Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cladistics 19 (2003) 213–223
Cladistics
www.elsevier.com/locate/yclad

Morphology versus molecules: resolution of the positions of *Nymphalis*, *Polygonia*, and related genera (Lepidoptera: Nymphalidae)

Niklas Wahlberg* and Sören Nylin

Department of Zoology, Stockholm University, S-106 91 Stockholm, Sweden

Accepted 21 March 2003

Abstract

The debate on whether to combine different data sets for simultaneous analysis has continued to the present day unabated. We have studied the effects of combining one morphological data set with four molecular data sets (two mitochondrial gene sequences and two nuclear gene sequences) for a group of butterflies belonging to the tribe Nymphalini using partitioned Bremer support. We particularly focus our attention on a group of species belonging to the genera *Aglais*, *Inachis*, *Roddia*, *Nymphalis*, *Kaniska*, and *Polygonia*. We find that, despite significant incongruence between most data partitions, all data partitions contribute positively to the support of most nodes in the most parsimonious trees found for the combined data set. We also find that the morphological data set resolves one particular node (*Kaniska* basal to *Polygonia*) with good support, while the molecular data sets are ambiguous about the existence of this node. We suggest that partitioned Bremer support allows one to critically appraise the robustness of each node in a given tree and thereby identify nodes that may change with the addition of new data and nodes that are likely to remain unchanged with new data. We also suggest that morphological data are still crucial to our being able to understand the relationships of extant organisms, despite published views to the contrary. Based on our results we suggest that *Inachis* should be synonymized with *Aglais*, *Roddia* with *Nymphalis*, and *Kaniska* with *Polygonia*.

A recent phylogenetic study of the tribe Nymphalini found that the Holarctic group of genera *Nymphalis*, *Polygonia*, *Aglais*, *Inachis*, *Kaniska*, and *Roddia* (termed the *Nymphalis* group in this paper) formed a strongly supported monophyletic clade (Nylin et al., 2001). This clade comprises 24 currently recognized species and has been instrumental in investigations of such phenomena as the evolution of host plant use (Janz and Nylin, 1997; Janz et al., 2001; Nylin, 1988), phenotypic plasticity in life-history traits (Nylin, 1992; Wiklund et al., 1992), effects of environment on distribution (Bryant et al., 2002), and insect physiology (Hiroyoshi and Mitsuhashi, 1999; Tanaka et al., 1997). The study by Nylin et al. (2001) was based on a morphological, ecological, and behavioral (MEB) data set comprising 97 characters and two DNA sequence data sets (695 bp of the mitochondrial gene ND1 and 379 bp of the nuclear gene *wingless*). That study showed convincingly that *Aglais* and *Inachis* form a monophyletic clade that is basal to the other genera of the *Nymphalis* group and that *Polygonia* is a strongly supported monophyletic clade. However, Nylin et al. (2001) were unable to ascertain the position of *Roddia l-album* (also known in the literature as *Nymphalis vaualbum*, a nomen nudum), the relationships of the species within *Polygonia* were unclear, and the position of *Kaniska* as sister to *Polygonia* depended solely on the MEB data set.

To clear up these ambiguities, we have generated sequence data from two additional genes; 1450 bp of the *cytochrome oxidase subunit I* (COI) mitochondrial gene and 1064 bp of the *elongation factor* 1α (EF-1α) nuclear gene for the 30 taxa included in the study by Nylin et al. (2001). In addition to the new sequences, almost all species were resequenced for *wingless*, producing a larger data set of 412 bp. The remaining two data sets (MEB and ND1) were taken from Nylin et al. (2001).

* Corresponding author. Fax: +46-8-167-715.
E-mail address: niklas.wahlberg@zoologi.su.se (N. Wahlberg).

0748-3007/03/$ - see front matter © 2003 The Willi Hennig Society. Published by Elsevier Science (USA). All rights reserved.
doi:10.1016/S0748-3007(03)00027-6

With the five separate data sets, we aim to investigate the contributions of each data set to the hierarchical patterns that we have discovered using partitioned Bremer support (PBS) (Baker and DeSalle, 1997; Baker et al., 1998). PBS has been widely used to investigate the effects of multiple data sets on the results of simultaneous analyses (DeSalle and Brower, 1997; Gatesy et al., 1999; Lambkin et al., 2002; O'Grady et al., 1998). The method takes advantage of the fact that each data partition in a combined data set contributes to the Bremer support (BS) value of a given node in a cladogram in an additive way, such that a positive value indicates that a given partition supports a given node and, conversely, a negative value indicates that a given partition does not support a given node (for details, see Baker and DeSalle (1997)). Note that a given partition can support some nodes in a cladogram and conflict with other nodes in the same cladogram.

Recently Lambkin et al. (2002) have shown that it is necessary to work with single trees rather than strict consensus trees to examine fully the effects of different data partitions on simultaneous analysis. This is because a given data partition may support a given node on one tree, while the same data partition may be in conflict with the same node on another equally parsimonious tree. Similarly, one should investigate the number of trees found when using anticonstraints (searching for the most parsimonious trees that do not contain a given node present in the most parsimonious tree(s) found in the combined analysis), as the different constrained trees can show different support for the same node and data partition (Lambkin et al., 2002).

PBS promises to be a very useful method for clearing up the controversy surrounding simultaneous analysis of data sets (e.g., de Queiroz et al., 1995; Kluge, 1989; Kluge and Wolf, 1993; Miller et al., 1997; Nixon and Carpenter, 1996), by allowing a researcher to identify both the particular nodes at which different data sets conflict and the nodes at which there is no conflict among data sets. PBS may also help settle the debate on whether morphological data sets have equal value to molecular data sets (e.g., Hedges and Maxson, 1996; Baker et al., 1998; Givnish and Sytsma, 1997; Miller et al., 1997), despite the more time-consuming data collection of the former. We aim to address both of these issues in our current study on the *Nymphalis* group. We are specifically interested in the relationships of the species currently placed in the genera *Nymphalis*, *Polygonia*, *Aglais*, *Inachis*, *Kaniska*, and *Roddia* and will treat the rest of the species included in our analyses as outgroups.

Material and methods

The species sampled are identical to those in the study by Nylin et al. (2001), except that we have excluded *Nymphalis cyanomelas* for which we had no material suitable for DNA extraction. The data set thus included almost all currently accepted species in the *Nymphalis* group (six species are missing: *Aglais ladakensis*, *A. kaschmirensis*, *Nymphalis cyanomelas*, *Polygonia giganteum*, *P. oreas*, and *P. haroldii*). In several cases DNA had to be extracted from different individuals because the extracts of Nylin et al. (2001) had become highly degraded or could not be found (Table 1). We extracted DNA mainly from one or two legs of freshly frozen or dried butterflies using QIAgen's DNEasy extraction kit. The spread voucher specimens can be viewed at http://www.zoologi.su.se/research/wahlberg/.

For each of the specimens we sequenced 1450 bp of COI, 1064 bp of the EF-1α, and 412 bp of *wingless*. Primers for COI were taken from Wahlberg and Zimmermann (2000), for EF-1α from Monteiro and Pierce (2001), and for *wingless* from Brower and DeSalle (1998). We performed all PCRs in a 20-μl reaction volume. The cycling profile for COI and *wingless* was 95 °C for 5 min, 35 cycles of 94 °C for 30 s, 47 °C for 30 s, 72 °C for 1 min 30 s, and a final extension period of 72 °C for 10 min. The cycling profile for EF-1α was 95 °C for 7 min, 35 cycles of 95 °C for 1 min, 55 °C for 1 min, 72 °C for 2 min, and a final extension period of 72 °C for 10 min. For all three genes, the PCR primers were also used for sequencing. In addition, we developed two internal primers for sequencing: EFmid (5′ CAA TAC CRC CRA TTT TGT 3′) for EF-1α and Patty (5′ ACW GTW GGW GGA TTA ACW GG 3′) for COI. Sequencing was done with a Beckman–Coulter CEQ2000 capillary sequencer. We checked the resulting chromatograms using the program BioEdit (Hall, 1999) and aligned the sequences by eye. The sequences have been submitted to GenBank (accession numbers in Table 1).

We tested the potential incongruence of the five data sets using the incongruence length difference (ILD) test of Farris et al. (1994), as implemented in the program Winclada (Nixon, 2002). We tested each pairwise combination using 1000 replicates of two random additions and tree bisection–reconnection (TBR) branch swapping.

We searched for the most parsimonious cladograms from the aligned data matrix consisting of 30 taxa using a heuristic search algorithm in the program NONA 2.0 (Goloboff, 1998). The molecular data were equally weighted and unordered with gaps treated as missing data. The MEB data were coded as in Nylin et al. (2001), although one character (character number 97 describing host plant family used with 17 states) had to be omitted due to limitations in the software used (Winclada can take only 10 character states per character, Nylin et al. coded their characters with the program PAUP*4b2). The heuristic searches were conducted with 10–1000 random-addition replicates using TBR branch swapping with up to 20 trees held during

Table 1
Collecting localities and GenBank accession numbers for sequences used in the combined analysis

Species	Locality	*wingless*	ND1	COI	EF-1α
Argynnis paphia	Stockholm, Sweden	AY090133	AF412758*	AY090200	AY090166
Junonia coenia	Tennessee, USA	AY248826	AF412756*	AY248777	AY248801
Hypolimnas bolina	Malaysia	AF412775	AF412751	AF187775	AY248802
Mynes geoffroyi	Australia	AF412760	AF412755	AY248778	AY248803
Symbrenthia hypatia	Malaysia	AF412784	AF412741	AY248779	AY248804
Araschnia levana	Estonia	AF412762	AF412730	AY248780	AY248805
Hypanartia lindigii	South America	AF412759	AF412757	AY248781	AY248806
Antanartia schaenia	Cameroon	AF412780	AF412746	AY218236	AY218255
Vanessa cardui	Missouri, USA	AF412770	AF412742	AY248782	AY248807
Vanessa virginiensis	Tennessee, USA	AY248827	AF412752*	AY248783	AY248808
Vanessa atalanta	Stockholm, Sweden	AF412772	AF412745	AY090221	AY090187
Vanessa gonerilla	New Zealand	AF412782	Missing	AY248784	AY248809
Inachis io	Stockholm, Sweden	AF412766	AF412737	AY248785	AY248810
Aglais urticae	Stockholm, Sweden	AF412777	AF412753	AY248786	AY248811
Aglais milberti	Washington, USA	AY248828	AF412738*	AY248787	AY248812
Nymphalis antiopa	Stockholm, Sweden	AY218284	AF412750*	AY218246	AY218266
Nymphalis polychloros	Öland, Sweden	AY248829	AF412743*	AY248788	AY248813
Nymphalis californica	Oregon, USA	AY248830	AF412740*	AY248789	AY248814
Nymphalis xanthomelas	Vladivostok, Russia	AY248831	AF412747*	AY248790	AY248815
Roddia l-album	British Columbia, Canada	AY248832	AF412739*	AY248791	AY248816
Kaniska canace	Japan	AY248833	AF412748*	AY248792	AY248817
Polygonia interrogationis	Tennessee, USA	AY248834	AF412731*	AY248793	AY248818
Polygonia comma	Tennessee, USA	AF412781	AF412732	AY248794	AY248819
Polygonia progne	New Hampshire, USA	AF412765	AF412736*	AY248795	AY248820
Polygonia satyrus	Oregon, USA	AY248835	AF412733*	AY248796	AY248821
Polygonia gracilis	Oregon, USA	AY248836	AF412735*	AY248797	AY248822
Polygonia faunus	Oregon, USA	AY248837	AF412734*	AY248798	AY248823
Polygonia c-album	Stockholm, Sweden	AY090154	AF412744*	AY090222	AY090188
Polygonia c-aureum	Japan	AF412786	AF412754	AY248799	AY248824
Polygonia egea	Greece	AY248838	AF412749*	AY248800	AY248825

ND1 sequences marked with an asterisk (*) are from an individual different from the other three gene sequences for a given species.

each step. The data sets were analyzed separately and combined.

We evaluated the robustness of the clades in the resulting cladograms with Bremer support values (Bremer, 1988, 1994). We used the program TreeRot (Sorensen, 1999) in conjunction with PAUP* 4.0b10 (Swofford, 1998) to calculate BS values. We assessed the contribution of each data partition to the BS values of the combined analyses using partitioned Bremer support (Baker and DeSalle, 1997; Baker et al., 1998) using the program TreeRot (Sorensen, 1999). In accordance with the recommendations of Lambkin et al. (2002), we calculated PBS values on each equally parsimonious tree separately and investigated the number and nature of suboptimal trees found when anticonstraints were enforced.

We have chosen to evaluate the robustness of the clades based only on BS values as we feel that they are more appropriate than the other commonly used indices of support such as bootstrap and jackknife values. BS values are free to vary between 0 and infinity, whereas the latter two values are constrained between 0 and 100. In many cases, nodes with BS values of, for instance, 10 and 100 will have bootstrap/jackknife values of 100. In such cases, the extra information given by the BS values is lost with the "statistical" support values. When discussing our results, we will refer to the BS values as giving weak, moderate, good, or strong support, even though delimiting such qualitative classes is an inherently subjective exercise. For this study we define weak support as BS values of 1–2, moderate support as BS values of 3–5, good support as BS values of 6–10, and strong support as BS values of >10. All BS values $\geqslant 7$ correspond to bootstrap values of $\geqslant 95\%$ in the current data set.

To investigate the implications of the new phylogenetic data for reconstructions of host plant utilization in *Nymphalis*, *Polygonia*, and related genera (Janz et al., 2001), we used the program MacClade (Maddison and Maddison, 1992). Use of the important host plant families Urticaceae, Ulmaceae, Cannabidaceae, Salicaceae, Grossulariaceae (*Ribes*), Betulaceae, and Ericaceae was traced onto the formerly proposed phylogeny (Nylin et al., 2001) and onto the best supported tree emerging from the present study. Attention was given only to the *Nymphalis* group clade (see Results), and host plant use was coded as discrete characters for each plant family (used/not used) with equal probabilities for gains and losses. This procedure may not result in the most accurate reconstruction of specific events, but we deemed it sufficient for a first comparison between phylogenies.

Results

General

The total combined data set consisted of 3621 nucleotides and 96 morphological, ecological, and behavioral characters. The basic statistics of the molecular data sets are presented in Table 2. The AT bias of the two mitochondrial genes (COI and ND1) is typical of insects. Three species showed an inferred codon deletion in the aligned *wingless* data set (*Inachis io*, *Aglais urticae*, and *A. milberti*) as discovered by Nylin et al. (2001). The number of parsimony-informative characters was about equal for each data set, with the exception of the COI, which had about two to three times more informative characters than the other data sets.

When analyzed separately, three of the five data sets (MEB, COI, and EF-1α) recovered the *Nymphalis* group as a monophyletic unit (Figs. 1A–C). The ND1 data set recovered *Nymphalis* + *Polygonia*, *Roddia* + *Kaniska*, and *Aglais* as monophyletic units, but these clades were interspersed with the outgroup genera (Fig. 1D). The *wingless* data set included *Antanartia schaenia* within the *Nymphalis* group (Fig. 1E).

The MEB data set showed significant incongruence with all of the molecular data sets according to the ILD test (Table 3). In the molecular data sets, there was significant incongruence between the EF-1α data set and the ND1 and COI data sets. *wingless* was not significantly incongruent with any of the molecular data sets, and neither were the mitochondrial genes with each other.

Parsimony analysis of the combined data set yielded two equally parsimonious trees (Fig. 2). These two trees differed in the relative positions of two outgroup clades, the *Vanessa* and *Symbrenthia* clades. In Tree 1, the *Vanessa* clade is the sister group to the *Nymphalis* group (Fig. 2A) and in Tree 2 it is the *Symbrenthia* clade that is sister to the *Nymphalis* group (Fig. 2B). The sister relationship of the *Vanessa* clade to the *Nymphalis* group is supported by two synapomorphies from the MEB data set (character numbers 85 and 96): all species that hibernate in this clade hibernate as adults (a unique trait among nymphalid butterflies) and most species lack microspines on larval setae (the ancestral state of microspines on larval setae is present in *I. io*, *Nymphalis xanthomelas*, and *N. antiopa*). The sister relationship of the *Symbrenthia* clade to the *Nymphalis* group is supported by two homoplastic characters from the MEB data set (character numbers 26 and 94). The first character describes the presence and shape of the discal spot, with the state rounded or occasionally elongate, not angled, present in *Araschnia*, *Antanartia*, *Aglais*, *Inachis*, *Nymphalis*, and (independently) *Vanessa gonerilla*. *Roddia*, *Kaniska*, and *Polygonia* have a further derived state of the discal spot, elongate and angled. *Symbrenthia* and *Mynes* show the ancestral state of no discal spot. In the second character, the pupal anterior projections project straight forward and to the sides, with the innersides forming a straight line in *Araschnia*, *Aglais*, *Inachis*, *Roddia*, most *Polygonia* species, *Nymphalis*, and the outgroup species *Argynnis paphia*; *Antanartia*, *Kaniska*, *Polygonia faunus*, *P. c-album*, and *P. gracilis* have the projections bending inward with the innersides forming a curve; *Mynes* has the ancestral state of only very slight projections (the state in *Symbrenthia* is unknown). We feel that the character support for the *Vanessa* clade being sister to the *Nymphalis* group is stronger and thus prefer Tree 1 in further discussion of the results.

The *Nymphalis* group forms a very strongly supported clade (BS of 17), within which there are three main clades. The most basal clade contains the genera *Inachis* and *Aglais*, and it is a strongly supported clade. The remaining two clades form a monophyletic unit with strong support. One of the clades contains species belonging to *Nymphalis* with *Roddia l-album* as the most basal species. This clade has very weak support. The second clade contains all species of *Polygonia* (a very strongly supported monophyletic taxon), with *Kaniska canace* as the most basal species. The second clade has moderate support from the combined data set (Fig. 2).

Within the three main clades there are several strongly supported groupings: the two species of *Aglais* have very strong support for their sister group relationship, as do *Polygonia faunus* and *P. c-album*. Within the *Polygonia* clade, the grouping of *P. progne*, *P. satyrus*, and *P. gracilis* receives good support, as does the grouping of *P. interrogationis* and *P. comma* with these three. The basal position of *P. c-aureum* with respect to the other species of *Polygonia* also receives good support.

Table 2
Basic statistics for the four molecular data sets in 30 species of Nymphalidae

Gene	No. of sites	No. variable	No. informative	Empirical base frequencies (%)			
				A	G	C	T
COI	1450	517	347	32.2	13.4	14.5	39.9
EF-1α	1064	261	155	26.1	23.3	27.1	23.5
ND1	695	287	157	30.8	14.9	7.7	46.6
Wingless	412	163	103	25.5	27.1	26.3	21.1

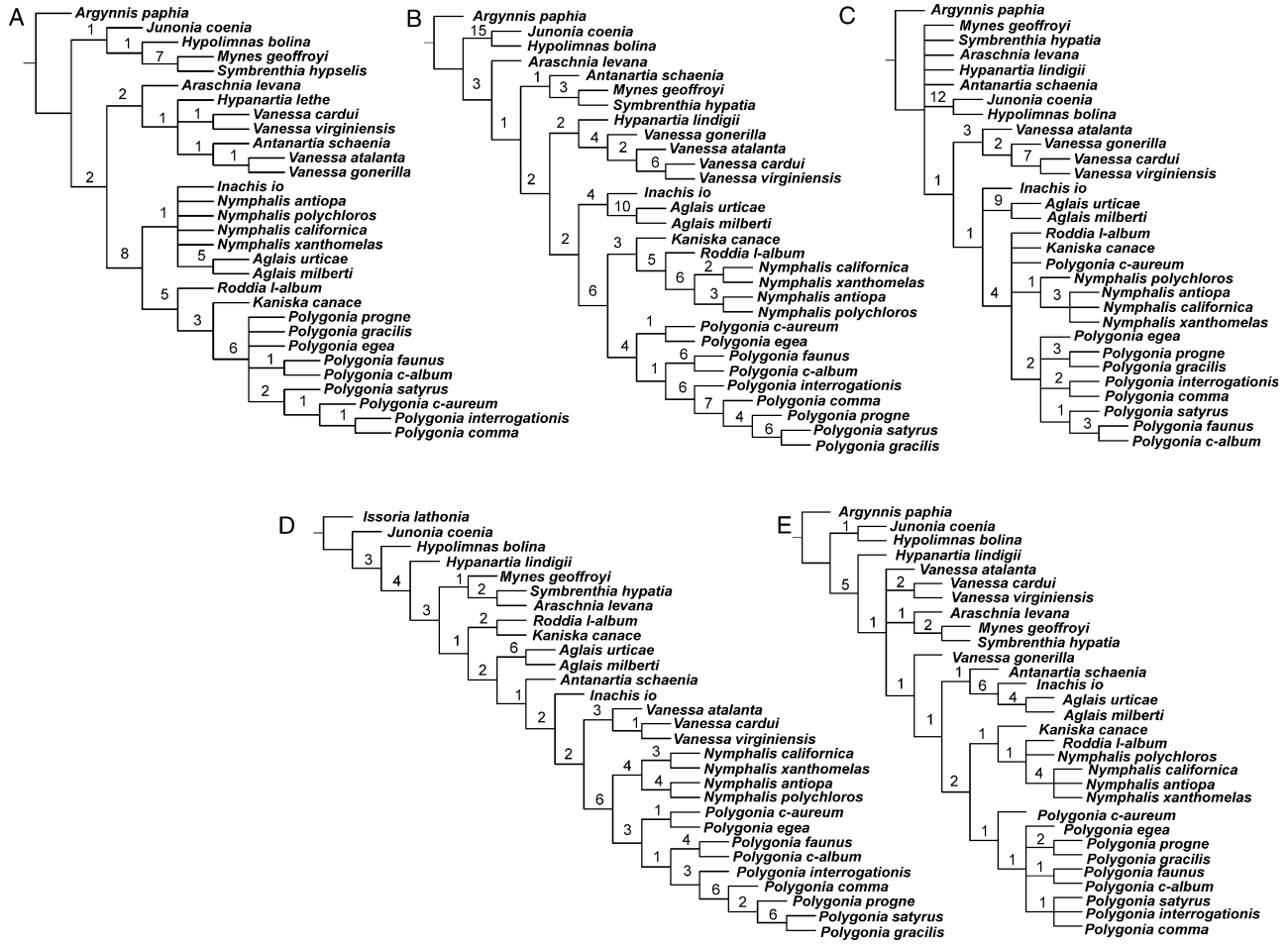

Fig. 1. Results from the analysis of each of the data sets separately. For each tree the CI and RI are calculated using only informative characters. Bremer supports are given for each node. (A) Strict consensus of 16 trees found for the morphological, ecological, and behavioral data set, length 302 steps, CI = 0.42, RI = 0.74. (B) The single most parsimonious tree found for the COI data set, length 1637 steps, CI = 0.34, RI = 0.42. (C) Strict consensus of 27 equally parsimonious trees found for the EF-1α data set, length 566 steps, CI = 0.46, RI = 0.57. (D) The single most parsimonious tree found for the ND1 data set (note that *Vanessa gonerilla* is missing from the data set); length 779, CI = 0.36, RI = 0.41. (E) Strict consensus of 3 equally parsimonious trees found for the *wingless* data set, length 363, CI = 0.50, RI = 0.56.

Table 3
p values of the ILD tests

Partition	MEB	COI	EF-1α	ND1
COI	0.0010*			
EF-1α	0.0010*	0.0090*		
ND1	0.0010*	0.1309	0.0200*	
Wingless	0.0080*	0.5015	0.3906	0.1518

Asterisks (*) denote significant incongruence between data partitions.

Partitioned Bremer support

Partitioned Bremer support values reveal that, despite the significant incongruence of the MEB data partition with the molecular data partitions, the former contributes positively to the BS values of (supports) 16 of 27 nodes, is neutral (has a value of 0) for 5 nodes, and is negative for (conflicts with) 6 nodes. Approximately the same ratio of positive/neutral/negative holds for the molecular data partitions (Fig. 2), though the ratio for each molecular partition is slightly different on the two equally parsimonious trees found from the combined analysis. Further comparison of PBS values is complicated by the fact that for several nodes the PBS values are averages (i.e., several equally parsimonious trees were found when searching with anticonstraints).

There are several nodes for which there is no conflict at all between the data partitions (i.e., there are no negative values) in Tree 1: the nodes defining the *Aglais* clade, the sister species relationship of *P. faunus* and *P. c-album*, the clade containing *Nymphalis*, *Polygonia*, *Roddia*, and *Kaniska* and the clade containing all the species belonging to the tribe Nymphalini (Fig. 2). Several nodes have weak conflict from one or two data partitions, for instance for the *Nymphalis* group, the nodes defining the clade containing *P. progne*, *P. satyrus*, and *P. gracilis*, the *Kaniska* + *Polygonia* clade, and the clade containing all of the *Nymphalis* group species.

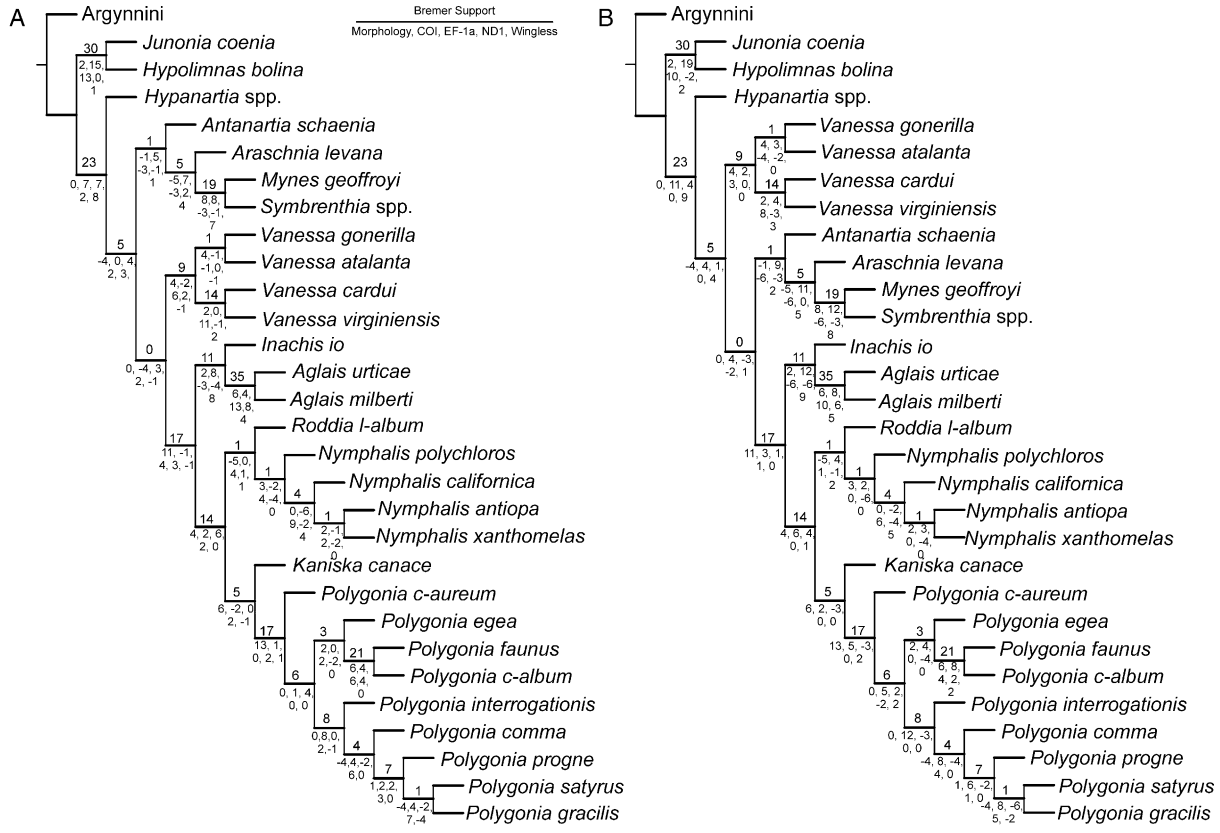

Fig. 2. Results from the analysis of the combined data set, both equally parsimonious trees shown, length 3745 steps, CI = 0.37, RI = 0.48 (the latter two indices calculated using only informative characters). Bremer support values are given above branches and partitioned Bremer support values are given below branches for the MEB, COI, EF-1α, ND1, and *wingless* data partitions. (A) Tree 1. (B) Tree 2.

Finally, several nodes show strong conflict between one or more data partitions and the rest of the partitions, such as the node defining the *Roddia* + *Nymphalis* clade, the *Inachis* + *Aglais* clade, all of the nodes within the *Nymphalis* clade, and the node defining *P. comma* as basal to three other species of *Polygonia*.

In Tree 1, there are only three nodes at which the molecular data partitions are in unanimous disagreement with the MEB data partition (Fig. 2A). The MEB data partition is in moderate conflict with the monophyly of all Nymphalini species excluding *Hypanartia* (the trees favored by the MEB data partition place *Hypanartia* within the Nymphalini, either sister to the *Symbrenthia*/*Nymphalis* clade in Fig. 2B, or sister to the *Vanessa*/*Nymphalis* clade in Fig. 2A) and with the position of *R. l-album* (the tree favored by the MEB data partition places *R. l-album* as basal to the *Polygonia* + *Kaniska* clade). The molecular data partitions are in weak conflict over the sister relationship of *Vanessa atalanta* and *V. gonerilla,* while the MEB data partition gives moderate support for this relationship. The mitochondrial data partitions are in conflict with the other data partitions at five nodes. MtDNA is in weak to moderate conflict with the relationships within the genus *Nymphalis*, while the other data partitions are either neutral or give moderate to strong support for the relationships shown in Fig. 2. MtDNA give good to strong support for the sister relationship of *Polygonia satyrus* and *P. gracilis* and for the clade containing *P. satyrus*, *P. gracilis*, *P. progne*, and *P. comma*, whereas the other data partitions are either neutral or are in weak to moderate conflict with these relationships. The nuclear genes do not conflict with the other data partitions together at any nodes. Within the *Nymphalis* group there are no other nodes at which one or more data partitions would be in moderate to strong conflict with the other data partitions.

For 12 of 27 nodes, PBS scores are calculated from only a single most parsimonious tree not containing that clade, and thus the PBS values shown are not averaged. For those nodes with averaged PBS values, the number of equally parsimonious trees that do not contain those nodes varied from two to eight (mode 2). Within the *Nymphalis* group clade, 11 nodes have PBS values that are averaged (Fig. 3). For the nodes with two to four equally parsimonious trees found without the node, the data partitions consistently give support or are in conflict regardless of the tree, although the magnitude differs. For 3 of the nodes in the *Polygonia* clade, we found six or eight trees and at these nodes the same data

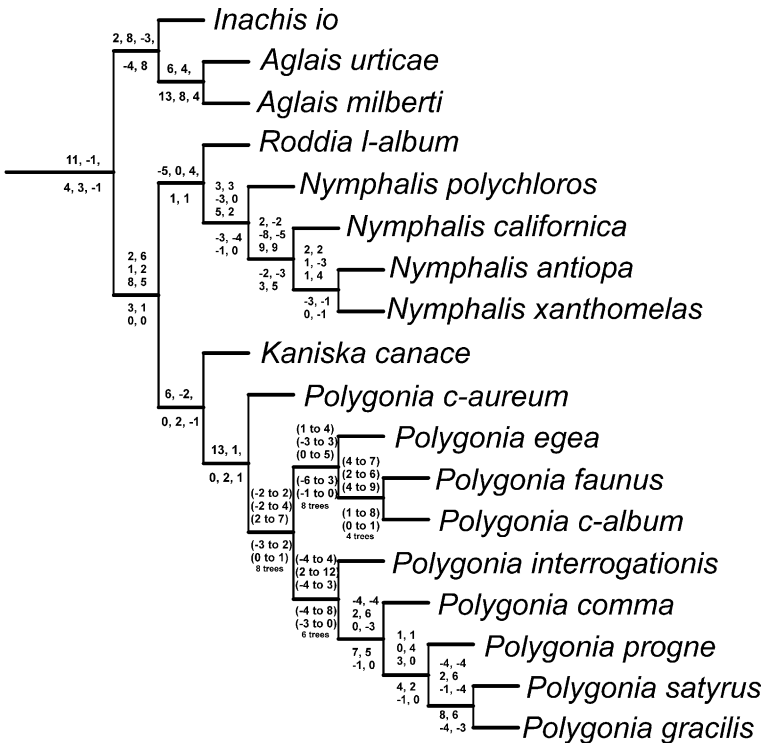

Fig. 3. Breakdown of the partitioned Bremer support values of the *Nymphalis* group clade. For the nodes with only five numbers, the PBS values are calculated from a single most parsimonious tree not containing that node (MEB, COI, EF-1α, ND1, and *wingless*, respectively). For the nodes containing two columns of numbers in five rows, the PBS values are calculated from two equally parsimonious trees and both values are shown (from top to bottom: MEB, COI, EF-1α, ND1, and *wingless*). For the nodes with ranges of numbers within parentheses, the PBS values are calculated from four to eight trees and only the range is shown and the number of trees is indicated (order of partitions remains the same).

partition can in several cases either give moderate support for or conflict moderately with the node (Fig. 3). For the node defining the clade containing *P. interrogationis* to *P. gracilis*, there are six equally parsimonious trees not containing that node. In all of these *P. interrogationis* is placed immediately after *P. c-aureum*. This position is given some support from all data partitions except the COI partition in some of the trees. On the other hand, the same data partitions support the most parsimonious position of *P. interrogationis* on the other trees. For the node defining the clade containing *P. egea* to *P. c-album,* there are eight equally parsimonious trees not containing that node. In four of the trees, *P. egea* is placed immediately after *P. c-aureum* and in the other four trees, *P. egea* is basal to the *P. interrogationis* clade. Both positions receive support from the three partitions that conflict with the most parsimonious position (COI, ND1, and *wingless*), yet the first two partitions also support the most parsimonious position of *P. egea*.

Reconstructions of host plant utilization

Tracing the use of host plant families in the *Nymphalis* group clade onto the formerly proposed phylogeny (Nylin et al., 2001) and onto the phylogeny now emerging as the best supported (Fig. 3) did not result in major differences (cf. Janz et al., 2001).

Urticaceae. Both phylogenetic reconstructions suggest that this host plant family was used by the ancestor of the *Nymphalis* group clade. It was lost by the ancestor of the clade *Nymphalis* + *Roddia* + *Kaniska* + *Polygonia* (henceforth the NP clade) and recolonized one or more times by species of *Polygonia*. The new reconstruction has fewer equivocal branches and suggests that recolonization did not happen until after *P. c-aureum* had branched off, early in the evolution of *Polygonia*.

Ulmaceae. Both reconstructions suggest one colonization by *Hypanartia* and one or more in the *Nymphalis* group clade. For the latter, the original reconstruction was equivocal as to details, allowing for the possibility of an early colonization by the ancestor of the NP clade, followed by losses and recolonizations. The new one is more clear in suggesting one colonization by the ancestor of *Nymphalis* + *Roddia*, one by *P. c-album*, and one or two by *P. comma* and *P. interrogationis*.

Cannabidaceae. Both reconstructions suggest one colonization by *Inachis* and at least one by *Polygonia*. In this case the new reconstruction is more equivocal and allows for a strong possibility (favored by weighting gains heavier) that the family was colonized already by the ancestor of *Polygonia* and later lost several times.

Salicaceae. For this family results were similar to those for Ulmaceae, in that the original reconstruction allowed for the possibility of a single colonization by the ancestor of the NP clade, followed by losses and recolonizations. Also for this family, the new reconstruction instead suggests one colonization by the ancestor of *Nymphalis* + *Roddia* and "independent" colonizations within *Polygonia*.

Grossulariaceae. This family is used in two sections of *Polygonia* that are well separated in both reconstructions; i.e., in both cases the optimization suggests at least two colonizations, one by *P. c-album* + *P. faunus* and one by *P. gracilis* + *P. progne*. However, in the new phylogeny *P. satyrus* intrudes as the possible sister species to *P. gracilis*, creating an equivocal reconstruction where use of *Ribes* may have evolved "independently" in *P. gracilis* and *P. progne* unless gains are given larger weight than losses.

Ericaceae. Similarly to Grossulariaceae, this family is used in two sections of *Polygonia*, by *P. faunus* and by *P. gracilis* + *P. progne*. The implications of an intruding *P. satyrus* are the same.

Betulaceae. The reconstructions are exactly the same, with four colonizations within the NP clade even if gains are weighted twice as heavily as losses.

Discussion

Our results are largely consistent with those of Nylin et al. (2001), yet they differ in certain important aspects. We found that the nodes with good to strong support in the previous study had even stronger support with the addition of new data. When comparing our favored phylogenetic hypothesis (Fig. 2A) with that of Nylin et al. (2001, their Fig. 4), we find the following similarities. Species belonging to the tribe Nymphalini form a very strongly supported monophyletic group, with *Hypanartia* coming out as the basal genus. *Araschnia*, *Symbrenthia*, and *Mynes* form a monophyletic group, as do the species of *Vanessa* (although the relationships within *Vanessa* differ between the studies). The *Nymphalis* group forms a strongly supported clade, within which the relationships of *Inachis*, *Aglais*, *Nymphalis*, and *Kaniska* remain the same. *Polygonia* remains a monophyletic group, although the relationships of most of the species change quite dramatically. The sister species relationships of *P. c-album* and *P. faunus* does not change, however. For these groups, the conclusions of Nylin et al. (2001) remain the same.

Our results differ from those of Nylin et al. (2001) in some crucial aspects. According to our results, *Antanartia* is the basal genus of the *Symbrenthia* clade, although its position has weak support. *Antanartia* still does not group with *Hypanartia*, in keeping with the conclusion of the previous study. It is clear that the position of *Antanartia* is likely to change in the future with the addition of more taxa and more data. In our study, we find that the sister group to the *Nymphalis* group may be the genus *Vanessa*. This arrangement has no Bremer support (an equally parsimonious tree places the *Symbrenthia* clade as sister to the *Nymphalis* group), but does have strong enough character support to tip the balance in its favor: the unique trait of adult hibernation. This trait is unlikely to have arisen twice, as a change in the hibernation stage would have great ramifications on the life cycle of a butterfly, involving changes in adult and larval physiology, host plant use, timing of reproduction, etc. Once again it is likely that a broader taxon sampling scheme would resolve the dilemma. Within the *Nymphalis* group, we find that the placement of *Roddia* has changed from basal to the *Kaniska* + *Polygonia* clade to basal to the *Nymphalis* clade. This arrangement is in moderate conflict with the MEB data partition, but has good support from the combined molecular data partitions, which thus overrides the morphological support for the alternative arrangement. It is very interesting to note that the molecular data partitions do not override the morphological data partition in any consistent way and in fact overall the MEB contributes equally to the BS values.

Within the *Polygonia* clade, the relationships of the species are much clearer than in the previous study and almost all nodes have good to strong support. *P. c-aureum* is the most basal species, with the rest of the species falling into two clades. One clade contains the European species (one Asian species, *P. giganteum*, is missing from the data set) and *P. faunus*, while the other clade is made up of only North American species (one Mexican species, *P. haroldii*, and one North American species, *P. oreas*, are missing from the data set). Within this second clade, the basal positions of *P. interrogationis* and *P. comma* are quite clear, but the relationships of the three species *P. progne*, *P. satyrus*, and *P. gracilis* are not entirely clear.

At this point, it can be noted that although reconstruction of host plant utilization by character optimization changes in details if the improved hypothesis of phylogeny (Fig. 3) is used rather than the one proposed by Nylin et al. (2001), the conclusions of Janz et al. (2001) are not challenged by the new information. The dynamic pattern of host plant utilization in the tribe remains, with frequent recolonizations of the same host plant families, suggesting that such colonizations are not truly "independent."

We have been able to dissect our combined data set in great detail to determine which partitions are contributing to the presence of each node in the most parsimonious trees using PBS. By combining our data sets we have been able to use the maximum amount of information about the relationships of species in the *Nymphalis* group that is available. Relying purely on the

morphological or any one molecular data set alone would lead to different results, and in the light of the current results, they would perhaps be misleading. The separate analysis of data sets and subsequent combining of the trees produced would in our case lead to an almost entirely unresolved tree (only the sister relationship of *A. urticae* + *A. milberti*, *P. c-album* + *P. faunus*, and *V. cardui* + *V. virginiensis* are common to all trees, see Fig. 1), as has been found in previous studies (e.g., Miller et al., 1997).

Through PBS analysis, we have been able to show that morphological data can have an important role in resolving nodes at which molecular data are ambiguous or even weakly disinformative. Such a node in our data set is the one defining *Kaniska canace* as basal to the *Polygonia* clade (Figs. 2 and 3). Two molecular data sets (COI and *wingless*) place *Kaniska* as basal to the *Roddia*+*Nymphalis* clade, but with weak support. It may be that the quick divergence of the lineages has not left a mark in the genes sequenced, but it has obviously left its mark in other genes coding for morphological characters. We thus suggest that morphological data can give very valuable information that may not be available in the restricted set of genes used normally for molecular systematic studies.

Our combined analysis has given us a fairly robust phylogeny for the *Nymphalis* group, one which we can use to critically appraise the current classification of the group. While the species concepts in the *Nymphalis* group have remained quite stable over the past 100 years, the species have bounced around in many genera depending on which field guide or checklist one peruses. For instance, *R. l-album* has often been included in *Nymphalis* or *Polygonia* and even *Vanessa*. *K. canace* is often placed in *Nymphalis* or *Vanessa*. *A. milberti* is usually placed in *Nymphalis*, and *Polygonia* has been split into several genera earlier. Recently, some checklists (Aarvik et al., 2000; Kullberg et al., 2001) have lumped all of the *Nymphalis* group genera into the genus *Nymphalis*, with the genera considered here being demoted to subgenus rank. However, over the past 20 years a consensus has emerged that the *Nymphalis* group consists of at least the following genera: *Nymphalis*, *Polygonia*, *Kaniska*, *Aglais*, and *Inachis*. The genus *Roddia* was described only recently (Korshunov, 1996) and has not been widely accepted yet.

A classification should reflect the phylogenetic relationships of species as closely as possible, but at the same time it should be stable. One way to confer stability to a classification is to base it on the most robust phylogenetic hypothesis available. Our preferred phylogenetic hypothesis for the *Nymphalis* group (Fig. 3) can be subdivided in several ways. First of all, the entire *Nymphalis* group has very strong support and it is unlikely that the addition of new data will cause it to become nonmonophyletic. Thus there is a certain justification in lumping all the genera into the genus *Nymphalis* (e.g. Aarvik et al., 2000; Kullberg et al., 2001). However, such an action will be quite unstabilizing as the large literature on species in the genera *Aglais* and *Polygonia* will likely be lost to future generations of evolutionary biologists, as these names become forgotten and electronic retrieval becomes the only method of examining past studies.

The opposite extreme, continuing to recognize the six current genera, also has some justification for the genera *Polygonia*, *Inachis*, *Aglais*, and *Kaniska*. However, the genera *Nymphalis* and *Roddia* have weak support and their relationships may change with the addition of new data, although our results suggest that more molecular data would strengthen the monophyly of this clade. Also, the retention of monotypic genera is uninformative in a classification, particularly when a monotypic genus is a sister lineage to a monophyletic, multispecies genus.

A solution may be to retain *Aglais* (including *Inachis*) and *Nymphalis* (including *Polygonia*, *Kaniska*, and *Roddia*). The male genitalia in these two groups are very distinct from each other (Niculescu, 1965). However, such an action would necessitate relegating the well-studied genus *Polygonia* to future oblivion. We thus find the best compromise between well-supported clades and nomenclatorial stability to be the recognition of three genera: *Aglais* (including *Inachis*), *Nymphalis* (including *Roddia*), and *Polygonia* (including *Kaniska*). We present our suggested classification for the *Nymphalis* group in the Appendix. Such a generic classification has the advantage of remaining stable should further evidence suggest new placements for *P. canace* and/or *N. l-album*. In such a case, the only taxon to be affected is the taxon with the new placement, with no need to resurrect the monotypic genera now being placed in synonymy.

In this study, we have been able to propose a phylogenetic hypothesis for the species in the genera *Aglais*, *Nymphalis*, and *Polygonia* that we believe will be robust to the addition of further data. Our current data set consists of morphological data, two mitochondrial gene sequences, and two (unlinked) nuclear gene sequences. By combining all of the data sets and using partitioned Bremer support, we have been able to dissect our data partitions to investigate which partitions conflict at which nodes and how they do so. We believe that we have been able to show the utility of combining all available data to arrive at a robust hypothesis rather than subjectively choosing some partitions over others. By using PBS, we can identify nodes that may potentially change with the addition of new data (nodes where different data partitions conflict strongly with each other), and nodes that are unlikely to change with future new data (nodes where all data partitions are in agreement). We have also shown the utility of combining morphological data with molecular data and suggest that there is a lot of extra information in morphology

that is not captured in the restricted set of genes being sequenced presently.

Acknowledgments

We are, as ever, extremely grateful to those who have helped us obtain butterfly specimens: Andy Brower, Anton Chichvarkhin, Konrad Fiedler, Darrell Kemp, Norbert Kondla, Mike Leski, Toomas Tammaru, and Christer Wiklund. We thank Elisabet Weingartner for help in the lab. We are grateful to Chris Lambkin and an anonymous referee for useful comments on a previous version of the manuscript.

Appendix A

Proposed generic classification of the *Nymphalis* group based on results presented in this paper (for a full synonymic list of the species see http://www.zoologi.su.se/research/wahlberg).

Aglais Dalman, 1816 (Type species *Papilio urticae* Linnaeus, 1758)
= *Inachis* Hübner, 1819 (Type species *Papilio io* Linnaeus, 1758) **n. syn.**
= *Ichnusa* Reuss, 1939 (Type species *Papilio (Vanessa) ichnusa* Bonelli, 1826)
Aglais urticae (Linnaeus, 1758)
Aglais ladakensis (Moore, 1882)
Aglais kaschmirensis (Kollar, 1844)
Aglais milberti (Godart, 1819)
Aglais io (Linnaeus, 1758) **n. comb.**

Nymphalis Kluk, 1802 (Type species *Papilio polychloros* Linnaeus, 1758)
= *Scudderia* Grote, 1873 (Type species *Papilio antiopa* Linnaeus, 1758)
= *Euvanessa* Scudder, 1889 (Type species *Papilio antiopa* Linnaeus, 1758)
= *Roddia* Korshunov, 1996 (Type species *Papilio l-album* Esper, 1781)
Nymphalis l-album (Esper, 1781)
Nymphalis polychloros (Linnaeus, 1758)
Nymphalis californica (Boisduval, 1852)
Nymphalis xanthomelas (Denis and Schiffermüller, 1775)
Nymphalis antiopa (Linnaeus, 1758)
Nymphalis cyanomelas (Doubleday, 1849)

Polygonia Hübner, 1819 (Type species *Papilio c-aureum* Linnaeus, 1758)
= *Eugonia* Hübner, 1819 (Type species *Papilio angelica* Stoll, 1782)
= *Comma* Rennie, 1832 (Type species *Papilio c-album* Linnaeus, 1758)
= *Grapta* Kirby, 1837 (Type species *Vanessa (Grapta) c-argenteum* Kirby, 1837)
= *Kaniska* Moore, 1899 (Type species *Papilio canace* Linnaeus, 1763)
Polygonia canace (Linnaeus, 1763)
Polygonia c-aureum (Linnaeus, 1758)
Polygonia giganteum (Leech, 1883)
Polygonia egea (Cramer, 1775)
Polygonia c-album (Linnaeus, 1758)
Polygonia faunus (Edwards, 1862)
Polygonia interrogationis (Fabricius, 1798)
Polygonia comma (Harris, 1842)
Polygonia progne (Cramer, 1776)
Polygonia satyrus (Edwards, 1869)
Polygonia oreas (Edwards, 1869)
Polygonia gracilis (Grote and Robinson, 1867)
Polygonia haroldii (Dewitz, 1877)

References

Aarvik, L., Berggren, K., Hansen, L.O., 2000. Catalogus Lepidopterorum Norvegiae. Lepidopterologisk Arbeidsgruppe, Zoologisk Museum, Universitet i Oslo, Norsk Institutt for Skogsforskning, Ås, Oslo.

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

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

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., DeSalle, R., 1998. Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: The utility of *wingless* as a source of characters for phylogenetic inference. Ins. Mol. Biol. 7, 73–82.

Bryant, S.R., Thomas, C.D., Bale, J.S., 2002. The influence of thermal ecology on the distribution of three nymphalid butterflies. J. Appl. Ecol. 39, 43–55.

de Queiroz, A., Donoghue, M.J., Kim, J., 1995. Separate versus combined analysis of phylogenetic evidence. Annu. Rev. Ecol. Syst. 26, 657–681.

DeSalle, R., Brower, A.V.Z., 1997. Process partitions, congruence, and the independence of characters: Inferring relationships among closely related Hawaiian *Drosophila* from multiple gene regions. Syst. Biol. 46, 751–764.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing the significance of incongruence. Cladistics 10, 315–319.

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.

Givnish, T.J., Sytsma, K.J., 1997. Consistency, characters, and the likelihood of correct phylogenetic inference. Mol. Phylogenet. Evol. 7, 320–330.

Goloboff, P.A., 1998. NONA. Ver. 2.0. published by author.

Hall, T.A., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids. Symp. Ser. 41, 95–98.

Hedges, S.B., Maxson, L.R., 1996. Re: Molecules and morphology in amniote phylogeny. Mol. Phylogenet. Evol. 6, 312–314.

Hiroyoshi, S., Mitsuhashi, J., 1999. Sperm reflux and its role in multiple mating in males of a butterfly *Polygonia c-aureum* Linnaeus. (Lepidoptera: Nymphalidae). J. Ins. Physiol. 45, 107–112.

Janz, N., Nylin, S., 1997. The role of female search behaviour in determining host plant range in plant feeding insects: A test of the information processing hypothesis. Proc. R. Soc. Lond. B 264, 701–707.

Janz, N., Nylin, S., Nyblom, K., 2001. Evolutionary dynamics of host plant specialization: A case study of the tribe Nymphalini. Evolution 55, 783–796.

Kluge, A.G., 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38, 7–25.

Kluge, A.G., Wolf, A.J., 1993. Cladistics: What's in a word? Cladistics 9, 183–199.

Korshunov, Y., 1996. Additions and Corrections to the Book ''Butterflies of the Asian Part of Russia''. ETA Group, Novosibirsk (in Russian).

Kullberg, J., Albrecht, A., Kaila, L., Varis, V., 2001. Checklist of Finnish Lepidoptera—Suomen perhosten luettelo. Sahlbergia 6, 45–190.

Lambkin, C.L., Lee, M.S.Y., Winterton, S.L., Yeates, D.K., 2002. Partitioned Bremer support and multiple trees. Cladistics 18, 436–444.

Maddison, W.P., Maddison, D.R., 1992. MacClade Ver. 3. Sinauer, Sunderland, MA.

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-1α gene sequences. Mol. Phylogenet. Evol. 18, 264–281.

Niculescu, E.V., 1965. Fauna Republicii populare Române: Insecta, Familia Nymphalidae.

Nixon, K.C., 2002. Winclada. Ver. 1.00.08. Published by author.

Nixon, K.C., Carpenter, J.M., 1996. On simultaneous analysis. Cladistics 12, 221–241.

Nylin, S., 1988. Host plant specialization and seasonality in a polyphagous butterfly, *Polygonia c-album* (Nymphalidae). Oikos 53, 381–386.

Nylin, S., 1992. Seasonal plasticity in life-history traits—Growth and development in *Polygonia c-album* (Lepidoptera, Nymphalidae). Biol. J. Linn. Soc. 47, 301–323.

Nylin, S., Nyblom, K., Ronquist, F., Janz, N., Belicek, J., Källersjö, M., 2001. Phylogeny of *Polygonia*, *Nymphalis* and related butterflies (Lepidoptera: Nymphalidae): A total-evidence analysis. Zool. J. Linn. Soc. 132, 441–468.

O'Grady, P.M., Clark, J.B., Kidwell, M.G., 1998. Phylogeny of the *Drosophila saltans* species group based on combined analysis of nuclear and mitochondrial DNA sequences. Mol. Biol. Evol. 15, 656–664.

Sorensen, M.D., 1999. TreeRot. Ver. 2.0. Boston University.

Swofford, D.L., 1998. PAUP*: Phylogenetic Analysis Using Parsimony. (*and Other Methods). Ver.4.0b10. Sinauer Associates, Sunderland, MA.

Tanaka, D., Sakurama, T., Mitsumasu, K., Yamanaka, A., Endo, K., 1997. Separation of bombyxin from a neuropeptide of *Bombyx mori* showing summer-morph-producing hormone (SMPH) activity in the Asian comma butterfly, *Polygonia c-aureum* L. J. Ins. Physiol. 43, 197–201.

Wahlberg, N., Zimmermann, M., 2000. Pattern of phylogenetic relationships among members of the tribe Melitaeini (Lepidoptera: Nymphalidae) inferred from mtDNA sequences. Cladistics 16, 347–363.

Wiklund, C., Wickman, P.O., Nylin, S., 1992. A sex difference in the propensity to enter direct/diapause development—A result of selection for protandry. Evolution 46, 519–528.