Supplementary Material for: Natural selection drives genome-wide evolution via chance genetic associations

Zachariah Gompert[*,1,2], Jeffrey L. Feder[3],
Patrik Nosil[4]

[1]Department of Biology, Utah State University, Logan, UT 84322, USA

[2]Ecology Center, Utah State University, Logan, UT 84322, USA

[3]Department of Biological Sciences, University of Notre Dame,
Notre Dame, Indiana 46556, USA

[4]CEFE, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3,
Montpellier, 34293, France

[*]To whom correspondence should be addressed;
E-mail: zach.gompert@usu.edu.

Supplementary Analyses and Results

Results for transplant to *Ceanothus*

The results presented in the main text focus on $\sim$7 million single nucleotide polymorphism (SNPs) from whole-genome sequence data from 246 T. cristinae transplanted to *Adenostoma* [65]. This approach focuses on survival in a natural habitat, as exemplified by studies of Darwin's finches and other birds [66], and thus avoids effects of transplantation to a novel host. Here, we briefly summarize largely parallel results based on 245 *T. cristinae* transplanted to *Ceanothus*, an alternative host used by *T. cristinae*.

We used the same ABC model described in the main text to obtain genomic predictions of expected fitness for the individuals released on *Ceanothus*. Similar to the *Adenostoma* treatment, cross-validation showed that our model for the *Ceanothus* treatment outperformed a null model, reducing errors for predicting survival by 20% (error rate for null model = 40%, error rate for fit model = 32%, 3 P < 0.001)(Supplementary Figure 1).

We tested for an association between the extent to which each SNP was associated with expected fitness and the total selection experienced by that SNP in the *Ceanothus* treatment. As with the *Adenostoma* treatment described in the main text, we first obtained Bayesian estimates of total selection as part of this analysis. We found that different alleles at loci were on average associated with a 11% difference in relative survival probabilities (i.e., a difference in marginal fitness of 11% or $|s| = 0.11$, 1st quartile = 0.049, 3rd quartile = 0.17)(Supplementary Figure2). Note, that while the distribution of s was very similar between the two host treatments, values of s for individual SNPs were only weakly correlated between treatments (Pearson $r = 0.044$, P < 0.001).

As with the *Adenostoma* treatment, the strength of association between SNP genotypes and expected fitness was significantly explained by our estimates of total selection (Pearson r for

2

29 log-odds of selection = 0.466, 95% CIs = 0.465-0.466, P < 0.0001)(Supplementary Figure 3B).

30 Thus, LD among (unknown) causal variants and SNPs, as captured by the association between

31 SNPs and genomic predictions of expected fitness, predicts the selection experienced by SNPs

32 during the experiment in this treatment as well.

LD between genome-wide SNPs and *Agouti* in a mouse experiment

34 To complement our primary analyses, we tested for genome-wide LD between SNPs and a

35 causal variant in an additional, independent data set. We did this to verify that our general

36 findings were not unique to our specific stick insect data set.

37   We used genomic data from a field enclosure experiment involving cryptically colored pop-

38 ulations of deer mice (*Peromyscus maniculatus*) transplanted to light or dark-soil enclosures

39 [67]. In this experiment, 481 deer mice were collected and released in light-soil (N = 217)

40 or dark-soil (N = 229) field-enclosures where they were subject to natural selection. Previ-

41 ous analyses of this experiment combined with genetic manipulations showed that a deletion

42 (*Ser*) within the *Agouti* gene affected the color of these mice and causally affected fitness

43 (i.e., was subject to direct selection). Here, we tested for LD (measured by r^2) between this

44 causal variant and 53,507 genome-wide SNPs that were sequenced for the same experimental

45 individuals. We estimated LD separately for the light-soil and dark-soil experimental popu-

46 lations. This was done using the same analytical approach and software as was used for the

47 analysis of the *Mel-Stripe* locus in *T. cristinae*.

48   We found evidence that SNPs across the genome were associated with *Ser* genotype in

49 both experimental populations (Fig. 3A,B). LD exceeded ~0.04 for the 1% of SNPs most as-

50 sociated with *Ser* in each population, and these high-LD SNPs were distributed across all

51 23 autosomes and the X chromosome. Nonetheless, a greater proportion of SNPs on chromo-

52 some 4 (which harbors *Agouti*) were among the 1% most associated SNPs than for any other

3

chromosome in the dark-soil population, and chromosome 4 has the second greatest propor-
tion of SNPs associated with *Ser* in the light-soil population. Moreover, about 20% of SNPs
were in LD with *Ser* to a non-trivial extent (i.e., $r^2 > 0.01$ for 20.0% and 23.0% of SNPs
in the light-soil and dark-soil experimental populations, respectively). Thus, similar to our
core results with *Mel-Stripe* in *Timema*, direct selection on *Ser* combined with long-range
LD, necessarily caused a degree of indirect selection on each of the 24 chromosomes in the *P.*
maniculatus populations.

LD between SNP markers and a fitness QTL in a stickleback experiment

To further assess the generality of our results, we tested for LD between SNP markers and a
QTL marker for fitness in an third data set. Specifically, we analyzed genetic data from an
experimental field population of threespine stickleback (*Gasterosteus aculeatus*) [68, 69]. In
this experiment, fitness was measured as the number of surviving F3 offspring attributable to
each F2 female fish released in an experimental pond. The experiment uncovered a single QTL
for fitness (offspring number), with the peak signal for a SNP on chromosome 4 near the *Eda*
locus (position 12,815,024), a gene known to affect bony armor and fitness [70, 71, 72]. We
treated this SNP marker as the causal locus for the purpose of our analysis, and tested for LD
with each of 387 additional genome-wide SNPs in the set of 224 F2 females. This was done
using the same analytical approach used for *T. cristinae* and the deer mouse data set.

LD with the QTL marker was exceptionally high for most SNPs on chromosome 4 (Fig.
3C). However, we also detected LD between this QTL marker and SNPs on other chromosomes
(Fig. 3D). For example, LD exceeded ∼0.04 for about 10% of SNPs, and for 1% of SNPs not
on chromosome 4. Moreover, 29% of SNPs, including 22% of SNPs not on chromosome 4,
exhibited non-trivial LD with the QTL marker (i.e., $r^2 > 0.01$). Consequently, we again found
evidence that, based on patterns of long-range LD, direct selection on a single locus would

4

77 necessarily result in indirect selection on physically unlinked loci distributed across multiple

78 chromosomes.

Detailed Methods

Field release-recapture experiment

81 In 2011, we collected 500 *T. cristinae* from a single population (4.51753°N, 119.80125°W,

82 named FHA), where the dominant host is *Adenostoma* and all individuals were collected from

83 this host [73]. A small tissue sample (i.e., a portion of one leg) was taken from each insect

84 using sterile scissors. Past work has shown no effect of such tissue sampling on survival in

85 the field or lab [73]. These individuals were then deployed in a mark-recapture experiment,

86 with ∼50 insects transplanted onto each of 10 experimental plants (five *Adenostoma* and five

87 *Ceanothus*) (see [73] for details). After eight days, surviving stick insects were collected to

88 quantify genome-wide allele frequency change between release and recapture, and thus test for

89 selection. Notably, past work has shown that recapture is a good proxy for survival, due to the

90 limited dispersal ability of these wingless insects [74, 75, 76, 73].

DNA sequence alignment, variant calling and genotype estimation for the experiment.

93 Whole genome DNA sequence data for individuals from this experiment were described in

94 [77, 65], where data was successfully obtained for 491 of the 500 transplanted insects. For

95 the current study, we aligned the whole genome DNA sequence data from each of these 491 *T.*

96 *cristinae* to the *T. cristinae* reference genome (version 1.3) using the bwa (version 07.10-r789)

97 mem algorithm with a band width of 100, a 20 bp seed length and a minimum score for output

98 of 30 [78]. We then used samtools (version 1.5) to compress, sort and index the align-

99 ments, and to remove PCR duplicates [79]. We then used the GATKHaplotypeCaller and

5

GenotypeGVCFs modules (version 3.5) to call variants and calculate genotype likelihoods [80]. We required a minimum base quality of 30, set the prior probability of heterozygosity to 0.001, and only called variants with a minimum phred-scaled confidence of 50.

The following filters were then applied using custom Perl scripts: minimum coverage of $1\times$ per individual (i.e., 491X coverage across all individuals), a minimum value of the base quality rank sum test of -8, a minimum value of the mapping quality rank sum test of -12.5, a minimum value of the read position rank sum test of -8, a minimum ratio of variant confidence to non-reference read depth of 2, a minimum mapping quality of 40, a maximum phred-scaled P-value of Fisher's exact test for strand bias of 60, and a minimum minor allele frequency of 0.01. Further, we only retained SNPs mapped to one of the 13 *T. cristinae* linkage groups. This resulted in 7,243,463 SNPs, which were used in subsequent analyses.

Next, we obtained maximum likelihood estimates of allele frequencies for all experimental samples using an expectation-maximization (EM) algorithm, as described in [81] and implemented in our own C++ program, estpEM [77]. For this, we used a convergence tolerance of 0.001 and allowed for a maximum of 30 EM iterations. We then used these allele frequency estimates and the genotype likelihoods from GATK to calculate empirical Bayesian genotype estimates. In particular, we computed $\Pr(g_{ij} = k | l_{ijk}; p_i) \propto l_{ijk} \Pr(g_{ij} = k | p_i)$ where $g_{ij} \in (0, 1, 2)$ denotes the genotype (i.e., count of non-reference alleles) for SNP i and stick insect j, l_{ijk} is the likelihood of genotype k, p_i is the non-reference allele frequency, and $\Pr(g_{ij} = k | p_i)$ is given by the binomial probability mass function with $n = 2$. Point estimates (posterior means) were then obtained as $\hat{g}_{ij} = \sum_{k \in (0,1,2)} k \times \Pr(g_{ij} = k | l_{ijk}; p_i)$. These point estimates range from zero to two, and are not constrained to be integer values.

Testing for indirect selection associated with *Mel-Stripe*

Following [65], we defined the *Mel-Stripe* color locus as spanning 4.1 Mbps on scaffold ∼702.1 and ∼6.4 Mbps on the adjacent scaffold 128, both on linkage group 8 (LG8). This is a large, complex structural variant that includes a ∼10 Mbp putative inversion and ∼1 Mbp deletion [82, 83].

As in past work [84], we inferred each individual's *Mel-Stripe* genotype from patterns of clustering in principle components analysis (PCA) space [85]. We first performed a PCA for all 491 *T. cristinae* based on 79,014 SNPs within the *Mel-Stripe* locus. This was done on the centered, but not standardized genotype matrix in R (version 3.5.1). We then used k-means clustering to group individuals into clusters based on the PC scores for the first PC axis; we assumed three groups. We delineated *Mel-Stripe* genotypes based on the k-means cluster (i.e., group) assignment. The three clusters recovered correspond with homozygotes for alternative inversion orientations (negative versus positive values of PC1, coded as genotype = 0 or 2) and heterozygotes (intermediate values of PC1, coded as genotype = 1) [84, 65] (Fig. 4).

We then measured pairwise LD between each of the ∼7 million SNPs and the *Mel-Stripe* locus as the square of the correlation between genotypes [86]. We wrote a computer program in C++ to efficiently perform this calculation for each SNP. The program relies heavily on functions from the Gnu Scientific Library (GSL) [87].

Estimating total selection

We next inferred the total selection experienced by each of the ∼7 million SNPs. Here, total selection reflects the combined effects of direct and indirect selection as captured by differences in marginal fitness (expected fitness averaged across genetic backgrounds) of genotypes at each SNP locus. Because we assume most SNPs do not causally effect fitness in this short-term experiment (but see [88]), we equate total selection with indirect selection for simplicity. We

7

146 used a Bayesian approach to estimate total selection on each SNP while accounting for genetic

147 drift. Information for this inference comes from mortality in the experiment and patterns of

148 allele frequency change. We first describe the model, and then document details pertaining to

149 how the model was fit.

150 We assume viability selection such that each individual's absolute fitness (i.e., survival

151 probability) is given by $w_j \in (0, 1)$. Here, we treat each SNP locus in a separate analysis and

152 assume absolute fitness is a function of an individual's genotype at a SNP and experimental

153 block (i.e., experimental bush). Thus, we allow survival to depend on genotype and also to vary

154 in space. We consider each SNP's association with survival independently with the expectation

155 that any such association mostly arises from LD with (unknown) causal variants not a direct

156 affect on fitness.

157 We model w as $w_{ij} = (1 - g_{ij} s_i) \div \left[\sum_{j \in m_j} (1 - g_{ij} s_i) v_{m_j} \right]$. Here, the numerator denotes

158 the expected fitness of individual j based on its genotype at locus i and the selection coefficient

159 (differential) s. The denominator converts this to an absolute fitness by dividing by the product

160 of the sum of the relative fitness values for all individuals in block m_{ij} (i.e., the same block

161 as individual j) and the survival proportion for block m_{ij} (denoted v_{m_j}). This ensures that

162 the mean absolute fitness in each block is equal to the survival proportion for that block. We

163 then assume that $y_j \sim \text{Bernoulli}(w_{ij})$, where y_j is a binary outcome variable denoting whether

164 the individual lived ($y_j = 1$) or died ($y_j = 0$). We assume a beta prior on the strength of total

165 selection experienced by each SNP, such that $s_i \sim \text{beta}(\ = 1, \ = 3)$. This is a semi-

166 conservative prior as it places greater prior probability on lower values of s.

167 We constructed a Markov chain Monte Carlo (MCMC) algorithm to generate samples from

168 the posterior distribution of each s_i. Proposal values for each s_i were generated from a uniform

169 distribution centered on the current value and accepted or rejected according to the Metropolis-

170 Hastings algorithm. A computer program implementing the MCMC analysis was written in

8

171 C++ with functions from the GSL [87]. We obtained posterior distributions for s_i for each

172 SNP based on 10,000 MCMC iterations following a 1000 iteration burnin. See "Supplementary

173 Analyses and Results" above for our complementary analysis of the *Ceanothus* treatment.

Genomic prediction of expected fitness

175 We obtained genomic predictions of expected fitness in an approximate Bayesian computation

176 (ABC) framework based on survival and genotype data from the sequenced SNPs for the 246

177 individuals transplanted to the five *Adenostoma* bushes (see "Supplementary Analyses and Re-

178 sults" above for our complementary analysis of the *Ceanothus* treatment). Our specific method

179 is tailored to our data and thus has the added benefits of allowing for variation in absolute fit-

180 ness among transplant sites, directly modeling the experimental design, and treating survival

181 explicitly as a binary phenotype. We first describe the model, and then second our procedure

182 for fitting the model.

183   We assume viability selection such that each individual's absolute fitness (i.e., survival

184 probability) is given by $w_j \in (0, 1)$. Absolute fitness is a function of an individual's multi-

185 locus genotype and experimental block (i.e., experimental bush). Thus, we allow survival to

186 depend on genotype but also to vary in space. We model $\mathbf{w}$ as follows. We specifically assume

187 multiplicative fitness such that,

$$w_j = \frac{\prod (1 - g_{ij} s_i)}{\prod_{j \in m_j} [\prod_i (1 - g_{ij} s_i)] v_{m_j}}, \tag{1}$$

188   Here, the numerator denotes the expected fitness of individual j based on its multilocus

189 genotype and the vector of selection coefficients $\mathbf{s}$. The denominator converts this to an abso-

190 lute fitness by dividing by the product of the sum of the relative fitness values for all individuals

191 in block m_j (i.e., the same block as individual j) and the survival proportion for block m_j (de-

192 noted v_{m_j}). This ensures that the mean absolute fitness in each block is equal to the survival

9

proportion for that block. We then assume that $y_j \sim \text{Bernoulli}(w_j)$, where y_j is a binary outcome variable denoting whether the individual lived ($y_j = 1$) or died ($y_j = 0$).

We specify a hierarchical model for the selection coefficients. We first specify a spike-and-slab prior for the number of causal variants, $n \sim \quad 0 + (1 - \quad)U(0, 100)$, where $U(0, 100)$ is an integer uniform distribution bounded by 0 and 100 (i.e., we assume between 0 and 100 causal variants), and $\quad = 0.5$ specifies the contribution of the uniform distribution and point-mass at zero. The n causal variants are then chosen randomly and with equal probability from a set of possible (i.e., candidate) casual variants (we discuss this in more detail in the section below on "Defining candidate causal variants for the direct selection analysis"). Next, we sample a value for the expected effect size for each casual variant from $\text{beta}(\quad; \quad)$, with $\quad = 1.0$ and $\quad = 9.0$ (this distribution has an expected value of 0.1). Effects (selection coefficients) for individual causal variants are then sampled from an exponential distribution with the mean effect set as the expectation; the sign of individual effects is specified at random and with equal probabilities for the non-reference allele increasing or decreasing fitness.

Our primary goal with this model is to use an ABC approach to estimate expected fitness while integrating over uncertainty in the model parameters (i.e., number, identity and effects of causal variants). As with ABC methods in general [89], we perform inference by (i) computing summary statistics from the observed data, (ii) sampling model parameters from their prior distributions (noted above), (iii) simulating evolution (here survival in the experiment), and (iv) retaining parameter values from a subset of simulations that generated summary statistics most similar to the observed data to form the (approximate) posterior distribution. We chose to fit the model based on the survival data vector ($\mathbf{y}$), rather than on some summary of allele frequency change. This provides a compact summary of the most critical information for estimating expected fitness, and necessarily accounts for the fact the multiple SNPs could provide equivalent and redundant information about expected fitness because of linkage disequilibrium

10

218 (LD).

219   We wrote a computer program in `C++` to simulate evolution based on the model described
220 above (this and other code associated with this manuscript is available from `GitHub`, `https:`
221 `//github.com/zgompert/TimemaPolygenicSelection.git`. The program re-
222 lies heavily on functions from the Gnu Scientific Library (GSL) [87]. We ran 400 million ($4e^8$)
223 simulations of evolution to fit the model. We then used a simple rejection algorithm to retain
224 parameter values from the 1000 simulations (0.00025% of the total) with the lowest discrep-
225 ancy with the observed data in terms of the survival vector. These were equated with a sample
226 from the (approximate) posterior distribution.

Defining candidate variants for the direct selection analysis

228 As noted above, we did not include all $\sim$7 million SNPs in the direct selection analysis, as this
229 would have been computationally infeasible (there would simply be too many combinations
230 of causal loci for even a cursory exploration via simulation in a realistic amount of time).
231 Thus, we used the results from the total selection model fit (which can be done for all $\sim$7
232 million SNPs) to delineate a subset of SNPs as candidates for the direct selection analysis. We
233 specifically identified the 49,133 SNPs with 95% equal-tail probability intervals on s_i from
234 the total selection analysis that excluded zero. That is, we chose the SNPs with the greatest
235 evidence of having experienced selection under the assumption that such SNPs are likely to
236 best represent candidates for those causally affecting fitness. By using this subset of SNPs
237 for the direct selection analysis, were are implicitly assigning zero prior probabilities for other
238 SNPs having direct effects on fitness in the direct selection model.

Testing the performance of the genomic prediction model with cross-validation

240 We next used cross-validation to quantify the predictive power of our genomic prediction model
241 given our data set. To do this, we ran an additional 400 million simulations with the computer

11

program described in the preceding section. However, in each simulation 80% of individuals were assigned to a training set, with the other 20% assigned to a test set. The training set was used to fit the model, and the test set allowed us to test the model on data not used for model fitting. Assignments to test versus training sets were random, and differed across simulations.

We chose the set of samples to retain to form the posterior as before (i.e., based on the prediction error), but based only on the discrepancy between observed and simulated survival for the 80% of individuals in the training set (that is, model fit was based solely on the training set). We then measured predictive performance based on the discrepancy between observed and simulated survival for the 20% of individuals in the test set, and considering only the subset of 1000 simulation outputs retained as part of the posterior (based on the training set). We then compared this to the expected error (i.e., discrepancy between observed and simulated) under a null model with no predictive power. To do this, we computed the test set discrepancy for random sets of 1000 simulations (i.e., these were not selected based on the match of the training set to the observed data). We repeated this procedure 100 times to obtain a null distribution of test set errors expected from a model with no actual predictive power.

Quantifying LD between SNPs and expected fitness

We asked whether the strength of the association (i.e., LD) between SNPs and genomic predictions of expected fitness could explain how strongly each SNP was impacted by indirect selection in the experiment. As noted in the main text, this involved three steps: (1) estimate expected fitness of each individual stick insect from our polygenic genomic prediction model, (2) quantify how strongly each of the ∼7 million SNPs was statistically associated with expected fitness at the organism level, (3) test whether the strength of the SNP-expected fitness relationship predicts independent estimates of total selection on each SNP.

Estimates of expected fitness for each stick insect were obtained from the ABC model

described above. In particular, each simulation used in the ABC fit of this model involves computing expected fitness for each stick insect (w_j). We thus obtained a posterior distribution for w_j from the 1000 simulations retained as part of the approximate posterior in the rejection step of the ABC fit as noted above.

We then fit linear regression models with expected fitness as the response variable and genotype at each SNP (one SNP at a time) as the independent variable. This was repeated 1000 times for each SNP, using the 1000 posterior samples for expected fitness. Thus, we obtained a posterior distribution on the coefficient of determination (r^2) for the linear model describing the association between each SNP and expected fitness. We used the mean of this distribution as the point estimate for each SNP. This was again done with code written in C++.

Lastly, we computed the Pearson correlation between the r^2 association of each SNP with expected fitness and the evidence that each SNP experienced selection. This was done in R. As evidence for selection, we used $log\left(\frac{x_i}{1-x_i}\right)$, where $x_i = 2|Pr(s_i > 0) - 0.5|$ (this treats evidence for $s_i > 0$ and $s_i < 0$ equally). We refer to this as the log-odds of selection. See "Supplementary Analyses and Results" above for our complementary analysis of the *Ceanothus* treatment.

Measuring LD and selection in windows

To capture local (rather than long range) patterns of LD and selection, we calculated the average LD and total (mostly indirect) selection in contiguous regions (windows) of the genome.

For LD, we calculated the mean LD (measured by the squared genotypic correlation, r^2) for all pairs of SNPs in 10 kilo base pair (kbp) windows along each genome scaffold (i.e., windows never spanned multiple scaffolds). This was done using a C++ program we wrote. For total selection, we instead calculated the mean log-odds of selection in 50 SNP windows. We chose to use windows defined by the number of SNPs here as we found that estimates for physical-

13

distance windows were very sensitive to SNP density (in contrast, because LD decays with physically distance, using windows based on physical distance for average LD was necessary). We computed the window-level summaries of total selection in R from the estimates of total selection on SNPs obtained from our total selection model described above (version 3.5.1).

We fit discrete state, homogeneous hidden Markov models (HMMs) to delineate large-scale, contiguous regions of the genome with elevated LD or total selection (as in [90]). We used our window-based summaries of LD and total selection (see previous paragraph) as input for the HMM analyses. For LD, we modeled logit mean LD (mean r^2) for each window assuming two hidden states, a background LD state (mean set to the empirical mean, $\mu = -4.49$ on the logit scale) and an elevated LD state (mean set to the 99th percentile of the empirical distribution $\mu = -2.62$). We assumed a Gaussian error distribution in each case with a standard deviation equal to the empirical standard deviation. The Baum-Welch algorithm was used to estimate the transition rate matrix between states [91], and the Viterbi algorithm was then used to predict the most likely sequence of hidden states. We modified the R package HiddenMarkov (version 1.8.11) to allow for fixed means and standard deviations for each hidden state, and then used the modified version of this package for our analysis [92]. We fit the model with a tolerance of $1e^{-4}$ and allowing for a maximum of 500 iterations (models converged before this maximum was reached).

We fit the HMM for total selection similarly. We fit the model based on the mean log-odds selection for each window. We again used the empirical mean as the mean for the background state ($\mu = -0.576$), but instead defined the high state as having a mean 0, which corresponds to more evidence for selection than against it when averaged across all SNPs in a window. In each case, we used the empirical standard deviation ($\sigma = 0.558$). We again used our modified version of the HiddenMarkov package, and again fit the model with a tolerance of $1e^{-4}$ and a maximum of 500 iterations (convergence was achieved prior to this cutoff).

14

References

[65] Nosil, P. *et al.* Natural selection and the predictability of evolution in *Timema stick* insects. *Science* **359**, 765–770 (2018).

[66] Grant, P. R. & Grant, B. R. *40 years of evolution: Darwin's finches on Daphne Major Island* (Princeton University Press, 2014).

[67] Barrett, R. D. *et al.* Linking a mutation to survival in wild mice. *Science* **363**, 499–504 (2019).

[68] Schluter, D. *et al.* Fitness maps to a large-effect locus in introduced stickleback populations. *Dryad, Dataset* **https://doi.org/10.5061/dryad.np5hqbzrc** (2020).

[69] Schluter, D. *et al.* Fitness maps to a large-effect locus in introduced stickleback populations. *Proceedings of the National Academy of Sciences* **118** (2021).

[70] Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of *ectodysplasin* alleles. *Science* **307**, 1928–1933 (2005).

[71] Barrett, R. D., Rogers, S. M. & Schluter, D. Natural selection on a major armor gene in threespine stickleback. *Science* **322**, 255–257 (2008).

[72] O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D. & Kingsley, D. M. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *eLife* **4**, e05290 (2015).

[73] Gompert, Z. *et al.* Experimental evidence for ecological selection on genome variation in the wild. *Ecology Letters* **17**, 369–379 (2014).

[74] Nosil, P. Reproductive isolation caused by visual predation on migrants between divergent environments. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**, 1521–1528 (2004).

[75] Nosil, P. & Crespi, B. J. Experimental evidence that predation promotes divergence in adaptive radiation. *Proceedings of the National Academy of Sciences* **103**, 9090–9095 (2006).

[76] Nosil, P. & Crespi, B. J. Ecological divergence promotes the evolution of cryptic reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences* **273**, 991–997 (2006).

[77] Riesch, R. *et al.* Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution* **1**, 1–13 (2017).

[78] Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

[79] Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

[80] McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–1303 (2010).

[81] Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

[82] Nosil, P. *et al.* Ecology shapes epistasis in a genotype–phenotype–fitness map for stick insect colour. *Nature Ecology & Evolution* 1–12 (2020).

[83] Villoutreix, R. *et al.* Large-scale mutation in the evolution of a gene complex for cryptic coloration. *Science* **369**, 460–466 (2020).

[84] Lindtke, D. *et al.* Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. *Molecular Ecology* **26**, 6189–6205 (2017).

[85] Li, H. & Ralph, P. Local PCA shows how the effect of population structure differs along the genome. *Genetics* **211**, 289–304 (2019).

[86] Sved, J. Correlation measures for linkage disequilibrium within and between populations. *Genetics Research* **91**, 183–192 (2009).

[87] Galassi, M. *et al.* *GNU Scientific Library: Reference Manual* (Network Theory Ltd., 2009).

[88] Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).

[89] Sisson, S. A., Fan, Y. & Beaumont, M. *Handbook of approximate Bayesian computation* (CRC Press, 2018).

[90] Soria-Carrasco, V. *et al.* Stick insect genomes reveal natural selection's role in parallel speciation. *Science* **344**, 738–742 (2014).

[91] Baum, L. E., Petrie, T., Soules, G. & Weiss, N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics* **41**, 164–171 (1970).

[92] Harte, D. *HiddenMarkov: Hidden Markov Models*. Statistics Research Associates, Wellington (2017). URL http://www.statsresearch.co.nz/dsh/sslib/. R package version 1.8-11.
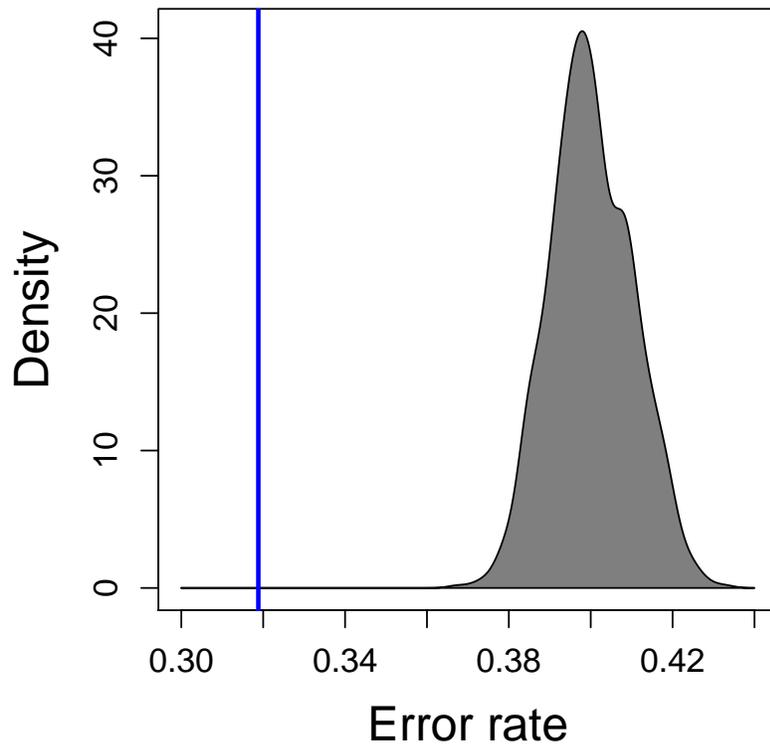
Figure 1: Plot gives the distribution of error rates for predicting individual survival over a series of null models (gray region), which is compared to the error rate for the fitted model (blue vertical line). The fitted model outperformed the null models (P < 0.001). Results are shown for *T. cristinae* transplanted to *Ceanothus*; see the main text for results on *Adenostoma*.
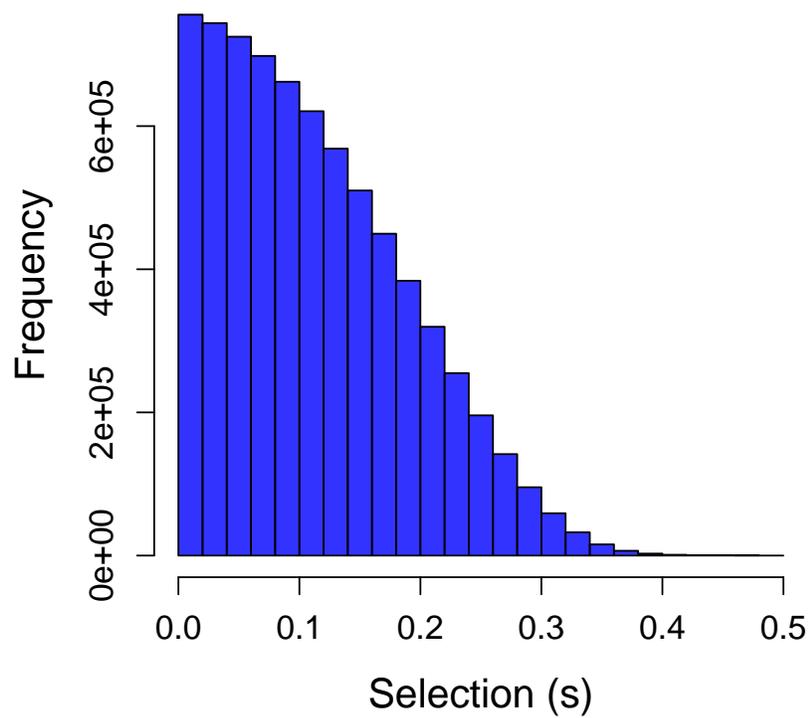
Figure 2: Histogram summarizes the distribution of point estimates of indirect selection across ~7 million SNPs. The absolute intensity of selection is summarized. As expected, selection estimates were strongly correlated with allele frequency change (Pearson $r = 0.75$, P < 0.001).
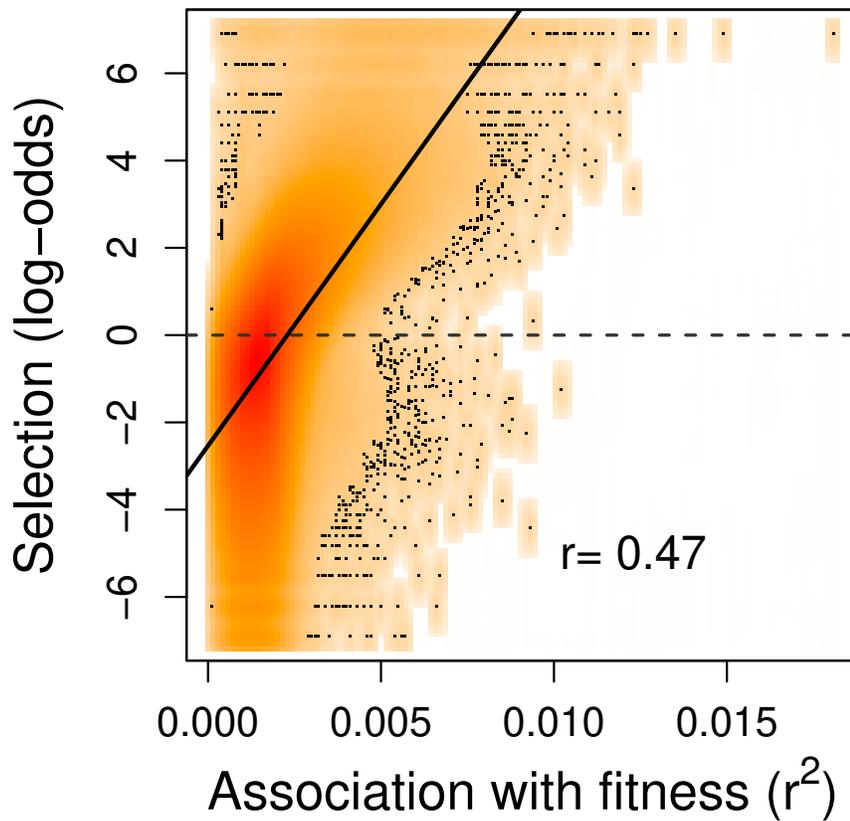
Figure 3: Plot shows that SNPs more strongly associated with expected fitness have a higher posterior probability of having experienced (indirect) selection. The heatmap depicts the relationship across the ~7 million SNPs with points for overlain in the regions of lowest density. A best fit line is shown (Pearson correlation = 0.47, r^2 = 0.22, P < 0.001). The dashed line denotes the threshold for a model of indirect selection explaining the SNP data better than a neutral model. Results are shown for *T. cristinae* transplanted to *Ceanothus*; see the main text for results on *Adenostoma*.
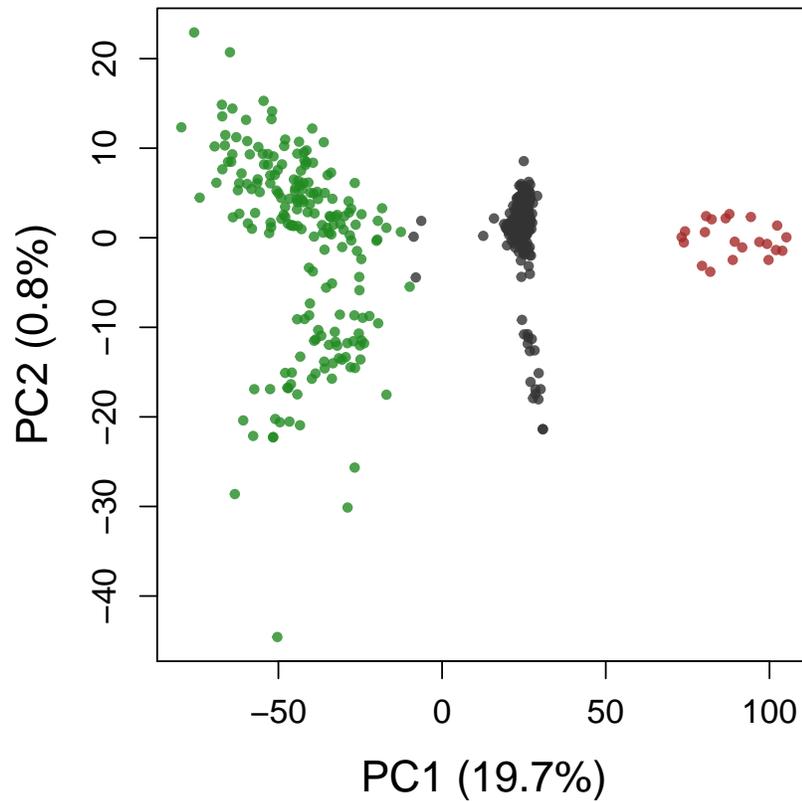
Figure 4: Statistical summary of genetic variation within the *Mel-Stripe* locus based on ~79 thousand SNPs. Points denote individuals PC1 and PC2 scores. Points are colored based on *Mel-Stripe* genotype (i.e., k-means assignment based on PC1 scores) and are colored to denote alternative inversion homozygotes (green versus brown) and heterozygotes (gray).
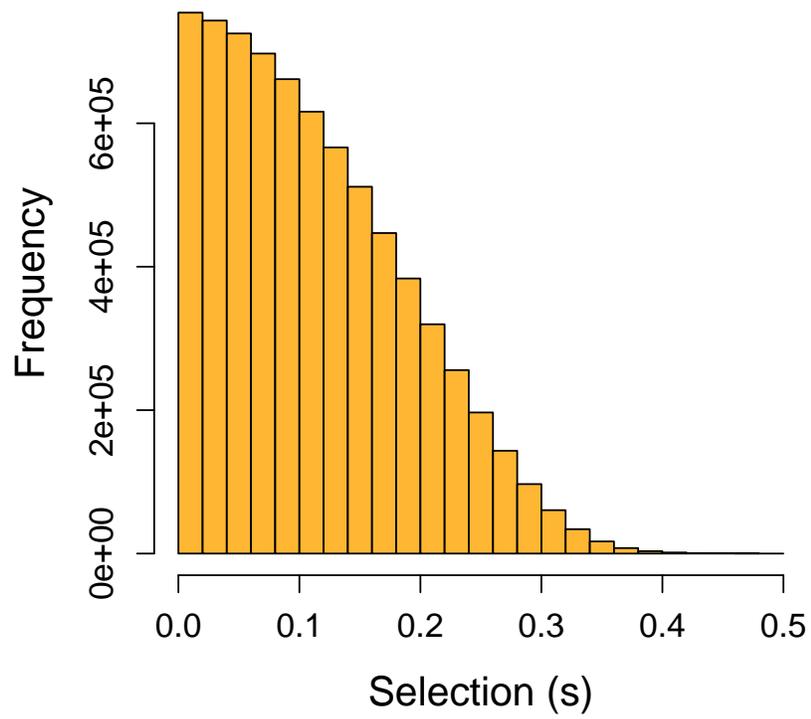
Figure 5: Histogram summarizes the distribution of point estimates of indirect selection across ~7 million SNPs. The absolute intensity of selection is summarized. As expected, selection estimates were strongly correlated with allele frequency change (Pearson $r = 0.75$, P < 0.001).