PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2005) **272**, 1577–1586
doi:10.1098/rspb.2005.3124
Published online 11 July 2005

Synergistic effects of combining morphological and molecular data in resolving the phylogeny of butterflies and skippers

Niklas Wahlberg[1,*], Michael F. Braby[2,5], Andrew V. Z. Brower[3], Rienk de Jong[4], Ming-Min Lee[3], Sören Nylin[1], Naomi E. Pierce[5], Felix A. H. Sperling[6], Roger Vila[5], Andrew D. Warren[3] and Evgueni Zakharov[6]

[1]*Department of Zoology, Stockholm University, 106 91 Stockholm, Sweden*
[2]*School of Botany and Zoology, The Australian National University, Canberra, ACT 0200, Australia*
[3]*Department of Zoology, Oregon State University, Corvallis, OR 97331, USA*
[4]*National Museum of Natural History, Department of Entomology, PO Box 9517, 2300 RA Leiden, The Netherlands*
[5]*Museum of Comparative Zoology Laboratories, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA*
[6]*Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2E9*

Phylogenetic relationships among major clades of butterflies and skippers have long been controversial, with no general consensus even today. Such lack of resolution is a substantial impediment to using the otherwise well studied butterflies as a model group in biology. Here we report the results of a combined analysis of DNA sequences from three genes and a morphological data matrix for 57 taxa (3258 characters, 1290 parsimony informative) representing all major lineages from the three putative butterfly super-families (Hedyloidea, Hesperioidea and Papilionoidea), plus out-groups representing other ditrysian Lepidoptera families. Recently, the utility of morphological data as a source of phylogenetic evidence has been debated. We present the first well supported phylogenetic hypothesis for the butterflies and skippers based on a total-evidence analysis of both traditional morphological characters and new molecular characters from three gene regions (*COI*, *EF-1α* and *wingless*). All four data partitions show substantial hidden support for the deeper nodes, which emerges only in a combined analysis in which the addition of morphological data plays a crucial role. With the exception of Nymphalidae, the traditionally recognized families are found to be strongly supported monophyletic clades with the following relationships: (Hesperiidae + (Papilionidae + (Pieridae + (Nymphalidae + (Lycaenidae + Riodinidae))))). Nymphalidae is recovered as a monophyletic clade but this clade does not have strong support. Lycaenidae and Riodinidae are sister groups with strong support and we suggest that the latter be given family rank. The position of Pieridae as the sister taxon to nymphalids, lycaenids and riodinids is supported by morphology and the *EF-1α* data but conflicted by the *COI* and *wingless* data. Hedylidae are more likely to be related to butterflies and skippers than geometrid moths and appear to be the sister group to Papilionoidea + Hesperioidea.

Keywords: molecular systematics; total evidence; Insecta

1. INTRODUCTION

Butterflies are arguably the best loved group of invertebrates and have been a source of inspiration for generations of natural historians and scientists. As a result, their generic- and specific-level classification is reasonably stable and the majority of taxa have been named (Ackery *et al.* 1999). Also unique for a large group of invertebrates is the immense biological knowledge amassed for many species, allowing butterflies to be used as a model group of organisms for wide ranging studies in ecology, evolution, population genetics and developmental biology (Boggs *et al.* 2003). Despite this wealth of information, the phylogenetic relationships and higher classification of the major groups of butterflies have remained contentious and competing hypotheses lack strong empirical support (Ehrlich 1958; Scott 1985; de Jong *et al.* 1996; Vane-Wright 2003).

The classification of butterflies and skippers has been based almost exclusively on the morphology of adult specimens for close to 250 years (Ackery *et al.* 1999), despite the rampant homoplasy in these morphologically variable insects. The usefulness of characters from immature stages has long been acknowledged (Müller 1886) but they have only recently been automatically incorporated in a phylogenetic context (Harvey 1991; Penz & Peggie 2003; Freitas & Brown 2004). Study of these features remains severely hampered by the lack of detailed descriptions or preserved specimens of larvae and pupae from which characters may be discerned.

Traditionally, the butterflies and skippers have been placed in two super-families and five families: Hesperiidae (skippers) are usually placed in their own super-family Hesperioidea while all other butterflies (Papilionidae, Pieridae, Lycaenidae and Nymphalidae) are placed in Papilionoidea. There is little agreement on the rank and monophyly of various groups and relationships among and

* Author for correspondence (niklas.wahlberg@zoologi.su.se).

Received 17 March 2005
Accepted 11 April 2005

within the families are largely unresolved (Vane-Wright 2003). For example, Riodinidae are sometimes separated from Lycaenidae, and Nymphalidae have been divided into as many as nine families. Papilionoidea and Hesperioidea have traditionally been considered sister taxa but recently the moth-like family Hedylidae (represented by the single genus *Macrosoma*) has been suggested to be more closely related to Papilionoidea than Hesperioidea (Scoble 1986, 1992), though this placement has been questioned (Weintraub & Miller 1987). Within Papilionoidea, Pieridae are either the sister group to the Papilionidae (Ehrlich 1958; Scott 1985) or to the (Nymphalidae + (Lycaenidae + Riodinidae)) (Kristensen 1976; de Jong *et al.* 1996; Weller *et al.* 1996; Ackery *et al.* 1999). Lycaenidae + Riodinidae are usually, but not always, considered to be the sister group of Nymphalidae (Scott 1985; Robbins 1988; de Jong *et al.* 1996).

DNA sequences have rarely been used to explicitly assess family level relationships of butterflies and skippers for exemplars of all families (Martin & Pashley 1992; Weller *et al.* 1996). These studies were based on short sequences of the nuclear 28S ribosomal DNA and the partial sequences of the mitochondrial gene *ND1*. The resulting trees were rooted with skippers, thus the monophyly of Papilionoidea was not tested. Molecular data have been utilized in a number of studies within lower taxa of butterflies, such as genera, tribes and subfamilies (Sperling 2003); however, there has been little coordination of effort among the studies cited in Sperling (2003). Thus, the many sequences available on public databases such as GenBank cannot be simply collated and analysed together because they do not represent homologous gene regions (Caterino *et al.* 2000). To alleviate this difficulty, we have agreed to sequence the same three genes for all studied taxa in our respective laboratories. The benefits of such cooperation are self-evident. The three genes we have chosen have been employed with great success in a variety of prior butterfly studies (Brower & Egan 1997; Brower 2000; Campbell *et al.* 2000; Caterino *et al.* 2001; Monteiro & Pierce 2001; Wahlberg *et al.* 2003; Megens *et al.* 2004), partly due to the availability of polymerase chain reaction primers that work well with all butterflies studied to date.

Recently, the utility of traditional morphological characters for inferring phylogenetic relationships has been heavily criticized and a purely molecular approach has been advocated (Hebert *et al.* 2003; Scotland *et al.* 2003). Such a stand is not supported by the fundamental characteristics of the two kinds of data (Miller *et al.* 1997; Baker & Gatesy 2002) nor by the actual use of morphological data in combination with molecular data (Jenner 2004). Here we investigate the relationships of higher taxa of butterflies and skippers using both traditional morphological data and new molecular data. We also demonstrate the synergistic impact of morphological data on the results of the simultaneous analysis of the combined dataset.

2. MATERIAL AND METHODS

Selection of taxa for sequencing was based on a published morphological study (de Jong *et al.* 1996) and covered all the major lineages in each butterfly and skipper family (see appendix A). In a few cases the genus coded for morphological characters was not available for molecular work. In such cases, a closely related genus was sequenced instead (see appendix A). Sequences of *Cytochrome Oxidase subunit I* (*COI*, 1531 bp), *Elongation Factor-1α* (*EF-1α*, 1225 bp) and *wingless* (403 bp) were generated in the laboratories of Brower, Pierce, Sperling and Wahlberg according to local protocols (see Brower 2000; Caterino *et al.* 2001; Monteiro & Pierce 2001; Wahlberg *et al.* 2003). Species identifications, voucher codes and deposition sites and GenBank accession numbers are given in appendix A. The morphological data of de Jong *et al.* (1996) were revised, with several characters being recoded, to yield a matrix of 99 characters (see electronic appendix for details). These were mainly scored from adult butterflies, comprising 39 wing venation, 19 leg, 14 head, 21 thoracic and two abdominal characters. In addition, four characters were taken from immature stages.

Heuristic parsimony searches were performed with NONA 2.0 (Goloboff 1998) via Winclada (Nixon 2002; 1000 random additions of taxa with 10 trees kept during each replication) for each dataset separately, for the combined molecular dataset and for the entire combined dataset. All characters were given equal weight. Out-groups were not specified *a priori* in any of the analyses and trees were rooted with a pyralid.

Bayesian phylogenetic analyses (Huelsenbeck *et al.* 2001) were performed using the programme MrBayes 3.1 (Ronquist & Huelsenbeck 2003). The GTR + G + I model of substitution (chosen using the programme MrModeltest; Nylander 2002) was fitted to each molecular partition and a rate variable model was fitted to the morphological data. Parameters were estimated for each of the genes and the morphology simultaneously (four partitions). Six chains (one cold and five heated) were run for 10 000 000 generations with trees sampled every 1000 generations. To check when stationarity was reached, likelihood values were graphically inspected and the first 1000 sampled trees were discarded as 'burn-in'. Similar analyses were run without the morphological data.

Nodal support for the cladogram was assessed using Bremer support (BS) (Bremer 1988, 1994) values and bootstrap analysis. BS values and related indices (see below) were calculated with the aid of TreeRot (Sorensen 1999) and PAUP* (Swofford 2001). Bootstrap values were calculated with NONA 2.0 using 1000 pseudo-replicates with 10 random addition replicates per pseudo-replicate.

There are several indices related to BS that describe the interactions of datasets in a combined analysis framework at a node-by-node and character-by-character basis (Baker & DeSalle 1997; Gatesy *et al.* 1999). Each data partition contributes additively to the total BS in the combined analysis framework, giving values known as partitioned Bremer support (PBS; Baker & DeSalle 1997; Gatesy *et al.* 1999). A positive PBS value indicates that a given data partition supports a given node while a negative value indicates conflict from a given data partition.

Patterns of homoplasy differ in different datasets (Barrett *et al.* 1991; Chippindale & Wiens 1994; Olmstead & Sweere 1994) and analysing the different datasets together can bring forth the underlying phylogenetic signal common to all datasets (Baker & DeSalle 1997; Baker *et al.* 1998). BS values can allow us to compare the effects of combined and separate analyses through an index known as hidden Bremer support (HBS; Gatesy *et al.* 1999). HBS is the difference between the BS of a given node in the combined analysis and the sum of the BS of each partition for the same node in the most parsimonious tree for that partition. This index allows us to

Table 1. Summary of character variation in the data partitions used in this study.

data partition	characters	parsimony informative	shortest tree(s)	CI	RI
morphology	99	88	506	0.32	0.68
COI	1531	566	4972	0.19	0.25
EF-1α	1225	403	4179	0.17	0.32
wingless	403	233	2508	0.18	0.42
total	3258	1290	12459	0.21	0.34

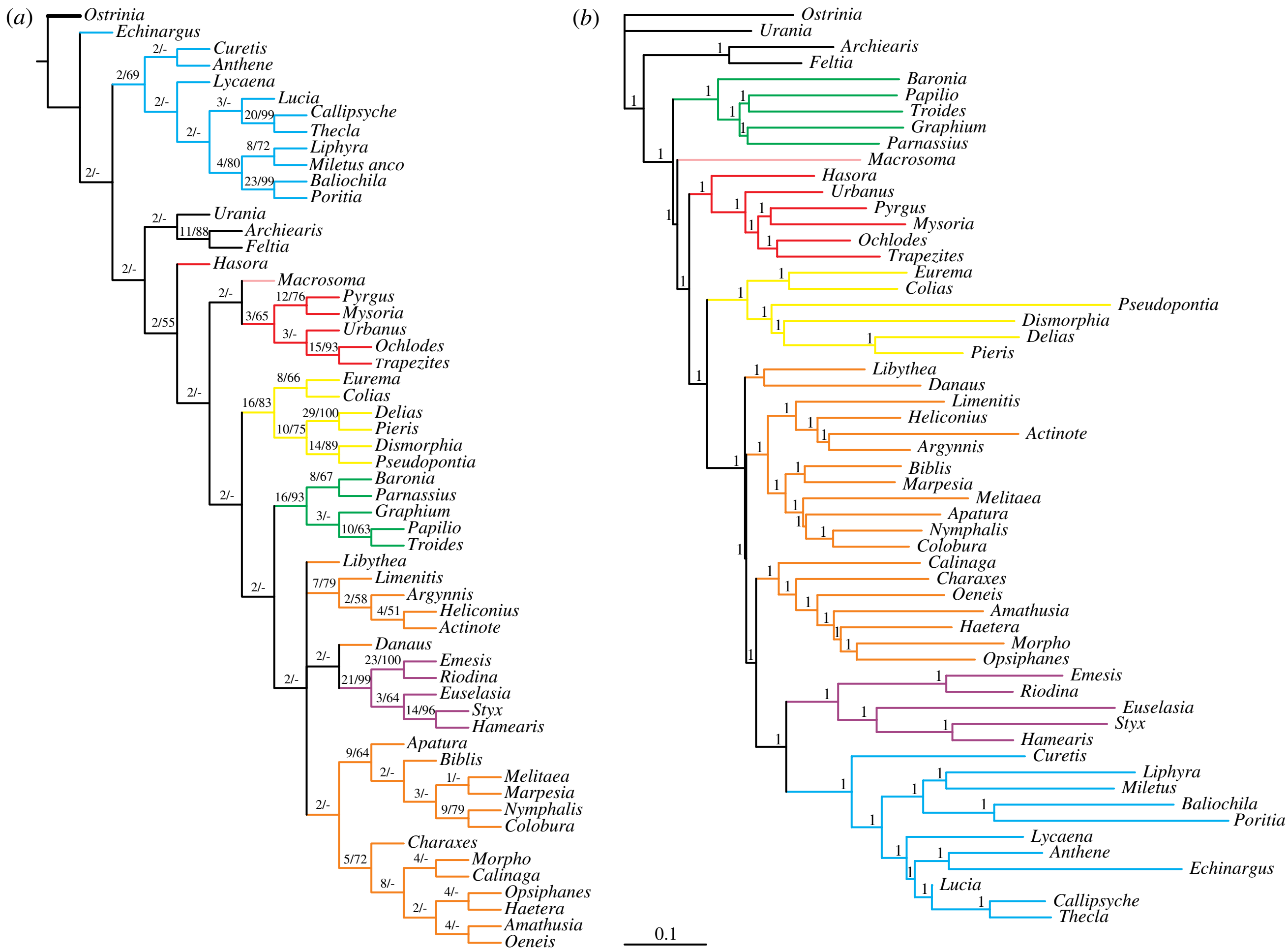

Figure 1. Phylogenetic analyses of the three molecular datasets combined. (*a*) Strict consensus of three equally maximum parsimonious trees (length 11 908, consistency index 0.18, retention index 0.32), numbers above branches are Bremer support, those to the right of the node are Bootstrap values; (*b*) tree resulting from Bayesian analysis (average likelihood = −49 408.8), numbers above or below branches are posterior probabilities for the node to the right of each number. Colour codes represent families as follows: pink, Hedylidae; red, Hesperiidae; green, Papilionidae; yellow, Pieridae; purple, Riodinidae; blue, Lycaenidae; orange, Nymphalidae.

evaluate whether there is increased character support for a clade in combined analysis of multiple datasets compared with the sum of support for that clade in the separate analyses of the different partitions. Positive values indicate increased character support in the combined analysis whereas negative values indicate increased character conflict in the combined analysis.

3. RESULTS

Our analysis is based on sequence data from two nuclear gene regions (1225 bp of *EF-1α* and 403 bp of *Wingless*) and one mitochondrial gene region (1531 bp of *COI*) for a total of 3159 base pairs and 99 morphological characters revised from a published study (de Jong *et al.* 1996) comprising a matrix of 57 taxa and 3258 characters (of which 1290 are parsimony informative; table 1). Analyses of the data partitions on their own results in many unconventional relationships (see electronic appendix). Our results for the morphology partition differ from those of de Jong *et al.* (1996), mainly due to our sparse out-group sampling (de Jong *et al.* sampled 15 out-group taxa).

Analyses based on the combined molecular data also fail to recover some of the traditionally recognized higher taxa as monophyletic groups (figure 1). The parsimony analysis finds many unconventional relationships (figure 1*a*). The low support for most nodes is probably due to large amounts of homoplasy and sparse sampling (there are about 20 000 recognized species of

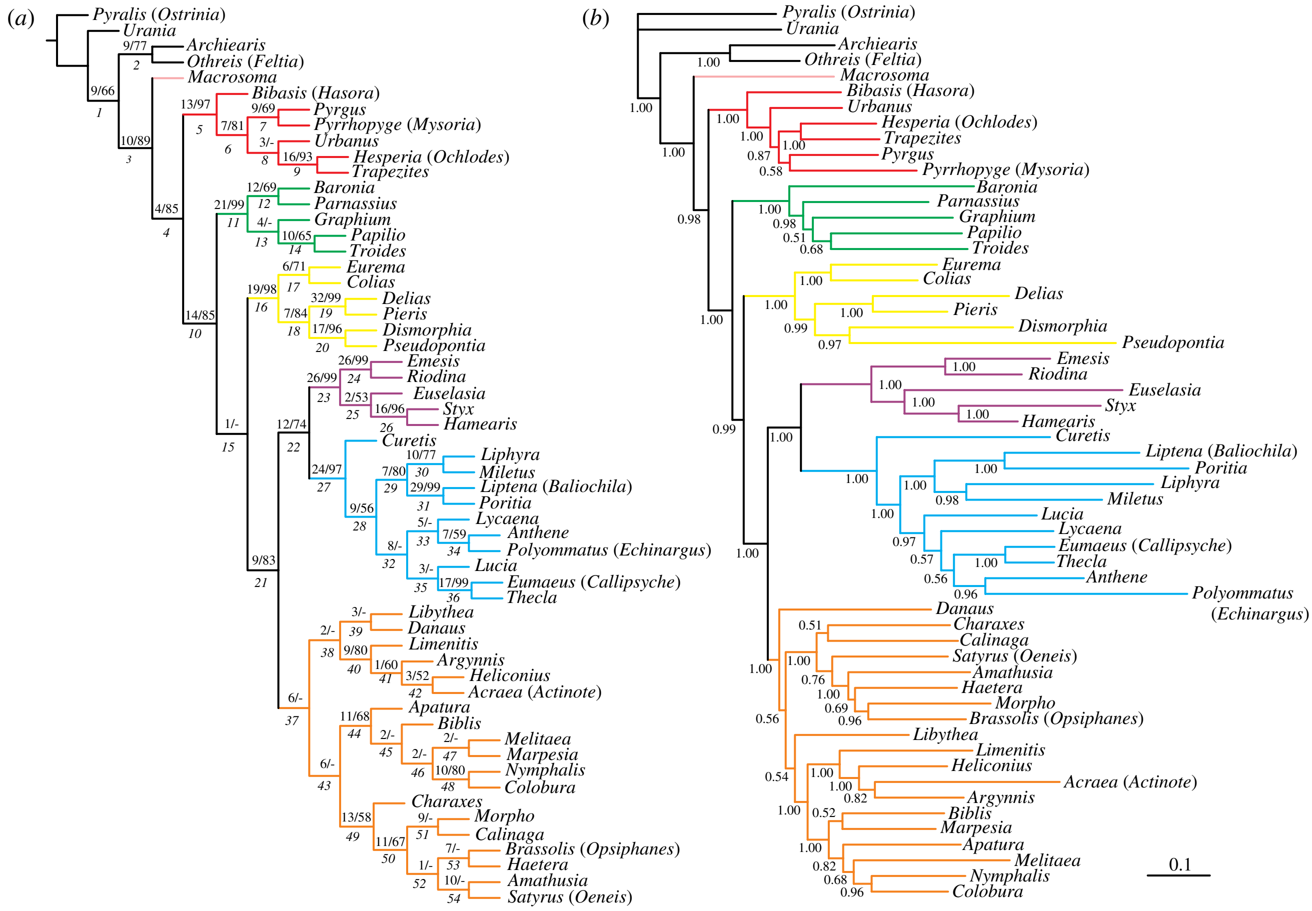

Figure 2. Phylogenetic analyses of the combined molecular and morphological datasets. (*a*) Single most parsimonious tree (length 12 459, consistency index 0.21, retention index 0.34). Numbers above branches are Bremer support/Bootstrap values for the node to the right of the numbers and italicized numbers below branches are node numbers referred to in table 1; (*b*) tree resulting from Bayesian analysis using mixed models (average likelihood = −51 999.2). Numbers below branches are posterior probabilities for the node to the right of each number. Taxa in parentheses are related substitutes from which sequence data were obtained. Colour codes as in figure 1.

butterflies; Robbins 1982). Attempting to take into account these problems by using a parameter rich model (each gene region having unique parameters; see electronic appendix) and Bayesian methodology to estimate the phylogenetic relationships results in a more conventional phylogeny, although there are still several highly unconventional clades (figure 1*b*). For instance, the Bayesian analysis places Papilionidae as sister to the rest of the butterflies, skippers and hedylids with a high posterior probability (100%) and also places the nymphalid satyrine clade (*sensu* Wahlberg *et al*. 2003) as sister to the Riodinidae + Lycaenidae clade with a posterior probability of 100%. Such relationships have never been suggested previously in the literature.

In contrast, the analyses of the combined molecular and morphology dataset provide strong support for the monophyly of all traditionally recognized higher taxa, except Nymphalidae which has moderate BS (6), high posterior probability (100%) but no bootstrap support (less than 50%; figure 2). Both analyses place Riodinidae as sister to Lycaenidae with a monophyletic Nymphalidae sister to the former two. Pieridae is placed as sister to (Nymphalidae + (Lycaenidae + Riodinidae)). Hedylidae is placed as sister to Hesperioidea + Papilionoidea. Support values for most of the nodes describing higher taxa are strong (BS $\geq$ 9, bootstrap $\geq$ 80%, posterior probability $\geq$ 95%; figure 2, table 2). The exceptions are the nodes describing Hesperioidea + Papilionoidea, Pieridae as sister to (Nymphalidae + (Lycaenidae + Riodinidae)) and monophyly of Nymphalidae which have BS values of 6 or less, although the same nodes have posterior probabilities higher than 95%.

PBS values show that in the combined analysis all the data partitions contribute positively to the support of the relationships in figure 2 (table 2). Since the magnitude of BS is contingent on the number of parsimony informative characters, dividing the total PBS for a partition by the number of parsimony informative characters of that partition gives us an index of the relative informativeness of each partition. The index clearly shows that the morphological (0.95) and *EF-1α* (0.56) partitions contribute most of the support for the relationships in figure 2*a*. These two partitions support almost all the major nodes and there is weak conflict derived from them at three nodes (major nodes shown in bold in table 2). The *COI* (0.29) and *wingless* (0.30) partitions show a higher level of conflict at many of the nodes yet both show strong support for the major nodes of interest.

Despite the strong support for most clades, we find that each data partition shows a great deal of homoplasy when mapped on to the tree found in combined parsimony analysis (table 1). However, combining the different data partitions reveals strong hidden support for almost all of the major nodes (table 2), indicating that combined

Table 2. Bremer support (BS), hidden Bremer support (HBS), partitioned Bremer support (PBS) and partitioned hidden Bremer support (PHBS) for nodes in the total evidence tree.

node	clade	BS total	HBS total	PBS morph	PBS *COI*	PBS *EF1α*	PBS *Wgl*	PHBS morph	PHBS *COI*	PHBS *EF1α*	PHBS *Wgl*	BS morph	BS *COI*	BS *EF1α*	BS *Wgl*
1		9	26	−6.0	−4.5	12.0	7.5	−3.0	3.5	18.0	7.5	−3	−8	−6	0
2		9	−13	−6.0	−4.5	12.0	7.5	−1.0	−2.5	13.0	−22.5	−5	−2	−1	30
3	**Hedy+Hesp+Papi**	**10**	**22**	**4.0**	**3.5**	**4.5**	**−2.0**	**5.0**	**−0.5**	**13.5**	**4.0**	**−1**	**4**	**−9**	**−6**
4	**Hesp+Papi**	**4**	**31**	**−0.8**	**−2.4**	**2.8**	**4.4**	**0.2**	**4.6**	**14.8**	**11.4**	**−1**	**−7**	**−12**	**−7**
5	**Hesperiidae**	**13**	**9**	**12.0**	**−0.8**	**2.5**	**−0.8**	**0.0**	**3.3**	**4.5**	**1.3**	**12**	**−4**	**−2**	**−2**
6		7	5	3.0	−2.0	1.0	5.0	1.0	−2.0	1.0	5.0	2	0	0	0
7		9	12	−1.0	2.0	3.0	5.0	3.0	−2.0	7.0	4.0	−4	4	−4	1
8		3	12	0.0	−1.0	0.0	4.0	2.0	4.0	2.0	4.0	−2	−5	−2	0
9		16	7	0.0	0.0	4.0	12.0	2.0	−2.0	4.0	3.0	−2	2	0	9
10	**Papilionoidea**	**14**	**40**	**9.0**	**4.5**	**−0.5**	**1.0**	**4.0**	**10.5**	**13.5**	**12.0**	**5**	**−6**	**−14**	**−11**
11	**Papilionidae**	**21**	**16**	**6.0**	**0.5**	**8.5**	**6.0**	**1.0**	**5.5**	**10.5**	**−1.0**	**5**	**−5**	**−2**	**7**
12		12	26	−1.0	6.0	8.0	−1.0	3.0	16.0	7.0	0.0	−4	−10	1	−1
13		4	13	2.0	6.0	−2.0	−2.0	1.0	11.0	−1.0	2.0	1	−5	−1	−4
14		10	14	1.3	6.0	1.7	1.0	1.3	9.0	1.7	2.0	0	−3	0	−1
15	**Pie+Nym+Lyc+Rio**	**1**	**35**	**2.0**	**−3.5**	**3.5**	**−1.0**	**4.0**	**−3.5**	**29.5**	**5.0**	**−2**	**0**	**−26**	**−6**
16	**Pieridae**	**19**	**32**	**2.5**	**−0.5**	**13.5**	**3.5**	**4.5**	**0.5**	**12.5**	**14.5**	**−2**	**−1**	**1**	**−11**
17		6	7	1.0	−1.0	4.0	2.0	−1.0	5.0	−2.0	5.0	2	−6	6	−3
18		7	22	−1.3	−2.7	7.7	3.3	−0.3	0.3	8.7	13.3	−1	−3	−1	−10
19		32	15	2.3	5.6	22.4	1.8	1.3	7.6	2.4	3.8	1	−2	20	−2
20		17	26	3.0	7.0	11.0	−4.0	4.0	10.0	10.0	2.0	−1	−3	1	−6
21	**Nym+Lyc+Rio**	**9**	**43**	**5.5**	**6.0**	**3.5**	**−6.0**	**3.5**	**12.0**	**22.5**	**5.0**	**2**	**−6**	**−19**	**−11**
22	**Lyc+Rio**	**12**	**41**	**8.0**	**17.0**	**−6.0**	**−7.0**	**9.0**	**17.0**	**14.0**	**1.0**	**−1**	**0**	**−20**	**−8**
23	**Riodinidae**	**26**	**1**	**4.0**	**6.0**	**10.0**	**6.0**	**3.0**	**2.0**	**1.0**	**−5.0**	**1**	**4**	**9**	**11**
24		26	1	3.5	17.5	−7.0	12.0	3.5	12.5	−12.0	−3.0	0	5	5	15
25		2	8	−1.0	2.0	6.0	−5.0	0.0	3.0	6.0	−1.0	−1	−1	0	−4
26		16	3	5.0	−12.0	20.0	3.0	4.0	−8.0	8.0	−1.0	1	−4	12	4
27	**Lycaenidae**	**24**	**23**	**5.0**	**4.0**	**6.0**	**9.0**	**8.0**	**4.0**	**20.0**	**−9.0**	**−3**	**0**	**−14**	**18**
28		9	45	3.5	10.0	1.0	−5.5	8.5	14.0	21.0	1.5	−5	−4	−20	−7
29		7	1	2.6	−1.4	3.4	2.4	6.6	−2.4	−0.6	−2.6	−4	1	4	5
30		10	6	3.0	5.0	1.0	1.0	7.0	1.0	1.0	−3.0	−4	4	0	4
31		29	9	3.0	5.0	8.0	13.0	5.0	2.0	2.0	0.0	−2	3	6	13
32		8	53	1.0	5.0	0.0	2.0	4.0	9.0	29.0	11.0	−3	−4	−29	−9
33		5	47	1.5	0.3	−3.0	6.3	2.5	5.3	26.0	13.3	−1	−5	−29	−7

(*Continued.*)

Table 2 (*Continued*.)

node	clade	BS total	HBS total	PBS morph	PBS *COI*	PBS *EF1α*	PBS *Wgl*	PHBS morph	PHBS *COI*	PHBS *EF1α*	PHBS *Wgl*	BS morph	BS *COI*	BS *EF1α*	BS *Wgl*
34		7	38	0.0	0.0	6.0	1.0	0.0	3.0	27.0	8.0	0	−3	−21	−7
35		3	15	2.3	−1.3	4.3	−2.3	5.3	0.7	8.3	0.7	−3	−2	−4	−3
36		19	4	1.3	0.7	18.3	−1.3	3.3	1.7	4.3	−5.3	−2	−1	14	4
37	**Nymphalidae**	**6**	**22**	**3.0**	**4.0**	**3.8**	**−4.8**	**−1.0**	**6.0**	**14.8**	**2.2**	**4**	**−2**	**−11**	**−7**
38		2	36	1.0	−6.0	10.0	−3.0	9.0	3.0	19.0	5.0	−8	−9	−9	−8
39		3	29	0.0	3.5	−1.0	0.5	4.0	9.5	6.0	9.5	−4	−6	−7	−9
40		9	16	1.5	−3.0	10.0	0.5	3.5	5.0	8.0	−0.5	−2	−8	2	1
41		1	9	1.0	−7.0	7.5	−0.5	2.0	3.0	5.5	−1.5	−1	−10	2	1
42		3	14	0.0	3.5	−1.0	0.5	1.0	6.5	7.0	−0.5	−1	−3	−8	1
43		6	39	0.0	14.5	−10.0	1.5	5.0	21.5	5.0	7.5	−5	−7	−15	−6
44		11	30	1.0	6.5	0.0	3.5	1.0	17.5	5.0	6.5	0	−11	−5	−3
45		2	29	0.0	7.5	−2.0	−3.5	2.0	14.5	9.0	3.5	−2	−7	−11	−7
46		2	20	0.0	7.5	−2.0	−3.5	3.0	7.5	9.0	0.5	−3	0	−11	−4
47		2	19	0.0	7.5	−2.0	−3.5	1.0	5.5	7.0	5.5	−1	2	−9	−9
48		10	10	2.0	2.5	7.0	−1.5	2.0	4.5	4.0	−0.5	0	−2	3	−1
49		13	23	0.0	14.8	−1.6	−0.2	0.0	14.8	5.4	2.8	0	0	−7	−3
50		11	15	−2.0	18.0	−6.0	1.0	−2.0	17.0	−3.0	3.0	0	1	−3	−2
51		9	13	3.0	−1.5	12.5	−5.0	2.0	−5.5	13.5	3.0	1	4	−1	−8
52		1	9	−2.0	1.0	−1.0	3.0	0.0	−1.0	8.0	2.0	−2	2	−9	1
53		7	14	−3.0	3.0	4.0	3.0	−2.0	4.0	3.0	9.0	−1	−1	1	−6
54		10	24	−2.0	9.0	2.0	1.0	−2.0	5.0	14.0	7.0	0	4	−12	−6
Total		543	1066	83.7	167.3	221.3	70.7	133.7	293.3	488.3	150.7	−50	−126	−267	−80

analysis enhances the inherent phylogenetic signal of the data compared with analysing the partitions separately. Partitioning the HBS values indicates that all four data partitions show substantial hidden signals that are revealed when analysed in combination with the other datasets.

4. DISCUSSION

In the discussion of their analysis of butterfly relationships based on morphology (the dataset employed here), de Jong *et al.* (1996) expressed a degree of dissatisfaction with the power of morphological features to resolve many clades. They questioned whether additional morphological data would improve the results and called for the pursuit of alternative sources of evidence. However, they qualified these controversial statements by arguing that the most productive avenue would be to combine those alternative data with morphological characters, as we have done here. Many current phylogenetic studies do not take morphological data into account for a variety of reasons (see Baker & Gatesy 2002; Scotland *et al.* 2003; Jenner 2004). As we have shown, ignoring the morphological data in this case resulted in several spurious clades, regardless of the method of analysis used (figure 1), including a paraphyletic Nymphalidae with regard to Lycaenidae+ Riodinidae, a sister relationship between Hedylidae and Papilionidae and a paraphyletic Papilionoidea with regard to Hesperioidea and Hedyloidea.

So we have an interesting problem: morphological data on their own provide weak support for, or fail to resolve, many nodes and molecular data on their own yield trees that contain implausible clades, some of which are strongly supported. A common claim from molecular systematists is that a large number of independent characters are needed to be able to estimate phylogenetic relationships robustly and reliably (Rokas *et al.* 2003). Yet for many groups of organisms, it is still not feasible to generate large amounts of DNA sequence data from a variety of gene regions. In Lepidoptera, for example, it has been a challenge to discover primers that are universal enough to amplify gene regions that are of broad phylogenetic utility in resolving all hierarchic taxonomic levels, from species to super-families (Friedlander *et al.* 1994). We feel that the propensity of limited molecular datasets to imply wrong, yet well supported, topologies is a problem that is not fully appreciated. Researchers now generally try to avoid the mistakes of early molecular systematics when dramatic rearrangements of phylogeny were often proposed on the basis of small and sparsely sampled datasets. However, there is still a tendency to reach very general conclusions from very specific and limited data. In particular, we feel that it should be more appreciated that adding more data is superior to adding ever more thorough analyses of existing data when the phylogeny is uncertain.

Our solution to this problem has been to combine our sequences with available morphological data. Morphological features are often uniquely derived complex structures that unequivocally unite major taxa, such as the osmeterial glands of papilionid larvae and the tricarinate antennae of nymphalid adults, and we see them as an intrinsic component of a robust treatment of this important group. Even though all characters are given equal weight and the morphological characters are vastly outnumbered by the molecular data, the low intrinsic homoplasy of the morphological features allows them to establish a structural framework in the combined analysis to which the morphological data contributes synergistically. Although the morphological data on their own were unable to give unequivocal answers in our analyses, the combined analyses (regardless of method used) have given us a robust phylogenetic hypothesis for the butterflies and skippers and the whole is greater than the sum of its parts, as is shown by the positive HBS values.

Our analyses of the combined dataset provide evidence to settle a long-standing question on the position and rank of Riodinidae. Our data give strong support for a Lycaenidae+Riodinidae sister group relationship, rather than embedding Riodinidae within Lycaenidae (de Jong *et al.* 1996) or placing them as sister to Nymphalidae (Robbins 1989). This result is concordant with an earlier molecular study that included a larger sample of taxa for the three groups but only used evidence from *wingless* (Campbell *et al.* 2000). The high BS and full congruence among the different datasets (table 2) indicates that only a great deal of contradictory new evidence could overturn this sister relationship. We believe that the current analysis provides a long-awaited empirical rationale for ranking Lycaenidae and Riodinidae as separate families.

The monophyly of Hesperiidae, Papilionidae and Pieridae is also strongly supported (figure 2) and there is little or no conflict between the data partitions at these nodes (table 2). These three families are generally considered to form monophyletic groups, though occasionally some small groups have been split off into their own families (e.g. Megathymidae from Hesperiidae and Baroniidae from Papilionidae). The case of Nymphalidae, however, is more complicated. Morphology, *COI* and *EF-1α* sequence data support the monophyly of the family while *wingless* conflicts moderately. The hidden BS of the partitions are positive, with the molecular partitions exhibiting larger values than the morphological partition. This suggests that there is a large amount of homoplasy (noise) in the molecular partitions for Nymphalidae that is overcome by the combined analysis of all data. The basal branch leading to Nymphalidae appears to be much shorter than the branches leading to the other families (figure 2*b*). One interpretation of such a pattern might be that the major lineages in Nymphalidae diverged very rapidly from one another. This may explain the difficulty in determining homologous states of morphological characters among nymphalids (Freitas & Brown 2004) and the difficulty in finding uniquely derived features to define the family (de Jong *et al.* 1996).

Relationships of taxa within families generally agree with recent family level studies (Brower 2000; Campbell *et al.* 2000; Caterino *et al.* 2001; Hall & Harvey 2002; Wahlberg *et al.* 2003; Freitas & Brown 2004), though some incongruent relationships can be noted. These incongruities are probably the result of insufficient taxon sampling in this study. The studies cited above included many more taxa per group and were generally able to resolve the relationships of major lineages within the families with higher support.

This is the first study to provide strongly supported answers to many of the questions about butterfly

Table 3. Information on specimens used for molecular studies.

family	subfamily	species	specimen ID	GenBank ID *COI*	GenBank ID *EF-1α*	GenBank ID *wingless*
Pyralidae		*Ostrinia nubilalis*	FS.b-5	AF170853	AF173392	—
Noctuidae		*Feltia jaculifera*	FS.b-150	U60990	AF173390	AY569039
Geometriidae		*Archiearis parthenias*	NW107-1	DQ018928	DQ018899	DQ018869
Uraniidae		*Urania leilia*	NW96-7	DQ018927	DQ018898	DQ018868
Hedylidae		*Macrosoma* sp.	FS.b-983	AF170854	AF173393	AY569042
Hesperiidae	Coeliadinae	*Hasora khoda*	AW97	DQ018930	DQ018901	DQ018871
Hesperiidae	Hesperiinae	*Ochlodes sylvanoides*	AW50	DQ018931	DQ018902	DQ018872
Hesperiidae	Trapezitinae	*Trapezites symmomus*	AW89	DQ018932	DQ018903	DQ018873
Hesperiidae	Pyrginae	*Urbanus dorantes*	AW280	DQ018929	DQ018900	DQ018870
Hesperiidae	Pyrginae	*Pyrgus communis*	FS.b-901	AF170857	AF173396	AY569043
Hesperiidae	Pyrrhopyginae	*Mysoria ambigua*	AW138	DQ018933	DQ018904	DQ018874
Papilionidae	Baroniinae	*Baronia brevicornis*	FS.a-167	AF170866	AF173406	AY569044
Papilionidae	Parnassiinae	*Parnassius phoebus*	FS.a-8	AF170872	AF173412	AY569045
Papilionidae	Papilioninae	*Graphium agamemnon*	FS.b-900	AF170874	AF173414	AY569046
Papilionidae	Papilioninae	*Papilion machaon*	FS.a-27	AF044006	AF044819	AY569124
Papilionidae	Papilioninae	*Troides helena*	FS.b-974	AF170878	AF173418	AY569047
Pieridae	Coliadinae	*Eurema hecabe*	MFB-00-P036	DQ018935	AY870587	DQ018876
Pieridae	Coliadinae	*Colias eurytheme*	FS.b-543	AF044024	AF173400	AY569040
Pieridae	Pierinae	*Delias belladonna*	DL-01-N104	DQ082773	AY870510	DQ082808
Pieridae	Pierinae	*Pieris napi*	FS.b-943	AF170861	AF173401	AY569041
Pieridae	Dismorphiinae	*Dismorphia zathoe*	MFB-00-P231	AY954566	AY870578	AY954596
Pieridae	Pseudopontiinae	*Pseudopontia paradoxa*	SC-01-T380	AY954564	AY870580	AY954594
Lycaenidae	Curetinae	*Curetis bulis*	MWT-93-A028	DQ018942	DQ018909	AF233549
Lycaenidae	Lipteninae	*Baliochila minima*	SP-93-P006	DQ018938	DQ018905	DQ018879
Lycaenidae	Miletinae	*Liphyra brassolis*	KD-94-T063	DQ018940	DQ018907	AF233551
Lycaenidae	Miletinae	*Miletus ancon*	KF-94-P022	DQ018941	DQ018908	AF233550
Lycaenidae	Poritiinae	*Poritia erycinoides*	MWT-93-B007	DQ018939	DQ018906	DQ018880
Lycaenidae	Theclinae	*Callipsyche behrii*	AS-92-Z034	DQ018943	DQ018910	DQ018881
Lycaenidae	Theclinae	*Thecla coelicolor*	DY-93-G038	DQ018945	DQ018912	DQ018883
Lycaenidae	Theclinae	*Lucia limbaria*	KD-94-Q002	DQ018944	DQ018911	DQ018882
Lycaenidae	Lycaeninae	*Lycaena helloides*	NP-99-W131	DQ018948	DQ018915	DQ018886
Lycaenidae	Polyommatinae	*Echinargus isola*	AS-92-Z185	DQ018947	DQ018914	DQ018885
Lycaenidae	Polyommatinae	*Anthene emolus*	MWT-93-A051	DQ018946	DQ018913	—
Lycaenidae	Polyommatinae	*Neurellipes staudingeri*	RD-98-U112	—	—	DQ018884
Riodinidae	Riodininae	*Riodina lysippus*	PDV-94-A050	DQ018952	DQ018919	AF233540
Riodinidae	Riodininae	*Emesis nr mandana*	PDV-94-T022	DQ018950	DQ018917	DQ018888
Riodinidae	Euselasiinae	*Euselasia nr orfita*	PDV-94-A022	DQ018951	DQ018918	DQ018889
Riodinidae	Styginae	*Styx infernalis*	GL-02-N259	DQ018949	DQ018916	DQ018887
Riodinidae	Hamearinae	*Hamearis lucina*	NW84-13	DQ018953	DQ018920	DQ018890
Nymphalidae	Libytheinae	*Libythea celtis*	NW71-1	AY090198	AY090164	AY090131
Nymphalidae	Danainae	*Danaus plexippus*	NW108-22	DQ018954	DQ018921	DQ018891
Nymphalidae	Calinaginae	*Calinaga buddha*	NW64-3	AY090208	AY090174	AY090141
Nymphalidae	Charaxinae	*Charaxes castor*	NW78-3	AY090219	AY090185	AY090152
Nymphalidae	Satyrinae	*Oeneis jutta*	EW4-1	DQ018958	DQ018925	DQ018896
Nymphalidae	Satyrinae	*Haetera piera*	CP01-84	DQ018959	DQ018926	DQ018897
Nymphalidae	Morphinae	*Amathusia phidippus*	NW114-17	DQ018956	DQ018923	DQ018894
Nymphalidae	Morphinae	*Morpho peleides*	NW66-5	AY090210	AY090176	AY090143
Nymphalidae	Morphinae	*Opsiphanes quiteria*	NW109-10	DQ018957	DQ018924	DQ018895
Nymphalidae	Limenitidinae	*Limenitis reducta*	NW67-2	AY090217	AY090183	AY090150
Nymphalidae	Heliconiinae	*Heliconius hecale*	NW70-6	AY090202	AY090168	AY090135
Nymphalidae	Heliconiinae	*Actinote stratonice*	NW90-14	AY218233	AY218252	DQ018892
Nymphalidae	Heliconiinae	*Argynnis paphia*	NW76-12	AY090200	AY090166	AY090133
Nymphalidae	Biblidinae	*Biblis hyperia*	NW106-3	DQ018955	DQ018922	DQ018893
Nymphalidae	Cyrestinae	*Marpesia orsilochus*	AB-RB250	AY788604	AY788706	AF246532
Nymphalidae	Apaturinae	*Apatura iris*	NW69-6	AY090199	AY090165	AY090132
Nymphalidae	Nymphalinae	*Melitaea cinxia*	NW73-14	AY788656	AY788776	AY788536
Nymphalidae	Nymphalinae	*Nymphalis antiopa*	NW70-2	AY218246	AY218266	AY218284
Nymphalidae	Nymphalinae	*Colobura dirce*	NW68-11	AY090228	AY090196	AY090162

relationships so eloquently raised by Vane-Wright and co-workers (de Jong *et al.* 1996; Vane-Wright 2003). Our data show that the six currently recognized skipper and butterfly families (Hesperiidae, Papilionidae, Pieridae, Lycaenidae, Riodinidae and Nymphalidae) are monophyletic entities, that Lycaenidae and Riodinidae are sister taxa, that Nymphalidae are sister to Lycaenidae + Riodinidae, that Pieridae are sister to (Nymphalidae +

(Lycaenidae + Riodinidae)) and lend support to the hypothesis that Hedylidae are more closely related to the butterflies and skippers than the geometrids within which they were once classified. Many of these taxa have been intuitively grouped based on striking morphological novelties since the time of Linnaeus but this study is the first to provide robust empirical support for relationships based on rigorous analysis of characters, both morphological and molecular. These results once again emphasize the importance of combining evidence from different sources as a means to dilute the bias of homoplasy within any individual data partition (Farris 1983; Miller *et al*. 1997).

APPENDIX A

See table 3.

This work has been supported in part by the Swedish Research Council (to S.N. and N.W.), USA NSF grants (to N.P. and A.B.), a Canadian NSERC grant (to F.S.), a Fulbright Foundation grant (to R.V.) and an Australian Research Council Fellowship and Fulbright Postdoctoral Fellow Award (to M.F.B.). We are very grateful to K. Brown Jr., A. Chichvarkhin, S. S. Collins, P. DeVries, A. Freitas, J. Hall, D. Janzen, D. Lohman, C. Stefanescu and K. Willmott for providing specimens used in this study. We thank Elisabet Weingartner and Carlos Peña for help in the laboratory.

REFERENCES

Ackery, P. R., de Jong, R. & Vane-Wright, R. I. 1999 The butterflies: Hedyloidea, Hesperoidea and Papilionoidea. In *Lepidoptera, moths and butterflies. 1. Evolution, systematics and biogeography. Handbook of zoology 4 (35), Lepidoptera* (ed. N. P. Kristensen), pp. 263–300. Berlin: de Gruyter.

Baker, R. H. & DeSalle, R. 1997 Multiple sources of character information and the phylogeny of Hawaiian drosophilids. *Syst. Biol.* **46**, 654–673.

Baker, R. & Gatesy, J. 2002 Is morphology still relevant?. In *Molecular systematics and evolution: theory and practice* (ed. R. DeSalle, W. Wheeler & G. Giribet), pp. 163–174. Basel: Birkhauser Verlag.

Baker, R. H., Yu, X. & DeSalle, R. 1998 Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. *Mol. Phyl. Evol.* **9**, 427–436.

Barrett, M., Donoghue, M. & Sober, E. 1991 Against consensus. *Syst. Zool.* **40**, 486–493.

Boggs, C. L., Watt, W. B. & Ehrlich, P. R. (eds) 2003 *Butterflies: evolution and ecology taking flight.* Rocky mountain biological lab symposium series. University of Chicago Press.

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

Bremer, K. 1994 Branch support and tree stability. *Cladistics* **10**, 295–304.

Brower, A. V. Z. 2000 Phylogenetic relationships among the Nymphalidae (Lepidoptera), inferred from partial sequences of the *wingless* gene. *Proc. R. Soc. B* **267**, 1201–1211. (doi:10.1098/rspb.2000.1129.)

Brower, A. V. Z. & Egan, M. G. 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. B* **264**, 969–977. (doi:10.1098/rspb.1997.0134.)

Campbell, D. L., Brower, A. V. Z. & Pierce, N. E. 2000 Molecular evolution of the *wingless* gene and its implications for the phylogenetic placement of the butterfly family Riodinidae (Lepidoptera: Papilionoidea). *Mol. Biol. Evol.* **17**, 684–696.

Caterino, M. S., Cho, S. & Sperling, F. A. H. 2000 The current state of insect molecular systematics: a thriving tower of babel. *Annu. Rev. Entomol.* **45**, 1–54.

Caterino, M. S., Reed, R. D., Kuo, M. M. & Sperling, F. A. H. 2001 A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera: Papilionidae). *Syst. Biol.* **50**, 106–127.

Chippindale, P. & Wiens, J. 1994 Weighting, partitioning, and combining characters in phylogenetic analysis. *Syst. Biol.* **43**, 278–287.

de Jong, R., Vane-Wright, R. I. & Ackery, P. R. 1996 The higher classification of butterflies (Lepidoptera): problems and prospects. *Entomol. Scand.* **27**, 65–101.

Ehrlich, P. R. 1958 The comparative morphology, phylogeny and higher classification of the butterflies (Lepidoptera: Papilionoidea). *The University of Kansas Science Bulletin* **39**, 305–370.

Farris, J. S. 1983 The logical basis of phylogenetic analysis. In *Advances in cladistics* (ed. N. I. Platnick & V. A. Funk), vol. 2, pp. 7–36. New York: Columbia University Press.

Freitas, A. V. L. & Brown, K. S. J. 2004 Phylogeny of the Nymphalidae (Lepidoptera: Papilionoidea). *Syst. Biol.* **53**, 363–383.

Friedlander, T. P., Regier, J. C. & Mitter, C. 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.

Gatesy, J., O'Grady, P. & Baker, R. H. 1999 Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. *Cladistics* **15**, 271–313.

Goloboff, P. A. 1998 *NONA*. New York: published by author.

Hall, J. P. W. & Harvey, D. J. 2002 A survey of androconial organs in the Riodinidae (Lepidoptera). *Zool. J. Linn. Soc.* **136**, 171–197.

Harvey, D. J. 1991 Higher classification of the Nymphalidae, appendix B. In *Higher classification of the Nymphalidae* (ed. H. F. Nijhout), pp. 255–273. Washington DC: Smithsonian Institution Press.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. 2003 Biological identifications through DNA barcodes. *Proc. R. Soc. B* **270**, 313–322. (doi:10.1098/rspb.2002.2218.)

Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. 2001 Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**, 2310–2314.

Jenner, R. A. 2004 Accepting partnership by submission? Morphological phylogenetics in a molecular millennium. *Syst. Biol.* **53**, 333–342.

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

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

Megens, H.-J., van Nes, W. J., van Moorsel, C. H. M., Pierce, N. E. & de Jong, R. 2004 Molecular phylogeny of the oriental butterfly genus *Arhopala* (Lycaenidae Theclinae) inferred from mitochondrial and nuclear genes. *Syst. Entomol.* **29**, 115–131.

Miller, J. S., Brower, A. V. Z. & DeSalle, R. 1997 Phylogeny of the neotropical moth tribe Josiini (Notodontidae: Dioptinae): comparing and combining evidence from DNA sequences and morphology. *Biol. J. Linn. Soc.* **60**, 297–316.

Monteiro, A. & Pierce, N. E. 2001 Phylogeny of *Bicyclus* (Lepidoptera: Nymphalidae) inferred from *COI*, *COII*, and *EF-1alpha* gene sequences. *Mol. Phyl. Evol.* **18**, 264–281.

Müller, W. 1886 Südamerikanische Nymphalidaeraupen: Versuch eines natürlichen Systems der Nymphaliden. *Zool. Jahrb.* **1**, 417–678.

Nixon, K. C. 2002 *Winclada*. Ithaca, NY: published by author.

Nylander, J.A.A. 2002 MrModeltest v2.1. Department of Systematic Zoology, Uppsala University: available from author.

Olmstead, R. & Sweere, J. 1994 Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Syst. Biol.* **43**, 467–481.

Penz, C. & Peggie, D. 2003 Phylogenetic relationships among Heliconiinae genera based on morphology (Lepidoptera: Nymphalidae). *Syst. Entomol.* **28**, 451–479.

Robbins, R. K. 1982 How many butterfly species? *News of the lepidopterists' society* **1982**, 40–41.

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

Robbins, R. K. 1989 Systematic implications of butterfly leg structures that clean antennae. *Psyche* **96**, 209–222.

Rokas, A., Williams, B. L., King, N. & Carroll, S. B. 2003 Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804.

Ronquist, F. & Huelsenbeck, J. P. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Scoble, M. J. 1986 The structure and affinities of the Hedyloidea: a new concept of butterflies. *Bull. Br. Museum Nat. Hist. (Entomol.)* **53**, 251–286.

Scoble, M. J. 1992 *The Lepidoptera: form, function and diversity.* Oxford University Press.

Scotland, R. W., Olmstead, R. & Bennett, J. R. 2003 Phylogeny reconstruction: the role of morphology. *Syst. Biol.* **52**, 539–548.

Scott, J. A. 1985 The phylogeny of butterflies (Papilionoidea and Hesperoidea). *J. Res. Lepid.* **23**, 241–281.

Sorensen, M. D. 1999 *TreeRot*. Boston University.

Sperling, F. A. H. 2003 Butterfly species and molecular phylogenies. In *Butterflies: evolution and ecology taking flight* (ed. C. L. Boggs, W. B. Watt & P. R. Ehrlich), pp. 431–458. University of Chicago Press.

Swofford, D. L. 2001 *PAUP*: phylogenetic analysis using parsimony (* and other methods) (version 4.0b10)*. Sunderland, MA: Sinauer Associates.

Vane-Wright, R. I. 2003 Evidence and identity in butterfly systematics. In *Butterflies: evolution and ecology taking flight* (ed. C. L. Boggs, W. B. Watt & P. R. Ehrlich), pp. 477–514. University of Chicago Press.

Wahlberg, N., Weingartner, E. & Nylin, S. 2003 Towards a better understanding of the higher systematics of Nymphalidae (Lepidoptera: Papilionoidea). *Mol. Phyl. Evol.* **28**, 473–484.

Weintraub, J. D. & Miller, J. S. 1987 The structure and affinities of the Hedyloidea: a new concept of butterflies. *Cladistics* **3**, 299–304.

Weller, S. J., Pashley, D. P. & Martin, J. A. 1996 Reassessment of butterfly family relationships using independent genes and morphology. *Ann. Entomol. Soc. Am.* **89**, 184–192.

The supplementary Electronic Appendix is available at doi://dx.doi.org/10.1098/rspb.2005.3124 or via http://www.journals.royalsoc.ac.uk.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.