Molecular Evolution of the *Wingless* Gene and Its Implications for the Phylogenetic Placement of the Butterfly Family Riodinidae (Lepidoptera: Papilionoidea)

*Dana L. Campbell,**[1] *Andrew V. Z. Brower,*† *and Naomi E. Pierce**

*Department of Organismic and Evolutionary Biology, Harvard University; and †Department of Entomology, Oregon State University

The sequence evolution of the nuclear gene *wingless* was investigated among 34 representatives of three lepidopteran families (Riodinidae, Lycaenidae, and Nymphalidae) and four outgroups, and its utility for inferring phylogenetic relationships among these taxa was assessed. Parsimony analysis yielded a well-resolved topology supporting the monophyly of the Riodinidae and Lycaenidae, respectively, and indicating that these two groups are sister lineages, with strong nodal support based on bootstrap and decay indices. Although *wingless* provides robust support for relationships within and between the riodinids and the lycaenids, it is less informative about nymphalid relationships. *Wingless* does not consistently recover nymphalid monophyly or traditional subfamilial relationships within the nymphalids, and nodal support for all but the most recent branches in this family is low. Much of the phylogenetic information in this data set is derived from first- and second-position substitutions. However, third positions, despite showing uncorrected pairwise divergences up to 78%, also contain consistent signal at deep nodes within the family Riodinidae and at the node defining the sister relationship between the riodinids and lycaenids. Several hypotheses about how third-position signal has been retained in deep nodes are discussed. These include among-site rate variation, identified as a significant factor by maximum likelihood analyses, and nucleotide bias, a prominent feature of third positions in this data set. Understanding the mechanisms which underlie third-position signal is a first step in applying appropriate models to accommodate the specific evolutionary processes involved in each lineage.

Introduction

The last 40 years have seen considerable activity in the higher systematics and classification of the ''true'' butterflies (Papilionoidea, including Hesperiidae). However, the multiple hypotheses that have arisen from these efforts do not provide a consistent interpretation of the evolution of the major butterfly lineages. In particular, assessing the monophyly and phylogenetic placement of the large family Riodinidae, which contains over 1,200 species, has been problematic, as can be seen by the conflicting phylogenetic hypotheses derived from morphological data. Although most morphological studies place the riodinids as the sister to the lycaenid butterflies and identify the nymphalids as the closest relatives to this riodinid + lycaenid clade (Ehrlich 1958; Ehrlich and Ehrlich 1967; Kristensen 1976; Scott and Wright 1990; DeVries 1991, 1997; Fiedler 1991; De Jong, Vane-Wright, and Ackery 1996), few characters support these relationships, and alternate relationships among these groups have been suggested (Robbins 1988*a*, 1988*b*; Martin and Pashley 1992; for review, see Campbell and Pierce 2000).

To determine the placement of riodinid butterflies, we examined a new set of molecular characters from *wingless* (*wg*), a nuclear gene which has shown utility in reconstructing species level to subfamily level relationships in nymphalids (Brower and DeSalle 1998). *Wingless* is a member of the wnt gene family, whose paralogs are easily distinguishable (Sidow 1992). Primers specific to lepidopteran *wingless* have been developed from the 3′ exon (Carroll et al. 1994; Brower and DeSalle 1998). Because this region of *wingless* showed a rapid rate of substitution in nymphalids, it holds promise for resolving family level relationships in the Lepidoptera (Brower and Egan 1997; Brower and DeSalle 1998). Previous molecular studies of the Papilionoidea have used few riodinid representatives (Martin and Pashley 1992; Weller, Pashley, and Martin 1996). Our work includes representative taxa from all the main lineages of riodinids, lycaenids, and nymphalids (*sensu* Harvey [1987], Eliot [1973], and Harvey [1991], respectively). The aims of this study are (1) to test the hypotheses that the Riodinidae and Lycaenidae are each monophyletic and determine the relationship between these two families and the family Nymphalidae, and (2) to describe patterns of the molecular evolution of *wingless* and demonstrate the utility of this gene for reconstructing family and subfamily level relationships among the Lepidoptera.

Materials and Methods

Specimens

Taxa were selected to represent each of the main lycaenid, riodinid, and nymphalid lineages (table 1). Previous studies of butterfly systematics agree on the Hesperiidae as the basal lineage of the Papilionoidea (but see Scoble 1986, 1988); for this reason, a hesperiid representative was included as an outgroup, as were two representatives of the Pieridae and one species of Papilionidae. Taxa were collected as fresh specimens, and the bodies were stored in 100% ethanol at −80°C.

[1] Present address: University Park, Maryland.

Key words: Riodinidae, Lycaenidae, Nymphalidae, molecular phylogenetics, *wingless*, gene utility, third codon positions, maximum likelihood.

Address for correspondence and reprints: Dana L. Campbell, Department of Biology, Biology/Psychology Building, University of Maryland, College Park, Maryland 20742. E-mail: dcampbell@oeb.harvard.edu.

Mol. Biol. Evol. 17(5):684–696. 2000

Table 1
Taxa Used in this Study and Their Classification According to Harvey (1987) (Riodinids), Eliot (1973) (Lycaenids), and Harvey (1991) (Nymphalids)

Family	Subfamily	Tribe	Species	Locality	GenBank Accession No.
Riodinidae	Euselasiinae		*Euselasia* sp.	Ecuador: Sucumbios Province	AF233534
	Hamearinae		*Taxila haquinus*	Malaysia: FRIM Kepong	AF233535
			Abisara saturata	Malaysia: Kuala Woh, Papah	AF233536
	Riodininae	Incertae sedis	*Cremna actoris*	Ecuador: Sucumbios Province	AF233537
		Eurybiini	*Eurybia* sp.	Ecuador: Sucumbios Province	AF233538
		Mesosemiini	*Mesosemia* sp.	Ecuador: Sucumbios Province	AF233539
		Riodinini	*Riodina lysippus*	Ecuador: Sucumbios Province	AF233540
		Charitini	*Sarota* sp.	Ecuador: Sucumbios Province	AF233541
		Emesini	*Emesis* sp.	Ecuador: Sucumbios Province	AF233542
		Nymphidiini	*Nymphidium cachrus*	Ecuador: Sucumbios Province	AF233543
		Helicopini	*Helicopis cupido*	Ecuador: Sucumbios Province	AF233544
		Lemoniini	*Thisbe irena*	Ecuador: Sucumbios Province	AF233545
Lycaenidae	Poritiinae	Poritiini	*Poritia phama*	Malaysia: Genting Tea Estate	AF233546
			Simiskina pheretia	Malaysia: Awana FR, Pahang	AF233547
		Liptenini	*Baliochila minima*	Kenya: Arabuko-Sokoke	AF233548
	Curetinae		*Curetis bulis*	Malaysia: FRIM Kepong	AF233549
	Miletinae	Miletini	*Miletis ancon*	Malaysia: FRIM Kepong	AF233550
		Liphyrini	*Liphyra brassolis*	Australia: Queensland	AF233551
			Spalgis epius	Malaysia: Genting Tea Estate	AF233552
	Theclinae		*Habrodais ganus*	USA: Nevada, Lang Crossing	AF233553
			Jalmenus daemeli	Australia: Queensland, Townsville	AF233554
	Polyommatinae		*Candalides geminus*	Australia: Queensland, Burra Range	AF233555
			Jamides alecto	Malaysia: FRIM Kepong	AF233556
	Lycaeninae		*Heliophorus kiana*	Malaysia: Kinabolu Park	AF233557
Nymphalidae...	Heliconiinae		*Heliconius erato*	French Guiana: Pointe Macouria	AF014127[a]
	Libytheinae		*Libytheana carienenta*	Brazil: Rondonia, Ariquemes	AF233558
	Satyrinae		*Cercyonis pegala*	USA: New York, Ithaca	AF014145[a]
	Morphinae		*Morpho helenor*	Brazil: Rondonia, Ariquemes	AF014144[a]
	Limenitidinae		*Limenitis arthemis*	USA: New York, Caroline	AF233559
			Diaethria clymena	Brazil: Rondonia, Ariquemes	AF014143[a]
	Nymphalinae		*Hypolimnas misippus*	Reared in captivity, ex. C. Clarke	AF233560
			Siproeta steneles	Brazil: Rondonia, Ariquemes	AF014142[a]
	Danainae		*Danaus plexippus*	Colombia: Meta, Villavisencia	AF233561
	Ithomiinae		*Melinaea maenius*	Brazil: Rondonia, Ariquemes	AF014146[a]
Pieridae	Pierinae		*Pieris rapae*	USA: New York, Ithaca	AF014148[a]
			Delias sp.	Australia: NSW, Lismore	AF233562
Papilionidae ...	Papilioninae		*Papilio glaucus*	USA: Colorado, Gunnison	AF233563
Hesperiidae ...			*Ancyloxypha numita*	USA: Massachusetts, Boston	AF233564

[a] From Brower and DeSalle (1998).

Wings were retained as vouchers in the Harvard Museum of Comparative Zoology (riodinids and lycaenids) and the American Museum of Natural History (nymphalids).

Molecular Methods

For the lycaenid, riodinid, and hesperiid species, genomic DNA was extracted using proteinase K digestion followed by phenol : chloroform (1:1) extraction (Maniatis, Fritsch, and Sambrook 1989). Genomic DNA for nymphalid, pierid, and papilionid species was prepared as described in Brower (1994).

The primers WG1, WG2, and WG2.1, described by Brower and DeSalle (1998), were used to amplify the 450-bp *wingless* fragment from all specimens. The amplification conditions consisted of 1 min denaturation at 94°C, 1 min annealing at 50°C, and 1 min extension at 72°C for 30 cycles. Final concentrations of reagents in a standard 100-μl reaction were 1× Taq buffer (50 mM KCl, 100 mM Tris-HCl, 0.1% Triton X-100), 3 mM $MgCl_2$, 0.5 μM each primer, and 2.5 U *Taq* polymerase. PCR products were excised from a 1.2% low-melt agarose gel and phenol : chloroform extracted. The purified products were cycle sequenced in both directions using the ABI dye terminator core kit in a Perkin-Elmer 480. Excess nucleotides were removed using CentriSep spin columns (Princeton Separations), and reactions were run out on an ABI 370 automated sequencer. PCR and sequencing of the *wingless* fragment for the nymphalids was performed as described in Brower and DeSalle (1998). Sequence chromatograms were edited and aligned by hand in the program SeqEd (Applied Biosystems, Inc. 1992, SeqEd 1.0.3). Edited nucleotide sequences were translated using GeneJockey (Taylor 1991). GenBank accession numbers for sequences are listed in table 1.

Data Analyses and Phylogenetic Methods

We used PAUP*, versions 4.0d54 and 4.0d59 (Swofford 1998), to calculate pairwise distance matrices, determine likelihood values, and conduct parsimony and LogDet distance analyses. Third-position Euclidean

Table 2
Distribution of Substitutions in *wingless* for 38 Butterfly Taxa

	All Sites	Nucleotide Position			Amino Acids
		First	Second	Third	
No. of characters	350	117	116	117	116
No. variable	214	70	50	117	91
No. informative	214	56	43	115	60

distances based on GC content were calculated as described by Ortí and Meyer (1996). Skewness ($-g1$) scores were calculated from 5,000 random trees generated in PAUP*, version 4.0d59. MacClade, version 3.06 (Maddison and Maddison 1996), was used to explore character evolution and patterns of nucleotide substitution along tree topologies. MEGA, version 1.01 (Kumar, Tamura, and Nei 1993) was used to determine proportions of synonymous and nonsynonymous differences. Codon preferences and measures of codon bias (ENC and scaled χ^2) were calculated using the program Molecular Evolutionary Analysis (MEA; generously provided by the author, Etsuko Moriyama, at Yale University).

Parsimony analyses consisted of heuristic searches with 10–50 random-addition replicates using tree bisection-reconnection (TBR) branch swapping. Analysis of amino acid characters employed a transformed Blosum 80 step matrix (see appendix) to weight substitutions among amino acid residues relative to their frequencies of change among observed protein sequences (Henikoff and Henikoff 1992). In parsimony analyses, nodal support was estimated using bootstrap analyses (100 replications, with 10 random-addition replicates each) and branch support (decay indices; Bremer 1988, 1994; Donoghue et al. 1992; Davis 1995). Distance analyses were carried out using the neighbor-joining algorithm (Saitou and Nei 1987); estimates of nodal support on distance trees were derived using bootstrap analyses (1,000 replications).

We compared the fit of 12 evolutionary models for this data set by estimating maximum-likelihood (ML) scores for the two trees obtained from the unweighted parsimony search. Likelihood scores for these trees were estimated under four models of evolution: the Jukes-Cantor (JC; Jukes and Cantor 1969), Kimura (1980) two-parameter (K2P), HKY85 (Hasegawa, Kishino, and Yano 1985), and general time reversible (GTR [=REV of Yang 1994]) models, using ML to estimate nucleotide frequencies for the HKY85 and GTR models. For each of these models, we evaluated likelihood scores under assumptions of (1) equal rates at all sites; (2) a proportion of sites estimated by ML being invariable (I); (3) rates at all sites evolving with rate heterogeneity as approximated by four discrete rate classes of the gamma distribution (Γ), estimated using ML; and (4) some sites being invariable, with variable sites evolving under a gamma distribution. Likelihood ratio tests were used to determine the model that best fit the data between pairs of nested models (Goldman 1993; Yang, Goldman, and Friday 1994; Felsenstein 1995).

Results and Discussion

Alignment and Substitution Patterns

Alignment of the 350-bp *wingless* fragment for all taxa required a single one-codon insertion in a pierid (*Delias*). Overall uncorrected sequence divergence percentages among taxa range from 10.3% to 41.6%, greater than those found by Brower and DeSalle (1998) for comparison of divergences in *wingless* and the mitochondrial *Cytochrome Oxidase I* (COI) gene between lepidopteran families (mean uncorrected pairwise difference between Pieridae vs. Nymphalidae reported as 22.8% for *wingless*, 16.6% for *COI*). The data set described here yielded 214 parsimony-informative nucleotides when sequences were compared across all taxa (not including outgroup taxa). Of these parsimony-informative nucleotide substitutions, 56 occurred at first-codon-position sites, 43 at second positions, and 115 at third positions (table 2). When reconstructed on the (two) most-parsimonious trees recovered from unweighted parsimony (trees discussed below), most second positions required smaller numbers of changes across the tree (up to 9 changes per character, averaging 1.3 changes per character) than did first and third positions (first positions made up to 12 changes per character, averaging 2.48; third positions made up to 19 changes per character, averaging 11). This is reflected in higher consistency index (CI) and retention index (RI) values when second positions alone were reconstructed on these trees than when first and third positions were reconstructed (table 3). The estimated proportion of nonsynonymous differences ($\pm$SD) ranged from 1.04 $\pm$ 0.73% to 38.08 $\pm$ 3.48% over all taxa. When translated, 79% of the 116 amino acid residues in this fragment are variable, showing up to eight character states, and 52% of amino acid sites were phylogenetically informative.

The ratio of transitions to transversions (ti/tv) estimated on the topologies recovered from an unweighted parsimony search was 1.3, reflecting a slightly faster overall accumulation of transitions than transversions

Table 3
Tree Statistics for Substitutions Reconstructed Onto the Two Trees Resulting from an Unweighted Parsimony Search (see fig. 3), Excluding Parsimony-Uninformative Characters

	All Sites	First Position	Second Position	Third Position
Consistency index	0.364	0.361	0.496	0.259
Retention index	0.665	0.660	0.786	0.522
Average no. of steps per character.....	4.94	2.48	1.30	11.00
Range	0–19	0–12	0–9	0–19

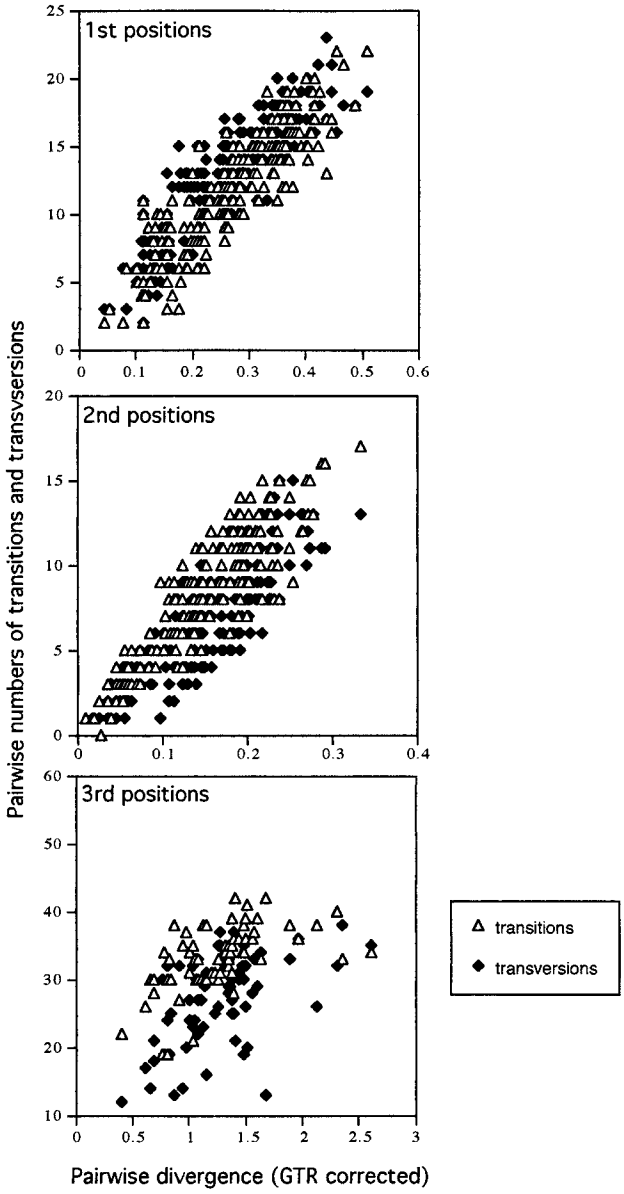

FIG. 1.—Abundances of transitions and transversions as a function of GTR-corrected sequence divergence for each pairwise comparison of taxa, broken down by codon position.

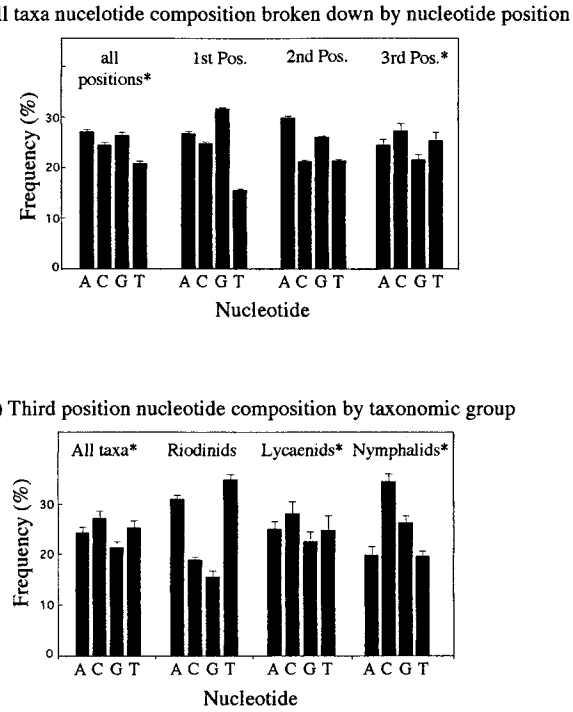

FIG. 2.—Nucleotide composition of (*a*) all nucleotides, broken down by codon position for all taxa, and (*b*) third-position nucleotides, broken down for riodinid, lycaenid, and nymphalid butterflies individually. Asterisks indicate statistically heterogeneous nucleotide composition ($P < 0.05$) in chi-square tests among taxa.

across all positions. To assess saturation effects in this data set, numbers of transitional and transversional pairwise changes were plotted against GTR-corrected pairwise distances. The GTR model was used, since it was found by likelihood ratio tests to be the optimal model for this data set (see below). For both first and second positions, numbers of transitions and transversions steadily accumulate as corrected pairwise divergences increase, indicating that saturation has not been reached (fig. 1). On the other hand, numbers of third-position transitions and transversions reach a plateau across the larger pairwise divergences, indicating possible accumulation of multiple hits in third positions. This plateau effect is more pronounced for transitions than for transversions, which is not surprising since at third codon positions only 3% of transitions cause amino acid replacements, compared with 41% for transversions (Crozier and Crozier 1993; Wakely 1996), and accumulate multiple hits more rapidly. In addition, third-position distances exceeded the maximum calculable value for the GTR correction method in 130 of the pairwise comparisons, further evidence that multiple hits occur at third positions in distant taxa. There were no undefined distances in GTR-corrected pairwise divergences of first or second positions.

Nucleotide and Amino Acid Composition

In the region of *wingless* examined, average nucleotide ratios were unbiased across all nucleotide positions, except for slightly lower proportions of thymidine (fig. 2*a*; see also Brower and DeSalle 1998). However, when codon positions were examined individually, nucleotide content showed more variation. Chi-square tests (implemented in PAUP*, version 4.0d59) indicate that *wingless* shows no significant nucleotide compositional variation among taxa in first or second positions ($\chi^2 = 34.49$, df = 111; $P = 1.00$). First-position nucleotides across all taxa show reduced T contents relative to A, C, and G contents (mean first-position T content = 15.7%; this possibly explains overall reduced thymidine levels). A similarly reduced T content in first positions was found by Friedlander et al. (1996) in the gene PEPCK in Lepidoptera, and by Ortí and Meyer (1996) in the *ependymin* gene in fish.

Mean base compositions among second-codon-position nucleotides show heightened A contents (mean =

Table 4
Likelihood Scores and Parameter Estimates for Various Models of Evolution, Incorporating Among-Site Rate Heterogeneity (Γ) and Invariant Sites (I), as Derived from the *wingless* Data and One of the Two Most-Parsimonious Topologies Shown in Figure 3

Model	−Log-Likelihood	Proporation of Invariable Sites (I)	Gamma Shape Parameter (α)	Transition/ Transversion Ratio	R-Matrix[a]					
JC	8,214.890									
K2P	7,931.090			1.729						
HKY85	7,927.719			1.752						
GTR	7,894.142				1.47,	3.05,	1.54,	0.76,	5.50,	1.00
JC+I	7,800.664	0.274								
K2P+I	7,502.863	0.274		1.847						
HKY85+I	7,493.452	0.274		1.893						
GTR+I	7,447.279	0.274			0.88,	2.45,	1.49,	0.43,	4.70,	1.00
JC+I+Γ	7,461.007	0.172	0.829							
K2P+I+Γ	7,124.189	0.191	0.784	2.330						
HKY85+I+Γ	7,113.122	0.189	0.774	2.348						
GTR+I+Γ	7,103.476	0.185	0.753		1.22,	3.86,	1.19,	0.69,	6.22,	1.00

[a] Relative proportion of substitution types A–C, A–G, A–T, C–G, C–T, G–T.

30.3%) and slightly lowered T and C contents (mean TC content = 42.0%). Ortí and Meyer (1996) describe a similar pattern in the *ependymin* gene, which, like *wingless*, codes for a secreted protein. This pattern may reflect compositional constraints imposed on second-position nucleotides. Specifically, hydrophobic amino acids (F, L, I, M, V, A, C; hydropathic indices as defined by Kyte and Doolittle [1982]) are never coded for by triplets with A in the second position, whereas hydrophilic amino acids do tend to have A in the second position. Thus, secreted proteins with a functional requirement for an overall hydrophilic nature (such as *ependymin*) have high A contents and low TC contents at second positions, whereas membrane-spanning proteins have the opposite condition (Naylor, Collins, and Brown 1995; Ortí and Meyer 1996). Similar requirements may be dictating high A contents in second-position nucleotides in *wingless*, which is a diffusible secreted glycoprotein with a high percentage of hydrophilic amino acids (Couso, Bishop, and Martinez Arias 1994; Perrimon 1996) (the butterflies examined here showed an average hydrophilic amino acid content of 68%).

Although base compositions across taxa are homogeneous at first and second positions, significant chi-square tests of third positions indicate among-taxa compositional heterogeneity ($\chi^2 = 433.33$; df = 111; $P < 0.001$). When third-position nucleotide content was broken down by family, we found riodinid base composition to be statistically homogeneous among all riodinid taxa examined ($\chi^2 = 25.30$; df = 33; $P = 0.829$), with all riodinids showing high AT contents (average AT = 54.7%; fig. 2*b*). On the other hand, significantly heterogeneous nucleotide compositions were found among lycaenid species ($\chi^2 = 134.17$; df = 33; $P < 0.001$), with four taxa having high AT (%AT > 67%) contents and two having high GC contents (%GC > 73%; average %AT = 48.6%). Nucleotide content among nymphalid taxa was found to be statistically homogeneous ($\chi^2 = 36.87$; df = 27; $P = 0.097$). All nymphalids examined have high GC contents (average AT = 42.7%; fig. 2*b*). Two different measures of codon bias, ENC and scaled χ^2, indicate similar (and low) average levels of codon bias in each of the three families, ranging from 0.180 to 0.208 (scaled χ^2) and from 55.553 to 56.802 (ENC; Shields et al. 1988; Wright 1990). However, codon preferences are notably different among the different families, and this may contribute to third-position nucleotide heterogeneity. The significance of nonstationary nucleotide heterogeneity and codon preferences among butterfly families for phylogenetic reconstruction is discussed below.

Maximum-Likelihood Analysis of Nucleotide Evolution

ML analyses were used to assess 16 models of sequence evolution using this data set and the two most-parsimonious trees resulting from the unweighted search. Both trees showed that the single parameter having the greatest effect on improving likelihood scores is among-sites rate heterogeneity (Γ). Table 4 shows scores and parameter estimates for the best (highest likelihood score) of these trees; both trees showed very similar likelihood scores and parameters. Likelihood ratio tests performed on nested models found the GTR+Γ+I model fit the data significantly better than all of the other models examined (all models shown in table 4 are nested within the GTR+Γ+I model, Swofford et al. 1996). Estimates of reversible rate parameters among substitution types (A-C, A-G, A-T, C-G, C-T, G-T) are distinct, with high rates between pyrimidines (A-G) and between purines (C-T; R-matrix, see table 4). A similar increase in rate parameters in these substitution classes was reported by Mason-Gamer, Weil, and Kellogg (1998) for *granule-bound starch synthase* in grasses and explains the improvement in likelihood scores with the use of the GTR model as compared with HKY85, which differs from GTR only in allowing different rates for all manners of transitions and transversions.

Phylogenetic Analyses

When all characters were weighted equally, maximum-parsimony analysis yielded two most-parsimoni-

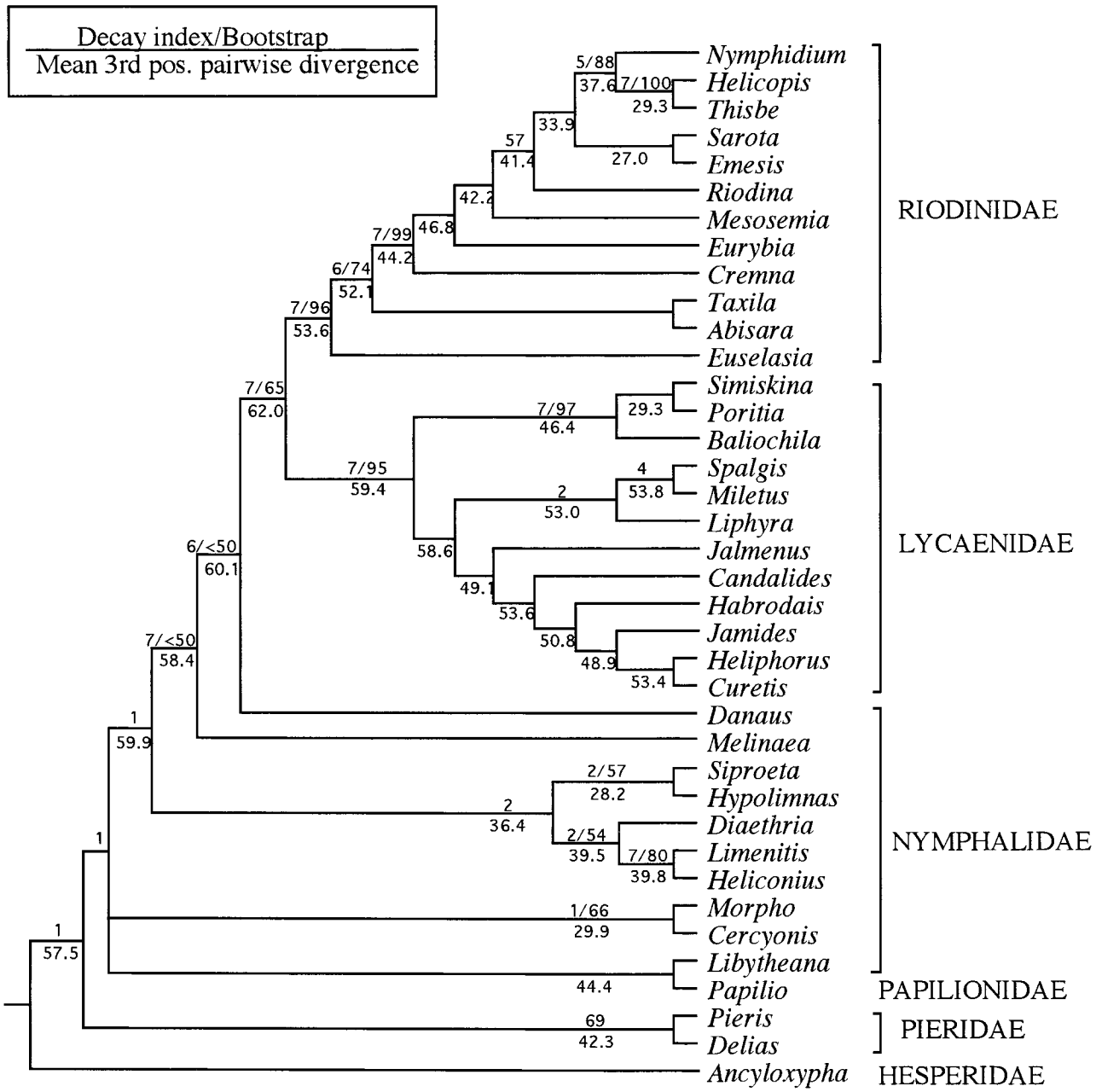

FIG. 3.—Strict consensus of two trees resulting from parsimony analysis of unweighted data set (length = 1,686; consistency index = 0.259). Statistics above branches are bootstrap values (≥50%)/decay indices; numbers below branches are mean third-position pairwise divergences.

ous trees (CI = 0.364; length = 1,686), differing only in the placement of the nymphalid clade *Cercyonis* + *Morpho* (consensus shown in fig. 3). Strong bootstrap and decay values support riodinid monophyly, lycaenid monophyly, and a sister relationship between the riodinids and the lycaenids. The sister of the riodinid + lycaenid lineage is less clear. Although all of the closest relatives to the riodinid + lycaenid clade are nymphalids, the nymphalids do not form a monophyletic group; rather, they are paraphyletic, both with respect to the riodinid + lycaenid lineage and with respect to the papilionids. However, basal relationships among the nymphalids are poorly resolved, and there is no bootstrap support ≥50% or decay index >1 supporting a paraphyletic Nymphalidae.

The above analyses of nucleotide evolution in this data set suggest that third positions and transitions are susceptible to multiple hits, which could confound phylogenetic analysis. Parsimony analyses using step matrices to downweight transitions with respect to transversions (ti : tv = 2:1, 5:1) recovered one to six most-parsimonious trees, in which consensus topologies were identical to unweighted analyses for the riodinid relationships. Although lycaenid monophyly remained highly supported, downweighting transitions had the effect of placing *Curetis* as the most basal lycaenid, with 67%

bootstrap support. This effect was also found when just third-position transitions were removed from analysis (see below). Although not supported with bootstrap values ≥50%, the nymphalid taxa *Diaethria, Limenitis, Heliconius, Siproeta,* and *Hypolimnas* formed a lineage with the same topology under all ti/tv weightings, which was consistent with the results of unweighted parsimony. Basal nymphalid relationships, however, were not stable among different ti/tv weightings.

A parsimony search using first and second positions alone yielded 111 trees (length = 389; CI = 0.414), and analysis of amino acids recovered 44 most-parsimonious trees (length = 5,119). Although some riodinid and lycaenid relationships were slightly less resolved than in the unweighted analysis, all nodes with bootstrap support >50% in the unweighted search (fig. 3) were also recovered with similar support in both analyses. In both weightings, strict consensus resulted in the collapse of all nymphalid nodes. Thus, at least for this sampling of representatives, *wingless* may not be a reliable marker for basal nymphalid relationships.

On the other hand, relationships within the Riodinidae and the Lycaenidae were remarkably well supported and stable except for two nodes which showed differences in patterns of support in response to the weighting of third-position transitions. One of these is the interpretation of the basal branching patterns of the lycaenids, especially the placement of *Curetis.* Downweighting or removing third-position transitions placed *Curetis* as the most basal branch of the lycaenids, with 66% bootstrap support when third position transitions were excluded (fig. 4) and 53% support in the analysis of amino acid characters.

An opposite trend is found in the node defining riodinid monophyly, which is well supported in the unweighted analysis (bootstrap value = 96%; fig. 3). As third-position transitions are downweighted with respect to other positions, support for this node decreases. For example, consensus of the four most-parsimonious trees recovered when third-position transitions were excluded from analysis shows 72% bootstrap support for riodinid monophyly (fig. 4). This suggests that signal from third-position transitions is consistent with support from first and second positions, and, when third-positions transitions are removed, bootstrap support for this node is reduced.

Third-codon-position variability above 20%–30% raw pairwise divergence has been considered to be at saturation level in analyses of other nuclear genes (Friedlander, Regier, and Mitter 1994). Such highly divergent third positions are often downweighted or removed from phylogenetic analyses because they contribute excessive noise and little to no signal (e.g., Friedlander et al. 1996; Fratí et al. 1997; but see Brower and DeSalle 1998). Although we observed divergence levels which, under these guidelines, implicate third-position saturation in *wingless* (27%–78%), a significant -g1 statistic (−0.250) calculated from third positions suggests that these substitutions are not all saturated, and at least some of them contain significant hierarchical signal at the $P < 0.01$ level (Hillis and Huelsenbeck 1992; but see Källersjö et al. 1992). Third positions also show increasing mean pairwise divergences at progressively deeper riodinid and lycaenid phylogenetic levels (fig. 3). Furthermore, a parsimony search on third positions alone resulted in one most-parsimonious tree (fig. 5) in which interpretations of riodinid relationships and monophyly were completely concordant with relationships recovered by first and second positions, with the exception of the placement of one riodinid, *Eurybia.* This search also recovered the sister relationship between the Riodinidae and the Lycaenidae. On the other hand, this tree did not recover lycaenid monophyly or basal lycaenid relationships as reconstructed by parsimony analyses employing first and second positions. Thus, while third positions appear to provide signal consistent for recovery of riodinid relationships and the riodinid + lycaenid node, third positions may contribute to contradictions or noise leading to reduced resolution and bootstrap support for lycaenids (such as that described above for *Curetis*).

Potential Mechanisms for Storing Signal in Third Positions at Deep Nodes

In euteleost fishes, Ortí and Meyer (1996) attribute third-position *ependymin* signal in deep nodes among orders and families to an accumulation of nonsynonymous substitutions (up to 35% of all third-position substitutions). This is in contrast to our findings: when we examined the characters supporting each node of the third-position most-parsimonious tree, we found that very few were nonsynonymous substitutions. The node supporting the riodinid lineage, for example, was supported by only 1 unambiguous third-position nonsynonymous substitution and 12 synonymous third-position substitutions. Likewise, the riodinid + lycaenid lineage is supported by only two nonsynonymous substitutions but nine synonymous substitutions. In addition, when a third-position-only search was conducted after removing these nonsynonymous characters, the same topology, with both of these nodes intact, was retrieved, indicating that these nonsynonymous changes were not sufficient to explain the signal underlying these nodes. Although *wingless* sequences show comparable (in fact, higher) levels of third-position divergence than do *ependymin* sequences, *wingless* shows less amino acid divergence than *ependymin* does. Ortí and Meyer (1996) estimated that ≥50% amino acid divergence is required before sufficient numbers of third-position nonsynonymous changes accumulate to provide reliable phylogenetic signal (their data show up to 67.5% pairwise divergence). Our data, which show a maximum of 47.7% amino acid divergence among taxa and very few nonsynonymous third positions, corroborate this estimate.

Another mechanism that might be involved in harboring phylogenetic signal in third positions is among-site rate heterogeneity. If third-position sites evolve at significantly heterogeneous rates, genetic distances between distantly related taxa include multiple hits at more quickly evolving sites which become saturated, as well as changes at more slowly evolving sites which retain

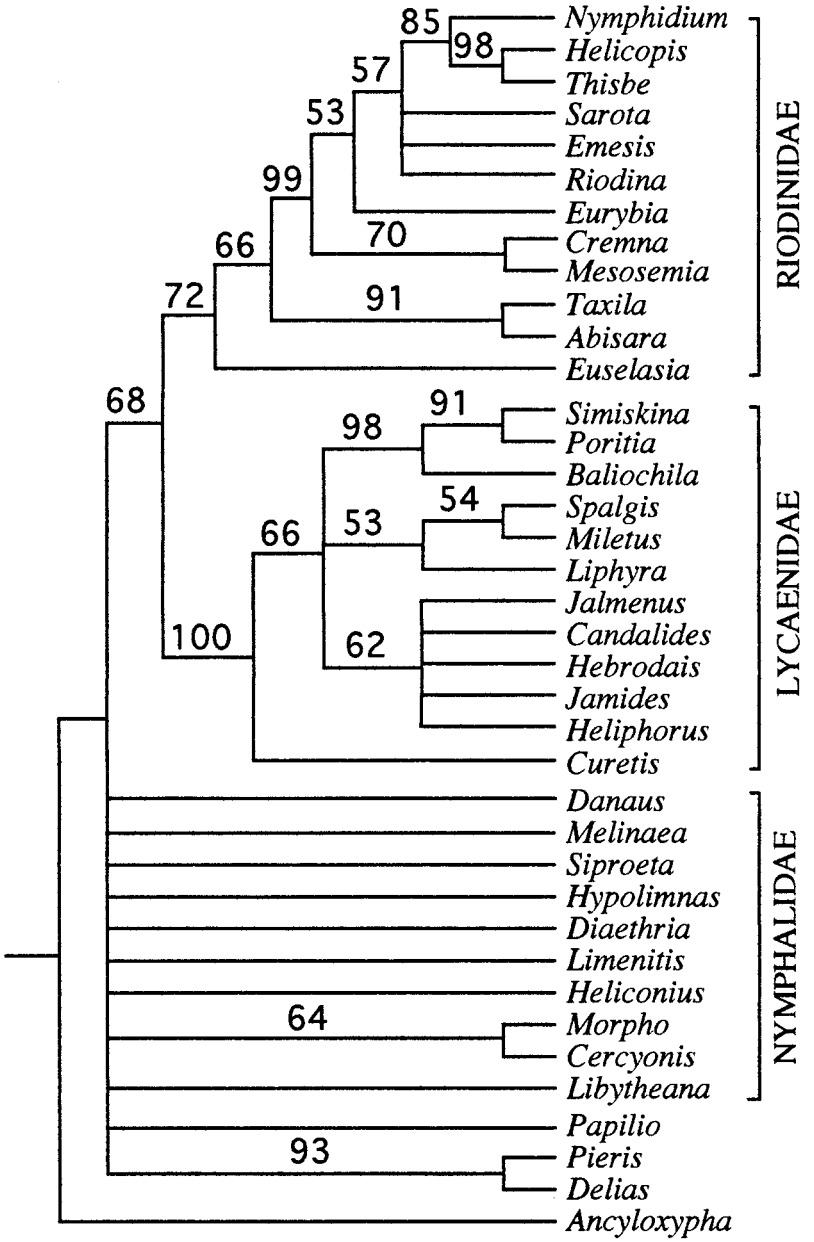

FIG. 4.—Bootstrap tree resulting from parsimony search excluding third-position transitions. A heuristic search found four most-parsimonious trees (length = 886; consistency index = 0.463). Bootstrap values are shown for nodes with support ≥50%; other nodes are collapsed.

phylogenetic signal. In other words, signal persists longer in the data if rate heterogeneity is a prominent feature, such that high third-position divergence creates useful variation in conservatively evolving sites while resulting in saturation of others (Yang 1998). Wide variation in number of steps per character when nucleotide substitutions were reconstructed onto the topology shown in figure 3 suggests that among-site rate variation exists in this data set (table 3). Likelihood ratio tests conducted using the unweighted tree topology and the GTR model of evolution to compare hypotheses of among-site rate variation in each nucleotide position individually further identify rate heterogeneity as a significant factor in third positions ($\chi^2 = 67.18$; $P < 0.005$; Γ shape parameter (α) = 3.785), as well as in first and second positions ($\chi^2 = 187.66$; 38.96, $\alpha = 0.443$ and 0.865 for first and second positions, respectively; table 5).

Methods of phylogenetic reconstruction based on an explicit model of evolution (i.e., likelihood and distance frameworks) offer sophisticated ways to compensate for rate heterogeneity; however, most available models do not yet accommodate nonstationary third-position base composition across taxa, which is a significant feature of this data set (fig. 2; discussed below). An alternative method allowing one to approach rate heterogeneity in parsimony is to reweight characters according to their consistencies (we used the rescaled consistency index) calculated with reference to a starting topology (Farris 1969; Carpenter 1988; Yang 1996; Sul-

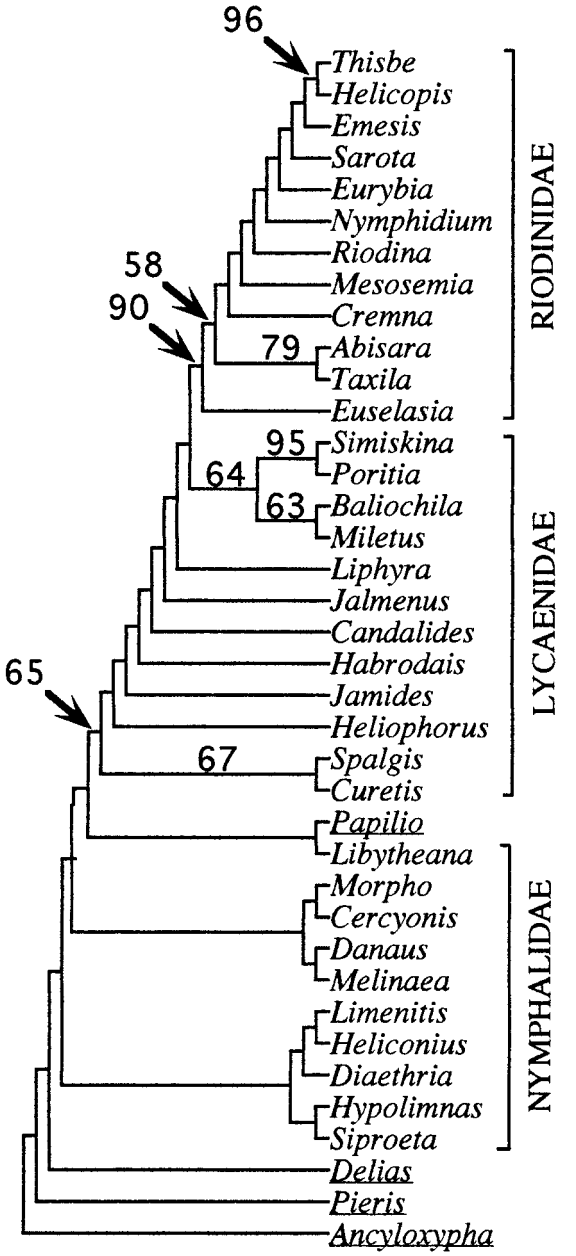

FIG. 5.—Single most-parsimonious tree resulting from analysis of third positions only (length = 1,264; consistency index = 0.218). The pierid, papilionid, and hesperiid outgroups are underlined. Numbers represent bootstrap values obtained from parsimony analysis of third characters reweighted with respect to the rescaled consistency index.

livan, Markert, and Kilpatrick 1997). In this way, we enhanced the signal from the most conservatively changing third-position characters based on the topology shown in figure 5. Analysis of reweighted third-position characters recovered a topology very similar to that generated by unweighted third positions, with increased bootstrap support for several nodes (see fig. 5), most notably (1) the node supporting riodinid monophyly (90% bootstrap support) and (2) the riodinid + lycaenid sister relationship (65% bootstrap support). This is consistent with the interpretation that in the riodinids the conservatively evolving third positions have retained useful phylogenetic information.

Table 5
Likelihood Ratio Test for Among-Site Rate Heterogeneity in Various Nucleotide Subsets of *wingless* under the General Time Reversible Model of Evolution

	Ln Likelihood[a]			
	Single Rate	Gamma	α[b]	χ^{2}[c]
All positions	−7,894.14	−7,113.35	0.495	1,561.58
First position	−1,573.30	−1,479.47	0.443	187.66
Second position...	−944.17	−924.69	0.865	38.96
Third position	−4,378.09	−4,344.5	3.785	67.18

[a] Calculated from one of the two topologies shown in figure 3.
[b] Gamma shape parameter.
[c] $\chi^2 = 2(\ln L_{gamma} - \ln L_{singlerate})$; df = 1; $P < 0.005$.

Nonstationary base composition across taxa is a third mechanism that may contribute to signal in third positions. At the third position, riodinid, lycaenid, and nymphalid lineages show significantly different nucleotide proportions, with high average AT content in riodinids, intermediate and heterogeneous in lycaenids, and low AT content in nymphalids (fig. 2*b*). Unequal nucleotide frequencies among taxa has often been cited as a cause of error in phylogenetic reconstruction because taxa with similar base compositions tend to cluster despite lack of shared ancestry (e.g., Saccone, Pesole, and Preparata 1989; Sidow and Wilson 1990; Lockhart et al. 1992, 1994; Hasegawa and Hashimoto 1993; Steel, Lockhart, and Penny 1993; unpublished data). However, if base compositional profiles evolved in a nonhomoplasious manner, they may contain phylogenetic information that reflects historical relationships.

To examine the potential contribution of third-position base composition heterogeneity to phylogenetic signal, trees were reconstructed using third-position distances based on the LogDet correction, which compensates for the effects of biased across-taxa nucleotide compositions in phylogenetic reconstruction (Lockhart et al. 1994). The resulting tree showed no resemblance to the third-position-only parsimony tree. Specifically, the LogDet tree did not recover riodinid, lycaenid, or riodinid + lycaenid monophyly; rather, taxa from all three lineages were grouped together and randomly distributed across the tree. The complementary experiment, in which neighbor-joining methods were used to reconstruct distances based only on third-position GC content (=''GC trees'' calculated from Euclidean distances as described by Ortí and Meyer 1996) recovered a topology far less random than the LogDet tree based on third positions. The GC tree recovered all of the riodinids in one clade (with four lycaenid interlopers which had high GC contents). The Lycaenidae, on the other hand, which show heterogeneous GC contents, were not recovered as monophyletic, and neither were the Nymphalidae. The results of these two analyses implicate the contribution of third-position nucleotide composition in the resolution of riodinid relationships.

Conclusions

Wingless is a single-copy nuclear gene that shows a level of variation appropriate for phylogenetic reconstruction of butterfly tribes and families. Despite this degree of variation, *wingless* is easy to PCR-amplify using the primers subscribing the region used here. Because this region is bounded on the 5′ end by an intron sequence, amplification of a longer segment has not been possible from genomic DNA to date. However, *wingless* sequences from other taxa (e.g., for *Bombyx*) are available from GenBank, and it may be possible to construct primers for PCR amplification of the exon 5′ of the region used here. Although the 5′ exon sequence is shorter than the fragment collected here (201 bp in *Drosophila*; Rijsewijk et al. 1987), the results from this study suggest that this second region of *wingless* may provide valuable characters for phylogenetic reconstruction.

Wingless has been cited as a useful source of phylogenetic characters for resolving relationships as deep as subfamily level in the Nymphalidae (Brower and DeSalle 1998). Our results here show that *wingless* provides an even greater degree of variation in the riodinids and lycaenids than in the nymphalids and resolves relationships in these two families at the tribal level to the among-families level with high nodal support.

In the riodinids, third-position substitutions continue to provide signal consistent with first- and second-position signal well past the 20%–30% divergence suggested as the saturation point by Friedlander, Regier, and Mitter (1994). In particular, third-position transitions harbor deep signal that supports riodinid monophyly, as well as resolving more recent riodinid relationships. Multiple factors probably contribute to this signal, including preservation of signal as similar nucleotide contents among related taxa and in more slowly evolving sites. However, third positions (especially transitions) also lend support for a paraphyletic Lycaenidae, an interpretation that conflicts with topologies derived from first and second positions. This contributes to unstable lycaenid relationships with decreased bootstrap values when third positions are included. The influence of third positions in defining basal nymphalid nodes appears to be minimal, although third positions do appear to contribute resolution to recent divergences in the nymphalids.

Different processes may be driving the evolution of *wingless* third positions in different butterfly families. Indications of this come in the form of different nucleotide compositions and codon preferences in different lineages. These phenomena have been documented in other data sets, both mitochondrial (e.g., Fratí et al. 1997) and nuclear (unpublished data), especially for higher-level phylogenetic problems. In cases such as these, close attention should be paid to any signals suggesting that third positions behave differently in the phylogenetic interpretation of different taxonomic groups. For this data set, excluding third-position transitions from the analysis (fig. 4) provides a conservative parsimony estimate for this overall analysis of relationships; however, this may exclude useful phylogenetic information, especially in the riodinids. New ML models are just beginning to accommodate nonstationary nucleotide and codon frequencies and may eventually help to solve these problems. To date, though, these models are few in number and rich in parameters (see Yang 1997; Galtier and Guoy 1998).

Acknowledgments

We are especially indebted to Phil DeVries for lending his vast expertise on riodinid biology, for his generous help in collecting and identifying butterflies, and for his improvements in the manuscript. We thank Kelvyn Dunn, Arthur Shapiro, Man-Wah Tan, and David Yeates for assistance in collecting and identifying specimens used in this study. André Mignault and Karen Nutt provided helpful laboratory assistance. Belinda Chang, Brian Farrell, Donald Harvey, Rodney Honeycutt, Jeff Jensen, Toby Kellogg, Carla Penz, Bob Reed, Chris Simon, and two anonymous reviewers provided helpful discussions and comments. Thanks to David Swofford for allowing us to publish results from test versions of PAUP*. This work was supported by a National Science Foundation Doctoral Dissertation Improvement Grant (DEB-9520824) and a National Institute of Health training grant to D.L.C. and by NSF DEB-9615760 to N.E.P. Grants from the Putnam Expedition Fund of the MCZ (to D.L.C. and N.E.P.), the Harvard Department of Organismic and Evolutionary Biology, and the Harvard Graduate Student Council, and support from the La Selva Lodge in Ecuador enabled us to collect many of these specimens. A.V.Z.B.'s contributions to this research were supported by NSF BSR-9220317 and DEB-9303251 and by a postdoctoral fellowship from the Smithsonian Institution and the American Museum of Natural History.

APPENDIX

Blosum80 Amino Acid Step Matrix Transformed[a] from the Log Odds Matrix of Henikoff and Henikoff (1992) for Use in Phylogenetic Reconstruction

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
[A] . .	—	22	23	19	25	16	25	20	17	19	20	22	21	20	22	10	15	16	33	26
[C] . .	22	—	37	35	31	34	39	24	33	25	28	32	37	32	34	24	25	24	39	34
[D] . .	23	37	—	14	32	25	26	31	22	30	31	15	28	21	25	19	22	29	42	33
[E] . .	19	35	14	—	30	25	20	27	14	26	25	19	24	11	19	17	20	23	36	29
[F] . .	25	31	32	30	—	31	26	19	28	16	19	31	34	29	29	25	26	21	26	13
[G] . .	16	34	25	25	31	—	29	30	23	29	28	20	31	26	26	18	23	28	37	32
[H] . .	25	39	26	20	26	29	—	31	22	28	29	19	32	19	21	23	26	29	36	17
[I] . . .	20	24	31	27	19	30	31	—	25	9	12	28	29	26	26	22	19	6	33	24
[K] . .	17	33	22	14	28	23	22	25	—	22	23	17	24	13	11	17	18	23	36	27
[L] . .	19	25	30	26	16	29	28	9	22	—	9	27	28	23	23	21	20	11	30	21
[M] .	20	28	31	25	19	28	29	12	23	9	—	26	29	20	24	22	19	14	31	26
[N] . .	22	32	15	19	31	20	19	28	17	27	26	—	29	18	20	14	17	26	39	28
[P] . .	21	37	28	24	34	31	32	29	24	28	29	29	—	27	27	23	26	27	42	35
[Q] . .	20	32	21	11	29	26	19	26	13	23	20	18	27	—	16	18	19	24	33	26
[R] . .	22	34	25	19	29	26	21	26	11	23	24	20	27	16	—	20	21	24	35	28
[S] . .	10	24	19	17	25	18	23	22	17	21	22	14	23	18	20	—	11	20	35	24
[T] . .	15	25	22	20	26	23	26	19	18	20	19	17	26	19	21	11	—	15	34	25
[V] . .	16	24	29	23	21	28	29	6	23	11	14	26	27	24	24	20	15	—	33	24
[W] .	33	39	42	36	26	37	36	33	36	30	31	39	42	33	35	35	34	33	—	21
[Y] . .	26	34	33	29	13	32	17	24	27	21	26	28	35	26	28	24	25	24	21	—

[a] Transformation consists of multiplying each log odds matrix value by −2 (so that values represent cost rather than score and are even). Then, for each value, the average diagonal element of the respective row and column is subtracted to make the diagonal zero while consistently adjusting all other values (Gary Olsen, personal communication).

LITERATURE CITED

BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution **42**:795–803.

———. 1994. Branch support and tree stability. Cladistics **1**: 295–304.

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

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**: 1–10.

BROWER, A. V. Z., and M. 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 Biol. Sci. **264**:969–977.

CAMPBELL, D. L., and N. E. PIERCE. 2000. Relationships of the Riodinidae (Lepidoptera) and phylogenetic placement with respect to other butterfly families: implications for the evolution of ant association? *In* C. BOGGS, W. WATT, and P. EHRLICH, eds. Ecology and evolution taking flight: butterflies as model study systems. University of Chicago Press, Chicago (in press).

CARPENTER, J. M. 1988. Choosing among equally parsimonious cladograms. Cladistics **4**:291–296.

CARROLL, S. B., J. GATES, D. N. KEYS, S. W. PADDOCK, G. E. F. PANGANIBAN, J. E. SELEGUE, and J. A. WILLIAMS. 1994. Pattern formation and eyespot determination in butterfly wings. Science **265**:109–114.

COUSO, J. P., S. A. BISHOP, and A. MARTINEZ ARIAS. 1994. The *wingless* signaling pathway and the patterning of the wing margin in *Drosophila*. Development **120**:621–636.

CROZIER, R. H., and Y. C. CROZIER. 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. Genetics **133**:97–117.

DAVIS, J. I. 1995. A phylogenetic structure for the monocotyledons, as inferred from chloroplast DNA restriction site variation, and a comparison of measures of clade support. Syst. Bot. **20**:503–527.

DE JONG, R., R. I. VANE-WRIGHT, and P. R. ACKERY. 1996. The higher classification of butterflies (Lepidoptera): problems and prospects. Entomol. Scand. **27**:65–101.

DEVRIES, P. J. 1991. Evolutionary and ecological patterns in myrmecophilous riodinid butterflies. Pp. 143–156 *in* C. R. HUXLEY and D. F. CUTLER, eds. Oxford surveys in evolutionary biology, Vol. 4. Ant-plant interactions. Oxford University Press, Oxford, England.

———. 1997. The butterflies of Costa Rica and their natural history, Vol. 2. Riodinidae. Princeton University Press, Princeton, N.J.

DONOGHUE, M. J., R. G. OLMSTEAD, J. F. SMITH, and J. D. PALMER. 1992. Phylogenetic relationships of dipsacales based on rbcL sequences. Ann. Mo. Bot. Gard. **79**:333–345.

EHRLICH, P. R. 1958. The comparative morphology, phylogeny and higher classification of the butterflies (Lepidoptera: Papilionoidea). Univ. Kans. Sci. Bull. **8**:305–370.

EHRLICH, P. R., and A. EHRLICH. 1967. The phenetic relationships of the butterflies I. Adult taxonomy and the nonspecificity hypothesis. Syst. Zool. **16**:301–317.

ELIOT, J. N. 1973. The higher classification of the Lycaenidae (Lepidoptera): a tentative arrangement. Bull. Br. Mus. Nat. Hist. **28**1–505.

FARRIS, J. S. 1969. A successive approximations approach to character weighting. Syst. Zool. **18**:374–385.

FELSENSTEIN, J. 1995. PHYLIP (phylogeny inference package). Version 3.57. Distributed by the author, Department of Genetics, University of Washington, Seattle.

FIEDLER, K. 1991. Systematic, evolutionary and ecological implications of myrmecophily within the Lycaenidae (Insecta:

Lepidoptera: Papilionoidea). Bonn. Zool. Monogr. **31**:1–210.

FRATÍ, F., C. Simon, J. Sullivan, and D. L. SWOFFORD. 1997. Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. J. Mol. Evol. **44**:145–158.

FRIEDLANDER, T. P., J. C. REGIER, and C. MITTER. 1994. Phylogenetic information content of five nuclear gene sequences in animals: initial assessment of character sets from concordance and divergence studies. Syst. Biol. **43**:511–525.

FRIEDLANDER, T. P., J. C. REGIER, C. MITTER, and D. L. WAGNER. 1996. A nuclear gene for higher level phylogenetics: phosphoenolpyruvate carboxykinase tracks Mesozoic-age divergences within Lepidoptera (Insecta). Mol. Biol. Evol. **13**:594–604.

GALTIER, N., and M. GOUY. 1988. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol. Biol. Evol. **15**:871–879.

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. **36**:345–361.

HARVEY, D. J. 1987. The higher classification of the Riodinidae. Ph.D. thesis, University of Texas at Austin.

———. 1991. Higher classification of the Nymphalidae. Pp. 255–273 *in* H. F. NIJHOUT, ed. The development and evolution of butterfly wing patterns. Smithsonian Institution Press, Washington, D.C.

HASEGAWA, M., and T. HASHIMOTO. 1993. Ribosomal RNA trees misleading? Nature **361**:23.

HASEGAWA, M., M. KISHINO, and T. YANO. 1985. Dating the human-ape split by a molecular clock of mitochondrial DNA. J. Mol. Evol. **19**:171–175.

HENIKOFF, S., and J. G. HENIKOFF. 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA **89**:10915–10919.

HILLIS, D. M., and J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. **83**:189–195.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

KÄLLERSJÖ, M., J. S. FARRIS, A. G. KLUGE, and C. BULT. 1992. Skewness and permutation. Cladistics **8**:275–287.

KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

KRISTENSEN, N. P. 1976. Remarks on the family-level phylogeny of butterflies (Insecta, Lepidoptera, Rhopalocera). Z. Zool. Syst. Evol. Forsch. **14**:25–33.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: molecular evolutionary genetics analysis. Version 1.01. Pennsylvania State University, University Park.

KYTE, J., and R. F. DOOLITTLE. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.

LOCKHART, P. J., C. J. HOWE, D. A. BRYANT, T. J. BEANLAND, and A. W. D. LARKUM. 1992. Substitutional bias confounds inference of cyanelle origins from sequence data. J. Mol. Evol. **34**:153–162.

LOCKHART, P. J., M. A. STEEL, M. D. HENDY, and D. PENNY. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. **11**:605–612.

MADDISON, W. P., and D. R. MADDISON. 1996. MacClade: analysis of phylogeny and character evolution. Version 3.06. Sinauer, Sunderland, Mass.

MANIATIS T. E., F. FRITSCH, and J. SAMBROOK. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y.

MARTIN, J. A., and D. P. PASHLEY. 1992. A molecular systematic analysis of butterfly family and some subfamily relationships (Lepidoptera: Papilionoidea). Ann. Entomol. Soc. Am. **85**:127–139.

MASON-GAMER, R. J., C. F. WEIL, and E. A. KELLOGG. 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. Mol. Biol. Evol. **15**:1658–1673.

NAYLOR, G. J. P., T. M. COLLINS, and W. M. BROWN. 1995. Hydrophobicity and phylogeny. Nature **373**:565–566.

ORTÍ, G., and A. MEYER. 1996. Molecular evolution of Ependymin and the phylogenetic resolution of early divergences among euteleost fishes. Mol. Biol. Evol. **13**:556–573.

PERRIMON, N. 1996. Serpentine proteins slither into the *wingless* and *hedgehog* fields. Cell **86**:513–516.

RIJSEWIJK, F., M. SCHUERMANN, E. WAGENAAR, P. PARREN, D. WEIGEL, and R. NUSSE. 1987. The Drosophila homolog of the mouse mammary oncogene *int*-1 is identical to the segment polarity gene *wingless.* Cell **50**:649–657.

ROBBINS, R. K. 1988*a*. Comparative morphology of the butterfly foreleg coxa and trochanter (Lepidoptera) and its systematic implications. Proc. Entomol. Soc. Wash. **90**:133–154.

———. 1988*b*. Male foretarsal variation in Lycaenidae and Riodinidae and the systematic placement of *Styx infernalis* (Lepidoptera). Proc. Entomol. Soc. Wash. **90**:356–368.

SACCONE, C., G. PESOLE, and G. PREPARATA. 1989. DNA microenvironments and the molecular clock. J. Mol. Evol. **29**: 407–411.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing trees. Mol. Biol. Evol. **4**: 406–425.

SCOBLE, M. J. 1986. The structure and affinities of the Hedyloidea: a new concept of the butterflies. Bull. Br. Mus. Nat. Hist. Entomol. Ser. **53**:251–286.

———. 1988. Hedylidae: a response to Weintraub and Miller. Cladistics **4**:93–96.

SCOTT, J. A., and D. M. WRIGHT. 1990. Butterfly phylogeny and fossils. Pp. 152–208 *in* O. KUDRNA, ed. Butterflies of Europe. Vol. 2. Aula-Verlag, Wiesbaden, Germany.

SHIELDS, D. C., P. M. SHARP, D. G. HIGGINS, and F. WRIGHT. 1988. ''Silent'' sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. **5**:704–716.

SIDOW, A. 1992. Diversification of the *Wnt* gene family on the ancestral lineage of vertebrates. Proc. Natl. Acad. Sci. **89**: 5098–5102.

STEEL, M. A., P. J. LOCKHART, and D. PENNY. 1993. Confidence in evolutionary trees from biological sequence data. Nature **364**:440–442.

SULLIVAN, J., J. A. MARKERT, and W. KILPATRICK. 1997. Phylogeography and molecular systematics of the *Peromyscus aztecus* species group (Rodentia: Muridae) inferred using parsimony and likelihood. Syst. Bio. **46(3)**:426–440.

SWOFFORD, D. L. 1998. PAUP*: phylogenetic analysis using parsimony and other methods. Test versions 4.0.0 d54, d59. Laboratory of Molecular Systematics, Smithsonian Institution, Washington, D.C.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic inference. Pp. 407–514 *in* D. HILLIS, C. MORITZ, and B. K. MABLE, eds. Molecular systematics. 2nd edition. Sinauer, Sunderland, Mass.

TAYLOR, P. L. 1991. Gene Jockey. Version 1.20. BIOSOFT, Cambridge, England.

WAKELEY, J. 1996. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. TREE **11**:158–163.

WELLER, S. J., D. P. PASHLEY, and J. A. MARTIN. 1996. Reassessment of butterfly family relationships using independent genes and morphology. Ann. Entomol. Soc. Am. **89**: 184–192.

WRIGHT, F. 1990. The 'effective number of codons' used in a gene. Gene **87**:23–29.

YANG, Z. 1993. Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. **10**:1396–1401.

———. 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. **39**:105–111.

———. 1996. Among-site rate variation and its impact on phylogenetic analyses. TREE **11**:367–372.

——— 1997. Phylogenetic analysis by maximum likelihood (PAML). Version 1.3. Department of Integrative Biology, University of California at Berkeley.

———. 1998. On the best evolutionary rate for phylogenetic analysis. Syst. Biol. **47**:125–133.

YANG, Z., N. GOLDMAN, and A. FRIDAY. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. Mol. Biol. Evol. **11**:316–324.

RODNEY HONEYCUTT, reviewing editor

Accepted January 6, 2000