bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TITLE:

Inferring the role of habitat dynamics in driving diversification: evidence for a species pump in Lake Tanganyika cichlids

AUTHORS: Thijs Janzen[1,2*], Rampal S. Etienne[2]

AUTHOR AFFILIATION

1. Max Planck Institute for Evolutionary Biology, Department of Evolutionary Theory, August-Thienemann-Straße 2, 24306, Plön, Germany.
2. University of Groningen, Groningen Institute for Evolutionary Life Sciences, Box 11103, 9700 CC Groningen, The Netherlands

*Corresponding author: Thijs Janzen, Max Planck Institute for Evolutionary Biology, Department of Evolutionary Theory, August-Thienemann-Straße 2, 24306, Plön, Germany, e-mail: thijsjanzen@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ABSTRACT

Geographic isolation that drives speciation is often assumed to slowly increase over time, for instance through the formation of rivers, the formation of mountains or the movement of tectonic plates. Cyclic changes in connectivity between areas may occur with the advancement and retraction of glaciers, with water level fluctuations in seas between islands or in lakes that have an uneven bathymetry. These habitat dynamics may act as a driver of allopatric speciation and propel local diversity. Here we present a parsimonious model of the interaction between cyclical (but not necessarily periodic) changes in the environment and speciation, and provide an ABC-SMC method to infer the rates of allopatric and sympatric speciation from a phylogenetic tree. We apply our approach to the posterior sample of an updated phylogeny of the *Lamprologini*, a tribe of cichlid fish from Lake Tanganyika where such cyclic changes in water level have occurred. We find that water level changes play a crucial role in driving diversity in Lake Tanganyika. We note that if we apply our analysis to the Most Credible Consensus tree, we do not find evidence for water level changes influencing diversity in the *Lamprologini*, suggesting that the MCC tree is a misleading representation of the true species tree. Furthermore, we note that the signature of habitat dynamics is found in the posterior sample despite the fact that this sample was constructed using a species tree prior that ignores habitat dynamics. However, in other cases this species tree prior might erase this signature. Hence we argue that in order to improve inference of the effect of habitat dynamics on biodiversity, phylogenetic reconstruction methods should include tree priors that explicitly take into account such dynamics.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Environmental changes such as the formation of mountain ridges, the formation of rivers and the movement of tectonic plates have long been known to be important drivers of speciation (Coyne and Orr 2004). Repeated environmental changes may thus lead to diversification patterns. For instance, the spatially explicit model of landscape fragmentation and hence splitting of geographic ranges of Pigot et al. (2010) explains a large part of the features of the reconstructed avian trees. Aguilée et al (2011, 2013) studied cyclic changes in the environment in particular and showed that they may present an alternative driver of diversity. Cyclic changes in the environment can cause populations to continuously switch between an allopatric and sympatric stage, providing a continuously renewed potential for speciation. These cyclic changes can in turn drive diversity towards levels unexpected given the current geography, and are sometimes referred to as a “species pump”. Examples of species pumps are environmental fluctuations fragmenting habitat on the slopes of mountains (Weir 2006; Sedano and Burns 2010; Hutter et al. 2013), glaciations and postglacial secondary contacts (Barnosky 2005), sea level changes causing the fusion and fragmentation of islands (Glor et al. 2004; Thorpe et al. 2008, but see Papadopoulou and Knowles 2015), and fluctuations in water level causing fragmentation and fusion of lakes with uneven bathymetry, as in the African Rift Lakes (Cohen et al. 1997b; Alin et al. 1999; Ivory et al. 2016).

The African Rift Lakes provide a good starting point in studying the interplay between cyclic habitat dynamics and speciation, because they have been subject to frequent water level changes (Cohen et al. 1997b; Alin and Cohen 2003; Ivory et al. 2016), which potentially influenced fish diversity in the lakes. An estimated number of 2000 cichlid fish species have

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

evolved in the African Rift Lakes over the past 10 million years (Turner et al. 2001), and comprise one of the most spectacular adaptive radiations (Seehausen 2006). The most prominent water level changes took place in Lake Tanganyika, where the water level has dropped over 600 meters on multiple occasions over the past million years, splitting the lake into multiple smaller lakes (Lezzar et al. 1996; Cohen et al. 1997a, 2007). Being the oldest lake of the three large rift lakes (Cohen et al. 1993), Lake Tanganyika contains the highest behavioral diversity (Konings 2007) and it is the only lake with a highly resolved phylogeny for cichlid fish. Micro-satellite analysis indicates that in several genera, populations are aligned with the location of the smaller lakes at low water level, suggesting the influence of water level changes on population segregation and speciation (Taylor et al. 2001; Koblmüller et al. 2011; Nevado et al. 2013).

Aguilée et al. (2013) developed a model for the African Rift Lakes in which populations at different locations diverge from each other depending on the local habitat, and at the same time allowed for sympatric speciation by implementing assortative mating that allows for a single branching point in trait values. Over time the different locations become separated or are reconnected, and this may drive the formation of new species. The authors conclude that stable numbers of diversity are best obtained by a fragmented habitat with recurrent merged states and rapid fluctuations. However, Aguilée et al. (2013) do not compare their results to empirical data.

Here, we provide a method to infer whether or how cyclic changes in the environment influence both the generation and the maintenance of biodiversity. We use an extension of the standard constant-rates birth-death model. Because deriving an expression for the likelihood of this model for a given set of phylogenetic branching times is difficult, but simulation of

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

phylogenies under the model is easy, we used approximate Bayesian computation (ABC) based on sequential Monte Carlo sampling (SMC) to estimate parameters from phylogenies. We applied our approach to an updated phylogeny of the *Lamprologini*, a tribe of cichlid fish from Lake Tanganyika in order to assess the importance of these habitat dynamics in shaping the current biodiversity of cichlids in Lake Tanganyika.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

METHODS

Model

To model the interaction between environmental change and speciation, we envisage a lake that consists of a single pocket at high water level, but that splits into two pockets when the water level drops. When the water level drops, we assume that all species distribute themselves equally over the two pockets; similarly, when the water level rises, all species previously contained in the two pockets are combined into the single pocket. Allopatric speciation can only occur when the water level is low. After an exponentially distributed waiting time, one of the two incipient species in either pocket can speciate into a new species. If this allopatric speciation does not occur before the water level rises again, i.e. reflecting that there has not been enough genetic divergence, the two incipient species in the two pockets merge back into one species. This is conceptually similar to the idea of protracted speciation (Etienne and Rosindell 2012): the water level drop initiates the speciation process whereas the allopatric speciation event is the completion of speciation under the protracted speciation model. Sympatric speciation can always occur in our model, either at high water level in the lake, or in both pockets when the water level is low. Extinction is considered to be a background process that occurs locally, i.e. within a pocket. If the water level is high, this causes extinction of a species, if the water level is low, this causes local extinction in one of the pockets.

We implemented our model using a Gillespie algorithm, where the time steps are chosen depending on the rate of possible events. In the model there are five possible events:

1) A water level change event, inducing incipient species or merger of incipient species.
2) Sympatric speciation event at high water level, with rate λ_s^h
3) Sympatric speciation event at low water level, with rate λ_s^l

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4) Allopatric speciation(-completion) event, at low water level, with rate λ_a^l

5) Extinction event, with rate μ

When the water level drops, all species distribute themselves over both pockets. Thus, immediately after a water level drop, the number of incipient species is equal to twice the number of species. When the water level rises, all incipient species that belong to the same species merge into a single species. During a sympatric speciation event, a single species splits into two new species, and the original (incipient) species is consumed in the process. If sympatric speciation occurs when the water level is low, the species in the other pocket is retained, and thus three new lineages arise: the first branching point occurs at the water level drop while the second occurs at the sympatric speciation event (Figure 1).

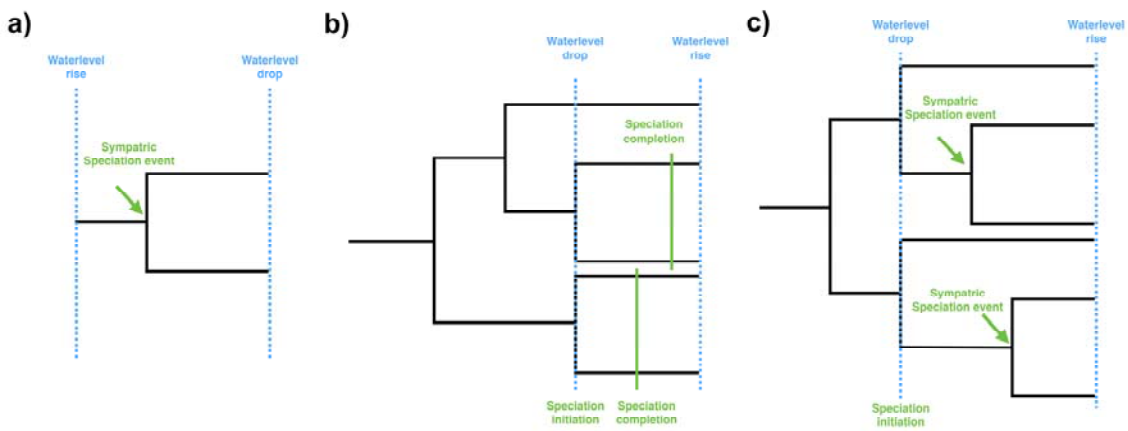

Figure 1. Schematic representation of the consequences of the three different types of speciation on phylogeny. Time proceeds from left to right. The dotted blue lines indicate water level changes. (a) During a sympatric speciation at high water level event, diversification is not aligned with any associated water level change. (b) During allopatric speciation at low water level, speciation initiation coincides with the water level drop, causing branching events (if speciation-completion occurs before water level rise) to line up in time. Branching events are independent of the time of speciation completion, hence, even when the actual speciation completion events occur at different time points, branching events in the species tree are identical. (c) During a sympatric speciation event at low water, the speciation event is independent of the water level changes. Because the original species is consumed in the process, a new branching event is also added at the water level change event. Hence, both speciation-completion (b) and sympatric speciation at low water level (c) cause branching times to line up at the time of water level drop.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

During an extinction event, one (possibly incipient) species is removed from the simulation. If the water level is low, this need not lead to the extinction of a species, because the sister incipient species might remain in the other pocket, ensuring survival of the species.

Maximum Likelihood

Without water level changes, our model reduces to the constant rates birth-death model (Nee et al. 1994). As a reference therefore, we estimated parameters of the standard birth-death model using Maximum Likelihood. The likelihood of the birth-death model was calculated using the function "bd_ML" from the R package DDD. (Etienne et al. 2012).

Fitting the model to empirical data

We performed two different fitting procedures: firstly, we performed a model selection procedure, where three different water level scenarios were fitted simultaneously to the data (more information about the chosen scenarios can be found in the next section). Hence, the model selection procedure simultaneously estimates parameter estimates and assesses the fit of the models (fraction of particles associated with each model). However, because all particles are often assigned only to the best fitting model, the model selection procedure does not always allow for the comparison of parameter estimates across different models. Therefore, we also fitted the three different water level scenarios independently to the empirical data, and obtained posterior distributions for the parameters relevant to these scenarios.

We fitted our model to 100 trees randomly sampled from the MCMC chain obtained from the *BEAST analysis (see below), and to the Most Credible Consensus (MCC) tree.

Water level scenarios

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The main focus of our approach is to assess the impact of water level changes on the diversification rate. Lake Tanganyika experienced low water level stands 35 - 40 kyears ago, 169 - 193 kya, 262 - 295 kya, 363 - 393 kya and 550 - 1100 kya (Lezzar et al. 1996; Cohen et al. 1997a). Consequently, high water levels occurred between 0 – 35 kya, 40 – 169 kya, 193 – 262 kya, 295 - 363 kya and 393 – 550 kya. Unfortunately the geological record does not reveal whether any low water level stands occurred beyond 1.1 million years ago. This leaves us with two alternative scenarios: either no low water level stands occurred beyond 1.1 million years ago, or these low water level stands have not been preserved accurately in the geological record.

In order to capture these two scenarios we performed inference using two alternative water level implementations. Firstly we used the exact literature values, assuming a high water level stand until 1.1 million years ago. We refer to this scenario as LW (Literature Waterlevels). Secondly we assumed that before 1.1. million years ago, water level changes occurred at the same *average* rate of water level change in the most recent 1.1 million years. In the recent 1.1 million years the lake experienced 5 high water level stands, and 5 low water level stands, which amounts to 10 water level changes in total. To extrapolate water level changes to more than 1.1 million years ago, we drew waiting times until the next water level change from an exponential distribution with rate 10. We refer to this scenario as EW (Extrapolated Waterlevels). Thirdly we also tested the null expectation: no water level changes at all, we refer to this scenario as NW (No Waterlevels). Without water level changes, the model reduces to the constant-rates birth death model.

Parameter estimation

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To fit the model to empirical data we used Approximate Bayesian Computation, in combination with a Sequential Monte Carlo scheme (ABC-SMC) (Toni et al. 2009). As summary statistics for the ABC analysis we chose the normalized LTT statistic (Janzen et al. 2015), tree size, Phylogenetic Diversity (Schweiger et al. 2008) and the γ statistic (Pybus and Harvey 2000). On all parameters (λ_s^h, λ_s^l, λ_a^l, μ) we chose uniform priors U(-3, 2), on a ^{10}log scale, such that the eventual prior distribution spans (10^{-3},10^{2}). A ^{10}log scale was chosen to avoid parameters to take on values below 0. The standard deviation of the normal distribution used to perturb the parameters was chosen to have a mean of 0, and a standard deviation of 0.05 (on the ^{10}log transformed parameter), and we updated one parameter each time (e.g. jumps were only made in one dimension, to avoid extremely low acceptance rates). The number of particles used was 10,000. To assess the fit of the model to the data we calculated the Euclidian distance between the summary statistic of the simulated data and the empirical data. To ensure that the differences in summary statistics were on the same scale, we normalized the differences. Differences were normalized by dividing each difference by the standard deviation of that summary statistic of 1,000,000 trees simulated using parameter values sampled from the prior.

Model selection

To identify which model best explains the data, we performed ABC model selection, as described in Toni et al. (2009; 2010). The main difference between standard ABC-SMC and ABC-SMC including model selection is that the latter adds one parameter, which keeps track of the model. As jumping kernel between models we assumed a 50% probability of staying at the same model, and a 25% probability of jumping to either other model. This reversible jump ABC-SMC model selection procedure results in a posterior distribution over the three

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

models, where the model with most support is selected most across all particles. We can calculate the Bayes factor, by taking the ratio of the number of particles assigned to the respective models (Toni et al. 2009). For example, the Bayes factor of LW/EW is the number of particles assigned to the model with literature water level changes divided by the number of particles assigned to the model with extrapolated water level changes. Because a model can receive zero particles, we set the Bayes factor for each model compared to the model with zero particles to the maximum support possible, which is the total number of particles: 10,000. We report the fraction of trees for which each respective model is selected, and we report the distribution of Bayes factors across all tested trees, as an indication of the support for each model.

Model selection validation

To assess whether our ABC-SMC method can accurately infer the correct model, we simulated 50 datasets for each model (NW, LW & EW), with parameter values drawn from the prior. From the accepted particles in the last iteration we then assessed the proportion of particles belonging to each model, and selected the most frequent model as the most likely model as inferred by the inference procedure. We report the selected models and the mean Bayes factors across the 50 replicates.

Measurement uncertainty

A phylogeny generated with a high rate of allopatric speciation and a high rate of water level changes tends to have multiple speciation events that are aligned in time (Figure 1, b). This is due to the fact that the onset of speciation is given by the time of water level change. Phylogenetic reconstruction methods such as BEAST currently do not allow for simultaneous

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

branching events. Hence, when fitting the model, trees are generated that are by definition dissimilar from the empirical tree constructed using BEAST, even if underlying events are close to the original events. To circumvent this we perturbed the branching time of each node in the simulated tree. In this way speciation events that were previously aligned in time now occur on slightly different time points, as in a tree from a BEAST analysis. We perturbed branching times by adding a random number drawn from a truncated normal distribution with mean 0, standard deviation σ, and as upper and lower limit the branching times of the parent and daughter lineages respectively. If there were no daughter lineages present, and the node gave rise to an extant species, the normal distribution was truncated to the present time. Nodes were perturbed from past to present (leaving the crown in place, to ensure a phylogenetic tree with the same age as the empirical tree). The standard deviation of the perturbation kernel was included as an extra parameter to be inferred, with a uniform prior on $(10^{-3},10^{2})$.

Empirical data

We fitted our model to the phylogenetic tree of the tribe of *Lamprologini*, the most diverse tribe within Lake Tanganyika, containing 78 species of cichlids (Day et al. 2007; Koblmüller et al. 2007; Sturmbauer et al. 2010). The *Lamprologini* are endemic to Lake Tanganyika, and all species are substrate brooders with shared paternal and maternal care. In contrast to the mouthbrooding species from the *Haplochromini*, the *Lamprologini* show little sexual dimorphism and dichromatism, which are well-known indicators for sexual selection (Kraaijeveld et al. 2011). We therefore expect that the *Lamprologini* is a good candidate for picking up signals from water level changes.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We reconstructed a new *Lamprologini* tree following the workflow of the most complete *Lamprologini* tree to date, which is a consensus tree based on the mitochrondrial ND2 gene (Sturmbauer et al. 2010), but we added three newly described species (*Lepidiolamprologus mimicus* (Schelly et al. 2007), *Neolamprologus timidus* (Kullander et al. 2014b) and *Chalinochromis cyanophleps* (Kullander et al. 2014a)). and 7 genes with partial coverage (4 mitochondrial genes: 18S (21 species), COI (9 species), ND2 (77 species) and cytochrome b (43 species) and 3 nuclear genes: 38A (21 species), RH1 (33 species) and ednrb1 (21 species)). Sequences were downloaded from GenBank using phyloGenerator (Pearse and Purvis 2013), and aligned with MAFFT using default settings (Katoh and Standley 2013). Genbank accession numbers for all used genes can be found in table S4. Substitution models for each gene were determined using PartitionFinder (Lanfear et al. 2012), yielding a shared substitution model for cytochrome b and COI (TN93+I+Γ), the HKY+Γ model for 18S and 38A, HKY+I for ednrb1, HKY+I+Γ for RH1 and TN93+I+Γ for ND2. As outgroup taxa we used the same species as used in Sturmbauer et al. 2010: 6 species from Lake Malawi, all members (4 species) of the *Eretmodini*, and 13 species from the C-lineage. For a full species list, see table S4.

In line with Sturmbauer et al. 2010 we applied 4 time calibrations: 0.57 - 1.0 million years for the Malawi clade, 5 - 6 million years for the *Lamprologini* clade, 5 - 6 million years for the C-lineage and 1.1 - 3.5 million years for the riverine *Lamprologini* species. All calibrations were implemented using a uniform prior distribution.

Using *BEAST (Heled and Drummond 2010) within the BEAST 2 package (Bouckaert et al. 2014), we inferred the time-calibrated species tree. We used a relaxed molecular clock with log-normally autocorrelated rates among branches. All priors were left at default settings, except for the tree prior, which was set to be a birth-death prior. The resulting MCMC chain contained ~1150M iterations. Convergence was assessed using ESS values (all > 100) and by

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

eye using AWTY (Nylander et al. 2008). We thinned the chain using the program LogCombiner (from the BEAST 2 suite) by selecting every 20,000th tree and removed the first 10% of the chain as burn-in. The resulting posterior sample consisted of 51800 trees.

Using TreeAnnotator (from the BEAST 2 suite) we constructed a Maximum Clade Credibility tree, storing the mean heights. Although this can sometimes lead to negative branch lengths, this was not the case for our *Lamprologini* tree.

We then pruned the tree to obtain the pure *Lamprologini* tree on which we fit our model. To prune the tree we removed all outgroup taxa and the riverine species using the function "drop.tip" from the ape package in R (Paradis et al. 2004).

Branching time uncertainty in the empirical tree

To account for uncertainty in the estimates of branching times in the *Lamprologini* tree we sampled 100 trees from the posterior distribution obtained by *BEAST. Sampling was performed at random, irrespective of the likelihood of the trees. Although ideally we would have liked to perform our analysis on the full posterior distribution, computational limitations forced us to only use a subset of the posterior distribution. In the Supplementary material we show that the distribution of summary statistics of the subset of 100 trees is similar to the distribution of the full collection of trees. The 100 sampled trees were, like the Maximum Clade Credibility tree, also pruned to remove the outgroup and riverine taxa and stored separately. For all 100 trees we performed both the ABC-SMC model selection algorithm and the ABC-SMC parameter estimation algorithm, to determine the impact of different branching times on the inferred water level model, and to determine whether the MCC tree is a good representation of the underlying variability.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Per tree we ran a single parameter estimation run, and 10 replicate model selection runs, each seeded with a different seed. We report the fraction of models selected across all 100 trees, where the model selected per tree, is the model being selected in the majority of replicates of that tree.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RESULTS

Lamprologini *phylogeny*

The onset of diversification within the *Lamprologini* is estimated to be around 5.45 MYA (95% HPD: [5, 5.92]), which is slightly earlier than previous estimates (5.28 MYA (Sturmbauer et al. 2010)). Species placement in the Maximum Clade Credibility tree is consistent with previous placement and/or taxonomic placement (figure 1). Placement of the *Altolamprologus* genus within the *Lamprologus callipterus* complex seems to re-iterate previously published evidence for introgressive hybridization between these species (Koblmüller et al. 2007). For the three species not previously included in the *Lamprologini* phylogeny *Lepidiolamprologus mimicus*, was placed within the other *Lepidiolamprologines*, *Chalinochromis cyanophleps* was placed as a sister species to *Julidochromis ornatus*, within the *Chalinochromis/Julidochromis* complex, and in agreement with previous analysis (Kullander et al. 2014b), *Neolamprologus timidus* is not placed as a sister species to *Neolamprologus furcifer*, but rather associates with *N. mondabu* and *N. falcicula*. In contrast to Gante et al. (2016), we place *N. olivaceous* outside the *Brichardi* complex. All other species in the *Brichardi* complex are inferred to be closely related, which is in agreement with the analysis by Gante et al., which is based on full genome sequences from only a small group of species, in contrast to the limited number of markers from a large number of species that we used.

As a reference we inferred speciation and extinction using the constant-rates birth-death model (Nee et al. 1994). Using Maximum Likelihood (the function bd_ML in the DDD package), we obtained an estimate of 1.004 for the speciation rate, and an estimate of 0.411 for the extinction rate. Estimates for the birth-death model obtained during reconstruction of

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the tree using BEAST could not be compared to our estimates here as these were obtained including the outgroups and riverine taxa.

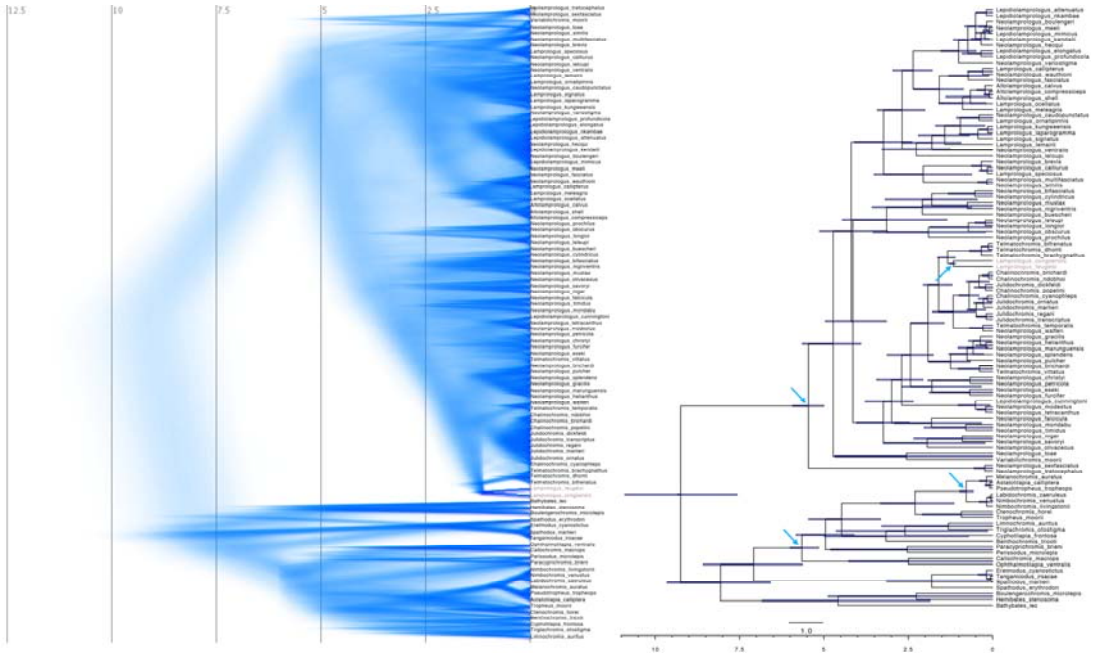

Figure 2. Phylogenetic hypothesis for the *Lamprologini* and outgroups, based on 3 mitochondrial and 4 nuclear genes, and four calibrations (indicated by arrows): 0.57 - 1.0 million years for the Malawi clade, 5 - 6 million years for the *Lamprologini* clade, 5-6 million years for the C-lineage and 1.1 - 3.5 million years for the riverine *Lamprologini* species. Left panel: Densitree representation of the MCMC chain obtained using *BEAST. Shown are trees from the thinned posterior chain, after selecting every 20,000th tree, after removing 10% of the chain as burn in. Riverine species are indicated in grey. Right panel: Maximum Clade Credibility tree. The tree was reconstructed using four calibrations (indicated by arrows): 0.57-1.0 million years for the Malawi clade, 5-6 million years for the *Lamprologini* clade, 5-6 million years for the C-lineage and 1.1-3.5 million years for the riverine *Lamprologini* species. Bars around the node span the 95% HPD for each node. Riverine species are indicated in grey. Please note that the order of species is different between the two panels.

Parameter estimation

Firstly, we estimated parameter values for the relevant models for all 100 trees sampled from the posterior. We report the average parameter values across the combined posterior across all 100 trees. The model without water level changes resembles the constant-rates birth-death model, and we find that estimates for sympatric speciation at high water level (λ_s^h) are

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

slightly higher than the estimated birth rate under the constant rates birth-death model (1.367 My^{-1} [0.599, 2.431] vs 1.001 respectively, see also Table 1). Similarly, we infer the extinction rate (μ) to also be slightly higher (0.796 My^{-1} [0.001, 2.013] vs 0.411, see also Table 1). Accurately inferring extinction rates is difficult (Nee and Holmes 1994), even for the constant-rates birth-death model, and typically diversification ($\lambda - \mu$) and turnover rates (μ/λ) are compared. We obtain estimates of 0.571 and 0.582 for diversification and turnover respectively, which are well in line with the estimates obtained using the constant rates birth-death model (a diversification rate of 0.59 and a turnover rate of 0.41 respectively). Taking into account the 95% confidence intervals on the obtained parameter estimates and the fact that the ABC-SMC estimates are potentially affected by the prior while the ML estimates are not, we are confident that the ABC-SMC method has inferred the correct parameter values for the model in the absence of water level changes.

Using the LW model, which implements water level changes following the literature (e.g. high water level until ~1.1 MYA, after which a series of water level changes took place), we infer a lower rate of sympatric speciation at high water level (0.571 My^{-1}), which is compensated with a high rate of allopatric speciation (13.326 My^{-1}) and sympatric speciation at low water level (2.894 My^{-1}), suggesting that water level dynamics are important drivers of biodiversity. Extinction is inferred to be low (0.050 My^{-1}).

Using the EW model, where water level changes are extrapolated to more than 1.1 MYA, we observe that the rate of sympatric speciation at high water level is inferred to be even higher than without water level changes (1.590 My^{-1} [0.842, 2.371]). Extinction, however, is lower (0.171 My^{-1} [0.001,0.724]), and allopatric speciation and sympatric speciation at low water level are both inferred to be lower than for the literature water scenario (4.948 My^{-1} and 0.069 My^{-1} respectively). This follows our expectations, given that with a higher number of water

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

level changes, the total time spent at low water level is higher, and too high rates of allopatric speciation and sympatric speciation at low water level would result in too much diversity.

Across all three water level models we observe that the distribution of the post-hoc perturbations σ differs considerably between the three scenarios: for the NW and EW model, especially values around $\sigma = 1$ are avoided, and only extremely large, or very small values occur. This seems to indicate that there is either a need for a very low degree of perturbation (very low values of σ), or for complete perturbation, which perturbates the node over the full available timescale (e.g. over the full distance between the parent and daughter node). For the LW model we observe a different pattern: predominantly high values of σ, which is possibly linked to high allopatric and sympatric speciation at low water rates, which generate branching times that are aligned in time. Correlating estimates for σ, λ_a^l and λ_s^l for the LW and EW models we generally recover extremely low R^2 values (all significant, but because of the high number of particles, significance is easily obtained), except for the LW model, where we recover an R^2 of 0.14 for the correlation of $\sigma \sim \lambda_a^l + \lambda_s^l$ (F-statistic: 8.279e4, 999904 df, p-value < 2e-16). This seems to indicate that there is an, albeit weak, relationship between a higher degree of perturbation and higher rates of allopatric and sympatric speciation at low water level.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Mean posterior density estimate, for sympatric speciation at high water (λ_s^h), extinction (μ), perturbation (σ), sympatric speciation at low water (λ_s^l) and allopatric speciation (λ_a^l). Shown are results for the model with no water level changes (NW), literature values for water level changes (LW) and waterlevel changes extrapolated beyond the literature range (EW). The 95% credibility interval is shown between square brackets. All values are rates per million years.

	λ_s^h	μ	σ	λ_a^l	λ_s^l
NW	1.367 [0.599, 2.431]	0.796 [0.001, 2.013]	7.049 [0.001, 44.235]		
LW	0.571 [0.293, 0.905]	0.050 [0.001, 0.218]	0.121 [0.001, 0.537]	13.326 [0.001, 65.364]	2.894 [0.001, 5.968]
EW	1.590 [0.842, 2.371]	0.171 [0.001, 0.724]	0.043 [0.001, 0.170]	4.948 [0.001, 35.398]	0.069 [0.001, 0.281]

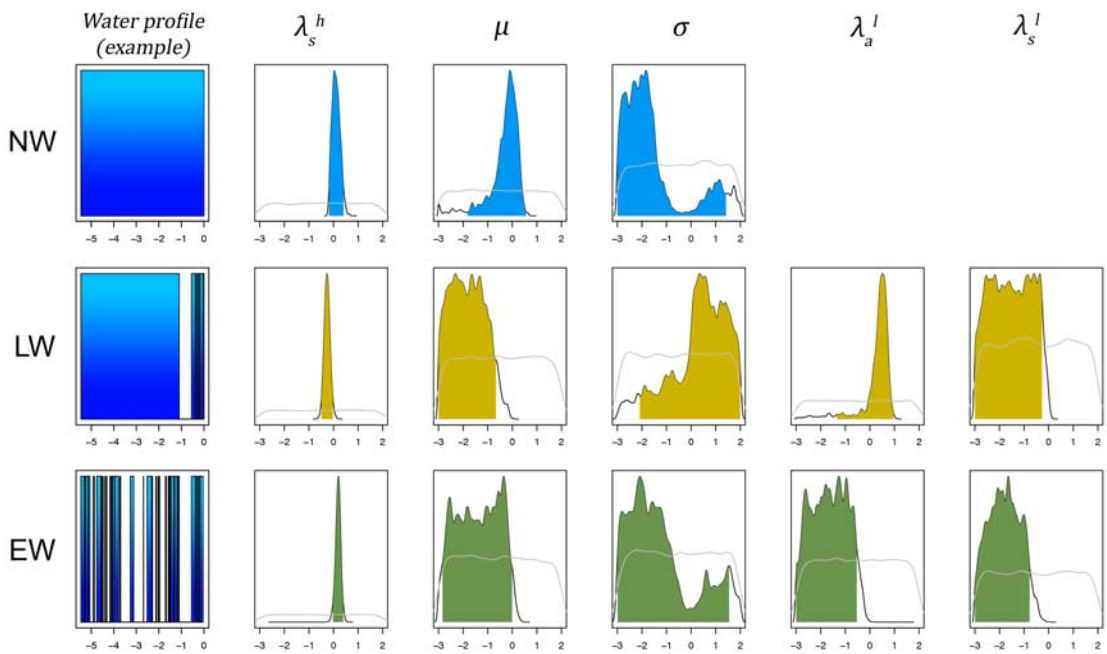

Figure 3. Posterior densities of the pooled posterior distribution across 100 randomly drawn trees from the posterior MCMC chain. Shown are estimates for the three water level scenarios (no water level changes (NW), literature values for water level changes (LW) and extrapolated values for water level changes (EW)). Shown are the posterior density (black line), 95% credibility interval (shaded area, blue for NW, gold for LW and green for EW), and the density of the prior (grey line). X-axes are on a 10-log scale. The first column shows a sample water level profile, with the water level on the y-axis, and the time before present (in million years) on the x-axis. Note that for the EW model, for each simulation a new profile was generated, and that the shown profile is only one example of such a profile. Because without water level changes, λ_s^l and λ_a^l have no meaning, their posterior distribution is not plotted for the NW scenario.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Model fitting

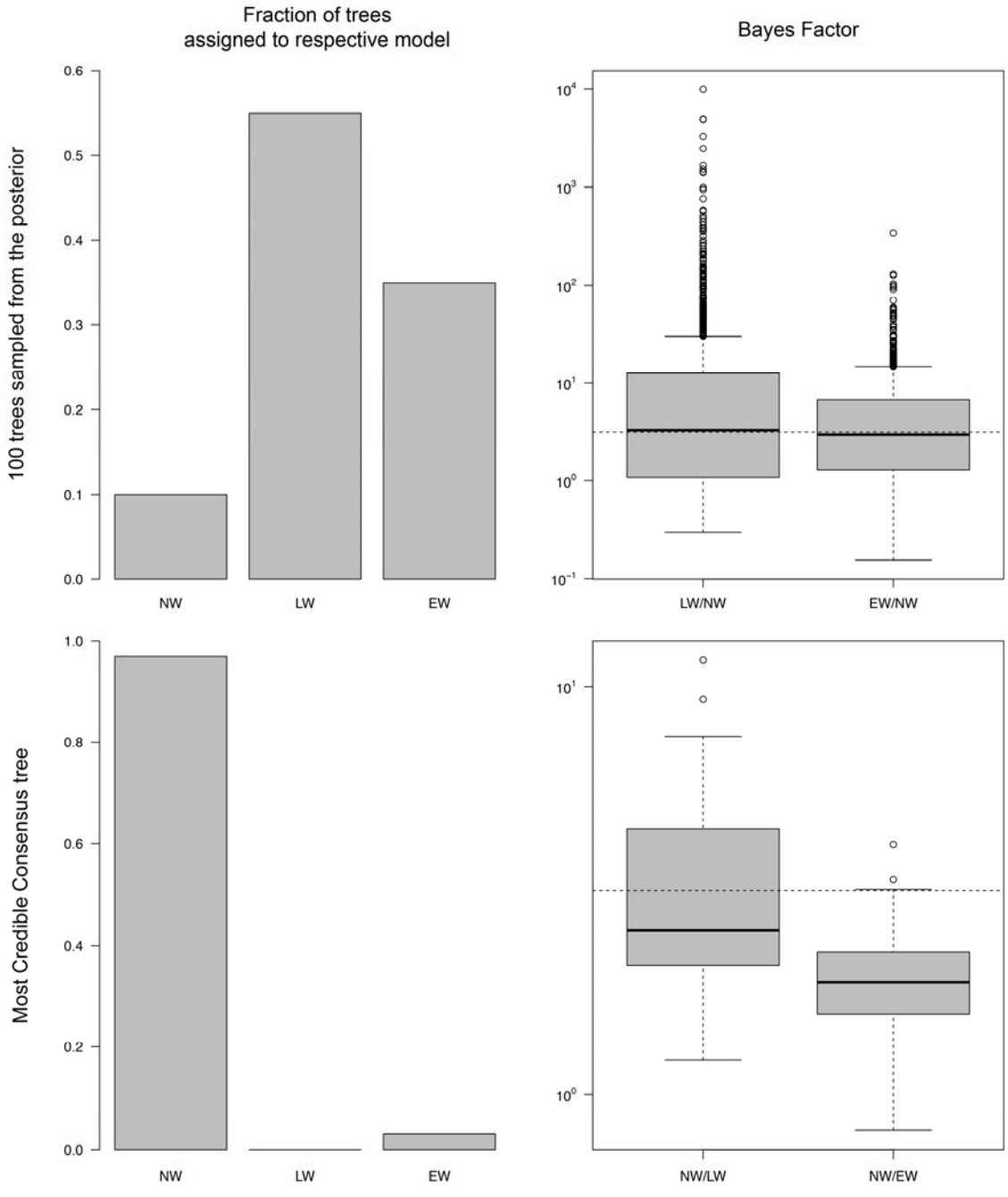

Figure 4. Model selection results on 100 trees drawn from the *BEAST posterior of the *Lamprologini* tree (top row), and on the *Lamprologini* MCC tree (bottom row). The left column shows the fraction of trees for which model NW, LW or EW was selected(NW: no water level changes. LW: literature water level changes, EW: extrapolated water level changes. The right column shows the Bayes factor for the LW relative to the NW model, the EW model relative to the NW model and the Bayes factor for the EW model relative to the

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LW model. The horizontal dotted line indicates a Bayes factor of $10^{0.5}$, above which support for the model in the numerator is considered to be high.

Model selection

When we apply the model selection algorithm to each selected tree (10 replicate model selection inference per tree) (Figure 4, top row), we find that in 10 out of 100 trees, the NW model receives most particles in most replicates. In 55 out of 100 trees, the LW model has most particles, and for the remaining 35 trees, the EW model is selected. The large support for the LW model is reflected in the Bayes factors as well, with a mean Bayes factor of 71.2 when comparing the LW model with the NW model, and a mean Bayes factor of 28.1 when comparing with the EW model. Using the Bayes factor interpretation of Jeffreys (1961), which postulates that any model with a Bayes factor higher than $10^{0.5} = 3.16$, has substantial support (and any model with a Bayes factor higher than $10^{1.0} = 10$ is considered to have strong support), we find that the mean Bayes factors for the LW model strongly support the LW model over the other two models. Median Bayes factors are lower, suggesting a strongly skewed distribution (see also Figure 4, top row, right panel), with median Bayes factors of 3.32 and 1.14 when comparing LW with NW and EW respectively. Support for the EW model over the NW model is high as well, with a mean Bayes factor of 16.22 (median 2.90). The low number of trees assigned to the NW model, and the high Bayes factors of the two other models when compared with the NW model together suggest that it is unlikely that branching patterns in the *Lamprologini* tree are caused by a model without water level changes.

When we apply the model selection algorithm to the MCC tree (100 independent runs) (Figure 4, bottom row), we find that in 97 out of 100 trees, the NW model receives most particles, and for the remaining 3 trees, the EW model is selected. Support for the NW model over the LW model is less overwhelming, with a mean Bayes factor of 3.36 (median 2.52).

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Support for the NW model over the EW model is even less, with a Bayes factor of 1.92 (median 1.87). These low Bayes factors suggest that within the replicates, the number of particles associated with the NW model was only marginally larger than the number of particles associated with the two other models. Nevertheless, we find that for the MCC tree, support for the LW and EW models is much lower than the support we found for trees sampled from the posterior.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Validation of the model selection procedure

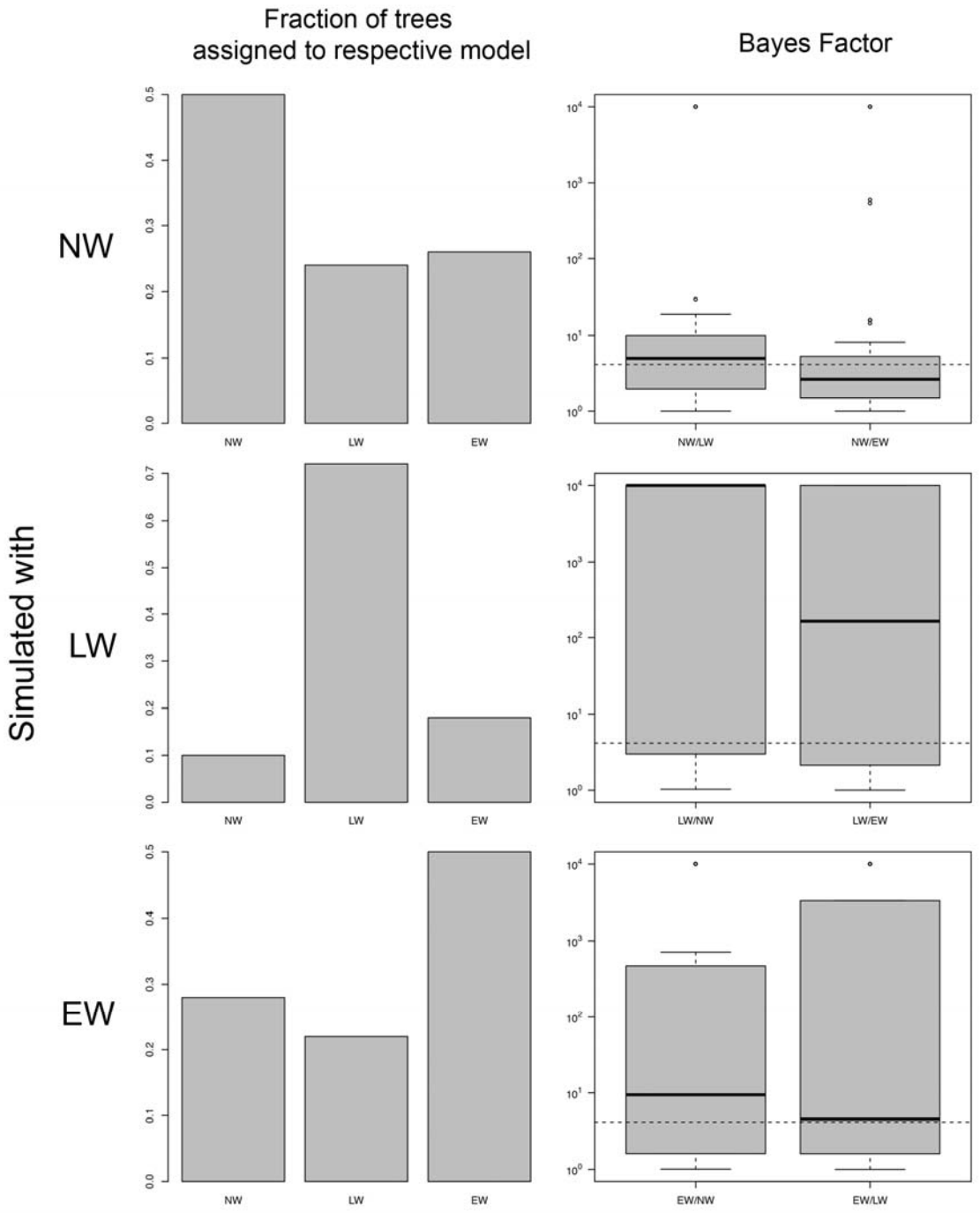

Figure 5. Validation of the ability of our ABC-SMC algorithm to infer the correct model. 50 replicate datasets were generated for each water level model (no water level changes NW, water level changes from the literature LW, or water level changes extrapolated beyond the literature range, EW). The left column shows the fraction of trees for which the respective model was selected. The right column shows the distribution of Bayes factors for the model used to simulate the data compared to the other two models. To accommodate Bayes factors

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of zero, shown is the transformation $\log_{10}(BF+1)$ The horizontal dotted line indicates a Bayes factor of $1+10^{0.5}$, above which support for the model in the numerator is considered to be high.

Model validation shows that when we simulated data using the NW model, the NW model was selected using our model validation algorithm more than the other two models (25 out of 50 replicates). The mean Bayes factors are both much higher than $10^{0.5}$, with a mean of 1024 and 841 versus the LW and EW model respectively (medians: 4.00 and 1.67). When data was simulated with the LW model, we selected the correct model in the majority of 50 replicates (36 out of 50 replicates). The Bayes factors reflect this, with a mean of 5648 and 4224 (medians: 10,000 and 163) versus the NW and EW model respectively. Lastly, when we simulated data using the EW model, we selected the correct model more than the other two models, in 25 out of 50 replicates. This was reflected by the Bayes factors as well, as the mean Bayes factor versus the NW model was 2323, and the mean Bayes factor versus the LW model was 2350 (medians: 8.43 and 3.55).

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DISCUSSION

We have presented a model that infers past speciation and extinction rates, and their interactions with changes in the environment, from a given phylogeny. We have shown that our model is able to accurately select between different scenarios, including or excluding environmental change. We applied our model to a phylogeny of the cichlid fish tribe of *Lamprologini* and found evidence that past water level changes have shaped current cichlid diversity in Lake Tanganyika, when we applied our model to a sample from the posterior distribution of trees of the *Lamprologini*, as inferred by *BEAST. We asked the model to select the best fitting of three scenarios: a scenario without any water level changes, a scenario using the values found in the literature, and a scenario using the mean rate of water level change found in the literature to extrapolate water level changes beyond the range of literature values available. We found that the model following literature water levels received most support, closely followed by the model extrapolating water level changes beyond the range of literature values. These results together indicate convincingly that the nullhypothesis, no effect of water level changes, should be rejected.

When we applied our model selection algorithm on the Most Credible Consensus (MCC) tree, we found highly contrasting results. Support for both models including water level changes diminished, and posterior support for the model without any water level changes increased dramatically. We conclude therefore that the MCC tree, at least for the *Lamprologini*, but probably more generally, provides a poor summary of the true species tree and of the underlying variation in branching patterns. Hence, we suggest to avoid reporting MCC trees, and instead to provide the reader with the full posterior distribution, for instance through a DensiTree plot. Posterior inference, for instance of speciation and extinction rates

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

should preferentially also be performed on multiple independent samples from the posterior, rather than on the MCC tree, as the underlying variation might lead to very different results, as we have shown.

When allopatric speciation rates are high, the resulting phylogenetic trees have internal nodes that have synchronized branching times, e.g. branching times that align with episodes of water level change. Phylogenetic reconstruction software does not allow for the alignment of branching times, and prior models used in phylogenetic reconstruction software explicitly exclude this. Our finding of evidence for a substantial role of habitat dynamics in diversification can therefore be regarded as conservative.

To improve the fit of trees generated by our model with trees generated by *BEAST we included an *a posteriori* perturbation parameter in our model. This parameter determines the standard deviation of a Gaussian perturbation kernel that is applied to each node after the simulation has completed. By perturbing each node, we ensured that branching times no longer align in time. A less *ad hoc* solution to deal with the alignment of branching times in the tree would be to incorporate the model presented here as a tree prior in phylogenetic reconstruction software. Although this need not introduce any significant differences in the tree topology, the distribution of branching times could be substantially influenced, and any subsequent inference focusing on such patterns could be very different. Including such models in tree reconstruction software may require incorporation of ABC methods, and will be extremely computationally demanding, but our results justify such an endeavor.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Although we refer in our model to the different implementations of speciation as sympatric and allopatric, care should be taken in interpreting these forms of speciation. We consider here allopatric speciation only on a large scale, where populations become allopatric over stretches of hundreds of kilometers (Sturmbauer et al. 2001). Large-scale isolation might not be necessary for cichlids, as some species can already be limited in gene flow by a sand stretch of 50 meters separating populations (Rico and Turner 2002). Such micro-allopatric speciation events are not captured by the allopatric speciation rate in our model. Rather, these local scale events are captured in our model by sympatric speciation. Hence, sympatric speciation in our model covers all degrees of speciation ranging from full sympatry to allopatry, providing that geographical isolation is smaller than that imposed by a water level change. Allopatric speciation in our model then solely refers to speciation events caused by geographical isolation over a large distance, driven by changes in water level, and inducing simultaneous branching events.

Our results are strongly in line with population genomic analyses in a number of cichlid species including *Eretmodus cyanostictus* (Verheyen et al. 1996), *Tropheus moorii* (Koblmüller et al. 2011; Nevado et al. 2013) and *Variabilichromis moorii* (Nevado et al. 2013), and resonate with population genomic findings across the three African Rift Lakes (Sturmbauer et al. 2001). Furthermore, population genetic studies have shown that water level fluctuations in Lake Malawi have been associated with population expansion in cichlid species (Arnegard et al. 1999; Sturmbauer et al. 2001; Genner et al. 2010), suggesting a potential role for water level changes in Lake Malawi as well. Considering that the geological record of Lake Malawi spans a much larger part of the total lifespan of the lake (Delvaux 1995; Lyons et al. 2015; Ivory et al. 2016), and we thus have much better information about water level fluctuations since the colonization of the lake by cichlids, we expect that modern

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

genetic developments will soon allow for a thorough understanding of the impact of water level changes on cichlids in Lake Malawi as well.

Our results potentially clarify previously recovered inconsistencies when studying diversification, for example in shrews in the Philippines. The Philippines have been subject to strong sea level fluctuations, causing the fission and fusion of several islands, primarily during the Pleistocene (Brown et al. 2013). Population genetic evidence has convincingly shown that the location of such fused islands correlates strongly with genetic divergence between populations in many different species (Evans et al. 2003; Linkem et al. 2010; Siler et al. 2010; Oaks et al. 2013). Phylogenetic analysis however, has failed to show any evidence of diversification associated with Pleistocene water level changes (Esselstyn and Brown 2009). The basis for this phylogenetic analysis however, was an MCC tree, for which we could also not find convincing evidence of water level changes driving diversification in cichlids. Repeating the analysis on the posterior distribution underlying the MCC tree could mitigate these problems, and could clarify the impact of Pleistocene water level changes on diversification in the Philippines archipelago.

Conclusion

Our model integrates standard constant-rate birth-death mechanics with environmental change and with speciation induced by geographical isolation. We analyzed the phylogeny of the tribe of *Lamprologini* to see whether past water level changes in Lake Tanganyika have contributed to the current diversity of cichlid fish in Lake Tanganyika. We find an important role for environmental changes in driving diversity, and find strong evidence that past water level changes have shaped current standing diversity in the triby of *Lamprologini*. However, we found that inference of past environmental changes from a single phylogeny, and more

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

specifically, from the MCC tree, tends to lead to unreliable results. We therefore advocate caution when using the MCC tree as a basis for further analysis. Furthermore, we argue for the inclusion of more detailed branching models in phylogenetic reconstruction software, which allow for the inclusion of an interaction between the environment, and speciation rates.

Acknowledgements

We would like to thank Lucas Molleman for useful discussions. We thank the Netherlands Organisation for Scientific Research (NWO) for financial support through VIDI and VICI grants awarded to RSE. We thank the Donald Smits Center for information Technology of the University of Groningen for their support and providing access to the Millipede high-performance computing cluster. We thank the Max Planck Institute for Evolutionary Theory for their support and providing access to their high-performance computing cluster.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

Aguilée R., Claessen D., Lambert A. 2013. Adaptive radiation driven by the interplay of eco-evolutionary and landscape dynamics. Evolution (N. Y). 67:1291–1306.

Aguilée R., Lambert A., Claessen D. 2011. Ecological speciation in dynamic landscapes. J. Evol. Biol. 24:2663–77.

Alin S., Cohen A., Bills R. 1999. Effects of landscape disturbance on animal communities in Lake Tanganyika, East Africa. Conservation. 13:1017–1033.

Alin S., Cohen A.S. 2003. Lake-level history of Lake Tanganyika, East Africa, for the past 2500 years based on ostracode-inferred water-depth reconstruction. Palaeogeogr. Palaeoclimatol. Palaeoecol. 199:31–49.

Arnegard M.E., Markert J.A., Danley P.D., Stauffer J.R., Ambali A.J., Kocher T.D. 1999. Population structure and colour variation of the cichlid fishes Labeotropheus fuelleborni Ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. Proc. R. Soc. Lond. B. 266:119–130.

Barnosky A. 2005. Effects of Quaternary climatic change on speciation in mammals. J. Mamm. Evol. 12:247–264.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 10:1–6.

Brown R.M., Siler C.D., Oliveros C.H., Esselstyn J. a., Diesmos A.C., Hosner P. a., Linkem C.W., Barley A.J., Oaks J.R., Sanguila M.B., Welton L.J., Blackburn D.C., Moyle R.G., Townsend Peterson a., Alcala A.C. 2013. Evolutionary processes of diversification in a model island archipelago. Annu. Rev. Ecol. Evol. Syst. 44:411–435.

Cohen A.S., Lezzar K.E., Tiercelin J.J., Soreghan M. 1997a. New palaeogeographic and lake-level reconstructions of Lake Tanganyika: implications for tectonic, climatic and biological evolution in a rift lake. Basin Res. 9:107–132.

Cohen A.S., Soreghan M., Scholz C.A. 1993. Estimating the age of formation of lakes: an example from Lake Tanganyika, East African Rift system. Geology. 21:511.

Cohen A.S., Stone J.R., Beuning K.R.M., Park L.E., Reinthal P.N., Dettman D., Scholz C.A., Johnson T.C., King J.W., Talbot M.R., Brown E.T., Ivory S.J. 2007. Ecological consequences of early Late Pleistocene megadroughts in tropical Africa. Proc. Natl. Acad. Sci. 104:16422–7.

Cohen A.S., Talbot M.R., Awramik S.M., Dettman D.L., Abell P. 1997b. Lake level and paleoenvironmental history of Lake Tanganyika, Africa, as inferred from late Holocene and modern stromatolites. Geol. Soc. Am. Bull. 109:444–460.

Coyne J., Orr H. 2004. Speciation. Sunderland, Massachusets U.S.A.: Sinauer Associates.

Day J.J., Santini S., Garcia-Moreno J. 2007. Phylogenetic relationships of the Lake Tanganyika cichlid tribe Lamprologini: the story from mitochondrial DNA. Mol. Phylogenet. Evol. 45:629–42.

Delvaux D. 1995. Age of Lake Malawi (Nyasa) and water level fluctuations. Mus. roy. Afrr. centr., Tervuren (Belg.), Dept. Geol. Min., Rapp. ann. 1993 1994. 108:99–108.

Esselstyn J.A., Brown R.M. 2009. The role of repeated sea-level fluctuations in the generation of shrew (Soricidae: Crocidura) diversity in the Philippine Archipelago. Mol. Phylogenet. Evol. 53:171–181.

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. R. Soc. B Biol. Sci. 279:1300–1309.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Etienne R.S., Rosindell J. 2012. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. Syst. Biol. 61:204–13.

Evans B.J., Brown R.M., McGuire J.A., Supriatna J., Andayani N., Diesmos A., Iskandar D., Melnick D.J., Cannatella D.C. 2003. Phylogenetics of fanged frogs: testing biogeographical hypotheses at the interface of the asian and Australian faunal zones. Syst. Biol. 52:794–819.

Gante H.F., Matschiner M., Malmstrøm M., Jakobsen K.S., Jentoft S., Salzburger W. 2016. Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. Mol. Ecol.

Genner M.J., Knight M.E., Haesler M.P., Turner G.F. 2010. Establishment and expansion of Lake Malawi rock fish populations after a dramatic Late Pleistocene lake level rise. Mol. Ecol. 19:170–82.

Glor R.E., Gifford M.E., Larson A., Losos J.B., Schettino L.R., Chamizo Lara A.R., Jackman T.R. 2004. Partial island submergence and speciation in an adaptive radiation: a multilocus analysis of the Cuban green anoles. Proc. R. Soc. B Biol. Sci. 271:2257–65.

Heled J., Drummond A.J. 2010. Bayesian Inference of Species Trees from Multilocus Data. Mol. Biol. Evol. 27:570–580.

Hutter C.R., Guayasamin J.M., Wiens J.J. 2013. Explaining Andean megadiversity: The evolutionary and ecological causes of glassfrog elevational richness patterns. Ecol. Lett. 16:1135–1144.

Ivory S.J., Blome M.W., King J.W., McGlue M.M., Cole J.E., Cohen A.S. 2016. Environmental change explains cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. Proc. Natl. Acad. Sci.:201611028.

Janzen T., Höhna S., Etienne R.S. 2015. Approximate Bayesian Computation of diversification rates from molecular phylogenies: introducing a new efficient summary statistic, the nLTT. Methods Ecol. Evol. 6:566–575.

Jeffreys H. 1961. The theory of probability. Oxford University Press.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Koblmüller S., Duftner N., Sefc K.M., Aibara M., Stipacek M., Blanc M., Egger B., Sturmbauer C. 2007. Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake Tanganyika--the result of repeated introgressive hybridization. BMC Evol. Biol. 7:7.

Koblmüller S., Salzburger W., Obermüller B., Eigner E., Sturmbauer C., Sefc K.M. 2011. Separated by sand, fused by dropping water: habitat barriers and fluctuating water levels steer the evolution of rock-dwelling cichlid populations in Lake Tanganyika. Mol. Ecol. 20:2272–90.

Kraaijeveld K., Kraaijeveld-Smit F.J.L., Maan M.E. 2011. Sexual selection and speciation: the comparative evidence revisited. Biol. Rev. Camb. Philos. Soc. 86:367–77.

Kullander S.O., Karlsson M., Norén M. 2014a. Chalinochromis cyanophleps, a new species of cichlid fish (Teleostei: Cichlidae) from Lake Tanganyika. Zootaxa. 3790:425–438.

Kullander S.O., Norén M., Karlsson M., Karlsson M. 2014b. Description of Neolamprologus timidus, new species, and review of N . furcifer from Lake Tanganyika (Teleostei: Cichlidae). Ichthyol.Explor.Freshwaters. 24:301–328.

Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.

Lezzar K.E., Tiercelin J.J., Batist M., Cohen A.S., Bandora T., Rensbergen P., Turdu C., Mifundu W., Klerkx J. 1996. New seismic stratigraphy and Late Tertiary history of the North Tanganyika Basin, East African Rift system, deduced from multichannel and high-resolution reflection seismic data and piston core evidence. Basin Res. 8:1–28.

Linkem C.W., Hesed K.M., Diesmos A.C., Brown R.M. 2010. Species boundaries and cryptic lineage diversity in a Philippine forest skink complex (Reptilia; Squamata; Scincidae: Lygosominae).

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mol. Phylogenet. Evol. 56:572–585.

Lyons R.P., Scholz C.A., Cohen A.S., King J.W., Brown E.T., Ivory S.J., Johnson T.C., Deino A.L., Reinthal P.N., Mcglue M.M., Blome M.W. 2015. Continuous 1.3-million-year record of East African hydroclimate, and implications for patterns of evolution and biodiversity. PNAS.:2–7.

Nee S., Holmes E. 1994. Extinction rates can be estimated from molecular phylogenies. Philos. Trans. Biol. Sci. 344:77–82.

Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 344:305–11.

Nevado B., Mautner S., Sturmbauer C., Verheyen E. 2013. Water-level fluctuations and metapopulation dynamics as drivers of genetic diversity in populations of three Tanganyikan cichlid fish species. Mol. Ecol. 22:3933–48.

Nylander J.A.A., Wilgenbusch J.C., Warren D.L., Swofford D.L. 2008. AWTY (are we there yet?): A system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics. 24:581–583.

Oaks J.R., Sukumaran J., Esselstyn J. a, Linkem C.W., Siler C.D., Holder M.T., Brown R.M. 2013. Evidence for climate-driven diversification? A caution for interpreting ABC inferences of simultaneous historical events. Evolution (N. Y). 67:991–1010.

Papadopoulou A., Knowles L.L. 2015. Genomic tests of the species-pump hypothesis: Recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. Evolution (N. Y). 69:1501–1517.

Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 20:289–290.

Pearse W.D., Purvis A. 2013. phyloGenerator: An automated phylogeny generation tool for ecologists. Methods Ecol. Evol. 4:692–698.

Pigot A.L., Phillimore A.B., Owens I.P.F., Orme C.D.L. 2010. The shape and temporal dynamics of phylogenetic trees arising from geographic speciation. Syst. Biol. 59:660–73.

Pybus O., Harvey P. 2000. Testing macro–evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. B Biol. Sci. 267:2267–72.

Rico C., Turner G.F. 2002. Extreme microallopatric divergence in a cichlid species from Lake Malawi. Mol. Ecol. 11:1585–90.

Schelly R., Takahashi T., Bills R., Hori M. 2007. The first case of aggressive mimicry among lamprologines in a new species of Lepidiolamprologus (Perciformes: Cichlidae) from Lake Tanganyika. Zootaxa. 49:39–49.

Schweiger O., Klotz S., Durka W., Kühn I. 2008. A comparative test of phylogenetic diversity indices. Oecologia. 157:485–95.

Sedano R.E., Burns K.J. 2010. Are the Northern Andes a species pump for Neotropical birds? Phylogenetics and biogeography of a clade of Neotropical tanagers (Aves: Thraupini). J. Biogeogr. 37:325–343.

Seehausen O. 2006. African cichlid fish: a model system in adaptive radiation research. Proc. R. Soc. B Biol. Sci. 273:1987–98.

Siler C.D., Oaks J.R., Esselstyn J.A., Diesmos A.C., Brown R.M. 2010. Phylogeny and biogeography of Philippine bent-toed geckos (Gekkonidae: Cyrtodactylus) contradict a prevailing model of Pleistocene diversification. Mol. Phylogenet. Evol. 55:699–710.

Sturmbauer C., Baric S., Salzburger W., Rüber L., Verheyen E. 2001. Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. Mol. Biol. Evol. 18:144–54.

Sturmbauer C., Salzburger W., Duftner N., Schelly R., Koblmüller S. 2010. Evolutionary history of the Lake Tanganyika cichlid tribe Lamprologini (Teleostei: Perciformes) derived from

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mitochondrial and nuclear DNA data. Mol. Phylogenet. Evol. 57:266–84.

Taylor M.I., Rüber L., Verheyen E. 2001. Microsatellites reveal high levels of population substructuring in the species-poor Eretmodine cichlid lineage from Lake Tanganyika. Proc. R. Soc. B Biol. Sci. 268:803–8.

Thorpe R.S., Surget-Groba Y., Johansson H. 2008. The relative importance of ecology and geographic isolation for speciation in anoles. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 363:3071–81.

Toni T., Stumpf M.P.H. 2010. Simulation-based model selection for dynamical systems in systems and population biology. Bioinformatics. 26:104–110.

Toni T., Welch D., Strelkowa N., Ipsen A., Stumpf M.P.H. 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface. 6:187–202.

Turner G.F., Seehausen O., Knight M.E., Allender C.J., Robinson R. 2001. How many species of cichlid fishes are there in African lakes? Mol. Ecol. 10:793–806.

Verheyen E., Rüber L., Snoeks J., Meyer A. 1996. Mitochondrial phylogeography of rock-dwelling cichlid fishes reveals evolutionary influence of historical lake level fluctuations of Lake Tanganyika, Africa. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 351:797–805.

Weir J.T. 2006. Divergent Timing and Patterns of Species Accumulation in Lowland and Highland Neotropical Birds. Evolution (N. Y). 60:842–855.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Information

Summary statistics

To show that the 100 trees sampled from the full posterior represent a reasonable subset of the full dataset, we compared the distribution of summary statistics for both datasets (e.g., either all 51858 trees, or only 100 trees). For the ABC-SMC algorithm we made use of 4 different summary statistics: number of tips, mean branch length (Phylogenetic Diversity (Schweiger et al. 2008)), the γ statistic (Pybus and Harvey 2000), and the normalized LTT statistic (Janzen et al. 2015). Because all trees have, by design, the same number of tips, and because the nLTT statistic only has meaning when compared to a specific tree, we only show results here for the mean branch length statistic, and the gamma statistic. Figure S1 shows that although the subset of 100 trees does not explore the full extremities of the full posterior distribution, the median and variance are highly similar between the two distributions. Furthermore, we find that summary statistics of the MCC tree (dotted line in Figure S1), are similar to the median for the mean branch length statistic, but are highly different for the gamma statistic.

bioRxiv preprint doi: https://doi.org/10.1101/085431; this version posted November 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

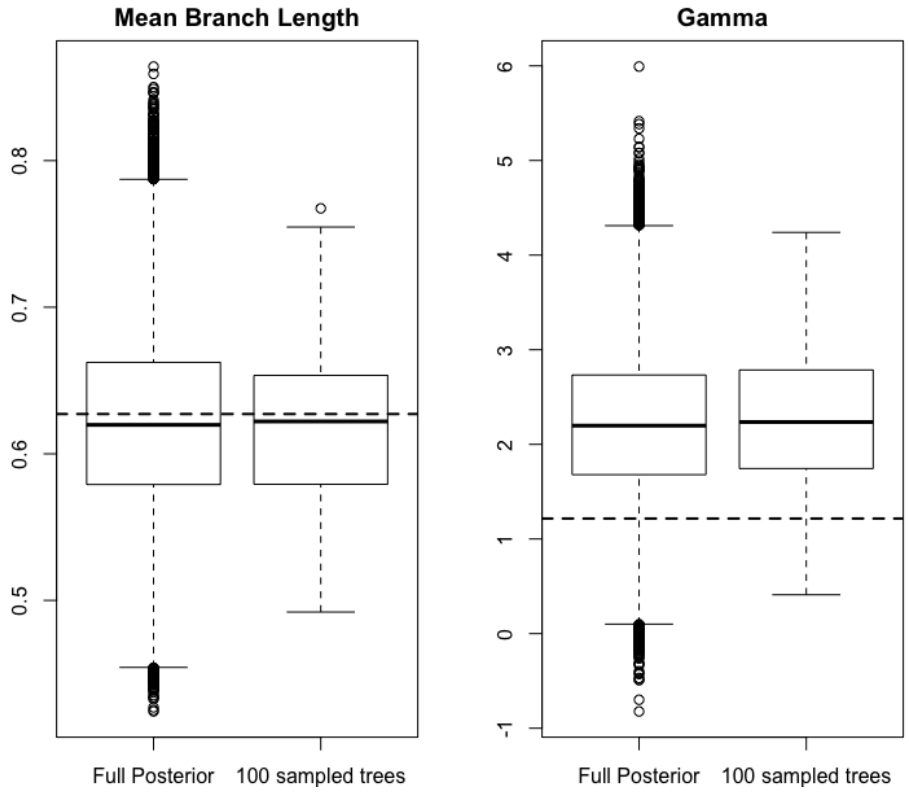

Figure S1. Comparison of the distribution of summary statistics for all trees from the *BEAST posterior ("Full Posterior"), and for 100 trees randomly sampled from the posterior ("100 sampled trees"). Shown are results for the mean branch length statistic (left panel), and the gamma statistic (right panel). The dotted line shows the summary statistic for the Most Credible Consensus tree.