Supplementary Materials for

Dominance hierarchy arising from the evolution of a complex small RNA regulatory network

Eléonore Durand[†], Raphaël Méheust[†], Marion Soucaze, Pauline M. Goubet[&], Sophie Gallina, Céline Poux, Isabelle Fobis-Loisy, Eline Guillon, Thierry Gaude, Alexis Sarazin, Martin Figeac, Elisa Prat, William Marande, Hélène Bergès, Xavier Vekemans, Sylvain Billiard, Vincent Castric[*].

[*]Correspondence to: Vincent.Castric@univ-lille1.fr
[†]These authors contributed equally to this work.
[&]Deceased

This PDF file includes:

Materials and Methods

Plant material and collection of S-genotypes

We used *Arabidopsis halleri* seeds collected in four French and Italian populations to constitute a set of individuals carrying 6 S-alleles selected to span the sequence diversity of the S-locus (*17*) including class I (Ah01), class II (Ah03), class III (Ah04) and class IV (Ah12, Ah13 and Ah20) alleles (*15*). We grew the seeds in the greenhouse and extracted DNA from leaves using the Nucleospin Multi-96 Plant extraction kit (Macherey-Nagel Gmbh, Duren, Germany). S-alleles were identified using PCR primers specific for the *SRK* alleles (*16*). Controlled pollinations were performed as described in ref. (*16)*, by manually depositing pollen from one individual chosen as the male parent on the pistil of the chosen female partner, within one day of flower opening. Plants were separated by at least 60 cm to avoid pollen contamination. The offspring were then genotyped for the 6 S-alleles using the same PCR protocol.

Controlled crosses and inference of the phenotypic dominance network

We determined the dominance relationships between each pair of S-alleles in pollen, say Sx and Sy*,* by using pollen from a heterozygous individual Sx/Sy to pollinate pistils of tester lines expressing either the Sx or Sy incompatibility type. When Sy is dominant, pollen from Sx/Sy should be compatible with the Sx tester line, but incompatible with the Sy tester line, whereas rejection by both tester lines indicates that Sx and Sy are codominant. Following ref. (*16*), compatibility was scored using pistil elongation 7 days after pollination. The tester lines were checked for proper rejection of pollen expressing their corresponding *SCR* allele. As negative controls (to assess fruit elongation for incompatible crosses), we self-pollinated each tester line. As positive controls, we crossed each tester line with an individual sharing no S-alleles to estimate fruit elongation values for compatible crosses. Based on these measures, we defined pollinations as successful if the fruit was at least 0.55 cm long 7 days after pollination. Using 4 to 15 replicates for each cross (Figure 2A), we defined a cross as compatible when more than 50% of the pollination replicates were successful; compatible crosses were clearly distinguishable from incompatible crosses (Fig. S12). Combining dominance phenotypes obtained in our experiment with those from ref. (*16*) allowed us to define the dominance phenotype for all 15 possible heterozygous combinations of the six S-alleles studied here.

BAC sequencing

Two BAC clones containing Ah04 and Ah12 were newly obtained, using the protocol of ref. (*9*). Briefly, high molecular weight DNA was prepared from young leaves of two *A. halleri* individuals carrying either Ah04 or Ah12 and used to construct two separate BAC libraries. Libraries were screened based on the two flanking genes, and positive clones were sequenced using a 454 multiplexing technology on Titanium sequencer (www.roche.com). *De-novo* assembly was performed by Newbler (www.roche.com) and only contigs representing the extremities of the BACs were organized. We were not able

to construct a BAC library for Ah01, and thus used a BAC containing the orthologous sequence in the closely related *A. lyrata* (Al01) (*9,17*).

Small RNA isolation, library preparation and sequencing

Total RNAs were extracted by TRI reagent (Sigma, USA) from 10-20 floral buds at various stages of development five to one days before flower opening. RNA quality and integrity were assessed with Nanodrop and Agilent Bioanalyzer. Samples were then sequenced on either a SOLiD4 or a PROTON platform. For SOLID4, RNAs were processed following SOLiD Total RNA-Seq Kit Part Number 4452437 Rev A 01/2010, starting with 1µg of total RNA without any enrichment. Adaptors were added by hybridization and ligation following the protocol, followed by reverse transcription and purification on minElute column (minElute PCR Purification Kit from Qiagen, USA). cDNAs were then run on an urea gel (10%), and the 60-80nt population (with adapters) was selected. cDNAs were amplified by 18 PCR cycles with barcoded-primers and the product was purified on purelink micro columns. The size was then controlled on the Bioanalyzer on HS DNA assay. The resulting cDNAs were then diluted and pooled with equal amounts from each library. The ePCR was done manually following the SOLiD 4 template bead preparation protocol (number 4448378 Rev B; Applied Biosystems USA) followed by terminal transferase. Each library was sequenced on a SOLiD 4 system following the Life Technologies instructions for single-end reads (50bp forward reads). Raw signals from the SOliD4 system were analyzed to obtain raw demultiplexed reads using bioscope 1.3.rBS131 from Life Technologies (USA). The protocol for PROTON sequencing was identical, except for the following steps: 1) RNA were processed using the ION total RNA-seq kit v2 MAN4476286 Rev E. 2) purification used a magnetic bead cleanup module (Life Technologies, USA). 3) cDNA were size selected with the Magnetic Cleanup Module. 4) cDNAs were amplified, and barcodes added, by PCR (Platinum PCR supermix high fidelity, 16 cycles). 5) a second round of SizeSelect with Magnetic Cleanup Module was done after Amplification. 6) The ePCR was done with a One Touch 2 (Life Technologies, USA) with OT2 200 kit v2. 7) Each library was sequenced on a PI v2 chip on a PROTON system following the Life Technologies instructions for Ion P1 sequencing chemistry 200 kit v2 with 250 cycles. Raw signals were analyzed to obtain raw demultiplexed reads (Fastq) by Torrent Server version TS3.6 from Life Technologies (USA).

Mapping criteria and identification of sRNA precursor genes

Adaptor sequences in SOLiD reads were trimmed using a python script from ABI LifeTechnologies (USA). Each sRNA sample was mapped against both S-alleles carried by an individual and the *Arabidopsis lyrata* genome (Araly1) using the *bowtie* program (*39*) version 0.12.7. For further analysis, we selected sRNAs with perfect genomic matches, that mapped to less than 10 sites in the genome and were specific to the S-locus region. Since the sRNA identified in Brassica is formed from a precursor with a hairpin structure (*12*), we searched for inverted repeats in the sequence. Hairpin structures were predicted using the *einverted* program in the EMBOSS package (*40*) version 6.2.0 with a window size of 350bp and the following scoring matrix: matches = +4; mismatches = -4;

gaps = -8 and a threshold of 50. The set of hairpins and sRNA alignments was then screened by using a set of criteria for the annotation of miRNA genes (*20*). Specifically, we retained hairpins in which >80% of at least 5 unique reads mapped with >99% strand bias and a maximal terminal loop size of 40bp. We then visually inspected the predicted RNA folding structure in *rnafold* (*41*). Additional precursor motifs within the S-locus region of all available S-alleles were then searched for using a similarity search based on the *YASS* program (*42*) version 1.14 starting from the 17 initially identified sRNA motifs plus *Smi* (*12*) and using an *e*-value threshold of 10^{-4}. Based on this first set of hits, we then iterated the procedure to identify further motifs that might have remained undetected because of their divergence.

Clustering of sRNA precursors into families

Hairpin precursors were then clustered into families using a similarity network. Each sRNA precursor sequence was used as query in a homology search against all other sRNA precursors using the *YASS* program (*42*) version 1.14. The results were used to construct an undirected graph in *Cytoscape* (*43*) version 3.0.1 in which each node corresponds to a sequence and two nodes were linked if they have sequence identity above a given threshold defined by: hits with an *e*-value < 10^{-4}, at least 60% sequence identity and covering at least 75% of the length of both the query and subject sequences. Each connected component (or connected subgraph) was considered a separate sRNA precursor gene family. Each sRNA gene family was then annotated by comparison with the miRBase database (*21*) version 19.

Target site predictions

Small RNA targets were predicted in *SCR* alleles including 1kb of flanking sequence ("*SCR*+/-1kb") using the Smith and Waterman algorithm (*28*), with the following scoring matrix: matches = +1; mismatches = -1; gaps = -2; G:U wobbles = -0.5. The full regulatory network was represented with Circos (*44*). To determine whether *SCR* +/- 1kb sequences are significantly enriched in target sites, they were randomized 1,000 times using the *shuffleseq* program (*40*) (EMBOSS package, version 6.2.0) and the average number of target sites per randomized sequence across the 1,000 replicates was compared to the actual number of targets observed in real sequences. To determine whether "silent" sRNA precursors (*i.e.* sequences belonging to the sRNA gene families based on sequence similarity but for which no sRNA production was detected) had retained intact targeting potential, we uniformly processed these potential precursors *in silico* into all possible 21nt factors, and predicted these sRNA targets in *SCR* alleles. To keep the comparison on a fair basis, the same *in silico* processing procedure was also applied to the "expressed" precursors (*i.e.* those for which sRNA production was detected); for these, their actual pattern of processing was ignored. Because the Al01 BAC clone was from *A. lyrata*, we sequenced the *A. halleri SCR* sequence +/- 1kb, and used this sequence for predicting the targets.

Construction of transgenic lines

We generated four series of transgenic plants: (1) *A. thaliana* C24 transformed with the female SI determinant (*AhSRK01* gene), (2) *A. thaliana* C24 transformed with the male SI determinant (*AhSCR01* gene), (3) *A. thaliana* C24 transformed with a mutated version of the *AhSCR01* gene (*AhSCR01*)*, and (4) *A. thaliana* C24 transformed with the *Ah20mirS3* gene. C24 *Arabidopsis thaliana* plants were grown in soil under long-day conditions at 21°C and 70% humidity. We used Gateway® vectors (Life Technologies, USA) for expression of transgenes in *Arabidopsis thaliana*. The DNA fragment containing the *Brassica oleracea SLR1* promoter (*pSLR1*, 1.5kb upstream of *SLR1* start codon)(*45*) was amplified using attB4-containing primer 5'-CCCCACAACTTTGTATAGAAAAGTTGTAGCTCTAGAACTAGTGGATCCC-3' and attB1-containing primer 5'-CCCCAC-TGCTTTTTTGTACAAACTTGTCTCTCTTCACCACTTTAATTTTC-3' and subsequently inserted by BP recombination into a pENTR-P4-P1R plasmid. *AhSRK01* genomic sequence was amplified from genomic DNA of an *Arabidopsis halleri* individual containing the *S01*-locus with specific primers that span the start or stop codon (AttB1 primer: 5'-ACAAGTTTGTACAAAAAAGCAGGCTATGAGAGGTGTAA-GAAGTATCTACC-3' and AttB2 primer: 5'-ACCACTTTGTACAAGAAA-GCTGGGTTTACCGAGGCTCAATGTCCGAAAAG-3'). 1,948 kb upstream and 803 bp downstream the *AhSCR01* allele were amplified from genomic DNA of the *Arabidopsis halleri* individual containing the *S01*-locus using the AttB1 primer: 5'-ACAAGTTTGTACAAAAAAGCAGGCTGTACGACGATGAGTAACAACTAC-3' and the AttB2 primer: 5'-ACCACTTTGTACAAGAAAGCTGGGTCCATT-GGGTGCCCTACAACACCTTC-3'. 1,936 kb upstream and 2,015 kb downstream the *Ah20mirS3* were amplified from a BAC clone containing the *Arabidopsis halleri S20*-locus (*9*) using the AttB1 primer: 5'-ACAAGTTTGTACAAAAAAGCAGGCTGAAC-CTCAACGTAAGATTCTACC-3' and AttB2 primer: 5'-ACCACTTTGTACAA-GAAAGCTGGGTGAGGAACAACTATACATTGTATG-3'.

AhSRK01, *AhSCR01* and *Ah20mirS3* were inserted by BP recombination into the pDONR-Zeo plasmid. The *SLR1* promoter, the genomic sequence *AhSRK01* and a 3'mock sequence were inserted in the pK7m34GW destination vectors by three fragment LR recombination. The *AhSCR01* and *Ah20mirS3* were respectively recombined in the pB7m34GW and the PH7m34GW plasmid with 5'mock and 3'mock sequences. PCR-mediated mutagenesis was realized according to the protocol described in ref. (*46*). To modify the putative *AhSCR01 Ah20mirS3* target site, we amplified the parental plasmid *AhSCR01*/pZeo with two complementary primers containing 4 mutations (underlined):

5'-CAAGATATTATAATATCAACATCCCTTTGGATCTTATTGGTTACTTT-GAAAACC-3' and 5'-GGTTTTCAAAGTAACCAATAAGATCCAAAGGGATGTT-GATATTATAATATCTTG-3'. The mutated *AhSCR01* (*AhSCR01**) was recombined in the pB7m34GW plasmid with 5'mock and 3'mock sequences. All DNA amplifications were performed with the PrimeSTAR® DNA polymerase (Takara, Japan). Arabidopsis C24 transgenic plants were generated using *Agrobacterium tumefaciens*-mediated transformation according to ref. (*47*).

Selection of *A. thaliana* transgenic lines

Eighteen independent plants containing the *AhSRK01* transgene were isolated on kanamycin containing medium. From segregation ratio of antibiotic resistance (Kanamycin) in their progeny, we determined that six lines had only one insertion site. For each single insertion line, T2 seeds were germinated on Kanamycin-containing medium and nine T2 resistant plants were transferred in soil. At T2 flowering, we carried out a test for pollen rejection using *A. halleri S*01 as pollen donor. Stage 13 stigmas, according to ref. (*48*), were pollinated with mature pollen and left for 6 h followed by fixing and staining with Aniline Blue. Stigmas were observed by epifluorescence microscopy and germinated pollen grains were manually counted. We considered that pollination is incompatible when less than 5 pollen tubes were able to overcome the stigmatic barrier (*49*). Among the six lines with a unique insertion, five were highly incompatible and one exhibited on average 15 pollen tubes per stigma (five replicates *per* tested line). We chose one *AhSRK01* line showing less than five pollen tubes *per* stigma and generated homozygous (T3) plants, which we considered as the female reference. We followed the procedure described above to identify five single-insertion *AhSCR01* lines and six *AhSCR01** lines, carrying Basta resistance. T2 individuals were used as pollen donors to test the SI reaction on stage 13 stigmas from the *AhSRK01* reference line. Pollen from all unique-insertion lines of *AhSCR01* and *AhSCR01** was strongly rejected on *AhSRK01* stigmas, with a single exception (one *AhSCR01** line). We chose one *AhSCR01* and one *AhSCR01** line showing a strong SI response as male references. We then selected seven single-insertion *Ah20mirS3* lines on Hygromycin-containing medium. Lines expressing *Ah20mirS3* cannot be identified using a pollination assay (as for *AhSRK01* and *AhSCR01* plants), so we decided to detect the presence of *Ah20mirS3* transcripts in T2 transgenic buds. We performed RT-PCR experiments using primers within the predicted stem-loop precursor (using primers 5'-GTTTTAGATTTTGCAAGTAACCG-3' and 5'-CGGGTAACCAATCAAAGC-3') on total RNA extracted from stage 12 developing buds. A PCR amplification was detected for four of the seven T2 lines. Homozygous plants (T3) were generated for two *Ah20mirS3* lines (named 6 and 12) exhibiting a PCR amplification.

Crossing design in *A. thaliana* transgenic lines

Hybrid plants containing both *AhSCR01* and *Ah20mirS3* (or both mutated *AhSCR01** and *Ah20mirS3*) were generated by manual crossing of *AhSCR01* T2 (or *AhSCR01**) plants with *Ah20mirS3* T3. We selected hybrid offspring that contain both *Ah20mirS3* and *AhSCR01* (or *AhSCR01**) by germination on a combined selection medium (Basta + Hygromycin). By construction, these plants are hemizygous for both transgenes. We also produced hemizygous plant for *AhSCR01* (or *AhSCR01**) by crossing them with C24 to compare their compatibility phenotypes with *AhSCR01 Ah20mirS3* and *AhSCR01* Ah20mirS3* hybrids. For all hybrid plants obtained from T2 parents, the presence of the *AhSCR01* (or *AhSCR01**) transgene was further verified by PCR on genomic DNA (using primers 5'-GTGTGTCTGTCCATAACTTAC-3' and 5'-CCCAAAA-TACTTAGCTCCATG-3'). We tested the SI reaction by following a crossing design (described in Table S1) in which we counted the number of germinated pollen grains as described above. When the number of germinated pollen grains was superior to 30, we reported a value equal to 30, indicating a compatible cross. For each cross, a minimum of

seven stigmas were pollinated on at least two different dates to ensure reproducibility of the results. We tested the effect of the *Ah20mirS3* on *AhSCR01* phenotype with a Mood's median test with threshold α=5%. Statistical analyses were performed with R (*50*).

S-alleles phylogeny based on *SRK* sequences.

Twenty *SRK* (exon 1) amino acid (AA) sequences from Arabidopsis species were aligned with *MAFFT* (*51*) version 7 using the default strategy. An *A. lyrata* paralog (Aly10) was added as outgroup. On the 471 AA alignment, *Gblocks* (*52*) version 0.91b was applied to remove poorly aligned regions using the options for low stringency. This resulted in a 418 AA dataset. The phylogenetic reconstruction was performed in a Bayesian framework under the site-heterogeneous CAT-GTR + Γ4 mixture model with *PhyloBayes* (*53*) version 3.3d. Two independent MCMC chains starting from a random tree were run for 30,000 cycles, sampling trees and associated model parameters every 10 cycles. At the end of the runs the largest discrepancy observed across all bipartitions of the two independent runs was smaller than 0.1. In addition the discrepancies and the effective sizes estimated for each parameters were < 0.1 and > 100 respectively. The initial 1,500 trees sampled in each MCMC run were discarded as the burn-in period. The 50% majority-rule Bayesian consensus tree and the associated posterior probability (PP) were then computed from the remaining 3,000 trees combined from the two independent runs. Phylogenetic classes were as defined by Prigoda *et al*. (*15*).

Molecular model *vs.* phenotypic observations.

To evaluate the congruence between the molecular model based on sRNA target predictions and the empirical observations of dominance, we first computed the proportion of phenotypically observed dominance relationships for which at least one sRNA produced by the dominant allele of the pair was predicted to target the *SCR* sequence from the recessive allele (a proxy for the "power" of the molecular model). We then determined how reliable the molecular predictions were by computing the proportion of molecular predictions involving a small RNA produced by a recessive allele but predicted to target a more dominant allele (a proxy for the "rate of false positives"). To test the significance of the latter figure, we explicitly took into account the fact that the non-recombining region of dominant S-alleles has larger average physical size than that of recessive alleles (*9*), and might therefore produce a larger total number of sRNAs than recessive S-alleles, so they would thus be expected to target recessive alleles more frequently by chance alone. 100,000 random networks were thus obtained by randomly rewiring targets across the set of *SCR* alleles, while keeping the set of sRNAs produced by each S-allele as a constant. To evaluate the potential contribution of siRNAs, the same procedure was applied to small RNAs produced by the S-locus but not from sRNA precursors. The robustness of the results to the threshold chosen to define putative targets was assessed by varying the threshold over the range 16-20. The proportion of predicted sRNA-target interactions that are consistent with the observed phenotype increases with the level of stringency of the target prediction procedure (Fig. S6), suggesting that the rate of false positives at high stringency is probably low. To test whether dominance is correlated with the repertoire of sRNAs and their targets, we took

explicit account of the S-allele phylogeny, and used a continuous random-walk MCMC procedure in *BayesTraits* (*54,55*) (Version 2, available at www.evolution.rdg.ac.uk) to compare the likelihoods of nested models in which the correlation was either estimated or fixed at zero. The first model (the independent model, $r = 0$) assumed that the characters vary independently of each other. The second model (the dependent model) allowed one character to vary based on the other. Specifically, we used the method designed for analysis of the evolution of continuously varying traits, and a Bayesian procedure to test whether the position of each allele in the dominance hierarchy correlates with variables including: 1) the number of expressed sRNA precursors alleles carry; 2) generalism of their sRNA precursors, defined as the average number of alleles predicted to be targeted per sRNA precursor; 3) the "targeting spectrum" of an allele (the total number of alleles predicted to be targeted by the complete set of sRNAs produced by its sRNA precursors); 4) the number of targets carried; 5) the "generalism" of its targets (defined as the average number of alleles predicted to "use" each of its targets) and 6) the total number of alleles targeting its *SCR*+/-1kb sequence. A population of 300 trees homogenously sampled from the 3,000 trees obtained from the Bayesian analyses was used for this analysis to take into account the phylogenetic uncertainty and we forced the chain to spend an equal amount of time on each tree in the sample. For each analysis, parameters priors were defined as uniform distributions bounded by the extreme parameters values (minimum and maximum values) obtained during preliminary ML analyses on each of the 300 trees. We repeated each analysis three times to check for stability of the results. The number of generations was set to 1,010,000 (with a burnin of 10,000 generations). Convergence was checked visually by evaluating changes in the log-likelihood in *Tracer* v1.4 (*56*). We then compared the harmonic mean of the log-likelihood values obtained during the analyses assuming independent evolution of the two characters or allowing for correlated evolution (the dependent model) using the Bayesian Factor (BF). In line with ref. (*38*), we took results with BF>2 as evidence favoring the dependent model, >5 as strong evidence, and >10 as very strong evidence.

Reconstruction of the ancestral repertoire of sRNA precursors and their targets.

To reconstruct the ancestral repertoire of sRNA precursors and targets we used the Multistate MCMC procedure of *BayesTraits* (*54,55*). For each node in the Bayesian consensus tree (Fig. 4, Fig. S10) we tested the presence or the absence of a given precursor or target by fixing its state to 0 (absence) or 1 (presence). Priors for the rates of character changes ($q0 \rightarrow 1$: rate of character gain and $q1 \rightarrow 0$: rate of character loss) were uniform distributions between 0 and 1 for all analyses. Some of the nodes present in the consensus tree were absent from the 300 sampled trees, therefore nodes were defined using the most recent common ancestor approach as described in the *BayesTraits* manual. "Absence" and "presence" scenarios were compared *via* BF, with BF>2 favoring a given ancestral state, >5 as strong evidence, and >10 representing very strong evidence (*38*).

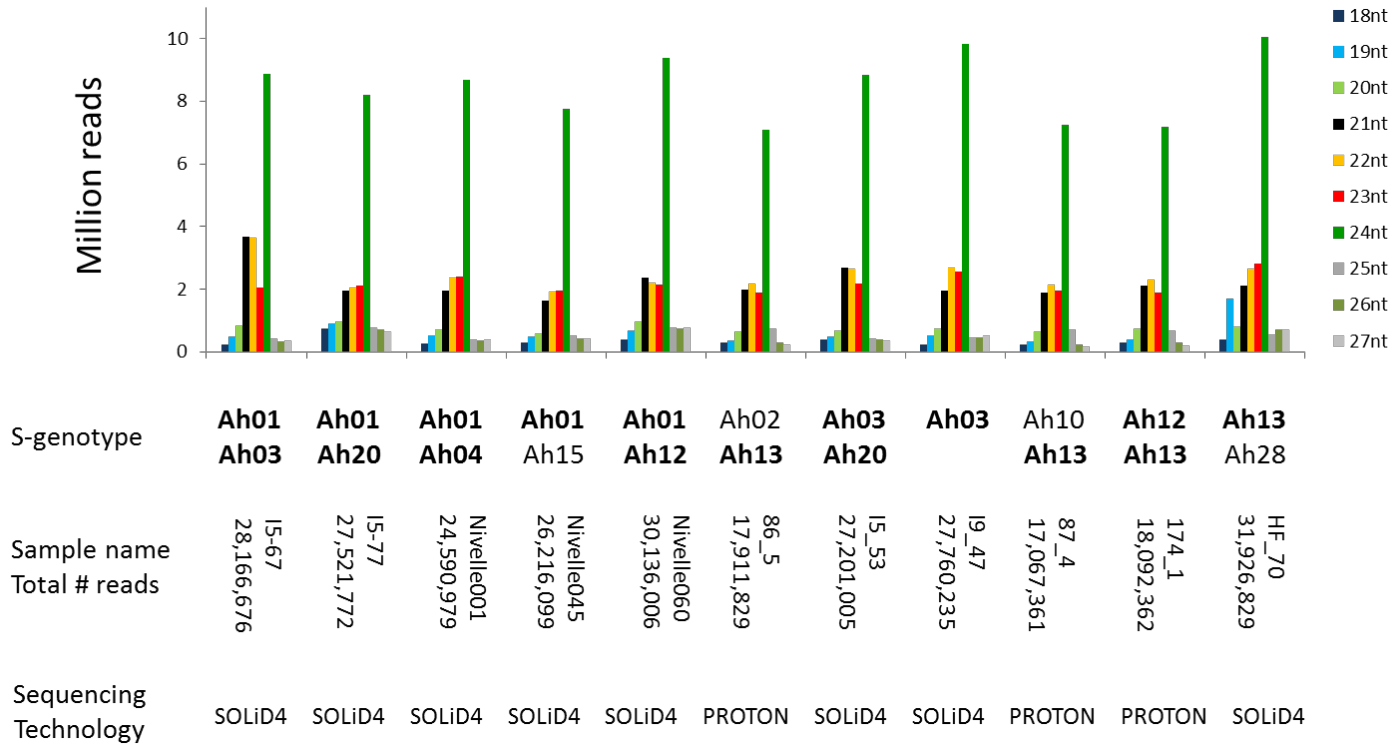
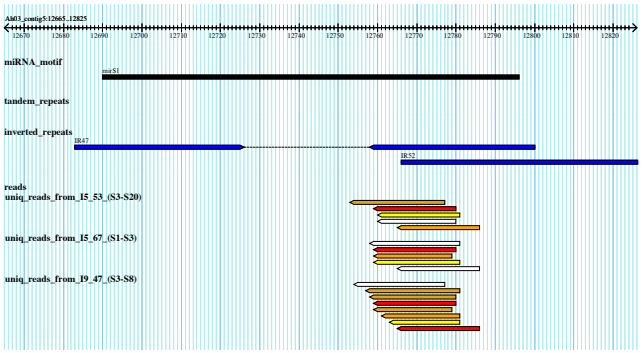
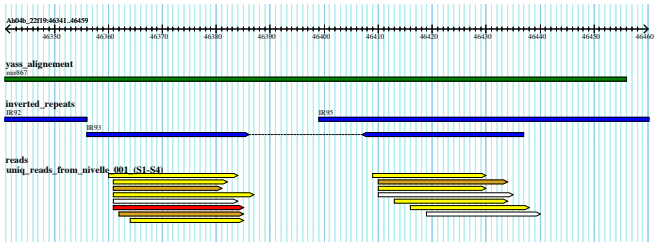


Figure S1. Size distribution of the small RNA reads and total number of reads obtained for each of the eleven plants analyzed. For one individual (I9-47), we identified a single S-allele, so this individual may either be homozygote for Ah03 or heterozygote with an additional S-allele that we were not able to identify.

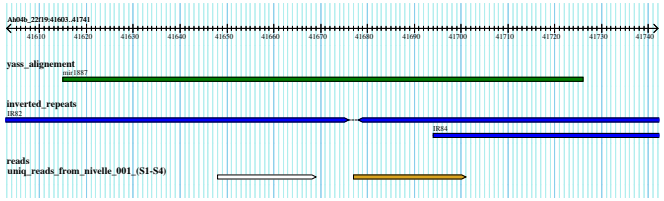
Ah03_mirS1



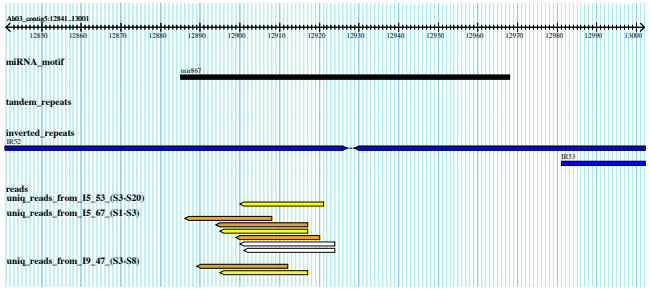
Ah04_mir867



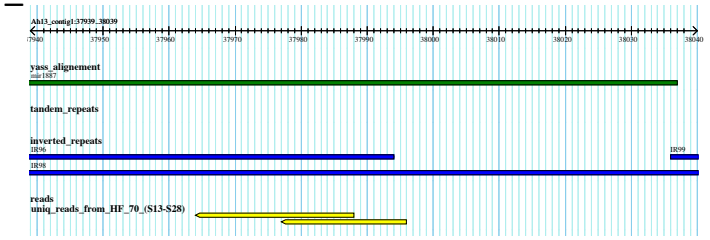
Ah04_mir1887



Ah03_mir867



Ah13_mir1887



Ah20_mir1887

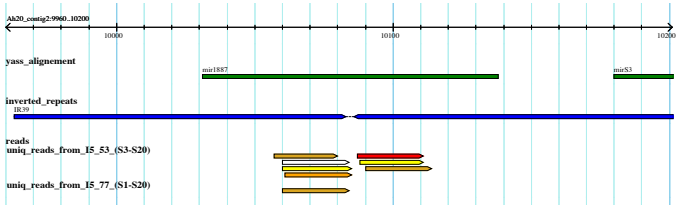
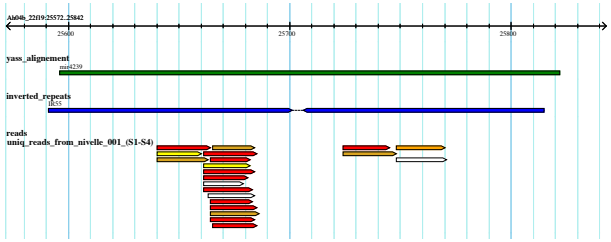
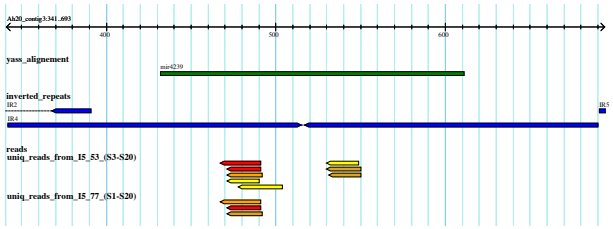


Figure S2. Mapping results for all predicted sRNA precursors on GBrowse. The predicted stemloop sequences are coloured in blue. Small RNA reads are coloured according their abundances (red: >20, orange : [11-20], light orange: [6-10], yellow: [2-5], white: 1).

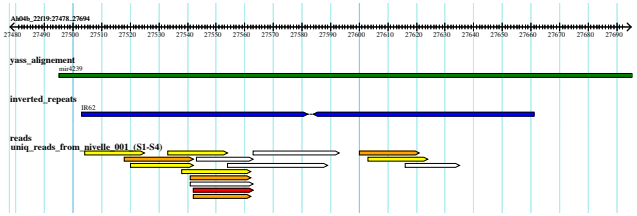
Ah04 mir4239



Ah20 mir4239



Ah04 mir4239b



Al1 mir4239

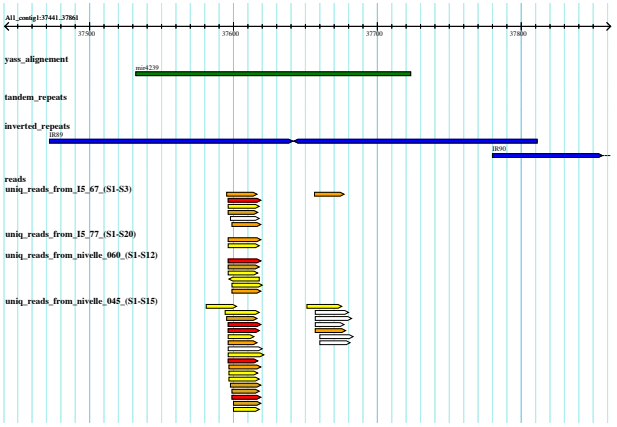
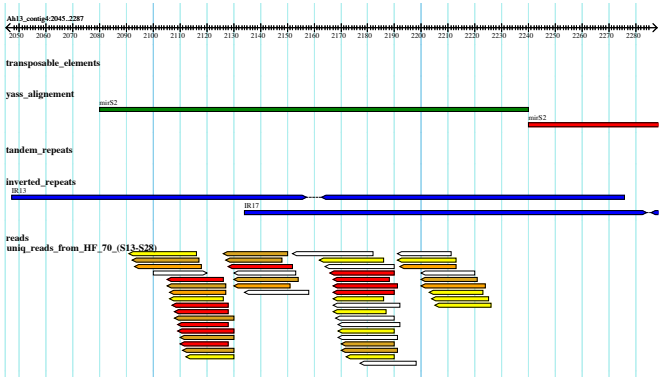


Figure S2. Continued.

Ah13_mirS2



Ah20_mirS2

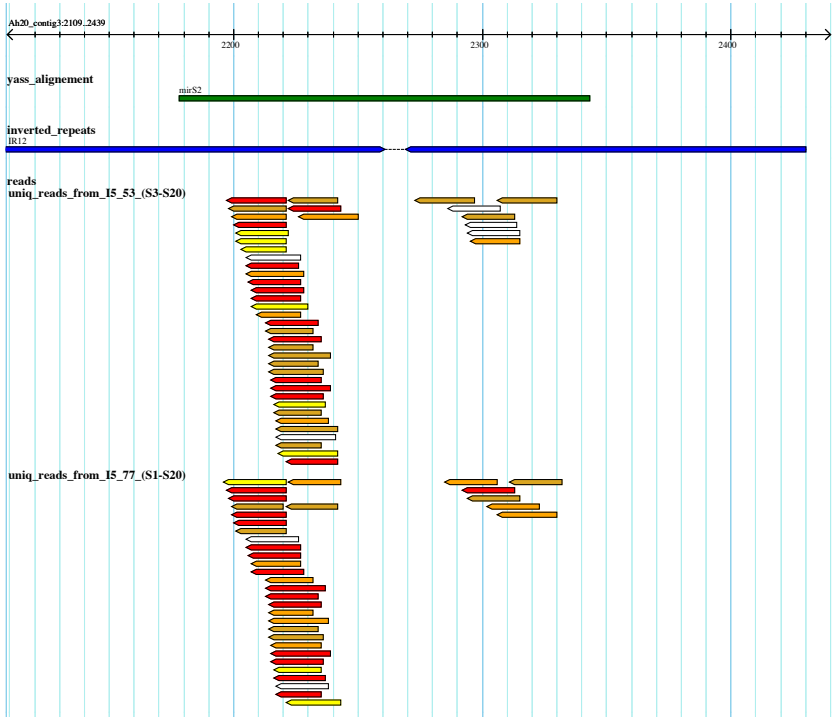
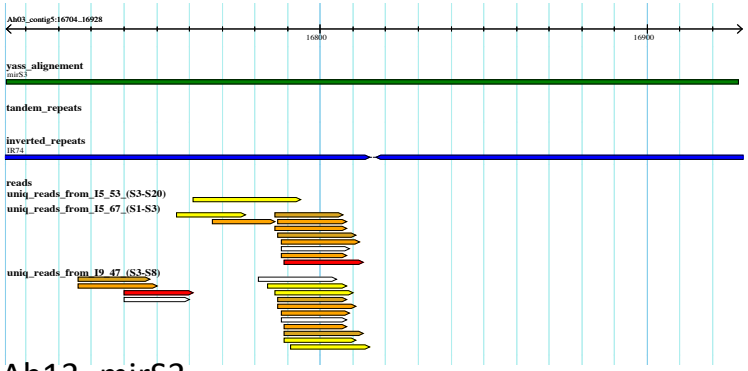
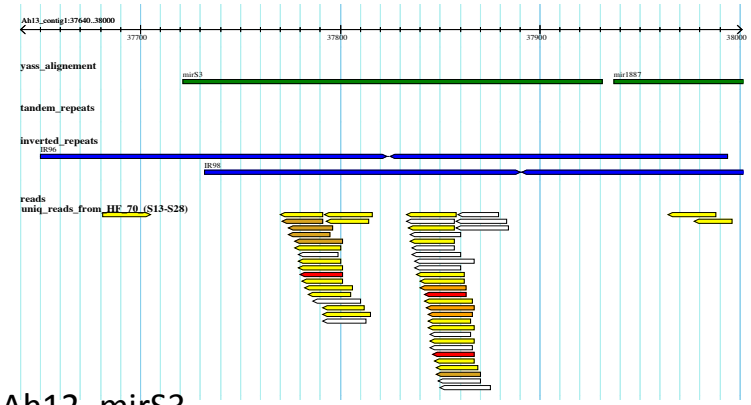


Figure S2. Continued.

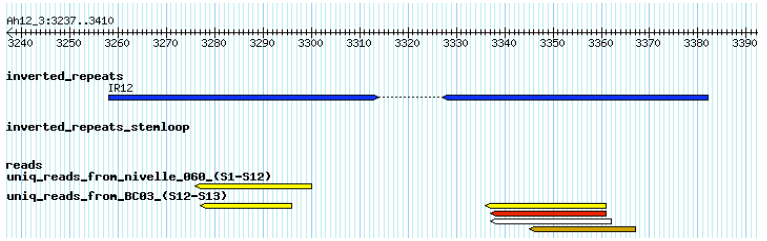
Ah03_mirS3



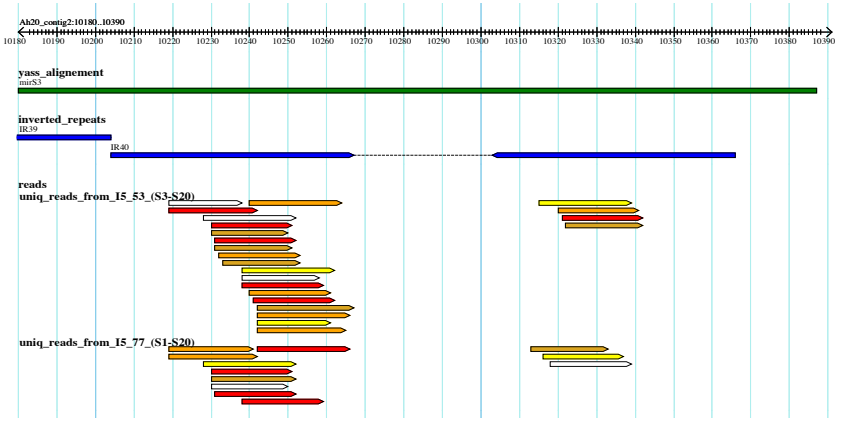
Ah13_mirS3



Ah12_mirS3



Ah20_mirS3



Ah04_mirS4

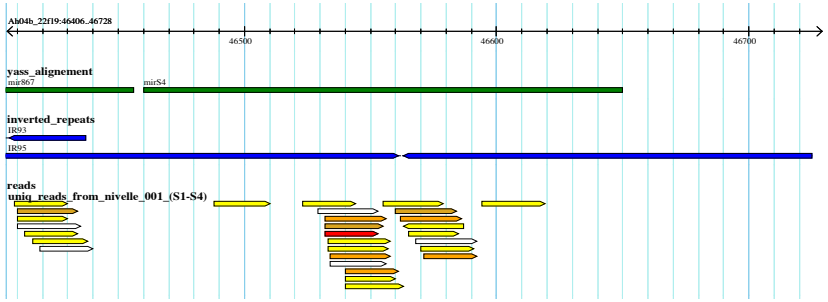


Figure S2. Continued.

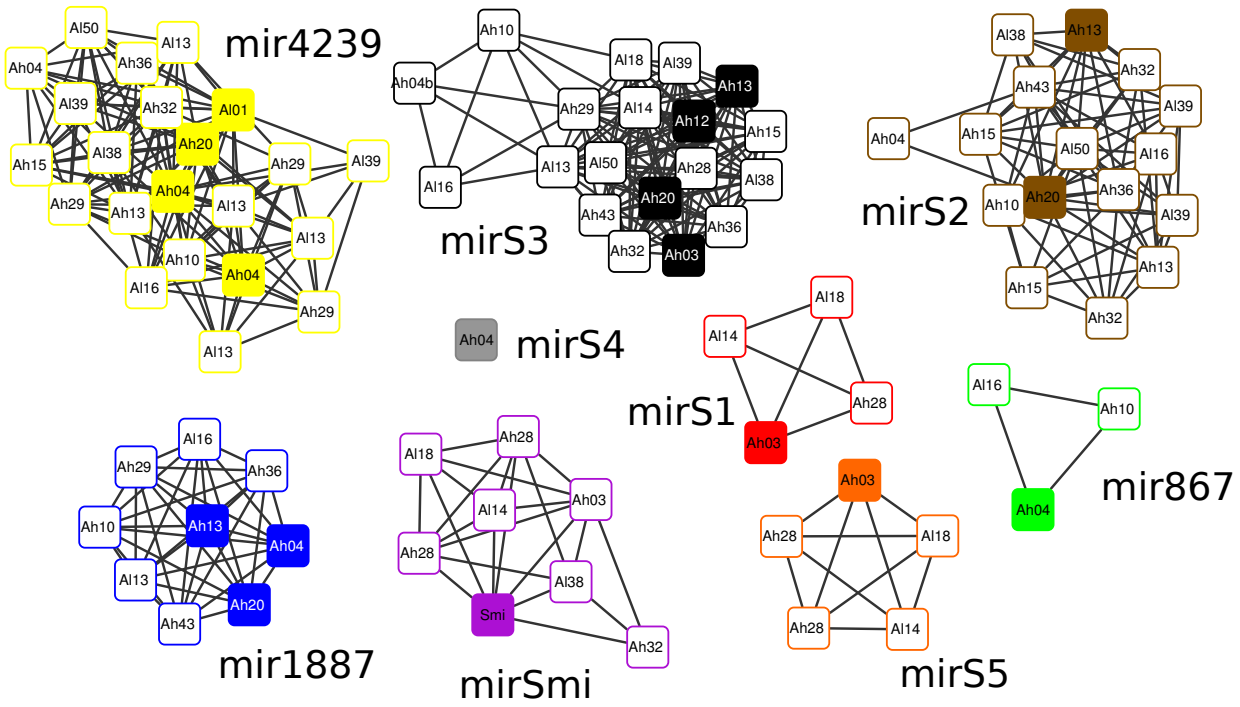


Figure S3. Results from the sequential *YASS* analysis. Two sRNA precursors are connected by a line if a reciprocal *YASS* hit is found between them spanning at least 50% of the sequence length with sequence identity above 70% and an *e*-value threshold of 10^{-4}. Solid boxes indicate the 17 sRNA precursors identified from the sRNAseq analysis on the 6 initial S-alleles, while empty boxes indicate additional sRNA precursors identified based on sequence similarity.

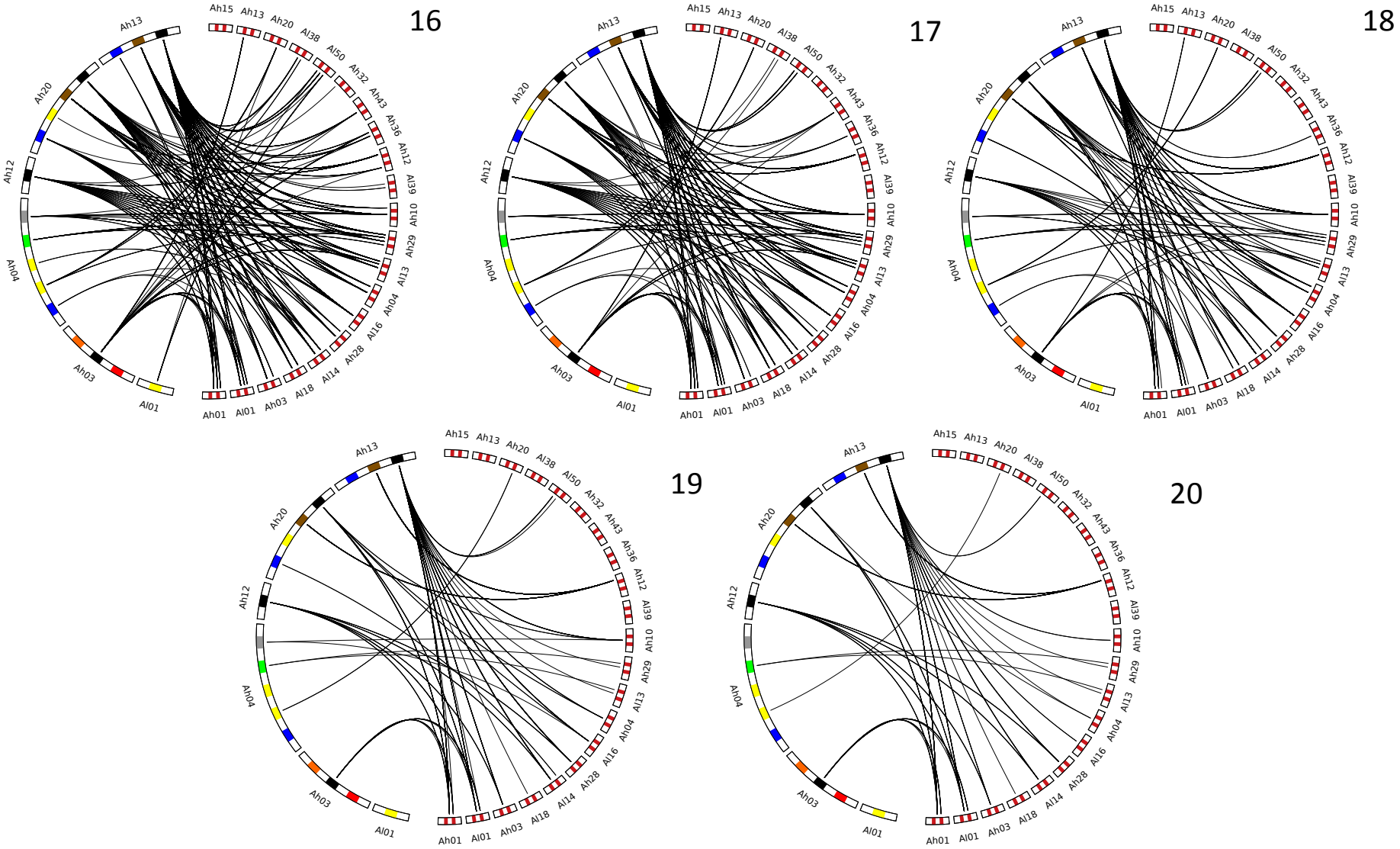


Figure S4. Sensibility analysis of the regulatory network to the identification threshold = 16, 17, 18, 19 and 20. For this analysis, all 20 *SCR* alleles in Fig. 4 were included.

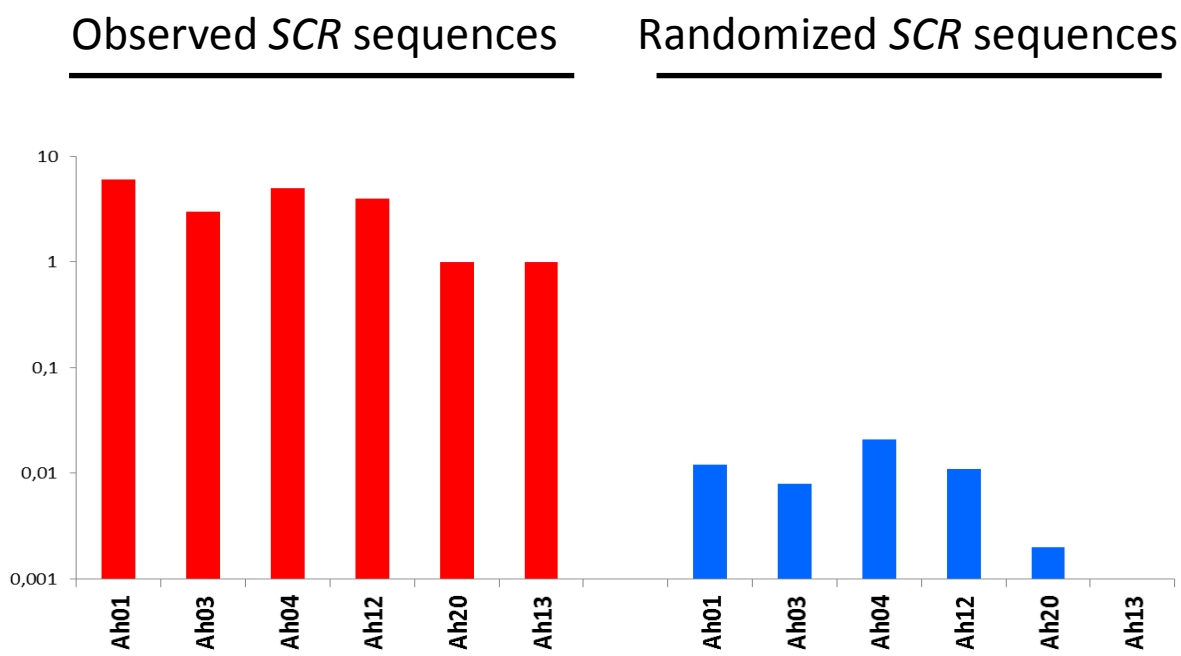


Figure S5. *SCR* sequences are enriched in sRNA targets. The figure shows the number of predicted targets on the observed *vs.* randomized *SCR* sequences.

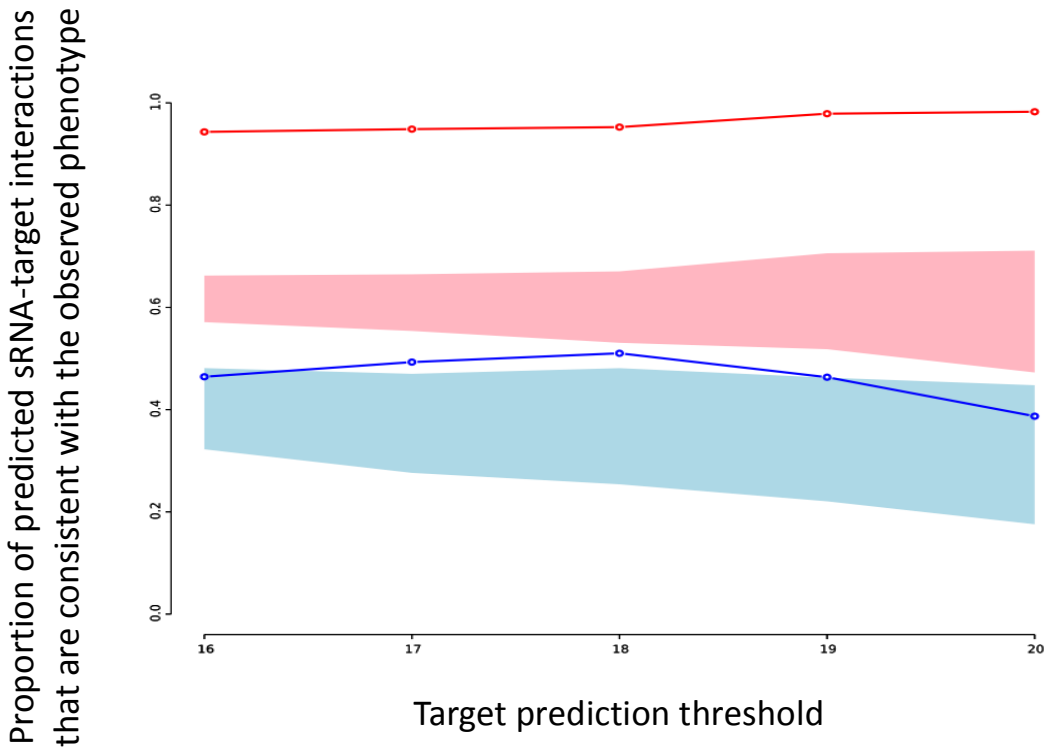


Figure S6. Proportion of predicted sRNA-target interactions that are consistent with the observed phenotype (from a dominant to a recessive allele) for different levels of stringency for the definition of the targets (from 16 = low stringency, leading to a large number of targets to 20 = high stringency, leading to a low number of targets). The red line represents sRNAs produced from hairpin precursors and the blue line represents siRNAs produced from the rest of the S-locus sequence for each allele (also unique and perfect matches). Blue and red shaded areas represent the random distribution obtained by 100,000 random rewiring of the predicted targeting network, keeping the number of sRNAs produced by each S-allele constant.

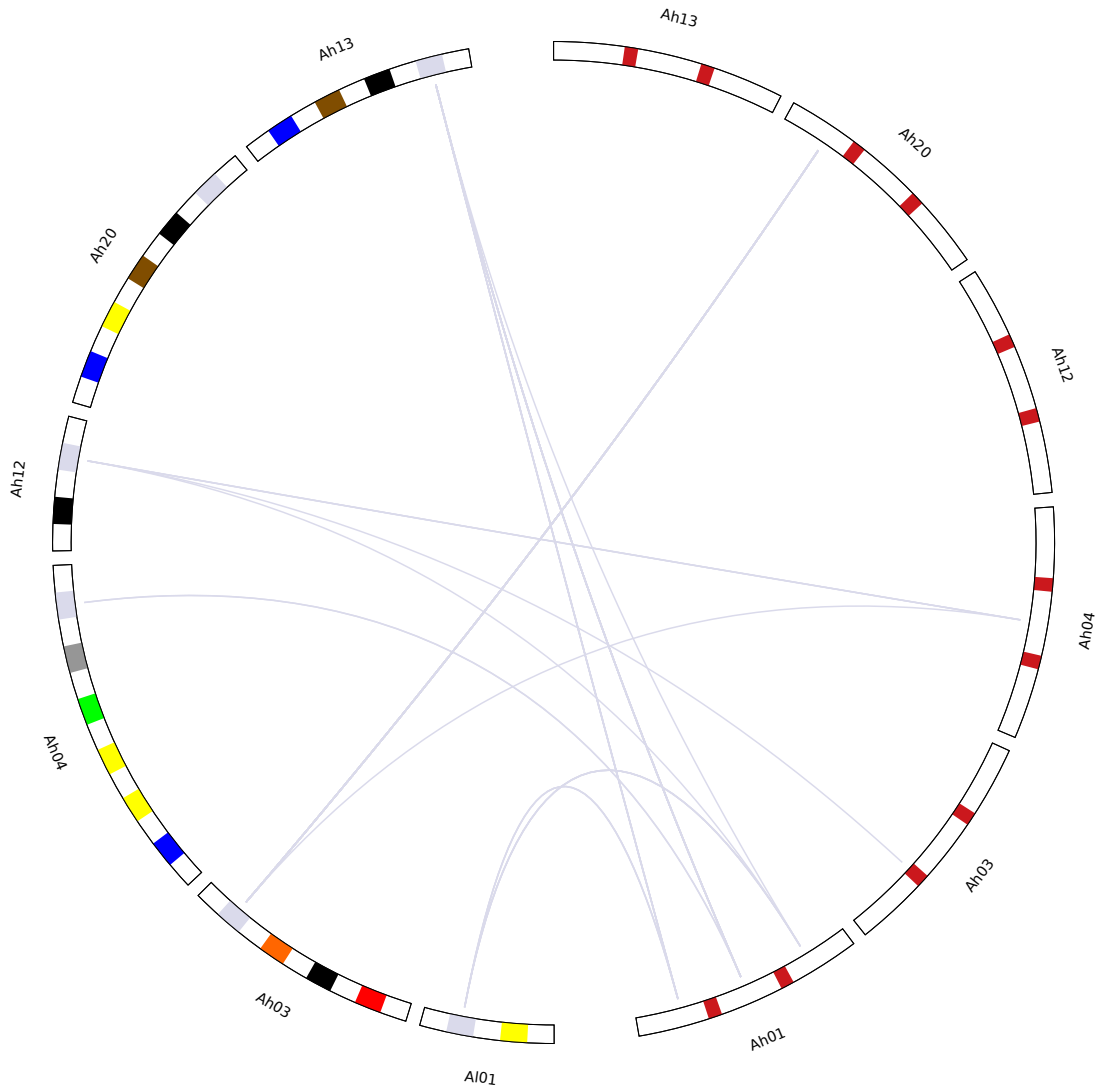


Figure S7. Predicted regulatory network based on sRNAs not produced by hairpin precursors (most likely siRNAs), showing no particular trend toward targeting of more recessive alleles.

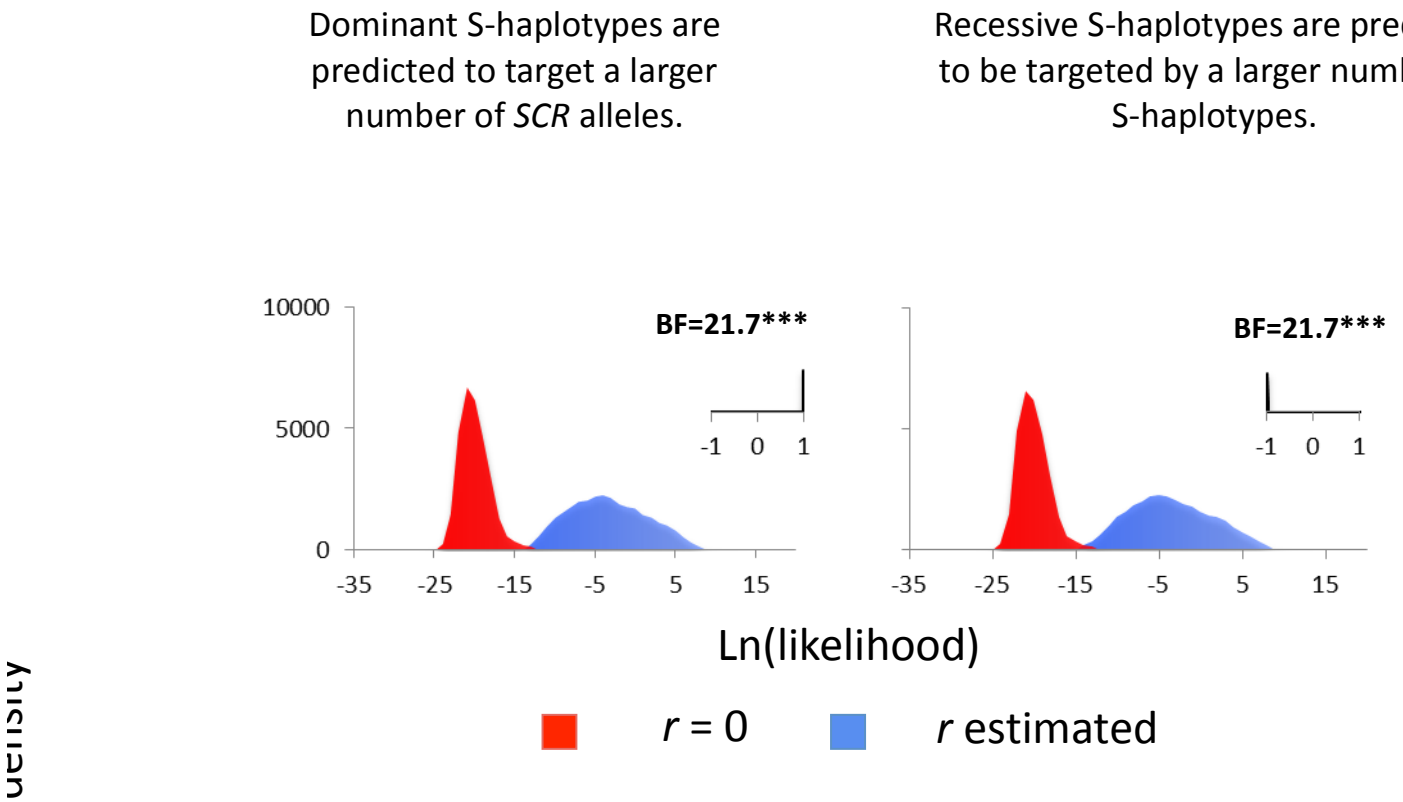


Figure S8. Likelihood distributions of Bayesian models assuming no correlation or allowing for some level of correlated evolution between: **A.** dominance and the number of alleles targeted by the sRNAs produced by a given allele. **B.** dominance and the number of alleles collectively targeting the target sites of a given allele. The Bayes factor comparing both models is given on each panel, as well as the distribution of correlation coefficients.

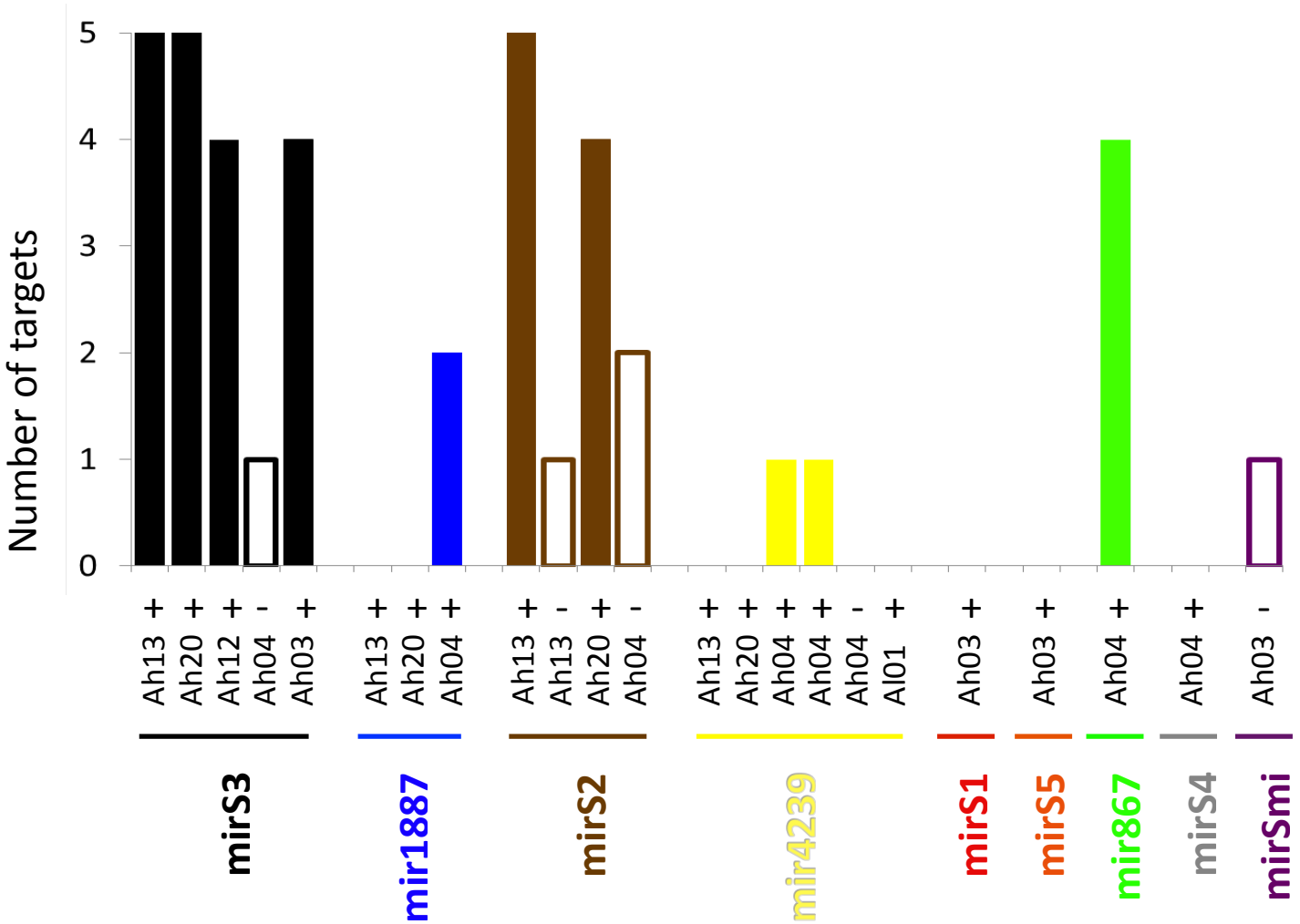


Figure S9. Number of predicted targets of expressed ("+", solid bars) *vs.* silent precursors ("-", open bars). Targets were predicted from *in silico* processing of complete precursors into 21nt sRNAs. The threshold score for target prediction was 18.

	mirS1			mirS2			mirS3			mirS4			mirS5		
	0	1	BF	0	1	BF	0	1	BF	0	1	BF	0	1	BF
Node 1	-4,292	-18,212	27,840	-21,632	-9,496	24,271	-17,981	-4,085	27,792	-5,061	-16,670	23,219	-4,289	-18,219	27,860
Node 2	-4,293	-11,873	15,159	-13,687	-9,233	8,909	-12,209	-4,079	16,260	-5,103	-8,241	6,277	-4,295	-13,118	17,647
Node 3	-4,289	-22,936	37,294	-9,221	-23,727	29,012	-22,834	-4,073	37,522	-5,055	-23,164	36,217	-4,294	-22,989	37,390
Node 4	-4,291	-10,733	12,885	-11,868	-9,388	4,961	-11,583	-4,079	15,006	-5,083	-9,427	8,689	-4,294	-10,658	12,728
Node 5	-13,437	-4,293	18,288	-9,264	-17,668	16,808	-14,003	-4,081	19,843	-5,063	-15,006	19,885	-13,478	-4,288	18,381
Node 6	-11,032	-4,291	13,483	-9,247	-15,264	12,034	-12,940	-4,077	17,725	-5,054	-13,867	17,626	-11,053	-4,294	13,518
Node 7	-10,193	-4,297	11,792	-9,221	-14,701	10,959	-13,008	-4,069	17,878	-5,065	-13,857	17,585	-10,184	-4,293	11,783
Node 8	-4,315	-8,384	8,137	-12,148	-9,294	5,709	-11,265	-4,071	14,390	-5,064	-11,372	12,615	-4,318	-8,485	8,335
Node 9	-4,291	-10,851	13,122	-14,221	-9,504	9,435	-12,012	-4,075	15,873	-5,051	-12,626	15,151	-4,291	-10,642	12,702
Node 10	-4,296	-10,597	12,602	-14,154	-9,176	9,956	-12,606	-4,088	17,035	-5,054	-12,920	15,732	-4,297	-10,574	12,553
Node 11	-4,289	-13,471	18,364	-17,950	-9,223	17,454	-13,171	-4,082	18,177	-5,061	-14,271	18,420	-4,297	-13,470	18,345
Node 12	-4,295	-13,967	19,344	-18,016	-9,261	17,509	-13,749	-4,076	19,347	-5,055	-14,699	19,288	-4,288	-14,012	19,449
Node 13	-4,282	-13,355	18,146	-16,847	-9,273	15,149	-13,035	-4,078	17,914	-5,047	-14,063	18,031	-4,297	-13,409	18,224
Node 14	-4,298	-12,351	16,106	-15,021	-9,261	11,521	-11,415	-4,088	14,655	-5,050	-13,227	16,355	-4,289	-12,469	16,361
Node 15	-4,288	-14,192	19,807	-14,722	-9,242	10,959	-13,126	-4,083	18,085	-5,059	-15,670	21,220	-4,295	-14,178	19,767
Node 16	-4,303	-10,902	13,198	-14,194	-9,188	10,011	-11,466	-4,077	14,779	-5,063	-12,391	14,656	-4,296	-10,821	13,051

sRNA precursors

	mirSmi			mir4239			mir 1887			mir 867		
	0	1	BF	0	1	BF	0	1	BF	0	1	BF
Node 1	-8,808	-22,266	26,915	-21,987	-7,992	27,990	-25,383	-13,051	24,663	-16,478	-5,551	21,854
Node 2	-8,827	-16,136	14,620	-15,464	-8,039	14,850	-17,660	-13,039	9,241	-8,056	-5,647	4,818
Node 3	-8,829	-27,537	37,417	-26,357	-7,994	36,726	-29,918	-13,053	33,729	-5,538	-21,473	31,869
Node 4	-8,823	-15,083	12,520	-14,157	-8,006	12,302	-15,996	-13,093	5,807	-5,945	-7,024	2,159
Node 5	-17,292	-8,845	16,895	-7,990	-16,464	16,948	-13,320	-22,052	17,464	-5,534	-15,228	19,387
Node 6	-14,996	-8,829	12,334	-7,992	-14,070	12,156	-13,047	-20,168	14,243	-5,539	-13,950	16,823
Node 7	-14,206	-8,838	10,735	-8,000	-13,413	10,826	-13,044	-19,743	13,397	-5,556	-13,875	16,638
Node 8	-8,898	-12,974	8,153	-11,896	-8,005	7,781	-13,302	-14,566	2,527	-5,568	-9,502	7,868
Node 9	-8,884	-15,619	13,470	-13,768	-7,997	11,541	-13,260	-14,819	3,118	-5,550	-11,914	12,728
Node 10	-8,831	-15,344	13,026	-13,678	-7,992	11,372	-13,240	-14,769	3,059	-5,543	-11,530	11,973
Node 11	-8,827	-12,929	8,204	-16,749	-8,005	17,487	-13,203	-15,004	3,602	-5,549	-14,551	18,004
Node 12	-8,810	-15,187	12,755	-16,857	-8,000	17,714	-13,170	-15,210	4,079	-5,540	-15,198	19,316
Node 13	-8,884	-16,129	14,492	-15,747	-8,057	15,379	-13,203	-15,075	3,743	-5,539	-14,445	17,813
Node 14	-8,861	-14,007	10,292	-11,910	-8,035	7,749	-13,213	-15,510	4,594	-5,527	-13,385	15,716
Node 15	-8,849	-17,031	16,364	-13,870	-7,989	11,762	-13,111	-15,428	4,635	-5,542	-15,247	19,411
Node 16	-8,813	-15,745	13,863	-13,569	-8,003	11,132	-13,360	-15,004	3,289	-5,544	-11,909	12,730

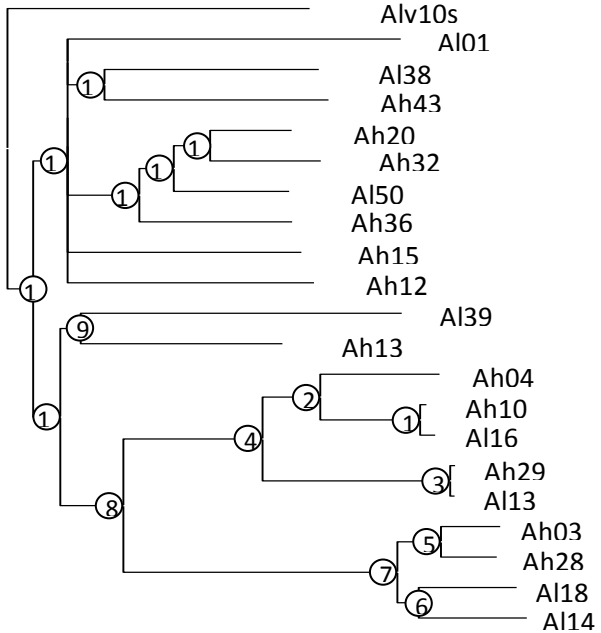


Figure S10. Ancestral states reconstruction of the presence /absence pattern of the different sRNA precursors and their predicted target sites. The harmonic mean of the log-likelihoods of the two alternative states analyses are reported (where "0" stands for absence and "1" for presence) and the associated Bayes Factor. Ancestral states inferences associated with BF>2 are highlighted in yellow. Ancestral nodes are numbered according to the phylogeny shown on the bottom right. BF>2: evidence, BF>5: strong evidence, and BF>10: very strong evidence (*38*).

	Tmir867IE			Tmir1887P			Tmir4239E2			TmirS2P1		
	0	1	BF	0	1	BF	0	1	BF	0	1	BF
Node 1	-4,612	-17,530	25,835	-6,927	-21,028	28,203	-11,713	-24,837	26,250	-11,092	-20,654	19,124
Node 2	-4,616	-9,861	10,489	-6,918	-14,726	15,616	-11,713	-18,621	13,816	-11,570	-12,162	1,184
Node 3	-19,628	-4,618	30,020	-6,936	-25,400	36,929	-12,324	-19,078	13,507	-26,261	-11,108	30,307
Node 4	-4,660	-8,134	6,949	-6,919	-14,102	14,367	-11,723	-16,946	10,448	-11,785	-12,135	0,699
Node 5	-4,610	-14,552	19,885	-6,916	-14,546	15,261	-11,729	-21,378	19,298	-11,102	-17,675	13,144
Node 6	-4,608	-13,482	17,747	-6,973	-10,775	7,605	-11,749	-20,135	16,774	-11,143	-14,986	7,686
Node 7	-4,619	-13,393	17,547	-6,993	-11,637	9,289	-11,796	-20,103	16,614	-11,093	-15,422	8,658
Node 8	-4,613	-10,295	11,364	-6,919	-13,401	12,965	-11,729	-18,028	12,599	-11,146	-14,533	6,774
Node 9	-4,622	-11,911	14,578	-6,910	-14,497	15,175	-11,715	-19,292	15,153	-11,114	-16,720	11,212
Node 10	-4,618	-12,250	15,262	-6,911	-14,963	16,103	-11,707	-19,602	15,789	-11,176	-17,134	11,917
Node 11	-4,614	-13,691	18,155	-6,921	-15,875	17,908	-11,761	-14,989	6,456	-11,055	-19,626	17,143
Node 12	-4,609	-14,408	19,596	-6,914	-16,222	18,615	-11,711	-17,262	11,101	-11,069	-20,216	18,293
Node 13	-4,603	-13,637	18,068	-6,907	-15,637	17,459	-11,702	-18,460	13,517	-11,080	-19,192	16,224
Node 14	-4,609	-12,772	16,327	-6,924	-14,098	14,347	-11,724	-19,529	15,609	-11,115	-18,024	13,817
Node 15	-4,620	-15,044	20,849	-6,918	-15,999	18,160	-11,710	-21,761	20,101	-11,090	-18,396	14,612
Node 16	-4,620	-11,745	14,250	-6,908	-14,226	14,636	-11,725	-19,046	14,643	-11,061	-17,131	12,139

	TmirS2P2			TmirS3I			TmirS3IE			TmirS4I		
	0	1	BF	0	1	BF	0	1	BF	0	1	BF
Node 1	-4,363	-18,268	27,809	-20,018	-6,896	26,245	-12,330	-24,677	24,694	-14,168	-5,213	17,910
Node 2	-4,368	-12,502	16,268	-12,389	-6,909	10,960	-12,329	-17,089	9,521	-5,283	-8,055	5,543
Node 3	-4,369	-23,407	38,077	-24,447	-6,945	35,003	-27,472	-12,385	30,174	-5,283	-22,865	35,163
Node 4	-4,359	-11,807	14,897	-10,716	-6,949	7,535	-12,519	-15,416	5,795	-5,225	-9,240	8,030
Node 5	-4,365	-14,303	19,876	-15,906	-6,919	17,974	-12,312	-21,796	18,968	-5,222	-15,075	19,707
Node 6	-4,380	-13,232	17,703	-13,856	-6,925	13,861	-12,469	-20,345	15,751	-5,206	-14,010	17,610
Node 7	-4,355	-13,576	18,443	-13,258	-6,966	12,584	-12,395	-20,132	15,474	-5,216	-14,130	17,828
Node 8	-4,363	-11,591	14,455	-8,012	-7,943	0,136	-12,327	-15,985	7,315	-5,212	-11,228	12,031
Node 9	-4,364	-12,117	15,506	-6,925	-10,851	7,852	-12,328	-15,485	6,313	-5,208	-12,670	14,923
Node 10	-4,366	-12,496	16,260	-6,928	-10,685	7,514	-12,343	-15,756	6,826	-5,217	-12,941	15,449
Node 11	-4,370	-13,454	18,168	-6,937	-15,959	18,044	-12,407	-16,031	7,247	-5,210	-14,332	18,245
Node 12	-4,367	-13,794	18,853	-6,955	-16,322	18,734	-12,502	-14,523	4,041	-5,209	-14,874	19,329
Node 13	-4,359	-12,947	17,176	-6,925	-15,461	17,073	-12,512	-14,380	3,736	-5,212	-14,241	18,058
Node 14	-4,355	-11,887	15,064	-6,901	-13,556	13,309	-12,436	-17,882	10,893	-5,216	-13,319	16,205
Node 15	-4,368	-12,586	16,437	-6,919	-13,527	13,215	-12,330	-16,192	7,723	-5,206	-15,863	21,313
Node 16	-4,375	-11,329	13,909	-6,927	-11,234	8,614	-12,378	-15,893	7,030	-5,210	-12,878	15,338

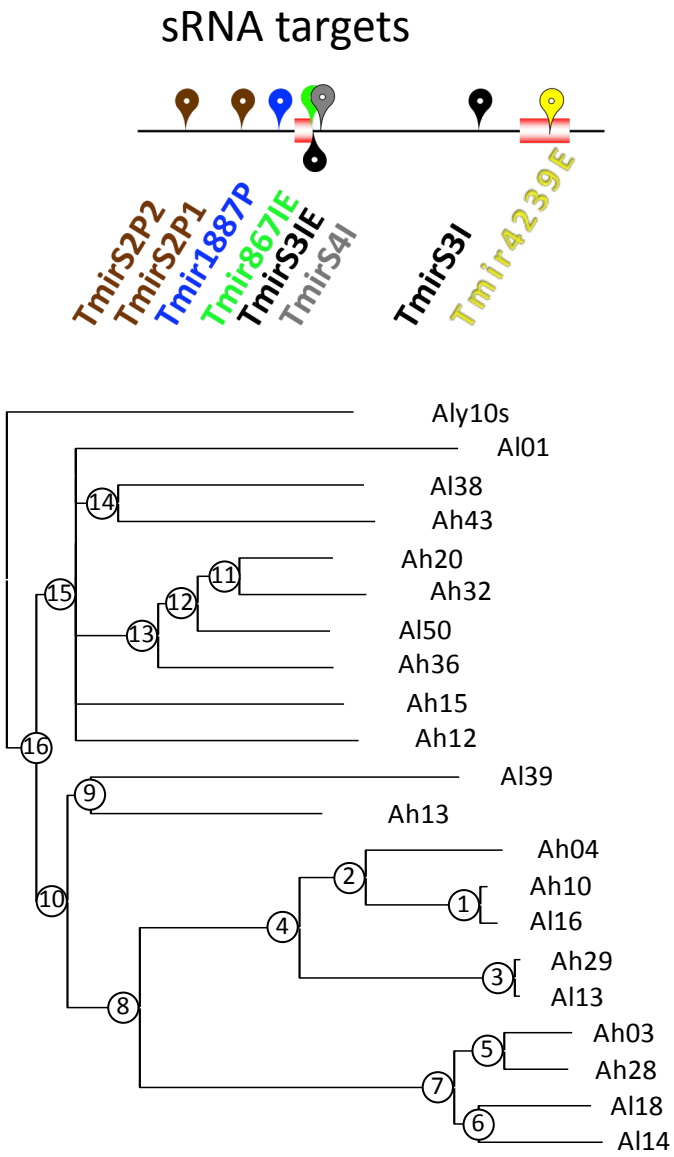


Figure S10. Continued.

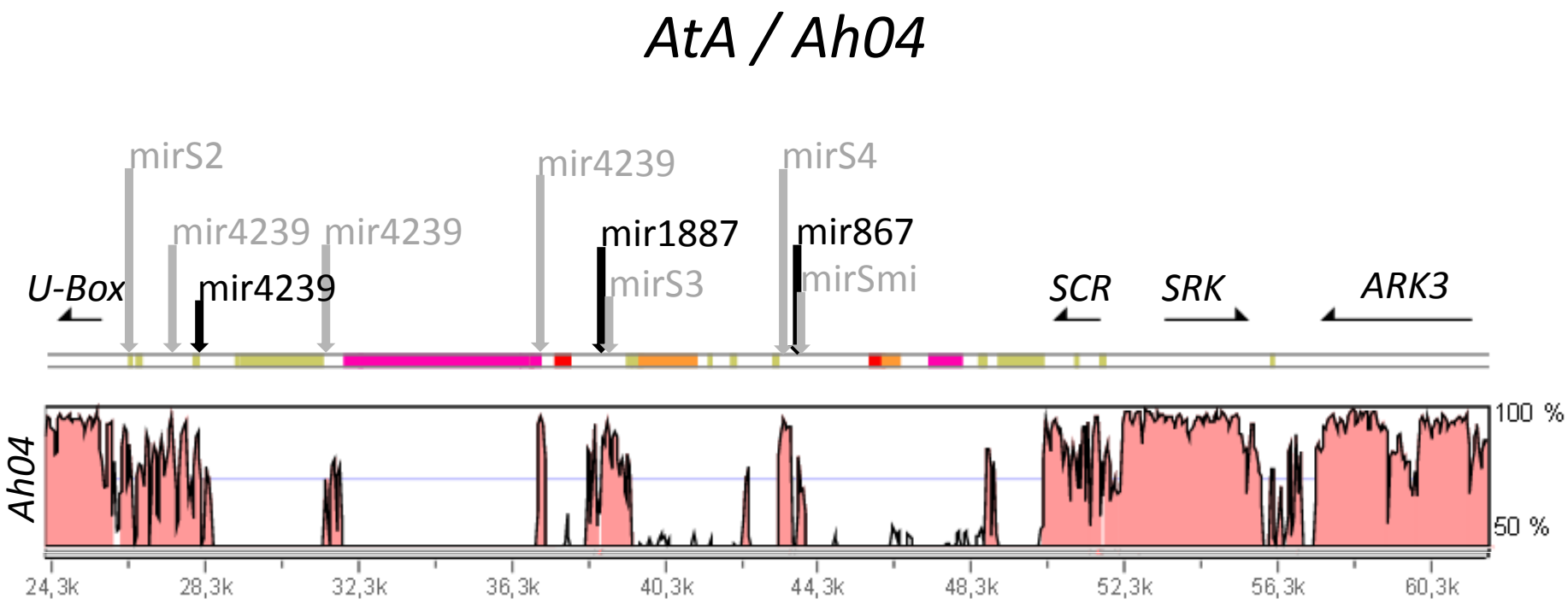


Figure S11. *VISTA* plots (*57*) showing the level of sequence conservation between the S-locus region of functional orthologs in *A. halleri*, *A. lyrata* and *A. thaliana*. Expressed sRNA genes are represented in black, while silent and sRNA precursors for which no expression data are available are represented in grey.

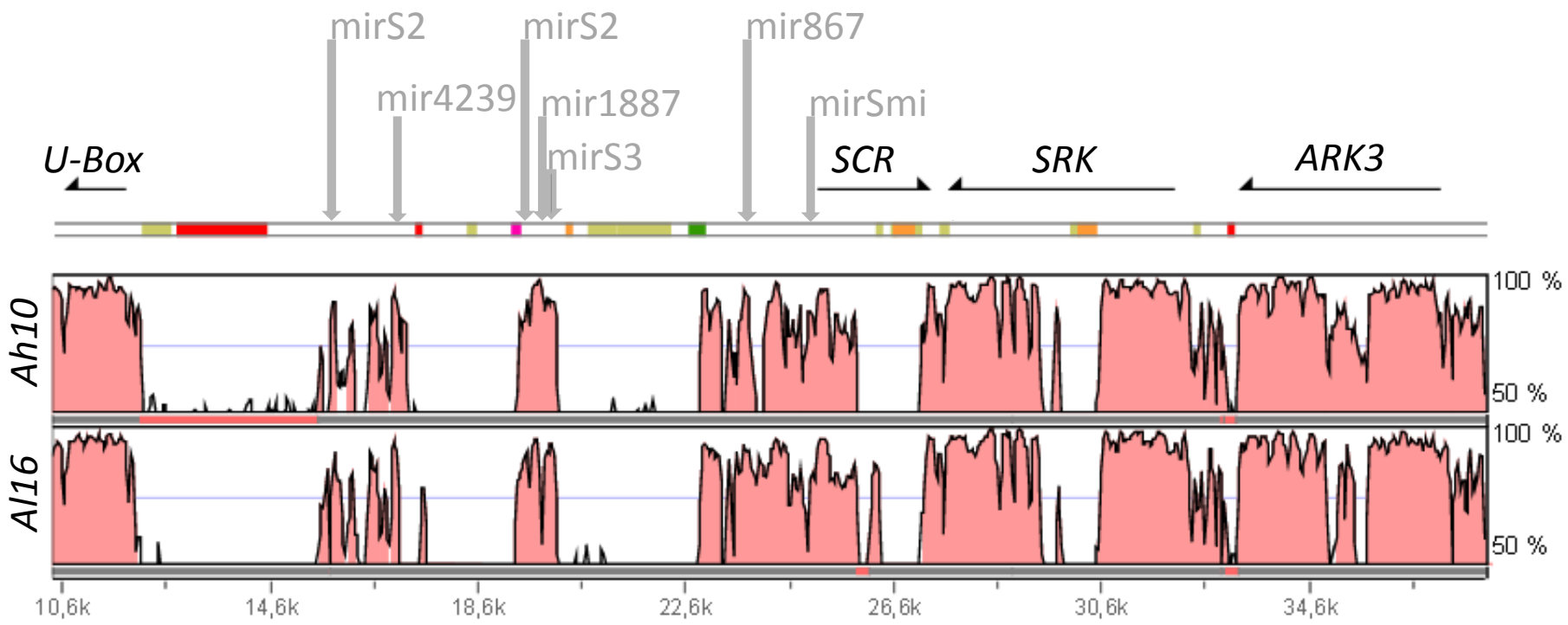


Figure S11. Continued.

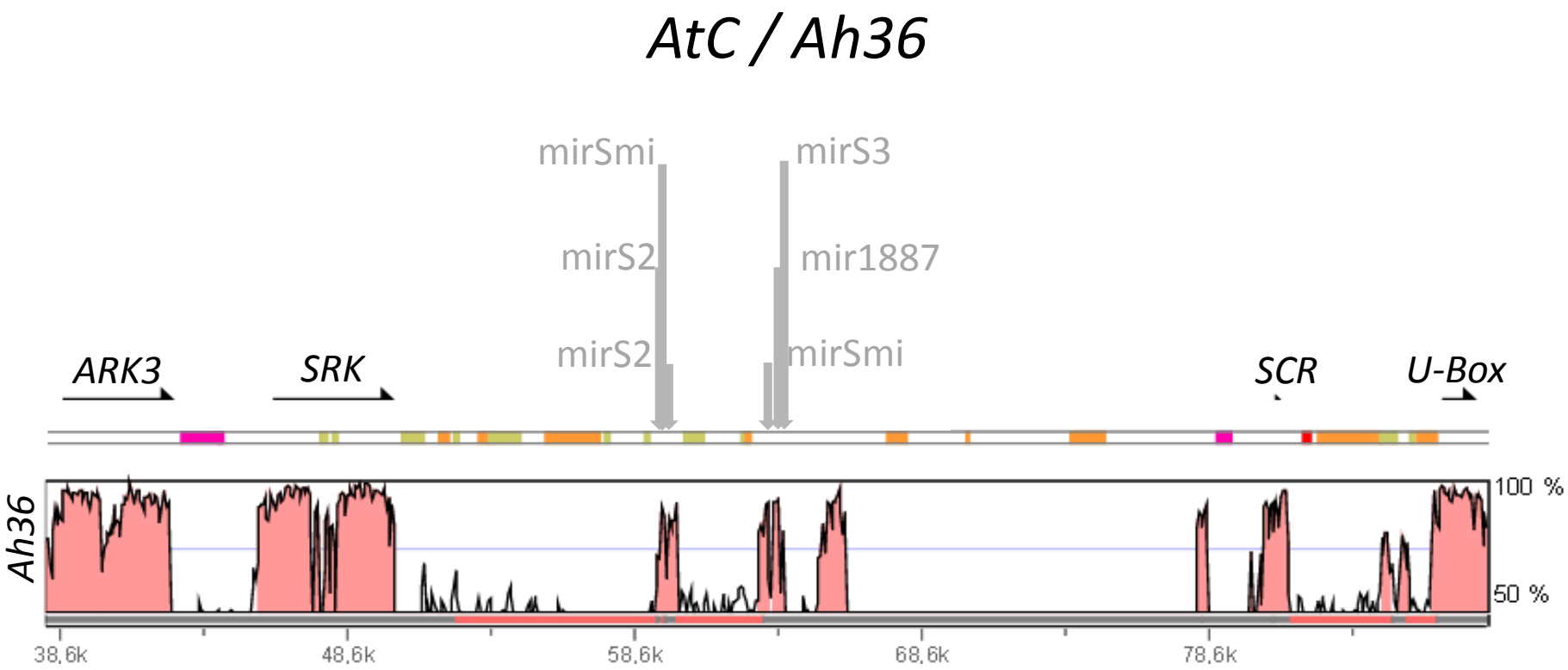


Figure S11. Continued.

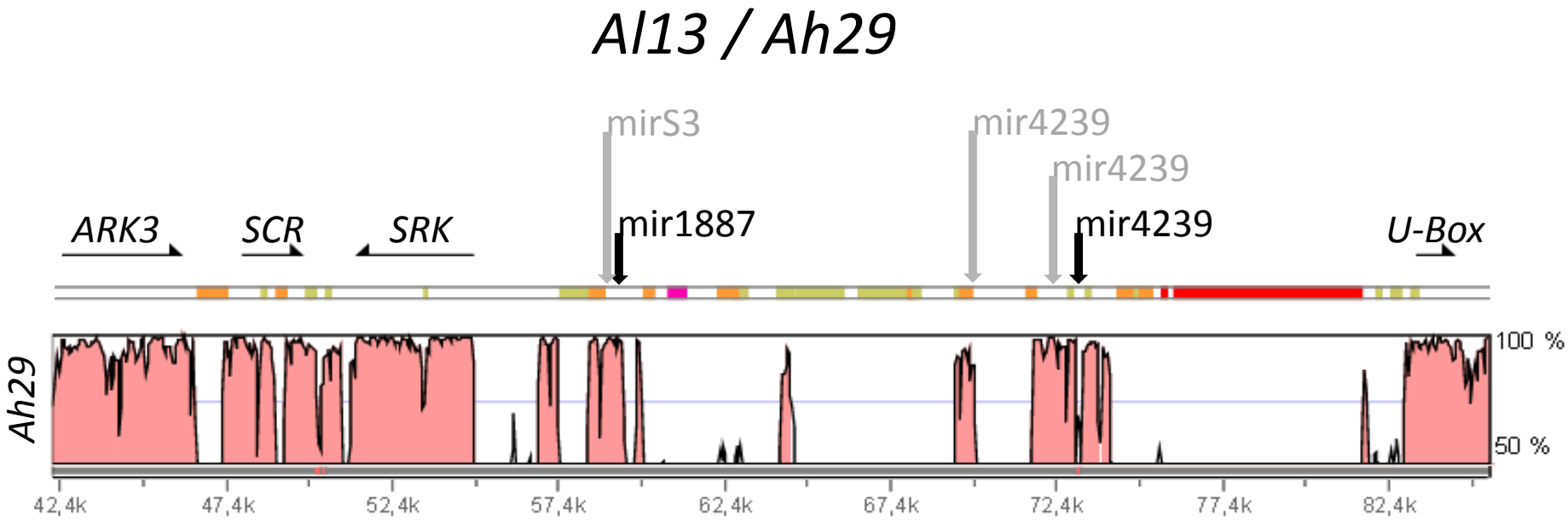


Figure S11. Continued.

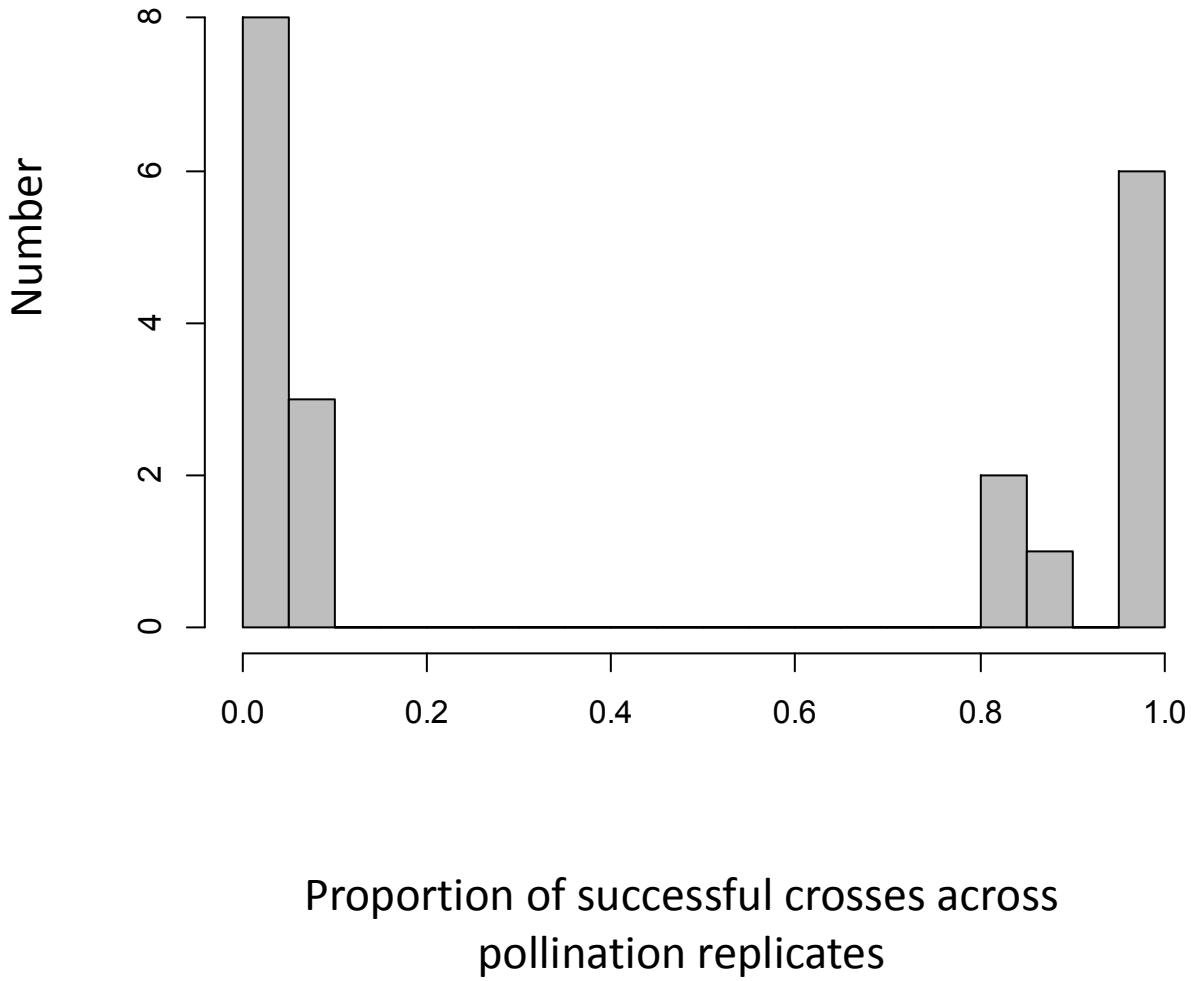


Figure S12. Distribution of the proportion of successful crosses across pollination replicates.

Table S1. Crossing design for functional validation of the *Ah20mirS3*/*AhSCR01* interaction.

		Males				
		$\frac{AhSCR01}{WT}\frac{WT}{WT}^{a}$	$\frac{AhSCR01}{WT}\frac{WT}{Ah20mirS3-6}^{b}$	$\frac{AhSCR01^{*}}{WT}\frac{WT}{WT}^{a}$	$\frac{AhSCR01^{*}}{WT}\frac{WT}{Ah20mirS3-6}^{b}$	$\frac{Ah20mirS3-6}{Ah20mirS3-6}$
Females	$\frac{AhSRK01}{AhSRK01}$	Incompatible	Compatible[c]	Incompatible	Incompatible[c]	Compatible
	$\frac{WT}{WT}(C24)$	Compatible	Compatible	Compatible	Compatible	

Pollinations were performed with pollen from:

[a] hemizygous individuals for *AhSCR01* or *AhSCR01**.

[b] Hybrids plants, hemizygous for both *Ah20mirS3*-6 and *AhSCR01*or *AhSCR01**.

[c] These crosses were made with 2 replicates insertion *Ah20mirS3* lines (*Ah20mirS3*-6 and *Ah20mirS3*-12).

Table S2. S-locus sequences (EMBL database)

Name	Accession	Reference
Ah03	KJ772378-KJ772385	Goubet *et al.* (2012)[a]
Ah04	KJ461484	this study
Ah10	KM592810-KM592817	this study
Ah12	KJ772373-KJ772377	this study
Ah13	KJ461479-KJ461483	Goubet *et al.* (2012)[a]
Ah15	KJ772386-KJ772395	Goubet *et al.* (2012)[a]
Ah20	KJ772396-KJ772400	Goubet *et al.* (2012)[a]
Ah28	KJ461475-KJ461478	Goubet *et al.* (2012)[a]
Ah29	KM592798-KM592803	this study
Ah32	KJ461470-KJ461474	Goubet *et al.* (2012)[a]
Ah36	KM592804-KM592809	this study
Ah43	KJ461485-KJ461492	Goubet *et al.* (2012)[a]
Al1	KJ772401-KJ772404	Goubet *et al.* (2012)[a]
Al13	genome Araly1	Hu *et al.* (2011)[b]
Al14	KJ772405-KJ772407	Goubet *et al.* (2012)[a]
Al16	HQ379629	Guo *et al.* (2011)[c]
Al18	KJ772408-KJ772414	Goubet *et al.* (2012)[a]
Al38	HQ379630	Guo *et al.* (2011)[c]
Al39	KJ772415-KJ772419	Goubet *et al.* (2012)[a]
Al50	HQ379631	Guo *et al.* (2011)[c]
Aly Sb	HQ379628	Guo *et al.* (2011)[c]
Small RNA sequencing (GEO database)		
I5-67	GSM1378102	this study
I5-77	GSM1378103	this study
Nivelle001	GSM1378105	this study
Nivelle045	GSM1378106	this study
Nivelle060	GSM1378107	this study
86-5 (BC01)	GSM1378109	this study
I5-53	GSM1378101	this study
I9-47	GSM1378104	this study
87-4 (BC02)	GSM1378110	this study
174-1 (BC03)	GSM1378111	this study
HF70	GSM1378100	this study
Trees (TREEBASE database)		
Trees	16394	this study

[a] Ref. (*9*), [b] ref. (*33*), [c] ref. (*10*).