Running head

TESTING THE RED QUEEN AND COURT JESTER

Title

Testing the role of the Red Queen and Court Jester as drivers of the macroevolution of Apollo butterflies

Authors

FABIEN L. CONDAMINE[1,2,3]*, JONATHAN ROLLAND[4], SEBASTIAN HÖHNA[5], FELIX A. H. SPERLING[3]† AND ISABEL SANMARTÍN[2]†

Authors' addresses

[1] *CNRS, UMR 5554 Institut des Sciences de l'Evolution (Université de Montpellier | CNRS | IRD | EPHE), Montpellier, France;*

[2] *Department of Biodiversity and Conservation, Real Jardín Botánico, CSIC, Plaza de Murillo, 2; 28014 Madrid, Spain;*

[3] *Department of Biological Sciences, University of Alberta, Edmonton T6G 2E9, AB, Canada;*

[4] *Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland;*

[5] *Division of Evolutionary Biology, Ludwig-Maximilian-Universität München, Grosshaderner Strasse 2, Planegg-Martinsried 82152, Germany*

† Co-senior authors.

Corresponding author (*): Fabien L. Condamine, CNRS, UMR 5554 Institut des Sciences de l'Evolution (Université de Montpellier), Place Eugène Bataillon, 34095 Montpellier, France. Phone: +336 749 322 96 | E-mail: fabien.condamine@gmail.com

***Abstract.* –** In macroevolution, the Red Queen (RQ) model posits that biodiversity dynamics depend mainly on species-intrinsic biotic factors such as interactions among species or life-history traits, while the Court Jester (CJ) model states that extrinsic environmental abiotic factors have a stronger role. Until recently, a lack of relevant methodological approaches has prevented the unraveling of contributions from these two types of factors to the evolutionary history of a lineage. Here we take advantage of the rapid development of new macroevolution models that tie diversification rates to changes in paleoenvironmental (extrinsic) and/or biotic (intrinsic) factors. We inferred a robust and fully-sampled species-level phylogeny, as well as divergence times and ancestral geographic ranges, and related these to the radiation of Apollo butterflies (Parnassiinae) using both extant (molecular) and extinct (fossil/morphological) evidence. We tested whether their diversification dynamics are better explained by a RQ or CJ hypothesis, by assessing whether speciation and extinction were mediated by diversity-dependence (niche filling) and clade-dependent host-plant association (RQ) or by large-scale continuous changes in extrinsic factors such as climate or geology (CJ). For the RQ hypothesis, we found significant differences in speciation rates associated with different host-plants but detected no sign of diversity-dependence. For CJ, the role of Himalayan-Tibetan building was substantial for biogeography but not a driver of high speciation, while positive dependence between warm climate and speciation/extinction was supported by continuously varying maximum-likelihood models. We find that rather than a single factor, the joint effect of multiple factors (biogeography, species traits, environmental drivers, and mass extinction) is responsible for current diversity patterns, and that the same factor might act differently across clades, emphasizing the notion of opportunity. This study confirms the importance of the confluence of several factors rather than single explanations in modeling diversification within lineages.

[Diversification; extinction; Himalayan orogeny; historical biogeography; host-plant shifts; integrative study; mountain building; Papilionidae; past climate change; speciation]

INTRODUCTION

Evolutionary biologists have long endeavored to determine which factors govern biodiversity dynamics, aiming to answer questions such as why some clades have diversified more than others, or why some lineages are widely distributed whereas others survive in restricted ranges (Ezard et al. 2016). Two different mechanistic macroevolutionary models have been proposed to explain the generation and maintenance of diversity. The Red Queen (RQ) model (Van Valen 1973), which stems from Darwin and Wallace, posits that diversification is driven by species-intrinsic, biotic factors such as interactions among species, species ecology, or life-history traits. The Court Jester (CJ) model, which builds on paleontological evidence (Barnosky 2001), argues that diversification dynamics result from historical abiotic forces such as abrupt changes in climate or geological tectonic events that drive speciation and extinction rates, usually acting clade-wide across lineages.

The CJ and RQ models are generally considered the two extremes of a continuum, operating over different geographic and temporal scales. Biotic factors such as species interactions shape ecosystems locally over short time spans, whereas abiotic factors such as climate and tectonic events shape large-scale patterns regionally and globally over millions of years (Benton 2009). However, biotic interactions can also be observed at large spatial and temporal scales (Liow et al. 2015; Silvestro et al. 2015), while Van Valen's (1973) original RQ hypothesis is now interpreted as accepting the role of a changing environment in shaping species evolution (Voje et al. 2015).

Although both abiotic (environmental) and biotic (species-intrinsic) drivers are recognized as fundamental for regulating biodiversity (Ezard et al. 2011), these two types of factors are often studied in isolation (Drummond et al. 2012a; Bouchenak-Khelladi et al. 2015; Lagomarsino et al. 2016), searching for correlations between shifts in diversification rates and the evolution of key innovations or the appearance of key opportunities (Maddison

et al. 2007; Alfaro et al. 2009; Rabosky 2014; Givnish et al. 2015). However, often no single factor but a confluence of biotic and abiotic factors is responsible for the diversification rate shift (Donoghue and Sanderson, 2015), and there could be an interaction effect. For example, C_4 grasses appeared in the Eocene but their expansion and explosive diversification started only after mid-Miocene aridification in Africa and Central Asia (Edwards et al. 2010).

Statistical assessment of the relative contributions of abiotic and biotic factors underlying diversity patterns has been made possible by the development of new probabilistic models in the field of diversification dynamics (Stadler 2013; Morlon 2014; Höhna 2015). One type of model estimates diversification rates that are clade-dependent and identifies differences in diversification rates among clades that can be explained by key innovations (Alfaro et al. 2009; Morlon et al. 2011; Rabosky et al. 2013), or by diversity-dependence and niche filling (i.e. diversification decreasing as the number of species increases, Rabosky and Lovette 2008; Etienne et al. 2012). A second type of model aims to detect statistical associations between diversification rates and changes in species traits (trait-dependent diversification models, Maddison et al. 2007; Ng and Smith 2014), or between geographic evolution and diversification, such as a change in continental connectivity allowing the colonization of a new region and a subsequent increase in allopatric speciation (Goldberg et al. 2011). A third type of model assumes continuous variation in diversification rates over time that depends on a paleoenvironmental variable and investigates whether diversification rates can be affected by abiotic environmental changes (e.g. paleotemperature, Condamine et al. 2013). Finally, episodic birth-death models search for tree-wide rate shifts that act concurrently across all lineages in a tree, for example a mass extinction event removing a fraction of lineages at a certain time in the past (Stadler 2011; Höhna 2015; May et al. 2016).

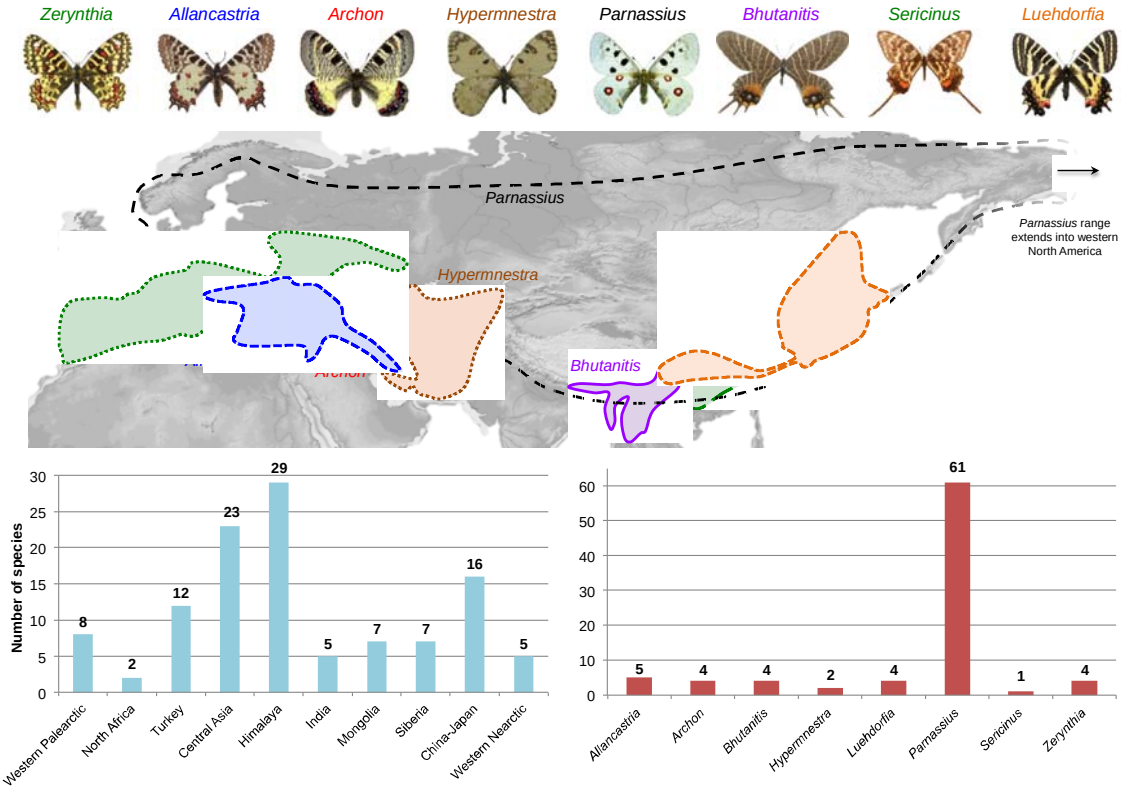
The first two types of models have been used to test RQ-like hypotheses on the effect of biotic interactions, while the other models are often used in the context of the CJ

hypothesis (environmental change). Studies using a subset of these models to address both abiotic and biotic factors are becoming more frequent, but are often used at a local or regional geographic scale (Schnitzler et al. 2011; Drummond et al. 2012a; Jønsson et al. 2012; Hutter et al. 2013; Bouchenak-Khelladi et al. 2015; Lagomarsino et al. 2016), and limited temporally. No study so far has addressed the full set of models using a large (geographic) and long (temporal) scale within the same lineage. Such a study would have substantial power to provide information on questions such as why some lineages diversify and others do not, and the extent that diversification is attributable to trait evolution and/or ecological opportunity (Wagner et al. 2012). It might also shed light on the current debate about whether biodiversity is at an equilibrium bounded by ecological limits (Rabosky 2013), although it is unclear how relevant such limits are compared to other biological explanations (Moen and Morlon 2014). Alternatively, diversity might be unbounded and controlled by rare but drastic environmental changes such as climatic mass extinction events (Antonelli and Sanmartín 2011; Meredith et al. 2011; Kergoat et al. 2014; Condamine et al. 2015a; May et al. 2016).

A key to such a study would be to find a group that: *(i)* has experienced both biotic and abiotic pressures at different temporal and spatial scales; *(ii)* shows large species diversity and high specialization; and *(iii)* has good information available about its taxonomic and evolutionary history. Nearly complete taxon sampling is crucial for accurate estimation of diversification rates (Cusimano and Renner 2010; Brock et al. 2011; Höhna et al. 2011; Davis et al. 2013), whereas rich fossil evidence is important for divergence time estimation and accurate reconstruction of biogeographic history (Meseguer et al. 2015). Phytophagous insects are especially good models because they present complex biotic (trophic) interactions with their host plants (Ehrlich and Raven 1964), and many of these lineages are old enough to have experienced dramatic past climatic changes.

Here, we address the role of biotic (RQ) and abiotic (CJ) factors that drive patterns of species richness in Apollo butterflies (Papilionidae: Parnassiinae). The Parnassiinae comprise eight genera and over 70 species (Ackery 1975; Weiss 1991-2005; Kocman 2009), grouped into three tribes – Luehdorfiini, Zerynthiini, and Parnassiini (Nazari et al. 2007). Apollo butterflies occur from the Western Palearctic to the Western Nearctic (Weiss 1991-2005). Most diversity is concentrated in the Palearctic, where we find two non-monophyletic lowland-flying communities separated by the Himalaya and Tibetan Plateau (HTP, **Figure 1**): an Eastern Palearctic group formed by *Luehdorfia* (Luehdorfiini) and *Bhutanitis* and *Sericinus* (Zerynthiini), and a Western Palearctic community including *Archon* (Luehdorfiini), *Allancastria* and *Zerynthia* (Zerynthiini), and *Hypermnestra* (Parnassiini). The largest genus, *Parnassius*, is mainly distributed in mountainous regions across the Holarctic, with its highest diversity in the HTP (Weiss 1991-2005; Kocman 2009; Nazari et al. 2007; Michel et al. 2008). Compared to the species-poor genera *Allancastria*, *Archon*, *Bhutanitis*, *Hypermnestra*, *Luehdorfia* and *Sericinus* (each with five species maximum), the 50+ species of *Parnassius* have been suggested as an example of rapid diversification. This unbalanced species richness, together with their disjunct pattern of distribution in the Himalayan region, suggests that the HTP orogeny might have been an important driver in the biogeographic and diversification history of Parnassiinae.

Figure 1. Distribution and species richness patterns of Apollo butterflies. Map shows the current separation of west and east communities of Parnassiinae, divided by the Himalayan-Tibetan Plateau. Lower portion of figure compares the current species diversity of each biogeographical unit (which may be non-monophyletic) or genus. Images of Parnassiinae (also in subsequent figures) are created by Fabien Condamine.



152

153 Parnassiinae are also remarkable for their (nearly) strict but diverse host-plant
154 specialization (Ehrlich and Raven 1964). Tribes Luehdorfiini and Zerynthiini feed on
155 Aristolochiaceae, the ancestral host of Parnassiinae (Condamine et al. 2012). Within
156 Parnassiini, *Hypermnestra* feeds on Zygophyllaceae, while *Parnassius* has experienced two
157 host-plant shifts, toward Crassulaceae+Saxifragaceae and Papaveraceae (Michel et al. 2008;
158 Condamine et al. 2012), and is regarded as an adaptive radiation (Rebourg et al. 2006).
159 Finally, Apollo butterflies constitute a model to understand the role of past environmental
160 change in lineage diversification. Phylogeographic studies pinpointed the effect of Pleistocene
161 glaciations on population-level dynamics (e.g. Keyghobadi et al. 2005; Gratton et al. 2008;
162 Schoville and Roderick 2009; Dapporto 2010; Todisco et al. 2010, 2012; Zinetti et al. 2013).
163 Time-calibrated phylogenies support an origin of the subfamily in the Early Cenozoic
164 (Condamine et al. 2012), which means that the lineage has experienced the dramatic cooling
165 and warming events of the last 50 Ma, including a drop in global temperatures at the end of

166 the Eocene (Eocene-Oligocene climate transition, EOCT, Liu et al. 2009) that led to the
167 demise of Boreotropical Holarctic vegetation (Pound et al. 2012).

168 In this study, we reconstruct a completely sampled phylogeny of Parnassiinae to
169 explore the role played by abiotic (CJ) and biotic (RQ) factors, isolated or in unison, in the
170 origin and fate of biological radiations, including paleotemperature, mountain orogeny, host
171 specialization, range expansion and diversity dependence with niche filling. If interactions
172 among species are the dominant drivers of evolution, diversification rates are expected to
173 show diversity-dependent dynamics or be dependent on ecological traits. For example, one
174 might expect diversification rates to decrease as a function of diversity or increase with
175 ecological opportunity. Conversely, if evolution is driven mainly by changes in the physical
176 environment, clade-wide responses due to abrupt abiotic perturbations should dominate
177 macroevolutionary dynamics, e.g. diversification rates may shift following major climatic
178 changes that extirpated certain lineages while favoring the radiation of others. We test these
179 hypotheses using trait-dependent, time-dependent, environment-dependent and episodic birth-
180 death models and compare the fit (explanatory power) of these models using maximum
181 likelihood and Bayesian inference. Our study is the first to address the CJ and RQ hypotheses
182 at such a large scale, and to compare the performance of these models within a single
183 inference framework.

184

185 **MATERIALS AND METHODS**

186 ***Taxon sampling and molecular dataset***

187 The latest effort to resolve phylogenetic relationships in Parnassiinae was Condamine
188 et al. (2012), but that study did not include all species. For our study, DNA sequences were
189 obtained from GenBank using *Taxonomy* browser searches (full dataset is provided in the
190 Supplementary Material that accompanies this article available on Dryad at

191 http://dx.doi.org/10.5061/dryad.[NNNN], **Appendix 1**). Sequences of less than 250
192 nucleotides (possibly microsatellites) and genes sequenced in very few specimens were not
193 included. Overall the dataset comprises >1,800 sequences, representing 69 species and five
194 genes (four mitochondrial and one nuclear) including 730 sequences for cytochrome oxidase I
195 (COI), 214 for NADH dehydrogenase 1 (ND1), 147 for NADH dehydrogenase 5 (ND5), 202
196 for rRNA 16S (16S), and 115 for elongation factor 1 alpha (EF-1α). This large molecular
197 dataset comprises all sequences generated by previous studies on Parnassiinae, especially by
198 F.L.C. and F.A.H.S. (Omoto et al. 2004, 2009; Katoh et al. 2005; Nazari et al. 2007; Nazari
199 and Sperling 2007; Michel et al. 2008; Condamine et al. 2012), and includes all currently
200 described species within the family, except for *Bhutanitis ludlowi* (Möhn 2005; Weiss 2005;
201 Churkin 2006). The dataset also includes the majority of described subspecies or geographic
202 races (Nazari and Sperling 2007; Gratton et al. 2008; Schoville and Roderick 2009; Dapporto
203 2010; Todisco et al. 2010, 2012; Zinetti et al. 2013). Additionally, sequences of nine species
204 of swallowtails were added to the dataset as outgroups, with *Baronia brevicornis* considered
205 as the sister group to all remaining Papilionidae, and two species for each of the four tribes
206 included within Papilioninae, the sister subfamily of Parnassiinae (Condamine et al. 2012).
207 Sequences were aligned using MAFFT 7.110 (Katoh and Standley 2013) with default
208 settings (E-INS-i algorithm). Reading frames of coding genes were checked in Mesquite 3.03
209 (www.mesquite.org). Datasets (species/subspecies names and GenBank accession numbers)
210 are available on Dryad (**Appendices 1 and 2**).
211
212 ***Assessment of fossil positions and calibrations***
213 Constraints on clade ages were enforced through fossil calibrations, whose systematic
214 position was assessed using phylogenetic analyses (Donoghue et al. 1989; Sauquet et al.
215 2012). Phylogenetic analyses of the dataset included both living and fossil taxa and

incorporated both morphological and molecular data using a total-evidence approach (Ronquist et al. 2012a).

Swallowtail fossils are scarce, but two unambiguously belong to Parnassiinae (Nazari et al. 2007). The first is †*Thaites ruminiana* (Scudder 1875), a compression fossil from limestone in the Niveau du gypse d'Aix Formation of France (Bouches-du-Rhone, Aix-en-Provence) within the Chattian (23.03-28.1 Ma) of the late Oligocene (Rasnitsyn and Zherikhin 2002; Sohn et al. 2012). The second is †*Doritites bosniaskii* (Rebel 1898), an exoskeleton and compression fossil from Italy (Tuscany) from the Messinian (5.33-7.25 Ma, late Miocene, Sohn et al. 2012). Absolute ages of geological formations were taken from Gradstein et al. (2012).

To assess the phylogenetic positions of fossils, our morphological dataset comprised 236 characters (Hiura 1980; Nazari et al. 2007) including adult external (160) and internal (54) morphology, egg (2) and larval (10) morphology, pupal morphology (5), and characters for pigment color and chromosome number (5). Morphological characters were coded for all extant species in each genus except *Parnassius* (only eight species: one per subgenus). Yet, *Parnassius* had little impact on our analyses since there are no fossils related to this genus (Nazari et al. 2007). Fossils were only coded for adult external morphology.

Morphological and molecular data were combined to construct a total-evidence phylogeny using MrBayes 3.2.6 (Ronquist et al. 2012b). Morphological data were modeled using the Mk model (Lewis 2001). The best-fitting partitioning scheme for the molecular data was selected with PartitionFinder 1.1.1 (Lanfear et al. 2012), using the *greedy* search algorithm and the Bayesian Information Criterion (BIC). Rather than using a single substitution model per molecular partition, we sampled across the entire substitution rate model space (Huelsenbeck et al. 2004) using the reversible-jump Markov Chain Monte Carlo (rj-MCMC) option. Two independent analyses with one cold and seven heated chains each

were run for 20 million generations, sampled every 2000th. A 50% majority rule consensus tree was built after discarding 25% samples as burn-in. The dataset used for this analysis is available on Dryad (**Appendix 3**).

Phylogeny and estimates of divergence times

We examined whether the rate of molecular evolution evolved in a clock-like pattern using PATHd8 (Britton et al. 2007). Since the hypothesis of a strict molecular clock was rejected for 68 of the 92 nodes (P<0.05), we estimated divergence times using Bayesian relaxed-clock methods accounting for rate variation across lineages (Drummond et al. 2006). MCMC analyses implemented in BEAST 1.8.2 (Drummond et al. 2012b) were employed to approximate the posterior distribution of rates and divergences times and infer their credibility intervals. We set the following settings and priors: a partitioned dataset (after the best-fitting PartitionFinder scheme) was analyzed using the uncorrelated lognormal distribution (UCLD) clock model, with the mean set to a uniform prior between 0 and 1, and an exponential prior (lambda=0.333) for the standard deviation. The branching process prior was set to either a Yule or a birth-death (Gernhard 2008) process (since tree priors may impact estimates of molecular dating, Condamine et al. 2015b), using the following uniform priors: the Yule birth rate and birth-death mean growth rate ranged between 0 and 10 with a starting value at 0.1, and the birth-death relative death rate ranged between 0 and 1 (starting value=0.5).

Calibration priors employed the fossil constraints indicated above (see Results for the phylogenetic positions), using a uniform prior with the minimum age equal to the youngest age of the geological formation where the fossil was found. Additionally, we constrained the crown of Papilionidae with a uniform distribution bounded by a minimum age of 47.8 Ma (Smith et al. 2003; Gradstein et al. 2012) based on two fossils †*Praepapilio colorado* and †*P. gracilis* (Durden and Rose 1978), both from the Lutetian (mid-Eocene) of the Green River

Formation (USA, Colorado). We could not place these fossils phylogenetically because of limited taxon sampling for Papilioninae. Instead, we placed these unambiguous fossils conservatively at the crown of the family as they share synapomorphies with all extant subfamilies (de Jong 2007), and have proven to be reliable calibration points for the crown group (Condamine et al. 2012). All uniform calibration priors were set with an upper bound equal to the estimated age of angiosperms (~140 Ma, *sensu* Magallón et al. 2015), which is six times older than the oldest Parnassiinae fossil. This upper age is intentionally set as ancient to allow exploration of potentially old ages for the clade. Since the fossil record of Lepidoptera is highly incomplete and biased (Sohn et al. 2015), caution is needed in using the few recent fossil calibrations.

To explore the impact of sampled diversity, tree prior, and morphological data on BEAST inferences, we ran four analyses as follows: *(i)* molecular data only and Yule model; *(ii)* molecular data only and birth-death process; *(iii)* molecular and morphological data (total-evidence analysis) and Yule model; and *(iv)* molecular and morphological data (total-evidence analysis) and birth-death process, with all other parameters set equal. The total-evidence analyses provide a completely sampled phylogeny of Parnassiinae due to the inclusion of *Bhutanitis ludlowi* (only available with morphological data).

Each MCMC analysis was run for 100 million generations, sampled every 10,000th, resulting in 10,000 samples in the posterior distribution of which the first 2,500 samples were discarded as burn-in. All analyses were performed on the computer cluster CIPRES Science Gateway (Miller et al. 2015), using BEAGLE (Ayres et al. 2012). Convergence and performance of MCMC runs were evaluated using Tracer 1.6 (Rambaut and Drummond 2009) and the effective sample size (ESS) criterion for each parameter. A maximum-clade credibility (MCC) tree was reconstructed, with median age and 95% height posterior density (HPD). Bayes factor (BF) comparisons using stepping-stone sampling (Xie et al. 2011),

291 which allows unbiased approximation of the marginal likelihood of Bayesian analyses (Baele
292 et al. 2013), was used to select among competing models. We considered 2lnBF values >10 to
293 significantly favor one model over another (Kass and Raftery 1995). The dataset and BEAST
294 xml files generated for this study are available on Dryad (**Appendix 4**).
295
296 ***Historical biogeography***
297 Definition of biogeographic units was based on the present-day distribution of species
298 but also informed by plate tectonics (Sanmartín et al. 2001) and alpine orogeny
299 reconstructions (Bouilhol et al. 2013). Ten areas were defined: WN=Western North America
300 (including the Rocky Mountains); WP=Europe (France, Spain, Pyrenees, Alps, Italy, Greece,
301 Crete, Balkans, Scandinavia, Western Russian, and Ural Mountains); SI=Eastern Russia,
302 Siberia, and Kamchatka; CA=Central Asia (Turkmenistan, Uzbekistan and Kazakhstan);
303 MO=Mongolian steppes and Altai Mountains; TU=Turkey, Caucasian region, Syria, Iraq,
304 Iranian Plateau and Zagros Mountains; HTP=Himalaya, Tibetan Plateau, and Pamir region;
305 IN=India, Bhutan, Yunnan, and Southern China; CJ=Northern China, Korea, and Japan;
306 AF=North Africa and Arabia.
307 Species ranges were defined by presence–absence in each region (**Appendix 5**).
308 Biogeographic analyses were performed on the MCC tree (outgroups removed), using the
309 Dispersal–Extinction–Cladogenesis model of range evolution (DEC; Ree and Smith 2008)
310 implemented in the R-package *BioGeoBEARS* 0.2.1 (Matzke 2014). Because of concerns with
311 statistical validity and model choice in BioGeoBEARS (Ree and Sanmartín in review), we did
312 not use the DEC+J model (Matzke 2014). We also preferred DEC to DEC+J because the
313 latter often infers null or extremely low extinction rates (Sanmartín and Meseguer 2016), an
314 effect of the model favoring direct dispersal over widespread ranges, which might not be
315 adequate for reconstructing the history of ancient groups. Biogeographic ranges larger than

four areas in size were disallowed as valid biogeographic states if they were not subsets of the terminal species ranges; widespread ranges comprising areas that have never been geographically connected (Sanmartín et al. 2001) were also removed. We constructed a time-stratified model (**Appendix 6** on Dryad) with four time slices that specified constraints on area connections and spanned the Parnassiinae evolutionary history; each interval represents a geological period bounded by major changes in tectonic and climatic conditions thought to have affected the distribution of these butterflies.

Testing the Red Queen and Court Jester hypotheses

To understand the relative contributions of RQ versus CJ -type factors, we ran a series of macroevolutionary models using both maximum likelihood (ML) and Bayesian inference (BI). Rather than on a single tree, analyses were performed on a sample of trees from the BEAST post-burnin posterior distribution (with outgroups removed); an exception was the diversity-dependent diversification models that, due to time constraints, were run only on the MCC tree. This approach allowed us to assess the robustness of results to phylogenetic uncertainty and estimate confidence intervals for the parameter estimates. We compared the fit of each model to the data using the corrected Akaike Information Criterion (AICc) for the ML analyses, and Bayes Factor comparisons (2lnBF) for BI.

Diversification analyses in a Red Queen model. – We first tested the hypothesis that diversity is bounded or at equilibrium, meaning that diversity expanded rapidly in early diversification, filled most niches and saturated toward the present. We explored the effect of diversity-dependence on speciation and extinction rates using the method of Etienne et al. (2012) implemented in the R-package *DDD* 2.7. We applied five different models: (*i*) speciation depends linearly on diversity without extinction, (*ii*) speciation depends linearly on

341 diversity with extinction, (*iii*) speciation depends exponentially on diversity with extinction,
342 (*iv*) speciation does not depend on diversity and extinction depends linearly on diversity, and
343 (*v*) speciation does not depend on diversity and extinction depends exponentially on diversity.
344 For each model, the initial carrying capacity was set to the current number of described
345 species.

346 We also tested the effect of host-plant association on diversification rates, since
347 previous study has pointed to a potential correlation between shifts in diversification and host-
348 plant switches (Ehrlich and Raven 1964). We used a diversification model of the state-
349 dependent speciation and extinction (SSE) family of models, in which extinction and
350 speciation rates are associated with phenotypic evolution of a trait along a phylogeny
351 (Maddison et al. 2007). In particular, we used the Multiple State Speciation Extinction model
352 (MuSSE; FitzJohn et al. 2009) implemented in the R-package *diversitree* 0.9-7 (FitzJohn
353 2012), which allows for multiple character states. Larval host plant data were taken from
354 previous work (Ehrlich and Raven 1964; Scriber 1984; Collins and Morris 1985; Tyler et al.
355 1994; Scriber et al. 1995). The following four character states were used: (1)
356 Aristolochiaceae; (2) Zygophyllaceae; (3) Papaveraceae; and (4) Crassulaceae+Saxifragaceae.
357 Data at a lower taxonomic level than plant family were not used because of the great number
358 of multiple associations exhibited by genera that could alter the phylogenetic signal. Thirty-
359 six different models were run to test whether speciation, extinction, or transition rates were
360 dependent on trait evolution. Models were built with increasing complexity, starting from a
361 model with no difference in speciation, extinction and transition rates (3 parameters) to the
362 most complex model with different speciation, extinction, and transition rates for each
363 character state (20 parameters). We estimated posterior density distribution with Bayesian
364 MCMC analyses (10,000 steps) performed with the best-fitting models and resulting
365 speciation, extinction and transition rates.

366 There have been concerns about the power of SSE models to infer diversification
367 dynamics from a distribution of species traits (Davis et al. 2013; Maddison and FitzJohn
368 2015; Rabosky and Goldberg 2015). First, SSE models tend to have a high type I error bias
369 related to the shape (topology) of the inferred tree (Rabosky and Goldberg 2015). To test
370 whether our diversification results were potentially biased, we estimated the difference of fit
371 (ΔAIC) between the best model from MuSSE and a null model (corresponding to the model
372 with no difference in speciation, extinction and transition rates between the states of a
373 character) and compared this with the difference between the same models as estimated from
374 simulated datasets. 1000 sets of traits were generated on the Parnassiinae phylogeny using the
375 *sim.history* function from the R-package *phytools* (Revell 2012). To keep the phylogenetic
376 signal in the trait, transition rates between states were simulated using values obtained from
377 the null model fitted on the observed phylogeny.

378 Second, there is concern whether SSE models are uncovering actual drivers of
379 diversification, or whether they are simply pointing to more complex patterns involving
380 unmeasured and co-distributed factors in the phylogeny (Beaulieu and O'Meara 2016). A
381 model with an additional hidden character may alleviate this issue. We applied the Hidden
382 SSE (HiSSE) model to specifically account for the presence of unmeasured factors that could
383 impact diversification rates estimated for the states of any observed trait. The analyses were
384 performed in the Bayesian software RevBayes (Höhna et al. 2016a).

385 To provide an independent assessment of the relationship between diversification rates
386 and host specificity, we used models that allow diversification rates to vary among clades.
387 BAMM 2.5 (Rabosky et al. 2013, 2014a) was used to explore for differential diversification
388 dynamic regimes among clades differing in their host-plant feeding. BAMM analyses were
389 run for 10 million generations, sampling every 10,000th and with four different values of the
390 compound Poisson prior (CPP) to ensure the posterior is independent of the prior (Moore et

391 al. 2016). Mixing and convergence among runs (ESS>200 after 15% burn-in) were assessed
392 with the R-package *BAMMtools* 2.1 (Rabosky et al. 2014b). BAMM has been criticized for
393 incorrectly modeling rate-shifts on extinct lineages, i.e. unobserved (extinct or unsampled)
394 lineages inherit the ancestral diversification process and cannot experience subsequent
395 diversification-rate shifts (Moore et al. 2016, but see Rabosky et al. 2017). To solve this, we
396 used here a novel approach implemented in RevBayes that models rate shifts consistently on
397 extinct lineages by using the SSE framework (Moore et al. 2016; Höhna et al. in prep.).
398 Although there is no information of rate shifts for unobserved/extinct lineages in a phylogeny
399 including extant species only, these types of events must be accounted for in computing the
400 likelihood. The number of rate categories is fixed in the analysis but RevBayes allows any
401 number to be specified, thus allowing direct comparison of different macroevolutionary
402 regimes.
403
404 *Diversification analyses in a Court Jester model.* – We evaluated the impact of abiotic
405 factors such as abrupt changes in tectonic or climatic settings or mass extinction events (e.g.
406 the EOCT at 33.9 Ma) using episodic birth-death models implemented in the R-packages
407 *TreePar* 3.3 (Stadler 2011) and *TESS* 2.1 (CoMET model, Höhna et al. 2016b; May et al.
408 2016). These models allow detection of discrete changes in speciation and extinction rates
409 concurrently affecting all lineages in a tree. Both approaches estimate changes in
410 diversification rates at discrete points in time, but can also infer mass extinction events,
411 modeled as sampling events in the past in which the extant diversity is reduced by a fraction.
412 Speciation and extinction rates can change at those points but remain constant within time
413 intervals. The underlying likelihood function is identical in the two methods (Stadler 2011;
414 Höhna 2015), but TreePar and TESS differ in the inference framework (ML vs. BI) and the
415 method used for model comparison (AICc vs. BF). In addition, TESS uses independent CPPs

to simultaneously detect mass extinction events and discrete changes in speciation and extinction rates, while TreePar estimates the magnitude and timing of speciation and extinction changes independently to the occurrence of mass extinctions (i.e. the three parameters cannot be estimated simultaneously due to parameter identifiability issues, Stadler, 2011). To compare inferences between CoMET and TreePar, we performed two independent TreePar analyses allowing and disallowing mass extinction events (argument ME=TRUE/FALSE). We compared models with 0, 1 or 2 rate shifts/mass extinction events in TreePar using the AICc, while BF comparisons were used to assess model fit between models with varying number and time of changes in speciation/extinction rates and mass extinctions in TESS. Finally, we used an implementation of CoMET in RevBayes (Höhna, unpublished), an episodic birth-death model in which speciation and extinction are allowed to vary at discrete points in time, but are autocorrelated (instead of independent/uncorrelated as in CoMET) across time intervals; a Brownian model with rates in the next time interval centered around the rates in the current time interval was used. Also, the number of diversification rate shifts was set *a priori* as in population skyline models (Strimmer and Pybus 2001), while in CoMET this is modeled through a CPP prior.

To examine the influence of elevational distribution (lowland, mountain, or both) on speciation and extinction rates, we used the trait-dependent diversification model GeoSSE (Geographic State Speciation and Extinction, Goldberg et al. 2011). Geographic characters require a third widespread state, because, unlike morphological traits, ancestors can be present in more than one state (area). This requires the modeling of cladogenetic state change, the probability associated with the division of an ancestral range between the two descendants, alongside anagenetic state change (i.e. the probability of character change along branches). Species were coded by their elevation zone, using data from literature, museum records, and field observations. We evaluated 12 models of increasing complexity to test the relationship

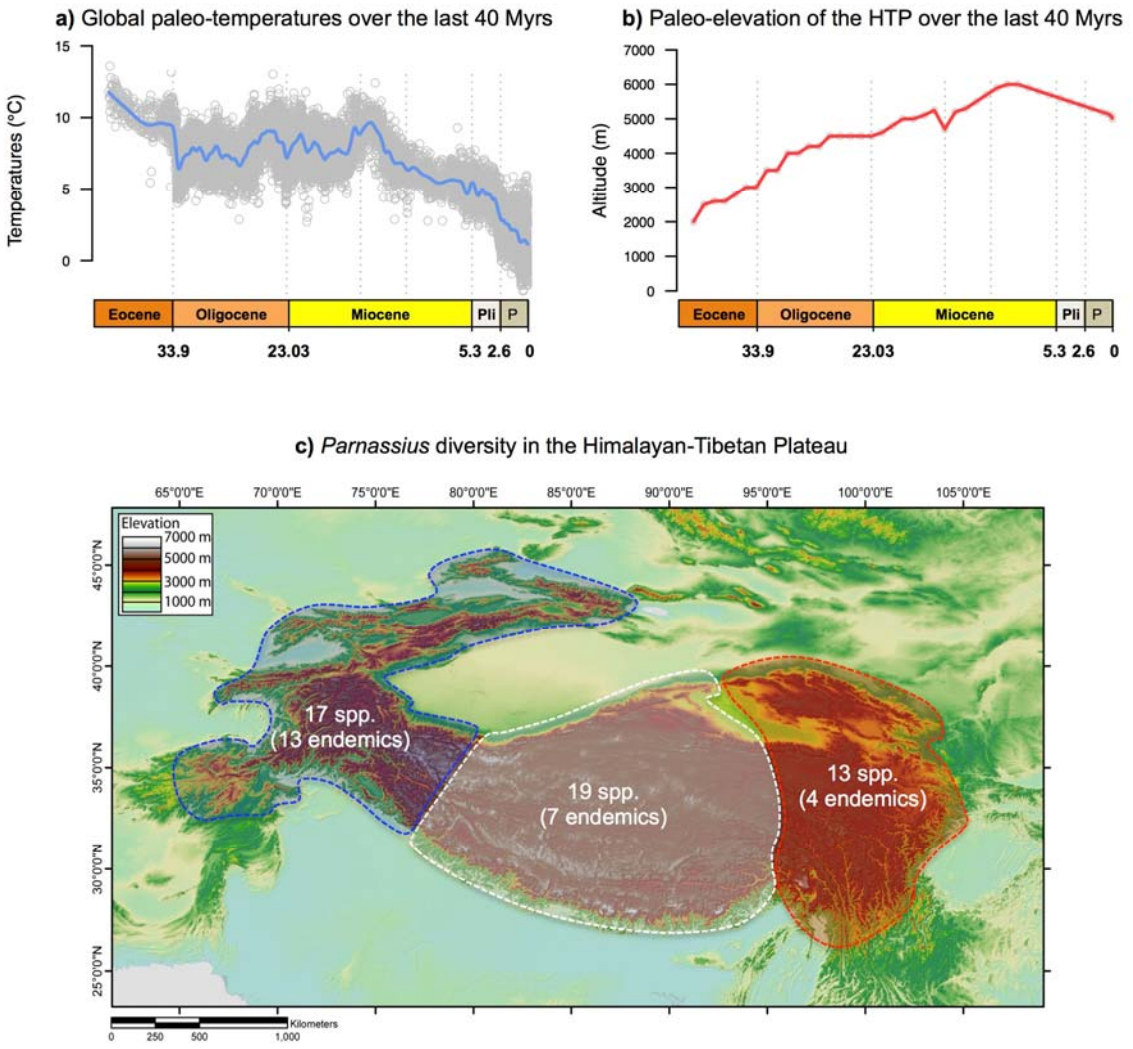
441 between elevational distributions and diversification. As in MuSSE above, we evaluated the
442 performance of GeoSSE using simulation tests. Trait simulation was more complex here
443 because there is no direct transition between states A and B. Instead, these transitions involve
444 range expansion to an intermediate widespread state AB (dAB, dBA) followed by local
445 extinction (xA, xB). Following Goldberg et al. (2011), we considered transition rates between
446 A to B as null because they involve more than one instantaneous event; the transition rate
447 from A to AB was coded as range expansion dA, from B to AB as dB; transitions from AB to
448 A and from AB to B ("extirpation rates") were coded as xB and xA, respectively. For our
449 1000 simulations, we used the transition rates of the model with equal speciation rates
450 (sA=sB=sAB), but different extinction (xA≠xB) and transition rates (dA≠dB).

451 Finally, we tested the impact of paleoenvironmental variables on diversification rates,
452 using a birth-death likelihood method in which rates may change continuously with an
453 environmental variable, itself varying through time (Condamine et al. 2013). We tested four
454 models: *(i)* a constant-rate birth-death model in which diversification is not associated to the
455 environmental variable; *(ii)* speciation rate varies according to environment and extinction
456 rate does not vary; *(iii)* speciation rate does not vary and extinction rate varies according to
457 environment; and *(iv)* both speciation and extinction rates vary according to environment. We
458 also tested the corresponding models in which speciation and/or extinction are allowed to
459 vary with time, but independently from the environmental variable (time-dependent birth-
460 death models, Morlon et al. 2011). When rates varied with the environment (E), we assumed
461 exponential variation, such that $\lambda(E) = \lambda_0 \times e^{\alpha E}$ and $\mu(E) = \mu_0 \times e^{\beta E}$, in which λ_0 and μ_0
462 are the speciation and extinction rates for a given environmental variable for which the value
463 is 0, and α and β are the rates of change according to the environment. Positive values for α or
464 β mean a positive effect of the environment on speciation or extinction (and conversely). As
465 environmental variables (**Figure 2**), we used paleotemperature (data retrieved from Zachos et

al. 2008) and paleo-elevation of the HTP (a proxy for mountain building), with data compiled from existing data in the literature available on Dryad (**Appendix 7**). The R-package *pspline* was used to build environmental vectors from the data as input for the birth-death models.

Alternatively, we used RevBayes and the CoMET framework (episodic birth-death models) to implement an analysis in which diversification rates are allowed to co-vary with, or evolve independently from, paleotemperatures and paleo-elevation across discrete time intervals (Palazzesi al. in review). A correlation coefficient, denoted β, is co-estimated with diversification rates in the model and describes the magnitude and direction of the relationship between the variable and changes in diversification rates: $\beta<0$ (negative correlation), $\beta>0$ (positive correlation), and $\beta=0$ (no correlation, in which case the model collapses to the episodic birth-death models; see **Appendix 8** on Dryad).

Figure 2. Environmental changes over the last 40 million years. a) Trends in global climate change estimated from relative proportions of different oxygen isotopes (Zachos et al. 2008), and b) trends in Himalayan-Tibetan Plateau (HTP) elevation for the last 40 Myrs, compiled from data from a literature survey; our compilation shows that the HTP was as high as 3000 m in the Eocene, congruent with recent reviews (Renner 2016; Appendix 7). c) Map of the current elevation of the Himalayan-Tibetan Plateau with the species diversity of *Parnassius* in the region.



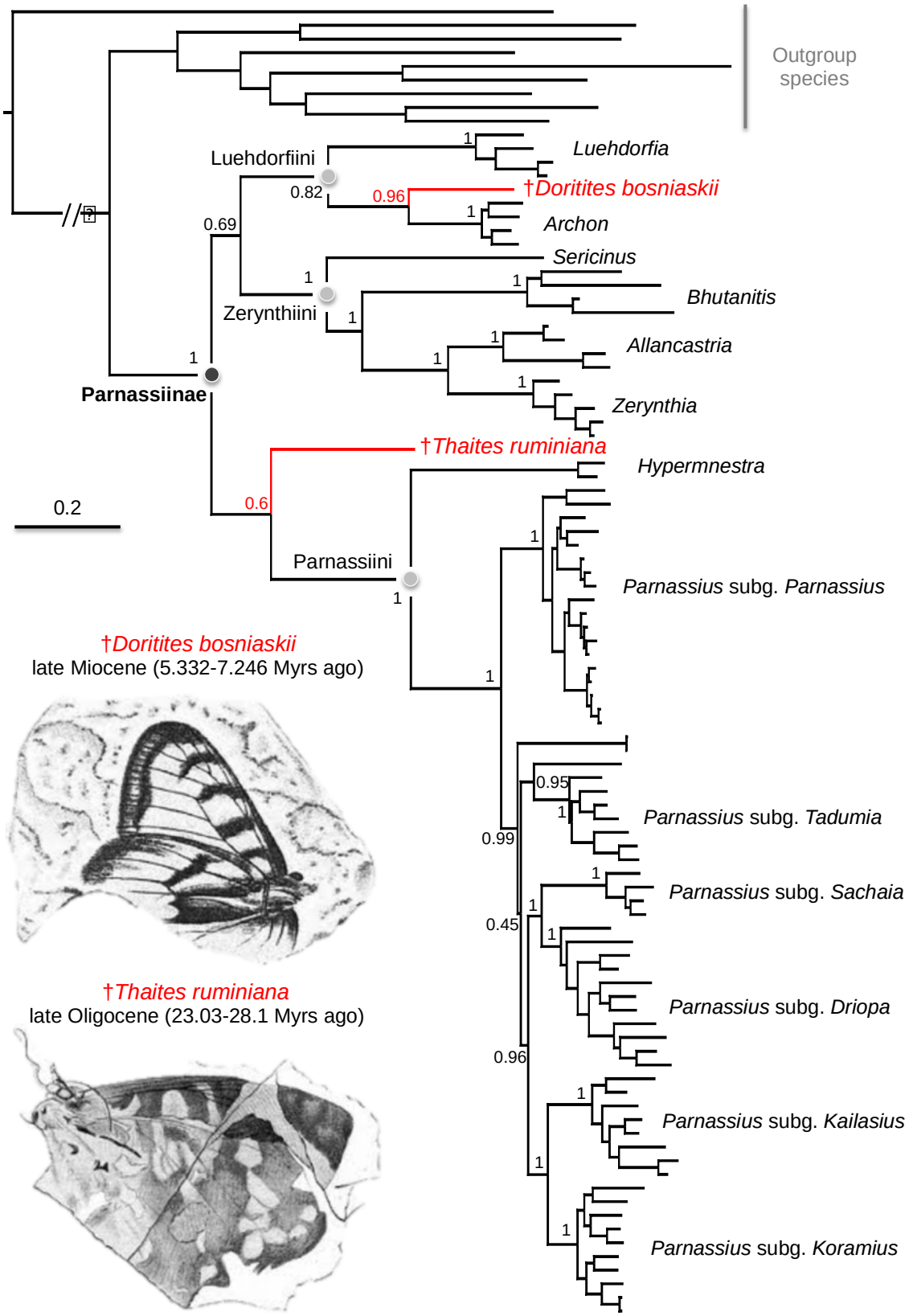
Results

Phylogeny and fossil placements

The molecular matrix comprised 4535 nucleotides and 85 extant species of Parnassiinae plus two fossil species, following recent taxonomic reviews (Frankenbach et al. 2012; Bollino and Racheli 2012) (**Appendices 1 and 2** on Dryad). PartitionFinder selected eight partitions as the best scheme for substitution models (**Appendix 9** on Dryad). Convergence among runs was supported by the low average standard deviation of split frequencies, PSRF values ≈1.0, and ESS>>1000 for many parameters. The resulting Bayesian trees were well resolved and robust: 74.4% of the nodes were recovered with PP>0.95 with MrBayes (**Figure 3**), and 83% in

495 BEAST (70% with PP=1) when fossil taxa were removed. Phylogenetic relationships agree
496 with previous studies (Nazari et al. 2007; Michel et al. 2008), showing
497 Luehdorfiini+Zerynthiini as sister-tribes and sister to Parnassiini; within the latter,
498 *Hypermnestra* is sister to the genus *Parnassius*. Most of the weak support is found within the
499 radiation of species complexes, such as *Parnassius apollo* or in the subgenus *Kailasius*, but
500 two basal nodes in *Parnassius* have PP<0.9.

501 The total-evidence analysis allowed us to test the phylogenetic placement of the
502 Parnassiinae fossils. †*Thaites* was often recovered as sister to Parnassiini (PP=0.6, **Figure 3**),
503 and occasionally as sister to Luehdorfiini+Zerynthiini. This fossil was used to provide a
504 minimum age for crown Parnassiinae, calibrated with a uniform prior bounded between 23.03
505 Ma (minimum) and 47.8 Ma (maximum). †*Doritites* was reconstructed as sister to *Archon*
506 (Luehdorfiini, PP=0.96), in agreement with Carpenter (1992), who tentatively synonymized
507 †*Doritites* with *Luehdorfia*. The crown of Luehdorfiini was thus constrained for divergence
508 time estimation using a uniform distribution bounded between 5.33 Ma (minimum) and 47.8
509 Ma (maximum).
510
511 **Figure 3. Phylogenetic relationships inferred with a Bayesian total-evidence approach**
512 **combining molecular and morphological data of extant and extinct Parnassiinae.** This
513 integrative approach allowed assessment of the phylogenetic placement of two fossils, which
514 were then used to calibrate the molecular dating analysis. Backbone nodes show posterior
515 probability values. Nodal support values within genera or subclades are shown in Appendix
516 10 for alternative dating analyses including, or not, morphological data.



517

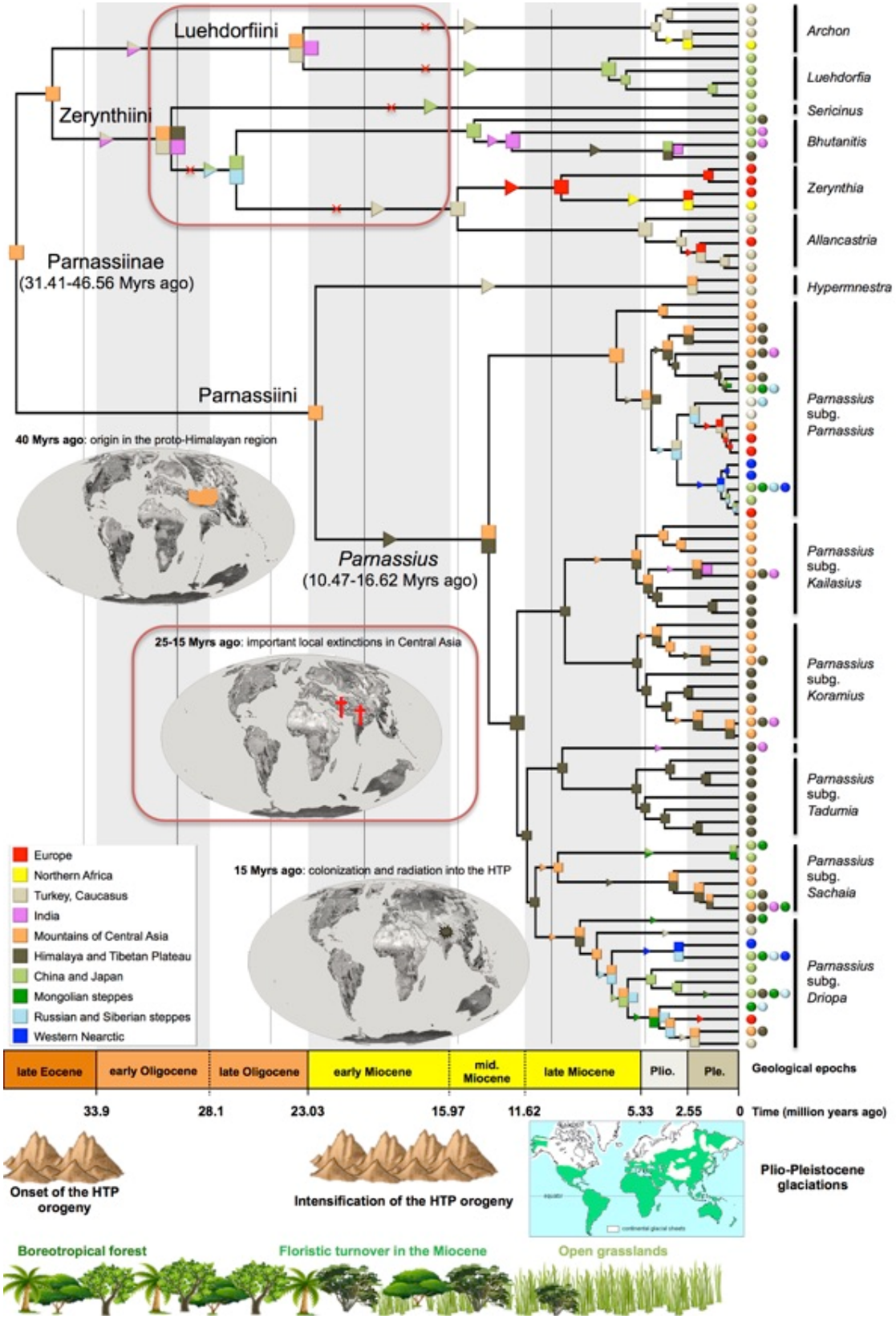
518

519 ***Divergence time and biogeographic estimates***

520 The BEAST analyses (Yule vs. birth-death, molecular vs. total-evidence) yielded
521 almost identical estimates of divergence times with less than one million years of difference
522 for most nodes (**Appendix 10** on Dryad). Bayes Factor comparisons were also not conclusive,
523 with little difference in the stepping-stone marginal likelihood estimate among analyses
524 (**Appendix 11** on Dryad). We present the results for the analysis with the birth-death prior
525 and the total-evidence dataset (**Figure 4**) because it offered slightly better convergence and
526 incorporating extinction into the tree prior seemed more realistic for such an old lineage
527 (results from the other analyses are available on Dryad, **Appendix 10**).

528 Based on this analysis and results from the DEC inference, Parnassiinae originated in
529 the late Eocene (38.6 Ma, 95% HPD 31.4-46.6 Ma) in Central Asia (**Figure 4**). The ancestor
530 of sister-tribes Luehdorfiini and Zerynthiini (36.7 Ma, 29.9-44.5 Ma) is also reconstructed in
531 Central Asia, followed by independent dispersal events to India + Caucasus-Turkey and the
532 HTP. Luehdorfiini diversified in the early Miocene at 23.3 Ma (17.4-30.2 Ma) within a broad
533 geographic range including India, Turkey-Caucasus, and Central Asia; this was followed by
534 extinction in Central Asia and India, and a later dispersal to China-Japan along the stem of
535 *Luehdorfia*. Zerynthiini originated in the Oligocene (30.3 Ma, 24.1-37.2 Ma) within a region
536 comprising India, Turkey-Caucasus, Central Asia, and the HTP, and also underwent
537 extinction events in Central Asia and India during the Miocene. In contrast, Parnassiini first
538 diversified within this region in the early Miocene (22.6 Ma, 17.2-28.5 Ma), followed by
539 dispersal to Turkey-Caucasus in *Hypermnestra*. The ancestor of *Parnassius* dispersed from
540 Central Asia to HTP around the mid-Miocene (13.4 Ma, 10.5-16.6 Ma), followed by
541 vicariance between subgenus *Parnassius* and the ancestor of the other subgenera. Dispersal
542 events to adjacent geographic regions and subsequent allopatric speciation are reconstructed

543 within each subgenus, especially in *Driopa*. A more detailed account of the biogeographic
544 history and alternative nodal reconstructions are available on Dryad (**Appendix 12**).
545
546 **Figure 4. Time-calibrated phylogeny and biogeography of Apollo butterfly radiation**
547 **and paleo-tectonic evolution at important periods.** Colored squares on nodes indicate the
548 most likely biogeographical estimate, as in the lower left inset (uncertainties in range
549 estimates are in Appendix 12). Triangles indicate range expansions, and X's denote local
550 extinctions. A timescale spans the full evolutionary history of the group. Earth maps represent
551 paleogeography at specific time periods. The red box highlights a notable period of local
552 extinctions in the clade Luehdorfiini+Zerynthiini. Panels at bottom represent the main
553 Cenozoic environmental changes occurring with the onset and rise of the Himalayan-Tibetan
554 Plateau (HTP) or floristic turnovers. Sensitivity analyses were conducted to test the effect of
555 adding or excluding morphological data; results are presented in Appendix 10. Plio., Pliocene;
556 Ple., Pleistocene.
557



558

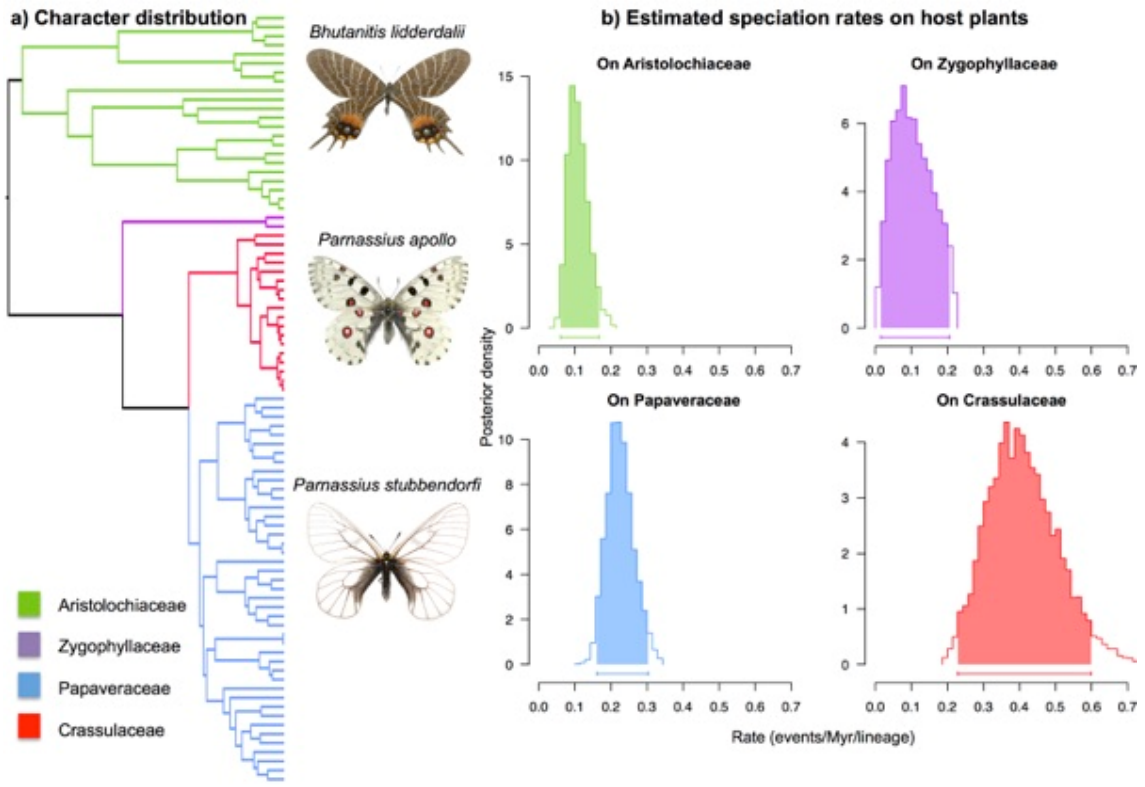
559 ***Macroevolutionary dynamics***

560 A summary of diversification rate analysis results is presented in **Table 1**.

561 *Diversification in a Red Queen model.* – Diversity-dependent diversification models
562 estimated an unlimited carrying capacity for the subfamily (**Appendix 13** on Dryad),
563 implying that accumulating diversity did not influence diversification rates. The best-fitting
564 DDD model was one in which extinction rate increased with diversity. For *Parnassius* the
565 best model was a diversity-dependence process with speciation decreasing with increasing
566 diversity. However, none of these models received significant support when compared to
567 other models (ΔAICc<2).

568 MuSSE analyses on host-plant preferences selected a model in which speciation rates
569 varied among clades feeding on different host plants; extinction and transition rates were
570 estimated as equal (**Appendix 14** on Dryad). Zygophyllaceae feeders (*Hypermnestra*) had the
571 lowest speciation rate, followed by Aristolochiaceae feeders (Luehdorfiini and Zerynthiini),
572 with the highest rates for clades feeding on Papaveraceae (all *Parnassius* except subgenus
573 *Parnassius*) and Crassulaceae+Saxifragaceae (subgenus *Parnassius*) (**Figure 5**). MCMC
574 credibility intervals showed significant differences between the speciation rates of
575 Papaveraceae feeders and Crassulaceae+Saxifragaceae feeders and those feeding on other
576 host plants (http://dx.doi.org/10.5061/dryad.[NNNN], **Appendix 15**). Transition rates for
577 host-plant switches were often estimated close to zero with strong niche conservatism in the
578 phylogeny; once a lineage shifted to a new host it did not switch back and only rarely
579 colonized another host plant (**Appendix 14**). We assessed the robustness of this pattern with a
580 simulation procedure, and Bayesian implementations of HiSSE and MuSSE in RevBayes
581 (http://dx.doi.org/10.5061/dryad.[NNNN], **Appendix 16**). All analyses supported the same
582 diversification pattern with speciation rates being different among lineages feeding on

583 different host plants. Our estimates were robust to known biases of SSE models, suggesting a
584 strong effect of host plant on diversification rates of Parnassiinae.
585
586 **Figure 5. Diversification of Parnassiinae linked to their host plant.** a) Phylogenetic
587 distribution of host-plant traits (four states) analyzed with MuSSE diversification models. b)
588 Bayesian inferences made with the best-fitting MuSSE model (see Appendix 14) showed that
589 speciation rates vary according to the host-plant trait: *Parnassius* clades feeding on
590 Papaveraceae and Crassulaceae+Saxifragaceae have higher speciation rates than their
591 relatives feeding on other families; extinction and transition rates were estimated as equal
592 across clades.



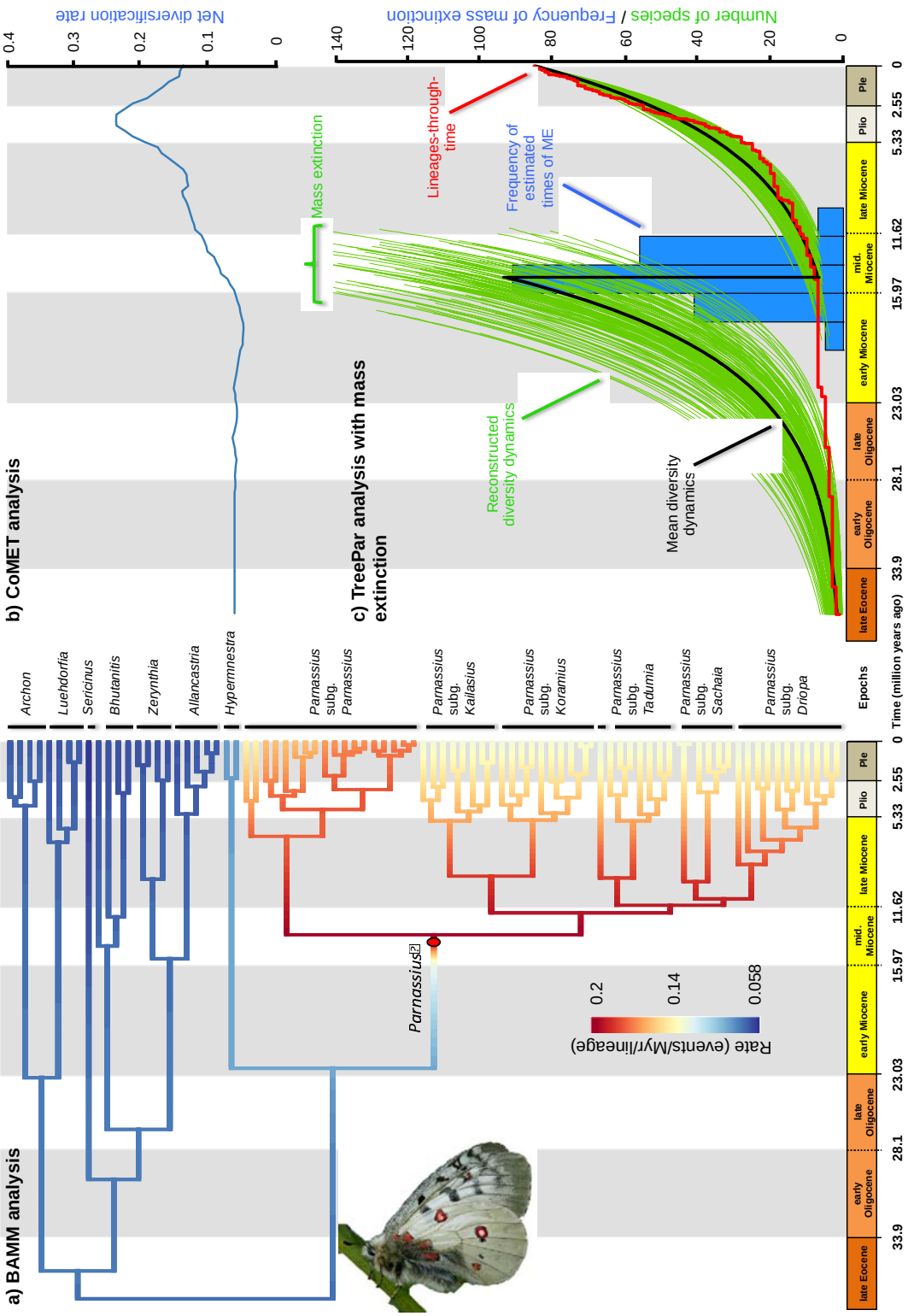
593
594 Bayesian inferences that model rate-heterogeneity across clades provided further
595 corroboration of the SSE inferences (**Figure 6a**). Both BAMM and an alternative
596 implementation of lineage-specific birth-death model in RevBayes supported a diversification

597 regime in which there is at least an increase in diversification rates along the stem of
598 *Parnassius* (and possibly a second upshift at the stem of subgenus *Parnassius*) and found
599 higher speciation rates for host-plant feeders on Crassulaceae+Saxifragaceae and
600 Papaveraceae than those on Aristolochiaceae or Zygophyllaceae (**Appendices 17-19** on
601 Dryad). BAMM was insensitive to the CPP prior.
602
603 **Figure 6. Diversification dynamics of Parnassiinae inferred with time-dependent models.**
604 a) Rate-heterogeneity-across-clades models implemented in BAMM and RevBayes identified
605 one significant shift in diversification rates along the stem leading to *Parnassius* (represented
606 by a red circle; see also Appendices 17-19), with a higher speciation rate in this clade than in
607 the rest of the tree. Episodic birth-death models implemented in CoMET (b) and TreePar (c)
608 identified a possible mass extinction event in the mid-Miocene (~15 Ma). The blue histogram
609 shows the uncertainty in the timing of the mass extinction event based on the estimates over
610 200 trees with TreePar (similar estimates are obtained for CoMET; Appendix 20). The two
611 models also estimated an increase in diversity around the early-mid Pliocene (b,c), but this
612 was followed by a decrease in diversity towards the present time in CoMET which was not
613 detected in TreePar. The red line shows the lineages-through-time plot. The green lines
614 represent the reconstructed past diversity before and after the mass extinction, based on the
615 net diversification rates and 7.5% of survival probability estimated in TreePar. Plio., Pliocene;
616 Ple., Pleistocene.
617

618

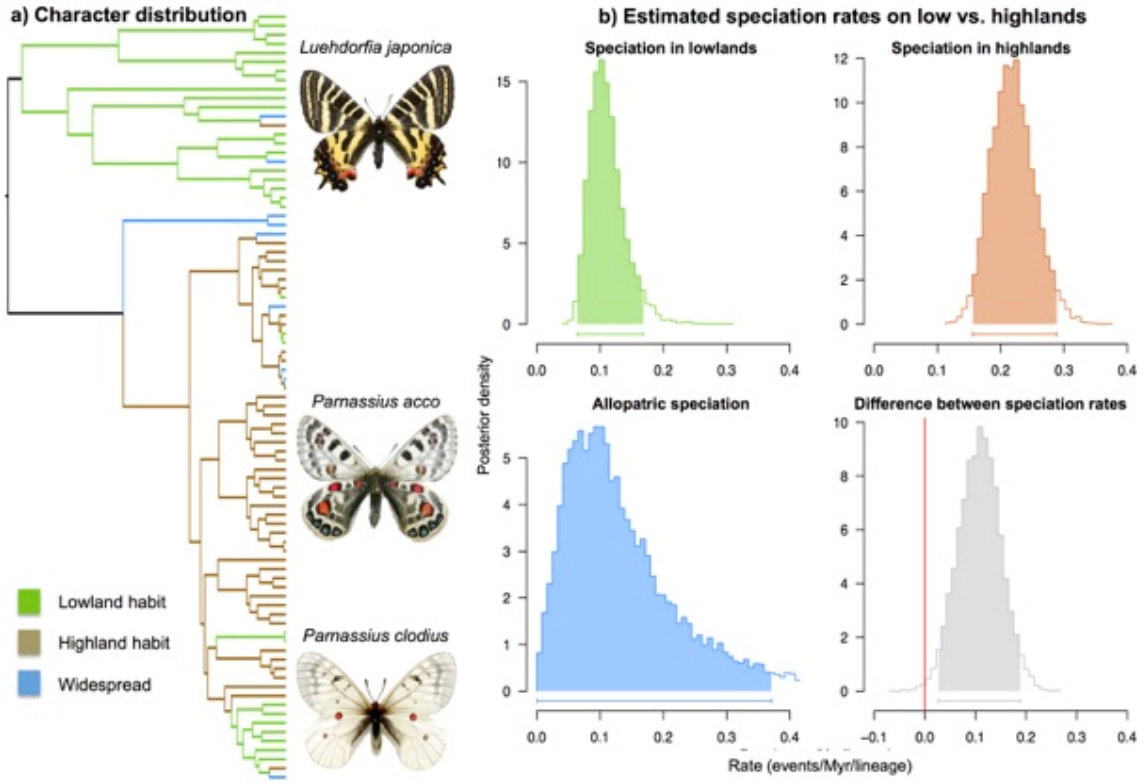
619

620



621 *Diversification in a Court Jester model.* – When mass extinctions are disallowed,
622 TreePar supported a model with a single shift at 3.2 Ma (**Table 2**), with a low initial rate of
623 diversification (r_2=0.031) and high turnover (ε_2=0.91), followed by 4-fold increase in
624 diversification rate (r_1=0.119) and a joint decrease of turnover (ε_1=0.17). The second best
625 model identified a second shift in the mid-Miocene (15.3 Ma) marked by an extinction period
626 surrounded by low and negative values for the net diversification rate, and values >1 for
627 turnover. When mass extinctions are allowed to occur, TreePar favored a significant mass
628 extinction at 14.8 Ma with an estimated survival probability after the mass extinction of 7.5%
629 (**Table 2**; **Figure 6c**). The CoMET model implemented in RevBayes with the autocorrelated
630 model also identified a mass extinction event in the mid-Miocene ~15 Ma (2lnBF=6,
631 **Appendix 20** on Dryad). There was also a significant increase in speciation rates around the
632 Pliocene, followed by a decrease in speciation towards the present (**Figure 6b**); similar results
633 were obtained with the TESS implementation of CoMET (**Appendix 20**).
634 GeoSSE analyses selected a best-fitting model with equal dispersal and extinction
635 rates between altitudinal regions (habitats), but higher speciation rates for highland regions
636 compared to lowland regions (**Appendix 14** on Dryad). Allopatric speciation (for
637 lowland/highland ancestors) was inferred to occur at a higher rate than within-lowland
638 speciation, but at a lower rate than within-highland speciation; these differences remained
639 significant in the Bayesian MCMC analyses, as the credibility interval did not overlap with
640 zero (**Figure 7**). Nevertheless, the difference in speciation rates between higher and lower
641 altitudes was not larger than expected by chance in our simulation tests (**Appendix 16** on
642 Dryad), so we cannot reject the observed effect of altitude on speciation being affected by the
643 shape of the phylogeny.
644

Figure 7. Inference of geographic mode of speciation between mountain and lowland species. a) The phylogenetic distribution of geographic traits (lowland, mountain or 'widespread') analyzed with GeoSSE models of diversification. b) Bayesian inferences made with the best-fitting GeoSSE model (selected over a series of 12 diversification models) showing that only speciation varies according to the geographic trait. The results indicate that the mountain-dwelling Parnassiinae species have significantly higher speciation rates than their counterparts. The allopatric (or biome divergence) speciation rate is also estimated to be elevated, congruent with our biogeographic estimates.



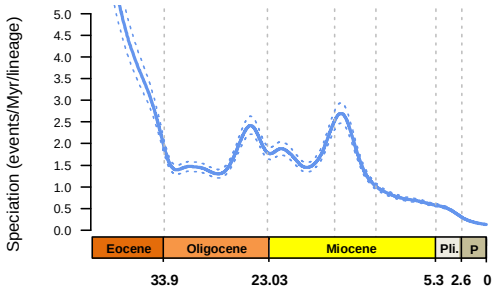
The ML-based paleoenvironment-dependent model implemented in RPANDA indicated positive dependence between diversification rates and past temperature (**Table 3**). In the best model, both speciation and extinction rates increased exponentially with temperature, with a general decrease in diversification rates toward the present, combined

658 with periods of high turnover rates in the late Oligocene and mid-Miocene (**Figure 8**). The
659 best paleo-elevation model had speciation increasing with increasing altitude, whereas
660 extinction remained constant (and not affected by elevation fluctuations), but this model does
661 not fit better than constant-rate or temperature-dependent models (**Table 3**). In contrast, the
662 Bayesian paleoenvironment-dependent model implemented in RevBayes indicated that both
663 past temperature and elevation did not impact the diversification of Parnassiinae, since the
664 correlation parameters α and β had credibility intervals that were not significantly different
665 from zero (**Appendix 21** on Dryad). Moreover, estimates with this model (**Figure 8**) showed
666 diversification dynamics that were similar to those obtained with CoMET and TreePar, but
667 unlike those from RPANDA. These differences might be attributed to the inference
668 framework (ML in RPANDA vs. BI in RevBayes) or to the different underlying birth-death
669 models (continuous in RPANDA vs. episodic in RevBayes). To test this, we implemented the
670 continuous-paleoenvironmental models (Condamine et al. 2013) in TESS. We then compared
671 these models against a constant birth-death model, episodic (piecewise constant) birth-death
672 models (1 to 5 shifts), and time-variable models. This procedure ensured that a single
673 statistical framework (ML and AIC) implemented in TESS was used to compare all models.
674 Results showed that the best-fitting model was one in which speciation and extinction rates
675 vary positively with paleotemperature, in agreement with RPANDA (**Table 4**). We therefore
676 concluded that the different results are related not to the inference framework but to the
677 underlying model: the use of an episodic birth-death model with autocorrelated rates across
678 time intervals in the RevBayes model rather than the uncorrelated episodic model
679 implemented in TESS.
680
681 **Figure 8. Paleoenvironment-dependent diversification processes in Apollo butterflies.**
682 ML continuous paleoenvironment-dependent models implemented in RPANDA (a-d) showed

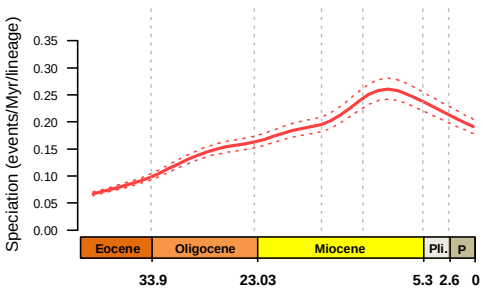
683 positive dependence between paleotemperatures and speciation/extinction; the paleo-
684 elevation-dependent diversification models did not fit the data better than the constant-rate or
685 time-variable models. Bayesian inference of environment-dependent diversification with
686 autocorrelated rates in RevBayes (e-h) supports a model in which both speciation and
687 extinction are not dependent on temperature and elevation changes (represented by dotted
688 lines, temperature and elevation values are given in Figure 2). Colored areas in speciation and
689 extinction plots indicate the credibility interval, with the continuous line representing the
690 median rate. Pli, Pliocene; and P, Pleistocene.

ML continuous paleoenvironment-dependent diversification

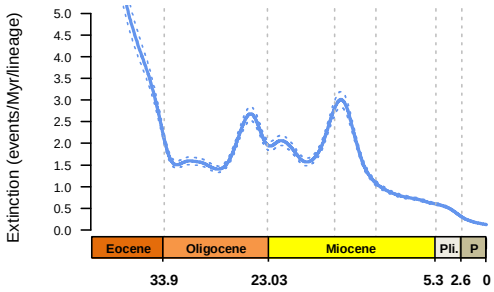
a) Speciation depends positively on past temperatures



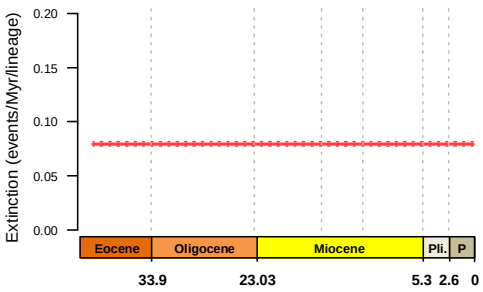
c) Speciation depends positively on past elevation



b) Extinction depends positively on past temperatures

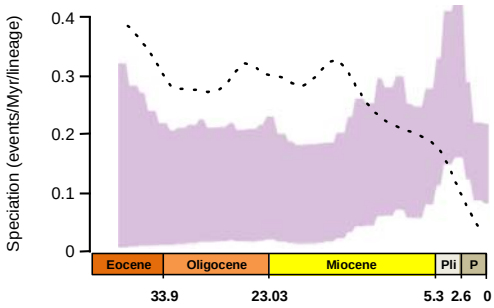


d) Extinction does not depend on past elevation

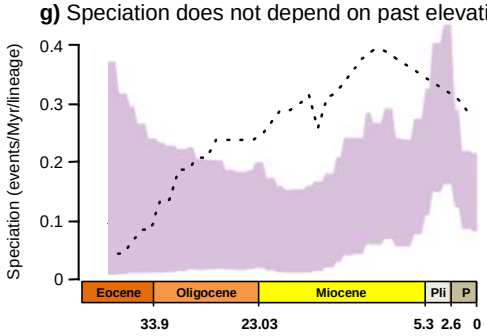


Bayesian autocorrelated paleoenvironment-dependent diversification

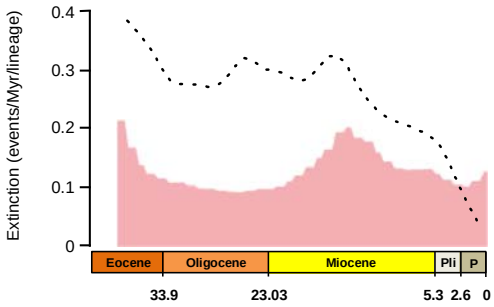
e) Speciation does not depend on past temperatures



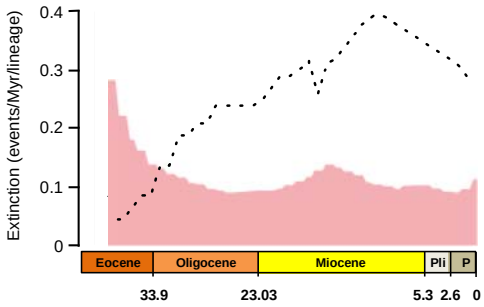
g) Speciation does not depend on past elevation



f) Extinction does not depend on past temperatures



h) Extinction does not depend on past elevation



691

692 ## Discussion

693 Seeking the determinants of species richness is a fundamental aspect of biology (Morlon
694 2014; Benton 2015). However, studies have generally focused on a single factor to explain a
695 given diversity pattern, whether by historical events promoting changes in diversification and
696 extinction (Barnosky 2001; Erwin 2009; Mayhew et al. 2008), or species ecology and clade
697 competitions as the dominant drivers of diversification (Rabosky 2013; Silvestro et al. 2015).
698 The combined influence of various factors has been tested less often (Drummond et al. 2012a;
699 Bouchenak-Khelladi et al. 2015; Lagomarsino et al. 2016), and at a lower geographic and
700 temporal scale or with a reduced set of methods. Further, these methods have mainly been
701 implemented within a ML framework. In this study, we used existing and some novel
702 methods under both ML and BI to evaluate potential factors driving diversification in
703 Parnassiinae: some are related to a RQ mechanistic model of evolution, others to a CJ
704 hypothesis.
705
706 ***Coupled effect of mountain building and climate change***
707 A long-held tenet in biology is that environmental stability leads to greater species richness
708 than a changing environment, suggesting that species interactions are a dominant force
709 (Wallace 1878). A recent debate has emerged between those arguing that physically dynamic
710 ecosystems play a major role in diversification (Hoorn et al. 2013; Condamine et al. 2015a)
711 and even promote adaptive radiation (Tan et al. 2013), and those maintaining that landscape
712 changes have not been important in diversification history, instead advocating the role of
713 large-scale dispersal events as drivers of speciation (Smith et al. 2014). In the Northern
714 Hemisphere, tectonic events such as the closure of Turgai Sea or the building of Alpine and
715 Himalayan orogenies have caused large-scale landscape and climatic changes (Sanmartín et
716 al. 2001). As a result many new habitats were formed that created potential ‘new ecological

717 spaces' and favored speciation (Xing and Ree 2017). Also, the newly created corridors
718 stimulated species exchange between communities, while also acting as barriers that separated
719 formerly continuous populations (Brikiatis 2014). Alternatively, landscape dynamics (i.e.
720 unlike long-term stable environments) may have shattered former habitats, leading organisms
721 to extinction. The rise of the HTP has often been invoked as providing important
722 opportunities to diversify (Fjeldså et al. 2012; Price et al. 2014; Wen et al. 2014; Favre et al.
723 2015; Hughes and Atchison 2015; Renner 2016), but its role as a driver has rarely been
724 formally tested with diversification models (Xing and Ree 2017).

725 Our analyses detected higher *in-situ* speciation rates for mountain species than their
726 lowland counterparts (with similar extinction rates), resulting in higher net diversification
727 rates for mountain species. Yet, this pattern was not robust to phylogenetic shape artifacts
728 (Rabosky and Goldberg 2015). Moreover, ML and BI-based paleo-elevation models found
729 that an explicit link between diversification rates and periods of HTP uplift could be rejected
730 in favor of a constant birth-death. These findings contrast with those of high-elevation taxa in
731 the Andes reported to have diversified faster due to the rise of the Andes as compared to their
732 low-elevation relatives (Lagomarsino et al. 2016; Pérez-Escobar et al. 2017). In contrast,
733 clade rate-heterogeneous models (BAMM, RevBayes) and biogeographic analyses supported
734 a link between diversification rates and the HTP orogeny within Parnassiini, particularly in
735 *Parnassius*. A significant upshift in speciation rates was detected along the stem of
736 *Parnassius*, coincident with the colonization of the HTP, while DEC inferred several
737 speciation/vicariance events associated with the HTP orogeny (**Figure 4**). The latter is
738 corroborated by the relatively high rate of allopatric (lowland/highland) speciation estimated
739 by GeoSSE. These different results can partially be explained by the HTP rise providing
740 novel, high-altitude habitats for colonization in *Parnassius*, and thereby promoting ecological
741 divergence and initial rapid diversification (followed later by allopatric speciation). Yet,

742 diversification was not strictly parallel to periods of mountain building. Conversely, the rise
743 of the HTP – and concomitant climatic changes such as regional cooling and aridification –
744 led to higher extinction rates in non-mountain-adapted Parnassiinae lineages, such as
745 Luehdorfiini and Zerynthiini that survived west and east of the Himalayan range; several
746 events of extinction are inferred within these lineages around the mid-Miocene. This mixed
747 effect is probably responsible for the lack of a significant association between orogeny and
748 speciation in tree-wide models, which is otherwise detected by clade-heterogeneous models.
749 Paleoclimate change is another abiotic factor postulated to be a major trigger of
750 diversification in terrestrial organisms (Erwin 2009). CoMET, TreePar, and the ML-based
751 paleotemperature-dependent model inferred a pattern of increased turnover (background
752 extinction) rates associated with periods of global climate warming, such as the mid-Miocene
753 Climate Optimum (MMCO, 15-17 Ma, Böhme 2003) or the late Oligocene Warming Event
754 (LOWE, ~25 Ma, Zachos et al. 2008, **Figure 2**). DEC also reconstructed several events of
755 (geographic) extinction in Central Asia and India along the stem branches subtending the
756 crown diversification of extant genera in Luehdorfiini and Zerynthiini (**Figure 4**). These long
757 branches are consistent with periods of high extinction rates related to the LOWE, which were
758 eventually followed by mid-Miocene radiations, especially in *Parnassius* (the latter probably
759 linked to the HTP rise). Paleotemperature and the HTP orogeny probably had a joint influence
760 in the diversification of Parnassiinae. Mountain building is known to be responsible for
761 regional climate change (Sepulchre et al. 2006; Armijo et al. 2015), and for changes in
762 biodiversity patterns, both directly as a geographic barrier and indirectly through climate
763 change (Hoorn et al. 2013). The HTP acted as a barrier to the influence of the Asian
764 monsoons (Zhisheng et al. 2001) and induced considerable drying of Central Asia (Quade et
765 al. 1989). In Parnassiinae, the HTP rise and subsequent periods of mountain building might
766 have promoted diversification through the provision of “climate refugia”: locations where

767 taxa survived periods of regionally adverse climate. Although usually associated with
768 maintaining biodiversity through glacial–interglacial climate changes (Gavin et al. 2014),
769 mountains can provide climate refugia during warming-aridification events (Migliore et al.
770 2013; Pokorny et al. 2015). Drastic range contractions within the non-mountain adapted
771 Luehdorfiini and Zerynthiini (**Figure 4**) could have been driven by aridification events
772 induced by HTP uplift and mid-Miocene climate warming, with *Hypermnestra* as a relict
773 group that survived in the Zagros Mountains. In *Parnassius*, colonization of the HTP
774 coincided with a boost of speciation, followed by dispersal to adjacent areas in the late
775 Miocene-Pliocene. The long stem branches leading to extant Parnassiinae genera are likely
776 the result of climate-driven extinction events associated with Cenozoic warming periods, such
777 as the LOWE and the MMCO. During these periods, Parnassiinae species probably migrated
778 toward cooler temperatures by colonizing the newly uplifted mountain ranges of the Zagros
779 and HTP. This agrees with the climate refugia hypothesis, a pattern currently observed in
780 Himalayan plants, which are shifting their distribution upward along with global warming
781 (Padma 2014). We argue that the HTP acted as climate refugia during global warming events,
782 in which pre-adapted parnassiines survived and diversified before re-colonizing ancestral
783 geographic ranges. It would be interesting to test this hypothesis in other extra-Himalayan
784 groups that diversified during the HTP orogeny (Favre et al. 2015), as well as in other
785 mountains around the world for which paleo-elevation data is available (Fjeldså et al. 2012;
786 Hoorn et al. 2013; Hughes and Atchison 2015).
787
788 ***Ecological interactions via insect-plant diversification and host-plant shifts***
789 Biological interactions among distantly related lineages probably played a major role in the
790 diversification of clades over geological time (Van Valen 1973; Liow et al. 2015; Silvestro et
791 al. 2015; Voje et al. 2015). Our analyses identified significant differences in speciation rates

associated with different host-plants. Several models (BAMM, MuSSE, and Bayesian implementations of MuSSE and HiSSE in RevBayes) indicated that *Parnassius* subgenera associated with Crassulaceae-Saxifragaceae and Papaveraceae have significantly higher rates of speciation than their relatives feeding in other host plants (**Figures 5, 6**). Interestingly, the lowest speciation rates are associated with butterflies feeding on Aristolochiaceae, the inferred ancestral host plant of the Parnassiinae (Condamine et al. 2012). Our results concur with the 'escape and radiate' hypothesis of Ehrlich and Raven (1964), which predicts that the evolution of a herbivore insect trait (i.e. new tolerance of plant defenses) that allows colonizing/feeding on a novel host plant lineage would lead to a burst of diversification. The highly unbalanced extant species richness between Parnassiini and the clade Luehdorfiini+Zerynthiini, explained above by geographic and climate drivers, might also be explained by differential speciation rates related to their feeding habit, as well as the overall low ability of Parnassiinae for shifting between host plants (i.e. transition rates estimated MuSSE are close to zero).

Testing this hypothesis would require reconstruction of the diversification history of the host-plant clades; this would allow us, for example, to identify shifts that led to reciprocal changes in diversification in the host plants and to test whether these shifts were accompanied by matching ancestral geographic distributions between herbivore insects and plants at key events (Cruaud et al. 2012). Unfortunately, so far there is no complete phylogeny for any of the large host families for Parnassiinae, including Aristolochiaceae (~500-600 spp.), Papaveraceae (~700-800 spp.), Crassulaceae (~1300-1500 spp.), and Saxifragaceae (~600-700 spp.). However, we found a positive correlation between species richness of these plant lineages and the speciation rates of their insect feeders (MuSSE: R^2=0.96, BAMM: R^2=0.87; **Appendix 22** on Dryad). This suggests that fast-diversifying clades are associated with families offering a higher number of potential hosts (akin to niche availability). Janz et al.

817 (2006) found a similar correlation between host-plant diversity and speciation rates in the
818 Nymphalidae; they posited that recurring oscillations between host-plant expansions (i.e.
819 incorporation of new plants into the feeding repertoire) and host specialization acted as a
820 major driving force behind the diversification of plant-feeding insects.

821 In adaptive radiations, the expectation is an initial rapid burst of diversification,
822 followed by a slowdown in speciation rates as niches are filled up; this is attributed to a
823 diversity-dependence effect (Phillimore and Price 2008; Etienne et al. 2012). A shift to a new
824 ecological resource (e.g. a new host plant) can be an adaptive breakthrough (Ehrlich and
825 Raven 1964), providing higher opportunities for speciation at the start of the radiation, which
826 significantly decrease as species accumulate. Though we did not detect this diversity-
827 dependence effect in Parnassiinae, a pattern of speciation decreasing with standing diversity
828 (and constant extinction rates) was found in *Parnassius*, supporting the hypothesis of this
829 genus as an example of an adaptive radiation (Rebourg et al. 2006), driven by the evolution of
830 a new host-plant association. The inference of similar extinction rates in clades of
831 Parnassiinae feeding on different host plants does agree with Van Valen's (1973) RQ
832 hypothesis, which assumes constant extinction probabilities shared by all members of any
833 given higher taxon.
834
835 ***Interplay between the Red Queen and the Court Jester***
836 Both abiotic and biotic factors drive species diversification, and biological radiations require
837 both extrinsic conditions and intrinsic traits – acting in unison or sequentially over time – for
838 success (Donoghue and Sanderson 2015). A lack of suitable methods and data has resulted in
839 few studies attempting to merge both types of factors to explain evolutionary radiations
840 (Bouchenak-Khelladi et al. 2015; Liow et al. 2015; Lagomarsino et al. 2016). Here, we argue
841 that RQ and CJ-type factors are intimately linked, and they both promote and prevent species

842 diversification. No single but multiple factors explain the diversification of Apollo butterflies,
843 and the effect of these factors differed across clades. Past changes in climate and elevation
844 during active orogenic periods jointly mediated the rapid radiation of mountain-dwelling
845 *Parnassius*, but also triggered extinction events in non-mountain adapted clades Luehdorfiini
846 and Zerynthiini. The ancestor of Luehdorfiini and Zerynthiini was reconstructed in the
847 Oligocene as occupying a broad distribution range extending from the Caucasus region, the
848 Iranian Plateau and the Zagros Mountains, through Central Asia to India (*red box* in **Figure**
849 **4**). Extinction in Central Asia and India around the Oligocene-Miocene, followed by
850 independent eastward/northward colonizations by Luehdorfiini and Zerynthiini, could explain
851 the current disjunct distribution pattern in these two tribes.

852 Our study also demonstrates that ecological traits and biotic interactions such as host-
853 plant associations and clade-specific diversity-dependence may confer differential
854 diversification capacity in closely related species co-occurring in the same rapidly changing
855 environment. In particular, the radiation of *Parnassius* in the HTP seems to have been
856 fostered by the joint effect of climate and geological changes associated with new biotic
857 interactions. Mountain uplift promoted speciation by allopatric speciation, but also increased
858 habitat heterogeneity, and led ultimately to the formation of new niches (ecological
859 divergence). The colonization of these niches led to new associations with mountain host
860 species, and potentially to the evolution of new morphological adaptive traits, e.g. denser
861 setae on the body or altitudinal melanism. Although it is difficult to tease apart the role of key
862 innovations (new traits that evolved allowing the invasion of a new niche) on ecological
863 opportunity (the formation of a new environment that confers selective advantage fitness for
864 the traits), a combination of extrinsic and intrinsic factors seems to have been necessary for
865 the success of the *Parnassius* radiation. Parnassiinae species feed mostly on plant lineages of
866 small size whose leaves grow close to the ground and which are characteristic elements of

867 modern temperate vegetation: *Aristolochia* and *Asarum* (Aristolochiaceae) for Luehdorfiini
868 and Zerynthiini, *Saxifraga* and *Sedum* (Saxifragaceae) and *Rhodiola* (Crassulaceae) for the
869 subgenus *Parnassius*, or *Corydalis* (Papaveraceae) for the other subgenera of *Parnassius*.
870 Zhang et al. (2014) inferred an origin of *Rhodiola* (70 spp., mid-Miocene; 12.1 Ma, 6.3-20.2
871 Ma) and *Sedum* (420 spp., late Oligocene) in the HTP, later spreading to the adjacent regions.
872 Similarly, Ebersbach et al. (2017) placed the colonization of the HTP by genus *Saxifraga* (370
873 spp.) in the late Oligocene (20-34 Ma). Therefore, there was a synchronous temporal and
874 geographic origin of *Parnassius* and its associated host plants in the HTP region. Conversely,
875 Luehdorfiini and Zerynthiini were unable to evolve new traits or establish novel host-plant
876 associations to invade the mountain and/or host-plant niches. This calls for further studies to
877 decipher the genomic baseline for the adaptations to elevation and host plants by comparing
878 Parnassiini and Luehdorfiini/Zerynthiini (Edger et al. 2015).
879
880 ***Underlying model and inference framework in diversification analyses***
881 Models of diversification have traditionally relied on time-continuous (independent) or
882 autocorrelated rates in the way they model rate-shifts, with few studies applying both types
883 for macroevolutionary inferences. Here we used both, in particular to estimate speciation and
884 extinction rates depending on an environmental variable that itself varies through time. The
885 first model is an ML implementation with a time-continuous birth-death model (Condamine et
886 al. 2013). For the purpose of this study we implemented Bayesian paleoenvironment-
887 dependent models but with autocorrelated (RevBayes) and time-continuous (TESS) rates of
888 diversification. Our comparison across methods has revealed new aspects on the behavior of
889 the models. The three methods are congruent on the role of orogeny (no effect of orogeny on
890 diversification *per se*), but they disagree on the effect of temperature (**Figure 8**). The time-
891 continuous models (RPANDA and TESS) support a positive dependence on speciation and

892 extinction, while the autocorrelated model (RevBayes) shows no dependence. The results
893 suggest that the lack of correlation found by the RevBayes paleoenvironmental model is
894 directly a consequence of the assumption of autocorrelation in rates across time intervals.
895 Thus, the difference lies in the underlying diversification model (autocorrelated vs. time-
896 continuous rates) and not due to the inference framework (Bayesian vs. ML).

897 Our study shows that multiple approaches must be combined to fully address the
898 diversification of clades in relation to abiotic and biotic factors. Rather than a single factor,
899 the joint effect of multiple factors (biogeography, species traits, environmental drivers, and
900 mass extinction) is responsible for current diversity patterns, and we recommend a
901 correspondingly multifaceted strategy for the study of these patterns.

902

903 **AUTHOR CONTRIBUTIONS**

904 F.L.C., F.A.H.S. and I.S. designed the study; F.L.C. and F.A.H.S collected the data; F.L.C.
905 analyzed the data with the help of J.R. on SSE models, S.H. on Bayesian implementation of
906 diversification models, and I.S. on biogeography; F.L.C. and I.S. wrote the paper with
907 significant contributions from J.R., S.H., and F.A.H.S.

908

909 **SUPPLEMENTARY MATERIAL**

910 Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.[NNNN].

911

912 **FUNDING**

913 This study has been funded by a Marie Curie Action (EU 7th Framework Programme)
914 (BIOMME project, IOF-627684) to F.L.C. (jointly supervised by F.A.H.S. and I.S.), by a
915 Natural Sciences and Engineering Research Council of Canada (NSERC), Discovery Grant to
916 F.A.H.S., and by MINECO/FEDER (CGL2015-67489-P) to I.S.

917

918 **ACKNOWLEDGMENTS**

919 We thank Sylvain Piry for help on the script for GenBank, Mark Miller for assistance on the
920 CIPRES cluster.

921

922 **REFERENCES**

923 Ackery P.R. 1975. A guide to the genera and species of Parnassiinae (Lepidoptera:
924 Papilionidae). Bull. British Mus. (Nat. Hist.), Entomol. 31:71-105.

925 Armijo R., Lacassin R., Coudurier-Curveur A., Carrizo D. 2015. Coupled tectonic evolution
926 of Andean orogeny and global climate. Earth-Sci. Rev. 143:1-35.

927 Ayres D.L., Darling A., Zwickl D.J., Beerli P., Holder M.T., Lewis P.O., Huelsenbeck J.P.,
928 Ronquist F., Swofford D.L., Cummings M.P., Rambaut A., Suchard M.A. 2012.
929 BEAGLE: an application programming interface and high-performance computing
930 library for statistical phylogenetics. Syst. Biol. 61:170-173.

931 Baele G., Li W.L.S., Drummond A.J., Suchard M.A., Lemey P. 2013. Accurate model
932 selection of relaxed clocks in Bayesian phylogenetics. Mol. Biol. Evol. 30:239–243.

933 Barnosky A.D. 2001. Distinguishing the effects of the Red Queen and Court Jester on
934 Miocene mammal evolution in the northern Rocky Mountains. J. Verteb. Paleontol.
935 21:172-185.

936 Beaulieu J.M., O'Meara B.C. 2016. Detecting hidden diversification shifts in models of trait-
937 dependent speciation and extinction. Syst. Biol. 65:583–601.

938 Benton M.J. 2009. The red queen and the court jester: species diversity and the role of biotic
939 and abiotic factors through time. Science 323:728–732.

940 Benton M.J. 2015. Exploring macroevolution using modern and fossil data. Proc. R. Soc. B
941 282:20150569.

942 Böhme M. 2003. The Miocene Climatic Optimum: evidence from ectothermic vertebrates of
943 Central Europe. Palaeogeogr. Palaeoclim. Palaeoecol. 195:389-401.
944 Bollino M., Racheli T. 2012. Parnassiinae (Partim). Parnassiini (Partim), Luehdorfiini,
945 Zerynthiini. In: Bauer E., Frankenbach T., editors. Butterflies of the World,
946 Supplement 20. Germany: Goecke und Evers, Keltern.
947 Bouchenak-Khelladi Y., Onstein R.E., Xing Y., Schwery O., Linder H.P. 2015. On the
948 complexity of triggering evolutionary radiations. New Phytol. 207:313–326.
949 Bouilhol P., Jagoutz O., Hanchar J.M., Dudas F.O. 2013. Dating the India–Eurasia collision
950 through arc magmatic records. Earth Planet. Sci. Lett. 366:163–175.
951 Brikiatis L. 2014. The De Geer, Thulean and Beringia routes: key concepts for understanding
952 early Cenozoic biogeography. J Biogeogr. 41:1036-1054.
953 Britton T., Anderson C.L., Jacquet D., Lundqvist S., Bremer K. 2007. Estimating divergence
954 times in large phylogenetic trees. Syst. Biol. 56:741–752.
955 Brock C.D., Harmon L.J., Alfaro M.E. 2011. Testing for temporal variation in diversification
956 rates when sampling is incomplete and nonrandom. Syst. Biol. 60:410-419.
957 Carpenter F.M. 1992. Treatise on Invertebrate Paleontology, Part R, Arthropoda 3-4.
958 Geological Society of America, Boulder, Colorado.
959 Churkin S. 2006. A new species of *Parnassius* Latreille, 1804 from Kyrgyzstan (Lepidoptera,
960 Papilionidae) Helios 7:142-158.
961 Collins N.M., Morris M.G. 1985. Threatened Swallowtail Butterflies of the World. The IUCN
962 Red Data Book, Cambridge.
963 Condamine F.L., Sperling F.A., Wahlberg N., Rasplus J.Y., Kergoat G.J. 2012. What causes
964 latitudinal gradients in species diversity? Evolutionary processes and ecological
965 constraints on swallowtail biodiversity. Ecol. Lett. 15:267–277.

966 Condamine F.L., Rolland J., Morlon H. 2013. Macroevolutionary perspectives to
967 environmental change. Ecol. Lett. 16:72–85.
968 Condamine F.L., Toussaint E.F.A., Clamens A.-L., Genson G., Sperling F.A.H., Kergoat G.J.
969 2015a. Deciphering the evolution of birdwing butterflies 150 years after Alfred
970 Russell Wallace. Sci. Rep. 5:11860.
971 Condamine F.L., Nagalingum N.S., Marshall C.R., Morlon H. 2015b. Origin and
972 diversification of living cycads: A cautionary tale on the impact of the branching
973 process prior on Bayesian molecular dating. BMC Evol. Biol. 15:65.
974 Cornell H.V. 2013. Is regional species diversity bounded or unbounded? Biol. Rev. 88:140–
975 165.
976 Cruaud A., Rønsted N., Chantarasuwan B., Chou L.S., Clement W., Couloux A., Cousins B.,
977 Forest F., Genson G., Harrison R.D., Hossaert-McKey M., Jabbour-Zahab R.,
978 Jousselin E., Kerdelhué C., Kjellberg F., Lopez-Vaamonde C., Peebles J., Pereira
979 R.A.S, Schramm T., Ubaidillah R., van Noort S., Weiblen G.D., Yang D.R, Yan-
980 Qiong P., Yodpinyanee A., Libeskind-Hadas R., Cook J.M., Rasplus J.Y., Savolainen
981 V. 2012. An extreme case of plant-insect co-diversification: figs and fig-pollinating
982 wasps. Syst. Biol. 61:1029-1047.
983 Cusimano N., Renner S.S. 2010. Slowdowns in diversification rates from real phylogenies
984 may be not real. Syst. Biol. 59:458-464.
985 Dapporto L. 2010. Speciation in Mediterranean refugia and post-glacial expansion of
986 *Zerynthia polyxena* (Lepidoptera, Papilionidae). J. Zoolog. Syst. Evol. Res. 48:229–
987 237.
988 Davis M.P., Midford P.E., Maddison W.P. 2013. Exploring power and parameter estimation
989 of the BiSSE method for analysing species diversification. BMC Evol. Biol. 13:38.

990 DeChaine E.G., Martin A.P. 2004. Historic cycles of fragmentation and expansion in
991 *Parnassius smintheus* (Papilionidae) inferred using mitochondrial DNA. Evolution
992 58:113–127.
993 Donoghue MJ, Sanderson MJ. 2015. Confluence, synnovation, and depauperons in plant
994 diversification. New Phytol. 207:260-274.
995 Donoghue M.J., Doyle J.A., Gauthier J., Kluge A.G., Rowe T. 1989. The importance of
996 fossils in phylogeny reconstruction. Annu. Rev. Ecol. Syst. 20:431–460.
997 Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and
998 dating with confidence. PLoS Biol. 4:e88.
999 Drummond C.S., Eastwood R.J., Miotto S.T., Hughes C.E. 2012a. Multiple continental
1000 radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key
1001 innovation with incomplete taxon sampling. Syst. Biol. 61:443–460.
1002 Drummond A., Suchard M.A., Xie D., Rambaut A. 2012b. Bayesian phylogenetics with
1003 BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969-1973.
1004 Durden C.J., Rose H. 1978. Butterflies from the middle Eocene: The earliest occurrence of
1005 fossil Papilionidae. Prarce-Sellards Ser. Tax. Mem. Mus. 29:1–25.
1006 Ebersbach J., Muellner-Riehl A.N., Michalak I., Tkach N., Hoffmann M.H., Röser M., Sun
1007 H., Favre A. 2017. In and out of the Qinghai-Tibet Plateau: divergence time estimation
1008 and historical biogeography of the large arctic-alpine genus *Saxifraga* L. J. Biogeogr.
1009 44:900-910.
1010 Edger P.P., Heidel-Fischer H.M., Bekaert M., Rota J., Glöckner G., Platts A.E., Heckel D. G.,
1011 Der J., Wafula E., Tang M., Hofberger J.A., Smithson A., Hall J.C., Blanchette M.,
1012 Bureau T.E., Wright S.I., dePamphilis C.W., Schranz M.E., Barker M.S., Conant G.C.,
1013 Wahlberg N., Vogel H., Pires J.C., Wheat C.W. 2015: The butterfly plant arms-race

escalated by gene and genome duplications. Proc. Natl. Acad. Sci. USA 112:8362-8366.

Edwards E.J., Osborne C.O., Stromberg C.A.E., Smith S.A., the C_4 Grasses Consortium. 2010. The origins of C_4 grasslands: integrating evolutionary and ecosystem science. Science 328:587–591.

Ehrlich P.R., Raven P.H. 1964. Butterflies and plants: a study in coevolution. Evolution 18:586–608.

Erwin D.H. 2009. Climate as a driver of evolutionary change. Curr. Biol. 19:R575–R583.

Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc. Roy. Soc. Lond. B 279:1300–1309.

Ezard T.H.G., Aze T., Pearson P.N., Purvis A. 2011. Interplay between changing climate and species' ecology drives macroevolutionary dynamics. Science 332:349–351.

Ezard T.H.G., Quental T.B., Benton M.J. 2016. The challenges to inferring the regulators of biodiversity in deep time. Philos. Trans. R. Soc. B 371:20150216.

Favre A., Päckert M., Pauls S.U., Jähnig S.C., Uhl D., Michalak I., Muellner-Riehl A.N. 2015. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. Biol. Rev. 90:236-253.

FitzJohn R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3:1084-1092.

FitzJohn R.G., Maddison W.P., Otto S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Syst. Biol. 58:595–611.

Fjeldså J., Bowie R.C.K., Rahbek C. 2012. The role of mountain ranges in the diversification of birds. Annu. Rev. Ecol. Evol. Syst. 43:249-265.

1038 Frankenbach T., Bollino M., Racheli T. 2012. Papilionidae XIV: Hypermnestra, Luehdorfiini,
1039 Zerynthiini. In: Bauer E., Frankenbach T., editors. Butterflies of the World, Part 36.
1040 Germany: Goecke und Evers, Keltern.
1041 Gavin D.G., Fitzpatrick M.C., Gugger P.F., Heath K.D., Rodríguez-Sánchez F., Dobrowski
1042 S.Z., Hampe A., Hu F.S., Ashcroft M.B., Bartlein P.J., Blois J.L., Carstens B.C., Davis
1043 E.B., de Lafontaine G., Edwards M.E., Fernandez M., Henne P.D., Herring E.M.,
1044 Holden Z.A., Kong W.S., Liu J., Magri D., Matzke N.J., McGlone M.S., Saltré F.,
1045 Stigall A.L., Tsai Y.H., Williams J.W. 2014. Climate refugia: joint inference from
1046 fossil records, species distribution models and phylogeography. New Phytol. 204:37-
1047 54.
1048 Gernhard T. 2008. The conditioned reconstructed process. J. Theor. Biol. 253:769–778.
1049 Givnish T.J. Spalink D., Ames M., Lyon S.P., Hunter S.J., Zuluaga A., Iles W.J.D., Clements
1050 M.A., Arroyo M.T.K., Leebens-Mack J., Endara L., Kriebel R., Neubig K.M., Whitten
1051 W.M., Williams N.H., Cameron K.M. 2015. Orchid phylogenomics and multiple
1052 drivers of their extraordinary diversification. Proc. R. Soc. B 282: 20151553.
1053 Goldberg E.E., Lancaster L.T., Ree R.H. 2011. Phylogenetic inference of reciprocal effects
1054 between geographic range evolution and diversification. Syst. Biol. 60:451–465.
1055 Gradstein F.M, Ogg J.G., Schmitz M., Ogg G. 2012. The geologic time scale 2012. Boston
1056 (USA): Elsevier. 1176 pp.
1057 Gratton P., Konopiński M.K., Sbordoni V. 2008. Pleistocene evolutionary history of the
1058 clouded apollo (*Parnassius mnemosyne*): genetic signatures of climate cycles and a
1059 'time-dependent' mitochondrial substitution rate. Mol. Ecol. 17:4248–4262.
1060 Hiura L. 1980. A phylogeny of the genera of Parnassiinae based on analysis of wing pattern,
1061 with description of a new genus (Lepidoptera: Papilionidae). Bull. Osaka Mus. Natural
1062 Hist. 33:71-85.

1063 Ho S.Y.W., Phillips M.J. 2009. Accounting for calibration uncertainties in phylogenetic
1064 estimation of evolutionary divergence times. Syst. Biol. 58:367–380.
1065 Höhna S. 2015. The time-dependent reconstructed evolutionary process with a key-role for
1066 mass-extinction events. J. Theoret. Biol. 380:321–331.
1067 Höhna S., Stadler T., Ronquist F., Britton T. 2011. Inferring speciation and extinction rates
1068 under different sampling schemes. Mol. Biol. Evol. 28:2577-2589.
1069 Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P.,
1070 Ronquist F. 2016a. RevBayes: Bayesian phylogenetic inference using graphical
1071 models and an interactive model-specification language. Syst. Biol. 65:726–736.
1072 Höhna S., May M.R., Moore B.R. 2016b. TESS: an R package for efficiently simulating
1073 phylogenetic trees and performing Bayesian inference of lineage diversification rates.
1074 Bioinformatics 32:789–791.
1075 Hoorn C., Mosbrugger V., Mulch A., Antonelli A. 2013 Biodiversity from mountain building.
1076 Nature Geosci. 6:154-154.
1077 Huelsenbeck J.P., Larget B., Alfaro M.E. 2004. Bayesian phylogenetic model selection using
1078 reversible jump Markov chain Monte Carlo. Mol. Biol. Evol. 21:1123–1133.
1079 Hughes C.E., Atchison G.W. 2015. The ubiquity of alpine plant radiations: from the Andes to
1080 the Hengduan Mountains. New Phytol. 207:275–282.
1081 Hutter C.R., Guayasamin J.M., Wiens J.J. 2013. Explaining Andean megadiversity: the
1082 evolutionary and ecological causes of glassfrog elevational richness patterns. Ecol.
1083 Lett. 16:1135–1144.
1084 Janz N., Nylin S., Wahlberg N. 2006. Diversity begets diversity: host expansions and the
1085 diversification of plant-feeding insects. BMC Evol. Biol. 6:4.
1086 de Jong R. 2007. Estimating time and space in the evolution of the Lepidoptera. Tijdschrift v.
1087 Entomol. 150:319–346.

1088 Jønsson K.A., Fabre P.H., Fritz S.A., Etienne R.S., Ricklefs R.E., Jørgensen T.B., Fjeldså J.,
1089 Rahbek C., Ericson P.G.P., Woogg F., Pasquet E., Irestedt M. 2012. Ecological and
1090 evolutionary determinants for the adaptive radiation of the Madagascan vangas. Proc.
1091 Natl. Acad. Sci. USA 109:6620-6625.
1092 Kass R.E., Raftery A.E. 1995. Bayes factors. J. Am. Stat. Assoc. 90:773–795.
1093 Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7:
1094 improvements in performance and usability. Mol. Biol. Evol. 30:772-780.
1095 Katoh T., Chichvarkin A., Yagi T., Omoto K. 2005. Phylogeny and evolution of butterflies of
1096 the genus *Parnassius*: inferences from mitochondrial 16S and ND1 sequences. Zool.
1097 Sci. 22:343–351.
1098 Keyghobadi N, Roland J, Strobeck C. 2005. Genetic differentiation and gene flow among
1099 populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape
1100 connectivity. Mol. Ecol. 14:1897-1909.
1101 Kocman S. 2009. Parnassius of Tibet and the Adjacent Areas. Vadim Tshikolovets,
1102 Pardubice, Czech Republic.
1103 Lagomarsino L.P., Condamine F.L., Antonelli A., Mulch A., Davis C.C. 2016. The abiotic
1104 and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae).
1105 New Phytol. 210:1430-1442.
1106 Lanfear R., Calcott B., Ho SYW, Guindon S. 2012. PartitionFinder: Combined selection of
1107 partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol.
1108 Evol. 29:1695-1701.
1109 Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological
1110 character data. Syst. Biol. 50:913–925.

1111 Liow L.H., Reitan T., Harnik P.G. 2015. Ecological interactions on macroevolutionary time
1112 scales: clams and brachiopods are more than ships that pass in the night. Ecol. Lett.
1113 18:1030-1039.
1114 Liu Z., Pagani M., Zinniker D., DeConto R., Huber M., Brinkhuis H., Shah S.R., Leckie
1115 R.M., Pearson A. 2009. Global cooling during the Eocene-Oligocene climate
1116 transition. Science 323:1187-1190.
1117 Maddison W.P., FitzJohn R.G. 2015. The unsolved challenge to phylogenetic correlation tests
1118 for categorical characters. Syst. Biol. 64:127-136.
1119 Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a binary character's effect on
1120 speciation and extinction. Syst. Biol. 56:701–710.
1121 Magallón S., Gómez-Acevedo S., Sánchez-Reyes L.L., Hernández-Hernández T. 2015. A
1122 metacalibrated time-tree documents the early rise of flowering plant phylogenetic
1123 diversity. New Phytol. 207:437-453.
1124 Matzke N.J. 2014. Model selection in historical biogeography reveals that founder-event
1125 speciation is a crucial process in island clades. Syst. Biol. 63:951-970.
1126 May M.R., Höhna S., Moore B.R. 2016. A Bayesian approach for detecting the impact of
1127 mass-extinction events on molecular phylogenies when rates of lineage diversification
1128 may vary. Methods Ecol. Evol. 7:947–959.
1129 Mayhew P.J., Jenkins G.N., Benton T.G. 2008. A long-term association between global
1130 temperature and biodiversity, origination and extinction in the fossil record. Proc. R.
1131 Soc. Lond. B. 275:47–53.
1132 Michel F., Rebourg C., Cosson E., Descimon H. 2008. Molecular phylogeny of Parnassiinae
1133 butterflies (Lepidoptera: Papilionidae) based on the sequences of four mitochondrial
1134 DNA segments. Ann. Soc. Entomol. Fr. 44:1-36.

1135 Migliore J., Baumel A., Juin M., Fady B., Roig A., Duong N., Médail F. 2013. Surviving in
1136 Mountain climate refugia: New insights from the genetic diversity and structure of the
1137 relict shrub *Myrtus nivellei* (Myrtaceae) in the Sahara Desert. PLoS One 8:e73795.
1138 Miller M.A., Schwartz T., Pickett B.E., He S., Klem E.B., Scheuermann R.H., Passarotti M.,
1139 Kaufman S., O'Leary M.A. 2015. A RESTful API for access to phylogenetic tools via
1140 the CIPRES Science Gateway. Evol Bioinformatics 11:43–48.
1141 Moen D., Morlon H. 2014. Why does diversification slow down? Trends Ecol. Evol. 29:190-
1142 197.
1143 Moore B.R., Höhna S., May M.R., Rannala B., Huelsenbeck J.P. 2016. Critically evaluating
1144 the theory and performance of Bayesian analysis of macroevolutionary mixtures. Proc.
1145 Natl. Acad. Sci. USA 113:9569–9574.
1146 Morlon H. 2014. Phylogenetic approaches for studying diversification Ecol. Lett. 17:508-525.
1147 Morlon H., Parsons T.L., Plotkin J. 2011. Reconciling molecular phylogenies with the fossil
1148 record. Proc. Natl. Acad. Sci. USA 108:16327–16332.
1149 Nazari V., Sperling F.A.H. 2007. Mitochondrial DNA divergence and phylogeography in
1150 western Palaearctic Parnassiinae (Lepidoptera: Papilionidae): how many species are
1151 there? Insect Syst. Evol. 38:121–138.
1152 Nazari V., Zakharov E.V., Sperling F.A.H. 2007. Phylogeny, historical biogeography, and
1153 taxonomic ranking of Parnassiinae (Lepidoptera: Papilionidae) based on morphology
1154 and seven genes. Mol. Phylogenet. Evol. 42:131-156.
1155 Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Phil. Trans. R.
1156 Soc. B 344:305–311.
1157 Ng J., Smith S.D. 2014. How traits shape trees: new approaches for detecting character state-
1158 dependent lineage diversification. J. Evol. Biol. 27:2035-2045.

1159 Omoto K., Katoh T., Chichvarkin A., Yagi T. 2004. Molecular evolution of the 'apollo'
1160 butterflies of the genus *Parnassius* (Lepidoptera: Papilionidae). Gene 326:141-147.
1161 Omoto K., Yonezawa T., Shinkawa T. 2009. Molecular systematics and evolution of the
1162 recently discovered "Parnassian" butterfly (*Parnassius davydovi* Churkin, 2006) and
1163 its allied species (Lepidoptera, Papilionidae). Gene 441:80-88.
1164 Padma T.V. 2014. Himalayan plants seek cooler climes. Nature 512:359.
1165 Palazzesi L., Hidalgo O., Barreda V.D, Forest F., Höhna S. submitted. Radiation of the
1166 Grassland biome during a falling CO_2 world.
1167 Pérez-Escobar O.A., Chomicki G., Condamine F.L., Karremans A.P., Bogarín D., Matzke
1168 N.J.,Silvestro D., Antonelli A. 2017. Recent origin and rapid speciation of Neotropical
1169 orchids in the world's richest plant biodiversity hotspot. New Phytol. 215:891-905.
1170 Phillimore A.B., Price T.D. 2008. Density-dependent cladogenesis in birds. PLoS Biol. 6:e71.
1171 Pokorny L., Riina R., Mairal M., Meseguer A.S., Culshaw V., Cendoya J., Serrano M.,
1172 Carbajal R., Ortiz S., Heuertz M., Sanmartín I. 2015. Living on the edge: timing of
1173 Rand Flora disjunctions congruent with ongoing aridification in Africa. Frontiers
1174 Genet. 6:154.
1175 Pound M.J., Haywood A.M., Salzmann U., Riding J.B. 2012. Global vegetation dynamics and
1176 latitudinal temperature gradients during the Mid to Late Miocene (15.97–5.33 Ma).
1177 Earth-Sci. Rev. 112:1–22.
1178 Price T.D., Hooper D.M., Buchanan C.D., Johansson U.S., Tietze D.T., Alström P., Olsson
1179 U., Ghosh-Harihar M., Ishtiaq F., Gupta S.K., Martens J., Harr B., Singh P., Mohan D.
1180 2014. Niche filling slows the diversification of Himalayan songbirds. Nature 509:222-
1181 225.
1182 Quade J., Cerling T., Bowman J. 1989. Development of Asian monsoon revealed by marked
1183 ecological shift during the latest Miocene in Northern Pakistan. Nature 342:163–166.

1184 Rabosky D.L., Santini F., Eastman J.M., Smith S.A., Sidlauskas B., Chang J., Alfaro M.E.
1185 2013. Rates of speciation and morphological evolution are correlated across the largest
1186 vertebrate radiation. Nat. Commun. 4:1958.
1187 Rabosky D.L., Lovette, I.J. 2008. Explosive evolutionary radiations: decreasing speciation or
1188 increasing extinction through time? Evolution 62:1866–1875.
1189 Rabosky D.L., Goldberg E.E. 2015. Model inadequacy and mistaken inferences of trait-
1190 dependent speciation. Syst. Biol. 64:340-355.
1191 Rabosky D.L., Donnellan S.C., Grundler M., Lovette I.J. 2014. Analysis and visualization of
1192 complex macroevolutionary dynamics: An example from Australian scincid lizards.
1193 Syst. Biol. 63:610-627.
1194 Rabosky D.L., Grundler M., Anderson C., Shi J.J., Brown J.W., Huang H., Larson J.G.
1195 2014b. BAMMtools: an R package for the analysis of evolutionary dynamics on
1196 phylogenetic trees. Methods Ecol. Evol. 5:701-707.
1197 Rabosky D.L., Mitchell J.S., Chang J. 2017. Is BAMM flawed? Theoretical and practical
1198 concerns in the analysis of multi-rate diversification models. Syst. Biol.
1199 doi.org/10.1093/sysbio/syx037.
1200 Rambaut A., Drummond A.J. 2009. Tracer v1.6. Available at http://beast.bio.ed.ac.uk/Tracer.
1201 Rasnitsyn A.P., Zherikhin V.V. 2002. Appendix: Alphabetic list of selected insect fossil sites.
1202 Impression fossils. In: Rasnitsyn A.P., Quicke D.L.J. (Ed.), History of Insects. Kluwer
1203 Academic Publishers, Dordrecht, Boston & London, pp. 437–444.
1204 Rebel H. 1898. Fossile Lepidopteren aus der Miocänformation von Gabbro. Sitzungsberichte
1205 der Kaiserlichen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche
1206 Classe 107:731-745.
1207 Rebourg C., Péténian F., Cosson E., Faure E. 2006. Patterns of speciation and adaptive
1208 radiation in *Parnassius* butterflies. J. Entomol. 3:204-215.

1209 Ree R.H., Smith S.A. 2008. Maximum-likelihood inference of geographic range evolution by
1210 dispersal, local extinction, and cladogenesis. Syst. Biol. 57:4–14.
1211 Renner S.S. 2016. Available data point to a 4-km-high Tibetan Plateau by 40 Ma, but 100
1212 molecular-clock papers have linked supposed recent uplift to young node ages. J.
1213 Biogeogr. 43:1479-1487.
1214 Revell, L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other
1215 things). Methods Ecol. Evol. 3:217-223.
1216 Rolland J., Condamine F.L., Jiguet F., Morlon H. 2014. Faster speciation and reduced
1217 extinction in the tropics contribute to the mammalian latitudinal diversity gradient.
1218 PLoS Biol. 12:e1001775.
1219 Roland J., Keyghobadi N., Fownes S. 2000. Alpine *Parnassius* butterfly dispersal: effects of
1220 landscape and population size. Ecology 81:1642–1653.
1221 Ronquist F., Klopfstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P.
1222 2012a. A total-evidence approach to dating with fossils, applied to the early radiation
1223 of the Hymenoptera. Syst. Biol. 61:973–999.
1224 Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu
1225 L., Suchard M.A., Huelsenbeck J.P. 2012b. MrBayes 3.2: Efficient Bayesian
1226 phylogenetic inference and model choice across a large model space. Syst. Biol.
1227 61:539-542.
1228 Ronquist F., Sanmartín I. 2011. Phylogenetic methods in biogeography. Annu. Rev. Ecol.
1229 Evol. Syst. 42:441-464.
1230 Ruane S., Bryson R.W. Jr., Pyron R.A., Burbrink F.T. 2014. Coalescent species delimitation
1231 in milksnakes (genus *Lampropeltis*) and impacts on phylogenetic comparative
1232 analyses. Syst. Biol. 63:231-250.

1233 Sanmartín I., Meseguer A.S. 2016. Extinction in phylogenetics and biogeography: From
1234 timetrees to patterns of biotic assemblage. Front. Genet. 7:35.
1235 Sanmartín I., Enghoff H., Ronquist F. 2001. Patterns of animal dispersal, vicariance and
1236 diversification in the Holarctic. Biol. J. Linn. Soc. 73:345–390.
1237 Sauquet H., Ho S.Y.W., Gandolfo M.A., Jordan G.J., Wilf P., Cantrill D.J., Bayly M.J.,
1238 Bromham L., Brown G.K., Carpenter R.J., Lee D.M., Murphy D.J., Sniderman J.M.K.,
1239 Udovicic F. 2012. Testing the impact of calibration on molecular divergence times
1240 using a fossil-rich group: the case of *Nothofagus* (Fagales). Syst. Biol. 61:289–313.
1241 Schnitzler J., Barraclough T.G., Boatwright J.S., Goldblatt P., Manning J.C., Powell M.P.,
1242 Rebelo T., Savolainen V. 2011. Causes of plant diversification in the Cape
1243 biodiversity hotspot of South Africa. Syst. Biol. 60:343–357.
1244 Schoville S.D., Roderick G.K. 2009. Alpine biogeography of Parnassian butterflies during
1245 Quaternary climate cycles in North America. Mol. Ecol. 18:3471–3485.
1246 Scriber J.M., Tsubaki Y., Lederhouse R.C. 1995. Swallowtail Butterflies: Their Ecology and
1247 Evolutionary Biology. Scientific Publishers, Gainesville.
1248 Scriber J.M. 1984. Larval foodplant utilization by the world Papilionidae (Lep.): latitudinal
1249 gradients reappraised. Acta Rhopaloc. 6:1–50.
1250 Scudder S.H. 1875. Fossil Butterflies. Memoirs of the American Association for the
1251 Advancement of Science I: I-XI. Salem, Massachusetts. pp. 1–99.
1252 Sepulchre P., Ramstein G., Fluteau F., Schuster M., Tiercelin J.-J., Brunet M. 2006. Tectonic
1253 uplift and Eastern Africa aridification. Science 313:1419-1423.
1254 Silvestro D., Antonelli A., Salamin N., Quental T.B. 2015. The role of clade competition in
1255 the diversification of North American canids. Proc. Natl. Acad. Sci. USA 112: 8684–
1256 8689.

1257 Smith B.T., McCormack J.E., Cuervo A.M., Hickerson M.J., Aleixo A., Cadena C.D., Pérez-
1258 Emán J., Burney C.W., Xie X., Harvey M.G., Faircloth B.C., Glenn T.C., Derryberry
1259 E.P., Prejean J., Fields S., Brumfield R.T. 2014. The drivers of tropical speciation.
1260 Nature 515:406-409.

1261 Smith M.E., Singer B., Carroll A. 2003. $^{40}Ar/^{39}Ar$ geochronology of the Eocene Green River
1262 Formation, Wyoming. Geol. Soc. Am. Bull. 115:549-565.

1263 Sohn J.-C., Labandeira C., Davis D., Mitter C. 2012. An annotated catalog of fossil and
1264 subfossil Lepidoptera (Insecta: Holometabola) of the world. Zootaxa 3286:1–132.

1265 Sohn J.-C., Labandeira C.C., Davis D.R. 2015. The fossil record and taphonomy of butterflies
1266 and moths (Insecta, Lepidoptera): implications for evolutionary diversity and
1267 divergence-time estimates. BMC Evol. Biol. 15:12.

1268 Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proc. Natl.
1269 Acad. Sci. USA 108:6187–6192.

1270 Stadler T. 2013. Recovering speciation and extinction dynamics based on phylogenies. J.
1271 Evol. Biol. 26:1203-1219.

1272 Strimmer K., Pybus O.G. 2001. Exploring the demographic history of DNA sequences using
1273 the generalized skyline plot. Mol. Biol. Evol. 18:2298-2305.

1274 Tan J., Kelly C.K., Jiang L. 2013. Temporal niche promotes biodiversity during adaptive
1275 radiation. Nature Comm. 4:2102.

1276 Todisco V., Gratton P., Cesaroni D., Sbordoni V. 2010. Phylogeography of *Parnassius*
1277 *apollo*: hints on taxonomy and conservation of a vulnerable glacial butterfly invader.
1278 Biol. J. Linn. Soc. 101:169–183.

1279 Todisco V., Gratton P., Zakharov E.V., Wheat C.W., Sbordoni V., Sperling F.A.H. 2012.
1280 Mitochondrial phylogeography of the Holarctic *Parnassius phoebus* complex supports
1281 a recent refugial model for alpine butterflies. J. Biogeogr. 39:1058-1072.

1282 Tyler H.A., Brown K.S., Wilson K. 1994. Swallowtail Butterflies of the Americas: A Study in
1283 Biological Dynamics, Ecological Diversity, Biosystematics and Conservation.
1284 Scientific Publishers, Gainsville.
1285 Van Valen L.M. 1973. A new evolutionary law. Evol. Theory 1:1-30.
1286 Voje K.L., Holen Ø.H., Liow L.H., Stenseth N.C. 2015. The role of biotic forces in driving
1287 macroevolution: beyond the Red Queen. Proc. Roy. Soc. Lond. B 282: 20150186.
1288 Wagner C.E., Harmon L.J., Seehausen O. 2012. Ecological opportunity and sexual selection
1289 together predict adaptive radiation. Nature 487:366-369.
1290 Wallace A.R. 1878. Tropical Nature and Other Essays. Macmillan and Co, London.
1291 Weiss J.C. 1991-2005. The Parnassiinae of the World. Part I 1991, II 1992, III 1999, IV 2005.
1292 Sciences Nat, Venette, 400 p.
1293 Wen J., Zhang J.-C., Nie Z.-L., Zhong Y., Sun H. 2014. Evolutionary diversifications of
1294 plants on the Qinghai-Tibetan Plateau. Front. Genet. 5:4.
1295 Xie W., Lewis P.O., Fan Y., Kuo L., Chen M.H. 2011. Improving marginal likelihood
1296 estimation for Bayesian phylogenetic model selection. Syst. Biol. 60:150–160.
1297 Xing Y., Ree, R.H. 2017. Uplift-driven diversification in the Hengduan Mountains, a
1298 temperate biodiversity hotspot. Proc. Natl Acad. Sci. USA 114:E3444-E3451.
1299 Yang Z., Rannala B. 2006. Bayesian estimation of species divergence times under a
1300 molecular clock using multiple calibrations with soft bounds. Mol. Biol. Evol. 23:212–
1301 226.
1302 Zachos J.C., Dickens G.R., Zeebe R.E. 2008. An early Cenozoic perspective on greenhouse
1303 warming and carbon-cycle dynamics. Nature 451:279–283.
1304 Zhang J.-Q., Meng S.-Y., Allen G.A., Wen J., Rao G.-Y. 2014. A rapid radiation and
1305 dispersal out of the Qinghai-Tibetan Plateau of an alpine plant lineage *Rhodiola*
1306 (Crassulaceae). Mol. Phylogenet. Evol. 23:147–158.

1307 Zhisheng A., Kutzbach J.E., Prell W.L., Porter S.C. 2001. Evolution of Asian monsoons and
1308 phased uplift of the Himalaya Tibetan plateau since Late Miocene times. Nature
1309 411:62-66.
1310 Zinetti F., Dapporto L., Vovlas A., Chelazzi G., Bonelli S., Balletto E., Ciofi C. 2013. When
1311 the rule becomes the exception. No evidence of gene flow between two *Zerynthia*
1312 cryptic butterflies suggests the emergence of a new model group. PLoS One 8:e65746.
1313

TABLES

Table 1. Summary of diversification analyses. Evolutionary models are divided by whether they are used to test hypotheses in favor of the Red Queen or to support the Court Jester. Abbreviations: MCC, maximum clade credibility; HTP, Himalaya and Tibetan Plateau.

Type of birth-death	Evolutionary model	Methods used	References	Data used	Settings	Main results
Time-dependence (rates vary continuously as a function of time and across clades)	Red Queen	BAMM RevBayes	Rabosky et al. (2014) Höhna et al. (2016a)	MCC tree of Parnassiinae	Bayesian model to test whether speciation rates have varied over time and/or across lineages (exponential variation)	A single shift: increase of diversification near the crown of the genus *Parnassius*
Diversity-dependence (rates vary as a function of the number of species)	Red Queen	DDD (*dd_ML*)	Etienne et al. (2012)	MCC tree of the clade and one of *Parnassius* (to test whether they reach their carrying capacity)	5 models to test whether speciation declines with diversity and/or extinction increases with diversity	Both Parnassiinae and Parnassius have not reached their carrying capacity, but show some degree of diversity dependence
Trait-dependence (rates vary as a function of a character state for a trait)	Red Queen	MuSSE (*make.musse*) RevBayes	Maddison et al. (2007) FitzJohn et al. (2009) Höhna et al. (2016a)	200 posterior trees + host plant preferences (four states)	36 models to test whether a character state impacted speciation and/or extinction and/or transition rates between host plants	Support for independent speciation rates for each host-specialist clade; no significant differences for extinction and transition rates among clades
Episodic birth-death model (rates vary discretely as a function of time)	Court Jester	TreePar (*bd.shifts.optim*) CoMET	Stadler (2011) May et al. (2016)	200 posterior trees MCC tree	4 models to test whether diversification rates and turnover episodically changed over time	Significant diversification rate and turnover changes in the late Pliocene (3.2 Ma) and a possible mass extinction in the mid-Miocene (15 Ma)
Trait-dependence (rates vary as a function of a character state for a trait)	Court Jester	GeoSSE (*make.geosse*) RevBayes	Maddison et al. (2007) Goldberg et al. (2011) Höhna et al. (2016a)	200 posterior trees + geographical occurrence (highland, lowland, or widespread)	12 models to test whether a character state impacted speciation and/or extinction and/or transition rates between mountains and lowlands	High speciation in mountain, medium allopatric speciation and low speciation in lowland (no effect of the trait on extinction and transition rates)
Environmental-dependence (rates vary as a function of a time-variable environment)	Court Jester	RPANDA (*fit_env*) TESS RevBayes	Condamine et al. (2013) Höhna et al. (2016b) Höhna (in prep.)	200 posterior trees + past global temperatures + past altitude of the HTP	8 models for each variable to test whether rates vary with the paleoenvironment	Speciation and extinction are positively associated with warm climate. Speciation or extinction is not significantly associated with past altitude

Table 2. Results from episodic diversification analyses in TreePar with or without mass extinction. When mass extinction is disallowed, the best-supported model has a single rate-shift, as determined by the lowest AICc and ΔAICc. When mass extinction is allowed, the best model also has a single mass-extinction event. Adding more diversification shifts or mass extinctions did not significantly improve the likelihood of the models. Abbreviations: BD, birth-death; ME, mass extinction; NP, number of parameters; logL, log-likelihood; AICc, corrected Akaike Information Criterion; ΔAICc, the difference in AICc between the model with the lowest AICc and the others. Parameter estimates: r, net diversification rate (speciation minus extinction); ε, turnover (extinction over speciation); st, shift time; sp, survival probability at a mass extinction event. 'r1' denotes the diversification rate and 'ε1' is the turnover, both inferred between Present and the shift time 1 ('st1').

	Mass extinction events disallowed					**Mass extinction events allowed**		
Models	**BD with no shift time**	**BD with 1 shift time**	**BD with 2 shift times**	**BD with 3 shift times**	**Models**	**BD with no ME**	**BD with 1 ME**	**BD with 2 MEs**
NP	2	**5**	8	11	**NP**	2	**4**	6
logL	-182.840 ±0.5640	**-175.099 ±0.5727**	-172.910 ±0.5716	-170.504 ±0.5778	**logL**	-240.789 ±0.5634	**-235.792 ±0.5667**	-234.082 ±0.5760
AICc	369.827 ±1.1281	**360.959 ±1.1454**	363.714 ±1.1432	366.624 ±1.1555	**AICc**	485.726 ±1.1267	**482.345 ±1.1334**	486.058 ±1.1520
ΔAIC	8.868	**0**	2.755	5.665	**ΔAIC**			
r1	0.0934 ±0.0008	**0.1188 ±0.0057**	0.0915 ±0.0067	0.0711 ±0.0105	**r**	0.0930 ±0.0008	**0.1672 ±0.0012**	0.1859 ±0.0014
ε1	0.5374 ±0.0047	**0.1699 ±0.0355**	0.3109 ±0.0411	0.5288 ±0.1224	**ε**	0.5363 ±0.0042	**0.0279 ±0.0038**	0.0025 ±0.0010
st1	-	**3.21 ±0.0573**	2.81 ±0.0693	2.56 ±0.0788	**st1**	-	**14.82 ±0.12**	7.37 ±0.17
r2	-	**0.0312 ±0.0007**	0.0254 ±0.0166	0.0031 ±0.0295	**sp1**	-	**0.0749 ±0.0021**	0.466 ±0.0144
ε2	-	**0.9142 ±0.0052**	0.8319 ±0.0529	0.9179 ±0.0795	**st2**	-	-	16.08 ±0.29
st2	-	-	15.29 ±0.8932	10.47 ±0.6951	**sp2**	-	-	0.0759 ±0.0028
r3	-	-	-0.3498 ±0.1026	0.0302 ±0.0190				
ε3	-	-	0.7616 ±0.0380	0.8461 ±0.0506				
st3	-	-	-	20.29 ±0.9277				
r4	-	-	-	-0.9929 ±0.1753				
ε4	-	-	-	0.8571 ±0.0403				

Table 3. Results from RPANDA analyses. The model with positive dependence of speciation and extinction on temperature has the lowest AICc and highest AICω. Macroevolutionary scenarios were compared in which speciation and extinction vary through time (Time), or by paleo-elevation changes of the Himalaya and Tibetan Plateau (Alti.), and global temperature changes (Temp.). Abbreviations: NP, number of parameters; logL, log-likelihood; AICc, corrected Akaike Information Criterion; ΔAICc, the difference in AICc between the model with the lowest AICc and the others; AICω, Akaike weight. Parameter estimates: λ_0 and μ_0, speciation and extinction for environmental value at present; α and β, parameter controlling variation of speciation (α) and extinction (β) with paleoenvironment.

Models	NP	logL	AICc	ΔAIC	AICω	λ_0	α	μ_0	β
Constant birth-death	2	-240.336 ±0.620	484.819 ±1.240	5.186	0.044	0.2048 ±0.0018	-	0.1123 ±0.0017	-
λ_{Time} and $\mu_{constant}$	3	-238.427 ±0.624	483.151 ±1.249	3.518	0.101	0.1962 ±0.0016	-0.0407 ±0.0004	5.00E-08 ±5.00E-08	-
$\lambda_{constant}$ and μ_{Time}	3	-239.089 ±0.621	484.474 ±1.242	4.841	0.052	0.1914 ±0.0016	-	0.0549 ±0.0010	0.0433 ±0.0006
λ_{Time} and μ_{Time}	4	-238.427 ±0.624	485.354 ±1.249	5.721	0.034	0.1962 ±0.0016	-0.0405 ±0.00004	8.535E-05 ±3.24E-05	0.0347 ±0.0024
$\lambda_{Alti.}$ and $\mu_{constant}$	3	-240.247 ±0.572	486.790 ±1.144	7.157	0.016	0.0310 ±0.0012	0.0004 ±6.35E-06	0.0794 ±0.0021	-
$\lambda_{constant}$ and $\mu_{Alti.}$	3	-241.542 ±0.577	489.379 ±1.155	9.746	0.004	0.1991 ±0.0017	-	0.5116 ±0.0676	-0.00024 ±1.35E-05
$\lambda_{Alti.}$ and $\mu_{Alti.}$	4	-240.253 ±0.582	489.005 ±1.164	9.372	0.005	0.0454 ±0.0021	0.0003 ±9.30E-06	0.1016 ±0.0058	-2.67E-05 ±6.29E-05
$\lambda_{Temp.}$ and $\mu_{constant}$	3	-240.605 ±0.632	487.505 ±1.264	7.872	0.011	0.2121 ±0.0021	-0.0242 ±0.0012	0.0834 ±0.0011	-
$\lambda_{constant}$ and $\mu_{Temp.}$	3	-238.058 ±0.649	482.412 ±1.298	2.779	0.146	0.1817 ±0.0013	-	0.0071 ±0.0005	0.4707 ±0.0138
$\lambda_{Temp.}$ and $\mu_{Temp.}$	**4**	**-235.566 ±0.595**	**479.633 ±1.189**	**0**	**0.586**	**0.0839 ±0.0017**	**0.3811 ±0.0073**	**0.0797 ±0.0019**	**0.3988 ±0.0041**

Table 4. Results from the TESS analyses. a) Paleoenvironment-dependent diversification models. b) Episodic birth-death models. The model with postive dependence of speciation and extinction with temperature has the lowest AICc and highest AICω. We compared macroevolutionary scenarios in which speciation and extinction vary through time (Time), or according to paleo-elevation changes of the Himalaya and Tibetan Plateau (Alti.), according to global temperature changes (Temp.), and episodic birth-death models (Shift). Abbreviations: NP, number of parameters; logL, log-likelihood; AICc, corrected Akaike Information Criterion; ΔAICc, the difference in AICc between the model with the lowest AICc and the others; AICω, Akaike weight. Parameter estimates: λ_0 and μ_0, speciation and extinction for the environmental value at present; α and β, parameter controlling the variation of the speciation (α) and extinction (β) with the paleoenvironment. λ_1 and μ_1, speciation and extinction for the first period of time (present to first shift).

a)

Models	NP	logL	AICc	ΔAIC	AICω	λ_0	α	μ_0	β
Constant birth-death	2	-240.687	485.425	3.668	0.044	0.2037	-	0.1113	-
λ_{Time} and $\mu_{constant}$	3	-238.811	483.775	2.018	0.101	0.1951	-0.0402	2.01E-06	-
$\lambda_{constant}$ and μ_{Time}	3	-239.331	484.816	3.059	0.060	0.1898	-	0.0515	0.0467
λ_{Time} and μ_{Time}	4	-238.802	485.916	4.159	0.035	0.1951	-0.0398	0.0002	0.0330
$\lambda_{Alti.}$ and $\mu_{constant}$	3	-239.570	485.294	3.537	0.047	0.0483	0.0003	0.09450	-
$\lambda_{constant}$ and $\mu_{Alti.}$	3	-240.416	486.985	5.228	0.020	0.1999	-	0.3730	-0.00014
$\lambda_{Alti.}$ and $\mu_{Alti.}$	4	-239.378	487.067	5.310	0.019	0.0472	0.0004	0.1169	0.0002
$\lambda_{Temp.}$ and $\mu_{constant}$	3	-239.570	485.294	3.537	0.047	0.0483	0.0003	0.0945	-
$\lambda_{constant}$ and $\mu_{Temp.}$	3	-237.826	481.806	0.049	0.271	0.1800	-	0.0064	0.5007
$\lambda_{Temp.}$ and $\mu_{Temp.}$	**4**	**-236.723**	**481.757**	**0**	**0.278**	**0.1383**	**0.1463**	**0.0458**	**0.3483**

b)

Models	NP	logL	AICc	ΔAIC	AICω	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
Shift 1	5	-237.008	484.542	2.785	0.069	0.052	0.176	-	-	-	-	0.004	0.018	-	-	-	-
Shift 2	8	-236.086	489.706	7.949	0.005	0.195	0.084	0.180	-	-	-	0.112	0.097	0.004	-	-	-
Shift 3	11	-236.118	497.379	15.622	0.0001	0.180	0.127	0.155	0.181	-	-	0.111	0.051	0.159	0.002	-	-
Shift 4	14	-234.251	501.935	20.178	0	0.164	0.159	0.049	0.161	0.190	-	0.136	0.062	0.074	0.058	0.008	-

APPENDICES

Appendix 1. All sequence data used for this study (a file is generated per gene).

Appendix 2. The individual gene alignments as recovered by MAFFT. Results of the Bayesian phylogenetic analyses for each gene, and an explanation of the results.

Appendix 3. The total-evidence matrix (including molecular and morphological data) used for the phylogenetic placement of fossils with MrBayes.

Appendix 4. The BEAST files for the Bayesian dating analyses (the tree prior can be a Yule process or a birth-death model, and the dataset can include or not the morphological data).

Appendix 5. The current geographic species distribution data of all Parnassiinae as coded present (1) or absent (0) in all ten geographic areas (Western Palearctic, North Africa, Turkey, Central Asia, Himalaya, India, Mongolia, Siberia, China-Japan, and Western Nearctic).

Appendix 6. The time-stratified biogeographic model used for DEC analyses (time slices represent geological epochs or stages in the Cenozoic).

Appendix 7. Paleo-elevation for the Himalayan and Tibetan compiled from the literature.

Appendix 8. Description of the Bayesian episodic environment-dependent birth-death model.

Appendix 9. Results of PartitionFinder performed on the concatenated molecular dataset.

Appendix 10. Time-calibrated trees of Parnassiinae as estimated by BEAST following four different analyses.

Appendix 11. Results of the model comparison for the dating analyses based on marginal likelihood estimates and Bayes factors.

Appendix 12. Biogeographic history of Parnassiinae as estimated by DEC.

Appendix 13. Results from the diversity-dependence diversification analyses in DDD.

Appendix 14. Results of the MuSSE and GeoSSE analyses performed on 200 trees randomly taken from the Bayesian dating analysis. Models are ranked by AICc.

Appendix 15. Plot of the difference between speciation rates between all traits. When the difference overlaps zero (vertical red bar), the speciation rates are not significantly different.

Appendix 16. Robustness of the SSE models with simulation tests, HiSSE analyses and an implementation in RevBayes. For the simulation, the difference of fit between the best model and the reference model is shown with the red vertical line for real data, and in black for simulated data. HiSSE and RevBayes agree with the MuSSE models on host plants.

Appendix 17. Summary of diversification models in BAMM compared across a gradient of values for the Poison process governing the number of rate shifts.

Appendix 18. Credible set of configuration shifts inferred with BAMM and five different values of the Poison prior. It shows the distinct shift configurations with the highest posterior probability. For each shift configuration, the locations of rate shifts are shown as black circles, with circle size proportional to the marginal probability of the shift.

Appendix 19. Rate-through-time plot as inferred with RevBayes for Parnassiinae. Net diversification rates significantly changed and increased along the stem of the genus *Parnassius*, in agreement with the rates as estimated with BAMM.

Appendix 20. Rate-through-time plot as inferred with CoMET for Parnassiinae. The analyses detected one possible mass extinction around 15 Ma and one speciation rate shift around 3.5 Ma, in agreement with two TreePar analyses allowing or not the mass extinction.

Appendix 21. Credibility intervals of the correlation parameters for the Bayesian (RevBayes) environment-dependent diversification models.

Appendix 22. Correlation (linear regression) between speciation rates as inferred with MuSSE (a) and BAMM (b) and the species richness of host plants on which each parnassiine clade is feeding. In both cases, a strong and positive correlation is found.