Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida

Bombarely, A.; Moser, M.; Amrad, A.; Bao, M.; Bapaume, L.; Barry, C.S.; Bliek, M.; Boersma, M.R.; Borghi, L.; Bruggmann, R.; Bucher, M.; D’Agostino, N.; Davies, K.; Druege, U.; Dudareva, N.; Egea-Cortines, M.; Delledonne, M.; Fernandez-Pozo, N.; Franken, P.; Grandont, L.; Heslop-Harrison, J.S.; Hintzschke, J.; Johns, M.; Koes, R.; Lv, X.; Lyons, E.; Malla, D.; Martinoia, E.; Mattson, N.S.; Morel, P.; Mueller, L.A.; Muhlemann, J.; Nouri, E.; Passeri, V.; Pezzotti, M.; Qi, Q.; Reinhardt, D.; Rich, M.; Richert-Pöggeler, K.R.; Robbins, T.P.; Schatz, M.C.; Schranz, M.E.; Schuurink, R.C.; Schwarzacher, T.; Spelt, K.; Tang, H.; Urbanus, S.L.; Vandenbussche, M.; Vijverberg, K.; Villarino, G.H.; Warner, R.M.; Weiss, J.; Yue, Z.; Zethof, J.; Quattrocchio, F.; Sims, T.L.; Kuhlemeier, C.

DOI 10.1038/nplants.2016.74

Publication date
2016

Document Version
Final published version

Published in
Nature Plants

License
CC BY

Citation for published version (APA):
Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C. S., Bliek, M., Boersma, M. R., Borghi, L., Bruggmann, R., Bucher, M., D’Agostino, N., Davies, K., Druege, U., Dudareva, N., Egea-Cortines, M., Delledonne, M., Fernandez-Pozo, N., Franken, P., ... Kuhlemeier, C. (2016). Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. Nature Plants, 2, Article 16074.
https://doi.org/10.1038/nplants.2016.74
Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida

Aureliano Bombarely16, Michel Moser21, Avichai Amrad2, Manzhu Bao3, Laure Bapaume4, Cornelius S. Barry5, Mattijs Bliek6, Maaike R. Boersma7, Lorenzo Borgi8, Rémy Bruggmann9, Marcel Bucher10, Nunzio D’Agostino11, Kevin Davies12, Uwe Druge13, Natalia Dudareva14, Marcos Egea-Cortines15, Massimo Delledonne16, Noe Fernandez-Pozo17, Philipp Franken13, Laurie Grandont18, J. S. Heslop-Harrison19, Jennifer Hintzsche20, Mitrick Johns20, Ronald Koes6, Xiaodan Lv21, Eric Lyons22, Diwa Malla23, Enrico Martinioa8, Neil S. Mattsson24, Patrice Morel25, Lukas A. Mueller17, Joëlle Muhlemann14, Eva Nouri26, Valentina Passeri6, Mario Pezzotti16, Qinzhou Qi23, Didier Reinhardt27, Melanie Rich28, Katja R. Richert-Poggeler29, Tim P. Robbins30, Michael C. Schatz31, M. Eric Schranz26, Robert C. Schuurink7, Trude Schwarzacher19, Kees Spelt6, Haibao Tang22, Susan L. Urbanus6, Michiel Vandenbussche25, Kitty Vijverberg32, Gonzalez H. Villarino24, Ryan M. Warner5, Julia Weiss15, Zhen Yue21, Jan Zetho32, Francesca Quattrocchio6, Thomas L. Sims23 and Cris Kuhlemeier2*

Petunia hybrida is a popular bedding plant that has a long history as a genetic model system. We report the whole-genome sequencing and assembly of inbred derivatives of its two wild parents, P. axillaris N and P. inflata S6. The assemblies include 91.3% and 90.2% coverage of their diploid genomes (1.4 Gb; 2n = 14) containing 32,928 and 36,697 protein-coding genes, respectively. The genomes reveal that the Petunia lineage has experienced at least two rounds of hexaploidization: the older gamma event, which is shared with most Eudicots, and a more recent Solanaceae event that is shared with tomato and other solanaceous species. Transcription factors involved in the shift from bee to moth pollination reside in particularly dynamic regions of the genome, which may have been key to the remarkable diversity of floral colour patterns and pollination systems. The high-quality genome sequences will enhance the value of Petunia as a model system for research on unique biological phenomena such as small RNAs, symbiosis, self-incompatibility and circadian rhythms.
The garden petunia, *Petunia hybrida*, with its diversity of colour and morphology is the world’s most popular bedding plant with an annual wholesale value exceeding US$130 million in the USA alone. *Petunia* has a long history as a model species for scientific research. To the scientific community, *Petunia* is best known for the discovery of RNAi. This breakthrough was the culmination of decades-long research on the synthesis and regulation of the floral pigments and as a consequence anthocyanin biosynthesis remains one of the best-known pathways of secondary metabolism in any plant species. Development, transposon activity, genetic self-incompatibility, and interactions with microbes, herbivores and pollinators have also been active research topics utilizing *Petunia* as model system.

The genus *Petunia* is a member of the Solanaceae family native to South America. It forms a separate and early branching clade within the family with a base chromosome number of 2 rather than the typical x = 12 found for most Solanaceae group crown–group species, including important crops such as tomato, potato, tobacco, pepper and eggplant. The commercial *P. hybrida* is derived from crosses between a white-flowered, moth-pollinated *P. axillaris*, and species of the *P. integrifolia* clade, a group of closely related bee-pollinated species and subspecies (Fig. 1). The first hybrids were produced by European horticulturists in the early nineteenth century, probably multiple times from different accessions of the two parent clades. The remarkable phenotypic diversity in today’s commercial garden petunias is the result of almost two centuries of intense commercial breeding. Here, we present the genome sequences of *P. axillaris* N and *P. inflata* S6, two inbred laboratory accesses representing the parents of *P. hybrida* (Fig. 1).

**Results and discussion**

**Sequencing, assembly and annotation.** For *P. axillaris*, we performed a hybrid de novo assembly using a combination of short read (Illumina; coverage 137X) and long read technologies (PacBio; coverage 21X), whereas for *P. hybrida*, we produced exclusively short reads (Illumina; coverage 135X) and performed a short read de novo assembly (for details see Supplementary Note 1). The resulting high-quality assemblies have a size of 1.26 Gb for *P. axillaris* and 1.29 Gb for *P. hybrida* (Table 1). The estimated size of both genomes is 1.4 Gb, using a k-mer size of 31, which is consistent with previous microdensity measurements. We have remapped Illumina reads to the assemblies and called single nucleotide polymorphism (SNPs) to estimate the level of heterozygosity, which is estimated as 0.03% for both accessions. Moreover, 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. The estimated unassembled fraction of the genome comprises ~140 Mb for *P. axillaris* (181 Mb if sequence gaps of 41 Mb are included) and ~110 Mb for *P. inflata* (197 Mb with sequence gaps of 87 Mb), which is likely to be due to the large numbers of repetitive sequences (see below). Genome annotation identified 32,928 protein-coding genes for *P. axillaris* and 36,697 protein-coding genes for *P. inflata* with an average of 5.2 and 5.1 exons per protein coding gene and an average predicted protein size of 393 and 386 amino acids, respectively.

**Repeat landscape of *Petunia* genomes.** *Petunia* genomes are rich in repetitive DNA (as are most other plant genomes), but its presence at 60–65% of the assembled genome is relatively low considering its genome size (Fig. 2a; Supplementary Note 2), indicating a larger gene, regulatory and low copy sequence space. Long terminal repeats (LTR)-retroelement-related sequences are abundant near centromeres (Fig. 2b), and within the assemblies, equal numbers of fragments and full-length Ty3/Gypsy-like and Ty1/Copia-like elements were detected. Repeat cluster analysis of unassembled reads supported the amount and complexity of the diverse and rearranged repeat landscape of *Petunia*. *Petunia* chromosomes

---

**Table 1 | Summary statistics of the genome assemblies.**

| Species       | Category     | Number | L50 (kb) | N50 (seqs) | Longest (Mb) | Size (Gb) |
|---------------|--------------|--------|----------|------------|--------------|-----------|
| *P. axillaris*| Total contigs| 109,892| 95.17    | 3,943      | 0.57         | 1.22      |
|               | Total scaffolds| 83,639| 1,236.73 | 309        | 8.56         | 1.26      |
| *P. inflata* S6| Total contigs| 213,254| 34.99    | 9,813      | 0.57         | 1.20      |
|               | Total scaffolds| 136,283| 884.43   | 406        | 5.81         | 1.29      |

The assemblies are version Paeax162 for *P. axillaris* N and version Paein101 for *P. inflata* S6.
average 200 Mb in length (three times that of *Solanum lycopersicum* or *S. tuberosum*), as a larger genome is distributed over 7 rather than 12 chromosomes (Fig. 2b). Chromosomal organization in *Petunia* is thus different compared to other Solanaceae and this together with high DNA transposon frequency and mobility has an effect on genome evolution, meiotic recombination and homogenization events10.

Figure 2 | Genome and repeat organization. **a**, Comparative genome organization of *Solanum lycopersicum*, *P. axillaris* and *Nicotiana tomentosiformis*. The circles are proportional to genome size; regulatory sequences and repeat classes are shown in the segments19,29. **b**, Fluorescent in situ hybridization (FISH) to *P. axillaris* chromosomes (grey). Red: four pericentromeric Petunia vein clearing virus (PVCV) sites; green: dispersed Gypsy-like retroelement junction probe at all centromeres (overlapping yellow signals); blue: 5S rDNA. Scale bar, 10 µm. **c**, Distribution of *dTph1*-like transposons in *P. axillaris* and *P. inflata*. **d**, Duplicated gene families in functional categories showing *Petunia*-specific and balanced families. **e**, Venn diagram based on the gene family cluster analysis from five Solanaceae species. The numbers below the species name indicate the number of protein-coding genes (top) and number of gene family clusters (bottom).
DNA transposons. DNA transposons are five times more abundant in the Petunia genome than in Nicotiana tomentosiformis and S. lycopersicum (Fig. 2a). The identification and cloning of the small endogenous non-autonomous hAT-like defective transposon of petunia hybrida1 (dTph1), which is highly mobile in the P. hybrida line W138 (Fig. 1d), has allowed the development of efficient tools for forward and reverse genetics11. The P. axillaris and P. inflata genomes contain 16 and 21 dTph1 copies, respectively (Fig. 2c and Supplementary Note 3). This is similar to the numbers in most P. hybridaccessions, but far fewer than in the old P. hybrida accession R27 or the hyperactive accession W138 with over 200 copies. Comparison of dTph1 insertion loci in P. axillaris and P. inflata with W138 provides evidence that both species indeed contributed to W138. dTph1 distribution patterns in wild P. axillaris accessions from Uruguay showed comparable low dTph1 copy numbers and a very low overall locus diversity, suggesting that dTph1 transposition activity is largely suppressed in natural populations, but was reactivated after the interspecific crosses leading to the domesticated P. hybrida. Seven previously identified dTph1-like elements and one newly discovered element, dTPh12, are present in both genomes, demonstrating their ancient origin (Fig. 2c). The expansion of different transposable elements—dTph1 in W138 and dTph7 in the two wild species—suggests that, despite extensive homology in their terminal inverted repeat regions, they may require different transacting factors for their mobility.

Endogenous pararetroviruses. Integrated copies of caulimoviridae are widespread in plant nuclear genomes including the Solanaceae12. These DNA viruses are characterized by a gag region with RNA binding domains and a pol region that codes for reverse transcriptase and RNase H (ref. 13). The P. axillaris and P. inflata genomes show near-complete but also degenerated and rearranged copies of Petunia vein clearing virus (PVVCV, a Petuvirus14, Supplementary Note 2). Their structures suggest that the behaviour and mode of integration are similar for both species, and parallel the types of complex rearrangements seen in the banana genome15. Fluorescent in situ hybridization of these sequences (Fig. 2b) showed signals near the centromeres of two chromosome pairs in P. axillaris adjacent to LTR retroelements. Phylogenetic analysis of single insertions showed repeated incidents of homogenization. Such homologous sequences contributed to the tandem array structures found in P. hybrida that are prerequisites of inducible and disease generating viruses14.

Gene families and tandem duplications. Polypeptide sequences from P. axillaris, P. inflata, S. lycopersicum, S. tuberosum, Nicotiana benthamiana and Arabidopsis thaliana were clustered into gene families. This analysis (Supplementary Note 4) grouped 39.2% of the genes into 27,600 gene families, ranging in size from 2 to 1,026 members. Most gene families followed the accepted evolutionary lineage (Fig. 3a), with the Petunia, Solanum and Solanaceae clades sharing gene families far more often than other species groupings (Fig. 2e). Two contrasting sets of gene families that are almost mutually exclusive were found: Petunia-specific families and balanced shared families (Fig. 2d). The size distributions of tandem gene arrays in P. axillaris, P. inflata and S. lycopersicum were quite similar, with each species containing about 8,000 genes in 3,000 tandem arrays.

Paleopolyploidy history of Petunia. Analysis of the Petunia data allowed us to infer the history of polyploidy not only for Petunia but for the entire Solanaceae. Paleopolyploidy is ubiquitous among angiosperms, with many independent lineage-specific paleopolyploidy events associated with changes in genome structure and gene retention and loss16,17. Most paleopolyploidy events are the result of ancient genome duplications (paleotetraploidies), but ancient triplications (paleohexaploidies) have also been identified, for example the gamma genome event near the origin of Eudicots (Fig. 3a) first detected by analysis of the Vitis vinifera (grape) genome18. Similarly, genome analysis of S. lycopersicum suggested that there was a triplication at some point during the evolution of the Solanaceae family19. Petunia as a sister to the x = 12 crown-group clade of the Solanaceae is an ideal species to investigate the timing and nature of this event (Fig. 3a).

Using whole-genome synteny analyses of our de novo assemblies, we identified genomic regions of collinearity between S. lycopersicum and P. axillaris, using V. vinifera as an outgroup (Supplementary Note 5). Inferring their relative timing by analysing synonymous changes (Ks), we show that Petunia shares the older gamma paleopolyploidy event with other higher eudicots, and the more recent paleohexaploidy event with S. lycopersicum. We then can infer that the Solanaceae event occurred at least 30 Myr ago (Fig. 3a). Microsynteny analysis shows the process of gene fractionation following the polyploidization event, and reveals that the S. lycopersicum genome has retained fewer genes than the Petunia genome, thus contributing to the relatively large genic fraction found in Petunia (Fig. 2a). From the fractionation patterns observed, (Fig. 3b), we predict a first and common incomplete gene
Figure 4 | A large fraction of P. hybrida genes may be the result of gene conversion. a, b. Two examples of genes with mixed parentage in P. hybrida accessions Mitchell, R143 and R27. a, PME inhibitor; Peaxi162Scf00002g00042 and Penf101Scf01857g01001 for P. axillaris N and P. inflata S6, respectively. b, Stress-induced phosphoprotein; Peaxi162Scf00002g00511 and Penf101Scf01857g08047 for P. axillaris N and P. inflata S6, respectively. Green and blue circles represent SNPs specific to P. axillaris N and P. inflata S6, respectively. Small black arrows represent SNPs present only in the P. hybrida lines.

Figure 5 | Pollinator attraction. a, Genome dynamics at different MYB gene regions. Genomic regions around AN2-like genes are highly rearranged with few conserved genes, whereas synteny is conserved around the MYB ODO1 involved in scent production. Black arrows, MYB genes. Different coloured arrows, other syntenic genes. Purple blocks, various repeat sequences. b, Biosynthesis of 2-phenylacetaldehyde is different in Petunia and S. lycopersicum. Red and blue arrows depict enzymatic steps characterized in S. lycopersicum and Petunia, respectively. The black arrow represents a predicted activity in S. lycopersicum. c, Biosynthesis of eugenol in Petunia. Although tomato also makes eugenol, homologues of the two genes involved seem to be absent. AADC, aromatic l-amino acid decarboxylase; PAAS, phenylacetaldehyde synthase; CFAT, CoA:coniferyl alcohol acetyltransferase; EGS, eugenol synthase.
fractionation step in both Petunia and S. lycopersicum and a second step after their divergence in S. lycopersicum only. This may have contributed to the separation of the lineages, similar to that observed in Saccharomyces yeasts60 but until now not yet described in flowering plants.

**Origin of the P. hybrida genomes.** Comparisons of the two genome sequences with transcriptomics data from three unrelated *P. hybrida* lines, namely Mitchell, R27 and R143 (Fig. 1d, see Supplementary Note 6) revealed a complex history of the garden petunia. The majority of the ∼20,000 analysed genes could be assigned to *P. axillaris* (~15,000), with only ~600 genes assigned to *P. inflata*. This indicates that the *P. inflata* parent makes only a minor contribution to the *P. hybrida* gene space. One possible explanation for this preponderance of the white parent genome could be that breeding for different colours and colour patterns required a background with recessive mutations in the pigmentation pathway. About 2,000 *P. hybrida* genes contain a high percentage of non-specific SNPs potentially derived from an unknown ancestor.

Approximately 1,500 genes of mixed parentage were identified, with blocks of SNPs similar to *P. axillaris* and other blocks similar to *P. inflata* (Fig. 4). These unusual constellations are conserved between the three *P. hybrida* accessions and may involve gene conversion, random repair of heteroduplexes, contributions of unknown parents or unknown mechanisms. Gene conversion events have been previously reported in plastids21 and polyploids22 but have not been reported before in hybrids (or species of hybrid origin). Definitive answers, especially to the question whether this phenomenon is restricted to transcribed regions will require transcriptome and whole-genome sequencing of multiple *P. hybrida* accessions.

**Genes encoding pollinator attraction traits.** Bee-pollinated *P. inflata* has purple flowers that produce only a limited amount of scent, whereas the flowers of the hawkmoth-pollinated *P. axillaris* are strongly scented and white (Fig. 1a). Colour and scent influence the attraction of pollinators and thereby cause reproductive isolation and ultimately speciation. Speciation of *P. axillaris* from a *P. inflata*-like ancestor involved the loss of anthocyanin pigments and the gain of volatiles4. Thus the genes that caused the changes in these two traits are potential speciation genes. The anthocyanin backbone is synthesized from phenylalanine by the core pathway (CHS, CHI, DFR, ANS, 3GT, 5GT and AAT); however, some of the decorating enzymes are compromised in *P. axillaris*. The steps in the pathway, from DFR on, are regulated by a ternary complex consisting of MYB, bHLH and WD40 transcription factors. The bHLH and WD40 components are functional, but in all *P. axillaris* accessions, the MYB factor AN2 has been inactivated because of independent mutations in the coding region23,24 (Fig. 1d). The only known function of AN2 is to regulate anthocyanin synthesis in petal lobes and this lack of pleiotropic effects makes AN2 a preferred target of selection in the natural habitat.

In *P. hybrida*, four related MYB factors activate the anthocyanin biosynthetic pathway in different tissues: AN2 controls anthocyanin deposition in the petal limb, AN4 in the anthers and DPL and PHZ in green tissues. Unlike AN2, the AN4, DPL and PHZ coding sequences have remained intact in *P. axillaris*. Based on *P. hybrida* data, differential expression of AN4 might be responsible for the shift in anther colour from purple in *P. inflata* to yellow in *P. axillaris*.

The genomic regions containing these four MYB genes have undergone massive rearrangements since the separation of the two species estimated at 0.9 Myr ago, possibly influenced by transposon or retroelement activities found in the vicinity (Fig. 5a). As a consequence, the synteny between the corresponding regions of *P. axillaris* and *P. inflata* has been largely destroyed and gene spacing altered. *P. axillaris* AN4 is duplicated and inactivated subsequently in anthers because of large insertions of transposon-like sequences in the promoter. Similar insertion events are visible around the other anthocyanin MYB genes. Instead, the genomic regions containing other anthocyanin regulators (AN1, JAF13, AN11) and other MYBs involved in vacuolar pH regulation and scent production show strong conservation of the synteny between the two *Petunia* species. Thus, the AN2-like MYBs reside in an exceptionally dynamic region of the genome. Although lack of pleiotropy makes AN2-like MYBs preferential targets of selection, genomic rearrangements may have provided the mechanism responsible for the remarkable spatial and temporal diversity of anthocyanin pigmentation patterns.

Exceptional dynamics of the regions containing the MYB regulators of the anthocyanin pathway is not restricted to *Petunia*. The regions in *S. lycopersicum* share little synteny with either of the two *Petunia* species indicating that large rearrangements occurred after the separation of the genera. In the more distantly related *Mimulus guttatus*, we also find duplications and rearrangements to have taken place after the separation of the ancestors of Solanaceae and Phrymaceae. Thus, genome dynamics of AN2-type MYB factors may be a general mechanism that caused the diversity of floral pigmentation patterns across angiosperms.

*P. axillaris* emits an abundant blend of floral benzenoid and phenylpropanoid volatiles whereas *P. inflata* only emits benzaldehyde. A comparison of all structural and regulatory genes known to be involved in floral scent synthesis indicates that all the known biosynthetic and regulatory genes encode functional proteins (Supplementary Note 8). Thus, the increase in complexity and concentration of volatiles accompanying the shift to moth pollination in *P. axillaris* involved mutations in cis-acting regulatory elements or the mutation of as yet unknown transcriptional regulators.

*Petunia* uses a single enzyme for the biosynthesis of 2-phenylacetaldehyde23 whereas *S. lycopersicum* utilizes an amino acid decarboxylase plus a yet unidentified amine oxidase (Fig. 5b). Interestingly, the *S. lycopersicum* genome does harbour a homologue of the *Petunia* gene, but this is predicted to be 124 amino acids shorter than its *Petunia* homologue and presumably inactive. Furthermore, although *S. lycopersicum* is also known to produce euugenol27, homologues of the two involved enzymes appear to be

---

**Box 1 | MicroRNAs.**

Small RNA sequencing of young flower buds and mapping of candidates in the genome sequences confirm the presence of 44 conserved miRNAs in *Petunia*, belonging to 30 families and corresponding to 140 MIR loci, in line with other species. Only two loci were unique, and an unannotated miRNA (MIR171d) and two unique and one truncated to *P. inflata* (MIR160d, MIR397b, MIR477), and MIR sequences and organization in the genome were largely conserved. MiRNA expression profiles were also highly similar between the two species and comparable to tomato and potato29, with highest expressions for miR166, miR159 and miR319, known for their involvement in floral organ development, and low expression for miRNA169c (BLIND), a well-studied regulator of floral whorl identity in *Petunia* (Fig. 1d) and Antirrhinum28. Predicted miRNA targets included genes involved in development and metabolic pathways. For additional information, see Supplementary Note 9.
Box 2 | Symbiosis with fungi.

A good example of Petunia as a model system is the study of the symbiotic interactions with fungi. In the arbuscular mycorrhiza (AM) with the Glomeromycota, the fungal hyphae function as an extension of the root system that enhances the acquisition of mineral nutrients, primarily phosphate. This symbiosis is widespread amongst land plants, but is lacking in the Brassicaceae (including Arabidopsis thaliana). Symbiosis signaling involves a family of lysine motive (LysM) receptor-like kinases (LysM-RLK) in the host, which perceive specific microbial signals44. LysM-RLKs activate a signalling pathway that is shared with root nodule symbiosis, a symbiotic interaction restricted to the Fabaceae. The comparison between Petunia and tomato versus the legumes Lotus japonicus and Medicago truncatula revealed that the two Solanaceae have considerably smaller gene families than the two legumes with 10/11 and 14 versus 17 and 18 LysM-RLK members, respectively. The expansion of the LysM-RLK family during the evolution of nodule symbiosis in the legumes45 will help to understand the evolution of two fundamentally different symbiotic interactions. Large-scale genomic and transcriptional analysis of two transcription factor families revealed that several GRAS genes are regulated during AM symbiosis, whereas AP2/ERF genes are induced during adventitious root formation. For additional information, see Supplementary Note 10.

Box 3 | Self-incompatibility.

In the Solanaceae, self-fertilization is prevented by S-RNase-mediated gametophytic self-incompatibility (GSI), which is based on the ability to reject pollen from a plant expressing a matching S-locus haplotype, while accepting pollen from individuals whose haplotypes do not match that of the styal parent46. During an incompatible pollination, the growth of self pollen tubes is inhibited by the action of an imported, cognate, ribonuclease, the S-RNase. In compatible pollinations, the action of non-self S-RNases is inhibited by ubiquitination and degradation by a SCF complexes. Our long-range sequencing strategy is an important mechanism to preserve co-adapted gene combination is an important mechanism to preserve co-adapted gene complexes. Our long-range sequencing strategy is an important step towards characterizing a complete S-locus, as a means to better understand the evolution of gametophytic self-incompatibility. For additional information, see Supplementary Note 11.

Box 4 | The circadian clock.

Floral volatiles serve to attract pollinators but will also be perceived by unwanted herbivores48. Volatile emission in P. axillaris is under circadian control, peaking at dusk when its nocturnal hawkmoth pollinator visits49. Although this specific output is known quite well, the genetic structure of the circadian clock itself is understudied in Petunia. In Arabidopsis, the clock consists of three loops, the morning, core and evening loop, based on their expression pattern. The current hypothesis is that gene dosage effect is important for clock function50. Comparison of the Petunia genomes with other Solanaceae indicates that the circadian clock has undergone a deep restructuring in the Solanaceae and each species seems to have a different set of genes (Supplementary Note 12). Petunia and the rest of the Solanaceae share single orthologues for LHY, TOC1, PRR3 and the MYB transcription factor LUX. In contrast PRR7, PRR5, GIGANTEA, ELF3 and ELF4 are in some cases duplicated or triplicated. P. inflata has a larger number of evening loop paralogues than P. axillaris. The GI gene is present in two copies in P. axillaris with three in P. inflata, and there are three ELF3 copies in P. axillaris and four in P. inflata. These results suggest strong purifying selection on some of the clock components but others may have undergone a rapid subfunctionalization or redeployment. Comparative analyses will help to understand how clock structure is adjusted to optimize specific outputs thus allowing adaptation to different environments. For additional information, see Supplementary Note 12.
with 454 and Illumina RNA-Seq reads and protein sequences from different protein databases. RNA-Seq Illumina data was mapped using TopHat2. RNAseq were annotated using IRNascn (http://lovelab.ucsd.edu/IRNAscan-SE/).

The gene functional annotation was performed by sequence homology search with different protein datasets using BlastP and protein domains search using InterProScan. Functional annotations were integrated using AHHRD (https://github.com/groupschoo/AHHRD). See Supplementary Note 1.

Repetitive elements analysis. Repeat annotation was performed using RepeatModeler (v1.0.8: http://www.repeatmasker.org/RepeatModeler.html), RepeatMasker (v4.0.5: http://www.repeatmasker.org/) with the repeat database Repbase (release 20140131; http://www.girinst.org/rebase/) and Genious (v7.1.4; http://www.geneious.com). Identification of PCVY-like and EPRV elements was performed using BlastN and TBLASTN. The identified sequences were aligned with ClustalW (MEGA5 package; http://www.mega-software.net/) and then manually curated. RepeatExplorer (http://www.repeatexplorer.org/) and other methods were used to extend the analysis to unannotated regions. Fluorescent in situ hybridization was performed in root tips from young P. axillaris and P. inflata plants for SS rDNA and three PCVY viral probes following the procedure described in Supplementary Note 2.

The detection of Dph1 loci in P. hybridum W138 was performed through a BLASTn search of the P. axillaris and P. inflata dph1 elements including the 500 bp of flanking sequence against the TTS W138 collection. Polymorphisms found in the genomic flanking regions were used to identify the species of origin.

FSH1 elements were identified in a P. axillaris population using a modification of the methodology described in Supplementary Note 3.

Whole-genome duplication, tandem duplications and gene family analysis. Whole-genome collinear analysis was performed using SynMap and microsynteny analysis were performed using GEvo in the comparative genomics platform, CoGe.

Received 12 October 2015; accepted 22 April 2016; published 27 May 2016

References

1. National Agricultural Statistics Service. Floriculture Crops 2014 Summary (US Department of Agriculture, 2015; http://usa.mannlib.cornell.edu/usa/nass/FlorCrop/2015/FlorCrop-06-04-2015.pdf)
2. Napoli, C., Lemieux, C. & Jorgensen, R. Introduction of chimeric chalcone synthase gene into Petunia results in reversible co-suppression of homologous genes in trans. Plant Cell 2, 279–289 (1990).
3. van der Krol, A. R., Mur, L. A., Beld, M., Mol, J. N. M. & Stuitje, A. R. Flavonoid genes in Petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell 2, 291–299 (1990).
4. Sheehan, H., Hermann, K. & Kuhlemeyer, C. Color and scent: how single genes influence pollinator attraction. Cold Spring Harb. Symp. Quant. Biol. 77, 117–133 (2012).
5. Saerkenin, T., Bohn, L., Olmeda, R. G. & Knapp, S. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae) a dated 1000-tip tree. BMC Evol Biol 13, 214 (2013).
6. Segato, A. L. A., Ramos-Fregonezi, A. M. C., Bonatto, S. L. & Freitas, L. B. Molecular insights into the purple-flowered ancestor of garden petunias. Am. J. Bot. 101, 119–127 (2014).
7. Stelmach, J. R., Lorenz-Lemke, A. P., Freitas, L. B. & Semir, J. in Petunia: Evolutionary, Developmental and Physiological Genetics 2nd edn (eds Gerats, T. & Strother, J.) 1–28 (Springer, 2009).
8. Sink, K. C. in Petunia: Morphology on Theoretical and Applied Genetics Vol 9 (ed. Sink, K. C.) 3–9 (Springer, 1984).
9. White, J. & Rees, H. Chromosome weights and measures in Petunia. Heredity 58, 139–143 (1987).
10. Heslop-Harrison, J. Genome evolution: extinction, continuation or explosion? Curr. Opin. Plant Biol. 15, 115–121 (2012).
11. Vandenbussche, M. et al. Generation of a 3D indexed Petunia insertion database for reverse genetics. Plant J. 54, 1105–1114 (2008).
12. Geering, A. D. W. et al. Endogenous floridoviruses are major components of plant genomes and hallmarks of virus evolution. Nature Commun. 5, 1269 (2014).
13. Hansén, C. & Heslop-Harrison, J. Sequences and phylogenies of plant pararetroviruses, viruses, and transposable elements. Adv. Bot. Res. 41, 165–193 (2004).
14. Richert-Poggeler, K. R., Nooren, F., Schwarzacher, T., Harper, G. & Hohn, T. Induction of infectious petunia vein clearing (pararetrovirus) virus from endogenous provirus in petunia. EMBO J. 22, 4836–4845 (2003).
15. D’Hont, A. et al. The banana (Musa acuminate) genome and the evolution of monocotyledonous plants. Nature 488, 213–217 (2012).
16. Iiao, Y. et al. Ancestral polyplody in seed plants and angiosperms. Nature 473, 97–100 (2011).
17. Solís, D. E., Bell, C. D., Kim, S. & Solís, P. S. Origin and early evolution of angiosperms. NY. Acad. Sci. 1133, 3–25 (2008).
18. Jaillon, O. et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449, 463–467 (2007).
19. Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485, 635–641 (2012).
20. Scanellier, D. R., Byrne, K. P., Gordon, J. L., Wong, S. & Wolfe, K. H. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature 440, 341–345 (2006).
21. Khakhlova, O. & Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J. 46, 85–94 (2006).
22. Salmon, A., Uddal, J. A., Jeddolah, J. A. & Wendel, J. Targeted capture of homoeologous coding and noncoding sequence in polyploid cotton. G3 2, 921–930 (2012).
23. Quattrocchio, F. et al. Molecular analysis of the anthocyanin2 gene of Petunia and its role in the evolution of flower color. Plant Cell 11, 1433–1444 (1999).
24. Hoshino, M. et al. Single-gene-mediated shift in pollinator attraction in Petunia. Plant Cell 19, 779–799 (2007).
25. Kaminaga, Y. et al. Plant phenylacetaldehyde synthase is a bifunctional homoeomeric enzyme that catalyzes phenylalanine decarboxylation and oxidation. J. Biol. Chem. 281, 23557–23566 (2006).
26. Verman, D. et al. Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. Proc. Natl Acad. Sci. USA 103, 8287–8292 (2006).
27. Tikuonu, Y. M. et al. Non-smoky glycosyltransferasel prevents the release of smoky aroma from tomato fruit. Plant Cell 25, 3067–3078 (2013).
28. Cartolano, M. et al. A conserved microRNA module exerts homeostatic control over Petunia hybrids and Antirrhinum majus floral organ identity. Nature Genet. 9, 901–905 (2007).
29. Sierro, N. et al. Reference genomes and transcriptomes of Nicotiana sylvestris and Nicotiana tomentosiformis. Genome Biol. 14, R60 (2013).
30. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read assembler. GigaScience 1, 18 (2012).
31. Bashir, A. et al. A hybrid approach for the automated finishing of bacterial genomes. Nature Biotechnol. 30, 701–707 (2012).
32. English, A. C. et al. Mind the gap: upgrading genomes with pacificbiosciences RS long-Read sequencing technology. PloS One 7, e47768 (2012).
33. Botterer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-mRNAs with RepeatModeler (v1.0.8) and Solanaceae genomes by gene conversion. Plant Cell 25, 827–839 (2013).
34. Botterer, M. et al. A conserved microRNA module exerts homeostatic control over Petunia hybrids and Antirrhinum majus floral organ identity. Nature Genet. 9, 901–905 (2007).
35. Sierro, N. et al. Reference genomes and transcriptomes of Nicotiana sylvestris and Nicotiana tomentosiformis. Genome Biol. 14, R60 (2013).
36. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read assembler. GigaScience 1, 18 (2012).
37. Bashir, A. et al. A hybrid approach for the automated finishing of bacterial genomes. Nature Biotechnol. 30, 701–707 (2012).
38. English, A. C. et al. Mind the gap: upgrading genomes with pacificbiosciences RS long-Read sequencing technology. PloS One 7, e47768 (2012).
39. Botterer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-mRNAs with RepeatModeler (v1.0.8) and Solanaceae genomes by gene conversion. Plant Cell 25, 827–839 (2013).
35. Cingolani, P. et al. A program for annotating and predicting the effects of nucleotide polymorphisms, SnpEff. Fly 6, 80–92 (2014).
Acknowledgements

We dedicate this work to Tom Gerats in honour of his lifelong contributions to petunia research, and we thank him for generously sharing his vast knowledge with all of us.

We thank H. Puchta for advice on Fig. 4, R. Koipli for IT and graphics support, and K. Esfeld for carefully reading the manuscript. This work was carried out without dedicated funding. We acknowledge following agencies who support our work on Petunia: NWO-ALW grant 022.001.018 (M.B); Marie Curie Independent Fellowship (L.G.); Swiss NSF grant 31003A_159493 and NCCR Plant Survival (C.K.); NWO-TOP grant 854.11.006 (R.K.); Borsa di Studio Lanzoni per Genetica Agraria, Accademia dei Lincei (V.P.); NWO-ALW grant 820.02.015 (K.V.); CNRS ATIP-AVENIR award (M.V.); Deutsche Forschungsgemeinschaft grant DR411/2-1 (U.D. and P.F.); Swiss NSF grant 31003A_135778 (D.R.).

Author contributions

The authors are listed in alphabetical order except for the first two and the last three. A.B., F.Q., T.Si. and C.K. conceived and planned the work. All authors wrote or commented on the main text and supplementary notes; A.B., M.M., A.A., L.B., C.B., M.Bl., M.Bo., D.B., N.D., N.F-P., L.G., I.H., I.H.-H., M.I., R.K., X.L., E.L., D.M., E.M., N.M., P.M., J.M., E.N., V.P., Q.Q., D.R., M.R., K.R.-P., T.R., E.S., R.S., T.Sc., C.S., H.T., S.U., M.V., K.V., G.V., R.W., J.W., Z.Y., J.Z. and F.Q. performed the experiments and analysed the data; R.B., M.D., X.L., M.P., M.S., Z.Y., T.Si. and C.K. contributed sequencing data and analysis tools.

Additional information

Supplementary information is available online. Reprints and permissions information is available online at www.nature.com/reprints. General correspondence should be addressed to C.K.; correspondence about sequencing and bioinformatics should be addressed to A.B. (aurebg@vt.edu); and requests for materials should be addressed to the indicated authors in each Supplementary Note.

Competing interests

The authors declare no competing financial interests.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/