

## ECOLOGICAL GENOMICS

# Large-scale mutation in the evolution of a gene complex for cryptic coloration

Romain Villoutreix<sup>1,2</sup>, Clarissa F. de Carvalho<sup>1</sup>, Víctor Soria-Carrasco<sup>1\*</sup>, Dorothea Lindtke<sup>3</sup>, Marisol De-la-Mora<sup>1,4</sup>, Moritz Muschick<sup>5,6</sup>, Jeffrey L. Feder<sup>7</sup>, Thomas L. Parchman<sup>8</sup>, Zach Gompert<sup>9</sup>, Patrik Nosil<sup>1,2,9†</sup>

The types of mutations affecting adaptation in the wild are only beginning to be understood. In particular, whether structural changes shape adaptation by suppressing recombination or by creating new mutations is unresolved. Here, we show that multiple linked but recombining loci underlie cryptic color morphs of *Timema chumash* stick insects. In a related species, these loci are found in a region of suppressed recombination, forming a supergene. However, in seven species of *Timema*, we found that a megabase-size “supermutation” has deleted color loci in green morphs. Moreover, we found that balancing selection likely contributes more to maintaining this mutation than does introgression. Our results show how suppressed recombination and large-scale mutation can help to package gene complexes into discrete units of diversity such as morphs, ecotypes, or species.

Genes that control adaptive traits have now been identified for defensive armor in stickleback (*Eda* and *Pitx*) (1, 2), coat color in mice (*Agouti*) (3) and Soay sheep (*Tyrrp1*) (4), beak size in Darwin’s finches (*Hmga*) (5), flower color in phlox (*Chs1*) (6), and flowering time in *Arabidopsis* (*Fri*) (7), among other examples (8). However, our understanding of the genetics of adaptation remains far from complete. This is because the type, number, and sequence of mutations that cause variation in adaptive genes are often unknown, yet these details can affect evolutionary dynamics (9, 10). Moreover, adaptation often involves multiple genes (11) that may interact. When this is the case, some gene combinations confer different fitness than others. This can lead to the evolution of coadapted gene complexes and make it difficult to predict evolution from analyses of single genes alone (9, 10).

Our understanding of adaptation is particularly poor when genes in regions of suppressed recombination control traits because this complicates fine-scale genetic mapping (12, 13). Examples of multiple traits mapping

to regions of suppressed recombination, such as those that are due to structural variants such as chromosomal inversions, are accumulating across diverse insect, plant, bird, and fish taxa (14–18). Many of these cases are thought to represent “supergenes,” in which suppressed recombination is favored to maintain adaptive gene combinations. This supergene hypothesis predicts that multiple genes control adaptive traits and that they are distributed throughout regions of suppressed recombination (Fig. 1).

Alternatively, structural variants create adaptive mutations (14, 19) such as inversion breakpoints that disrupt a reading frame or alter expression at a developmental switch gene. This mutation hypothesis predicts that genes affecting adaptive traits will be more localized at regions such as breakpoints. These two hypotheses are not mutually exclusive, but the processes involved may counteract each other because reduced recombination can have a subsequent effect on suppressing mutation.

Here, we found evidence related to each hypothesis, but also identified a large-scale deletion that eliminates dozens of genes, affecting multiple adaptive color loci (i.e., a “supermutation”; Fig. 1). We studied wingless, phytophagous *Timema* stick insects, which rely on crypsis for protection against visual predators while resting on their host plants (20–22). Body coloration in many *Timema* species has evolved into green and brown (i.e., melanistic) morphs that approximate the colors of the leaves and stems of their hosts, respectively (Fig. 1). Green versus melanistic morphs of *T. cristinae* are distinguished by a major locus on linkage group eight (LG8 hereafter) named *Mel-Stripe* (22, 23) (Fig. 1). This locus spans ~10 megabases of sequence and exhibits suppressed recombination, putatively caused by an inversion. Additionally, one edge of the locus exhibits a large-scale (~1 megabase pair)

insertion/deletion (indel) polymorphism that harbors 83 predicted genes (table S1).

Whether one, few, or many loci within *Mel-Stripe* affect color is unknown, as are the locations and identities of such loci. We aimed to resolve these issues here, in the context of the hypotheses above, including testing the core assumption of the supergene and supermutation hypotheses, i.e., that multiple linked loci rather than a single developmental switch affect trait variation. To achieve this, we studied 10 of the 11 sexual species of *Timema* found in California, including all those in the Northern California and Santa Barbara clades (Northern California clade = *T. poppensis*, *T. knulli*, *T. curi* spp., *T. petita*, *T. californicum*, and *T. landelsensis*; Santa Barbara clade = *T. cristinae*; Southern California clade = *T. chumash*, *T. bartmani*, and *T. podura*; nine species were used in both sequencing coverage and phylogenetic analyses, whereas data for coverage and phylogenetics were not collected for *T. podura* and *T. petita*, respectively; Fig. 1).

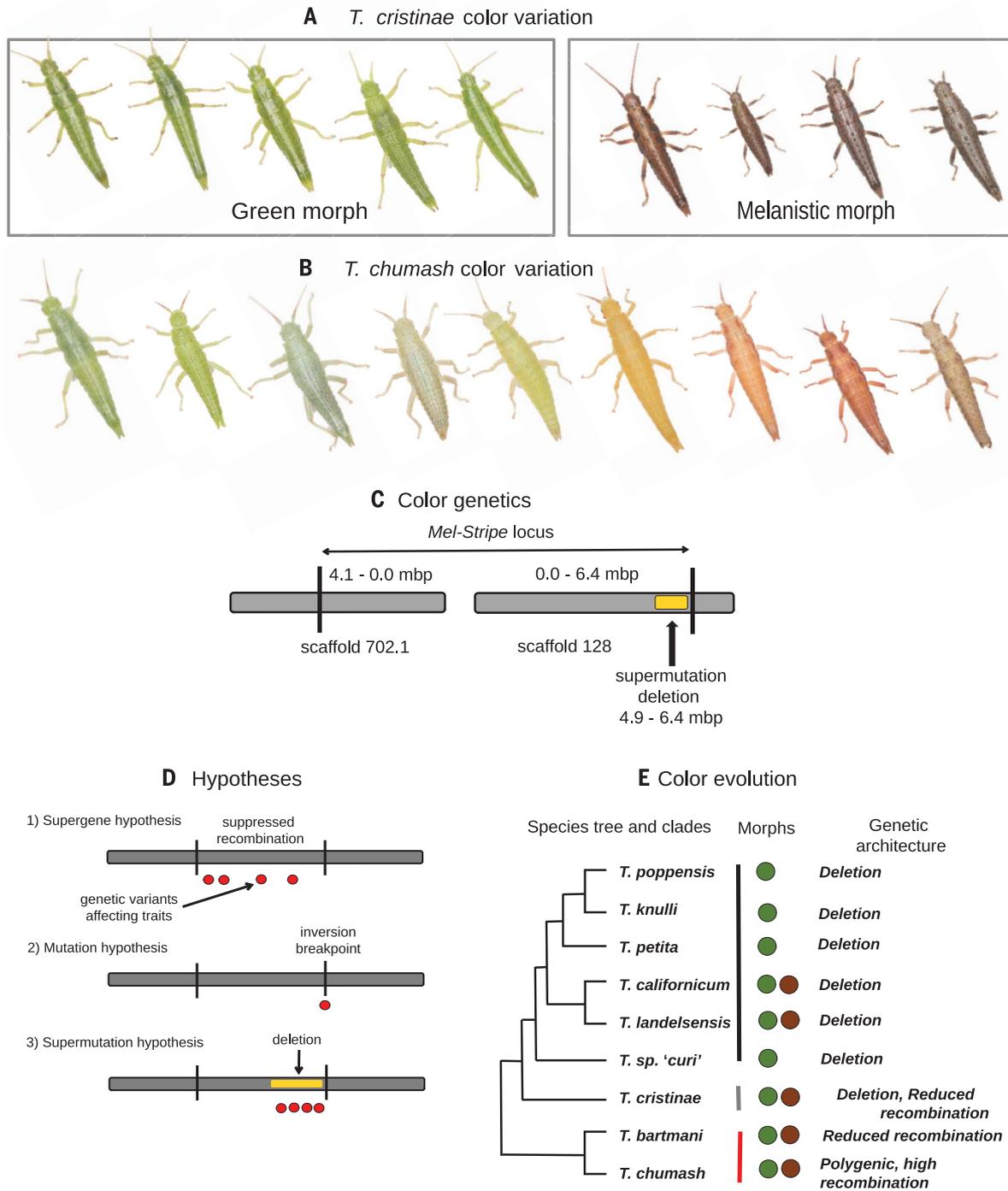
We first collected genotyping-by-sequencing (GBS) data from thousands of individuals from nine *Timema* species distributed across the three clades found in California (24). We used these data to compare sequencing coverage between the indel region and the rest of LG8. This allowed us to test whether the indel represents an insertion or a deletion, and to quantify how widespread the mutation is (Fig. 1 and tables S2 to S6). This revealed that sequencing coverage across the indel region is comparable to that for the remainder of LG8 for the melanistic morph in all species in the Santa Barbara and Northern California clades, but reduced relative to the rest of LG8 for the green morph in all species in these two clades (including polymorphic species and those monomorphic for green) (Fig. 2). This pattern alone could represent an insertion in the melanistic morph or a deletion in the green morph. However, if melanistic individuals in the Santa Barbara and Northern California clades harbored an insertion, then we would expect this region to be missing in the Southern California clade, resulting in low coverage in both morphs in this clade. However, coverage across the indel region for both morphs in the Southern California clade was comparable to that for the rest of LG8.

Our data thus support that the indel is a deletion in green individuals in the Santa Barbara and Northern California clades (for convenience, we at times refer to this as “the deleted region,” with the understanding that it is not deleted in the Southern California clade). Because of the observed low sequencing coverage of the deleted region for green morphs of different species, we hypothesize that this region controls color across species (Fig. 2), and that genes in this region maintain synteny among species [see linkage

<sup>1</sup>Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK. <sup>2</sup>Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175 CNRS-Université de Montpellier, École Pratique des Hautes Études, Université Paul Valéry, 34293 Montpellier 5, France. <sup>3</sup>Department of Biological Sciences, University of Bern, Bern CH-3012, Switzerland. <sup>4</sup>Escuela Nacional de Estudios Superiores, Unidad Juriquilla, Querétaro 76230, México. <sup>5</sup>Institute of Ecology and Evolution, University of Bern, Bern 3012, Switzerland. <sup>6</sup>Department of Fish Ecology & Evolution, EAWAG, Swiss Federal Institute for Aquatic Science and Technology, Kastanienbaum CH-6047, Switzerland. <sup>7</sup>Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA. <sup>8</sup>Department of Biology, University of Nevada–Reno, Reno, NV 89557, USA. <sup>9</sup>Department of Biology, Utah State University, Logan, UT 84322, USA.

\*Present address: John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK.

†Corresponding author. Email: patrik.nosil@cefe.cnrs.fr

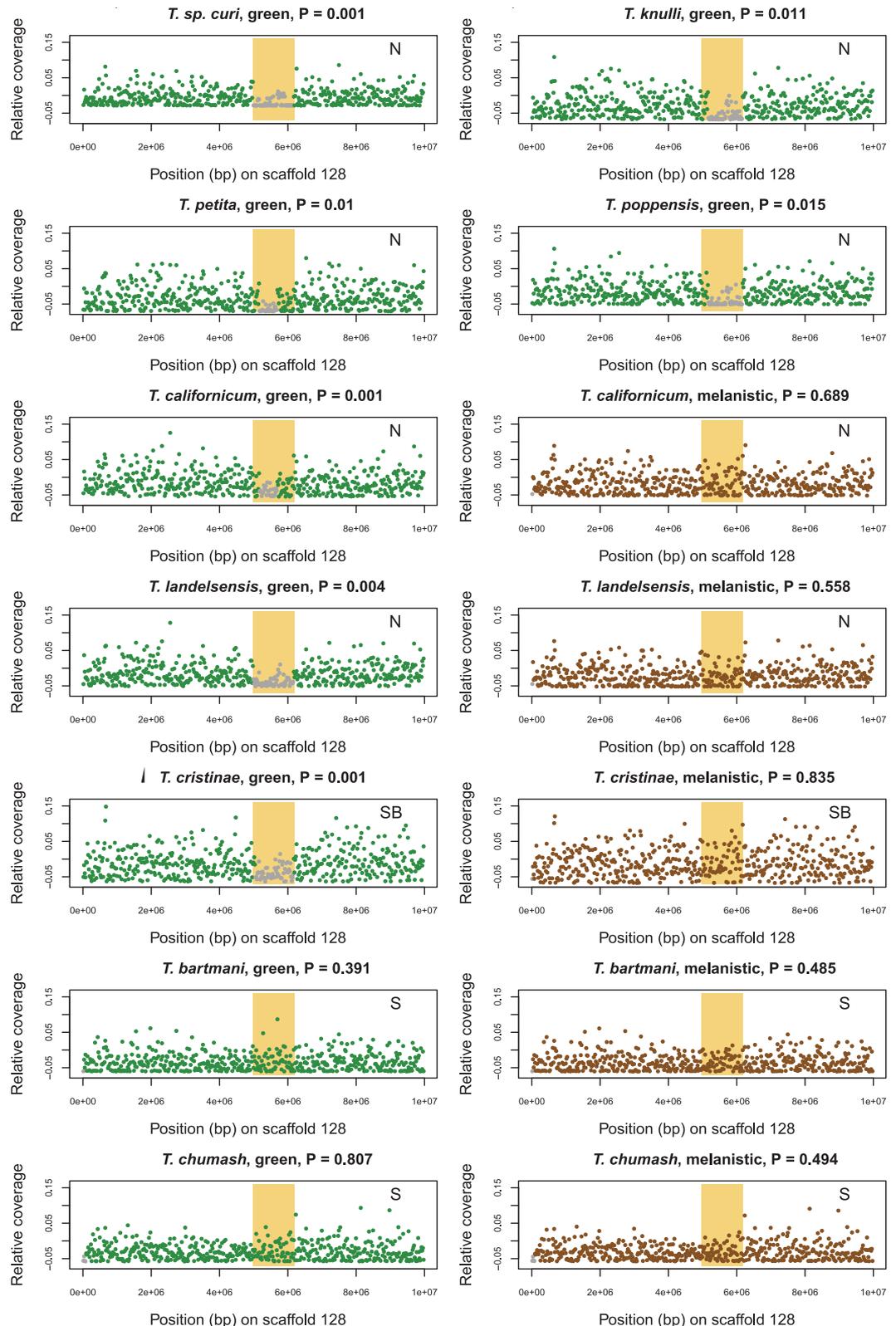


**Fig. 1. Genetics of coloration in *Timema* species and associated hypothesis testing.** (A) Photographs of *T. cristinae* exhibiting discrete green (left five individuals) and melanistic (right four individuals) color morphs. (Photographs by Moritz Muschick.) (B) Photographs of *T. chumash*, which exhibit a range of cryptic coloration. (Photographs by Moritz Muschick.) (C) Schematic of the *Mel-Stripe* locus on LG8, including the indel polymorphism. (D) The three core hypotheses considered in the present study. Gray bars are chromosomes and vertical black lines denote the breakpoint of an inversion. Red circles are causal variants affecting trait variation. The yellow box is a

large-scale mutation, in this case, a deletion. See the text for further details. (E) Phylogenetic relationships among the *Timema* species (24) studied and a summary of the core results. Center vertical lines delineate the three main extant California clades (black line, Northern California clade; gray line, Santa Barbara clade; and red line, Southern California clade). Green and brown filled circles depict green and melanistic morphs, respectively. Species with only one circle are monomorphic, and those with two circles are polymorphic. *T. podura* is also found in California but is not included because data were not collected for this species.

**Fig. 2. Evidence for a large indel polymorphism in *Timema*.**

Shown are sequencing coverage values for a section of LG8 for green versus melanistic individuals (mapped against a reference genome of a melanistic *T. cristinae*). Divergence from the reference genome was 24 million years ago for the Northern California clade (N), 30 million years ago for the Southern California clade (S), and zero for the Santa Barbara clade (SB) (24). Species monomorphic for green are: *T. curi* spp., *T. knulli*, *T. petita*, and *T. poppensis*. All other species are polymorphic for green and melanistic individuals, as shown by the coverage results. Melanism is recessive such that melanistic individuals are likely homozygous. Shown is relative coverage determined as coverage (sequence depth) minus the mean across all 20-kilobase pair windows. More specifically, we centered (subtracted the mean) and standardized (divided by the SD) the read depth measures such that they had a mean of 0 and an SD of 1. The boxed region is the indel in the *T. cristinae* genome. *P* values above each panel give the probability of observing coverage lower than that in *Mel-Stripe* more generally by chance based on a randomization test. Shown are results from a two-state hidden Markov model (green and brown indicate a normal- to high-coverage state; gray indicates a low-coverage state). bp, base pair.

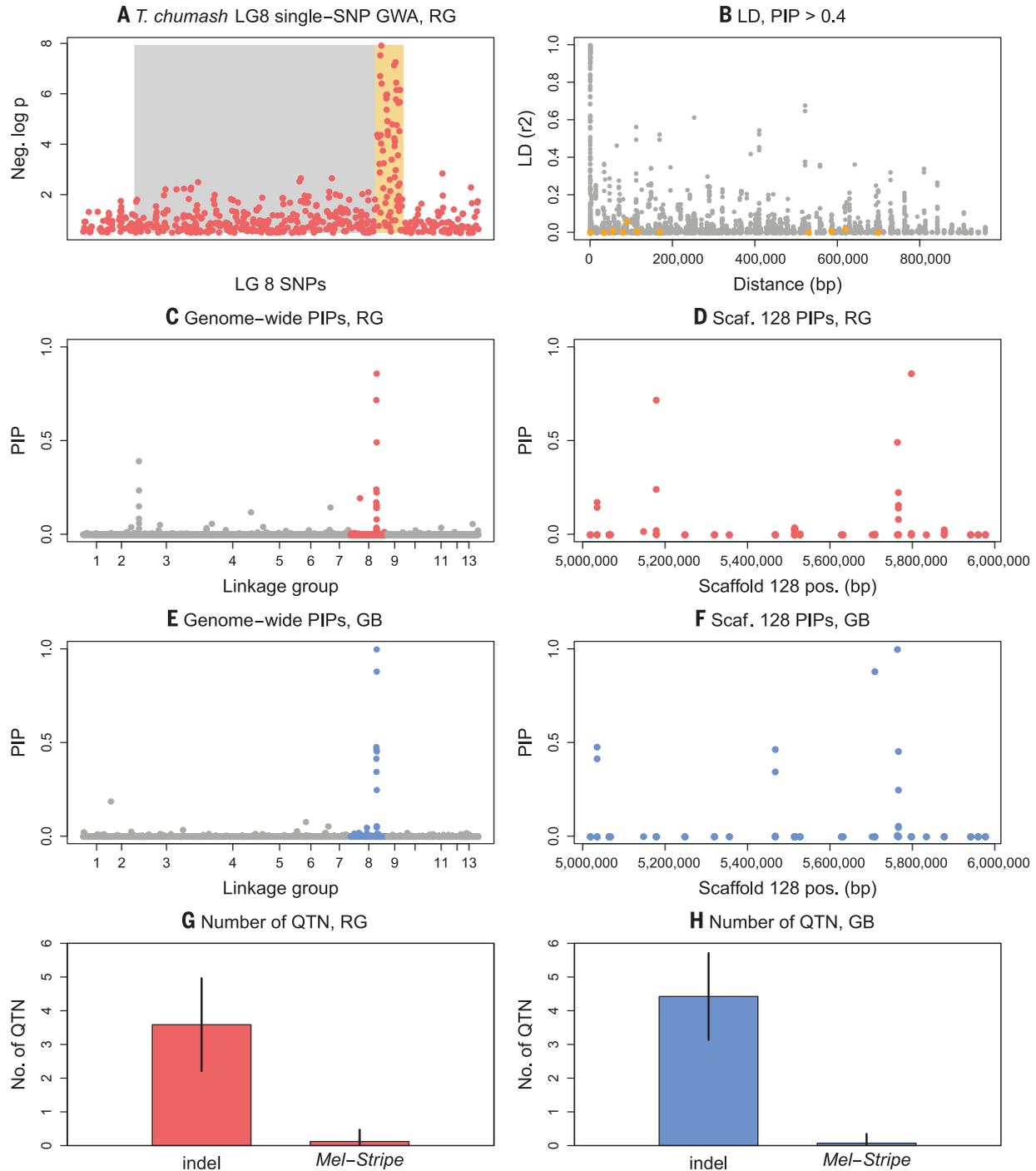


disequilibrium (LD) results below for further support for this claim].

To test the synteny prediction, we generated a new reference genome of a striped *T. cristinae* to compare with the melanistic

reference. We did so by de novo assembly of Illumina data followed by Hi-C technology, yielding a chromosome-level assembly. This revealed that the block of genes that is deleted in the green morphs from the Santa Barbara

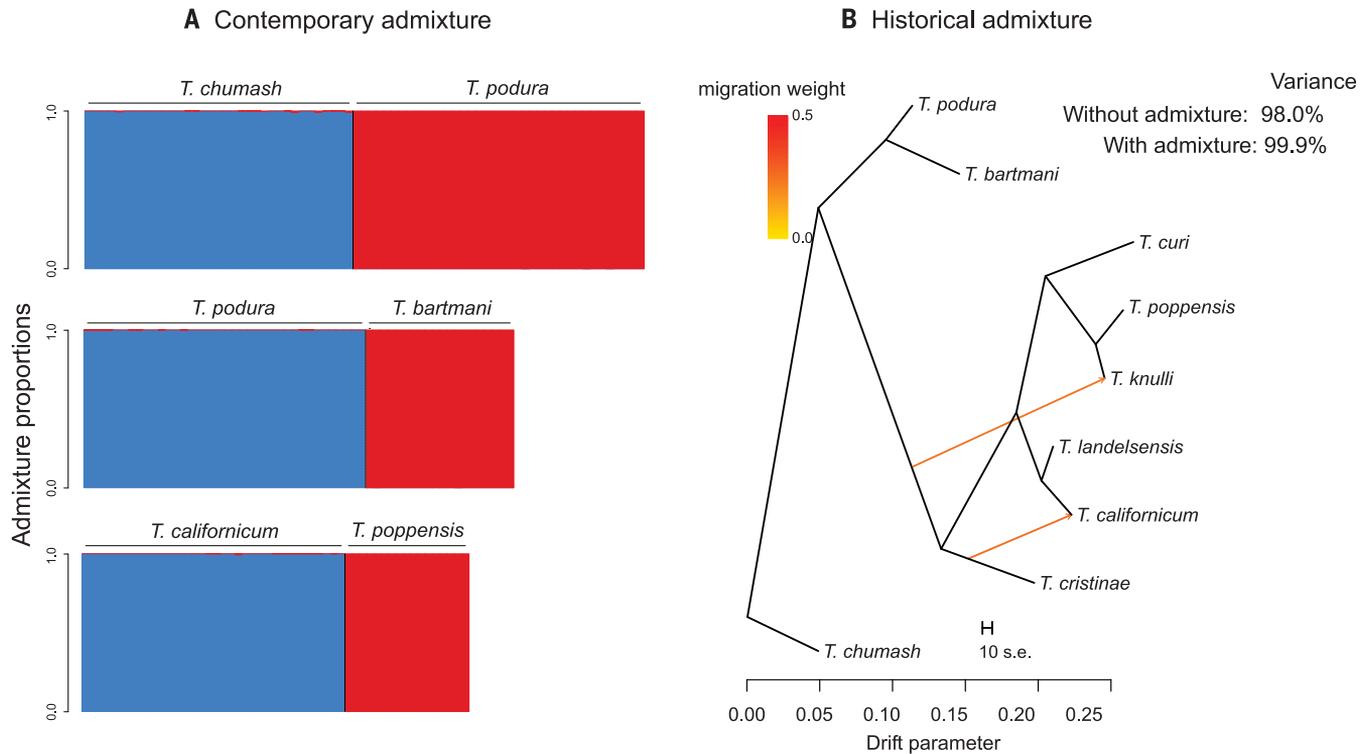
and Northern California clades forms a contiguous segment in the genome of the striped *T. cristinae*, supporting the synteny hypothesis. Notably, *T. cristinae*, which exhibits the deletion, has three genotypes at color loci that



**Fig. 3. Evidence for multigenic control of color variation in *T. chumash*.**

(A) Results of single-locus GWA mapping of RG color variation in *T. chumash* without correction for population structure in GENABEL. For analogous results with correction for population structure and for the GB trait, see figs. S2 to S4. The y-axis shows the negative  $\log_{10} P$  value (Neg. log p) for each test for which an SNP is associated with color variation. A peak of association is evident in the deleted region (yellow shading) but not in the remainder of the *Mel-Stripe* locus (gray shading). (B) LD, measured as the squared correlation coefficient  $r^2$  between color-associated SNPs on LG8 in *T. chumash*, shown as a function of the bp distances between such SNPs. Orange dots are SNPs with PIPs > 0.4. LD between these SNPs is low, and no higher than between

other SNPs (i.e., those with PIPs < 0.4, gray dots) in this genomic region. (C to F) Results from multilocus GWA mapping in GEMMA, genome wide and for the scaffold showing the bulk of strong associations, i.e., scaffold 128 on LG8, zoomed into the 5- to 6-Mbp position, which represents the deleted region. RG color variation is shown as red dots and GB color variation as blue dots, with coloring restricted to LG8 because it shows the bulk of association. pos. position. (G and H) Number of QTNs estimated to reside in the deleted region versus the remainder of *Mel-Stripe* for RG and GB, respectively (indel, 5.0 to 6.0 megabase pairs on scaffold 128; *Mel-Stripe* excluding indel, 1 to 4,139,489 base pairs on scaffold 702 and 1 to 5.0 megabase pairs on scaffold 128).



**Fig. 4. Inference of contemporary and historical admixture and introgression between *Timema* species.** (A) Colored bars showing the proportion of the genome for each individual inherited from each of two source populations (colored red versus blue) on the basis of the admixture model in ENTROPY (34). Species pairs shown are *T. chumash* versus *T. podura* (locality = BMT,  $N = 73$ , SNPs = 4194), *T. podura* versus *T. bartmani* (locality = BMCG3,  $N = 58$ , SNPs = 3108), and *T. californicum* versus *T. poppensis* (locality = SM,  $N = 53$ , SNPs = 2243). These were chosen because they are currently

co-occurring (e.g., sympatric or parapatric) species pairs in which the opportunity for admixture exists; other cases are shown in fig. S9. Data were taken directly from (24). (B) Historical admixture inferred from whole-genome resequencing data analyzed in TREEMIX. Here, *T. chumash* was designated as the outgroup. The two migration edges for the graph explaining 99.9% of the variation in the population allele frequency covariance matrix are shown and colored to denote the extent of admixture. Arrows on these colored edges indicate the direction of introgression.

occur at appreciable frequencies across the species range: green homozygotes (GG), melanistic homozygotes (MM), and heterozygotes (GM) (22). Thus, the deletion does not have lethal effects.

Given these results, we performed genetic mapping of color in two species in the Southern California clade, which lacks the deletion. We collected GBS data for 531 *T. chumash* and 132 *T. bartmani* insects using standardized photographs to quantify body color in the red-to-green color spectrum (a trait referred to as “RG” hereafter) and in the green-to-blue color spectrum (a trait referred to as “GB” hereafter). Single-nucleotide polymorphism (SNP) genome-wide association (GWA) mapping revealed that color maps to the ~1-megabase-pair region that is deleted in Northern California, and not to the larger region of suppressed recombination adjacent to it (Fig. 3 and figs. S2 to S4; see multilocus mapping below for statistical support; tables S2 and S5). Thus, color loci (or a color locus) were deleted in the Santa Barbara and Northern California clades, consistent with the supermutation hypothe-

sis. We did detect two additional minor peaks of color association on LGs other than LG8. Examination of the striped *T. cristinae* reference confirmed that these peaks do not reside on LG8 and thus do not represent assembly errors in the melanistic reference. However, we focused subsequent analyses on LG8, which accounts for the overwhelming bulk of association.

We next inferred evolutionary history to better understand the dynamics by which *Mel-Stripe* evolved, and this revealed four core points (Fig. 4 and figs. S8 and S9). First, stochastic character state reconstruction revealed that the deletion most likely arose once in the ancestor of the Santa Barbara and Northern California clades (posterior probability = 0.73), and almost certainly arose no more than twice (posterior probability > 0.99; fig. S8). Second, as previously reported (24), there is no evidence for contemporary admixture inferred from a STRUCTURE-like analyses (Fig. 4 and fig. S9). Third, we inferred potential historical admixture at the *Mel-Stripe* locus in an explicitly phylogenetic context using TREEMIX

(25). This revealed that a phylogenetic tree lacking admixture for this locus explained 98% of the covariance in allele frequencies (Fig. 4). Thus, historical admixture for the *Mel-Stripe* locus in *Timema* is low or absent. Fourth, allowing migration revealed two potential admixture events involving the Santa Barbara and Northern California clades, and increasing the variance explained up to 99.9% of the covariance in allele frequencies. Critically, these two migration events were from deeper branches in the phylogeny to two terminal branches such that they cannot explain overall color variation among species (Figs. 1 and 3). Moreover, these inferred admixture events would have been so ancient that the deletion polymorphism would have been maintained for extended periods of time through processes other than introgression. Indeed, the chromosomal forms (i.e., morphs) of *T. cristinae* are ancient and maintained by long-term balancing selection (22, 24). Collectively, the weight of the evidence implies that balancing selection is a more important mechanism for the maintenance of polymorphism

in *Timema* than is introgression, contrasting other systems in which introgression strongly affects evolution (e.g., mimetic *Heliconius* butterflies, Darwin's finches, *Helianthus* sunflowers, and African cichlids) (18, 26–28).

Finally, we sought to distinguish whether multiple linked color genes or a single developmental switch control color, the former being a key assumption of both the supergene and supermutation hypotheses. We were drawn to *T. chumash* to test this assumption because it displays a wide range of color variations in addition to green and brown, including shades of yellow, pink, tan, beige, and blue (Fig. 1), and this species lacks the deletion observed in Northern California clade green-colored *Timema*. This led us to speculate that *T. chumash* may exhibit recombination among multiple color loci. This hypothesis predicts low genetic structure in the region harboring color genes (i.e., a lack of morph-associated genetic clusters), with low LD among color loci. As reported below, GBS data support these predictions, revealing that multiple linked but recombining loci affect color (Fig. 3 and figs. S2 to S5).

To infer the number of discrete genetic regions influencing color, we used a Bayesian multilocus GWA mapping approach in GEMMA that accounts for LD among SNPs (29). In contrast to past work in *T. cristinae* (22), we found evidence for multilocus control of color in *T. chumash* (Fig. 3 and figs. S6 and S7). Specifically, the region to which color maps contains multiple, distinct peaks of phenotype–genotype association, generally separated from each other by several kilobase pairs. Some peaks were associated with variation in only one trait (RG or GB), and accordingly the genetic correlation between RG and GB was moderate [ $R = -0.34$ , 95% confidence interval (CI) =  $-0.41$ ,  $-0.26$ ]. These results suggest that a contiguous region controls color but that control is multigenic.

Our mapping approach allowed us to explicitly estimate the number of genetic variants (i.e., quantitative trait nucleotides, QTNs) controlling each trait by considering how often SNPs were retained as trait associated across different Markov chain Monte Carlo (MCMC) steps in the GWA (the proportion of such steps is called the posterior inclusion probability, PIP hereafter; Fig. 3). In the case of multigenic control with recombination among loci, the one or few SNPs that best tag (by LD) each causal variant are expected to consistently be trait associated across MCMC steps (i.e., exhibit high PIP values). Thus, PIP values across such SNPs sum to an estimate of the number of total causal variants (i.e., even if causal variants are not unambiguously identified, the number of such variants can be estimated).

The above analysis revealed that in the region that is deleted in the Northern California clade, ~3 to 6 genetic variants control GB in

*T. chumash* (mean = 4.4, posterior SD  $\pm$  1.3) and ~2 to 5 control RG (mean = 3.6, posterior SD  $\pm$  1.4; Fig. 3, G and H, and fig. S6). By contrast, the number of causal variants predicted to be in the remainder of the *Mel-Stripe* locus in *T. chumash* was near zero (GB = 0.07, posterior SD  $\pm$  0.26; RG = 0.13, posterior SD  $\pm$  0.36). PIPs for SNPs in the deleted region for RG and GB were mostly uncorrelated (Pearson's  $r = 0.085$ , 95% CI =  $-0.092$ ,  $0.256$ ,  $P = 0.35$ ). Nine SNPs had PIPs > 0.1 for RG (7.1%), eight for GB (6.3%), and four for both (3.1%); this overlap exceeds null expectations from a randomization test (expected = 0.6%,  $P = 0.001$ ). Thus, whereas PIPs for most SNPs in this chromosomal region were not associated between the two different body colors, approximately half of the SNPs most associated with RG were also highly associated with GB.

Also consistent with a multigenic model, effect sizes were moderate and fairly uniformly distributed among the most strongly color-associated SNPs (fig. S6). Moreover, phenotypic color scores increasingly became more melanistic (defined by high scores for RG and low scores for GB) as the number of melanica-associated alleles an individual harbored increased (across the 10 most strongly color-associated SNPs), and we did not detect evidence for epistasis (fig. S7). Finally, LD among the top color-associated SNPs was low, indicative of recombination between them (Fig. 3B; see also fig. S10 for LD-based inferences on synteny). Thus, multiple linked but recombining variants affect GB and RG coloration in *T. chumash*.

Although the identities of the genes causally affecting color remain to be resolved, several genes are promising candidates (table S1), and their identities provide hints at how deletion of color loci could hypothetically lead to or maintain a green phenotype in *Timema*. One hypothesis is that the deletion prevents color pigment precursors from entering tegument color cells and constrains the development of such cells. This could lead to transparency of the tegument, giving the insect a green appearance by exposing the green hemolymph underneath. Future tests of this idea are required.

Our collective results support the supermutation hypothesis but they do not preclude a role for suppressed recombination because these hypotheses are not mutually exclusive. Indeed, closer examination of the data from *T. bartmani* (which lacks the deletion) revealed suppressed recombination in the region harboring coloration loci, as supported by genetic clustering by morph in principal components analysis, morph-associated haplotype blocks in phased structure analyses, high LD, and “block-like” patterns of association in single-SNP GWAS (note that these patterns all contrast with *T. chumash*, which exhibits recombination among color loci;

Fig. 3 and fig. S5). Thus, although large-scale mutation affected color loci in the Northern California clade, a supergene scenario is likely in *T. bartmani* in the Southern California clade. Our results highlight that whether the key role of structural variants in adaptation is to suppress recombination should be tested rather than assumed. Indeed, an example of a supergene controlling heterostyly in *Primula* involves not only suppressed recombination but also complex patterns of gene duplication (30, 31).

In conclusion, we demonstrate here multiple mechanisms controlling large evolutionary changes in *Timema* body coloration. Although developmental switches through gene regulation have been described in other systems (32), our results indicate that in *Timema*, discrete color morphs likely evolve through large-scale mutation of multiple genes. These mechanisms help to explain rapid evolutionary leaps and difficulties observing intermediate or transitional forms, thus helping to resolve one of Darwin's greatest concerns with his theory of evolution by natural selection (33). Moreover, our results complement, but are distinct from, accumulating examples of introgression leading to evolutionary novelty. Collectively, our findings illustrate how the evolution of gene complexes can package variation into discrete units of biodiversity. Although our focus here was on color morphs, similar processes could underlie the evolution of other forms of diversity, such as ecotypes or species.

#### REFERENCES AND NOTES

- R. D. H. Barrett, S. M. Rogers, D. Schluter, *Science* **322**, 255–257 (2008).
- Y. F. Chan et al., *Science* **327**, 302–305 (2010).
- R. D. H. Barrett et al., *Science* **363**, 499–504 (2019).
- J. Gratten et al., *Science* **319**, 318–320 (2008).
- S. Lamichhaney et al., *Science* **352**, 470–474 (2016).
- L. C. Coberly, M. D. Rauscher, *Evolution* **62**, 1076–1085 (2008).
- T. M. Korves et al., *Am. Nat.* **169**, E141–E157 (2007).
- R. D. H. Barrett, H. E. Hoekstra, *Nat. Rev. Genet.* **12**, 767–780 (2011).
- D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, *Science* **312**, 111–114 (2006).
- R. C. Lewontin, *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, 1974).
- M. V. Rockman, *Evolution* **66**, 1–17 (2012).
- L. H. Rieseberg, *Trends Ecol. Evol.* **16**, 351–358 (2001).
- M. A. F. Noor, K. L. Grams, L. A. Bertucci, J. Reiland, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12084–12088 (2001).
- M. Kirkpatrick, *PLOS Biol.* **8**, e1000501 (2010).
- T. Schwander, R. Libbrecht, L. Keller, *Curr. Biol.* **24**, R288–R294 (2014).
- D. Ayala, R. F. Guerrero, M. Kirkpatrick, *Evolution* **67**, 946–958 (2013).
- C. R. Lee et al., *Nat. Ecol. Evol.* **1**, 0119 (2017).
- P. Jay et al., *Curr. Biol.* **28**, 1839–1845.e3 (2018).
- R. F. Guerrero, F. Rousset, M. Kirkpatrick, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 430–438 (2012).
- C. P. Sandoval, *Biol. J. Linn. Soc. Lond.* **52**, 341–356 (1994).
- A. A. Comeault et al., *Curr. Biol.* **25**, 1975–1981 (2015).
- D. Lindtke et al., *Mol. Ecol.* **26**, 6189–6205 (2017).
- P. Nosil et al., *Science* **359**, 765–770 (2018).
- R. Riesch et al., *Nat. Ecol. Evol.* **1**, 82 (2017).
- J. K. Pickrell, J. K. Pritchard, *PLOS Genet.* **8**, e1002967 (2012).
- S. Lamichhaney et al., *Science* **359**, 224–228 (2018).
- J. I. Meier et al., *Nat. Commun.* **8**, 14363 (2017).

28. K. D. Whitney, R. A. Randell, L. H. Rieseberg, *New Phytol.* **187**, 230–239 (2010).
29. X. Zhou, P. Carbonetto, M. Stephens, *PLOS Genet.* **9**, e1003264 (2013).
30. J. Li *et al.*, *Nat. Plants* **2**, 16188 (2016).
31. B. McClure, *Nat. Plants* **2**, 16184 (2016).
32. H. F. Nijhout, *Evol. Dev.* **5**, 579–592 (2003).
33. C. Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, 1859).
34. Z. Gompert *et al.*, *Mol. Ecol.* **23**, 4555–4573 (2014).
35. Data for: R. Villoutreix, C. F. de Carvalho, V. Soria-Carrasco, D. Lindtke, M. De-la-Mora, M. Muschick, J. L. Feder, T. L. Parchman, Z. Gompert, P. Nosil, Large-scale mutation in the evolution of a gene complex for cryptic coloration, *Dryad* (2020); <https://doi.org/10.5061/dryad.pk0p2ngkf>.

#### ACKNOWLEDGMENTS

We thank T. Reimchen, M. Joron, and D. Ayala for discussions and comments on previous versions of the manuscript; T. Oakley for lab space; and Ó. Mira Pérez, L. Lloyd, J. Thakrar, F. Whiting, and A. Štambuk for assistance with lab work. The support and

resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged, as well as access to the High Performance Computing Facilities, particularly to the Iceberg and ShARC HPC clusters, from the Corporate Information and Computing Services at the University of Sheffield. **Funding:** This work was supported by a grant from the European Research Council (NatHisGen R/129639, <https://erc.europa.eu/>) and by a fellowship from the Royal Society of London to P.N. V.S.-C. was supported by a Leverhulme Trust Early Career Fellowship (ECF-2016-367). M.M. was supported by an SNSF Advanced Postdoc.Mobility fellowship (P300P3\_147888). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. **Author contributions:** R.V., C.F.d.C., Z.G., and P.N. conceived the project. R.V., C.F.d.C., V.S., D.L., M.M., and P.N. collected data. R.V., C.F.d.C., V.S., M.D., and Z.G. led data analysis, aided by all authors. All authors contributed to writing. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** DNA sequences have been deposited in the NCBI SRA (PRJNA638692). Data processing and analysis scripts have been archived in the Dryad digital repository (35). This includes bash and Perl scripts for DNA sequence alignment, variant

calling and variant filtering, scripts for running GEMMA and TREEMIX, and R scripts for fitting hidden Markov models to the coverage data. SNP genotype matrices, population allele frequency data, and nongenetic data (i.e., color data for association mapping) were also archived in the same Dryad repository (35). Correspondence and requests for materials should be sent to P.N.

#### SUPPLEMENTARY MATERIALS

[science.sciencemag.org/content/369/6502/460/suppl/DC1](https://science.sciencemag.org/content/369/6502/460/suppl/DC1)  
Materials and Methods  
Supplementary Text  
Figs. S1 to S10  
Tables S1 to S7  
References (36–81)  
MDAR Reproducibility Checklist

[View/request a protocol for this paper from Bio-protocol.](#)

17 September 2019; resubmitted 29 February 2020  
Accepted 19 May 2020  
10.1126/science.aaz4351

## Large-scale mutation in the evolution of a gene complex for cryptic coloration

Romain Villoutreix, Clarissa F. de Carvalho, Víctor Soria-Carrasco, Dorothea Lindtke, Marisol De-la-Mora, Moritz Muschick, Jeffrey L. Feder, Thomas L. Parchman, Zach Gompert and Patrik Nosil

*Science* **369** (6502), 460-466.  
DOI: 10.1126/science.aaz4351

### How stick insects got their colors

Stick insects of the genus *Timema* show variations in color that are likely due to selection for camouflage on their plant hosts. Some species are only brown or only green, whereas others span from reddish-pink hues to green to brown. To identify the genetics underlying this variation, Villoutreix *et al.* sequenced most extant California *Timema* species and identified a deletion that correlated with green body coloration in some species. However, this deletion appeared to be limited to the Northern California clade, with more species in the Southern California clade retaining this locus. In these southern species, a nongreen body color is likely caused by mutation. Even among closely related species, adaptive evolution can converge on the same phenotype from different evolutionary processes.

*Science*, this issue p. 460

#### ARTICLE TOOLS

<http://science.sciencemag.org/content/369/6502/460>

#### SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2020/07/22/369.6502.460.DC1>

#### REFERENCES

This article cites 74 articles, 14 of which you can access for free  
<http://science.sciencemag.org/content/369/6502/460#BIBL>

#### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

---

*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works