How Important is Budding Speciation for Comparative Studies?

Daniel S. Caetano[1,2] and Tiago Bosisio Quental[2]

[1] *Department of Biological Sciences, Towson University, Towson, MD, 21252, USA*

[2] *Department of Ecology, University of São Paulo, São Paulo, SP, 05508-090, Brazil*

contact: dcaetano@towson.edu

Abstract: The acknowledgment of evolutionary dependence among species has changed how we ask biological questions in a fundamental way. Phylogenetic models became the standard approach for studies with three or more lineages, in particular those using extant species. Most phylogenetic comparative methods (PCMs) translate relatedness into covariance, meaning that evolutionary changes before lineages split should be interpreted together whereas after the split lineages are expected to change independently. This clever realization has shaped decades of research. Here we investigate one element of the comparative method often ignored or assumed as unimportant: if nodes of a phylogeny represent the dissolution of the ancestral lineage into two new ones or if the ancestral lineage can survive speciation events (i.e., budding). Budding speciation is reported in paleontological studies, due to the nature of the evidence for budding in the fossil record, but it is surprisingly absent in comparative methods. Here we show that many PCMs assume that divergence happens as a symmetric split, even if these methods don't explicitly mention this assumption. We discuss the properties of trait evolution models for continuous and discrete traits and their adequacy under a scenario of budding speciation. We test the effects of budding speciation under a series of plausible evolutionary scenarios and show when and how these can influence our estimates. We also propose that long-lived lineages that have survived through a series of budding speciation events and given birth to multiple new lineages (here called "progenitor lineages") can produce evolutionary patterns that challenge our intuition about the most parsimonious history of trait changes in a clade. We hope our discussion can help bridge comparative approaches in paleontology and neontology as well as foster awareness about the assumptions we make when we use phylogenetic trees.

Keywords: Diversification; evolutionary lineages; fossil record; model adequacy; phylogenetic comparative methods; speciation mode; trait evolution

Phylogenetic trees are the main representation of evolutionary relationships among lineages and stand as a symbol of evolutionary thought. However, their relatively simple structure is not capable of informing us about all aspects of evolution. A fully-resolved phylogeny is composed of nested bifurcations, each internal node representing ancestral-descendant relationships as well as divergence times. However, processes such as horizontal gene transfer, hybridization, and introgression can produce complex evolutionary relationships that challenge the explanatory power of bifurcating phylogenies (Philippe and Douady, 2003; Mallet et al., 2016; Bastide et al., 2018). Interestingly, the graphical representation of bifurcating trees could even influence how we think (Baum et al., 2005). A node connecting one ancestral branch to two new ones can suggest evolutionary histories much simpler than what we observe in nature and potentially downplay important aspects of macroevolution. For instance, if read literally, bifurcating phylogenies can be seen as speciation events that happened due to the split of an ancestral lineage into two new ones, coincident with the extinction, or dissolution, of the ancestral lineage (see discussion in Meier and Willmann, 2000). This view makes the concept of a lineage synonym to a branch of a phylogenetic tree. However, evidence from empirical systems shows the mode of speciation can be varied and often complex (Rosenblum et al., 2012), including frequent instances of budding speciation (Funk and Omland, 2003; Gottlieb, 2004; Crawford, 2010; Otero et al., 2022; Patsis et al., 2021) in which the ancestral lineage survives the speciation event and co-occurs (Anacker and Strauss, 2014) with the newly formed species.

Divergent views on the definition of lineages with respect to cladogenetic events can influence how we read phylogenetic trees. Hennig (1950; also see Willman, 1986) defines a lineage as an internodal branch, meaning that nodes of a phylogeny represent a dissolution of the ancestor lineage into two new ones—a literal view of a bifurcating tree. In contrast, Simpson's evolutionary lineage concept recognizes the continuation of a lineage through events of speciation (Simpson 1961; also see Meier and Willmann, 2000), a scenario commonly referred to as "budding speciation" (e.g., Raup, 1985; Foote, 1996; Gottlieb, 2004). Through the lenses of Simpson's evolutionary lineages, we can imagine lineages spanning two or more connected branches on a phylogenetic tree, each of those with their own evolutionary trajectory. Many recent comparative models overcame the assumption that each branch of the tree is an independently evolving lineage. There is a current trend of approaches that allow for more

realistic hypotheses of macroevolution, such as the role of species competition (Nuismer and Harmon, 2015), lineage-age dependent processes (Hagen et al., 2015; Hagen et al., 2018), and a dynamic adaptive landscape (Uyeda and Harmon, 2014). Each of these introduced biological processes in which the identity of the lineage is an important factor to predict the dynamics of trait evolution (i.e., lineages are not all equal for the model). It is worth emphasizing that if phylogenies are correctly "read" to represent nested groups of lineages that share a common ancestral, then the distinction between budding and bifurcation is usually not relevant, but other information (e.g. species longevity) is lost.

The paleontological literature often adopts the representation of budding trees (e.g., Raup, 1985; Foote, 1996; Benton and Pearson, 2001) which differs from phylogenies of extant lineages by informing which branches are new lineages and which are the continuation of the ancestral. In this case, the assumption of symmetric or asymmetric speciation is made clear. Budding is recognized in the fossil record as a cladogenetic event in which a new lineage appears as a branching of an older lineage that can still be found after the speciation event (e.g., Foote, 1996). Although budding speciation might be commonly reported in the fossil record, it is rarely incorporated into phylogenetic comparative models. Raup (1985) mentioned that budding should not influence estimates of diversification rates because the addition or subtraction of lineages at any given time would be perceived similarly if we represent a phylogeny either by budding or bifurcation. This question has been revisited decades later by Bapst and Hopkins (2017) and Crouch et al. (2021), both showing that budding can change divergence time estimation and, as a result, alter our estimate of the accumulation of lineages through time. Thus, budding should not influence diversification rates only if the true dated tree is known, otherwise changes in divergence time estimation can potentially impact estimates of diversification.

Up to date, there has not been any attempt to investigate if, and how, budding speciation can affect our estimates of the tempo and mode of trait evolution. De Lisle and colleagues (2021) explored the effect of shifts in the adaptive optima on extinction rates using a model based on population-level dynamics and showed that lineages would rarely survive peak shifts and those occupying stable optima are expected to have greater chance of survival. In turn, Duchen and colleagues (2021) studied how cladogenesis changes the average species phenotype using individual-level simulations and showed that new lineages budding off an ancestral population

can show a significant phenotypic deviation due to neutral processes. Combined, these results point to the idea that long-lived evolutionary lineages might occupy stable optima (also see Goldberg and Foo, 2020). In contrast, new lineages may budd off with distinct phenotypes, due to the effect of cladogenesis on the species trait and selection to occupy a distinct region of the morphospace. Most of the new lineages might be relatively short-lived because selection towards a novel adaptive peak is associated with a greater chance of extinction (also see Rosenblum et al., 2012) whereas the ones that arrive at stable optima are more likely to survive.

Here we discuss in which circumstances budding speciation can affect our estimates of the tempo and mode of trait evolution. More specifically, we review the properties of phylogenetic comparative models of trait evolution that dictate whether or not budding speciation can influence our conclusions. We introduce a lineage-age dependent process that emulates the budding of new lineages from stable long-lived progenitor lineages (Hagen et al., 2015; Hagen et al., 2018; De Lisle et al., 2021; Goldberg and Foo, 2020; Duchen et al., 2021). Finally, we test how budding speciation can bias our estimates using simulations of biologically plausible scenarios of diversification and trait evolution. In our view, the acknowledgment of budding speciation in comparative approaches can help bridge decades of research in paleontology with modern phylogenetic comparative methods, and improve our perspective of macroevolution.

What is budding and how to recognize it?

Budding is defined as a speciation event in which the new species co-occurs in time with its ancestral lineage (Foote, 1996). In the absence of the fossil record, budding has been defined as a new lineage formed within or at the edge of the ancestral lineage (Anacker and Strauss, 2014) or as a biological cause of paraphyletic species (i.e., incomplete lineage sorting associated with recent speciation; see Funk and Omland, 2003). Budding speciation is opposed by bifurcation—the split of an ancestral lineage into two new ones (Fig. 1). Hagen and colleagues (2015) utilized the term *asymmetrical* speciation for budding and *symmetrical* speciation for bifurcation, here we use these terms interchangeably. To help differentiate the parent from the daughter lineage we use the term progenitor lineage (see Gottlieb, 2004). A progenitor lineage is a lineage that has given birth to one or more new lineages through budding. Progenitor lineages

vary in duration and in how prolific they are.

The paleontological literature suggests budding is a common evolutionary pattern and some argue it represents the majority of speciation events observed in deep time. For example, Aze and colleagues (2011) reconstructed a large phylogeny of macroperforate foraminifera in which the majority of cladogenetic events were recognized as budding through analysis of morphological characters. Bapst and Hopkins (2017) applied an explicit probabilistic model to date a phylogeny of trilobites and also show that budding events are often supported by the fossil record. Similarly, Parins-Fukuchi (2021) re-evaluated the diversification of hominins, also using a probabilistic model, and suggests the occurrence of budding speciation events. In all these cases, budding was recognized in the fossil record when a new, and morphologically distinct, lineage was detected while its ancestral remains with the same, or similar, phenotype after the event (e.g., Aze et al., 2011; Bapst and Hopkins 2017). Of course, such an estimate depends on evidence from fossil occurrences that both sister lineages co-occurred in time for at least part of their longevity (Foote, 1996).

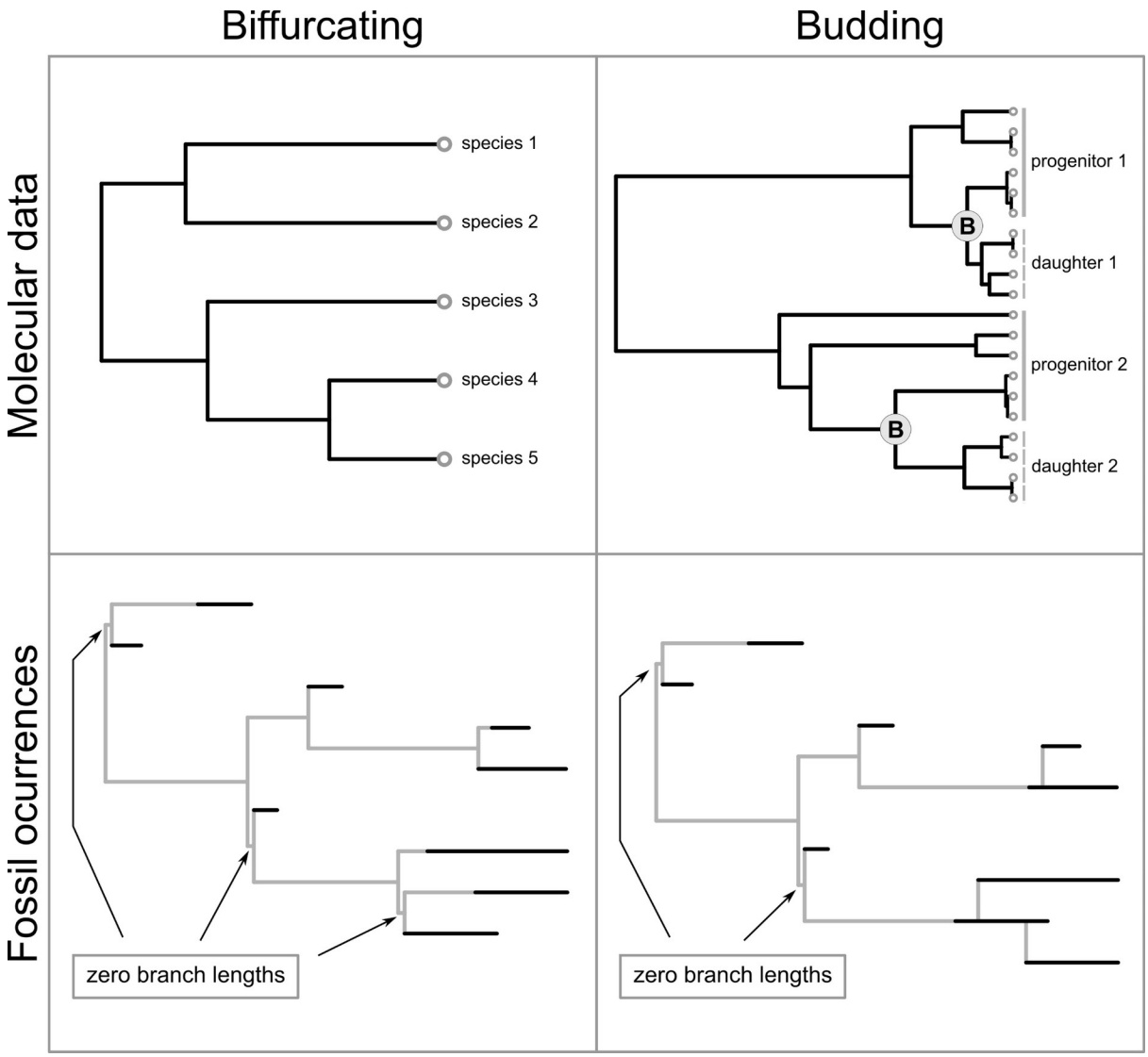


Figure 1: Branching history of lineages interpreted through bifurcation or budding. Top row shows two distinct phylogenetic trees estimated with molecular sequences. Bottom row shows the same topology reconstructed using fossil data (dark lines represent the observed longevity of fossils and gray lines show "ghost" lineages). Budding can only be identified with molecular data if more than one population is sampled per species (Top-right). Bifurcating phylogenies of fossil taxa requires the addition of "ghost" lineages (bottom row), which represent estimated longevity beyond fossil occurrences. Budding decreases the total length of "ghost" branches and minimizes the number of zero branch lengths (bottom-right) added due to incongruence between topology and age of fossil lineages.

If budding speciation is frequent, we also expect to recover more recent events using molecular data. However, the genetic signal left by asymmetrical speciation has a much shorter shelf life than the signal stored in the fossil record. When a new lineage buds off from its ancestral, the progenitor species becomes paraphyletic (Fig. 1) due to incomplete lineage sorting (Funk and Omland, 2003). If the new lineage maintains cohesion and does not go extinct (see De Lisle et al., 2021), the alleles should complete sorting and the ancestral become monophyletic (Funk and Omland, 2003 and references therein), erasing the signal for budding. Otero and colleagues (2022) show an interesting case in *Iberodes* plants which underwent two events of budding speciation within the last 5 million years. In both instances, the new lineage evolved distinct morphological and ecological traits (Otero et al., 2022). *Iberodes* has inland and coastal species and the potential change in selective pressure together with the peripheral distribution of the younger coastal lineages likely were key factors for budding divergence. Similarly, Papuga and colleagues (2018) show peripheral plant populations that have lower niche breadth (i.e., are more specialized) than central populations as well as divergence in ecological traits (i.e., soil parameters), both factors that can cause budding by ecological speciation. Strong evidence for budding speciation was also detected from molecular phylogenies by Baldwin (2005), showing that *Layia glandulosa* (a Compositae plant) is the progenitor species for *L. discoidea*. Anacker and Strauss (2013) tested 71 sister pairs and demonstrated that young divergences frequently show overlapping and asymmetrical ranges, an indication of budding speciation. This asymmetry was not detected among older clades, suggesting the signal of budding on geographic distribution of sister pairs is lost as lineages get older.

Although the emphasis on budding speciation in paleobiology is more present than in neontological studies, there is strong evidence for its role in species diversification. Many cases of incomplete lineage sorting might be resulting from budding speciation events (Funk and Omland, 2003; Gottlieb, 2004; Baldwin, 2005; Papuga et al., 2018; Otero et al., 2022), but are sometimes not referred to with these terms. This creates a false perception that budding is a pattern that can only be recovered in deep time, using the fossil record (see Benton and Pearson, 2001). If frequent, and we suspect it is, budding can be an important factor in understanding the patterns of trait evolution because peripheral populations can show distinct mean phenotypic values (Papuga et al., 2018) and, if divergence happens, can originate new lineages with distinct

average phenotypes (Gottlieb, 2004; Duchen et al., 2021) and evolutionary trajectories (De Lisle et al., 2021).

Hereon we explore the potential effects of budding speciation in phylogenetic comparative methods of trait evolution. We discuss diverse evolutionary scenarios and the properties of statistical models used to test for them. First we investigate under which circumstances the existence of asymmetric speciation would not influence our inferences, then we enumerate a series of scenarios in which taking into account the possibility of budding speciation could change how we study trait evolution using phylogenetic trees.

When budding doesn't matter

Put simply, budding will not influence our understanding of phenotypic evolution if changes happening at any point in time, and at any branch of the phylogeny, are independent of the prior history of the lineage and their ancestors. If evolutionary changes are truly memoryless, then the recognition of a progenitor species will not matter for our estimates of trait evolution. Perhaps not surprisingly, the application of memoryless trait evolution models (such as the Markov model and other diffusion-based models) is common in comparative studies.

Two important simplifications were introduced when Felsenstein described the method of independent contrasts (1981; 1985). First, trait changes happen independently in each branch of the phylogeny. This allows for the translation of a phylogenetic tree into a variance-covariance matrix describing the expected covariance in trait observations for any species pair. Second, evolutionary changes at each point along a branch are independent and identically distributed (iid). This guarantees the tractability of the model, because the same distribution can be used to model expected trait changes at any branch. Most models of trait evolution developed after Felsenstein's seminal work share these assumptions (see review in O'Meara, 2012; Pennell and Harmon, 2013). However, most, if not all, phylogenetic comparative models were not created with the intent to accurately describe evolution in a mechanistic way, and the use of these simplistic approaches does not mean we assume evolution follows these rules. Comparative models originated from a necessity to incorporate evolutionary history into statistical tests—the simplest way we can adequately achieve this goal, the better.

If models of trait evolution that assume a homogeneous process across all branches of the tree are adequate representations of macroevolution, the incorporation of budding speciation will not change our estimates. This is because differentiating lineages in the phylogenetic tree will have no influence on the underlying model—lineages become effectively interchangeable. However, this is not the trend that we are currently observing in phylogenetic comparative methods. After the description of the first models of trait evolution, extensions allowing heterogeneity in the process, often associated with some predictor, were implemented and shown to better capture the variation of empirical data (e.g., Eastman, 2011; Uyeda and Harmon, 2014; Rabosky and Goldberg, 2015; Caetano et al., 2018). More recently, studies have demonstrated that rate heterogeneity should be taken into account even when no a priori predictors are present (e.g., Rabosky and Goldberg, 2015; Beaulieu and O'Meara, 2016; Caetano et al., 2018). In fact, the debate on the importance of model adequacy took the spotlight in the comparative methods community. Many investigations have shown models that fail to capture important temporal or branch-specific variation observed in empirical data can severely bias our biological interpretations (e.g., Rabosky and Goldberg, 2015; Pennell et al., 2015; Caetano et al., 2018). Development of more adequate models often means the increase in model complexity to reflect the dynamic nature of macroevolution and, as a result, hint that the condition of homogeneous and memoryless evolutionary changes across the tree of life—under which budding would not matter—is unlikely. Below we discuss how budding could be generating heterogeneity in the phylogenetic history of phenotypes and in which ways the results affect our conclusions about trait evolution.

When budding matters

In direct contrast to the previous cases, under which budding speciation would not influence our estimates, budding is expected to be important in any evolutionary scenario in which the identity of evolutionary lineages is relevant. This might be the case if lineage-age influences the tempo and/or mode of trait evolution (Pennell et al., 2014; Hagen et al., 2018; Goldberg and Foo, 2020) or if the age of competing lineages is important to predict their competitive strength and/or risk of extinction (Ezard et al., 2011; Rosenblum et al., 2012; Carrillo et al., 2020; Januario and Quental, 2021). Although there are other evolutionary

processes under which the identity of lineages are important, here we focus on these two scenarios for simplification. In contrast, there are special cases that generate heterogeneity in trait evolution but under which budding is likely not relevant. For example, if trends in rates of trait evolution are due to abiotic causes, such as response to climatic changes or mass extinctions, then all living lineages would be under the same effect and, everything else being equal, we would expect responses to be independent of lineage identity (e.g., Clavel and Morlon, 2017). Heterogeneity in trait evolution can also be due to a predictor variable which is also modeled across the branches of the tree using ancestral state reconstruction. In this case, budding speciation could be the cause of a change in the mapping of the predictor, especially for discrete variables (see "*Effects of budding speciation and lineage-age dependent processes on ancestral reconstruction"*).

Below we enumerate scenarios in which budding speciation could influence our conclusions about the tempo and mode of trait evolution.

1) When evolutionary changes are concentrated at or near lineage origination

The central distinction between symmetric and asymmetric speciation is the age contrast between progenitor and daughter lineage immediately after divergence. The daughter species will usually have a smaller population size and geographic distribution (Foote et al., 2007; Liow and Stenseth, 2007) and it will often be under higher extinction risk due to the change in adaptive optima relative to the progenitor species (De Lisle et al., 2021). If lineage-age is related to the tempo of trait evolution, such that younger lineages are expected to show faster rates of trait change, we would expect relatively more evolution to happen in daughter lineages when compared to their progenitors close or near speciation events. Thus, disparity between two descendants of a budding node in a phylogenetic tree cannot be attributed to equal amounts of change at each branch because evolution has been concentrated in the daughter lineage. Furthermore, if the same progenitor lineage generates another daughter by budding, the starting point in the morphospace for the new lineage (i.e., the ancestral state) will be closer to the ancestral state of the steam node for the clade than expected under a homogeneous, lineage-age independent, trait evolution model. In other words, a lineage-age dependent process can produce a scenario in which the progenitor lineage "carries" its ancestral phenotype through one or more

budding speciation events.

Younger lineages might undergo quick phenotypic change as they move towards a new adaptive peak (Hunt et al., 2008; De Lisle et al., 2021) whereas prolonged time under a stable adaptive zone might be associated with a slowdown in trait evolution—often linked to specialization (Goldberg and Foo, 2020; De Lisle et al., 2020). Budding has been shown to happen with peripheral populations which have restricted geographic range and limited population size (Otero et al., 2022) and hence might predispose, or even drive, evolutionary changes at speciation. The reduced population size of young species contributes to neutral evolutionary processes, such as genetic drift (Rosenblum et al., 2012; Dunchen et al., 2019), and increases vulnerability to stochastic geographical and ecological events that could cause extinction. The asymmetry in trait evolution expected under budding speciation might reflect these differences. Undeniably, recently diverged lineages face evolutionary challenges distinct from older, more established, ones.

2) *When daughter lineage survival depends on being ecologically different from its progenitor*

The asymmetry in age generated by budding speciation can influence the competitive strength of daughter lineages relative to their progenitors and, as result, also the extinction risk of the younger lineage (Ezard et al., 2011; Rosenblum et al., 2012; Carrillo et al., 2020; Januario and Quental, 2021). Progenitor lineages are expected to have larger population sizes and geographic ranges (Anacker and Strauss, 2014) which, everything else being equal, improves their chance of survival in interspecific competition with newly formed species. When competition between progenitor and daughter lineages is present, daughter lineages that have lived enough to be sampled, either in the fossil record or still living today, are expected to be sufficiently distinct from their progenitors to have escaped competitive exclusion (De Lisle et al., 2021). Of course, competition is not dependent on budding and its effect is equally expected if speciation would be symmetric. However, budding potentially intensifies the effect of interspecific competition, and might increase heterogeneity in trait evolution. Symmetric speciation produces two new lineages equally distinct from the ancestral lineage, such that it is not possible to recognize either as the continuation of the ancestral lineage after speciation. Thus,

interspecific competition between lineages after symmetric speciation, at least a priori, would not
284 favor either lineage because they are effectively interchangeable. Under the asymmetric scenario, the survival of the daughter species is expected to depend on character displacement, which can
286 differentiate daughter lineages from their progenitors and alleviate competition. Alternatively, differentiation between progenitor and daughter lineages can be due to ecological speciation, in
288 which ecological divergence resulting from adaptation to different environments is the key aspect that underlies the speciation process (Nosil 2012). Under this speciation scenario the
290 daughter species, by definition, would be ecologically different from the progenitor species.

Diversification in islands is a flagship example of character displacement due to
292 interspecific competition among closely related species. These have been studied in a variety of organisms and the evolution of distinct ecomorphs has been frequently reported (e.g., Rundle et
294 al., 2000; Gillespie, 2004; Mahler et al., 2013). Ecomorphs are defined by a unique set of morphological and behavioral traits that are important to habitat use and exploitation of
296 resources. Distinct ecomorphs, by definition, show little overlap in these traits, supporting the idea that ecomorphs represent separated evolutionary optima, and, more importantly, there is
298 evidence that character displacement in these groups is an indication of interspecific competition. Phylogenetic reconstructions of adaptive radiations in islands often show support for convergent
300 evolution of ecomorphs within a clade. Budding speciation could help us explain the frequency of convergence observed in these systems. For instance, a well-established progenitor lineage
302 representing one ecomorph type—thus, occupying an stable adaptive optimum—could displace daughter lineages into distinct optima (and/or exclude those that fail to do so). As a result,
304 daughter lineages converge into similar ecomorphs that can or not co-occur in space and time (Fig. 2). Under this scenario, convergence would be a result of a finite number of potential
306 ecomorphs, due to the characteristics of the environment, phylogenetic inertia, and developmental constraints.

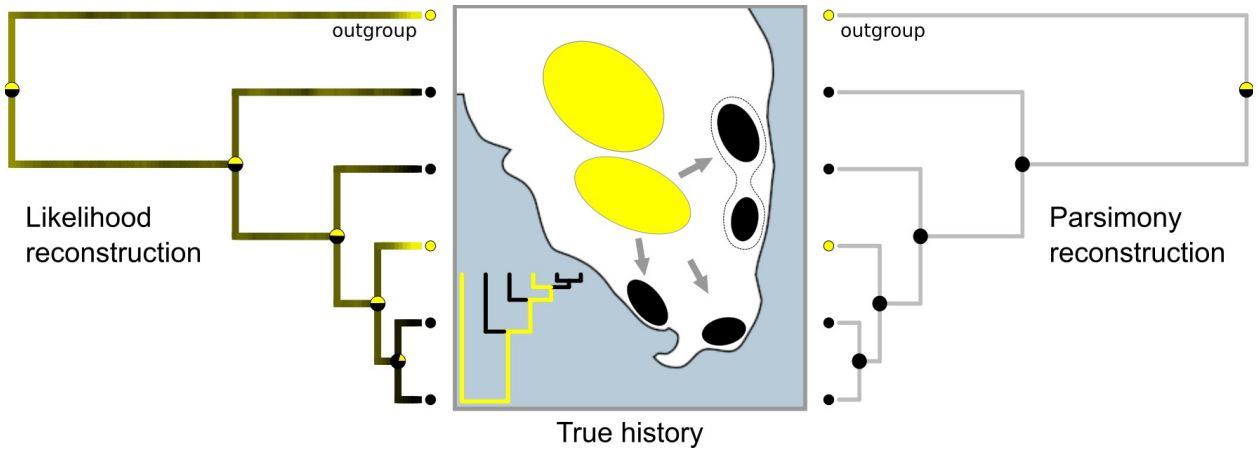


Figure 2: Conflict between ancestral state reconstruction and the true history of lineage evolution under budding. The progenitor lineage is the sister lineage to the outgroup, both showing a mainland distribution (in yellow). New species appeared through budding and converged three times into a different niche, associated with the coastal distribution (in black). Neither likelihood or parsimony reconstruction recovered the true history of the lineages. However, the most parsimonious reconstruction yields the incorrect history with high certainty. Left: Marginal ancestral state reconstruction for the best fit Markov model. Branches are painted following a posterior distribution of 100 stochastic maps. Right: Most parsimonious ancestral reconstruction following Sankoff's (1975) algorithm. Center: True history of the trait. Geographical distribution of five species at the present. Mainland distributions (in yellow) are ancestral, but parsimony reconstruct one of them as a recent reversal. Phylogeny in bottom left shows the continuation of the progenitor lineage through multiple budding events.

Although we predict an intensifying effect of asymmetric speciation on interspecific competition, other nature of interactions might be more complex. Nuismer and Harmon (2015) demonstrated mathematically the effect of the mode of trait evolution and phylogenetic diversity (PD) in the outcome of interspecific interactions in communities of closely related taxa. They showed that PD is a good predictor of interspecific interactions if these are dependent on phenotypic matching, such as competition, with more closely related lineages showing stronger interspecific interactions. Budding could change the relationship between PD and expected trait similarity, because long-lived progenitor species would accumulate less evolutionary changes than expected under a homogeneous trait evolution model, such as Brownian motion, causing the role of the phylogeny as a predictor to become less prevalent. In contrast, Nuismer and Harmon (2015) show that under stabilizing coevolution the phylogeny is a poor predictor of interactions, and we do not expect that budding would influence this result. Even though Nuismer and Harmon (2015) have not considered asymmetric speciation events, its potential effect on the

outcome of interspecific interactions likely depends on many factors and deserves further study.

How budding speciation and lineage-age dependent processes influence estimates of trait evolution?

We use simulations to evaluate scenarios in which budding speciation can deviate our expectations and, more importantly, test if these deviations have a significant impact in our understanding of phenotypic evolution using phylogenetic trees. We explored the impact of budding on continuous and discrete trait evolution models widely used in the literature. We also investigated how likely is budding speciation to produce erroneous estimates of ancestral states when long-lived progenitor lineages are present. We focused our attention on phylogenies of extant species, which are most often estimated using molecular data, and in the absence of fossil information. There are methods to incorporate fossils not only for node dating, but also to help on estimates of trait evolution (Slater et al., 2012). However, here we want to emulate scenarios in which comparative methods are most frequently employed.

As we discussed above, budding speciation will be important for evolutionary processes in which the identity of the lineage is relevant. Here we chose one of such processes, the case in which rates of trait evolution are dependent on lineage age. We modeled lineage-age dependent rates of trait evolution with an exponentially distributed slowdown process. This emulates a scenario of fast evolution in younger lineages with a rate slowdown as the lineages get older, producing more evolutionary novelties early and tending towards phenotypic stability later on. We produced a continuum from a mild effect of lineage-age to a strong concentration of rates close to the nodes. We standardized the average rate across the tree throughout the simulated scenarios to differentiate the rate slowdown within lineages from the confounding effect of an overall change in the average rate across replicates. We controlled the frequency of budding speciation by assigning a probability of budding to each node of the phylogeny. Phylogenetic trees were simulated independently of the traits and the history of budding using a constant rate birth-death model ($\lambda = 0.2$, $\mu = 0.1$) over a 40 My interval. To reduce variation in tree size we used rejection sampling to keep only phylogenies with 250 to 350 extant lineages (mean tree size of unconstrained simulation is 300 species). Phylogenies can be simulated independently of the presence of budding because the birth-death model controls the events of lineage origination and

extinction, regardless of the mode of speciation. We simulated data using the complete phylogenies but pruned all extinct lineages to emulate the use of molecular trees. A detailed report of the simulation is available in the Supplementary Materials.

Effects of budding and lineage-age dependent processes on the adequacy of continuous trait evolution models

After simulating continuous traits under a variety of scenarios of budding speciation (see Supplementary Materials) we estimated parameters for Brownian motion, Ornstein-Uhlenbeck, and Early-Burst models. Neither of these models are fully congruent with the process we used to simulate the data, but our goal was to evaluate the impact of budding speciation on models widely used in the literature. To account for rate variation, we used a variable rates BM model (Eastman et al. 2011) to rescale the branches of the tree according to the rates identified by the reversible jump MCMC approach. This method accommodates—at least part of—the heterogeneity in the variance component of these models (i.e., σ^2) which were introduced by the lineage-age dependent function used to simulate the data. We compared all results using homogeneous and variable rate models (Figs. S1-3).

In overall, variable rate BM models showed significant improvement in model adequacy. Model adequacy tests for a homogeneous rate BM model (Pennel et al. 2015) detected a negative relationship between node height and the size of the phylogenetic independent contrasts (PICs) indicating that larger trait changes are estimated to have happened closer to the tips (Fig. S2; S_{hgt}). However, a variable rate BM model (also Fig. S2) does not show evidence for such deviation, suggesting this is an effect of underestimating the rate variation introduced by budding. A regression of phylogenetic independent contrasts (PICs) and their expected variance (S_{var}) shows that nodes connected to shorter branches are associated with more trait change (Fig. S1), independently if the fitted BM model was homogeneous or not. Inadequacy of S_{var} is expected due to the concentration of rates early in the history of lineages and the stronger effect of slowdown on the more longevous lineages when compared with short-lived ones. The deviances for S_{hgt} and S_{var} are only detectable when the rate slowdown is strong (> 50% rate slowdown when lineages become 2.5 My old), meaning that a relaxed rates model (Eastman et al. 2011) is able to adequately describe trait variation if lineage-age effects are mild.

390 Inadequacies in both S_{hgt} and S_{var} point to trait changes concentrated close to the tips, which is expected since molecular trees have an accumulation of nodes near the present, some of these
392 generated by budding, producing new lineages with high rates of trait evolution.

Model inadequacy could be an artifact of unobserved speciation events deeper in the tree
394 but results remain constant even when we replicate analyses using the full history of lineage diversification of the clade (i.e., including the extinct lineages). Deviations of S_{hgt}, S_{var}, and C_{var}
396 get stronger as the intensity of the lineage-age dependent slowdown factor increases. On the other hand, the frequency of budding speciation (i.e., if rare or common) has no detectable
398 important effect on model adequacy under all summary statistics we reported above (Figs. S1-3). We did not verify any deviation on the remaining summary statistics adopted by Pennel and
400 colleagues (2015).

With respect to the support for alternative trait evolution models (i.e., Brownian motion,
402 Ornstein-Uhlenbeck, and Early-Burst) in function of lineage-age dependent rate variation, the OU model shows a marked increase in AIC weights in response to stronger slowdown factors
404 (Fig. 3). In the majority of cases, phylogenetic half-life was estimated to be multiple times longer than the age of the clade (40 My) indicating processes indistinguishable from Brownian motion
406 (Cooper et al. 2016). Average phylogenetic half-life estimated across replicates was only shorter than clade age when the strongest slowdown factor was applied (Fig. S4). This means that an OU
408 process is only supported when progenitor lineages are practically in stasis whereas virtually all trait evolution is concentrated on the origination of new lineages following budding speciation.
410 Similarly to the model adequacy summary statistics reported above, these results remain constant regardless of the frequency of budding speciation used to simulate the data (Fig. S4). Results are
412 also robust to the inclusion of extinct lineages, suggesting a lack of influence from unobserved speciation events at deep time. This is expected, since lineage-age dependent slowdowns will
414 concentrate changes in relatively young lineages which are more abundant towards the present—the same pattern observed under OU models.

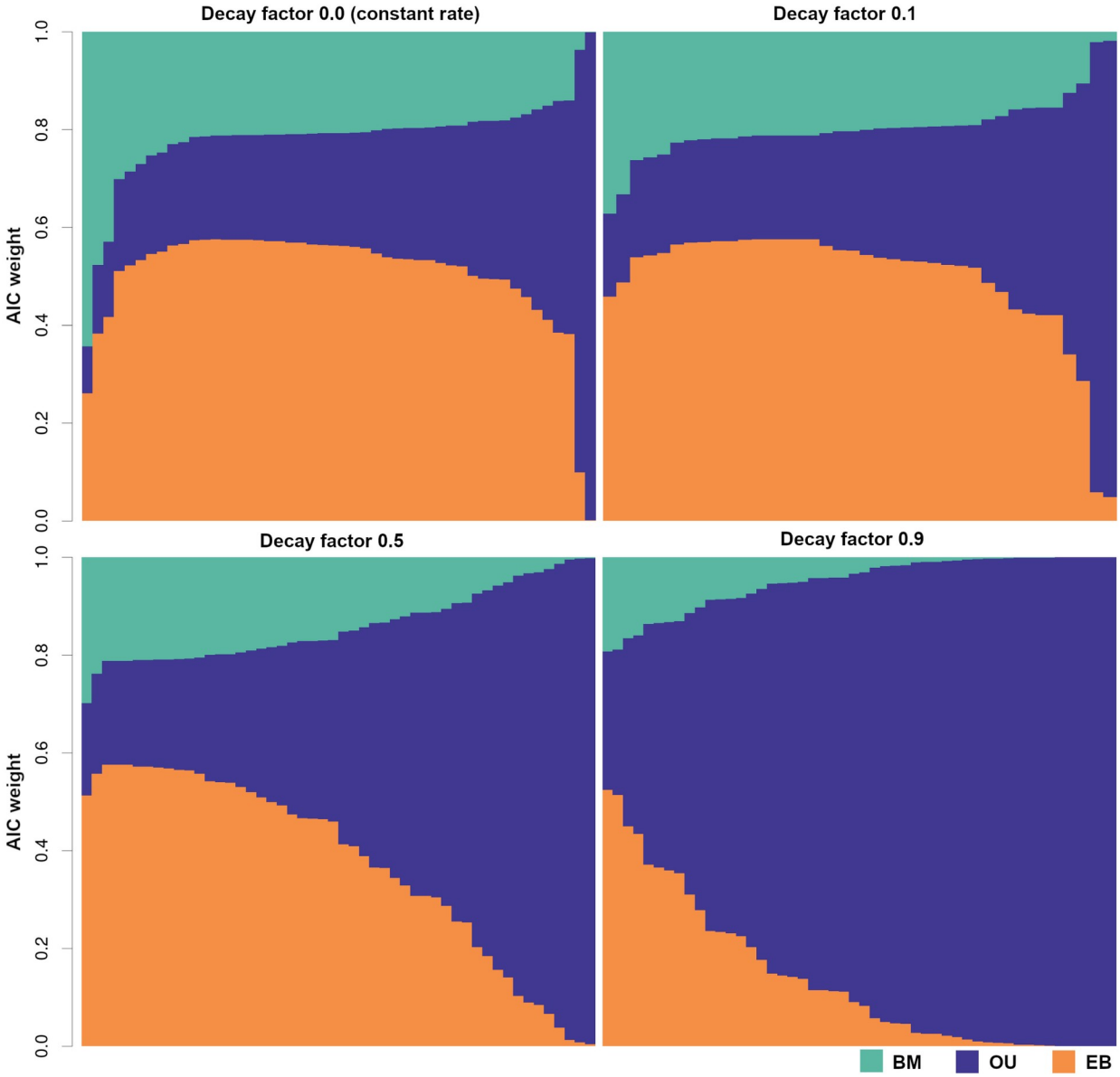


Figure 3: Support for alternative trait evolution models in function of the lineage-age dependent process in the absence of budding. Panels show the distribution of AIC weight for Brownian motion (BM), Ornstein-Uhlenbeck (OU), and Early Burst (EB) models. Slowdown factor on the BM rate of 0.0 (constant rate), 0.1, 0.5, and 0.9 means that a lineage had a rate reduced by that proportion when it reached 2.5 My ($\boldsymbol{\sigma}_{2.5}^2 = \boldsymbol{\sigma}_0^2 - \boldsymbol{\sigma}_0^2 *$ decay factor). Support for OU models increases as relative rates of trait evolution concentrate at speciation events. Results remain the same if we use a variable rates model (rjMCMC) instead of a homogeneous one.

When we simulate continuous traits under a lineage-age dependent process, the Brownian motion model with varying rates adequately describes most characteristics of the data but fails to capture the concentration of trait changes on shorter branches (S_{var}). If lineage-age dependent processes happen in nature, our results reinforce the cautionary note that parameter estimates can be more informative than model choice alone (Cooper et al. 2016). Although the majority of scenarios supported OU models, only the strongest case of rate slowdown resulted in patterns distinguishable from Brownian motion. Fortunately, the deviance of the slope of absolute contrasts in function of their expected variance (S_{var} - Pennel et al. 2015) proved to be effective in indicating the presence of a concentration of trait change towards shorter branches, even when controlled for rate variation. In the absence of a more specific approach, we suggest that S_{var} can be used to identify potential cases of lineage-age dependent rates in continuous traits. Finally, we recommend that estimates of Brownian motion using empirical data, including ancestral state reconstruction of continuous traits, should be conducted with models that allow for rate heterogeneity.

Effects of budding speciation and lineage-age dependent processes on ancestral reconstruction

Ancestral reconstruction without evidence from the fossil record is challenging and sometimes impossible. Estimates of the phenotype of extinct lineages depend heavily on the capacity of the model to capture the evolutionary trends that have happened in the past. Nevertheless, simulation studies allow us to evaluate these methods under a “best-case” scenario. By introducing budding speciation to a background of relatively simple trait evolution we can investigate how asymmetric speciation is expected to affect our estimations of trait change along the branches of the tree. Here we focus on discrete traits because most studies employing continuous traits are not interested in the ancestral value for a particular node but in evolutionary trends affecting clades, such as shifts in rates of trait change and the estimation of evolutionary optima.

To evaluate the effect of budding speciation in ancestral reconstruction we used an index of how incorrect the estimate at a node is with respect to the true history of the trait. We measured the highest marginal probability among all states excluding the true state for that node as an estimate of “wrongness”. This metric reflects how likely an internal node will be estimated

as the wrong ancestral state with strong support from the model. Note that this is distinct from uncertainty. When uncertainty is maximized, then each of the n states will have marginal ancestral state probability equal to 1/n, meaning there is no information to determine the state of the node and any of the possible states are equally likely. In contrast, wrongness is maximized when we have certainty of ancestral reconstruction but the state is incongruent with the true (simulated) history. Uncertainty is a lesser problem than wrongness because we will not, or at least we should not, support or refute evolutionary hypotheses based on uncertain estimates. Wrongness, on the other hand, can result in misleading interpretations.

We used linear mixed models to test for the association between node age and wrongness, taking into account the frequency of budding speciation, the degree of lineage-age dependent rates, and the presence of cladogenetic changes (see Table 1). All fitted models show a strong positive association between node age and wrongness (Fig. 4), meaning that ancestral reconstruction of old nodes (i.e., those closer to the root of the tree) are more likely to be misleading. The chosen model suggests that budding speciation has a significant effect on wrongness when cladogenetic trait changes are also present (see Table 1 and Fig. 4). This means the effect of long-lived progenitor lineages in rates of trait evolution significantly impact ancestral state reconstruction if cladogenetic changes are present. Figure 4 shows a graphical representation of the increase in the frequency of wrong reconstructions in function of node age for each treatment. There is no detectable difference between the null model, representing homogeneous rates of trait evolution across the tree, and the presence of budding without cladogenetic changes. Budding speciation associated with cladogenetic changes in the trait increases the chance of misleading ancestral state reconstruction, especially for younger nodes. This result is somewhat unexpected and important because younger nodes are often expected to have more information than older ones (Schultz et al., 1996; Boyko and Beaulieu, 2021). The case where budding is associated with cladogenetic changes represents the conceptual scenario described above as "*When daughter lineage survival depends on being ecologically different from its progenitor*".

The ancestral state of progenitor lineages can be mistakenly reconstructed as a novelty

Budding can increase the chance of reconstructing the wrong state closer to the tips of the phylogeny (see Fig. 4), especially if it is associated with cladogenetic events. Wrong ancestral reconstructions are more likely to happen if the true history of the trait is incongruent with the most parsimonious reconstruction (Fig. 2). One potential cause for such a pattern is competition between progenitor and daughter lineages resulting in a displacement of the newly formed species, which either evolves traits that help escape competition or gets extinct, likely without leaving a record (Rosenblum et al., 2012; De Lisle et al., 2020). If a progenitor lineage produces several daughters that evolve a distinct phenotype, the frequency of trait states among extant species can mislead ancestral state estimation. For instance, if the true ancestral state for the clade (i.e., the state of the progenitor lineage) becomes rare among extant species because of budding speciation, ancestral reconstruction can estimate a recent reversal instead of multiple events of convergent evolution (see Fig. 2), especially under maximum parsimony. Convergence could simply be due to a limitation of the possible states for the trait. For example, in plants, the evolution of coastal lineages from an inland ancestor, a trait with only two possible states, has been recorded to occur multiple times, likely due to the isolation of populations at the edge of the ancestral range (Papuga et al., 2018). Many ecological traits have similar broad definitions that restrict the pool of possible states and increase the chance of convergence. It is important to note that such generalizations are necessary to recognize patterns in nature as well as to provide adequate statistical power to comparative tests.

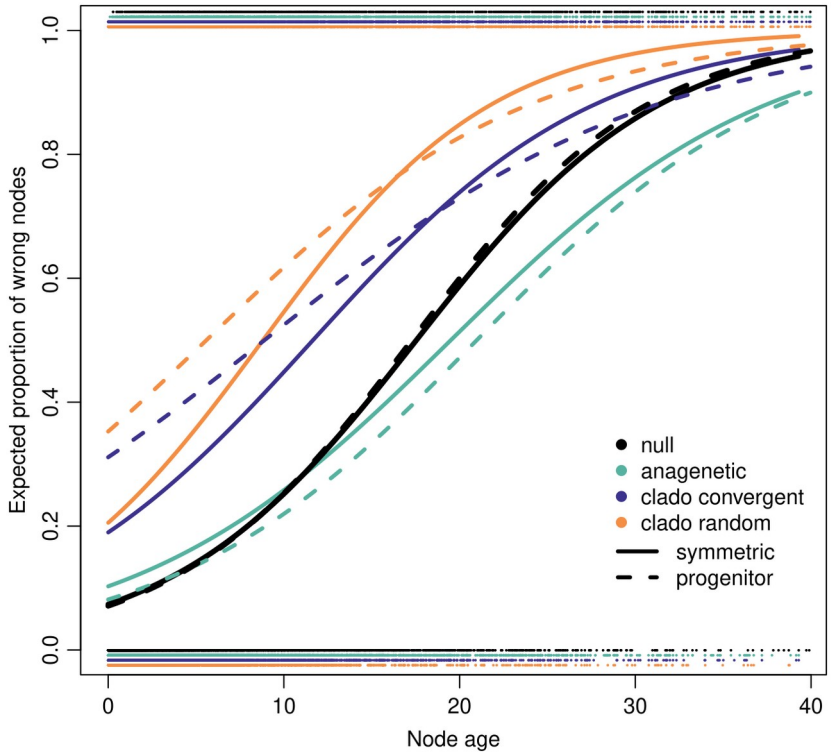


Figure 4: Relationship between the presence of nodes with wrongly reconstructed states and node age. Nodes were considered wrongly reconstructed if the marginal ancestral state probability for any state distinct from the correct state were > 0.5 at that node. Logistic regressions were performed with the pooled results from the 50 simulation replicates per group and independent for each type of node and study group (see **Table 1**). Solid lines represent nodes with symmetric speciation whereas dashed lines represent progenitor lineages (i.e., nodes with budding speciation).

Wrongly reconstructing the ancestral state of a clade as a recent reversal (Fig. 2) becomes especially relevant when evaluating scenarios of Dollo's law. Dollo's law is often evoked as the hypothesis that a complex character is unlikely to be regained once it is lost (see discussion in Collin & Miglietta, 2008). Testing Dollo's law requires integrated investigation of phylogenetics with genetics and development in order to ascertain whether the character was genuinely regained or independently (re)evolved. The role of comparative methods is to provide evidence for reversals using model testing and ancestral state reconstruction (Collin & Miglietta, 2008). In some cases, inferences from comparative approaches are incongruent with what is known about the genetics, development, and physiology of organisms. Pyron (2015), for example, discusses

514 the inference of multiple transitions from viviparity back to oviparity in snakes despite the external evidence against so (Griffith et al., 2015). Pyron (2015) suggests that comparative
516 approaches should not ignore external evidence but also that findings from phylogenetic inferences should be further investigated integratively. We agree with these suggestions,
518 however, the models we apply today can, and should, be re-evaluated in light of what we learn about the patterns and processes of macroevolution. An unlikely ancestral reconstruction of
520 parity might simply mean that our current modeling paradigm is not able to adequately capture the evolutionary history of the trait. For instance, budding speciation, as we represent in this
522 study, could help explain oviparous lineages nested deep into viviparous clades as descendants of long-lived progenitor lineages (see discussion in Pyron 2015). This scenario would indeed
524 require many additional evolutionary transitions, which are heavily penalized under a parsimony paradigm. However, this alternative history would support the extensive knowledge about
526 genetics, development, and physiology of snakes (see discussion in Griffith et al., 2015). In our view, when there is a clash between model estimates and biological knowledge, it is wise to
528 review our models and ponder which important processes the model is failing to capture. More importantly, this is an example of how external biological information can be used to support the
530 inclusion of a "more complex" evolutionary history, such as budding speciation, as a plausible scenario in comparative studies.

532 **Can budding help us understand complex histories of trait evolution?**

Here we used simulations to investigate the effect of budding in models of trait evolution
534 with a focus on the frequency of asymmetric speciation, rates of trait evolution varying in function of lineage-age, and cladogenetic trait changes. Our results show the introduction of
536 budding speciation has an impact but does not completely hinder the utility of the most popular models of trait evolution. The effects of budding we report here, for the most part, can be
538 translated as heterogeneity in trait evolution among lineages. Different from other sources of heterogeneity, such as predictor regimes (e.g., Uyeda and Harmon, 2014), budding produces
540 variation dependent on lineage identity, requiring the identification of progenitor lineages comprising multiple contiguous branches of a phylogeny. From this perspective, we are
542 optimistic about incorporating budding speciation into phylogenetic comparative models. In fact,

it is plausible that a portion of the heterogeneity captured by rate varying models, such as hidden rates models for discrete traits (Beaulieu and O'Meara, 2016; Caetano et al., 2018; Boyko and Beaulieu, 2021) and those applying reversible jump MCMC for continuous traits (Eastman et al., 2011; Uyeda and Harmon, 2014), is due to the effect of budding speciation.

Table 1: Relationship between the presence of nodes with wrongly reconstructed states and node age. Logistic regressions were performed with the pooled results from the 50 simulation replicates (see **Figure 4**). *wrong_node* : Nodes were considered wrongly reconstructed if the marginal ancestral state probability for any state distinct from the correct state were > 0.5 at that node. *node_age* : The age of the node in the phylogeny. *sim_type* : The type of trait history simulated. The different types include a simple single rate MK model (i.e., the null model), a model with rates decreasing in function of lineage-age and no cladogenetic change, a model with rates decreasing in function of lineage-age and cladogenetic changes associated with budding events, and a model with rates decreasing in function of lineage-age and cladogenetic changes associated with budding events in the presence of convergent evolution among all daughter lineages of the same progenitor lineage. *node_type* : The type of speciation event at the node (if symmetric or budding). *descendant_history* : The type of history of the clade descendant of a budding event. We separated clades by the presence or absence of further budding events among the descendants of a budding node and if the budding event leads to a terminal branch on the phylogeny or not.

Model (as used in R)	**Included effects**	**AIC**	**DF**
glmer(wrong_node ~ node_age + sim_type + (1| node_type) + (1|descendant_history) , family = binomial)	Simulation type, if nodes represent budding events, and the history of the descendants of budding events	30119.47	7
glmer(wrong_node ~ node_age + sim_type + (1| node_type) , family = binomial)	Simulation type and whether the nodes represent budding or symmetric speciation	30267.49	6
glmer(wrong_node ~ node_age + sim_type + (1| descendant_history) , family = binomial)	Simulation type and the history of clades descendant from budding events	30160.42	6
glm(wrong_node ~ node_age + sim_type , family = binomial)	Simulation type	30323.25	5
glm(wrong_node ~ node_age , family = binomial)	No effects	32690.81	2

When we relaxed rates of trait evolution for continuous traits, we were able to significantly improve the adequacy of the Brownian motion model in the presence of budding, meaning that rate varying models might already be capturing, at least partially, the changes in the mode of trait evolution. In general, models that allow the tempo and mode of trait evolution to change across the tree, thus letting lineages differ from one another, are likely to be more adequate in the presence of budding than approaches which assume homogeneous rates across the tree. Distinct from other scenarios, in which some predictor trait is responsible for rate shifts, budding speciation is expected to affect trait evolution dependent on the mode of speciation. This introduces a complication because we need to reconstruct the budding history of lineages, which does not easily leave a trace on extant species (e.g., it needs samples of multiple populations of recently diverged lineages). Budding history can be recovered either from a detailed fossil record, which is absent for the majority of groups, or from molecular phylogenies, but the latter is restricted to recent events of asymmetric speciation. One potential solution is to use data augmentation (e.g., Quintero and Landis, 2020) to co-estimate budding history and trait evolution model parameters using simulations. This approach could be challenging because both the frequency of budding speciation as well as the location of the progenitor lineages would need to be sampled. It is not clear, however, how much signal for a specific budding history there would be in a molecular phylogeny, so the utility of explicitly simulating budding histories remains to be explored. Nevertheless, our results show that budding is likely one of the factors contributing to heterogeneity in trait evolution and, in the absence of a more specific approach, comparative models that allow for the tempo and mode of evolution to vary across the tree are the recommended approach.

A challenge in the incorporation of budding speciation is that progenitor lineages can produce scenarios incongruent with the most parsimonious history for a trait. On the one hand, statistical models of trait evolution are the most used approaches to understand how traits have changed in deep time, many of which incorporate fairly complex modes of trait evolution. On the other hand, these models are not bound to match the history of trait evolution we intuitively expect, especially when our expectation is highly congruent with a parsimony paradigm. Stochastic character mapping, for instance, is an essential tool to visualize the realization of trait changes along the branches of the tree under fitted models. However, there is a general sense that

590 mappings with multiple substitutions along the same branch are less likely, even though all mappings are fully supported by the chosen model. For instance, it is commonplace for multiple
592 trait changes along a branch to be regarded as "statistical artifacts", because a single shift, which is expected under parsimony, is closer to our *a priori* expectation. This creates a paradox where
594 we adopt statistical models to incorporate a potentially complex macroevolutionary history but we apply *a posteriori* "parsimony checks" to evaluate the verisimilitude of estimates. Budding
596 speciation, as represented here, would not pass on a parsimony check.

The idea that long-lived progenitor lineages can carry the state of the ancestral through
598 multiple speciation events and produce a clade in which the ancestral state is rare goes against the parsimony criteria (since it requires multiple additional steps). Yet, budding speciation can be
600 supported by biologically informed patterns and processes we presented in this study. When we intuitively imagine a parsimonious trait history, we are doing so independently of what is known
602 about the evolutionary history of the system. Are trait reconstructions parsimonious if they require the adoption of more complex explanations for the biological knowledge available for the
604 system? If a most parsimonious reconstruction challenges what is known about the evolution of development, anatomy and physiology of a group, can we consider it parsimonious? Why has the
606 paradigm of parsimony in phylogenetics grown roots so strong it can challenge evidence from other disciplines? These are open questions. Reflecting on the role of comparative approaches
608 and recognizing their limitations, especially when testing scenarios of complex trait evolution, is key to the development of alternative models that help the study of macroevolution to become a
610 more integrative endeavor. The incorporation of budding speciation is one example of the direction we can take improving comparative studies.

612 **Closing remarks**

Budding speciation might indirectly or directly impact both estimates of lineage
614 diversification and trait evolution under phylogenetic comparative models widely used in the literature. In both cases, investigation of budding speciation has not taken place until fairly
616 recently, decades after these methods have been consolidated in the study of macroevolution by neontologists. This fact is even more striking if we contrast with the paleontological literature,
618 which has employed the concept of budding, as well as progenitor lineages, consistently through

many generations. It is true that molecular phylogenies only rarely provide evidence for budding events. However, comparative studies have been trying to incorporate macroevolutionary processes into models even when there is no direct evidence from molecular phylogenies, such as hidden rates of trait evolution, ancient shifts in diversification, or past events of introgression. Our results suggest that it might be possible, although challenging, to incorporate budding speciation into the comparative models of trait evolution for both discrete and continuous traits. The introduction of budding speciation in comparative approaches, however, depends on the departure from the parsimony paradigm which we suspect is a barrier for the development of macroevolutionary models that can fully integrate external biological information about trait evolution. Finally, we hope our contribution helps bring to attention the role of budding speciation in understanding patterns of trait evolution and to strengthen the bridge between paleontological and neontological evolutionary studies.

References

Anacker, B. L., and S. Y. Strauss. 2014. The geography and ecology of plant speciation: Range overlap and niche divergence in sister species. *Proc. R. Soc. B.* 281:20132980.

Aze, T., T. H. G. Ezard, A. Purvis, H. K. Coxall, D. R. M. Stewart, B. S. Wade, and P. N. Pearson. 2013. Identifying anagenesis and cladogenesis in the fossil record. *PNAS* 110:E2946–E2946.

Baldwin, B. G. 2005. Origin of the serpentine-endemic herb *Layia discoidea* from the widespread *L. glandulosa* (Compositae). *Evolution* 59:2473–2479.

Bapst, D. W. 2013. When can clades be potentially resolved with morphology? *PLOS ONE* 8:e62312.

Bastide, P., C. Solís-Lemus, R. Kriebel, K. W. Sparks, and C. Ané. 2018. Phylogenetic comparative methods on phylogenetic networks with reticulations. *Syst. Bio.* 67:800-820.

Baum, D. A., S. D. Smith, and S. S. S. Donovan. 2005. The tree-thinking challenge. *Science* 310:979–980.

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

Benton, M. J., and P. N. Pearson. 2001. Speciation in the fossil record. *Trends Ecol. Evol.* 16:405–411.

Boyko J.D., Beaulieu J.M. 2021. Generalized hidden markov models for phylogenetic comparative datasets. *Methods Ecol. Evol.* 12:468–478.

Caetano, D. S., B. C. O'Meara, and J. M. Beaulieu. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution* 72:2308–2324.

Carrillo, J. D., S. Faurby, D. Silvestro, A. Zizka, C. Jaramillo, C. D. Bacon, and A. Antonelli. 2020. Disproportionate extinction of South American mammals drove the asymmetry of the Great American Biotic Interchange. *PNAS* 117:26281–26287.

Clavel, J., and H. Morlon. 2017. Accelerated body size evolution during cold climatic periods in the Cenozoic. *PNAS* 114:4183–4188.

Cooper, N., G. H. Thomas, C. Venditti, A. Meade, and R. P. Freckleton. 2016. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biol. J. Linn. Soc.* 118:64–77.

Collin, R., and M. P. Miglietta. 2008. Reversing opinions on Dollo's Law. *Trends Ecol. Evol.* 23:602–609.

Crawford, D. J. 2010. Progenitor-derivative species pairs and plant speciation. *Taxon* 59:1413–1423.

Crouch, N. M. A., S. M. Edie, K. S. Collins, R. Bieler, and D. Jablonski. 2021. Calibrating phylogenies assuming bifurcation or budding alters inferred macroevolutionary dynamics in a densely sampled phylogeny of bivalve families. *Proc. R. Soc. B.* 288:20212178.

De Lisle, S. P., D. Punzalan, N. Rollinson, and L. Rowe. 2021. Extinction and the temporal distribution of macroevolutionary bursts. *J. Evol. Biol.* 34:380–390.

Duchen, P., M. L. Alfaro, J. Rolland, N. Salamin, and D. Silvestro. 2021. On the effect of asymmetrical trait inheritance on models of trait evolution. *Syst. Bio.* 70:376–388.

Eastman, J. M., M. E. Alfaro, P. Joyce, A. L. Hipp, and L. J. Harmon. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution* 65:3578–3589.

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

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17:368–376.

Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.

Foote, M. 1996. On the probability of ancestors in the fossil record. *Paleobiology* 22:141–151.

Foote, M., J. S. Crampton, A. G. Beu, B. A. Marshall, R. A. Cooper, P. A. Maxwell, and I. Matcham. 2007. Rise and fall of species occupancy in Cenozoic fossil mollusks. *Science* 318:1131–1134.

Funk, D. J., and K. E. Omland. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.* 34:397–423.

Gillespie, R. 2004. Community assembly through adaptive radiation in hawaiian spiders. *Science* 303:356–359.

Gottlieb, L. D. 2004. Rethinking classic examples of recent speciation in plants. *New Phytol.* 161:71–82.

Gould, N. E. S. J., and N. Eldredge. 1972. Punctuated equilibria: An alternative to phyletic gradualism. *Models in Paleobiology*, 82–115, edited by Schopf, TJM Freeman, Cooper & Co, San Francisco.

Goldberg, E. E., and J. Foo. 2020. Memory in trait macroevolution. *Am. Nat.* 195:300–314.

Griffith, O. W., D. G. Blackburn, M. C. Brandley, J. U. Van Dyke, C. M. Whittington, and M. B. Thompson. 2015. Ancestral state reconstructions require biological evidence to test evolutionary hypotheses: A case study examining the evolution of reproductive mode in squamate reptiles. *J. Exp. Zool. B. Mol. Dev.* Evol. 324:493–503.

Hagen, O., K. Hartmann, M. Steel, and T. Stadler. 2015. Age-dependent speciation can explain the shape of empirical phylogenies. *Syst. Bio.* 64:432–440.

Hennig, W. 1950. Grundzüge einer theorie der phylogenetischen systematik. Berlin: Aufbau.

Hunt, G., M. A. Bell, and M. P. Travis. 2008. Evolution toward a new adaptive optimum: Phenotypic evolution in a fossil stickleback lineage. *Evolution* 62:700–710.

Januario, M., and T. B. Quental. 2021. Re-evaluation of the "law of constant extinction" for ruminants at different taxonomical scales. *Evolution* 75:656–671.

Liow L.H. & Stenseth N.C. 2007 The rise and fall of species: Implications for macroevolutionary and macroecological studies. *Proc. R. Soc. B.* 274:2745–2752.

Mahler, D. L., T. Ingram, L. J. Revell, and J. B. Losos. 2013. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science* 341:292–295.

Mallet, J., N. Besansky, and M. W. Hahn. 2016. How reticulated are species? *BioEssays* 38:140–149.

Meier, R., and R. Willmann. 2000. The Hennigian species concept. In Wheeler, Q. D., and R. Meier (eds). *Species concepts and phylogenetic theory: A debate* (pp. 30–43). Columbia University Press, New York.

Nosil, P. 2012. *Ecological Speciation*, Oxford: Oxford University Press, p. 280, ISBN 978–0199587117

Nuismer, S. L., and L. J. Harmon. 2015. Predicting rates of interspecific interaction from phylogenetic trees. *Ecol. Lett.* 18:17–27.

O'Meara, B. C. 2012. Evolutionary inferences from phylogenies: A review of methods. *Annu. Rev. Ecol. Evol. Syst.* 43:267–285.

Otero, A., P. Vargas, V. Valcárcel, M. Fernández-Mazuecos, P. Jiménez-Mejías, V. Valcárcel, I. Villa-Machío, and A. L. Hipp. 2022. A snapshot of progenitor-derivative speciation in *Iberodes* (Boraginaceae). *Mol. Ecol.* In press (https://doi.org/10.1111/mec.16459)

Papuga, G., P. Gauthier, V. Pons, E. Farris, and J. D. Thompson. 2018. Ecological niche differentiation in peripheral populations: A comparative analysis of eleven Mediterranean plant species. *Ecography* 41:1650–1664.

Parins-Fukuchi, C. 2021. Morphological and phylogeographic evidence for budding speciation: an example in hominins. *Biol. Lett.* 17:20200754.

Patsis, A., R. P. Overson, K. A. Skogen, N. J. Wickett, M. G. Johnson, W. L. Wagner, R. A. Raguso, J. B. Fant, and R. A. Levin. 2021. Elucidating the evolutionary history of oenothera sect. *Pachylophus* (Onagraceae): A phylogenomic approach. *Syst. Bot.* 46:799–811.

Pennell, M. W., and L. J. Harmon. 2013. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. *Ann. N. Y. Acad. Sci.* 1289:90–105.

Pennell, M. W., L. J. Harmon, and J. C. Uyeda. 2014. Is there room for punctuated equilibrium in macroevolution? *Trends Ecol. Evol.* 29:23–32.

Pennell, M. W., R. G. FitzJohn, W. K. Cornwell, and L. J. Harmon. 2015. Model adequacy and the macroevolution of angiosperm functional traits. *Am. Nat.* 186:E33–E50.

Philippe, H., and C. J. Douady. 2003. Horizontal gene transfer and phylogenetics. *Curr. Opin. Microbiol.* 6:498–505.

Pyron, R. A. 2015. Advancing perspectives on parity-mode evolution. *J. Exp. Zool. B Mol. Dev. Evol.* 324:562–563.

Quintero, I., and M. J. Landis. 2020. Interdependent phenotypic and biogeographic evolution driven by biotic interactions. *Syst. Biol.* 69:739–755.

Rabosky, D. L., and E. E. Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst. Biol.* 64(2):340–355.

Raup, D. M. 1985. Mathematical models of cladogenesis. *Paleobiology* 11:42–52.

Rosenblum, E. B., B. A. J. Sarver, J. W. Brown, S. Des Roches, K. M. Hardwick, T. D. Hether, J. M. Eastman, M. W. Pennell, and L. J. Harmon. 2012. Goldilocks meets Santa Rosalia: An ephemeral speciation model explains patterns of diversification across time scales. *Evol. Biol.* 39:255–261.

Rundle, H. D., L. Nagel, J. W. Boughman, and D. Schluter. 2000. Natural selection and parallel speciation in sympatric sticklebacks. *Science* 287:306–308.

Sankoff, D. 1975. Minimal mutation trees of sequences. *SIAM J. Appl. Math.* 28:35–42.

Schultz T.R., Cocroft R.B., Churchill G.A. 1996. The reconstruction of ancestral character states. *Evolution*. 50:504–511.

Simpson, G. G. 1961. *Principles of animal taxonomy*. New York: Columbia University Press.

Slater, G. J., L. J. Harmon, and M. E. Alfaro. 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution* 66:3931–3944.

Uyeda, J. C., and L. J. Harmon. 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Syst. Bio.* 63:902–918.

Willmann, R. 1986. Reproductive isolation and the limits of the species in time. *Cladistics* 2:356–358

Supplementary Information for:

How important is budding speciation for comparative studies?

Daniel S. Caetano[1,2] and Tiago Bosisio Quental[2]
[1] *Department of Biological Sciences, Towson University, Towson, MD, 21252, USA*
[2] *Department of Ecology, University of São Paulo, São Paulo, SP, 05508-090, Brazil*

TABLE OF CONTENTS

List of Figures

Figure S1: Effect of lineage-age dependent trait evolution and frequency of budding speciation on expected variance (S_{var}).
Figure S2: Effect of lineage-age dependent trait evolution and frequency of budding speciation on node depth (S_{hgt}).
Figure S3: Effect of lineage-age dependent trait evolution and frequency of budding speciation on absolute contrasts (C_{var}).
Figure S4: Phylogenetic half-life in function of budding frequency and strength of the lineage-age dependent slowdown factor.
Figure S5: Example of budding simulation on phylogenetic tree including fossil lineages.
Figure S6: Model expectations for the interaction between age-dependent trait evolution and events of symmetric or budding speciation.
Figure S7: Rate slowdown in function of the exponential decay rate parameter.

List of Sections

Section 1: Simulating budding speciation
Section 2: Lineage-age dependent models and branch length scaling
Section 3: Cladogenetic changes associated with budding speciation
Section 4: Simulation scenarios and parameter values
Section 5: R Markdown with simulation example
References

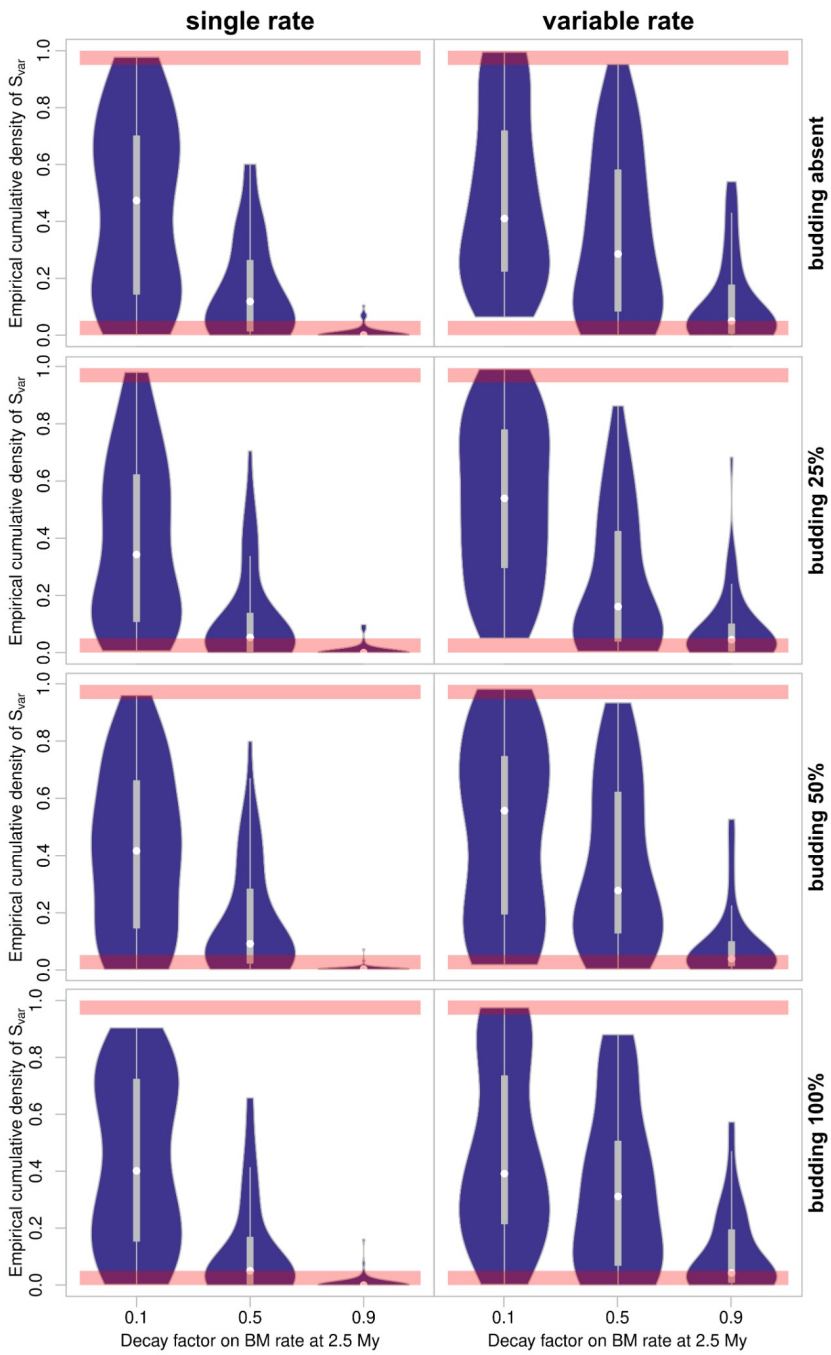


Figure S1: Effect of lineage-age dependent trait evolution and frequency of budding speciation on estimates of a Brownian Motion with homogeneous rate (left column) and a variable rates BM model (right column). Plots show the distribution of the observed slope of absolute contrasts in function of their expected variance (S_{var}) across 50 simulation replicates. Empirical cumulative densities were computed with respect to a null distribution (Pennel *et al*., 2015). Concentration towards extreme values (regions highlighted in salmon) is evidence of model inadequacy. The x axes show the slowdown factor ($\boldsymbol{\sigma}_{2.5}^2 = \boldsymbol{\sigma}_0^2 - \boldsymbol{\sigma}_0^2 *$ decay factor). Small values of S_{var} indicate slowdown causes a concentration of trait changes in short branches.

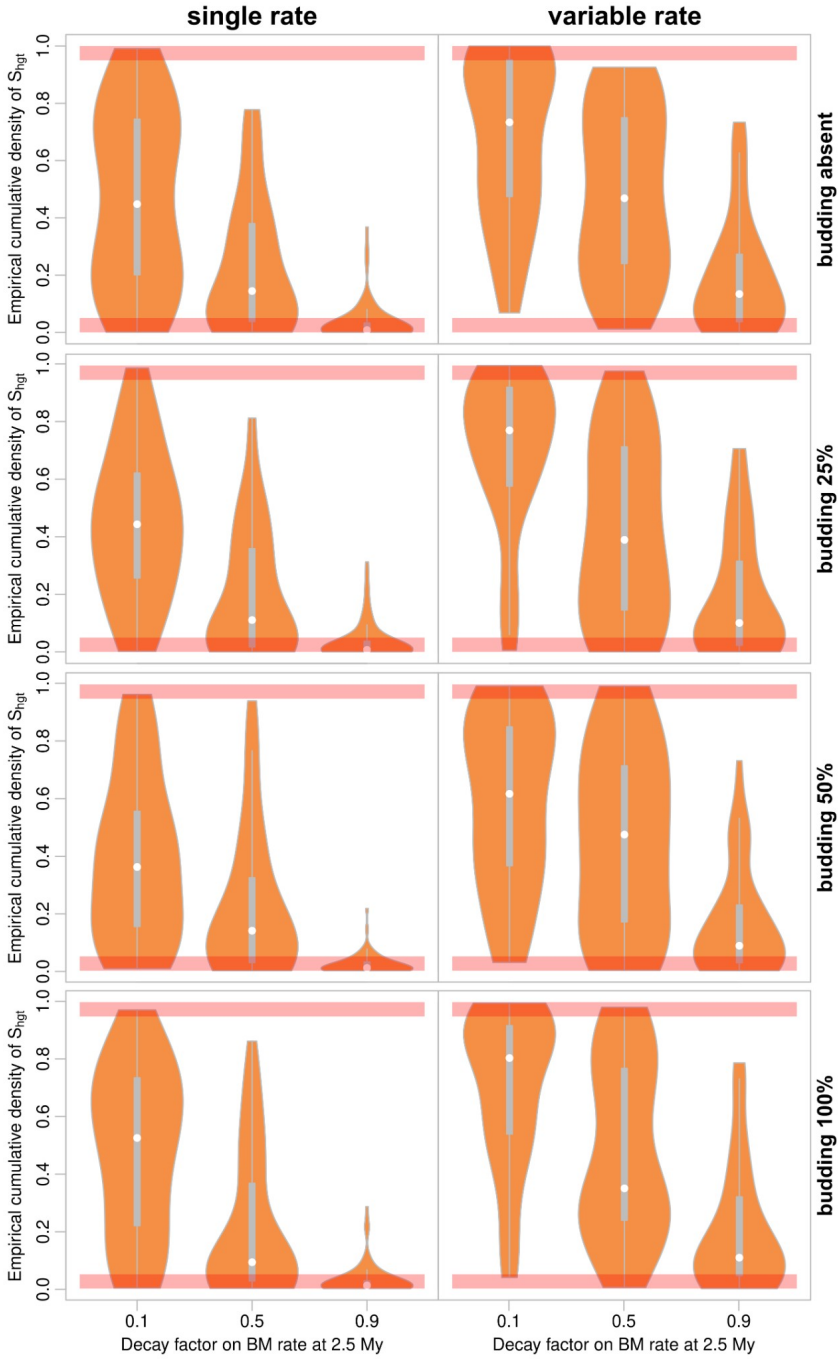


Figure S2: Effect of lineage-age dependent trait evolution and frequency of budding speciation on estimates of a Brownian Motion with homogeneous rate (left column) and a variable rates BM model (right column). Plots show the distribution of the slope of absolute contrasts in function of node depth (S_{hgt}) across 50 simulation replicates. Empirical cumulative densities were computed with respect to a null distribution (Pennel *et al*., 2015). Concentration towards extreme values (regions highlighted in salmon) is evidence of model inadequacy. The x axes show the slowdown factor ($\boldsymbol{\sigma}_{2.5}^2 = \boldsymbol{\sigma}_0^2 - \boldsymbol{\sigma}_0^2 *$ decay factor). Small values represent negative slopes and indicate that more trait change is estimated to have happened at younger nodes. When allowing rates to vary, this relationship fades away.

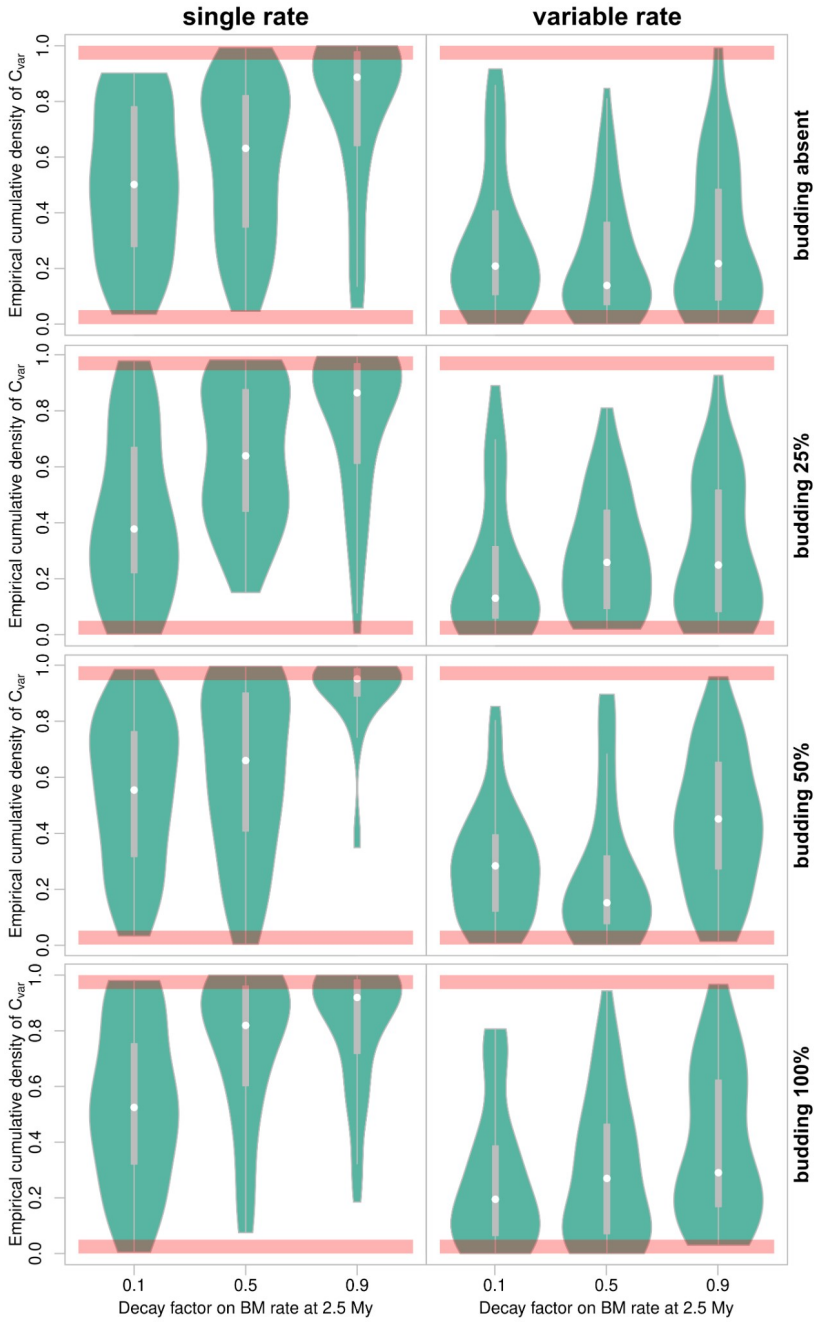


Figure S3: Effect of lineage-age dependent trait evolution and frequency of budding speciation on estimates of a Brownian Motion with homogeneous rate (left column) and a variable rates BM model (right column). Plots show the coefficient of variation of absolute contrasts (C_{var}) across 50 simulation replicates. Empirical cumulative densities were computed with respect to a null distribution (Pennel *et al*. 2015). Concentration towards extreme values (regions highlighted in salmon) is evidence of model inadequacy. The x axes show the slowdown factor ($\boldsymbol{\sigma}_{2.5}^2 = \boldsymbol{\sigma}_0^2 - \boldsymbol{\sigma}_0^2 *$ decay factor). High values indicate the model is not properly accounting for rate variation.

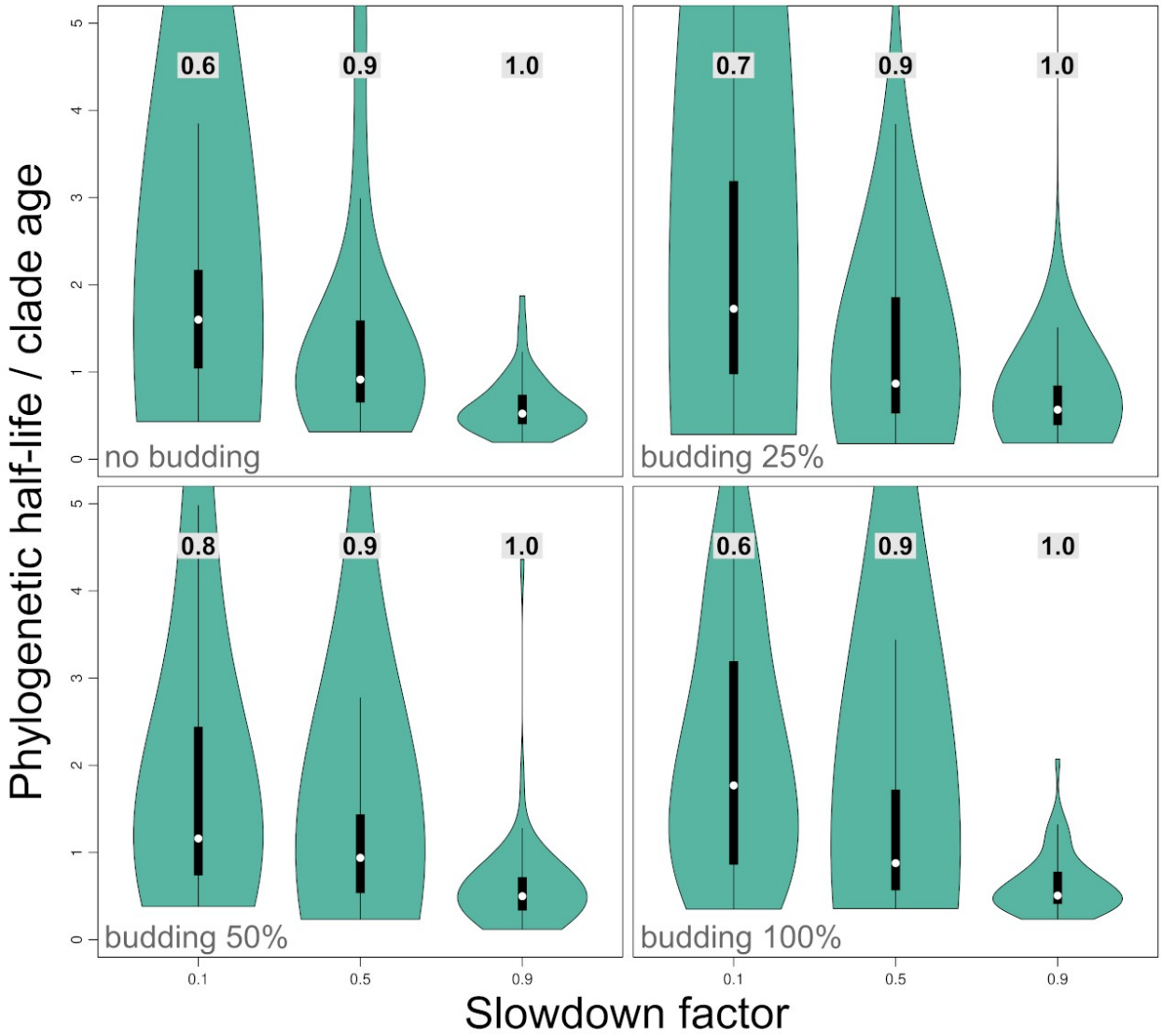


Figure S4: Distribution of phylogenetic half-life estimated from Ornstein-Uhlenbeck models fitted to multiple simulation replicates with increasing budding frequency and strength of the lineage-age dependent slowdown factor. Phylogenetic half-life was rescaled to represent the proportion of the clade age, thus a value of 1 indicates a half-life equal to the age of the tree (40 My). Values > 1 represent very weak pull towards the optima, with half-lifes multiple times the age of the clade indicating a process indistinguishable from unbounded trait evolution. Numbers in gray boxes show the proportion of the total pool of replicates that favored OU models.

Section 1: Simulating budding speciation

We simulate budding speciation by assigning a binary variable to each node of a phylogeny which has value equal to 1 when the node represents a divergence through budding (asymmetric) speciation and 0 when the node is a symmetric (bifurcation) event. We used a binary distribution to control the frequency of budding speciation. The probability of budding is independent for each node of the tree, so long-lived progenitor lineages are the result of random events rather than autocorrelation in the probability of budding speciation after a previous budding event happened.

Budding speciation does not affect the simulation of lineages under a birth-death model. Rates of birth and death remain the same regardless of the mode of speciation (if symmetric or budding). As a result, here we simulate phylogenetic trees independent of budding and create budding scenarios a posteriori by assigning each node of the tree as a budding event or not. Because budding histories are created using fixed trees, here we make all 50 simulation replicates across the scenarios pairwise, meaning that a single tree can be used to create budding scenarios with varying characteristics. The advantage of this approach is that we can compare scenarios of budding speciation and trait evolution while keeping the phylogeny constant.

When a budding speciation event happens, one of the descendant lineages is a daughter lineage and the other is the progenitor lineage. We randomly selected which of the lineages are the progenitor (either the branch on the left or on the right of a node). The progenitor lineage is the continuation of the ancestral lineage which survives the speciation event and inherits the age of the ancestor whereas the daughter lineage represents a newly formed species with age of 0 time units. The progenitor lineage can generate another daughter lineage if the descendant node is also defined as a budding speciation event by the simulation. When that happens, then the progenitor lineage branch is the continuation of the ancestral branch, creating a long-lived continuous progenitor lineage that can span multiple nodes of the tree. Figure S5 shows one example of a simulation with colored progenitor lineages. Multiple adjacent branches with the same color represent the continuation of the same progenitor lineage through multiple events of budding speciation.

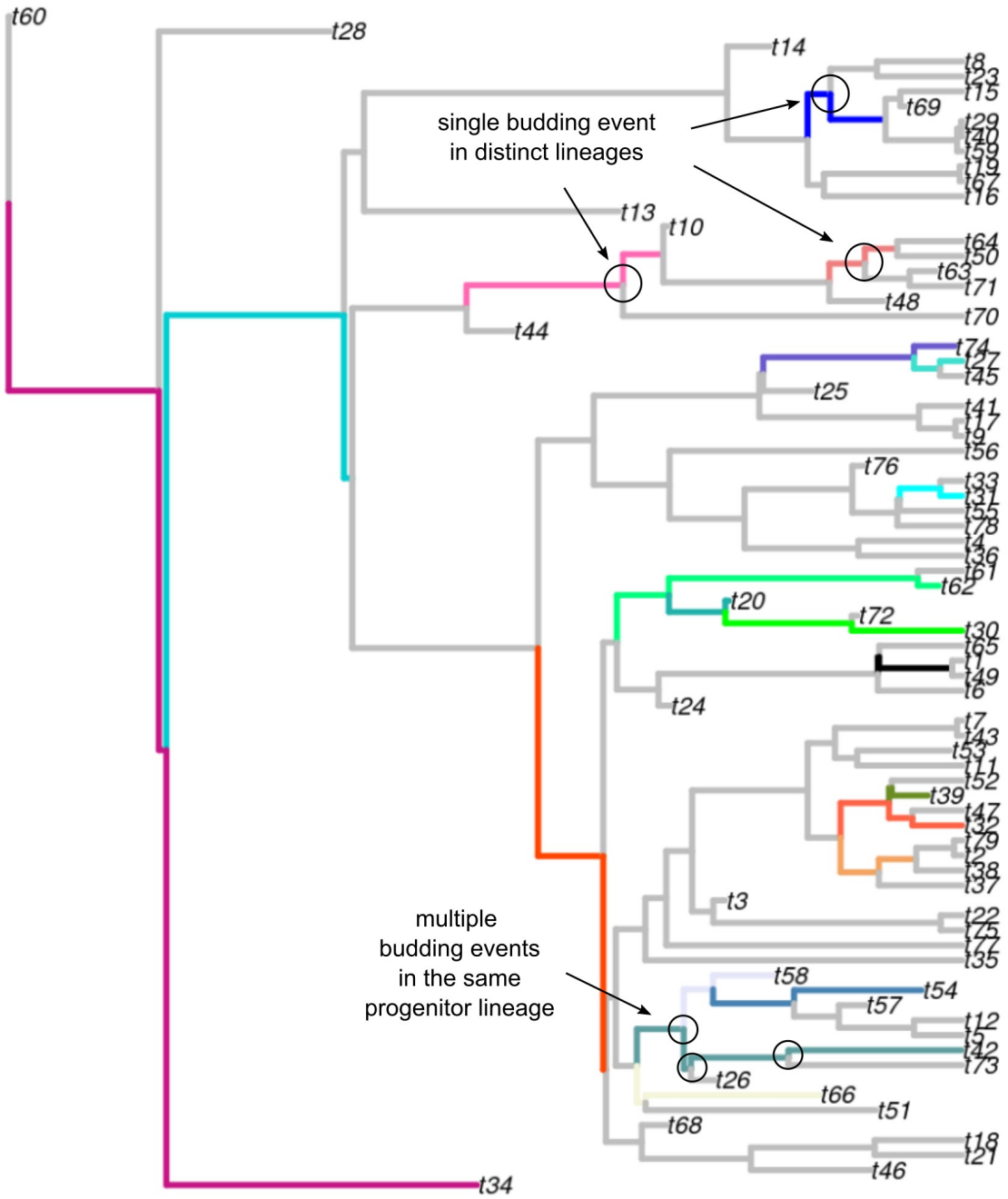


Figure S5: Example of budding simulation on phylogenetic tree including fossil lineages. Colored branches represent progenitor lineages. Each progenitor lineage has a unique color (assigned randomly). Multiple adjacent branches with the same color indicate a progenitor lineage that generated multiple daughters through budding.

Section 2: Simulation of lineage-age dependent trait evolution

Here we implemented lineage-age dependent models of trait evolution. This process simulates the scenario in which lineages are more likely to show rapid trait change earlier after speciation and then gradually reduce the pace of trait evolution as they approach a stable evolutionary optima. For example, when an ancestral lineage survives a speciation event, one of the descendants of the internal node (i.e., the new lineage) will be assigned an age of 0 MY (Million years) whereas the other (i.e., the ancestral or mother lineage) will inherit the age of the ancestor (Figure S6). Put simply, the rate of trait evolution is a function of lineage age—as lineage age increases the rate of trait evolution decreases (Figure S6). This relationship is not restricted to be linear and more complex functions can also be implemented.

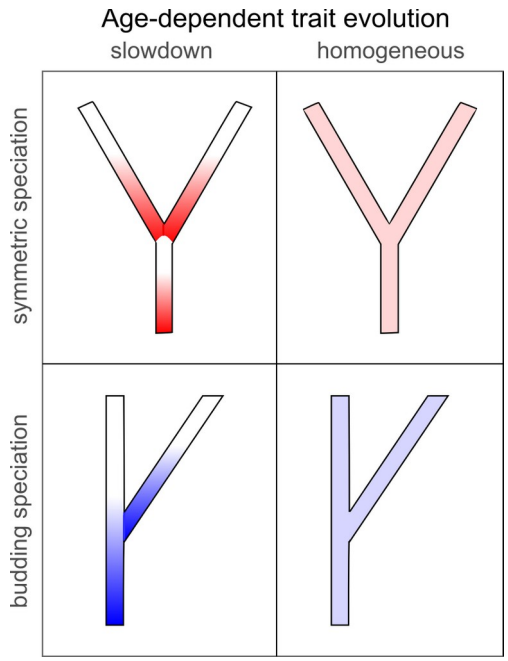


Figure S6: Model expectations for the interaction between age-dependent trait evolution and events of symmetric or budding speciation. Color gradients show the pace of trait evolution. Stronger colors represent faster rates. Top row: After symmetric speciation events two new lineages are generated and the ancestral lineage is dissolved (i.e., becomes extinct). Bottom row: Ancestral lineages survive budding speciation events and create asymmetry in lineage-age.

We implemented lineage-age dependent rate of trait evolution by rescaling the branches of the tree. Faster rates of trait change are emulated by extending the branch whereas shortening branches reduce the expected amount of trait change. We rescaled branches based on the average lineage-age at a very small time interval, computed with the age of the lineage at the start and

end points of the time interval. Then, we modeled the slowdown in rates of trait evolution in function of lineage-age with an exponential distribution. We controlled the mean of the exponential distribution in order to create scenarios with varying levels of rate slowdown, from patterns that are indistinguishable from homogeneous rate to strong concentration of rates after the birth of new lineages, similar to punctuated equilibrium. This approach allows for the use of most univariate models of trait evolution available for both discrete and continuous traits without any modification to existing software packages.

Section 3: Cladogenetic changes associated with budding speciation

We implemented cladogenetic models of trait evolution for discrete traits associated with budding speciation. Cladogenetic changes emulates the scenario of *quantum speciation* (Gottlieb, 2003 and references therein) in which a new species is formed through budding and undergoes rapid trait change whereas the progenitor species remains relatively unchanged. Rapid trait change of newly formed lineages can be a reflection of competition, because only lineages that are distinct enough in key traits to explore a new adaptive optima might survive enough time to be sampled in the fossil record or be still alive today. We modeled this scenario by assigning a binary variable to each budding node to control whether there is a cladogenetic change associated with the birth of the new species.

Note that cladogenetic changes are not present in every budding node in the simulation. Adding a cladogenetic change to every budding node is problematic due to two main reasons 1) it introduces a large number of "jumps" in trait space and in many cases erases the phylogenetic history in the trait and 2) the simulation reflects the evolutionary history of a single trait and not the entire phenome of the organism. It is possible, for example, that at least one trait in the organism does show a cladogenetic change associated with a budding speciation event, as predicted by the *quantum speciation* framework. However, our simulations represent a small sample of an organism's entire poll of traits, and not their complete set of phenotypes. Thus, even if every budding speciation event is associated with some cladogenetic change, we can, and should, control the frequency of cladogenetic changes along the tree as a representation of the particular trait under study.

We performed all simulations using 50% of cladogenetic change probability at every budding node of the tree. Meaning that only half of the budding events are associated with a

cladogenetic change in the particular trait being simulated. We represented a cladogenetic change as a random draw from the potential states for the (discrete) trait, excluding the current trait at that point of the simulation (i.e., a jump to another state). A random draw is adequate in this case because all discrete trait simulations used symmetric transition matrices (Q), but it is possible to weigh transitions based on asymmetric transition matrices under more complex scenarios. The new lineage was allowed to change immediately after the cladogenetic event. The cladogenetic event happens independent of time, at the node, thus, depending on the rates of trait evolution following a budding event, there is a chance the trait changes soon after the cladogenetic "jump". We did not control for such cases in our simulation scenarios.

Section 4: Simulation scenarios and parameter values

Our simulation study can be broadly divided into continuous trait simulations and discrete trait simulations. For continuous traits, we simulated a single continuous trait following a Brownian Motion model of trait evolution with base rate (σ^2) of 0.2. For discrete traits, we used a symmetric rate Makov transition matrix for three states

$$Q = \begin{matrix} -0.06 & 0.03 & 0.03 \\ 0.03 & -0.06 & 0.03 \\ 0.03 & 0.03 & -0.06 \end{matrix}$$

. Preliminary tests with binary states suggested that the rapid decay of information for ancestral state reconstruction of older nodes would significantly constrain the window of parameter values that we could explore using simulations. Thus we chose to use three states instead of two. Parameter values for continuous and discrete trait evolution models reported above reflect base rates, or initial rates, of trait evolution. Under lineage-age dependent models of trait evolution these base rates are scaled dependent on the age of the lineage (see Figure S7). Thus, multiple simulation scenarios show varying rates of trait evolution across the branches of the tree (see Figure S6).

In order to implement the changes in rates of trait evolution produced by the lineage-age dependent model, we transformed the branches of the tree instead of changing the rates of trait change directly (following Eastman *et al.*, 2011). It is possible to emulate faster rates of trait evolution by extending the branch such that we have a fixed rate of trait change, but more time

for the trait to evolve. The same can be done to slow down changes since the trait would have less time to change if branch lengths are shortened. We used this approach both for simulation and estimation under lineage-age dependent rates of trait evolution.

Phylogenetic trees were simulated using a birth-death process ($\lambda = 0.2$, $\mu = 0.1$) and fixing the root age to 40 Mya. This set of parameters will produce, in average, 300 extant lineages after 40 million years of evolution when excluding all simulations in which the group went completely extinct. We used rejection sampling to keep only trees between 250 to 350 extant lineages (interval centered on the average diversity expected from the simulation). Each of the scenarios were replicated 50 times with the same poll of 50 phylogenetic trees. We controlled the lineage-age dependent rate of trait evolution using an exponential decay factor. Higher values for the rate of the exponential decay function produces a stronger rate slowdown. We used rate parameters equal to 0.042, 0.279, and 0.925. Figure S7 shows the decay of the initial trait evolution rate in function of lineage age for each of the scenarios tested in this study. Note that, when the lineage is 2.5 My old, the rate parameters produce a reduction to 90%, 50%, and 10% of the initial rate of trait evolution, respectively.

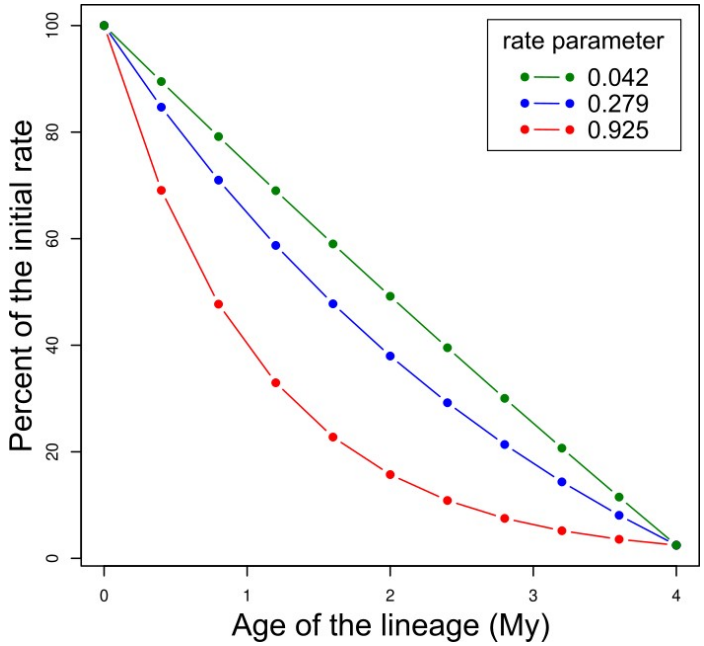


Figure S7: Rate slowdown in function of the exponential decay rate parameter. Stronger rate parameters cause a faster decay in rates of trait evolution. The percent of the initial rate means the percent of the rate of trait evolution when the lineage has age of 0 My. Note that the percent of the initial rate decay as the lineage gets older.

Budding speciation (as detailed in section "*Simulating budding speciation*") was simulated using a probability of asymmetric speciation at each node of the tree. We varied the frequency of budding speciation from absent (0%, our null model), to rare (25%), medium (50%), and a case in which every node represented a budding event (100%). We repeated simulations with every combination of lineage-age dependent rate and budding probability, for the total of 16 distinct scenarios of trait evolution for both continuous and discrete traits. We also repeated each of these 16 scenarios with and without cladogenetic changes associated with budding (see section "*Cladogenetic changes associated with budding speciation*").

References cited in this supplement:

Eastman, J. M., M. E. Alfaro, P. Joyce, A. L. Hipp, and L. J. Harmon. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution* 65:3578–3589.

Pennell, M. W., R. G. FitzJohn, W. K. Cornwell, and L. J. Harmon. 2015. Model adequacy and the macroevolution of angiosperm functional Traits. *Am. Nat.* 186:E33–E50.

R Markdown with simulation example

Daniel Caetano and Tiago Quental

5/16/2022

This Markdown document will show how to install and use the "buddPhy" package for R. This package allows for simulation of multiple scenarios of budding speciation for both discrete and continuous traits. The package also offers functions for plotting the budding history of lineages and the distribution of rates of trait evolution under several scenarios of lineage-age dependent rates, budding and cladogenetic events associated with asymmetric speciation. The package is currently available through github.

Package installation

First install the package dependencies:

```
install.packages("viridisLite")
install.packages("ape")
install.packages("expm")
install.packages("FossilSim")
install.packages("ratematrix")
install.packages("phytools")
```

The package can be installed from the github repository.

```
library( devtools )
install_github(repo = "Caetanods/buddPhy")
```

```
library( buddPhy )
```

Generating a phylogeny

The first step is to obtain a phylogenetic tree. `buddPhy` will trace the evolutionary history of the lineages through the branches of an existent phylogeny, but it will not generate a new one. This is useful because the user can produce lineage histories based on their empirical phylogenetic tree without the need to modify it in anyway.

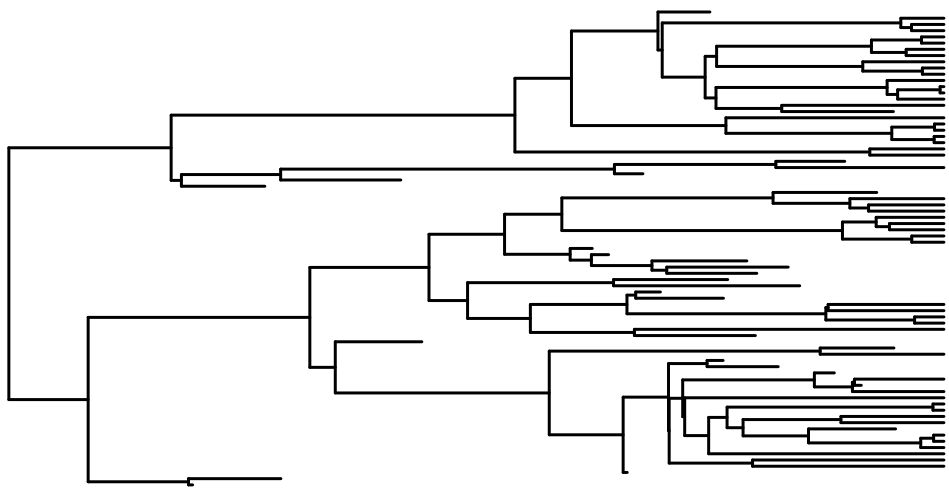
If a phylogenetic tree is not available, then one can be simulated using a variety of packages. Here we are using `TreeSim`, but note that other packages such as `geiger`, `phytools`, and `ape` could be used as well.

```
library( TreeSim )
```

```
## Loading required package: ape
```

```
## Loading required package: geiger
```

```
phy <- sim.bd.taxa(n = 50, numbsim = 1, lambda = 0.02, mu = 0.01, complete = TRUE)[[1]]
plot.phylo(x = phy, show.tip.label = FALSE, edge.width = 1.5)
```



Simulating discrete trait histories

`buddPhy` will generate the history of the lineages together with the history of a discrete or a continuous trait. Let's start by simulating a discrete trait. For this first we need to create a Q matrix with the rates of transition between states:

```
Q <- rbind(c(-0.06, 0.03, 0.03), c(0.03, -0.06, 0.03), c(0.03, 0.03, -0.06))
colnames(Q) <- rownames(Q) <- 1:3
```

We will use this matrix to generate the history for 3 states. Note that we have symmetric evolution. Now we can use the `buddPhy::sim_Mk_budding_exp` to generate the trait history:

```
sim <- sim_Mk_budding_exp(tree = phy, Q = Q, budding_prob = 0.25, change_rate = 0.2
                    , cladogenetic_change = "prob", cladogenetic_prob = 0.5
                    , cladogenetic_state = "other")
```

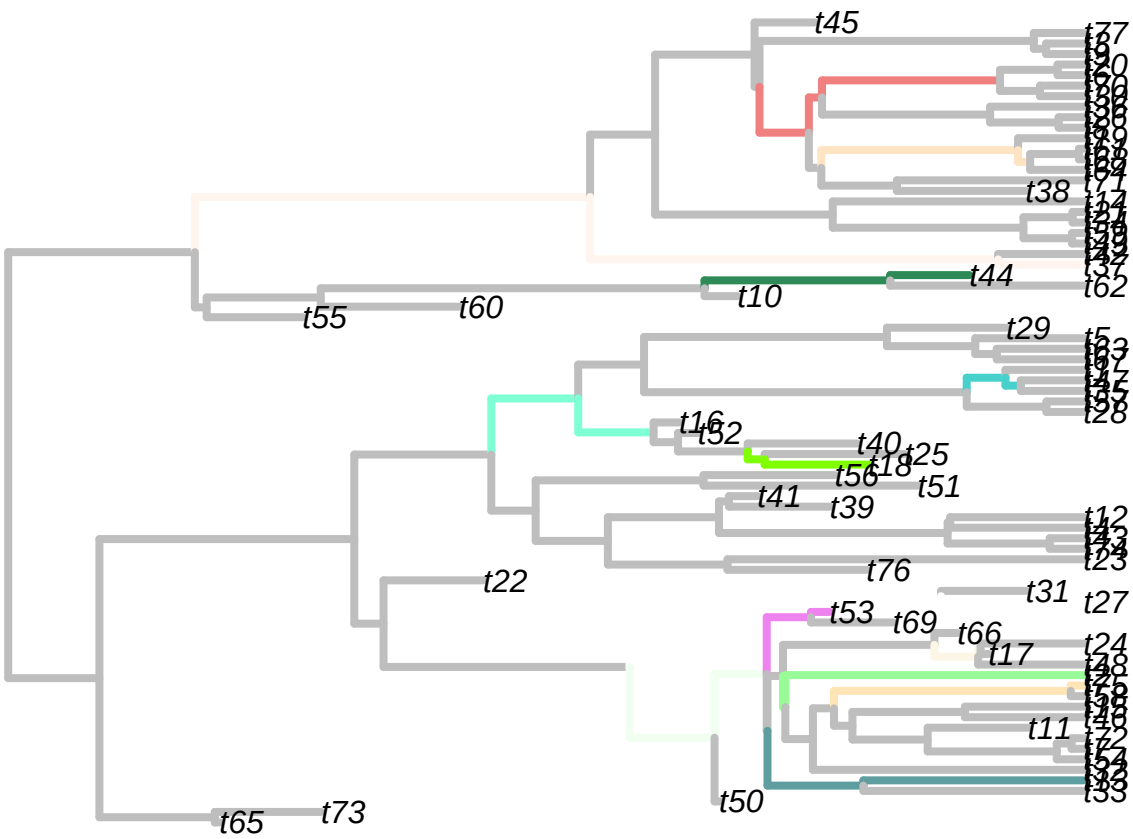
The line of code above is controlling many aspects of the simulation. `budding_prob` controls the probability that a budding event will happen at a given node of the tree. `change_rate` is the mean value of the exponential decay function which reduces rates of trait evolution in function of lineage-age. In this model the older lineages show slower rates of trait evolution than the new lineages. `cladogenetic_change`

is the type of trait change that happens at the event of budding. Here the user can choose between making a cladogenetic change ("prob" or "flat") or nothing special at the nodes (i.e., anagenetic evolution only). `cladogenetic_prob` is the probability that a cladogenetic event will happen when a node is assigned to be associated with a budding speciation event. Setting this parameter to 0.5 means that an event of cladogenetic change will only happen about half of the times that a budding event takes place. Finally, the `cladogenetic_state` parameter controls which state can appear after an event of budding. Some of the options are "other" (a state distinct from the state of the progenitor lineage) and "any" (any state, including the same state of the progenitor lineage).

Please refer to the help page for the `sim_Mk_budding_exp` function for more information about the options.

After performing the simulation we can plot the history of the lineages. Note that lineages with distinct colors are different. Colored branches indicate progenitor lineages which produced one or more daughter lineages through budding. The collection of colors are randomized.

```
plot_mother_lineages(sim_obj = sim, background_color = "gray", edge_width = 4)
```



We can also summary the history of the internal nodes and indicate which of the nodes were budding speciation events:

```
budd_hist <- buddPhy::get_Budding_History(sim = sim)
## The list of budding nodes:
budd_hist$budd_nodes
```

```
##  [1]  98 118 149 109 130 111 124  88 105 136 120  89 134 137 110 114
```

```r
## A logic vector that can be used to perform further analyses:
budd_hist$budd_edges
```

```
##   [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE  TRUE
##  [13] FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [25] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE  TRUE
##  [37] FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE FALSE
##  [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE
##  [73]  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
##  [85]  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [97] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [121] FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE
## [133] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

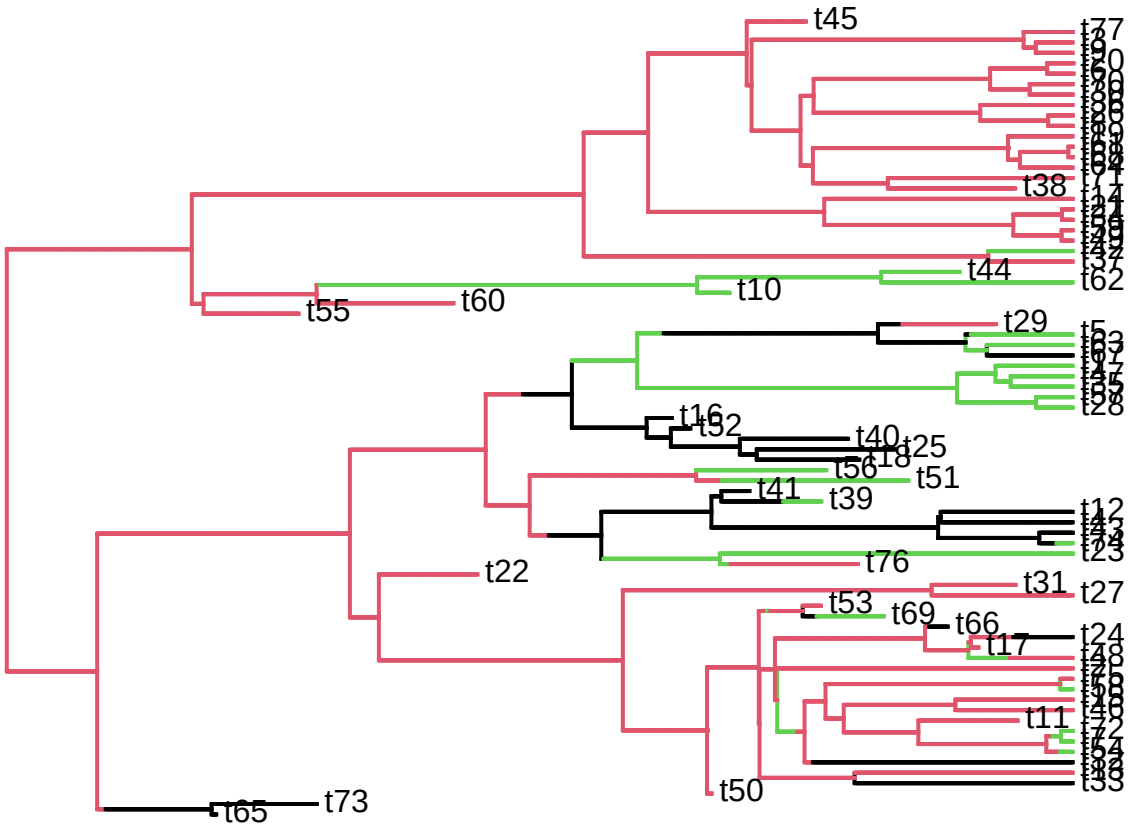
We can also plot the history of the traits. This is similar to a stochastic map, but here we are showing the true history of the trait and not an estimate of ancestral states.

```r
library( phytools )
```

```
## Loading required package: maps
```

```r
## The sim object is a list with many elements, including a stochastic map
##      with the true history of the simulation.
plotSimmap(tree = sim$simmap)
```

```
## no colors provided. using the following legend:
##         1         2         3
##   "black" "#DF536B" "#61D04F"
```



Of course, we can also extract the states for each tip of the tree:

```
sim$tip_state
```

```
##  t5 t21  t4 t72 t67 t54 t61 t19 t26 t24 t58  t2 t20 t37 t13 t34  t1 t35 t71 t14
##   3   2   1   3   1   3   2   2   2   1   3   2   2   2   2   2   3   3   2   2
## t23 t33 t77 t43 t36 t70 t57  t7 t46 t63  t9 t64 t47 t68 t59 t48 t15 t12 t49 t28
##   3   1   2   1   2   2   3   3   2   3   2   2   3   2   2   2   2   1   2   3
## t74 t27 t75  t3  t6 t30  t8 t62 t42 t32 t11 t31 t38 t29 t17 t44 t66 t51 t25 t69
##   3   2   2   2   2   2   2   3   3   1   2   2   2   2   2   3   1   3   1   3
## t18 t76 t40 t56 t39 t53 t45 t41 t10 t50 t52 t16 t22 t60 t73 t55 t65
##   1   2   1   3   3   2   2   1   3   2   1   1   2   2   1   2   1
## Levels: 1 2 3
```

Plotting the rates of trait evolution

In `buddPhy`, the age of the lineages influence the rate of trait evolution. Thus, rates of trait evolution vary across the tree. This is not a rate-homogeneous model of trait evolution. The distribution of rates will vary in function of the exponential decay factor and the probabilities of budding speciation.

We can easily plot the variation in the rates of trait evolution and also extract this information to make further analyses.

```
## The function will produce a phylogenetic tree in a simmap format and also
##      information for the colors.
```

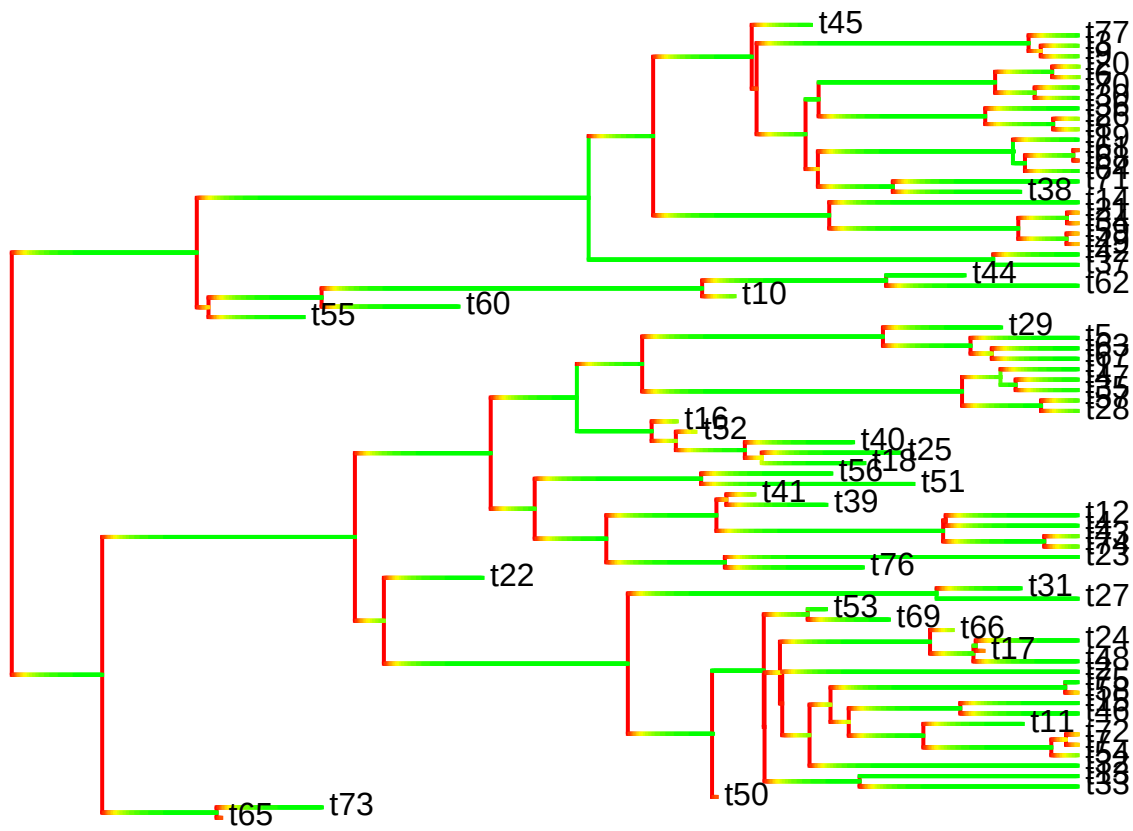
```
rate_tree <- make_scaler_simmap(sims = sim, ncat = 20)
class( rate_tree$simmap )
```

```
## [1] "phylo"  "simmap"
```

```
head( rate_tree$col_vec )
```

```
##         A         B         C         D         E         F
## "#00FF00" "#1AFF00" "#35FF00" "#50FF00" "#6BFF00" "#86FF00"
```

```
## Then we can use the plotSimmap function to produce the desired figure:
plotSimmap(tree = rate_tree$simmap, colors = rate_tree$col_vec)
```



In the figure above red colors indicate the fastest rate, which is equal to the original Q matrix used for the simulation. Then colors approaching green are slower rates of trait evolution. The decay function works by reducing the original rate of trait evolution by a factor.

We can check the actual values for the slowdown factor. These are also produced by the same function used above and stored as an element of the list:

```
rate_tree$legend_matrix
```

```
##          from         to label     col
## 1  0.00000000 0.04909519     A #00FF00
```

```
## 2  0.04909519 0.09819039     B #1AFF00
## 3  0.09819039 0.14728558     C #35FF00
## 4  0.14728558 0.19638078     D #50FF00
## 5  0.19638078 0.24547597     E #6BFF00
## 6  0.24547597 0.29457117     F #86FF00
## 7  0.29457117 0.34366636     G #A1FF00
## 8  0.34366636 0.39276156     H #BBFF00
## 9  0.39276156 0.44185675     I #D6FF00
## 10 0.44185675 0.49095195     J #F1FF00
## 11 0.49095195 0.54004714     K #FFF100
## 12 0.54004714 0.58914233     L #FFD600
## 13 0.58914233 0.63823753     M #FFBB00
## 14 0.63823753 0.68733272     N #FFA100
## 15 0.68733272 0.73642792     O #FF8600
## 16 0.73642792 0.78552311     P #FF6B00
## 17 0.78552311 0.83461831     Q #FF5000
## 18 0.83461831 0.88371350     R #FF3500
## 19 0.88371350 0.93280870     S #FF1A00
## 20 0.93280870 1.00000000     T #FF0000
```

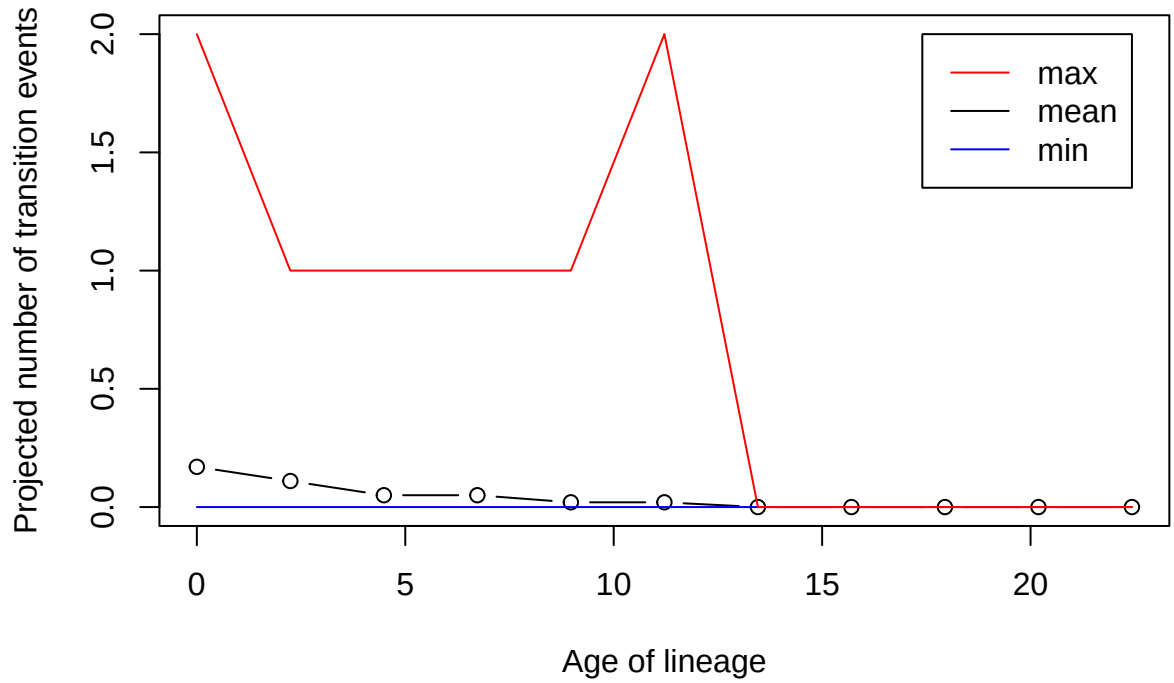
```r
## The columns "from" and "to" have the limits for the categories associated with
##      each color on column "col". The "label" column is just a sequence of letters
##      to help plotting the information.
## The values of "from" and "to" are rate scalers. So the true transition rate for
##      each of these categories are:
ft_rate <- data.frame(from_rate = rate_tree$legend_matrix$from * Q[1,2]
                      , to_rate = rate_tree$legend_matrix$to * Q[1,2])
ft_rate
```

```
##      from_rate     to_rate
## 1  0.000000000 0.001472856
## 2  0.001472856 0.002945712
## 3  0.002945712 0.004418568
## 4  0.004418568 0.005891423
## 5  0.005891423 0.007364279
## 6  0.007364279 0.008837135
## 7  0.008837135 0.010309991
## 8  0.010309991 0.011782847
## 9  0.011782847 0.013255703
## 10 0.013255703 0.014728558
## 11 0.014728558 0.016201414
## 12 0.016201414 0.017674270
## 13 0.017674270 0.019147126
## 14 0.019147126 0.020619982
## 15 0.020619982 0.022092838
## 16 0.022092838 0.023565693
## 17 0.023565693 0.025038549
## 18 0.025038549 0.026511405
## 19 0.026511405 0.027984261
## 20 0.027984261 0.030000000
```

We can investigate the pattern created by the slowdown function by plotting the scaling factor in function of the age of the lineage. Below we will use the average branch length as the age of the lineage as a point of reference.

```
mean_br <- mean(phy$edge.length)
rate_decay <- simulate_rate_decay(change_rate = 0.2, start_rate = Q
                                  , time_elapsed = mean_br, plot = TRUE
                                  , nsims = 100)
```

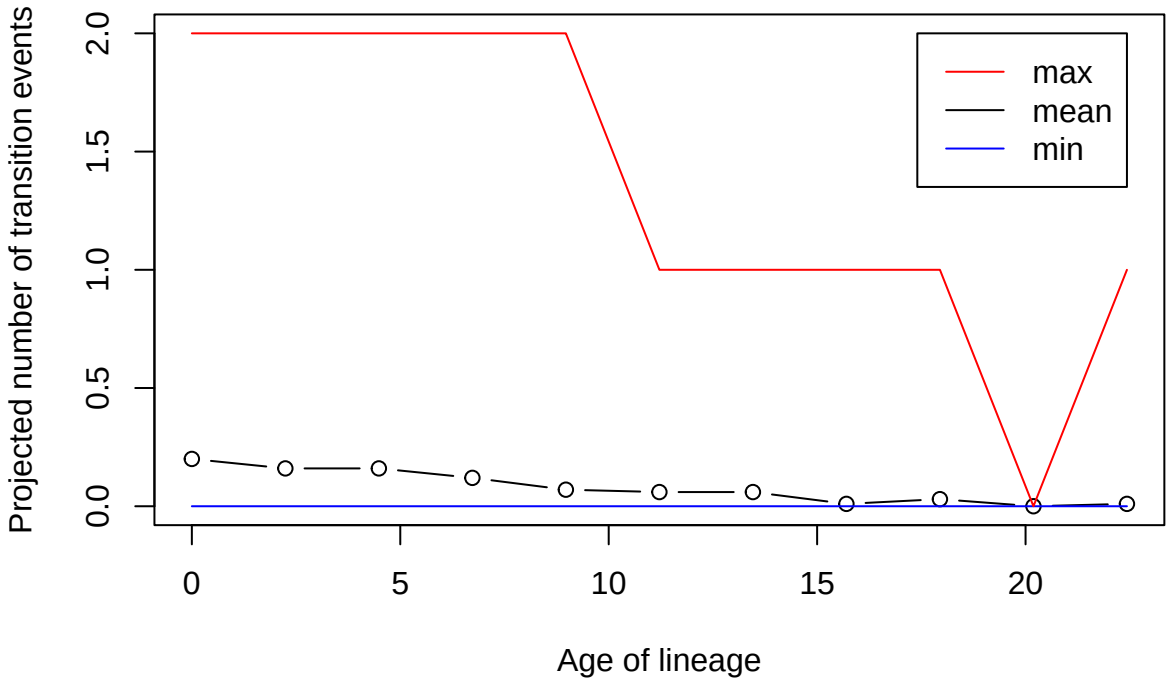
```
## [1] "Projecting expected number of transitions after 22.4324713275105 time units using a MK model."
```



The figure above shows how the expected number of transitions will decrease over time. If we change the `change_rate` we can control how fast this slowdown happens:

```
slow_rate_decay <- simulate_rate_decay(change_rate = 0.1, start_rate = Q
                                       , time_elapsed = mean_br, plot = TRUE
                                       , nsims = 100)
```

```
## [1] "Projecting expected number of transitions after 22.4324713275105 time units using a MK model."
```



Other functions in `buddPhy`

This was just a quick example of the type of functions that buddPhy has. Here are a list of the functions in the package. Please refer to their help pages for instructions of usage and what type of thing can be done with them.

```
## List of the functions and their parameters:
lsf.str("package:buddPhy")
```

```
## drop_fossils_track_node : function (phy)
## get_Budding_History : function (sim)
## make_phylip_data : function (data, states, file = "biogeobears_data.txt")
## make_scaler_simmap : function (sims, ncat = 10, low_cut, high_cut)
## mcmc_mk_budd_exp : function (phy, trait, k, prior_beta_Q = c(2, 5), prior_beta_budd = c(2,
##     5), prior_exp_rate = 1, sample_prob = c(0.25, 0.25, 0.25, 0.25), Q_prop = 0.2,
##     budd_prop = 0.2, rate_prop = 0.2, gen = 100)
## optim_base_rate : function (sim, target_mean_rate)
## plot_cont_sim : function (sim_obj, edge_width = 1.5, bg = "white", jump_col = "coral",
##     jump_width = 1, main = "", legend = TRUE)
## plot_mother_lineages : function (sim_obj, background_color = "gray", edge_width = 3, no_margin = TRU
## sim_BM_budding_exp : function (tree, sigma, anc = 0, budding_prob = 0, budding_mother = NA,
##     change_rate = 2, decay_fn = TRUE, jump_size = NA)
## sim_BM_trace_history : function (sim_BM, sigma, anc = 0, change_rate = 2, decay_fn = TRUE, jump_size
##     jump_type = "fixed", jump_all = FALSE)
## sim_Mk_budding_exp : function (tree, Q, anc = NULL, budding_prob = 0, budding_mother = NA, change_ra
```

```
##     decay_fn = TRUE, cladogenetic_change = "none", cladogenetic_state = "any",
##     cladogenetic_prob = 1)
## sim_Mk_trace_history : function (sim_MK, Q, anc = NULL, change_rate = 2, decay_fn = TRUE, cladogenet
##     cladogenetic_state = "any", cladogenetic_prob = 1)
## simulate_rate_decay : function (change_rate, start_rate = NULL, time_elapsed, plot = TRUE, npoints =
##     nsims = 1000)
## tree_rescale_exp : function (change_rate = 0.05, budding_prob = 0.3, tree, chunk_fraction = 0.01,
##     decay_fn = TRUE)
```

Some highlights are `sim_Mk_trace_history` which can be used to simulate trait histories while holding the history of the lineages constant. Other interesting functions are `sim_BM_budding_exp` and `sim_BM_trace_history` which work similarly to the example above, but simulate a continuous trait under Brownian motion model.