Taming the dragon: genetic variation in wild and domesticated *Antirrhinum majus*

Hilde Schneemann[1]*, Apollonia Palmer[2], David L. Field[3], Nick Barton[1], Sean Stankowski[5]*

[1]ISTA Austria, Am Campus 1, 3400 Klosterneuburg, Austria
[2]Department of Ecology and Evolution, The University of Sussex, Brighton, United Kingdom
[3]Applied BioSciences, Macquarie University, Sydney, New South Wales, Australia
[4]John Innes Centre, Norwich, United Kingdom
[5]Department of Genetics, Evolution and Environment, University College London, United Kingdom

*Corresponding authors
hilde.schneemann@ist.ac.at; s.stankowski@ucl.ac.uk

Running title: genetics of wild and domesticated snapdragons

Keywords: artificial selection, flower colour, horticultural varieties, snapdragon

Abstract

Domesticated species and their wild relatives provide powerful case studies for examining the processes that shape rapid diversification. Here, we conduct a population genetic analysis of 31 domesticated varieties and 33 natural populations of the common snapdragon, *Antirrhinum majus*, a species in which closely related flower colour varieties form hybrid zones and display extensive variation in floral colour and patterning. We find that domesticated snapdragons form a single genetic group, suggesting that these lines were derived from the same set of ancestral founders. Despite their strikingly diverse flowers, domesticated snapdragons show substantially lower genetic diversity compared to wild populations. Genetic markers strongly associated with known colour loci in wild populations show only weak associations across domesticated lines. This may be because recombination during selective breeding has weakened associations between casual mutations and linked markers, and/or because different mutations control colour variation in domesticated snapdragons.

Introduction

Domesticated species and their wild relatives provide powerful case studies for examining the processes that shape rapid diversification (Ross-Ibarra et al., 2007). Domestication often causes strong reductions in genetic variation owing to strong founder effects and increased rates of drift and inbreeding over many generations of artificial selection. Yet, some domesticated species are characterised by high levels of phenotypic diversity, often exceeding levels of variation found in nature. While this diversity has been shaped by strong diversifying selection targeting heritable traits of human interest, its origins remain less certain: to what extent does it arise during domestication itself, and how much derives from standing variation present in ancestral populations?

Most insights into the genetics of domestication come from major crops and livestock, where selection has targeted traits related to yield, growth, and reproduction. Because these traits are often highly polygenic, the identification of the loci that underpin variation under domestication is exceedingly difficult (Rockman, 2012). In contrast, ornamental plants and animals have been shaped primarily through selection on conspicuous aesthetic traits like colouration, which often has a relatively simple genetic architecture consisting of relatively few loci of large effect (Cieslak et al., 2011; Yang et al., 2025). Despite their historical importance in the development of genetics, ornamental systems remain comparatively understudied from a population genetic perspective, especially in the context of variation present in their wild relatives.

The common snapdragon, *Antirrhinum majus*, provides an excellent system for studying variation both under domestication and in nature (Fig. 1). Native to the western Mediterranean, *A. majus* has a long history of association with humans, with evidence suggesting it was grown in ancient Roman gardens and subsequently cultivated across Europe from at least the eighteenth century (Hudson et al., 2008; Schwarz-Sommer et al., 2003). Early horticultural descriptions emphasise both the remarkable diversity within cultivated forms and the tendency of varieties not to reproduce true from seed, with novel phenotypes frequently arising following sowing and desirable forms maintained through cuttings (e.g. Vilmorin-Andrieux & Cie, 1885; Stubbe, 1966). Subsequent breeding by commercial and amateur horticulturists across Europe, North America, and Asia has produced modern domesticated snapdragons that together encompass substantial phenotypic diversity, including variation in flower colour and pattern, floral morphology, flowering time, and growth form, with distinct varieties adapted for dwarf habit or for tall, long-stemmed cut flower production. The domestication process has also resulted in the loss of self-incompatibility, which prevents self-pollination in wild populations (Li et al., 2019; Surendranadh et al., 2022). This has presumably been favoured in breeding programs as autonomous selfing reduces the need for controlled pollination and also eases the stabilisation of desirable phenotypes.

Figure 1. Floral diversity in wild and domesticated snapdragons. (a & b) The two varieties of *A. majus*, subspecies *majus* show a striking difference in their flower colour. (c - i) Natural hybridization between the magenta- flowered var. *pseudomajus* and yellow-flowered *striatum* produces a wide range of colour phenotypes owing to the segregation and recombination of alleles at 7 loci that affect the expression of anthocyanin (magenta) and aurone (yellow) pigments in floral tissue. Colour variation in domesticated snapdragons is equally striking, with some lines exhibiting patterns and intensities not typical of the phenotypes found in either variety, or in natural hybrid zones (e.g., L11, L30) (see Fig. S1 for multiple photos of each line). Also note the variation in floral morphology in domesticated lines, including peloric flowers (L7 L10, L41), which are radially symmetric, and peloric double flowered mutants (L42, L46), contrasting with the bilaterally symmetrical wild-type morphology. Photo credits: a and b by Mitch Olsen, c by David Field, d to i by Daria Shipilina.

In parallel, *A. majus* has served as an important model for understanding the genetic basis of phenotypic variation in both laboratory and natural populations (Hudson, Critchley, and Erasmus 2008). In the wild, closely related taxa within *A. majus* display striking polymorphism in floral colour and patterning (Fig. 1). In particular, *A. m. m.* var. *pseudomajus* and var. *striatum*, which are native to France and Spain, differ primarily in the distribution of magenta anthocyanin and yellow aurone pigments in the corolla tissue (Whibley et al., 2006). Detailed molecular genetic studies have shown that the colour differences between the varieties are caused by at least 7 loci of moderate to large effect, including *Rosea*, *Eluta*, and *Rubia*, which control anthocyanin production (Field et al., 2025; Tavares et al., 2018), and *Sulfurea*, *Cremosa*, *Aurina*, and *Flavia*, which control aurone production (Bradley et al., 2017, 2025; Richardson et al., 2025). Detailed studies of hybrid zones between these taxa, primarily the hybrid zone in the Val de Ribes near the town Planoles, have shown that spatially varying

selection maintains sharp transitions in flower colour over short geographic distances, with strong differentiation at pigmentation loci but relatively low divergence across much of the genome (Pal et al., 2025; Surendranadh et al., 2025; Tavares et al., 2018; Whibley et al., 2006). However, frequent hybridization between the varieties leads to the shuffling of causal alleles, producing a diverse range of colour patterns, many of which are similar to those observed in domesticated varieties (Fig. 1). Together, these findings suggest that much of the colour variation in domesticated varieties could be underpinned by pre-existing genetic variation.

The combination of extensive variation in nature, the well-characterised genetic basis of flower colour, and large range of phenotypic variation in horticultural varieties make *A. majus* an attractive species for investigating the genetic changes that accompany domestication, and the relationship between variation in natural and domesticated plants. Here, we analyse patterns of genetic variation in 31 domesticated varieties of *A. majus* and 33 natural populations from France and Spain. Using a panel of genetic markers we ask: (i) is there evidence that horticultural varieties have multiple independent origins, or do they represent diversification from a single domestication event? (ii) How has the domestication history of *A. majus* shaped patterns of genetic diversity? and (iii) How much of the colour variation in domesticated varieties is controlled by loci that are known to affect colour in nature?

Methods

Plant material

We obtained leaf material from 289 individuals across 33 populations of *A. majus ssp. majus* from southern France and northern Spain with an emphasis on the Spanish Pyrenees where the geographic ranges of the varieties come into close proximity (See Fig. S2 for a map; Table S1). This included 17 populations of magenta-flowered var. *pseudomajus*, 15 populations of yellow-flowered var. *striatum*, and 1 population from the centre of a well-studied hybrid zone near the town of Planoles where flower colour is variable. A small amount of leaf tissue was taken from each plant, placed in paper envelopes and desiccated in a bag of silica gel to preserve DNA.

We sourced seeds for 31 domesticated *Antirrhinum majus* lines, representing broad phenotypic variation (Fig 1; Fig. S2; Table S2). Seeds were placed in a refrigerator at 4°C for two weeks to simulate winter dormancy. Plants were grown in a randomised block design, consisting of 10 trays, with each tray containing 1 individual of each line. Seeds were sown on the surface of 8cm pots containing 3:1 mix of Melcourt peat-free multi-purpose potting soil and Perlite. Following germination, seedlings were thinned to one individual per pot. Plants were grown in the University of Sussex greenhouse at 24°C under a 06:00-21:00 light cycle and bottom-watered as needed. Plants were fertilised every 2 weeks using a ¼ strength preparation of MiracleGrow All Purpose Soluble Plant Food. Leaf material was collected and stored as described above.

Photography and scoring of flower colour

We took photos of 1 flower from 5 individuals of each line under standard conditions using a Nikon Coolpix S7000 16.0MP set to 20x Optical Zoom and macro mode, with an exposure of -1.3. Photos were taken in a Lightdow Light Box (30 cm) on a black fabric background.

We also quantified the colouration of 1 flower from 5 plants per domesticated line using a quantitative scoring system, described in Surendranadh et al. (2025), to separately quantify the intensity of magenta anthocyanin pigment and yellow aurone pigment in the corolla tissue. In nature, the magenta score ranges from 0 (no magenta) to 5 (intense magenta) while the yellow score ranges from 0 (no yellow) to 3 (intense yellow). Because some domesticated plants show more intense pigmentation than wild ones, we increased the range up to 6 for magenta and 4 for yellow.

SNP genotyping and preparation of genetic data

Samples were genotyped at a panel of 103 biallelic SNP markers which were previously developed as described in Ringbauer et al. (2018). The panel included 30 SNP markers that were selected because their alternative alleles are in strong linkage disequilibrium with causal SNPs that contribute to colour differences in natural populations. Hereafter we refer to these as 'colour associated SNPs'. This included 1 SNP in LD with the *Cremosa* locus, 11 SNPs associated with *Flavia* (spread over 2 centimorgans), 3 loci associated with *Sulfurea* (all at the same map position), and 14 SNPs associated with the linked loci *Rosea* and *Eluta* (spread over 1 cM). The remaining 73 SNPs are not associated with flower colour and were instead selected for their high levels of within-population diversity. Dried leaf tissue was sent to LGC Genomics who performed DNA extractions and genotyping using Kompetitive Allele Specific PCR (KASP). Details of each marker and information needed to genotype them can be found on GitHub (see data accessibility statement).

Genotypes were initially encoded as 0 and 2 for the alternative homozygotes and 1 for the heterozygote. Alleles at each locus were then polarised based on whether allele 1 was more common in var. *pseudomajus* or var. *striatum*. This was done by pooling individuals from the 17 *pseudomajus* (ps) populations and 15 *striatum* (st) populations, calculating the frequency of allele 1 in each variety, and then finding the difference in frequency as $\Delta p = p_{ps} - p_{st}$. For loci where Δp was negative, we flipped the identity of the alternative homozygotes so that the allele more common in *pseudomajus* is always encoded as homozygous state 2.

Analysis of genetic divergence and diversity

We used a combination of model-based and model-free methods to explore the relationships between the sequenced samples based on the 73 SNPs not causally associated with flower colour. For the model-based analysis, we used the program *Admixture* (Alexander et al., 2009) We tested a range of values of *K* between 2 and 30, with 10 replicate runs conducted for each K to ensure convergence.

We used 2 different model free approaches to study divergence. First, for the 73 SNPs not associated with flower colour, we calculated the pairwise genetic distance, π, between all pairs of individuals according to Nei (1987). π measures the expected probability of mismatch between alleles randomly drawn from samples *x* and *y* (here, pairs of individuals) so is analogous to π_w when calculated within populations and lines, and π_b (or d_{xy}) when calculated between them. We then used the *R* package *ggtree* to build an unrooted neighbour-joining tree from the matrix of pairwise distances. Second, we conducted a principal components analysis (PCA) on the genotype matrix using the *prcomp* function in *R*. As PCA does not allow missing data, we imputed missing genotypes by sampling a genotype from another randomly selected individual from the same wild population or domesticated line.

To evaluate genetic diversity at the individual level, we calculated the per-sample heterozygosity (H_i), which is the proportion of heterozygous sites across an individual's multilocus genotype. We used the *R* package *hierfstat* (Goudet, 2005) to calculate mean inbreeding coefficient, F_{IS}, for each population and line over all loci from the averaged observed heterozygosity (H_o) and corrected expected heterozygosity (H_s) for each wild population/line as described by Nei (1987). For these analyses, we used all 103 SNPs.

Genetic differentiation and genotype-phenotype associations at known colour loci

We used *hierfstat* (Goudet, 2005) to calculate Weir and Cockerham F_{ST} for each of the 103 SNPs between the wild individuals of var. *pseudomajus* and var. *striatum*. This was done by pooling individuals across all populations of the same variety.

We used Spearman-rank correlation to test for associations between our magenta and yellow scores and the genotypes at the 30 markers that are linked to five of the known colour loci.This analysis was conducted separately for wild and domesticated individuals. For the wild samples, we obtained colour scores for a random sample of 289 wild individuals (i.e., the same

number of samples as in the domesticated set) from the hybrid zone in Planoles where the colour loci segregate and recombine in hybrids.

Results

Relationships between between wild and domesticated plants

All three analyses–including *Admixture*, PCA, and analyses of genetic distance revealed a clear, consistent picture of the relationships between the wild and domesticated snapdragons. In the *Admixture* analysis, wild and domesticated samples showed strong separation at $K = 2$, with wild individuals being primarily assigned to K1 and domesticated samples were assigned to K2. The same patterns were observed in the PCA, where wild and domesticated lines showed clear separation on PC1 (Fig. S5). Higher values of K revealed additional structure within the wild or domesticated groups (Fig. 2A; Fig S3). Within the wild samples, we did not observe a clear subdivision of plants by flower colour variety, with structure instead being driven primarily by geography (Fig. S4). This finding is consistent with other recent work (Richardson et al. 2025; Pal et al. 2025).

Wild and domesticated samples also formed distinct groups in a neighbour-joining tree based on pairwise genetic distances (π) (Figure 2B). Specifically, all domesticated individuals formed a single group that was clearly separated from the wild samples. Within the domesticated clade, individuals from the same line tended to cluster together, though the degree of clustering varied among lines (Fig. S6). Distributions of pairwise π, clearly show that the distances between domesticated and wild plants (mean $\pi_b = 0.5$) are much greater than distances between plants from different domesticated lines ($\pi_b = 0.18$), yet similar to the distances observed between plants from different wild populations ($\pi_b = 0.44$) (Fig. 2C).

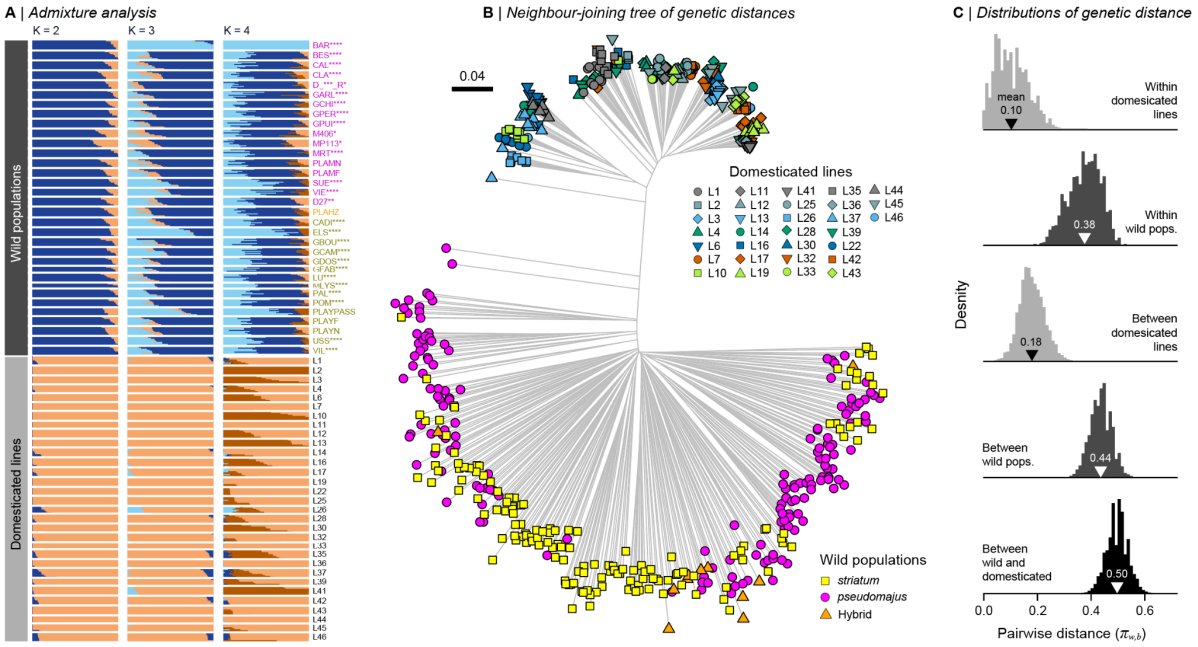

Figure 2. Genetic divergence between wild and domesticated snapdragons. (A) *Admixture* plots showing inferred structure assuming $K = 2$, *3*, and 4 clusters. Each horizontal bar represents 1 individual, and the fraction of each colour indicates the proportion of its genome inferred to derive from each cluster. (B) Neighbour-joining tree based on estimates of genetic distance (π) calculated between all pairs of individuals. Different symbol and colour combinations distinguish individuals from the domesticated lines, wild populations of *pseudomajus* and *striatum*, and the hybrid population. (C) Distributions of pairwise genetic distance (π) between all pairs of individuals within each of the domesticated lines ($n = 1213$), within each of the populations ($n = 1396$), between individuals from different lines ($n = 40{,}403$), between individuals from different populations ($n = 49{,}325$), and between wild and domesticated individuals ($n = 92{,}191$). The mean pairwise distance is shown on each plot. All analyses are based on the 73 SNPs not causally associated with flower colour.

Diversity within natural populations and domesticated lines

Diversity within wild populations was 3.8 times higher than in the domesticated lines (mean $\pi_w$wild = 0.38 v.s. mean $\pi_w$domesticated = 0.1) (Fig. 2c). This striking difference was especially clear from the visual inspection of the genotype matrix (Fig. 3A) where many markers are near fixation for a single allele across the domesticated samples.

Levels of per-sample heterozygosity H_i, calculated across all 103 markers, are also lower in domesticated lines, ranging from near 0 (e.g., L2, L41) to 0.16 (mean 0.08) (Fig. 1B), compared with 0.22 (BAR) to 0.44 (Planoles HZ) in wild populations (Fig. 3B; Table S3). On average, populations of magenta *pseudomajus* tended to have slightly lower H_i than yellow *striatum* populations, regardless of whether or not colour loci were included (mean $H_i$ps = 0.36 vs. H_ist 0.38 without selected loci; mean $H_i$ps = 0.31 vs. H_ist 0.34 with all loci) (Fig. 1B). However, this difference was not statistically significant (Wilcoxon rank-sum tests $p > 0.05$).

Across the wild populations, F_{IS} values were centred around 0 (mean = 0.019 ± 0.04), ranging from -0.07 (pop. GBOU) to 0.11 (pop. LU) (Fig. 3C; Table S4). In contrast, F_{IS} showed strong deviations from 0 in many domesticated lines, with a mean absolute value of F_{IS} of 0.37 ± 0.27. Despite a tendency of values to be positive, consistent with a deficit of heterozygotes within lines, 4 lines–all of which were documented F_1 horticultural varieties–showed highly negative F_{IS}, suggesting a strong excess of heterozygosity at polymorphic sites (L11, L22, L33 & 36). This can be seen in the plot of the genotype matrix (Fig. 3A) where some loci are heterozygous in most or all individuals.

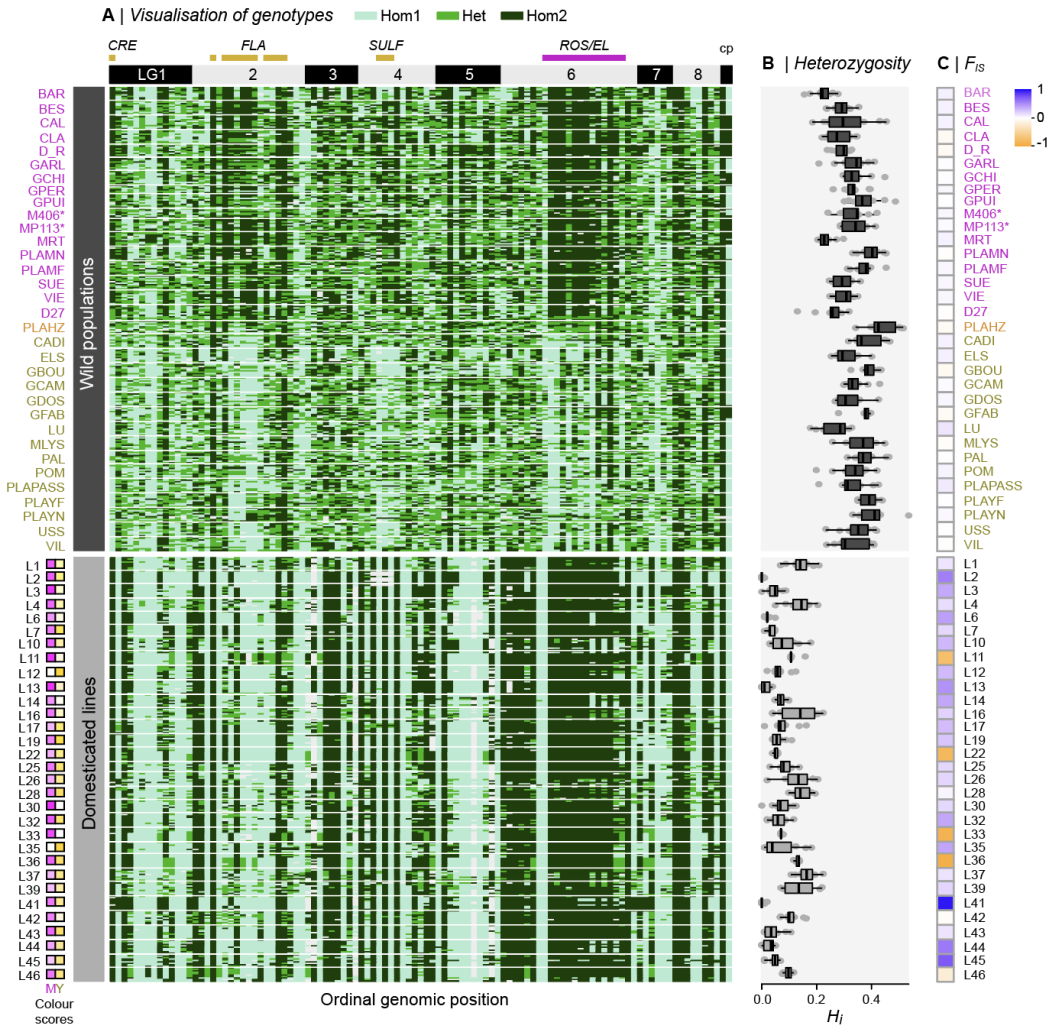

Figure 3. Genetic diversity in wild and domesticated snapdragons. (A) Visualisation of genotypes for 578 individuals across 33 natural populations and 39 domesticated lines. Each column represents one of the 103 SNP

markers, ordered by their position along the 8 linkage groups. The location of the 30 markers associated with the known colour loci, *CRE*, *FLA*, *SULF*, *ROS* and *EL* are shown at the top of the plot. Each row represents the multilocus genotype for one individual. The three shades of green represent the possible genotypes at each site (light grey represents missing data). Alleles are polarised so that the allele more common in var. *pseudomajus* than in var. *striatum* the dark green homozygote. White lines are boundaries between individuals from different populations/lines. A visual representation of the mean flower colour scores for magenta (M) and yellow pigment (Y) is shown next to the line ID. (B) Levels of per individual heterozygosity (H_i) in wild and domesticated snapdragons. The boxplots show the distribution of H_i for each population/line. (C) Mean F_{IS} over all loci for each wild population/line. The colour of the box indicates the value of F_{IS}, but see Table S4 for values for each line.

Genetic differentiation and genotype-phenotype associations at known colour loci

As expected we found strong genetic differentiation at markers previously shown to be in strong LD with known colour loci when wild *pseudomajus* and *striatum* were compared (Fig. 4A; also see fig. 3A). The highest F_{ST} value was 0.93 near *ROS*, with substantial values observed at *CRE* (0.76), *EL* (0.72), *FLA* (0.70), and *SULF* (0.44) as well (Fig. 4A; Table S5). In contrast, non-colour loci generally showed very low F_{ST} (mean 0.02 ± s.d. 0.06), with the exception of one marker near *SULF*, which showed a substantial F_{ST} of 0.53.

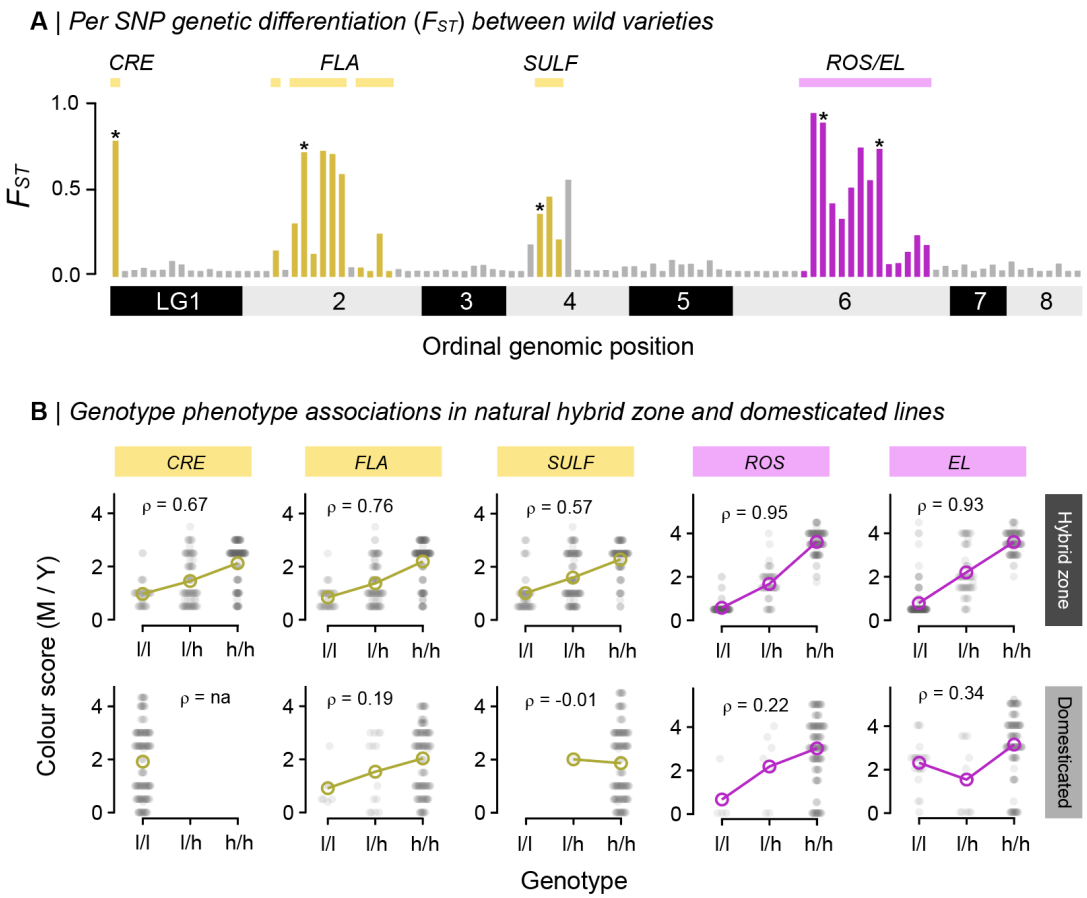

Figure 4. Genetic differentiation between wild flower-colour varieties of *A. m. majus* and genotype-phenotype associations for wild and domesticated plants at markers linked to 5 known flower colour loci. (A) Weir and Cockerham F_{ST} for each of the 103 SNPs. SNPs tightly linked to colour loci are coloured yellow or magenta. Asterisks mark the SNPs shown in part B. (B) Each plot shows the relationship between the 3 genotypes and relevant colour score for 5 of the colour-associated SNPs. The top row shows the relationship for 289 individuals from the natural hybrid zone at planoles, while the bottom row shows the relationship for the 289 domesticated samples. The alleles have been relabelled according to whether they are associated with low (l) or high (h) levels of pigmentation. The spearman rank correlation (ρ) between genotype and phenotype is indicated in each plot.

We did, however, find clear differences in the strength of genotype-phenotype associations in wild and domesticated samples. For the samples from the hybrid zone near

Planoles, the 30 SNP markers linked to five of the known colour loci showed strong associations between genotype and their relevant colour score (Fig. 4; Table S5). For example, for the 14 markers within the region on LG6 containing the colour loci *ROS* and *EL*, the highest Spearman correlation coefficient (ρ) between genotype and magenta score was 0.95. Strong correlations were also observed for markers near the three loci that affect yellow colouration, including *CRE* (ρ = 0.68), *FLA* (ρ = 0.77), and *SULF* (ρ = 0.58).

Within the domesticated samples, the relationships between genotype and phenotype were much weaker (Fig. 4; Table S5). Eight out of 30 colour linked markers were monomorphic in the domesticated samples. For other markers the relationships between genotype and phenotype were substantially lower compared with the wild populations. For example, within the *ROS*/*EL* region, the largest value of Spearman's ρ was 0.36, while the strongest relationship with the yellow score was for a marker in the *FLA* region, where the correlation was 0.34.

Discussion

Origins of domesticated the domesticated lines

In our analyses, domesticated individuals from 31 lines, originating from different countries and breeding companies, formed a single genetic group, rather than multiple distinct clusters embedded within the wild samples. This pattern is consistent with these domesticated lines being derived from the same set of founding individuals, followed by further within-line divergence, rather than multiple, independent domestication events. However, more complex histories could also generate a similar pattern. For example, early horticultural varieties may have been independently derived from different sets of founders, and then subsequently mixed, eroding the signal of multiple origins (Meyer & Purugganan, 2013; Sang & Ge, 2007).

Our conclusions about origins of domesticated snapdragons are necessarily limited to the natural populations and domesticated lines included in this study. *Antirrhinum majus* spans a broader geographic and ecological range than is represented here, and we have examined a relatively small number of domesticated lines from a limited set of suppliers. Expanding sampling across both wild populations and cultivated material, along with whole-genome sequencing, will be important for testing the generality of these patterns. In addition, more detailed analysis of historical records could provide valuable context for reconstructing the origins and subsequent diversification of domesticated snapdragons.

Impact of domestication on neutral genetic diversity

Our results suggest that the process of domestication has had a marked effect on levels of genetic diversity in domesticated snapdragons. The difference in diversity observed between wild and domesticated snapdragons is strong, and equates to an average reduction of more than 70% in domesticated lines. Many domesticated species show similar reductions in diversity relative to their wild progenitors, reflecting strong founder effects and increased rates of drift and inbreeding. However, the magnitude varies widely among species. For example, studies show that maize and barley retain ~70–80% of the diversity present in their wild progenitors (Morrell et al., 2014; Ross-Ibarra et al., 2007), whereas stronger reductions have been observed in rice, where domesticated populations retain ~20–40% of wild diversity (Zhu et al., 2007), and tomato, where some cultivated varieties retain less than 5% of the genetic diversity present in wild relatives (100 Tomato Genome Sequencing Consortium et al., 2014).

One factor that has almost certainly contributed to the reduced diversity in domesticated snapdragons is the loss of self-incompatibility, which promotes outcrossing, thereby maintaining high levels of heterozygosity in nature (Li et al., 2019; Surendranadh et al., 2022). By contrast, horticultural varieties readily produce selfed seed without the need for a pollinator, which would lead to high inbreeding. Consistent with this, many of our domesticated lines showed a deficit of heterozygotes, indicated by a positive F_{IS}, which provides direct evidence for recent inbreeding.

In interpreting our results, it is important to acknowledge that this marker panel was designed to include SNPs that show high diversity in nature which could bias our estimate for the reduction in diversity. A further caveat regarding the magnitude of the reduction is that the ancestral source population(s) of domesticated snapdragons remains unknown. If domestication began from a genetically depauperate population, the difference in diversity could partly reflect low standing variation in the founders rather than a reduction during domestication. However, this scenario seems unlikely given the broadly similar levels of diversity across natural populations, and the diversity of flower colour phenotypes present in domesticated lines, suggesting contributions from multiple source populations.

Relationships between colour variation in nature and under domestication

Detailed studies of natural populations of *A. majus* have identified multiple loci underlying variation in flower colour, and molecular markers in linkage disequilibrium with these causal variants have allowed researchers to study their dynamics in natural populations (Bradley et al., 2017, 2025; Field et al., 2025; Richardson et al., 2025; Tavares et al., 2018). Consistent with this, we observed strong genetic differentiation at colour loci between the wild populations of vars. *pseudomajus* and *striatum*, and we also confirm that colour-linked SNPs strongly predict colour variation in samples from the hybrid zone at Planoles.

We were, however, surprised to find that the same loci explain very little of the variation in colour observed among domesticated snapdragons. There are two non-mutually exclusive explanations for this result. First, it is possible that extensive recombination during selective breeding has weakened the associations between marker loci and functional variants, meaning the marker genotypes are no longer informative about an individual's genotype at causal loci in domesticated varieties. The alternative explanation is that different mutations controlling colour variation have arisen in domesticated varieties. Some aspects of colouration—such as extremely high levels of anthocyanin pigmentation, peloric floral morphology, double flowers, and variegated colour patterns—are rare or absent in natural populations. This too might suggest that some new mutations have arisen and been targeted by selection during domestication. Both processes have likely contributed to this finding, and more work is needed to evaluate their relative importance.

Conclusions

Our comparative analysis of wild and domesticated *A. majus* provides a first step towards understanding the genetic consequences of domestication in *A. majus* and highlights its value as a model for additional studies in this area of research. Future work combining whole-genome sequencing, broader sampling across natural populations and domesticated lines, the inclusion of additional *Antirrhinum* species, and a focus on a wider range of traits will help to clarify the origins of domesticated snapdragons, the sources of genetic variation underlying phenotypic diversity.

Acknowledgments

We thank Enrico Coen and members of his research group for ongoing collaborative research on *A. majus*. Monique Burrus and Christophe Andalo helped with the collection of samples. Lucy Unwin, Maria Clara Catellanos and Alba Edwards helped maintain the plants. Members of the Barton groups provided useful discussion and comments on an earlier version of the manuscript.

Data Accessibility

Data and code used to create the manuscript are available on GitHub at https://github.com/seanstankowski/snapdragonDomestication2026

References

100 Tomato Genome Sequencing Consortium, Aflitos, S., Schijlen, E., de Jong, H., de Ridder, D., Smit, S., Finkers, R., Wang, J., Zhang, G., Li, N., Mao, L., Bakker, F., Dirks, R., Breit, T., Gravendeel, B., Huits, H., Struss, D., Swanson-Wagner, R., van Leeuwen, H., … Peters, S. (2014). Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. *The Plant Journal: For Cell and Molecular Biology*, *80*(1), 136–148.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Bradley, D., Boell, L., Richardson, D., Copsey, L., Whibley, A., Xu, T., Zhang, Y., Xue, Y., Field, D., & Coen, E. (2025). Shaping of developmental gradients through selection on multiple loci in*Antirrhinum*. In *bioRxiv*. https://doi.org/10.1101/2025.02.12.637738

Bradley, D., Xu, P., Mohorianu, I.-I., Whibley, A., Field, D., Tavares, H., Couchman, M., Copsey, L., Carpenter, R., Li, M., Li, Q., Xue, Y., Dalmay, T., & Coen, E. (2017). Evolution of flower color pattern through selection on regulatory small RNAs. *Science*, *358*(6365), 925–928.

Cieslak, M., Reissmann, M., Hofreiter, M., & Ludwig, A. (2011). Colours of domestication. *Biological Reviews of the Cambridge Philosophical Society*, *86*(4), 885–899.

Field, D. L., Stankowski, S., Reiter, T., Polechova, J., Bradley, D., Richardson, D., Pal, A., Shipilina, D., Boell, L., Pickup, M., Xue, Y., Coen, E., & Barton, N. (2025). Genome-wide cline analysis identifies new locus contributing to a barrier to gene flow across an*Antirrhinum*hybrid zone. In *bioRxiv* (p. 2025.02. 17.638607). https://doi.org/10.1101/2025.02.17.638607

Goudet, J. (2005). hierfstat, a package for r to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes*, *5*(1), 184–186.

Hudson, A., Critchley, J., & Erasmus, Y. (2008). The genus antirrhinum (snapdragon): a flowering plant model for evolution and development. *CSH Protocols*, *2008*, db.emo100.

Li, M., Zhang, D., Gao, Q., Luo, Y., Zhang, H., Ma, B., Chen, C., Whibley, A., Zhang, Y. 'e, Cao, Y., Li, Q., Guo, H., Li, J., Song, Y., Zhang, Y., Copsey, L., Li, Y., Li, X., Qi, M., … Xue, Y. (2019). Genome structure and evolution of Antirrhinum majus L. *Nature Plants*, *5*(2), 174–183.

Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nature Reviews. Genetics*, *14*(12), 840–852.

Morrell, P. L., Gonzales, A. M., Meyer, K. K., & Clegg, M. T. (2014). Resequencing data indicate a modest effect of domestication on diversity in barley: a cultigen with multiple origins. *Journal of Heredity*, *105*(2), 253–264.

Nei, M. (1987). *Molecular evolutionary genetics*. https://scholar.google.co.uk/citations?user=VxOmZDgAAAAJ&hl=en&oi=sra

Pal, A., Shipilina, D., Le Moan, A., McNairn, A. J., Grenier, J. K., Kucka, M., Coop, G., Chan, Y. F., Barton, N. H., Field, D. L., & Stankowski, S. (2025). Genealogical analysis of replicate flower colour hybrid zones in Antirrhinum. *Molecular Ecology*, e70067.

Richardson, Bradley, D., Copsey, L., Whibley, A., Burrus, M., Andalo, C., Zhu, S., Field, D. L., Xue, Y., & Coen, E. (2025). Genomic tree scans identify loci underlying adaptive peaks in*Antirrhinum*. In *bioRxiv*. https://doi.org/10.1101/2025.02.12.637406

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution; International Journal of Organic Evolution*, *66*(1), 1–17.

Ross-Ibarra, J., Morrell, P. L., & Gaut, B. S. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, *104 Suppl 1*(suppl_1), 8641–8648.

Sang, T., & Ge, S. (2007). Genetics and phylogenetics of rice domestication. *Current Opinion in Genetics & Development*, *17*(6), 533–538.

Schwarz-Sommer, Z., Davies, B., & Hudson, A. (2003). An everlasting pioneer: the story of Antirrhinum research. *Nature Reviews. Genetics*, *4*(8), 657–666.

Surendranadh, P., Arathoon, L., Baskett, C. A., Field, D. L., Pickup, M., & Barton, N. H. (2022). Effects of fine-scale population structure on the distribution of heterozygosity in a long-term study of *Antirrhinum majus*. *Genetics*, *221*(3). https://doi.org/10.1093/genetics/iyac083

Surendranadh, P., Stankowski, S., Field, D. L., & Barton, N. H. (2025). Extreme long-range migration distorts flower colour clines in an *Antirrhinum* hybrid zone. In *bioRxiv* (p. 2025.06.25.661303). https://doi.org/10.1101/2025.06.25.661303

Tavares, H., Whibley, A., Field, D. L., Bradley, D., Couchman, M., Copsey, L., Elleouet, J., Burrus, M., Andalo, C., Li, M., Li, Q., Xue, Y., Rebocho, A. B., Barton, N. H., & Coen, E. (2018). Selection and gene flow shape genomic islands that control floral guides. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(43), 11006–11011.

Whibley, A. C., Langlade, N. B., Andalo, C., Hanna, A. I., Bangham, A., Thébaud, C., & Coen, E. (2006). Evolutionary paths underlying flower color variation in Antirrhinum. *Science*, *313*(5789), 963–966.

Yang, W., Liu, M., Feng, L., Jiao, P., Jiang, J., Huang, L., Liu, J., López-Pujol, J., & Hu, Q. (2025). Domestication history and genetic changes for the newly evolved flower color in the ornamental plant Lobularia maritima (Brassicaceae). *Horticulture Research*, *12*(4), uhae355.

Zhu, Q., Zheng, X., Luo, J., Gaut, B. S., & Ge, S. (2007). Multilocus analysis of nucleotide variation of Oryza sativa and its wild relatives: severe bottleneck during domestication of rice. *Molecular Biology and Evolution*, *24*(3), 875–888.

Supplementary information for

Taming the dragon: genetic variation in wild and domesticated *Antirrhinum majus*

Tables:

Figures:

Table S1. Information about each of the wild populations used in this study.

Population ID	Latitude	Longitude	Sample size	Variety
BAR****	42.124	0.3216	10	*pseudomajus*
BES****	42.2114	2.667	10	*pseudomajus*
CAL****	42.1038	1.8318	10	*pseudomajus*
CLA****	43.17665	3.09585	10	*pseudomajus*
D_***_R*	43.3427	1.4925	10	*pseudomajus*
GARL****	42.4505	2.617	10	*pseudomajus*
GCHI****	42.77	1.86	9	*pseudomajus*
GPER****	42.47621	2.84878	10	*pseudomajus*
GPUI****	42.522385	2.61729	10	*pseudomajus*
M406*	42.3543812	2.16898442	10	*pseudomajus*
MP113*	42.3308333	2.17	10	*pseudomajus*
MRT****	43.64537	3.8901	10	*pseudomajus*
PLAMN	42.3221454	2.0879949	10	*pseudomajus*
PLAMF	42.3204901	2.09630819	10	*pseudomajus*
SUE****	42.4153	0.7373	10	*pseudomajus*
VIE****	42.7377	0.7571	9	*pseudomajus*
D27**	42.5067	2.1218	10	*striatum*
CADI****	42.2913825	1.86464051	10	*striatum*
ELS****	42.28883	1.36765	10	*striatum*
GBOU****	42.7203	2.2582	10	*striatum*
GCAM****	42.8	1.92	10	*striatum*
GDOS****	42.965	2.5258	9	*striatum*
GFAB****	42.59347	2.6152	7	*striatum*
LU****	42.97	2.26	10	*striatum*
MLYS****	42.83	2.2	6	*striatum*
PAL****	42.3073	1.8556	10	*striatum*
POM****	43.1122	2.2715	9	*striatum*
YP044*	42.360097	1.926686	9	*striatum*
PLAYF	42.3234887	2.04835431	10	*striatum*
PLAYN	42.3246171	2.06534717	10	*striatum*
USS****	42.7444	2.0858	10	*striatum*
VIL****	42.59	2.3649	10	*striatum*
PLAHZ	42.3226764	2.07446891	11	hybrid

572 **Table S2. Information about each of the domesticated lines used in this study.**

Line	Line name	Supplier	Type
L1	Beautiful Mixture	Dehner	Mixed Height
L2	Mango Twist	Suttons	Tall Border
L3	Defiance	Chiltern Seeds	Cut Flowers
L4	High Mix	Dehner	Mixed Height
L6	Lucky Lips	Thompson And Morgan	Cut Flowers
L7	Sweet Duet F1	Thompson And Morgan	Cut Flowers
L10	Mini Cherry Cola	Thompson And Morgan	Short Border
L11	Purple Twist' F1 Hybrid	Chiltern Seeds	Tall Border
L12	Canary Bird	Chiltern Seeds	Cut Flowers
L13	Day And Night	Suttons	Tall Border
L14	Brighton Rock	Happy Green Shop Seeds	Cut Flowers
L16	Half-Height Mix	Dehner	Mixed Height
L17	Serenade F2	Kiepenkerl	Mixed Height
L19	Rembrandt	Thompson And Morgan	Cut Flowers
L22	Liberty Lavender	Chiltern Seeds	Cut Flowers
L25	Orange Wonder	Plant Gang	Cut-Flowers
L26	Trailing Mix	Thompson And Morgan	Mixed Height
L28	Dwarf Mix	Kiepenkerl	Dwarf
L30	Black Prince	Thompson And Morgan	Cut Flowers
L32	Mendel's Garden	NA	NA
L33	Potomac Royal F1	Plants Of Distinction	Cut Flowers
L35	Royal Bride	Thompson And Morgan	Cut Flowers
L36	Snaptastic Scarlet Orange	Chiltern Seeds	Cut Flowers
L37	Tetra Ruffled Giants	Chiltern Seeds	Cut Flowers
L39	Cut-Flower Mix	Kiepenkerl	Cut Flowers
L41	Antiquity Sunset Mix	Suttons	Dwarf
L42	Double Madame Butterfly Mixed F1	Thompson And Morgan	Cut Flowers
L43	Rembrandt Bicolour	Mr Fothergills	Tall Border
L44	Bizarre Hybrids	Sarah Raven	Cut Flowers
L45	Nanum Crown Mix	Flower Seeds UK	Mixed Height
L46	Madame Butterfly Bronze White F1	Chiltern Seeds	Cut Flowers

573
574
575
576
577
578
579
580

Table S3. Mean and standard deviation of per-individual heterozygosity for each population/line.

Group type	group	H_I all loci	s.d. H_I all loci	mean H_I non-colour loci	s.d. H_I non-colour loci	Variety
wild	BAR****	0.243	0.049	0.216	0.039	pseudomajus
wild	BES****	0.351	0.043	0.286	0.037	pseudomajus
wild	CAL****	0.359	0.102	0.309	0.087	pseudomajus
wild	CLA****	0.305	0.053	0.279	0.050	pseudomajus
wild	D_***_R*	0.326	0.037	0.289	0.031	pseudomajus
wild	GARL****	0.369	0.065	0.327	0.060	pseudomajus
wild	GCHI****	0.373	0.034	0.337	0.052	pseudomajus
wild	GPER****	0.383	0.056	0.330	0.038	pseudomajus
wild	GPUI****	0.383	0.086	0.372	0.056	pseudomajus
wild	M406*	0.400	0.067	0.330	0.050	pseudomajus
wild	MP113*	0.404	0.063	0.335	0.046	pseudomajus
wild	MRT****	0.287	0.041	0.233	0.029	pseudomajus
wild	PLAMN	0.449	0.041	0.397	0.043	pseudomajus
wild	PLAMF	0.411	0.062	0.371	0.040	pseudomajus
wild	SUE****	0.334	0.048	0.292	0.038	pseudomajus
wild	VIE****	0.323	0.050	0.298	0.035	pseudomajus
wild	PLAHZ	0.445	0.066	0.441	0.051	hybrid
wild	D27**	0.300	0.070	0.249	0.053	striatum
wild	CADI****	0.410	0.076	0.383	0.055	striatum
wild	ELS****	0.308	0.058	0.307	0.046	striatum
wild	GBOU****	0.426	0.033	0.388	0.032	striatum
wild	GCAM****	0.379	0.057	0.335	0.042	striatum
wild	GDOS****	0.345	0.076	0.318	0.052	striatum
wild	GFAB****	0.426	0.055	0.367	0.039	striatum
wild	LU****	0.280	0.045	0.262	0.052	striatum
wild	MLYS****	0.378	0.067	0.359	0.070	striatum
wild	PAL****	0.409	0.050	0.381	0.048	striatum
wild	POM****	0.354	0.070	0.329	0.067	striatum
wild	YP044*	0.358	0.083	0.326	0.062	striatum
wild	PLAYF	0.432	0.036	0.385	0.035	striatum
wild	PLAYN	0.453	0.054	0.405	0.056	striatum
wild	USS****	0.388	0.076	0.348	0.057	striatum
wild	VIL****	0.359	0.078	0.324	0.065	striatum
domesticated	L1	0.169	0.064	0.138	0.040	NA
domesticated	L2	0.003	0.006	0.002	0.004	NA
domesticated	L3	0.056	0.037	0.048	0.026	NA
domesticated	L4	0.152	0.051	0.133	0.047	NA
domesticated	L6	0.031	0.017	0.022	0.012	NA
domesticated	L7	0.045	0.019	0.034	0.014	NA
domesticated	L10	0.087	0.055	0.080	0.049	NA
domesticated	L11	0.088	0.024	0.105	0.021	NA
domesticated	L12	0.067	0.036	0.063	0.031	NA
domesticated	L13	0.008	0.010	0.014	0.015	NA
domesticated	L14	0.069	0.032	0.069	0.017	NA
domesticated	L16	0.126	0.066	0.130	0.067	NA
domesticated	L17	0.095	0.043	0.079	0.041	NA
domesticated	L19	0.071	0.032	0.060	0.024	NA
domesticated	L22	0.071	0.009	0.051	0.006	NA
domesticated	L25	0.093	0.029	0.079	0.026	NA
domesticated	L26	0.129	0.059	0.127	0.059	NA
domesticated	L28	0.170	0.040	0.147	0.031	NA
domesticated	L30	0.067	0.039	0.071	0.039	NA
domesticated	L32	0.061	0.031	0.061	0.031	NA
domesticated	L33	0.069	0.000	0.072	0.003	NA
domesticated	L35	0.077	0.072	0.063	0.061	NA
domesticated	L36	0.105	0.010	0.133	0.007	NA
domesticated	L37	0.177	0.040	0.161	0.035	NA
domesticated	L39	0.135	0.057	0.136	0.058	NA
domesticated	L41	0.004	0.009	0.003	0.007	NA
domesticated	L42	0.105	0.022	0.108	0.031	NA
domesticated	L43	0.057	0.048	0.041	0.034	NA
domesticated	L44	0.038	0.029	0.027	0.021	NA
domesticated	L45	0.050	0.027	0.043	0.021	NA
domesticated	L46	0.108	0.015	0.097	0.015	NA

Table S4. Mean expected heterozygosity (*Ho*), observed heterozygosity (*Hs*), and the inbreeding coefficient (F_{IS}) for each population/line over all loci and for non-colour loci.

group	sample type	*Ho* all loci	*Hs* all loci	F_{IS} all loci	H_o non-colour loci	H_s non-colour loci	F_{IS} non-colour loci
BAR****	wild	0.220	0.235	0.065	0.245	0.255	0.041
BES****	wild	0.292	0.310	0.059	0.360	0.367	0.021
CADI****	wild	0.391	0.416	0.061	0.422	0.451	0.065
CAL****	wild	0.315	0.335	0.059	0.373	0.395	0.057
CLA****	wild	0.285	0.270	-0.058	0.313	0.303	-0.034
D27**	wild	0.256	0.253	-0.012	0.311	0.301	-0.033
D_***_R*	wild	0.294	0.278	-0.060	0.333	0.311	-0.073
ELS****	wild	0.314	0.335	0.064	0.315	0.347	0.093
GARL****	wild	0.334	0.338	0.013	0.377	0.382	0.014
GBOU****	wild	0.398	0.372	-0.068	0.439	0.412	-0.065
GCAM****	wild	0.341	0.349	0.022	0.381	0.398	0.042
GCHI****	wild	0.345	0.349	0.012	0.381	0.390	0.024
GDOS****	wild	0.325	0.339	0.042	0.354	0.367	0.033
GFAB****	wild	0.375	0.359	-0.045	0.437	0.403	-0.083
GPER****	wild	0.336	0.349	0.038	0.393	0.385	-0.020
GPUI****	wild	0.381	0.383	0.007	0.391	0.408	0.043
LU****	wild	0.269	0.303	0.112	0.292	0.331	0.118
M406*	wild	0.336	0.349	0.038	0.407	0.426	0.044
MLYS****	wild	0.365	0.356	-0.025	0.386	0.382	-0.012
MP113*	wild	0.341	0.348	0.022	0.413	0.428	0.035
MRT****	wild	0.236	0.250	0.059	0.293	0.307	0.045
PAL****	wild	0.389	0.395	0.014	0.424	0.425	0.003
POM****	wild	0.336	0.354	0.050	0.366	0.382	0.043
YP044*	wild	0.332	0.364	0.088	0.367	0.392	0.064
PLAYF	wild	0.393	0.397	0.010	0.438	0.448	0.023
PLAYN	wild	0.410	0.422	0.030	0.462	0.459	-0.008
PLAHZ	wild	0.451	0.431	-0.047	0.459	0.457	-0.004
PLAMN	wild	0.405	0.400	-0.010	0.462	0.461	-0.002
PLAMF	wild	0.379	0.393	0.036	0.420	0.443	0.052
SUE****	wild	0.298	0.310	0.037	0.341	0.364	0.062
USS****	wild	0.354	0.355	0.005	0.399	0.400	0.001
VIE****	wild	0.304	0.315	0.036	0.331	0.353	0.061
VIL****	wild	0.330	0.324	-0.019	0.365	0.359	-0.015
L1	domesticated	0.139	0.157	0.113	0.168	0.176	0.043
L2	domesticated	0.002	0.004	0.561	0.003	0.006	0.561
L3	domesticated	0.048	0.078	0.375	0.057	0.083	0.310
L4	domesticated	0.132	0.155	0.149	0.149	0.158	0.055
L6	domesticated	0.022	0.038	0.417	0.031	0.054	0.417
L7	domesticated	0.034	0.040	0.170	0.046	0.056	0.179
L10	domesticated	0.081	0.117	0.310	0.085	0.117	0.271
L11	domesticated	0.114	0.066	-0.738	0.085	0.050	-0.690
L12	domesticated	0.065	0.092	0.299	0.066	0.089	0.256
L13	domesticated	0.014	0.027	0.480	0.008	0.023	0.667
L14	domesticated	0.069	0.112	0.383	0.067	0.119	0.438
L16	domesticated	0.131	0.162	0.194	0.125	0.167	0.253
L17	domesticated	0.080	0.113	0.291	0.096	0.134	0.278
L19	domesticated	0.061	0.081	0.251	0.073	0.102	0.282
L22	domesticated	0.052	0.028	-0.816	0.074	0.040	-0.816
L25	domesticated	0.080	0.097	0.174	0.090	0.112	0.196
L26	domesticated	0.127	0.156	0.184	0.129	0.164	0.209
L28	domesticated	0.148	0.155	0.047	0.173	0.180	0.041
L30	domesticated	0.077	0.093	0.172	0.069	0.094	0.265
L32	domesticated	0.061	0.099	0.383	0.063	0.111	0.438
L33	domesticated	0.076	0.040	-0.875	0.072	0.036	-1.000
L35	domesticated	0.064	0.105	0.393	0.073	0.117	0.375
L36	domesticated	0.135	0.071	-0.918	0.108	0.058	-0.859
L37	domesticated	0.163	0.185	0.120	0.179	0.190	0.059
L39	domesticated	0.137	0.168	0.183	0.134	0.175	0.233
L41	domesticated	0.003	0.072	0.959	0.004	0.042	0.899
L42	domesticated	0.110	0.107	-0.028	0.109	0.112	0.031
L43	domesticated	0.041	0.047	0.119	0.058	0.064	0.084
L44	domesticated	0.028	0.067	0.589	0.039	0.073	0.465
L45	domesticated	0.042	0.153	0.724	0.049	0.175	0.721
L46	domesticated	0.096	0.080	-0.198	0.108	0.078	-0.388

Table S5. Genotype-phenotype associations for SNPs linked to known colour loci in the natural populations. The strength of the relationship between the colour scores and the genotypes at each SNP is indicated by the rho value estimated from the Spearman Rank correlation and the intensity of the coloured shading. Where the Rho value is 'NA' it could not be calculated because the locus was monomorphic.

Locus name	LG	Map pos (cM)	Colour locus	Rho wild	Rho domesticated	Pigment
ros_assembly_473914	6	49.06	ROS/EL	0.041	NA	Magenta
ros_assembly_543443	6	49.29	ROS/EL	0.907	NA	Magenta
ros_assembly_567004	6	49.37	ROS/EL	0.950	0.223	Magenta
ros_assembly_575837	6	49.40	ROS/EL	0.770	NA	Magenta
ros_assembly_576271	6	49.40	ROS/EL	0.275	0.079	Magenta
ros_assembly_620992	6	49.55	ROS/EL	0.732	0.045	Magenta
ros_assembly_653015	6	49.66	ROS/EL	0.868	0.115	Magenta
ros_assembly_670530	6	49.72	ROS/EL	0.752	0.102	Magenta
ros_assembly_715001	6	49.86	ROS/EL	0.936	0.339	Magenta
ros_assembly_737420	6	49.94	ROS/EL	0.268	NA	Magenta
ros_assembly_744403	6	49.96	ROS/EL	0.014	0.289	Magenta
ros_assembly_748981	6	49.98	ROS/EL	0.596	NA	Magenta
ros_assembly_758578	6	50.01	ROS/EL	0.522	0.369	Magenta
ros_assembly_849332	6	50.31	ROS/EL	0.091	0.191	Magenta
s1187_290152	1	2.17	CRE	0.678	NA	Yellow
s154_504353	2	41.99	FLA	0.100	NA	Yellow
s2338_45429	2	43.18	FLA	0.178	0.344	Yellow
s829_8371	2	43.18	FLA	0.764	0.219	Yellow
s829_204463	2	43.18	FLA	0.454	0.206	Yellow
s316_93292	2	43.18	FLA	0.768	0.200	Yellow
s316_257789	2	43.18	FLA	0.429	0.137	Yellow
s444_38909	2	43.18	FLA	0.399	0.157	Yellow
s992_223854	2	43.18	FLA	0.030	0.165	Yellow
s148_425797	2	44.31	FLA	0.212	0.216	Yellow
s1140_224946	2	44.91	FLA	0.136	NA	Yellow
s155_1201194	2	48.05	FLA	0.123	0.207	Yellow
s91_78256	4	9.36	SULF	0.577	0.068	Yellow
s91_122561	4	9.36	SULF	0.465	0.124	Yellow
s91_181717	4	9.36	SULF	0.533	0.003	Yellow

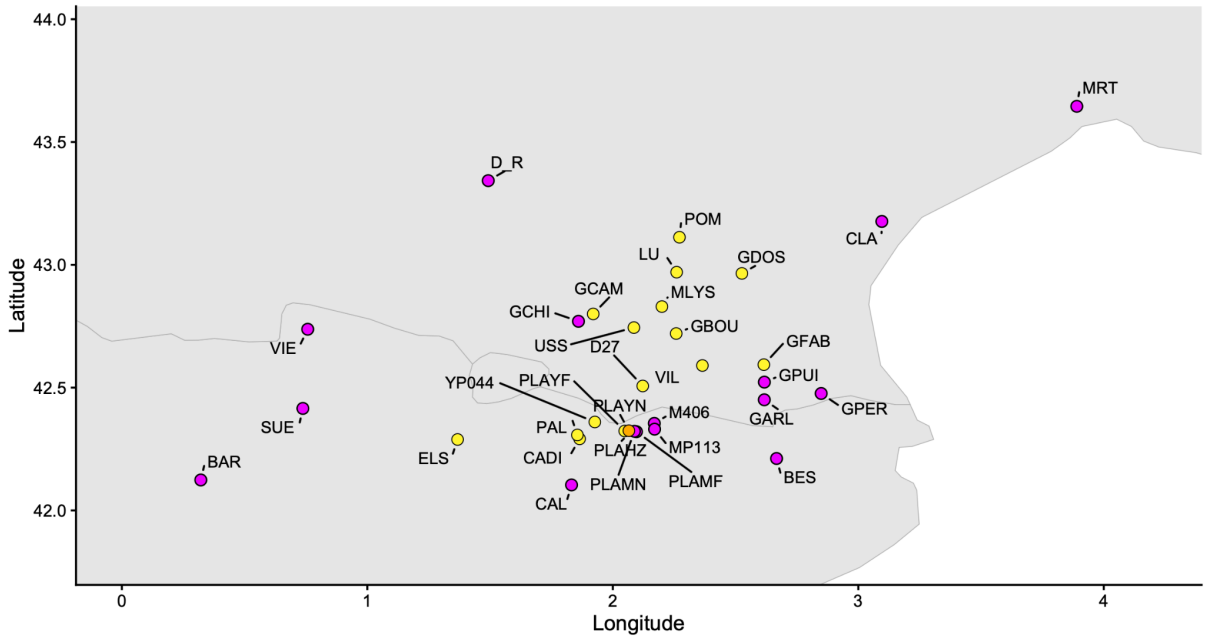

Figure S1. Map of wild populations of A. *majus* ssp. *majus* sampled in this study. The colour of the circles indicates whether the population was the magenta flowered var. *pseudomajus*, the yellow flowered var. *striatum*, or a hybrid population.

Fig. S2. Images of one flower from replicate plants for each of the domesticated lines.

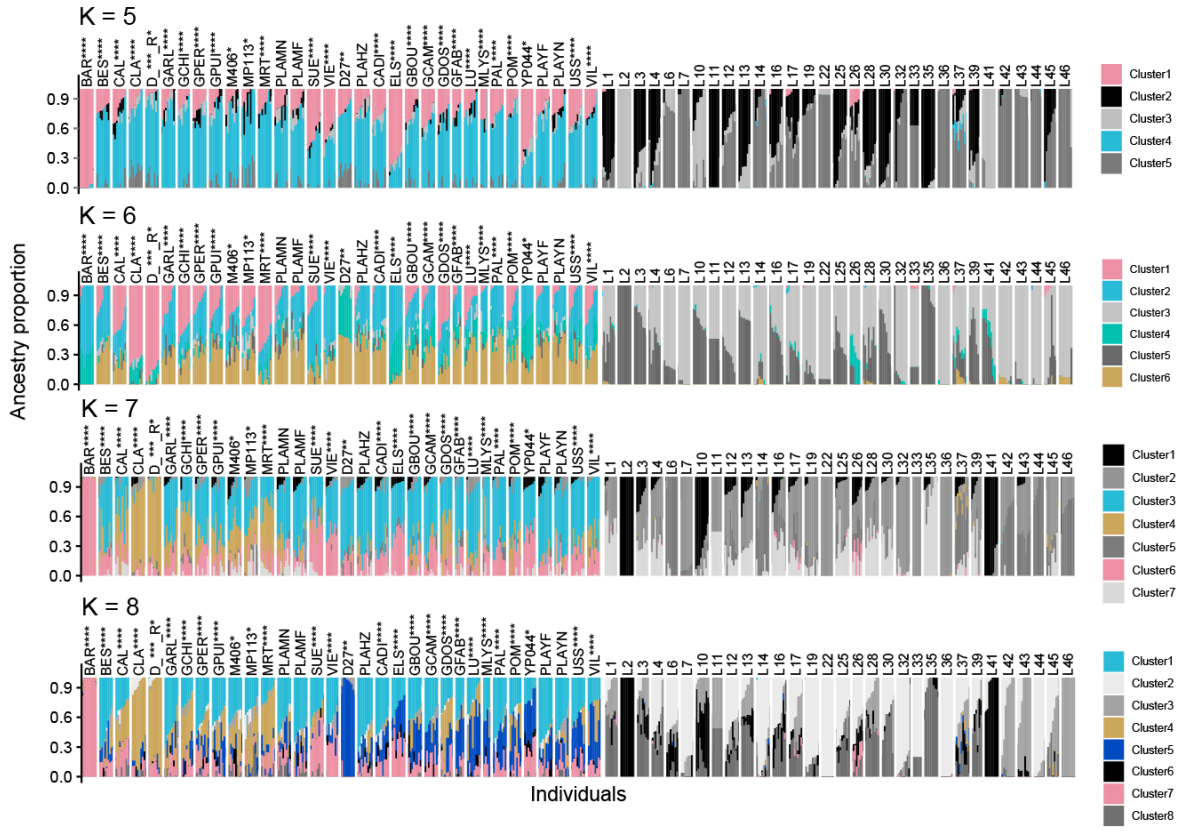

Fig. S3. Results of Admixture analysis for different values of K. Clusters primarily associated with the wild populations are shown as colours, while those primarily associated with domesticated samples are shown as shades of grey.

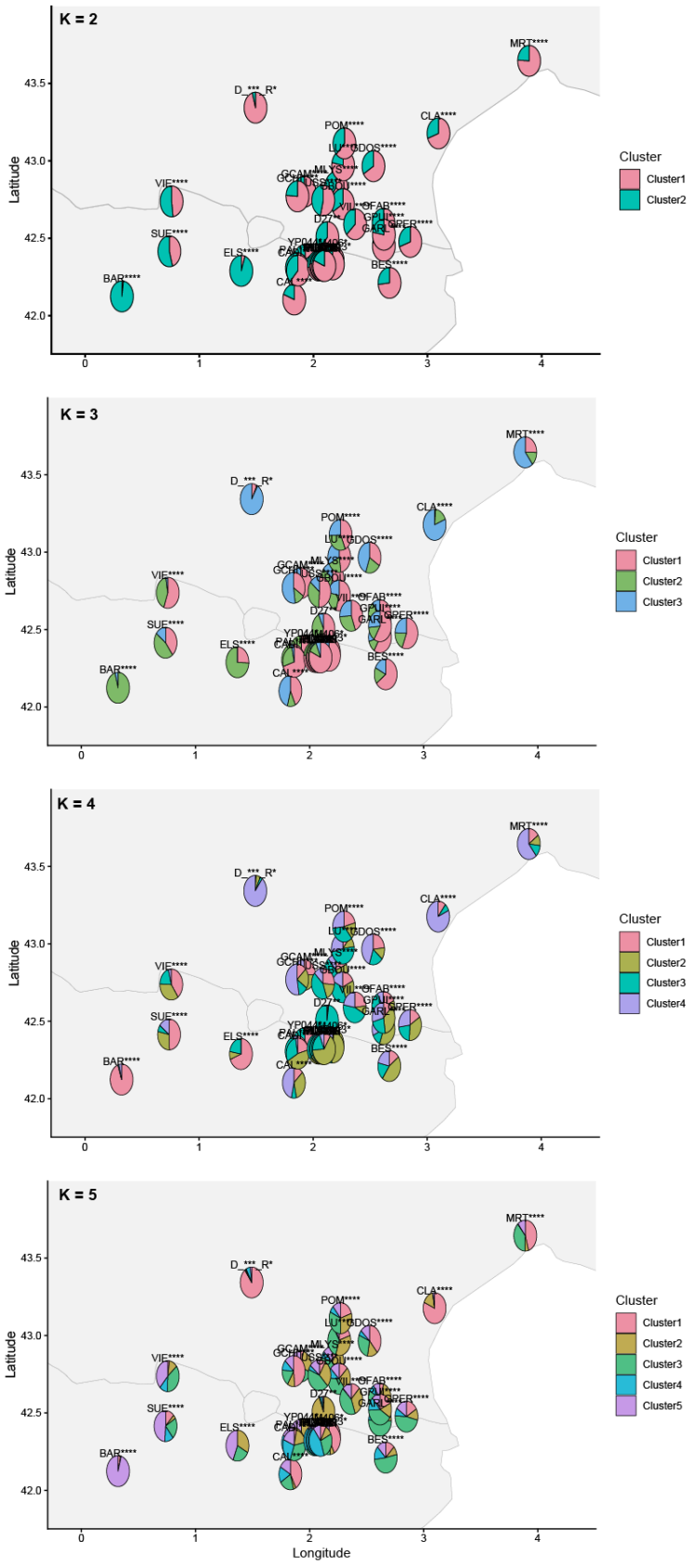

Fig S4. Maps of admixture scores for the wild populations. The pie charts show the total proportion of genome assigned to each K across all individuals from within a population. For these analyses, *Admixture* was run only on the wild individuals.

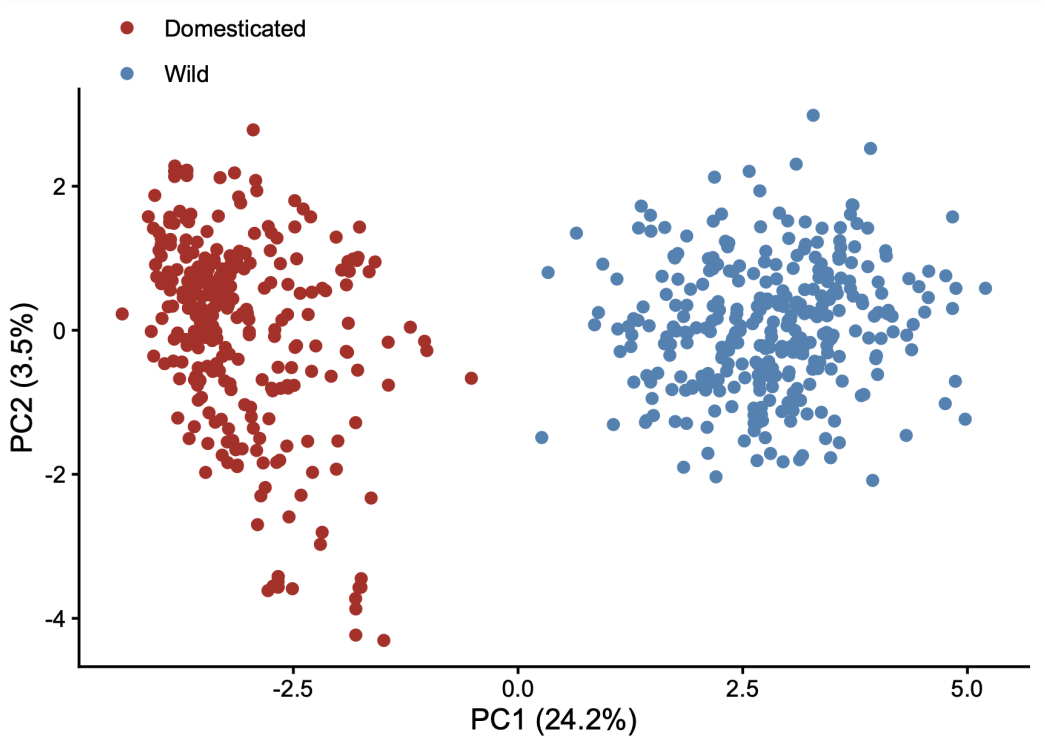

Figure S5. Principal Component Analysis (PCA) of the genotype matrix for the domesticated and wild samples. The PCA is based on the 70 loci not associated with flower colour.

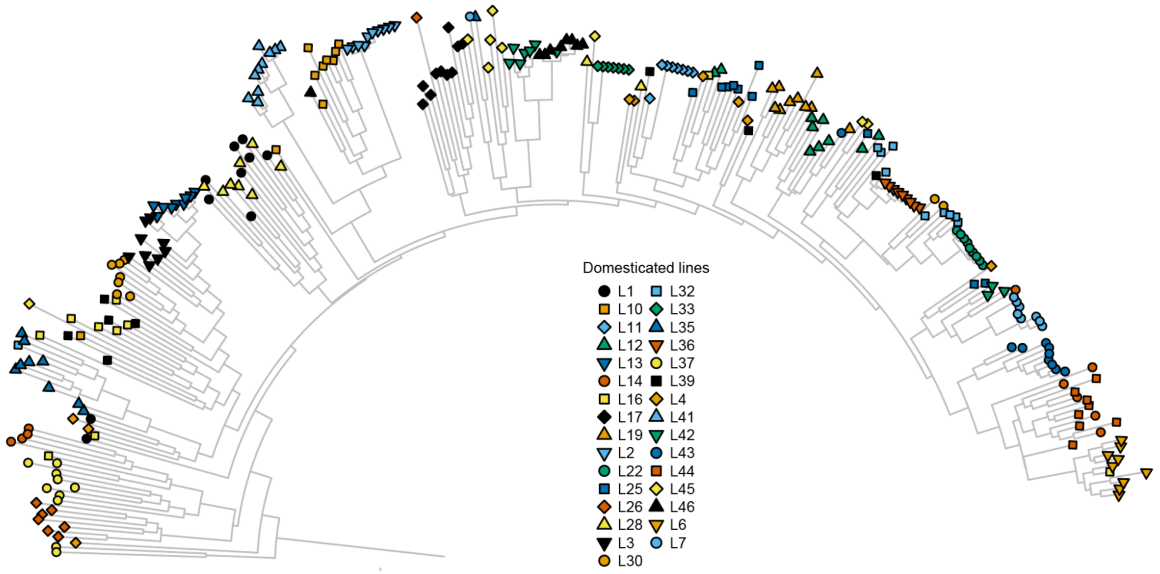

Figure S6. A radial view of the clade of domesticated samples. The topology is identical to that shown in Fig. 1B.