PNAS
www.pnas.org

Supplementary Information for

Complex evolutionary processes maintain an ancient chromosomal inversion

Patrik Nosil, Victor Soria-Corrasco, Romain Villoutreix, Marisol De la Mora, Clarissa F. de Carvalho, Thomas Parchman, Jeffrey L. Feder, Zachariah Gompert

Patrik Nosil and Zachariah Gompert
E-mail: patrik.nosil@cefe.cnrs.fr and zach.gompert@usu.edu

This PDF file includes:

Supporting Information Text

Materials and Methods

Genome annotation. We annotated the *Timema knulli* genome using the `BRAKER2` (version 2.1.6) pipeline (1). Prior to genome annotation, we identified and masked repeat sequences using `RepeatMasker` (version 4.0.7). This was done using more sensitive slow search option, the NCBI search engine, and an existing repeat library developed for *Timema* stick insects (2). Repetitive regions were soft masked (set to lowercase letters).

`BRAKER2` was used to automate training of the gene predictions tools `GeneMark-EP+` (3) and `AUGUSTUS` (4, 5) using protein homology information. The pipeline begins with self-training using `GeneMark-ES` (6) to create an initial set of seed genes as described in the `ProtHint` pipeline, which uses `DIAMOND` (v0.9.24.125) and `Spaln` (version 2.3.3d) (3, 7, 8) and a protein data base. We used the set of 2,601,995 arthropod proteins from `OrthDB` (arthropod data set version 10, `odb10_arthropoda_fasta.tar.gz`) (9) as protein evidence. Output from this pipeline is then used for iterative model training with `GeneMark-EP+` (3). A set of anchored genes from `GeneMark-EP+` are then used to train `AUGUSTUS` (version 3.5.0) and predict the final set of genes. Functional annotation of the predicted *T. knulli* genes, specifically each CDS (coding sequence) was conducted using `InterProScan` (version 5.60-92.0) (10). Gene density was summarized based on the gene (GENE) annotations. Here, we determined the number of annotated genes in 5 megabase pair (mbp) windows across each *T. knulli* chromosome and then for the bounds of the *Perform* locus (the lower and upper bound plus or minus 5 mbps) and the remainder of the *Perform* locus.

Assigning chromosome numbers. To identify chromosomes (scaffolds) homologous to the *T. cristinae* linkage groups in our *T. knulli* genome, we first compared the *T. cristinae* reference plus linkage map to a more recent yet published *T. cristinae* genome from a green striped stick insect, which was constructed based on proximity ligation of DNA in chromatin and reconstituted chromatin (Hi-C) and comprised 13 large scaffolds, each corresponding to one of the 13 *T. cristinae* chromosomes (2). Specifically, we constructed a blast database from the 13 scaffolds of the newer (green striped) genome and then identified homologous scaffolds from the older melanic genome (and linkage map) by blasting each of these scaffolds against the database. This was done with `blastn` (version 2.11.0) with a minimum e-value of $1e^{-50}$ and a minimum percent identity of 92. Only matches of >10,000 bps were considered (11). Then, in `R` (4.0.2), we computed the total length of matches between each of the 13 linkage groups from the melanic *T. cristinae* genome and the 13 large scaffolds from the newer, green striped genome. In most cases, there was an unambiguous correspondence between linkage groups and chromosome scaffolds. However, our linkage groups 9 and 13 were under-assembled on the linkage map as both corresponded to a single scaffold, and much of the new scaffold 14101 was not mapped to any linkage group. Thus, our old linkage groups 9 and 13 were combined and are hereafter referred to as chromosome 9, and our new scaffold 14101 was denoted chromosome 13 (Table S2).

We then used `cactus` (version 1.0.0) to align the *T. knulli* genome to the green striped *T. cristinae* genome (12, 13). For this, we first used `RepeatMasker` (version 4.0.7) to mask repetitive regions of the genome (14); this was done using a repeat library developed for *Timema* (2). We then performed a pairwise alignment between the genomes with `cactus`. The `HalSynteny` tool was then used to extract syntenic alignment blocks from the comparative alignment (15) (https://github.com/ComparativeGenomicsToolkit/hal). We then identified homologous chromosomes by summing the total length of syntenic segments between each *T. cristinae* and *T. knulli* genome.

Nanopore sequencing and structural variant calling in *T. knulli*. We used Oxford Nanopore long-read sequencing (16) to obtain additional evidence that the *Perform* locus is a segregating inversion within *T. knulli*. We chose this approach as we expected long DNA sequence reads to have a substantial chance of spanning and accurately detecting the expected large inversion (17). To do this, we extracted high-molecular weight DNA from a single *T. knulli* collected from BCE C (on *Ceanothus*) where the *Ceanothus* allele (that is the expected ancestral, non-inverted allele) occurs at high frequency. This was done with Qiagen's MagAttract HMW DNA kit (Qiagen, Inc.) in accordance with the manufacturer's protocol. We extracted DNA from two samples taken from the thorax of this individual, which yielded 803 and 1018 nanograms of DNA respectively on a dsDNA HS (high sensitivity) assay with a Qubit f4 fluorometer (Thermo Fisher). We then repaired and polished the DNA molecules with the NEBNEXT FFPE DNA Repair Mix and NEBNEXT Ultra II end-repair/dA-tailing module in accordance with Oxford Nanopore's suggested protocol. The two DNA samples were then pooled and adaptor oligos for sequencing were added with the Oxford Nanopore ligation sequencing kit (SQK-LSK109). We sequenced the resulting library on a R9.4 flow cell with a MiniION using a 72 hour run time. We used `guppy_basecaller` (version 6.1.7_gpu) to call nucleotides from the raw output. This generated 471,648 sequences with a total length of 863 megabases (about 0.5× genome coverage). Note that while this is low coverage, it proved sufficient to validate the expected inversion as described below.

We first used `NanoFilt` (18) to remove bases with quality scores less than 6 and then aligned the filtered nanopore DNA sequences to the *T. knulli* reference genome with `minimap2` (version 2.23-r1117) (19). We used the preset option for mapping Nanopore reads against a reference (`-x map-ont`) and used soft clipping for supplementary alignments. `samtools` (version 1.12) was used to compress, sort and index the alignments (20). We then used `sniffles2` (version 2.0.3) to call structural variants (21). We required an alignment length of at least 100 bps, a mapping quality of at least 15, and a minimum structural variant length of 35 bps supported by at least one read for variant calling. We then focused specifically on inversions on chromosome 11 that were 1 mbp or greater in length; there were five of these, one of which spanned the *Perform* locus (see Figure S6 for details).

Population genomic analyses of LD and heterozygosity. We summarized patterns of linkage disequilibrium (LD) and observed heterozygosity on chromosome 11 within and outside of the *Perform* locus to further test the hypothesis that the *Perform* locus is a polymorphic inversion within *T. knulli*. For LD, we focused on common SNPs (minor allele frequency > 0.05) on chromosome 11 in the BCE population (where Redwood and *Ceanothus* co-occur). We then computed pairwise LD, as the squared genotypic correlation (r^2) for each pair of the 1537 SNPs that met the above criteria. This was done separately for $P_{RW}P_{RW}$ homozygotes, $P_{RW}P_C$ heterozygotes, P_CP_C homozygotes, and the combination of $P_{RW}P_{RW}$ and P_CP_C homozygotes. If *Perform* is a segregating inversion, we would expect elevated LD within *Perform* for the combined sample of $P_{RW}P_{RW}$ and P_CP_C homozygotes relative to other parts of chromosome 11 and samples comprising only a single genotype. Patterns of LD were summarized using heatmaps. This was done in R (version 4.1.3). For observed heterozygosity, we first converted our SNP genotype estimates to the nearest integer value, where 1 thus denotes a heterozygote for a SNP. We then computed the mean heterozygosity in each *Perform* genotype cluster ($P_{RW}P_{RW}$ homozygotes, $P_{RW}P_C$ heterozygotes, and P_CP_C homozygotes) for each SNP on chromosome 11 (including the *Preform* locus). We expect elevated heterozygosity for $P_{RW}P_C$ heterozygotes within the *Perform* locus if this locus is a segregating inversion in *T. knulli*. This analysis was conducted in R (version 4.1.3) as well.

ABC inference of gene flow and selection. We used approximate Bayesian computation (ABC) to fit and compare alternative models for selection with gene flow in the *T. knulli*-*Ceanothus*-Redwood system (22, 23). We first fit a Bayesian F-model to estimate (putative neutral) migration rates, Nm (number of migrants per generation), between our three main populations: BCE C (BCE on *Ceanothus*), BCE RW (BCE on Redwood; parapatric with BCE C) and BCTURN (on *Ceanothus*, allopatric with respect to BCE C and BCE RW) (24, 25). This statistical model approximates several population genetic models, including an island model of drift-gene flow equilibrium (26–29). Estimates of gene flow were based on allele frequencies in each population, but excluding chromosome 11 (i.e., the chromosome harboring the *Perform* locus). For this analysis, we placed Cauchy priors on Nm (the number of migrants) with bounds of 0 and 50, a location parameter of 0 and a scale parameter of 10, and Jeffery's beta priors on the migrant allele frequencies (lower bounds = 0, upper bounds = 1, $\alpha = 0.5$, $\beta = 0.5$). We fit this model in R using Hamiltonian Monte Carlo via the `R` interface with `STAN` (`rstan` version 2.21.2) (30). Posteriors were inferred from 10 independent Markov chain Monte Carlo (MCMC) analyses, with each chain using a random subset of 5000 (out of 62,093) SNPs (this was done to increase computation speed and reduce LD among loci). For each run, we used 4 independent chains, each comprising 2000 iterations and a 1000 iteration burnin. The No-U-Turn sampler (NUTS) was used for updates (31). The Gelman-Rubin convergence diagnostic was computed to verify likely convergence of the MCMC algorithm to the posterior distribution.

We next fit ABC models for selection on *Perform*, with gene flow based on our estimates from the F-model described in the preceding section. Our goal here was to compare models of divergent selection (directional selection in opposing directions on different hosts) to heterozygote advantage while accounting for drift and gene flow, and to estimate the strength of selection under these models. Here, we assumed three populations, BCE C (on *Ceanothus*), BCE RW (on Redwood) and BCTURN (on *Ceanothus*) with host-dependent selection on *Perform*, that is, we assumed one set of selection coefficients for BCE C and BCTURN and a second set of selection coefficients for BCE RW. We allowed for one of two models for selection on each host: (i) directional selection, where one homozygote was the most fit, or (ii) heterozygote advantage (overdominance), where the *Perform* heterozygotes were the most fit. With directional selection, we assumed $w_{11} = 1 + s$, $w_{12} = 1 + hs$, and $w_{22} = 1$, where w_{11}, w_{12} and w_{22} are relative fitnesses for the *Perform* genotypes, s is the selection coefficient, and h is the heterozygote effect, and w_{11} refers to the genotype that was more fit on Redwood in the experiments and that was at higher frequency in BCE RW (i.e., the $P_{RW}P_{RW}$ homozygote). We placed uniform priors on h (lower bounds = 0 upper bounds = 1) and log uniform priors on the absolute value of s with bounds of 0.001 and 0.9 (-6.91 and -0.11 on the natural-log scale). We assumed s was positive on Redwood and negative on *Ceanothus* (i.e., alternative homozygotes favored on each host). For heterozygote advantage (i.e., overdominance), we assumed $w_{11} = 1 - s1$, $w_{12} = 1$, and $w_{22} = 1 - s2$, where $s1$ and $s2$ denote the decrease in relative fitness of the two alternative homozygotes ($P_{RW}P_{RW}$ and P_CP_C, respectively). We used the same log-uniform priors on $s1$ and $s2$ as were used for s, with the added constraint of $s1 < s2$ on Redwood and $s1 > s2$ on *Ceanothus*. We placed equal prior probabilities of directional selection versus heterozygote advantage on each host (i.e., 0.5 each) and allowed for the models to differ on the two hosts.

We modeled evolution following a generalized Wright-Fisher model with selection and gene flow. Specifically, the expected allele frequency change at *Perform* for each population was $E[\Delta p] = \Delta p_s + \Delta p_m$, where Δp_s and Δp_m are the expected change caused by selection and gene flow respectively. We assumed $\Delta p_s = sp(1-p)[p + h(1-2p)]$ for directional selection or $\Delta p_s = p(1-p)[s2 - p(s1 + s2)]$ for heterozygote advantage, and $\Delta p_m = m_{ba}(p_a - p_b) + m_{ca}(p_a - p_c)$ where m_{ba} and m_{ca} are the migration rates (proportions) from populations b and c to population p_a, and p_b and p_c are the corresponding migrant and source population allele frequencies (32). We then assumed that the actual allele frequency in each population following selection, gene flow and drift was $p_{t+1} \sim \text{binomial}(p = p_t + E[\Delta p], 2N_e)$, where N_e is the variance effective population size for the relevant population (BCE C, BCE RW, or BCTURN). We did not attempt to estimate N_e, but rather to integrate over uncertainty in contemporary N_e (i.e., we treat this as a nuisance parameter). Specifically, we assumed re-scaled beta priors on N_e for each population with a lower bound of 50, an upper bound of 1000 and α and β set to 6 (symmetrical about the mean of $N_e = 525$ and relatively flat over the range). We allowed for asymmetric gene flow with expectations set by the neutral gene flow Bayesian F-model defined above. Specifically, for population pair i and j, we assumed re-scaled beta priors on Nm_{ij} and Nm_{ji} with lower and upper bounds set to the 2.5th and 97.5th percentiles of the posterior from the neutral F-model and α and β set to 10 (again symmetrical and relatively but slightly less flat over the range). We then solved for, e.g.,

m_{ij} as $m_{ij} = \frac{Nm_{ij}}{N_e}$ (with N_e denoting N_e for population j).

We conducted 25 million simulations of evolution to estimate the model and parameter posterior probabilities. In each case, the selection models and all relevant parameters were sampled from their priors. We then simulated evolution for 2500 generations starting from *Perform* allele frequencies of 0.5 for all populations (this was sufficient time to remove sensitivity to our initial allele frequency but not so long to ensure one allele was lost, as will always ultimately be the case given sufficient time without recurrent mutation). Simulations were performed using a custom program written in `C++` with functions from the Gnu Scientific Library (33). We used the vector of final (at generation 2500) *Perform* allele frequencies for the three populations as the output (summary statistics) from the simulations. Thus, the allele frequency vectors from the 25 million simulations were compared to the actual *Perform* allele frequency vector for the three populations. Using the rejection algorithm from the `abc R` package (version 2.1; R version 4.0.2) (34) we identified the 0.004% (1000 out of 25 million) of simulations resulting in the smallest Euclidean distance between the simulated and observed allele frequencies. Model posteriors were computed as the proportion of these retained simulations arising from each model (i.e., each combination of directional selection versus heterozygote advantage for the two host plants). The model of heterozygote advantage on both host plants had the highest posterior probability. Consequently, we estimated the selection coefficients ($s1$ and $s2$) on each host plant under the heterozygote advantage model (model-averaging is not appropriate as the selection coefficients do not have a consistent definition across models). This was done by considering the 0.015% (∼1000) of simulations with heterozygote advantage on both hosts with smallest distance between observed and simulated summary statistics, and performing ridge regression for parameter adjustment on the log-transformed selection coefficients. This was also done with the `abc R` package (version 2.1) (34).

Dating the chromosomal inversion with δaδi. We estimated the divergence time between the *Perform* chromosomal variants in a population genetic context with the diffusion approximation approach implemented in `δaδi` (35). We specifically followed an approach inspired by (36), which modeled recombination between subgenomes (in polyploids) as being analogous to gene flow between populations. We focused on the BCE population and designated two "populations", each comprising individuals homozygous for one of the *Perform* inversion alleles. We assumed these populations were descended from a single ancestral population with a mutation-scaled effective population size of θ ($4N_{anc}\mu$, where μ is the locus mutation rate) that diverged at time T_{split} (measured in $2N_{anc}$ generations), which corresponds with the origin time for the inversion (i.e., the creation of the two distinct inversion haplotypes). We allowed for the relative effective sizes of the two inversion "populations" to increase or decrease over time based on population growth parameters $\nu_1 = N_1/N_{anc}$ and $\nu_2 = N_2/N_{anc}$); this could reflect selection or drift in inversion allele frequencies. We modeled potential genetic exchange (recombination or gene conversion) between inversion alleles (populations) using the migration rate parameter from `δaδi`, $M_{12} = 2N_{anc}m_{12}$ and $M_{21} = 2N_{anc}m_{21}$ (here m_{12} and m_{21} are proportions), as in (36). Thus, our estimate of the divergence time between inversion alleles accounts for possible reduced DNA sequence divergence resulting from recombination within the inversion.

We used `δaδi` (with Python 3.9.7) to first infer the joint site frequency spectrum for our two populations, one comprising 25 $P_{RW}P_{RW}$ homozygotes (the allele more common on Redwood) and one comprising 33 P_CP_C homozygotes (the all more common on *Ceanothus*) (all from BCE); this was done within `δaδi` directly from the filtered vcf file. We down-sampled the data at this stage to 70% of the smaller size (i.e., 70% of 25 diploids). We then used `δaδi` to estimate the model parameters, specifically θ, T_{split}, ν_1, ν_2, and the genetic exchange parameters $M_{12} = 2N_{anc}m_{12}$ and $M_{21}2N_{anc}m_{21}$ (here m_{12} and m_{21} are proportions) (see https://github.com/zgompert/TimemaRW/blob/main/im_dadi_old.py). We used three rounds of numerical optimization, comprising 20, 10 and five iterations each to estimate the model parameters.

We then used the average of two published per-base mutation rates for insects, $2.9e^{-9}$ for *Heliconius* and $2.8e^{-9}$ for *Drosophila* (37), to convert our estimates of divergence time to time in years (or equivalently generations as *Timema* are univoltine). This conversion also required an estimate of the number of sequenced bases for the *Perform* locus so that we could compute the per-locus mutation rate (μ in `δaδi`). Importantly, this is not the same as the total length of the locus as not all bases were sequenced, and even considering only sequenced bases not all were sequenced to high coverage or exhibited properties that would have allowed a SNP to have been called at a nucleotide position even if it were variable. Thus, we first used the `samtools` (version 1.5) `depth` command to determine the number of bases within *Perform* sequenced to at least 2× coverage (average per individual), which was the same threshold used for variant calling. We then tried to account for the fact that a subset of these sites would not pass other filtering criteria. Specifically, we calculated the proportion of SNPs that passed the coverage filter but failed quality control based on other filters (about 2/3rds of the initial SNPs) and assumed that this same proportion of non-SNPs would have been filtered out if they had been variable. This gave us an effective number of sequenced bases of 53,538.3, which we used in combination with the per-base mutation rate to calculate the divergence time in years (see https://github.com/zgompert/TimemaRW/blob/main/ComputeDate.R). Confidence intervals on the divergence time were inferred using a block-jackknife procedure to account for the non-independence among SNPs within *Perform*. Specifically, the SNPs within *Perform* were divided into 100 contiguous 18 SNP windows and divergence time estimates were obtained for each unique data subset of 99 of the 100 SNP windows.

186 **Supplemental Tables.**

Table S1. Summary of samples and genetic data used to measure host-associated genetic differentiation. Host abbreviations are: C = *Ceanothus*, P = *Pseudotsuga menziesii* (Douglas fir), A = *Arctostaphylos* (Manzanita), Pi = *Pinus*, Q = *Quercus*, RW = *Sequoia sempervirens* (Redwood). N1 and N2 denote the sample sizes for populations 1 and 2, respectively. See Riesch et al. (38) for additional information about these populations.

Species	Population 1	Population 2	N1	N2	Number of SNPs
T. californicum	SM on A	SM on Q	17	20	7858
T. knulli	BCE on RW	BCWP on C	15	12	1139
T. knulli	BCTUR on C	BCTUR on Pi	17	16	1139
T. landelsensis	BCBOG on C	BCBOG on Q	23	20	8548
T. landelsensis	BCSUM on C	BCSUM on Q	20	11	8548
T. poppensis	TBARN on P	TBARN on RW	20	20	7157

Table S2. Summary of the relationships between our current chromosome number system (based on *T. cristinae*), our previous *T. cristinae* genome that used linkage groups (39), and the numbers (IDs) for the chromosome-scale scaffolds in *T. cristinae* (GS = green striped morph), *T. knulli*, and *T. chumash* genomes here based on whole genome-alignments. Note that our earlier linkage groups corresponded with *T. cristinae* chromosomes with two exceptions, one chromosome (now 9 = *T. cristinae* GS scaffold 16151) was split between two linkage groups, and the sex chromosome, X (now 13 = *T. cristinae* GS scaffold 14101) was not assigned to any linkage group.

T. cristinae chromosome number	Old *T. cristinae* linkage group	*T. cristinae* (GS) scaffold number	*T. knulli* scaffold number	*T. chumash* scaffold number
1	1	8483	29	43
2	2	14640	813	1392
3	3	42935	29	43
4	4	42912	6886	43
5	5	18722	6895	56
6	6	9928	6839	1469
7	7	10660	934	1510
8	8	7748	6852	113
9	9,13	16151	1305	43
10	10	14160	30	1213
11	11	12033	500	48
12	12	12380	6840	1403
13	NA	14101	775	1308

Table S3. Summary of samples for the *T. knulli* performance experiment and associated genetic analyses. Host abbreviations are: C = *Ceanothus* and RW = Redwood (*Sequoia sempervirens*). BCE, BCSH and BCXD are very near each other and all treated as "BCE" for analyses the simply exclude BCTURN. N denotes sample size.

Population	Host	Latitude (°N)	Longitude (°W)	N
BCE	C	36.07	121.60	68
BCE	RW	36.07	121.60	24
BCSH	RW	36.07	121.60	1
BCTURN	C	36.08	121.61	37
BCXD	C	36.07	121.60	8

Table S4. Sample sizes from population BCE in the rearing experiment used in the regression analysis connnecting *Perform* genotype to performance. Counts are given for each combination of source host (RW = Redwood, C = *Ceanothus*), *Perform* genotype, host treatment (columns, Redwood or *Ceanothus*) and sex. Variability in sample sizes for different source and genotype combinations reflects the different frequencies of *Perform* alleles on different hosts.

		Redwood		*Ceanothus*	
Source	Genotype	female	male	female	male
RW	$P_{RW}P_{RW}$	5	1	5	6
	$P_{RW}P_C$	3	3	2	0
	P_CP_C	0	0	0	0
C	$P_{RW}P_{RW}$	3	2	2	1
	$P_{RW}P_C$	11	5	14	5
	P_CP_C	12	4	13	4

Table S5. Summary of regression models for 15-day weight in the *T. knulli* experiment. Results are shown for the Redwood (RW) and *Ceanothus* host treatments for models with *Perform* genotype (G), host source (H), or both factors and their interaction (G:H). Effects of sex and developmental stage were removed prior to analysis. Regression parameter estimates (β) and associated *P*-values are reported along with the overall model r^2, *P*-value and Akaike information criterion metric (AIC). The model with the lowest AIC (i.e., the best model) for each host treatment is denoted with an asterisk.

	Genotype		Host		Genotype:Host		Model		
Model	β	P	β	P	β	P	r^2	P	AIC
RW G*	-0.009	0.033					0.106	0.033	-213.43
RW H			0.013	0.071			0.077	0.071	-212.09
RW G:H	-0.004	0.428	0.038	0.075	0.020	0.120	0.179	0.051	-213.09
C G*	-0.018	0.010					0.132	0.010	-184.45
C H			0.013	0.339			0.019	0.339	-178.35
C G:H	-0.027	0.006	-0.056	0.192	0.030	0.323	0.169	0.035	-182.65

Table S6. Summary of regression models for 21-day weight in the *T. knulli* experiment. Results are shown for the Redwood (RW) and *Ceanothus* host treatments for models with *Perform* genotype (G), host source (H), or both factors and their interaction (G:H). Effects of sex and developmental stage were removed prior to analysis. Regression parameter estimates (β) and associated *P*-values are reported along with the overall model r^2, *P*-value and Akaike information criterion metric (AIC). The model with the lowest AIC (i.e., the best model) for each host treatment is denoted with an asterisk.

	Genotype		Host		Genotype:Host		Model		
Model	β	P	β	P	β	P	r^2	P	AIC
RW G	-0.004	0.418					0.016	0.418	-186.15
RW H*			0.026	0.004			0.189	0.004	-194.48
RW G:H	0.004	0.469	0.038	0.162	-0.006	0.709	0.201	0.032	-191.08
C G*	-0.019	0.019					0.116	0.019	-163.68
C H			0.006	0.703			0.003	0.703	-158.04
C G:H	-0.023	0.030	0.033	0.495	-0.044	0.196	0.184	0.032	-163.44

Table S7. Summary of regression models for survival in the *T. knulli* experiment. Results are shown for the Redwood (RW) and *Ceanothus* host treatments for models a null model (intercept only) and models with *Perform* genotype (G), host source (H), sex (S) or combinations of two factors and their interaction (G:H and G:S). The overall model r^2 and *P*-value from a simple linear model fit and the Akaike information criterion metric (AIC) assuming a normal or binomial error distribution are reported. The model with the lowest AIC based on the binomial error distribution (i.e., the best model) for each host treatment is denoted with an asterisk.

Model	r^2	P	AIC (normal)	AIC (binomial)
RW null*	NA	NA	25.95	34.30
RW G	0.002	0.759	27.85	36.20
RW H	0.001	0.810	27.89	36.23
RW S	0.006	0.596	27.65	35.98
RW G:H	0.020	0.818	30.94	39.83
RW G:S	0.009	0.934	31.48	38.73
C null	NA	NA	32.90	39.19
C G*	0.115	0.014	28.52	34.60
C H	0.121	0.012	28.21	35.83
C S	0.079	0.044	30.63	37.44
C G:H	0.169	0.029	29.25	37.52
C G:S	0.188	0.018	28.08	36.61

Table S8. Number and proportion (in parantheses) of *T. knulli* surviving to the end of the experiment as a function of *Perform* genotype, host treatment (Redwood or *Ceanothus*) and sex.

	Redwood		*Ceanothus*	
Genotype	female	male	female	male
$P_{RW}P_{RW}$	8/8 (1.00)	3/3 (1.00)	6/7 (0.86)	4/7 (0.57)
$P_{RW}P_{C}$	11/14 (0.79)	7/8 (0.88)	15/16 (0.94)	4/5 (0.80)
$P_{C}P_{C}$	11/12 (0.92)	4/4 (1.00)	13/13 (1.00)	4/4 (1.00)

Table S9. Model parameter estimates from δaδi. The scaled parameters presented are defined as follows: $\theta = 4N_{anc}\mu$, $\nu_1 = N_1/N_{anc}$, $\nu_2 = N_2/N_{anc}$, $T_{split} = 2N_{anc}t_{split}$, $M_{12} = 2N_{anc}m_{12}$, $M_{21} = 2N_{anc}m_{21}$, where N and t denote actual effective population sizes and time in generations (years) and μ is the total mutation rate for the locus. 95% CIs denote 95% block-jackknife confidence intervals. Here, populations 1 and 2 refer to homozygotes for P_s (more common on *Ceanothus*) and P_g (more common on Redwood), respectively.

Parameter	Estimate	95% CI lower	95% CI upper
θ	111.184	41.845	234.364
ν_1	0.231	0.106	0.631
ν_2	3.901	1.875	10.386
T_{split}	13.833	2.573	28.976
M_{12}	0.367	0.134	0.782
M_{21}	0.110	0.040	0.248

187 **Supplemental Figures.**

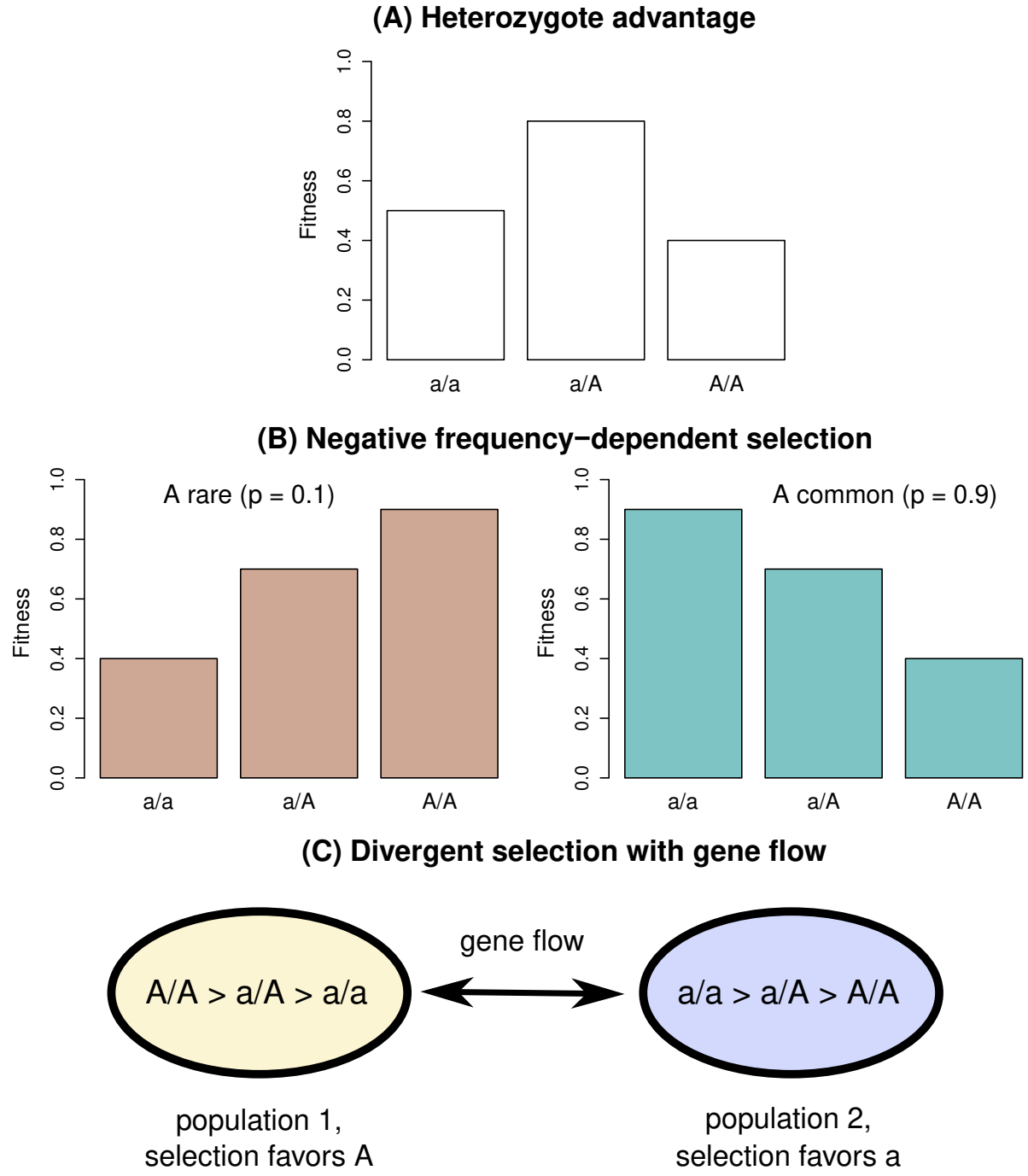

Fig. S1. Graphical overview of forms of balancing selection that can maintain genetic variation. Panel (A) shows the case of heterozygote advantage (overdominance) where the heterozygote (a/A) has the highest Darwinian fitness. This can occur for several reasons, including life-history trade-offs where the heterozygote has the highest mean fitness across life-history stages. Panel (B) shows negative frequency-dependent selection where selection is always directional (favoring a/a or A/A homozygotes) but where the rarer homozygote is most fit. The illustration contrasts cases where the frequency of A (p) is 0.1 versus 0.9. Panel (C) shows balancing selection caused by divergent selection, that is directional selection favoring alternative alleles or genotypes in different populations, combined with gene flow between populations occupying different habitats. Thus, unlike the first two cases, this third case necessarily involves multiple populations. All forms of balancing selection shown here can maintain polymorphism at the population and species level.

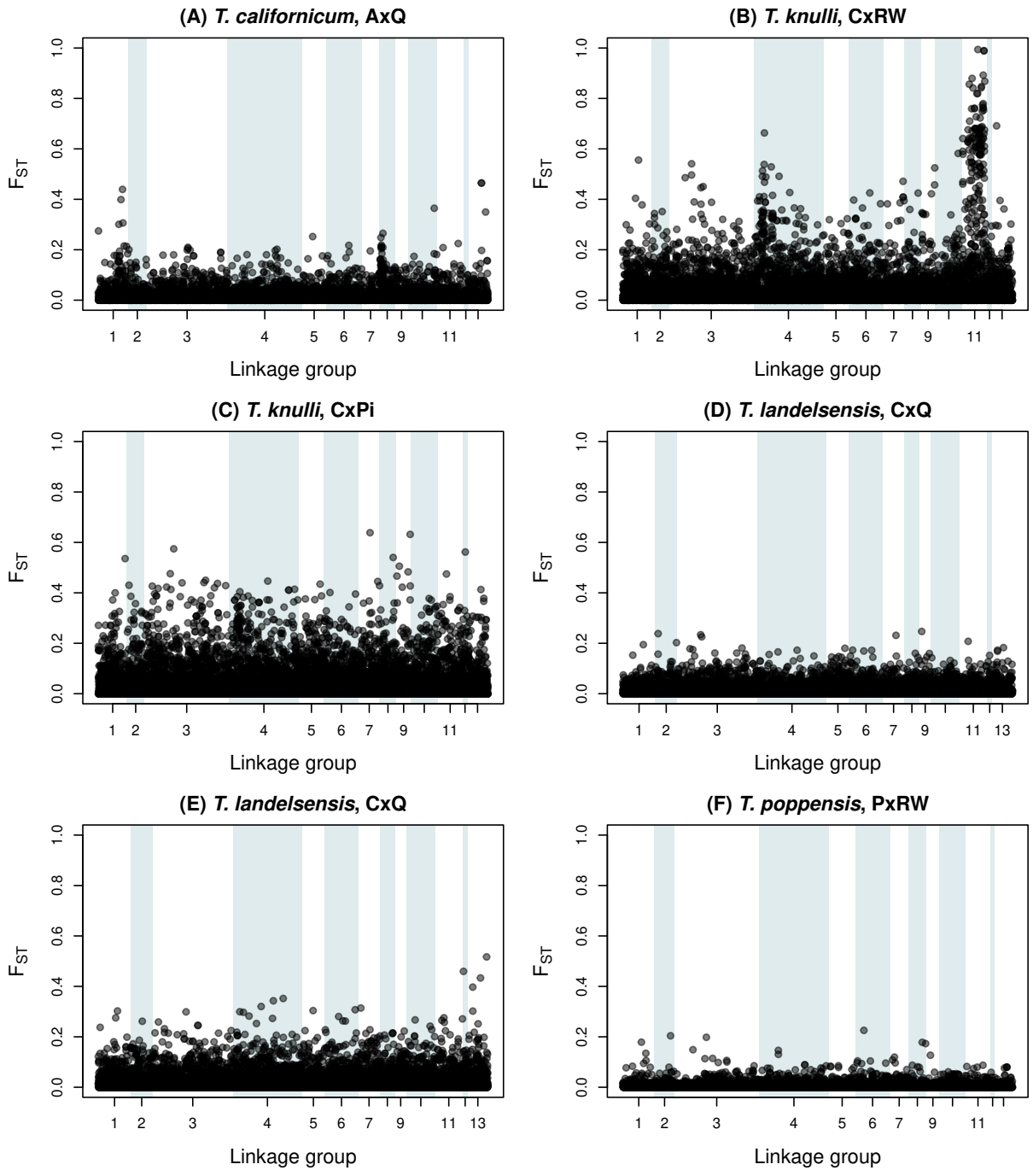

Fig. S2. Manhattan plots of genome-wide genetic differentiation for parapatric *Timema* populations on different hosts. Points denote F_{ST} for individual SNPs organized by *T. cristinae* linkage groups. Host abbreviations are A = *Adenostoma*, C = *Ceanothus*, P = *Pseudotsuga menziesii* (Douglas Fir), Pi = *Pinus* (pine), Q = *Quercus* (oak), and RW = *Sequoia sempervirens* (Redwood).

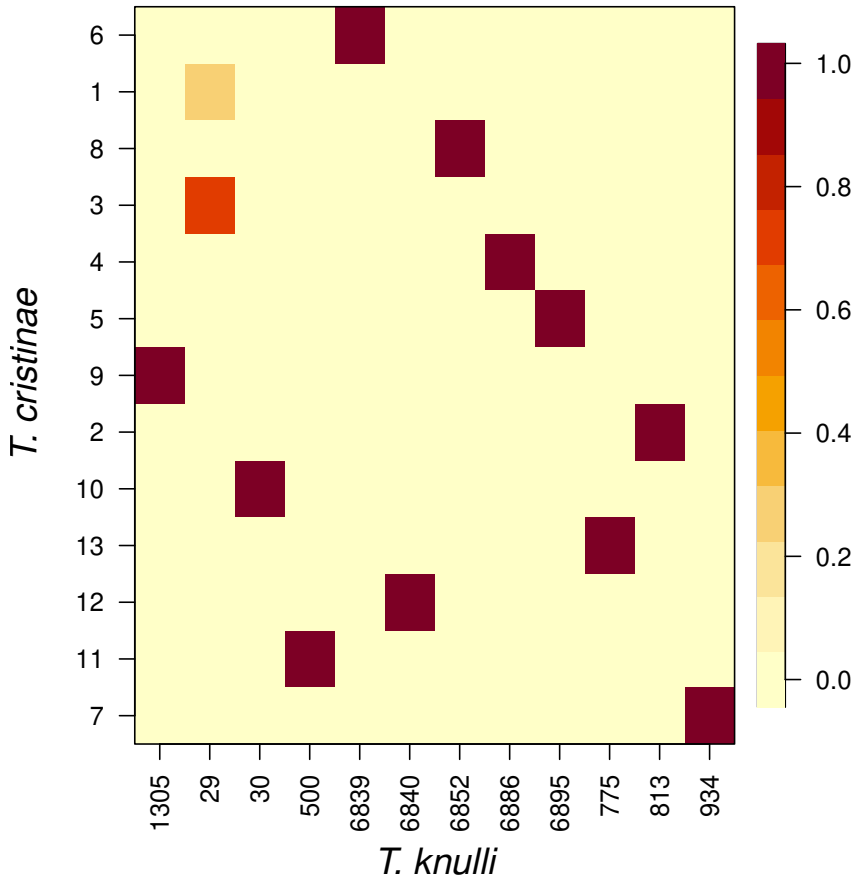

Fig. S3. Heatmap shows the proportion synteny blocks on of each *T. knulli* scaffold that aligned to each of the *T. cristinae* chromosomes.

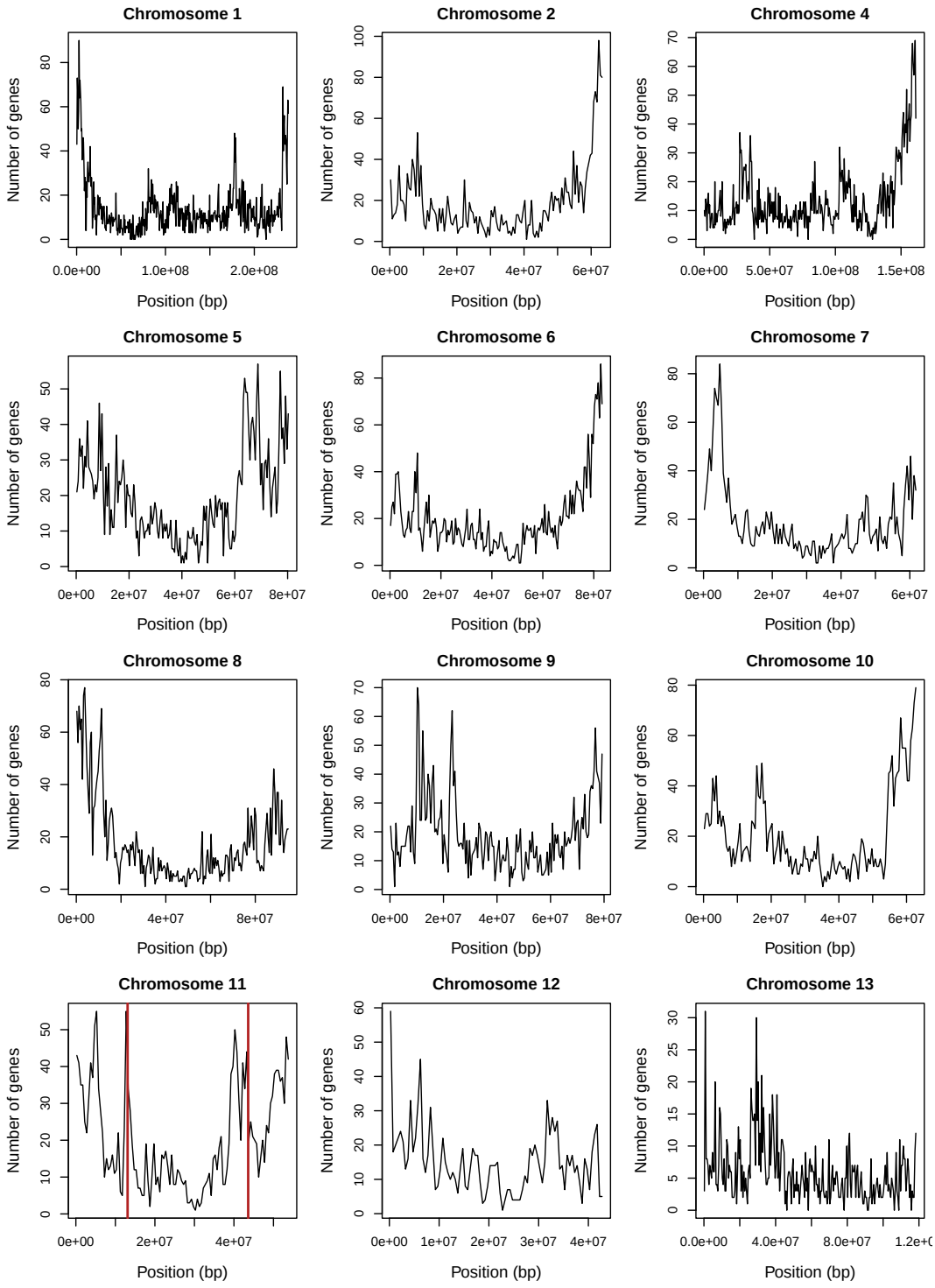

Fig. S4. Plots show gene density along *T. knulli* chromosomes. Gene density was calculated as the number of genes overlapping 500 mbp windows. Numerous factors likely contribute to variation in gene density along chromosome, including the location of centromeres (which is not known in *T. knulli*), but notable spikes in gene density are apparent near the bounds of the *Perform* inversion on chromosome 11, as indicated by vertical red lines.

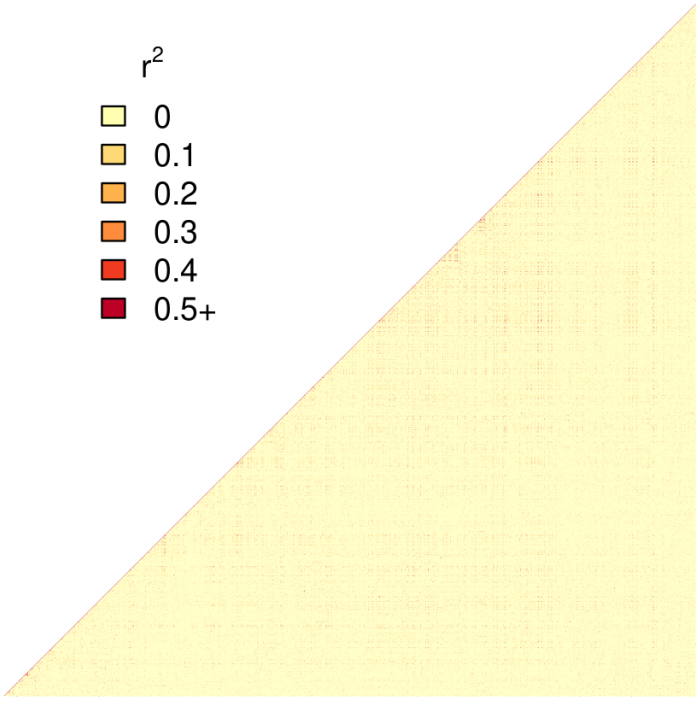

Fig. S5. Heatmap showing patterns of pairwise linkage disequilibrium (LD) between SNPs for chromosome 11 in $P_{RW}\,P_C$ *T. knulli* heterozygotes. LD was quantified using the squared genotypic correlation (r^2). Heatmaps for additional genotypes and genotype combinations are shown in the main text.

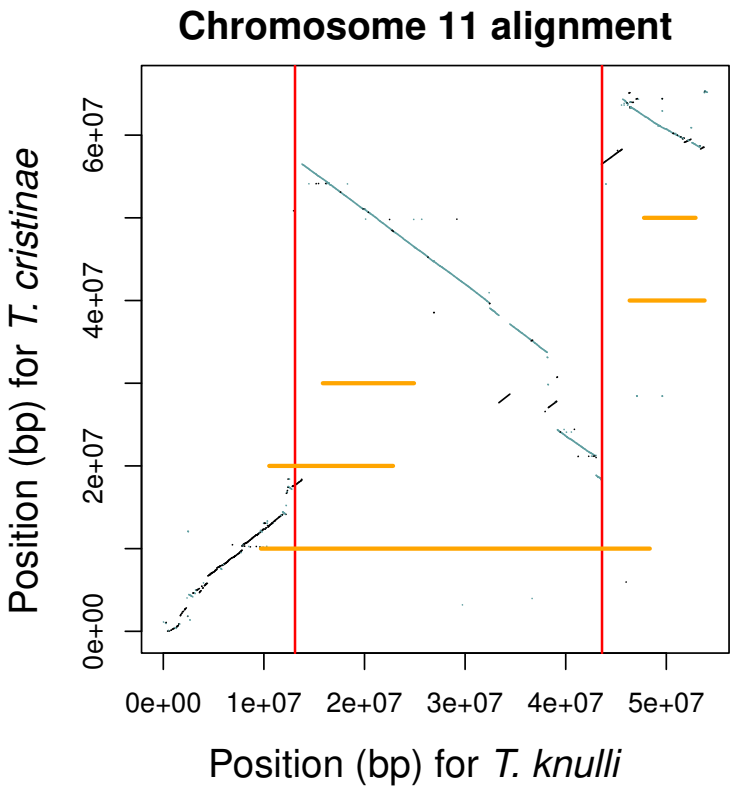

Fig. S6. The dot plot shows the alignment of chromosome 11 for *T. knulli* (from Redwood) and *T. cristinae*. Line segments denote aligned genome regions with the orientation of the alignment shown by the direction of the lines. The bounds of the *Perform* locus in the *T. knulli* genome are denoted by the vertical red lines. The location of the five inversions detected on chromosome 11 for the *T. knulli* genome from *Ceanothus* relative to the Redwood *T. knulli* genome are shown with horizontal orange lines. The location of these along the y-axis is arbitrary. These inversions were delineated based on nanopore DNA sequence data. The main text focuses on the largest of these five inversions.

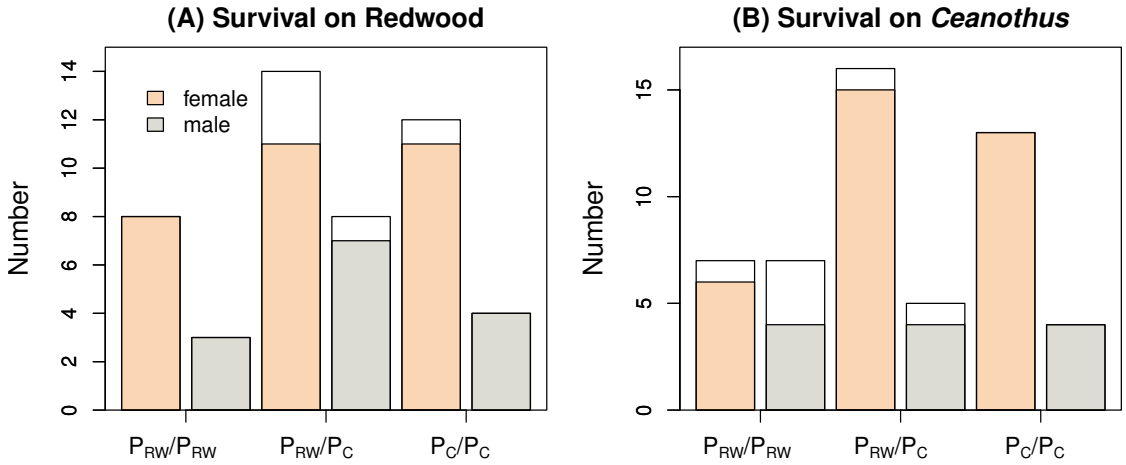

Fig. S7. Barplots show the number of individuals with each *Perform* genotype at the start (bar) and end (colored portion of bar) of the rearing experiment on Redwood (A) and *Ceanothus* (B).

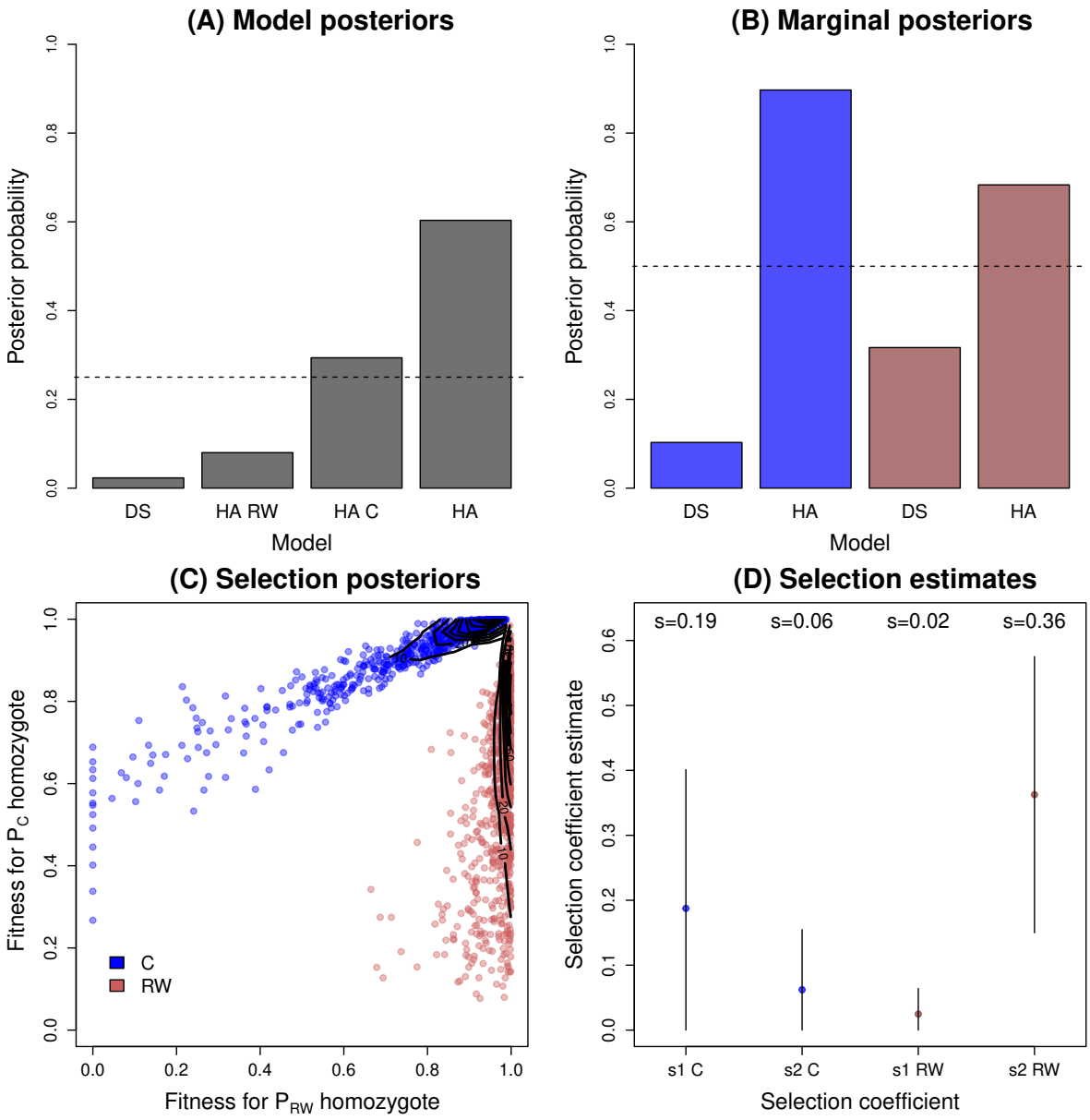

Fig. S8. Summary of the ABC results. Panels (A) and (B) given model posterior probabilities. In (A) posteriors are given for DS = divergent directional selection on both hosts, HA RW = heterozygote advantage on Redwood and directional selection on *Ceanothus*, HA C = directional selection on Redwood and heterozygote advantage on *Ceanothus*, and HA = heterozygote advantage on both hosts. In (B) marginal posteriors are shown for directional (DS) versus heterozygote advantage (HA) on each host (indicated by color). Panel (C) shows the joint posterior for the fitness of P_{RW} versus P_C homozygotes on each host, where points denote individual samples from the posterior with contours overlain. (D) gives posterior estimates of the selection coefficients $s1$ and $s2$ (heterozygote advantage) on *Ceanothus* (C) versus Redwood (RW). Points and numbers denote posterior medians and vertical bars indicate 95% credible intervals.

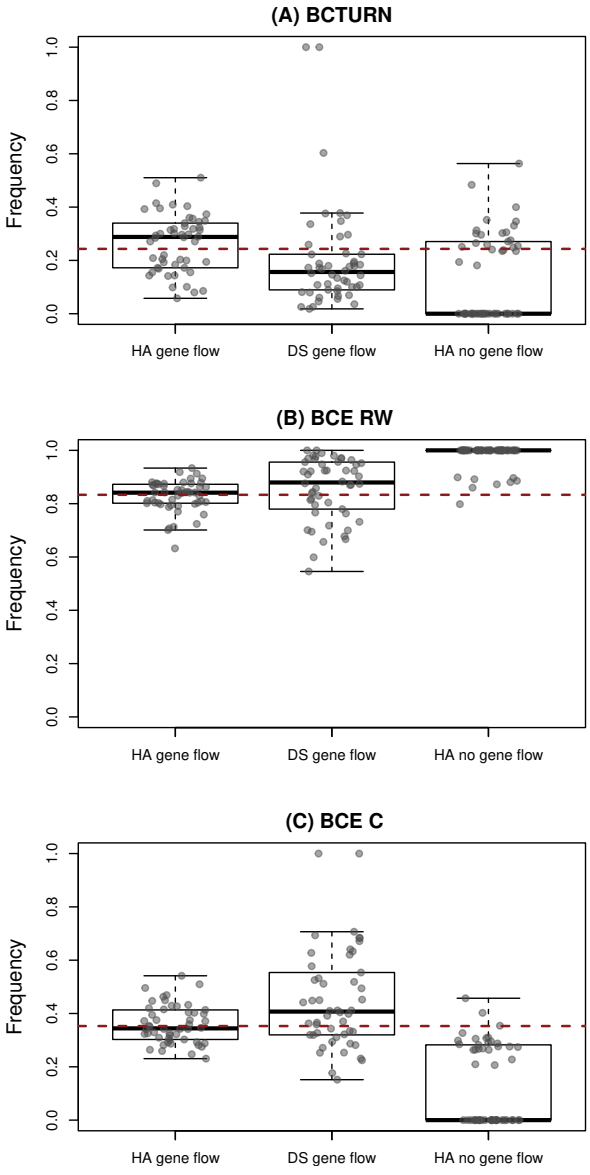

Fig. S9. Boxplots show the P_{RW} allele frequency from simulations for BCTURN (A), BCE RW (B) and BCE C (C) after 250,000 generations of evolution with heterozygote advantage (HA) and gene flow, directional selection (DS) and gene flow, or HA without gene flow. Results are shown for 50 replicate simulations under each set of conditions. Boxes denote the 1st and 3rd quartile, with the median given by the midline and whiskers extending to the minimum and maximum value or 1.5$\times$ the interquartile range. Points show the allele frequency for each replicate simulation. The observed P_{RW} allele frequency in each population is shown with a red, dashed line.

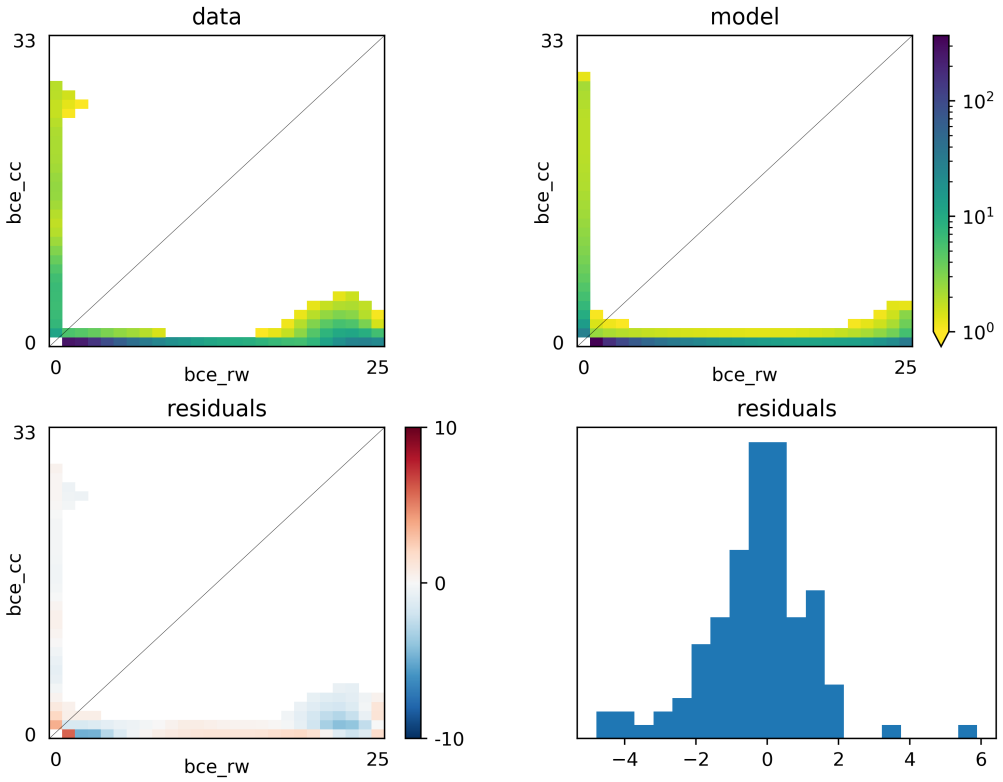

Fig. S10. Summary of model fit for divergence time models in $\delta a \delta i$. The top panels show the observed (data) and predicted (model) joint site frequency spectra for *Perform* locus. The bottom panels show the corresponding residuals, that is the deviation between the observed and model-predicted joint site frequency spectra.

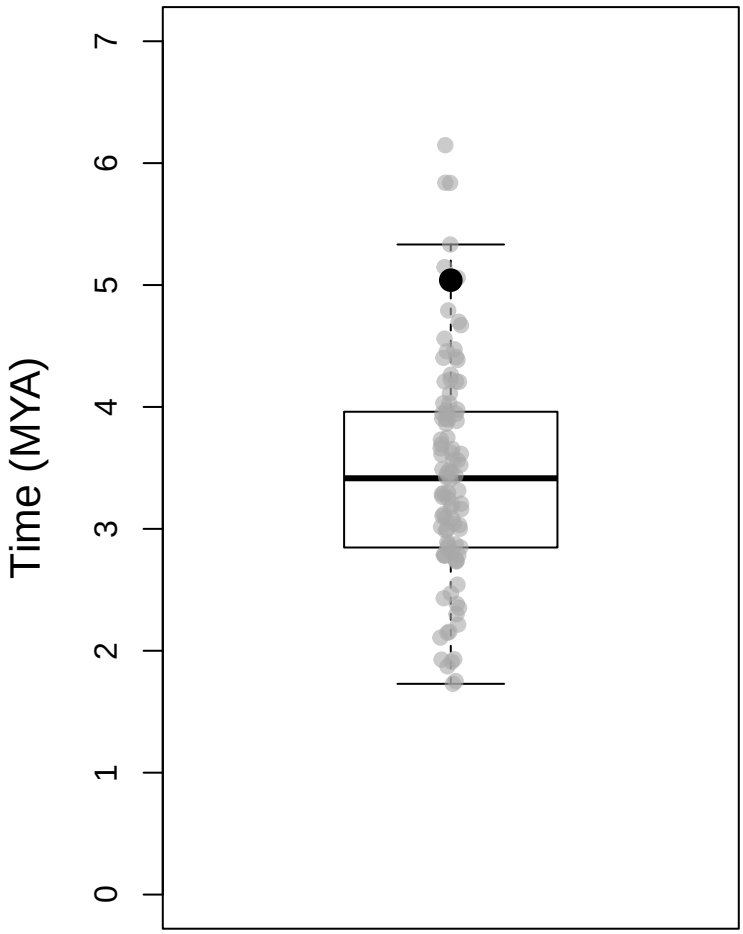

Fig. S11. Inversion divergence time estimates from δaδi. The boxplot summarizes estimates of divergence time (in millions of years ago = MYA) for each of 100 block-jackknife replicates. Boxes denote the 1st and 3rd quartile, with the median given by the midline and whiskers extending to the minimum and maximum value or 1.5$\times$ the interquartile range. Gray points denote the estimate for each replicate; the larger black dot indicates the estimate from the full data set.

References

1. Brua T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M (2021) BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics* 3(1):lqaa108.
2. Villoutreix R, et al. (2020) Large-scale mutation in the evolution of a gene complex for cryptic coloration. *Science* 369(6502):460–466.
3. Bruna T, Lomsadze A, Borodovsky M (2020) GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics and Bioinformatics* 2(2):lqaa026.
4. Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* 33(suppl_2):W465–W467.
5. Stanke M, et al. (2006) AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research* 34(suppl_2):W435–W439.
6. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M (2005) Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research* 33(20):6494–6506.
7. Gotoh O (2008) A space-efficient and accurate method for mapping and aligning cdna sequences onto genomic sequence. *Nucleic Acids Research* 36(8):2630–2638.
8. Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12(1):59–60.
9. Kriventseva EV, et al. (2019) OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research* 47(D1):D807–D811.
10. Jones P, et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.
11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215(3):403–410.
12. Paten B, et al. (2011) `Cactus`: Algorithms for genome multiple sequence alignment. *Genome Research* 21(9):1512–1528.
13. Armstrong J, et al. (2020) Progressive `cactus` is a multiple-genome aligner for the thousand-genome era. *Nature* 587(7833):246–251.
14. Smit AF (2010) Repeat-masker open-3.0. *http://www. repeatmasker. org*.
15. Krasheninnikova K, et al. (2020) `halSynteny`: a fast, easy-to-use conserved synteny block construction method for multiple whole-genome alignments. *GigaScience* 9(6):giaa047.
16. Lu H, Giordano F, Ning Z (2016) Oxford Nanopore MinION sequencing and genome assembly. *Genomics, Proteomics & Bioinformatics* 14(5):265–279.
17. Zhang L, Chaturvedi S, Nice CC, Lucas LK, Gompert Z (2022) Population genomic evidence of selection on structural variants in a natural hybrid zone. *Molecular Ecology*.
18. De Coster W, D'hert S, Schultz DT, Cruts M, Van Broeckhoven C (2018) `NanoPack`: visualizing and processing long-read sequencing data. *Bioinformatics* 34(15):2666–2669.
19. Li H (2018) `Minimap2`: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18):3094–3100.
20. Li H, et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
21. Smolka M, et al. (2022) Comprehensive structural variant detection: from mosaic to population-level. *BioRxiv*.
22. Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025–2035.
23. Sisson SA, Fan Y, Beaumont M (2018) *Handbook of approximate Bayesian computation*. (CRC Press).
24. Gaggiotti OE, Foll M (2010) Quantifying population structure using the F-model. *Molecular Ecology Resources* 10(5):821–830.
25. Gompert Z, Springer A, Brady M, Chaturvedi S, Lucas LK (2021) Genomic time-series data show that gene flow maintains high genetic diversity despite substantial genetic drift in a butterfly species. *Molecular Ecology* 30(20):4991–5008.
26. Balding DJ, Nichols RA (1995) A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* 96(1):3–12.
27. Nicholson G, et al. (2002) Assessing population differentiation and isolation from single-nucleotide polymorphism data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4):695–715.
28. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4):1567–1587.
29. Gompert Z, et al. (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution* 66(7):2167–2181.
30. Stan Development Team (2020) `RStan`: the `r` interface to `stan`. R package version 2.21.2.
31. Hoffman MD, Gelman A, , et al. (2014) The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 15(1):1593–1623.
32. Ewens WJ (2004) *Mathematical Population Genetics: Theoretical Introduction*. (Springer) Vol. 1.
33. Galassi M, et al. (2002) *GNU Scientific Library*. (Network Theory Limited Godalming).
34. Csillery K, Francois O, Blum MGB (2012) `abc`: an `r` package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*.
35. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* 5(10):e1000695.
36. Blischak PD, Sajan M, Barker MS, Gutenkunst RN (2022) Demographic history inference and the polyploid continuum.

bioRxiv.
37. Liu H, et al. (2017) Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. *Molecular Biology and Evolution* 34(1):119–130.
38. Riesch R, et al. (2017) Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution* 1(4):1–13.
39. Nosil P, et al. (2018) Natural selection and the predictability of evolution in *Timema* stick insects. *Science* 359(6377):765–770.