

SI Appendix: Standing genetic variation in a tissue-specific enhancer underlies selfing-syndrome evolution in *Capsella*.

Adrien Sicard^{1,*}, Christian Kappel¹, Young Wha Lee², Natalia Wozniak¹, Cindy Marona¹, John R. Stinchcombe², Stephen I. Wright², Michael Lenhard^{1,*}.

¹ Institut für Biochemie und Biologie, Universität Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam-Golm, Germany

² Department of Ecology & Evolutionary Biology, University of Toronto, 25 Willcocks St., Toronto, ON, Canada M5S 3B2

*Authors for correspondence: michael.lenhard@uni-potsdam.de; tel.: +49-331-9775580
adrien.sicard@uni-potsdam.de; tel.: +49-331-9775583

SI Materials and Methods.

Growth conditions

Plants were grown in a growth chamber under a 16 h day/8 h night photoperiod at a temperature of 22°C during the day and 16°C during the night. The light intensity was set at 150 mmol m⁻² s⁻¹ during the day and the humidity level was maintained at 70 % throughout the cycle.

Molecular cloning and plant transformation

Two *C. rubella* (*SAPr* and *SAPr_2*) and two *C. grandiflora* (*SAPg* and *SAPg_2*) alleles were used for the transgenic experiments. A 10 kb fragment starting 4.5 kb before the start codon of *SAP* and ending 1 kb after the stop codon was amplified from *NILrr* DNA (*SAPr*) or from a *C. rubella* BAC clone containing the *SAP* locus (Bio S&T, Montreal- *SAPr_2*) using the primers oAS286 and oAS295. *SAPg* and *SAPg_2* were respectively amplified from *NILgg* DNA and a *C. grandiflora* BAC clone containing the *SAP* locus (1) using the same primer pair. The *C. grandiflora* *SAP* genomic fragments were 2 kb smaller than *SAPr* and *SAPr_2*. This discrepancy between the sizes of these alleles was due to an insertion of a transposable element in the upstream region of *C. rubella* *SAP*. The PCR-amplified genomic fragments were subcloned into a modified version of pBluescript II KS (StrataGen, pBlueMLAPUCAP) using the In-Fusion® HD Cloning Plus (Clontech), sequenced and transferred into the *AscI* site of the plant transformation vector pBarMAP, a derivative of pGPTVBAR (2). Genomic chimeric constructs as well as reporter constructs were also generated and subcloned into pBlueMLAPUCAP by ligation independent cloning using the In-Fusion® HD Cloning Plus (Clontech) as indicated in **Table S1**. The fragments were then transferred into the *AscI* site of the pBarMAP vector. These genomic constructs were then used to transform *A. thaliana* Col-0 by floral dip (3). The sequences of the primers used are presented in **Table S2**.

Morphological measurements.

To characterise the effect of the QTL and the different genomic constructs as well as to map the mutations underlying PAQTL_6 we measured the size and the cellular properties of different organs.

Leaf size was measured on the fully expanded 4th or 6th leaf. The leaves were flattened out on a white paper before being scanned at 300 dpi. Sepals, petals, stamen were measured from the 15th to 20th fully open flowers on the main stem. Carpels were measured from unfertilised flowers. The different floral

organs were flattened out and scanned at a resolution of 3600 dpi. Area, width and length were measured using ImageJ (<http://rsbweb.nih.gov/ij/>) from the digitalised images of the dissected organs. For each plant, the size of 6 petals, 6 sepals, 6 stamens and 2 carpels from 2 different flowers were measured.

Morphometric analysis of petal outlines was performed using Elliptic Fourier Descriptors (EDF) for closed outlines as described (4). We first converted the digital images into binary images using imageJ (<http://rsbweb.nih.gov/ij/>). Outline coordinates were then extracted using the `bwboundaries` function in Matlab. Outlines were Fourier transformed using adapted functions (5). We used the base of the petals as starting point and did not perform any size normalisation. A principal component analysis using the `R` function `prcomp` was then performed on the EFD coefficients. The principal components (PC) linked to the genotype were identified by a Kruskal-Wallis test with the formula “principal component score ~ genotype”. To illustrate the variation caused by each principal component, we calculated the mean shape by averaging all coefficients for the chosen number of harmonics and added the principal components scores itself multiplied by the corresponding eigenvectors. Outlines were then reconstructed using inverse elliptical Fourier transformations (5). The effect of the PC could therefore be visualised by reconstituting the petal outlines using the maximum and minimum scores.

To determine the size of the petal primordia, we performed an mPS-PI (6) staining on young flower buds of *qILrr* and *qILgg*. Briefly, young inflorescence tissues were fixed in fixative (50% methanol and 10% acetic acid), destained using Ethanol, treated with 1% periodic acid, stained with propidium iodide (100 ug/ml) in Schiff reagent (100 mM metabisulphite, 0.15 N HCL) and cleared in a Chloral-hydrate solution (4 g chloral hydrate, 1 ml glycerol and 2 ml water). The samples were then mounted on microscope slides in Hoyer's solution and imaged with a confocal laser scanning microscope Zeiss LSM710 (Zeiss) using an excitation wavelength of 488 nm, with emission collected between 520 to 720 nm. The section of the z-stack with the longest optical transect through the petal primordia was used to quantify the number of epidermal cells recruited into the primordia. The latter were then plotted against the flower bud diameter. Since *qILrr* and *qILgg* only differ by the size of their petals, the diameter of the developing flower could be used as a proxy of developmental stage.

Kinematic analysis of petal development was performed by measuring two developing petals from each manually dissected bud starting from the oldest unopened flower and extending to the youngest bud from which petals could be dissected. The time interval between the formation of two successive floral buds (the plastochron) was calculated from the number of flowers opening up within 7 days. Average petal area was then plotted against time taking into account the calculated plastochron for each genotype.

Petal cell size and cell number were determined from a dried-gel agarose print (7) of whole petals from fully open flower. Cell outlines were imaged under Differential Interference Contrast (DIC) on an Olympus BX51 microscope using an AxioCam ICc3 camera (Zeiss). For each petal, the cell-outline images were merged using the `photomerge` function of Adobe Photoshop. Resulting images were further processed using the Python module `scikit-image` (8). Images were transformed to grayscale and median filtered. Cell borders were segmented using adaptive thresholding, small objects were removed. Binarized cell borders were then dilated and skeletonized before saving them for further processing. Segmentations were roughly curated by overlaying initial and skeleton images in Adobe Photoshop and/or Gimp. Whole petals were horizontally aligned to the central axis, area and centroid coordinates were extracted. Analyses and illustration were done using R.

Genetic mapping.

To refine the position of the PAQTL_6, we screened about 300 *NILrg* progenies for plants having a recombinant chromosome between the markers G08 and G11 (**Table S3**). The recombinants were then

phenotyped for petal size and genotyped at different positions along the focal region (**Table S3**). These data were then used in a QTL mapping analysis using MAPQTL 6.0 (Kyazma BV, Wageningen, Netherlands). Genome-wide permutations (1000 permutation) were used to determine the LOD-score threshold ($\alpha=0.05$) and a two-LOD support interval was used to determine the position of the QTL within a 95% confidence interval (9).

Based on the above analysis, we screened an additional 3000 *NILrg* progenies for plants having a recombination breakpoint between G09 and G09_20 (**Table S3**). The selected recombinants were selfed and genotyped to identify between 3 and 6 plants homozygous for the *C. grandiflora* allele and 3 to 6 plants homozygous for the *C. rubella* allele in the remaining segregating region. We termed these plants “sister lines”. These plants were then selfed for another generation and the petal size of four replicates per progeny plant was measured as described above. For each petal analysed, the additive effect was calculated by subtracting the recombinant petal size mean from each petal size value. The average additive values of each sister line were then compared between the genotype groups using a Student’s t-test. The position of the recombination breakpoint for each of these recombinants was determined by genotyping the selected recombinants with additional markers in the focal region; these markers are presented in **Table S3**. The location of the recombination breakpoint was further refined by Sanger sequencing of the transitions in the most informative recombinants (NIL_275 and NIL_79). This identified the first polymorphic nucleotide at 14,058,690 bp and 14,061,824 bp, respectively, delimiting this QTL region to a 3,134 bp interval on scaffold_7.

Confocal imaging and analysis of the dual reporter lines.

To analyse the expression pattern of *SAPr* and *SAPg*, we imaged *pSAPr::YFP* and *pSAPg::YFP* lines with a confocal laser scanning microscope Zeiss LSM710 (Zeiss) using an excitation wavelength of 514 nm, with emission collected between 630 to 720 nm for FM4-64 and between 520 - 570 for YFP.

To follow simultaneously the accumulation of *SAPr* and *SAPg* during plant development, we transformed *A. thaliana* Col-0 plants with either *SAPr-CFP*, *SAPg-YFP* or *SAPg-CFP*. T1 transgenic plants were then crossed to obtain “dual-reporter” line expressing either *SAPr-CFP* and *SAPg-YFP* or *SAPg-CFP* and *SAPg-YFP*. CFP and YFP fluorescence was imaged on a confocal laser scanning microscope Zeiss710 (Zeiss) and a 40X water immersion lens (C-Apochromat 40x/1.20 W Korr M27, Zeiss) in stage 2 flower meristems as well as in the oldest petal primordia showing clear *SAPg-YFP* expression. The fluorescence signals were collected on two different tracks. The first one used an excitations of 405 and 561 nm and collected CFP emission between 460-520 nm and FM6-64 signal between 630 and 760 nm. The second used a excitation at 488 nm and collected YFP fluorescence between 510 and 570 nm. For each tissue analysed, the images were taken in a z-stack of 2 μ m with an interval of 0.2 μ m between each focal plane. The microscope settings (laser intensity, pinhole, gain, objectives, etc.) were kept constant between the tissue types. Three F1 plants from 3 independent crosses of each cross type were imaged. A maximum z projection of the 2 μ m sections was used to quantify the CFP and YFP signal in 5 nuclei for each tissue type with ImageJ (<http://rsbweb.nih.gov/ij/>). The CFP/YFP ratio was used to compare the expression of *SAPr* and *SAPg* in flower meristem and petal primordia.

Quantitative reverse-transcription PCR

To quantify *SAP* expression in the inflorescence of *qILrr* and *qILgg*, we extracted total RNA from young inflorescences (containing the inflorescence meristem and the flower buds from stage 1 to 9) using Trizol (Life Technologies). RNAs were treated with Turbo RNase (Ambion) and reverse transcribed with the Superscript III Reverse Transcriptase (Invitrogen). This cDNA were used as template to quantify the relative *SAP* and *AG* mRNA abundance using the SensiMix SYBR Low-ROX kit (Bioline), a

LightCycler® 480 (Roche) and the primers described in **Supplementary Table 2**. For each genotype we used three biological replicates and for each of them three technical replicates.

***In situ* hybridisation.**

In situ hybridisation of *SAP* mRNA during flower development was performed as described (10). The full length open reading frame of *SAPr* was used to synthesize sense and antisense probes. No signal could be detected when hybridising *NILrr* or *NILgg* inflorescences with the sense probes. The sections were photographed with an Olympus BX51 microscope equipped with an AxioCam ICc3 camera (Zeiss).

Allele frequencies and population-genetic analysis

The protein sequences of *SAP* orthologs within Viridiplantae were retrieved from a blast against NCBI (<http://www.ncbi.nlm.nih.gov/>) and phytozome (<http://phytozome.jgi.doe.gov/pz/portal.html>) databases. Protein sequences were aligned using MUSCLE in MEGA5 (11). The relationship between the different proteins was investigated by constructing a Neighbor-Joining phylogenetic tree with MEGA5 using a Poisson model with complete deletion data treatment.

The population genetic analysis was performed on a data set including 180 re-sequenced *C. grandiflora* individuals from a single population (12), a species-wide sample of 13 *C. grandiflora* individuals (13), and a *C. rubella* species-wide sample of 73 individuals. The latter include the sequencing data for 51 *C. rubella* individuals, which were downloaded from the European Nucleotide Archive (<http://www.ebi.ac.uk/ena>, data made publicly available by Daniel Koenig and Detlef Weigel, study number PRJEB6689) as well as *SAP* sequences resequenced on an Ion Torrent platform from 22 *C. rubella* accession (**Table S4**). Note that some of the sequences from the two *C. rubella* data set may be redundant.

Haplotypes were reconstructed combining local assembly and multiple paired-end based phasing approaches. Obtained sequences were corrected manually by visual inspection using IGV (14). For this read mapping was done using bwa mem (15), reads mapping to the region of interest were extracted using samtools (16). Initial phasing was attempted using samtools phase, hierarchical clustering of read pairs based on samtools called variants and co-occurrence graph clustering of the same called variants. Phased reads were then assembled using MIRA version 4 [(17), http://www.chevreux.org/projects_mira.html]. Local variant calling was done using samtools (16). Hierarchical clustering of *Cr1504/Cg926* variants was done based on Euclidean distances. For this *Cr1504* nucleotides were coded as 0, heterozygous ones as 1 and *Cg926* ones as 2.

We conducted a candidate gene association mapping analysis using the *C. grandiflora* population genomics data from Josephs et al (2016) (12). For 180 of these individuals, we measured average petal area from 3-4 flowers. Candidate gene association mapping will increase the power to detect a local association over a genome-wide association mapping due to multiple testing, and we therefore focused on the *SAP* intron only. Association mapping was conducted using plink version 1.07, where we tested all SNPs in the region with a minor allele frequency greater than 10%. Significance was assessed using the Benjamini Hochberg false discovery rate correction (FDR).

Statistical analysis

Statistical analyses were conducted in R or Microsoft Excel 7. Trait values were assumed to be normally distributed. We performed a Tukey's HSD test using the agricolae package add-ons implemented in R software for multi-comparison tests. For two-sample comparisons we used a two-sided Student's t-test assuming unequal variances. Data are presented as mean \pm s.e.m. and p values below 0.05 were

considered statistically significant. Association between petal size and genotype were test using Welch t-test and the Benjamini-Hochberg method was used to correct for multiples testing.

SI Text

Evolutionary history of *CrSAP*.

To test for more recent, post-divergence hybridization between *C. grandiflora* and *C. rubella* re-introducing a derived *SAP* haplotype from the selfing species into *C. grandiflora*, we used two approaches. First, clustering the individuals based on their genotype within the causal *SAP* intronic interval identified two distinct clades. One, named here the *Cg* clade, included the *Cg926* allele, while the *Cr1504* allele fell into a second clade, termed the *Cr* clade (**Fig. S9**). The two clades mostly separate the two species: *C. rubella* clustered in the *Cr* clade (with the exception of one *Cr* individual whose haplotype at this locus and other loci differed from all other *Cr* suggesting that the sample might have been misannotated and may belong to another species) and most of the *C. grandiflora* individuals clustered in the *Cg* clade. Six *C. grandiflora* individuals fell into the *Cr* clade, however, suggesting that the *SAPr* allele could have been reintgressed in these individuals through post-divergence hybridisation between the two species. Second, to estimate the length of *Cr* like haplotypes, we counted the occurrence of *Cr* (reference) k-mers of increasing length in all *Cg* and *Cr* individuals (the *Cr* outlier individual was filtered out of this analysis - **Fig S9A**). Re-introgression events would in this case be detected through an increase in the prevalence of long *Cr* k-mers. In all *Cr* individuals k-mers up to 80 bp were largely abundant (>80% of the total read numbers), whereas the abundance of the 50 bp *Cr*-like K-mer corresponded to less than 60% in most of the *Cg* individuals. Only 6 *Cg* individuals show an increase in haplotype length suggestive of a potential recent re-introgression of a *SAPr* haplotype. These individuals correspond to the *Cg* samples that fell into the *Cr* clade in **Fig S9B**. To avoid confounding more recently re-introgressed variants with standing variation, we removed these six individuals from the data set we used to calculate *Cr1504* alleles frequencies. Even after removing all samples with potential introgression, the *Cr1504* haplotypes were shown to occur in this *C. grandiflora* population at least once in all polymorphic sites.

SI References

1. Sicard A, et al. (2014) Repeated Evolutionary Changes of Leaf Morphology Caused by Mutations to a Homeobox Gene. *Curr Biol* 24(16):1880–1886.
2. Becker D, et al. (1992) the left T-DNA border. 1195–1197.
3. Clough SJ, Bent a F (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16(6):735–43.
4. Kuhl FP, Giardina CR (1982) Elliptic Fourier features of a closed contour. *Comput Graph image Process* 18:236 – 258.
5. Claude J (2010) Morphometrics with R. *Biometrics* 66, pp 656–664.
6. Truernit E, et al. (2008) High-Resolution Whole-Mount Imaging of Three-Dimensional Tissue Organization and Gene Expression Enables the Study of Phloem Development and Structure in *Arabidopsis*. *Plant Cell Online* 20(6):1494–1503.
7. Horiguchi G, Fujikura U, Ferjani A, Ishikawa N, Tsukaya H (2006) Large-scale histological analysis of leaf mutants using two simple leaf observation methods: Identification of novel genetic pathways governing the size and shape of leaves. *Plant J* 48(4):638–644.
8. van der Walt S, et al. (2014) scikit-image: image processing in Python. *PeerJ* 2:e453.
9. van Ooijen JW (1992) Accuracy of mapping quantitative trait loci in autogamous species. *Theor Appl Genet* 84(7-8):803–11.
10. Wahl V, et al. (2013) Regulation of flowering by trehalose-6-phosphate signaling in *Arabidopsis thaliana*. *Science* 339(6120):704–7.
11. Tamura K, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–9.
12. Josephs EB, Lee YW, Stinchcombe JR, Wright SI (2015) Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proc Natl Acad Sci USA* 112(50):15390–15395.
13. Agren JA, et al. (2014) Mating system shifts and transposable element evolution in the plant genus *Capsella*. *BMC Genomics* 15(1):602.
14. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24–26.
15. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Prepr arXiv* 00(00):3.
16. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–9.
17. Chevreur, B., Wetter, T. and Suhai S (1999) Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. *Comput Sci Biol Proc Ger Conf Bioinforma* 99:45–56.
18. Sicard A, et al. (2011) Genetics, evolution, and adaptive significance of the selfing syndrome in the genus *Capsella*. *Plant Cell* 23(9):3156–71.

SI Figures

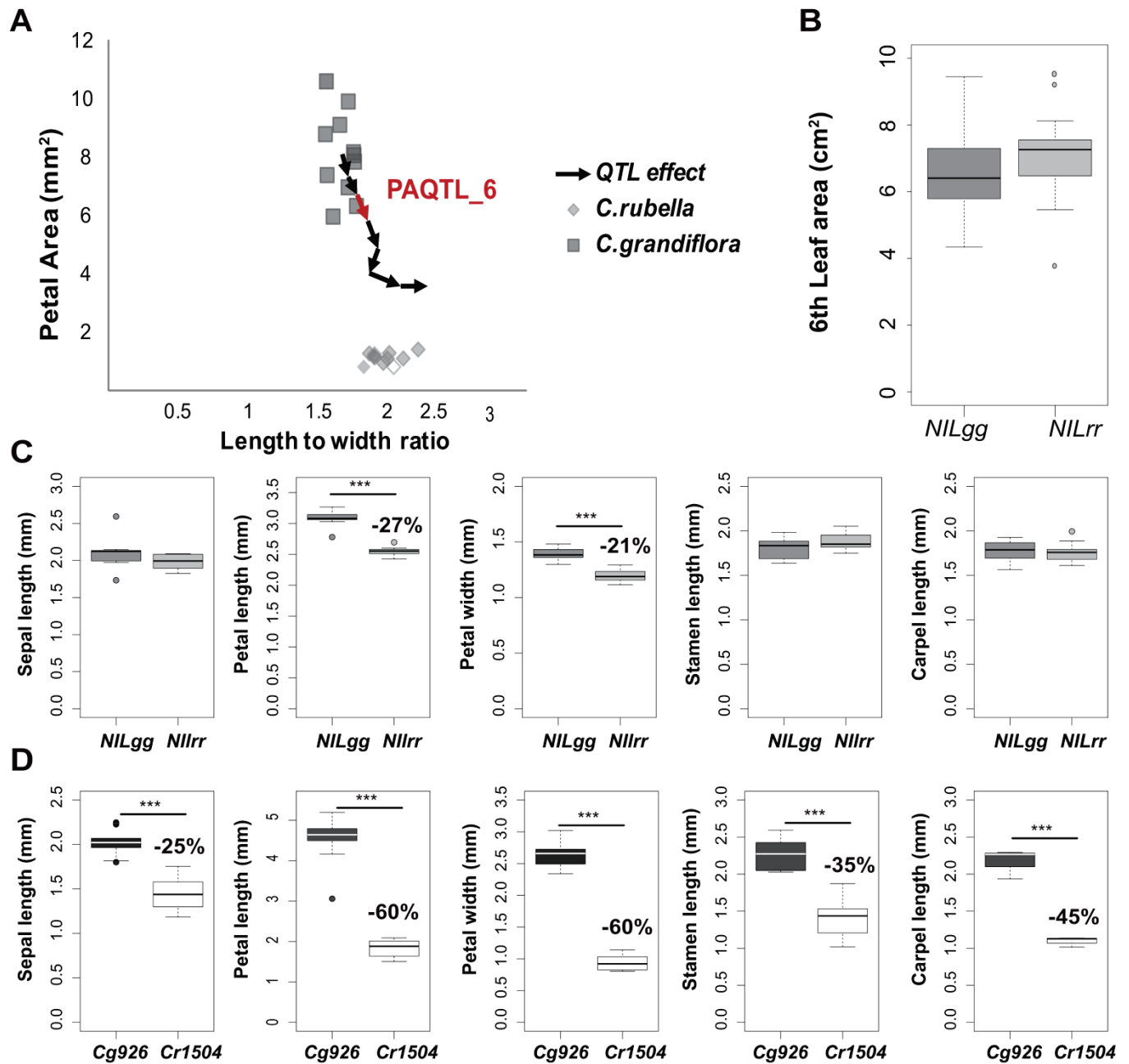


Fig. S1: Allelic variation at PAQTL_6 has contributed to reducing the size of *C. rubella* petals after its transition to selfing.

(A) QTL analysis of petal morphology in the *Cg926* x *Cr1504* RILs. The average petal sizes of different *C. grandiflora* and *C. rubella* accessions are plotted against their average petal width-to-length ratios. A plausible evolutionary path is shown as a succession of arrows, where each arrow represents the predicted effect of one QTL on petal size and width/length ratio, starting from the average value of the *C. grandiflora* accessions. The effect of each QTL was predicted from the percentage of species variation they explained as determined in (18). Note that the order of the QTL is arbitrary. PAQTL_6, highlighted in red, has the strongest influence on petal size.

(B) Box-plot illustrating the distribution of the 6th leaf area from 18 individuals in *NILgg* and *NILrr*. In all box-plots, the middle lines represent the median, the upper and lower boxes represent the upper and lower quartile respectively, while the upper and lower whiskers add or subtract 1.5 times the Interquartiles ranges to/from the 75 and 25 percentile respectively. Possible outliers are displayed as a circle.

(C,D) Comparison of flower organ sizes between *NILgg* and *NILrr* (C) or between *Cg926* and *Cr1504* (D). (n= 18 (C) and 5 (D) individuals per genotype). Only the size of the petals differs between *NILgg* and *NILrr*, whereas all flower organs are smaller in *C. rubella* compared to *C. grandiflora*. Asterisks: Significantly different at * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ as determined with a Student's t-test.

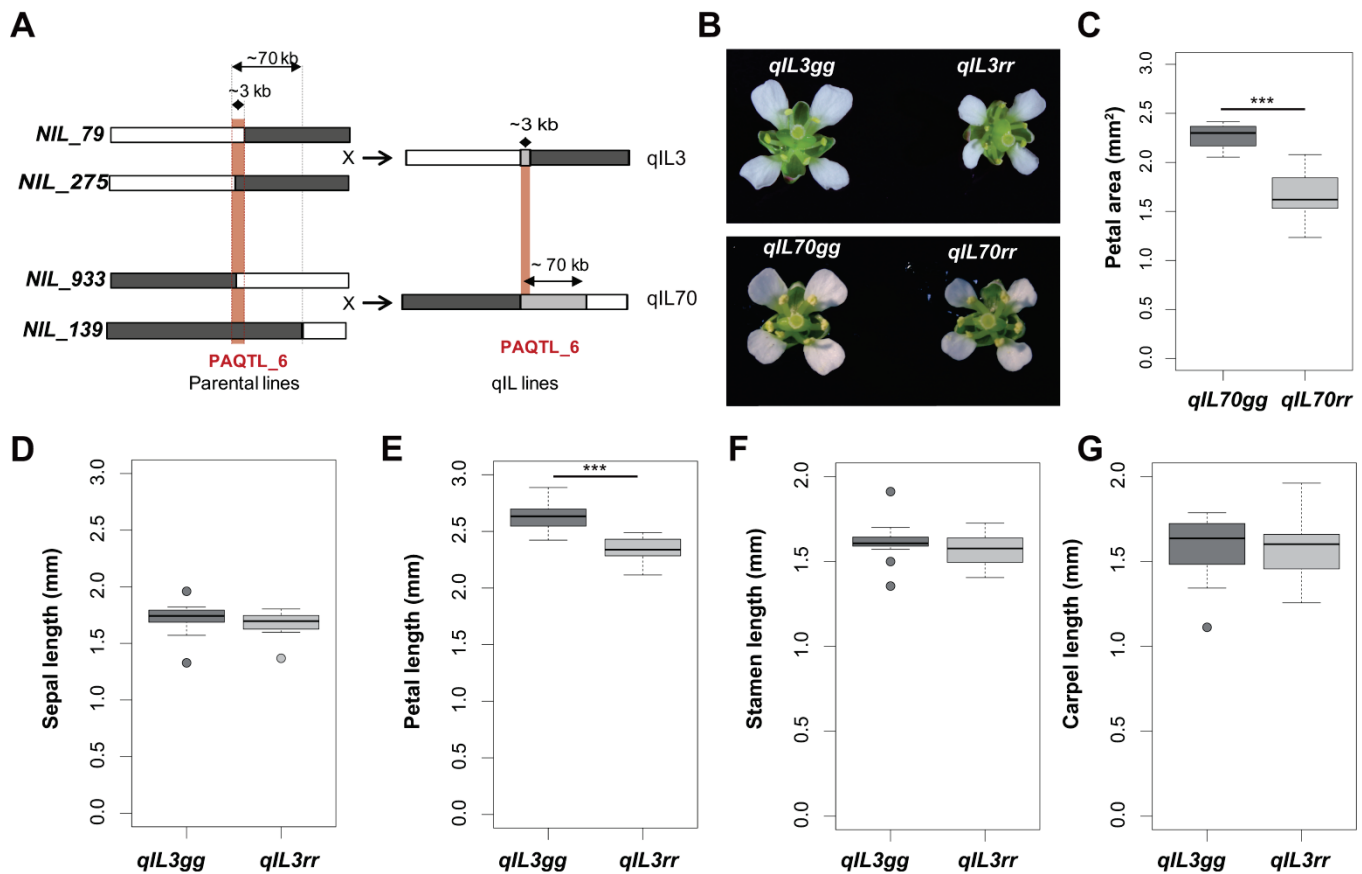


Fig. S2: Quasi-isogenic lines localize the causal region to the *SAP* intron.

(A) Scheme of the crosses used to generate the quasi-isogenic lines segregating for a 3.1 kb (qIL3) or 70 kb (qIL70) interval around PAQTL_6. The diagrams illustrate the genotype of the plants along scaffold_7 around the PA_QTL6 region. The dark grey color indicates regions homozygous for *C. rubella*; chromosomal fragments homozygous for the *C. grandiflora* allele are shown in white and heterozygous regions are shown in light grey. PAQTL_6 position is indicated in red. Note the different widths of the red boxes indicating the QTL interval are due to different scales.

(B) Photographs of the flowers of *qIL3rr*, *qIL3gg*, *qIL70rr* and *qIL70gg* genotypes. Swapping the genotype from homozygous *C. grandiflora* to homozygous *C. rubella* decreases the petal size to a similar extent as in the NILs, confirming that the polymorphisms underlying PAQTL_6 are comprised with the 3.1 kb region.

(C) Petal area of *qIL70gg* is significantly larger than that of *qIL70rr*. Distribution of petal area means from 7 individuals for each genotype are shown.

(D-G), Effect of the segregating 3.1 kb interval on the length of flower organs in the *qIL3*. Distributions of organ dimension means from 13 individuals for each genotype are shown. Only petal size is affected.

Asterisks: Significantly different at * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ as determined with a Student's t-test.

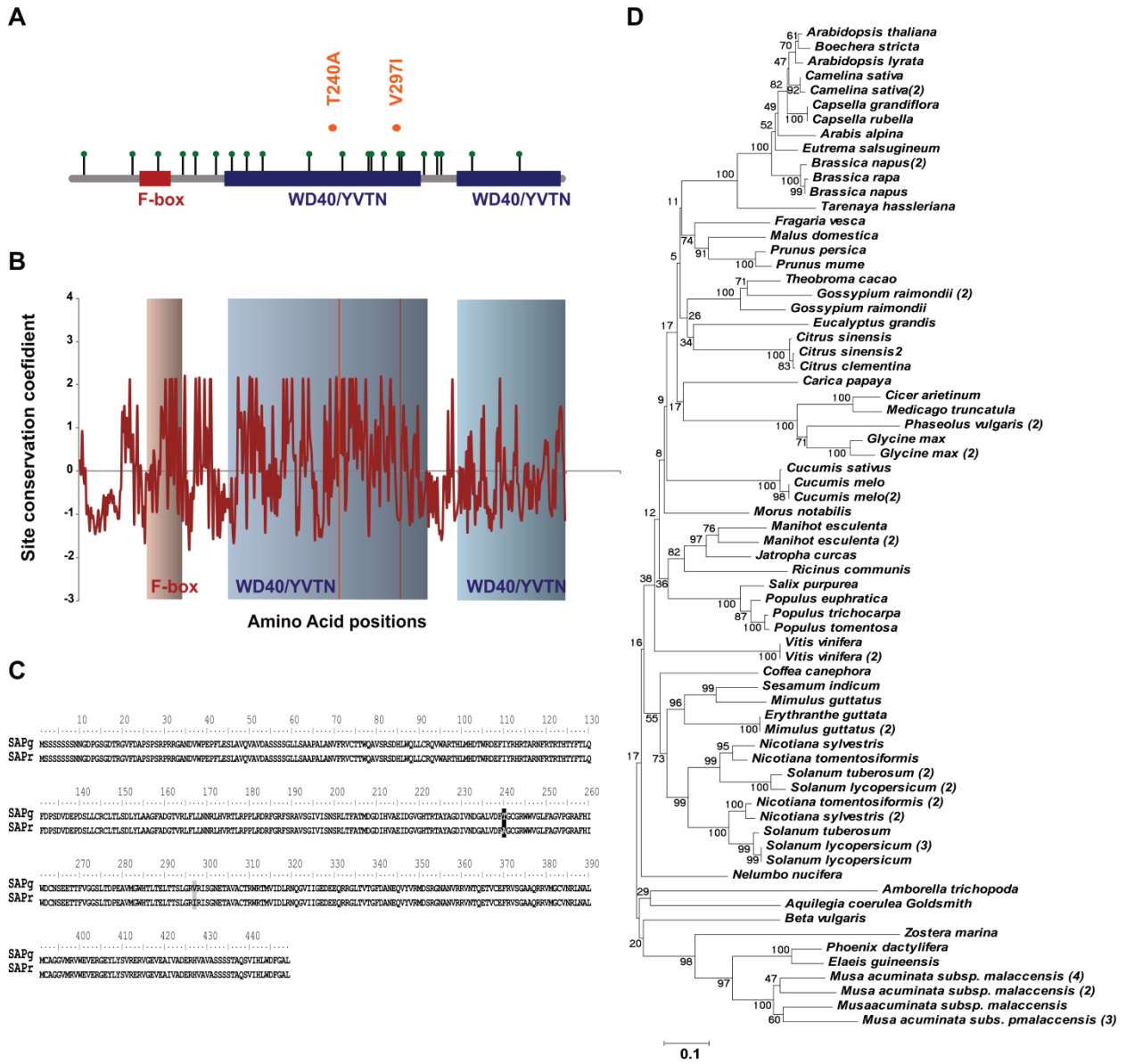


Fig. S3: SAP protein structure, motif homologies and phylogeny.

(A) Protein representation of SAP. The positions of the motifs are indicated in red for the F-box domain and in blue for the WD40/YVTN domain. The positions of the *C. grandiflora*/*C. rubella* polymorphisms are indicated in orange. The positions of putative posttranslational modifications are indicated by green dots.

(B) SAP sequence conservation within the Eudicot clade. The positions of the amino-acid exchanges due non-synonymous SNPs between the two *Capsella* species are indicated by orange lines. None of these polymorphisms are located in highly conserved positions.

(C) Alignment of *Cg926* and *Cr1504* SAP protein sequence.

(D) Neighbor joining tree of SAP protein homologs within Viridiplantae. Numbers indicate bootstrap support from 1,000 runs.

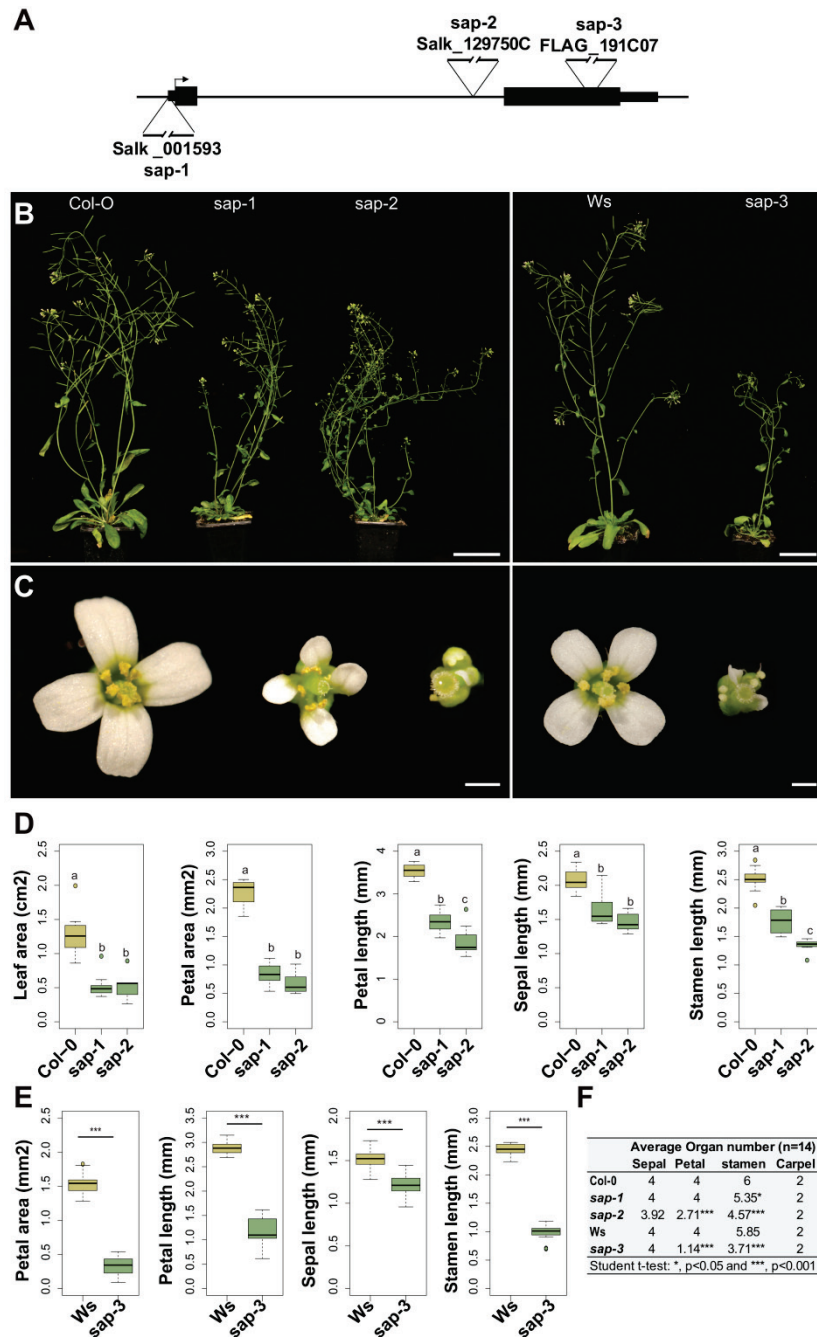


Fig. S4: SAP functions as a general growth regulator.

(A) Schematic representation of T-DNA insertions in the *A. thaliana* *SAP* locus. *sap-1* and *sap-2* have been generated in a Col-0 background, while *sap-3* has been produced in a Ws background.

(B,C) Whole-plant (B) and flower (C) phenotypes of *sap* T-DNA insertion lines and their respective wild-type backgrounds. Scale bars are 4 cm (B) and 1 mm (C).

(D,E) All measured organs are smaller in *sap-1* and *sap-2* than in Col-0 (D) and in *sap-3* compared to Ws-0 (E). Distributions of organ dimension means from 6 to 16 (D) and 11 plants per genotype (E) are shown.

(F) *SAP* regulates floral-organ initiation. Average floral-organ numbers in *sap-1*, *sap-2*, *sap-3*, Col-0 and Ws. Letters indicate significant differences as determined by Tukey's HSD test. Asterisks: Significantly different at * p<0.05, ** p<0.01, and *** p<0.001 as determined with a Student's t-test.

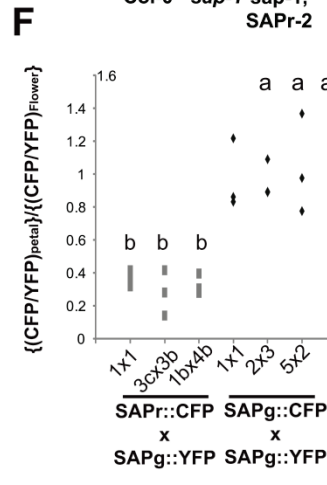
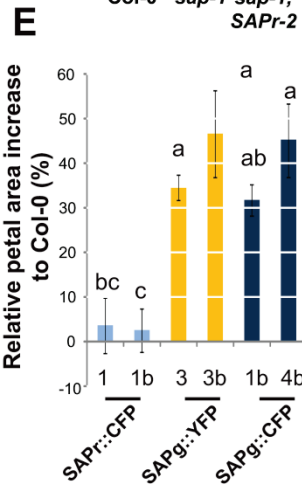
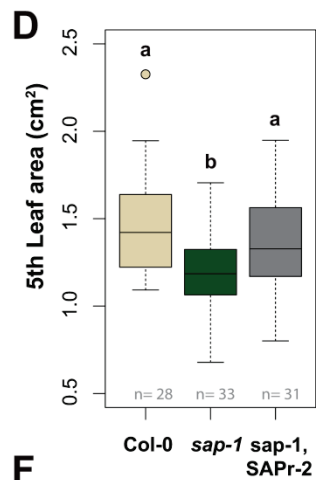
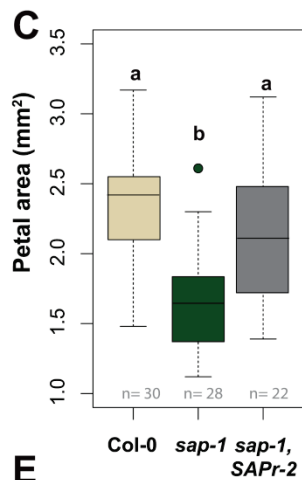
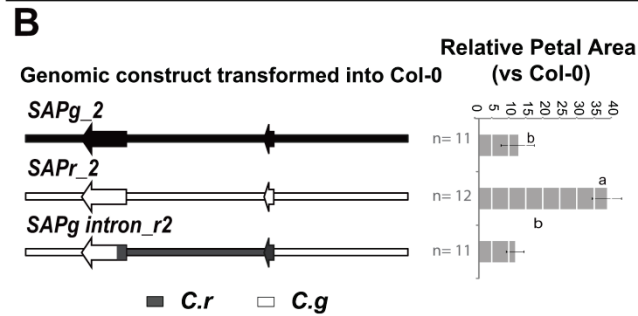


Fig. S5: *A. thaliana* transformants reproduce the PAQTL_6 effect.

(A) Transforming *SAPr2* and *SAPg2* in *A. thaliana* Col-0 recapitulates the effect of PAQTL_6. Flower phenotype of Col-0, Col-0 transformed with *SAPr2* or with *SAPg2*.

(B) Schematic representation of the genomic constructs used to transform *A. thaliana* Col-0 are shown in the left panel, while their effects on petal area is quantified in the right panel. Values are mean \pm SEM. The number (n) of independently transformed lines used to calculate the average value is indicated on the figure.

(C, D) Complementation of *sap-1* phenotypes with the *SAPr2* allele. Distribution of organ dimension means from n individuals per genotype. The number (n) of individuals analyzed is indicated on the figure.

(E) *SAPr-CFP*, *SAPg-YFP* and *SAPg-CFP* are functional and recapitulate the PAQTL_6 effect on *A. thaliana* petal size. Values are mean \pm SEM from 5 T2 individuals per line.

(F) Average YFP/CFP ratio in *SAPr::CFP* x *SAPg::YFP* and *SAPr::CFP* x *SAPgYFP* F1 plants. Each dot represents the average of 5 nuclei from one plant. Letters indicate significant differences as determined by Tukey's HSD test.

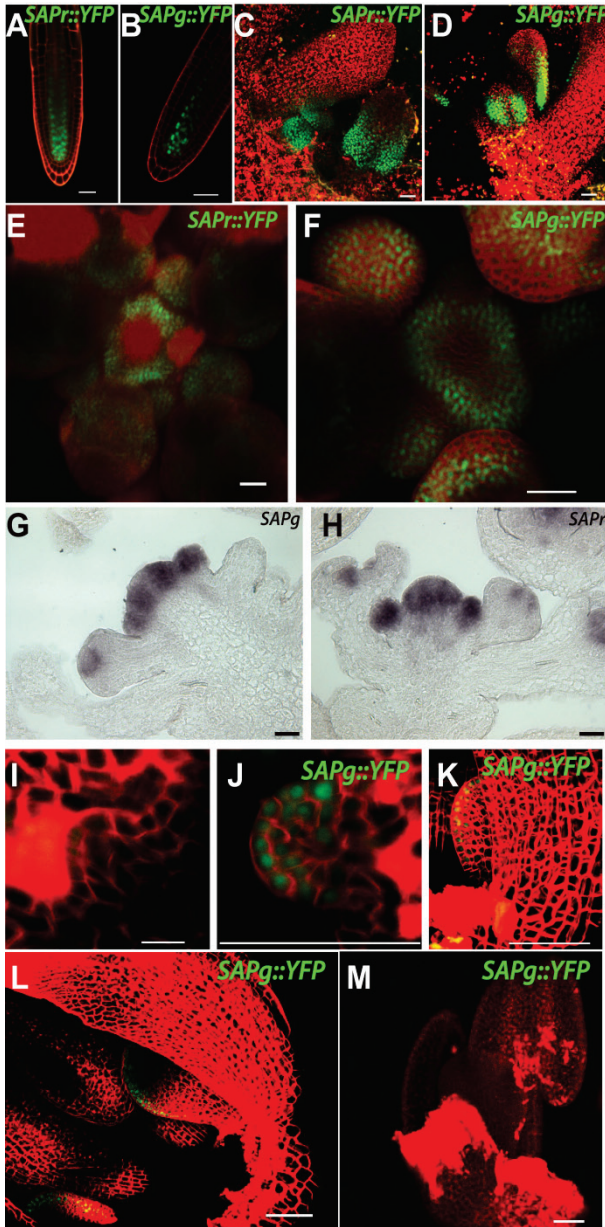


Fig. S6: Expression pattern of *SAPr* and *SAPg* during plant development.

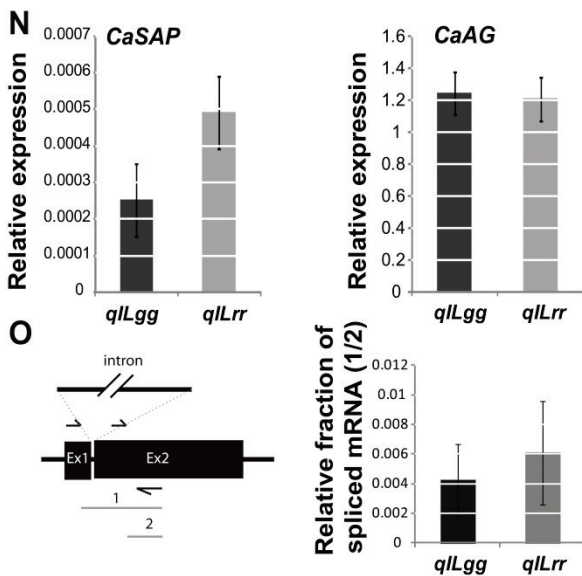
(A-F) Expression of *SAPr::YFP* and *SAPg::YFP* could be detected in Col-0 roots (A and B), young leaves (C and D) and inflorescence meristems (E and F). Scale bar represents 30 μ m.

(G and H) *In situ* hybridisation with *SAP* antisense probe in *Capsella* inflorescences of *NILrr* and *NILgg* plants reveals a similar pattern of expression to the fluorescent reporters. Scale bar represents 30 μ m.

(I-M) *SAPg::YFP* is expressed in the distal region at early stages of petal development. Scale bar represents 10 μ m (I-K) and 50 μ m for (L and M).

(N) Relative expression levels of *Capsella SAP* and *Capsella AGAMOUS (CaAG)*, as determined by quantitative RT-PCR (qRT-PCR), normalized to the constitutively expressed gene *Capsella TUB6*. Values are mean \pm SEM from 3 biological replicates.

(O) Relative fraction of spliced *SAP* mRNA as determined by quantitative RT-PCR (qRT-PCR) is identical in *qILgg* and *qILrr*. The position of the primer pairs used and of the corresponding PCR product is shown in the scheme on the left corner. The values are the mean of the ratio between relative abundance of PCR product 1 to the relative abundance of PCR product 2, both normalised to *Capsella TUB6*, \pm SEM from 3 biological replicates.



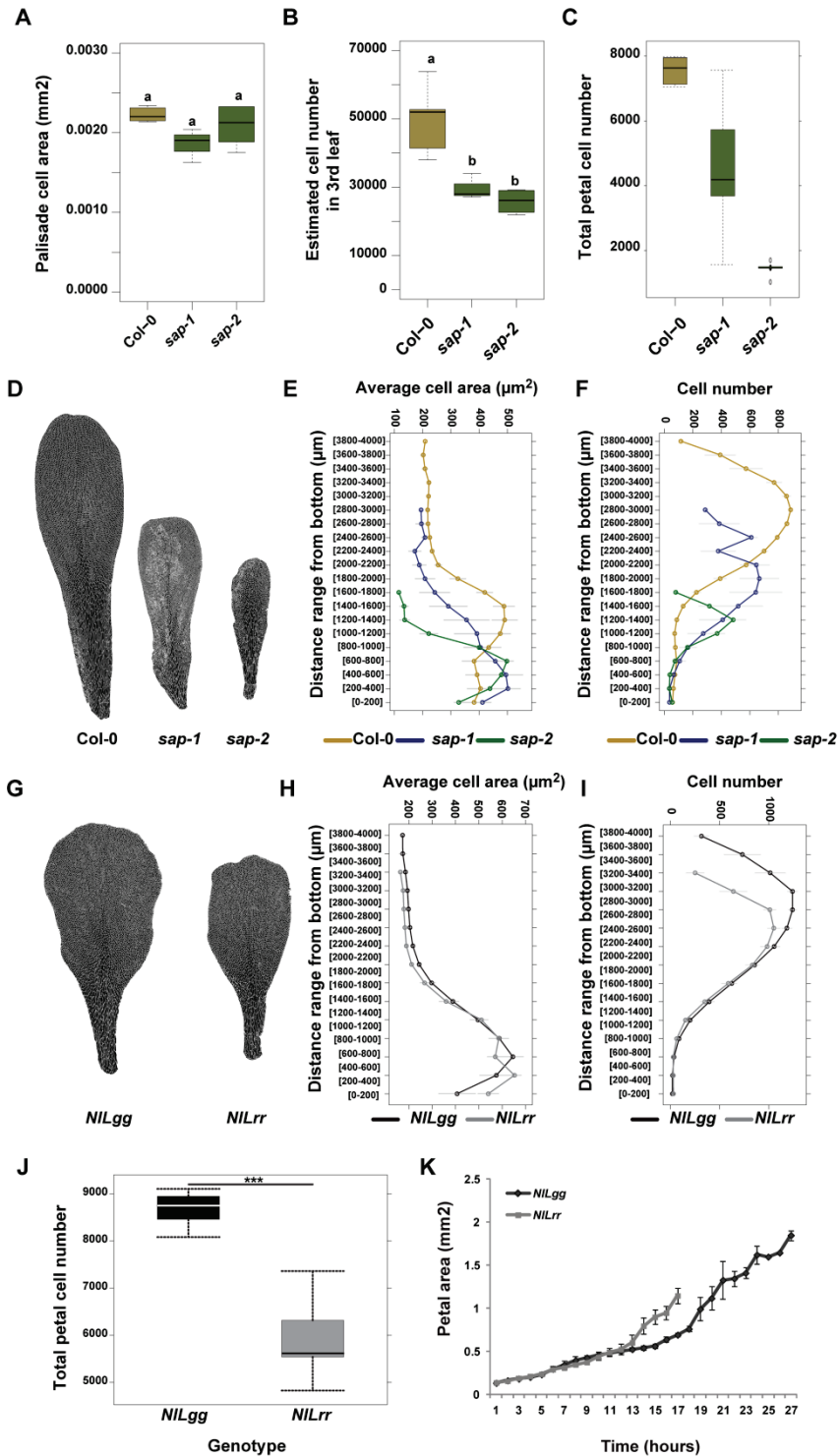


Fig. S7: *SAP* regulates organ size by controlling cell division.

(A) Palisade cell area in the fully developed 3rd leaf does not differ between Col-0, *sap-1* and *sap-2*. Letters indicate significant differences as determined by Tukey's HSD test.

(B) Estimated total cell number in the third leaf is reduced in *sap-1* and *sap-2*. Letters indicate significant differences as determined by Tukey's HSD test

(C) Total petal cell number is reduced in *sap-1* and *sap-2* when compared to Col-0 (n=6).

(D) The distal region of the petals is strongly affected by *sap-1* and *sap-2* mutations.

(E) The maximum average cell size in segments along the longitudinal petals axes do not differ between Col-0, *sap-1* and *sap-2*. However, the minimal cell size is lower in *sap-2* and the general cell size pattern is shifted towards the proximal side of the petal. Values are mean ± SEM (n=6)

- (F) The distribution of cell numbers in Col-0, *sap-1* and *sap-2* petals in sections along their longitudinal axes indicates that cell proliferation is strongly diminished in *sap* mutants, especially in the distal region.
- (G) Segmented petals of *NILgg* and *NILrr*.
- (H) No difference in average cell size per segments along the longitudinal axes could be observed between *NILgg* and *NILrr* petals. Values are mean \pm SEM (n=6)
- (I) Distribution of cell numbers in *NILgg* and *NILrr* petals along their longitudinal axes. Note that the profile is identical for the two genotypes in the proximal part of the petals. However *NILgg* reaches higher values in the distal area.
- (J) Total petal cell number is lower in *NILrr* petals compared to *NILgg* (n=6). Asterisks: Significantly different at * p<0.05, ** p<0.01, and *** p<0.001 as determined with a Student's t-test.
- (K) The petal growth rate is identical in *NILgg* and *NILrr* plants (n=4).

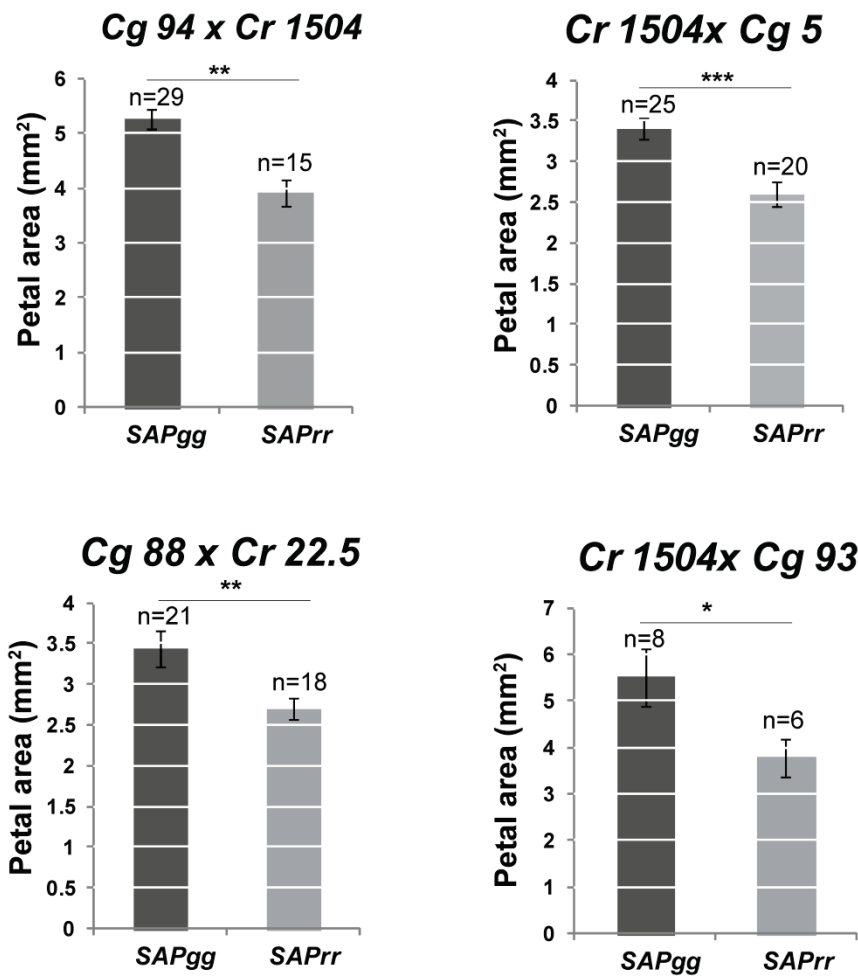


Fig. S8: *SAP* co-segregates with petal size in other *C. grandiflora* x *C. rubella* crosses.

Petal size segregation analysis in F2 populations of different *C. grandiflora* x *C. rubella* crosses genotyped for the *SAP* locus. In all of the crosses, plants homozygous for the *C. grandiflora* allele (*SAPgg*) develop larger petals than plants homozygous for the *C. rubella* allele (*SAPrr*). Values are mean \pm SEM (the number of individuals used to calculate the mean values is indicating on the figures). Asterisks: Significantly different at * p<0.05, ** p<0.01, and *** p<0.001 as determined by Student's t-test.

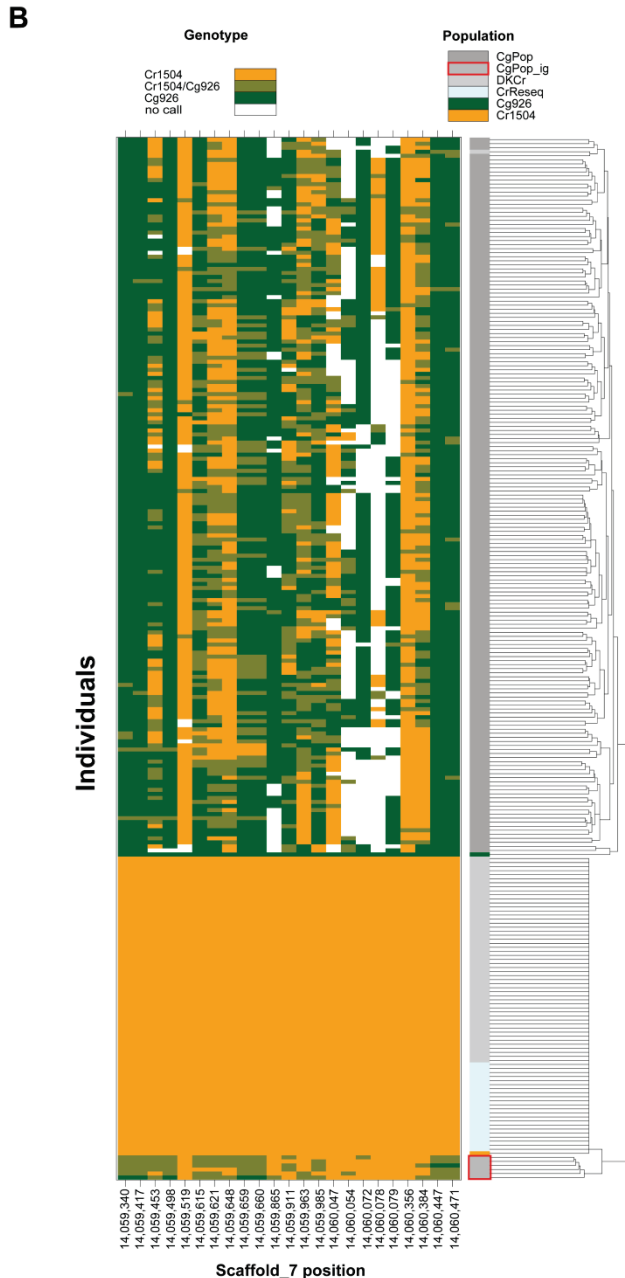
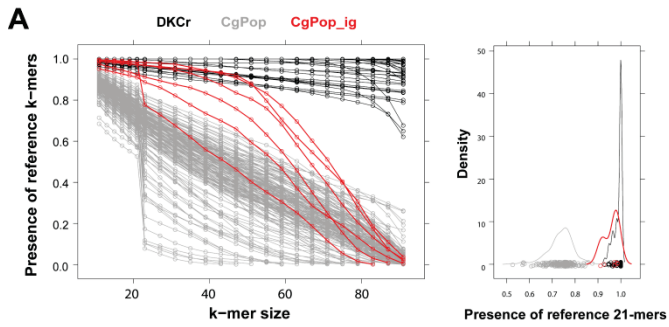
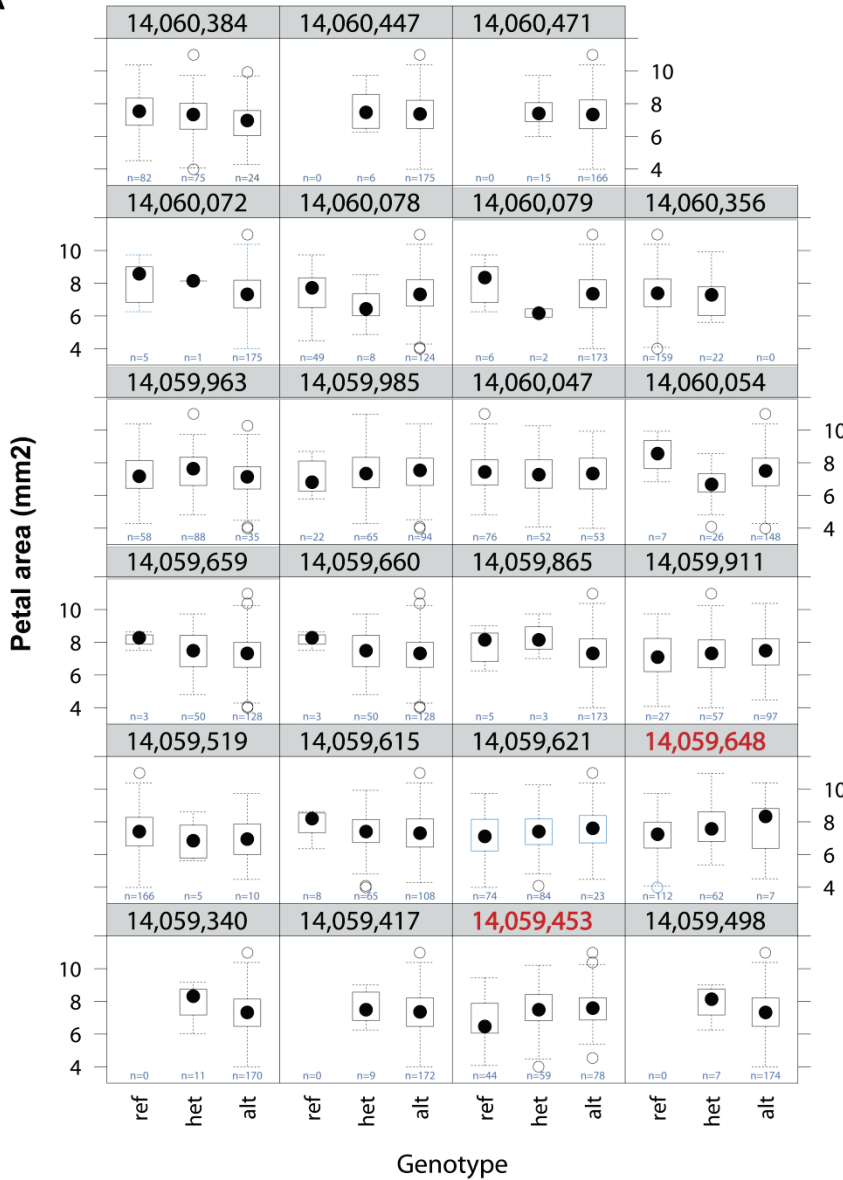
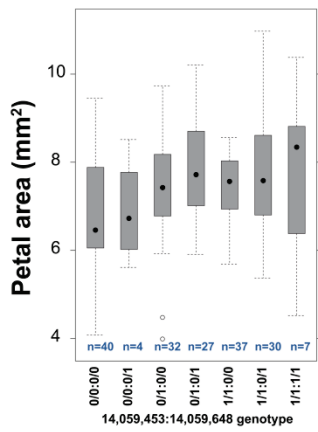


Fig. S9: All *Cr1504*-like alleles within the PAQTL_6 causal region are segregating in *C. grandiflora*. (A) Putative recent *C. rubella* *SAP* introgression in the *C. grandiflora* population was estimated by counting the presence of *C. rubella* reference like k-mers of increasing length in genome re-sequencing raw data. Each line represents an individual. Those from the *C. grandiflora* population are colored in gray (CgPop), putatively introgressed ones in red (CgPop_ig) and publicly available *C. rubella* accessions in black (DKCr). Six individuals with putative introgression were identified based on a high presence of 21-mers as shown in the distribution on the right. Higher k-mer values were not considered as they have a strong coverage dependence. Additional haplotype investigations for selected samples using paired end data further supports this approach (B) Haplotype clustering and genotype at the candidate polymorphisms of all the resequenced individuals. The red square indicates individuals with exceptionally long *C. rubella* like haplotype blocks suggestive of more recent introgression identified in (A). This data set includes sequences from all *C. grandiflora* resequenced individuals (CgPop), from the parental *SAP*gg (*Cg926*) and *SAP*rr (*Cr1504*) alleles, from targeted re-sequencing of the *SAP* intron in 22 *C. rubella* accessions (CrReseq) as well as from all publicly available *C. rubella* genomes (DKCr). The *C. grandiflora* individuals with possible introgression events (CgPop_ig) are indicated by a red square on the figure. Note that this analysis failed to identify polymorphisms private to *C. rubella* in this interval even when individuals with putative recent *C. rubella* *SAP* introgression are not considered.

A



B



C

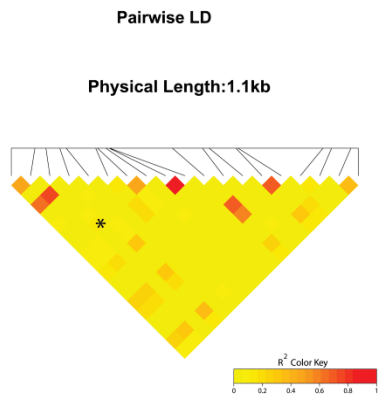


Fig. S10: Single marker analysis of the effect of the candidate polymorphisms on petal size.
(A) Distribution of organ dimension means from n individuals depending on the presence of *Cr reference* (or *Cr1504*) allele at candidate polymorphic sites. The number (n) of individuals analysed is indicated on the figure. The two polymorphisms having a significant effect (as determined by Welch t-test, p-values were Benjamini-Hochberg corrected and considered significant if adjusted p-value <0.05) on petal size are highlighted in red font.
(B) The distribution of the petal dimension means depending on the genotype at the SNPs 14059453 and 14059648 suggests that these polymorphisms act in an additive manner. The number (n) of individuals analysed is indicated on the figure. 0/0 indicates homozygosity for the *Cr* allele, 1/1 indicates homozygosity for the *Cg* allele, while 0/1 are heterozygotes.
(C) Pairwise linkage disequilibrium (LD) is low among 21 of the candidates SNPs. The asterisk marks the LD between the SNPs 14059453 and 14059648. Note that SNPs 14060054 and 14060078 were not included due to a high number of missing genotype values.

SI Tables

Table S1: Summary of the Ligation Independent Cloning reactions used to generate chimeric *SAP* genomic fragments and *SAP* reporter constructs.

Constructs	DNA fragment 1			DNA fragment 2			DNA fragment 3		
	Corresponding genomic fragment	Primer used	DNA Template	Corresponding genomic fragment	Primer used	DNA Template	Corresponding genomic fragment	Primer used	DNA Template
<i>SAPg intron_r</i>	Promoter	oAS286- oAS289	<i>SAPg</i>	1st Exon; intron and 0.3kb of Exon2	oAS290- oAS293	<i>SAPr</i>	1kb Exon2 and 3'UTR	oAS294- oAS295	<i>SAPg</i>
<i>SAPg intron_r2</i>	Promoter	oAS286- oAS289	<i>SAPg_2</i>	1st Exon; intron and 0.3kb of Exon2	oAS290- oAS293	<i>SAPr_2</i>	1kb Exon2 and 3'UTR	oAS294- oAS295	<i>SAPg_2</i>
<i>SAPg intron_rb</i>	Promoter; 1st Exon and 1.3 kb of its intron	oAS286- oAS323	<i>SAPg</i>	2.1 kb intron and 0.3kb of Exon2	oAS326- oAS293	<i>SAPr</i>	1kb Exon2 and 3'UTR	oAS294- oAS295	<i>SAPg</i>
<i>SAPg intron_rc</i>	Promoter; 1st Exon and 2.5 kb of its intron	oAS286- oAS324	<i>SAPg</i>	0.7 kb intron and 0.3kb of Exon2	oAS327- oAS293	<i>SAPr</i>	1kb Exon2 and 3'UTR	oAS294- oAS295	<i>SAPg</i>
<i>pSAPg::2xYFP-NLS</i>	SAPg promoter intron and 0.4kb EXon2	oAS286- oAS328	<i>SAPg_2</i>	<i>2xYFP-NL-nosTerm</i>	oAS329- oAS1118	<i>pAS004</i>			
<i>pSAPr::2xYFP-NLS</i>	SAPr promoter intron and 0.4kb EXon2	oAS286- oAS328	<i>SAPr_2</i>	<i>2xYFP-NL-nosTerm</i>	oAS330- oAS1118	<i>pAS004</i>			
<i>SAPrCFP</i>	SAPr promoter ; Exon1; intron and Exon2	oAS286- oAS1216	<i>SAPr</i>	CFP	oAS1214- oAS1222	<i>pAS0198</i>	SAPr 3'UTR	oAS1220- oAS295	
<i>SAPgCFP</i>	SAPg promoter ; Exon1; intron and Exon2	oAS286- oAS1216	<i>SAPg</i>	CFP	oAS1214- oAS1222	<i>pAS0198</i>	SAPg 3'UTR	oAS1220- oAS295	
<i>SAPrYFP</i>	SAPg promoter ; Exon1; intron and Exon2	oAS286- oAS1216	<i>SAPg</i>	YFP	oAS1214- oAS1222	<i>pAS0135</i>	SAPg 3'UTR	oAS1220- oAS295	

Table S2: Primers used in this study.

Name	Usage	Sequence (5'-3')
oAS286	molecular cloning SAP	CTACCATGGTGAATTCAAACAAAAACCTAGCCCCTTGC
oAS289	molecular cloning SAP	GAAGGAACAGGGGAAGAAGAAGAAG
oAS290	molecular cloning SAP	TTCCCCTGTTTCCTTCCTTCTCAT
oAS293	molecular cloning SAP	CGTGGAGACGGTTGTTCAAGAG
oAS294	molecular cloning SAP	ACAACCGTCTCCACGTAAGGAC
oAS295	molecular cloning SAP	ATGCCTGCAGGTCGACAAATGTTATGCCAACATCATTTCATTA
oAS323	molecular cloning SAP	GACTAACATTGGCAAGGCAAAA
oAS324	molecular cloning SAP	TGATTTCTCTATACATCTCTCTCCCA
oAS326	molecular cloning SAP	TTGCCAATGTTAGTCTTATATTACCAC
oAS327	molecular cloning SAP	TGTATAGAGAAATCAAATGTACTTACTTC
oAS328	molecular cloning SAP	GAAAGCTTGCATGCCGGGTCACTGGATTTTGGTT
oAS329	molecular cloning SAP reporter construct	TTATATCCAACCTCGAGAATGGCTCCCAAGAAGAA
oAS330	molecular cloning SAP reporter construct	TTATATCAAACCTCGAGAATGGCTCCCAAGAAGAA
oAS1118	molecular cloning SAP reporter construct	ATGCCTGCAGGTCGACGGGTCACTGGATTTTGGTT
oAS1214	molecular cloning SAP reporter construct	TCCTGCTCCTGCTCCCCACGGGTTAGGCGGCT
oAS1216	molecular cloning SAP reporter construct	TCCTGCTCCTGCTCCCAGTGCACCGAAATCCCAAAGA
oAS1220	molecular cloning SAP reporter construct	ATTCTTCTCGCCGACACTAGAC
oAS1222	molecular cloning SAP reporter construct	TTAAAGCTCATCATGTAACCTTGTACAGCTC
oAS420	qPCR <i>CaAG</i>	AGCAGTTTGGTCTTGGCG
oAS421	qPCR <i>CaAG</i>	AAACTATTTCCAAGTCGCCG
oAS407	qPCR <i>CaSAP</i>	TGAAAGACATGTGGCGGTTG
oAS408	qPCR <i>CaSAP</i>	AGTGCACCGAAATCCCAAAG
oAS1086	qPCR <i>caTUB6</i>	TTCGACCAGCTGATGAACTG
oAS1087	qPCR <i>caTUB6</i>	CCTTCACCAAAGGTGTCAGAC
oAS416	qPCR caSAP-Exon1	TGTCTTCCTCCTCCTCCTTCCA
oAS419	qPCR caSAP-Exon2	ATTTACCGTCATCGGACGGCTAGA
oAS418	qPCR caSAP-Exon2	GCCGGACAGTTCCGTCGGCG
oAS297	genotyping sap t-DNA insertion	ATGTCACCATACTGAACCAATACTTTTCCTAA
oAS298	genotyping sap t-DNA insertion	TTAACGGCAACTTGAACGGCGAG
oAS299	genotyping sap t-DNA insertion	GCGTGGACCGCTTGCTGCAACT
oAS1205	genotyping sap t-DNA insertion	CATAGATGGTCGGAGCGAGAGA
oAS1204	genotyping sap t-DNA insertion	GTAGGTCCACGGATTGTTGGAG
oHB246	genotyping sap t-DNA insertion	CGTGTGCCAGGTGCCACGGAATAGT
oAS357	genotyping sap t-DNA insertion	ACCGTCGTTTACTATATCTCCAGCG

oAS140	genotyping sap t-DNA insertion	ATCTGGTACGCTACAGTGCACCGAAAT
--------	--------------------------------	-----------------------------

Table S3: Molecular markers.

Name	Primer name	Position;	Polymorphism Cr/Cg	Sequence (5'-3')	Polymorphism revealed with
G08	G08F	10149356	A/G	TCGCTAGATCCTTAACTGTGTTCTCC	Nde I
	G08 R			AGAAACAACCACAGGAGGAAGAGTTAG	
G09	G09 F	13427057	T/G	gagatgaggatgacaatatcg	Taq I
	G09 R			CTAAGCCGTGGTGAGTGTAGGCTA	
G09_1	oAS371	13829358	G/C	ATTAGTCTCGACTGCTGATCTGAACTCGA	HaeIII
	oAS372			GATAGTATATGCAACCAAGGTTGTTGTCTGAA	
G09_2	oAS345	13869338	A/ATATATTTGTATAA	GCGCTCACAAATACTTCCACA	Indel
	oAS346			TCCCGAAAGGTCAAAAATCAAGA	
G09_3	oAS341	13878216	CTCCCCCAAAAATATTATAAAC/C	AAGCAGCTCCTTGTCTCCCTTA	Indel
	oAS342			GCCATGACAGATTTACACAGTC	
G09_4	oAS335	13881521	A/AGTAGTTGCAACAGTTAAAACCTAA AAGGAGTCGCCAAA	AGTTTAAAGTGTGCTGAAGTCG	Indel
	oAS336			GCGCCAACGACACAAAAAGATA	
G09_5	oAS353	13886380	T/G	CCAGTCTTCTGATGTGTTTATAAGAAATTGT	XmnI
	oAS354			CCTCAAATCCTCCTTCTCTCATGC	
G09_6	oAS351	13959305	T/C	TAATTCATTGAATTTGGTTAATTCCTGA	XmnI
	oAS352			ACACCTCATGATGTTTAGCATCACCCAC	
G09_7	oAS361	13972319	G/A	CAAGAAGAGGCAACCAAACAGAAAGC	Hinf I
	oAS362			CATTTCCAATGCAGCCTTGCCACT	
G09_8	oAS339	14000868	A/ACATACGTAACATGATGCATAA	TGGGTCATGTGATAGAGAGGGT	Indel
	oAS340			TAAAAGCGGATTGTTGGCTCG	
G09_9	oAS343	14044317	T/TGTAGACACATTTGTTATTTTATAC ACGTATAAACTGTATTTTTTTG	CCCTGCGATGTAGTGTATCTGT	Indel
	oAS344			CGTGTTAAAATTAGACAGGGTTAGAGG	
G09_10	oAS355	14053975	C/A	AAATTCCTCCTAAGAAACTGGTCAA	HindII
	oAS356			CGTCTTGCATTCGACGTCGTC	
G09_11	oAS357	14058109	A/G	TCACGAGCCGTCAGGCATTGTT	EcoRV
	oAS358			ACCGTCGTTACTATATCTCCAGCG	
G09_12	oAS359	14061421	T/A	GCGGTTACCTTTCAAAGGTCAGTTTA	NlaIII
	oAS360			TCGTGGGTCTATATAACTTAAAGTGAGGG	
G09_13	oAS305	14065975	aaaaaaataaaaaaat_2.5kb_tatgattgatctaagacaa/A	ATAATTAATAAATATAACAACCCGTGCA	oAS305-oAS303: Cr allele
	oAS303			CTATATCAAAACCAAGTGCAGAGTTTGG	oAS304-oAS303: Cg allele
	oAS304			TTTGTTGAATGATATATGATTAAGTCCTTTCATT	
G09_13	oAS347	14066555	T/C	AGACGGCATGTTACATGTAATGA	Hinfl
	oAS348			AAACAAAAACCTAGCCCCCTTGC	
G09_14	oAS349	14072064	T/C	TCAACACAAAAGTTTTAGATTTTTCATCT	HindII
	oAS350			TGGCTCATTGAGTTGGTCTTT	
G09_15	oAS363	14080879	A/G	GCAGCATGGGATCTATTCACACC	DdeI

	oAS364			TATCTACAATTTGAGTCATGTATGTATAA GTTGG	
G09_16	oAS391	14133754	C/T	CAGACCTGGGCTTTTGAAAATT	HinfI
	oAS392			CATGTAGGTGGAGGTGTAGGTG	
G09_17	oAS369	14097819	A/C	CTCTAGCGTAGTCTTCTGAAACCATCATC CTTA	DdeI
	oAS370			CTATGTAACATGTGGTGTATGGAGATGC	
G09_18	oAS365	14170392	G/A	TTTACCGTAAATTCGTTATAATAAGT G GA TC	BamHI
	oAS366			TTGCTAATTAACATATAGGTGTGGATG	
G09_19	oAS367	14341953	T/A	GAGCGGACAGCATCGCAGATAG	HindIII
	oAS368			TAGTCGAACGATCTAGCTCGCTTG	
G09_20	oAS213	14283848	G/A	CGCACTCCAATATAACTATGTGTTGTTA	HindII
	oAS214			GAGACTTGACCTCCTCTTTCTG	
G11	G11F	16968217	G/A	GGAGTGGAGTGGCCCTCC	TaqI
	G11R			CTAAACGAGAACTAGAACCAGAACCCAA	

Table S4: List of the *C. rubella* accessions used for the targeted re-sequencing of *SAP*.

Sample number	Accession name	Geographical origin
1	Cr1Gr1	Samos, Greece
2	TAAL	Taguement, Algeria
4	39-5	Bacia, Italy
5	78-3	Greece
7	84-14	Greece
8	79.12	Greece
9	23-9	Senerchia, Italy
10	82.14	Nemea, Greece
11	72.12	Parinthus, Greece
12	1504.2	La Palma Spain
14	34.11	Italy
15	77-21	Greece
16	76-3	Greece
17	75-2	Greece
18	1215	Tenerife
19	100-7	Greece
21	6-26	France
23	86IT1-C	Italy
24	CrGo665	unknown
25	1377	Buenos Aires
26	83.1	Greece
27	103	Greece