Systematic Entomology (2004) **29**, 115–131

Molecular phylogeny of the Oriental butterfly genus *Arhopala* (Lycaenidae, Theclinae) inferred from mitochondrial and nuclear genes

HENDRIK-JAN MEGENS[1,2], W. JOHAN VAN NES[2], COLINE H. M. VAN MOORSEL[2], NAOMI E. PIERCE[3] and RIENK DE JONG[1]

[1]National Museum of Natural History, Leiden, The Netherlands, [2]Institute of Evolutionary and Ecological Sciences, Leiden University, Leiden, The Netherlands and [3]Department of Organismic and Evolutionary Biology (Museum of Comparative Zoology Laboratories), Harvard University, Cambridge, Massachusetts, U.S.A.

Abstract. We present a phylogeny for a selection of species of the butterfly genus *Arhopala* Boisduval, 1832 based on molecular characters. We sequenced 1778 bases of the mitochondrial genes Cytochrome Oxidase 1 and 2 including $tRNA^{Leu}$, and a 393-bp fragment of the nuclear *wingless* gene for a total of 42 specimens of 33 species, representing all major species groups. Analyses of mtDNA and *wingless* genes show congruent phylogenetic signal. The phylogeny presented here confirms the monophyly of the *centaurus*, *eumolphus*, *camdeo* and *epimuta* groups and the *amphimuta* subgroup. It confirms close relationships between species within the *agelastus* group, that together with the *amphimuta* subgroup, *centaurus* and *camdeo* groups form a monophyletic group. However, incongruencies with previous taxonomic studies also occur; the *amphimuta* and *silhetensis* groups are not monophyletic, as is the genus *Arhopala* itself. One enigmatic species, *A. kinabala*, was evaluated further for topology and the support for basal placement of this species is due mainly to the *wingless* gene. However, in the Parsimony analysis, and subsequent Maximum Likelihood evaluations, certain nodes could not be resolved due to insufficient support. The mtDNA shows extreme AT bias with compositional heterogeneity at 3rd codon positions, which may result in saturation. By contrast, the *wingless* gene does not show compositional bias, suggesting that poor support is not due solely to saturation. The evaluation of morphological characters used in previous studies on *Arhopala* systematics on the molecular tree indicates that the macular pattern and the absence of tails at the hind wings show extensive homoplasy. A significant phylogenetic signal (as indicated by T-PTP tests) is present in several of these morphological characters, which are nevertheless of limited use in phylogenetic studies due to their labile nature.

Introduction

The Sunda region, which consists of the Malay Peninsula, Borneo, Sumatra and Java, has the largest number of butterfly species in regions of the Old World (approximately 1300). Many species are largely or entirely restricted to the area (around 60%), but local endemism is relatively low even though the area is highly fragmented into islands. This is particularly evident in the large island of Borneo, where only approximately 20% of the butterflies are endemic (R. de Jong, unpublished data). Of those few endemics that exist, most are confined to mountainous areas. The spread of lowland species across the Sunda area can be explained by the lowering of sea level during Pleistocene glaciations, when the islands and the Malay Peninsula were occasionally connected by land (Voris, 2000).

Correspondence: Hendrik-Jan Megens, Cell Biology and Immunology, Department of Animal Sciences, Wageningen University, PO Box 338, 6700AH Wageningen, The Netherlands. E-mail: Hendrik-Jan.Megens@wur.nl

© 2004 The Royal Entomological Society

The repeated breaking up of the area during the Pleistocene has been proposed to be a driving force in speciation (Corbet, 1946), which would be congruent with patterns observed in the African (Cobb *et al.*, 2000) and Amazonian (Bush, 1994) tropics. However, if these periods of isolation are important factors, then in the Sunda region one would expect higher local endemicity. In all, the high diversity in this region remains largely unexplained. Many factors undoubtedly have played a role, abiotic (geological and climatological history) as well as biotic (such as habitat partitioning and host-plant selection) (Moritz *et al.*, 2000).

Prominent in the butterfly diversity of the Sunda region is the genus *Arhopala* Boisduval, 1832 (Lycaenidae: Theclinae) with approximately 120 species making up almost 10% of all butterfly species. Thus it is an excellent group of organisms to analyse possible causes for the high diversity of the butterflies and other group of organisms in this particular area.

Diversity in the genus Arhopala

The distribution of *Arhopala* ranges from Sri Lanka to Japan, and from southern Nepal, throughout the Archipelago, to the northern tropics of Australia. With approximately 200 species described, it is among the largest genera of butterflies. However, the true diversity cannot be summarized by simple numbers as most species are further subdivided in subspecies or forms (for revisions and species lists, see Evans, 1957; Eliot, 1963, 1972; D'Abrera, 1986, 1990; Bridges, 1988; Corbet & Pendlebury, 1992; Parsons, 1998). Such forms often coincide with geographical regions, such as northernmost limits of the Sunda region, and are most likely expressions of ecotypic variation (Corbet, 1946). Further east, in New Guinea and its associated islands, there are approximately 40 species that are virtually absent elsewhere (D'Abrera, 1990; Parsons, 1998) because of isolation in the absence of past or present land bridges.

Arhopala is also diverse in life-history traits as research on foodplant use has shown (reviewed in Fiedler, 1991; Parsons, 1998). Both oligophagous and polyphagous preferences exist, and wide ranges of plant families are used. Furthermore, as is usual in Lycaenidae, all *Arhopala* species appear to have very specific symbiotic relationships with ants (reviewed in Fiedler, 1991). There is a wide variety of myrmecophilous relationships throughout *Arhopala* including obligatory myrmecophily (Maschwitz *et al.*, 1984) and even carnivory (*A. wildei*; King & Ring, 1996).

Despite the wide variety in habitat preference and distribution, *Arhopala* morphology is remarkably uniform throughout the genus. Usually the wings of the male are bright lustrous blue to violet-blue on the upperside with a straight black margin. The females are slightly duller violet-blue in colour and the black margins are irregular and broad. The underside is brown with darker brown markings that in many species are rather faint. Although there are some conspicuous exceptions in which, for instance, the males have a green upperside or the underside markings are extraordinarily dark or absent all together, most species are not easily identified. Although the wing markings and genitalia usually provide adequate information, it is normally only by a combination of characters that most species can be identified with certainty (Corbet, 1946; D'Abrera, 1986).

Taxonomy of Arhopala

Several butterfly systematists have studied and revised *Arhopala* thoroughly, and all of them had trouble dividing the genus into smaller groups in a satisfactory manner. As early as 1889, Doherty (in Evans, 1957) remarked that *Arhopala* was a 'cumbrous genus and that every opportunity should be taken to divide it'. At that time many *Arhopala* species were put in the genus *Amblypodia* Horsfield, 1829. Among the first to recognize that *Arhopala* constituted a natural group of species was Corbet (1941) who was also the first to make an attempt to split up the Indo-Malayan species of the regions into smaller groups.

Evans (1957) made the first revision of the whole genus, including the species of the Australian region. He further split the *Arhopala* group into the genera *Arhopala s. str.* (five species from New Guinea), *Flos* Doherty, 1889 (thirteen Oriental species), *Aurea* Evans, 1957 (four Oriental species), *Panchala* Moore, 1882 (eight Oriental species) and *Narathura* Moore, 1878 (158, mainly Oriental, but including some Australian/New Guinean species and species groups). These divisions were based mainly on the shape (*Arhopala s. str.*) and size (*Aurea*) of the hindwing cell, and shape of male (*Flos*) and female (*Panchala*) genitalia.

Because *Narathura* still contained the vast majority of species, Evans (1957) divided this genus into a large number of groups. Apparently struggling to use the few structural characters available, he tried to make meaningful combinations with other characters such as the positions and shape of the spots on the underside of the wings. This led to a few strange divisions. For example, *Narathura epimuta* and *Narathura atosia* were put in different species groups on the basis of the latter having a tail on the hindwing, yet they are otherwise indistinguishable. Other divisions were made on the basis of presence or absence of a tail, the shape of the hindwing termen, etc., all in combination with other characters. However, these characters were very inconsistently used. Hindwing cell shape and form was important in establishing *Arhopala s. str.* and *Aurea* as genera, but in the *centaurus* and *hercules* groups, which both have an unusually short hindcell, it was not. These butterflies have such an obvious '*Narathura*-like' appearance that Evans did not believe they really were separate groups. The same is true for the genus *Flos*, which shares many characters with *Arhopala* but does not possess the normal macular underwing appearance. In addition, it has an aberrant male genitalia feature. However, it does share the same unusually short 9th forewing vein with the *abseus* group, which Evans confined to *Narathura*.

The problems with the definitions of the genera and groups by Evans, led Eliot (1963) to conclude that many of these groups were artificial, and subsequently he revised *Arhopala* again. In his view only *Flos* deserved generic status, based mainly on the shape of the uncus of the male genitalia, and he grouped the rest into *Arhopala* again. Eliot also divided *Arhopala* into a large number of what he called 'natural' species groups. However, he too was confronted with the problem of having to define groups on the basis of combinations of characters, characters that could in no way be used in a genus-wide scope. As none of the groups was formally described as taxa, it is not always clear as to what characters were used in particular to define certain groups.

As an example, the shape and position of post discal spots in hindwing spaces 5, 6 and 7 aids in defining eight small, mainly monotypic groups, and two large groups, the *camdeo* group consisting of about ten species, and the large *amphimuta* group consisting of about twenty-four species. Yet, although *Arhopala antimuta* was assigned to the *amphimuta* group, it does not share this particular character state. Apparently other characters were deemed more important, such as the dislocated post discal band of the forewing at vein 4 and the lack of a tail and shape of hindwing termen. Consequently, these particular characters all have their own exceptions in the *amphimuta* group.

Corbet, Evans and Eliot are well-known butterfly systematists and renowned for their contributions to butterfly systematics. Their work is invaluable and many of their conclusions are confirmed by recent research. Despite these important contributions to the systematics of *Arhopala*, they had often radically different views on grouping, and all agreed that the relationships among the major groups within *Arhopala* remained a mystery, as Eliot (1963: p. 189) stated 'as a result of an all-directional evolution'.

Research aims

Arhopala provides the opportunity to study the causes for the high diversity of butterflies and possibly other groups of insects in the Sunda region. Careful evaluation of key ecological and biogeographical changes in a historical framework allows for comparative analyses of evolutionary changes. Such an approach is carried out using knowledge of the historical relationships of the organisms involved. Here we report on a molecular phylogeny of the genus *Arhopala*. Future papers will deal with particular questions in detail, such as biogeographical and ecological implications.

As previous studies (Corbet, 1946; Evans, 1957; Eliot, 1963, 1972) have shown that morphology is a difficult tool for phylogenetic analysis in *Arhopala,* we use DNA sequences. We focus on genes that have proved to be useful in previous studies on butterfly phylogeny (Caterino & Sperling, 1999; Campbell *et al*., 2000; Rand *et al*., 2000), i.e. the mitochondrial genes Cytochrome Oxidase 1 and 2 (Brower, 1994; Brown *et al*., 1994; Brower & Egan, 1997; Brower & DeSalle, 1998; Caterino & Sperling, 1999; Monteiro & Pierce, 2000; Rand *et al*., 2000), and part of a nuclear gene called *wingless* (Brower & DeSalle, 1998; Brower, 2000; Campbell *et al*., 2000). We investigate the substitution patterns of these markers and determine the sources of ambiguity between and within sets of molecular characters. We discuss how far the results are congruent with classification schemes postulated previously and how far they broaden our understanding of *Arhopala* phylogeny. Finally, we address the ambiguities encountered in morphology using cladistic methods, determine levels of homoplasy and phylogenetic content and discuss the problems using these characters in phylogenetic systematics.

Materials and methods

Selection of taxa

Corbet (1946) pointed out that many species of *Arhopala* are extremely rare. We managed to collect material from twenty-eight species of *Arhopala* (including *Flos anniella*) from the Malayan region. In addition, three Australian species were obtained. We believe that these taxa represent a good sample of the taxonomic diversity found within *Arhopala* for the Oriental region as all the groups and subgroups present in the Sunda region that are represented by more than one species are included in this study. We failed to obtain specimens from two New Guinean species groups of interest: the *Arhopala s. str*. (*sensu* Evans, 1957) and the *hercules* group. We managed to collect sequence data from forty-two specimens (Table 1). We chose *Semanga* Distant, 1884 and *Surendra* Moore, 1879 as outgroup taxa to *Arhopala* based on Eliot's (1973) assumption that the genera *Arhopala* s.l., *Flos*, *Thaduka* Moore, 1878, *Mahathala* Moore, 1878 and *Apporasa* Moore, 1884 form a monophyletic group (which he called the *Arhopala* section; this group is synonymous with the subtribe Arhopalina of Corbet & Pendlebury, 1992). The *Semanga* section and *Surendra* section together with the subtribe Arhopalina constitute the tribe Arhopalini (Eliot, 1973).

Collecting, DNA extraction, PCR and sequencing

The specimens were collected as adults in the wild at various sites in Malaysia, Indonesia and Australia (see Table 1 for locality data). Specimens were killed and the whole bodies immediately stored in 100% EtOH and wings were saved dry as reference material. In some instances we only preserved a few legs to keep whole specimens for the museum collection.

Genomic DNA was obtained from single individuals using a salt extraction method. Legs or the first abdominal segment were homogenized in an extraction buffer (50 mM Tris-HCl, 20 mM EDTA, pH 8.0, 2% SDS and 250 $\mu g\, ml^{-1}$ proteinase K). The samples were then incubated at 60 °C for

Table 1. Collection data for the specimens used in this study. Specimens collected by M.-W. Tan (MWT), N. Pierce (NP), D. Cleary (DC), K. Dunn (KD) and H.-J. Megens (HJM). Kaltim refers to the province of East Kalimantan, Kalten to Central Kalimantan, Indonesia

Species and species group	Collector and sample code	Collection date and locality
Arhopala Boisduval, 1832		
abseus group		
A. abseus Hewitson, 1862	MWT 93-E051	VII.1993 Gn. Serapi, Malaysia
agelastus group		
A. labuana Bethune-Baker, 1896	NP 95-Z298	IX.1995 Sarawak, Malaysia
	HJM 97240522	V.1997 Berau, Kaltim, Indonesia
A. alaconia Hewitson, 1869	HJM 97280512	V.1997 Berau, Kaltim, Indonesia
A. barami Bethune-Baker, 1903	NP 95-Z318	IX.1995 Sarawak, Malaysia
	HJM 9725054	V. 997 Berau, Kaltim, Indonesia
agesias group		
A. kinabala Druce, 1895	NP 95-Y263	IX.1995 Pahang, Malaysia
alitaeus group		
A. denta Evans, 1957	HJM 9728051	V.1997 Berau, Kaltim, Indonesia
amphimuta group, ***amphimuta*** subgroup		
A. major Staudinger, 1889	MWT 93-B055	VII.1993 FRIM, Malaysia
	HJM 9729057	V.1997 Berau, Kaltim, Indonesia
A. amphimuta Felder, 1860	MWT 93-C034	VII.1993 FRIM, Malaysia
A. moolaiana Moore, 1879	NP 95-Y246	IX.1995 Pahang, Malaysia
	HJM 97280518	V.1997 Berau, Kaltim, Indonesia
amphimuta group, ***muta*** subgroup		
A. muta Hewitson, 1862	NP 95-Z297	IX.1995 Sarawak, Malaysia
amphimuta group, ***perimuta*** subgroup		
A. antimuta Felder, 1865	NP 95-Y312	IX.1995 Pahang, Malaysia
	HJM 9729053	V.1997 Berau, Kaltim, Indonesia
anthelus group		
A. achelous Hewitson, 1862	MWT 93-D017	VII.1993 Kokol, Sabah, Malaysia
	HJM 97270514	V.1997 Berau, Kaltim, Indonesia
A. anthelus Westwood, 1851	NP 95-Z316	IX.1995 Sarawak, Malaysia
aurea group		
A. aurea Hewitson, 1862	DC 97733	1997 Sangai, Kalten, Indonesia
camdeo group		
A. opalina Moore, 1884	MWT 93-C071	VII.1993 Genting, Malaysia
A. hellada Fruhstorfer, 1914	MWT 93-E050	VII.1993 Gn. Serapi, Malaysia
A. wildei Miskin, 1891	KD 94-T069	VII.1994 Queensland, Australia
centaurus group		
A. centaurus Fabricius, 1775	KD 94-T070	VII.1994 Queensland, Australia
A. madytus Fruhstorfer, 1914	KD 95-Z559	XI.1995 Queensland, Australia
A. pseudocentaurus Doubleday, 1847	NP 95-Z291	IX.1995 Sarawak, Malaysia
cleander group		
A. ace De Nicéville, 1892	NP 95-Y261	IX.1995 Pahang, Malaysia
A. agrata De Nicéville, 1890	MWT 93-E068	VII.1993 Gn. Serapi, Malaysia
A. silhetensis Hewitson, 1862	HJM 97240523	V.1997 Berau, Kaltim, Indonesia
democritus group		
A. democritus Fabricius, 1793	MWT 93-E075	VII.1993 Gn. Serapi, Malaysia
	HJM 97240526	V.1997 Berau, Kaltim, Indonesia
epimuta group		
A. epimuta Moore, 1857	MWT 93-C014	VII.1993 Papah, Malaysia
	HJM 9727058	V.1997 Berau, Kaltim, Indonesia
A. atosia Hewitson, 1863	HJM 97270523	V.1997 Berau, Kaltim, Indonesia
eumolphus group		
A. eumolphus Cramer, 1780	NP 95-Y211	VIII.1995 Pahang, Malaysia
A. horsfieldi Pagenstecher, 1890	NP 95-Y249	IX.1995 Pahang, Malaysia
ganesa group		
A. paraganesa De Nicéville, 1882	MWT 93-D026	VII.1993 Kokol, Sabah, Malaysia
rama group		
A. buddha Bethune-Baker, 1903	DC 979268	1997 Sangai, Kalten, Indonesia
Flos *anniella* Hewitson, 1862	DC 977714	1997 Sangai, Kalten, Indonesia
	MWT 93-C048	VII.1993 Genting Ridge, Malaysia
Semanga *superba* Druce, 1873	MWT 93-C070	VII.1993 Genting,Malaysia
Surendra *vivarna* Horsfield, 1829	MWT 93-B049	VII.1993 Tapah, Malaysia

Table 2. Primers used in this study. Numbers refer to the *Drosophila yakuba* mitochondrial genome (Clary & Wolstenholme, 1985). Primers are taken from (1) Simon *et al.* (1994), (2) Brower (1994) and (3) Brower & DeSalle (1998). Primers marked with '-M' were slightly modified by the authors.

Position (mtDNA)	Name	Source	Sequence (5′ to 3′)
C1-J-1751	Ron-M	1	GGA GCT CCT GAC ATA GCA TTC CC
C1-N-2191	Nancy	1	CCC GGT AAA ATT AAA ATA TAA ACT TC
C1-J-2164	Tonya	1	GAA GTT TAT ATT TTA ATT TTA CCG GG
C1-N-2757	Hobbes	1	AAA TGT TGN GGR AAA AAT GTT A
C1-J-2792	George	2	ATA CCT CGA CGT TAT TCA GA
C2-N-3275	Phyllis	2	GTA ATA GCI GGT AAR ATA GTT CA
C2-J-3291	Strom	2	TAA TTT GAA CTA TYT TAC CIG C
C2-N-3778	Eva-M	2	ATT ACT TGC TTT CAG TCA TCT
	LepWG1	3	GAR TGY AAR TGY CAY GGY ATG TCT GG
	LepWG2	3	ACT ICG CAR CAC CAR TGG AAT GTR CA

1 h, and proteins and salts were precipitated by adding a quarter volume of saturated NaCl. Samples were put on ice (30 min) and after spinning (15 min, >14 000 **g**) the supernatant was saved and DNA was ethanol-precipitated.

Parts of the mitochondrial genes Cytochrome Oxidase 1 and 2 were amplified by PCR using the primers listed in Table 2. The following PCR profile was used: thirty-five cycles of 1 min denaturation at 94 °C, 1 min annealing at 50 °C and extension for 1 min at 72 °C. For partial amplification of the nuclear *wingless* gene, conditions were slightly different with forty cycles at an annealing temperature of 52 °C.

PCR products were gel purified (1% agarose) using the Qiagen gel extraction kit. Forward and reverse sequencing was done with the Dye Terminator biochemicals (Perkin Elmer, Foster City, CA, U.S.A.) using the PCR primers, according to the manufacturer's protocols. Sequence reactions were purified on Cephadex columns, and electrophoresis was performed on ABI 373 and 377 automated sequencers.

Sequence comparisons and phylogenetic inference

The Cytochrome Oxidase 1 and 2 sequences (including the $tRNA^{Leu}$) were aligned to the *Drosophila yakuba* mitochondrial genome (Clary & Wolstenholme, 1985) and compared with the complete CO1 and 2 DNA and amino-acid sequences for *Papilio machaon* (Caterino & Sperling, 1999). The *wingless* sequences were compared with the *Drosophila melanogaster wg* sequence (Rijsewijk *et al.*, 1987) and the *Heliconius erato wingless* sequence (Brower & DeSalle, 1998). Translations in amino acids were made using MACCLADE 3.08 (Maddison & Maddison, 1999) using the *Drosophila* mitochondrial translation table for the Cytochrome Oxidase sequences, and the universal code for *wingless*. Codon usage tables were calculated using Molecular Evolutionary Analysis (MEA) software (Moriyama, 2000).

Uncorrected pairwise sequence divergence was calculated using PAUP* 4.04a (Swofford, 2000). Nucleotide composition tables were calculated with MEA. Base composition heterogeneity was calculated as χ^2 distances using PAUP and MS Excel.

CO1, CO2 and *wingless* datasets were tested for incongruence with the Incongruence Length Difference test (Farris *et al.*, 1995) implemented in PAUP* as the 'partition homogeneity test'. All phylogenetic inferences were done using unweighted parsimony with PAUP* 4.04a using the heuristic search option (1000 replicates, steepest descent). Bootstrap analyses (Felsenstein, 1988; Hillis & Bull, 1993) were done using informative characters only, with 1000 replicates.

As the method of phylogenetic estimation can influence the degree of systematic error (Reed & Sperling, 1999), we assume that although there is no reason to reject the hypothesis of one partition history under a maximum parsimony (MP) analysis this is not automatically true when other methods of analysis are conducted. Obvious differences in evolution between the mitochondrial and nuclear DNA exist, such as evolutionary speed and base composition. When an explicit model of evolution is assumed, each partition must be evaluated under its own parameters of evolution (Yang, 1996b). Maximum Likelihood methods were used for the purpose of evaluating topologies found in the combined MP analysis (Yang, 1996a), and for evaluating the position of one particular species within a tree. The tRNA and the 1st, 2nd and 3rd codon positions of the mitochondrial and nuclear sequences were analysed separately. The –ln likelihoods were then added because partitions can be assumed to be independent (Yang, 1996a). We implemented a GTR model (Lanave *et al.*, 1984; substitution rates estimated), with base frequencies estimated empirically, using a Γ-shape (four rate categories, α estimated; Yang, 1996b), and invariable sites (I estimated). For testing the significance of differences in likelihood between topologies we could not use the Kishino–Hasegawa, (1989) test because of its parametric nature (M. Sanderson, pers. comm.). Therefore, we used the non-parametric Templeton test (Templeton, 1983) using the differences in likelihood

© 2004 The Royal Entomological Society,

per site. The Templeton test was carried out using the Wilcoxon sign rank test as implemented in SPSS 8.0 for Windows.

Evaluation of morphological characters

Ten characters that were used by Corbet (1941), Evans (1957) and Eliot (1963, 1972) to recognize groups and/or species within *Arhopala* s.l. were chosen for this study. These characters and their states were explicitly implemented as defined and interpreted by these authors. Characters and states are drawn in Fig. 1. These characters are:

1. tail at hind wing vein 2 absent (0); present (1),
2. venation as compared to the usual *Arhopala* type with hindwing cell being half hindwing length, long 9th forewing vein, and origin of forewing vein 5 closer to vein 6 than to 4 ('usual' *Arhopala* venation; 0); deviating from this condition by having: long hindwing cell (1); short hindwing cell (2); short 9th forewing vein (3); origin forewing vein 4, 5, and 6 equidistant at their origins (4),
3. underside forewing post discal spots in spaces 3 and 4 continuous at vein 4 (0); more or less completely dislocated (2); intermediate (1),
4. underside forewing post discal spots in spaces 4 compared to those in spaces 5 and 6; in line (0); slightly dislocated (1),
5. underside forewing post discal spot in space 10 absent (0); present (1),
6. underside hindwing post discal spots dislocated (0); more or less continuous at vein 2 (1),
7. underside hindwing post discal spots 5, 6 and 7 with centres in line, more or less macular and in echelon, yes (0); no (1),
8. upperside forewing in the males with the black border a thread of no more than 1 mm (0); very broad (sexual dimorphisms absent or almost absent) (2); intermediate (1).

Where possible we used Evans' and Eliot's interpretations of character states for the taxa involved (Evans, 1957; Eliot, 1963, 1972). However, as they did not make formal descriptions of species and species groups, the character states could only be extracted for a limited number of taxa for each of the characters. We examined each of the character states for all taxa using the specimens from Table 1, and also confirmed them using plates from D'Abrera (1986, 1990), Seki *et al.* (1991) and Corbet & Pendlebury (1992).

These characters were evaluated subsequently on a selected tree topology (see Results), using MACCLADE 3.08. Character state changes were treated as unordered. For evaluating non-random fit on the phylogeny we used a topology-dependent (based on the same selected tree) PTP test (Archie, 1989; Faith & Cranston, 1991) as implemented in PAUP*. Each character was evaluated individually (Wahlberg, 2001). The consistency index was calculated using PAUP*. As character numbers 3 and 8 each have three character states and a logical intermediate could be assumed (character state 2), these values were calculated under both ordered and unordered state changes.

Results

Sequences and alignment

We managed to obtain and align sequences from Cytochrome Oxidase 1, $tRNA^{Leu}$ and Cytochrome Oxidase 2 in four separate PCR reactions, homologous with the *Drosophila yakuba* (Clary & Wolstenholme, 1985) sites 1777–2170

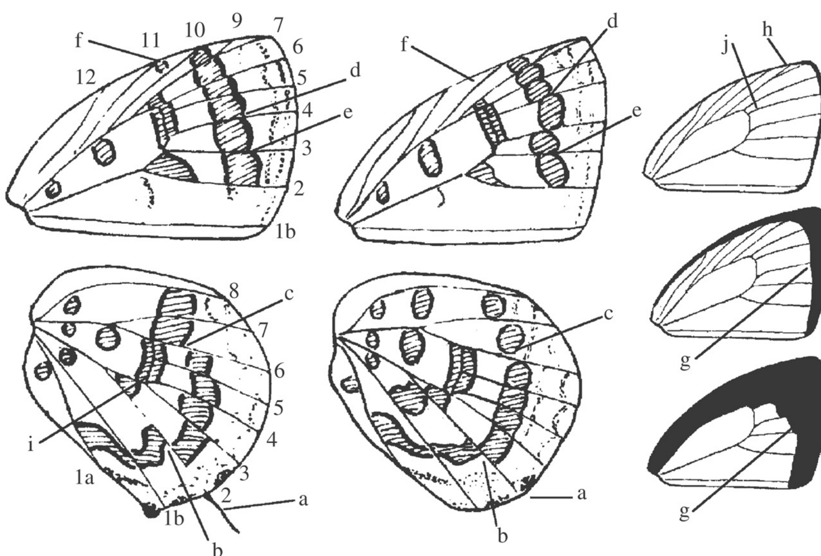


Fig. 1. Characters used for morphological evaluation: (a) tails present/absent; (b) underside hindwing post discal spots dislocated/continuous at vein 2; (c) post discal spots 5, 6 and 7 of hindwing out of line/in echelon; (d) position post discal spot 4 compared with 5 and 6 in line/dislocated; (e) forewing post discal spots 3 and 4 continuous at vein 4/completely dislocated; (f) forewing costal spot 10 present/absent; (g) male upperside forewing band a thread/intermediate/broad (sexual dimorphism absent). Also indicated is the predominant venation in *Arhopala*: long forewing vein 9 (h), length of hindwing cell about half the length of hindwing (i) and origin of forewing vein 5 closer to vein 6 than to vein 4 (j). Wing vein numbers are indicated; numbers of the spots are assigned by the vein below the spot. Redrawn and modified from Eliot (1972).

(GenBank accession numbers AY235861 to AY235900), sites 2229–2712 (numbers AY235955 to AY235996), sites 2804–3265 (numbers AY235812 to AY235852) and sites 3292–3769 (numbers AY235904 to AY235944). We failed to obtain sites 1777–2170 for the taxa *A. eumolphus* and *A. buddha*, 2804–3265 for *A. antimuta* from Malaysia and positions 3292–3769 for *A. wildei*.

No stop codons were found in the protein coding sequences. No obvious aberrations were found in the amino-acid sequences or $tRNA^{Leu}$ sequences when compared with other butterflies except for the following. *Semanga* appears to have a terminal stop codon (TAA) in CO1, where other butterflies have a T that has to be polyadynelated in the mRNA (Caterino & Sperling, 1999). *Semanga* also has an insertion of four bases (TATC) at the start of the $tRNA^{Leu}$ gene that probably will not significantly alter the secondary structure because it will only result in a larger internal loop. Combined, this resulted in a 7-bp insertion in *Semanga superba* at position 3010 (junction CO1–$tRNA^{Leu}$). As it is a unique insertion in one outgroup taxon only, it was excluded from further analyses.

At positions 3440–3470 there is a highly variable region where in several taxa complete codons are inserted or deleted, notably codons coding for asparagine. Despite the fact that indels corresponded with complete codons, the alignment for this region could not be unambiguously assessed and therefore has also been omitted from further analysis.

For the nuclear *wingless* gene, a 393-bp fragment was obtained for all specimens (GenBank accession numbers AY236007 to AY236048), homologous to sites 1381–1744 of the *Drosophila melanogaster wg* gene (Rijsewijk *et al.*, 1987). No indels were detected either among the forty-two specimens under consideration, or in comparing these with the *Heliconius erato wingless* sequence. An aligned dataset for all five gene fragments is deposited in Treebase (www.treebase.org; study accession number S888, matrix accession number M1439).

Nucleotide composition bias and substitution pattern

A striking feature when comparing the nuclear and mitochondrial DNA is the very unequal base composition of the latter (Table 3). This becomes most clear when we take the four-fold degenerate 3rd codon positions into consideration. For many species the AT content is over 95%, and in some species even over 98% (most extreme is *A. major*, with 98.8% in CO1, and 100% in CO2). Compared with other insects, *Arhopala* ranks among the most AT-biased groups. In CO2, AT content is on average at 78% (considering all codon positions), with a minimum of 77% for *A. kinabala*, and a maximum of 80% for *A. barami*. These numbers are even slightly higher than averages for Hymenoptera and other butterfly taxa (Jermiin & Crozier, 1994).

Base composition variation within *Arhopala* appears to be evenly distributed at all positions in the *wingless* gene, and in 1st and 2nd positions as well as tRNA in the mtDNA. There is, however, a marked heterogeneity in the AT content at 3rd codon positions (Table 4). Four-fold degenerate sites show some heterogeneity but this is not statistically significant. At the two-fold degenerate sites the AG sites do not show any heterogeneity, but the CT sites do. This could indicate that the variation in base composition is not due merely to stochastic effects in the mtDNA, and therefore has been subject to some kind of

Table 3. Base composition for each of the codon positions. Base composition for each of the codon positions, per gene. Values are taken as averages of all ingroup species. Numbers in parentheses are standard errors. For the 3rd codon positions a distinction was made between two-fold (partly) and four-fold (completely) degenerate sites in the dataset, with the percentage of occurrence in parentheses. Note that for two-fold degenerate sites A→G and C→T each add up to 100%.

Codon position	A	C	G	T
CO1				
1st	31.60 (0.58)	12.33 (0.53)	23.11 (0.66)	32.95 (0.61)
2nd	18.88 (0.48)	25.58 (0.52)	13.91 (0.23)	41.64 (0.21)
3rd	44.95 (1.56)	4.30 (1.49)	0.89 (0.58)	49.86 (2.42)
two-fold (56%)	98.71 (1.09)	8.08 (3.16)	1.29 (1.09)	91.92 (3.16)
four-fold (44%)	47.53 (3.60)	3.96 (1.84)	1.32 (1.09)	47.19 (4.30)
CO2				
1st	39.05 (0.92)	13.63 (0.52)	17.02 (1.28)	30.31 (0.91)
2nd	28.40 (0.45)	15.98 (0.55)	12.65 (1.02)	42.97 (1.93)
3rd	40.96 (2.06)	3.71 (1.67)	1.17 (0.57)	54.16 (2.60)
two-fold (70%)	97.42 (1.56)	6.72 (2.81)	2.58 (1.56)	93.28 (2.81)
four-fold (30%)	50.60 (6.24)	2.61 (2.14)	1.58 (1.41)	45.21 (6.08)
wingless				
1st	29.53 (0.62)	24.40 (0.48)	31.57 (0.44)	14.50 (0.44)
2nd	26.06 (0.41)	22.83 (0.35)	28.15 (0.39)	22.94 (0.29)
3rd	23.66 (1.91)	22.02 (3.11)	23.63 (1.73)	30.69 (3.35)
two-fold (44%)	30.27 (6.34)	53.09 (5.31)	69.73 (6.34)	46.91 (5.31)
four-fold (56%)	34.82 (3.18)	13.99 (3.53)	20.22 (2.36)	30.97 (3.58)

Table 4. Base composition heterogeneity. Base composition heterogeneity across the ingroup species, calculated per codon position, using a χ^2 test (not accounting for phylogenetic structure). For the 3rd codon position a distinction was made between four-fold and two-fold degenerate sites.

	mtDNA			*wingless*		
	χ^2	d.f.	*P*	χ^2	d.f.	*P*
all 1st	5.05	105	1.000	2.17	117	1.000
all 2nd	0.50	105	1.000	0.53	117	1.000
all 3rd	135.78	105	0.023*	41.38	117	1.000
Four-fold	118.98	105	0.166	37.51	117	1.000
Two-fold AG	22.40	35	0.951	13.95	39	1.000
Two-fold CT	59.73	35	0.006**	11.35	39	1.000
tRNA	2.26	105	1.000			

Degrees of freedom = d.f.; $*P < 0.05$, $**P < 0.01$.

selection. In this case it would mean an evenly strong selection against G, but considerable variation in the bias against C.

Even the amino-acid composition is affected by the strong AT bias. The ratio of AT-rich codons (those coding for amino acids F, Y, M, I, N and K) to those rich in GC (codons for amino acids G, A, R and P) is on average 2.4 in CO1 and 4.1 in CO2. These values are high even when compared with other insects (Foster & Hickey, 1999). By comparison, this ratio is 0.94 in *wingless*. The selective pressure on the amino acid composition would be most notable on the hydrophobic amino acids and could cause substantial degrees of parallel evolution even in the polypeptide chain (Foster & Hickey, 1999).

Observed genetic sequence divergence per site for the CO2 and *wingless* genes was plotted against the CO1 gene (Fig. 2) to visualize differences in substitution speed. CO2 appears to have a similar substitution pattern as CO1, although it is becoming slightly less diver-

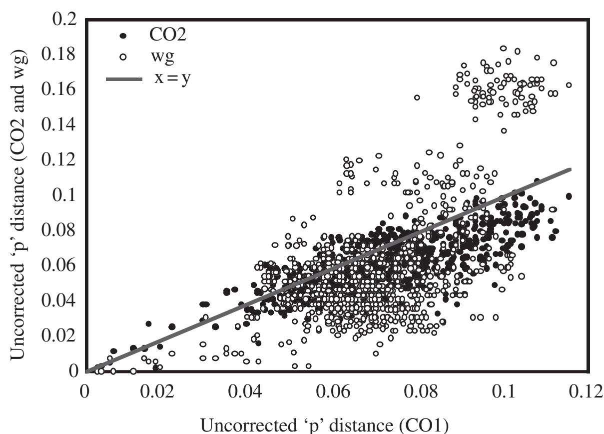


Fig. 2. The uncorrected pairwise distances of CO1 (horizontal axis) are plotted against the uncorrected pairwise distances of CO2 (●) and *wingless* (○). The line corresponds to $x = y$.

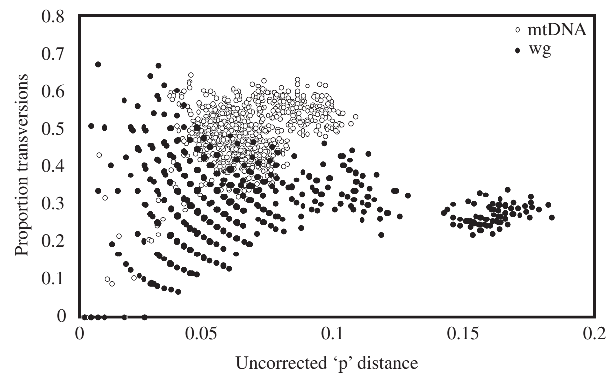


Fig. 3. Relationship between the uncorrected pairwise distances of CO1 and CO2 combined (○) and *wingless* (●), vs. proportion of transversions.

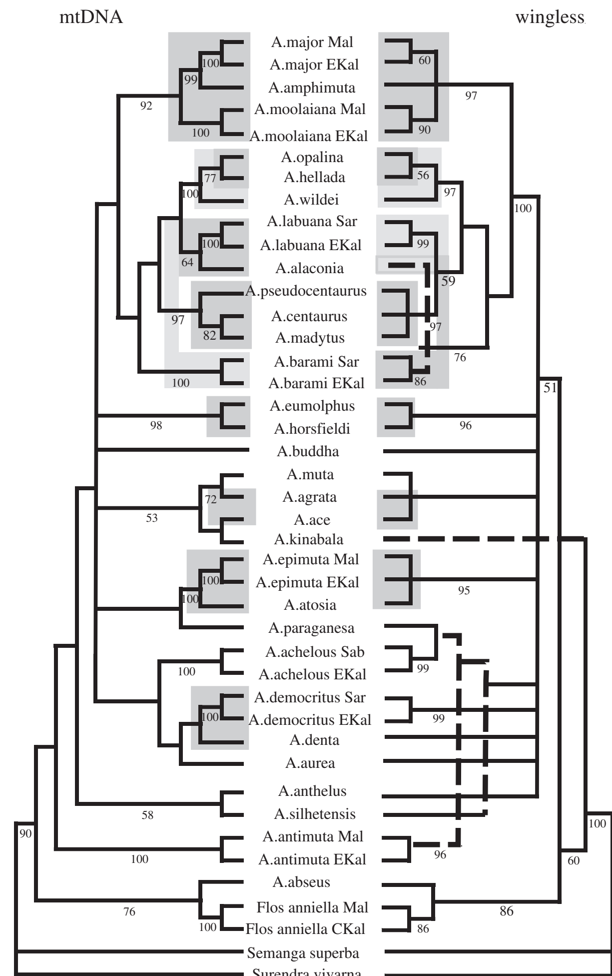


Fig. 4. Strict consensus trees of CO1 and 2 combined (A) and *wingless* (B). Figures below branches are bootstrap values. Shaded squares indicate clades that are congruent with previous taxonomic hypotheses according to Eliot (1963).

Table 5. Summary of tree statistics from the Maximum Parsimony analyses.

	Seq. length	Var. outg.	Var. ing.	Info. outg.	Info. ing.	No. steps. MPT	No. MPT	CI	RI	RC
mtDNA	1778	549	508	413	386	1866	9	0.38	0.48	0.18
wg	393	140	112	98	83	324	172	0.58	0.69	0.40
combined	2171	689	620	511	469	2215	58	0.401	0.50	0.20

Seq. Length = total sequence length; var. outg. = number of variable sites, outgroup taxa included; var. ing. = number of variable sites among ingroup taxa only; info. outg. = number of informative characters including outgroup taxa; info. ing. = number of informative characters among ingroup taxa only. No. steps MPT = number of steps in Most Parsimonious Tree(s); no. MPT = number of Most Parsimonious Trees; CI = Consistency Index; RI = Retention Index; RC = Rescaled Consistency index.

gent at the most dissimilar sequences. For *wingless*, the initial rate of evolution appears to be in the order of 30–50% of CO1, but then appears to 'overtake' the mitochondrial genes in time (here represented as pairwise divergence).

The proportion of transversions compared with the total number of observed pairwise differences levels off in distantly related species around 50% for mtDNA, and around 30% for *wingless* (Fig. 3). The transition–transversion ratio should best be estimated between very closely related

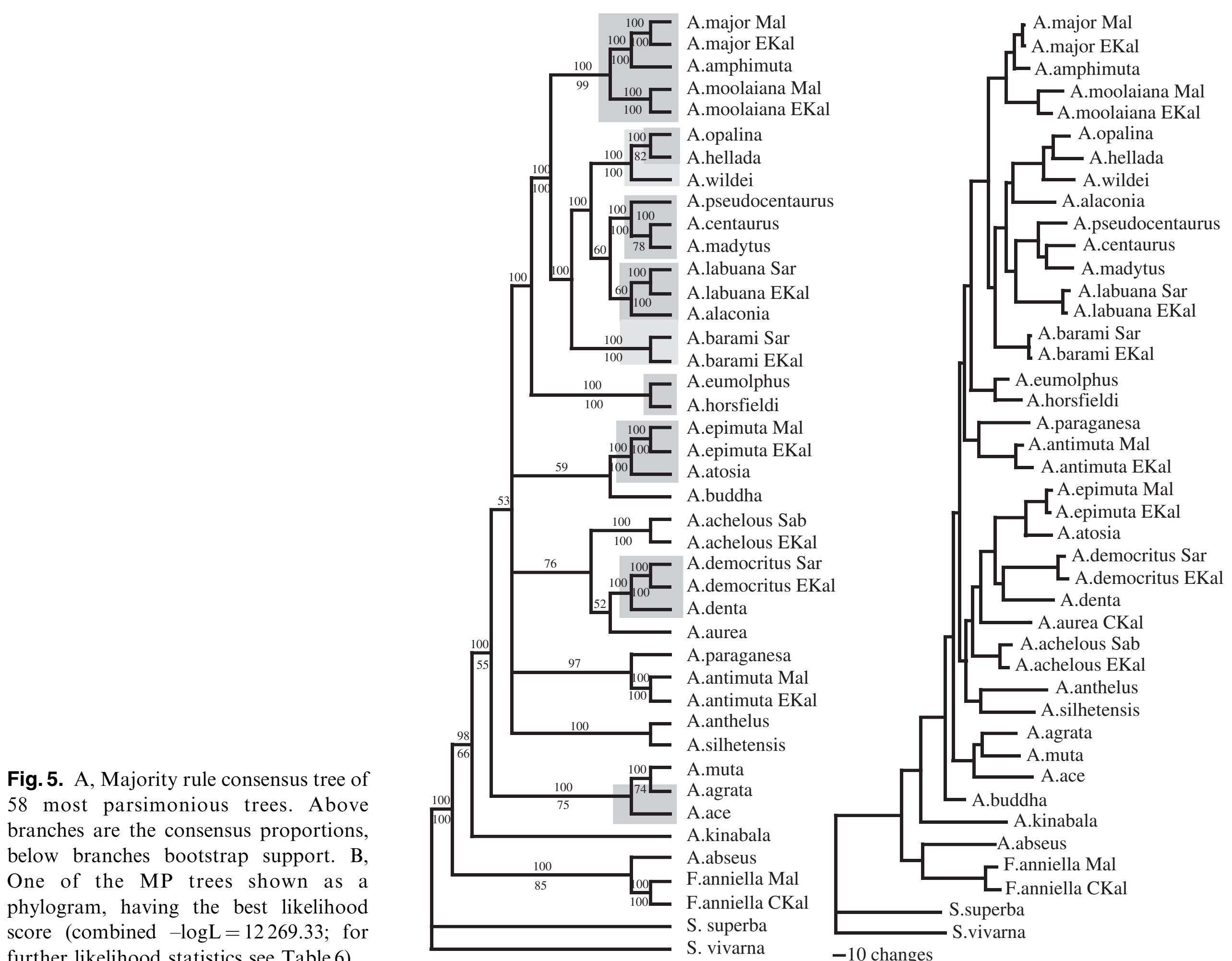


Fig. 5. A, Majority rule consensus tree of 58 most parsimonious trees. Above branches are the consensus proportions, below branches bootstrap support. B, One of the MP trees shown as a phylogram, having the best likelihood score (combined –logL = 12 269.33; for further likelihood statistics see Table 6).

Table 6. Summary of tree statistics from the Maximum Likelihood evaluation.

	–logL	AC	AG	AT	CG	CT	Γ-shape	p. inv.
1st CO	2093.1	2.29	19.12	2.4	0	80.44	0.78	0.72
2nd CO	925.0	0	22.66	5.38	4.39	10.09	0.75	0.94
3rd CO	7039.8	0.87	33.28	0.06	0	27.53	0.3	0.00
tRNA	164.0	1.9×10^{8}	1.0×10^{10}	0	0	4.0×10^{8}	0.14	0.65
1st *wg*	332.8	0.94	1.2	1.6	0.1	3.12	0.37	0.40
2nd *wg*	238.1	9.4×10^{6}	3.0×10^{7}	0	8.0×10^{6}	4.0×10^{7}	infinity	0.67
3rd *wg*	1476.6	2.33	14.21	5.12	2.14	21.11	2.85	0.03

Values are taken from the 'most likely' of all the trees found in the MP analysis (see Fig. 4B). Substitution rates are relative to the GT (= 1) transversion; AC indicates inferred substitution rate from A to C or C to A (symmetrical) etc. Rates are not comparable between (codon) positions. –logL = negative log Likelihood value; Γ-shape is inferred value for this parameter; p. invar. is inferred proportion of invariable sites.

species in order to observe the instantaneous substitution pattern, but this poses problems as the small number of nucleotide changes causes a large variation in the estimation (Yang & Yoder, 1999). The instantaneous transition–transversion ratio for the mtDNA appears to be higher, possibly around 2, as is the case with *wingless*.

Parsimony analysis

We reconstructed separate phylogenies based on mtDNA and *wingless*. Results are shown and compared in Fig. 4 (for tree statistics see Table 5). All specimens belonging to the same species are grouped together. There is agreement between mtDNA and *wingless* with regard to those parts of the phylogeny that are well supported. Particularly important is that one clade, consisting of the *amphimuta* subgroup, the *centaurus* group, *A. barami* and several other species, is clearly distinguishable in both phylogenies. There is also good agreement about the relationships within this clade. Furthermore, the sistergroup relationship between *Flos anniella* and *A. abseus*, *A. epimuta* and *A. atosia*, and *A. eumolphus* and *A. horsfieldi* is confirmed, as well as the close relationship of *A. agrata* and *A. ace*. The relationship of the latter species with *A. muta* is well supported in the molecular phylogeny, but has not previously been hypothesized. Additionally, there appears to be agreement on the (lack of) resolution at the base of the phylogenies. Disagreement is especially notable on the status of *A. kinabala*. In the *wingless* tree it is the most basal taxon in the *Arhopala–Flos* group, whereas it groups well within *Arhopala* in the CO tree.

The partition homogeneity test (Farris *et al.*, 1995) did not reveal any reason to reject the hypothesis that CO1, CO2 and *wingless* datasets represent the same process partition ($P = 0.54$ for combined mtDNA vs. *wingless*, and $P = 0.47$ for CO1, CO2 and *wingless* all separately; at 500 replicates). A combined phylogenetic analysis revealed the majority rule consensus tree shown in Fig. 5A. The result of the combined analysis does give slightly higher bootstrap values but the values remain low on the deeper nodes. The position of *A. kinabala* is resolved, but with a low bootstrap value of 55, at the base of a monophyletic group to which *A. abseus* and *F. anniella* form the sistergroup.

Maximum likelihood evaluation

Further evaluation of mtDNA and *wingless* under Maximum Likelihood (ML) yielded one best tree with the highest likelihood score (topology shown in Fig. 5B, evaluation statistics can be found in Table 6). The Templeton test showed that forty-four of fifty-eight MP trees were statistically ($P < 0.05$) different from the best tree. Interestingly, a consensus of the remaining fourteen trees did not resolve any better than the consensus of all fifty-eight MP trees, indicating that further distinction between topologies could not be made.

The position of *A. kinabala* is among the most conspicuous anomalies between the trees. Yet this particular taxon is of interest to the systematics of *Arhopala* because it features many unusual characters compared with most other *Arhopala* species, such as the absence of sexual dimorphism, the strong maculate ventral surface patterns and hind wing shape. Therefore, we further investigated its position in the tree (see Discussion below). On the MP tree with the highest likelihood score (see Fig. 5B), we investigated all seventy-nine possible positions of *A. kinabala*. The procedure was two-fold because of computational limitations. All seventy-nine possible trees were evaluated by a parsimony-based Templeton test (Templeton, 1983); those trees significantly different ($P < 0.05$) from the best tree (which was the original topology) were discarded. The seven remaining trees were then evaluated under an ML model as described above.

When nuclear and mitochondrial data are evaluated together, the preferred placing of *A. kinabala* is as in the tree of Fig. 5, although there is no statistical difference (using the ML-based Templeton test) with it being placed at the very basis ($P = 0.94$) or as a sistergroup of *Flos* and *A. abseus* ($P = 0.06$). *Wingless* alone also marks this topology as most likely, preferred statistically over every other topology. A parsimony-based analysis based on mtDNA, however, would place *A. kinabala* as a sistergroup of the clade including the *agelastus*, *amphimuta* and *centaurus*

groups, different ($P < 0.05$) from any of the basal positions. By contrast, an ML evaluation on all positions of the mtDNA combined (GTR-I-Γ with seven rate categories and empirical base frequencies; parameters estimated) favours the basal position of Fig. 5 (other topologies not significantly different using a KH-test). This indicates that in particular in the mtDNA there is substantial ambiguous phylogenetic signal with regard to the position of *A. kinabala*.

Evaluation of morphological characters

The character states for each of the eight morphological characters examined can be read from the boxes at the tips of the phylogenies in Fig. 6. The topology of the phylogeny is identical to Fig. 5. The putative ancestral character states are inferred by MACCLADE 3.08. In several instances the character states could not be determined as the characters were absent (missing data, indicated with a question mark). In other cases we could not unambiguously decide between character states. These ambiguities are shown as 'polymorphic' character states in Fig. 6.

The phylogenetic content of the characters in the form of a T-PTP test and CI scores are also shown in Fig. 6. In the case of characters 3 and 8, there are two scores because these were evaluated both ordered as well as unordered. We observed significant ($P < 0.05$) T-PTP scores for four out of eight characters: the venation (Fig. 6B), the relative position of the post discal spots in spaces 3 and 4 of the forewing (Fig. 6C), the position of forewing space 4 compared with 5 and 6 (Fig. 6D), and the dislocation of the hind wing post discal spots at vein 2. Of these four characters, all but one (venation; $CI = 1.0$) showed considerable homoplasy ($CI < 0.2$). Regardless, venation is not an informative character for the higher phylogeny of *Arhopala*, as the character states are not shared among groups of species.

Discussion

Molecular systematics and Arhopala *phylogeny*

In both the separate and the combined analysis of the mtDNA and *wingless* sequence data, there is considerable congruence in topology, indicating that certain clades are well differentiated phylogenetically (Figs 4, 5). In a number of cases, the species groups as recognized by Eliot (1963) are validated. The *amphimuta* subgroup and the *centaurus* group are monophyletic. *Arhopala eumolphus* and *A. horsfieldi* (*eumolphus* group), and *A. epimuta* and *A. atosia* (*epimuta* group) clearly are very closely related. The representatives of the *camdeo* group (*A. opalina* and *A. hellada*) group together, although they are not as closely related as the former species. *Arhopala barami*, *A. labuana* and *A. alaconia* are assigned to the *agelastus* species group by Eliot (1963), and although their relationships are not clearly resolved they apparently do form a group of closely related species. *Arhopala denta* seems to group with *A. democritus*, although this is not a very well supported relationship. Evans (1957) grouped the latter species in the *democritus* subgroup of the *democritus* group, whereas Eliot did not recognize these two species as belonging to a single group.

In other cases the groups are not validated. For example, *A. muta* and *A. antimuta* are part of the *amphimuta* group (*muta* and *perimuta* subgroups, respectively) according to Eliot (1963), and therefore should be closely related to each other and to the *amphimuta* subgroup. However, the three subgroups do not form a monophyletic group according to the molecular hypothesis. Instead, *A. muta* appears to be very closely related to *A. agrata* and *A. ace*, members of the *cleander* group. In Eliot's (1972) view the *cleander* group comprises one group with large and one group with small species. At least one member of the 'large' *cleander* group (*A. silhetensis*) does not seem closely related to these 'small' *cleander* species, although relationships remain unclear. The same problem is presented by *A. anthelus* and *A. achelous*, which are grouped together by Corbet (1941), Evans (1957) and Eliot (1963) but fail to group together in the molecular phylogeny.

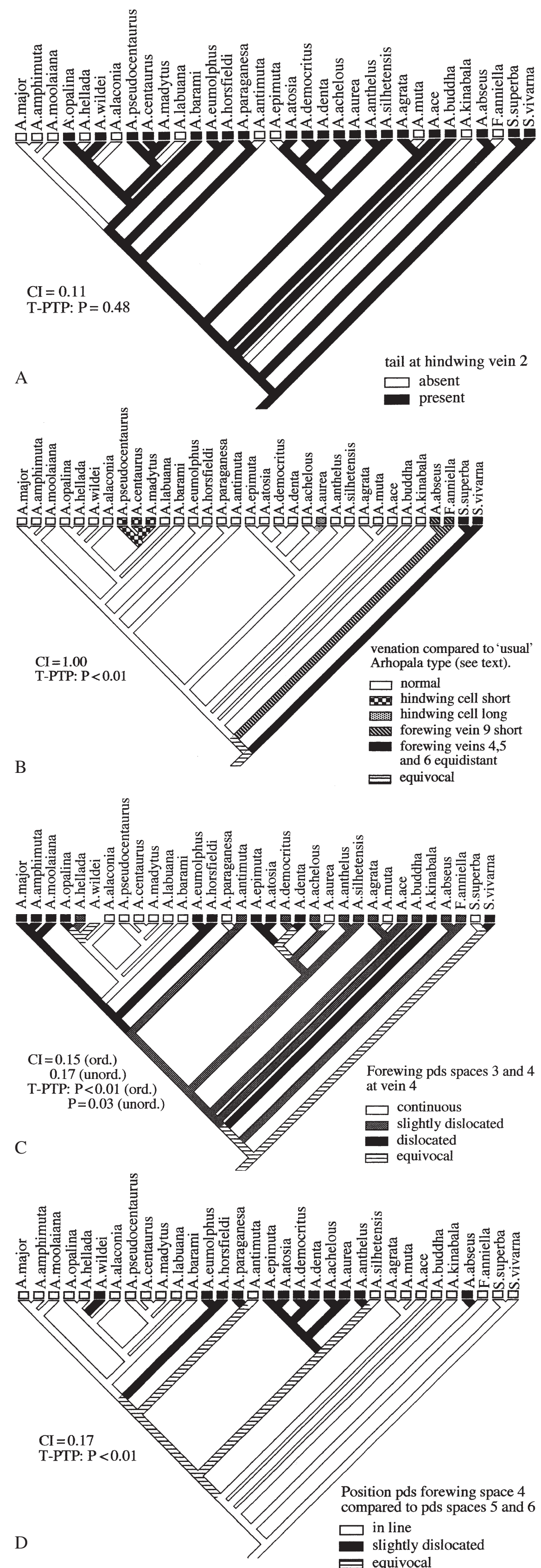
Eliot (1963) argued that the genera *Panchala* and *Aurea*, erected by Evans (1957) on the basis of aberrant female ovipositor shape and venation, respectively, did not deserve generic status and re-assigned them to *Arhopala* (species groups *ganesa* and *aurea*). From our molecular phylogeny it would seem that Eliot was correct in his view, as the species *A. paraganesa* and *A. aurea* reliably cluster within the major *Arhopala* group. Artefactual clustering with other *Arhopala* species due to long branch attraction is unlikely as both the mtDNA and the *wingless* sequence data show congruent patterns for these species.

In terms of congruence with previous morphologically based studies, it is striking to note that the relationships between species groups were especially difficult to ascertain (Eliot, 1963). Although most of the groups considered by Eliot appear to be closely related units in the molecular trees, the inter-relationships in many cases remain uncertain. By contrast, the *amphimuta* subgroup, the *centaurus* group, the *camdeo* group, *A. wildei*, *A. barami*, *A. labuana* and *A. alaconia* (the latter three belonging to the *agelastus* group; Eliot, 1963) together form a well-supported monophyletic group. Within this clade the *amphimuta* group appears to be sister group to the others combined, where *A. barami* comes out at the base. This relationship has never been proposed based on morphology, because there are no characters that explicitly support it.

Arhopala sensu Eliot is not monophyletic; both in the mitochondrial data and the nuclear data it becomes apparent that *A. abseus* is a sistergroup to *Flos* (as exemplified by *F. anniella*). This is not surprising because *A. abseus* shares the short 9th forewing vein with *Flos*. The spiny uncus defining *Flos* appears to be an autapomorphous character of this group, whereas the venation is a good indicator for the monophyly of *Flos* together with *A. abseus* (Fig. 6B).

Arhopala kinabala is most likely a basal taxon of the *Arhopala* group of species studied here (including *Flos*), when *wingless* only is considered. Alternatively, *Flos* and *A. abseus* constitute a sistergroup to all the other *Arhopala* species under consideration, with *A. kinabala* being the most basal taxon of the latter group. Interestingly, *A. kinabala* is atypical in having no sexual dimorphism, very distinct brown markings and a slightly irregular hindwing margin. These characters are shared with some other genera in the Arhopalina: *Thaduka*, *Mahathala* and *Apporasa*.

The Australian species do not form a monophyletic group, and considering the phylogeny of the other *Arhopala* species, it seems most likely that there have been two independent dispersals from the Oriental region to the Papuan region. However, a few species of the *centaurus* group also occur in the Oriental region. In this respect it would be very useful to know the phylogenetic position of the *Arhopala s. str.* (*sensu* Evans, 1957) group.

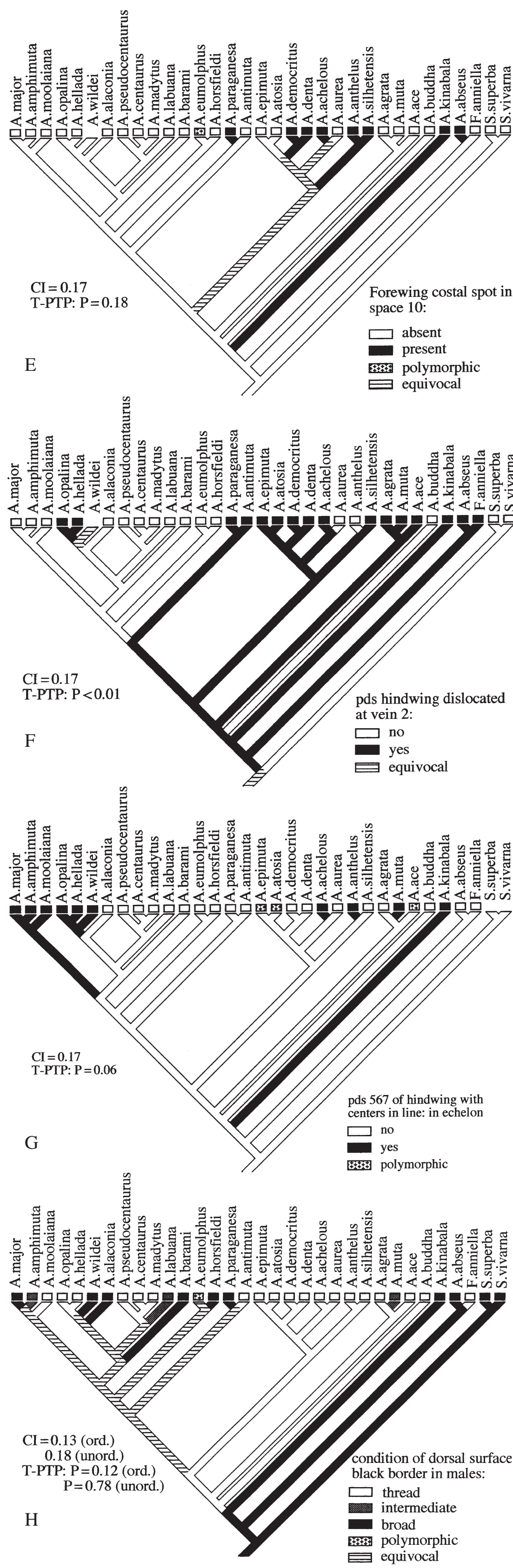


Substitution pattern

When comparing the separate analyses for mtDNA and *wingless,* we notice considerable incongruence in tree topology in particular at the deeper nodes of the trees. Because this largely concerns those parts of the phylogeny that are poorly resolved, we assume that this is due to an overall lack of phylogenetic information in both datasets at this particular level. This is further supported by the character incongruence test, which shows that in general there is not sufficient character incongruence to make the two datasets incompatible. This means that although there might be very little information in favour of a certain hypothesis, there is also little to disprove it.

The AT bias is a well-known and well-studied phenomenon, but there is still much debate about its cause (Rand, 1994; Francino & Ochman, 1997; Galtier & Lobry, 1997). The phenomenon appears to be almost universal throughout the animal kingdom, and is most pronounced among neopteran insects (Jermiin *et al.*, 1994; Jermiin & Crozier, 1994; Lunt *et al.*, 1996), although some exceptions exist (Wirth *et al.*, 1999). Although the level of AT bias in *Arhopala* is possibly higher than in other cases, it is not strikingly unusual. Yet, it does affect the conclusions drawn in this paper, in particular with regard to those drawn from the mtDNA. Particularly affected are assessments about substitution pattern and genetic divergence on the one hand, and the reliability of the phylogenetic inference on the other.

In very strongly biased genomes the number of character states at silent sites is effectively reduced from four to two for four-fold degenerate sites, and to one for two-fold degenerate sites. This results in a smaller effective proportion of neutral states in mtDNA than in nuclear DNA that has no strong compositional bias (cf. Lyons-Weiler & Hoelzer, 1999). This effect may be further enhanced by the relatively large proportion of codons that are two- rather than four-fold degenerate in insect mitochondrial genomes (Jermiin & Crozier, 1994; in *Arhopala* most marked in CO2



where 70% of codons are two-fold degenerate). The pattern of a high instantaneous rate of evolution in mtDNA compared with nuclear DNA followed by an apparent slowing down between more distantly related taxa (as observed in Fig. 2; see also Brower & DeSalle, 1998) is therefore a result of the skewed base composition. The same is true for the observed transition–transversion ratios (Fig. 3). The transition bias in *wingless* is maintained in comparisons of distantly and closely related taxa alike (although the variation is very large between closely related species). In the mtDNA, however, an initial transition bias appears to be disappearing over time, which is indicative of saturation (e.g. Gomez-Zurita *et al*., 2000).

The saturation effect may have consequences for the phylogenetic inference. For parsimony analysis it can result in underestimation of the number of rare-to-common changes (Eyre-Walker, 1998), leading to underestimation of the true genetic divergence and homoplasy. In theory, an ML analysis should be able to deal with compositional bias (Yang, 1996b), at least to a certain extent. However, when the assumptions of the model used are not met, the subsequent analysis becomes flawed. From Fig. 3 we can conclude that the observed transition–transversion ratio is not stationary in time, an observation also made in other studies (e.g. Brown *et al*., 1994). The estimated substitution rates are given in Table 6. Transitions are generally more likely than transversions. For *wingless* we note that there are no substitution types that are extremely unlikely compared with others. In the mtDNA, however, in particular mutations to G are deemed extremely unlikely. The estimates in general do not agree well with the observed instantaneous substitution pattern.

A further problem is the heterogeneous base composition present in the mitochondrial CO 3rd codon positions when the ML model used assumes a stationary base composition (as is the case in PAUP*). This may result in an incorrect phylogenetic inference (Lyons-Weiler & Hoelzer, 1999), especially at deeper nodes. Long branch attraction occurs even sooner with compositional bias (Eyre-Walker, 1998), and may well be aggravated further because of the compositional heterogeneity. An example may be *A. kinabala*, whose basal placement is difficult to confirm in the mtDNA analyses. This possibly also relates to the fact that this taxon has one of the highest GC contents.

A combined analysis of mtDNA and *wingless* shows the same ambiguities at basal nodes (Fig. 5); interestingly, the incongruence between the *wingless* and mtDNA trees (as observed in Fig. 4) may actually represent an agreement concerning the failure of resolving this part of the phylogeny. Although this could be due to the saturation effects

Fig. 6. Inferred morphological character states (indicated by boxes at tip of the tree) and changes based on the topology of Fig. 5B. The CI and T-PTP scores of each character are shown. Missing data resulted in the absence of a box at the tip of the tree. The designation pds stands for post discal spot.

mentioned above, it is conceivable that the actual reason for the lack of support at exactly this particular level is more complex. The low phylogenetic signal in the extensive genetic variation could be a reflection of *Arhopala* history, such as instances of accelerated speciation rates, or aspects of demography (Guyer & Slowinski, 1993; Pagel, 1998).

Evaluation of morphological characters on the molecular phylogeny

Cladistic methods become seriously flawed when characters and character states are improperly called homologous, and when homoplasy is common. Although cladistic methodology and terminology was not formally used by Corbet (1941, 1946), Evans (1957) and Eliot (1963, 1972), they nevertheless hinted that both these problems play a role in establishing a stable higher taxonomy of *Arhopala* s.l. In order to re-evaluate a selection of the characters used by these workers in a cladistic context, we plotted the character states on the independently derived phylogeny (Fig. 6).

The absence of tails is a good example of a character state with several independent origins during the history of *Arhopala* (Fig. 6A). We can deduce from the tree that the ancestral state probably had a tail present at vein 2 of the hindwing. At several instances in evolution the tails became lost subsequently. There is no evidence that once a tail is lost, it can re-appear. Apparently, the loss of a tail will happen readily, and probably even more often than indicated in Fig. 6A. For example, most species of *Flos* have tails, but *F. anniella*, the only species of *Flos* included in this study, does not. It is clear that the absence of a tail is not a very reliable character state in phylogeny reconstruction of *Arhopala*, as is confirmed by both CI and T-PTP scores (Fig. 6A). Therefore, Eliot (1963) was correct in not sharing Evans' (1957) assignment of *A. epimuta* to a different group as *A. atosia* on the basis of the presence of tails, as these two species are indeed closely related.

However, in other instances the tails were apparently important to Eliot as well, for instance in assigning *A. antimuta* and *A. muta* to the *amphimuta* group. Other characters were important for this assignment as well, such as the post discal spots in hindwing spaces 5, 6 and 7 with their centres being in line, and macular (Fig. 6G). *Arhopala muta* has these spots in line, but they are not as macular as those of the *amphimuta* subgroup. Likewise, *A. antimuta* has bandlike spots in these spaces and they are very much out of line. Here we clearly see how Eliot and Evans were struggling to find a set of characters that congruently could define a group. Based on the general appearance and the shape of the post discal band of the forewing, *A. antimuta* and *A. muta* appeared related to the '*amphimuta* types', yet this proves to be incorrect. A careful analysis of the macular pattern shows that the post discal spots are usually more rounded when they are lined up in one row. In *A. antimuta* and *A. muta* this correlation is absent, indicating the independent origin of this character state.

Characters deduced from the maculation pattern appear to be rather prone to homoplasy, as is confirmed by their consistency indices, which are all less than 0.2. Yet, a phylogenetic component is present as is indicated by a number of the T-PTP scores showing significant ($P < 0.05$) values. For example, an important character in identification and grouping is whether the post discal band on the hindwing underside is dislocated or not (Figs 1, 6F). Again, as in the former characters, the distribution of the dislocated hindwing band appears to be constrained phylogenetically, and seems to support the grouping of the *amphimuta* subgroup with the *agelastus* and *centaurus* groups, and in addition also the sistergroup relation of this clade with the *eumolphus* group. In spite of its clear phylogenetic constraints, this character is not always unambiguously interpreted. For example, Eliot (1972) disputed the claims of Evans (1957) that in *A. alica* the band was dislocated; Eliot found that several specimens showed what he called 'partly dislocated' bands.

One other important character for placing *A. antimuta* and *A. muta* in the *amphimuta* group concerns the position of the spot in space 4 compared with the band formed by spaces 5 and 6 on the forewing underside. The continuous state seems widespread. Interestingly, the dislocated state is present in a clade that is very weakly supported by the molecular data (Fig. 6D). By contrast, the relative positions of the post discal spots in spaces 3 and 4 at vein 4 of the forewing (Fig. 6C) were not regarded at all as being valuable for establishing groups within *Arhopala*. In part this is due to the difficulty in assigning the character state; there appears to be various ways in which the post discal spots can become dislocated. This is why an 'intermediate' state is added. In addition, there appears to be some interspecific variation in position and form of post discal spots 3 and 4, but we have not evaluated sufficient specimens to confirm this. Nevertheless, Fig. 6C shows that this character is uncorrelated with the former (Fig. 6D), but still holds considerable phylogenetic signal.

A costal spot in space 10 of the underside of the forewing (Fig. 6E) is present in only a small number of groups; it is shared by *A. anthelus* and *A. silhetensis*, as well as by *A. democritus* and *A. denta*. The relationships among these taxa have largely been ignored, although Evans (1957) placed *A. democritus* and *A. denta* together, which seems to be confirmed through this shared character state. However, no significant T-PTP score can be assigned to this character, possibly because the inter-relationships between the species sharing the presence of the costal spot are very weakly supported and therefore possibly incorrectly inferred.

A character important for species identification but less for grouping is width of the black marginal band of the dorsal side of the forewing in males. In most species this band is extremely narrow, not more than a black thread. In other species it becomes more pronounced (see also Fig. 1). A phylogenetic signal may exist in this character (Fig. 6H) that has previously been ignored. Several basal groups show no sexual dimorphism and retain the broad irregular bands in the male phenotype (although this is not true for *Flos*).

For all the other *Arhopala* species, only those of the *amphimuta*, *wildei* and *agelastus* groups show broader black bands, but here this condition often differs still from the female phenotype because the bands are better defined. Furthermore, no phylogenetic signal appears to be present according to the T-PTP test result, even when the character states are 'ordered'. This is due to the high variability of this character in those groups; some species have narrow bands, others broad ones. Possibly in these groups females recognize conspecific males on the basis of shape and width of this black margin. When interpreted as a mate recognition feature, variability in this character could (within the *Arhopala* group) be unique to one particular clade.

Characters that are derived from venation appear to be phylogenetically informative (T-PTP: $P < 0.01$) and perfectly parsimonious ($\mathrm{CI} = 1.0$), indicating that venation creates very stable character states. However, these characters cannot be used to define larger groups within *Arhopala* (Fig. 6B). The short hindwing cells of the *centaurus* group and the long hindwing cell of the *aurea* group were used to define these groups and appear autapomorphies here. However, the *hercules* group from the Papuan region also has a short hindwing cell. Possibly, this group is closely related to the predominantly Papuan *centaurus* group.

The evaluation of these eight morphological characters demonstrates the problems past systematists had when searching for clues concerning higher phylogeny of the genus *Arhopala* s.l. The macular patterns, which have been the most important source of character states, show in a number of cases phylogenetically constrained conditions that nevertheless have a very low consistency index and seem therefore very homoplasious. Wing patterns are often subject to selection pressure, which could explain low phylogenetic inertia. In *Arhopala* it is not clear what external factors may determine the faint macular patterns normally present.

The genetic background of the wing pattern character states in *Arhopala* is currently unclear, but it seems likely that homoplasy sometimes is due to non-homologous character state changes. For instance, a post discal band can become dislocated because spots move distally, but also because spots become more rounded and hence more defined. In addition, spots may move because wing shape changes, which in turn may cause covariation in other characters (Monteiro *et al.*, 1997). In general, changes in subsets of the wing patterns, such as position, size and shape of spots, are more difficult to achieve than changes on the whole pattern (Brakefield & French, 1993). In summary, the extensive homoplasy, the difficulty with homologizing characters and the problem of interdependence of characters all limit the use of wing pattern data in cladistic methods.

Conclusions

The basal position of the apparent sister taxa *Flos* and *A. abseus*, the basal position of *A. kinabala*, and the largely unresolved relationships between many species groups compared with the very well resolved clade consisting of the *amphimuta* subgroup, *camdeo*, *centaurus* and *agelastus* groups all give important new insights in to the systematics of this large group of butterflies. Although many groupings made by Evans and Eliot are confirmed, some incongruencies occur that in part can be attributed to misinterpretation of morphological characters. There are clearly historical constraints in morphology, although many characters show considerable homoplasy, which limits their use in phylogenetic methods.

The molecular markers used in this study do not appear to evolve very differently in *Arhopala* when compared with similar studies. Their usefulness is apparent for resolving the relationships within certain parts of the *Arhopala* phylogeny. The relationships between many groups, however, remain obscure at present. Saturation effects clearly play a role, and in the mtDNA this effect is enhanced because of relatively fast evolutionary rates combined with a very marked AT bias. It seems probable, however, that relatively rapid speciation rates contributed as well to the loss of phylogenetic information; this aspect needs to be explored in future studies.

Acknowledgements

We thank Andre Mignault for technical assistance. Man-Wah Tan collected many specimens and assisted in identifications. Danny Cleary provided specimens and together with Ferry Slik assisted in the fieldwork. Thanks to William Piel and Antonia Monteiro for discussion and reviewing the manuscript. The research was funded through NWO Priority Program Biodiversity in Disturbed Ecosystems (grant 895-10-002P) and further supported through grants from NWO (grant SIR14-2218), Jan Joost ter Pelkwijk Fonds en Uyttenboogaart-Eliasen Stichting (all to the first author).

References

Archie, J.W. (1989) A randomization test for phylogenetic information in systematic data. *Systematic Zoology*, **38**, 239–252.

Brakefield, P.M. & French, V. (1993) Butterfly wing patterns: developmental mechanisms and evolutionary change. *Acta Biotheoretica*, **41**, 447–468.

Bridges, C.A. (1988) Catalogue of Lycaenidae and Riodinidae (Lepidoptera: Rhopalocera). Published by C. A. Bridges.

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

Brower, A.V.Z. (2000) Phylogenetic relationships among the Nymphalidae (Lepidoptera) inferred from partial sequences of the *wingless* gene. *Proceedings of the Royal Society of London B*, **267**, 1201–1211.

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. *Insect Molecular Biology*, **7**, 73–82.

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. *Proceedings of the Royal Society of London B*, **264**, 969–977.

Brown, J.M., Pellmyr, O., Thompson, J.N. & Harrison, R.G. (1994) Phylogeny of *Greya* (Lepidoptera: Prodoxidae), based on nucleotide sequence variation in mitochondrial Cytochrome Oxidase I and II: congruence with morphological data. *Molecular Biology and Evolution*, **11**, 128–141.

Bush, M. (1994) Amazonian speciation: a necessarily complex model. *Journal of Biogeography*, **21**, 5–17.

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: Papilionoidae). *Molecular Biology and Evolution*, **17**, 684–696.

Caterino, M.S. & Sperling, F.A.H. (1999) *Papilio* phylogeny based on Mitochondrial Cytochrome Oxidase I and II genes. *Molecular Phylogenetics and Evolution*, **11**, 122–137.

Clary, D.O. & Wolstenholme, D.R. (1985) The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evolution*, **22**, 252–271.

Cobb, M., Huet, M., Lachaise, D. & Veuille, M. (2000) Fragmented forests, evolving flies: molecular variation in African populations. *Molecular Ecology*, **9**, 1591–1597.

Corbet, S.A. (1941) A key to the Indo-Malayan species of *Arhopala* Boisduval (Lepidoptera: Lycaenidae). *Proceedings of the Royal Entomological Society of London (B)*, **10**, 149–170.

Corbet, S.A. (1946) Observations on the Indo-Malayan species of the genus *Arhopala* Boisduval (Lepidoptera: Lycaenidae). *Transactions of the Royal Entomological Society of London*, **96**, 73–88.

Corbet, S.A. & Pendlebury, H.M. (1992) *The Butterflies of the Malay Peninsula*, 4th revised edn. Malayan Nature Society, Kuala Lumpur.

D'Abrera, B. (1986) *Butterflies of the Oriental Region, Part III Lycaenidae*. Hill House, Melbourne.

D'Abrera, B. (1990) *Butterflies of the Australian Region*, 2nd edn. Hill House, Melbourne.

Eliot, J.N. (1963) A key to the Malayan species of *Arhopala*. *Malayan Nature Journal*, **17**, 188–217.

Eliot, J.N. (1972) Some *Arhopala* from Borneo, with a revision of the *Arhopala cleander* group. (Lepidoptera: Lycaenidae). *Journal of Natural History*, **6**, 1–15.

Eliot, J.N. (1973) The higher classification of the Lycaenidae (Lepidoptera): a tentative arrangement. *Bulletin of the British Museum (Natural History), Entomology*, **28**, 373–505.

Evans, W.H. (1957) A revision of the *Arhopala* group of Oriental Lycaenidae (Lepidoptera: Rhopalocera). *Bulletin of the British Museum (Natural History), Entomology*, **5**, 85–141.

Eyre-Walker, A. (1998) Problems with parsimony in sequences of biased base composition. *Journal of Molecular Evolution*, **47**, 686–690.

Faith, D.P. & Cranston, P.S. (1991) Could a cladogram this short have arisen by chance alone?: on permutation tests for cladistic structure. *Cladistics*, **7**, 1–28.

Farris, J.S., Källersjö, M., Kluge, A.G. & Bult, C. (1995) Testing significance of incongruence. *Cladistics*, **10**, 315–319.

Felsenstein, J. (1988) Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics*, **22**, 521–565.

Fiedler, K. (1991) Systematic, evolutionary, and ecological implications of myrmecophily within the Lycaenidae (Insecta: Lepidoptera: Papilionoidea). *Bonner Zoologische Monographien*, **31**.

Foster, P.G. & Hickey, D.A. (1999) Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *Journal of Molecular Evolution*, **48**, 284–290.

Francino, M.P. & Ochman, H. (1997) Strand asymmetries in DNA evolution. *Trends in Genetics*, **13**, 240–245.

Galtier, N. & Lobry, J.R. (1997) Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperatures in prokaryotes. *Journal of Molecular Evolution*, **44**, 632–636.

Gomez-Zurita, J., Juan, C. & Petitpierre, E. (2000) The evolutionary history of the genus *Timarcha* (Coleoptera, Chrysomelidae) inferred from mitochondrial COII gene and partial 16S rDNA sequences. *Molecular Phylogenetics and Evolution*, **14**, 304–317.

Guyer, C. & Slowinski, J.B. (1993) Adaptive radiations and the topology of large phylogenies. *Evolution*, **47**, 253–263.

Hillis, D.M. & Bull, J.J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, **42**, 182–192.

Jermiin, L.S. & Crozier, R.H. (1994) The Cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger*: sequence divergence in Hymenoptera may be associated with nucleotide content. *Journal of Molecular Evolution*, **38**, 228–294.

Jermiin, L.S., Graur, D., Lowe, R.M. & Crozier, R.H. (1994) Analysis of directional mutation pressure and nucleotide content in mitochondrial Cytochrome b genes. *Journal of Molecular Evolution*, **39**, 160–173.

King, A.J. & Ring, L.R. (1996) The life history of *Arhopala wildei wildei* Miskin (Lepidoptera: Lycaenidae). *Australian Entomologist*, **23**, 117–120.

Kishino, H. & Hasegawa, M. (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidae. *Journal of Molecular Evolution*, **29**, 170–179.

Lanave, C., Preparata, G., Saccone, C. & Serio, G. (1984) A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution*, **20**, 86–93.

Lunt, D.H., Zhang, D.-X., Szymura, J.M. & Hewitt, G.M. (1996) The insect cytochrome oxidase I genes: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Molecular Biology*, **5**, 153–165.

Lyons-Weiler, J. & Hoelzer, G.A. (1999) Null model selection, compostional bias, character state bias, and the limits of phylogenetic information. *Molecular Biology and Evolution*, **16**, 1400–1405.

Maddison, D.R. & Maddison, W.P. (1999) *MacClade, Version 3.08*. Sinauer Associates, Sunderland, MA.

Maschwitz, U., Schroth, M., Hänel, H. & Pong, T.Y. (1984) Lycaenids parasitizing symbiotic plant–ant partnerships. *Oecologia*, **64**, 78–80.

Monteiro, A., Brakefield, P.M. & French, V. (1997) The relationship between eyespot shape and wing shape in the butterfly *Bicyclus anynana*: a genetic and morophometrical approach. *Journal of Evolutionary Biology*, **10**, 787–802.

Monteiro, A. & Pierce, N.E. (2000) Phylogeny of *Bicyclus* (Lepidoptera: Nymphalidae) inferred from COI, COII and EF-1α gene sequences. *Molecular Phylogenetics and Evolution*, **18**, 264–281.

Moritz, C., Patton, J.L., Schneider, C.J. & Smith, T.B. (2000) Diversification of rainforest faunas: an integrated molecular approach. *Annual Review of Ecology and Systematics*, **31**, 533–563.

Moriyama, E.N. (2000) *Molecular Evolutionary Analysis Package (MEA)*. Yale University, New Haven.

Pagel, M. (1998) Inferring evolutionary processes from phylogenies. *Zoologica Scripta*, **26**, 331–348.
Parsons, M. (1998) *The Butterflies of Papua New Guinea. Their Systematics and Biology*. Academic Press, New York.
Rand, D.M. (1994) Thermal habit, metabolic rate and the evolution of mitochondrial DNA. *Trends in Ecology and Evolution*, **9**, 125–131.
Rand, D.B., Heath, A., Suderman, T. & Pierce, N.E. (2000) Phylogeny and life history evolution of the genus *Chrysoritis* within the Aphnaeini (Lepidoptera: Lycaenidae), inferred from mitochondrial Cytochrome Oxidase I sequences. *Molecular Phylogenetics and Evolution*, **17**, 85–96.
Reed, R.D. & Sperling, F.A.H. (1999) Interaction of process partitions in phylogenetic analysis: an example from the swallowtail butterfly genus *Papilio*. *Molecular Biology and Evolution*, **16**, 286–297.
Rijsewijk, F., Schuermann, M., Wagenaar, E., Parren, P., Weigel, D. & Nusse, R. (1987) The *Drosophila* homolog of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless. *Cell*, **50**, 649–657.
Seki, Y., Takanami, Y. & Otsuka, K. (1991) *Butterflies of Borneo, 2: Lycaenidae*. Tobishima Corp., Tokyo.
Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.
Swofford, D.L. (2000) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.04a*. Sinauer Associates, Sunderland, MA.
Templeton, A.R. (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. *Evolution*, **37**, 221–244.
Voris, H.K. (2000) Maps of pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, **27**, 1153–1167.
Wahlberg, N. (2001) The phylogenetics and biochemistry of host-plant specialization in Melitaeine butterflies (Lepidoptera: Nymphalidae). *Evolution*, **55**, 522–537.
Wirth, Th., Le Guellec, R. & Veuille, M. (1999) Directional substitution and evolution of nucleotide content in the Cytochrome Oxidase II gene in Earwigs (Dermapteran Insects). *Molecular Biology and Evolution*, **16**, 1645–1653.
Yang, Z.H. (1996a) Maximum likelihood models for combined analyses of multiple sequence data. *Journal of Molecular Evolution*, **42**, 587–596.
Yang, Z.H. (1996b) Among site rate variation and its impact on phylogenetic analysis. *Trends in Ecology and Evolution*, **11**, 367–372.
Yang, Z.H. & Yoder, A.D. (1999) Estimation of the Transition/Transversion rate bias and species sampling. *Journal of Molecular Evolution*, **48**, 274–283.

Accepted 15 May 2003