Phylogenetic Discordance at the Species Boundary: Comparative Gene Genealogies Among Rapidly Radiating *Heliconius* Butterflies

Margarita Beltrán,*† Chris D. Jiggins,* Vanessa Bull,† Mauricio Linares,‡ James Mallet,† W. Owen McMillan,§ and Eldredge Bermingham*

*Smithsonian Tropical Research Institute; †The Galton Laboratory, Department of Biology, University College London; ‡Instituto de Genética, Universidad de los Andes; and §Department of Biology, University of Puerto Rico

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 ($\approx$550 bp) of sex-linked triose-phosphate isomerase (*Tpi*), intron 3 ($\approx$450 bp) of autosomal mannose-phosphate isomerase (*Mpi*), and 1,603 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, *Heliconius erato-Heliconius himera* and *Heliconius melpomene-Heliconius 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 (mtDNA) 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 noncoding nuclear regions and describe their mode and tempo of evolution relative to the mitochondrial protein-coding genes, cytochrome oxidases I and II (COI and 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 hybridize 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 10 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 recognized species groups and show a number of species pairs separated by <4% mtDNA sequence divergence, implying divergence within the last two million years (Brower 1994; Brower and Egan 1997; Brower and DeSalle 1998). Two *Heliconius* species in particular, *Heliconius erato* and *Heliconius melpomene*, are recognized as excellent model systems for the study of both intraspecific 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 America 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 hybridization still occurs (Jiggins et al. 1996, 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 *Hel-*

iconius (Brower 1994). (2) What are the relationships between sister species and populations in the *melpomene* and *erato* species groups. The mtDNA data imply that *H. melpomene* is paraphyletic with respect to its sister species *Heliconius cydno* and fails to resolve relationships between *H. erato* and the closely related *Heliconius himera* (Brower 1996*a*). Studies of pre- and post-mating isolation between these sister species suggest they speciated recently as a result of divergence in habitat and color pattern (McMillan, Jiggins, and Mallet 1997; Jiggins et al. 2001*b*). 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 multilocus phylogeny of species pairs can produce insights into speciation processes as well as clarify 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 with those in the mitochondrial COI and COII genes. The two mtDNA genes evolve rapidly and are already known to be informative in studies of divergence ranging from intraspecific biogeography of races to relationships among tribes and subfamilies in *Heliconius* (Brower 1994, 1996*b*). Specifically, our goals are (1) to compare evolutionary patterns between mtDNA and nuclear genes and between introns and exons within nuclear genes, (2) to determine phylogenetic levels at which these genes are informative, and (3) to describe genealogical relationships between races and sister species at the loci studied.

Materials 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). *Heliconius heurippa* and *Heliconius 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 taxa within the Heliconiini. Butterflies were collected and 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, one-third 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
mtDNA

A region of mtDNA spanning the 3' end of COI, leucine tRNA, and COII was amplified from individual genomic DNA using polymerase chain reaction (PCR). The identity of this region was confirmed by comparison with sequences of *H. cydno* (GenBank accession number U0851) and *Drosophila yakuba* (GenBank accession number 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 (Simon et al. 1994; table 2). For both pairs of primers we used a cycling profile of 48°C for 45 s and 72°C for 60 s (four cycles), followed by 94°C for 45 s, 52°C for 45 s, and 72°C for 1 min and 30 s for 29 cycles. Twenty-five–microliter reactions contained 2 μl of DNA, 1× buffer, 2 mM $MgCl_2$, 0.8 mM dNTPs, 0.5 mM of each primer, and 0.025 U/μl of Amplitaq polymerase.

The PCR products were electrophoretically separated on 1.5% low–melting point agarose with ethidium bromide (1 μg/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 10-μl cycle sequence reaction mixture contained 2 μl dRhodamine, 2 μl Half-term, 1 μl primer, and 5 μl template. The cycle profile was 96°C for 15 s, cooling at 1°C/s to 50°C, and heating at 1°C/s to 60°C for 4 min, repeated for 24 cycles. The product was cleaned over Centrisep columns filled with 700 μl G-50 Sephadex and dried. The samples were resuspended in 0.9 μl of a 5:1 deionized formamide: bluedextran-EDTA (pH 8.0) solution, denatured at 90°C for 2 min and loaded into 6% acrylamide gels. Gels were run on an ABI Prism 377 Sequencer (PE Applied Biosystems) for 7 h.

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*

Table 1
Heliconiini Taxa Included in the Study

			SPECIES			NUMBER OF ALLELES			ALIGNMENT NUMBER	
Group	Identifica-tion Number	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	F	*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	—	—	—

NOTE.—Identification numbers are Smithsonian collection numbers and are prefixed by "stri-b-." "Number of 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 and 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 Number of Alleles column, where * indicates an agreement between the results derived from cloning and TTGE; in two cases of disagreement the number of bands visible on the TTGE gel is given in parentheses.

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 *Heli-*

conius specific primers were designed that consistently amplified the *Tpi* region out of genomic DNA preparations.

For *Tpi*, the primers were situated in exons 3 and 4 of *Heliothis* (GenBank accession number 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. 2001*a*; A. Tobler et al. personal communication). Double-stranded DNA was synthesized in 25-µl reactions containing 2 µl of genomic DNA, 1× buffer, 3 mM MgCl$_2$, 0.8 mM dNTPs, 0.5 mM of each primer, and 0.03 U/µl of *Taq* gold polymerase. DNA was amplified using the following step-cycle profile: 94°C for 7 min, 94°C for 45 s, 58°C for 45 s, 72°C for

Table 2
Primers Used to Amplify *Heliconius* mtDNA and Nuclear DNA

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

NOTE.—Cytochrome oxidase (COI and COII) 3′ end positions are given relative to *D. yakuba* (X03240). Positions for *Tpi* and *Mpi* are given relative to *Heliothis* (U23080) and *H. 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.

1 min and 45 s for 10 cycles with the annealing temperature reduced 0.5°C per cycle, then 25 cycles with annealing temperature of 53°C. 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 three to five 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 (*H. sapiens* GenBank accession numbers AF227216 and 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 25-μl PCR reaction mixture contained 2 μl of DNA, 1× buffer, 3 mM MgCl₂, 0.8 mM dNTPs, 0.5 mM of each primer, and 0.03 u/μl of Amplitaq. Amplification was carried out using the following step-cycle profile: 94°C for 3 min, 94°C for 40 s, 55°C for 40 s, and 72°C for 45 s 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 tris-acetic acid-EDTA and were run from 46 to 53°C at a temperature ramp of 1°C/h. The results are summarized 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* groups (see *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 minimized 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 comparisons of different sequences inferred to represent a single allele were compatible with this assumption. For *Tpi*, where more than 1 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

Table 3
Variability in Each Gene Region

Gene	Region	Number of Species	Number of Individuals	Number of Alleles	Alignment Number	Number of Base Pairs	Number of 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

NOTE.—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 (first/second/third). To facilitate comparison between gene regions, the CO data are given for all taxa and separately for 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.

among codon positions. Phylogenetic analyses and genetic distances (figs. 1–4) were calculated 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 with 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 (1,000 replicates, heuristic search with TBR branch swapping). In figures 3 and 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 with 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 66 bp of the *Mpi* exon.

Results and Discussion
Patterns of Molecular Evolution in *Heliconius* mtDNA, *Tpi,* and *Mpi* Genes
mtDNA

The final aligned mitochondrial sequences yielded 1,603 characters including nucleotides and gaps from 37 individuals (GenBank accession numbers AF413672–AF413708). This represents 822 bp of the COI gene corresponding to positions 2191 to 3009 of the *D. yakuba* sequence (X03240), the complete leucine-tRNA gene (70 bp), 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 region, which shows a 1-bp indel (in *Heliconius elevatus* and *Heliconius 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 number 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 number 126 in *Dryadula phaetusa,* number 127

in *Heliconius sara*, and number 129 in *D. iulia* (see also Brower 1994).

Of the 1,603 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 with 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 10 times more frequent than were transversions (table 4).

Tpi

We obtained 155 bp 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 155 bp observed were variable. The *Tpi* exon had a higher GC content (49%) than did 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 *Heliconius 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 66 bp 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 base pairs corresponded to positions 1601 to 1627 of exon 3 in the *H. sapiens* reference sequence (AF227216), and 39 bp to positions 417 to 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 nonsynon-

Table 4
Best-Supported Models of Molecular Evolution and Estimated Parameter Values for the Mitochondrial COI and COII Genes and the *Tpi* and *Mpi* Combined Intron and Exon Regions

Gene Region	Alignment Number	Model	Base Composition						Substitution Rates						
									Tr		Tv				Tr/Tv
			a	c	g	t	I	α	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	1.91
	2	HKY + G	0.34	0.15	0.19	0.32	0	0.79	—	—	—	—	—	—	1
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

NOTE.—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.

ymous 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 with 13 amino acid changes across only 31 amino acid positions in *Mpi*. The different amino acid replacement pattern between the two genes coincides with the higher proportion of first and second 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. erato 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*-silvaniform 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 (figs. 3 and 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 (fig. 3*a*). However, in the *erato* group, some indels at both *Tpi* and *Mpi* were not concordant with the sequence-based ML 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. 3*b*). However,
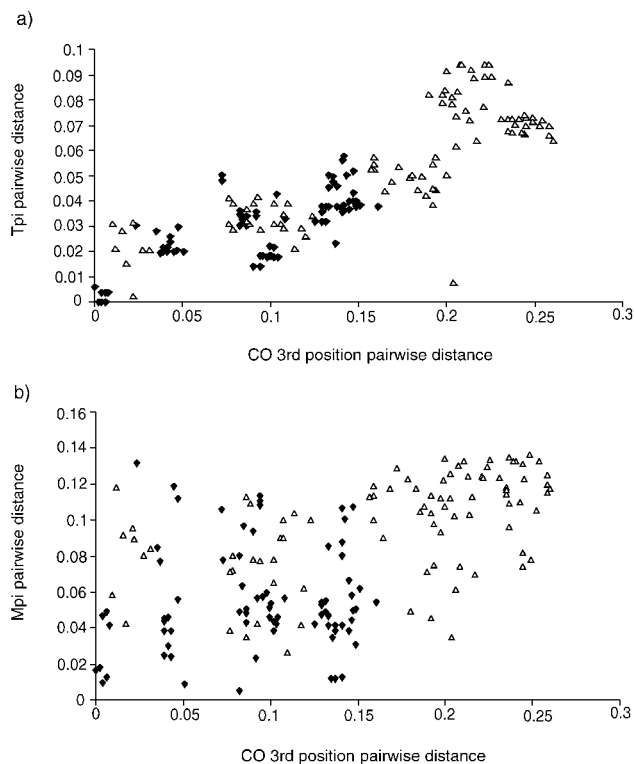

FIG. 1.—Uncorrected genetic distances between *Heliconius Tpi* and *Mpi* sequences plotted as a function of divergence in third codon positions of the mitochondrial COI and COII genes. Open triangles show comparisons in the *erato-sapho* clade (alignments 2 and 4) and closed diamonds the *melpomene*-silvaniform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

in the case of *Mpi* in the *erato* group there was no way to make the phylogeny consistent with all indels. The 7-bp deletion at position 473 (shown as a filled oval in fig. 4*b*) grouped alleles 2842#1 (*H. himera*), 590 (*H. hecalesia*) and 2923 (*H. charithonia*), whereas a 2-bp deletion at position 222 grouped *H. hecalesia* and *H. charithonia* with *H. sara*, *Heliconius eleuchia*, and *H. sapho*. The 2-bp indel was consistent with the ML tree, whereas the 7-bp indel appears to be genuinely homoplasious (fig. 4*b*).

   Natural recombination was, in contrast, apparently rare. Another apparently homoplasious indel, a 5-bp deletion at position 353 (open oval), appears to result from natural recombination. The *erato* 2980#1 allele appears to be a recombinant between 2980#2 and an allele which was not sampled, similar to 2981#1 (see *Supplementary Material*). Recombination could generate high levels of homoplasy, reducing phylogenetic signal. However, MP-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 in the mitochondrial CO genes (0.36, fig. 2). Thus, homoplasy was lower in the nuclear genes.
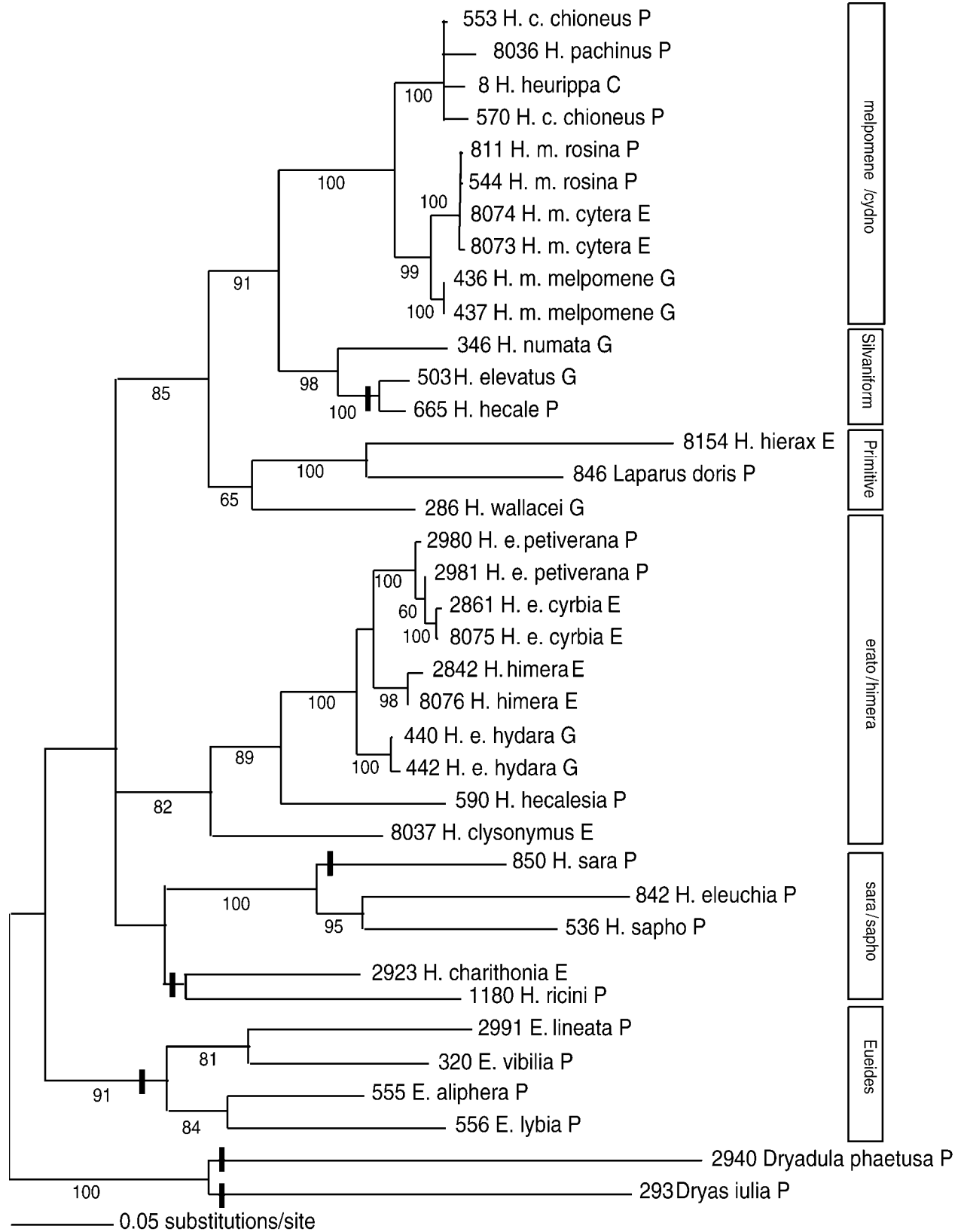
FIG. 2.—ML phylogeny for Heliconiina species based on combined mitochondrial COI and COII sequences. All branches were collapsed which were not supported with either >95% confidence using likelihood ratio tests (implemented in Paup*) or >50 bootstrap support or by phylogenetically informative indel characters. 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 1,000 replicate bootstraps. Branches without parsimony bootstrap support reflect differences between the MP and ML trees. Sample numbers correspond to those given in table 1. P = Panama, E = Ecuador, G = French Guiana, and C = Colombia.
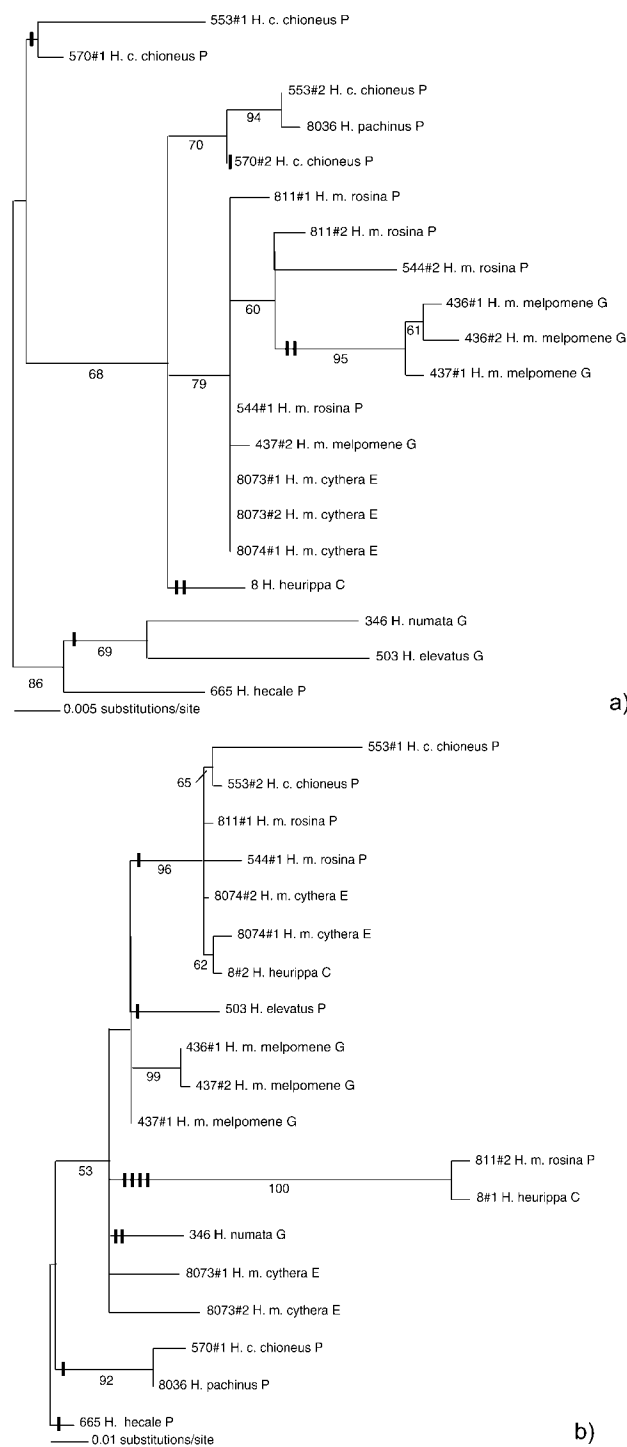
553#1 H. c. chioneus P
570#1 H. c. chioneus P
94 — 553#2 H. c. chioneus P
70 — 8036 H. pachinus P
570#2 H. c. chioneus P
811#1 H. m. rosina P
811#2 H. m. rosina P
544#2 H. m. rosina P
60 — 436#1 H. m. melpomene G
61 — 436#2 H. m. melpomene G
95 — 437#1 H. m. melpomene G
68 79 — 544#1 H. m. rosina P
437#2 H. m. melpomene G
8073#1 H. m. cythera E
8073#2 H. m. cythera E
8074#1 H. m. cythera E
8 H. heurippa C
346 H. numata G
69 — 503 H. elevatus G
86 — 665 H. hecale P
0.005 substitutions/site

a)

553#1 H. c. chioneus P
65 — 553#2 H. c. chioneus P
811#1 H. m. rosina P
96 — 544#1 H. m. rosina P
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
53 — 811#2 H. m. rosina P
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

b)

FIG. 3.—ML phylogenies for the *melpomene-cydno* group (alignments 1 and 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, identification number 811. P = Panama, E = Ecuador, G = French Guiana, and C = Colombia.

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, 1996*b*). 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 third positions were compared with *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-in-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 six-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 (fig. 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): (1) the three-parameter TrN + G model (Tamura and Nei 1993) for *Tpi* in the *melpomene*-silvaniform group (alignment 1); (2) 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 (3) 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 with 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-
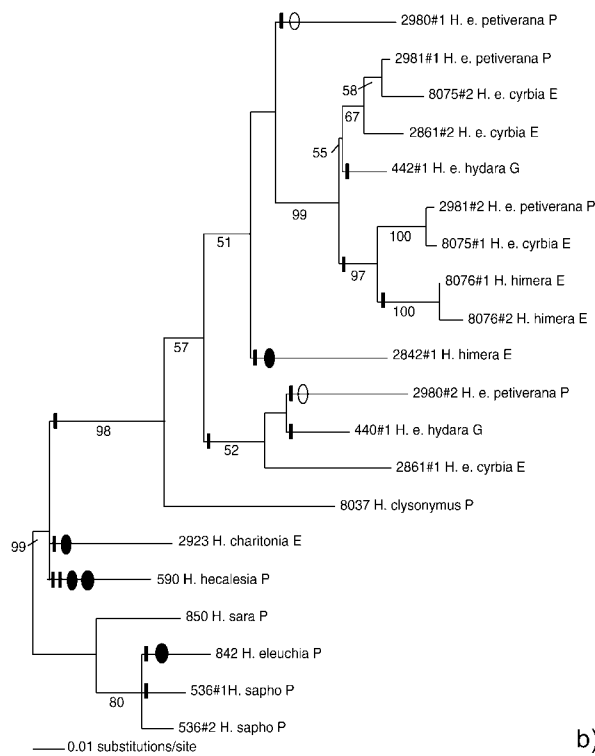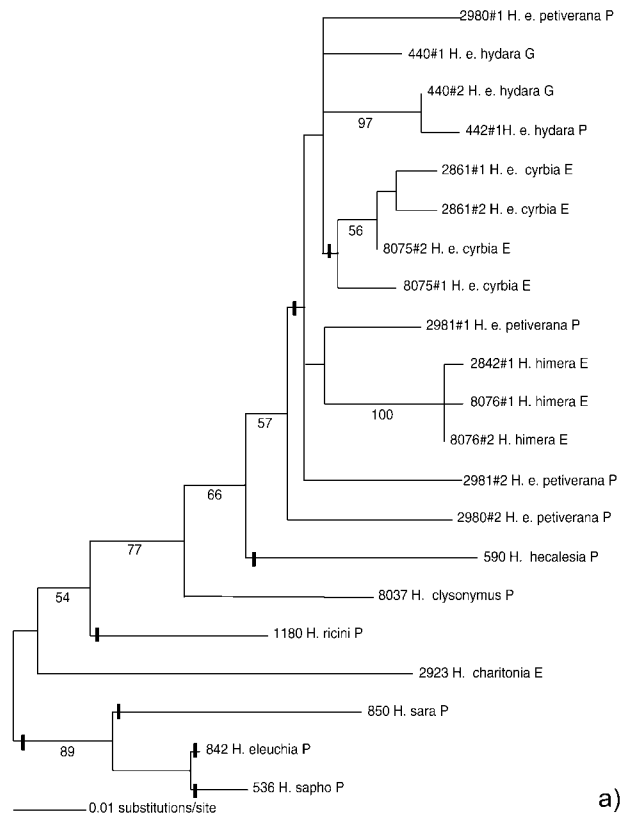
a)

b)

FIG. 4.—Maximum-likelihood phylogenies for *erato-himera* group (alignments 2 and 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

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 (figs. 3 and 4). Phylogenetic resolution was somewhat less at nuclear loci compared with the mtDNA (compare fig. 2 with figs. 3 and 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 (figs. 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, *H. hecalesia*, *Heliconius hierax*, and *Eueides lineata*.

In our ML tree (fig. 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

←

example 2980#1 and 2980#2 are two *Tpi* separate alleles cloned from the same individual *H. erato petiverana*, male, identification number 2980. P = Panama, E = Ecuador, and G = French Guiana.

Table 5
Results of SH Tests of Alternative *a priori* Hypotheses for the Phylogenetic Relationships Between *melpomene-cydno* and *erato-himera*

	Align-ment Number	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)

NOTE.—For this test, the ML 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 $P \le 0.05$ are shown with a *.

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. Analyses 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) 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 *Heliconius wallacei* (fig. 2; SH test; Delta = 62.47, P = 0.002). Combined analysis of the nuclear coding sequence from the *wingless, Tpi,* and *Mpi* genes provides 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 (fig. 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 *heu-*

rippa) varied depending on the gene region analyzed (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 1996*b*). In Brower's 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 (fig. 2). However, it is interesting to note that despite strong bootstrap support, ML 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, four transition substitutions (three C-T and one A-G) and three A-T transversions provide no significant support for *cydno* group monophyly (table 5); the SH test would seem to be overly conservative.

In contrast, the ML tree inferred from sequence variation in the *Mpi* intron (fig. 3*b*) shows *melpomene* and *cydno* paraphyletic with respect to the silvaniforms *H. elevatus*, *H. hecale*, and *Heliconius 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. 3*a*) 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* (fig. 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. 4*a*) *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 an 18-bp deletion at position 211 (fig. 4*a*) shared by all *himera* and *erato* alleles with the exception of *H. erato petiverana* 2980#2. In contrast, in the *Mpi* genealogy (fig. 4*b*) 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% to 1.2% per lineage per million years (Brower 1994) this suggests that both species pairs diverged from each other within the last 1½ 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* ge-

nealogies in both species pairs failed to show structure consistent with species boundaries, despite considerable resolution and well-supported nodes within the trees (figs. 3*b* and 4*b*). 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 harbor ancestral shared variation for longer time periods than sex-linked or maternally inherited genes such as *Tpi* and CO. In particular, maternally inherited mitochondrial genes will coalesce on average four times faster than autosomal genes do. 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 interspecific 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 was 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 hybridize in the

Table 6
Evaluation of the Three-Times Rule for *Heliconius* Species

Group	Mitochondrial Branch Length (l)	mtDNA Diversity (d)	Coalescence Ratio (l/d)	PREDICTED PERCENTAGE 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

NOTE.—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 five alleles, adjusting for the smaller population size of sex-linked genes in the case of *Tpi*.

wild. *Heliconius melpomene* and *cydno* are broadly sympatric, with hybrids forming perhaps 0.1% of overlapping populations (Jiggins et al. 2001*b*). 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, whereas the former would favor the "capture" of new alleles following rare interspecific hybridization.

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 highlight 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 recognizable 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.

Supplementary Material

Supplementary material is available at the MBE website, www.molbiolevol.org.

Acknowledgments

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 (UK) and National Science Foundation (USA) grants to JM and WOM respectively.

LITERATURE CITED

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. Pp. 289–303 *in* P. C. HOCH and A. G. STEPHENSON, eds, Experimental and molecular approaches to plant biosystematics. Missouri Botanical Garden, St. Louis.

BELTRÁN, M. 1999. Evidencia genética (Alozimas) para evaluar el posible origen híbrido de *Heliconius heurippa* (Lepidoptera: Nymphalidae). Masters dissertation, 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. J. Gen. Virol. **79**:2921–2928.

BROWER, A. V. Z. 1994. Phylogeny of *Heliconius* butterflies inferred from mitochondrial DNA sequences (Lepidoptera: Nymphalidae). Mol. Phylogenet. Evol. **3**:159–174.

———. 1996*a*. A new mimetic species of *Heliconius* (Lepidoptera: Nymphalidae), from southeastern Colombia, as revealed by cladistic analysis of mitochondrial DNA sequences. Zool. J. Linn. Soc. **116**:317–332.

———. 1996*b*. 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 Mol. Biol. **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. Proc. R. Soc. Lond. B **264**:969–977.

BROWN, K. S. 1981. The biology of *Heliconius* and related genera. Annu. Rev. Entomol. **26**:427–456.

CATERINO, M. S., and F. A. H. SPERLING. 1999. *Papilio* phylogeny based on mitochondrial cytochrome oxidase I and II genes. Mol. Phylogenet. Evol. **11**:122–137.

CRACRAFT, J. 1989. Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. Pp. 28–59 *in*D. OTTE and J. A. ENDLER, eds. Speciation and its consequences. 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 (N Y) **50**:191–254.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA-sequences—a maximum-likelihood approach. J. Mol. Evol. **17**:368–376.

GOLDMAN, N., J. P. ANDERSON, and A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. **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. Mol. Biol. Evol. **4**:144–158.

HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **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. 2001*a*. 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). Biol. J. Linn. Soc. **59**:221–242.

JIGGINS, C. D., R. E. NAISBIT, R. L. COE, and J. MALLET. 2001*b*. Reproductive isolation caused by colour pattern mimicry. Nature (Lond.) **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. Biochem. Genet. **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. Zool. J. Linn. Soc. **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. Proc. Natl. Acad. Sci. USA **92**:8507–8511.

MADDISON, W. P., and D. R. MADDISON. 1997. MacClade: analysis of phylogeny and character evolution. 3.07 edition. Sinauer Associates, Sunderland, Mass.

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. Pp. 390–403 *in*D. J. HOWARD and S. H. BERLOCH-ER, eds. Endless forms: species and speciation. Oxford University Press, New York.

MCMILLAN, W. O., C. D. JIGGINS, and J. MALLET. 1997. What initiates speciation in passion vine butterflies? Proc. Natl. Acad. Sci. 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 **161**: 1517–1526.

NICHOLS, R. 2001. Gene trees and species trees are not the same. Trends Ecol. Evol. **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. Mol. Biol. Evol. **5**:568–583.

PENZ, C. M. 1999. Higher level phylogeny for the passion-vine butterflies (Nymphalidae; Heliconiinae) based on early stage and adult morphology. Zool. J. Linn. Soc. **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. Mol. Biol. Evol. **15**:1172–1182.

RIFFARTH, H. 1901. Die Gattung Heliconius Latr.: Neu bearbeitet und Beschreibung neuer Formen. Berl. Entomol. Z. **46**:25–183.

SCHMIDT, P. 2001. The effects of diet and physiological stress on the evolutionary dynamics of an enzyme polymorphism. Proc. R. Soc. Lond. 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. Philos. Trans. R. Soc. Lond. B **308**:433–613.

SHIMODAIRA, H., and M. HASEGAWA. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **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 reaction primers. Ann. Entomol. Soc. Am. **87**:651–702.

SPERLING, F. A. H. 1994. Sex-linked genes and species differences in Lepidoptera. Can. Entomol. **126**:807–818.

SWOFFORD, D. L. 2000. PAUP*: phylogenetic analysis using parsimony (*and other methods). 4th edition. Sinauer Associates, Sunderland, Mass.

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. Mol. Biol. Evol. **10**:512–526.

TAVARÉ, S. 1984. Line of descent and genealogical processes and their applications in population genetic models. Theor. Popul. Biol. **26**:164–199.

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. Appl. Environ. Microbiol. **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. J. Mol. Evol. **39**:105–111.