*Supplementary Materials and Methods*

**Cline-fitting procedures**

One-dimensional cline analysis was used to estimate a series of parameters that describes the pattern of spatial variation for each marker/trait. We used *Analyse 1.3* (Barton & Baird 1995) to fit three nested one-dimensional cline models to the *MaMyb2* allele frequencies and mean trait values for each population. The simplest model is a sigmoid cline described by a four-parameter hyperbolic tangent, *y* = (1 + tanh[2(*xi* - *c*)/*w*])/2, where *y* is the frequency of an allele or the mean value of a trait, *c* is the cline center, and *w* is the cline width, defined as the ratio between the change in the trait across the zone and gradient at the cline center (∆*p*/(∂*z*/∂*x*). ∆*p* is estimated as the difference between the maximum and minimum allele frequencies or trait values in the tails of the cline (*P*max and *P*min, respectively), which are estimated during the fit. The other two models are ‘stepped’ clines, which consist of a central sigmoid step flanked by two exponential tails, where *y* ∝ exp(– 4*xθ*1/2/*w*), that describes the pattern of gene flow into the foreign gene pool; *θ* is the rate of decay, and the strength of the barrier to gene flow, *B*, can be estimated as the ratio between the difference in the allele frequency/mean trait value and the initial gradient along distance *x* at the edges of the central segment (Szymura & Barton 1989). In the symmetrical ‘S-step’ model, *θ* and *B* are equal on both sides. In the asymmetrical ‘A-step’ model, the pattern of gene flow is different on the left and right side.

 *Analyse 1.3* (Barton & Baird 1995) uses the Metropolis algorithm to search the likelihood surface to find the optimal solution to the model. To ensure that the likelihood surface was thoroughly explored, multiple runs were conducted using different initial parameter estimates and random seeds. *Analyse 1.3* allows likelihood values to be calculated assuming either a binomial distribution (‘frequency’ model) or normal distribution (‘quantitative trait’ model). With the exception of stigma exertion, which was analyzed using the ‘frequency’ model based on the proportion of flowers with exerted stigmas in each population, likelihood values were calculated using the ‘quantitative trait’ model for all traits. After obtaining ML solutions for the three models, the best fitting model was identified using Likelihood Ratio Tests (α = 0.05) by comparing the test statistic (2Δ*ML*) to a chi-squared distribution where the degrees of freedom were equal to the difference in the number of free parameters between more complex and simpler models. As the minimum and maximum mean allele frequencies or trait values (pmin, *p*max) were allowed to vary in the tails of the cline, the sigmoid, A-step, and S-step models were described by 4 (*pmin*, *pmax*, *c*, *w*), 6 (*pmin*, *pmax*, *c*, *w*, *θ, B*) and 8 parameters (*pmin*, *pmax*, *c*, *w*, *θ0, B0, θ1, B1*), respectively. To meet the requirements of the program, trait data were scaled between 0 and 1 according to the formula *zi* = *xi* – *xmin* /*xmax* – *xmin*, where *xi*is the *i*th observed data and *xmin* and *xmax* are the observed minimum and maximum values of the data, respectively.

To confirm that the populations in the common garden experiment captured the same pattern of spatial genetic variation observed in natural populations, we compared the shapes of the *MaMyb2-*M3 genetic marker clines between natural and greenhouse-raised populations. After genotyping each greenhouse raised individual at the *MaMyb2*-M3 marker according to Streisfeld et al. (2013) we used *Analyse 1.3* (Barton & Baird 1995) to fit one-dimensional cline models to the three allele frequency datasets. Pairwise likelihood ratio tests were used to determine if cline shapes differed significantly among the three datasets. This involved constraining all model parameters for dataset *i* to the estimates obtained for dataset *j* and recording the change in the likelihood. Two-times the differences in the log-likelihood values (2Δ*ML*) were compared to a chi-square distribution where the degrees of freedom were equal to the number of constrained parameters.

**Floral trait measurements**

All measurements were taken from two flowers per plant on the first day that they were fully open. Flower color was quantified by extracting and measuring anthocyanin pigment content according to Streisfeld and Kohn (2005). Nectar volume, corolla length, and corolla tube width were measured with digital calipers as described in Streisfeld and Kohn (2005). Pedicel length was measured as the distance from the stem to the base of the calyx. Corolla width was measured as the widest width across the petals. The length of the tallest stamen, shortest stamen, and style length were measured from the base of the corolla tube. Stigma exertion was calculated as the difference between style length and corolla length. Stigma-anther distance was calculated as the difference between style length and length of the tallest anther.

**Tests for coincidence and concordance**

The coincidence of cline centers (*c*) and concordance of cline widths (*w*) was tested using a composite likelihood method (Phillips et al. 2004; Kawakami et al. 2004). The method involves obtaining high-resolution likelihood profiles for a set of traits across a range of parameter values, with the parameter of interest (*c* or *w*) constrained and the other parameter values free to vary. The log-likelihood estimates for each parameter value were summed over traits to obtain a composite likelihood profile. The parameter value with the lowest log likelihood is taken as *MLcomp*. *MLsum* is then obtained by summing the lowest log-likelihood values obtained for each trait. The test statistic, Δ*ML*, is obtained by subtracting *MLcomp* from *MLsum*. If clines are coincident, *MLsum* is not significantly smaller (more likely) than *MLcomp*,as determined by a chi-squared test (α = 0.05) with *n*-1 degrees of freedom where *n* is the number of traits included in the test. Log-likelihood profiles were constructed for each trait across the full range of available parameter space (75 km) with a resolution of 20 m using the ‘cross section’ routine in *Analyse 1.3.*

**Molecular Methods**

DNA isolation, RADseq library preparation, and Illumina HiSeq 2000 sequencing followed the methods described in Sobel and Streisfeld (2015). Data processing occurred using *Stacks v. 1.12* (Catchen et al. 2013). Reads were filtered based on quality, and errors in the barcode sequence or RAD site were corrected using the *process\_radtags* script in *Stacks*. Loci were created using the *denovo\_map.pl* function of *Stacks*, with two identical raw reads required to create a stack, two mismatches allowed between loci for an individual, and two mismatches allowed when processing the catalog. Single nucleotide polymorphisms (SNPs) were determined and genotypes called using a maximum-likelihood statistical model implemented in *Stacks* (Hohenlohe et al. 2010, 2012; Catchen et al. 2011). We performed several independent runs in *Stacks* using a range of parameters for stack building, genotype calling. Preliminary analyses conducted on the resulting datasets produced qualitatively similar results. To include a locus in the final dataset, which was constructed using the same parameters values as Sobel and Streisfeld (2015), we required it to be present in at least 80% of all individuals, genotyped to a depth of at least 15x, and present at a minor allele frequency > 0.1. To avoid analyzing multiple SNPs from the same 100-bp locus, only the first SNP on each tag was output by *Stacks.*

**Structure analysis**

We tested for population genetic structure using the Bayesian model-based clustering algorithm implemented in the program *Structure* *2.3.4* (Pritchard et al. 2000). Four replicate *Structure* runs were conducted with varying numbers of inferred clusters (*K*1…16), assuming admixture and correlated allele frequencies. Each run consisted of 50 000 iterations of burn-in, followed by 150 000 iterations of sampling. The optimal *K* was determined using the Δ*K* method (Evanno et al. 2005), implemented in *Structure Harvester* (Earl & van Holdt 2011). The probability of membership to each cluster (*Q*) was obtained for each individual at the optimal *K* by averaging values obtained from 20 independent runs using *Clummp* (Jackobsson & Rosenberg 2007).

**Permutation test for comparing mean estimates of within and among ecotype *FST***

We compared mean pairwise of *FST* between the intra-ecotype and inter-ecotype treatments. The permutation test was conducted using custom scripts in *R* (scripts deposited on DRYAD). We first calculated the difference between the observed mean estimates of intra- and inter-ecotype pairwise *FST*.After pooling the data from each treatment group, we randomly resampled the data into two groups with sample sizes equal to the treatment groups (100,000) and calculated the difference between the two subsamples for each permutation. The probability of observing the mean difference between the intra- and inter-ecotype treatments by chance was calculated as the proportion of the permutations in which the mean difference between the groups was greater than or equal to the observed mean difference.

*Literature cited*

Barton, N. H., and S. J. E. Baird. 1995. Analyse: an application for analysing hybrid zones. Freeware, Edinburgh.

Catchen, J., A. Amores, P. A. Hohenlohe, W. A. Cresko, and J. H. Postlethwait. 2011. *Stacks*: Building and genotyping loci *de novo* from short-read sequences. *G3: Genes Genomes Genetics* 1:171-182.

Earl, D. A., and B. M. vonHoldt. 2012. Structure harvester: A website and program for visualizing Structure output and implementing the Evanno method. *Conservation Genetics Resources* 4:359-361.

Evanno, G., S. Regnaut., and J. Goudet. 2005. Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology* 14:2611-2620.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PloS Genetics* 6:e1000862.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806.

Kawakami, T., R. K. Butlin, M. Adams, D. Paull, and S. J. Cooper. 2009. Genetic analysis of a chromosomal hybrid zone in the Australian morabine grasshoppers (Vandiemenella, viatica species group). *Evolution* 63:139-152.

Phillips, B. L., S. J. Baird, and C. Moritz. 2004. When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. *Evolution* 58:1536-1548.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.

Sobel, J.M., and M. A Streisfeld. 2015. Strong premating isolation exclusively drives insipient speciation in *Mimulus aurantiacus. Evolution* 69:447-461.

Streisfeld, M. A., and J. R. Kohn. 2005. Contrasting patterns of floral and molecular variation across a cline in *Mimulus aurantiacus*. *Evolution* 59:2548-2559.

Streisfeld, M. A., W. N. Young, and J. M. Sobel. 2013. Divergent selection drives genetic differentiation in an R2R3-Myb transcription factor that contributes to incipient speciation in *Mimulus aurantiacus*. *PLoS genetics* 9:e1003385.

Szymura J.M., and Barton, N.H. 1986. Genetic analysis of a hybrid zone between the fire bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution* 4:1141-1159.

*Supplementary tables*

**Table S1.** Geographic coordinates for the 30 localities sampled in Streisfeld et al. (2013) and used to generate the *MaMyb2-*M3 *red* allele frequency dataset (*P*). Also indicated are the 16 populations used in the common garden experiment, along with the allele frequency at the same marker from seedlings grown in a common environment.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Pop. ID | Latitude(Degrees) | Longitude(Degrees) | Ecotype | *P*nature | RADseqSample size | Geography ID (west to east) | *P*garden |
| BCG | 32.63108 | -117.02170 | R | 1.00 |  |  |  |
| CLC | 32.92045 | -117.16942 | R | 1.00 |  |  |  |
| CRS | 33.13037 | -117.30717 | R | 0.93 |  |  |  |
| DLR | 33.16818 | -117.05237 | R | 1.00 | 9 | 6 | 1 |
| EHP | 32.72317 | -117.07298 | R | 1.00 |  |  |  |
| ELF | 33.08595 | -117.14530 | R | 0.98 | 11 | 3 | 0.94 |
| ELT | 32.89422 | -117.08982 | R | 1.00 |  |  |  |
| FLP | 32.80582 | -116.98670 | R | 1.00 |  |  |  |
| LDG | 32.72752 | -116.97810 | R | 1.00 |  |  |  |
| LH | 33.06088 | -117.11877 | R | 1.00 | 12 | 4 | 0.97 |
| MT | 32.82095 | -117.06175 | R | 1.00 | 12 | 2 | 1.00 |
| OSP | 33.10197 | -117.03508 | R | 1.00 |  |  |  |
| PMD | 32.93787 | -117.05913 | R | 1.00 |  |  |  |
| SDP | 32.99810 | -117.23538 | R | 1.00 |  |  |  |
| SXN | 33.07717 | -117.28523 | R | 1.00 |  |  |  |
| UCSD | 32.88940 | -117.23618 | R | 1.00 | 12 | 1 | 1.00 |
| BC | 33.12262 | -116.80468 | H | 0.04 | 14 | 12 | 0.04 |
| BS | 33.01480 | -117.01643 | H | 0.88 |  |  |  |
| DLZ | 32.65250 | -116.78597 | H | 0.28 |  |  |  |
| JMC | 32.73732 | -116.95410 | H | 0.96 | 18 | 5 | 0.98 |
| LKW | 33.16372 | -117.01610 | H | 0.22 | 15 | 8 | 0.13 |
| MW | 33.00718 | -116.95978 | H | 0.37 | 17 | 9 | 0.60 |
| OAK | 32.91407 | -116.88932 | H | 0.53 | 12 | 10 | 0.53 |
| WM | 32.82133 | -116.90228 | Y | 0.69 | 17 | 7 | 0.72 |
| BCRD | 32.94958 | -116.63795 | Y | 0.00 | 6 | 13 | 0.00 |
| INJ | 33.09785 | -116.66432 | Y | 0.00 | 11 | 14 | 0.00 |
| LO | 32.67670 | -116.33123 | Y | 0.00 | 12 | 16 | 0.00 |
| PCT | 32.73258 | -116.46983 | Y | 0.00 | 9 | 15 | 0.03 |
| POTR | 32.60380 | -116.63392 | Y | 0.00 | 12 | 11 | 0.00 |
| PVT | 32.83340 | -116.54860 | Y | 0.04 |  |  |

**Table S2.** Maximum likelihood parameter estimates obtained from one-dimensional cline analysis of the *MaMyb2*-M3 marker from the three datasets under the best fitting cline model (A-step).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Populations | *c* (km) | *w* (km) | *pmin* | *pmax* | *θ0* | *B0/w* | *θ1* | *B1/w* | -lnl |
| 30, nature | -0.765 | 0.958 | 0.0001 | 0.9829 | 14.70 | 0.0032 | 2.91e+09 | 0.1249 | -12.38 |
| 16, nature | -0.781 | 0.905 | 0.0001 | 0.9894 | 13.65 | 0.0030 | 5.90e+09 | 0.7730 | -11.40 |
| 16, garden | -0.618 | 0.909 | 0.0017 | 0.9796 | 12.60 | 0.0032 | 4.04e+09 | 0.5293 | -12.28 |

*c*, position of cline center; *w*, cline width; *Pmin*, allele frequency in the right tail, *Pmax*, allele frequency in the left tail; *B0/w*, barrier strength standardized by *w*; *θ*, rate of exponential decay; *-lnl*, log-likelihood.

**Table S3.** Results of linear regression, testing for a relationship between variation in mean trait value for each population in the common garden experiment and distance along the one-dimensional transect. Statistically significant values in bold.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trait** | **Slope** | **Intercept** | ***r*2** | ***F*-ratio** | ***P*** |
| Nectar volume (μl) | -0.0054 | 0.477 | 0.132 | 3.29 | n.s.  |
| **Anthocyanin abs.** | **-0.0117** | **0.546** | **0.737** | **43.05** | **< 0.0001** |
| **Pedicel length (mm)** | **-0.0127** | **0.466** | **0.685** | **33.59** | **<0.0001** |
| **Tube width (mm)** | **0.0143** | **0.458** | **0.736** | **42.85** | **<0.0001** |
| **Corolla length (mm)** | **0.0128** | **0.449** | **0.761** | **44.713** | **<0.0001** |
| **Corolla width (mm)** | **0.0119** | **0.401** | **0.633** | **24.14** | **0.0002** |
| Style length (mm) | -0.0061 | 0.661 | 0.223 | 4.02 | n.s.  |
| Tallest anther (mm) | -0.0057 | 0.411 | 0.117 | 2.98 | n.s.  |
| Shortest anther (mm) | -0.0048 | 0.343 | 0.085 | 2.40 | n.s.  |
| **Stigma exertion (mm)** | **-0.0136** | **0.614** | **0.874** | **105.01** | **<0.0001** |
| Stigma-anther distance (mm) | -0.0029 | 0.596 | -0.009 | 0.86 | n.s.  |

**Table S4.** Results of the Evanno et al. (2005) method for inferring the optimum number of population clusters (*K*) following analysis in the program *Structure* (Pritchard et al 2000). The optimum value of *K* is highlighted in bold text.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***K*** | **n runs** | **Mean LnP(K)** | **SD LnP(K)** | **Ln'(K)** | **|Ln''(K)|** | **Delta K** |
| 1 | 3 | -391987.53 | 3.51 | NA | NA | NA |
| **2** | **4** | **-367339.83** | **3.30** | **24647.71** | **10223.63** | **3101.99** |
| 3 | 4 | -352915.75 | 4.66 | 14424.08 | 9496.05 | 2038.67 |
| 4 | 4 | -347987.73 | 98.66 | 4928.03 | 56.68 | 0.57 |
| 5 | 4 | -343003.03 | 16.97 | 4984.70 | 1551.50 | 91.42 |
| 6 | 4 | -339569.83 | 17.41 | 3433.20 | 261.48 | 15.02 |
| 7 | 4 | -336398.10 | 569.88 | 3171.73 | 405.08 | 0.71 |
| 8 | 4 | -333631.45 | 656.07 | 2766.65 | 299.78 | 0.46 |
| 9 | 4 | -331164.58 | 693.03 | 2466.88 | 369.35 | 0.53 |
| 10 | 4 | -329067.05 | 1091.13 | 2097.53 | 272.78 | 0.25 |
| 11 | 4 | -326696.75 | 81.71 | 2370.30 | 904.20 | 11.07 |
| 12 | 4 | -325230.65 | 94.20 | 1466.10 | 292.68 | 3.11 |
| 13 | 4 | -324057.23 | 183.26 | 1173.43 | 31.35 | 0.17 |
| 14 | 4 | -322852.45 | 392.65 | 1204.78 | 18090.30 | 46.07 |
| 15 | 4 | -339737.98 | 15907.63 | -16885.53 | 1933.00 | 0.12 |
| 16 | 4 | -358556.50 | 71367.94 | -18818.53 | NA | NA |

**Table S5.** Composite likelihood tests comparing the cline width between the genomic PC1 and six floral traits. Δ*ML,* test statistic.

|  |  |  |
| --- | --- | --- |
| Trait | Δ*ML* | *P* |
| Anthocyanin | 37.86 | < 0.0001 |
| Pedicel length | 32.98 | < 0.0001 |
| Corolla width | 24.06 | < 0.0001 |
| Corolla length | 32.80 | < 0.0004 |
| Tube width | 12.8 | < 0.0001 |
| Stigma exertion | 17.28 | < 0.0003 |

*Supplementary figures*



**Figure S1.** Mean values from the first two principal components of a PC analysis conducted on the SNP genotype matrix (5382 loci). The horizontal and vertical error bars are ± 1 standard deviation of PCs 1 and 2 for each sample site, respectively. Sites are color coded by ecotype (red, yellow, or hybrid). Numbers show the order of sample sites across the collapsed transect from left to right.

****

**Figure S2.** Variation in population structure across the one-dimensional transect revealed by a Bayesian clustering analysis (5382 loci). Top panel: Individual *Q* scores from *Stucture* (*K* = 2) within each population, arranged in geographic order from left to right. Each vertical line represents the proportion of ancestry to one of two clusters (black or white shading) for each individual. Sample locations are as in Fig. 1. Center panel: Variation in mean *QK1*score across the collapsed transect. The solid line is the least-squares regression, and dotted lines are the 95% confidence intervals. Bottom panel: standard deviation of *QK1* in each sample site. The position of the hybrid zone (hz) is indicated by the horizontal line.

****

**Figure S3.** Relationship between geographic and genetic distance (*FST*, averaged over loci) among sample locations. Intra-ecotype (red × red ecotype and yellow × yellow ecotype) and inter-ecotype (i.e., red ecotype × yellow ecotype) comparisons are coded using red and blue symbols, respectively. The dashed black line shows the least-squares regression for all pairs of comparisons, while the red and blue lines show the similar scaling-relationships for the intra-ecotype and inter-ecotype comparisons, respectively.