Nymphalid butterflies diversify following near demise at the Cretaceous/Tertiary boundary

Niklas Wahlberg1,2,* , Julien Leneveu1 , Ulassa Kodandaramaiah2, Carlos Peña2, Sören Nylin2, André V. L. Freitas3 and Andrew V. Z. Brower4

1Laboratory of Genetics, Department of Biology, University of Turku, 20014 Turku, Finland
2Department of Zoology, Stockholm University, 10691 Stockholm, Sweden
3Departamento de Biología Animal and Museu de Zoología, Instituto de Biología, Universidade Estadual de Campinas, CP 6109, Campinas, SP 13083-970, Brazil
4Department of Biology, Middle Tennessee State University, Murfreesboro, TN 37132, USA

The butterfly family Nymphalidae contains some of the most important non-drosophilid insect model systems for evolutionary and ecological studies, yet the evolutionary history of the group has remained shrouded in mystery. We have inferred a robust phylogenetic hypothesis based on sequences of 10 genes and 235 morphological characters for exemplars of 400 of the 540 valid nymphalid genera representing all major lineages of the family. By dating the branching events, we infer that Nymphalidae originated in the Cretaceous at 90 Ma, but that the ancestors of 10–12 lineages survived the end-Cretaceous catastrophe in the Neotropical and Oriental regions. Patterns of diversification suggest extinction of lineages at the Cretaceous/Tertiary boundary (65 Ma) and subsequent elevated speciation rates in the Tertiary.

Keywords: Lepidoptera; Nymphalidae; mass extinction; times of divergence; diversification

1. INTRODUCTION

Butterflies in the family Nymphalidae are among the most charismatic insects in many habitats, and their beauty and diversity inspire a lifelong passion for the natural world among scientists and enthusiasts alike. As adults, nymphalids can be visually striking, and their behaviours are amenable to easy observation. These attributes have led scientists to use many nymphalid species as model systems in evolutionary and ecological studies (Boggs et al. 2003). Species of Nymphalidae are among the first taxa that helped us begin to understand the complex relationships between insects and their host plants (Ehrlich & Raven 1964), the effects of habitat fragmentation on the population dynamics of endangered species (Hanski 1999), the genetic mechanisms behind the developmental pathways of morphological features (Beldade & Brakefield 2002), and the coevolutionary interactions between organisms in mimicry rings and aposematic coloration (Brower 1996; Mallet et al. 1998). However, such important studies have been impeded by a lack of knowledge of the time scale involved in evolutionary processes leading to the patterns we see today (Vane-Wright 2004; Wahlberg 2006), be it the evolution of vagility or the evolution of host plant defence resistance.

The age of butterflies (including Nymphalidae) has been the subject of longstanding debate (Forbes 1947; Shields & Dvorak 1979; Scott & Wright 1990; de Jong 2007), with the consensus being that the five families of Papilionoidea originated and diversified in the Tertiary (Vane-Wright 2004). This mainly based on the occurrence of the few butterfly fossils known from Oligocene and Eocene deposits (Scott & Wright 1990; Emmel et al. 1992). However, several recent studies on subgroups within butterflies, extrapolating rates of DNA divergence from minimum age estimate calibrations of the fossils, suggest that the butterflies may be older, with most subfamilies originating before the KT boundary (Wahlberg 2006; Nazari et al. 2007; Wheat et al. 2007; Peña & Wahlberg 2008). The implications of such an age for the evolutionary history of Nymphalidae have not been investigated yet.

The timeframe over which a group of species has evolved has a strong influence on their implied historical biogeography. The distributions of extant butterflies have variously been attributed to vicariance following the break up of Gondwana (Miller & Miller 1997; Braby et al. 2005) and to dispersal events in more recent times (Kodandaramaiah & Wahlberg 2007; 2009). Obviously, the period of time during which a group of species has diversified will dictate which geological processes have possibly affected their evolution—for example, if a group is much younger than the splitting of the continents on which it is found today, trans-oceanic dispersal is likely to be favoured over vicariance as an explanation of current distributions (de Queiroz 2005).

Here we provide a robust phylogenetic hypothesis for the butterfly family Nymphalidae based on the most comprehensive sampling of genera to date (400 out of 540...
valid genera) and the largest dataset currently available (up to 10 genes sequenced and up to 235 morphological characters coded for each exemplar). Based on these data we also infer times of diversification, historical biogeography and the tempo of diversification in this family of approximately 6000 species.

2. MATERIAL AND METHODS

Butterfly samples were collected by the authors and contributed by colleagues from around the world (see electronic supplementary material, table S1 for the list of taxa). In most cases, genomic DNA was extracted from two legs of dried specimens; for some specimens the legs were preserved in 95 per cent ethanol in the field prior to extraction. DNA was extracted using the Qiagen DNEasy extraction kit following a slightly modified insect protocol. Polymerase chain reaction and sequencing protocols have been published previously (Wahlberg & Zimmermann 2000; Wahlberg et al. 2003, 2005b; Wahlberg & Wheat 2008), with the majority of new sequences being generated according to Wahlberg & Wheat (2008). Morphological characters were coded according to Freitas & Brown (2004) for those genera that had sufficient available material (electronic supplementary material, table S1).

We compiled a data matrix of 10 gene regions and 235 morphological characters, providing taxonomic representation of all subfamilies, tribes and subtribes of Nymphalidae, as well as 75 per cent of all currently accepted genera. Of the missing genera, almost all are considered to be closely related to sampled genera, and thus their exclusion is not expected to affect inferred relationships substantially. All taxa (400 genera plus 29 outgroups) were sequenced for 3127 base pairs of the gene regions COI, EF-1α and wingless, the standard gene regions employed for molecular systematics of Lepidoptera (Sperling 2003). In addition, seven novel gene regions (Wahlberg & Wheat 2008) were sequenced for multiple exemplars representing the phylogenetic diversity of the family: 86 taxa were sequenced for a 596 bp of ArgKin, 131 for 850 bp of CAD, 104 for 710 bp of IDH, 168 for 733 bp of MDH, 107 for 411 bp of RpS2 and 199 for 651 bp of RpS5. Furthermore, 238 taxa (including outgroups) were coded for up to 235 morphological characters (Freitas & Brown 2004). Molecular data for outgroups were taken from Wahlberg et al. (2005a), with the seven novel gene regions being sequenced for at least one representative of each outgroup family.

The combined morphological and molecular data were analysed using maximum parsimony in the program TNT (Goloboff et al. 2008). The data were treated with equal weights and morphological characters were coded as ordered or unordered according to Freitas & Brown (2004). Molecular data were treated as unordered. We used the New Search Technology algorithms in TNT; using level 10 of search intensity, holding up to 10 000 trees in each cycle, using default values for tree fusing, drifting and sectorial searches, and 50 ratchet cycles. The trees obtained underwent branch-swapping to obtain additional equally parsimonious trees. Stability of the nodes of the most parsimonious trees was assessed using 1000 bootstrap pseudoreplicates.

In addition, the molecular data were subjected to maximum likelihood analyses using the online version of RAxML (Stamatakis et al. 2008). The data were modelled with the default GTR+G model, and analysed unpartitioned, partitioned into mitochondrial and nuclear gene regions, and partitioned by individual gene regions. For the unpartitioned data, 10 separate analyses were done, each including a search for the maximum likelihood topology, as well as 100 bootstrap pseudoreplicates. The bootstrapped trees were combined into a single file to calculate bootstrap values for all nodes.

Times of divergences and topology were estimated simultaneously using the program BEAST v. 1.4.8 (Drummond & Rambaut 2007). Two kinds of analyses were done—one that included outgroup sequences and one that included only nymphalid sequences. Both analyses were repeated twice (each run took two to three months of CPU time on a 64-bit dual core desktop computer). Both analyses had the same calibration points.

The data were partitioned into the mitochondrial gene region (COI) and the combined nuclear genes (ArgKin, CAD, EF1α, GAPDH, IDH, MDH, RpS2, RpS5 and wingless), owing to the high proportion of A-T nucleotides in the mitochondrial gene compared with the approximately equal base ratios of the nuclear genes. The tree prior was set to the Yule process, the independent models for the two partitions were both set to the GTR+G model, while all other priors were left to defaults. The branch lengths were allowed to vary under a relaxed clock model with an uncorrelated log-normal distribution. The analyses were run for 40 million generations with every 4000th generation sampled. Using the TRACER program (part of the BEAST package) we confirmed that the two runs had converged to a stationary distribution after 5000 samples (burn-in), which left 5000 samples in each run describing the posterior distribution.

Maximum times of divergence were constrained by the ages of plant families for six clades of butterflies whose larvae are oligotrophic on those families. The maximum age of the clade Danainae was constrained to be 83 Ma based on the estimated age of Gentianales (Bremer et al. 2004). This plant clade is not used by any other nymphalid butterflies, but is commonly used by the danaines and inferred to be the ancestral host plant family of the group (Janz et al. 2006). The maximum age of the heliconine+nymphaline clade was constrained to be 110 Ma based on the estimated age of Rosales (Anderson et al. 2005), which is a plant clade very commonly used in these two clades and inferred to be the ancestral host plant family (Janz et al. 2006; Nylin & Wahlberg 2008). The maximum age of Biblidinae was constrained to be 65 Ma based on the estimated age of the split between Ricinus and Dalechampia (Davis et al. 2005), which represents the clade of Euphorbiaceae on which species of Biblidinae are highly specialized (DeVries 1987). The maximum age of the nymphaline clade including Kallimini, Junonini, Victorinini and Melitaeini was constrained to be 74 Ma based on the age estimate for the family Acanthaceae (Bremer et al. 2004), which is a plant family not used outside this clade (Nylin & Wahlberg 2008). The maximum age of Melitaeini, excluding Euphydryas, was constrained to be 40 Ma based on the estimated age of Asteraceae (Kim et al. 2005), which is a plant family very rarely used outside of this clade (Nylin & Wahlberg 2008). Finally, the maximum age of Satyrinae was constrained to be 65 Ma based on the estimated age of Poaceae (Prasad et al. 2003), which is a plant family very rarely used outside this clade (Peña & Wahlberg 2008).
Minimum age constraints were based on seven fossils, four of which are from the Florissant formation from the Eocene/Oligocene boundary (Emmel et al. 1992), two are from Dominican amber from the Miocene (Peñalver & Grimaldi 2006) and one is from Oligocene deposits in southeastern France (Nel et al. 1993). The minimum age of the split between Libythea and Libytheana was constrained to be 34 Ma based on two fossils, Libytheana vagabunda and L. florissanti, both from the Florissant formation in Colorado and recently found to be sister to Extant Libytheana in a cladistic analysis (Kawahara 2009). The minimum age of the split between Vanessa and its sister group was constrained to be 34 Ma based on the fossil Vanessa ameraisilicida (Emmel et al. 1992), which has been placed in the extant genus (Miller & Brown 1989). This fossil was also used as a calibration point (Kawahara 2009). The minimum age of the split between Hypearna and its sister group was constrained to be 34 Ma based on the fossil Prodrys persephone (Emmel et al. 1992), which is clearly related to Hypearna based on morphological features of the wings (K. Willmott 2006, personal communication). This fossil was also used as a calibration point by Wahlberg (2006). The minimum age of the split between Dynamine and its sister group was constrained to be 20 Ma based on the Dominican amber fossil of Dynamine alyacea, which clearly belongs to this extant genus (Peñalver & Grimaldi 2006). The minimum age of the split between Smyrna and its sister group was constrained to be 20 Ma based on the Dominican amber fossil larva of Smyrna (Hammond & Poinar 1998). The minimum age of the split between Libethe and its sister group was constrained to be 25 Ma based on the fossil Lethe corbieri found in the Oligocene deposits of southeast France (Nel et al. 1993). This fossil has been used as a calibration point in the study by Peña & Wahlberg (2008).

The BEAST analysis resulted in an ultrametric tree with 400 terminal taxa for the ingroup and 399 internal nodes. To analyse the tempo of speciation during the evolutionary history of Nymphalidae, the most probable age and 95 per cent credibility intervals for each node were transferred into a spreadsheet program. The cumulative number of lineages through time was plotted using a logarithmic scale: a constant rate of speciation in the lineage together with a lower extinction rate of speciation should produce an approximately exponentially increasing number of lineages until the final stages of diversification (Harvey et al. 1994), which are not of concern here owing to incomplete lineage sampling within genera. According to Harvey et al. (1994), a mass extinction leads to a severe drop in the lineages through time (LTT) plot, and this would be observed as a slowdown in speciation rate of extant lineages earlier on in the LTT plot.

Dispersal-vicariance optimization of ancestral areas implemented in DIVA v. 1.2 (Ronquist 1997) was used to infer the biogeographic history of the group. The distributions of genera were divided into six areas corresponding to the zoogeographic realms—Nearctic, Neotropical, Palearctic, Oriental, Australasian and Afro tropical. Nodes subtending genera with the same distributions were collapsed and treated as a single terminal. Wherever published hypotheses were available about the origin of a clade—for instance Junonia (Kodandaramaiah & Wahlberg 2007), Coenonympha (Kodandaramaiah & Wahlberg 2009) and Phyciodina (Wahlberg & Freitas 2007)—the inferred area of origin of the clade was used instead of the current distribution of its extant species. DIVA assigns a cost of one for dispersals and extinctions, while vicariance and within-area speciation events receive no cost. The ancestral optimization with the smallest cost is chosen as the preferred scenario. The optimization was run on the BEAST topology without constraints on the maximum number of inferred ancestral areas.

3. RESULTS AND DISCUSSION

All methods of analysis recovered generally similar topologies (figure 1; nodes with asterisks were very stable in all analyses). The trees were rooted with Hesperiidae, the putative sister taxon to the five families of Papilionoidea (Wahlberg et al. 2005a). We find the monophyly and relationships among 12 nymphalid subfamilies to be strongly supported in all analyses (figure 1; electronic supplementary material, figs S1–S3). Contrary to earlier studies based on molecular data (Brower 2000; Wahlberg et al. 2005a; Wahlberg & Wheat 2008), the position of Danainae is now stable and well-supported as the sister to all Nymphalidae except Libytheinae, a result congruent with a previous morphological study (Freitas & Brown 2004). The subfamily Libytheinae appears to be the sister taxon to the rest of Nymphalidae, a hypothesis suggested 50 years ago (Ehrlich 1958; Kristensen 1976), but never corroborated by molecular work thus far (Brower 2000; Wahlberg et al. 2005a; Wahlberg & Wheat 2008). The subfamily Pseudergolinae is very clearly an independent lineage sister to the other nymphaline subfamilies (figure 1) and not the sister group of Cyrestinae, as found in a previous study (Wahlberg et al. 2003).

The analysis of times of divergence suggests that Nymphalidae began diversifying in the late Cretaceous around 90 Ma, with the libytheine lineage branching off from the common ancestor of the rest of Nymphalidae (figure 2). Not long after that, the danaine and satyrine lineages diverged at about 89 and 85 Ma, respectively, and finally the heliconine and nymphaline lineages diverged from each other at about 78 Ma. Given the inferred credibility intervals of the times of divergence, it appears that all the cladogenesis leading to the lineages currently recognized as subfamilies occurred in the late Cretaceous or early Tertiary (figure 2), suggesting that Nymphalidae was already morphologically and ecologically diversified before the end of the Cretaceous period. The estimated origin of the family is coincident with the global rise of the angiosperms at about 100 Ma, which is likely to have been particularly favourable for a radiation of an angiosperm-feeding group such as Nymphalidae.

It is thought that the KT event that led to the extinction of non-avian dinosaurs also had a large impact on phytophagous insects (such as butterflies; Labandeira et al. 2002). Our inferred slowing down of diversification rate prior to the KT boundary (figure 2) can be explained by an increased number of lineage extinctions (Harvey et al. 1994) occurring at 65 Ma. In the lineage accumulation curve of figure 2, the red region corresponds to the period when all the extant subfamilies originated (94–64 Ma). The yellow region corresponds to the time interval when the subfamily lineages diversified, giving rise to tribes; notice that none of the divergence events leading to the tribes is older than a divergence event...
leading to a subfamily. The black zone of the curve comprises the period after the beginning of diversification into modern genera. As the sampling is not exhaustive in this time period (one sampled species per genus), the curve underestimates the cumulative number of lineages.

The shape of the curve found in the LTT plot for Nymphalidae (figure 2) is very similar to that found in the simulations of a mass extinction event in Harvey et al. (1994). The rapid increase in diversification rate at the beginning is expected in this kind of analysis; the lineages which survived to the present day are the ones, on average, that got off to a flying start (Nee et al. 1994). This ‘push of the past’ is followed by a period of stability (90–64 Ma) with a low net rate of diversification. This rate changes dramatically at 64 Ma and the time of lineage doubling halves (10 instead of 20 Myr) right after the KT boundary. This dramatic change in the net rate of diversification can reflect a change in the rate of speciation or extinction.

As argued above, a major extinction event in Nymphalidae that took place at about 64 Ma, such as the KT event, would explain the shape of the curve. Extrapolating the average rate of net diversification after the crisis (grey line over yellow part of the curve in figure 2) to the apparent origin of the period preceding the KT boundary (blue dotted line in figure 2), we find that 60 per cent of all the lineages have gone extinct under this hypothesis, suggesting that the KT event may have had a major impact on nymphalid butterflies. Harvey et al. (1994) suggested that only a large mass extinction event would leave such a signal in the LTT plot. It is possible that the ancestral lineages of extant subfamilies of Nymphalidae survived in small populations in geographically restricted areas and became on average more genetically diversified from one another than before the event.

Further support for extinction of nymphalid butterflies around the time of the KT event comes from the historical biogeography of the family. A dispersal-vicariance
Figure 2. Bayesian ultrametric tree showing the phylogenetic relationships of subfamilies (clade colours as in figure 1) and tribes of Nymphalidae, with ages given in millions of years ago (Ma). The red and yellow stars show the nodes of the divergence events that have led to the subfamilies and the tribes, respectively. The grey ellipses indicate the time of the first divergence event within subfamilies/tribes (based on electronic supplementary material, fig. S4). The curve below the tree shows the cumulative number of lineages through time (the analysis includes the complete dataset, up to the genus level). Coloured parts of the curves refer to the inferred divergences leading to the major extant taxonomic subdivisions. The subfamilies (in red) all originated before or immediately after the Cretaceous–Tertiary (KT) event, whereas all the lineages leading to the tribes (in yellow) diversified several million years after the KT boundary. The average rate of net diversification from the Paleocene to the mid-Eocene (grey line) has been traced on top of the lineage through time plot (in yellow). The average rate was then extrapolated as the potential rate of diversification in the late Cretaceous (blue line) where the observed rate of diversification is observed to be twice as slow (probably owing to extinctions), leading to an estimate of up to 60 per cent loss of diversity at 65 Ma. The last part of the curve (in black) represents the time period in which genera have diversified to produce extant species (47 Ma to present). Owing to non-exhaustive species-level sampling, this part of the curve underestimates the net rate of diversification.
analysis with no restriction on maximum ancestral areas shows the surprising result that the implied distributions of the ancestral taxa living between 94 and 33 Ma, a span of nearly 60 million years, were almost all in the Neotropical and/or Oriental regions (figure 2; electronic supplementary material, figs S3 and S4). Given that the Neotropics and the Oriental region were never in geological contact during this period, the implied absence of butterflies in the other major biomes is puzzling. We hypothesize that the ancestors of the various lineages in Nymphalidae were widespread prior to the KT boundary, but only 10–12 lineages in the Neotropical and Oriental regions survived the great extinction wave. These lineages then diversified in the Neotropical and Oriental regions before dispersing to the rest of the world 57–40 Myr (electronic supplementary material, fig. S5).

Subsequent diversification of subfamilies occurred in the Tertiary, with almost all lineages leading to currently recognized tribes diverging before the Oligocene (about 34 Ma; figure 2) and almost all sampled genera diverging from sister genera before the Late Miocene (about 11 Ma; electronic supplementary material, fig. S4). Of note here are the divergence times of the genus Heliconius from its sister genus Eueides at about 18 Ma (95% credibility interval 12–23 Ma) and of the genus Bicyclus from its sister genus Heteropsis at about 24 Ma (95% credibility interval 18–28 Ma). Heliconius and Bicyclus are the subjects of intense research and are candidates for genomic sequencing (Papa et al. 2008). Also noteworthy is that the estimated times of divergence for genera in the subfamily Nymphalinae are somewhat younger here (between 1 and 13 Ma depending on the clade) than in a previous study (Wahlberg 2006), and the estimated times of divergence for the subfamily Satyrinae are somewhat older (about 10 Myr older) than in a previous study (Peña & Wahlberg 2008), both of which used only butterfly fossils as calibration points. These 10 to 15 per cent fluctuations imply that inferred dates may shift further with the addition of data to the matrix or more external calibrations.

The optimal biogeographical hypothesis in DIVA indicates a widespread tropical ancestor at the base of Nymphalidae. The sister clade to Libytheinae may have arisen either in the Oriental or in the Neotropical region or both. As mentioned above, these two regions were inferred to be the ancestral areas from the base of the tree up to the nodes defining subfamilies as well as many of the tribes (electronic supplementary material, fig. S5). During the time span 57–34 Ma (the Eocene/Oligocene boundary), the Afrotopical region was colonized independently at least three times from the Oriental region: once by the common ancestor of Limenitidae at around 57 Ma, once or twice by the common ancestor of the Vagrantini+Argynnini tribes in Heliconiinae (between 40 and 50 Ma) and once by the common ancestor of Charaxini+Pallini in Charaxinae (also between 40 and 50 Ma). In addition, it appears that the Afrotopical region was colonized from the Neotropical region between 40 and 30 Ma by the common ancestors of Biblidini and Catonephelini (both Biblidinae). The Palaearctic region appears to have been colonized before the Eocene/Oligocene boundary by the ancestor of Nymphalini around 48 Ma and by the ancestor of Apaturinae around 40 Ma, both from the Oriental region. Also, the Nearctic region appears to have been colonized from the Oriental region by the ancestor of Melitaeini around 40 Ma. Extant distributions of genera have largely formed after the Eocene/Oligocene boundary (around 34 Ma), which is known as a period of cooling and drying on Earth, when large swaths of moist forests were replaced with dry grasslands (Willis & McElwain 2002), mainly because of the separation of remnant Gondwana into South America, Australia and Antarctica, allowing cold ocean currents to cool down the southern polar region. Species groups associated with open landscapes and grasslands appear to have diversified at this period, such as Satyrini (Peña & Wahlberg 2008) and Melitaeini (Leneveu et al. 2009).

Examining the temporal and spatial scale of macroevolutionary processes in phytophagous insects has only recently become possible with comprehensively sampled phylogenetic hypotheses and robust estimates of patterns and times of divergence (McKenna et al. 2009). We have been able to infer novel aspects of the deep evolutionary history of a species-rich group of phytophagous insects, suggesting that the event(s) causing mass extinction of vertebrates and other organisms with a well-preserved fossil record around the KT boundary also affected the taxa with a scant fossil record. These extinctions were putatively caused by a massive bolide impact, which could have cooled down Earth substantially for a long period of time. This clearly would have affected ephemeral, ectothermic insects dependent on the warmth of the sun for daily activities such as feeding and reproduction, and here we show for the first time that phylogenetic evidence supports this hypothesis. Extinctions could have been further compounded as a result of widespread extinctions of angiosperm host plants in the same period, disrupting obligate butterfly–plant interactions (Wilf et al. 2006). The robust, comprehensive phylogenetic hypothesis, and historical, biogeographical and divergence time estimates for this model group, will provide a long-awaited framework for comparative studies and open up new avenues of research that can capitalize on the extensive knowledge about the ecology and natural history of Nymphalidae.

We are grateful to the numerous people who have contributed specimens to this study, too numerous to name individually, although we would like to point out the importance of contributions by the African Butterfly Research Institute and Area de Conservación Guanacaste, Costa Rica. We are grateful to Folmer Botma and an anonymous reviewer for useful comments on a previous version of the manuscript. This work has been supported by grants from the Academy of Finland to N.W. (no. 118369), the Swedish Research Council to S.N. and N.W., and NSF (no. DEB0640301) to A.V.Z.B. A.V.L.F. acknowledges FAPESP (no. 98/05101-8 and no. 04/05269-9), Brazilian CNPq (fellowship no. 300315/2005-8) and NSF (DEB grant no. 0527441).

REFERENCES


