Complex evolutionary processes maintain an ancient chromosomal inversion

Patrik Nosil,[1*] Victor Soria-Corrasco,[2] Romain Villoutreix,[1]
Marisol De la Mora Curiel,[1,3] Clarissa F. de Carvalho,[1]
Thomas Parchman,[4] Jeffrey L. Feder,[5] Zachariah Gompert[6*]

[1]DCEFE, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier, France
[2]John Innes Centre, Norwich, NR4 7UH UK
[3]Escuela Nacional de Estudios Superiores, Unidad Juriquilla, UNAM. Querétaro 76230, México
[4]Department of Biology, University of Nevada, Reno, Nevada 89557, USA
[5]Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556, USA
[6]Department of Biology, Utah State University, Logan, Utah 84322, USA

*To whom correspondence should be addressed;
E-mail: patrik.nosil@cefe.cnrs.fr and zach.gompert@usu.edu.

Abstract

Genome re-arrangements such as chromosomal inversions are often involved in adaptation. As such, they experience natural selection, which can erode genetic variation. Thus, whether and how inversions can remain polymorphic for extended periods of time remains debated. Here we combine genomics, experiments, and evolutionary modeling to elucidate the processes maintaining an inversion polymorphism associated with the use of a challenging host plant (Redwood trees) in *Timema* stick insects. We show that the inversion is maintained by a combination of processes, finding roles for life-history trade-offs (i.e., balancing selection), local adaptation to different hosts, and gene flow. We use models to show how such multi-layered regimes of selection and gene flow provide resilience to help buffer populations against the loss of genetic variation, maintaining potential for future evolution. We further show that the inversion polymorphism has persisted for millions of years and is not a result of recent introgression. We thus find that rather than being a nuisance, the complex interplay of evolutionary processes provides a mechanism for the long-term maintenance of genetic variation.

Significance statement

Variation is the fuel for evolution. How genetic variation is maintained is one of the central questions in biology. This is an especially striking question for chromosomal inversions (a change in the structure of an organism's genome), as inversions are often subject to natural selection, which can erode variation. We studied stick insects that have an inversion that helps them live on Redwood trees. We found that this inversion has been present for millions of years, and that a suite of factors, including environmental heterogeneity and gene exchange, contribute to the persistence of this polymorphism. Our results show how a complex interplay of evolutionary processes offers a bulwark against the loss of variation allowing for the potential

for future evolution.

Introduction

Genetic variation is the ultimate fuel for evolution. However, many forms of natural selection (e.g., directional and purifying selection) and random genetic drift are expected to result in the loss of genetic variation, depleting the reservoir of fuel for evolution. Whether and how genetic variation can be maintained over long periods of time thus remains a central question in biology [1–6]. We address this question here by studying the maintenance of an ancient chromosomal inversion. Since their discovery by Sturtevant $\sim$100 years ago [7], chromosomal inversions have been central to the development of evolutionary biology. For example, they served as the first genetic markers, motivated ideas by Dobzhansky, Ford, and others concerning co-adapted gene complexes and balancing selection, and they underlie several modern theories of adaptation that involve suppressed recombination [8–11]. Inversions also serve as powerful models for studying the maintenance of genetic variation, because their age can be estimated and they are often subject to natural selection [12–14].

Although inversions are now known to vary along environmental clines and to be associated with adaptive traits [8, 9, 13, 15–17], studies that directly estimate selection on inversions are few, some notable exceptions aside [18–20]. Thus, the mode and strength of selection acting on inversions remains poorly quantified, making it difficult to infer how and why inversion polymorphism is maintained. For example, balancing selection can maintain inversion polymorphisms [9, 21], especially if strong enough to counteract drift, but this is not true of many forms of selection. Similarly, the role of other processes, such as gene flow, in maintaining polymorphism requires further study [6, 22–24].

Determining the age of an inversion is also important for explaining the maintenance of inversion polymorphisms. For example, one hypothesis is that the inversion is young and still in the process of sweeping to fixation. In other words, it could be that the inversion will not

be maintained as polymorphic in the long term. If the inversion polymorphism is found to be ancient such that this 'young inversion' hypothesis is refuted, then studies of the processes maintaining variation, particularly natural selection, are required to explain the inversion polymorphism (Figure 1A). Here we combine field data, genomics, experimental estimates of fitness, and evolutionary modeling to to elucidate the processes driving the long-term maintenance of an inversion polymorphism with fitness consequences across populations using different hosts (Figure 1).

Our study system is the genus *Timema*, a group of plant-feeding stick insects distributed throughout southwestern North America (Figure 1). *Timema* are well studied for their cryptic colors and patterns, which help them avoid predation by visual predators such as birds and lizards [25, 26]. These traits are highly heritable and controlled by a modest number ($\sim$5) of linked loci on linkage group (LG) 8 (LG8 hereafter), which often exhibit strongly reduced recombination due to structural genomic features including chromosomal inversions and deletions [12, 27, 28]. *Timema* are also known to use a particularly wide range of host-plant species, including both conifers and flowering plants (i.e., angiosperms) [29]. This host-plant use, in the context of local adaptation (i.e., growth and survival on different hosts; 'performance' hereafter), is our focus here. Notably, the genetic basis of performance variation in *Timema* was previously unknown, but as we report here also involves a chromosomal inversion (on a different chromosome from color, LG11).

Results

Genome scans reveal exceptional host-associated differentiation on linkage group 11. During the 30-million year diversification of the *Timema* genus, host shifts have occurred frequently between plant families (within conifers and within flowering plants), and several times even between these plant divisions [29]. Indeed, *Timema* are broadly generalized in diet, often feeding

on multiple plant families in nature and surviving in the lab on novel hosts [30]. One exception involves the use of Redwood (*Sequoia sempervirens*); very few *Timema* species and populations use Redwood in nature–only *T. knulli* and *T. poppensis*–and most exhibit poor performance on this host in laboratory experiments [30].

We thus initiated our investigation by quantifying patterns of genetic differentiation for the sexual species of *Timema* that live in the vicinity of Redwood in northern California and that use multiple hosts in nature. Specifically, we study *T. californicum* and *T. landelsensis*, which do not use Redwood, *T. poppensis*, which is specialized on conifers including Redwood in some localities, and *T. knulli*, which uses both Redwood and flowering plant hosts (i.e., angiosperms). In this context, *T. knulli* is of particular interest as it is polymorphic in host-plant use, living on Redwood (a conifer) as well as other more commonly-used hosts such as *Ceanothus* (an angiosperm) (in contrast, *T. poppensis* uses only conifer hosts). We did so using published genotyping-by-sequencing (GBS) data [31, 32]. Our core interest was whether the use of a certain host was associated with genetic differentiation, and if so whether this was genome-wide or restricted to individual chromosomes. Due to the known strong effects of geographic isolation on genetic structure in *Timema* [33], we restricted our survey to the six pairwise comparisons involving nearby populations using different hosts (broadly speaking, 'parapatry', Table S1, Figure 1). This revealed that genetic differentiation between parapatric, conspecific populations was generally weak. The exception to this trend was LG11 for populations of *T. knulli* using *Ceanothus* versus Redwood: LG11 was strongly differentiated in this comparison. We thus focused our study on *T. knulli*, with particular reference to the use of Redwood.

Redwood *T. knulli* populations are distinguished by a chromosomal inversion. The results above were based on mapping GBS reads to the published *T. cristinae* reference genome [31, 32]. *Timema knulli* is known from cytological work to have one chromosome pair fewer than *T. cristinae* [34], and we suspected structural variation on LG11 within *T. knulli*. Thus, to increase

the accuracy and precision of the current work and test explicitly for structural variation, we generated a high-quality *de novo* reference genome assembly for *T. knulli*. We did so using an individual collected from Redwood and a combination of PacBio and Illumina reads with Hi-C technology for scaffolding. The *T. knulli* genome comprised 12 large scaffolds corresponding to the 13 known *T. cristinae* chromosomes, but with a fusion between *T. cristinae* chromosomes 1 and 3 (we refer to the fused chromosome as chromosome 1 and retain the *T. cristinae* linkage group numbering for the other chromosomes; total assembly length = 1,322,373,696 base pairs; scaffold N50 = 83,614,905 base pairs) (Table S2, Figure S1). We then used this reference genome for further population genetic and trait mapping analyses, with new data collected to allow larger sample sizes for *T. knulli* than what was available from published data.

Using new GBS data from 138 *T. knulli* collected on *Ceanothus* and Redwood (Table S3) we detected a large block of differentiation (e.g., highly accentuated $\mathrm{F_{ST}}$) on chromosome 11, whose boundaries were delimited using a Hidden Markov Model (HMM) approach applied to the results of a principal components analysis (PCA) (Figure 2). This block spanned genomic positions 13,093,370 to 43,606,674 on chromosome 11 ($\sim$30 mega-base pairs, mbps). We hereafter refer to this region as the '*Perform*' locus, as polymorphism at this regions was associated with performance variation in an experiment reported below. A PCA of SNPs within the *Perform* locus revealed three genetic clusters segregating within populations (Figure 2). In contrast, PCA of genome-wide genetic variation exhibited structure by geography. This result is consistent with the *Perform* locus being a structural genomic variant that segregates within populations, differs in frequency among populations (as we reported in more detail below, one allele is at 84% frequency on Redwood but only at 34% frequency on *Ceanothus*), and exhibits reduced recombination between the two chromosomal variants.

To more formally test the existence of a chromosomal inversion on *T. knulli* chromosome 11, we aligned the *T. knulli* genome with published chromosome-level assemblies of *T. cristinae*

and *T. chumash* genomes [27, 28]. These alignments identified an inversion on chromosome 11 in the Redwood *T. knulli* genome relative to both *T. cristinae* and *T. chumash*. Most critically, the breakpoints of this inversion coincided with the identified bounds of the *Perform* locus (Figure 3). In contrast, this genomic region was co-linear between *T. cristinae* and *T. chumash*. The collective results are most consistent with the *Perform* locus being a polymorphic chromosomal inversion in *T. knulli*. To explicitly test this hypothesis, we gathered nanopore long-read DNA sequence data from a second *T. knulli* collected on *Ceanothus*. This revealed a large inversion (9,706,606 to 48,357,002 bps on chromosome 11) relative to the Redwood *T. knulli* genome (Figures 3, S2). Critically, this inversion spanned the *Perform* locus and the inversion boundaries identified between species, consistent with expectations if the inversion also segregates within *T. knulli*.

The *Perform* locus inversion affects performance on different hosts. We next considered the evolutionary processes potentially maintaining the inversion polymorphism. Specifically, to connect the inversion polymorphism to fitness, we tested if performance on *Ceanothus* and Redwood are affected by the *Perform* locus. Such an association with fitness would firmly refute strict neutrality with regard to the evolution of inversion frequencies. To do so, we collected *T. knulli* and reared them in the laboratory on either *Ceanothus* or Redwood, measuring growth and survival (notably these are the same individuals analyzed above to delimit the *Perform* locus). We focus our analyses here on specimens from the vicinity of the locality BCE, where *T. knulli* uses both *Ceanothus* and Redwood (we thus exclude population BCTURN, which uses only *Ceanothus*)(Table S3). These experiments revealed that the *Perform* locus explains appreciable and significant variation in both growth and survival, but with a trade-off between these fitness components that suggests balancing selection, especially on *Ceanothus* (Figure 4).

Specifically, our experiments revealed that one allele at the *Perform* locus was associated with increased growth on both *Ceanothus* and Redwood (hereafter '*Pg*', this is the allele at a

high-frequency on Redwood and referred to as the 'Redwood' allele above; linear regression on residuals after removing effect of sex; *Ceanothus* 15 day weight, β = 0.018, r^2 = 0.178, $P = 0.002$; *Ceanothus* 21 day weight, β = 0.020, r^2 = 0.233, $P < 0.001$; Redwood 15 day weight, β = 0.0086, r^2 = 0.109, $P = 0.031$; Redwood 21 day weight, β = 0.0072, r^2 = 0.053, $P = 0.138$). Critically, this same allele negatively affected survival on *Ceanothus* (linear regression, β = -0.14, r^2 = 0.115, $P = 0.014$), representing a host-specific life-history trade-off. Notably, this latter result was sex-dependent, with *Pg* most markedly decreasing male survival (for individuals homozygous for this allele, 86% of females survived but only 57% of males survived; Table S4). Comparable results were observed using generalized linear models (GLM) for survival rather than simple linear regression (GLM survival on *Ceanothus*, β = -1.71, P = 0.031), demonstrating that the results are robust to methods of analysis. Thus, there is a fitness trade-off between growth and survival at the *Perform* locus suggesting that the locus could be under balancing selection, at least on *Ceanothus*.

The inversion is maintained by complex selection and gene flow. The results above suggest that genetic variation at the *Perform* locus could be maintained, in part, due to life-history trade-offs that vary with host and sex resulting in balancing selection. Moreover, selection appears to be shifted between populations feeding on different hosts. Specifically, the *Pg* allele that confers higher growth (but reduced survival) is at higher frequency in nature on Redwood (84%) than on *Ceanothus* (34%). Thus, there is a marked ($\sim$50%) allele frequency difference between populations on different hosts. We suspect that this reflects the previously documented difficulties *Timema* have using Redwood in laboratory experiments [30]; use of Redwood favors the growth allele to make 'a go of it' on this challenging host (at the same time this growth allele does not appear to compromise survival on Redwood *per se*). In contrast, growing on *Ceanothus* is easy for *Timema* such that the survival advantage is more important than a growth benefit, leading to a high frequency of the allele associated with increased survival on *Ceanothus* (hereafter '*Ps*').

Thus, a shift in selection appears to result in divergent allele frequencies (i.e., adaptation) between hosts with evidence consistent with balancing selection on at least *Ceanothus*. However, gene flow between hosts could also play a role in maintaining variation, especially on Redwood. In principle, gene flow could maintain variation even without balancing selection, that is via a balance between directional selection that acts in divergent directions between hosts and gene flow. We next used evolutionary modeling to quantify these possibilities and their effects on the maintenance of variation.

Specifically, we used approximate Bayesian computation (ABC) to estimate the probability of population genetic models that included genetic drift and gene flow (as inferred from putative neutral loci; see Figure 2C) and either balancing or divergent (between hosts) directional selection on the *Perform* locus. We modeled evolution of the *Perform* locus inversion alleles, not the DNA sequence variation within this genomic region. We did this because of the evidence for selection on the inversion alleles and our interest in the maintenance of this inversion polymorphism rather than on nucleotide variation within the inversion. These models included adjacent (i.e., parapatric) *Ceanothus* (BCE C) and Redwood (BCE RW) populations and an allopatric *Ceanothus* population (BCTURN) (Table S3), with the latter being important to help parse the roles of balancing selection versus gene flow in maintaining variation. Models with balancing selection (i.e., over-dominance at the *Perform* locus) were most probable on both *Ceanothus* (posterior probability = 0.897) and Redwood (posterior probability = 0.683), and a model of (divergent) directional selection on both hosts was very unlikely (posterior probability = 0.023)(Figure 5). Under the most probable model of balancing selection on both hosts (posterior probability = 0.603), relative fitnesses of *Perform* homozygotes (when heterozygote fitness is set to 1.0) were 0.81 for the *PgPg* homozygote (i.e., the homozygote for the allele conferring the growth advantage, that is the Redwood allele) and 0.94 for the *PsPs* homozygote (where *Ps* denotes the allele conferring increased survival on *Ceanothus*, that is the *Ceanothus* allele) on

Ceanothus versus 0.98 and 0.64 for the *PgPg* and *PsPs* homozygotes on Redwood (Figure 5). Thus, this population genetic model-fitting analysis strongly supports balancing selection on *Ceanothus*, consistent with the experiment, and also suggests possible balancing selection on Redwood, though this was not evident from the experiment and the estimated relative fitnesses of the heterozygote (1.0) and the fitter homozygote (0.98 for *PgPg*) were very similar.

The combination of processes buffers populations against the loss of variation. We have shown that gene flow and balancing selection (at least on *Ceanothus*) together can explain the observed polymorphism at the *Perform* locus, but it is unclear whether both processes are necessary for the maintenance of variation. In other words, does this combination of processes maintain variation that would be lost with either process in isolation? To address this question, we simulated evolution under our best model of balancing selection (on both hosts) and gene flow and under two counterfactual models–one with gene flow and divergent directional selection between hosts and one with balancing selection but no gene flow. These two models thus eliminate balancing selection or gene flow, respectively. For all models, we used selection coefficients estimated from the ABC analysis (assuming either balancing selection or directional selection) and gene flow inferred from neutral models based on genome-wide SNP data (except where gene flow was set to 0).

Replicate simulations, each spanning 250,000 generations, showed that the balancing selection with gene flow model routinely maintains variation and predicts the observed data extremely well (Figures 6, S3). Directional selection with gene flow also maintained variation over this moderate time interval in all but a few simulations, but failed to recover the observed *Perform* allele frequencies as well as the balancing selection with gene flow model (Figure S3). Finally, variation was lost in many of the balancing selection without gene flow simulations. Thus, these simulations suggest that gene flow among populations feeding on different hosts and experiencing different selection pressures is important for the long-term maintenance of

variation at *Perform*, and that this combined with balancing selection is particularly effective at preventing the loss of polymorphism.

The chromosomal inversion is ancient. Lastly, we estimated the age of the Redwood inversion to test the hypothesis that it might be young and in the act of sweeping rather than a polymorphism maintained over the long-term (Figure 1). To do so, we first used a phylogenetic approach to estimate the divergence time between the *T. knulli* chromosomal variants. Our inferences were based on SNP data within the *Perform* locus for *T. knulli*, *T. poppensis*, *T. petita* and *T. californicum* and species divergence time estimates from a published, time-calibrated phylogeny [31]. This revealed that the inversion is ancient, inconsistent with the young and sweeping hypothesis. Specifically, the divergence time between Redwood and *Ceanothus* chromosomal variants in *T. knulli* was estimated as 7.5 million years ago, MYA hereafter (90% equal-tail probability intervals [ETPI] = 3.4-13.5 MYA) (Figure 3E,F). Next, we generated a complementary estimate of this divergence time using a population genetic approach based on the site-frequency spectrum and allowing for recombination between inversion haplotypes. Our estimate of the divergence time using this approach (implemented in δaδi [35]) was 5.0 MYA (95% block-jackknife confidence interval lower bound = 1.9 MYA) (see Figure S4 for model performance and Table S5 and Figure S5 for model parameter estimates), which is broadly consistent with the phylogenetic estimate above.

Furthermore, our results from the phylogenetic analysis suggest that the deep divergence between Redwood and *Ceanothus* alleles in *T. knulli* is not due to recent introgression from the closely related species *T. poppensis*, wich feeds on Redwood and other confiers. Specifically, the *T. poppensis Perform* DNA sequences were more closely related to the inverted Redwood *T. knulli* alleles than the *Ceanothus T. knulli* alleles, but the divergence time between *T. poppensis* and *T. knulli* Redwood alleles was 4.7 MYA (90% ETPI = 2.1 to 9.8 MYA). This corresponds roughly to the previously inferred divergence time between these two species based on genome-

wide SNP data (4.1 MYA, 90% ETPI = 2.3 to 6.7 MYA) [31]. Thus, while *T. poppensis* appears to share DNA sequence similarity at the *Perofrm* locus with the *T. knulli* Redwood alleles and there is uncertainty regarding whether the origin of the inverted Redwood allele predates the split between *T. knulli* and *T. poppensis* (the posterior probability of this is 0.60), our results suggest that the Redwood allele in *T. knulli* diverged from *T. poppensis* millions of years ago. More importantly, the inversion appears to have been maintained as a polymorphism within *T. knulli* for millions of years (based on both phylogenetic and population genetic models), whether or not allelic divergence predates the species divergence.

Discussion

Genetic variation is the ultimate fuel for evolution, but it remains unclear whether and how it can be maintained for extended periods of time. The maintenance of variation is particularly puzzling given that drift and many forms of natural selection tend to erode variation, depleting evolution's fuel reservoir. Our results have broad implications for understanding the long-term maintenance of genetic variation, and the capacity to adapt to challenging environments. Specifically, we discovered an ancient chromosomal inversion in *Timema* stick insects, and reported that it has been maintained as polymorphic for millions of years and that it likely facilitates the use of a challenging host plant (Redwood). We combined genomics, experiments, and evolutionary modeling to elucidate the processes maintaining variation, finding a role for life-history trade-offs (i.e., balancing selection), local adaptation to different hosts, and gene flow among populations (we ruled out recent introgression from another species). We then used models to show how such multi-layered regimes of selection and gene flow provide resilience that buffers populations against the loss of genetic variation, maintaining future evolutionary potential.

Beyond the maintenance of variation, our results have implications for understanding local adaptation. This process is a hallmark of evolution and is known to be common, but its dynamics

remain poorly understood because studying such dynamics often requires genetic analyses of adaptive mutations, whose identification has only recently become more feasible [36–38]. In this context, adaptation might involve the fixation of mutations that are beneficial in a new environment. This is a 'directional selection' hypothesis, often invoked in classical population genetics thinking and models [39, 40]. Alternatively, adaptation may involve shifts in allele-frequencies rather than fixation, representing disruption of a pre-existing evolutionary balance [41–43]. Such shifts might stem from standing genetic variation, and could occur via changes in the weight of balancing selection that maintains alternative alleles, slightly towards one allele or the other. This is a 'shifting balancing selection' hypothesis, often emphasized in the ecological genetics literature [9]. Our results are broadly consistent with this latter hypothesis. Still, the functional significance of the *Perform* inversion remains to be resolved. For example, the inversion could contribute to adaptation by suppressing recombination among linked loci that affect performance thereby creating a supergene, or the breakpoint mutations of the inversion could themselves be responsible for the observed fitness effects of this structural variant [10, 28, 44, 45].

Our results also have relevance for understanding the spatial context of evolution, namely the potential for gene flow and recombination between populations. Specifically, at spatial scales allowing gene flow, recombination will occur between populations. This can result in the breakdown of adaptive gene combinations, frustrating the ability of divergent selection to generate multi-locus local adaptation [46]. Thus, factors that reduce recombination, such as chromosomal inversions, are predicted to evolve when gene flow occurs [10, 47]. Gene flow is also relevant as it can modulate the degree to which alleles can move around in space and time, as increasingly documented in cases of adaptive introgression [16, 48–50]. We here demonstrated a key role for gene flow in the maintenance of genetic variation, but further work is required to test its role in the initial origin of inversion polymorphism.

In conclusion, although inversion evolution has received much recent attention [13–17, 28, 51], studies that directly elucidate the processes affecting inversions are still few [18–20]. This makes it difficult to connect data and theory, and precludes objective evaluation of ideas that have emerged over the last century concerning the evolutionary dynamics and role of inversions. Studies estimating selection on inversions are needed, and we provided such a study here, thereby elucidating how genetic variation can be maintained for millions of years. We find that rather than being a nuisance, complexity of evolutionary processes can generate resilience that buffers populations against the loss of variation. Further studies of the maintenance of ancient genetic variants, including inversions, are required to solidify the general importance of combinations of multiple evolutionary processes for maintaining genetic variation.

Methods

Measuring host-associated genetic differentiation. We used previously published single nucleotide polymorphism (SNP) data, obtained by genotyping-by-sequencing (GBS), to quantify host plant-associated genetic differentiation in four *Timema* species–*T. californicum*, *T. knulli*, *T. landelsensis*, and *T. poppensis*. All four are sexual species from the monophyletic "Northern" *Timema* clade that live in the vicinity of Redwood and use multiple hosts in nature [31]. We focused our analyses to six pairwise comparisons of nearby (i.e., parapatric) populations on different hosts (Table S1). Genomic data from these populations were originally described by [31]. Here, we used SNPs and associated genotype likelihoods (from `vcf` files) generated through a more recent re-analysis of these genomic data by [32].

We first estimated allele frequencies in each population at each of 1139 to 8548 genome-wide SNPs (Table S1). This was done using the program `estpEM` (version 0.1) [52] (Dryad, `https://doi.org/10.5061/dryad.nq67q`), which implements the expectation-maximization (EM) algorithm from [53] to estimate allele frequencies while accounting for uncertainty in

genotypes as expressed by genotype likelihoods. We used a convergence tolerance of 0.001 and allowed for a maximum of 40 EM iterations. Then, for each pair of populations, we computed $\mathrm{F_{ST}} = (H_T - H_S)/H_T$ for each SNP, and then summarized the distribution of $\mathrm{F_{ST}}$ for each linkage group (as defined by the *T. cristinae* genome to which these data were aligned) by computing the mean and various percentiles. Here, H_S and H_T denote the expected heterozygosities for the (sub)populations and the total, respectively. These calculations were performed in `R` (version 4.0.2) (see `https://github.com/zgompert/TimemaFusion/blob/main/hostFst.R`).

Generating the *T. knulli* genome and assigning chromosome numbers. We generated a *de novo* reference genome for *T. knulli* using a combination of PacBio and Illumina reads from a Hi-C genomic library. DNA extraction, library preparation, DNA sequencing, and *de novo* genome assembly were performed by Dovetail Genomics. A single female stick insect was used for the assembly, and the individual was chosen based on a preliminary analysis that suggested it was homozygous for the *Perform* Redwood allele. The final assembly was created using Dovetail's HiRise Assembly pipeline. It comprised 1,322,373,696 base pairs (bps) with an N50 of 83,614,905 bps. Using `BUSCO` version 4.0.5 with 255 BUSCOs, the assembly included 216 complete BUSCOs (212 single copy and four duplicated), 15 fragmented BUSCOs and 24 missing BUSCOs.

Much of our recent work in *Timema*, including the analyses of host-associated genetic differentiation described in the previous section, has relied on a *T. cristinae* reference genome and associated linkage map, with each linkage group comprising multiple moderately large scaffolds (version 1.3c2; this genome comes from a melanic stick insect; see [5, 27]). We wished to identify chromosomes (scaffolds) homologous to the *T. cristinae* linkage groups in our *T. knulli* genome for consistency in chromosome (linkage group) names and numbering. To do this, 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 [28]. 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 [54]. 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 [55, 56]. For this, we first used `RepeatMasker` (version 4.0.7) to mask repetitive regions of the genome [57]; this was done using a repeat library developed for *Timema* [28]. 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 [58] (`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. There was a one-to-one correspondence between *T. cristinae* and *T. knulli* chromosomes with one exception, *T. cristinae* chromosomes 1 and 3 were represented by a single fused chromosome in *T. knulli* (hereafter chro-

mosome 1), consistent with cytological work showing that *T. knulli* has one fewer chromosome than *T. cristinae* [34]. Thus, we were able to map our older *T. cristinae* linkage map numbers to the large scaffolds (i.e., chromosomes) in *T. knulli*.

***Timema knulli* sample collection.** In spring 2019 (March 16-18), we collected 138 *T. knulli* for population genomic analyses and for use in a performance rearing experiment (described below). Most stick insects were collected at one of two localities–BCTURN, where *Ceanothus* is the main host and *Timema* are not found on Redwood (N = 37), and BCE where both *Ceanothus* and Redwood are hosts (N = 68 and N = 24, respectively) (Table S3). Ten additional *T. knulli* were collected from three localities near BCE; BCOG (N = 1 on *Ceanothus*), BCSH (N = 1 on Redwood) and BCXD (N = 8 on *Ceanothus*). Stick insects were collected in sweep nets by beating host plants with a stick, as in past work [31, 59]. Captured insects were placed in plastic tubes, and kept in a cooler with ice for 1-2 days during transplantation to the laboratory for use in the performance experiment, as detailed below.

DNA extraction, library preparation and sequencing. After the performance experiment (see details below), we isolated DNA from each of 138 *T. knulli*. Frozen legs from each individual were ground into powder form using a Qiagen TissueLyser (Qiagen Inc., Valencia, CA). Genomic DNA was then extracted using Qiagen DNeasy Blood and Tissue kits, using a protocol with slightly altered incubation temperatures and times. We used a reduced-representation technique (i.e., genotyping-by-sequencing or GBS) to construct DNA sequencing libraries following the protocol detailed in [60]. Genomic DNA from each individual was digested with two restriction endonucleases, MseI (four base recognition site) and EcoRI (six base recognition site). Illumina adaptors with unique 8-10 bp DNA barcodes for each individual were ligated to EcoRI cut sites, and a base Illumina adaptor was ligated to MseI cut sites. Barcoded fragment libraries were then PCR amplified using Illumina primers and a high-fidelity proofreading polymerase (Iproof, BioRad, Hercules, CA). PCR products were pooled into a single library which

was then quality screened using an Agilent BioAnalzyer automated electrophoresis device. To reduce the portion of the genome targeted for sequencing, the reduced-representation library was then size-selected for DNA fragments 350-450 bp in length using a Pippin Prep quantitative electrophoresis unit (Sage Science, Beverly, MA) at the University of Texas Genome Sequencing and Analysis Facility (UTGSAF). The size-selected library was then sequenced using S2 chemistry and a single lane on an Illumina NovaSeq 4000 at UTGSAF.

DNA sequence alignment, variant calling, filtering and genotype estimation. We aligned the newly acquired *T. knulli* GBS reads to our new *T. knulli* reference genome. This was done with the `aln` and `samse` algorithms from `bwa` (version 0.7.17-r1188) [61]. For alignment, we set the maximum number of allowed mismatches to 4, allowed only 2 mismatches in the first 20 bp of the alignment, trimmed bases with quality scores $<$10, and only output alignments for reads with a single, best alignment. We then used `samtools` (version 1.5) to compress, sort and index the alignments [62]. We then used `samtools` (version 1.5) and `bcftools` (version 1.6) for variant calling [62]. Here, we used the consensus caller (`-c`), applied the recommended mapping quality adjustment for Illumina data (`-C 50`), and only output SNPs when the probability of all individuals being homozygous for the reference allele conditional on the data was $<$0.01. We then used a series of `Perl` scripts to filter the variant set. Specifically, we only retained SNPs that met the following criteria: $2\times$ minimum coverage per individual, a minimum of 10 reads supporting the non-reference allele, Mann-Whitney P-values for base quality, mapping quality and read position rank-sum tests $>$ 0.005, a minimum ratio of variant confidence to non-reference read depth of 2, a minimum mapping quality of 30, no more than 20% of individuals with missing data, only two alleles observed, and coverage not exceeding 3 SDs of the mean coverage (at the SNP level). This left us with 64,650 SNPs for further analysis.

We then used the (ad)mixture model implemented in `entropy` (version 1.2) to obtain Bayesian estimates of genotypes [63, 64]. This model uses a mixture prior on genotypes for

each locus and individual based on co-estimated allele frequencies from a series of k hypothetical source populations (similar to the admixture model from [65]). The model also accounts for uncertainty in genotypes arising from finite sequence coverage and possible sequencing errors as captured by the genotype likelihoods computed with `bcftools`. We estimated genotypes using Markov chain Monte Carlo (MCMC) and assuming either 2 or 3 source populations (i.e., our estimates integrate over these two possibilities). We ran 10 MCMC chains total (5 each for 2 and 3 source populations), each comprising 8000 steps, a 5000 step burnin and a thinning interval of 3. MCMC output was visually inspected to ensure (probable) convergence of the chains to the posterior distribution. Bayesian genotype estimates were then obtained by taking the posterior mean of the number of non-reference alleles (0, 1, or 2) for each locus and individual (these estimates are not constrained to integer values).

Delineating the *Perform* locus. We used principal component analysis (PCA) to delineate the region of *T. knulli* chromosome 11 associated with host-plant use (feeding on *Ceanothus* versus Redwood), i.e., the *Perform* locus. First, we conducted separate PCA ordinations of the genetic data (centered but not standardized genotype matrixes) for the 62,093 SNPs not on chromosome 11 and the 2557 SNPs on chromosome 11. Only the PCA of chromosome 11 showed host-associated genetic structure, and thus we then focused on chromosome 11. To localize the portion of chromosome 11 exhibiting this pattern, we performed PCA in 100-SNP sliding windows along chromosome 11. We summarized each PCA by the eigenvalue associated with the first eigenvector. Larger values coincide with greater genetic structure along this first PCA axis. All PCAs were done with the `prcomp` function in `R` (version 4.0.2). Visual inspection of the eigenvalues indicated a broad peak of high eigenvalues (accentuated structure) spanning much of chromosome 11. We fit a Hidden Markov model to the eigenvalues in `R` with the `HiddenMarkov` package (version 1.8.13) [66]. We allowed for two hidden states, which we initialized with expected values equal to the 25th and 75th percentiles of the empirical eigen-

value distribution across chromosome 11. We assumed a normal distribution for the observed eigenvalues with standard deviations initialized at half the empirical standard deviation. We then estimated the hidden state means, standard deviations, and transitions between hidden states using the Baum-Welch algorithm (i.e., we set initial values for means and standard deviations but these were then refined with the Baum-Welch algorithm) [67]. For this, we allowed a maximum of 500 iterations and set the tolerance to $1e^{-4}$. This procedure identified high (mean eigenvalue = 4.6, SD = 0.45) and low (mean eigenvalue = 2.8, SD = 0.29) states. We then used the Viterbi algorithm for decoding, that is for inferring the most likely hidden state for each 100 SNP window [68]. A single contiguous set of 100 SNP windows was assigned to the high state, which we hereafter refer to as the *Perform* locus. This region (i.e., the *Perform* locus) includes base positions 13,093,370 to 43,606,674 (i.e., $\sim$30 megabases) of *T. knulli* chromosome 11.

Determining *Perform* is a chromosomal inversion. We used a series of comparative genome alignments to test the hypothesis that the Perform locus is an inversion. Specifically, we performed pairwise whole-genome alignments for our *de novo* chromosome-level reference genomes for *T. knulli* (described in this paper), *T. cristinae* (the green striped morph) [28], and *T. chumash* [27]. Repetitive genomic regions were masked prior to genome alignment using `RepeatMasker` (version 4.0.7) and a *Timema* repeat library from [28]. We ran `RepeatMasker` using the slow/sensitive search (`-s`) with the NCBI engine. We then used `cactus` (version 1.0.0) to align each pair of genomes [55,56]. `cactus` creates genome alignment graphs, which can represent genome rearrangements and copy number variation. We then used `HAL` (Hierarchical Alignment) `tools` (version 2.1) to extract synteny blocks from the genome graphs. This was specifically done with `HalSynteny` with the default lower bound for synteny blocks of 5000 bps [58]. We then constructed sequence alignment dot plots from the synteny blocks using `R` (version 4.0.2) to visualize inversions and other structural variation between species. These patterns of structural variation were compared to the bounds of the *Perform* locus, delimited

within *T. knulli* as described above using the PCA approach .

The comparative alignments described above demonstrated that *Perform* coincides with an inversion in the Redwood *T. knulli* allele relative to *T. cristinae* and *T. chumash*. We used Oxford Nanopore long-read sequencing [69] to verify that this genomic region is a segregating inversion within *T. knulli* (as strongly suggested by the PCA results). 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 [70]. To do this, we extracted high-molecular weight DNA from a single *T. knulli* collected from BCEC (on *Ceanothus*) where the *Ceanothus* allele (that is the expected ancestral, non-inverted allele) occurs at high frequency. This was done with Qiagen's MatAttract 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$\times$ genome coverage). Note that while this is low coverage, it proved sufficient to validate the expected inversion as described below.

We first used `NanoFilt` [71] 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) [72]. We used the preset option for mapping Nanopore reads against a reference (`-x map-ont`) and used soft clipping for supplementary alignments. `samtools` (ver-

sion 1.12) was used to compress, sort and index the alignments [62]. We then used `sniffles2` (version 2.0.3) to call structural variants [73]. 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 our results above and Figure S2 for details).

Performance experiment. We conducted a laboratory experiment to test for a potential effect of the *Perform* locus on performance (here growth and survival) in *T. knulli* reared on *Ceanothus* or Redwood. For this experiment, each of the 138 *T. knulli* collected (see "*T. knulli* sample collection" above) were placed individually in 500 millimeter plastic containers, with air holes for breathing punched into the lid containers using a needle. Each stick insect was then fed fresh plant material from either *Ceanothus* or Redwood every second day (when survival was recorded, see below). Host-plant treatment was determined randomly and was independent of the host from which the stick insect was collected. We then measured weight and survival at 15 and 21 days as metrics of performance, and survival (dead or alive) was monitored every second day for the course of the 21-day experiment.

Testing for associations between *Perform* and *T. knulli* performance. We next tested for an association between *Perform* genotype and weight and survival on *Ceanothus* and Redwood during the performance experiment. We used PCA and k-means clustering to assign *Peform* genotypes (following, e.g., [12]). Specifically, we performed a PCA of the SNP genotypes for SNPs within the *Perform* locus; this was done on the centered but not standardized genotype matrix. We then used k-means clustering with three centers to assign each individual to a cluster based on the first PC from the ordination of SNPs in the *Perform* locus. We then fit models for 15 and 21 day weight (linear models) and survival (generalized linear model with binomial response and logit link) on each host plant as a function of *Perform* genotype (i.e., we

fit distinct models for each of the two host-plant treatments). Here, genotype corresponds to the assigned cluster number with homozygous clusters coded as 0 and 2 and the heterozygous (intermediate on PC1) cluster coded as 1 [12]. We removed the effects of sex and developmental stage on weight prior to the analyses, and dropped *T. knulli* from BCTURN to avoid possible confounding effects of population structure. Models were fit in `R` with the `lm` and `glm` functions (version 4.0.2).

Modeling 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 [74, 75]. 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) [6, 76]. This statistical model approximates several population genetic models, including an island model of drift-gene flow equilibrium [77–80]. 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, α = 0.5, β = 0.5). We fit this model in R using Hamiltonian Monte Carlo via the `R` interface with `STAN` (`rstan` version 2.21.2) [81]. 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 reduced linkage disequilibrium 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 [82]. 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 balancing selection 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) 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 $PgPg$ 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 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 ($PgPg$ and $PsPs$, 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 versus balancing selection 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 overdominance, 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 [83]. 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 [84]. We used the vector of final (at generation 2500) *Perform* allele frequencies for the three popula-

tions 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) [85] 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 versus balancing selection for the two host plants). The model of balancing selection on both host plants had the highest posterior probability. Consequently, we estimated the selection coefficients ($s1$ and $s2$) on each host plant under the balancing selection 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% ($\sim$1000) of simulations with balancing selection 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) [85].

Additional simulations testing if a combination of processes buffers populations against the loss of genetic variation. We conducted an additional set of forward-time simulations of evolution to determine whether and to what extent our best fit model (balancing selection with gene flow) maintained variation at the *Perform* locus (i.e., we conducted a predictive check of this model) and how this compared to two counterfactual models–one with directional selection and gene flow and one with balancing selection and no gene flow. We used the same general model described above. For the balancing selection with gene flow simulations, we sampled effective population size and gene flow parameters from the same prior distributions used above and then sampled selection coefficients from the posterior distributions inferred from ABC. For the balancing selection without gene flow simulations, we did the same thing, except we set the

migration rates to 0. Lastly, for directional selection with gene flow, we sampled selection coefficients from a posterior inferred from the ABC model when only considering the directional selection model (i.e., forcing directional selection). We ran 50 simulations (50 samples from the prior or posterior distributions depending on the parameters) under each of the three models with each running for 250,000 generations. Initial *Perform* allele frequencies were set to 0.5 for all simulations. These simulations were conducted in `R` (version 4.1.3). We then compared the outcome of these simulations to the observed variability at the *Perform* locus.

Dating the chromosomal inversion. We first used a phylogenetic approach to estimate the divergence time between the *Perform* chromosomal variants, i.e., alleles (as in [12]. For this, we used GBS data from 138 *T. knulli* described above, along with 69 newly sequenced *T. petita* (from site 101S, latitude = 35.73°N, longitude = 121.31°W), and 329 *T. poppensis* and 86 *T. californicum* originally described in [28]. These data were aligned to the *T. knulli* reference genome using the `bwa aln` algorithm (version 0.7.17-r1198) and alignments were compressed, sorted and indexed with `samtools` as described above for the *T. knulli* samples [61, 62]. We identified variable nucleotides (SNPs) across this full set of samples but only within the *Perform* locus using `samtools` (version 1.5) and `bcftools` (version 1.6). Other than considering only the *Perform* locus, variant calling options and subsequent filtering were as described above for *T. knulli*. We then determined the number of invariant bases of each type (A, C, G, or T) within the *Perform* locus, as this information is part of the phylogenetic model. Specifically, using the `samtools depth` command, we determined coverage for each individual at each site within *Perform* that was not called as a SNP (even before filtering). We counted the site as invariant if we had data for at least 80% of the individuals with a mean coverage of at least 2× per individual. This resulted in 789 variable sites (SNPs) and 18,425, 11,610, 12,007, and 18,570 invariant As, Cs, Gs, and Ts, respectively. We then used `Perl` scripts to convert the variant file to a nexus alignment and to choose a subset of indi-

viduals for phylogenetic analysis (the conversion scripts are from [12] and are available from GitHub, `https://github.com/zgompert/TimemaFusion`). Specifically, for the outgroup taxa *T. californicum*, *T. poppensis* and *T. petita*, we chose the 8-10 (10 for *T. petita* only) individuals with the least missing data for the aligned SNPs, and for *T. knulli* we retained 33 individuals from BCE C (host = *Ceanothus*) and 25 from the parapatric population BCE RW (host = Redwood).

We then used BEAST2 (version 2.6.6) [86] to estimate the divergence times between the *Perform* chromosomal variants in *T. knulli*. We encoded information on the invariant sites using the `constantSiteWeights` option. We fit the GTR sequence evolution model with rate heterogeneity that approximated a gamma distribution using four rate categories. We assumed a relaxed log-normal clock [87] with a coalescent extended Bayesian skyline tree prior [88]. Following [12], we fit a gamma distribution to the previously inferred divergence time for all four of our taxa–*T. knulli*, *T. petita*, *T. californicum*, and *T. poppensis*–using the `fitdistr` function in `R`. This gives a gamma with α = 10.8509 and β = 0.973, which has a mean of 11.5 million years and standard deviation of 3.4 million years. We used this as the prior on the root divergence time and thus as a calibration point for our key divergence time of interest, that is between the two chromosomal variants in *T. knulli*. Our input `xml` file (`tknulli_perform_og.xml`) is available from GitHub (`https://github.com/zgompert/TimemaFusion`). We estimated the tree and associated divergence times based on 3 chains each comprising 10 million iterations. Posteriors were summarized in `R`.

Second, we estimated the divergence time in a population genetic context with the diffusion approximation approach implemented in δaδi [35]. We specifically followed an approach inspired by [89], 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 propotions), as in [89]. 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 $PgPg$ homozygote (the allele more common on Redwood) and one comprising 33 $PsPs$ 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 exchnage parameters $M_{12} = 2N_{anc}m_{12}$ and $M_{21}2N_{anc}m_{21}$ (here m_{12} and m_{21} are propotions) (see `https://github.com/zgompert/TimemaFusion/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* [90], 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 $2\times$ 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/TimemaFusion/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.

References

[1] Lewontin, R. C. & Hubby, J. L. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**, 595 (1966).

[2] Kimura, M. *The neutral theory of molecular evolution* (Cambridge University Press,

1983).

[3] Gillespie, J. H. *The causes of molecular evolution*, vol. 2 (Oxford University Press On Demand, 1994).

[4] Wittmann, M. J., Bergland, A. O., Feldman, M. W., Schmidt, P. S. & Petrov, D. A. Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. *Proceedings of the National Academy of Sciences* **114**, E9932–E9941 (2017).

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

[6] Gompert, Z., Springer, A., Brady, M., Chaturvedi, S. & Lucas, L. K. Genomic time-series data show that gene flow maintains high genetic diversity despite substantial genetic drift in a butterfly species. *Molecular Ecology* **30**, 4991–5008 (2021).

[7] Sturtevant, A. H. A case of rearrangement of genes in *Drosophila*. *Proceedings of the National Academy of Sciences* **7**, 235–237 (1921).

[8] Dobzhansky, T. Adaptive changes induced by natural selection in wild populations of *Drosophila*. *Evolution* 1–16 (1947).

[9] Ford, E. B. Ecological genetics. In *Ecological Genetics*, 1–11 (Springer, 1977).

[10] Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).

[11] Hoffmann, A. A. & Rieseberg, L. H. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics* **39**, 21 (2008).

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

[13] Ayala, D. *et al.* Association mapping desiccation resistance within chromosomal inversions in the african malaria vector *Anopheles gambiae*. *Molecular Ecology* **28**, 1333–1342 (2019).

[14] Jay, P. *et al.* Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nature Genetics* **53**, 288–293 (2021).

[15] Durmaz, E., Benson, C., Kapun, M., Schmidt, P. & Flatt, T. An inversion supergene in *Drosophila* underpins latitudinal clines in survival traits. *Journal of Evolutionary Biology* **31**, 1354–1364 (2018).

[16] Todesco, M. *et al.* Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**, 602–607 (2020).

[17] Koch, E. L. *et al.* Genetic variation for adaptive traits is associated with polymorphic inversions in *Littorina saxatilis*. *Evolution Letters* **5**, 196–213 (2021).

[18] Lowry, D. B. & Willis, J. H. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology* **8**, e1000500 (2010).

[19] Ayala, D., Guerrero, R. F. & Kirkpatrick, M. Reproductive isolation and local adaptation quantified for a chromosome inversion in a malaria mosquito. *Evolution* **67**, 946–958 (2013).

[20] Mérot, C., Llaurens, V., Normandeau, E., Bernatchez, L. & Wellenreuther, M. Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. *Nature Communications* **11**, 1–11 (2020).

[21] Llaurens, V., Whibley, A. & Joron, M. Genetic architecture and balancing selection: the life and death of differentiated variants. *Molecular Ecology* **26**, 2430–2448 (2017).

[22] Felsenstein, J. The theoretical population genetics of variable selection and migration. *Annual Review of Genetics* **10**, 253–280 (1976).

[23] Yeaman, S. & Jarvis, A. Regional heterogeneity and gene flow maintain variance in a quantitative trait within populations of lodgepole pine. *Proceedings of the Royal Society B: Biological Sciences* **273**, 1587–1593 (2006).

[24] Reid, J. M. *et al.* Immigration counter-acts local micro-evolution of a major fitness component: Migration-selection balance in free-living song sparrows. *Evolution Letters* **5**, 48–60 (2021).

[25] 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).

[26] 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).

[27] Nosil, P. *et al.* Ecology shapes epistasis in a genotype–phenotype–fitness map for stick insect colour. *Nature Ecology & Evolution* **4**, 1673–1684 (2020).

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

[29] Muschick, M., Soria-Carrasco, V., Feder, J. L., Gompert, Z. & Nosil, P. Adaptive zones shape the magnitude of premating reproductive isolation in *Timema* stick insects. *Philosophical Transactions of the Royal Society B* **375**, 20190541 (2020).

[30] Larose, C., Rasmann, S. & Schwander, T. Evolutionary dynamics of specialisation in herbivorous stick insects. *Ecology Letters* **22**, 354–364 (2019).

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

[32] Chaturvedi, S. *et al.* Climatic similarity and genomic background shape the extent of parallel adaptation in *Timema* stick insects. *Nature Ecology and Evolution* (2022).

[33] Nosil, P. *et al.* Genomic consequences of multiple speciation processes in a stick insect. *Proc. R. Soc. B* rspb20120813 (2012).

[34] Schwander, T. & Crespi, B. J. Multiple direct transitions from sexual reproduction to apomictic parthenogenesis in *Timema* stick insects. *Evolution* **63**, 84–103 (2009).

[35] Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* **5**, e1000695 (2009).

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

[37] Archambeault, S. L., Bärtschi, L. R., Merminod, A. D. & Peichel, C. L. Adaptation via pleiotropy and linkage: association mapping reveals a complex genetic architecture within the stickleback *Eda* locus. *Evolution letters* **4**, 282–301 (2020).

[38] Villoutreix, R. *et al.* Testing for fitness epistasis in a transplant experiment identifies a candidate adaptive locus in *Timema* stick insects. *Philosophical Transactions of the Royal Society B* **377**, 20200508 (2022).

[39] Fisher, R. A. *The genetical theory of natural selection* (Oxford, 1930).

[40] Sabeti, P. C. *et al.* Positive natural selection in the human lineage. *Science* **312**, 1614–1620 (2006).

[41] Burke, M. K. *et al.* Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* **467**, 587–590 (2010).

[42] Pritchard, J. K. & Di Rienzo, A. Adaptation–not by sweeps alone. *Nature Reviews Genetics* **11**, 665–667 (2010).

[43] Rudman, S. M. *et al.* Direct observation of adaptive tracking on ecological time scales in *Drosophila*. *Science* **375**, eabj7484 (2022).

[44] Villoutreix, R. *et al.* Inversion breakpoints and the evolution of supergenes. *Molecular Ecology* **30**, 2738–2755 (2021).

[45] Zhang, L., Reifová, R., Halenková, Z. & Gompert, Z. How important are structural variants for speciation? *Genes* **12**, 1084 (2021).

[46] Felsenstein, J. Skepticism towards santa rosalia, or why are there so few kinds of animals? *Evolution* 124–138 (1981).

[47] Yeaman, S. & Whitlock, M. C. The genetic architecture of adaptation under migration–selection balance. *Evolution* **65**, 1897–1911 (2011).

[48] Dasmahapatra, K. K. *et al.* Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012).

[49] Huerta-Sánchez, E. *et al.* Altitude adaptation in tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014).

[50] Giska, I. *et al.* Introgression drives repeated evolution of winter coat color polymorphism in hares. *Proceedings of the National Academy of Sciences* **116**, 24150–24156 (2019).

[51] Schaal, S. M., Haller, B. C. & Lotterhos, K. E. Inversion invasions: when the genetic basis of local adaptation is concentrated within inversions in the face of gene flow. *Philosophical Transactions of the Royal Society B* **377**, 20210200 (2022).

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

[53] 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).

[54] Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).

[55] Paten, B. *et al.* `Cactus`: Algorithms for genome multiple sequence alignment. *Genome Research* **21**, 1512–1528 (2011).

[56] Armstrong, J. *et al.* Progressive `cactus` is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).

[57] Smit, A. F. Repeat-masker open-3.0. *http://www. repeatmasker. org* (2010).

[58] Krasheninnikova, K. *et al.* `halSynteny`: a fast, easy-to-use conserved synteny block construction method for multiple whole-genome alignments. *GigaScience* **9**, giaa047 (2020).

[59] Sandoval, C. P. The effects of the relative geographic scales of gene flow and selection on morph frequencies in the walking-stick *Timema cristinae*. *Evolution* **48**, 1866–1879 (1994).

[60] Parchman, T. L. *et al.* Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* **21**, 2991–3005 (2012).

[61] Li, H. & Durbin, R. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

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

[63] Gompert, Z. *et al.* Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology* **23**, 4555–4573 (2014). URL `http://dx.doi.org/10.1111/mec.12811`.

[64] Shastry, V. *et al.* Model-based genotype and ancestry estimation for potential hybrids with mixed-ploidy. *Molecular Ecology Resources* **21**, 1434–1451 (2021).

[65] Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).

[66] Harte, D. `HiddenMarkov`*: Hidden Markov Models*. Statistics Research Associates, Wellington (2021). URL `https://www.statsresearch.co.nz/dsh/sslib/`. `R` package version 1.8-13.

[67] 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).

[68] Forney, G. D. The Viterbi algorithm. *Proceedings of the IEEE* **61**, 268–278 (1973).

[69] Lu, H., Giordano, F. & Ning, Z. Oxford Nanopore MinION sequencing and genome assembly. *Genomics, Proteomics & Bioinformatics* **14**, 265–279 (2016).

[70] Zhang, L., Chaturvedi, S., Nice, C. C., Lucas, L. K. & Gompert, Z. Population genomic evidence of selection on structural variants in a natural hybrid zone. *Molecular Ecology* (2022).

[71] De Coster, W., D'hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. `NanoPack`: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).

[72] Li, H. `Minimap2`: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

[73] Smolka, M. *et al.* Comprehensive structural variant detection: from mosaic to population-level. *BioRxiv* (2022).

[74] Beaumont, M. A., Zhang, W. & Balding, D. J. Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035 (2002).

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

[76] Gaggiotti, O. E. & Foll, M. Quantifying population structure using the F-model. *Molecular Ecology Resources* **10**, 821–830 (2010).

[77] Balding, D. J. & Nichols, R. A. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* **96**, 3–12 (1995).

[78] Nicholson, G. *et al.* Assessing population differentiation and isolation from single-nucleotide polymorphism data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 695–715 (2002).

[79] Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).

[80] Gompert, Z. *et al.* Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution* **66**, 2167–2181 (2012). URL `http://dx.doi.org/10.1111/j.1558-5646.2012.01587.x`.

[81] Stan Development Team. `RStan`: the `r` interface to `stan` (2020). URL `https://mc-stan.org/`. R package version 2.21.2.

[82] Hoffman, M. D., Gelman, A. *et al.* The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* **15**, 1593–1623 (2014).

[83] Ewens, W. J. *Mathematical Population Genetics: Theoretical Introduction*, vol. 1 (Springer, 2004).

[84] Galassi, M. *et al.* *GNU Scientific Library* (Network Theory Limited Godalming, 2002).

[85] Csillery, K., Francois, O. & Blum, M. G. B. `abc`: an `r` package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* (2012).

[86] Bouckaert, R. *et al.* `BEAST 2`: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**, e1003537 (2014).

[87] Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**, e88 (2006).

[88] Heled, J. & Drummond, A. J. Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology* **8**, 1–15 (2008).

[89] Blischak, P. D., Sajan, M., Barker, M. S. & Gutenkunst, R. N. Demographic history inference and the polyploid continuum. *bioRxiv* (2022).

[90] Liu, H. *et al.* 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**, 119–130 (2017).

Acknowledgments

This study is part of a project that has received funding from the European Research Council (ERC) to PN, under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 770826 EE-Dynamics). ZG was supported by the US NSF (DEB

1844941). The support and resources from the Center for High-Performance Computing at the University of Utah are gratefully acknowledged.

Author contributions

PN and ZG designed the study. PN, RV, MC, CC and TP conducted the experiments and generated the data. ZG and VSC analyzed the data. PN and ZG drafted the manuscript. All authors helped edit and revise the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

DNA sequence data are available from the NCBI SRA (Accession numbers pending) and data from the performance experiment are available from Dryad (DOI pending).

Code availability

Computer code used for analyses is available from GitHub (`https://github.com/zgompert/TimemaFusion`).

Figures

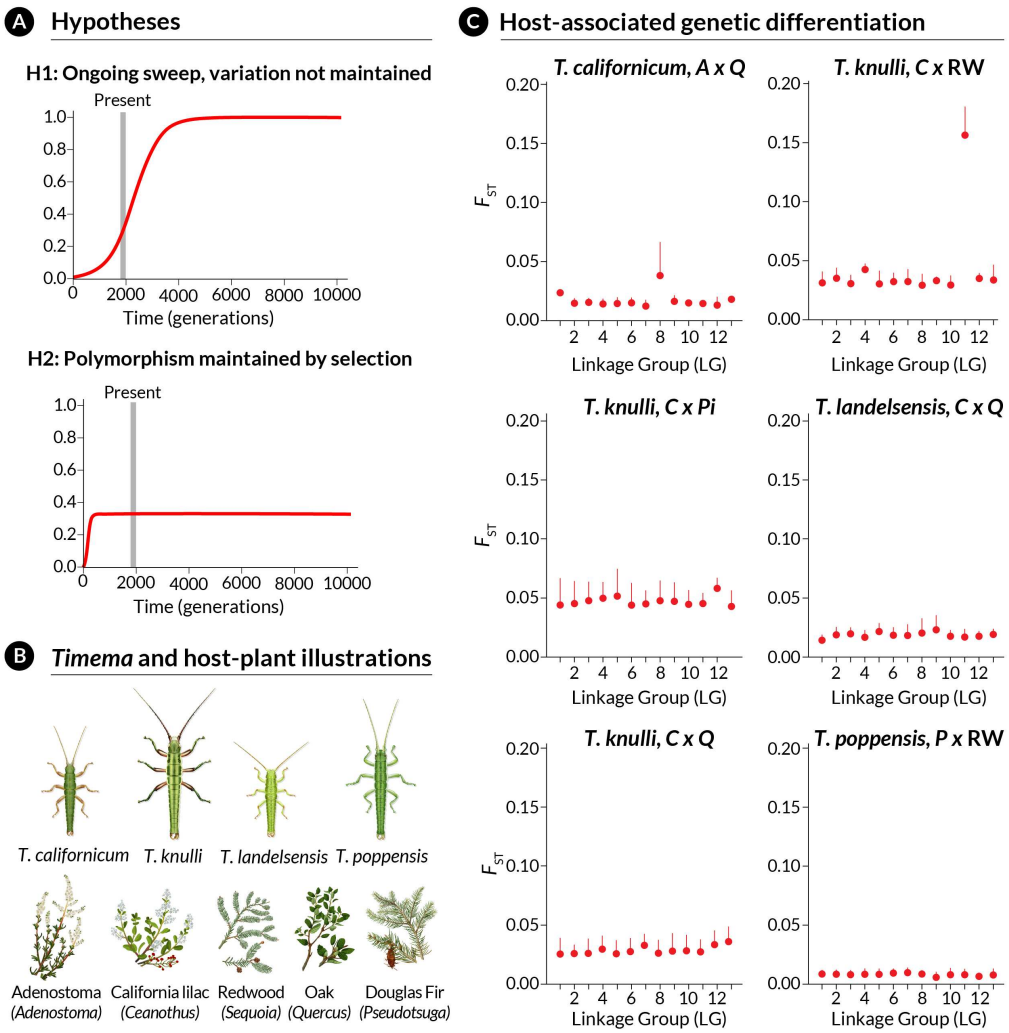

Figure 1: Conceptual overview and evidence of host-associated differentiation. Panel (A) illustrates our two alternative hypotheses of (H1) an ongoing sweep and (H2) polymorphism maintained by selection. Under the first hypothesis the inversion is young and in the process of sweeping; variation will not be maintained. Under the second hypothesis balancing selection promotes the long-term maintenance of inversion polymorphism. (B) Shows illustrations of *Timema* stick insects and their host plants, for the taxa studied here. Panel (C) summarizes genome-wide genetic differentiation for parapatric *Timema* populations on different hosts. Points denote mean F_{ST} for each of 13 *T. cristinae* linkage groups with horizontal lines extending to the 75th percentile of F_{ST} for that linkage group. Host abbreviations are A = *Adenostoma*, C = *Ceanothus*, P = *Pseudotsuga menziesii* (Douglas Fir), Pi = *Pinus* (pine), Q = *Quercus* (oak), and RW = *Sequoia sempervirens* (Redwood).

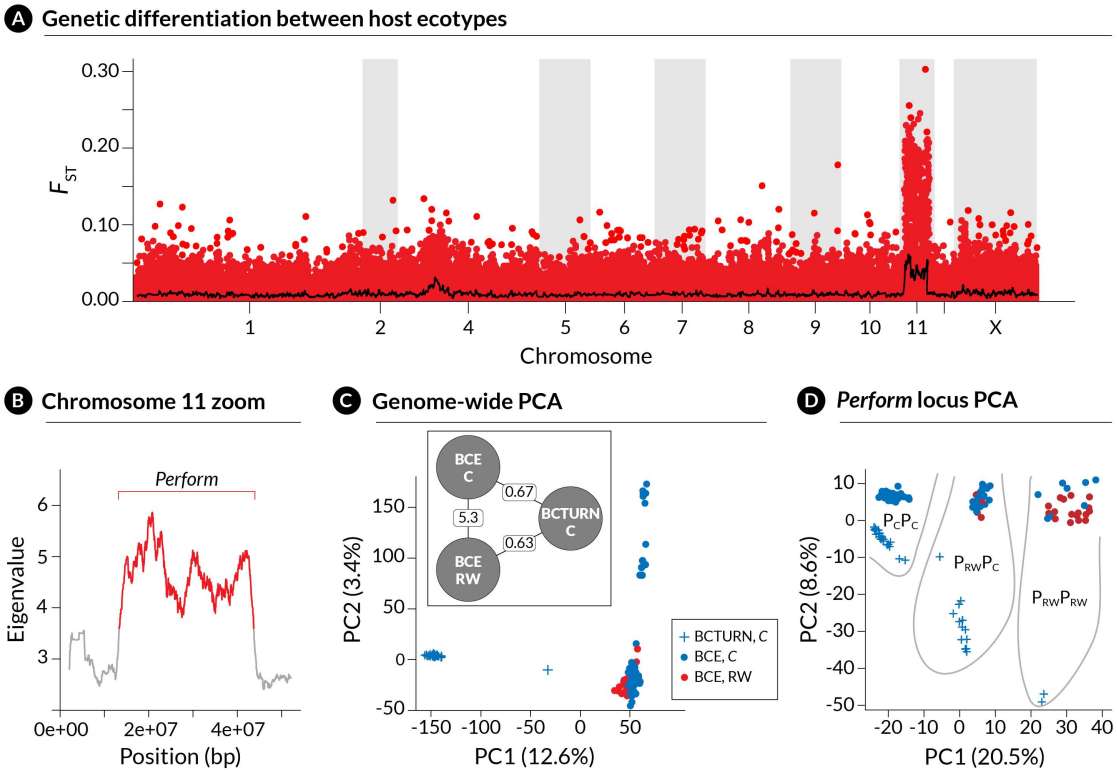

Figure 2: Genetic differentiation and structure associated with Redwood feeding in *Timema knulli*. These results are all based on the new reference genome for *T. knulli*. (A) Manhattan plot of $\mathrm{F_{ST}}$ between stick insects collected on *Ceanothus* versus Redwood at BCE. Points denote $\mathrm{F_{ST}}$ for individual SNPs. *Timema knulli* chromosomes are used here (chromosome 3 from *T. cristinae* is fused to chromosome 1; X = the X sex chromosome). Panel (B) shows eigenvalues for the first principal component of genetic variation in *T. knulli* (excluding BCTURN) in 100 SNP overlapping, sliding windows along chromosome 11. Colors denote alternative states as identified by a Hidden Markov model (HMM), with red denoting the elevated eigenvalue state and defining the bounds for the '*Perform*' locus on chromosome 11 (text for details). Panels (C) and (D) show summaries of genetic variation in *T. knulli* based on principal components analysis (PCA) for all SNPs not on chromosome 11 (C) and for the *Perform* locus only (D). Values for the first two principal components are shown with colors and symbols denoting locations and hosts. The inset in (C) is a schematic for the model used to infer neutral rates of gene flow among populations: BCE C (on *Ceanothus*), BCE RW (on Redwood) and BCTURN C (on *Ceanothus*). Point estimates of Nm, that is the number of migrants exchanged per generation, are shown on lines connecting the populations, and are consistent with a pattern of isolation by geographic distance.

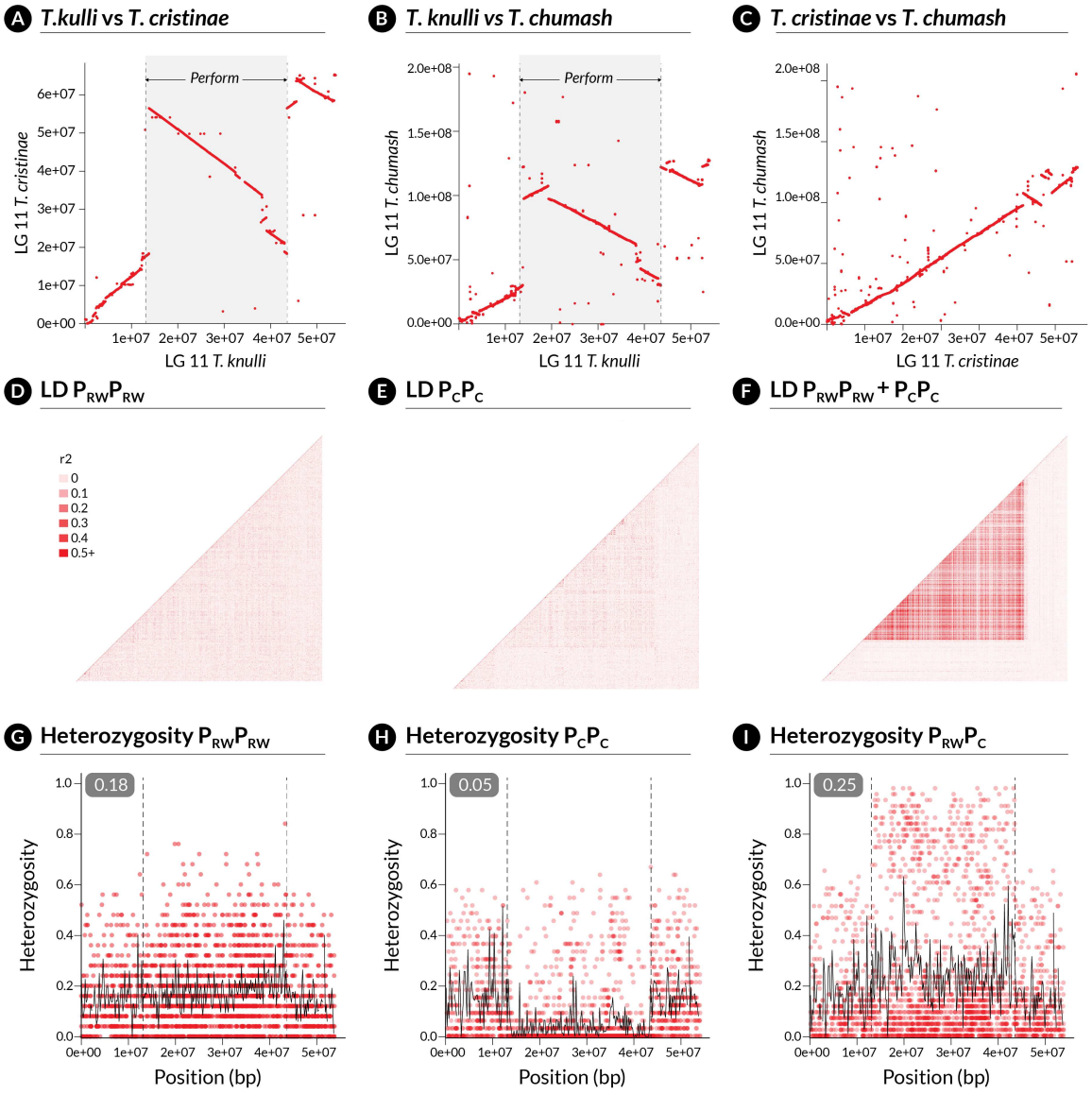
A T.kulli vs T. cristinae
B T. knulli vs T. chumash
C T. cristinae vs T. chumash
Perform
LG 11 T. cristinae
LG 11 T. chumash
LG 11 T. knulli
D LD PRWPRW
E LD PCPC
F LD PRWPRW + PCPC
r2
0
0.1
0.2
0.3
0.4
0.5+
G Heterozygosity PRWPRW
H Heterozygosity PCPC
I Heterozygosity PRWPC
0.18
0.05
0.25
Heterozygosity
Position (bp)

Figure 3: Genome alignments and evidence *Perform* is an ancient inversion. Dot plots show alignments of chromosome 11 for *T. knulli* and *T. cristinae* (A), *T. knulli* and *T. chumash* (B), and *T. cristiane* and *T. chumash* (C). Red 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 gray shaded region. A large inversion coinciding with the *Perform* locus is evident between *T. knulli* and both *T. cristiane* (A) and *T. chumash* (B), but no such inversion is found for *T. cristinae* versus *T. chumash*. Panel (D) summarizes the evidence for and estimated bounds of the *Perform* inversion within *T. knulli*. Specifically, horizontal black lines show the inferred bounds based on the *T. knulli* nanopore data, the comparative alignments in (A) and (B), and the eigenvalues from a PCA in *T. knulli* (see Figure 2B). Panel (E) shows the phylogeny for the *Perform* locus estimated with `BEAST2`. Colored points indicate taxa and inversion alleles (for *T. knulli* only) (Redwood = RW, *Ceanothus* = C). Bifurcations with posterior probability > 0.5 are shown with pie charts colored to denote posterior probabilities. Panel (F) shows the corresponding Bayesian posterior distributions for divergence times for *T. knulli* and *T. poppensis* based on genome-wide SNP data, for the *T. knulli* RW chromosomal variant and *T. poppensis* based on SNPs within the *Perform* locus, and for SNPs within the *T. knulli* RW and C *Perform* chromosomal variants. Points and horizontal lines denote posterior medians and 95% equal-tail probability intervals [ETPIs], respectively.

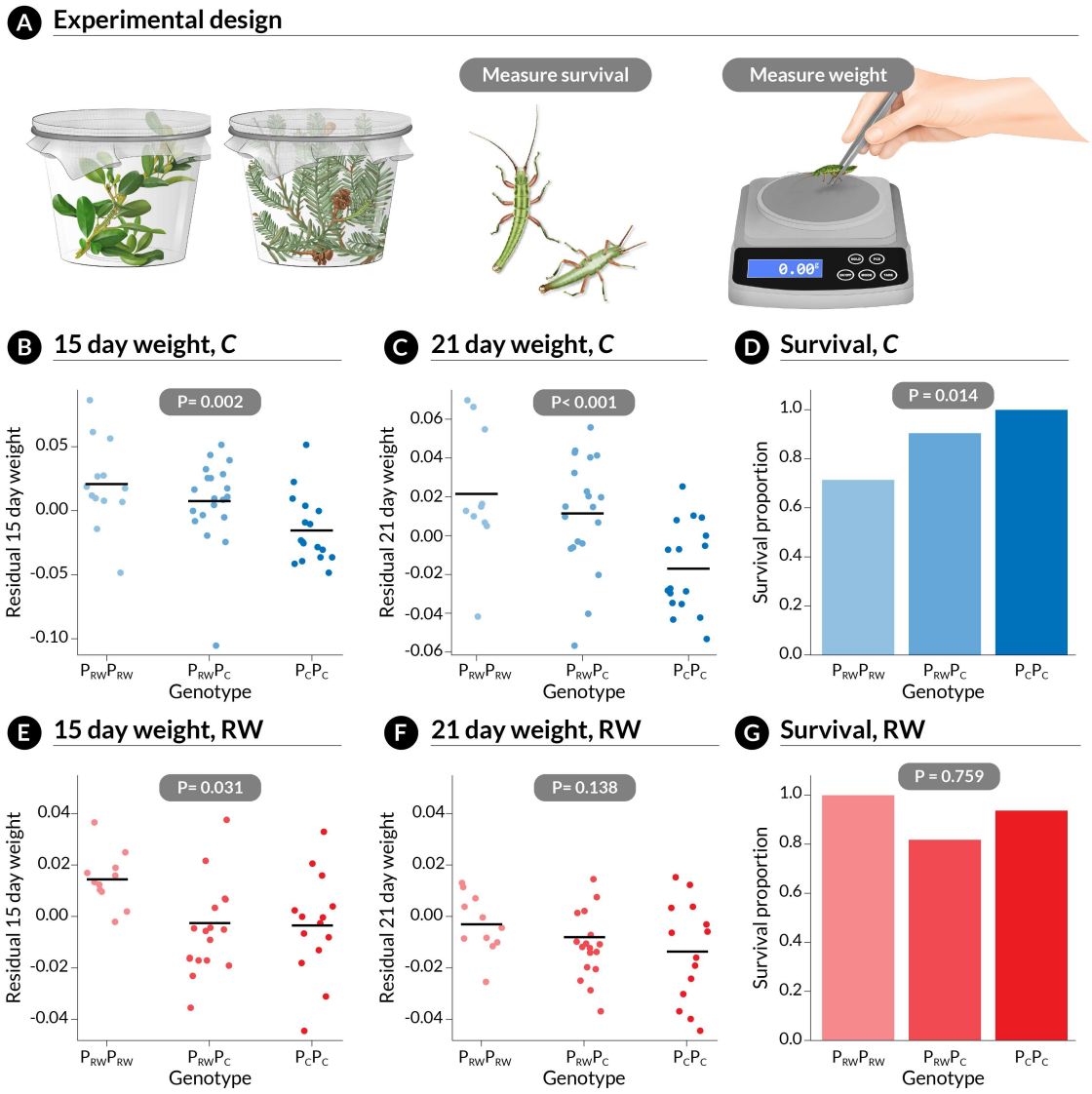

Figure 4: Summary of the rearing and genetic mapping experiments. Panel (A) illustrates the experimental design. Panels (B) and (C) show 15-and 21-day weight for *T. knulli* reared on *Ceanothus* based on their *Perform* genotype (i.e., 0 and 2 are alternative homozygotes, with 0 being homozygous for the Redwood allele, and 1 is the heterozygote). Points denote individuals (with a small jitter applied to the x-axis), horizontal lines give means for each genotype. The P-value for the null hypothesis of no effect of *Perform* is shown. A barplot (D) shows survival proportions on *Ceanothus* along with the P-value for the null model of no effect of genotype on survival. Analogous results are shown for *T. knulli* on Redwood (RW, *Sequoia*) in (E) (15-day weight), (F) (21-day weight) and (G) (survival).

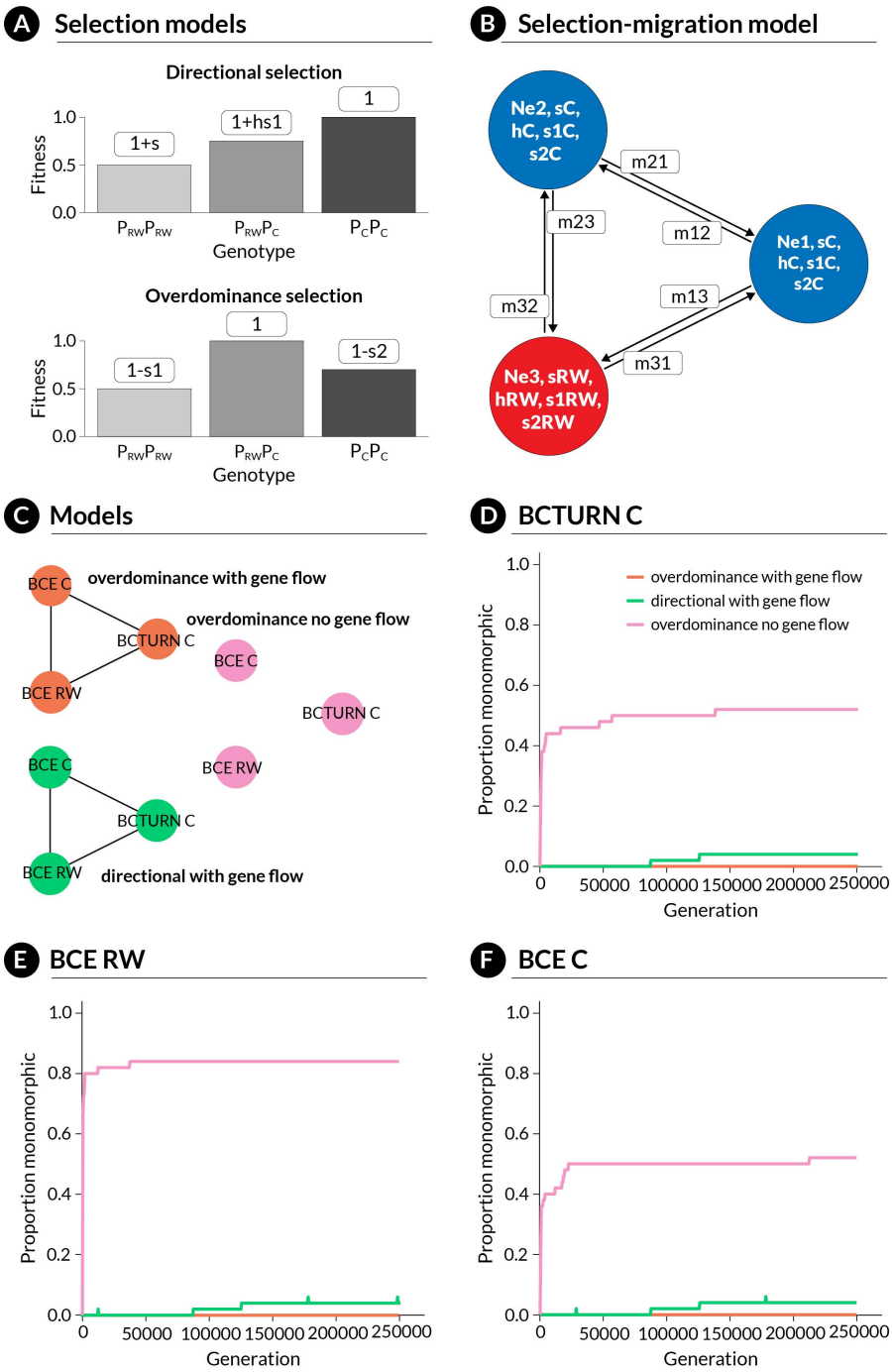
A Selection models
Directional selection
1.0
0.5
0.0
Fitness
1+s
1+hs1
1
PRWPRW
PRWPC
PCPC
Genotype
Overdominance selection
1.0
0.5
0.0
Fitness
1-s1
1
1-s2
PRWPRW
PRWPC
PCPC
Genotype
B Selection-migration model
Ne2, sC, hC, s1C, s2C
Ne1, sC, hC, s1C, s2C
Ne3, sRW, hRW, s1RW, s2RW
m21
m12
m23
m32
m13
m31
C Models
BCE C
overdominance with gene flow
BCTURN C
overdominance no gene flow
BCE C
BCE RW
BCTURN C
BCE C
BCE RW
BCTURN C
BCE RW
directional with gene flow
D BCTURN C
overdominance with gene flow
directional with gene flow
overdominance no gene flow
Proportion monomorphic
1.0
0.8
0.6
0.4
0.2
0.0
0
50000
100000
150000
200000
250000
Generation
E BCE RW
Proportion monomorphic
1.0
0.8
0.6
0.4
0.2
0.0
0
50000
100000
150000
200000
250000
Generation
F BCE C
Proportion monomorphic
1.0
0.8
0.6
0.4
0.2
0.0
0
50000
100000
150000
200000
250000
Generation

Figure 5: Summary of the ABC model and inferences. Panel (A) illustrates the fitness schemes and definitions of selection coefficients under directional versus balancing selection. Specifically, for directional selection s denotes the difference in relative fitness for alternative homozygotes and h gives the heterozygote effect (with $0 < h < 1$), whereas for balancing selection $s1$ and $s2$ denote the reductions in fitness for homozygotes relative to the heterozygote. Panel (B) summarizes the demographic component of the model. Colored circles correspond with populations with colors denoting host, red = Redwood and blue = *Ceanothus*. Populations have distinct effective population sizes (Ne) and selection coefficients (either s and h or $s1$ and $s2$) dictated by host (RW or C). Asymmetric gene flow is allowed as indicated by the migration edges. Panels (C) and (D) given model posterior probabilities. In (C) posteriors are given for DS = divergent directional selection on both hosts, BS/RW = balancing selection on Redwood and directional selection on *Ceanothus*, BS/C = directional selection on Redwood and balancing selection on *Ceanothus*, and BS = balancing selection on both hosts. In (D) marginal posteriors are shown for directional (DS) versus balancing selection (BS) on each host (indicated by color). Panel (E) shows the joint posterior for the fitness of RW versus C allele homozygotes on each host, where points denote individual samples from the posterior with contours overlain. (F) gives posterior estimates of the selection coefficients $s1$ and $s2$ (balancing selection) on *Ceanothus* (C) versus Redwood (RW). Points and numbers denote posterior medians and vertical bars indicate 95% credible intervals.

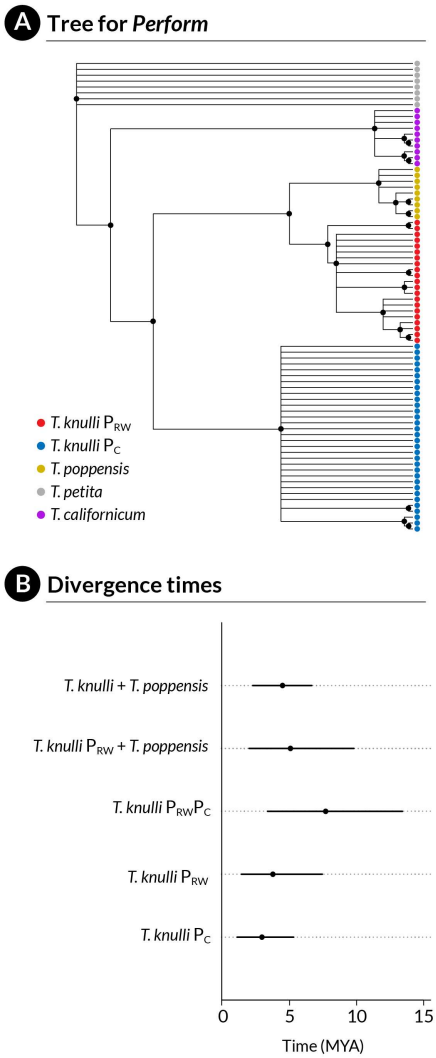

Figure 6: Summary of simulations testing the effects of balancing selection and gene flow on the maintenance of polymorphism at the *Perform* locus. Panel (A) illustrates the three models–balancing selection with gene flow (our best model), directional selection with gene flow, and balancing selection without gene flow–that we consider for the three focal populations analyzed with the ABC model, BCE on *Ceanothus* (BCE C), BCE on Redwood (BCE RW) and BCTURN (an allopatric *Ceanothus* population). Panels (B)–(D) show the proportion of replicate simulations in which variation at *Perform* was lost over time in BCTURN (B), BCE RW (C) and BCE C (D).

Supplement Tables and Figures

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. [31] 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 [5], 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*). 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	c3.07	121.60	1
BCTURN	C	36.08	121.61	37
BCXD	C	36.07	121.60	8

Table S4: Proportion 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
PgPg	1.00	1.00	0.86	0.57
PgPs	0.79	0.88	0.94	0.80
PsPs	0.92	1.00	1.00	1.00

Table S5: Model parameter estimates from $\delta\mathtt{a}\delta\mathtt{i}$. 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 homozygoes 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

Supplemental Figures

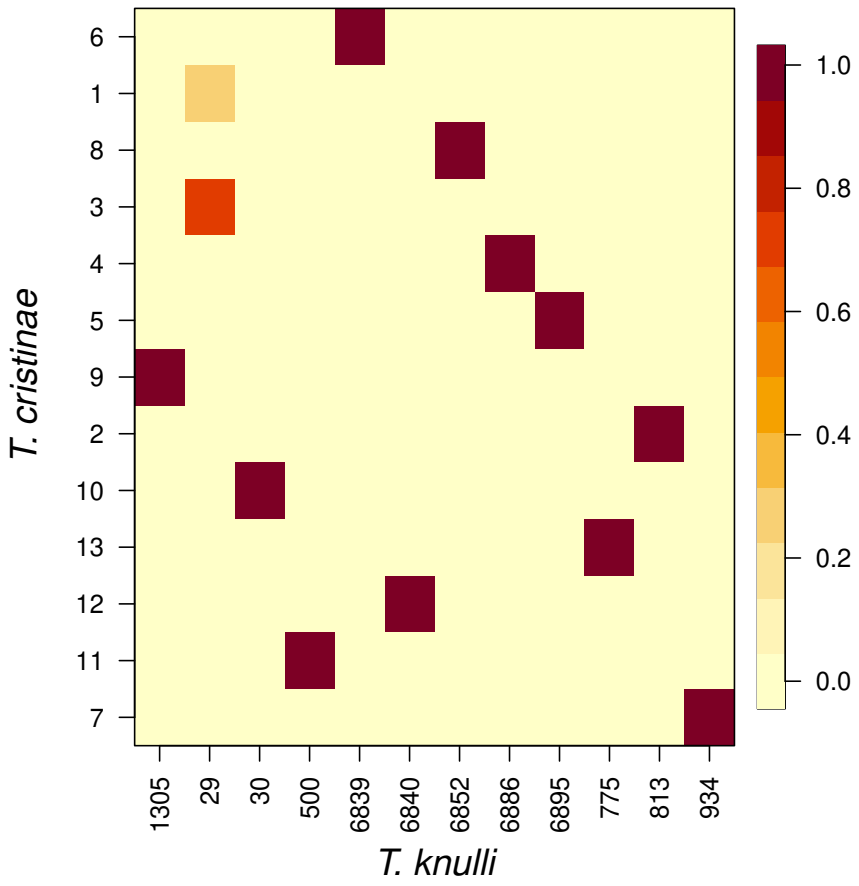

Figure S1: Heatmap shows the proportion synteny blocks on of each *T. knulli* scaffold that aligned to each of the *T. cristinae* chromosomes.

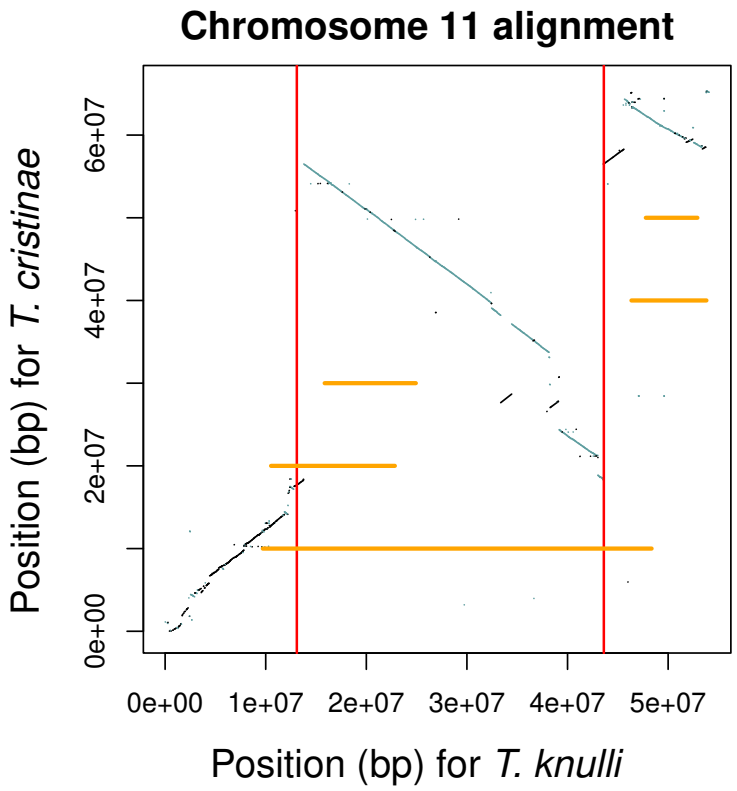

Figure S2: The dot plot shows the alignment of chromosome 11 for *T. knulli* (from Redwood) and *T. cristinae*. Line segments denote aligned genome regions with the orientation of the alignment shown by the direction of the lines. The bounds of the *Perform* locus in the *T. knulli* genome are denoted 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.

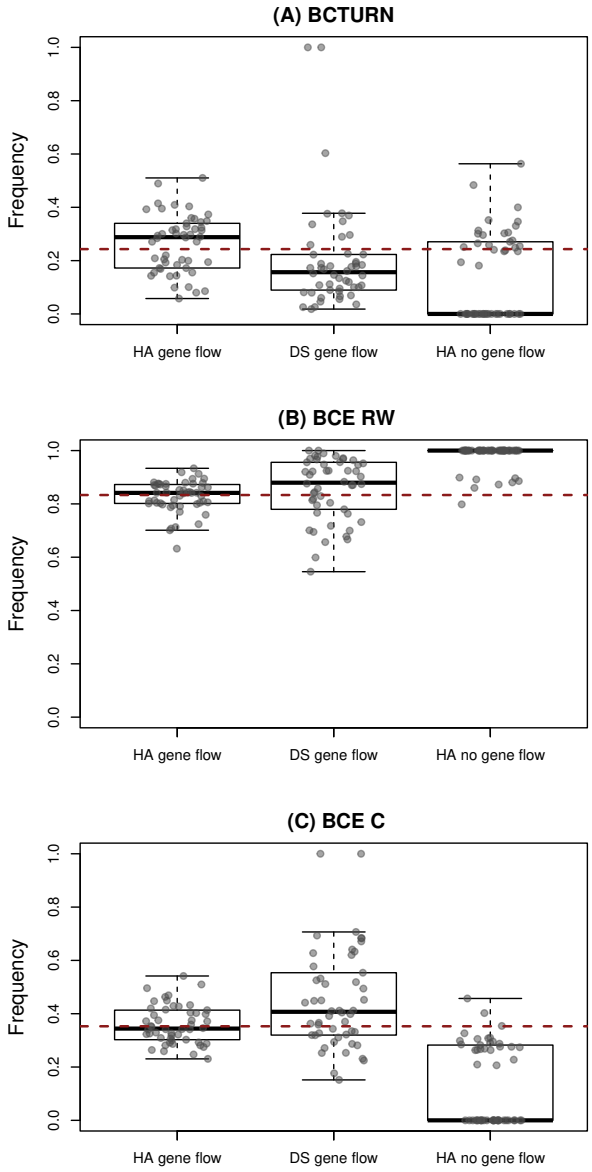

Figure S3: Boxplots show the RW (*Pg*) allele frequency from simulations for BCTURN (A), BCE RW (B) and BCE C (C) after 250,000 generations of evolution with balancing selection (BS) and gene flow, directional selection (DS) and gene flow, or BS without gene flow. Results are shown for 50 replicate simulations 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× the interquartile range. Points show the allele frequency for each replicate simulation. The observed RW allele frequency in each population is shown with a red, dashed line.

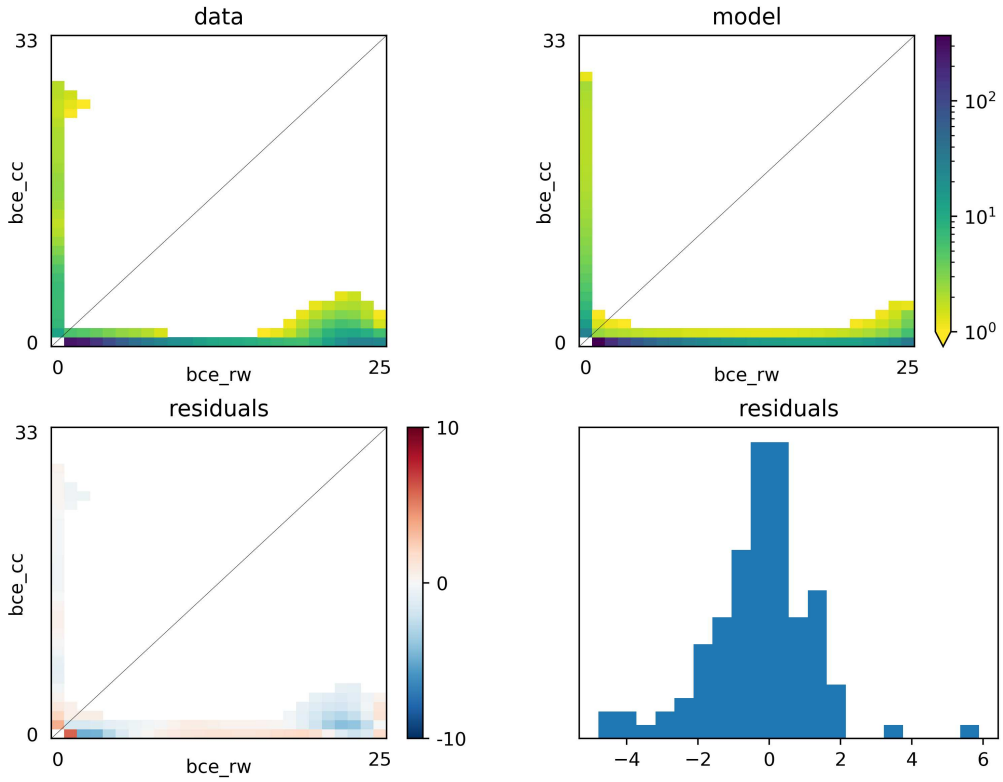

Figure S4: Summary of model fit for divergence time models in δaδi. The top panels show the observed (data) and predicted (model) joint site frequency spectra for *Perform* locus. The bottom panels show the corresponding residuals, that is the deviation between the observed and model-predicted joint site frequency spectra.

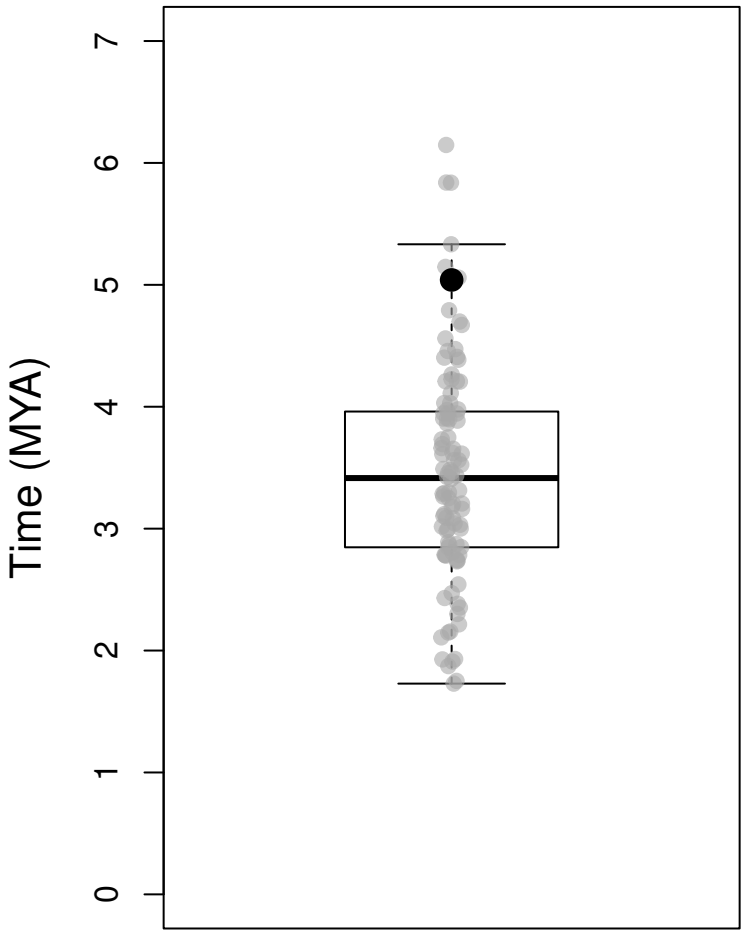

Figure S5: Inversion divergence time estimates from δaδi. The boxplot summarizes estimates of divergence time (in millions of years ago = MYA) for each of 100 block-jackknife replicates. Boxes denote the 1st and 3rd quartile, with the median given by the midline and whiskers extending to the minimum 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.