Phylogenetic Codivergence Supports Coevolution of Mimetic *Heliconius* Butterflies

Jennifer Hoyal Cuthill*, Michael Charleston

School of Information Technologies, University of Sydney, Sydney, NSW, Australia

*E-mail: jenhc@it.usyd.edu.au

Abstract

The unpalatable and warning-patterned butterflies *Heliconius erato* and *Heliconius melpomene* provide the best studied example of mutualistic Müllerian mimicry, thought – but rarely demonstrated – to promote coevolution. Some of the strongest available evidence for coevolution comes from phylogenetic codivergence, the parallel divergence of ecologically associated lineages. Early evolutionary reconstructions suggested codivergence between mimetic populations of *H. erato* and *H. melpomene*, and this was initially hailed as the most striking known case of coevolution. However, subsequent molecular phylogenetic analyses found discrepancies in phylogenetic branching patterns and timing (topological and temporal incongruence) that argued against codivergence. We present the first explicit cophylogenetic test of codivergence between mimetic populations of *H. erato* and *H. melpomene*, and re-examine the timing of these radiations. We find statistically significant topological congruence between multilocus coalescent population phylogenies of *H. erato* and *H. melpomene*, supporting repeated codivergence of mimetic populations. Divergence time estimates, based on a Bayesian coalescent model, suggest that the evolutionary radiations of *H. erato* and *H. melpomene* occurred over the same time period, and are compatible with a series of temporally congruent codivergence events. This evidence supports a history of reciprocal coevolution between Müllerian co-mimics characterised by phylogenetic codivergence and parallel phenotypic change.

Introduction

The Neotropical butterfly genus *Heliconius* is highly diverse, with 54 species [1], many of which can be subdivided into multiple wing pattern morphs, or races [2]. These unpalatable [3,4] butterflies have diversified to form regional Müllerian [5] mimicry complexes [6], each involving multiple species with a convergently evolved [7] predator warning pattern [4,8]. For almost 150 years, biologists have debated whether the remarkable adaptive radiation of the *Heliconius* was driven by reciprocal ecological associations, a process we would now call coevolution [9]. Unpalatable Müllerian [5] co-mimics share the cost of educating inexperienced predators [4] (unlike palatable Batesian [6] mimics, which may "parasitize" their unpalatable models [10]). According to Müller's original model [5], two unpalatable co-mimics will both gain in fitness by their resemblance, though the ratio of these fitness gains will be proportionate to the ratio of their population sizes; giving greater fitness benefits to a rarer population (since a more abundant co-mimic is predicted to lose a greater number of individuals to encounters with inexperienced predators) [10]. It has been suggested that mimicry (and, particularly, mutualistic Müllerian mimicry [10]) may provide some of the most favourable conditions for coevolution, which has been defined (in the strict sense) as reciprocal evolutionary change [11] under mutualistic or competitive selection [12]. Therefore, mimetic wing pattern evolution among *Heliconius* butterflies may provide key evidence regarding the importance of coevolution in adaptive radiation [9,13].

The parallel wing pattern radiations of *H. erato* and *H. melpomene* have been the primary case study in the debate over coevolution between Müllerian co-mimics [2,13-15]. Across the Americas, each species is divided into approximately 30 morphs [2]. With few exceptions, the wing patterns of *H. melpomene* and *H. erato* match in every region where

they co-occur [2] (Figure 1). The two species are reciprocally monophyletic [16] and do not hybridise [17]. Therefore mimicry between them has involved convergence at the genetic and phenotypic level [7,18]. Purifying selection against intra-species hybrids with unusual wing patterns [4] acts with assortative mating [19] to generate partial reproductive isolation between the parapatric morphs [14], potentially showing speciation in action [2,6,20].

Codivergence is the parallel divergence of ecologically associated lineages within two phylogenies [21], and is one predicted outcome of coevolution [10,22]. Codivergence may not, in itself, prove coevolution in the strict sense [13]. However, codivergence can be considered some of the strongest available evidence for coevolution [13,22], since, as Page [21] puts it, "it is difficult to imagine that [codivergence] can occur without at least some degree of coevolution". Topological and temporal congruence (similarity of branching pattern and timing, respectively) between the phylogenies of *H. erato* and *H. melpomene*, compatible with a history of codivergence, would therefore support their coevolution [2,13,14,23] (in reference to the Pleistocene refugium hypothesis, of Brown et al.1974, Sheppard et al. [2] suggested that coevolution between *H. erato* and *H. melpomene* may have been aided by population isolation, but see [24] for a critical review). In contrast, a lack of topological or temporal congruence would suggest that coevolution did not occur (as previously suggested [14,25,26]).

Despite considerable discussion of phylogenetic branching patterns [2,14,26] and timing [14,25,26], and an early biogeographic character-based analysis [14], there has been no previous test of codivergence between the mimetic populations of *H. erato* and *H. melpomene* using methods from cophylogenetic analysis (reviewed in [21]), which were developed specifically for this purpose. Cophylogenetic analysis seeks to reconstruct histories between associated entities that can be represented by a pair of phylogenies (such as genes and species, parasites and their hosts, populations and biogeographic regions [21], or mimics

and models [27]) to determine whether there is statistically significant evidence for codivergence between these associated phylogenies (as described below). We present the first explicit cophylogenetic test for codivergence between mimetic populations of *H. erato* and *H. melpomene*, using multilocus coalescent [28,29] phylogenies, and re-examine divergence times, based on the Bayesian multilocus coalescence model [29] and a recently published fossil-calibrated butterfly molecular clock [30].

One of the greatest challenges for phylogenetic reconstruction of recent radiations with low sequence divergence, such as that of the *Heliconius,* is the incomplete sorting of ancestral polymorphisms among divergent populations [31]. Incomplete lineage sorting can cause individual gene trees to conflict with each other and with the true population tree. A related problem is that individuals sampled from divergent lineages may not form reciprocally monophyletic clades on an individual gene tree, or on a population tree built using gene sequence concatenation or gene tree consensus methods [32]. Coalescent methods are designed to take individual gene histories into account by modelling the processes of mutation and inheritance, specifically, the coalescence of sampled genes, back through a gene tree, to their most recent common ancestor (reviewed by Rosenberg and Nordborg [31]). Coalescent phylogenetic methods reconstruct the relationships between divergent populations, that are partially to completely genetically isolated [33,34], by optimally reconciling the histories of multiple gene loci within one population-level tree [28,29]. Coalescent methods have rarely been applied to heliconian population genetics [25], and have not previously been used to reconstruct the phylogenies of *H. erato* and *H melpomene*. Here we use coalescent [29] and character support [35] methods to delimit monophyletic populations of *H. erato* and *H. melpomene,* among individuals sampled at the level of country, biogeographic region, and morph (Table S1). Phylogenetic relationships [28,29] and divergence times [29] for these populations are reconstructed using Bayesian [29] and

parsimony based (Minimise Deep Coalescence, MDC [28]) coalescent methods. These phylogenies provide the basis for cophylogenetic tests of topological congruence, conducted across the set of phylogenetic estimates returned by the coalescent analyses. Estimated branching patterns, cophylogenetic histories, and divergence times are presented and discussed, based primarily on phylogenies of the country level mimetic populations, which received the highest support for monophyly from the character based [35] and Bayesian coalescent [29] analyses.

Materials and methods

Study system

The taxon set (Table S1) included eight pairs of mimetic wing pattern morphs from *H. erato* and *H. melpomene* (as well as the morphs *H. erato chestertonii* and *H. melpomene plesseni*, whose co-mimics were not included in this study), sampling populations from the major South American biogeographic regions (East and West of the Andes) and seven countries [25,26] (Figure 1). Twenty-two related species [16,25,26,36] were also included, to place and date the population radiations of *H. erato* and *H. melpomene* within the wider radiation of the *Heliconius*. These species, and morphs of *H. erato* and *H. melpomene*, were selected for their coverage of the four included gene loci.

Molecular data and analyses

Phylogenetic reconstructions were based on a DNA sequence dataset sampled from four unlinked gene loci (mitochondrial cytochrome c oxidase, COI and COII, and nuclear mannose 6-phosphate isomerase, Mpi, and triose-phosphate isomerase, Tpi) spanning 3533 base pairs. DNA sequences were downloaded from GenBank (accession numbers and source

studies Table S1). Sequences of each gene locus were aligned using MUSCLE [37]. Alignments for individual gene loci were used in Maximum likelihood (ML) phylogenetic analyses, implemented in TREEFINDER [38], to produce individual gene trees as input for the MDC [28] coalescent phylogenetic analyses (described below). The ML phylogenetic analyses were conducted using the best-fit substitution model for each gene, selected by jModelTest 0.1.1 [39] under the Akaike Information Criterion (AIC) [40]. Sequences for the four included gene loci were also concatenated to produce a multilocus alignment (Table S3).

Population genetic statistics were estimated based on the multilocus alignment, using SITES [41]. These statistics included the average pairwise sequence divergence within a population and an effective population size parameter, θ, estimated as the product of effective population size and mutation rate ($\theta = 4 N_e\mu$ where N_e = effective population size and μ is the neutral mutation rate [42,43]). Two tests of population monophyly were performed, on the multilocus alignment, for specimens of *H. erato* and *H. melpomene* grouped at the level of country, biogeographic region, morph or species. First, monophyly was assessed with Shimodaira-Hasegawa (SH) tests [35] on monophyly-constrained ML trees, using TREEFINDER. This test compared the support for each population level using the AIC, which evaluates the fit of a statistical model to the data against the number of parameters imposed by that model – in this case, the number of constraints required for monophyly at the given population level. Since the AIC is a measure of information loss, the preferred phylogenetic hypothesis will be the one with the lowest AIC value. Incongruence length difference (ILD) tests, conducted using PAUP* 4.0b10 [44], indicated significant incongruence between the nuclear loci for both *H. erato* and *H. melpomene* ($P = 0.01$ each case), so the nucleotide substitution model was partitioned by gene locus (COI, GTR+I+G; COII, HKY+I+G; Mpi, GTR+G; Tpi, HKY+G). The second test of population monophyly

was based on Bayesian multilocus coalescent phylogenies reconstructed for each geographical sampling level (and is described below).

Coalescent population phylogenies were reconstructed by minimising the number of deep gene coalescences [28] in Mesquite [45] and using a Bayesian multi-population coalescent model in *BEAST [29], part of the BEAST 1.6.1 package [46]. Such methods, which are based on a explicit model of gene lineage coalescence, have been found to accurately reconstruct population level phylogenies and are robust to low levels of gene flow [33,34]. MDC [28,45] phylogenies were each reconstructed using a heuristic, population level tree search, which incorporated the branch lengths of the four gene trees, did not auto-resolve gene tree polytomies, used subtree pruning and regraft (SPR) branch-swapping, and stored up to 100 equally good trees at each search step. Bayesian coalescent analyses were based on partitioned nucleotide substitution models, selected under the Bayesian Information Criterion (COI, HKY+I+G; COII, HKY+I+G; Mpi, HKY+G; Tpi, HKY+G). The focal *BEAST analyses were run using a relaxed log-normal molecular clock (selected based on Bayes factor comparisons against an, otherwise identical, analysis run with a strict molecular clock), allowing the mutation rate to vary between branches of the phylogeny [46]. A Yule prior was specified for the branching process of the population tree. Since the two mitochondrial gene loci (COI and COII) are non-recombining, a linked tree was specified for these loci in the *BEAST analyses [46]. Each *BEAST analysis was run with a Markov chain Monte Carlo (MCMC) chain length of 10^8 steps, parameter sampling every 10^4 steps, and a conservative burn-in of 25%. Effective sample size (ESS) values, for the posterior distribution of each parameter, were assessed to check chain convergence in each *BEAST run. Output from *BEAST was analysed using the programs Tracer [46] and FigTree 1.3.1 (A. Rambaut, http://tree.bio.ed.ac.uk/software/figtree/). The MDC and Bayesian coalescent population phylogenies were used as input for the cophylogenetic analyses (described below).

The Bayesian phylogenies also provided the basis for the second test of population monophyly. In this test, hypotheses of population divergence (at the level of country or morph versus species) were tested by comparing the coalescent likelihood [29] and population tree posterior [46] calculated under the Bayesian multi-population coalescent model [29]. The coalescent likelihood calculates the probability of each gene tree *g* given the population tree *S*, as $P(g|S) = \prod_{b \in S} P(L_b(g)|N_b(t))$, where *b* denotes the branches of the species tree *S*, $L_b(g)$ is the implied history of *g* over *b*, and $N_b(t)$ is the function for effective population size through time [29]. The population tree posterior is the sum of that tree's log likelihood and log prior probability, plus the log prior probability densities for any other included priors [46]. For each of these parameters, Tracer was used to calculate the mean value across the MCMC samples (excluding the burn-in) as well as the 95% Bayesian credibility interval (BCI), which is the shortest interval containing 95% of the sampled values. This preferred phylogenetic hypothesis, in this test, is the one with the highest coalescent likelihood and tree posterior. Comparing these parameter values between population trees allowed us to evaluate independent phylogenetic estimates for each geographical sampling level (and so did not require nested hypotheses of population monophyly, as does the Bayesian coalescent method for population delimitation of Yang and Rannala [47] for example). The Bayesian coalescent analysis was based on a reduced dataset consisting of those country level populations which were sampled at all four gene loci, according to the requirements of *BEAST. Biogeographic region was not included as a population level in the Bayesian coalescent analyses, due to the unavailability of gene sequences with sufficient coverage of the four included loci.

Population divergence times were estimated under the Bayesian coalescent model [29], which estimates and incorporates phylogenetic branching patterns and effective population sizes, using a relaxed log-normal molecular clock with a fossil-calibrated rate of

0.01909 substitutions per site per million years [30]. Such methods, which explicitly model gene lineage coalescence, are expected to give relatively accurate estimates of divergence times compared, for example, to estimates from individual or concatenated gene loci [48]. The time-calibrated substitution rate was set for one reference locus (COI) and specified as a prior for the 3 remaining loci (COII, Mpi, Tpi), after Heled and Drummond [29]. This analysis produced tree topologies identical to those of the *BEAST analyses (described above), in which no time-calibrated substitution rate was specified (there, the reference locus rate and priors for the other loci were set to 1, giving branch lengths in units of substitutions per site).

Cophylogenetic analyses

The MDC and Bayesian coalescent populations phylogenies of *H. erato* and *H. melpomene* were used in cophylogenetic analyses, conducted using TreeMap 3 [49] and Jane 3 [50]. These analyses tested for statistically significant topological congruence between the two phylogenies, compatible with a history of codivergence between the mimetic populations. Cophylogeny mapping reconstructs histories that explain the similarities and differences between associated phylogenies given a cost regime for the recoverable historical events [50,51]. This is achieved by mapping current ecological associations (e.g. between mimics and models) back into the internal nodes of one phylogeny (e.g. that of the model) to reconstruct a cophylogenetic history (e.g. the history of mimicry between two species). In our context, the recoverable historical events are *codivergence* (parallel divergence of mimetic lineages), *duplication* (divergence of mimic lineages without model divergence), *model switch* (divergence of a mimic lineage onto an additional model lineage), and *loss* (absence of a mimic on a model lineage where it would otherwise be expected).

Cophylogeny mapping in Jane uses heuristics to find solutions that minimise the overall cost of a historical reconstruction given a cost regime. The default event costs are zero (0) for a codivergence event, one (1) for duplication and model switch events, and two (2) for loss events. TreeMap 3 attempts to find a Pareto set of solutions, that is, all the histories that could be optimal, given the input phylogenies and associations, under a range of event cost regimes. This range is very permissive: codivergence is set at 0 cost and all the other costs are assumed to be positive, but do not need to be specified. Statistical analysis can then be performed (in both programs) to test whether the cost of a historical reconstruction is significantly lower than expected by chance, by generating a pseudo-random sample of minimal costs from a null distribution of problem instances with the same model phylogeny. The null distribution is generated by randomising repeatedly either the leaf associations or the branching order of the associate (mimic) tree. Thus there are two null hypotheses we might reject: either (a) that the current associations between model and mimic are not a consequence of a history of coevolution with the model phylogeny, and (b) the branching order of the mimic tree is not dependent on the branching order of the model tree. We prefer the latter test as it accounts for differences in probability of different tree shapes, but we conducted both tests for completeness and comparability with other studies.

Müllerian co-mimics may benefit from a shared warning pattern to different degrees [2,15]. *H. erato* has several characteristics, independent of hypothetical divergence times, which suggest that it has had the dominant role in its mimicry relationship with *H. melpomene* (Eltringham, 1916 cited in [15]). These include greater current [15], and possibly historical [25], abundance, greater gregariousness, a wider geographic distribution, and pupal mating [15]. Therefore, we treated *H. erato* as the model and *H. melpomene* as the mimic in our main cophylogeny mapping analyses, conducted using TreeMap and Jane. For comparison, these analyses were also repeated with a reversed model-mimic relationship.

An additional pair-wise distance correlation test of topological congruence was performed in TreeMap 3. This test compares the significance of correlations of pair-wise distances between leaves, for associated clades in the two phylogenies, against a distribution of such measures estimated by randomising subtrees of the mimic phylogeny.

To ensure that the results of the cophylogenetic analyses were not exclusive to our coalescent population phylogenies, a cophylogenetic pair-wise distance correlation test (as described above) was performed on recent, genome-wide, amplified fragment length polymorphism (AFLP) phylogenies of *H. erato* and *H. melpomene* [26] and on phylogenies reconstructed from a recently published ten-gene dataset [18] which included five linked genes involved in heliconian wing colour pattern determination. To produce the input for these cophylogenetic analyses, the topologies of the published AFLP phylogenies [26] were replicated and the ten-gene dataset [18] was downloaded from GenBank (accession numbers from Table S1 in Hines et al. [18]) and re-analysed. For the ten-gene dataset, MUSCLE was used to produce separate gene locus alignments for *H. erato* (plus its relatives *H. himera* and *H. clysonymus*) and *H. melpomene* (plus its relatives *H. cydno*, *H. ismenius* and *H. numata*). Alignments for each gene locus were then concatenated to produce two multilocus alignments: one including all ten genes and the other containing only the five colour pattern genes. A maximum likelihood phylogeny was then reconstructed for each multilocus alignment in TREEFINDER, using a partitioned nucleotide substitution model with the best-fit substitution model for each gene selected by jModelTest under the AIC. For each of these phylogenies, monophyletic clades of each wing pattern morph were then collapsed to a single leaf, to avoid pseudo-replication of mimicry associations.

Results

Population phylogenetics

Shimodaira-Hasegawa (SH) character support tests [35] on monophyly-constrained maximum likelihood (ML) trees could not reject population monophyly at the level of country, biogeographic region, or morph for *H. erato* or *H. melpomene* ($P > 0.4$ in all cases). Taking into account the number of parameters imposed by each monophyly constraint, using the Akaike Information Criterion [40] (AIC), *country* was the favoured monophyly level for the sampled within-species populations (AIC scores: species = 44033, countries = 47500, regions = 47780, morphs = 47502). Coalescent likelihood mean (clm) and population tree posterior (ptp) values under a Bayesian multi-population coalescent model [29] also favoured population divergence at the level of *country* (corresponding phylogeny Figure S1) over divergence at the level of *species* or *morph* (corresponding phylogeny Figure S2): country level clm = 1774 [1685 to 1862]; ptp = -17478 [-17598 to -17359]; species level clm = 1220 [1130 to 1310], ptp = -18287 [-18399 to -18179]; morph level clm = 1701 [1613 to 1788], ptp = -17617 [-17732 to -17503].

Monophyly of sampled morph populations (at least at the level of country) is supported by the gene sequence data and provides the most probable coalescent history for the sampled gene loci. This concurs with the greater clustering of individuals into monophyletic country level populations observed on recent genome-wide AFLP phylogenies, relative to phylogenetic estimates based on three concatenated mitochondrial loci [26]. These results (see also [14,26]) suggest that the wing pattern morphs sampled from multiple countries (here *H. erato hydara*, *H. erato petiverana*, *H. melpomene melpomene* and *H. melpomene rosina*) may be non-monophyletic. However, neutral markers for recently diverged populations that can experience ongoing low-level gene flow (including those used in this study) may show relatively low levels of phylogenetic structure [18] and we note that the character support analyses were unable to reject monophyly of the higher population

levels of biogeographic region or morph. Based on the character support and Bayesian coalescent analyses, we therefore focussed our cophylogenetic analyses on the country level populations, which received the highest monophyly support. However, phylogenies for region and morph level populations, which received lesser support, were also analysed.

Topological congruence between the population phylogenies of *H. erato* and *H. melpomene*

To account for phylogenetic uncertainty, we conducted cophylogenetic analyses across the set of phylogenies returned by the coalescent analyses (listed in Table 1). Across these MDC and Bayesian coalescent population phylogenies, the overwhelming indication is of significant topological congruence (Table 1). The cophylogenetic analyses suggest that, in almost all cases, there are more codivergence events between the mimetic populations of *H. erato* and *H. melpomene* than would be expected by chance if their phylogenies were independent (e.g. Figure 3). Figure 3 shows an example pair of phylogenies from this set of phylogenetic estimates with similarly high congruence (see Table 1). Reconciling the phylogeny of *H. melpomene* with that of *H. erato* indicates remarkable topological congruence, with eight codivergence events out of out of a possible eleven, two duplications followed by model switches, and one loss. Interestingly, we still obtain highly significant congruence between the phylogenies when the mimic-model relationship is reversed (Table S2).

The pair-wise distance correlation test also showed significant congruence at the roots of the *H. erato* and *H. melpomene* phylogenies, for the majority of our phylogenetic estimates (Table 1), suggesting that the phylogenies have been highly dependent on each other throughout their history.

Pair-wise distance correlation tests conducted on the recent AFLP phylogenies of Quek et al. [26] (Figure S3) and on phylogenies reconstructed from a recently published ten-

gene dataset which includes five colour pattern genes [18] (Figures S4 and S5) also suggest significant topological congruence between the radiations of *H. erato* and *H. melpomene* (Table 1), contrary to the conclusions of these authors.

Temporal congruence between the population phylogenies of *H. erato* and *H. melpomene*

We estimated average uncorrected sequence divergence (average pairwise substitutions per site, excluding gaps and indels) at 0.0245 within *H. erato* and 0.0153 within *H. melpomene* (excluding other, putatively incipient, species). Thus, the average uncorrected sequence divergence among individuals sampled from *H. melpomene* is considerably lower than (62% of) that estimated for *H. erato*, as previously suggested [14]. However, the effective population size parameter for *H. melpomene*, measured as the product of effective population size and mutation rate ($\theta = 4 N_e\mu$), was estimated at 41% to 63% of that for *H. erato*, see also [25]: Watterson's [42] estimate of θ was 0.1058 for *H. erato* and 0.0429 for *H. melpomene*, the pairwise nucleotide diversity estimate of θ was 0.0245 for *H. erato* and 0.0153 for *H. melpomene*, with each estimate of θ calculated as an average per base pair, with gaps, indels and sequences for putatively incipient species excluded [41]. Effective population size is known to be positively correlated with average genetic diversity. Therefore, the lower average pairwise genetic diversity of *H. melpomene*, relative to *H. erato*, is an expected consequence of a lower effective populations size (e.g. see [43]), and is compatible with similar origination dates for the sampled clades of *H. melpomene* and *H. erato*, as discussed below.

In results similar to those of Flanagan et al. [25], average uncorrected sequence divergence between *H. erato* and the closely related species *H. hecalesia* (0.0605) was greater than that between *H. melpomene* and its close relative, *H. cydno* (0.0303), or that between *H. melpomene* and the silvaniforms (0.0512), which form the outgroup to *H.*

melpomene plus the *H. cydno* group [16]. Divergence times estimated on our Bayesian coalescent phylogenies of the *Heliconius* (Figures S1 and S2) using a fossil-calibrated butterfly molecular clock [30], also suggest that the split between *H. hecalesia* and *H. erato* is older than that between *H. cydno* and its *H. melpomene* sister clade, as previously suggested [25] (as well as that between *H. melpomene* – plus the *H. cydno* group – and the silvaniforms). However, like the population genetic results discussed above, the Bayesian divergence estimates suggest temporal congruence between the radiations of *H. erato* and *H. melpomene* (Figure 3), contrary to [25]. Of the codivergence events reconstructed for the MDC phylogenies of Figure 3, for example, 95% Bayesian Credibility Intervals overlap where they are available. A historical reconstruction incorporating the estimated divergence times on the Bayesian country level phylogeny finds five codivergence events out of a maximum of eight (Figure S6).

Discussion

Evidence for codivergence

Our coalescent population phylogenies for *H. erato* and *H. melpomene* (e.g. Figure 2; Figures S1 and S2) have many features in common with previous phylogenetic estimates [14] [26], including a strong signal from biogeographic region (East or West of the Andes). The MDC coalescent phylogenies represented in Figure 2, for example, shares major topological features with recent, genome-wide, AFLP phylogenies of *H. erato* and *H. melpomene* [26] (Figure S3). These features include a relatively basal split, within each species, between two clades; one clade containing eastern and western populations of, mimetic, *H. erato hydara / H. melpomene melpomene* plus the populations of the other western morphs (mimetic, *H. erato petiverana / H. melpomene rosina* and *H. erato cyrbia / H. melpomene cythera*), the

other clade containing the remaining eastern populations of *H. e. hydara / H. m. melpomene* plus populations of the other eastern mimetic morphs. Within the solely eastern clades, a basal split between two major sub-clades is also shared with the recent AFLP phylogenies; one sub-clade containing *H. e. hydara* / *H. m. melpomene* and the French Guianan population of, mimetic, *H. erato erato* / *H. melpomene thelxiopeia*, the other sub-clade containing populations of, mimetic: *H. erato lativitta* / *H. melpomene malleti*, *H. erato emma* / *H. melpomene aglaope*, *H. erato etylus* / *H. melpomene ecuadoriensis* and *H. erato favorinus* / *H. melpomene amaryllis*.

Cophylogenetic analyses, conducted across the set of coalescent phylogenetic reconstructions, give an overall picture of statistically significant topological congruence between the evolutionary radiations of *H. erato* and *H. melpomene* co-mimics (contrary to previous suggestions [14,26]) (Table 1). In particular, all phylogenetic estimates for the, best supported, country level populations are compatible with a history of repeated codivergence between mimetic populations.

In the interpretation of their AFLP phylogenies, Quek et al. [26] emphasised elements of incongruence between the topologies for *H. erato* and *H. melpomene*. They [26] noted, specifically, that the earliest branching lineages within each species did not represent co-mimetic morphs (these were *H. erato etylus* sampled from East Ecuador, which instead falls within the eastern clade of our Figure 2, and *H. melpomene nanna* sampled from Brazil, which was not included in our coalescent analyses).

However, a cophylogenetic analysis conducted on these recent AFLP phylogenies [26] also indicates that patterns of evolutionary branching among co-mimics are significantly more similar than expected by chance (Table 1; Figure S3), despite elements of incongruence such as those described above. This suggests that an early lack of phylogenetic resolution [14] and as well as the complexity of more recent estimates of phylogenetic branching

patterns [26] have previously concealed significant topological congruence between the phylogenies of *H. erato* and *H. melpomene*, which is revealed by quantitative cophylogenetic analysis.

Phylogenies of *H. erato* and *H. melpomene* based on a total-evidence reanalysis of the complete ten-gene dataset of Hines et al. [18] also show significant topological congruence (Table 1; Figure S4), and biogeographic clustering patterns complementary to those of the coalescent population phylogenies illustrated in Figures 2. Similar reanalyses of only the five colour pattern genes from this dataset [18] are less able to cluster individuals of the same morph and show reduced biogeographic signal but also indicate statistically significant topological congruence between phylogenies of *H. erato* and *H. melpomene* co-mimics (Table 1; Figure S5).

As expected for recent evolutionary radiations of populations that still experience low level gene flow, phylogenetic reconstructions for *H. erato* and *H. melpomene* are subject to some uncertainty, and there are differences between phylogenetic estimates based on different gene partitions (e.g. Figures S4 and S5), taxon partitions, and methodologies (e.g. the MDC and Bayesian phylogenetic estimates shown in Figures 2 and S1, respectively). However, several topological features are common to phylogenetic estimates based on different methodologies and data partitions (as discussed above) and the consistent result that emerges when we consider these various phylogenetic estimates is one of statistically significant topological congruence in the branching patterns of co-mimetic populations within these two species.

To be compatible with codivergence, ecologically associated phylogenies must be both topologically and temporally congruent [49]. For example, the phylogenies of *H. erato* and *H. melpomene* might show topological but not temporal congruence if wing patterns arising from an earlier radiation (previously suggested to be that of *H. erato* [14,25]) were

secondarily converged upon during a later, but topologically similar, radiation by a mimic (previously suggested to be *H. melpomene*) [15]. Previous analyses have generally suggested that the phylogenies of *H. erato* and *H. melpomene* were temporally incongruent. In his influential paper of 1996 [14], Brower estimated that two eastern clades within *H. erato* and *H. melpomene* were of similar ages (150,000 – 200,000 Y), based on uncorrected average within-clade sequence divergence. However, his estimation that a key divergence between populations East and West of the Andes occurred earlier in *H. erato* (1.5-2 MYA) than in *H. melpomene* (65,000 YA) [14] has been taken as evidence against codivergence of the two species [14,23]. In the same vein, Flanagan et al. [25] suggested that *H. erato* was approximately twice as old as *H. melpomene*, based on corrected genetic divergences from their nearest relatives (thought to be *Heliconius hecalesia* and *Heliconius cydno*, respectively). These apparent discrepancies in divergence times have previously been taken as evidence against coevolution of *H. erato* and *H. melpomene* [14,23].

As in previous studies [14], we find that *H. melpomene* shows lower genetic diversity than *H. erato*, as measured by the average uncorrected pair-wise divergence between individuals. However, population genetic comparisons indicate that the effective population size (estimated as θ, the product of effective population size and mutation rate) of *H. melpomene* is smaller than that of *H. erato* (observed here and also by Flanagan et al. [25]). The effective population size is the size of an idealized breeding population that would experience the same effects of random mutation as a real population under study [52]. Effective population size is generally positively related to, but less than, the census population size [53]. Therefore field observations suggesting that *H. melpomene* generally has a census population size approximately half that of *H. erato* (e.g. see [15]) are compatible with the difference in effective population size estimated from sampled genetic variation. Within-species genetic diversity is positively correlated with effective population size [43,53-

55]. Indeed, the population size parameter θ determines the average genetic diversity of the population, because it takes into account both effective population size and mutation rate (these parameters can be separated using independent estimates of the mutation rate, for example from fossil calibrated divergence times [54], however such information is not available for *H. erato* or *H. melpomene*). The estimated difference in effective population size predicts that average genetic variation could be lower within the less abundant species *H. melpomene*, even if its radiation was temporally congruent with that of *H. erato* (e.g. see [43]).

Like Flanagan et al. [25], we estimated splits between *H. melpomene* and closely related clades (the *H. cydno* clade or the silvaniforms) to be younger than the split between *H. erato* and its close relative *H. hecalesia* (Figures S1 and S2). This finding was supported both by differences in average sequence divergence (here and also Flanagan et al. [25]) and by the Bayesian coalescent phylogenetic analyses (which are more robust, since they estimate and take into account effective population sizes [29]). However, the crucial test for codivergence is not whether *H. erato* and *H. melpomene* first diverged from respective outgroups at similar times but whether their internal population radiations were temporally congruent.

Our Bayesian coalescent reconstructions date the bases of the sampled clades at approximately 350,000 years ago for both *H. erato* and *H. melpomene* (Figure 3), with overlapping 95% Bayesian Credibility Intervals. This is compatible with a contemporaneous codivergence event at the start of the sampled radiations of these species (e.g. Figure 3). Thus we concur with Brower's [14] suggestion that much of the phenotypic diversity within *H. erato* and *H. melpomene* evolved relatively recently, but estimate the origin of the sampled morphs of *H. erato* to be considerably more recent than his estimate of 1.5-2 MY, and contemporaneous with that of *H. melpomene*, contrary to his conclusions [14]. The first divergence between eastern and western populations (see Figure 3; Figures S1 and S2) is

dated at approximately 200,000 years BP for *H. erato* and at approximately 40,000 years BP for *H. melpomene* (though a smaller sample of western morphs were included for this species). However, 95% Bayesian Credibility Intervals overlap, suggesting that contemporaneous codivergence of western and eastern populations, within the two species, cannot be rejected, contrary to previous suggestions [14].

Overall, the population genetic and Bayesian coalescent divergence time estimates strongly suggests that the parallel phenotypic radiations of *H. erato* and *H. melpomene* occurred over an overlapping time period, contrary to previous suggestions [14,25,26]. The phylogenetic reconstructions and divergence time estimates are compatible with a series of contemporaneous codivergence events, occurring during a Müllerian mimicry relationship sustained over at least 350,000 years.

Codivergence and coevolution

Congruent phylogenies are often considered necessary to sustain hypotheses of (strictly reciprocal) coevolution [10,14]. Thus, our finding of significant topological and temporal congruence between the phylogenies of *H. erato* and *H. melpomene* demonstrates that coevolution between the two species was possible (contrary to some previous suggestions [14,26]). Furthermore, codivergence can be considered some of the strongest evidence that coevolution did occur [21,22]. In the case of *H. erato* and *H. melpomene*, the codivergent populations identified by the cophylogenetic reconstructions frequently represent distinct mimetic wing patterns (e.g. Figure 3). Thus, population codivergence is correlated with parallel genetic [2,56] and phenotypic [2] variation. When sustained codivergence is accompanied by multiple examples of parallel phenotypic change (as in the co-mimetic morphs of *H. erato* and *H. melpomene* illustrated in Figure 1), reciprocal coevolution can be considered a more probable mechanism than, for example, entirely one-sided evolutionary

change by one species (previously suggested to be the less abundant *H. melpomene*) to match its co-mimic (previously suggested to be the more abundant *H. erato*) [13].

While theory suggests that the mutual fitness benefits of Müllerian mimicry will promote coevolution [2,5,10], evidence for this has previously been rare [15] (though biogeographic comparisons suggesting that the dominant model *H. erato* has sometimes converged towards *H. melpomene* [13] offer another line of evidence for reciprocal evolution). Therefore, our demonstration of sustained codivergence between mimetic populations of *H. erato* and *H. melpomene* represents a key case for the study of coevolution.

Coevolution is a powerful concept because it describes a mechanism for the coordination of evolutionary change in genetically separate populations [12]. Consequently, evidence for coevolution has fundamental implications for ecology, population genetics and wider evolutionary theory [12]. We present evidence for phylogenetic codivergence between mimetic populations of *H. erato* and *H. melpomene*. Such codivergence represents some of the strongest evidence for coevolution [13,21]. Therefore, the parallel radiations of *H. erato* and *H. melpomene* support a hypothesis of reciprocal coevolution between Müllerian co-mimics characterised by population codivergence and parallel phenotypic change [2]. Consequently, we suggest that these parallel radiations deserve to be reinstated (after [13,57]) as the most striking known example of coevolution.

Acknowledgements

Satellite image of Central and South America courtesy of NASA Earth Observatory. Photographs of *Heliconius* type specimens courtesy of Butterflies of America (specimens held by The Natural History Museum, London) and The Linnean Society of London. This research was supported by a Discovery Project grant from the Australian Research Council.

References

1. Brown JKS (1981) The biology of *Heliconius* and related genera. Ann Rev Entomol 26: 427-456.

2. Sheppard PM, Brown KS, Benson WW, Singer MC (1985) Genetics and the evolution of Müllerian mimicry in *Heliconius* butterflies. Phil Trans Roy Soc Lond 308: 433-613.

3. Engler H, Spencer KC, Gilbert LE (2000) Preventing cyanide release from leaves. Nature 406: 144-145.

4. Kapan DD (2001) Three-butterfly system provides a field test of Müllerian mimicry. Nature 409: 338-340.

5. Müller F (1879) *Ituna* and *Thyridia*; a remarkable case of mimicry in butterflies. Trans Entomol Soc Lond 1879: xx-xxix.

6. Bates HW (1862) Contributions to an insect fauna of the Amazon valley (Lepidoptera: Heliconidae). Trans Linnean Soc 23: 495-556.

7. Reed RD, Papa R, Martin A, Hines HM, Counterman BA, et al. (2011) *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. Science 333: DOI: 10.1126/science.1208227.

8. Benson WW (1972) Natural selection for Müllerian mimicry in *Heliconius erato* in Costa Rica. Science 176: 936-939.

9. Thompson JN (1994) The coevolutionary process. Chicago: University of Chicago Press.

10. Joron M, Mallet JL (1998) Diversity in mimicry: paradox or paradigm? Trends Ecol Evol 13: 461-466.

11. Janzen DH (1980) When is it coevolution? Evolution. 34: 611-612.

12. Thompson JN (1989) Concepts of coevolution. Trends Ecol Evol 4: 179-183.

13. Gilbert LE (1983) Coevolution and mimicry. In: Futuyma DM, Slatkin M, eds. Coevolution. Sunderland: Sinauer Associates Inc. pp. 263-281.

14. 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.

15. Mallet J (1999) Causes and consequences of a lack of coevolution in Müllerian mimicry. Evol Ecol 13: 777-806.

16. Beltrán 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 92: 221-239.

17. Counterman BA, Araujo-Perez F, Hines HM, Baxter SW, Morrison CM, et al. (2010) Genomic hotspots for adaptation: the population genetics of Müllerian mimicry in *Heliconius erato*. PLoS Genet 6 DOI: 10.1371/journal.pgen.1000796.

18. Hines HM, Counterman BA, Papa R, de Moura PA, Cardoso MZ, et al. (2011) Wing patterning gene redefines the mimetic history of Heliconius butterflies. Proc Natl Acad Sci USA 49: 19666-19671.

19. Kronforst MR, Young LG, Kapan DD, McNeely C, ONeill RJ, 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.

20. Arias CF, Muñoz AG, Jiggins CD, Mavárez J, Bermingham E, et al. (2008) A hybrid zone provides evidence for incipient ecological speciation in *Heliconius* butterflies. Mol Ecol 17: 4699-4712.

21. Page, RDM (2003). Introduction. In: Page RDM, ed. Tangled trees: Phylogeny, cospeciation, and coevolution. Chicago: The University of Chicago Press. pp. 1-21.

22. Futyma DJ, Slatkin M (1983) Introduction. In: Coevolution. Sunderland : Sinauer Associates Inc. pp. 1-13.

23. Mallet J, Jiggins CD, McMillan WO (1996) Mimicry meets the mitochodrian. Curr Biol 6: 937-940.

24. Knapp S, Mallet J (2003) Refuting refugia? Science 300: 71-72.

25. Flanagan NS, Tobler A, Davison A, Pybus OG, Kapan DD, et al. (2004) Historical demography of Müllerian mimicry in the neotropical *Heliconius* butterflies. Proc Natl Acad Sci USA 101: 9704-9709.

26. Quek S-P, Counterman BA, de Moura PA, Cardoso MZ, Marshall CR, et al. (2010) Dissecting comimetic radiations in *Heliconius* reveals divergent histories of convergent butterflies. Proc Natl Acad Sci USA 107: 7365–7370.

27. Ceccarelli FS, Crozier RH (2007) Dynamics of the evolution of Batesian mimicry: moelcular phylogenetic analysis of ant-mimicking *Myrmarachne* (Aranae: Salticidae) species and their ant models. J Evolution Biol 20: 286-295.

28. Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. Syst Biol 55: 21-30.

29. Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. Mol Biol Evol 27: 570-580.

30. Simonsen TJ, Zakharov EV, Djernaes M, Cotton AM, Vane-Wright RI, et al. (2011) Phylogenetics and divergence times of Papiloninae (Lepidoptera) with special reference to the enigmatic genera *Teinopalpus* and *Meandrusa.* Cladistics 27: 113-137.

31. Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat Rev Genet 3: 380-390.

32. Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. Syst Biol 55: 21-30.

33. Eckert AJ, Carstens BC (2008) Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. Mol Phylogenet Evol 49: 832-842.

34. Zhang C, Zhang D-X, Zhu T, Yang Z (2011) Evaluation of a Bayesian coalescent method of species delimitation. Syst Biol 60: 747-761.

35. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16: 1114-1116.

36. Beltrán M, Jiggins CD, Bull V, Linares M, Mallet J, et al. (2002) Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies. Mol Biol Evol 19: 2176-2190.

37. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797.

38. Jobb G, Haeseler AV, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol 4: 18.

39. Posada D (2008) jModeltest: phylogenetic model averaging. Mol Biol Evol 25: 1253-1256.

40. Akaike H (1974) A new look at statistical model identification. IEEE T Automat Contr 19: 716-723.

41. Hey J, Wakely J (1997) A coalescent estimator of the population recombination rate. Genetics 145: 833-846.

42. Watterson GA (1975) On the number of segregating sitres in genetical models without recombination. Theor Popul Biol 7: 256-276.

43. Wakely J, Hey J (1997) Estimating ancestral population parameters. Genetics 145: 847-855.

44. Swofford DL (2000) PAUP* Phylogenetic analysis using parsimony (*and other methods) version 4. Sunderland: Sinauer Associates.

45. Maddison WP, Maddison DR (2010) Mesquite: a modular system for evolutionary analysis version 2.73. http://mesquiteproject.org.

46. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.

47. Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. Proc Natl Acad Sci USA 107: 9264-9269.

48. McCormack JE, Heled J, Delaney KS, Townsend Peterson A, Lacey Knowles L (2010) Calibrating divergence times on species trees versus gene trees: implications for speciation history of *Aphelocoma* jays. Evolution 65: 184-202.

49. Charleston MA, Robertson DL (2002) Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. Syst Biol 51: 528-535.

50. Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R (2010) Jane: a new tool for the cophylogeny reconstruction problem. Algorithm Mol Biol 5: 16.

51. Charleston MA (1998) Jungles: a new solution to the host/parasite phylogeny reconciliation problem. Math Biosci 149: 191-223.

52. Wright S (1931) Evolution in Mendelian populations. Genetics 16: 97-159.

53. Frankham R (1996) Relationship of genetic variation to population size in wildlife. Conserv Biol 10: 1500-1508.

54. Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. Nat Rev Genet 10: 195-205.

55. Kronforst MR, Gilbert LE (2008) The population genetics of mimetic diversity in *Heliconius* butterflies. P Roy Soc B 275: 493-500.

56. Baxter SW, Nadeau NJ, Maroja LS, Wilkinson P, Counterman BA, et al. (2010) Genomic hotspots for adaptation: the population genetics of Müllerian mimicry in the *Heliconius melpomene* clade. PLoS Genet 6: 1-12.

57. Futuyma DJ (1986) Evolutionary biology: Second edition. Sunderland: Sinauer Associates Inc.

58. Wilson EB (1927) Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 22: 209-212.

59. Chamberlain NL, Hill RI, Kapan DD, Gilbert LE, Kronforst MR (2009) Polymorphic butterfly reveals the missing link in ecological speciation. Science 326: 847-850.

60. Bull V, Beltrán M, Jiggins CD, McMillan WO, Bermingham E, et al. (2006) Polyphyly and gene flow between non-sibling *Heliconius* species. BMC Biol 4: 11-27.

Figures and Figure legends

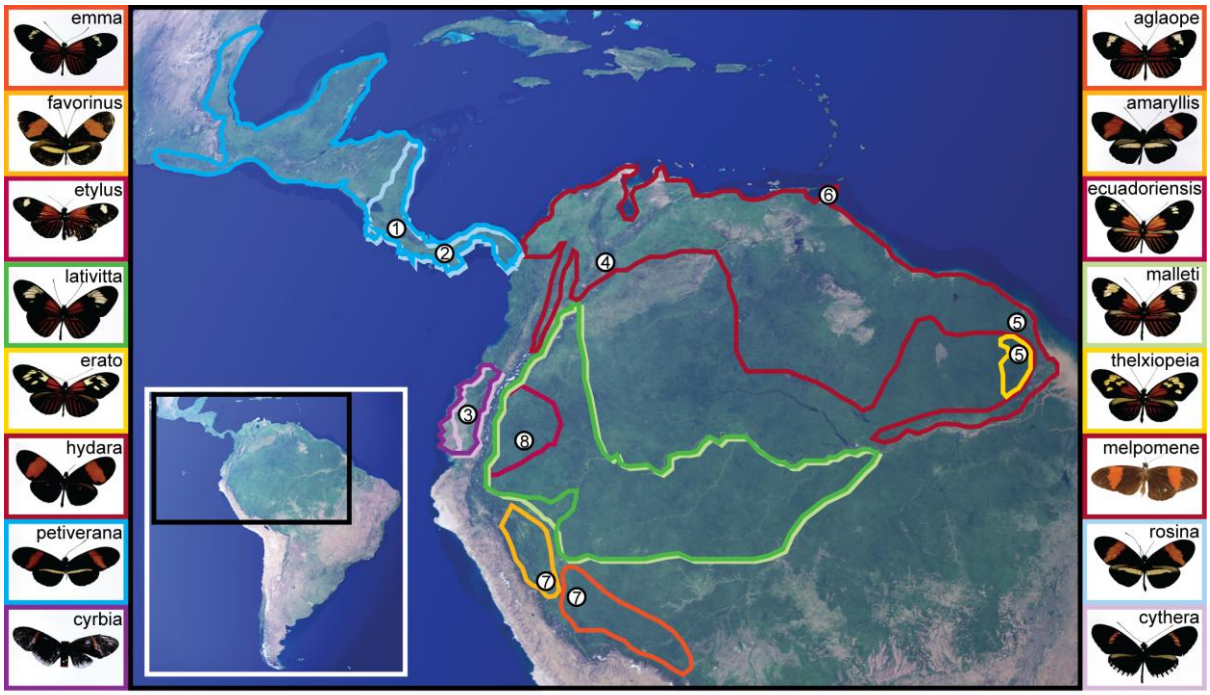

Figure 1. Wing patterns morphs and geographic distributions of the Müllerian co-mimics *H. erato* and *H. melpomene*. Mimetic morphs [2,14,26] (aligned rows), from *H. erato* (right) and *H. melpomene* (left), included in this study (photographs show the type specimens). Coloured boundaries on a satellite image of Central and South America indicate the geographic range of each morph [2]. Numbers indicate countries where morphs were sampled (by [25,26]): West of the Andes, 1 Costa Rica, 2 Panama, 3 West Ecuador; East of the Andes, 4 Colombia, 5 French Guiana, 6 Trinidad, 7 Peru and 8 East Ecuador.

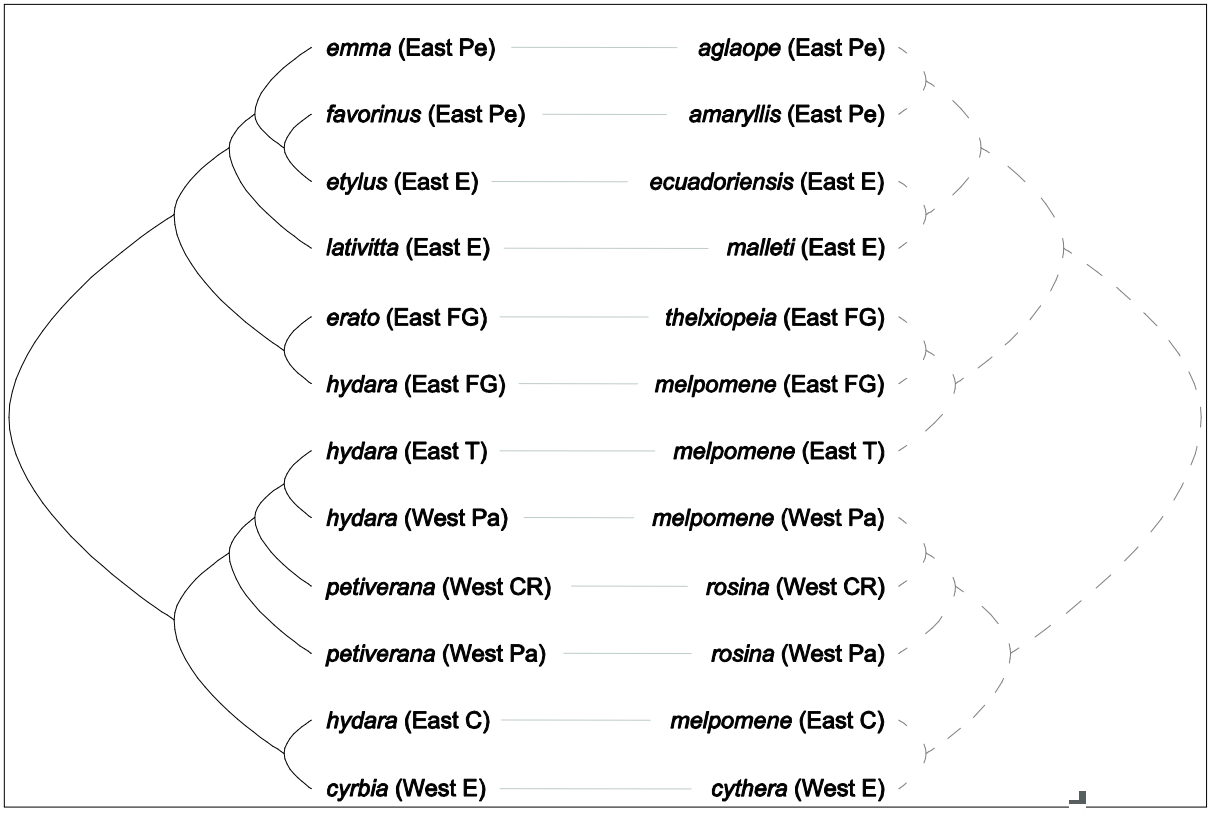

Figure 2. Phylogenies of *H. erato* and *H. melpomene* illustrating branching orders of co-mimetic country-level populations within each species. Example phylogenies independently estimated for *H. erato* (black, left) and *H. melpomene* (grey dashed, right) using the Minimise Deep Coalescence (MDC) method [28]. These correspond to cophylogenetic analysis "separate MDC countries 1" in Table 1. *H. erato* / *H. melpomene*

co-mimics (see Figure 1) are indicated by grey lines. This is one of several possible phylogeny pairs with similarly high congruence (see Table 1). Taxon labels indicate the sampled biogeographic region (East or West of the Andes), and country (abbreviations are: CR Costa Rica, Pa Panama, E Ecuador; C Colombia, FG French Guiana, T Trinidad and Pe Peru).

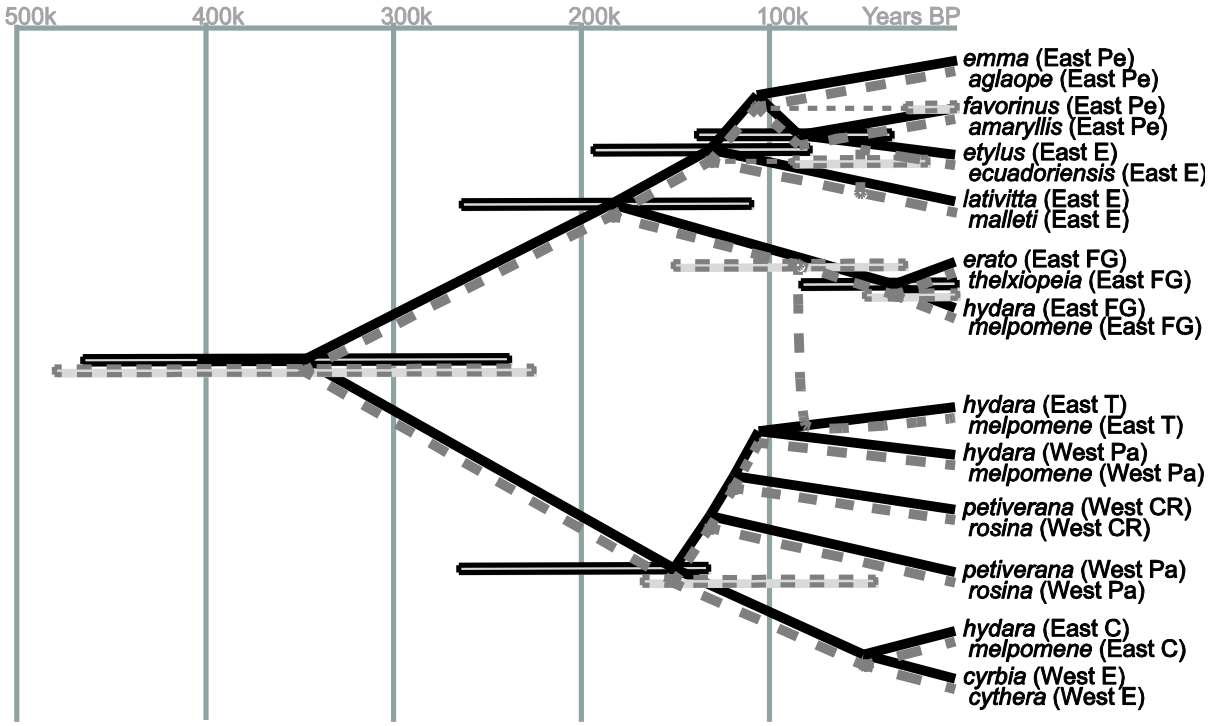

Figure 3. Cophylogenetic reconstruction of the history of mimicry between country-level populations of *H. erato* and *H. melpomene*. Example cophylogenetic history of the mimicry relationship between *H. erato* and *H. melpomene* based on MDC phylogenetic estimates (shown in Figure 2), and reconstructed using TreeMap 3. Bars indicate 95% Bayesian Credibility Intervals for divergence times. Solid grey dots correspond to reconstructed codivergence events; white dots represent duplication events which, in this case, are both followed by model switch events; the only mimicry loss event is indicated at

the most recent common ancestor of *H. e. hydara* populations from Trinidad and Panama. Taxon labels correspond to those in Figure 2.

Phylogenies	Minimum Cost (p [95% max])		Distance Correlation (p [95% max])	
	Random Associations	Random Mimic Tree	Root of Target (*melpomene*)	Root of Mimic (*erato*)
separate MDC countries 1	0 [0.0024]	0 [0.0024]	0.003 [0.0074]	0.001 [0.0042]
separate MDC countries 2	0 [0.0024]	0 [0.0024]	0.002 [0.0059]	0 [0.0024]
separate MDC regions 1	0.200 [0.2254]	0.181 [0.2054]	0 [0.0024]	0 [0.0024]
separate MDC morphs 1	0.025 [0.0354]	0.016 [0.0245]	0.004 [0.0089]	0.001 [0.0042]
combined MDC countries 1	0.005 [0.0103]	0.004 [0.0089]	0.001 [0.0042]	0 [0.0024]
combined MDC countries 2	0.003 [0.0074]	0 [0.0024]	0 [0.0024]	0 [0.0024]
combined MDC countries 3	0.005 [0.0103]	0 [0.0024]	0.002 [0.0059]	0.004 [0.0089]
combined MDC countries 4	0 [0.0024]	0 [0.0024]	0.001 [0.0042]	0.001 [0.0042]
combined MDC regions 1	0.225 [0.2514]	0.212 [0.2379]	0.017 [0.0257]	0.015 [0.0233]
combined MDC morphs 1	0.547 [0.5784]	0.529 [0.5605]	0.054 [0.0686]	0.086 [0.1040]
combined *BEAST countries	0.029 [0.0401]	0.029 [0.0401]	0.088 [0.1061]	0.077 [0.0941]
combined *BEAST morphs	0.003 [0.0074]	0.019 [0.0282]	0.007 [0.0130]	0.005 [0.0103]
Quek et al., 2010 AFLP	N / A		0 [0.0024]	0 [0.0024]
Hines et al. 2011 ten-genes	N / A		0 [0.0001]	0 [0.0001]
Hines et al. 2011 colour pattern genes	N / A		0 [0.0001]	0 [0.0001]

Table 1. Significance of congruence between *H. erato* and *H. melpomene* phylogenies. Phylogenies were estimated separately for the clades *H. erato* and *H. melpomene* and in a combined analysis of the *Heliconius*. The significance of the Cophylogeny mapping analyses was estimated based on minimising total reconstruction costs with Jane 3, either by randomising the leaf associations (column 2) or by randomizing the *H. erato* phylogeny (column 3). The significance of the pair-wise distance correlation was calculated, at the root of the *H. melpomene* phylogeny (column 5) and the root of the *H. erato* phylogeny (column 6), using TreeMap 3. Each *p*-value was estimated with 1000 Monte-Carlo replicates, and the 95% confidence upper bound was calculated for each using Wilson's score interval for binomial proportions [58].

Supporting Information

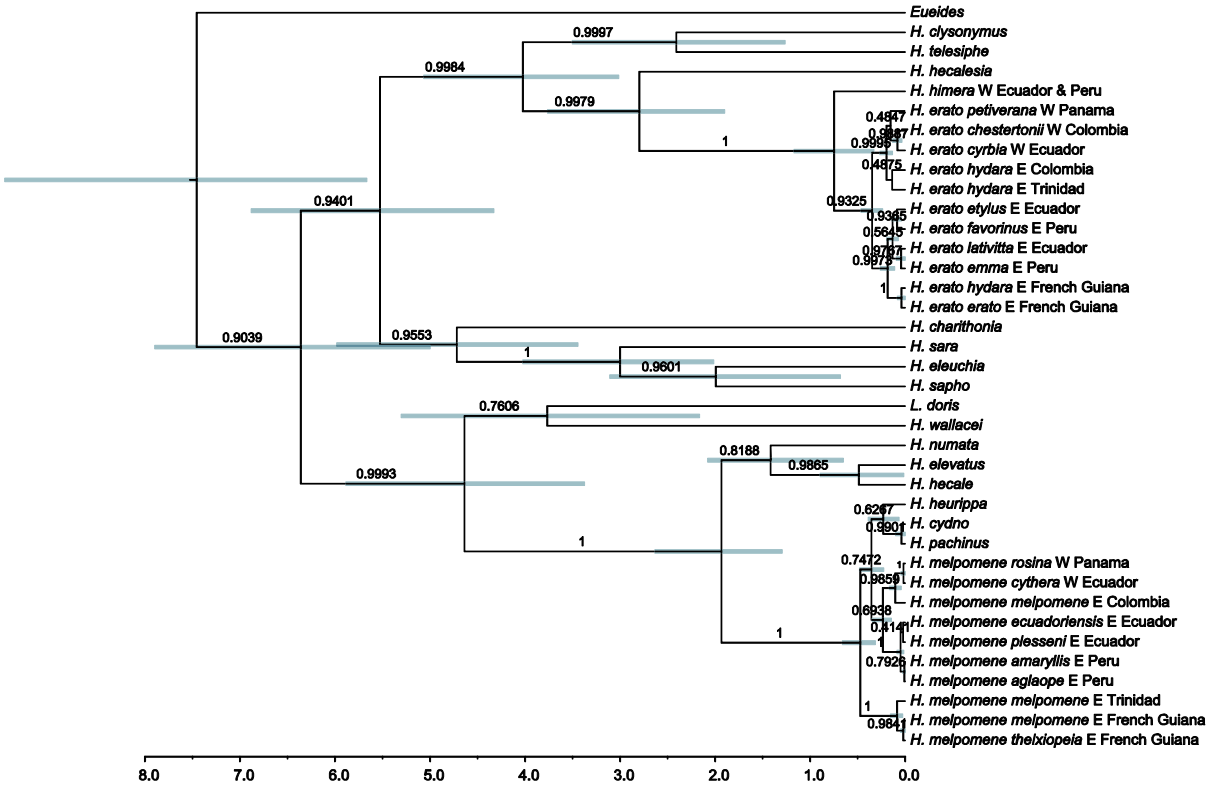

Figure S1. Bayesian coalescent phylogeny for the *Heliconius* with country level populations of *H. erato* and *H. melpomene*: branch labels give posterior probabilities, the axis indicates time (MY BP), and scale bars show 95% Bayesian Credibility Intervals for the mean node age.

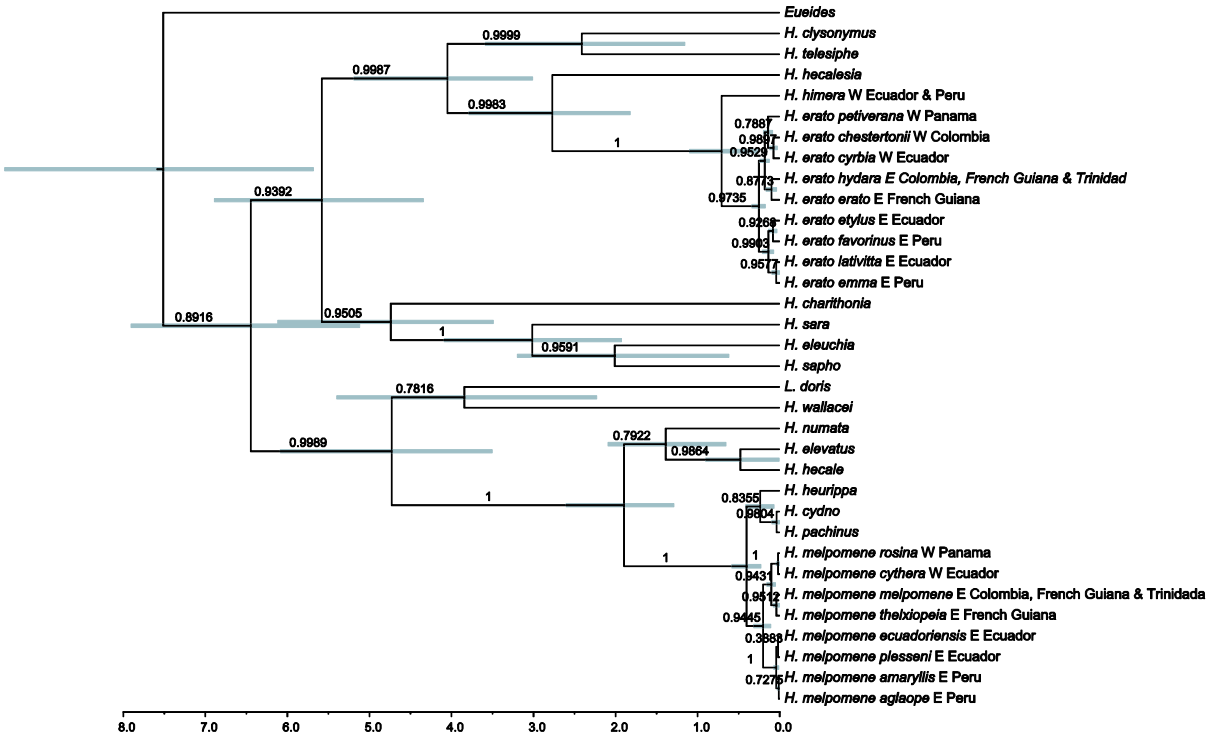

Figure S2. Bayesian coalescent phylogeny for the *Heliconius* with morph level populations of *H. erato* and *H. melpomene*, labelled as for Figure S1.

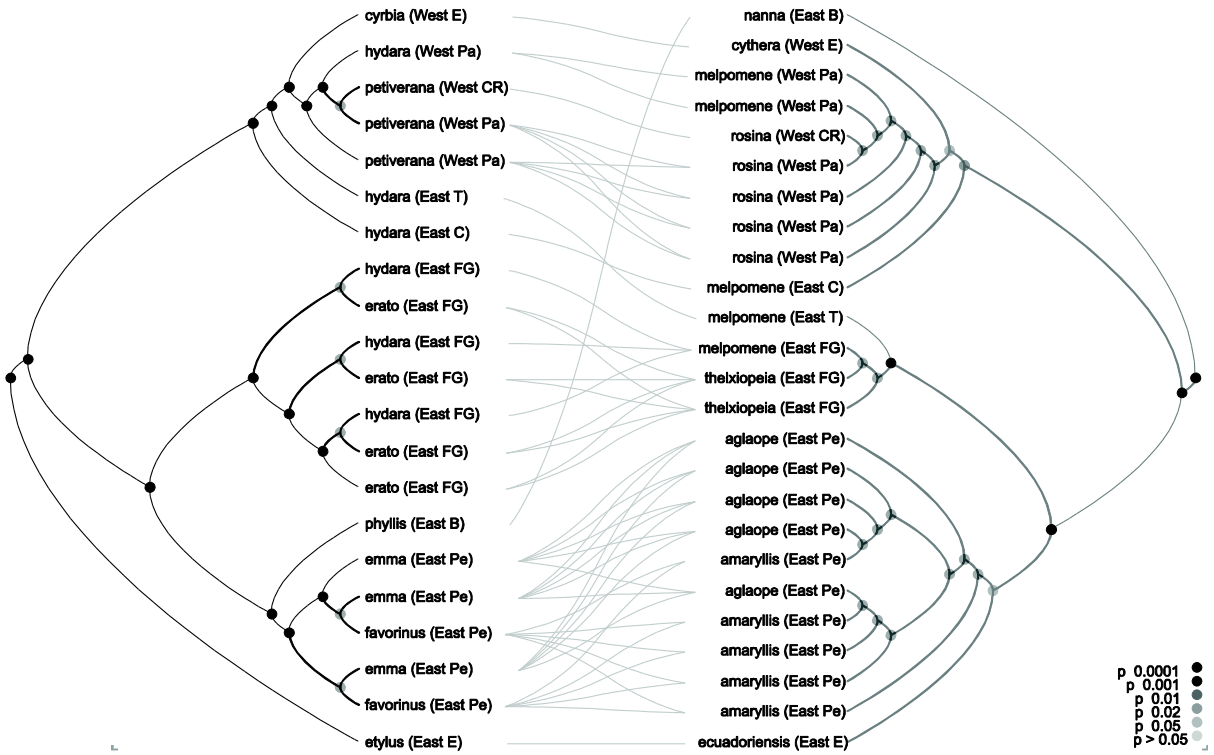

Figure S3. Phylogenies for *H. erato* (left) and *H. melpomene* (right) reproduced from [26] and corresponding to cophylogenetic analysis "Quek et al., 2010 AFLP" in Table 1. *H. erato* / *H. melpomene* co-mimics sampled from the same country are indicated by grey lines. Taxon labels indicate the sampled biogeographic region (East or West of the Andes), and country (abbreviations correspond to Figure 2). Shaded circles indicate the significance of a pairwise correlation test conducted for the shaded node (with p values corresponding to the key).

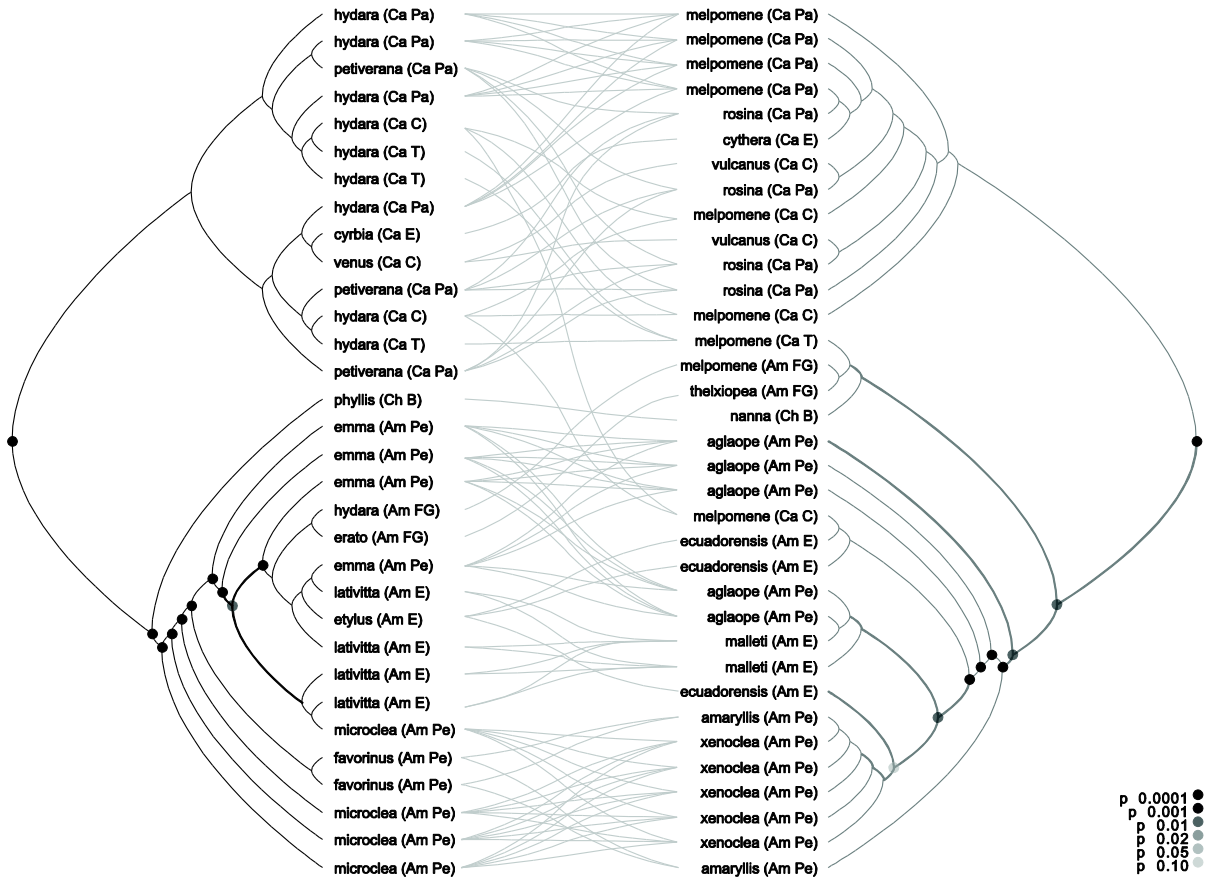

Figure S4. Maximum likelihood phylogenies independently estimated for *H. erato* (left) and *H. melpomene* (right) based on the ten-gene dataset of [18], corresponding to cophylogenetic analysis "Hines et al. 2011 ten-genes" in Table 1. Taxon labels indicate the sampled biogeographic region (abbreviations are: Am Amazon, Ca Caribbean, Ch Chocoan-Parana), and country (abbreviations correspond to Figure 2 with additional abbreviation: B Brazil). Further annotation corresponds to Figure S3.

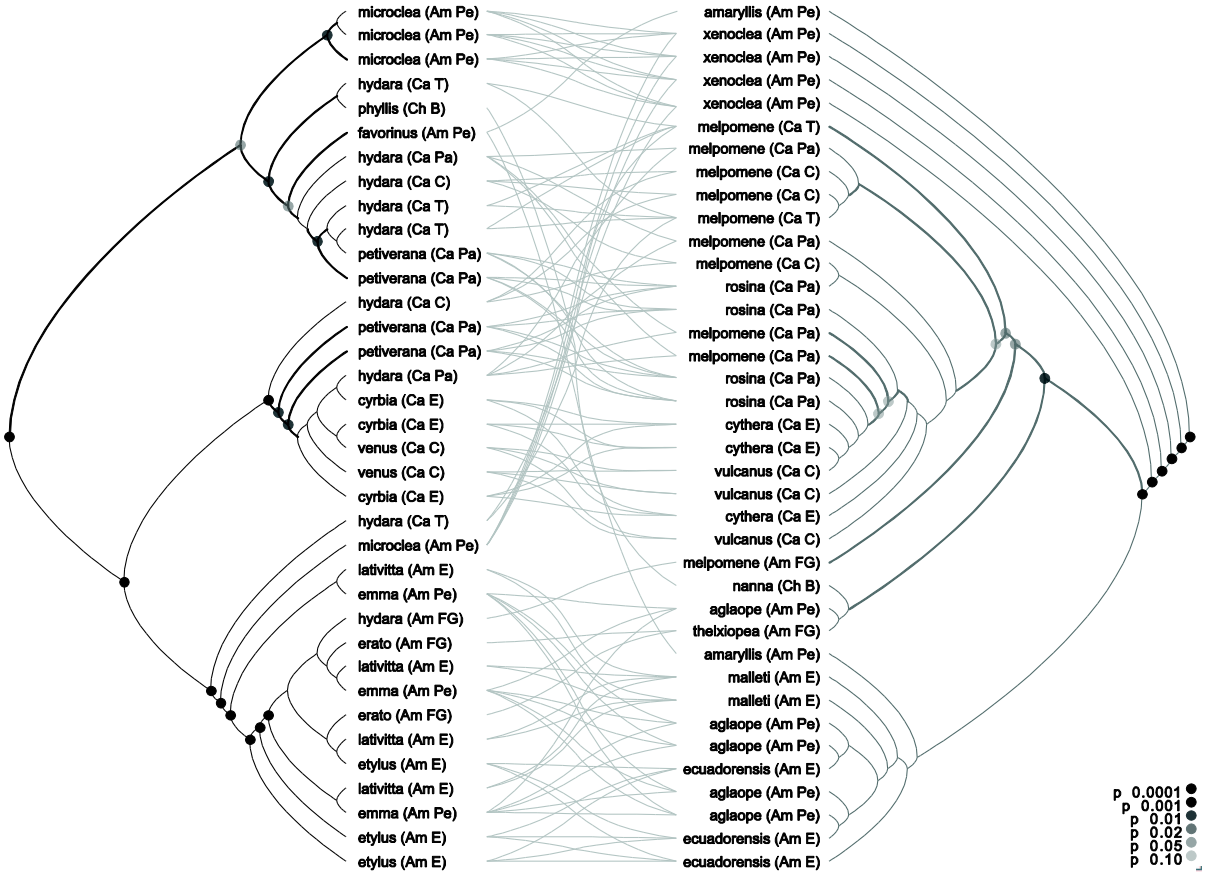

Figure S5. Phylogenies reconstructed as for those of Figure S4 except based on only the five colour pattern genes of [18] and corresponding to cophylogenetic analysis "Hines et al. 2011 colour pattern genes" in Table 1. Further annotation corresponds to Figure S3.

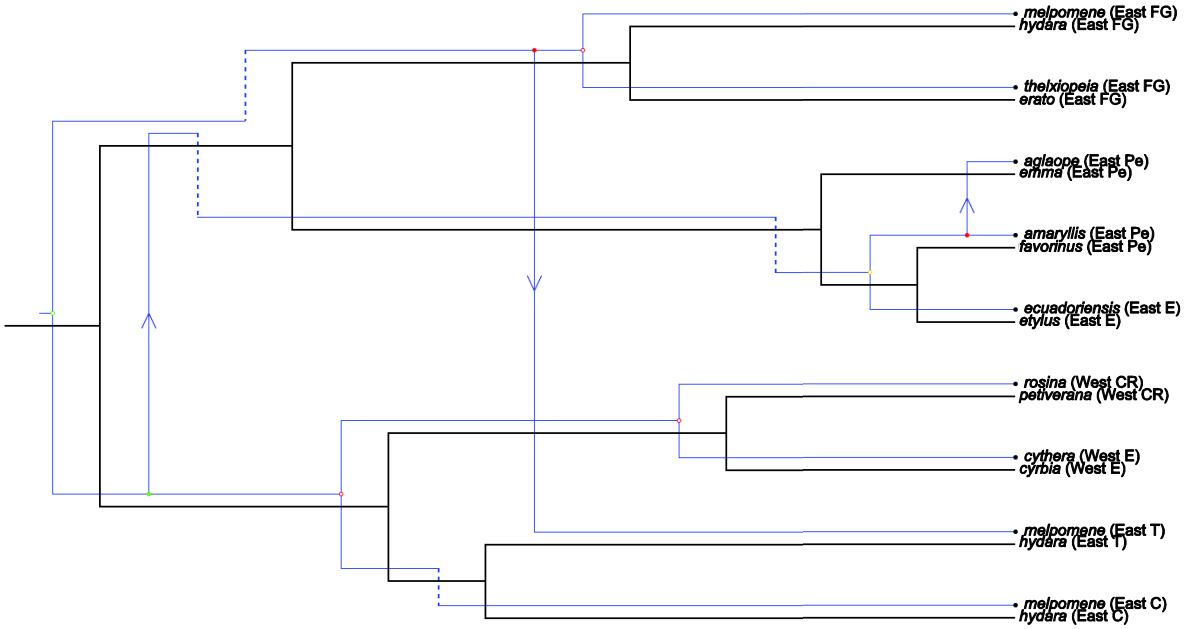

Figure S6. History of mimicry between *H. erato* (black phylogeny) and *H. melpomene* (blue phylogeny), reconstructed using Jane 3, based on the phylogeny shown in Figure S1: white-filled circles represent codivergence, solid circles represent duplications, arrows represent model switches, and dashed lines represent losses.

Table S1. Sampling information for the study system, including accession numbers (not available in preprint version).

Table S2. Significance of congruence between phylogenies with a model (*H. melpomene*) to mimic (*H. erato*) relationship reversed relative to Table 1.

Phylogenies	Minimum Cost (p [95% max])		Distance Correlation (p [95% max])	
	Random Associations	Random Mimic Tree	Root of Target (*melpomene*)	Root of Mimic (*erato*)
separate MDC countries 1	0 [0.0024]	0 [0.0024]	0.002 [0.0059]	0.001 [0.0042]
separate MDC countries 2	0 [0.0024]	0 [0.0024]	0 [0.0024]	0 [0.0024]
separate MDC regions 1	0.002 [0.0059]	0.007 [0.0130]	0 [0.0024]	0 [0.0024]
separate MDC morphs 1	0.018 [0.0270]	0.017 [0.0257]	0.002 [0.0059]	0.001 [0.0042]
combined MDC countries 1	0.007 [0.0103]	0.007 [0.0130]	0 [0.0024]	0 [0.0024]
combined MDC countries 2	0.001 [0.0042]	0.003 [0.0074]	0 [0.0024]	0.001 [0.0042]
combined MDC countries 3	0.004 [0.0089]	0.004 [0.0089]	0 [0.0024]	0 [0.0024]
combined MDC countries 4	0.005 [0.0103]	0.004 [0.0089]	0 [0.0024]	0.003 [0.0074]
combined MDC regions 1	0.123 [0.1439]	0.121 [0.1418]	0.01 [0.0170]	0.01 [0.0170]
combined MDC morphs 1	0.665 [0.6948]	0.591 [0.6220]	0.07 [0.0864]	0.062 [0.0776]
combined *BEAST countries	0.009[0.0157]	0.009 [0.0157]	0.096 [0.1148]	0.095 [0.1138]
combined *BEAST morphs	0.012 [0.0195]	0.007 [0.0130]	0.01 [0.0170]	0.009 [0.0157]

Table S3. DNA sequences used in this study: alignment in Nexus format (not available in preprint version).