation of 4-coumaroyl-CoA with 3 molecules of malonyl-CoA by the enzyme Chalcone Synthase (CHS) to yield a yellow-colored chalcone that is converted by Chalcone Isomerase (CHI) into a colorless flavanone (naringenin). Isomerization happens also spontaneously in the absence

of the CHI enzyme, but this yields a mixture of the two stereoisomers of which only one can be converted by Flavanone 3-Hydroxylase (F3H) into a dihydroflavonol (dihydrokaempferol). Two additional hydroxylases, Flavonoid 3'Hydroxylase (F3H) (encoded by the *Ht1* locus) and F3'5'H (encoded by *HF1* and *HF2*), produce the 3' and/or 5'-hydroxylated dihydroflavonols (dihydroquercetin and dihydromyricetin respectively) that are the precursors of different anthocyanins.

The first reaction that is specific for the side-branch that yields anthocyanins and proanthocyanindins (condensed tannins) is the conversion of the colorless dihydroflavonols into colored anthocyanidins by Dihydroflavonol Reductase (DFR) and Anthocyanin Synthase (ANS). From this point on, the subsequent enzymatic steps add different decorations, such as sugars and acyls groups, to the basal anthocyanin skeleton, resulting in more stable products with different color shades (hues). These decorations include the addition of glucose moieties by 3-Glucosyl Transferase (3GT) and 5- Glucosyl Transferase (5GT), a rhamnose moiety by an Anthocyanin Rhamnosyl Transferase [ART] encoded by *RHAMNOSYLATION AT THREE* [*RT*]), and methyl groups by Anthocyanin 3'methyltransferase (A3'MT) and A3'5'MT encoded by *METHYLATION AT THREE* [*MT*] and *METHYLATION AT FIVE 5* [*MF*]) genes respectively) and acyl groups (by Anthocyanin Acyl Transferase, AAT). The addition of these decorations occurs in a precise order that is dictated by the substrate specificities of the enzymes catalyzing the single reactions (Provenzano et al., 2014).

Consequently, mutations that inactivate enzymes involved in an early step of decoration like, for example *rt*, block not only the addition of a rhamnosyl group, but also the subsequent addition of 5-glucose, acyl and methyl groups. Furthermore, the activity of some genes and enzyme activities has been reported to be restricted to certain tissues. The genes HF2, MF1 and MF2, for instance, do not affect pigmentation in tubes and anthers (de Vlaming et al., 1984; Provenzano et al., 2014).

Most of the enzymes of the anthocyanin pathway reside in the cytoplasm with the exception of the cytochrome P450 hydroxylases C4H, F3'H and F3'5'H which are in the endoplasmatic reticulum, whereas the anthocyanin end products accumulate in the central vacuole. The mechanism that transports anthocyanins across the tonoplast into the vacuole is however poorly understood. The accumulation of anthocyanins in the lumen of the vacuole is affected by the *AN9* locus, which encodes a Glutathione Transferase (GST) thought to be required for the transport of the pigments across the tonoplast (Alfenito et al., 1998). The activity of *AN9* is required for proper pigmentation of the petals as *an9* mutants show very pale flowers. The *AN9* protein has been shown to be functionally interchangeable with the maize *Bz2* (Alfenito et al., 1998). Although it has been proposed that glutathionation is required for sequestration of anthocyanins in the vacuole (Marrs et al., 1995), direct demonstration of this is still lacking. Therefore, the mode of action of AN9 remains unclear.

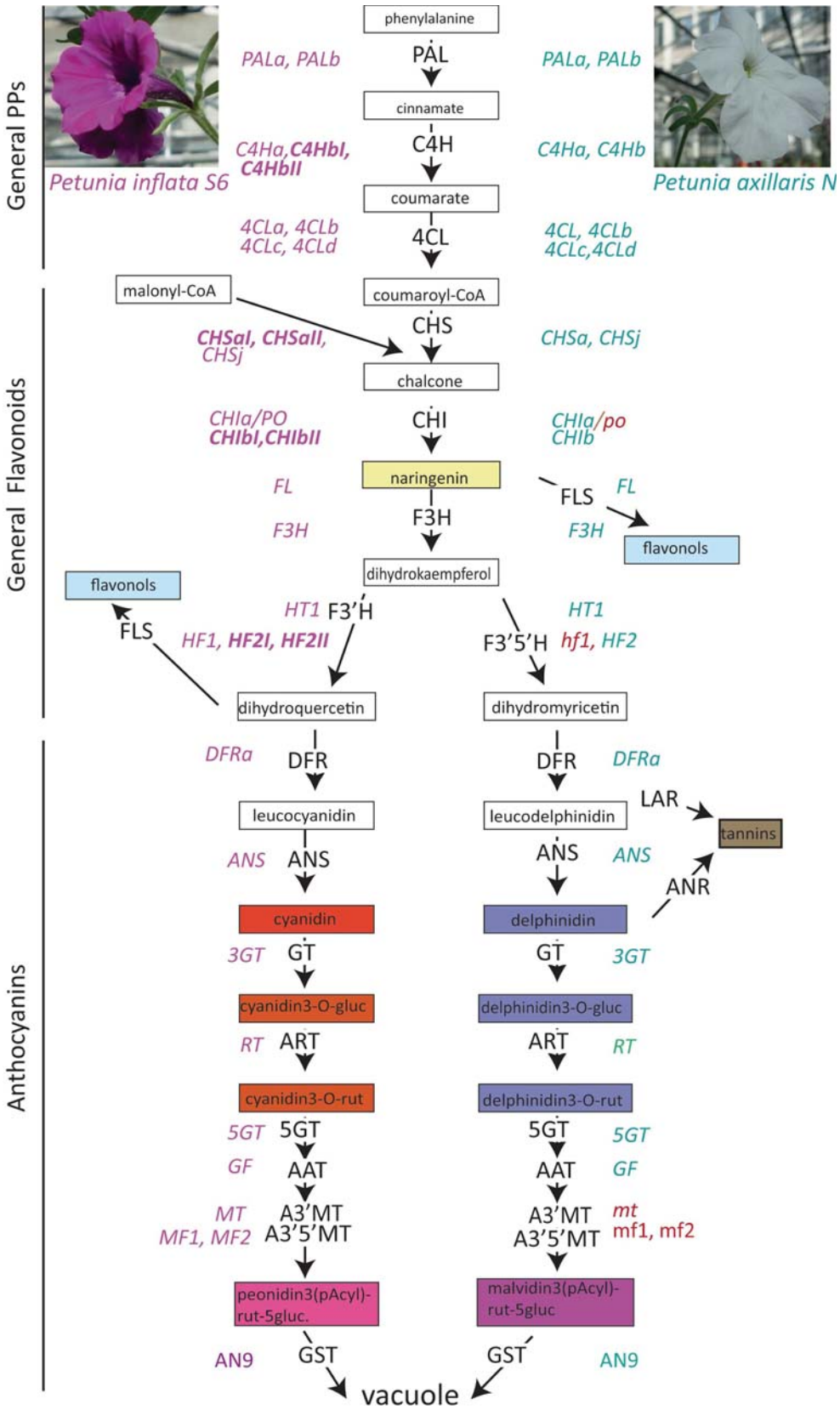


Figure 1. The anthocyanin biosynthetic pathway in petunia. The synthesis of colored anthocyanins from phenylalanine involves some 15 enzymes (indicated in black). The genes/alleles encoding these enzymes in *PinfS6* and in *PaixN* in purple and green letters respectively. Recessive (non-functional) alleles are in lower case and in red, while gene duplications are indicated in bold face.

Regulatory network controlling anthocyanin accumulation.

In Petunia, the transcriptional regulation of early biosynthetic genes (EBGs) encoding enzymes of general phenylpropanoid (PAL, C4H, 4CL) and general flavonoid metabolism (CHS, CHI, F3H) is largely independent from regulation of late biosynthetic genes (LBGs) encoding the enzymes that are specifically involved in the synthesis (and decoration) of proanthocyanidins and anthocyanins. This split regulation makes it possible for unpigmented tissues to accumulate flavonols and flavanones (Koes et al., 2005). Till now no regulators of the early steps of the pathway have been identified in petunia.

The transcriptional activation of LBGs requires in all plant species analyzed a WD40 repeat, a bHLH and a MYB protein, which form a protein complex (MBW) directly activating the promoters of downstream LBGs such as *DFR* (Spelt et al., 2000; Chan et al., 2010). In petunia, different MBW complexes are responsible for anthocyanin accumulation in different plant parts. The WDR protein ANTHOCYANIN 11 (AN11) (de Vetten et al., 1997) is more or less ubiquitously expressed throughout the plant and *an11* mutations abolish anthocyanin synthesis in all pigmented tissues. Instead the bHLH protein ANTHOCYANIN1 (AN1) is only expressed in tissues that accumulate anthocyanin or tannins, and *an1* mutants loose anthocyanins or proanthocyanidins in all tissue (Spelt et al., 2000). Although AN1 is structurally and functionally similar to bHLH regulators of the anthocyanin pathway from maize (R-family) and Antirrhinum (DELILA, DEL), it belongs to a distinct clade of bHLH proteins that diverged from the R/DEL clade before the separation of monocot and dicot species (Spelt et al., 2000). The apparent R/DEL ortholog of *P. hybrida*, *JAF13*, can interact with the MYB and WD40 partners of the MBW complex and can activate transcription of downstream structural anthocyanin genes when ectopically expressed in leaf and stem tissues (Quattrocchio et al., 1998). However, the function of *JAF13* is not simply redundant with *AN1*, as *an1* mutants lose pigmentation in all tissues, despite the presence of an active *JAF13* gene. Since loss of function *jaf13* mutants are lacking the exact role of *JAF13* has remained mysterious.

In *P. hybrida* the MYB component of the MBW complex is encoded by multiple partially redundant genes, which have different expression patterns and control pigmentation in distinct subsets of tissues. The encoded MYB proteins are highly similar in sequence and function and can induce anthocyanin synthesis in a plethora of tissues, when expressed ectopically (Quattrocchio et al., 1998; Quattrocchio et al., 1999; Kroon, 2004; Albert et al., 2011). *AN2* is expressed in the petal limb and tube, but not in other pigmented tissues like anthers and pedicels. It encodes a MYB protein that can interact with AN1 and JAF13 and is sufficient to activate anthocyanins genes like *DFR* when ectopically expressed (Quattrocchio et al., 1999; Quattrocchio et al., 2006). The paralog *AN4* encodes a highly similar MYB protein that is essential for pigmentation of anthers and certain cells in the petal tube, but has no obvious role in pigmentation of the petal limb (de Vlaming et al., 1984; Kroon, 2004; Povero, 2011). *AN4* is expressed in anthers at early developmental stages to activate transcription of *AN1* and structural anthocyanin genes (LBGs). Curiously, *AN4* is also expressed in the petal limb, but only during late stages of flower development

when the expression of *AN1*, *DFR* and other structural anthocyanin genes has already ceased (Kroon, 2004; Povero, 2011). Two additional paralogs, *DEEP PURPLE* (*DPL*)/*MYBb* and PURPLE HAZE (PHZ) are expressed in floral and vegetative tissues and have been implicated in the accumulation of pigment in vegetative tissues (Kroon, 2004; Albert et al., 2011).

Both *an1* and *an11* mutants loose the brown coloration of the seed coat epidermis, indicating that these two factors are required for the accumulation of proanthocyanidins (condensed tannins) in this tissue (Koes et al., 1990; Spelt et al., 2002; Zenoni et al., 2011). Proanthocyanidin synthesis is, however, not affected in *an2 an4* mutants. Hence, the MYB partner of AN1 and AN11 required for activation of the proanthocyanidin pathway, if any, remains to be identified.

Two MYBs belonging to a different clade, MYBX and MYB27, have been shown to act as repressors of anthocyanin gene when overexpressed, but their function in planta is not fully understood (Kroon, 2004; Albert et al., 2011).

Regulation of the vacuolar pH in petal epidermal cells.

Anthocyanins behave as pH indicators as their absorption spectrum and color depends on the pH of the solvent. In epidermal petal cells of *P. hybrid*a vacuoles are moderately acidic, which gives a reddish color to anthocyanins that are stored there. Mutations in any of seven *PH* loci (*PH1-PH7*) result in more bluish petals and reduce the acidity of crude petal extracts and of the vacuolar lumen, suggesting that these genes are involved in a pathway that acidifies vacuoles (de Vlaming et al., 1983; van Houwelingen et al., 1998; Verweij et al., 2008; Faraco et al., 2014). In most plant cells ubiquitous V-ATPase pumps acidify the vacuoles and other endomembrane compartments. Pigmented epidermal cells have vacuoles that are more acidic than those from vegetative tissues, which relies on a distinct and more powerful proton pump consisting of two distinct P-ATPases encoded by *PH1* and *PH5*. PH5 belongs to the P_{3A} ATPase subfamily of proton pumps, which includes the main plasma membrane proton transporters. Instead, PH5 is localized on the tonoplast to pump protons into the vacuole (Verweij et al., 2008). However, PH5 alone has a rather modest proton pumping activity, which can be boosted by interaction with the P_{3B}-ATPase PH1. This protein, the first of this family to be found in eukaryotes, is very similar to bacterial magnesium transporter, but the mechanism by which it enhances the H^+ pumping activity of PH5 remains to be solved (Faraco et al., 2014).

Expression of *PH1* and *PH5* and vacuolar acidification requires a regulatory MBW complex that includes AN1, AN11, and the MYB protein PH4 plus the WRKY protein PH3, and is consequently directly linked to anthocyanin synthesis through the use of several shared regulators (Verweij et al., 2008; Faraco et al., 2014).

Results

The differences in pigmentation pattern are thought to be a major factor for the distinct pollination syndromes of moth-pollinated *P. axillaris* flowers and bee-pollinated *P. inflata* flowers. In *P. inflata* flowers corolla limb and tube as well as anthers are colored by anthocyanins, whereas *P. axillaris* flowers have a cream-white petal limb, which lacks anthocyanins but contains copious amounts of flavonols (quercetin), and yellow anthers (Figure 1). *P. axillaris*N is, however, not completely acyanic, because it accumulates anthocyanins in the corolla tube and some vegetative tissues (such as pedicels) and proanthocyanidins in the seed coat epidermis.

To understand how the different pigmentation patterns evolved, we analyzed *P. axillaris* N and *P. inflata* S6 homologs of pigmentation genes known from *P. hybrida* and, in some cases, pigmentation genes that are known from other species.

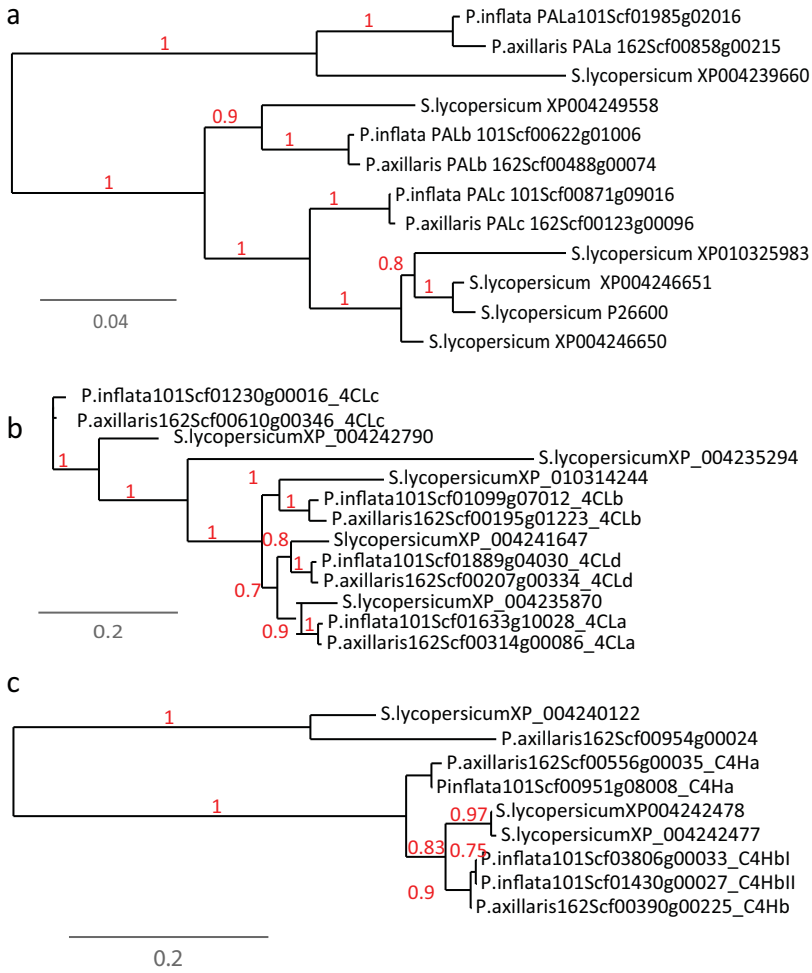


Figure 2. Phylogenetic analysis of PAL, C4H and 4CL proteins. (a) Phylogenetic tree of the PAL family, (b) of the cytochrome P450 proteins C4H and (c) of the 4CL proteins in *PaxiN (P. axillaris)*, *PinfS6 (P. inflata)* and tomato (*S.lycopersicum).*

Structural genes involved in the general phenylpropanoid pathway

Structural genes and encoded enzymes of the general phenylpropanoid pathway were not extensively studied by biochemical or genetic analyses in *P. hybrida* and, hence, were all inferred by similarity with homologs form other species. We found that the *PaxiN* and *PinfS6* genomes contain 2-4 paralogs encoding PAL-like, C4H-like and 4CL-like proteins (Figures 1 and 2). For each of the *PAL* and *4CL* homologs in *P. axillaris N* we found clear

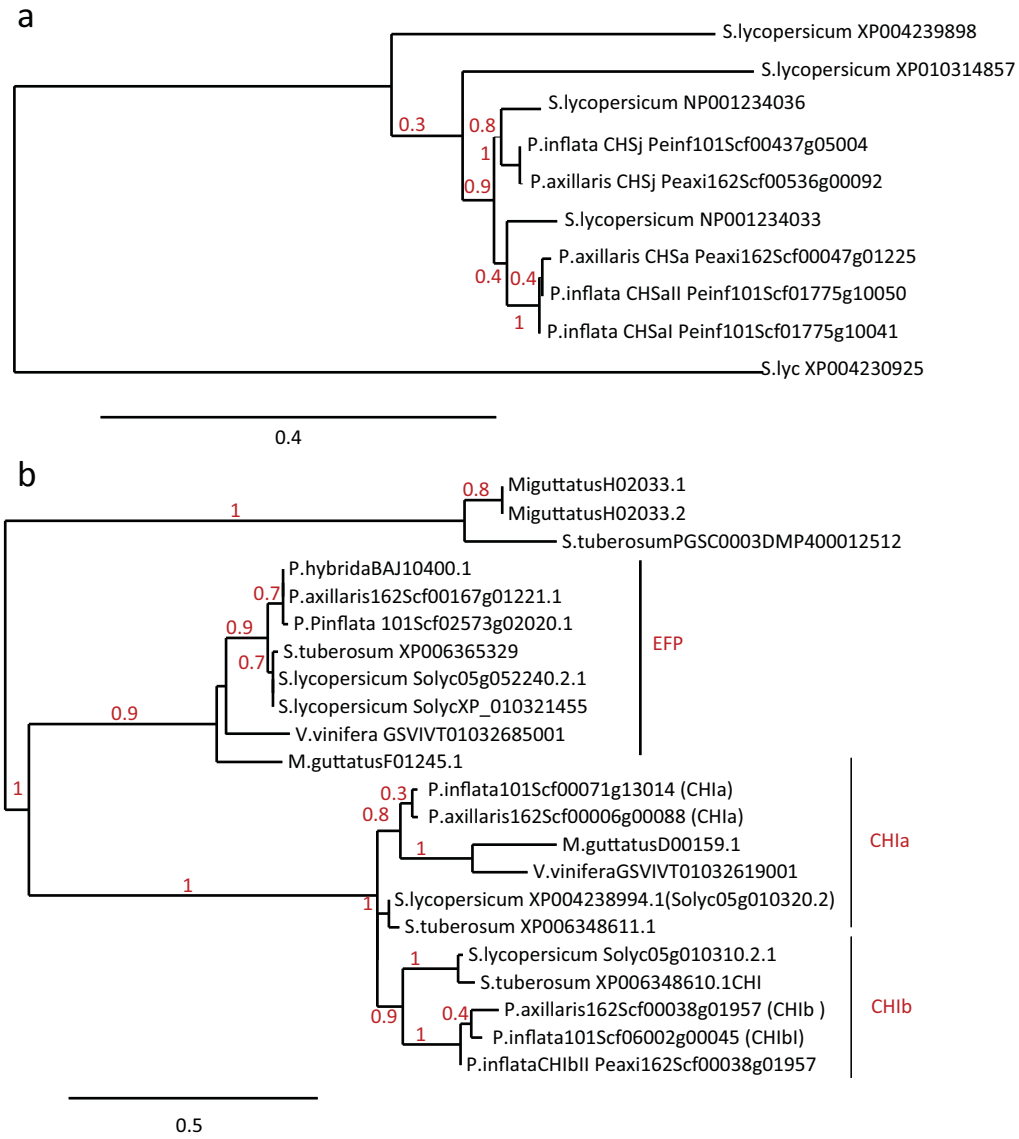


Figure 3. Phylogenetic analysis of the CHS and CHI families in petunia and tomato. (a) tree of the genes encoding CHSa and CHSj in petunia and tomato. The outgroup is here a CHS gene from tomato belonging to a different group. (b) CHI family in petunia(*P.inflata*, *P.axillaris* and *P. hybrida*), potato (*S. tuberosum*), tomato (*S. lycopersicum*), grape (*V. vinifera*) and Mimulus. The three different types (CHIa, CHIb and EFP) are present in all species. The proteins in the outgroup are putative "fatty acid binding proteins" closely related to CHI.

orthologs in *P. inflata S6* and in tomato, which belongs like petunia to the *Solanaceae* family (Figure 2). This indicates that the gene duplications that gave rise to these small gene families predate the separation of *Petunia* and tomato. The *C4H* gene family also expanded by gene duplication, but this happened more recently, after the separation of *Petunia* and *tomato*, and once more after the separation of *P. axillaris* from *P. inflata*, giving rise to *C4HbI* and *C4HbII* in the *P. inflata* genome (Figure 2).

Structural genes involved in general flavonoid metabolism

Structural genes encoding enzymes from general flavonoid metabolism have been extensively studied in *P. hybrida*. *P. hybrida* contains some 10-20 genes encoding CHS or CHS-like proteins (Koes et al., 1989b). Only two of these (*CHSa* and *CHSj*) are expressed at significant levels in floral tissues and upon induction with UV-light in seedlings and are required for pigmentation of floral tissues and seeds, while the others are putative pseudogenes or gene fragments and are hardly expressed (van der Krol et al., 1988; Koes et al., 1989a; Koes et al., 1990; Napoli et al., 1990; Napoli et al., 1999).

We found that tomato also contains distinct *CHSa* and *CHSj*-like genes, suggesting that these genes originate from duplication events that preceded the divergence of *Petunia* and tomato (Figure 3). Although *CHSa* and *CHSj* of *P. hybrida* encode functionally similar enzymes (Napoli et al., 1999) and are expressed in very similar patterns, their transcription is regulated differently, since *CHSj* expression is dependent on the regulatory MBW complex that also activates LBGs, whereas *CHSa* is expressed independently (Quattrocchio et al., 1993). It will be interesting to find if this also holds true for *CHSa* and *CHSj* homologs from tomato (and other species) as this can provide insight in the evolution of cis-regulatory gene sequences, a process that is thought to underlie much of the morphological information.

The *PinfS6* genome contains two linked copies of *CHSa* (designated *CHSaI* and *CHSaII*) that apparently stem from a recent duplication in the *P. inflata* lineage, as the *PaxiN* genome contains only a single *CHSa* copy. Distinct *P. hybrida* lines contain either two duplicated *CHSa* genes, an apparent introgression from *P. inflata*, or a single *CHSa* gene, that apparently originates from *P. axillaris* without any obvious effect on the pigmentation of floral tissues or the accumulation of anthocyanin (Koes et al., 1987). Hence, the functional relevance of this *CHSa* gene duplication, if any, is unclear.

In *P. hybrida* a single gene, *CHIa*, which is located at the *POLLEN* (*PO*) locus, is required for expression of enzymatically active CHI in the petals (limb and tube) and anthers (van Tunen et al., 1988). The *PaxiN6* and *PinfS6* genomes both contain a single *CHIa* homolog with a highly similar coding sequence. However, the *PaxiN* and *PinfS6* alleles do differ in their 5' flanking sequences, which affects their expression in anthers (see below). The *P. hybrida CHIb* gene encodes a CHI-like protein of unknown function. It is expressed in late stages of anther development, well after expression of other flavonoid and anthocyanin synthesis stops, and does not direct expression of a CHI enzyme that is active on a tetrahydroxychalcone substrate (van Tunen et al., 1988). The *PaxiN* genome contains a single *CHIb* gene, similar to the few *P. hybrida* lines that were analyzed, whereas *CHIb* has been duplicated in the *P. inflata S6* lineage. The diversification of *CHIa* from *CHIb* is also

found in the genus *Solanum*, while grape and Mimulus seem to miss an ortholog of *CHIb* (Figure 3).

Another CHI-like gene, encoding a type IV CHI protein, was recently discovered in *Ipomea* by analysis of mutants with diminished pigmentation of flowers. In the petals of these mutants, the expression of several structural genes is reduced compared to wild types. The silencing of the petunia homolog also results in petunia pale flowers (Morita et al., 2014). The precise function of this gene, called *EFP* (*Enhancer of Flavonoid Production*) is unknown, but since the encoded protein lacks several amino acids that are conserved in active CHI enzymes, it is unlikely to function as a CHI enzyme. Both *PaxiN* and *PinfS6* contain a copy of *EFP*. Tomato, potato, grape and Mimulus all have an ortholog of *EFP*, which shares a common ancestor with *CHIa* and *CHIb* (Figure 3).

In *P. hybrida* F3H and F3'H are encoded by the loci *AN3* and *HT1* (van Houwelingen et al., 1998; Brugliera et al., 1999). The *PaxiN* and *PinfS6* genomes both contain single *AN3* and *HT1* orthologs (Figure 1). Genetic analyses suggested that *P. hybrida* contains an additional locus, *HT2*, that is involved in 3' hydroxylation of dihydrokaempferol in the petal tube (de Vlaming et al., 1984). However, in the *PaxiN* and *PinS6* genomes we could not find an obvious paralog of the *F3'H*/*HT1* gene that might represent *HT2*. This suggests that *HT2*, might encode the cytochrome b5 protein DIFf (de Vetten et al., 1999) or another protein that is needed for F3'H enzyme activity. Alternatively, the genetic data suggesting the existence of two *HT* loci may have resulted from a translocation of *HT1* in some of the *P. hybrida* lines used for mapping, rather than a gene duplication.

Taken together these findings show that the *PaxiN* and *PinfS6* genomes both encode a complete set of enzymes involved in general phenylpropanoid and flavonoid metabolism, consistent with the observation that both species can synthesize anthocyanins and flavonols. Curiously, duplication of genes involved in the general phenylpropanoid and general flavonoid metabolism seems to have occurred repetitively in *P. inflata*, but they only involve early genes of the pathway: *C4Hb*, *CHSa*, *CHIb* and *HF2* (Figure 1). The similarity of the *PinS6* paralogs, indicates that they arose by recent gene

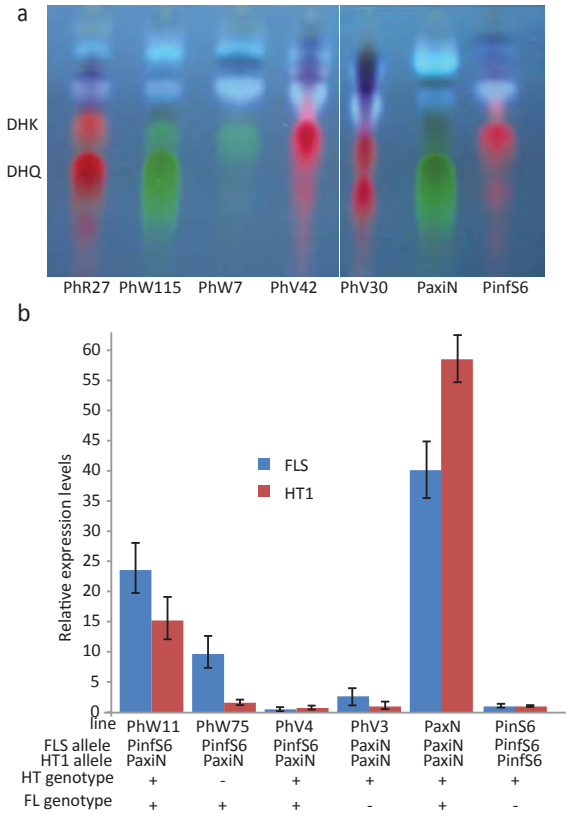


Figure 4. Regulation of genes involved in flavonols biosynthesis.
(a)Thin Layer Chromatography of flavonoids in petals of different P. hybrida lines lines. The picture is taken under exposure to UV light. In this conditions the anthoycnains fluoresce red and flavonols yellow/green. The position of dihydrokaempferol (DHK) and dihydroquercetin (DHQ) are indicated. Note that petals of *PaxiN* (*FL HT1*) and *PhW115* (*FL HT1*) contain large amounts of dihydroquercetin and a lower amount of dihydrokaempferol, while total flavonol amount is reduced in the fl lines PhR27, PhW75, *PhV42*, *PhV30* and PinfS6 and that in *PhW75* (fl ht1) kaempferol is the main flavonol, rather than quercetin. (b) Relative gene expression (referred to the expression of SAND). The alleles for FLS and HT1 (originating from PinfS6 or PaxiN) in the different lines have been determined by PCR amplification with allele specific primers. + and – indicate the presence of wild type or mutated alleles. All *Petunia hybrida* lines are from the Amsterdam collection.

duplication in the *P. inflata* lineage after the separation from *P. axillaris,* rather than the loss of a copy of an earlier duplicated gene in the *P. axillaris* lineage. This might be related to the necessity of producing enough precursors for the synthesis of anthocyanins and other flavonoids.

Regulatory and structural genes involved in flavonol synthesis

The first step that is specific for the synthesis of flavonols is the conversion of a dihydroflavonol into the corresponding flavonol by the enzyme Flavonol Synthase (FLS). In *P. hybrida* lines with dominant alleles for (nearly) all pigmentation genes the main flavonols are quercetin derivatives. Strikingly myricetin is only found in very small amounts, if at all, presumably because dihydromyricetin is a poor substrate for FLS.

In *ht1 ht2* mutants, which cannot synthesize dihydroquercetin, the main flavonols are derivatives of kaempferol (de Vlaming et al., 1984). The *amount* of flavonols that accumulate in various tissues is controlled by the locus *FLAVONOL* (*FL*) (de Vlaming et al., 1984). The dominant *FL* allele (*FL*) directs the accumulation of high amounts of flavonols, whereas in homozygotes for the recessive *fl* allele flavonol accumulation is strongly reduced, though not completely abolished.

Flowers of *P. inflata* accessions (and related species with colored flowers) accumulate low levels of flavonols, suggesting that they are *fl fl*, whereas petals of nearly all *P. axillaris* accessions accumulate copious amounts of flavonols, indicating that they harbor the dominant *FL* allele (Wijsman, 1983) (Figure 4A). Real time PCR analysis showed that *FLS* mRNA is >40-fold more abundant in *P. axillaris* N petals than in *P. inflata* S6 petals (Figure 4B). Interestingly, expression of *HT1*, which encodes F3'H, is also much higher in *P. axillaris N* than in *P. inflata S6* petals. However, we found no obvious sequence differences between the *FLS* and *HT1* alleles of *PaxiN* and *PinfS6* that might account for the different amounts of *FLS* and *HT1* mRNAs and flavonol end products that are synthesized.

Because in crosses of *P.axillaris*N and *P. exserta*, recombinants can be found between *FL* and *FLS* at low frequency (Hermann et al., 2013), we asked whether *FL* controls *FLS* and *HT1* expression *in trans*. Therefore, we analyzed *FLS* and *HT1* expression in petals of several inbred *P. hybrida* lines by real-time PCR. The *P. hybrida* lines used differ with respect to the *FL* genotype and consequently the amount of flavonols that accumulate in petals (Figure 4A). Importantly the amount of *FLS* mRNA in petals correlated with the *FL* genotype and seemed much less dependent, if at all, on the *FLS* allele (from *P. inflata* or *P. axillaris*) (Figure 4B). For example, the genomic sequence of the *FLS* allele of *P. hybrida* V30 indicates that it originates from *P. axillaris* rather than *P. inflata*, but it is nevertheless expressed at relatively low level, consistent with the *fl* genotype of V30. Similarly, the *FLS* allele of *P. hybrida* W115 originated from *P. inflata* rather than *P. axillaris*, but is in W115 relatively strongly expressed, consistent with the *FL* genotype of this line. A similar result was obtained for *HT1*. The *HT1* alleles of all four analyzed *P. hybrida* lines originated from *P. axillaris*, but were expressed at different levels in accordance with the *fl* genotype of the host. The only exception is *P. hybrida* W75; this line expresses little or no *HT1* mRNAs despite the *FL* genotype, because it contains a recessive (mutant) *ht1* allele (Figure 4B), and consequently accumulates kaempferol rather than quercetin as the major flavonol (Figure 4A).

These results suggest the coordinate expression of groups of genes involved in specific branches of the flavonoid pathway. This also postulates regulatory networks independent from *AN2* for the control of the expression such group(s) of genes. Recently, a mutation in the promoter of *MYB-FL*, a gene encoding for a transcription factor, was shown to be responsible for the increase of flavonol production observed in *P. axillaris* as compared to *P. inflata* (Sheehan et al., 2015). As *PinfS6* does not accumulate flavonols, the duplication of multiple early genes of the pathway is not likely to be related to the need to synthesize a larger amount of precursors for the production of these co-pigments.

WD40 and bHLH transcription regulators controlling anthocyanin synthesis

Genetic analyses identified a regulatory complex consisting of a WD40, bHLH and MYB protein that is required for the transcriptional activation of structural genes involved in anthocyanin and/or proanthocyanidin synthesis in a broad range of Angiosperms (Koes et al., 2005; Petroni and Tonelli, 2011). In *P. hybrida* the WD40 protein is encoded by a single gene, *AN11*, that is homologous to similar regulators from Arabidopsis (TTG1) and maize (PALE ALEURONE COLOR, PAC) (de Vetten et al., 1997; Walker et al., 1999; Carey et al., 2004). However, the bHLH partners identified in distinct species belong to two distinct clades. One of these clades is defined by proteins of the maize R-family, which are encoded by multiple paralogs with diverse expression patterns (Ludwig and Wessler, 1990), and DELILA (DEL) from *Antirrhinum* which is required for pigmentation of the flower tube (Goodrich et al., 1992). By contrast, AN1 of *P. hybrida*, which is required for anthocyanin synthesis in all pigmented tissues, belongs to a distinct bHLH clade that diverged from the R/DEL clade before the split of monocot and dicot species (Spelt et al., 2000). *P. hybrida* possesses in addition at least one ortholog of *R*/*DEL* genes. This gene, *JAF13*, is expressed in all pigmented tissues, can activate transcription of *DFR* (and other genes of the pathway) when ectopically expressed in leaves (Quattrocchio et al., 1998) and can, like AN1, interact with the WD40 (AN11) and MYB regulators (AN2, PH4) of the anthocyanin and vacuolar acidification pathways (Quattrocchio et al., 2006). The genomes of closely related *Solanaceae* (tomato and potato) also encode bHLH proteins belong to the AN1 and JAF13 clades (Figure 7). Since the *an1* phenotype shows that *JAF13* cannot replace *AN1*, this raised the question whether *jaf13* mutants are lacking because of redundancy or because *JAF13* has no role in activating anthocyanin genes.

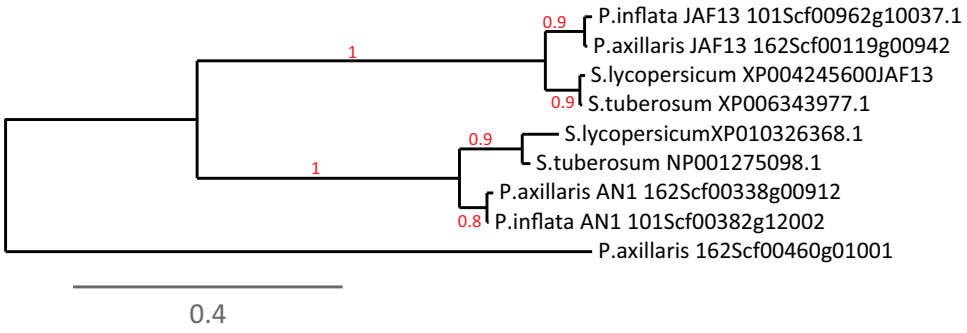


Figure 5. bHLH anthocyanin regulatory proteins in petunia, potato and tomato. Two distinct clades of proteins (AN1 and JAF13 homologs) are present in all three species. In the genomes of both PinfS6 (P. inflata) and PaxiN (P. axillaris) is only one JAF13 gene as supported by the phylogeny of the protein with the highest homology to JAF13 after AN1 (Peaxi162Scf00460g01001), which is not part of the JAF13 clade.

In both *PaxiN* and *PinfS6* genomes we found only one copy of *JAF13*. The second most similar protein in *PaxiN* is AN1, and the third is Peaxi162Scf00460g01001 (Figure 5). In a phylogenetic tree, this last protein groups together with the Arabidopsis bHLH factor AtMYC2, which was shown to be involved in abscisic acid signaling (Abe et al., 2005)

(Figure 6a). This suggests the *JAF13* function is unique in petunia, and that the absence of *jaf13* mutants is not due to redundancy with paralogous genes.

To identify *jaf13* mutants we screened populations of the *P. hybrida* line W138 by PCR for transposon insertions in *JAF13* (Koes et al., 1995). This yielded the *jaf13*^{V2032} allele. Homozygotes for *jaf13*^{V2032} did not show any phenotype. RNA analyses revealed that *jaf13*^{V2032} petals express a longer RNA that retains the *dTPH1* insertion 253 bp before the end of the coding sequence (Figures 6b and 6c). As it was unclear whether *jaf13*^{V2032} was a mutant allele with reduced activity, we screened 500 progeny plants by PCR for intragenic transpositions and found a new allele, *jaf13*^{B2128}, with a *dTPH1* insertion upstream of the bHLH domain (Figure 6c). Petals of *jaf13*^{B2128} homozygotes had a paler color, than those from wild type siblings, and often contained (revertant) spots and sectors with more intense pigmentation, which might result from transposon excisions (Figure 6d). Quantitative analysis of the anthocyanin content in the petals confirms that the *jaf13* mutant accumulates 25 to 30% less anthocyanin than wild type petals (Figure 6d).

To address whether the reduced pigmentation was due to the *dTPH1* insertion in *jaf13*^{B2128} we analyzed DNA from pale and dark petal sectors of *jaf13*^{B2128} homozygotes and, as a control, from wild type petals. PCR analysis with two primers flanking the B2128 transposon insertion, amplified from *jaf13*^{b2128} homozygous petals a ~ 800-bp fragment containing the *dTPH1* insertion and a ~500-bp fragment, which is similar to wild type size, and results from (multiple) somatic excisions (Figure 6f, top panel). To analyze these somatic excisions at higher resolution, we used radioactively labeled primers that amplified a smaller fragment and analyzed the products with 1 bp resolution on (denaturing 6% polyacrylamide) sequencing gels, as described previously (van Houwelingen et al., 1999). Using DNA from petal tissue without any phenotypic selection showed that most excision products contain a 7-bp footprint, which does not restore the JAF13 reading frame, whereas smaller footprints were formed less frequently. However, analysis of two full colored revertant spots showed that those cells contain high amounts of an excision product with either a 6 or 0-bp footprint respectively (and at lower abundance an array of different footprints that result from excision of the sister allele).

Therefore we concluded that the reduce pigmentation of *jaf13*^{B2128} petals is indeed caused by the transposon insertion in *jaf13*.

RT-PCR analysis of *jaf13*^{B2128} petals showed no significant decrease of *JAF13* transcripts, however, mRNAs of *DFR* and *CAC*, a target gene of AN1, PH3 and PH4 (Verweij, 2007), are about 40% less abundant compared to wild type petals (Figure 6g). By screening phage display libraries (de Bruin et al., 1999), we isolated 7 phages able to bind to the JAF13 protein. We used the three phages (Ab2, Ab3 and Ab5) giving the best signals in ELISA assays to analyze native protein extracts from leaves (which do not express JAF13 mRNA) and petals of different genotypes. ELISA assays of native protein extracts from different plant tissues, showed that wild type and *an11* (*P. hybrida* line W134) petals give stronger signal with all three antibodies, than *jaf13*^{B2128} petals, while no signal was detected in leaves (Figure 6h). This indicates that the truncated protein encoded by *jaf13*^{B2128} accumulates at slightly lower levels than wild type JAF13 and, hence, may have some residual activity, analogous to similar truncations of AN1 (Spelt et al., 2002).

Taken together these results suggest that JAF13 has a minor role in the activation of anthocyanin synthesis and possibly vacuolar acidification, and acts as an “intensifier” that enhances the action of the AN1/AN11/AN2 complex.
The MYB partners of the MBW complex are discussed in the next section

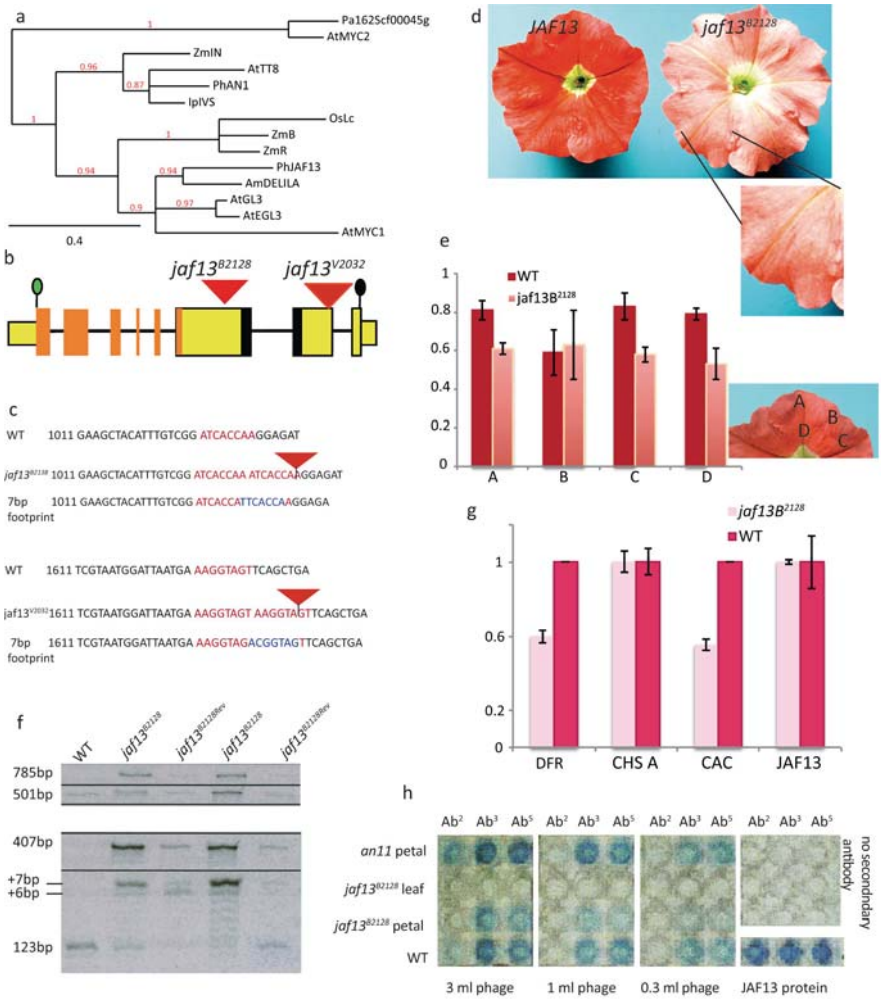


Figure 6. Analysis of *jaf13* mutant.
a) Phylogenetic analysis of bHLH proteins related to JAF13 and AN1. Mining the PaxiN protein collection with the JAF13 sequence yields, in order of similarity, JAF13, AN1 and Pa162scf00045g protein sequences. b) Structure of the *JAF13* gene. Introns are given as thin lines and exons as rectangles. In region encoding conserved N-terminal domain and the bHLH domain is indicated by orange and black shading, a green oval indicates the START of translation and a black one the translation STOP. Red triangles indicate the transposon insertion in *V2032* and *B2128*. c) Sequences of the alleles *jaf13*^{B2128} and *jaf13*^{V2032} at the site of insertion of the transposons compared to the *JAF13* sequence at the same position. The target site duplication is indicated in red. The sequences of excision alleles containing 7-bp footprints are shown for both *jaf13*^{B2128} and *jaf13*^{V2032}. In blue the extra nucleotides compared to the wild type sequences. d) Phenotype of a wild type flower compared to a *jaf13*^{B2128} mutant. The petals of the mutant are lighter in color and in the insert darker reversion spots are recognizable. e) Analysis of anthocyanin content in different areas (A, B, C and D) of petals from wild type and *jaf13*^{B2128} flowers (n=3 ±SD). The value indicate OD_{530}. f) Analysis of footprints generated by excision of dTPh1 from *jaf13*^{B2128} with two combinations of primers. These primer sets amplify respectively a small (lower panel) and a larger (upper panel) fragment of the *JAF13* gene, both containing the site of insertion of the transposon in *jaf13*^{B2128}. A footprint of 6 base pairs and a perfect excision are detected in 2 different revertant sectors. In mutant tissue, footprints of 7 base pairs represent the most frequent excisions of the transposon. g) RT-PCR of anthocyanin gene expression in petals of wild type and *jaf13*^{B2128} flowers. Values

were acquired by radio-activity quantification (by ImageQuant) of blots hybridized with probes for each of the genes. Expression values are given in arbitrary units standardized for the expression of actin. h) quantification of the JAF13 protein in *jaf13*^{B2128} mutant flowers by three different phage antibodies (Ab^{1}, Ab^{2} and Ab^{3}). All antibodies detect a mild decrease of the JAF13 protein in the mutant as compared to wild type. The same protein extract show no blue staining when no secondary antibody is added. The positive control is the recombinant JAF13 protein used to select the antibodies from a phage library

MYB transcription regulators controlling anthocyanin synthesis

The MYB component of the MBW complex has a special role, because it is a major factor in determining the specificity of the complex and the downstream pathway(s) that is (are) activated (Koes et al., 2005; Ramsay and Glover, 2005; Serna and Martin, 2006). In *P. hybrida* petals, for example, the activation of genes involved in vacuolar acidification requires the MYB protein PH4 (Quattrocchio et al., 2006), whereas activation of anthocyanin synthesis requires a distinct MYB protein encoded by *AN2* or, in other tissues, the paralogs *AN4*, *DPL* or *PHZ* (Quattrocchio et al., 1999; Kroon, 2004; Albert et al., 2011).

Flowers of *P. inflata* and related (sub) species accumulate anthocyanins in both petals and anthers. Because many plant species, including several *Solanaceae*, bear colored petals, this is apparently an ancient trait. However, species with blue or purple colored anthers are rare, suggesting that this trait arose much later during evolution, possibly on multiple independent occasions. Although the loss of anther pigmentation in *P. hybrida* mutants causes no obvious deleterious effects, no such mutants have been reported in natural population of *P. inflata* or related species. This suggests that colored anthers are important for fitness in a natural habitat, presumably because they aid in the attraction of pollinators and thereby reproductive success.

In *P. hybrida* the four genes, *AN2*, *AN4*, *DPL* (also known as *MYBb*) and *PHZ*, encode the MYB component of MBW complexes that activate anthocyanin (Quattrocchio et al., 1998; Quattrocchio et al., 1999; Kroon, 2004; Albert et al., 2011). The proteins AN2, AN4, PHZ and DPL have highly similar sequences and function, but are expressed in distinct, partially overlapping, patterns and control pigmentation in distinct subsets of tissues (Quattrocchio et al., 1999; Kroon, 2004; Albert et al., 2011). *DPL* and *PHZ* are expressed in floral and vegetative tissues and have been implicated in the accumulation of pigment in vegetative tissues (Kroon, 2004; Albert et al., 2011). *AN2* is exclusively expressed in the petal limb and tube and encodes a MYB protein that interacts with AN1 and JAF13 (Quattrocchio et al., 1999; Quattrocchio et al., 2006) and activates all known LBGs and one EBG (*CHSJ*). *an2* null alleles strongly reduce, but do not completely abolish, pigmentation of the petal limb, but have no clear effect on pigmentation of the petal tube, anthers, and vegetative tissues like pedicels (de Vlaming et al., 1984; Quattrocchio et al., 1999). *AN4* encodes a highly similar MYB protein that is essential for the pigmentation of anthers and certain cells in the petal tube, but has no obvious role in pigmentation of the petal limb (de Vlaming et al., 1984; Kroon, 2004; Povero, 2011). *AN4* mRNA is expressed in anthers at early developmental stages to activate transcription of *AN1* and structural anthocyanin genes (LBGs). Curiously, *AN4* is also expressed in the petal limb, but only during late stages of flower development when the expression of *AN1*, *DFR* and other structural anthocyanin

genes has already ceased (Kroon, 2004; Povero, 2011). Because ectopic expression of *AN2* and *AN4* in *P. hybrida* is sufficient for the induction of ectopic expression of *AN1* and, subsequently, downstream structural anthocyanin genes (Spelt et al., 2000; Povero, 2011), it is conceivable that the gain of pigmentation in novel tissues, such as *P. inflata* anthers, originated from alterations in expression of the MYB-type anthocyanin regulators.

We found in the genomes of *P. inflata*S6 and *P. axillaris*N homologs of the *AN2*, *AN4*, *DPL* and *PHZ* from *P. hybrida*, but no additional paralogs. *PiAN4 (P.inflata*S6 *AN4)*, *PiDPL* and *PiPHZ* reside in ~ 270 kb region that was covered by a single scaffold, whereas *PiAN2* was found in a distinct scaffold (Figure 7). The latter was not surprising as *AN2* and *AN4* were previously mapped to two different chromosomes in *P. hybrida* (de Vlaming et al., 1984).

To assess when the gene duplications occurred by which *AN2*, *AN4*, *DPL* and *PHZ* arose, we identified putative homologs in the genomes of other *Solanaceae*. We found that tomato, potato, bell pepper, and eggplant all possess multiple genes with high similarity to *AN2*. Phylogenetic analyses show that these (putative) AN2 homologs often cluster in species-specific clades, indicating that the expansion of this MYB gene (sub)family happened independently in each lineage (Figure 7a). Because *AN4*, *DPL* and *PHZ* are closely related, and more similar to a *Nicotiana benthamiana* homolog than to PiAN2 (Figure 7), it appears that *AN2* and the ancestor of the *AN4-DPL-PHZ* clade existed prior to the divergence of *Petunia* and *Nicotiana*, and that the further expansion of the *AN4*/*DPL*/*PHZ* cluster occurred after the separation from *Nicotiana*. However, similar expansions of this gene family occurred independently in the tomato, eggplant, potato, pepper and *Mimulus* lineages.

The *AN2*/*AN4* homologs of tomato and *Mimulus* are also clustered in the genome (Figure 8). Comparison of the 100-kb genomic region containing the tomato cluster of MYBs with the *P. inflata*S6 scaffolds containing *AN2* (650Kb) and the *AN4*, *DPL PHZ* cluster (over 1Mbp), shows that both are syntenic, although the similarity is limited to one side of the MYB gene cluster. Because the *P. inflata*S6 genomic fragment contains multiple insertions often consisting of several transposons and nearly no genes (Figure 8b), the *AN4-DPL-PHZ* cluster encompasses a much larger region (~300kb) than the tomato cluster (65Kb). All together, the syntenic region extends for 170Kb in tomato and over 600Kb in *P. inflata*S6. The region containing *PiAN2*, by contrast, displays no synteny with the tomato scaffold containing the MYB cluster, but is syntenic with another region of the tomato genome, consistent with the idea that in the Petunia lineage a translocation separated *AN2* from the *AN4-DPL-PHZ* cluster (Figure 8). This translocation is accompanied by the duplication of a fragment and the insertion of several DNA chunks.

After duplication, the roles of *AN4*, *DPL* and *PHZ* as determinants of pigmentation pattern subsequently diverged by *cis*-regulatory changes as they are in *P. hybrida* expressed in different patterns. *DPL* and *PHZ* are expressed in certain cells within petals and after light induction in vegetative tissues (Albert et al., 2011), whereas *AN4* is expressed in anthers of young (stage 1-3) flowers buds, where it is needed to activate transcription of *AN1* and downstream anthocyanin genes, such as *DFR* (Povero, 2011). Given that this apparently happened well after the divergence of *Petunia* from *Solanum* and *Nicotiana* species, it looks like it that the appearance and neo-functionalization were the key by which Petunia acquired colored anthers.

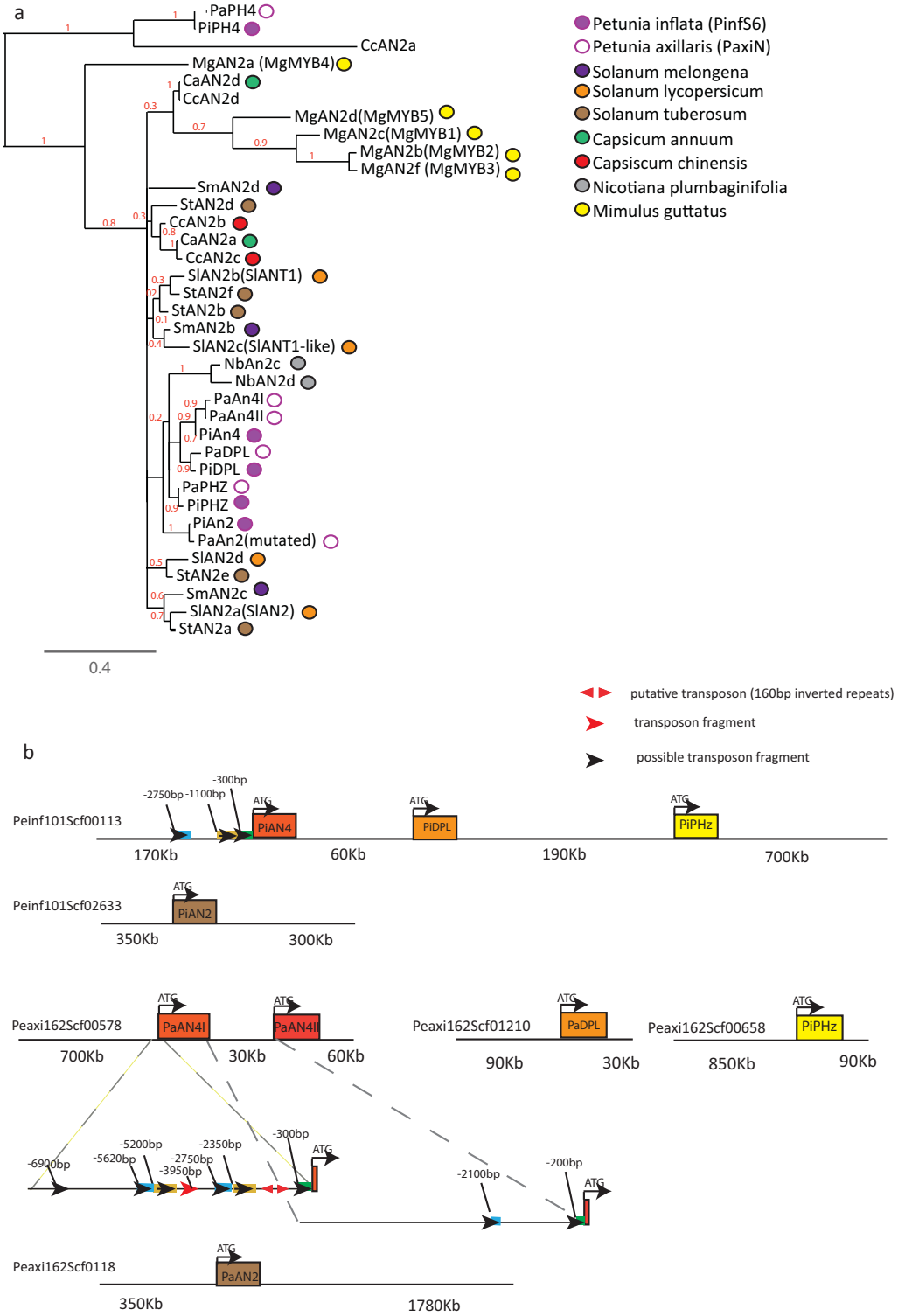


Figure 7. Anthocyanin MYBs in *PinfS6* and *PaxiN.* a) Phylogenetic tree of MYB proteins in different species. The tree was built after alignment with Muscle and G blocks curation. Reliability values are given on base of 400 bootstraps. The *P.axillarisN* An2 (PaAN2) protein sequence used in this analysis was obtained by correcting the mutation that results in a frame shift. b) Map of PinfS6 and PaxiN scaffolds containing AN4, DPL and PHZ. Note that all three *Pinf*S6 alleles map to a single scaffold, whereas the *Paxi*N alleles map to 3 distinct ones. Distances between genes and/or scaffold ends are indicated below the maps. The blowup of the promoter regions of the two *an4* duplicates (*an4* I and *an4* II) shows the position of regions that are similar to the *PiAN4* promoter, and additional transposon-like sequences with 160bp inverted repeats.

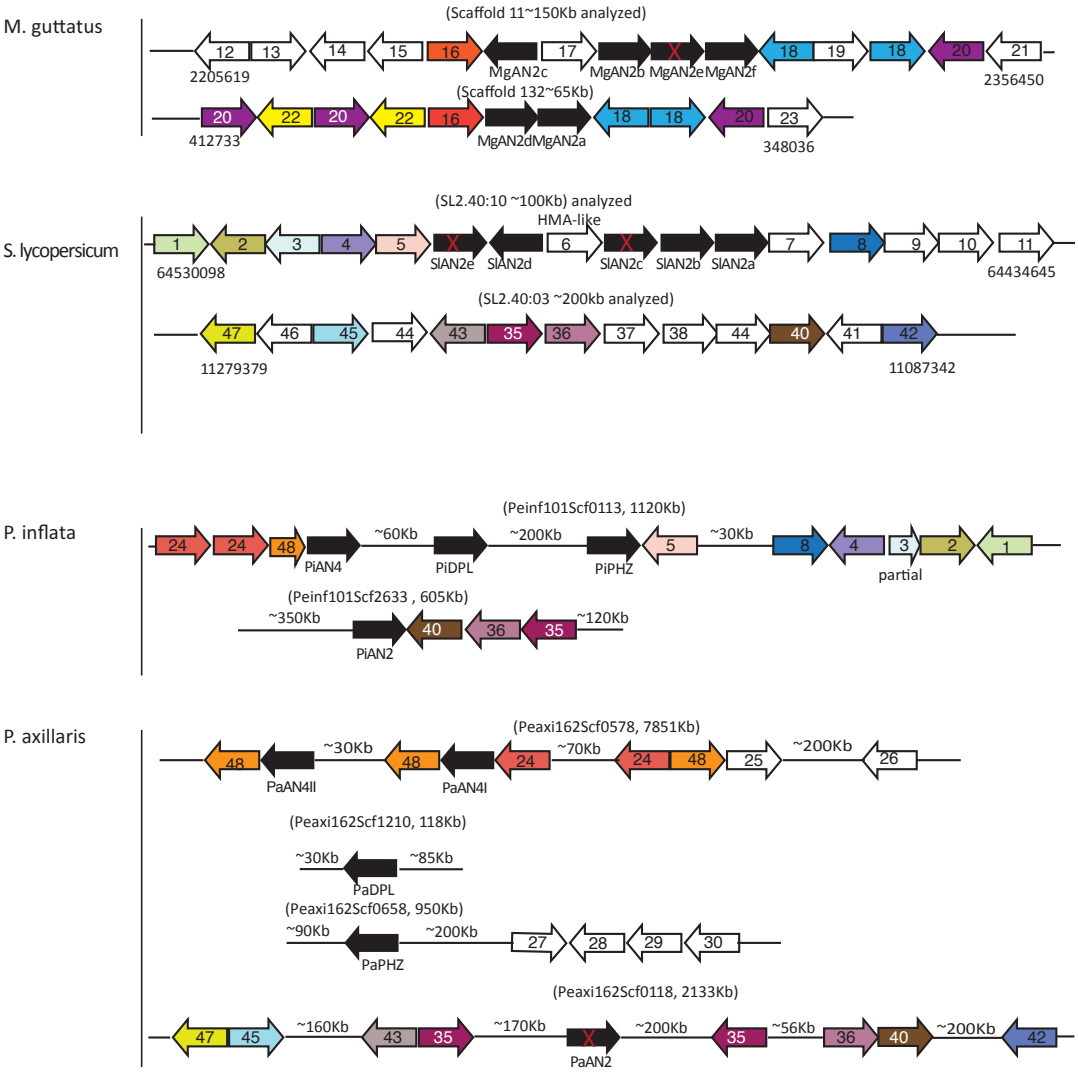


Figure 8. Anthocyanin regulatory genes of the MYB family. Synteny analysis of the genomic regions containing the anthocyanin MYB genes (black arrows) in different species. More details about the genes appearing in this graph are in Tables 1 and 2. For *Mimulus guttatus*, and *Solanum lycopersicum*, fragments of the indicated chromosomes were analyzed (the length and position of the analyzed fragment is given in the figure). For *Petunia axillaris* and *Petunia inflata* complete scaffolds, of the dimensions reported in the figure, were analyzed for the presence of genes homologous to those flanking anthocyanin MYBs in *Mimulus* and tomato. The lines indicate gene-poor regions. Drawings are not to scale. Red crosses indicate non functional genes

Figure 9. Genes encoding non-MYB regulators of anthocyanins and MYBs regulating other pathways in *P. axillaris* and *P. inflata*. Synteny analysis of the genomic regions containing the anthocyanin transcription regulators AN1, JAF13, AN11 and MYBx, the vacuolar pH regulator PH4 and the scent regulators ODO1 and EOB1. On the contrary of what observed for the MYB genes of the anthocyanin pathway, these genes present good syntenic conservation between the two petunia species. Small arrows indicate truncated sequences (this could in some cases due to the position in the scaffold).

Table 1. MYB genes in the phylogenetic and synteny analysis in Figure 8.

Gene	Accession number	Position	Remarks/name in literature
SlAN2e	Solyc10g086300	SL2.50ch10:65177075..65177380	10bp insertion at nt 130 from ATG
SlAn2d	Solyc10g086290	SL2.50ch10:65167328..65168683	
SlAN2c	Solyc10g086270	SL2.50ch10:65,144,661..65,145,610	ANT1-like
SlAN2b	Solyc10g086260	SL2.50ch10:65137652..65138664	ANT1
SlAN2a	Solyc10g086250	SL2.50ch10:65133410..65134967	AN2
MgAN2c	mgv1a023671m	scaffold11:2,239,234..2,241,196	Myb1
MgAN2b	mgv1a024703m	Scaffold 11:2,267,743..2,271,070	Myb2
MgAN2e	mgv1a023318m	Scaffold 11:2,313,496..2,313,867	partial
MgAN2f	mgv1a024996m	Scaffold 11:2,324,046..2,325,041	Myb3
MgAN2d	mgv1a019326m	Scaffold 132:374,122..375,741	Myb5
MgAN2a	mgv1a025765m	Scaffold 132:366,987..368,604	Myba
PiAN2	Peinf101Scf02633g05002	Scaffold 2633:356538…358106	AN2
PiDPL/MYBb	Peinf101Scf00113g02010	Scaffold 113:236820…239718	Deep Purple/ MYBB
PiAN4	Peinf101Scf00113g01010	Scaffold 113:175827…177046	AN4
PiPHZ	Peinf101Scf00113g04004	Scaffold 113:433409…435357	Purple Haze
PaAN2	Peaxi162Scf00118g00310	Scaffold 118:356538…358106	AN2
PaDPL/MYBb	Peaxi162Scf01210g00002	Scaffold 1210:29800…18502	Deep Purple/ MYBB
PaPHZ	Peaxi162Scf00658g00110	Scaffold 658:93094…91155	Purple Haze
PaAN4I	Peaxi162Scf00578g00008	Scaffold 578:89754…88554	AN4
PaAN4II	Peaxi162Scf00578g00007	Scaffold 578:57794…56593	AN4
PaPH4	Peaxi162Scf00349g00057		PH4
PiPH4	Peinf101Scf02429g00002		PH4

Table 2. Genes appearing in the synteny analysis reported in Figure 9.

number	gene	accession number
1	Ras-related small GTP-binding family protein	Solyc10g086350
2	ATP binding microtubule motor family protein	Solyc10g086340
3	RPL23AB ribosomal protein L23AB	Solyc10g086330
4	PIP5K9 phosphatidyl inositol monophosphate 5-kinase	Solyc10g086320
5	Uncharacterized protein (GTP-binding)	Solyc10g086310
6	Heavy metal associated protein (HMA)	Solyc10g086280
7	UDP-Glycosyltransferase superfamily protein	Solyc10g086240
8	Unknown Protein	Solyc10g086230
9	OPR2, ATOPR2 12-oxophytodienoate reductase 2	Solyc10g086220
10	Translin family protein	Solyc10g086210
11	SAUR-like auxin-responsive protein family	Solyc10g086200
12	Serine Hydroxymethyltransferase	mgv1a004699m
13	Molecular chaperone Hsp40	mgv11b021190m
14	AGL21 (AGAMOUS-like 21)	mgv1a021032m
15	Unknown Protein	mgv1a024095m
16	Calcium-dependent lipid-binding protein	mgv1a015143m, mgv1a015009m
17	Unknown Protein	mgv1a024765m
18	EF-hand calcium-binding domain containing protein	mgv1a015576m, mgv1a015814m, mgv1a021228m
19	Fructokinase	mgv1a010551m
20	Leucine-rich repeat receptor-like protein kinase	mgv1a004681m, mgv1a017960m, mgv1a004918m, mgv1a017960m
21	PPR repeat	mgv1a004507m
22	5'-AMP-activated protein kinase , beta subunit	mgv1a025479m, mgv1a024703m
23	HYL1, DRB1 dsRNA-binding domain-like	mgv1a003102m
24	MAT3 (Methionine Adenosyltransferase3)	
25	P-loop containing nucleoside triphosphate hydrolases	
26	pentatricopeptide repeat-containing protein	
27	serine/threonine-protein phosphatase PP1	
28	GDP-mannose-pyrophosphorylase/mannose-1-pyrophosphatase	
29	actin-related protein 5	
30	aspartyl protease family protein	
35	Serine/threonine protein kinase	Solyc03g043710
36	ARMADILLO/BETA-CATENIN repeat family protein	Solyc03g043700
37	Retrovirus-related Pol polyprotein from transposon TNT 1-94	Solyc03g043690
38	zinc ion binding protein	Solyc03g043680
39	Transposon Ty1-BL Gag-Pol polyprotein	Solyc03g043670
40	Serine Protease family	Solyc03g043660
42	Copper Transport Protein ATOX1-related	Solyc03g043640
43	Trimeric coiled-coil oligomerisation domain of matrilin	Solyc03g043720
44	Unknown protein	Solyc03g043730
45	Hydroxyproline-rich glycoprotein	Solyc03g043740
46	Pyridine nucleotide-disulphide oxidoreductase	Solyc03g043750
47	SPT2 chromatin protein	Solyc03g043760
48	transposon	
49	Zeta toxin	Solyc06g053810
50	40S ribosomal protein S15	Solyc06g053820
51	AUX/IAA family	Solyc06g053830
52	AUX/IAA family	Solyc06g053840

Mutations in MYB genes that reduced pigmentation in *P. axillaris* flowers

To understand how the pigmentation patterns of *P. inflata* and *P. axillaris* diverged, in particular the pigmentation of anthers and petal limbs, we analyzed the *P. axillari*s alleles of all known regulatory (this section) and structural anthocyanin genes (next section). The *P. axillaris* alleles of *AN1*, *JAF13* and *AN11* lack obvious defects, such as mutations that disrupt the coding sequence or splice sites. This was not surprising because *P. axillaris* can synthesize anthocyanins in several tissues, such as the petal limb, pedicels, and seeds, which is known to be AN1- and AN11-dependent. Consistent with previous data on other *P. axillaris* accessions (Quattrocchio et al., 1999), we found that the coding sequence of *PaAN2* is interrupted by a frameshift mutation that resembles a transposon footprint. This inactivation of *AN2* was a major cause for the loss of anthocyanins in *P. axillaris* petal limbs. Interestingly, we found that the 2-MB genomic region surrounding *PaAN2* was heavily rearranged as compared to the *P inflata* region and had surprising few genes in common (Figure 8).

The *P. axillaris* genome contains two copies of *AN4*, designated *PaAN4I* and *PaAN4II*, that are separated by 30 kb (Figures 7 and 8) and apparently arose by a gene duplication in the *P. axillaris* lineage, because *PaAN4I* is more similar to *PaAN4II* (4 SNPs) than to *PiAN4* (16 SNPs). The *P. axillaris* paralogs *PaDPL* and *PaPHZ* are in two distinct scaffolds of 118 Kb and 950 Kb respectively, which have no overlap with the 7815 Kb scaffold containing *PaAN4I* and *PaAN4II*. Hence, the clustering of *AN4*, *DPL* and *PHZ* seen in *P. inflata,* was reduced or lost in *P. axillari*s lineage, either through large insertions in the intergenic regions, or by translocations. Moreover, the genomic regions surrounding the *PaDPL*, *PaPHZ* and the two *PaAN4 copies*, which cover nearly 8 Mb, show hardly any similarity with those surrounding the *P inflata* homologs (Figure 8). The only exception is a duplicated *MAT3* gene (gene 24 in Figure 8). The two copies of this gene are positioned next to *PiAN4* and also next to *PaAN4I*, although separated from it by a 10-Kb transposon-like insertion. Also between the two MAT3 copies we found a large insertion (~30Kb) of gene poor sequence. This indicates that this genomic region was subject to massive rearrangements, since the separation of the *P. inflata* and *P. axillaris* lineages.

In contrast, analysis of the genomic regions containing other genes encoding anthocyanin regulators (*AN1*, *AN11*, *JAF13*) revealed good synteny conservation between *P. inflata* and *P. axillaris*. The same conclusions come from the analysis of the genomic fragments containing other MYBs regulating vacuolar pH (*PH4*), scent emission (*ODO1* and *EOB1*) and the homologue of the Arabidopsis *Caprice* gene *MYBx* (Kroon, 2004) (Fig.9).

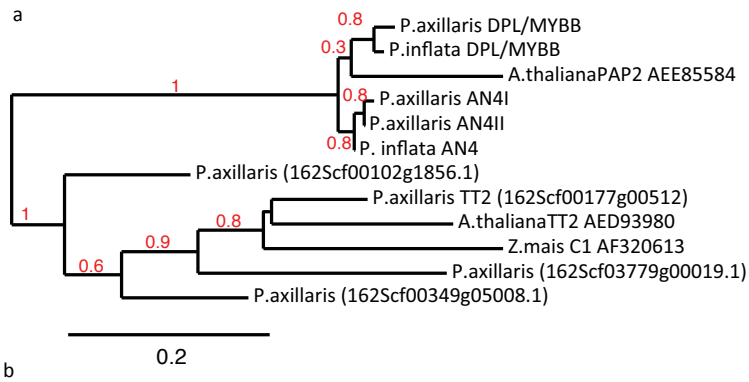
Closer examination of the *PaxiN AN4I* allele shows that it contains multiple rearrangements in the 5' flanking region, which apparently originate from the insertion of a transposon followed by an inverted duplication of part of the transposon and *AN4* promoter sequences (Figure 7). This rearrangement is very similar, if not identical, to that found in recessive (inactive) *an4* alleles of *P. hybrida* (Povero, 2011). This indicates that these rearrangements caused the loss of anthocyanin pigmentation in *P. axillaris* anthers, and that the aforementioned inactive *P. hybrida an4* alleles are introgressions from the parental *P. axillaris* accessions.

Regulation of proanthocyanidin accumulation in petunia seeds

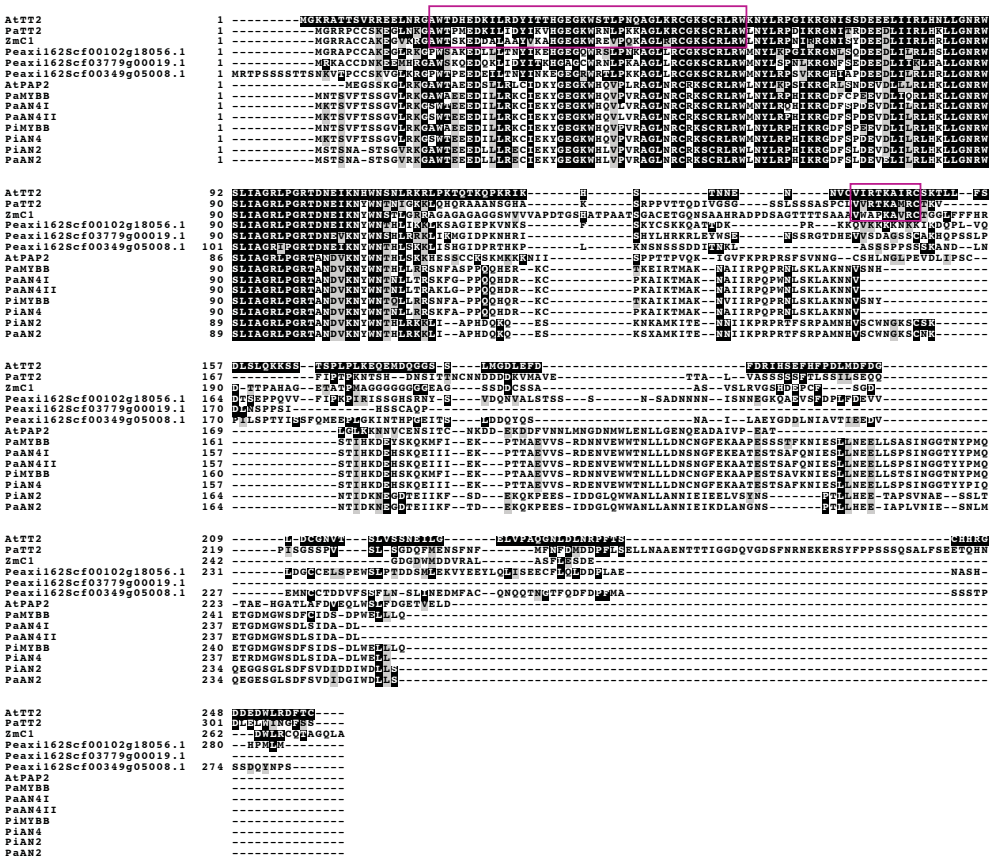
Mutations in *AN2* and *AN4* do not affect the coloration of the seed coat, while *an1* and *an11* seeds are paler than those of wild type plants (Zenoni et al., 2011). Furthermore, *35S:AN4-RNAi* and *35S:MYBB-EAR* constructs cause loss of anthocyanin pigmentation in anthers and tube of the flower, but do not have effect on the seed coat phenotype (Povero, 2011). Although this last might be (in part) due to the poor activity of the CaMV35S promoter in the seed coat, all these observations together strongly suggest that another MYB controls proanthocyanidin accumulation in the seeds together with AN1 and AN11.

Figure 10a shows that the petunia genome (*PaxiN*) encodes 4 MYB proteins with high similarity to TRANSPARENT TESTA 2 (TT2), which is required for tannin accumulation in Arabidopsis seeds. In a phylogenetic tree, these 4 proteins group together with AtTT2 in a clade that is distinct from the one containing AN2, AN4 and AtPAP2, all MYBs involved in anthocyanin deposition. One of the proteins (Peaxi162Scf00177g05011.1) clearly shows higher similarity to AtTT2 than the other three (this is why we named it PaTT2). We can however, not exclude that (some of) the other proteins are functionally redundant with PaTT2.

None of the petunia TT2 related proteins contains the DExWRLxxT motif that was proposed to be characteristic for this class of MYBS (Kranz et al., 1998). Instead, by running MEME (Bailey and Elkan, 1994), we found a different motif (VxRTKAxRC) that is present in AtTT2 and PaTT2 outside the R2R3 repeats, and which is shared (although less conserved) with the ZmC1 protein that according to the MYB inventory of Kranz and colleagues groups in the same class of TT2. Another motif recognized by MEME is present within the R2 repeat and essentially consists of a few residues conserved among AtTT2, PaTT2 and ZmC1. The two motifs shared by AtTT2, PaTT2 and ZmC1 (Figure 8 C) are not present in any of the other petunia MYBs in the clade of TT2. The isolation of mutants will be necessary to better understand if these factors indeed regulate proanthocyanidin accumulation in the seed coat of petunia.



b



c

```
PaTT2   19    KEGLNKGAWTPMEDKILIDYIKVHGEGKWRNLPKKAGLKRCGKSCRLRWLNYLRPDIKR
AtTT2   21    REELNRGAWTDHEDKILRDYITTHGEGKWSTLPNQAGLKRCGKSCRLRWKNYLRPGIKR
ZmC1    19    KEGVKRGAWTSKEDDALAAYVKAHGEGKWREVPQKAGLRRCGKSCRLRWLNYLRPNIRR

PaTT2   155   SLSSSASPCLVVRTKAMRCTKVFIPTPKN
AtTT2   141   HSTNNENNVCVIRTKAIRCSKTLLFSDLS
ZmC1    172   AGTTTTSAAAVWAPKAVRCTGGLFFFHRD
```

Figure 10. Putative homologs of AtTT2 encoded by the *PaxiN* genome.

a) Four proteins group together with AtTT2, but the protein we named PaTT2 is more closely related to it than the other 3 members of this small group of MYBs.

b) Alignment of the proteins in the phylogenetic tree in a). Shade was given to any residue position identical to the AtTT2 sequence. (**c**) Motifs common to AtTT2, the suspected petunia homologue (PaTT2) and the maize C1 protein, which according to a classification of MYB proteins (Kranz et al., 1998). based on phylogeny, is part of group 5 like TT2.

Mutations in structural anthocyanin genes that altered the pigmentation of petals and anthers in *P. axillaris.*

*P. axillaris*N flowers synthesize in the petal tube and in vegetative tissues (pedicel, light-induced leaves) anthocyanins with a similar glycosylation and acylation pattern as *P. inflata*S6 flowers. Hence, it is not surprising that we found intact alleles for *DFR*, *ANS*, *3GT*, *5GT* in the *PaxiN* genome (Figure 1). However, *P. axillari*s contains mutant alleles for several other structural genes, which compromise anthocyanin methylation (*mt*, *mf1*, *mf2*) or the pigmentation of specific tissues (*hf1*, *po*/*chia*).

The paralogous genes *MF1* and *MF2,* which encode A3'5'MT, are both inactivated in *P. axillaris*N by multiple insertions and deletions (Provenzano et al., 2014). The inactive *mf1* alleles of *PaxiN* and two additional *P. axillaris*N accessions (S1 and S2) share a 22-bp deletion in exon 6, an insertion of a ~7-kb retrotransposon–like element in intron 2, and a 1421-bp insertion in intron 3, compared to the functional *MF1* alleles from *P. hybrida* and *PinS6*, suggesting that (one of) these mutations caused the loss of anthocyanin 5' methylation in *P. axillaris* after the separation from *P. inflata*. The *mf* alleles of the *P. axillaris* accessions N, S1 and S2 contain several additional deletions that differ between *mf1* alleles, suggesting that these arose after the gene has been inactivated in distinct *P. axillaris* populations (Provenzano et al., 2014). *PaxiN* also contains a mutant allele at *MT*, with a double transposon insertion in the last intron, similar to (introgressed) *P. hybrida mt1* alleles. However, some other *P. axillaris* accessions (line S2) contain a functional *MT* allele lacking this insertion, suggesting that the recessive *mt* allele arose relatively recently in *P. axillaris* and is not yet fully fixed in this species (Provenzano et al., 2014).

Genetic data suggested that *P. axillaris* accessions harbor a peculiar allele at *HF1*, designated *hf1-1,* that has a reduced phenotypic expression in the limb (Wijsman, 1983; de Vlaming et al., 1984). The *hf1* allele in the *PaxiN* genome contains three insertions in the second intron, similar to the *hf1* alleles of several related petunia species and *P. hybrida* lines (Chen et al., 2007). To examine whether these insertions affect the expression and activity of *HF1*, we performed RT-PCR analyses, using a primer complementary to the second exon in combination with an oligo dT primer, and found that *Paxi*N petals express a mutant *hf1* transcript that is polyadenylated within the first insertion and consequently lacks the last exon (Figure 11). The truncated protein encoded by this mRNA apparently has some remaining activity, as *hf1-1* is not a complete null allele (de Vlaming et al., 1984). The *HF2* gene of PaxiN does not present any mutation in the coding sequence, suggesting that *PaHF2* is a wild type allele, however, we did not check its expression.

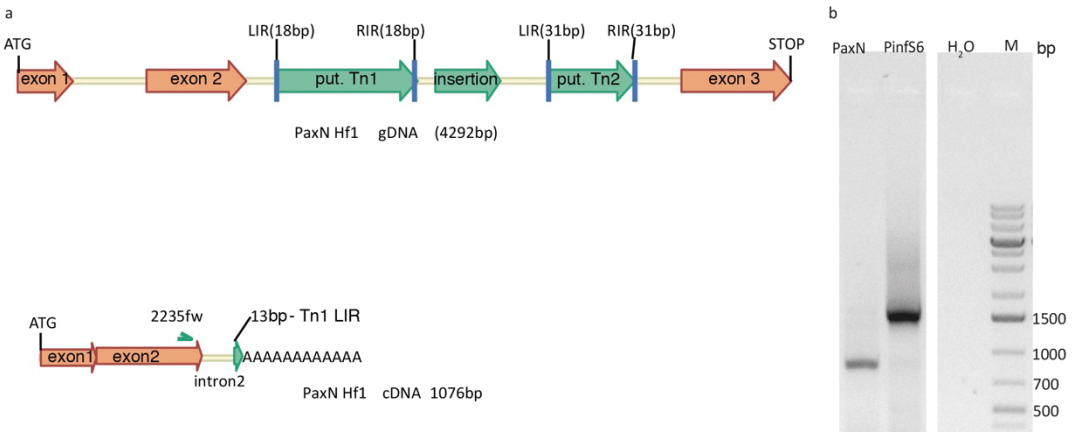


Figure 11. The *hf1-1* allele of *PaxiN*. In a) a scheme of the genomic fragment containing the coding sequence of the *HF1* gene in *PaxiN*. Intron II carries three insertions of which two have the characteristics of transpons. The putative Tn1 (~800bp) has inverted repeats of 18bp (the LIR sequence presents 3 mismatches with the RIR) and putative Tn2 (~500bp) has inverted repeats of 31bp (3 mismatches between LIR and RIR). The smaller insertion in the middle (~200bp) does not have the characteristics of a transposon. Petal cDNA from *PinfS6* and *PaxiN* was amplified in a RACE reaction using the primer #2235fw (CTCCAATCGTCCACCTAATGCA) and an oligo dT primer, the resulting products are shown in b). The RACE product from the *PaxiN* cDNA is much shorter than the one obtained from *PinfS6*. The sequence of the amplification product shows poly-adenylation in the putative Tn1 (immediately after the 13bp LIR).

In *P. hybrida* mutations in the partially redundant *HF1 and HF2* genes reduce the synthesis of the violet-colored petunidin and malvidin derivatives and results in the accumulation of magenta-colored peonidins. In an *fl* background, which synthesizes small amounts of flavonols, *hf1 hf2* petals accumulate abundant amounts of peonidins, resulting in a bright magenta color. However in a genetic background (*FL*) that synthesizes much higher amounts of flavonols, such as *P. axillaris*N*, hf1* severely reduces anthocyanin synthesis, presumably due to competition for the dihydroquercetin substrate, resulting in a very pale magenta petal color. This suggests that the occurrence of *hf1-1* allele in *P. axillaris*N also contributed to the reduced pigmentation of the petal limb. For unknown reasons *P. hybrida* anthers cannot synthesize cyanidin or peonidin derivatives, and consequently mutation of *HF1* blocks anthocyanin synthesis in anthers (de Vlaming et al., 1984). Hence, *hf1-1* may have contributed to the loss of anthocyanins in *P. axillaris*N anthers as well.

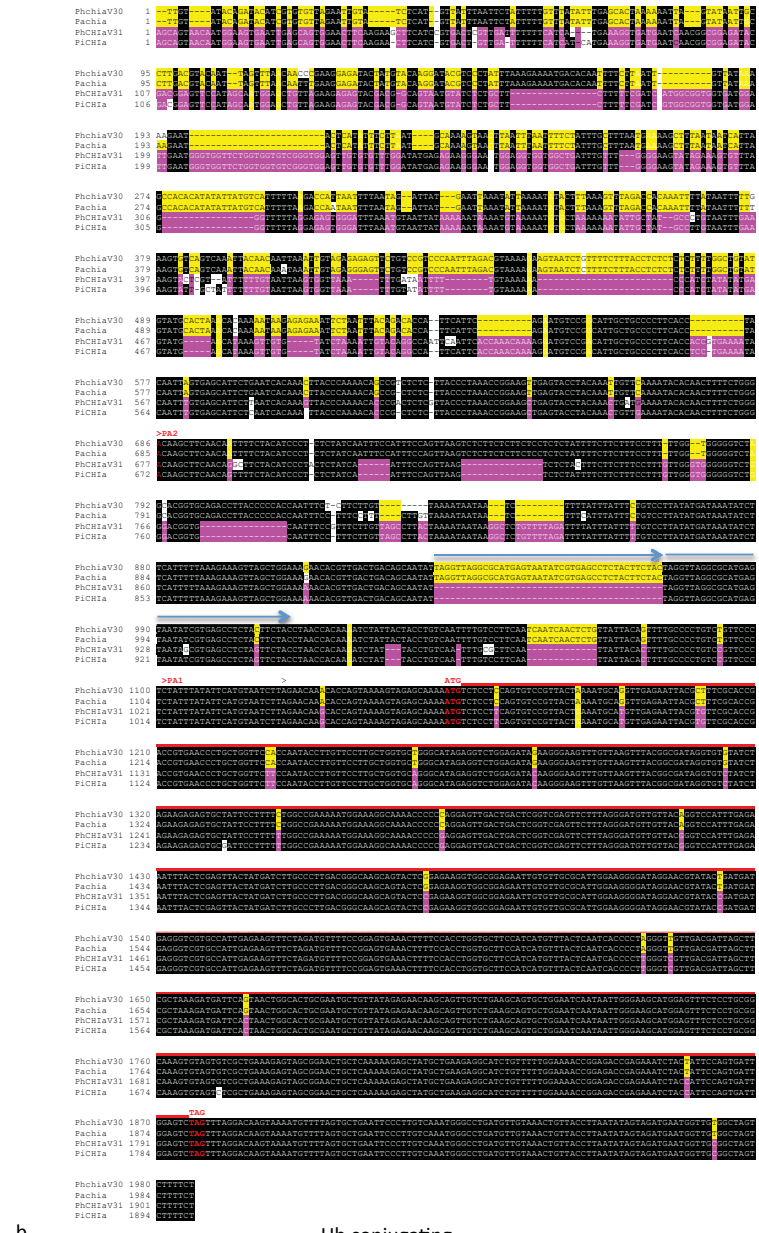
The color of *P. hybrida* anthers is determined in part by the *POLLEN* (*PO*) locus (de Vlaming et al., 1984), which contains *CHIa* (van Tunen et al., 1991). In *P. hybrida* lines with a dominant *PO* allele, *CHIa* is expressed in both petals and anthers, whereas lines containing the recessive *po* allele express *CHIa* only in petals, but not in anthers. The divergent expression pattern of the *P. hybrida PO* and *po* alleles results from differences in cis-regulatory elements (van Tunen et al., 1991). We found that the *CHIa* allele of *P. inflata* is highly similar to the dominant *CHIa* alleles *PO of P. hybrida* lines, whereas the *CHIa* allele of *P. axillaris* is most similar to the recessive *chia* alleles of *po* lines (Figure 11). This indicates that the dominant and recessive *PO* alleles of *P. hybrida* arose in nature and were introgressed from *P. inflata* and *P. axillaris* accessions respectively.

In the absence of CHI enzyme activity part of the yellow chalcone substrate accumulates, presumably as a glycoside, and part isomerizes spontaneously and (one of the

stereoisomers) is converted into anthocyanin or flavonol end products. Hence, anthers of *PO AN4* genotypes, like *P. inflata,* have a blue color, whereas the anthers and pollen of *po/po AN4* genotypes are greenish (mix of yellow and blue pigments). In combination with homozygous *an4*, the recessive *po* allele specifies a yellow anther color resulting from the accumulation of yellow chalcones and colorless dihydroflavonols. Hence the bright yellow color of *P. axilaris* anthers and pollen is caused by the combined effect of the *po* allele and mutations that block anthocyanin synthesis (*an4*, *hf1-1*).

To address whether the distinct pigmentation of *P. inflata* and *P. axillaris* anthers resulted from a gain of function mutation in the *P. inflata* lineage or a loss of function in the *P. axillaris* lineage we compared the upstream sequences of the $CHIA^{PO}$ allele of *PinS6* and $chiA^{po}$ from *PaxiN* (Figure 12). The dominant *P. hybrida CHIa* gene can be expressed from two distinct, differentially regulated promoters. The proximal PA1 promoter of the dominant *CHIa* allele of *PO* lines is active in both petals and young anthers and responsible for expression of CHI enzyme activity, whereas PA1 of the recessive *chia* allele is only expressed in petals, but not in anthers (van Tunen et al., 1988; van Tunen et al., 1991). The function of the distal PA2 promoter is unclear. PA2 of both the *chia* and the *CHIa* allele is active in pollen during late stages of anther development, though the longer mRNA that is produced does not result in CHI enzyme activity. The 600 bp region upstream the PA1 transcription start is highly similar in the *chia* and *CHIA* alleles, and only small indels are seen. The largest of these is a ~40nt direct repeat, which presumably resulted from a duplication in the *PaxiN* lineage, because this sequence occurs only once in *PinS6*. The region upstream -600 diverged completely in the *PaxiN* and *PinS6* alleles (figure 11a). The gene region upstream the *CHI* coding sequence of the PinS6 and PaxiN allele displays clear similarity (synteny) over an extended region and harbors homologous genes, but contains several big indels (and numerous smaller ones) because of which the intergenic regions differ substantially between *PaxiN* and *PinS6* (Figure 11b). The sequence divergence between the *chia* and *CHIa* alleles upstream from -600 apparently is due to the insertion of an ~4775 bp sequence in the *PinS6 CHIa* allele. This sequence is moderately repetitive in the *PaxiN* and *PinS6* genomes, but lacks obvious features of transposons, suggesting that this polymorphism is caused by the insertion of this element in the *CHIa PInfS6* allele (and possibly a gain of CHI expression in *P. inflata* anthers) rather than a deletion in *PaxiN* (and a loss of CHI expression in *P. axillaris* anthers).

a



b

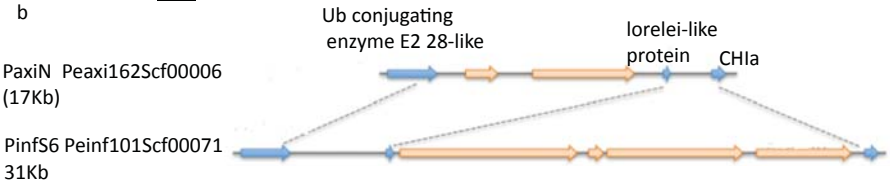


Figure 12. Comparison of CHIa/CHIb alleles in *PinS6* and *PaxiN*. a**)** Alignment of *CHIa* alleles of *PaxiN* and *PinS6* and *P. hybrida* lines harboring either a dominant (*PO*) *CHIa*^{V30} allele (line V31) or the recessive (*po*) *chia*^{V31} allele. PA1 denotes the transcription start of an mRNA expressed in young anthers and PA2transcription start of an alternative promoter that is active in pollen during late stage of anther development

b) Diagram of genomic PaxiN and Pinf6 regions upstream of CHIa. Blue arrows denote the position and orientation of CHIa and two upstream genes. Beige arrows denote larger insertion/deletions that differ between *PaxiN* and *Pinf*.

Genes controlling vacuolar pH

In *P. hybrida* a violet petal color requires the accumulation of malvidin or petunidin derivatives, in combination with the hyper-acidification of vacuoles in petal cells which is controlled by seven *PH* genes (*PH1* to *PH7*). Given that *P. inflata* petals have a similar violet color, it is not surprising that the *PinS6* genome contains intact copies of *PH1*, *PH3*, *PH4* and *PH5*.

ph mutants arise frequently In *P. hybrida* populations and are in greenhouse and breeding conditions (which includes reproduction by hand pollination) perfectly viable. However, such mutants are not found in populations of *P. inflata* and related species in natural habitats, suggesting that *PH* genes confer a strong selective advantage in nature. Strikingly, we found that *P. axillaris* contains, despite its white petal color, also functional alleles of these *PH* genes. In natural habitats *ph* mutations may be counter-selected in *P. inflata* because blue colored mutant flowers attract less visits of the pollinators. However, this cannot explain the conservation of PH genes in white flowering *P. axillaris* species. In *P. hybrida PH5* and *PH4* are also required for the accumulation of tannins in seeds (Verweij et al., 2008), which may affect seed viability and could explain the conservation of functional alleles also in *P. axillaris*. However, *PH1* is not essential for tannin accumulation (Faraco et al., 2014), suggesting that yet other functions in other cell types for this vacuolar hyper-acidification machinery exist, for which also PH1 function is conserved in *P. axillaris*.

Discussion

The divergent "pollination syndromes" of *P. axillaris* and *P. inflata* species are a major factor determining their genetic separation. Flower color is an important component of the pollination syndrome as flowers with distinct colors have been shown to attract distinct pollinating species (Galliot et al., 2006; Hernandez-Garcia et al., 2010). Although (field) experiments with genotypes that differ in flower shape, scent, or color may help to identify potential reproductive isolation barriers that operate today (Grandin et al., 1998; Hoballah et al., 2007; Hernandez-Garcia et al., 2010; Klahre et al., 2011), they cannot tell whether such genetic alterations were actually important in the evolutionary past unless accompanied by genome-wide analysis of the underlying genetic polymorphisms.

Because *P. inflata* and *P. axillaris* are closely related species with very different pigmentation patterns, they offer a unique possibility to analyze the primary genetic changes associated wit the pollinator shift, without being swamped by secondary changes that continue to accumulate after the genetic separation.

Our data show that several genes involved in the synthesis of anthocyanin pigments or their common flavonoid and phenylpropanoid precursors were duplicated in either the *P. axillaris* or the *P. inflata* lineage. In genome studies increases in the numbers of certain (classes of) genes is often taken as an indication for the importance of those genes in the evolution of certain traits of the species and genome that was studied (e.g. (Perry et al., 2007) and (Denoeud et al., 2014). We suggest that such reasoning should be used with

caution, because at least for anthocyanin genes, for which a substantial amount of functional data are available, we found that gene numbers correlate poorly if at all, with the products (pigments) that are made. For example the duplication of *CHSa* genes in *P. inflata* has no clear effect on flower pigmentation when introgressed in *P. hybrida*. As decades of genetic analyses using a wide variety of *P. hybrida* lines and cultivars did not identify flower pigmentation loci associated with phenylpropanoid synthesis, it is unlikely that the duplication of genes from the general phenylpropanoid metabolism have a clear effect on flower pigmentation. Moreover, the duplication of *AN4* in *P. axillaris* is associated with reduced pigmentation of anthers.

Instead, the evolution of flower pigmentation in *Petunia* relied apparently mostly on expansion and subsequent alterations in a gene (sub)family encoding MYB transcription activators. We infer that the ancestral *Petunia* species, from which *P. inflata* and *P. axillaris* originate, resembled *P. inflata* in that it had flowers with pigmented anthers and petals and that the uncolored anthers and petal limbs of *P. axillaris* are derived traits resulting from loss of function mutations in *AN2*, *AN4* and *HF1*.

Most Angiosperm species have anthers and pollen with either a white (no pigment) or a bright yellow color caused by the accumulation of carotenoids (Endress, 1996; Willmer, 2011). The yellow color of *P. axillaris* anthers and pollen, by contrast, results from flavonoid pigments, in particular chalcones, (de Vlaming et al., 1984; van Tunen et al., 1991; Napoli et al., 1999), similar to yellow anthers and pollen from non-angiosperms and spores from ferns and mosses (Willmer, 2011). Purple or blue colored anthers and pollen are more rare but found in a range of Angiosperm lineages, suggesting that these traits evolved multiple times. Phylogenetic data indicate that Petunia gain of blue anther color by an expansion of the family of AN2 paralogs and the subfunctionalization of AN4 by a change in expression pattern, presumably by alterations in cis-regulatory elements. The appearance of PHZ and DPL presumably contributed to novel pigmentation patterns, primarily in vegetative tissues. It is striking that similar expansions occurred independently in other lineages of the *Solanaceae*, suggesting that these MYB genes were a major player in the divergence of pigmentation patterns and gain of pigmentation on new sites.

The analysis of the genomic regions containing the anthocyanin MYB genes reveals a great plasticity that is not shared by the regions containing other regulators of the anthocyanin pathway or MYBs involved in the control of other pathways.

The findings that (i) these MYB genes reside in a relatively variable region that is subject to frequent rearrangements and (ii) that they can activate transcription of their bHLH partners (Spelt et al., 2002; Povero, 2011) explain the large variation seen in pigmentation patterns of Angiosperm species.

Material and Methods

Mining of data, phylogeny and synteny analysis.

To obtain sequence data from *Solanaceae* and other species we have mined the NCBI (http://www.ncbi.nlm.nih.gov/), Phytozome (http://www.phytozome.net/) and Sol Genomics Network (http://solgenomics.net/) databases. In many cases we have manually improved the annotation of genes, so the sequences corresponding to the accession number that we report are sometimes not completely correct in the databases.

Table 3. Accession numbers of petunia genes mentioned in this work.

gene	PaxiN	PinfS6
4CL a	Peinf101Scf01633g10028	Peinf101Scf01633g10028
4CL b	Peaxi162Scf00195g01223	Peinf101Scf01099g07012
4CL c	Peaxi162Scf00610g00346	Peinf101Scf01230g00016
4CL d	Peaxi162Scf00207g00334	Peinf101Scf01889g04030
AN1	Peaxi162Scf00338g00912	Peinf101Scf00382g12002
AN2	Peaxi162Scf00118g00310.1	Peinf101Scf02633g05002
AN4 I	Peaxi162Scf00578g00008	Peinf101Scf00113g01010
AN4 II	Peaxi162Scf00578g00007	-
AN9	Peaxi162Scf00713g00038	Peinf101Scf00861g02014
AN11	Peaxi162Scf00912g00146	Peinf101Scf01482g12034
ANS	Peaxi162Scf00620g00533	Peinf101Scf01166g06033
C4H a	Peaxi162Scf00556g00035	Peinf101Scf00951g08008
C4H b	Peaxi162Scf00390g00225	Peinf101Scf03806g00033
CHI a	Peaxi162Scf00006g00088	Peinf101Scf00071g13014
CHI bI	Peaxi162Scf00038g01957	Peinf101Scf03176g00016
CHI bII	-	Peinf101Scf06002g00045
CHS a I	Peaxi162Scf00047g01225	Peinf101Scf01775g10041
CHS a II	-	Peinf101Scf01775g10050
CHS j	Peaxi162Scf00536g00092	Peinf101Scf00437g05004
DFR	Peaxi162Scf00366g00630	Peinf101Scf00073g04027
DPL/MYBB	Peaxi162Scf01210g00002	Peinf101Scf00113g02010
F3H	Peaxi162Scf00328g01214	Peinf101Scf01063g07016
FLS	Peaxi162Scf00927g00035	Peinf101Scf00491g01007
3GT I	Peaxi162Scf00163g00081	Peinf101Scf00192g01001
3GT II		Peinf101Scf00086g03003
5GT	Peaxi162Scf00378g00113	Peinf101Scf00459g02012
HF1	Peaxi162Scf00150g00218	Peinf101Scf00872g01009
HF2 I	Peaxi162Scf00108g00417	Peinf101Scf00586g11035.2
HF2 II	Peaxi162Scf00108g00417.2	Peinf101Scf00586g11036.1
HT1	Peaxi162Scf00201g00243	Peinf101Scf01556g04023
JAF13	Peaxi162Scf00119g00942	Peinf101Scf00962g10037
MF1	Peaxi162Scf00089g00427.1	Peinf101Scf02008g02013
MF2	Peaxi162Scf00316g00055	Peinf101Scf02093g03016
MYB27	Peaxi162Scf03779g00019	Peinf101Scf00522g02001.1
MYBX	Peaxi162Scf00521g00814	Peinf101Scf00877g03014.1
MT	Peaxi162Scf00518g00430	Peinf101Scf00400g05023
ODO1	Peaxi162Scf00002g00037	Peinf101Scf00284g00012
PAL a	Peaxi162Scf00858g00215	Peinf101Scf01985g02016

PAl b	Peaxi162Scf00488g00074	Peinf101Scf00622g01006
PAl c	Peaxi162Scf00123g00096	Peinf101Scf00871g09016
PHZ	Peaxi162Scf00658g00110	Peinf101Scf00113g04004
PH1	Peaxi162Scf00569g00024.1	Peinf101Scf01822g03014.1
PH4	Peaxi162Scf00349g00057	Peinf101Scf02429g00002
PH5	Peaxi162Scf00177g00620	Peinf101Scf00968g02017
RT	Peaxi162Scf00487g00064	Peinf101Scf00778g02002
TT2	Peaxi162Scf00177g00512	Peinf101Scf01252g02023

Phylogenetic trees were built from G-box cured alignments (MUSCLE) and phylogeny analysis by Maximum Likehood with the online tool of Philogeny.fr (http://phylogeny.lirmm.fr/phylo_cgi/index.cgi). For all trees branches support is calculated on the basis of 500 bootstrap.

The accession number of petunia genes mentioned in this paper are in table 3.

Plant material and transgenic production

All petunia lines and mutants mentioned in this work come from the Amsterdam mutant collection. The plants were grown at standard greenhouse conditions and transformants where obtained by Agrobacterium mediated leaf disk transformation.

ELISA assay with phage antibodies

The phage isolation of antibodies against the recombinant JAF13 protein from a library of of human scFv fragments and the ELISA quantification of the protein in plant extracts were done according to a procedure previously described (de Bruin et al., 1999).

Anthocyanin quantification and TLC analysis

Anthocyanins were extracted for 30 minutes (under agitation) from punches of petal tissue using MeOH containing HCl 0.05%. The absorbance of the solution was then determined by absorbance at 530nm. TLC analysis of petal anthocyanins and flavonols was performed as previously described (van Houwelingen et al., 1998).

Real Time PCR

Expression of FLS and HT1 in flowers of different *P. hybrid*a lines and in *P. axillaris* and *P. inflata* was determined by Real Time PCR using the following primers:

#5651	qPCR_SAND_Fw	CTTACGACGAGTTCAGATGCC
#5652	qPCR_SAND_Rv	TAAGTCCTCAACACGCATGC
#6862	FLSqPCR-FW	AGGATGCGAGAAGTTGTAGACAG
#6883	FLSqPCR-RV	ACAACACCAAGAGCCAAATCGG
#6881	Ht1qPCR-FW	GCCGCTAAAATGAAGAAGCTCC
#6880	Ht1qPCR-RV	TTTCCCTCCATCATTATCCGCATC

PCR amplification was performed with a ECO P1180-Illumina and the kit I-Taq Universal Syber Green (BioRad catalogue number 172-5124). The results were analyzed with the Software Eco version 4.0.

Acknowledgments

V.P. was supported by a fellowship of the "Accademia Nazionale dei Lincei per Genetica Agraria, fondazione Valeria Vincenzo Landi".

References

Abe, H., Kobayashi, M., Urao, T., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2005). Functional analyses of Arabidopsis transcription factors, AtMYC2/RD22BP1 and AtMYB2. Plant and Cell Physiology **46,** S214-S214.

Albert, N.W., Lewis, D.H., Zhang, H., Schwinn, K.E., Jameson, P.E., and Davies, K.M. (2011). Members of an R2R3-MYB transcription factor family in Petunia are developmentally and environmentally regulated to control complex floral and vegetative pigmentation patterning. The Plant journal : for cell and molecular biology **65,** 771-784.

Alfenito, M.R., Souer, E., Goodman, C.D., Buell, R., Mol, J., Koes, R., and Walbot, V. (1998). Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. Plant Cell **10,** 1135-1149.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol **2,** 28-36.

Baumann, K., Perez-Rodriguez, M., Bradley, D., Venail, J., Bailey, P., Jin, H., Koes, R., Roberts, K., and Martin, C. (2007). Control of cell and petal morphogenesis by R2R3 MYB transcription factors. Development **134,** 1691-1701.

Brugliera, F., Barri-Rewell, G., Holton, T.A., and Mason, J.G. (1999). Isolation and characterization of a flavonoid 3'-hydroxylase cDNA clone corresponding to the Ht1 locus of Petunia hybrida. The Plant journal : for cell and molecular biology **19,** 441-451.

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.

Chan, H., Babayan, V., Blyumin, E., Gandhi, C., Hak, K., Harake, D., Kumar, K., Lee, P., Li, T.T., Liu, H.Y., Lo, T.C., Meyer, C.J., Stanford, S., Zamora, K.S., and Saier, M.H., Jr. (2010). The p-type ATPase superfamily. J Mol Microbiol Biotechnol **19,** 5-104.

Chen, S., Matsubara, K., Omori, T., Kokubun, H., Kodama, H., Watanabe, H., Hashimoto, G., Marchesi, E., Bullrich, L., and Ando, T. (2007). Phylogenetic analysis of the genus Petunia (Solanaceae) based on the sequence of the Hf1 gene. J Plant Res **120,** 385-397.

de Bruin, R., Spelt, K., Mol, J., Koes, R., and Quattrocchio, F. (1999). Selection of high-affinity phage antibodies from phage display libraries. Nature Biotechnology **17,** 397-399.

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.

de Vetten, N., ter Horst, J., van Schaik, H.-P., den Boer, B., Mol, J., and Koes, R. (1999). A cytochome b5 is required for full acivity of flavonoid 3'5'-hydroxylase, a cytochrome P450 involved in the formation of blue flower colors. Proc. Natl. Acad. Sci. USA. **96,** 778-783.

de Vlaming, P., Schram, A.W., and Wiering, H. (1983). Genes affecting flower colour and pH of flower limb homogenates in *Petunia hybrida*. Theor. Appl. Genet. **66,** 271-278.

de Vlaming, P., Cornu, A., Farcy, E., Gerats, A.G.M., Maizonnier, D., Wiering, H., and Wijsman, H.J.W. (1984). *Petunia hybrida*: A short description of the action of 91 genes, their origin and their map location. Plant Mol. Biol. Rep. **2,** 21-42.

Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., Zheng, C., Alberti, A., Anthony, F., Aprea, G., Aury, J.M., Bento, P., Bernard, M., Bocs, S., Campa, C., Cenci, A., Combes, M.C., Crouzillat, D., Da Silva, C., Daddiego, L., De Bellis, F., Dussert, S., Garsmeur, O., Gayraud, T., Guignon, V., Jahn, K., Jamilloux, V., Joet, T., Labadie, K., Lan, T., Leclercq, J., Lepelley, M., Leroy, T., Li, L.T., Librado, P., Lopez, L., Munoz, A., Noel, B., Pallavicini, A., Perrotta, G., Poncet, V., Pot, D., Priyono, Rigoreau, M., Rouard, M., Rozas, J., Tranchant-Dubreuil, C., VanBuren, R., Zhang, Q., Andrade, A.C., Argout, X., Bertrand, B., de Kochko, A., Graziosi, G., Henry, R.J., Jayarama, Ming, R., Nagai, C., Rounsley, S., Sankoff, D., Giuliano, G., Albert, V.A., Wincker, P., and Lashermes, P. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science **345,** 1181-1184.

Endress, P.K. (1996). Diversity and evolutionary biology of tropical flowers. (Cambridge, UK: Cambridge University Press).

Faraco, M., Spelt, C., Bliek, M., Verweij, W., Hoshino, A., Espen, L., Prinsi, B., Jaarsma, R., Tarhan, E., de Boer, A.H., Di Sansebastiano, G.P., Koes, R., and Quattrocchio, F.M. (2014). Hyperacidification of vacuoles by the combined action of two different P-ATPases in the tonoplast determines flower color. Cell Rep **6,** 32-43.

Galliot, C., Stuurman, J., and Kuhlemeier, C. (2006). The genetic dissection of floral pollination syndromes. Curr Opin Plant Biol **9,** 78-82.

Goodrich, J., Carpenter, R., and Coen, E.S. (1992). A common gene regulates pigmentation pattern in diverse plant species. Cell **68,** 955-964.

Grandin, N., de Almeida, A., and Charbonneau, M. (1998). The Cdc14 phosphatase is functionally associated with the Dbf2 protein kinase in Saccharomyces cerevisiae. Molecular & general genetics : MGG **258,** 104-116.

Hermann, K., Klahre, U., Moser, M., Sheehan, H., Mandel, T., and Kuhlemeier, C. (2013). Tight genetic linkage of prezygotic barrier loci creates a multifunctional speciation island in Petunia. Curr Biol **23,** 873-877.

Hernandez-Garcia, C.M., Bouchard, R.A., Rushton, P.J., Jones, M.L., Chen, X., Timko, M.P., and Finer, J.J. (2010). High level transgenic expression of soybean (Glycine max) GmERF and Gmubi gene promoters isolated by a novel promoter analysis pipeline. BMC Plant Biol **10,** 237.

Hoballah, M.E., Gubitz, T., Stuurman, J., Broger, L., Barone, M., Mandel, T., Dell'Olivo, A., Arnold, M., and Kuhlemeier, C. (2007). Single gene-mediated shift in pollinator attraction in Petunia. Plant Cell **19,** 779-790.

Klahre, U., Gurba, A., Hermann, K., Saxenhofer, M., Bossolini, E., Guerin, P.M., and Kuhlemeier, C. (2011). Pollinator choice in Petunia depends on two major genetic Loci for floral scent production. Curr. Biol. **21,** 730-739.

Koes, R., Verweij, C.W., and Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci. **5,** 236-242.

Koes, R., Souer, E., van Houwelingen, A., Mur, L., Spelt, C., Quattrocchio, F., Wing, J.F., Oppedijk, B., Ahmed, S., Maes, T., Gerats, T., Hoogeveen, P., Meesters, M., Kloos, D., and Mol, J.N.M. (1995). Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants. Proc. Natl. Acad. Sci. USA **81,** 8149-8153.

Koes, R.E., Spelt, C.E., and Mol, J.N.M. (1989a). The chalcone synthase multigene family of *Petunia hybrida* (V30): Differential, light-regulated expression during flower development and UV light induction. Plant Mol. Biol. **12,** 213-225.

Koes, R.E., Spelt, C.E., Mol, J.N.M., and Gerats, A.G.M. (1987). The chalcone synthase multigene family of *Petunia hybrida* (V30): Sequence homology, chromosomal localization and evolutionary aspects. Plant Mol. Biol. **10,** 375-385.

Koes, R.E., Spelt, C.E., Van den Elzen, P.J.M., and Mol, J.N.M. (1989b). Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida*. Gene. **81,** 245-257.

Koes, R.E., Van Blokland, R., Quattrocchio, F., Van Tunen, A.J., and Mol, J.N.M. (1990). Chalcone synthase promoters in petunia are active in pigmented and unpigmented cell types. Plant Cell **2,** 379-392.

Kranz, H.D., M., D., Greco, R., Jin, H., Leyva, A., R.C., M., Petroni, K., Urzainqui, A.B., M., Martin, C., Smeekens, S., Tonelli, C., Paz-Ares, J., and Weisshaar, B. (1998). Towards functional characterisation of the members of the *R2R3-MYB* gene family from *Arabidopsis thaliana*. Plant J. **16,** 263-276.

Kroon, A.R. (2004). Transcription regulation of the anthocyanin pathway in *Petunia hybrida* (Amsterdam: Vrije Universiteit).

Ludwig, S.R., and Wessler, S.R. (1990). Maize *R* gene family: tissue specific helix-loop-helix proteins. Cell **62,** 849-851.

Marrs, K.A., Alfenito, M.R., Lloyd, A.M., and Walbot, V. (1995). A glutathione *S*-transferase involved in vacuolar transfer encoded by the maize gene *Bronze-2*. Nature **375,** 397-400.

Morita, Y., Takagi, K., Fukuchi-Mizutani, M., Ishiguro, K., Tanaka, Y., Nitasaka, E., Nakayama, M., Saito, N., Kagami, T., Hoshino, A., and Iida, S. (2014). A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. The Plant journal : for cell and molecular biology **78,** 294-304.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-supression of homologous genes in trans. Plant Cell **2,** 279-289.

Napoli, C.A., Fahy, D., Wang, H.Y., and Taylor, L.P. (1999). white anther: A petunia mutant that abolishes pollen flavonol accumulation, induces male sterility, and is complemented by a chalcone synthase transgene. Plant Physiol **120,** 615-622.

Perry, G.H., Dominy, N.J., Claw, K.G., Lee, A.S., Fiegler, H., Redon, R., Werner, J., Villanea, F.A., Mountain, J.L., Misra, R., Carter, N.P., Lee, C., and Stone, A.C. (2007). Diet and the evolution of human amylase gene copy number variation. Nature genetics **39,** 1256-1260.

Petroni, K., and Tonelli, C. (2011). Recent advances on the regulation of anthocyanin synthesis in reproductive organs. Plant science : an international journal of experimental plant biology **181,** 219-229.

Povero, G. (2011). Physiological and genetic control of anthocyanin pigmentation in different species. In Developmental Genetics, Molecular Cell Biology (Amsterdam Vrije Universiteit), pp. 200.

Provenzano, S., Spelt, C., Hosokawa, S., Nakamura, N., Brugliera, F., Demelis, L., Geerke, D.P., Schubert, A., Tanaka, Y., Quattrocchio, F., and Koes, R. (2014). Genetics and evolution of anthocyanin methylation. Plant Physiol **165,** 962-977.

Quattrocchio, F., Wing, J.F., Leppen, H.T.C., Mol, J.N.M., 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., Wing, J.F., van der Woude, K., Mol, J.N.M., and Koes, R. (1998). Analysis of bHLH and MYB-domain proteins: species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes. Plant J. **13,** 475-488.

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.

Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J., and Koes, R. (1999). Molecular analysis of the *anthocyanin2* gene of *Petunia* and its role in the evolution of flower color. Plant Cell **11,** 1433-1444.

Ramsay, N.A., and Glover, B.J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends in plant science **10,** 63-70.

Serna, L., and Martin, C. (2006). Trichomes: different regulatory networks lead to convergent structures. Trends Plant Sci **11,** 274-280.

Sheehan, H., Moser, M., Klahre, U., Esfeld, K., Dell'Olivo, A., Mandel, T., Metzger, S., Vandenbussche, M., Freitas, L., and Kuhlemeier, C. (2015). MYB-FL controls gain and loss of floral UV absorbance, a key trait affecting pollinator preference and reproductive isolation. Nat Genet.

Spelt, C., Quattrocchio, F., Mol, J., and Koes, R. (2000). *anthocyanin1* of petunia encodes a basic-Helix Loop Helix protein that directly activates structural anthocyanin genes. Plant Cell **12,** 1619-1631.

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.

Tornielli, G., Koes, R., and Quattrocchio, F. (2009). The genetics of flower color. In Petunia: Evolutionary, Developmental and Physiological Genetics, T. Gerats and J. Strommer, eds (Heidelberg: Springer), pp. 269-300.

van der Krol, A.R., Lenting, P.E., Veenstra, J., Van der Meer, I.M., Koes, R.E., Gerats, A.G.M., Mol, J.N.M., and Stuitje, A.R. (1988). An anti-sense chalcone synthase gene in transgenic plants inhibits flower pigmentation. Nature **333,** 866-869.

van Houwelingen, A., Souer, E., Mol, J.N.M., and Koes, R.E. (1999). Epigenetic interactions between three *dTph1* transposons in two homologous chromosomes activates a new excision-repair mechanism in petunia. Plant Cell **11,** 1319-1336.

van Houwelingen, A., Souer, E., Spelt, C., 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.

van Tunen, A.J., Mur, L.A., Recourt, K., Gerats, A.G.M., and Mol, J.N.M. (1991). Regulation and manipulation of flavonoid gene expression in anthers of petunia: the molecular basis of the *po* mutation. Plant Cell **3,** 39-48.

van Tunen, A.J., Koes, R.E., Spelt, C.E., Van der Krol, A.R., Stuitje, A.R., and Mol, J.N.M. (1988). Cloning of the two chalcone flavanone isomerase genes from *Petunia hybrida*: Coordinate, light-regulated and differential expression of flavonoid genes. EMBO J. **7,** 1257-1263.

Verweij, C.W. (2007). Vacuolar acidification: Mechanism, regulation and function in petunia flowers. (Amsterdam: VU-University).

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.

Walker, A.R., Davison, P.A., Bolognesi-Winfield, A.C., James, C.M., Srinivasan, N., Blundell, T.L., Esch, J.J., Marks, M.D., and Gray, J.C. (1999). The *TRANSPARENT TESTA GLABRA1* locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. Plant Cell **11,** 1337-1350.

Wijsman, H.J.W. (1983). On the interrelationships of certain species of petunia II. Experimental data: crosses between different taxa. Acta Bot. Neerl. **32,** 97-107.

Willmer, P. (2011). Pollination and floral ecology. (Princeton, New Jersey: Princeton University Press).

Zenoni, S., D'Agostino, N., Tornielli, G., B., Quattrocchio, F., Chiusano, M.L., Koes, R., Zethof, J., Guzzo, F., Delledonne, M., Frusciante, L., Gerats, T., and Pezzotti, M. (2011). Revealing impaired pathways in the an11 mutant by high-throughput characterization of Petunia axillaris and Petunia inflata transcriptomes. Plant Journal **68,** 11-27.

Supplementary Note 8

Extreme variation in volatile production in wild petunias: what can we learn from their genomes?

Maaike R. Boersma[a], Joëlle K. Muhlemann[b1], Avichai Amrad[c], Natalia Dudareva[b], Cris Kuhlemeier[c], and Robert C. Schuurink[a2]

[a] Department of Plant Physiology, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands.
[b] Department of Biochemistry, Purdue University, West Lafayette, IN, 47907-2063, USA.
[c] Institute of Plant Sciences, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland.

[1] Current address: Department of Biology, Wake Forest University, 1834 Gulley Dr., Winston-Salem, NC 27109, USA.
[2] Correspondence to R.C.Schuurink@uva.nl

ABSTRACT

The genus *Petunia* displays wide variation in the production and emission of floral benzenoid and phenylpropanoid (FBP) volatiles. Variation in volatile production is part of coordinated changes in multiple floral traits that lead to shifts in pollinator visitation. The molecular nature of the genetic changes is mostly unknown and might involve mutations in structural genes, *cis*-acting regulatory elements as well as transcription factors. Although the variation in FBP emission in some *Petunia* species and hybrids can be explained by the mutation of the regulatory gene *ODORANT1*, a general mechanism underlying shifts in FBP emission remains unknown. The sequencing of the genomes of two closely related *Petunia* species, *P.axillaris axillaris N* and *P. integrifolia inflata S6*, provides the tools to study the genomic mechanisms underlying shifts in FBP emission: *P.axillaris axillaris N* emits an abundant blend of FBP whereas *P. integrifolia inflata S6* only emits benzaldehyde. Here we present an annotated catalogue of all known structural and regulatory genes involved in FBP production in these two species. We show that dissimilarity in volatile emission cannot be explained by differences in the coding sequences of a single FBP biosynthetic, *L*-phenylalanine biosynthetic or regulatory gene. We conclude that the major increase in the amount, complexity and regulation of volatile emission during the evolution of *P. axillaris* has involved either hitherto unknown genes or *cis*-acting regulatory mutations in known genes. Interestingly, the genes involved in phenylacetaldehyde and (iso)eugenol production have no close homologs in other *Solanaceae*.

INTRODUCTION

Over 1700 volatile compounds have been identified from the floral headspace of 90 plant families, and the differences in floral volatiles at the species level is a promising tool for evolutionary studies in angiosperms (Knudsen, 2006). Pollination of many insect-pollinated plants is dependent on floral scent (Dudareva et al., 2004; Kessler et al., 2008; Klahre et al., 2011) and the abundance, ratio and composition of the volatiles in the floral headspace are suggested to be involved in reproductive isolation of closely related species (Dudareva et al., 2013). *Petunia* floral scent is mainly composed of phenylpropanoids and benzenoids (Verdonk et al., 2003), which are the second most abundant and widespread class of plant volatiles (Knudsen, 2006) (Figure 1).

Floral phenylpropanoids and benzenoids (FBPs) are derived from the precursor *L*-phenylalanine (*L*-Phe), which originates from the shikimate and predominantly arogenate pathway (Maeda et al., 2010; Tzin and Galili, 2010; Yoo et al., 2013). The phenylpropanoid related (C_6-C_2) compounds 2-phenylethanol, 2-phenylacetaldehyde and 2-phenylethylacetate are synthesized via an unusual pathway in *Petunia* in which 2-phenylacetaldehyde is directly produced from *L*-Phe by phenylacetaldehyde synthase (PAAS) (Kaminaga et al., 2006). The biosynthesis of both benzenoids (C_6-C_1) and phenylpropanoids (C_6-C_3) starts with the deamination of *L*-Phe to *trans*-cinnamic acid (CA) by *L*-phenylalanine ammonia lyase (PAL). The C_6-C_1 pathway proceeds through the β-oxidative and non-β-oxidative pathway and yields benzaldehyde, benzylalcohol, methylbenzoate (MeBA), benzylacetate, methylsalicylate (MeSA) and benzylbenzoate (Boatright et al., 2004; Orlova et al., 2006; Van Moerkercke et al., 2009; Klempien et al., 2012; Qualley et al., 2012). The biosynthetic pathway leading to the C_6-C_3 volatiles isoeugenol and eugenol shares the precursor monolignol coniferyl alcohol with the lignin pathway (Muhlemann et al., 2014b),

whose oxygen group is eliminated during isoeugenol and eugenol formation (Koeduka et al., 2006). The exact biosynthetic route leading to vanillin is still unknown (Figure 1).

FBP emission and biosynthesis in *Petunia* is tissue-specific and can exhibit diurnal and developmental variation. It is intricately regulated by an array of MYB transcription factors that control different branches of the FBP biosynthetic pathway. The R2R3-MYB transcription factors ODORANT1 (ODO1), EMISION OF BENZENOIDSI/II/V (EOBI/II/V) and MYB4 are known to play a role in regulating FBPs biosynthesis (Verdonk et al., 2005; Spitzer-Rimon et al., 2010; Colquhoun et al., 2011; Van Moerkercke et al., 2011a; Spitzer-Rimon et al., 2012). The key regulator ODO1 regulates precursor supply to the FBP pathway and has been suggested to be in a feedback loop with EOBI and EOBII (Verdonk et al., 2005; Van Moerkercke et al., 2011a; Spitzer-Rimon et al., 2012; Van Moerkercke et al., 2012). MYB4 regulates *cinnamate-4-hydroxylase* (C4H) and thereby eugenol and isoeugenol biosynthesis (Colquhoun et al., 2011). The exact role of EOBV is not yet clear (Spitzer-Rimon et al., 2012), although a role as a negative regulator has recently been suggested (Cna'ani et al., 2015b). The clock gene LATE ELONGATED HYPOCOTYL (PhLHY) seems to directly determine the diurnal rhythm of FBP production by regulating *ODO1* expression (Fenske et al., 2015). Emission has recently been shown to be regulated by the R2R3-MYB transcription factor PH4 (Cna'ani et al., 2015a), which is predominantly known for its role in vacuolar acidification in relation to anthocyanin biosynthesis (Quattrocchio et al., 2006).

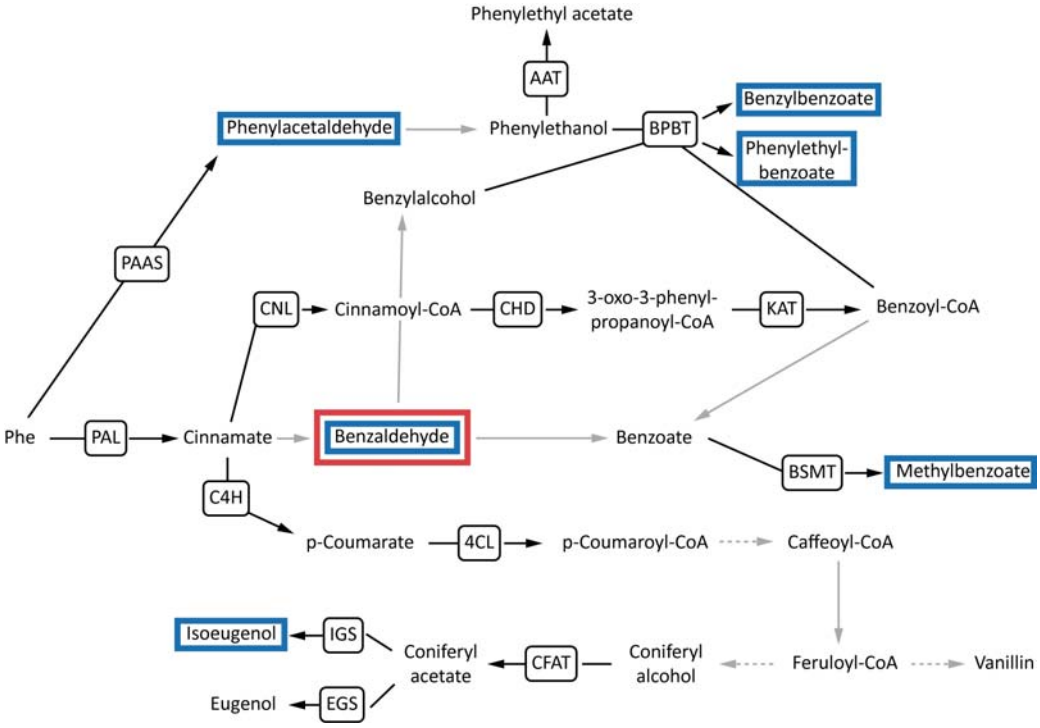


Figure 1. Biosynthetic pathway leading to the formation of floral phenylpropanoids and benzenoids in *Petunia*. Volatiles emitted by *PinfS6* are indicated by a red box and *PaxiN* volatiles by blue boxes. Grey arrows represent biochemical steps for which enzymes/genes have not been published in *Petunia* and dashed arrows represent multiple biochemical steps between metabolites. The precursor Phe is derived from the shikimate and arogenate pathway. Abbreviations: Phe: *L*-Phenylalanine; PAL: *L*-phenylalanine ammonia lyase; CNL: cinnamate-CoA ligase; CHD: cinnamoyl-CoA hydratase-dehydrogenase; C4H: cinnamate 4-hydoxylase; 4CL, 4-coumarate CoA-ligase; PAAS: phenylacetaldehyde synthase; KAT: 3-ketoacyl-CoA thiolase; BPBT: benzoyl-CoA:benzylalcohol/2-phenylethanol benzoyltransferase; BSMT: S-adenosyl-L-methionine:benzoic acid/salicylic acid carboxyl methyltransferase; CFAT: coniferylalcohol acetyl-CoA:coniferyl alcohol acetyltransferase; IGS: isoeugenol synthase; EGS: eugenol synthase; AAT: alcohol acetyltransferase.

Ultimately, the composition of the floral headspace is shaped by the selective pressures exerted by pollinators and florivores (Muhlemann et al., 2014a). Plants have evolved elaborate strategies to attract pollinators while preventing visitation by florivores. Within the *Petunia* floral headspace, certain FBPs were shown to attract pollinators, while others function in deterring of florivores (Kessler et al., 2013). Although little is known about the function of individual volatiles and specific volatiles mixes on insect behavior, *Petunia* species display typical moth or bee pollinated phenotypes. *P. axillaris* flowers are white and scented which is characteristic for nocturnal hawkmoth pollination (Ando et al., 2001; Klahre et al., 2011) whereas *P. inflata* has purple flowers which make few volatiles and are pollinated by bees (Stehmann, 1987; Ando et al., 2001). Scent is an important cue for hawkmoth visitation of *Petunia* flowers (Klahre et al., 2011), and the emission of volatiles is timed with pollinator activity (Hoballah et al., 2005). Different *P. axillaris* lines emit mixes of FBPs in which MeBA is dominant (Kondo et al., 2006; Koeduka et al., 2009). The *P. axillaris* line whose genome is sequenced in this study, *P.axillaris axillaris N* (*PaxiN*), emits benzylbenzoate, 2-phenylacetaldehyde, MeBA, benzaldehyde, phenylethylbenzoate and isoeugenol (Hoballah et al., 2005; Van Moerkercke et al., 2011b). In contrast, *P. inflata S6* (*PinfS6*) only emits benzaldehyde (Stuurman et al., 2004; Hoballah et al., 2005; Hoballah et al., 2007) (Figure 1).

A recent phylogenetic analysis of the genus robustly puts *P. inflata* and related bee-pollinated species into the ancestral short-tube clade, from which long-tubed species such *P. axillaris* are derived (Reck-Kortmann et al., 2014). Therefore, the evolution of a hawkmoth-pollination syndrome from a bee-pollinated ancestor has involved major changes in the amounts, complexity and diurnal regulation of FBPs. The availability of high quality genome sequences of representatives of the two clades provides new tools to study the genetic and genomic basis of this evolutionary transition.

Quantitative trait locus (QTL) analysis of a segregating population derived from crosses *PinfS6* or *P.axillaris* ssp *parodii* with a non-fragrant *P. hybrida* has shown that in, both populations, two loci can explain a large part of the differences in volatile production between the parents. This may suggest that only a few genes are responsible for the difference in the volatile profiles of *PaxiN* and *PinfS6* (Stuurman et al., 2004). The single-gene hypothesis can explain even faster adaptations: the rapid adaption of floral traits can be controlled by mutation of a single regulator (Dell'Olivo and Kuhlemeier, 2013). For instance, corolla color is under control of such a single regulator. The loss of color in *Petunia* is explained by mutations in the R2R3 MYB transcription factor ANTHOCYANIN2 (AN2) (Quattrocchio et al., 1999). It has also been shown that the mutation of two MYB binding sites in the promoter of *ODO1* is responsible for the loss of FBPs emission in the non-fragrant hybrid R27 (Van Moerkercke et al., 2011a). Furthermore, *ODO1* was identified as one of the two major QTLs that explain the difference in FBPs emission between *PaxiN* and *P. exserta* (Klahre et al., 2011). This points to a function for ODO1 as a single regulator of fragrance, similar to the function of AN2 in color biosynthesis.

The difference in floral scent production between closely related species of *Petunia* makes this an excellent model system to study evolutionary changes underlying shifts in FBP biosynthesis (Muhlemann et al., 2014a). Here, we show that the shift in volatile emission observed in *PaxiN* and *PinfS6* cannot be explained by the accumulation of mutations in the

coding regions of biosynthetic or regulatory genes. The shift in volatile emission is likely caused by the mutation of *cis*-acting regulatory elements or the mutation of a yet unknown transcriptional regulator, *L*-phenylalanine biosynthetic gene or transporter.

RESULTS

38 FBP genes have been annotated in the *PaxiN* and *PinfS6* genome

To annotate FBP genes in the *PaxiN* and *PinfS6* genome, *Petunia* sequences of five transcription factors, three *L*-phenylalanine biosynthetic genes and twelve biosynthetic genes acting downstream of *L*-Phe were obtained from Genbank (Table 1). These sequences were blasted against the *PaxiN* genome and the obtained sequences of *PaxiN* were subsequently used to identify the homologous genes in *PinfS6.* In total 38 gene models have been annotated for *PaxiN* and *PinfS6*. The transcription factors –but EOBI - and the FBP biosynthetic genes 3-ketoacyl-CoA thiolase 1 (*KAT1*), coniferylalcohol acetyl-CoA:coniferyl alcohol acetyltransferase (*CFAT*) and eugenol synthase (*EGS*) are encoded by single genes in both genomes. The *PinfS6* genome harbors an additional gene model of EOBI. The remaining biosynthetic genes have two (cinnamoyl-CoA hydratase-dehydrogenase (*CHD*), *C4H*, isoeugenol synthase (*IGS*)*, PAAS*) three (*PAL*, benzoyl-CoA:benzylalcohol/2-phenylethanol benzoyltransferase (*BPBT*), 4-coumarate CoA-ligase (*4CL*), cinnamate-CoA ligase (*CNL*)) or four (S-adenosyl-L-methionine:benzoic acid/salicylic acid carboxyl methyltransferase (*BSMT*)) copies. In addition, the *PinfS6* genome harbors a third gene model for *C4H*, whereas it has two gene models in *PaxiN.* Conversely, the *PaxiN* genome harbors a fifth and sixth gene model for *BSMT* and a third gene model for *PAAS*. *PAL*, *C4H* and *4CL* are relevant for both the FBP pathway and the anthocyanin pathway, and these genes have been discussed in (Supplementary Note 7). Thus we conclude that the complexity of the gene families is very similar between the two species.

Defective FBP or *L*- Phe biosynthetic genes cannot explain the difference in floral volatiles emitted by *PaxiN* and *PinfS6*

To investigate whether divergence of genes underlies the difference between benzaldehyde emission by *PinfS6* and emission of many FBPs by *PaxiN* (Figure 1), the coding sequences (CDS) of FBP genes were compared after translation (Table 2). Interestingly, the similarity between putative orthologs of *PaxiN* and *PinfS6* was higher than between their paralogs. For example, *PaxiN*PAAS1 was more similar to *PinfS6*PAAS1 than to the other *PaxiN*PAAS genes (Figure 3a). However, there are some exceptions. For example, the similarity between the BPBT1 and 2 paralogs is higher than between their putative orthologs.

It is well known that an (early) stop codon or a frameshift in the CDS of a biosynthetic gene can cause the loss of a volatile from the floral headspace, as was observed for IGS and isoeugenol synthesis in *Petunia axillaris subsp. parodii* (Koeduka et al., 2009). Thus we first investigated all known biosynthetic genes, i.e. those involved in benzylbenzoate, 2-phenylacetaldehyde, MeBA, phenylethylbenzoate and isoeugenol biosynthesis (Figure 1). Overall, these genes are 95% similar at the protein level in *PaxiN* and *PinfS6* (Table 2). However, none of the known *PinfS6* FBP biosynthetic genes have additional stop codons or frameshifts. Moreover, we found mostly conservative amino acid changes suggesting that the massive differences in volatile emission between the two species are unlikely to be due to functional differences at the protein level of biosynthetic enzymes.

Another explanation for the very limited FBPs biosynthesis by *PinfS6* could be the mutation of one or a few *L*-phenylalanine biosynthetic genes (Stuurman et al., 2004). The functional mutation of a shikimate or arogenate pathway biosynthetic gene could prevent the precursor supply to the FBP pathway. 3-Deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) and chorismate mutase (CM) are the only two enzymes from the shikimate pathway of which the roles in FBP biosynthesis have been characterized (Colquhoun et al., 2010; Langer et al., 2014). *DAHP1* and *2* and *CM1* and *2* are all expressed in *Petunia* petals, but only DAHP1 and CM1 are shown to be involved in *L*-Phe biosynthesis for FBPs. The biosynthesis of *L*-Phe from shikimate in petunia can proceed through two alternative pathways, still silencing of an arogenate dehydratase (*ADT1*) leads to a great decrease in FBP emission (Maeda et al., 2010; Yoo et al., 2013). *PinfS6* and *PaxiN* have no frame shifts or additional stop codons in the CDS of the *L*-phenylalanine biosynthetic genes *CM1*, *DAPH1* or *ADT1*. *PinfS6* has one to three non-conserved amino acids changes in these proteins compared to *PaxiN* (Table 2), meaning that these proteins are likely functional in *PinfS6*. Interestingly, the *PinfS6* genome harbors a second gene model of *ADT1* separated by 4 kbp, *ADT1b*. The predicted amino acid sequence of ADT1b is identical to that of ADT1a, however unlike the intronless *ADT1a* it contains a small intron.

PaxiN and _PinfS6_ have only minor differences in currently known FBP regulators

Because no functional mutations were found in the protein coding regions of genes responsible for FBP biosynthesis, the subsequent comparison focused on transcriptional regulators since, in principle, a lack of transcriptional regulators can shut down FBP biosynthesis (Verdonk et al., 2005; Langer et al., 2014). However, although functional mutations in transcriptional regulators could explain the loss of FBP biosynthesis it cannot explain why *PinfS6* emits benzaldehyde. Perhaps benzaldehyde emission could in this case be explained by precursor (cinnamic acid, Figure 1) outflow from other pathways (e.g. flavonoid, lignin) that is still converted to benzaldehyde as if this was by default, or the activity of a yet unknown transcription factor that regulates only this part of the FBP pathway.

Since the R2R3 MYB ODO1, a key regulator of FBP biosynthesis, can have a major role in shifting FBP biosynthesis (Klahre et al., 2011), this gene was studied in more detail. The protein sequences of ODO1 of *PaxiN* and *PinfS6* are identical in the R2R3 domain but differ by seven amino acids downstream and contains no frame shifts or stop codons. Two of these seven amino acids are identical in *PinfS6* and the fragrant *P. hybrida cv*. W115 and thus cannot contribute to changes in ODO1 functionality (Figure 2). A165G, S243N and T232I are the only non-conservative amino acid alterations in *PinfS6* compared to *PaxiN* and *P. hybrida cv. W115*. It remains to be tested if these mutations affect ODO1 functionality. *ODO1* is expressed at a lower level in *PinfS6* than *PaxiN* (Klahre et al., 2011), however this might lead to an overall reduction of FBP emission, and cannot explain a complete loss of FBP biosynthesis. It should also be noted here that mutation in two MYB-binding sites (MBS) in the promoter of *ODO1* leads to loss of FBP biosynthesis in *P. hybrida* (Van Moerkercke et al., 2011a), but these MBSs are present and intact in the promoters of *PaxiN* and *PinfS6*. In addition, the synteny between the corresponding regions (10kbp) of *PaxiN* and *PinfS6* around *ODO1* remained intact.

Table 1. Original Genbank numbers and geneID of *PaxiN* and *PinfS6* FBP genes

Gene	Genbank	*PaxiN* geneID	*P. Inf6* geneID
Biosynthetic genes			
PAAS1	DQ243784.1	Peaxi162Scf00561g02002	Peinf101Scf00985g02002
PAAS2	DQ243784.1	Peaxi162Scf00152g02037	
PAAS3	DQ243784.1	Peaxi162Scf00152g02036	Peinf101Scf01065g00017
BPBT1	AY611496.1	Peaxi162Scf00007g00012	Peinf101Scf01180g01014
BPBT2	AY611496.1	Peaxi162Scf00007g00011	Peinf101Scf01180g01013
BPBT3	AY563157.1	Peaxi162Scf00169g00922	Peinf101Scf00276g03003
BSMT1	AY233465.1/AY233466/DQ494491	Peaxi162Scf00047g11023	Peinf101Scf00686g02008
BSMT2	AY233465.1/AY233466/DQ494491	Peaxi162Scf00047g11028	
BSMT3	AY233465.1/AY233466/DQ494491	Peaxi162Scf00047g11029	
BSMT4	AY233465.1/AY233466/DQ494491	Peaxi162Scf03967g00005	Peinf101Scf00437g04005
BSMT5	AY233465.1/AY233466/DQ494491	Peaxi162Scf00423g00021	Peinf101Scf00437g02026
BSMT6	AY233465.1/AY233466/DQ494491	Peaxi162Scf00423g00119	Peinf101Scf00437g02028
CNL1	JN120848	Peaxi162Scf00294g04015	Peinf101Scf00099g03010
CNL2	JN120848	Peaxi162Scf00784g00010	Peinf101Scf00099g04016
CNL3	JN120848	Peaxi162Scf00294g00032	Peinf101Scf00123g10004
CHD1	JX142126.1	Peaxi162Scf00231g00330	Peinf101Scf00961g08042
CHD2	JX142126.1	Peaxi162Scf01363g00003	Peinf101Scf05015g00020
KAT1	FJ657663.1	Peaxi162Scf00052g08016	Peinf101Scf00191g27016
CFAT	DQ767969.1	Peaxi162Scf00474g02020	Peinf101Scf00221g12012
EGS	EF467241.1	Peaxi162Scf00020g17016	Peinf101Scf00318g05021
IGS1	DQ372813.1	Peaxi162Scf00889g02031	Peinf101Scf01349g08001
IGS2	DQ372813.1	Peaxi162Scf00060g00024	Peinf101Scf00001g26039
***L*-phenylalanine biosynthetic genes**			
CM1	EU751616	Peaxi162Scf00166g09045	Peinf101Scf01969g05044
DAHP1	JQ955569	Peaxi162Scf00030g17016	Peinf101Scf00007g07016
ADT1a	FJ790412	Peaxi162Scf00114g00001	Peinf101Scf00392g01004
ADT1b			Peinf101Scf00392g01021
Transcription factors			
EOBIa	KC182628.1	Peaxi162Scf00129g12034	Peinf101Scf02365g00039
EOBIb			Peinf101Scf05157g00020
EOBII	EU360893.1	Peaxi162Scf00080g06005	Peinf101Scf00394g13005
EOBV	GQ449250.1	Peaxi162Scf00362g00831	Peinf101Scf01468g01015
MYB4	HM447143.1	Peaxi162Scf01221g00035	Peinf101Scf00661gX
ODO1	AY705977.1	Peaxi162Scf00002g03010	Peinf101Scf00284g00012

* PAL, C4H and 4CL have been described in Supplementary Note 7, (for abbreviations, see Figure 1; CM1: chorismate mutase 1 ; DAHP1: 3-deoxy-D-arabino-heptulosonate-7-phosphate; ADT1: arogenate dehydratase, EOB: EMISSION OF BENZENOIDS; MYB: myeloblastosis; ODO1; ODORANT1)

Since the R2R3 MYB ODO1, a key regulator of FBP biosynthesis, can have a major role in shifting FBP biosynthesis (Klahre et al., 2011), this gene was studied in more detail. The protein sequences of ODO1 of *PaxiN* and *PinfS6* are identical in the R2R3 domain but differ by seven amino acids downstream and contains no frame shifts or stop codons. Two of these seven amino acids are identical in *PinfS6* and the fragrant *P. hybrida cv*. W115 and thus cannot contribute to changes in ODO1 functionality (Figure 2). A165G, S243N and T232I are the only non-conservative amino acid alterations in *PinfS6* compared to *PaxiN* and *P. hybrida cv. W115*. It remains to be tested if these mutations affect ODO1 functionality. *ODO1* is expressed at a lower level in *PinfS6* than *PaxiN* (Klahre et al., 2011), however this might lead to an overall reduction of FBP emission, and cannot explain a complete loss of FBP biosynthesis. It should also be noted here that mutation in two MYB-binding sites (MBS) in the promoter of *ODO1* leads to loss of FBP biosynthesis in *P. hybrida* (Van Moerkercke et al., 2011a), but these MBSs are present and intact in the promoters of *PaxiN* and *PinfS6*. In addition, the synteny between the corresponding regions (10kbp) of *PaxiN* and *PinfS6* around *ODO1* remained intact.

Two other R2R3 MYBs that play a role in FBPs biosynthesis, EOBI and EOBII, have only one amino acid change in *PinfS6* compared to *PaxiN*, A144V and V165M, respectively. In contrast to *PaxiN, PinfS6* harbors an additional *EOBI, EOBIb.* The predicted amino acid sequence of EOBIb has in comparison to EOBIa one amino acid less and eight amino acid changes, of which three conserved (PAM 250 >250). Similar as with ODO1, these amino acid changes are located outside the R2R3 domain of EOBI and EOBII. The amino acid changes in EOBII are conservative and moreover it is known that *EOBII* is expressed in both *PinfS6* and *PaxiN* (data not shown). The amino acid changes in EOBI are only weakly similar (PAM 250 =<0.5), however, it is unlikely that this leads to changes in the functionality of EOBI because the fragrant *Petunia cv.* P720 has the same amino acid substitution as *PinfS6*. Expression data for *EOBI* in *PinfS6* and *PaxiN* are unavailable yet.

Besides the loss of function of positive regulators, the gain of function of negative regulators could explain loss of FBP biosynthesis. The known negative regulators of FBP biosynthesis, MYB4 and EOBV (Colquhoun et al., 2011; Spitzer-Rimon et al., 2012), both have four amino acid changes in *PinfS6* compared to *PaxiN*, two of which are conservative. Interestingly, EOBV has a deletion of six amino acids in *PaxiN* towards the C-terminus. Whether these substitutions and deletions in EOBV lead to its reduced activity in *PaxiN* and increased FBP biosynthesis is unknown and requires further investigations, as well as *EOBV* transcript levels. Remarkably, *PaxiN* EOBV protein is completely identical to EOBV from the fragrant hybrid cv. P720.

Table 2. Comparison of *PaxiN* and *PinfS6* FBP genes

Biosynthetic genes	*PaxiN* amino acids	*P.infS6 amino acids*	changes	of which conserved	similarity
PAAS1	506	506	24	16	95.26
PAAS2	506				
PAAS3	506	506	30	17	94.07
BPBT1	460	460	9	4	98.04
BPBT2	460	460	7	4	98.48
BPBT3	456	456	9	8	98.03
BSMT1	357	357	29	18	91.88
BSMT2	357				
BSMT3	336				
BSMT4	360	360	9	7	97.5
BSMT5	358	358	13	5	96.38
BSMT6	358	356	10	6	97.19
CNL1	570	570	16	8	97.19
CNL2	528	570	16	8	96.59
CNL3	196	587	57	23	68.37
CHD1	724	724	8	6	98.9
CHD2	724	713	8	2	98.88
KAT1	462	462	25	15	94.59
CFAT	454	458	23	10	93.61
EGS	308	308	6	2	98.05
IGS1	323	323	9	4	97.21
IGS2	319	319	17	10	94.67
***L*-phenylalanine precursor genes**	*PaxiN* amino acids	*P.infS6 amino acids*	changes	of which conserved	similarity
CM1	324	324	5	3	98.46
DAHP1	533	533	3	1	99.44
ADT1a	424	424	5	2	98.82
ADT1b		424	5	2	98.82
Transcription factors	*PaxiN* amino acids	*P.infS6 amino acids*	changes	of which conserved	similarity
EOBIa	202	202	1	0	99.5
EOBIb		201	7	5	96.52
EOBII	197	197	1	1	99.49
EOBV	267	273	4	2	98.5
MYB4	258	258	4	2	98.45
ODO1	294	294	7	4	97.62

***** For abbreviations, see Figure 1 and table 1. Amino acid sequences have been aligned with ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Conserved is PAM 250 >250).

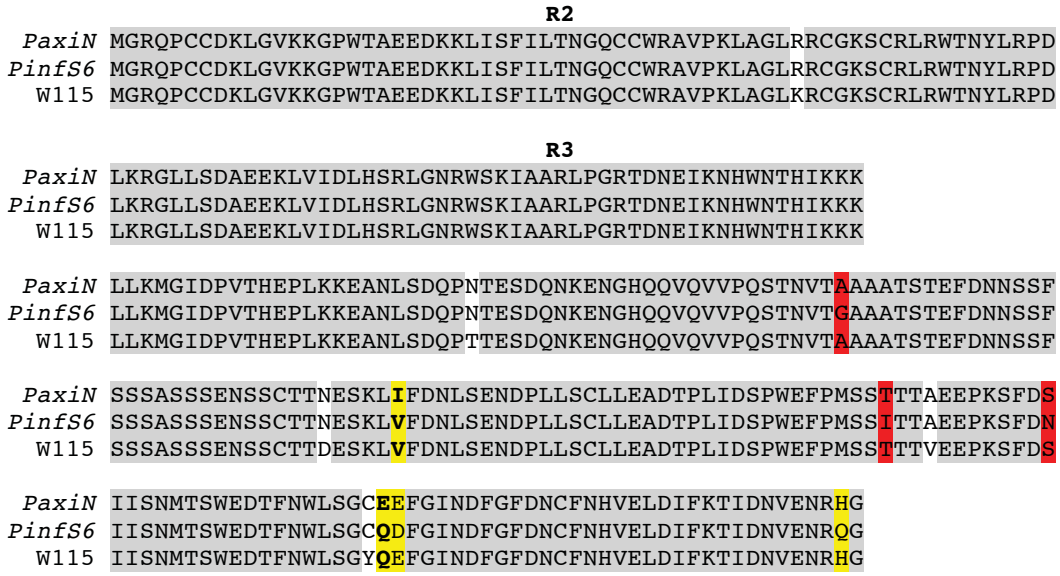


Figure 2. Amino acid sequence alignment of Petunia ODO1.

ODORANT1 (ODO1) from *P.axillaris axillaris N* (*PaxiN*), *P. integrifolia inflata S6* (*PinfS6*) and *P. hybrida cv.* Mitchell (W115) were aligned with ClustalW2. Yellow indicates amino acids that are conserved between *PaxiN* and *PinfS6* (Gonnet PAM 250 scoring >250), these amino acids are bold if they are similar between *PinfS6* and W115. Red indicates amino acids that are not conserved between *PaxiN* and *PinfS6*.

PAAS, CFAT and IGS have no close homologs in other *Solanaceae*

Even though the phenylalanine pathway is highly conserved (Tohge et al., 2013a) there are differences between *Petunia* and other Solanaceae in biosynthesis of phenylpropanoids. For example, 2-phenylacetaldehyde is synthesized by PAAS in *Petunia* whereas tomato requires two enzymes for this conversion (Kaminaga et al., 2006; Tieman et al., 2006). We thus explored some differences in this regard between *Petunia* and other Solanaceae, using the genomes of *Nicotiana benthamiana*, *Solanum lycopersicum* and *Solanum tuberosum*. PAAS, CFAT and IGS, which are likely to be unique for *Petunia,* were selected for further analysis. The three *PaxiN* genes were used to blast the genomes. The protein sequences of the three best hits for PAAS, CFAT and IGS were obtained from *N. benthamiana*, *S. lycopersicum* and *S. tuberosum* and compared to *PaxiN* and *PinfS6*.

P. hybrida PAAS synthesizes 2-phenylacetaldehyde by oxidative decarboxylation and oxidation of *L*-Phe. PAAS also has been identified in rose but not in tomato (Kaminaga et al., 2006). In tomato two independent enzymes presumably catalyze the decarboxylation and oxidation of *L*-Phe to form 2-phenylacetaldehyde. A small family of decarboxylases in tomato catalyzes the first step in phenethylamine production. The second step is likely performed by a yet unidentified amine oxidase (Tieman et al., 2006). The most homologous gene model in tomato to *Petunia* PAAS is 76% similar (Solyc03g0455020.2.1 to *PaxiN*PAAS2). However, this putative tomato PAAS has an early stop codon that makes the tomato protein 124 amino acids shorter than *PaxiN*PAAS2. Additionally, all the Petunia PAAS copies are more similar to each other then to *N. benthamiana*, *S. lycopersicum* and *S. tuberosum* genes (Figure 3a).

Coniferyl acetate is synthesized by CFAT and is the direct substrate for IGS and EGS, which produce isoeugenol and eugenol, respectively. The two *Petunia* genes directly responsible for isoeugenol synthesis, CFAT and IGS, gave very poor blast hits in all the three

Solanaceae studied (Figure 3b and c). *Petunia* IGS is 47% identical to *Petunia* EGS (Koeduka et al., 2008) and the *Petunia* IGSs copies fall in a separate clade. In contrast, *Petunia* EGSs have higher similarity to putative reductases from *N. benthamiana*, *S. lycopersicum* and *S. tuberosum* (Figure 3b). Likewise, *Petunia* CFAT does not group together with genes from the other Solanaceae species. This implies that although these enzymes act in the same pathway, EGS is more conserved within the Solanaceae than IGS and CFAT.

DISCUSSION

Petunia has proven to be a suitable model system to study biosynthesis and regulation of the floral benzenoid phenylpropanoid (FBP) pathway (Verdonk et al., 2003; Muhlemann et al., 2014a). The sequence of the genomes from two closely related *Petunia* species, *PaxiN* and *PinfS6*, provides the scientific community with a valuable tool to study changes in FBP biosynthesis at the genomic level. *PaxiN* and *PinfS6* have distinct floral volatile profiles, and moreover, they are the ancestors of a comprehensive *P. hybrida* collection that display wide variety in FBP biosynthesis.

The analysis of both genomes shows that there are no frame shifts or stop codons present in *PinfS6* FBP biosynthetic genes compared to *PaxiN*. Moreover, *PinfS6* and *PaxiN* FBP biosynthetic proteins overall differ only 5% and most of those changes are silent or conservative. Small functional differences caused by changes in the protein sequences cannot be excluded, but it seems unlikely that these can account for the massive and complex increase in volatiles in *P. axillaris* (Table 2). Conversely, the loss of FBP biosynthesis in *PinfS6* could be the result of a mutation in upstream biosynthetic genes preventing precursor supply to the FBP pathway. *L*-Phe is an essential constituent of proteins as well as the precursor for many different secondary metabolites (Tohge et al., 2013b). Functional mutation of this pathway would probably lead to serious defects in plant, and this could explain the presence of multiple copies of shikimate and arogenate pathway and *PAL* genes in plants. Nevertheless, there are examples in which silencing of specific shikimate and arogenate pathway genes (*CM1, DAPH1* and *ADT1*) leads to reduced FBP emission in *Petunia* without causing severe phenotypes (Verdonk et al., 2005; Colquhoun et al., 2010; Maeda et al., 2010; Langer et al., 2014). Interestingly, the functionality of the known *L*-phenylalanine biosynthetic genes *CM1, DAPH1* and *ADT1*, is likely not altered in *PinfS6*. CM1, DAPH1 and ADT1a/ADT1b have only three, five and three amino acid changes compared to *PaxiN*, respectively (Table 2). Although *ADT1b* from PinfS6 is identical to *ADT1a* at the protein level, it contains an intron unlike *ADT1a*. It cannot be excluded that the changes in the protein sequences collectively underlie the phenotypic differences as polygenic trait as argued by Rockman (Rockman, 2012). However, the relatively simple genetic structure suggested by QTL analysis (Stuurman et al., 2004; Klahre et al., 2011) indicates the presence of major functional differences in as yet unidentified genes or in promoters of the genes analyzed here. Potential candidates also include transporters such as the proposed but unidentified plastidial *L*-Phe transporter, which delivers *L*-Phe from the plastid to the cytosol (Yoo et al., 2013).

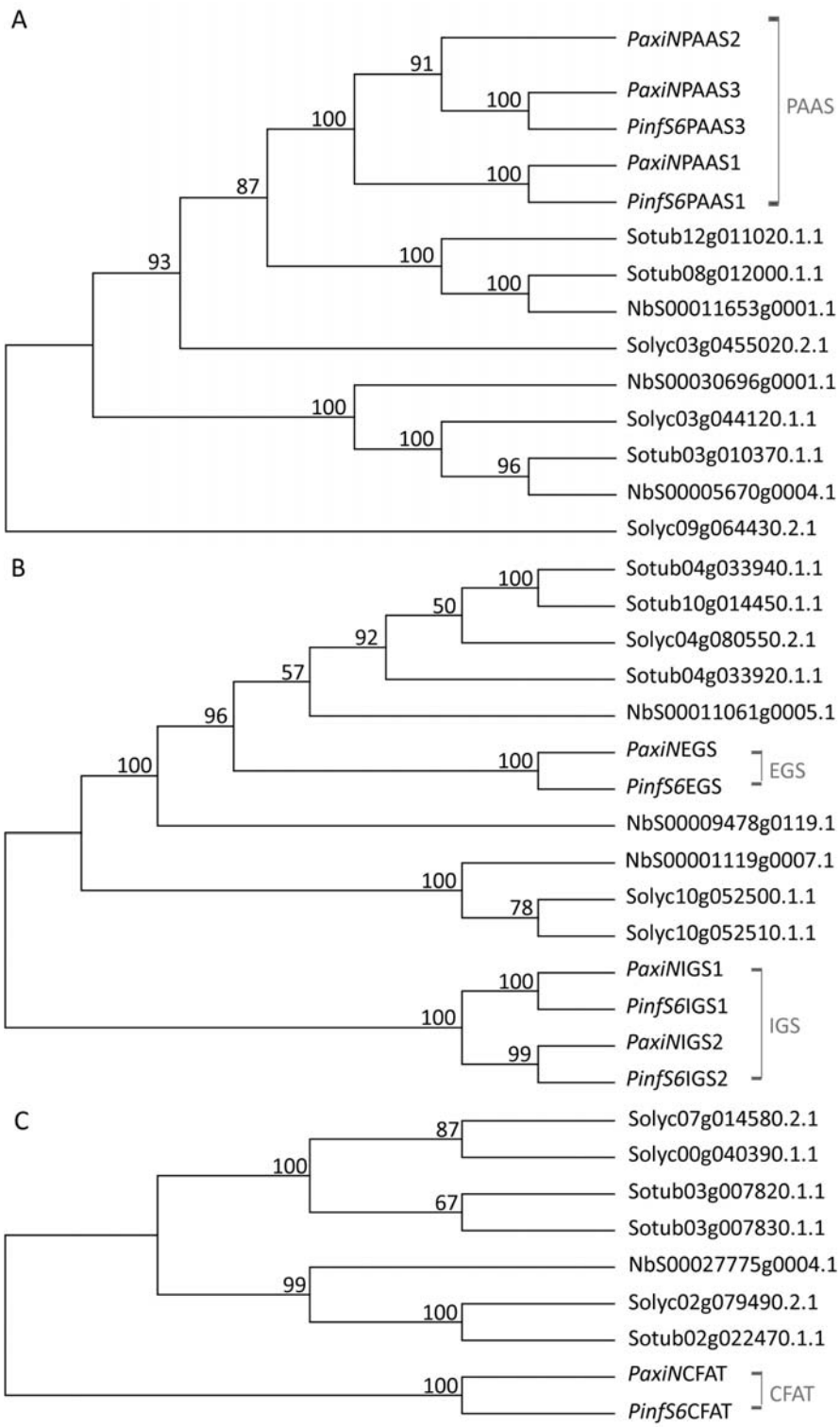


Figure 3. Unrooted neighbor joining trees of PAAS (a), IGS and EGS (b) and CFAT (c).
Unrooted neighbor joining trees from the protein sequence of different Solanaceae were created with MEGA6.06. Numbers represent Bootstrap values for 1000 replicates. Petunia identifiers are listed in table 1. Abbreviations: PAAS: phenylacetaldehyde synthase; CFAT: coniferylalcohol acetyl-CoA:coniferyl alcohol acetyltransferase; IGS: isoeugenol synthase; EGS: eugenol synthase; PaxiN: *Petunia axillaris axillaris N*; PinfS6: *Petunia integrifolia infalata S6*; Solyc: *Solanum lycopersicum*; Sotub: *Solanum tuberesum*; Nb: *Nicotiana benthamiana*.

We also investigated whether the difference in volatile emission by *PaxiN* and *PinfS6* can be explained by genetic variation in key transcriptional regulators of FBP emission. Prior evidence points to such a role for ODO1. Mutations in EOBII binding sites of the *ODO1* promoter are sufficient to lose the ability to synthesize FBPs (Van Moerkercke et al., 2011a). Moreover, two QTLs explain the difference in FBP biosynthesis in *PaxiN* and *P. exserta,* and *ODO1* was found on one of these two QTLs on chromosome VII (Klahre et al., 2011). Interestingly, *ODO1* is expressed in both *PaxiN* and *PinfS6,* although at a lower level in *PinfS6* (Klahre et al., 2011). It is expected that a lower level of *ODO1* expression would result in lower levels of all FBPs, which is not the case in *PinfS6*. Further analysis of ODO1 revealed that there are three non-conserved amino acid substitutions in *PinfS6* compared to the fragrant *PaxiN* and *P. hybrida cv.* W115 (Figure 2). It remains to be determined if these mutations have consequences for the activity of ODO1 in *PinfS6*, however, it's unlikely that these mutations alter ODO1 activity.

From our analysis of the genome it has become clear that *PinfS6* is likely carrying functional biosynthetic FBP genes. It has a dysfunctional fragrance system which can be restored by the replacement of chromosome III and a segment of chromosome IV with those from the non-fragrant hybrid W138 (Stuurman et al., 2004). Interestingly, W138 is non-fragrant since is lacks *ODO1* expression, thus ODO1 seems still functional in *PinfS6* (Verdonk et al., 2005). Replacement of chromosome II in *P. exserta* as well as in *PinfS6* with that of *PaxiN* restores FBP biosynthesis *(Stuurman et al., 2004; Klahre et al., 2011; Hermann et al., 2013).* Thus it seems that a major, unknown regulator is located on chromosome II that is absent in *PinfS6* (and *P. exserta*). One of the other regulators of FBP biosynthesis, EOBII, is located on chromosome II (Hermann et al., 2013). EOBII and EOBI are more conserved between *PinfS6* and *PaxiN* than ODO1 and both have only one amino acid substitution. This amino acid change is conservative in EOBII. In addition, the *PinfS6* genome harbors *EOBIb*, which has eight amino acid changes and one deletion compared to EOBIa. Altogether there seem no differences in ODO1, EOBI and EOBII between *PaxiN* and *PinfS6* that can explain the difference in FBP biosynthesis *(Sablowski et al., 1994; Tohge et al., 2013a)*.

The presence of a negative regulator in *PinfS6*, downregulating biosynthesis of benzylbenzoate, 2-phenylacetaldehyde, MeBA, phenylethylbenzoate and isoeugenol, could explain the differences between *PinfS6* and *PaxiN.* The only two known negative regulators of FBP biosynthesis to date are MYB4 and EOBV. MYB4 specifically regulates isoeugenol and eugenol biosynthesis and does not control the biosynthesis of the other volatiles (Colquhoun et al., 2011). The exact role of EOBV is still unknown, but silencing of EOBV leads to increased overall volatile emission, that is not further specified (Spitzer-Rimon et al., 2012). Recently a role as a wide-range transcriptional repressor of the pathway was suggested for EOBV (Cna'ani et al., 2015b). EOBV has four amino acid changes and a deletion of six amino acids in *PaxiN* compared to *PinfS6*. Since expression and activity data of EOBV are not available, the potential effect of mutations in EOBV on FBP biosynthesis remains elusive.

In order to study if the mutation of other *L*-phenylalanine biosynthetic genes could lead to diminutive FBP biosynthesis more understanding of this biosynthetic pathway is necessary. To target the search for mutations in *L*-phenylalanine biosynthetic genes causing the shift in FBP biosynthesis, expression data revealing which biosynthetic genes are differently expressed in *PaxiN* and *PinfS6* petals is essential or perhaps even protein levels and

enzymatic activities. It might also shed light on why *PinfS6* can still emit benzaldehyde. Its potential precursor, cinnamic acid, is also a precursor for anthocyanin and lignin biosynthesis and it is possible that cinnamic acid not utilized by these pathways could be converted to benzaldehyde by default. Interestingly, silencing of one of the FBP regulators, EOBI, in *P. hybrida* Mitchell (W115) does not alter benzaldehyde emission, whereas all other volatiles that are also emitted by *PaxiN* are affected by EOBI silencing (Spitzer-Rimon et al., 2012). It suggests that benzaldehyde emission could be regulated differently than emission of other FBPs.

Even though changes in the floral headspace are believed to be a promising tool for studying evolutionary mechanism at the species level (Knudsen, 2006), there are examples of studies using floral volatiles for studying evolution between species. A study comparing IGS and EGS between *Ocimum basilicum*, *Clarkia breweri* and *P. hybrida* led to the hypothesis that EGS and IGS have evolved independently at least twice in different plant lineages (Koeduka et al., 2008). Interestingly, *IGS* from *Petunia* seems to be more different from other Solanaceae genes than *Petunia* EGS. Tomato is not known to produce isoeugenol, and the lack of *IGS* genes in *N. benthamiana*, *S. lycopersicum* and *S. tuberosum* genomes implies that *IGS* has been lost or gained in different Solanaceae genera. Surprisingly, the *Petunia* biosynthetic gene directly upstream of IGS, CFAT, also falls in a separate clade than genes from *N. benthamiana*, *S. lycopersicum* and *S. tuberosum* (Figure 3c). Although *Petunia* EGS has homologous genes in other Solanaceae (Figure 3b), and tomato is reported to synthesize eugenol (Ortiz-Serrano and Gil, 2010), genera within the Solanaceae probably have a different acyltransferase responsible for the conversion of coniferyl alcohol. A similar situation appears to have occurred for 2-phenylacetaldehyde biosynthesis. *Petunia* 2-phenylacetaldehyde is synthesized by the enzyme PAAS, whereas tomato requires two enzymatic steps for the conversion of *L*-Phe (Kaminaga et al., 2006; Tieman et al., 2006). *Petunia* PAAS genes do clade separately from putative *N. benthamiana*, *S. lycopersicum* and *S. tuberosum* PAAS genes (Figure 3a). The tomato gene most homologous to *PaxiN*PAAS2, encodes a protein that is 124 amino acid shorter than its *Petunia* homolog. This might implicate that PAAS function has been lost in tomato and two other genes evolved in tomato to catalyze 2-phenylacetaldehyde formation. The presence of a limited number of FBP-producing species in the Solanaceae makes it difficult to unravel how benzenoid/phenylpropanoid biosynthesis has evolved in this family. However, our results clearly show that the Solanaceae have evolved multiply strategies for (iso)eugenol and 2-phenylacetaldehyde biosynthesis.

METHODS

FBP CDS were extracted from Genbank and annotated in WebApollo as described in the main article. Protein sequences were aligned with ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Amino acids are considered conserved when the Gonnet PAM 250 scoring is >250. Unrooted neighbor joining trees were created with MEGA6.06 using ClustalW for the alignment and 1000 Bootstrap replicates for the phylogeny analysis (http://www.megasoftware.net/).

REFERENCES

Ando, T., Nomura, M., Tsukahara, J., Watanabe, H., Kokubun, H., Tsukamoto, T., Hashimoto, G., Marchesi, E., and Kitching, I.J. (2001). Reproductive isolation in a native population of Petunia sensu Jussieu (Solanaceae). Ann Bot-London **88,** 403-413.

Boatright, J., Negre, F., Chen, X., Kish, C.M., Wood, B., Peel, G., Orlova, I., Gang, D., Rhodes, D., and Dudareva, N. (2004). Understanding in vivo benzenoid metabolism in petunia petal tissue. Plant Physiol **135,** 1993-2011.

Cna'ani, A., Spitzer-Rimon, B., Ravid, J., Farhi, M., Masci, T., Aravena-Calvo, J., Ovadis, M., and Vainstein, A. (2015a). Two showy traits, scent emission and pigmentation, are finely coregulated by the MYB transcription factor PH4 in petunia flowers. The New phytologist.

Cna'ani, A., Muhlemann, J.K., Ravid, J., Masci, T., Klempien, A., Nguyen, T.T., Dudareva, N., Pichersky, E., and Vainstein, A. (2015b). Petunia x hybrida floral scent production is negatively affected by high-temperature growth conditions. Plant, cell & environment **38,** 1333-1346.

Colquhoun, T.A., Schimmel, B.C., Kim, J.Y., Reinhardt, D., Cline, K., and Clark, D.G. (2010). A petunia chorismate mutase specialized for the production of floral volatiles. The Plant Journal **61,** 145-155.

Colquhoun, T.A., Kim, J.Y., Wedde, A.E., Levin, L.A., Schmitt, K.C., Schuurink, R.C., and Clark, D.G. (2011). PhMYB4 fine-tunes the floral volatile signature of Petunia x hybrida through PhC4H. J Exp Bot **62,** 1133-1143.

Dell'Olivo, A., and Kuhlemeier, C. (2013). Asymmetric effects of loss and gain of a floral trait on pollinator preference. Evolution **67,** 3023-3031.

Dudareva, N., Pichersky, E., and Gershenzon, J. (2004). Biochemistry of plant volatiles. Plant Physiol **135,** 1893-1902.

Dudareva, N., Klempien, A., Muhlemann, J.K., and Kaplan, I. (2013). Biosynthesis, function and metabolic engineering of plant volatile organic compounds. New Phytol **198,** 16-32.

Fenske, M.P., Hewett Hazelton, K.D., Hempton, A.K., Shim, J.S., Yamamoto, B.M., Riffell, J.A., and Imaizumi, T. (2015). Circadian clock gene LATE ELONGATED HYPOCOTYL directly regulates the timing of floral scent emission in Petunia. Proceedings of the National Academy of Sciences of the United States of America **112,** 9775-9780.

Hermann, K., Klahre, U., Moser, M., Sheehan, H., Mandel, T., and Kuhlemeier, C. (2013). Tight Genetic Linkage of Prezygotic Barrier Loci Creates a Multifunctional Speciation Island in Petunia. Current Biology **23,** 873-877.

Hoballah, M.E., Stuurman, J., Turlings, T.C.J., Guerin, P.M., Connetable, S., and Kuhlemeier, C. (2005). The composition and timing of flower odour emission by wild Petunia axillaris coincide with the antennal perception and nocturnal activity of the pollinator Manduca sexta. Planta **222,** 141-150.

Hoballah, M.E., Gubitz, T., Stuurman, J., Broger, L., Barone, M., Mandel, T., Dell'Olivo, A., Arnold, M., and Kuhlemeier, C. (2007). Single gene-mediated shift in pollinator attraction in Petunia. The Plant cell **19,** 779-790.

Kaminaga, Y., Schnepp, J., Peel, G., Kish, C.M., Ben-Nissan, G., Weiss, D., Orlova, I., Lavie, O., Rhodes, D., Wood, K., Porterfield, D.M., Cooper, A.J., Schloss, J.V., Pichersky, E., Vainstein, A., and Dudareva, N. (2006). Plant phenylacetaldehyde synthase is a bifunctional homotetrameric enzyme that catalyzes phenylalanine decarboxylation and oxidation. J Biol Chem **281,** 23357-23366.

Kessler, D., Gase, K., and Baldwin, I.T. (2008). Field experiments with transformed plants reveal the sense of floral scents. Science **321,** 1200-1202.

Kessler, D., Diezel, C., Clark, D.G., Colquhoun, T.A., and Baldwin, I.T. (2013). Petunia flowers solve the defence/apparency dilemma of pollinator attraction by deploying complex floral blends. Ecol Lett **16,** 299-306.

Klahre, U., Gurba, A., Hermann, K., Saxenhofer, M., Bossolini, E., Guerin, P.M., and Kuhlemeier, C. (2011). Pollinator choice in Petunia depends on two major genetic Loci for floral scent production. Curr Biol **21,** 730-739.

Klempien, A., Kaminaga, Y., Qualley, A., Nagegowda, D.A., Widhalm, J.R., Orlova, I., Shasany, A.K., Taguchi, G., Kish, C.M., Cooper, B.R., D'Auria, J.C., Rhodes, D., Pichersky, E., and Dudareva, N. (2012). Contribution of CoA ligases to benzenoid biosynthesis in petunia flowers. The Plant cell **24,** 2015-2030.

Knudsen, J.T.E., R.; Gershenzon, J; Ståhl, B. (2006). Diversity and Distribution of Floral Scent. The Botanical Review **72,** 120.

Koeduka, T., Orlova, I., Baiga, T.J., Noel, J.P., Dudareva, N., and Pichersky, E. (2009). The lack of floral synthesis and emission of isoeugenol in Petunia axillaris subsp parodii is due to a mutation in the isoeugenol synthase gene. Plant J **58,** 961-969.

Koeduka, T., Louie, G.V., Orlova, I., Kish, C.M., Ibdah, M., Wilkerson, C.G., Bowman, M.E., Baiga, T.J., Noel, J.P., Dudareva, N., and Pichersky, E. (2008). The multiple phenylpropene synthases in both Clarkia breweri and Petunia hybrida represent two distinct protein lineages. The Plant Journal **54,** 362-374.

Koeduka, T., Fridman, E., Gang, D.R., Vassao, D.G., Jackson, B.L., Kish, C.M., Orlova, I., Spassova, S.M., Lewis, N.G., Noel, J.P., Baiga, T.J., Dudareva, N., and Pichersky, E. (2006). Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. Proceedings of the National Academy of Sciences of the United States of America **103,** 10128-10133.

Kondo, M., Oyama-Okubo, N., Ando, T., Marchesi, E., and Nakayama, M. (2006). Floral scent diversity is differently expressed in emitted and endogenous components in Petunia axillaris lines. Ann Bot-London **98,** 1253-1259.

Langer, K.M., Jones, C.R., Jaworski, E.A., Rushing, G.V., Kim, J.Y., Clark, D.G., and Colquhoun, T.A. (2014). PhDAHP1 is required for floral volatile benzenoid/phenylpropanoid biosynthesis in Petunia x hybrida cv 'Mitchell Diploid'. Phytochemistry **103,** 22-31.

Maeda, H., Shasany, A.K., Schnepp, J., Orlova, I., Taguchi, G., Cooper, B.R., Rhodes, D., Pichersky, E., and Dudareva, N. (2010). RNAi suppression of Arogenate Dehydratase1 reveals that phenylalanine is synthesized predominantly via the arogenate pathway in petunia petals. The Plant cell **22,** 832-849.
Muhlemann, J.K., Klempien, A., and Dudareva, N. (2014a). Floral volatiles: from biosynthesis to function. Plant, cell & environment.
Muhlemann, J.K., Woodworth, B.D., Morgan, J.A., and Dudareva, N. (2014b). The monolignol pathway contributes to the biosynthesis of volatile phenylpropenes in flowers. The New phytologist **204,** 661-670.
Orlova, I., Marshall-Colon, A., Schnepp, J., Wood, B., Varbanova, M., Fridman, E., Blakeslee, J.J., Peer, W.A., Murphy, A.S., Rhodes, D., Pichersky, E., and Dudareva, N. (2006). Reduction of benzenoid synthesis in petunia flowers reveals multiple pathways to benzoic acid and enhancement in auxin transport. The Plant cell **18,** 3458-3475.
Ortiz-Serrano, P., and Gil, J.V. (2010). Quantitative comparison of free and bound volatiles of two commercial tomato cultivars (Solanum lycopersicum L.) during ripening. J Agric Food Chem **58,** 1106-1114.
Qualley, A.V., Widhalm, J.R., Adebesin, F., Kish, C.M., and Dudareva, N. (2012). Completion of the core beta-oxidative pathway of benzoic acid biosynthesis in plants. P Natl Acad Sci USA **109,** 16383-16388.
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. The Plant cell **18,** 1274-1291.
Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J., and Koes, R. (1999). Molecular analysis of the anthocyanin2 gene of petunia and its role in the evolution of flower color. The Plant cell **11,** 1433-1444.
Reck-Kortmann, M., Silva-Arias, G.A., Segatto, A.L., Mader, G., Bonatto, S.L., and de Freitas, L.B. (2014). Multilocus phylogeny reconstruction: New insights into the evolutionary history of the genus Petunia. Mol Phylogenet Evol **81,** 19-28.
Rockman, M.V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution **66,** 1-17.
Sablowski, R.W., Moyano, E., Culianez-Macia, F.A., Schuch, W., Martin, C., and Bevan, M. (1994). A flower-specific Myb protein activates transcription of phenylpropanoid biosynthetic genes. Embo J **13,** 128-137.
Spitzer-Rimon, B., Marhevka, E., Barkai, O., Marton, I., Edelbaum, O., Masci, T., Prathapani, N.K., Shklarman, E., Ovadis, M., and Vainstein, A. (2010). EOBII, a gene encoding a flower-specific regulator of phenylpropanoid volatiles' biosynthesis in petunia. The Plant cell **22,** 1961-1976.
Spitzer-Rimon, B., Farhi, M., Albo, B., Cna'ani, A., Ben Zvi, M.M., Masci, T., Edelbaum, O., Yu, Y., Shklarman, E., Ovadis, M., and Vainstein, A. (2012). The R2R3-MYB-Like Regulatory Factor EOBI, Acting Downstream of EOBII, Regulates Scent Production by Activating ODO1 and Structural Scent-Related Genes in Petunia. The Plant cell **24,** 5089-5105.
Stehmann, J.R. (1987). Petunia exserta (Solanaceae): Uma nova especie do rio grande do sul, brasil. . Napaea Rev. Bot. **2,** 19-21.
Stuurman, J., Hoballah, M.E., Broger, L., Moore, J., Basten, C., and Kuhlemeier, C. (2004). Dissection of floral pollination syndromes in Petunia. Genetics **168,** 1585-1599.
Tieman, D., Taylor, M., Schauer, N., Fernie, A.R., Hanson, A.D., and Klee, H.J. (2006). Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. Proceedings of the National Academy of Sciences of the United States of America **103,** 8287-8292.
Tohge, T., Watanabe, M., Hoefgen, R., and Fernie, A.R. (2013a). The evolution of phenylpropanoid metabolism in the green lineage. Crit Rev Biochem Mol Biol **48,** 123-152.
Tohge, T., Watanabe, M., Hoefgen, R., and Fernie, A.R. (2013b). Shikimate and phenylalanine biosynthesis in the green lineage. Front Plant Sci **4,** 62.
Tzin, V., and Galili, G. (2010). New Insights into the Shikimate and Aromatic Amino Acids Biosynthesis Pathways in Plants. Mol Plant **3,** 956-972.
Van Moerkercke, A., Haring, M.A., and Schuurink, R.C. (2011a). The transcription factor EMISSION OF BENZENOIDS II activates the MYB ODORANT1 promoter at a MYB binding site specific for fragrant petunias. Plant J **67,** 917-928.
Van Moerkercke, A., Haring, M.A., and Schuurink, R.C. (2012). A model for combinatorial regulation of the petunia R2R3-MYB transcription factor ODORANT1. Plant Signal Behav **7,** 518-520.
Van Moerkercke, A., Galván-Ampudia, C.S., Haring, M.A., and Schuurink, R.C. (2011b). A plasma membrane-localised ABC-transporter associated with floral fragrance in petunia petals (University of Amsterdam).
Van Moerkercke, A., Schauvinhold, I., Pichersky, E., Haring, M.A., and Schuurink, R.C. (2009). A plant thiolase involved in benzoic acid biosynthesis and volatile benzenoid production. Plant J **60,** 292-302.
Verdonk, J.C., Haring, M.A., van Tunen, A.J., and Schuurink, R.C. (2005). ODORANT1 regulates fragrance biosynthesis in petunia flowers. The Plant cell **17,** 1612-1624.
Verdonk, J.C., Ric de Vos, C.H., Verhoeven, H.A., Haring, M.A., van Tunen, A.J., and Schuurink, R.C. (2003). Regulation of floral scent production in petunia revealed by targeted metabolomics. Phytochemistry **62,** 997-1008.

Yoo, H., Widhalm, J.R., Qian, Y., Maeda, H., Cooper, B.R., Jannasch, A.S., Gonda, I., Lewinsohn, E., Rhodes, D., and Dudareva, N. (2013). An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. Nat Commun **4,** 2833.

Supplementary Note 9

Identification of conserved miRNAs in *Petunia axillaris* and *P. inflata* young flower buds and their verification in the *Petunia* genome sequence

Kitty Vijverberg[1,2]*,Nunzio D'Agostino[3], Tom Gerats[1]

[1]Radboud University, FNWI, IWWR, Heyendaalseweg 135, NL-6525AJ Nijmegen, The Netherlands
[2]Wageningen University, Biosystematics, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[3]Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Centro di ricerca per l'orticoltura, via Cavalleggeri 25, 84098 Pontecagnano, SA, Italy

*for correspondence: kitty.vijverberg@wur.nl

Running title: A catalog of conserved miRNAs in *Petunia*

Abstract

MicroRNAs (miRNAs) are short, endogenous RNAs that play key roles in the regulation of numerous biological processes, including development and metabolic pathways. Here, we present a catalog of conserved miRNAs identified in small RNA data sets from young flower buds of *Petunia axillaris* (*PaxiN*) and *P. inflata* (*PinfS6*) on the basis of identity to miRNAs known from *Arabidopsis* and Solanaceae, and confirmed by the presence of (a) corresponding *MIR loc(i)(us)* in the *Petunia* genome sequences. We identified 44 miRNAs, belonging to 30 families, and covering 140 *MIR loci*, representing 120 miRNA*s. An additional 13 miRNAs overlapped at *loci*, supposedly representing additional family members. Our results are consistent with those in tomato and potato that showed 34 miRNA families, corresponding to 96 and 120 *MIR* genes, respectively, and other plant species, including monocots. The results are highly comparable between *P. axillaris* and *P. inflata*, that showed all but four *MIR loci* present in both species, and one quarter of the *MIRs* having identical sequences, whereas the remainder showed minor (< 5 %) sequence variation only. MiRNA frequencies were also highly similar between *P. axillaris* and *P. inflata* and in good agreement with those reported for tomato and potato buds. MiRNA target genes included well-known candidates, such as miRNA156-*Squamosa Promoter Binding-like protein* (*SPL*), and hints to flower development and anthocyanin biosynthesis regulatory pathways. The presented *MIR* sequences allow for functional analysis by using mutants. The high conservation between *P. axillaris* and *P. inflata* indicates the usability of our miRNA catalog in the entire genus.

Keywords: flower buds, *MIR*, miRNA, miRNA*, miRNA frequency, miRNA target, *Petunia*

Introduction

MicroRNAs (miRNAs) are small ~ 20-22 nucleotides (nts) long RNAs that play key roles in the regulation of gene expression in most eukaryotes (Bartel, 2004; He and Hannon, 2004). They are particularly notable as regulators of growth and development by controlling the expression of genes that encode transcription factors (TFs) and other regulatory proteins (Dugas and Bartel, 2004; Chen, 2009; Chuck et al., 2009). MiRNAs are transcribed from *MIR genes* in the nuclear DNA, followed by processing of the primary transcript (*pri-miRNA*) into an intermediate (*pre-miRNA*, *MIR*) and subsequent miRNA/miRNA*-duplex (Papp et al., 2003; Park et al., 2005; Bologna et al., 2013; Budak and Akpinar, 2015). After transport to the cytoplasm, the miRNA acts via guiding an RNA-induced silencing complex (RISC) to its complementary target transcript to down-regulation its expression (Voinnet, 2009; Rogers and Chen, 2013), whereas the miRNA* usually is degraded.

Characteristics of the stem-loop (hairpin) structures of pri-miRNAs are useful for the (*de novo*) identification of *MIR loci* in the genome sequence, the miRNA*-sequence and miRNA subtype. Stem-loops of plant *MIR*s range between ~ 55-930 nt (146 nt on average), have a minimum free energy (MFE) lower than -20 kcal/mol and show conserved miRNA/miRNA* sequences (1-5 mismatches maximum) (Ambros et al., 2003; Meyer et al., 2008; Thakur et al., 2012). The expression also of an accompanying miRNA* supports the existence of the miRNA and its functionality in the species and tissue under study. Since the miRNA*s usually degrade, their frequency will be low as compared to the miRNA, although, examples exist in which the miRNA* also functions in gene regulation (Manavella et al., 2013) and the functional strand varies in different tissues (Kozomara and Griffith-Jones, 2014).

Petunia has since long been investigated for the mechanisms that underlie plant development, particularly of the flower and inflorescence (Rijpkema et al., 2006; Gerats and Strommer, 2008; Castel et al., 2010), and metabolic pathways, including the biosynthesis of anthocyanin and fragrance (Koes et al., 2005; Van Moerkercke et al., 2012; Albert et al., 2014). Most of these mechanisms involve regulation by TFs, and a role for miRNAs in the fine-tuning of these processes is obvious. Empirical evidence for a role for miRNAs in *Petunia* flower development comes from the involvement of *Phy-BLIND* (*BL*; *Phy-miR169c*) in the spatial restriction of homeotic class C genes to the inner two floral whorls (Cartolano et al., 2007). The regulation of the C genes by *BL* is indicated to be indirect, possibly via the targeting of a nuclear TF Y-*alpha* (*NF-YA*) that would activate the expression of the C genes. In *Arabidopsis*, evidence for the involvement of miRNAs in the regulation of anthocyanin production comes from the targeting of *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) genes by miR156, at least shown for *SPL*9 (Gou et al., 2011). Through down-regulation of the *SPL* expression, a *WDM*-transcriptional activation complex (β-transducin WD40 repeat / basic helix-loop-helix [*bHLH*] / myeloblastosis [*MYB*]) is stabilized, resulting in the accumulation of anthocyanin biosynthetic genes. *SPL* genes are deeply conserved in the plant kingdom and known to have important roles in the regulation of developmental phase change as well as flowering (Huijser and Schmid, 2011; Poethig, 2013). This might imply a link between the transition to flowering and secondary metabolism. Approving the presence of miRNA target sites in TFs that are known to be involved in development or metabolic pathways in plants is, therefore, important evidence for the functionality of miRNAs in these processes.

In the present study, we determined miRNAs and their frequencies in young flower buds of *Petunia axillaris* (*PaxiN*) and *P. inflata* (*PinfS6*), the ancestral parental species of the common *P. hybrida*, via high-

throughput sequencing technology and bioinformatics analysis. We focused on a catalogue of *conserved miRNAs* in *Petunia*, based on identity of small RNA reads to miRNAs known from *Arabidopsis thaliana* and related Solanaceae species: *Solanum lycopersicum* (tomato), *S. tuberosum* (potato), and *Nicotiana tabacum* (tobacco). Putative miRNAs were confirmed by the presence of (a) corresponding *MIR loc(i)(us)* in the *P. axillaris* and *P. inflata* genome sequences, a good secondary structure of the *MIR*s, and the presence of accompanying miRNA*s in the small RNA sequence datasets. We also predicted miRNA target sites in the genes annotated in *Petunia*, with emphasis on TFs and other proteins involved in flower development and anthocyanin pathways. We characterized a total of 44 conserved miRNAs, belonging to 30 families and covering 140 *MIR loci*, and in addition 13 variants that overlapped at one of the *MIR loci*, and present them as the conserved miRNA part of the miRNA-catalogue of *Petunia*.

Results

Characteristics of the small RNA data

To identify and characterize the miRNAs in *P. axillaris* and *P. inflata* flower buds, and to confirm their presences in the *Petunia* genome sequence, we sequenced ten small RNA-libraries: four of young buds (YB; Fig. 1) and one of larger buds (LB) per species. The libraries were each prepared from a pool of buds originating from four-five plants, and the sequences were obtained via Illumina high throughput technology. We generated ~ 5-10 million trimmed reads (Mrds) of between 15-34 nts per sample (Table 1), covering 65-80 % of the total number of sequences in the libraries (Appendix Table 1B).

The overall size distribution of trimmed reads showed a clear overrepresentation (55-60 %) of 24 nt sequences (Fig. 2). Second highest were 21 nt, 22 nt and 23 nt long reads, each representing 7-9 % of the reads. The samples, including the two species and two bud sizes, showed similar distribution patterns, with most variation found in the smallest size classes of 15-16 nts, the most abundant class of 24 nts, and the 28 nt class. This variation was mainly due to one *P. axillaris* sample that showed a reduced number of 24 nt reads and increased number of 15 nt reads (Fig. 2), possibly as a result of the younger plants used in this one sample (Appendix Table 1B). The data suggested a species-specific lower amount of 18 nt, 19 nt, 20 nt, 22 nt and 23 nt reads in *P. inflata* buds as compared to *P. axillaris* (two tailed t-test, $p < 0.05$), whereas the amount of 24 nt reads was slightly higher in *P. inflata*. These differences are apparently related to other than miRNAs small RNAs.

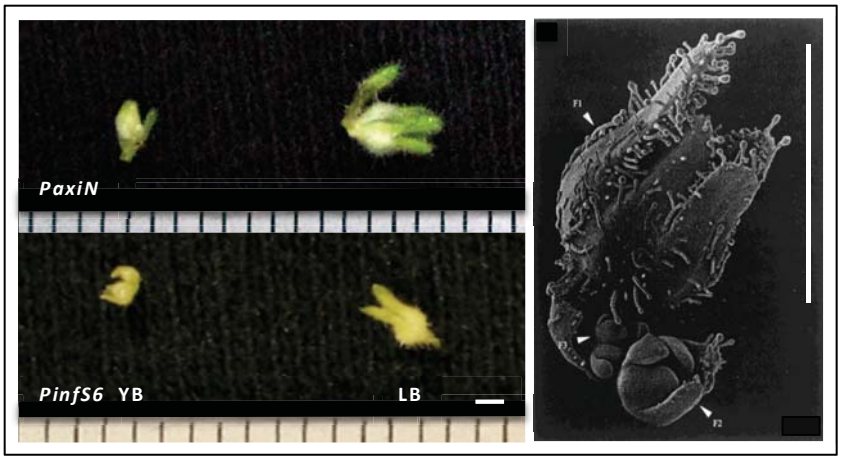


Figure 1. Developmental and size stages of *P. axillaris* (*PaxiN*) and *P. inflata* (*PinfS6*) young (YB) and larger (LB) flower buds used for small RNA sequencing. Right an electron microscopy picture of a *P. hybrida* bud with bracts removed (Maes *et al*., 1999), visualizing the stage used. In order to analyze the same stage, slightly smaller buds were analyzed for *PinfS6* (YB < 2 mm; LB 3-4 mm) then *PaxiN* (YB < 3 mm; LB 4-5 mm).Scale bars in mm.

Table 1. Read summary of small RNA libraries of *Petunia* young flower buds.

		Petunia axillaris					*Petunia inflata*				
Sampled ID		**YB1.1**[1]	**YB2.2**	**YB2.3**	**YB3.3**	**LB4.3**	**YB1.2**	**YB2.2**	**YB2.3**	**YB3.3**	**LB2.3**
15-34 nt reads	x1000 [2]	4895	8578	9579	5179	2694	4536	44522	9999	22527	10445
r,t,cp,mtRNAs	(%) [3]	1.9	3.0	1.7	1.8	2.8	3.0	4.1	3.7	2.5	3.7
Cleaned reads	x1000 [4]	4801	8318	9417	5085	2618	4399	42690	9626	21954	10063
Collapsed rds	x1000 [5]	2666	4301	4805	2894	1669	2658	16096	4963	10428	4713
miRNA reads	x1000 [6]	66	130	143	75	36	65	754	174	331	166
miRNA reads	(%)	1.4	1.6	1.5	1.5	1.4	1.5	1.8	1.8	1.5	1.7

See for more information on the reads and miRNA frequencies: Appendix Table 1B and Online Table 9.1.
[1] Numbers indicate plant selection and plant age, e.g., 1.1 = plants 1-5, age 9 weeks, see Appendix Table 1B
[2] Reads trimmed on quality, maximum of 2 ambiguities, adaptors, and length.
[3] Contaminating small RNAs in trimmed reads; [4] Reads cleaned from small RNA contaminants.
[5] Reads collapsed on the basis of sequence *and* length identity.
[6] Based on annotation against miRNAs of *Ath*, *Sly*, *Stu* and *Nta*, and also including c. 5 % miRNA*s

The small RNA contaminants concerned a small proportion of the reads only, 1.7-4.1 % (Table 1), and included the ribosomal small subunit RNAs (rRNAs), transfer RNAs (tRNAs), chloroplast RNAs (cpRNAs), and mitochondrial RNAs (mtRNAs). The larger part of these contaminants were cpRNAs and tRNAs, on average 38 % and 26 % in YBs of *PaxiN* and 27 % and 43 % in YBs of *PinfS6*, respectively. In LBs, the cpRNAs were increased to 47 % and tRNAs decreased to 21 %, likely reflecting a relatively higher amount of green parts in these larger buds.

Collapsing of identical reads, both in length and sequence, resulted in ~ 2-5 million unique reads per sample (~ 52 %; Table 1; Appendix Table 1B). Most of these reads, 86 %, occurred only once, but a reduced number of singletons was found in the most abundant size classes: 82 %, 80 %, 82 % and 72 % in the 21 nt, 22 nt, 23 nt, and 24 nt classes, respectively. Probably, the lower numbers of singletons in the high abundant classes reflect a larger proportion of functionally relevant small RNAs in these classes.

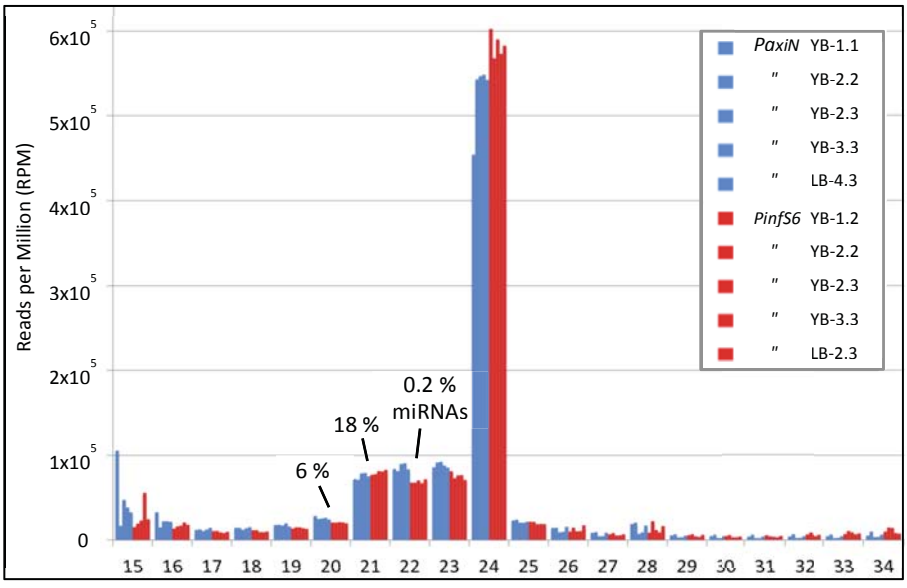


Figure 2. Size distribution of small RNA sequences of *P. axillaris* (*PaxiN*; blue) and *P. inflata* (*PinfS6*; red) young (YB) and larger (LB) buds, with average percentages of miRNAs per class indicated. The ten samples showed similar distribution patterns. The 24 nt reads were, by far, the most abundant class, but hardly included any miRNAs (< 0,01%). MiRNAs virtually all resided in the 21 nt and 20 nt classes. Sample numbers indicate plant selection and plant age (see Appendix Table 1B).

On average 1.5 % of the small RNA sequences in *P. axillaris* and 1.6 % in *P. inflata* were annotated as conserved miRNAs by showing a perfect match to miRNAs known from *Arabidopsis* (*Ath*), *Solanum spp. (Sly*, *Stu*) and/or tobacco (*Nta)* or, for 5.7 % of the annotated reads, to an accompanying miRNA* (Table 1; Appendix Table 1B). The miRNAs and miRNA*s resided almost exclusively in the 21 nt and 20 nt classes, covering 18 % and 6 % of the total number of reads in these classes, respectively (Fig. 2). In the 22 nt size class, only 0.2 % of the reads were annotated as miRNA(*)s, whereas miRNAs were absent from the 23 nt class and very low (< 0.01 %) in the 24 nt class. The high numbers of non-singletons/non-miRNAs in the 24 nt and other high abundant classes likely correspond to the other large class of small RNAs, the small interfering RNAs (siRNAs). Of the annotated miRNAs, 90.7 % measured 21 nts and 8.5 % 20 nts. Three miRNAs and two miRNA*s were found of 22 nts, and one each of 18 nts, 19 nts and 24 nts.

Identification of conserved miRNAs in *Petunia*

In order to identify and characterize the conserved miRNAs in *Petunia*, we searched for 100 % matches (CPM + BPM, see methods) of our small RNA sequence data of young flower buds to miRNAs known from *Arabidopsis* and related Solanaceae: tomato, potato, and tobacco (miRBase versions 20, 21), and the presence of accompanying *MIR loci (MIR genes)* in the petunia genome sequences. We found 81 miRNA candidates that met our criteria of being present in at least two of the ten libraries analyzed by a minimum of two reads, of which 44 showed *MIR loci* in the *P. axillaris* genome sequence. These 44 miRNAs belonged to 30 miRNA families and covered a total of 140 *MIR loci* in addition to five duplicated *loci* (Table 2; Appendix Table 2B). One additional candidate, miR6164, covered 20 *MIR loci* distributed over the genome sequence, all corresponding to long pre-miRNAs (~ 250 nts), most of them with a moderately supported secondary structure (Fig. 3; Appendix Table 2B). Since miR6164 has also a (very) low expression (< 1 RPM), it might be discutable. Of the remaining candidates, 13 overlapped at one of the already defined *MIR loci* (referred to as *variant2* and *v3*), seven lacked support in the genome sequence based on one or two mismatch(es) (indicated as *var*), and 16 apparently were miRNA*-sequences (Table 2; Appendix Table 2B).

To further confirm the identified miRNAs in *Petunia*, we assessed the presence of the miRNA*s in our small RNA sequence datasets. The 140 *MIR loci* in *PaxiN* represented 120 miRNA*s, including fifteen that were shared by more than one *MIR*, namely, by two (n = 11), three (n = 3) or four (n = 1) (Appendix Table 2B). The sharing of miRNA*s usually occurred by *MIRs* of the same miRNA, except for miRNA*168a-c that was shared by both family members. Since miRNA*168a-c also showed a high expression, it might be that this species actually is the (or also a) miRNA. The presence of a miRNA* was confirmed for all but two of the 25 miRNAs with a high (> 20 RPM) expression (Table 2; Appendix Tables 1B, 2B). One of the two highly expressed miRNAs that lacked confirmation by miRNA*s, miR319h,I, showed a high frequent miRNA* variant of one nt shorter, possibly being the 'true' miRNA* in this case. Of the 19 miRNAs with a low (< 20 RPM) expression, 13 were confirmed by showing at least two miRNA*s in a minimum of two of the ten libraries investigated. Confirmation of the other six miRNAs remained unresolved due to their low frequency. Of the 13 miRNA variants, the six with a moderate to high expression (1 - 70 RPM) were confirmed by the presence of miRNA*s in the small RNA datasets, whereas the two with the highest expression (80 - 300 RPM), miR319a-h *v2* and miR319h *v3*, were not (Appendix Table 2B). Possibly, the lack of miRNA* expression of these two variants, as also of miR394,

Table 2. Catalog of Conserved miRNAs in *Petunia*

Family	Member	Mature Sequence	nt	Pre # [1]	Star # [2]	Pre # [3]	*Ath*	*Sly*	*Stu*	*Nta*	RPM [4]	* [5]	N [6]
(*P. axillaris*)						(*P. inflata*)							
miR0156	a-g	UGACAGAAGAGAGUGAGCAC	20	7	5		a-g	d	e,g-k	a-e	1-5	±	
	j,l	UGACAGAAGAGAGAGAGCAC	20	3	3		h	-	-	-	~ 1	-	
miR0157	a-e	UGACAGAAGAUAGAGAGCAC	20	5	5		a-d	156 a-c	156 a-d	g-j	20-200	Y	T3t3
miR0159	a-c	UUUGGAUUGAAGGGAGCUCUA	21	3	3		a	Y	-	Y	>1000	Y	T
miR0160	a-d	UGCCUGGCUCCCUGUAUGCCA	21	3	2	+1 (+1)	a-c	a	a,b	a-c	200-1000	Y	T
miR0162	a	UCGAUAAACCUCUGCAUCCAG	21	1	1		a,b	Y	a,b	a,b	~ 200	Y	
miR0164	a-d	UGGAGAAGCAGGGCACGUGCA	21	4	4		a,b	a,b	-	a,b	20-200	Y	T
	g,h	UGGAGAAGCAGGGCACAUGCU	21	2	2	(+1)	-	-	Y	c	< 1	±	
miR0166	a-h	UCGGACCAGGCUUCAUUCCCC	21	8	6		a-g	a,b	a,c, d	a-h	>1000	Y!	T3t5
	k	UCGGACCAGGCUUCAUUCCUC	21	1	1		-	c	b	-	>1000	Y!	t
miR0167	a-e	UGAAGCUGCCAGCAUGAUCUA	21	5	5	(+1)	a,b	a	a-d	d,e	20-200	Y	T t
miR0168	a	UCGCUUGGUGCAGGUCGGGAA	21	1	1		a,b	-	-	d,e	200-1000	Y!	T3t4
	b,c	UCGCUUGGUGCAGGUCGGGAC	21	2	"		-	a,b	-	a-c	~ 200	"	t4 [8]
miR0169	a,b	CAGCCAAGGAUGACUUGCCGA	21	2	2		a	c	-	a-p	< 1	-	T
	c,d	CAGCCAAGGAUGACUUGCCGG	21	2	2		b,c	a	-	q-s	1-5	±	T6 t
	g-x	UAGCCAAGGAUGACUUGCCU (A)	20 (21)	17 (+2)	17	-3 +1 (+2)	*var*	d,e	a-h	t	< 1	±	
miR0171	a-g	UGAUUGAGCCGUGCCAAUAUC	21	7	7	-1	-	a	a,e	c	20-200	Y!	t
	j,k	UUGAGCCGCGCCAAUAUCACU	21	2	1		-	d	-	b	1-5	Y	
	n,o	UUGAGCCGCGUCAAUAUCUCU	21	2	2		-	-	b	-	< 1	-	
miR0172	a-g	AGAAUCUUGAUGAUGCUGCAU	21	7	6	(+2)	a,b	a,b	a,b, e	c-i	20-200	Y	t
	j,k	GGAAUCUUGAUGAUGCUGCAG	21	2 (+1)	2		-	-	d	-	20-200	Y	T t
miR0319	a-e	UUGGACUGAAGGGAGCUCCCU	21	5	5		a,b	b	a	a,b	>1000	Y	t
	h,i	UUGGACUGAAGGGAGCUCCU	20	2	2	(+2)	c	c	b	-	~ 200	±	[9]
miR0390	a-c	AAGCUCAGGAGGGAUAGCGCC	21	3	2		a,b	b	-	b,c	200-1000	Y	T5t2
miR0393	a-c	UCCAAAGGGAUCGCAUUGAUCC	22	3	2	(+1)	a,b	-	Y	-	~ 20	±	T
miR0394	a-c	UUGGCAUUCUGUCCACCUCC	20	3	2		a,b	Y	384 Y	Y	200-1000	- !	t
miR0395	a-m	CUGAAGUGUUUGGGGGAACUC	21	12 (+2)	6	-4 +1 (+13)	a,d, e	a,b	a-j	a-c	1-5	±	

Family	Member	Mature Sequence	nt	Pre # [1]	Star # [2]	Pre # [3]	*Ath*	*Sly*	*Stu*	*Nta*	RPM [4]	* [5]	N [6]
(*P. axillaris*)						(*P. inflata*)							
miR0396	a,b	UUCCACAGCUUUCUUGAACUG	21	2	2		a	a	-	a	200-1000	Y	t1
	c,d	UUCCACAGCUUUCUUGAACUU	21	2	2		b	b	Y	b,c	20-200	Y	
miR0397	a,b	AUUGAGUGCAGCGUUGAUGA	20	1	1	+1	-	Y	Y	-	< 1	±	
miR0398	a,b	UGUGUUCUCAGGUCACCCCUU	21	2	2	-1	a	-	-	-	< 1	-	[10]
	e	UAUGUUCUCAGGUCGCCCCUG	21	1	1	(+1)	-	-	a	-	1-5	±	
	h	CAGGGGCGACCUGAGAACACA	21	1	1		-	-	-	-	~ 1	-	
miR0399	a-c	UGCCAAAGGAGAGUUGCCCUG	21	3	3		b,c	-	-	-	< 1	±	[11]
	c-g	GGGCUACUCUCUAUUGGCAUG	21	4	2		-	-	a-h k-o	-	< 1	±	[11]
miR0403	a	UUAGAUUCACGCACAAACUCG	21	1	1		Y	-	-	-	200-1000	Y	
miR0408	a	UGCACUGCCUCUUCCCUGGCU	21	1	1		-	-	b	Y	~ 1	±	
miR0477	a	CCUCUCCCUCAAGGGCUUCUC	21	1	1	part	-	-	a	-	~ 1	±	[12]
miR0479	a	UGAGCCGAACCAAUAUCACUC	21	1	1		-	-	Y	-	~ 1	±	
miR0482	a	UUUCCAAUUCCACCCAUUCCUA	22	1	1		-	a	a	a,c	20-200	Y	
miR0827	a	UUAGAUGAACAUCAACAAACA	21	1	1		-	-	Y	Y	20-200	Y	
miR2111	a,b	UAAUCUGCAUCCUGAGGUUUA	21	2	2		a,b	-	-	-	~ 1	-	
miR6149	a	UUGAUACGCACCUGAAUCGGC	21	1	1		-	-	-	a,b	20-200	±	
miR8016	a	AUUUUUGAAUGGAAGGCCCAUGUG	24	1	1		-	-	Y	-	5-20	Y	
Possible miRNA in *Petunia*													
miR6164	a-t	CCUCCGUUUCAAUUUAUGUGA	21	20 (+1)	19	-11 +10 (+4)		-	-	a,b	< 1	±	
MicroRNA variants in *Petunia*						**Variation [7]**							
miR0157	a-e *v2*	**U**UGACAGAAGAUAGAGAGCAC	21	5		+ 5'-U	a-c	156 a-c	156 b-d	g-j	20-200	Y	t [13]
miR0171	a-g *v2*	UUGAGCCGUGCCAAUAUCACG	21	3		3 nt shift	b,c	b	d	-	20-200	Y	
	j,k *v2*	UUGAGCCGCGCCAAUAUCAC–	20	2		- 3'-U	-	d	-	b	~ 1	Y	
	j *v3*	UGAUUGAGCCGCGCCAAUAUC	21	1		3 nt shift	a	-	-	-	< 1	-	
miR0172	a-g *v2*	AGAAUCUUGAUGAUGCUGC––	19	7		- 3'-AU	a,b	a,b	a-c,e	c-i	~ 1	Y	
	j,k *v2*	GGAAUCUUGAUGAUGCUG–––	18	2		- 3'-CAG	e	-	d	j	1-5	Y	
miR0319	a-h *v2*	CUUGGACUGAAGGGAGCUCC	20	6		1 nt shift	b	a,b	a,c	a,b	200-1000	-	
	h *v3*	UUGGACUGAAGGGAGCUCCU**U**	21	1		+ 3'-U	c	c	-	-	20-200	-	
miR0390	a-c *v2*	AAGCUCAGGAGGGAUAGCGC–	20	3		- 3'-C	a,b	b	-	b,c	5-20	Y	

Family	Member	Mature Sequence	nt	Pre # [1]	Variation [7]	*Ath*	*Sly*	*Stu*	*Nta*	RPM [4]	* [5]	N [6]
miR0397	a *v2*	UCAUUGAGUGCAGCGUUGAUG	21	1	2 nt shift	a	-	-	-	< 1	±	
miR0398	a,b *v2*	UUGUGUUCUCAGGUCACCCCU	21	2	1 nt shift	a	-	b	-	~ 1	-	
	h *v2*	UGUGUUCUCAGGUCGCCCCUG	21	1	RC [14]	b	-	-	Y	< 1	±	
miR2111	a,b *v2*	UAAACCUCAGGAUGCAGAUUA	21	1	RC [14]	-	-	-	-	<< 1	-	

See for more information on the miRNA Catalog of Petunia: Appendix Tables 2B to 2F and Online Table 9.2.

[1] Number of pre-miRNA types found in *P. axillaris*, with values in brackets indicating exact duplicates, either true or, most probable, as a result of the assembly

[2] Number of miRNA* sequences, reflecting that some of the miRNA* are shared by more than one pre-miRNAs

[3] Differences in the number of pre-miRNA types in *P. inflata* as compared to *P. axillaris*, with values between brackets being (near) exact duplicates, either true or as a result of the assembly.

[4] Summary of Expressions of the miRNAs, see also Appendix Table 1B

[5] Summary of Expressions of the miRNA*s, see also Appendix Table 1B; Y! = Yes, > 100 RPM; Y = Yes, 0,4 – 15 RPM; ± = Yes, by at least two reads in two of the ten small RNA libraries analyzed in case of low expressed miRNAs

[6] Notes; summary of frequencies of conserved miRNAs in *Petunia* reported by Tedder et al., 2009: T < 5 clones; t = idem of one nt longer; prefix 1 ~ 10 clones, 2 ~ 20 clones, etc. of a total of 1 Mrds

[7] miRNA variant of a family member already defined in the upper part of the Table, localizing at the same *MIR locus*, with the type of variation indicated: either a 1-3 nt length difference or a 1-3 nt shifted localization.

[8] miR*168 a-c is shared by both family members and highly expressed, suggesting that this miRNA*, actually, is the or also a miRNA

[9] miR*319 h,I is not found in the small RNA sequence data, but an alternative read of one nt shorter is present at high frequency, possibly representing the miRNA* in this case

[10] miR398 is absent from *P. inflata*, showing one mismatch in the mature miRNA sequence, whereas an exact copy of the miR*398 is found, suggesting that miR*398 actually is the miRNA in *Petunia*

[11] miR399 *c* is found as a miRNA as well as a miRNA*

[12] MIR477 is present as a pseudo-locus in *P. inflata*, showing only the first 40 nt, confirmed by the absence of miR477 in the small RNA sequence data

[13] Most miRNA variants were confirmed by the presence of miRNA*s of one or more family members in our small RNA sequence datasets, suggesting that some of them are likely the miRNAs encoded by that particular *MIR locus* instead or, for some others, both can be expressed from the same locus

[14] Reverse complement

with high expressions, is also explained by a shorter or slightly shifted miRNA*, which is possibly related to the miRNA length of 20 nt of the miRNAs concerned. Apart from these few exceptions, all miRNAs identified in *Petunia* are well supported by the genome sequences as well as availability of miRNA*s in the small RNA sequence data.

Conservation of the identified miRNAs between *P. axillaris* and *P. inflata*

To investigate the level of conservation of the identified miRNAs within the genus *Petunia*, we analyzed their presence also in the *P. inflata* genome sequence in comparison to *P. axillaris.* The results were in great agreement with each other in all, numbers of *MIR loci*, *MIR* sequences, and localization in the genome (Table 2; Appendix Tables 2B to 2F). This strongly supported the conserved nature of the identified miRNAs in general as well as in the two parental *Petunia* species, and indicated a high degree of similarity between the two petunia genome sequences as well as their valid assemblies. Two of the 140 *MIR loci* defined in *PaxiN* were absent in *PinfS6*, *MIR*171g and *MIR*398b, one occurred partly / as pseudogene, *MIR*477, and two were found in addition, *MIR*160d and *MIR*397b (Table 2; Appendix Table

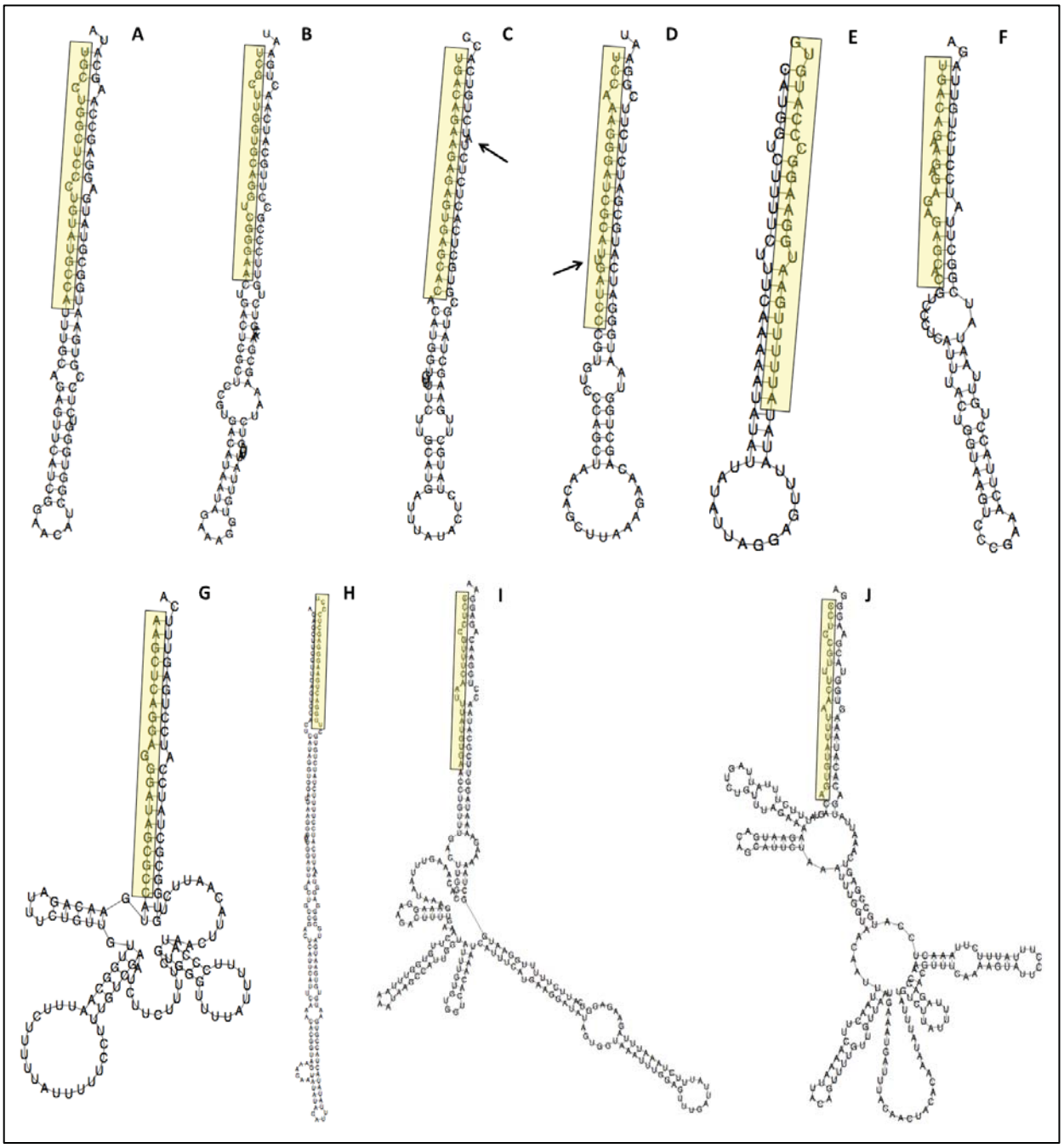


Figure 3. Examples of secondary structures of *MIR*s in *Petunia* (all *P. axillaris*), with miRNAs indicated in yellow. **(A, B)** *MIRs* with a short length and good structure, **(A)** *MIR*160b: 82 nt, -52 kcal/mol, **(B)** *MIR*168a: 102 nt, -41 kcal/mol; **(C, D)** *MIRs* of 20 nt and 22 nt, showing the characteristic bulge in the miRNA/miRNA*duplexes (arrows) to adjust for the 21 nt needed for excision by *DICERlike*1 (*DCL*1) (Ming-Hsuan and Shu-Hsing, 2012), **(C)** *MIR*156b (20 nt): 90 nt, -49 kcal/mol, **(D)** *MIR*393c (22 nt): 83 nt, -44 kcal/mol; **(E, F)** *MIRs* with the *highest* MFE-values, **(E)** *MIR*8016: 68 nt, -24 kcal/mol, short *MIR*, good structure, miRNA at second arm, **(F)** *MIR*156j (20 nt): 82 nt, -22 kcal/mol, several bulges, weak support; **(G)** *MIR* with a *medium* MFE-value and a bifurcating loop, *MIR*390b: 140 nt, -52 kcal/mol, moderate support; **(H)** *MIR* with the *lowest* MFE-value, MIR319e: 175 nt, -81 kcal/mol, good support, miRNA at second arm; **(I, J)** two *MIRs* of miR6164, long, with bifurcating loops, **(I)** *MIR*6164t: 233 nt, -76 kcal/mol, moderate support, **(J)** *MIR*6164n: 235 nt, -45 kcal/mol, weak support. **A**, **B**, **D**, **E** and **G** are well-supported by expression of the miRNA and miRNA*, **C** is supported by a low expression, **H** has a high expression but lacks miRNA* expression, and **F**, **I**, and **J** have a low support by expressions (see Appendix Tables 1B and 2B).

2C). *MIR*398a showed the same miRNA* for both *Petunia* species, but differed at the miRNA-site by one mismatch. This was supported by the absence / presence of the accompanying miRNA(*)s in the small RNA sequence datasets (Appendix Table 1B), suggesting that miR*398a might be the functional transcript in *P. inflata* instead. Some variation was also found between *PaxiN* and *PinfS6* for miRNAs with > 10 *loci*, indicating a higher rate of evolution at (or due to) these high copy numbers. In addition, more exact and near exact duplicated *MIR loci* were found in *P. inflata*, 14 and 10, respectively, versus a total of five in *P. axillaris*, either being true duplicates or, probably more likely, a result of the more fragmented *P. inflata* genome assembly.

Complete sequence conservation between *PinfS6* and *PaxiN* was found at all mature miRNA sites, except for miR398a (see former paragraph), and at 90 % of the miRNA*-sites, whereas the other 10 % miRNA*s showed one or two mismatches (indicated by ', Appendix Table 2C). Some more variation was found in the loop sequences of which one quarter was identical between *PinfS6* and *PaxiN*, 65 % showed one to three mismatches or an indel (indicated by '), and 10 % showed up to six ('') or more (''') mismatches. The increased variation in the loop sequences is congruent with a reduced selection pressure at this region as compared to the miRNA/miRNA* duplex. Entirely conserved *MIRs* between *P. axillaris* and *P. inflata* were most prevalent for miR399, of which six *loci* were without divergence and the seventh showed one indel, and *MIR*479. All other miRNAs usually showed one to two *MIR*s without sequence divergence between *PaxiN* and *PinfS6* and an additional two to five *MIRs* with some variation, suggesting different levels of selection at different *MIR loci* from within the same family member. The miRNA* sequences that were unique for one or the other *Petunia* species were confirmed by the small RNA sequence data, e.g., miR*169d and miR397*a were found in *P. axillaris* only and miR*171c'-*v2* in *P. inflata* only.

Conservation between *P. axillaris* and *P. inflata* was also found in the organization of *MIR loci* in the genome sequences. Most of the *MIR loci* were localized separately, but 14 *clusters*, usually including two direct or indirect *MIR loci* from within the same family at a distance of less than 1000 to up to 200.000 nts were found (Appendix Tables 2B and 2C). These clusters particularly included sets of two *loci* from miR156, miR157 and miR159; arrays of five and six of the 17 *loci* of miR169g-x; all 12 *loci* of miR395a-m; two inverted sets each of two *loci* of miR396; and the seven *loci* of miR399. All these clusters were present in *P. axillaris* as well as *P. inflata*, including the same *MIR loci* and organized in the same way, although, often with some variation in the distances between the *loci* in a *cluster*. This either could be a result of a lack of selection pressure at the intergenic regions or possibly of differences in the assemblies between the two species.

Overall, the miRNA catalog presented here in Table 2, with additional information in Appendix Tables 2B to 2F and Online Table 9.2, is well supported by our data and highly conserved between *P. axillaris* and *P. inflata*.

Characteristics of the conserved miRNAs identified in *Petunia*

The conserved miRNAs identified in *Petunia* showed a number of general characteristics, including: (1) Usually one to two, and occasionally three, *family members* (different miRNAs of the same family) were found (Table 2). For those families for which also one or more variant(s) were detected, the actual number of family members could be slightly higher, e.g., four for miR171. (2) Most miRNAs represented

one to three *MIR loci* in the genome, whereas nine covered four to eight *loci*, and three showed > 10 *loci*: miR169g-x, miR395a-m, and miR6164. (3) The *length* of the *MIRs* usually measured between 70-130 nts, with ~ 15 % being longer, up to ~ 250 nts. The longest *MIRs* were found for miR159 and miR319 (~ 170 nts), one copy of miR166, and all 20 copies of miR6164 (200-250 nts). (4) The *minimum free energy (MFE)* of *MIRs* varied to a certain level with length, usually being -30 to -55 kcal/mol for 70-130 nt long sequences and down to -80 kcal/mol for 170-250 nt long sequences (Appendix Tables 2B and 2C; Fig. 3). Only two pre-miRNAs showed a MFE > -28 kcal/mol, *MIR*8016, likely as a result of its short length of 68 nts, and *MIR*156j, apparently because it is weakly supported. Of the 20 200-250 nt long *MIRs* found for miR6164, six showed a MFE of < -60 kcal/mol, whereas the others showed higher values, up to -33 kcal/mol, indicating that at least part of the *MIR*6164*s* could be functional. (5) A total of 14 *MIR clusters* was found in the genome sequences, each including two or more direct or inverted *MIR loci* from within the same or a related family at a maximum distance of 200,000 nts. All these clusters were present in *P. axillaris* as well as *P. inflata*, similarly organized, indicating their evolutionary importance / conservation.

MiRNA expression in *Petunia* young flower buds

To investigate the involvement of the identified miRNAs in flower development and metabolic pathways, we analyzed their frequencies in young flower buds of *PaxiN* and *PinfS6*. The expression of 65 miRNAs, including the 44 conserved ones, 13 variants thereof, and eight possible candidates (Table 2), are given in Appendix Table 1B and summarized in Figure 4. The overall pattern of miRNA expressions was largely similar between the ten bud samples, including the two different species and two bud sizes. Four miRNAs showed a (very) high expression of > 1000 RPM: miR166a-h, miR159, miR319a-e, and miR166k, all well known for their involvement in floral organ development in other species (Nag and Jack, 2010; Luo et al., 2013). Relatively high expression of ~ 200-1000 RPM was found for seven additional miRNAs: miR160a, miR168a, miR390, miR394, miR396a,b, miR403, and miR319a-h *v2*, with miR160 also known for its role in flower development, whereas others have been associated to leaf and root development (Jin et al., 2013), and miR168 and miR403 are known for their regulation of *ARGONAUT*1 (*AGO*1) and *AGO*2, respectively, proteins that are essential for miRNAs and siRNAs to function (Mallory and Vaucheret, 2011). Eighteen miRNAs showed an expression between 20-200 RPM, including, for example, miR157 and miR172 that are known to decrease and increase, respectively, during juvenile to adult growth (Huijser and Schmid, 2011). The remaining half of the miRNAs showed expressions < 5 RPM, with among them miR169c (*BLIND*; Cartolano et al., 2007), a miRNA known in *Petunia* for its involvement in the restriction of C gene expression to the inner two flower whorls. The relatively low expression of *BLIND* illustrates the possible importance of low-abundant miRNAs in small RNA sequence data, possibly as a result of expression in only a subset of the entire bud sample.

MiRNAs most differently expressed between *bud sizes* included miR157 and miR172 (see former paragraph), miR159, miR319a-e, miR168a, miR8016, and miR171j,k-*v2* (indicated *b* in Fig. 4; two tailed t-test: * = $p < 0.05$, ** = $p < 0.001$). Most differently expressed miRNAs between *species* also included miR157 and miR172 and, in addition, miR827 and some of the lower expressed miRNAs, e.g., miR408, miR398e and miR156j,l (indicated *y* in Fig. 4). Some of the low expressed miRNAs were (virtually) uniquely expressed in one or the other *Petunia* species, e.g., miR2111, miR477, miR169g-x, and miR398a in *P. axillaris,* and miR164g,h, miR172-*var*, miR171j-*v3*, and miR2111-*v2* in *P. inflata*. For some miRNAs,

either one or the other variant was expressed per species, e.g., miR398a and miR2111. The overall miRNA expression pattern suggested that the differences between bud size were correlated to the differences between species, e.g., miR157 showed a high expression in larger buds and in *P. inflata* (Fig.

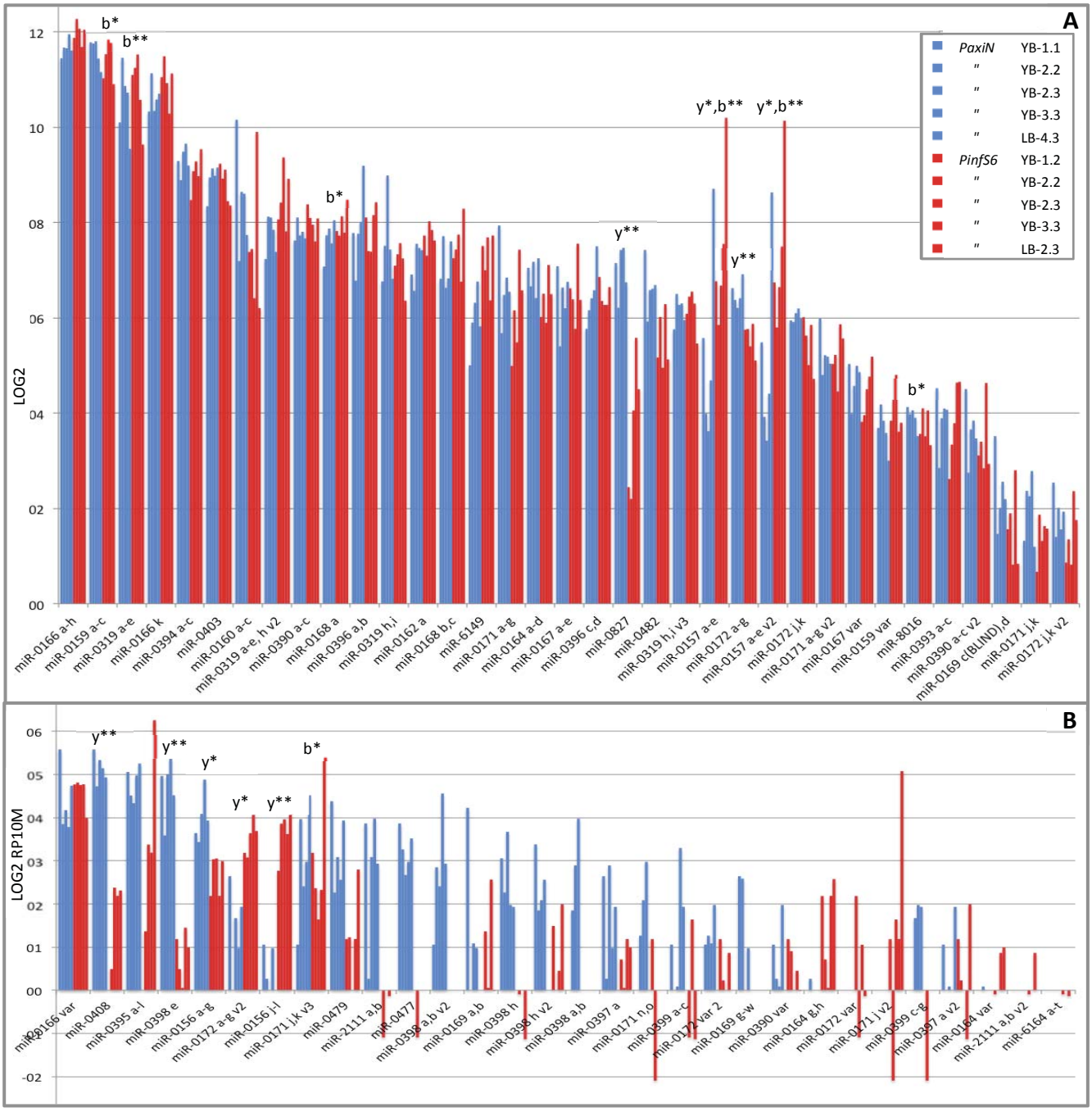


Figure 4. **Frequencies of conserved miRNAs in *P. axillaris* (*Pax*; blue) and *P. inflata* (*Pin*; red) young (YB) and larger (LB) flower buds. (A)** the 35 miRNAs with highest expressions (in Log2 RPM), **(B)** the 30 miRNAs with lower expression (in Log2 RP10M). Differences between YBs and LBs (*b*) and YBs of *Pax versus* YBs of *Pin* (*y*) are indicated for the more abundant miRNAs (two tailed t-test: * = $p < 0.05$; ** = $p < 0.001$). Highest expressions were found for miR166, miR159, and miR319, all known for their role in floral organ development. Most variation between the two species corresponded to the use of buds of a slightly later developmental stage in *PinfS6* as compared to *PaxiN*, e.g., miR157 and miR159.

4). Possibly, the bud sizes used for *P. inflata* corresponded to a slightly later developmental stage than for *P. axillaris*, coinciding with the differences in plant and flower size between the two species, and despite our efforts to correct for this by harvesting smaller buds for *P. inflata* (Fig. 1).

Expression of miRNA*s *versus* miRNAs in *Petunia* young flower buds

To verify the expression levels of miRNAs in young flower buds, and to resolve the miRNA species expressed, we also investigated the frequencies of miRNA* in the small RNA sequence data. Overall, the miRNA* expressions corresponded to ~ 5-10 % of the expression of miRNAs (Table 2; Appendix Table 1B; Fig. 5). Highest expressions of 80-300 RPM were found for miR*168a-c, miR*166a,b, miR166*k, and miR*171d, all in accordance with highly expressed miRNAs (Fig. 5 *versus* Fig. 4). The high expression of miR*168a-c was interpreted as the possibility that it (also) could have a miRNA function (see above). The high expression of miR*166a,b pointed to miR166a and/or miR166b as the highly expressed species from within miR166a-h in young flower buds. Evidence for expression of also (some of the) other species of miR166a-h comes from the detection of miR*166c, miR*166d, and miR166*e,f, although, with much lower frequencies (Appendix Table 1B). The high abundance of miRNA*171d similarly pointed to miR171d as the highly expressed species in young flower buds from miR171a-g, whereas the detection

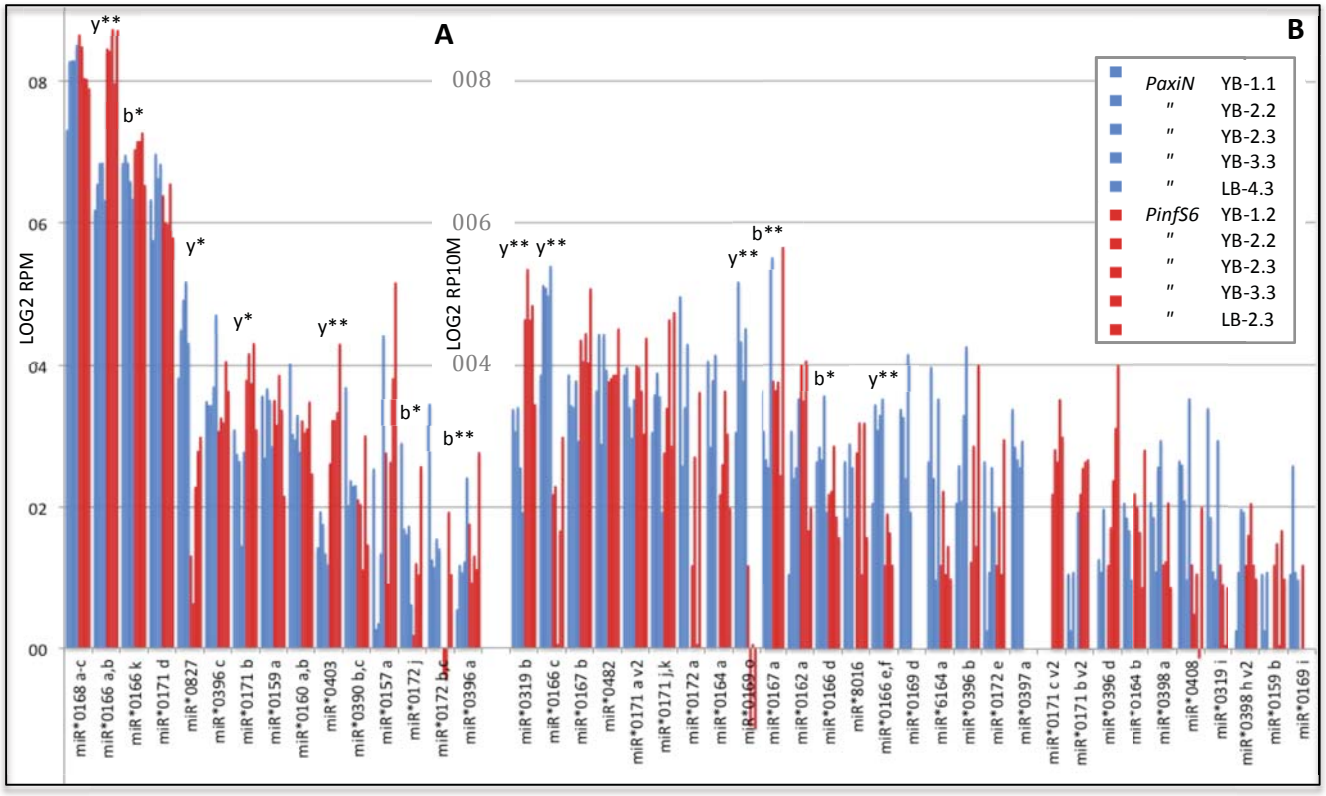


Figure 5. Frequencies of miRNA*s in *P. axillaris* (*Pax*; blue) and *P. inflata* (*Pin*; red) young (YB) and larger (LB) flower buds. (A) the 15 miRNA*s with highest expression (in RPM), **(B)** 29 miRNA*s with a lower expression (in RP10M). Differences between YBs and LBs (*b*) and YBs of *Pax versus* YBs of *Pin* (*y*) are indicated for the more abundant miRNA*s (two tailed t-test: * = $p < 0.05$; ** = $p < 0.001$). Highest expressions were found for miR*168a-c, miR*166a-b, miR*166k, and miR*171d, corresponding to miRNAs with a high expression (Fig. 4). The species-specific absences of miRNA*s, e.g., of miR*169d and miR*397a in *PinfS6* and miR*171c-*v2* in *PaxiN*, confirmed the results from sequences, showing a mismatch in these miRNA*s between the two species (Appendix Table 5.2).

of miR*171a and miR*171b at lower levels indicated that these species were also expressed. In general, the expression levels of the miRNA*s supported those of the miRNAs, and usually one of the miRNA species of a family member was identified at high frequency and one or more of the others species of that family member with a low frequency in addition.

Differences in miRNA* expression found between *bud sizes* or between *species* did not always correspond to the differences found in miRNA expression (*b* and *y* in Fig. 5 *versus* Fig. 4). For example, miR*166a,b and miR*403 were significantly increased in *P. inflata* as compared to *P. axillaris*, whereas miR166a-g and miR403 showed similar expression in the two species, and miR*167a was higher expressed in larger buds, while miRNA167 was not. The expression of miR827* and miR827, on the other hand, both showed a high expression in *P. axillaris*. Possibly, the differences in miRNA* *versus* miRNA expression become more pronounced at lower expression levels or remain undiscovered due to the inclusion of more than one miRNA species in the miRNA expression. Another possibility is that the differences between miRNA* and miRNA expression correlate to the developmental stage of the buds, e.g., that miR*166a,b, miR*403, and miR*167a are less efficiently degraded in buds of a later developmental stage (*P. inflata*), a possibility that has to be confirmed.

Putative miRNA targets in *Petunia*

In order to support the functional relevance of the miRNAs identified, we predicted their putative targets in the genes annotated in *Petunia* (see Supplementary Note 1, Bombarely et al.), using TargetFinder (Fahlgren et al., 2007). We found targets for most miRNA family members (Table 3; Online Table 9.3), except for miR160, miR169c,d, miR398e and miR403. Additional targets were found after adding 3'-untranscribed regions (3'-UTR) sequence data to the analysis, if available (e.g., for *BLIND*, Cartolano et al. 2007). These included, apart from the known interaction of miR169c,d with *NF-YA*, a target of miR403, and additional candidates of miR167, miR394, miR395, and miR482 (Online Table 9.3 end of Table). In plants, miRNA target sites are often presented in the coding sequences, but also occur in 3' UTRs (Bartel 2004; Chen 2009). Additional miRNA target genes are, therefore, expected after isolating and analyzing the entire gene sequences, including UTRs. The target genes yet presented are valuable starting points for further analysis and verification, e.g., by detecting cleaved products through 5'-rapid amplification of cDNA ends (5'-RACE) or degradome sequencing.

The predicted miRNA target genes in *P. axillaris* and *P. inflata* were virtually identical (Online Table 9.3). Among them were the large transcription factor families known from other studies, e.g., miRNA156 - *SQUAMOSA PROMOTOR BINDING-like PROTEIN* and miRNA171 - *GRAS family TF*s interactions (Table 3; Gou et al., 2011; Huijser and Schmid, 2011; Sun et al., 2012; Chorostecki et al., 2012; Poethig, 2013). A few well-known target genes were lacking, e.g., the miR160 - *AUXIN RESPONSE FACTOR* (*ARF*) and miR162 - *DICER*-like interactions, possibly as a result of the absence of the particular (or part of the) sequence in our data sets. A number of interactions potentially involved in flower development was resolved (indicated in *light orange* in Online Table 9.3), such as miR157 - *MADS*-box TFs, or in anthocyanin pathways (indicated in *light purple*), including miR159 - *UDP-Glycosyltransferase* (3*GT* or 5*GT*), miR164g,h - *4-Coumarate CoA Ligase* (4*CL*), and miR172j,k - *CHalcone Isomerase* (*CHI*). These candidates are interesting, initial clues for further investigations of a possible role of miRNA regulation in flower developmental and metabolic pathways.

Table 3. Selection of putative miRNA Target Genes in *Petunia*

Family	Target Gene (0-2) [1]	Function [2]	Target Gene (2-4) [3]	Function
miR0156	squamosa promoter binding protein	leaf development, vegetative phase change, flower and fruit development, plant architecture, sporogenesis	*LIGULELESS*1 protein (*Pax*)	development of the ligule
miR0157	squamosa promoter binding protein	idem	MADS-box transcription factor	control of all major aspects of development
miR0159	-		myb domain protein 40	flavenoid synthetic pathway, stress response
miR0162	-		mediator of RNA polymerase II	mediates to bind *Pol* II to synthesize precursors of mRNA, most snRNA, miRNA
miR0164	NAC domain containing protein	stress perception regulator, developmental programs	4-coumarate:CoA ligase 2	biosynthesis of plant secondary compounds
miR0166	Homeobox-leucine zipper family	maintenance of floral organ identity		
miR0167	MATE efflux family protein	antiporter/trans membrane transporter activity	CYP71D5v2, mono-oxygenase (*Nta*)	metabolic pathways
miR0168			Argonaute family protein	defense response, gene silencing by RNA targeting
miR0169			basic helix-loop-helix (bHLH); CCAAT-binding TF subunit A (Nta, Phy)	regulation of many pathways during development; flower development, restriction C gene expression
miR0171	GRAS family transcription factor	root and shoot development, defense response, gibberellic acid signaling		
miR0172	AP2-like ethylene-responsive	cell differentiation, flower development, meristem maintenance		
miR0319	TCP family transcription factor	cell proliferation, circadian rhythm, negative regulation of leaf senescence	acetyl-CoA carboxylase 1	fatty acid metabolism
miR0390	Leucine-rich receptor-like protein	protein kinase activity, ATP binding		
miR0393	Leucine-rich receptor-like protein	protein kinase activity, ATP binding		
miR0394	FLAVIN-BINDING (F)-box protein	circadian rhythm, flower development, ubiquitination		
miR0395			sulfate transporter	
miR0396	pentatricopeptide repeat 336	chromatin assembly or disassembly (RNA-editing)	growth-regulating factor	
miR0397	laccase	lignin catabolic process, oxidation-reduction process, glucuronoxylan metabolism	pentatricopeptide repeat 336	chromatin assembly or disassembly, (RNA-editing)

Family	Target Gene (0-2)	Function	Target Gene (2-4)	Function
miR0398	GATA type zinc finger TF factor	gene expression regulation	4-coumarate: CoA ligase 2	biosynthesis of plant secondary compounds
miR0399	phosphate transporter 1	phosphate transport		
miR0403	-		AGO3 (*Sly*)	RNA silencing
miR0408	-		Copper-exporting P-type ATPase	anther development, pollination
miR0477	ATPase family		Zinc finger C-x8-C-x5-C-x3-H	zinc ion binding, nucleic acid binding,
miR0479	*GRAS* family transcription factor	root and shoot development, defense response, gibberellic acid signaling, intracellular signal transduction		
miR0482	-		K+ channel inward rectifying (*Stu*)	
miR0827	-		Cytochrome P450 superfamily	electron carrier activity, mono-oxygenase activity, iron ion binding
miR2111	Galactose oxidase /kelch repeat	hydrolase activity, manganese ion binding, phosphatase activity	AT-hook motif nuclear-localized	embryo and flower development, meristem structural organization
miR6149	F-box family protein	circadian rhythm, flower development, ubiquitination		
miR8016	*MATE* efflux family protein	antiporter/trans membrane transporter activity	DnaJ homolog subfamily C member	ATP-hydrolysis, stress response

See for more information on the putative miRNA Target Genes in *Petunia*, including DNA identification numbers and score values, Online Table 9.3

[1] The miRNA Target Genes with best support (Score 0 to 2) found among the genes annotated in *Petunia*

[2] Global description of the function, mainly based on The Arabidopsis Information Resource (TAIR)

[3] A selection of miRNA Target Genes with moderate support (Score 3 to 4) found among the genes annotated in *Petunia*, listed for their putative involvement in flower development or metabolic pathways

Discussion

Here, we present the first part of a miRNA catalog in *Petunia*, including the conserved miRNAs expressed in young flower buds (Table 2; Appendix Tables 2B to 2F; Online Table 9.2), covering *the large part of conserved* miRNAs in plants (Cuperus et al., 2011; Chorostecki et al., 2012). The data is well supported by the presence of *MIR loci (MIR genes*) in the *P. axillaris* and *P. inflata* genome sequences and miRNA and miRNA* sequences in the ten small RNA datasets analyzed. The catalog includes 44 miRNAs, belonging to 30 families, and covering 140 *MIR loci*, representing 120 miRNA*s. An additional 13 miRNAs overlap at one of the *MIR loci* and probably represent extra family members, and one miRNA that covers 20 *MIR loci* might be doubtful. Our results confirm 12 of the 15 conserved miRNAs reported in an earlier study on miRNAs in *Petunia* floral buds by Tedder et al. (2009; indicated by *T* in Table 2), and an additional eight that were among the miRNAs reported with one nt difference in that study (indicated by

t in Table 2). Two members of miR167 and one of miR390, each covered by one or two sequenced clones in the previous study, were not supported by our data. The current availability of the petunia genome sequences allowed us to provide the *MIR* sequences in addition (Appendix Table 2E).

The large part of the *MIR loci* is comparable between *P. axillaris* and *P. inflata*, in numbers, sequences, and organization in the genome (Table 2; Appendix Tables 2B, 2C and 2E), supporting their conservation and importance in *Petunia*. Only two of the 140 *MIR loci* were unique to *P. axillaris*: MIR171g and MIR398b, and two were unique and one partial to *P. inflata*: MIR160d, MIR397b and MIR477. This uniqueness was confirmed by the expression of miRNA*s. Complete sequence identity between *P. axillaris* and *P. inflata* was usually found for one or two *MIR loci* per family member, whereas one or more additional *MIRs* showed some sequence divergence (Appendix Table 2C, indicated by '). Apparently, there are different levels of selection at different *MIR loci* from within one family member, possibly reflecting preservation of function for some miRNAs and allowing functional divergence for others. A total of 14 *clusters* of two or more *MIR loci* was found in the petunia genome sequences and conserved between the two species (Appendix Tables 2B and 2C), indicating functional relevance, shared control, and/or collaborative functioning of these clustered miRNAs.

Petunia **miRNAs** ***versus*** **other species**

From the 44 conserved miRNAs found in *Petunia*, 25 are known in *Arabidopsis* (Table 2; Fahlgren et al., 2007), indicating their conservation at a wider taxonomic range, whereas the remaining 19 occur in tomato, potato and/or tobacco, suggesting a more Solanaceae specific nature (Table 2; miRBase 21). Some miRNAs found in tomato and potato: miR3627, miR4376, and miR4414, or reported to be Solanaceae- or Solanum-specific: miR1919, miR5300, the multi-locus miR5303, and miR5304 (Tomato Genome Consortium, 2012; Gu et al., 2014), were not found in *Petunia*. Possibly, these miRNAs are not expressed in *Petunia* young flower buds, although, this is less likely since they show moderately to high expression in flower buds of tomato and/or potato (Appendix Table 4). Another possibility is that these miRNAs have a (slightly) different sequence in *Petunia* and, therefore, remained undetected in our annotation based on exact similarity, or they do not exist at all in *Petunia*, contradicting to being Solanaceae specific. Current analysis of our small RNA datasets in order to catalog also the *Petunia* specific miRNAs will further shed light on this.

The numbers of *MIR loci* in *Petunia*, summed over the miRNA families, are in good agreement with those found in tomato, potato, and *Arabidopsis* (Fig. 6; Table 2; Tomato Genome Consortium, 2012; Chorostecki et al., 2012; miRBase 21). *Petunia* showed relatively many *loci* for miR164, miR169, miR171, miR319, and miR6164, and few for miR162 (one instead of two), miR399 (in comparison to potato), and miR482 (absent in *Ath*). Despite these differences, e.g., a single copy for miR162 and up to five copies for miR164 are not unique in the plant kingdom (miRBase 21). Apparently, some variation in *MIR* copy number is common among plants, possibly in order to allow regulation by miRNAs to evolve.

Interestingly, most of the *MIR clusters* in the petunia genome sequences are conserved also outside the genus, some of them up to the monocots, including clusters of two or more *loci* of miR156 and miR159, miR167, miR169, miR395, miR396, and miR399 (Cui et al., 2009; Gu et al., 2014). The conservation of these clusters suggests an origin from before the split of the monocots and dicots, and indicates their evolutionary importance. Among these '*ancient miRNA clusters*', there are the four

families with the highest numbers of *MIR loci* (indicated *ac* in Fig. 6), suggesting that a long evolutionary history apparently is congruent with an increased number of *MIR loci*. It would be interesting to further investigate whether the miRNAs in a cluster share transcriptional control, particularly when they are in tandem and not too far apart, and/or whether they function in collaboration.

MiRNAs in young flower buds

The miRNA expressions in *Petunia* young flower buds correspond to a large extend to those found in tomato and potato (Appendix Table 4), supporting the importance of the particular miRNAs in aspects of flower development in the Solanaceae. Among the highest expressed miRNAs, frequencies of miR156/157, miR166 and miR167 were somewhat higher in tomato and potato as compared to *Petunia*, and of miR159, and particularly miR319, were lower. These differences could well be explained by the analysis of younger flower buds in *Petunia* (see Figs 1 and 4), as is also conform the use of slightly younger buds in tomato as compared to potato (Tomato Genome Consortium, 2012). Of the low expressed miRNAs, 25 were absent from the *Petunia* small RNA data sets based on 100 % identical sequences, whereas four were present in *Petunia* but not found in tomato and potato flower buds: miR477, miR479, miR6149 and miR8916. It would be interesting to compare and discuss these results again after having cataloged the *Petunia* specific miRNAs.

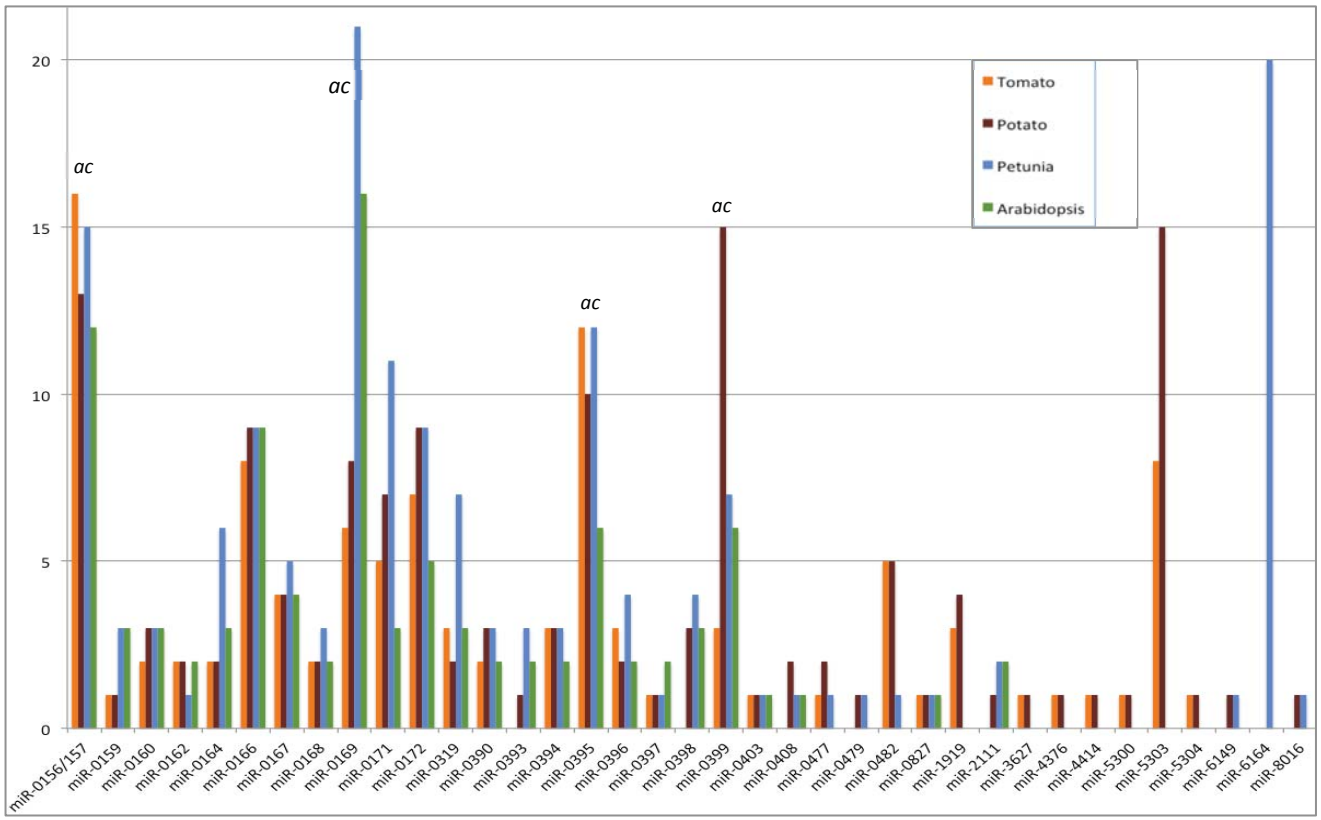


Figure 6. Number of *MIR loci* per miRNA summed per family, in *Petunia* (blue; this study), tomato (orange) and potato (brown; Tomato Genome Consortium, 2012) and *Arabidopsis* (green; Chorostecki et al., 2012). Note that 1/3rd of the families include *MIR loci* of more than one family member (Table 2). The results show congruence as well as some variation in the number of *loci* between the four species. *ac* = family that also contains an *ancient MIR cluster*, apparently corresponding to a high number of *MIR loci*.

In comparison to the miRNA frequencies found in *Petunia* flower buds by Tedder et al. (2009), most of our highly expressed miRNAs (200-5000 RPM, Table 2; Appendix Table 1B) were represented by the highest numbers of clones (60-100) in their study, including miR156/157, miR166, miR168, and miR390 (see also Table 2 last column). In terms of RPM, however, the expression levels of these miRNAs were ~ 10-100 times lower in Tedder et al. Possibly, the comparison of the two studies is influenced by the analysis of a lower number of sequences by Tedder et al. (~ 1 Mrds in total from two different libraries) and/or methodological differences. Strikingly low expressions were found for miR159 and miR319, each represented by one clone only, possibly implying the analysis of (much) older buds by Tedder et al., since the expression of miR159 and miR319 reduces significantly in larger flower buds (Fig. 4). Expression of miR169c (*BLIND*; Cartolano et al., 2007), a well-studied regulator of whorl identity in *Petunia* and *Antirrhinum*, was among the highest expressed miRNAs in Tedder et al. (58 clones, ~ 60 RPM), whereas it was moderately to low expressed in our study (~ 5 RPM). Moderate expression of miR169 was also found in tomato and potato (Appendix Table 4; Tomato Genome Consortium, 2012), and supported by our small RNA data in *P. hybrida* (not shown), and could possibly be expected if expressed in only a subset of the bud tissue. According to the function of miR*BL*, early in flower development, increased expression in (much) older buds is not expected, and not supported by our data (Fig. 4; Appendix Table 1B). In any case, it would be interesting to analyze the expression of miR169c and other miRNAs in series of bud stages in *Petunia* to further explore their role in flower development.

The miRNAs that are highly expressed in young flower buds of *Petunia* include those known for their involvement in flower development, e.g., miR166 in floral organ polarity, miR159 in flowering time and floral transition, and miR319 in cell differentiation, and more general regulatory aspects, e.g., miR394 in transcription and cell cycle control and miR403 in gene silencing (Sun et al, 2012; Luo et al, 2013). The highly expressed miR168 and miR403, both targeting a member of the *ARGONAUT* family (Mallory and Vaucheret, 2011), point to the importance of small RNA regulatory pathways in the (very) young flower buds, e.g., to bring about many and rapid transitions. In this context, it would be interesting to also investigate the siRNAs in the small RNA sequence datasets available.

The miRNA target prediction resolved many interactions known from other plant species, supporting a similar functioning of miRNAs in *Petunia* in most cases (Table 3; Online Table 9.3; Sun et al, 2012). The results offer a wealth of starting material for further investigation and confirmation of miRNA - target gene interactions in *Petunia*. A number of putative candidates in flower development and anthocyanin pathways are included in the list (Table 3; indicated in *light orange* and *light violet*, respectively in Online Table 9.3). Since the miRNAs presented in this paper concern the conserved miRNAs only, additional miRNA - target gene interactions are expected after resolving the *Petunia* specific miRNAs and adding them to the catalog. Also, additional miRNA target genes may be found after including additional 3'-UTR regions to the analysis. The available *MIR* sequences (Appendix Table 2E) allow for searching for mutants in the *Petunia* transposon library (Vandenbussche et al., 2008), to confirm the proposed miRNA - target gene interactions and investigate their potential role in regulatory pathways. The presented data, therefore, has a high potential to further understand regulatory aspects of flower and plant development in general and in *Petunia* in particular.

Methods

Plant material

Seeds of *Petunia axillaris* line S26 (*Pax*; *PaxiN; Paxi*S26) and *P. inflata* line S6 (*Pin*; *Pinf*S6) were obtained from R. Koes (Free University, Amsterdam, The Netherlands), and grown in soil under standard greenhouse conditions: ~ 21 °C and 16 h light/8 h dark. Young buds (YB) and larger buds (LB) were harvested (nine) 12 and 18 weeks after seeding, at c. similar developmental stages (Fig. 1): for *PaxiN*, YB < 3 mm, LB = 4-5 mm, for *PinfS6*, YB < 2 mm, LB = 3-4 mm, with bracts removed. Each sample consisted of a pool of buds from four-five plants, collected on ice, and transferred to liquid N_2, and RNA was isolated at the same day as harvesting. A total of four YB and one LB samples were used per species for the small RNA library preparation and sequencing.

RNA extraction and small RNA library preparation

Total RNA was isolated by a standard method using TRIzol® Reagent (Life Technologies), including a second precipitation step to increase RNA purity, and using 20 minutes centrifugation, 14.000 rpm, to optimize small RNA recovery. Samples were stored in the second precipitation, in 0.3 M Sodium Acetate (NaAc)/70 % Ethanol (EtOH), at -80 °C, upon library preparation. RNA was suspended in 10-30 µl DEPC-MQ to a final concentration of ~ 1 µg/µl, and checked on agarose for a 25S to 18S ribosomal peak ratio of > 4:5 to be used for library preparation.

Small RNA libraries were prepared from 1.2 µg total RNA, using the Illumina compatible NextFlex™ Small RNA Sequencing Kit (version 1; Bioo Scientific; Sanbio BV, Uden, The Netherlands; Cat. no. 5132) and accompanying Small RNA Barcodes Primers (Cat. no. 5133). This protocol included an 18 h long 3'-adapter ligation of pre-adenylated oligonucleotides, with a modified ligase, to optimize binding to plant miRNAs and other small RNAs that contain 2'-O-methylated 3'-ends (Munafó and Robb, 2010; McCormick et al., 2011). After the 5'-adapter ligation, the recommended purification step was included, using the RNA Clean and Concentrator Kit (Zymo Research; BaseClear, Leiden, The Netherlands; Cat. no. R1015). We amplified the cDNAs by 14-16 PCR cycli. The ~ 150 nt long small RNA libraries were isolated from 3 % NuSieve™ (Lonza, Basel, Switzerland)/1 % agarose gels, and extracted by the QIAquick Gel Extraction Kit (Qiagen Benelux BV, Venlo, The Netherlands; Cat. no. 28706). Libraries were eluted in 40 µl pre-warmed buffer of the Kit, and precipitated in NaAc/EtOH, at -20°C, after addition of 1 µl co-precipitant of the NextFlex Kit.

Small RNA sequencing and data cleanup

After re-suspension of the small RNA libraries in 10 µl MQ, they were sent to BaseClear (Leiden, The Netherlands) for Illumina HiSeq 2500 sequencing. Quality Control included a BioAnalyzer measurement (Agilent Technologies) and qPCR, on which basis the samples were pooled in equal amounts to 18 samples/lane in order to provide a minimum of ~ 5 Million Reads/library. Sequence data was de-multiplexed and reads delivered as '.fastq' files together with the run-statistics.

Raw small RNA sequence data was trimmed on quality at a cut-off value of > 0.05, a maximum of 2 ambiguities, adapters, and length between 15-34 nt, using the CLC Genomic Workbench (version 7.0; http://www.clcbio.com). Small RNA contaminants were filtered-out on the basis of a BLAST search

against a database including tRNAs from *Arabidopsis thaliana* (*Ath*) (http://lowelab.ucsc.edu/GtRNAdb/Athal) and Rfam 11.0 (http://www.sanger.ac.uk; subset RF00005), small subunit rRNAs from Rfam (subsets RF00001;5S rRNA and RF00002;5_8S_rRNA), mtDNAs from the tomato mitochondrial genome (http://www.mitochondrialgenome.org) and GenBank: *Ath*, *Solanum lycopersicum* (*Sly*), *Nicotiana tabacum* (*Nta*), and *Vitis vinifera* (*Vvi*), and cpDNAs from GenBank: *Ath*, *Sly*, *Solanum tuberosum* (Stu), *Nta*, *Glycine max* (*Gma*), *Medicago truncatula* (*Mtr*) and *Vvi*, using an identity of ≥ 80 %.

MicroRNA annotation and verification

Annotation of microRNAs was performed by using a suite of custom Perl scripts and, in addition, miRDeep-P (version 1.3; Yang and Li, 2011) combined with Bowtie (version 1.0.1; Langmead et al., 2009), and CLCbio, included the following steps: (1) merge and count reads of identical sequence and length, (2) annotate unique reads based on miRNAs known from related species: *Ath*, *Sly, Stu,* and *Nta*, downloaded from miRBase (version 20; http://www.miRBase.org), and (3) verify annotated reads on the basis of the *PaxiN* and *PinfS6* genome sequences and the analysis of miRNA* presences in the small RNA sequence data. At step (2), we included the *core perfect matches* (CPM = identical in sequence and length) and *big perfect matches* (BPM = identical in sequence, but 1-2 nt longer), whereas the *small perfect matches* (SPM = identical in sequence, but 1-2 nt shorter) were left out, since they could not properly be assigned to a miRNA family member. As such, we identified the *conserved* miRNA complement in the small RNA sequence data of *Petunia* young flower buds.

For step (3), we first identified the pre-miRNA sequence(s) (*MIRs*) in the *P. axillaris* genome sequence, using CLCbio, the criteria for miRNA annotation (Ambros et al., 2003; Meyers et al., 2008; Kozomara and Griffiths-Jones, 2014), and the following pipeline: a BLAST search for 100 % matches of the mature miRNA sequence in the genome sequence; a motif search for > 70 % (> 60 % in case of 20 and 22 nt miRNAs; Ming-Hsuan and Shu-Hsing, 2012), but < 100 % identical inverted repeats at a maximum distance of 600 nt from the miRNAs in the scaffolds involved; and the determination of secondary structures in RNAfold of the Vienna RNA package (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi, Lorenz at al., 2011). Sequences that showed a correct miRNA/miRNA* duplex, a hairpin structure, and with a minimum free energy (MFE) of < -25 kcal/mol (Thakur et al., 2011) were saved: 2nd structure, MFE-value, and miRNA* sequence, including the characteristic two nt 3'-overhang. Secondly, we identified the *MIR*s by using miRDeep-P, with a window of 400 nt and build-in settings, providing a list of 400 nt sequences with a miRNA-match, sequences of the miRNA, loop, miRNA*, pre-miRNA, and flanks. MFE-values of the pre-miRNAs were calculated in RNAfold, and their positions in the scaffolds determined in a separate analysis. We defined the cut-off values for 'true' pre-miRNAs in miRDeep on the basis of comparison to results of CLCbio: (1) both, those with a star-score of 3.9 as well as -1.3; (2) one, if two different 400 nt selections were made at the same miRNA location, the one with the best secondary structure, usually the shorter one; (3) those with a reliable secondary structure after visual inspection (Fig. 3), particularly excluding sequences > 250 nt with a low MFE more as a result of length than of proper annealing; (4) the one suggested by miRDeep, if more than one mature miRNA matched to the same location (the other was then listed as a *variant2* or *v3*); (5) the CLCbio result, if a pre-miRNA based on a miRNA* was given rather than a miRNA. This definition was extrapolated to the subsequent analyses, including the identification of pre-miRNAs in the *P. inflata* genome.

MicroRNA naming

The miRNAs were named according to the guidelines from miRBase (Kozomara and Griffiths-Jones, 2010, 2014) and annotation to miRNAs known from other species. Accordingly, all miRNA species from one miRNA family member shared the mature miRNA sequence, but represented a unique *MIR locus* in the genome sequence. In cases that more than one family member per family was found, a suffix was given to distinguish them from the other(s), equaling the number of *loci* found (Table 2), e.g., family miR156 includes two family members (two different miRNA sequences): miR156a-g and miR156j-l, including seven and three miRNAs/*MIR loci*, respectively: miRNA156a/*MIR*156a, miRNA156b/*MIR*156b, etc. If different, slightly shifted or shorter miRNAs mapped to the same *locus*, they were named a variant (*v2* or *v3*; Table 2), being potential candidates to be an additional family member, either by splitting one miRNA member into two or by non-exclusive use of the same *locus* under different circumstances. The prefixes *Pax-*, *Pin-* and *Phy-*, represent the miRNAs in *P. axillaris, P. inflata*, and *P. hybrida*, respectively.

MiRNA and miRNA* frequency analysis

Annotated sequences with a minimum frequency of two in at least two of the ten small RNA libraries analyzed were regarded candidates for 'true' miRNAs or miRNA*s and included in our analysis. The frequency data was collected in an Excel table, normalized to Reads Per Million cleaned reads (RPM), and Log2 transformed for representation in graphs. Tests for significant differences were performed on the normalized data, using t-tests that take into account the variance, and either tested between YB of *PaxiN* and YB of *PinfS6* (*y* in the graphs) or between the YB samples and LB samples (*b* in the graphs).

MiRNA target prediction

The analysis for miRNA targets was performed on the gene annotation list presented in this paper (Supplementary Note 1, Bombarely et al.) that was based on transcript data of *P. axillaris*, *P. inflata*, and *P. hybrida* (Zenoni et al., 2011), protein sequence data from the tomato genome (Tomato Genome Consortium, 2012), and SwissProt for all Solanaceae, by using TargetFinder (version 1.6; Carrington lab; Fahlgren et al., 2007). C. 20 % of this data included 5'-UTRs and 30 % 3'UTRs. TargetFinder uses a plant based scoring metric with a cutoff value of 4 and the following penalties for miRNA/target duplexes: '+1' for a mismatch, single nt gap or single nt bulge, '+0.5' for a G:U base pair, and 'double scores' at the 'seed' region: positions 2-13 relative to the 5'-end of the miRNA (Liu et al., 2014). Duplexes were rejected if they contained more than one single nt bulge or gap, or seven total mismatches, G:U base pairs, bulges and gaps, or four total mismatches or total G:U base pairs.

Supplementary data

Additional Tables in the Appendix are:

Table 1B: Read characteristics, miRNA frequencies and miRNA* frequencies, in ten small RNA libraries of young flower buds of *Petunia axillaris* and *P. inflata*.

Tables 2B to 2F: Catalog of conserved miRNAs in *Petunia*, including names, localization in the genome, and characteristics of pre-miRNAs (*MIR*s) and miRNA*s in *P. axillaris* (**2B**) and *P. inflata* (**2C**), and sequences of the miRNAs (**2D**), *MIR*s (**2E**), and miRNA*s (**2F**).

Table 4: Comparison of miRNA frequencies in flower buds of *P. axillaris*, *P. inflata,* tomato and potato.

Additional Tables and Sequences Online are / can be accessed via:

ftp://ftp.solgenomics.net/genomes/Petunia_axillaris/annotation/ and ftp://ftp.solgenomics.net/genomes/Petunia_inflata/annotation/ with one of the following suffixes:

Peaxi_Peinfl_v1_miRNA.catalog: **Online Tables 9.1, 9.2, and 9.3,** including extra information on the miRNA frequencies in young flower buds (**Table 9.1**), the catalog of conserved miRNAs (**Table 9.2**), and the predicted miRNA target genes with DNA identification numbers (**Table 9.3**) in *Petunia*.

Peaxi_v1_miRNA.fasta and Peinfl_v1_miRNA.fasta: **Fasta Sequences** of the conserved **miRNAs, *pre*-miRNAs (*MIR*s),** and **miRNA*s** in *P. axillaris* and *P. inflata*, respectively (see also Appendix Tables 2D-F).

Acknowledgements

This work was funded by the Netherlands Scientific Organization (NWO-ALW) as part of project number: 820.02.015.

Author Contributions

TG supervised the research and provided knowledge on *Petunia* and flower development, TG and KV designed the experiments, KV performed the experiments, analyzed the data, participated in bioinformatics analysis, and wrote the manuscript, NDA performed the bioinformatics analysis, designed and wrote the custom Perl scripts, and provided input on the text.

References

Albert, N.W., Davies, K.M., Lewis, D.H., Zhang, H., Montefiori, M., Brendolise, C., Boase, M.R., Ngo, H., Jameson, P.E., and Schwinn, K.E. (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. The Plant Cell **26**: 962–980.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., and Tuschl T. (2003). A uniform system for microRNA annotation. RNA **9**: 277–279.

Bartel, D.P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell **116**: 281–297.

Bologna, N.G., Schapire, A.L., and Palatnik, J.F. (2013). Processing of plant microRNA precursors. Briefings in Functional Genomics **12**: 37–45.

Budak, H., N.G., and Akpinar B.A. (2015). Plant miRNAs: biogenesis, organization and origins. Functional and Integrative Genomics **15** (4).

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. Nature Genetics **39**: 901–905.

Castel, R., Kusters, E., and Koes, R. (2010). Inflorescence development in petunia: through the maze of botanical terminology. Journal of Experimental Botany **61**: 2235–2246.

Chen, X. (2009). Small RNAs and their roles in Plant Development. Annual Review of Cell Development and Biology **35**: 21–44.

Chorostecki, U., Crosa, V.A., Lodeyro, A.F., Bologna, N.G., Martin, A.P., Carrillo, N., Schommer, C., and Palatnik, J.F. (2012). Identification of new microRNA-regulated genes by conserved targeting in plant species. Nucleic Acids Research **40**: 8893–8904.

Chuck, G., Candela, H., and Hake, S. (2009). Big impacts by small RNAs in plant development. Current Opinion in Plant Biology **12**: 81–86.

Cui, X., Xu, S.M., Mu, D.S., and Yang, Z.M. (2009). Genomic analysis of rice microRNA promoters and clusters. Gene **431**: 61–66.

Cuperus, J.T., Fahlgren, N., and Carrington, J.C. (2011) Evolution and Functional Diversification of MIRNA Genes. The Plant Cell **23**: 431–442

Dugas, D.V., and Bartel, B. (2004). MicroRNA regulation of gene expression in plants. Current Opinion in Plant Biology **7**: 512–520.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L., and Carrington, J.C. (2007). High-Throughput Sequencing of *Arabidopsis* microRNAs: Evidence for Frequent Birth and Death of MIRNA Genes. PLoS ONE **2**: e219.

Gerats, T. and Strommer, J. (ed.) (2009). Petunia. Evolutionary, Developmental and Physiological Genetics. 2nd edition, Springer, New York.

Gou, J.Y., Felippes, F.F., Liu, C.J., Weigel, D., and Wang, J.W. (2011). Negative Regulation of Anthocyanin Biosynthesis in *Arabidopsis* by a miR156-Targeted *SPL* Transcription Factor. The Plant Cell **23**: 1512–1522.

Gu, M., Liu, W., Meng, Q., Zhang, W., and Chen, A., Sun, S., Xu, G. (2014). Identification of microRNAs in six solanaceous plants and their potential link with phosphate and mycorrhizal signalings. Journal of integrative Plant Biology **56**: 1164-1178

He, L. and Hannon, G.J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. Nature Reviews Genetics **5**: 522–531.

Huijser, P. and Schmid, M. (2011). The control of developmental phase transitions in plants. Development **138**: 4117–4129.

Jin, D., Wang, Y., Zhao, Y., and Chen, M. (2013). MicroRNAs and Their Cross-Talks in Plant Development. Journal of Genetics and Genomics **40**: 161–170.

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

Kozomara, A. and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Research **42**: D68–73.

Kozomara, A. and Griffiths-Jones, S. (2010). miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Research **39**: D152–D157.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology **10:** R25

Liu, Q., Wang, F., and Axtell, M.J. (2014). Analysis of Complementarity Requirements for Plant MiRNA Targeting Using a *Nicotiana benthamiana* Quantitative Transient Assay. The Plant Cell **26**: 741–753.

Lorenz, R., Bernhart, S.H., Siederdissen, C.H.Z., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms for Molecular Biology **6**: 26.

Luo, Y., Guo, Z., and Li, L. (2013). Evolutionary conservation of microRNA regulatory programs in plant flower development. Developmental Biology **380**: 133–144.

Mallory, A. and Vaucheret, H. (2011). Form, Function, and Regulation of *ARGONAUTE* Proteins. The Plant Cell **22**: 3879–3889.

Manavella, P.A., Koenig, D., Rubio-Somoza, I., Burbano, H.A., Becker, C., and Weigel, D. (2013). Tissue-specific silencing of Arabidopsis *SU(VAR)*3-9 *HOMOLOG*8 by miR171a. Plant Physiology **161**: 805–812.

McCormick, K.P., Willmann, M.R., and Meyers, B.C. (2011). Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. Silence **2**: 2.

Meyers, B.C. et al. (2008). Criteria for Annotation of Plant MicroRNAs. The Plant Cell **20**: 3186–3190.

Ming-Hsuan, L., and Shu-Hsing W. (2012). Unique Size of Plant miRNA. Biogenesis and Evolution. PhD-thesis, Academia Sinica, Taiwan: 1–36.

Munafó, D.B., and Robb, G.B. (2010). Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. RNA **16**: 2537–2552.

Nag, A., and Jack, T. (2010). Sculpting the Flower; the Role of microRNAs in Flower Development. Current Topics in Developmental Biology **91**: 349–378.

Papp, I., Mette, M.F., Aufsatz, W., Daxinger, L., Schauer, S.E., Ray, A., van der Winden, J., Matzke, M., and Matzke, A.J.M. (2003). Evidence for nuclear processing of plant microRNA and short interfering RNA precursors. Plant Physiology **132**: 1382–1390.

Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., and Poethig, R.S. (2005). Nuclear processing and export of microRNAs in *Arabidopsis*. Proceedings of the Nat. Academy of Sci. U.S.A. **102**: 3691–3696.

Poethig, R.S. (2013). Vegetative Phase Change and Shoot Maturation in Plants. Current Topics in Developmental Biology **105**: 125–152.

Rijpkema, A., Gerats, T., and Vandenbussche, M. (2006). Genetics of Floral Development in *Petunia*. Advances in Botanical Research **44**: 237–278.

Rogers, K. and Chen, X. (2013). MicroRNA Biogenesis and Turnover in Plants. Cold Spring Harbor Symposia on Quantitative Biology **77**: 183–194.

Sun, G. (2011). MicroRNAs and their diverse functions in plants. Plant Molecular Biology **80**: 17–36.

Tedder, P., Zubko, E., Westhead, D.R., and Meyer, P. (2009). Small RNA analysis in *Petunia hybrida* identifies unusual tissue-specific expression patterns of conserved miRNAs and of a 24mer RNA. RNA **15**: 1012–1020.

Thakur, V., Wanchana, S., Xu, M., Bruskiewich, R., Quick, W.P., Mosig, A., and Zhu, X.-G. (2011). Characterization of statistical features for plant microRNA prediction. BMC Genomics **12**: 108.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635–641.

Van Moerkercke, A., Galvan-Ampudia, C.S., Verdonk, J.C., Haring, M.A., and Schuurink, R.C. (2012). Regulators of floral fragrance production and their target genes in petunia are not exclusively active in the epidermal cells of petals. Journal of Experimental Botany **63**: 3157–3171.

Vandenbussche, M., Janssen, A., Zethof, J., van Orsouw, N., Peters, J., van Eijk, M.J.T., Rijpkema, A.S., Schneiders, H., Santhanam, P., de Been, M., van Tunen, A., and Gerats, T. (2008). Generation of a 3D indexed *Petunia* insertion database for reverse genetics. The Plant Journal **54**: 1105–1114.

Voinnet, O. (2009). Origin, Biogenesis, and Activity of Plant MicroRNAs. Cell **136**: 669–687.

Yang, X. and Li, L. (2011). miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics **27**: 2614-2615.

Zenoni, S., D'Agostino, N., Tornielli, G.B., Quattrocchio, F., Chiusano, M.L., Koes, R., Zethof, J., Guzzo, F., Delledonne, M., Frusciante, L., Gerats, T., and Pezzotti, M. (2011). Revealing impaired pathways in the *an*11 mutant by high-throughput characterization of *Petunia axillaris* and *Petunia inflata* transcriptomes. The Plant Journal **68**: 11–27.

Appendix: Additional Tables

Table 1B. Read characteristics, miRNA- and miRNA* frequencies in small RNA libraries of buds of *P. axillaris* and *P. inflata*

Core + Big Perfect Matches (CPM + BPM) in *Ath, Sly, Stu* and/or *Nta*						RPM Cleaned reads					Averages			
miRNA \ Sample	*Paxi*N [1] YB-1.1	*Paxi*N YB-2.2	*Paxi*N YB-2.3	*Paxi*N YB-3.3	*Paxi*N LB-4.3	*PInfl*S6 YB-1.2	*PInfl*S6 YB-2.2	*PInfl*S6 YB-2.3	*PInfl*S6 YB-3.3	*PInfl*S6 LB-2.3	Average YB	Average LB	Average *Pax* YB	Average *Pin* YB
Plant Selection (nrs)	1-5	6-10	6-10	11-15	16-20	1-5	6-10	6-10	11-15	6-10	**SORT**			
Plant Age (weeks)	9	12	18	18	18	12	12	18	18	18				
Input Rds (x1000)	7295	10680	14520	7955	3965	6015	62040	13620	32620	13860	19343	8913	10113	28574
Cleaned Reads (x1000)	**4801**	**8318**	**9417**	**5085**	2618	**4399**	**42690**	**9626**	**21954**	**10063**	13286	6341	6905	19667
Cleaned Rds (% of input)	66	78	65	64	66	73	69	71	67	73	69	69	68	70
Collapsed Reads (x1000)	2666	4301	4805	2894	1669	2658	16096	4963	10428	4713	6101	3191	3666	8536
Collapsed (% of cleaned)	56	52	51	57	64	60	38	52	48	47	52	55	54	49
Singletons (x1000)	2370	3689	4132	2535	1499	2339	12871	4278	8894	4030	5138	2764	3181	7095
Singlet's (% of collapsed)	89	86	86	88	90	88	80	86	85	85	86	88	87	85
miR-0166 a-h [2]	2791,3	3262,0	3226,1	3959,3	3122,2	3753,8	4946,5	4286,8	3286,1	4231,8	3689,0	3677,0	3309,7	4068,3
miR-0159 a-c	3520,5	3459,4	3574,1	2781,9	2284,2	2084,8	2960,9	3680,0	3506,4	1914,6	3196,0	2099,4	3334,0	3058,0
miR-0319 a-e	1097,1	2814,0	1864,3	1693,0	747,9	2189,6	2427,4	2950,4	1525,1	796,0	2070,1	771,9	1867,1	2273,1
miR-0166 k	1286,8	2244,5	1298,0	1530,0	1667,3	2120,7	2874,4	1944,7	1247,3	2236,1	1818,3	1951,7	1589,8	2046,8
miR-0394 a-c	626,7	474,2	717,5	806,3	586,7	355,1	540,1	622,7	502,3	743,5	580,6	665,1	656,2	505,0
miR-0403	324,5	493,0	562,1	505,2	569,1	602,2	484,8	552,4	349,2	327,7	484,2	448,4	471,2	497,1
miR-0160 a-c	1141,6	146,2	400,0	389,4	213,5	166,6	174,0	85,1	957,1	73,7	432,5	143,6	519,3	345,7
miR-0319 a-e, h *v2*	150,6	278,8	275,4	230,5	166,9	267,8	342,2	660,0	225,1	483,5	303,8	325,2	233,8	373,8
miR-0390 a-c	197,0	275,5	212,1	223,8	203,2	333,5	273,3	247,9	194,3	271,6	244,7	237,4	227,1	262,3
miR-0168 a	135,0	212,4	234,6	189,2	264,7	226,2	211,5	280,5	221,0	356,3	213,8	310,5	192,8	234,8
miR-0396 a,b	219,7	109,9	217,7	258,0	584,8	275,5	168,9	167,5	285,2	343,7	212,8	464,3	201,3	224,3
miR-0319 h,i	108,7	182,5	507,4	172,9	113,1	136,6	161,2	189,6	151,8	82,1	201,3	97,6	242,9	159,8
miR-0162 a	120,2	94,9	188,1	176,6	171,1	211,9	158,2	260,6	229,7	197,0	180,0	184,0	144,9	215,1
miR-0168 b,c	113,1	210,0	99,1	113,5	194,8	152,3	172,8	214,8	108,3	312,7	148,0	253,8	133,9	162,1
miR-6149	32,1	59,7	79,6	108,6	56,5	182,1	127,8	206,3	82,4	212,0	109,8	134,2	70,0	149,7
miR-0171 a-g	245,2	51,2	89,3	114,8	93,6	31,8	71,4	44,8	172,4	95,4	102,6	94,5	125,1	80,1
miR-0164 a-d	132,5	101,2	144,6	85,3	152,0	64,8	90,9	59,5	138,0	90,3	102,1	121,2	115,9	88,3
miR-0167 a-e	135,6	42,3	99,5	73,5	108,1	98,2	83,7	54,5	188,2	83,1	97,0	95,6	87,7	106,2
miR-0396 c,d	54,6	71,5	85,0	95,6	181,1	115,9	81,7	77,2	77,3	100,1	82,3	140,6	76,7	88,0
miR-0827	142,1	74,1	171,8	177,2	107,3	5,5	4,6	16,6	47,9	22,7	80,0	65,0	141,3	18,6
miR-0482	171,8	60,7	95,9	97,7	103,1	35,9	64,8	31,0	78,1	35,0	79,5	69,1	106,5	52,5
miR-0319 h,i *v3*	54,2	90,6	77,3	78,9	61,9	68,0	87,1	93,7	78,7	44,0	78,6	53,0	75,2	81,9
miR-0157 a-e	47,7	15,9	12,3	25,8	418,3	108,9	57,7	102,3	187,4	1174,3	69,8	796,3	25,4	114,1
miR-0172 a-g	98,5	83,0	73,9	85,2	120,7	53,9	54,4	42,3	58,7	34,4	68,7	77,5	85,1	52,3
miR-0157 a-e *v2*	44,8	15,1	10,7	21,2	396,9	107,1	55,6	100,0	180,4	1125,8	66,9	761,3	23,0	110,8
miR-0172 j,k	61,7	60,2	68,5	73,4	63,4	64,6	49,4	32,2	57,8	26,3	58,5	44,9	65,9	51,0
miR-0171 a-g *v2*	63,3	27,9	37,2	36,4	32,8	32,7	37,3	21,9	58,2	47,4	39,4	40,1	41,2	37,5
miR-0167 *var*	32,7	15,9	23,8	31,9	29,0	14,1	15,5	22,6	27,2	36,4	23,0	32,7	26,1	19,9
miR-0159 *var*	12,9	18,2	14,3	12,0	8,0	14,3	19,4	27,9	12,2	13,9	16,4	11,0	14,3	18,5
miR-8016	17,5	15,7	16,7	14,9	11,5	11,8	17,2	11,4	16,6	10,0	15,2	10,7	16,2	14,3
miR-0393 a-c	22,9	7,2	14,9	17,1	16,8	6,1	10,1	13,8	25,0	25,2	14,6	21,0	15,5	13,8
miR-0390 a-c *v2*	22,7	6,7	12,6	14,4	11,1	8,6	10,6	7,2	24,8	7,7	13,5	9,4	14,1	12,8
miR-0169 c(*BL*),d	11,5	2,8	4,0	5,9	4,6	3,0	3,7	1,8	7,0	1,8	4,9	3,2	6,0	3,9
miR-0171 j,k	2,5	5,2	4,8	6,9	2,3	1,6	3,7	2,5	3,1	3,0	3,8	2,6	4,8	2,7
miR-0172 j,k v2	5,8	2,6	4,0	2,9	3,8	1,8	2,6	1,8	5,1	3,4	3,3	3,6	3,9	2,8
miR-0166 *var*	4,8	1,4	1,8	1,4	2,7	2,7	2,8	2,7	2,7	1,6	2,5	2,1	2,4	2,7
miR-0408	4,8	2,6	4,0	3,5	3,1	0,0	0,1	0,5	0,5	0,5	2,0	1,8	3,8	0,3
miR-0395 a-l	3,3	2,3	2,0	3,1	3,8	0,0	0,3	1,0	0,9	7,7	1,6	5,7	2,7	0,6
miR-0398 e	3,1	1,2	3,2	4,1	2,3	0,2	0,1	0,1	0,3	0,2	1,5	1,2	2,9	0,2
miR-0156 a-g	1,2	1,1	1,7	2,9	1,5	0,5	0,8	0,8	0,5	0,8	1,2	1,2	1,7	0,6
miR-0172 a-g *v2*	0,6	0,0	0,3	0,2	0,4	0,9	0,8	1,2	1,7	1,3	0,7	0,8	0,3	1,2

miRNA \ Sample	*Paxi*N [1] YB-1.1	*Paxi*N YB-2.2	*Paxi*N YB-2.3	*Paxi*N YB-3.3	*Paxi*N LB-4.3	*PInfl*S6 YB-1.2	*PInfl*S6 YB-2.2	*PInfl*S6 YB-2.3	*PInfl*S6 YB-3.3	*PInfl*S6 LB-2.3	Average YB	Average LB	Average *Pax* YB	Average *Pin* YB
miR-0156 j-l	0,2	0,1	0,0	0,2	0,0	0,7	1,5	1,6	1,2	1,7	0,7	0,8	0,1	1,2
miR-0171 j,k *v2*	0,2	1,6	0,5	0,8	2,3	0,9	0,5	0,3	0,5	4,3	0,7	3,3	0,8	0,6
miR-0479	2,1	0,5	0,8	0,6	1,5	0,2	0,2	0,0	0,2	0,7	0,6	1,1	1,0	0,2
miR-2111 a,b	1,5	0,1	0,8	1,6	0,8	0,0	0,0	0,0	0,1	0,0	0,5	0,4	1,0	0,0
miR-0477	1,5	1,0	0,6	0,8	1,1	0,0	0,0	0,0	0,0	0,0	0,5	0,6	1,0	0,0
miR-0398 a,b *v2*	0,2	0,7	0,5	2,4	0,8	0,0	0,0	0,0	0,0	0,0	0,5	0,4	1,0	0,0
miR-0169 a,b	1,9	0,0	0,2	0,2	0,0	0,0	0,3	0,1	0,6	0,1	0,4	0,0	0,6	0,2
miR-0398 h	0,8	0,5	1,3	0,4	0,4	0,0	0,1	0,0	0,0	0,0	0,4	0,2	0,7	0,0
miR-0398 h *v2*	1,0	0,4	0,4	0,6	0,0	0,0	0,3	0,0	0,1	0,4	0,4	0,2	0,6	0,1
miR-0398 a,b	0,0	0,4	0,7	1,6	0,0	0,0	0,0	0,0	0,0	0,0	0,3	0,0	0,7	0,0
miR-0397 a,b	0,6	0,1	0,7	0,2	0,4	0,0	0,2	0,1	0,2	0,2	0,3	0,3	0,4	0,1
miR-0171 n,o	0,0	0,2	0,4	0,8	0,0	0,2	0,0	0,0	0,0	0,0	0,2	0,0	0,4	0,1
miR-0399 a-c	0,2	0,0	0,1	1,0	0,4	0,0	0,0	0,3	0,0	0,1	0,2	0,2	0,3	0,1
miR-0172 *var 2*	0,2	0,2	0,2	0,4	0,0	0,2	0,1	0,0	0,2	0,0	0,2	0,0	0,3	0,1
miR-0169 g-x	0,6	0,6	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,2	0,0	0,4	0,0
miR-0390 *var*	0,2	0,1	0,1	0,4	0,0	0,2	0,2	0,0	0,1	0,1	0,2	0,0	0,2	0,1
miR-0164 g,h	0,0	0,1	0,0	0,0	0,0	0,5	0,2	0,1	0,5	0,6	0,2	0,3	0,0	0,3
miR-0172 *var*	0,0	0,0	0,0	0,0	0,0	0,5	0,0	0,2	0,1	0,0	0,1	0,0	0,0	0,2
miR-0171 j *v3*	0,0	0,0	0,0	0,0	0,0	0,2	0,0	0,3	0,2	3,4	0,1	1,7	0,0	0,2
miR-0399 c-g	0,0	0,0	0,3	0,4	0,4	0,0	0,0	0,0	0,0	0,0	0,1	0,2	0,2	0,0
miR-0397 a,b v2	0,2	0,0	0,1	0,0	0,4	0,2	0,1	0,0	0,0	0,4	0,1	0,4	0,1	0,1
miR-0164 *var*	0,0	0,0	0,1	0,0	0,0	0,0	0,1	0,0	0,2	0,2	0,0	0,1	0,0	0,1
miR-2111 a,b *v2*	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,2	0,0	0,0	0,0	0,0	0,1
miR-6164 a-t	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,1	0,0	0,0	0,0	0,0	0,0
miR*0168 a-c	158,3	309,7	311,8	311,1	362,5	405,3	356,9	262,7	260,1	237,4	297,0	299,9	272,7	321,2
miR*0166 a,b	72,7	93,5	114,7	115,0	79,8	349,6	342,6	427,2	249,6	423,3	220,6	251,6	99,0	342,3
miR*0166 k	114,8	124,1	115,0	96,0	81,0	131,2	141,9	142,1	154,3	92,5	127,4	86,7	112,5	142,4
miR*0171 d	80,4	54,1	125,4	99,1	113,8	83,7	64,3	63,3	93,9	55,5	83,0	84,6	89,8	76,3
miR*0827	14,2	22,6	30,3	36,2	19,9	2,5	1,5	4,9	6,9	7,9	14,9	13,9	25,8	4,0
miR*0396 c	11,2	10,8	10,8	13,0	26,0	8,4	9,6	9,1	16,6	12,4	11,2	19,2	11,5	10,9
miR*0171 b	8,5	6,7	6,3	2,8	6,9	13,9	17,9	13,4	19,8	8,5	11,2	7,7	6,1	16,3
miR*0159 a	11,9	6,5	12,7	11,4	7,3	11,4	8,9	14,5	10,3	4,5	11,0	5,9	10,6	11,3
miR*0160 a,b	16,2	8,2	7,8	9,8	6,9	9,3	8,3	8,6	11,2	5,6	9,9	6,2	10,5	9,4
miR*0403	2,7	3,8	3,4	2,6	2,3	6,1	9,3	9,3	10,1	19,7	5,9	11,0	3,1	8,7
miR*0390 b,c	12,9	4,1	5,2	4,9	5,0	4,3	4,1	2,2	8,1	2,8	5,7	3,9	6,8	4,7
miR*0157 a	5,8	1,2	1,3	2,6	21,4	6,8	1,9	6,2	14,1	35,8	5,0	28,6	2,7	7,3
miR*0172 j	7,5	3,2	3,1	3,3	1,5	1,1	2,3	2,1	6,0	1,0	3,6	1,3	4,3	2,9
miR*0172 b,c	10,8	2,4	2,2	2,9	2,7	0,0	0,8	0,7	3,8	2,1	3,0	2,4	4,6	1,3
miR*0396 a	1,5	2,3	2,1	2,4	5,3	3,4	1,9	2,5	2,2	6,9	2,3	6,1	2,1	2,5
miR*0319 b	1,0	0,8	1,1	0,6	0,4	2,5	4,1	2,5	2,9	1,1	1,9	0,7	0,9	3,0
miR*0166 c	1,5	3,5	3,4	3,1	4,2	0,5	0,5	0,1	0,3	0,8	1,6	2,5	2,9	0,3
miR*0167 b	1,5	1,1	1,1	1,4	0,8	2,0	1,7	2,2	1,6	3,4	1,6	2,1	1,2	1,9
miR*0482	1,2	2,2	0,7	2,2	1,5	1,4	1,4	1,5	1,5	2,3	1,5	1,9	1,6	1,4
miR*0171 a *v2*	1,5	1,6	1,1	0,8	1,1	1,6	1,6	1,2	0,8	2,1	1,3	1,6	1,2	1,3
miR*0171 j,k	0,8	1,2	1,5	1,2	0,4	0,7	1,1	2,5	0,7	2,7	1,2	1,5	1,2	1,2
miR*0172 a	3,1	0,6	1,1	2,0	0,0	0,2	0,7	0,1	1,2	0,1	1,1	0,0	1,7	0,6
miR*0164 a	1,7	0,7	1,4	1,8	0,0	0,5	0,6	1,2	0,8	0,4	1,1	0,2	1,4	0,8
miR*0169 o	0,8	3,6	2,0	1,4	2,3	0,2	0,0	0,1	0,0	0,0	1,0	1,1	2,0	0,1
miR*0167 a	1,2	0,8	0,6	0,6	4,6	1,4	1,2	1,4	0,5	5,1	1,0	4,8	0,8	1,1
miR*0162 a	0,2	0,8	0,5	0,6	1,1	1,6	1,1	1,7	0,3	0,4	0,9	0,8	0,5	1,2
miR*0166 d	0,6	0,7	0,6	1,2	0,4	0,5	0,5	0,7	0,4	0,3	0,6	0,3	0,8	0,5
miR*8016	0,6	0,4	0,7	0,6	0,0	0,7	0,9	0,2	0,9	0,3	0,6	0,1	0,6	0,7
miR*0166 e,f	0,4	1,1	0,8	1,0	1,1	0,2	0,4	0,3	0,2	0,0	0,6	0,6	0,8	0,3
miR*0169 d	1,0	1,0	0,5	1,8	0,4	0,0	0,0	0,0	0,0	0,0	0,5	0,2	1,1	0,0
miR*6164 a	0,6	1,6	0,5	0,2	1,1	0,2	0,5	0,2	0,3	0,2	0,5	0,7	0,7	0,3
miR*0396 b	0,4	0,6	0,4	1,0	1,9	0,0	0,2	0,7	0,3	1,6	0,5	1,7	0,6	0,3

miRNA \ Sample	*Paxi*N [1] YB-1.1	*Paxi*N YB-2.2	*Paxi*N YB-2.3	*Paxi*N YB-3.3	*Paxi*N LB-4.3	*PinfI*S6 YB-1.2	*PinfI*S6 YB-2.2	*PinfI*S6 YB-2.3	*PinfI*S6 YB-3.3	*PinfI*S6 LB-2.3	Average YB	Average LB	Average *Pax* YB	Average *Pin* YB
miR*0172 e	0,6	0,1	0,2	0,6	0,4	0,2	0,4	0,2	0,8	0,1	0,4	0,2	0,4	0,4
miR*0397 a	1,0	0,7	0,6	0,6	0,8	0,0	0,0	0,0	0,0	0,0	0,4	0,4	0,7	0,0
miR*0171 c' *v2*	0,0	0,0	0,0	0,0	0,0	0,5	0,7	0,6	1,1	0,8	0,4	0,4	0,0	0,7
miR*0171 b *v2*	0,2	0,1	0,2	0,0	0,4	0,5	0,6	0,6	0,6	0,1	0,4	0,2	0,1	0,6
miR*0396 d	0,0	0,2	0,2	0,4	0,0	0,2	0,3	0,5	0,9	1,6	0,3	0,8	0,2	0,5
miR*0164 b	0,4	0,4	0,3	0,2	0,0	0,5	0,4	0,3	0,2	0,7	0,3	0,3	0,3	0,3
miR*0398 a	0,4	0,4	0,2	0,6	0,8	0,2	0,2	0,4	0,2	0,1	0,3	0,4	0,4	0,3
miR*0408	0,6	0,6	0,4	0,2	1,1	0,2	0,1	0,2	0,1	0,4	0,3	0,8	0,5	0,2
miR*0319 i	1,0	0,4	0,2	0,2	0,8	0,2	0,2	0,1	0,2	0,0	0,3	0,4	0,5	0,2
miR*0398 h *v2*	0,0	0,1	0,2	0,4	0,4	0,2	0,3	0,4	0,2	0,2	0,2	0,3	0,2	0,3
miR*0159 b	0,2	0,1	0,2	0,0	0,0	0,2	0,3	0,1	0,3	0,2	0,2	0,1	0,1	0,2
miR*0169 i	0,2	0,6	0,2	0,2	0,0	0,2	0,0	0,0	0,0	0,0	0,2	0,0	0,3	0,1
miR*0398 e	0,4	0,2	0,4	0,2	0,4	0,0	0,0	0,0	0,0	0,1	0,2	0,2	0,3	0,0
miR*0477	0,4	0,0	0,5	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,2	0,0	0,3	0,0
miR*0172 d	0,2	0,1	0,0	0,0	0,0	0,7	0,1	0,0	0,1	0,1	0,2	0,0	0,1	0,2
miR*0164 c	0,2	0,4	0,0	0,4	0,0	0,0	0,1	0,1	0,0	0,1	0,1	0,0	0,2	0,0
miR*0390 a	0,0	0,1	0,1	0,0	0,0	0,2	0,1	0,1	0,5	0,1	0,1	0,0	0,1	0,2
miR*0171 c *v2*	0,4	0,2	0,4	0,0	0,4	0,0	0,0	0,0	0,0	0,0	0,1	0,2	0,3	0,0
miR*0479	0,0	0,6	0,2	0,0	0,4	0,0	0,0	0,1	0,0	0,4	0,1	0,4	0,2	0,0
miR*6149	0,2	0,1	0,0	0,2	0,0	0,0	0,0	0,2	0,1	0,0	0,1	0,0	0,1	0,1
miR*0169 n	0,0	0,1	0,3	0,4	0,4	0,0	0,0	0,0	0,0	0,0	0,1	0,2	0,2	0,0
*miR*6164 e*	*0,0*	*0,1*	*0,0*	*0,2*	*0,0*	*0,5*	*0,0*	*0,0*	*0,0*	*0,1*	*0,1*	*0,0*	*0,1*	*0,1*
*miR*0319 d*	*0,2*	*0,0*	*0,2*	*0,2*	*0,0*	*0,0*	*0,0*	*0,0*	*0,0*	*0,0*	*0,1*	*0,0*	*0,2*	*0,0*
*miR*0166 h*	*0,0*	*0,0*	*0,2*	*0,4*	*0,0*	*0,0*	*0,0*	*0,0*	*0,0*	*0,0*	*0,1*	*0,0*	*0,2*	*0,0*
miR*0395 k,l	0,0	0,4	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,1	0,0	0,1	0,0
miR*0169 b'	0,0	0,0	0,0	0,0	0,0	0,5	0,1	0,0	0,0	0,8	0,1	0,4	0,0	0,1
*miR*0166 h''*	*0,0*	*0,0*	*0,0*	*0,0*	*0,0*	*0,5*	*0,0*	*0,0*	*0,1*	*0,0*	*0,1*	*0,0*	*0,0*	*0,1*
*miR*0169 c*	*0,0*	*0,0*	*0,1*	*0,4*	*0,0*	*0,0*	*0,0*	*0,0*	*0,0*	*0,0*	*0,1*	*0,0*	*0,1*	*0,0*
miR*0393 c	0,2	0,0	0,1	0,2	1,9	0,0	0,0	0,0	0,0	0,1	0,1	1,0	0,1	0,0
*miR*0395 b-e*	*0,0*	*0,0*	*0,1*	*0,4*	*0,4*	*0,0*	*0,0*	*0,0*	*0,0*	*0,1*	*0,1*	*0,2*	*0,1*	*0,0*
miR*0157 b	0,0	0,0	0,0	0,0	0,0	0,2	0,2	0,0	0,1	0,0	0,1	0,0	0,0	0,1
*miR*0399 e-g*	*0,0*	*0,0*	*0,2*	*0,2*	*0,0*	*0,0*	*0,0*	*0,0*	*0,0*	*0,0*	*0,1*	*0,0*	*0,1*	*0,0*
miR*0397 a *v2*	0,0	0,0	0,0	0,0	0,0	0,0	0,2	0,0	0,3	0,6	0,1	0,3	0,0	0,1
*miR*0399 b*	*0,0*	*0,0*	*0,2*	*0,2*	*0,4*	*0,0*	*0,0*	*0,0*	*0,0*	*0,0*	*0,1*	*0,2*	*0,1*	*0,0*
*miR*0156 e*	*0,0*	*0,0*	*0,0*	*0,4*	*0,0*	*0,0*	*0,0*	*0,0*	*0,0*	*0,0*	*0,1*	*0,0*	*0,1*	*0,0*
miR*0319 e	0,0	0,0	0,1	0,0	0,0	0,0	0,1	0,1	0,1	0,0	0,1	0,0	0,0	0,1
miR*0156 b-d	0,0	0,0	0,2	0,0	0,0	0,0	0,1	0,0	0,0	0,2	0,0	0,1	0,1	0,0
miR*0164 h	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,2	0,0	0,1	0,0	0,0	0,0	0,1

[1] **Sample names** are: *Paxi*N, *Pax* = *Petunia axillaris*; *Pinf*S6, *Pin* = *Petunia inflata*; YB = young buds; LB = Larger buds (see Fig. 1)

[2] **Upper half** presents miRNAs, **lower half** miRNA*s;
S**uffixes** are: *v2* and *v3* = variant of an already defined miRNA and locus; *var* = variant NOT confirmed by the *Petunia* genome sequences
Grey values are < 1 RPB (one read per Billion ~ absent)

Table 2B. DNA identification numbers, MIR characteristics and miRNA*-type of conserved miRNAs in *Petunia axillaris*

Family Name / Member Name	Species Name	Petunia Scaffold	Mature Begin [1]	Mature End [1]	Star Begin	Star End	Mature Strand [2]	Pre Length	MFE kcal/mol	Pre [3] Type	Star [4] Type	Star [5] Present	Cluster [6]
Petunia axillaris													
miR-0156													
miR-0156 a-g	miR-0156 a	Peaxi162Scf00064	337067	337086	337132	337153	+ / -	87	-42,6	**a**	**a**	-	
miR-0156 a-g	miR-0156 b	Peaxi162Scf00297	229751	229770	229820	229841	+ / +	90	-48,4	**b**	**b-d**	< 1	I
miR-0156 a-g	miR-0156 c	Peaxi162Scf00297	1165227	1165246	1165295	1165316	+ / -	90	-48,6	**c**	**b-d**	"	I
miR-0156 a-g	miR-0156 d	Peaxi162Scf00070	895396	895415	895470	895491	+ / -	96	-47,6	**d**	**b-d**	"	II
miR-0156 a-g	miR-0156 e	Peaxi162Scf00070	913547	913566	913611	913632	+ / -	86	-45,3	**e**	**e**	<<1	II
miR-0156 a-g	miR-0156 f	Peaxi162Scf00062	2065222	2065241	2065290	2065311	+ / -	90	-48,6	**f**	**f**	-	
miR-0156 a-g	miR-0156 g	Peaxi162Scf00347	1098665	1098684	1098729	1098750	+ / +	86	-47,8	**g**	**g**	-	
miR-0156 j-l	miR-0156 j	Peaxi162Scf00001	916641	916660	916702	916722	+ / -	82	-22,3	**j**	**j**	-	
miR-0156 j-l	miR-0156 k	Peaxi162Scf00343	445321	445340	445385	445405	+ / -	85	-48,7	**k**	**k**	-	III
miR-0156 j-l	miR-0156 l	Peaxi162Scf00139	1287557	1287576	1287619	1287639	+ / +	83	-46,4	**l**	**l**	-	
miR-0157													
miR-0157 a-e	miR-0157 a	Peaxi162Scf00332	48589	48608	48648	48665	+ / +	80	-44,0	**a**	**a**	1-20*	IV
miR-0157 a-e	miR-0157 b	Peaxi162Scf00343	445739	445758	445805	445822	+ / -	84	-41,9	**b**	**b**	<1	III
miR-0157 a-e	miR-0157 c	Peaxi162Scf00343	446099	446118	446163	446182	+ / -	85	-42,7	**c**	**c**	-	III
miR-0157 a-e	miR-0157 d	Peaxi162Scf00683	220492	220511	220566	220583	+ / -	92	-44,0	**d**	**d**	-	V
miR-0157 a-e	miR-0157 e	Peaxi162Scf00877	188044	188063	188112	188129	+ / +	89	-43,4	**e**	**e**	-	
miR-0159													
miR-0159 a-c	miR-0159 a	Peaxi162Scf00683	297394	297414	297246	297266	+ / -	169	-81,6	**a**	**a**	5-20	V
miR-0159 a-c	miR-0159 b	Peaxi162Scf00332	89542	89562	89391	89411	+ / +	172	-72,1	**b**	**b**	<1	IV
miR-0159 a-c	miR-0159 c	Peaxi162Scf00074	2106807	2106827	2106655	2106675	+ / +	173	-77,9	**c**	**c**	-	
miR-0160													
miR-0160 a-d	miR-0160 a	Peaxi162Scf00156	104259	104279	104322	104342	+ / -	84	-48,7	**a**	**a,b**	5-20*	
miR-0160 a-d	miR-0160 b	Peaxi162Scf00747	193786	193806	193848	193868	+ / -	83	-52,4	**b**	**a,b**	"	
miR-0160 a-d	miR-0160 c	Peaxi162Scf00267	1134909	1134929	1134977	1134998	+ / +	89	-46,1	**c**	**c**	-	
miR-0160 a-d	miR-0160 d	*not in Pax*									cannot be tested		
miR-0162													
miR-0162 a	miR-0162 a	Peaxi162Scf00017	3063481	3063501	3063549	3063568	+ / +	87	-38,0	**a**	**a**	~ 1*	
miR-0164													
miR-0164 a-d	miR-0164 a	Peaxi162Scf00037	1320016	1320036	1320082	1320102	+ / +	87	-42,9	**a**	**a**	~ 1	
miR-0164 a-d	miR-0164 b	Peaxi162Scf00744	158353	158373	158481	158501	+ / +	149	-52,6	**b**	**b**	< 1	
miR-0164 a-d	miR-0164 c	Peaxi162Scf00009	956899	956919	957067	957087	+ / +	189	-61,0	**c**	**c**	< 1	bp
miR-0164 a-d	miR-0164 d	Peaxi162Scf00071	533618	533638	533659	533679	+ / +	62	-34,3	**d**	**d**	-	
miR-0164 g,h	miR-0164 g	Peaxi162Scf00038	830686	830706	830740	830762	+ / -	77	-35,8	**g**	**g**	-	
miR-0164 g,h	miR-0164 h	Peaxi162Scf00207	1518018	1518038	1518075	1518097	+ / +	78	-38,7	**h**	**h**	< 1	
miR-0166													
miR-0166 a-h	miR-0166 a	Peaxi162Scf00074	1924124	1924144	1924149	1924169	+ / -	137	-51,5	**a**	**a,b**	200-1000*	bp
miR-0166 a-h	miR-0166 b	Peaxi162Scf00087	1336896	1336916	1336751	1336771	+ / -	166	-52,7	**b**	**a,b**	"	
miR-0166 a-h	miR-0166 c	Peaxi162Scf00572	478977	478997	478879	478899	+ / +	97	-40,0	**c**	**c**	1-5*	
miR-0166 a-h	miR-0166 d	Peaxi162Scf00740	517413	517433	517277	517297	+ / +	157	-46,3	**d**	**d**	~ 1	bp
miR-0166 a-h	miR-0166 e	Peaxi162Scf00325	885810	885830	885737	885757	+ / +	94	-41,5	**e**	**e,f**	~ 1	
miR-0166 a-h	miR-0166 f	Peaxi162Scf00813	126148	126168	125917	125937	+ / -	252	-86,9	**f**	**e,f**	"	bp
miR-0166 a-h	miR-0166 g	Peaxi162Scf00618	727464	727484	727405	727425	+ / -	80	-42,0	**g**	**g**	-	
miR-0166 a-h	miR-0166 h	Peaxi162Scf00719	575322	575342	575253	575273	+ / +	90	-34,9	**h**	**h**	±	
miR-0166 k	miR-0166 k	Peaxi162Scf00011	422595	422615	422460	422480	+ / +	156	-63,2	**k**	**k**	20-200	
miR-0167													
miR-0167 a-e	miR-0167 a	Peaxi162Scf00608	28148	28168	28201	28221	+ / -	73	-31,7	**a**	**a**	1-5*	
miR-0167 a-e	miR-0167 b	Peaxi162Scf00259	904705	904725	904761	904781	+ / -	77	-34,2	**b**	**b**	1-5*	VI
miR-0167 a-e	miR-0167 c	Peaxi162Scf00259	891163	891183	891238	891258	+ / -	96	-36,2	**c**	**c**	-	VI
miR-0167 a-e	miR-0167 d	Peaxi162Scf00316	976970	976990	977031	977047	+ / -	78	-35,7	**d**	**d**	-	
miR-0167 a-e	miR-0167 e	Peaxi162Scf00572	549649	549669	549736	549750	+ / -	102	-43,3	**e**	**e**	-	
miR-0168													
miR-0168 a	miR-0168 a	Peaxi162Scf00014	1136483	1136503	1136564	1136584	+ / -	102	-41,8	**a**	**a-c**	200-1000*	
miR-0168 b,c	miR-0168 b	Peaxi162Scf00876	272722	272742	272779	272799	+ / -	78	-36,1	**b**	**a-c**	"	
miR-0168 b,c	miR-0168 c	Peaxi162Scf00045	1076588	1076608	1076693	1076713	+ / +	126	-42,1	**c**	**a-c**	"	
miR-0169													
miR-0169 a,b	miR-0169 a	Peaxi162Scf00186	1723459	1723479	1723545	1723565	+ / +	107	-51,3	**a**	**a**	-	bp
miR-0169 a,b	miR-0169 b	Peaxi162Scf00324	645904	645924	645952	645972	+ / -	69	-43,5	**b**	**b**	-	VII
miR-0169 c,d	miR-0169 c(*BL*	Peaxi162Scf00129	1722354	1722374	1722453	1722473	+ / -	120	-50,3	**c**	**c**	±	
miR-0169 c,d	miR-0169 d	Peaxi162Scf00324	541993	542013	542052	542072	+ / -	80	-33,3	**d**	**d**	< 1	VII
miR-0169 g-x	miR-0169 g	Peaxi162Scf00418	209592	209611	209663	209683	+ / -	92	-45,8	**g**	**g**	±*	
miR-0169 g-x	miR-0169 h	Peaxi162Scf00008	1742180	1742199	1742237	1742270	+ / -	91	-38,7	**h**	**h**	-	
miR-0169 g-x	miR-0169 i	Peaxi162Scf00010	3013709	3013728	3013768	3013801	+ / -	92	-40,0	**i**	**i**	< 1	

Family Name / Member Name	Species Name	Petunia Scaffold	Mature Begin [1]	Mature End [1]	Star Begin	Star End	Mature Strand [2]	Pre Length	MFE kcal/mol	Pre [3] Type	Star [4] Type	Star [5] Present	Cluster [6]
miR-0169 g-x	miR-0169 j	Peaxi162Scf00088	729351	729371	729434	729454	+ / +	104	-32,8	j	j	-	VIII
miR-0169 g-x	miR-0169 k	Peaxi162Scf00088	729560	729579	729634	729653	+ / +	93	-46,0	k	k	-	VIII
miR-0169 g-x	miR-0169 l	Peaxi162Scf00088	735506	735525	735585	735604	+ / +	99	-43,1	l	l	-	VIII
miR-0169 g-x	miR-0169 m	Peaxi162Scf00088	735713	735733	735778	735796	+ / +	86	-40,8	m	m	-	VIII
miR-0169 g-x	miR-0169 n	Peaxi162Scf00088	1113509	1113529	1113601	1113619	+ / -	110	-45,6	n	n	< 1	VIII
miR-0169 g-x	miR-0169 o	Peaxi162Scf00822	31248	31267	31356	31389	+ / +	127	-47,7	o	o	~ 1	IX
miR-0169 g-x	miR-0169 p	Peaxi162Scf00822	60831	60851	60906	60924	+ / +	95	-36,1	p	p	-	IX
miR-0169 g-x	miR-0169 p2	Peaxi162Scf00822	112803	112823	112881	112899	+ / +	98	-33,2	p2	p	-	IX
miR-0169 g-x	miR-0169 q	Peaxi162Scf00822	124190	124210	124293	124311	+ / +	123	-52,7	q	q	-	IX
miR-0169 g-x	miR-0169 q2	Peaxi162Scf00822	135135	135155	135237	135255	+ / +	122	-49,3	q2	q	-	IX
miR-0169 g-x	miR-0169 r	Peaxi162Scf00822	169932	169952	170028	170046	+ / +	116	-53,0	r	r	-	IX
miR-0169 g-x	miR-0169 s	Peaxi162Scf00171	111004	111024	111094	111116	+ / -	113	-31,1	s	s	-	
miR-0169 g-x	miR-0169 t	Peaxi162Scf00171	110772	110792	110844	110863	+ / -	91	-39,7	t	t	-	
miR-0169 g-x	miR-0169 u	Peaxi162Scf00227	903121	903141	903215	903233	+ / -	113	-49,2	u	u	-	
miR-0169 g-x	miR-0169 v	Peaxi162Scf00413	266232	266252	266334	266352	+ / -	121	-39,0	v	v	-	bp
miR-0169 g-x	miR-0169 w	Peaxi162Scf01064	234235	234255	234304	234322	+ / -	88	-38,4	w	w	-	
miR-0171													
miR-0171 a-g	miR-0171 a	Peaxi162Scf00007	835621	835641	835564	835584	+ / +	78	-32,1	a	a	-	
miR-0171 a-g	miR-0171 b	Peaxi162Scf00047	658148	658168	658094	658114	+ / -	75	-33,3	b	b	5-20*	
miR-0171 a-g	miR-0171 c	Peaxi162Scf01204	193189	193209	193134	193154	+ / -	76	-28,8	c	c	-	
miR-0171 a-g	miR-0171 d	Peaxi162Scf00328	1280146	1280166	1280084	1280102	+ / -	83	-29,4	d	d	20-200*	
miR-0171 a-g	miR-0171 e	Peaxi162Scf00089	1825540	1825560	1825562	1825582	+ / -	75	-37,8	e	e	-	
miR-0171 a-g	miR-0171 f	Peaxi162Scf00439	1018479	1018499	1018456	1018476	+ / -	94	-32,4	f	f	-	
miR-0171 a-g	miR-0171 g	Peaxi162Scf00207	1655318	1655338	1655260	1655280	+ / -	79	-29,3	g	g	-	
miR-0171 j,k	miR-0171 j	Peaxi162Scf00088	210321	210341	210256	210276	+ / -	86	-40,2	j	j,k	~ 1	
miR-0171 j,k	miR-0171 k	Peaxi162Scf00313	238798	238818	238733	238753	+ / -	86	-37,9	k	j,k	"	
miR-0171 n,o	miR-0171 n	Peaxi162Scf00007	724100	724120	724039	724059	+ / +	82	-43,4	n	n	-	
miR-0171 n,o	miR-0171 o	Peaxi162Scf00462	503513	503533	503447	503467	+ / -	87	-38,6	o	o	-	
miR-0172													
miR-0172 a-g	miR-0172 a	Peaxi162Scf00627	459738	459758	459670	459690	+ / -	89	-34,2	a	a	~1 *	
miR-0172 a-g	miR-0172 b	Peaxi162Scf00029	1503238	1503258	1503156	1503176	+ / -	103	-51,8	b	b,c	1-5	
miR-0172 a-g	miR-0172 c	Peaxi162Scf00115	714683	714703	714605	714625	+ / +	99	-42,0	c	b,c	"	
miR-0172 a-g	miR-0172 d	Peaxi162Scf00145	1980648	1980668	1980564	1980584	+ / +	105	-39,5	d	d	< 1	X
miR-0172 a-g	miR-0172 e	Peaxi162Scf00145	2027353	2027373	2027270	2027290	+ / +	104	-36,9	e	e	< 1	X
miR-0172 a-g	miR-0172 f	Peaxi162Scf00145	117018	117038	116930	116950	+ / -	109	-43,1	f	f	-	X
miR-0172 a-g	miR-0172 g	Peaxi162Scf00123	3697	3717	3616	3636	+ / -	102	-40,0	g	g	-	
miR-0172 j,k	miR-0172 j	Peaxi162Scf00164	524657	524677	524569	524589	+ / -	109	-51,2	j	j	1-5*	
miR-0172 j,k	miR-0172 j	Peaxi162Scf28957	1930	1950	1842	1862	+ / +	109	-51,2	j	j	"	
miR-0172 j,k	miR-0172 k	Peaxi162Scf00760	437890	437910	437787	437807	+ / +	124	-49,7	k	k	-	
miR-0319													
miR-0319 a-e	miR-0319 a	Peaxi162Scf00074	1427029	1427049	1426880	1426900	+ / +	170	-82,3	a	a	-	
miR-0319 a-e	miR-0319 b	Peaxi162Scf00117	468077	468097	467924	467944	+ / -	174	-81,8	b	b	1-5	
miR-0319 a-e	miR-0319 c	Peaxi162Scf00205	1162559	1162579	1162408	1162428	+ / -	172	-86,1	c	c	-	
miR-0319 a-e	miR-0319 d	Peaxi162Scf00761	312727	312747	312579	312599	+ / +	169	-81,5	d	d	±	
miR-0319 a-e	miR-0319 e	Peaxi162Scf01126	153718	153738	153564	153584	+ / +	175	-87,3	e	e	< 1	
miR-0319 h,i	miR-0319 h	Peaxi162Scf00073	344655	344674	344505	344524	+ / -	172	-73,9	h	h	-	
miR-0319 h,i	miR-0319 i	Peaxi162Scf00015	942312	942331	942166	942185	+ / +	166	-80,1	i	i	<1 (200-1000)	
miR-0390												miRNA* variant?	
miR-0390 a-c	miR-0390 a	Peaxi162Scf00053	1522421	1522441	1522555	1522568	+ / -	148	-44,9	a	a	< 1	bp
miR-0390 a-c	miR-0390 b	Peaxi162Scf00055	2136472	2136492	2136598	2136611	+ / -	140	-52,1	b	b,c	5-20	bp
miR-0390 a-c	miR-0390 c	Peaxi162Scf00744	142676	142696	142776	142789	+ / -	114	-50,8	c	b,c	"	
miR-0393													
miR-0393 a-c	miR-0393 a	Peaxi162Scf00164	1615576	1615597	1615630	1615650	+ / +	75	-34,2	a	a,b	-	
miR-0393 a-c	miR-0393 b	Peaxi162Scf01010	290865	290886	290937	290957	+ / -	93	-38,0	b	a,b	-	
miR-0393 a-c	miR-0393 c	Peaxi162Scf00444	791670	791691	791729	791750	+ / +	83	-44,0	c	c	< 1*	
miR-0394													
miR-0394 a-c	miR-0394 a	Peaxi162Scf00011	3217427	3217446	3217514	3217533	+ / -	107	-40,6	a	a	-	
miR-0394 a-c	miR-0394 b	Peaxi162Scf00174	1333306	1333325	1333380	1333399	+ / +	94	-36,0	b	b,c	-	
miR-0394 a-c	miR-0394 c	Peaxi162Scf00064	912723	912742	912817	912836	+ / +	114	-42,5	c	b,c	-	
miR-0395												no miRNA*?	
miR-0395 a-m	miR-0395 a	Peaxi162Scf01030	218112	218132	218146	218166	+ / +	83	-28,6	a	a	-	XI
miR-0395 a-m	miR-0395 b	Peaxi162Scf01030	25579	25599	25513	25533	+ / -	87	-35,2	b	b-e	±	XI
miR-0395 a-m	miR-0395 c	Peaxi162Scf01030	94806	94826	94740	94760	+ / -	87	-34,9	c	b-e	"	XI
miR-0395 a-m	miR-0395 d	Peaxi162Scf01030	65179	65199	65113	65133	+ / -	87	-36,8	d	b-e	"	XI
miR-0395 a-m	miR-0395 e	Peaxi162Scf01030	103353	103373	103286	103306	+ / -	88	-32,6	e	b-e	"	XI
miR-0395 a-m	miR-0395 f	Peaxi162Scf01030	50836	50856	50770	50790	+ / -	87	-30,1	f	f	-	XI

Family Name / Member Name	Species Name	Petunia Scaffold	Mature Begin [1]	Mature End [1]	Star Begin	Star End	Mature Strand [2]	Pre Length	MFE kcal/mol	Pre [3] Type	Star [4] Type	Star [5] Present	Cluster [6]
miR-0395 a-m	miR-0395 g	Peaxi162Scf01030	71897	71917	71835	71855	+ / -	83	-39,6	**g**	**g-h**	-	XI
miR-0395 a-m	miR-0395 h	Peaxi162Scf01030	111565	111585	111502	111522	+ / -	82	-38,4	**h**	**g-h**	-	XI
miR-0395 a-m	miR-0395 i	Peaxi162Scf01030	77686	77706	77618	77638	+ / -	89	-39,8	**i**	**I-j**	-	XI
miR-0395 a-m	miR-0395 i	Peaxi162Scf01030	132252	132272	132184	132204	+ / -	89	-39,8	**i**	**I-j**	-	XI
miR-0395 a-m	miR-0395 i	Peaxi162Scf49057	571	591	503	523	+ / +	89	-39,8	**i**	**I-j**	-	XI
miR-0395 a-m	miR-0395 j	Peaxi162Scf01030	169957	169977	169890	169910	+ / -	88	-47,3	**j**	**I-j**	-	XI
miR-0395 a-m	miR-0395 k	Peaxi162Scf01030	71682	71702	71623	71643	+ / -	80	-39,8	**k**	**k-l**	< 1	XI
miR-0395 a-m	miR-0395 l	Peaxi162Scf01030	111350	111370	111292	111312	+ / -	79	-39,3	**l**	**k-l**	< 1	XI
miR-0396													
miR-0396 a,b	miR-0396 a	Peaxi162Scf00088	1421103	1421123	1421170	1421190	+ / +	88	-38,8	**a**	**a**	1-5	XII
miR-0396 a,b	miR-0396 b	Peaxi162Scf00313	689364	689384	689434	689454	+ / +	91	-35,6	**b**	**b**	~ 1	XIII
miR-0396 c,d	miR-0396 c	Peaxi162Scf00088	413076	413096	413166	413186	+ / -	111	-44,6	**c**	**c**	5-20*	XII
miR-0396 c,d	miR-0396 d	Peaxi162Scf00313	318074	318094	318151	318171	+ / -	98	-46,0	**d**	**d**	< 1	XIII
miR-0397													
miR-0397 a,b	miR-0397 a	Peaxi162Scf01216	63013	63032	63093	63102	+ / -	100	-36,4	**a**	**a**	< 1	
miR-0397 a,b	miR-0397 b	*not in Pax*										cannot be tested	
miR-0398													
miR-0398 a,b	miR-0398 a	Peaxi162Scf00314	1043473	1043493	1043409	1043429	+ / -	86	-36,1	**a**	**a**	-	
miR-0398 a,b	miR-0398 b	Peaxi162Scf00451	916623	916643	916560	916580	+ / +	84	-33,5	**b**	**b**	-	
miR-0398 e	miR-0398 e	Peaxi162Scf00004	4330613	4330633	4330548	4330568	+ / +	86	-34,8	**e**	**e**	< 1	
miR-0398 h	miR-0398 h	Peaxi162Scf00272	876761	876781	876833	876853	+ / +	93	-32,0	**h**	**h**	-	bp
miR-0399													
miR-0399 a-c	miR-0399 a	Peaxi162Scf00192	437679	437699	437615	437635	+ / +	85	-33,2	**a**	**a**	±	XIV
miR-0399 a-c	miR-0399 b	Peaxi162Scf00192	397850	397870	397809	397829	+ / +	62	-37,6	**b**	**b**	-	XIV
miR-0399 a-c	miR-0399 c	Peaxi162Scf00192	460001	460021	459936	459956	+ / +	86	-37,9	**c**	**c**	miR399c-j	XIV
miR-0399 c-g	miR-0399 d	Peaxi162Scf00192	382039	382059	382107	382127	+ / +	89	-37,8	**d**	**d**	-	XIV
miR-0399 c-g	miR-0399 e	Peaxi162Scf00192	391466	391486	391539	391559	+ / +	96	-35,1	**e**	**e-g**	±*	XIV
miR-0399 c-g	miR-0399 f	Peaxi162Scf00192	377993	378013	378048	378068	+ / +	78	-41,7	**f**	**e-g**	"	XIV
miR-0399 c-g	miR-0399 g	Peaxi162Scf00192	372149	372169	372206	372226	+ / +	80	-37,8	**g**	**e-g**	"	XIV
miR-0403													
miR-0403	miR-0403 a	Peaxi162Scf00770	819177	819197	819111	819131	+ / -	87	-39,8	**a**	**a**	5-20	
miR-0408													
miR-0408	miR-0408 a	Peaxi162Scf00071	263010	263030	262956	262976	+ / +	75	-31,7	**a**	**a**	< 1	
miR-0477													
miR-0477	miR-0477 a	Peaxi162Scf00786	33423	33443	33491	33512	+ / -	90	-40,4	**a**	**a**	< 1	
miR-0479													
miR-0479	miR-0479 a	Peaxi162Scf00740	288718	288738	288665	288685	+ / -	74	-33,4	**a**	**a**	< 1	
miR-0482													
miR-0482	miR-0482 a	Peaxi162Scf00201	1129749	1129770	1129694	1129714	+ / -	77	-30,0	**a**	**a**	1-5	
miR-0827													
miR-0827	miR-0827 a	Peaxi162Scf00169	982228	982248	982166	982186	+ / -	83	-31,4	**a**	**a**	5-20	
miR-2111													
miR-2111 a,b	miR-2111 a	Peaxi162Scf00094	92512	92532	92559	92579	+ / +	68	-29,0	**a**	**a**	-	
miR-2111 a,b	miR-2111 b	Peaxi162Scf00318	276851	276871	276902	276922	+ / +	72	-36,3	**b**	**b**	-	
miR-6149													
miR-6149	miR-6149 a	Peaxi162Scf00562	868808	868828	868882	868902	+ / +	95	-36,1	**a**	**a**	~ 1	
miR-8016													
miR-8016	miR-8016 a	Peaxi162Scf00068	610327	610350	610283	610306	+ / -	68	-23,8	**a**	**a**	~ 1	
n = 30 miRNA families, 44 family members, 140 loci + 5 (near) exact duplicates													
miR-6164	**miR-6164**	*low support by frequencies and 2nd structure*											
miR-6164	miR-6164 a	Peaxi162Scf00164	733618	733638	733831	733851	+ / +	234	-54,3	**a**	**a**	< 1	bp
miR-6164	miR-6164 b	Peaxi162Scf00714	594507	594527	594720	594740	+ / +	234	-57,2	**b**	**b**	-	bp
miR-6164	miR-6164 c	Peaxi162Scf00442	637516	637536	637729	637749	+ / +	234	-75,0	**c**	**c**	-	bp
miR-6164	miR-6164 d	Peaxi162Scf00397	252952	252972	253167	253187	+ / -	236	-64,9	**d**	**d**	-	bp
miR-6164	miR-6164 e	Peaxi162Scf00978	247786	247806	247968	247988	+ / +	203	-60,4	**e**	**e**	±	bp
miR-6164	miR-6164 f	Peaxi162Scf00597	889597	889617	889795	889813	+ / +	218	-63,5	**f**	**f**	-	bp
miR-6164	miR-6164 g	Peaxi162Scf00016	1293308	1293328	1293439	1293513	+ / +	205	-53,6	**g**	**g**	-	bp
miR-6164	miR-6164 h	Peaxi162Scf00972	507422	507442	507636	507656	+ / +	235	-49,1	**h**	**h**	-	bp
miR-6164	miR-6164 i	Peaxi162Scf00856	266091	266111	266305	266325	+ / +	235	-54,7	**i**	**i**	-	bp
miR-6164	miR-6164 j	Peaxi162Scf00063	682321	682341	682532	682552	+ / +	232	-50,8	**j**	**j,k**	-	bp
miR-6164	miR-6164 k	Peaxi162Scf00795	110704	110724	110921	110941	+ / +	238	-49,3	**k**	**j,k**	-	bp
miR-6164	miR-6164 l	Peaxi162Scf00064	1646757	1646777	1646948	1646968	+ / -	212	-45,1	**l**	**l**	-	bp
miR-6164	miR-6164 m	Peaxi162Scf00003	4635233	4635253	4635448	4635468	+ / -	236	-47,7	**m**	**m**	-	bp
miR-6164	miR-6164 n	Peaxi162Scf00203	1361450	1361470	1361646	1361666	+ / +	216	-36,3	**n**	**n**	-	bp
miR-6164	miR-6164 o	Peaxi162Scf00233	576719	576739	576935	576955	+ / +	235	-44,5	**o**	**o**	-	bp
miR-6164	miR-6164 p	Peaxi162Scf00011	2183277	2183297	2183471	2183491	+ / -	215	-50,6	**p**	**p**	-	bp

Family Name / Member Name	Species Name	Petunia Scaffold	Mature Begin [1]	Mature End [1]	Star Begin	Star End	Mature Strand [2]	Pre Length	MFE kcal/mol	Pre [3] Type	Star [4] Type	Star [5] Present	Cluster [6]
miR-6164	miR-6164 q	Peaxi162Scf00481	358092	358112	358274	358294	+ / -	203	-60,6	**q**	**q**	-	bp
miR-6164	miR-6164 r	Peaxi162Scf00538	299958	299978	300175	300195	+ / -	238	-48,3	**r**	**r**	-	bp
miR-6164	miR-6164 s	Peaxi162Scf00526	495241	495261	495545	495275	+ / +	235	-33,6	**s**	**s**	-	bp
miR-6164	miR-6164 t	Peaxi162Scf00070	47771	47791	47966	47986	+ / +	215	-51,9	**t**	**t**	-	bp
miR-6164	miR-6164 u	Peaxi162Scf01078	16122	16142	16334	16354	+ / -	233	-76,2	**u**	**u**	-	bp
n = one miRNA family, one family member, 20 loci + 1 exact duplicate													
miRNAs overlapping at loci		*information about localisation in the genome and characteristics of pre-miRNAs can be found above*											
miR-0157 a-e v2	miR-0157 a v2		All five loci		extra 5'-U							1-5	
"	miR-0157 b v2											-	
"	miR-0157 c v2											-	
"	miR-0157 d v2											-	
"	miR-0157 e v2											*?*	
miR-0171 a-g v2	miR-0171 a v2		Three of seven loci		3nt shifted							1-5	
"	miR-0171 b v2											~ 1	
"	miR-0171 c v2											< 1	
miR-0171 j,k v2	miR-0171 j v2		Both loci		minus 3'-U							1-5	
"	miR-0171 k v2											"	
miR-0171 j,k v3	miR-0171 j v3		One of two loci		3nt shifted							-	
miR-0172 a-g v2	miR-0172 a v2		All seven loci		minus 3'-AU							1-5	
"	miR-0172 b v2											1-20	
"	miR-0172 c v2											"	
"	miR-0172 d v2											1-5	
"	miR-0172 e v2											~ 1	
"	miR-0172 f v2											-	
"	miR-0172 g v2											1-5	
miR-0172 j,k v2	miR-0172 j v2		Both loci		minus 3'-CAG							~ 5	
"	miR-0172 k v2											-	
miR-0319 a-i v2	miR-0319 a v2		Six of seven loci		1nt shifted							-	
"	miR-0319 b v2											-	
"	miR-0319 c v2											-	
"	miR-0319 d v2											-	
"	miR-0319 e v2											-	
"	miR-0319 h v2											-	
miR-0319 h,i v3	miR-0319 h v3		One of two loci		extra 3'-U							-	
miR-0390 a-c v2	miR-0390 a v2		All three loci		minus 3'-C							-	
"	miR-0390 b v2											1-20	
"	miR-0390 c v2											"	
miR-0397 a,b v2	miR-0397 a v2		One locus		2nt shifted							< 1	
miR-0398 a,b v2	miR-0398 a v2		Both loci		1nt shifted							-	
"	miR-0398 b v2											-	
miR-0398 h v2	miR-0398 h v2		The one locus		reverse complement							< 1	
miR-2111 a,b v2	miR-2111 a v2		Both loci		reverse complement							-	
"	miR-2111 b v2												
n = 13 miRNA variants, overlapping at 36 loci													
miRNAs not supported by Petunia genome sequence							*miR sequence*				*cannot be tested*		
miR-0159 var	miR-0159 var	not in *Pax/Pin* (is one longer)			3'-end mismatch		UUUGGAUUGAAGGGAGCUCUU						
miR-0164 var	miR-0164 var	not in *Pax/Pin* (one mismatch)			3'-end mismatch		UGGAGAAGCAGGGCACGUGCG						
miR-0166 var	miR-0166 var	not in *Pax/Pin* (is one longer)			internal mismatch		UCGGACCAGGCUUCAUCCCCC						
miR-0167 var	miR-0167 var	no pre-miR in *Pax*			3'-end mismatch		UGAAGCUGCCAGCAUGAUCUGG						
miR-0172 var	miR-0172 var1	not in *Pax/Pin* (is one longer)			end-nt exchange 1		AGAAUCUUGAUGAUGCUGCAG						
miR-0172 var	miR-0172 var2	not in *Pax/Pin* (is one longer)			end-nt exchange 1		GGAAUCUUGAUGAUGCUGCAU						
miR-0390 var	miR-0390 var	not in *Pax/Pin* (is one longer)			internal mismatch		AAGCUCAGGAGGGAUAGCACC						
n = 7 miR variants, showing one-two mismatches to the genome sequence													

[1] **Mature Begin and End numbers in color** indicate localisation of the miRNA at the 3'-side (second arm)

[2] **Mature Strand** +/- indicate the loclisation of the *MIR* in the Reverse Complement of the Scaffold

[3] ***MIR* types that occur more than once** are either exact duplicates or a result of incomplete assembly; near exact duplicates are denoted 2

[4] **Star Types** with more than one letter are shared by the particular *MIR*s of that family

[5] **Star Presences** (in RPM) support the existence of the miRNA; **E**xact frequencies are in Table 1B; * = annotated as miRNAs in miRBase but miRNA*s in our data light orang*e* = unexpected absence of miRNA*; *green* = variants that are likely to exist as the main form *(dark)* or in addition (*light*)

[6] **Clusters *of MIR loci*** in the *Petunia* genome: I. miR156b +156c, inverted at 180.000 nt; II. miR156d +156e, direct at 20.000 nt; III. miR156k +157b +157c, direct in 1000 nt; IV. miR157a +159b, direct at 40.000 nt; V. miR157d +159a, inverted in 80.000 nt; VI. miR167b +167c, direct at 15.000 nt; VII. miR169b +169d, direct at 100.000 nt; VIII. miR169j-n, four direct and one indirect copies in 100.000 nt; IX. miR169o-r, four direct copies at distances 10.000 - 60.000 nt; X. miR172d-f, two direct and one indirect copies in 160.000 nt; XI. miR395a-l, all but one direct copies, within 150.000 nt; XII. miR396a +396c, inverted at 10.000 nt; XIII. miR396b +396 d, inverted at 20.000 nt; XIV. miR399a-g, direct copies within 90.000 nt bp = bifurcation in precursor

Table 2C. DNA identification numbers, *MIR* characteristics and miRNA*-type of conserved miRNAs in *Petunia inflata*

Family Name / Member Name	Species Name	Petunia Scaffold	Mature Begin [1]	Mature End [1]	Star Begin	Star End	Mature Strand [2]	Pre Length	MFE kcal/mol	Pre [3] Type	Star [4] Type	Star [5] Present	Cluster [6]
Petunia inflata													
miR-0156												see *Pax*	
miR-0156 a-g	miR-0156 a	Peinf101Scf00248	153076	153095	153142	153162	+ / +	87	-42,6	**a**	**a**	(Table 2B)	
miR-0156 a-g	miR-0156 b	Peinf101Scf00416	1348261	1348280	1348330	1348350	+ / +	90	-48,4	**b**	**b-d**		I
miR-0156 a-g	miR-0156 c	Peinf101Scf00416	1238183	1238202	1238252	1238271	+ / -	90	-48,6	**c**	**b-d**		I
miR-0156 a-g	miR-0156 d	Peinf101Scf00416	1434688	1434707	1434763	1434782	+ / -	96	-44,1	**d'**	**b-d**		II
miR-0156 a-g	miR-0156 e	Peinf101Scf00416	1441309	1441328	1441374	1441394	+ / -	86	-45,3	**e'**	**e**		II
miR-0156 a-g	miR-0156 f	Peinf101Scf00716	672868	672887	672936	672957	+ / +	90	-47,8	**f'**	**f'**	-	
miR-0156 a-g	miR-0156 g	Peinf101Scf00594	2288027	2288046	2288091	2288112	+ / +	86	-46,9	**g'**	**g**		
miR-0156 j-l	miR-0156 j	Peinf101Scf00883	1289315	1289334	1289380	1289396	+ / -	82	-22,3	**j**	**j**		
miR-0156 j-l	miR-0156 k	Peinf101Scf01139	41406	41425	41470	41490	+ / -	85	-48,7	**k**	**k**		III
miR-0156 j-l	miR-0156 l	Peinf101Scf00055	1676777	1676796	1676836	1676856	+ / -	80	-47,6	**l'**	**l**		
miR-0157													
miR-0157 a-e	miR-0157 a	Peinf101Scf01889	60248	60267	60307	60327	+ / +	80	-44	**a**	**a**		IV
miR-0157 a-e	miR-0157 b	Peinf101Scf01139	41828	41847	41891	41911	+ / -	84	-44,5	**b'**	**b-c'**	-	III
miR-0157 a-e	miR-0157 c	Peinf101Scf01139	42198	42217	42263	42283	+ / -	85	-44,8	**c**	**b-c'**	-	III
miR-0157 a-e	miR-0157 d	Peinf101Scf01633	1101129	1101148	1101207	1101227	+ / -	99	-43,6	**d'**	**d**		V
miR-0157 a-e	miR-0157 e	Peinf101Scf01190	91147	91166	91215	91235	+ / -	89	-42,7	**e'**	**e**		
miR-0159													
miR-0159 a-c	miR-0159 a	Peinf101Scf01633	1178326	1178346	1178178	1178198	+ / -	169	-81,6	**a**	**a**		V
miR-0159 a-c	miR-0159 b	Peinf101Scf01889	120677	120697	120526	120546	+ / +	172	-67,1	**b'**	**b**		IV
miR-0159 a-c	miR-0159 c	Peinf101Scf00962	1432242	1432262	1432089	1432110	+ / -	174	-77,83	**c''**	**c**		
miR-0160													
miR-0160 a-d	miR-0160 a	Peinf101Scf00824	127821	127841	127884	127904	+ / +	84	-45,5	**a**	**a,b**		
miR-0160 a-d	miR-0160 b	Peinf101Scf02350	13466	13486	13528	13548	+ / -	83	-51,8	**b'**	**a,b**		
miR-0160 a-d	miR-0160 b	Peinf101Ctg1210064	168	188	230	250	+ / -	83	-51,8	**b'**	**a,b**		
miR-0160 a-d	miR-0160 c	Peinf101Scf01590	477840	477860	477908	477928	+ / -	89	-46,1	**c**	**c**		
miR-0160 a-d	miR-0160 d	Peinf101Scf00665	1581880	1581900	1581942	1581962	+ / +	83	-38,2	**d**	**d**		
miR-0162													
miR-0162 a	miR-0162 a	Peinf101Scf04403	266013	266033	265945	265965	+ / +	87	-36,9	**a'**	**a**		
miR-0164													
miR-0164 a-d	miR-0164 a	Peinf101Scf00633	695805	695825	695871	695891	+ / +	87	-42,9	**a**	**a**		
miR-0164 a-d	miR-0164 b	Peinf101Scf00782	893255	893275	893387	893407	+ / +	153	-50,1	**b''**	**b**		
miR-0164 a-d	miR-0164 c	Peinf101Scf00116	780187	780207	780353	780373	+ / +	187	-62,72	**c''**	**c**		bp
miR-0164 a-d	miR-0164 d	Peinf101Scf00650	1687678	1687698	1687748	1687768	+ / -	91	-38,8	**d'''**	**d**		
miR-0164 g,h	miR-0164 g	Peinf101Scf03112	24232	24252	24288	24308	+ / +	77	-39,2	**g'**	**g**		
miR-0164 g,h	miR-0164 g	Peinf101Scf03915	27279	27299	27335	27355	+ / -	77	-39,5	**g'2**	**g**		
miR-0164 g,h	miR-0164 h	Peinf101Scf01889	1640407	1640427	1640463	1640483	+ / +	77	-43	**h'**	**h**		
miR-0166													
miR-0166 a-h	miR-0166 a	Peinf101Scf01533	451746	451766	451630	451650	+ / -	137	-49,3	**a'**	**a,b**		bp
miR-0166 a-h	miR-0166 b	Peinf101Scf00559	37371	37391	37226	37246	+ / +	166	-50,92	**b'**	**a,b**		
miR-0166 a-h	miR-0166 c	Peinf101Scf01061	857530	857550	857453	857473	+ / +	98	-36,6	**c'**	**c**		bp
miR-0166 a-h	miR-0166 d	Peinf101Scf01482	77652	77672	77516	77536	+ / -	157	-46,26	**d**	**d**		bp
miR-0166 a-h	miR-0166 e	Peinf101Scf04383	33545	33565	33472	33492	+ / -	94	-37,8	**e'**	**e,f**		
miR-0166 a-h	miR-0166 f	Peinf101Scf00034	1024474	1024494	1024222	1024242	+ / -	273	-89,44	**f''**	**e,f**		bp
miR-0166 a-h	miR-0166 g	Peinf101Scf04519	14801	14821	14742	14762	+ / +	80	-42	**g**	**g**		
miR-0166 a-h	miR-0166 h	Peinf101Scf00571	220914	220934	220833	220853	+ / -	102	-30,97	**h''**	**h''**	< 1	
miR-0166 k	miR-0166 k	Peinf101Scf00255	773552	773572	773413	773433	+ / +	160	-61,3	**k''**	**k**		
miR-0167													
miR-0167 a-e	miR-0167 a	Peinf101Scf10575	8228	8248	8287	8306	+ / -	78	-32,6	**a'**	**a**		
miR-0167 a-e	miR-0167 a	Peinf101Scf10576	11219	11239	11278	11297	+ / -	78	-32,6	**a'**	**a**		
miR-0167 a-e	miR-0167 b	Peinf101Scf00571	2707851	2707871	2707908	2707927	+ / -	77	-34,2	**b'**	**b**		VI
miR-0167 a-e	miR-0167 c	Peinf101Scf00571	2694062	2694082	2694140	2694162	+ / -	101	-34	**c'**	**c**		VI
miR-0167 a-e	miR-0167 d	Peinf101Scf00309	86400	86420	86448	86467	+ / +	68	-32,9	**d''**	**d**		
miR-0167 a-e	miR-0167 e	Peinf101Scf01061	408456	408476	408537	408559	+ / -	104	-39,56	**e'**	**e'**	-	
miR-0168													
miR-0168 a	miR-0168 a	Peinf101Scf01093	1076002	1076022	1076088	1076108	+ / +	107	-43,82	**a''**	**a-c**		
miR-0168 b-c	miR-0168 b	Peinf101Scf00276	558697	558717	558754	558774	+ / -	78	-37,3	**b**	**a-c**		
miR-0168 b-c	miR-0168 c	Peinf101Scf00146	1120631	1120651	1120715	1120735	+ / -	105	-40,7	**c'''**	**a-c**		
miR-0169													
miR-0169 a,b	miR-0169 a	Peinf101Scf01940	99479	99499	99566	99586	+ / -	107	-50,6	**a'**	**a**		bp
miR-0169 a,b	miR-0169 b	Peinf101Scf01349	1084079	1084099	1084127	1084147	+ / -	69	-41,3	**b'**	**b'**	< 1	VII
miR-0169 c,d	miR-0169 c(*BL*	Peinf101Scf01969	55267	55287	55366	55386	+ / -	120	-50,3	**c**	**c**		
miR-0169 c,d	miR-0169 d	Peinf101Scf01349	895351	895371	895410	895430	+ / -	80	-37,2	**d'**	**d'**	-	VII

Family Name / Member Name	Species Name	Petunia Scaffold	Mature Begin [1]	Mature End [1]	Star Begin	Star End	Mature Strand [2]	Pre Length	MFE kcal/mol	Pre [3] Type	Star [4] Type	Star [5] Present	Cluster [6]
miR-0169 g-x	miR-0169 g	Peinf101Scf00073	1192289	1192308	1192363	1192380	+ / -	92	-50,5	g'	g		
miR-0169 g-x	miR-0169 h	Peinf101Scf00930	1419064	1419083	1419137	1419157	+ / +	93	-34,1	h'2	h		
miR-0169 g-x	miR-0169 h	Peinf101Scf00930	1311589	1311608	1311660	1311680	+ / +	92	-33,9	h'	h		
miR-0169 g-x	miR-0169 i	Peinf101Scf00590	2904866	2904885	2904940	2904958	+ / -	92	-39,5	i'	i		
miR-0169 g-x	miR-0169 j	Peinf101Scf00594	2274134	2274153	2274219	2274238	+ / -	104	-33,5	j''	j		VIII
miR-0169 g-x	miR-0169 k	Peinf101Scf00594	2274361	2274380	2274441	2274460	+ / -	99	-49,7	k'	k'	-	VIII
miR-0169 g-x	miR-0169 l	Peinf101Scf00594	2277637	2277656	2277716	2277735	+ / -	99	-43,6	l'	l		VIII
miR-0169 g-x	miR-0169 m	Peinf101Scf00594	2277844	2277864	2277909	2277930	+ / -	86	-40,8	m	m		VIII
miR-0169 g-x	miR-0169 n	Peinf101Scf00594	56302	56322	56392	56412	+ / +	110	-43,8	n'	n		VIII
miR-0169 g-x	miR-0169 o	Peinf101Scf01210	337550	337569	337657	337675	+ / +	126	-46,5	o'	o		IX
miR-0169 g-x	miR-0169 o	Peinf101Scf01210	362721	362740	362827	362847	+ / +	126	-42,7	o'2	o		IX
miR-0169 g-x	miR-0169 p	Peinf101Scf01210	375144	375164	375222	375242	+ / +	98	-34,9	p'	p		IX
miR-0169 g-x	miR-0169 q	Peinf101Scf01210	381669	381689	381771	381790	+ / +	122	-54,5	q'	q		IX
miR-0169 g-x	miR-0169 r	Peinf101Scf01210	397783	397803	397879	397898	+ / +	116	-56,5	r'	r		IX
miR-0169 g-x	miR-0169 t	Peinf101Scf04980	94555	94575	94627	94645	+ / -	91	-37,7	t'	t		
miR-0169 g-x	miR-0169 u	Peinf101Scf00728	503985	504005	504078	504096	+ / -	112	-45,1	u'	u		
miR-0169 g-x	miR-0169 v	Peinf101Scf01363	53018	53038	53119	53139	+ / -	121	-39,25	v	v'	-	
miR-0169 g-x	miR-0169 w	Peinf101Scf00482	493565	493585	493635	493653	+ / -	89	-33,6	w'	w'	-	
miR-0169 g-x	miR-0169 x	Peinf101Scf00637	125588	125608	125686	125706	+ / -	118	-42	x'''	x'''		
miR-0171													
miR-0171 a-g	miR-0171 a	Peinf101Scf01020	293400	293420	293343	293363	+ / +	78	-33,3	a'	a		
miR-0171 a-g	miR-0171 b	Peinf101Scf01775	480758	480778	480704	480724	+ / +	75	-32,7	b'	b		
miR-0171 a-g	miR-0171 c	Peinf101Scf00351	1216991	1217011	1216936	1216956	+ / +	76	-28,8	c'	c'	~ 1	
miR-0171 a-g	miR-0171 d	Peinf101Scf03843	21984	22004	21922	21942	+ / -	83	-31,2	d'	d		
miR-0171 a-g	miR-0171 e	Peinf101Scf02008	585237	585257	585184	585204	+ / +	74	-37,8	e'	e		
miR-0171 a-g	miR-0171 f	Peinf101Scf00562	667676	667696	667602	667622	+ / +	95	-29,6	f'	f'	-	
miR-0171 a-g	miR-0171 g	not in *Pin*									cannot be tested		
miR-0171 j,k	miR-0171 j	Peinf101Scf03005	427270	427290	427205	427225	+ / +	86	-40,2	j	j,k		
miR-0171 j,k	miR-0171 k	Peinf101Scf05332	60132	60152	60067	60087	+ / +	86	-37,9	k	j,k		
miR-0171 n,o	miR-0171 n	Peinf101Scf01020	200868	200888	200807	200827	+ / +	82	-43,4	n	n		
miR-0171 n,o	miR-0171 o	Peinf101Scf00039	1634165	1634185	1634102	1634122	+ / -	84	-32,1	o'	o		
miR-0172													
miR-0172 a-g	miR-0172 a	Peinf101Scf01075	299092	299112	299024	299044	+ / +	89	-34,1	a'	a		
miR-0172 a-g	miR-0172 b	Peinf101Scf02125	117960	117980	117878	117898	+ / +	103	-51	b'	b,c		
miR-0172 a-g	miR-0172 c	Peinf101Scf05429	117872	117892	117878	117898	+ / +	99	-42,4	c'	b,c		
miR-0172 a-g	miR-0172 c	Peinf101Scf10256	642	662	564	584	+ / +	99	-42,4	c'	b,c		
miR-0172 a-g	miR-0172 c	Peinf101Scf16754	2052	2072	1974	1994	+ / +	99	-42,4	c'	b,c		
miR-0172 a-g	miR-0172 d	Peinf101Scf00406	82005	82025	81919	81939	+ / -	107	-44,87	d'	d		X
miR-0172 a-g	miR-0172 e	Peinf101Scf00406	121634	121654	121551	121571	+ / -	104	-33,1	e'	e		X
miR-0172 a-g	miR-0172 f	Peinf101Scf00406	574428	574448	574340	574360	+ / +	109	-43,1	f	f		X
miR-0172 a-g	miR-0172 g	Peinf101Scf00871	497681	497701	497598	497618	+ / +	102	-43,1	g	g		
miR-0172 j,k	miR-0172 j	Peinf101Scf00244	1647202	1647222	1647114	1647134	+ / +	109	-48,4	j'	j		
miR-0172 j,k	miR-0172 k	Peinf101Scf00972	45262	45282	45159	45179	+ / +	124	-49,4	k'	k		
miR-0319													
miR-0319 a-e	miR-0319 a	Peinf101Scf00962	716287	716307	716139	716159	+ / -	170	-82,9	a'	a		
miR-0319 a-e	miR-0319 b	Peinf101Scf00664	982813	982833	982660	982680	+ / +	174	-81,84	b'	b		
miR-0319 a-e	miR-0319 c	Peinf101Scf00522	166445	166465	166294	166314	+ / -	172	-83,89	c'	c		
miR-0319 a-e	miR-0319 d	Peinf101Scf00528	35684	35704	35536	35556	+ / -	169	-75,3	d'	d		
miR-0319 a-e	miR-0319 e	Peinf101Scf00198	278501	278521	278347	278367	+ / -	175	-84,9	e'	e		
miR-0319 h,i	miR-0319 h	Peinf101Scf01961	137932	137951	137782	137801	+ / +	172	-73,9	h	h		
miR-0319 h,i	miR-0319 h	Peinf101Scf03109	19299	19318	19147	19167	+ / -	172	-73,9	h	h		
miR-0319 h,i	miR-0319 i	Peinf101Scf04017	201189	201208	201043	201062	+ / +	166	-80,1	i	i		
miR-0319 h,i	miR-0319 i2	Peinf101Scf03422	62307	62326	62149	62160	+ / -	166	-80,6	i'	i		
miR-0390													
miR-0390 a-c	miR-0390 a	Peinf101Scf00507	393294	393314	393413	393433	+ / -	140	-41,3	a'	a		bp
miR-0390 a-c	miR-0390 b	Peinf101Scf00085	2505672	2505692	2505791	2505811	+ / -	140	-54,52	b''	b,c		bp
miR-0390 a-c	miR-0390 c	Peinf101Scf00782	938203	938223	938308	838328	+ / -	126	-54,5	c'''	b,c		bp
miR-0393													
miR-0393 a-c	miR-0393 a	Peinf101Scf00244	1044694	1044715	1044748	1044769	+ / -	76	-34,02	a'	a,b		
miR-0393 a-c	miR-0393 b	Peinf101Scf02361	188989	189010	189060	189080	+ / -	93	-37,99	b'	a,b		
miR-0393 a-c	miR-0393 b2	Peinf101Scf20721	1322	1343	1394	1415	+ / -	94	-39,5	b'2	a,b		
miR-0393 a-c	miR-0393 c	Peinf101Scf01452	244708	244729	244771	244791	+ / -	84	-35	c''	c		
miR-0394													
miR-0394 a-c	miR-0394 a	Peinf101Scf00255	722319	722338	722406	722425	+ / -	107	-40,6	a	a		bp
miR-0394 a-c	miR-0394 b	Peinf101Scf01043	258816	258835	258886	258905	+ / -	89	-34,9	b'	b,c		bp
miR-0394 a-c	miR-0394 c	Peinf101Scf00471	913877	913896	913977	913996	+ / -	119	-39,72	c''	b,c		bp

Family Name / Member Name	Species Name	Petunia Scaffold	Mature Begin [1]	Mature End [1]	Star Begin	Star End	Mature Strand [2]	Pre Length	MFE kcal/mol	Pre [3] Type	Star [4] Type	Star [5] Present	Cluster [6]
miR-0395													
miR-0395 a-m	miR-0395 a	Peinf101Scf00077	885984	886004	885922	885943	+ / -	83	-29,1	a'	a		XI
miR-0395 a-m	miR-0395 a	Peinf101Scf00077	942296	942316	335557	335578	+ / -	83	-29,1	a'	a		XI
miR-0395 a-m	miR-0395 a2	Peinf101Scf01368	335619	335639	335557	335578	+ / -	83	-33,4	a''2	a		XI
miR-0395 a-m	miR-0395 b	Peinf101Scf00077	140823	140843	140758	140778	+ / +	86	-37,44	b'	b-e		XI
miR-0395 a-m	miR-0395 b2	Peinf101Scf00077	84198	84218	84132	84152	+ / +	87	-35,2	b'2	b-e		XI
miR-0395 a-m	miR-0395 c	Peinf101Scf00077	124547	124567	124481	124501	+ / +	87	-34,86	c	b-e		XI
miR-0395 a-m	miR-0395 c	Peinf101Scf00077	158247	158267	158181	158201	+ / +	87	-34,86	c	b-e		XI
miR-0395 a-m	miR-0395 c2	Peinf101Scf00077	168835	168855	168769	168789	+ / +	87	-34,4	c'	b-e'	-	XI
miR-0395 a-m	miR-0395 d	Peinf101Scf00077	174173	174193	174111	174131	+ / +	83	-39,6	d'	b-e		XI
miR-0395 a-m	miR-0395 e	Peinf101Scf00077	259619	259639	259551	259571	+ / +	89	-37,92	e''	b-e		XI
miR-0395 a-m	miR-0395 e2	Peinf101Scf00077	237139	237159	237072	237092	+ / +	88	-37,9	e''2	b-e		XI
miR-0395 a-m	miR-0395 g	Peinf101Scf00077	187490	187510	187428	187448	+ / +	83	-39	g	g-h		XI
miR-0395 a-m	miR-0395 g	Peinf101Scf00077	245917	245937	245855	245875	+ / +	83	-39	g	g-h		XI
miR-0395 a-m	miR-0395 g	Peinf101Scf00077	270286	270306	270224	270244	+ / +	83	-39	g	g-h		XI
miR-0395 a-m	miR-0395 g	Peinf101Ctg1323080	549	569	487	507	+ / +	83	-39	g	g-h		XI
miR-0395 a-m	miR-0395 j	Peinf101Scf00077	204484	204504	204417	204437	+ / +	88	-46,5	j'	l-j'	-	XI
miR-0395 a-m	miR-0395 k	Peinf101Scf00077	187274	187294	187215	187235	+ / +	80	-42,7	k'	k-l'	-	XI
miR-0395 a-m	miR-0395 k	Peinf101Scf00077	245701	245721	245642	245662	+ / +	80	-42,7	k'	k-l'	-	XI
miR-0395 a-m	miR-0395 k	Peinf101Scf00077	268155	268175	268096	268116	+ / +	80	-42,7	k'	k-l'	-	XI
miR-0395 a-m	miR-0395 k	Peinf101Scf00077	270070	270090	270010	270030	+ / +	80	-42,7	k'	k-l'	-	XI
miR-0395 a-m	miR-0395 k	Peinf101Ctg1323080	333	353	274	292	+ / +	80	-42,7	k'	k-l'	-	XI
miR-0395 a-m	miR-0395 m	Peinf101Scf01368	657235	657255	657196	657216	+ / +	60	-30,3	m'''	m'''		XI
miR-0396													
miR-0396 a,b	miR-0396 a	Peinf101Scf03005	256859	256879	256926	256946	+ / -	88	-38,8	a	a		XII
miR-0396 a,b	miR-0396 b	Peinf101Scf01701	420835	420855	420903	420923	+ / -	89	-39,4	b'	b		XIII
miR-0396 c,d	miR-0396 c	Peinf101Scf03005	596153	596173	596243	596265	+ / +	112	-38,8	c'	c		XII
miR-0396 c,d	miR-0396 d	Peinf101Scf01701	49549	49569	49626	49646	+ / +	98	-46	d	d		XIII
miR-0397													
miR-0397 a,b	miR-0397 a	Peinfl101Scf01175	522323	522342	522403	522422	+ / -	100	-36.70	a'	a'	-	
miR-0397 a,b	miR-0397 b	Peinfl101Scf00490	215282	215301	215370	215389	+ / -	108	-34.90	b	b		bp
miR-0398													
miR-0398 a,b	miR-0398 a	not in *Pin*, but... [7]	129900	129920	129837	129921	+ / -		-28.50	a'	a		
miR-0398 a,b	miR-0398 b	not in *Pin*									cannot be tested		
miR-0398 e	miR-0398 e	Peinf101Scf01151	38374	38394	38309	38329	+ / +	86	-36,8	e'	e		
miR-0398 e	miR-0398 e2	Peinf101Scf12538	94	114	29	49	+ / -	86	-33,7	e'2	e		
miR-0398 h	miR-0398 h	Peinf101Scf01933	749017	749037	749097	749117	+ / -	101	-30,1	h'	h		bp
miR-0399													
miR-0399 a-c	miR-0399 a	Peinf101Scf01452	1297730	1297750	1297666	1297686	+ / -	85	-33,19	a	a		XIV
miR-0399 a-c	miR-0399 b	Peinf101Scf01452	1235200	1235220	1235159	1235179	+ / -	62	-37,6	b	b		XIV
miR-0399 a-c	miR-0399 c	Peinf101Scf01452	1323875	1323895	1323810	1323830	+ / -	86	-37,9	c	c		XIV
miR-0399 c-f	miR-0399 d	Peinf101Scf01452	1217627	1217647	1217695	1217715	+ / -	89	-39	d	d		XIV
miR-0399 c-f	miR-0399 e	Peinf101Scf01452	1209064	1209084	1209122	1209142	+ / -	96	-36,51	e	e-g		XIV
miR-0399 c-f	miR-0399 f	Peinf101Scf01452	1214685	1214705	1214742	1214762	+ / -	78	-43,6	f	e-g		XIV
miR-0399 c-f	miR-0399 g	Peinf101Scf01452	1228870	1228890	1228945	1228965	+ / -	79	-40,1	g'	e-g		XIV
miR-0403													XIV
miR-0403	miR-0403 a	Peinf101Scf01724	11346	11366	11280	11300	+ / +	87	-39,45	a'	a		
miR-0408													
miR-0408	miR-0408 a	Peinf101Scf00650	1398476	1398496	1398422	1398442	+ / -	75	-32,9	a'	a		
miR-0477													
miR-0477	miR-0477 a	Peinfl101Scf01602	276450	276470	only first 40 nt found						cannot be tested		
miR-0479													
miR-0479	miR-0479 a	Peinf101Scf03679	54512	54532	54459	54479	+ / -	74	-33,4	a	a		
miR-0482													
miR-0482	miR-0482 a	Peinf101Scf01556	190143	190164	190085	190105	+ / +	80	-33,8	a'	a		
miR-0827													
miR-0827	miR-0827 a	Peinf101Scf00276	1240887	1240907	1240825	1240845	+ / -	83	-31	a'	a		
miR-2111													
miR-2111 a,b	miR-2111 a	Peinf101Scf01173	859985	860005	860032	860052	+ / -	68	-29.00	a	a		
miR-2111 a,b	miR-2111 b	Peinf101Scf00144	306642	306662	306689	306709	+ / +	68	-32.60	b'	b		
miR-6149													
miR-6149	miR-6149 a	Peinf101Scf00947	818451	818471	818525	818545	+ / -	95	-36,1	a'	a		
miR-8016													
miR-8016	miR-8016 a	Peinf101Scf00778	1181847	1181870	1181803	1181826	+ / -	68	-23,1	a'	a		
n = 30 miRNA families, 44 family members, 140 loci + 24 (near) exact duplicated loci													

Family Name / Member Name	Species Name	Petunia Scaffold	Mature Begin [1]	Mature End [1]	Star Begin	Star End	Mature Strand [2]	Pre Length	MFE kcal/mol	Pre [3] Type	Star [4] Type	Star [5] Present	Cluster [6]
miR-6164	**miR-6164**	*low support by frequencies and 2nd structure*											
miR-6164	miR-6164 a	Peinf101Scf03477	1473	1493	32995	33015	+ / -	235	-61,8	a''	**a**		bp
miR-6164	miR-6164 a	Peinf101Scf00487	712171	712191	23839	23859	+ / -	235	-58,37	a'''	**a**		bp
miR-6164	miR-6164 b	Peinf101Scf00962	2634356	2634376	2634549	2634569	+ / -	214	-55,2	b''	**b**		bp
miR-6164	miR-6164 c	Peinf101Scf03671	16505	16525	476	496	+ / +	235	-77,4	c''	**c**		bp
miR-6164	miR-6164 d	Peinf101Scf00251	2086136	2086156	165861	165881	+ / +	235	-74,03	d''	**d**		bp
miR-6164	miR-6164 e	Peinf101Scf01189	537520	537540	873279	873299	+ / +	234	-81	e''	**e**		bp
miR-6164	miR-6164 f	Peinf101Scf00789	947527	947547	947729	947748	+ / +	222	-74,7	f''	**f**		bp
miR-6164	miR-6164 g	Peinf101Scf03539	97277	97297	974542	974561	+ / -	207	-65,7	g''	g'	-	bp
miR-6164	miR-6164 h2	Peinf101Scf00637	464624	464644	120876	120896	+ / +	235	-43,76	h''2	h''	-	bp
miR-6164	miR-6164 h	Peinf101Scf00269	380147	380167	626877	626897	+ / +	243	-45,91	h''	h''	-	bp
miR-6164	miR-6164 h3	Peinf101Scf03661	42548	42568	118097	118118	+ / +	236	-54,82	h''3	h''	-	bp
miR-6164	miR-6164 i	Peinf101Scf00230	379582	379602	379796	379816	+ / -	235	-46,9	l''	l''		bp
miR-6164	miR-6164 v	Peinf101Scf02215	178742	178762	178945	178965	+ / +	224	-60,9	v'''	v'''		bp
miR-6164	miR-6164 w	Peinf101Scf01734	43972	43992	44073	44095	+ / -	124	-24,26	w'''	w'''		bp
miR-6164	miR-6164 x	Peinf101Scf01142	687399	687419	687578	687598	+ / -	200	-41,3	x'''	x'''		bp
miR-6164	miR-6164 y	Peinf101Scf00782	2023610	2023630	2023769	2023790	+ / +	181	-32,2	y'''	y'''		bp
miR-6164	miR-6164 z	Peinf101Scf01538	260363	260383	260636	260656	+ / +	294	-76,5	z'''	z'''		bp
miR-6164	miR-6164 ae	Peinf101Scf10071	2376	2396	1334405	1334425	+ / -	220	-68,5	ae'''	ae'''		bp
miR-6164	miR-6164 aa	Peinf101Scf05020	9140	9160	9327	9346	+ / -	207	-55,46	aa'''	aa'''		bp
miR-6164	miR-6164 ab	Peinf101Scf00129	149212	149232	1781541	1781562	+ / -	278	-49,2	ab'''	ab'''		bp
miR-6164	miR-6164 ac	Peinf101Scf00692	166592	166612	166767	166787	+ / -	196	-49,79	ac''	ac,ad'''		bp
miR-6164	miR-6164 ad	Peinf101Scf01617	29714	29734	29885	29904	+ / +	191	-45,1	ad'''	ac,ad'''		bp
n = one miRNA family, one family members, 19 loci + three near exact duplicated loci													
miRNAs overlapping at loci [8]		*see information about miRNA variants Table 2B (PaxiN) and localisation in the genome and pre-miRNAs above*											
miRNAs not supported by Petunia genome sequence [8]				*see information about miRNA variants Table 2B (PaxiN)*									

[1,2,3,4] **See at the end of Table 2B**. Sequence differences between *P. inflata* and *P. axillaris* are denoted with ' (*light orange*) = 1-3 mismatches or one indel in Pre-miRNA / 1-2 mismatches in miRNA*; '' (*medium orange*) = up to six mismatches in Pre-miRNA; ''' (*dark orange*) = more than six mismatches in Pre-miRNA or unknown to PaxiN

[5] **Presences of miRNA*s are given in Table 2B**. Here, only the frequencies of *P. inflata* specific miRNA*s are given

[6] **See at the end of Table 2B**. In *P. inflata*, the same *MIR* clusters as in *P. axilaris* were found, but often within different distances, namely, I. 110.000 nt; II. 20.000 nt; IV. 60.000 nt; X. 10.000 nt; XI. 200.000 nt; XIII. 35.000 nt

[7] **miRNA398a** is not found in *P. inflata*, but a miRNA*398a sequence identical to *P. axillaris* occurs, possibly indicating that this is the miRNA instead

[8] **See Lower part of Table 2B** for inormation about miRNA variants

Table 2D. Sequences of mature miRNAs in *Petunia*

*Petunia*_(*Pax*,*Pin*)_miR-0156a-g	UGACAGAAGAGAGUGAGCAC
*Petunia*_(*Pax*,*Pin*)_miR-0156j-l	UGACAGAAGAGAGAGAGCAC
*Petunia*_(*Pax*,*Pin*)_miR-0157a-e	UGACAGAAGAUAGAGAGCAC
*Petunia*_(*Pax*,*Pin*)_miR-0159a-c	UUUGGAUUGAAGGGAGCUCUA
*Petunia*_(*Pax*,*Pin*)_miR-0160a-d	UGCCUGGCUCCCUGUAUGCCA
*Petunia*_(*Pax*,*Pin*)_miR-0162a	UCGAUAAACCUCUGCAUCCAG
*Petunia*_(*Pax*,*Pin*)_miR-0164a-d	UGGAGAAGCAGGGCACGUGCA
*Petunia*_(*Pax*,*Pin*)_miR-0164g,h	UGGAGAAGCAGGGCACAUGCU
*Petunia*_(*Pax*,*Pin*)_miR-0166a-h	UCGGACCAGGCUUCAUUCCCC
*Petunia*_(*Pax*,*Pin*)_miR-0166k	UCGGACCAGGCUUCAUUCCUC
*Petunia*_(*Pax*,*Pin*)_miR-0167a-e	UGAAGCUGCCAGCAUGAUCUA
*Petunia*_(*Pax*,*Pin*)_miR-0168a	UCGCUUGGUGCAGGUCGGGAA
*Petunia*_(*Pax*,*Pin*)_miR-0168b,c	UCGCUUGGUGCAGGUCGGGAC
*Petunia*_(*Pax*,*Pin*)_miR-0169a,b	CAGCCAAGGAUGACUUGCCGA
*Petunia*_(*Pax*,*Pin*)_miR-0169c,d	CAGCCAAGGAUGACUUGCCGG
*Petunia*_(*Pax*,*Pin*)_miR-0169g-x	UAGCCAAGGAUGACUUGCCU(A)
*Petunia*_(*Pax*,*Pin*)_miR-0171a-g	UGAUUGAGCCGUGCCAAUAUC
*Petunia*_(*Pax*,*Pin*)_miR-0171j,k	UUGAGCCGCGCCAAUAUCACU
*Petunia*_(*Pax*,*Pin*)_miR-0171n,o	UUGAGCCGCGUCAAUAUCUCU
*Petunia*_(*Pax*,*Pin*)_miR-0172a-g	AGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pax*,*Pin*)_miR-0172j,k	GGAAUCUUGAUGAUGCUGCAG
*Petunia*_(*Pax*,*Pin*)_miR-0319a-e	UUGGACUGAAGGGAGCUCCCU
*Petunia*_(*Pax*,*Pin*)_miR-0319h,i	UUGGACUGAAGGGAGCUCCU
*Petunia*_(*Pax*,*Pin*)_miR-0390a-c	AAGCUCAGGAGGGAUAGCGCC
*Petunia*_(*Pax*,*Pin*)_miR-0393a-c	UCCAAAGGGAUCGCAUUGAUCC
*Petunia*_(*Pax*,*Pin*)_miR-0394a-c	UUGGCAUUCUGUCCACCUCC
*Petunia*_(*Pax*,*Pin*)_miR-0395a-m	CUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pax*,*Pin*)_miR-0396a,b	UUCCACAGCUUUCUUGAACUG
*Petunia*_(*Pax*,*Pin*)_miR-0396c,d	UUCCACAGCUUUCUUGAACUU
*Petunia*_(*Pax*,*Pin*)_miR-0397a,b	AUUGAGUGCAGCGUUGAUGA
*Petunia*_(*Pax*,*Pin*)_miR-0398a,b	UGUGUUCUCAGGUCACCCCUU
*Petunia*_(*Pax*,*Pin*)_miR-0398e	UAUGUUCUCAGGUCGCCCCUG
*Petunia*_(*Pax*,*Pin*)_miR-0398h	CAGGGGCGACCUGAGAACACA
*Petunia*_(*Pax*,*Pin*)_miR-0399a-c	UGCCAAAGGAGAGUUGCCCUG
*Petunia*_(*Pax*,*Pin*)_miR-0399c-g	GGGCUACUCUCUAUUGGCAUG
*Petunia*_(*Pax*,*Pin*)_miR-0403	UUAGAUUCACGCACAAACUCG
*Petunia*_(*Pax*,*Pin*)_miR-0408	UGCACUGCCUCUUCCCUGGCU
*Petunia*_(*Pax*,*Pin*)_miR-0477	CCUCUCCCUCAAGGGCUUCUC

*Petunia*_(*Pax*,*Pin*)_miR-0479	UGAGCCGAACCAAUAUCACUC
*Petunia*_(*Pax*,*Pin*)_miR-0482	UUUCCAAUUCCACCCAUUCCUA
*Petunia*_(*Pax*,*Pin*)_miR-0827	UUAGAUGAACAUCAACAAACA
*Petunia*_(*Pax*,*Pin*)_miR-2111a,b	UAAUCUGCAUCCUGAGGUUUA
*Petunia*_(*Pax*,*Pin*)_miR-6149	UUGAUACGCACCUGAAUCGGC
*Petunia*_(*Pax*,*Pin*)_miR-8016	AUUUUUGAAUGGAAGGCCCAUGUG
*Petunia*_(*Pax*,*Pin*)_miR-6164	CCUCCGUUUCAAUUUAUGUGA
*Petunia*_(*Pax*,*Pin*)_miR-0157a-e(v2)	UUGACAGAAGAUAGAGAGCAC
*Petunia*_(*Pax*,*Pin*)_miR-0171a-g(v2)	UUGAGCCGUGCCAAUAUCACG
*Petunia*_(*Pax*,*Pin*)_miR-0171j,k(v2)	UUGAGCCGCGCCAAUAUCAC-
*Petunia*_(*Pax*,*Pin*)_miR-0171j,k(v3)	UGAUUGAGCCGCGCCAAUAUC
*Petunia*_(*Pax*,*Pin*)_miR-0172a-g(v2)	AGAAUCUUGAUGAUGCUGC--
*Petunia*_(*Pax*,*Pin*)_miR-0172j,k(v2)	GGAAUCUUGAUGAUGCUG---
*Petunia*_(*Pax*,*Pin*)_miR-0319a-e,h,i(v2)	CUUGGACUGAAGGGAGCUCC
*Petunia*_(*Pax*,*Pin*)_miR-0319h,i(v3)	UUGGACUGAAGGGAGCUCCUU
*Petunia*_(*Pax*,*Pin*)_miR-0390a-c(v2)	AAGCUCAGGAGGGAUAGCGC
*Petunia*_(*Pax*,*Pin*)_miR-0397a,b(v2)	UCAUUGAGUGCAGCGUUGAUG
*Petunia*_(*Pax*,*Pin*)_miR-0398a,b(v2)	UUGUGUUCUCAGGUCACCCCU
*Petunia*_(*Pax*,*Pin*)_miR-0398h(v2)	UGUGUUCUCAGGUCGCCCCUG
*Petunia*_(*Pax*,*Pin*)_miR-2111a,b(v2)	UAAACCUCAGGAUGCAGAUUA

Table 2E. Sequences of pre-miRNAs in *Petunia*

*Petunia*_(*Pax*,*Pin*)_*MIR*-0156a	UGACAGAAGAGAGUGAGCACACAUGGUGUUUUCUUUCAUGAUAUAUUCAUGCUUGAAGCUAUGA GUGCUCACCCUCUUUCUGUCACC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0156b	UGACAGAAGAGAGUGAGCACACAUGGUGUUUUCUUGCAUGAUUUAUACUCUAUGCUUGAAGCUA UGCGUGCUCACUCUCUAUCUGUCACC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0156c	UGACAGAAGAGAGUGAGCACACAUGGUAUUUUCUUGCAUAAUGUUUAUCCUAUGCUUGAAGCUA UGCGUGCUCACUCUCUAUCUGUCACC
*Petunia*_(*Pax*)_*MIR*-0156d	UGACAGAAGAGAGUGAGCACAUAUGGUGUUUUCUUGCAUAAUAUAUUUGUAUUAUUUAUGCUUG AAGCUAUGCGUGCUCACUCUCUAUCUGUCACC
*Petunia*_(*Pin*)_*MIR*-0156d	UGACAGAAGAGAGUGAGCACAUAUGGUGUUUUCUUGCAUAAUAUAUUUGUAUUAUUUAUGCUUG AAGUUAUGCGUGCUCACUCUCUAUCUGUCACC
*Petunia*_(*Pax*)_*MIR*-0156e	UGACAGAAGAGAGUGAGCACACAUGAUAUUUUCUUGCAUCUAUGCAUAUGCUUGAAGUUAUGUG UGCUCACGCUCUAUCUGUCACC
*Petunia*_(*Pin*)_*MIR*-0156e	UGACAGAAGAGAGUGAGCACACAUGAUAUUUUCUUGCAUCAAUGCAUAUGCUUGAAGUUAUGUG UGCUCACGCUCUAUCUGUCACC
*Petunia*_(*Pax*)_*MIR*-0156f	UGACAGAAGAGAGUGAGCACACGCAGACAAAUGUAUAGACUGUUUAUGCCUUUGUGAUAUUUGC GUGUGCUCACUUCUCAUUCUGUCACC
*Petunia*_(*Pin*)_*MIR*-0156f	UGACAGAAGAGAGUGAGCACACGCAGACAAUUGUAUAGACUGUUUAUGCCUUUGUGAUAUUUGC GUGUGCUCACUUCUCUUUCUGUCACC
*Petunia*_(*Pax*)_*MIR*-0156g	UGACAGAAGAGAGUGAGCACACGCAGGCAGUUGUAUAGAGUGUAUACCUUUGUUAUUUGCGUGU GCUCACUUCUCUUUCUGUCAGC
*Petunia*_(*Pin*)_*MIR*-0156g	UGACAGAAGAGAGUGAGCACACGCAGGCAUUUGUAUAGAGUGUAUACUUUUGUUAUUUGCGUGU GCUCACUUCUCUUUCUGUCAGC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0156j	UGACAGAAGAGAGAGAGCACGUCCCUCAUUUACUGGUAAGUCCCGAAACUUACCUGUUAAUAUC GGCUUAUCCUCUGUUAGA
*Petunia*_(*Pax*,*Pin*)_*MIR*-0156k	UGACAGAAGAGAGAGAGCACAGAUGAUGAAGUGCAUGGAAACUUUUUGCACCUCACUCCUUUGU GCUCUCUAUUCUUCUGUCAUC
*Petunia*_(*Pax*)_*MIR*-0156l	UGACAGAAGAGAGAGAGCACAAUCUGCCUUCAGCAAAAUGAUGUUGCUGCUUGUUGGAAUGUGC UCUCUCUGCUUCUGUCAAU
*Petunia*_(*Pin*)_*MIR*-0156l	UGACAGAAGAGAGAGAGCACAAUCUGCCAUCAGCAAAAUGAUGCUGCUUGUUGGAAUGUGCUCU CUCUGCUUCUGUCAAU
*Petunia*_(*Pax*,*Pin*)_*MIR*-0157a	UGACAGAAGAUAGAGAGCACUGAUGAUGAUUUGCUAAAGUAGCAUCUCAAUUCAUUUGUGCUCU CUAUGCUUCUGUCAUC
*Petunia*_(*Pax*)_*MIR*-0157b	UGACAGAAGAUAGAGAGCACAGAUGAUGAAGUACAUGGAAACUUCUGUACCUCACUCCUUUGUG CUCUUUAUUUUUCUGUCAUC
*Petunia*_(*Pin*)_*MIR*-0157b	UGACAGAAGAUAGAGAGCACAGAUGAUGAAGUGCAUGGAAACUUCUGUACCUCACUCUUUUGUG CUCUUUAUUCUUCUGUCAUC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0157c	UGACAGAAGAUAGAGAGCACAGAUGAUGAAGUGCACGGAAGCUUUAUGCACCUCACUCCUUUGU GCUCUUUAUCCUUCUGUCAUC
*Petunia*_(*Pax*)_*MIR*-0157d	UGACAGAAGAUAGAGAGCACACAUGAUGAAAAUGCUAAAUUUGGAAGGCACAAAGCAUCUUAAUU CAUGUGUGCUCUCUAUGCUUCCGUCAUC
*Petunia*_(*Pin*)_*MIR*-0157d	UGACAGAAGAUAGAGAGCACAGAUGAUGAAAUGCUAAAAGGCACUGAAUAAACUGCAAAAGCAU CUUAAUUCAUUUGUGCUCUCUAUGCUUCCGUCAUC
*Petunia*_(*Pax*)_*MIR*-0157e	UGACAGAAGAUAGAGAGCACAGAUGAUGAGAUGUUUAAUUGGAAGCUAUCUGCAUCUCACUCCU UUGUGCUCUCUAUUCUUCUGCCAUC
*Petunia*_(*Pin*)_*MIR*-0157e	UGACAGAAGAUAGAGAGCACAGAUGAUGAGAUGUUUAAUUGGAAGCUAGCCGCAUCUCACUCCU UUGUGCUCUCUAUUCUUCUGCCAUC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0159a	GAGCUCCUUGAAGUCCAACAGAGGAUCUAACAGGUAAGAUUGAGCUGCUGACCUAUGGAUCCCU CAGCCCUAUCUGUUGCUAUUUCAAGAAUGGUAGGUUUGUGGGUUGCAUAUGUCAGGAGCUUCAU UACCCUAUGUUGGAUCCCUUUUUGGAUUGAAGGGAGCUCUA
*Petunia*_(*Pax*)_*MIR*-0159b	GAGCUUCUUUAAGUCCAACAGAGGAUCUAACUGGUAAAAUUAAGCUGCUGAUAUAUGGAUCCCU CAGCCCUAUAUAAUUUAUACUUUGUUGAAAUAAUAGGGUUGUGGCUUGCAUAUACCAGGAGCUU UAUUACCCUAUGUUUGAUCCCUUUUUGGAUUGAAGGGAGCUCUA
*Petunia*_(*Pin*)_*MIR*-0159b	GAGCUUCUUUAAGUCCAACAGAGGAUCUAACUGGUAAAAUUAAGCUGCUCAUAUAUGGAUCCCU CAGCCCUAUAUAAUUUAUACUUUCUUGAAAUAAUAGGGUUGUGGCUUGCAUAUACCAGGAGCUU UAUUAUCCUAUGUUUGAUCCCUUUUUGGAUUGAAGGGAGCUCUA

*Petunia*_(*Pax*)_*MIR*-0159c	GAGCUUCUUUGAAGUCCAAAAGAGGAUCUAAGUGGGAAAAUUAAGCUGCUGAUCUAUGGAUUCC UCAGUCCUCUCUAUUCAUAAUUCAUUGUUUGGGUAGGUUUGUGGUUUGCAUAUAUCAGGAGCUG CAUUUACCCUUACUGAGAUCCUCGUUUGGAUUGAAGGGAGCUCUA
*Petunia*_(*Pin*)_*MIR*-0159c	GAGCUUCUUUGAAGUCCAAAAGAGGAUCUAAGUGGGAAGAUUGAGCUGCUGAUCUAUGGAUCCC UCAGUCCUAUCUAUUCUUAAUUCAUUGUUUUGGAUAGGUUUGUGGUUUGCAUAUAUCAGGAGCU GCAUUUACCCUUACUAGGAUCCUUGUUUGGAUUGAAGGGAGCUCUA
*Petunia*_(*Pax*,*Pin*)_*MIR*-0160a	UGCCUGGCUCCCUGUAUGCCAUUUGCAAAGCUCAUCGUAAUAUGCCAUGGGUCUUCGUGAAUGG CGUAUGAGGAGCCAAGCAUA
*Petunia*_(*Pax*)_*MIR*-0160b	UGCCUGGCUCCCUGUAUGCCAUUUGCAGAGUUCAUCGGAACAUCGGUGGGUCUCCGUGAAUGGC GUAUGAGGAGCCAAGCAUA
*Petunia*_(*Pin*)_*MIR*-0160b	UGCCUGGCUCCCUGUAUGCCAUUUGCAGAGUUCACCGUAACAUCGGUGGGUCUCCGUGAAUGGC GUAUGAGGAGCCAAGCAUA
*Petunia*_(*Pax*,*Pin*)_*MIR*-0160c	UGCCUGGCUCCCUGUAUGCCACAUGCUUUCACCAAUCUUGUAUUCUUCGAUUGGCUGAUCAGUG GGUGGCGUGCGAGGAGCCAAGCAUA
*Petunia*_(*Pin*)_*MIR*-0160d	UGCCUGGCUCCCUGUAUGCCAUCUAAGAGCAGCACUGUCAAAAUUGACACACUUCCUAGUUGGC ACCAGAGGAGUCGGGCAGA
*Petunia*_(*Pax*)_*MIR*-0162a	GGAGGCAGCGGUUCAUCGAUCUGUUCCCUGAAAAGCUAUAAACUAAAUAUAGCAAACAGGAAUC GGUCGAUAAACCUCUGCAUCCAG
*Petunia*_(*Pin*)_*MIR*-0162a	GGAGGCAGCGGUUCAUCGAUCUGUUCCCUGAAAAGCUAUAAACAAAAUAUAGCAAACAGGAAUC GGUCGAUAAACCUCUGCAUCCAG
*Petunia*_(*Pax*,*Pin*)_*MIR*-0164a	UGGAGAAGCAGGGCACGUGCAAGUUCUUCAUUUGAACAUAUAUGCAUCGGUCAUGCAUGAAAAU UGCACGUGUUCUCCUUCUCCAAC
*Petunia*_(*Pax*)_*MIR*-0164b	UGGAGAAGCAGGGCACGUGCAUUACUAACUCAUGCACAAGAAUUGUCGGUCAGUUAGUAAAAGU UUUAUGGCAUUCACUUGCUAUGAUAGUAGUAGUUGUACCAUUAAUGGAAAAUGAGUUAGUUCUU CAUGUGCCUGUCUUCCCCAUC
*Petunia*_(*Pin*)_*MIR*-0164b	UGGAGAAGCAGGGCACGUGCAUUACUAACUCAUGCACAAAGAAUUGUCGGUCAGUUAAUAUAAG UUUUUAUGGCAUUCACUUGCUAUCAUAGUAGUAGUUGUACCCAAUUAAUGGGAAAUGAGUUAGU UCUUCAUGUGCCUGUCUUCCCCAUC
*Petunia*_(*Pax*)_*MIR*-0164c	UGGAGAAGCAGGGCACGUGCAUUACUAACUCAUCUCACAAGUCACAAAGUGUUAAAUAUGGUAU UAGUCGGCCAUUAAUUCACCUUAUGGGAGCUGCCAUGCAUAUUGGAGUAUCAAACAAACCUCCU AUGUAUUGUUUCUGUACGUAUGAUAAAUGAGUUAGUUCUUCAUGUGCCCCUCUUCCCCCAUC
*Petunia*_(*Pin*)_*MIR*-0164c	UGGAGAAGCAGGGCACGUGCAUUACUAACUCACAUCUCACAAGUCACAAAGUAUUAAAUAUGGU AUUAGUCGGCCAUUAAUUCACUUUAUGGGAGCUGCCAUGCAUAUUGGAGUAUCACACAAACCUC CAAUGUGUUGUUUCUGUAUGAUAAAUGAGUUAGUUCUUCAUGUGCCCCUCUUCCCCAUC
*Petunia*_(*Pax*)_*MIR*-0164d	UGGAGAAGCAGGGCACGUGCAAGUUCUCUACGUGAAUUGUGCAUGUGUUCUCCUUCUCCAAC
*Petunia*_(*Pin*)_*MIR*-0164d	UGGAGAAGCAGGGCACGUGCAAGUUGUCUGCAUGCGUGCCCGUUUUACUGUCAUUCAUGAAUGA AUUGUGCAUGUGUUCUCCUUCUCCAAC
*Petunia*_(*Pax*)_*MIR*-0164g	UGGAGAAGCAGGGCACAUGCUGGAUUACUUGGACUUGUUAUCCUCUCAGAAUUUAGCAUGUGCU CUUGCCCUCCAGC
*Petunia*_(*Pin*)_*MIR*-0164g	UGGAGAAGCAGGGCACAUGCUGGAUUACUUGGACUUGCUUUCCUCUGAGAAUUUAGCAUGUGCU CUUGCCCUCCAGC
*Petunia*_(*Pin*)_*MIR*-0164g2	UGGAGAAGCAGGGCACAUGCUGGAUUACUUGGACUUGCUAGCCUCUGAGAAUUUAGCAUGUGCU CUUGCCCUCCAGC
*Petunia*_(*Pax*)_*MIR*-0164h	UGGAGAAGCAGGGCACAUGCUAAAUUACUUGGAACUCUGAUCUUCACAAGAAUUUUGCAUGUGC UCUUGCUCUCCAGC
*Petunia*_(*Pin*)_*MIR*-0164h	UGGAGAAGCAGGGCACAUGCUGAAUUACUUGGAACUCUGAUCUUCCAAGAAUUUUGCAUGUGCU CUUGCUCUCCAGC
*Petunia*_(*Pax*)_*MIR*-0166a	GGAAUGUUGUCUGGCUCGAGGUCACCAACUAGAUCCAUGACCUUCAUUAAGUAACAUAUACAUA UACAUUAGUUGAUUAUAUGAAAAGAGUGGUCAUGGAUCAUGUGUUUGUGUUGUCGGACCAGGCU UCAUUCCCC
*Petunia*_(*Pin*)_*MIR*-0166a	GGAAUGUUGUCUGGCUCGAGGUCACCAACUAGAUCCAUGACCUUCAUUAAGUAACAUAUACAUA UACAUUAGUUGAUUAUAUGAAAAGAGUGGUUAUGGAUCAUGUGUUUGUGUUGUCGGACCAGGCU UCAUUCCCC
*Petunia*_(*Pax*)_*MIR*-0166b	GGAAUGUUGUCUGGCUCGAGGAUCUUCAAUCUCGAUCAAAUUCUUAUACAAAUCUAUUAUACAU AUGUUUUACAAAAUCCAUUUCUAUGUAGAUCCAUUAAUGUUUUUAAUAGAUCUAUAAGUUUUGA UCAUGUUUUUGGUGUCGUCGGACCAGGCUUCAUUCCCC
*Petunia*_(*Pin*)_*MIR*-0166b	GGAAUGUUGUCUGGCUCGAGGAUCUUCAAUCUUGAUCAAAUUCUUAUACAAAUCUAUUAUACAU

	AUUUUUUAUAAAAUCCAUUUCUAUGUAGAUCCAUUAAUGUUUUUAAUAGAUCUAUAAGUUUUGA UCAUGUUUUUGGUGUCGUCGGACCAGGCUUCAUUCCCC
*Petunia*_(*Pax*)_*MIR*-0166c	AGAAUGUCGUCUGGUUCGAGAUCUUUCAUGACCCGAAAAUUGACGAGGUAGAGAAGUUUUCAUC AUCUGAAUGAUUUCGGACCAGGCUUCAUUCCCC
*Petunia*_(*Pin*)_*MIR*-0166c	AGAAUGUCGUCUGGUUCGAGAUCUUUCAUGAACUCGAAAAUUGAGGAGGUAGAGAAGUUGUCAU CAUCUGAAUGAUUUCGGACCAGGCUUCAUUCCCC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0166d	GGAAUGUUACCUGGCUCGAAGUCAUUAUCUCUUCUUGAUUAACUUAACUACUGCUUCUUCUCAC UUCACAGUAUUUUGACAAGUAUCCGCACUUGAGCUCGCUGUCUUAAAAUUCAGAAAUCAAGAGA UAGUGUCGUCGGACCAGGCUUCAUUCCCC
*Petunia*_(*Pax*)_*MIR*-0166e	GGAAUGCUGUCUGGUUCGAAACCAUUCAAGAAAUCAACUGUCAAUUAUCAAAGUUAUGAUGAUU UGAAUGAUUUCGGACCAGGCUUCAUUCCCC
*Petunia*_(*Pin*)_*MIR*-0166e	GGAAUGCUGUCUGGUUCGAAACCAGUCAAGAAAUCAACUGCCAAUUAUCAAAGUUAUGAUGAUU UGAAUGAUUUCGGACCAGGCUUCAUUCCCC
*Petunia*_(*Pax*)_*MIR*-0166f	GGAAUGCUGUCUGGUUCGAAACCAUUCUAUAAAGUGGCUGAUCAAGGUAAAGGAUAGUCACUAG GGGAUCCAACAGAAUCAAAUGUUUCGAUGUCUAACUGCAUAUAUAUACAAAAAAAAUAUGUAUA UAUAGAUAGAGUUAGAUAUACGUUACGUUGUCAGAACCCACAACGUCUAAAAUAAUCUGGAUCU GCCACACUUGAAUCAGUUGAUCAUUUGAUAUGAGUGAUUUCGGACCAGGCUUCAUUCCCC
*Petunia*_(*Pin*)_*MIR*-0166f	GGAAUGCUGUCUGGUUCGAAACCAUUCUAUAAAGUGGCUGAUCAAGGUAAAGGAGUCACUAGGG GAUCCAACAAUAAGUUCAUGGAUUCAAUAGAAUCAAAUGUUUCCAUGUCUAACUGCAUAUAUAU ACUAUACAAAAAAGAUGUAUAUAUAGAUAGAGUUAGAUAUACGUUACGUUGUCAGAACCCACAA AGUCUAAAAUAAUCUGGAUCUGCCACACUUGAAUCAGUUGAUCAUUUGAUAUGAGUGAUUUCGG ACCAGGCUUCAUUCCCC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0166g	GGAGUGUUGCCUGGUUCGAAGACACCAUCUUUUUUGUUAUCUAUUAUGAGAUAAUGUCUUCGGA CCAGGCUUCAUUCCCC
*Petunia*_(*Pax*)_*MIR*-0166h	GGAAUGUCGUCUGGUUCAAGAUCUUUCAUGUCAUGUGAAAGAAAAAAAAUGAAGAUCACAUGAA UGAUUUCGGACCAGGCUUCAUUCCCC
*Petunia*_(*Pin*)_*MIR*-0166h	AAAAUGUCGUCUGGUUCAAGAUCUUUCAUGUCAGGUGAAAGAAAAAAAAAUGAAGAUCCAGAUA UCAUCACAUGAAUGAUUUCGGACCAGGCUUCAUUCCCC
*Petunia*_(*Pax*)_*MIR*-0166k	GGAAUGUUGGCUGGCUCGACACAAUCACUCUCCUAUAUAUGGUCUUGUUUGUGGUUUUGUUUUU GUUUUUCAUGAUGGUGAUGAAAAACAAAACAAGAUUCACAUGCUUAGAUGAAAAUAAUGGGAGU GAAGGAGUCGGACCAGGCUUCAUUCCUC
*Petunia*_(*Pin*)_*MIR*-0166k	GGAAUGUUGGCUGGCUCGACACAAUCACUCUCCUUUAUAUACAGUCUUGUUUGUGGUUUUGUCU UUGUUUUUUUCACUAUGGUGAUGAAAAACAGAACAAGAUUCACAUAUUUAGAUGAAAAUAAUGG AAGUGAAGGAGUCGGACCAGGCUUCAUUCCUC
*Petunia*_(*Pax*)_*MIR*-0167a	UGAAGCUGCCAGCAUGAUCUAAACUUCCCUUUUGUUUCAUUGAGGAAAGAUCAGAUCAUGUGGC AGCCUCACC
*Petunia*_(*Pin*)_*MIR*-0167a	UGAAGCUGCCAGCAUGAUCUAAACUUCCCUUUUGUUUAAUUUCACUGAGGAAAGAUCAGAUCAU GUGGCAGCCUCACC
*Petunia*_(*Pax*)_*MIR*-0167b	UGAAGCUGCCAGCAUGAUCUAAACUUUUCUACUUCUCCGUGGAUCGUGGAAAGAUCAGAUCAUG UGGCAGCAUCACC
*Petunia*_(*Pin*)_*MIR*-0167b	UGAAGCUGCCAGCAUGAUCUAAACUUUUCUACUUCUCCAUGGAUCGUGGAAAGAUCAGAUCAUG UGGCAGCAUCACC
*Petunia*_(*Pax*)_*MIR*-0167c	UGAAGCUGCCAGCAUGAUCUAAUCUUGGCCAGAUAAAUACAUAUGGAAAGAAACAAGUGUUGCU GACUACGGUUAGGUCAUGCCCUGACAGCCUCA
*Petunia*_(*Pin*)_*MIR*-0167c	UGAAGCUGCCAGCAUGAUCUAAUCUUGGCCAGAUAAAUACAUAUGGAAAGAAACAACAAGUAUU GUUGACUACGGUUAGGUCAUGCCCUGACAGCCUCACU
*Petunia*_(*Pax*)_*MIR*-0167d	UGAAGCUGCCAGCAUGAUCUAAACUUACCUCUUUUUAAUUAUGUUUGAGGGAAGAUCAGAUCAU GUGGUAGCUUCACC
*Petunia*_(*Pin*)_*MIR*-0167d	UGAAGCUGCCAGCAUGAUCUAAACUUACCUAUGUUUGAGGGAAGAUCAGAUCAUGUGGUAGCUU CACC
*Petunia*_(*Pax*)_*MIR*-0167e	UGAAGCUGCCAGCAUGAUCUAACUUUGGCCAAAAUAAAGCUUCAACGGUUGAAGAUAAUAACUA GAUUUUGACCAUGGUUAGGUCAUGCUCCGACAGCUUCA
*Petunia*_(*Pin*)_*MIR*-0167e	UGAAGCUGCCAGCAUGAUCUAACUUUGGCCAAAAUAAAGCUUCAACGGAUGAAGAUAAUAACCA AAUUUUGACCAUGGUUAGGUCAUGUUCCGACAGCUUCACU
*Petunia*_(*Pax*)_*MIR*-0168a	UCGCUUGGUGCAGGUCGGGAACUGACUCGCUCCGUGACAUAAUAGAAAGGUGUUAUAUUUAGUC UAAAGCGAACGUCUGUUCCCGCCUUGCAUCAACUGAAU
*Petunia*_(*Pin*)_*MIR*-0168a	UCGCUUGGUGCAGGUCGGGAACUGACUCGCUCUGUGACAUAAUAAAACAAGUGUGUUACAUAUU UAGUCUAAAGCGAACGUCUGUUCCCGCCUUGCAUCAACUGAAU

*Petunia*_(*Pax*,*Pin*)_*MIR*-0168b	UCGCUUGGUGCAGGUCGGGACCUGCUUCGCCAGAUACGACGGCGUUUCGAUGAGGGUCCCGCCU UGCAUCAACUGAAU
*Petunia*_(*Pax*)_*MIR*-0168c	UCGCUUGGUGCAGGUCGGGACUUAUCACCCUUUAUUUAUCCUAUAUAUAUAUAUAUAUAUAUAU AUUCACAAUAUAUAUAUGUUUGAAUAUCGCCAAGUGUAAGUCCCGCCUUGCAUCAACUGAAU
*Petunia*_(*Pin*)_*MIR*-0168c	UCGCUUGGUGCAGGUCGGGACUUAAUUCAUCAAUAUAUAUUUUUUUCAAAUGAAUAUAUAUGUUU GAAUAUCACGAAGUAUAAGUCCCGCCUUGCAUCAACUGAAU
*Petunia*_(*Pax*)_*MIR*-0169a	CAGCCAAGGAUGACUUGCCGACUAUUGCAUUAACGAUAAAGAGUAUGGUUAUACAUUCAUGCGA UGGAACUCGGAAGAGAAAAGUUGGCAAGUGGUCCUUGGCUACA
*Petunia*_(*Pin*)_*MIR*-0169a	CAGCCAAGGAUGACUUGCCGACUACUGCAUUAACGGUAAACAGUAUGGUUAUAUAUUCAUGCGA UGGAACUCGGAAGAGAAAAGUUGGCAAGUGGUCCUUGGCUACA
*Petunia*_(*Pax*)_*MIR*-0169b	CAGCCAAGGAUGACUUGCCGAGAUAUUGUAGCAAAGCAAUGAUCUAUCGGCAAGUCGUCCUUGG CUACA
*Petunia*_(*Pin*)_*MIR*-0169b	CAGCCAAGGAUGACUUGCCGAGAUAUUGAAGCAAAGCAAUGAUCUAUCGGCAGGUCGUCCUUGG CUACA
*Petunia*_(*Pax*,*Pin*)_*MIR*-0169c(BL)	CAGCCAAGGAUGACUUGCCGGCAAGCUUGAUCAACUCUACUUUUUGAGUACUUUUUAUUAUAGU UAAAAACUAAAAAGUGGAGAUUAAUAACUUUGACUGGCAAGUCAUUUUUGGCUACA
*Petunia*_(*Pax*)_*MIR*-0169d	CAGCCAAGGAUGACUUGCCGGCAAUUUCAAAGCAAAAUUGCAGAGGUUAUUCGUAUUCCGGCAA GUUGUUCCUGGCUACA
*Petunia*_(*Pin*)_*MIR*-0169d	CAGCCAAGGAUGACUUGCCGGCAAUUUCAAAGCAAAAUUGCAGAGAUUAUUCGUAUUCCGGCAA GUUGUUCUUGGCUACA
*Petunia*_(*Pax*)_*MIR*-0169g	UAGCCAAGGAUGACUUGCCUGCUUCGAGGCUAGGUGUUGAUAUGUGAGUAUCAGCAUUCUUGUU UGAUUGCCAGGCAGUCUCCUUGGCUACU
*Petunia*_(*Pin*)_*MIR*-0169g	UAGCCAAGGAUGACUUGCCUGAUUCGAGGCUAGGUGUUGAUACGUGAGUAUCAGCAUUCUUGUU UGAUUGCCAGGCAGUCUCCUUGGCUACU
*Petunia*_(*Pax*)_*MIR*-0169h	UAGCCAAGGAUGACUUGCCUGAAUCAUUCUUGAGGCUUGUAGUGCUGACAGCCUUAAUUUUGAU CACCAGGCAGUCUCUCCUUUGGCUUUC
*Petunia*_(*Pin*)_*MIR*-0169h	UAGCCAAGGAUGACUUGCCUGAGUCAUUCUUGUGGCUUGUAAUUGCUGACAGCCUUAAUUUUGA UCACCAGGCAGUCUCUCCUUUGGCUUUC
*Petunia*_(*Pin*)_*MIR*-0169h2	UAGCCAAGGAUGACUUGCCUGAGUCAUUCUUGAGGCUUGUAAUGUGCUGACAGCCUUAUUUUUU AUCACCAGGCAGUCUCUCCUUUGGCUUUC
*Petunia*_(*Pax*)_*MIR*-0169i	UAGCCAAGGAUGACUUGCCUGCUUCAUUCUCGAGGCUGAUUGUUAUGUCUUCAGCAUUCUUGUU UGAUCACCAGGCAGUCUCCUUGGCUAAU
*Petunia*_(*Pin*)_*MIR*-0169i	UAGCCAAGGAUGACUUGCCUGCUUCAUUCUUGAGGAUGAUUGUUAUGUCUUCAGCAUUCUUGUU UGAUCACCAGGCAGUCUCCUUGGCUAAU
*Petunia*_(*Pax*)_*MIR*-0169j	UAGCCAAGGAUGACUUGCCUGUUCCAUUAUUCUUUAAAAGGUUUAUAUAUAAACCAUCAAUGGU GGCCUUUUUUACUUGGUUUCAGGCAGUCUCCUUGACUACC
*Petunia*_(*Pin*)_*MIR*-0169j	UAGCCAAGGAUGACUUGCCUGCUCCAUUAUUAUUUGAAAGGUUUAUAUAUAAACCAUCAAUGGU GGCCUUUUUUAUUUGGUUUCAGGCAGUCUCCUUGACUACC
*Petunia*_(*Pax*)_*MIR*-0169k	UAGCCAAGGAUGACUUGCCUGCACCACUUGGAAUUUACACCAUAUCGGUGGUGUCCUUUCUUUC UUGGUUUCAGGCAUGUCAUCUUGGCUAGC
*Petunia*_(*Pin*)_*MIR*-0169k	UAGCCAAGGAUGACUUGCCUGCACCACUUGGAAAGGUCUUCUACACCAUUUUGAUGGUGUCCUU UCUUUCUUGGUUUCAGGCAUGUCAUCUUGGCUAAC
*Petunia*_(*Pax*)_*MIR*-0169l	UAGCCAAGGAUGACUUGCCUGCUCCAUUAUUAUUUAAAAGGUAUAUACACCAUCAAUGGUGGCC UUUUUUAUUGGUUUCAGGCAGUCUCCUUGGCUACC
*Petunia*_(*Pin*)_*MIR*-0169l	UAGCCAAGGAUGACUUGCCUGCUCCAUUAGUAUUUAAAAGGUAUAUACACCAUCAAUGGUGGCC UUUUUGAUUGGUUUCAGGCAGUCUCCUUGGCUACC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0169m	UAGCCAAGGAUGACUUGCCUACGUACACCACUUGGAAAGGUCUUCUCCUUUCUUUUUUGGUUUA AGGCAGGUCAUCUUUAGCUAAC
*Petunia*_(*Pax*)_*MIR*-0169n	UAGCCAAGGAUGACUUGCCUACUCCAUUUGACAGUUGAAACAACCAAAAAGGAUACUAUUAUAA GUAACGAUCCUUUUCAUUUGGUUGUAGGCAGGUCAUCCUAGCUAAC
*Petunia*_(*Pin*)_*MIR*-0169n	UAGCCAAGGAUGACUUGCCUACUCCAUUUGACAGUUGAAACAACCAAAAAGGAUGCUAUUACAA GGAACGAUCCUUUUUAUUUGGUUGUAGGCAGGUCAUCCUAGCUAAC
*Petunia*_(*Pax*)_*MIR*-0169o	UAGCCAAGGAUGACUUGCCUUUCAUCAAUGCCUUGCAAAAGGGGUGACAGAUAGUAUAAUAUUU UUUUACACUAUUGAGUUAAAUCCUUUUCAUGGUUGGAAUGAACGGCAAGCAUCUGAGGCGACU
*Petunia*_(*Pin*)_*MIR*-0169o	UAGCCAAGGAUGACUUGCCUUUCAUCAAUGCCUUGCAAAAGGGGUGACAGAUGGUAUAAUAUUU UUUUACACUAUGAGUUAAAUCCUUUUCAUGGUUGGAAUGAACGGCAAGCAUCUGAGGCGACU

*Petunia*_(*Pin*)_*MIR*-0169o2	UAGCCAAGGAUGACUUGCCUUUCAUCAAUGCCUUGCAAAAGGGGUGACAGAUGGUAUAAUAUUU UUUUACACAAUGAGUUAAAUCCUUUUCAUGGUUGGAAUGAACGGCAAGCAUCUGAGGCGACU
*Petunia*_(*Pax*)_*MIR*-0169p	UAGCCAAGGAUGACUUGCCUACAUAUAUAUACCAAACAAGCUGAAAAUUAAUCCUUGUUAUUGG UUGACAUGUUGGGCAGUCAUCCAUGGUUAUG
*Petunia*_(*Pax*)_*MIR*-0169p2	UAGCCAAGGAUGACUUGCCUACAUAUAUAUACUAUCAAACAAGCUGAAAAUUAACUCUUGUUAU UGGUUGACAUGUUAGGCAGUCAUCCAUGGUUAUG
*Petunia*_(*Pin*)_*MIR*-0169p	UAGCCAAGGAUGACUUGCCUACAUAUAUAUACUAUCAAACAAGCUGAAAAAUAACUCUUGUUAU UGGUUGACAUGUUAGGCAGUCAUCCAUGGUUAUG
*Petunia*_(*Pax*,*Pin*)_*MIR*-0169p2	UAGCCAAGGAUGACUUGCCUACAUAUAUAUACUAUCAAACAAGCUGAAAAUUAACUCUUGUUAU UGGUUGACAUGUUAGGCAGUCAUCCAUGGUUAUG
*Petunia*_(*Pax*)_*MIR*-0169q	UAGCCAAGGAUGACUUGCCUAAUGCACCACUUUCCUCCCCUAAAAGGGAAGUUUGUACGAGCUC AGAAACACUUCCCCCUUUUGUGGAGUAAGUUCAGUAGCUAGGCAGUCGUCUUUGGCUAUA
*Petunia*_(*Pax*)_*MIR*-0169q2	UAGCCAAGGAUGACUUGCCUAAUGUACCACUUUCCUCCCUAAAAGGGAAGUUCGUACGAGCUCA GAAGUAUUUCCCCUUUUGUAGAGUAAGUUCAGUAGCUAGGCAGUCGUCUUUGGCUAUA
*Petunia*_(*Pin*)_*MIR*-0169q	UAGCCAAGGAUGACUUGCCUAAUGCACCACUGUCCUCCCUAAAAGGGAAGUUUGUACGAGCUCA GAAACACUUCCCCUUUUGUGGAGUAAGUUCAGUAGUUAGGCAGUCGUCUUUGGCUAUA
*Petunia*_(*Pax*)_*MIR*-0169r	UAGCCAAGGAUGACUUGCCUAAUGCACUAUUUUUUCCUCCCUAAAAGGGGAUGUUGUACAAACA GUUCCCCUUCUGUGGAGCAAGUAUAAUAGCUAGGCAGUCGUCUUUGGCUACA
*Petunia*_(*Pin*)_*MIR*-0169r	UAGCCAAGGAUGACUUGCCUAAUGCACUAUUUUUUCCUCCCUAAAAGGGGAUGUUGUACAAACA GUCCCCCUUCUGUGGAGCAAGUAUAGUAGCUAGGCAGUCGUCUUUGGCUACA
*Petunia*_(*Pax*)_*MIR*-0169s	UAGCCAAGGAUGACUUGCCUAAACUCAUAAUUCAAGAUUUCAUAAACUUAAGCUAGUGAGCGAG CUGUUGUAUAUAUAAUAGUAUGUGUUUUGGGCGGUGAUCCGAGGUUACC
*Petunia*_(*Pax*)_*MIR*-0169t	UAGCCAAGGAUGACUUGCCUAAAACCUACUCAAUUUUGCAGGGUUUAUAUGAUCCUACAAAAUA GGGUUGUAGGCGUCAUCCCAGGCUAUU
*Petunia*_(*Pin*)_*MIR*-0169t	UAGCCAAGGAUGACUUGCCUAAAACCUACUCAAUUUUGCAGGGUUUAUAUGAUCCUACAAAAUA GGAUUGUAGGCGUCAUCCCAGGCUAUU
*Petunia*_(*Pax*)_*MIR*-0169u	UAGCCAAGGAUGACUUGCCUACAUACCCUUCUCCCAAAAGGGGUAGAAUUUUCAUCUUUAAAAC GAAACUUACCUUUUGGGGGUUGCAAUGCUAGGCAGGCAUCCUGGCUAUA
*Petunia*_(*Pin*)_*MIR*-0169u	UAGCCAAGGAUGACUUGCCUACAUACCCUCCUCCCAAAAGGGGUAGAAUUUUCAACUUUAAAAC GAAACUUACCUUUUGGGGUUGCAAUGCUAGGCAGGCAUCCUGGCUAUA
*Petunia*_(*Pax*,*Pin*)_*MIR*-0169v	UAGCCAAGGAUGACUUGCCUACUCCAUUUGACAGUUCCGAAAAAAUAAAAAAAUAACAAACUAA AGAAGGUUACAAUAAUGAUCCUUUUCAUUUGGUUGCAGGCGAGUUAUCCUGGCUAUA
*Petunia*_(*Pax*)_*MIR*-0169w	UAGCCAAGGAUGACUUGCCUAAAAAUCUCUUAUGUUGGAUUUGUCUACAAUCCUCUAAAGUGGU UUUUAGGUGUCAUCCUUGGAUAAC
*Petunia*_(*Pin*)_*MIR*-0169w	UAGCCAAGGAUGACUUGCCUAAAAAUCUCUUACGUGAGAUUUGAUCUACAAUCCUUUAAAGUGG UUUUUAGGUGUCAUCCAUGGGUAAC
*Petunia*_(*Pin*)_*MIR*-0169x	UAGCCAAGGAUGACUUGCCUACUCUACUUGACAGUUCAAAAGAAGAAGGAAAAAAAACUGGAGA AUGUGAUAAUAAUGUUCCAUUUCAGUUGGUUGCAGGCAAGUUAUCCUGGCUAUA
*Petunia*_(*Pax*)_*MIR*-0171a	UGUUGGUGCGGUUCAAUGAGAAAGUAUCGCUCAACAAGUAAAUUUGACCCUAUUUUUUGAUUGA GCCGUGCCAAUAUC
*Petunia*_(*Pin*)_*MIR*-0171a	UGUUGGUGCGGUUCAAUGAGAAAGUAUCGCUCAACAAGUAUAUUUGACCCUACUUUUUGAUUGA GCCGUGCCAAUAUC
*Petunia*_(*Pax*)_*MIR*-0171b	UAUUGGUGCGGUUCAAUGAGAAAGCAGUAAUCGAGAAGUUUUGACUCUACUUUUUGAUUGAGCC GUGCCAAUAUC
*Petunia*_(*Pin*)_*MIR*-0171b	UAUUGGUGCGGUUCAAUGAGAAAGCAGUACUCGAGAAGUUUUGACUCUACUUUUUGAUUGAGCC GUGCCAAUAUC
*Petunia*_(*Pax*)_*MIR*-0171c	UAUUGAUGCGGUUCAAUUAGAAAGCCGAAUUCUUUGUGUUUAGAAUCCUGUUAUUUGAUUGAGC CGUGCCAAUAUC
*Petunia*_(*Pin*)_*MIR*-0171c	UAUUGGUGCGGUUCAAUUAGAAAGCCGAAUUCUUUGUGUUUAGAACUCUGUUAUUUGAUUGAGC CGUGCCAAUAUC
*Petunia*_(*Pax*)_*MIR*-0171d	UAUUGGCCUGGUUCACUCAGACGACAAACGAAACUAUUUGAAUAAGUGGUGGAAUUUCAGUGUG AUUGAGCCGUGCCAAUAUC
*Petunia*_(*Pin*)_*MIR*-0171d	UAUUGGCCUGGUUCACUCAGACGACAAAUGAAACUAUUUGAAUGAGUGGUGGAAUUUCGGUGUG AUUGAGCCGUGCCAAUAUC
*Petunia*_(*Pax*)_*MIR*-0171e	UGUUGGAAUGGCUCAAUCAAAUCAAAUUCCCCAAAGUUUUUGGGUCAUUUAAUUUGAUUGAGCC GUGCCAAUAUC

*Petunia*_(*Pin*)_*MIR*-0171e	UGUUGGAAUGGCUCAAUCAAAUCAAAUUCCCCAAAGUUUUGGGUCAUUUAAUUUGAUUGAGCCG UGCCAAUAUC
*Petunia*_(*Pax*)_*MIR*-0171f	UAUUGGCCAGGUUCACUCAGACAGAAACACAACUUAUUCUUUUUUGAUGAAUAUGUUGGAGUUU AAUUUGCUUUGAUUGAGCCGUGCCAAUAUC
*Petunia*_(*Pin*)_*MIR*-0171f	UAUUGGCCAGGUUCACUCAAACAGAAACACAACUUUUUUCUUUUUUGAUGAAUAUGUUGGAGUU UAAUUUGCUUUGAUUGAGCCGUGCCAAUAUC
*Petunia*_(*Pax*)_*MIR*-0171g	UAUUGGUGAGGUUCAAUUAGAUUACUAUUGUCCACACAAGUUUGAACAAUUUAUAAUUUGAUUG AGCCGUGCCAAUAUC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0171j	CGAUGUUGGUGAGGUUCAAUCUGAAGACGGGUUUACGUUUUGUUUGCGUAAAGAACGAUCUCUG AUUGAGCCGCGCCAAUAUCACU
*Petunia*_(*Pax*,*Pin*)_*MIR*-0171k	CGAUGUUGGUGAGGUUCAAUCCGAAGACGGGUUUACGUUUUGUUUUCGUAAAGAACGAUCUCAG AUUGAGCCGCGCCAAUAUCACU
*Petunia*_(*Pax*,*Pin*)_*MIR*-0171n	AGAUGUUGAUGCGACUCAAUCUAAAGAACAUGGUUAAAGACCAUUAGCCAUGUGAUUUUGAUUG AGCCGCGUCAAUAUCUCU
*Petunia*_(*Pax*)_*MIR*-0171o	AGAUAUUGAUGAGGCUCAAUCUGAAGACACAUGAUCAAAAUCAUUAAAUAGGCAUGUGGUUUUU GAUUGAGCCGCGUCAAUAUCUCU
*Petunia*_(*Pin*)_*MIR*-0171o	AGAUAUUGAUGAGGCUGAAUCUGAAGAGACAUGAAAAUCAUUAAAUAGGCAUGUGGUUUUUGAU UGAGCCGCGUCAAUAUCUCU
*Petunia*_(*Pax*)_*MIR*-0172a	GCAGCACCAUCAAGAUUCACAUAGAAAAGUUGAGCAGAAAUUGAAAUCCGCCCAAAAGUUUGAU CAUGAGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pin*)_*MIR*-0172a	GCAGCACCAUCAAGAUUCACAUUGAAAAGUUGAGCAGAAAUUGAAAUCCGCCCAAAAGUUUGAU CAUGAGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pax*)_*MIR*-0172b	GUAGCAUCAUCAAGAUUCACAUGCAAAAGGCAAAGUGGUGAGUCUGAUGAAAUUAUGACAUAGC CAUGGCUUUUUGAAGGUGAGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pin*)_*MIR*-0172b	GUAGCAUCAUCAAGAUUCACAUGCAAAAGGCAAAGAGGUGAGUGUGAUGAAAUUAUGACACAGC CAUGGCUUUUUGAAGGUGAGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pax*)_*MIR*-0172c	GUAGCAUCAUCAAGAUUCACAUACUGAAGGCAAGGUUAAUGAAAUGAAAUAGAAAUGACCAUGG CCUUAUUGAAAGUGAGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pin*)_*MIR*-0172c	GUAGCAUCAUCAAGAUUCACAUAUUGAAGGCAAGGUUAAUGAAAUGAAUUAGAAAUGACCAUGG CCUUAUUGAAAGUGAGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pax*)_*MIR*-0172d	GCAGCAUUAUCAAGAUUCACAUACAAUAUUAAUGUGGAGAAAAAAAUAUUACUUCUAAAAUCUG CCUCCAUGUUUUUCAACAUGAGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pin*)_*MIR*-0172d	GCAGCAUUAUCAAGAUUCACAUACAAUAUUAAUGUGGAGCAGAAAAAAUAUUACUUCUAAAAUC UGCAUCCAUGUUUUUCAACAUGAGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pax*)_*MIR*-0172e	GCAUCAUCAUCAAGAUUCACAUAGACAUGUGGAGCACAAAAAGAAUGCUAGUAUUAAAUUAUGC UUCCAAGUUUCUGAACAUGAGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pin*)_*MIR*-0172e	GCAUCAUCAUCAAGAUUCACAUAGACAUGUUGAGCACAAAAAGAAUGCUAGUAUUAAAUUAUGC UUCCAAGUUUCUGAACAUGAGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pax*,*Pin*)_*MIR*-0172f	GCAGCAUUAUUAAGAUUCACAUAUAAAUUAAUGUGGAACAGAGAGAAAAUAUAUACUUCUAACA UCUGCCUCCAUGUUUUUGAACAUGAGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pax*,*Pin*)_*MIR*-0172g	GCAGCAUCCUCAAGAUUCACAUACAUAUAUGUGCAGUGACGUGCCAUUAUAUUGUCAUACUUUA UUCUUAACUAGAGUAUGAGAAUCUUGAUGAUGCUGCAU
*Petunia*_(*Pax*)_*MIR*-0172j	GGAGCAUCAUCAAGAUUCACAUAGCCUUGUUAGGGUUUCAUAGGGGUGAGGAUAAUUACAUUUU UGCCCCUAUUGCUCAUUGAUUGUGGGAAUCUUGAUGAUGCUGCAG
*Petunia*_(*Pin*)_*MIR*-0172j	GGAGCAUCAUCAAGAUUCACAUGGCCUUGUUAGGGUUUCAUAGGGGUGAGAAUAAUUAUAUUUU UUCCCCUAUUGCUCAUUGAUUGUGGGAAUCUUGAUGAUGCUGCAG
*Petunia*_(*Pax*)_*MIR*-0172k	GCAGCAUCUUCAAGAUUCACAUAGCUUUAUAUAGAGUUCCAUGGUUGGUGAAAGUAAAAGAUGU ACUACUCAUAUUCUUUUGCUACUAUGGCUCUUUGAUGUGGGAAUCUUGAUGAUGCUGCAG
*Petunia*_(*Pin*)_*MIR*-0172k	GCAGCAUCUUCAAGAUUCACAUAGCUUUAUAUAGAGUUCCAUGGUUGGUGAAAGUAAAAUAUGU UGUACUAAUAUUCUUUUGCUACUAUGGCUCUUUGAUGUGGGAAUCUUGAUGAUGCUGCAG
*Petunia*_(*Pax*)_*MIR*-0319a	AGAGCUUUCUUCAGUCCACUCAUGGGGGGCAAUAGGGUUCAAUUUGCUGCUGACUCAUUCAUCC AAAUGCUGAGGUUUUAUAGUUGCUAGCACCUUAGUAGCUGAGUGAAUGAAGUGGGAGACAAGUU GGAUCAUAAGCUUCCUGUACUUGGACUGAAGGGAGCUCCCU
*Petunia*_(*Pin*)_*MIR*-0319a	AGAGCUUUCUUCAGUCCACUCAUGGGGAGCAAUAGGGUUCAAUUUGCUGCUAACUCAUUCAUCC AAAUGCUGAGGUUUUUAAGUUGCUAGCACCUUAGUAGCUGAGUGAAUGAAGUGGGAGACAAGUU GGAUCAUAAGCUUCCUGUACUUGGACUGAAGGGAGCUCCCU
*Petunia*_(*Pax*)_*MIR*-0319b	AGAGCUUUCUUCAGUCCACAUAUAGGGGGCAAUAUGGUUCAAUUAGCUGCUGACUCAUUCACAC

	AAAUGCUAAGGCCUUCAAUUGAAAUUUAUAAGCCCUUAGUAGCUGAGUGAAUGAAGUGGGAGAC AAGUUGAAUCUUAUGCUUUCUGUGCUUGGACUGAAGGGAGCUCCCU
*Petunia*_(*Pin*)_*MIR*-0319b	AGAGCUUUCUUCAGUCCACAUAUAGGGGGCAAUAUGGUUCAAUUAGCUGCUGACUCAUUCACAC AAAUGCUAAGGCCUUCAGUUGAAAUUUAUAAGCCCUUAGUAGCUGAGUGAAUGAAGUGGGAGAC AAGUUGAAUCUUAUGCUUUCUGUGCUUGGACUGAAGGGAGCUCCCU
*Petunia*_(*Pax*)_*MIR*-0319c	GGAGCUCUCUCCAGUCCAGUCCGAGGCAGAUCGAAGGCUAUAAAAACAGCUGCUGACUCGUUGA UUCUUAAGCACAUCAAUAAGUGUAAAGAAAUUGAGGUGUUUGGGAUUCAACGAUGCAUGAGCUG UAUUUAGCUAUCGCUGUCGCGUCUUGGACUGAAGGGAGCUCCCU
*Petunia*_(*Pin*)_*MIR*-0319c	GGAGCUCUCUCCAGUCCAGUCCGAGGCAGAUCGAAGGCUAUAAAAACAGCUGCUGACUCGUUGA UUCUUAAGCACAUCAUUAAGUGUAAAGAAAUUGAGGUGUUUGGGAUUCAACGAUGCAUGAGCUG UACUUAGCUAUCGCUGUCACGUCUUGGACUGAAGGGAGCUCCCU
*Petunia*_(*Pax*)_*MIR*-0319d	AGAGCUUCCUUUAGUCCACUCAUAGGUGGAUAAAGGAUUUGAAUUAUCUGCCGACUCAUUCAUU CAAACACAGUAGGAUAUCUUUGUGUUUACAGUACUGUGAAUGUGUGAAUGAUGCGGGAGAUAAA UCAUCCUUUUCUAUCUUUGCUUGGACUGAAGGGAGCUCCCU
*Petunia*_(*Pin*)_*MIR*-0319d	AGAGCUUCCUUUAGUCCACUCAUAGGUGGAUAAAGGAUUUGAAUUAUAUGCCGACUCAUUCAUU CAAACACAGUAGGAUAUCUUUGUGUUUACAGUACUGUGAAUGUGUGAAUGAUGCGGGAGAUAAA UCAUCCUUUUCUAUCUUUGCUUGGACUGAAGGGAGCUCCCU
*Petunia*_(*Pax*)_*MIR*-0319e	AGAGCUUCCUUCAGUCCACUCAUAGGUGGAUGAAGGGAUUUGGAUUAGCUGCCGACUCAUUCAU UCAAACACGGUAGAAACAAUAUAUACAUUUAUAUACUACCGUGAAUGUGUGAAUGAUGCGGGAG GUAAAUUCAUCCUUUUCUAUCUGUGCUUGGACUGAAGGGAGCUCCCU
*Petunia*_(*Pin*)_*MIR*-0319e	AGAGCUUCCUUCAGUCCACUCAUAGGUGGAUGAAUGGAUUUGGAUUAGCUGCCGACUCAUUCAU UCAAACACGGUAGAAACAAUAUAUACAUUUAUAUACUACCGUGAAUGUGUGAAUGAUGCGGGAG GUAAAUUCAUCCUUUUCUAUCUGUGCUUGGACUGAAGGGAGCUCCCU
*Petunia*_(*Pax*,*Pin*)_*MIR*-0319h	GAGCUUCCUUCAGCCCACUCAUGGAGGAGAAUUGGGGUUGAACUAGCUGCCGACUCAUUCACCC AACCACUCAGUAGAAAAGGAUAGAUUUUGUGCUACUGUGAUUGAGUGAAUGAUGCGGGAGAUAG UUUUCUAUUCCUCUCUUUCUUUGCUUGGACUGAAGGGAGCUCCU
*Petunia*_(*Pax*,*Pin*)_*MIR*-0319i	GAGCUCCUUUCAGGCCAAGACCGAGGGUCGAGAAGCGGCAAGAGCUGCCAUGUCAUGCAUUUUG GUUAAACUUAACUAGCAAAAGCGUUGAGAAUAACCAGUUGCAUGACAGGGGAGCAACUACUUCC GCUAACUUGCUCCACUUAUUGGACUGAAGGGAGCUCCU
*Petunia*_(*Pin*)_*MIR*-0319i2	GAGCUCCUUUCAGGCCAAGACCGAGGGUCGAGGAGCGGCAAGAGCUGCCAUGUCAUGCAUUUUG GUUAAACUUAACUAGCAAAAGCGUUGAGAAUAACCAGUUGCAUGACAGGGGAGCAACUACUUCC GCUAACUUGUUCCACUAAUUGGACUGAAGGGAGCUCCU
*Petunia*_(*Pax*)_*MIR*-0390a	AAGCUCAGGAGGGAUAGCGCCAUUGAAAAUGAUGUGAUGUUUAAUUUAUUAACUUUUGUUGCUU UUGCUUCUUUUUUCCUUCUCUUUUUCUGAGUUUGUCUUUUCCAUAAAACCAUCAAUUAUUUGGC GCUAUCCAUCCUGAGUUCUA
*Petunia*_(*Pin*)_*MIR*-0390a	AAGCUCAGGAGGGAUAGCGCCAUGAAAAAUGAUGUGAUGUUUAAUUUGUUAACUUUUGUCUUCU UUUUUCCUUCUUUUUUUCUGAGUUUGUCUUUUCCAUAAAACCAUCAAUUAUUUGGCGCUAUCCA UCCUGAGUUCUA
*Petunia*_(*Pax*)_*MIR*-0390b	AAGCUCAGGAGGGAUAGCGCCAUGAACAGAUUUCUGUUGUUGGCAAUUUCUUUUUAUUUUUCCU UUGUCGAAUCUUCUUUUCGUGGGUUUUAUUUUUCCCAUAACUUACAAUUCUGUGGCGCUAUCCA UCCUGAGUUUCA
*Petunia*_(*Pin*)_*MIR*-0390b	AAGCUCAGGAGGGAUAGCGCCAUGAACAGAUUUAUGUUGUUGGGAAUUUCUUUUUAUCUUUCCC UUGUUGAAUCUUAUUUUCGUGGGUUUUAUUUUUCCCAUAGCUUACAAUUCUGUGGCGCUAUCCA UCCUGAGUUUCA
*Petunia*_(*Pax*)_*MIR*-0390c	AAGCUCAGGAGGGAUAGCGCCAUCGAUGACUACGGUACAUUAGCUUAAUAGAAAAAAAAUAUAU UUAAGCUGUGUUACGUUGGACAUCUGUAGCGCUAUCCAUCCUGAGUUUCA
*Petunia*_(*Pin*)_*MIR*-0390c	AAGCUCAGGAGGGAUAGCGCCAUGGAUGAUUAUUGUACAAAAAUGUACAUGAUGAGCUUAAUUA AAACUAUUGUUUUUAGGCUGUCUUCAUUUAACCAUCUGUAGCGCUAUCCAUCCUGAGUUUCA
*Petunia*_(*Pax*)_*MIR*-0393a	UCCAAAGGGAUCGCAUUGAUCCCAUUUUACUUCUGACAAAAAAUAGUGAGUUUGGAUCAUGCUAU CCCUUUGGACU
*Petunia*_(*Pin*)_*MIR*-0393a	UCCAAAGGGAUCGCAUUGAUCCCAUUUUACUUCUGACAAAAAAUAGUGAGUUUGGAUCAUGCUA UCCCUUUGGACU
*Petunia*_(*Pax*)_*MIR*-0393b	UCCAAAGGGAUCGCAUUGAUCCCAUUUCAUUUGUUGUACUAUUUCUUUCAAGAAAUUGAAUAGU GAAUUCGGAUCAUGCUAUCCCUUUGGACU
*Petunia*_(*Pin*)_*MIR*-0393b	UCCAAAGGGAUCGCAUUGAUCCCAUUUCAUUUGUUGUAGUAUUUCUUUCAAGAAAUUGAAUAGU GAAUUCGGAUCAUGCUAUCCCUUUGGACU
*Petunia*_(*Pin*)_*MIR*-0393b2	UCCAAAGGGAUCGCAUUGAUCCCAUUUCAUUUAGGUUUACUAUUUCUUUCAAGAAAUUGAAUAG UGAAUUCGGAUCAUGCUAUCCCUUUGGACU

*Petunia*_(*Pax*)_*MIR*-0393c	UCCAAAGGGAUCGCAUUGAUCCCGUGUCCCAGCUAACAGCUUAAAGAACAGCUGGUAAUGGGAU CAUGCGAUCUCUUCGGAAU
*Petunia*_(*Pin*)_*MIR*-0393c	UCCAAAGGGAUCGCAUUGAUCCUGUGUCCCAACUAAUUAACAGCUUAAAUAACUGGUAAUGGGA UCGUGCGAUCUCUUCGGAAU
*Petunia*_(*Pax*,*Pin*)_*MIR*-0394a	UUGGCAUUCUGUCCACCUCCAUCGUUAUAGCUCGAUCUCUAAUGGGUGCUUUAUCAUAGUGUUU GGUAGCAUCUCCAUGAGAUGAGGAGGGGGCCAAAGUGCCAAAC
*Petunia*_(*Pax*)_*MIR*-0394b	UUGGCAUUCUGUCCACCUCCGUGGAUCUUAAUUUCCUUUCAAGAAAAUACACAAAUUAAAGGAG UGUAAAGUGGAGGUGGGCAUACUGCCAACA
*Petunia*_(*Pin*)_*MIR*-0394b	UUGGCAUUCUGUCCACCUCCGUAGAUCUUAAUUUCCUUUCAAGAAAAUACACAAAGAAGUGUAA AGUGGAGGUGGGCAUACUGCCAACA
*Petunia*_(*Pax*)_*MIR*-0394c	UUGGCAUUCUGUCCACCUCCGCUUUUUCUUGUUUCAAGAAUUAUUGAAAGAGGGACCAAAAAAA GGACAAAAAAGAAAAGAAGAUUUAACGUGGAGGUGGGCAUACUGCCAACA
*Petunia*_(*Pin*)_*MIR*-0394c	UUGGCAUUCUGUCCACCUCCGCUUUUUCUUGUUUCAAGAAUUGAAAGAGGGACCAAGUAAAAUA AAAGGACAAAAAAGAAAAAAGAUGAUUUAACGUGGAGGUGGGCAUACUGCCAACA
*Petunia*_(*Pax*)_*MIR*-0395a	GUUCUCUCAAAUCACUUCAUUGAGUCACUAUUCCUUCUUUUAAAGGGAAAUGAAUCUACCCACU GAAGUGUUUGGGGGAACUC
*Petunia*_(*Pin*)_*MIR*-0395a	GUUCUCUCAAAUCACUUCAUUGAGUUACUAUUCCUUCUUUUAAGGGGAAAUGAAUCUACCCACU GAAGUGUUUGGGGGAACUC
*Petunia*_(*Pin*)_*MIR*-0395a2	GUUCUCUCAAAUCACUUCAUUGAGUUAUUAUUCCAUUUUCUAAGUGGAAAUGAAUGUACCCACU GAAGUGUUUGGGGGAACUC
*Petunia*_(*Pax*)_*MIR*-0395b	GUUCUCCUGAUCACUUCAUUGGGAUUUUAUAAAUGCCCUAAUUUGAAAGGGUAUAAAAUAUAAC UACUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pin*)_*MIR*-0395b	GUUCUCCUGAUCACUUCAUUGGAUUUUAUAAAUGCCCUAAUUUGAAAGGGUAUAAAAUAUAUCU ACUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pin*)_*MIR*-0395b2	GUUCUCCUGAUCACUUCAUUGGGAUUUUAUAAAUGCCCUAAUUUAAAAGGGUAUAAAAUAUAAC UACUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0395c	GUUCUCCUGAUCACUUCAUUGGGAUUUUAUAAAUGCCUUAAUUUGAAAGGGUAUUAAAUAUACC UACUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pin*)_*MIR*-0395c2	GUUCUCCUGAUCACUUCAUUAGGAUGUUAUAAAUGCCUUAAUUUGAAAGAGUAUUAAAUAUACC UACUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pax*)_*MIR*-0395d	GUUCUCCUGAUCACUUCAUUGGGAUUUUAUAAAUGCCAUAACUUGGAAGGGUAUUAAGAAUACC CACUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pin*)_*MIR*-0395d	GUUCUCCUGAUCACUUCAUUGGGAUUUUAUAAAUGCCAUAACUUGGAAGGGUAGAAUACCCACU GAAGUGUUUGGGGGAACUC
*Petunia*_(*Pax*)_*MIR*-0395e	GUUCUCCUGAUCACUUCAUUGGGAUUUUUAUGAAUGCCAUCACCUGGAAGGGUAUUAAGAAUAA UCACUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pin*)_*MIR*-0395e	GUUCUCCUGAUCACUUCAUUGGGAUUUUUAUGAAUACCAUAACUUUGGAAGGGUAUUAAGAAUA CCCACUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pin*)_*MIR*-0395e2	GUUCUCCUGAUCACUUCAUUGGGAUUUUUAUGAAUGCCAUAAUUUGGAAGGGUAUUAAAAAUAC CCACUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pax*)_*MIR*-0395f	GUUCUCCUGAUCACGUCAUUGGGAUUUUAUAAAUGCCCUAAUUUGAAAGGGUAUAAAACAUAAC UACUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0395g	GUUUCCCUGACCGCUUCAUGAGGGCUUAUUUCUCCACAAGUUGUUGCUUGAAAUUAGCCCUGCU GAAGUGUUUGGGGGAACUC
*Petunia*_(*Pax*)_*MIR*-0395h	GUUUCCCUGACCGCUUCAUGAGGGCUUAUUUCUCCCAAGUUGUUGCUCGAAAUUAGCCCUGCUG AAGUGUUUGGGGGAACUC
*Petunia*_(*Pax*)_*MIR*-0395i	GUUCCCUUAAAUGCUUCAUGAGAGUCUUAAUUACUGAGUACUUCAUAACUUUGUAGUUGGAUGC CCUCCUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pax*)_*MIR*-0395j	GUUCCCUUAAAUGCUUCAUGAGGGCUUAAUUACUGAGUACUUCAUAACUUUGUAGUUGGAUGCC CUCCUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pin*)_*MIR*-0395j	GUUCCCCUAAAUGCUUCAUGAGGGCUUAAUUACUGAGUACUUCAUAACUUUGUAGUUGGAUACC CUCCUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pax*)_*MIR*-0395k	GUUCCCUUGACCACUUCAUGAGGGCUUAUGAUCUUCACAAGUUGUUGAUUGAAGCCCUGCUGAA GUGUUUGGGGGAACUC
*Petunia*_(*Pin*)_*MIR*-0395k	GUUCCCUUGAUCACUUCAUGAGGGCUUAUGAUCUUCACAAGUUGUUGAUCGAAGCCCUGCUGAA GUGUUUGGGGGAACUC

*Petunia*_(*Pax*)_*MIR*-0395l	GUUCCCUUGACCACUUCAUGAGGGCUUAUAUCUUCACAAGUUGUUGAUCGGAGCCCUGCUGAAG UGUUUGGGGGAACUC
*Petunia*_(*Pin*)_*MIR*-0395m	GUUCUCCUCAGCACUUCAUUGGGACUGAAAAGAUACCUACUGAAGUGUUUGGGGGAACUC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0396a	UUCCACAGCUUUCUUGAACUGCAUCUUUCAAAAUUAACCACCAAUAUGGUUACGAGAAUAUGUU GCGGUUCAAUAAAGCUGUGGGAAG
*Petunia*_(*Pax*)_*MIR*-0396b	UUCCACAGCUUUCUUGAACUGCAUCUCUGAAAAAAACUCAUCACUAUGAGCAAAGACAGAAAUA GUUGCGGUUCAAUAUAGCUGUGGGAAG
*Petunia*_(*Pin*)_*MIR*-0396b	UUCCACAGCUUUCUUGAACUGCAUCUCUCAAAAAACUCAUCACUAUGAGCAAGAGAGAAAUUGU UGCGGUUCAAUAUAGCUGUGGGAAG
*Petunia*_(*Pax*)_*MIR*-0396c	UUCCACAGCUUUCUUGAACUUCUUCUUGCUGAAUUUGAUCUCUAAUUGGCAAUUUUGGAAGCAG UUUGAGAUGAGAUUAAAGCUAUGAAAGUCCAAGAAAGCUGUGGGAAA
*Petunia*_(*Pin*)_*MIR*-0396c	UUCCACAGCUUUCUUGAACUUCUUCUUGCUGAAUUUUGAUCUCUAAUUGGCAAUUAUGGGAGCA GUUUGAGAUGAGAUUAAAGCUAUGAAAGUCCAAGAAAGCUGUGGGAAA
*Petunia*_(*Pax*,*Pin*)_*MIR*-0396d	UUCCACAGCUUUCUUGAACUUCUCUUUACUUUUCAUCUCUGGCAUAAUCGACCAGUUGAUGAGG UUUAGCUCUGAAAGUUCAAGAAAGCUGUGGGAAA
*Petunia*_(*Pax*)_*MIR*-0397a	AUUGAGUGCAGCGUUGAUGAAAUAUUUCUUUAAUUUCUGUCAAAAUGUUGUCAUUUGGGCAUUC CCUCCAGUUGGUUUUCAUCUACGUUGCACUCAAUUA
*Petunia*_(*Pin*)_*MIR*-0397a	AUUGAGUGCAGCGUUGAUGAAAUAUUUCUUUAAUUUCAGUCAAAAUGUUGCCAUUUGGGCAUUC CCUCCAGUUAGUUUUCAUCUACGCUGCACUCAAUUA
*Petunia*_(*Pin*)_*MIR*-0397b	AUUGAGUGCAGCGUUGAUGAAGAUGCCAAAUAUCCACCAAGUUAUUACUGACUGCUGCUCUAUC AUUAUUCAUUUCACAGAGCGUUUCAUCUACGUUGCACUCAAUCA
*Petunia*_(*Pax*)_*MIR*-0398a	GGAGUGUACCAGGGAACACAUGUGCAUUUUGGCUAAUUUGUUAAUGGUUGGAUCCAAAAUCCAC UUGUGUUCUCAGGUCACCCCUU
*Petunia*_(*Pin*)_*MIR*-0398a	GGAGUGUACCAGGGAACACAUGUGCAUUUUGGCUUUUGUUAAUGGUUGGAUCCAAAAUCCACUU GUGUUCUCAGUUCACCCCUU
*Petunia*_(*Pax*)_*MIR*-0398b	GGAGUGUUCAUGGGAACACAUGUGCAUUUUGGUUAUUGAUAAUGGCUAUUUUAUAAAUGCACUU GUGUUCUCAGGUCACCCCUU
*Petunia*_(*Pax*)_*MIR*-0398e	GGGACGACUUGAGAUCAUAUGUAUGAGCGUUUUUUGUUUUUAACAAUUUUUGUAAAUGUGAGAC AUAUGUUCUCAGGUCGCCCCUG
*Petunia*_(*Pin*)_*MIR*-0398e	GGGACGACUUGAGAUCAUAUGUAUGAGCGUUUUUUGUUUUUAACAAUUUUUAUAAAUGUGAUAC AUAUGUUCUCAGGUCGCCCCUG
*Petunia*_(*Pin*)_*MIR*-0398e2	GGGACGACUUGAGAUCAUAUGUAUGAGCGUUUUUUGUUUUUAACAAUUUUUGUAAAUGUGAAAC AUAUGUUCUCAGGUCGCCCCUG
*Petunia*_(*Pax*)_*MIR*-0398h	CAGGGGCGACCUGAGAACACAUAUUGAAUCACACCAUUCAAUUAAAUUGAAAAAAGAAAAAGUC UCAACAUGUGAUCUCAAGUGGCCCCUAUU
*Petunia*_(*Pin*)_*MIR*-0398h	CAGGGGCGACCUGAGAACACAUAUUGAAUCACACAUUCAAUUAAAUUGGAAAAAAGAAAAAAAA AGAAAAGUGCAACAUGUGAUCUCAAGUGGCCCCUAUU
*Petunia*_(*Pax*,*Pin*)_*MIR*-0399a	GGGCUUCUCUCUAUUGGCAUGCAGUUGUCUAGUAAUUCCACUUCAUCACACAUAUUUCGGCGGACA UGCCAAAGGAGAGUUGCCCUG
*Petunia*_(*Pax*,*Pin*)_*MIR*-0399b	GGGAUACUCUCUAUUGGCAUGCAGUUGUGUUUCAGCUGACAUGCCAAAGGAGAGUUGCCCUG
*Petunia*_(*Pax*,*Pin*)_*MIR*-0399c	GGGCUACUCUCUAUUGGCAUGCAGUUAUGUAUGUGACUUCACUUCAUCUCAUAUUUCAAAUGAC AUGCCAAAGGAGAGUUGCCCUG
*Petunia*_(*Pax*,*Pin*)_*MIR*-0399d	GGGCUACUCUCUAUUGGCAUGCAGUUAAUUAUCUAGUGAUUCCACUUCACAACAUAUUUUAGCU GACACGCCAAAGGAGAGCUGCCCUG
*Petunia*_(*Pax*,*Pin*)_*MIR*-0399e	GGGCUACUCUCUAUUGGCAUGCAGUUUUUUGGGUUGCUCCAUAUAUAUAUAUUUGAUCACAUAU UUCUACUGACAUGCCAAAGGAGAGCUGCCCUG
*Petunia*_(*Pax*,*Pin*)_*MIR*-0399f	GGGCUACUCUCUAUUGGCAUGCAGUUAUGUAUGUGACUUCACAUGUUUCAACUGACAUGCCAAA GGAGAGCUGCCCUG
*Petunia*_(*Pax*)_*MIR*-0399g	GGGCUACUCUCUAUUGGCAUGCAGUUUUUUUGGUGGCUCCAAUGCUUUUUCUACUGACAUGCCA AAGGAGAGCUGCCCUG
*Petunia*_(*Pin*)_*MIR*-0399g	GGGCUACUCUCUAUUGGCAUGCAGUUUUUUGGCGGCUCCAAUGCUUUUUCUACUGUCAUGCCAA AGGAGAGCUGCCCUG
*Petunia*_(*Pax*)_*MIR*-0403a	CGUUUGUGCGUGAAUCUAACAACCCCUUUACAUCAUUAAAACUGUUUCAUUGAUGGGGUGUGUU UGUUAGAUUCACGCACAAACUCG
*Petunia*_(*Pin*)_*MIR*-0403a	CGUUUGUGCGUGAAUCUAACAACCCCUUCACAUCAUUAAAACUGUUUCAUUGAUGGGGUGUGUU AGUUAGAUUCACGCACAAACUCG

*Petunia*_(*Pax*)_*MIR*-0408a	ACAGGGACGAGACAGAGCAUGAGAUGUGCAAUUCUUCAAUUCCUGCCUAUUCCAUGCACUGCCU CUUCCCUGGCU
*Petunia*_(*Pin*)_*MIR*-0408a	ACAGGGACGAGACAGAGCAUGAGAUAUGCAAUUCUUCAAUUCUUGCCUAUUCCAUGCACUGCCU CUUCCCUGGCU
*Petunia*_(*Pax*)_*MIR*-0477a	CCUCUCCCUCAAGGGCUUCUCUCCUGCAUGUUUAUGUGACUUGUUAAAGUAGAAAGCAUGCAAG GAAAGAAACUCUGGCAGGGAGAGCCA
*Petunia*_(*Pax*,*Pin*)_*MIR*-0479a	GUGAUAUUGGGUUGGCUCAUUAUUUUAUGAUGAAUUCAUAAGACAGAGAGAGAUGAGCCGAACC AAUAUCACUC
*Petunia*_(*Pax*)_*MIR*-0482a	GGGAUUGGUGGGUUGGAAAGCUUUUAAGUUUUUUGGUUCUUUUUAAUGCAUUAGCUUUCCAAUU CCACCCAUUCCUA
*Petunia*_(*Pin*)_*MIR*-0482a	GGGAUUGGUGGGUUGGAAAGCUUUUAAGUUUUUUGGUUCUUUUCAAUGCAUAAUUAGCUUUCCA AUUCCACCCAUUCCUA
*Petunia*_(*Pax*)_*MIR*-0827a	UUUGUUGAUGGUCAUCUAGCUAGUCAUCAUUCUGCAAAAUUUUUGCACCCAUUAUGCCAUGGUU AGAUGAACAUCAACAAACA
*Petunia*_(*Pin*)_*MIR*-0827a	UUUGUUGAUGGUCAUCUAGCUAGUCAUCAUUCUGCAUAAUUUUUGCACCCAUUAUGCCAUGGUU AGAUGAACAUCAACAAACA
*Petunia*_(*Pax*,*Pin*)_*MIR*-2111a	UAAUCUGCAUCCUGAGGUUUAGAUCAGAAUAUUUUAACUGCUUUCUAGUCCUUGGGAUGUAGAU UACC
*Petunia*_(*Pax*)_*MIR*-2111b	UAAUCUGCAUCCUGAGGUUUAGAUCGGCUACUUUAAUUAGUACUGCUUCUAGCCCUUGGGAUGC AGAUUACU
*Petunia*_(*Pin*)_*MIR*-2111b	UAAUCUGCAUCCUGAGGUUUAGAUCGGCUUCUUUAGUACUGCUUCUAGCCCUUGGGAUGCAGAU UACU
*Petunia*_(*Pax*)_*MIR*-6149a	UUGAUACGCACCUGAAUCGGCAAAGAAUAUUGGAUGAGACAGUCCAAUUAGAUCUUAUUAUCUA UGUCUUAUGUCGAUUUAGGUUCGUAUUUAAA
*Petunia*_(*Pin*)_*MIR*-6149a	UUGAUACGCACCUGAAUCGGCAAAGAAUAUUGGAUGAGAUAGUCCAAUUAGAUCUUAUUAUCUA UGUCUUAUGUCGAUUUAGGUUCGUAUUUAAA
*Petunia*_(*Pax*)_*MIR*-8016a	CAUGGUCUUUUCUUUCAAAAAUAUAUUAUAUUAGGAGUUUAUAUAUUUUUGAAUGGAAGGCCCA UGUG
*Petunia*_(*Pin*)_*MIR*-8016a	CAUGGUCUUUUCUUUCAAAAAUAUAUUAUAUUAGAAGUUUAUAUAUUUUUGAAUGGAAGGCCCA UGUG
*Petunia*_(*Pax*,*Pin*)_*MIR*-0157a(v2)	UUGACAGAAGAUAGAGAGCACUGAUGAUGAUUUGCUAAAGUAGCAUCUCAAUUCAUUUGUGCUC UCUAUGCUUCUGUCAUCA
*Petunia*_(*Pax*)_*MIR*-0157b(v2)	UUGACAGAAGAUAGAGAGCACAGAUGAUGAAGUACAUGGAAACUUCUGUACCUCACUCCUUUGU GCUCUUUAUUUUUCUGUCAUCA
*Petunia*_(*Pin*)_*MIR*-0157b(v2)	UUGACAGAAGAUAGAGAGCACAGAUGAUGAAGUGCAUGGAAACUUCUGUACCUCACUCUUUUGU GCUCUUUAUUCUUCUGUCAUCA
*Petunia*_(*Pax*)_*MIR*-0157c(v2)	UUGACAGAAGAUAGAGAGCACAGAUGAUGAAGUGCACGGAAGCUUUAUGCACCUCACUCCUUUG UGCUCUUUAUCCUUCUGUCAUCA
*Petunia*_(*Pin*)_*MIR*-0157c(v2)	UUGACAGAAGAUAGAGAGCACAGAUGAUGAAGUGCACGGAAGCUUUAUGCACCUCACUCCUUUG UGCUCUUUAUUCUUCUGUCAUCA
*Petunia*_(*Pax*)_*MIR*-0157d(v2)	UUGACAGAAGAUAGAGAGCACACAUGAUGAAAUGCUAAAUUUGGAAGGCACAAAGCAUCUUAAU UCAUGUGUGCUCUCUAUGCUUCCGUCAUCA
*Petunia*_(*Pin*)_*MIR*-0157d(v2)	UUGACAGAAGAUAGAGAGCACAGAUGAUGAAAUGCUAAAAGGCACUGAAUAAACUGCAAAAGCA UCUUAAUUCAUUUGUGCUCUCUAUGCUUCCGUCAUCA
*Petunia*_(*Pax*)_*MIR*-0157e(v2)	UUGACAGAAGAUAGAGAGCACAGAUGAUGAGAUGUUUAAUUGGAAGCUAUCUGCAUCUCACUCC UUUGUGCUCUCUAUUCUUCUGCCAUCA
*Petunia*_(*Pin*)_*MIR*-0157e(v2)	UUGACAGAAGAUAGAGAGCACAGAUGAUGAGAUGUUUAAUUGGAAGCUAGCCGCAUCUCACUCC UUUGUGCUCUCUAUUCUUCUGCCAUCA
*Petunia*_(*Pax*)_*MIR*-0171a(v2)	AGAUGUUGGUGCGGUUCAAUGAGAAAGUAUCGCUCAACAAGUAAAUUUGACCCUAUUUUUUGAU UGAGCCGUGCCAAUAUCACG
*Petunia*_(*Pin*)_*MIR*-0171a(v2)	AGAUGUUGGUGCGGUUCAAUGAGAAAGUAUCGCUCAACAAGUAUAUUUGACCCUACUUUUUGAU UGAGCCGUGCCAAUAUCACG
*Petunia*_(*Pax*)_*MIR*-0171b(v2)	AGAUAUUGGUGCGGUUCAAUGAGAAAGCAGUAAUCGAGAAGUUUUGACUCUACUUUUUGAUUGA GCCGUGCCAAUAUCACG
*Petunia*_(*Pin*)_*MIR*-0171b(v2)	AGAUAUUGGUGCGGUUCAAUGAGAAAGCAGUACUCGAGAAGUUUUGACUCUACUUUUUGAUUGA GCCGUGCCAAUAUCACG

*Petunia*_(*Pax*)_*MIR*-0171c(v2)	AGAUAUUGAUGCGGUUCAAUUAGAAAGCCGAAUUCUUUGUGUUUAGAAUCCUGUUAUUUGAUUG AGCCGUGCCAAUAUCACG
*Petunia*_(*Pin*)_*MIR*-0171c(v2)	AGAUAUUGGUGCGGUUCAAUUAGAAAGCCGAAUUCUUUGUGUUUAGAACUCUGUUAUUUGAUUG AGCCGUGCCAAUAUCACG
*Petunia*_(*Pax*,*Pin*)_*MIR*-0171j(v2)	GAUGUUGGUGAGGUUCAAUCUGAAGACGGGUUUACGUUUUGUUUGCGUAAAGAACGAUCUCUGA UUGAGCCGCGCCAAUAUCAC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0171j(v3)	UGUUGGUGAGGUUCAAUCUGAAGACGGGUUUACGUUUUGUUUGCGUAAAGAACGAUCUCUGAUU GAGCCGCGCCAAUAUC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0171k(v2)	GAUGUUGGUGAGGUUCAAUCCGAAGACGGGUUUACGUUUUGUUUUCGUAAAGAACGAUCUCAGA UUGAGCCGCGCCAAUAUCAC
*Petunia*_(*Pax*)_*MIR*-0172a(v2)	AGCACCAUCAAGAUUCACAUAGAAAAGUUGAGCAGAAAUUGAAAUCCGCCCAAAAGUUUGAUCA UGAGAAUCUUGAUGAUGCUGC
*Petunia*_(*Pin*)_*MIR*-0172a(v2)	AGCACCAUCAAGAUUCACAUUGAAAAGUUGAGCAGAAAUUGAAAUCCGCCCAAAAGUUUGAUCA UGAGAAUCUUGAUGAUGCUGC
*Petunia*_(*Pax*)_*MIR*-0172b(v2)	AGCAUCAUCAAGAUUCACAUGCAAAAGGCAAAGUGGUGAGUCUGAUGAAAUUAUGACAUAGCCA UGGCUUUUUGAAGGUGAGAAUCUUGAUGAUGCUGC
*Petunia*_(*Pin*)_*MIR*-0172b(v2)	AGCAUCAUCAAGAUUCACAUGCAAAAGGCAAAGAGGUGAGUGUGAUGAAAUUAUGACACAGCCA UGGCUUUUUGAAGGUGAGAAUCUUGAUGAUGCUGC
*Petunia*_(*Pax*)_*MIR*-0172c(v2)	AGCAUCAUCAAGAUUCACAUACUGAAGGCAAGGUUAAUGAAAUGAAAUAGAAAUGACCAUGGCC UUAUUGAAAGUGAGAAUCUUGAUGAUGCUGC
*Petunia*_(*Pin*)_*MIR*-0172c(v2)	AGCAUCAUCAAGAUUCACAUAUUGAAGGCAAGGUUAAUGAAAUGAAUUAGAAAUGACCAUGGCC UUAUUGAAAGUGAGAAUCUUGAUGAUGCUGC
*Petunia*_(*Pax*)_*MIR*-0172d(v2)	AGCAUUAUCAAGAUUCACAUACAAUAUUAAUGUGGAGAAAAAAAUAUUACUUCUAAAAUCUGCC UCCAUGUUUUUCAACAUGAGAAUCUUGAUGAUGCUGC
*Petunia*_(*Pin*)_*MIR*-0172d(v2)	AGCAUCAUCAAGAUUCACAUAUUGAAGGCAAGGUUAAUGAAAUGAAUUAGAAAUGACCAUGGCC UUAUUGAAAGUGAGAAUCUUGAUGAUGCUGC
*Petunia*_(*Pax*)_*MIR*-0172e(v2)	AUCAUCAUCAAGAUUCACAUAGACAUGUGGAGCACAAAAAGAAUGCUAGUAUUAAAUUAUGCUU CCAAGUUUCUGAACAUGAGAAUCUUGAUGAUGCUGC
*Petunia*_(*Pin*)_*MIR*-0172e(v2)	AGCAUCAUCAAGAUUCACAUAUUGAAGGCAAGGUUAAUGAAAUGAAUUAGAAAUGACCAUGGCC UUAUUGAAAGUGAGAAUCUUGAUGAUGCUGC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0172f(v2)	AGCAUUAUUAAGAUUCACAUAUAAAUUAAUGUGGAACAGAGAGAAAAUAUAUACUUCUAACAUC UGCCUCCAUGUUUUUGAACAUGAGAAUCUUGAUGAUGCUGC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0172g(v2)	AGCAUCCUCAAGAUUCACAUACAUAUAUGUGCAGUGACGUGCCAUUAUAUUGUCAUACUUUAUU CUUAACUAGAGUAUGAGAAUCUUGAUGAUGCUGC
*Petunia*_(*Pax*)_*MIR*-0172j(v2)	GCAUCAUCAAGAUUCACAUAGCCUUGUUAGGGUUUCAUAGGGGUGAGGAUAAUUACAUUUUUGC CCCUAUUGCUCAUUGAUUGUGGGAAUCUUGAUGAUGCUG
*Petunia*_(*Pin*)_*MIR*-0172j(v2)	GCAUCAUCAAGAUUCACAUGGCCUUGUUAGGGUUUCAUAGGGGUGAGAAUAAUUAUAUUUUUUC CCCUAUUGCUCAUUGAUUGUGGGAAUCUUGAUGAUGCUG
*Petunia*_(*Pax*)_*MIR*-0172k(v2)	GCAUCUUCAAGAUUCACAUAGCUUUAUAUAGAGUUCCAUGGUUGGUGAAAGUAAAAGAUGUACU ACUCAUAUUCUUUUGCUACUAUGGCUCUUUGAUGUGGGAAUCUUGAUGAUGCUG
*Petunia*_(*Pin*)_*MIR*-0172k(v2)	GCAUCUUCAAGAUUCACAUAGCUUUAUAUAGAGUUCCAUGGUUGGUGAAAGUAAAAUAUGUUGU ACUAAUAUUCUUUUGCUACUAUGGCUCUUUGAUGUGGGAAUCUUGAUGAUGCUG
*Petunia*_(*Pax*)_*MIR*-0319a(v2)	AGCUUUCUUCAGUCCACUCAUGGGGGGCAAUAGGGUUCAAUUUGCUGCUGACUCAUUCAUCCAA AUGCUGAGGUUUUAUAGUUGCUAGCACCUUAGUAGCUGAGUGAAUGAAGUGGGAGACAAGUUGG AUCAUAAGCUUCCUGUACUUGGACUGAAGGGAGCUCC
*Petunia*_(*Pin*)_*MIR*-0319a(v2)	AGCUUUCUUCAGUCCACUCAUGGGGAGCAAUAGGGUUCAAUUUGCUGCUAACUCAUUCAUCCAA AUGCUGAGGUUUUUAAGUUGCUAGCACCUUAGUAGCUGAGUGAAUGAAGUGGGAGACAAGUUGG AUCAUAAGCUUCCUGUACUUGGACUGAAGGGAGCUCC
*Petunia*_(*Pax*)_*MIR*-0319b(v2)	AGCUUUCUUCAGUCCACAUAUAGGGGGCAAUAUGGUUCAAUUAGCUGCUGACUCAUUCACACAA AUGCUAAGGCCUUCAAUUGAAAUUUAUAAGCCCUUAGUAGCUGAGUGAAUGAAGUGGGAGACAA GUUGAAUCUUAUGCUUUCUGUGCUUGGACUGAAGGGAGCUCC
*Petunia*_(*Pin*)_*MIR*-0319b(v2)	AGCUUUCUUCAGUCCACAUAUAGGGGGCAAUAUGGUUCAAUUAGCUGCUGACUCAUUCACACAA AUGCUAAGGCCUUCAGUUGAAAUUUAUAAGCCCUUAGUAGCUGAGUGAAUGAAGUGGGAGACAA GUUGAAUCUUAUGCUUUCUGUGCUUGGACUGAAGGGAGCUCC
*Petunia*_(*Pax*)_*MIR*-0319c(v2)	AGCUCUCUCCAGUCCAGUCCGAGGCAGAUCGAAGGCUAUAAAAACAGCUGCUGACUCGUUGAUU CUUAAGCACAUCAAUAAGUGUAAAGAAAUUGAGGUGUUUGGGAUUCAACGAUGCAUGAGCUGUA UUUAGCUAUCGCUGUCGCGUCUUGGACUGAAGGGAGCUCC

*Petunia*_(*Pin*)_*MIR*-0319c(v2)	AGCUCUCUCCAGUCCAGUCCGAGGCAGAUCGAAGGCUAUAAAAACAGCUGCUGACUCGUUGAUU CUUAAGCACAUCAUUAAGUGUAAAGAAAUUGAGGUGUUUGGGAUUCAACGAUGCAUGAGCUGUA CUUAGCUAUCGCUGUCACGUCUUGGACUGAAGGGAGCUCC
*Petunia*_(*Pax*)_*MIR*-0319d(v2)	AGCUUCCUUUAGUCCACUCAUAGGUGGAUAAAGGAUUUGAAUUAUCUGCCGACUCAUUCAUUCA AACACAGUAGGAUAUCUUUGUGUUUACAGUACUGUGAAUGUGUGAAUGAUGCGGGAGAUAAAUC AUCCUUUUCUAUCUUUGCUUGGACUGAAGGGAGCUCC
*Petunia*_(*Pin*)_*MIR*-0319d(v2)	AGCUUCCUUUAGUCCACUCAUAGGUGGAUAAAGGAUUUGAAUUAUAUGCCGACUCAUUCAUUCA AACACAGUAGGAUAUCUUUGUGUUUACAGUACUGUGAAUGUGUGAAUGAUGCGGGAGAUAAAUC AUCCUUUUCUAUCUUUGCUUGGACUGAAGGGAGCUCC
*Petunia*_(*Pax*)_*MIR*-0319e(v2)	AGCUUCCUUCAGUCCACUCAUAGGUGGAUGAAGGGAUUUGGAUUAGCUGCCGACUCAUUCAUUC AAACACGGUAGAAACAAUAUAUACAUUUAUAUACUACCGUGAAUGUGUGAAUGAUGCGGGAGGU AAAUUCAUCCUUUUCUAUCUGUGCUUGGACUGAAGGGAGCUCC
*Petunia*_(*Pin*)_*MIR*-0319e(v2)	AGCUUCCUUCAGUCCACUCAUAGGUGGAUGAAUGGAUUUGGAUUAGCUGCCGACUCAUUCAUUC AAACACGGUAGAAACAAUAUAUACAUUUAUAUACUACCGUGAAUGUGUGAAUGAUGCGGGAGGU AAAUUCAUCCUUUUCUAUCUGUGCUUGGACUGAAGGGAGCUCC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0319h (v2	AGCUUCCUUCAGCCCACUCAUGGAGGAGAAUUGGGGUUGAACUAGCUGCCGACUCAUUCACCCA ACCACUCAGUAGAAAAGGAUAGAUUUUGUGCUACUGUGAUUGAGUGAAUGAUGCGGGAGAUAGU UUUCUAUUCCUCUCUUUCUUUGCUUGGACUGAAGGGAGCUCC
*Petunia*_(*Pax*,*Pin*)_*MIR*-0319h(v3	AGAGCUUCCUUCAGCCCACUCAUGGAGGAGAAUUGGGGUUGAACUAGCUGCCGACUCAUUCACC CAACCACUCAGUAGAAAAGGAUAGAUUUUGUGCUACUGUGAUUGAGUGAAUGAUGCGGGAGAUA GUUUUCUAUUCCUCUCUUUCUUUGCUUGGACUGAAGGGAGCUCCUU
*Petunia*_(*Pax*)_*MIR*-0390a(v2)	AAGCUCAGGAGGGAUAGCGCCAUUGAAAAUGAUGUGAUGUUUAAUUUAUUAACUUUUGUUGCUU UUGCUUCUUUUUUCCUUCUCUUUUUCUGAGUUUGUCUUUUCCAUAAAACCAUCAAUUAUUUGGC GCUAUCCAUCCUGAGUUCUA
*Petunia*_(*Pin*)_*MIR*-0390a(v2)	AAGCUCAGGAGGGAUAGCGCCAUGAAAAAUGAUGUGAUGUUUAAUUUGUUAACUUUUGUCUUCU UUUUUCCUUCUUUUUUUCUGAGUUUGUCUUUUCCAUAAAACCAUCAAUUAUUUGGCGCUAUCCA UCCUGAGUUCUA
*Petunia*_(*Pax*)_*MIR*-0390b(v2)	AAGCUCAGGAGGGAUAGCGCCAUGAACAGAUUUCUGUUGUUGGCAAUUUCUUUUUAUUUUUCCU UUGUCGAAUCUUCUUUUCGUGGGUUUUAUUUUUCCCAUAACUUACAAUUCUGUGGCGCUAUCCA UCCUGAGUUUCA
*Petunia*_(*Pin*)_*MIR*-0390b(v2)	AAGCUCAGGAGGGAUAGCGCCAUGAACAGAUUUAUGUUGUUGGGAAUUUCUUUUUAUCUUUCCC UUGUUGAAUCUUAUUUUCGUGGGUUUUAUUUUUCCCAUAGCUUACAAUUCUGUGGCGCUAUCCA UCCUGAGUUUCA
*Petunia*_(*Pax*)_*MIR*-0390c(v2)	AAGCUCAGGAGGGAUAGCGCCAUCGAUGACUACGGUACAUUAGCUUAAUAGAAAAAAAAUAUAU UUAAGCUGUGUUACGUUGGACAUCUGUAGCGCUAUCCAUCCUGAGUUUCA
*Petunia*_(*Pin*)_*MIR*-0390c(v2)	AAGCUCAGGAGGGAUAGCGCCAUGGAUGAUUAUUGUACAAAAAUGUACAUGAUGAGCUUAAUUA AAACUAUUGUUUUUAGGCUGUCUUCAUUUAACCAUCUGUAGCGCUAUCCAUCCUGAGUUUCA
*Petunia*_(*Pax*)_*MIR*-0397a(v2)	UCAUUGAGUGCAGCGUUGAUGAAAUAUUUCUUUAAUUUCUGUCAAAAUGUUGUCAUUUGGGCAU UCCCUCCAGUUGGUUUUCAUCUACGUUGCACUCAAUUAUG
*Petunia*_(*Pin*)_*MIR*-0397a(v2)	UCAUUGAGUGCAGCGUUGAUGAAAUAUUUCUUUAAUUUCAGUCAAAAUGUUGCCAUUUGGGCAU UCCCUCCAGUUAGUUUUCAUCUACGCUGCACUCAAUUAUG
*Petunia*_(*Pax*)_*MIR*-0398a(v2)	GAGUGUACCAGGGAACACAUGUGCAUUUUGGCUAAUUUGUUAAUGGUUGGAUCCAAAAUCCACU UGUGUUCUCAGGUCACCCCU
*Petunia*_(*Pin*)_*MIR*-0398a(v2)	GAGUGUACCAGGGAACACAUGUGCAUUUUGGCUUUUGUUAAUGGUUGGAUCCAAAAUCCACUUG UGUUCUCAGUUCACCCCU
*Petunia*_(*Pax*,*Pin*)_*MIR*-0398b(v2	GAGUGUUCAUGGGAACACAUGUGCAUUUUGGUUAUUGAUAAUGGCUAUUUUAUAAAUGCACUUG UGUUCUCAGGUCACCCCU
*Petunia*_(*Pax*)_*MIR*-0398h(v2)	AAUAGGGGCCACUUGAGAUCACAUGUUGAGACUUUUUCUUUUUUCAAUUUAAUUGAAUGGUGUG AUUCAAUAUGUGUUCUCAGGUCGCCCCUG
*Petunia*_(*Pin*)_*MIR*-0398h(v2)	AAUAGGGGCCACUUGAGAUCACAUGUUGCACUUUUCUUUUUUUUUCUUUUUUCCAAUUUAAUUG AAUGUGUGAUUCAAUAUGUGUUCUCAGGUCGCCCCUG
*Petunia*_(*Pax*,*Pin*)_*MIR*-2111a(v2)	GGUAAUCUACAUCCCAAGGACUAGAAAGCAGUUAAAAUAUUCUGAUCUAAACCUCAGGAUGCAG AUUA
*Petunia*_(*Pax*)_*MIR*-2111b(v2)	AGUAAUCUGCAUCCCAAGGGCUAGAAGCAGUACUAAUUAAAGUAGCCGAUCUAAACCUCAGGAU GCAGAUUA
*Petunia*_(*Pin*)_*MIR*-2111b(v2)	AGUAAUCUGCAUCCCAAGGGCUAGAAGCAGUACUAAAGAAGCCGAUCUAAACCUCAGGAUGCAG AUUA

Table 2F. Sequences of miRNA*s in *Petunia*

*Petunia*_(*Pax*,*Pin*)_miR*0156a	GCUCACCCUCUUUCUGUCACC
*Petunia*_(*Pax*,*Pin*)_miR*0156b-d	GCUCACUCUCUAUCUGUCACC
*Petunia*_(*Pax*,*Pin*)_miR*0156e	GCUCACGCUCUAUCUGUCACC
*Petunia*_(*Pax*)_miR*0156f	GCUCACUUCUCAUUCUGUCACC
*Petunia*_(*Pin*)_miR*0156f	GCUCACUUCUCUUUCUGUCACC
*Petunia*_(*Pax*,*Pin*)_miR*0156g	GCUCACUUCUCUUUCUGUCAGC
*Petunia*_(*Pax*,*Pin*)_miR*0156j	GCUUAUCCUCUGUUAGA
*Petunia*_(*Pax*,*Pin*)_miR*0156k	GCUCUCUAUUCUUCUGUCAUC
*Petunia*_(*Pax*,*Pin*)_miR*0156l	GCUCUCUCUGCUUCUGUCAAU
*Petunia*_(*Pax*,*Pin*)_miR*0157a	GCUCUCUAUGCUUCUGUCAUC
*Petunia*_(*Pax*)_miR*0157b	GCUCUUUAUUUUUCUGUCAUC
*Petunia*_(*Pin*)_miR*0157b-c	GCUCUUUAUUCUUCUGUCAUC
*Petunia*_(*Pax*)_miR*0157c	GCUCUUUAUCCUUCUGUCAUC
*Petunia*_(*Pax*,*Pin*)_miR*0157d	GCUCUCUAUGCUUCCGUCAUC
*Petunia*_(*Pax*,*Pin*)_miR*0157e	GCUCUCUAUUCUUCUGCCAUC
*Petunia*_(*Pax*,*Pin*)_miR*0159a	GAGCUCCUUGAAGUCCAACAG
*Petunia*_(*Pax*,*Pin*)_miR*0159b	GAGCUUCUUUAAGUCCAACAG
*Petunia*_(*Pax*,*Pin*)_miR*0159c	GAGCUUCUUUGAAGUCCAAAAG
*Petunia*_(*Pax*,*Pin*)_miR*0160a-b	GCGUAUGAGGAGCCAAGCAUA
*Petunia*_(*Pax*,*Pin*)_miR*0160c	GCGUGCGAGGAGCCAAGCAUA
*Petunia*_(*Pin*)_miR*0160d	GCACCAGAGGAGUCGGGCAGA
*Petunia*_(*Pax*,*Pin*)_miR*0162a	GGAGGCAGCGGUUCAUCGAUC
*Petunia*_(*Pax*,*Pin*)_miR*0164a	CACGUGUUCUCCUUCUCCAAC
*Petunia*_(*Pax*,*Pin*)_miR*0164b	CAUGUGCCUGUCUUCCCCAUC
*Petunia*_(*Pax*,*Pin*)_miR*0164c	CAUGUGCCCCUCUUCCCCAUC
*Petunia*_(*Pax*,*Pin*)_miR*0164d	CAUGUGUUCUCCUUCUCCAAC
*Petunia*_(*Pax*,*Pin*)_miR*0164g	CAUGUGCUCUUGCCCUCCAGC
*Petunia*_(*Pax*,*Pin*)_miR*0164h	CAUGUGCUCUUGCUCUCCAGC
*Petunia*_(*Pax*,*Pin*)_miR*0166a-b	GGAAUGUUGUCUGGCUCGAGG
*Petunia*_(*Pax*,*Pin*)_miR*0166c	AGAAUGUCGUCUGGUUCGAGA
*Petunia*_(*Pax*,*Pin*)_miR*0166d	GGAAUGUUACCUGGCUCGAAG
*Petunia*_(*Pax*,*Pin*)_miR*0166e-f	GGAAUGCUGUCUGGUUCGAAA
*Petunia*_(*Pax*,*Pin*)_miR*0166g	GGAGUGUUGCCUGGUUCGAAG
*Petunia*_(*Pax*)_miR*0166h	GGAAUGUCGUCUGGUUCAAGA
*Petunia*_(*Pin*)_miR*0166h	AAAAUGUCGUCUGGUUCAAGA
*Petunia*_(*Pax*,*Pin*)_miR*0166k	GGAAUGUUGGCUGGCUCGACA
*Petunia*_(*Pax*,*Pin*)_miR*0167a	GAUCAUGUGGCAGCCUCACC

*Petunia*_(*Pax*,*Pin*)_miR*0167b	GAUCAUGUGGCAGCAUCACC
*Petunia*_(*Pax*,*Pin*)_miR*0167c	GGUCAUGCCCUGACAGCCUCA
*Petunia*_(*Pax*,*Pin*)_miR*0167d	GAUCAUGUGGUAGCUUCACC
*Petunia*_(*Pax*)_miR*0167e	GGUCAUGCUCCGACAGCUUCA
*Petunia*_(*Pin*)_miR*0167e	GGUCAUGUUCCGACAGCUUCA
*Petunia*_(*Pax*,*Pin*)_miR*0168a-c	CCCGCCUUGCAUCAACUGAAU
*Petunia*_(*Pax*,*Pin*)_miR*0169a	GGCAAGUGGUCCUUGGCUACA
*Petunia*_(*Pax*)_miR*0169b	GGCAAGUCGUCCUUGGCUACA
*Petunia*_(*Pin*)_miR*0169b	GGCAGGUCGUCCUUGGCUACA
*Petunia*_(*Pax*,*Pin*)_miR*0169c(BL)	GGCAAGUCAUUUUUGGCUACA
*Petunia*_(*Pax*)_miR*0169d	GGCAAGUUGUUCCUGGCUACA
*Petunia*_(*Pin*)_miR*0169d	GGCAAGUUGUUCUUGGCUACA
*Petunia*_(*Pax*,*Pin*)_miR*0169g	GCAGUCUCCUUGGCUACU
*Petunia*_(*Pax*,*Pin*)_miR*0169h	GCAGUCUCUCCUUUGGCUUUC
*Petunia*_(*Pax*,*Pin*)_miR*0169i	GCAGUCUCCUUGGCUAAU
*Petunia*_(*Pax*,*Pin*)_miR*0169j	GGCAGUCUCCUUGACUACC
*Petunia*_(*Pax*)_miR*0169k	GCAUGUCAUCUUGGCUAGC
*Petunia*_(*Pin*)_miR*0169k	GCAUGUCAUCUUGGCUAAC
*Petunia*_(*Pax*,*Pin*)_miR*0169l	GGCAGUCUCCUUGGCUACC
*Petunia*_(*Pax*,*Pin*)_miR*0169m	GGCAGGUCAUCUUUAGCUAAC
*Petunia*_(*Pax*,*Pin*)_miR*0169n	GGCAGGUCAUCCUAGCUAAC
*Petunia*_(*Pax*,*Pin*)_miR*0169o	GCAAGCAUCUGAGGCGACU
*Petunia*_(*Pax*,*Pin*)_miR*0169p	GGCAGUCAUCCAUGGUUAUG
*Petunia*_(*Pax*,*Pin*)_miR*0169q	GGCAGUCGUCUUUGGCUAUA
*Petunia*_(*Pax*,*Pin*)_miR*0169r	GGCAGUCGUCUUUGGCUACA
*Petunia*_(*Pax*)_miR*0169s	GGCGGUGAUCCGAGGUUACC
*Petunia*_(*Pax*,*Pin*)_miR*0169t	GGCGUCAUCCCAGGCUAUU
*Petunia*_(*Pax*,*Pin*)_miR*0169u	GGCAGGCAUCCUGGCUAUA
*Petunia*_(*Pax*)_miR*0169v	GGCGAGUUAUCCUGGCUAUA
*Petunia*_(*Pin*)_miR*0169v	GGCAAGUUAUCCUGGCUAUA
*Petunia*_(*Pax*)_miR*0169w	GGUGUCAUCCUUGGAUAAC
*Petunia*_(*Pin*)_miR*0169w	GGUGUCAUCCAUGGGUAAC
*Petunia*_(*Pin*)_miR*0169x	GGCAAGUUAUCCUGGCUAUA
*Petunia*_(*Pax*,*Pin*)_miR*0171a	UGUUGGUGCGGUUCAAUGAGA
*Petunia*_(*Pax*,*Pin*)_miR*0171b	UAUUGGUGCGGUUCAAUGAGA
*Petunia*_(*Pax*)_miR*0171c	UAUUGAUGCGGUUCAAUUAGA
*Petunia*_(*Pin*)_miR*0171c	UAUUGGUGCGGUUCAAUUAGA
*Petunia*_(*Pax*,*Pin*)_miR*0171d	UAUUGGCCUGGUUCACUCAGA
*Petunia*_(*Pax*,*Pin*)_miR*0171e	UGUUGGAAUGGCUCAAUCAAA

*Petunia*_(*Pax*)_miR*0171f	UAUUGGCCAGGUUCACUCAGA
*Petunia*_(*Pin*)_miR*0171f	UAUUGGCCAGGUUCACUCAAA
*Petunia*_(*Pax*)_miR*0171g	UAUUGGUGAGGUUCAAUUAGA
*Petunia*_(*Pax*,*Pin*)_miR*0171j-k	CGAUGUUGGUGAGGUUCAAUC
*Petunia*_(*Pax*,*Pin*)_miR*0171n	AGAUGUUGAUGCGACUCAAUC
*Petunia*_(*Pax*,*Pin*)_miR*0171o	AGAUAUUGAUGAGGCUCAAUC
*Petunia*_(*Pax*,*Pin*)_miR*0172a	GCAGCACCAUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172b-c	GUAGCAUCAUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172d	GCAGCAUUAUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172e	GCAUCAUCAUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172f	GCAGCAUUAUUAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172g	GCAGCAUCCUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172j	GGAGCAUCAUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172k	GCAGCAUCUUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0319a	AGAGCUUUCUUCAGUCCACUC
*Petunia*_(*Pax*,*Pin*)_miR*0319b	AGAGCUUUCUUCAGUCCACAU
*Petunia*_(*Pax*,*Pin*)_miR*0319c	GGAGCUCUCUCCAGUCCAGUC
*Petunia*_(*Pax*,*Pin*)_miR*0319d	AGAGCUUCCUUUAGUCCACUC
*Petunia*_(*Pax*,*Pin*)_miR*0319e	AGAGCUUCCUUCAGUCCACUC
*Petunia*_(*Pax*,*Pin*)_miR*0319h	GAGCUUCCUUCAGCCCACUC
*Petunia*_(*Pax*,*Pin*)_miR*0319i	(G)AGCUCCUUUCAGGCCAAGA
*Petunia*_(*Pax*,*Pin*)_miR*0390a	CGCUAUCCAUCCUGAGUUCUA
*Petunia*_(*Pax*,*Pin*)_miR*0390b-c	CGCUAUCCAUCCUGAGUUUCA
*Petunia*_(*Pax*,*Pin*)_miR*0393a-b	AUCAUGCUAUCCCUUUGGACU
*Petunia*_(*Pax*,*Pin*)_miR*0393c	AUCAUGCGAUCUCUUCGGAAU
*Petunia*_(*Pax*,*Pin*)_miR*0394a	AGGGGGCCAAAGUGCCAAAC
*Petunia*_(*Pax*,*Pin*)_miR*0394b-c	AGGUGGGCAUACUGCCAACA
*Petunia*_(*Pax*,*Pin*)_miR*0395a	GUUCUCUCAAAUCACUUCAUUG
*Petunia*_(*Pax*,*Pin*)_miR*0395b-e	GUUCUCCUGAUCACUUCAUUG
*Petunia*_(*Pin*)_miR*0395b-e2	GUUCUCCUGAUCACUUCAUUA
*Petunia*_(*Pax*)_miR*0395f	GUUCUCCUGAUCACGUCAUUG
*Petunia*_(*Pax*,*Pin*)_miR*0395g-h	GUUUCCCUGACCGCUUCAUGA
*Petunia*_(*Pax*)_miR*0395i-j	GUUCCCUUAAAUGCUUCAUGA
*Petunia*_(*Pin*)_miR*0395j	GUUCCCCUAAAUGCUUCAUGA
*Petunia*_(*Pax*)_miR*0395k-l	GUUCCCUUGACCACUUCAUGA
*Petunia*_(*Pin*)_miR*0395k	GUUCCCUUGAUCACUUCAUGA
*Petunia*_(*Pin*)_miR*0395m	GUUCUCCUCAGCACUUCAUUG
*Petunia*_(*Pax*,*Pin*)_miR*0396a	GUUCAAUAAAGCUGUGGGAAG
*Petunia*_(*Pax*,*Pin*)_miR*0396b	GUUCAAUAUAGCUGUGGGAAG

*Petunia*_(*Pax*,*Pin*)_miR*0396c	GUCCAAGAAAGCUGUGGGAAA
*Petunia*_(*Pax*,*Pin*)_miR*0396d	GUUCAAGAAAGCUGUGGGAAA
*Petunia*_(*Pax*)_miR*0397a	AUCUACGUUGCACUCAAUUA
*Petunia*_(*Pin*)_miR*0397a	AUCUACGCUGCACUCAAUUA
*Petunia*_(*Pin*)_miR*0397b	AUCUACGUUGCACUCAAUCA
*Petunia*_(*Pax*,*Pin*)_miR*0398a	GGAGUGUACCAGGGAACACA
*Petunia*_(*Pax*)_miR*0398b	GGAGUGUUCAUGGGAACACA
*Petunia*_(*Pax*,*Pin*)_miR*0398e	GGGACGACUUGAGAUCAUAUG
*Petunia*_(*Pax*,*Pin*)_miR*0398h	UGAUCUCAAGUGGCCCCUAUU
*Petunia*_(*Pax*,*Pin*)_miR*0399a	GGGCUUCUCUCUAUUGGCAUG
*Petunia*_(*Pax*,*Pin*)_miR*0399b	GGGAUACUCUCUAUUGGCAUG
*Petunia*_(*Pax*,*Pin*)_miR*0399c	GGGCUACUCUCUAUUGGCAUG
*Petunia*_(*Pax*,*Pin*)_miR*0399d	CGCCAAAGGAGAGCUGCCCUG
*Petunia*_(*Pax*,*Pin*)_miR*0399e-g	UGCCAAAGGAGAGCUGCCCUG
*Petunia*_(*Pax*,*Pin*)_miR*0403a	CGUUUGUGCGUGAAUCUAACA
*Petunia*_(*Pax*,*Pin*)_miR*0408a	ACAGGGACGAGACAGAGCAUG
*Petunia*_(*Pax*)_miR*0477a	GAAACUCUGGCAGGGAGAGCCA
*Petunia*_(*Pax*,*Pin*)_miR*0479a	GUGAUAUUGGGUUGGCUCAUU
*Petunia*_(*Pax*,*Pin*)_miR*0482a	GGGAUUGGUGGGUUGGAAAGC
*Petunia*_(*Pax*,*Pin*)_miR*0827a	UUUGUUGAUGGUCAUCUAGCU
*Petunia*_(*Pax*,*Pin*)_miR*2111a	GUCCUUGGGAUGUAGAUUACC
*Petunia*_(*Pax*,*Pin*)_miR*2111b	GCCCUUGGGAUGCAGAUUACU
*Petunia*_(*Pax*,*Pin*)_miR*6149a	CGAUUUAGGUUCGUAUUUAAA
*Petunia*_(*Pax*,*Pin*)_miR*8016a	CAUGGUCUUUUCUUUCAAAAAUAU
*Petunia*_(*Pax*,*Pin*)_miR*0157a(v2)	GCUCUCUAUGCUUCUGUCAUCA
*Petunia*_(*Pax*)_miR*0157b(v2)	GCUCUUUAUUUUUCUGUCAUCA
*Petunia*_(*Pin*)_miR*0157b-c(v2)	GCUCUUUAUUCUUCUGUCAUCA
*Petunia*_(*Pax*)_miR*0157c(v2)	GCUCUUUAUCCUUCUGUCAUCA
*Petunia*_(*Pax*,*Pin*)_miR*0157d(v2)	GCUCUCUAUGCUUCCGUCAUCA
*Petunia*_(*Pax*,*Pin*)_miR*0157e(v2)	GCUCUCUAUUCUUCUGCCAUCA
*Petunia*_(*Pax*,*Pin*)_miR*0171a(v2)	AGAUGUUGGUGCGGUUCAAUG
*Petunia*_(*Pax*,*Pin*)_miR*0171b(v2)	AGAUAUUGGUGCGGUUCAAUG
*Petunia*_(*Pax*)_miR*0171c(v2)	AGAUAUUGAUGCGGUUCAAUU
*Petunia*_(*Pin*)_miR*0171c(v2)	AGAUAUUGGUGCGGUUCAAUU
*Petunia*_(*Pax*,*Pin*)_miR*0171j-k(v2)	GAUGUUGGUGAGGUUCAAUC
*Petunia*_(*Pax*,*Pin*)_miR*0171j(v3)	UGUUGGUGAGGUUCAAUCUGA
*Petunia*_(*Pax*,*Pin*)_miR*0172a(v2)	AGCACCAUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172b-c(v2)	AGCAUCAUCAAGAUUCACA
*Petunia*_(*Pax*)_miR*0172d(v2)	AGCAUUAUCAAGAUUCACA

*Petunia*_(*Pin*)_miR*0172d(v2)	AGCAUCAUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172e(v2)	AUCAUCAUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172f(v2)	AGCAUUAUUAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172g(v2)	AGCAUCCUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172j(v2)	GCAUCAUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0172k(v2)	GCAUCUUCAAGAUUCACA
*Petunia*_(*Pax*,*Pin*)_miR*0319a(v2)	AGCUUUCUUCAGUCCACUCC
*Petunia*_(*Pax*,*Pin*)_miR*0319b(v2)	AGCUUUCUUCAGUCCACAUA
*Petunia*_(*Pax*,*Pin*)_miR*0319c(v2)	AGCUCUCUCCAGUCCAGUCC
*Petunia*_(*Pax*,*Pin*)_miR*0319d(v2)	AGCUUCCUUUAGUCCACUCC
*Petunia*_(*Pax*,*Pin*)_miR*0319e(v2)	AGCUUCCUUCAGUCCACUCA
*Petunia*_(*Pax*,*Pin*)_miR*0319h(v2)	AGCUUCCUUCAGCCCACUCA
*Petunia*_(*Pax*,*Pin*)_miR*0319h(v3)	AGAGCUUCCUUCAGCCCACUC
*Petunia*_(*Pax*,*Pin*)_miR*0390a(v2)	GCUAUCCAUCCUGAGUUCUA
*Petunia*_(*Pax*,*Pin*)_miR*0390b(v2)	GCUAUCCAUCCUGAGUUUCA
*Petunia*_(*Pax*,*Pin*)_miR*0390c(v2)	GCUAUCCAUCCUGAGUUUCA
*Petunia*_(*Pax*,*Pin*)_miR*0397a(v2)	UCUACGCUGCACUCAAUUAUG
*Petunia*_(*Pin*)_miR*0397b(v2)	UCUACGUUGCACUCAAUCAUG
*Petunia*_(*Pax*,*Pin*)_miR*0398a(v2)	GAGUGUACCAGGGAACACAUGU
*Petunia*_(*Pax*)_miR*0398b(v2)	GAGUGUUCAUGGGAACACAUGU
*Petunia*_(*Pax*,*Pin*)_miR*0398h(v2)	AAUAGGGGCCACUUGAGAUCA
*Petunia*_(*Pax*,*Pin*)_miR*2111a(v2)	GGUAAUCUACAUCCCAAGGAC
*Petunia*_(*Pax*,*Pin*)_miR*2111b(v2)	AGUAAUCUGCAUCCCAAGGGC

Table 4. MiRNA frequencies in *P. axillaris* and *P. inflata* flower buds *versus* tomato and potato

	Sum of all Variants				
	Tomato [1]	Potato [1]	Petunia	Petunia	
miRNA family			*Pax* All Average	*Pin* All Average	miR Family
156_S01			204	642	miR-0156/7
159_S01			3137	2847	miR-0159
160_S01			458	291	miR-0160
162_S01			150	211	miR-0162
164_S01			123	89	miR-0164
166_S01			4880	6188	miR-0166
167_S01			118	125	miR-0167
168_S01			353	451	miR-0168
169_S01			6	4	miR-0169
171_S01			164	128	miR-0171
172_S01			162	99	miR-0172
319_S01			2153	2592	miR-0319
390_S01			236	276	miR-0390
393_S01			16	16	miR-0393
394_S01			642	553	miR-0394
395_S01			3	2	miR-0395
396_S01			376	339	miR-0396
397_S01			1	0	miR-0397
398_S01			5	0	miR-0398
399_S01			1	0	miR-0399
403_S01			491	463	miR-0403
408_S01			4	0	miR-0408
477_S01			1	0	miR-0477
479_S01			1	0	miR-0479
482_S01			106	49	miR-0482
827_S01			134	19	miR-0827
2111_S02			1	0	miR-2111
			67	162	miR-6149
			15	13	miR-8016
845_S01					
858_S01					
894_S01					
1511_S01					
1863_S01					
1916_S01					
1919_S01					
2089_S01					
2118_S01					
2911_S01					
2916_S01					
3627_S01					
4371_S01					
4376_S01					
4414_S01					
5021_S01					
5054_S01					
5059_S01					
5072_S01					
5077_S01					
5139_S01					
5300_S01					
5301_S01					
5303_S01					
5304	not tested	not tested			

10,001 - 100,000
1001 - 10,000
101 - 1000
11 - 100
1-20

[1] **MiRNA frequencies** in Tomato a and Potato buds are from the Tomato Genome Consertium (2012)

Supplementary Note 10

Genomic insight into the pathways that control adventitious root formation and arbuscular mycorrhiza in petunia

Philipp Franken[1], Uwe Druege[1], Marcel Bucher[2], Lorenzo Borghi[3], Enrico Martinoia[3], Laure Bapaume[4], Mélanie K. Rich[4], Eva Nouri[4], and Didier Reinhardt[4]

[1]Dept. of Plant Propagation, Leibniz Institute of Vegetable and Ornamental Crops (IGZ), Kühnhäuserstr. 101, 99090 Erfurt, Germany

[2]Cologne Biocenter, Cluster of Excellence on Plant Sciences (CEPLAS), University of Cologne, Zuelpicher Straße 47b, 50674 Cologne, Germany

[3]Institute of Plant and Microbiology, University of Zürich, Zollikerstr. 107, CH-8008 Zürich, Switzerland

[4]Dept. of Biology, University of Fribourg, Chemin du Musée 10, CH-1700 Fribourg, Switzerland

Abstract

The roots of plants are optimized for nutrient and water uptake, symbiotic interactions, and anchoring in the soil. In addition to the basic developmental programs that govern root growth, plants have different forms of root specialization which represent important adaptive traits. A central trait, which is influenced by complex interactions between plant genotype and environmental conditions, is root system architecture. It is influenced by both, the endogenous programs that regulate root branching and root growth, as well as by exogenous cues such as nutrient availability and interactions with mycorrhizal fungi. Adventitious rooting, i.e. the production of roots from aerial organs, is an additional adaptive phenomenon, in particular in clonally growing plants, that represents an important trait for vegetative propagation of agriculturally important crops. Using the genome sequences of two representative petunia species, *Petunia axillaris* and *P. inflata*, the parents of most *P. hybrida* lines, we systematically analyzed several root-related pathways in the context of adventitious rooting, strigolactone biology, arbuscular mycorrhiza, and phosphate starvation. Phylogenomic and gene expression analysis provides evidence for an involvement of transcription factors of the APETALA2-like and GRAS family of transcriptional regulators in adventitious rooting. Furthermore, we systematically explore the genes involved in symbiosis signalling including strigolactone biosynthesis and transport, the LysM receptor-like kinase family, and the common symbiosis signalling pathway. Our data give new insights into the evolution and regulation of these complex regulatory pathways, and provide new targets for future functional analysis by reverse genetic approaches.

Abbreviations:

AM	Arbuscular mycorrhiza
SL	Strigolactone
RNS	Root nodule symbiosis
AR	Adventitious root
CSSP	Common symbiosis signaling pathway
LCO	Lipochitooligosaccharide
GRAS	GAI, RGA, SCR

Introduction

Plant productivity is often limited by nutrient and water uptake (Hell and Hillebrand, 2001; Elser et al., 2007). In order to improve their nutrient and water supply, plants can either change their root architecture, or they can engage in symbiotic interactions such as arbuscular mycorrhiza (AM) with soil fungi (*Glomeromycota*), and root nodule symbiosis (RNS) with bacterial symbionts (rhizobia). In order to orchestrate these distinct nutritional strategies, plants have two regulatory programs: the basic developmental program for the generation, expansion and function of the root system, and the program for accommodation of symbiotic microbes which is based on bi-directional exchange of signals and nutrients. Although these programs may appear to be complementary, they compete for the same resources, and therefore, plants have evolved regulatory mechanisms to control resource allocation to the sites where new organs are formed. For example, a regulatory pathway involving the receptor kinase HAR1 restricts the number of nodules in RNS. *Har1* mutants form excessively abundant nodules, resulting in an inefficient and underdeveloped root system (Krusell et al., 2002; Nishimura et al., 2002). Hence, the regulatory mechanisms in root development and symbiosis are of central importance for plant fitness and survival.

Roots can be formed in different contexts. After germination and outgrowth of the primary root, which is of embryonic origin, lateral root formation contributes to the expansion of the root system. Lateral roots are initiated by auxin in the pericycle, a process that has been extensively studied in *Arabidopsis thaliana* (Lavenus et al., 2013; Van Norman et al., 2013). In addition to root branching, plants can form adventitious roots (AR) from aerial tissues in stems or leaves. AR formation is frequently observed in detached or excised shoot segments as a survival strategy (da Costa et al., 2013), a process that is utilized in vegetative plant propagation through cuttings. However, despite its enormous economic importance in the production of ornamental plants (Druege, 2009), AR formation has been much less investigated than lateral root and primary root development.

Formation of ARs has been observed and studied in pre-etiolated intact seedlings of *A. thaliana* (Sorin et al., 2005; Gutierrez et al., 2012), where they originate from pericycle cells in the hypocotyl. This process resembles lateral root formation in roots as opposed to AR formation in excised shoot tissues (Correa et al., 2012; da Costa et al., 2013). *Petunia hybrida* has in recent years been established as a convenient model to study AR formation from mature stem tissues at the molecular and physiological level (Ahkami et al., 2009; Klopotek et al., 2010). AR development in *Petunia* is induced at the bases of shoot cuttings by local accumulation of the auxin indole-3-acetic acid, which is mediated by basipetal polar auxin transport (Ahkami et al., 2013). This highlights commonalities in the formation of lateral and adventitious roots (Bellini et al., 2014). However, in contrast to lateral root formation in intact plants, AR formation in cuttings depends on the establishment of a strong sink that attracts carbon flux to the new rooting zone, thereby providing the resources to establish a new root system. In *Petunia* cuttings, the establishment of a new sink is accompanied by the activation of cell wall invertases (Ahkami et al., 2009), a phenomenon that depends on polar auxin transport and auxin accumulation (Ahkami et al., 2013) while a high photosynthetic activity of the leaves provides carbon for root formation (Klopotek et al. 2012). Although auxin has been identified as an important regulatory signal, the molecular-genetic control of AR formation remains poorly understood (Ludwig-Müller, 2009; Pop et al., 2011; da Costa et al., 2013).

Formation of new root meristems can involve re-programming of differentiated cells in the pericycle (in the case of lateral root formation), or in the vascular tissue (in the case of AR formation), or start directly with division of meristematic cambial cells (in the case of AR formation). The meristem of primary root and lateral root consists of the quiescent center (QC), and the surrounding stem cells that produce the characteristic cell types of the root, i.e. epidermis, cortex, endodermis, pericycle, and the central vascular stele (Jiang and

Feldman, 2005; Perilli et al., 2012; Petricka et al., 2012). During the establishment of the embryonic root, the differentiation and patterning of these cell types is controlled by a regulatory network of transcription factors including the GRAS-type transcription factors SHORT ROOT (SHR) and SCARECROW (SCR) (Sparks et al., 2013), and APETALA 2 (AP2) domain transcription factors such as PLETHORA (PLT) (Horstman et al., 2014). Hence, the question arises to which degree these developmental factors are also involved in AR formation.

Besides their basic root developmental program, plants have evolved a dedicated signaling pathway for the establishment of arbuscular mycorrhiza (AM), symbiotic associations with soil fungi which can improve their nutritional status (Smith and Read, 2008). *Petunia* has become an important model for genetic and transcriptomic analysis of this mutualistic interaction (Wegmüller et al., 2008; Breuillin et al., 2010; Feddermann et al., 2010; Feddermann and Reinhardt, 2011; Kretzschmar et al., 2012; Nouri et al., 2014). The first known element in AM interaction is the root-borne signal strigolactone (SL), which stimulates the metabolism and hyphal branching of AM fungi in the vicinity of the host (Gutjahr and Parniske, 2013). In addition, SL has been identified as a central hormone in shoot development (Gomez-Roldan et al., 2008; Umehara et al., 2008), with conserved functions from mosses to angiosperms (Delaux et al., 2012; Ruyter-Spira et al., 2013). Key factors in SL biosynthesis (Snowden et al., 2005), SL sensing (Hamiaux et al., 2012), and SL transport (Kretzschmar et al., 2012) have been identified and characterized in *Petunia*, which therefore represents one of the leading model species in the elucidation of SL biology (Ruyter-Spira et al., 2013; Seto and Yamaguchi, 2014; Al-Babili and Bouwmeester, 2015).The first fungal signals in the AM interaction consist of N-acetyl glucosamine oligomers (Genre et al., 2013), that can carry lipidic side chains (lipochitooligosaccharides, LCOs), and further decorations such as sulfate or acetyl groups (Maillet et al., 2011). RNS involves similar LCOs that are collectively referred to as nod factors because of their central role in nodulation (Dénarié et al., 1996). In RNS, the bacterial nod factors are perceived by nod factor receptors (NFRs), that belong to the group of lysine motif receptor-like kinases (LysM-RLKs) (Madsen et al., 2003; Radutoiu et al., 2003). In analogy to the closely related chitin receptors (Liu et al., 2012; Hayafune et al., 2014), the NFRs are thought to mediate nod factor perception as heterodimers (Gough and Cullimore, 2011; Gust et al., 2012). LjNFR1 and LjNFR5, as their respective orthologues MtLyk3 and MtNFP in *M. truncatula*, belong to two separate LysM-RLK subfamilies. Interestingly, close relatives of LjNFR1/MtLyk3 in rice (*Oryza sativa*) and *A. thaliana* function as receptors for chitin oligosaccharides (CO), that play a role in the recognition of pathogens (Kaku et al., 2006; Miya et al., 2007; Petutschnig et al., 2010; Shimizu et al., 2010; Shinya et al., 2012; Wan et al., 2012). Recent evidence suggests that some LysM-RLKs could be involved in both RNS and AM (Op den Camp et al., 2011). Moreover, some LysM-RLKs recognize both AM fungal signals as well as bare chitin oligomers (Miyata et al., 2014; Zhang et al., 2015). It remains to be seen whether this apparent promiscuity reflects the involvement of heterodimeric receptor complexes with overlapping ligand specificities.

Mutual recognition of the symbiotic partners in AM results in the establishment of cellular compatibility and the formation of the symbiotic machinery. Most information about these processes comes from the legumes *Lotus japonicus* and *Medicago truncatula*, which also engage in root nodule symbiosis (RNS) with rhizobia (Parniske, 2008). Genetic analysis in these species has revealed a series of symbiosis-related genes that are required for both, AM and RNS, and which therefore are referred to as common symbiosis (SYM) genes. Together they constitute the common SYM signaling pathway (CSSP) (Harrison, 2012; Gutjahr and Parniske, 2013; Oldroyd, 2013). Interestingly, the establishment of AM also involves regulatory transcription factors of the GRAS-type family that are related to the regulators of basic root development (see above) (Gutjahr and Parniske, 2013).

Here we investigate a set of components of the regulatory network underlying cell fate decisions and root specialization in the context of AR and AM formation, and we compare them with the orthologues in other model species. In particular, we analyze the genes involved in SL biology, all members of the LysM-RLK family, the componenets of the CSSP, and finally the complete families of transcription factors in the AP2-like and GRAS families. The predicted proteins were compared with members of the corresponding protein families in other species in order to analyze the substructures of the families and their phylogeny. We further screened an EST library for the cDNA fragments belonging to the transcription factor genes and collected all expression data obtained so far by array hybridization (Breuillin et al., 2010; Ahkami et al., 2014). Taken together, the combination of genomics and transcriptomics identifies targets for future studies, such as putative regulators of AR and AM formation, which will be subject to further evolutionary genomics and functional analysis by transposon insertion mutagenesis. These studies will provide new insight into the basis of these complex root traits**.**

Results

Genome-wide analysis of the AP2- and GRAS families of transcription factors

Based on BLAST searches among the virtual transcripts of the *P. axillaris* genome 109 APETALA2 (AP2)-like and 69 GRAS transcription factors (TFs), respectively, were discovered (**Tables S1 and S2, available online at** http://www.igzev.de/projects/Genome-petunia-roots/; user: igz; pw: RHizoGEnPEt). These sequences were compared with 24'816 unique sequences in an EST database of *P. hybrida* W115 (also known as Mitchell diploid) (Breuillin et al., 2010), resulting in a set of 42 AP2-like and 21 GRAS genes, that are represented on a *P. hybrida* W115 microarray (Breuillin et al., 2010). The predicted intron-exon structure of these 63 genes was confirmed by manual curation in the PaxiN genome, and their respective orthologues were identified in the PinfS6 genome. In most cases, the *P. hybrida* W115 EST sequences were related more closely to the orthologues in the PaxiN genome than to the respective orthologues of the second parent PinfS6 (**Table 1**), indicating that the breeding of *P. hybrida* W115 involved several backcrosses to *P. axillaris*.

Phylogenetic analysis at the amino acid level with a representative set of AP2-like proteins from *A. thaliana* (Dietz et al., 2010) showed that the 42 *Petunia* AP2-like proteins comprise members of all major clades of AP2-related proteins (**Fig. 1**). Similarly, the phylogenetic comparison with all 32 predicted *A. thaliana* GRAS TFs (Tian et al., 2004) revealed a good coverage of the 21 *Petunia* homologues over the entire tree of GRAS proteins (**Fig. 2**). A number of *Petunia* AP2 genes represented close homologues of known developmental regulators such as PLETHORA (PLT) (Xu et al., 2006), SHINE1 (SHN1) (Shi et al., 2011), AINTEGUMENTA (ANT) (Mizukami and Fischer, 2000) and WRINKLED1 (WRI1) (To et al., 2012). Similarly, several predicted *Petunia* GRAS-like transcription factors represented homologues of well-known developmental regulators such as SCARECROW (SCR) (DiLaurenzio et al., 1996), SHORT ROOT (SHR) (Helariutta et al., 2000), HAIRY MERISTEM (HAM) (Stuurman et al., 2002), GIBBERLIC ACID INSENSITIVE (GAI) (Peng et al., 1997), and REPRESSOR of GAI-LIKE (RGL) (Tyler et al., 2004).

Introducing in the phylogenetic analysis GRAS proteins with a known function in root symbioses, such as MtRAM1 (Gobbato et al., 2012), LjRAD1 (Xue et al., 2015), MtDELLA1 and MtDELLA2 (Floss et al., 2013), MtNSP, MtNSP1 (Smit et al., 2005), MtNSP2 (Oldroyd and Long, 2003; Kalo et al., 2005), the rice genes OsSLR1 and OsDIP1 (Itoh et al., 2002; Yu et al., 2014), and two AM-related GRAS genes from *L. japonicus* (Xue et al., 2015) revealed that

Table 1. Similarity of *P. hybrida* EST sequences with the respective orthologues in PaxiN and PinfS6.

Apetala2-like				**GRAS-like**		
ESTs	**PaxiN**	**PinfS6**		**ESTs**	**PaxiN**	**PinfS6**
cn1423	99.3*	95.2		cn3308	99.7	99.2
cn2123	98.8	98.1		cn3635	98.7	99.7
cn3335	99.6	99.6		GO_dr001P0019E06_F_ab1	99.8	99.3
cn456	99.8	98.3		cn5145	97.7	96.5
cn4574	99.7	98.6		cn7628	99.4	98.9
cn4615	98.8	97.4		cn7986	99.6	97.2
cn4698	99.1	98.3		DC242984_1	99.7	97.5
cn5374	80.1	80.5		GO_drpoolB-CL2239Contig1	100.0	99.7
cn594	99.2	96.4		DY396009_1	99.0	96.9
cn7109	100.0	98.9		GO_dr004P0023M03_F_ab1	99.7	98.4
cn7169	99.5	96.0		GO_drpoolB-CL2831Contig1	99.7	98.0
cn7885	99.1	97.4		GO_drpoolB-CL3877Contig1	99.7	98.9
cn8462	100.0	98.8		GO_drpoolB-CL4978Contig1	99.5	96.8
cn9604	98.6	96.1		GO_drpoolB-CL8772Contig1	100.0	95.3
DC240070_1	100.0	100.0		GO_drpoolB-CL9151Contig1	94.5	93.5
GI_NP1240021	99.1	95.9		GO_drs12P0009O19_F_ab1	98.7	98.3
GO_dr001P0004B23_F_ab1	99.8	96.4		GO_drs13P0003F13_R_ab1	99.7	99.2
GO_dr001P0005O12_F_ab1	97.6	85.5		IP_PHBS004A21u	98.8	98.2
GO_dr001P0010M10_F_ab1	99.8	98.0		IP_PHBS005D08u	99.5	99.0
GO_dr001P0011E21_F_ab1	98.5	97.6		IP_PHBS012D08u	100.0	99.0
GO_dr001P0015C16_F_ab1	99.8	95.5		SG_SGN-U211047	98.2	99.0
GO_dr001P0017B08_F_ab1	99.4	94.6				
GO_dr004P0008C18_F_ab1	99.7	97.9		**Mean**	**99.1**	**98.0**
GO_dr004P0020P19_F_ab1	98.5	96.4				
GO_dr004P0024N21_F_ab1	99.6	95.4				
GO_dr004P0029M08_F_ab1	100.0	99.6				
GO_drpoolB-CL1692Contig1	99.6	98.5				
GO_drpoolB-CL2213Contig1	99.9	99.1				
GO_drpoolB-CL336Contig1	99.1	96.8				
GO_drpoolB-CL3796Contig1	98.5	94.4				
GO_drpoolB-CL4794Contig1	96.0	93.2				
GO_drpoolB-CL4842Contig1	99.8	98.3				
GO_drpoolB-CL6307Contig1	97.5	96.8				
GO_drpoolB-CL6822Contig1	99.0	98.8				
GO_drpoolB-CL7001Contig1	100.0	96.1				
GO_drpoolB-CL7079Contig1	99.5	92.1				
GO_drpoolB-CL724Contig1	96.5	97.0				
GO_drpoolB-CL9655Contig1	99.0	99.7				
GO_drs12P0020L05_F_ab1	99.4	89.8				
SG_SGN-U210374	94.2	92.2				
Mean	**98.5**	**96.1**				

*Similarity is expressed as % identity at the nucleotide level between the open reading frames of *P. hybrida* vs. PaxiN and PinfS6, respectively.

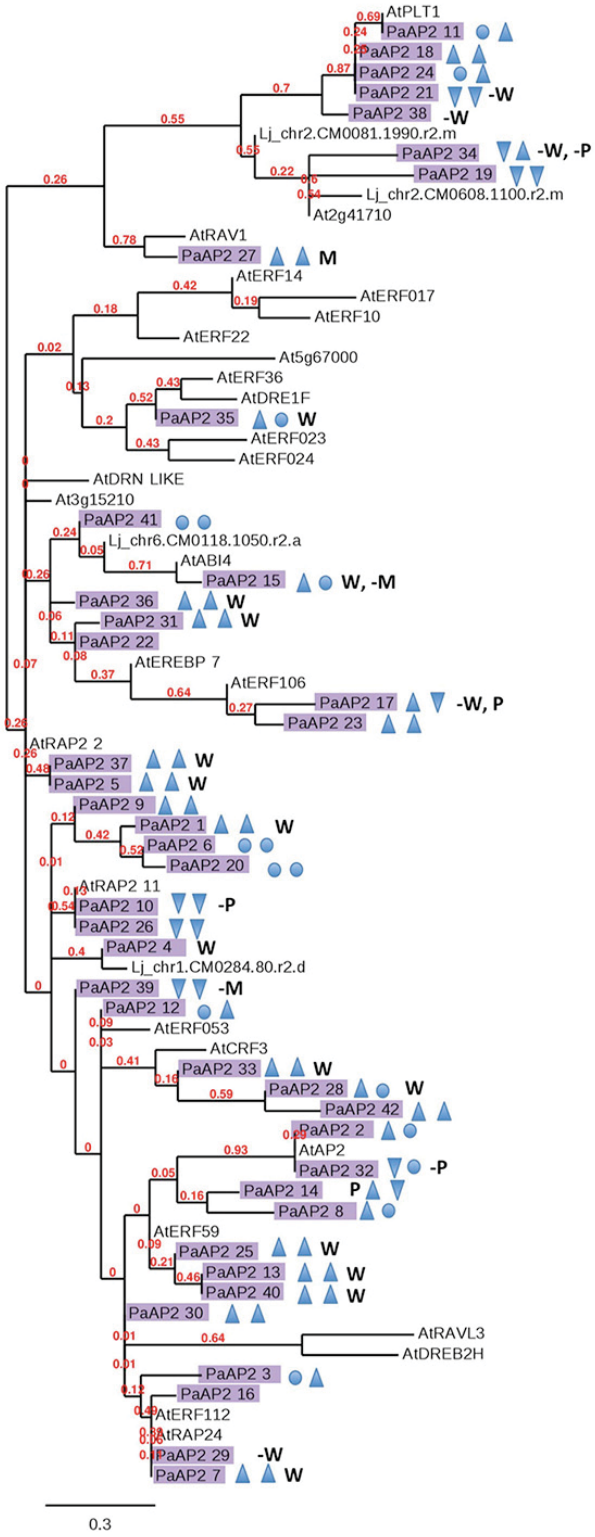


Figure 1. Phylogenetic analysis of the AP2-like family of transcription factors.
Amino acid sequences of 42 *Petunia* AP2-like sequences that are represented on the microarray were clustered with a selected set of 30 AP2-like sequences from *Arabidopsis*. *Petunia* sequences are highlighted in purple, blue signs behind gene names indicate early and late regulation, respectively, in AR formation (induced: arrowhead up, repressed: arrowhead down, unaltered: circle), as well as by wounding (W, -W), in mycorrhizal roots (M, -M), and by phosphate depletion (P, -P). Alignment and tree construction was carried out as described (Dereeper et al., 2008), with the advanced settings. Bootstrap values represent 100 replicates.

Petunia harbors close homologues of all these sequences. The closest petunia homologue of MtRAM1 (PaGRAS46) has recently been identified in a genetic screen as ATYPICAL ARBUSCULE (ATA), an essential factor in AM symbiosis and a central regulator of symbiotic gene expression (Rich et al., 2015). Interestingly, some clades lacked an *Arabidopsis* homologue (**Fig. 2**), suggesting that these clades represent TFs with a specific role in AM symbiosis (Delaux et al., 2014; Favre et al., 2014; Xue et al., 2015). Unfortunately, the *Petunia* orthologues of some of these AM-related GRAS genes are not represented on the microarray (**Table S2, Fig. 2**).

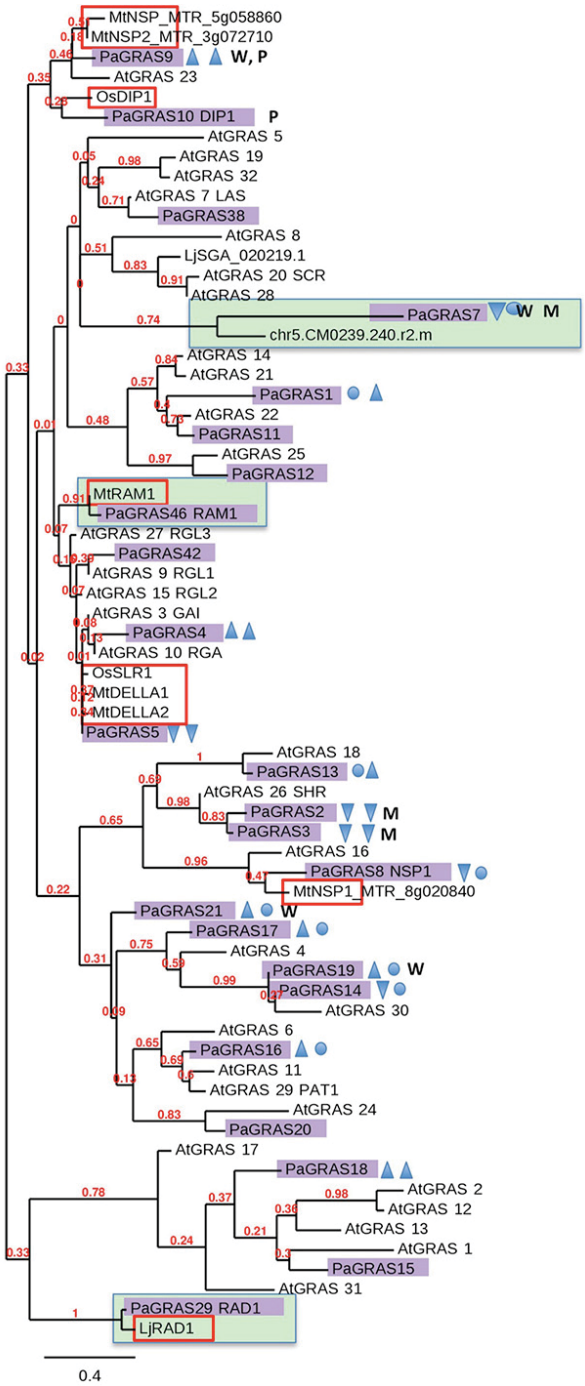


Figure 2. Phylogenetic analysis of the GRAS family of transcription factors.

Amino acid sequences of 24 *Petunia* GRAS genes, including 21 that are represented on the microarray, were clustered with all 32 GRAS genes from *Arabidopsis*, and with selected symbiosis-related genes from *M. truncatula* (Mt), *L. japonicus* (Lj), and rice (Os). *Petunia* sequences are highlighted in purple, green blocks indicate branches that lack *Arabidopsis* homologues (bootstrap support >0.9), blue signs behind gene names indicate early and late regulation in AR formation (up, down, unaltered), as well as by wounding (W, -W), in mycorrhizal roots (M, -M), and by phosphate depletion (P, -P). Alignment and tree construction was carried out as described (Dereeper et al., 2008), with the advanced settings. Bootstrap values represent 100 replicates.

Expression patterns of AP2- and GRAS-like transcription factors during adventitious root formation

Next, microarray analysis of all genes encoding AP2 and GRAS proteins was carried out under three conditions representing central developmental programs of angiosperm root systems: AR formation, mycorrhizal roots, and phosphate-starved roots (**Tables S1 and S2**). AR formation was assessed throughout a time course experiment with time points between 2 hours and 8 days after excision. Controls for wounding included leaves 2 hours after wounding. Microarray analysis was performed as described (Breuillin et al., 2010; Ahkami et al., 2014; Nouri et al., 2014). According to Druege et al. (Druege et al., 2014), AR formation can be subdivided into two phases: i) the induction phase (Ai) which involves reprogramming of target cells towards establishment of meristemoids before the onset of

visible cell divisions, and, ii.) the formation phase (Af) from approximately 3 days post-excision (dpe) with first cell divisions, to 8 dpe with fully differentiated roots at the time of emergence from the stem base. Most of the regulated AP2-like genes were induced during the whole process of AR formation (**Fig. 1, Table S1**). However, seven AP2-like genes were induced exclusively or predominantly during the induction phase (time points 2h, 6h, and 24h in **Table S1;** first arrowheads or circles in **Fig. 1**), whereas five were induced primarily during the formation phase (time points 48h, 72h, 96h, 144h, and 192h in **Table S1;** second arrowheads or circles in **Fig. 1**), suggesting phase-specific functions. Many genes that were induced during AR formation were also induced by wounding (**Table S1**, sign "W" in **Fig. 1**), indicating overlapping functions, perhaps related to ethylene signaling (Druege et al., 2014). As in the case of AP2-like genes, some of the GRAS genes were expressed predominantly during the early or late phase of AR formation (**Table S2**). Notably, two SHORT ROOT-like proteins were down-regulated during the whole process of AR formation, whereas several SCARECROW-like proteins were induced during the initiation phase.

Expression patterns of AP2- and GRAS-like transcription factors during arbuscular mycorrhiza and in response to phosphate starvation

In order to gain insight into their potential function in symbiosis, the expression patterns of genes encoding AP2-like transcription factors were analyzed in mycorrhizal roots and in response to phosphate starvation. In general, *AP2-like* genes were only poorly responsive to mycorrhizal colonization (**Fig. 1; Table S1**). Similarly, only few members of the *AP2-like* gene family exhibited altered expression under phosphate-depleted conditions (**Fig. 1; Table S1**).

Since several GRAS-type TFs are involved in establishment and functioning of root symbioses (Oldroyd and Long, 2003; Smit et al., 2005; Gobbato et al., 2012; Floss et al., 2013; Yu et al., 2014; Xue et al., 2015), the expression of all GRAS TFs was also assessed in mycorrhizal *Petunia* roots. Surprisingly, AM development did not change the expression levels of most of the 21 GRAS TFs, and members such as *PaRAM1* and *PaRAD1*, which would be expected to be induced in mycorrhizal roots, are not represented on the microarray (**Fig. 2 and Table S2**). Phosphate deficiency had only mild effects on the expression levels of the GRAS transcription factor genes, but two of them encoding NSP2-like proteins were up-regulated (**Table S2)**.

The strigolactone pathway in *Petunia*

In order to systematically explore the constituents of the SL pathway in *Petunia*, we assessed the occurrence of SL biosynthetic and sensing genes in the genomes of PaxiN and PinfS6. Previous studies have already identified several SL-related genes in *P. hybrida* by forward mutant screens and by homology searches based on known sequences from other angiosperm species (Snowden et al., 2005; Drummond et al., 2009; Drummond et al., 2012; Hamiaux et al., 2012). Consistent with these studies, we found potential orthologs in both *Petunia* genomes for all SL-biosynthetic genes (**Fig. S3a-d**). Clear single orthologs were confirmed for *PhMAX1*, *PhCCD7*/*DAD3* (corresponding to *AtMAX3*), and *PhCCD8*/*DAD1* (corresponding to *AtMAX4*) (**Table S3**; **Figs. 3b-d**). The initial reaction in SL biosynthesis is catalyzed by the carotenoid isomerase D27, which has so far been studied only in rice and *Arabidopsis* (Ruyter-Spira et al., 2013). *Petunia* contains two potential D27 orthologs (D27-L1 and D27-L2) (**Fig. 3a**), and an additional homolog was identified in PinfS6 (not shown).

SL sensing involves the alpha/beta-hydrolase D14, and the F-box protein MAX2 (Ruyter-Spira et al., 2013). *Petunia* contains the putative D14 ortholog DAD2 (Hamiaux et al., 2012) (**Fig. 4a**), in addition to three more distant homologues that cluster between AtD14 and the karrikin receptor AtKAI2 (D14-L1 to D14-L3) and a PaxiN-specific homologue (D14-4) (**Fig. 4a**). MAX2, which represents a single gene in pea, *Arabidopsis*, and rice, is represented by a pair of homologous genes in *Petunia*. Their functional redundancy may be the reason

for the fact that the MAX2 homolog(s) of *Petunia* were not identified by forward genetic mutants screens. Tomato has a similar pair of MAX2 homologs (**Fig. 4b**), indicating that a general duplication of this gene has occurred in the Solanaceae.

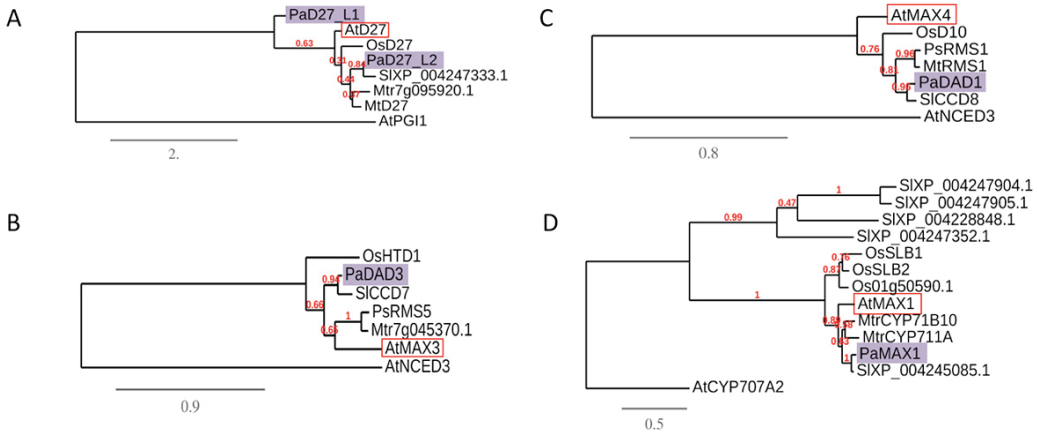


Figure 3. Phylogenetic analysis of proteins involved in strigolactone biosynthesis
Amino acid sequences of the homologs of D27 (A), DAD3/CCD7 (B), DAD1/CCD8 (C), and MAX1 (D) from *P. axillaris*, tomato (*S. lycopersicum*), rice (*Oryza sativa*), and *A. thaliana* were analyzed as described (Dereeper et al., 2008), with the advanced settings. Bootstrap values represent 100 replicates. *Arabidopsis* sequences are framed in red, *Petunia* homologs are highlighted in purple.

The phylogenetic characterization of the SL transporter PhPDR1 was more complex, because plant genomes in general contain large families of full-size ABCG proteins (Kang et al., 2011). In order to identify close homologs (and perhaps orthologs), we retrieved the 5 closest homologs of PhPDR1 from *P. axillaris*, tomato, rice, *Arabidopsis*, and *Medicago*. Surprisingly, phylogenetic analysis of these sequences revealed that the PDRs of rice and *Arabidopsis* tended to cluster species-wise, whereas the homologs from tomato, *Medicago*, and *Petunia* showed a more intermingled phylogeny (**Fig. 5**). Tomato and *Medicago* contain pairs of close homologs of PhPDR1, but also *Petunia* harbors a closely related gene (*PaPDR4*) (**Fig. 5**). It will be interesting to explore its function and test it for SL transport activity.

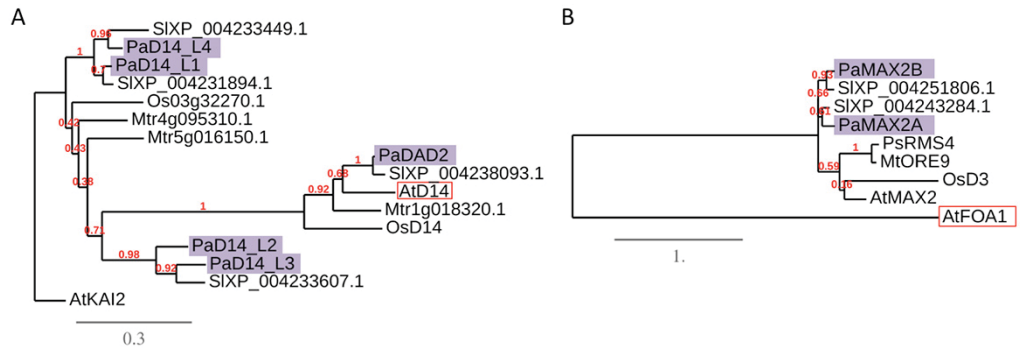


Figure 4. Phylogenetic analysis of proteins involved in strigolactone sensing
Amino acid sequences of the homologs of D27 (A) and MAX2 (B) from *P. axillaris*, tomato (*S. lycopersicum*), rice (*Oryza sativa*), and *A. thaliana* were analyzed as described (Dereeper et al., 2008), with the advanced settings. Bootstrap values represent 100 replicates. *Arabidopsis* sequences are framed in red, *Petunia* homologs are highlighted in purple.

SL-related genes exhibited a rather complex expression pattern. The carotene isomerase D27, which catalizes the first biosynthetic step in SL biosynthesis (Al-Babili and Bouwmeester, 2015), was strongly repressed during adventitious rooting, and was expressed at lower levels in roots than in aerial tissues (**Table S3**). DAD1, in contrast, was slightly repressed during the induction phase and induced moderately during the formation phase of AR formation (**Table S3**). Interestingly, this pattern was reciprocal to the level of

IAA found in the same tissues in dependence on PAT (Ahkami et al., 2013), which was shown to be sensitive to strigolactone action (Shinohara et al., 2013; Koltai, 2014). Both, DAD1 and DAD3 were expressed at higher levels in roots than in the aerial tissues, and were induced by phosphate starvation (**Table S3**). DAD2, which is involved in SL sensing, was rather repressed during AR formation. Wounding and mycorrhizal colonization had no consistent effect on the expression of SL-related genes. Taken together, the tendency of SL-related genes to be downregulated during AR formation is consistent with the inhibitory effect of SL in AR formation in *Arabidopsis* and pea (Rasmussen et al., 2012).

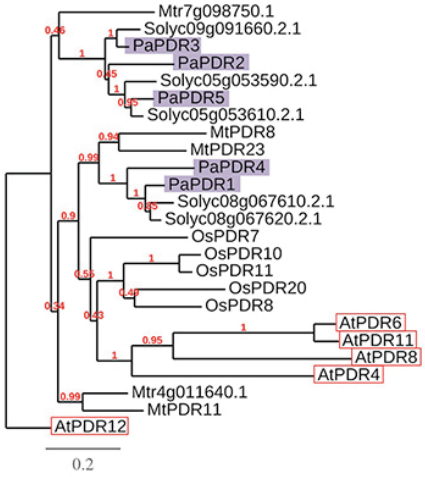


Figure 5. Phylogenetic analysis of the strigolactone transporter PDR1

Amino acid sequences of the homologs of PhPDR1 from *P. axillaris*, tomato (*S. lycopersicum*), rice (*Oryza sativa*), and *A. thaliana* were analyzed as described (Dereeper et al., 2008), with the advanced settings. Bootstrap values represent 100 replicates. *Arabidopsis* sequences are framed in red, *Petunia* homologs are highlighted in purple.

The LysM family of receptors for microbial signals

We further screened the *Petunia* genome for homologues of the lysine motif receptor-like kinases (LysM-RLKs) that are essential for RNS (Madsen et al., 2003; Op den Camp et al., 2011), and that are thought to be required for the initial recognition of the fungal partner in AM (Op den Camp et al., 2011; Zhang et al., 2015). We used the 17 *L. japonicus* LysM family members (Lohmann et al., 2010) to search the genomes of both *Petunia* species, PaxiN and PinfS6. Careful manual curation was required in the case of the members of the NFR1 subfamily, which have 12 exons, of which some are as short as 8 bp (Lohmann et al., 2010), and therefore cannot be reliably predicted with automated procedures.

Open reading frames were reconstructed according to genomic information from *L. japonicus*, *M. truncatula*, and the tomato genome sequence, resulting in a total of 10 LysM receptor kinases in *P. axillaris* and 11 in *P. inflata* (**Table S4**). The *Petunia* LysM family exhibits a clearly distinct Lys-I group with three members in *P. axillaris* and four members in *P. inflata* (**Fig. 6, Table S4**), of which Lys-IIc and Lys-IId encode a pair of very similar proteins (>99% identity at the amino acid level, and a 20 aa extension at the C-terminus of Lys-IId), that are organized as a tandem repeat. The remaining 7 members resolved rather poorly despite their distinct gene structures (compare **Fig. 6 and Table S4**). Although *Petunia* exhibits two LYS-III members, based on exon-intron structure, these cluster far from the legume LYS-III clade (**Fig. 5**, asterisks). Considering the entire LysM-RLK families per species, the LysM families of the legumes *L. japonicus* and *M. truncatula*, with 17 and 18 members (Arrighi et al., 2006; Lohmann et al., 2010), respectively, turned out to be considerably larger than the solanaceous LysM-RLK families with 10/11 members in *Petunia* and 14 members in tomato (**Fig. 6, Table S4**). In general, the LysM-RLKs were expressed at higher levels in roots than in aerial tissues, and were induced by wounding and during AR formation, except for Lys-IIIb, which was repressed during AR formation (**Table S4**). None of the LysM-RLKs showed a significant regulation in response to AM colonization (**Table S4**), whereas PaLYS-IIa, the closest petunia homolog of LjNFR5 and MtNFP, was repressed by high phosphate supply phosphate (**Table S4**) (Breuillin et al., 2010).

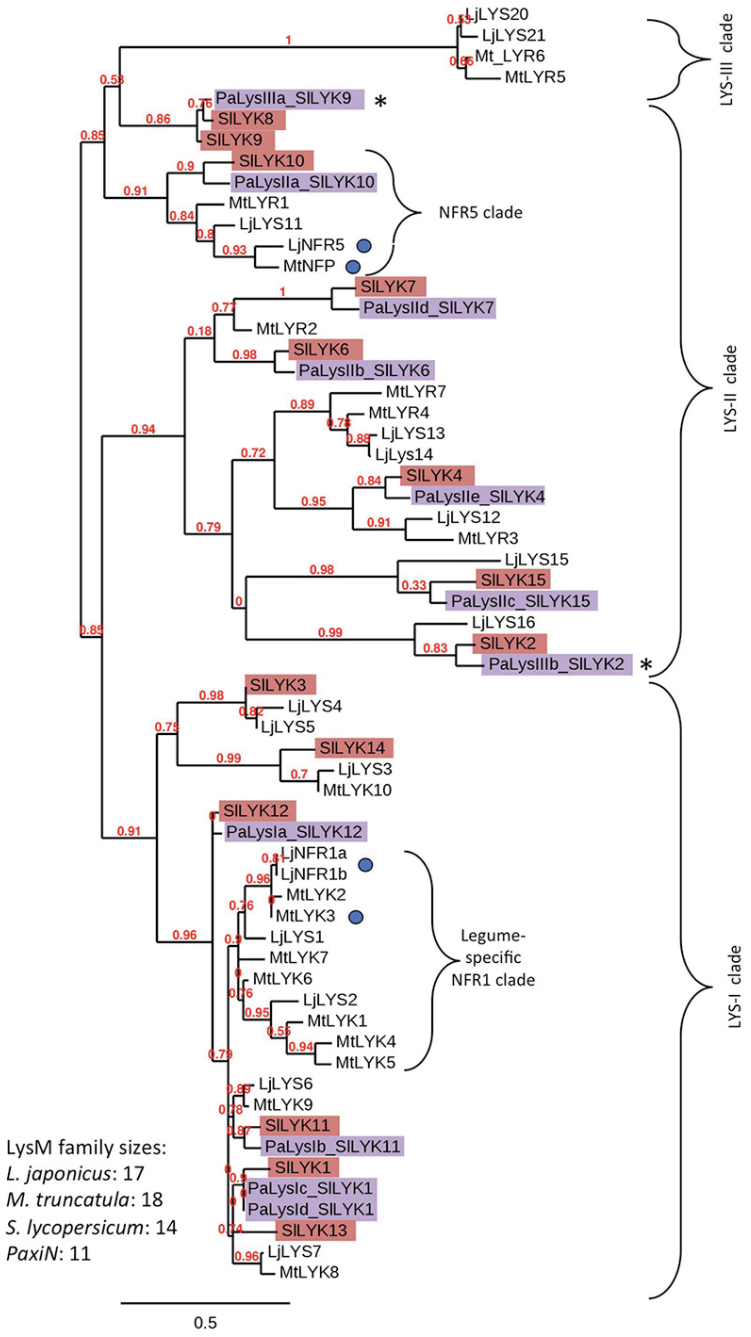


Figure 6. Phylogenetic analysis of the LysM family of receptor-like kinases.

Amino acid sequences of all LysM-RLKs of M. truncatula (Mt), L. japonicus (Lj), P. axillaris (Pa), and tomato (Sl) was clustered together. Blue circles indicate functionally characetrized nod factor receptors, asterisks denote *Petunia* members of the LYS-III clade (based on exon-intron structure), that clustered with homologues of the LYS-II clade (compare with **Table S5**). *Petunia* sequences are highlighted in purple, tomato sequences in red. Annotation of clades is based on (Lohmann et al., 2010). Alignment and tree construction was carried out as described (Dereeper et al., 2008), with the basic settings. Bootstrap values represent 100 replicates.

The large size of the LysM-RLK families in *L. japonicus* and *M. truncatula*, is due primarily to the extended Lys-I clade in the legumes with 9 and 10 members, respectively. Notably, the NFR1 group in the Lys-I clade contained no solanaceous homologues, suggesting that this branch may play a specific role in nodulating plant species, although recent evidence indicates a role for closely related legume type-I LysM-RLKs in AM (Zhang et al., 2015). Another important difference between legumes and the *Solanaceae* is the fact that the NFR5 clade of the legumes has two members each, whereas *Petunia* and tomato have only one (**Figure S4**), indicating that the legumes have undergone a specific duplication of NFR5. Taken together, these results indicate that molecular evolution of the LysM-RLK gene family in legumes has been considerably influenced by the evolution of RNS, relative to other angiosperms, like *Petunia*, that engage only in AM.

General and AM-specific symbiosis signalling components in *Petunia*

In order to explore the constituents of the CSSP in *Petunia*, we searched the predicted *Petunia* transcriptome by tblastn using functionally characterized proteins from *M. truncatula* and *L. japonicus* as queries. This approach revealed homologues for all known components of the CSSP, and homologs of four additional genes known to be essential for AM, namely VAPYRIN, STR, STR2, and RAM2 (**Table S5**) (Feddermann et al., 2010; Pumplin et al., 2010; Zhang et al., 2010; Wang et al., 2012). Three CSSP genes, namely SYMRK, NENA and CCaMK, were induced by AR formation and were expressed at higher levels in roots than in aerial tissues (**Table S5**). Only two genes were induced in mycorrhizal roots, namely the homologs of STR and RAM2, which were shown to be induced in mycorrhizal petunia roots in previous studies (Breuillin et al., 2010; Rich et al., 2015).

Discussion

Phylogeny of AP2-like and GRAS genes

Organogenesis in multicellular organisms is often controlled by networks of transcription factors (TFs), as for example the ABC system in floral development (Causier et al., 2010). Similarly, root inception and maintenance involves the action of a series of TFs including AP2-like, and GRAS-type TFs, which control cell fate decisions and cellular identity (Petricka et al., 2012). The *P. axillaris* genome encodes a total of 109 AP2-like transcription factors and 69 GRAS transcription factors. In comparison, 147 AP2-like genes (many referred to as ethylene response factors, ERFs) have been detected in *Arabidopsis* (Dietz et al., 2010). The higher number of AP2-like genes in *Arabidopsis* may be based on less stringent criteria applied in this study. On the other hand, the *Petunia* GRAS family has approximately double the size of the family in *Arabidopsis* (with 32 members), and is in the range of the rice GRAS family (57 members), as estimated by Tian and colleagues (Tian et al., 2004), although recent estimations are slightly higher (33 and 60, respectively) (Hirsch and Oldroyd, 2009). For each TF in *P. axillaris*, the corresponding orthologue was found in the genome of the sister species *P. inflata*, except for one AP2-like transcription factor. This indicates high quality of the two genome sequences.

The phylogenomic analysis of the AP2-like genes revealed that homologues of essentially all the clades known from *Arabidopsis* were found in *Petunia* (**Fig. 1**). On the other hand, the phylogenetic analysis of the *Petunia* GRAS family, together with symbiosis-related GRAS proteins from other species (*Medicago*, *Lotus*, rice), and with all GRAS genes of *Arabidopsis* revealed three branches that lack an *Arabidopsis* homologue (**Fig. 2,** green boxes), indicating that the members of these branches may play a specific role in symbiosis, and therefore were secondarily lost in the non-symbiotic species *Arabidopsis* (Delaux et al., 2013; Delaux et al., 2014; Favre et al., 2014). Indeed, the analysis of the syntenic regions containing *RAM1*, *RAD1*, and the *L. japonicus* GRAS gene Lj_chr5CM0239.240.r2.m (AM-GRAS3) revealed that these regions had undergone major rearrangements in *A. thaliana*, that resulted in the deletion of the three GRAS genes (**Fig. 7**). In the case of the *RAD1* locus, the corresponding syntenic region in *Arabidopsis* could not be identified with certainty, since two regions exhibited a remote similarity with the syntenic regions containing the *RAD1* locus in *Vitis* and the *Solanaceae* (**Fig. 7b**).

Regulation of AP2-like and GRAS genes in adventitious rooting and arbuscular mycorrhiza

During the process of AR formation, 34 of the 41 AP2-like genes that are represented on the microarray, and 15 of the 21 GRAS genes showed up-regulation or repression (**Figs. 1 and 2, Tables S1 and S2**) and may therefore be involved in AR development. A significant number of AP2 genes, representing several ethylene response factor (ERF) clades but also other branches, were induced both by wounding and during AR formation (**Fig. 1, Table S1**),

pointing to a role of ethylene in both processes (Druege et al., 2006; Druege et al., 2014). However, little overlap was detected between AR formation and mycorrhizal roots or Pi-starved roots (**Figs. 1 and 2, Table S1**). AP2-like genes that were differentially regulated during AR formation included homologues of known developmental regulators such as PLT1, CRF3 and AP2 (**Fig. 1**), pointing to specific developmental programs related to cell fate decisions during AR induction, differentiation, and growth. Recently, two *PLT*-like genes were shown to be induced during AR formation in poplar, and the positive regulatory role of an Aintegumenta-like (AIL) TF in AR formation was proven by overexpression and knockdown (Rigal et al., 2012). Our transcriptomic analysis of AR formation shows differential induction of early and late genes (**Table S1**), which may be related to the initiation and formation phase, respectively, in AR development (Druege et al., 2014).

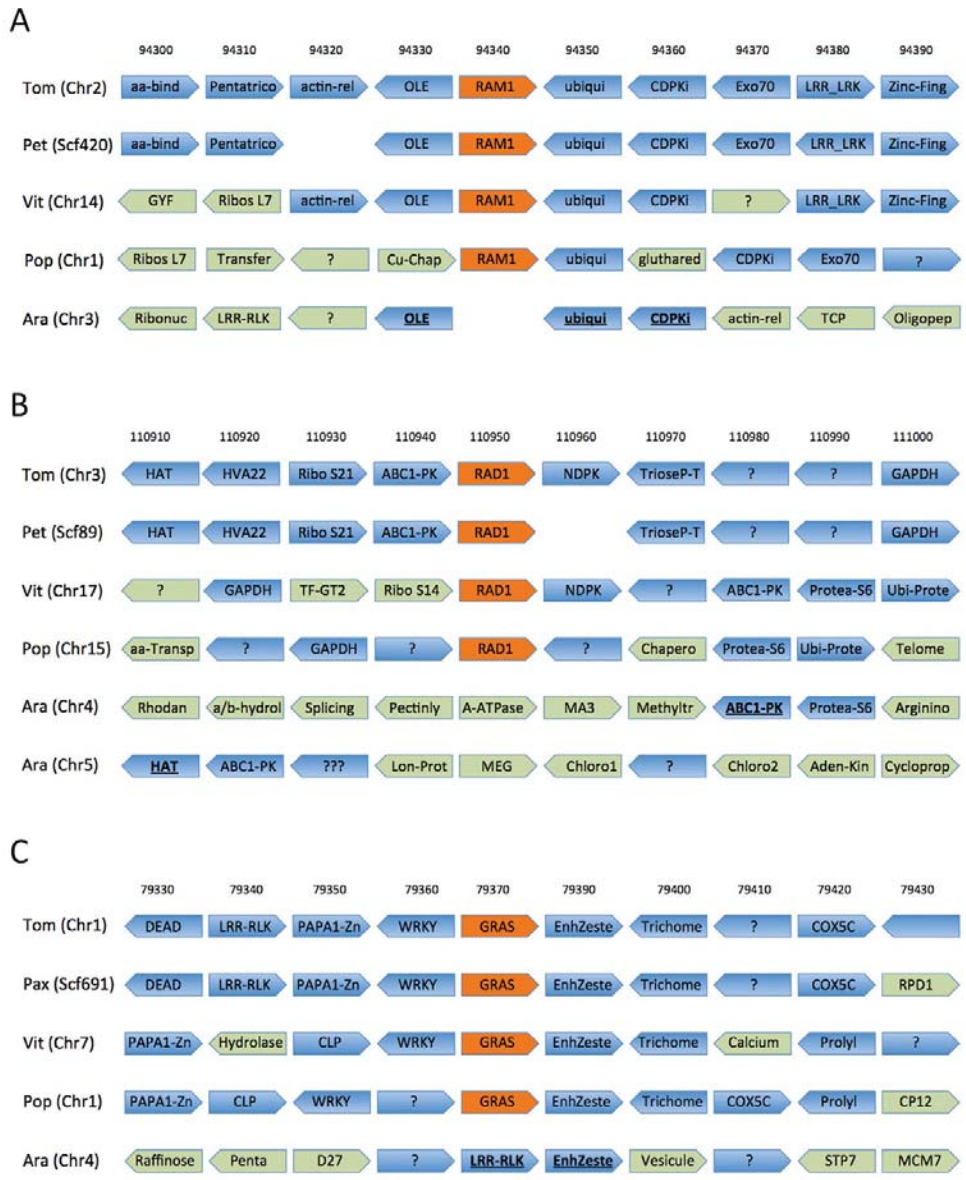


Figure 7. Synteny of the regions containing RAM1, RAD1 and AM-GRAS3.
The genomic regions containing RAM1, RAD1, and AM-GRAS3 were first identified in the genome of *S. lycopersicum* (Tom). Subsequently, the corresponding regions were identified in the genomes of *P. axillaris* N (Pax), vine (Vit, *Vitis vinifera*), poplar (Pop, *Populus trichocarpa*), and *A. thaliana* (Ara). Genes with at least two hits appear in blue, gene substitutions with only one occurrence are green. Sizes of genes and distances between

genes are not drawn to scale. Numbers above the genes represent the gene number in tomato. Question marks represent open reading frames (ORFs) with uncertain annotation.
(A) The respective gene names of the RAM1 homologues are (from top to bottom): Solyc02g094340.1.1, Peaxi162Scf00420g06006.1, GSVIVT01033046001, Potri.001G326000.1. Abbreviations: aa-bind, amino acid-binding protein; Pentatrico, pentatricopeptide repeat protein; actin-rel, actin-related protein; OLE, OLE e 1 pollen protein; ubiqui, ubiquitin family protein; CDPKi, calcium-dependent protein kinase inhibitor; Exo70, exocyst subunit Exo70; LRR-RLK, leucine-rich repeat receptor-like kinase; Zinc-Fing, zinc-finger DHHC domain protein; GYF, GYF-domain protein; Ribos L7, ribosomal protein L7; Transfer, ω-hydroxypalmitate *O*-feruloyl transferase; Cu-Chap, chloroplast copper chaperone; glutharedg glutharedoxin; Ribonuc, ribonuclease; TCP, TCP family transcription factor; Oligopep, oligopeptide transporter.
(B) The respective gene names of the RAD1 homologues are (from top to bottom): Solyc03g110950.1.1, Peaxi162Scf00089g03001.1, GSVIVT01007532001, Potri.015G091200.1. Abbreviations: HAT, histone-acetyl transferase; HVA22, HVA22-like protein; Ribo S21, ribosomal protein S21; ABC1-PK, ABC1 family protein kinase; NDPK, nucleoside diphosphate kinase; TrioseP-T, triosephosphate transporter; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; TF-GT2, transcription factor GT2; Ribo S14, ribosomal protein S14; Protea-S6, 26S proteasome regulatory subunit6; Ubi-Prote, ubiquitin protease; aa-Transp, amino acid transporter; Chapero, chaperon BC1; Telome, telomere length regulation protein; Rhodan, rhodanese-like; a/b-hydrol, alpha/beta-hydrolase; Splicing, splicing factor 3B; Pectinly, pectin lyase; A-ATPase, AAA_ATPase replication factor; MA3, MA3-domain-containing protein; Methyltr, methyltransferase; Arginino, argininosuccinate synthase; Lon-Prot, Lon peptidase S16; MEG, maternally expressed gene; Chloro1, chloroplast protein1; Chloro2, chloroplast protein2; Aden-Kin, adenylate kinase; Cycloprop, cyclopropyl isomerase.
(C) The respective gene names of the GRAS homologues are (from top to bottom): Solyc01g079370.2.1, Peaxi162Scf01160g01023.1, GSVIVT01028126001, Potri.002G195600.1. Abbreviations: DEAD, DEAD-domain protein; PAPA1-Zn, PAPA1 zinc finger protein; WRKY, WRKY transcription factor; Enh_Zeste, enhancer of zeste; Trichome, trichome birefringence protein; COXSC, cytochrome C oxidase subunit VC family protein; RPD1, root primordium defective1; Hydrolase, glycosyl hydrolase; CLP, CLP protease; Calcium, calcium-dependent phosphodiesterase; Prolyl, prolyl oligopeptidase; CP12, CP12 domain-containing protein 2; Raffinose, raffinose synthase; Penta, pentatricopeptide protein; D27, beta-carotene isomerase D27; Vesicule, vesicule transport protein; STP7, sugar transporter 7; MCM7, minichromosome maintenance7 protein.

Among GRAS-type transcription factors, SCARECROW (SCR) and SHORT-ROOT (SHR) are important regulators of cell identity during development of the primary root in *Arabidopsis* (Sparks et al., 2013). They are involved in the establishment of quiescent center identity and maintenance of the stem cell status of the surrounding cells (Sabatini et al., 2003), and would therefore be expected to be induced during AR formation. However, the SCR clade, which comprises three members in *Arabidopsis* (**Fig. 2**), does not have a close *Petunia* homologue, and *Petunia* SHR homologues were repressed rather than induced during AR formation (**Fig. 2, Table S2**). While SCARECROW-LIKE (SCL) genes have been implicated at early stages of AR formation in *Pinus radiata* and *Castanea sativa* (Sanchez et al., 2007), it remains to be shown whether the role of SCR and SHR is conserved between *Arabidopsis* and *Petunia*.

Several GRAS genes have been shown to promote AM colonization or to be essential for AM, namely MtRAM1, MtNSP, MtNSP1, MtNSP2, LjRAD1, OsDIP1, and OsSLR1. Notably, their closest homologues in *Petunia* were either not induced during AM, or they were not represented on the micorarray (**Fig. 2**).

***Petunia* as a model for strigolactone biology**

Strigolactones (SL) regulate apical dominance and lateral branching of the shoot (Domagalska and Leyser, 2011), and are the first signal in the AM interaction (Gutjahr and Parniske, 2013). SLs are released from roots to stimulate AM fungal metabolism and hyphal branching, thereby promoting the interaction. *Petunia* has been one of the leading genetic systems to analyze SL biosynthesis and signaling (Snowden and Napoli, 2003; Snowden et al., 2005; Drummond et al., 2009; Drummond et al., 2012; Hamiaux et al., 2012), and it is the only system, in which a SL transporter, PDR1, has been identified and functionally studied (Kretzschmar et al., 2012). Based on *pdr1* loss- and gain-of-function mutant phenotypes (Kretzschmar et al., 2012; Sasse et al., 2015), PhPDR1 is likely to represent the

main, if not the only SL transporter in *Petunia*. Whether PDR4, the closest homolog of PDR1, contributes to SL transport remains to be shown.

Our analysis on the SL-related genes in the *P. axillaris* and *P. inflata* genomes complements earlier studies and establishes the entire set of SL-related genes known to date (**Figs. 4-6**). The analysis of the PDR1 family and its closest homologues in *Petunia*, tomato, *Medicago*, *Arabidopsis*, and rice suggests a complex pattern of evolution in this gene family. Phylogenetic analysis failed to reveal close homologs of PDR1 in *Arabidopsis* and rice, in contrast to tomato and *Medicago* that each carry a pair of potential PDR1 orthologs (**Fig. 5**). So far, direct evidence for a PDR1 ortholog was obtained only in *Nicotiana tabacum* (Xie et al., 2015). Further research in *Arabidopsis* and rice will be needed to show, which transporters mediate SL flux within these species, and SL exudation from roots, at least in the case of the AM-competent species rice.

The LysM receptor-like kinase family in *Petunia*

Recognition of the symbiotic signal from AM fungi and rhizobia is mediated by LysM-RLK family members. The symbiotic signals are represented by lipochitooligo-saccharides (LCO) consisting of an *N*-acetylglucosamin backbone with 4-5 monomers, and various lateral substitutions, among them a fatty acid (Dénarié et al., 1996; Maillet et al., 2011). Additional substitutions in case of the Nod factors result in a great variability, which is believed to contribute to the high degree of host specificity in RNS. In contrast, AM shows very little specificity at the level of infection and colonization, although a high degree of functional diversity (Klironomos, 2003) points to a certain level of specificity at the functional level (Smith and Read, 2008).

In the case of AM, two main types of LCO signals have been described, which are distinguished by a sulfate group at a characteristic position (Maillet et al., 2011). In addition, non-decorated *N*-acetylglucosamin oligosaccharides (i.e. chitin oligosaccharides) can trigger some aspects of the symbiotic program in plants (Genre et al., 2013). How the legumes distinguish between the signals from rhizobia and AM fungi is not fully understood, but is thought to be related to the specific set of LysM-RLKs, which can heterodimerize and therefore may give rise to a high number of diverse dimeric receptor complexes with a wide range of ligand specificities (Antolin-Llovera et al., 2014).

The similarity of the microbial signals in AM and RNS is mirrored in overlapping functions of LysM-RLK in these two mutualistic interactions (Op den Camp et al., 2011). Furthermore, some LysM-RLK of rice exhibit a dual function as chitin receptors involved in disease resistance and as receptors of symbiotic signals in AM (Miyata et al., 2014; Zhang et al., 2015). Interestingly, the LysM-RLK gene families in legumes have considerably expanded relative to tomato and *Petunia* (**Fig. 6**) and other angiosperms (Zhang et al., 2009), indicative of functional diversification as a consequence of the evolution of RNS. Based on this evidence, a non-nodulating species such as *Petunia* offers the possibility to functionally analyze the LysM-RLK family in a simpler context that has been shaped exclusively by the interaction with AM fungi (and perhaps chitinous pathogens), but not rhizobia.

The LYS-I clade in the legumes is particularly large, with 9 and 10 members in *L. japonicus* and *M. truncatula*, respectively, compared to 3 or 4 in *Petunia* and 6 in tomato, and it involves a legume-specific subclade (**Fig. 6**). Together with a legume-specific Lys-III clade (**Fig. 6**), they may represent a specific set of receptors for RNS. In this context, it is worth to note that the promiscuous chitin receptors of rice that are involved in AM development, also fall into the LYS-I clade (Miyata et al., 2014; Zhang et al., 2015).

In contrast to the NFR1 subfamily, the NFR5 clade contains only 2 members in the legumes, and only a single one in *Petunia* and tomato, indicative of a legume-specific gene duplication (**Fig. 6**). Analysis of microsynteny at the NFP locus revealed that *PaLYSIIa* and *Sl_LYK10* represent the orthologues of *NFP* in *M. truncatula* (and hence also of *NFR5* in *L.*

japonicus) (**Fig. 8**). While the orthologous regions could be identified in the genomes of the AM-competent species poplar (*Populus trichocarpa*) grapevine (*Vitis vinifera*), common bean (*Phaseolus vulgaris*), and *M. truncatula*, no obvious syntenic region could be identified in the *Arabidopsis* genome.

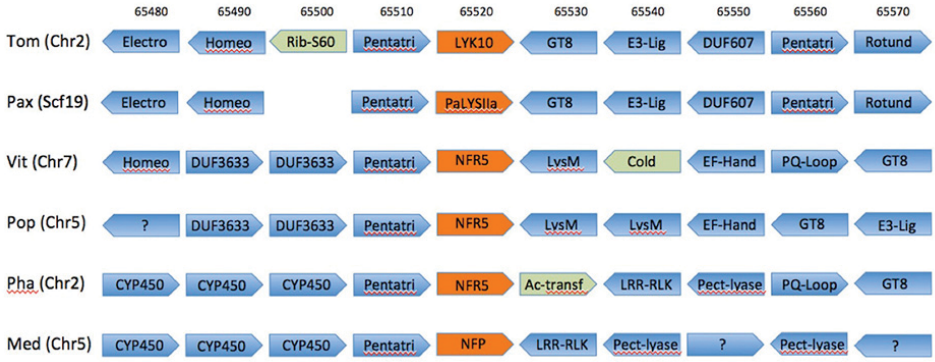


Figure 8. Synteny of the region containing the nod factor receptor NFP and its orthologues.
The genomic region containing the closest homologue of *NFP* (*LYK10*) was identified in *S. lycopersicum* (Tom), and compared with the corresponding regions from *P. axillaris* N (Pax), *V. vinifera* (Vit), poplar (Pop), common bean (Pha, *Phaseolus vulgaris*), and *M. truncatula* (Med). Genes with at least two hits appear in blue, gene substitutions with only one occurrence are green. Sizes of genes and distances between genes are not drawn to scale. Numbers above the genes represent the gene number in tomato. Question marks represent open reading frames (ORFs) with uncertain annotation. The respective gene names of the NFP homologues are (from top to bottom): Solyc02g065520.1.1, Peaxi162Scf00019g15005.1, GSVIVT01019039001, Potri.005G128400.1, Phvul.002G025500.1, Medtr5g019040.1. Abbreviations: Electro, electron transport protein; Homeo, Bel1-like homeodomain protein; Rib-S60, ribosome 60S biogenesis protein; Pentatri, pentatricopeptide repeat protein; GT8, glycosyl transferase8; E3-Lig, E3-Ligase C3HC4; DUF607, protein containing domain of unknown function 607; Rotund, Rotundifolia; EF-Hand, EF-hand-containg calcium binding protein; PQ-Loop, PQ-loop repeat family protein; CYP450, cytochrome P450; Pect-lyase, pectate lyase.

Interestingly, *PaLYSIIa* is repressed in *Petunia* by high phosphate, i.e. under conditions that repress AM (see microarray ID drpoolB-CL8626Contig1 in (Breuillin et al., 2010)). Given that it represents the orthologue of the nod factor receptor in the legumes and of *PaNFP* from *Parasponia andersonii*, which functions as a receptor for an AM fungal signal, its repression by phosphate could contribute to the strong inhibitory effect of phosphate on AM. Surprisingly, two *Petunia* members of the LYS-III clade (based on exon-intron structure, see **Table S4**) clustered within the LYS-II branch instead of the LYS-III branch (**Fig. 6**, asterisks).

Conservation of genes required for AM symbiosis

Upon recognition by LysM-RLKs, the symbiotic signal from AM fungi and rhizobia is transmitted to the nucleus by a dedicated pathway known as the common symbiosis (SYM) signaling pathway (CSSP) (Oldroyd et al., 2011; Harrison, 2012; Gutjahr and Parniske, 2013). The genes constituting the CSSP are widely conserved among AM-competent plants comprising monocots and dicots (including *Petunia*) (Delaux et al., 2014; Favre et al., 2014), and even lower land plants such as liverworts, mosses, hornworts, and lycophytes (Wang et al., 2010; Xue et al., 2015), suggesting that they have a common evolutionary origin in early land plants (Kistner and Parniske, 2002). In addition to the genes constituting the CSSP, which are also involved in nodulation, several other genes have been identified as essential specifically in AM symbiosis (Feddermann et al., 2010; Pumplin et al., 2010; Zhang et al., 2010; Wang et al., 2012). They are induced in cortical cells that host the fungal arbuscules (Feddermann et al., 2010; Pumplin et al., 2010; Zhang et al., 2010; Wang et al., 2012),

suggesting that they may play essential roles at later stages of symbiosis. These were also found in the Petunia genome, and are induced (**Table S5**) as in other species.

Conclusions

We report a systematic genomic survey of *Petunia* genes involved in AM symbiosis and in AR formation, with the focus on genes encoding AP2-like proteins, GRAS transcription factors, and LysM-RLKs, besides other genes that have been functionally studied by mutant analysis. ARs share several anatomical and developmental commonalities with lateral roots (Bellini et al., 2014), including the involvement of auxin as a central inducing signal (Ahkami et al., 2013). This fact, and the expression patterns of AP2-like and GRAS genes, suggests that AR formation may involve a conserved developmental program controlling cell specification via specific transcriptional networks. Concerning AM, we conclude that *Petunia* has highly conserved sets of genes related to SL biology and the CSSP. On the other hand, the comparative phylogenomic analysis of LysM-RLKs in *Petunia* and tomato vs. the legumes *M. truncatula* and *L. japonicus*, revealed that the latter have significantly larger gene families, indicative of accelerated evolution of this type of receptor-like kinases that are implicated in the perception of AM fungi and rhizobia.

Materials and Methods

Sequence retrieval and analysis

The original AP2/PLETHORA-like and GRAS-like proteins from *A. thaliana* were used as query sequences to search the *P. hybrida* EST database (http://webblast.ipk-gatersleben.de/petunia/). Sequence identifiers with E-values inferior of e^{-10} were used to retrieve the corresponding full length coding sequences (predicted transcripts) in the *Petunia* genome database (http://petuniasp.sgn.cornell.edu/blast/blast.html). Additional sequences, which were not represented in the EST database, were identified by direct blast searches against the petunia genome database with *A. thaliana* sequences as queries. Proteins from other plant species (*M. truncatula*, *L. japonicus*, *S. lycopresicum*, and *O. sativa*) were identified in the NCBI database by blastp (http://blast.ncbi.nlm.nih.gov/Blast.cgi). SL-related sequences were retrieved from http://jcvi.org/medicago, http://rice.plantbiology.msu.edu, and http://solgenomics.net.

To identify the petunia homologues of common SYM genes and LysM-RLKs, the legume sequences were used as queries to search the *Petunia* predicted proteomes of PaxiN. To retrieve the complete set of LysM-RLKs, both *Petunia* genomes were searched by tblastn with all *L. japonicus* LysM-RLK sequences at the Solanaceae Genome Network (SGN; http://petuniasp.sgn.cornell.edu/blast/blast.html). The completeness of the *Petunia* LysM-RLK families was confirmed by similar results from both *Petunia* species. Protein sequence of the members of the subclades LYS-II and LYS-III were predicted by Augustus (http://bioinf.uni-greifswald.de/augustus/submission.php), whereas the protein predictions for the complex exon-intron structure of the LYS-I members was resolved manually according to the published LYS-I genes of *L. japonicus*, *M. truncatula*, and *S. lycopersicum*. The accession numbers of all sequences used in this study are listed in **Table S6**.

Phylogenetic analysis

Phylogenetic analysis was performed with Phylogeny.fr (Dereeper et al., 2008), involving previously described tools (Castresana, 2000; Guindon and Gascuel, 2003; Edgar, 2004; Anisimova and Gascuel, 2006; Chevenet et al., 2006; Dereeper et al., 2010). Phylograms were produced with the "Advanced" mode (see http://www.phylogeny.fr) using the bootstrap function to calculate support for branch separation. Bootstrap values reflect 100 replicates.

Gene expression analysis

Total RNA was extracted from the rooting zone (stem base) of leafy cuttings at different developmental stages after excision, from roots and from fresh and wounded leaves. In addition, total RNA was isolated from mycorrhizal and non-mycorrhizal control roots 5, 7 and 8 weeks after inoculation and from roots of plants fertilized with low or high phosphate concentrations. Total RNA from different developmental stages of cutting stem bases and control tissues (leaves and roots) was extracted with the QIAGEN kit (Qiagen, Hilden, Germany). After DNAse treatment following the protocol of the DNAse supplier (Qiagen), 20 µg of total RNA was used for probe synthesis. Array hybridization, data analysis and normalization was carried out by Roche Nimblegene (Waldkraiburg, Germany) as described (Breuillin et al., 2010; Ahkami et al., 2014; Nouri et al., 2014). The array was based on sequences of cDNA libraries from RNA of leafy cuttings at different developmental stages and of mycorrhizal and non-mycorrhizal roots, and on publicly available sequences from public databases (Breuillin et al. 2010).

Analysis of synteny

The regions of the genes of interest in petunia were identified by blastn in the *P. axillaris* scaffold database (http://petuniasp.sgn.cornell.edu). For the remaining species, the regions of interest were identified by blastn in phytozome (http://phytozome.jgi.doe.gov). The regions of interest in the *P. axillaris* genome were manually curated to determine the identity and orientation of genes adjacent to the genes of interest. For the other species, the identity and orientation of the corresponding genes were directly extracted from phytozome.

Supplementary Tables

Supplemental Tables S1-S6 are available online at the following URL:
http://www.igzev.de/projects/Genome-petunia-roots/; user: igz; pw: RHizoGEnPEt

Table S1. Virtual transcripts of AP2-like genes deduced from the two *Petunia* genomes and their expression in different tissues of *P. hybrida*, during AR formation, in mycorrhizal roots, and in response to phosphate starvation.

Table S2. Virtual transcripts of GRAS transcription factor genes deduced from the two *Petunia* genomes and their expression in different tissues of *P. hybrida*, during AR formation, in mycorrhizal roots, and in response to phosphate starvation.

Table S3. Virtual transcripts of strigolactone-related genes deduced from the two *Petunia* genomes and their expression in different tissues of *P. hybrida*, during AR formation, in mycorrhizal roots, and in response to phosphate starvation.

Table S4. Virtual transcripts of LysM-RLK genes deduced from the two *Petunia* genomes and their expression in different tissues of *P. hybrida*, during AR formation, in mycorrhizal roots, and in response to phosphate starvation.

Table S5. Virtual transcripts of genes required for AM symbiosis deduced from the two *Petunia* genomes and their expression in different tissues of *P. hybrida*, during AR formation, in mycorrhizal roots, and in response to phosphate starvation genome.

Table S6. List of sequence identifiers and accession numbers.

References

Ahkami, A.H., Melzer, M., Ghaffari, M.R., Pollmann, S., Javid, M.G., Shahinnia, F., Hajirezaei, M.R., and Druege, U. (2013). Distribution of indole-3-acetic acid in *Petunia hybrida* shoot tip cuttings and relationship between auxin transport, carbohydrate metabolism and adventitious root formation. Planta **238,** 499-517.

Ahkami, A.H., Lischewski, S., Haensch, K.T., Porfirova, S., Hofmann, J., Rolletschek, H., Melzer, M., Franken, P., Hause, B., Druege, U., and Hajirezaei, M.R. (2009). Molecular physiology of adventitious root formation in Petunia hybrida cuttings: involvement of wound response and primary metabolism. New Phytol. **181,** 613-625.

Ahkami, A.H., Scholz, U., Steuernagel, B., Strickert, M., Haensch, K.-T., Druege, U., Reinhardt, D., Nouri, E., von Wirén, N., Franken, P., and Hajirezaei, M.R. (2014). Comprehensive transcriptome analysis unravels the existence of crucial genes regulating primary metabolism during adventitious root formation in *Petunia hybrida*. PLoS ONE **9,** e100997.

Al-Babili, S., and Bouwmeester, H.J. (2015). Strigolactones, a novel carotenoid-derived plant hormone. Annual review of plant biology **66,** 161-186.

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Systematic Biology **55,** 539-552.

Antolin-Llovera, M., Petutsching, E.K., Ried, M.K., Lipka, V., Nurnberger, T., Robatzek, S., and Parniske, M. (2014). Knowing your friends and foes - plant receptor-like kinases as initiators of symbiosis or defence. New Phytol. **204,** 791-802.

Arrighi, J.F., Barre, A., Ben Amor, B., Bersoult, A., Soriano, L.C., Mirabella, R., de Carvalho-Niebel, F., Journet, E.P., Gherardi, M., Huguet, T., Geurts, R., Denarie, J., Rouge, P., and Gough, C. (2006). The *Medicago truncatula* lysine motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. Plant Physiol. **142,** 265-279.

Bellini, C., Pacurar, D.I., and Perrone, I. (2014). Adventitious roots and lateral roots: Similarities and differences. Annual Review Of Plant Biology **65,** 639-666.

Breuillin, F., Schramm, J., Hajirezaei, M., Ahkami, A., Favre, P., Druege, U., Hause, B., M., B., Kretzschmar, T., Bossolini, E., Kuhlemeier, C., Martinoia, E., Franken, P., Scholz, U., and Reinhardt, D. (2010). Phosphate systemically inhibits development of arbuscular mycorrhiza in *Petunia hybrida* and represses genes involved in mycorrhizal functioning. Plant J. **64,** 1002-1017.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution **17,** 540-552.

Causier, B., Schwarz-Sommer, Z., and Davies, B. (2010). Floral organ identity: 20 years of ABCs. Seminars in Cell & Developmental Biology **21,** 73-79.

Chevenet, F., Brun, C., Banuls, A.L., Jacq, B., and Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics **7**.

Correa, L.D., Troleis, J., Mastroberti, A.A., Mariath, J.E.A., and Fett, A.G. (2012). Distinct modes of adventitious rooting in *Arabidopsis thaliana*. Plant Biol. **14,** 100-109.

da Costa, C.T., de Almeida, M.R., Ruedell, C.M., Schwambach, J., Maraschin, F.S., and Fett-Neto, A.G. (2013). When stress and development go hand in hand: main hormonal controls of adventitious rooting in cuttings. Frontiers in Plant Science **4**.

Delaux, P.-M., Varala, K., Edger, P.P., Coruzzi, G.M., Pires, J.C., and Ane, J.-M. (2014). Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. Plos Genetics **10,** e1004487.

Delaux, P.-M., Xie, X., Timme, R.E., Puech-Pages, V., Dunand, C., Lecompte, E., Delwiche, C.F., Yoneyama, K., Becard, G., and Sejalon-Delmas, N. (2012). Origin of strigolactones in the green lineage. New Phytol. **195,** 857-871.

Delaux, P.M., Sejalon-Delmas, N., Becard, G., and Ane, J.M. (2013). Evolution of the plant-microbe symbiotic 'toolkit'. Trends Plant Sci. **18,** 298-304.

Dénarié, J., Debellé, F., and Promé, J.C. (1996). Rhizobium lipo-chitooligosaccharide nodulation factors: Signaling molecules mediating recognition and morphogenesis. Annual Review of Biochemistry **65,** 503-535.

Dereeper, A., Audic, S., Claverie, J.M., and Blanc, G. (2010). BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evolutionary Biology **10**.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., and Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. **36,** W465-W469.

Dietz, K.-J., Vogel, M.O., and Viehhauser, A. (2010). AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. Protoplasma **245,** 3-14.

DiLaurenzio, L., WysockaDiller, J., Malamy, J.E., Pysh, L., Helariutta, Y., Freshour, G., Hahn, M.G., Feldmann, K.A., and Benfey, P.N. (1996). The *SCARECROW* gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. Cell **86,** 423-433.

Domagalska, M.A., and Leyser, O. (2011). Signal integration in the control of shoot branching. Nature Reviews Molecular Cell Biology **12,** 211-221.

Druege, U. (2009). Involvement of carbohydrates in survival and adventitious root formation of cuttings within the scope of global horticulture. In Adventitious root formation of forest trees and horticultural plants - From genes to application, K. Niemi and C. Scagel, eds (ResearchSignpost Kerala), pp. 187–208.

Druege, U., Xylaender, M., Zerche, S., and von Alten, H. (2006). Rooting and vitality of poinsettia cuttings was increased by arbuscular mycorrhiza in the donor plants. Mycorrhiza **17,** 67-72.

Druege, U., Franken, P., Lischewski, S., Ahkami, A.H., Zerche, S., Hause, B., and Hajirezaei, M.R. (2014). Transcriptomic analysis reveals ethylene as stimulator and auxin as regulator of adventitious root formation in petunia cuttings. Frontiers in Plant Science **5**.

Drummond, R.S.M., Sheehan, H., Simons, J.L., Martines-Sanchez, N.M., Turner, R.M., Putterill, J., and Snowden, K.C. (2012). The expression of petunia strigolactone pathway genes is altered as part of the endogenous developmental program. Frontiers in Plant Science **2,** 1-14.

Drummond, R.S.M., Martinez-Sanchez, N.M., Janssen, B.J., Templeton, K.R., Simons, J.L., Quinn, B.D., Karunairetnam, S., and Snowden, K.C. (2009). *Petunia hybrida CAROTENOID CLEAVAGE DIOXYGENASE7* is involved in the production of negative and positive branching signals in petunia. Plant Physiol. **151,** 1867-1877.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32,** 1792-1797.

Elser, J.J., Bracken, M.E.S., Cleland, E.E., Gruner, D.S., Harpole, W.S., Hillebrand, H., Ngai, J.T., Seabloom, E.W., Shurin, J.B., and Smith, J.E. (2007). Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecology Letters **10,** 1135-1142.

Favre, P., Bapaume, L., Bossolini, E., Delorenzi, L., Falquet, L., and Reinhardt, D. (2014). A novel bioinformatics pipeline to discover genes related to arbuscular mycorrhizal symbiosis based on their evolutionary conservation pattern among higher plants. BMC Plant Biology **14,** 333.

Feddermann, N., and Reinhardt, D. (2011). Conserved residues in the ankyrin domain of VAPYRIN indicate potential protein-protein interaction surfaces. Plant Signaling and Behavior **6,** 680-684.

Feddermann, N., Duvvuru Muni, R.R., Zeier, T., Stuurman, J., Ercolin, F., Schorderet, M., and Reinhardt, D. (2010). The *PAM1* gene of petunia, required for intracellular accommodation and morphogenesis of arbuscular mycorrhizal fungi, encodes a homologue of VAPYRIN. Plant J. **64,** 470-481.

Floss, D.S., Levy, J.G., Levesque-Tremblay, V., Pumplin, N., and Harrison, M.J. (2013). DELLA proteins regulate arbuscule formation in arbuscular mycorrhizal symbiosis. Proc. Natl. Acad. Sci. U. S. A. **110,** E5025-E5034.

Genre, A., Chabaud, M., Balzergue, C., Puech-Pagès, V., Novero, M., Rey, T., Fournier, J., Rochange, S., Bécard, G., Bonfante, P., and Barker, D.G. (2013). Short-chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear Ca2+ spiking in *Medicago truncatula* roots and their production is enhanced by strigolactone. New Phytol. **198,** 179-189.

Gobbato, E., Marsh, J.F., Vernie, T., Wang, E., Maillet, F., Kim, J., Miller, J.B., Sun, J., Bano, S.A., Ratet, P., Mysore, K.S., Dénarie, J., Schultze, M., and Oldroyd, G.E.D. (2012). A GRAS-type transcription factor with a specific function in mycorrhizal signaling. Current Biology **22,** 2236-2241.

Gomez-Roldan, V., Fermas, S., Brewer, P.B., Puech-Pagès, V., Dun, E.A., Pillot, J.P., Letisse, F., Matusova, R., Danoun, S., Portais, J.C., Bouwmeester, H., Bécard, G., Beveridge, C.A.,

Rameau, C., and Rochange, S.F. (2008). Strigolactone inhibition of shoot branching. Nature **455,** 189-194.

Gough, C., and Cullimore, J. (2011). Lipo-chitooligosaccharide signaling in endosymbiotic plant-microbe interactions. Mol. Plant-Microbe Interact. **24,** 867-878.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology **52,** 696-704.

Gust, A.A., Willmann, R., Desaki, Y., Grabherr, H.M., and Nürnberger, T. (2012). Plant LysM proteins: modules mediating symbiosis and immunity. Trends Plant Sci. **17,** 495-502.

Gutierrez, L., Mongelard, G., Flokova, K., Pacurar, D.I., Novak, O., Staswick, P., Kowalczyk, M., Pacurar, M., Demailly, H., Geiss, G., and Bellini, C. (2012). Auxin controls *Arabidopsis* adventitious root initiation by regulating jasmonic acid homeostasis. Plant Cell **24,** 2515-2527.

Gutjahr, C., and Parniske, M. (2013). Cell and developmental biology of arbuscular mycorrhiza symbiosis. Annual Review of Cell and Developmental Biology **29,** 593-617.

Hamiaux, C., Drummond, R.S.M., Janssen, B.J., Ledger, S.E., Cooney, J.M., Newcomb, R.D., and Snowden, K.C. (2012). DAD2 Is an alpha/beta hydrolase likely to be involved in the perception of the plant branching hormone, strigolactone. Current Biology **22,** 2032-2036.

Harrison, M.J. (2012). Cellular programs for arbuscular mycorrhizal symbiosis. Curr. Opin. Plant Biol. **15,** 691-698.

Hayafune, M., Berisio, R., Marchetti, R., Silipo, A., Kayama, M., Desaki, Y., Arima, S., Squeglia, F., Ruggiero, A., Tokuyasu, K., Molinaro, A., Kaku, H., and Shibuya, N. (2014). Chitin-induced activation of immune signaling by the rice receptor CEBiP relies on a unique sandwich-type dimerization. Proc. Natl. Acad. Sci. U. S. A. **111,** E404-E413.

Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., Hauser, M.T., and Benfey, P.N. (2000). The *SHORT-ROOT* gene controls radial patterning of the *Arabidopsis* root through radial signaling. Cell **101,** 555-567.

Hell, R., and Hillebrand, H. (2001). Plant concepts for mineral acquisition and allocation. Curr. Opin. Biotechnol. **12,** 161-168.

Hirsch, S., and Oldroyd, G.E.D. (2009). GRAS-domain transcription factors that regulate plant development. Plant signaling & behavior **4,** 698-700.

Horstman, A., Willemsen, V., Boutilier, K., and Heidstra, R. (2014). AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks. Trends Plant Sci. **19,** 146-157.

Itoh, H., Ueguchi-Tanaka, M., Sato, Y., Ashikari, M., and Matsuoka, M. (2002). The gibberellin signaling pathway is regulated by the appearance and disappearance of SLENDER RICE1 in nuclei. Plant Cell **14,** 57-70.

Jiang, K., and Feldman, L.J. (2005). Regulation of root apical meristem development. Annual Review of Cell and Developmental Biology **21,** 485-509.

Kaku, H., Nishizawa, Y., Ishii-Minami, N., Akimoto-Tomiyama, C., Dohmae, N., Takio, K., Minami, E., and Shibuya, N. (2006). Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. Proc. Natl. Acad. Sci. U. S. A. **103,** 11086-11091.

Kalo, P., Gleason, C., Edwards, A., Marsh, J., Mitra, R.M., Hirsch, S., Jakab, J., Sims, S., Long, S.R., Rogers, J., Kiss, G.B., Downie, J.A., and Oldroyd, G.E.D. (2005). Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. Science **308,** 1786-1789.

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y., and Martinoia, E. (2011). Plant ABC transporters. In The Arabidopsis book, A.S.o.P. Biologists, ed, pp. e0153.

Kistner, C., and Parniske, M. (2002). Evolution of signal transduction in intracellular symbiosis. Trends Plant Sci. **7,** 511-518.

Klironomos, J.N. (2003). Variation in plant response to native and exotic arbuscular mycorrhizal fungi. Ecology **84,** 2292-2301.

Klopotek, Y., Haensch, K.T., Hause, B., Hajirezaei, M.R., and Druege, U. (2010). Dark exposure of petunia cuttings strongly improves adventitious root formation and enhances carbohydrate availability during rooting in the light. J. Plant Physiol. **167,** 547-554.

Koltai, H. (2014). Receptors, repressors, PINs: a playground for strigolactone signaling. Trends Plant Sci. **19,** 727-733.

Kretzschmar, T., Kohlen, W., Sasse, J., Borghi, L., Schlegel, M., Bachelier, J.B., Reinhardt, D., Bours, R., Bouwmeester, H.J., and Martinoia, E. (2012). A petunia ABC protein controls strigolactone-dependent symbiotic signalling and branching. Nature **483,** 341-346.

Krusell, L., Madsen, L.H., Sato, S., Aubert, G., Genua, A., Szczyglowski, K., Duc, G., Kaneko, T., Tabata, S., de Bruijn, F., Pajuelo, E., Sandal, N., and Stougaard, J. (2002). Shoot control of root development and nodulation is mediated by a receptor-like kinase. Nature **420,** 422-426.

Lavenus, J., Goh, T., Roberts, I., Guyomarc'h, S., Lucas, M., De Smet, I., Fukaki, H., Beeckman, T., Bennett, M., and Laplaze, L. (2013). Lateral root development in *Arabidopsis*: fifty shades of auxin. Trends Plant Sci. **18,** 455-463.

Liu, T.T., Liu, Z.X., Song, C.J., Hu, Y.F., Han, Z.F., She, J., Fan, F.F., Wang, J.W., Jin, C.W., Chang, J.B., Zhou, J.M., and Chai, J.J. (2012). Chitin-induced dimerization activates a plant immune receptor. Science **336,** 1160-1164.

Lohmann, G.V., Shimoda, Y., Nielsen, M.W., Jorgensen, F.G., Grossmann, C., Sandal, N., Sorensen, K., Thirup, S., Madsen, L.H., Tabata, S., Sato, S., Stougaard, J., and Radutoiu, S. (2010). Evolution and regulation of the *Lotus japonicus* LysM receptor gene family. Mol. Plant-Microbe Interact. **23,** 510-521.

Ludwig-Müller, J. (2009). Molecular basis for the role of auxins in adventitious rooting. In Adventitious root formation of forest trees and horticultural plants - From genes to application, K. Niemi and C. Scagel, eds (ResearchSignpost Kerala), pp. 1-29.

Madsen, E.B., Madsen, L.H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T., Tabata, S., Sandal, N., and Stougaard, J. (2003). A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature **425,** 637-640.

Maillet, F., Poinsot, V., André, O., Puech-Pagès, V., Haouy, A., Gueunier, M., Cromer, L., Giraudet, D., Formey, D., Niebel, A., Andres Martinez, E., Driguez, H., Bécard, G., and Dénarié, J. (2011). Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. Nature **469,** 58-64.

Miya, A., Albert, P., Shinya, T., Desaki, Y., Ichimura, K., Shirasu, K., Narusaka, Y., Kawakami, N., Kaku, H., and Shibuya, N. (2007). CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. **104,** 19613-19618.

Miyata, K., Kozaki, T., Kouzai, Y., Ozawa, K., Ishii, K., Asamizu, E., Okabe, Y., Umehara, Y., Miyamoto, A., Kobae, Y., Akiyama, K., Kaku, H., Nishizawa, Y., Shibuya, N., and Nakagawa, T. (2014). The bifunctional plant receptor, OsCERK1, regulates both chitin-triggered immunity and arbuscular mycorrhizal symbiosis in rice. Plant Cell Physiol. **55,** 1864-1872.

Mizukami, Y., and Fischer, R.L. (2000). Plant organ size control: *AINTEGUMENTA* regulates growth and cell numbers during organogenesis. Proc. Natl. Acad. Sci. U. S. A. **97,** 942-947.

Nishimura, R., Hayashi, M., Wu, G.J., Kouchi, H., Imaizumi-Anraku, H., Murakami, Y., Kawasaki, S., Akao, S., Ohmori, M., Nagasawa, M., Harada, K., and Kawaguchi, M. (2002). HAR1 mediates systemic regulation of symbiotic organ development. Nature **420,** 426-429.

Nouri, E., Breuillin-Sessoms, F., Feller, U., and Reinhardt, D. (2014). Phosphorus and nitrogen regulate arbuscular mycorrhizal symbiosis in *Petunia hybrida*. PLoS ONE **9,** e90841.

Oldroyd, G.E.D. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. Nature Reviews Microbiology **11,** 252-263.

Oldroyd, G.E.D., and Long, S.R. (2003). Identification and characterization of Nodulation-Signaling Pathway 2, a gene of *Medicago truncatula* involved in nod factor signaling. Plant Physiol. **131,** 1027-1032.

Oldroyd, G.E.D., Murray, J.D., Poole, P.S., and Downie, J.A. (2011). The rules of engagement in the legume-rhizobial symbiosis. Annual Review of Genetics **45,** 119-144.

Op den Camp, R., Streng, A., De Mita, S., Cao, Q., Polone, E., Liu, W., Ammiraju, J.S.S., Kudrna, D., Untergasser, A., Bisseling, T., and Geurts, R. (2011). LysM-type mycorrhizal receptor recruited for *Rhizobium* symbiosis in nonlegume *Parasponia*. Science **331,** 909-912.

Parniske, M. (2008). Arbuscular mycorrhiza: the mother of plant root endosymbioses. Nature Reviews Microbiology **6,** 763-775.

Peng, J.R., Carol, P., Richards, D.E., King, K.E., Cowling, R.J., Murphy, G.P., and Harberd, N.P. (1997). The *Arabidopsis GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. Genes Dev. **11,** 3194-3205.

Perilli, S., Di Mambro, R., and Sabatini, S. (2012). Growth and development of the root apical meristem. Curr. Opin. Plant Biol. **15,** 17-23.

Petricka, J.J., Winter, C.M., and Benfey, P.N. (2012). Control of *Arabidopsis* root development. Annual Review of Plant Biology **63,** 563-590.

Petutschnig, E.K., Jones, A.M.E., Serazetdinova, L., Lipka, U., and Lipka, V. (2010). The lysin motif receptor-like kinase (LysM-RLK) CERK1 Is a major chitin-binding protein in Arabidopsis thaliana and subject to chitin-induced phosphorylation. J. Biol. Chem. **285,** 28902-28911.

Pop, T.I., Pamfil, D., and Bellini, C. (2011). Auxin control in the formation of adventitious roots. Notulae Botanicae Horti Agrobotanici Cluj-Napoca **39,** 307-316.

Pumplin, N., Mondo, S.J., Topp, S., Starker, C.G., Gantt, J.S., and Harrison, M.J. (2010). *Medicago truncatula* Vapyrin is a novel protein required for arbuscular mycorrhizal symbiosis. Plant J. **61,** 482-494.

Radutoiu, S., Madsen, L.H., Madsen, E.B., Felle, H.H., Umehara, Y., Gronlund, M., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., and Stougaard, J. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature **425,** 585-592.

Rasmussen, A., Mason, M.G., De Cuyper, C., Brewer, P.B., Herold, S., Agusti, J., Geelen, D., Greb, T., Goormachtig, S., Beeckman, T., and Beveridge, C.A. (2012). Strigolactones suppress adventitious rooting in *Arabidopsis* and pea. Plant Physiol. **158,** 1976-1987.

Rich, M.K., Schorderet, M., Bapaume, L., Falquet, L., Morel, P., Vandenbussche, M., and Reinhardt, D. (2015). A petunia GRAS transcription factor controls symbiotic gene expression and fungal morphogenesis in arbuscular mycorrhiza. Plant Physiol **in press**.

Rigal, A., Yordanov, Y.S., Perrone, I., Karlberg, A., Tisserant, E., Bellini, C., Busov, V.B., Martin, F., Kohler, A., Bhalerao, R., and Legue, V. (2012). The AINTEGUMENTA LIKE1 homeotic transcription factor PtAIL1 controls the formation of adventitious root primordia in poplar. Plant Physiol. **160,** 1996-2006.

Ruyter-Spira, C., Al-Babili, S., van der Krol, S., and Bouwmeester, H. (2013). The biology of strigolactones. Trends Plant Sci. **18,** 72-83.

Sabatini, S., Heidstra, R., Wildwater, M., and Scheres, B. (2003). *SCARECROW* is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. Genes Dev. **17,** 354-358.

Sanchez, C., Vielba, J.M., Ferro, E., Covelo, G., Sole, A., Abarca, D., De Mier, B.S., and Diaz-Sala, C. (2007). Two *SCARECROW-LIKE* genes are induced in response to exogenous auxin in rooting-competent cuttings of distantly related forest species. Tree Physiology **27,** 1459-1470.

Sasse, J., Simon, S., Guebeli, C., Liu, G.-W., Cheng, X., Frimi, J., Bouwmeester, H., Martinoia, E., and Borghi, L. (2015). Asymmetric localizations of the ABC transporter PaPDR1 trace paths of directional strigolactone transport. Current Biology **25,** 647-655.

Seto, Y., and Yamaguchi, S. (2014). Strigolactone biosynthesis and perception. Curr. Opin. Plant Biol. **21,** 1-6.

Shi, J.X., Malitsky, S., De Oliveira, S., Branigan, C., Franke, R.B., Schreiber, L., and Aharoni, A. (2011). SHINE transcription factors act redundantly to pattern the archetypal surface of *Arabidopsis* flower organs. Plos Genetics **7,** e1001388.

Shimizu, T., Nakano, T., Takamizawa, D., Desaki, Y., Ishii-Minami, N., Nishizawa, Y., Minami, E., Okada, K., Yamane, H., Kaku, H., and Shibuya, N. (2010). Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. Plant J. **64,** 204-214.

Shinohara, N., Taylor, C., and Leyser, O. (2013). Strigolactone can promote or inhibit shoot branching by triggering rapid depletion of the auxin efflux protein PIN1 from the plasma membrane. Plos Biology **11**.

Shinya, T., Motoyama, N., Ikeda, A., Wada, M., Kamiya, K., Hayafune, M., Kaku, H., and Naoto, S. (2012). Functional characterization of CEBiP and CERK1 homologs in *Arabidopsis* and rice reveals the presence of different chitin receptor systems in plants Plant Cell Physiol. **53,** 1696-1706.

Smit, P., Raedts, J., Portyanko, V., Debelle, F., Gough, C., Bisseling, T., and Geurts, R. (2005). NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. Science **308,** 1789-1791.

Smith, S.E., and Read, D.J. (2008). Mycorrhizal Symbiosis. (New York: Academic Press).

Snowden, K.C., and Napoli, C.A. (2003). A quantitative study of lateral branching in petunia. Functional Plant Biology **30,** 987-994.

Snowden, K.C., Simkin, A.J., Janssen, B.J., Templeton, K.R., Loucas, H.M., Simons, J.L., Karunairetnam, S., Gleave, A.P., Clark, D.G., and Klee, H.J. (2005). The *Decreased apical dominance 1*/*Petunia hybrida carotenoid cleavage dioxygenase8* gene affects branch production and plays a role in leaf senescence, root growth, and flower development. Plant Cell **17,** 746-759.

Sorin, C., Bussell, J.D., Camus, I., Ljung, K., Kowalczyk, M., Geiss, G., McKhann, H., Garcion, C., Vaucheret, H., Sandberg, G., and Bellini, C. (2005). Auxin and light control of adventitious rooting in *Arabidopsis* require *ARGONAUTE1*. Plant Cell **17,** 1343-1359.

Sparks, E., Wachsman, G., and Benfey, P.N. (2013). Spatiotemporal signalling in plant development. Nature Reviews Genetics **14,** 631-644.

Stuurman, J., Jäggi, F., and Kuhlemeier, C. (2002). Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. Genes Dev. **16,** 2213-2218.

Tian, C.G., Wan, P., Sun, S.H., Li, J.Y., and Chen, M.S. (2004). Genome-wide analysis of the GRAS gene family in rice and *Arabidopsis*. Plant Mol.Biol. **54,** 519-532.

To, A., Joubes, J., Barthole, G., Lecureuil, A., Scagnelli, A., Jasinski, S., Lepiniec, L., and Baud, S. (2012). WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in *Arabidopsis*. Plant Cell **24,** 5007-5023.

Tyler, L., Thomas, S.G., Hu, J.H., Dill, A., Alonso, J.M., Ecker, J.R., and Sun, T.P. (2004). DELLA proteins and gibberellin-regulated seed germination and floral development in *Arabidopsis*. Plant Physiol. **135,** 1008-1019.

Umehara, M., Hanada, A., Yoshida, S., Akiyama, K., Arite, T., Takeda-Kamiya, N., Magome, H., Kamiya, Y., Shirasu, K., Yoneyama, K., Kyozuka, J., and Yamaguchi, S. (2008). Inhibition of shoot branching by new terpenoid plant hormones. Nature **455,** 195-200.

Van Norman, J.M., Xuan, W., Beeckman, T., and Benfey, P.N. (2013). To branch or not to branch: the role of pre-patterning in lateral root formation. Development **140,** 4301-4310.

Wan, J., Tanaka, K., Zhang, X.-C., Son, G.H., Brechenmacher, L., Nguyen, T.H.N., and Stacey, G. (2012). LYK4, a LysM receptor-like kinase, is important for chitin signaling and plant innate immunity in *Arabidopsis*. Plant Physiol. **160,** 396-406.

Wang, B., Yeun, L.H., Xue, J.Y., Liu, Y., Ane, J.M., and Qiu, Y.L. (2010). Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. New Phytol. **186,** 514-525.

Wang, E.T., Schornack, S., Marsh, J.F., Gobbato, E., Schwessinger, B., Eastmond, P., Schultze, M., Kamoun, S., and Oldroyd, G.E.D. (2012). A common signaling process that promotes mycorrhizal and oomycete colonization of plants. Current Biology **22,** 2242-2246.

Wegmüller, S., Svistoonoff, S., Reinhardt, D., Stuurman, J., Amrhein, N., and Bucher, M. (2008). A transgenic dTph1 insertional mutagenesis system for forward genetics in mycorrhizal phosphate transport of Petunia. Plant J. **54,** 1115-1127.

Xie, X., Wang, G., Yang, L., Cheng, T., Gao, J., Wu, Y., and Xia, Q. (2015). Cloning and characterization of a novel *Nicotiana tabacum* ABC transporter involved in shoot branching. Physiol. Plant. **153,** 299-306.

Xu, J., Hofhuis, H., Heidstra, R., Sauer, M., Friml, J., and Scheres, B. (2006). A molecular framework for plant regeneration. Science **311,** 385-388.

Xue, L., Cui, H., Buer, B., Vijayakumar, V., Delaux, P.-M., Junkermann, S., and Bucher, M. (2015). Network of GRAS transcription factors involved in the control of arbuscule development in *Lotus japonicus*. Plant Physiol **167,** 854-871.

Yu, N., Luo, D.X., Zhang, X.W., Liu, J.Z., Wang, W.X., Jin, Y., Dong, W.T., Liu, J.Y., Liu, H., Yang, W.B., Zeng, L.J., Li, Q., He, Z.H., Oldroyd, G.E.D., and Wang, E. (2014). A DELLA protein complex controls the arbuscular mycorrhizal symbiosis in plants. Cell Research **24,** 130-133.

Zhang, Q., Blaylock, L.A., and Harrison, M.J. (2010). Two *Medicago truncatula* half-ABC transporters are essential for arbuscule development in arbuscular mycorrhizal symbiosis. Plant Cell **22,** 1483-1497.

Zhang, X., Dong, W., Sun, J., Feng, F., Deng, Y., He, Z., Oldroyd, G.D.E., and Wang, E. (2015). The receptor kinase CERK1 has dual functions in symbiosis and immunity signalling. Plant J. **81,** 258–267.

Zhang, X.C., Cannon, S.B., and Stacey, G. (2009). Evolutionary genomics of LysM genes in land plants. BMC Evolutionary Biology **9,** 183.

Supplementary note 11

Characterization of SI Loci in *P. inflata* and *P. axillaris* identifies multiple linked S-locus F-box genes.

Diwa Malla[a], Qinzhou Qi[a] Timothy P. Robbins[b] and Thomas L. Sims[a1]

[a]Department of Biological Sciences, Northern Illinois University, DeKalb IL 60115
[b]Department of Plant and Crop Sciences, University of Nottingham, Sutton Bonington, Leicestershire, UK LE12 5RD

[1]Address correspondence to tsims@niu.edu

Running title: S-locus sequences in *Petunia*

ABSTRACT

Petunia species exhibit S-RNase-based gametophytic self-incompatibility (GSI), in which plants are capable of recognizing and rejecting their own (self) pollen, while accepting pollen from a different (non-self) individual. Gametophytic self-incompatibility is found in both progenitor species (*Petunia inflata* and *Petunia axillaris* as well as in the cultivated hybrid *Petunia hybrida*). Genes critical for self versus non-self recognition are the S-locus ribonuclease (S-RNase), which is the style-recognition component of GSI, and S-locus F-box (SLF) genes, which encode the pollen-recognition component of GSI. Although any individual plant possesses at most two S-RNase alleles, multiple SLF genes appear to be involved in recognizing an overlapping range of non-self S-RNase proteins. Both S-RNase and SLF genes are thought to be tightly linked at the S-locus as pollen-recognition and style-recognition specificities do not recombine. Here, we have used BLASTn and BLASTp searches of the assembled sequences of *Petunia inflata* S6 and *Petunia axillaris* N to identify and characterize S-RNase and SLF genes in both species. Each genome contains a single S-RNase allele, as expected for these homozygous lines. The S-RNase allele in *Petunia inflata* appears to be a previously uncharacterized S-RNase allele, whereas the S-RNase allele in *Petunia axillaris* is identical to the previously reported S_1-RNase of *Petunia hybrida.* The latter observation confirms that self-incompatibility in *Petunia hybrida* can be inherited from both progenitor species. We have identified 29 putative S-Locus F-box genes in *Petunia inflata*, the majority of which are present as two closely related copies and 19 SLF genes in the genome of *Petunia axillaris*. The SLF genes in these two species represent at least 20 different SLF variants. In both species, multiple SLF genes are linked on the same scaffold, and in *Petunia axillaris* SLF10 is linked to the S_{ax1}-RNase. These data provide a valuable resource for future studies to assemble and characterize an entire S-locus, as well as to determine the molecular basis of self versus non-self recognition in GSI.

INTRODUCTION

Gametophytic self-incompatibility (GSI), a genetic mechanism that acts to prevent self-fertilization in many angiosperms, is based on the ability of the pistil to selectively inhibit growth of “self” pollen while allowing the growth of “non-self” pollen. During pollination, pollen grains germinate on the stigmatic surface, producing pollen tubes that enter the pistil and begin growth through the transmitting tract of the style. In incompatible pollinations in *Petunia*, pollen tube growth is arrested in the upper third of the style via the action of the S-RNase, an abundant, style-expressed ribonuclease that is imported into both incompatible and compatible pollen tubes. Current models for GSI propose that pollen-tube growth inhibition is due to the cytotoxic action of the S-RNase, which degrades pollen tube RNA, inhibiting protein synthesis. During a compatible pollination, the action of the S-RNase is inhibited by a SCF^{SLF} E3 ubiquitin ligase complex. This complex is proposed to contain one or more F-box proteins (SLF), a Culllin protein, the Skp1-like protein SSK1 and either Rbx1 or SBP1 RING-domain proteins (Zhao et al., 2010; Meng et al., 2011; Sims, 2012; Li et al., 2014).

Estimated to occur in up to three-quarters of eudicot families (Igic & Kohn 2001), S-RNase-based Gametophytic self-incompatibility (GSI) was first described in *Petunia hybrida*, by Darwin:

> “....protected flowers with their own pollen placed on the stigma never yielded nearly a full complement of seed; whilst those left uncovered produced fine capsules, showing that pollen from other plants must have been brought to them, probably by moths. Plants growing vigorously and flowering in pots in the green-house, never yielded a single capsule; and this may be attributed, at least in chief part, to the exclusion of moths.” (Darwin, 1891).

Petunia, primarily *Petunia hybrida, Petunia inflata* and *Petunia axillaris* has been one of the primary systems used to study the genetics and mechanism of S-RNase-based self-incompatibility ever since that first description (Sims & Robbins 2009). Researchers such as Mather (1943), Linskens (1975), and de Nettancourt (1977), determined that gametophytic SI in Petunia was governed by a single, multiallelic S-locus, and that recognition and rejection of self-pollen was controlled gametophytically by alleles expressed in pollen. Mutations unilaterally inactivating self-incompatibility in pollen (pollen-part mutations) have been identified in *Petunia inflata* and were associated with centric chromosomal fragments (Brewbaker and Natarajan, 1960). Tetraploid plants with diploid heteroallelic pollen have been shown to be self-compatible, due to "competitive interaction' in pollen. Shivanna and Rangaswamy (1969) demonstrated that pollination of immature styles could be used to overcome self-incompatibility, a phenomenon now understood to result from low-level expression of the S-RNase early in the development of the style. Ascher (1984) demonstrated quantitative variation in the strength of the self-incompatibility reaction, which he termed pseudo-self-compatibility. More recently, key genes involved in regulating recognition and rejection of pollen in GSI were either first discovered or functionally tested in *Petunia* (Clark et al., 1990; Ioerger et al., 1991; Sims & Ordanic 2001; Lee et al., 1994; Qiao et al., 2004; Sijacic et al., 2004; Kubo et al., 2010; Zhao et al., 2010; Li et al., 2014; Willams et al., 2014a, 2014b; Kubo et al., 2015).

To date, progress on investigating certain aspects of GSI function and mechanisms has been limited due to the lack of assembled and annotated genome and transcriptome sequences for *Petunia.* Kubo et al., (2010) and Williams et al., (2014a) demonstrated that multiple S-locus F-box (SLF) proteins acted to recognize and inhibit overlapping sets of non-self S-RNase. Although different approaches had previously identified up to 10 SLF variants, several questions regarding the role of SLF proteins cannot be answered without a high-quality assembled genome. Among these questions are: How may SLF genes are found in the *Petunia* genome? Are all SLF variants present and expressed in different S-haplotypes and/or *Petunia* species? Are all SLF variants linked to the S-locus and to each other? What is the sequence variability of different SLF alleles and SLF variants, and can that sequence variability be used to identify protein regions involved in binding to the S-RNase? Other questions about the organization and expression of GSI require a high-quality assembled genome. In previous work, Wang et al (2003, 2004) attempted to use chromosome walking approaches to characterize the S-locus of S_2-*Petunia inflata*. Although these authors estimated the size of the S-locus in *Petunia inflata* as at least 4.4 Mb, and were able to identify and partially sequence an 881 kb contig containing the S_2-SLF1 and S_2-RNase genes linked within 160 kb, they were unable to fully assemble the complete S-locus due to the presence of highly repetitive DNA sequences. The S-locus in the Solanaceae has been shown to be subcentromeric (Tanksley & Loaiza-Figueroa 1985; ten Hoopen et al., 1998; Entani et al., 2000) and is in a region of suppressed recombination, which in *Petunia*, is located on chromosome III. Thus, having a fully sequenced and assembled genome should aid in approaches to fully assemble and characterize a complete S-locus for *Petunia.*

We used BLASTn and BLASTp queries of the assembled genomic DNA sequences for *Petunia inflata* S6 and *Petunia axillaris* N followed by manual annotation and comparison to known sequences to identify S-RNase and SLF genes in both genomes. These approaches confirmed the existence of known SLF variants in *Petunia* and identified a number of new SLF variants. We identified 29 putative S-Locus F-box genes in *Petunia inflata* and 19 SLF genes in the genome of *Petunia axillaris*, representing at least 20 different SLF variants. In both species, multiple SLF genes are linked on the same scaffold, and in *Petunia axillaris* SLF10 is linked to the S_{ax1}-RNase. The

S-RNase in S6 *Petunia inflata* appears to be a previously uncharacterized S-RNase. In a surprising and fortuitous observation, we found that the S-RNase gene in *Petunia axillaris* N is identical to that in S_1-*Petunia hybrida* (Clark et al., 1990). This finding strongly suggests that the S-haplotypes are identical in these two lines, and indicates that self-incompatibility in *Petunia hybrida* can be inherited from both parents (Ando et al., 1998). Together with the characterization of other genes involved in GSI interactions, the sequence data obtained from the Petunia Genome Project will provide invaluable resources for further investigations into the organization, expression and function of genes governing self versus non-self recognition in gametophytic self-incompatibility.

RESULTS

Identification of the S-RNase allele in *Petunia axillaris*

BLASTn and BLASTp searches of the *Petunia axillaris* assembled genome gave a single hit that showed 100% DNA sequence identity to the S_1-RNase mRNA of *Petunia hybrida* (Clark et al., 1990). Comparison of this genomic DNA region with the sequenced S_1-RNase gene of *Petunia hybrida* (U07362) showed nearly 100% DNA sequence identity (3119 of 3221 bases) across the entire coding region, and 99% identity when 5' and 3' flanking regions are included (Figure 1). The position of the intron and splice sites were identical between the S_1-RNase gene of *Petunia hybrida* and its homolog in *Petunia axillaris.* In spite of these minor differences, the predicted amino acid sequence for S_1-*Petunia hybrida* and the S-RNase of *Petunia axillaris* are identical (Figure 2). We have therefore tentatively named the S-RNase gene in *Petunia axillaris* N as S_{ax1}-RNase to illustrate this identity. Because the S-RNase genes in S_1 *Petunia hybrida* and S_{ax1} *Petunia axillaris* are identical, this finding demonstrates that self-incompatibility in *Petunia hybrida* can be inherited from both of the progenitor species (e.g., Ando et 1998). This finding also suggests that the S-locus of S_1 *Petunia hybrida* and that of S_{ax1} *Petunia axillaris* encode the same haplotype. Indeed, PCR primers based on SLF gene sequences from *Petunia axillaris* (see below) have been successfully used to clone multiple SLF variant genes from a pollen cDNA library of S_1S_1 *Petunia hybrida* (Sims and Ordanic 2001; Qi & Sims, in preparation). Figures 1 and 2 show DNA and protein sequence alignments of the S_{ax1} and S_1-RNases.

```
Sax1    1   AAGCTTTGCACTTTCTGTTGACTGTATGCCATTGTTACCACCTTAAAGGAGCTATCCATT  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1    1   AAGCTTTGCACTTTCTGTTGACTGTATGCCATTGTTACCACCTTAAAGGAGCTATCCATT  60

Sax1   61   TTTTCCATGAACATGCTGAGAAAAGGACCAAATGAAGGAAACTGGACAAGATCAGAGTCA  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1   61   TTTTCCATGAACATGCTGAGAAAAGGACCAAATGAAGGAAACTGGACAAGATCAGAGTCA  120

Sax1  121   AAGGTGTATTACAAAAGATTATCCTTCTCTCCTGCTATGGACTTCCATCTCAGGAAAAAG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  121   AAGGTGTATTACAAAAGATTATCCTTCTCTCCTGCTATGGACTTCCATCTCAGGAAAAAG  180

Sax1  181   ATAGAGCTTTGTTCATGTACTCACATCCCAACTCTAGAAAATGTTCCAACTTAGTTGTAA  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  181   ATAGAGCTTTGTTCATGTACTCACATCCCAACTCTAGAAAATGTTCCAACTTAGTTGTAA  240

Sax1  241   ACCTGATATAGTTTACAAATTATCGTATTGAATATTCCCTCAGTGATACTTCATTATCTT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  241   ACCTGATATAGTTTACAAATTATCGTATTGAATATTCCCTCAGTGATACTTCATTATCTT  300

Sax1  301   CTCTATGTAGTTGACCGACCATTTAATTGTAAAAGTATTATGTTTCAATGAAAGGTGAAG  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  301   CTCTATGTAGTTGACCGACCATTTAATTGTAAAAGTATTATGTTTCAATGAAAGGTGAAG  360

Sax1  361   AGTTAAATTCTTGTACAGGATGTCGAGGGAGGAAAATGATTCATCTCCCATAGCAACTCA  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  361   AGTTAAATTCTTGTACAGGATGTCGAGGGAGGAAAATGATTCATCTCCCATAGCAACTCA  420

Sax1  421   TGAAGCCTTGAAATTTATCGAGTAAATATAttttttctttttttACTTTGTGCTGATCAATT  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  421   TGAAGCCTTGAAATTTATCGAGTAAATATATTTTTCTTTTTTACTTTGTGCTGATCAATT  480

Sax1  481   AACATTTTCTTAATCTTGTATAATCTACTACCACTCATATTATGTGTAAATGTTTCTATT  540
            |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
PhS1  481   AACATTTTCTTAATCTTGTATAA-CTACTACCACTCATATTATGTGTAAATGTTTCTATT  539
                                  *
Sax1  541   CTAAATCTTAAAATTCTCGTGTTGCACGGGCATCAAAAATATTAATTTATTTTTCTAAAT  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  540   CTAAATCTTAAAATTCTCGTGTTGCACGGGCATCAAAAATATTAATTTATTTTTCTAAAT  599

Sax1  601   TTACATAATTTAATCATGCTACAACCATTCGTAACTTTCTATAATGATTTACAatatata  660
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
PhS1  600   TTACATAATTTAATCATGCTACAACCATTTGTAACTTTCTATAATGATTTACAATATATA  659
                                        *
Sax1  661   tatatatatatatatatatata----------------------tatatatatatata  696
            ||||||||||||||||||||||                      ||||||||||||||
PhS1  660   TATATATATATATATATATATATATATATATATATATATATATTTATATATATATATA  719

Sax1  697   tatttatatatatataaaacatactatctattcaatatatatatgtgatatttttaaatt  756
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  720   TATTTATATATATATAAAACATACTATCTATTCAATATATATATGTGATATTTTTAAATT  779



Sax1  757   gaatatatatatttatataatatttttaacatattgtatatatagtatttattcaattaa  816
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  780   GAATATATATATTTATATAATATTTTTAACATATTGTATATATAGTATTTATTCAATTAA  839
```

```
Sax1  817   tatttttaacatactatatattcaatatttatttaatattttaacacaatatatcacttt  876
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  840   TATTTTTAACATACTATATATTCAATATTTATTTAATATTTTAACACAATATATCACTTT  899

Sax1  877   ttgtatatgttgtgtattCAACATTTATTTACATATGTTTTCTTCCCACACAGCTAAATT  936
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  900   TTGTATATGTTGTGTATTCAACATTTATTTACATATGTTTTCTTCCCACACAGCTAAATT  959

Sax1  937   GTAAGGTATTTGAAATAAAATCAGGGCCATCTGTTGTCGGCAACTGAGCCATCCACCTGT  996
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  960   GTAAGGTATTTGAAATAAAATCAGGGCCATCTGTTGTCGGCAACTGAGCCATCCACCTGT  1019

Sax1  997   CTACAACAACAACAACATACCTAGTGTAATTCCATAAGTGGGTTTAGGAAACTGAGATGT  1056
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
PhS1  1020  CTACAACAACAACAACATACCTAGTGTAATTCCATAAGTGGATTTAGGAAACTGAGATGT  1079
                                               *
Sax1  1057  ACGCAGAACTTACCCCACCAGAATGAAGAGATTGTTTCCGAAAGACCCTCGGCTaaaaaa  1116
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1080  ACGCAGAACTTACCCCACCAGAATGAAGAGATTGTTTCCGAAAGACCCTCGGCTAAAAAA  1139

Sax1  1117  aCATATTTGAAAttttttttAAGACAAACCAAATATTTTGaaaaaagataaaaacgaata  1176
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1140  ACATATTTGAAATTTTTTTTAAGACAAACCAAATATTTTGAAAAAAGATAAAAACGAATA  1199

Sax1  1177  agttaaaaaaTACCATTAATGCTCAAGTTCTCATATAAAAACTACATAACCAGGGAAAAC  1236
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1200  AGTTAAAAAATACCATTAATGCTCAAGTTCTCATATAAAAACTACATAACCAGGGAAAAC  1259

Sax1  1237  ACAAGACGTCAATAGCGATAAGGACAAGATAAAACTACTATGCATAAGAATACTACCGCT  1296
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1260  ACAAGACGTCAATAGCGATAAGGACAAGATAAAACTACTATGCATAAGAATACTACCGCT  1319

Sax1  1297  AAAATGTCAACAATCAATCGTCTTCTACCTAACCTTCTACCATAATCCCAGACCTCCACG  1356
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1320  AAAATGTCAACAATCAATCGTCTTCTACCTAACCTTCTACCATAATCCCAGACCTCCACG  1379

Sax1  1357  CTTTCCTGTCAATGGTCATGTCCTCGGTGATCTAGATGTGAGTCATATCATGTCGAATCA  1416
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1380  CTTTCCTGTCAATGGTCATGTCCTCGGTGATCTAGATGTGAGTCATATCATGTCGAATCA  1439

Sax1  1417  CCTCGCCCCAATTCTTCTTTGGTCTACCTCTACCTCTCTGCAGACCTAGCACAACCAGCC  1476
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1440  CCTCGCCCCAATTCTTCTTTGGTCTACCTCTACCTCTCTGCAGACCTAGCACAACCAGCC  1499

Sax1  1477  TCTCACACCTCCTCACTGGCGCATCGGTGCCCCTTCTCTTTGCATGGCCGAACCATCGCA  1536
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1500  TCTCACACCTCCTCACTGGCGCATCGGTGCCCCTTCTCTTTGCATGGCCGAACCATCGCA  1559

Sax1  1537  ATCTCACTTCTCGCATCTTGTCCACTACTGAGGCCACTCTCACCTTATCACGAATGGCCT  1596
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1560  ATCTCACTTCTCGCATCTTGTCCACTACTGAGGCCACTCTCACCTTATCACGAATGGCCT  1619

Sax1  1597  CATTCCTAATCTTATCCAACTCGTGTGCCCACACAGCCATCTAAGCTGTTAACATACATA  1656
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1620  CATTCCTAATCTTATCCAACTCGTGTGCCCACACAGCCATCTAAGCTGTTAACATACATA  1679

Sax1  1657  TTATTTTAACATATCCAACCTCATGTGTCCACACATCCACGTGTCTCATATATATTATTT  1716
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1680  TTATTTTAACATATCCAACCTCATGTGTCCACACATCCACGTGTCTCATATATATTATTT  1739
```

```
Sax1  1717  TTAAAAGGCTGTTAACACTATTTAAGTGCCCTAATCCACTTCTTCTAGTACAGATTCTAG  1776
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1740  TTAAAAGGCTGTTAACACTATTTAAGTGCCCTAATCCACTTCTTCTAGTACAGATTCTAG  1799

Sax1  1777  CTTGaaaaaaaaaTTAGTTAGGTTAGTGAAGTGATAAGTCCTATATTAACCCATTCCACT  1836
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1800  CTTGAAAAAAAAAATTAGTTAGGTTAGTGAAGTGATAAGTCCTATATTAACCCATTCCACT  1859

Sax1  1837  GAAAATACGATTATTATATATTGCCATATAGCAAAAGGAAGGAACTAAACATGAGTTGTT  1896
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1860  GAAAATACGATTATTATATATTGCCATATAGCAAAAGGAAGGAACTAAACATGAGTTGTT  1919

Sax1  1897  CAAACTTTAGAATGTTCAAGTTACAGCTGGCGTCAGTTTTATGTGTTTTTCTTTTTGCTT  1956
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1920  CAAACTTTAGAATGTTCAAGTTACAGCTGGCGTCAGTTTTATGTGTTTTTCTTTTTGCTT  1979

Sax1  1957  GCTCTCCAATTTCTGGGTCTTTCGACCACTGGCAACTCGTTTTAACATGGCCTGCAGGTT  2016
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1980  GCTCTCCAATTTCTGGGTCTTTCGACCACTGGCAACTCGTTTTAACATGGCCTGCAGGTT  2039

Sax1  2017  ATTGCAAAGTTAAAGGTTGTCCGAGACCAGTAATTCCGAACGACTTTACTATTCATGGTC  2076
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2040  ATTGCAAAGTTAAAGGTTGTCCGAGACCAGTAATTCCGAACGACTTTACTATTCATGGTC  2099

Sax1  2077  TTTGGCCAGATAGCATTTCCGTCATAATGAATAACTGCGATCCGACTAAAACGTTTGTGA  2136
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2100  TTTGGCCAGATAGCATTTCCGTCATAATGAATAACTGCGATCCGACTAAAACGTTTGTGA  2159

Sax1  2137  CGATCACTGTAAGTTTATAACATTATCTTCTTAAGCGATTGTAAttttttttttctcatt  2196
            |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
PhS1  2160  CGATCACTGTAAGTTTATAACATTATCTTCTTAAGCGATTGTAAATTTTTTTTTCTCATT  2219
                                                        *
Sax1  2197  tattgttttgctttttcctttctttttatttttgtttcttgAATAACCTGCAGCCTAATGT  2256
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2220  TATTGTTTTGCTTTTTCCTTTCTTTTTATTTTGTTTCTTGAATAACCTGCAGCCTAATGT  2279

Sax1  2257  TTATAGGAAATAAATCAAATAACCGAACTGGAGAAGCGCTGGCCTGAATTGACTACTACC  2316
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2280  TTATAGGAAATAAATCAAATAACCGAACTGGAGAAGCGCTGGCCTGAATTGACTACTACC  2339

Sax1  2317  GCACAATTTGCTTTAACGAGTCAATCTTTCTGGAGATATCAATACGAAAAGCATGGAACA  2376
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2340  GCACAATTTGCTTTAACGAGTCAATCTTTCTGGAGATATCAATACGAAAAGCATGGAACA  2399

Sax1  2377  TGTTGTTTTCCTGTCTACAGTCAATCAGCATATTTTGATTTTGCTATAAAATTAAAAGAC  2436
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2400  TGTTGTTTTCCTGTCTACAGTCAATCAGCATATTTTGATTTTGCTATAAAATTAAAAGAC  2459

Sax1  2437  AAGACTGATCTGTTGAGTATTCTCAGAAGTCAAGGTGTTACTCCGGGATCAACTTATACT  2496
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2460  AAGACTGATCTGTTGAGTATTCTCAGAAGTCAAGGTGTTACTCCGGGATCAACTTATACT  2519

Sax1  2497  GGAGAAAGAATCAACAGTTCCATCGCGTCAGTAACCCGAGTGAAACCTAACCTCAAGTGC  2556
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2520  GGAGAAAGAATCAACAGTTCCATCGCGTCAGTAACCCGAGTGAAACCTAACCTCAAGTGC  2579
```

```
Sax1  2557  CTTTATTATCGAGGCAAATTGGAATTAACTGAGATAGGAATATGTTTTGACCGAACGACA  2616
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2580  CTTTATTATCGAGGCAAATTGGAATTAACTGAGATAGGAATATGTTTTGACCGAACGACA  2639

Sax1  2617  GTTGCTATGATGTCGTGTCCTCGGATTAGTACGTCATGCAAATTCGGGACAAATGCGAGG  2676
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2640  GTTGCTATGATGTCGTGTCCTCGGATTAGTACGTCATGCAAATTCGGGACAAATGCGAGG  2699

Sax1  2677  ATTACGTTTCGACAGTGAGAAACGTTCGATTTCATGTTCTTTCTTTCTAATTTTATGCAG  2736
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2700  ATTACGTTTCGACAGTGAGAAACGTTCGATTTCATGTTCTTTCTTTCTAATTTTATGCAG  2759

Sax1  2737  AGTATAATAAAGGAGGTTTTACACTGTATACCCAAATTTTATAAAATATTACTATATATT  2796
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2760  AGTATAATAAAGGAGGTTTTACACTGTATACCCAAATTTTATAAAATATTACTATATATT  2819

Sax1  2797  TTTACTCATTTTAAACTTCATTTCACTTCATGCCCATTTTGACATAAATATTTATACTCC  2856
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2820  TTTACTCATTTTAAACTTCATTTCACTTCATGCCCATTTTGACATAAATATTTATACTCC  2879

Sax1  2857  ATGTCCATTTTGACATAAATATTTACACAACTATCCATGGTACAAAAAGTTCACCTTTTA  2916
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2880  ATGTCCATTTTGACATAAATATTTACACAACTATCCATGGTACAAAAAGTTCACCTTTTA  2939

Sax1  2917  GCCACGATGAACCGTAGATAGTGCCACCGTGGCTACCGGTCGCCATTAGCCACGATTAAT  2976
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2940  GCCACGATGAACCGTAGATAGTGCCACCGTGGCTACCGGTCGCCATTAGCCACGATTAAT  2999

Sax1  2977  CATGGCTAATGGCCGCGTTTAGCCGTGGTCTTACAATAGGTTCTGGAACCGAGGCTAACA  3036
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  3000  CATGGCTAATGGCCGCGTTTAGCCGTGGTCTTACAATAGGTTCTGGAACCGAGGCTAACA  3059

Sax1  3037  ACCTATAGATTTCATGACCCAATTAACATTTACCGTGTAAGACCCATGTCTAatttaatt  3096
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  3060  ACCTATAGATTTCATGACCCAATTAACATTTACCGTGTAAGACCCATGTCTAATTTAATT  3119

Sax1  3097  tttatattttactatttttatattaATCACGGTTCAACCGTGGCTTAGTAACCCCTTCTT  3156
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  3120  TTTATATTTTACTATTTTTATATTAATCACGGTTCAACCGTGGCTTAGTAACCCCTTCTT  3179

Sax1  3157  aaaaaaaaTTTAATATTTAAATGAAAGACAGTTGCTGAATTC  3198
            |||||||||||||||||||||||||||||||||||||||||||
PhS1  3180  AAAAAAAATTTAATATTTAAATGAAAGACAGTTGCTGAATTC  3221
```

Figure 1. Alignment of the genomic DNA sequences of the S_1-RNase gene of *Petunia hybrida* (PhS1) and the homologous region for S_{ax1}-*Petunia axillaris* (Sax1). Start and stop codons for the S_1-RNase are shown in boldface and the exon regions of the S_1-RNase are underlined. Individual base differences are shown with an asterisk.

```
Sax1    1  MFKLQLASVLCVFLFACSPISGSFDHWQLVLTWPAGYCKVKGCPRPVIPNDFTIHGLWPD  60
           MFKLQLASVLCVFLFACSPISGSFDHWQLVLTWPAGYCKVKGCPRPVIPNDFTIHGLWPD
PhS1     1  MFKLQLASVLCVFLFACSPISGSFDHWQLVLTWPAGYCKVKGCPRPVIPNDFTIHGLWPD  60

Sax1   61  SISVIMNNCDPTKTFVTITEINQITELEKRWPELTTTAQFALTSQSFWRYQYEKHGTCCF  120
           SISVIMNNCDPTKTFVTITEINQITELEKRWPELTTTAQFALTSQSFWRYQYEKHGTCCF
PhS1    61  SISVIMNNCDPTKTFVTITEINQITELEKRWPELTTTAQFALTSQSFWRYQYEKHGTCCF  120

Sax1  121  PVYSQSAYFDFAIKLKDKTDLLSILRSQGVTPGSTYTGERINSSIASVTRVKPNLKCLYY  180
           PVYSQSAYFDFAIKLKDKTDLLSILRSQGVTPGSTYTGERINSSIASVTRVKPNLKCLYY
PhS1   121  PVYSQSAYFDFAIKLKDKTDLLSILRSQGVTPGSTYTGERINSSIASVTRVKPNLKCLYY  180

Sax1  181  RGKLELTEIGICFDRTTVAMMSCPRISTSCKFGTNARITFRQ  222
           RGKLELTEIGICFDRTTVAMMSCPRISTSCKFGTNARITFRQ
PhS1   181  RGKLELTEIGICFDRTTVAMMSCPRISTSCKFGTNARITFRQ  222
```

Figure 2. Alignment of the predicted amino acid sequence from S_{ax1}-*Petunia axillaris* and the known amino acid sequence for S_1-*Petunia hybrida*

Identification of the S-RNase allele in *Petunia inflata*

BLASTp and BLASTn queries of the assembled *Petunia inflata* genome gave a single strong hit, as expected for a homozygous line. Identification of open reading frames and comparison with known S-RNase alleles demonstrated that the identified sequence region encoded a S-RNase that was most closely related (91% amino acid sequence identity) to the S_3-RNase of *Petunia hybrida* (Clark et al., 1990; Genbank AAA60466). Alignment of the protein sequences of the S6-*Petunia inflata* S-RNase with that of S_3-RNase of *Petunia hybrida* (Figure 3) showed that the majority of amino acid sequence differences were in the HVa and HVb regions demonstrated to be sufficient to distinguish S-RNase alleles in some cases (Matton et al., 1997, 1999). Thus, the S-RNase in this line of *Petunia inflata* likely represents a previously uncharacterized S-locus haplotype.

```
                                        C1                                C2
PinfS6   1    MVRLQLLSALFILLFSLSPVSANFDYFQLVLTWPASFCYPKNKCQRRSNNFTIHGLWPEK  60
              M RLQL+SA FILLFSLSPVSANFDYFQLVLTWPASFCYPKNKCQRRSNNFTIHGLWPEK
PhybS3   1    MFRLQLISAFFILLFSLSPVSANFDYFQLVLTWPASFCYPKNKCQRRSNNFTIHGLWPEK  60

                          HVa                       HVb              C3
PinfS6   61   KRFRLEFCTGDEYARFLKEDSIINDLERHWIQMRFDEKYAKDKQPLWEHEYTKHGICCSN  120
              KRFRLEFCTGD+Y RFL+ED+IIN LERHWIQMRFDE YA  KQPLWEHEY +HGICC N
PhybS3   61   KRFRLEFCTGDKYKRFLEEDNIINVLERHWIQMRFDETYANTKQPLWEHEYNRHGICCKN  120

                        C4
PinfS6   121  LYKQREYFLLAMRLKDKLDLLTILRNHGITPGTKHTFGEIQKAIKTVTNNNDPDLKCVEN  180
              LY Q+ YFLLAMRLKDKLDLLT LR HGITPGTKHTFGEIQKAIKTVT+NNDPDLKCVEN
PhybS3   121  LYDQKAYFLLAMRLKDKLDLLTTLRTHGITPGTKHTFGEIQKAIKTVTSNNDPDLKCVEN  180

                      C5
PinfS6   181  IKGVMELNEIGICYTPAADRFDRCRHSNTCDETSSTKILFRG  222
              IKGVMELNEIGICYTPAADRFDRCRHSNTCDETSSTKILFRG
PhybS3   181  IKGVMELNEIGICYTPAADRFDRCRHSNTCDETSSTKILFRG  222
```

Figure 3. Alignment of the predicted amino acid sequence for the S-RNase of S6 *Petunia inflata* and the S_3-RNase of *Petunia hybrida*. Conserved domains C1 through C5 are underlined and labeled. The two hypervariable domains are labeled and shaded in yellow.

Identification and Characterization of SLF genes in *Petunia axillaris*

Multiple sequence alignment of previously identified SLF sequences was used to produce consensus sequences for SLF1 through SLF8 variants (Figure 4). Those consensus sequences along with previously identified SLF and SLFL DNA sequences were used as BLASTn queries to identify putative SLF homologs on numerous scaffolds. Queries with SLF2, SLF3, SLF5, SLF6, SLF8 and S_2-SLFx showed strong homology to only a single gene region on different scaffolds. Manual annotation and BLASTp queries of the NCBI protein database confirmed that each of these regions corresponded to previously identified SLF variants from *Petunia*. Conversely, SLF1, SLF4, SLF7, and SLF-S3B queries identified from two to five separate gene sequences on different scaffolds. Manual annotation and characterization of these hits demonstrated that all appeared to be true SLF genes. Criteria included the presence of a single open reading frame with no introns (SLF genes lack introns), the presence of a well-defined N-terminal F-box domain, the absence of any other protein interaction domains and significant amino acid sequence identity to previously identified SLF proteins.

SLF1 **AAGAATAAAGGATGGCGAATGGTATTTTAAAGAAATTGCC**

SLF2 **CTTTCAGATGTTTATTGGGATCCTCCTA**

SLF3 **GATTAATATATTATTTAGGATTCCCGTGAAATCTCT**

SLF4 **ACACCATTCTCCAAAGTGCAATGAAATTATATTGTAAAGAATACAA**

SLF5 **GCCGTTACAGTGCCAATATTATGAAGATGCCA**

SLF6 **GAGATGAATATATTCTGTTAAAGCGTTGCTTATACAAGAAAACAACCAAT**

SLF7 **AGCACAGAGACTTTTCGCAATATGAAAATGCCGGATGCGTGTCATTTCAAA**

SLF8 **TGAATACGCGAGGAATAAGCTTTTGCGTCAAACTCAAAGGATC**

Figure 4: Consensus sequences from aligned SLF gene variants used as BLASTn probes. Multiple sequence alignment of previously-identified SLF variants SLF1 through SLF8 were used to produce a consensus sequence region unique to each SLF. These consensus sequences were used as the initial probes for BLASTn queries of the assembled genomes.

Table 1. Summary of SLF and S-RNase BLAST queries in *Petunia axillaris*. Scaffold identities and coordinates and best BLASTp hits are shown for identified SLF and S-RNase genes for "N/S26" *Petunia axillaris*. Linked regions are color-coded

Query	Scaffold	Size	Coordinates	BLASTp Hit	BLASTp %ID	Class
SLF1-1	00326	940666	508809-509978	AGL76530, S5-SLF1	99	SLF1
SFL1-2	00715	718562	578400-579572	ADD21612, S1-SLF1	99	SLF1
SLF2	00070	1471679	699640-700806	BAJ24857, S19-SLF2	95	SLF2
SLF3	00326	940666	596118-597275	BAJ24858, S5-SLF3	98	SLF3
SLF4-1	00671	726586	152315-153526	BAJ24865, S7-SLF4	94	SLF4
SLF4-2*(SLF12)	00326	940666	531044-532225	BAJ24865, S7-SLF4	78	SLF12
SLF5	03085	28205	6171-7340	BAJ24871, S7-SLF5	95	SLF5
SLF6	00172	1583766	619079-620257	BAJ24881, S19-SLF6	96	SLF6
SLF7-1(SLF14)	00172	1583766	708846-710139	ABR18788, S2-DD8	77	SLF14
SLF7-2(SLF16)	01162	482111	154759-155919	ABR18785, S2-DD5	76	SLF16
SLF8	00326	940666	464377-465552	ABX82525, S2-SLF8	90	SLF8
SLF-S3B-1 (SLF10)	00336	1804933	636123-637262	ABR15914,A-134	98	SLF10
SLF-S3B-2 (SLF18)	00671	726586	549469-550611	ABR15914, A-134	74	SLF18
SLF-S3B-3**(SLF19_1)	75437	1274	113-1273	ABR18782, S1-DD2	70	SLF19_1
SLF-S3B-4** (SLF19_2)	01590	192725	30722-31882	ABR18782, S1-DD2	70	SLF19_2
SLF-S3B-5 (SLF20)	01394	65038	63647-64564	ABR18782, S1-DD2	69	SLF20
S_2-SLFx (SLF13)	00326	940666	686239-687402	AHF49538, S2-SLFx	99	SLF13
SLF11	00426	634794	185428-186405	BAQ18962	95	SLF11
SLF15	01310	413452	331303-332496	AIK66473	94	SLF15
SLF17	01162	482111	154759-155919	AIK66479	99	SLF15
S-RNase	00336	1804933	449226-450052	AAA60465, S_1-RNase	100	N/A

*SLF4-2 sequence is in FBA_1 superfamily but lacks a clear N-terminal F-box domain.
**Identical sequences

Several of the identified genes appeared to be new SLF variants, including SLF4_2, SLF7-1, SLF7-2, SLF-S3B-1, SLF-S3B-2, SLF-S3B-3/4, SLF-S3B-5 and S2-SLFx. These have tentatively been assigned new SLF variant numbers by comparison with assignments suggested by Williams et al (2014b) and are listed in Table 1. Together these data indicate that there appear to be a minimum of 17 SLF variants present in *Petunia axillaris.* Several of the SLF genes are linked on the same scaffold assembly, as shown in Table 1 and Figure 5. Scaffold 00326 encodes five SLF gene sequences including SLF1_1, SLF3, SLF8, SLF13 and SLF12. SLF4 and SLF18 are also linked on a separate scaffold, as are SLF6 and SLF14. Finally, the S_{ax1}-RNase of *Petunia axillaris* is linked to SLF10 at a distance of approximately 187 kb. Two hits on different scaffolds (00326 and 00715) showed extremely strong (99%) BLASTp identity to previously identified SLF1 variants from *Petunia hybrida*, S5-SLF1 and S1-SLF1. These two SLF proteins differ by 11 out of 389 amino acids. Interestingly, *Petunia axillaris* N is completely self-compatible whereas S_1-*Petunia hybrida* is completely self-incompatible. As one common cause of the breakdown of self-incompatibility is a partial duplication of the S-locus that results in heteroallelic pollen expressing different SLF alleles (Golz et al, 2000; Sijacic et al., 2004) it is possible that these sequences reflect a partial duplication of the S-locus in this line. In fact, a naturally occurring example of SLF gene duplication associated with self-incompatibility has been described previously in *Petunia axillaris* (Tsukamoto et al., 2005).

Identification and Characterization of SLF genes in *Petunia inflata*

BLAST queries of the *Petunia inflata* genome assembly, carried out in the same manner as for *Petunia axillaris*, gave an even greater number of putative SLF hits. At least 29 identified genes appear to encode true SLF proteins, using the same criteria as above. Three gene regions (scaffold 01586, SLF7-3; scaffold 01751, SLF12_2 and scaffold 01501, SLF15_2) are apparent pseudogenes, as none encode a functional open reading frame. In the case of SLF7_3 a single nucleotide insertion at position 360 results in a frame-shift. As was the case for *Petunia axillaris*, several of the SLF genes are linked on the same scaffold. We have identified five linkage clusters, including SLF1-1 with SLF8-1 and SLF12_2(ψ), SLF1-2 with SLF12_1, SLF5-2 with SLF14_1, SLF7-3(Ψ) with SLF16_2 and SLF16_1 with SLF13_1.

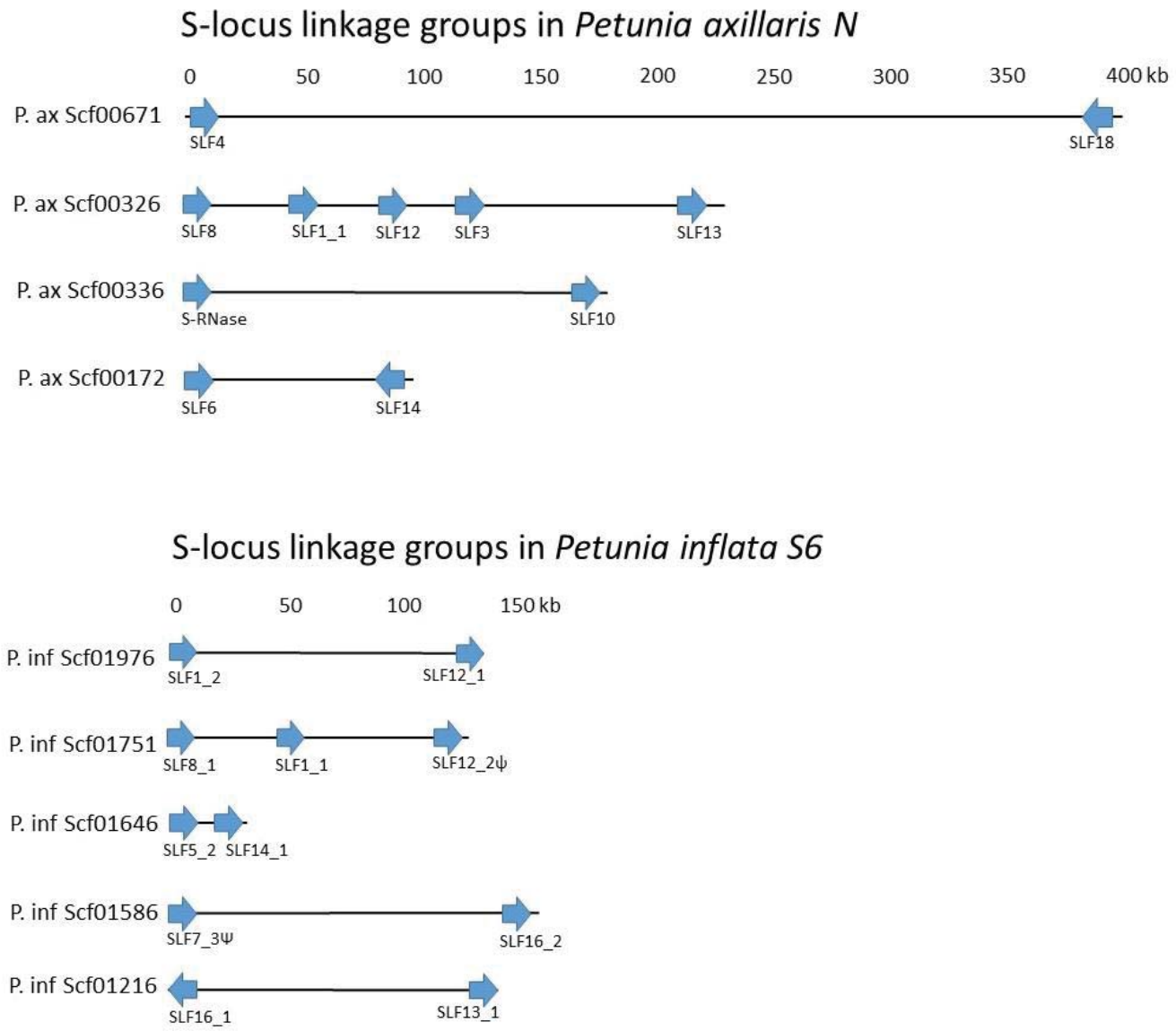


Figure 5. Linkage of SLF variants in *Petunia axillaris* N and *Petunia inflata* S6. The diagrams show a schematic of linked SLF variants found on the same scaffolds for *Petunia axillaris N* and for *Petunia inflata* S6. Individual scaffolds are numbered with the identity and orientation of different SLF variants found on the scaffolds shown by arrowheads. Depiction of the size of individual SLF variants is not to scale. Approximate distances between SLF variants is depicted by the scale at the top of the figure. Only that portion of individual scaffolds containing SLF variants is shown.

Table 2. Summary of SLF and S-RNase BLAST queries in *Petunia inflata*. Scaffold identities and coordinates and best BLASTp hits are shown for identified SLF and S-RNase genes for "S6" *Petunia inflata*. Linked regions are color-coded.

Gene	Scaffold	Size	Coordinates	Best BLASTp Hit	BLASTp %ID	Class
SLF1-1	01751	453812	144819-146092 (+)	AIK66522, S6a-SLF1	99	SLF1
SLF1-2	01976	214490	4694-5969 (+)	ADD21612, S1-SLF1	95	SLF1
SLF2-1	01991	197077	278634-279788 (+)	AIK66498, S3-SLF2	90	SLF2
SLF2-2	01973	296567	153541-154695 (-)	BAQ19037, Sm-SLF2	83	SLF2?
SLF3	01923	159365	5575-6582 (+)	BAQ18957, S10-SLF3	96	SLF3
SLF4-1	00922	184426	64033-65244 (-)	BAJ24865, S7-SLF4	96	SLF4
SLF4-2	05340	224001	129825-131051 (+)	BAJ24865, S7-SLF4	93	SLF4
SLF5-1	01516	657630	644401-645570 (+)	BAQ19039, Sm-SLF5	96	SLF5
SLF5-2	01646	115327	3200-4369 (+)	BAQ19039, Sm-SLF5	96	SLF5
SLF6-1	01506	144469	61311-62492 (+)	BAJ24881, S19-SLF6	95	SLF6
SLF6-2	00996	461697	107014-108195 (-)	BAJ24881, S19-SLF6	96	SLF6
SLF7-1	00080	221173	3893-5071 (-)	ABX82526, S2-SLFLa	99	SLF7
SLF7-2	00526	752603	259704-260882 (-)	BAQ18970, S11-SLF7	98	SLF7
SLF7-3Ψ	01586	859704	334991-336167 (+)	BAQ19025, S0m-SLF7	79	pseudogene
SLF8-1	01751	453812	50755-51939 (+)	BAQ18935, S7-SLF8	97	SLF8
SLF8-2	03408	108618	69122-70303 (+)	BAQ18935, S7-SLF8	88	SLF8?
SLF9-1	01995	82553	62151-63281 (-)	BAQ18945, S9-SLF9	97	SLF9
SLF9-2	01799	514228	140243-141373 (-)	BAQ18945. S9-SLF9	98	SLF9
SLF10	01608	234268	115426-116565 (+)	AAR15915, S2-A134	99	SLF10
SLF11-1	04466	242965	204665-205837 (+)	AIK66454, S6a-SLF11	99	SLF11
SLF11-2	00431	165049	14066-15178 (+)	BAQ18974, S11-SLF11	95	SLF11
SLF12-1	01976	214490	142523-143704 (+)	AIK66458, S6a-SLF12	100	SLF12
SLF12-2Ψ	01751	453812	177587-179429 (+)	AIK66458, S5-SLF12	71	pseudogene
SLF13-1	01216	514760	360226-361392 (+)	AIK66464, S6a-SLF13	99	SLF13
SLF13-2	02268	142811	121932-123095 (-)	BAQ19014, S22-SLF13	96	SLF13
SLF14-1	01646	115327	30658-31851 (+)	AIK66469, S6a-SLF14	99	SLF14
SLF14-2	01246	26952	6131-7324 (-)	AIK66469, S6a-SLF14	99	SLF14
SLF15-1	16932	4714	563-1723 (+)	AIK66475, S12-SLF15	96	SLF15
SLF15-2Ψ	01501	1043979	485372-486154 (-)	AIK66472, S3-SLF15	84	pseudogene

SLF16-1	01216	514760	214995-216155 (-)	AIK66479, S6a-SLF16	100	SLF16
SLF16-2	01586	859704	501814-502983 (+)	BAQ18940, S7-SLF16	77	SLF16
SLF17	00041	2522227	1710310-1711476 (+)	BAQ18941, S7-SLF17	99	SLF17
S6-RNase	01010	272190	153855-154627 (+)	AAA60466, S3-RNase	91	S6-RNase

Phylogenetic comparison of SLF gene sequences

To analyze similarities among different SLF variants, we used MEGA6 (Tamura et al., 2013) to carry out a phylogenetic comparison of all of the SLF variants identified in *Petunia inflata* S6 and *Petunia axillaris* N. This analysis (Figure 6) identified 6 clades supported by bootstrap values greater than 72% of 1000 bootstrap replicates. The identified clades contained the following SLF varants. Clade 1: SLF 7, SLF14, SLF15, SLF16, SLF17. Clade 2: SLF3, SLF11, SLF13. Clade 3: SLF9, SLF10, SLF18, SLF19, SLF20. Clade 4: SLF1, SLF2. Clade 5: SLF8. Clade 6: SLF4, SLF5, SLF6.

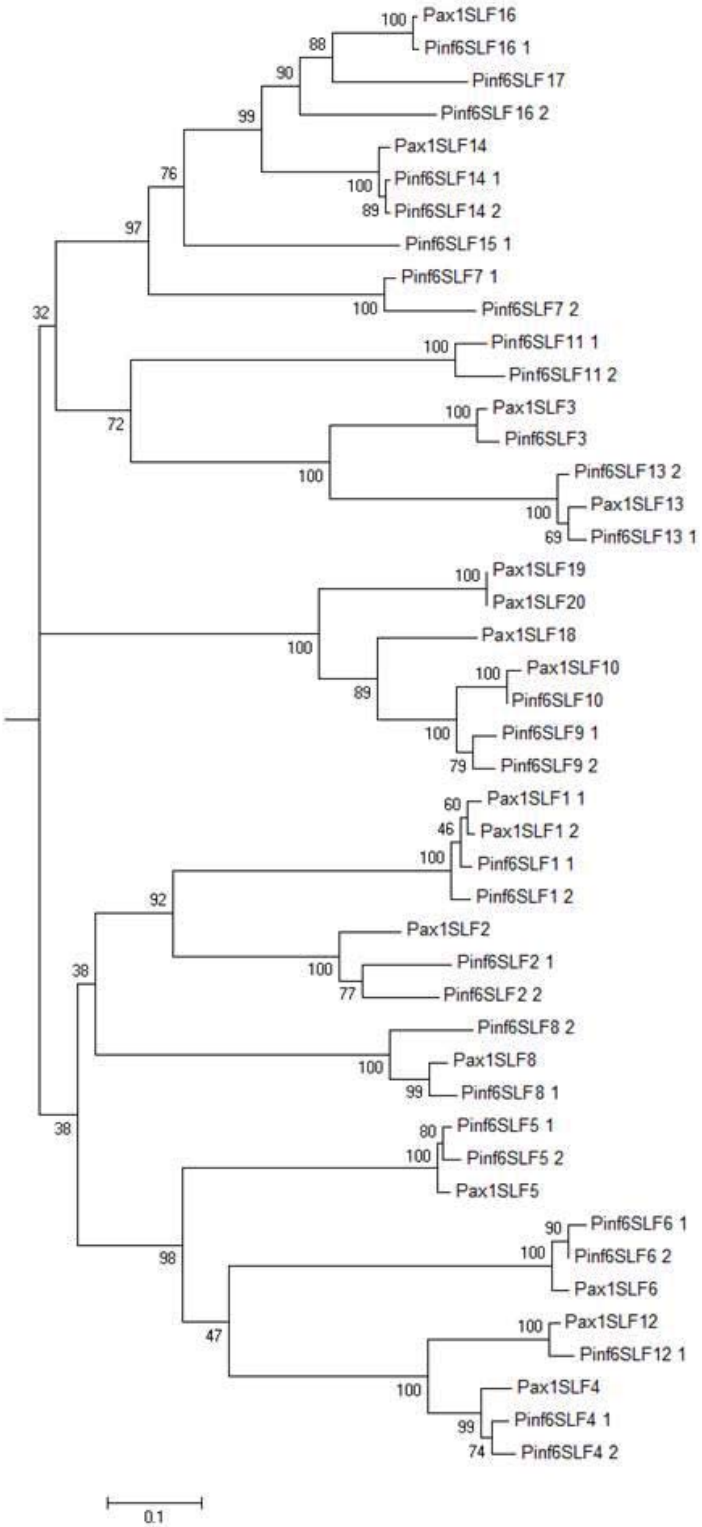


Figure 6. Maximum likelihood phylogenetic relationships between SLF variants from *Petunia axillaris N* and *Petunia inflata* S6 were produced using MEGA6. The scale bar at the bottom shows the number of substitutions per site. Bootstrap values are shown for individual nodes.

DISCUSSION

Having completely assembled and annotated genomic DNA sequences for *Petunia inflata* and *Petunia axillaris* will provide an abundance of information that should enhance investigations on the organization and expression of the S-locus and on the functional mechanisms involved in self versus non-self recognition in gametophytic self-incompatibility. In a preliminary analysis of the sequence assemblies for the two genomes, we carried out BLAST queries to identify S-RNase and SLF genes presumed to be encoded and linked to the S-locus. Numerous SLF genes (19 in *Petunia axillaris* and 29 in *Petunia inflata*) were identified. These included sequences nearly identical to previously identified SLF genes as well as several new SLF genes which likely encode additional SLF variants beyond those currently identified. One of the challenges of GSI research will be to determine if all of the putative SLF genes are expressed and if some or all of them function in self-incompatibility recognition (via interactions with S-RNase proteins). As predicted from current models for the structure of the S-locus, many of the SLF genes identified were linked to each other; at least one SLF gene in *Petunia axillaris* was found on the same DNA scaffold as the S-RNase gene in this species. Although the quality of the sequence assembly for these two genomes did not allow assembly of a complete "virtual" S-locus, it is likely that a complete S-locus can be assembled using these data combined with screens of BAC libraries. Recently, an analysis of SLF genes in the completed tomato and potato genomes has indicated that the approximate size of the S-locus is 14.5 Mb and 17.9 Mb, respectively (Kubo et al., 2015). Adding together all the SLF-bearing scaffold sizes in *Petunia axillaris* gives a total minimum size of approximately 8.0 Mb. A similar analysis of SLF-bearing scaffold sizes in *Petunia inflata* gives a total minimum size of approximately 11.3 Mb; this is consistent with a significant duplication in *Petunia inflata*. Until the scaffolds are connected, it is not possible to accurately compare the scale of these S-loci, but in *Petunia* it would appear to be of comparable size to that observed in *Solanum*.

Analysis of S-RNase-encoding genes in *Petunia inflata* S6 and *Petunia axillaris* N showed single copies of the S-RNase, as expected from the fact that both of these lines were homozygous. The S-RNase in *Petunia inflata* appears to be an as yet uncharacterized S-RNase. The most closely related characterized S-RNase to the one identified is the S_3-RNase of *Petunia hybrida* (Q40875, Okuley and Sims, 1994). As these two S-RNases differ by 20 amino acids, with 10 of the differences occurring in the HVa and HVb hypervariable regions known to play a major role in determining allele specificity (Matton et al, 1997), it is extremely likely that this S-RNase encodes a different recognition specificity than any previously published and characterized S-RNase. By contrast, we found that the S-RNase in N *Petunia axillaris* was nearly identical (99% nucleotide sequence identity across the entire gene and 5' and 3' flanking regions) to that of the S_1-RNase of *Petunia hybrida*. That result shows that self-incompatibility in the cultivated garden petunia can be inherited from both of the progenitor species to *Petunia hybrida*, *Petunia inflata* and *Petunia axillaris*. Because S_1-*Petunia hybrida* is completely self-incompatible and never sets seed on self-pollination, whereas the sequenced *Petunia axillaris* N line is completely self-compatible, setting large seed capsules on selfing (Sims, unpublished), it will be instructive to closely compare the structure and expression of the S-loci in these two lines to determine the molecular basis for GSI breakdown.

We identified at least 20 different SLF variants in the two species analyzed, but not all SLF variants were present in each species. For example, *Petunia axillaris* N lacked SLF7, SLF9 and SLF11 whereas *Petunia inflata* S6 lacked SLF18, SLF19 and SLF20. Whereas the SLF variants in *Petunia* axillaris were mostly single genes, nearly all of the SLF variants in *Petunia inflata* were found as two closely-related copies that mapped to different scaffolds. This suggests that the S-locus in *Petunia inflata* S6 may have undergone a recent duplication event compared with

Petunia axillaris. Supporting that suggestion is the finding that the estimated size of S-locus-containing scaffolds in *Petunia inflata* S6 exceeds 11 Mb whereas that for *Petunia axillaris* N spans just over 8 Mb of DNA.

METHODS

Plant Material

Seeds of *Petunia inflata* line S6 and *Petunia axillaris* line N were obtained from the Free University of Amsterdam. Plants used for DNA extraction were grown axenically in tissue culture containers (Ball Horticulture). Mature plants (leaves and stems) were harvested, flash-frozen in liquid nitrogen and stored at -80C until used for DNA extractions.

DNA Extractions

Plant material (approximately 15 g) was extracted using a modification of methods designed to isolate high molecular weight DNA from nuclei (Fisher and Goldberg, 1982; Carrier et al., 2011). Briefly, frozen plant material was homogenized in a blender with liquid nitrogen until a fine powder was obtained. Powdered material was thawed in 1X SEB plus mercaptoethanol (10 mM Tris pH 8.0, 100 mM KCl, 10 mM Na_2EDTA, 0.5 M sucrose, 4 mM spermidine, 1 mM spermine, 0.13% carbamic acid, 0.25% PVP-40, 0.2% β-mercaptoethanol), then filtered through Nitex mesh. Triton X-100 was added to 0.5% and nuclei isolated and washed by repeated low speed centrifugation and washing with SEB. Nuclei were lysed by adding an equal volume of NLB (2% Sodium N-lauryl sarcosine, 40 mM Na_2EDTA, 0.1 M Tris-HCl pH 8.0 and 1mg/ml proteinase K) followed by incubation at 55 C for 1 hour. Cesium chloride was added to 50% w/w along with ethidium bromide to a final concentration of 0.4%. DNA gradients were centrifuged in a 70.1 Ti rotor at 40,000 rpm for 36 hours followed by re-banding in a VTi 65.2 rotor at 60,000 rpm for 6 hours. Ethidium bromide was removed by extraction with SSC-saturated isopropanol and the remaining solution dialyzed against TNE (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.1 mM EDTA) for 24 hours. DNA was precipitated, washed with 70% ethanol, dried and resuspended in EB (10 mM Tris pH 8.0) to a final concentration of 150 µg/ml.

DNA Sequencing

DNA libraries (two 1000 bp paired-end, two 8 kb and two 15 kb mate-pair) were constructed at the University of Illinois Roy J. Carver Biotechnology Center. Paired-end libraries were sequenced on two lanes of HiSeq 2000 and the mate-pair libraries sequenced on a single HiSeq 2500 lane.

DNA Assembly and Annotation

DNA sequences for both *Petunia inflata* and *Petunia axillaris* were combined with previous raw reads produced by BGI-Shenzen for the Petunia Genome Project and assembled using SOAPdenovo2 (Luo et al., 2010). For *Petunia axillaris*, the Illumina data were combined with PacBio data to produce the final genome assembly. MAKER (Cantarel et al 2008) was used to annotate predicted transcript and protein sequences.

Identification and manual annotation of S-RNase and SLF gene sequences

Known S-RNase sequences (S_1-RNAse, S_3-RNase) from *Petunia hybrida* (Clark et al., 1990) were used in BLASTn and BLASTp (Altschul et al., 1990) searches of the assembled scaffolds for *Petunia axillaris* and *Petunia inflata*. Individual scaffold regions were used in pairwise alignment against known S-RNase mRNA and gene sequences and analyzed for open reading frames. Multiple approaches were used to identify potential SLF coding genes. Initially, all known SLF genes of a particular variant class (SLF1 to SLF8, Kubo et al., 2010) were aligned to identify conserved regions unique to a particular SLF variant class (Figure 4). Those conserved sequences were then used

as BLASTn probes against the assembled genomes to identify scaffold regions homologous to a particular SLF variant. In addition, other known SLFL genes (Wang et al 2003), DDX genes from *Nicotiana alata* (Wheeler and Newbigin 2000), SLF-S3B (Qiao et al 2004) and S2-SLFx (Li et al 2014) were used as BLASTn probes. BLASTn queries used an Expect value of 10. Scaffold regions identified by BLASTn queries were used as input into the Translate program of ExPASy to identify open reading frames. Translated protein sequences were then used in BLASTp queries of GenBank to identify the most closely-related proteins. Conserved domains were identified using a combination of BLASTp and InterPro Scan 5 (Jones et al. 2014).

Phylogenetic comparison of SLF variants

Manually annotated SLF variants from *Petunia axillaris* N and *Petunia inflata* S6 were used as input to the MEGA6 phylogenetics analysis program (Tamura et al., 2013). Sequences were aligned using ClustalW and the alignment used as input to the MEGA6 maximum likelihood phylogenetics analysis program, using the bootstrap method with a value of 1000 iterations.

REFERENCES

Altschul, S.F., Gish, W., Miller., W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403-410

Ando, T., Tsukamoto, T., Akiba, N., Kokubin, H., Watanabe, H., Ueda, Y. and Marchesi, E. (1998). Differentiation in the Degree of Self-incompatibility in *Petunia axillaris* (Solanaceae) Occurring in Uruguay. Acta. Phytotax. Geobot. 49, 37-47.

Ascher, P.D. (1984). Self-incompatibility. In K.C. Sink, Ed., *Petunia: Monographs on theoretical and applied genetics Vol. 9.* Berlin, Springer-Verlag 92-109

Brewbaker, J.L., Natarajan, A.T. (1960). Centric fragments and pollen-part mutation of incompatibility alleles in Petunia. Genetics 45, 699-704

C**antarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A.S., and Yandell, M.** (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research 18, 188-196.

C**arrier, G., Santoni, S., Rodier-Goud, M., Canaguier, A., de Kochko, A., Dubreuil-Tranchant, C., This, P., Boursiquot, J.-M., and le Cunff, L.** (2011). An efficient and rapid protocol for plant nuclear DNA preparation suitable for next generation sequencing methods. American Journal of Botany e13-315.

Clark, K.R., Okuley, J., Collins, P.D. and Sims, T.L. (1990). Sequence variability and developmental expression of S-alleles in self-incompatible and pseudo-self-compatible *Petunia.* The Plant Cell 2, 815-826.

Darwin, C. The effects of cross and self fertilisation in the vegetable kingdom. 2nd ed. New York (NY), USA: D. Appleton and Company; 1898

De Nettancourt, D. (1977). *Incompatibility in angiosperms*. New York USA: Springer-Verlag.

Entani, T., Iwano, M., Shiba, H., Takayama, S., Fukui, K. and Isogai, A. (1999). Centromeric localization of an S-

RNase gene in *Petunia hybrida* Vilm. Theoretical and Applied Genetics 99: 391-397

Fischer, R.L. and Goldberg, R.B. (1982). Structure and flanking regions of soybean seed protein genes. Cell 29, 651-660.

Golz, J.F., Su, V., Clarke, A.E., Newbigin, E. (1999). A molecular description of mutations affecting the pollen component of the *Nicotiana alata* S locus. Genetics 152, 1123-1135.

Hua, Z. and Kao, T.-H. (2006). Identification and characterization of components of a putative *Petunia* S-locus F-box-containing E3 ligase complex involved in S-RNase-based self-incompatibility. The Plant Cell 18, 2531-2553.

Igic, B. and Kohn, J.R. (2001). Evolutionary relationships among self-incompatibility S-RNases. Proceedings of the National Academy of Sciences USA 98, 13167-13171

Ioerger, T.R., Gohlke, J.R., Xu, B., Kao, T.-H. (1991). Primary structural features of the self-incompatibility protein in Solanaceae. Sexual Plant Reproduction 4, 81-87.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R. and Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236-1240.

Kubo, K., Entani, T., Takara, A., Wang, N., Fields, A.M., Hua, Z., Toyoda, M., Kawashima, S., Ando, T., Isogai, A., Kao, T.-H. and Takayama, S. (2010). Collaborative Non-Self Recognition System in S-RNase-Based Self-Incompatibility. Science 330, 796-799.

Kubo, K., Paape, T., Hatakeyama, M., Entani, T., Takara, A., Kajihara, K., Tsukahara, M., Shimizu-Inatsugi, R., Shimizu, K.K. and Takayama, S. (2015). Gene duplication and genetic exchange drive the evolution of S-RNase-based self-incompatibility in *Petunia*. Nature Plants 1. doi:10.1038/nplants.2014.5

Lee, H.S., Huang, S., Kao, T.-H. (1994). S proteins control rejection of incompatible pollen in *Petunia inflata* Nature 367, 560-563

Li, R., Zhu, H,, Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Chan, G., Kristiansen, K., Li, S., Yang, H., Wang, J. and Wang, J. (2010). De novo assembly of human genomes with massively parallel short read sequencing. Genome Research 20, 265-272.

Li, S., Sun, P., Stephen-Williams, J. and Kao, T.-H. (2014). Identification of the self-incompatibility locus F-box protein-containing complex in *Petunia inflata.* Plant Reproduction 27, 31-45.

Linskens, H.F. (1975). Incompatibility in *Petunia*. Proceedings of Royal Society London Biological Sciences 188, 299-311.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, X., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, S., Yu, C., Wang, B., Lu, Y., Han, C., Shi, Y., Liu, S., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S,-M,, Peng, S., Xiaoqian, Z., Liu, G., Lia, X., Li, Y., Yang, H., Wang, J., Lam, T.-W. and Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short read *de novo* assmbler. Giga Science 1, 18-23.

Mather, K. (1943). Specific differences in *Petunia* I. Journal of Genetics 45, 215-235

Matton, D.P., Maes, O., Laublin, G., Xike, Q., Bertrand, C., Morse, D. and Cappadocia, M. (1997). Hypervariable domains of self-incompatibility RNases mediate allele-specific pollen recognition. The Plant Cell 9, 1757-1766.

Matton, D.P., Luu, D.T., Xike, Q,, Laublin, G., O'Brien, M., Maes, O., Morse, D. and Cappadocia, M. (1999). Production of an S-RNase with dual specificity suggests a novel hypothesis for the generation of new S alleles. The Plant Cell 11, 2087-2097.

McClure, B., Cruz-García, F. and Romero, C. (2011). Compatibility and incompatibility in S-RNase-based systems. Annals of Botany 108, 647-658

Meng, X., Sun, P. and Kao, T.-H. (2011). S-RNase-based self-incompatibility in *Petunia inflata*. Annals of Botany 108, 637-646.

Qiao, H., Wang, F., Zhao, L., Zhou, J.L., Lai, Z., Zhang, Y.S., Robbins, T.P. and Xue, Y. (2004). The F-box protein AhSLF-S_2 controls the pollen function of S-RNase-based self-incompatibility. Plant Cell 16, 2307-2322.

Shivanna, K.R. and Rangaswamy, N.S. (1969). Overcoming self-incompatibility in *Petunia axillaris* I. Delayed pollination, pollination with stored pollen and bud pollination. Phytomorphology 19, 372-380.

Sijacic P., Wang, X., Skirpan, A.L., Wang, Y., Dowd P.E., McCubbin, A.G., Huang, S., Kao, T.-H. (2004). Identification of the pollen determinant of S-RNase-mediated self-incompatibility. Nature 429, 302-305.

Sims, T.L. and Ordanic, M. (2001). Identification of a S-ribonuclease binding protein in *Petunia hybrida*. Plant Molecular Biology 47, 771-783.

Sims, T.L., Patel, A. and Shrestha, P. (2010). Protein interactions and subcellular localization in S-RNase-based self-incompatibility. Biochemical Society Transactions 38, 622-626.

Sims, T.L. and Robbins, T.P. (2009). Gametophytic Self-Incompatibility in *Petunia*. in Petunia: Evolutionary, Developmental and Physiological Genetics. Gerats T and Strommer J (eds.) ISBN 978-0-387-84795-5, Springer Life Sciences pp.85-106

Sims, T.L. (2012). Protein Interactions in S-RNase-based Gametophytic Self-Incompatibility. In Jianfeng Cai ed. Protein Interactions. In Tech Web www.intechopen.com

Tamara, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). MEGA6 , Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30 , 2725-2729.

Tanksley, S.D., Loaiza-Figueroa, F. (1985). Gametophytic self-incompatibility is controlled by a single major locus on chromosome 1 in *Lycopersicon peruvianum.* Proceedings of the National Academy of Sciences USA 82 , 5093-5096.

ten Hoopen, R., Harbord, R.M. Maes, T., Nanninga, N. and Robbins, T.P. (1998). The self-incompatibility (S) locus in *Petunia hybrida* is located on chromosome III in a region syntenic for the Solanaceae. Plant Journal 16, 729-734

Tsukamoto, T., Ando, T., Watanabe, H., Marchesi, E. and Kao, T.-H. (2005) Duplication of the S-locus F-box gene is associated with breakdown of pollen function in an S-haplotype identified in a natural population of self-incompatible *Petunia axillaris.* Plant Molecular Biology 57,141–153

Wang, Y., Wang, X., McCubbin, A.G. and Kao, T.-H. (2003) Genetic mapping and molecular characterization of the self-incompatibility (S) locus in *Petunia inflata* Plant Molecular Biology 53, 565-580.

Wang, Y., Tsukamoto, T., Yi, K,-W., Wang, X., Huang, A., McCubbin, A.G. and Kao, T.-H. (2004) Chromosome walking in the *Petunia inflata* self-incompatibility (S-) locus and gene identification in an 881-kb contig containing S2-RNase. Plant Molecular Biology 54, 727-742.

Wheeler, D. and Newbigin, E. (2007) Expression of 10 S-Class SLF-*like* Genes in *Nicotiana alata* Pollen and Its Implications for Understanding the Polen Factor of the S Locus. Genetics 177, 2171-2180

Williams, J.S., Natale, C.A., Wang, N., Li, S., Brubaker T.R., Sun, P. and Kao, T.-H. (2014a) Four previously identified *Petunia inflata* S-locus F-box genes are involved in pollen specificity in self-incompatibility. Molecular Plant 7, 567-569.

Williams, J.S., Der, J.P., dePamphilis, C.W. and Kao, T.-H. (2014b). Transcriptome analysis reveals the same 17 S-Locus F-box genes in two haplotypes of the self-incompatibility locus of *Petunia inflata*. Plant Cell 26, 2873-2888.

Zhao, L., Huang, J., Zhao, Z., Li, Q., Sims, T.L. and Xue, Y. (2010) The Skp1-like Protein SSK1 is Required for Cross-Pollen Compatibility in S-RNase-Based Self-Incompatibility. Plant Journal 65, 52-63.

Supplementary Note 12

Genetic structure of the circadian clock in Petunia

Marcos Egea-Cortines[1] and Julia Weiss[1]

[1]Genética Molecular, Instituto de Biotecnología Vegetal, Universidad Politécnica de Cartagena, 30202 Cartagena, Spain

Address correspondence: Marcos.egea@upct.es

Abstract

The circadian clock is a complex network comprising three layers of transcriptional, post-translational and metabolic control. Together they play a role in integrating environmental cues and metabolism to create coherent outputs for growth, adaptation and fine-tuning of secondary metabolism. We surveyed the Petunia genomes and compared the genetic structure of the clock with Arabidopsis and with other *Solanaceae*. We found a complex situation whereby the general configuration of the circadian clock is conserved in *Solanaceae* compared to Arabidopsis. However, we did not find a complete set of duplicated clock genes that would have been expected based on the paleohexaploidy shared by *Solanaceae* and a gene dosage hypothesis. Instead we found gene duplications for some of the genes involved in posttranslational control such as *GI*, *ELF3*, *ELF4* or proteins with dual function in DNA-binding and protein complex formation such as *TOC1*, *PRR7* and *PRR5*. The transcription factors *LHY* and *LUX*/*PCL1* were retained mostly as a single copy gene in Petunia and the rest of the *Solanaceae*. When compared to genes from other families, *LHY*, *GI* and *LUX* show subclades grouping the *Solanaceae* genes indicating a family specific evolution of the complete pathway that occurs both by changes in gene copy and coding region. The set of genes taking part in the circadian clock seems to be species-specific, and might explain the diversity of ecological niches colonized by the *Solanaceae*.

Introduction

Day to day coordination of basic biological functions within the environment is achieved by the entrainment of an endogenous circuit of control known as the circadian clock. It comprises three different parts, one is formed by a group of genes involved in entrainment and coordination with the environment, a second one is the so-called core clock, and finally those genes that act in the output of the clock. Physical signals such as light and temperature act as inputs of the clock. Endogenous chemical signals such as sugar or nitrogen levels can also modify its function (Gutierrez et al., 2008; Haydon et al., 2013; Salomé et al., 2010; Jarillo et al., 2006). The core of the clock is a set of genes comprising three loops known as morning, midday or evening loop (Pokhilko et al., 2012). Some core clock has components that also play direct roles in activation and repression of target genes as the outputs of the clock (Egea-Cortines et al., 2013; Yakir et al., 2007; Mas et al., 2013). The major processes known to be controlled as outputs of the clock are flowering time, response to cold, growth and basic metabolic processes like photosynthesis, starch metabolism and secondary metabolism such as scent emission (Farré and Weise, 2012; de Montaigu et al., 2010; Graf and Smith, 2011). New evidence links clock outputs to a variety of important traits like adaptation to environment, pathogen resistance or fitness (Egea-Cortines et al., 2013).

Work in Arabidopsis, crops and non-model species has shown that adaption to environmental signals might have occurred by modifications within the clock, and not in downstream processes (Mallona et al., 2011; Weller et al., 2012; Brachi et al., 2010; Zakhrabekova et al., 2012). This indicates that the fine-tuning and divergences found in the clock in different species might show their evolutionary history and will help us understand and improve plant adaptation. Indeed recent evidence shows that circadian clock tuning maybe important during plant domestication in tomato (Müller et al., 2016).

The clock structure and function in higher plants is best understood in Arabidopsis. Five members of the small gene family of *PSEUDORESPONSE REGULATOR* form part of the clock (Mas et al., 2013). The genes *PSEUDORESPONSE REGULATOR 9*, *7* and *5* form the morning loop, *PRR3* is apparently a vascular-specific component of the clock (Para et al., 2007). The core loop is formed in Arabidopsis by two paralogous *MYB* genes *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) and *LONG ELONGATED HYPOCOTYL* (*LHY*) (Takata et al., 2010) together with PRR1 known as *TIMING OF CAB1 (TOC1)* (Alabadi et al., 2001). Finally the evening loop comprises three genes *EARLY FLOWERING 3* and *4* (*ELF3* and 4) and *LUX ARRYTHMO* (Nusinow et al., 2011).

Understanding the evolution of the plant circadian clock is becoming important as the major traits of relevance in agriculture are directly controlled or affected by the clock (Panda et al., 2002). A current hypothesis is that circadian clock genes may be under special selective pressure as they form a coordinated genetic network comprising mutual activation and repression via direct interactions. These are sometimes the result of several independent gene products that form protein complexes (McClung and Gutiérrez, 2010; Lou et al., 2012). As the genome of the *Solanaceae* lineage has undergone at least a paleohexaploidy event (Tomato and Consortium, 2012; Kim et al., 2014; Sierro et al., 2014; Xu et al., 2011), a starting

hypothesis would be that the complete set of genes may have undergone a duplication and we may find at least one additional paralog specific for the Solanaceae.

Petunia has been used in establishing the circadian clock as a master regulator of development in plants. Early work showed that the *CHLOROPHYL A/B BINDING PROTEIN* gene (*CAB*) has a rhythmic expression (Stayton et al., 1989). This confirmed previous works that identified rhythmic expression of *CAB* in tomato and tobacco plants (Paulsen and Bogorad, 1988; Piechulla, 1988). Amongst the outputs of the clock, the circadian emission of floral scent has been studied with detail in Petunia and Antirrhinum (Kolosova et al., 2001; Verdonk et al., 2003). Furthermore, the *MYB* transcription factor *ODORANT1*, involved in quantitative control of scent emission in Petunia, displays a rhythmic expression indicating a direct regulation of the clock on scent output in Petunia (Verdonk et al., 2005). Recent studies using artificial vision image analysis show that *Petunia x hybrida* flowers open in a short time frame right after dawn indicating that important aspects of morphogenesis, beyond hypocotyl elongation, might have a light and circadian clock control (Navarro et al., 2012). A recent work overexpressing the *Petunia x hybrida LHY* gene shows the control of scent emission timing by the clock (Fenske et al., 2015), confirming the role of the clock in controlling secondary metabolism.

We have performed a comprehensive search for the genes involved in the clock of *Petunia axillaris* and *P. inflata* using Arabidopsis as a template. As the circadian clock genes are not explored with detail in the *Solanaceae* and there is a genome duplication specific for the family, we have analysed and compared the findings in Petunia with those of other sequenced *Solanaceae* genomes. Our results indicate a basic structure similar to that of Arabidopsis, but the number of specific paralogs found for a given gene seems to be species specific and does not follow a simple rule, based on genome duplications, and maybe related to specific biological and molecular functions.

Results

We surveyed the Petunia genomes using proteins identified by BLAST in tomato corresponding to circadian clock components. We found seven genes in *P. axillaris* and *P. inflata* with high similarity to the *PRR9, 7, 5, 3* and *TOC1* genes from Arabidopsis. This increased number was due to gene duplications in *PRR7* and *PRR5* resulting in two paralogs present in Petunia. As we found gene duplications for *PRR7* and *5* in different *Solanaceae*, we assumed that the complete set of clock related *PRR* genes has been be duplicated (Table 1-list of genes). This would be in agreement with the genome duplications found in the *Solanaceae* (Tomato and Consortium, 2012; Xu et al., 2011). However this hypothesis was not correct. The gene *PRR3* appeared as single copy genes in the species surveyed (Figure 1).

Table 1: A list of genes of the circadian clock from Petunia and Solanaceae

Arabidopsis	PinfS6	PaxilN	Other Solanaceae
PRR9	P.inflataPRR9	P.axillarisPRR9	Duplicated in N.benthamiana
PRR7	P.inflataPRR7a P.inflataPRR7b	P.inflataPRR7a P.axillaris PRR7b	Single or double copy
PRR5	P.inflataPRR5a P.inflataPRR5b	P.axillarisPRR5a P.axillaris PRR5b	Single or double copy
PRR3	P.inflataPRR3	P.axillarisPRR3	Single copy
TOC1	P.inflataTOC1	P.axillarisTOC1	Duplicated in N.benthamiana
LHY	P.inflataLHY	P.axillarisLHY	Single copy in Solanaceae
CCA1	Absent	Absent	Absent
ELF3	P.inflataELF3 P.inflataELF3a P.inflataELF3b P.inflataELF3c	P.axillarisELF3 P.axillarisELF3a P.axillarisELF3b	Three to four copies in Solanaceae
ELF4	P.inflataELF4a; P.inflataELF4b	P.axillarisELF4a; P.axillarisELF4b	Duplicated in Solanaceae
LUX	P.inflataLUX	P.axillarisLUX	Duplicated in S.lycopersicum and S.tuberosum
GI	P.inflataGI1; P.inflataGI2; P.inflataGI3	P.axillarisGI1; P.axillarisGI2	Two to four copies in Solanaceae

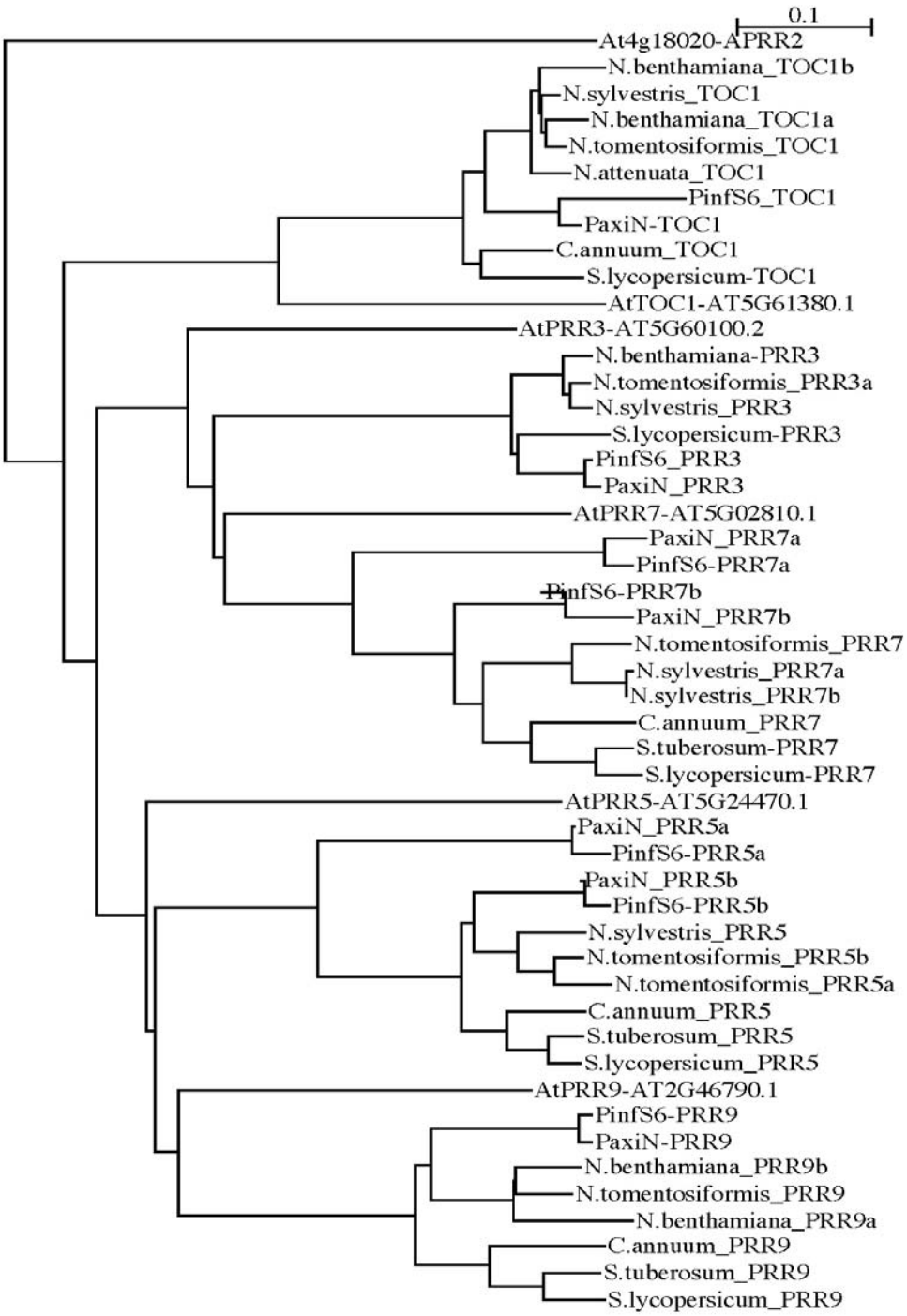


Figure 1. Phylogenetic analysis of *PSEUDORESPONSE REGULATOR* genes from Petunia and other Solanaceae. We used Arabidopsis genes to establish the families corresponding to PRR),7,5,3 and TOC1. Plant species correspond to: At, *Arabidopsis thaliana*; *Nicotiana sylvestris; Nicotiana tomentosiformis; Nicotiana benthamiana; Nicotiana attenuata, Solanum lycopersicum* and *Solanum tuberosum*.

The survey of the Petunia genomes for *LHY* and *CCA1* orthologs showed several genes with high degree of homology. Both Arabidopsis *CCA1* and *LHY* gave identical BLAST hits with one scaffold showing higher scores that the rest. This indicated a possible structure of midday loop that would comprise a single *LHY/CCA1* gene as compared to Arabidopsis that has *CCA1* and *LHY* (Figure 2). A comparative analysis with the rest of the *Solanaceae* sequenced genomes identified several *LHY/CCA1* homologous genes with different BLAST e values. However, the inclusion of the identified clones in the phylogenetic framework supported the previous preliminary finding as we identified a single *LHY* gene present in *Solanum tuberosum*, *Nicotiana benthamiana*, *Capsicum annuum*, *P. axillaris* and *P. inflata* (Figure 2). The other genes found clustered with the *REVEILLE* gene family of circadian clock related *MYB* genes (Rawat et al., 2009; Farinas and Mas, 2011; Rawat et al., 2011). Our current data indicates that there is a single *LHY* ortholog in the Solanaceae.

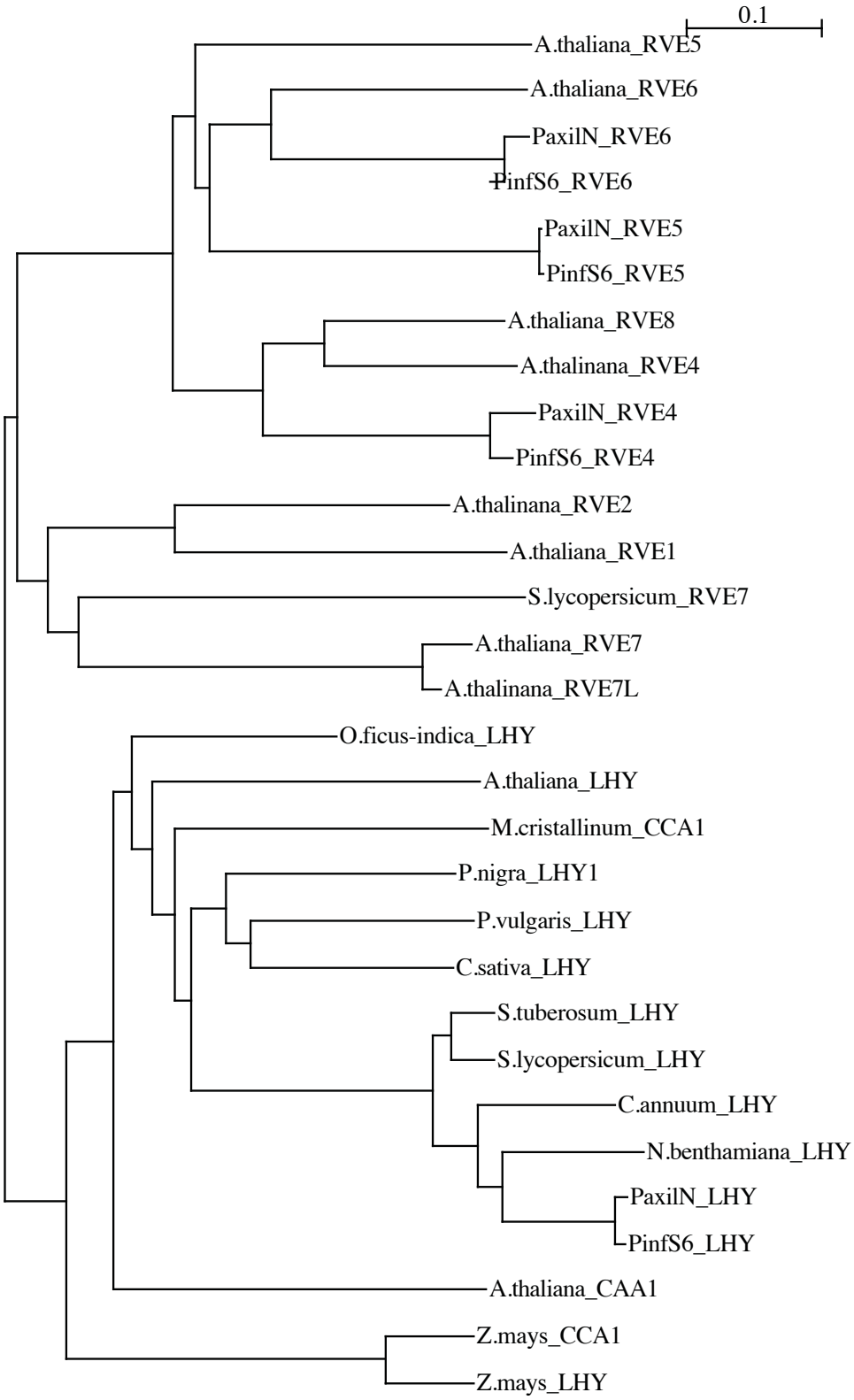


Figure 2. Phylogenetic analysis of MYB genes involved in circadian clock in Petunia and other *Solanaceae*. Species correspond to *Opuntia ficus-indica*; *Mesembryanthemum crystallinum; Populus nigra, Phaseolus vulgaris; Castanea sativa* and *Zea mays*.

EARLY FLOWERING 3 forms together with *ELF4* and *LUX ARRYHTMO (LUX)* a protein complex involved in repression of the morning loop comprising *PRR9*, *7* and *5*. The *ELF3* gene is single copy gene in Arabidopsis. We identified three *ELF3* in *P. axillaris* and four in *P. inflata*, forming two distinct clades (Figure 4). One comprising a cluster of genes together with S.*lycopersicum* and S.*tuberosum*, showed a separate phylogenetic position between two paralogs from rice and the rest of the ELF3 genes from Eudicots, including Arabidopsis. This indicates that this gene may be an ancient ortholog. The second clade comprises what appears to be a further gene duplication with two closely related paralogs in all the species except for *P. axillaris* that presents one gene. The *Physcomitrella patents* genome comprises three *ELF3*-like genes albeit with low homology (blast e-14) forming a distinct clade, indicating that the set of genes aroused in different event.

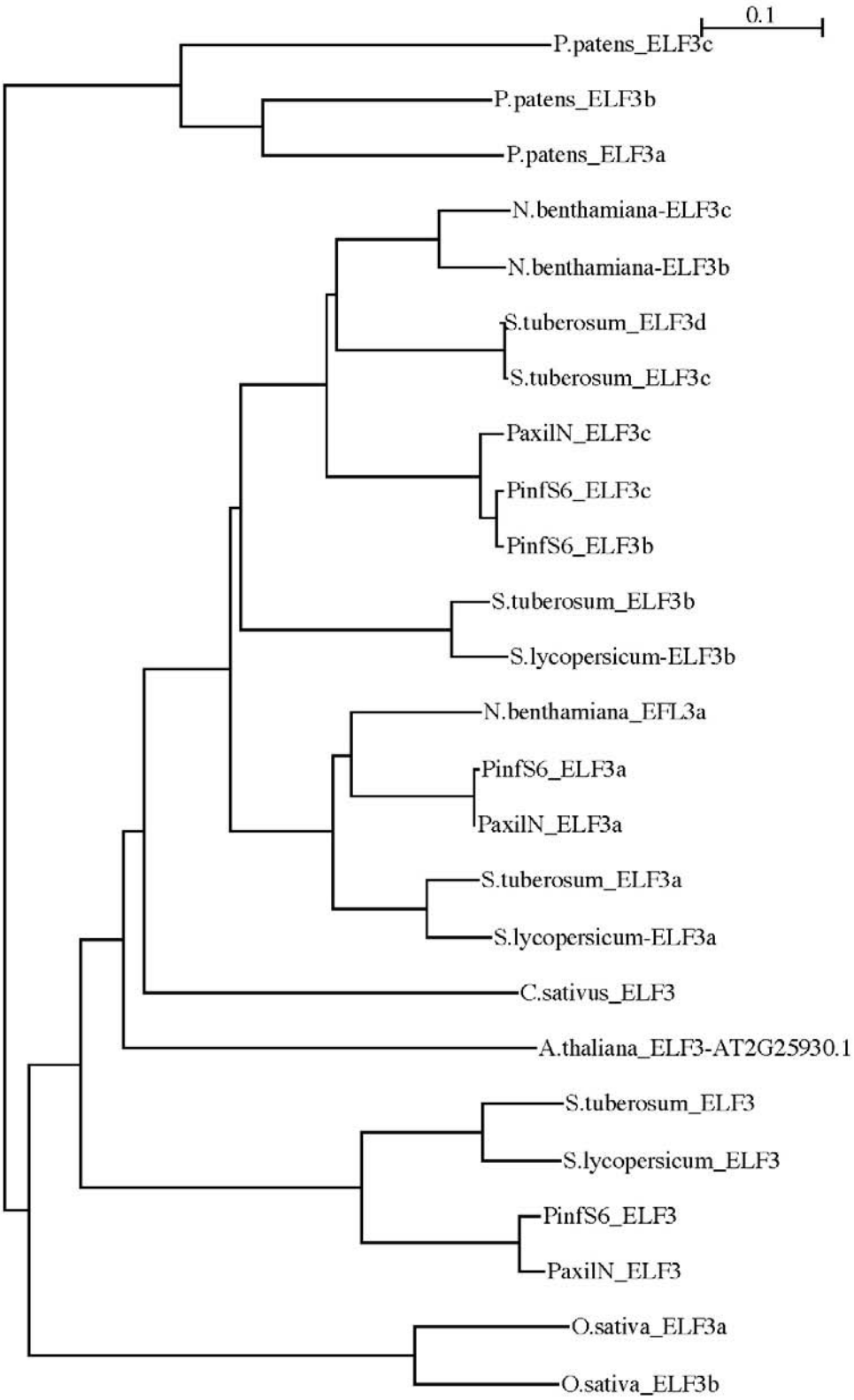


Figure 3. Phylogeny of ELF3 genes in Petunia, *Solanaceae*, *Physcomitrella patens* and *Oryza sativa.*

The *ELF4* gene belongs to a small gene family of five members in Arabidopsis (*ELF4, ELF4-L1 ELF4-L2 ELF4L3*and *ELF4-L4*). *ELF4* encodes a nuclear-localized protein of 111 amino acids with unknown molecular function, and the protein family lacks introns. A biological function is known in detail only for *ELF4*. It is involved in photoperiod perception and flowering time (Doyle et al., 2002; Nusinow et al., 2011; Kikis et al., 2005). We found two genes that clustered with *ELF4* in *P. inflata* and *P.axillaris*, indicating a gene duplication of *ELF4* (Figure 4). We also found duplications in *S.lycopersicum* and *C.anuum* indicating a conservation of this duplication in the *Solanaceae*. There were additional ORFs in Petunia that were closer to the *ELF4-L2 ELF4-L3 and ELF4-L4* from Arabidopsis of unknown biological function (Figure 4).

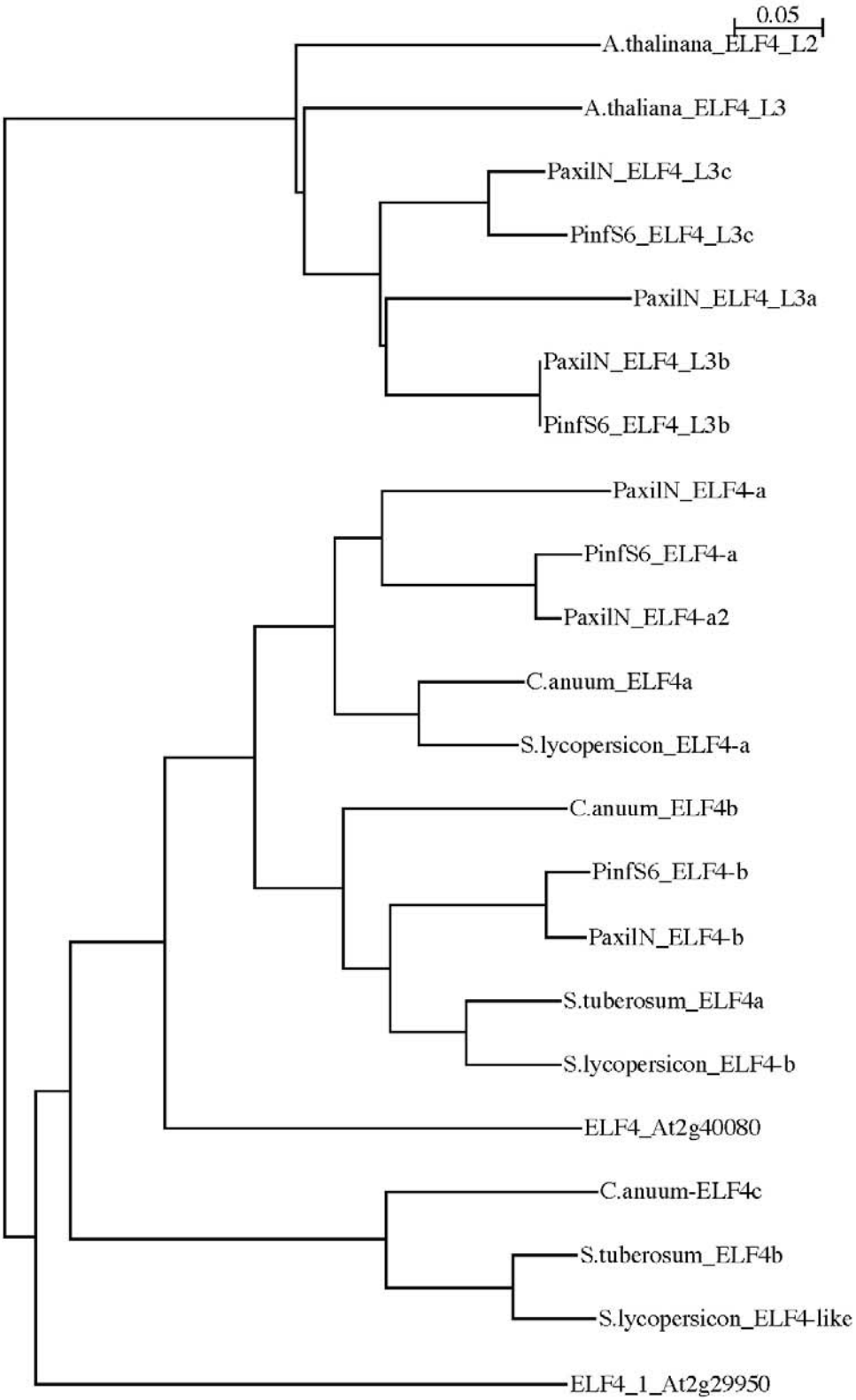


Figure 4. Phylogeny of *ELF4* and *ELF4-like* genes in *Solanaceae* and Arabidopsis.

The *LUX* gene also known as *PHYTOCLOCK1* (*PCL1*) (Hazen et al., 2005; Onai and Ishiura, 2005) is a single *MYB* domain transcription factor of the GARP family (Dubos et al., 2010). The Arabidopsis genome has two copies, *LUX* and *BROTHER OF LUX ARRYHTHMO (BOA)*, also involved in circadian clock regulation (Dai et al., 2011). The so-called evening complex is formed by ELF3, ELF4 and LUX proteins whereby the LUX DNA-binding domain is thought to provide DNA binding capacity (Helfer et al., 2011). A survey of the Petunia genomes showed a single scaffold with high similarity to *LUX* (Figure 5). Additional BLAST searches found other genes but the phylogenetic analysis placed them outside the *LUX* and *BOA* clade. We also found that the *Solanaceae* genes clustered in a subclade while the *Fabaceae* genes formed a second subclade suggesting the acquisition of family specific features.

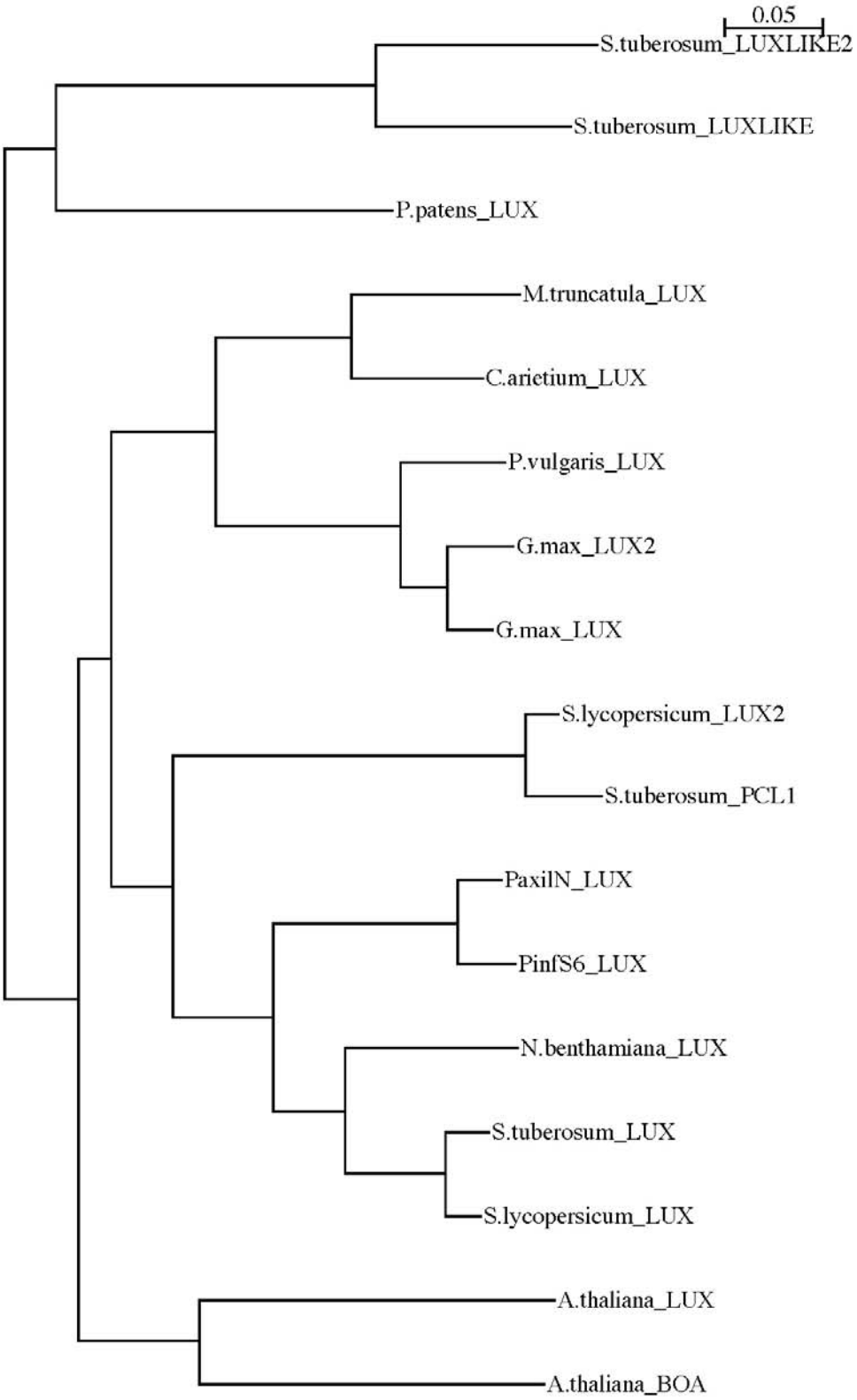


Figure 5. Phylogenetic analysis of *LUX ARRHYTHMO* in Petunia and Solanaceae and comparison to genes from Leguminosae and Arabidopsis.

The *GI* locus is a single copy gene in Arabidopsis. An initial survey in the *Petunia* and other *Solanaceae* genomes indicated the presence of at least two copies of *GI*. The *Solanaceae* genes clustered together, away from Arabidopsis separating the *Fabaceae* clade. The *Solanaceae* clade showed a middle split indicating a conserved genome duplication that seems to be maintained at the *GI* loci. Further gene duplications within one of the subclades have occurred in *PinfS6* and tomato that show an additional gene duplication whereas *N.benthamiana* shows duplications in both (Figure 6).

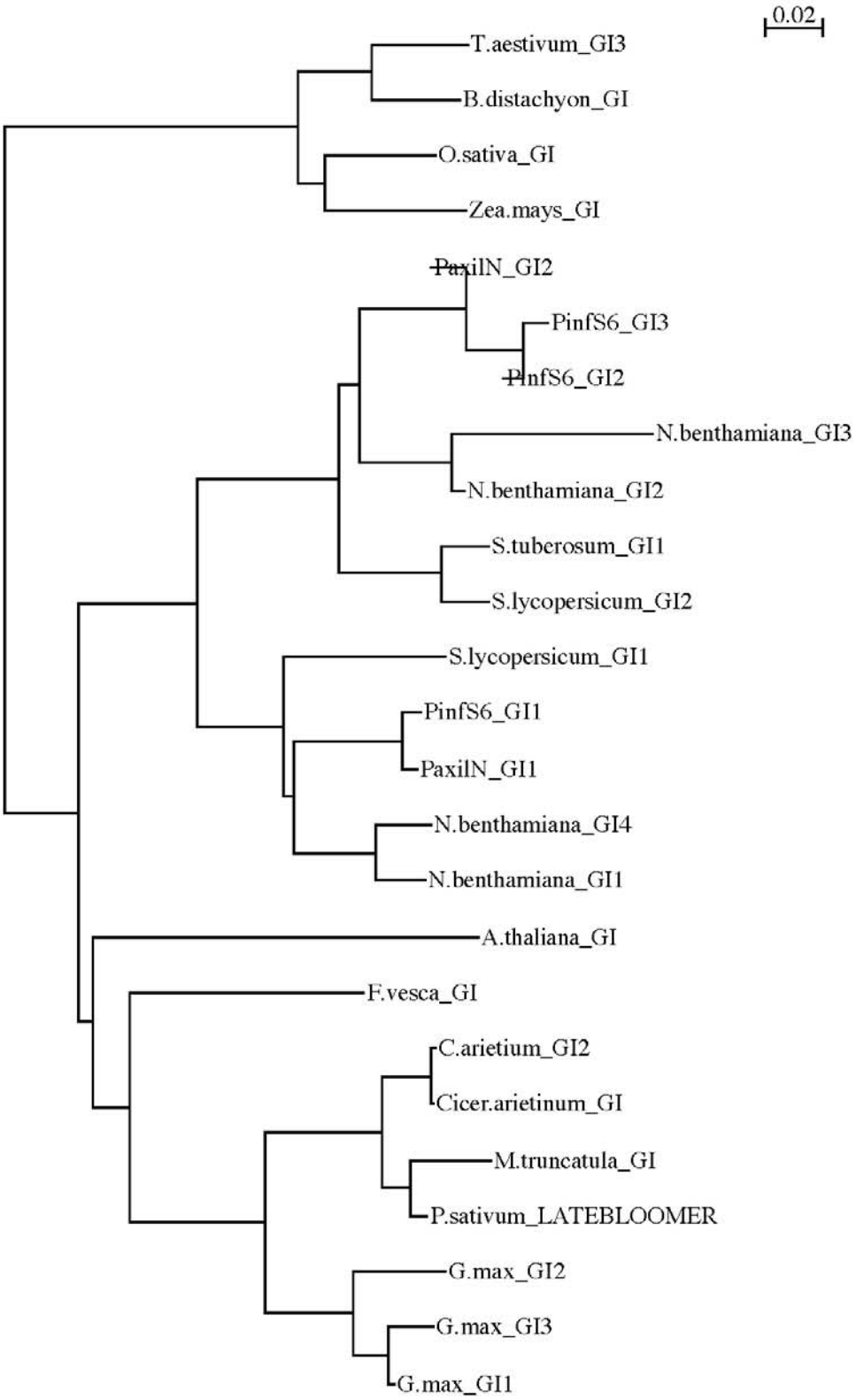


Figure 6. Phylogenetic analysis of GIGANTEA genes from Petunia.

Altogether the clock genes are conserved in Petunia, detailed analysis indicates a departure from the Arabidopsis structure, with some features that are probably common to the *Solanaceae*.

Discussion

The simplest known circadian clock gene regulatory network currently identified in plants is found in the single cell algae *Ostreococcus tauri* and it comprises a core loop formed by a *CCA1* and a *TOC1* ortholog (Corellou et al., 2009). An additional protein with a LOV-domain similar to ZTL-FKF-LKP2 is important for circadian control indicating further similarities in the structure (Djouani-Tahri et al., 2011). Three loops that are apparently conserved in Arabidopsis and other plants form the structure of the clock. Here we have studied the main genes of the morning, core and evening loop of *P.inflata* and *P. axillaris* and performed a comprehensive comparison with other *Solanaceae* that have sequenced genomes available. As the morning, core and evening loops function at the molecular level via protein complexes, a current hypothesis is that gene dosage and protein stoichiometry play a fundamental role in robust pacing (Kim and Forger, 2012; Lou et al., 2012). The *Solanaceae* genomes have undergone genome duplications and they have retained conserved duplicated genes in floral homeotic genes in tomato and Petunia (Quinet et al., 2014; Vandenbussche et al., 2003, 2004). We found that this is not the case for clock genes in *Petunia* and other *Solanaceae*. The morning loop comprises the *PRR* paralogs *PRR9*, *7* and *5*. They form a protein complex with *TOPLESS* and sequentially repress the expression of the core loop genes *CCA1* and *LHY* (Wang et al., 2013). Our analysis shows that *Solanaceae* tend to have a single *PRR9*. In contrast, *PRR7* and *PRR5* are mostly duplicated in the *Solanaceae* analyzed. The *PRR3* gene, involved in selective degradation of *TOC1* (Para et al., 2007) was also found as a single copy gene. A single copy gene in *P.axillaris*, and *P.inflata* represents the core element TOC1. Our data indicates that the *PRR* genes, involved in control of the circadian clock, display a different type of evolution, with *PRR5* and *7* showing ubiquitous conserved duplications and *PRR9*, 3 and TOC1 appearing mostly as single copy genes.

The *CCA1*/*LHY* belong to the *R1/R2* single *MYB* family of transcription factors (Dubos et al., 2010). From an evolutionary perspective, plants like *Lemna*, *Physcomitrella patens*, Arabidopsis or Poplar have two *CCA1*/*LHY* genes resulting from taxon specific genome duplications (Miwa et al., 2006; Dubos et al., 2010; Okada et al., 2009; Takata et al., 2009). In contrast the *Solanaceae* appear to have a single copy of *LHY/CCA1* in agreement with other families like *Oryza, Carica papaya* and *Vitis vinifera* (Lou et al., 2012; Murakami et al., 2007). Furthermore our results support the hypothesis that *CCA1* belongs to a set of genes that originated in the alfa and beta duplication from the Brassicales, as *CCA1* appears in this family but not in others (Lou et al., 2012).

The dose sensitive theory for clock genes may apply to *PRR3* and *LHY*, but may not be true for the rest of the genes analysed, *ELF3, ELF4, FKF* and *GI* that are involved in protein complexes. We found that overall the fate of the different gene families in the *Solanaceae* cannot be predicted based on this simple hypothesis. The *ELF3* gene is found as a single copy gene in Arabidopsis while we found three in the different *Solanaceae* genomes analysed. As a gene with high homology to *ELF3* is involved in heteromorphic self incompatibility in *Fagopyrum esculentum* (Yasui et al., 2012), we can assume a diverging biological function of the *ELF3* paralogs, but functional work is required to test this hypothesis.

The *Solanaceae* genomes apparently share a similar number of *ELF4* genes. However both *P. axillaris* and *P.inflata* have a gene duplication for the closest ortholog of *ELF4*. The *GI* gene plays a key role in adaptation to cold and temperature compensation (Gould et al., 2006; Cao et al., 2007). It is duplicated across the *Solanaceae* and *P.inflata* has three genes suggesting a possible role for *GI* in adaptation to local environments in *P.inflata*, or acquisition of new biological functions.

Altogether the data presented show that the *Solanaceae* has a common genetic structure of the circadian clock. However, the precise set-up in terms of genes present in each species seems to be diverse. This may explain the distinct ecological niches colonized by this family of plants.

Materials and methods

Identification of clock genes

We obtained the translated proteins from Arabidopsis of the clock genes using gene names and accession numbers from (www.arabidopsis.org). The Arabidopsis proteins were used to obtain putative orthologs from tomato using BLASTP (solgenomics.net). The tomato translated protein sequences were used to obtain contigs, scaffolds and proteins from *P. axillaris* and proteins. The contigs and scaffolds obtained were manually trimmed to fragments encompassing 1 kb up and downstream of the putative start and stop codons. The tomato protein and the Petunia genomic fragment were aligned using GeneWise, obtaining the predicted complete coding region and ORF (Birney et al., 2004). Correspondence of exon/intron number to the Arabidopsis genomic structure was visually inspected. We used a simple in-house software (KaraTeka), to manually annotate the exon and intron sequences of the genes analysed.

Phylogenetics

Phylogenetic analysis was performed using CLUSTALX using the Phylip algorithm for Neighbour joining tree using 1000 bootstraps (Larkin et al., 2007) and trees were rendered with NJplot. In

order to verify main trees, we used a pipeline comprising a sequence alignment set for maximum accuracy with MUSCLE (Edgar, 2004). Aligned sequences were automatically curated from poorly aligned positions and blocks with Gblocks (Castresana, 2000). The phylogenetic analysis was performed by PhyML that uses a Near Neighbour Interchange algorithm (Guindon and Gascuel, 2003) and trees were rendered by Treedyn (Chevenet *et al.*, 2006). Gene to gene differences were identified using ALIGN (www.ebi.ac.uk).

Gene ontology annotation

The Gene ontology (Ashburner et al., 2000) descriptors of the genes involved in the clock of Petunia were manually annotated for GO terms using the Arabidopsis homepage (www.arabidopsis.org) and comparing it to GO terms annotated for orthologs of the annotated legume genomes of *Medicago truncatula*, *Glycine max* and *Lotus japonicus* (http://plantgrn.noble.org/LegumeIP/).

Identification of polymorphism

We identified DNA polymorphisms between *PaxilN* and *PinfS6* using KALIGN. Protein differences were identified using multiple alignments obtained with CLUSTALX.

Acknowledgments

This work was funded by Fundación Séneca -11895/PI/09 and 19398/PI/14, Proyecto BFU2013-45148-R. and FEDER. We would like to thank Pedro Navarro for the KaraTeka application and Jose María Jiménez for comments on the manuscript.

References

Alabadi, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Mas, P., and Kay, S.A. (2001). Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. Science (80-.). **293**: 880–883.

Ashburner, M. et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. **25**: 25–9.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res. **14**: 988–995.

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J., and Roux, F. (2010). Linkage and Association Mapping of Arabidopsis thaliana Flowering Time in Nature. Plos Genet. **6**: e1000940.

Cao, S.Q., Song, Y.Q., and Su, L. (2007). Freezing sensitivity in the gigantea mutant of Arabidopsis is associated with sugar deficiency. Biol. Plant. **51**: 359–362.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. **17**: 540–552.

Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B., and Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics **7**: 439.

Corellou, F., Schwartz, C., Motta, J.P., Djouani-Tahri, E., Sanchez, F., and Bouget, F.Y. (2009). Clocks in the Green Lineage: Comparative Functional Analysis of the Circadian Architecture of the Picoeukaryote Ostreococcus. Plant Cell **21**: 3436–3449.

Dai, S., Wei, X., Pei, L., Thompson, R.L., Liu, Y., Heard, J.E., Ruff, T.G., and Beachy, R.N. (2011). BROTHER OF LUX ARRHYTHMO is a component of the Arabidopsis circadian clock. Plant Cell **23**: 961–72.

Djouani-Tahri, E.B., Christie, J.M., Sanchez-Ferandin, S., Sanchez, F., Bouget, F.Y., and Corellou, F. (2011). A eukaryotic LOV-histidine kinase with circadian clock function in the picoalga Ostreococcus. Plant J. **65**: 578–588.

Doyle, M.R., Davis, S.J., Bastow, R.M., McWatters, H.G., Kozma-Bognar, L., Nagy, F., Millar, A.J., and Amasino, R.M. (2002). The ELF4 gene controls circadian rhythms and flowering time in Arabidopsis thaliana. Nature **419**: 74–77.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. Trends Plant Sci. **15**: 573–581.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**: 1792–1797.

Egea-Cortines, M., Ruíz-Ramón, F., and Weiss, J. (2013). Circadian Regulation of Horticultural Traits: Integration of Environmental Signals in Plants. In Horticultural Reviews, J. Janick, ed (Wiley).

Farinas, B. and Mas, P. (2011). Functional implication of the MYB transcription factor RVE8/LCL5 in the circadian control of histone acetylation. Plant J. **66**: 318–329.

Farré, E.M. and Weise, S.E. (2012). The interactions between the circadian clock and primary metabolism. Curr. Opin. Plant Biol. **15**: 293–300.

Fenske, M.P., Hewett Hazelton, K.D., Hempton, A.K., Shim, J.S., Yamamoto, B.M., Riffell, J. a., and Imaizumi, T. (2015). Circadian clock gene *LATE ELONGATED HYPOCOTYL* directly regulates the timing of floral scent emission in *Petunia*. Proc. Natl. Acad. Sci.: 201422875.

Gould, P.D., Locke, J.C.W., Larue, C., Southern, M.M., Davis, S.J., Hanano, S., Moyle, R., Milich, R., Putterill, J., Millar, A.J., and Hall, A. (2006). The molecular basis of temperature compensation in the Arabidopsis circadian clock. Plant Cell **18**: 1177–1187.

Graf, A. and Smith, A.M. (2011). Starch and the clock: the dark side of plant productivity. Trends Plant Sci. **16**: 169–175.

Guindon, S. and Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst. Biol. **52**: 696–704.

Gutierrez, R.A., Stokes, T.L., Thum, K., Xu, X., Obertello, M., Katari, M.S., Tanurdzic, M., Dean, A., Nero, D.C., McClung, C.R., and Coruzzi, G.M. (2008). Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene CCA1. Proc. Natl. Acad. Sci. U. S. A. **105**: 4939–4944.

Haydon, M.J., Mielczarek, O., Robertson, F.C., Hubbard, K.E., and Webb, A.A.R. (2013). Photosynthetic entrainment of the Arabidopsis thaliana circadian clock. Nature.

Hazen, S.P., Schultz, T.F., Pruneda-Paz, J.L., Borevitz, J.O., Ecker, J.R., and Kay, S.A. (2005). LUX ARRHYTHMO encodes a Myb domain protein essential for circadian rhythms. Proc. Natl. Acad. Sci. U. S. A. **102**: 10387–10392.

Helfer, A., Nusinow, D.A., Chow, B.Y., Gehrke, A.R., Bulyk, M.L., and Kay, S.A. (2011). LUX ARRHYTHMO encodes a nighttime repressor of circadian gene expression in the Arabidopsis core clock. Curr Biol **21**: 126–133.

Jarillo, J. a., Piñeiro, M. a., and Pineiro, M.A. (2006). The molecular basis of photoperiodism. Biol. Rhythm Res. **37**: 353–380.

Kikis, E.A., Khanna, R., and Quail, P.H. (2005). ELF4 is a phytochrome-regulated component of a negativefeedback loop involving the central oscillator components CCA1 and LHY. Plant J. **44**: 300–313.

Kim, J.K. and Forger, D.B. (2012). A mechanism for robust circadian timekeeping via stoichiometric balance. Mol. Syst. Biol. **8**: 630.

Kim, S. et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat. Genet. **46**: 270–8.

Kolosova, N., Gorenstein, N., Kish, C.M., and Dudareva, N. (2001). Regulation of circadian methyl benzoate emission in diurnally and nocturnally emitting plants. Plant Cell **13**: 2333–2347.

Larkin, M.A. et al. (2007). Clustal W and clustal X version 2.0. Bioinformatics **23**: 2947–2948.

Lou, P., Wu, J., Cheng, F., Cressman, L.G., Wang, X., and McClung, C.R. (2012). Preferential retention of circadian clock genes during diploidization following whole genome triplication in Brassica rapa. Plant Cell **24**: 2415–26.

Mallona, I., Egea-Cortines, M., and Weiss, J. (2011). Conserved and divergent rhythms of CAM-related and core clock gene expression in the cactus Opuntia ficus-indica. Plant Physiol. **156**: 1978–1989.

Mas, D.P., Wu, D.D., and McClung, C.R. (2013). Beyond Arabidopsis: The circadian clock in non-model plant species. Semin. Cell Dev. Biol. **24**: 430–436.

McClung, C.R. and Gutiérrez, R. a (2010). Network news: prime time for systems biology of the plant circadian clock. Curr. Opin. Genet. Dev. **20**: 588–98.

Miwa, K., Serikawa, M., Suzuki, S., Kondo, T., and Oyama, T. (2006). Conserved expression profiles of circadian clock-related genes in two Lemna species showing long-day and short-day photoperiodic flowering responses. Plant Cell Physiol. **47**: 601–612.

De Montaigu, A., Toth, R., Coupland, G., and Tóth, R. (2010). Plant development goes like clockwork. Trends Genet **26**: 296–306.

Müller, N.A. et al. (2016). Domestication selected for deceleration of the circadian clock in cultivated tomato. Nat. Genet. **48**: 89–93.

Murakami, M., Tago, Y., Yamashino, T., and Mizuno, T. (2007). Comparative overviews of clock-associated genes of Arabidopsis thaliana and Oryza sativa. Plant Cell Physiol. **48**: 110–121.

Navarro, P.J., Fernández, C., Weiss, J., and Egea-Cortines, M. (2012). Development of a configurable growth chamber with a vision system to study circadian rhythm in plants. Sensors **12**: 15356–15375.

Nusinow, D.A., Helfer, A., Hamilton, E.E., King, J.J., Imaizumi, T., Schultz, T.F., Farré, E.M., Kay, S.A., and Farre, E.M. (2011). The ELF4-ELF3-LUX complex links the circadian clock to diurnal control of hypocotyl growth. Nature **475**: 398–U161.

Okada, R., Kondo, S., Satbhai, S.B., Yamaguchi, N., Tsukuda, M., and Aoki, S. (2009). Functional characterization of CCA1/LHY homolog genes, PpCCA1a and PpCCA1b, in the moss Physcomitrella patens. Plant J. **60**: 551–563.

Onai, K. and Ishiura, M. (2005). PHYTOCLOCK 1 encoding a novel GARP protein essential for the Arabidopsis circadian clock. Genes to Cells **10**: 963–972.

Panda, S., Hogenesch, J.B., and Kay, S.A. (2002). Circadian rhythms from flies to human. Nature **417**: 329–335.

Para, A., Farre, E.M., Imaizumi, T., Pruneda-Paz, J.L., Harmon, F.G., and Kay, S.A. (2007). PRR3 is a vascular regulator of TOC1 stability in the Arabidopsis circadian clock. Plant Cell **19**: 3462–3473.

Paulsen, H. and Bogorad, L. (1988). Diurnal and Circadian Rhythms in the Accumulation and Synthesis of mRNA for the Light-Harvesting Chlorophyll a/b-Binding Protein in Tobacco. Plant Physiol. **88**: 1104–9.

Piechulla, B. (1988). Plastid and Nuclear Messenger-Rna Fluctuations in Tomato Leaves - Diurnal and Circadian-Rhythms during Extended Dark and Light Periods. Plant Mol. Biol. **11**: 345–353.

Pokhilko, A., Fernandez, A.P., Edwards, K.D., Southern, M.M., Halliday, K.J., Millar, A.J., and Fernández, A.P. (2012). The clock gene circuit in Arabidopsis includes a repressilator with additional feedback loops. Mol. Syst. Biol. **8**: 574.

Quinet, M., Bataille, G., Dobrev, P.I., Capel, C., Gómez, P., Capel, J., Lutts, S., Motyka, V., Angosto, T., and Lozano, R. (2014). Transcriptional and hormonal regulation of petal and

stamen development by STAMENLESS, the tomato (Solanum lycopersicum L.) orthologue to the B-class APETALA3 gene. J. Exp. Bot. **65**: 2243–2256.

Rawat, R., Schwartz, J., Jones, M.A., Sairanen, I., Cheng, Y.F., Andersson, C.R., Zhao, Y.D., Ljung, K., and Harmer, S.L. (2009). REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. Proc. Natl. Acad. Sci. U. S. A. **106**: 16883–16888.

Rawat, R., Takahashi, N., Hsu, P.Y., Jones, M.A., Schwartz, J., Salemi, M.R., Phinney, B.S., and Harmer, S.L. (2011). REVEILLE8 and PSEUDO-REPONSE REGULATOR5 form a negative feedback loop within the Arabidopsis circadian clock. PLoS Genet. **7**: e1001350.

Salomé, P.A., Weigel, D., McClung, C.R., and Salome, P.A. (2010). The Role of the Arabidopsis Morning Loop Components CCA1, LHY, PRR7, and PRR9 in Temperature Compensation. Plant Cell **22**: 3650–3661.

Sierro, N., Battey, J.N.D., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., Goepfert, S., Peitsch, M.C., and Ivanov, N. V (2014). The tobacco genome sequence and its comparison with those of tomato and potato. Nat. Commun. **5**: 3833.

Stayton, M.M., Brosio, P., and Dunsmuir, P. (1989). Photosynthetic Genes of Petunia (Mitchell) Are Differentially Expressed during the Diurnal Cycle. PLANT Physiol. **89**: 776–782.

Takata, N., Saito, S., Saito, C.T., Nanjo, T., Shinohara, K., and Uemura, M. (2009). Molecular phylogeny and expression of poplar circadian clock genes, LHY1 and LHY2. New Phytol. **181**: 808–819.

Takata, N., Saito, S., Saito, C.T., and Uemura, M. (2010). Phylogenetic footprint of the plant clock system in angiosperms: evolutionary processes of pseudo-response regulators. BMC Evol. Biol. **10**: 126.

Tomato, T. and Consortium, G. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635–41.

Vandenbussche, M., Theissen, G., Van de Peer, Y., and Gerats, T. (2003). Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. Nucleic Acids Res. **31**: 4401–4409.

Vandenbussche, M., Zethof, J., Royaert, S., Weterings, K., and Gerats, T. (2004). The duplicated B-class heterodimer model: Whorl-specific effects and complex genetic interactions in Petunia hybrida flower development. Plant Cell **16**: 741–754.

Verdonk, J.C., Haring, M.A., van Tunen, A.J., Schuurink, R.C., Flowers, P., and Tunen, A.J. Van (2005). ODORANT1 regulates fragrance biosynthesis in petunia flowers. Plant Cell **17**: 1612–1624.

Verdonk, J.C., Ric de Vos, C.. H., Verhoeven, H.A., Haring, M.A., van Tunen, A.J., Schuurink, R.C., and de Vos, C.H.R. (2003). Regulation of floral scent production in petunia revealed by targeted metabolomics. Phytochemistry **62**: 997–1008.

Wang, L., Kim, J., and Somers, D.E. (2013). Transcriptional corepressor TOPLESS complexes with pseudoresponse regulator proteins and histone deacetylases to regulate circadian transcription. Proc. Natl. Acad. Sci. U. S. A. **110**: 761–6.

Weller, J.L. et al. (2012). A conserved molecular basis for photoperiod adaptation in two temperate legumes. Proc. Natl. Acad. Sci. U. S. A. **109**: 21158–63.

Xu, X. et al. (2011). Genome sequence and analysis of the tuber crop potato. Nature **475**: 189–95.

Yakir, E., Hilman, D., Harir, Y., and Green, R.M. (2007). Regulation of output from the plant circadian clock. FEBS J **274**: 335–345.

Yasui, Y., Mori, M., Aii, J., Abe, T., Matsumoto, D., Sato, S., Hayashi, Y., Ohnishi, O., and Ota, T. (2012). S-LOCUS EARLY FLOWERING 3 is exclusively present in the genomes of short-styled buckwheat plants that exhibit heteromorphic self-incompatibility. PLoS One **7**: e31264.

Zakhrabekova, S. et al. (2012). Induced mutations in circadian clock regulator Mat-a facilitated short-season adaptation and range extension in cultivated barley. Proc. Natl. Acad. Sci. U. S. A. **109**: 4326–4331.