

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

Running title: Diversity dynamics in Nymphalidae butterflies

Authors: Carlos Peña<sup>1,\*</sup> and Marianne Espeland<sup>2,\*</sup>

*1* Laboratory of Genetics, Department of Biology, University of Turku, Turku, Finland. Email: [mycalesis@gmail.com](mailto:mycalesis@gmail.com)

*2* Museum of Comparative Zoology and Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, USA. Email: [marianne.espeland@gmail.com](mailto:marianne.espeland@gmail.com)

Corresponding author: [mycalesis@gmail.com](mailto:mycalesis@gmail.com)

Comments: 6283 words, 7 figures, 2 tables and 15 supplementary material files.

Both authors contributed equally to this work.

Data Archival Location: <http://dx.doi.org/10.6084/m9.figshare.639208>

**Pre-peer review:** This manuscript has been peer-reviewed by two anonymous referees using Peerage of Science. The peer-review process can be accessed via <https://www.peerageofscience.org/?link=71445> using the following credentials:

## Abstract

The speciose butterfly family Nymphalidae has been used to develop hypotheses explaining evolutionary interactions between plants and insects. Theories of insect-hostplant dynamics predict accelerated diversification in some scenarios. 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 three out of 13 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 responsible for the increased diversification rate in Ithomiini, congruent with the diffuse cospeciation hypothesis. Our results show that taking phylogenetic uncertainty into account when estimating 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 (Mitter et al. 1988). The study of the evolution of hostplant use has spawned several theories explaining

the evolutionary interactions between plants and insects (reviewed by Nyman et al. 2012). The “escape-and-radiate” hypothesis (Ehrlich and Raven 1964) states that plants and herbivore butterflies are involved in an evolutionary arms race in which newly acquired adaptive characters in plants act as defense against herbivores. This would allow the plant lineage to escape the herbivorous pressure and diversify. Eventually, the butterflies would also acquire a character to overcome this barrier and radiate onto the available plant resource. An alternative hypothesis of herbivore diversification is the “oscillation hypothesis” (Janz 2011) or “diffuse cospeciation” (Nyman et al. 2012) in which range expansions of plants and insects facilitate allopatric speciation and cross-colonization of insects onto related plants. This hypothesis predicts near-simultaneous diversification of plants and insects that feed on them (Nyman et al. 2012). The “resource abundance-dependent diversity dynamics” hypothesis states that plant groups that are common and widely distributed will host a higher diversity of herbivores by facilitating their speciation over wider geographic distributions. According to this hypothesis, there should be a time lag between the diversification of hostplants and insects that feed on them (Nyman et al. 2012).

The butterfly family Nymphalidae has been an important taxon for developing some of the mentioned hypotheses. Nymphalidae contains around 6000 species (van Nieukerken et al. 2011), 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 (Heikkilä et al. 2012). Several studies have used 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 (Elias et al. 2009; Wahlberg et al. 2009; Fordyce 2010; Heikkilä et al. 2012). 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 (Peña and Wahlberg 2008) due to the abundance of grasses over extensive geographic areas (“resource abundance-dependent diversity dynamics” hypothesis). Fordyce (2010) found increased diversification rates in some Nymphalidae lineages after a major hostplant shift, which appears to be in agreement with Ehrlich and Raven (1964) “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 (“radiate-and-escape” hypothesis, e.g. Fordyce 2010), climatic events and shifts to closely related hostplants (“diffuse cospeciation hypothesis” e.g. McLeish et al. 2007; Voje et al. 2009) or resource-abundance dynamics hypothesis (e.g. Peña and Wahlberg 2008). If patterns of diversification bursts are identified, it is necessary to test whether this is caused by a shift onto chemically different hostplant groups, climatic change, geography, ecological opportunity, or other factors.

In this study, we used a time calibrated genus-level phylogenetic hypothesis for Nymphalidae butterflies (taken from Wahlberg et al. 2009) to investigate patterns of diversification. We applied MEDUSA (modelling evolutionary diversity using stepwise AIC, Alfaro et al. 2009; Harmon et al. 2011), a recently developed statistical method, to study the diversification pattern of Nymphalidae butterflies. Particularly, we wanted to study the effects of phylogenetic uncertainty and modified the current MEDUSA method to take this into account (Multi-MEDUSA). 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 (FitzJohn 2012).

## Methods

### Data

For analyses, we used the phylogenetic trees from the study of Wahlberg et al. (2009) that were generated using DNA sequence data from 10 gene regions for 398 of the 540 valid genera in Nymphalidae. We employed Wahlberg et al. (2009) maximum clade credibility tree (MCC tree) (supp. mat. 01) 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. 02) in order to correct for phylogenetic uncertainty when performing the diversification analyses.

We compiled species richness data for Nymphalidae genera from several sources including the specialist-curated lists on <http://tolweb.org>, Lamas (2004) 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 Wahlberg et al. (2009) study (supp. mat. 03).

Hostplant data for Nymphalidae species were compiled from several sources including Ackery (1988), HOSTS database (<http://bit.ly/YI7nwW>), Dyer and Gentry (2002) and others (supp. mat. 04–05) 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 (supp. mat. 04).

### Analyses of Diversification

We used the statistical software R version 2.15.1 (Core Team 2013) in combination with the APE (Paradis et al. 2004), GEIGER (Harmon et al. 2008) and diversitree (FitzJohn 2012) packages along with our own scripts to perform the analyses (included as supplementary materials). All analyses were run on the 1000 random trees from Wahlberg et al. (2009) as well as on the maximum clade credibility tree.

### LTT plots

We obtained semilogarithmic lineages-through-time (LTT) plots after removing outgroups and including the maximum clade credibility tree and 95% credibility interval obtained by random sampling of 1000 trees from the posterior distribution of the BEAST run after burnin from Wahlberg et al. (2009).

### Detecting diversification shifts on phylogenetic trees

We analyzed patterns of diversification in Nymphalidae by using Turbo-MEDUSA version 0.1 (Harmon et al. 2011) on the maximum clade credibility tree from Wahlberg et al. (2009). Turbo-MEDUSA, and the original algorithm MEDUSA, fit 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 (**b**: speciation and **d**: extinction). 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 (Alfaro et al. 2009). To our knowledge, studies using MEDUSA and Turbo-MEDUSA have so far 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 diversification rate shifts and therefore used custom-made scripts to run Turbo-MEDUSA across 1000 random genus-level trees from the posterior distribution (Multi-MEDUSA, supp. mat. 06) and summarize the estimated changes in diversification rates for nodes across all trees. Patterns of change in diversification rates are significant if they are found at the same node in at least 95% 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. We let MEDUSA estimate up to 25 turnover points in our trees as the maximum number of shifts found in all trees was only 21.

We also tested for a possible effect of non-random sampling of genera on the estimated speciation and extinction rates of our MCC tree. It has been found that the estimation of diversification rates can be biased in incomplete phylogenies if species were sampled non-randomly and rates are estimated assuming random sampling. Stadler and Bokma (2013) presented a method to estimate birth and death rates in the constant birth-death model for strict higher level phylogenies by applying a cutoff limit and collapsing the subclades younger than the cutoff into existing older clades. We defined a strict higher level phylogeny by applying a cutoff limit at 24 MYA (Fig. S1). The number of species for each clade were recalculated and assigned to remaining clades. Speciation and extinction rates were estimated for the resulting tree following the methods in Stadler and Bokma (2013).

## **Effect of the quality of posterior distribution of trees on MEDUSA and Multi-MEDUSA**

We tested the effect of topology uncertainty and width of confidence intervals of ages on the number of diversification splits inferred by the MEDUSA and Multi-MEDUSA approaches. We created a thousand sets of 1000 trees with increasingly better support for the topology by replacing each of the trees in the posterior distribution with our MCC tree. We measured the degree of topology uncertainty for each set as the ratio of nodes with high posterior probability ( $> 0.95$ ) on the MCC tree of each set of thousand trees. We run Multi-MEDUSA on all the sets of trees, and counted the number of consistently recovered diversification splits inferred across more than 95% of the trees.

To test for an effect of confidence interval widths of node ages, we created a thousand sets of 1000 trees with narrower confidence intervals on the estimated ages for all the nodes. We achieved this by running a BEAST analysis and forcing a fixed topology (our MCC tree) so different branch lengths were allowed. We randomly selected 1000 trees from the posterior distribution of trees after burning (supp. mat. 07) and replaced each of the trees with the MCC tree (supp. mat. 08) to create a thousand sets of 1000 trees. We plotted the relationship between width of confidence intervals and the number diversification splits inferred across more than 95% of the posterior distribution of trees.

## Estimation of trait-dependent speciation rates

We tested whether diversification of Nymphalidae lineages is driven by hostplant by using the “binary state speciation and extinction” (BiSSE; Maddison, Midford, Otto, 2007) Bayesian approach as implemented in the R package diversitree (FitzJohn 2012). MuSSE (FitzJohn 2012) 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 diversification rate for a clade (Maddison et al. 2007). We used our compiled data of hostplant use to produce binary datasets for the characters “feeding on the plant family Solanaceae or Apocynaceae” (supp. mat. 09) which are the main hostplants of the diverse Ithomiini butterflies and closest relatives (Willmott and Freitas 2006) and Danaini (Apocynaceae). 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 likelihood ratio tests to find out whether the hypothesis of effect on diversification has a significantly better likelihood than the null hypothesis (no effect). Due to computational constraints, we could not run BiSSE on the 1000 trees from the posterior distribution. However, we were able to run it on a sample of 250 trees.

## Results

### LTT plot

LTT plots of the 1000 randomly chosen trees, along with the maximum clade credibility tree, from the posterior distribution of trees from Wahlberg et al. (2009) are shown in Figure 1. This figure shows that the accumulation of lineages in Nymphalidae is congruent with a pattern of increase in diversification rate starting at around 50 MYA. However, it should be noted that we used a genus level phylogeny with very incomplete sampling at the species level. Thus, the slowdown of diversification that is apparent during the last 20 MYA is an artifact and should be interpreted very carefully.

### Detecting diversification shifts on the maximum clade credibility tree

The MEDUSA analysis on the MCC tree in combination with richness data estimated 13 changes in the tempo of diversification in Nymphalidae history (Figure 2; Table 1; supp. mat. 10). The estimated corrected threshold of AICc scores for selecting the optimal model was estimated as 7.8 units. In all MEDUSA analyses, the maximum number of inferred diversification splits in all trees was 21. Thus a `model.limit` of 25 splits was adequate enough. The background diversification rate for Nymphalidae was estimated as  $r = 0.081$  lineages per Million of years and the AICc score for the best fit model was 5158.6 (Table 1). MEDUSA also estimated the speciation and extinction rates under the basic constant birth–death model (AICc = 5449.3,  $\lambda = 0.3405$ ,  $\mu = 0.2664$ ). These values are lower than the estimated rates for the constant birth–date model estimated on a strict higher level phylogeny of Nymphalidae as defined by Stadler and Bokma (2013) and fall outside the 95% confidence interval (Fig S2;  $\lambda = [0.35\text{--}0.56]$ ;  $\mu = [0.27\text{--}0.51]$ ).

Some of the 13 changes in diversification correspond to rate increases in very species-rich genera: *Ypthima* ( $r = 0.264$ ), *Charaxes* ( $r = 0.251$ ), *Callicore* + *Diaethria* ( $r = 0.220$ ), *Pedaliodes* ( $r = 0.196$ ) and *Taenaris* ( $r = 0.238$ ). We found rate increases for other clades as well such as: Lethina + Mycalesina ( $r = 0.130$ ), Oleriina + Ithomiina + Napeogenina + Dircennina + Godyrina ( $r = 0.181$ ), Euptychiina + Pronophilina + Maniolina + Satyrina ( $r = 0.114$ ), Phyciodina in part ( $r = 0.227$ ) and Satyrina ( $r = 0.220$ ).

### Phylogenetic uncertainty in the Multi-MEDUSA approach

We found that the analyses by MEDUSA on the 1000 trees did not estimate the same diversification shifts as in the MCC tree (the background diversification and all shifts found by MEDUSA on the 1000 trees are provided in supp. mat. 11). In order to obtain the diversification shifts that were estimated in most of the 1000 trees, we plotted the diversification shifts versus number of trees containing that particular diversification shift (Figure 3, supp. mat. 12) as estimated by MEDUSA. There were three diversification shifts found in more than 95% of the trees: (i) diversification rate increase in the genus *Charaxes*; (ii) rate increase in Ithomiini subtribes Oleriina + Ithomiina + Napeogenina + Dircennina + Godyrina, and (iii) slowed diversification in part of Danaini (Figure 3).

We obtained mean and standard deviation statistics for the diversification values found on the shifts on the 1000 trees (supp. mat. 11). We found that some of the changes in diversification rate values had great variation across the posterior distribution of trees. A boxplot of the 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 (Figure 4). This variation is especially wide for the clade formed by the genera *Magneuptychia* and *Caeruleuptychia* because MEDUSA estimated diversification values from three times the background diversification rate ( $r = 0.2755$ ) to almost zero ( $r = 2.4e-07$ ). The diversification rates estimates for the root (background diversification rate) and the clades (*Tirumala* + *Danaus* + *Amauris* + *Parantica* + *Ideopsis* + *Euploea* + *Idea*) and (Oleriina + Ithomiina + Napeogenina + Dircennina + Godyrina) are relatively consistent across the 1000 trees (Figure 4). 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 (Euptychiina + Pronophilina + Satyrina + Maniolina) (see supp. mat. 12).

### MEDUSA performance due to phylogenetic uncertainty

We found that most of the diversification splits were consistently recovered across more than 95% of the trees from the posterior distribution when there was a high ratio of nodes with high posterior probability (Figure 5). For our data, it would be necessary to obtain a set of trees from the Bayesian run with almost 99% of the nodes with posterior probability higher than 0.95 (almost no phylogenetic uncertainty) in order to obtain most of the diversification splits from the MCC tree on the posterior distribution of trees. For example the MCC tree of the set of trees number 950 had 99.75% of the nodes with posterior probability > 0.95, however it was possible to recover only 12 diversification shifts (out of a total of 14) across more than 95% of the trees (supp. mat. 13–14).

The figure 6 shows that the width of confidence intervals on the estimated ages of diversification is correlated with the number of diversification splits that are consistently recovered across the posterior distribution of

trees. Very narrow confidence intervals are needed in order to recover all 14 diversification splits across more than 95% of trees from the posterior distribution (supp. mat. 15).

### Estimation of trait-dependent speciation rates

The MEDUSA analyses taking into account phylogenetic uncertainty estimated a 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 diversification rate on part of Ithomiini (Oleriina + Ithomiina + Napeogenina + Dircennina + Godyrina) (Figure 5). The Markov Chain Monte Carlo algorithm was run for 10000 generations discarding the first 7500 as burnin. The estimated mean diversification rate for taxa that do not feed on Solanaceae was  $r = 0.10$  while the diversification rate for the Solanaceae feeders was  $r = 0.14$  (see Figure 6 for a boxplot of speciation and extinction values for the 95% credibility intervals). 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 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 2; character states available in supp. mat. 07, code in supp. mat. 11, and mcmc run in supp. mat. 12). 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. 13; profiles plot of speciation rates in supp. mat. 14; boxplot of 95% credibility intervals in supp. mat. 15). A BiSSE analysis to test whether the trait “feeding on Apocynaceae” had any effect on increased diversification rates found similar speciation rates for lineages feeding on Apocynaceae and other plants (Figure 7). It has been shown that BiSSE performs poorly under certain conditions (Davis et al. 2013). 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 (Davis et al. 2013).

## Discussion

### Effects of phylogenetic uncertainty on the performance of MEDUSA

The MEDUSA method has been used to infer changes in diversification rates along a phylogenetic tree. Since its publication (Alfaro et al. 2009) the results of using MEDUSA on a single tree, the maximum clade credibility tree, have been used for generation of hypotheses and discussion (e.g. Litman et al. 2011; Heikkilä et al. 2012; Ryberg and Matheny 2012). 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 in a very small percentage of the 1000 randomly sampled trees from the posterior distribution (Figure 3). We also found that, even though MEDUSA could estimate the same diversification splits on two or more trees, the estimated diversification rates could vary widely (Figure 4). For example, in our Nymphalidae trees, we found that the split for *Magneptychia* and *Caeruleptychia* had a variation

from  $r = 0.2755$ , higher than the background 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 (Peña et al. 2011). 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 (0.5 to 0.6; Wahlberg et al. 2009). As a result, MEDUSA inferred a diversification rate increase for part of Satyrini in the maximum clade credibility tree, but this was recovered only in 17% of the trees from the posterior distribution.

We run Multi-MEDUSA on sets of thousand trees that gradually decrease in phylogenetic uncertainty (degree of conflicting topology and width of confidence intervals). It appears that the number of nodes with high posterior probability and width of confidence intervals for the estimated ages of diversification on the MCC tree greatly affect the number of diversification splits from the MCC tree that are consistently found on most trees from the posterior distribution. In order to recover most diversification splits, it is necessary to obtain a set of trees from a BEAST run with almost no phylogenetic uncertainty, i.e. very narrow confidence intervals and almost no conflicting topologies so that most of the nodes are recovered with high posterior probability.

If there is strong phylogenetic signal for increases or decreases in 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 diversification rate splits in more than 95% of the posterior distribution of trees for only three splits: the genus *Charaxes*, part of Danaini and part of Ithomiini (Figure 3), while estimating splits for other lineages in only a fraction of the posterior distribution of trees.

The clade Ithomiini and the non-basal danuids are well supported by high posterior probability values in Wahlberg et al. (2009, supplementary information, fig. XX3S). Therefore our MEDUSA analyses recovered an increase in diversification rate in more than 95% of the posterior distribution of trees (Figure 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 (Brown Jr. 1987). Ithomiini butterflies are exclusively Neotropical and most species feed on Solanaceae hostplants during larval stage (supp. mat. 04; Willmott and Freitas 2006). Optimizations of the evolution of hostplant use on phylogenies evidence a probable shift from Apocynaceae to Solanaceae (Willmott and Freitas 2006; Brower et al. 2006). Fordyce (2010) 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.

In this study, we investigated whether the strong signal for an increase in 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 (BiSSE; FitzJohn et al. 2009) 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, shows a significantly higher net diversification rate for Ithomiini taxa, which can be attributed to the trait “feeding on Solanaceae hostplants” (Figure 5). This is in agreement with the findings of Fordyce (2010) using other statistical methods. Due to the fact that Ithomiini are virtually the only nymphalids using Solanaceae as hostplants (except for *Hypanartia*, *Vanessa* and *Acraea*; supp. mat.04), it is possible that the trait responsible for a higher diversification of Ithomiini might not be the hostplant character. As noted by Maddison et al. (2007), 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 (Fordyce 2010) and radiation scenarios among chemically different lineages of Solanaceae plants (Brown Jr. 1987; Willmott and Freitas 2006). 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. A detoxification mechanism has been found in Pieridae larvae feeding on Brassicales hosts (Wheat et al. 2007).

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 (Ehrlich and Raven 1964; Janz 2011; Nyman et al. 2012). Wheat et al. (2007) 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 (Bell et al. 2010), the family Solanaceae split from its sister group about 59 (49-68) MYA and diversification started (crown group age) around 37 (29-47) MYA. Wahlberg et al. (2009) 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 is congruent with the diffuse cospeciation hypothesis.

The uplift of the Andes was a major tectonic event that underwent higher activity during the Oligocene (Somoza 1998). 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 (Hoorn et al. 2010) including Ithomiini butterflies and Solanaceae hostplants. Moreover, all Solanaceae clades currently present in New World originated in South America (Olmstead 2013) as well as Ithomiini butterflies (Wahlberg et al. 2009). This suggests a process of “diffuse cospeciation” of Ithomiinae and hostplants.

**Danaini** Our Multi-MEDUSA approach gave a significant slowdown in 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 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 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 (Wahlberg and Rubinoff 2011).

**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 (Ackery 1988) and some species appear to be polyphagous (Müller et al. 2010). Aduse-Poku et al. (2009) 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. Müller et al. (2010) 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 (Peña and Wahlberg 2008). Thus, it is somewhat surprising that part of Satyrini (the subtribes Euptychiina, Satyrina and Pronophilina) were found to have accelerated diversification in only 17% of the trees from the posterior distribution. Although this can be attributed to low phylogenetic signal (posterior probability value = 0.6 for this clade in Wahlberg et al. 2009), the clade Satyrini is very robust (posterior probability value = 1.0 for this clade in Wahlberg et al. 2009) and MEDUSA failed to identify any significant accelerated 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 (Peña et al. 2011) 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.

## Conclusions

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 95% 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 Multi-MEDUSA 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 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 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 Multi-MEDUSA approach to perform analyses on the posterior distribution of trees found strong support for an increase in diversification rate in the tribe Ithomiini and the genus *Charaxes*, and for a decrease in 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 found that the trait “feeding on Solanaceae”, or a codistributed character, is responsible for the higher diversification rate of Ithomiini, but 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 (Janz 2011; Nyman et al. 2012).

## 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, to Folmer Bokma for in using the strict high level phylogeny method. 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. Lineage through time plot for the maximum clade credibility tree (black) and 1000 random trees from the posterior distribution (coloured lines) of the nymphalid phylogeny. Since this is a genus level phylogeny, the observed slowdown of diversification in the last 20 MYA is an artifact and should be disregarded.

Figure 2. Results of the MEDUSA analysis run on the maximum clade credibility tree. Rate shifts were estimated for the following nodes (besides the background rate): 2) Limenitidinae + Heliconiinae, 3) Ypthima, 4) Charaxes, 5) Mycalesina + Lethina, 6) Ithomiini in part, 7) Satyrini in part, 8) Phyciodina in part, 9) Danaini in part, 10) Caeruleptychia + Magneptychia, 11) Satyrina, 12) Callicore + Diaethria, 13) Pedaliodes, 14) Taenaris.

Figure 3. Results of the Multi-MEDUSA analysis on 1000 random trees from the posterior distribution of the Nymphalidae phylogeny. Bars show the number of trees where MEDUSA found significant increases or decreases in diversification rates for tips or clades in Nymphalidae. Tip and clade indexes are detailed in supplementary material 10.

Figure 4. Boxplot of the range of diversification values for tips or clades estimated by MEDUSA on the 1000 random trees from the posterior distribution of the Nymphalidae phylogeny. The tips or clades shown are those present on the maximum clade credibility tree.

Figure 5. BiSSE analysis of diversification of nymphalids due to feeding on Solanaceae hostplants. Speciation and diversification rates are significantly higher in Solanaceae feeders.

Figure 6. Boxplot of speciation ( $\lambda$ ), extinction ( $\mu$ ) and transition ( $q$ ) parameter values of the BiSSE analysis on diversification due to feeding on Solanaceae hostplants.

Figure 7. BiSSE analysis of diversification of nymphalids due to feeding on Apocynaceae hostplants. There is no effect of this trait either on speciation or diversification rates.

Table 1. Significant diversification rate shifts found in the Turbo-MEDUSA analysis of the Nymphalid maximum clade credibility tree. Split.Node = node number,  $r$  = net diversification rate,  $\epsilon$  = relative extinction rate, LnLik.part = log likelihood value

Table 2. Likelihood ratio test between the model of increased diversification of nymphalids feeding on Solanaceae against a model forcing equal speciation rates (no effect on diversification).

Figure S1. MCC tree defined as strict higher level phylogeny but collapsing all clades younger than the cutoff value at 24 MYA.

Figure S2. Boxplot of the 95% confidence intervals for estimated speciation and extinction rates on an enforced strict higher level phylogeny of our MCC tree by collapsing all clades younger than the cutoff value at 24 MYA (see figure S1). Estimated rates using the method in Stadler and Bokma (2013).

## References

- Ackery P.R.** 1988. Hostplants and classification: a review of nymphalid butterflies. *Biological Journal of the Linnean Society*. 33:95–203.
- 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. *Molecular Phylogenetics and Evolution*. 53:463–478.
- Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon L.J.** 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*. 106:13410–13414.
- Bell C.D., Soltis D.E., Soltis P.S.** 2010. The age and diversification of the angiosperms re-revisited. *American Journal of Botany*. 97:1296–1303.
- Brower A.V.Z., Freitas A.V.L., Lee M.-M., Silva-Brandão K.L., Whinnett A., Willmott K.R.** 2006. Phylogenetic relationships among the Ithomiini (Lepidoptera: Nymphalidae) inferred from one mitochondrial and two nuclear gene regions. *Systematic Entomology*. 31:288–301.
- Brown Jr. K.S.** 1987. Chemistry at the Solanaceae/Ithomiinae Interface. *Annals of the Missouri Botanical Garden*. 74:359–397.
- Core Team R.** 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing
- Davis M.P., Midford P.E., Maddison W.** 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evolutionary Biology*. 13:38.
- Dyer L.A., Gentry G.L.** 2002. Caterpillars and parasitoids of a tropical lowland wet forest.

- Ehrlich P.R., Raven P.H.** 1964. Butterflies and plants: a study in coevolution. *Evolution*. 18:586–608.
- Elias M., Joron M., Willmott K.R., Silva-Brandão K.L., Kaiser V., Arias C.F., Gomez Piñerez L.M., Uribe S., Brower A.V.Z., Freitas A.V.L., Jiggins C.D.** 2009. Out of the Andes: patterns of diversification in clearwing butterflies. *Molecular Ecology*. 18:1716–1729.
- FitzJohn R.G.** 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*. 3:1084–1092.
- FitzJohn R.G., Maddison W.P., Otto S.P.** 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology*. 58:595–611.
- Fordyce J.A.** 2010. Host shifts and evolutionary radiations of butterflies. *Proceedings of the Royal Society B*. 277:3735–3743.
- Harmon L.J., Rabosky D.L., FitzJohn R.G., Brown J.W.** 2011. turboMEDUSA: MEDUSA: Modeling Evolutionary Diversification Using Stepwise AIC.
- Harmon L.J., Weir J.T., Brock C.D., Glor R.E., Challenger W.** 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics*. 24:129–131.
- Heikkilä M., Kaila L., Mutanen M., Peña C., Wahlberg N.** 2012. Cretaceous origin and repeated tertiary diversification of the redefined butterflies. *Proceedings of the Royal Society B: Biological Sciences*. 279:1093–9.
- Hoorn C., Wesselingh F.P., ter Steege H., Bermudez M.A., Mora A., Sevink J., Sanmartín I., Sanchez-Meseguer A., Anderson C.L., Figueiredo J.P., Jaramillo C., Riff D., Negri F.R., Hooghiemstra H., Lundberg J., Stadler T., Särkinen T., Antonelli A.** 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*. 330:927–931.
- Janz N.** 2011. Ehrlich and Raven Revisited: Mechanisms Underlying Codiversification of Plants and Enemies. *Annual Review of Ecology, Evolution, and Systematics*. 42:71–89.
- Lamas G.** 2004. Checklist: Part 4A. Hesperioidea-Papilionoidea. Gainesville:
- Litman J.R., Danforth B.N., Eardley C.D., Praz C.J.** 2011. Why do leafcutter bees cut leaves? New insights into the early evolution of bees. *Proceedings of the Royal Society B*. 278:3593–3600.
- Maddison W.P., Midford P.E., Otto S.P.** 2007. Estimating a binary character’s effect on speciation and extinction. *Systematic Biology*. 56:701–710.
- McLeish M.J., Chapman T.W., Schwarz M.P.** 2007. Host-driven diversification of gall-inducing Acacia thrips and the aridification of Australia. *BMC Biology*. 13:1–13.
- Mitter C., Farrell B., Wiegmann B.** 1988. The phylogenetic study of adaptive zones: Has phytophagy promoted insect diversification? *The American Naturalist*. 132:107–128.
- Müller C.J., Wahlberg N., Beheregaray L.B.** 2010. “After Africa:” the evolutionary history and systematics of the genus *Charaxes* Ochsenschlager (Lepidoptera: Nymphalidae) in the Indo-Pacific region. *Biological Journal of the Linnean Society*. 100:457–481.
- van Nieukerken E.J., Kaila L., Kitching I.J., Kristensen N.P., Lees D.C., Minet J., Mitter C., Mutanen M., Regier C., Simonsen T.J., Wahlberg N., Yen S.-h., Zahiri R., Adamski D.,**

- Baixeras J., Bartsch D., \textbackslashAA B., Brown J.W., Bucheli S.R., Davis D.R., Prins J.D., Prins W.D., Epstein M.E., Gentili-poole P., Hättenschwiler P., Hausmann A., Holloway J.D., Kallies A., Karsholt O., Kawahara A.Y., Koster S.J.C., Kozlov M.V., Lafontaine J.D., Lamas G., Landry J.-F., Lee S., Nuss M., Park K.-t., Penz C., Rota J., Schintlmeister A., Schmidt B.C., Sohn J.-c., Solis M.A., Tarmann G.M., Warren A.D. 2011. Order Lepidoptera. Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. *Zootaxa*. p. 212–221.
- Nyman T., Linder H.P., Peña C., Malm T., Wahlberg N. 2012. Climate-driven diversity dynamics in plants and plant-feeding insects. *Ecology Letters*. 15:889–898.
- Olmstead R.G. 2013. Phylogeny and biogeography in Solanaceae, Verbenaceae and Bignoniaceae: a comparison of continental and intercontinental diversification patterns. *Botanical Journal of the Linnean Society*. 171:80–102.
- Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. 20:289–290.
- Peña C., Nylin S., Wahlberg N. 2011. The radiation of Satyrini butterflies (Nymphalidae: Satyrinae): a challenge for phylogenetic methods. *Zoological Journal of the Linnean Society*. 161:64–87.
- Peña C., Wahlberg N. 2008. Prehistorical climate change increased diversification of a group of butterflies. *Biology Letters*. 4:274–278.
- Ryberg M., Matheny P.B. 2012. Asynchronous origins of ectomycorrhizal clades of Agaricales. *Proceedings of the Royal Society B*. 279:2003–2011.
- Somoza R. 1998. Updated Nazca (Farallon) - South America relative motions during the last 40 My: implications for mountain building in the central Andean region. *Journal of South American Earth Sciences*. 11:211–215.
- Stadler T., Bokma F. 2013. Estimating Speciation and Extinction Rates for Phylogenies of Higher Taxa. *Systematic Biology*. 62:220–230.
- Voje K.L., Hemp C., Flagstad Ø., Saetre G.-P., Stenseth N.C. 2009. Climatic change as an engine for speciation in flightless Orthoptera species inhabiting African mountains. *Molecular Ecology*. 18:93–108.
- Wahlberg N., Leneveu J., Kodandaramaiah U., Pena C., Nylin S., Freitas A.V.L., Brower A.V.Z. 2009. Nymphalid butterflies diversify following near demise at the Cretaceous/Tertiary boundary. *Proceedings of the Royal Society B: Biological Sciences*. 276:4295–4302.
- Wahlberg N., Rubinoff D. 2011. Vagility across Vanessa (Lepidoptera: Nymphalidae): mobility in butterfly species does not inhibit the formation and persistence of isolated sister taxa. *Systematic Entomology*. 36:362–370.
- Wheat C.W., Vogel H., Wittstock U., Braby M.B., Underwood D., Mitchell-Olds T. 2007. The genetic basis of a plant-insect coevolutionary key innovation. *PNAS*. 104:20427–20431.
- Willmott K.R., Freitas A.V.L. 2006. Higher-level phylogeny of the Ithomiinae (Lepidoptera: Nymphalidae): classification, patterns of larval hostplant colonization and diversification. *Cladistics*. 22:297–368.