The speciation history of *Heliconius*: inferences from multilocus DNA sequence data

by

Margarita Sofia Beltrán

A thesis submitted for the degree of
Doctor of Philosophy of the University of London

September 2004

Department of Biology
University College London

Abstract

Heliconius butterflies, which contain many intermediate stages between local varieties, geographic races, and sympatric species, provide an excellent biological model to study evolution at the species boundary. *Heliconius* butterflies are warningly coloured and mimetic, and it has been shown that these traits can act as a form of reproductive isolation. I present a species-level phylogeny for this group based on 3834bp of mtDNA (COI, COII, 16S) and nuclear loci (*Ef1α*, *dpp*, *ap*, *wg*). Using these data I test the geographic mode of speciation in *Heliconius* and whether mimicry could drive speciation. I found little evidence for allopatric speciation. There are frequent shifts in colour pattern within and between sister species which have a positive and significant correlation with species diversity; this suggests that speciation is facilitated by the evolution of novel mimetic patterns. My data is also consistent with the idea that two major innovations in *Heliconius*, adult pollen feeding and pupal-mating, each evolved only once. By comparing gene genealogies from mtDNA and introns from nuclear *Tpi* and *Mpi* genes, I investigate recent speciation in two sister species pairs, *H. erato*/*H. himera* and *H. melpomene*/*H. cydno*. There is highly significant discordance between genealogies of the three loci, which suggests recent speciation with ongoing gene flow. Finally, I explore the phylogenetic relationships between races of *H. melpomene* using an AFLP band tightly linked to the *Yb* colour pattern locus (which determines the yellow bar in the hindwing). At this locus, races group according to geographical location rather than clades inferred from colour pattern phenotype. This and similar loci can be used to facilitate comparative mapping with other *Heliconius* species. In summary, the patterns at each phylogenetic level using different gene regions are consistent with a scenario of rapid, adaptively driven divergence and speciation in this group.

Querido Eduardo, poeta de Colombia:

Cuando por muchos años y por muchas regiones mi pensamiento se detenía en Colombia, se me aparecía tu vasta tierra verde y forestal, el río Cauca hinchado por las lágrimas de María y planeando sobre todas las tierras y los ríos, como pañuelos de terciopelo celestial, las extraordinarias mariposas amazónicas, las mariposas de Muzo. Siempre vi tu país al través de una luz azul de mariposas bajo este enjambre de alas ultravioleta, y vi también los caseríos desdoblados en este tembloroso vaivén de alas, y luego vi la historia de Colombia seguida por un cometa de mariposas azules: sus grandes capitanes, Santander, Bolívar con una mariposa luminosa posada en cada hombro, como la más deslumbrante charretera, y a tus poetas, infortunados como José Asunción o como Porfirio o soberbios como Valencia, perseguidos hasta el fin de su vida por una mariposa, que olvidaban de pronto en el sombrero o en un soneto, mariposa que voló cuando Silva consumó su romántico suicidio para posarse más tarde tal vez sobre tus sienes, Eduardo Carranza.

***Pablo Neruda** poeta chileno en homenaje*
al poeta colombiano Eduardo Carranza (1946)

Nada me apartará del corazón verde de Colombia... Mi poesía seguirá celebrándote, Esmeralda. Luego el Museo de Oro Precolombino: con sus máscaras, collares, caracoles, mariposas, ranitas refulgentes……

Pablo Neruda

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

Appendix 3.1

Chapter 4

Chapter 5

Table of Figures

Chapter 1

Chapter 2

Chapter 3

Appendix 3.1

Chapter 4

Acknowledgments

Many thanks to Chris Jiggins who has been my supporter, advisor and editor throughout my graduate career. Eldredge Bermingham for giving me the opportunity to work in his lab, for discussions and guidance throughout the project and for a wonderful time in his beloved Panamá. Jim Mallet for invaluable supervision and collaboration, and for having faith in me. Owen McMillan for shared primers, techniques, comments and lovely weekends in Culebra. Stella Betancourt, Crisanto Velandia and La Family Happiness, who are always there for me. Silvia and Roger Jiggins for improving my English, my cooking and for bird watching weekends. Maribel Gonzales, Liliana Cortez, Miriam 'La Dra. Cocodrilo' de Anaya, Oris Sanjur, Nimidiana Gomez for invaluable help in the Naos lab and for always making me laugh. And thanks to all the many other people in STRI for friendship, barbeques, parrandas, and lots of fun. Alaine Whinnett for cookies and smiles. Marie Zimmermann, Mathieu Joron and Fraser Simpson for chasing butterflies with me in Peru. Carla Penz, Andrew Brower, Mauricio Linares, Keith Willmott, Gerardo Lamas and Luis Mendoza Cuenca for sharing samples and taxonomic expertise. Alexandra Tobler for allowing me to use unpublished primers, and helping to develop protocols. Grethel Grajales and Gustavo Concheiro for help with GIS and maps. Nick Isaac for help using MacroCAIC. Michael Turelli for ideas about the geographical analysis. Catalina Estrada and Roxana Yockteng, for always being there as my friends. Rahul Shah for a floor to sleep on in London, and Mauricio Sarmiento for making the days more cheerful in London. Finally, for financial support I thank the Overseas Research Scheme, the Bogue Fellowships and the Smithsonian Tropical Research Institute, and for collecting permits ANAM in Panama and Ministerio del Ambiente in Ecuador.

Preface

Colombia, my country, is the second most diverse on the planet, the third most diverse for butterflies 10% of which are endemic (Mittermeier, 1999). My main interest is to understand the patterns and processes of speciation that are fundamental to explaining the diversity of life. In this study I use phylogenies of *Heliconius* butterflies, which contain many intermediate stages between local varieties, geographic races, and sympatric species to study evolution at the species boundary.

The first chapter describes a species-level phylogeny for subtribe Heliconiina based on mtDNA (COI, COII, 16S) and conserved regions of nuclear loci (*Ef1α*, *dpp*, *ap*, *wg*). This well-supported phylogenetic hypothesis shows *Heliconius* to be paraphyletic with respect to its sister genera *Laparus* and *Neruda*, but these three genera are monophyletic with respect to *Eueides*. Clarification of the relationships between these sisters genera helps to understand how unique traits such as pollen feeding and pupal mating behaviour have evolved in *Heliconius*. There is a well-supported monophyletic 'pupal mating clade' suggesting that this trait evolved once. *Heliconius*, *Laparus doris* and *Neruda* form a single clade, demonstrating a single origin for pollen feeding, but with a loss of the trait in *Neruda*. However, the position of *Neruda* within *Heliconius* is not strongly supported, and *Neruda* can be placed basal to *Heliconius* without significant loss of support. I therefore cannot reject the most parsimonious hypothesis for the evolution of pollen feeding, that of a single origin in a sister taxa to *Neruda* that went on to diversify into present day *Laparus* and *Eueides*. Pollen feeding may therefore have evolved uniquely and never have been lost.

In the second chapter I use the species-level phylogeny described in the first chapter combined with information on the geographical and ecological attributes of *Heliconius*, to detect the geographical pattern of speciation in this group and to test if changes in colour pattern are important in speciation. The data suggest that allopatric speciation is not the predominant mode of speciation in *Heliconius*.

Mapping of the colour traits onto the phylogeny shows that there are frequent shifts in colour pattern within and between sister species and these changes can occur in sympatry or allopatry. Additionally, using a new methodology of comparative analysis I show that clades with greater colour pattern diversification also have greater species richness than expected at random. This suggests that speciation is facilitated by the evolution of novel mimetic patterns.

In chapter three I use fast-evolving markers to explore the relationships between sister species in *Heliconius*. Here I describe two new nuclear markers for *Heliconius* (*Tpi* and *Mpi*) and their use to investigate the phylogeny of recent speciation of two sister species pairs, *H. erato* and *H. himera* and *H. melpomene* and *H. cydno*. The nuclear intron sequences evolved at rates similar to those of mitochondrial coding sequences, but the phylogenetic utility of introns is restricted to closely related geographic populations and species due to high levels of indel variation. Genealogies show that genetic variation in maternally inherited mtDNA and sex-linked *Tpi* gene clusters by species. However, the *Mpi* genealogies in both species pairs fail to show structure consistent with species boundaries and suggested recent introgression between species. The discordance between gene genealogies is consistent with models of adaptive speciation with ongoing gene flow proposed for recently diverged sister species in *Heliconius* (Beltran *et al*., 2002).

In chapter four I investigate the relationships between races within *H. melpomene*. Until now attempts to understand the evolution of colour pattern between races were based inferences from either crossing experiments or neutral markers such as allozymes and mtDNA. Here, I describe the identification and phylogenetic history of a marker (a41) linked to a colour pattern gene (*Yb*) in races of *H. melpomene*. The resulting cladogram matches biogeographical distributions better than wing pattern similarity. This might imply multiple origins of similar colour patterns in different geographic regions. This approach is not really conclusive about the colour pattern history and to improve this it may be necessary to identify regions closer to the locus controlling the colour pattern. However, the conversion

of a41 into a codominant marker is an important step towards positional cloning and molecular characterization of the regulatory mechanisms that produce this colour pattern phenotype.

Finally, in chapter 5 (conclusions) I discuss the use of different genes to resolve relationships at different phylogenetic levels and the utility of well-resolved phylogenetic hypotheses to understand the biological and ecological processes involved in speciation.

References

Beltran M, Jiggins CD, Bull V*, et al*. (2002) Phylogenetic discordance at the species boundary: Comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Molecular Biology and Evolution* **19**, 2176-2190.

Mittermeier CG (1999) *Hotspots: Earth's biologically richest and most endangered terrestrial ecoregions* CEMEX, Conservation International and Agrupacion Sierra Madre, Mexico.

Declaration

The chapters include data from collaborative work, in which I alone performed all the laboratory work. I wrote all the chapters, some of which may form the basis of co-authored publications of which I will be the first author.

The work described in chapter 1, 2 and 4 is entirely my own work. Appendix 3.1 of chapter 3 has already been published in *Molecular Biology and Evolution* and was co-authored with Chris D. Jiggins and Eldredge Bermingham, at the Smithsonian Tropical Research Institute, Panama, James Mallet and Vanessa Bull at The Galton Laboratory, Department of Biology, University College London, W. Owen McMillan at Department of Biology, University of Puerto Rico and Mauricio Linares at Instituto de Genética, Universidad de los Andes, Bogotá. However, all the molecular laboratory work and data analysis was my own work. Co-authors assisted in providing samples, discussing experimental strategy and in preparing the manuscript.

Margarita Beltrán	Prof. James Mallet
Candidate	Supervisor

Chapter 1

Do pollen feeding and pupal-mating have a single origin in *Heliconius* butterflies? Inferences from multilocus DNA sequence data

Abstract

A well-resolved phylogenetic hypothesis for *Heliconius* butterflies is needed to understand how unique traits such as pollen feeding and pupal-mating behaviour have evolved. Therefore, phylogenetic relationships among 60 Heliconiina species (86% of the whole subtribe) were inferred from partial DNA sequences of mitochondrial cytochrome oxidase I, II and 16S and fragments of nuclear genes *elongation factor-1α*, *apterous*, *decapentaplegic* and *wingless* (3834bp in total). My results suggest that *Heliconius* is paraphyletic, with *Laparus doris* and *Neruda* falling within the genus, demonstrating a single origin for pollen feeding but with a loss of the trait in *Neruda*. However different genes do not produce a clear agreement as regards relationships of *Neruda* within *Heliconius*, suggesting a rapid basal radiation of *Heliconius* and monophyly of pollen feeding is not entirely ruled out. There is also a highly supported monophyletic 'pupal-mating clade' suggesting that pupal mating evolved only once in the Heliconiina.

Introduction

Unique traits of Heliconius

The genus *Heliconius* or passion-vine butterflies together with the related *Laparus*, *Eueides* and *Neruda* are one of the best-known groups of Neotropical butterflies. These four genera have undergone rapid speciation and divergence. Heliconiines show amazing examples of mimicry and have been important in studies of ecological processes such as coevolution between insects and plants (Brown, 1981). *Heliconius* butterflies have two unique traits that may have facilitated rapid adaptive radiation, pollen feeding and pupal-mating behaviour (Gilbert, 1991). There is disagreement over the origin of these two unique

characteristics due to discordance among phylogenetic hypotheses (see systematics in *Heliconius* for more discussion). Gilbert (1991) used Brown's (1981) phylogeny of Heliconiina to describe important ecological events in the evolution of this group, such as pollen feeding and pupal-mating, and explain life-history innovations (Fig. 1B). For example, most adult lepidopterans feed on fluid resources such as nectar, decomposing animals and fruit, and dung. However, Gilbert (1972) showed that *Heliconius* butterflies collect pollen for its nutritive value, rather than as an indirect result of visits for nectar as had previously been suggested. The butterflies collect and accumulate large loads of pollen and the production of abundant saliva helps keep pollen attached to the proboscis for long periods allowing butterflies to obtain amino acids (Gilbert, 1972). Amino acids assimilated from pollen increase egg production and enable a long adult life span of up to six months (Boggs *et al.,* 1981; Gilbert, 1972; Mallet *et al.,* 1998). Also, pollen can provide nitrogen and precursors for synthesis of cyanogenic glycosides that may increase the concentration of defensive chemicals in adult butterflies (Cardoso, 2001; Nahrstedt and Davis, 1981).

Morphological studies have revealed no unique structures among the species that use pollen in their diets (Krenn *et al*., 2001; Penz and Krenn, 2000). However, there are a combination of features that assist collection and processing of pollen. For example, *Laparus*, *Neruda*, *Eueides* and *Heliconius* have the second segment of labial palpi in their proboscis cylindrical rather than club-shaped as the rest of the Heliconiina and Penz (1999) suggested that narrow labial palpi helps *Heliconius* and *Laparus* to keep pollen attached to their proboscis. Behaviour is important too and pollen-feeding species manipulate *Lantana* flowers faster and more thoroughly compared to non-pollen-feeding relatives (Krenn and Penz, 1998).

The evolution of pollen feeding in *Heliconius* has affected the evolution of their pollen sources, plants from the Cucurbitaceae family such as *Gurania* and *Psiguria* (Gilbert, 1975). For instance *Psiguria* vines develop small flowers that contain a high amount of pollen and nectar. Those flowers are produced in an

inflorescence and male flowers last only one day, and then drop off the inflorescence. Thus, male flowers are a reliable pollen source that guarantees daily visits from pollen feeding but also pollinating butterflies (Gilbert, 1975). *Heliconius* butterflies mainly visit male flowers because female flowers do not have pollen. Most *Psiguria* vines are apparently male, but infrequently when the vine is large and has plenty of resources to generate a fruit it can produce female flowers (Murawski & Gilbert 1986). *Heliconius*, while attempting to find pollen, visit female flowers, and leave pollen behind gathered during visits to other flowers; thus pollination takes place (Gilbert, 1975). The amount of pollen dispersed by *Heliconius* is greater than by hummingbirds, the other pollinator of *Psiguria* (Murawski and Gilbert, 1986).

Differences in patterns of pollen exploitation in *Heliconius* provide evidence of habitat segregation. Among comimics, Boggs (1981) and Murawski and Gilbert (1986) showed that *H. erato* and *H. hewitsoni* tend to collect smaller pollen grains such as those of *Lantana*, while the comimic sympatric species, such as *H. melpomene* and *H. pachinus* and their relatives tend to collect bigger pollen grains such as those of *Psiguria*. Between sister species, the sympatric species *H. cydno* and *H. melpomene* differed significantly in pollen load composition for three of the five most commonly collected pollen species as a result of differences in habitat preferences (Estrada and Jiggins, 2002). *H. cydno* is found in closed-canopy areas while *H. melpomene* is found in open areas, suggesting that the chances of encounter between the sister species are reduced in nature contributing to pre-mating isolation (Estrada and Jiggins, 2002). Diversification in microhabitat, larval food use and adult pollen-donor plant use, all contribute to diversification and speciation of the genus (Gilbert, 1975).

A second unusual trait of *Heliconius* is a unique mating behaviour. *Heliconius* adults have a variety of mating strategies. In one strategy, males hold territories and court females that enter these areas (Hernandez and Benson, 1998). Another is to follow ‘traplines’ between larval food plants looking for newly emerged females, then males may sit on female pupae and mating occurs while the female

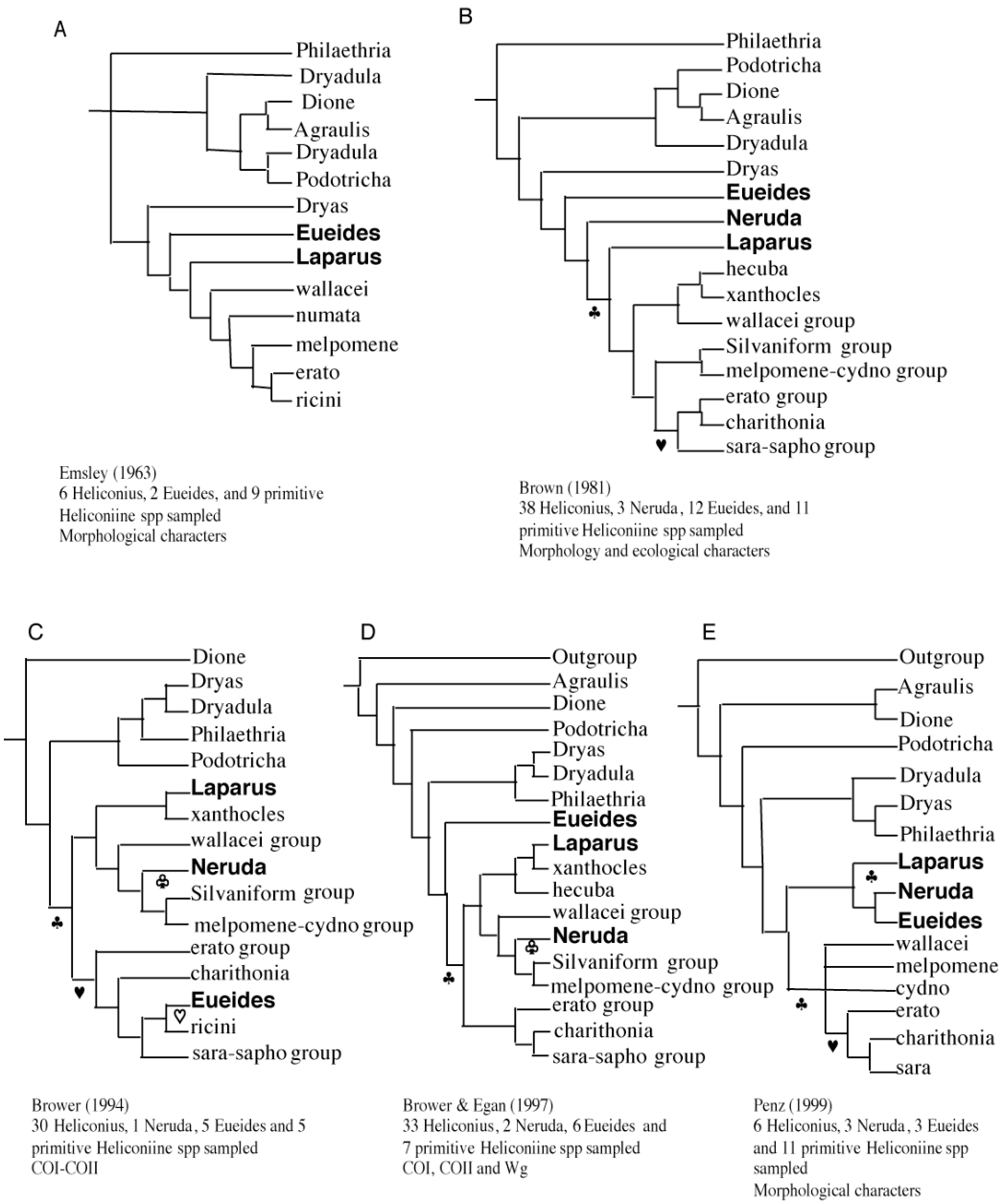
is eclosing ("pupal-mating"). Pupal-mating occurs in several insect orders (Thornhill and Alcock, 1993) and in passion-vine butterflies almost half the *Heliconius* species (42%) are pupal-maters (Gilbert, 1991). For example males of *H. hewitsoni* search the larval host plant, *Passiflora pittieri*, for female pupae, inspect these pupae regularly, then compete with each other for a position on a pupa and for access to the female as she ecloses. Mating takes place as the female begins to eclose, and females mate only once (Deinert *et al*., 1994).

Gilbert (1991) suggested that pupal-mating might play an important role in the radiation of *Heliconius* as well as in the packing of *Heliconius* species into local habitats. Pupal-mating might enhance the possibility of intrageneric mimicry because in almost all cases, each mimetic species pair consists of a pupal-mating and a non-pupal visually mating species. The strikingly different mating tactics of these groups allow phenotypically identical species to occupy the same habitats without mate recognition conflict. Second, this mating tactic may influence host-plant specialisation, because pupal-mating species may displace other heliconiines from their hosts by interference competition (Gilbert, 1991). Males of these species sit on, mate with, and disrupt eclosion of other *Heliconius* species of both mating types. This aggressive behaviour may prevent any other heliconiine species from evolving preference for host plants used by pupal-mating species.

Systematics in Heliconius butterflies

In the last sixty years seven major studies have addressed the systematics of the passion-vine butterflies or Heliconiina (Brower, 1994a; Brower and Egan, 1997; Brown, 1981; Emsley, 1963; Emsley, 1965; Michener, 1942; Penz, 1999) (Fig. 1). The current taxonomy places the "passion vine butterflies" as subtribe Heliconiina within the tribe Heliconiini. This tribe includes various other Asian genera. The tribe is placed in the nymphaline subfamily Heliconiinae, which also includes the Argynnini or fritillaries, and the Acraeini – see Lamas et al. (2004) for details.

Figure 1. Summary of major phylogenetic hypotheses for heliconiine butterflies published in the last sixty years. Note that the genus name is omitted for *Heliconius*. A. Emsley (1963). B. Brown (1981). C. Brower (1994). D. Brower and Egan (1997). E. Penz (1999). Black clubs mark pollen feeding behaviour, black hearts mark pupal-mating clades. White clubs and hearts represent loss of the trait.



Using 16 characters of adult morphology Michener, (1942) proposed a key to the identification of eight genera of passion-vine butterflies: *Heliconius*, *Eueides*, *Philaethria*, *Podotricha*, *Dryas*, *Dryadula*, *Dione* and *Agraulis*. However, Emsley

was the first to propose a phylogenetic hypothesis at a higher level for passion-vine butterflies (1963) and for the genus *Heliconius* (1965). His conclusions were based on the Michener (1942) generic revision and the addition of morphological characters considered useful for higher-level systematics in this group. He used 18 different characters such us the distribution of androconical scales on veins to establish a 'primitive' ground plan, genitalia and traits of the adult morphology (Emsley, 1963; Emsley, 1965). The result was a highly unresolved dendrogram that contained a small number of species groups and reflects morphological divergence but not necessarily the evolutionary history of the group (Fig. 1A Emsley, 1963).

In the Michener (1942) revision the genus *Heliconius* included species that are presently classified in the genera *Laparus* Billberg and *Neruda* Turner. Turner (1976) formally defined the three genera, *Neruda*, *Laparus* and *Heliconius*. *Neruda* was characterised by a distinct wing shape, particularly the broad triangular forewings with very extensive friction patches in the male, although the females have wings of more normal shape for this genus. Other unique characters of *Neruda* were lack of scoli on the head of the larva and short antennae. Turner (1976) also considered *Laparus* sufficiently distinct to be a candidate for generic rank, due to traits such as pupae that lack the gold spots and flanges and well developed antennal spines of other species. Additionally, *Laparus* has a marked colour polymorphism as an adult, is the only species with marked morphological polymorphism as a pupa, and is the only species apart from *N. metharme*, to produce blue colour not by iridescence but by laying white scales over black (Turner, 1976). Following this work, the passion-vine butterflies contained 10 recognised genera: *Philaethria*, *Podotricha*, *Dryas*, *Dryadula*, *Dione*, *Agraulis*, *Eueides*, *Laparus*, *Neruda* and *Heliconius*.

Brown (1981) made a number of changes at species level taxonomy and used characters derived from pupal morphology (Turner, 1968), adult behaviour (Gilbert, 1972), chemistry, and genetics. Brown viewed *Dryas* as closely related to *Eueides* and *Heliconius*, and *Dryadula* as closely related to *Agraulis*, *Dione*,

Philaethria and *Podotricha* following Emsley, but split *Heliconius* into four genera *Neruda* (3 species), *Laparus* (1 species), *Eueides* (12 species) and *Heliconius* (38 species) following Turner (1976) (Fig. 1B). Brown (1981) used characters to justify species groupings but did not perform any statistically based phylogenetic analysis. For example: he used *Cethosia*, the nymphaline old world genus related to new world heliconiines, to root the tree and placed *Agraulis*, *Dione*, *Podotricha*, *Dryadula*, and *Dryas* basally to the rest of *Heliconius* because, as in *Cethosia*, the wing venation of the hindwing in these groups does not form a closed discal cell. The wing patterns in this basal group possess some elements of the standard nymphaline ground plan (e.g. *Agraulis*, *Dione*) and/or mimetic patterns that can be fairly simply derived from them (e.g. *Philaethria*, *Podotricha, Dryadula* and *Dryas*) (Nijhout, 1991). The 'open cell' heliconiines are generally fast flying to avoid predation because these genera are relatively edible (Brower, 1995). Also, their highly dispersive populations are associated with open sunny habitats, where they visit unspecialised butterfly-pollinated flowers with short corollas and large floral displays (e.g. *Lantana*) (Gilbert, 1991).

The remaining group called by Brown the 'advanced genera' is the most diverse in terms of numbers of species, and all four genera share a closed discal cell. Their wing pattern differs more from the general nymphaline ground plan by great simplification and loss of many elements, as well as by the appearance of several novel mimetic patterns (Nijhout, 1991). The 'closed-cell' genera, *Eueides*, *Neruda*, *Heliconius* and *Laparus* are relatively unpalatable, aposematic, slow flying, and *Heliconius* and *Laparus* feed on pollen from specialised butterfly-pollinated flowers such us *Psiguria* (Gilbert, 1991). Within *Heliconius*, Brown used the absence of a signa on the female bursa copulatrix as a character to define the pupal-mating group (*erato*+*sara*/*sapho* group in Fig. 1B).

Recent contributions (Brower, 1994a; Brower and Egan, 1997; Penz, 1999) have proposed new phylogenetic hypotheses for passion-vine butterflies. All these analyses employed only parsimony or weighted parsimony analysis. Brower (1994) presented a cladogram based on successive approximations weighting for

35 species of *Heliconius* and the related genera *Eueides*, *Laparus*, and *Neruda*, based on mitochondrial DNA (mtDNA) sequences from cytochrome oxidase subunits I and II (950 bp). This study focused on the relationships between these 'advanced' genera and the five heliconiine genera considered primitive to this clade, *Dione*, *Podotricha*, *Philaethria*, *Dryadula*, and *Dryas* (Fig. 1C). The data supported the most traditionally recognised species groups and also the monophyly of the four closed-cell genera with respect to other heliconiine outgroups. However, in Brower's phylogeny *Heliconius* was made paraphyletic by the internal placement of *Eueides*, *Laparus* and *Neruda*. Most surprisingly *Eueides* was nested within (Brown's, 1981) *Heliconius* pupal-mating group.

Three years later Brower and Egan (1997) revised the position of *Eueides* after adding to the mtDNA data a nuclear protein-coding gene *wingless* (*wg*, 375 bp). This study also added 7 new taxa including *Agraulis* to complete the ten recognised passion-vine butterfly genera and used *Speyeria* (from the heliconiine tribe Argynnini) as an outgroup to assess relationships among basal Heliconiina. Separate analysis of the COI-COII sequences showed *Eueides* still nested within the pupal-mating *Heliconius* clade and *wg* provided a poorly resolved picture of the relationships. Neither of these two gene regions alone supported the monophyly of *Heliconius* with respect to *Eueides*. Nonetheless, simultaneous parsimony analysis of the two regions together supported a topology largely in agreement with traditional views of heliconiine relationships based on morphology where *Eueides* is basal to *Heliconius*, *Neruda* and *Laparus*. However, *Heliconius* remained paraphyletic because *Neruda* and *Laparus* still branched internally to the genus (Fig. 1D). These results suggested that pollen-feeding behaviour evolved in the common ancestor of *Laparus* and *Heliconius* and was subsequently lost in an ancestor of *Neruda*.

Finally, Penz (1999) proposed a higher-level phylogeny for the passion-vine butterflies based on 146 morphological characters from early stages and adults. In this study 24 species were analysed representing the ten currently accepted genera of Heliconiina. The phylogeny derived from the combined analysis of character

sets gathered from different life stages supported the monophyly of all genera but differed in topology from all previously proposed hypotheses (Fig. 1E). At the basal level the results were congruent with the molecular hypothesis (Brower, 1994a; Brower and Egan, 1997), but the relationships among *Laparus*, *Eueides* and *Neruda* differed significantly. *Heliconius* was monophyletic with respect to *Laparus*, *Eueides*, and *Neruda*, supported by three pupal morphology characters. Penz (1999) suggested that pollen feeding behaviour either evolved independently in *Laparus* and the ancestor of *Heliconius*, or it evolved in the common ancestor of genera *Laparus*, *Neruda*, *Eueides*, and *Heliconius* and was subsequently lost by the ancestor of *Neruda* and *Eueides*.

Conflict between phylogenies

Three analytical attributes might contribute to the conflict among recent Heliconiina phylogenies: Taxon sampling, number of informative characters and methods of tree reconstruction (Brower *et al.*, 1996).

Taxon sampling

Many phylogenetic analyses, particularly morphological studies, use higher taxa (e.g. genera, families) rather than species as terminal taxa. However, sampling only a single species per higher taxon may result in low accuracy of phylogenetic reconstruction. A recent study reviewed different parsimony methods for coding and sampling higher taxa and compared their relative accuracies using computer simulations, and showed that using species as terminal taxa is the most accurate under almost all conditions, and is often superior to the other methods by a large margin (Wiens, 1998). Therefore, broad species sampling is a positive aspect of the DNA analysis by Brower (1994) and Brower and Egan (1997), in contrast with the morphological analysis of Penz (1999) where just one species per genus was used.

Number of informative characters

In molecular systematics the estimation of phylogenies can benefit from the analysis of multiple genetic data sets and the combination of data sets that evolve

at different rates in order to clarify relationships at different levels of phylogeny (Huelsenbeck *et al*., 2001). Brower and Egan (1997) clarified the position of *Eueides*, the traditionally recognised sister genus of *Heliconius*, based on mtDNA sequence data combined with a region of the nuclear gene *wg*. The two mtDNA genes COI and COII evolve rapidly (Brown *et al*., 1979) and are already known to be informative in studies of divergence ranging from intra-specific biogeography of races (Brower, 1994b; Huelsenbeck and Ronquist, 2001) to relationships among tribes and subfamilies (Stern, 1994). The nuclear gene *wg* has been suggested as a useful source of characters for phylogenetic studies of butterflies, and perhaps other insect taxa, with divergence times up to 60 million years ago (Brower and DeSalle, 1998). Also, due to its uniform base composition *wg* appears to become saturated more slowly than mtDNA. However, the number of characters informative for the basal branches of the Heliconiina provided by the two genes is still low, due to saturation at the third position in COI and COII (Brower, 1996a) and short *wg* sequences (375bp) that contain little variation to resolve relationships.

Methods of tree reconstruction

Incongruence between mtDNA trees and nuclear gene trees have been reported before and may arise because of homoplasy in the data, differences in analytical and methodological procedures used to build the phylogenies, or simple statistical error (Brower *et al*., 1996); they may also be due to differences in gene history between genes (e.g. Beltran et al. 2002), although this is generally less likely at the relatively high phylogenetic levels studied here.

There are available four primary methods for reconstructing phylogenies from protein and nucleic acid sequence alignments: distance-based methods, or phenetics (e.g. neighbour joining), maximum parsimony (MP), maximum likelihood (ML), and Bayesian methods (BAY). In distance methods, the character data from aligned sequences are converted into a matrix of pairwise genetic distances between the sequences, while ML, MP and BAY use the alignment directly by comparing characters at each nucleotide or amino acid site.

These methods are called character-based. MP is based on the view that the optimum tree requires the fewest number of changes to explain the data in the alignment. ML finds an optimal phylogenetic hypothesis by finding the tree that maximises the probability of observing the data given the tree (hypothesis). Finally, BAY analysis is based on the notion of posterior probability distribution of trees, which is the probability of a tree conditioned on the observations (Hall, 2001). The posterior probability is obtained by combining the prior probability (all the trees are considered equally probable) and likelihood for each tree using Bayes theorem (Huelsenbeck and Ronquist, 2001). ML and BAY are similar as both analyses postulate a model of evolution and the program searches for the most likely trees and models that are consistent with the data. However, BAY differs somewhat from ML, ML seeks the tree that maximises the likelihood of observing data given the model of evolution and tree hypotheses, while in BAY analysis, the prior probability becomes a factor, and in certain circumstances may choose trees that are non-optimal from the point of view of the likelihood alone. Additionally, BAY, as implemented by the MrBayes, uses Markov chain Monte Carlo (MCMC) to do the likelihood optimization (Huelsenbeck and Ronquist, 2001), whereas likelihood analyses re-evaluate the full likelihood for every tree selected via branch swapping; this makes MrBayes analysis much faster for larger datasets than for example PAUP doing a likelihood search.

The parsimony approach has been extensively used in studies of molecular and morphological character evolution (Sober, 1991). However, MP will always underestimate the number of substitutions, especially where saturation is a problem as in basal branches of mtDNA phylogenies and information can by lost by assuming an incorrect model of substitutions (Routman *et al*., 1994). ML is advantageous because it allows the evolutionary model for a particular data set to be estimated and used in phylogeny reconstruction, and allows for the multiple substitutions that lead to saturation, but it is time consuming. BAY using the MCMC method is faster and is gaining acceptance in phylogenetics (Huelsenbeck *et al*., 2001). MP has been the only methodology used previously to reconstruct phylogenies for *Heliconius*. Thus, using other methodologies, such as ML and

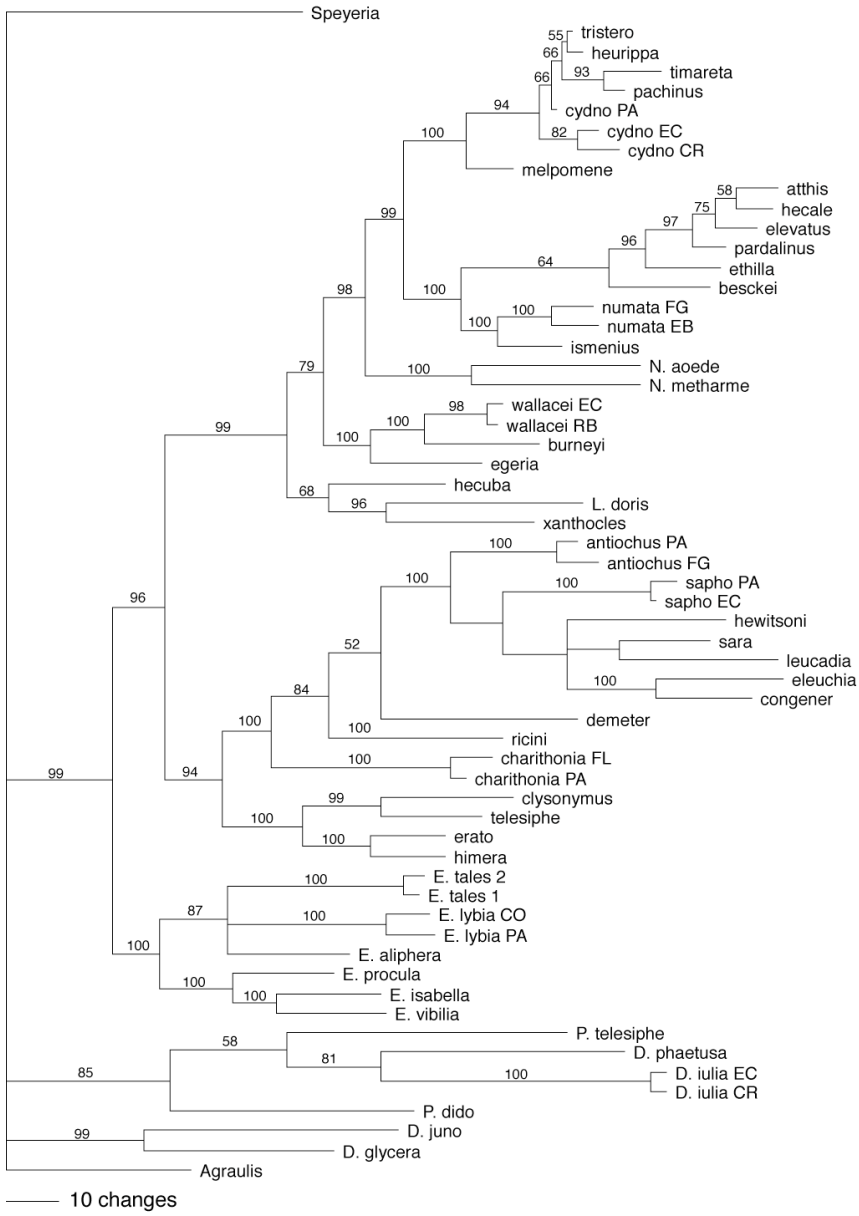
BAY should strengthen inferences proposed for the group. If sequences have a strong phylogenetic signal, different methods will produce the same phylogenetic tree.

Brower's (1994) parsimony analysis placed *Eueides* as sister to *H. charithonia* within the *erato-sara/sapho* group of *Heliconius*, so I decided to repeat this analysis using the more modern methodologies. I used Brower's (1994) mtDNA data to calculate a ML tree based on general-time-reversible time model of nucleotide substitution (GTR+G+I) (Yang, 1994) suggested as the best fit for these data by likelihood (Posada and Crandall, 1998). The ML tree showed *Eueides* basal to *Heliconius*. To confirm this pattern I ran a BAY analysis (Huelsenbeck and Ronquist, 2001), and the tree showed strong support for placing *Eueides* basal to *Heliconius* (Fig. 2). Additionally, I applied the method of Shimodaira and Hasegawa (Shimodaira and Hasegawa, 1999) to compare different tree topologies. I compared Brower's (1994) parsimony tree (*Eueides, Laparus* and *Neruda* as part of *Heliconius*) with four different scenarios. First, *Heliconius* was forced to be monophyletic by placing *Eueides*, *Laparus*, and *Neruda* basally; then three topologies were tried in which *Heliconius* was paraphyletic but with *Laparus*, *Neruda* or *Eueides* basal. Brower's (1994) and the last three topologies where significantly worse ($p<0.001$) compared with the scenario of *Eueides* basal to *Heliconius* and *Laparus* and *Neruda* as part of *Heliconius*. These results suggest that different methodologies are giving rise to these different phylogenetic signals. Even before the use of *wg*, Brower's data did not provide strong support for his placement of *Eueides*; instead it appears to have been an artefact of successively weighted parsimony analysis.

In summary, to establish a useful phylogenetic hypothesis for the heliconiines it would be helpful to add more taxa, more genes, and compare different methods of tree reconstruction. Once a hypothesis is well supported it will be possible to make better inferences about the biology of the group. For this reason, the principal goal of my study was to construct a species level phylogeny using mitochondrial DNA and exons of nuclear genes. This hypothesis will be used to

answer the following questions. Is *Heliconius* monophyletic? How many times has pollen-feeding arisen in the *Heliconius* group? What are the relationships within major clades of *Heliconius*? Is the pupal-mating group monophyletic?

Figure 2. Phylogenetic hypothesis for heliconiine species based on Brower's (1994) sequences (COI and COII) with Bayesian analysis (BAY). Values above branches show BAY probabilities.



Material and methods

Sampling methods

I sampled 122 individual butterflies, representing 38 *Heliconius*, 10 *Eueides*, and 10 outgroup species. According to the Lamas *et al*. (2004) classification, only 10 species of Heliconiina are missing from this study: 4 species of *Heliconius* (*H. astraea*, *H. lalitae*, *H. tristero* and *H. luciana*), 1 *Neruda* (*N. godmani*), 1 rare *Eueides* (*E. emsleyi*), and 5 outgroup heliconiines (*Podotricha judith*, *Philaethria constantinoi*, *P. ostara*, *P. pygmalion*, and *P. wernickei*). Thirty-seven individuals used in this study were the same as those used in Beltran et al. 2002 (Table 1 Appendix 3.1). Additional species and additional individuals of previously sampled species are listed in table 1. To evaluate relationships between basal Heliconiina, I included *Castilia perilla* (Nymphalidae: Nymphalinae: Melitaeini: *Phyciodina*) as outgroup. Butterflies were collected, preserved in liquid nitrogen and are stored in the Smithsonian Tropical Research Institute in Panama. Wings of voucher specimens are preserved in glassine envelopes; pictures are available at http://neruda.cap.ed.ac.uk/phylogeny. From each individual, 1/6 of the thorax was used and the genomic DNA was extracted using the DNeasy Kit (Qiagen) following the manufacturer's recommended protocols. Additional samples were obtained as DNA aliquots from Carla Penz and Andrew Brower (Table 1).

Molecular regions and sequencing methods

Mitochondrial DNA

Two mitochondrial DNA regions were used: first, a region of cytochrome oxidase (CO), spanning cytochrome oxidase I (COI), the gene for leucine transfer RNA (tRNA-leu), cytochrome oxidase II (COI and COII); and second, the region coding for 16S ribosomal RNA (16S). Both regions have been used to explore phylogenetic relationships in insects (e.g. Brower 1994, Smith *et al*., 2002). Two different sampling strategies were followed, for CO at least 2 individuals per species were sequenced and for 16S just 12 individuals were sequenced in order

to check the relationships within *Heliconius* (811 *H. m. rosina*, 346 *H. numata*, 8560 *H. burneyi*, 8549 *H. hecuba*, 846 *L. doris*, 8569 *N. aoede*, 440 *H. e. hydara*, 8037 *H. clysonimus*, 842 *H. eleuchia*, 8562 *H. demeter*, 320 *E. vibilia*, and 293 *D. iulia*).

Table 1. Heliconiina included in the study. ID numbers are STRI collection numbers and should be prefixed by 'STRI-B-'except ^= collection at UCL, *=Carla Penz collection, £= Mauricio Linares and +=Andrew Brower collection. French Guiana is abbreviated to F. Guiana.

GROUP	ID No.	Species	Location	GROUP	ID No.	Name	Location
melpomene/ cydno	8023	*H. pachinus*	Panama	**"primitive" *Heliconius***	8561	*H. burneyi*	Peru
	40	*H. heurippa*	Colombia		G-31-2+	*H. egeria*	
	8520	*H. timareta*	Ecuador		8549	*H. hecuba*	Ecuador
	8521	*H. timareta*	Ecuador		8550	*H. hecuba*	Ecuador
	M-26£	*H. m. mocoa*	Colombia		8609	*H. xanthocles*	Peru
	M-82£	*H. m. mocoa*	Colombia		8610	*H. xanthocles*	Peru
	8158	*H. m. malleti*	Ecuador	***Laparus***	847	*L. doris*	Panama
	C-1*	*H. besckei*	Brazil		8645	*L. doris*	Peru
	B-3-2+	*H. besckei*	Brazil	***Neruda***	2949	*N. metharme*	Ecuador
silvaniforms	8130	*H. numata*	Ecuador		RB257+	*N. metharme*	
	434	*H. elevatus*	F. Guiana		8569	*N. aoede*	Peru
	608	*H. hecale*	Panama		8198	*N. aoede*	F. Guiana
	666	*H. ismenius*	Panama		8570	*N. aoede*	Peru
	656	*H. ismenius*	Panama	***Eueides***	8699	*E. vibilia*	Peru
	8127	*H. atthis*	Ecuador		875	*E. aliphera*	Ecuador
	8162	*H. atthis*	Ecuador		8595	*E. lybia*	Peru
	JM-23-1+	*H. pardalinus*			780	*E. isabella*	Panama
	8014	*H. ethilla*	Argentina		8740	*E. isabella*	Peru
	8729	*H. ethilla*	Peru		2933	*E. procula*	Ecuador
	C-2*	*H. nattereri*	Brazil		C-7*	*E. procula*	
erato/ himera	8024	*H. clysonimus*	Panama		2948	*E. tales*	Ecuador
	856	*H. hecalesia*	Panama		C-8*	*E. tales*	Ecuador
	2928	*H. telesiphe*	Ecuador		C-9*	*E. pavana*	Brazil
	8525	*H. telesiphe*	Ecuador		8750	*E. lampeto*	Peru
	9111	*H. hortense*	Mexico		8748	*E. heliconioides*	Peru
charithonia	EC-3971^	*H. charithonia*	Ecuador		8747	*E. heliconioides*	Peru
	EC-190^	*H. peruvianus*	Ecuador	***Dryadula***	2921	*D. phaetusa*	Panama
	EC-288^	*H. peruvianus*	Ecuador	***Dryas***	2924	*D. iulia*	Ecuador
	C-3*	*H. hermathena*	Amazon	***Philaethria***	8700	*P. dido*	Peru
sara/sapho	308	*H. sara*	F. Guiana		690	*P. dido*	Panama
	8197	*H. ricini*	F. Guiana		8749	*P. diatonica*	Peru
	857	*H. eleuchia*	Ecuador	***Podotricha***	2929	*P. telesiphe*	Ecuador
	2739	*H. sapho*	Panama		8146	*P. telesiphe*	Ecuador
	P-32-6+	*H. antiochus*	Panama	***Dione***	2939	*D. juno*	Ecuador
	C-4*	*H. hewitsoni*			2970	*D. juno*	Panama
	C-6*	*H. hewitsoni*			8727	*D. juno*	Peru
	C19-1+	*H. congener*			C-10*	*D. glycera*	
	RB119+	*H. leucadia*			Pe57+	*D. glycera*	
	8562	*H. demeter*	Peru		8748	*D. moneta*	Peru
	8563	*H. demeter*	Peru		Mex2-1+	*D. moneta*	
"primitive" *Heliconius*	8154	*H. hierax*	Ecuador	***Agraulis***	8554	*A. vanillae*	Panama
	8147	*H. hierax*	Ecuador		8725	*A. vanillae*	Peru
	286	*H. wallacei*	F. Guiana	**Outgroup**	8632	*C. perilla*	Peru
	8212	*H. wallacei*					
	8560	*H. burneyi*	Peru				

The mitochondrial CO region was amplified using primers and protocols described by Beltran *et al*. 2002 (Appendix 3.1). A *Drosophila yakuba* sequence

(GenBank accession number X03240) was used as a reference. The clean template obtained was sequenced in a 10μl cycle sequence reaction mixture containing 1μl BigDye, 0.3x buffer, 2mM primer, and 2μl of template. The cycle profile was 96 ºC for 30 sec., then 96ºC for 10 sec., 50ºC for 15 sec, 60ºC for 4 min. for 30 cycles. This product was cleaned by precipitation using 37.5μl of 70% EtOH and 0.5mM $MgCl_2$. The samples were re-suspended in 4μl of a 5:0.12 deionized formamide: crystal violet solution, denatured at 85ºC for 2 min. and loaded into 5.5% acrylamide gels. Gels were run on BaseStation (MJ Research) for 3 hours.

The additional mitochondrial region used was the 16S ribosomal RNA gene. This region was amplified using 16Sar1 5'-CCC GCC TGT TTA TCA AAA ACA T-3' and Ins16Sar 5'-CCC TCC GGT TTG AAC TCA GAT C -3'. Primers were obtained by modifying Palumbi (1996) to improve amplification in Lepidoptera. The identity of this region was confirmed by comparison with *Eresia burchellii* (GenBank accession AF186861). Double-stranded DNA was synthesised in 10μl reactions containing 2μl of genomic DNA, 1x buffer, 1mM $MgCl_2$, 0.8mM dNTPs, 0.5mM of each primer, and 0.05 u/μl of Qiagen Taq polymerase. DNA was amplified using the following step-cycle profile: 94ºC for 5 min., 94ºC for 30 sec., 55ºC for 30 sec., 72ºC for 1 min. for 34 cycles. These products were sequenced as described for COI and COII.

Nuclear loci

Four nuclear loci were used, *elongation factor*-1α (*Ef1α*), *apterous* (*ap*), *decapentaplegic* (*dpp*) and *wingless* (*wg*). *Ef1α* is a key factor in protein synthesis playing a central role in protein chain elongation (Bischoff *et al*., 2002). This gene has been using in many phylogenetic studies and results showed informativeness of synonymous nucleotide substitutions up to divergences of 60 million years (Cardoso, 2001; Cho *et al*., 1995; Mitchell *et al*., 1997; Reed and Sperling, 1999). The genes *ap* and *dpp*, are involved in wing development in *Drosophila* and were isolated in *Heliconius* by the Owen Mcmillan laboratory in Puerto Rico (Tobler *et al*., 2003), but there is no report of their phylogenetic utility. The sampling for *Ef1α* was the same as COI and II, for *ap* and *dpp* was the same that 16S (just 12

individuals representing the major clades in *Heliconius*). In addition, *wg* sequences were included in the analysis although not for the same individuals. Sequences of *wg* were loaded from Brower's GenBank accessions AY090135, UO08554, AF014126 TO AF014135 and AF169869 to AF169921).

The *elongation factor 1α* (*Ef1α*) region was initially amplified and sequenced from genomic DNA using a mix of primers from *Papilio* (*Ef1*-5) (Reed and Sperling, 1999) and bumble bees (F2-rev) (Walldorf and Hovemann, 1990). The primers were situated at position 15 (*Ef1*-5) and 955 (F2-rev) of *Papilio glaucus* (GenBank accession AF044826). Then, initial *Heliconius* sequences were aligned and *Heliconius* specific primers were designed to amplify the region consistently using genomic DNA extracts. The specific primers designed were *Ef1*-H-f 5'-GAG AAG GAA GCC CAG GAA AT-3' and *Ef1*-H-r 5'- CCT TGA CRG ACA CGT TCT TT-3'. DNA was amplified using the step cycle profile described for 16S and sequenced as for the mitochondrial region.

The other two nuclear genes sequenced were *apterous* (*ap*) and *decapentaplegic* (*dpp*). The gene *apterous* was amplified using primers *ap*-f35 5'-TGA ATC CTG AAT ACC TGG AGA-3' and *ap*-r224 5'-GGA ACC ATA CCT GTA AAA CCC-3' and *decapentaplegic* using *dpp*-f34 5'-AGA GAA CGT GGC GAG ACA CTG-3' and *dpp*-r327 5'-GAG GAA AGT TGC GTA GGA ACG-3' (Tobler *et al*., 2003). The identities of the regions were verified by contrast with *Precis coenia* GenBank accession numbers L42140 and L42141 respectively. The products from *ap* and *dpp* were sequenced as a described before.

Alignment and phylogenetic analyses

Chromatograms were edited and base calls checked using SEQUENCHER 4.1 (Gene Codes Corporation, Inc). The protein-coding mtDNA and nuclear DNA sequences were checked for reading-frame errors and unexpected stop codons by translating the nucleotide sequences to peptides using MacClade 4.0 (Maddison and Maddison, 1997). Maximum likelihood models of sequence evolution for each gene were estimated using ModelTest 3.04 (Posada and Crandall, 1998).

Bayesian analysis (BAY) run in MrBayes (Huelsenbeck and Ronquist, 2001), was used to infer the phylogeny based on the best-fit model selected by ModelTest with one million generations, a sample frequency of 100 and 2% of the trees burned (by generation 2000 the likelihoods of the trees converged on a stable value, and trees generated before this were excluded to calculate the consensus tree). For comparison, maximum parsimony (MP) trees were obtained using PAUP* version 4.0b8 (Swofford, 2000) in a heuristic search with TBR branch swapping. The consensus tree was calculated using majority rule. Bootstrap (1000 replicates, heuristic search with TBR branch swapping) was used to assess confidence in each node.

The Incongruence for Length Difference test (ILD; Farris *et al*., 1994) implemented by PAUP was used to test incongruence between the different partitions, (e.g. COI/COII vs. *Ef1α*; mtDNA (COI, tRNA-leu, COII, 16S) vs. nuclear (*Ef1α*, *ap*, *dpp*, *wg*); COI vs. *ap*; *Ef1α* vs. *ap*; etc). This test was applied to a matrix including the twelve individuals sequenced for COI, COII, *Ef1α*, *ap*, *dpp* adding *wg* sequences of GenBank for these individuals. Additionally, to test specific hypotheses alternative *a priori* scenarios were compared using the method of Shimodaira and Hasegawa (Goldman *et al*., 2000; Shimodaira and Hasegawa, 1999) and implemented using PAUP* version 4.0b8. For each genus (i.e. *Heliconius, Laparus, Neruda, Eueides*), maximum likelihood for two or three topologies was compared in the same test. To seek each scenario, the topology shown in Figure 4 was modified using MacClade (Maddison and Maddison, 1997).

Results

Characterisation of the nucleotide data

The final nucleotide data set contained 3834 positions (2119 mitochondrial, 1716 nuclear), translating to 1083 amino acids (511 mitochondrial, 572 nuclear). The individual sequences are available at GenBank (accession numbers to be obtained) and the alignment of full data is available on request.

Mitochondrial DNA

For mitochondrial DNA (mtDNA) 1611bp were obtained from the CO region including nucleotides and gaps. These represent 822 bp of COI corresponding to position 2191 to 3009 of the *D. yakuba* sequence (X03240), the complete tRNA-leu gene (78bp) and 711 bp representing the entire COII coding sequence, matching positions 3012-3077 and 3083-3766 in *D. yakuba* respectively. For 16S ribosomal RNA 512bp were amplified corresponding to positions 26 to 541 in *Eresia burchellii* (AF186861). Length variation was concentrated in tRNA-leu and in 16S. At the beginning of tRNA-leu an insertion of 12bp was found in one individual of *H. demeter* (STRI-B-8563) while 7bp of the same insertion was shared by *H. charithonia*, *H. peruvianus*, *H. ricini* and the second individual of *H. demeter* (STRI-B- 8562). Another 3bp insertion was observed at position 71 in *H. ismenius*. In 16S 29 gaps were found at positions 51 to 63, 241 to 280, and 337 to 351. Additionally, codon deletions were found. In COI the third codon of the alignment, corresponding to amino acid position #243 in *D. yakuba,* X03240), was deleted in some *Eueides* species (*E. lineata, E. vibilia*, *E. lybia*, *E. aliphera*, *E. isabella*, and *E. tales*). There was another codon deletion in *H. ismenius* just before the COI stop codon. In COII three closely adjacent codon deletions were observed at amino acid position #126 in *Dryadula phaetusa*, #127 in *H. sara* and at position #129 in *Dryas iulia*.

Of the 2123 nucleotide sites examined for mtDNA, 755 (40%) were variable (Table 2). Almost all the variation occurred in the protein-coding region (641 were variable in COI and COII). As expected the variation per position was higher at third positions (70%) (Table 2). In total for mtDNA 30% of the sites were informative.

Nuclear DNA

The nuclear genes *Ef1*α 876bp, *ap* 195 bp and *dpp* 270bp were aligned with *Papilio glaucus* (GenBank accession AF044826) at positions 50 to 925, *Precis coenia* (L42140) at positions 193 to 387 and *Precis coenia* (LA42141) at positions 145 to 414, respectively. Only *dpp* showed length variation with respect

to the reference sequence, a codon deletion at position 196 of *Precis coenia* (LA42141) was observed in *H. cydno chioneus*, *H. numata* and *H. burneyi*.

Table 2. Nucleotide variability over genes and codon position. The values were calculated for the whole data set in CO (COI and COII) and *Ef1α* and just for the 12 species for the remaining genes. CI = consistency index. RI = retention index.

CO	**All sites**	**Codon pos. 1**	**Codon pos. 2**	**Codon pos. 3**
No. of characters	1611	511	511	511
No. of invariant	955	375	454	66
No. variable	656	136	57	445
No. informatives	587	106	32	417
Tree length	3751	460	109	3086
CI	0.262	0.361	0.596	0.236
RI	0.715	0.797	0.799	0.704
Ef1α	**All sites**	**Codon pos. 1**	**Codon pos. 2**	**Codon pos. 3**
No. of characters	876	292	292	292
No. of invariant	615	259	271	84
No. variable	261	33	20	208
No. informatives	186	12	5	169
Tree length	734	53	34	647
CI	0.47	0.66	0.67	0.44
RI	0.83	0.83	0.6	0.83
Gene	**16S**	***ap***	***dpp***	***wg***
No. of characters	512	195	270	375
No. of invariant	413	163	211	281
No. variable	99	32	59	94
No. informatives	38	15	27	34
Tree length	155	54	86	157
CI	0.748	0.722	0.837	0.669
RI	0.426	0.423	0.745	0.212

There were similar levels of variability across nuclear genes and codon positions (Table 2). *Ef1α* showed 21% of sites variable and most of the variation was at third positions (79%), while *ap*, *dpp*, and *wg* showed similar levels of variation overall (16%, 21% and 25% respectively) (Table 2), and for *ap*, *dpp*, and *wg* at third positions (85%, 77% and 61% respectively). Also, the relative divergence of first, second and third position were similar among nuclear and mtDNA (Table 2).

Among the 12 species sampled for *dpp*, *ap*, *wg*, and 16S divergence ranged from 0.1% between sister groups (*H. m. rosina* vs. *H. numata*) to 26% between

Heliconius species and outgroups (*H. m. rosina* vs. *D. iulia*). The complete data for CO and *Ef1α* showed divergences ranging from 0.3% (between sister groups) to 39% (between *Heliconius* species and outgroups) and 0.3% to 32% respectively. The lowest divergence was at *Ef1α* second position, which had a maximum pairwise divergence of <3% within Heliconiina, whereas comparing nuclear and mtDNA regions nearly all third positions exhibited divergences >10%. Additionally, divergences at the first two codon positions were higher in the mitochondrial data than in *Ef1α*, indicating greater rates of protein evolution as well as nucleotide evolution.

Table 3. Best supported models of molecular evolution and estimated parameter values for the different data sets.

Data set	**CO**	**16S**	***Ef1α***	***ap***	***dpp***	***wg***
Model	GTR+I+G	F81+G	GTR+I+G	K2P+G	K2P+G	TrNef+G
Base frequencies						
A	0.374	0.4421	0.28	0.25	0.25	0.25
C	0.1081	0.0649	0.244	0.25	0.25	0.25
G	0.0647	0.1234	0.2447	0.25	0.25	0.25
T	0.4532	0.3697	0.2313	0.25	0.25	0.25
Substitution model		All equal rates				
Tr/tv ratio				1.9162	2.1976	
Tr [A-G]	15.2471		6.2591			6.2038
[C-T]	27.3268		13.4501			12.7904
Tv [A-C]	2.9031		1.6729			1
[A-T]	1.7001		3.3225			1
[C-G]	2.8548		1.7038			1
[G-T]	1		1			1
Invariable sites	0.5001	0	0.543	0	0	0
Gamma parameter	0.5187	0.1236	0.8076	0.0639	0.2419	0.3561

Models of sequence evolution

There was some variation between models of sequence evolution selected for each region (Table 3). Evolution at CO and *Ef1α* were best explained by the six-parameter general time reversible model of nucleotide substitution (GTR) (Yang, 1994). For 16S and the rest of nuclear genes, simpler models fitted the data (Table 3): TrNef with just three substitution types (Tamura and Nei, 1993), F81 with equal base frequencies (Felsenstein, 1981) and K2P with a single substitution type (Kimura, 1980). It is perhaps unsurprising that smaller data matrices supported

simpler models of evolution. As is usual for insects the mitochondrial DNA showed a strong AT bias (Caterino and Sperling, 1999), AT content for CO was 82% and 16S 81% compared with nuclear DNA (50%) (Table 3). Also, transitions were almost ten times more frequent than transversions in CO, *Ef1*, and *wg*, while other nuclear and mitochondrial regions did not have significant transition: transversion rate bias, probably due to the lower sample sizes (Table 3). The gamma parameter varied between 0.06 and 0.80 (Table 3). Only in CO and *Ef1α* was a significant non-zero parameter for proportion of invariable sites (0.50 and 0.54 respectively) suggesting some site-specific constraints on sequence evolution (Table 3).

Congruence test

ILD tests between mitochondrial data (COI+tRNA-leu+COII+16S) vs. nuclear data (*Ef1α+ap+dpp+wg*) provided no evidence for incongruence between partitions based on nucleotides (p=0.08), nor between amino acid sequence partitions (p=0.23). The lack of incongruence between amino acid partitions results in part from low resolution in the nuclear partition because only 13 amino acids were informative. Comparisons within mtDNA did not show any incongruence either (e.g. COI vs. COII p=0.18 and COI+COII vs. 16S p=0.40), and neither did mtDNA vs. individual partitions of nuclear genes. Within nuclear genes only one comparison showed significant incongruence, *Ef1α* vs. *wg* (p=0.01) and it was the only significant test out of 18 comparisons in total. Therefore, there was no strong evidence for significant incongruence between data sets and it seems reasonable to use the total data to calculate a combined evidence phylogenetic hypothesis.

Phylogenetic analyses

Topologies for individual data sets are shown in Figure 3 (A. mtDNA, B. *Ef1α*, and C. *ap*, *dpp* and *wg*) and the combined hypothesis using all genes described before is shown in Figure 4. Phylogenetic resolution was somewhat weaker at nuclear loci compared with the mtDNA, for example mtDNA and *Ef1α* showed a monophyletic clade that included the sister clades *cydno-melpomene* and the

silvaniforms, but in *Ef1α* there was no resolution of species relationships within that clade (Fig. 3A and B). However resolution increased for clades in which species are more distantly related such as *sara*/*sapho* and *erato*.

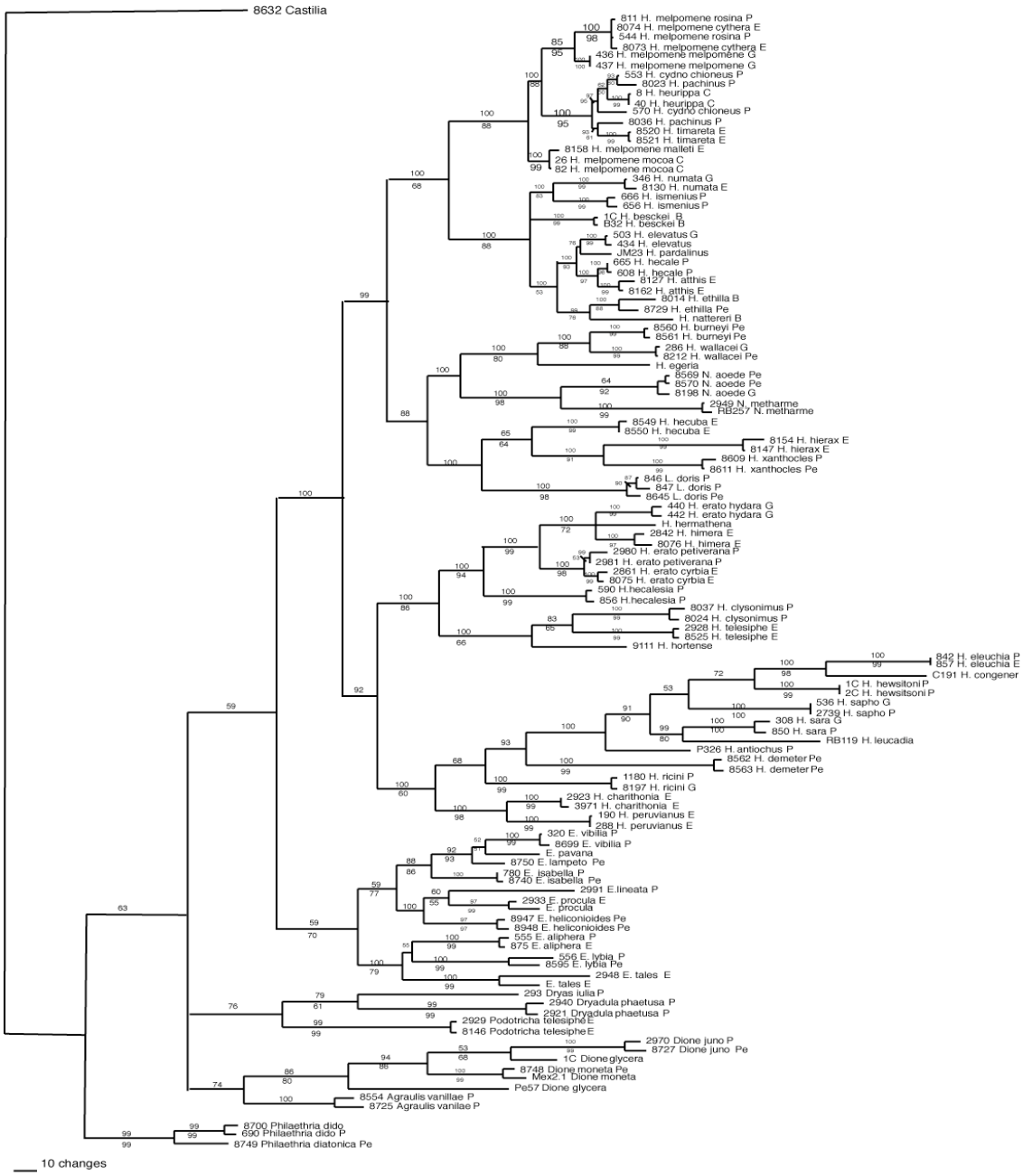
The data also produced well- resolved relationships among genera (Fig. 4). *Heliconius*, *Laparus* and *Neruda* formed a well-supported monophyletic clade (100% with MrBayes) with *Eueides* basal to this group, in agreement with traditional relationships (Brown, 1981). *Laparus* fell in a well-supported clade with *H. hierax*, *H. wallacei*, and *H. hecuba* with 100% Bayesian probability. Also, *Neruda* fell within *Heliconius* closely related to *cydno*/*melpomene* and the silvaniform group, with 95% Bayesian support. Various alternative topologies were compared using SH tests (Table 4). The results showed that a tree constrained to have both *L. doris* and *Neruda* basal to *Heliconius* was a significantly worse fit both for mtDNA and nuclear DNA alone and also for combined evidence ($p<0.001$). When *Laparus* was forced to be basal to *Heliconius* the resulting tree was a worse fit to the data based on combined evidence ($p=0.039$). In contrast, even combined evidence could not exclude the possibility that *Neruda* was basal to *Heliconius* (mtDNA $p= 0.507$, nuclear $p=0.365$, combined $p=0.329$) (Table 4). This might be a result of the different placements suggested by different genes: mtDNA placed *Neruda* inside the *Heliconius* primitive group (Fig 3a), while nuclear data from *Ef1α*, *ap*, *dpp*, and *wg* placed *Neruda* basal to the silvaniforms + *cydno*/*melpomene* group (Fig. 3B and C). The only hypothesis that could be ruled out is for *Neruda* to be part of the pupal-mating clade (mtDNA $p=0.234$, nuclear $p=0.007$, combined $p=0.008$) (Table 4).

Table 4. Results of SH tests of alternate *a priori* hypotheses for phylogenetic relationships between *Heliconius*, *Laparus doris*, *Neruda* and *Eueides*. CI=consistency index. RI=retention index.

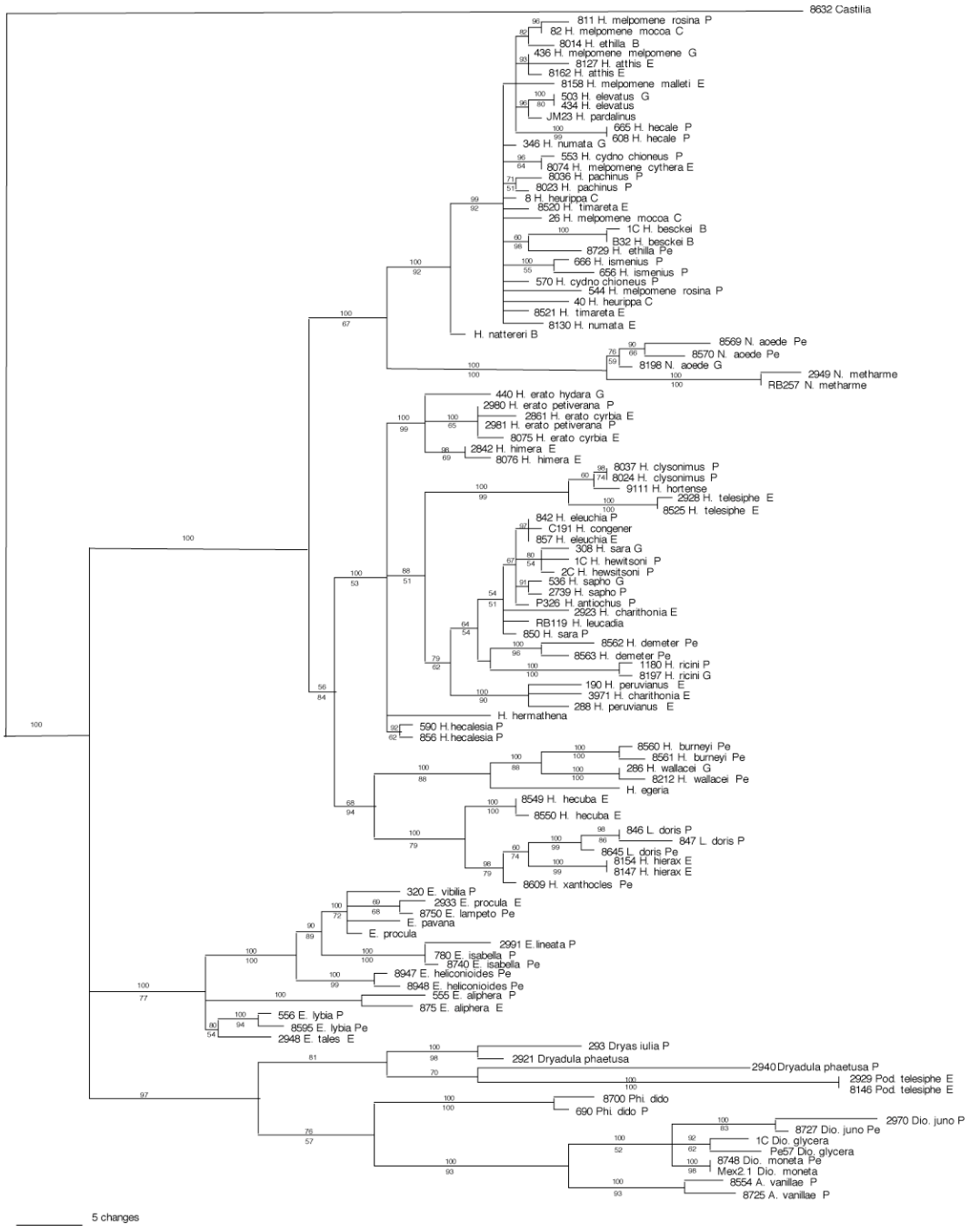
Data set	Hypotheses	SH	P	Length	CI	RI
mtDNA	Figure 3	19105.7	(Best)	3641	0.295	0.489
	Laparus basal to *Heliconius*	19135.6	0.092	3662	0.294	0.485
	Neruda basal to *Heliconius* and *Eueides*	19117	0.507	3647	0.295	0.488
	Eueides inside pupal-mating clade	19127.3	0.266	3655	0.294	0.487
	Laparus and *Neruda* basal to *Heliconius*	19197.9	**0.000***	3696	0.291	0.478
	Neruda basal to *cydno-melpomene* and "Primitives"	19108.8	0.843	3642	0.295	0.489
	Neruda inside pupal-mating clade	19126.5	0.234	3654	0.294	0.487
	Neruda basal to pupal-mating clade	19117.6	0.476	3645	0.295	0.489
	Neruda basal to *Heliconius*	19117	0.507	3647	0.295	0.488
Nuclear	Figure 3	9238.92	(Best)	1271	0.505	0.669
	Laparus basal to *Heliconius*	9265.13	0.154	1284	0.5	0.663
	Neruda basal to *Heliconius* and *Eueides*	9255.47	0.365	1281	0.501	0.664
	Eueides inside pupal-mating clade	9287.03	**0.017***	1306	0.492	0.651
	Laparus and *Neruda* basal to *Heliconius*	9340.87	**0.000***	1332	0.482	0.637
	Neruda basal to *cydno-melpomene* and "Primitives"	9244.86	0.748	1273	0.504	0.668
	Neruda inside pupal-mating clade	9295.76	**0.007***	1298	0.495	0.655
	Neruda basal to pupal-mating clade	9255.44	0.365	1281	0.501	0.664
	Neruda basal to *Heliconius*	9255.47	0.365	1281	0.501	0.664
Combined	Figure 3	29376.1	(Best)	4912	0.35	0.539
	Laparus basal to *Heliconius*	29434.4	**0.039***	4946	0.347	0.534
	Neruda basal to *Heliconius* and *Eueides*	29402.9	0.329	4928	0.348	0.537
	Eueides inside pupal-mating clade	29463.6	**0.005***	4961	0.346	0.532
	Laparus and *Neruda* basal to *Heliconius*	29592.6	**0.000***	5028	0.341	0.522
	Neruda basal to *cydno-melpomene* and "Primitives"	29384.3	0.777	4915	0.349	0.538
	Neruda inside pupal-mating clade	29458.4	**0.008***	4952	0.347	0.533
	Neruda basal to pupal-mating clade	29403.5	0.318	4926	0.349	0.537
	Neruda basal to *Heliconius*	29402.9	0.329	4928	0.348	0.537

Figure 3. BAY phylogenies for heliconiine species based on separate data partitions mtDNA (CO and 16S) and nuclear data. A. mtDNA. B. *Ef1α*. C. CO, 16S, *Ef1α*, *dpp*, *ap* and *wg* for 12 species. Branch lengths were estimated using likelihood. Values above branches show BAY probabilities and below show parsimony bootstrap support for the equivalent node, after 1000 replicates. Branches without bootstrap support reflect differences between BAY and MP trees. P= Panama, E= Ecuador, G= French Guiana, C= Colombia, and Pe= Peru.

A. mtDNA.



B. BAY phylogeny based on *elongation factor 1 α* (*Ef1α*).



C. BAY phylogeny based on CO, 16S, *Ef1*α, *dpp*, *ap* and *wg* for12 species representing the major clades within *Heliconius* and *Dryas* as a root.

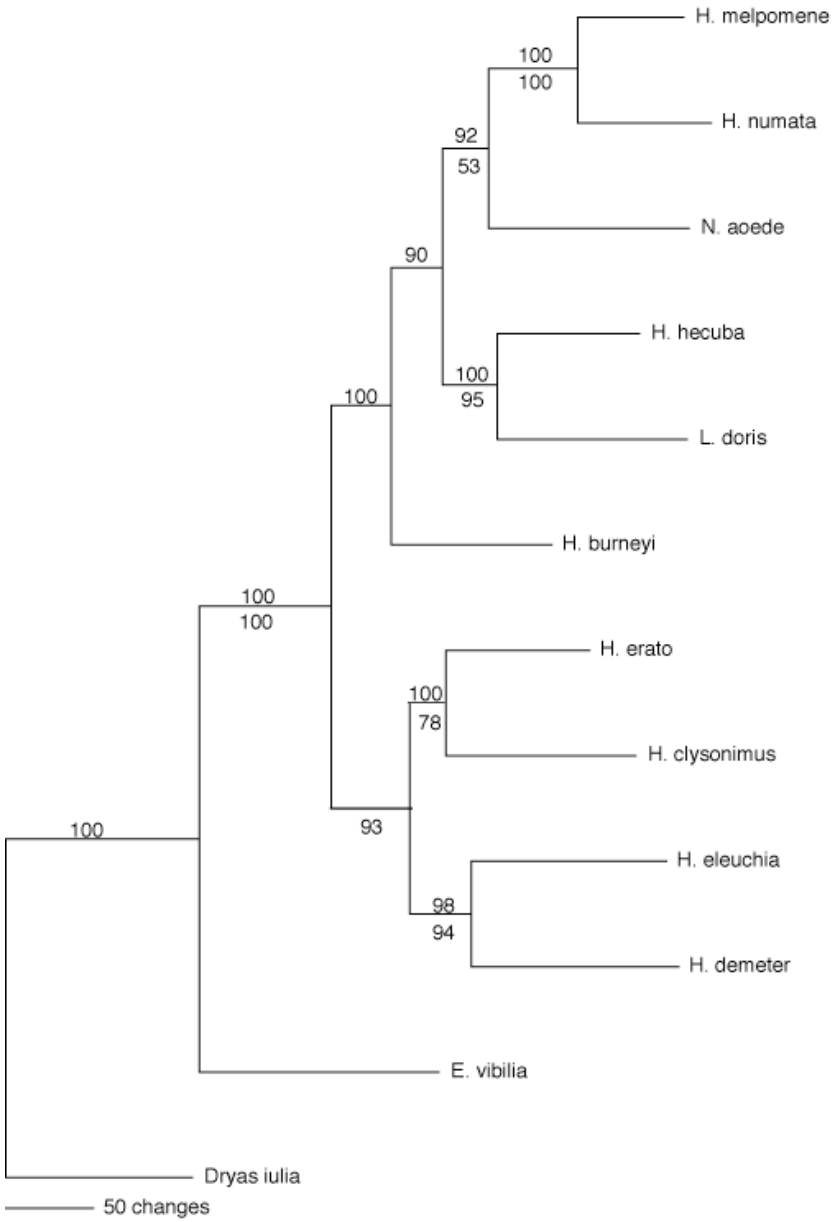
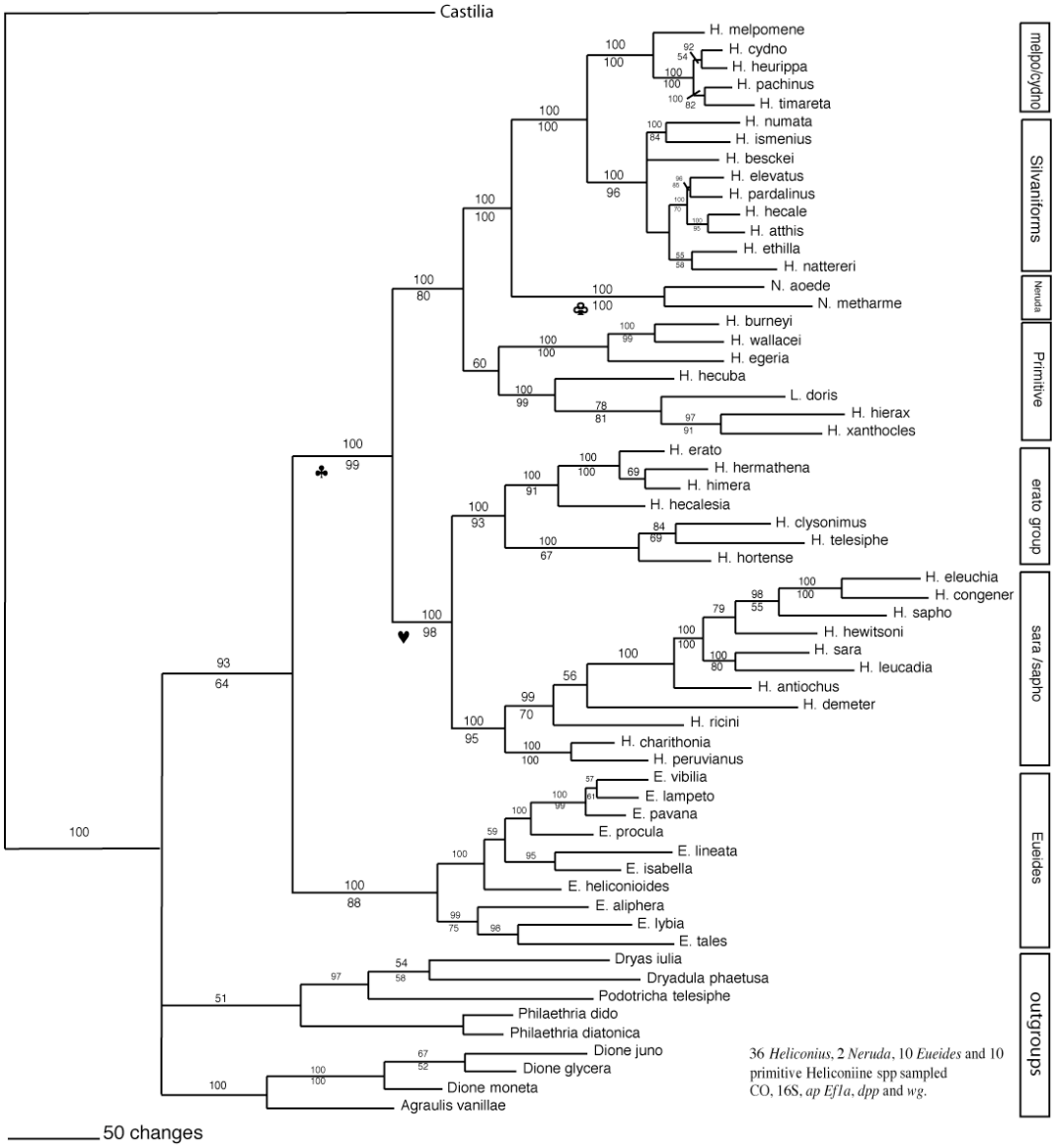


Figure 4. BAY phylogenetic hypothesis for heliconiine species based on combined mitochondrial (CO and 16S) and nuclear data (*Ef1*α, *dpp*, *ap* and *wg*). Only one individual per species was used and the *wg* sequences aligned correspond to different individuals. Branch lengths were estimated using likelihood. Values above branches show BAY probabilities and below show parsimony bootstrap support for the equivalent node, after 1000 replicates. Branches without support reflect differences between BAY and MP trees. P= Panama, E= Ecuador, G= French Guiana, C= Colombia, and Pe= Peru. Black clubs mark pollen feeding behaviour, black hearts mark pupal-mating. White clubs and hearts represent loss of the trait.



Discussion

I have presented a novel phylogenetic hypothesis for the Heliconiina, using more species, improved phylogenetic methodology and more phylogenetic information than heretofore attempted. Nine new species were added to the total used by Brower and Egan (1997): five *Heliconius*, *H. nattereri*, *H. hierax*, *H. hecalesia, H. peruvianus*, and *H. hermathena*; four *Eueides*, *E. lampeto*, *E. pavana*, *E. lineata*, *E. heliconioides,* and finally one outgroup species *Dione moneta*. Following Lamas *et al*. (2004) the species included represent 36 of 40 *Heliconius* species (90%), and 60 of 69 (86%) of the species in the subtribe Heliconiina. One of the missing species is *H. tristero*, recently defined by Brower (1996b). However, recent attempts to collect this species in the type localities have failed (Moreno-Fonseca *et al*., 2004). The remaining missing species *H. astraea*, *H. lalitae*, and *H. luciana*, were difficult to obtain as they are restricted to small areas of Brazil, French Guiana, and Venezuela respectively. Additionally, three new nuclear regions were reported for this subtribe *Ef1α* (876bp), *ap* (195 bp) and *dpp* (270bp) and 659bp were added to the COI region reported by Brower (1994) for *Heliconius*.

The phylogenetic hypothesis from combined evidence (Fig. 4) mostly agreed with the cladogram given by Brower and Egan (1997). Of 25 nodes at or above the species level 23 are concordant including the position of the genera *Eueides*, *Neruda* and *L. doris*. The position of *Laparus* as a basal member of *Heliconius* was well supported, but that of *Neruda* remains unclear, as it was not possible to rule out the hypothesis that *Neruda* is sister to *Heliconius*. However, *Neruda* basal to silvaniforms + *melpomene*/*cydno* group was independently supported by *Ef1α*, mtDNA and combined *Ef1α*, mtDNA*, ap*, *dpp* and *wg* tree (Fig. 3). Using this data, I therefore cannot reject the most parsimonious hypothesis for the evolution of pollen feeding, that of a single origin in a sister taxa to *Neruda* that went on to diversify into present day *Laparus* and *Heliconius*. Pollen feeding may therefore have evolved uniquely and never have been lost.

Relationships in the melpomene/cydno and silvaniform group

The *melpomene/cydno* group and the silvaniform complex are a rapidly radiating group of species with little differentiation at nuclear loci. The combined analysis showed two monophyletic groups *melpomene/cydno* and silvaniforms both with 100% BAY support (Fig. 4). This result was mostly due to information from mtDNA (Fig. 3A), because *Ef1*α has little informative variation (Fig. 3B). Brower and Egan (1997) observed the same pattern comparing COI, COII vs. *wg*. This suggests that nuclear markers (mainly coding regions) used previously to assess *Heliconius* relationships are insufficiently variable to resolve relationships at or near the species boundary.

In the *melpomene/cydno* group, races of *H. melpomene* cluster into two different clades. *H. melpomene* races from west of the Andes clustered with the *H. cydno* clade, while races of *H. melpomene* from east were clustered with *H. m. melpomene* from French Guiana. *H. cydno* appeared paraphyletic with respect to *H. heurippa*, *H. pachinus*, and *H. timareta* (Fig. 3A). Brower (1994) suggested that *H. heurippa*, *H. pachinus*, and *H. timareta* might represent well-differentiated races of *H. cydno* rather than distinct species, because they are parapatric or allopatric. However, analyses of genitalia, allozymes, RAPDs, and mating behaviour have suggested that *H. heurippa* is a good species (Beltrán, 1999). Also, studies of assortative mating in *H. pachinus* suggest species status (Gilbert, 2003), meanwhile for *H. timareta* detailed morphological and biological studies remain to be carried out.

The composition of the silvaniform complex agreed with Brower (1994), (Fig. 4), but the exact topology differed. The *H. numata* + *H. ismenius* and *H. atthis* + *H. hecale* species pairs were the only nodes in agreement with Brower and Egan (1997). It has been considered that *H. ethilla* is a sister to *H. atthis*, but here *H. ethilla* clustered with *H. nattereri*, one of the new species included. *H. atthis* and *H. hecale* are sympatric in Ecuador and it seems possible that their apparent sister species relationship could be a result of some gene flow between *H. atthis* and *H. hecale*. Additionally, it is clear that *H. elevatus* and *H. besckei* are part of this

complex rather than of the *melpomene*/*cydno* group as proposed by Brown (1981). Most of the silvaniforms have a typical 'tiger' colour pattern and Brower (1994) proposed that the 'postman' pattern (red forewing patches and yellow hindwing stripes on a black background) of *H. besckei* might be the ancestral colour pattern of this clade. This idea was supported here as *H. besckei* is placed basal to the silvaniforms.

Pupal-mating clade

The combined results provided strong support for a monophyletic 'pupal-mating' clade, suggesting that this trait may have played an important role in the phylogenetic expansion of *Heliconius* as well as in the packing of *Heliconius* species into local habitats. This group includes *sara*/*sapho*, *erato*/*himera* and *H. charithonia* groups (Fig 4). All the species in this clade were part of the pupal-mating clade in Brown (1981), Brower (1994) and Brower and Egan (1997). The *sara*/*sapho* clade had the largest sister species genetic distances in *Heliconius*, suggesting that species such us *H. eleuchia*, *H. congener*, and *H. sapho* are relatively ancient. In this clade the topology remainede largely unchanged as compared to previous hypothesis (Brower and Egan, 1997), and the additional species *H. peruvianus* was placed as sister to *H. charithonia* as expected (Jiggins and Davies, 1998).

In the *erato* clade three additional species were included, *H. hermathena*, *H. hecalesia*, and *H. hortense*. The placements of *H. hermathena* and *H. hortense* were different to the those previously suggested (Brown, 1981). *H. hermathena* is restricted to certain non-forest habitats in the Brazilian Amazon. One of its four subspecies, *H. h. vereatta*, is mimetic of sympatric *H. m. melpomene* and *H. e. hydara* and is very restricted geographically (Brown and Benson, 1977). The other three are non-mimetic, little differentiated and apparently widespread but the populations are patchy and in low densities. Their wing colour pattern is black, yellow and red; the forewing is black with red, resembling *H. m. melpomene* and *H. e. hydara*, while the hindwing is black with yellow bars and spots, resembling *H. charithonia*. Brown and Benson (1977) suggested, based on

adult morphology and pupal-mating behaviour, that this species is closely related to *H. erato* and *H. charithonia*. However, pupal characters (short head appendages) suggested that the species is relatively primitive, near to the *melpomene* group, in which all the members show the head appendages strongly shortened. The results shown here demonstrate that *H. hermathena* is a member of the pupal-mating clade, but with some discordance between nuclear and mtDNA results. The nuclear data (*Ef1*α) suggested *H. hermathena* basal to the pupal-mating clade (Fig 3B), while mtDNA place *H. hermathena* as a sister to *H. erato* (Fig 3A). Perhaps *H. hermathena* is a basal member of this clade, which would explain the unusual pupal morphology shared with *H. melpomene*, and has acquired colour pattern elements and mtDNA haplotypes via recent hybridization with *H. erato*.

H. hortense and *H. clysonimus* have similar colour patterns (red band on forewing, yellow in hindwing) and *H. hortense* is a geographical replacement of *H. clysonimus* found in the cloud forests of Mexico and Nicaragua, with *H. clysonimus* distributed in similar habitats through Costa Rica and North Andes. For this reason, *H. hortense* was expected to be a sister species of *H. clysonimus* (Brown, 1981). However, in the combined evidence *H. hortense* appeared basal to the sister species *H. telesiphe* and *H. clysonimus* (Fig 4). This result is due to the mtDNA topology (Fig 3A), because in the *Ef1*α tree the relationship of these species was not well resolved (Fig 3B). The colour pattern of *H. telesiphe* is similar but in different positions (two red bands on the forewing, and yellow or white on the hindwing). Thus the ancestral colour pattern for this group was probably a yellow band on the forewing and red on the hindwing and *H. telesiphe* has diverged recently in the Southern Andes. Although there are no extant *H. telesiphe* with yellow bars on the forewings as in *H. clysonimus* and *H. hortense*, one race, *H. telesiphe cretacea*, has white forewing bands.

Paraphyletic taxa

Paraphyly was observed at several different levels. Paraphyly of species relative to their sisters was observed in the *melpomene*/*cydno* group, *H. melpomene* was

paraphyletic with respect to a clade that includes *H. cydno* and related species. In the *erato* group, *H. erato* was paraphyletic with respect to *H. himera* and *H. hermathena* (Fig 3A and B see also Flanagan *et al*., 2004). Second at the genus level, *Heliconius* was paraphyletic with respect to *Laparus* and *Neruda* (Fig 4).

At the species level, this paraphyly is expected due to hybridization and recent speciation. Many wild hybrids between *H. cydno* and *H. melpomene* (Mallet *et al*., 2003) and *H. erato* and *H. himera* (Jiggins *et al*., 1996; Mallet *et al*., 2003) have been found, and it is known that these species have strong but incomplete reproductive isolation (McMillan *et al*., 1997; Naisbit *et al*., 2002). Recent studies of pre- and post-mating isolation between the sister species *H. cydno* and *H. melpomene* suggested that these species separated recently as a result of divergence in habitat and colour pattern (Jiggins *et al*., 2001b). Additionally, it was suggested that the speciation process between these species might have occurred in sympatry or parapatry with ongoing gene flow, but the alternative allopatric divergence cannot be ruled out (Jiggins *et al*., 1996; Jiggins *et al*., 2001b). For this reason, and because ancestral polymorphisms may persist after speciation, phylogenies of recently evolved species, which may still exchange genes, are inevitably difficult to resolve and likely to produce paraphyletic taxa, even in cases when the initial split was a simple bifurcation.

Paraphyletic patterns for the closely related species were observed where a number of races for each species were included in this study. This paraphyly might be observed in more pairs of sister species if more geographic populations were sequenced, because around 26% of *Heliconius* species hybridize in the wild (Mallet *et al*., 2003). However, the markers used in this study, also failed to resolve relationships between close relatives partly because of slow rates of evolution. Faster evolving markers as introns of nuclear genes might help to resolve the species boundaries (Beltran *et al*., 2002).

At the genus level, it is clear that *Laparus doris* is part of *Heliconius*, suggesting only a single origin for pollen feeding in the *Heliconius* group. *Laparus doris* was

suggested as a different genus by Turner (1976), in part due to the marked colour polymorphism as an adult (red, yellow and the unique blue or green ray pattern in hindwing). It is the only species within *Heliconius* with morphological polymorphism as a pupa, and the pupa do not have the gold spots and flanges and well developed antennal spines of other species. Also, it is the only species apart from *N. metharme*, to produce blue colour not by iridescence but by laying white scales over black (Turner, 1976), but these morphological traits provide weak support for a basal position in *Heliconius*. Colour patterns are known to be rapidly evolving and pupal characters may be derived adaptations to gregarious larval ecology. *Neruda* was also defined as a genus by Turner (1976), due to its short antennae and wing shape. In particular the broad triangular forewings with extensive friction patches of the male are very distinctive, although the females have wings of more normal shape for the genus *Heliconius*. Additionally, the *Neruda* larva does not have scoli on the head, as do other *Heliconius* species. *Neruda* was therefore basal to *Heliconius* according to morphology but again the support is weak for this placement. I have retained traditional nomenclature but it now seems likely that the genus *Laparus*, at least, should be subsumed within *Heliconius*.

Conclusions

Overall, there is a good concordance of the molecular hypothesis presented here with previous molecular phylogenies. However, the inclusion of more species and the addition of more sequence information, were useful to clarify some relationships in the Heliconiina. The hypothesis shows that *Heliconius* is not monophyletic because *Laparus doris* and possibly *Neruda* fall within the genus. These results suggest that pollen-feeding behaviour evolved in the common ancestor of *Laparus* and *Heliconius*. Pollen-feeding may have been subsequently lost by the ancestor of *Neruda*, but the position of *Neruda* remains somewhat unclear. Additionally, the results provided strong support for exclusion of *Eueides* from *Heliconius* and for a monophyletic 'pupal-mating clade' including the *erato*/*sara*/*sapho* groups. This phylogenetic hypothesis can now be used to test hypotheses regarding evolutionary patterns of diversification across the genus.

References

Beltran M, Jiggins CD, Bull V, *et al*. (2002) Phylogenetic discordance at the species boundary: Comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Molecular Biology and Evolution* **19**, 2176-2190.

Beltrán M (1999) *Evidencia genética (Alozimas) para evaluar el posible origen híbrido de Heliconius heurippa (Lepidoptera: Nymphalidae)* Masters, Universidad de los Andes.

Bischoff C, Kahns S, Lund A, *et al*. (2002) The human elongation factor 1 A-2 gene (EF1A2): complete sequence and characterization of gene structure and promoter activity. *Genomics* **68**, 63-70.

Boggs CL, Smiley JT, Gilbert LE (1981) Patterns of pollen exploitation by *Heliconius* butterflies. *Oecologia* **48**, 284-289.

Brower AVZ (1994a) Phylogeny of *Heliconius* butterflies inferred from mitochondrial DNA sequences (Lepidoptera: Nymphalidae). *Molecular Phylogenetics and Evolution* **3**, 159-174.

Brower AVZ (1994b) 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* **91**, 6491-6495.

Brower AVZ (1995) Locomotor mimicry in butterflies? A critical review of the evidence. *Philosophical Transactions of the Royal Society of London (B)* **347**, 413-425.

Brower AVZ (1996a) A new mimetic species of *Heliconius* (Lepidoptera:Nymphalidae), from southeastern Colombia, revealed by cladistic analysis of mitochondrial DNA sequences. *Zoological Journal of the Linnean Society* **116**, 317-332.

Brower AVZ (1996b) Parallel race formation and the evolution of mimicry in *Heliconius* butterflies: A phylogenetic hypothesis from mitochondrial DNA sequences. *Evolution* **50**, 195-221.

Brower AVZ, DeSalle R, Vogler A (1996) Gene trees, species trees, and systematics: A cladistic perspective. *Annual Review of Ecology and Systematics* **27**, 423-450.

Brower AVZ, Egan MG (1997) Cladistic analysis of *Heliconius* butterflies and relatives (Nymphalidae: Heliconiiti): a revised phylogenetic position for *Eueides* based on sequences from mtDNA and a nuclear gene. *Proceedings of the Royal Society of London Series B-Biological Sciences* **264**, 969-977.

Brower AVZ, DeSalle R (1998) Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. *Insect Molecular Biology* **7**, 73-82.

Brown KS, Benson WW (1977) Evolution in modern amazonian non-forest islands: *Heliconius hermathena*. *Biotropica* **9**, 95-117.

Brown KS (1981) The biology of *Heliconius* and related genera. *Annual Review of Entomology* **26**, 427-456.

Brown WM, George MJ, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 1967-1971.

Cardoso NZ (2001) Patterns of pollen collection and flower visitation by *Heliconius* butterflies in southeastern Mexico. *Journal of Tropical Ecology* **17**, 763-768.

Caterino MS, Sperling FAH (1999) *Papilio* phylogeny based on mitochondrial cytochrome oxidase I and II genes. *Molecular Phylogenetics and Evolution* **11**, 122-137.

Cho SW, Mitchell A, Regier JC*, et al.* (1995) A highly conserved nuclear gene for low-level phylogenetics - Elongation factor-1-alpha recovers morphology based tree for Heliothine moths. *Molecular Biology and Evolution* **12**, 650-656.

Deinert EI, Longino JT, Gilbert LE (1994) Mate competition in butterflies. *Nature* **370**, 23-24.

Emsley MG (1963) A morphological study of imagine Heliconiinae (Lep.:Nymphalidae) with a consideration of the evolutionary relationships within the group. *Zoologica* **48**, 85-130.

Emsley MG (1965) Speciation in *Heliconius* (Lep., Nymphalidae): morphology and geographic distribution. *Zoologica, New York* **50**, 191-254.

Estrada C, Jiggins CD (2002) Patterns of pollen feeding and habitat preference among *Heliconius* species. *Ecological Entomology* **27**, 448-456.

Farris JS, Kallersjo M, Kluge AG, Bult C (1994) Testing significance of incongruence. *Cladistics-the International Journal of the Willi Hennig Society* **10**, 315-319.

Felsenstein J (1981) Evolutionary trees from DNA sequences a maximum-likelihood approach. *Journal of Molecular Evolution* **17**, 368-376.

Flanagan N, Tobler A, Davison A*, et al.* (2004) Historical demography of Mullerian mimicry in the Neotropical *Heliconius* butterflies. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9704-9709.

Gilbert LE (1972) Pollen feeding and reproductive biology of *Heliconius* butterflies. *Proceedings of the National Academy of Sciences of the United States of America* **69**, 1403-&.

Gilbert LE (1975) Ecological consequences of a coevolved mutualism between butterflies and plants. In: *Coevolution of Animals and Plants* (eds. Gilbert LE, Raven PH), p. 246. University of Texas Press, Austin.

Gilbert LE (1991) Biodiversity of a Central American *Heliconius* community: pattern, process, and problems. In: *Plant-Animal Interactions: Evolutionary Ecology in Tropical and Temperate Regions* (eds. Price PW, Lewinsohn TM, Fernandes TW, Benson WW), pp. 403-427. John Wiley & Sons, New York.

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. In: *Ecology and Evolution Taking Flight: Butterflies as Model Systems* (eds. Boggs CL, Watt WB, Ehrlich PR). Univ. of Chicago Press, Chicago.

Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* **49**, 652-670.

Hall B (2001) *Phylogenetic trees made easy* Sinauer Associates, Inc., Sunderland, Massachusetts.

Hernandez MIM, Benson WW (1998) Small-male advantage in the territorial tropical butterfly *Heliconius sara* (Nymphalidae): A paradoxical strategy? *Animal Behaviour* **56**, 533-540.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Evolution - Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**, 2310-2314.

Jiggins CD, McMillan WO, Neukirchen W, Mallet J (1996) What can hybrid zones tell us about speciation? The case of *Heliconius erato* and *H. himera* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society* **59**, 221-242.

Jiggins CD, Davies N (1998) Genetic evidence for a sibling species of *Heliconius charithonia* (Lepidoptera; Nymphalidae). *Biological Journal of the Linnean Society* **64**, 57-67.

Jiggins CD, Naisbit RE, Coe RL, Mallet J (2001) Reproductive isolation caused by colour pattern mimicry. *Nature* **411**, 302-305.

Kimura Y (1980) A simpler method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111-120.

Krenn HW, Penz CM (1998) Mouthparts of *Heliconius* butterflies (Lepidoptera : Nymphalidae): A search for anatomical adaptations to pollen- feeding behavior. *International Journal of Insect Morphology & Embryology* **27**, 301-309.

Krenn HW, Zulka KP, Gatschnegg T (2001) Proboscis morphology and food preferences in nymphalid butterflies (Lepidoptera : Nymphalidae). *Journal of Zoology* **254**, 17-26.

Lamas G, Callaghan C, Casagrande MM*, et al.* (2004) *Hesperioidea and Papilionoidea* Association for Tropical Lepidoptera, Gainesville, Florida.

Maddison WP, Maddison DR (1997) MacClade: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, Massachussets.

Mallet J, McMillan WO, Jiggins CD (1998) Mimicry and warning colour at the boundary between races and species. In: *Endless Forms: Species and Speciation* (eds. Howard DJ, Berlocher SH), pp. 390-403. Oxford University Press, New York.

Mallet J, Neukirchen W, Linares M (2003) Hybrids between species of *Heliconius* and *Eueides* butterflies: a database. http://www.ucl.ac.uk/taxome/hyb/helichyb.html.

McMillan WO, Jiggins CD, Mallet J (1997) What initiates speciation in passion-vine butterflies? *Proceedings of the National Academy of Sciences of the United States of America* **94**, 8628-8633.

Michener CD (1942) A generic version of the Heliconiinae (Lepidoptera, Nymphalidae). *American Museum Novitates* **1197**, 1-8.

Mitchell A, Cho S, Regier JC*, et al.* (1997) Phylogenetic utility of elongation factor-1 alpha in Noctuoidea (Insecta: Lepidoptera): The limits of synonymous substitution. *Molecular Biology and Evolution* **14**, 381-390.

Moreno-Fonseca LP, Beltran M, Salazar C, *et al*. (2004) The status of *Heliconius tristero* Brower (Lepidoptera: Nymphalidae), a doubtful example of mimetic convergence between closely related passion-vine butterflies. *Biological Journal of the Linnean Society. In review*.

Murawski DA, Gilbert LE (1986) Pollen flow in *Psiguria warscewiczii* - a comparison of *Heliconius* butterflies and hummingbirds. *Oecologia* **68**, 161-167.

Nahrstedt A, Davis RH (1981) The occurrence of the cyanoglucosides, linamarin and lotaustralin, in *Acraea* and *Heliconius* Butterflies. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **68**, 575-577.

Naisbit RE, Jiggins CD, Linares M, Salazar C, Mallet J (2002) Hybrid sterility, Haldane's rule and speciation in *Heliconius cydno* and *H. melpomene*. *Genetics* **161**, 1517-1526.

Nijhout HF (1991) *The Development and Evolution of Butterfly Wing Patterns* Smithsonian Institution Press, Washington, DC.

Penz CM (1999) Higher level phylogeny for the passion-vine butterflies (Nymphalidae, Heliconiinae) based on early stage and adult morphology. *Zoological Journal of the Linnean Society* **127**, 277-344.

Penz CM, Krenn HW (2000) Behavioral adaptations to pollen-feeding in *Heliconius* butterflies (Nymphalidae, Heliconiinae): An experiment using *Lantana* flowers. *Journal of Insect Behavior* **13**, 865-880.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.

Reed RD, Sperling FAH (1999) Interaction of process partitions in phylogenetic analysis: An example from the swallowtail butterfly genus *Papilio*. *Molecular Biology and Evolution* **16**, 286-297.

Routman E, Wu R, Templeton AR (1994) Parsimony, molecular evolution, and biogeography - the case of the North-American giant salamander. *Evolution* **48**, 1799-1809.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**, 1114-1116.

Smith P, Kambhampati S, Armstrong K (2002) Phylogenetic relationships among *Bactrocera* species (Diptera: Tephritidae) inferred from mitochrondrial DNA sequences. *Molecular Phylogenetics and Evolution* **26**, 8-17.

Sober E (1991) *Reconstructing the past: parsimony, evolution, and inference* MIT Press, Cambridge, MA, U.S.A.

Stern DL (1994) A phylogenetic analysis of soldier evolution in the aphid family Hormaphididae. *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* **256**, 203-209.

Swofford DL (2000) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**, 512-526.

Thornhill R, Alcock J (1993) *The evolution of insect mating systems* Harvard University Press, Cambrigde.

Tobler A, Flanagan N, Jiggins CD, Heckel DG, McMillan WO (2003) A genetic map of allozyme, AFLP and color pattern genes in *Heliconius erato* and *H. himera*. *Genetics. In prep*.

Turner JRG (1968) Some new *Heliconius* pupae - their taxonomic and evolutionary significance in relation to mimicry (Lepidoptera, Nymphalidae). *Journal of Zoology* **155**, 311-&.

Turner JRG (1976) Adaptive radiation and convergence in subdivisions of butterfly genus *Heliconius* (Lepidoptera-Nymphalidae). *Zoological Journal of the Linnean Society* **58**, 297-308.

Walldorf U, Hovemann BT (1990) *Apis mellifera* cytoplasmic Elongation-factor 1-alpha (Ef-1-alpha) is closely related to *Drosophila melanogaster* Ef-1-alpha. *Febs Letters* **267**, 245-249.

Wiens J (1998) The accuracy of methods for coding and sampling higher-level taxa for phylogenetic analysis: a simulation study. *Syst. Biol.* **47**, 397-413.

Yang Z (1994) Estimating the pattern of nucleotide substitution. *Journal of Molecular Evolution* **39**, 105-111.

Chapter 2

Species diversity and speciation in *Heliconius* butterflies: inferences from species level phylogeny

Abstract

Heliconius butterflies, which contain many intermediate stages between local varieties, geographic races, and sympatric species, provide an excellent biological model to study speciation. Using a species-level phylogeny for this group based on 3834bp of mtDNA (COI, COII, 16S) and nuclear loci (*Ef1α*, *dpp*, *ap*, *wg*), I test the geographic mode of speciation in *Heliconius* and whether mimicry is associated with speciation. The most recently diverged sister species presented a broad spread of range overlap values. Of 28 comparisons evaluated, four were closely related and showed sympatry close to 100% and five closely related comparisons displayed a sympatry of 0%; the remaining closely related comparisons showed intermediate degrees of overlap. This suggests that allopatric speciation is not the predominant mode of speciation in *Heliconius*. Mapping of the colour traits onto the phylogeny showed that there are frequent shifts in colour pattern within and between sister species and that these changes can occur in sympatry or allopatry. Clades with greater colour pattern diversification also have greater species richness than expected at random; this suggests that speciation is facilitated by the evolution of novel mimetic patterns.

Introduction

Speciation is the creative process leading to build up of species diversity, so that understanding the general patterns and processes of speciation is fundamental to explaining the diversity of life (Mayr, 1963). In the last few years speciation has received much attention in many fields, from genetics through ecology to paleontology (e.g. see special issue 2001 Vol.16 Trends in Ecology and Evolution). The central question is how can one species split into two reproductively isolated groups of organisms? To answer this question one issue is

whether new species evolve in geographic isolation from their ancestors (allopatric speciation), in a geographically contiguous population (parapatric speciation), or within the geographical range of their ancestor (sympatric speciation). Given enough time, speciation is an inevitable consequence of populations evolving in allopatry. As selection does not promote reproductive compatibility between geographically isolated populations, they will eventually become reproductively incompatible. Thus, allopatric speciation requires only geographical isolation and time. In contrast, sympatric speciation requires natural selection to drive divergence (Turelli *et al.*, 2001).

Sympatric speciation has become increasingly accepted in the past decade, as a result of new models confirming its plausibility and new evidence that the conditions specified by the models are found in many natural populations (Jiggins and Bridle, 2004; Via, 2001). Sympatric speciation is most likely if selection first establishes a stable polymorphism and then favours assortative mating between each polymorphic type. One of the best examples are host races in the apple maggot fly *Rhagoletis pomonella*, where ecologically driven reproductive isolation is associated with adaptation to alternative host plants (Bush, 1969). In the last 15 years research has verified the genetic differentiation between the sympatric host races of *Rhagoletis* (Feder *et al.*, 1988; McPheron *et al.*, 1988), and shown strong disruptive selection on habitat use. For example, there is a trade-off because although apples are nutritionally inferior to hawthorn fruits, they represent enemy free space (Grant, 1981). Also, in adult behaviour, mating occurs exclusively on or near the fruit of their host plant causing pre-mating isolation (Feder, 1998). Finally, there is isolation due to the phenology of the host plant, as flies with deeper diapauses are more favoured in apple than in the hawthorn race (Feder, 1998).

Evidence for sympatric speciation also comes from laboratory experiments. A study using *Drosophila melanogaster* tested the hypothesis that disruptive selection on habitat preferences can lead to sympatric speciation when individuals mate locally within the selected habitat. The conclusion was that the reproductive

isolation that develops over the course of the experiment was a result of offspring returning to mate in the habitat type selected by their parents. Thus, reproductive isolation can evolve in a laboratory setting between sympatric populations with no physical barriers to gene flow (Rice and Salt, 1990). Also, a review of forty years of laboratory experiments on speciation found that speciation generally failed in laboratory studies characterised by a single strong force of disruptive selection (Rice and Hostert, 1993), but that multiple selective forces, particularly when disruptive selection had a pleiotropic effect on assortative mating, were very effective in promoting reproductive isolation between sympatric populations.

Heliconius butterflies: a model to study speciation

The passion-vine butterflies have played a key role in understanding evolutionary biology (e. g. mimicry) and ecology (e. g. mutualism between insects and plants) and it would be difficult to point a group of neotropical butterflies that have contributed more to our knowledge of the biological processes in the tropics. Brown (1981) presented an excellent review of the biology of *Heliconius* and related genera that is still considered the most comprehensive compilation of references about these butterflies. Since this review research has continued in many areas. Some examples are studies of the genetics, ecology and evolution of Müllerian mimicry (Brower, 1994b; Brower, 1995; Brower, 1996; Joron and Mallet, 1998; Joron *et al.*, 1999; Linares, 1997; Mallet, 1986a; Mallet, 1986b; Mallet and Gilbert, 1995; Mallet, 1999; Mallet and Joron, 1999; Naisbit *et al.*, 2003; Sheppard *et al.*, 1985; Speed, 1999; Srygley, 1999; Srygley and Ellington, 1999; Turner and Mallet, 1996), reproductive isolation (Jiggins *et al.*, 2001a; Naisbit *et al.*, 2002), reproductive biology (Bissoondath and Wiklund, 1995; Boggs, 1990; Garcia-Barros, 2000; Passos and Sazima, 1995), population genetics (Kronforst and Fleming, 2001), and molecular evolution (Hsu *et al.*, 2001). However, here I want to focus on studies of speciation in *Heliconius* facilitated by mimicry.

Heliconius butterflies show a continuum of geographic divergence and speciation, they are unpalatable and exhibit inter- and intraspecific diversification of colour

and patterns. Bates' classic paper (Bates, 1862), reflecting observations during his stay in the Amazon, showed a geographical pattern for the different colour forms: similar between species within any one area of the Amazon basin, but the mimetic colour patterns themselves changed every 100-200 miles. Beside this geographic divergence, closely related species within an area often belonged to mimicry "rings" (groups of unpalatable species, together with some palatable species, that have converged on the same warning colour pattern) (Mallet and Gilbert, 1995). Bates' system has all the intermediate stages between local varieties, geographic races, and sympatric species that make it an excellent biological model to study selection at the species boundary.

Multiple forms of selection are acting in the divergence of *Heliconius* species, a good example being the sister species *H. cydno* and *H. melpomene*. They are broadly sympatric, from Central America to Andean South America, but they differ in habitat, host plant use and mimicry (Mallet *et al*., 1998). In mimicry, although both species are unpalatable and warningly coloured, they belong to two different mimetic rings, *H. cydno* is typically black with yellow and white and usually mimics species from the group *sapho/eleuchia*, that inhabit deep forest interiors. *H. melpomene* is black with red and yellow and is a mimic of *H. erato*, which inhabits open areas. The two species occasionally hybridise and backcross in nature. Hybrid females are sterile, but males are fertile. The few hybrids produced are expected to have low fitness because they are non- mimetic (Naisbit *et al*., 2002).

Additionally, recent work shows that a shift in colour pattern mimicry was crucial in generating pre-mating isolation (Jiggins *et al*., 2001b). To test whether males use mimetic colour pattern as a cue in choosing mates, the response of males to moving model made with either natural wings or coloured paper was investigated. *H. melpomene* males approached *H. cydno* colour patterns about half as frequently as those of their own type. Similarly, *H. cydno* males were a third as likely to court a *H. melpomene* pattern as their own type. The results suggested that colour is important for choosing mates and that assortative mating is greater in sympatry.

Assortative mating reduces hybrid mating and the few hybrids produced are poorly adapted because they are non-mimetic. Thus, the mimetic shift can cause both pre-mating and post-mating isolation (Jiggins *et al*., 2001b). This study is a clear example of speciation facilitated by ecological adaptation. However, the role of mimicry and habitat use in speciation has never been tested in a phylogenetic context for this group of butterflies.

Phylogenetic tests of speciation

Species-level phylogenies derived from molecular data provide an indirect record of speciation and together with information on the geographical and ecological attributes of species, they can present information on the causes of speciation (Barraclough and Nee, 2001). One approach is to evaluate the relative frequency that different geographical modes of speciation occur. A study using species-level phylogenies based on the geographical distributions of recently formed sister species of several bird, insect, and fish groups, shows that virtually all the most recent sister species had no range overlap, suggesting that allopatric speciation is the norm (Barraclough and Vogler, 2000). This result supports the view of Mayr that allopatric speciation is more common (Mayr, 1963). This study has been widely cited as empirical evidence of allopatric speciation, and it has been suggested that the test may be conservative, as related sympatric species might falsely appear to be sister species because of genetic homogenization resulting from hybridization after divergence in allopatry (Turelli *et al*., 2001). However, a phylogenetic study of Cameroon crater- lake cichlids (Schliewen, 1994), showed monophyly of taxa from a small isolated environment, in this case crater lakes, and provided powerful evidence that speciation has occurred sympatrically.

Another approach is to use phylogenies in comparative studies. The most widely used comparative method for continuous characters is independent contrasts (Felsenstein, 1985). This method is designed to detect whether change in one trait is associated with change in another, and uses phylogenies to correct for statistical non independence caused by shared history (Felsenstein, 1985). Estimates of phylogenetic relationships are employed to transform *n* species mean values into

n-1 contrast values. Under a specified mode of evolution these contrasts are independent and association between traits can be tested using regression (Felsenstein, 1985; Garland *et al*., 1992)

This comparative methodology can also be used to look for correlates of species diversity. For example, traits such as plant-feeding in insects (Mitter *et al*., 1988), and sexual dichromatism in birds and lizards (Barraclough *et al*., 1995; Stuart-Fox and Owens, 2003) were found as factors associated with high diversity. Two alternative methods can be used in such studies. If the selected trait changes only rarely, its phylogenetic distribution is mapped to identify sister clades having different character states (Mitter *et al*., 1988). If the character is changing more frequently, a particular taxonomic level is chosen for analysis, and pairs of sister taxa compared to see whether the more diverse clade has a higher or lower occurrence of the character state proposed to increase diversity (Barraclough *et al*., 1995). Both methods test the null hypothesis of no association between the character and diversity.

MacroCAIC is a computer program that extends the method of phylogenetically independent contrasts to include species richness data (Agapow and Isaac, 2002). It examines a user-selected character for correlation with species richness, thus allowing identification of the factors driving large-scale patterns of diversity (Agapow and Isaac, 2002). The supplied phylogenies do not need to have branch lengths and can also represent taxa higher than species level (e.g. genera), which is important where the lower relationships have not been clearly resolved. This program has already been used to study species richness over a range of phylogenies, including hoverflies (Katzourakis *et al*., 2001), carnivores and primates (Gittleman and Purvis, 1998), lizards (Stuart-Fox and Owens, 2003), and a study of body size using 38 complete species-level phylogenies from vertebrates and invertebrate groups (Orme *et al*., 2002).

In a conventional analysis, MacroCAIC compares differences in character X with differences in species number. The nodal value (the one used in the next

comparison up the tree) is estimated as the mean of the two daughter nodes. The underlying model of evolution in this analysis is Brownian motion (random drift). The alternate hypothesis is that lineages where X increases will show higher (or lower) speciation rates, ignoring extinction (Isaac *et al*., 2003). With mimicry, there is a different underlying idea of how evolution has occurred. Mimicry is a qualitative trait, but taxa seem to vary in their tendency to produce distinct mimetic forms. I am therefore interested in testing whether there is any association between clades having more colour pattern forms and their species richness. In this study, I developed with help of Nick Isaac a novel method using MacroCAIC to test this association (see methods).

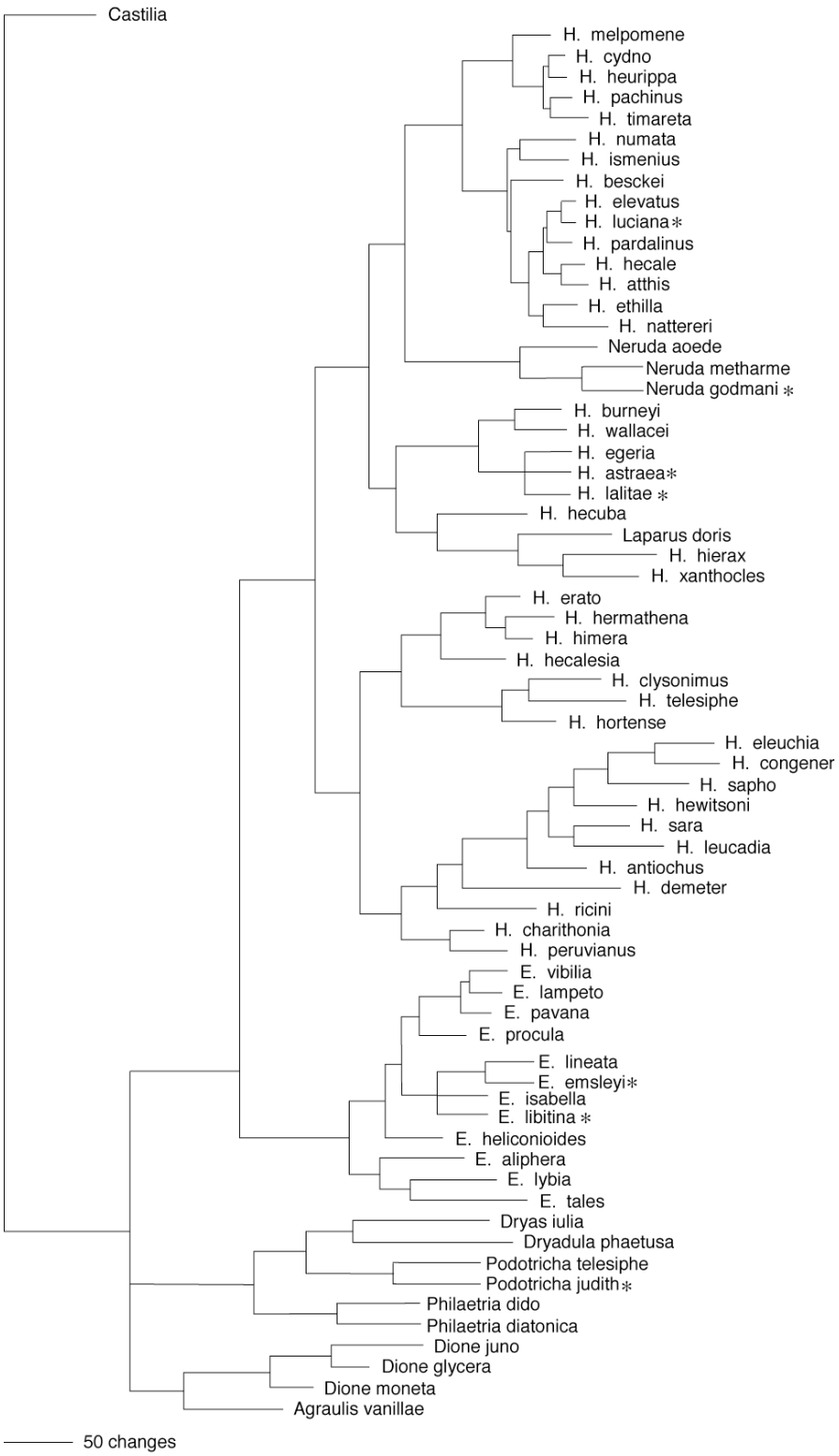
The aims of this chapter are to use my species-level phylogenetic hypothesis for the heliconiine butterflies 1) to test for the geographical pattern of speciation in *Heliconius* following the methodology described in Barraclough and Vogler (2000) and 2) to test whether changes in colour pattern are important in the speciation of *Heliconius*.

Methods

Phylogeny

I based my analysis on the molecular phylogeny of *Heliconius* described in the first chapter of this thesis (Fig 1 in this chapter), based on 3834bp of mtDNA (COI, COII, 16S) and nuclear loci (*Ef1α*, *dpp*, *ap*, *wg*). The phylogeny is nearly complete and included one or more representatives of all 10 genera of the subtribe Heliconina, equivalent to 60 of 69 species currently recognized. For the genus *Heliconius* 36 of 39 species are included. In summary, the phylogeny showed that the genus *Heliconius* is paraphyletic, with *Laparus doris* and *Neruda* falling within the genus. The support is high to place *Laparus* as part of *Heliconius*, however different genes did not produce a clear agreement as regards relationships of *Neruda* within *Heliconius*. The possibility that *Neruda* was basal to *Heliconius* could not be ruled out (Fig. 1).

Figure 1. Bayesian phylogeny for heliconiine species based on combined mitochondrial (CO and 16S) and nuclear data (*Ef1α*, *dpp*, *ap* and *wg*). Asterisks indicate species without sequence data that have been inserted for analysis, not included in the original phylogeny. The placement of species is based in the topology of Brown (1981).



Geographical isolation

The methodology described by Barraclough and Vogler (2000) was used to detect the geographic mode of speciation in *Heliconius*, with some modifications. This methodology plots range overlap against node age (estimated via genetic distance) in order to provide a summary of the pattern of geographic overlap subsequent to speciation. If speciation is predominantly allopatric, recently diverged sister species will initially overlap little in their geographical ranges. As geographical range changes after speciation, overlap between anciently split species should then increase. If speciation is predominantly sympatric, recently diverged sister species share a great deal of their geographical ranges, and range overlap is likely to decrease for relatively older nodes (Barraclough and Vogler, 2000).

In order to estimate node ages I first tested for a molecular clock using a likelihood ratio test (Felsenstein, 1981). The relative ages of nodes for the phylogeny were calculated using the *r8s* program (Sanderson, 2003). This is a program for estimating absolute rates of molecular evolution and divergence times on a phylogenetic tree in the absence of a molecular clock (Sanderson, 2003). The *Heliconius* phylogenetic tree including the estimated branch lengths was run using PL method (penalized likelihood). This method combines a parametric model having a different substitution rate on every branch with a non-parametric roughness, which costs the model more if rates change too quickly from branch to branch. The contribution of the two components is determined by a "smoothing parameter" (Sanderson, 2002). To find the optima of the PL method, the Powell algorithm was used (Sanderson, 2003). An age of 1.5 million years for the *H. cydno-H. melpomene* split (age estimated from the mtDNA data – see Beltran *et al*. (2002) and (Brower, 1994a), and a smoothing parameter of 1 were fixed for this analysis.

To test the geographical mode of speciation only species in the *Heliconius* clade were included (Fig. 1 clade including *Heliconius*, *Laparus* and *Neruda*). Species ranges were obtained from (Brown, 1979). Ranges were redrawn as shaded regions onto equal area maps in the computer. In ArcView®, a desktop mapping

program (Enviromental Systems Reseach Institute, Inc.), shaded regions were drawn around dot maps by eye, following text descriptions of the ranges of each species. Additionally, altitudinal ranges were included using the GIS program GRASS (Geographic Resources Analysis Support System). Shaded areas drawn in ArcView, were edited in GRASS deleting areas above and below the altitudinal range of the species, generating a new set of maps. The area of geographic overlap between all possible pairs of species was calculated in GRASS using the maps.

Degree of sympatry between sister species was calculated by measuring the overlap area, and dividing this by the total area of the species with smaller range. In this study the calculation of overlap area between clades was modified from Baraclough and Vogler (2000). Instead of summing the ranges of all the species included in the clade, the overlap was calculated as the average of the overlap area between pairs of species in the two clades (modification suggested by M. Turelli, pers. comm.). Overlaps varied from 0% (no range overlap) to 100% (range of one clade is entirely overlapped by its sister). Levels of sympatry were then plotted against relative node age and a line fitted using linear regression on arcsine-transformed values of sympatry. The rank correlation coefficient was determined using Spearman's method.

Comparative analysis

In order to illustrate changes in colour pattern between species, colour patterns described for *Heliconius* in Brown (1981) were mapped onto the phylogeny using MacClade 4.0 (Maddison and Maddison, 1997). Brown (1981) described seven different colour pattern groups 1) Dennis pattern, orange rays in hindwing, 2) postman pattern, yellow band forewing and red bar in hindwing and vice versa, 3)'tiger' pattern: yellow, orange and black mimics of ithomiine species, 4) other mimetic, mimics of other groups e.g. *Actinote*, *Tithorea* and *Elzunia*, 5) iridescent blue with white or yellow forewing markings, 6) white or yellow on hindwing and usually in forewing, 7) black with yellow bars. In order to include outgroups two more categories were added 8) orange and black, patterns that are mostly orange with black bars and 9) green and brown e.g. *Philaethria* (Fig. 2). For species that

contain more than one category, all the different states where included as a polymorphic character (Fig. 2 e.g. *H. melpomene* races). Character changes were optimized using 'accelerate changes' (ACCTRAN) implemented in MacClade 4.0.

Figure 2. Examples of colour patterns in passion-vine butterflies, and classification for this study (1-10). *H. melpomene rosina*, *H. m. cythera*, *H. m. malleti* and *H. m. melpomene* are geographical races, the different colour patterns were included as a polymorphic character state.

1) Dennis pattern
H. xanthocles

2) Postman pattern
H. melpomene rosina

3) 'tiger' pattern
H. ismenius

4) Other mimetic
H. hecuba

5) Blue and yellow
H. wallacei

6) White or yellow
H. sapho

7) Black with yellow bars
H. pachinus

8) Orange and black
Dryadula phaetusa

9) Green and brown
Philaethria dido

H. m . cythera

H. m. malleti

H. m. melpomene

Additionally I investigated whether changes in colour pattern are associated with high species diversity in *Heliconius* using MacroCAIC (Agapow and Isaac, 2002). The number of subspecies recognised in a recent revision of the group were used as a proxy for the colour pattern diversity. This assumes that subspecies definition is entirely based on colour pattern characters, which is generally true in this

group. I tested the prediction that clades with a greater tendency to produce colour patterns also have greater species diversity.

In a conventional analysis MacroCAIC calculates the richness of higher clades as the average of their constituent clades. For example, to test the correlation in a clade of three species here one daughter is a single species with 3 races and the other is a sister species pair with 1 and 4 races respectively. MacroCAIC takes the average of the two sister species (2.5) and compares it with the single species (3). This comparison therefore goes against the alternative hypothesis that the clade with more species also has more races. However, the total number of races in the pair of sister species (5) does support the alternate hypothesis. Therefore, I want to sum the total number of races in each clade rather than averaging between clades in order to obtain nodal values.

The methodology was based on the null hypothesis that if each species had the same number of races this would give a positive correlation between number of races and number of species with a slope of 1.0. In contrast, the alternative hypothesis that producing more races leads to speciation, would produce a slope significantly greater than 1. In order to test the hypothesis, I used MacroCAIC to calculate the relative rate difference (RRD) for both race and species diversity. RRD is calculated as $[\ln (n_i / n_j)]$ where n_i and n_j are the number of species (or races) in each clade and is a measure of the difference in the rate at which the two clades have accumulated species and has a normal distribution (Agapow and Isaac, 2002). First, I used diversity of colour pattern races as a species richness value. Thus, the species richness file was a list of the number of races per species and RRD for racial diversity was calculated for each node (Appendix 2.1). Then, the program was run again, but using the actual species diversity data (i.e. without a separate species richness file) to calculate RRD for species diversity at each node. In both cases a random number file was used as a data file. I am not actually interested in comparing any trait values, but the program requires this file, and by using the same random number file in each analysis this facilitates the alignment of the RRD values for species and racial diversity at each node. Then the RRD

values for species richness and colour pattern richness were aligned and a regression through the origin was performed in Excel. I then tested whether the slope was different from +1. The slope of the regression was subtracted from 1 and divided by the standard error. The t-score from this calculation was checked for significance using a Student's t distribution (two-tailed).

MacroCAIC analysis is likely to be very sensitive to taxonomic decisions regarding species status. The *Heliconius* taxonomy is very stable, but there is one group in which there is some uncertainty. Therefore in addition to the original topology and the number of races described by Lamas *et al*. (2004) (Appendix 2.1), a modified topology was also investigated in which *H. heurippa*, *H. pachinus*, *H. timareta* were considered races of *H. cydno* as proposed by some authors (Brower, 1994a). I also carried out an analysis with the original topology but with Caribbean races of *H. charithonia* and *Dryas iulia* excluded, as these races do not show marked differences in mimetic colour pattern.

In the comparative analysis missing data also may greatly affect the calculation of independent contrasts. Following previous comparative studies (e. g. Katzourakis et al. 2001), I repeated the analysis including missing taxa. The missing species in the phylogeny were placed in the tree based on the topology from morphological data (Brown, 1981), using MacClade 4.0 (Maddison and Maddison, 1997) (Fig. 1 species labeled with *). The different topologies were converted to MacroCAIC format using TreeEdit v1.0 (Rambaut and Charleston, 2001).

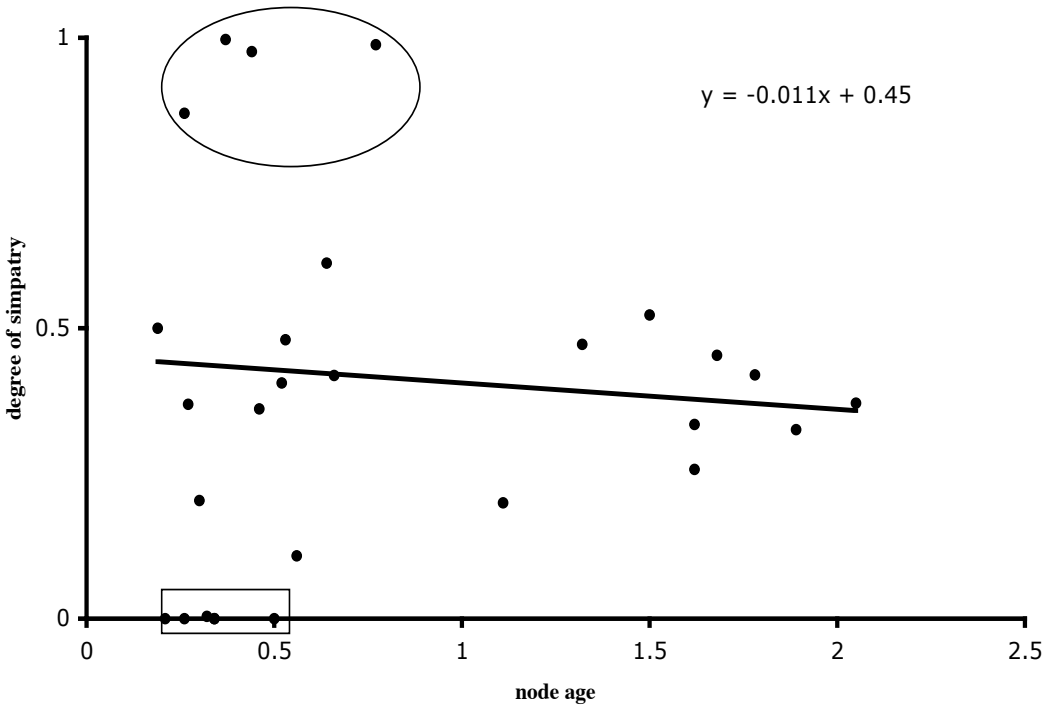
Results

Geographical isolation

The plot of degree of sympatry against relative node age is shown in Figure 3. The null hypothesis, that rate of evolution is homogeneous among all branches in the phylogeny, was rejected (LR= 277.49 $p<0.01$). This suggests that rates of substitution vary significantly among branches and a molecular clock is inappropriate. Thus, the relative ages of nodes were estimated using the *r8s*

program (Appendix 2.2). The most recently diverged sister species displayed a broad spread of range overlap values. Three species pairs, *H. elevatus* and *H. pardalinus*, *H. ethilla* and *H. nattereri*, and *H. wallacei* and *H. burneyi* and one clade of three species *H. wallacei*, *H. burneyi* and *H. egeria* were closely related and showed sympatry close to 1 (Fig. 3 see oval). In constrast, *H. cydno* and *H. heurippa*, *H. timareta* and *H. pachinus*, *H. numata* and *H. ismenius*, *H. telesiphe* and *H. hortense*, and *H. congener* and *H. eleuchia* displayed a sympatry of zero (Fig. 3 see rectangle). All the remaining closely related comparisons showed intermediate degrees of overlap. In addition sympatry remained fairly constant with relative node age suggesting that range changes have not occurred in these groups. The slope was -0.011 with an intercept of 0.45. The Spearman's rank correlation coefficient was not significant (r_S=0.04, P=0.26).

Figure 3. Plot of the degree of sympatry against relative node age in *Heliconius*. The *Y*-axis is the degree of sympatry, and the *X*-axis is relative node age. Regression analysis in text based on Arsine transformed data. The y-intercept suggests geographical overlap at time of speciation (0.45). Oval and rectangle highlight outliers (see text).



Comparative analysis

Many sister species differ in colour pattern, 10 of twenty sister species pairs showed shifts in colour (Fig. 4). For example, between the sister species *H. cydno* and *H. heurippa*, *H. cydno* has two character states black with blue iridescence and white or yellow in the forewing and white or yellow on hindwing, while *H. heurippa* has a pattern similar to the postman pattern (yellow and red in a black background).

MacroCAIC results are shown in Table 1. In all cases where colour pattern diversity of clades was plotted against species richness, the slope was greater than one. The null hypothesis that species have the same number of races (slope=1) was rejected in the analysis for the original phylogeny (Fig. 5 Table 1a).

Additionally, analysis of the same phylogeny changing the topology with *H. cydno* combined with *H. timareta*, *H. pachinus* and *H. heurippa* as a single species, and where the Caribbean races of *H. charithonia* and *D. iulia* where excluded, showed significant results (Table 1b and c). The other analyses showed slopes greater than one, but the values were not significant (Table 1d and e).

Table 1. Results of test for association between changes in colour pattern and species richness based on MacroCAIC. *v* represents degrees of freedom (number of contrast –1), p- value was calculated from a Student's t distribution (2-tailed). Significant tests are indicated as *p<0.05.

Phylogeny	Slope	*v*	*t* (diff from 1)
a. Species sequenced (Fig. 1)	y=1.27x	38	2.19*
b. Species sequenced; *H. cydno* including more races	y=1.30x	37	2.69*
c. Species sequenced; number of races modified for *H. charitonia* and *D. iulia*	y=1.26x	38	2.15*
d. Complete phylogeny	y=1.16x	38	1.37
e. Complete phylogeny; *H. cydno* and races of *H. charitonia* and *D. iulia* modified.	y=1.25x	37	1.79

Figure 4. Representation of shifts in colour between sister species. Colour pattern key in the figure and examples in figure 2. Acctran was used for character optimization.

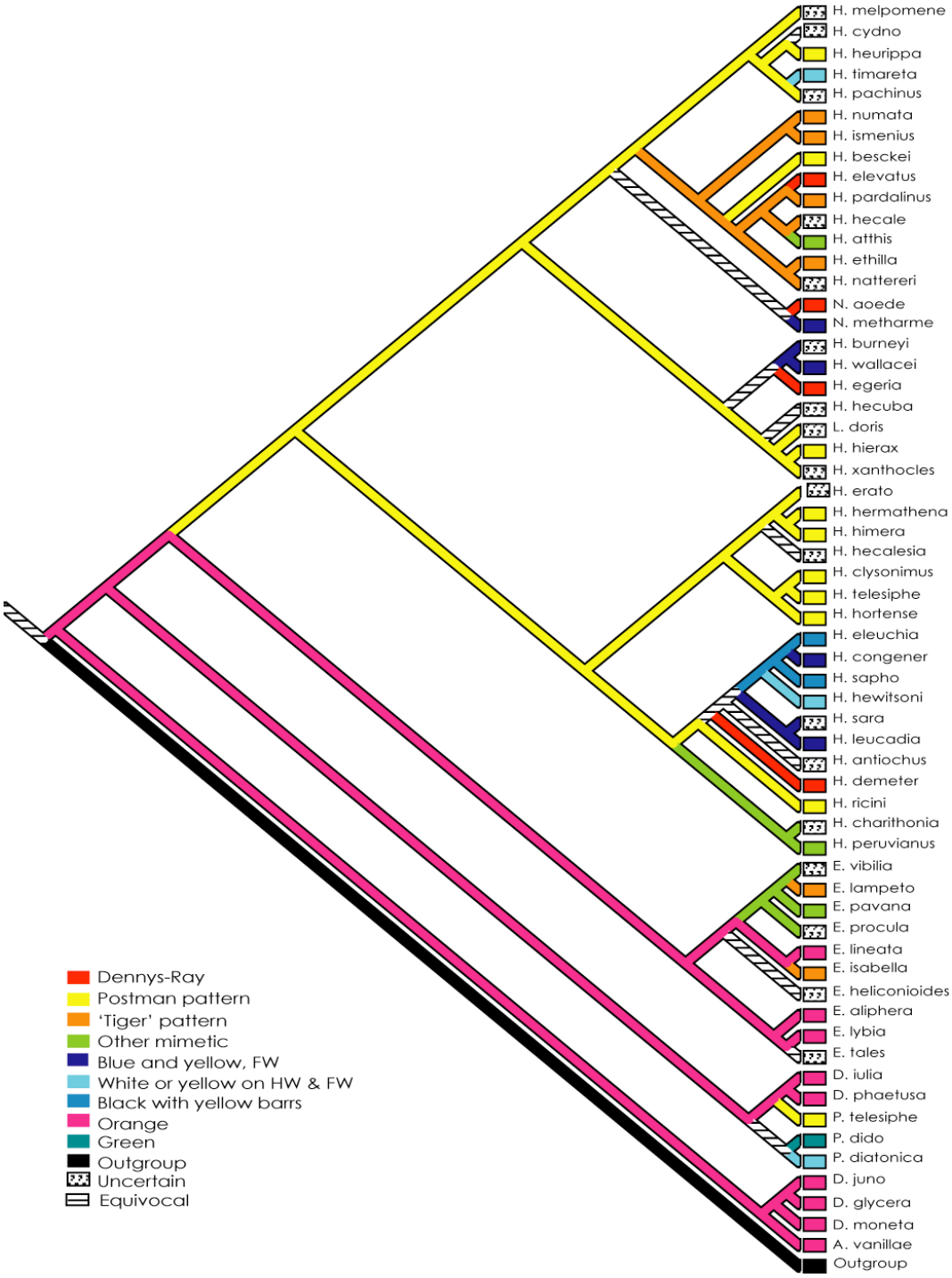
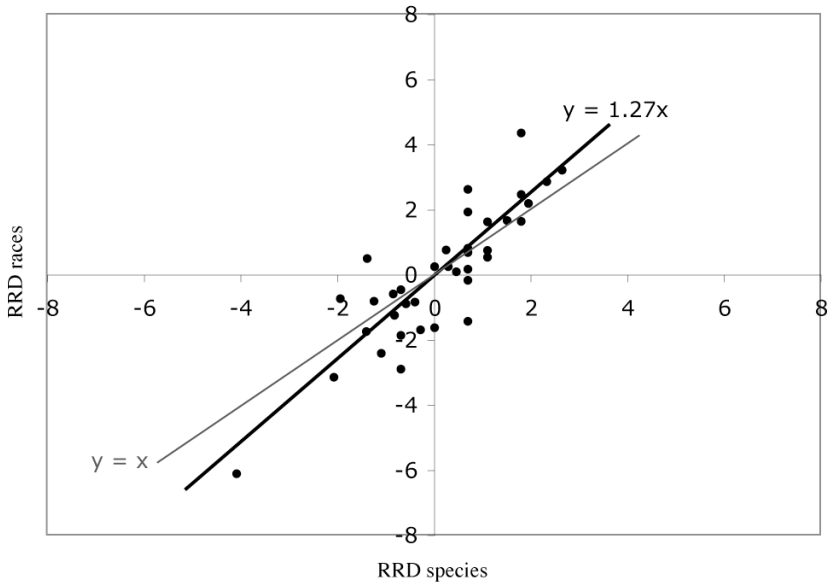


Figure 5. Results of the regression based on MacroCAIC analysis, for the original phylogeny including branch lengths. The species richness contrasts were calculated as the relative rate difference (RRD). Null hypothesis: each species had the same number of races (y=x, slope=1.0). Alternative hypothesis clades with more species tend to have more races (slope significantly greater than 1).



Discussion

This study is the first comprehensive phylogenetic analysis of speciation in a group of well-studied mimetic butterflies. For more than 100 years *Heliconius* have been subject of studies in mimicry, ecology, population genetics, reproductive biology and evolutionary biology (see references in introduction). This study used a well-established phylogenetic hypothesis to understand the geographical mode of speciation of this group and to investigate whether shifts in colour pattern between sister species can play an important role in the diversification of these butterflies.

Geographical mode of speciation in Heliconius

Barraclough and Vogler (2000) found evidence for a predominantly allopatric mode of speciation in birds, fishes and insects. In all cases, intercepts were

distributed around 0, and in most cases, almost all recent nodes displayed zero sympatry. Only two cases were the exceptions to the generalization: the fruit fly genus *Rhagoletis*, which has one very recent node with sympatry of 1.0 (between *Rhagoletis chionanthi* and *Rhagoletis osmanthi*), and the tiger beetle subgenus *Ellipsoptera*, which has one recent node with sympatry around 0.47. They suggested that despite one and possibly two recent sympatric splits in *Rhagoletis*, the pattern in the remainder of the group is not predominantly of sympatric speciation, because the other recent nodes display sympatries near zero.

The *Heliconius* data show an even more extreme pattern than *Rhagoletis*, with many cases of recently diverged species in sympatry (Fig. 3). The results also showed several 'allopatric' recently diverged species pairs. In fact 'allopatric' species pairs are largely parapatric; for example, *H. numata* is replaced by *H. ismenius* in the tall forest of Central America, Venezuela and Ecuador. The *H. cydno* group includes *H. cydno*, *H. pachinus*, *H. heurippa* and *H. timareta*. The latter is the only allopatric species: *H. cydno* is distributed widely across Central America, Colombia, Venezuela, and Ecuador; *H. heurippa* is found in East Colombia; *H. pachinus* is present in SE Costa Rica and SW Panama and *H. timareta* is found in East Ecuador. Finally, the *H. clysonimus* clade includes *H. hortense*, *H. clysonimus* and *H. telesiphe*. *H. clysonimus* is parapatric with *H. hortense* and narrowly sympatric with *H. telesiphe*; these species are distributed in cloud forests of the mountains of Mexico and Nicaragua, Costa Rica, Panama and the North Andes, and finally Southern Andes respectively.

One problem in the Barraclough and Vogler (2000) methodology is that the potential for parapatric speciation is not taken into account. The model considers that at each speciation event, the ranges of descendent species are allocated according to the one of the two chosen modes of speciation. Allopatric speciation involves a split of the area at a uniform random position across any axis of the species range. Sympatric speciation involves the appearance of a new species within the range of an existing species. Thus, parapatric speciation (speciation at geographically contiguous populations) is not considered. However, theory has

shown that rapid parapatric speciation on the time scale of a few hundred to a few thousand generations is probable even when neighbouring subpopulations exchange several individuals each generation (Gavrilets *et al*., 2000). Speciation in parapatry has been proposed for species such *H. cydno* and *H. melpomene* (Jiggins *et al*., 2001b).

The methodology also predicts that if allopatric speciation is the norm, then recently diverged sister taxa will tend to be geographically nonoverlapping, and the degree of overlap is expected to increase between deeper phylogenetic clades as a result of geographical range shifts. On the other hand, if sympatric speciation is predominant, then recent sister taxa will be entirely overlapping, but sister clades deeper in the phylogeny will be more likely to have shifted their ranges and thus overlap less. *Heliconius* data not only showed a high intercept (0.45 that represents geographical overlap at time of speciation), but also showed no increase in the degree of sympatry over time. This suggests that allopatric speciation is not ubiquitous in *Heliconius* indeed almost all recent speciation events are either in parapatry or sympatry.

Another criticism of the Barrachough and Vogler approach is that the current distribution of the recently speciated taxa is not necessarily the distribution at the time of speciation (Losos and Glor, 2003). For example, sympatric species today might have speciated in allopatry or parapatry and subsequently expanded their ranges to come into sympatry. In contrast, today's allopatric species might have speciated in sympatry followed by range shifts that could produce an allopatric pattern today. The *Heliconius* data do suggest fast range movement, but sympatry between so many recently split species seems unlikely to have arisen through range changes occurring after speciation.

Comparative analysis

The mapped colour patterns showed that 10 of 20 sister species differed in colour pattern. However, the numbers of shifts in colour are in fact greater than

suggested by this analysis because of the rather broad colour pattern categories suggested by Brown (1981) and used in this study.

First, some species are classified in the same broad colour pattern category, but in fact differ in their pattern. For instance, *H. charithonia* is present in North America, Venezuela and the Andes and is replaced by *H. peruvianus* in Peru. *H. charithonia* is black with yellow stripes and the pattern is non- mimetic, while *H. peruvianus* mimics *Elzunia pavonii* (Ithomiinae). Another example is the inclusion of both types of postman pattern in just one category (red band in forewing-yellow bar in hindwing vs. yellow band in forewing-red bar in hindwing). Species such as *H. hermathena* and *H. himera* both have postman-like patterns, but *H. himera* is non-mimetic with yellow in the forewing and the red in the hindwing, while *H. hermathena* is the other way round, and is mimetic with *H. erato*. Similarly, *H. clysonimus* (red on hindwing yellow in forewing) is non-mimetic, whereas *H. telesiphe* mimics *Podotricha telesiphe* (red in forewing and yellow in hindwing), although both are classed as postman patterns.

Second, when one species has more than one character state and shares one of these states with the sister species it is not shown as a change. For example, *H. hierax* is non-mimetic and has a postman pattern; its sister species *H. xanthocles* has some races with a similar postman pattern as *H. hierax*, although most of the *H. xanthocles* races belong to the Amazonian dennis-ray ring and look very different to *H. hierax*.

Finally, there are some sister species that do not show changes in colour. These represent geographical replacements or are sympatric species that perhaps occupy different habitats. For instance, the sister species *H. numata* is distributed in the tall forest of Venezuela to south Brazil and is replaced by *H. ismenius* in the Andes and central America; both have the same generalized 'tiger' pattern, although they mimic different ithomiine species. *H. ethilla* and *H. nattereri* belong to the same mimicry ring ('tiger'), but *H. nattereri* is distributed in the humid forest of east Brazil, while *H. ethilla* occupies marginal forest between

Panama and south Brazil. *H. sara* is widespread and occupies forest and edges and *H. leucadia* is present in the Amazon basin but occupies only forest, both are blue and yellow in the forewing.

In summary, is difficult to classify the colour patterns to show a clear picture of changes between sister species due to the great diversity of pattern and the number of races within species. However, most *Heliconius* species present shifts in colour pattern with respect to their sisters although there are also some clear cases of speciation without shifts. These results can be added to the evidence that changes in colour are important in the speciation of *Heliconius*.

I have also shown that across the genus as a whole there is an association of changes in colour pattern with high species diversity. The analysis based on MacroCAIC results is different from previous independent contrast analysis. Usually, such an analysis is used to show that increase in species richness is associated with increase in a trait, for example body size (Orme *et al*., 2002) or size dimorphism and dichromatism (Stuart-Fox and Owens, 2003). Instead I have here shown that there is a tendency for more colour pattern diversity to be associated with higher species richness. The causal factors that underlie differences between clades in tendency to produce pattern diversity remains unclear, but perhaps local species diversity in *Heliconius* might be partly due to the capacity to generate different wing patterns. In *Heliconius* wing pattern diversity seems most connected to local species diversity in groups that use visual signals in courtship (e.g. in the *cydno*/*melpomene*/silvaniform clade). Their success in a local community depends on their capacity to recognize their own pattern and that pattern being sufficiently distinct from close relatives to avoid hybridization (Gilbert, 2003).

Additionally, colour pattern is involved in predator defence through Müllerian mimicry, and as well as being important for choosing mates. Different species of *Heliconius* converge in warning coloration due to mimicry. The most impressive example are the co-mimics *H. erato* and *H. melpomene*, which have each evolved

into almost 30 parapatric colour pattern races (Brown, 1979). Colour pattern radiation in both species is mostly due to divergence at a few genetic loci (Sheppard *et al*., 1985). In a similar way, colour pattern differences between close related species, such us *H. erato* and *H. himera* (Jiggins and McMillan, 1997) and *H. cydno* and *H. melpomene* (Naisbit *et al*., 2003), involve just a few genes of major effect in each case. This suggests that the morphological divergence of species and races has a similar genetic architecture.

Methodological issues of comparative analysis

One methodological problem can be the limits to the inferences that can be described from phylogenetic comparative studies. The most important limitation is that phylogenetic correlations alone cannot clearly identify a causal relationship. However, here I am using traits that are already known to play a role in reproductive isolation (Jiggins *et al*., 2001b). For example, a recent study of mate choice between four parapatric races of *H. melpomene* showed that males are more likely to approach and court their own colour patterns as compared with those of other races (Jiggins *et al*., 2004). This demonstrates that there is a casual link between colour pattern change and speciation.

Inadequate sampling can dramatically alter the conclusion of comparative studies. An early analysis of body size frequency distributions for Australian mammals, that included 30% of species, showed a negative correlation between number of species and body mass. In contrast, the same analysis performed later with the inclusion of more species, showed a positive correlation (Blackburn and Gaston, 1998). In my analysis when missing species were added there was a reduction in the significance of the correlation (Table 1). However, when the *H. egeria* node, and either *Neruda* or *E. emsleyi* node were removed from the analysis, the slope was significantly greater than one. I added the species to the tree following Brown (1981), and the topology of all three nodes goes against the hypothesis that clades with more species have more races. For example, the clade of *N. godmani* and *N. metharme* has 4 races while *N. aoede* alone has 15 races (Fig. 1, Appendix 2.1). If the topology were changed to cluster *N. godmani* with *N. aoede* the results would

support the alternative hypothesis. The exact relationships at this level are not well supported by morphological data for any of the species that were added. Thus, molecular data for these species is needed in order to carry out a complete species level analysis.

Changes in topology highlight another limitation of this study, which is that any conclusions are necessarily dependent on the accuracy of the phylogenetic hypothesis used. The results are likely to be sensitive to taxonomic decisions. Here I have tested an alternative classification, in which both cases when *H. timareta*, *H. pachinus* and *H. heurippa* were included as races rather than species or when races from the Caribbean for *H. charithonia* and *D. iulia* were excluded. The slopes values were different for each comparison but remained significant.

Conclusions

The data supports adaptive speciation in *Heliconius* mediated by colour pattern changes. The possible route of speciation in *Heliconius* is a scenario where racial divergence in parapatry is produced by shifting balance (Joron *et al.*, 1999; Mallet, 1993). Racial divergence might be driven and maintained by selection on wing patterns, because once a warning colour pattern becomes abundant, the local predator community learns to avoid it. This favours the common pattern and causes frequency-dependent selection against rarer patterns. Rare hybrids and recombinants between divergent colour patterns are not recognised as unpalatable and form an adaptive trough between two adaptive peaks (Mallet and Barton, 1989). Each peak represents races of the same species with different colour patterns and they remain separated only by narrow hybrid zones.

Differences in colour pattern among races can cause coevolutionary divergence in colour pattern preferences. For example, male butterflies from four recently diverged parapatric populations of *H. melpomene* were more likely to approach and court their own colour patterns as compared with those of other races (Jiggins *et al.*, 2004). Additionally, races can have some minor differences in ecology or host plants (Benson, 1978) and show some allozyme and molecular differentiation

(Brower, 1994c; Jiggins and Davies, 1998). Warning colour diverges most rapidly in geographic races long before speciation. However, assortative mating together with divergence in warning colour and ecology might lead to speciation between races.

Related species generally differ in colour pattern and in their mimicry rings (Turner, 1976). Mimicry differences between sister species acts as a form of pre- and post-mating isolation. There is strong assortative mating due to colour pattern differences (Jiggins *et al*., 2001b). Rare nonmimetic hybrids are strongly disfavoured, and can be sterile or have a reduction in fertility and mating success (Naisbit *et al*., 2001, 2002). Sister species also show strong differences species in habitat and host plant (Estrada and Jiggins, 2002). Additionally, there is some molecular differentiation (Beltran *et al*., 2002; Flanagan *et al*., 2004). Divergence processes between sister species and that between races of *Heliconius* is similar. This suggests that the processes occurring in parapatry between races is a precursor of speciation in *Heliconius*.

Darwin (1959) was among the first to suggest that some groups of organisms contain disproportionately large number of species, and that the unusual species richness of these groups may be because of characteristics of either the organisms or their environment. The comparative analysis performed in MacroCAIC, suggests that clades with a greater tendency to produce colour patterns (more races) also have greater species diversity (more species). This result supports Darwin's conclusion that "... it has invariably been the case that a larger proportion of the species ... of the large genera present varieties than ... the smaller genera. Moreover, the species of the large genera which present any varieties invariably present a larger average number of varieties than do the species of the smaller genera" (Darwin, 1859). Darwin therefore believed that the formation of varieties or races was part of the same process that eventually leads to the formation of species. This study suggests that Darwin's hypothesis is correct for the genus *Heliconius*.

References

Agapow PM, Isaac NJB (2002) MacroCAIC: revealing correlates of species richness by comparative analysis. *Diversity and Distributions* **8**, 41-43.

Barraclough TG, Harvey PH, Nee S (1995) Sexual selection and taxonomic diversity in Passerine birds. *Proceedings of the Royal Society of London Series B-Biological Sciences* **259**, 211-215.

Barraclough TG, Vogler AP (2000) Detecting the geographical pattern of speciation from species-level phylogenies. *American Naturalist* **155**, 419-434.

Barraclough TG, Nee S (2001) Phylogenetics and speciation. *Trends in Ecology & Evolution* **16**, 391-399.

Bates HW (1862) Contributions to an insect fauna of the Amazon valley. Lepidoptera: Heliconidae. *Transactions of the Linnean Society of London* **23**, 495-566.

Beltran M, Jiggins CD, Bull V*, et al*. (2002) Phylogenetic discordance at the species boundary: Comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Molecular Biology and Evolution* **19**, 2176-2190.

Benson WW (1978) Resource partitioning in passion vine butterflies. *Evolution* **32**, 493-518.

Bissoondath CJ, Wiklund C (1995) Protein content of spermatophores in relation to monandry polyandry in butterflies. *Behavioral Ecology and Sociobiology* **37**, 365-371.

Blackburn TM, Gaston KJ (1998) Some methodological issues in macroecology. *American Naturalist* **151**, 68-83.

Boggs CL (1990) A general model of the role of male-donated nutrients in female insects reproduction. *American Naturalist* **136**, 598-617.

Brower AVZ (1994a) Phylogeny of *Heliconius* butterflies inferred from mitochondrial DNA sequences (Lepidoptera: Nymphalidae). *Molecular Phylogenetics and Evolution* **3**, 159-174.

Brower AVZ (1994b) 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, USA* **91**, 6491-6495.

Brower AVZ (1994c) 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* **91**, 6491-6495.

Brower AVZ (1995) Locomotor mimicry in butterflies? A critical review of the evidence. *Philosophical Transactions of the Royal Society of London (B)* **347**, 413-425.

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.

Brown KS (1979) *Ecologia Geográfica e Evolução nas Florestas Neotropicais. Vol 1.*, Universidade Estadual de Campinas.

Brown KS (1981) The biology of *Heliconius* and related genera. *Annual Review of Entomology* **26**, 427-456.

Bush GL (1969) Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera: Tephritidae). *Evolution* **23**, 237-251.

Darwin C (1859) *On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life* Murray, London.

Estrada C, Jiggins CD (2002) Patterns of pollen feeding and habitat preference among *Heliconius* species. *Ecological Entomology* **27**, 448-456.

Feder JL, Chilcote CA, Bush GL (1988) Genetic differentiation bewteen sympatric host races of apple maggot fly *Rhagoletis pomonella*. *Nature (London)* **336**, 61-64.

Feder JL (1998) The apple maggot fly, *Ragolethis pomonella*: flies in the face of conventional wisdom about speciation? In: *Endless Forms: Species and Speciation* (eds. Howard DJ, Berlocher SH), pp. 130-144. Oxford University Press, New York.

Felsenstein J (1981) Evolutionary trees from DNA sequences - a Maximum-likelihood approach. *Journal of Molecular Evolution* **17**, 368-376.

Felsenstein J (1985) Phylogenies and the comparative method. *American Naturalist* **125**, 1-15.

Flanagan N, Tobler A, Davison A*, et al*. (2004) Historical demography of Mullerian mimicry in the Neotropical *Heliconius* butterflies. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9704-9709.

Garcia-Barros E (2000) Body size, egg size, and their interspecific relationships with ecological and life history traits in butterflies (Lepidoptera : Papilionoidea, Hesperioidea). *Biological Journal of the Linnean Society* **70**, 251-284.

Garland T, Harvey PH, Ives AR (1992) Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.* **41**, 18-32.

Gavrilets S, Li H, Vose MD (2000) Patterns of parapatric speciation. *Evolution* **54**, 1126-1134.

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. In: *Ecology and Evolution Taking Flight: Butterflies as Model Systems* (eds. Boggs CL, Watt WB, Ehrlich PR). Univ. of Chicago Press, Chicago.

Gittleman JL, Purvis A (1998) Body size and species-richness in carnivores and primates. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 113-119.

Grant PR (1981) Speciation and adaptative radiation of Darwin's finches. *American Scientist* **69**, 653-663.

Hsu R, Briscoe A, Chang B, Pierce N (2001) Molecular evolution of a long wavelength-sensitive opsin in mimetic *Heliconius* butterflies (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society* **72**, 435-449.

Isaac NJB, Agapow PM, Harvey PH, Purvis A (2003) Phylogenetically nested comparisons for testing correlates of species richness: A simulation study of continuous variables. *Evolution* **57**, 18-26.

Jiggins CD, McMillan WO (1997) The genetic basis of an adaptive radiation: warning colour in two *Heliconius* species. *Proceedings of the Royal Society of London Series B-Biological Sciences* **264**, 1167-1175.

Jiggins CD, Davies N (1998) Genetic evidence for a sibling species of *Heliconius charithonia* (Lepidoptera; Nymphalidae). *Biological Journal of the Linnean Society* **64**, 57-67.

Jiggins CD, Linares M, Naisbit RE*, et al.* (2001a) Sex-linked hybrid sterility in a butterfly. *Evolution* **55**, 1631-1638.

Jiggins CD, Naisbit RE, Coe RL, Mallet J (2001b) Reproductive isolation caused by colour pattern mimicry. *Nature* **411**, 302-305.

Jiggins CD, Bridle JR (2004) Speciation in the apple maggot fly: a blend of vintages? *Trends in Ecology & Evolution* **19**, 111-114.

Jiggins CD, Estrada C, Rodrigues A (2004) Mimicry and the evolution of premating isolation in *Heliconius melpomene Linnaeus*. *Journal of Evolutionary Biology* **17**, 680-691.

Joron M, Mallet JLB (1998) Diversity in mimicry: paradox or paradigm? *Trends in Ecology & Evolution* **13**, 461-466.

Joron M, Wynne IR, Lamas G, Mallet J (1999) Variable selection and the coexistence of multiple mimetic forms of the butterfly *Heliconius numata*. *Evolutionary Ecology* **13**, 721-754.

Katzourakis A, Purvis A, Azmeh S, Rotheray G, Gilbert F (2001) Macroevolution of hoverflies (Diptera: Syrphidae): an analysis of phylogeny and adult characters. *Journal of Evolutionary Biology* **14**, 219-227.

Kronforst MR, Fleming TH (2001) Lack of genetic differentiation among widely spaced subpopulations of a butterfly with home range behaviour. *Heredity* **86**, 243-250.

Lamas G, Callaghan C, Casagrande MM*, et al.* (2004) *Hesperioidea and Papilionoidea* Association for Tropical Lepidoptera, Gainesville, Florida.

Linares M (1997) The ghost of mimicry past: Laboratory reconstitution of an extinct butterfly 'race'. *Heredity* **78**, 628-635.

Losos JB, Glor RE (2003) Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology & Evolution* **18**, 220-227.

Maddison WP, Maddison DR (1997) MacClade: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, Massachussets.

Mallet J (1986a) Gregarious roosting and home range in *Heliconius* butterflies. *National Geographic Research* **2**, 198-215.

Mallet J (1986b) Hybrid zones of *Heliconius* butterflies in Panama and the stability and movement of warning color clines. *Heredity* **56**, 191-202.

Mallet J, Barton N (1989) Inference from clines stabilized by frequency-dependent selection. *Genetics* **122**, 967-976.

Mallet J (1993) Speciation, raciation, and color pattern evolution in *Heliconius* butterflies: evidence from hybrid zones. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 226-260. Oxford University Press, New York.

Mallet J, Gilbert LE (1995) Why are there so many mimicry rings? Correlations between habitat, behaviour and mimicry in *Heliconius* butterflies. *Biological Journal of the Linnean Society* **55**, 159-180.

Mallet J, McMillan WO, Jiggins CD (1998) Mimicry and warning colour at the boundary between races and species. In: *Endless Forms: Species and Speciation* (eds. Howard DJ, Berlocher SH), pp. 390-403. Oxford University Press, New York.

Mallet J (1999) Causes and consequences of a lack of coevolution in mullerian mimicry. *Evolutionary Ecology* **13**, 777-806.

Mallet J, Joron M (1999) Evolution of diversity in warning color and mimicry: Polymorphisms, shifting balance, and speciation. *Annual Review of Ecology and Systematics* **30**, 201-233.

Mayr E (1963) *Animal species and evolution* Hardvard University Press, Cambrigde, Massachusetts.

McPheron BA, Smith DC, Berlocher SH (1988) Genetic differences between host races of *Rhagoletis pomonella*. *Nature (London)* **336**, 64-66.

Mitter C, Farrell B, Wiegmann BM (1988) The phylogenetic study of adaptative zone: has phytophagy promoted insect diversification? *American Naturalist* **132**, 107-128.

Naisbit RE, Jiggins CD, & Mallet J (2001) Disruptive sexual selection against hybrids contributes to speciation between *Heliconius cydno* and *H. melpomene*. *Proceedings of the Royal Society B* **268**, 1849-1854.

Naisbit RE, Jiggins CD, Linares M, Salazar C, Mallet J (2002) Hybrid sterility, Haldane's rule and speciation in *Heliconius cydno* and *H. melpomene*. *Genetics* **161**, 1517-1526.

Naisbit RE, Jiggins CD, Mallet J (2003) Mimicry: developmental genes that contribute to speciation. *Evolution & Development* **5**, 269-280.

Orme CDL, Isaac NJB, Purvis A (2002) Are most species small? Not within species-level phylogenies. *Proceedings of the Royal Society of London Series B-Biological Sciences* **269**, 1279-1287.

Passos L, Sazima M (1995) Reproductive biology of the *Distylous manettia luteo*-Rubra (Rubiaceae). *Botanica Acta* **108**, 309-313.

Rambaut A, Charleston M (2001) Phylogenetic Tree Editor. Department of Zoology, University of Oxford, Oxford.

Rice WR, Salt GW (1990) The evolution of reproductive isolation as a correlated character under sympatric condition: experimental evidence. *Evolution* **44**, 1140-1152.

Rice WR, Hostert EE (1993) Perspective: Laboratory experiments on speciation: what have we learned in forty years? *Evolution* **47**, 1637-1653.

Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* **19**, 101-109.

Sanderson MJ (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301-302.

Schliewen UK (1994) Sympatric speciation suggests by monophyly of crater lake cichlids. *Nature (London)* **368**, 629-632.

Sheppard PM, Turner JRG, Brown KS, Benson WW, Singer MC (1985) Genetics and the evolution of mullerian mimicry in *Heliconius* butterflies. *Philosophical Transactions of the Royal Society of London (B)* **308**, 433-613.

Speed MP (1999) Robot predators in virtual ecologies: the importance of memory in mimicry studies. *Animal Behaviour* **57**, 203-213.

Srygley RB (1999) Locomotor mimicry in *Heliconius* butterflies: contrast analyses of flight morphology and kinematics. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **354**, 203-214.

Srygley RB, Ellington CP (1999) Estimating the relative fitness of local adaptive peaks: the aerodynamic costs of flight in mimetic passion-vine butterflies *Heliconius*. *Proceedings of the Royal Society of London Series B-Biological Sciences* **266**, 2239-2245.

Stuart-Fox D, Owens IPF (2003) Species richness in agamid lizards: chance, body size, sexual selection or ecology? *Journal of Evolutionary Biology* **16**, 659-669.

Turelli M, Barton N, Coyne JA (2001) Theory and speciation. *Trends in Ecology & Evolution* **16**, 330-343.

Turner JRG (1976) Adaptive Radiation and Convergence in Subdivisions of Butterfly Genus *Heliconius* (Lepidoptera-Nymphalidae). *Zoological Journal of the Linnean Society* **58**, 297-308.

Turner JRG, Mallet JLB (1996) 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* **351**, 835-845.

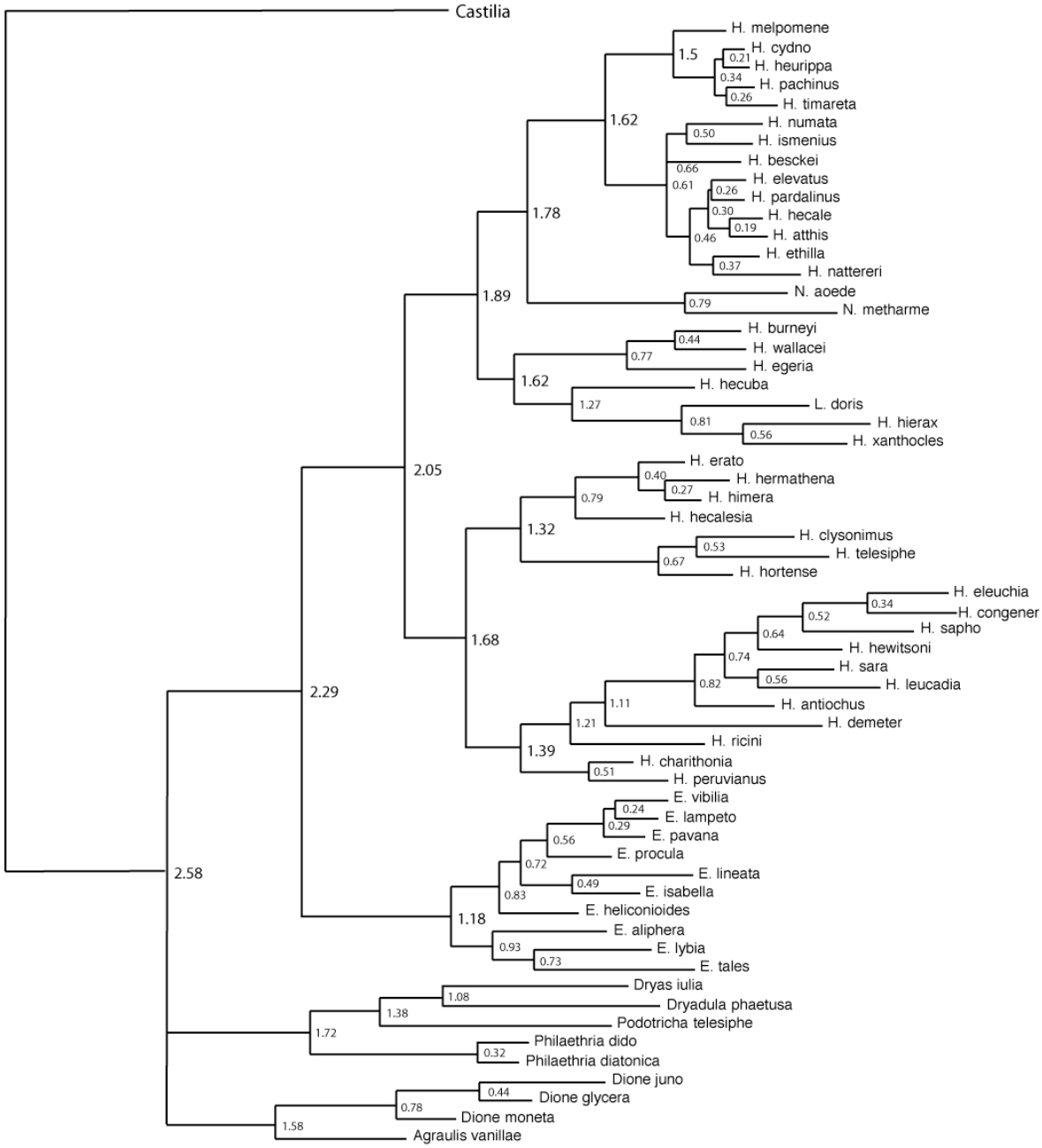
Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution* **16**, 381-390.

Appendices

Appendix 2.1. Number of races per species taken from Lamas *et al.*, (2004). Number of races were used as richness values for MacroCAIC analysis.

Species	No. races	Species	No. races
Heliconius melpomene	30	*Heliconius eleuchia*	3
Heliconius cydno	14	*Heliconius congener*	3
Heliconius heurippa	1	*Heliconius sapho*	5
Heliconius pachinus	1	*Heliconius hewitsoni*	1
Heliconius timareta	2	*Heliconius sara*	10
Heliconius numata	28	*Heliconius leucadia*	4
Heliconius ismenius	8	*Heliconius antiochus*	5
Heliconius besckei	1	*Heliconius demeter*	15
Heliconius elevatus	12	*Heliconius ricini*	2
Heliconius luciana	2	*Heliconius charithonia*	8
Heliconius pardalinus	11	*Heliconius peruvianus*	1
Heliconius hecale	29	*Eueides vibilia*	6
Heliconius atthis	1	*Eueides lampeto*	8
Heliconius ethilla	25	*Eueides pavana*	1
Heliconius nattereri	1	*Eueides libitina*	2
Neruda aoede	15	*Eueides procula*	7
Neruda metharme	3	*Eueides lineata*	1
Neruda godmani	1	*Eueides emsleyi*	2
Heliconius burneyi	10	*Eueides isabella*	13
Heliconius wallacei	6	*Eueides heliconine*	3
Heliconius egeria	7	*Eueides aliphera*	3
Heliconius astraea	2	*Eueides lybia*	6
Heliconius lalitae	2	*Eueides tales*	13
Heliconius hecuba	11	*Dryas iulia*	13
Laparus doris	1	*Dryadula phaetusa*	1
Heliconius hierax	2	*Podotricha telesiphe*	2
Heliconius xanthocles	16	*Podotricha judith*	4
Heliconius erato	29	*Philaetria dido*	6
Heliconius hermathena	6	*Philaetria diatonica*	1
Heliconius himera	1	*Dione juno*	5
Heliconius hecalesia	7	*Dione glycera*	1
Heliconius clysonimus	4	*Dione moneta*	3
Heliconius telesiphe	3	*Agraulis vanillae*	8
Heliconius hortense	1		

Appendix 2.2. Node ages for *Heliconius* phylogeny, estimated using *r8s* program.



Chapter 3

Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies

Abstract

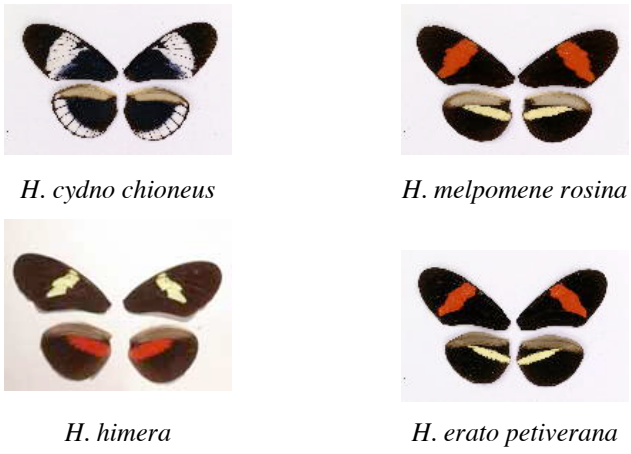
Recent adaptive radiations provide excellent model systems for understanding speciation, but rapid diversification can cause problems for phylogenetic inference. Here I develop two nuclear markers to complement mtDNA sequences and use them to study the phylogenetic relationships between the recently diverged sister species pairs, *H. erato* and *H. himera*, and *H. melpomene* and *H. cydno*. Genealogies were inferred from DNA sequences of mitochondrial cytochrome oxidase (COI and COII 1603bp), and introns of two nuclear genes, the sex-linked triose-phosphate isomerase (*Tpi* ≈550 bp) and the autosomal mannose-phosphate isomerase (*Mpi* ≈450 bp). My results show marked discordance between gene genealogies and species boundaries at different loci. At mtDNA and *Tpi*, *H. erato* and *H. himera* and *H. melpomene* and *H. cydno* cluster together by species. In constrast, the *Mpi* genealogies in both species pairs failed to show structure consistent with the species boundaries. These results suggest that speciation does not necessarily isolate all regions of the genome at the same time. Additionally, I compare these results with recent studies, that use *Mpi* and *Tpi* to explore in detail the relationships between the sister species pairs, *H. melpomene* and *H. cydno* and the demographic history of the comimics *H. melpomene* and *H. erato*. These studies showed that the fast evolving nuclear loci developed here have proven informative in understanding the relationships of sister species and within species radiation.

Introduction

The passion-vine butterflies (Heliconiina) have undergone a recent radiation, with many closely related species that hybridise in the wild at low levels. Natural interspecific hybrids are known between 25-28% of all *Heliconius* species (Mallet

et al., 1998). For example, the sister pairs *H. melpomene* and *H. cydno*, and *H. erato* and *H. himera* (Fig. 1). *H. melpomene* and *H. cydno* are sympatric across most of Central and Andean South America, and coexist with a low level of hybridization. Hybrids are known from across their entire range (Mallet *et al*., 1998). However, the two species have diverged in habitat use that matches the distribution of their co-mimics. *H. cydno* mimics species from the group *H. sapho H. eleuchia*, that inhabit deep forest interiors, while *H. melpomene* mimics *H. erato*, which inhabits open areas (Estrada and Jiggins, 2002; Mallet and Gilbert, 1995). They also differ in their degree of host plant specialization (Smiley, 1978). A second related species pair, *H. himera* and *H. erato* are parapatric. *H. erato*, is present in wet forest all the way through south and central America, while *H. himera* is endemic to dry forest in southern Ecuador and northern Peru. In three known zones of contact, the two species remain distinct, while hybrids are found at low frequency (Jiggins *et al*., 1996). *H. himera* feeds on the same *Passiflora* host plants as *H. erato*, and differs in ecology from its sister species only in its drier habitats and in lacking close mimics (Jiggins *et al*., 1996).

Figure 1. Example of sister species used in this study.



H. cydno chioneus *H. melpomene rosina*

H. himera *H. erato petiverana*

Studies of pre- and post-mating isolation between these sister species suggest that they speciated recently as a result of divergence in habitat and colour pattern (Jiggins *et al*., 2001b; McMillan *et al*., 1997). This might have occurred in

sympatry or parapatry with ongoing gene flow, but the alternative, allopatric divergence cannot be ruled out.

Theory predicts that in rapidly speciating taxa gene genealogies will vary between loci. Therefore, inferring the demographic history of populations or closely related species benefits from comparisons of multiple loci (Edwards and Beerli, 2000). Especially if species are hybridising, there is clearly a need to sample many gene regions. For this reason I here develop new fast-evolving nuclear loci for the study of closely related species.

Gene regions to study species boundary

Mitochondrial DNA (mtDNA) sequences have been widely used to infer intraspecific gene genealogies and determine relationships between closely related species (Brower, 1994; Brower, 1996). A rapid rate of evolution and short coalescence times mean that phylogenies are often well resolved even between recently separated populations and species complexes (Avise, 1994). However, mtDNA studies in *Heliconius* implied that *H. melpomene* is paraphyletic with respect to its sister species *H. cydno* and fails to resolve relationships between *H. erato* and the closely related *H. himera* (Brower, 1996).

Until recently few nuclear genes have been systematically explored for rates of nucleotide substitution and other gene features central to phylogenetic utility (Regier *et al.*, 1998). Nuclear sequences with slower mutations rates and higher effective population sizes are expected to evolve more slowly, but show higher population diversity than mtDNA. Nuclear markers that have been used in *Heliconius*, *wg* (Brower and Egan, 1997), *EF1α*, *dpp* and *ap* (first chapter of this thesis), were not sufficiently variable to be informative at the level of populations and sister species (Brower and DeSalle, 1998). Thus, it is clear from the nuclear loci used, that exons are conserved such that they provide little information between sister species. Therefore, fast-evolving markers are needed to resolve relationships between closely related species.

Introns of nuclear genes are attractive candidates for independent sources of rapidly evolving DNA for phylogentetic studies (Adamczyk *et al*., 1996; Prychitko and Moore, 2000; Rokas and Holland, 2000; Slade *et al*., 1994). Two potential single copy nuclear genes are Triosephospate isomerase (*Tpi*) and Mannose 6 phosphate isomerase (*Mpi*). *Tpi* is an important enzyme for carbohydrate metabolism encoded by a sex-linked nuclear gene in Lepidoptera (Hasson *et al*., 1998). Sequences of this gene have been used in studies of evolution and population genetics in *Drosophila* (Hasson *et al*., 1998), and in understanding the evolutionary origins of introns (Katzourakis *et al*., 2001; Logsden *et al*., 1995). *Mpi*, encoded by an autosomal gene is highly variable in Lepidoptera and the expressed protein is highly polymorphic in Lepidoptera (Beltrán, 1999; Jiggins *et al*., 1997; Raijmann *et al*., 1997).

The main goals of this chapter are to 1) develop two nuclear gene regions to complement sequence data from mtDNA in order to understand relationships between sister species in *Heliconius* and 2) to compare the results obtained here with studies that used these new markers to explore in detail the relationships between the sister species pairs, *H. melpomene* and *H. cydno* and the demographic history of the comimics *H. melpomene* and *H. erato*.

Materials and methods

Taxa at different levels of evolutionary divergence, from geographic populations of the same species, through sister species within *Heliconius*, and finally outgroup taxa were sequenced. The total experimental design included 37 individuals representing 18 *Heliconius* species and 7 outgroup species (Table 1 in Appendix 3.1). Additionally, two new nuclear regions were developed for *Heliconius* (*Tpi* and *Mpi*). New primers, amplification and sequencing methods are described in Appendix 3.1 (see materials and methods). Phylogenetic analyses are also described in Appendix 3.1.

Results and discussion

Two fast-evolving markers were successfully developed for *Heliconius* butterflies (Beltran *et al*., 2002, Appendix 3.1). The nuclear intron sequences evolved at rates similar to those of mitochondrial coding sequences (Fig. 1 in Appendix 3.1). However, the phylogenetic utility of introns was restricted to closely related geographic populations and species due to high levels of indel variation (Appendix 3.1).

Genealogical patterns from mtDNA and two nuclear regions were described for 18 different species of *Heliconius* (Figs. 3, 4 and 5 in Appendix 3.1). Data showed marked discordance between genes in the pattern of relationships between sister species. Discordance between gene genealogies is consistent with models of adaptive speciation with ongoing gene flow as proposed for *Heliconius* species (Appendix 3.1). Since publication of this work, several other studies have used the same loci to explore relationships between sister species in *Heliconius* focusing in the *H. melpomene* and *H. cydno* and *H. erato* and *H. himera* pairs.

Recent studies in *Drosophila* have highlighted the value of multiple gene genealogies in differentiating between speciation models, as allopatric divergence is more likely to produce phylogenetic concordance at different loci (Kliman *et al*., 2000; Wang *et al*., 1997). Thus a multi-locus phylogeny of species pairs can produce insights into speciation processes as well as clarifying the relationships between taxa. Bull *et al*. (2004) used mtDNA, *Tpi*, *Mpi* and two more nuclear genes *Ci* (cubitus interruptus) and *Ddc* (dopa decarboxylase), to test for gene flow between sympatric populations of *H. melpomene* and *H. cydno*.

For mtDNA and *Tpi*, data were consistent with an ancient split between *H. melpomene* and *H. cydno*, and a much more recent split, or ongoing introgression between the two *H. melpomene* races. In contrast, data for *Mpi* demonstrated no fixed allelic differences, with no French Guiana alleles found within Panama clades. This suggested recent introgression between sympatric populations of *H. cydno* and *H. melpomene* in Panama (Bull *et al*., 2004). The *Ci* genealogy showed

a similar pattern to *Mpi*, alleles from allopatric *H. m. melpomene* from French Guiana form a distinct clade, but there is no separation between *H. cydno* and *H. m. rosina* in Panama. The *Ddc* genealogy was inconclusive due to lack of resolution, probably due to a mixture of low levels of divergence and some recombination-induced homoplasy (Bull *et al.*, 2004).

The data showed coalescence after divergence between *H. cydno* and *H. melpomene* in mtDNA and *Tpi* (and probably also *Ddc*), but a lack of coalescence in *Mpi* and *Ci* (Bull *et al.*, 2004). This suggests that there is a selective introgression of *Mpi* and *Ci* alleles since speciation or shared ancestral polymorphism. Genealogies from *Tpi* and mtDNA may be similar because female hybrids between the two species are sterile, an example of Haldane's rule (Naisbit *et al.*, 2002). This prevents transfer of mitochondria (Sperling, 1994) and of *Tpi*; the latter is sex-linked and strongly associated with female sterility in *Heliconius* crosses (Jiggins *et al.*, 2001a; Naisbit *et al.*, 2002). In contrast, introgression of autosomal *Mpi* and *Ci* haplotypes between *H. cydno* and *H. melpomene* may be possible within Panama because they are unlinked to any chromosomes known to be associated with sterility (Bull *et al.*, 2004), but it is difficult to distinguish gene flow from ancestral polymorphism. Speciation does not necessarily isolate all regions of the genome, and therefore cannot be expected to produce instantaneous reciprocal monophyly. The example of *H. cydno* and *H. melpomene* demonstrate that the boundaries between animal species can remain porous to gene flow long after speciation, and introgression can sometimes be an important factor in animal evolution (Bull *et al.*, 2004).

The nuclear data also helped to clarify our understanding of the relationships between *H. himera* and *H. erato* (Flanagan *et al.*, 2004). Genealogies from mtDNA and *Tpi* showed similar topologies, with *H. himera* forming a monophyletic group nested within different geographic races of *H. erato*, while in the *Mpi* genealogy alleles from both species were mixed (Flanagan *et al.*, 2004).

Additionally, the genealogies of *Tpi* and *Mpi* were used to test for the

demographic history of the two comimics *H. erato* and *H. melpomene* (Flanagan *et al.*, 2004). There was evidence for some genetic differences at both nuclear loci between *H. erato* populations separated by the Andes. However, both eastern and western populations retained high levels of genetic diversity, suggesting high historical gene flow. For *H. melpomene* genetic variation was divided into major biogeographic regions, the west of the Andes, Amazonia, and the Guyanan Shield, but the phylogenetic relationships among regional clades varied between loci. For example, the Guyana Shield lineage was basal to the rest of the *H. melpomene* radiation for mtDNA, but was the most derived lineage at *Tpi*. At *Mpi*, the relationship between the three lineages was unresolved. Thus, the phylogeographic patterns at the nuclear loci were not concordant between the comimics (Flanagan *et al.*, 2004).

It has been suggested that *H. melpomene* and *H. erato* are examples of divergence in Pleistocene refuges (Brower, 1996; Sheppard *et al.*, 1985). These sequence data can be used to test the timing and pattern of divergence. In both species, genealogies were most consistent with a history of population growth rather than with constant effective population size. However, the demographic histories suggested different time scales of divergence. In *H. erato* the genealogies of both nuclear loci were most consistent with population growth across the species and within major biogeographic regions and the most-recent common ancestor at both loci occurred within the Pliocene period. Although *H. melpomene* showed similar evidence for population growth, the most-recent common ancestor was much more recent than in *H. erato* at both *Tpi* and *Mpi*, occurring near the Pliocene–Pleistocene boundary (Flanagan *et al.*, 2004).

The understanding of the evolutionary history of the parallel mimetic radiations within *H. erato* and *H. melpomene* was improved significantly by the addition of high-resolution genealogical information for two unlinked nuclear loci developed here (*Tpi* and *Mpi*). The differing evolutionary and demographic histories shown by nuclear data provided strong evidence against simultaneous diversification of *H. erato* and *H. melpomene* wing patterns. Instead, the older origin, greater

abundances, and earlier population expansion in *H. erato* suggested that the wing-pattern radiation occurred first in this species and that *H. melpomene* diverged later to mimic the pre-existing patterns in *H. erato* (Flanagan *et al.*, 2004). They also suggest, specifically in *H. erato*, that divergence took place before the Pleistocene period.

References

Adamczyk JJ, Silvain JF, Prowell DP (1996) Intra- and interspecific DNA variation in a sodium channel intron in Spodoptera (Lepidoptera: Noctuidae). *Annals of the Entomological Society of America* **89**, 812-821.

Avise JC (1994) *Molecular Markers, Natural History and Evolution* Chapman and Hall, London.

Beltran M, Jiggins CD, Bull V, *et al.* (2002) Phylogenetic discordance at the species boundary: Comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Molecular Biology and Evolution* **19**, 2176-2190.

Beltrán M (1999) *Evidencia genética (Alozimas) para evaluar el posible origen híbrido de Heliconius heurippa (Lepidoptera: Nymphalidae)* Masters, Universidad de los Andes.

Brower AVZ (1994) 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* **91**, 6491-6495.

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.

Brower AVZ, Egan MG (1997) Cladistic analysis of *Heliconius* butterflies and relatives (Nymphalidae: Heliconiiti): a revised phylogenetic position for *Eueides* based on sequences from mtDNA and a nuclear gene. *Proceedings of the Royal Society of London Series B-Biological Sciences* **264**, 969-977.

Brower AVZ, DeSalle R (1998) Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. *Insect Molecular Biology* **7**, 73-82.

Bull V, Beltrán M, Bermingham E, *et al.* (2004) Molecular evidence for gene flow between species of *Heliconius*. *In prep*.

Edwards SV, Beerli P (2000) Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**, 1839-1854.

Estrada C, Jiggins CD (2002) Patterns of pollen feeding and habitat preference among *Heliconius* species. *Ecological Entomology* **27**, 448-456.

Flanagan N, Tobler A, Davison A, *et al*. (2004) Historical demography of Mullerian mimicry in the Neotropical *Heliconius* butterflies. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9704-9709.

Hasson E, Wang IN, Zeng LW, Kreitman M, Eanes WF (1998) Nucleotide variation in the triosephosphate isomerase (Tpi) locus of *Drosophila melanogaster* and *Drosophila simulans*. *Molecular Biology and Evolution* **15**, 756-769.

Jiggins CD, McMillan WO, Neukirchen W, Mallet J (1996) What can hybrid zones tell us about speciation? The case of *Heliconius erato* and *H. himera* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society* **59**, 221-242.

Jiggins CD, McMillan WO, King P, Mallet J (1997) The maintenance of species differences across a *Heliconius* hybrid zone. *Heredity* **79**, 495-505.

Jiggins CD, Linares M, Naisbit RE, *et al*. (2001a) Sex-linked hybrid sterility in a butterfly. *Evolution* **55**, 1631-1638.

Jiggins CD, Naisbit RE, Coe RL, Mallet J (2001b) Reproductive isolation caused by colour pattern mimicry. *Nature* **411**, 302-305.

Katzourakis A, Purvis A, Azmeh S, Rotheray G, Gilbert F (2001) Macroevolution of hoverflies (Diptera: Syrphidae): an analysis of phylogeny and adult charaters. *Journal of Evolutionary Biology* **14**, 219-227.

Kliman RM, Andolfatto P, Coyne JA, *et al*. (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**, 1913-1931.

Logsden JM, Tyshenko MG, Dixon C, *et al*. (1995) Seven newly discovered intron positions in the triose-phosphate isomerase gene: Evidence for the introns-late theory. *Proceedings of the National Academy of Sciences, USA* **92**, 8507-8511.

Mallet J, Gilbert LE (1995) Why are there so many mimicry rings? Correlations between habitat, behaviour and mimicry in *Heliconius* butterflies. *Biological Journal of the Linnean Society* **55**, 159-180.

Mallet J, McMillan WO, Jiggins CD (1998) Mimicry and warning colour at the boundary between races and species. In: *Endless Forms: Species and Speciation* (eds. Howard DJ, Berlocher SH), pp. 390-403. Oxford University Press, New York.

McMillan WO, Jiggins CD, Mallet J (1997) What initiates speciation in passion-vine butterflies? *Proceedings of the National Academy of Sciences of the United States of America* **94**, 8628-8633.

Naisbit RE, Jiggins CD, Linares M, Salazar C, Mallet J (2002) Hybrid sterility, Haldane's rule and speciation in *Heliconius cydno* and *H. melpomene*. *Genetics* **161**, 1517-1526.

Prychitko T, Moore W (2000) Comparative evolution of the mitochondrial cytochrome b gene and nuclear β-fibrinogen intron 7 in woodpeckers. *Molecular Biology and Evolution* **17**, 1101-1111.

Raijmann L, Ginkel WE, Heckel DG, Menken SBJ (1997) Inheritance and linkage of isozymes in *Yponomeuta padellus* (Lepidoptera, Yponomeutidae). *Heredity* **78**, 645-654.

Regier JC, Fang QQ, Mitter C, *et al*. (1998) Evolution and phylogenetic utility of the *period* gene in Lepidoptera. *Molecular Biology and Evolution* **15**, 1172-1182.

Rokas A, Holland PWH (2000) Rare genomic changes as a tool for phylogenetics. *Trends in Ecology & Evolution* **15**, 454-459.

Sheppard PM, Turner JRG, Brown KS, Benson WW, Singer MC (1985) Genetics and the evolution of mullerian mimicry in *Heliconius* butterflies. *Philosophical Transactions of the Royal Society of London (B)* **308**, 433-613.

Slade RW, Moritz C, Heideman A (1994) Multiple nuclear gene phylogenies - Application to Pinnipeds and comparison with a mitochondrial DNA gene phylogeny. *Molecular Biology and Evolution* **11**, 341-356.

Smiley JT (1978) Plant chemistry and the evolution of host specificity: new evidence from *Heliconius* and *Passiflora*. *Science* **201**, 745-747.

Sperling FAH (1994) Sex-linked genes and species differences in Lepidoptera. *Canadian Entomologist* **126**, 807-818.

Wang RL, Wakeley J, Hey J (1997) Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* **147**, 1091-1106.

Appendix 3.1.

Paper published by Margarita Beltrán, Chris Jiggins, Vanessa Bull, Mauricio Linares, James Mallet, W, Owen McMillan and Eldredge Bermingham (2002). Phylogenetic discordance at the species boundary: Comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Molecular Biology and Evolution* **19**, 2176-2190.

Abstract

Recent adaptive radiations provide excellent model systems for understanding speciation, but rapid diversification can cause problems for phylogenetic inference. Here we use gene genealogies to investigate the phylogeny of recent speciation in the heliconiine butterflies. We sequenced three gene regions, intron 3 (≈550 bp) of sex-linked triose-phosphate isomerase (*Tpi*), intron 3 (≈450 bp) of autosomal mannose-phosphate isomerase (*Mpi*), and 1603 bp of mitochondrial cytochrome oxidase subunits I and II (COI and COII), for 37 individuals from 25 species of *Heliconius* and related genera. The nuclear intron sequences evolved at rates similar to those of mitochondrial coding sequences, but the phylogenetic utility of introns was restricted to closely related geographic populations and species due to high levels of indel variation. For two sister species pairs, *H. erato*/*H. himera* and *H. melpomene*/*H. cydno*, there was highly significant discordance between the three genes. At mtDNA and *Tpi*, the hypotheses of reciprocal monophyly and paraphyly of at least one species with respect to its sister could not be distinguished. In contrast alleles sampled from the third locus, *Mpi*, showed polyphyletic relationships between both species pairs. In all cases, recent coalescence of mtDNA lineages within species suggests that polyphyly of nuclear genes is not unexpected. In addition, very similar alleles were shared between *melpomene* and *cydno* implying recent gene flow. Our finding of discordant genealogies between genes is consistent with models of adaptive speciation with ongoing gene flow and highlights the need for multiple locus comparisons to resolve phylogeny among closely related species.

Introduction

Mitochondrial DNA sequences have been widely used to infer intraspecific gene genealogies and determine relationships between closely related species. A rapid rate of evolution and short coalescence times mean that phylogenies are often well resolved even between recently separated populations and species complexes (Avise, 1994). However, theory predicts that in rapidly speciating taxa gene genealogies will vary between loci. Hence, inferring the demographic history of populations or closely related species benefits from comparisons of multiple loci (Edwards and Beerli, 2000). In addition to narrowing confidence intervals around demographic parameters such as historical population sizes and divergence times, combined nuclear and mtDNA genealogies will help to test between models of species formation and detect gene flow between taxa (Wang, Wakeley and Hey, 1997; Kliman et al., 2000).

The development of fast-evolving loci to complement mtDNA has not proved easy in spite of an explosive growth of available sequence data. In this study, we develop two non-coding nuclear regions and describe their mode and tempo of evolution relative to the mitochondrial protein-coding genes, Cytochrome oxidase I and II (COI & COII). These loci were developed as a tool to understand speciation and geographical differentiation in *Heliconius* butterflies. The passion-vine butterflies (Heliconiini) have undergone a recent radiation, with many closely related species that hybridise in the wild at low levels (Mallet, McMillan and Jiggins, 1998). The group is well studied due to their bright coloration, impressive mimicry, close relationships with *Passiflora* host plants and geographic variability (Brown, 1981).

Phylogenetic relationships among the Heliconiini have been reworked at least ten times in the last century, using morphological and ecological characters (Brown, 1981; Penz, 1999). More recently phylogenetic hypotheses based on mtDNA and nuclear sequences generally support most of the traditionally recognised species groups and show a number of species pairs separated by < 4% mtDNA sequence

divergence, implying divergence within the last 2 million years (Brower, 1994; Brower and Egan, 1997; Brower and DeSalle, 1998). Two *Heliconius* species in particular, *H. erato* and *H. melpomene*, are recognised as excellent model systems for the study of both intra-specific morphological differentiation and speciation (Sheppard et al., 1985; Brower, 1996b; Mallet, McMillan and Jiggins, 1998). Both species show parallel divergence into more than 20 geographic races across forests in Central and South America and their hybrid zones provide natural systems for the study of selection in the wild (Mallet and Barton, 1989). Furthermore, both taxa have very closely related sister species, which show strong but incomplete reproductive isolation, permitting the study of speciation while hybridisation still occurs (Jiggins et al., 1996; Jiggins et al., 2001b).

There are a number of questions in *Heliconius* systematics and evolution that require the development of rapidly evolving nuclear markers. In particular, (1) Is *Heliconius* monophyletic? Monophyly is supported by recent morphological evidence (Penz, 1999), but has been challenged by mtDNA sequence data, which places *Laparus doris* and the entire genus *Eueides* within *Heliconius* (Brower, 1994). (2) What are the relationships between sister species and populations in the *melpomene* and *erato* species groups. The mtDNA data implies that *H. melpomene* is paraphyletic with respect to its sister species *H. cydno* and fails to resolve relationships between *H. erato* and the closely related *H. himera* (Brower, 1996a). Studies of pre- and post-mating isolation between these sister species suggest they speciated recently as a result of divergence in habitat and colour pattern (McMillan, Jiggins and Mallet, 1997; Jiggins et al., 2001b). This might have occurred in sympatry or parapatry with ongoing gene flow, but the alternative, allopatric divergence cannot be ruled out. Recent studies in *Drosophila* have highlighted the value of multiple gene genealogies in differentiating between speciation models, as allopatric divergence is more likely to produce phylogenetic concordance at different loci (Wang, Wakeley and Hey, 1997; Kliman et al., 2000). Thus a multi-locus phylogeny of species pairs can produce insights into speciation processes as well as clarifying the relationships between taxa.

Our major aim is to develop nuclear gene regions to complement sequence data from the mitochondrial protein-coding genes. The only nuclear DNA marker used in *Heliconius* systematics to date, *wingless,* (Brower and Egan, 1997), evolves slowly and is not sufficiently variable to address questions at or near the species boundary (Brower and DeSalle, 1998). As a consequence, we have developed primers for introns of two genes, *triose-phospate isomerase* (*Tpi*) and *mannose 6-phosphate isomerase* (*Mpi*), known to be highly variable as allozyme loci. Our experimental design takes advantage of the established phylogeny of the Heliconiini. We sampled taxa at different levels of evolutionary divergence, from geographic populations of the same species, through sister species within *Heliconius*, and finally outgroup taxa and compared the relative rate of sequence change in these genes to the mitochondrial cytochrome oxidase I and II (COI & COII) genes. The two mtDNA genes evolve rapidly and are already known to be informative in studies of divergence ranging from intra-specific biogeography of races to relationships among tribes and subfamilies in *Heliconius* (Brower, 1994; 1996b). Specifically, our goals are: (1) to compare evolutionary patterns between mtDNA and nuclear genes (2) between introns and exons within nuclear genes, (3) to determine phylogenetic levels at which these genes are informative and (4) to describe genealogical relationships between races and sister species at the loci studied.

Material and methods

Sampling methods

We sampled 37 individual butterflies, representing 18 *Heliconius* and 7 outgroup species (Table 1). We used a replicated design, with three geographic populations of both the widespread species *H. erato* and *H. melpomene* from Ecuador, Panama and French Guiana, and sister species comparisons between *H. melpomene* and its close relative *H. cydno*, and between *H. erato* and *H. himera*. From each population, two individuals were sampled for the main groups (*melpomene*/*cydno* and *erato*/*himera*) and both alleles of the nuclear genes were sequenced (Table 1).

H. heurippa and *H. pachinus*, two parapatric sister taxa of *H. cydno*, whose specific status remains to be tested, were also sampled. Deeper phylogenetic divisions were represented by a single sequence for more distantly related species both within *Heliconius* and for outgroup taxon within the Heliconiini. Butterflies were collected, preserved in liquid nitrogen and are stored in the Smithsonian Tropical Research Institute in Panama. Wings of voucher specimens are preserved in glassine envelopes. From each individual, 1/3 of a thorax was ground in liquid nitrogen and the genomic DNA was extracted following Harrison, Rand and Wheeler (1987).

Molecular regions and sequencing methods

Mitochondrial DNA

A region of mtDNA spanning the 3' end of cytochrome oxidase subunit I (COI), leucine tRNA, and cytochrome oxidase subunit II (COII) was amplified from individual genomic DNA using PCR. The identity of this region was confirmed by comparison to sequences of *H. cydno* (GenBank accession U0851) and *D. yakuba* (GenBank accession X03240). The region was amplified from genomic DNA in two parts using primers C1-J-2183 and TL2-N-3014 for COI and C1-J-2783 and C2-N-3812 for COII (Table 2). For both pairs of primers we used a cycling profile of 48ºC for 45 sec. and 72ºC for 60 sec. (4 cycles), followed by 94ºC for 45 sec., 52ºC for 45 sec. and 72ºC for 1:30 sec., for 29 cycles. 25µl reactions contained 2µl of DNA, 1x buffer, 2 mM $MgCl_2$, 0.8mM dNTPs, 0.5mM of each primer, 0.025 u/μl of Amplitaq polymerase.

The PCR products were electrophoretically separated on 1.5% low melting point agarose with ethidium bromide (1µg/ml). Bands were cut from the gel and dissolved in gelase. This template was sequenced using the external primers mentioned above and a number of internal primers (Table 2). The 10µl cycle sequence reaction mixture contained 2µl dRhodamine, 2µl Halfterm, 1µl primer, and 5µl template. The cycle profile was 96ºC for 15 sec., then cooling at 1 ºC / sec to 50ºC, then heating at 1 ºC / sec to 60ºC for 4 minutes, repeated for 24 cycles. This product was cleaned over Centrisep columns filled with 700µl G-50

Sephadex and dried. The samples were re-suspended in 0.9µl of a 5:1 deionized formamide: bluedextran/EDTA (pH 8.0) solution, denatured at 90ºC for two min. and loaded into 6% acrylamide gels. Gels were run on an ABI Prism 377 Sequencer (PE Applied Biosystems) for 7 hours.

Nuclear loci development

Primers for the genes of two enzymes, *triose-phosphate isomerase* (*Tpi*) and *mannose-6-phosphate isomerase* (*Mpi*) were developed in the laboratories of D. Heckel and W. O. McMillan. *Tpi* is an important enzyme for carbohydrate metabolism encoded by a sex-linked nuclear gene in Lepidoptera (Logsden et al., 1995). *Mpi* is encoded by an autosomal gene and the expressed protein is highly polymorphic in Lepidoptera (Jiggins et al., 1997; Raijmann et al., 1997; Beltrán, 1999). For both genes, we first designed degenerate PCR primers (Table 2) around conserved amino acid positions identified by comparing published sequence data from *Drosophila* and *Heliothis* for *Tpi* and *Homo sapiens* and *Drosophila* for *Mpi*. For *Mpi*, the region was amplified and sequenced from genomic DNA using these degenerate primers. For *Tpi*, the degenerate primers were then used to amplify the region from *Heliconius* cDNA made via reverse transcriptase from total mRNA. Amplified products in the targeted size range were cloned using pGEM®-T Easy Vector System (Promega) and sequenced as described above. The initial *Heliconius* sequence was aligned to published sequences and *Heliconius* specific primers were designed that consistently amplified the *Tpi* region out of genomic DNA preparations.

For *Tpi*, the primers were situated in exon 3 and 4 of *Heliothis* (GenBank accession U23080) and spanned intron 3 of the *Tpi* gene (Table 2). Previous work has shown that the region amplified is inherited in a Mendelian manner and is sex-linked in both *melpomene* and *erato,* as expected for the *Tpi* allozyme (Jiggins et al., 2001a; Tobler et al., 2001). Double-stranded DNA was synthesised in 25µl reactions containing 2µl of genomic DNA, 1x buffer, 3mM $MgCl_2$, 0.8mM dNTPs, 0.5mM of each primer, and 0.03 u/µl of Taq gold polymerase. DNA was amplified using the following step-cycle profile: 94ºC for 7 min., 94ºC for 45 sec., 58ºC for 45 sec., 72ºC for 1:45 min. for 10 cycles with the annealing temperature

reduced 0.5 ºC per cycle, then 25 cycles with annealing temperature of 53 ºC. The products obtained from genomic DNA were run in a low-melting point agarose gel and the bands excised and dissolved in gelase. For population and sister species samples in the *H. melpomene* and *H. erato* group, the gelase products were cloned to obtain the sequence for each allele, using pGEM®-T Easy Vector System II (Promega). The templates obtained from 3-5 colonies per individual were sequenced as above. For the remaining taxa gelase products were sequenced directly as for mtDNA.

The *Mpi* primers were situated in exons 3 and 4 (*Homo sapiens* GenBank accession #s AF227216 & AF227217) and amplified intron 3 (Table 2). The region amplified by these primers segregates in a Mendelian manner and is inherited in complete linkage with the *Mpi* allozyme in broods of *H. erato* (A. Tobler, personal communication). The 25µl PCR reaction mixture contained 2µl of DNA, 1x buffer, 3mM $MgCl_2$, 0.8mM dNTPs, 0.5 mM of each primer, and 0.03u/μl of Amplitaq. Amplification was carried out using the following step-cycle profile: 94 ºC for 3 min., then 94ºC for 40 sec., 55ºC for 40 sec, 72ºC for 45 sec. for 34 cycles. These products were cloned and sequenced as described above. For *Mpi*, 8 µl of double stranded PCR product was run on a temporal temperature gradient gel using the BioRad 'Dcode' system to confirm that no more than two alleles were amplified per individual. Gels contained 8% acrylamide and 1.75 TAE, and were run from 46 to 53°C at a temperature ramp of 1°C per hour. The results are summarised in Table 1.

Sequence alignment

Chromatograms of mtDNA were edited and base calls checked using SEQUENCHER 3.1 (Gene Codes Corporation, Inc). Following the verification of each sequence for an individual, protein-coding regions were aligned in SEQUENCHER across all taxa. Introns of nuclear sequences were aligned in Clustal W (Higgins and Sharp, 1988) and then adjusted manually to increase overall similarity. Due to strong sequence divergence and many indels, introns

could only be aligned within the *melpomene*/silvaniform and the *erato*/*sapho* group (Supplementary material).

Taq error and allele selection

Sequencing of cloned PCR products is known to produce errors due to both single base substitution and recombination occurring during the PCR reaction (Wang and Wang, 1997; Bracho, Moya and Barrio, 1998; Kobayashi, Tamura and Aotsuka, 1999). We minimised this problem by sequencing at least three, and in most cases five clones per individual and selecting a 'consensus' sequence for each allele based on the parsimonious assumption that single base taq induced error was likely to occur only once. In all cases comparison of different sequences inferred to represent a single allele were compatible with this assumption. For *Tpi*, where more than one clone was compared with the deduced allele sequence, the distribution of errors was 15, 12, 9 and 2 clones with 0, 1, 2 and 3 single base pair errors respectively. For *Mpi* the same distribution was 12, 11, 1 and 1. If undetected, such errors are unlikely to affect phylogenetic analyses as they would most likely be autapomorphic. One case of recombination between the two alleles in a single individual was also observed. In that case, when clone sequences were aligned the pattern was that expected following a single recombination event, which presumably occurred during the PCR reaction and was by chance selected for sequencing (Wang and Wang, 1997).

Phylogenetic analysis

The nucleotide sequences for protein-coding mtDNA and nuclear DNA sequences were checked for reading-frame errors and termination codons and translated to functional peptide sequences in MacClade 4.0 (Maddison and Maddison, 1997). This program was also used to compute various sequence statistics including nucleotide transformation frequencies and variation among codon positions. Phylogenetic analyses were performed with PAUP* version 4.0b8 (Swofford, 2000). Models of sequence evolution were compared by means of likelihood ratio tests (G-tests) using ModelTest 3.04 (Posada and Crandall, 1998). PAUP* was then used to search for the maximum likelihood (ML) tree, based on the best fit

model and parameter estimates given by ModelTest using a heuristic search with tree bisection reconnection (TBR). Confidence in each node was tested using the likelihood-ratio test implemented by PAUP*, which sequentially collapses branch lengths to zero and compares resulting topologies to the ML tree. For comparison, maximum parsimony (MP) trees were obtained using a heuristic search with TBR branch swapping. The consensus tree was calculated using majority rule. Confidence in each node was assessed by bootstrapping (1000 replicates, heuristic search with TBR branch swapping). In figures 3-4, branches were collapsed if they had less than 95% likelihood support (see above), bootstrap support of less than 50, or were not supported by an indel character.

In order to test specific hypotheses, alternative *a priori* scenarios were compared to the ML tree using the method of Shimodaira and Hasegawa (Shimodaira and Hasegawa, 1999; Goldman, Anderson and Rodrigo, 2000) and implemented using either PAUP* version 4.0b8 or the program *SHTests v1.0* written by A. Rambaut (when compared both programs gave very similar results). For each species pair (i.e. *melpomene*/*cydno* and *erato*/*himera*), four tree topologies were compared in the same test: reciprocal monophyly, paraphyly of species 1 with species 2 monophyletic, paraphyly of species 2 with species 1 monophyletic and polyphyly. In each case the shortest tree for each scenario was sought using MacClade, starting with the ML tree as presented in Figures 2-4. In order to test phylogenetic hypotheses at the generic level, we also constructed a data matrix including our exon and mtDNA data with previously published *wingless* sequences. Twenty species were included in this analysis, 19 of which had a complete data set for all genes and one outgroup, *Dryas iulia*, which was complete except for the 66bp of *Mpi* exon.

Results and discussion

Patterns of molecular evolution in Heliconius mtDNA, Tpi and Mpi genes

Mitochondrial DNA

The final aligned mitochondrial sequences yielded 1603 characters including nucleotides and gaps from 37 individuals (GenBank accession numbers

AF413672-AF413708). This represents 822 bp of the cytochrome oxidase I gene corresponding to position 2191 to 3009 of the *D. yakuba* sequence (X03240), the complete leucine-tRNA gene (70bp), and 711 bp representing the entire COII coding sequence. Our mtDNA analysis is based on 659 bp more of the COI gene than used by Brower (1994). Most length variation was concentrated in the tRNA-leucine, which shows a 1-bp indel (in *H. elevatus* and *H. hecale*) and a 7-bp insertion immediately following the COI termination codon (in *H. charithonia* and *H. ricini*). This region shows length variation in other Lepidopteran species (Brower, 1994; Caterino and Sperling, 1999). A codon deletion in COI (in the third codon of our alignment, corresponding to amino acid position # 243 in *D. yakuba,* X03240) was shared by all *Eueides* species, and in the COII gene we observed three nearby codon deletions, at amino acid position #126 in *Dryadula phaetusa*, #127 in *H. sara* and at position #129 in *Dryas iulia* (see also Brower 1994).

Of the 1603 nucleotide sites examined, 440 (27%) were variable. Most of the variation occurred in the protein-coding regions. Twenty five percent of sites were phylogenetically informative in COI and 22% in COII as compared to 6% in the tRNA leucine. Within coding regions the total variability and variation per position is similar between the two CO subunits (Table 3). The GC content of COI + COII was 26%, comparable to that observed in other insects (Caterino and Sperling, 1999). As expected >75% of the variation occurs at third positions (Table 3) and transitions were almost ten times more frequent than transversions (Table 4).

Tpi

We obtained 155bp of *Tpi* exon sequence corresponding to positions 425 to 455 (31 bp) of Exon 3 and 536 to 659 (124 bp) of Exon 4 in *Heliothis* (U23080) for 46 alleles from 33 individuals representing 21 species (GenBank accession numbers AF413752-AF413797, notation Tpi-1 and Tpi-2 refer to alleles). There was no length variation between the *Heliothis* and the *Heliconius* exon sequences. In the *Tpi* exons, 55 of the 155bp observed were variable. The *Tpi* exon had a higher GC

content (49%) than the mitochondrial CO genes, similar to previously observed differences between mtDNA and *wingless* (Brower and DeSalle, 1998). The general pattern of divergence was quite similar between mtDNA and the *Tpi* exon, with most of the substitutions concentrated in third positions (Table 3).
The *Tpi* intron 3 exhibited considerable length variation and alignment was only possible between closely related species. In the *melpomene*/silvaniform group, the *Tpi* intron 3 was completely absent in *H. elevatus*, and ranged from 345 bp in *H. melpomene cythera* (allele 8073#2) to 457 bp in *H. hecale*. In the *erato*/*sapho* group (Table 3) the same intron varied from 216 bp in *H. hecalesia* to 244 bp in *H. charithonia* and *H. clysonymus*. As we were unable to align the *Tpi* intron across groups, analysis of this region was restricted to within group comparisons. The *melpomene*/silvaniform group (alignment 1) and *erato*/*sapho* group (alignment 2) *Tpi* alignments are given in the supplementary material and are available upon request or at http//nmg.si.edu/cj/.

Mpi

We obtained 66bp of *Mpi* exon for 43 alleles from 31 individuals representing 19 butterfly species (GenBank accession numbers AF413709-AF413751, notation Mpi-1 and Mpi-2 refer to alleles). Twenty-seven bp corresponded to positions 1601-1627 of exon 3 in the *Homo sapiens* reference sequence (AF227216), and 39 bp to positions 417-455 of exon 4 in the *H. sapiens* sequence (AF227217). As with *Tpi* exons, there were no indels in the *Mpi* exons and GC content was 40%. However, in stark contrast to *Tpi*, *Mpi* exons showed an unexpectedly high rate of non-synonymous changes. Across 40 alleles from 19 species sequenced for both *Mpi* and *Tpi*, there were 6 amino acid replacements across 51 codon positions in *Tpi*, compared to 13 amino acid changes across only 31 AA positions in *Mpi*. The different amino acid replacement pattern between the two genes coincides with the higher proportion of 1st and 2nd position changes recorded for *Mpi* (Table 2). *Mpi* intron 3 also exhibited considerable length variation and again intron alignment was only possible between closely related species. In the *melpomene*/silvaniform group, *Mpi* intron 3 ranged from 114 bp in *H. m. melpomene* (allele 437#1) to 388 bp in *H. pachinus*; however, length variation

within even single populations of each species was almost as great. In the *erato*/*sapho* group (Table 3) the same intron varied from 411 bp in *H. e. hydara* (allele 440#1) to 464 bp in *H. himera* (allele 8076#2). The difficulty of alignment between distantly related taxa meant that all intron-based analysis was restricted to within group comparisons. The *melpomene*/silvaniform group (alignment 3) and *erato*/*sapho* group (alignment 4) *Mpi* alignment files are given in the supplementary material and are available upon request or at http//nmg.si.edu/cj/.

Patterns of length variation at nuclear loci

There were two kinds of length variation in *Mpi* and *Tpi* introns. First, there was variation in the length of repeated elements. Some of these repeats were homopolymers, for example there was a poly-A repeat starting at position 34 of the *Tpi* intron in the *melpomene*/sylvaniform group that varied from five to nine bases in length (Supplementary material Alignment # 1). In other cases there were repeated elements that were interrupted or complex. For example in the *Mpi* intron there was a microsatellite region showing extensive variation around a CACACA motif. In general this type of indel variation provided little phylogenetic information and could not be easily mapped onto the *Mpi* and *Tpi* likelihood trees (see below).

Second, we observed insertions and deletions that were not associated with repeated elements. In virtually all cases, these indels were synapomorphic and therefore provided additional support for nuclear gene phylogenies based only on nucleotide substitutions (Figures 3-4). For example, a 7 bp insertion at *Tpi* position 263 in alignment 1 was common to three *melpomene* alleles from French Guiana that represent a monophyletic clade based on the sequence data (Figure 3a). However in the *erato* group, some indels at both *Tpi* and *Mpi* were not concordant with the sequence-based maximum likelihood trees. Nonetheless, a tree constrained such that each *Tpi* indel represented a unique evolutionary event was not a significantly worse fit to the *Tpi* sequence data than the initial ML tree (SH test; Delta = 14.83, p = 0.06). We therefore recalculated the ML tree with a constraint based on the two discordant indels, which forced *erato* to be

paraphyletic with respect to *himera*, and grouped the *H. erato cyrbia* alleles to form a monophyletic group (Fig. 3b). However, in the case of *Mpi* in the *erato* group there was no way to make the phylogeny consistent with all indels. The 7bp deletion at position 473 (shown as a filled oval in Figure 4b) grouped alleles 2842#1 (*H. himera*), 590 (*H. hecalesia*) and 2923 (*H. charithonia*), whereas a 2bp deletion at position 222 grouped *H. hecalesia* and *H. charithonia* with *H. sara*, *H. eleuchia* and *H. sapho*. The 2bp indel was consistent with the ML tree, while the 7bp indel appears to be genuinely homoplasious (Fig. 4b). Another apparently homoplasious indel, a 5bp deletion at position 353 (open oval), may be an artefact. The 2980#1 allele appears to be a cloned PCR recombinant between 2980#2 and an allele which was not sampled, similar to 2981#1, which would explain this homoplasy (Supplementary material).

Natural recombination was, in contrast, apparently rare. Recombination could generate high levels of homoplasy, reducing phylogenetic signal. However, maximum-parsimony based consistency indices calculated for the two species groups pictured in Figures 3 and 4, (excluding uninformative characters) were higher in the *Tpi* (0.77 and 0.65 respectively) and *Mpi* (0.80 and 0.64) regions than the in mitochondrial CO genes (0.36, Fig. 2). Thus, homoplasy was lower in the nuclear genes.

Rate comparisons between mitochondrial and nuclear genes

Rates of molecular evolution in the intron region of both *Tpi* and *Mpi* were high and similar to the mitochondrial coding genes that are typically used in inter- and intraspecific phylogenetic studies (Brower, 1994; 1996b). Mean divergence between the *melpomene*/*cydno* group and the silvaniform species is 5.4% at mitochondrial COI and COII, 4.1% at *Tpi* and 5.4% at *Mpi*; divergence for the same genes between the *sapho* and *erato* clades is 9%, 7.9% and 12.3% respectively.

Rates of evolution between genes were compared by plotting pairwise uncorrected sequence divergence of mitochondrial COI and COII against nuclear allele

divergence for the same individuals (Fig. 1). In the case of heterozygotes, one allele was randomly selected for each individual. Sequence divergence was very similar between CO and *Tpi* when all codon positions are included (data not shown). When only CO 3rd positions were compared to *Tpi*, the intron was evolving at approximately one third the rate of CO suggesting that the neutral substitution rate was faster in the mitochondrion (Fig. 1). In contrast, *Mpi* showed cases of high divergence between individuals carrying closely related mtDNA haplotypes, in part reflecting much higher within-population polymorphism at this nuclear locus. Nonetheless, there was a crude correlation between CO and *Mpi* distance, with the slope suggesting that the *Mpi* intron is evolving at approximately half the synonymous rate of CO.

Models of sequence evolution

There was little concordance between the models of sequence evolution selected for each region, in part reflecting the size of the data matrices, which differed due to sequence length and the number of taxa that could be aligned. Both COI and COII were best explained by the 6 parameter general-time-reversible model of nucleotide substitution (GTR+I+G) (Yang, 1994). The two genes were therefore combined in order to estimate parameters of sequence evolution (Table 4) and for phylogenetic analysis (Figure 2). For the nuclear sequence data, simpler models with fewer parameters were an equally good fit to the data, which is perhaps unsurprising given the smaller size of these data matrices. The following models were selected (Table 4): i) The three parameter TrN+G model (Tamura and Nei, 1993) for *Tpi* in the *melpomene*/silvaniform group (alignment 1); ii) the HKY+G model (Hasegawa, Kishino and Yano, 1985) for the *Tpi* (alignment 2) and *Mpi* (alignment 4) data in the *erato*/*sapho* group; and iii) the even simpler F81+G (Felsenstein, 1981) model for the *Mpi* data in the *melpomene*/silvaniform group (alignment 3). In all cases the nuclear regions had lower estimated transition: transversion rate ratios as compared to mitochondrial sequences. In contrast, the estimated gamma-shape parameters (α) were similar in all gene regions and varied from 0.41 to 1.0 (Table 4). Thus, even within the intron sequences, there

was considerable site-to-site rate variation suggesting some site-specific constraints on sequence evolution.

Phylogenetic analysis

The level of variation observed in the nuclear loci produced well-supported genealogical relationships between alleles sampled from closely related species and geographic races of the same species (Figure 3 and 4). Phylogenetic resolution was somewhat less at nuclear loci compared to the mtDNA (compare Fig. 2 with Figs 3-4), primarily due to the shorter length of these sequences. In addition, there is a large amount of phylogenetically-informative indel variation in our intron sequences, which increases confidence in the tree topologies presented (Figures 3 and 4). We found only two cases where indels were not concordant with our ML trees, both in the *Mpi* data for the *erato* group. These discrepancies may indicate recombination, but the generally low levels of homoplasy in our data suggest that recombination is rare and does not inhibit phylogenetic reconstruction.

Nonetheless, the high rate of molecular evolution, particularly the extensive length variation in the intron of both gene regions, restricts the phylogenetic utility of these loci to very closely related species or populations. Sequences are impossible to align among more distantly related taxa and, even among those sequences that can be aligned, large deletions occasionally destroy phylogenetic signal. The use of introns therefore proves to be something of a lottery, since the region of interest may have been lost in some taxa. However, levels of variation were fairly high even in the short regions of exon examined in this study, suggesting that longer fragments of nuclear coding sequence would provide considerable phylogenetic information for resolving deeper level phylogenetic relationships (Brower and DeSalle, 1998; Regier et al., 1998).

Analysis of Heliconius and related genera

Although our nuclear sequence data are not ideally suited for phylogenetic analysis at the generic level, these data, in combination with the additional

mitochondrial sequence, warrant a reassessment of some outstanding questions in *Heliconius* phylogeny. The ML tree based on COI+COII accords reasonably well with previous phylogenetic analyses based on sequence morphological and ecological data (Brown, 1981; Brower, 1994; Brower and Egan, 1997) and includes three taxa not studied in previous molecular analyses, *Heliconius hecalesia*, *Heliconius hierax* and *Eueides lineata*.

In our ML tree (Figure 2) based on COI and COII, *Eueides* is a sister taxon to *Heliconius*, as expected on the basis of ecology, morphology and combined mitochondrial and nuclear gene sequences (Brower and Egan, 1997). This contrasts with an earlier parsimony analysis of a smaller portion of the mitochondrial COI and COII regions, which placed *Eueides* within the genus *Heliconius*. *Mpi* was uninformative at this level as it could not be amplified in outgroup taxa, but trees inferred from the 155 bp of *Tpi* exon data using both ML and parsimony methods show a monophyletic *Heliconius* clade with *Eueides* outside *Heliconius*, with bootstrap support of only 56% (data not shown). However, the overall support for reciprocal monophyly of the two genera is weak. Analysis based either on the mtDNA data alone (SH Test; Delta = 7.27, p = 0.47), or including previously published *wingless* sequences with our *Mpi*, *Tpi* and CO data (SH Test; Delta = 8.21, p = 0.22) both fail to exclude the possibility that *Eueides* falls within *Heliconius*.

Laparus doris has traditionally been considered a monotypic sister genus to *Heliconius* (Brown, 1981; Penz, 1999). However, our mtDNA data provide strong statistical support for previous results (Brower and Egan 1997) based on COI, COII and *wingless* which group *doris* within *Heliconius* in a clade which includes *H. wallacei* (Figure 2; SH Test; Delta = 62.47, p = 0.002). Combined analysis of nuclear coding sequence from the *wingless*, *Tpi* and *Mpi* genes provide no further resolution, as the likelihood test based on these data fails to reject a tree in which *doris* lies outside *Heliconius* (SH Test; Delta = 2.42, p = 0.41).

Within *Heliconius*, there was an unresolved trichotomy at the base of the genus (Figure 2), with no compelling support at COI and COII for two major divisions (Riffarth, 1901; Emsley, 1965). *Heliconius sapho*, *H. sara*, and *H. eleuchia* and *H. charithonia* are considered to share the characteristic pupal-mating behavior with the *erato* group (Brown, 1981; Lee et al., 1992). However, our ability to align *Tpi* and *Mpi* introns in *H. ricini*, *H. charithonia*, *H. sapho* and allies, with those of the *erato*/*himera* group, but not the *melpomene*/silvaniform group gives support for the traditional grouping.

The melpomene *group*

The evolutionary relationships between *H. melpomene* and the *H. cydno* group (*cydno*, *pachinus* and *heurippa*) varied depending on the gene region analysed (Table 5). *Heliconius melpomene* and *H. cydno* were reciprocally monophyletic sister taxa in the mtDNA ML tree, with an average uncorrected divergence of 3.3%. This contrasts with the paraphyly of *melpomene* with respect to *cydno* described previously on the basis of a shorter region of the COI + COII genes (Brower, 1996b). In that study, paraphyly was due to a single branch, *melpomene* from French Guiana, placed basally to *cydno* and the rest of *melpomene*. Parsimony analysis also supports reciprocal monophyly of *melpomene* and *cydno* and branches leading to both groups show high bootstrap support (Figure 2). However, it is interesting to note that despite strong bootstrap support, maximum-likelihood topology tests were only able to reject the hypothesis that both groups were polyphyletic (Table 5). This is in part because the monophyly of *H. melpomene* was supported by only two transitions. More surprisingly, 4 transition substitutions (3, C-T and 1, A-G) and 3 A-T transversions provide no significant support for *cydno* group monophyly (Table 5); the SH test would seem to overly conservative.

In contrast, the ML tree inferred from sequence variation in the *Mpi* intron (Fig 3b) shows *melpomene* and *cydno* paraphyletic with respect to the silvaniforms *H. elevatus*, *H. hecale*, and *H. numata*, traditionally considered more distantly

related, with a mean divergence of 5.4% between the two groups. However support for this topology is not significant with the silvaniforms constrained to be outgroups (SH Test; Delta = 24.72, p = 0.22). Nonetheless, there was very strong support for polyphyly of alleles between *melpomene* and the *cydno* group (*cydno*, *heurippa* and *pachinus*). All topologies in which either species is monophyletic with respect to the other are significantly worse than the ML tree (Table 5). Average divergence between the species was 6.4%. However, very similar alleles were also shared between these two species, differing by as little as 3 bp (between *melpomene* allele 8074#2 and *cydno* allele 553#2). There was a great deal of allelic diversity within populations, with some very divergent alleles sampled from within the Ecuador and Panama populations of *melpomene* and *cydno*. More detailed surveys of these populations show that the divergent sequences represented here by alleles 811#2 and 8#1 are present at low frequency in both *melpomene* and *cydno* in Panama, and thus our results do not represent sequencing or mistaken identity errors (V. Bull, personal communication).

Lastly, the *Tpi* ML tree (Fig 3a) showed *melpomene* as a monophyletic group, with *cydno* alleles basal and paraphyletic with respect to *melpomene*, with an average uncorrected divergence of 3.3% between the species. The silvaniforms form a distinct clade and were used to root the *melpomene* + *cydno* clade, with an average divergence of 4.1% between the two groups. However, between *melpomene* and *cydno* there was very little phylogenetic resolution provided by the *Tpi* data and it was not possible to reject the alternative hypotheses of polyphyly or paraphyly between the species at this locus (Table 5).

The erato *group*

The phylogenetic relationship between *H. erato* and *H. himera* also varied depending on the gene region examined. In the mtDNA tree, *himera* forms a monophyletic group nested within different geographic races of *H. erato* (Figure 2). Mean uncorrected divergence between *himera* and all other *erato* was 3.2%. Nonetheless, a tree where the two species were forced to be reciprocally monophyletic could not be rejected in likelihood tests (Table 5). A similar pattern

was demonstrated in our *Tpi* sequence data (Table 5). In the *Tpi* tree (Fig 4a) *H. himera* forms a monophyletic group within *erato*. The unconstrained ML tree showed both species monophyletic, and support for the paraphyly of *erato* came from a 18bp deletion at position 211 (Figure 4a) shared by all *himera* and *erato* alleles with the exception of *H. erato petiverana* 2980#2. In contrast, in the *Mpi* genealogy (Fig 4b) alleles from both species are clearly mixed. There were highly divergent alleles in both groups and topologies that forced either species to be monophyletic were not supported by the data (Table 5).

Conclusions regarding relationships between sister species

In conclusion, genetic variation in the maternally inherited mitochondrial genome and the sex-linked *Tpi* gene clustered together by species. *Heliconius melpomene* and *H. cydno* showed an average of 3.3% uncorrected sequence divergence at COI and COII genes and 3.1% uncorrected divergence at the *Tpi* region. *Heliconius erato* and *H. himera* showed similar levels of divergence at both loci (3.2% at both CO and *Tpi*). Assuming a rate of mitochondrial evolution of 1.1-1.2 % per lineage per million years (Brower, 1994) this suggests that both species pairs diverged from each other within the last one and a half million years. However, both *H. erato* and *H. melpomene* are more widely distributed than their respective sister species and neither the mtDNA or *Tpi* genealogies exclude the possibility that geographic populations of one species are paraphyletic with respect to the sister species (Figs. 2 - 4). In contrast, the *Mpi* genealogies in both species pairs failed to show structure consistent with species boundaries, despite considerable resolution and well supported nodes within the trees (Figs. 3b and 4b). Our data, therefore, show marked discordance between gene genealogies and species boundaries at different loci.

Why are the genealogies discordant?

Gene trees are not the same as species trees and the discordance between allelic genealogies observed may simply reflect differences in expected coalescence times among loci (Tajima, 1983; Pamilo and Nei, 1988; Takahata, 1989; Nichols, 2001). Of the three loci examined in this study, the autosomal *Mpi* locus has the

largest genetic effective population size and is expected to harbour ancestral shared variation for longer time periods than sex-linked or maternally inherited genes such as *Tpi* and CO respectively. In particular, maternally inherited mitochondrial genes will coalesce on average 4 times faster than autosomal genes. We can use the mtDNA data to predict the coalescence time of nuclear genes following the three-times rule (Tavaré, 1984; Palumbi, Cipriano and Hare, 2001). In the absence of gene flow, coalescence theory predicts nuclear allele coalescence within a species for a majority of autosomal nuclear loci when the branch length leading to the mtDNA sequences of that species is three times longer than the average within species mtDNA sequence diversity (or two times as long for an X-linked locus). Our data show that all species in the *erato*/*himera* and *melopomene*/*cydno* groups have mtDNA branch length to diversity ratios less than 2 (Table 6). Therefore, a majority of nuclear loci are expected to show polyphyletic patterns between these species pairs.

Although we cannot rule out a purely neutral explanation of the polyphyly observed at nuclear loci based on coalescence theory, the striking pattern at *Mpi* of high diversity within and shared alleles between species suggests, in addition, balancing selection and inter-specific gene flow. Unusually high allelic diversity in allozyme studies of *Heliconius* (Jiggins et al., 1997; Beltrán, 1999) and direct evidence that *Mpi* is under balancing selection in other organisms (Schmidt, 2001), both suggest that this locus might be under selection. Furthermore, there were an unusually high number of amino acid changes in our *Mpi* exon sequence (13 changes among 40 alleles from 19 species), supporting the suggestion that this variation is maintained by natural selection. The high variation in the intron region could therefore be explained by hitchhiking with nearby selected exon polymorphism.

Even in the absence of nuclear allele coalescence within species, it should still be possible to detect the signature of recent introgression between species. Because the three gene regions studied here are evolving at similar rates (Fig. 1), the

observation of far more closely related alleles between *melpomene* and *cydno* at *Mpi* than at the other two loci likely results from recent introgression between species. Indeed, both of the sister species pairs studied here are known to hybridise in the wild. *H. melpomene* and *cydno* are broadly sympatric with hybrids forming perhaps 0.1% of overlapping populations (Jiggins et al., 2001b). Furthermore, there is hybrid female sterility between the species, associated with the sex-linked *Tpi* gene in one direction of backcross (Naisbit et al., 2001). This hybrid sterility might be expected to prevent introgression at both mtDNA and *Tpi* (Sperling, 1994), while allowing the flow of some nuclear genes. At least for *melpomene* and *cydno*, the pattern observed is therefore consistent with the genetic architecture of reproductive isolation. It seems likely that both gene flow and balancing selection have played a role: the latter could maintain high allelic diversity within populations while the former would favour the 'capture' of new alleles following rare inter-specific hybridisation.

In conclusion, phylogenies of recently evolved species, which may still exchange genes, are inevitably difficult to resolve. The markers studied here provide well-supported gene genealogies, but the general lack of concordant reciprocal monophyly between closely related species and the disagreements between loci highlights the importance of multiple locus comparisons in resolving sister species relationships. Fast-evolving nuclear genes such as those described here are likely to become an important tool for phylogenetic analysis. Furthermore, it is clear that biologically and ecologically relevant species may sometimes not be recognisable under phylogenetic (Cracraft, 1989) or genealogical species concepts (Baum and Shaw, 1995). Speciation does not necessarily isolate all regions of the genome, and therefore cannot be expected to produce instantaneous reciprocal monophyly.

Acknowledgements

We would like to thank the Autoridad Nacional del Ambiente in Panama and the Ministerio del Ambiente in Ecuador for permission to collect butterflies; Maribel

Gonzalez, Nimiadina Gomez, Oris Sanjur and Alexandra Tobler for help in the laboratory. This work was funded by the Smithsonian Tropical Research Institute, and Natural Environment Research Council (U.K.) and National Science Foundation (USA) grants to JM and WOM respectively.

References

Avise, J. C. 1994. Molecular Markers, Natural History and Evolution. Chapman and Hall, London.

Baum, D. A. and K. L. Shaw. 1995. Genealogical perspectives on the species problem. 289-303 in Experimental and Molecular Approaches to Plant Biosystematics (Hoch, P. C. and A. G. Stephenson, eds.). Missouri Botanical Garden, St. Louis, Missouri.

Beltrán, M. 1999. Evidencia genética (Alozimas) para evaluar el posible origen híbrido de *Heliconius heurippa* (Lepidoptera: Nymphalidae). Masters, Universidad de los Andes, Santafé de Bogota.

Bracho, M. A., A. Moya and E. Barrio. 1998. Contribution of Taq polymerase-induced errors to the estimation of RNA virus diversity. Journal of General Virology 79: 2921-2928.

Brower, A. V. Z. 1994. Phylogeny of *Heliconius* butterflies inferred from mitochondrial DNA sequences (Lepidoptera: Nymphalidae). Molecular Phylogenetics and Evolution 3: 159-174.

Brower, A. V. Z. 1996a. A new mimetic species of *Heliconius* (Lepidoptera: Nymphalidae), from southeastern Colombia, as revealed by cladistic analysis of mitochondrial DNA sequences. Zoological Journal of the Linnean Society 116: 317-332.

Brower, A. V. Z. 1996b. Parallel race formation and the evolution of mimicry in *Heliconius* butterflies: a phylogenetic hypothesis from mitochondrial DNA sequences. Evolution 50: 195-221.

Brower, A. V. Z. and R. DeSalle. 1998. Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of *wingless* as a source of characters for phylogenetic inference. Insect Molecular Biology 7: 73-82.

Brower, A. V. Z. and E. G. Egan. 1997. Cladistic analysis of *Heliconius* butterflies and relatives (Nymphalidae: Heliconiiti): a revised phylogenetic position for *Eueides* based on sequences from mtDNA and a nuclear gene. Proceedings of the Royal Society of London, B 264: 969-977.

Brown, K. S. 1981. The biology of *Heliconius* and related genera. Annual Review of Entomology 26: 427-456.

Caterino, M. S. and F. A. H. Sperling. 1999. *Papilio* phylogeny based on mitochondrial cytochrome oxidase I and II genes. Molecular Phylogenetics and Evolution 11: 122-137.

Cracraft, J. 1989. Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. 28-59 in Speciation and its Consequences (Otte, D. and J. A. Endler, eds.). Sinauer Associates, Sunderland, Mass.

Edwards, S. V. and P. Beerli. 2000. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54: 1839-1854.

Emsley, M. G. 1965. Speciation in *Heliconius* (Lep., Nymphalidae): morphology and geographic distribution. Zoologica, New York 50: 191-254.

Felsenstein, J. 1981. Evolutionary Trees from DNA-Sequences - a Maximum-Likelihood Approach. Journal of Molecular Evolution 17: 368-376.

Goldman, N., J. P. Anderson and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. Systematic Biology 49: 652-670.

Harrison, R. G., D. M. Rand and W. C. Wheeler. 1987. Mitochondrial-DNA Variation in Field Crickets across a Narrow Hybrid Zone. Molecular Biology and Evolution 4: 144-158.

Hasegawa, M., H. Kishino and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 21: 160-174.

Higgins, D. G. and P. M. Sharp. 1988. Clustal - a Package for Performing Multiple Sequence Alignment on a Microcomputer. Gene 73: 237-244.

Jiggins, C. D., P. King, W. O. McMillan and J. Mallet. 1997. The maintenance of species differences across a *Heliconius* hybrid zone. Heredity 79: 495-505.

Jiggins, C. D., M. Linares, R. E. Naisbit, C. Salazar, Z. Yang and J. Mallet. 2001a. Sex-linked hybrid sterility in a butterfly. Evolution 55: 1631-1638.

Jiggins, C. D., W. O. McMillan, W. Neukirchen and J. Mallet. 1996. What can hybrid zones tell us about speciation? The case of *Heliconius erato* and *H. himera* (Lepidoptera: Nymphalidae). Biological Journal of the Linnean Society 59: 221-242.

Jiggins, C. D., R. E. Naisbit, R. L. Coe and J. Mallet. 2001b. Reproductive isolation caused by colour pattern mimicry. Nature (London) 411: 302-305.

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman, A. J. Berry, J. McCarter, J. Wakeley and J. Hey. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex. Genetics 156: 1913-1931.

Kobayashi, N., K. Tamura and T. Aotsuka. 1999. PCR error and molecular population genetics. Biochemical Genetics 37: 317-321.

Lee, C. S., B. A. McCool, J. L. Moore, D. M. Hillis and L. E. Gilbert. 1992. Phylogenetic study of heliconiine butterflies based on morphology and restriction analysis of ribosomal RNA genes. Zoological Journal of the Linnean Society 106: 17-31.

Logsden, J. M., M. G. Tyshenko, C. Dixon, J. D. Jafari, V. K. Walker and J. D. Palmer. 1995. Seven newly discovered intron positions in the triose-phosphate isomerase gene: Evidence for the introns-late theory. Proceedings of the National Academy of Sciences, USA 92: 8507-8511.

Maddison, W. P. and D. R. Maddison (1997). MacClade: Analysis of phylogeny and character evolution. 3.07 ed. Sunderland, Massachussets: Sinauer Associates.

Mallet, J. and N. H. Barton. 1989. Strong natural selection in a warning color hybrid zone. Evolution 43: 421-431.

Mallet, J., W. O. McMillan and C. D. Jiggins. 1998. Mimicry and warning colour at the boundary between races and species. 390-403 in Endless Forms: Species and Speciation (Howard, D. J. and S. H. Berlocher, eds.). Oxford University Press, New York.

McMillan, W. O., C. D. Jiggins and J. Mallet. 1997. What initiates speciation in passion vine butterflies? Proceedings of the National Academy of Sciences, USA 94: 8628-8633.
Naisbit, R., C. D. Jiggins, M. Linares, C. Salazar and J. Mallet. 2001. Haldane's rule for sterility in crosses between *Heliconius cydno* and *H. melpomene*. Genetics in press.
Nichols, R. 2001. Gene trees and species trees are not the same. Trends in Ecology & Evolution 16: 358-364.
Palumbi, S. R., F. Cipriano and M. P. Hare. 2001. Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. Evolution 55: 859-868.
Pamilo, P. and M. Nei. 1988. Relationships between gene trees and species trees. Molecular Biology and Evolution 5: 568-583.
Penz, C. M. 1999. Higher level phylogeny for the passion-vine butterflies (Nymphalidae; Heliconiinae) based on early stage and adult morphology. Zoological Journal of the Linnean Society 127: 277-344.
Posada, D. and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817-818.
Raijmann, L., W. E. Ginkel, D. G. Heckel and S. B. J. Menken. 1997. Inheritance and linkage of isozymes in *Yponomeuta padellus* (Lepidoptera, Yponomeutidae). Heredity 78: 645-654.
Regier, J. C., Q. Q. Fang, C. Mitter, R. S. Peigler, T. P. Friedlander and M. A. Solis. 1998. Evolution and phylogenetic utility of the *period* gene in Lepidoptera. Molecular Biology and Evolution 15: 1172-1182.
Riffarth, H. 1901. Die Gattung Heliconius Latr.: Neu bearbeitet und Beschreibung neuer Formen. Berl. Entomol. Zeit. 46: 25-183.
Schmidt, P. 2001. The effects of diet and physiological stress on the evolutionary dynamics of an enzyme polymorphism. Proceedings of the Royal Society of London, B 268: 9-14.
Sheppard, P. M., J. R. G. Turner, K. S. Brown, W. W. Benson and M. C. Singer. 1985. Genetics and the evolution of muellerian mimicry in *Heliconius* butterflies. Philosophical Transactions of the Royal Society of London (B) 308: 433-613.
Shimodaira, H. and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution 16: 1114-1116.
Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reacion primers. Annals of the Entomological Society of America 87: 651-702.
Sperling, F. A. H. 1994. Sex-linked genes and species differences in Lepidoptera. Canadian Entomologist 126: 807-818.
Swofford, D. L. (2000). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). 4. ed. Sunderland, Massachusetts: Sinauer Associates.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105: 437-460.

Takahata, N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics 122: 957-966.

Tamura, K. and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10: 512-526.

Tavaré, S. 1984. Line of descent and genealogical processes and their applications in population genetic models. Theoretical Population Biology 26: 164-199.

Tobler, A., N. Flanagan, C. D. Jiggins, D. G. Heckel and W. O. McMillan. 2001. A genetic map of allozyme, AFLP and color pattern genes in *Heliconius erato* and *H. himera*. Genetics in prep.

Wang, G. C. Y. and Y. Wang. 1997. Frequency of formation of chimeric molecules is a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. Applied and Environmental Microbiology 63: 4645-4650.

Wang, R. L., J. Wakeley and J. Hey. 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. Genetics 147: 1091-1106.

Yang, Z. 1994. Estimating the pattern of nucleotide substitution. Journal of Molecular Evolution 39: 105-111.

Table 1. Heliconiina taxa included in the study. ID numbers are Smithsonian collection numbers and are prefixed by 'stri-b-'. 'No. alleles' refers to the number of distinct allelic sequences identified for each individual. Alignment numbers indicate aligned groups and correspond to numbers given in Tables 3 & 4. For *Mpi*, alleles of most *melpomene* and *erato* group individuals were run on TTGE gels to confirm the number of alleles present. Results are indicated in the No. alleles column, where * indicates an agreement between the results derived from cloning and TTGE, or in two cases of disagreement the number of bands visible on the TTGE gel is given in brackets.

GROUP	**SPECIES**				**No. alleles**			**Alignment no.**	
	ID No.	**Sex**	**Name**	**Location**	CO	*Tpi*	*Mpi*	*Tpi*	*Mpi*
melpomene/ cydno	553	M	*H. cydno chioneus*	Panama	1	2	2	**1**	**3**
	570	M	*H. cydno chioneus*	Panama	1	2	1	**1**	**3**
	811	M	*H. melpomene rosina*	Panama	1	2	2*	**1**	**3**
	544	M	*H. melpomene rosina*	Panama	1	2	1*	**1**	**3**
	436	M	*H. melpomene melpomene*	French Guiana	1	2	1*	**1**	**3**
	437	M	*H. melpomene melpomene*	French Guiana	1	2	2*	**1**	**3**
	8074	M	*H. melpomene cythera*	Ecuador	1	1	2*	**1**	**3**
	8073	M	*H. melpomene cythera*	Ecuador	1	2	2*	**1**	**3**
	8036	M	*H. pachinus*	Panama	1	1	1	**1**	**3**
	8	M	*H. heurippa*	Colombia	1	1	2	**1**	**3**
silvaniforms	346	F	*H. numata*	French Guiana	1	1	1	**1**	**3**
	503	M	*H. elevatus*	French Guiana	1	1	1	**1**	**3**
	665	M	*H. hecale*	Panama	1	1	1	**1**	**3**
erato/himera	2980	M	*H. erato petiverana*	Panama	1	2	2*	**2**	**4**
	2981	M	*H. erato petiverana*	Panama	1	2	2(1)	**2**	**4**
	440	M	*H. erato hydara*	French Guiana	1	2	1*	**2**	**4**
	442	M	*H. erato hydara*	French Guiana	1	1	1(2)	**2**	**4**
	2861	M	*H. erato cyrbia*	Ecuador	1	2	2*	**2**	**4**
	8075	M	*H. erato cyrbia*	Ecuador	1	2	**2***	**2**	**4**
	2842	M	*H. himera*	Ecuador	1	1	1*	**2**	**4**
	8076	M	*H. himera*	Ecuador	1	2	2*	**2**	**4**
	8037	M	*H. clysonymus*	Panama	1	1	1	**2**	**4**
	590	M	*H. hecalesia*	Panama	1	1	1	**2**	**4**
charithonia	2923	M	*H. charithonia*	Ecuador	1	1	1	**2**	**4**
sara/sapho	850	M	*H. sara*	Panama	1	1	1	**2**	**4**
	1180	M	*H. ricini*	Panama	1	1		**2**	
	842	F	*H. eleuchia*	Panama	1	1	1	**2**	**4**
	536	M	*H. sapho*	Panama	1	1	2	**2**	**4**
Primitive	8154	M	*H. hierax*	Ecuador	1	-	-	**-**	**-**
	286	M	*H. wallacei*	French Guiana	1	1	1	**-**	**-**
	846	M	*L. doris*	Panama	1	1	1	**-**	**-**
Eueides	2991	M	*E. lineata*	Panama	1	-	1	**-**	**-**
	320	M	*E. vibilia*	Panama	1	1	-	**-**	**-**
	555	F	*E. aliphera*	Panama	1	1	-	**-**	**-**
	556	M	*E. lybia*	Panama	1	-	1	**-**	**-**
Dryadula	2940	M	*D. phaetusa*	Panama	1	-	-	**-**	**-**
Dryas	293	F	*D. iulia*	Panama	1	1	-	**-**	**-**

Table 2. Primers used to amplify *Heliconius* mtDNA and nuclear DNA. Cytochrome oxidase (COI and COII) 3' end positions are given relative to *Drosophila yakuba* (X03240). Positions for *Tpi* and *Mpi* are given relative to *Heliothis* (U23080) and *Homo sapiens* (AF227216 and AF227217) respectively. Mitochondrial primers were designed in the Harrison laboratory (Simon et al., 1994). Nuclear primers were designed by D. Heckel and W.O. McMillan.

Gene	Name	Position	Sequence (5' to 3')
COI	C1-J-2183	2183	CAACATTTATTTTGATTTTTTGG
	C1-J-2441	2442	CCAACAGGAATTAAAATTTTTAGATGATTAGC
	C1-J-2783	2783	TAGGATTAGCTGGAATACC
tRNA-Leucine	TL2-N-3014	3014	TCCAATGCACTAATCTGCCATATTA
	TL2-J-3039	3039	TAATATGACAGATTATATGTAATGGA
COII	C2-J-3297	3297	TGAACTATTTTACC(A/G/T)GC
	C2-N-3812	3812	CATTAGAAGTAATTGCTAATTTACTA
Tpi	Tpi-1	424	GGTCACTCTGAAAGGAGAACCATCTT
	Tpi-2	660	CACAACATTTGCCCAGTTGTTGCCAA
Mpi	Mpi 4+	1600	TTTAAGGTGCTCTATATAAGRAARGC
	Mpi 5-	456	TTCTGGTTTGTGATTTGGATCYTTRTA

Table 3. Variability in each gene region. The species included in each alignment are given in Table 1. The number of variable sites (excluding length variation) is given for the whole gene region and then by codon position ($1^{st}/2^{nd}/3^{rd}$). To facilitate comparison between gene regions, the CO data are given for all taxa and separately those taxa included in each intron alignment. Note that phylogenetic analysis for nuclear genes was carried out using combined intron + exon sequence for alignments 1-4.

Gene	Region	No. species	No. indivs.	No. alleles	Alignment no.	No. base pairs	No. variable sites
COI	Exon	25	37	-	All taxa	822	240 39/7/194
	Exon	7	13	-	*melpomene*/ silvaniform	822	56 6/1/49
	Exon	9	15	-	*erato*/*sapho*	822	137 14/3/120
tRNA		25	37	-	All taxa	70	8
		7	13	-	*melpomene*/ silvaniform	70	1
		9	15	-	*erato*/*sapho*	70	3
COII	Exon	25	37	-	All taxa	711	184 30/10/144
	Exon	7	13	-	*melpomene*/ silvaniform	711	27 2/1/24
	Exon	9	15	-	*erato*/*sapho*	711	87 10/6/71
Tpi	Exon	21	33	46	All taxa	155	55 9/7/39
	Intron	7	13	20	1 *melpomene*/ silvaniform	499	52
	Intron	9	15	21	2 *erato*/*sapho*	249	81
Mpi	Exon	19	31	43	All taxa	66	31 7/5/19
	Intron	7	13	19	3 *melpomene*/ silvaniform	415	90
	Intron	8	15	20	4 *erato*/*sapho*	500	179

Table 4. Best-supported models of molecular evolution and estimated parameter values for the mitochondrial CO I and II genes and the *Tpi* and *Mpi* combined intron and exon regions. Differing numbers of parameters are estimated for each locus, according to the best-fit model selected via likelihood using model test (Posada and Crandall, 1998). Similarly values for those parameters allowed to vary under the best-supported evolutionary model were estimated using likelihood. I is the proportion of invariant sites and α is the shape parameter for the gamma distribution. Parameter values of 1 are fixed values in the respective model rather than estimated parameters. To facilitate comparison between gene regions, the CO data are given for all taxa and separately for those taxa included in each intron alignment.

Gene region	Alignment no.	Model	Base composition				I	α	Substitution rates						
									Tr		Tv				Tr/Tv
			a	c	g	t			a-g	c-t	a-c	a-t	c-g	g-t	
COI + COII	All taxa included	GTR+I+G	0.35	0.1	0.1	0.45	0.6	0.66	21.6	90.5	6.75	4.17	0	1	-
	1 and 3	GTR+I+G	0.34	0.12	0.12	0.42	0.8	2.65	72.4	198.2	5.4	13.8	0	1	-
	2 and 4	GTR+I+G	0.34	0.12	0.12	0.42	0.7	0.85	63.4	244.4	18.7	7.40	0	1	-
Tpi	1	TrN+G	0.36	0.14	0.18	0.32	0	0.41	1.35	4.37	1	1	1	1	-
	2	HKY+G	0.34	0.15	0.19	0.32	0	0.79	-	-	-	-	-	-	1.91
Mpi	3	F81+G	0.37	0.14	0.12	0.37	0	1.0	-	-	-	-	-	-	1
	4	HKY+G	0.4	0.14	0.13	0.33	0	0.89	-	-	-	-	-	-	1.02

Table 5. Results of SH tests of alternative *a priori* hypotheses for the phylogenetic relationships between *melpomene*/*cydno* and *erato*/ *himera*. For this test, the maximum-likelihood trees for the *erato*/ *himera* and *melpomene*/*cydno* species groups were simultaneously evaluated against constrained trees representing the alternative phylogenetic hypotheses. Tests were carried out in Paup* using the RELL method with 1000 replicates (Swofford, 2000). The log likelihood (p-value) is given for each tree and topologies which are not consistent with the data at p≤ 0.05 are shown with a *.

		Alignment no.	Reciprocal monophyly	Paraphyly I	Paraphyly II	Polyphyly
melpomene (I) and *cydno* (II)	COI + COII		**3252.7 (best)**	3253.0 (0.70)	3259.0 (0.39)	3304.3 (0.00)*
	Tpi	**1**	1442.1 (0.12)	1448.9 (0.05)*	**1425.3 (best)**	1441.6 (0.18)
	Mpi	**3**	1571.1 (0.00)*	1546.0 (0.00)*	1520.4 (0.00)*	**1378.2 (best)**
erato (I) and *himera* (II)	COI + COII		5046.2 (0.76)	**5046.2 (best)**	5056.7 (0.048)*	5066.4 (0.021)*
	Tpi	**2**	1403.01 (0.77)	**1441.0 (best)**	1427.1 (0.047)*	1433.4 (0.013)*
	Mpi	**4**	2561.7 (0.00)*	2546.5 (0.00)*	2519.5 (0.002)*	**2481.9 (best)**

Table 6. Evaluation of the three-times rule for *Heliconius* species. Branch lengths were obtained from a neighbor-joining tree constructed from mitochondrial sequences using the Kimura two-parameter model. The mtDNA diversity is the average pairwise distance within each group. To predict monophyly we used Figure 2 in Palumbi, Cipriano and Hare (2001), for a sample size of 5 alleles, adjusting for the smaller population size of sex-linked genes in the case of *Tpi*.

Group	**Mitochondrial branch length (l)**	**mtDNA diversity (d)**	**Coalescence ratio (l/d)**	**Predicted % monophyletic**		**Nuclear monophyly observed?**	
				Tpi	***Mpi***	***Tpi***	***Mpi***
melpomene	0.009	0.008	1.1	30%	20%	yes	no
cydno	0.01	0.015	0.69	20%	10%	no	no
erato petiverana + cyrbia	0.012	0.008	1.54	50%	30%	no	no
himera	0.011	0.008	1.45	50%	30%	yes	no

Figure 1. Uncorrected genetic distances between *Heliconius Tpi* and *Mpi* sequences plotted as a function of divergence in 3rd codon positions of the mitochondrial CO I and II genes. Open triangles show comparisons in the *erato*/*sapho* clade (alignments 2 and 4) and closed diamonds the *melpomene*/silvaniform clade (alignments 1 and 3). For nuclear genes a single allele was randomly chosen to represent each individual.

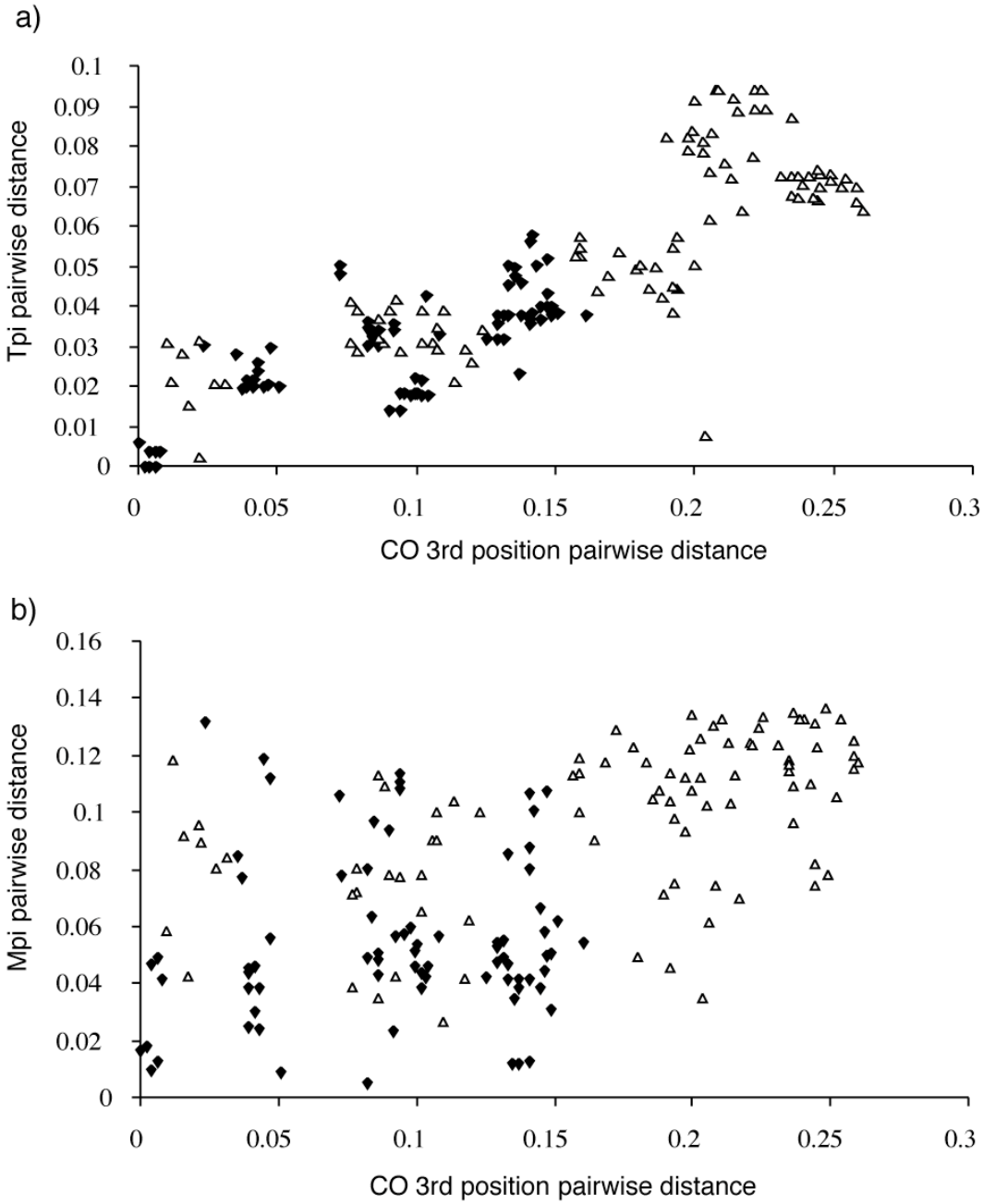


Figure 2. Maximum-likelihood phylogeny for Heliconiina species based on combined mitochondrial COI and COII sequences. All branches were collapsed which were not supported with either >95% confidence using likelihood ratio tests (implemented in Paup*), or >50 bootstrap support, or by phylogenetically informative indel characters. Branch lengths were estimated using likelihood. Bold vertical lines show synapomorphic indel variation. Values on branches show parsimony bootstrap support for the equivalent node, after 1000 replicate bootstraps. Branches without parsimony bootstrap support reflect differences between the MP and ML tree. Sample numbers correspond to those given in Table 1. P = Panama, E = Ecuador, G = French Guiana and C= Colombia.

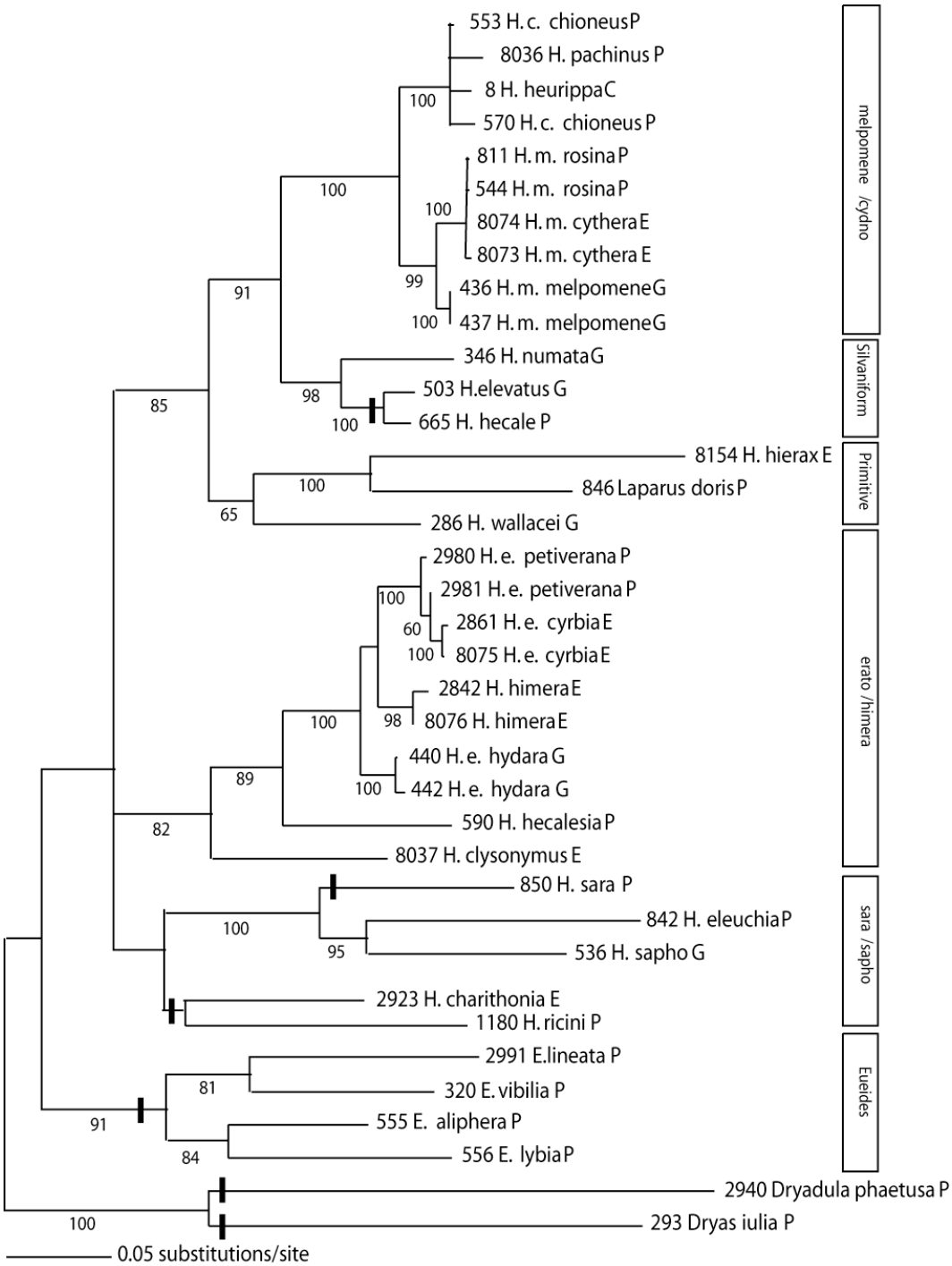
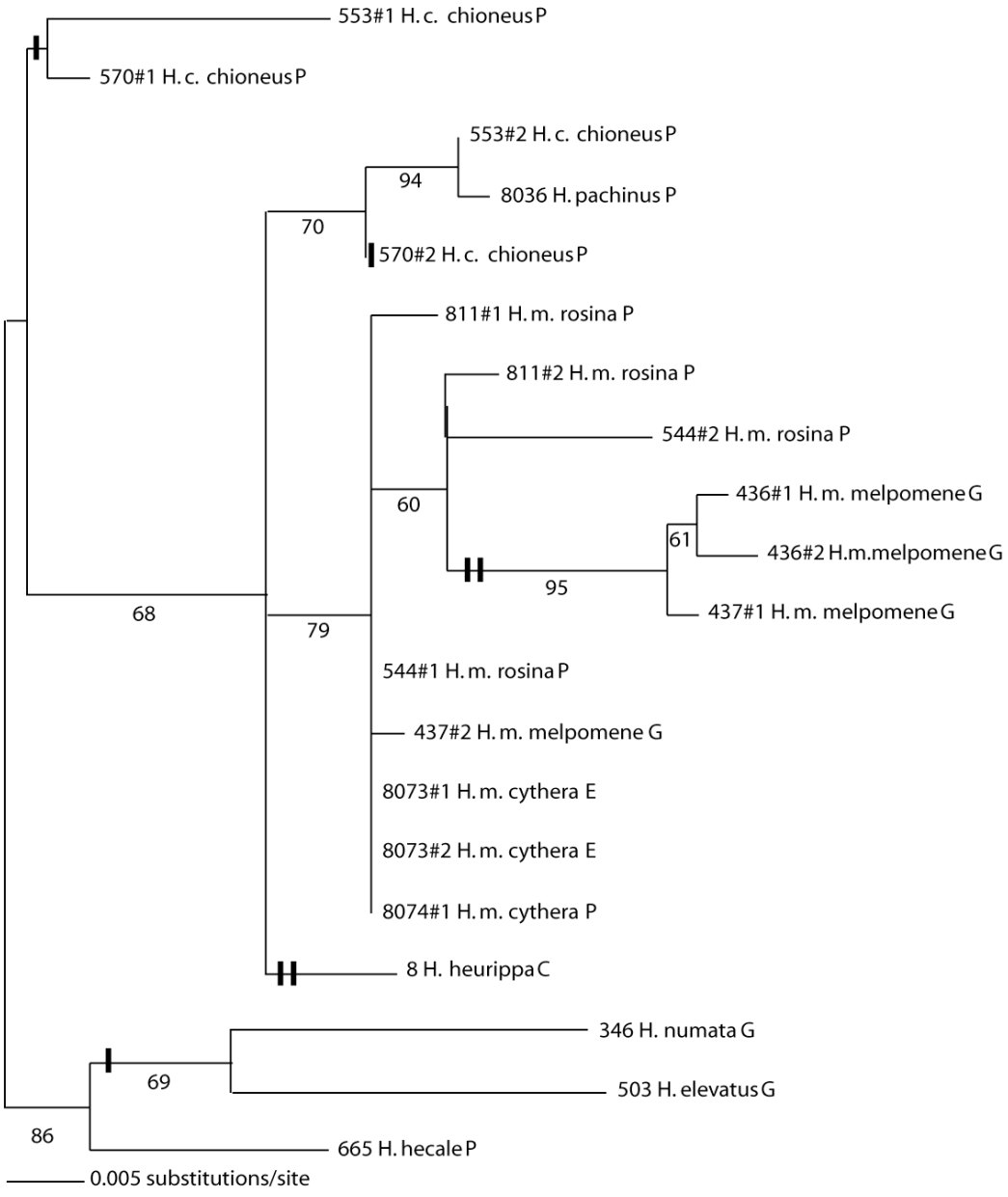


Figure 3. Maximum-likelihood phylogenies for the *Heliconius melpomene/cydno* group (alignment 1 & 3), based on *Tpi* exon+intron (a) and *Mpi* exon+intron (b). Tree descriptions follow the conventions presented in Figure 2. Sample numbers correspond to those given in Table 1 and are followed by an allele number in the case of heterozygotes. Thus, for example 811#1 and 811#2 are two *Tpi* separate alleles cloned from the same individual *H. melpomene rosina*, male ID No. 811. P = Panama, E = Ecuador, G = French Guiana and C= Colombia.

a.



b.

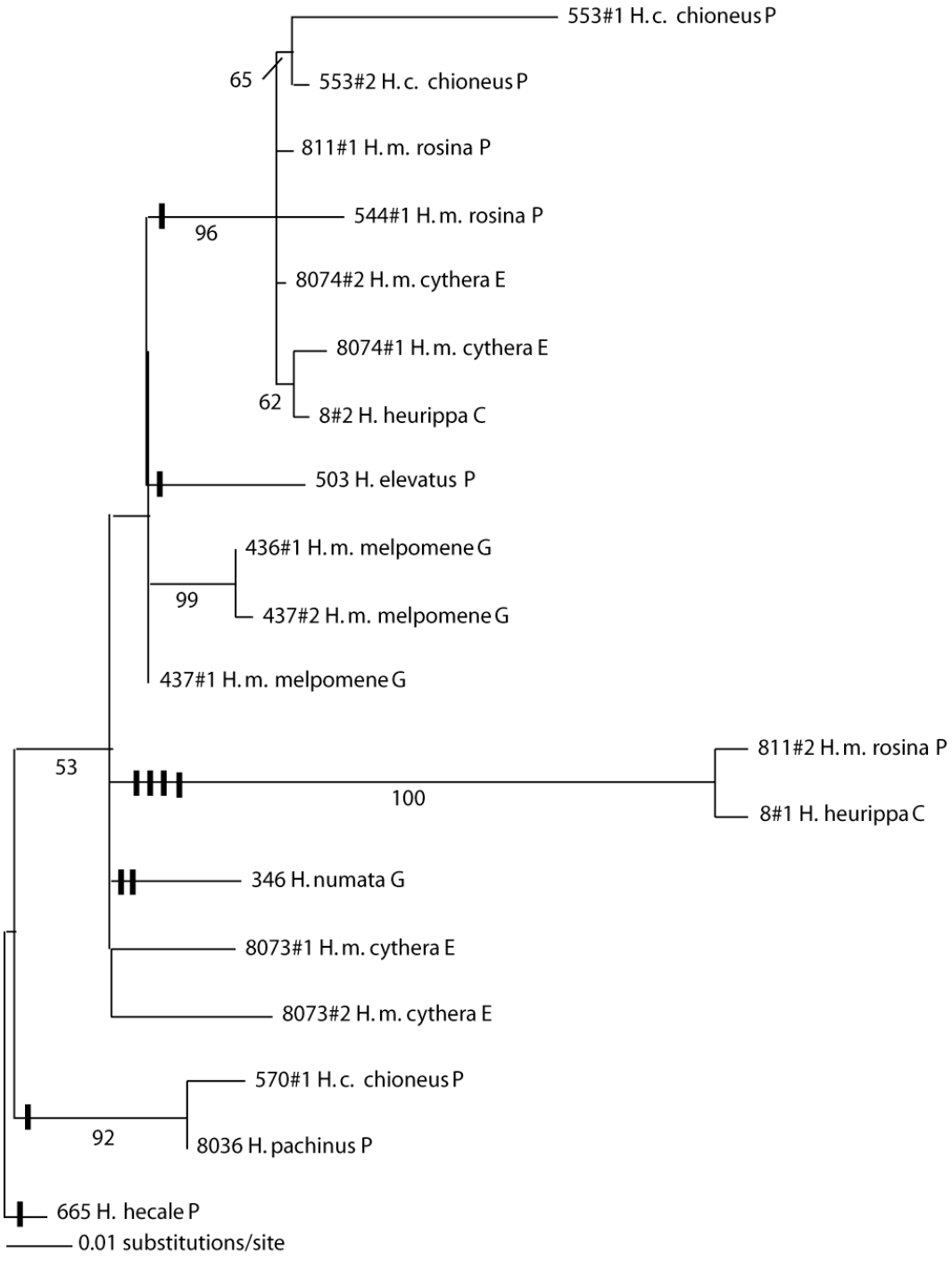
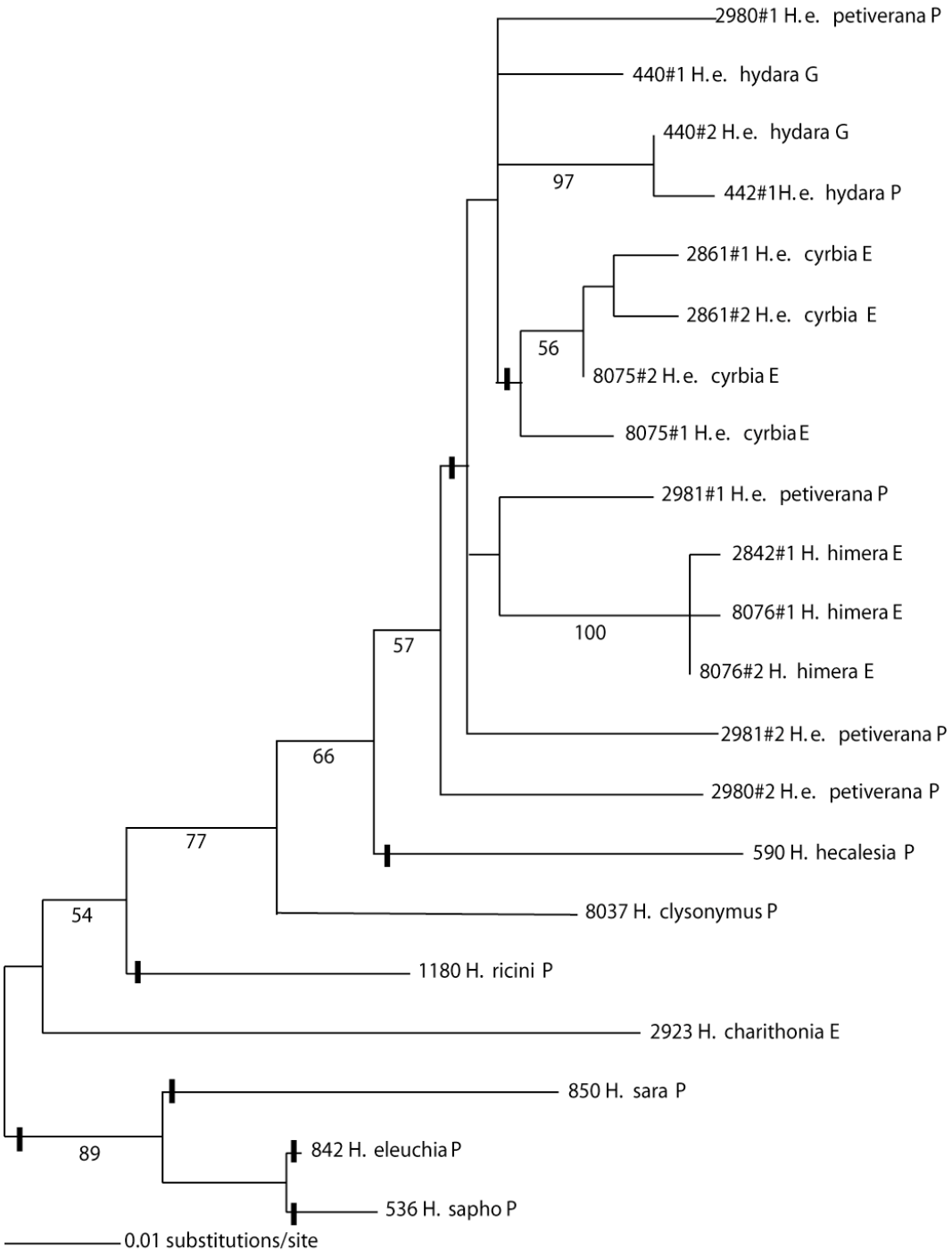
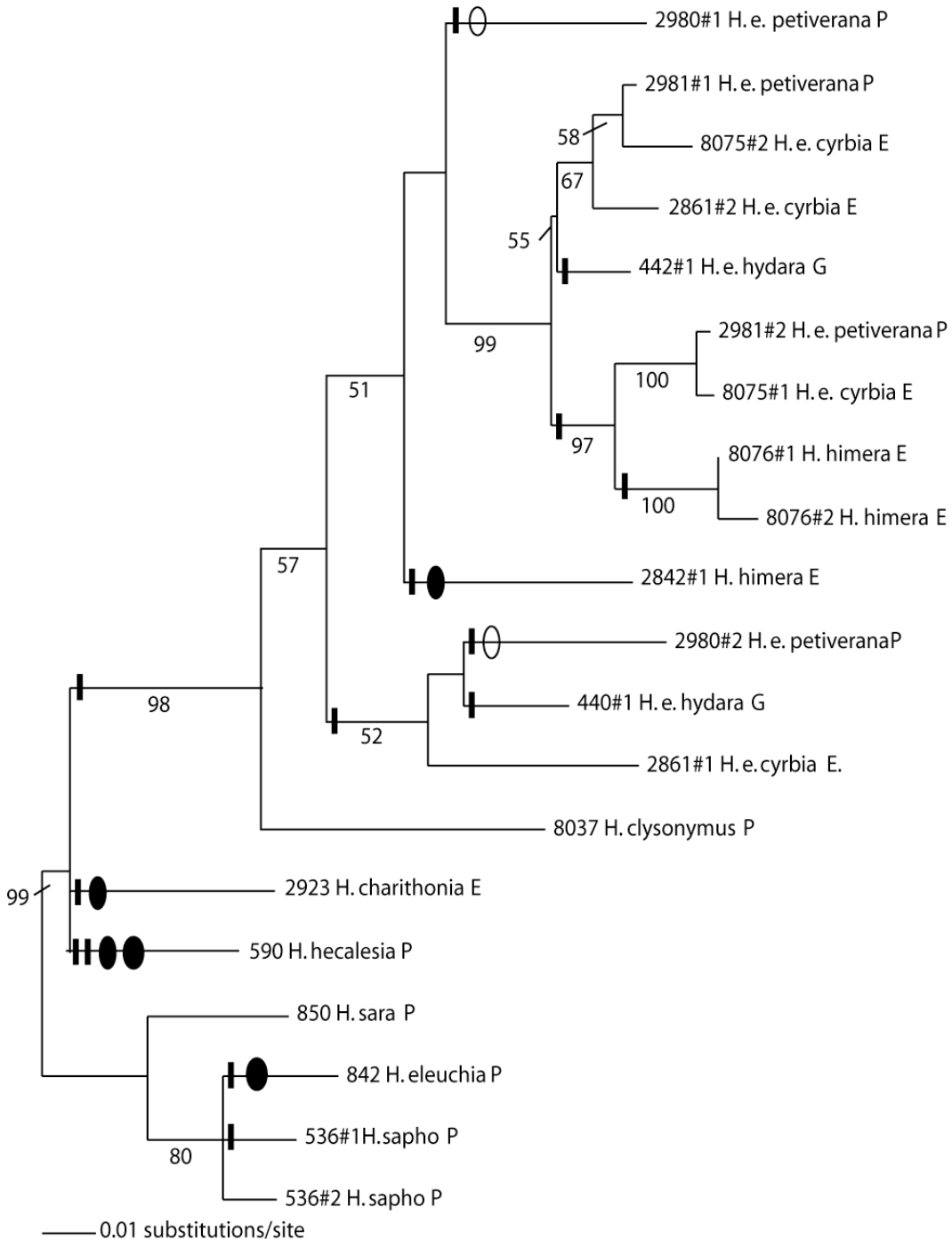

553#1 H.c. chioneus P
65
553#2 H.c. chioneus P
811#1 H.m. rosina P
544#1 H.m. rosina P
96
8074#2 H.m. cythera E
8074#1 H.m. cythera E
62
8#2 H. heurippa C
503 H. elevatus P
436#1 H.m. melpomene G
99
437#2 H.m. melpomene G
437#1 H.m. melpomene G
811#2 H.m. rosina P
53
100
8#1 H. heurippa C
346 H.numata G
8073#1 H.m. cythera E
8073#2 H.m. cythera E
570#1 H.c. chioneus P
92
8036 H. pachinus P
665 H. hecale P
0.01 substitutions/site

Figure 4. Maximum-likelihood phylogenies for *Heliconius erato/himera* group (alignment 2 & 4) based on *Tpi* exon+intron (a) and *Mpi* exon+intron (b). Tree descriptions follow the conventions presented in Figure 2. Bold vertical lines show unique indel changes. Ovals mark the three indels that were not synapomorphic on the sequence-based ML tree. Sample numbers correspond to those given in Table 1 and are followed by an allele number in the case of heterozygotes. Thus, for example 2980#1 and 2980#2 are two *Tpi* separate alleles cloned from the same individual *H. erato petiverana*, male ID No. 2980. P = Panama, E = Ecuador and G = French Guiana.

a.



b.


2980#1 H. e. petiverana P
2981#1 H. e. petiverana P
8075#2 H. e. cyrbia E
2861#2 H. e. cyrbia E
442#1 H. e. hydara G
2981#2 H. e. petiverana P
8075#1 H. e. cyrbia E
8076#1 H. himera E
8076#2 H. himera E
2842#1 H. himera E
2980#2 H. e. petiverana P
440#1 H. e. hydara G
2861#1 H. e. cyrbia E.
8037 H. clysonymus P
2923 H. charithonia E
590 H. hecalesia P
850 H. sara P
842 H. eleuchia P
536#1H. sapho P
536#2 H. sapho P
58
67
55
99
51
100
97
100
57
98
52
99
80
0.01 substitutions/site

Chapter 4

Identification and phylogenetic history of a marker linked to a colour pattern gene in *Heliconius melpomene*

Abstract

Mimicry in *Heliconius* butterflies is an interesting example of adaptive radiation. Although it is known that few genes of major effect control differences in colour pattern within and between races in *Heliconius* the molecular basis of these genes is not known. Here I isolate and sequence an AFLP marker closely linked to the *Yb* locus, that codes for presence or absence of a yellow band in the hindwing in *H. melpomene*. The AFLP band was separated on an acrylamide gel, visualised by silver staining, and isolated. The fragment was amplified, sequenced and locus specific primers designed. Sequences for this region were then obtained from populations representing different genotypes at the *Yb* locus. The resulting cladogram was similar to the cladogram shown by mtDNA, and both cladograms matched biogeographical distributions better than wing pattern similarity. This might imply multiple origins of similar colour patterns in different geographic regions. Isolation of this marker is the first step towards positional cloning of the colour pattern gene. In addition, the locus-specific primers designed here can be used for mapping in further segregating populations of *H. melpomene*, and to test for homology of colour pattern genes between *Heliconius* species.

Introduction

Mimicry and divergence in colour pattern

Mimetic colour pattern races of *Heliconius* butterflies provide an example of adaptive radiation. Bates' classic paper (Bates, 1862), showed a geographical order for the different colour patterns: similar between species within any one area of the Amazon basin, but the mimetic colour patterns themselves changed every 100-200 miles. Beside this geographic divergence, closely related species within an area often belonged to mimicry "rings" (groups of unpalatable species,

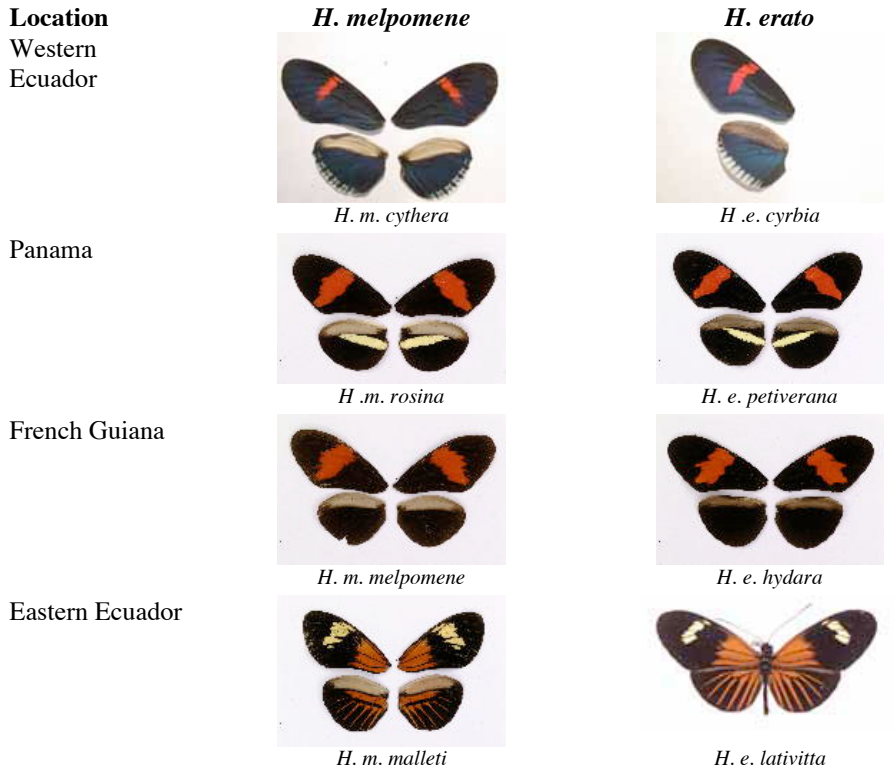
together with some palatable species, that have converged on the same warning colour pattern) (Mallet and Gilbert, 1995).The most impressive example are the co-mimics *H. erato* and *H. melpomene*, which show up to 30 parapatric colour pattern races (Brown, 1979).

Mimicry and warning colour are also good examples of traits under strong selection. Müllerian mimics usually have monomorphic populations separated by narrow hybrid zones, and colour patterns diverge geographically to form a diversity of colour pattern races (e.g. Fig. 1). Evidence collected from hybrid zones showed that racial divergence might be driven and maintained by selection on wing patterns (Mallet, 1993). The shifting balance hypothesis predicts that once a warning colour pattern becomes abundant, the local predator community learns to avoid it. This favours the common pattern and causes frequency-dependent selection against rarer patterns. Rare hybrids and recombinants between divergent colour patterns are not recognised as unpalatable and form an adaptive trough between two adaptive peaks, each peak representing races of the same species (Mallet, 1993).

Colour pattern radiation between and within species is mostly due to the divergence between few genes of major effect (Jiggins and McMillan, 1997; Linares, 1997; Naisbit *et al*., 2003; Sheppard *et al*., 1985). Natural selection can maintain species that are distinct from each other at some genes, in spite of persistent gene flow at other genes (Machado *et al*., 2002). For example, data from mtDNA and allozymes of a hybrid zone between *H. erato* and *H. himera,* showed that ecological selection prevents the spread of neutral alleles between the species (Jiggins *et al*., 1997). In contrast, a study of two sister species pairs *H. erato* and *H. himera*, and *H. melpomene* and *H. cydno*, showed that the genetic variation in the maternally inherited mtDNA and the sex-linked *Tpi* gene clustered together by species. However, the *Mpi* genealogies in both species pairs failed to show structure consistent with species boundaries and suggested recent introgression between species (Appendix 3.1, Beltran *et al*., 2002). The discordant genealogies between genes are consistent with models of adaptive speciation with

ongoing gene flow proposed for *Heliconius*. Thus, genealogies of regions linked to colour pattern are expected to differ from unlinked genes, because natural selection might prevent gene flow at regions of the genome that are tightly linked to genes for species-specific adaptations such as colour pattern genes.

Figure 1. Müllerian mimicry and geographic variation in *Heliconius melpomene* and *Heliconius erato*. *H. melpomene* races are shown in the left and their co-mimics within *H. erato* are shown in the right. Pictures are from individuals used in this thesis except for *H. e. lativitta* (downloaded from Jim Mallet web page). *H. m. mocoa* and *H. amaryllis* have similar pattern to *H. m. rosina*.



Phylogenetic methods to understand pattern radiation in Heliconius

The first attempt to understand the evolution of colour pattern in the mimic species *H. erato* and *H. melpomene* was a parsimony tree constructed from colour pattern alleles. The tree was based in the dominance-sieve theorem, which predicts that newly arisen advantageous mutations have a much greater chance of establishing in a finite population if they are dominant or at least expressed to some degree in heterozygotes (Sheppard *et al.*, 1985; Turner, 1984). Thus, recessive alleles are likely to be primitive to dominant alleles (Turner, 1983;

Turner, 1984). The results suggested that *H. melpomene* and *H. erato* both started as a black butterfly with longitudinal yellow stripes and a solid yellow forewing patch, and have been mutual Müllerian mimics during the whole of their adaptive radiation (Sheppard *et al.*, 1985; Turner, 1984). The ancestral pattern was also similar to the present extra-amazonian races, with a red band on the forewing and a yellow bar on the hindwing, while the Amazonian patterns (e.g. orange rays) are secondarily derived. The new patterns are supposed to have arise in the centre of the species range distribution while the older ones persist in multiple populations at the periphery (Mallet, 1993; Sheppard *et al.*, 1985). This is supported by the fact that the genetic basis of colour patterns in distinct races with red bands and yellow hindwings bars is apparently homologus (Mallet, 1989; Sheppard *et al.*, 1985; Turner, 1984). The shifting balance hypothesis would therefore predict that a phylogeny of genes controlling common elements between these distinct populations would group the populations from the edge of the Amazon.

Turner et al. (1979) used allozymes to investigate variation in different species of *Heliconius* and found that the frequency of some electromorphs shifts between populations while fixed differences were rare. The results showed a lack of diagnostic differences between races despite differentiation in colour pattern (Turner *et al.*, 1979). Later, Brower (1996) used mtDNA genealogies and showed parallel results. Similar patterns found in different clades (e.g. the *H. m. melpomene* pattern shared by three different clades), and similar haplotypes shared by different pattern forms (e.g. clade including *H. e. hydara* and *H. e. lativitta* colour patterns in Fig. 1). He suggested that convergent phenotypes had arisen independently in each lineage, whereas others have been maintained by selection while substantial mtDNA divergence was taking place (Brower, 1996). These results seem to contradict prior theories of the evolution of colour pattern based on analysis of wing- pattern genetics (Sheppard *et al.*, 1985; Turner, 1984).

More recently Flanagan *et al.* (2004) using mtDNA and two nuclear markers (*Tpi* and *Mpi*) provided strong evidence against simultaneous diversification of *H. erato* and *H. melpomene* wing patterns. Instead, the older origin, greater

abundances, and earlier population expansion in *H. erato* suggested that the wing-pattern radiation occurred first in this species and that *H. melpomene* diverged later to mimic the pre-existing patterns in *H. erato* (Flanagan *et al.*, 2004). This result does not support the earlier studies of racial diversification between *H. melpomene* and *H. erato* that suggested that the diversification occurred simultaneously in response to extrinsic forces over the Pleistocene period (Brower, 1996; Sheppard *et al.*, 1985). However, they suggest that additional comparative demographic analysis incorporating variation at the colour-pattern loci themselves will clarify further the evolutionary history of specific pattern elements.

Although markers such as mtDNA, *Tpi*, *Mpi* and allozymes have helped to resolve relationships between species, they are not very conclusive about the relationships of colour pattern races. In order to understand the history of the patterns, the evolutionary history of the genes controlling pattern, or markers tightly linked to them needs to be studied. Linkage mapping can be used to identify markers linked to colour pattern genes.

Searching for regions linked to colour pattern genes

AFLP methods rapidly generate hundreds of highly replicable markers from DNA of any organism (Muller and Wolfenbarger, 1999). These markers have been widely employed to construct linkage maps of a variety of agronomically important organisms, such as barley (Becker *et al.*, 1995), melon (Wang *et al.*, 1997), soybean (Keim *et al.*, 1997), chicken (Herbergs *et al.*, 1999; Knorr *et al.*, 1999), rice (Zhu *et al.*, 1999), oats (Zhu and Kaeppler, 2003) and silkworm (Tan *et al.*, 2001). They have also been successfully employed to map agronomically important genes such as the *cyst* nematode resistance locus *Gpa2* in potato (VanderVoort *et al.*, 1997), the *Mlo* gene in barley (Simons *et al.*, 1997), the *Rx* gene for extreme resistance to potato virus X in tetraploid potato (Bendahmane *et al.*, 1997), a QTL for resistance to *Bacillus thuringiensis* toxins in the diamondback moth (Heckel *et al.*, 1999) and a QTL for sclerotinia resistance in sunflowers (Ronicke *et al.*, 2004). Additionally, when added to existing maps,

AFLP markers have greatly extended the total genomic coverage and decreased the average distance between markers. Therefore, AFLPs are emerging as a marker of choice for genetic mapping for systems with few existing markers and for systems where additional markers are needed to augment existing RFLP and RAPD markers (Mueller and Wolfenbarger, 1999).

The construction of a molecular linkage map represents the first step in the genetic analysis of a target trait of interest. When the genomic regions involved in the expression of the target trait are identified, experiments to trace the alleles of interest at these key regions can be conducted. Once AFLP linkage with the target loci has been identified, bands can be excised from the gels, cloned and sequenced (Cho *et al*., 1996). The sequences obtained allow primers to be designed to amplify the same region from genomic DNA, converting dominant AFLP markers into codominant markers.

Crosses between races of *H. melpomene* and *H. erato* have shown that most of the differences in the colour pattern between races can be explained by changes in only a few genes of large effect (Sheppard *et al*., 1985). Recently, crosses between *H. cydno* and *H. melpomene* have identified genes of major effect that are very likely homologous to those controlling pattern differences between geographic races within species (Naisbit *et al*., 2003). Here I concentrate on a genomic region containing tightly linked genes known as *Yb*, *N*, and *Sb*.
The early study of the genetics of colour pattern in *H. melpomene*, using interracial crosses, showed that the recessive allele *yb* adds a broad yellow band across the hindwing. The yellow band appeared in different forms, one example is a strong yellow bar on both the upper and undersides of the hindwing (e.g. *H. m. rosina* Fig.1) (Mallet, 1986). Tightly linked to *Yb* is *N*, which controls presence ($N^N N^N$) or absence ($N^B N^B$) of an area of white or yellow in the forewing band (Sheppard *et al*., 1985). The *Sb* locus controls the presence (Sb_2) or absence (Sb_1) of the white submarginal band on the hindwing of *cydno* (Linares, 1997) and *H. m. cythera* (Sb^c). Finally, a recent study using hybrids between *H. cydno* and *H. melpomene*, showed that the *Yb* locus was closely linked both to *N* and *Sb* (Naisbit

et al., 2003). These results confirmed the homology of these loci between species with those previously described from crosses within *melpomene* (Sheppard et al. 1985) and within *cydno* (Linares, 1997).

A previous linkage study of *H. melpomene* has identified one AFLP band tightly linked (within 1cM) to the gene that controls the yellow band pattern in the hindwing of *H. melpomene* (*Yb*) (Jiggins *et al*., 2004). The main goals of this chapter are 1) to describe how the AFLP band was isolated sequenced, and converted into a codominant marker and 2) to test the hypothesis that distinct races with shared colour patterns have a common history, using sequences sampled from populations with distinct genotypes at the *Yb*/*N*/*Sb* complex of loci.

Material and methods

Isolation and amplification of AFLP band

Screening of fluorescently labelled AFLP bands is easy and fast, but it is not possible to excise the target band directly from a sequencing gel. However, the same selective reaction can be separated in a denaturing acrylamide gel and stained with silver nitrate in order to detect and cut the specific band (Cho *et al*., 1996). One difficulty is that the fluorescent marker banding patterns might differ from the silver staining gel. Fluorescent markers are labelled on a single primer, while silver staining will show all DNA fragments. For this reason a series of ladders and other markers such as nuclear genes *apterous* (*ap*) and *decapentaplegic* (*dpp*), and microsatellites were run alongside the AFLP PCR reactions to provide size standards. Six PCR products from the selective reaction (*Eco*-CA, *Mse*-CTA) were chosen that contained the band of interest (a41 288bp) and the closest band to it (a40 280bp), AFLP protocols are described in (Jiggins *et al*., 2004). These samples were run in the order shown in Table 1.

Table 1. Experimental design for identification of the AFLP band linked to *Yb*. The notation a40 and a41 are identification codes for particular bands from this AFLP selective reaction.

Lane	Sample	AFLP Bands	Size
1	Ladder Rox-400		200-250-300pb
2	ap-dpp		212-312bp
3	Microsatellites		294-300-330-350bp
4	35 CA-CTA	a40	280bp
5	81 CA-CTA	a40-a41	280-288pb
6	82 CA-CTA	a40-a41	280-288pb
7	83 CA-CTA	a40-a41	280-288pb
8	84 CA-CTA	a40-a41	280-288pb
9	36CA-CTA	a40	280bp
10	Microsatellites		294-300-330-350bp
11	ap-dpp		212-312bp
12	Ladder Rox-400		200-250-300pb

To concentrate the samples (4 to 9 table 1) an aliquot of 120µl of diluted selective PCR reaction was dried and resuspended in 1µl of water. The denaturing acrylamide gel was prepared with 6% acrylamide solution, 7M urea, and 1 TBE buffer (89mM Tris-borate plus 2.5mM EDTA pH 8.3%). The gel was pre-run at 80w until the temperature reached 50ºC. Then, the samples were prepared by mixing 1µl of each sample with 3µl of 3X loading buffer (10mM NaOH, 95% formamide, 0.05% bromophenol blue, and 0.05% xylene cyanol FF), denaturing at 95 ºC for 3 min, followed immediately by chilling on ice. Four microlitres of each sample were loaded into wells. To prevent the gel from cooling, gel-loading time did not exceed 15 min. The gel was run at a constant 70W for 3 hours. Gel were fixed to the small glass plate, and stained with silver nitrate using the Promega Silver sequence™ DNA kit.

The specific bands were excised directly from the dried gel on the glass plate using a razor blade. The bands were diluted in 10µl of water and 2µl of this dilution was used as a template in a total PCR reaction volume of 25µl. The PCR profile and reaction components were the same as those used for the selective amplification (Jiggins *et al*., 2004). The PCR products were electrophoretically separated on 1.5% low melting point agarose with ethidium bromide (1µg/ml). Bands were cut from the gel and dissolved in gelase. The clean template obtained

was sequenced using the un-labelled pre-selective primers (Jiggins *et al.*, 2004). The 10µl cycle sequence reaction mixture contained 1µl BigDye, 0.6µl 5x buffer, 2.5µl primer, and 2µl of template. The cycle profile was 96ºC for 30 sec., then 96ºC for 10 sec., 50ºC for 15 sec, 60ºC for 4 min. for 30 cycles. This product was cleaned by precipitation using 37.5µl of 70% EtOH and 0.5mM $MgCl_2$. The samples were re-suspended in 4µl of a 5:0.12 deionized formamide: crystal violet solution, denatured at 85ºC for two min. and loaded into 5.5% acrylamide gels. Gels were run on a BaseStation (MJ Research) for 3 hours. Sequences were aligned and specific primers were designed using Primer3. The specific primers designed were a41f 5'- CCGTCTACGAAAGATATGCCA-3' and a41r 5'- TTTTGTTCCACTATTACGCAAAAA-3'.

Butterfly material

Sequences for this region were obtained from 18 individuals from 8 races of *H. melpomene*, 3 individuals of the sister species *H. cydno* and one individual of *H. numata* used as an outgroup (Table 2). Butterflies were collected, preserved in liquid nitrogen and stored in the Smithsonian Tropical Research Institute in Panama. Wings of voucher specimens are preserved in envelopes. From each individual, 1/6 of the thorax was ground and the genomic DNA was extracted from individuals in the STRI collection using the DNeasy Kit (Qiagen) following the recommended protocols.

Sequencing methods and phylogenetic analyses

The a41 region was amplified using the specific primers. Double-stranded DNA was synthesised in 10µl reactions containing 2µl of genomic DNA, 1x buffer, 1mM $MgCl_2$, 0.8mM dNTPs, 0.5mM of each primer, and 0.05 u/μl of Qiagen Taq polymerase. DNA was amplified using the following step-cycle profile: 94ºC for 5 min., 94ºC for 30 sec., 55ºC for 30 sec., 72ºC for 1 min. for 34 cycles. PCR products were cleaned using low melting point gels as described before. Cleaned products were sequenced in a 10µl cycle sequence reaction mixture contained 2µl BigDye, 2µl 1x buffer, 0.075µl primer, and 2µl of template. The cycle profile was

96 ºC for 30 sec., then 96ºC for 1 min., 50ºC for 5 sec, 60ºC for 4 min. for 24 cycles. Products were sequenced in a 3100 ABI sequencing machine.

Table 2. Individuals used in the phylogentic analysis. ID numbers are Smithsonian collection numbers and are prefixed by 'STRI-B-'. *N* locus: presence ($N^N N^N$) or absence ($N^B N^B$) of an area of white or yellow in the forewing band. *Yb* locus: presence (*ybyb*) or absence (*YbYb*) of the hindwing yellow bar of melpomene, Sb^c is the presence of a white submarginal band in the hindwing (Linares, 1997; Naisbit et al., 2003; Sheppard et al., 1985). The $yb^c yb^c$ locus is a new notation that represents presence of the hindwing yellow bar on the underside only. Clade numbers were assigned according to assumed colour pattern genotypes for topology for SH test.

Individual	Locality	Genotype	Yellow bar (hindwing)	Clade
1391 *H. m. rosina*	Panama	$N^B N^B$ *ybyb*	Present	1
1137 *H. m. rosina*	Panama	$N^B N^B$ *ybyb*	Present	1
8749 H. m. amaryllis	Peru (East Andes)	$N^B N^B$ *ybyb*	Present	1
8747 *H. m. amaryllis*	Peru (East Andes)	$N^B N^B$ *ybyb*	Present	1
m150 *H. m. mocoa*	Colombia (East Andes)	$N^B N^B$ *ybyb*	Present	1
m75 *H. m. mocoa*	Colombia (East Andes)	$N^B N^B$ *ybyb*	Present	1
59 *H. m. vulcanus*	Colombia (West Andes)	$N^B N^B$ $yb^c yb^c$	Present (underside)	2
60 *H. m. vulcanus*	Colombia (West Andes)	$N^B N^B$ $yb^c yb^c$	Present (underside)	2
8510 *H. m. cythera*	Ecuador (West Andes)	$N^B N^B$ $yb^c yb^c$ Sb^c	Present (underside)	2
8511 *H. m. cythera*	Ecuador (West Andes)	$N^B N^B$ $yb^c yb^c$ Sb^c	Present (underside)	2
8542 *H. m. plesseni*	Ecuador (East Andes)	$N^B N^B$ *YbYb*	Absent	3
8543 *H. m. plesseni*	Ecuador (East Andes)	$N^B N^B$ *YbYb*	Absent	3
249 *H. m. melpomene*	French Guiana	$N^B N^B$ *YbYb*	Absent	3
8159 *H. m. malleti*	Ecuador (East Andes)	$N^N N^N$ *YbYb*	Absent	4
9035 *H. c. weymeri*	Colombia (East Andes)	$N^N N^N$ *ybyb*	Present	5
593 *H. c. chioneus*	Panama	$N^N N^n$ $Yb_c Yb_c$	Absent	5
594 *H. c. chioneus*	Panama	$N^N N^N Yb_c Yb_c$	Absent	5
N235 *H. numata*	Peru		Absent	Root

Chromatograms were edited and base calls checked using SEQUENCHER 4.1 (Gene Codes Corporation, Inc). Maximum parsimony (MP) trees were obtained using PAUP* version 4.0b8 (Swofford, 2000) in a heuristic search with TBR branch swapping. The consensus tree was calculated using majority rule. Confidence in each node was assessed by bootstrapping (1000 replicates, heuristic search with TBR branch swapping). In order to test specific hypotheses, first a maximum likelihood (ML) tree was calculated using PAUP* version 4.0b8, based on the model that best fit the data ModelTest 3.04; (Posada and Crandall, 1998). The parsimony cladogram (Fig. 3A) and ML tree (tree not shown) had very similar topology. Thus, alternative *a priori* scenarios were compared to the ML

tree using the method of Shimodaira and Hasegawa (Goldman *et al.*, 2000; Shimodaira and Hasegawa, 1999) and implemented using PAUP* version 4.0b8. Four tree topologies were compared, ML tree of a41, a hypothesis where the races were grouped by mtDNA results, races grouped according to Turner (1984), and races assumed according colour pattern genotypes (see Table 1 for clades-Fig. 5). In each case the shortest tree for each scenario was sought using MacClade, starting with the ML tree.

For comparison mtDNA sequences of two individuals of *H. melpomene amaryllis*, two of *H. melpomene plesseni*, two of *H. melpomene vulcanus* and one of *H. cydno weymeri* sequences were downloaded from GenBank (Brower, 1996) (Accession number UO8497, UO8496, UO8493, UO8485, UO8479, UO8487, and AY54116 respectively), and from the first chapter of this thesis.

Results

Isolation and amplification of AFLP band linked to Yb locus

The AFLP band of interest was cut from a thin acrylamide gel with a razor blade and amplified directly using two rounds of PCR. Lanes 4 to 9 in Fig. 2 show the band pattern of AFLP products visualized by silver staining (a41 band at 288pb). Lanes 1 to 3 in Fig. 3A and B show amplified fragments of equal molecular weight (288pb), showing that identical DNA fragments previously visualized by fluorescence can be isolated and directly amplified from silver stained acrylamide gels. Several methods of Blast search implemented by GenBank were used to find homologies with sequences already published (e.g. blast for short sequences and tBlastX that translates in all possible directions and reading frames). I did not find any significant hits, suggesting that this region has not been reported before. Additionally, the identity of the region was confirmed by mapping the locus using the specific primers described here (Jiggins *et al.*, 2004).

Figure 2. Detection of a41 on 6% denaturing polyacrylamide gel using silver staining. Samples correspond to the experimental design showed in table 2 (sample number 1 is indicated with *). Arrow shows position of a41 (288bp).

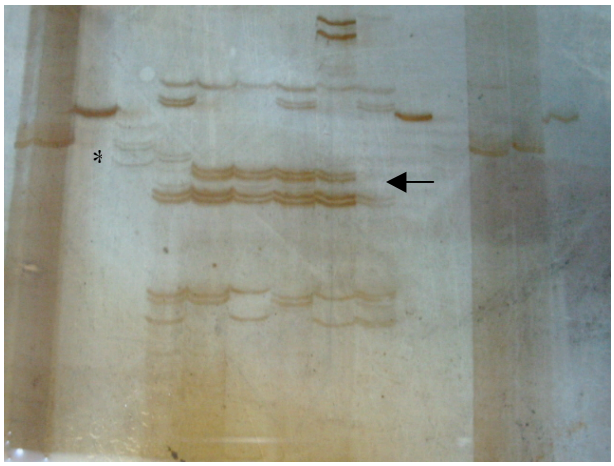
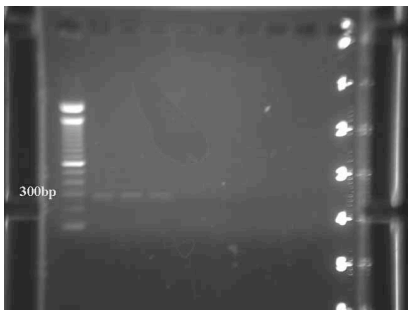
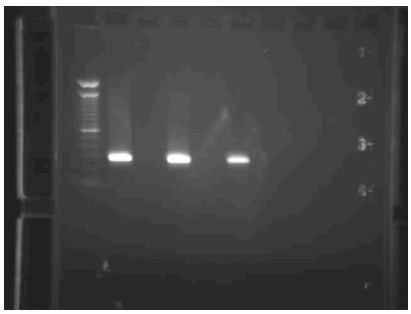


Figure 3. A. First amplification of a41 (288bp) using AFLP selective primers (*Eco*-CA, *Mse*-CTA) and conditions. B. Re-amplification.

A.



B.



Eighteen individuals representing eight races of *H. melpomene*, two races of *H. cydno* and *H. numata*, were sequenced for a41. Amplification was successful and in some cases individuals showed two alleles (e.g. *H. cydno* one allele at 325bp and the other at 194bp). Alleles were isolated from agarose gels, cleaned and sequenced (following protocols describe above). However, sequences from the shortest allele could not be aligned with the other individuals, suggesting that it is not an allelic form of this locus. In general a41 fragments were variable in size between 380bp and 306pb.

The final alignment for this region contained 436 positions including gaps (Appendix 4.1). Sixty-five (15%) of the characters were variable and 35 (8%) were parsimony-informative. This region exhibited considerable length variation and the indel variation provided phylogenetic information that was easily mapped onto the tree (Fig. 4A). The mtDNA final alignment contained 1595 positions including gaps. Eighty-one (5%) of the characters were variable and 67 (4.2%) were parsimony-informative (Fig. 4B). The GC content was similar between mtDNA (24%) and a41 (28%). The model of evolution that best fitted the data from the a41 marker was HKY+G (Hasegawa *et al*., 1985), with two substitution types and gamma-shape parameters (α=0.5) and for mtDNA the best fit model was GTR (Yang, 1994). The a41 region had lower estimated transition: transversion rate ratios as compared to mitochondrial sequences.

The mtDNA cladogram shows a clear biogeographic pattern consisting of *H. melpomene* races from the west of the Andes with *H. m. melpomene* from French Guiana basal to this clade and a second clade formed by *H. melpomene* races from east Andes (Fig. 4B). The a41 cladogram shows a similar biogeographical pattern, the western clade remains the same as the mtDNA clade. The eastern clade is divided into two. One clade contains *H. m. mocoa* from Colombia and *H. m. malleti* from Ecuador. The second clade included *H. m. plesseni*, *H. m. amaryllis* and *H. cydno* (Fig. 4A). When the divergence within geographical clades was compared a41 showed less than 1% among the western races and less than 7% pairwise divergence among the eastern clade including *H. cydno*. In general divergence in mtDNA was less than 3% among all *H. melpomene* races. An SH test based on the a41 ML tree (Fig. 5A) showed that the topologies constrained according to mtDNA (Fig. 5B) or by colour pattern (Fig. 5 C, D) were significantly worse fit to the data.

Discussion

The a41 band is the first isolated and sequenced marker linked to a colour pattern region in *Heliconius* and the conversion of this AFLP band into a codominant marker is an important step towards molecular characterization of the colour

pattern gene. Primers designed using this and similar loci can be used to identify Bac clones containing the colour pattern locus itself. Additionally, although there are great differences in colour pattern between *H. cydno*, *H. melpomene* and *H. numata* (Fig. 4), this region was successfully amplified in all three species. This region has already been mapped in *H. numata*, where it is tightly linked to the supergene *P*, which controls all the major colour pattern polymorphisms in this very different-looking, tiger-patterned species. This result suggests homology between the *Yb* or *N* locus in *H. melpomene* and the supergene *P* of *H. numata* (Joron *et al.*, 2004).

The genealogies of regions linked to colour pattern were expected to differ from neutral markers, such as mtDNA. Thus, the phylogenetic pattern that might have been expected was a tree following colour pattern similarity. For example, the races *H. m. rosina*, *mocoa* and *amaryllis* share a common phenotype (Fig. 4). If the phenotypes shared a common genetic origin, they might be predicted to form a monophyletic clade in the phylogeny. However, the phylogenetic relationships inferred from the a41 sequences showed conflicts with the expected pattern and with the hypothesis of a single origin of hindwing yellow bar as previously proposed (Sheppard *et al.*, 1985; Turner *et al.*, 1979; Turner, 1984). One explanation is that yellow hindwing bar pattern, has evolved numerous times. The data therefore support Browers (1996) argument that the phylogenetic distribution of wing patterns requires multiple reappearance of some recessive alleles such as the hindwing yellow bar allele. This suggests that the dominance-sieve theorem (Turner, 1984), might be an over-simplification of the process of phenotypic diversification of colour pattern (Brower, 1996).

An alternative is that the marker sequenced here is distant enough from the colour pattern gene so that the two do not share common history. In particular, neutral gene flow is possible across hybrid zones (Mallet, 1993). Thus, parts of the genome can introgress easily between races of *H. melpomene*. The a41 band appeared tightly linked to *Yb* gene (1 cM), however 1% recombination could

result in a high fraction of recombinants between a4a and the colour pattern loci over many generations.

In conclusion the a41 cladogram shows some geographical structure within races of *H. melpomene*, especially for races from the western Andes. The cladogram fails to corroborate the wing-pattern parsimony hypotheses of Sheppard *et al.* (1985) and Turner (1984), and perhaps suggests a multiple origin for the yellow bar colour pattern instead of just a single origin. This approach is not really conclusive and to improve this it may be necessary to identify regions closer to the locus controlling the colour pattern. However, the conversion of a41 into a codominant marker is an important step towards positional cloning and molecular characterization of the regulatory mechanisms that produce colour pattern switch.

Figure 4. Phylogeny for races of *H. melpomene* using A. a41 region B. mtDNA. Branch lengths were estimated using parsimony. Values above branches show parsimony bootstrap support. P= Panama, E= Ecuador, FG= French Guiana, C= Colombia, and Pe= Peru. Black bars represent indels.

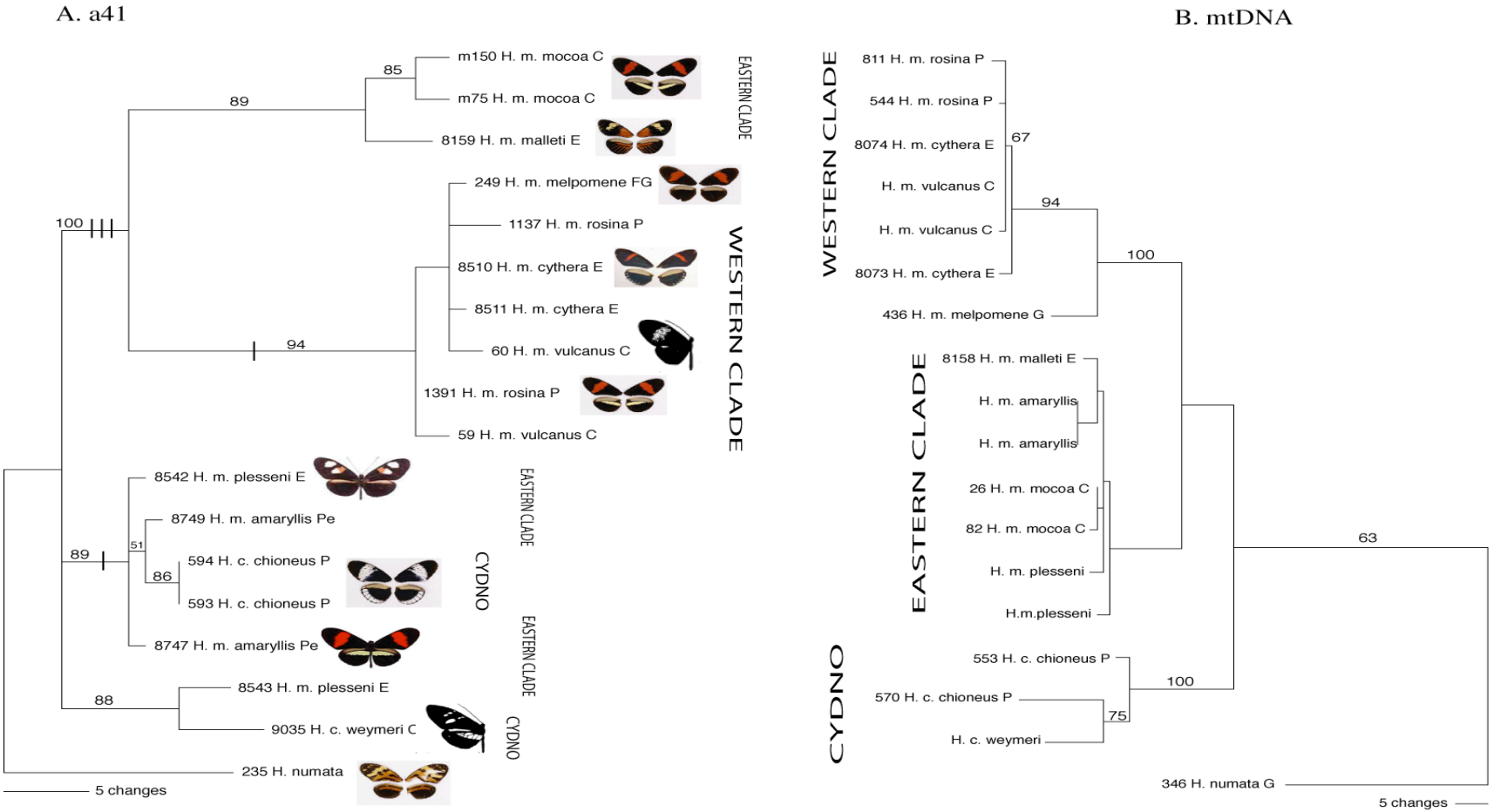
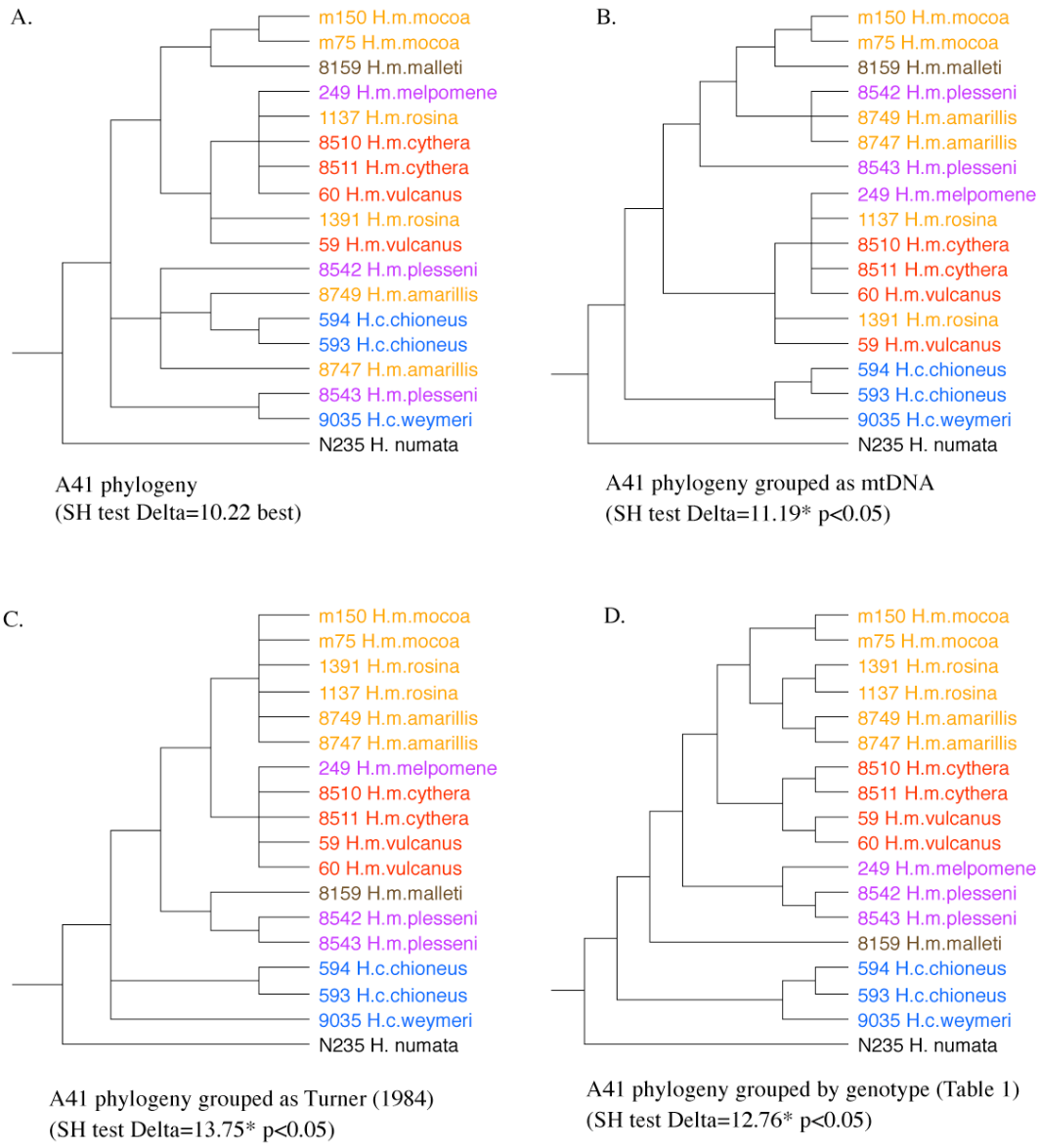


Figure 5. Tree topologies tested in SH test. In D. topology was grouped according to assumed colour pattern genotypes for topology showed in table 1. Topologies with * were significantly worse fit to the data.



References

Bates HW (1862) Contributions to an insect fauna of the Amazon valley. Lepidoptera: Heliconidae. *Transactions of the Linnean Society of London* **23**, 495-566.

Becker J, Vos P, Kuiper F, Salamani F, Heun M (1995) Combined mapping of AFLP and RFLP markers in barley. *Mol. Gen. Genet.* **249**, 65-73.

Beltran M, Jiggins CD, Bull V, *et al.* (2002) Phylogenetic discordance at the species boundary: Comparative gene genealogies among rapidly radiating Heliconius butterflies. *Molecular Biology and Evolution* **19**, 2176-2190.

Bendahmane A, Kanyuka K, Baulcombe DC (1997) High-resolution genetical and physical mapping of the Rx gene for extreme resistance to potato virus X in tetraploid potato. *Theoretical and Applied Genetics* **95**, 153-162.

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.

Brown KS (1979) *Ecologia Geográfica e Evolução nas Florestas Neotropicais. Vol 1.*, Universidade Estadual de Campinas.

Cho YG, Blair MW, Panaud O, McCouch SR (1996) Cloning and mapping of variety-specific rice genomic DNA sequences: amplified fragment length polymorphisms (AFLP) from silver-stained polyacrylamide gels. *Genome* **39**, 373-378.

Flanagan N, Tobler A, Davison A, *et al.* (2004) Historical demography of Mullerian mimicry in the neotropical *Heliconius* butterflies. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9704-9709.

Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* **49**, 652-670.

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **21**, 160-174.

Heckel DG, Gahan LJ, Liu YB, Tabashnik BE (1999) Genetic mapping of resistance to *Bacillus thuringiensis* toxins in diamondback moth using biphasic linkage analysis. *Proc. Natl. Acad. Sci. U. S. A.* **96**.

Herbergs J, Siwek M, Crooijmans R, Van der Poel JJ, Groenen MAM (1999) Multicolour fluorescent detection and mapping of AFLP markers in chicken (*Gallus domesticus*). *Animal Genetics* **30**, 274-285.

Jiggins CD, McMillan WO (1997) The genetic basis of an adaptive radiation: warning colour in two *Heliconius* species. *Proceedings of the Royal Society of London Series B-Biological Sciences* **264**, 1167-1175.

Jiggins CD, McMillan WO, King P, Mallet J (1997) The maintenance of species differences across a Heliconius hybrid zone. *Heredity* **79**, 495-505.

Jiggins CD, Mavarez J, Beltran M, McMillan W, Bermingham E (2004) A genetic linkage map of the mimetic butterfly, *Heliconius melpomene*. *In prep.*

Joron M, Beltran M, Mavarez J, *et al*. (2004) Evidence for a conserved colour pattern locus in *Heliconius*. *In prep*.

Keim P, Schupp JM, Travis SE, *et al*. (1997) A high-density soybean genetic map based on AFLP markers. *Crop Science* **37**, 537-543.

Knorr C, Cheng HH, Dodgson JB (1999) Application of AFLP markers to genome mapping in poultry. *Animal Genetics* **30**, 28-35.

Linares M (1997) The ghost of mimicry past: Laboratory reconstitution of an extinct butterfly 'race'. *Heredity* **78**, 628-635.

Machado CA, Kliman RM, Markert JA, Hey J (2002) Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution* **19**, 472-488.

Mallet J (1986) Hybrid zones of *Heliconius* butterflies in Panama and the stability and movement of warning color clines. *Heredity* **56**, 191-202.

Mallet J (1989) The genetics of warning colour in Peruvian hybrid zones of *Heliconius erato* and *H. melpomene*. *Proceedings of the Royal Society of London, B* **236**, 163-185.

Mallet J (1993) Speciation, raciation, and color pattern evolution in *Heliconius* butterflies: evidence from hybrid zones. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 226-260. Oxford University Press, New York.

Mallet J, Gilbert LE (1995) Why are there so many mimicry rings? Correlations between habitat, behaviour and mimicry in *Heliconius* butterflies. *Biological Journal of the Linnean Society* **55**, 159-180.

Muller U, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution* **14**, 389-394.

Naisbit RE, Jiggins CD, Mallet J (2003) Mimicry: developmental genes that contribute to speciation. *Evolution & Development* **5**, 269-280.

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.

Ronicke S, Hahn V, Horn R, *et al*. (2004) Interspecific hybrids of sunflower as a source of Sclerotinia resistance. *Plant Breeding* **123**, 152-157.

Sheppard PM, Turner JRG, Brown KS, Benson WW, Singer MC (1985) Genetics and the evolution of mullerian mimicry in *Heliconius* butterflies. *Philosophical Transactions of the Royal Society of London (B)* **308**, 433-613.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**, 1114-1116.

Simons G, vanderLee T, Diergaarde P, *et al*. (1997) AFLP-based fine mapping of the Mlo gene to a 30-kb DNA segment of the Barley genome. *Genomics* **44**, 61-70.

Swofford DL (2000) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Tan Y-D, Wan C, Zhu Y, *et al*. (2001) An amplified fragment length polymorphism map of the silkworm. *Genetics* **157**, 1277-1284.

Turner JRG, Johnson MS, Eanes WF (1979) Contrasted modes of evolution in the same genome: allozymes and adaptive change in *Heliconius*. *Proceedings of the National Academy of Sciences, USA* **76**, 1924-1928.

Turner JRG (1983) Mimetic butterflies and punctuated equilibria: some old light on a new paradigm. *Biological Journal of the Linnean Society* **20**, 277-300.

Turner JRG (1984) Mimicry: the palatability spectrum and its consequences. In: *The Biology of butterflies (Symposia of the Royal Entomological Society of London, No. 11)* (eds. Vane-Wright RI, Ackery PR), pp. 141-161. Academic Press, London.

VanderVoort JR, Wolters P, Folkertsma R, *et al*. (1997) Mapping of the cyst nematode resistance locus Gpa2 in potato using a strategy based on comigrating AFLP markers. *Theoretical and Applied Genetics* **95**, 874-880.

Wang GH, Thomas CE, Dean RA (1997) A genetic map of melon (Cucumis melo L.) based on amplified fragment length polymorphism (AFLP) markers. *Theoretical and Applied Genetics* **95**, 791-798.

Yang Z (1994) Estimating the pattern of nucleotide substitution. *Journal of Molecular Evolution* **39**, 105-111.

Zhu JH, Stephenson DA, Laurie DA, Li W, Tang D (1999) Towards rice genome scanning by map-base AFLP fingerprinting. *Mol. Gen. Genet.* **261**, 184-195.

Zhu S, Kaeppler HF (2003) A genetic linkage map for hexaploid, cultivated oat (*Avena sativa* L.) based on an intraspecific cross 'Ogle/MAM17-5'. *Theoretical and Applied Genetics* **107**, 26-35.

Appendix 4.1.

Nexus file containing the alignment for a41.

```
#NEXUS a41

BEGIN DATA;
       DIMENSIONS  NTAX=18 NCHAR=436;
       FORMAT DATATYPE=DNA  MISSING=? GAP=-  INTERLEAVE ;
MATRIX

[                              10        20        30        40        50        60]
[                              .         .         .         .         .         .]

m150_H.m.mocoa        ?????????????ACTATTACGCAAAAACTAGTGAACAGATTTACCTATATGAAATTTCA   [60]
m75_H.m.mocoa         ????????????TCCACTATTACGCAAAAACTAGTGAACAGATTTACCTATATGAAATTTCA   [60]
249_H.m.melpone       ?????????????????????????AAAACTAGTGAACAGATTTACTTAGATGAAATTTCA   [60]
1391_H.m.rosina       ?????????????????????????AAAACTAGTGAACAGATTTACTTAGATGAAATTTCA   [60]
1137_H.m.rosina       ??????????????????????CAAAAACTAGTGAACAGATTTACTTAGATGAAATTTCA   [60]
8510_H.m.cythera      ???????????????????TACGCAAAAACTAGTGAACAGATTTACTTAGATGAAATTTCA   [60]
8511_H.m.cythera      ??????????????????????CAAAAACTAGTGAACAGATTTACTTAGATGAAATTTCA   [60]
59_H.m.vulcanus       ?????????????????????AAAAACTAGTGAACAGATTTACTTAGATGAAATTTCA   [60]
60_H.m.vulcanus       ??????????????????????CAAAAACTAGTGAACAGATTTACTTAGATGAAATTTCA   [60]
8159_H.m.malleti      ??????????CCACTATTACGCAAAAACTAGTGAA--GATTTACCTATATGAAATTTCA   [58]
8542_H.m.plesini      ???TTTTTGTTCCACTATTACGCAAAAACTAGTGAACAGATTTAC---------------   [45]
8543_H.m.plesini      ??????????????????????CAAAAACTAGTGAACAGATTTA----GATGAAATTTCA   [56]
8749_H.m.amarillis    ???????????????????CGCAAAAACTAGTGAACAGATTTAC---------------   [45]
8747_H.m.amarillis    ??????????????????????CAAAAACTAGTGAACAGATTTAC---------------   [45]
594_H.c.chioneus      ??????????CCACTATTACGCAAAAACTAGTGAACAGATTTAC---------------   [45]
593_H.c.chioneus      ??????????CCACTATTACGCAAAAACTAGTGAACAGATTTAC---------------   [45]
9035_H.c.weymeri      ????TTTTGTTCCACTATTACGCAAAAACTAGTGAATAGATTTAC---------------   [45]
N235_H._numata        CAGTTTTTGTTCCACTATTACGCAAAAACTAGTAAACAGATTTA----GATGAAATTTCA   [56]

[                              70        80        90        100       110       120]
[                              .         .         .         .         .         .]

m150_H.m.mocoa        CAGCTTA--------TATATCAGAATAACACATATACGACAACAAGTTTAAGCAGGTAAA   [112]
m75_H.m.mocoa         CAGCTTA--------TATATCAGAATAACACATATACGACAACAAGTTTAAGCAGGTAAA   [112]
249_H.m.melpone       CAGCTTATAAG----TATATCAGAATAACACATATTCTACAACAAGTTTAAGCAAGTAAA   [116]
1391_H.m.rosina       CAGCTTATAAG----TATATCAGAATAACACATATTCTACAACAAGTTTAAGCAAGTAAA   [116]
1137_H.m.rosina       CAGCTTATAAG----TATATCAGAATAACACATATTCTACAACAAGTTTAAGCAAGTAAA   [116]
8510_H.m.cythera      CAGCTTATAAG----TATATCAGAATAACACATATTCTACAACAAGTTTAAGCAAGTAAA   [116]
8511_H.m.cythera      CAGCTTATAAG----TATATCAGAATAACACATATTCTACAACAAGTTTAAGCAAGTAAA   [116]
59_H.m.vulcanus       CAGCTTATAAG----TATATCAGAATAACACATATTCTACAACAAGTTTAAGCAAGTAAA   [116]
60_H.m.vulcanus       CAGCTTATAAG----TATATCAGAATAACACATATTCTACAACAAGTTTAAGCAAGTAAA   [116]
8159_H.m.malleti      CACCTTATA--------TATCAGAATAACACATATACGACAACAAGTTTAAGCAGGTAAA   [110]
8542_H.m.plesini      ------------------------------------------------------------   [45]
8543_H.m.plesini      CAGCTTATAAGTAAGTATATCAGAATAACACATATACTACAACAAGTTTAAGCAGGTAAA   [116]
8749_H.m.amarillis    ------------------------------------------------------------   [45]
8747_H.m.amarillis    ------------------------------------------------------------   [45]
594_H.c.chioneus      ------------------------------------------------------------   [45]
```

```
593_H.c.chioneus      ------------------------------------------------------------   [45]
9035_H.c.weymeri      ------------------------------------------------------------   [45]
N235_H._numata        CAGTTTAT----AAGTATATCAGAATAACACATATACTACAACAAGTTTAAGCAGGTAAA   [112]

[                             130       140       150       160       170       180]
[                             .         .         .         .         .         .]

m150_H.m.mocoa        TCCGCATTTCGGTCGCTAGTTCTAAATATGATACAATTTTATTTTCTTAATTTTC-----   [167]
m75_H.m.mocoa         TCCGCATTTCGGTCGCTAGTTCTAAATATGATACAATTTTATTTTCTTAATTTTC-----   [167]
249_H.m.melpone       TCCACATTTCGGTCGCTAGTTCTAAATATGATACAATTTTATTTTCTTAATTTTC-----   [171]
1391_H.m.rosina       TCCACATTTCGGTCGCTAGTTCTAAATATGATACAATTTTATTTTCTTAATTTTC-----   [171]
1137_H.m.rosina       TCCACATTTCGGTCGCTAGTTCTAAATATGATACAATTTTATTTTCTTAATTTTC-----   [171]
8510_H.m.cythera      TCCACATTTCGGTCGCTAGTTCTAAATATGATACAATTTTATTTTCTTAATTTTC-----   [171]
8511_H.m.cythera      TCCACATTTCGGTCGCTAGTTCTAAATATGATACAATTTTATTTTCTTAATTTTC-----   [171]
59_H.m.vulcanus       TCCACATTTCGGTCGCTAGTTCTAAATATGATACAATTTTATTTTCTTAATTTTC-----   [171]
60_H.m.vulcanus       TCCACATTTCGGTCGCTAGTTCTAAATATGATACAATTTTATTTCCTTAATTTTC-----   [171]
8159_H.m.malleti      TCCGCATTTCGGTCGCTAGTTCTAAATATGATACAATTTTATTTTCTTAATTTTC-----   [165]
8542_H.m.plesini      ------------------CTTCTAAATATGATACAATTTTATATTCTTAATACTCCATCA   [87]
8543_H.m.plesini      TCCGCATTTCGGTCGCTAG-------TATGATACAATTTCATATTCTTAATACTCCATCA   [169]
8749_H.m.amarillis    ------------------CTTCTAAATATGATACAATTTTATATTCTTAATACTCCATCA   [87]
8747_H.m.amarillis    ------------------CTTCTAAATATGATACAATTTTATATTCTTAATACTCCATCA   [87]
594_H.c.chioneus      ------------------CTTCTAAATATGATACAATTTTATATTCTTAATACTCCATCA   [87]
593_H.c.chioneus      ------------------CTTCTAAATATGATACAATTTTATATTCTTAATACTCCATCA   [87]
9035_H.c.weymeri      ------------------CTTCTAAATATGATACAATTTTATATTCTTAATACTCCATCA   [87]
N235_H._numata        KCAGCATTTCGGTTGCTAGTKCTAAATA------------CTCTACTTAATACTCCRTCA   [160]

[                             190       200       210       220       230       240]
[                             .         .         .         .         .         .]

m150_H.m.mocoa        ---------------ATAATTTAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [212]
m75_H.m.mocoa         ---------------ATAATTTAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [212]
249_H.m.melpone       ---------------ATAATTTAAGCTTTTTTATTATAGCTGTTCSTGTATACATTTTTA   [216]
1391_H.m.rosina       ---------------ATAATTTAAGCTYTTTTATTATAGCTGTTCGTGTATACATTTTTA   [216]
1137_H.m.rosina       ---------------ATAATTTAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [216]
8510_H.m.cythera      ---------------ATAATTTAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [216]
8511_H.m.cythera      ---------------ATAATTTAAGCTCTTTTATTATAGCTGTTCGTGTATACATTTTTA   [216]
59_H.m.vulcanus       ---------------ATAATTTAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [216]
60_H.m.vulcanus       ---------------ATAATTTAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [216]
8159_H.m.malleti      ---------------ATAATTTAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [210]
8542_H.m.plesini      TCTTTATGTTTTCCCATAGTTAAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [147]
8543_H.m.plesini      TCTTTATGTTTTCCCATTGTTTAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [229]
8749_H.m.amarillis    TCTTTATGTTTTCCCATAGTTTAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [147]
8747_H.m.amarillis    TCTTTATGTTTTCCCATAGTTTAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [147]
594_H.c.chioneus      TCTTTATGTTTTCCCATAGTTTAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [147]
593_H.c.chioneus      TCTTTATGTTTTCCCATAGTTTAAGCTTTTTTATTATAGCTGTTCGTGTATACATTTTTA   [147]
9035_H.c.weymeri      TCTTTATGTTTTCCCATTGTTTAAGCTTYTTTATTATAGCTGTTCGTATATACATTTTTA   [147]
N235_H._numata        TCTTTATGTTTTTCCATAGTTTAAGCCTCTTTATTACAGCTGTTCGTGTATACATTTATA   [220]

[                             250       260       270       280       290       300]
[                             .         .         .         .         .         .]
```

```
m150_H.m.mocoa        AAGCTTACATGATTATGATACTT-------------GTTCTTGCCTTGTTTATTAAAATT   [259]
m75_H.m.mocoa         AAGCTTACATGATTATGATACTT-------------GTTCTTGCCTTGTTTATTAAAATT   [259]
249_H.m.melpone       AAGCTTACATGATTATGATACTT-------------GTTCTTGCCTTGTCTATTAAAATT   [263]
1391_H.m.rosina       AAGCTTACATGATTATGATACTT-------------GTTCTTGCCTTGTTTATTAAAATT   [263]
1137_H.m.rosina       AACCTTATATGATTATGATACTT-------------GTTCTTGCCTTGTTTATTAAAATT   [263]
8510_H.m.cythera      AAGCTTACATGATTATGATACTT-------------GTTCTTGCCTTGTTTATTAAAATT   [263]
8511_H.m.cythera      AAGCTTACATGATTATGATACTT-------------GTTCTTGCCTTGTTTATTAAAATT   [263]
59_H.m.vulcanus       AAGCTTACATGATTATGATACTT-------------GTTCTTGCCTTGTCTATTAAAATT   [263]
60_H.m.vulcanus       AAGCTTACATGATTATGATACTT-------------GTTCTTGCCTTGTTTATTAAAATT   [263]
8159_H.m.malleti      AAGCTTACATGATTATGATACTT-------------GTTCTTGCCTTGTTTATTAAAATT   [257]
8542_H.m.plesini      AAGCTTACATGATTTATAGTTTCGGTATCAAATAAAGTTCTTGCCTTGTTTATTAAAATT   [207]
8543_H.m.plesini      AAGCTTACATGATTTATAGTTTCAGTATCTAATAAAGTTCTTGCGTTGTTTATTAAAATT   [289]
8749_H.m.amarillis    AAGCTTACATGATTTATAGTTTCGGTATCAAATAAAGTTCTTGCCTTGTTTATTAAAATT   [207]
8747_H.m.amarillis    AAGCTTACATGATTTATAGTTTCGGTATCAAATAAAGTTCTTGCCTTGTTTATTAAAATT   [207]
594_H.c.chioneus      AAGCTTACATGATTTATAGTTTCGGTATCAAATAAAGTTCTTGCCTTGTTTATTAACATT   [207]
593_H.c.chioneus      AAGCTTACATGATTTATAGTTTCGGTATCAAATAAAGTTCTTGCCTTGTTTATTAACATT   [207]
9035_H.c.weymeri      AAGCTTACATGATTTATAGTTTCAGTATCTAATAAAGTTCTTGCGTTGTTTATTAAAATT   [207]
N235_H._numata        AAGCTTACATGATTTATAGTTTCGGTATCAAATAAAGTTCTTGCCTTGTTTATAAAAATT   [280]

[                             310       320       330       340       350       360]
[                             .         .         .         .         .         .]

m150_H.m.mocoa        TTTACCCTTTATTAAGAATACTAATAATAACTTATAATTATTTTTCGACTTTTTTT-CAA   [318]
m75_H.m.mocoa         TTTACCCTTTATTAAGAATACTAATAATAACTTATAATTATTTTTCGACTTTTTTT-CAA   [318]
249_H.m.melpone       TTTACCCTTTATTAAGAATACTAATAATAACTTATAATTAYTTTTCGACTTTTTTT-CAA   [322]
1391_H.m.rosina       TTTACCCTTTATTAAGAATACTAATAATAACTTATAATTATTTTTCGACTTTTTTT-CAA   [322]
1137_H.m.rosina       TTTACCCTTTATTAAGAATACTAATAATAACTTATAATTACTTTTCGACTTTTTTT-CAA   [322]
8510_H.m.cythera      TTTACCCTTTATTAAGAATACTAATAATAACTTATAATTAYTTTTCGACTTTTTTT-CAA   [322]
8511_H.m.cythera      TTTACCCTTTATTAAGAATACTAATAATAACTTATAATTAYTTTTCGACTTTTTTT-CAA   [322]
59_H.m.vulcanus       TTTACCCTTTATTAAGAATACTAATAATAACTTATAATTATTTTTCGACTTTTTTT-CAA   [322]
60_H.m.vulcanus       TTTACCCTTTATTAAGAATACTAATAATAACTTATAATTAYTTTTCGACTTTTTTT-CAA   [322]
8159_H.m.malleti      TTTACCCTTTATTAAGAATACTAATAATAACTTATAATTATTTTTCGACTTTTTTT-CAA   [316]
8542_H.m.plesini      TTTACCCTTAAT-AAAAATACTAATAATAACTTAT--AACTTTTTCGATTTTTTTTTCAA   [264]
8543_H.m.plesini      TTTACCCTTTATTAAAAATACTAATAATAACTTATAATTCTTATTCGATTTTTTTTCAA   [349]
8749_H.m.amarillis    TTTACCCTTAAT-AAAAATACTAATAATAACTTAT--AACTTTTTCGACTTTTTTT-CAA   [263]
8747_H.m.amarillis    TTTACCCTTAAT-AAAAATACTAATAATAACTTAT--AACTTTTTCGATTTTTTTTTCAC   [264]
594_H.c.chioneus      TTTACCCTTAAT-AAAAATACTAATAATAACTTAT--AACTTTTTCGACTTTTTTT-CAA   [263]
593_H.c.chioneus      TTTACCCTTAAT-AAAAATACTAATAATAACTTAT--AACTTTTTCGACTTTTTTT-CAA   [263]
9035_H.c.weymeri      TTTACCCTTTATTAAAAATACTAATAATAACTTATAATTCTTTTTCGATATTTTTTTCAA   [267]
N235_H._numata        TTTACTCTTTATTAAAAAAACTAAAAATAACTTATAATTCTTTTTCTATTTTTTTTTCAA   [340]

[                             370       380       390       400       410       420]
[                             .         .         .         .         .         .]

m150_H.m.mocoa        AATATAAGATTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGCATATCTTTCGAAG   [378]
m75_H.m.mocoa         AATATAAGATTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGCATATCTTTCGAAG   [378]
249_H.m.melpone       AATATAAGGTTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGTATATCTTTCGTAG   [382]
1391_H.m.rosina       AATATAAKATTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGTATATCTTTCGTAG   [382]
1137_H.m.rosina       AATATAAGGTTTTTTTACATCTCCACAGGTGTTCAATAGAAGCTGGTATATCTTTCGTAG   [382]
8510_H.m.cythera      AATATAAGGTTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGTATATCTTTCGTAG   [382]
8511_H.m.cythera      AATATAAGGTTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGTATATCTTTCGTAG   [382]
```

```
59_H.m.vulcanus       AATATAAGATTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGCATATCTTTCGTAG   [382]
60_H.m.vulcanus       AATATAAGGTTTTTYTACATCTCCACAGGTGTTCAATAGAAGCTGGCATATCTTTCGTAG   [382]
8159_H.m.malleti      AATATAAGATTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGTATATCTTTCGTAG   [376]
8542_H.m.plesini      AATATAAGATTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGTATATCTTTCGTAG   [324]
8543_H.m.plesini      AATATAAGATTTTTCTACATCTCCGCAGGTGTTCAATAGAAGCTGGTATATCTTTCGTAG   [409]
8749_H.m.amarillis    AATATAAGATTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGCATATCTTTCGTAG   [323]
8747_H.m.amarillis    AATATAAGATTTTTCTACATCTCCACAGGNGTTCAATAGAAGCTGGTATATCTTTCGTAG   [324]
594_H.c.chioneus      AATATAAGATTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGTGTATCTTTCGTAG   [323]
593_H.c.chioneus      AATATAAGATTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGTGTATCTTTCGTAG   [323]
9035_H.c.weymeri      AATATAAGATTTTTCTACATCTCCACAGGTGTTCAATAGAAGCTGGCATATCTTTCGTAG   [327]
N235_H._numata        AATATAAGGTTTTTTTACATCTCCACAGGGGTTCAATAGAAGCTGGTATATCTTTCGTAG   [400]

[                             430    ]
[                             .      ]

m150_H.m.mocoa        ACGGTGGTCAGGA???   [394]
m75_H.m.mocoa         ACGGTGGTCAGGAG??   [394]
249_H.m.melpone       ACG????????????   [398]
1391_H.m.rosina       ACGG???????????   [398]
1137_H.m.rosina       ACGG???????????   [398]
8510_H.m.cythera      ACGG???????????   [398]
8511_H.m.cythera      ACGGT??????????   [398]
59_H.m.vulcanus       ACG????????????   [398]
60_H.m.vulcanus       ACGGT??????????   [398]
8159_H.m.malleti      ACGGTGGTCAGGAG??   [392]
8542_H.m.plesini      ACGG???????????   [340]
8543_H.m.plesini      ACG????????????   [425]
8749_H.m.amarillis    ACGG???????????   [339]
8747_H.m.amarillis    ACG????????????   [340]
594_H.c.chioneus      ACGGTGGTCAGGA???   [339]
593_H.c.chioneus      ACGG???????????   [339]
9035_H.c.weymeri      ACGGTGGTCAGGAGTT   [343]
N235_H._numata        AC?????????????   [416]
;
END;
```

Chapter 5

Conclusions: What can phylogenies tell about speciation?

Phylogeny of *Heliconius*

Heliconius butterflies have been considered model organisms for ecological and evolutionary studies, but lack of a well-supported phylogeny has left some of the proposed evolutionary hypotheses untestable. The molecular phylogeny presented here represents the best-supported species-level hypothesis of relationships proposed for *Heliconius* to date. This phylogenetic hypothesis can now be used to test hypotheses regarding evolutionary patterns of diversification across the genus. The *Heliconius* phylogeny is consistent with the proposed scenario of rapid speciation for this group and the data suggested that unique ecological traits of *Heliconius* such as pupal mating and pollen feeding behaviour evolved only once.

A recent study of *Papilio* used a well-resolved phylogeny based on mitochondrial DNA (COI-COII) and nuclear (*Ef1α*) DNA sequence data to estimate divergence rates for these two genes (Zakharov *et al*., 2004). They estimated that the early diversification within *Papilio* occurred 55–65 million years ago, based on a combination of biogeographic calibration points rather than fossils, and proposed that the substitution rates for the mtDNA genes *COI* and *COII* and the nuclear gene *Ef1α* are 7.8–10.2 x 10^{-3} and 1.32–2.0 x 10^{-3} per site per million year, respectively. These rates are 2–4 (for mtDNA) and up to 30 (for *Ef1α*) times slower than the "standard" substitution rates for *COI* in *Drosophila* (20.0–29.0 x 10^{-3} substitutions per site per million years, Beckenbach *et al*., 1993). However, there is now abundant evidence for a diversity of mtDNA rates in insects (Zakharov *et al*., 2004). The amount of sequence divergence between *Papilio* subdivisions is equivalent to divergences between genera in other tribes of the Papilionidae, and between genera of moths of the noctuid subfamily Heliothinae (Zakharov *et al*., 2004).

Using the Brower (1994) and Zakharov *et al*. (2004) substitution rates I estimated the dates of pairwise divergences representing the major splits in *Heliconius* (Table 1). The first striking observation is that in *Heliconius* there is no overlap between the dates estimated from the two genes. This means that there must be a difference in the relative rate of evolution for both genes between *Papilio* and *Heliconius*, with mtDNA evolving even faster than *Ef1α* in *Heliconius* as compared to *Papilio*. This suggests that there might be not a common molecular clock for butterflies.

Table 1. Percentage divergence and age estimates in millions of years of internal nodes in *Heliconius* using substitution rates suggested by Brower (1994) and Zakharov *et al*. (2004) (Upper and lower values).

			MtDNA				***Ef1α***	
		% Div.	Brower	Upper	Lower	**% Div.**	Upper	Lower
H. m. rosina	*H. m. rosina*	0.2	0.1	0.2	0.2	1.0	7.3	4.8
H. m. rosina	*H. m. melpomene*	1.6	0.7	2.1	1.6	1.0	7.4	4.9
H. m. rosina	*H. cydno*	4.4	2.0	5.7	4.3	1.0	7.5	4.9
H. m. rosina	*H. numata*	10.1	4.6	12.9	9.9	0.7	5.5	3.6
H. m. rosina	*H. erato*	21.4	9.7	27.5	21	4.1	31.1	20.5
H. m. rosina	*Neruda*	15.5	7.0	19.9	15.2	4.5	34.4	22.7
H. m. rosina	*L. doris*	18.7	8.5	23.9	18.3	5.6	42.8	28.2
H. m. rosina	*Eueides*	15.9	7.2	20.4	15.6	7.8	59.4	39.2
H. m. rosina	*Dryas*	22	10	28.2	21.6	6.3	47.8	31.5

The dates estimated for *Heliconius* suggests that divergence between species is older than Pleistocene. For example, *H. melpomene* and *H. cydno* one of the youngest sister species in the genus, showed a divergence of 2 million years according to Brower (1994) and from 5.7 to 4.3 million of years according to Zakharov *et al*. (2004).

It is clear that estimation of divergence times remains difficult in *Heliconius*, due to a lack of fossil record or well-dated vicariance events in the group. Nonetheless, mitochondrial (COI-COII) and nuclear (*Ef1α*) DNA sequence data have provided considerable new resolution for the phylogeny of *Heliconius*, a large and important genus for research in ecology, genetics, and evolutionary

biology. This phylogeny provides a number of insights into the evolutionary history of the group.

Ecological speciation

Colour pattern plays an important role in mate detection and assortative mating in many organisms. For example, in vertebrates, ultraviolet sexual dimorphism in blue tits (Andersson *et al*., 1998), face-colour polymorphism in Gouldian finches (Fox *et al*., 2002), carotenoid-based yellow plumage colour in American goldfinches (MacDougall and Montgomerie, 2003), colour of the tail feathers in northern flickers (Wiebe and Bortolotti, 2001), male courtship colour in cichlid fishes (Knight and Turner, 2004; Van Oppen *et al*., 1998; Wilson *et al*., 2000), polychromatisms in tilapia (Nxomani *et al*., 1999), and male colour in sticklebacks (Scott, 2004) are known to play a role in mate choice. In butterflies, male ultraviolet wing colour in sulphur butterflies (Silberglied and Taylor, 1978) play a similar role, and even in humans, blue eye colour among US whites (Grant and Lauderdale, 2002), and hair and eye colour (Little *et al*., 2003) have been shown to play a role in selection of partners. Finally, in plants there is evidence for phenotypic assortative mating for flower colour in periwinkle, where the observed higher frequency of white x white flower matings compared with white x pink flower matings appeared to be due to the constancy of flower colour exhibited by butterfly pollinators during flower visits (Kulkarni, 1999). Similarly, bumblebees significantly preferred blue flowers over yellow flowers, inducing positive frequency-dependent selection, assortative mating, and directional selection on different corolla colour morphs of the plant population visited (Smithson and Macnair, 1996).

Ecological speciation is the specific case where genes or phenotypic traits are both under divergent selection between environments and also cause reproductive isolation (Schluter, 2001). For example, the colour patterns of *Heliconius* butterflies are under selection for mimicry leading to selection against hybrids, but are also involved in mate choice leading to assortative mating (Jiggins *et al*., 1996; Jiggins *et al*., 2001; Mallet *et al*., 1998). Other examples occur in threespine

sticklebacks in postglacial lakes, where body size and nuptial coloration are under ecological selection and also influence mate choice (Boughman, 2001; Nagel and Schluter, 1998). In Darwin's ground finches, beak and body size are under divergent selection, and influence vocal signals that are used as cues in interspecific mate discrimination (Podos, 2001; Ratcliffe and Grant, 1985). In all these cases the ecological selection reduces hybrid fitness in the traits that are also used as mating cues. In threespine sticklebacks, F1 hybrids grew more slowly in both parental habitats (Hatfield and Schluter, 1999). In Darwin's ground finches F1 hybrid survival fluctuates with resource distribution (Grant and Grant, 1993; Grant and Grant, 1992).

Here I have shown that phylogeny supports the hypothesis that changes in colour pattern are important in the speciation of *Heliconius*: clades with more species tend to have more colour pattern diversity. Furthermore, an analysis of geographic patterns implies that speciation commonly occurred in parapatry or even sympatry. This suggests that speciation is facilitated by the evolution of novel mimetic colour patterns.

Relationships between sister species

Heliconius contains many closely related species that hybridise in the wild at low levels (Mallet *et al*., 1998). This implies rapid diversification with on-going gene flow, which can cause problems for phylogenetic inference, because speciation does not necessarily isolate all regions of the genome at a single instant. Therefore, instantaneous reciprocal monophyly between recently diverged sister species might be not expected. For this reason, multiple locus comparisons between fast-evolving genes are needed to resolve phylogenies among closely related species.

I compared mtDNA and two nuclear introns (*Tpi* and *Mpi*) for the sister species pairs *H. melpomene*/*H. cydno*, and provided evidence of recent introgression between these species. The three gene regions studied are evolving at similar rates, and the observation of more closely related *Mpi* alleles found between *H.*

melpomene and *H. cydno* than at the other two loci, in spite of similar rates of molecular evolution suggests recent introgression. There is hybrid female sterility between the species, associated with the sex-linked *Tpi* locus in backcross experiments. This hybrid sterility might be expected to prevent introgression at both mtDNA and *Tpi*, while allowing the flow of some nuclear genes. Thus, it seems likely that gene flow occurs during speciation in these species.

The *Tpi* and *Mpi* markers also provided strong evidence against simultaneous diversification of *H. erato* and *H. melpomene* wing patterns. In both species, genealogies were most consistent with a history of population growth rather than with constant effective population size, but the demographic histories suggested different time scales of divergence. In *H. erato* the genealogies of both nuclear loci suggested a most-recent common ancestor in the Pliocene period, while divergence in *H. melpomene* implied a more recent origin near the Pliocene–Pleistocene boundary (Flanagan *et al*., 2004). These data tell us that divergence was likely not simultaneous in the two species, but is not very informative in resolving relationships between colour pattern races.

Genetic basis of adaptive traits

The evolutionary relationships of colour pattern in *Heliconius* might be resolved by a phylogeny based on the molecular regions that control colour pattern polymorphism. However, there are to date very few examples of adaptive traits in wild populations whose molecular basis has been identified. For example, recent work on the molecular genetic basis of melanic phenotypes in hair and feather colour has identified several candidate genes in mammals and birds. One example is melanocortin-1 receptor (MC1R) that controls melanin distribution. Reconstruction of the evolution of MC1R sequences in primates showed the presence of mutations at important functional sites in several lineages (Mundy *et al*., 2003). In bananaquit MC1R alleles showed that melanism is a derived trait in this species (Mundy *et al*., 2003). In rock pocket mice, results indicate that adaptive melanism has arisen at least twice and that these similar phenotypic changes have a different genetic basis (Hoekstra and Nachman, 2003). Finally,

phylogenetic reconstructions of variant MC1R alleles in geese and skuas showed that melanism is a derived trait that evolved in the Pleistocene (Mundy *et al*., 2004). These results confirm the utility of a candidate gene approach to understanding colour evolution in vertebrates.

Colour pattern radiation in *Heliconius* species similarly has a simple genetic basis, with divergence mostly due to a few genes of major effect (Jiggins and McMillan, 1997; Linares, 1997; Naisbit *et al*., 2003; Sheppard *et al*., 1985), but the molecular basis of these genes is unknown. Here I have used sequences for a marker linked to the *Yb* locus, that codes for presence or absence of a yellow band in the hindwing in *H. melpomene* to study the relationships of these races. The resulting cladogram matched biogeographical distributions better than wing pattern similarity, suggesting multiple origins of the yellow bar colour pattern. However, until the actual genes controlling pattern can be identified, this conclusion remains tentative. Nonetheless, the primers designed here are the first step towards identification and positional cloning of this colour pattern gene.

In summary, species-level phylogenies derived from molecular data provide an indirect record of the speciation events that have produced the diversity today. Phylogenetic trees contain information about the relative timing of reconstructed speciation events, and can be used to estimate speciation rates. Also, with information on the geographical and ecological attributes of species, they can provide information on the causes of speciation within clades. This approach can be complemented with the use of genealogies reconstructed from different genes. Genealogies allow accurate reconstruction of species histories and insights into population processes.

References

Andersson S, Ornborg J, Andersson M (1998) Ultraviolet sexual dimorphism and assortative mating in blue tits. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 445-450.

Beckenbach AT, Wei Y-W, Liu H (1993) Relationships in the *Drosophila obscura* species group inferred from mitochondrial cytochrome oxidase II sequences. *Molecular Biology and Evolution* **10**, 619-634.

Boughman JW (2001) Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature* **411**, 944-948.

Brower AVZ (1994) Phylogeny of *Heliconius* butterflies inferred from mitochondrial DNA sequences (Lepidoptera: Nymphalidae). *Molecular Phylogenetics and Evolution* **3**, 159-174.

Flanagan N, Tobler A, Davison A, *et al*. (2004) Historical demography of Mullerian mimicry in the Neotropical *Heliconius* butterflies. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9704-9709.

Fox S, Brooks R, Lewis MJ, Johnson CN (2002) Polymorphism, mate choice and sexual selection in the Gouldian finch (*Erythrura gouldiae*). *Australian Journal of Zoology* **50**, 125-134.

Grant BR, Grant PR (1993) Evolution of darwin finches caused by a rare climatic event. *Proceedings of the Royal Society of London Series B-Biological Sciences* **252**, 253-253.

Grant MD, Lauderdale DS (2002) Cohort effects in a genetically determined trait: eye colour among US whites. *Annals of Human Biology* **29**, 657-666.

Grant PR, Grant BR (1992) Hybridization of bird species. *Science* **256**, 193-197.

Hatfield T, Schluter D (1999) Ecological speciation in sticklebacks: Environment-dependent hybrid fitness. *Evolution* **53**, 866-873.

Hoekstra HE, Nachman MW (2003) Different genes underlie adaptive melanism in different populations of rock pocket mice. *Molecular Ecology* **12**, 1185-1194.

Jiggins CD, McMillan WO, Neukirchen W, Mallet J (1996) What can hybrid zones tell us about speciation? The case of *Heliconius erato* and *H. himera* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society* **59**, 221-242.

Jiggins CD, McMillan WO (1997) The genetic basis of an adaptive radiation: warning colour in two *Heliconius* species. *Proceedings of the Royal Society of London Series B-Biological Sciences* **264**, 1167-1175.

Jiggins CD, Naisbit RE, Coe RL, Mallet J (2001) Reproductive isolation caused by colour pattern mimicry. *Nature* **411**, 302-305.

Knight ME, Turner GF (2004) Laboratory mating trials indicate incipient speciation by sexual selection among populations of the cichlid fish *Pseudotropheus zebra* from Lake Malawi. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, 675-680.

Kulkarni RN (1999) Evidence for phenotypic assortative mating for flower colour in periwinkle. *Plant Breeding* **118**, 561-564.

Linares M (1997) The ghost of mimicry past: Laboratory reconstitution of an extinct butterfly 'race'. *Heredity* **78**, 628-635.

Little AC, Penton-Voak IS, Burt DM, Perrett DI (2003) Investigating an imprinting-like phenomenon in humans Partners and opposite-sex parents have similar hair and eye colour. *Evolution and Human Behavior* **24**, 43-51.

MacDougall AK, Montgomerie R (2003) Assortative mating by carotenoid-based plumage colour: a quality indicator in American goldfinches, Carduelis tristis. *Naturwissenschaften* **90**, 464-467.

Mallet J, McMillan WO, Jiggins CD (1998) Mimicry and warning colour at the boundary between races and species. In: *Endless Forms: Species and Speciation* (eds. Howard DJ, Berlocher SH), pp. 390-403. Oxford University Press, New York.

Mundy NI, Kelly J, Theron E, Hawkins K (2003) Evolutionary genetics of the melanocortin-1 receptor in vertebrates. In: *Melanocortin System*, pp. 307-312.

Mundy NI, Badcock NS, Hart T*, et al*. (2004) Conserved genetic basis of a quantitative plumage trait involved in mate choice. *Science* **303**, 1870-1873.

Nagel L, Schluter D (1998) Body size, natural selection, and speciation in sticklebacks. *Evolution* **52**, 209-218.

Naisbit RE, Jiggins CD, Mallet J (2003) Mimicry: developmental genes that contribute to speciation. *Evolution & Development* **5**, 269-280.

Nxomani C, Ribbink AJ, Kirby R (1999) DNA profiling of *Tilapia guinasana*, a species endemic to a single sinkhole, to determine the genetic divergence between color forms. *Electrophoresis* **20**, 1781-1785.

Podos J (2001) Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature* **409**, 185-188.

Ratcliffe LM, Grant PR (1985) Species recognition in Darwin finches (*Geospiza, Gould*) .3. Male responses to playback of different song types, dialects and heterospecific songs. *Animal Behaviour* **33**, 290-307.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution* **16**, 372-380.

Scott RJ (2004) Assortative mating between adjacent populations of threespine stickleback (Gasterosteus aculeatus). *Ecology of Freshwater Fish* **13**, 1-7.

Sheppard PM, Turner JRG, Brown KS, Benson WW, Singer MC (1985) Genetics and the evolution of mullerian mimicry in *Heliconius* butterflies. *Philosophical Transactions of the Royal Society of London (B)* **308**, 433-613.

Silberglied RE, Taylor OR (1978) Ultraviolet reflection in courtship of the sulphur butterflies, *Colias eurytheme* and *C. philodice* (Lepidoptera, Piridae). *Behavior Ecology and Sociobiology* **3**, 203-243.

Smithson A, Macnair MR (1996) Frequency-dependent selection by pollinators: Mechanisms and consequences with regard to behaviour of bumblebees *Bombus terrestris* (L) (Hymenoptera: Apidae). *Journal of Evolutionary Biology* **9**, 571-588.

Van Oppen MJH, Turner GF, Rico C*, et al*. (1998) Assortative mating among rock-dwelling cichlid fishes supports high estimates of species richness from Lake Malawi. *Molecular Ecology* **7**, 991-1001.

Wiebe KL, Bortolotti GR (2001) Variation in colour within a population of northern flickers: a new perspective on an old hybrid zone. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **79**, 1046-1052.

Wilson AB, Noack-Kunnmann K, Meyer A (2000) Incipient speciation in sympatric Nicaraguan crater lake cichlid fishes: sexual selection versus ecological diversification. *Proceedings of the Royal Society of London Series B-Biological Sciences* **267**, 2133-2141.

Zakharov EV, Caterino MS, Sperling FAH (2004) Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera : Papilionidae). *Systematic Biology* **53**, 193-215.