Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect

Dorothea Lindtke[1,2], Kay Lucek[2,3], Víctor Soria-Carrasco[2], Romain Villoutreix[2], Timothy E. Farkas[4], Rüdiger Riesch[5], Stuart R. Dennis[6], Zach Gompert[7] & Patrik Nosil[2]

[1] Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary AB, T2N 1N4, Canada
[2] Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK
[3] Department of Environmental Sciences, University of Basel, Hebelstrasse 1, CH–4056 Basel, Switzerland
[4] Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06369, USA
[5] School of Biological Sciences, Royal Holloway, University of London, Egham Hill, Egham, TW20 0EX, UK
[6] Department of Aquatic Ecology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, CH–8600 Dübendorf, Switzerland
[7] Department of Biology, Utah State University, Old Main Hill, Logan, Utah 84322, USA

Keywords: color polymorphism, heterozygote advantage, chromosomal inversion, adaptation, divergence, gene flow

Corresponding author: Dorothea Lindtke
Department of Biological Sciences
University of Calgary
2500 University Drive NW
Calgary AB, T2N 1N4
Canada
dorothea.lindtke@ucalgary.ca
(403) 220–5361

Running title: Chromosomal variants in a stick insect

Abstract

How polymorphisms are maintained within populations over long periods of time remains debated, because genetic drift and various forms of selection are expected to reduce variation. Here, we study the genetic architecture and maintenance of phenotypic morphs that confer crypsis in *Timema cristinae* stick insects, combining phenotypic information and genotyping-by-sequencing data from 1360 samples across 21 populations. We find two highly divergent chromosomal variants that span megabases of sequence and are associated with color polymorphism. We show that these variants exhibit strongly reduced effective recombination, are geographically widespread, and probably diverged millions of generations ago. We detect heterokaryotype excess and signs of balancing selection acting on these variants through the species' history. A third chromosomal variant in the same genomic region likely evolved more recently from one of the two color variants and is associated with dorsal pattern polymorphism. Our results suggest that large-scale genetic variation associated with crypsis has been maintained for long periods of time by potentially complex processes of balancing selection.

Introduction

Crypsis is a widespread trait that reduces the risk of prey or predators from becoming initially detected when in plain sight, for example through background matching (Stevens & Merilaita, 2009). It is a central element in prey-predator interactions, and its selective advantage can be substantial and involve reduced metabolic costs and higher survival probability. However, much remains unknown about the details of the genetic basis of crypsis, and the evolutionary processes involved in its origin and maintenance (Stevens & Merilaita, 2009; Skelhorn & Rowe, 2016). Here we study the genetic basis and evolutionary processes that maintain different cryptic morphs within populations of a stick insect.

Although some species only exist as a single, highly optimized cryptic form, others are polymorphic. New morphs might recurrently evolve but be transient through their subsequent replacement, for example if predators initially avoid unfamiliar morphs (predator wariness; Mappes *et al.*, 2005). Polymorphisms might also be maintained by gene flow-selection balance (e.g., King & Lawson, 1995; Hoekstra *et al.*, 2004), negative assortative mating (e.g., Tuttle *et al.*, 2016; Hedrick *et al.*, 2016), or various mechanisms of balancing selection (Hedrick *et al.*, 1976). Balancing selection can result from: (i) variable microhabitats that can induce spatially or temporally varying selection (Charlesworth & Charlesworth, 2010), (ii) frequency-dependent selection, for example based on predator behavior (apostatic selection; Clarke, 1969; Allen, 1988; Bond & Kamil, 1998), or (iii) heterozygote advantage. Despite this, changing selection pressures, allele turnover, or genetic drift in finite populations are expected to eventually remove existing variants (Charlesworth, 2006; Charlesworth & Charlesworth, 2010). Consequently, balancing selection is often regarded a common, albeit predominantly short-term, mechanism for maintaining variation (Asthana *et al.*, 2005; Charlesworth, 2006; Fijarczyk & Babik, 2015).

The nature of selection on color polymorphisms can also affect their genetic architecture.

For example, heterozygotes that exhibit intermediate trait values can be selected against if selection is divergent between discrete environments. Accordingly, many polymorphic species show dominant trait expression for color patterns, often realized as dominance hierarchies in cases with more than two morphs (Clarke & Sheppard, 1972; Joron *et al.*, 2011; Johannesson & Butlin, 2017). Given the potentially high selective pressure against maladapted color morphs, finely tuned genetic architectures that result in strongly dominant trait expression can evolve, with heterozygotes being phenotypically similar or even identical to one homozygote (Le Poul *et al.*, 2014). Alternatively, phenotypically different heterozygotes might not be selected against, for example if intermediate niches exist, or if effectively no intermediates occur because one of the homozygotes is lethal (Hedrick, 2012; Le Poul *et al.*, 2014; Kuepper *et al.*, 2016; Tuttle *et al.*, 2016).

If more than one locus is required to generate alternative morphs and if recombinant phenotypes are selected against, genetic architectures that keep multiple adaptive alleles in linkage disequilibrium (LD) can evolve (Charlesworth & Charlesworth, 1975; Yeaman, 2013; Kirkpatrick & Barrett, 2015; Charlesworth, 2016). This can for instance be achieved by tight physical linkage, genetic modifiers of recombination, or structural changes such as chromosomal rearrangements. Chromosomal rearrangements have the advantage that they reduce recombination only in heterokaryotypes, thus facilitating purging of deleterious mutations through normal recombination in homokaryotypes (Otto & Lenormand, 2002; Kirkpatrick, 2010). However, chromosomal rearrangements can also reduce fitness in heterokaryotypes (Rieseberg, 2001; Kirkpatrick & Barton, 2006; Faria & Navarro, 2010), a situation that will act against the maintenance of polymorphisms within populations.

Such selective processes acting on color polymorphisms and the genetic architecture of the traits involved have been investigated in a variety of organisms, providing evidence consistent with the wide range of ecological and genetic outcomes described above (e.g., Cook, 1998; Nachman *et al.*, 2003; Joron *et al.*, 2011; Richards *et al.*, 2013; Kunte *et al.*,

2014; Kuepper *et al.*, 2016; Lamichhaney *et al.*, 2016; Tuttle *et al.*, 2016; van't Hof *et al.*, 2016). However, less is known about the extent to which cryptic polymorphisms differ from the frequently studied colorful outcomes of sexual selection or mimicry (Stevens & Merilaita, 2009). Crypsis differs from mimicry as morph frequencies are independent of the population dynamics or evolution of a model species (Endler, 1981), and some morphs might become fixed by directional selection if they go undetected by predators (Bond & Kamil, 1998). Thus, processes other than apostatic selection or predator wariness might drive the maintenance of variation in species exhibiting cryptic phenotypes. In addition, it remains unclear how often cryptic polymorphisms are maintained within populations over long periods of time by balancing selection (Gray & McKinnon, 2007), versus being an ephemeral outcome of environmental changes (e.g., van't Hof *et al.*, 2016) or a balance between gene flow and selection (e.g., King & Lawson, 1995; Hoekstra *et al.*, 2004).

Here, we address these unresolved issues in understanding the evolution and maintenance of cryptic color morphs by studying populations of the stick insect *Timema cristinae*. This species has three color and color-pattern morphs that are adapted to different microhabitats (details below). Combining genotyping-by-sequencing (GBS) and phenotypic data from hundreds of samples across 21 populations we investigate the maintenance and genetic architecture of this polymorphism.

Study system

The genus *Timema* comprises 21 described species of herbivorous stick insects in southwestern North America (Sandoval *et al.*, 1998; Vickery & Sandoval, 2001; Law & Crespi, 2002; Nosil *et al.*, 2002). All *Timema* are wingless and rely on crypsis as protection against avian predators while resting on their host plants (Sandoval, 1994a). Their body coloration approximate colors of stems, fruits, needles or leaves of their hosts, and most

species exhibit color or color-pattern polymorphisms (Sandoval *et al.*, 1998; Crespi & Sandoval, 2000; Comeault *et al.*, 2015, 2016). In two species (*T. cristinae* and *T. podura*), variation in color or color-pattern has been experimentally linked to fitness variation in the face of visual predation, supporting the adaptive nature of the polymorphisms and their role in crypsis (Sandoval, 1994a,b; Nosil, 2004; Sandoval & Nosil, 2005; Nosil & Crespi, 2006).

We focus here on *T. cristinae*, which is endemic to chaparral vegetation in a mountainous region surrounding Santa Barbara, California. *T. cristinae* is polymorphic for two distinct body color morphs found within populations: a common green morph matching coloration of leaves, and a rarer melanistic (i.e., dark gray or red) morph approximating coloration of stems or fruits of their host plants, or soil (Fig. 1a; Sandoval, 1994a,b; Comeault *et al.*, 2015). These color morphs will be referred to as 'green' and 'melanistic' hereafter. Compared to green morphs, melanistic *T. cristinae* are more cryptic to avian predators on the woody tissue of their host plants but less so on leaves, and show higher resistance to fungal infections (Comeault *et al.*, 2015).

T. cristinae primarily uses *Adenostoma fasciculatum* and *Ceanothus spinosus* as host plants. As an adaptation to the differently shaped and colored leaves of these host species, the green morph exhibits an additional polymorphism for the presence or absence of a white longitudinal dorsal stripe ('green-striped' and 'green-unstriped' pattern morphs hereafter). The stripe visually divides the body into two slim parts and increases survival on the narrow needle-like leaves of *Adenostoma*. In contrast, the white stripe is conspicuous on the broad leaves of *Ceanothus* and reduces survival on this host (Sandoval, 1994a; Sandoval *et al.*, 1998; Nosil, 2004; Nosil & Crespi, 2006).

Pattern morph frequencies vary across the landscape according to gene flow-selection balance between the often patchily distributed host plants (Sandoval, 1994b). Thus, the green-striped morph tends to be more common on *Adenostoma* and the green-unstriped morph more common on *Ceanothus* (Sandoval, 1994a,b; Nosil *et al.*, 2002). By contrast, the

melanistic morph, which does not express the stipe, occurs at ∼11% within populations of either host plant and, although slightly more common in hot and dry climates, does not vary substantially in frequency across the landscape (Comeault *et al.*, 2015). Thus, intra-population polymorphism of color morphs is likely not strongly driven by gene flow between populations differing in morph frequency.

Previous studies suggest that the genetic architecture for either color or pattern involves one or a few loci of large effect, and that color and pattern loci are distinct but physically linked. These studies further showed that the green variant is fully dominant to melanistic, while the stripe is partially recessive (Sandoval, 1994a,b; Comeault *et al.*, 2015). Thus, intermediates for color do not exist but incompletely or faintly green-striped *T. cristinae* are occasionally observed in the field (Sandoval, 1994a), reflecting the imperfect dominance for pattern (Comeault *et al.*, 2015), or recombination among multiple loci controlling this trait.

Despite this background, numerous fundamental aspects of the evolution of these polymorphisms remain unresolved, which we investigated here. For example, whether different cryptic morphs have existed over extended periods of time is unknown, as is the potential contribution of different mechanisms of balancing selection or negative assortative mating in maintaining variation. It is also unknown whether the genetic architecture of cryptic traits involves reduced recombination between potentially many loci or is more aligned with a single locus. Our results show that the color polymorphism in *T. cristinae* is not recent and involves a large genomic region under balancing selection that almost completely lacks genetic exchange between divergent variants. Nevertheless, heterokaryotypes are in excess, possibly caused by heterokaryotype advantage selection. We discuss our results in light of general issues concerning the long-term maintenance of adaptive polymorphism.

Materials and Methods

Samples

We analyzed data from 1360 *T. cristinae* from 21 populations throughout the species range, which were sampled in spring 2013 and preserved in ethanol (Fig. 2; Table S1). To study in detail the mechanisms maintaining color and pattern morphs within populations and their genomic outcomes, we first focused analyses on a single site for which we had a large sample size (n = 435) and where *Adenostoma* and *Ceanothus* host plants co-occur. This population, named N1 (N34°31.034', W119°47.790'), comprises an area of about 50 x 70 m and has not previously been studied. Using sweep nets we collected from N1 a total of 94 and 341 *T. cristinae* on 32 *Adenostoma* and 64 *Ceanothus* plant individuals, respectively. We then tested if our findings can be replicated by re-analyzing a second population with a large sample size (FHA), using data from 600 previously published samples (Comeault *et al.*, 2015). *Adenostoma* dominates this site and all *T. cristinae* were collected from this host. We detected major chromosomal variants associated with color morphs in both populations. We thus investigated if these were geographically widespread using previously published data from 19 additional populations sampled on either *Adenostoma* or *Ceanothus* throughout the species range (5–20 individuals per population, 325 in total; Fig. 2; Table S1; Riesch *et al.*, 2017).

Phenotype characterization

Using digital photographs, we scored dorsal color as 'melanistic' or 'green', and dorsal pattern in green individuals as 'green-striped' or 'green-unstriped'. Because photographs were taken of most, but not all, collected *T. cristinae*, sample sizes were lower than for genetic data: 409 mostly sexually-immature *T. cristinae* from N1, 588 adult samples from

FHA (Comeault *et al.*, 2015), and 305 adult samples from 18 additional populations (Riesch *et al.*, 2017). For detailed analyses in N1 and FHA, we further classified phenotypes as 'green-incomplete' if a dorsal stripe was present but faint or not developed along the full body length. Depending on the markedness of the stripe, these phenotypes were scored as either 'green-striped' or 'green-unstriped' otherwise.

For FHA, we also analyzed a number of previously published continuous measurements on sexually-mature individuals (Comeault *et al.*, 2015; Riesch *et al.*, 2017): percent of the dorsal body area striped (% striped), body length (BL), and the following six continuous traits on color channels: lateral green-blue (latGB), lateral red-green (latRG), lateral luminance (i.e., brightness; latL), dorsal green-blue (dorGB), dorsal red-green (dorRG), and dorsal luminance (dorL). We could not obtain these measurements for N1 because standardized photographs of adult individuals were not taken.

Genotyping-by-sequencing (GBS)

We obtained genomic DNA from all 435 specimens from N1 and prepared individually barcoded restriction-site associated DNA libraries using protocols as for the other previously published samples (Comeault *et al.*, 2015; Riesch *et al.*, 2017). Libraries were single-end sequenced on three Illumina HiSeq2000 lanes at the National Center for Genome Research (Santa Fe, New Mexico, USA). We filtered raw sequences and used Bowtie2 (Langmead & Salzberg, 2012) to map reads to the *T. cristinae* reference genome (Soria-Carrasco *et al.*, 2014; Riesch *et al.*, 2017), which comprises 13 linkage groups (LGs), likely corresponding to the chromosomes of the species ($2n = 25/26$, X0/XX; Schwander & Crespi, 2009). We called variants using Samtools and Bcftools (Li, 2011), and after discarding variants where less than 90% of samples were covered, we retained 304 168 bi-allelic SNPs with mean coverage depth per SNP per individual of ∼5x.

We re-analyzed sequence data for 600 individuals from FHA (Comeault *et al.*, 2015, NCBI BioProject PRJNA284835), after excluding two individuals with low sequence coverage. We called and filtered variants as above and retained 384 611 bi-allelic SNPs with mean coverage depth per SNP per individual of ∼7x.

We combined sequences from 325 *T. cristinae* that were sampled from 19 additional populations distributed across the species range (Riesch *et al.*, 2017, NCBI BioProject PRJNA356885) with 20 individuals from each N1 and FHA. Samples from N1 and FHA were chosen such that all main karyotypes (below) were included, and served as references to determine whether the same karyotypes were present across the species range. We used settings for sequence filtering, mapping and variant calling as above and retained 626 854 bi-allelic SNPs with mean coverage depth per SNP per individual of ∼5x. Further details are provided in the Supplemental Information.

Identification of genomic clusters

Principal component analysis (PCA):

We conducted PCA separately for populations N1, FHA, and the combined data set of 21 populations. Because missing genotype information can affect principal components, we first re-called SNPs for PCA analyses, requiring at least 99% of individuals to be covered, retaining 62 542, 168 020, and 99 008 SNPs, respectively. To account for genotype uncertainty, we used a Bayesian model and Markov Chain Monte Carlo (MCMC) to obtain joint posterior probabilities for genotypes and allele frequencies given the genotype likelihoods estimated by BCFTOOLS, along with Hardy-Weinberg priors, as in past work (Nosil *et al.*, 2012; Gompert *et al.*, 2012, 2014). We further excluded variants with a minor allele frequency (MAF) $< 1\%$, and randomly selected SNPs to achieve at least 100 bp distance among variants, retaining 11 751, 30 297, and 8 758 SNPs for N1, FHA, and the

combined data set, respectively.

We collapsed posterior genotype probabilities into a single value per individual and locus (i.e., posterior mean of alternative allele dosage, ranging from zero to two), centered values for each SNP by subtracting the mean over all individuals, and conducted a PCA on the genotype covariance matrix on the centered but unscaled values using the *prcomp* function in R (R Core Team, 2016). Visual inspection of PCA scatter plots revealed three striking genotypic clusters on the first two PC axes (Figs. S1a and S2a). To investigate this clustering in more detail, we sequentially removed 48 and 37 genome-wide PCA outliers as in Price *et al.* (2006) for populations N1 and FHA, respectively. Briefly, PCAs were visually inspected after each iteration of outlier removal and remaining samples were subjected to a new iteration until individuals peripheral of the main clusters were eliminated (Figs. S1 and S2; more details on the process of outlier removal and discussion of outliers in Supplemental Information). PCA applied separately to each LG revealed that genomic clustering could be attributed to variation on LG8 only, and that the three main clusters further split on LG8 into a total of six clusters (Fig. 1b; Figs. S1 to S5). We defined these PCA clusters for N1 and FHA by first grouping individuals by k-means clustering on the first 10 PC axes computed from SNPs on LG8 only (*kmeans* function in R, with 10 initial centers). We then obtained assignment probabilities for individuals per cluster by applying linear discriminant analysis of the first 10 PC axes as explanatory variables and cluster assignment as grouping factor (*lda* function in R, MASS library; leave-one-out cross-validation), and retained samples with at least 80% assignment probability for any cluster for further analyses (Fig. 1c; Tables S2 and S3; Fig. S4b). We then tested for an association between PCA clusters and phenotypic morphs using χ^2 tests with the *chisq.test* function in R and significance values computed by Monte Carlo simulation with 100 000 replicates.

Model-based cluster assignment:

Given the distinctive arrangement of PCA clusters on only one LG, their association with color and pattern morphs, and the known dominance relationships for color and pattern loci (Sandoval, 1994a,b; Comeault *et al.*, 2015), we suspected that clusters were caused by divergent chromosomal variants existing as homo- and heterokaryotypes (i.e., heterokaryotypes are located in-between the homokaryotypes in the PCA plot). We thus predicted that by assigning diploid genomic ancestry to each locus and individual, hetero- or homozygous ancestries would prevail for genomic regions causing these clusters (e.g., three main PCA clusters could then be described by diploid combinations of two ancestry clusters 'melanistic' and 'green', Fig. 1d; and six PCA clusters by diploid combinations of three ancestry clusters 'melanistic', 'green-striped' and 'green-unstriped', Fig. 1c).

We used the software STRUCTURE to obtain locus-specific estimates of ancestry for SNPs on LG8 (the site-by-site output from the linkage model; Pritchard *et al.*, 2000; Falush *et al.*, 2003). To test if individuals from different PCA clusters represent homozygous and heterozygous combinations of two main ancestry clusters of which one is further subdivided, we set the number of ancestry clusters to $k = 2$ or $k = 3$. To obtain karyotype assignments for tests of Hardy-Weinberg Equilibrium (HWE) and phenotypic differences among karyotypes (below), we also ran STRUCTURE using all individuals from population N1 and FHA (i.e., including PCA outliers; 435 and 600 samples) using SNPs on three adjacent scaffolds on LG8 that showed a particularly strong signal of genetic clustering ('scaffolds 931, 318, and 1440' hereafter). We set $k = 2$ as we were interested in karyotype estimates for the main axis of variation (i.e., 'melanistic' versus 'green' variants). We repeated this analysis for the combined data set of 21 populations to test if variants are geographically spread and in HWE. Further details of preparation of STRUCTURE input files and settings are provided in the Supplemental Information.

Multi-locus genome-wide association mapping

To map color and pattern traits, we used population FHA where we had a better record of phenotypic traits and a larger number of samples than for N1. Although these traits have previously been mapped in FHA (Comeault *et al.*, 2015, 2016; Riesch *et al.*, 2017), we re-mapped them here using the sets of SNPs and individuals used in our other analyses to make results compliant (i.e., using the same version of the *T. cristinae* reference genome and excluding PCA outliers). We excluded individuals with ambiguous phenotype data (i.e., classified differently by two researchers), and scored both color and pattern as binary traits. We restricted mapping to SNPs assigned to linkage groups and excluded SNPs with MAF $< 1\%$, retaining 180 512 SNPs and 552 samples for color, and 180 506 SNPs and 498 samples for pattern mapping (only green individuals were used for the latter). We mapped traits using Bayesian sparse linear mixed models (BSLMMs) with the probit model implemented in the software GEMMA (Zhou *et al.*, 2013), as in previous work (Comeault *et al.*, 2015, 2016; Riesch *et al.*, 2017, details in Supplemental Information).

Population genomic statistics

We found that the six PCA clusters were indeed associated with homo- and heterokaryotypic combinations of three chromosomal variants (see Results). To obtain information regarding the evolutionary processes affecting these variants and the time scales involved, we computed and compared different population genomic statistics. This was done across the genome within or between PCA clusters, depending on the prediction being tested. Specifically, we estimated relative and absolute between-cluster divergence (F_{ST} and D_{xy}, respectively), within-cluster nucleotide diversity (π), and a measure of between-cluster, intra-locus LD ($\mathrm{Z_g}$; Storz & Kelly, 2008). In addition, we surveyed chromosomal variants for signals of recent positive selection ('selective sweeps') by

estimating extended haplotype homozygosity within and between clusters (iES and Rsb; Tang *et al.*, 2007). All statistics were computed in non-overlapping 20-kb windows. We provide below an overview of the logic behind our analyses, with details of how the statistics were calculated provided in the Supplemental Information.

Balancing selection might target a single locus, multiple loci, or structural genomic changes such as chromosomal inversions. The genomic processes of balancing selection and their expected outcomes arising at or linked to inversion breakpoints are highly similar to those expected for a single or multiple linked selected loci. Loci subject to long-term varying selection are expected to show elevated nucleotide diversity between alleles sampled from different subpopulations (or here, different chromosomal variants) relative to diversity within them (Hudson & Kaplan, 1988; Charlesworth *et al.*, 1997; Kelly & Wade, 2000; Nordborg & Innan, 2003; Storz & Kelly, 2008). In contrast, new alleles that were rapidly driven to high or intermediate frequencies by selection will show reduced diversity compared to ancestral alleles or neutral loci (Sabeti *et al.*, 2002; Voight *et al.*, 2006). Similarly, for a sufficiently old inversion polymorphism maintained by balancing selection, alleles of sites linked to inversion breakpoints are expected to show longer coalescent times (i.e., elevated π; Wakeley, 2008) when sampled from heterokaryotypes compared to either genome-wide expectations or alleles sampled from any homokaryotype (Navarro *et al.*, 2000; Guerrero *et al.*, 2012). In contrast, the evolution of a new inversion will eliminate diversity within inversion homokaryotypes, which will only slowly recover through genetic exchange with the standard type. As recombination is more likely in the center of the inversion, reduced diversity will remain near the breakpoints until new mutations accumulate (Navarro *et al.*, 1997, 2000; Guerrero *et al.*, 2012).

In addition, increased levels of LD are expected to build up at and closely linked to selected loci or inversion breakpoints that are maintained by balancing selection (Charlesworth *et al.*, 1997; Kelly, 1997; Storz & Kelly, 2008; Peischl *et al.*, 2013; Wallace

et al., 2013). LD can also extend over larger genomic regions following a selective sweep, but is expected to decay over time with increasing physical distance from single selected loci (Sabeti *et al.*, 2002; Slatkin, 2008), and at a slower rate for multiple linked selected loci or inversions where recombination is suppressed (Navarro & Barton, 2002; Wallace *et al.*, 2013; Peischl *et al.*, 2013).

We were interested in detecting genetic regions subject to balancing selection (including putative inversion breakpoints that might be associated with divergent chromosomal variants), and in investigating whether chromosomal variants show indications for recent or ancient evolution. We thus examined the genome for heterogeneity in Z_g, D_{xy}, and Rsb between homokaryotypic clusters, and compared π in homokaryotypic clusters to π in their corresponding heterokaryotypic cluster along the genome. To facilitate comparison among the different statistics we used, we defined 'high-differentiation scaffolds' for each pair of homokaryotypic cluster as scaffolds with mean F_{ST} above or equal the 97.5% quantile of all scaffolds from the 13 LGs. Scaffolds 931, 318, and 1440 (above) are a subset of these high-differentiation scaffolds (Fig. 3 shows their positions on LG8).

To further investigate the mechanisms and the history of balancing selection, we computed additional statistics for whole populations, irrespective of chromosomal variants (i.e., genetic clusters). We contrasted statistics that are informative regarding balancing selection in the sampled generation (i.e., HWE), and those indicative of such selection in the recent or distant past (i.e., LD and Tajima's D; Garrigan & Hedrick, 2003; Hedrick, 2006). Specifically, we computed Burrow's composite measure of within-population LD (Δ; Weir, 1979) between pairs of SNPs, and Tajima's D (Tajima, 1989) in non-overlapping 20-kb windows. Increased LD can indicate that multi-locus balancing selection acted consistently during the recent history of populations (Navarro & Barton, 2002; Garrigan & Hedrick, 2003; Hedrick, 2006). Positive values of Tajima's D can arise if mutations accumulated independently among polymorphic variants, which is expected if balancing selection acted

over extended periods of time in the distant past (Simonsen *et al.*, 1995; Garrigan & Hedrick, 2003; Hedrick, 2006). As positive Tajima's D can potentially also result from directional positive selection that differs from the standard full-sweep model (Przeworski *et al.*, 2005), we used linear regression to test whether increased Tajima's D is associated with decreased π or increased iES in any of the homokaryotypic clusters, indicating that recent positive selection on one chromosomal variant might have caused an excess of intermediate-frequency alleles in the whole population and increased Tajima's D. Because Tajima's D or LD can equally be affected by recent population dynamics, we compared these statistics computed for different scaffolds on LG8 to genome-wide expectations.

Divergence dating

We estimated divergence time between chromosomal variants by two different methods. First, we used the program BEAST 2 (Bouckaert *et al.*, 2014), including previously published genetic data and divergence times of four related *Timema* species (Riesch *et al.*, 2017, NCBI BioProject PRJNA356405). We based estimations on scaffolds 931, 318, and 1440, or on high-differentiation scaffolds common to all three pairwise combinations of homokaryotypic clusters, without assuming a chromosomal inversion. Second, we used Approximate Bayesian Computation (ABC), based on scaffolds 931, 318, and 1440, assuming the presence of a chromosomal inversion (details in Supplemental Information).

Tests for HWE and heterokaryotype excess

To further test for balancing selection in the sampled generation, we tested for heterokaryotype excess relative to HWE. We classified individuals as homo- or heterokaryotypes of the two main chromosomal variants 'melanistic' and 'green', given their diploid genomic ancestry on scaffolds 931, 318, and 1440 assigned by STRUCTURE and

$k = 2$ (above). To define karyotypes we used thresholds for STRUCTURE admixture proportions (q) that best delimited clusters for each data set ($0.3 < q < 0.7$, $0.32 < q < 0.68$, and $0.38 < q < 0.62$ for N1, FHA, and the combined data set, respectively; Fig. S6). We used the obtained karyotype counts to apply an exact test for HWE (Wigginton *et al.*, 2005), using R code from `http://csg.sph.umich.edu/abecasis/Exact/r_instruct.html`. We further measured the direction of deviation from HWE using the fixation index $F = (H_E - H_O)/H_E$, where H_E and H_O are the expected and observed heterokaryotype frequencies.

Mating preference models

We used a modeling approach to test to what extent heterokaryotype advantage selection or mating preferences likely contribute to the observed frequencies of main karyotypes. We considered two models, one of negative assortative mating and one of universal mating advantage of the melanistic morph (Comeault *et al.*, 2015). We assumed that m and G variants correspond to a single locus with two alleles associated with melanistic and green morphs and controlling mating preferences, and with selection acting on one or both homokaryotypes. We first computed equilibrium genotype proportions for various strengths of mating preference and selection, and then obtained the probability of sampling the genotype counts observed in N1 or FHA from these proportions (details in Supplemental Information).

Test for phenotypic differences among karyotypes

We used linear modeling to investigate whether phenotypic traits (i.e., body length and six continuous color traits measured for individuals from FHA) are associated with karyotype. Individuals were classified as homo- and heterokaryotypes for the 'green' and 'melanistic'

variants as before. We determined whether adding either karyotype or binary color state as explanatory variable improved models by analysis-of-deviance and by difference in Akaike's Information Criterion (ΔAIC; including sex and % striped as covariates; details in Supplemental Information).

Results

We first report results of genomic analyses in population N1, and then complement them with analyses involving phenotype measurements, or karyotype assignments across the species range. Unless stated otherwise, results reported in the main article were obtained from N1 as only this population had all three chromosomal variants present as homokaryotypes in adequate quantities. Comparable results from FHA are provided in the Supplemental Information.

Phenotypic morphs are associated with highly divergent genetic clusters

The first axis of variation in the genome-wide PCA analysis showed three striking genotypic clusters that were almost perfectly associated with color morph (i.e., two green clusters and one melanistic cluster; x-axis in Fig. S1f). This clustering by color was explained by LG8 only (Fig. S3), and PCA restricted to variants on LG8 revealed additional clustering on the second axis of variation that was associated with pattern morphs (Fig. 1b).

We assigned individuals to the six conspicuous clusters on the first two PC axes using a model-free clustering algorithm. These clusters showed a pronounced non-random association with phenotypic morphs ($\chi^2_{(15,N=357)} = 647$, p-value = 0.00001, Cramér's V = 0.78; Fig. 1c; Table S2). We suspected that the observed structure was caused by three

divergent chromosomal variants associated with color and pattern and existing in homo- and heterokaryotypic combinations. We refer to these variants and resulting clusters as follows: (i) three chromosomal variants m, U, and S (i.e., 'melanistic', 'green-unstriped', and 'green-striped'); (ii) six PCA clusters mm, UU, SS, mU, mS, and US, resulting from all six possible homo- or heterokaryotypic combinations of these chromosomal variants (Fig. 1c); (iii) three main karyotypes mm, mG, and GG, resulting from STRUCTURE analysis with $k = 2$ (i.e., 'green-unstriped' and 'green-striped' variants pooled as 'green'; Fig. 1d). We provide further support for our notations below.

Genetic differentiation between clusters

To more finely determine the genomic regions generating genetic clustering, we estimated genome-wide F_{ST} between pairwise combinations of homokaryotypic clusters (mm, UU, and SS). This revealed that F_{ST} between clusters ranged among 20-kb windows from values near zero to one, but that strong differentiation was almost entirely restricted to one region of ~13 megabases of sequence covering ~29% of LG8 (Fig. 3). Absolute divergence measured as D_{xy} (Fig. 4) and joint allele frequency spectra (Fig. S7) further confirmed strong differentiation and putatively evolutionary independence between m and U or m and S variants. The observed block-like pattern of high F_{ST} is reminiscent of the genomic outcome of a chromosomal inversion that suppresses effective recombination and genetic exchange among chromosomal variants (e.g., Cheng *et al.*, 2012; Corbett-Detig & Hartl, 2012; Lamichhaney *et al.*, 2016; Tuttle *et al.*, 2016), although future work is needed to test this hypothesis directly.

Genetic clusters represent distinct chromosomal variants

STRUCTURE analyses with $k = 2$ confirmed that the three main clusters on PC axis 1 represent homo- and heterozygous combinations of two ancestry clusters (i.e., m and G; Fig. 1d; Figs. S8 and S9). By defining $k = 3$, we could also support that genomic clusters on PC axis 2 represent homo- and heterozygous combinations of a further subdivision of G identified with $k = 2$ (Figs. S10 and S11). This is consistent with our expectation of three chromosomal variants m, U, and S. Their associations with phenotype are in agreement with previous results on dominance and linkage within and among color and pattern loci mapped to LG8 (Comeault *et al.*, 2015). We further confirmed that our second population, FHA, showed patterns of genomic clustering and differentiation that were very similar to N1, although the UU karyotype was not identified and likely not strongly represented (Table S3; Figs. S2, S4, S5 and S12), which prevented comparisons among all three homokaryotypic clusters.

Multi-locus genome-wide association mapping

Multi-locus genome-wide association mapping in population FHA confirmed that candidate SNPs for color and pattern are located within the boundaries defining divergent chromosomal variants on high-differentiation scaffolds on LG8 (except one pattern candidate SNP on LG4; Fig. 3; Tables S4 to S7). However, the specific positions of candidate SNPs should be interpreted very cautiously given the particularly high levels of LD in this region (below; Figs. 4 and 5b; Fig. S13b). For this reason, we did not pursue further functional annotation of candidate SNPs.

Chromosomal variants are ancient and present throughout the species range

The high level of divergence among chromosomal variants suggests that they have coexisted for a sufficient amount of time to build up genetic differentiation. To investigate the evolutionary history of the three chromosomal variants in more detail, we determined several statistics for homokaryotypic (mm, UU, and SS) and heterokaryotypic clusters (mU, mS, and US) that are informative regarding divergence time and the processes of selection. We restricted our analyses to LG8 given that genetic clustering and association with phenotypic morphs was largely confined to this part of the genome.

We found that the genomic region of high F_{ST} between chromosomal variants also showed considerably elevated D_{xy} and $\mathrm{Z_g}$ between the melanistic variant (m) and either green variant (U or S), compared to genome-wide expectations (Fig. 4). This suggests the region was subject to varying selection between variants (Storz & Kelly, 2008). Similarly, π within mU and mS heterokaryotypes was elevated, while π within mm and UU homokaryotypes approached background levels of diversity, resembling the expected outcomes for an old inversion polymorphism maintained by balancing selection (Fig. 6; Navarro *et al.*, 2000; Guerrero *et al.*, 2012). Nevertheless, mm and UU homokaryotypes also showed deviations in π and Rsb along LG8, consistent with the effects of more recent differential selection (Figs. 4 and 6; Fig. S14).

In contrast with the pronounced differentiation between the melanistic and either green variant, when we compared the two green variants, U and S, we found that D_{xy} and $\mathrm{Z_g}$ were only slightly elevated compared to genome-wide expectations (Fig. 4). Increased haplotype homozygosity and significantly reduced levels of π within SS homozygotes further suggest that the green-striped variant experienced a considerable recent selective sweep (Figs. 4 and 6; Fig. S14). Our results thus support a recent evolution of the green-striped

variant, consistent with a young polymorphism or new chromosomal inversion (Navarro *et al.*, 2000; Guerrero *et al.*, 2012; DeGiorgio *et al.*, 2014).

We further evaluated the age of the chromosomal variants using BEAST 2 and ABC. BEAST 2 analyses estimated m and U variants to have split 13.5 or 8.0 million years (Ma) ago, based on scaffolds 931, 318, and 1440 or on high-differentiation scaffolds, respectively (95% highest posterior density intervals: 2.3–20.0 or 2.3–15.2 Ma; Fig. S15; one year corresponds to one generation in *T. cristinae*). By contrast, we estimated U and S variants to have split more recently, 2.7 or 1.8 Ma ago, based on the two sets of scaffolds, respectively (95% highest posterior density intervals: 0.6–5.7 or 0.7–3.3 Ma). When using ABC, meaningful time estimates were not possible due to a wide spread of the posterior distribution (median: 0.87 million generations, 2.5 and 97.5% quantiles: 0.018 and 39.3 million generations; Figs. S16 and S17). This might reflect uncertainty in parameter estimates affecting divergence time, or an old polymorphism has reached equilibrium and thus provides little information on divergence time under a simple mutation model.

A sufficiently old and balanced polymorphism might be spread through large parts of the species range. To test this expectation, we re-analyzed samples from 19 localities across the species distribution. Isolation-by-distance contributes to pronounced genetic divergence among *T. cristinae* populations (Nosil *et al.*, 2012; Riesch *et al.*, 2017), rendering analyses of differentiation between the only slightly differentiated U and S chromosomal variants difficult (Fig. S18). We thus considered here only the two main chromosomal variants (m and G) identified by STRUCTURE with $k = 2$. We found that m variants were indeed present in all populations at considerable and similar frequencies (mean 0.37, s.d. 0.12; Fig. 2; Table S1). This finding is consistent with geographically widespread balancing selection within populations, as opposed to gene flow-selection balance among divergent populations. It also suggests that balancing selection may have acted for an extended period of time, particularly given the low dispersal distance of *T. cristinae* (i.e., meters to dozens of meters

per generation; Sandoval, 2000) and the patchiness of the habitat.

Further consideration of mechanisms of balancing selection

We investigated the mechanisms and the history of balancing selection using additional population genetic parameters that are informative regarding balancing selection at different time scales. To determine the potential mechanisms of balancing selection in the current generation, we tested for deviations from HWE for the two main chromosomal variants m and G. In N1, we found that heterokaryotypes were in considerable excess relative to their expected frequency (15.5% more heterokaryotypes than expected; fixation index $F = -0.16$, p-value = 0.00119; Fig. 5a; Table S1). FHA showed a similar heterokaryotype excess (16.7% more than expected; $F = -0.17$, p-value = 0.00004; Table S1; Fig. S13a). Estimates from additional populations across the species range suggest that heterokaryotype excess could be widespread, although not necessarily ubiquitous (10 out of 19 populations showed $F < 0$; Table S1). Although low sample sizes preclude a definitive test for HWE in all 19 populations, heterokaryotype excess is clearly evident in both populations for which we had large sample sizes.

To examine if balancing selection acted in the past, we calculated within-population LD (Δ) and Tajima's D. We found increased Δ ranging over several hundreds of kb distance between SNPs for high-differentiation scaffolds, relative to other scaffolds (Fig. 5b; Fig. S13b). This is consistent with high levels of between-cluster, intra-locus LD (Z_g) determined above, which measures a different aspect of LD. Further, Tajima's D was elevated for high-differentiation scaffolds compared to other scaffolds on LG8 or genome-wide expectation (Fig. 5c; Fig. S13c; p-value = 5.076 x 10^{-15} or < 2.2 x 10^{-16}, Mann-Whitney U tests). We found no evidence that increased Tajima's D in high-differentiation scaffolds was caused by recent positive selection on any chromosomal

variant (Figs. S19 and S20). Our results thus indicate that balancing selection maintained the polymorphism on LG8 during the past. As an extended genomic region shows these signals of balancing selection, our results further indicate that selection targets many linked loci or a region of strongly reduced recombination (Kelly & Wade, 2000; Navarro & Barton, 2002; Nordborg & Innan, 2003).

Testing for potential causes of heterokaryotype excess

The observed heterokaryotype excess could arise through two main and potentially overlapping mechanisms: negative assortative mating or heterozygote advantage selection. Our data suggest that negative assortative mating between melanistic and green morphs is unlikely to have caused the heterokaryotypes excess. Specifically, the equilibrium frequency of the recessive color allele is expected to be ∼0.71 for various strengths of negative assortative mating (Hedrick *et al.*, 2016). However, we observe much lower frequencies of 0.33 and 0.36 for m variants in populations N1 and FHA, respectively (assuming here that m and G chromosomal variants are perfectly associated with color morphs, which is largely consistent with our results). Moreover, mating trials do not support negative assortative mating in *T. cristinae*, where, if anything, melanistic morphs have a universal mating advantage (Comeault *et al.*, 2015). We further tested which strengths of heterokaryotype advantage selection and mating preferences could explain the observed karyotype frequencies in N1 and FHA. A model for negative assortative mating and one for universal mating advantage of the melanistic morph both indicate that heterokaryotype advantage selection likely contributes generating the observed frequencies, although we cannot fully exclude alternative scenarios (Figs. S21 and S22).

Given some evidence for heterokaryotype advantage, we tested if heterokaryotypes differ from homokaryotypes in traits known to affect survival in *T. cristinae* (Nosil & Crespi,

2006). We found that karyotypic state had a minor but significant effect on body length and all continuous color traits tested (after Benjamini and Hochberg adjustment; Table S8; Fig. S23). The effect of karyotype remained significant with models only addressing variation between green morphs with mG versus GG karyotypes (Table S9). Future work is required to determine whether phenotypic differences among karyotypes affect fitness and contribute to heterokaryotype excess.

Discussion

T. cristinae exhibits three color and pattern morphs that are cryptic on different plant parts and on different plant species. The frequent co-occurrence of melanistic and green color morphs on the same host plants allowed us to address the putative duration and evolutionary mechanisms maintaining the cryptic polymorphism within populations of this species. Despite genetic drift and changing selection pressures being expected to eventually lead to the loss of existing variants (Charlesworth & Charlesworth, 2010), our results support that color morphs have been maintained over extended periods of time by balancing selection. We have revealed that the color polymorphism is associated with highly divergent chromosomal variants involving several megabases of sequence. Interestingly, our results suggest that heterokaryotype advantage might contribute to maintaining this chromosomal polymorphism. This is surprising because incomplete dominance or recombination in heterokaryotypes might result in maladaptive intermediate phenotypes that do not match either stems or leaves of either host plant. We here discuss four aspects of our results: (i) the genetic architecture of crypsis; (ii) the maintenance of polymorphisms through time; (iii) the mechanisms of that maintenance; and (iv) the implications for adaptation and speciation.

Genetic architecture of crypsis

The genetic architecture of cryptic color and pattern polymorphism in *T. cristinae* agrees with two main observations of the genetic basis of discrete color polymorphisms in a variety of organisms (Llaurens *et al.*, 2017). First, color and pattern exhibit dominance hierarchies (also see Sandoval, 1994a,b; Comeault *et al.*, 2015), in line with findings in other organisms (Clarke & Sheppard, 1972; Joron *et al.*, 2011; Le Poul *et al.*, 2014; Johannesson & Butlin, 2017). Second, in several species color polymorphisms were mapped to regions of reduced recombination such as chromosomal inversions or supergenes (Joron *et al.*, 2011; Richards *et al.*, 2013; Kunte *et al.*, 2014; Wellenreuther *et al.*, 2014; Kuepper *et al.*, 2016; Lamichhaney *et al.*, 2016; Tuttle *et al.*, 2016). Our results revealed that in *T. cristinae* color and pattern traits are associated with an extended genomic region, consistent with highly reduced recombination. Patterns of Z_g and π further indicate that recombination is reduced between different karyotypes relative to within them, consistent with the presence of a chromosomal inversion, as are the particular patterns of F_{ST} and π along the genome (Figs. 3, 4 and 6). Nevertheless, if selection targets many linked loci so that recombination among them is effectively reduced through low fitness of recombinants, similar outcomes are expected without an inversion (Kelly & Wade, 2000; Navarro & Barton, 2002; Nordborg & Innan, 2003). Future work is required that explicitly test for the presence of a chromosomal inversion, and whether different variants are the result of several linked inversions (e.g., Joron *et al.*, 2011) or evolved by rare recombination events in heterokaryotypes (e.g., Imsland *et al.*, 2012). Further, studies on the cytogenetics of the genus that extend previous work (Schwander & Crespi, 2009) and that determine fitness effects of crossover events in inversion heterozygotes are needed.

The agreement in the genetic architecture of crypsis in *T. cristinae* with that of color polymorphic traits in various other species suggests that architectures that prevent

formation of maladapted phenotypic intermediates without reducing gene flow genome-wide might be a common evolutionary outcome. The presumably opposing selective advantage of melanistic and green morphs on different plant parts and the seeming absence of positive assortative mating between them is theoretically expected to select for chromosomal rearrangements that reduce recombination between locally adapted alleles at multiple loci (Charlesworth & Charlesworth, 1975; Yeaman, 2013; Kirkpatrick & Barrett, 2015; Charlesworth, 2016). However, it remains to be determined whether chromosomal variants in *T. cristinae* indeed contain multiple genes controlling color or pattern, which mutations are causal, and whether additional traits adaptive to different plant parts map to the same genomic region.

Maintenance of polymorphisms through time

Our results are consistent with chromosomal variants in *T. cristinae* having been maintained by balancing selection through the recent and distant history of the species. LD is expected to decay or build up over tens to thousands of generations by recombination, gene flow, or genetic drift (Garrigan & Hedrick, 2003; Hedrick, 2012). Increased LD in high-differentiation scaffolds relative to the genomic background (Figs. 4 and 5b; Fig. S13b) indicates that recombination or gene flow was reduced, or drift increased in this genomic region. This is consistent with balancing selection acting during the recent history of populations, but can also be caused by a selectively neutral inversion polymorphism. However, in this latter case we would not expect the polymorphism to be present in all populations, which we observed here (Fig. 2; Table S1). Tajima's D is affected by mutation and selection, where a signal of balancing selection might require many thousands or millions of generations to be generated or lost, although the statistic can also be influenced by more recent population dynamics such as genetic drift and population structure (Simonsen *et al.*, 1995; Garrigan & Hedrick, 2003; Hedrick, 2012). As recent population

dynamics will affect Tajima's D genome-wide (i.e., not restricted to high-differentiation scaffolds), we conclude that past balancing selection enabling the accumulation of independent substitutions within chromosomal variants is more likely (Fig. 5c; Figs. S7 and S13c). Further, the particular patterns in π, D_{xy}, and Z_g along the genome between color variants are consistent with an old inversion polymorphism (Navarro *et al.*, 2000; Guerrero *et al.*, 2012; Peischl *et al.*, 2013) or long-term multi-locus balancing selection (Kelly & Wade, 2000; Navarro & Barton, 2002; Nordborg & Innan, 2003; Storz & Kelly, 2008).

The long-term maintenance of polymorphisms is considered to be probably unusual (Asthana *et al.*, 2005; Charlesworth, 2006; Fijarczyk & Babik, 2015). In *Drosophila melanogaster*, for example, inversion polymorphisms are commonly short-lived and frequently less than hundreds of thousands of years old (< 1 N_e generations; Andolfatto *et al.*, 1999, 2001; Corbett-Detig & Hartl, 2012). Similarly, polymorphic inversions in *Anopheles gambiae* were maintained for less than 11 000 years (< 2.7 N_e generations; White *et al.*, 2007, 2009). However, polymorphisms have also been shown to persist for millions of years in some species, such as *Drosophila pseudoobscura* (up to 2 Ma; Schaeffer, 2008; Wallace *et al.*, 2013), the ruff (*Philomachus pugnax*: 3.8 Ma; Lamichhaney *et al.*, 2016), or in the form of sex chromosomes (Charlesworth, 2016), and can even be shared across species boundaries (Wiuf *et al.*, 2004; White *et al.*, 2009; Leffler *et al.*, 2013; Novikova *et al.*, 2016). Our results for *T. cristinae* are compatible with the examples for old polymorphisms, and indicate that the forces of balancing selection have likely been strong and continuous over time to prevent the loss of this variation. As several *Timema* species related to *T. cristinae* are polymorphic for color, it will be interesting to test in future work whether color alleles pre-date speciation events. Alternatively, polymorphisms can be acquired by introgression from a related taxon (e.g., Besansky *et al.*, 2003; Feder *et al.*, 2003), which can falsely suggest their long-term maintenance within a species. We regard recent introgression as unlikely given that speciation events in the genus occurred millions

of generations ago and *T. cristinae* being geographically isolated from other *Timema* species (Law & Crespi, 2002; Riesch *et al.*, 2017), although we cannot exclude introgression from a now extinct species (e.g., Tuttle *et al.*, 2016).

Mechanisms of the maintenance of polymorphisms

We detected a pronounced excess of heterokaryotypes within several populations that cannot easily be explained by negative assortative mating or universal mating advantage of the melanistic morph alone (Fig. 5a; Table S1; Figs. S13a, S21 and S22). Instead, heterokaryotypes might have a fitness advantage over homokaryotypes, for example because of selective trade-offs (e.g., Johnston *et al.*, 2013), associative overdominance (Pamilo & Palsson, 1998; Charlesworth & Willis, 2009), improved crypsis resulting from differences in body color compared to homokaryotypes (Tables S8 and S9; Fig. S23), or a combination of these processes.

Although heterozygote advantage can constitute a simple mechanism of balancing selection, it remains controversial whether it maintains variants over extended periods of time (Clarke, 1979; De Boer *et al.*, 2004; Charlesworth & Charlesworth, 2010; Spurgin & Richardson, 2010; Sellis *et al.*, 2011; Hedrick, 2012). Indeed, few empirical examples exist where heterozygote advantage selection is considered to maintain polymorphisms (reviewed by Gemmell & Slate, 2006; Hedrick, 2006, 2011, 2012), often due to a mutant allele that confers improved fitness but is lethal in homozygotes. However, such a polymorphism is expected to be short-lived as it will be lost once a new allele evolves that is not associated with a fitness cost (Clarke, 1979; Charlesworth & Charlesworth, 2010; Hedrick, 2012).

Similarly, heterokaryotype excess involving highly differentiated chromosomal variants often includes lethality of one homokaryotype, where the polymorphism is commonly maintained by negative assortative mating (e.g., Wang *et al.*, 2013; Kuepper *et al.*, 2016;

669 Lamichhaney *et al.*, 2016; Tuttle *et al.*, 2016). However in *T. cristinae*, despite the
670 presumably long divergence time between color variants, both homokaryotypes are
671 represented. The system might however eventually transition to the more commonly
672 observed situation described above, for example through the accumulation of recessive
673 deleterious mutations in the rare variant.

674 In addition to heterokaryotype excess, additional processes of balancing selection
675 probably contribute to the maintenance of color polymorphism in *T. cristinae*. In
676 particular, the availability of micro-niches (i.e., stems and leaves) on each host plant likely
677 support the maintenance of two color morphs (Levene, 1953; Nagylaki, 2009). Universal
678 mating advantage of the melanistic morph might further prevent the stochastic or selective
679 loss of the less common melanistic variant (Fig. S22). Thus, although apostatic selection
680 and predator wariness are often considered important mechanisms maintaining
681 polymorphisms in species that use color traits as protection against predation (Clarke,
682 1979; Allen, 1988; Mappes *et al.*, 2005; Bond, 2007; Wellenreuther *et al.*, 2014), our results
683 suggest that the long-term maintenance of the cryptic polymorphism within *T. cristinae*
684 populations can instead be driven by several other, collectively acting forms of balancing
685 selection.

686 Implications for adaptation and speciation

687 Polymorphisms might also be lost through speciation, as divergent selection or reduced
688 recombination between distinct chromosomal variants can drive the evolution of
689 reproductive isolation (Coyne & Orr, 1998; Navarro & Barton, 2003; Butlin, 2005;
690 Kirkpatrick & Barton, 2006; Schluter & Conte, 2009; Faria & Navarro, 2010; Hugall &
691 Stuart-Fox, 2012; Charron *et al.*, 2014). The fine-scaled, temporally and spatially highly
692 heterogeneous habitats of *T. cristinae*, however, might instead promote the long-term

maintenance of polymorphisms (Gray & McKinnon, 2007; Svardal *et al.*, 2015; Gulisija & Kim, 2015), perhaps amplified by small local population sizes that can further delay speciation (Claessen *et al.*, 2008). Reproductively isolated small populations, if they arise, might also not persist as they are expected to rapidly accumulate deleterious mutations (Lynch *et al.*, 1995) and are less likely to hold the phenotypic variation necessary to withstand rapidly changing selection pressures (Nei *et al.*, 1975; Forsman & Wennersten, 2016).

In conclusion, our work indicates that several interacting mechanisms of balancing selection may maintain adaptive polymorphisms over extended periods of time, despite individual mechanisms often being regarded to maintain variation only short-term. It remains to be determined whether and how often processes of balancing selection other than apostatic selection or predator wariness, as suggested by our work, are important drivers in maintaining cryptic color polymorphisms in other species. Finally, the melanistic versus green color morphs of *T. cristinae* illustrate how long-term maintenance of adaptive polymorphisms in micro-niches might constitute an alternative evolutionary outcome to speciation, particularly in fine-scaled heterogeneous environments (e.g., Gray & McKinnon, 2007). Most broadly, our results show how population genomics can yield powerful insights into evolutionary processes and dynamics when combined with ecological data, ideally from multiple traits and across the species range.

Acknowledgments

We thank Aaron Comeault for assistance during data analysis and for discussions, Alex Buerkle for support with simulations, and Rosa Marin for drawings. We thank Subject Editor Brent Emerson and three reviewers for valuable comments on an earlier version of this manuscript. The work was funded by a grant from the European Research Council

(NatHisGen R/129639) to PN. The Royal Society of London provided salary support to PN via a University Research Fellowship. KL was supported by a Swiss National Science Foundation Early Postdoc Mobility Grant (P2BEP3_152103), VSC was supported by a Leverhulme Early Career Fellowship, and RR was supported by the Human Frontier Science Program.

References

Allen J (1988) Frequency-dependent selection by predators. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **319**, 485–503.

Andolfatto P, Depaulis F, Navarro A (2001) Inversion polymorphisms and nucleotide variability in *Drosophila*. *Genetical Research*, **77**, 1–8.

Andolfatto P, Wall J, Kreitman M (1999) Unusual haplotype structure at the proximal breakpoint of In(2L)t in a natural population of *Drosophila melanogaster*. *Genetics*, **153**, 1297–1311.

Asthana S, Schmidt S, Sunyaev S (2005) A limited role for balancing selection. *Trends in Genetics*, **21**, 30–32.

Besansky N, Krzywinski J, Lehmann T, *et al.* (2003) Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: Evidence from multilocus DNA sequence variation. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 10818–10823.

Bond A, Kamil A (1998) Apostatic selection by blue jays produces balanced polymorphism in virtual prey. *Nature*, **395**, 594–596.

Bond AB (2007) The evolution of color polymorphism: Crypticity, searching images, and apostatic selection. *Annual Review of Ecology Evolution and Systemtics*, **38**, 489–514.

Bouckaert R, Heled J, Kuehnert D, *et al.* (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Butlin R (2005) Recombination and speciation. *Molecular Ecology*, **14**, 2621–2635.

Charlesworth B, Charlesworth D (2010) *Elements of Evolutionary Genetics*. Roberts & Company Publishers.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research*, **70**, 155–174.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, **2**, 379–384.

Charlesworth D (2016) The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evolutionary Applications*, **9**, 74–90.

Charlesworth D, Charlesworth B (1975) Theoretical genetics of Batesian mimicry II. Evolution of supergenes. *Journal of Theoretical Biology*, **55**, 305–324.

Charlesworth D, Willis JH (2009) Fundamental concepts in Genetics: The genetics of inbreeding depression. *Nature Reviews Genetics*, **10**, 783–796.

Charron G, Leducq JB, Landry CR (2014) Chromosomal variation segregates within incipient species and correlates with reproductive isolation. *Molecular Ecology*, **23**, 4362–4372.

Cheng C, White BJ, Kamdem C, *et al.* (2012) Ecological genomics of *Anopheles gambiae* along a latitudinal cline: A population-resequencing approach. *Genetics*, **190**, 1417–1432.

Claessen D, Andersson J, Persson L, de Roos AM (2008) The effect of population size and recombination on delayed evolution of polymorphism and speciation in sexual populations. *American Naturalist*, **172**, E18–E34.

Clarke B (1969) Evidence for apostatic selection. *Heredity*, **24**, 347–352.

Clarke B (1979) Evolution of genetic diversity. *Proceedings of the Royal Society Series B-Biological Sciences*, **205**, 453–474.

Clarke C, Sheppard P (1972) Genetics of mimetic butterfly *Papilio polytes* L. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **263**, 431–458.

Comeault AA, Carvalho CF, Dennis S, Soria-Carrasco V, Nosil P (2016) Color phenotypes are under similar genetic control in two distantly related species of *Timema* stick insect. *Evolution*, **70**, 1283–1296.

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

Cook L (1998) A two-stage model for *Cepaea* polymorphism. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **353**, 1577–1593.

Corbett-Detig RB, Hartl DL (2012) Population genomics of inversion polymorphisms in *Drosophila melanogaster*. *PLoS Genetics*, **8**, e1003056.

Coyne J, Orr H (1998) The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **353**, 287–305.

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

De Boer R, Borghans J, van Boven M, Kesmir C, Weissing F (2004) Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics*, **55**, 725–731.

DeGiorgio M, Lohmueller KE, Nielsen R (2014) A model-based approach for identifying signatures of ancient balancing selection in genetic data. *PLoS Genetics*, **10**, e1004561.

Endler J (1981) An overview of the relationships between mimicry and crypsis. *Biological Journal of the Linnean Society*, **16**, 25–31.

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

Faria R, Navarro A (2010) Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends in Ecology & Evolution*, **25**, 660–669.

Feder J, Berlocher S, Roethele J, *et al.* (2003) Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 10314–10319.

Fijarczyk A, Babik W (2015) Detecting balancing selection in genomes: limits and prospects. *Molecular Ecology*, **24**, 3529–3545.

Forsman A, Wennersten L (2016) Inter-individual variation promotes ecological success of populations and species: evidence from experimental and comparative studies. *Ecography*, **39**, 630–648.

Garrigan D, Hedrick P (2003) Perspective: Detecting adaptive molecular polymorphism: Lessons from the MHC. *Evolution*, **57**, 1707–1722.

Gemmell NJ, Slate J (2006) Heterozygote advantage for fecundity. *PLoS One*, **1**, e125.

Gompert Z, Comeault AA, Farkas TE, *et al.* (2014) Experimental evidence for ecological selection on genome variation in the wild. *Ecology Letters*, **17**, 369–379.

Gompert Z, Lucas LK, Nice CC, Fordyce JA, Forister ML, Buerkle CA (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, **66**, 2167–2181.

Gray SM, McKinnon JS (2007) Linking color polymorphism maintenance and speciation. *Trends in Ecology & Evolution*, **22**, 71–79.

Guerrero RF, Rousset F, Kirkpatrick M (2012) Coalescent patterns for chromosomal inversions in divergent populations. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **367**, 430–438.

Gulisija D, Kim Y (2015) Emergence of long-term balanced polymorphism under cyclic selection of spatially variable magnitude. *Evolution*, **69**, 979–992.

Hedrick P, Ginevan M, Ewing E (1976) Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology and Systematics*, **7**, 1–32.

Hedrick PW (2006) Genetic polymorphism in heterogeneous environments: The age of genomics. *Annual Review of Ecology Evolution and Systematics*, **37**, 67–93.

Hedrick PW (2011) Population genetics of malaria resistance in humans. *Heredity*, **107**, 283–304.

Hedrick PW (2012) What is the evidence for heterozygote advantage selection? *Trends in Ecology & Evolution*, **27**, 698–704.

Hedrick PW, Smith DW, Stahler DR (2016) Negative-assortative mating for color in wolves. *Evolution*, **70**, 757–766.

Hoekstra H, Drumm K, Nachman M (2004) Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution*, **58**, 1329–1341.

van't Hof AE, Campagne P, Rigden DJ, *et al.* (2016) The industrial melanism mutation in British peppered moths is a transposable element. *Nature*, **534**, 102–105.

Hudson R, Kaplan N (1988) The coalescent process in models with selection and recombination. *Genetics*, **120**, 831–840.

Hugall AF, Stuart-Fox D (2012) Accelerated speciation in colour-polymorphic birds. *Nature*, **485**, 631–634.

Imsland F, Feng C, Boije H, *et al.* (2012) The Rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. *PLoS Genetics*, **8**.

Johannesson K, Butlin RK (2017) What explains rare and conspicuous colours in a snail? A test of time-series data against models of drift, migration or selection. *Heredity*, **118**, 21–30.

Johnston SE, Gratten J, Berenos C, *et al.* (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, **502**, 93–95.

Joron M, Frezal L, Jones RT, *et al.* (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, **477**, 203–206.

Kelly J (1997) A test of neutrality based on interlocus associations. *Genetics*, **146**, 1197–1206.

Kelly J, Wade M (2000) Molecular evolution near a two-locus balanced polymorphism. *Journal of Theoretical Biology*, **204**, 83–101.

King R, Lawson R (1995) Color-pattern variation in Lake Erie watersnakes: the role of gene flow. *Evolution*, **49**, 885–896.

Kirkpatrick M (2010) How and why chromosome inversions evolve. *PLoS Biology*, **8**, e1000501.

Kirkpatrick M, Barrett B (2015) Chromosome inversions, adaptive cassettes and the evolution of species' ranges. *Molecular Ecology*, **24**, 2046–2055.

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

Kuepper C, Stocks M, Risse JE, *et al.* (2016) A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, **48**, 79–83.

Kunte K, Zhang W, Tenger-Trolander A, *et al.* (2014) *doublesex* is a mimicry supergene. *Nature*, **507**, 229–232.

Lamichhaney S, Fan G, Widemo F, *et al.* (2016) Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nature Genetics*, **48**, 84–88.

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

Law J, Crespi B (2002) The evolution of geographic parthenogenesis in *Timema* walking-sticks. *Molecular Ecology*, **11**, 1471–1489.

Le Poul Y, Whibley A, Chouteau M, Prunier F, Llaurens V, Joron M (2014) Evolution of dominance mechanisms at a butterfly mimicry supergene. *Nature Communications*, **5**, 5644.

Leffler EM, Gao Z, Pfeifer S, *et al.* (2013) Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science*, **339**, 1578–1582.

Levene H (1953) Genetic equilibrium when more than one ecological niche is available. *American Naturalist*, **87**, 331–333.

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993.

Llaurens V, Whibley A, Joron M (2017) Genetic architecture and balancing selection: the life and death of differentiated variants. *Molecular Ecology*, pp. n/a–n/a.

Lynch M, Conery J, Burger R (1995) Mutation accumulation and the extinction of small populations. *American Naturalist*, **146**, 489–518.

Mappes J, Marples N, Endler J (2005) The complex business of survival by aposematism. *Trends in Ecology & Evolution*, **20**, 598–603.

Nachman M, Hoekstra H, D'Agostino S (2003) The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5268–5273.

888 Nagylaki T (2009) Polymorphism in multiallelic migration-selection models with
889 dominance. *Theoretical Population Biology*, **75**, 239–259.

890 Navarro A, Bardadilla A, Ruiz A (2000) Effect of inversion polymorphism on the neutral
891 nucleotide variability of linked chromosomal regions in *Drosophila*. *Genetics*, **155**,
892 685–698.

893 Navarro A, Barton N (2002) The effects of multilocus balancing selection on neutral
894 variability. *Genetics*, **161**, 849–863.

895 Navarro A, Barton N (2003) Accumulating postzygotic isolation genes in parapatry: A new
896 twist on chromosomal speciation. *Evolution*, **57**, 447–459.

897 Navarro A, Betran E, Barbadilla A, Ruiz A (1997) Recombination and gene flux caused by
898 gene conversion and crossing over in inversion heterokaryotypes. *Genetics*, **146**, 695–709.

899 Nei M, Maruyama T, Chakraborty R (1975) Bottleneck effect and genetic variability in
900 populations. *Evolution*, **29**, 1–10.

901 Nordborg M, Innan H (2003) The genealogy of sequences containing multiple sites subject
902 to strong selection in a subdivided population. *Genetics*, **163**, 1201–1213.

903 Nosil P (2004) Reproductive isolation caused by visual predation on migrants between
904 divergent environments. *Proceedings of the Royal Society B-Biological Sciences*, **271**,
905 1521–1528.

906 Nosil P, Crespi B (2006) Experimental evidence that predation promotes divergence in
907 adaptive radiation. *Proceedings of the National Academy of Sciences of the United States*
908 *of America*, **103**, 9090–9095.

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

911 Nosil P, Gompert Z, Farkas TE, *et al.* (2012) Genomic consequences of multiple speciation
912 processes in a stick insect. *Proceedings of the Royal Society B-Biological Sciences*, **279**,
913 5058–5065.

914 Novikova PY, Hohmann N, Nizhynska V, *et al.* (2016) Sequencing of the genus *Arabidopsis*
915 identifies a complex history of nonbifurcating speciation and abundant trans-specific
916 polymorphism. *Nature Genetics*, **48**, 1077–1082.

917 Otto S, Lenormand T (2002) Resolving the paradox of sex and recombination. *Nature*
918 *Reviews Genetics*, **3**, 252–261.

919 Pamilo P, Palsson S (1998) Associative overdominance, heterozygosity and fitness. *Heredity*,
920 **81**, 381–389.

Peischl S, Koch E, Guerrero R, Kirkpatrick M (2013) A sequential coalescent algorithm for chromosomal inversions. *Heredity*, **111**, 200–209.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, **38**, 904–909.

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

Przeworski M, Coop G, Wall J (2005) The signature of positive selection on standing genetic variation. *Evolution*, **59**, 2312–2323.

R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Richards PM, Liu MM, Lowe N, Davey JW, Blaxter ML, Davison A (2013) RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Molecular Ecology*, **22**, 3077–3089.

Riesch R, Muschick M, Lindtke D, *et al.* (2017) Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution*, **1**, 0082.

Rieseberg L (2001) Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, **16**, 351–358.

Sabeti P, Reich D, Higgins J, *et al.* (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**, 832–837.

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

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

Sandoval C (2000) Persistence of a walking-stick population (Phasmatoptera : Timematodea) after a wildfire. *Southwestern Naturalist*, **45**, 123–127.

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

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

Schaeffer SW (2008) Selection in heterogeneous environments maintains the gene arrangement polymorphism of *Drosophila pseudoobscura*. *Evolution*, **62**, 3082–3099.

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

957 Schwander T, Crespi BJ (2009) Multiple direct transitions from sexual reproduction to
958 apomictic parthenogenesis in *Timema* stick insects. *Evolution*, **63**, 84–103.

959 Sellis D, Callahan BJ, Petrov DA, Messer PW (2011) Heterozygote advantage as a natural
960 consequence of adaptation in diploids. *Proceedings of the National Academy of Sciences*
961 *of the United States of America*, **108**, 20666–20671.

962 Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality
963 fro DNA polymorphism data. *Genetics*, **141**, 413–429.

964 Skelhorn J, Rowe C (2016) Cognition and the evolution of camouflage. *Proceedings of the*
965 *Royal Society B-Biological Sciences*, **283**, 20152890.

966 Slatkin M (2008) Linkage disequilibrium - understanding the evolutionary past and
967 mapping the medical future. *Nature Reviews Genetics*, **9**, 477–485.

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

970 Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC,
971 mechanisms and misunderstandings. *Proceedings of the Royal Society B-Biological*
972 *Sciences*, **277**, 979–988.

973 Stevens M, Merilaita S (2009) Animal camouflage: current issues and new perspectives.
974 *Philosophical Transactions of the Royal Society B-Biological Sciences*, **364**, 423–427.

975 Storz JF, Kelly JK (2008) Effects of spatially varying selection on nucleotide diversity and
976 linkage disequilibrium: Insights from deer mouse globin genes. *Genetics*, **180**, 367–379.

977 Svardal H, Rueffler C, Hermisson J (2015) A general condition for adaptive genetic
978 polymorphism in temporally and spatially heterogeneous environments. *Theoretical*
979 *Population Biology*, **99**, 76–97.

980 Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA
981 polymorphism. *Genetics*, **123**, 585–595.

982 Tang K, Thornton KR, Stoneking M (2007) A new approach for using genome scans to
983 detect recent positive selection in the human genome. *PLoS Biology*, **5**, 1587–1602.

984 Tuttle EM, Bergland AO, Korody ML, *et al.* (2016) Divergence and functional degradation
985 of a sex chromosome-like supergene. *Current Biology*, **26**, 344–350.

986 Vickery VR, Sandoval CP (2001) Descriptions of three new species of *Timema*
987 (Phasmatoptera: Timematodea: Timematidae) and notes on three other species. *Journal*
988 *of Orthoptera Research*, **10**, 53–61.

Voight B, Kudaravalli S, Wen X, Pritchard J (2006) A map of recent positive selection in the human genome. *PLoS Biology*, **4**, 446–458.

Wakeley J (2008) *Coalescent Theory: An Introduction*. W. H. Freeman.

Wallace AG, Detweiler D, Schaeffer SW (2013) Molecular population genetics of inversion breakpoint regions in *Drosophila pseudoobscura*. *G3: Genes, Genomes, Genetics*, **3**, 1151–1163.

Wang J, Wurm Y, Nipitwattanaphon M, *et al.* (2013) A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*, **493**, 664–668.

Weir B (1979) Inferences about linkage disequilibrium. *Biometrics*, **35**, 235–254.

Wellenreuther M, Svensson EI, Hansson B (2014) Sexual selection and genetic colour polymorphisms in animals. *Molecular Ecology*, **23**, 5398–5414.

White BJ, Cheng C, Sangare D, Lobo NF, Collins FH, Besansky NJ (2009) The population genomics of trans-specific inversion polymorphisms in *Anopheles gambiae*. *Genetics*, **183**, 275–288.

White BJ, Hahn MW, Pombi M, *et al.* (2007) Localization of candidate regions maintaining a common polymorphic inversion (2La) in *Anopheles gambiae*. *PLoS Genetics*, **3**, 2404–2414.

Wigginton J, Cutler D, Abecasis G (2005) A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, **76**, 887–893.

Wiuf C, Zhao K, Innan H, Nordborg M (2004) The probability and chromosomal extent of trans-specific polymorphism. *Genetics*, **168**, 2363–2372.

Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, E1743–E1751.

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

Data Accessibility

The genetic data from population N1 has been deposited in the NCBI Short Read Archive (BioProject PRJNA386212). Phenotypic and processed genetic data used for analysis as well as simulation code have been archived in Dryad (doi:10.5061/dryad.jt644).

Author Contributions

DL, KL, and PN conceived the project. DL, VSC, RV, and SRD performed data analyses. TEF and RR collected data. ZG contributed to computer code and discussions. DL and PN wrote the manuscript, and all authors contributed to further writing and revisions.

Tables and Figures

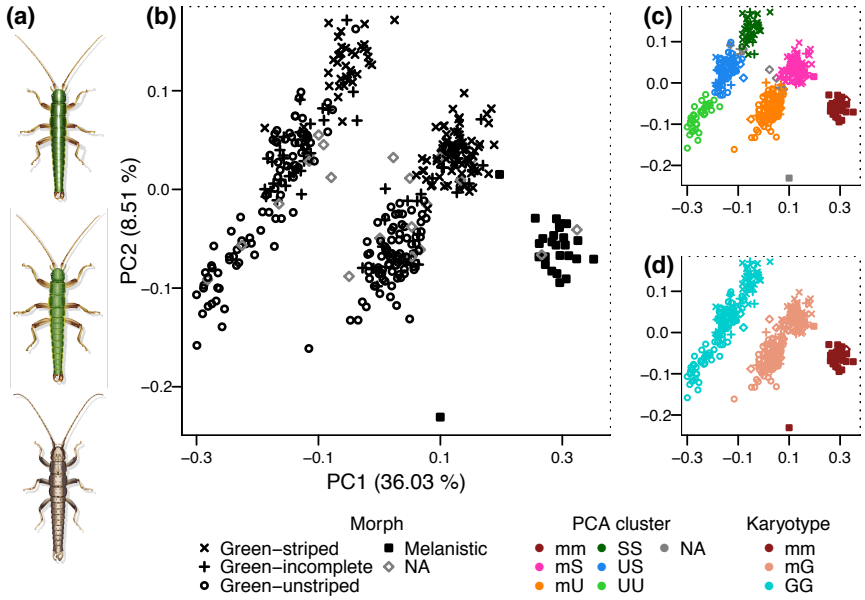

Figure 1: Genetic structure on LG8 associated with phenotypic morphs in population N1. (a) Three *T. cristinae* morphs cryptic on leaves or stems of their host plants: green-striped, green-unstriped, and melanistic. (b) Principal component axis one (PC1) shows clustering by color morph, with two distinct clusters for green morphs (crosses, pluses, and circles) and one cluster for the melanistic morph (filled squares). Principal component axis two (PC2) shows a gradient by pattern morph, from green-unstriped (circles) to green-striped morphs (crosses). (c) K-means clustering and linear discriminant analysis were used to define six PCA clusters, corresponding to diploid combinations of three chromosomal variants 'melanistic' (m), 'green-striped' (S), and 'green-unstriped' (U). (d) STRUCTURE with $k = 2$ identified two main chromosomal variants 'melanistic' (m) and 'green' (G), resulting in three main karyotypes by their diploid combinations. PCA outlier individuals were excluded in (b) and (c) and are not shown in (d).

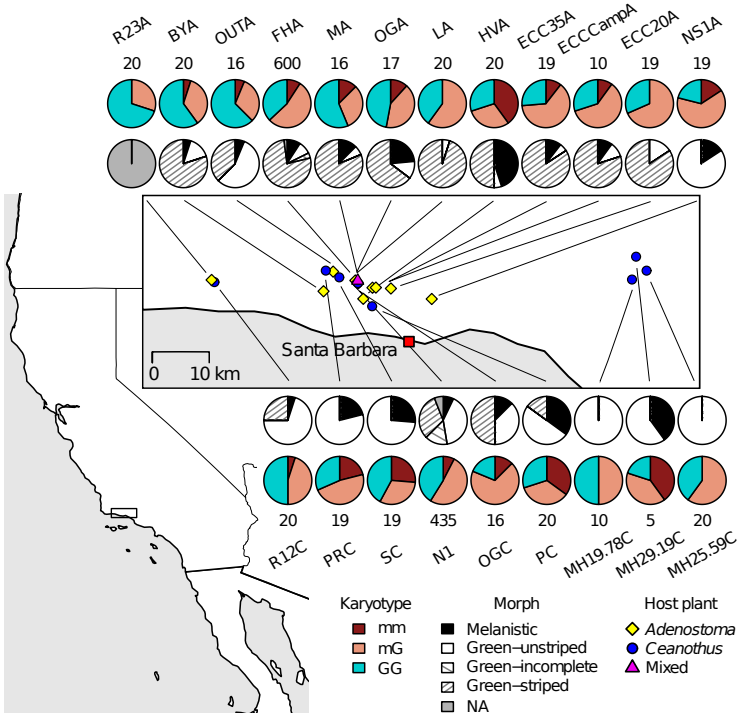

Figure 2: Sampling localities, karyotype and morph frequencies for 21 *T. cristinae* populations. Samples were obtained from 12 populations on *Adenostoma* host plants (yellow diamonds), eight populations on *Ceanothus* host plants (blue circles), and one mixed population (N1, magenta triangle) across the species distribution around Santa Barbara, California. Pie charts in the top and bottom rows show karyotype frequencies estimated by genetic clustering using STRUCTURE with $k = 2$ (i.e., corresponding to m and G chromosomal variants). Pie charts in the second and third rows give morph frequencies for *Adenostoma* hosts (top) and *Ceanothus* or mixed hosts (bottom). Population names and sample sizes are indicated.

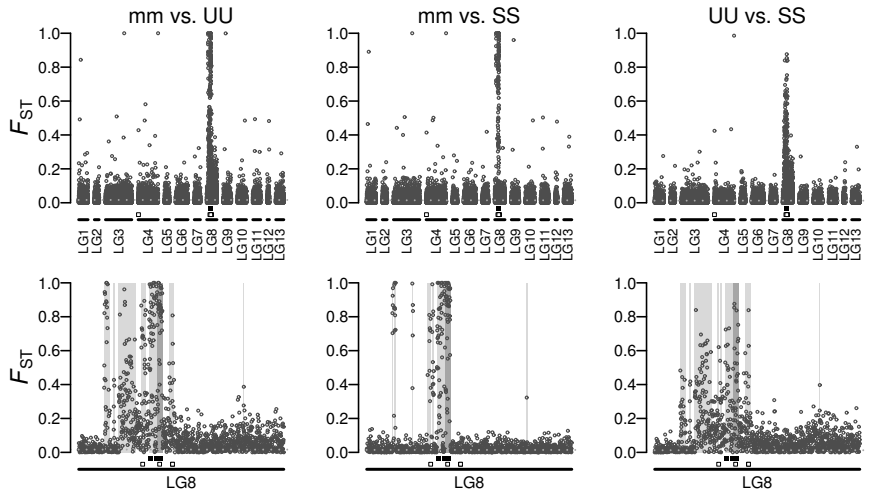

Figure 3: Genome-wide differentiation between pairs of homokaryotypic clusters in population N1. Top row, F_{ST} for non-overlapping 20-kb windows and all LGs; bottom row, LG8 only. Gray dotted lines show genome-wide 50% quantiles, and squares on the x-axis indicate positions of candidate SNPs for color (closed symbols) and pattern (open symbols). High-differentiation scaffolds are highlighted in gray, with a darker shade indicating a subset of three scaffolds that were treated separately in some analyses (scaffolds 931, 318, and 1440). The distinctive blocks of high F_{ST} on LG8 span approximately 13 Mb.

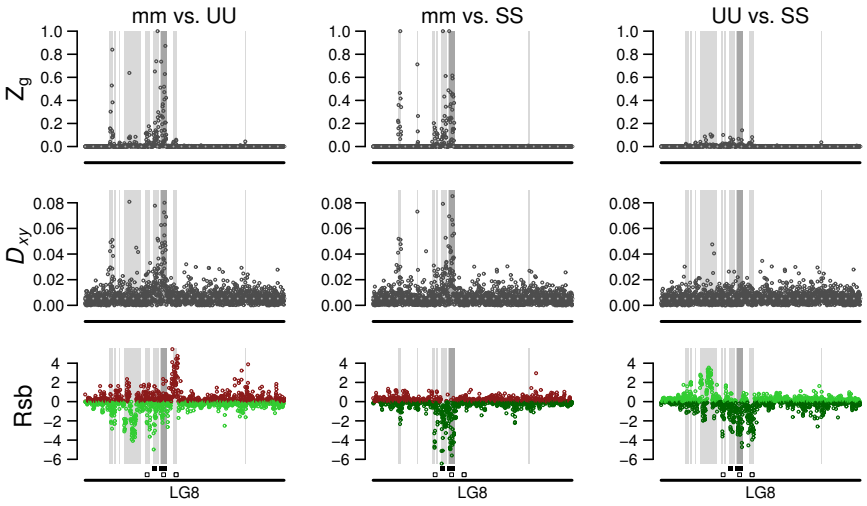

Figure 4: Population genomic parameters along LG8 for pairs of homokaryotypic clusters in population N1. Regions of high intra-locus LD (Z_g) and increased absolute genetic divergence (D_{xy}) between clusters coincide with high-differentiation scaffolds (Fig. 3; highlighted in gray). Extreme values in Rsb indicate that regions of extended haplotype homozygosity differ among clusters. All statistics were calculated in non-overlapping 20-kb windows. Gray dotted lines show genome-wide 50% quantiles for Z_g and D_{xy}, and squares on the x-axis indicate positions of candidate SNPs for color (closed symbols) and pattern (open symbols).

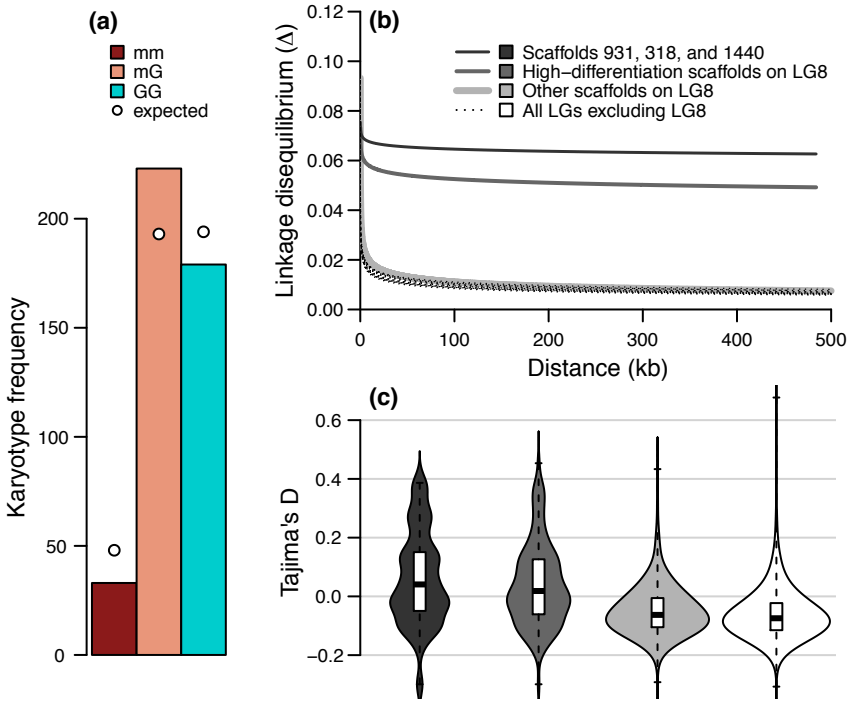

Figure 5: History of balancing selection in population N1. (a) Karyotype frequencies. Bars show observed karyotype counts, and circles indicate expected counts for a population in HWE. (b) Decay of LD with physical distance between pairs of SNPs summarized for different sets of scaffolds on LG8 and for all other LGs. The y-axis shows Burrow's composite measure of Hardy Weinberg and LD (Δ). Lines were fitted by non-linear regression. (c) Tajima's D statistic for non-overlapping 20-kb windows for different sets of scaffolds on LG8 and for all other LGs combined. White boxes range from the first to third quartile, black horizontal bars give the median, whiskers extend to the data extremes, and shapes are Gaussian kernel densities.

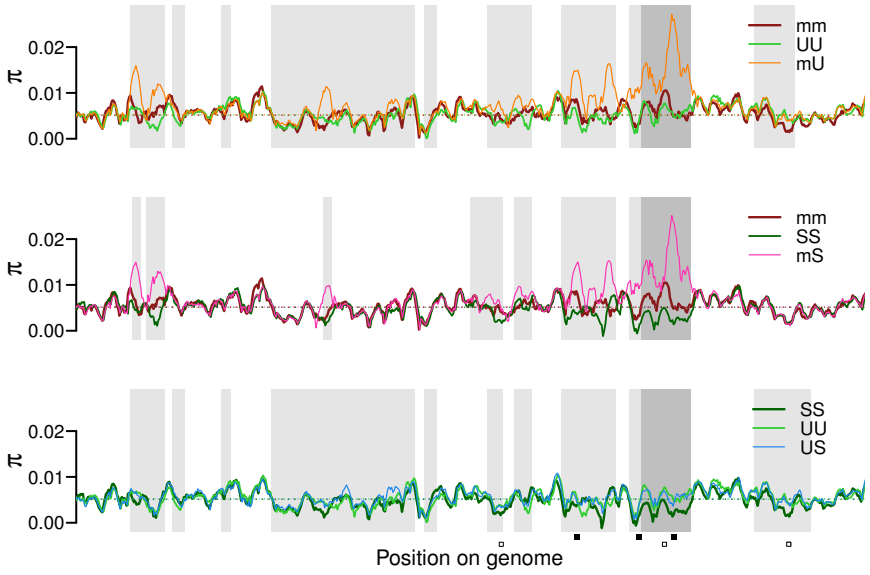

Figure 6: Nucleotide diversity (π) along LG8 for all clusters in population N1. Approximately one third of LG8 is shown. Each panel shows π for two homokaryotypic clusters and their corresponding heterokaryotypic cluster. Orange and pink lines in the top two panels show regions of increased diversity in heterokaryotyes between melanistic and green variants (mU or mS) relative to diversity within homokaryotypes (mm, UU, or SS), coinciding with high-differentiation scaffolds (Fig. 3; highlighted in gray). Increased diversity is not observed for heterokaryotypes between green-unstriped and green-striped variants (US; bottom panel, blue line). Lines depict smoothed estimates from non-overlapping 20-kb windows. Dotted lines show genome-wide 50% quantiles, and squares on the x-axis indicate positions of candidate SNPs for color (closed symbols) and pattern (open symbols).