Published in final edited form as:

Mol Ecol. 2013 May ; 22(10): 2760–2770. doi:10.1111/mec.12290.

Cryptic genetic and wing pattern diversity in a mimetic *Heliconius* butterfly

R.I. Hill¹, L.E. Gilbert², and M.R. Kronforst³

¹University of the Pacific, Department of Biological Sciences, 3601 Pacific Ave, Stockton, CA 95211

²University of Texas at Austin, Section of Integrative Biology, Austin, TX 78712

³University of Chicago, Department of Ecology & Evolution, 1101 E. 57th St., Chicago, IL 60637

Abstract

Despite rampant color pattern diversity in South America, *Heliconius erato* exhibits a “postman” wing pattern throughout most of Central America. We examined genetic variation across the range of *H. erato*, including dense sampling in Central America, and discovered a deep genetic break, centered on the mountain range that runs through Costa Rica. This break is characterized by a novel mitochondrial lineage, which is nearly fixed in northern Central America, that branches basal to all previously described mitochondrial diversity in the species. Strong genetic differentiation also appears in Z-linked and autosomal markers and it is further associated with a distinct, but subtle, shift in wing pattern phenotype. Comparison of clines in wing phenotype, mtDNA and nuclear markers indicate they are all centered on the mountains dividing Costa Rica, but that cline width differs among datasets. Phylogeographic analyses, accounting for this new diversity, rewrite our understanding of mimicry evolution in this system. For instance, these results suggest that *H. erato* originated west of the Andes, perhaps in Central America, and as many as 1 million years before its co-mimic, *H. melpomene*. Overall our data indicate that neutral genetic markers and color pattern loci are congruent and converge on the same hypothesis – *H. erato* originated in northwest South America or Central America with a “postman” phenotype and then radiated into the wealth of color patterns present today.

Keywords

Central America; *Heliconius erato*; Müllerian mimicry; phylogeography; South America

Introduction

Heliconius butterflies display colorful wing patterns that warn predators of their unpalatability (Brown 1981). They are also a well-known example of Müllerian mimicry, a form of mimicry in which multiple, defended species converge on the same warning pattern (Eltringham 1916; Müller 1879; Turner 1971, 1976). One particularly striking aspect of *Heliconius* mimicry is the widespread color pattern variation exhibited by many species (Sheppard *et al.* 1985). For instance, the wing pattern of *H. erato* shifts drastically every few hundred kilometers, producing a patchwork of over 25 named color pattern forms across

Corresponding author: rhill@pacific.edu.

Data Accessibility Detailed data for samples, including wing phenotype data, and accession data for all sequence data are included in Supplementary Table ST1.

AFLP data are provided in Supplementary Table ST2.

Latin America (Brown *et al.* 1974; Turner & Mallet 1996). A second species, *H. melpomene*, is a co-mimic of *H. erato* and the two species converge on the same warning pattern wherever they are sympatric (Brown *et al.* 1974; Turner & Mallet 1996).

Interestingly, previous work has found that the striking color pattern shifts in *H. erato* and *H. melpomene* are generally accompanied by very little genetic differentiation among populations. For instance, Brower (1994; 1996) sequenced portions of mtDNA from a variety of *H. erato* and *H. melpomene* races and found very little phylogeographic structure, aside from large-scale regional clustering. Similarly, Flanagan *et al.* (2004) analyzed portions of two nuclear genes and found similar, low levels of genetic divergence and little spatial structure. More recently, Quek *et al.* (2010) used thousands of molecular markers and they were able to distinguish many geographic populations and subspecies in both *H. erato* and *H. melpomene*. In contrast to these studies of neutral markers, recent work has shown that genetic variation around color patterning genes themselves is strongly differentiated among color pattern races (Hines *et al.* 2011; Nadeau *et al.* 2012; Reed *et al.* 2011), as expected given a history of divergent natural selection and reduced gene flow.

The combined results of these previous analyses suggest that the history of diversification in *H. erato* and *H. melpomene* differ considerably. For instance, *H. erato* appears to have originated on the western side of South America approximately 2.8 million years ago, while *H. melpomene* may have originated in the east, around 2.1 million years ago (Quek *et al.* 2010). Furthermore, *H. erato* underwent rapid diversification and expansion resulting in widely dispersed sister taxa with *H. melpomene* experiencing a slower pace of diversification, producing a stepwise directional expansion from east to west (Quek *et al.* 2010). For both species, the combined actions of recent diversification and on-going gene flow between color pattern races prevent genetic differentiation across most of the genome, except for those narrow intervals that control phenotypic differences (Baxter *et al.* 2010; Counterman *et al.* 2010; Nadeau *et al.* 2012).

Critical factors that influence the power of empirical population genetics and phylogeography to unlock evolutionary history are the amount and utility of molecular data, and the distribution and number of samples. While the molecular tools brought to bear on *Heliconius* population genetics continue to expand, moving from mtDNA sequences to now thousands of nuclear markers (Quek *et al.* 2010), color patterning genes (Hines *et al.* 2011), and even full genome sequencing (Heliconius Genome Consortium 2012), sampling is still quite sparse. For instance, *Heliconius* samples from Central America have been underrepresented in population genetic studies historically, perhaps because highly variable species like *H. erato* and *H. melpomene* display a single wing pattern phenotype throughout the region. However, an accurate depiction of diversification and mimicry evolution in *Heliconius* requires broader sampling. Here we provide a striking example of this by showing that expanded sampling of *H. erato* across Central America reveals unprecedented, cryptic genetic diversity and divergence. We discovered a new, basal and highly divergent mtDNA lineage in *H. erato* that is nearly fixed north of the mountains that run through Costa Rica. We further show that this complete mtDNA break is associated with strong genetic differentiation throughout the nuclear genome as well as a distinct shift in wing pattern phenotype. These results expand, enrich, and potentially rewrite our understanding of the evolutionary history of a classic biological system.

Materials and Methods

Molecular data and analysis

Genomic DNA was extracted using a DNeasy Blood and Tissue Kit (QIAGEN) following the manufacturer's protocol except that two 100 µl elution steps were used. All PCR products were sequenced directly, in both directions.

Mitochondrial Genes—A total of 233 individuals of *Heliconius erato* were analyzed, comprising 12 subspecies from across the species' range (Mexico to Brazil; Table 1). For the taxonomy of *H. erato* we followed Lamas (2004). We analyzed 1611 bp spanning the 3' end of *cytochrome oxidase* subunit I (*CoI*), the tRNA-Leucine, and all of *cytochrome oxidase* subunit II (*CoII*) for 220 *H. erato* individuals, three individuals of *H. himera*, one individual of *H. clysonymus* and one *H. hecalesia*, (Genbank accession numbers for mtDNA are those used in Quek *et al.* (2010) plus the following: JX512041 - JX512175). This 1611 bp span does not overlap with the “barcode” region at the 5' end of *CoI* for which data were available for 13 individuals from Costa Rica and Panama (Genbank accession: GU157231, GU157232, GU334001–GU334005, HEU08588; Barcode of Life Database Sample ID: YB-BCI4969, YB-BCI4507, YB-BCI7538, YB-BCI14436, YB-BCI10716). We generated an additional 658 bp at the 5' end of *CoI* for eight individuals from the larger dataset (Genbank Accession: JX512070, JX512085, JX512088, JX512091, JX512098, JX512136, JX512137, JX512175) making a “barcode” dataset of 21 samples. PCR conditions and primers for *CoI-CoII* followed Beltran *et al.* (2002) for the 1611 bp segment, and Dasmahapatra *et al.* (2010) for the 5' “barcode” region.

PHYML was used with 1000 bootstrap replicates to estimate a maximum likelihood gene tree phylogeny for the *CoI-II* dataset. The GTR+I+G model was selected from among 24 models (JC, HKY, GTR option) by jModelTest using Akaike Information Criterion (Akaike 1974). For the *CoI* “barcode” dataset we used MEGA 5.05 (Tamura *et al.* 2011) to generate a Neighbor Joining tree using raw percent sequence differences and 1000 bootstrap replicates. We estimated the age of origin of *H. erato* by examining uncorrected pairwise divergence per million years among the mtDNA clades following Quek *et al.* (2010; 2004).

Nuclear Genes—We examined a subset of samples from across Central America for differentiation at two Z-linked nuclear genes: *Triose phosphate isomerase*, hereafter *Tpi* (444 bp, 59 individuals, 72 haploid sequences, Genbank Accession: JX507569 - JX507627, AF413752-AF413755, AY319203, AY319205, AY319202, AY319201, AY319204) and *Tyrosine Hydroxylase*, hereafter *TH* (736 bp, 55 individuals, 65 haploid sequences, Genbank Accession: JX507511-JX507568). PCR conditions and primers for *Tpi* followed Beltran *et al.* (2002). We designed primers for a section of *TH* (exon 5) using *Heliconius* EST data (TH2-fwd primer attatactctgaccgaagaag; TH2-rev primer tgtgtagattggaaacacgg). This region of *TH* was PCR amplified using methods similar to Beltran *et al.* (2002). Haplotype networks for *Tpi* and *TH* were generated using the program Network 4.6 (Fluxus Technology, Suffolk, England).

AFLP Markers—We genotyped a subset of 144 individuals from across Central America with AFLP markers. AFLPs were generated using the Applied Biosystems Plant Mapping Kit and scored using Applied Biosystems Genemapper software. Additional details of AFLP primers and scoring followed Quek *et al.* (2010). Genetic clustering of the AFLP data was performed with Structure 2.2 (Falush *et al.* 2007; Pritchard *et al.* 2000). Structure runs were conducted with 200,000 burn-in generations followed by 1,000,000 generations of data collection. We tested for the optimal number of clusters by varying the number of groups (*K*) from two to five, and comparing the likelihood scores.

Genetic Differentiation—We tested for significant genetic differentiation among populations in *COI-II*, *Tpi*, *TH* and AFLP data using the AMOVA method implemented in Arlequin 3.1 (Excoffier *et al.* 2005). We grouped individuals in Central America by whether they were north or south of the mountains that divide Costa Rica from Northwest to Southeast, the Cordilleras de Guanacaste, de Tilarán, de Central and de Talamanca. Within Costa Rica, localities from the Pacific slope were considered “south” and localities from the Atlantic slope were considered “north.” The north/south assignment of localities is generally easy to see in Figure 1, except in the Central Valley near San Jose (site 8) which was considered south because it is in a valley draining to the Pacific, whereas sites 9 and 10 were north, being in a valley draining to the Atlantic. To rule out the possibility that a geographic pattern was driven by samples outside of Costa Rica, two sets of analyses were conducted, one with all samples included (Mexico to Panama), and one limited to samples from Costa Rica.

Inferring Ancestral Geographic Ranges—To infer the ancestral geographic range for *H. erato*, we used the Bayesian binary MCMC (BBM) method implemented in the program RASP 2.0 beta (Yu *et al.* 2011). This analysis was based on gene trees for *COI-II* and *Tpi*, the two genes for which we had data from across the range of the species. For this analysis, each gene was analyzed separately and only unique haplotypes were considered. Gene trees were estimated using BEAST 1.6 (Drummond & Rambaut 2007), under the GTR+I+G model, with 10 million generations of MCMC, sampling every 1000 generations. *H. clysonymus* served as an outgroup for both genes. Uncertainties in tree topology were accounted for by using 9000 trees for BBM. Each unique haplotype was assigned to one or more of the following geographic regions based on sampling location: A) north of Costa Rica mountains, B) west of the Andes (including Panama & southern CR), C) Guiana Shield, D) Amazon, and E) coastal Brazil (*COI-II* analysis only). BBM analyses utilized 10 MCMC chains, run simultaneously for 5 million generations, sampling every 100 generations.

Wing color pattern

To examine fine-scale changes in wing pattern across Central America, we took measurements of the yellow hindwing bar from digital photographs of 172 *H. erato* (26 sites). Individuals from our genetic analyses were used when available ($n = 62$). Our analysis also included re-measurement of individuals analyzed by Gilbert (1983). Specimens were oriented in the same plane as the camera using spirit levels. Size was calibrated by including a scale bar in each image. Measurements were taken using ImageJ 1.43u (<http://rsbweb.nih.gov/ij/download.html>). The width of the dorsal hindwing yellow bar was measured at its widest point by taking a measurement perpendicular to the bar at the location where it crosses vein Cu1. Associations with body size were accounted for by dividing by forewing length to give relative width of the hindwing yellow bar. A one-tailed Student's t-test assuming equal variances (Microsoft Excel v.14.2.2) was used to test the hypothesis that within Costa Rica, Pacific-side localities (samples from the E1 and E2 clades) have greater relative width of the dorsal hindwing yellow bar.

Cline analysis

To investigate the geographic variation in wing phenotype, mtDNA genotype and AFLP genotype, we compared the fit of a linear, a step cline and a four-parameter logistic (here after “sigmoid”) model for each dataset. For the wing phenotype models, the y-variable (Y_i) was relative width of the dorsal hindwing yellow bar data for the i individuals scaled to values between 0 and 1, with one minus the scaled value used in analyses in order to facilitate comparison with the mtDNA and AFLP datasets. For mtDNA the y-variable (Y_i) was the percent of individuals in each population belonging to clade E3 for each of i populations. For the AFLP data the y-variable (Y_i) was the fraction of each of i individual's

genotype that was “northern” as inferred from our Structure analysis. For all three data sets, the x-variable (X_i) for each sample locality i was geographic distance. Geographic distance was calculated relative to San Jose, Costa Rica with Torti, Panama set to 0 km.

$$\text{The step cline model was: } Y_i = \begin{cases} y_1 & \text{if } X_i \leq x_{mid} \\ y_2 & \text{if } X_i > x_{mid} \end{cases} \quad (1)$$

$$\text{The sigmoid model was: } Y_i = y_1 + \frac{y_2 - y_1}{1 + e^{(x_{mid} - X_i)/w}} \quad (2)$$

where y_1 is a lower y-axis asymptote, y_2 is an upper y-axis asymptote, x_{mid} is the cline midpoint, and w is the cline width (Fitzpatrick & Shaffer 2007).

Models were fit with maximum likelihood in R using `mle2()`. The corrected Akaike information criterion (AICc in R) was used to compare the linear model to the step model. The difference in AICc value between the models was used to assess their relative fit to the data, with a difference of 6.0 or greater indicating that a more complex model was favored with 95% certainty (Motulsky & Christopoulos 2003). Likelihood ratio tests (LRT) with one degree of freedom were used to compare the step and sigmoid models, since as the value of w becomes very small, the sigmoid model approaches the step model (Fitzpatrick & Shaffer 2007). Confidence intervals (95%) for cline midpoint (x_{mid}) and cline width (w) for each dataset were constructed from the standard errors of parameter values output from `mle2()`. These confidence intervals and the likelihood profiles were used to investigate the correspondence of cline midpoint and cline width for wing phenotype, mtDNA genotype and AFLP genotype.

Results

Genetic differentiation

We discovered deep mitochondrial divergence, represented by a new “north” / E3 clade, in *H. erato* in northern Central America (Figure 1). All but one of the 72 samples (98.6%) belonging to the new E3 clade originate north of the mountain chain that divides Costa Rica into Pacific and Atlantic drainages (Figure 1). The single exception is a sample from San Jose located near the northern crest in the Central Valley (sample LG-7, site 8 in Figure 1). Samples north of the mountains that divide Costa Rica were significantly and highly differentiated from samples south of the mountains at mtDNA genes *CoI-CoII* (MX - PA Fst = 0.75, $P < 0.001$; $N = 85$ north and $N = 69$ south; see Table 2 for results using samples limited to Costa Rica). The same pattern was seen when analyzing individuals with only *CoI* “barcode” sequence (Supplementary Figure S1).

The other two mtDNA lineages were made up almost entirely of samples from south of the mountains that divide Costa Rica. The exceptions to perfect geographic association were entirely limited to samples from Costa Rica. The “western” or E1 mtDNA clade was mainly comprised (89.1%) of samples from southern Costa Rica and southward (Panama and into northwest South America west of the Andes mountains). Individuals in the E1 clade that were not from southern localities were mostly from mountain valleys near San Jose (10 individuals from sites 9 and 10) with a single individual from the Atlantic Coast (site 11; Figure 1). The “eastern” or E2 clade was primarily comprised (94%) of samples from South America east of the Andes mountains, but had three individuals from northern Costa Rica localities (sites 14, 16, 17; Figure 1).

Genetic differentiation associated with geography was also evident throughout the nuclear genome (Table 2, Figure 2, Figure S2). *Heliconius erato* was strongly differentiated into north and south populations in Central America at the nuclear loci *Tpi* (MX - PA $F_{st} = 0.34$, $P < 0.001$; $N=36$ north and $N=36$ south) and *TH* (MX - PA $F_{st} = 0.26$, $P < 0.001$; $N=36$ north, $N=27$ south) (See Table 2 for results limited to Costa Rica). Moreover, AFLP markers (2201 polymorphic AFLP loci) indicated significant differentiation between north and south populations ($F_{st} = 0.12$, $P < 0.001$). Limiting the analyses for these nuclear data to just Costa Rica also strongly rejected the presence of a single population (Table 2). Genotypic clusters analyzed with Structure always found a clear genetic break across the mountains of Costa Rica, even with $K = 2$ groups. The best fitting number of groups was $K=4$ corresponding to Mexico, Northern Costa Rica to Mexico, Southern Costa Rica southward, with a fourth group representing variation from South America (Figure 2).

Col dating results

The mtDNA maximum likelihood tree indicated that the new E3 lineage in *H. erato* branched before the origin of the other two main mtDNA lineages (E1 & E2; Figure 1). Using sequence data for *CoI* (766 bp), the mean pairwise divergence between the main mtDNA clades were: E3 - *H. himera* = 4.5, E1 - E3 = 4.6, E2 - E3 = 4.8, with an overall average of 4.63%. Applying a rate of 1.5% uncorrected pairwise divergence per My (Quek et al 2004, 2010) resulted in a minimum age for the origin of *H. erato* of 3.09 million years.

Ancestral Geographic Ranges

BBM analysis of both *CoI-CoII* and *Tpi* yielded similar probability estimates for the ancestral range of *H. erato*. For *CoI-CoII*, the highest probability (0.73) ancestral range encompassed two regions, north of the Costa Rica mountains and west of the Andes. The next highest probability range was north of the Costa Rica mountains by itself (0.16), followed by west of the Andes by itself (0.10). For *Tpi*, the highest probability (0.70) ancestral range was again north of the Costa Rica mountains combined with west of the Andes. The next highest probability range was west of the Andes, by itself (0.28) and there was no third estimate.

Cline analysis

We observed a marked shift in the yellow bar wing phenotype, mtDNA and AFLP markers from west Panama to Mexico (Figure 3). For the relative width of the dorsal hindwing yellow bar (Figure 3A) the greatest change appeared within Costa Rica. The relative width of the dorsal hindwing yellow bar changed across the mountains, being significantly smaller for *H. erato* from north localities ($p < 0.0001$, $t = 4.89$; north: mean = 0.073, s.d. = 0.010, $N = 33$; south: mean = 0.082, s.d. = 0.008, $N = 66$). For the cline in this yellow bar phenotype, the sigmoid model (using 1-scaled value of relative bar width) was significantly better than the step cline (likelihood ratio = 21.3, P -value < 0.00001, Table 3), and had a much greater probability (>0.999) compared with the linear model ($\Delta AIC = 58.3$). For the cline in mtDNA, the step cline model had a much greater probability (>0.999) compared with the linear model ($\Delta AIC = 24.4$), but the sigmoid model was not significantly better than the step cline (likelihood ratio = -0.2, P -value > 0.99, Table 3, Figure 3B). For the cline in AFLP data, the sigmoid model was significantly better than the step cline (likelihood ratio = 84.5, P -value < 0.00001, Table 3, Figure 3C), and had a much greater probability (>0.999) compared with the linear model ($\Delta AIC = 311.3$).

The cline midpoints of the three datasets were broadly coincident with each other and were all estimated to lie near the northern crest of the mountains crossing Costa Rica (Table 3, Figure 3). The north crest lies at 655 km on the transect when measured at the east-west midpoint of the mountain range (San Jose), with the south crest at 609 km. Both crests are

>2200 m elevation at this position. The 95% confidence interval for cline midpoint of the wing phenotype overlapped both mtDNA and AFLP estimates (Table 3). However, the 95% confidence intervals for cline midpoint of mtDNA and AFLP indicated that the midpoint for mtDNA is 21-33 km farther north (Table 3, Figure 4). Estimates of cline midpoint were very similar for mtDNA and AFLP sigmoid models (Table 3). The similarity in cline midpoint estimates for each of the best fitting models are shown by their likelihood profiles in Figure 4. The likelihood surface for the mtDNA step model shows a second peak estimate for cline midpoint between 618-634 km, but the peak likelihood from 668-680 km was a significantly better fit (likelihood ratio = 9.0, P -value = 0.003).

The cline widths for mtDNA and AFLP were much narrower than for the wing phenotype (Figure 4). The mtDNA dataset had the narrowest cline width (0 km) as a result of the favored step cline model. Estimates of cline width were very similar for mtDNA and AFLP sigmoid models (Table 3).

Discussion

We discovered deep mtDNA divergence within *H. erato* in Central America (Figures 1B and S1) and this genetic break is corroborated by nuclear differentiation at two Z-linked genes as well as genome-wide AFLP markers (Figures 2 and S2). There is a striking geographic pattern to this genetic divergence (Table 2). The new mtDNA lineage is nearly fixed in *H. erato* populations from northern Costa Rica and northward but virtually absent from populations farther south (Figure 1A). This cryptic lineage is novel, despite previous widespread sampling and extensive research focused on *H. erato*.

Previous work using mtDNA has identified two main clades in *H. erato* – one east and one west of the Andes (Brower 1994; Brower 1996; Quek *et al.* 2010). The new mtDNA lineage from northern Central America branches basally to these other clades and has important implications for our understanding of mimicry evolution in *Heliconius*. First, the new mitochondrial lineage suggests that *H. erato* originated approximately 3.1 million years ago, making it as much as 1 million years older than its co-mimic, *H. melpomene*. The older age for *H. erato* lends strong support to the hypothesis that *H. erato* radiated prior to *H. melpomene*, as opposed to the two species diversifying in parallel. Second, ancestral state reconstructions of historical ranges, based on both mtDNA and a nuclear gene, suggest that *H. erato* originated west of the Andes, perhaps even in northern Central America. This is in contrast to prior work, which indicated a center of origin for *H. erato* on the western slope of the Andes (Quek *et al.* 2010). This is also very different than *H. melpomene*, which appears to have originated on the opposite side of the continent in coastal Brazil (Quek *et al.* 2010). Our results highlight the very different evolutionary histories of two taxa that today share nearly identical phenotypic variation and participate in an intricate mimetic relationship throughout their range.

If *H. erato* did originate in Central America, its ancestral phenotype may have been the red and yellow banded “postman” phenotype that exists there now. Interestingly, this is also reflected in recent work on the red wing patterning gene *optix*. Phylogeographic analyses based on *optix* suggest that the red banded phenotype is ancestral to other wing pattern phenotypes in *H. erato* (Hines *et al.* 2011). This scenario is consistent with the geographic distribution of *H. erato* subspecies (Mallet 2010). Today, peripheral populations of postman and red-banded phenotypes surround rayed races, which occupy the Amazon basin. It has been assumed that in systems like *Heliconius*, where there is substantial gene flow among phenotypically distinct subspecies, markers unlinked to the traits themselves would be insufficient to reveal the history of phenotypic evolution (Hines *et al.* 2011). Here we have demonstrated that this is not the case; comprehensive sampling, combined with robust

analytical methods, uncovers patterns in unlinked genetic markers that corroborate and extend the story emerging from color patterning genes.

The distribution of genetic variation in *H. erato* suggests two possible evolutionary histories for this species. First, *H. erato* may have originated north of the mountains in Costa Rica and spread southward, leaving ancestral haplotypes behind. A second possibility is that *H. erato* originated farther south, and then colonized Central America in two waves, an original event that populated all of Central America and Mexico with ancestral genetic variation, followed by a second event that only moved as far north as the Costa Rican mountains. It is worth noting that the Isthmus of Panama is estimated to have been in place by 3 mya (Coates & Obando 1996; Kirby *et al.* 2008), at about the same time as our date for the origin of *H. erato* (3.1 mya). The rising isthmus may have led to increased connectivity among populations, and could have facilitated either of these scenarios.

The first scenario described above is the simplest explanation for our results, but the second scenario could produce similar patterns if new mtDNA haplotypes were able to completely replace ancestral ones south of the mountains. Distinguishing between these scenarios is difficult but at least one observation supports the first scenario – the presence of diverse mtDNA haplotypes in northern Central America. For instance, we find mtDNA haplotypes that fall into E1, E2, and E3 clades all present in northern Costa Rica, with those from the E2 clade (otherwise present only east of the Andes) being quite divergent from the rest. This observation supports the idea that the area north of the Costa Rican mountains may be the ultimate source of *H. erato* and much of its genetic variation.

Our analyses also revealed a subtle, but distinct, phenotypic transition in *H. erato* that is associated with the genetic break in Costa Rica. Gilbert (1983) studied the cline of narrowing hindwing band width from Costa Rica northward and suggested the band in *H. erato* narrowed as a result of a lack of selection to match its wide-banded *H. melpomene* co-mimic, which becomes less abundant in northern Costa Rica and beyond. Our phenotypic analyses, which are based on widespread sampling across Central America and include Gilbert's (1983) original specimens, verify the presence of a band width cline but show that the sharpest transition is associated with the transition in mtDNA and AFLP markers in Costa Rica (Figure 3).

Indeed, the cline midpoint estimates of the three datasets were all near the north crest of the mountains dividing Costa Rica (Figure 3, Table 3). It is important to note that this mountain range serves as a transition point for a variety of other butterflies as well. For instance, northern populations of *Heliconius sapho* are replaced by *H. hewitsoni* on the southwest side of the mountains, and their co-mimics, *H. cydno* and *H. pachinus* have matching distributions. Similarly, *H. sara* exists as different subspecies on the two sides, as do a variety of ithomiines (DeVries 1987). Generally, these patterns point to a significant role for the Cordilleras de Guanacaste, de Tilarán, de Central and de Talamanca in shaping the biogeography of Central America, similar to the role of the Andes in South America.

One final point to note is that the results presented here complicate the taxonomy of *H. erato* in Central America. Currently, four subspecies of *H. erato* are described in Central America. The subspecies are recognized by variation in the yellow hindwing bar as well as additional wing pattern characters (Lamas 1998). *H. e. cruentus* inhabits the Pacific side of Mexico, Guatemala and El Salvador. *H. e. petiverana* inhabits the Caribbean side of northern Central America from Mexico to Nicaragua, with *H. e. demophaon* occurring on both coasts from Nicaragua to W. Panama. These three subspecies all display the hindwing yellow bar. *H. e. hydara*, a subspecies lacking the yellow bar, occurs in eastern Panama and farther south into South America (Lamas 2004). While our results support four genetic groups in Central

America, the distributions do not coincide with these subspecies. In particular, the distribution of *H. e. demophoon* currently encompasses the genetic break in Costa Rica, and the type locality for this subspecies is in Nicaragua. Our data are consistent with populations north and south of the Costa Rican mountains being different subspecies. Thus, *demophoon* is the most appropriate name for the northern subspecies inhabiting southern Nicaragua and northern Costa Rica (Lamas 2004). However, the yellow-banded form from southern Costa Rica and northern Panama may eventually need to be described as a new subspecies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. Abbott for specimens from Nicaragua, and L. Go, L. Cabour, N. Gillingham, and G. Johnson for help with lab work. Thank you to R. Papa for providing artwork and images used in Figure 1, to J. Mayberry for discussing the cline analysis, and to University of the Pacific Experiential Learning students for discussions of the manuscript. We also thank Y. Bassett, of the Smithsonian Tropical Research Institute, for permission to use *H. erato* *Col* data from Barro Colorado Island, and N. Chamberlain, B. Counterman, K. Kunte, O. McMillan, S. Mullen, S-P. Quek, and W. Savage for help with various aspects of this project. This work was funded by NIH NIGMS Grant GM068763 and NSF grant DEB-1020355 to MRK. We thank MINAET (Costa Rica), Universidad Zamorano (Honduras), and MECN & Ministerio Ambiente (Ecuador) for permits.

Literature Cited

- Akaike H. A new look at statistical model identification. *IEEE Trans Automat Control*. 1974; 19:716–723.
- Baxter SW, Nadeau N, Maroja L, et al. Genomic hotspots for adaptation: the population genetics of Müllerian mimicry in the *Heliconius melpomene* clade. *PLoS Genetics*. 2010; 6:e1000794. [PubMed: 20140188]
- Beltran M, Jiggins CD, Bull V, et al. Phylogenetic discordance at the species boundary: Comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Molecular Biology and Evolution*. 2002; 19:2176–2190. [PubMed: 12446809]
- Brower AVZ. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America*. 1994; 91:6491–6495. [PubMed: 8022810]
- Brower AVZ. Parallel race formation and the evolution of mimicry in *Heliconius* butterflies: a phylogenetic hypothesis from mitochondrial DNA sequences. *Evolution*. 1996; 50:195–221.
- Brown KS Jr. The Biology of *Heliconius* and related genera. *Annual Review of Entomology*. 1981; 26:427–456.
- Brown KS, Sheppard PM, Turner JRG. Quaternary Refugia in Tropical America: Evidence from Race Formation in *Heliconius* Butterflies. *Proceedings of the Royal Society of London Series B-Biological Sciences*. 1974; 187:369–378.
- Coates, AG.; Obando, JA. Geological evolution of the Central American Isthmus. In: Jackson, JBC.; Budd, AF.; Coates, AG., editors. *Evolution and Environment in Tropical America*. University of Chicago Press; Chicago: 1996. p. 21-56.
- Counterman BA, Araujo-Perez F, Hines HM, et al. Genomic hotspots for adaptation: the population genetics of Müllerian mimicry in *Heliconius erato*. *PLoS Genetics*. 2010; 6:e1000796. [PubMed: 20140239]
- Dasmahapatra KK, Elias M, Hill RI, Hoffman JI, Mallet J. Mitochondrial DNA barcoding detects some species that are real, and some that are not. *Molecular Ecology Resources*. 2010; 10:264–273. [PubMed: 21565021]
- DeVries, PJ. *The butterflies of Costa Rica and their natural history Volume I: Papilionidae, Pieridae, Nymphalidae*. Princeton University Press; Princeton, New Jersey: 1987.

- Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*. 2007; 7:214. [PubMed: 17996036]
- Eltringham H. On specific and mimetic relationships in the genus *Heliconius*. *Transactions of the Entomological Society of London*. 1916; 1916:101–148.
- Excoffier L, Laval G, Schneider S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*. 2005; 1:47–50. [PubMed: 19325852]
- Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*. 2007; 7:574–578. [PubMed: 18784791]
- Fitzpatrick BM, Shaffer HB. Introduction history and habitat variation explain the landscape genetics of hybrid tiger salamanders. *Ecological Applications*. 2007; 17:598–608. [PubMed: 17489263]
- Flanagan NS, Tobler A, Davison A, et al. Historical demography of Müllerian mimicry in the neotropical *Heliconius* butterflies. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101:9704–9709. [PubMed: 15210977]
- Gilbert, LE. Coevolution and mimicry. In: Futuyma, DJ.; Slatkin, M., editors. *Coevolution*. Sinauer Associates, Sunderland; Massachusetts: 1983. p. 263–281.
- Heliconius_Genome_Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*. 2012; 487:94–98. [PubMed: 22722851]
- Hines HM, Counterman BA, Papa R, et al. Wing patterning gene redefines the mimetic history of *Heliconius* butterflies. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:19666–19671. [PubMed: 22084094]
- Kirby MX, Jones DS, MacFadden BJ. Lower Miocene Stratigraphy along the Panama Canal and Its Bearing on the Central American Peninsula. *PLoS ONE*. 2008; 3:e2791. [PubMed: 18665219]
- Lamas G. Comentarios taxonomicos y nomenclaturales sobre Heliconiini neotropicales, con designacion de lectotipos y descripcion de cuatro subespecies nuevas (Lepidoptera: Nymphalidae: Heliconiinae). *Revista Peruana de Entomologia*. 1998; 40:111–125.
- Lamas, G. *Atlas of Neotropical Lepidoptera - Checklist: Part 4A Hesperioidea - Papilionoidea*. Scientific Publishers; Gainesville, FL: 2004.
- Mallet J. Shift happens! Shifting balance and the evolution of diversity in warning colour and mimicry. *Ecological Entomology*. 2010; 35(Suppl. 1):90–104.
- Motulsky, HJ.; Christopoulos, A. *Fitting models to biological data using linear and non-linear regression A practical guide to curve fitting*. GraphPad Software, Inc.; San Diego, CA: 2003.
- Müller, F. *Ituna* and *Thyridia*, a remarkable case of mimicry in butterflies. In: Meldola, R., translator. *Proceedings of the Entomological Society of London*. Vol. 1879. 1879. p. xx-xxix.
- Nadeau NJ, Whibley A, Jones RT, et al. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B-Biological Sciences*. 2012; 367:343–353.
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155:945–959. [PubMed: 10835412]
- Quek SP, Counterman BA, Albuquerque de Moura P, et al. Dissecting comimetic radiations in *Heliconius* reveals divergent histories of convergent butterflies. *Proceedings of the National Academy of Sciences, USA*. 2010; 107:7365–7370.
- Quek SP, Davies SJ, Itino T, Pierce NE. Codiversification in an ant-plant mutualism: Stem texture and the evolution of host use in *Crematogaster* (Formicidae: Myrmicinae) inhabitants of *Macaranga* (Euphorbiaceae). *Evolution*. 2004; 58:554–570. [PubMed: 15119439]
- Reed RD, Papa R, Martin A, et al. *optix* Drives the Repeated Convergent Evolution of Butterfly Wing Pattern Mimicry. *Science*. 2011; 333:1137–1141. [PubMed: 21778360]
- Sheppard PM, Turner JRG, Brown KS Jr, Benson WW, Singer MC. Genetics and the evolution of Müllerian mimicry in *Heliconius* butterflies. *Philosophical Transactions of the Royal Society of London B Biological Sciences*. 1985; 308:433–613.
- Tamura K, Peterson D, Peterson N, et al. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*. 2011; 28:2731–2739. [PubMed: 21546353]

- Turner, JRG. Studies of Müllerian mimicry and its evolution in burnet moths and heliconid butterflies. In: Creed, ER., editor. Ecological Genetics and Evolution. Blackwell, Oxford; UK: 1971. p. 224-260.
- Turner JRG. Adaptive radiation and convergence in subdivisions of butterfly genus *Heliconius* (Lepidoptera-Nymphalidae). Zoological Journal of the Linnean Society. 1976; 58:297–308.
- Turner JRG, Mallet JLB. Did forest islands drive the diversity of warningly coloured butterflies? Biotic drift and the shifting balance. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences. 1996; 351:835–845.
- Yu, Y.; Harris, AJ.; He, XJ. RASP (Reconstruct Ancestral State in Phylogenies) 20 beta. 2011. Available at <http://mnh.scu.edu.cn/soft/blog/RASP>

Figure 1.

Wing pattern diversity and mtDNA *Col-II* results. In panel A the distributions of nearly all *Heliconius erato* color pattern races are outlined with black lines. The map is color coded to indicate which of five major color pattern groups is present, with tan representing the "postman" phenotype. The inset of Costa Rica shows numbered sampling localities and the transition in mtDNA across the mountains, as indicated by pie charts with black portions representing the fraction of each population belonging to clade E3. The "X" on the map represents the location (P.N. Guanacaste) of several *Col* barcode individuals analyzed in Figure S1. In panel B, the novel mitochondrial clade discovered here is labeled as E3, with E2 and E1 east and west of the Andes respectively. Sample size is indicated in parentheses for each country, with colored dots corresponding to the major color pattern group included in the analysis. For samples from Costa Rica, the number preceding the sample size is the locality indicated on the map in panel A. Numbers at nodes are ML bootstrap values. Scale bar is average number of substitutions per site.

Figure 2.

AFLP Structure plot. Each vertical line is an individual, with colors indicating genotypic clusters. The value K corresponds to the number of *a priori* groups in each analysis and lnL is the ln-likelihood of the data given that value of K . The asterisk by $K=4$ indicates this is the best fitting number of genotypic clusters. Nine sampling locations, in geographic order, are highlighted: Mexico, Honduras, Nicaragua, northern Costa Rica, northward draining portions of Costa Rica's Central Valley, southward draining portions of the Central Valley, southern Costa Rica, *H. e. demophoon* from Panama, and *H. e. hydara* from Panama.

Figure 3.

Clinal variation in *Heliconius erato* hindwing yellow bar, mtDNA and AFLP markers across Central America. Torti, Panama is at 0 km and populations in Costa Rica lie between 400 and 800 km. The vertical gray shading represents the location of the mountains dividing Costa Rica, as measured by the location of the crests due north and due south of San Jose. San Jose lies in the central valley between the crests which are >2200m elevation, and represents the approximate center of the mountains from east to west. The south crest is at 609 km along the transect and the north crest is at 655 km. Representative phenotypes above the graphs are from: the Osa Peninsula in southern Costa Rica (left), near Arenal in Northern Costa Rica (middle), and near San Juan de Lima (Michoacan) in western Mexico (right). The fitted sigmoid and step cline models are shown. For the cline in hindwing yellow bar shown in A each point represents a single individual. Relative yellow bar width was scaled between 0 and 1, with 1 minus this scaled value plotted to facilitate comparison with the genetic data. For mtDNA data in B, each point represents a population for which the fraction of members belonging to the E3 clade was calculated. For AFLP data in C, each point

represents an individual for which the proportion of northern ancestry based on the Structure analysis was calculated.

Figure 4.

Maximum likelihood profiles for the cline midpoint (A), and cline width (B) in hindwing yellow bar width, mtDNA and AFLP data. The likelihood profile for each of the best fitting models is shown. Cline width is broadly coincident for the three datasets in A. The three datasets have different cline widths as indicated in B. Note that there is no cline width curve for mtDNA in B since the step model was favored for that dataset and cline width is 0 km by definition.

Summary of taxonomy, geographic location and sample size of *H. erato* individuals used in this study.

Table 1

<i>H. erato</i> subspecies	Country	Sample size (# of individuals)			
		<i>Col-Col</i>	TH	Tpi	AFLP
<i>chestertonii</i>	Colombia	7			
<i>cruentus</i>	Mexico	16	4	4	14
<i>cyrbia</i>	Ecuador	4			
<i>denophoon</i>	Costa Rica	109	44	42	100
	Nicaragua	1	1	1	1
	Panama	14		7	9
<i>emma</i>	Peru	5			
<i>erato</i>	French Guiana	8			
<i>erylus</i>	Ecuador	10			
<i>favorinus</i>	Peru	4			
<i>hydara</i>	Colombia	5			
	French Guiana	4			
	Panama	5			4
	Trinidad	6			
<i>lativitta</i>	Ecuador	6			
<i>petiverana</i>	Honduras	20	6	6	16
	Mexico	2			
<i>phyllis</i>	Brazil	7			
Total =		233	55	60	144

Table 2

Results of AMOVA's testing for genetic differentiation with different data sets and geographic regions. Sample size (*N*) here refers to number of individuals for mtDNA and AFLP, and number of sequences for *Tpi* and *TH*.

Genetic Marker	Mexico to Panama		Costa Rica only	
	<i>F_{st}</i>	<i>P</i>	<i>F_{st}</i>	<i>P</i>
		<i>N</i> (north; south)		<i>N</i> (north; south)
<i>Col-Col</i>	0.75	< 0.001	0.66	< 0.001
AFLP	0.12	< 0.001	0.10	< 0.001
<i>Tpi</i>	0.34	< 0.001	0.21	0.002
<i>TH</i>	0.26	< 0.001	0.19	< 0.001

Cline analysis results for yellow bar phenotype, AFLP genotype and mitochondrial clade. For cline midpoint and cline width, the 95% confidence intervals are given in parentheses. The best fitting model is indicated in the “Model” column with three asterisks (***). Note that $y1$ and $y2$ values for hindwing yellow bar width are the 1-scaled values (see Methods).

Table 3

Model	df	Fitted asymptote values ($y1/y2$)	Midpoint (km)	Width (km)	Log likelihood	Likelihood ratio test P -value §
Hindwing yellow bar						
Step	3	0.32 / 0.62	679 - 681.5	NA	77.66	
Sigmoid***	4	0.22 / 0.66	667.8 (719.5, 616.1)	90.2 (132.7, 47.7)	88.32	< 0.00001
AFLP genotype (% north)						
Step	3	0.27 / 0.90	668.0 - 680.4	NA	108.86	
Sigmoid***	4	0.14 / 0.93	646.9 (653.4, 640.4)	31.7 (36.8, 26.6)	151.15	< 0.00001
mtDNA genotype (% E3 clade)						
Step***	3	0.12 / 0.94	668.0 - 680.4	NA	2.94	
Sigmoid	4	0 / 1	661.5 (688.7, 634.3)	26.7 (46.9, 6.5)	2.84	> 0.999

§ step cline vs sigmoid model