Current Biology, Volume *26*

Supplemental Information

Gain and Loss of Floral Scent Production through Changes in Structural Genes during Pollinator-Mediated Speciation

Avichai Amrad, Michel Moser, Therese Mandel, Michel de Vries, Robert C. Schuurink, Loreta Freitas, and Cris Kuhlemeier

Supplemental Experimental Procedures

Genotyping and marker association analyses

Genotyping of IL3-2 accessions for marker regression analysis was performed with the EOBII CAPS marker. Genotypic classes were not normally distributed (Shapiro-Wilk test), so tests for association of means with *EOBII* genotype were carried out using the Kruskal-Wallace one-way analysis of variance in R. Wild accessions were genotyped using Sanger sequencing with AA001 & AA004 primers (see Oligonucleotides Table below). Progeny of putative hybrid wild accessions were genotyped using a CNL1 CAPS marker based on a polymorphism located 18 bp upstream of the nonsense mutation. All details for CAPS markers can be found on http://www.botany.unibe.ch/deve/caps/index.html

Quantitative RT-PCR

Flower limb samples were collected one day after anthesis at 15:00, from three plants representing three biological replicates. RNA extraction was carried out using the RNeasy Plant Mini Kit (Qiagen). After extraction, RNA samples were treated with DNAse I (Sigma-Aldrich) to remove any remaining genomic DNA, and the quality of the RNA was subsequently measured on a 2100 Bioanalyzer (Agilent). Samples with RNA Integrity Numbers (RIN) less than 7.5 were discarded. First strand synthesis was performed using Transcriptor Universal cDNA Master (Roche) containing random hexamer primers, according to the manufacturer's recommendations. Quantitative RT-PCR experiments were performed using a LightCycler® 96 Real-Time PCR System (Roche) with FastStart Essential DNA Green Master qPCR Kit (Roche), according to the manufacturer's recommendations. Reactions were run in triplicates for each gene. Primers are given in the Oligonucleotides Table. Cycle of quantification (Cq) thresholds and normalization calculations were determined by the LightCycler® 480 Software (v.1.1.0.1320; Roche). Each biological replicate of each species was analyzed on a separate PCR plate and a single *P. axillaris* 15:00 sample from one biological replicate was included in each plate as a normalizer. Integration of data from biological replicates was conducted manually. Standard curves were performed to determine the PCR efficiency of each set of primers. No reverse-transcription controls were performed for each sample. Previously identified reference genes [S1] were tested for stability of expression across different *P. axillaris* and *P. exserta* samples and NormFinder [S2] was used to perform stability tests. *EF1α* and *SAND* were shown to be the most stable reference genes and were included for each sample.

RNA sequencing

For all experiments, three biological replicates were used for each genotype, and flower limb tissue was harvested one day after anthesis around 14:50 (at 15:00 is the transition from light to dark). For experiment one (comparisons of *P. axillaris*, *P. exserta* and IL3-2), biological replicates of *P. axillaris*, *P. exserta*, IL3-2$^{P.axi}$, IL3-2^{het} and IL3-2$^{P.exs}$ were sequenced. One of the *P. exserta* replicates was proven to be contaminated and was discarded from further analysis. For experiment two RNA was sequenced from *P. axillaris*, *P. inflata*, *P. exserta* and *P. axillaris* x *P. inflata* F1. RNA extraction was carried out using the RNeasy Plant Mini Kit (Qiagen).

RNA was sequenced in the Lausanne Genomic Technologies Facility (Lausanne, Switzerland). Quality of RNA was checked using a Fragment Analyzer (Advanced Analytical). For experiment one, RNA Quality Numbers (RQN) ranged from 5.9 to 8.8; for experiment two, RQN ranged from 6.6 to 9.8. Cluster generation was performed with the sequencing libraries using the Illumina TruSeq PE Cluster Kit (v.3). Samples from experiment one were paired-end sequenced and samples from experiment two single-end sequenced for 100 bp. Sequencing data were processed using the Illumina Pipeline Software v.1.82. For all samples, raw reads were checked for contamination by aligning them against rRNA sequences from *P. axillaris*, the *Escherichia coli* genome and the human transcriptome using bowtie2 (v.2.2.1) [S3]. Reads aligning to mentioned sequences were discarded. FastqMcf (v.1.1.2-686; http://code.google.com/p/ea-utils/wiki/FastqMcf) was used to remove Illumina adapter sequences and trim low quality regions. After trimming, reads shorter than 60 bp were discarded. These pre-processed reads were mapped against the draft reference genome of *P. axillaris* (v.1.6.2) [S4] using the SNP-tolerant and splice-aware aligner GSNAP (v.2015-06-12) [S5]. To allow for variation between reference and reads, the "–m" option was set to 0.04. Variant calling was carried out in GATK (v.3.4.0) [S6] according to GATK best practices for RNA-seq data. After duplicate marking and splitting reads with N in their CIGAR string, local realignment

around indels was undertaken, and base quality scores were recalibrated, using a set of high quality SNPs determined by an initial run of the GATK HaplotypeCaller. Those SNPs were used to determine genes being inside or outside of the introgressed region in IL3-2. Only exonic positions with maximal upstream distance of 100 bp from start codon and maximal downstream distance of 100 bp from stop codon were considered to detect allele-specific expression.

Differential expression analyses

Reads mapping to the genes of interest were counted using HTseq (v.0.6.1) in the union mode. Differential expression analysis was performed in R (http://www.R-project.org/) using DESeq2 (v.1.8.1) [S7] with the default parameters, including the Cook's distance treatment to remove outliers. Normalized read counts are presented in figures. Allelic coverage for variant positions was detected with ASEReadCounter implemented in GATK [S6,S8]. Analyses of allelic imbalance were conducted in R (http://www.R-project.org/) with the package MBASED (v.1.2.0) [S9]. Parameters of read count overdispersion were estimated with a custom R script provided by the author of the MBASED package (available on request).

Transient expression experiment

Transformation assays were done according to Van Moerkercke et.al 2011 [S10]. One-day-old flowers were syringe-infiltrated and two days later flower limbs were harvested. Tissue of four flowers were combined and homogenized in liquid nitrogen and transferred to a glass vial containing 2 ml of 1M NaCl 60°C. Immediately and following a short vortex samples were analyzed with PTR-MS for 30 cycles with an air inlet of approximately 20 L/h. Values were averaged and normalized to tissue weight before statistical analysis.

List of primers

Purpose	Gene/element	Primer name	Primer sequence (5' to 3')	Reference
Quantitative RT-PCR, reference gene	*SAND*	SAND-F	CTTACGACGAGTTCAGATGCC	[S1]
Quantitative RT-PCR, reference gene	*SAND*	SAND-R	TAAGTCCTCAACACGCATGC	[S1]
Quantitative RT-PCR, reference gene	*EF1α*	EF1α-F	CCTGGTCAAATTGGAAACGG	[S1]
Quantitative RT-PCR, reference gene	*EF1α*	EF1α-R	CAGATCGCCTGTCAATCTTGG	[S1]
Quantitative RT-PCR	*CNL*	qCNL_UF	GGGTACTTCAAGAATGACAAGGC	
Quantitative RT-PCR	*CNL*	qCNL_UR	CACCTAGGATGTGGCATGGC	
Genotyping by CNL1 sequencing	*CNL1*	AA001	TTCCGGCCAGTAAGTTATGG	
Genotyping by CNL1 sequencing	*CNL1*	AA004	TGTTGAAGTGTCTGACTGGTCA	
P. axillaris/P.exserta sequence	*CNL1*	cdsCNL_F	ATGGACGAGTTACCAAAATGTGG	
P. axillaris sequence	*CNL1*	cdsCNL_R	CTACAGACGAGCTGGCAAAT	
P. exserta sequence	*CNL1*	exsCNL1_R3	TCTGGGACGTTTGATTTGCA	
Gibson Assembly	*CNL1* cds	B470	CAACAAACAACATTACAATTTACTATTCTAGTCGAATGG ACGAGTTACCAAAATGTG	
Gibson Assembly	*P. axillaris CNL1* cds	B471	AGCTCAGACTAGGTGGATCTCTACAGACGAGCTGGCAA	
Gibson Assembly	*P. exserta CNL1* cds	B570	CTCAGACTAGGTGGATCTCTAGTAGTGACACATTGAATC TGG	
Gibson Assembly	35S promoter	B468	GGCCTCTTCGCTATTACGCCAGACTAGAGCCAAGCTGAT	
Gibson Assembly	35S promoter	B469	CATAGTTTGCTCCACATTTTGGTAACTCGTCCATTCGACT AGAATAGTAAATTGTAATGTTG	
Gibson Assembly	pGreenII vector	B477	GTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGG ATGGCAAGATCCTG	
Gibson Assembly	pGreenII vector	B473	GCAAAGGAGATCAGCTTGGCTCTAGTCTGGCGTAATAG CGAAGAG	
Gibson Assembly	pGreenII vector	B472	AGATTCTTGATTTGCCAGCTCGTCTGTAGAGATCCACCT AGTCTGAG	
Gibson Assembly	pGreenII vector	B478	GTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAAC AAG	
Gibson Assembly	pGreenII vector	B569	CCAGATTCAATGTGTCACTACTAGAGATCCACCTAGTCT GAG	

Supplemental References

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

2. Andersen, C.L. (2004). Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. *64*, 5245–5250.

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

4. Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C.S., Bliek, M., Boersma, M.R., Borghi, L., Bruggmann, R., et al. (2016). Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. Nat. Plants *2*, 16074.

5. Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics *26*, 873–881.

6. DePristo, M.A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

7. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

8. Castel, S.E., Levy-Moonshine, A., Mohammadi, P., Banks, E., and Lappalainen, T. (2015). Tools and best practices for allelic expression analysis. bioRxiv, 016097.

9. Mayba, O., Gilbert, H.N., Liu, J., Haverty, P.M., Jhunjhunwala, S., Jiang, Z., Watanabe, C., and Zhang, Z. (2014). MBASED: allele-specific expression detection in cancer tissues and cell lines. Genome Biol. *15*, 405.

10. Van Moerkercke A., Haring M.A. and Schuurink R.C. (2011) 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.