

Diversity dynamics in Nymphalidae butterflies: Effect of phylogenetic uncertainty on diversification rate shift estimates

Carlos Peña^{1,*}, Marianne Espeland²

1 Carlos Peña Laboratory of Genetics, Department of Biology, University of Turku, Turku

2 Museum of Comparative Zoology and Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, USA. Email: marianne.espeland@gmail.com.

Current address: McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida, Gainesville, FL, US

* E-mail: Corresponding mycalesis@gmail.com

Abstract

The species rich butterfly family Nymphalidae has been used to study evolutionary interactions between plants and insects. Theories of insect-hostplant dynamics predict accelerated diversification due to key innovations. We investigated whether phylogenetic uncertainty affects a commonly used method (MEDUSA, modelling evolutionary diversity using stepwise AIC) for estimating shifts in diversification rates in lineages, by extending the method across a random sample of trees from the posterior distribution of a Bayesian run. We found that phylogenetic uncertainty greatly affects diversification rate estimates. Different trees produced diversification rates ranging from high values to almost zero for the same clade, and both significant rate increase and decrease in some clades. Only four out of 18 significant shifts found on the maximum clade credibility tree were consistent across most of the sampled trees. Among these, we found accelerated diversification for Ithomiini butterflies. We used the binary speciation and extinction model (BiSSE) and found that a hostplant shift to Solanaceae is correlated with increased net diversification rates in Ithomiini, congruent with the diffuse cospeciation hypothesis. Our results show that taking phylogenetic uncertainty into account when estimating net diversification rate shifts is of great importance, and relying on the maximum clade credibility tree alone can potentially give erroneous results.

Keywords: diversification analysis, MEDUSA, BiSSE, speciation rate, insect-hostplant dynamics

Introduction

Hostplant shifts have been invoked to be responsible for a great part of the biodiversity of herbivorous insects [1]. The study of the evolution of hostplant use has spawned several theories explaining the evolutionary interactions between plants and insects [2]. The “escape-and-radiate” hypothesis [3], the “oscillation hypothesis” [4,5] or “diffuse cospeciation” [2].

The butterfly family Nymphalidae has been an important taxon for developing some of the mentioned hypotheses. Nymphalidae contains around 6000 species [6], and is the largest family within the true butterflies. The family most likely originated around 94 MYA in the mid Cretaceous. Diversification of the group began in the Late Cretaceous and most major radiations (current subfamilies) appeared shortly after the Cretaceous-Paleogene (K-Pg) boundary [7]. Several studies have used time-calibrated phylogenies and diversification models to reconstruct the evolutionary history of the group to identify patterns of accelerated or decelerated diversification of some Nymphalidae clades [7–10]. For example, it has been suggested that climate change in the Oligocene and the subsequent diversification of grasses has led to diversification of the subfamily Satyrinae [11] due to the abundance of grasses over extensive geographic areas (“resource abundance-dependent diversity dynamics” hypothesis). Reference 9 found increased net diversification rates in some Nymphalidae lineages after a major hostplant shift, which appears to be in agreement with Reference 3 “escape-and-radiate” model of diversification.

Although it has been suggested that part of the great diversity of Nymphalidae butterflies is a

result of hostplant-insect dynamics, it is necessary to use modern techniques to investigate whether the diversification patterns of Nymphalidae are in agreement with the theoretical predictions. It is necessary to test whether the overall diversification pattern of Nymphalidae is congruent with events of sudden diversification bursts due to hostplant shift, climatic events or shifts to closely related hostplants [5,12].

In this study, we used a time-calibrated genus-level phylogenetic hypothesis for Nymphalidae butterflies [10] to investigate patterns of diversification. We applied MEDUSA [13,14], a recently developed statistical method, to study the diversification pattern of Nymphalidae butterflies. MEDUSA fits likelihood models of diversification into a time-calibrated tree and tests whether allowing increases or decreases in speciation and extinction rates within the tree produces better fit of the models. MEDUSA is able to take into account unsampled extant species diversity during model fitting and it is normally applied to the maximum clade credibility phylogenetic tree. Particularly, we wanted to study the effects of phylogenetic uncertainty and by using the extended MEDUSA method called MultiMEDUSA [13]. We also tested whether hostplant association dynamics can explain the diversification patterns of component Nymphalidae lineages by testing whether character states of hostplant use affected the diversification pattern of those lineages employing the method BiSSE as implemented in the R package `diversitree` [15].

Materials and Methods

Data

For analyses, we used the phylogenetic trees from the study of Reference 10 that were generated using DNA sequence data from 10 gene regions for 398 of the 540 valid genera in Nymphalidae. We employed Reference 10 maximum clade credibility tree (MCC tree) (Fig. 1) as well as a random sample of 1000 trees from their BEAST run after burnin. Their original BEAST run was for 40 million generations. We used a burnin of 25 million generations and took a random sample of 1000 trees using Burntrees v.0.1.9 <http://www.abc.se/~nylander/> (supp. mat. 5) in order to correct for phylogenetic uncertainty when performing the diversification analyses.

Species richness data for Nymphalidae genera were compiled from several sources including the specialist-curated lists on <http://tolweb.org>, Reference 16 and curated lists of Global Butterfly Names project (<http://www.ucl.ac.uk/taxome/gbn/>). We assigned the species numbers of genera not included in the phylogeny to the closest related genus that was included in Reference 10 study according to available phylogenetic studies [17–28], taxonomical classification and morphological resemblance when no phylogenies were available (supp. mat. 3).

Hostplant data for Nymphalidae species were compiled from several sources including Reference 29, HOSTS database (<http://bit.ly/YI7nwW>), Reference 30, Reference 31, Reference 32 and others (supp. mat. 15–16) for a total of 6586 hostplant records, including 428 Nymphalidae genera and 143 plant families and 1070 plant genera. It was not possible to find any hostplant data for 35 butterfly genera.

Analyses of Diversification

We used the statistical software R version 3.0.1 [33] in combination with the APE [34], GEIGER [35], MEDUSA [13] and `diversitree` [15] packages along with our own scripts to perform the analyses (included as supplementary materials 2, 6, 10). All analyses were run on the 1000 random trees from Reference 10, on the maximum clade credibility tree derived from these and the MCC tree from Reference 10.

Detecting diversification shifts on phylogenetic trees

Patterns of diversification in Nymphalidae were analyzed by using MEDUSA version 093.4.33 [13] on the maximum clade credibility tree from Reference 10. MEDUSA fits alternative birth-death likelihood models to a phylogenetic tree in order to estimate changes in net diversification rates along branches.

MEDUSA estimates likelihood and AIC scores for the simplest birth-death model, with two parameters, the rates r : net diversification = speciation (b) - extinction (d) and ε : relative extinction = d/b . The AIC scores of the two-parameter model are then compared with incrementally more complex models until the addition of parameters do not improve the AIC scores beyond a cutoff value. MEDUSA finds the likelihood of the models after taking into account branch lengths and number of species per lineage [13]. Most studies using MEDUSA only run the method on a single tree, usually the maximum clade credibility tree, which makes the assumption that this tree is correct. We wanted to study the effects of phylogenetic uncertainty on estimation of net diversification rate shifts and therefore selected 1000 random genus-level trees from the posterior distribution (MultiMEDUSA, supp. mat. 5). We calculated a new MCC tree from the selection of trees, ran MEDUSA and MultiMEDUSA on the selection of 1000 trees and summarized the estimated changes in net diversification rates for nodes across all trees. Patterns of change in net diversification rates are considered significant if they are found at the same node in at least 90% of the trees. We also expected to find similar values of net diversification ($r = b - d$) and relative extinction rate ($\varepsilon = d/b$) values across the 1000 trees for the nodes where changes in diversification tempo occurs. We used the AICc threshold of 7.8 units, as estimated by MEDUSA, as the limit for a significantly better fit to select among increasingly complex alternative models.

Estimation of trait-dependent speciation rates

As the MEDUSA and MultiMEDUSA approaches estimated an increase in net diversification in the clade Ithomiini, we tested whether this pattern can be explained by increase in birth-rate due to hostplant use and performed analyses with the “binary state speciation and extinction” [36] Bayesian approach as implemented in the R package `diversitree` [15]. MuSSE [15] is designed to examine the joint effects of two or more traits on speciation. Because most of Nymphalidae butterflies are restricted to use one plant family as hostplant, the character states can be coded as presence/absence, for which the BiSSE analysis is better suited. BiSSE was designed to test whether a binary character state has had any effect on increased net diversification rate for a clade [36]. We used our compiled data of hostplant use to produce a binary dataset for the character “feeding on the plant family Solanaceae” with two states (absence = 0; presence = 1) (supp. mat. 17), which is the main hostplants of the diverse Ithomiini butterflies and closest relatives [37]. Other hostplant shifts were not tested for effect on net diversification rates due to the low support for diversification shifts found in the MEDUSA analyses. We analyzed the data using BiSSE employing the Markov Chain Monte Carlo algorithm on the maximum clade credibility tree, taking into account missing taxa by using the parameter “sampling factor” (`sampling.f`) in `diversitree`. We also used constrained analyses forcing no effect of hostplant use on diversification and used likelihood ratio tests to find out whether the hypothesis of effect on diversification has a significantly better likelihood than the null hypothesis (no effect). The analyses were run across a sample of 250 trees from the posterior distribution. The records of *Vanessa* and *Hypanartia* feeding on Solanaceae [30,38] might be incorrect as it is unlikely that these species can be feeding on this plant family. We tested whether coding these two genera as not feeding on Solanaceae affected our results.

Results

Detecting diversification shifts on the maximum clade credibility tree

The MEDUSA analysis on the MCC tree in combination with richness data estimated 18 changes in the tempo of diversification in Nymphalidae history (Fig. 1; Table 1). The corrected AICc acceptance threshold for adding subsequent piecewise birth-death processes to the overall model was set to 7.8 units, as prescribed by MEDUSA. In all MEDUSA analyses, the maximum number of inferred diversification splits in all trees was 26. The background net diversification rate for Nymphalidae was estimated as r

= 0.092 lineages per Million of years and the AICc score for the best fit model was 5090.5 (Table 1). MEDUSA also estimated that the basic constant birth–death model was not a better explanation for our data (AICc = 5449.3).

Some of the 18 changes in diversification correspond to rate increases in very species-rich genera: *Ypthima* ($r = 0.311$), *Charaxes* ($r = 0.291$), *Callicore* + *Diaethria* ($r = 0.220$), *Pedaliodes* ($r = 0.196$) and *Taenaris* ($r = 0.234$). We found rate increases for other clades as well such as: Mycalesina ($r = 0.191$), Oleriina + Ithomiina + Napeogenina + Dircennina + Godyridina ($r = 0.187$), Satyrini ($r = 0.116$), Phyciodina in part ($r = 0.241$) and Satyrina ($r = 0.221$), *Coenonympha* ($r = 0.209$), *Caeruleptychia* + *Magneptychia* ($r = 0.312$) and *Taenaris* ($r = 0.312$). We also found decreases in net diversification rates for Limenitidinae + Heliconiinae ($r = 0.0541$), part of Danaini ($r = 0.0423$), Pseudergolinae ($r = 0.024$) and Coenonymphina ($r = 0.065$) (Table 1).

Phylogenetic uncertainty in the MultiMEDUSA approach

We tested MEDUSA to find out whether taking into account the phylogenetic signal from a random sample of 1000 trees from the posterior distribution can return similar estimates of diversification to the values obtained from the MCC tree. We run MultiMEDUSA on a random sample of 1000 trees (supp. mat. 5) from the posterior distribution and compared the results with a MEDUSA analysis on the MCC tree derived from this sample (supp. mat. 9).

We found that the analysis by MultiMEDUSA on the 1000 trees estimated lower median net diversification rates for the diversification shifts found by MEDUSA on the MCC tree from Reference 10 (Table 2). Although the diversification pattern found by MEDUSA and MultiMEDUSA was the same, the latter consistently estimated lower rates. Thus, the shifts recovered with low net diversification rate on the MCC were recovered with negative net diversification rate by MultiMEDUSA. The background diversification and all shifts found by MEDUSA on the 1000 trees are provided as an R object in supp. mat. 7).

We also compared the results from MultiMEDUSA (derived from the sample of 1000 trees) with the splits found by MEDUSA on the MCC tree derived from this random sample. In the summary statistics, MultiMEDUSA reports the frequency of the diversification shifts found in the trees (parameter `sum.prop`). Thus, if a node is found in only half of the 1000 trees, but the phylogenetic signal was strong enough to be picked up by MEDUSA and a node shift was found most of the time, then the `sum.prop` should be close to 1. For example the node 625 consists of the clade *Charaxes* + *Polyura* and it was found in only 256 trees, however MultiMEDUSA was consistently able to find a diversification shift for that node and the `sum.prop` value is 0.996.

We found different rates for diversification shifts even for the nodes that appear in most of the sample of 1000 trees (Table 2) as many of the nodes also found in the MCC tree had a probability of being recovered higher than 0.90.

There were four diversification shifts found with more than 90% of probability in the trees from the random sample (Table 2): (i) net diversification rate increase in the genus *Ypthima*; (ii) net diversification rate increase in the genus *Charaxes*; (iii) rate increase in Ithomiini subtribes Oleriina + Ithomiina + Napeogenina + Dircennina + Godyridina; and (iv) rate increase in *Callicore* + *Diaethria*.

MultiMEDUSA provided mean and standard deviation statistics for the diversification values found on the shifts on the 1000 trees (supp. mat. 13–14), and found that some of the changes in net diversification rate values had great variation across the posterior distribution of trees. A boxplot of the net diversification rate values estimated for the clades that appear in the MCC tree shows that some shifts are estimated as increased or slowed diversification pace depending on the tree used for analysis (Fig. 2). This variation is especially wide for the clade formed by the genera *Magneptychia* and *Caeruleptychia* because MEDUSA estimated diversification values from six times the background net diversification rate ($r = 0.5234$) to almost zero ($r = 1.22e-01$). The rates for *Taenaris* were between 0.14 and 0.44 (mean value 0.25). Similar degrees of variation were found in the nodes for *Ypthima*, *Charaxes* and *Coenonympha* (Fig. 2). The net diversification rates estimates for the clades (Oleriina + Ithomiina + Napeogenina + Dircennina

+ Godyridina), Limenitidinae + Heliconiinae and Pseudergolinae are relatively consistent across the 1000 trees (Fig. 2).

It is also evident that not all the diversification shifts estimated on the MCC tree are consistently recovered in most of the 1000 trees. Some of the splits in the MCC tree are recovered in very few trees, for example the split for the clade Satyrini (Euptychiina + Pronophilina + Satyrina + Maniolina) is recovered with a probability of 18% (Fig. 3).

Estimation of trait-dependent speciation rates

The MEDUSA analyses taking into account phylogenetic uncertainty estimated a net diversification rate increase in part of the clade Ithomiini across more than 95% of the trees. Our BiSSE analysis found a positive effect of the character state “feeding on Solanaceae” on the net diversification rate on part of Ithomiini (Oleriina + Ithomiina + Napeogenina + Dircennina + Godyridina) (Fig. 4). The Markov Chain Monte Carlo algorithm was run for 10000 generations discarding the first 7500 as burnin. The estimated mean net diversification rate for taxa that do not feed on Solanaceae was $r = 0.11$ while the net diversification rate for the Solanaceae feeders was $r = 0.16$ (see Fig. S1 for a boxplot of speciation and extinction values for the 95% credibility intervals). The same analysis considering *Vanessa* and *Hypanartia* as non-Solanaceae feeders due to dubious records produced the same pattern and net diversification rates (Fig. S2). Therefore, the rest of the analyses were performed assuming these two genera as Solanaceae feeders. We constrained the BiSSE likelihood model to force equal rates of speciation for both character states in order to test whether the model of different speciation rates is a significantly better explanation for the data. A likelihood ratio test found that the model for increased net diversification rate for nymphalids feeding on Solanaceae is a significantly better explanation than this character state having no effect on diversification ($p < 0.001$) (Table 3; character states available in supp. mat. 17, code in supp. mat. 18, and mcmc run in supp. mat. 19). We combined the post-burnin mcmc generations from running BiSSE on 250 trees from the posterior distribution and found the same pattern as the BiSSE analysis on the maximum clade credibility tree (combined mcmc run in supp. mat. 20; profiles plot of speciation rates in Fig. S3; boxplot of 95% credibility intervals in Fig. S4). A BiSSE analysis to test whether the trait “feeding on Apocynaceae” had any effect on increased net diversification rates found similar speciation rates for lineages feeding on Apocynaceae and other plants (Fig. S5). It has been shown that BiSSE performs poorly under certain conditions [39]. However, our data has adequate number of taxa under analysis (more than 300 tips), adequate speciation bias (between 1.5x and 2.0x), character state bias (around 8x) and extinction bias (around 4x) for the analysis of Solanaceae hostplants. Thus, BiSSE is expected to produce robust results [39].

Discussion

Effects of phylogenetic uncertainty on the performance of MEDUSA

The MEDUSA method has been used to infer changes in net diversification rates along a phylogenetic tree. Since its publication [13] the results of using MEDUSA on a single tree, the maximum clade credibility tree, have been used for generation of hypotheses and discussion [7,40,41]. However, we found that MEDUSA estimated different diversification shifts and different rates of diversification for certain lineages when phylogenetic uncertainty was taken into account by using MEDUSA on a random sample of trees from the posterior distribution of a Bayesian run. We found that some diversification splits, estimated on the Nymphalidae maximum clade credibility tree, were found with a very low probability in a random sample of 1000 trees from the posterior distribution (Fig. 3, Table 2). We also found that, even though MEDUSA could estimate the same diversification splits on two or more trees, the estimated net diversification rates could vary widely (Fig. 2). For example, in our Nymphalidae trees, we found that the split for

Magneptychia and *Caeruleptychia* had a variation from $r = 0.5234$, higher than the background net diversification rate, to almost zero. This means that observed patterns and conclusions can be completely contradictory depending on tree choice.

In this study, the effect of phylogenetic uncertainty on the inferred diversification splits by MEDUSA is amplified because some Nymphalidae taxa appear to be strongly affected by long-branch attraction artifacts [27]. Thus, the Bayesian runs are expected to recover alternative topologies on the posterior distribution of trees, resulting in low support and posterior probability values for the nodes. For example, posterior probability values for clades in Satyrini are very low [10]. As a result, MEDUSA inferred a net diversification rate increase for part of Satyrini in the maximum clade credibility tree, but this was recovered only in 13% of probabilities in the MultiMEDUSA analysis on the random sample of trees.

If there is strong phylogenetic signal for increases or decreases in net diversification rates for a node, it is expected that these splits would be inferred by MEDUSA in most of the posterior distribution of trees. However, weak phylogenetic signal for some nodes can cause some clades to be absent in some trees and MEDUSA will be unable to estimate any diversification shift (due to a non-existent node). This is the reason why MEDUSA estimated net diversification rate splits with more than 90% of probabilities in the sample of trees for only four splits: the genus *Charaxes*, the genus *Ypthima*, part of Ithomiini and the clade *Callicore* + *Diaethria* (Fig. 3), while estimating splits for other lineages with much lower probability.

The clade Ithomiini and the non-basal danaids are well supported by high posterior probability values in Reference 10. Therefore our MEDUSA analyses recovered an increase in net diversification rate with probability higher than 90% in the posterior distribution of trees (Fig. 3).

Hostplant use and diversification in Nymphalidae

Ithomiini

Keith Brown suggested that feeding on Solanaceae was an important event in the diversification of Ithomiini butterflies [42]. Ithomiini butterflies are exclusively Neotropical and most species feed on Solanaceae hostplants during larval stage [37]. Optimizations of the evolution of hostplant use on phylogenies evidence a probable shift from Apocynaceae to Solanaceae in the ancestor of the tribe [37,43]. Reference 9 found that the Gamma statistics, a LTT plot of an Ithomiini phylogeny and the fit of the density-dependent model of diversification are consistent with a burst of diversification in Ithomiini following the shift from Apocynaceae to Solanaceae.

We investigated whether the strong signal for an increase in net diversification rate for Ithomiini (found by MEDUSA) can be explained due to the use of Solanaceae plants as hosts during larval stage. For this, we used a Bayesian approach [44] to test whether the trait “feeding on Solanaceae” had any effect on the diversification of the group.

Our BiSSE analysis, extended to take into account missing taxa and phylogenetic uncertainty, shows a significantly higher net diversification rate for Ithomiini taxa, which can be attributed to the trait “feeding on Solanaceae hostplants” (Fig. 4). This is in agreement with the findings of Reference 9 using other statistical methods. Due to the fact that Ithomiini are virtually the only nymphalids using Solanaceae as hostplants, it is possible that the trait responsible for a higher diversification of Ithomiini might not be the hostplant character. As noted by Reference 36, the responsible trait might be a codistributed character such as a trait related to the ability to digest secondary metabolites.

Solanaceae plants contain chemical compounds and it has been suggested that the high diversity of Ithomiini is consistent with the “escape-and-radiate scenario” due to a shift onto Solanaceae [9] and radiation scenarios among chemically different lineages of Solanaceae plants [37,42]. According to this theory, the shift from Apocynaceae to Solanaceae allowed Ithomiini to invade newly available resources due to a possible key innovation that allowed them cope with secondary metabolites of the new hosts. Additional studies are needed to identify the actual enzymes that Ithomiini species might be using for

detoxification of ingested food as they have been found in other butterfly groups [45].

The increase in diversification inferred by MEDUSA occurred after the probable shift from Apocynaceae to Solanaceae, as the Solanaceae feeders in the subtribes Melinaeina and Mechanitina are not included in the diversification shift (shift number 5 in Fig. 1). The apparent conflicting results from MEDUSA and BiSSE can be explained by the low species-richness of the subtribes Melinaeina and Mechanitina compared to the other subtribes included in the shift (52 versus 272 species). It can be that MEDUSA is more conservative than BiSSE and is not including Melinaeina and Mechanitina in the shift due to low species numbers.

Although the Solanaceae genera used by the Ithomiini clades are well known [37], we do not have any understanding on the physiological routes involved in the detoxification of Solanaceae compounds by the several lineages of Ithomiini. We can speculate that older lineages exploiting a novel toxic resource [10,37] might not be too efficient in metabolizing plant toxins and that younger lineages are able to deal with toxins more efficiently, so that host switching events within Solanaceae are possible, which can lead to higher diversification. Studies in *Papilio* species have reported that detoxification enzymes can become more efficient in metabolizing toxins than ancestral configurations of the proteins, providing more opportunities for hostplant switch [46]. This might be the reason why the basal Ithomiini subtribes Melinaeina and Mechanitina are so species-poor and restricted to few Solanaceae hosts [37], while recent subtribes are species-rich and have expanded their host range into several Solanaceae lineages [37]. It might be that the switch to feeding in Solanaceae was an important event in the evolutionary history of Ithomiini, but the actual radiation occurred after critical physiological changes (a probable key innovation) allowed efficient detoxification of Solanaceae toxins.

The diffuse cospeciation hypothesis predicts almost identical ages of insects and their hostplants, while the “resource abundance-dependent diversity” and the “escape-and-radiate” hypotheses state that insects diversify after their hostplants [2–4]. Reference 45 found strong evidence for a model of speciation congruent with Ehrlich and Raven’s hypothesis in Pieridae butterflies due to, in addition to the identification of a key innovation, a burst of diversification in glucosinolate-feeding taxa shortly afterwards (with a lag of ~10 MY). According to a recent dated phylogeny of the Angiosperms [47], the family Solanaceae split from its sister group about 59 (49–68) MYA and diversification started (crown group age) around 37 (29–47) MYA. Reference 10 give the corresponding ages for Ithomiini as 45 (39–53) and 37 (32–43) MYA, respectively. Thus, current evidence shows that Solanaceae and Ithomiini might have diversified around the same time, during the Late Eocene and Oligocene, and this would be congruent with the diffuse cospeciation hypothesis.

The uplift of the Andes was a major tectonic event that underwent higher activity during the Oligocene [48]. This caused climatic changes in the region that affected the flora and fauna of the time, which coincides with the diversification of modern montane plant and animal taxa [49] including Ithomiini butterflies and Solanaceae hostplants. Moreover, all Solanaceae clades currently present in New World originated in South America [50] as well as Ithomiini butterflies [10]. This suggests a process of “diffuse cospeciation” of Ithomiini and hostplants.

Danaini

Our MultiMEDUSA approach gave a significant slowdown in net diversification rate in the subtribe Danaina of the Danini. Both Danaina and the sister clade Euploeina feed mainly on Apocynaceae and thus a hostplant shift should not be responsible for the observed slowdown of diversification in the Danaina. As expected, our BiSSE analysis of Apocynaceae feeders shows that there is no effect of feeding on this plant family on the net diversification rates of Nymphalidae lineages. Many of the Danaina are large, strong fliers, highly migratory and involved in mimicry rings. Among them is for example the monarch (*Danaus plexippus*), probably the most well known of all migratory butterflies. The causes for a lower net diversification rate in the Danaina remains to be investigated, but their great dispersal power might be involved in preventing allopatric speciation. It has been found in highly vagile species in the nymphalid

genus *Vanessa* that dispersal has homogenized populations due to gene flow, as old and vagile species seem to be genetically homogeneous while younger widespread species show higher genetic differentiation in their populations [51].

Charaxes

The genus *Charaxes* contain 193 species distributed in the Old World with highest diversity in the Afrotropical region. These butterflies are also very strong fliers, but contrary to *Danaina*, which are specialized Apocynaceae feeders, *Charaxes* are known to feed on at least 28 plant families in 18 orders [29] and some species appear to be polyphagous [52]. Reference 53 showed that most of the diversification of the genus occurred from the late Oligocene to the Miocene when there were drastic global climatic fluctuations, indicating that the diversity mainly is driven by climate change. Reference 52 found that climatic changes during the Pliocene, Pleistocene as well as dispersal and vicariance might have been responsible for the high diversification of the genus.

Satyrini

The diverse tribe Satyrini radiated simultaneously with the radiation of their main hostplant, grasses, during the climatic cooling in the Oligocene [11]. Thus, it is somewhat surprising that part of Satyrini (the subtribes Euptychiina, Satyrina and Pronophilina) were found to have accelerated diversification in only 13% of probabilities in the trees from the posterior distribution. Although this can be attributed to low phylogenetic signal [10], the clade Satyrini is very robust [10] and MEDUSA failed to identify any significant accelerated net diversification rate for Satyrini. It appears that the radiation of Satyrini as a whole was not remarkably fast and therefore not picked up by MEDUSA.

The origin of the tribe Satyrini is not completely clear (originated either in the Neotropical or Eastern Palaearctic, Oriental and/or Indo-Australian regions [27] and their radiation involved colonizing almost all continents starting from their place of origin. For butterflies, a prerequisite for colonizing new areas is that suitable hostplants are already present. Therefore, it should be expected that the diversification of Satyrinae occurred in a stepwise manner, with pulses or bursts of diversification for certain lineages but unlikely for the tribe Satyrini as a whole.

We found that even though MEDUSA estimated several diversification shifts in the maximum clade credibility tree of Nymphalidae, only a few of these splits were found in more than 90% of the trees from the posterior distribution. In the literature, it is common practice that conclusions are based on the splits estimated on the maximum clade credibility tree. However, by using a MultiMEDUSA approach, we found that some of this splits might be greatly affected by phylogenetic uncertainty. Moreover, some of these splits can be recovered either as increases or decreases in net diversification rate depending on the tree from the posterior distribution that was used for analysis. This means that contradictory conclusions would be made if only the maximum clade credibility tree was used for analysis. MEDUSA appears to be sensitive to the number of nodes with high posterior probability and width of age confidence intervals. For our data, it would be necessary to obtain a posterior distribution of trees with no conflicting topology, and very similar estimated ages for nodes in order to consistently recover most of the diversification splits on the posterior distribution of trees that were inferred by MEDUSA on the MCC tree.

Our MultiMEDUSA approach to perform analyses on the posterior distribution of trees found strong support for an increase in net diversification rate in the tribe Ithomiini and the genus *Charaxes*, and for a decrease in net diversification rate in the subtribe *Danaina*. Due to phylogenetic uncertainty, we did not obtain strong support for other diversification splits in Nymphalidae. Our BiSSE analysis corroborated other studies in that the trait “feeding on Solanaceae”, or a codistributed character, was important in the diversification of Ithomiini butterflies. However, by applying MEDUSA we found that a critical character in the radiation of the group might have appeared after the shift from Apocynaceae to Solanaceae. We also found that the trait “feeding on Apocynaceae” is not responsible for the slowdown of diversification

in *Danaina*. Ithomiini and Solanaceae diversified near simultaneously, which is in agreement with the diffuse co-speciation hypothesis [2,4].

Acknowledgments

We are thankful to Mark Cornwall for help with the script to extend MEDUSA to include phylogenetic uncertainty, Niklas Wahlberg for commenting on the manuscript and giving us the posterior distribution of trees, Luke Harmon for commenting on the manuscript and anonymous reviewers for their comments, which greatly improved the manuscript, Jessica Slove Davidson and Niklas Janz for access to their hostplant data. The study was supported by a Kone Foundation grant (awarded to Niklas Wahlberg), Finland (C. Peña) and the Research Council of Norway (grant no. 204308 to M. Espeland). We acknowledge CSC-IT Center for Science Ltd. (Finland) for the allocation of computational resources.

Figure legends

Figure 1. MEDUSA analysis run on the MCT from Reference 10. Estimated rate shifts in Table 1.

Figure 2. Range of diversification values for tips or clades estimated by MEDUSA on the 1000 random trees.

Figure 3. Results of the MultiMEDUSA analysis showing the probability for nodes of being recovered as diversification events.

Figure 4. BiSSE analysis of diversification of Solanaceae feeding nymphalids shows higher speciation and net diversification rates in Solanaceae feeders.

References

1. Mitter C, Farrell B, Wiegmann B (1988) The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *Am Nat* 132: 107–128. doi:10.1086/284840.
2. Nyman T, Linder HP, Peña C, Malm T, Wahlberg N (2012) Climate-driven diversity dynamics in plants and plant-feeding insects. *Ecol Lett* 15: 889–898. doi:10.1111/j.1461-0248.2012.01782.x.
3. Ehrlich PR, Raven PH (1964) Butterflies and plants: A study in coevolution. *Evolution* 18: 586–608. doi:10.2307/2406212.
4. Janz N (2011) Ehrlich and Raven Revisited: Mechanisms Underlying Codiversification of Plants and Enemies. *Annu Rev Ecol Evol S* 42: 71–89. doi:10.1146/annurev-ecolsys-102710-145024.
5. Nylin S, Slove J, Janz N (2014) Host plant utilization, host range oscillations and diversification in Nymphalid butterflies: A phylogenetic investigation. *Evolution* 68: 105–124. doi:10.1111/evo.12227.
6. Nieukerken EJ van, Kaila L, Kitching IJ, Kristensen NP, Lees DC, et al. (2011) Order Lepidoptera. In: *Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness*. Zootaxa. pp. 212–221.
7. Heikkilä M, Kaila L, Mutanen M, Peña C, Wahlberg N (2012) Cretaceous origin and repeated tertiary diversification of the redefined butterflies. *P Roy Soc B* 279: 1093–1099. doi:10.1098/rspb.2011.1430.
8. Elias M, Joron M, Willmott KR, Silva-Brandão KL, Kaiser V, et al. (2009) Out of the Andes: Patterns of diversification in clearwing butterflies. *Mol Ecol* 18: 1716–1729. doi:10.1111/j.1365-294X.2009.04149.x.
9. Fordyce JA (2010) Host shifts and evolutionary radiations of butterflies. *P Roy Soc B* 277: 3735–3743. doi:10.1098/rspb.2010.0211.
10. Wahlberg N, Leneveu J, Kodandaramaiah U, Peña C, Nylin S, et al. (2009) Nymphalid butterflies diversify following near demise at the Cretaceous/Tertiary boundary. *P Roy Soc B* 276: 4295–4302. doi:10.1098/rspb.2009.1303.

11. Peña C, Wahlberg N (2008) Prehistorical climate change increased diversification of a group of butterflies. *Biol Lett* 4: 274–278. doi:10.1098/rsbl.2008.0062.
12. Ferrer-Paris JR, Sánchez-Mercado A, Vilorio AL, Donaldson J (2013) Congruence and Diversity of Butterfly-Host Plant Associations at Higher Taxonomic Levels. *PloS one* 8: e63570. doi:10.1371/journal.pone.0063570.
13. Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, et al. (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *P Natl Acad Sci USA* 106: 13410–13414. doi:10.1073/pnas.0811087106.
14. Harmon LJ, Rabosky DL, FitzJohn RG, Brown JW (2011) turboMEDUSA: MEDUSA: Modeling Evolutionary Diversification Using Stepwise AIC.
15. FitzJohn RG (2012) Diversitree: comparative phylogenetic analyses of diversification in R. *Method Ecol Evol* 3: 1084–1092. doi:10.1111/j.2041-210X.2012.00234.x.
16. Lamas G (2004) Checklist: Part 4A. Hesperioidea-Papilionoidea. Gainesville.
17. Matos-Maraví PF, Peña C, Willmott KR, Freitas AV, Wahlberg N (2013) Systematics and evolutionary history of butterflies in the “Taygetis clade” (Nymphalidae: Satyrinae: Euptychiina): Towards a better understanding of Neotropical biogeography. *Mol Phylogenet Evol* 66: 54–68. doi:10.1016/j.ympev.2012.09.005.
18. Brower AVZ, Wahlberg N, Ogawa JR, Boppré M, Vane-Wright RI (2010) Phylogenetic relationships among genera of danaine butterflies (Lepidoptera: Nymphalidae) as implied by morphology and DNA sequences. *Syst Biodivers* 8: 75–89. doi:10.1080/14772001003626814.
19. Kodandaramaiah U, Lees DC, Müller CJ, Torres E, Karanth KP, et al. (2010) Phylogenetics and biogeography of a spectacular Old World radiation of butterflies: the subtribe Mycalesina (Lepidoptera: Nymphalidae: Satyrini). *BMC Evol Biol* 10: 172. doi:10.1186/1471-2148-10-172.
20. Kodandaramaiah U, Peña C, Braby MF, Grund R, Müller CJ, et al. (2010) Phylogenetics of Coenonymphina (Nymphalidae: Satyrinae) and the problem of rooting rapid radiations. *Mol Phylogenet Evol* 54: 386–394. doi:10.1016/j.ympev.2009.08.012.
21. Ortiz-Acevedo E, Willmott KR (2013) Molecular systematics of the butterfly tribe Preponini (Nymphalidae: Charaxinae). *Syst Entomol* 38: 440–449. doi:10.1111/syen.12008.
22. De-Silva DL, Day JJ, Elias M, Willmott KR, Whinnett A, et al. (2010) Molecular phylogenetics of the Neotropical butterfly subtribe Oleriina (Nymphalidae: Danainae: Ithomiini). *Mol Phylogenet Evol* 55: 1032–1041. doi:10.1016/j.ympev.2010.01.010.
23. Freitas AVL, Brown Jr. KS (2004) Phylogeny of the Nymphalidae (Lepidoptera). *Syst Biol* 53: 363–383. doi:10.1080/10635150490445670.
24. Peña C, Wahlberg N, Weingartner E, Kodandaramaiah U, Nylin S, et al. (2006) Higher level phylogeny of Satyrinae butterflies (Lepidoptera: Nymphalidae) based on DNA sequence data. *Mol Phylogenet Evol* 40: 29–49. doi:10.1016/j.ympev.2006.02.007.
25. Penz CM (1999) Higher level phylogeny for the passion-vine butterflies (Nymphalidae, Heliconiinae) based on early stage and adult morphology. *Zool J Linn Soc* 127: 277–344. doi:10.1111/j.1096-3642.1999.tb00680.x.
26. Silva-Brandão KL, Wahlberg N, Francini RB, Azeredo-Espin AML, Brown Jr. KS, et al. (2008) Phylogenetic relationships of butterflies of the tribe Acraeini (Lepidoptera, Nymphalidae, Heliconiinae) and the evolution of host plant use. *Mol Phylogenet Evol* 46: 515–531. doi:10.1016/j.ympev.2007.11.024.
27. Peña C, Nylin S, Wahlberg N (2011) The radiation of Satyrini butterflies (Nymphalidae: Satyrinae): a challenge for phylogenetic methods. *Zool J Linn Soc* 161: 64–87. doi:10.1111/j.1096-3642.2009.00627.x.
28. Peña C, Nylin S, Freitas AVL, Wahlberg N (2010) Biogeographic history of the butterfly subtribe Euptychiina (Lepidoptera, Nymphalidae, Satyrinae). *Zool Scr* 39: 243–258. doi:10.1111/j.1463-6409.2010.00421.x.
29. Ackery PR (1988) Hostplants and classification: a review of nymphalid butterflies. *Biol J Linn Soc* 33: 95–203. doi:10.1111/j.1095-8312.1988.tb00446.x.
30. Dyer LA, Gentry GL (2002) Caterpillars and parasitoids of a tropical lowland wet forest. Available: <http://www.caterpillars.org>. Accessed 17 April 2013.

31. Beccaloni G, Vilorio Á, Hall S, Robinson G (2008) Catalogue of the hostplants of the Neotropical butterflies. London: The Natural History Museum.
32. Janzen DH, Hallwachs W (2009) Dynamic database for an inventory of the macrocaterpillar fauna, and its food plants and parasitoids, of Area de Conservacion Guanacaste (ACG), northwestern Costa Rica. Available: <http://janzen.sas.upenn.edu>. Accessed 17 April 2013.
33. Core Team R (2013) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available: <http://www.r-project.org>.
34. Popescu A-A, Huber KT, Paradis E (2012) ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* 28: 1536–1537. doi:10.1093/bioinformatics/bts184.
35. Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics* 24: 129–131. doi:10.1093/bioinformatics/btm538.
36. Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character's effect on speciation and extinction. *Syst Biol* 56: 701–710. doi:10.1080/10635150701607033.
37. Willmott KR, Freitas AVL (2006) Higher-level phylogeny of the Ithomiinae (Lepidoptera: Nymphalidae): classification, patterns of larval hostplant colonization and diversification. *Cladistics* 22: 297–368. doi:10.1111/j.1096-0031.2006.00108.x.
38. Scott JA (1986) The Butterflies of North America. Stanford University Press.
39. Davis MP, Midford PE, Maddison W (2013) Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evol Biol* 13: 38. doi:10.1186/1471-2148-13-38.
40. Litman JR, Danforth BN, Eardley CD, Praz CJ (2011) Why do leafcutter bees cut leaves? New insights into the early evolution of bees. *P Roy Soc B* 278: 3593–3600. doi:10.1098/rspb.2011.0365.
41. Ryberg M, Matheny PB (2012) Asynchronous origins of ectomycorrhizal clades of Agaricales. *P Roy Soc B* 279: 2003–2011. doi:10.1098/rspb.2011.2428.
42. Brown Jr. KS (1987) Chemistry at the Solanaceae/Ithomiinae Interface. *Ann Mo Bot Gard* 74: 359–397.
43. Brower AVZ, Freitas AVL, Lee M-M, Silva-Brandão KL, Whinnett A, et al. (2006) Phylogenetic relationships among the Ithomiini (Lepidoptera: Nymphalidae) inferred from one mitochondrial and two nuclear gene regions. *Syst Entomol* 31: 288–301. doi:10.1111/j.1365-3113.2006.00321.x.
44. FitzJohn RG, Maddison WP, Otto SP (2009) Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst Biol* 58: 595–611. Available: doi:10.1093/sysbio/syp067.
45. Wheat CW, Vogel H, Wittstock U, Braby MB, Underwood D, et al. (2007) The genetic basis of a plant-insect coevolutionary key innovation. *P Natl Acad Sci USA* 104: 20427–20431. doi:10.1073/pnas.0706229104.
46. Li W, Schuler MA, Berenbaum MR (2003) Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: Specificity and substrate encounter rate. *P Natl Acad Sci USA* 100: 14593–14598. doi:10.1073/pnas.1934643100.
47. Bell CD, Soltis DE, Soltis PS (2010) The age and diversification of the angiosperms re-revisited. *Am J Bot* 97: 1296–1303. doi:10.3732/ajb.0900346.
48. Somoza R (1998) Updated Nazca (Farallon) - South America relative motions during the last 40 My: implications for mountain building in the central Andean region. *J S Am Earth Sci* 11: 211–215. doi:10.1016/S0895-9811(98)00012-1.
49. Hoorn C, Wesselingh FP, Steege H ter, Bermudez MA, Mora A, et al. (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* 330: 927–931. doi:10.1126/science.1194585.
50. Olmstead RG (2013) Phylogeny and biogeography in Solanaceae, Verbenaceae and Bignoniaceae: a comparison of continental and intercontinental diversification patterns. *Bot J Linn Soc* 171: 80–102. doi:10.1111/j.1095-8339.2012.01306.x.
51. Wahlberg N, Rubinoff D (2011) Vigility across *Vanessa* (Lepidoptera: Nymphalidae): mobility in butterfly species does not inhibit the formation and persistence of isolated sister taxa. *Syst Entomol* 36: 362–370. doi:10.1111/j.1365-3113.2010.00566.x.

52. Müller CJ, Wahlberg N, Beheregaray LB (2010) ‘After Africa’: the evolutionary history and systematics of the genus *Charaxes* Ochsenheimer (Lepidoptera: Nymphalidae) in the Indo-Pacific region. Biol J Linn Soc 100: 457–481. doi:10.1111/j.1095-8312.2010.01426.x.

53. Aduse-Poku K, Vingerhoedt E, Wahlberg N (2009) Out-of-Africa again: a phylogenetic hypothesis of the genus *Charaxes* (Lepidoptera: Nymphalidae) based on 5 gene regions. Mol Phylogenet Evol 53: 463–478. doi:10.1016/j.ympev.2009.06.021.

Supp. mat.

Supporting_Information.S1.nex: MCC Nymphalidae tree from Reference 10.

Supporting_Information.S2.R: Run MEDUSA on MCC tree from Reference 10.

Supporting_Information.S3.csv: Species richness for lineages in Nymphalidae.

Supporting_Information.S4.txt: Results of running MEDUSA on the MCC tree.

Supporting_Information.S5.zip: 1000 random trees from Reference 10.

Supporting_Information.S6.R: Run MultiMEDUSA on 1000 random trees from Reference 10.

Supporting_Information.S7.zip: Raw results from the MultiMEDUSA run on the random sample of trees from the posterior distribution.

Supporting_Information.S8.pdf: Summary of MultiMEDUSA analysis on the 1000 trees from the posterior distribution.

Supporting_Information.S9.tree: MCC tree from the 1000 random trees selected from the posterior distribution.

Supporting_Information.S10.R: Run MEDUSA on MCC tree from the 1000 random trees selected from the posterior distribution.

Supporting_Information.S11.txt: Summary results from a MEDUSA run on the MCC tree from the 1000 random trees selected from the posterior distribution.

Supporting_Information.S12.pdf: Figure for MEDUSA run on MCC tree from random 1000 trees.

Supporting_Information.S13.txt: Summary results from a MultiMEDUSA run on the 1000 random trees selected from the posterior distribution.

Supporting_Information.S14.pdf: Probability of nodes with estimated rates from a MultiMEDUSA run on the 1000 random trees selected from the posterior distribution.

Supporting_Information.S15.csv: Hostplants of Nymphalidae butterflies recorded from the literature.

Supporting_Information.S16.csv: References for hostplants data.

Supporting_Information.S17.csv: Data matrix with character states for hostplant use.

Supporting_Information.S18.R: R code for running the BiSSE analysis.

Supporting_Information.S19.csv: Raw results for the BiSSE analysis.

Supporting_Information.S20.csv: Raw results for the combined BiSSE analysis.