**Online Supplementary Materials**

**Updating the *T. cristinae* reference genome with additional linkage mapping.** We generated GBS data to improve (relative to1) the clustering of scaffolds into linkage groups (LGs) for the *T. cristinae* reference genome. In particular, we generated two additional lanes of Illumina hi-seq DNA sequence data with V3 reagents (128 million high-quality sequences) for individuals from three mapping families (192 individuals), which we combined with the data reported in1. We aligned these data to the reference genome scaffolds using BWA 0.7.5a-r4052 (the aln and samse algorithms) with a maximum number of allowed mismatches of 4, and a minimum base quality score of 10. We placed only sequences with a unique best hit. We then used the Bayesian variant caller in SAMTOOLS and BCFTOOLS 0.1.19 to identify variable nucleotides and to calculate genotype likelihoods based on the combined new and previous data1. We performed variant calling separately for each family with minimum base and alignment quality scores of 15 and 10, respectively, data required for 80% or more of the individuals in a family, and the full prior with θ set to 0.001. We called variants only if the posterior probability of the data under the null model of no variation was less than 0.001. We then estimated recombination rates between all pairs of SNPs within each family, as previously described1. We included data only from individuals and loci where the posterior probability of the most likely genotype was 0.95 (offspring posterior probabilities were calculated using the genotype likelihoods from BCFTOOLS and a prior based on the parental genotypes and Mendelian inheritance; see1 for details). We then constructed LGs from the recombination rate estimates using a heuristic clustering algorithm, as described in1. Overall the new LGs were similar to those generated previously. In particular, it was highly correlated between assemblies whether scaffolds were on the same or different LGs (r = 0.8). We treat the resulting new LG designation and ordering on the new draft genome (v0.3) as our best working hypothesis for the genome organization of *T. cristinae* that will iteratively be improved over time. The new draft genome (v0.3) includes 1413 scaffolds assigned to 13 LGs (551 Mb of the genome were assigned to linkage groups [53%]), with an average of 109 scaffolds per linkage group.

**Sampling survey of *Timema* populations across California to study stages of speciation.** In April and May 2012, we obtained *Timema* samples at 47 geographic sites across California, and stored them in 96% ethanol. Detailed information on samples is given in Table S1 and Figure S4 at the bottom of this document.

**Table S1.** Details about the populations of 12 Californian *Timema* (*T.*) species sampled. Number of individuals refer to the total number of individuals sampled and include those discarded for further analyses because of a low number of sequence reads.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Species** | **Locality** | **Latitude (N)** | **Longitude (W)** | **Host code** | **Host** | **No individuals** |
| *T. bartmani* | BMCG3 | 33.83 | -116.74 | IC | *Calocedrus decurrens* | 1 |
| *T. bartmani* | BMCG3 | 33.83 | -116.74 | WF | *Abies concolor* | 20 |
| *T. bartmani* | BMP90 | 33.80 | -116.70 | P | *Pinus sp.* | 40 |
| *T. bartmani* | BMP90 | 33.80 | -116.70 | WF | *Abies concolor* | 1 |
| *T. bartmani* | BMPCT | 33.84 | -116.74 | IC | *Calocedrus decurrens* | 1 |
| *T. bartmani* | BMPCT | 33.84 | -116.74 | WF | *Abies concolor* | 39 |
| *T. bartmani* | JL | 34.16 | -116.90 | P | *Pinus sp.* | 20 |
| *T. bartmani* | JL | 34.16 | -116.90 | WF | *Abies concolor* | 20 |
| *T. bartmani* | PCT8000ft | 33.83 | -116.72 | P | *Pinus sp.* | 15 |
| *T. bartmani* | PCTCR | 33.83 | -116.71 | P | *Pinus sp.* | 19 |
| *T. bartmani* | PCTCR | 33.83 | -116.71 | WF | *Abies concolor* | 19 |
| *T. boharti* | SRTH | 32.98 | -116.52 | C | *Ceanothus sp.* | 8 |
| *T. californicum* | LICK | 37.34 | -121.65 | Q | *Quercus sp.* | 20 |
| *T. californicum* | LP | 37.10 | -121.88 | Q | *Quercus sp.* | 20 |
| *T. californicum* | SM | 37.02 | -121.73 | M | *Arctostaphylos sp.* | 17 |
| *T. californicum* | SM | 37.02 | -121.73 | Q | *Quercus sp.* | 20 |
| *T. chumash* | BALD | 34.22 | -117.67 | C | *Ceanothus sp.* | 7 |
| *T. chumash* | BALD | 34.22 | -117.67 | MM | *Cercocarpus betuloides* | 4 |
| *T. chumash* | BALD | 34.22 | -117.67 | Q | *Quercus sp.* | 7 |
| *T. chumash* | BMT | 33.83 | -116.80 | C | *Ceanothus sp.* | 20 |
| *T. chumash* | BMT | 33.83 | -116.80 | Q | *Quercus sp.* | 15 |
| *T. chumash* | BS | 33.82 | -116.79 | C | *Ceanothus sp.* | 20 |
| *T. chumash* | BS | 33.82 | -116.79 | Q | *Quercus sp.* | 20 |
| *T. chumash* | DZ243 | 33.86 | -116.83 | M | *Arctostaphylos sp.* | 20 |
| *T. chumash* | GR104 | 34.23 | -117.68 | Q | *Quercus sp.* | 20 |
| *T. chumash* | GR603 | 34.22 | -117.74 | Q | *Quercus sp.* | 17 |
| *T. chumash* | GR806 | 34.22 | -117.71 | MM | *Cercocarpus betuloides* | 16 |
| *T. chumash* | GR806 | 34.22 | -117.71 | Q | *Quercus sp.* | 20 |
| *T. chumash* | HF4 | 34.27 | -118.10 | C | *Ceanothus sp.* | 1 |
| *T. chumash* | HF4 | 34.27 | -118.10 | Q | *Quercus sp.* | 7 |
| *T. chumash* | HF6 | 34.27 | -118.12 | Q | *Quercus sp.* | 5 |
| *T. chumash* | HFRBP | 34.26 | -118.11 | M | *Arctostaphylos sp.* | 21 |
| *T. chumash* | HFRBP | 34.26 | -118.11 | Q | *Quercus sp.* | 20 |
| *T. chumash* | HFRS | 34.36 | -118.01 | M | *Arctostaphylos sp.* | 20 |
| *T. chumash* | HFRS | 34.36 | -118.01 | MM | *Cercocarpus betuloides* | 20 |
| *T. chumash* | HFRS | 34.36 | -118.01 | Q | *Quercus sp.* | 20 |
| *T. chumash* | HFTP | 34.34 | -117.98 | C | *Ceanothus sp.* | 18 |
| *T. chumash* | PF243 | 33.86 | -116.84 | A | *Adenostoma fasciculatum* | 1 |
| *T. chumash* | PF243 | 33.86 | -116.84 | C | *Ceanothus sp.* | 20 |
| *T. chumash* | PF243 | 33.86 | -116.84 | M | *Arctostaphylos sp.* | 14 |
| *T. chumash* | PF243 | 33.86 | -116.84 | Q | *Quercus sp.* | 5 |
| *T. cristinae* | BY | 34.50 | -119.86 | A | *Adenostoma fasciculatum* | 20 |
| *T. cristinae* | BY | 34.50 | -119.86 | C | *Ceanothus sp.* | 20 |
| *T. cristinae* | BY | 34.50 | -119.86 | MM | *Cercocarpus betuloides* | 20 |
| *T. cristinae* | BY | 34.50 | -119.86 | Q | *Quercus sp.* | 10 |
| *T. cristinae* | ECCAMP | 34.51 | -119.76 | A | *Adenostoma fasciculatum* | 19 |
| *T. cristinae* | ECCAMP | 34.51 | -119.76 | M | *Arctostaphylos sp.* | 20 |
| *T. cristinae* | ECCAMP | 34.51 | -119.76 | Q | *Quercus sp.* | 20 |
| *T. cristinae* | OUT | 34.53 | -119.84 | A | *Adenostoma fasciculatum* | 3 |
| *T. cristinae* | OUT | 34.53 | -119.84 | C | *Ceanothus sp.* | 2 |
| *T. cristinae* | R23 | 34.52 | -120.08 | A | *Adenostoma fasciculatum* | 20 |
| *T. cristinae* | R9 | 34.51 | -120.07 | C | *Ceanothus sp.* | 7 |
| *T. cristinae* | VP | 34.53 | -119.85 | C | *Ceanothus sp.* | 20 |
| *T. cristinae* | VP | 34.53 | -119.85 | M | *Arctostaphylos sp.* | 4 |
| *T. cristinae* | VP | 34.53 | -119.85 | Q | *Quercus sp.* | 20 |
| *T.* sp. ‘cuesta ridge’ | CR | 35.36 | -120.65 | A | *Adenostoma fasciculatum* | 20 |
| *T.* sp. ‘cuesta ridge’ | CR | 35.36 | -120.65 | C | *Ceanothus sp.* | 20 |
| *T.* sp. ‘cuesta ridge’ | CR | 35.36 | -120.65 | CY | *Cupressus sargentii* | 20 |
| *T.* sp. ‘cuesta ridge’ | CR | 35.36 | -120.65 | M | *Arctostaphylos sp.* | 19 |
| *T.* sp. ‘cuesta ridge’ | CR | 35.36 | -120.65 | Q | *Quercus sp.* | 6 |
| *T. knulli* | BCE | 36.07 | -121.60 | RW | *Sequoia sempervirens* | 15 |
| *T. knulli* | BCTUR | 36.08 | -121.61 | C | *Ceanothus sp.* | 17 |
| *T. knulli* | BCTUR | 36.08 | -121.61 | P | *Pinus sp.* | 16 |
| *T. knulli* | BCWP | 36.07 | -121.60 | C | *Ceanothus sp.* | 12 |
| *T. knulli* | BCWP | 36.07 | -121.60 | Q | *Quercus sp.* | 1 |
| *T. knulli* | H1M37 | 36.17 | -121.68 | C | *Ceanothus sp.* | 4 |
| *T. knulli* | H1M37 | 36.17 | -121.68 | Q | *Quercus sp.* | 1 |
| *T. knulli* | HB | 36.16 | -121.67 | C | *Ceanothus sp.* | 20 |
| *T. knulli* | HB | 36.16 | -121.67 | Q | *Quercus sp.* | 3 |
| *T. landelsensis* | BCBOG | 36.07 | -121.58 | C | *Ceanothus sp.* | 23 |
| *T. landelsensis* | BCBOG | 36.07 | -121.58 | Q | *Quercus sp.* | 20 |
| *T. landelsensis* | BCHC | 36.06 | -121.57 | M | *Arctostaphylos sp.* | 3 |
| *T. landelsensis* | BCHC | 36.06 | -121.57 | Q | *Quercus sp.* | 20 |
| *T. landelsensis* | BCOG | 36.07 | -121.58 | C | *Ceanothus sp.* | 5 |
| *T. landelsensis* | BCOG | 36.07 | -121.58 | Q | *Quercus sp.* | 20 |
| *T. landelsensis* | BCSUM | 36.06 | -121.56 | C | *Ceanothus sp.* | 20 |
| *T. landelsensis* | BCSUM | 36.06 | -121.56 | M | *Arctostaphylos sp.* | 3 |
| *T. landelsensis* | BCSUM | 36.06 | -121.56 | Q | *Quercus sp.* | 11 |
| *T. petita* | 101SS | 35.73 | -121.31 | C | *Ceanothus sp.* | 20 |
| *T. podura* | BMCG3 | 33.83 | -116.74 | IC | *Calocedrus decurrens* | 20 |
| *T. podura* | BMCG3 | 33.83 | -116.74 | Q | *Quercus sp.* | 20 |
| *T. podura* | BME | 33.80 | -116.76 | A | *Adenostoma fasciculatum* | 20 |
| *T. podura* | BME | 33.80 | -116.76 | C | *Ceanothus sp.* | 4 |
| *T. podura* | BMLC | 33.81 | -116.75 | M | *Arctostaphylos sp.* | 1 |
| *T. podura* | BMLC | 33.81 | -116.75 | Q | *Quercus sp.* | 20 |
| *T. podura* | BMOKC | 33.82 | -116.75 | Q | *Quercus sp.* | 18 |
| *T. podura* | BMPCT | 33.84 | -116.74 | IC | *Calocedrus decurrens* | 19 |
| *T. podura* | BMPCT | 33.84 | -116.74 | M | *Arctostaphylos sp.* | 1 |
| *T. podura* | BMPCT | 33.84 | -116.74 | WF | *Abies concolor* | 1 |
| *T. podura* | BMT | 33.83 | -116.80 | C | *Ceanothus sp.* | 20 |
| *T. podura* | BMT | 33.83 | -116.80 | Q | *Quercus sp.* | 18 |
| *T. podura* | BS | 33.82 | -116.79 | C | *Ceanothus sp.* | 3 |
| *T. podura* | BS | 33.82 | -116.79 | Q | *Quercus sp.* | 3 |
| *T. podura* | DZ243 | 33.86 | -116.83 | A | *Adenostoma fasciculatum* | 20 |
| *T. podura* | DZ243 | 33.86 | -116.83 | M | *Arctostaphylos sp.* | 10 |
| *T. podura* | PCT8000ft | 33.83 | -116.72 | P | *Pinus sp.* | 5 |
| *T. podura* | PCTCR | 33.83 | -116.71 | P | *Pinus sp.* | 1 |
| *T. podura* | PF243 | 33.86 | -116.84 | A | *Adenostoma fasciculatum* | 20 |
| *T. podura* | PF243 | 33.86 | -116.84 | C | *Ceanothus sp.* | 8 |
| *T. podura* | PF243 | 33.86 | -116.84 | M | *Arctostaphylos sp.* | 5 |
| *T. podura* | PF243 | 33.86 | -116.84 | Q | *Quercus sp.* | 13 |
| *T. podura* | SRHWY | 32.82 | -116.51 | A | *Adenostoma fasciculatum* | 5 |
| *T. poppensis* | FROCK | 38.89 | -123.38 | C | *Ceanothus sp.* | 1 |
| *T. poppensis* | FROCK | 38.89 | -123.38 | DF | *Pseudotsuga menziesii* | 20 |
| *T. poppensis* | LP | 37.10 | -121.88 | DF | *Pseudotsuga menziesii* | 16 |
| *T. poppensis* | MM | 37.00 | -121.71 | RW | *Sequoia sempervirens* | 20 |
| *T. poppensis* | SM | 37.02 | -121.73 | RW | *Sequoia sempervirens* | 19 |
| *T. poppensis* | TBARN | 38.62 | -123.29 | DF | *Pseudotsuga menziesii* | 20 |
| *T. poppensis* | TBARN | 38.62 | -123.29 | RW | *Sequoia sempervirens* | 20 |
| *T. shepardi* | FROCK | 38.89 | -123.38 | C | *Ceanothus sp.* | 12 |

**Genotyping-by-sequencing (GBS) and stages of speciation.** We obtained 1,157,803,056 Illumina single-end 100 bp reads from 1545 individuals sequenced across seven sequencing lanes on the Hiseq2000 platform, which were parsed using custom Perl scripts based on code from a previous study3. We identified and removed the in-line barcodes, including those that were 1 bp away due to synthesizing or sequencing errors, and relabelled the sequences with the corresponding sample identifiers. In addition, we removed the following six base pairs of the EcoRI cut site and the adapters at the 3' end when present. We discarded sequences that were shorter than 16 bp after parsing or those lacking barcodes, as well as all reads of the asexual species *Timema shepardi* (n = 53,569,163). The total number of reads retained for the remaining 1533 individuals was 1,104,233,893, and the mean number of reads per individual was 720,309 (95% interval = 161,807-1,576,215). The average length of the sequences was 73 bp (95% interval = 67-78 bp). We aligned 71.6% of the reads (790,903,445) to the *T. cristinae* reference genome previously published1 using BOWTIE2 2.1.04 with the local model and the ‘--very-sensitive-local’ preset (-D 20 -R 3 -N 0 -L 20 -i S,1,0.50). The average number of mapped reads per individual was 512,235 (95% interval = 109,824 – 1,166,144).

Following mapping, we excluded from further analyses an additional 28 individuals that had fewer than 100,000 mapped reads. We used SAMTOOLS 0.1.195 to sort and index the alignments of the remaining 1505 individuals, which we used in further analyses. Variants were called using SAMTOOLS mpileup and BCFTOOLS using the full prior and requiring the probability of the data to be less than 0.5 under the null hypothesis that all samples were homozygous for the reference allele to call a variant. We ignored insertion and deletion polymorphisms. We identified 726,955 single nucleotide variants (SNVs) with an average depth across all individuals of ~4768x (mean coverage per variant per individual ~ 3×, median coverage ~ 1×). We applied further, more stringent filtering schemes specific to particular downstream analyses, which are described in the corresponding sections below.

We measured genome-wide genetic differentiation between pairs of populations using the Hudson’s *F*ST estimator6. For each population and variant, we inferred maximum-likelihood allele frequencies from the genotype likelihoods by means of the iterative soft expectation-maximization algorithm (EM), as before. We developed a Perl script to calculate Hudson’s *F*ST from bcf files. All the code is distributed under the GNU General Public License 3.0 and will be deposited in the Dryad repository. We estimated *F*ST for every pair of populations (defined as the pool of individuals from the same species, locality, and host plant), but excluded populations with less than two individuals beforehand. For each comparison, we further filtered out the SNVs that were present in less than 50% of the individuals from the two populations and had a pooled MAF estimate below 5%. This resulted in a variable number of variants used for *F*ST estimation in each comparison, ranging from 6404 to 62,603 (mean = 38,356, median = 40,313, 95% interval = 14,716 – 57,103). A table with details about all the comparisons and *F*ST estimates will be made available from the Dryad repository.

We also estimated genetic structure and potential admixture using a hierarchical Bayesian model that jointly estimates genotypes and admixture proportions as implemented in the program ENTROPY 1.2b7. This model is similar to the popular STRUCTURE8 algorithm, but accounts for sequencing errors and genotype uncertainties inherent to next-generation sequencing methods. We estimated parameters for a model with K=2 population clusters for every pair of populations found at the same geographic locality but belonging to different species, and for K=number-of-hosts-plants for conspecific populations found at the same geographic locality (Table S2). In addition, we fitted a model with K=1 in both cases and evaluated what model fit the data better using the difference in Deviance Information Criterion (ΔDIC, negative values indicate data favours K=1). DIC penalizes model complexity by adding to the posterior mean deviance the effective number of parameters (approximated as half the posterior variance of the deviance) 9. For each pair, we used the bi-allelic SNVs for which there were sequence data from at least 85% of the samples involved in each comparison and that were not fixed within any of the populations compared. We set the scalar of the Dirichlet initial value of q to 50. As starting admixture proportions, we used values obtained by applying linear discriminant analysis on a covariance matrix of composite genotypes estimated assuming Hardy-Weinberg equilibrium. We ran two independent Markov Chain Monte Carlo (MCMC) analyses for 35,000 generations and took samples every 10th iteration. We assessed mixing and convergence by visually inspecting the posterior deviance traces. We discarded the first 1,000 samples (10,000 iterations) from each chain as a burn-in, and combined the two chains (5,000 samples in total) to estimate model parameters. Details are provided in Table S2.

**Table S2.** Details on the number of individuals, single nucleotide variants (SNVs), admixture proportions, and difference in Deviance Information Criterion (ΔDIC) estimated with ENTROPY for conspecific populations on different hosts (among populations) and for different species ignoring hosts (among species), sampled from the same locality in both cases. Admixture proportions are given for a number of clusters K=number-of-hosts-plants for among populations comparisons (only species sampled from 2 or more host plants from the same locality are shown) and K=number of species for among species comparisons. Sample sizes and admixture proportions are showed in the same order than hosts and species. **Δ**DIC is the difference in DIC between K=1 and the K used for estimating admixture proportions (negative values indicate K=1 is a better fit). Host codes are A for chamise (*Adenostoma fasciculatum*), C for California lilac (*Ceanothus spinosus*), CY for Sargent’s cypress (*Cupressus sargentii*), DF for Douglas fir (*Pseudotsuga menziesii*), IC for incense cedar (*Calocedrus decurrens*), LP for lodgepole pine (*Pinus contorta*) , M for manzanita (*Arctostaphylos sp.*), MM for mountain mahogany (*Cercocarpus betuloides*), P for other pine (*Pinus sp.*), Q for oak (*Quercus sp.*), RW for redwood (*Sequoia sempervirens*), and WF for white fir (*Abies alba*).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Locality** | **Species** | **Hosts** | **Sample size** | **SNVs** | **Admixture** | **ΔDIC** |
|  |  |  |  |  |  |  |  |
| Among populations | BMCG3 | *T. bartmani* | IC,WF | 1,19 | 1138 | 0.52,0.51 | -832 |
| BMP90 | *T. bartmani* | LP,WF,P | 20,1,15 | 1103 | 0.33,0.21,0.23 | 651 |
| BMPCT | *T. bartmani* | IC,WF | 1,39 | 1354 | 0.70,0.36 | 700 |
| JL | *T. bartmani* | WF,P | 20,20 | 783 | 0.97,0.98 | -14616 |
| PCTCR | *T. bartmani* | WF,P | 19,18 | 1361 | 0.71,0.28 | -14939 |
| SM | *T. californicum* | M,Q | 17,19 | 1199 | 0.16,0.53 | -16836 |
| BALD | *T. chumash* | C,MM,Q | 7,4,7 | 1315 | 0.22,0.20,0.57 | -10777794 |
| BMT | *T. chumash* | C,Q | 20,15 | 678 | 0.99,0.93 | -8143 |
| BS | *T. chumash* | C,Q | 20,19 | 944 | 0.03,0.05 | -12213 |
| GR806 | *T. chumash* | MM,Q | 16,20 | 2462 | 0.83,0.64 | -12557 |
| HF4 | *T. chumash* | C,Q | 1,7 | 1107 | 0.17,0.22 | -26002 |
| HFRBP | *T. chumash* | M,Q | 21,20 | 1827 | 0.98,0.24 | -6890 |
| HFRS | *T. chumash* | M,MM,Q | 20,20,19 | 919 | 0.59,0.79,0.38 | -96398 |
| PF243 | *T. chumash* | A,C,M,Q | 1,19,13,5 | 693 | 0.40,0.34,0.30,0.47 | -13003 |
| BY | *T. cristinae* | A,C,MM,Q | 20,19,19,10 | 2739 | 0.00,0.00,0.00,0.20 | -59632 |
| OUT | *T. cristinae* | A,C | 3,2 | 805 | 0.50,0.50 | -4704 |
| VP | *T. cristinae* | C,M,Q | 20,4,20 | 1891 | 0.10,0.18,0.12 | -17356 |
| CR | *T. 'Cuesta Ridge'* | A,C,CY,M,Q | 20,18,20,19,6 | 2114 | 0.11,0.03,0.03,0.03,0.11 | -409688 |
| BCTUR | *T. knulli* | C,P | 17,16 | 1933 | 0.59,0.00 | 9784 |
| BCWP | *T. knulli* | C,Q | 10,1 | 1223 | 0.26,0.27 | -53569 |
| H1M37 | *T. knulli* | C,Q | 4,1 | 785 | 0.50,0.57 | -32921 |
| HB | *T. knulli* | C,Q | 20,3 | 1702 | 0.72,0.72 | -230375 |
| BCBOG | *T. landelsensis* | C,Q | 23,20 | 1913 | 0.88,0.71 | -12760 |
| BCHC | *T. landelsensis* | M,Q | 3,20 | 1814 | 0.45,0.62 | -15401597 |
| BCOG | *T. landelsensis* | C,Q | 4,18 | 1597 | 0.11,0.10 | -60971 |
| BCSUM | *T. landelsensis* | C,M,Q | 20,3,11 | 1657 | 0.54,0.02,0.54 | -37446 |
| BMCG3 | *T. podura* | IC,Q | 19,19 | 1950 | 0.63,0.52 | -10337 |
| BME | *T. podura* | A,C | 20,4 | 1903 | 0.95,0.24 | -4103 |
| BMLC | *T. podura* | M,Q | 1,20 | 1898 | 0.15,0.11 | -89279 |
| BMPCT | *T. podura* | IC,M,WF | 19,1,1 | 2225 | 0.32,0.00,0.00 | -12950 |
| BMT | *T. podura* | C,Q | 20,18 | 2573 | 0.49,0.36 | -2216 |
| BS | *T. podura* | C,Q | 3,3 | 1201 | 0.48,0.49 | -26228 |
| DZ243 | *T. podura* | A,M | 20,8 | 2196 | 0.69,0.57 | -15758 |
| PF243 | *T. podura* | A,C,M,Q | 20,8,5,13 | 2335 | 0.28,0.16,0.37,0.15 | -228622 |
| FROCK | *T. poppensis* | C,DF | 1,20 | 1467 | 0.01,0.13 | -164578 |
| TBARN | *T. poppensis* | DF,RW | 20,20 | 935 | 0.48,0.52 | 1083 |
|  | | | | | | | |
| Among species | LP | *T. californicum,*  *T. poppensis* | DF,Q | 20,16 | 2152 | 0.33,0.21,0.23 | 139240 |
| SM | *T. californicum,*  *T. poppensis* | M,Q,RW | 36,17 | 2243 | 1.000,0.000 | 151576 |
| BMT | *T. chumash,*  *T. podura* | C,Q | 35,38 | 4194 | 0.999,0.000 | 560870 |
| BS | *T. chumash,*  *T. podura* | C,Q | 39,6 | 3105 | 0.003,1.000 | 190033 |
| DZ243 | *T. chumash,*  *T. podura* | A,M | 20,28 | 4098 | 0.996,0.001 | 322521 |
| PF243 | *T. chumash,*  *T. podura* | A,C,M,Q | 38,46 | 4040 | 1.000,0.000 | 569373 |
| BMCG3 | *T. bartmani,*  *T. podura* | IC,Q,WF | 20,38 | 3108 | 0.000,1.000 | 160123 |
| BMPCT | *T. bartmani,*  *T. podura* | IC,M,WF | 40,21 | 3517 | 0.000,1.000 | 187304 |
| PCTCR | *T. bartmani,*  *T. podura* | WF,P | 37,1 | 2015 | 0.998,0.030 | -3928 |
|  |  |  |  |  |  |  |  |

**Maximum-likelihood phylogenetic inference and genealogical sorting index (GSI).** We removed, from the raw dataset of SNVs described above, variants with sequence data for less than 85% of the individuals, a depth greater than 10,000, a phred-scale quality score lower than 20, or more than two alleles. We kept 28,701 variants with a mean depth across all individuals of ~6930× (mean coverage per variant per individual ~ 4.5×). We used a custom Perl script to generate a multiple alignment that encoded heterozygous genotypes as IUPAC ambiguities. We partitioned the alignment by linkage group (LG) and excluded the positions in genomic regions not assigned to any linkage group. This resulted in a multiple alignment of a total of 1505 individuals and 19,556 positions distributed in 13 partitions as follows: LG1: 1185, LG2: 1642, LG3: 2097, LG4: 220, LG5: 1291, LG6: 622, LG7: 1367, LG8: 4033, LG9: 1273, LG10: 2866, LG11: 466, LG12: 2030, LG13: 464. We inferred 1000 maximum-likelihood bootstrap trees using the rapid heuristic algorithm implemented in RAxML 8.2.910,11. For each partition, we used a GTR substitution model, rate heterogeneity was incorporated using the CAT model with 25 categories, and likelihood was corrected for ascertainment bias using the Lewis approach12. Maximum-likelihood optimizations were started from random starting trees. We performed an *a posteriori* bootstrapping analysis using the extended majority-rule consensus (autoMRE) criterion with the recommended cutoff threshold of 0.03. This analysis indicated convergence after 500 bootstrap replicates. We used the R package ape13 to root bootstrap trees between the North Clade + Santa Barbara Clade and the South Clade (following root placement inferred in Bayesian phylogenetic analyses below). Subsequently, we used the R package genealogicalSorting to calculate, for each bootstrap tree, the Genealogical Sorting Index (GSI) for each of the 166 groups with at least 2 individuals delimited by species (11), species and locality (56), and species, locality, and host (98). Bootstrap trees and tables with all GSI values are deposited in Dryad.

**Genotyping-by-sequencing (GBS) for tests on the effect of colour-pattern and CHC on genome-wide differentiation.** We sequenced individuallybarcoded restriction-site associated DNA libraries of 325 samples from 19 *T. cristinae* populations on three Illumina lanes, using molecular and analytical methods described above (section GBS and stages of speciation). We combined these new sequences with 17 randomly chosen samples (10 males and 7 females) from the FHA mapping population, resulting in sequences from 342 individuals spanning 20 populations (5-20 individuals per population, mean = 17) for population genetic analyses of genetic differentiation. After filtering raw sequences (minimum read and base quality score 20, minimum read length 50 bp after trimming), we obtained 286,357,541 DNA sequences for all 342 samples (mean 837,303 reads per individual with mean read length 83.9 bp). We mapped 94.2% (269,643,356) reads to the *T. cristinae* draft genome v0.3a (90.9% of LG designation and ordering identical to new v0.3) using BOWTIE2 2.2.3 with the '--very-sensitive-local' preset. We used SAMTOOLS to sort and index alignments, and identified SNPs with SAMTOOLS mpileup and BCFTOOLS using the full prior and requiring the probability of the data being homozygous for the reference allele to be less than 0.01. We further discarded variants with low quality (score below 20) and where less than 90% of samples were covered. We retained 613,261 bi-allelic SNPs with mean coverage depth per SNP per individual ~5× (per SNP average ranging from 2.2 to 28.7; per individual average ranging from 1.0 to 10.3).

We estimated genome-wide Hudson's *F*ST for all 190 population pairs as *F*ST = 1 - Hw/Hb. Hw is the mean number of differences among sequences from the same population, and Hb the mean number of differences among sequences from different populations, averaged over loci. We calculated Hw and Hb for each locus from population allele frequencies estimated using genotype probabilities obtained with SAMTOOLS and BCFTOOLS, as in1. For each population pair, we excluded loci with a MAF less than 0.05, or where less than 50% of individuals were covered.

**Whole-genome re-sequencing of 10 population pairs spanning eight species.** We sequenced an additional 384 *Timema* genomes using the same protocols as for *T. cristinae*1. Of these, five were not appropriate for population genetic analyses, as they were single specimens each from a single locality. We did not analyse them further here following assembly and variant calling and will use them in future work. The other 379 genomes, which we do analyse here in a population genetic framework, stem from 16-20 individuals per population sampled from 10 parapatric pairs of host-associated populations. We sampled each of these pairs at the same general locality, usually directly adjacent to one another (the one exception was HFRS, where each population was sampled in the same general locality but separated by a slightly larger distance of ~3 km rather than the usual hundreds of metres). As with the genomes from the transplant experiment, we aligned the paired-end sequences to the *T. cristinae* reference genome (v0.3) using the BWA-MEM algorithm in BWA 0.7.5a-r4052. We used a minimum seed length of 20 bp, searched for internal seeds in seeds longer than 1.3 \* 20 base-pairs, discarded chains if the seeded bases were shorter than 100 bp, and set the minimum score to output an alignment to 30. We then used SAMTOOLS to compress, sort, and index the alignments and to remove potential PCR duplicates. We then identified variant nucleotides using the UnifiedGenotyper in GATK with the prior probability of heterozygosity set to 0.001, a minimum base quality score of 20, a call confidence threshold of 50, and a maximum of 2 alleles allowed. We considered only SNPs that mapped to one of the 13 identified LGs (i.e., due to our interest in genetic architecture we ignored the scaffolds not assigned to a LG). We further filtered the initial set of variants by retaining only those with (*i*) a minimum total sequencing depth of 384, (*ii*) a minimum of 10 reads supporting the non-reference allele, (*iii*) no more than 1% of reads spanning an insertion-deletion, (*iv*) no more than 5 mapping quality 0 reads, (*v*) a maximum absolute value of the base quality rank sum test of 3, (*vi*) a maximum absolute value of the mapping quality rank sum test of 2, (*vii*) a maximum absolute value of the read position rank sum test of 2, and (*viii*) a minimum ratio of the variant confidence score to the non-reference read depth of 2. We then discarded SNPs with a MAF less than 1% (across all individuals), which left us with 5.07 million SNPs for subsequent analyses.

**Population genetics using whole genomes from eight species.** We obtained maximum likelihood allele frequency estimates for each of the 20 populations (10 population pairs) for each of the 5.07 million SNPs identified above. We did this using an expectation-maximization (EM) algorithm that accounts for uncertainty in the underlying genotypes of individuals and that thus can work directly with the relative genotype likelihoods from GATK's UnifiedGenotyper. We implemented the previously described algorithm14 in a stand-alone C++ program written using the Gnu Scientific Library. We set the tolerance for EM convergence to 0.001 and the maximum number of EM iterations to 20. We then used these maximum likelihood allele frequency estimates to calculate sequence-based estimates of *F*ST between each of the 10 pairs of ecotypes, as described above. The set of 10-population pairs included four populations from each of two species (and two populations from the other six). Additionally, Nei’s measure of absolute divergence (*D*XY)15 was determined for each 20-kb window for the two hetero-specific population pairs (LP and SM).

We used Approximate Bayesian Computation (ABC) to estimate parameters of a Wright-Fisher model with migration to quantify gene flow (the number of migrants per generation or *Nem*) between the 10 pairs (Table S3, Figs. S1, S4). We inferred gene flow from random sets of 5000 SNPs with MAF >5% in order to obtain an estimate of the genome-average effective gene flow for each pair. We assumed discrete generations, where the pair of populations (with constant sizes *N0* and *N1*)diverged *t* generations in the past and has experienced constant, symmetric migration at rate *m* (where *m* is the proportion of migrant individuals). Ancestral allele frequencies are fixed at the mean for the pair of populations. We then used ABC to infer *t* and *Nem* (the product of the mean of *N0* and *N1,* and the migration rate). We used this approach rather than a coalescent-based ABC analysis or other methods based on diffusion approximations because it is a valid and efficient way to make inferences about recent evolutionary dynamics as in 16), and does not make the assumption that gene flow is a weak force in contrast to 17).

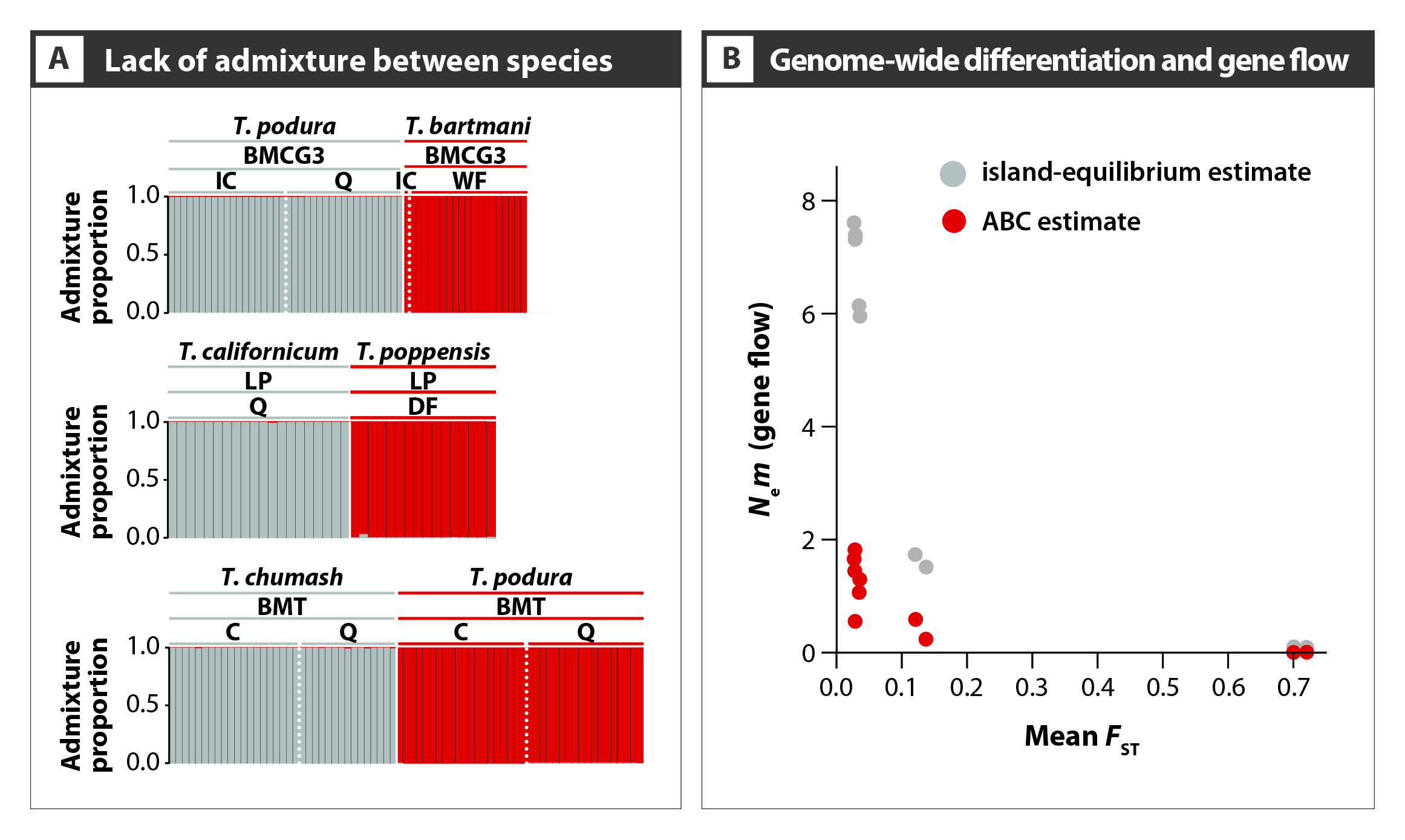
We placed log uniform priors with lower and upper bounds of 200 and 20,000 on the population sizes, and lower and upper bounds of 10 and 50,000 on the split time between taxon pairs. We placed a uniform prior bounded by 0 and 0.2 on the migration rate. We chose these priors to focus computational efforts on reasonable portions of parameter space. We selected summary statistics that are informative about divergence time and migration rate: a multilocus estimator of *F*ST, and the proportion of alleles that were rare or absent (MAF < 5%) in one population but not the other (MAF > 5%). We calculated the latter separately for each population in the pair. We then ran 500,000 ABC simulations from the Wright-Fisher model with parameter values sampled from their priors. We wrote the code to conduct the simulations and to calculate the summary statistics in C++ using the Gnu Scientific Library. We then used the rejection method with local-linear adjustment implemented in the R package abc to estimate the posterior distribution for each parameter18,19. We retained only the 0.1% of samples with summary statistics closest to the observed values, and we log- (all but *m*) or logit- (*m*) transformed the parameters during the inference procedure.

We detected non-negligible levels of gene flow between most taxon pairs (the primary exception involved the heterospecific species pair; range of *Nem* for conspecific pairs = 0.24 - 1.82, range of *Nem* for heterospecific pairs = 0.0067 - 0.0078). As expected, rates of gene flow estimated using ABC declined with average genome-wide *F*ST. We obtained similar results if we estimated *Nem* simply based on an equilibrium island model as *Nm* = - (*FST*-1)/(4*FST*))20.

**Table S3.** Population pairs used for whole genome re-sequencing and their characteristics. (A) Sample information: n1= sample size on host 1, n2 = sample size on host 2. (B) Gene flow estimates (*Nem* or number of migrants per generation) between 10 taxon pairs of *Timema*. Parameter estimates from the Approximate Bayesian Computation (ABC) approach and an equilibrium island model (right-hand column) are given. (C) Summary of patterns of genetic differentiation for the 10 population pairs from Hidden Markov Models (HMMs).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **(A) Pair** | **Locality** | **Species** | **Host 1** | **Host 2** | | | **n1/n2** | | **Latitude**  **(N)** | | | **Longitude (W)** | | |
| 1 | BCBOG | *T. landelsensis* | *Ceanothus* | *Quercus* | | | 19/20 | | 36.0660 | | | -121.5806 | | |
| 2 | BCTURN | *T. knulli* | *Ceanothus* | *Pinus* | | | 16/16 | | 36.0762 | | | -121.6064 | | |
| 3 | BMCG3 | *T. podura* | *Quercus* | *Calocedrus* | | | 20/20 | | 33.8313 | | | -116.7412 | | |
| 4 | BMT | *T. podura* | *Ceanothus* | *Quercus* | | | 20/18 | | 33.8273 | | | -116.7955 | | |
| 5 | BS | *T. chumash* | *Ceanothus* | *Quercus* | | | 20/20 | | 33.8164 | | | -116.7902 | | |
| 6 | CR | *T. sp.* ‘cuesta ridge’ | *Ceanothus* | *Cupressus* | | | 20/20 | | 35.3562 | | | -120.6543 | | |
| 7 | HFRS | *T. chumash* | *Ceanothus* | *Quercus* | | | 17/18 | | 34.3556 | | | -118.0120 | | |
| 8 | LP | *T. californicum / T. poppensis* | *Quercus* | *Pseudotsuga* | | | 20/16 | | 37.1019 | | | -121.8756 | | |
| 9 | SM | *T. californicum / T. poppensis* | *Quercus* | *Sequoia* | | | 20/19 | | 37.0188 | | | -121.7256 | | |
| 10 | VP | *T. cristinae* | *Ceanothus* | *Quercus* | | | 20/20 | | 34.5325 | | | -119.8467 | | |
|  | | | | | | | | | | | | | | |
| **(B)** | **Locality** | **Posterior median of *Nem*** | **Lower bound**  **90 % CI** | | **Upper bound**  **90 % CI** | | | **Equilibrium**  **estimate** | | |  | | | |
| 1 | BCBOG | 1.3037 | 0.4112 | | 3.1517 | | | 5.9647 | | |
| 2 | BCTUR | 0.5919 | 0.2450 | | 3.0173 | | | 1.7486 | | |
| 3 | BMCG3 | 1.4501 | 0.2338 | | 3.7673 | | | 7.3929 | | |
| 4 | BMT | 1.8287 | 0.5793 | | 4.4588 | | | 7.3281 | | |
| 5 | BS | 0.5668 | 0.0620 | | 7.8986 | | | 7.4081 | | |
| 6 | CR | 1.0743 | 0.2888 | | 2.7388 | | | 6.1451 | | |
| 7 | HFRS | 0.2438 | 0.0107 | | 19.0472 | | | 1.5236 | | |
| 8 | LP | 0.0078 | 0.0021 | | 0.0380 | | | 0.0972 | | |
| 9 | SM | 0.0067 | 0.0003 | | 0.1233 | | | 0.1072 | | |
| 10 | VP | 1.6659 | 0.6283 | | 3.8307 | | | 7.6160 | | |
|  | | | | | | | | | | | | | |
| **(C)** | **Locality** | **Number of regions of accentuated *F*ST** | **Mean size of**  **regions of accentuated**  ***F*ST (# 20-kb windows)** | | | **Mean *F*ST of regions of accentuated *F*ST** | | | | **Mean background *F*ST** | | |
| 1 | BCBOG | 2 | 980 | | | 0.056 | | | | 0.039 | | |
| 2 | BCTUR | 0 | N/A | | | N/A | | | | 0.125 | | |
| 3 | BMCG3 | 3 | 690 | | | 0.040 | | | | 0.032 | | |
| 4 | BMT | 0 | N/A | | | NA | | | | 0.033 | | |
| 5 | BS | 1 | 241 | | | 0.045 | | | | 0.033 | | |
| 6 | CR | 2 | 262 | | | 0.080 | | | | 0.038 | | |
| 7 | HFRS | 0 | N/A | | | N/A | | | | 0.141 | | |
| 8 | LP | 5 | 408 | | | 0.789 | | | | 0.714 | | |
| 9 | SM | 10 | 242 | | | 0.784 | | | | 0.691 | | |
| 10 | VP | 0 | N/A | | | N/A | | | | 0.032 | | |

**Figure S1. Genetic structure and gene flow.** (A) Lack of admixture between species in analyses of genetic structure (codes below species names are for locality name and host). (B) Mean genetic differentiation (*F*ST) between conspecific ecotypes and species studied using whole-genome re-sequencing (these are a subset of those studied using genotyping-by-sequencing data). Gene flow estimates are shown from Approximate Bayesian Computation and from island-equilibrium estimates.



**Table S4**.**Results of the Hidden Markov Model (HMM) analyses of genetic differentiation between 14 pairs of *Timema* taxa.** These are the same taxa depicted in Figure 2 of the main text.LG = linkage group. Numbers in parentheses following taxon pair codes represent the summed number of regions of accentuated *F*ST across LGs. Values in the body of the table refer to the number of regions of accentuated differentiation per LG, followed in italics by the proportion of the LG involved (i.e., for LG with at least one such region of accentuated differentiation).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Taxon pair** | **LG1** | **LG2** | **LG3** | **LG4** | **LG5** | **LG6** | **LG7** |
| HV (0) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MR1 (0) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| R12 (3) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| LAPRC (2) | 0 | 1, *0.055* | 0 | 0 | 0 | 0 | 0 |
| VP (0) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| BS (1) | 0 | 0 | 1, *0.047* | 0 | 0 | 0 | 0 |
| BMCG3 (3) | 0 | 0 | 3, *0.407* | 0 | 0 | 0 | 0 |
| BMT (0) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| BCBOG (3) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| BCTUR (0) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CR (2) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| HFRS (0) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| LP (5) | 0 | 0 | 1, *0.057* | 0 | 1, *0.107* | 0 | 0 |
| SM (10) | 0 | 0 | 1, *0.053* | 4, *0.338* | 1, *0.071* | 0 | 1, *0.084* |
|  |  |  |  |  |  |  |  |
|  | **LG8** | **LG9** | **LG10** | **LG11** | **LG12** | **LG13** |  |
| HV (0) | 0 | 0 | 0 | 0 | 0 | 0 |  |
| MR1 (0) | 0 | 0 | 0 | 0 | 0 | 0 |  |
| R12 (3) | 3, *0.590* | 0 | 0 | 0 | 0 | 0 |  |
| LAPRC (2) | 1, *0.226* | 0 | 0 | 0 | 0 | 0 |  |
| VP (0) | 0 | 0 | 0 | 0 | 0 | 0 |  |
| BS (1) | 0 | 0 | 0 | 0 | 0 | 0 |  |
| BMCG3 (3) | 0 | 0 | 0 | 0 | 0 | 0 |  |
| BMT (0) | 0 | 0 | 0 | 0 | 0 | 0 |  |
| BCBOG (3) | 2, *0.653* | 1, *0.357* | 0 | 0 | 0 | 0 |  |
| BCTUR (0) | 0 | 0 | 0 | 0 | 0 | 0 |  |
| CR (2) | 2, *0.247* | 0 | 0 | 0 | 0 | 0 |  |
| HFRS (0) | 0 | 0 | 0 | 0 | 0 | 0 |  |
| LP (6) | 1, *0.117* | 0 | 1, *0.260* | 2, *0.584* | 0 | 0 |  |
| SM (10) | 1, *0.022* | 0 | 0 | 2, *0.415* | 0 | 0 |  |

**Genetic differentiation and allele frequencies.** We tested whether HMM regions of accentuated differentiation (*F*ST) had MAF that differed from the genomic background (*i.e.*, for the eight of 14 taxon pairs with at least one such region). The approach was as follows: for each population, we inferred maximum-likelihood site MAFs from the genotype likelihoods by means of the iterative soft expectation-maximization algorithm (EM), as before. We then evaluated whether MAFs for the sets of HMM regions of accentuated differentiation were different from MAFs of similar sets of random genomic regions. For each population (defined by the host plant, “H. 1” and “H. 2” in Table S5), we calculated an observed multi-region MAF across the HMM regions of accentuated differentiation as the mean MAF of all SNPs with a MAF greater than 1% within such regions (“Obs. 1” and “Obs. 2” in Table S5, one observed value for each population in the pair). We obtained the empirical null distribution of MAFs for random genomic regions by means of a randomisation procedure. For each population pair, we randomly sampled without replacement the same number of genomic regions of the same size as the corresponding HMM regions of accentuated differentiation. Subsequently, we calculated multi-region mean MAFs for each population, as before. We repeated this procedure 1000 times in order to obtain the null distribution of 1000 multi-region MAFs (“Null expectation 1” and “Null expectation 2” in Table S5; mean and 95% confidence intervals are shown). We assessed significance by computing empirical cumulative distributions and calculating two-tail *P*-values (“*P* 1” and “*P* 2” in Table S5).

We found that MAFs in the HMM regions of accentuated differentiation sometimes differed from the genomic background, but not in a consistent way. When we detected differences, they were weak in magnitude (on the order of ~1%) and varied in sign among taxon pairs (i.e., sometimes being higher and sometimes lower than the genomic background). Thus, we did not observe a strong and consistent overall association between allele frequency and *F*ST, as may be expected when *F*ST is estimated for larger windows containing many SNPs, rather than for individual SNPs.

**Table S5.** Minor allele frequency (MAF) of HMM regions of accentuated differentiation compared to null genomic background expectations. H = host plant. NA = not applicable (*i.e.*, population pairs that did not have at least one region of accentuated differentiation in the HMM analyses). Obs. = empirically observed MAFs for each population in a population pair). *P* 1and *P* 2 are the significance values for each population.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Populations** | | |  | **Null expectation 1** | |  |  | **Null expectation 2** | |  |
| **Locality** | **H 1** | **H 2** | **Obs. 1** | **mean** | **95% CI** | ***P* 1** | **Obs. 2** | **mean** | **95% CI** | ***P* 2** |
| HV | A | C | NA | NA | NA | NA | NA | NA | NA | NA |
| MR1 | A | C | NA | NA | NA | NA | NA | NA | NA | NA |
| R12 | A | C | 0.159 | 0.146 | 0.138-0.159 | 0.042 | 0.171 | 0.154 | 0.145-0.171 | 0.058 |
| LA/PRC | A | C | 0.178 | 0.148 | 0.140-0.164 | 0.014 | 0.177 | 0.163 | 0.155-0.174 | 0.004 |
| BCBOG | C | Q | 0.197 | 0.222 | 0.202-0.229 | 0.006 | 0.263 | 0.250 | 0.242-0.261 | 0.014 |
| BCTUR | C | P | NA | NA | NA | NA | NA | NA | NA | NA |
| BMCG3 | IC | Q | 0.196 | 0.215 | 0.205-0.230 | 0.000 | 0.215 | 0.225 | 0.219-0.238 | 0.000 |
| BMT | C | Q | NA | NA | NA | NA | NA | NA | NA | NA |
| BS | C | Q | 0.155 | 0.172 | 0.144-0.203 | 0.280 | 0.218 | 0.230 | 0.202-0.267 | 0.498 |
| CR | C | CY | 0.225 | 0.223 | 0.212-0.236 | 0.682 | 0.224 | 0.225 | 0.214-0.236 | 0.928 |
| HFRS | M | Q | NA | NA | NA | NA | NA | NA | NA | NA |
| LP | DF | Q | 0.214 | 0.213 | 0.186-0.233 | 0.962 | 0.232 | 0.235 | 0.227-0.247 | 0.554 |
| SM | Q | RW | 0.184 | 0.208 | 0.186-0.251 | 0.010 | 0.223 | 0.230 | 0.223-0.237 | 0.058 |
| VP | C | Q | NA | NA | NA | NA | NA | NA | NA | NA |

**Quantifying colour-pattern and CHCs** We recorded digital images of 873 adult *T. cristinae* using previously described methods21 (Table S6, Fig. S4 for map of localities).

**Table S6.** Identity, locality, and sample sizes of populations and species used to study phenotypic variation in cuticular hydrocarbons (CHCs) and colour pattern (% body area striped). Note that we used some datasets in several analyses such that sample sizes are not unique to just one analysis. In addition to the numbers in the table, the perfuming experiment included 96 insects from the FHA population (24 males, 72 females), 24 females from the SMRW population, and 24 females from the SMHCRW population. Abbreviations for host plant (*A* = *Adenostoma*, *C* = *Ceanothus*, *S* = *Sequoia*).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Population code** | **Latitude (N)** | **Longitude (W)** | **Host** | **CHC variation**  ***N* (males, females)** | **% body area striped**  ***N* (males, females)** |
| *T. cristinae* |  |  |  |  |  |
| FHA | 34.517644 | -119.800989 | *A* | 20 (10, 10) | 20 (10, 10) |
| ECC20A | 34.504972 | -119.73285 | *A* | 18 (10, 8) | 18 (10, 8) |
| PC | 34.476789 | -119.768839 | *C* | 20 (10, 10) | 20 (10, 10) |
| MH19.78C | 34.519144 | -119.270992 | *C* | 10 (5, 5) | 10 (5, 5) |
| MH25.59C | 34.533242 | -119.243072 | *C* | 20 (10, 10) | 20 (10, 10) |
| R12C | 34.515031 | -119.071031 | *C* | 20 (10, 10) | 20 (10, 10) |
| R23A | 34.519111 | -119.077511 | *A* | 20 (10, 10) | - |
| HVA | 34.488586 | -119.785839 | *A* | 20 (10, 10) | 20 (10, 10) |
| LA | 34.512586 | -119.796203 | *A* | 20 (10, 10) | - |
| PRC | 34.533308 | -119.857644 | *C* | 20 (10, 10) | 20 (10, 10) |
| OGC | 34.513442 | -119.796086 | *C* | 15 (8, 7) | 15 (8, 7) |
| NS1A | 34.488361 | -119.654611 | *A* | 19 (9, 10) | 18 (9, 9) |
| MA | 34.515103 | -119.797133 | *A* | 16 (10, 6) | 16 (10, 6) |
| ECCCampA | 34.506411 | -119.761644 | *A* | 9 (5, 4) | 9 (5, 4) |
| OGA | 34.513406 | -119.796322 | *A* | 18 (8, 10) | 17 (8, 9) |
| ECC35A | 34.5062 | -119.768136 | *A* | 18 (10, 8) | 18 (10, 8) |
| OUTA | 34.531683 | -119.843517 | *A* | 16 (9, 7) | 16 (9, 7) |
| BYA | 34.5006 | -119.86195 | *A* | 20 (10, 10) | 20 (10, 10) |
| SC | 34.5226 | -119.83175 | *C* | 19 (10, 9) | 19 (10, 9) |
| MH29.19C | 34.555367 | -119.263167 | *C* | 5 (0, 5) | 5 (0, 5) |
|  |  |  |  |  |  |
| *T. poppensis* |  |  |  |  |  |
| SMRW | 37.01876 | -121.72556 | *S* | - | - |
| SMHCRW | 37.01074 | -121.71508 | *S* | - | - |

For CHC variation, we sampled 20 different populations of *Timema cristinae* (eight on *Ceanothus* and 12 on *Adenostoma*) for a total of 915 individuals (559 males and 356 females), and two populations of *Timema* *poppensis* (48 females).To extract CHCs from the body surface of individual insects, we euthanized insects by 1-h freezing, and then submerged each insect for 10 minutes in 1 ml of HPLC-grade hexane in separate vials. Subsequently, we removed the insect from each vial, concentrated the sample to dryness by hexane evaporation at room temperature, and re-constituted the CHC extract by adding 100 µl of hexane containing (*E*)-9-octadecenyl acetate as an internal standard (IS). We then analysed an aliquot of each sample on a 6890 Hewlett Packard (now Agilent) gas chromatograph (GC) equipped with a DB-5 MS column (50 m × 0.25 mm i.d.), using the following temperature program: 100 °C for 1 min, then 20 °C per min to 280 °C. The final temperature of 280 °C was held for 40 min. Temperatures of the GC injector and the flame ionization detector (FID) were set to 280 °C.

In total, we quantified 26 different mono- and di-methylated CHCs for each individual *T*. *cristinae*. Specifically, we quantified eight methylated pentacosanes, eight methylated heptacosanes (including the six monomethyl heptacosanes previously described22), and 10 methylated nonacosanes. As is standard practice in studies of CHC variation22, we analysed proportional rather than absolute abundances of CHCs; this allowed us to reduce experimental error and to remove individual differences in CHCs stemming from insect body size variation 23,24. We determined the total amount of each target CHC by multiplying the area count of the respective FID peak with 200 ng of the IS and by dividing the product by the FID area count of the IS. We calculated proportional CHCs by dividing the amount of each CHC in a given sample by the sum of all CHCs in that sample. We then transformed these CHC proportions using log-contrasts23,25 to remove the non-independence among analysed variables. We calculated log-contrasts by dividing the value for each CHC by the value of the CHC 5-methylheptacosane (5Me27), and then taking the log10 of these new variables, resulting in 25 log-contrast transformed values for every insect. We obtained similar results when we divided the value for each CHC by the value of a CHC other than 5Me27.

**Repeatability of CHC measurements.** To test the repeatability of our phenotypic measures (i.e., proportional CHCs), we randomly chose hexane extracts of six males and six females each from two different populations (FHA and MH25.59C), and analysed them once each on two consecutive days, using the protocols described above. We again calculated log-contrasts for proportions of all 25 CHCs (contrasting against the 26th CHC, 5Me27) and calculated intra-class correlation coefficients (ICC, *n* = 24)26 in IBM SPSS Statistics 21 (IBM Corporation). ICC analyses revealed very high repeatability for every single CHC (ICC, *r* ≥ 0.859 in all cases, as follows for each compound: LogC\_C25-1, 0.964; LogC\_C25-2, 0.891; LogC\_C25-3, 0.972; LogC\_C25-4, 0.998; LogC\_C25-5, 0.989; LogC\_C25-6, 0.973; LogC\_C25-7, 0.976; LogC\_C25-8, 0.933; LogC\_C27-1, 0.986; LogC\_C27-2, 0.947; LogC\_C27-4, 0.955; LogC\_C27-5, 0.859; LogC\_C27-6, 0.933; LogC\_C27-7, 0.987; LogC\_C27-8, 0.968; LogC\_C29-1, 0.929; LogC\_C29-2, 0.952; LogC\_C29-3, 0.940; LogC\_C29-4, 0.998; LogC\_C29-5, 0.965; LogC\_C29-6, 0.950; LogC\_C29-7, 0.931; LogC\_C29-8, 0.874; LogC\_C29-9, 0.980; LogC\_C29-10, 0.919).

**Differences between ecotypes in CHCs.** Weconducted this analysis on CHC extracts of 343 *T. cristinae* from 20 different populations sampled in 2013 (174 males and 169 females from 12 populations adapted to *Adenostoma* and eight populations adapted to *Ceanothus*). Because we measured many more individuals in the FHA population than in other populations to accomplish GWA mapping, we randomly chose 10 males and 10 females from FHA for this analysis. Across all populations, we detected five samples that were extreme multivariate CHC outliers based on Mahalanobis distance as calculated in the SPSS ‘Regression’ procedure, and we thus removed them from subsequent analyses (i.e., one male from OGC, one female each from ECC35A and ECCCampA, and two females from ECC20A).

To reduce data dimensionality and to account for multicollinearity, we conducted a principal components analysis (on a covariance matrix with promax rotation) on the remaining 338 samples (172 males and 166 females). We retained principal component (PC) axes with an eigenvalue larger than the mean eigenvalue as variables for subsequent analyses (resulting in six axes retained, which accounted for 89.5% of the total variation; Table S7). We then conducted multivariate analysis of variance (MANOVA) on these six PCs as our primary test of phenotypic differences between sexes and ecotypes, by testing for effects due to ‘sex’, ‘host plant’, and the interaction of ‘sex-by-host plant’.

**Table S7.** Loadings for principal components (PC) analyses carried out on cuticular hydrocarbon (CHC) profiles of male and female *T. cristinae* from 20 different populations (host plants: *Adenostoma, N* = 12; *Ceanothus*, *N* = 8).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **PC1** | **PC2** | **PC3** | **PC4** | **PC5** | **PC6** |
| **Eigenvalues** | 13.478 | 4.579 | 3.791 | 2.091 | 1.955 | 1.246 |
| **% Variance** | 44.555 | 15.102 | 12.503 | 6.897 | 6.449 | 4.110 |
| **CHC compound** |  |  |  |  |  |  |
| LogC\_C25-1 | -0.001 | 0.719 | -0.083 | 0.091 | 0.059 | 0.017 |
| LogC\_C25-2 | 0.009 | 1.038 | 0.022 | -0.197 | -0.024 | 0.068 |
| LogC\_C25-3 | -0.044 | 0.580 | 0.016 | 0.027 | 0.072 | -0.066 |
| LogC\_C25-4 | 0.041 | 0.101 | 0.070 | 0.915 | 0.041 | 0.018 |
| LogC\_C25-5 | -0.142 | 0.103 | -0.097 | 0.042 | 0.925 | 0.049 |
| LogC\_C25-6 | 0.188 | 0.617 | 0.060 | 0.129 | -0.118 | -0.091 |
| LogC\_C25-7 | 0.000 | 0.731 | -0.012 | -0.041 | 0.156 | 0.040 |
| LogC\_C25-8 | -0.237 | 0.132 | 0.945 | 0.043 | -0.079 | 0.144 |
| LogC\_C27-1 | 0.388 | 0.491 | -0.041 | -0.008 | 0.050 | -0.108 |
| LogC\_C27-2 | 0.187 | 0.504 | -0.025 | 0.126 | -0.007 | -0.088 |
| LogC\_C27-4 | 0.524 | 0.297 | 0.064 | 0.081 | -0.082 | 0.143 |
| LogC\_C27-5 | 0.509 | 0.218 | 0.003 | 0.041 | -0.030 | 0.079 |
| LogC\_C27-6 | -0.073 | 0.281 | 0.140 | 0.025 | 0.028 | -0.115 |
| LogC\_C27-7 | 0.550 | 0.183 | -0.115 | 0.010 | -0.011 | -0.073 |
| LogC\_C27-8 | -0.041 | 0.221 | 0.329 | -0.034 | 0.000 | -0.110 |
| LogC\_C29-1 | 0.775 | -0.067 | -0.072 | -0.068 | 0.009 | 0.016 |
| LogC\_C29-2 | 0.580 | -0.002 | -0.227 | -0.046 | -0.036 | 0.076 |
| LogC\_C29-3 | 0.575 | 0.076 | -0.163 | 0.016 | -0.062 | 0.081 |
| LogC\_C29-4 | 0.235 | -0.177 | -0.001 | 0.017 | 0.051 | 0.853 |
| LogC\_C29-5 | 0.703 | -0.081 | -0.083 | 0.044 | 0.002 | 0.013 |
| LogC\_C29-6 | 0.666 | 0.019 | -0.065 | 0.042 | -0.064 | 0.021 |
| LogC\_C29-7 | 0.094 | 0.176 | -0.264 | -0.048 | 0.034 | 0.086 |
| LogC\_C29-8 | -0.151 | 0.466 | -0.022 | -0.037 | -0.089 | 0.008 |
| LogC\_C29-9 | 0.551 | -0.130 | -0.122 | -0.016 | -0.018 | 0.025 |
| LogC\_C29-10 | 0.867 | -0.197 | 0.479 | -0.083 | 0.126 | -0.086 |

**Genomic data from population FHA for GWA mapping.** We obtained genotypes for mapping from publicly available sequence data for 602 *T. cristinae* individualsfrom the FHA population (NCBI BioProject PRJNA284835)21. From the 524,832 SNPs obtained in that previous study, we created subsets for the 592 individuals for which CHC data were available, as well as for males and females separately. We discarded variants with a MAF below 1% in each subset. The resulting datasets comprised 246,258 variants (all 592 individuals), 246,293 variants for females (197 individuals), and 245,778 for males (395 individuals). As in past work1, we used a custom Perl script to calculate empirical Bayesian posterior probabilities for the genotypes of each individual and locus using the genotype likelihoods and allele frequencies estimated by BCFTOOLS along with Hardy-Weinberg priors (i.e. *p*(AA) = *pi2*; *p*(aa) = (1-*pi*)2; *p*(Aa) = 2*pi*(1-*pi*); ‘A’ is the major allele, ‘a’ is the minor allele, and ‘*p*’is the major allele frequency). Finally, we calculated the posterior mean genotype for each individual, at each locus, defined as the minor allele dosage (i.e., *g-hatij* as Σ*k* = {0, 1, 2}*k* \* Pr(*gij* = *k*| data, *pi*), where *gij* is the genotype for locus *i* and individual *j,* and *k* are the variants). We used these imputed genotypes for all GWA mapping analyses.

**Genome-wide association (GWA) mapping and cross-validation.** We used the software GEMMA 0.94 for GWA mapping27. We used GEMMA to implement Bayesian sparse linear mixed models (BSLMMs) using a multiple-SNP Bayesian approach to model the genetic architecture of traits while considering relatedness of individuals. In BSLMMs implemented in GEMMA the effects of SNPs are modelled as coming from a mixture of two normal distributions. Thus, effects of SNPs that individually have infinitesimal effects (‘polygenic distribution’) and SNPs with measurable (i.e., ‘larger’ or ‘sparse’) effects can be estimated. GEMMA also provides posterior inclusion probabilities (PIPs, also called γ parameter) that reflect the weight of evidence that individual SNPs are associated with the trait of interest.

We estimated the above-mentioned hyper-parameters and PIP values for the following seven traits: % striped, the proportion of methylated pentacosanes, heptacosanes, and nonacosanes in females (fpenta, fhepta, and fnona, respectively), and the proportion of methylated pentacosanes, heptacosanes, and nonacosanes in males (mpenta, mhepta, and mnona, respectively). We treated sexes separately for CHCs due to strong sexual dimorphism in CHCs. For % striped, we ran GEMMA on residuals that corrected for differences between sexes by regressing each trait against sex. We report for each trait the point estimates (median) and 95% equal-tail probability intervals (ETPIs) of hyper-parameters, calculated across 10 independent MCMC runs per trait. For each chain, we ran 20,000,000 iterations with a recording pace of one record state in every 100 steps and discarded the first 5,000,000 iterations as burn-in. We excluded SNPs with a MAF less than, or equal to, one percent.

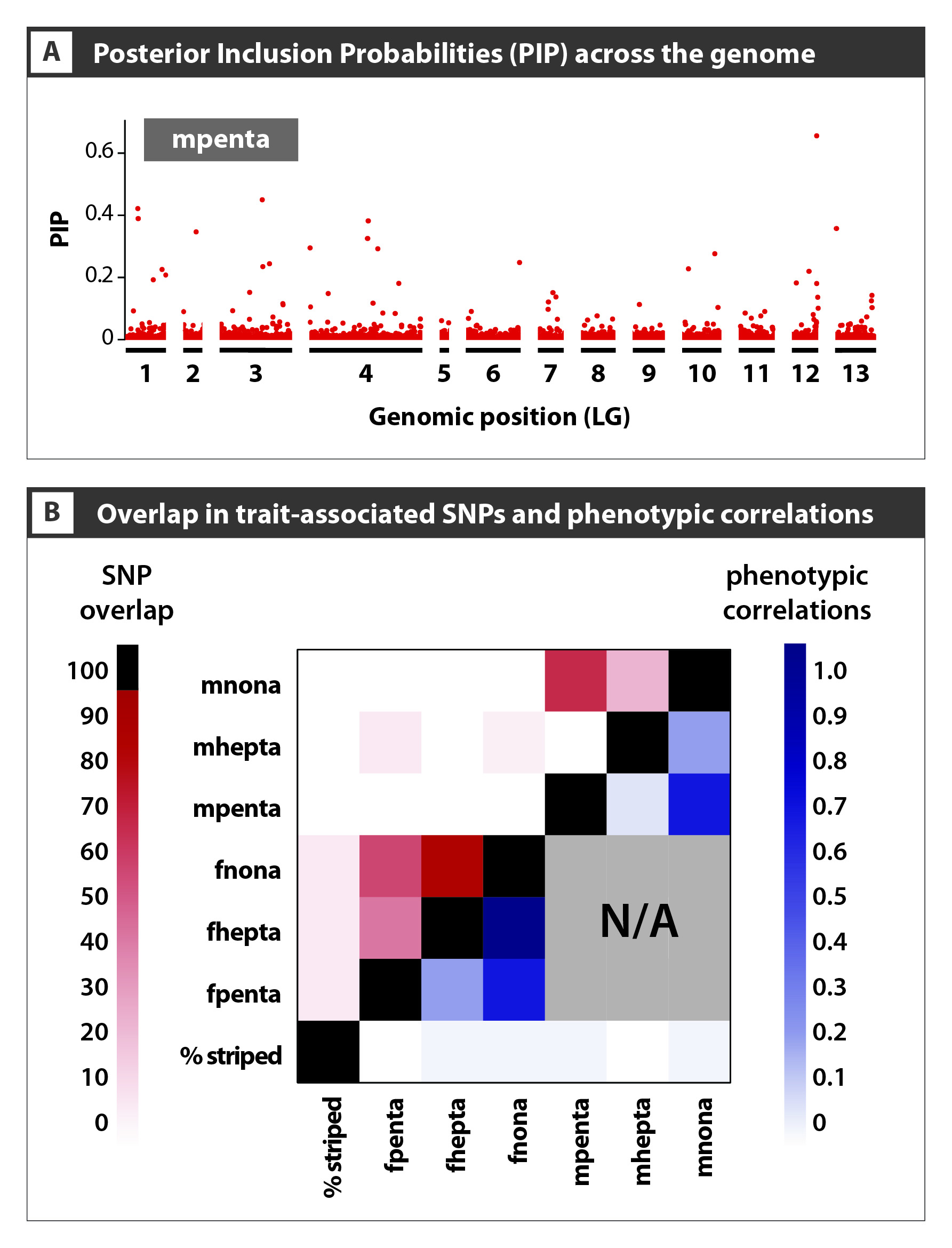
Following these standard GWA runs, we performed cross-validation analyses to test the predictive power of our GWA. The approach is akin to that commonly taken in genomic prediction/genomic selection studies28. For each trait, we estimated a predicted phenotype (based on genotype) for each individual by randomly masking 10% of individual phenotypes (‘test set’) 27 and using the remaining 90% of phenotypes (‘training set’) to obtain model parameters in GEMMA using the same parameters as in the standard runs. We then used these parameters in GEMMA to obtain predicted phenotypic values using the ‘–predict’ option27. In each instance, we ran 10 replicate MCMC chains for each training set and repeated this procedure 10 times (i.e., until we had obtained predicted values for every individual). We repeated the entire process 10 times with different random combinations of individuals in each training set to avoid any potential ‘training set’ biases, resulting in a total of 100 predicted phenotypes for each observed phenotype.

We then estimated the reliability of genomic prediction by correlating the mean predicted phenotypic values against the observed individual phenotypic values. For CHCs, GEMMA predicted values were logit-transformed because the CHC phenotypes are proportional data29; this transformation provides a more conservative estimate of the correlation. We report the square of the correlation coefficient (*r*) and its significance. This *r*-squared value estimates the phenotypic variation due to estimated additive genetic effects, with an upper limit being the ratio of observed genetic variance (VG) to phenotypic variance28, which is reported by GEMMA as PVE.

**Table S8**.**Genetic architecture of the seven traits studied.** PVE: proportion of phenotypic variance explained by genetic data (all SNPs); PGE: proportion of genetically explained phenotypic variance due to sparse (measurable) genetic effects; n-SNP: number of SNPs with measurable effect. CHCs = cuticular hydrocarbons. Given are median values and 95% equal-tail probability intervals (ETPIs). Values of *r*2 are predictive power from cross-validation runs.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Genetic architecture of non-CHC traits** | | | |  | **Genetic architecture of methylated CHCs** | | | |
| **parameter** | **median** | **2.5%**  **bound** | **97.5%**  **bound** |  | **parameter** | **median** | **2.5%**  **bound** | **97.5%**  **bound** |
|  | | |  |  |  | | |  |
| 1) % body area striped (% striped) | | | |  | 2) proportion of female pentacosanes (fpenta) | | | |
| PVE | 0.68 | 0.62 | 0.77 |  | PVE | 0.43 | 0.04 | 0.76 |
| PGE | 0.97 | 0.88 | 1.00 |  | PGE | 0.22 | 0.00 | 0.92 |
| n-SNP | 12 | 6 | 28 |  | n-SNP | 17 | 0 | 259 |
| *r*2 = 0.582, *P* < 0.001 | | | |  | *r*2 = 0.010, *P* = 0.153 | | | |
|  | | | |  |  | | | |
| 3) proportion of female heptacosanes (fhepta) | | | |  | 4) proportion of female nonacosanes (fnona) | | | |
| PVE | 0.27 | 0.01 | 0.68 |  | PVE | 0.32 | 0.02 | 0.71 |
| PGE | 0.29 | 0.00 | 0.95 |  | PGE | 0.26 | 0.00 | 0.94 |
| n-SNP | 29 | 0 | 275 |  | n-SNP | 30 | 0 | 276 |
| *r*2 = 0.040, *P* = 0.005 | | | |  | *r*2 = 0.032, *P* = 0.012 | | | |
|  | | | |  |  | | | |
| 5) proportion of male pentacosanes (mpenta) | | | |  | 6) proportion of male heptacosanes (mhepta) | | | |
| PVE | 0.74 | 0.17 | 1.00 |  | PVE | 0.64 | 0.12 | 1.00 |
| PGE | 0.15 | 0.00 | 0.83 |  | PGE | 0.16 | 0.00 | 0.80 |
| n-SNP | 51 | 0 | 281 |  | n-SNP | 17 | 0 | 266 |
| *r*2 = 0.021, *P* = 0.003 | | | |  | *r*2 = 0.016, *P* = 0.012 | | | |
|  | | | |  |  | | | |
| 7) proportion of male nonacosanes (mnona) | | | |  |  | | | |
| PVE | 0.70 | 0.21 | 1.00 |  |  |  |  |  |
| PGE | 0.29 | 0.00 | 0.93 |  |  |  |  |  |
| n-SNP | 47 | 0 | 277 |  |  |  |  |  |
| r2 = 0.042, *P* < 0.001 | | | |  |  | | | |
|  | | | |  |  | | | |

**Figure S2. Details of genome-wide association mapping.** (A) Posterior inclusion probabilities (PIP) of SNPs along the genome for a representative trait (mpenta = male methylated pentacosanes). Higher PIP values are indicative of stronger association with phenotypic variation. (B) Heat map showing the number of the 100 most strongly trait-associated SNPs per trait shared between different traits (above diagonal) and phenotypic associations between traits (*r*2 values; traits were transformed as for GWA). N/A = Not applicable. % striped = percent body area striped, fpenta = female methylated pentacosanes, fhepta = female methylated heptacosanes, fnona = female methylated nonacosanes, mpenta = male methylated pentacosanes, mhepta = male methylated heptacosanes, mnona = male methylated nonacosanes.



**Linkage group partitioning analysis.** A prediction for polygenic traits is that the number of trait-associated SNPs per LG will be positively correlated with the size of a LG30. We tested and largely supported this prediction, as reported in the main text. We calculated a point estimate for the number of trait-associated SNPs per LG by summing the PIPs across all SNPs on the LG, and defined LG size as the number of SNPs in the GWA analysis for that LG for each trait. For individual traits, number of trait-associated SNPs per LG was significantly, positively correlated with LG size for all six CHC traits (fpenta, r = 1.00; fhepta, r = 1.00; fnona, r = 1.00; mpenta, r = 1.00; mhepta, r = 1.00; mnona, r = 1.00, all *P* < 0.05), but not for % striped (r = 0.36, *P* = 0.22).

**Perfuming trials with no-choice mating experiments.** We conducted perfuming experiments to test if CHCs are important in *T. cristinae* mate choice. We collected juvenile *T. cristinae* from one study site (FHA) in the Santa Ynez Mountains, California, USA, between March and April 2014, and juvenile *T. poppensis* in late April and early May 2014 from two study sites (SMRW and SMHCRW) in the Santa Cruz Mountains, California, USA, where they feed on coastal redwood, *Sequoia sempervirens*. We captured insects as early instars using sweep nets and reared them to maturity in separate-sex containers in the laboratory on the foliage of their native host plant collected at the site of population origin (*A. fasciculatum* for *T. cristinae* and *S. sempervirens* for *T. poppensis*). For the no-choice copulation trials, we randomly selected individual *T. cristinae* from the laboratory population, tested them once, and then euthanized them. Protocols for the no-choice copulation trials used in this study are based on previously published protocols31, but were slightly altered to gain information not only on whether copulation occurred, but also when it occurred. We confined one male and one female *T. cristinae* in a 10-cm Petri dish for 4 h. For the first 15 min, we took an ‘all occurrence continuous sampling’ approach and during the remaining 225 min, we conducted ‘scan sampling’ at 15-min intervals to record for each interval if copulation occurred32. Based on *Timema* mating behaviour33, we specifically scored whether or not a pair was copulating (male sits on the female’s back with their genitals interlocked).

Each individual perfume consisted of CHCs extracted and pooled from six adult females. We created ‘conspecific’ native population CHC perfumes using hexane-extracted CHCs from six randomly selected virgin females from the same population (FHA), approximately 24 h after females had molted into sexually mature adults modified from 34,35. We created ‘heterospecific’ CHC perfumes using hexane-extracted CHCs from six randomly selected virgin females of *T. poppensis,* again approximately 24 h after they had molted into mature adults. To make a perfume, we euthanized six live females by 1-h freezing, and submerged them, one female at a time, in the same 1 ml of HPLC-grade hexane for 10 min to extract the CHCs from their body surface. We removed each female before adding the next.

We let the hexane extract passively evaporate to dryness at room temperature, inserted a live trial female into the vial containing the residual CHCs of the six extracted females, and gently hand-vortexed the vial for 1 min to facilitate CHC transfer from the vial’s walls to the body surface of the trial female. We applied the same procedure for females of the control (no perfume) treatment, except that we hand-vortexed these females in clean vials. We allowed all trial females to recover for 10 min from the perfuming procedure before the onset of a mating trial.

In total, we conducted 24 no-choice copulation trials (eight trials each with ‘conspecific native population perfume’, ‘heterospecific perfume’, and ‘no perfume’) between one male and one female *T. cristinae* from the FHA population*.* We conducted perfuming trials during the same time of day (8:45 am – 12:45 pm) on different days, and on each day ran the same number of ‘conspecific’ and ‘heterospecific’ perfuming trials simultaneously. We conducted all ‘no perfume’ trials during the last two days of testing. We analysed the latency to copulate (i.e., minutes until copulation) by means of a Kaplan-Meyer analysis in IBM SPSS Statistics 21 (IBM Corporation). Our perfuming protocol led to strong effects on mate choice (Table S9), which is congruent with previous studies in other insect systems34-36.

**Table S9. Treatment comparisons from the perfuming experiment.** All-pairwise comparisons (Log Rank tests) of the three treatments (‘conspecific perfume’, ‘heterospecific perfume’, and ‘no perfume’) for the no-choice copulation trials between one male and one female *T. cristinae*. Significant results are in bold.

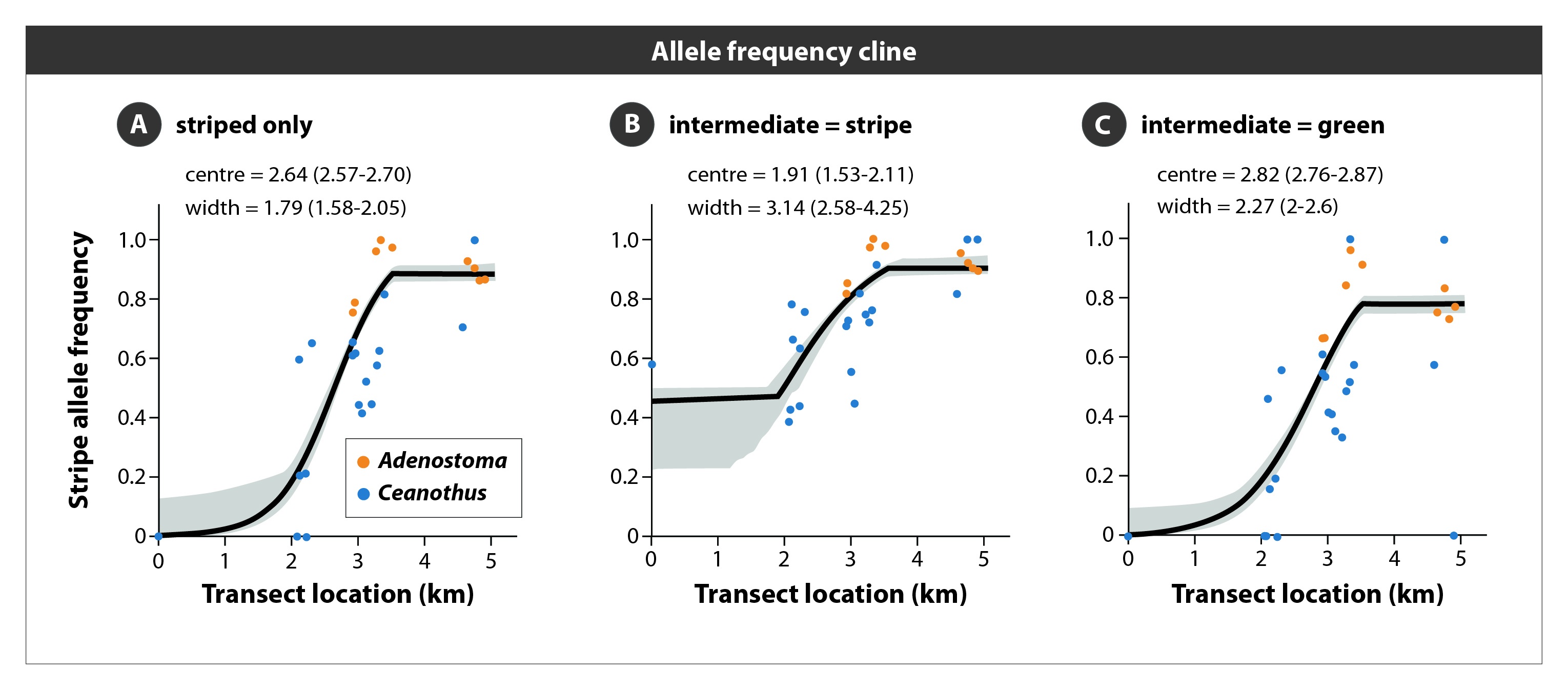
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Conspecific perfume** | | **Heterospecific perfume** | | **No perfume** | |
| **Treatment** | ***Chi*2** | ***P*** | ***Chi*2** | ***P*** | ***Chi*2** | ***P*** |
| Conspecific perfume |  |  | 16.512 | **<0.001** | 8.364 | **0.004** |
| Heterospecific perfume | 16.512 | **<0.001** |  |  | 14.681 | **<0.001** |
| No perfume | 8.364 | **0.004** | 14.681 | **<0.001** |  |  |

**Morph frequency cline in *T. cristinae.***Sampling and analytical details are contained in the methods section and a map of localities is available in Figure S4. Full data and locality information is provided in Table S10 and results in the Figure S3 bellow Table S10.

**Table S10. Sample sites and morph frequencies for the cline analysis.** G = green-unstriped, S = green-striped, I = intermediate, M = melanistic. C = *Ceanothus*. A = *Adenostoma.* Zero values for a locality across all morphs are true zeros, not due to lack of sampling the locality (N/A = not applicable due to no individuals being collected).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Locality number** | **2001** | | | |  | **1996** | | | | **Host** | **Latitude (N)** | **Longitude (W)** |
|  | **G** | **S** | **I** | **M** |  | **G** | **S** | **I** | **M** |  |  |  |
| 1 | 0 | 0 | 0 | 0 |  | 4 | 0 | 2 | 1 | C | 34.493 | -120.066 |
| 2 | 7 | 0 | 0 | 1 |  | 2 | 0 | 2 | 0 | C | 34.508 | -120.065 |
| 3 | 0 | 0 | 0 | 0 |  | 23 | 0 | 4 | 2 | C | 34.509 | -120.065 |
| 4 | 2 | 0 | 0 | 0 |  | 20 | 1 | 16 | 2 | C | 34.510 | -120.069 |
| 5 | 4 | 0 | 0 | 0 |  | 17 | 1 | 4 | 5 | C | 34.512 | -120.069 |
| 6 | 0 | 0 | 0 | 0 |  | 3 | 0 | 2 | 0 | C | 34.513 | -120.072 |
| 7 | 3 | 1 | 0 | 1 |  | 6 | 4 | 9 | 2 | C | 34.513 | -120.074 |
| 8 | 3 | 1 | 1 | 0 |  | 12 | 10 | 8 | 6 | C | 34.514 | -120.075 |
| 9 | 4 | 2 | 0 | 1 |  | 1 | 1 | 2 | 2 | C | 34.514 | -120.072 |
| 10 | 0 | 0 | 0 | 0 |  | 4 | 3 | 1 | 2 | C | 34.515 | -120.071 |
| 11 | 0 | 0 | 0 | 0 |  | 3 | 4 | 2 | 1 | A | 34.515 | -120.071 |
| 12 | 17 | 15 | 2 | 2 |  | 88 | 50 | 50 | 33 | C | 34.515 | -120.071 |
| 13 | 1 | 10 | 1 | 0 |  | 74 | 111 | 73 | 13 | A | 34.515 | -120.071 |
| 14 | 0 | 0 | 0 | 0 |  | 16 | 4 | 3 | 7 | C | 34.515 | -120.072 |
| 15 | 4 | 1 | 0 | 0 |  | 24 | 5 | 1 | 4 | C | 34.516 | -120.073 |
| 16 | 6 | 1 | 4 | 1 |  | 23 | 10 | 44 | 14 | C | 34.516 | -120.073 |
| 17 | 6 | 1 | 0 | 2 |  | 10 | 3 | 16 | 5 | C | 34.517 | -120.074 |
| 18 | 10 | 5 | 2 | 0 |  | 2 | 1 | 5 | 0 | C | 34.517 | -120.074 |
| 19 | 14 | 6 | 2 | 1 |  | 9 | 9 | 15 | 7 | C | 34.517 | -120.075 |
| 20 | 0 | 0 | 0 | 0 |  | 0 | 2 | 0 | 0 | C | 34.517 | -120.075 |
| 21 | 1 | 7 | 1 | 0 |  | 4 | 55 | 19 | 8 | A | 34.517 | -120.075 |
| 22 | 0 | 8 | 1 | 0 |  | 0 | 36 | 2 | 1 | A | 34.517 | -120.076 |
| 23 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | C | N/A | N/A |
| 24 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | A | N/A | N/A |
| 25 | 0 | 0 | 0 | 0 |  | 1 | 2 | 3 | 4 | C | 34.517 | -120.076 |
| 26 | 4 | 91 | 7 | 1 |  | 2 | 25 | 9 | 5 | A | 34.518 | -120.077 |
| 27 | 0 | 6 | 3 | 1 |  | 4 | 19 | 12 | 1 | A | 34.529 | -120.073 |
| 28 | 0 | 0 | 0 | 0 |  | 11 | 11 | 11 | 4 | C | 34.529 | -120.074 |
| 29 | 0 | 1 | 0 | 0 |  | 0 | 0 | 0 | 0 | C | 34.529 | -120.075 |
| 30 | 2 | 1 | 0 | 0 |  | 8 | 44 | 9 | 9 | A | 34.529 | -120.075 |
| 31 | 0 | 0 | 0 | 0 |  | 11 | 32 | 17 | 6 | A | 34.530 | -120.080 |
| 32 | 0 | 0 | 2 | 1 |  | 0 | 0 | 0 | 0 | C | 34.530 | -120.083 |
| 33 | 1 | 3 | 1 | 0 |  | 0 | 0 | 0 | 0 | A | 34.530 | -120.083 |

**Figure S3. Cline in allele frequency, inferred from morph frequencies (grey shaded areas are ± 95% credible intervals).** (A) Excluding intermediate phenotypes. (B) Treating intermediates as green-striped morphs. (C) Treating intermediates as green-unstriped morphs.



**Within-generation transplant experiment.** As the procedures for implementing this experiment have been previously described37, we provide here only a brief overview. We induced host shifts in nature. To do this, we collected individual *T. cristinae* (n = 500) from *Adenostoma* in an area dominated by *Adenostoma*,but in which some *Ceanothus* also occurs (population FHA). We collected individuals on 14 April 2011 and placed them in 500-mL plastic containers at a density of 50 individuals per container. The following day (15 April 2011), we randomly assigned individuals to one of 10 experimental bushes (five of each host species). Each individual then had a portion of one leg removed as a tissue sample using sterile scissors (no effect of tissue sampling on survival was seen in either lab or field experiments)37. We moved each group of 50 individuals onto either an individual of their native host plant (*Adenostoma*) or the alternative host plant (*Ceanothus*) on 16 April 2011. We recaptured surviving experimental insects using sweep nets and visual surveys during 24 and 25 April 2011, and took a second tissue sample from these insects (n = 140). Past mark recapture work and surveys conducted in this specific experiment have shown that this protocol is highly effective at recapturing the overwhelming majority of surviving individuals and that dispersal across ‘bare ground’ (grassy regions not containing suitable hosts) is near absent37. Thus, mortality resulted in the recaptured individuals in each population at the end of the experiment being a subset of those initially released (range of surviving individuals across experimental bushes = 7 – 23).

**Whole-genome re-sequencing of *T. cristinae* from the FHA population.** We re-sequenced the genomes of 473 of the 500 individuals from the FHA population using previously published protocols to extract DNA, to prepare individually-barcoded sequencing libraries, and to conduct whole-genome re-sequencing1 (we could not obtain data for 27 individuals which were distributed across blocks and treatments). We aligned the paired-end sequences to the *T. cristinae* reference genome (v0.3) using the BWA-MEM algorithm in BWA 0.7.5a-r4052. We used a minimum seed length of 20 bp, set -r to 1.3 to look for internal seeds in seeds longer than 1.3 \* 20-bp seeds, discarded chains if the seeded bases were shorter than 100 bp, and set the minimum score to output an alignment to 30. We then used SAMTOOLS to compress, sort, and index the alignments and to remove potential PCR duplicates. We then identified variant nucleotides using the UnifiedGenotyper in GATK with the prior probability of heterozygosity set to 0.001, a minimum base quality score of 20, a call confidence threshold of 50, and a maximum of 2 alleles allowed. In subsequent analyses, we considered only SNPs that mapped to one of the 13 identified LGs (*i.e.*, due to our interest in genetic architecture, we ignored the scaffolds not assigned to a LG). We further filtered the initial set of variants by retaining only those with (*i*) a minimum total sequencing depth of 500, (*ii*) a minimum of 10 reads supporting the non-reference allele, (*iii*) no more than 1 % of reads spanning an insertion-deletion, (*iv*) no more than 5 mapping quality 0 reads, (*v*) a maximum absolute value of the base quality rank sum test of 3, (*vi*) a maximum absolute value of the mapping quality rank sum test of 2, (*vii*) a maximum absolute value of the read position rank sum test of 2, and (*viii*) a minimum ratio of the variant confidence score to the non-reference read depth of 2. We then discarded SNPs with MAF less than 1%, which left us with 8.15 million SNPs for subsequent analyses.

We used an empirical Bayesian approach to estimate genotypes for the called SNPs. In particular, we calculated the posterior probability of *gij* = 0, 1, or 2 non-reference alleles as Pr(*gij* | data, *pi*) = (Pr(data | *gij*) Pr(*gij* | *pi)*)/Pr(data), where *i* and *j* index a locus and individual, Pr(data | *gij*) is the genotype likelihood calculated with GATK's UnifiedGenotyper, and Pr(*gij* | *pi*) is the probability of the genotype given Hardy-Weinberg expectations and the maximum likelihood allele frequency estimate from GATK. We then calculated the mean of the posterior (i.e., *g-hatij*) as Σ*k* = {0, 1, 2} *k* \* Pr(*gij* = *k*| data, *pi*). Finally, we obtained maximum likelihood estimates of the treatment-specific allele frequencies from the genotype estimates.

**Estimation of morphological differentiation within and between species.** Methods are described above, and full results tabulated here in Tables S11 and S12.

**Table S11**: Species, locality, and sample sizes of populations used for studying phenotypic distances between populations and species. Abbreviations for the host plants (A = *Adenostoma fasciculatum*, AC = *Acer macrophyllum*, C = *Ceanothus spinosus*, CY = *Cupressus sargentii*, DF = *Pseudotsuga menziesii*, M = *Arctostaphylos* sp., MM = *Cercocarpus betuloides*, P = *Pinus* sp., Q = *Quercus* sp., RW = *Sequoia sempervirens*).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Population Locality Code** | **Latitude (N)** | **Longitude**  **(W)** | **Host** | ***N* total**  **(males, females)** |
| *T. boharti* |  |  |  |  |
| SRHWY | 32.8223 | -116.505 | *C* | 2 (1, 1) |
| *T. californicum* |  |  |  |  |
| Lick | 37.3424 | -121.648 | *Q* | 30 (15, 15) |
| LP | 37.10186 | -121.876 | *AC, M, Q* | 38 (18, 20) |
| SM | 37.01876 | -121.726 | *M, P, Q* | 16 (3, 13) |
| *T. chumash* |  |  |  |  |
| BALD | 34.22108 | -117.668 | *C, Q* | 50 (24, 26) |
| BS | 33.81641 | -116.79 | *C* | 7 (6, 1) |
| GR10.43 | 34.22505 | -117.68 | *Q* | 48 (25, 23) |
| GR8.06 | 34.22046 | -117.707 | *MM, Q* | 80 (40, 40) |
| HF4 | 34.26536 | -118.098 | *C* | 4 (2, 2) |
| HF6 | 34.26695 | -118.117 | *Q* | 4 (2, 2) |
| HFDPD | 34.3406 | -118.016 | *M, Q* | 40 (21, 19) |
| HFRB | 34.25807 | -118.105 | *M, Q* | 3 (1, 2) |
| HFRS | 34.35558 | -118.012 | *MM, Q* | 35 (18, 17) |
| HFTP | 34.34355 | -117.983 | *C* | 32 (15, 17) |
| *T. cristinae* |  |  |  |  |
| FH | 34.51764 | -119.801 | *A* | 40 (20, 20) |
| FIGMT | 34.72803 | -119.951 | *Q* | 4 (1, 3) |
| HV | 34.48859 | -119.786 | *A, C* | 25 (15, 10) |
| L | 34.51258 | -119.796 | *A* | 31 (20, 11) |
| M | 34.51511 | -119.797 | *M* | 1 (0, 1) |
| NH | 34.51554 | -119.797 | *A* | 18 (13, 5) |
| PR | 34.53331 | -119.858 | *C* | 41 (20, 21) |
| R12 | 34.51503 | -120.071 | *A, C* | 12 (5, 7) |
| SC | 34.5226 | -119.832 | *C* | 23 (14, 9) |
| *T.* sp. ‘cuesta ridge’ |  |  |  |  |
| CRH | 35.36192 | -120.658 | *C, CY* | 42 (21, 21) |
| CRL | 35.35064 | -120.647 | *A, C, M, MM* | 51 (25, 26) |
| *T. knulli* |  |  |  |  |
| BCE | 36.0713 | -121.599 | *RW* | 36 (21, 15) |
| BCTUR | 36.06215 | -121.562 | *C* | 3 (1, 2) |
| BCXRD | 36.0706 | -121.591 | *C* | 14 (4, 10) |
| HB | 36.16438 | -121.675 | *C* | 9 (6, 3) |
| *T. landelsensis* |  |  |  |  |
| BCBOG | 36.06599 | -121.581 | *Q* | 30 (17, 13) |
| BCHC | 36.06266 | -121.573 | *M, Q* | 21 (10, 11) |
| BCHR | 36.06225 | -121.565 | *Q* | 12 (6, 6) |
| BCOG | 36.06266 | -121.573 | *C, Q* | 20 (9, 11) |
| BCSUM | 36.06544 | -121.578 | *C, Q* | 14 (7, 7) |
| *T. petita* |  |  |  |  |
| 101SS | 35.73057 | -121.314 | *C* | 33 (20, 13) |
| *T. podura* |  |  |  |  |
| BMTB | 33.82714 | -116.781 | *Q* | 6 (0, 6) |
| BS | 33.81641 | -116.79 | *C* | 4 (2, 2) |
| DZ243 | 33.85644 | -116.835 | *A* | 10 (3, 7) |
| *T. poppensis* |  |  |  |  |
| LP | 37.10186 | -121.876 | *DF, RW* | 9 (3, 6) |
| SM | 37.01876 | -121.726 | *RW* | 40 (20, 20) |
| SMHC | 37.01002 | -121.714 | *RW* | 40 (20, 20) |

**Table S12**. Trait loadings in principal components (PC) analyses. I-IV are the first four PCs. Abbreviations for the traits: BL = body length, BW = body width, HW= head width, latRG = lateral red-green colour channel, latGB = lateral green-blue colour channel, latL = lateral luminance, dorRG = dorsal red-green colour channel, dorGB = dorsal green-blue colour channel, dorL = dorsal luminance.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***Females*** | | | | ***Males*** | | | |
| ***Variable*** | **I** | **II** | **III** | **IV** | **I** | **II** | **III** | **IV** |
| BL | 0.040 | 0.535 | -0.292 | -0.007 | 0.049 | -0.523 | 0.390 | -0.047 |
| BW | 0.270 | 0.475 | 0.143 | -0.018 | 0.381 | -0.406 | -0.102 | -0.200 |
| HW | 0.237 | 0.563 | -0.048 | 0.051 | 0.323 | -0.488 | -0.067 | -0.106 |
| latRG | 0.397 | -0.185 | -0.400 | 0.440 | 0.401 | 0.124 | 0.228 | 0.571 |
| latGB | -0.444 | -0.014 | 0.153 | 0.268 | -0.371 | -0.084 | -0.401 | 0.188 |
| latL | 0.303 | 0.050 | 0.556 | 0.415 | 0.222 | -0.195 | -0.632 | 0.096 |
| dorRG | 0.397 | -0.290 | -0.308 | 0.273 | 0.436 | 0.188 | 0.011 | 0.451 |
| dorGB | -0.375 | 0.160 | 0.075 | 0.688 | -0.316 | -0.342 | -0.230 | 0.541 |
| dorL | 0.353 | -0.156 | 0.547 | -0.110 | 0.332 | 0.332 | -0.410 | -0.283 |
| Proportion variance | 34.28% | 24.34% | 13.76% | 11.01% | 35.47% | 25.54% | 14.98% | 11.25% |

**Phylogenetic inference and molecular dating.** As with the maximum-likelihood inference of bootstrap trees, we used a custom Perl script to generate a multiple alignment of 19,556 SNVs, but this time we produced consensus sequences with IUPAC ambiguities by pooling the individuals of every species and localities sampled (47 geographic populations in total). Mean coverage per variant per locality was ~133×. As before, we partitioned the alignment by linkage group (LG) and excluded the positions in genomic regions not assigned to any linkage group.

In order to infer the relationships among the populations and the position of the root without an outgroup, we first inferred a calibration-free tree using BEAST 2.1.338. We used a reversible-jump substitution model (RBS)39 for each partition, which allows sampling a mixture of models, in combination with a Γ distribution of rates to account for rate heterogeneity among sites. We used the clockstaR 1.0 R package40 to select the optimal number of relaxed molecular clocks. This computationally efficient method computes the K-tree distance metric41 between all partition trees and uses a clustering algorithm along with a goodness-of-clustering measure (the Gap statistic) to estimate the optimal number of clusters in the data42. We ran the analysis using maximum-likelihood trees of the partitions inferred with RAxML 8.0.2010 and used 1000 bootstrap replicates to estimate the Gap statistic. We found the optimal model was a single molecular clock model shared by all partitions. We used a Γ(α = 0.001, β = 1000.0) prior distribution for the clock mean rate (ucld.mean) and an Exp(λ = 0.3333) prior distribution for the clock rate standard deviation (ucld.stdev). We used a birth-death tree prior with a unif(a = 0, b = 1000) prior distribution of birth rates and a unif(a = 0, b = 1) prior distribution of relative death rates. We applied Γ(α = 0.2, β = 5.0) prior distributions for RBS rates and Exp(λ = 1.0) prior distributions for the shape (α) parameter of the Γ distributions of substitution rates. We ran four chains for 200,000,000 generations sampling every 5,000 generations, using the BEAGLE library43 to speed up analyses by using NVIDIA Tesla M2070 and K40 GPGPU cards for parallel computation. We assessed stationarity and convergence comparing visually the parameter traces with Tracer 1.6. We discarded one of the runs because it converged to a lower likelihood local optimum. We removed the first 50% of samples as burn-in of the other three runs and combined them with LogCombiner. Effective sample size (ESS) was over 200 for all the parameters and above 4000 for the posterior and likelihood. We obtained the maximum credibility tree with TreeAnnotator and summarized divergence times using the common ancestor (CA) tree approach44.

We recovered with great support (Bayesian posterior probability, BPP > 0.9) the three clades reported in previous studies45,46: Southern Clade (comprising *T. bartmani*, *T. boharti*, *T. chumash*, and *T. podura*), Santa Barbara Clade (comprising *T. cristinae*) and Northern Clade (comprising *T. californicum*, *T. knulli*, *T. landelsensis*, *T. petita*, *T. poppensis*, and a putative new species named *T. ‘cuesta ridge’*). However, differently from some of the previous studies, we inferred the root so that *T. chumash* was placed in the Southern Clade. Previous studies failed to clearly resolve species relationships, especially among the species within the Northern and Southern Clades, likely because the data were limited to one of a few genes45-49. In contrast, all the species in our tree showed strongly supported reciprocal monophyly (BPP = 1), except for the relationship between *T. knulli* and *T. poppensis*, where *T. poppensis* appears nested within *T. knulli* as a consequence of the poorly supported (BPP = 0.63) basal placement of the *T. knulli* BCE locality.

*Strategy for estimating divergence times for secondary calibration.* The fossil record of stick insect is poor, and very few fossils can be unequivocally classified into any specific extant lineage, and none into *Timema* in particular50-52. Consequently, we devised a strategy for calibrating the tree of *Timema* using secondary calibrations derived from a time-calibrated tree of insects. Use of secondary calibrations is preferable to extrapolating evolutionary rates, because molecular rates of evolution vary across the genome53, among lineages54, and through time55. Moreover, similar approaches have been successfully used for mammals beforehand56. Our strategy consisted of: (1) assembling multiple alignments of several molecular markers with sequences of the main order of insects and the main clades of *Timema* retrieved from public databases, (2) gathering a set of calibrations based on insect fossil data from the literature, (3) inferring a time-calibrated tree of insects including divergence events for the main clades of *Timema*, and (4) using divergence time estimates for such events as secondary calibrations for the inference of the tree of *Timema* using the SNVs obtained from GBS data.

*Sequence retrieval and multiple alignment.* We selected the genera to be included in the phylogenetic analysis of insects on the basis of: (1) their belonging to phylogenetically well-supported groups (~ orders) according to the most recent review of evidence57, (2) the availability of DNA sequence data in GenBank58, and (3) their adequacy to place calibrations based on availability of reliable fossil data50-52,59. Likewise, we chose a range of molecular markers previously used for phylogenetic inference of deep relationships among insect orders, among stick insects, and among *Timema* species. In particular, we used two ribosomal mitochondrial genes (12S, 16S), two protein-coding mitochondrial genes (COI, COII), two ribosomal nuclear genes (18S, 28S), and three protein-coding nuclear genes (*actin*, *h3*, and *hsp70*). We downloaded DNA sequences for 41 genera of insects of 13 orders. In some cases, we used sequences from different species of the same genus for different markers. For every genus, we chose the longest sequence when multiple accessions were available. In the case of *Timema*, we used the sequences from different species to generate consensus sequences of the Northern and Southern clades recovered with strong support in the previous calibration-free Bayesian inference. As previously, we generated consensus sequences using a custom Perl script that encoded variable positions as IUPAC ambiguities. We carried out multiple alignment of coding genes using MACSE 1.01b60, which aligns DNA sequences considering their amino acid translation, followed by the elimination of all codons with over 75% gaps. We aligned rDNA sequences using MAFFT 7.164b61 with the X-INS-i alignment framework, which uses the SCARNA pairwise alignment algorithm to account for RNA secondary structure62,63. We filtered rDNA alignments with GBLOCKS 0.91b64 setting the 'Minimum Number Of Sequences For A Conserved Position' to 50% of the number of sequences + 1, the 'Minimum Number Of Sequences For A Flank Position' to 50% of the number of sequences + 1, the 'Maximum Number Of Contiguous Nonconserved Positions' to 10, the 'Minimum Length of a Block' to 5, and 'Allowed Gap Positions' to all (-b1 = 0, -b2 = 0, -b3 = 10, -b4 = 5, -b5=a). The length of the alignments were: 578 bp for 12S, 651 bp for 16S, 1534 bp for COI, 696 bp for COII, 1985 bp for 18S, 2592 bp for 28S, 1396 bp for *actin*, 329 bp for *H3*, and 1948 bp for *Hsp70* (a total of 11,866 bp). We will deposit the alignments in the Dryad repository (pending acceptance).

We used PARTITIONFINDER 1.1.165 to select the best fit partition scheme and molecular evolution model. We tested the 16 input schemes along with JC, HKY and GTR substitution models, with and without Γ-distributed substitution rates, and with and without a proportion of invariants (details on Dryad). We used the Bayesian Information Criterion (BIC) to select the partitioning strategy best fitting the data, which was the following 6-partition scheme: (1) mitochondrial rDNA (12S and 16S, 1309 bp), (2) first and second positions of mitochondrial coding DNA (COI and COII, 1487 bp), (3) third positions of mitochondrial coding DNA (743 bp), (4) nuclear rDNA (18S and 28S, 4647 bp), (5) first and second positions of nuclear coding DNA (actin, H3, and Hsp70; 2454 bp), and (6) third positions of nuclear coding DNA (1226 bp). Subsequently, we used the clockstaR as before to select the optimal number of relaxed molecular clocks. The optimal model was a single molecular clock model shared by all partitions.

*Calibrations.* We chose six calibrations for phylogenetically well-supported groups based on robust fossil data (Table S13). We excluded implicitly uninformative calibrations. In this regard, we did not consider the only known timematodean fossil66, which we could use to calibrate the stem of *Timema*, because the stem age of *Timema* is already accounted for by including an older calibration for the stem of Euphasmatodea. We defined age intervals for the calibration using unequivocal fossil data to set hard lower bounds, and fossils from external or more inclusive (and necessarily older) groups to set conservative soft upper bounds. For each calibration, we modelled uncertainty as a Γ-probability distribution with an offset equal to the minimum, a fixed-shape parameter that concentrates the mass of the distribution towards the minimum (α = 2), and a variable rate parameter (β) so that 95% of the area lies below the maximum.

*Inference of divergence times.* We inferred a time-calibrated tree of insects with BEAST using the partitioning scheme selected with PARTITIONFINDER before, but using, for each partition, a reversible-jump substitution model (RBS)39, which allows sampling a mixture of models, in combination with a Γ distribution of rates to account for rate heterogeneity among sites. In accordance with the results of clockStart, we used a single uncorrelated lognormal molecular clock shared among all partitions. We used a Γ(α = 0.001, β = 1000.0) prior distribution for the clock mean rate (ucld.mean) and an Exp(λ = 0.3333) prior distribution for the clock rate standard deviation (ucld.stdev). We used a birth-death tree prior with a Γ(α = 0.001, β = 1000.0) prior distribution of birth rates and a Γ(α = 2.0, β = 2.0) prior distribution of relative death rates. We applied Γ(α = 0.2, β = 5.0) prior distributions for RBS rates and Exp(λ = 1.0) prior distributions for the shape (α) parameter of the Γ distributions of substitution rates. To date the tree, we placed calibrations on the stem of five well-supported groups of insects based on fossil evidence from the literature (described above, see also Table S15). Phylogenetic relationships among the orders of insects are subject to intense research and some are still under scrutiny. Therefore, we used the tree based on reviewed evidence from recent literature57 as a topological backbone for phylogenetic inference. We constrained the monophyly of every clade supported by all five kinds of data: morphological, rDNA, mtDNA, nuclear protein-coding DNA, and phylogenomic (backbone tree to be deposited in the Dryad repository). In addition, we constrained the topological relationships between the three clades of *Timema* that were strongly supported in Bayesian inferences using GBS data (see above). We evaluated the joint prior calibration distributions (i.e., effective priors) to ensure there were not unexpected interactions among the calibrations, the birth-death tree prior, and the monophyly constraints67. We ran four chains for 10,000,000 generations adding the tag 'sampleFromPrior="true"' and sampling parameters every 5,000 steps. We combined the log files with LogCombiner (part of the BEAST package), after removing the first 50% of samples as burn-in, and confirmed that the 95% confidence intervals (CI) obtained were very similar to those of the initial Γ prior distributions (Table S13). Subsequently, we ran four chains for 100,000,000 generations, sampling parameters and trees every 5,000 generations, as before. We assessed stationarity and convergence comparing visually the parameter traces with Tracer. We combined the four runs with LogCombiner after removing the first 50% of samples as burn-in. Effective sample size (ESS) was above 300 for the posterior distributions of trees and all divergence times. In particular, ESS was above 800 for the distributions of divergence times we used for secondary calibrations subsequently.

We estimate the median divergence time for the most recent common ancestor (MRCA) of *Timema* to be 30.0 Ma (95% High Posterior Density (HPD) interval: 15.3-49.8). We estimate the split between the Northern clade and the Santa Barbara clade (i.e., *T. cristinae*) to have happened 24.4 Ma ago (95% HPD: 10.6-42.0). This pushes the origin of *Timema* (crown-group sense) back by 10 Ma when compared to the previous study, which was based on cytochrome oxidase I (COI) data and the extrapolation of a generic mtDNA molecular clock rate for arthropods47.

*Bayesian inference and divergence time estimation of localities of Timema.* We used BEAST as before to carry out Bayesian inference and divergence time estimation of the populations of each locality. We used the same models and priors, with the exception of the tree prior, which was set to a calibrated Yule with a Γ(α = 0.001, β = 1000.0) prior distribution of birth rates to ensure the marginal distributions of the calibrated nodes reflect the calibration priors densities68. We fitted a Γ distribution to the posterior distributions of divergence times estimated previously with the function “fitdistr” of the MASS R package 7.3-2969 and used them as calibration priors. Specifically, we placed a Γ(α = 12.757, β = 2.326) distribution on the split between the Southern clade and the Northern and Santa Barbara clade (effectively the root), and a Γ(α = 9.791, β = 2.432) distribution on the split between the Southern clade and the Santa Barbara clade (see Table S14 for details). We ascertained the effective prior distribution of the calibrations as before, but because of problems related to chains being invariably trapped in infinite log-likelihood values after a few million generations, we ran instead 40 chains for 1 million of generations, sampling every 5000 and combined them using LogCombiner after removing the first 10% as burn-in. We confirmed that there were no unexpected interactions among the priors (Table S14). We then ran 14 chains for 200,000,000 generations, sampling every 5,000 generations. After careful visual examination of the traces with Tracer, we retained four runs that consistently converged onto the same stationary distribution for multiple parameters and showed the highest mean posterior probability and likelihood. We removed the first 75% of samples as burn-in and combined them with LogCombiner. ESS was over 200 for most of the parameters of the posterior distribution, and, in particular, it was above 3000 for the tree likelihood distributions. We obtained the maximum credibility tree with TreeAnnotator and summarized divergence times using the common ancestor (CA) tree approach44.

**Table S13.** Fossil evidence used for calibration. Uncertainty was modelled as a Γ-distribution with an offset (see text for details). The effective prior distributions are summarized given the 95% confidence interval. Ma = millions of years.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Stem group** | **Justification** | | **Age range (Ma)** | **BEAST calibration** | **Effective prior (Ma)** | **References** |
|  | Minimum | Maximum |  |  |  |  |
| Entognatha | Oldest hexapodan fossil: *Rhyniella praecursor* (Entognatha) from the Lochkovian-Pragian (Lower Devonian) | Cambrian explosion | 419-541 | Γ(α=2, β=25.7); o=419 | 420.6-530.0 | 51,70,71 |
| Holometabola | Oldest insect gall fossil trace from the Upper Pennsylvanian (Upper Carboniferous) | Split between Entognatha and Insecta | 302-419 | Γ(α=2, β=24.6); o=302 | 304.9-398.3 | 72 |
| Diptera | Oldest dipteran fossil: *Grauvogelia arzvilleriana* from the Lower Anisian (Middle Triassic) | Split between Holometabola and rest of Insecta | 247-302 | Γ(α=2, β=11.5); o=247 | 247.1-283.5 | 73 |
| Holophasmatodea | Oldest stem-phasmatodean fossil: *Cretophasmomima melanogramma* from the Yixian formation (Lower Cretaceous) | Split between Entognatha and Insecta | 129-419 | Γ(α=2, β=40.0); o=129 | 131.5-272.5 | 59 |
| Euphasmatodea | Oldest euphasmatodean fossil eggs from the Cenomanian (Upper Cretaceous) | Oldest stem-phasmatodean fossil: *Cretophasmomima melanogramma* from the Yixian formation (Lower Cretaceous) | 95-129 | Γ(α=2, β=7.1); o=95 | 96.1-150.5 | 50,51,59,74,75 |
| Phylliidae | Oldest leaf insect fossil: *Eophyllium meselensis* from the lower Middle Eocene | Origin of the MRCA of Euphasmatodea | 47-95 | Γ(α=2, β=10.1); o=47 | 47.7-82.3 | 52 |

**Table S14.** Secondary calibrations used to date the tree of *Timema* populations. Ma = millions of years.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Split** | **BEAST primary**  **estimation (Ma)** | **BEAST secondary calibration** | **BEAST secondary calibration range (Ma)** | **Effective prior (Ma)** | **BEAST secondary**  **estimation (Ma)** |
| *Timema* root | 15.3-49.8 | Γ(α=12.2, β=2.6) | 17.0-53.1 | 19.6-51.9 | 19.1-47.6 |
| Northern Clade – Santa Barbara Clade | 10.6-42.0 | Γ(α=9.6, β=2.7) | 12.6-45.8 | 10.9-35.4 | 13.2-35.0 |

**Figure S4. Maps of the study localities used in the different analysis.** (A) *T. cristinae*. (B) Genus-wide. (C) Cline. (D) Map of species ranges. Modified from 45.

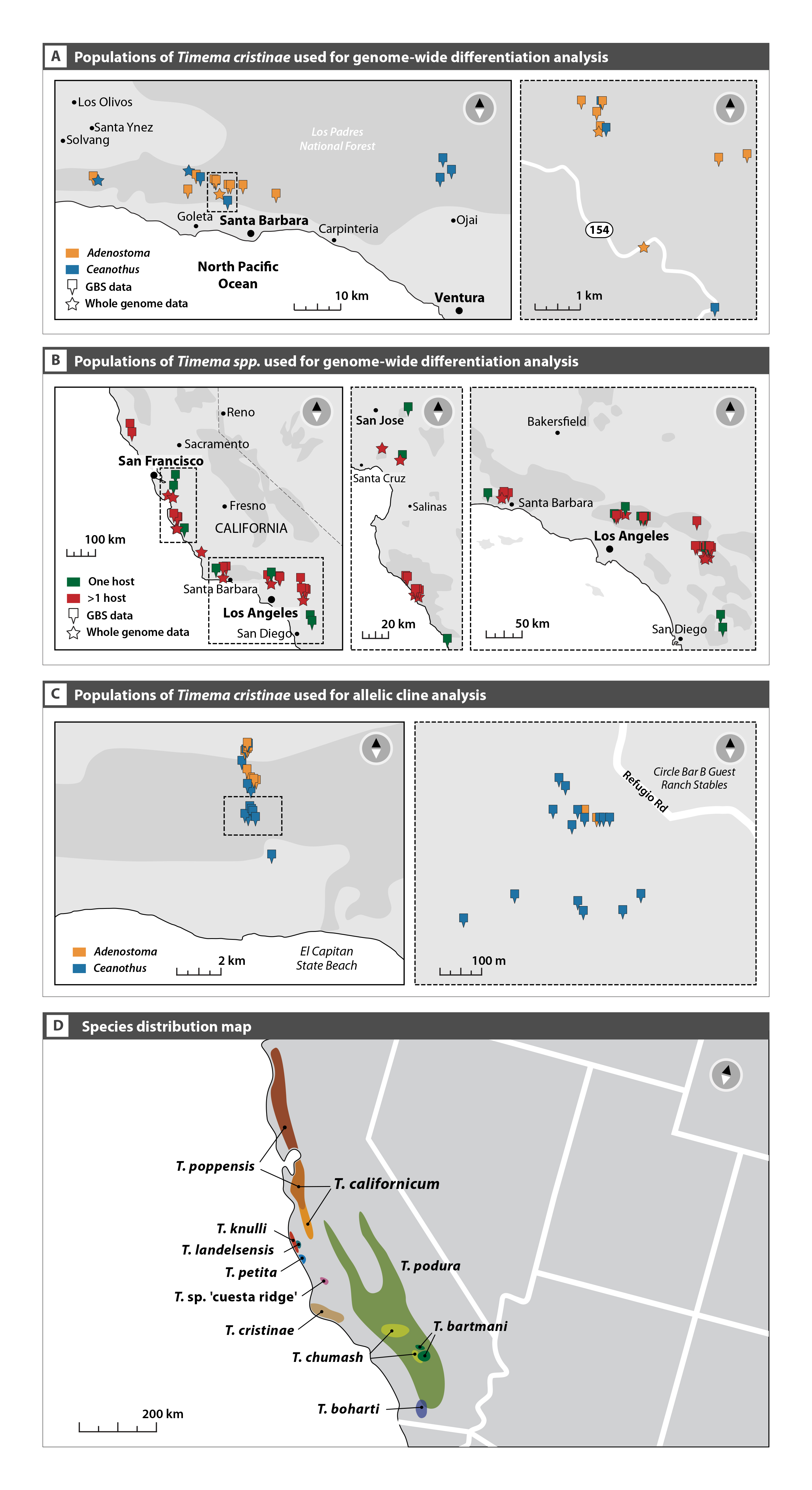
****

Table S15. Summary of data that were re-analysed from previously published studies and that are new to this study. GWA = Genome wide association. HMM = Hidden Markov Model.

|  |  |  |
| --- | --- | --- |
| Analysis | Phenotypic data | Genomic data |
| Quantification of cline in morph frequencies | New to this study | N/A |
| GWA mapping of colour-pattern | From21 | From21 |
| Whole genome HMM analysis of accentuated and background differentiation in *T. cristinae* (160 genomes) | N/A | From1 |
| Whole genome HMM analysis of transplant-and-sequence experiment (473 genomes) | From37 | New to this study |
| Association of population differentiation in colour-pattern with genomic differentiation | New to this study | New to this study |
| GWA mapping of CHCs | From21 | From21 |
| Perfuming experiment | New to this study | N/A |
| Association of population differentiation in CHCs with sexual isolation | CHCs new to this study; sexual isolation data from76 | N/A |
| Association of population differentiation in CHCs with genomic differentiation | New to this study | New to this study |
| Whole genome HMM analysis if accentuated and background differentiation in multiple *Timema* ecotypes and species (379 genomes) | N/A | New to this study |
| Genome-wide differentiation between sympatric ecotypes and species using GBS data | N/A | New to this study |
| Temporal evolution of sexual isolation between species | From22 | New to this study |
| Temporal evolution of morphological differentiation between species | New to this study | New to this study |

**References is OSM**

1 Soria-Carrasco, V. *et al.* Stick Insect Genomes Reveal Natural Selection’s Role in Parallel Speciation. *Science* **344**, 738-742, doi:10.1126/science.1252136 (2014).

2 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

3 Nosil, P. *et al.* Genomic consequences of multiple speciation processes in a stick insect. *Proceedings of the Royal Society B: Biological Sciences*, doi:10.1098/rspb.2012.0813 (2012).

4 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-U354, doi:10.1038/nmeth.1923 (2012).

5 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

6 Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting F-ST: The impact of rare variants. *Genome Research* **23**, 1514-1521, doi:10.1101/gr.154831.113 (2013).

7 Gompert, Z. *et al.* Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Mol. Ecol.* **23**, 4555-4573, doi:10.1111/mec.12811 (2014).

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

9 Gelman, A. *et al.* in *Bayesian Data Analysis* (eds A. Gelman *et al.*) 165–193 (CRC Press, 2103).

10 Stamatakis, A. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, btu033 %U <http://bioinformatics.oxfordjournals.org/content/early/2014/2001/2021/bioinformatics.btu2033> (2014).

11 Stamatakis, A., Hoover, P. & Rougemont, J. A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology* **57**, 758-771, doi:10.1080/10635150802429642 (2008).

12 Lewis, P. O. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**, 913-925, doi:10.1080/106351501753462876 (2001).

13 Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289-290, doi:10.1093/bioinformatics/btg412 (2004).

14 Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993, doi:10.1093/bioinformatics/btr509 (2011).

15 Nei, M. *Molecular Evolutionary Genetics*. (Columbia University Press, 1987).

16 Foll, M., Shim, H. & Jensen, J. D. WFABC: a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular Ecology Resources* **15**, 87-98, doi:10.1111/1755-0998.12280 (2015).

17 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**, doi:10.1371/journal.pgen.1000695 (2009).

18 Beaumont, M. A., Zhang, W. Y. & Balding, D. J. Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025-2035 (2002).

19 Csillery, K., Francois, O. & Blum, M. G. B. abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* **3**, 475-479, doi:10.1111/j.2041-210X.2011.00179.x (2012).

20 Wright, S. Isolation by distance. *Genetics* **28**, 114-138 (1943).

21 Comeault, A. A. *et al.* Selection on a Genetic Polymorphism Counteracts Ecological Speciation in a Stick Insect. *Current Biology* **25**, 1-7, doi:10.1016/j.cub.2015.05.058 (2015).

22 Schwander, T. *et al.* Hydrocarbon divergence and reproductive isolation in Timema stick insects. *BMC Evol. Biol.* **13**, doi:10.1186/1471-2148-13-151 (2013).

23 Blows, M. W. & Allan, R. A. Levels of mate recognition within and between two Drosophila species and their hybrids. *Am. Nat.* **152**, 826-837, doi:10.1086/286211 (1998).

24 Rundle, H. D., Chenoweth, S. F., Doughty, P. & Blows, M. W. Divergent selection and the evolution of signal traits and mating preferences. *Plos Biology* **3**, 1988-1995, doi:ARTN e368DOI 10.1371/journal.pbio.0030368 (2005).

25 Aitchison, J. *The statistical analysis of compositional data*. 12th edn, (Chapman and Hall, 1986).

26 Bartko, J. J. The intraclass correlation coefficient as a measure of reliability. *Physiol. Rep.* **19**, 3–11 (1966).

27 Zhou, X., Carbonetto, P. & Stephens, M. Polygenic Modeling with Bayesian Sparse Linear Mixed Models. *Plos Genetics* **9**, doi:e100326410.1371/journal.pgen.1003264 (2013).

28 Wray, N. R., Yang, J., Goddard, M. E. & Visscher, P. M. The Genetic Interpretation of Area under the ROC Curve in Genomic Profiling. *Plos Genetics* **6**, doi:10.1371/journal.pgen.1000864 (2010).

29 Crawley, M. J. *Statistics: An introduction using R.* . (Wiley, NY, 2005).

30 Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* **43**, 519-U544, doi:10.1038/ng.823 (2011).

31 Nosil, P., Crespi, B. J. & Sandoval, C. P. Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature* **417**, 440-443 (2002).

32 Martin, P. & Bateson, P. *Measuring Behaviour: An Introductory Guide*. 3rd Edition edn, (Cambridge University Press, 2007).

33 Arbuthnott, D. & Crespi, B. J. Courtship and mate discrimination within and between species of Timema walking-sticks. *Anim. Behav.* **78**, 53-59, doi:10.1016/j.anbehav.2009.02.028 (2009).

34 Thomas, M. L. & Simmons, L. W. Male-derived cuticular hydrocarbons signal sperm competition intensity and affect ejaculate expenditure in crickets. *Proc. R. Soc. B-Biol. Sci.* **276**, 383-388, doi:10.1098/rspb.2008.1206 (2009).

35 Weddle, C. B. *et al.* Cuticular hydrocarbons as a basis for chemosensory self-referencing in crickets: a potentially universal mechanism facilitating polyandry in insects. *Ecol. Lett.* **16**, 346-353, doi:10.1111/ele.12046 (2013).

36 Dyer, K. A., White, B. E., Sztepanacz, J. L., Bewick, E. R. & Rundle, H. D. REPRODUCTIVE CHARACTER DISPLACEMENT OF EPICUTICULAR COMPOUNDS AND THEIR CONTRIBUTION TO MATE CHOICE IN DROSOPHILA SUBQUINARIA AND DROSOPHILA RECENS. *Evolution* **68**, 1163-1175, doi:10.1111/evo.12335 (2014).

37 Gompert, Z. *et al.* Experimental evidence for ecological selection on genome variation in the wild. *Ecol. Lett.* **17**, 369-379, doi:10.1111/ele.12238 (2014).

38 Bouckaert, R. *et al.* BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput Biol* **10**, e1003537, doi:10.1371/journal.pcbi.1003537 (2014).

39 Bouckaert, R., Alvarado-Mora, M. V. & Pinho, J. R. R. Evolutionary rates and HBV: issues of rate estimation with Bayesian molecular methods. *Antiviral Therapy* **18**, 497-503, doi:10.3851/IMP2656 (2013).

40 Duchêne, S., Molak, M. & Ho, S. Y. W. ClockstaR: choosing the number of relaxed-clock models in molecular phylogenetic analysis. *Bioinformatics* **30**, 1017-1019, doi:10.1093/bioinformatics/btt665 (2014).

41 Soria-Carrasco, V., Talavera, G., Igea, J. & Castresana, J. The K tree score: quantification of differences in the relative branch length and topology of phylogenetic trees. *Bioinformatics* **23**, 2954-2956, doi:10.1093/bioinformatics/btm466 (2007).

42 Tibshirani, R., Walther, G. & Hastie, T. Estimating the Number of Clusters in a Data Set via the Gap Statistic. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **63**, 411-423 (2001).

43 Ayres, D. L. *et al.* BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics. *Systematic Biology*, doi:10.1093/sysbio/syr100 (2011).

44 Heled, J. & Bouckaert, R. Looking for trees in the forest: summary tree from posterior samples. *BMC Evol. Biol.* **13**, 221 (2013).

45 Law, J. H. & Crespi, B. J. The evolution of geographic parthenogenesis in Timema walking-sticks. *Mol. Ecol.* **11**, 1471–1489, doi:10.1046/j.1365-294X.2002.01547.x (2002).

46 Law, J. H. & Crespi, B. J. Recent and ancient asexuality in Timema walkingsticks. *Evolution* **56**, 1711-1717, doi:10.1111/j.0014-3820.2002.tb01484.x (2002).

47 Sandoval, C., Carmean, D. A. & Crespi, B. J. Molecular phylogenetics of sexual and parthenogenetic Timema walking-sticks. *Proc. R. Soc. B-Biol. Sci.* **265**, 589-595 (1998).

48 Crespi, B. J. & Sandoval, C. P. Phylogenetic evidence for the evolution of ecological specialization in Timema walking-sticks. *J. Evol. Biol.* **13**, 249-262 (2000).

49 Schwander, T., Henry, L. & Crespi, B. J. Molecular Evidence for Ancient Asexuality in Timema Stick Insects. *Current Biology* **21**, 1129-1134, doi:10.1016/j.cub.2011.05.026 (2011).

50 Tilgner, E. The fossil record of Phasmida (Insecta: Neoptera). *Insect Systematics & Evolution* **31**, 473-480 (2000).

51 Grimaldi, D. & Engel, M. S. *Evolution of the Insects*. (Cambridge University Press, 2005).

52 Wedmann, S., Bradler, S. & Rust, J. The first fossil leaf insect: 47 million years of specialized cryptic morphology and behavior. *Proceedings of the National Academy of Sciences* **104**, 565-569, doi:10.1073/pnas.0606937104 (2007).

53 Hodgkinson, A. & Eyre-Walker, A. Variation in the mutation rate across mammalian genomes. *Nature Reviews Genetics* **12**, 756-766, doi:10.1038/nrg3098 (2011).

54 Lanfear, R., Welch, J. J. & Bromham, L. Watching the clock: Studying variation in rates of molecular evolution between species. *Trends Ecol. Evol.* **25**, 495-503, doi:10.1016/j.tree.2010.06.007 (2010).

55 Ho, S. Y. W., Phillips, M. J., Cooper, A. & Drummond, A. J. Time Dependency of Molecular Rate Estimates and Systematic Overestimation of Recent Divergence Times. *Mol Biol Evol* **22**, 1561-1568, doi:10.1093/molbev/msi145 (2005).

56 Soria-Carrasco, V. & Castresana, J. Diversification rates and the latitudinal gradient of diversity in mammals. *Proc. R. Soc. B-Biol. Sci.* **279**, 4148-4155, doi:10.1098/rspb.2012.1393 (2012).

57 Trautwein, M. D., Wiegmann, B. M., Beutel, R., Kjer, K. M. & Yeates, D. K. Advances in Insect Phylogeny at the Dawn of the Postgenomic Era. *Annu. Rev. Entomol.* **57**, 449-468, doi:10.1146/annurev-ento-120710-100538 (2012).

58 Benson, D. A. *et al.* GenBank. *Nucleic Acids Research* **41**, D36-D42, doi:10.1093/nar/gks1195 (2013).

59 Wang, M. *et al.* Under Cover at Pre-Angiosperm Times: A Cloaked Phasmatodean Insect from the Early Cretaceous Jehol Biota. *PLoS One* **9**, e91290, doi:10.1371/journal.pone.0091290 (2014).

60 Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E. J. P. MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. *PLoS One* **6**, e22594, doi:10.1371/journal.pone.0022594 (2011).

61 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

62 Katoh, K. & Toh, H. Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics* **9**, 212, doi:10.1186/1471-2105-9-212 (2008).

63 Tabei, Y., Kiryu, H., Kin, T. & Asai, K. A fast structural multiple alignment method for long RNA sequences. *BMC Bioinformatics* **9**, 33, doi:10.1186/1471-2105-9-33 (2008).

64 Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540-552 (2000).

65 Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.*, doi:10.1093/molbev/mss020 (2012).

66 Zompro, O. *Inter- and intra-ordinal relationships of the Mantophasmatodea, with comments on the phylogeny of polyneopteran orders (Insecta: Polyneoptera)*. (2005).

67 Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, doi:21410.1186/1471-2148-7-214 (2007).

68 Heled, J. & Drummond, A. J. Calibrated Tree Priors for Relaxed Phylogenetics and Divergence Time Estimation. *Systematic Biology* **61**, 138-149, doi:10.1093/sysbio/syr087 (2012).

69 Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. (Springer New York, 2002).

70 Hirst, S. On some Arthropod Remains from the Rhynie Chert (Old Red Sandstone). *Geological Magazine* **63**, 69 - 71, doi:10.1017/S0016756800083692 (1926).

71 Whalley, P. & Jarzembowski, E. A. A new assessment of Rhyniella, the earliest known insect, from the Devonian of Rhynie, Scotland. *Nature* **291**, 317-317, doi:10.1038/291317a0 (1981).

72 Labandeira, C. C. & Phillips, T. L. A Carboniferous insect gall: insight into early ecologic history of the Holometabola. *Proceedings of the National Academy of Sciences* **93**, 8470-8474 (1996).

73 Krzeminski, W., Krzeminska, E. & Papier, F. Grauvogelia arzvilleriana sp. n. - the oldest Diptera species (Lower/Middle Triassic of France). *Acta Zoologica Cracoviensia* **37**, 95-99 (1994).

74 Rasnitsyn, A. P. & Ross, A. J. A preliminary list of arthropod families present in the Burmese amber collection at The Natural History Museum, London. *BULLETIN-NATURAL HISTORY MUSEUM GEOLOGY SERIES* **56**, 21–24 (2000).

75 Zherikhin, V. V. & Ross, A. J. A review of the history, geology and age of Burmese amber (Burmite). *Bulletin Natural History Museum London (Geology)* **56**, 3–10 (2000).

76 Nosil, P. & Hohenlohe, P. A. Dimensionality of sexual isolation during reinforcement and ecological speciation in Timema cristinae stick insects. *Evolutionary Ecology Research* **14**, 467–485 (2012).