

Supporting Information

Hines et al. 10.1073/pnas.1110096108

SI Text

Population Structure Inferred Using STRUCTURE. Methods. Individuals of both *Heliconius melpomene* and *Heliconius erato* were assigned to populations in separate analyses for *optix* and the combined nuclear unlinked markers using the program STRUCTURE 2.2. For these analyses we ran a linkage model (1) (LOG10RSTART = -7, LOG10RMAX = -2, LOG10RMIN = -10, LOG10RPROPSD = 0.2) to account for proximity of SNPs within genes to each other and ran 70,000 generations, with 20,000 generations as burn-in. As with the analysis of molecular variance (AMOVA), outgroups, *Heliconius himera*, and *Heliconius erato chesteronii* were excluded. Analyses were run for *optix* three times for each of one to six populations for each species. We used the point where likelihoods plateaued and additional populations provided only minor contributions to choose an optimal number of populations ($n = 3$) to compare and display for all analyses. Three separate runs assuming three populations were run for unlinked markers. Each replicate gave nearly identical results so a single run was chosen to display results.

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

Results. Population assignment based on the combined unlinked nuclear markers revealed the same geographic patterns seen in the phylogenetic analysis (Fig. S2). For *H. erato*, the structure is weak, but there is a tendency to distinguish the Amazonian+Chacoan/Parana lineages from Caribbean lineages. In *H. melpomene*, populations are highly structured into Caribbean, Amazonian, and Guiana Shield (French Guiana, Trinidad)+Chacoan/Parana lineages. In contrast, *optix* sequences split individuals of both species into lineages structured by both color pattern and geography. For *optix* this includes, (i) all Amazonian rayed lineages + some *Heliconius erato microclea* + some French Guianan *Heliconius erato hydara*; (ii) some Caribbean and non-rayed Amazonian lineages; and (iii) remaining Caribbean individuals. For *H. melpomene* this includes: (i) the rayed races + *Heliconius melpomene xenoclea* + *Heliconius melpomene plesseni*; (ii) *Heliconius melpomene amaryllis* + *Heliconius melpomene nanna*, which are the yellow banded Amazonian races; and (iii) the Caribbean lineages + Amazonian *Heliconius melpomene melpomene*.

Fig. S1. (A–J) Neighbor-joining (B–E and G–J) and Bayesian (A and F) phylogenies of each gene for haplotypes of *H. erato* and *H. melpomene*. Branches and taxa are colored by phenotypes indicated in A. Inset in A is a neighbor-joining phylogeny using *optix* sequences combined across both species, demonstrating that the lineages do not share alleles. Bars at the right indicate geographic distribution in either the Amazon, Caribbean, or Chacoan/Parana, and distinguishes *H. himera* and outgroups separately. Diagonal dividers are used for multiphenotype haplotypes/polytomies. Support values on Bayesian trees are posterior probabilities. Slanted lines on branches indicate that some of the branch length was removed for presentation. Taxa are represented by their voucher number, the first three letters of the race, and haplotype within each individual. Bayesian phylogenies were generated in MrBayes v.3.1.2 using a separate model for each gene assigned using the hierarchical likelihood ratio test in Modeltest 3.7 (1). Models used for inferring Bayesian trees included mitochondrial models HKY+I+G for *H. erato* and GTR+I+G for *H. melpomene* and *optix* models, HKY+G for *H. erato* and HKY+I+G for *H. melpomene*. For the Bayesian analyses, we performed three runs with four chains and 3,000,000 million generations for each analysis. For all runs, the three trees converged after a 10% burn-in, ascertained using AWTY (2) and Tracer (3). Nexus and tree files are available in Dryad.

Fig. S1

1. Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
2. Wilgenbusch JC, Warren DL, Swofford DL (2004) AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. <http://ceb.csit.fsu.edu/awty>. Accessed February 2011.
3. Rambaut A, Drummond AJ (2007) Tracer v1.4. Available from <http://beast.bio.ed.ac.uk/Tracer>.

Fig. S2. Population assignment of each individual for *optix* and unlinked nuclear markers using STRUCTURE. CP, color pattern phenotype (rayed vs. nonrayed); GEO, Geographic regions; GS&C, Guianan Shield (Trinidad, French Guiana) and Chacoan/Parana. Black lines within bar graphs separate different color pattern races, indicated by their first three letters: agl, aglaope; ama, amaryllis; amp, amphitrite; cyr, cyrbia; cyt, cythera; dig, dignus; ecu, ecuadorensis; emm, emma; era, erato; ety, etylus; fav, favorinus; hydE, French Guiana *H. erato hydara*; hydW, Caribbean *H. erato hydara*; lat, lativitta; meE, French Guiana and Trinidad *H. melpomene melpomene*; melW, remaining Caribbean *H. melpomene melpomene*; mic, microclea; pet, petiverana; nan, nanna; phy, phyllis; ple, plesseni; ros, rosina; the, thelxiopea; ven, venus; vul, vulcanus; xen, xenoclea.

Fig. S2

Table S2. Diversity indices for each species and wing-pattern phenotype

<i>H. erato</i>					<i>H. melpomene</i>						
	θ_W /bp	θ_π /bp				θ_W /bp	θ_π /bp		θ_W /bp	θ_π /bp	
Linked*			Unlinked*			Linked*			Unlinked*		
<i>H. erato</i>	0.0255	0.0153	<i>H. erato</i>	0.0263	0.0125	<i>H. melpomene</i>	0.0124	0.0063	<i>H. melpomene</i>	0.0154	0.0086
Rayed	0.0107	0.0074	Rayed	0.0171	0.0096	Rayed	0.0056	0.0031	Rayed	0.0089	0.0069
Red band	0.0197	0.0137	Red band	0.0179	0.0111	Red band	0.0076	0.0063	Red band	0.0096	0.0077
<i>bves</i>			<i>2654</i>			<i>bves</i>			<i>2654</i>		
<i>H. erato</i>	0	0.0091	<i>H. erato</i>	0.0202	0.0092	<i>H. melpomene</i>	0.0109	0.0027	<i>H. melpomene</i>	0.0121	0.0059
Rayed	0.0099	0.0053	Rayed	0.0126	0.0087	Rayed	0.0019	0.0006	Rayed	0.0045	0.0032
Red band	0.0136	0.0096	Red band	0.0109	0.0078	Red band	0.0052	0.0028	Red band	0.0075	0.0054
<i>kinesin</i>			<i>cat</i>			<i>kinesin</i>			<i>cat</i>		
<i>H. erato</i>	0.0282	0.0138	<i>H. erato</i>	0.0220	0.0135	<i>H. melpomene</i>	0.0114	0.0096	<i>H. melpomene</i>	0.0133	0.0059
Rayed	0.0128	0.0069	Rayed	0.0116	0.0061	Rayed	0.0055	0.0037	Rayed	0.0079	0.0047
Red band	0.0183	0.0134	Red band	0.0144	0.0123	Red band	0.0088	0.0100	Red band	0.0058	0.0041
<i>gpcr</i>			<i>sumo</i>			<i>gpcr</i>			<i>sumo</i>		
<i>H. erato</i>	0.0242	0.0174	<i>H. erato</i>	0.0367	0.0161	<i>H. melpomene</i>	0.0092	0.0037	<i>H. melpomene</i>	0.0193	0.0139
Rayed	0.0173	0.0119	Rayed	0.0257	0.0143	Rayed	0.0042	0.0029	Rayed	0.0146	0.0136
Red band	0.0197	0.0177	Red band	0.0309	0.0148	Red band	0.0077	0.0035	Red band	0.0131	0.0128
<i>optix</i>			<i>suz12</i>			<i>optix</i>			<i>suz12</i>		
<i>H. erato</i>	0.0265	0.0190	<i>H. erato</i>	0.0296	0.0104	<i>H. melpomene</i>	0.0145	0.0079	<i>H. melpomene</i>	0.0197	0.0107
Rayed	0.0057	0.0029	Rayed	0.0225	0.0104	Rayed	0.0065	0.0032	Rayed	0.0094	0.0073
Red band	0.0233	0.0134	Red band	0.0172	0.0084	Red band	0.0080	0.0072	Red band	0.0167	0.0121
<i>Van Gogh</i>						<i>Van Gogh</i>					
<i>H. erato</i>	0.0255	0.0142				<i>H. melpomene</i>	0.0137	0.0061			
Rayed	0.0107	0.0105				Rayed	0.0079	0.0042			
Red band	0.0206	0.0137				Red band	0.0076	0.0064			

Estimates of nucleotide diversity within species and among the rayed and red-banded lineages of both species were conducted for each gene using SITES (1). For nucleotide diversity we calculated the Watterson's θ_W (2), an estimator of $4N_e\mu$ using the number of segregating sites and θ_π , the average number of differences between all haplotypes. θ_W /bp = Watterson's estimator of $4N_e\mu$, per base pair; θ_π /bp = average number of pairwise differences, per base pair. *Multilocus estimates of nucleotide diversity based on concatenated sequences.

1. Wakeley J, Hey J (1997) Estimating ancestral population parameters. *Genetics* 145:847–855.
2. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276.

Table S3. Recombination indices for each species and wing-pattern phenotype

<i>H. erato</i>				<i>H. melpomene</i>							
	γ /bp	Min		γ /bp	Min		γ /bp	Min			
Linked*			Unlinked*			Linked*			Unlinked*		
<i>H. erato</i>	0.1224	36	<i>H. erato</i>	0.1150	51	<i>H. melpomene</i>	0.0509	13	<i>H. melpomene</i>	0.0606	16
Rayed	0.0404	13	Rayed	0.0739	15	Rayed	0.1042	6	Rayed	0.0272	12
Red band	0.0995	31	Red band	0.1308	38	Red band	0.0389	13	Red band	0.0744	11
<i>bves</i>			<i>2654</i>			<i>bves</i>			<i>2654</i>		
<i>H. erato</i>	0.0191	7	<i>H. erato</i>	0.0144	10	<i>H. melpomene</i>	0.0000	0	<i>H. melpomene</i>	0.0026	1
Rayed	0.0026	1	Rayed	0.0061	3	Rayed	0.0000	0	Rayed	0.0096	2
Red band	0.0147	5	Red band	0.0104	6	Red band	0.0000	0	Red band	0.0010	2
<i>kinesin</i>			<i>cat</i>			<i>kinesin</i>			<i>cat</i>		
<i>H. erato</i>	0.0054	6	<i>H. erato</i>	0.0044	12	<i>H. melpomene</i>	0.0037	4	<i>H. melpomene</i>	0.0092	8
Rayed	0.0014	1	Rayed	0.0013	3	Rayed	0.0000	0	Rayed	0.0063	6
Red band	0.0035	3	Red band	0.0041	6	Red band	0.0060	4	Red band	0.0080	3
<i>gpcr</i>			<i>sumo</i>			<i>gpcr</i>			<i>sumo</i>		
<i>H. erato</i>	0.0563	7	<i>H. erato</i>	0.0779	22	<i>H. melpomene</i>	0.0000	0	<i>H. melpomene</i>	0.0310	2
Rayed	0.0119	4	Rayed	0.0547	7	Rayed	0.0000	0	Rayed	0.0113	4
Red band	0.0455	5	Red band	0.0889	20	Red band	0.0000	0	Red band	0.0355	2
<i>optix</i>			<i>suz12</i>			<i>optix</i>			<i>suz12</i>		
<i>H. erato</i>	0.0065	10	<i>H. erato</i>	0.0183	7	<i>H. melpomene</i>	0.0053	3	<i>H. melpomene</i>	0.0179	5
Rayed	0.0000	0	Rayed	0.0118	2	Rayed	0.0288	2	Rayed	0.0000	0
Red band	0.0069	8	Red band	0.0274	6	Red band	0.0036	3	Red band	0.0300	4
<i>Van Gogh</i>						<i>Van Gogh</i>					
<i>H. erato</i>	0.0351	6				<i>H. melpomene</i>	0.0419	6			
Rayed	0.0245	7				Rayed	0.0754	4			
Red band	0.0289	10				Red band	0.0293	6			

Estimates of recombination within species and among the rayed and red-banded lineages of both species were conducted for each gene using SITES (1). A coalescent estimator of the population recombination rate per generation per base pair, γ /bp (1), was used to estimate the number of crossing over events per generation. We also used the Hudson's four-gamete test (2) in SITES to determine the minimum number of recombination intervals (Min).

*Multilocus γ /bp is an average of individual gene values, and the multilocus min is the sum of the individual gene values.

1. Wakeley J, Hey J (1997) Estimating ancestral population parameters. *Genetics* 145:847–855.

2. Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–164.

Table S4. Primer and PCR information for each gene

Gene	Aligned fragment size (bp)	Primer name	Sequence (5' to 3')	Annealing temperature (°C)
<i>optix</i>	794/802*	Optix_coding_D_387853_F3 Optix_coding_D_388709_R3	AATGCGTCCAGAAGGCATAC CCGAGAGCTCTACTCGATCC	53
<i>kinesin</i>	501/504*	GPCRInt_Kinesin_E1_F GPCRInt_Kinesin_E1_R	TTTATTGAACCTAGGCCACC AGTTAAGTCTATCAGCACCTCT	50–51
<i>VanGogh</i>	715	Hm_21P16_Gn5_Gn4_F1 Hm_21P16_Gn5_Gn4_R1	TAGCTTGTGCCTTCTCTGAGA GGGATGGGCAAGAAGTTA	50–51
<i>GPCR</i>	522	Hmel_28L23_con1_Gn15_Fb Hmel_28L23_con1_Gn15_Rb	GTTACACATGCCCGGTGATAA CGTCTCTCAGCCTCATTG	50–51
<i>bves</i>	385	bves11_F bves11_R	CAACAAGTAAAACCTGCACAGCA ACTGGCTTGCAGAATGTCAC	52
<i>SUMO</i>	805	SUMO_F SUMO_R	CCAAATCCGCTTATGG GAAGAAAAACATGTTATTAT	50–51
<i>Suz12</i>	520	SUZ12-F115 SUZ12-R678	ACGAGTTCACGGATGTCA ATATGGAGGACCGTTTGC	50–51
<i>2654</i>	872	2654_F 2654_R	AAAATGGTATTGGAAAGTAC GTAGACATAGCATTCTGAT	50
<i>CAT</i>	1,081	CAT_F CAT_R	TCAAGACTGCGATTCAAACA TGTCTTCAGTTTGCCACT	50–51
<i>mt (COI-tRNA^{leu})</i>	1,510	C1-J-2183F [†] TL2-N3014R [†]	CAACATTTATTTTGATTTTTGG TCCAATGCACTAATCTGCCATATTA	50–54
<i>(tRNA^{leu}-COII)</i>		C1-J-2783_HH_F [†] C2-N-3812R [†]	TAGGITTAGCTGGWATACCTCG CATTAGAAGTAATTGCTAATTTACTA	50–51

PCRs were run as 8- μ L reactions using Qiagen Mastermix and with the following protocol: 94 °C for 4 min; 35 cycles of 94 °C for 30 s, annealing temperatures listed in the table for 30 s, and 72 °C for 1 min, followed by 72 °C for 5 min. Gene choice and primer design for *VanGogh*, *kinesin*, *GPCR*, and *Suz12* was based on Counterman et al. (1). *SUMO*, *CAT*, and *2654* were based on Salazar et al. (2). *bves* is orthologous to *Drosophila* FBgn0031150 and includes newly designed primers. The information for *optix* is provided in Reed et al. (3).

*Gene with indel variation; *optix/melpomene*.

[†]Beltran et al. (4).

[‡]Modified from Beltran et al. (4) and Brower (5).

- Counterman BA, et al. (2010) Genomic hotspots for adaptation: The population genetics of Müllerian mimicry in *Heliconius erato*. *PLoS Genet* 6:e1000796.
- Salazar C, et al. (2010) Genetic evidence for hybrid trait speciation in *Heliconius* butterflies. *PLoS Genet* 6:e1000930.
- Reed RD, et al. (2011) *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* 333:1137–1141.
- Beltrán M, et al. (2002) Phylogenetic discordance at the species boundary: Comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Mol Biol Evol* 19:2176–2190.
- Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proc Natl Acad Sci USA* 91:6491–6495.