Wing patterning gene redefines the mimetic history of *Heliconius* butterflies

Heather M. Hines[a,1], Brian A. Counterman[b], Riccardo Papa[c], Priscila Albuquerque de Moura[d], Marcio Z. Cardoso[d], Mauricio Linares[e], James Mallet[f,g], Robert D. Reed[h], Chris D. Jiggins[i], Marcus R. Kronforst[j], and W. Owen McMillan[a,k]

[a]Department of Genetics, North Carolina State University, Raleigh, NC 27695; [b]Department of Biological Sciences, Mississippi State University, Mississippi State, MS 39762; [c]Department of Biology and Center for Applied Tropical Ecology and Conservation, University of Puerto Rico-Rio Piedras, Rio Piedras, Puerto Rico PR 00931; [d]Departamento de Botânica, Ecologia e Zoologia, Universidade Federal do Rio Grande do Norte, Natal RN 59072-970, Brazil; [e]Facultad de Ciencias Naturales y Matemáticas, Universidad del Rosario, Carrera 24 No. 63c-69, Bogotá Colombia; [f]Department of Genetics, Evolution and Environment, University College London, London, WC1E 6BT United Kingdom; [g]Department of Organismal and Evolutionary Biology, and [j]Faculty of Arts and Sciences Center for Systems Biology, Harvard University, Cambridge MA 02138; [h]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697; [i]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, United Kingdom; and [k]Smithsonian Tropical Research Institute, Apartado Postal 2072, Balboa, Panama

Edited by May R. Berenbaum, University of Illinois at Urbana–Champaign, Urbana, IL, and approved October 12, 2011 (received for review June 22, 2011)

The mimetic butterflies *Heliconius erato* and *Heliconius melpomene* have undergone parallel radiations to form a near-identical patchwork of over 20 different wing-pattern races across the Neotropics. Previous molecular phylogenetic work on these radiations has suggested that similar but geographically disjunct color patterns arose multiple times independently in each species. The neutral markers used in these studies, however, can move freely across color pattern boundaries, and therefore might not represent the history of the adaptive traits as accurately as markers linked to color pattern genes. To assess the evolutionary histories across different loci, we compared relationships among races within *H. erato* and within *H. melpomene* using a series of unlinked genes, genes linked to color pattern loci, and *optix*, a gene recently shown to control red color-pattern variation. We found that although unlinked genes partition populations by geographic region, *optix* had a different history, structuring lineages by red color patterns and supporting a single origin of red-rayed patterns within each species. Genes closely linked (80–250 kb) to *optix* exhibited only weak associations with color pattern. This study empirically demonstrates the necessity of examining phenotype-determining genomic regions to understand the history of adaptive change in rapidly radiating lineages. With these refined relationships, we resolve a long-standing debate about the origins of the races within each species, supporting the hypothesis that the red-rayed Amazonian pattern evolved recently and expanded, causing disjunctions of more ancestral patterns.

Müllerian mimicry | population genetics | phylogeography

Researchers typically rely on neutrally evolving loci to generate a phylogenetic and population genetic history of adaptive divergence. The rationale is that these markers provide an unbiased view of the relationships among divergent phenotypes and a better understanding of the evolutionary processes generating variation. However, the genome is a complicated mosaic shaped by an interplay of mutation, drift, selection, and recombination. Recombination allows different regions of the genome to experience alternative restrictions to gene flow, and thus develop different evolutionary trajectories. The closer a genetic marker is to the alleles responsible for adaptive differences, the more likely that it will trace the history of phenotypic change.

Understanding how phenotypic variation is generated in nature is greatly enhanced by studying groups that are actively undergoing diversification. By deciphering the history of such diverse phenotypes we gain a clearer understanding of the evolutionary process, including the tempo and mode of phenotypic change. *Heliconius* butterflies present one of the most striking examples of a recent phenotypic radiation. The 40 species in the genus exhibit hundreds of wing patterns that are involved in Müllerian mimicry complexes, where distasteful species converge on a shared warning signal to avoid predation. This convergence is particularly remarkable in two species, *Heliconius erato* and *Heliconius melpomene.* These species are phylogenetically distant and do not hybridize (1), yet they have converged to share over 20 different mimetic color patterns across the Neotropics (Fig. 1) (2, 3). Most of the color-pattern diversity in these species can be partitioned into two major groups: "rayed" patterns, involving orange-red rayed hindwing patterns with orange-red basal forewings, and "red-banded" patterns, involving crimson-banded forewings and hindwings that are black and may have a yellow bar (Fig. 1). The rayed phenotypes are comimetic with several other *Heliconius* species across a broad contiguous Amazonian distribution. In contrast, the red-banded phenotypes are mostly restricted to just the two comimics *H. erato* and *H. melpomene*, and are found in multiple disjunct regions around the periphery of the Amazon.

A long history of research has been devoted to understanding the historical processes generating the wing pattern diversity within *H. erato* and *H. melpomene*. Earlier investigators proposed an allopatry-based Pleistocene refugium hypothesis to explain these patterns, whereby the identical patchwork of color patterns that characterize *H. erato* and *H. melpomene* arose simultaneously when populations of the two species became isolated together in forest refugia during Pleistocene cooling (2–4). Mallet proposed an alternative parapatric-based hypothesis, where color patterns evolved through a process similar to Wright's shifting balance (5). Under his hypothesis, novel color patterns became common enough locally to be fixed by frequency dependent selection. If advantageous, the pattern would spread, resulting in shifting of the boundaries of mimicry complexes over time (6, 7). Mallet proposed that the disjunct color patterns we see today may have had a single origin and been created when a rayed pattern originated and spread from the Amazon, displacing and fragmenting a previously contiguous red-banded population (8).

As a test of these hypotheses, molecular markers unlinked to color pattern, including mtDNA (9–11), nuclear sequences (12), and amplified fragment length polymorphism (AFLPs) (11), have been used to more carefully dissect the timing of the two parallel radiations and the relationships among color pattern races within the two comimics. These studies inferred an older (10–12) and different phylogenetic history for *H. erato* than

Author contributions: H.M.H., B.A.C., R.D.R., C.D.J., M.R.K., and W.O.M. designed research; H.M.H., B.A.C., R.P., and W.O.M. performed research; P.A.d.M., M.Z.C., M.L., C.D.J., and M.R.K. contributed new reagents/analytic tools; H.M.H., B.A.C., and W.O.M. analyzed data; H.M.H., B.A.C., R.P., J.M., R.D.R., C.D.J., M.R.K., and W.O.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. JN897400 and JN898803) and in the Dryad database (doi:10.5061/dryad.8h154h65).

[1]To whom correspondence should be addressed. E-mail: heather_hines@ncsu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1110096108/-/DCSupplemental.

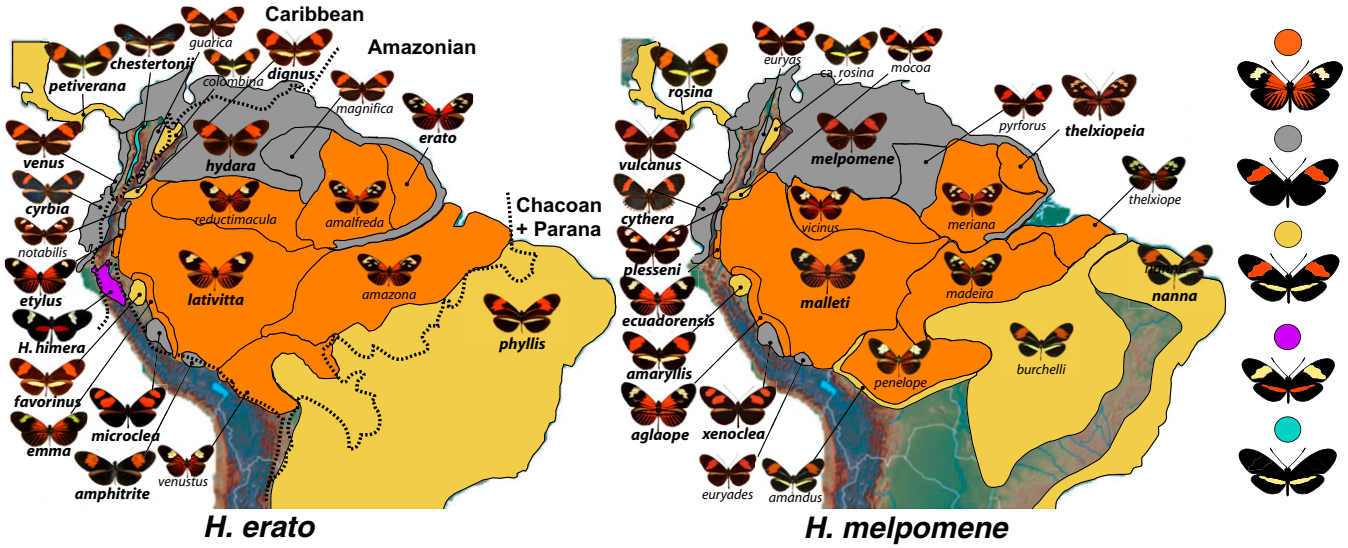

Fig. 1. The parallel radiation of mimetic color-pattern complexes in *H. erato* and *H. melpomene*. Each region demarcates the boundaries of a different color pattern and is colored based on broad categorization of the color patterns. The highly divergent races/incipient species are indicated with separate colors. This characterization excludes differences in the shape of the forewing band, which is quite variable among races. Races sampled have larger images and names. Dashed lines are boundaries of the major biotic domains for this region recognized by Morrone (25). This figure was redrawn using the range maps in refs. 2 and 8.

H. melpomene, supporting earlier hypotheses (13) that the more abundant *H. erato* was the "model," and the rarer *H. melpomene* diverged more recently to mimic it. This theory refuted a simple refugium hypothesis, as both species did not evolve simultaneously. Furthermore, rather than clustering by color pattern, phylogenies from these studies cluster individuals within both species by geographic proximity, with the major lineages dividing the Amazon from the Central American/West Andean region (9–12). Given the geographically disjunct nature of the phenotypes, these data led to the intriguing suggestion that similar color-pattern types within each species were acquired through multiple independent origins. However, even a large sample of color pattern-independent markers may not be adequate to refute the possibility of a single origin within each comimic. Turner et al. (14) argued, using allozyme data, that markers unlinked to color pattern exhibit a different history to color-pattern loci and are poor indicators of the evolutionary history of these phenotypic races. This notion is supported by recent association data showing that only loci tightly linked to the adaptive color patterns have restricted gene flow at racial hybrid zones within both *H. erato* and *H. melpomene* (15–17). Phylogenetic data from the genomic regions near color-pattern loci should provide a more accurate picture of the history of these phenotypes.

The color-pattern radiation of *Heliconius* butterflies is one of only a handful of cases (e.g., refs. 18–20), where it is now possible to examine the history of a radiation at the loci responsible for the adaptive changes. The genomic interval controlling red pattern variation (Fig. 1) has recently been localized to a homologous region in both comimics (16, 17, 21). Further population genetic and comparative gene-expression work on this region has identified a gene, *optix*, that controls red color-pattern variation in both *H. erato* and *H. melpomene* (22). Although no other genes show the pattern-specific expression of *optix*, other genes in this interval have shown significant genotype-by-phenotype associations (16, 17, 23) and expression differences between divergent color-pattern races (16, 17, 24).

Here we examine sequence variation in *optix* and four other candidate genes linked to red pattern elements (*bves*, *kinesin*, *GPCR*, and *VanGogh*) across the mimetic races in both *H. erato* and *H. melpomene*. We compare these data to sequences from four unlinked nuclear genes (*SUMO*, *Suz12*, *2654*, and *CAT*) and a mitochondrial COI-COII fragment, to address whether the color-pattern history inferred for *optix* and nearby genes is different from neutral genomic signal. We hypothesize that the phylogeny of a color-pattern locus, compared with unlinked markers, will show more population structure by red color-pattern phenotypes and demonstrate the history of adaptive phenotype evolution more clearly. Specifically, if shared red patterns have a common origin within each species, they should form a single cluster in phylogenetic analyses of *optix* alleles. If convergently derived, their *optix* alleles should cluster in multiple independent lineages. We use these data to examine the historical events leading to extant color-pattern complexes within each species and to compare the timing and patterns of these radiations between species.

Results

To infer the history of the color-pattern radiations in *H. erato* and *H. melpomene*, we examined patterns of genetic diversity, phylogenetic relationships, and population structure in markers linked and unlinked to color pattern among individuals of 14 distinct color pattern races of *H. erato* ($n = 73$) and 11 comimetic races of *H. melpomene* ($n = 61$) (Fig. 1 and Table S1).

Genetic Diversity Within Comimics. There were marked differences within species and among different phenotypes in standing levels of genetic variation. *H. erato* had over twofold higher estimates of nucleotide diversity (θ_π)(Table S2) and recombination (γ) (Table S3) than *H. melpomene* across nuclear loci. Within species, patterns of overall genetic diversity (Table S2), recombination (Table S3), and phylogenetic signal (Table 1) did not differ consistently between color pattern-linked and -unlinked nuclear markers. However, color pattern-linked genes showed a reduction in diversity (Table S2) and recombination (Table S3) among rayed individuals of both species. This difference is particularly strong in *optix* for *H. erato*, which showed a fourfold reduction in diversity in rayed versus nonrayed individuals. This discrepancy can be visualized using mismatch plots (Fig. 2). At unlinked loci—and increasingly so for linked markers and *optix*—there were greater pairwise genetic distances among red-banded phenotypes than among rayed phenotypes for both species. These plots also highlight the greater pairwise genetic distances in *H. erato* than in *H. melpomene* and show that the difference between them is more pronounced for red-banded phenotypes, with rayed phenotypes having more similar genetic distances between species.

Population Structure of Unlinked Markers. As previously observed (11, 12), nucleotide variation at markers unlinked to color pattern

Table 1. Population structure inferred from AMOVA and phylogenetic signal in color pattern

		AMOVA: population structure						Phylogeny
		Color pattern		Geography		Color pattern: Amazon		
Gene	Type (kb)	% Var	*P* value	% Var	*P* value	% Var	*P* value	CP Steps*
H. erato								
optix	Target (0)	55.19	0.0010	21.70	0.0675	52.51	0.0147	3
kinesin	Linked (−181)	8.19	0.0323	7.84	0.0411	4.24	0.1281	14–18
GPCR	Linked (−172)	8.24	0.0547	26.37	0.0010	5.77	0.1017	17
bves	Linked (−220)	9.06	0.0147	2.53	0.1799	14.73	0.0284	12–20
VanGogh	Linked (+87)	−0.94	0.4135	7.98	0.0147	−3.95	0.9013	21
SUMO	Unlinked	−0.30	0.5591	0.94	0.0391	−1.29	0.9775	30
Suzy	Unlinked	−1.50	0.6305	0.15	0.2239	1.16	0.2160	24–33
2654	Unlinked	−0.19	0.4230	−1.33	0.7146	−0.65	0.5494	24–34
CAT	Unlinked	15.91	0.2102	72.14	0.0010	−1.54	0.9091	17–23
mt	Unlinked	20.09	0.0459	72.06	0.0010	−1.94	0.6090	8*
H. melpomene								
optix	Target	21.84	0.0147	31.67	0.0029	14.68	0.0557	5
kinesin	Linked	19.50	0.0381	18.40	0.0538	15.24	0.2307	4–18
GPCR	Linked	1.61	0.2893	3.94	0.0997	3.93	0.1144	7–28
bves	Linked	4.95	0.1965	34.80	0.0049	0.57	0.5572	11–27
VanGogh	Linked	3.37	0.1799	12.69	0.0156	10.50	0.0244	15–22
SUMO	Unlinked	−0.54	0.4438	12.89	0.0254	−6.89	0.8534	19–22
Suzy	Unlinked	−4.04	0.5093	32.10	0.0068	−6.58	0.6227	13–39
2654	Unlinked	1.87	0.2581	24.10	0.0020	1.32	0.3597	10–31
CAT	Unlinked	−0.84	0.4438	13.18	0.0117	−4.60	0.9707	21–24
mt	Unlinked	−0.51	0.3744	35.21	0.0078	−15.08	0.9306	13*

CP Steps represents the number of color-pattern changes inferred on the neighbor-joining trees using parsimony, with ranges representing alternative reconstruction of polytomies; Var, variance. Gray shading represents color pattern linked genes. Unshaded rows represent unlinked genes.

*Because *mt* has only one haplotype for each individual, approximately half the number of steps are expected.

were structured mostly by geography. In *H. erato*, nucleotide variation at three (*2654*, *Suz12*, and *SUMO*) of the four unlinked nuclear markers was broadly distributed among populations with very little evidence of population structure among races, geographic regions, or color-pattern phenotypes in the analysis of molecular variance (AMOVA) (Table 1) or phylogenetic analyses (Fig. S1 *H–J*). In contrast, the mitochondrial fragment and the unlinked nuclear marker *CAT* had clear population structure, largely reflecting geography (Figs. 3*A*, Table 1, and Fig. S1*G*). Both genes recognized distinct lineages that match closely to the Amazonian, Caribbean, and Chacoan/Paranan biotic domains of Morrone (25) (Fig. 1). In the AMOVA, a large amount of variation in both of these genes was explained by geographic divisions into Amazon +Chacoan and Caribbean domains (72.1% each). A fair amount of variation was also explained by color pattern (rayed vs. nonrayed patterns; 15.9–20.1%). However, this pattern was likely the result of regional differences in color-pattern phenotypes, as rayed phenotypes are found only in the Amazon region (Fig. 1). When we reduce the effect of geographic structure by examining color-pattern structure within the Amazonian region, none of the genetic variation was explained by color pattern (Table 1).

In *H. melpomene*, there was a strong geographic pattern to the distribution of variation across unlinked loci. The mitochondrial fragment demarcated four major populations, including the same three lineages for *H. erato* (Amazon, Caribbean, and Chacoan/Parana) but it differed in resolving a Guiana Shield lineage consisting of Trinidad and French Guianan specimens (Figs. 3*A* and Fig. S1*F*). The unlinked nuclear markers have less straightforward phylogenetic clustering of populations (Figs. 3*A* and Fig. S1 *G–J*). However, all of these genes demonstrated significant geographic signal with a fair amount of variation explained by the geographic division of the Amazon+Chacoan from the Caribbean region (12.9–35.2%) in the AMOVA. These same genes exhibited virtually no variation that could be explained by color pattern (Table 1).

***optix* Exhibits Population Structure by Color Pattern.** In contrast to unlinked markers, *optix* showed strong population structure based on color pattern. This structure was most apparent in *H. erato*, where the Bayesian phylogeny (Fig. S1*A*), neighbor-joining tree (Fig. 3*A*), and haplotype network (Fig. 3*B*) of the inferred haplotypes for *optix* place nearly all haplotypes of rayed races into a single derived lineage. Over half of the *optix* variation was explained by color pattern phenotype (55.2%, $P = 0.00098$), a much larger portion than any other gene (Table 1). This large and significant contribution of color pattern remained when the effect of geographic structure is removed (52.5%, $P = 0.01466$).

There was less intraspecific variation in *optix* in *H. melpomene*, making phylogenetic and network inferences more difficult. Nonetheless, individuals clustered by color-pattern phenotype. In the Bayesian tree, most of the rayed phenotypes formed a basal polytomy (Fig. S1*A*). In contrast, the neighbor-joining tree clustered most of the rayed alleles together, but the alleles fell on a derived lineage. In the haplotype network (Fig. 3*B*) the rayed haplotypes clustered together near the origin of *H. melpomene*. AMOVA similarly supported color-pattern clustering in *optix* for *H. melpomene*. Using all populations, *optix* had the highest variation of any locus explained by color pattern (21.8%, $P = 0.01466$). When removing the effect of geographic structure, the variation explained by color pattern structure for *optix* is reduced, but nearly significant (14.7%, $P = 0.0557$). The inferred patterns of population structure for unlinked markers and *optix* were further supported by STRUCTURE analyses (*SI Text* and Fig. S2). Similar to every locus thus far examined, a neighbor-joining tree combining *optix* sequences for both *H. melpomene* and *H. erato* resulted in two distinct clades, with no sharing of alleles between the comimics (Fig. S1*A*).

There were a few exceptions to the complete clustering by color pattern in both the *H. erato* and *H. melpomene optix* data (Fig. 3*B* and Fig. S1*A*). In many of these exceptions, an individual possessed both a rayed and nonrayed haplotype. For example, three of the five *H. erato microclea* individuals possessed

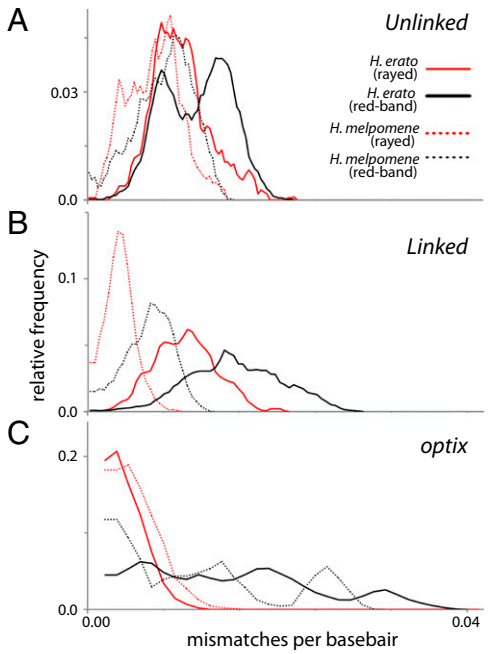

Fig. 2. Mismatch distributions of comimics and major color-pattern lineages. These plots display the relative frequency of pairwise distances using a three-difference sliding-window average, including distributions within *H. erato* (solid lines) and *H. melpomene* (dotted lines) for rayed and nonrayed phenotypes using combined unlinked nuclear markers (*Top*), combined linked markers excluding *optix* (*Middle*), and *optix* (*Bottom*).

haplotypes that grouped with both rayed and nonrayed clusters. Similarly, in *H. melpomene*, a single haplotype of a rayed *H. melpomene aglaope* individual was resolved in a nonrayed clade and a single haplotype in a nonrayed *H. melpomene amaryllis* individual fell within a rayed clade. Furthermore, we observed evidence for recombination between the rayed and nonrayed haplotypes in an *H. erato hydara* individual from Trinidad (TT05_hyd_b) (Fig. S1*A*), which possessed one haplotype that matched other nonrayed haplotypes and another that appeared to be a recombinant between nonrayed and rayed haplotypes.

Decay of Color-Pattern Signal in Other Linked Genes. Relationships between allelic variation and color pattern for the other loci linked to color pattern were less pronounced. Phylogenetic trees of these linked markers had considerable admixture of rayed and nonrayed patterns among lineages in both *H. erato* and *H. melpomene* (Fig. 3*A* and Fig. S1 *B*–*E*). This admixture resulted in many more phylogenetically inferred evolutionary shifts in color pattern for these linked markers than for *optix* in *H. erato*, and a similar but less extreme difference in *H. melpomene* (Table 1). Despite such reduced signal, unlinked markers generally had more inferred evolutionary steps in color pattern than linked markers, and less variation explained by color pattern after correcting for geographic signal for both comimics (Table 1). Among these linked genes, *kinesin* for *H. melpomene* had the most notable color-pattern clustering. It had the highest levels of variation explained by color pattern after correcting for geography (15.2%) and the fewest evolutionary transitions in pattern of all other *optix*-linked markers (Table 1).

Discussion

Contrasted Modes of Evolution in the Same Genome. In the two comimics, *H. erato* and *H. melpomene*, distinctly different pictures of the history of adaptive change emerge depending on where variation is sampled in the genome. Variation at markers broadly scattered across the genome and unlinked to loci underlying phenotypic differences consistently support the hypothesis that similar color-pattern phenotypes have evolved multiple times within both radiations (10–12). In contrast, genetic variation in the red-determining transcription factor *optix* structures lineages primarily by color pattern and supports a common origin for similar patterns within each species. Thus, although *optix* has been independently co-opted to produce red wing patterns in each comimetic species (22), our genetic data suggest that each species has acquired red patterning alleles only once.

Our data also highlight the remarkable extent to which recombination and gene flow disassociate evolutionary histories within the genome. *optix* is the only locus surveyed that maintains strong genotype-by-phenotype associations across races in both lineages. The weak color-pattern signal at the other candidate loci closely linked (<250 kb) to the red pattern was unexpected, as several of these loci have previously shown significant associations with color patterns across hybrid zones. For example, in hybrid zones in Peru, *kinesin*, *VanGogh*, and *GPCR* showed strong genotype-by-phenotype association and restricted gene flow (16, 17). Even *kinesin*, which demonstrated expression differences between divergent phenotypes of both *H. erato* and *H. melpomene* (16, 17, 23) and was strongly implicated in the origin of the red forewing band of *H. heurippa* (24), was only weakly associated with color pattern in our analyses. These patterns likely reflect the physical distances between these loci and the functional differences that are driving phenotypic diversity. The nature of the functional changes responsible for the highly variable red patterns in *H. erato* and *H. melpomene* remain unknown. However, they are almost certainly to be found in unsampled *cis*-regulatory regions around *optix*, which shows highly pattern-specific gene-expression differences but almost no amino acid changes within or between *H. melpomene* and *H. erato* (22). Attesting to the very localized scale of phylogenetic phenotypic association, a few red-banded individuals were found to possess *optix* rayed alleles, suggesting recombination events between the *optix* coding region and functional *cis*-regulatory sites. This finding raises hopes that further genetic analysis will more narrowly define the exact functional regions responsible for phenotypic change and demonstrates the scale at which gene flow and recombination can decouple the evolutionary history of different parts of the genome. These contrasting modes of evolution serve as a cautionary example to those attempting to understand the history of adaptive change in rapidly radiating lineages using only neutral markers.

***Optix* Reveals the History of a Mimetic Radiation.** The ability to sample variation near the loci that are the targets of selection permits deeper insights into the directionality of color-pattern change and the historical processes leading to extant mimetic diversity. For *H. erato*, all red-banded phenotypes fell together in a highly diverse, ancestral lineage relative to the genetically less-variable rayed lineage (Fig. 3*B*). The derived position and lower levels of variation of the rayed lineage provide evidence for a recent history and rapid spatial spread of the rayed mimicry complex, in line with the centrifugal spread hypothesis (8), whereby the rayed phenotype evolved recently and spread rapidly across the Amazon Basin, fracturing and marginalizing the range of the ancestral red-banded phenotypes and creating the disjunct distributional patterns we observe today. Haplotype clustering of subphenotypes of the geographically disjunct red-forewing band populations (black vs. yellow-barred hindwing) provides further evidence that the red-forewing band phenotypes have been fragmented from once contiguous populations. Collectively, these patterns support a fluid history for the mimicry complexes, with boundaries that have ebbed and flowed through time, a concept put forth by Mallet (7, 8) and supported by evidence of rapidly moving hybrid zones between races (26).

Deciphering the origins of different color-pattern phenotypes in *H. melpomene* is more difficult, in part because lower allelic variation limits phylogenetic resolution. In contrast to *H. erato*, the genetic evidence suggests that the rayed pattern was present early in *H. melpomene* evolution. The origin of the rayed phenotype in *H. melpomene* is reconstructed near the ancestral node in the haplotype network of *optix*, which divides the Amazonian rayed and nonrayed lineages. The haplotype network thus suggests that

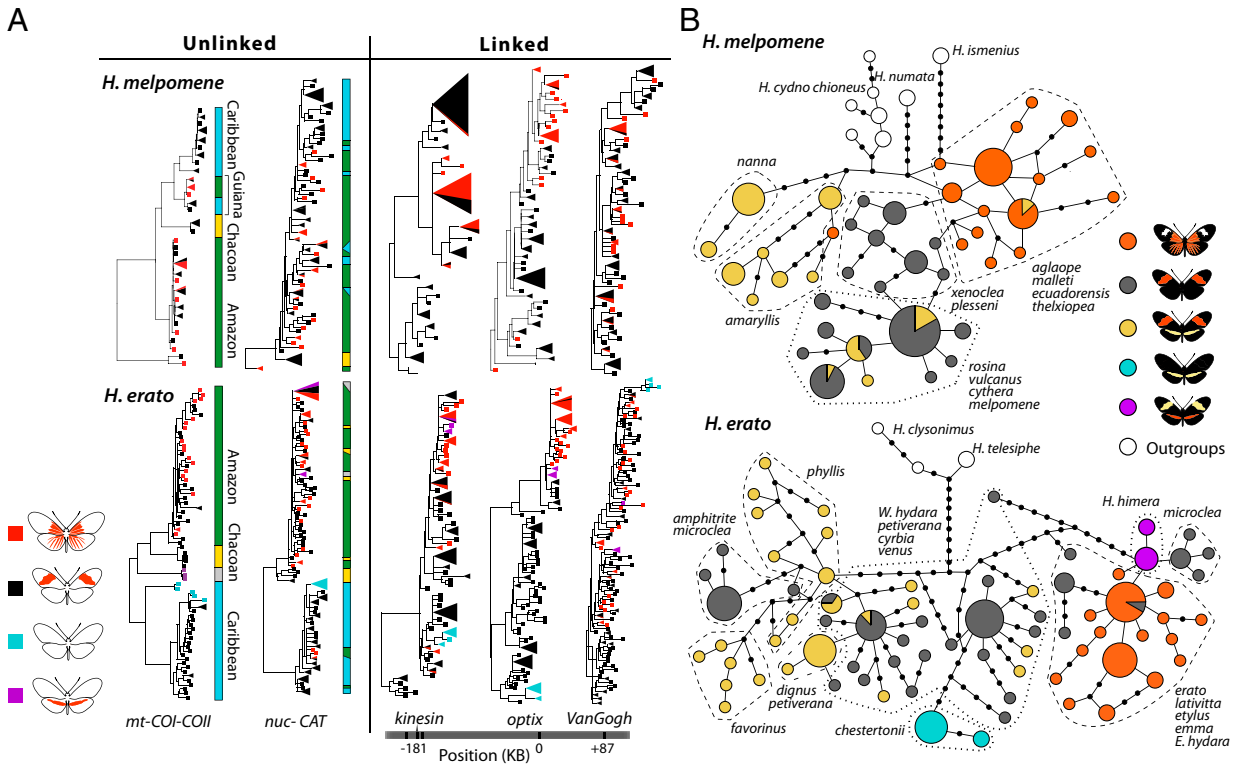

Fig. 3. Phylogeographic relationships of *H. erato* and *H. melpomene*. (*A*) Phylogenetic clustering of color pattern in select color-pattern linked and unlinked markers. Trees were constructed using neighbor-joining methods with terminal nodes colored by the major red color-pattern phenotypes. Unlinked markers displayed are those with greater population structure, and linked markers include *kinesin*, which has the greatest color pattern association besides *optix*, and the closest gene, *VanGogh*. Triangular clades represent shared haplotypes, with size reflecting the frequency of the haplotype. Colored bars indicate geographic distribution. (*B*) *optix* haplotype networks. Each node represents a haplotype, lines between nodes represent a single base change, and the size of the node represents haplotype frequency. Nodes are colored by the color patterns containing that haplotype. Major racial groups are indicated using dashed polygons, with dashes differing by geographic domain (large dash –, Amazon/Chacoan; small dash -, Caribbean).

the *H. melpomene* radiation may have originated and spread from the Amazon and colonized other regions of Central and South America, a finding consistent with results from large-scale AFLP data (10). Support for this hypothesis, however, is weak, and there is more variation within nonrayed phenotypes than rayed phenotypes. Furthermore, although patterns are not transferred between *H. erato* and *H. melpomene*, *H. melpomene* may have acquired different red pattern phenotypes through hybridization with its closest relatives. *H. melpomene* is a member of a larger complex of species that are known to hybridize in nature, including a number of species in the Amazon region with rayed phenotypes (1, 27). The acquisition of new color patterns by hybridization is thought to play an important role in the evolution of pattern variation in *Heliconius* (24, 28, 29). Additional sequence data around *optix* across the *H. melpomene* species-complex should help resolve the origins of red patterns within the *melpomene* group.

Advergence and Convergence and the Origins of Mimicry Between *H. erato* and *H. melpomene*. These data allow us to reassess hypotheses about the timing of the parallel coradiations and the origins of the mimetic relationship between *H. erato* and *H. melpomene*. Our data provide additional evidence against simultaneous diversification of *H. erato* and *H. melpomene*, but do not rule out the idea that the rayed patterns may have diverged more recently in parallel in the Amazon. Overall, *H. erato* harbors substantially more variation and has deeper intraspecific genealogies than *H. melpomene*. The two species are also inferred to have different phylogeographic history, with eastward spread of color pattern in *H. erato* and westward spread in *H. melpomene*. Although exact dating is difficult given that selection will impact rates of genetic divergence (16, 17), the discrepancies in timing and geographic pattern of the radiations appear to be driven mostly by the considerably older origins of the nonrayed *H. erato* patterns. Extant levels of variation in *optix* and unlinked markers in rayed phenotypes are more similar between comimics, suggesting that rayed patterns diverged at about the same time in the two species. The parallel radiation can thus be explained by *H. erato* establishing the red-banded populations first, *H. melpomene* adverging on these patterns, and both species acquiring the Amazonian rayed mimetic pattern around the same time, with subsequent spread of this pattern fragmenting established red-banded populations. It is also possible that both *H. erato* and *H. melpomene* converged together on the rayed pattern to mimic other Amazonian heliconiines. Unlike the nonrayed forms of *H. erato* and *H. melpomene*, the rayed phenotypes are part of a larger mimicry complex composed of over a dozen species (mostly *Heliconius* but also other butterfly genera, and even day-flying moths) that share the same pattern (4, 13, 30). The deeper phylogenetic origins of this convergence can now be more thoroughly explored using the genes underlying the phenotypic change.

Conclusions and Future Directions. Nucleotide variation at regions tightly linked to the functional sites driving adaptive change provide unique insights into the origins of the patchwork of mimetic color pattern races in *H. erato* and *H. melpomene* that has intrigued biologists for over 150 y. Contrary to the history reflected in the majority of the genome, data around the functional sites driving phenotypic variation suggest that similar wing-pattern phenotypes share a common origin within each of the two parallel radiations. These data also suggest that the rayed Amazonian phenotype evolved recently and around the same time in *H. erato* and *H. melpomene*, and spread rapidly, replacing the ancestral red-banded phenotype. Although variation in red is a major aspect of the complex story of mimicry between *H. erato* and *H. melpomene*, a number of other loci interact with the red locus to generate the phenotypic variation that characterizes the

two radiations. Current research is identifying these loci, including loci that modulate the shape of the forewing band and the presence of a yellow hindwing bar (21). A combination of sequence information across the different color-pattern loci promises an even more complete picture of the history of these coradiations and a deeper understanding about how novel variation arises and spreads during adaptive change.

Materials and Methods

Taxonomic Sampling and DNA Data Collection. We sequenced 137 individuals of *H. erato* (including three *Heliconius himera*) and *H. melpomene*, and two outgroups for each species (Fig. 1 and Table S1). Specimens were sequenced for 10 gene fragments, including four unlinked nuclear markers [*CAT* (1081 bp), *SUMO* (805 bp), *Suz12* (520 bp), *Gene 2654* (872 bp)], a mitochondrial region *COI-tRNAleu-COII* (1,510 bp), and five genes within the red color-pattern interval. These genes included an 800-bp *optix* transcript (432 bp of coding sequence and 361–370 bp of 3′ UTR) and coding regions of four genes within 250 kb of *optix* (Table 1), including *kinesin* (501 bp), *GPCR* (522 bp), *VanGogh* (715 bp), and *bves* (385 bp). Further gene information is available in Table 1 and Table S4.

DNA was extracted from thoracic tissue using the Qiagen DNA Plant kit, PCR-amplified (Table S4), and purified using ExoSAP-IT (USB, Affymetrix). We sequenced both forward and reverse strands using ABI Big Dye Terminator v3.1 reactions and PCR primers, and called SNPs based on consistent double peaks in chromatograms. For *optix*, we cloned (TOPO TA Cloning kit; Invitrogen) a few individuals (Table S1) to facilitate phasing. Haplotypes were inferred from polymorphic sequences using PHASEv2.1 (31), allowing recombination (32) with a recombination rate prior of 0.04, and declaring known phases from cloned individuals.

Reconstructing Phylogenetic Trees and Haplotype Networks. We constructed neighbor-joining trees of phased haplotypes for each gene and species in PAUP* 4.0b10 (33). We also performed a combined neighbor-joining analysis on *optix* for the two species to test for allele sharing. For *optix* and *mt COI-COII* we constructed Bayesian phylogenies of haplotypes using MrBayes v.3.1.2 (34) (Fig. S1). We generated parsimony-based haplotype networks for *optix* using TCS 1.2.1 (35), relaxing the number of steps to allow placement of outgroups.

Tests for Population Structure. We inferred population structure for each gene and species using an AMOVA implemented in Arlequin 3.1 (36). Population structure was assessed relative to geographic region (Caribbean, Amazon +Chacoan) (Fig. 1 and Table S1), color pattern (rayed, nonrayed), and race within these groups by comparing variance in uncorrected genetic distances of phased haplotypes. Outgroups, *H. himera* and *H. erato chestertonii* were excluded, as these lineages show strong reproductive isolation from other races. We also assigned individuals to populations based on Hardy–Weinberg equilibrium assumptions using STRUCTURE 2.2 (37) (*SI Text*). As a phylogenetic measure of color-pattern signal for each gene, we inferred the number of evolutionary steps between rayed and nonrayed color patterns using parsimony-based character reconstruction on neighbor-joining trees in MacClade (38).

Genetic Diversity and Recombination Estimates. Methods for estimating genetic diversity and recombination are outlined in Tables S2 and S3. Mismatches (nucleotide differences) between pairs of haplotypes were calculated within species and within rayed and red-banded sets (Table S1) of individuals using concatenated sequences of the nuclear unlinked genes for each individual in Arlequin 3.1.

ACKNOWLEDGMENTS. We thank Stephanie Ruzsa, Vincent Izzi, Nicola Chamberlain, and Felix Araujo-Perez for assistance with sample preparation and sequencing; Matt Bertone for graphics assistance; Jeff Thorne for analytical advice; and the Peruvian Ministerio de Agricultura and Instituto Nacional De Recursos Naturales (004-2008-INRENA-IFFS-DCB and 011756-AG-INRENA), the Ecuadorian Ministerio del Ambiente Ecuadorian (013-09 IC-FAU-DNB/MA), and the Brazilian Ministério do Meio Ambiente (permit 10894-1) for collection permits. This work was funded by a Ruth L. Kirschstein National Research Service Award F32 GM889942 (to H.M.H.), Fapern/Conselho Nacional de Pesquisas Grant PPP/2007 (to M.Z.C.); National Institutes of Health/National Institute of General Medical Sciences Grant GM068763 and National Science Foundation Division of Environmental Biology (DEB) Grant 1020355 (to M.R.K.); Biotechnology and Biological Sciences Research Council Grant G006903 (to J.M.); and National Science Foundation Grants DEB-0844244 and DEB-0715096 (to R.D.R. and W.O.M.).

1. Mallet J, Beltrán M, Neukirchen W, Linares M (2007) Natural hybridization in heliconiine butterflies: The species boundary as a continuum. *BMC Evol Biol* 7:28.
2. Brown KS, Sheppard PM, Turner JRG (1974) Quaternary refugia in tropical America: Evidence from race formation in Heliconius butterflies. *Proc R Soc Lond B Biol Sci* 187:369–378.
3. Sheppard PM, Turner JRG, Brown KS, Benson WW, Singer MC (1985) Genetics and the evolution of Müllerian mimicry in *Heliconius* butterflies. *Philos Trans R Soc Lond, B* 308:433–613.
4. Turner JRG (1983) Mimetic butterflies and punctuated equilibria: Some old light on a new paradigm. *Biol J Linn Soc Lond* 20:277–300.
5. Wright S (1982) Character change, speciation, and the higher taxa. *Evolution* 36:427–443.
6. Mallet J, Singer MC (1987) Individual selection, kin selection, and the shifting balance in the evolution of warning colours: The evidence from butterflies. *Biol J Linn Soc Lond* 32:337–350.
7. Mallet J (2010) Shift happens! Shifting balance and the evolution of diversity in warning colour and mimicry. *Ecol Entomol* 35:90–104.
8. Mallet J (1993) *Hybrid Zones and the Evolutionary Process*, ed Harrison RG (Oxford University Press, New York), pp 226–260.
9. 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.
10. Brower AVZ (1996) Parallel race formation and the evolution of mimicry in *Heliconius* butterflies: A phylogenetic hypothesis from mitochondrial DNA sequences. *Evolution* 50:195–221.
11. Quek S-P, et al. (2010) Dissecting comimetic radiations in *Heliconius* reveals divergent histories of convergent butterflies. *Proc Natl Acad Sci USA* 107:7365–7370.
12. Flanagan NS, et al. (2004) Historical demography of Mullerian mimicry in the neotropical *Heliconius* butterflies. *Proc Natl Acad Sci USA* 101:9704–9709.
13. Eltringham H (1917) On the specific and mimetic relationships of the genus *Heliconius*. *Trans Entomol Soc Lond* 1916:101–148.
14. Turner JRG, Johnson MS, Eanes WF (1979) Contrasted modes of evolution in the same genome: Allozymes and adaptive change in Heliconius. *Proc Natl Acad Sci USA* 76:1924–1928.
15. Mallet J, McMillan WO, Jiggins CD (1998) *Endless Forms: Species and Speciation*, ed Howard DJ (Oxford University Press, New York), pp 390–403.
16. Counterman BA, et al. (2010) Genomic hotspots for adaptation: The population genetics of Müllerian mimicry in *Heliconius erato*. *PLoS Genet* 6:e1000796.
17. Baxter SW, et al. (2010) Genomic hotspots for adaptation: The population genetics of Müllerian mimicry in the *Heliconius melpomene* clade. *PLoS Genet* 6:e1000794.
18. Terai Y, Morikawa N, Okada N (2002) The evolution of the pro-domain of bone morphogenetic protein 4 (Bmp4) in an explosively speciated lineage of East African cichlid fishes. *Mol Biol Evol* 19:1628–1632.
19. Colosimo PF, et al. (2005) Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307:1928–1933.
20. Shapiro MD, et al. (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428:717–723.
21. Joron M, et al. (2006) A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biol* 4:e303.
22. Reed RD, et al. (2011) *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* 333:1127–1141.
23. Nadeau NJ, et al. (2011) Evidence for genomic islands of divergence among hybridizing species and subspecies of *Heliconius* butterflies obtained by large-scale targeted sequencing. *Phil Trans Roy Soc B*, in press.
24. Salazar C, et al. (2010) Genetic evidence for hybrid trait speciation in *Heliconius* butterflies. *PLoS Genet* 6:e1000930.
25. Morrone JJ (2006) Biogeographic areas and transition zones of Latin America and the Caribbean islands based on panbiogeographic and cladistic analyses of the entomofauna. *Annu Rev Entomol* 51:467–494.
26. Blum MJ (2002) Rapid movement of a *Heliconius* hybrid zone: Evidence for phase III of Wright's shifting balance theory? *Evolution* 56:1992–1998.
27. Beltran M, Jiggins CD, Brower AVZ, Bermingham E, Mallet J (2007) Do pollen feeding, pupal-mating and larval gregariousness have a single origin in *Heliconius* butterflies? Inferences from multilocus DNA sequence data. *Biol J Linn Soc Lond* 92:221–239.
28. Gilbert LE (2003) *Ecology and Evolution take Flight*, eds Boggs CL, Watt WB, Ehrlich PR (University of Chicago Press, Chicago), pp 281–318.
29. Mallet J (2009) *Speciation and Patterns of Diversity*, eds Butlin R, Bridle J, Schluter D (Cambridge University Press, Cambridge, U.K.; New York), pp 177–194.
30. Mallet J (2001) Causes and consequences of a lack of coevolution in Müllerian mimicry. *Evol Ecol* 13:777–806.
31. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989.
32. Li N, Stephens M (2003) Modeling linkage disequilibrium, and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165:2213–2233.
33. Swofford DL (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4* (Sinauer Associates, Sunderland, Massachusetts).
34. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
35. Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. *Mol Ecol* 9:1657–1659.
36. Excoffier LGL, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.
37. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
38. Maddison DR, Maddison WP (2005) *MacClade 4.08* (Sinauer, Sunderland, MA).