PLANT SCIENCE

Evolution of flower color pattern through selection on regulatory small RNAs

Desmond Bradley,[1] Ping Xu,[2] Irina-Ioana Mohorianu,[2,3] Annabel Whibley,[1] David Field,[4,5] Hugo Tavares,[1]* Matthew Couchman,[1] Lucy Copsey,[1] Rosemary Carpenter,[1] Miaomiao Li,[6,7] Qun Li,[6] Yongbiao Xue,[6,7,8] Tamas Dalmay,[2]† Enrico Coen[1]†

Small RNAs (sRNAs) regulate genes in plants and animals. Here, we show that population-wide differences in color patterns in snapdragon flowers are caused by an inverted duplication that generates sRNAs. The complexity and size of the transcripts indicate that the duplication represents an intermediate on the pathway to microRNA evolution. The sRNAs repress a pigment biosynthesis gene, creating a yellow highlight at the site of pollinator entry. The inverted duplication exhibits steep clines in allele frequency in a natural hybrid zone, showing that the allele is under selection. Thus, regulatory interactions of evolutionarily recent sRNAs can be acted upon by selection and contribute to the evolution of phenotypic diversity.

A convenient system for studying selection in natural populations is afforded by hybrid zones, where closely related species or populations come into contact (*1*). Such a hybrid zone has been described for two subspecies of *Antirrhinum majus* (snapdragon) that differ in flower color (*2*), a trait involved in pollinator attraction (*3*–*7*). Both subspecies are pollinated by bees but have alternate patterns for guiding flower entry: *A. m. pseudomajus* flowers are magenta, with a patch of yellow highlighting the bee entry point (Fig. 1A), whereas *A. m. striatum* flowers are yellow with magenta veins at the entry point (Fig. 1B). The magenta and yellow flower color intensities show sharp clines at a hybrid zone (*2*) where the subspecies come into contact. Production of magenta is regulated by *ROSEA* (*ROS*) and *ELUTA* (*EL*) (*8*–*10*). *ROS* encodes a MYB-like transcription factor that promotes anthocyanin biosynthetic gene expression in *A. m. pseudomajus* and exhibits a steep cline in allele frequencies at the hybrid zone (*2*, *9*). Distribution of yellow pigment is regulated by *SULF* (Fig. 1, B and C), which represses production of the yellow flavonoid aurone in *A. m. pseudomajus* (Fig. 1D) (*2*, *9*, *10*). Here, we study the molecular nature of *SULF*.

To isolate *SULF*, we first mapped it to an interval of ~3 Mb on chromosome 4 by sequencing pools of *sulf* and *SULF* phenotypes from a segregating population (fig. S1). In parallel, we carried out a transposon mutagenesis experiment in *A. majus* (*SULF*) and isolated a mutant, *sulf-660*, that was both somatically and genetically unstable (fig. S2A and supplementary materials). Comparing the genome sequence of *sulf-660* and its revertants revealed a single insertion site, within the mapped region of *SULF*, specific to *sulf-660*. Three independent revertants had different excision footprints at this site, confirming that the transposon was responsible for the *sulf* phenotype (fig. S2B).

Basic Local Alignment Search Tool searches of the sequence flanking the transposon insertion site revealed regions of 74 to 88% nucleotide sequence identity to *A. majus* chalcone 4′-O-glucosyltransferase (*Am4'CGT*), which encodes an enzyme involved in synthesis of the yellow pigment aurone (Fig. 2A and table S1) (*11*). The regions of *Am4'CGT* homology were organized as an inverted duplication in the *A. majus SULF* genome. Both the left and right arms of the duplication carried deletions relative to intact *Am4'CGT*, suggesting they had independently degenerated from a more complete precursor. A contiguous region of inverted homology between the left and right arms spanned a ~590–base pair (bp) region (red arrows, Fig. 2A), separated by a ~600-bp spacer region, which contained the transposon insertion site of *sulf-660*. Phylogenetic analysis indicated that the *SULF* inverted repeats were likely generated from *Am4'CGT* recently in the evolution of the *Antirrhinum* lineage (Fig. 2B and fig. S3).

To determine whether the inverted duplication at *SULF* might be under selection, we compared *A. m. pseudomajus* and *A. m. striatum* populations sampled from either side of a hybrid zone. Polymerase chain reaction (PCR) using oligos flanking the inverted repeats gave bands in the range 1.5 to 2.5 kb for all individuals from the *A. m. pseudomajus* (n = 96) but not the *A. m. striatum* populations (n = 95), suggesting that the inverted duplication was present at higher frequency in *A. m. pseudomajus* (fig. S4). Sequencing pools of ~50 individuals from each population revealed reduced depth of sequence for *A. m. striatum* compared with *A. m. pseudomajus* over a ~145-kb region around *SULF*, suggesting that *A. m. striatum* carried deletions relative to *A. m. pseudomajus* in this chromosome region (Fig. 2C).

This conclusion was supported by PCR amplification assays using a range of oligos. Deletion alleles were also observed in resequenced individuals, including a 1.3-kb deletion that removed the left arm of the inverted repeat and part of the spacer sequence in *A. m. striatum*. Thus, the inverted duplication present in *SULF* of *A. m. pseudomajus* is absent or at low frequency in *A. m. striatum* populations, further demonstrating the requirement for the inverted duplication for *SULF* function.

Single-nucleotide polymorphisms (SNPs) in a ~300-kb interval containing *SULF* showed steep clines in allele frequency (Fig. 2D and fig. S5) centered at the same geographic location as clines for *ROS* and flower color (*2*). SNPs sampled from other positions along chromosome 4 either showed no clines or showed clines centered at different geographic locations (Fig. 2D and fig. S5). The significance of the clines at *SULF* was confirmed by comparing DNA sequences from pools of individuals sampled from a transect covering ~20 km on either side of the hybrid zone. Of the ~7×10^5 polymorphic SNPs on the *SULF* chromosome, 99% showed no allele frequency differences across the transect, and of those that did, more than 99% did not yield steep clines aligned with *ROS*. Thus, there is likely to be strong selection acting on *SULF*.

The coincidence of the *SULF* and *ROS* clines suggests that these loci interact. In *A. m. pseudomajus*, where *ROS* confers magenta color, *SULF* could be favored because it restricts yellow to create a contrasting highlight at the bee entry point (Fig. 1A). In *A. m. striatum*, where *ros* confers reduced magenta intensity for much of the flower, *sulf* could be favored because it confers both a striking yellow color and a contrasting background to the magenta veins (Fig. 1B). Thus, selection acting on different allele combinations at *SULF* and *ROS* allows alternate floral guides to be maintained on either side of a hybrid zone. The situation is comparable to selection acting on loci controlling yellow and red coloration of mimetic patterns in *Heliconius* butterflies (*12*, *13*).

Given the structure of the inverted duplication at *SULF* and its homology to *Am4'CGT*, we hypothesized that *SULF* represses *Am4'CGT* and thus restricts yellow flower color via regulatory

[1]Department of Cell and Developmental Biology, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK. [2]School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK. [3]School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK. [4]Department of Botany and Biodiversity Research, University of Vienna, Faculty of Life Sciences, Rennweg 14, A-1030 Vienna, Austria. [5]Institute of Science and Technology Austria, Klosterneuburg, Austria. [6]State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, and National Center for Plant Gene Research, Beijing 100101, China. [7]University of Chinese Academy of Sciences, Beijing 100190, China. [8]Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China.
*Present address: Sainsbury Laboratory, University of Cambridge, Cambridge CB2 1LR, UK.
†Corresponding author. Email: enrico.coen@jic.ac.uk (E.C.); t.dalmay@uea.ac.uk (T.D.)

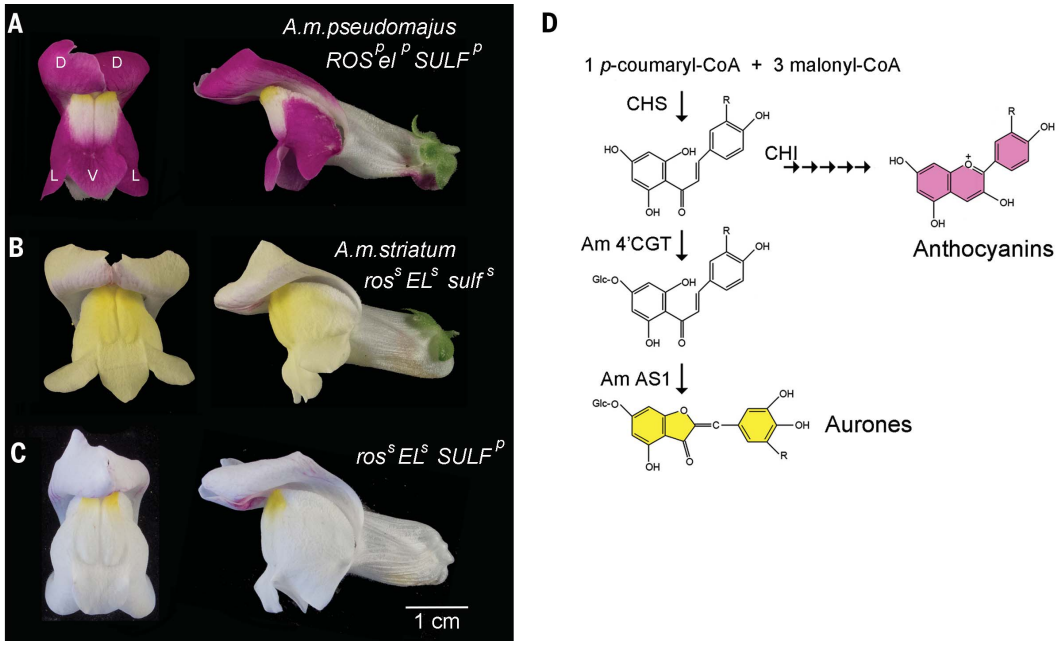

Fig. 1. Flower color pattern phenotypes. Flower face (left) and side (right) views of *A. majus* (*A. m.*) species, showing lower ventral (V), lateral (L), and upper dorsal (D) lobes. Bee vision is sensitive to both yellow and the blue components of magenta reflectance. (**A**) *A. m. pseudomajus*. Magenta with yellow highlight at the bee entry point. (**B**) *A. m. striatum*. Yellow with magenta highlights. (**C**) Flowers from plants with *ros EL* from *A. m. striatum* (ros^S EL^S) and *SULF* from *A. m. pseudomajus* ($SULF^P$). (**D**) Schematic showing the pathways to anthocyanin and aurone pigments. Chalcone synthase, CHS; chalcone isomerase, CHI; *A. m.* chalcone 4′-*O*-glucosyltransferase, Am4′CGT; *A. m.* aureusidin synthase, AmAS1.

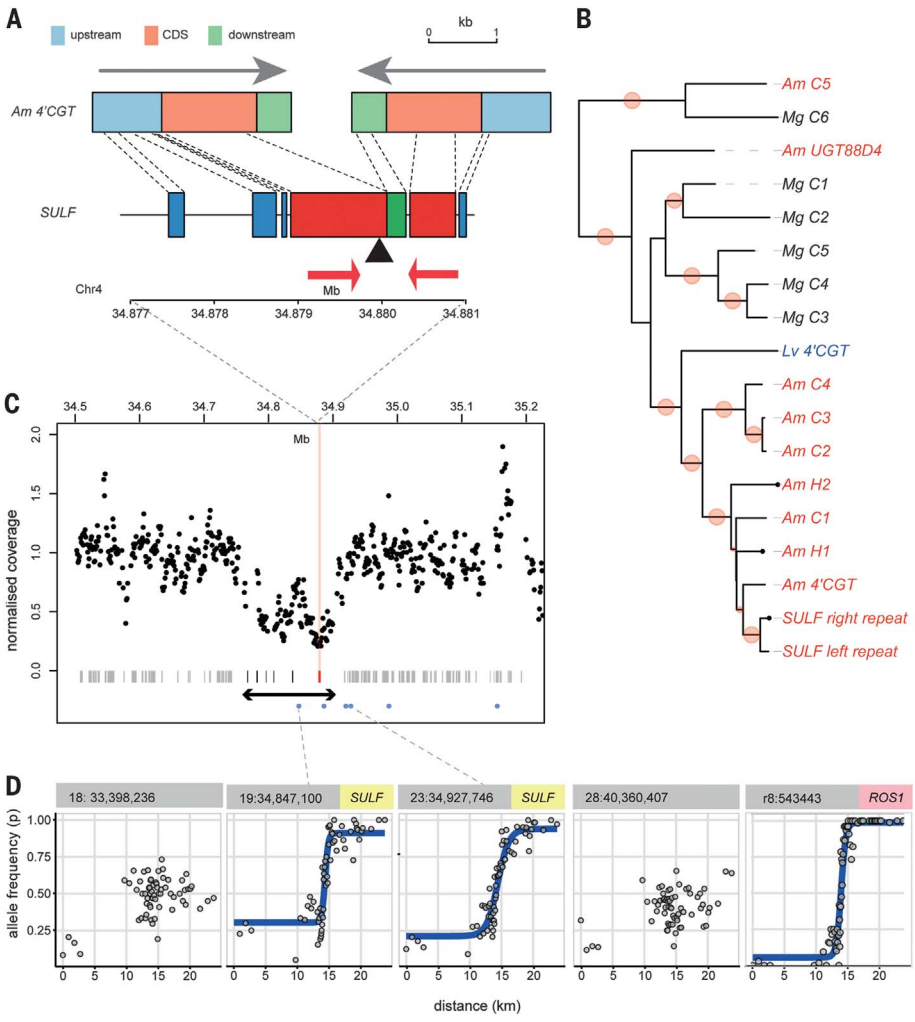

Fig. 2. *SULF* locus shows homology to *Am4′CGT* and signatures of selection. (**A**) *SULF* inverted duplication. Organization of *Am4′CGT* is shown twice (gray arrows) to indicate regions of homology with *SULF* (CDS, coding sequence). The left and right inverted repeats at *SULF* (red arrows) flank the transposon insertion site of *sulf-660* (black triangle). (**B**) Maximum likelihood phylogeny of CGT-related DNA sequences from *Antirrhinum majus* (red), *Mimulus guttatus* (black), and *Linaria vulgaris* (blue). Bootstrap support for nodes with >85% support (red circles, scaled by strength). For extended clade, see fig. S2. (**C**) Plot of *A. m. striatum* sequence coverage normalized against *A. m. pseudomajus* for pools located at either end of the hybrid zone. Bars indicate genes, with *SULF* locus in red. Double-headed arrow shows region underrepresented in *A. m. striatum*. Positions of KASP SNPs used for cline analysis (blue dots). (**D**) Clines for KASP (Kompetitive Allele Specific PCR) markers across the hybrid zone transect. SNP index and chromosome position is indicated above each plot. Markers from *SULF* show steep clines at the hybrid zone, aligned with clines for *ROS1* (right). Markers further away from *SULF* either show no clines (two examples shown) or clines centered at other geographic locations (fig. S4).

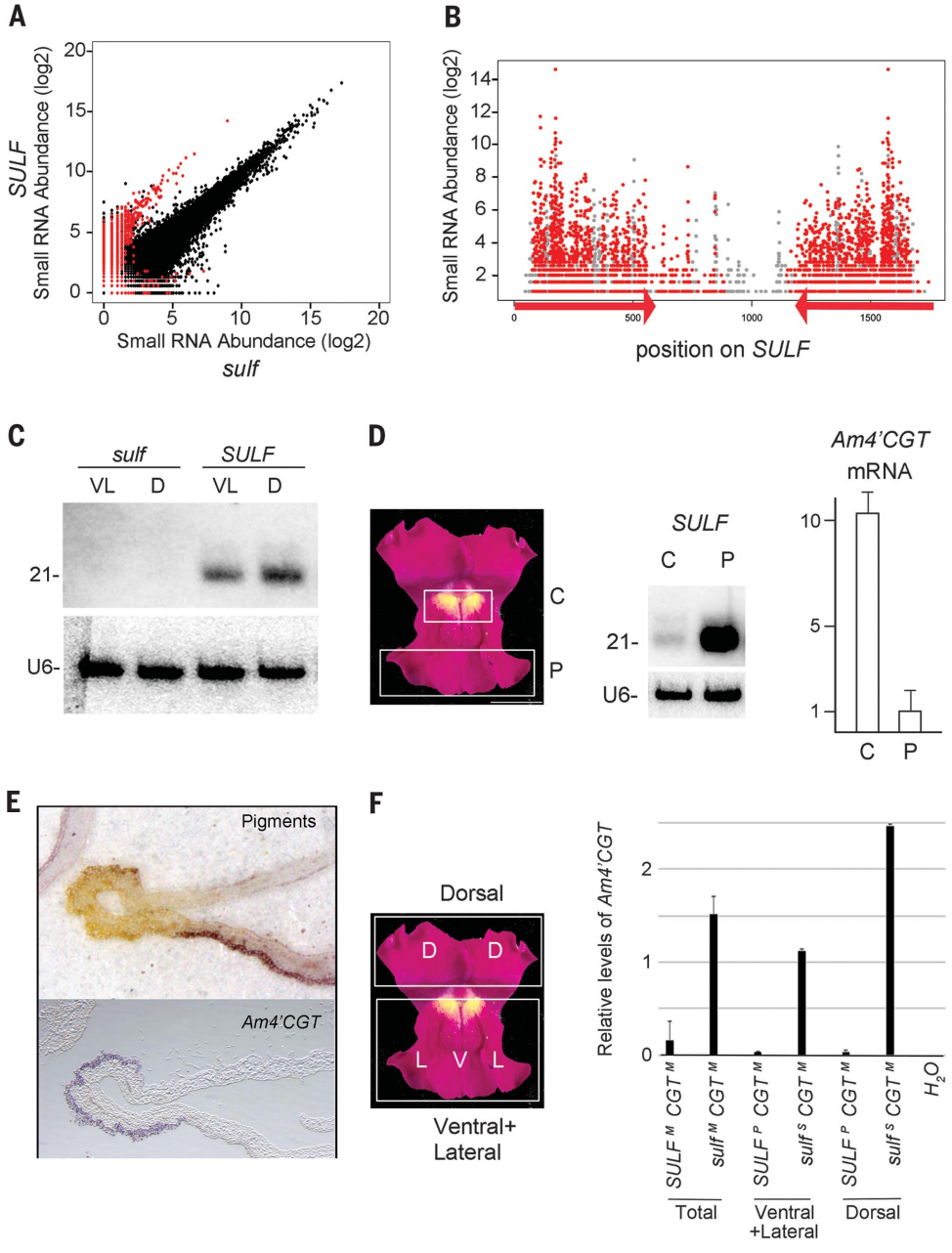

Fig. 3. *SULF* locus makes sRNAs targeting *Am4′CGT*. (**A**) Comparison of total read abundance for sRNAs isolated from libraries of *sulf-660* and *SULF-661*. sRNAs mapping to the *SULF* locus in red. (**B**) Abundance of sRNAs mapping to *SULF* from the *SULF-661* libraries. Reads with potential to target *Am4′CGT* (red) and those unable to target (too many mismatches) (gray). (**C**) Blot of petal RNA probed with an oligo matching one of the abundant 21–nucleotide oligomers, showing signal in ventral and lateral (VL) or dorsal (D) petals in *SULF-661* but not *sulf-660*. U6, ubiquitin control. (**D**) Complementary expression pattern of *SULF* sRNAs and *Am4′CGT* expression. Petals (left) were dissected into a central (C) yellow region and a peripheral (P) nonyellow region. For *SULF* expression, sRNA blots were probed with *SULF*, revealing stronger expression in the peripheral compared to the central region (middle). For *Am4′CGT*, RNA was subject to quantitative real-time PCR (qRT-PCR), showing lower expression in the peripheral region (right). (**E**) Floral bud of *A. majus* was sectioned to reveal the pigments (top), and similar sections were probed to reveal, by in situ RNA hybridization, the expression pattern of *Am4′CGT* (purple stain, bottom). (**F**) qRT-PCR on petal RNA (total or dissected into upper and lower regions). Expression of *Am4′CGT* is reduced in genotypes carrying *SULF* from *A. majus* ($SULF^M$) or *A. m. pseudomajus* ($SULF^P$) compared with those carrying *sulf* from *A. majus* ($sulf^M$) or *A. m. striatum* ($sulf^S$). Standard errors were calculated from the means of three independent biological samples, each analyzed in triplicate.

small RNAs (sRNAs). To determine whether *SULF* generated sRNAs, sRNA libraries were prepared from petals of *A. majus SULF* and *sulf-660*. The biggest differences in sRNA abundance mapped to the *SULF* inverted repeats and corresponded to predominantly 21–nucleotide oligomers (Fig. 3, A and B). RNA blots probed with *SULF* confirmed that sRNAs from the inverted repeat were present in *SULF* and absent in *sulf* genotypes, including *A. m. striatum* (Fig. 3C and fig. S6). The sRNAs likely derive from processed transcripts predicted to generate long-foldback hairpin RNAs (fig. S7).

If the sRNAs generated by *SULF* restrict yellow pigmentation by targeting *Am4′CGT*, then *SULF* and *Am4′CGT* should exhibit complementary expression patterns. Analysis of RNA extracted from yellow and nonyellow regions of the petals of *A. majus* showed that *SULF* was preferentially expressed in the nonyellow region, whereas *Am4′CGT* was mainly expressed in the yellow region (Fig. 3D). The spatial restriction of *Am4′CGT* was confirmed by RNA in situ hybridization (Fig. 3E).

Overall expression of *Am4′CGT* was lower in petals of *SULF* compared with *sulf-660* (Fig. 3F). 5′ RACE (rapid amplification of cDNA ends) on *SULF* genotypes revealed products for *Am4′CGT* terminating at a range of positions, suggesting cleavage at multiple sites (fig. S8). No clear bands at the size expected for cleavage products were found in *sulf*. The lack of a single cleavage site in *SULF* genotypes is consistent with the *SULF* inverted duplication generating multiple sRNAs targeting *Am4′CGT* (Fig. 3B). To determine whether *SULF* alleles from the subspecies also varied in their ability to repress *Am4′CGT*, we introgressed *SULF* from *A. m. pseudomajus* ($SULF^P$) or *A. m. striatum* ($sulf^s$) into an *A. majus* background with the same *Am4′CGT* target allele. *Am4′CGT* expression was reduced in both dorsal and ventral petals of $SULF^P$ compared with $sulf^s$ (Fig. 3F). Thus, *SULF* acts by repressing transcript levels of the target *Am4′CGT* gene in *A. m. pseudomajus* but not in *A. m. striatum*.

If selection on inverted duplications is a common mechanism for establishing regulatory interactions, we might expect the genome to contain a large number of inverted duplications similar to *SULF*. Scanning the *A. majus*

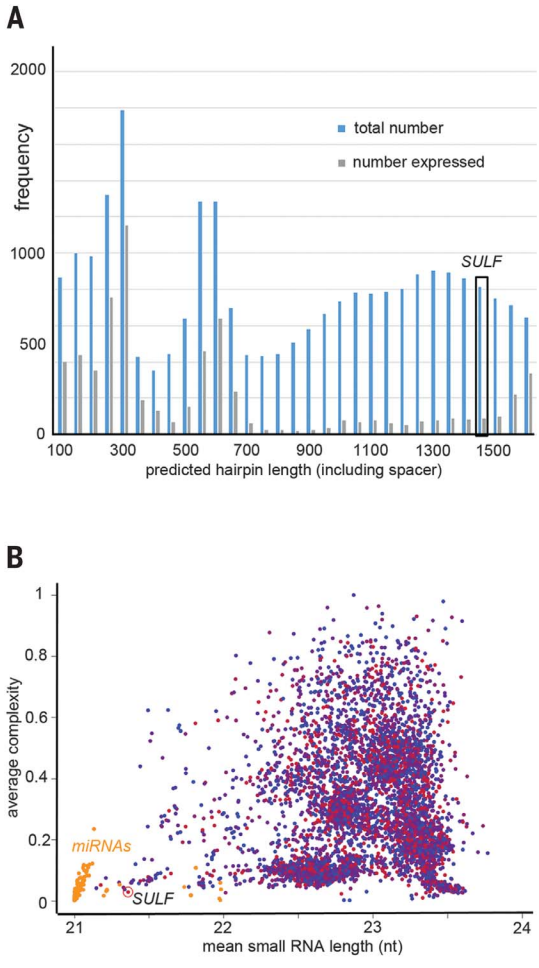

Fig. 4. Expression and frequency distribution of inverted repeats and microRNA genes in *Antirrhinum majus*. (**A**) Frequency and expression levels of inverted repeats with folding energies similar to *SULF*, as a function of length of predicted hairpin RNA (including spacer). An inverted repeat is considered expressed if the maximum overall abundance of incident sRNAs in any library is above a noise threshold (*20*). Boxed region shows class to which *SULF* belongs. (**B**) Average complexity and mean length of sRNAs mapping to inverted repeats [as in (A)] and microRNA hairpins. Each point corresponds to a predicted transcript with a hairpin-like structure. *SULF* hairpin is circled in red. Only sRNAs in the 21- to 24-nt range are considered. Average complexity is the number of different reads (unique) divided by the total number of reads mapping to the hairpin (*29*). Although *SULF* generates sRNAs throughout the inverted repeats, the high abundance of some leads to a low overall complexity. For inverted repeats, transcript abundance is color coded on a log scale and varies from blue (low abundance, 20) to red (high abundance, 160,000). Orange indicates microRNA hairpins.

genome for inverted duplications with a similar adjusted folding energy to *SULF* revealed many such regions, some of which generated sRNAs (Fig. 4A). However, most of these sRNAs were >21 nucleotides (nt) long, unlike those generated by *SULF* (circled, Fig. 4B), which were ~21 nt. Moreover, the sRNA population generated by *SULF* was of relatively low complexity (ratio of the number of unique reads to total reads) because of the high abundance of a subset of sRNAs. Based on size and complexity, the profile of sRNAs generated by *SULF* was similar to that of conserved microRNA loci (orange spots, Fig. 4B). Given that the *SULF* hairpin is about five times as long as a typical conserved microRNA hairpin, these findings suggest that *SULF* generates a functioning long regulatory hairpin RNA.

If only a subset of sRNAs generated by *SULF* are required to inhibit target gene activity, selection would not be able to maintain homology with the target gene *Am4'CGT* over the extended length observed (590 bp). This argument implies that *SULF* is of recent evolutionary origin, consistent with the phylogenetic analysis (Fig. 2B). With respect to its young age, *SULF* is similar to other inverted duplications with extended similarity to protein coding regions that encode sRNAs (*14*–*17*). Over evolutionary time, functional inverted duplications such as *SULF* might be lost, maintained, or become shorter microRNA hairpins (*14*, *15*, *18*–*21*). The deletions observed in both the left and right arms of the inverted repeat at *SULF*, relative to *Am4'CGT* (Fig. 2A), suggest that the process of size reduction may have already occurred to some extent.

Among the many documented cases of loci contributing to natural variation (*22*), several examples of small regulatory RNAs have been described (*23*–*26*). However, these examples involve changes in expression pattern of preexisting microRNAs or creation of new target sites, rather than de novo generation of a small regulatory RNA, as observed with *SULF*. The unusual nature of *SULF* may be a matter of chance or may reflect constraints on regulatory mechanisms (*27*). For example, the biosynthetic pathway to yellow aurone pigment synthesis has fewer steps and has a more limited taxonomic distribution than the magenta anthocyanin pigment synthesis pathway (*11*, *28*). Variation in transcription factors, such as *ROS*, may therefore not be available specifically to modulate yellow patterning. Inverted duplications that generate regulatory RNAs may thus provide a flexible mechanism, complementing that based on transcription factor or cis-regulatory variation (*22*), for modulating or creating novel expression patterns upon which natural selection may act to generate evolutionary change.

REFERENCES AND NOTES

1. N. H. Barton, G. M. Hewitt, *Nature* **341**, 497–503 (1989).
2. A. C. Whibley *et al.*, *Science* **313**, 963–966 (2006).
3. R. Hopkins, M. D. Rausher, *Nature* **469**, 411–414 (2011).
4. H. D. Bradshaw Jr., D. W. Schemske, *Nature* **426**, 176–178 (2003).
5. Y. W. Yuan, K. J. Byers, H. D. Bradshaw Jr., *Curr. Opin. Plant Biol.* **16**, 422–428 (2013).
6. H. Sheehan *et al.*, *Nat. Genet.* **48**, 159–166 (2016).
7. M. T. Clegg, M. L. Durbin, *Nat. Rev. Genet.* **4**, 206–215 (2003).
8. J. Hackbarth, P. Michaelis, G. Scheller, *Z. Indukt. Abstamm. Vererbungsl.* **80**, 1 (1942).
9. K. Schwinn *et al.*, *Plant Cell* **18**, 831–851 (2006).
10. H. Stubbe, *Genetik und Zytologie von Antirrhinum L. sect. Antirrhinum* (Veb Gustav Fischer Verlag, Jena, Germany, 1966).
11. E. Ono *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11075–11080 (2006).
12. C. D. Jiggins, *The Ecology and Evolution of Heliconius Butterflies* (Oxford Univ. Press, 2017).
13. N. J. Nadeau, *Curr. Opin. Insect Sci.* **17**, 24–31 (2016).
14. N. Fahlgren *et al.*, *Plant Cell* **22**, 1074–1089 (2010).
15. E. Allen *et al.*, *Nat. Genet.* **36**, 1282–1290 (2004).
16. M. J. Axtell, *Biochim. Biophys. Acta* **1779**, 725–734 (2008).
17. O. Voinnet, *Cell* **136**, 669–687 (2009).
18. J. Cui, C. You, X. Chen, *Curr. Opin. Plant Biol.* **35**, 61–67 (2017).
19. J. Piriyapongsa, I. K. Jordan, *PLOS ONE* **2**, e203 (2007).
20. M. Nozawa, S. Miura, M. Nei, *Genome Biol. Evol.* **4**, 230–239 (2012).
21. F. Borges, R. A. Martienssen, *Nat. Rev. Mol. Cell Biol.* **16**, 727–741 (2015).
22. A. Martin, V. Orgogozo, *Evolution* **67**, 1235–1250 (2013).
23. S. Arif *et al.*, *Curr. Biol.* **23**, 523–528 (2013).
24. A. Clop *et al.*, *Nat. Genet.* **38**, 813–818 (2006).
25. S. K. Nair *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 490–495 (2010).
26. J. M. Debernardi, H. Lin, G. Chuck, J. D. Faris, J. Dubcovsky, *Development* **144**, 1966–1975 (2017).
27. K. Chen, N. Rajewsky, *Nat. Rev. Genet.* **8**, 93–103 (2007).
28. Y. Tanaka, N. Sasaki, A. Ohmiya, *Plant J.* **54**, 733–749 (2008).
29. I. Mohorianu *et al.*, *Plant J.* **67**, 232–246 (2011).

ACKNOWLEDGMENTS

The sRNA sequencing data presented in this study is publicly available on Gene Expression Omnibus 56 under accession number GSE91378. Data sets for genomic DNAs are available at the European Nucleotide Archive, accession number PRJEB22668, and scripts at linked sites. The authors have no competing interests. We thank M.-E. Mannarelli for technical support, N. Barton for suggestions on the manuscript, and A. Rebocho for helpful discussions. This work was supported by the U.K. Biotechnology and Biological Sciences Research Council grant BB/G009325/1, awarded to E.C. H.T. was supported by a Ph.D. scholarship from the Portuguese Science Foundation (FCT) (SFRH/BD/60982/2009), through the European Social Fund (Programa Operacional Potencial Humano program). The supplementary materials contain additional data. For further background, see antspec.org.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/358/6365/925/suppl/DC1
Materials and Methods
Figs. S1 to S8
Table S1
References (*30*–*59*)

11 July 2017; accepted 6 October 2017
10.1126/science.aao3526

Science

Evolution of flower color pattern through selection on regulatory small RNAs

Desmond Bradley, Ping Xu, Irina-Ioana Mohorianu, Annabel Whibley, David Field, Hugo Tavares, Matthew Couchman, Lucy Copsey, Rosemary Carpenter, Miaomiao Li, Qun Li, Yongbiao Xue, Tamas Dalmay and Enrico Coen

Science **358** (6365), 925-928.
DOI: 10.1126/science.aao3526

How the snapdragon chooses its color

In some snapdragons, a yellow spot in a field of magenta shows the bee the best place to go. Flowers of a related subspecies are mainly yellow with magenta veins marking the target. Bradley *et al.* analyzed a locus that regulates the pattern of color. The locus contains an inverted gene duplication that encodes small RNAs that repress pigment biosynthesis. Analysis of flowers derived from a region of the Pyrenees where the subspecies coexist indicates that natural selection is operating upon the locus.

Science, this issue p. 925