Online Supporting Information fro Gompert et al. 'Experimental evidence for ecological selection on genome variation in the wild'.

Genotype-by-sequencing (GBS): Genomic library preparation and sequencing protocols. See (Nosil *et al.* 2012) for complete details. Briefly, genomic DNA from each individual was digested in the presence of the restriction enzymes EcoRI and MseI and the resulting fragments were ligated to adaptors containing a unique 8, 9, or10-bp barcode sequence and the Illumina sequencing adaptor. These fragments were then amplified by PCR using the Illumina sequencing primers. Following PCR, we pooled all genomic libraries and size selected DNA fragments ranging in size from 250 – 400 bp by isolating the indicated range of fragments from a 2.5% agarose gel. Genomic DNA fragments collected in the gel punches were then purified using QiaQuick gel extraction kits (Qiagen, Inc.). The pooled and size-selected libraries were then sequenced on eight lanes of the Illumina HiSeq 2000 system using V3 reagents at SeqWright DNA Technology Services, Houston, TX, USA. Sequencing reads were a total of 100 bp, each consisting of an 8 to 10-bp barcode at the EcoR1 end of our amplified fragments, 6-bp corresponding to the EcoRI cut site, and 84 to 86 bp of informative genomic sequence.

Details of SNP calling and estimation of genotype and allele frequencies. We used the full prior in bcftools, only called SNPs if reads occurred in a minimum of 30% of the individuals, and required the probability of the data to be less than 0.05 under the null model that all samples were homozygous for the reference allele. We ignored insertions and deletions and removed any genetic regions where individuals appeared to have more than two haplotypes, as such regions likely correspond to multiple-copy repeats, paralogous sequences, or poor alignments. Lastly, we discarded any variable site where the observed allele counts from apparent heterozygous individuals were very unlikely given a binomial distribution with p = 0.5 to avoid any loci that did not behave as Mendelian units.

As in past work (Nosil *et al.* 2012) we used a Bayesian model and Markov chain Monte Carlo (MCMC) to estimate genotypes and allele frequencies in each experimental population. Specifically, we generated samples from the unnormalized posterior distribution $\Pr(\mathbf{g}, \mathbf{p} \mid \mathbf{x}) = \prod_i \Pr(x_{ij} \mid g_{ij}) \Pr(g_{ij} \mid p_i) \Pr(p_i)$ where x_{ij} (the sequence data) and g_{ij} (the unknown genotype) are the

number of sequence reads and gene copies containing the reference allele for locus i and individual j, and p_i is the reference allele frequency. Thus, the marginal posterior probability distribution of the allele frequencies (**p**) incorporates uncertainty in the genotypes caused by finite sequence coverage, and we obtained posterior probabilities for each genotype that reflected our allele frequency estimates. We summarized the joint posterior probability distribution of genotypes and allele frequencies based on two 15,000 iteration chains with 1,000 iteration burn-in and 5 iteration thinning intervals. We verified likely convergence to the stationary distribution by visual inspection of sample histories.

Details of whole genome sequencing and assembly. Genomic libraries were constructed from genomic DNA that was isolated from a total of 10 individual female *T. cristinae* using Qiagen DNeasy Blood and Tissue kits (Qiagen, Inc.). To minimize genetic differences that might affect assembly of sequences all individuals used in DNA extractions were collected from the same sample population during spring 2011 (population code: PC). All library construction and sequencing was carried out at BGI genomics institute, Hong Kong. Sequencing was accomplished on 7 lanes of an Illumina HiSeq 2000 machine with V3 reagents (one lane each for all libraries except the 5000 bp library, which was sequenced in two lanes). Raw sequencing reads totaled 303.073 Gb and following initial quality control sequencing reads totaled 182.015 Gb (see Table S1 for number of reads before and after initial quality control per library). Initial quality control included removing reads with greater than 5% N's or with evidence of polyA regions, removing reads where 20% or more of the calls were considered low quality bases, removing adaptor polluted reads, removing reads with overlapping reads, and removing duplicated reads. Three assemblies with different choices of mer-size (31, 47, and 71) and the software SOAPdenovo (version 1.05) resulted in many small scaffolds that were evidently under-assembled relative to expected genome size (their cumulative length exceeded the genome size from flow cytometry). This may be due to heterozygosity within and among the outbred individuals in the sequencing libraries. In contrast, assembly of the reads from the 170, 2000, and 5000 bp insert libraries with ALLPATHS-LG (version 43375) yielded a much smaller number of larger scaffolds (the 500 and 800 bp libraries cannot be used in ALLPATHS-LG). Invoking the HAPLOIDIFY=T option in ALLPATHS-LG, to better utilize reads from heterozygous individuals, yielded a longer assembly with fewer scaffolds than without this option. We used

this latter assembly for all further analyses. This assembly included 190,773 contigs in 14,221 scaffolds with the N50 for the scaffolds being 312,000bp (with gaps). The estimated fraction (total length of assembly in bp / estimated genome size from flow through cytometry) of the genome represented in this assembly is ~80% (1027063217 / (1.3*10^9) = 0.7900486).

Mapping GBS reads to the whole genome assembly. We assembled the GBS contig consensus sequences to the draft *T. cristinae* genome to determine whether allele frequency change varied among genomic regions. We performed this reference-based assembly using SeqMan Ngen (DNASTAR) with a 23bp mer size and a 90% minimum match percentage. 231,182 of the 276,968 (83%) consensus sequences for GBS contigs assembled to the draft genome, and scaffold length was highly correlated with the number of GBS sequences that assembled to a scaffold ($r = 0.96$). We mapped 155,920 of the 186,576 SNPs identified in the GBS data onto the genome assembly, and 140,549 of these SNPs mapped to 3950 scaffolds that were greater than 50,000bp in length (these large scaffolds represent 941.7 megabases of the draft genome).

Details of linkage disequilibrium analyses. We used a Monte Carlo algorithm to incorporate uncertainty in genotype into our estimates of Δ and estimated Δ separately for each experimental population. Specifically, we first sampled genotypes for each individual based on their posterior probabilities (these posterior probabilities were obtained from the allele frequency estimation model described previously). We then calculated $\Delta_{ii'}$ for each locus pair based on the sampled genotypes. We iterated this procedure 100 times for each locus pair and used the mean value of $\Delta_{ii'}$ as an estimate of $\Delta_{ii'}$. We summarized estimates of $\Delta_{ii'}$ for all 17,405,208,600 locus pairs within each population. We used C to implement this Monte Carlo estimation procedure.

Details of null models: Absence of selection on individual loci (null model 1). We used a Monte Carlo-Bayesian method to estimate the probability that the allele frequency change at each locus exceeded neutral expectations under random genotype-independent mortality.

We obtained the distribution of expected allele frequency change at each locus under the null hypothesis that survival (λ) and genotype ($\mathbf{g}$) were independent by permuting the elements of λ within each experimental population, and calculating the allele frequency change based on the

permuted survival vector (as described previously, see 'Estimation of allele frequency change'). We generated the null distribution of expected allele frequency change (Δp^{drift}) by repeatedly (1000 times) sampling genotypes (**g**) according to their posterior distributions and permuting λ. Thus, this Monte Carlo procedure incorporated uncertainty in genotypes and the stochastic nature of genetic drift. Specifically, we estimated the Monte Carlo-Bayesian posterior probability that a reference allele at each locus increased ($Pr[\Delta p_i > \Delta p^{drift}{}_i]$) or decreased ($Pr[\Delta p_i < \Delta p^{drift}{}_i]$) in frequency more than expected under null model 1. Here Δp_i and $\Delta p^{drift}i$ denote the observed and null expectations for allele frequency change at locus i. We then defined a selection index $S^{index}{}_i$ = max($Pr[\Delta p_i > \Delta p^{drift}{}_i]$, $Pr[\Delta p_i < \Delta p^{drift}{}_i]$), and equated 'exceptional change' at a locus with a selection index of 97.5 or greater (this is equivalent to a two-tailed probability of 95% or more that the allele frequency change was greater than expected under null model 1). Thus, exceptional allele frequency change indicates a significant statistical association between an individual's genotype at a locus and survival in the experiment.

Parallel evolution is often used to infer selection, as it is unlikely to arise by drift. We thus estimated the probability that each locus exhibited exceptional parallel allele frequency change at the treatment level (i.e., transplant to *Adenostoma* or *Ceanothus*). Specifically, we summed Δp_i and $\Delta p^{drift}{}_i$ across the set of experimental populations (we still limited permutations of λ to individuals within the same experimental population to maintain population-specific survival rates). We then estimated $Pr(\Delta p_i > \Delta p^{drift}{}_i)$, $Pr(\Delta p_i < \Delta p^{drift}{}_i)$, calculated selection indexes, and identified loci with exceptional allele frequency change as previously described. This is a more powerful test for host plant-dependent selection as it combines evidence from allele frequency change across each set of five replicate populations.

Absence of selection on any loci (null model 2). Next, we asked whether the number of loci that exhibited exceptional allele frequency change within an experimental population or treatment was different than expected if survival was wholly independent of genotype. This is a null model of genome wide genetic drift. We did this analysis considering number of loci exhibiting exceptional change by generating 100 permutations of the survival vector λ (at the level of individuals, and thus genotypes at all loci within individuals), and estimating the number of loci exhibiting exceptional allele frequency change based on each of these permuted data sets

using the methods described in the previous paragraphs. In other words we treated each permuted data set as the observed data and estimated the number of loci with exceptional allele frequency change using 1000 additional permutations of each permuted survival vector. This procedure generated a null distribution for the number of loci exhibiting exceptional allele frequency change.

Details of selection coefficient estimation. We assumed that whether individual j survived (λ_j = {0, 1} represented a Bernoulli trial with p = w_j. We specified a linear model for the absolute fitness of individual j in population k with genotype g_{ij} at locus i as:

$w_j = \mu_k (1 + \sigma_i)$ if $g_{ij} = 0$
$w_j = \mu_k$ if $g_{ij} = 1$
$w_j = \mu_k (1 - \sigma_i)$ if $g_{ij} = 2$

This is an additive model for fitness where σ is half the difference in absolute fitness between the two homozygous genotypes, and μ_k is the absolute fitness of heterozygous individuals in population k. We defined the selection response for locus i as $s_i = 1/K \sum_k 2 \mu_k \sigma_i$, which is the average difference in mean fitness between the two homozygous genotypes in the ten experimental populations. We specified an independent uniform prior on each μ_k, U(0,1). Rather than specify a prior on σ_i we placed a prior on the ratio of the absolute fitness of the two homozygous genotypes $\alpha_i = (\mu_k (1 + \sigma_i)) / (\mu_k (1 - \sigma_i))$. We placed a prior on α_i because, unlike the difference in absolute fitness between the two homozygous genotypes, α_i is independent of μ_k and consequently does not vary among populations. We assumed the following prior for α_i, which is equivalent to specifying independent uniform priors on the fitness of each homozygous genotype:

$P(\alpha_i) = 0.5$ if $0 < \alpha_i < 1$
$P(\alpha_i) = 1 / (2 * \alpha_i^2)$ if $\alpha_i >= 1$

We used Markov chain Monte Carlo (MCMC) to obtain a sample from the joint posterior probability distribution of the model parameters for each locus $P(\alpha_i, \mu_k \mid \lambda \mathbf{g}) \propto \Pi_j$

P(λ _j | α_i, μ, g_ij) Π_k P(μ_k) P(α_i) P(g_ij). Note that P(g_ij) denotes the genotype posterior probability distribution. We developed software using the C++ language and the GNU Scientific Library to implement a MCMC algorithm for this model. We ran three 10,000 iteration chains with 5,000 iteration burn-ins and 5 iteration thinning intervals to estimate the model parameters for each SNP. We generated posterior probability distributions for the *s*_i based on posterior samples of α_i and μ_k. We assessed convergence to the stationary distribution using quantitative convergence diagnostics.

Details of tests for genomic clustering. We tested whether the frequency of SNPs with exceptional allele frequency change varied among that scaffolds comprising the draft genome assembly (we restricted all these analyses to the 3950 largest scaffolds in the assembly, which were all >50,000bp in length). We used Bayesian methods to determine whether the frequency of SNPs with exceptional change on each scaffold better fit a model with a binomial sampling distribution with (i) each scaffold characterized by the same probability parameter (constrained model), or (ii) each scaffold characterized by its own probability parameter (hierarchical model). The hierarchical model should be favored if some regions of the genome were affected more by selection than others (notably, the binomial likelihood includes the number of SNPs on a scaffold, and consequently our method accounts for the effect of scaffold length on the number of SNPs putatively affected by selection on a scaffold).

We placed an uninformative Beta prior (Jeffery's prior) on the single probability parameter in the constrained model, but assumed p_i ~ Beta(π V, (1 - π) V)), π ~ Beta(0.5, 0.5), and V ~ Uniform(1e-2, 1e6) for the hierarchical model. We contrasted the two models for the probability of exceptional change using deviance information criterion (DIC). We fit models and estimated DIC using the rjags interface with the JAGS software. Specifically, we compared full non-hierarchical generalized linear models with each scaffold allowed its own binomial parameter to reduced models with a single binomial parameter shared by all scaffolds.

Results of tests for genomic clustering. Bayesian model comparison methods indicated that, after accounting for the number of loci on each scaffold, the genome-wide distribution of loci with exceptional allele frequency change was better explained by a model with a binomial

sampling distribution with each scaffold characterized by the same probability parameter (constrained model of 'no genomic clustering', D = Deviance =946.9, pD = effective number of parameters = 1.0, DIC = 947.9) than a model with each scaffold characterized by its own probability parameter (hierarchical model of 'genomic clustering', D = 920.2, pD = 31.6, DIC = 951.8, DIC difference = 3.89 ± 2.77).

Connecting genotypic and phenotypic changes. These analyses were restricted to recaptured individuals (i.e., survivors) because individual phenotypes and genotypes of released individuals could not be matched up (i.e., tissues from released individuals were not stored in individual tubes with corresponding phenotypic information, whereas those from recaptured individuals were). We first asked whether the SNPs most strongly associated with the stripe phenotype in the survivors showed greater evidence of exceptional allele frequency change than random sets of SNPs. We chose the 10 SNPs with the greatest allele frequency difference between striped and green stick insects (mean allele frequency difference = 0.173). We quantified the joint evidence of selection on this set of 10 loci as the sum of selection indexes over these loci. We then randomly sampled other sets of 10 loci and calculated the joint evidence of selection on these sets of loci to obtain a null distribution (we generated 10,000 random sets). We then asked whether any of the *Ceanothus* exceptional parallel change loci were associated with the stripe phenotype. We did this by calculating the difference in allele frequency (or equivalently, mean genotype) between striped and non-striped individuals. We then used a permutation test (10,000 permutations) to test whether these allele frequencies differences were greater than expected under the null hypothesis of no association between genotype at each locus and stripe phenotype.

Tests for dispersal. In addition to sampling individual experimental plants, we also sampled the five nearest suitable host plants for the presence of experimental *T. cristinae*. Surrounding plants ranged from 2.2-7.6m from the experimental plants. Of the 50 surrounding host plants we sampled, only a single experimental *T. cristinae* was captured, indicating limited dispersal off of experimental plants, as documented in previous studies (Nosil 2004; Nosil & Crespi 2006).

Lab and field-based tests for potential effects of tissue sampling. We tested whether our method of non-lethal tissue sampling influenced survival by comparing the survival of individual

T. cristinae that were versus were not subjected to tissue sampling in a controlled laboratory experiment. Paired replicates were set up using individuals of a similar size and age and 6 – 20 individuals were used per treatment in each replicate pair (5 replicate pairs total). We found no effect of tissue sampling on survival probability after either three days or six days (three days: mean proportion of tissue-sampled individuals surviving = 0.85, mean proportion of non-tissue-sampled individuals surviving = 0.87, paired t-test, t = -1.00, d.f. = 4, $p = 0.37$; six days: mean proportion of tissue-sampled individuals surviving = 0.50, mean proportion of non-tissue-sampled individuals surviving = 0.59, t=-2.06, d.f. = 4, $p = 0.11$).

We then tested whether tissue sampling had an effect on recapture probability in the field. To do so, we placed 20 individual *T. cristinae* that had a tissue sample taken (treatment individuals, same protocol as main experiment above) and 20 individuals that did not (controls) onto one each of an *Adenostoma* and *Ceanothus* host plant near our experimental site. All individuals were each marked on the abdomen with a Sharpie pen to enable identification upon recapture. Individuals were placed on their respective bush and recaptured six days later using the same methods as described for the main field experiment. On the *Adenostoma* bush we recaptured nine treatment and no control individuals and on the *Ceanothus* bush we recaptured six treatment and three control individuals. The direction of these trends indicates that the method of tissue sampling we employed did not negatively influence recapture probabilities in the wild. Thus, tissue sampling is unlikely to have caused the non-host related selection in the experiment. This argument is bolstered by the additional lines of evidence for selection and treatment effects in the main field experiment (e.g., correlations of selection response with host plant and elevation – these are unlikely to arise by drift and treatment effects cannot be explained by taking a tissue sample because animals in all treatments were treated in the same manner) outlined in the main text.

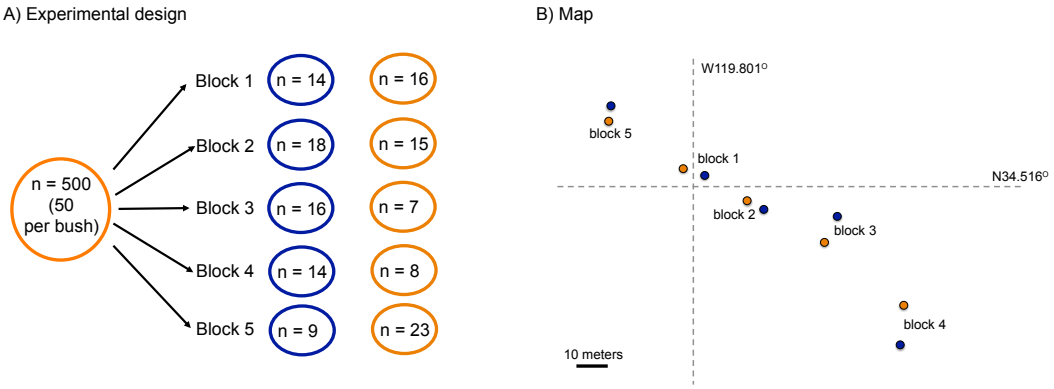

Figure S1. Description of the experiment. Blue circles represent *Ceanothus* and orange circles *Adenostoma.* A) A total of 500 individuals were transplanted into five replicate blocks. Numbers within circles denote number of recaptured individuals (out of 50 released). B) Map of the experiment.

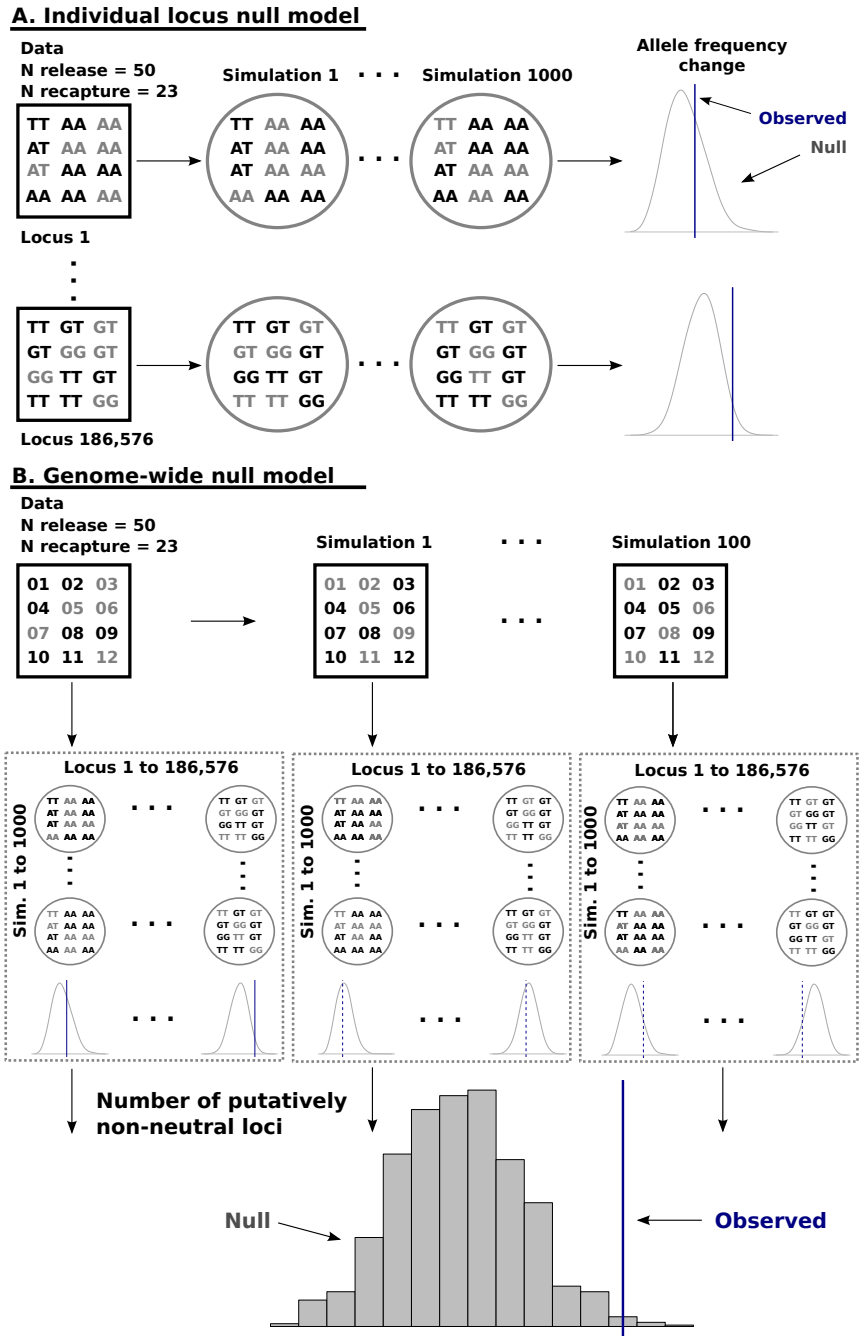

Figure S2. Schematic of the null models of the absence of selection (null models 1 and 2). Each box or circle contains the genetic data for a hypothetical locus. Black numbers or letters denote individuals that survived and gray characters denote individuals that lived. A) In null model 1 the squares denote the true survivors and circles denote simulated sets of survivors.

Density plots show the simulated distribution of allele frequency change for each locus and the vertical blue lines denote the observed allele frequency change. We reject the null hypothesis in model 1 for loci where the observed change (blue line) is an extreme tail of this null distribution. B) In null model 2 the real and replicate simulated data sets are show in boxes and numbers rather than genotypes are used to highlight the fact that the survival data is applied to all loci. Each of these data sets (the true data set and 100 replicate simulated data sets) are subjected to the null model 1 analysis (i.e., pane A) and the number of exceptional allele frequency change loci designated from the true survival data (vertical blue line) and replicate simulated data sets (gray histogram) are tabulated to determine whether the null hypothesis of no selection across the genome (null model 2) can be rejected.

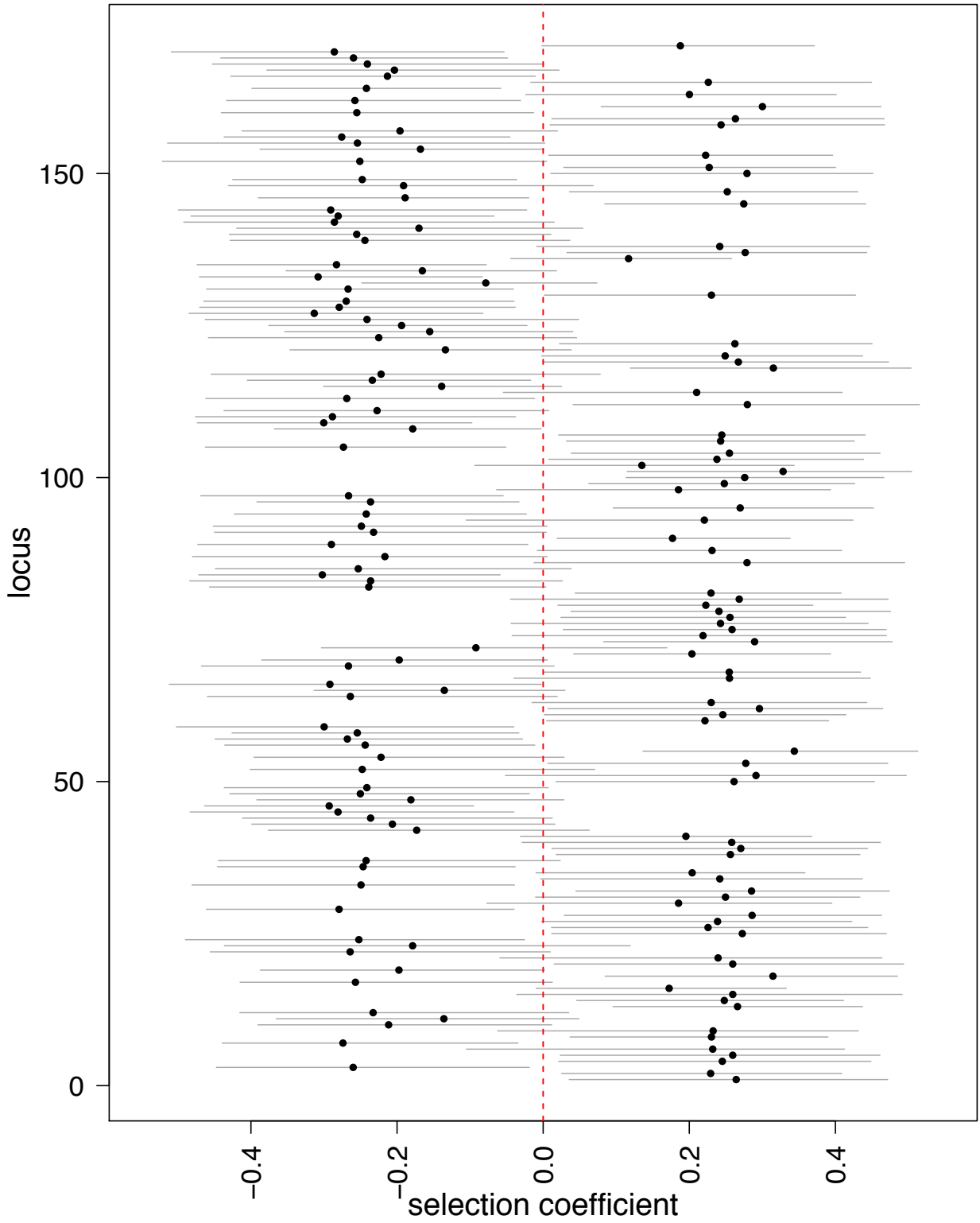

Figure S3. Selection coefficients and 95% credible intervals for the 171 loci exhibiting exceptional parallel change in the experimental populations transplanted to *Adenostoma*.

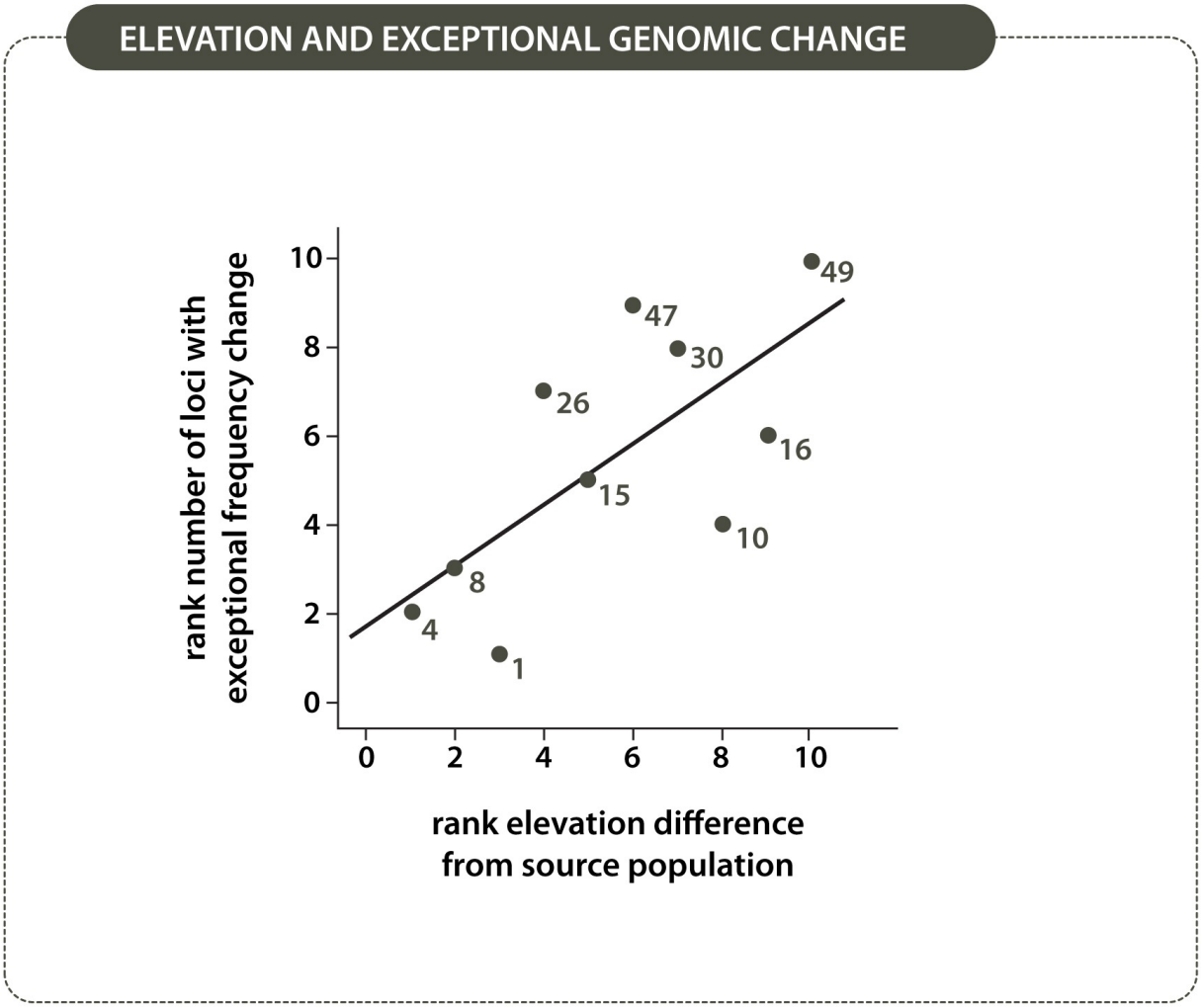

Figure S4. Elevation and genomic change. Show is the relationship between the number of loci with exceptional change within individual experimental populations and the difference in elevation between an experimental population and the source population. Rank correlation was used due to highly non-normal distributions.

Table S1. Total number of sequenced bases per library construction both before and after initial quality control (QC) steps.

Library (size in bp)	Average read length (bp)	Total nucleotides sequenced (Gb)	Nucleotides sequenced following initial QC (Gb)
170	100	42.171	38.148
500	100	73.09	58.219
800	100	33.634	26.720
2000a	100	40.446	19.888
2000b	100	41.923	16.755
5000	100	71.809	22.285
Total		303.073	182.015

Table S2. Morph frequencies (i.e., % striped individuals) upon release and recapture for each of the ten experimental populations.

	Adenostoma treatment		*Ceanothus* treatment	
	% striped upon release	% striped upon recapture	% striped upon release	% striped upon recapture
Block 1	82	94	74	71
Block 2	74	93	74	78
Block 3	82	86	84	88
Block 4	80	100	74	64
Block 5	76	82	82	78

Table S3. Minor allele frequency distribution (in each experimental population).

Population	Mean	s.d.	1%	25%	50%	75%	99%
1A	0.108	0.111	0.006	0.032	0.067	0.142	0.472
1C	0.108	0.111	0.006	0.033	0.067	0.141	0.473
2A	0.109	0.111	0.006	0.033	0.067	0.141	0.473
2C	0.108	0.111	0.006	0.032	0.067	0.142	0.473
3A	0.109	0.111	0.007	0.03	0.068	0.144	0.472
3C	0.109	0.111	0.007	0.03	0.067	0.143	0.472
4A	0.108	0.111	0.006	0.03	0.066	0.142	0.472
4C	0.109	0.111	0.007	0.03	0.068	0.144	0.473
5A	0.108	0.111	0.007	0.03	0.067	0.143	0.472
5C	0.108	0.111	0.007	0.03	0.067	0.143	0.473

Table S4. Fixation of alleles in the experiment. Shown for loci with minor allele frequency >5% at the onset of the experiment the number of SNPs where one allele was fixed during the course of the experiment within individual populations, blocks, treatments, and across the entire experiment. The analyses account for uncertainty in genotype and allele frequencies. Note that many more loci would exhibit fixation, including some at the level of the whole experiment, if SNPs with lower starting minor allele frequencies were considered. Also shown are median (and quantiles) and maximum allele frequencies for loci that fixed in the experiment (see also Figure 3b). A = *Adenostoma*. C = *Ceanothus*. Treat. = treatment.

comparison	median number	5% lower bound	95% upper bound	median allele frequency	75th quantile allele frequency	95th quantile allele frequency	maximum allele frequency
1A	3903	3789	4017	0.06	0.07	0.10	0.31
1C	4687	4569	4807	0.06	0.07	0.10	0.33
2A	4607	4499	4726	0.06	0.07	0.10	0.33
2C	3119	3017	3219	0.06	0.07	0.10	0.28
3A	20967	20743	21214	0.07	0.09	0.15	0.62
3C	3844	3727	3957	0.06	0.07	0.10	0.31
4A	17564	17359	17757	0.07	0.07	0.14	0.56
4C	5684	5548	5812	0.06	0.07	0.11	0.35
5A	922	866	980	0.05	0.06	0.08	0.21
5C	14261	14063	14462	0.06	0.08	0.13	0.52
Treat. A	4	1	8	0.05	0.06	0.07	0.09
Treat. C	2	0	5	0.05	0.06	0.06	0.09

Online SI references

Nosil P. (2004). Reproductive isolation caused by visual predation on migrants between divergent environments. *Proc. R. Soc. B-Biol. Sci.*, 271, 1521-1528.

Nosil P. & Crespi B.J. (2006). Experimental evidence that predation promotes divergence in adaptive radiation. *Proc. Natl. Acad. Sci. U. S. A.*, 103, 9090-9095.

Nosil P., Gompert Z., Farkas T.E., Comeault A.A., Feder J.L., Buerkle C.A. & Parchman T.L. (2012). Genomic consequences of multiple speciation processes in a stick insect. *Proc. R. Soc. B-Biol. Sci.*, 279, 5058-5065.