
Article:
Comeault, A.A., Carvalho, C.F., Dennis, S. et al. (2 more authors) (2016) Color phenotypes are under similar genetic control in two distantly related species of Timema stick insect. Evolution, 70 (6). pp. 1283-1296. ISSN 0014-3820

https://doi.org/10.1111/evo.12931

Full Title: Color phenotypes are under similar genetic control in two distantly related species of Timema stick insect

Aaron A. Comeault[1,2], Clarissa Ferreira[1], Stuart Dennis[1,3], Víctor Soria-Carrasco[1], and Patrik Nosil[1]

[1]Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK, [2]Department of Biology, University of North Carolina, Chapel Hill, N.C., USA, 27516, [3]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dept. of Aquatic Ecology, Dübendorf, Switzerland

Corresponding author:
Aaron A. Comeault
Department of Biology, University of North Carolina,
Chapel Hill, NC, USA, 27599
Email: aacomeault@gmail.com

Running title: Genetic architecture of color in Timema

Keywords: genome-wide association mapping, natural selection, polymorphism, crypsis, convergence

This manuscript contains 7,377 words, 2 tables, and 5 figures. All data will be deposited at Dryad (phenotypic data and scripts) or NCBI's short read archive (Illumina sequence reads).

Abstract

Ecology and genetics are both of general interest to evolutionary biologists as they can influence the phenotypic and genetic response to selection. The stick insects Timema podura and T. cristinae exhibit a green/melanistic body color polymorphism that is subject to different ecologically-based selective regimes in the two species. Here we describe aspects of the genetics of this color polymorphism in T. podura, and compare this to previous results in T. cristinae. We first show that similar color phenotypes of the two species cluster in phenotypic space. We then use genome-wide association mapping to show that in both species, color is controlled by few loci, dominance relationships between color alleles are the same, and SNPs associated with color phenotypes co-localize to the same linkage group. Regions within this linkage group that harbor genetic variants associated with color exhibit elevated linkage disequilibrium relative to genome wide expectations, but more strongly so in T. cristinae. We use these results to discuss predictions regarding how the genetics of color could influence levels of phenotypic and genetic variation that segregate within and between populations of T. podura and T. cristinae, drawing parallels with other organisms.

Introduction

Recent advances in sequencing technologies have facilitated a proliferation of studies describing genomic patterns of differentiation between species or populations found in different geographical or ecological contexts (Hohenlohe et al. 2010; Nadeau et al. 2012; Jones et al. 2012; Ellegren et al. 2012; Soria-Carrasco et al. 2014; Poelstra et al. 2014). While in some cases genetic regions showing accentuated differentiation harbor genes that are known to underlie traits involved in adaptation (Dasmahapatra et al. 2012; Poelstra et al. 2014), the phenotypic effects of the genes contained within such 'outlier' regions is typically unknown. Identifying genetic regions harboring adaptive loci is thus a key goal in evolutionary biology, and can facilitate subsequent tests of how selection affects patterns of genetic differentiation.

Even if the specific genes controlling adaptive phenotypes are unknown, general aspects of their genetics, such as numbers of loci underlying phenotypic variation, dominance relationships between alleles, and the genomic distribution of adaptive genes can provide insight into the evolutionary process (Rausher & Delph 2015). For example, when adaptation is the result of many loci, each having small effects on phenotypic variation, the genetic response to selection is expected to result in minor and even transient shifts in allele frequencies across loci (Pritchard et al. 2010; Berg & Coop 2014; Yeaman 2015). This scenario can contrast one where strong genetic differentiation can be observed as a result of selection acting on traits controlled by few loci, each having large phenotypic effects (Nadeau et al. 2012; Poelstra et al. 2014). Thus, whether traits are

highly polygenic versus controlled by few loci of large-effect has implications for patterns of genomic differentiation.

Additional aspects of genetic architecture, such as dominance relationships among alleles and localized patterns of linkage disequilibrium (LD), will also affect the response to selection. For example, dominance relationships between alleles can affect whether an allele's phenotypic effects are expressed and thus visible to selection (Haldane 1927; Charlesworth 1992; Rosenblum et al. 2010). Patterns of LD in genomic regions harboring alleles involved in adaptation will influence the genomic extent to which selection will affect genetic differentiation: if LD is high, selection can have impacts across a broad genomic region, while low LD is expected to result in more localized effects (Maynard Smith & Haigh 1974; Barton 2000). Understanding these aspects of the genetics of traits therefore has important implications for understanding patterns of differentiation and segregation among and within populations, respectively.

Here we study the genetics of a green/melanistic color polymorphism found in two species of Timema stick insects. The genus Timema is comprised of ~21 species of herbivorous insects that are endemic to southwestern North America and show a wide range of within- and among-species variation in body coloration (Sandoval et al. 1998). This variation is known to be of adaptive significance, for example in crypsis and the avoidance of visual predation by lizards and birds (Sandoval 1994; Sandoval & Nosil 2005). Similar color phenotypes frequently segregate as polymorphisms in distantly related species of Timema (Crespi & Sandoval 2000), providing a well suited system for addressing questions regarding the ecology and genetics of the evolution of adaptive color phenotypes. Timema podura and T. cristinae are two such species that both display

an intraspecific polymorphism in color (Fig. 1). These species are estimated to have diverged from a common ancestor approximately 20 million years ago (Timema have a single generation per year; Sandoval et al. 1998) and represent an interesting opportunity to study the evolution of color because it is unclear whether a similar genetic basis underlying phenotypic variation is expected between species that diverged so long ago (Conte et al. 2012). The goals of this study are therefore to (1) quantitatively describe similarities (and differences) in color between T. podura and T. cristinae, (2) determine aspects of the genetic control of color in T. podura – including the number of loci underlying this variation and dominance relationships between alleles – to facilitate comparisons with that of T. cristinae, (3) compare patterns of LD observed within genomic regions containing candidate SNPs associated with color in both T. podura and T. cristinae, and (4) generate predictions regarding how aspects of genetics might influence genetic differentiation in each species, which can be tested in future work.

Methods

Study system

Timema cristinae is endemic to the coastal chaparral of the westernmost mountains of the Transverse Ranges of southern California and is primarily found on the two host plants Ceanothus spinosus (Rhamnaceae) and Adenostoma fasciculatum (Rosaceae). Within T. cristinae, green and melanistic color phenotypes segregate as a polymorphism and the frequency of the color phenotypes does not differ between populations inhabiting different species of host plants (Comeault et al. 2015). Green

individuals of this species can also express a dorsal white stripe, however this trait is not expressed in melanistic individuals (Comeault et al. 2015) and thus we do not deal with it here. The color phenotypes of T. cristinae are maintained within populations due to a balance of selective agents that are similar between hosts and include selection for crypsis in leafy (green favored) and woody (melanistic favored) plant microhabitats, differences in fungal infection rates, and potential fitness differences associated with climatic variation. Classical genetic crosses and genome wide association (GWA) mapping indicate that T. cristinae color phenotypes are under simple Mendelian control with most variation in color being explained by a single region on linkage group 8 (LG 8), with the green allele dominant to the melanistic allele (Comeault et al. 2015).

The other species we consider, T. podura, is endemic to the San Bernardino, Santa Rosa, and San Jacinto Mountains of central southern California and also inhabits host plant species in the genus Ceanothus (C. leucodermis) and Adenostoma (A. fasciculatum). Like T. cristinae, T. podura display either a green or a melanistic color phenotype (Fig. 1; melanistic individuals have also been referred to as "grey" or "red"; Sandoval & Nosil, 2005). Unlike T. cristinae, the frequency of T. podura color phenotypes is different between populations living on different host species: green T. podura are, to current knowledge, not found on A. fasciculatum (Sandoval & Nosil 2005). Experiments have shown that avian predators preferentially depredate melanistic individuals when on C. leucodermis (potentially due to the light green color of C. leucodermis branches) and green individuals on Adenostoma (Sandoval & Nosil 2005). Thus, in contrast to T. cristinae, there is evidence for divergent selection acting on T. podura color phenotypes between host species. The maintenance of melanistic T. podura

131 on C. leucodermis could be due to gene flow between populations found on different
132 hosts, as documented in T. cristinae at spatial scales similar to those separating
133 populations of T. podura on different hosts (Nosil et al. 2012; Sandoval and Nosil 2005).
134 Aspects of the genetics of color (such as dominance relationships among alleles) could
135 also contribute to the maintenance of maladaptive phenotypic variation, but until now the
136 have remained unknown.
137
138 Quantifying variation in color
139 To quantitatively measure color we recorded digital images of 42 adult T. podura
140 collected from a phenotypically variable population found on C. leucodermis plants
141 (population code: BSC; 33.816°N, -116.790°W) and 602 T. cristinae found on A.
142 fasciculatum (FHA; 34.518°N, -119.801°W). The 602 T. cristinae images have
143 previously been used to qualitatively describe color (classified as green versus non-green;
144 Comeault et al. 2015), while here we report novel analyses that quantitatively describe
145 color in a manner that allows direct comparison between the two species. All digital
146 images were recorded under standard conditions and color and exposure were corrected
147 in post-processing (SI).
148 For each image we first recorded RGB values of the lateral margin of the second
149 thoracic segment for each individual using the color histogram plugin in ImageJ
150 (Abràmoff et al. 2004). For each individual we then summarized variation in red-green
151 color (RG) using the relationship (R-G)/(R+G), green-blue color (GB) as (G-B)/(G+B),
152 and luminance (i.e. brightness; L) as (R+G+B) (Endler 2012). While this method of
153 measuring color does not take into account the visual system of the receiver or the light

environment an object is viewed in, it does represent an unbiased measurement of color that can be useful in a comparative context. Because T. cristinae does not reflect UV light (Comeault et al. 2015) the digital images we use here likely capture a majority of the biologically relevant differences between the color phenotypes.

We quantified phenotypic overlap between T. podura and T. cristinae color morphs using RG, GB, and L values. First, we used linear models assuming normally distributed error to compare RG, GB, and L values between green and melanistic T. podura color phenotypes, green and melanistic T. cristinae color phenotypes, green T. podura and green T. cristinae, and melanistic T. podura and melanistic T. cristinae. We also analyzed the position of the different colors in phenotypic space using an approach analogous to that used by Beuttell & Losos (1999) to quantify clustering of Anolis ecomorphs in multivariate phenotypic space. Specifically, we first calculated the Euclidean distance between all individuals in our sample (i.e. all pairwise comparisons) in RG – GB color space. We then used Wilcoxon signed rank tests to determine whether phenotypic distances observed between the same color phenotypes of the two species (i.e. T. podura and green T. cristinae or melanistic T. podura and melanistic T. cristinae) were less than the phenotypic distance between different colored individuals of the same species. These analyses enabled us to ask whether the same color phenotypes of the two species are closer to each other, in phenotypic space, than to the alternate color phenotype of their own species. All statistical analyses were carried out in R (Core Team 2013).

Genomic sampling of T. podura

We extracted whole genomic DNA from 50 T. podura (19 green and 31 melanistic) that included the same 42 individuals used to quantify color and 8 additional individuals sampled from the same population that were not photographed (but were qualitatively scored as "green" or "melanistic") using Qiagen DNeasy blood and tissue kits (Qiagen). We then used the method of Parchman et al. (2012) to generate individually barcoded restriction-site associated DNA libraries for each of these 50 individuals (SI). We pooled these 50 libraries with an additional 48 uniquely barcoded libraries that were part of another study, selected for fragments ranging in size from 300 to 500 bp with Pippin-prep targeted size selection (Sage Science, Inc., MA, USA), and sequenced on a single lane of the Illumina HiSeq2000 platform using V3 reagents at the National Center for Genome Research (Santa Fe, NM, USA).

We removed barcodes and the following six bp of the EcoRI cut site from raw sequence reads, while allowing for single bp errors in the barcode sequence due to synthesis or sequencing error, using a custom Perl script developed and implemented in Nosil et al. (2012). Following removal of barcode sequences this resulted in a total of 130,280,785 raw sequence reads with an average of 2,605,616 reads per individual (95% interval = 1,351,013 – 3,356,050) and an average length of 83 bp (95% interval = 63 – 86). We aligned 90,923,479 of these reads (69.8%) to the reference genome sequence of T. cristinae (Soria-Carrasco et al. 2014) using BOWTIE2 version 2.1.0 (Langmead & Salzberg 2012) with the local model and the '--very-sensitive-local' preset (-D 20 -R 3 -N 0 -L 20 -i S,1,0.50). We used SAMTOOLS version 0.1.19 (Li et al. 2009) to sort and index alignments. We used the reads mapped to the T. cristinae genome to generate a reference consensus sequence of T. podura using SAMTOOLS "mpileup", and

BCFTOOLS. We used vcfutils.pl with the "vcf2fq" command to filter out positions with a number of reads below 8 and above 500, as well as those with a phred-scale mapping quality score lower than 20. Filtered sites were coded as missing data. Subsequently, we used BOWTIE2 with the same arguments used above to align 100,095,223 raw reads (76.8%) to this reference consensus. As before, the alignments were sorted and indexed with SAMTOOLS.

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. Insertion and deletion polymorphisms were discarded. We identified 638,828 single nucleotide polymorphisms (SNPs) that were reduced to 137,650 SNPs after discarding SNPs for which there were sequence data for less than 40% of the individuals, low confidence calls with a phred-scale quality score lower than 20, and SNPs with more than two alleles. Average depth of the retained SNPs across all individuals was ~460x (mean coverage per SNP per individual ~9x).

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. $p(A)=p^2$; $p(a)=(1-p)^2$; $p(Aa)= 2p(1-p)$). We then computed the posterior mean genotype scores for each individual, at each locus, by multiplying the probability of the homozygous minor allele genotype by two and adding the probability of the heterozygous genotype. These imputed genotype scores range from zero to two and represent the dosage of the minor allele in a given genotype. All imputed genotype scores were saved

in bimbam file format. These imputed genotypes were used for multi-locus GWA mapping analyses and principal components analyses (PCAs) described below. Because other analyses (e.g. analyses of linkage disequilibrium and single-SNP GWA mapping) required discrete genotypic data, we collapsed imputed genotype scores into three discrete genotypic values: imputed genotypes ranging from 0 to 0.6 (inclusive) were scored as homozygous for the minor allele, imputed genotypes between 0.6 and 1.4 were scored as heterozygous, and imputed genotypes greater than or equal to 1.4 were scored as homozygous for the major allele.

Genetic structure within the T. podura sample

Population structure can confound GWA mapping studies (Freedman et al. 2004; Price et al. 2006). Although this is unlikely to be a major issue in our data set because T. podura were sampled in a single locality at the scale of only hundreds of meters, we nonetheless tested for genetic structure using two approaches. First, we used a hierarchical Bayesian model that jointly estimates genotypes and admixture proportions as implemented in the program ENTROPY (available from Gompert et al. 2014). This model is similar to the popular STRUCTURE algorithm (Pritchard et al. 2000), but accounts for sequencing error and genotype uncertainties inherent to next-generation sequencing methods (for comparable approach see Skotte et al. 2013). We estimated parameters for models with K = 1-4 population clusters and used the deviance information criterion (DIC) to determine the number of clusters most appropriately represented by our data (Spiegelhalter et al. 2002; see SI for details).

In addition to hierarchical Bayesian modeling, we carried out a PCA on the matrix of imputed genotype scores using the "pca" function in the R library PCAMETHODS. We then assessed the number of PCs that significantly described genetic variation using the Q^2 cross-validation statistic (Krzanowski 1987) as calculated using the argument "cv = 'q2'" within the "pca" function. The value of Q^2 represents a measure of the explained variation of a given PC relative to random expectations and is calculated as 1 – (predicted residual sum of squares / residual sum of squares) (Krzanowski 1987; Abdi & Williams 2010). We interpreted PCs with $Q^2 > 0.05$ as capturing a significant amount of variation in our data (Abdi & Williams 2010). To determine whether phenotypic variation in color was concordant with genetic variation described by significant PCs, we fit generalized linear models with binomial error terms for each PC, where the PC score was the predictor variable and color was the response variable. If a PC explained a significant amount of variation in color (as determined using likelihood ratio tests and a Bonferroni-corrected alpha = 0.0036), we assessed the strength of that PC's association with color phenotypes using the proportional increase in residual deviance explained by that model relative to the null (i.e. pseudo R^2; Dobson 2002). As described in the Results, these analyses show there is no major axis of genetic variation that is correlated with color phenotypes in our data set. Nonetheless, we account for relatedness and population structure in our GWA analyses as described below.

Genetic control of T. podura color phenotypes estimated through GWA mapping

We estimated aspects of the genetic basis of color phenotypes in T. podura using multi-locus Bayesian sparse linear mixed models (BSLMMs) as implemented in the

267 software package GEMMA (Zhou & Stephens 2012; Zhou et al. 2013). Because T.
268 podura color phenotypes were completely non-overlapping in two-dimensional color
269 space (Fig. 1b) we unambiguously scored each of the 50 genotyped individuals as green
270 (n = 19) or melanistic (n = 31) and ran probit BSLMMs in GEMMA (as done previously
271 for green and melanistic phenotypes of T. cristinae: Comeault et al. 2015). Multi-locus
272 association mapping in GEMMA accounts for both relatedness among individuals and LD
273 between SNPs by including a genomic kinship matrix as a random effect and estimating
274 SNP effect sizes while controlling for other SNPs included in the model, respectively
275 (Zhou et al. 2013).
276 Bayesian sparse linear mixed models as implemented in GEMMA also provide
277 useful estimates of hyperparameters that quantitatively describe the genetics of traits
278 (Zhou & Stephens 2012; Zhou et al. 2013; see Discussion). These hyperparameters
279 include the total phenotypic variation explained by all SNPs (proportion of phenotypic
280 variation explained; PVE), the proportion of PVE that can be explained by 'measurable-
281 effect' SNPs that have non-zero, and detectable, effects on phenotypic variation (PGE)
282 that are independent of the kinship matrix included in the model, and the number of
283 independent genomic regions needed to explain the PVE (n-SNPs; the number of SNPs
284 where the relationship between genotype and phenotype [β] is estimated to be greater
285 than zero).
286 We implemented BSLMMs in GEMMA using 10 independent Markov-chain
287 Monte Carlo (MCMC) chains ran for 25 million steps with an initial burn-in period of 5
288 million steps. Parameter values estimated by BSLMMs were recorded every 100 steps
289 and written every 10,000 steps. All additional options in GEMMA remained at default

values and SNPs with minor allele frequencies < 0.01 were excluded from these analyses (121,435 SNPs retained). Here we report the median and 95% credible interval (95% equal tail posterior probability intervals [95% ETPPIs]) for PVE, PGE, PVE x PGE (an estimate of the total phenotypic variation explained by only SNPs with large phenotypic effects), and n-SNP. To assess the strength of the genetic signal in our data set to accurately estimate hyperparameters we carried out both permutation tests and cross-validation using genomic prediction (SI).

In addition to the hyperparameters described above, GEMMA provides the posterior inclusion probability (PIP) and estimates the phenotypic effect (β) of each SNP that is identified as having a non-zero effect on phenotypic variation in at least one model iteration. PIP is computed as the proportion of model iterations that a given SNP is identified as having a non-zero β. SNPs that are more strongly associated with phenotypic variation are therefore expected to have large PIPs and these SNPs are the strongest candidates of being linked to the functional variant(s) underlying phenotypic variation. Thus, the magnitude of the PIP of a SNP reflects the weight of evidence that that SNP is associated with variation in T. podura color phenotypes.

For comparison with multi-locus GWA mapping analyses, we also implemented single-SNP GWA mapping. This analysis was carried out following the EIGENSTRAT method of Price et al. (2006) as implemented in the GENABEL R library (Aulchenko et al. 2007). Prior to single-SNP GWA mapping we remove SNPs with minor allele frequencies less than 0.01, individuals with call rates < 0.95, individuals with the proportion of alleles identical-by-state (IBS) > 0.95, and individuals with abnormally high levels of heterozygosity (false discovery rate < 0.01) with the "check.marker"

function in GENABEL (Aulchenko et al. 2007). We also excluded SNPs that were out of Hardy-Weinberg equilibrium using the "check.marker" function, setting the "p-level" option to 0.0001. These conditions resulted in all 50 individuals and 85,291 SNPs being retained for single-SNP GWA mapping. We adjusted for population structure in this analysis by including the first 14 axes of genetic variation generated from a PCA of the genomic kinship matrix (14 axes is the number that describe a significant amount of genetic variation in our sample, see Results).

Co-localization of regions associated with color in the two species

Because we found SNPs mapping to LG 8 to have the largest mean PIP in both T. cristinae and T. podura (see Results), we tested whether this pattern is expected by chance using permutation tests. The purpose of this analysis was to determine the probability of co-localization of SNPs with high PIPs to LG 8 while accounting for: (1) the genomic distribution of SNPs in our data set and (2) the distribution of PIPs observed for these SNPs. We therefore randomly permuted PIPs (without replacement) 10,000 times for both the T. podura and T. cristinae SNP data sets. During this permutation procedure the number and location of SNPs along each linkage group was maintained. We then calculated the proportion of permuted data sets for which LG 8 had the largest mean PIP in both species as our null expectation.

Dominance relationships at candidate loci

We next determined dominance relationships at the T. podura candidate SNPs identified by GWA mapping by calculating the ratio of dominant to additive effects of

alleles at each of these SNPs (for parallel analysis in T. cristinae see Comeault et al. 2015). Because color phenotypes are discrete and unambiguously scored (Fig. 1), each green individual was assigned a score of 0 and each melanistic individual a score of 1. Dominance effects (d) are calculated as the difference between the mean phenotype of heterozygotes and half difference between the mean phenotypes of the two homozygous genotypes. Additive effects (a) were calculated as half the phenotypic difference between the mean phenotype of the two homozygous genotypes. The ratio d/a represents the deviance of the phenotypes of heterozygotes from those expected under additivity (Burke et al. 2002; Miller et al. 2014). The expected value of d/a for additive alleles is 0 while completely dominant or recessive alleles will be 1 or -1. Here we follow previous conventions (Burke et al. 2002; Miller et al. 2014) and classify alleles as being dominant if d/a is greater than 0.75, recessive if d/a is less than -0.75, partially dominant or partially recessive if d/a is between 0.75 and 0.25 or -0.75 and -0.25, respectively, and additive if d/a is between -0.25 and 0.25.

Linkage disequilibrium between candidate SNPs and within candidate genomic regions

To quantify levels of LD for candidate genomic regions identified by GWA mapping, we computed genotypic correlations (r^2) for the regions spanned by all candidate SNPs mapping to LG 8 of the T. cristinae genome (i.e., the entire region between the 'left-most' and 'right-most' SNP on this LG, considering a linear genomic organization). We focused on LG 8 because this linkage group contained the strongest evidence for containing variants associated with color phenotypes in both species (see Results). We were carried out all LD analyses described below in parallel for T. podura

359 and T. cristinae using SNPs that passed the same filters described for those used in
360 single-SNP GWA mapping. For T. cristinae we used a previously published data set used
361 to identify candidate SNPs associated with color (Comeault et al. 2015) with the same
362 filtering applied to the T. podura data set. Prior to LD analyses in T. cristinae we
363 randomly down-sampled the number of individuals to match that of T. podura (i.e. 19
364 green and 31 melanistic individuals). All LD analyses were carried out using the "r2fast"
365 function of the GENABEL R library (Aulchenko et al. 2007)

366 Following filtering we computed LD between each candidate SNP (all pairwise
367 comparisons), the candidate genomic region spanning all LG 8 candidate SNPs, regions
368 on LG 8 that did not contain candidate SNPs (hereafter "non-candidate region"), and the
369 genome as a whole. Within candidate and non-candidate regions we retained a single
370 SNP per sequence read (i.e., 100 bp) as to not inflate estimates of LD due to mapped
371 sequences containing multiple SNPs. Following this procedure, we calculated r^2 between
372 all SNPs located on the same scaffold for each scaffold within a given region. We
373 restricted LD comparisons to SNPs found on the same scaffold because we were
374 interested in localized LD and the absolute distance between SNPs on different scaffolds
375 of the current draft of the T. cristinae (v0.3) genome is unknown. To estimate
376 'background' levels of LD within the genome we randomly sampled 1000 SNPs from
377 across the genome (i.e., using all LGs) and calculated LD for all pairwise comparisons.

378 To determine whether levels of LD between the candidate SNPs, within candidate
379 regions, and within non-candidate regions were greater than null genomic expectations,
380 we compared the proportion of pairwise LD comparisons for a given class of SNPs with
381 median LD of the random genomic sample of 1000 SNPs using binomial tests. The

genomic expectation for this analysis is that 50% of LD comparisons within a given class will be below and above median genomic LD.

In addition to quantifying LD within defined genomic regions, we measured the decay of LD with distance for each of the 13 linkage groups of the T. cristinae genome by computing the mean and 99% empirical quantile of r^2 as a function of the distance between SNPs. Measurements of LD were binned into 100 bp bins depending on the distance between the two SNPs used to calculate LD (e.g., estimates of LD for all SNPs 301 to 400 bp apart were binned into one bin).

Results

Quantifying variation in color

Within T. podura the green and melanistic phenotypes differ with respect to RG and GB color ($F_{1,40} = 158.92$, $P < 0.001$; $F_{1,40} = 126.66$, $P < 0.001$) but not luminance ($F_{1,40} = 3.76$, $P = 0.06$). Within T. cristinae the color phenotypes differ in RG color, GB color, and luminance ($F_{1,600} = 1050.90$, $P < 0.001$; $F_{1,600} = 52.07$, $P < 0.001$, respectively). Comparing color phenotypes between species reveal that melanistic T. podura do not differ from melanistic T. cristinae in GB color ($F_{1,82} = 1.68$, $P = 0.20$) but have significantly different RG color ($F_{1,82} = 4.371$, $P = 0.04$) and luminance ($F_{1,82} = 29.05$, $P < 0.0001$). Green T. podura differ from green T. cristinae in RG color, GB color, and L ($F_{1,558} = 25.14$, $P = 0.004$; $F_{1,558} = 44.28$, $P < 0.001$; $F_{1,558} = 41.53$, $P < 0.001$, respectively).

Despite some difference in color between T. podura and T. cristinae, both green and melanistic color phenotypes broadly overlap in RG – GB color space and the Euclidean distances between similarly colored individuals of each species were much less than the Euclidean distances between differently colored individuals within species (mean [SE] Euclidean distance between T. podura and T. cristinae having the same color = 0.193 [0.0011] and between differently colored T. podura = 0.377 [0.0050] or T. cristinae = 0.501 [0.0006]; Fig. 1b). Therefore, while there are slight differences in the color phenotypes of T. podura compared to those of T. cristinae, similar color phenotypes cluster tightly in phenotypic space and are more similar to each other than to differently colored individuals of their own species ($U = 315{,}985{,}777$, $P < 0.0001$; Fig. 1b).

Genetic structure within the T. podura sample

To test for potential genetic structure within our sample of 50 T. podura, we carried out hierarchical Bayesian modeling and PCA on the imputed genotype matrix. DIC increased with the number of clusters in hierarchical Bayesian model ran with K = 1 – 4 and the best model was K=1 (Table S1). When models were run with K > 1, we did not observe any distinct clustering of individuals based on color phenotype (Fig. S1). Principal Components Analysis of genotype likelihoods identified 14 axes that describe a significant amount of genetic variation based on a threshold of $Q^2 > 0.05$ (Table S2). Together, these 14 PCs explained a cumulative 53.59% of the variation in genotypes and PC1 accounted nearly half (26.59%) of this variation. Binomial regressions of color phenotype against PC scores revealed that only two of the 14 PCs (PC4, and PC7) explain a significant amount of variation in color phenotypes (Table S2); however, these

PCs each account for a small fraction of total genetic variation in our data set (2.44% and 2.04%, respectively). Taken together, these results indicate that there is no major axis of genetic variation correlated with color phenotype. Nonetheless, all GWA mapping analyses we describe below implement methods to correct for minor levels of genetic structure among individuals (see Methods).

Genetic control of T. podura color phenotypes estimated through GWA mapping

Hyperparameters estimated from BSLMMs indicate that color variation in T. podura is controlled by a simple genetic architecture with 97% of phenotypic variation being explained by genotype and 94% of this explained variation being due to only two SNPs with measurable phenotypic effects (median estimates; Fig. 2 for complete posterior distributions). Similar results were obtained for T. cristinae with 95% of phenotypic variation in color being explained by genotype and 95% of this explained variation being due to 7 SNPs with measurable phenotypic effects (median estimates; Fig. 3; Comeault et al. 2015).

Two SNPs in the T. podura data set were identified as having measurable effects on color phenotypes in > 10% of BSLMM iterations (i.e., PIPs > 0.10; blue points in Fig. 3b). Both of these SNPs map to LG 8 of the T. cristinae genome: one at position 10972 of scaffold 1806, 13.6 kb from the nearest gene annotation and the second at position 349343 of scaffold 284, 4.3 kb from the nearest gene annotation (see Supplementary File 1 for InterPro or GO annotations for each predicted gene located on these two scaffolds and the candidate scaffold identified by single-SNP GWA mapping [results presented below]). The PIPs of these SNPs are 0.295 and 0.102, their model-averaged estimates of

β are 9.92 and 4.25, respectively, and melanistic alleles are recessive to green alleles (d/a = -1 and -0.95, respectively; Fig. 4).

Cross-validation analyses revealed that hyperparameter estimates and effect sizes reported above are unlikely due to chance. For example, BSLMM analyses repeated using randomly permuted phenotypic data sets did not recover any SNPs having measurable effects on phenotypic variation in > 10% of model iterations and confidence intervals for hyperparameter estimates spanned nearly the entire interval [0,1], indicating a strong genetic signal within our observed data (Fig. S2). This strong genetic signal was also confirmed by our ability to accurately predict the phenotype of individuals from genotypic information alone (prediction accuracy = 96.8%).

Single-SNP GWA mapping in T. podura identified two SNPs that are associated with color phenotypes that also map to the T. cristinae genome assembly (significance level: $P < 0.000001$; Table S3). One of these SNPs mapped to LG 8 (scaffold 1154; position 30072) and the second to LG 10 (scaffold 380; position 189546). Dominance relationships between alleles at these two SNPs mirror those of the SNPs identified by multi-SNP mapping with green alleles being dominant to melanistic alleles (d/a = -0.95 and -0.94, respectively; Fig. 4). Because LG 8 has the highest density of candidate SNPs identified by both multi-locus and single-SNP GWA mapping in both T. podura and T. cristinae (Table S3 for results from T. podura and Comeault et al. 2015 for results for T. cristinae), we focus our remaining analyses on this LG.

Co-localization of regions associated with color in the two species

We explored whether SNPs associated with color variation were statistically concentrated on LG 8 by calculating the mean PIP for SNPs within each LG. Previous work in T. cristinae suggests that SNPs associated with color were concentrated on LG 8 (Comeault et al. 2015). We confirmed this result (Fig. 3a). In T. podura, mean PIP also differs significantly across the 13 LGs (proportion test; $\chi^2 = 21731.33$, d.f. = 12, $P < 0.001$) and SNPs mapping to LG 8 had the highest mean PIP of all LGs (mean PIP = 0.000111; Fig. 3a). This mean PIP was nearly an order of magnitude greater than the LG with the second largest mean PIP (LG 1; mean PIP = 0.0000194). The two candidate scaffolds we identify for T. podura were both located on LG 8 and had mean PIPs of 0.0118 and 0.00161 (scaffolds 1806 and 284). Randomization tests showed that the co-localization of candidate SNPs in T. podura and T. cristinae to LG 8 is unlikely to happen by chance ($P = 0.0067$); however, within LG 8, candidate SNPs mapped to different scaffolds in the two species and we do not have the resolution to determine whether there is further co-localization of functional variation.

Linkage disequilibrium between candidate SNPs and within candidate genomic regions

Genotypes at LG 8 candidate SNPs are in strong LD within T. podura and within T. cristinae (median $r^2 = 0.81$ and 0.46, respectively), and all estimates of LD between candidate SNPs are greater than the 97.5% empirical quantile of genome-wide LD (Table 2). The higher LD observed between T. podura candidate SNPs could be due to there being fewer candidate SNPs identified for T. podura compared to T. cristinae (3 versus 26) and the fact that the T. podura candidate region spans a shorter genomic distance than

494 the T. cristinae candidate region (combined scaffold lengths of candidate region = 8.1 Mb
495 and 12.7 Mb, respectively).

496 Linkage disequilibrium within the candidate genomic region that contains
497 candidate SNPs in T. podura is 28.3% greater than median genomic LD ($P < 1 \times 10^{-15}$;
498 Table 2) while LD within the non-candidate region is not elevated relative to median
499 genomic LD (P = 1; Table 2). Linkage disequilibrium within the T. cristinae candidate
500 genomic region is also greater than median genomic LD, but even more strongly so than
501 in T. podura (i.e., 113.4% greater than mean genomic LD; $P < 1 \times 10^{-15}$; Table 2). This
502 large difference in LD within the 'candidate' versus 'background' regions was observed
503 despite the T. cristinae candidate region spanning 12,739 Kb (versus 8,135 Kb in T.
504 podura), containing roughly twice as many SNPs as the T. podura candidate region (1171
505 and 499 SNPs, respectively), and the average mean-distance between SNPs contained on
506 candidate scaffolds being roughly equal (112 [SD = 67] Kb in T. cristinae and 109 [94]
507 Kb in T. podura). In contrast to T. podura, LD within the non-candidate region of LG 8
508 in T. cristinae is also elevated ($P < 1 \times 10^{-15}$; Table 2). Linkage disequilibrium is
509 therefore elevated within the candidate region on LG 8 in both species, however this LD
510 is more pronounced, and extends across a longer genomic distance, in T. cristinae
511 compared to T. podura.

512 Supporting this finding, the decay of LD with distance was the same for each
513 linkage group in the T. podura sample, with LD falling to genomic background levels
514 within ~100 bp (Fig. 5). By contrast, in T. cristinae LD within LG 8 remains elevated
515 over larger genomic distances when compared to the genomic background (Fig. 5).
516

Discussion

Our results show that similar color phenotypes of T. podura and T. cristinae largely overlap in two-dimensional color space, with strong divergence between color morphs within species (Fig. 1b). In addition to the similarities we observe at the phenotypic level, we show that color phenotypes in T. podura and T. cristinae share at least three aspects of genetics. First, color phenotypes in both species are controlled by major effect loci (Fig. 2). Second, dominance relationships of alleles associated with color phenotypes are the same between these two species, with green alleles dominant to melanistic alleles (Comeault et al. 2015). Third, the same LG is implicated in each species, with genotype – phenotype associations co-localizing to LG 8. These results generate the testable hypothesis that the same gene (or group of genes) might control color in these two species. Future work is required to test this hypothesis, for example using fine scale mapping and analyses of synteny. Such tests could allow interesting parallels (or differences) to be drawn with other species, such as Heliconius butterflies, where genetic variation affecting aposematic color phenotypes found in multiple species has been shared through introgression (Dasmahapatra et al. 2012; Wallbank et al. 2016). Below we discuss the implications of our current findings, including those that do not rely on resolving the causal variants affecting color, along with additional questions that could be resolved by elucidating such variants.

Implications of genetic architecture for the response to selection

539 Important insights into the evolutionary process can be gained through an
540 understanding of quantitative aspects of the genetics of traits involved in adaptation and
541 speciation (Rausher & Delph 2015). As we describe in the methods of this manuscript,
542 multi-locus GWA mapping using BSLMMs provides advantages over single-SNP GWA
543 analyses because it provides estimates of three hyperparameters that quantitatively
544 describe aspects of genetics of traits while accounting for uncertainty in the specific
545 SNPs (and genes) causally associated with phenotypic variation (Zhou & Stephens 2012;
546 Zhou et al. 2013). These hyperparameters – namely the number of genetic regions
547 underlying phenotypic variation, the 'polygenic' component of phenotypic variation, and
548 the amount of phenotypic variation explained by SNPs with measurable effects on
549 phenotypic variation – can be useful in helping predict the phenotypic and genetic
550 response to selection. Our results predict that selection acting on color in populations of
551 T. podura and T. cristinae will result in strong divergence at the genetic regions
552 underlying those color phenotypes. Moreover, patterns of LD suggest that selection
553 acting on color phenotypes in T. podura could have less of an effect on neighboring sites
554 in the genome than in T. cristinae, because LD within the genomic region controlling
555 color is low in T. podura when compared to T. cristinae.
556 LD affects the genomic response to selection and can be generated by several
557 mechanisms. For example, elevated LD can represent regions of reduced recombination
558 (e.g., due to structural variation such as chromosomal re-arrangement; Lowry & Willis
559 2010) or positive, correlated, or epistatic selection (e.g., Kim & Nielsen 2004). These are
560 not mutually exclusive mechanisms because selection can favor structural rearrangements
561 that capture multiple alleles that positively affect fitness (Kirkpatrick & Barton 2006;

562 Feder et al. 2013). In T. cristinae, the mechanisms generating high LD on LG 8 are
563 unknown, but the size of the region affected (28.25% of this linkage group) hints at the
564 possibility of a large-scale inversion polymorphism. In T. podura, the genomic extent and
565 magnitude of LD within the candidate region is less than in T. cristinae, suggesting a lack
566 of structural variation, more ancient structural variation (i.e., allowing more time for
567 recombination), or recent, but weaker, selection (Table 2). Future work could usefully
568 test these explanations for variation in LD in these and other Timema species.
569 Dominance relationships at the locus that controls color in the studied species will
570 result in melanistic alleles being hidden from selection in heterozygous individuals. This
571 will have two general effects on the evolutionary response to selection: (1) recessive
572 melanistic alleles will be maintained within populations when they are maladaptive
573 longer than green alleles and (2) dominant green alleles will be able to respond to
574 selection more quickly than melanistic alleles when found at low frequencies in a
575 population. In T. podura the melanistic phenotype, to our knowledge, is fixed within
576 populations living on Adenostoma (Sandoval & Nosil 2005), suggesting that there is
577 strong selection acting against the green phenotype on Adenostoma. This idea is
578 supported by predation experiments that have shown that green T. podura are more
579 heavily depredated than melanistic T. podura on Adenostoma, while the opposite is true
580 on Ceanothus (Sandoval & Nosil 2005). Another explanation for the lack of green
581 individuals within Adenostoma populations is that the green allele has never reached
582 these populations. This however seems unlikely based on the geographic proximity of T.
583 podura populations found on either host (i.e., scale of a few kilometers) and high rates of
584 gene flow among adjacent populations of other species of Timema at similar or even

585 larger scales (Nosil et al. 2012). Given the T. podura population analyzed for this study
586 was from Ceanothus, it is surprising that we find green alleles at a much lower frequency
587 than melanistic alleles (Fig. 5). A combination of factors could contribute to the higher
588 frequency of melanistic alleles we observe in the population of T. podura studied here,
589 including recent colonization, unmeasured sources of selection favoring melanistic
590 individuals (differential survival measured by Sandoval & Nosil 2005 was based on
591 short-term predation by a single predator: Western scrub jays [Aphelocoma californica]),
592 the ability of melanistic alleles to hide from selection in heterozygotes, or high rates of
593 directional gene flow from Adenostoma to Ceanothus.
594 Influences of genetics on evolution have been shown in T. cristinae (Comeault et
595 al. 2015) and other systems (Rosenblum et al. 2010). For two species of lizard living on
596 the white sands of New Mexico (Sceloporus undulatus and Aspidoscelis inornata),
597 Rosenblum et al. (2010) showed that dominance relationships between derived 'white'
598 alleles were dominant to 'brown' alleles at the melanocortin receptor 1 locus (Mc1R) in
599 S. undulatus but recessive in A. inornata. These differences in dominance relationships
600 underlie different patterns in the segregation of genetic variation within populations of
601 these lizards living in white-sand environments. Moreover, this example helps illustrate
602 how understanding the genetic basis of phenotypic variation can help us understand how
603 selection structures genetic and phenotypic variation in natural populations. In turn,
604 genetic architecture itself can evolve, as might occur for dominance relations in
605 Heliconius butterflies (Le Poul et al. 2014). The results we present here will help inform
606 such predictions in populations of Timema and can be used to develop a better

607 understanding of speciation through integrating data describing links between
608 phenotypes, genotypes, and fitness.
609
610 Conclusions and future directions
611 While a quantitative understanding of the genetic basis of color in T. cristinae and
612 T. podura helps generate predictions regarding patterns of genetic differentiation,
613 identifying the causal alleles (and mutations) controlling these color phenotypes would
614 facilitate a better understanding of the evolutionary history of this variation (e.g.,
615 Colosimo et al. 2005; Linnen et al. 2009; Wallbank et al. 2016). For example, do color
616 phenotypes represent an ancestral polymorphism segregating within populations that may
617 have been differentially and independently sorted in the different species? While a
618 phylogeny does exist for Timema (Sandoval et al. 1998), green/melanistic-like color
619 polymorphisms are pervasive across species (Crespi & Sandoval 2000), making it
620 difficult to infer the ancestral color (or colors) of this group. If color alleles are
621 segregating from ancestral variation, Timema color phenotypes could share similarities
622 with lateral armor plates in stickleback where low-plated alleles at the Ecotdysplasin
623 locus (Eda) have been re-used during adaptation to fresh-water environments from
624 standing genetic variation segregating in marine populations (Schluter & Conte 2009).
625 Such examples would suggest a bias towards the recurrent evolution of the same color
626 phenotypes across different environments. Alternatively, color phenotypes could be the
627 result of independent evolution occurring at different sites in the same locus or in
628 different loci (Steiner et al. 2009). If the same locus or type of mutation (e.g. cis-
629 regulatory mutations) is involved in the evolution of color in Timema, this could suggest

630 a role of mutational biases in influencing evolutionary trajectories. Streisfeld & Rausher
631 (2011) showed that the evolution of floral pigment intensity is biased towards mutations
632 occurring in transcription factors while the evolution of floral hue is biased towards
633 mutations occurring in coding regions of pathway genes. In light of these examples,
634 identifying causal variants affecting color in Timema would help to inform key debates in
635 molecular evolution, such as whether constraints exist in the genetic changes leading to
636 adaptation (Stern & Orgogozo 2009), and the extent to which genes involved in
637 adaptation have pleiotropic effects (Rennison et al. 2015).
638 The recent increase in our understanding of the genetic basis of adaptive traits in
639 Timema stick insects (Comeault et al. 2014, 2015), genomic resources in this system
640 (Soria-Carrasco et al. 2014), and genome editing methods in general (Bono et al. 2015),
641 could help to facilitate the discovery of the specific gene or genes underlying these
642 phenotypes.
643

Acknowledgements

We thank J. Wolf, R. Snook, K.E. Delmore, members of the D. Matute lab, two anonymous reviewers and the associate editor A. Sweigart for helpful comments and/or discussion that improved previous versions of this manuscript. This research was supported by the European Research Council (Grant R/129639 to P.N.) and the Natural Sciences and Engineering Research Council of Canada (PGS-D3 to AAC). All authors declare no conflicts of interest. In addition to public repositories, all data are available from the authors upon request.

652 **References:**

653 Abdi H, Williams LJ (2010) Principal component analysis. WIREs Computational
654 Statistics, **2**, 433–459.

655 Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image Processing with ImageJ.
656 Biophotonics International, **11**, 36–42.

657 Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for
658 genome-wide association analysis. Bioinformatics, **23**, 1294–1296.

659 Barton NH (2000) Genetic hitchhiking. Philosophical Transactions of the Royal Society
660 B: Biological Sciences, **355**, 1553–1562.

661 Berg JJ, Coop G (2014) A Population Genetic Signal of Polygenic Adaptation. PLoS
662 Genetics, **10**, e1004412.

663 Beuttell K, Losos JB (1999) Ecological morphology of Caribbean Anoles. Herpetological
664 Monographs, **13**, 1–28.

665 Bono JM, Olesnicky EC, Matzkin LM (2015) Connecting genotypes, phenotypes and
666 fitness: harnessing the power of CRISPR/Cas9 genome editing. Molecular Ecology,
667 **24**, 3810–3822.

668 Burke JM, Tang S, Knapp SJ, Rieseberg LH (2002) Genetic analysis of sunflower
669 domestication. Genetics, **161**, 1257–67.

670 Charlesworth B (1992) Evolutionary rates in partially self-fertilizing species. The
671 American Naturalist, **140**, 126–148.

672 Colosimo PF, Hosemann KE, Balabhadra S et al. (2005) Widespread parallel evolution in

673 sticklebacks by repeated fixation of Ectodysplasin alleles. Science, **307**, 1928–1933.

674 Comeault AA, Flaxman SM, Riesch R et al. (2015) Selection on a genetic polymorphism
675 counteracts ecological speciation in a stick insect. Current Biology, **25**, 1975–1981.

676 Comeault AA, Soria-Carrasco V, Gompert Z et al. (2014) Genome-wide association
677 mapping of phenotypic traits subject to a range of intensities of natural selection in
678 Timema cristinae. The American Naturalist, **183**, 711–727.

679 Conte GL, Arnegard ME, Peichel CL, Schluter D (2012) The probability of genetic
680 parallelism and convergence in natural populations. Proceedings of the Royal
681 Society B: Biological Sciences, **279**, 5039–5047.

682 Core Team R (2013) R: A language and environment for statistical computing. R
683 Foundation for Statistical Computing, **1**.

684 Crespi BJ, Sandoval CP (2000) Phylogenetic evidence for the evolution of ecological
685 specialization in Timema walking-sticks. Journal of Evolutionary Biology, **13**, 249–
686 262.

687 Dasmahapatra KK, Walters JR, Briscoe AD et al. (2012) Butterfly genome reveals
688 promiscuous exchange of mimicry adaptations among species. Nature, **487**, 94–98.

689 Dobson AJ (2002) Introduction to generalized linear models. Chapman & Hall/CRC
690 Press.

691 Ellegren H, Smeds L, Burri R et al. (2012) The genomic landscape of species divergence
692 in Ficedula flycatchers. Nature, **491**, 756–760.

693 Endler JA (2012) A framework for analysing colour pattern geometry: adjacent colours.

694 Biological Journal of the Linnean Society, **107**, 233–253.

695 Feder JL, Flaxman SM, Egan SP, Comeault A a., Nosil P (2013) Geographic mode of
696 speciation and genomic divergence. Annual Review of Ecology, Evolution, and
697 Systematics, **44**, 73–97.

698 Freedman ML, Reich D, Penney KL et al. (2004) Assessing the impact of population
699 stratification on genetic association studies. Nature Genetics, **36**, 388–393.

700 Gompert Z, Lucas LK, Buerkle CA et al. (2014) Admixture and the organization of
701 genetic diversity in a butterfly species complex revealed through common and rare
702 genetic variants. Molecular Ecology, **23**, 4555–4573.

703 Haldane JBS (1927) A Mathematical Theory of Natural and Artificial Selection, Part V:
704 Selection and Mutation. Mathematical Proceedings of the Cambridge Philosophical
705 Society, **23**, 838.

706 Hohenlohe PA, Bassham S, Etter PD et al. (2010) Population genomics of parallel
707 adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics, **6**,
708 e1000862.

709 Jones FC, Grabherr MG, Chan YF et al. (2012) The genomic basis of adaptive evolution
710 in threespine sticklebacks. Nature, **484**, 55–61.

711 Kim Y, Nielsen R (2004) Linkage disequilibrium as a signature of selective sweeps.
712 Genetics, **167**, 1513–1524.

713 Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation.
714 Genetics, **173**, 419–434.

715 Krzanowski W (1987) Cross-validation in principal component analysis. Biometrics, **43**,
716 575–584.

717 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature
718 Methods, **9**, 357–359.

719 Li H, Handsaker B, Wysoker A et al. (2009) The Sequence Alignment/Map format and
720 SAMtools. Bioinformatics, **25**, 2078–2079.

721 Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE (2009) On the origin and spread of an
722 adaptive allele in deer mice. Science, **325**, 1095–1098.

723 Lowry DB, Willis JH (2010) A widespread chromosomal inversion polymorphism
724 contributes to a major life-history transition, local adaptation, and reproductive
725 isolation. PLoS Biology, **8**, e1000500.

726 Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genetical
727 Research, **23**, 23–35.

728 Miller CT, Glazer AM, Summers BR et al. (2014) Modular skeletal evolution in
729 sticklebacks is controlled by additive and clustered quantitative trait loci. Genetics,
730 **197**, 405–420.

731 Nadeau NJ, Whibley A, Jones RT et al. (2012) Genomic islands of divergence in
732 hybridizing Heliconius butterflies identified by large-scale targeted sequencing.
733 Philosophical Transactions of the Royal Society B: Biological Sciences, **367**, 343–
734 353.

735 Nosil P, Gompert Z, Farkas TE et al. (2012) Genomic consequences of multiple

736 speciation processes in a stick insect. Proceedings of the Royal Society B: Biological
737 Sciences, **279**, 5058–5065.

738 Parchman TL, Gompert Z, Mudge J et al. (2012) Genome-wide association genetics of an
739 adaptive trait in lodgepole pine. Molecular Ecology, **21**, 2991–3005.

740 Poelstra JW, Vijay N, Bossu CM et al. (2014) The genomic landscape underlying
741 phenotypic integrity in the face of gene flow in crows. Science, **344**, 1410–1414.

742 Le Poul Y, Whibley A, Chouteau M et al. (2014) Evolution of dominance mechanisms at
743 a butterfly mimicry supergene. Nature Communications, **5**, 5644.

744 Price AL, Patterson NJ, Plenge RM et al. (2006) Principal components analysis corrects
745 for stratification in genome-wide association studies. Nature Genetics, **38**, 904–909.

746 Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: hard
747 sweeps, soft sweeps, and polygenic adaptation. Current Biology, **20**, R208–15.

748 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
749 multilocus genotype data. Genetics, **155**, 945–959.

750 Rausher MD, Delph LF (2015) When does understanding phenotypic evolution require
751 identification of the underlying genes? Evolution, **69**, 1655–1664.

752 Rennison DJ, Heilbron K, Barrett RDH, Schluter D (2015) Discriminating selection on
753 lateral plate phenotype and its underlying gene, Ectodysplasin, in Threespine
754 Stickleback. The American Naturalist, **185**, 150–156.

755 Rosenblum EB, Römpler H, Schöneberg T, Hoekstra HE (2010) Molecular and
756 functional basis of phenotypic convergence in white lizards at White Sands.

757 Proceedings of the National Academy of Sciences of the United States of America,
758 **107**, 2113–7.

759 Sandoval CP (1994) Differential visual predation on morphs of Timema cristinae
760 (Phasmatodeae:Timemidae) and its consequences for host range. Biological Journal
761 of the Linnean Society, **52**, 341–356.

762 Sandoval C, Carmean DA, Crespi BJ (1998) Molecular phylogenetics of sexual and
763 parthenogenetic Timema walking-sticks. Proceedings of the Royal Society B:
764 Biological Sciences, **265**, 589–595.

765 Sandoval CP, Nosil P (2005) Counteracting selective regimes and host preference
766 evolution in ecotypes of two species of walking-sticks. Evolution, **59**, 2405–2413.

767 Schluter D, Conte GL (2009) Genetics and ecological speciation. Proceedings of the
768 National Academy of Sciences of the United States of America, **106**, 9955–9962.

769 Skotte L, Korneliussen TS, Albrechtsen A (2013) Estimating individual admixture
770 proportions from next generation sequencing data. Genetics, **195**, 693–702.

771 Soria-Carrasco V, Gompert Z, Comeault AA et al. (2014) Stick insect genomes reveal
772 natural selection's role in parallel speciation. Science, **344**, 738–742.

773 Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of
774 model complexity and fit. Journal of the Royal Statistical Society: Series B
775 (Statistical Methodology), **64**, 583–639.

776 Steiner CC, Römpler H, Boettger LM, Schöneberg T, Hoekstra HE (2009) The genetic
777 basis of phenotypic convergence in beach mice: similar pigment patterns but

778 different genes. Molecular Biology and Evolution, **26**, 35–45.

779 Stern DL, Orgogozo V (2009) Is Genetic Evolution Predictable? Science, **323**, 746–751.

780 Streisfeld MA, Rausher MD (2011) Population genetics, pleiotropy, and the preferential
781 fixation of mutations during adaptive evolution. Evolution, **65**, 629–642.

782 Wallbank RWR, Baxter SW, Pardo-Díaz C et al. (2016) Evolutionary Novelty in a
783 Butterfly Wing Pattern through Enhancer Shuffling. PLoS Biology, 1–16.

784 Yeaman S (2015) Local adaptation by alleles of small effect. The American Naturalist,
785 **186**.

786 Zhou X, Carbonetto P, Stephens M (2013) Polygenic modeling with Bayesian sparse
787 linear mixed models. PLoS Genetics, **9**, e1003264.

788 Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association
789 studies. Nature Genetics, **44**, 821–4.

790

791

Tables and Figure Legends

Table 1. Summary of the ecology of the two species of Timema stick insects included in this study.

species	Location	host plants considered here	selection on color phenotypes
T. cristinae	Coastal western Transverse Range, Southern California	C. spinosus, A. fasciculatum	Balance of multiple sources of selection, often within host species, maintains polymorphism.
T. podura	San Bernardino, Santa Rosa and San Jacinto Mountains, Central Southern California	C. leucodermis, A. fasciculatum	Divergent selection acting between host plants.

Table 2. Linkage disequilibrium, calculated as genotypic correlations (r^2) between pairs of SNPs. Median r^2 and confidence intervals are reported for groups of SNPs sampled at different genomic scales (see methods for details). The confidence interval reported for candidate SNPs represents the minimum and maximum LD observed between any pair of candidate SNPs, while for all other SNP classes confidence intervals are reported as 95% equal tail-probability intervals. "P > genome" represents the probability that the proportion of LD within a given class of SNP with r^2 greater than median genomic LD was observed by chance.

a) T. podura

genomic scale	r^2	P > genome
candidate SNPs	0.8116 (0.7715-0.8931)	< 0.00001
candidate region	0.0179 (0.0000-0.2423)	< 0.00001
non-candidate region	0.0144 (0.0000-0.2083)	1
LG 8	0.0150 (0.0000-0.2162)	< 0.00001
genome	0.0139 (0.0000-0.1992)	n/a

b) T. cristinae

genomic scale	r^2	
candidate SNPs	0.4602 (0.0619-1.0000)	< 0.00001
candidate region	0.0323 (0.0001-0.4281)	< 0.00001
non-candidate region	0.0189 (0.0000-0.2602)	< 0.00001
LG 8	0.0211 (0.0000-0.2941)	< 0.00001
genome	0.0151 (0.0000-0.2054)	n/a

Figure 1. (a) Representative images of melanistic and green phenotypes for T. podura and T. cristinae. (b) Phenotypic position of 42 T. podura and 602 T. cristinae in RG – GB color space. Hashed lines in 'b' represent the range of RG (horizontal line) and GB (vertical line) values for T. podura phenotypes and the size of the symbols is proportional to an individual's luminance.

Figure 2. Posterior probability distributions of parameter estimates describing the genetic architecture for color in T. podura (red lines) and T. cristinae (blue lines). The total amount of phenotypic variation explained by genotype (PVE) and the proportion of that variation that can be explained by SNPs with non-zero affects on phenotypic variation (PGE) are given, along with the number of SNPs in our data set that have non-zero affects on phenotypic variation (N-SNP).

Figure 3. Genome wide association mapping of SNPs associated with color variation in T. podura and T. cristinae. (a) Mean posterior inclusion probabilities (PIPs) for SNPs mapping to each of the 13 T. cristinae linkage groups (LGs). Error bars represent one standard error. b) Manhattan plots showing associations between SNPs and color phenotypes in T. podura and T. cristinae. SNPs significantly associated with color in the single-SNP analyses ($P < 0.00001$) are shown as red points and the LG 8 candidate SNPs identified by multi-locus GWA mapping are shown as solid blue points.

Figure 4. Dominance relationships between alleles at candidate SNPs associated with color variation in T. podura. Mean phenotype (bars) and 95% binomial confidence

intervals (vertical lines; computed using the 'binconf' function in R) are shown for genotypes at each of four candidate SNPs identified by multi-SNP (left two panels) and single-SNP (right two panels) GWA mapping. The location of the candidate SNPs are given above each panel: linkage group (lg) and scaffold (scaf) are given before the position (in bp). Ratios above each bar report the number of melanistic individuals that have that genotype over the total number of individuals with that genotype. Green individuals are scored as "0" and melanistic individuals as "1". Based on allele frequencies within this sample of individuals, segregation of genotypes at each SNP did not significantly differ from Hardy-Weinberg expectations (all $P > 0.1$).

Figure 5. Decay of LD with distance in both T. podura and T. cristinae. The median (solid lines) and 99% quantile (dashed lines) of r^2 is plotted for SNPs binned by the distance between them. Distances were binned every 100 bp from 1 to 1000bp. Each linkage group is plotted independently and LG 8 is highlighted in red.