# Supporting Information

## Quek et al. 10.1073/pnas.0911572107

### SI Materials and Methods

**AFLP Scoring Details.** Eight AFLP primer pairs were used for selective amplification: ACT-CAT, ACA-CAT, ACT-CTG, ACA-CTG, ACT-CTT, ACA-CAG, ACT-CTT, and ACA-CAG. Fragments were electrophoresed on an ABI 3730XL DNA analyzer (Applied Biosystems) with GeneScan 500 ROX size standard (Applied Biosystems). To ensure consistency among electrophoretic runs, a control comprising the same sample amplified with the same primer pair was included in every run, and the fragment profile from this control was compared across runs by eye. Fragments were scored separately for *H. erato* and *H. melpomene* using GeneMapper 4.0 (Applied Biosystems). Only fragments ranging from 50 to 500 bp were scored, and within-project normalization was enabled. The advanced peak detection algorithm was used, with light smoothing turned on and all other settings left at defaults. As suggested by Holland et al. (1), we explored scoring with various bin widths, namely 0.5, 0.7, and 0.9. All other scoring parameters were left at default settings. In preliminary analyses, bin widths of 0.5 produced topologies with the best resolution, so we used this bin width for our final analyses. In fast-evolving molecular markers such AFLPs in which scoring is binary, allele scoring within the ingroup can be affected by similarly sized but homoplasious peaks in the outgroups. The tendency for homoplasious peaks to be scored as homologues is expected to increase with narrower bin widths, and the degree of homoplasy is expected to increase with increasing distance to the outgroups. For this reason, we scored the ingroup both with and without outgroups. Likewise, certain groups within the ingroup were isolated and additionally scored on their own (after preliminary analyses confirmed that those groups were monophyletic).

**Structure-Based Clustering.** For each species, we first performed two series of short clustering runs ($10^5$ burn-in, $10^5$ data collection) using the admixture model and varying the number of clusters from K = 1 to K = 13. These initial runs revealed maxima in the log-likelihood profiles at K = 6 and K = 7 for both species. We then ran two series of long clustering runs ($2 \times 10^5$ burn-in, $10^6$ data collection) across the range of K = 5 to K = 8. All Structure analyses were based only on genetic data (no prior population information was used). Clusters that consisted of more than one population were also analyzed separately by limiting each dataset to the individuals and polymorphisms in the group under investigation and setting K equal to the number of populations in the cluster.

### SI Discussion

**Geographic Origins of *H. erato* and *H. melpomene*.** There is additional but indirect evidence supporting the inferred geographic origins of *H. erato* and *H. melpomene* from their AFLP data (Figs. 1 *A* and *C* and 2 *A* and *C* in the main text). Apart from *H. himera* (previously inferred to have peripatrically speciated and maintained genetic isolation from *H. erato*), and apart from the comimetic races *H. e. cyrbia* and *H. m. cythera* (which each form their own clades apart from other Ecuadorian samples), the remaining *H. erato* races that occur in Ecuador are monophyletic (*etylus*) or nearly monophyletic (*lativitta*), whereas the *H. melpomene* races (excluding *cythera*) occurring in Ecuador (*plesseni, ecuadoriensis*, and *aglaope*) do not form their own distinct clades but are interspersed with one another within a single clade (Figs. 1*A* and 2*A* in the main text). A similar phenomenon is seen in Colombia: *H. erato* entities are phylogenetically clumped but *H. melpomene* races (*vulcanus* and *melpomene*) are interspersed with each other. Structure analyses also concur with the presence of well separated *H. erato* but not *H. melpomene* races in Ecuador and Colombia (Figs. 1*C* and 2*C* in the main text). These observations suggest that *H. erato* has been present in Ecuador and Colombia long enough for its races to diverge at a sufficient number of loci, and that *H. melpomene* has a more recent history there. These findings, in combination with the basal phylogenetic position(s) of some Ecuadorian races, point to an origin of *H. erato* in or near Ecuador.

In French Guiana, the converse is seen. *H. melpomene* races are phylogenetically clumped, whereas *H. erato* races are interspersed with each other. *H. melpomene* in the eastern tip of the continent (coastal Brazil) also exhibits more fine-scale population genetic structure than *H. erato* there (P.A.d.M. et al., unpublished observations). These findings suggest a greater antiquity of *H. melpomene* relative to *H. erato* in the eastern parts of the continent and are consistent with an origin of *H. melpomene* in the east. Future work with more comprehensive sampling will be required to fully resolve the geographic origins of *H. melpomene* and *H. erato*.

**Colombian Ancestry for Isthmus Populations and for the Comimics *H. e. cyrbia* and *H. m. cythera*.** A common geographic trend between the AFLP phylogenies of *H. erato* and *H. melpomene* is that Costa Rica specimens are nested within Panama (Figs. 1*A* and 2*A* in the main text), and that specimens from the isthmus (Costa Rica and Panama) are of Colombian origin. In *H. melpomene*, this is evident in the paraphyletic grade of Colombian samples basal to isthmus specimens (Fig. 2*A* in the main text). Applying a parsimonius ancestral area reconstruction to the *H. erato* phylogeny also results in Colombian ancestry for specimens from the isthmus. This is perhaps unsurprising, as Colombia is at present the only terrestrial gateway to the isthmus. By the same measure, Colombian origins are also inferred for *H. e. cyrbia* and *H. m. cythera* in Ecuador.

**Final Uplift of the Northern Andes and the Biogeography of *H. erato* and *H. melpomene*.** The uplift of the Andes has been a potent force in shaping the distribution and diversification of many neotropical taxa (2). In the duration between approximately 2 and 4 Mya, the eastern cordillera of the Colombian Andes achieved at least 60% of its current elevation (3), and this likely explains the mtDNA distributions for *H. erato* and *H. melpomene* noted previously (4, 5). The mtDNA trees presented here also show ancestral Andean divisions, but our data additionally reveal secondary eastward migration in both species: clades E1 and M2d [i.e., western clades in *H. erato* and *H. melpomene*, respectively, described by Brower (4)] contain individuals from the eastern cordillera of the Colombian Andes (and Trinidad for E1).

For *H. erato*, the AFLP tree also shows a well supported clade residing west of the Andes. However, unlike the mtDNA tree, it does not create a basal split in the phylogeny but appears in a more derived position (note, however, that nodes along the backbone of the AFLP tree are poorly supported). This ambiguity is reflected in other nuclear regions: based on two nuclear genes, Flanagan et al. (6) found some genetic differentiation, as well as substantial shared variation, between *H. erato* populations on either side of the Andes, and concluded that this reflected high levels of historical gene flow before the establishment of the Andes as a significant dispersal barrier.

In the *H. melpomene* AFLP tree, all of the samples from west of the Andes are contained in one large clade (gray box in Fig. 2*A* in the main text) which also contains two individuals from the eastern cordillera of the Colombian Andes. Because most of the nodes along the backbone of the AFLP tree are well supported, "west of Andes" can be unambiguously placed as a derived biogeographic region for *H. melpomene*. The clear Andean division seen in the AFLP data are also broadly reflected in the investigation by Flanagan et al. (6), who noted three biogeographic regions (west of Andes, Amazonia, and Guiana Shield).

**Inferring Color Pattern Evolution.** Although our genome-wide AFLP data are consistent with multiple origins of the major color pattern phenotypes in each species, there is one caveat associated with our analyses of color pattern evolution. Our approach is based on the standard phylogenetic method of inferring trait evolution by mapping the trait onto a phylogeny of closely related taxa. However, there is substantial hybridization among color pattern races within both *H. erato* and *H. melpomene* (7), and hybridization between each species and its closest relatives (8). There are two different ways in which hybridization and resulting gene flow could cause the evolutionary history of the wing pattern to become disconnected from the genome-wide background inferred from our AFLP data. First, extensive gene flow among geographically proximate races could homogenize genetic variation on a local scale and effectively erase the historical information in the AFLP data. Hence, it is possible that that each phenotype originated once and subsequent gene flow among neighboring populations has erased the signature of a single origin. However, our data suggest that this is not the case. For instance, in the *H. erato* AFLP tree, sister taxa tend to be widely dispersed geographically, and in the *H. melpomene* AFLP tree, there is relatively good bootstrap support for higher-level relationships. Both of these results are unlikely in the face of extensive, recent gene flow.

A second way in which the evolutionary history of color pattern could become disconnected from the rest of the genome is if hybridization selectively transferred alleles at color patterning loci across racial or species boundaries. Although good evidence for this phenomenon is lacking, it is strongly suspected that introgression of color patterning traits has occurred among closely related *Heliconius* species (9–11). Furthermore, some of the apparent discordance between our AFLP and mtDNA trees could be the result of introgression (as described later). It is important to note that color pattern introgression is expected to cause geographically proximate populations to look more similar, so this phenomenon could apply to the rayed or red patch phenotypes. For both of these, adjacent populations have similar phenotypes but fall out in different locations on the AFLP trees. In contrast, this explanation is unlikely to apply to the postman pattern because different populations with this phenotype are geographically disjunct in both *H. erato* and *H. melpomene*.

**Mitochondrial Versus Nuclear Genomes: Discordance or Resolution?** Our results clearly show that the AFLP data yielded better resolution, as judged by the geographic or racial cohesiveness of clades. In both *H. melpomene* and *H. erato*, it is generally true that individuals within AFLP clades were also found to be strongly clustered in the mtDNA tree, more so in *H. melpomene* than in *H. erato* (Figs. 1 and 2 in the main text). This suggests that, at this level, the differences between the AFLP and mtDNA trees may be in part caused by differences in the resolving powers of the two data types (i.e., varying coalescence depths, incomplete lineage sorting). At higher level relationships in *H. melpomene*, however, the mtDNA and AFLP trees are discordant, suggesting that incomplete lineage sorting is unlikely to account for the differences. For example, Brazilian samples form the earliest branch in the AFLP tree but not in the mtDNA tree, and the *H. cydno* lineage is closer to isthmus specimens in the AFLP tree but closer to Peru and Ecuador specimens in the mtDNA tree. In *H. erato*, it is not clear whether the deepest level relationships are discordant between the mtDNA and AFLP trees. One of the major clades in the mtDNA tree, E1, is found in the AFLP tree, and because relationships among AFLP clades are poorly supported, we cannot rule out the possibility that the individuals in mtDNA clade E2 actually form an AFLP clade.

Hybridization may be a large part of the reason for discordance between the AFLP and mtDNA topologies. Hybridization and introgression are widespread across *Heliconius* butterflies (12–15) and although known racial hybrid zones were avoided in our selection of sampling locations, mtDNA leakage can still occur. Indeed, results from this study and others (6) indicate continued gene flow in the midst of racial diversification. Selective sweeps acting on mitochondrial genes may also play a role in creating disagreement between the mtDNA and AFLP data.

**New Insights from the Matriline.** The mtDNA phylogenies of *H. erato* and *H. melpomene* reported here present a somewhat modified view from that of Brower (4, 5). Our mtDNA phylogeny for *H. erato* clearly shows three well supported clades: E1, E2, and *H. himera* (respectively, the Western clade and Eastern clade of *H. erato* described by Brower (4), and *H. himera*). The relationship among these three clades, however, is not well supported by bootstrap analysis, which places *H. himera*, rather than E1, as the sister to E2. Brower (4) noted that the Western and Eastern mtDNA clades of *H. erato* represent a "basal split between east and west of the Andes, reflecting a vicariant separation at the base of the Pleistocene." Our results indicate a more complex scenario than a trans-Andean vicariant event. Our mtDNA clade E1 in Fig. 1*B* in the main text (west of the Andes) contains specimens from the eastern cordillera of the Colombian Andes, and from Trinidad, approximately 1,000 km due east of the eastern limits of Brower's Western clade (4). Based on parsimonious ancestral state reconstruction in the clade E1, these eastern vagrants could be matrilineal descendants of ancestors from west of the Andes (see Fig. 1*B* in the main text, in which samples that are truly west of the Andes are highlighted in gray boxes). Trinidad may have been colonized via the coastal lowlands in the north of the continent at a time when sea levels were lower, such as during one of the glacial phases of the Pleistocene.

The mtDNA study of *H. melpomene* of Brower (5) noted five clades: Brazil, Guiana, *cydno,* and two clades, a Western clade and Eastern clade, separated by the Andes. These correspond, respectively, to M2a, M2b, M1b, M2d, and M1a, in Fig. 2*B* in the main text. The present study has revealed an additional clade, M2c (comprising French Guiana and Trinidad samples), that was not present in the study of Brower (5), and that the western clade of Brower (4) (M2d in Fig. 2*B* in the main text) now contains specimens from the eastern cordillera of the Colombian Andes (thus, like the western clade in *H. erato*, it is not truly west of the Andes). Additionally, the branching order among our mtDNA clades differs from the strict consensus parsimony tree reported by Brower (5), in which the Guiana clade is sister to a clade containing all other clades in an unresolved polytomy. Here, the mtDNA maximum likelihood tree shows a clear basal split in the ingroup, between M1 (*cydno* clade plus Eastern clade) and M2 (the remainder, i.e., Brazil, Guiana, Western and French Guiana plus Trinidad clades). Additionally, our M1b [i.e., *H. cydno* clade of Brower (4)] also includes an individual of *H. m. melpomene* from Colombia.

**How Did *H. erato* Diversify?** Although this research is primarily focused on examining the potential for coevolution between *H. erato* and *H. melpomene*, the finding of independent radiations in the two species raises an important follow-up question:

how did *H. erato* diversify in the first place? This is a challenging question because stabilizing selection on warning patterns is expected to prevent the evolution of new color pattern phenotypes. For this reason, the issue of diversity in warning coloration and Müllerian mimicry has received substantial attention (16–18). A very brief summary of a few potential explanations for *H. erato* diversification are as follows.

***Wright's shifting balance.*** Mallet and coworkers (16, 17, 19–21) have proposed that color pattern diversification in *Heliconius* may be a good example of evolution via Wright's shifting balance hypothesis. Although the shifting balance has received substantial criticism as a general mechanism for evolutionary change (22, 23), there are multiple features of the *Heliconius* system which lend support to this argument, namely, fine-scale population genetic subdivision within species that could facilitate the local drift necessary for the first phase of the process (18), and evidence for moving color pattern hybrid zones that may represent the last stage (24). Recently, Mallet (19) suggested that the geographic distribution of color patterns in *H. erato* provides evidence for the shifting balance. In particular, the patchy distribution of the postman pattern around the periphery of a contiguous block of the rayed pattern in the Amazon could be the result of a new, rayed phenotype emerging and spreading out in the middle of the ancestral, postman phenotype. Our AFLP data allow us to test this prediction because, if true, we might expect postman populations to form an early-branching clade with rayed populations also forming a clade which is in a more derived position. Our AFLP tree for *H. erato* actually reveals the opposite results—phenotypes do not form monophyletic clades and rayed populations are basal. However, this is not a strong test of the shifting balance hypothesis because (*i*) support for relationships in the *H. erato* AFLP tree is not very strong, and (*ii*) this centrifugal radiation hypothesis is not necessarily an integral part of the shifting balance argument for *Heliconius* diversification.

***Natural selection.*** It is possible that divergent natural selection played a role in generating color pattern diversity in *H. erato*. Two possible mechanisms could be (*i*) selection to mimic different species in different locations or (*ii*) selection on the phenotype to match local environments, e.g., to maximize signaling efficiency in different light environments. However, as Mallet points out (19), the available data suggest that these specific explanations are unlikely.

***Sexual selection.*** There is good evidence for color pattern based mate preference in *Heliconius* (25–27). It is possible that diversification in the earliest radiations could have been caused, at least in part, by divergent sexual selection.


1. Holland BR, Clarke AC, Meudt HM (2008) Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. *Syst Biol* 57:347–366.
2. Miller MJ, et al. (2008) Out of Amazonia again and again: Episodic crossing of the Andes promotes diversification in a lowland forest flycatcher. *Proc R Soc Lond B Biol Sci* 275:1133–1142.
3. Gregory-Wodzicki KM (2000) Uplift history of the central and northern Andes: A review. *Geol Soc Am Bull* 112:1091–1105.
4. 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.
5. 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.
6. Flanagan NS, et al. (2004) Historical demography of Mullerian mimicry in the neotropical *Heliconius* butterflies. *Proc Natl Acad Sci USA* 101:9704–9709.
7. Mallet J (1989) The genetics of warning color in Peruvian hybrid zones of *Heliconius erato* and *H. melpomene*. *Proc R Soc Lond B Biol Sci* 236:163–185.
8. Mallet J, Beltran M, Neukirchen W, Linares M (2007) Natural hybridization in heliconiine butterflies: The species boundary as a continuum. *BMC Evol Biol* 7:28.
9. Gilbert LE (2003) Adaptive novelty through introgression in Heliconius wing patterns: Evidence for shared genetic "tool box" from synthetic hybrid zones and a theory of diversification. *Ecology and Evolution Taking Flight: Butterflies as Model Systems*, eds Boggs CL, Watt WB, Ehrlich PR (Univ of Chicago Press, Chicago), pp 281–318.
10. Jiggins CD, Salazar C, Linares M, Mavarez J (2008) Hybrid trait speciation and *Heliconius* butterflies. *Philos Trans R Soc Lond B* 363:3047–3054.
11. Mavarez J, et al. (2006) Speciation by hybridization in Heliconius butterflies. *Nature* 441:868–871.
12. Beltran M, et al. (2002) Phylogenetic discordance at the species boundary: Comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Mol Biol Evol* 19:2176–2190.
13. Bull V, et al. (2006) Polyphyly and gene flow between non-sibling *Heliconius* species. *BMC Biol* 4:11.
14. Kronforst MR (2008) Gene flow persists millions of years after speciation in *Heliconius* butterflies. *BMC Evol Biol* 8:98.
15. Kronforst MR, Young LG, Blume LM, Gilbert LE (2006) Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies. *Evolution* 60: 1254–1268.
16. Joron M, Mallet JLB (1998) Diversity in mimicry: Paradox or paradigm? *Trends Ecol Evol* 13:461–466.
17. Mallet J, Joron M (1999) Evolution of diversity in warning color and mimicry: Polymorphisms, shifting balance, and speciation. *Annu Rev Ecol Syst* 30:201–233.
18. Kronforst MR, Gilbert LE (2008) The population genetics of mimetic diversity in *Heliconius* butterflies. *Proc R Soc Lond B Biol Sci* 275:493–500.
19. Mallet J (2010) Shift happens! Shifting balance and the evolution of diversity in warning colour and mimicry. *Ecol Entomol* 35:90–104.
20. Mallet J, Singer MC (1987) Individual selection, kin selection, and the shifting balance in the evolution of warning colors: The evidence from butterflies. *Biol J Linn Soc* 32: 337–350.
21. Turner JRG, Mallet JLB (1996) Did forest islands drive the diversity of warningly coloured butterflies? Biotic drift and the shifting balance. *Philos Trans R Soc Lond B Biol Sci* 351:835–845.
22. Coyne JA, Barton NH, Turelli M (1997) Perspective: A critique of Sewall Wright's shifting balance theory of evolution. *Evolution* 51:643–671.
23. Coyne JA, Barton NH, Turelli M (2000) Is Wright's shifting balance process important in evolution? *Evolution* 54:306–317.
24. Blum MJ (2002) Rapid movement of a Heliconius hybrid zone: Evidence for phase III of Wright's shifting balance theory? *Evolution* 56:1992–1998.
25. Jiggins CD, Estrada C, Rodrigues A (2004) Mimicry and the evolution of premating isolation in Heliconius melpomene Linnaeus. *J Evol Biol* 17:680–691.
26. Jiggins CD, Naisbit RE, Coe RL, Mallet J (2001) Reproductive isolation caused by colour pattern mimicry. *Nature* 411:302–305.
27. Kronforst MR, et al. (2006) Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proc Natl Acad Sci USA* 103: 6575–6580.

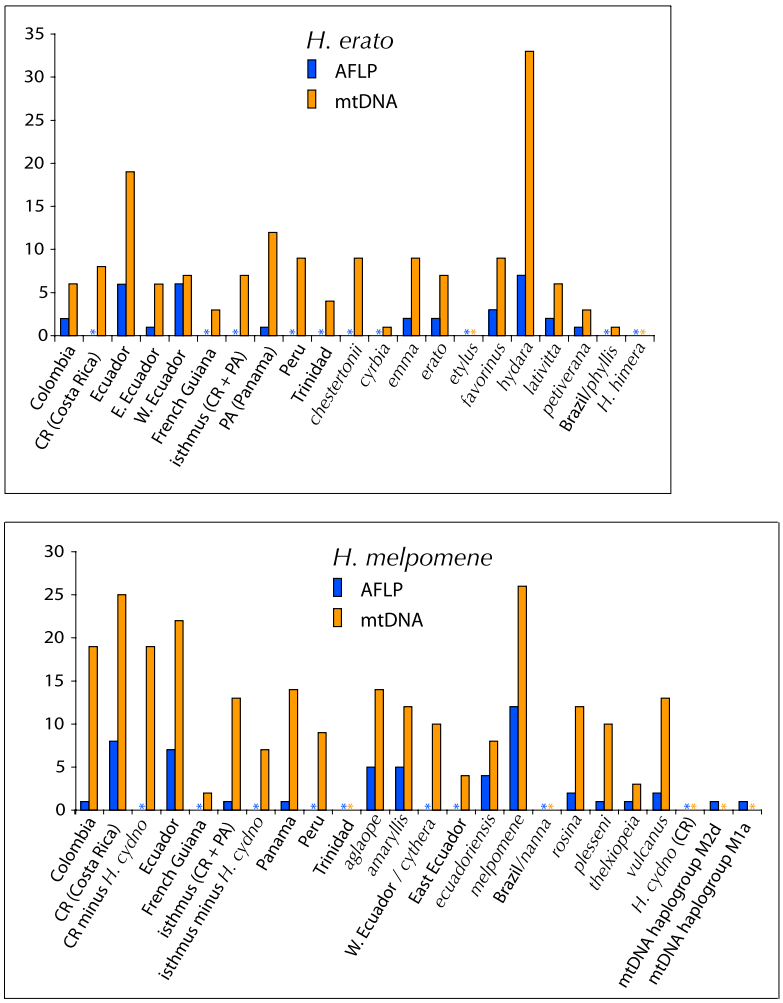


**Fig. S1.** Clade Construction Indexes (CCIs) for groups within *H. erato* and *H. melpomene* calculated from the AFLP neighbor joining and mtDNA maximum likelihood trees shown in Figs. 1 and 2 in the main text. The CCI is the number of lineages that must be removed to achieve monophyly. Asterisks denote CCI values of 0 (i.e., monophyly).

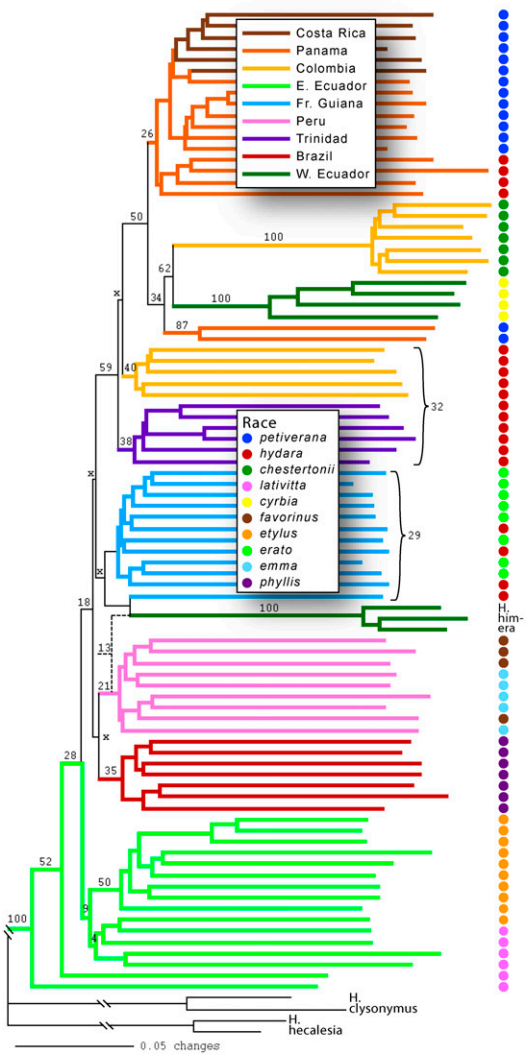


**Fig. S2.** Phylogeny of the races of *H. erato* based on neighbor joining analysis of AFLP data scored with outgroups. Bootstrap support values (%) are shown at nodes or next to curly brackets. An "x" denotes a clade not present in the bootstrap consensus topology. Dotted lines indicate clades recovered only in the bootstrap consensus and not in the original topology.

## Other Supporting Information Files

Table S1 (DOC)
Table S2 (DOC)