Supplementary Information

Comeault *et al*., "Color phenotypes are under similar genetic control in two distantly related species of *Timema* stick insect"

This file contains Figures S1 and S2, Tables S1 through S3, Supplementary Methods, and Supplementary References.

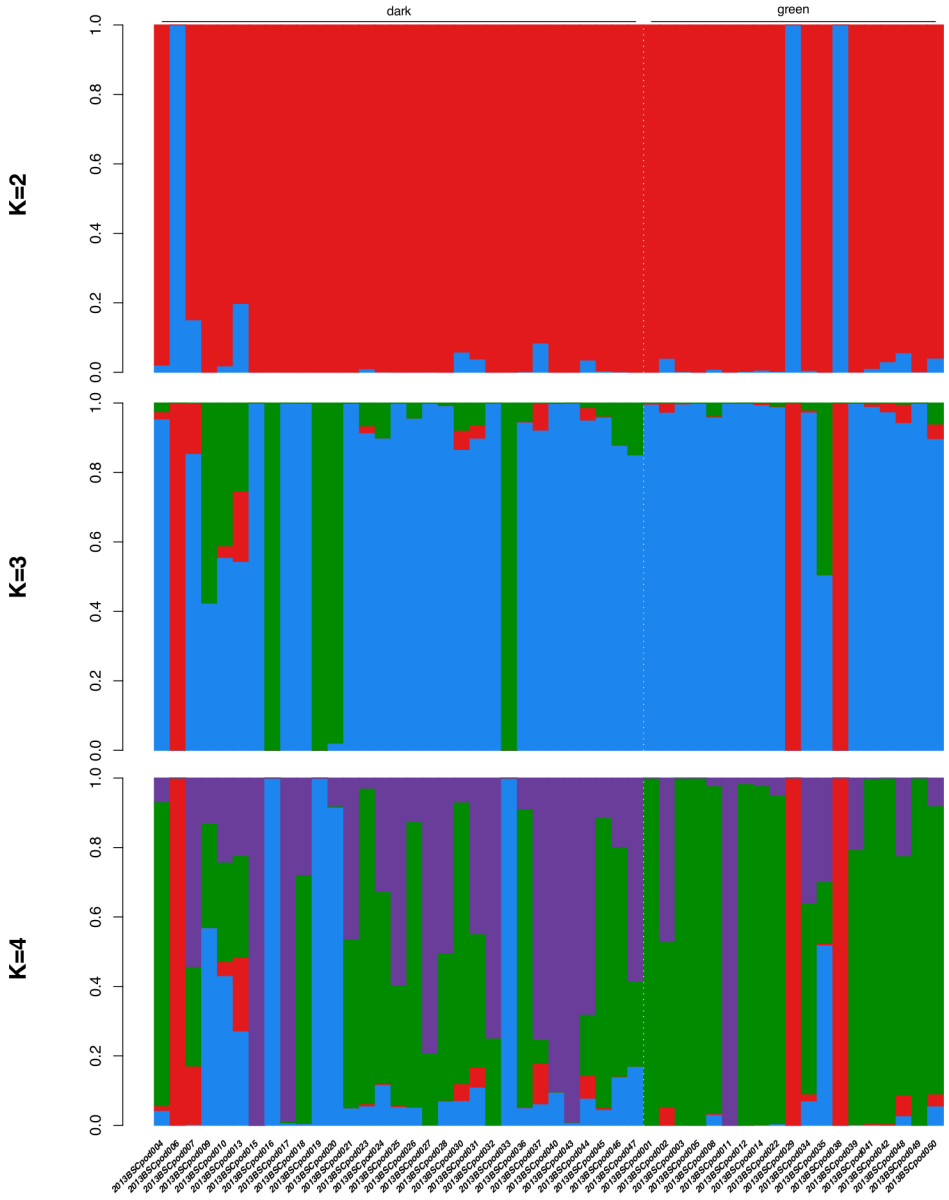

Figure S1. Admixture proportions estimated for 50 individuals genotyped at 56,149 SNPs using hierarchical Bayesian models with *K*=2-4. Colors depict population classifications for each of the *K* genetic clusters. Individuals are separated by color phenotype (color is indicated by horizontal lines along the top of the figure).

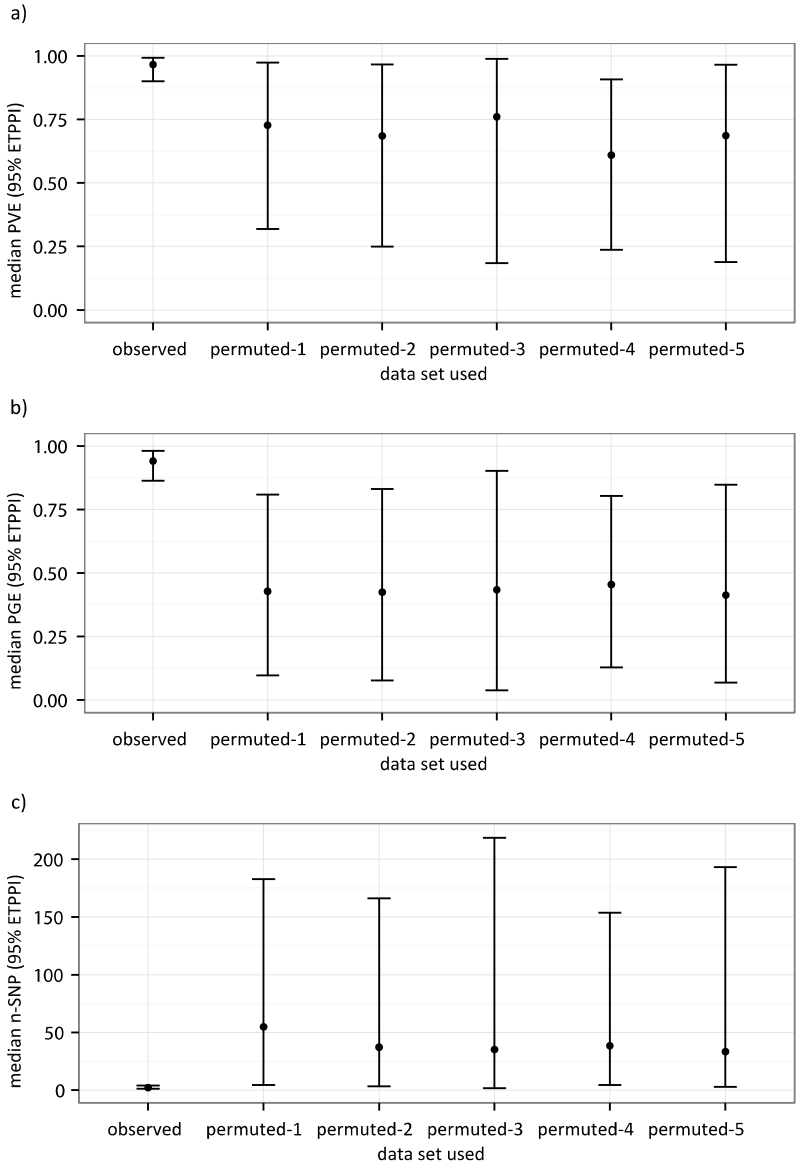

Figure S2. Median and 95% equal-tail posterior probability interval (ETPPI) of hyperparameter estimates from Bayesian sparse linear mixed model GWA mapping carried out on our original *T. podura* GBS data sets (‘observed’) and five data sets where phenotypic values was randomly permuted among individuals (permuted-1 through permuted-5). a) PVE, b) PGE, and c) n-SNP.

Table S1. Deviance information criterion (DIC) model selection results of the admixture analysis for *K*=1-4 based on 56,149 SNPs from 50 individuals.

K	**Mean deviance**	**Effective number of parameters**	**DIC**
1	4622904.8	759313	5382217.8
2	4554698.4	1931507	6486205.4
3	4486048.0	3263618	7749666.0
4	4384014.2	5432806	9816820.2

Table S2. Results of PCA carried out on the matrix of imputed genotypes of the 50 *T. podura*. Cumulative R^2 represents the total proportion of variation in the genetic data set explained by the given number of PCs, while pseudo R^2 represents the proportional increase in residual deviance of color phenotype explained by including a given PC as an explanatory variable in a generalized linear model predicting color relative to a null model lacking that PC. Whether a PC was associated with color phenotype is based on a likelihood ratio test with a Bonferroni-corrected significance level of $p = 0.0036$.

PC	Q^2	cumulative R^2	associated with phenotype?	psuedo R^2 (%)
1	0.234	0.2659	no	n/a
2	0.223	0.2948	no	n/a
3	0.209	0.3210	no	n/a
4	0.196	0.3454	$p < 0.0001$	25.33
5	0.184	0.3680	no	n/a
6	0.168	0.3891	no	n/a
7	0.154	0.4095	$p = 0.0009$	16.66
8	0.139	0.4291	no	n/a
9	0.125	0.4480	no	n/a
10	0.112	0.4664	no	n/a
11	0.098	0.4845	no	n/a
12	0.084	0.5020	no	n/a
13	0.070	0.5191	no	n/a
14	0.055	0.5359	no	n/a

Table S3. Candidate SNPs associated with color phenotypes in *T. podura* as identified through multi-locus and single-SNP GWA mapping. SNP ID denotes the position in v0.3 of the *T. cristinae* genome assembly and gives the position of the SNP in the form of linkage group_order_scaffold-position. Strengths of associations are given as posterior inclusion probabilities for the multi-SNP analysis and *p*-values from the single-SNP analysis.

Analysis	SNP ID	PIP/*p*-value	Effect size
multi-locus	lg8_ord104_scaf1806-10972	0.295	9.92
multi-locus	lg8_ord126_scaf284-349343	0.102	4.25
single-SNP	lg8_ord101_scaf1154-30072	9.34×10^{-07}	0.130
single-SNP	lg10_ord41_scaf380-189546	1.73×10^{-07}	0.117

Supplementary Methods

Collecting standardized digital images and recording color

Images were recorded in RAW format with a Canon 600d camera equipped with an EF-S 60mm F2.8 macro lens (Canon (UK) Ltd., Surry, UK) with an aperture of f/14, a shutter speed of 1/250s, and two wireless external flashes (Yongnuo YN560-II speedlight, Yongnuo Digital, www.yongnuo.eu) set to manual. Each image was captured with 1 cm grid paper as a background and included a standard color chip (Colorgauge Micro, Image Science Associates LLC, Williamson, NY, USA). Following image capture each individual was placed in an individually labeled 1.5 ml microcentrifuge tube and preserved in 95% ethanol for subsequent DNA extraction. Images were linearized and corrected to 80% reflectance based on a neutral gray color target (target #10 of the cologauge micro color chip) with Adobe's Photoshop Lightroom 4 software (Adobe Systems Software Ireland Ltd), and saved as a .tif file.

We measured color for each individual from the lateral margin of the second thoracic segment using the polygon selection tool in ImageJ (Abràmoff et al. 2004). We then recorded mean RGB values for this region using the color histogram plugin in ImageJ. For each individual measurement we converted raw RGB values to variables representing two color channels and one luminance channel as suggested by Endler (2012).

Library preparation for sequencing on the Illumina HiSeq 2000 platform

Briefly, restriction-site associated DNA libraries were generated by digesting genomic DNA in the presence of the restriction enzymes *Eco*RI and *Mse*I (New England Biolabs), ligating double stranded adapters containing the Illumina priming site and one of 50 unique 8 to 10 base pair (bp) barcode sequence to the restriction fragments, and amplifying ligated fragments using the polymerase chain reaction (PCR). For a detailed protocol refer to Parchman et al. (2012).

Determining genetic structure within the T. podura population sample

We used a hierarchical Bayesian model that jointly estimates genotypes and admixture proportions as implemented in the program 'entropy' (Gompert et al. 2014).We ran five independent Markov chain Monte Carlo (MCMC) chains for each $K = 1$ through 4 using 56,149 SNPs for which we have sequence data from at least 95% of the 50 samples. We set the scalar of the Dirichlet initial value of q to 50 and used as starting admixture proportions the values obtained by applying linear discriminant analysis on a covariance matrix of composite genotypes estimated assuming Hardy-Weinberg equilibrium. We ran each chain for 50,000 steps and took samples every 10^{th} iteration. We examined visually the posterior deviance traces to ensure convergence and mixing. We discarded the first 4,000 samples (40,000 iterations) from each chain as burn-in, and combined the five chains (5,000 samples in total) to estimate model parameters.

We then used DIC to determine the most appropriate number of genetic clusters represented in our data. DIC penalizes model complexity by adding to the posterior mean deviance the effective number of parameters, which can in turn be approximated as half the posterior variance of the deviance (Gelman et al. 2013), and has been shown to outperform other methods for identifying the optimal number of clusters (Gao et al. 2011).

Cross-validation of multi-locus genome-wide association mapping analysis.

First, we conducted a permutation test using GWA mapping in *gemma* as described in the main text using five phenotypic data sets generated by randomly permuting the phenotypic scores for each individual.

Second, we performed cross validation using the genomic prediction function in *gemma* to predict phenotypes of individuals whose color phenotype was randomly masked. Following standard GWA runs in *gemma* we performed cross-validation analyses to test the predictive power of our GWA mapping. The approach is akin to that commonly taken in genomic prediction/genomic selection studies (Wray et al., 2010). A predicted phenotype was estimated for each individual by randomly masking 10% of individual phenotypes ('test set') (Zhou et al., 2013) and using the remaining 90% of phenotypes ('training set') to obtain model parameters in *gemma* using the same parameters as in the standard runs. These parameters were then used in *gemma* to obtain predicted phenotypic values using the '–predict' option (Zhou et al., 2013). In each instance, we ran five replicate MCMC chains for each training set and repeated this procedure 10 times (i.e., until predicted values were obtained for every individual). The entire process was then repeated five times with different random combinations of individuals in each training set to avoid any potential 'training set' biases, resulting in a total of 25 predicted phenotypes for each observed phenotype.

The reliability of genomic prediction was then estimated by calculating the area under the receiver operating characteristic curve (AUC). Briefly, the method plots correctly predicted phenotypes – i.e. the true positive rate (TPR or sensitivity) - against incorrectly assigned phenotypes – i.e. the false-positive rate (FPR or 1-specificity). The AUC is then calculated as the proportion of phenotypes correctly assigned, or falling under the ROC curve (Wray et al., 2010). Calculation of AUC was performed using function performance from the package ROCR (Sing et al., 2005).

Supplementary References:

Abràmoff, M. D., P. J. Magalhães, and S. J. Ram. 2004. Image Processing with ImageJ. Biophotonics Int. 11:36–42.

Aulchenko, Y. S., S. Ripke, A. Isaacs, and C. M. van Duijn. 2007. GenABEL: an R library for genome-wide association analysis. Bioinformatics 23:1294–1296.

Comeault, A. A., S. M. Flaxman, R. Riesch, C. Emma, V. Soria-Carrasco, Z. Gompert, T. E. Farkas, M. Muschick, T. L. Parchman, J. Slate, and P. Nosil. 2015. Selection on a genetic polymorphism counteracts ecological speciation in a stick insect. Curr. Biol. 25:1975–1981.

Endler, J. A. 2012. A framework for analysing colour pattern geometry: adjacent colours. Biol. J. Linn. Soc. 107:233–253.

Gao, H., K. Bryc, and C. Bustamante. 2011. On identifying the optimal number of population clusters via the deviance information criterion. PLoS One 6:e21014.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. Evaluating, comparing, and expanding models. Pp. 165–193 *in* Bayesian Data Analysism(3rd Edn.). CRC Press.

Gompert, Z., L. K. Lucas, C. A. Buerkle, M. L. Forister, J. a Fordyce, and C. C. Nice. 2014. Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. Mol. Ecol. 23:4555–4573.

Parchman, T. L., Z. Gompert, J. Mudge, F. D. Schilkey, C. W. Benkman, and C. A. Buerkle. 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. Mol. Ecol. 21:2991–3005.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. a Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904–909.

Sing, T., Sander, O., Beerenwinkel, N., T. Lengauer 2005 ROCR: visualizing classifier performance in R. *Bioinformatics* 21(20):3940-3941

Wray, N.R., et al. 2010 The Genetic Interpretation of Area under the ROC Curve in Genomic Profiling. *Plos Genetics* **6**(2)

Zhou, X., P. Carbonetto, and M. Stephens. 2013 Polygenic Modeling with Bayesian Sparse Linear Mixed Models. *Plos Genetics* **9**(2).