Phylogenetic effective sample size

Krzysztof Bartoszek

July 25, 2015

Abstract

In this paper I address the question — *how large is a phylogenetic sample?* I propose a definition of a phylogenetic effective sample size for Brownian motion and Ornstein–Uhlenbeck processes — the *regression effective sample size*. I discuss how mutual information can be used to define an effective sample size in the non–normal process case and compare these two definitions to an already present concept of effective sample size (the mean effective sample size). Through a simulation study I find that the AIC_c is robust if one corrects for the number of species or effective number of species. Lastly I discuss how the concept of the phylogenetic effective sample size can be useful for biodiversity quantification, identification of interesting clades and deciding on the importance of phylogenetic correlations.

Keywords : Biodiversity, effective sample size, measurement error, Ornstein–Uhlenbeck process, phylogenetic comparative methods, quantitative trait evolution

1 Introduction

One of the reasons to introduce phylogenetic comparative methods in the words of Martins and Hansen [1996], was to address the problem of statistical dependence. This they called the "degrees of freedom" or "effective sample size" problem. If we have n species related by a phylogenetic tree then unless this is a star phylogeny then our effective sample size is less than n (in extreme cases even one). This is important in evaluating the accuracy of parameter estimation or hypothesis tests as these depend on the number of independent data points and not on the observed number of data points [Martins and Hansen, 1996]. Ignoring the correlations (and hence inflating the sample size) results in too narrow confidence intervals, inflated p–values and power. All of this leads to type I and II errors of which the user may be oblivious of.

In a phylogenetic context the calculation of the effective number of observations has not been often addressed directly. In statistical literature effective sample size (ESS) is usually parameter specific, it can be understood as "the number of independent measurements one would need to reach the same amount of information about a parameter as in the original data" [Faes et al., 2009] — in other words how many independent points do we have for

estimating a particular parameter. Nunn [p. 145 2011] points out that often phylogenetic comparative methods have been viewed in a restricted manner as a "degrees of freedom" correction procedure to reduce (due to the nonindependence) the statistical power of an analysis. Most phylogenetic comparative methods work in this way — one assumes a model and maximizes the likelihood under that model. Hence the issue of ESS as mentioned above has been taken care of but only for this problem. In other situations as Nunn [2011] following Pagel [1993] reminds the "degrees of freedom analogy can be misleading". It is more important how the variance is partitioned among species. In fact in the case of model selection, or when one wants to know how many "independent" taxa one has e.g. for conservation purposes the situation becomes much more complex. As we will see it is more important how the covariance is structured.

Smith [1994] directly approached the problem of effective sample size. He studied interspecies phenotypic data by a nested ANOVA and "Determination of the taxonomic levels that account for most of the variation can be used to select a single level at which it is most reasonable to consider the data points as independent". From the perspective of modern phylogenetic comparative methods this is a "hack", as Smith [1994] himself wrote "the method improves the nonindependence problem but does not eliminate it". From our perspective his work is important as from the nested ANOVA setup he partitioned the variance into components from different levels of the phylogeny and then defined the effective sample size as

$$\begin{aligned} n_e &= (\#\text{of superfamilies})(\text{PVC for superfamilies}) + (\#\text{of families})(\text{PVC for families}) \\ &\quad +(\#\text{of genera})(\text{PVC for genera})(\#\text{of species})(\text{PVC for species}) \end{aligned} \tag{1}$$

where PVC is percentage of variance component. Citing Smith [1994] again "The method does not require that levels of the nested hierarchy are defined by taxonomic categories. It is possible, and would be preferable, to organize the species of an allometric data set cladistically, using higher and lower levels of nodes from a cladogram to define levels of the nested ANOVA. For some simplified cladograms this is straightforward, but in general, it is often difficult (or impossible) to organize "real–life" cladograms into 3 or 4 levels and apply a nested ANOVA to them". In this work I build up upon the main idea of Smith [1994] to "consider each species as some fraction of a free observation varying between 0 and 1.0, a value could be computed ... that would reflect the balance between constraint and independent evolution. This value is defined as the effective sample size (effective N) for the data set and trait, as opposed to the traditionally used observed sample size (observed N)." Building up on the modern development of stochastic models for phylogenetic comparative methods I do not have to restrict myself to partitioning the data into hierarchical levels containing different fractions of the variance but rather look at the dependence pattern in tree and model of evolution as a whole. This might make it impossible (but maybe not always) to assign to each species (or taxonomic level) its fraction of free observations but as we shall see it will allow me to calculate the sum of fractions of free observations.

2 Effective sample size

Effective sample size is intuitively meant to represent the number of independent particles of data in the sample. If the sample is correlated then each observation will only have a certain fraction of the information it carries particular to itself. The rest of the information will be shared with one/some/all other points in the sample. We would like to quantify what proportion of the whole sample is made up of these independent bits of information. If this proportion is p then our phylogenetic effective sample size (pESS) will be $n_e = pn$. However our situation is a bit different. It is reasonable to assume that we have a least one observation — at least one species described by at least a single trait. One way is to define p to be between 1 and $1/n$. Alternatively we can define as

$$n_e = 1 + p(n - 1), \tag{2}$$

where $p \in [0, 1]$. I will call this p of Eq. (2) the phylogenetic ESS factor. The value n_e/n is useful in practice to compare between different sized phylogenies and I will call it the relative phylogenetic ESS.

Martins and Hansen [1996] point out that in the discrete trait case the ESS cannot be greater than the number of independent evolutionary changes regardless of the number of observed species. Maddison and FitzJohn [2015] very recently remind us of this again. Phylogenetic comparative methods are there to take care of "pseudoreplicates" due to the tree induced correlations. However, especially in the discrete case, tests of significance might have inflated power as one uses the number of species instead of the (unknown) number of independent evolutionary changes. Unfortunately at the moment there does not seem to be any solution for this problem [Maddison and FitzJohn, 2015]. Hopefully the phylogenetic effective sample size concept presented here could indicate a direction for finding one.

Statistical definitions of effective sample size are commonly introduced in the context of parameter estimations — what is the ESS for a given parameter/set of parameters. I am in a different situation — I want to quantify how many independent particles do I observe. In this situation one has to propose one's own definition of effective sample size that will be useful from a practical point of view. This is not an obvious task in the situation of n dependent observations. The case of multivariate observations where individual components are dependent between each other and correlations between traits can be negative will be even more complicated. Below I will discuss a couple of possible approaches for defining effective sample size and in the next section discuss how they can be applied in the phylogenetic comparative methods field.

Ané [2008] defined an effective sample size for estimating the root state under a Brownian motion model of evolution. She noticed that it can be very small — 6 for a phylogeny of 49 species [mammal phylogeny of Garland, T., Jr. et al., 1993]. In fact my simulations give very similar numbers. She defined the effective sample size as

$$n_e^{\mathrm{E}} := \mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}, \tag{3}$$

where $\mathbf{R}$ is the between species correlation matrix. I call this the *mean effective sample size*

(mESS) as it is related to the loss of information when estimating the expected value of a correlated sample.

This definition of an effective sample size has a fault. It does not say how much independent signal there is in the sample but only how much information we have about the expected value. The likelihood function also depends on other moments so I consider other alternatives.

Currently Ornstein–Uhlenbeck process are the state of the art in modelling trait evolution [Bartoszek et al., 2012, Beaulieu et al., 2012, Clavel et al., Cressler et al., 2015, Ingram and Mahler, 2013, Uyeda and Harmon, 2014]. This process on a phylogeny is multivariate normal one. Therefore all the information will be contained in the mean vector and covariance matrix. In fact we have a natural multiple regression approach and each species, y_i, can be represented as

$$y_i = \mathrm{E}\left[y_i|\vec{y}_{-i}\right] + \epsilon_i,$$

where $\vec{y}_{-i}$ is the measurement vector without i. This will be of course of the form

$$y_i = a_i + \vec{b}_i \cdot \vec{y}_{-i} + \epsilon_i,$$

where ϵ_i will be independent of $\vec{y}_{-i}$. The variance of ϵ_i is $\sigma_i^2 - \mathbf{V}_{i,-i}\mathbf{V}_{-i,-i}^{-1}\mathbf{V}_{-i,i}$, where $-i$ notation again means removing the appropriate rows and columns. As the variance of y_i is σ_i^2 then the independent of the other species part of this variance equals $(1 - \mathbf{V}_{i,-i}\mathbf{V}_{-i,-i}^{-1}\mathbf{V}_{-i,i}/\sigma_i^2)\sigma_i^2$. Standardizing every species to variance 1 will mean that each species carries $1 - \mathbf{V}_{i,-i}\mathbf{V}_{-i,-i}^{-1}\mathbf{V}_{-i,i}/\sigma_i^2$ signal specific to itself. Therefore I propose to define a phylogenetic effective sample size, called *regression effective sample size* (rESS), as

$$n_e^R = \sum_{i=1}^{n}(1 - \mathbf{V}_{i,-i}\frac{\mathbf{V}_{-i,-i}^{-1}}{\sigma_i^2}\mathbf{V}_{-i,i}). \tag{4}$$

This formula has the unfortunate property that it can be lesser than 1 especially when the sample size is small. This is as there is no constraint in the formula to ensure it is greater than 1. In such case I take $n_e^R = 1$.

The rESS, just as the mESS, can be calculated for any process evolving on a phylogenetic tree. However just as the mESS does not catch everything, the rESS will not catch everything in the non–normal process situation. In the non–normal process situation it is necessary to reach for more complicated mathematical tools. The motivation behind the multiple regression approach is to measure how much signal is contained about each species in other species and how much is specific to that species. Another way of formulating the problem is to ask how much information is contained in the joint distribution of all of the species when compared with only the marginal distributions. The natural mathematical framework for this is information theory and the concept of mutual information.

As the name itself suggests mutual information quantifies how much information do different probabilistic objects contain about each other. I will briefly introduce a few concepts from information theory pointing the reader to e.g. Koch [2014] for a more detailed discussion.

Definition 1. *[Koch, 2014] Let $\vec{X} \in \mathbb{R}^n$ be a random vector with density f such that it has mean $\vec{\mu}$ and covariance $\mathbf{V}$. Further let f_j $(j = 1, \ldots, n)$ be the marginal densities of f and f_G be a Gaussian density with the same mean $\vec{\mu}$ and covariance $\mathbf{V}$, i.e. for $x \in \mathbb{R}^n$*

$$f_G(\vec{x}) = \left(\sqrt{(2\pi)^n \det(\mathbf{V})}\right)^{-1} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \mathbf{V}^{-1}(\vec{x} - \vec{\mu})\right).$$

We then define the following.

1. *The entropy of f as*

$$\mathcal{H}(f) = -\int_{\mathrm{supp}(f)} f(\vec{x}) \log f(\vec{x}) \mathrm{d}\vec{x},$$

   where $\mathrm{supp}(f) = \{\vec{x} \in \mathbb{R}^n : f(\vec{x}) > 0\}$ is the support of f.

2. *The negentropy of f as*

$$\mathcal{J}(f) = \mathcal{H}(f_G) - \mathcal{H}(f).$$

3. *The mutual information of f as*

$$\mathcal{I}(f) = \sum_{j=1}^{n} \mathcal{H}(f_j) - \mathcal{H}(f).$$

Intuitively the entropy of a density (or rather random variable behaving according to its law) is the measure of uncertainty about the value of this random variable prior to observation. The negentropy from our perspective is more of a technical term however the mutual information between two densities (or random variables) will be very important in proposing an effective sample size definition.

The maximum sample size attained is n, when all species are independent of each other (we have a star phylogeny). In this situation the density function of our n dimensional vector of observations will be the product of the marginal n densities. No observation contains any information about any other. Therefore to quantify how much information do sample points contain about each other we will consider in Lemma 1 the mutual information between the sample's n–dimensional density and the density defined as the product of the marginal densities. If we recall that all the considered evolutionary models here (Brownian motion, Ornstein–Uhlenbeck) are multivariate normal then we should expect that the entropy based measures be dependent only on the covariance matrix and marginal variances. In the Gaussian case all shared knowledge is coded in the covariance structure. As we see in Lemma 1 this will indeed be the case.

Lemma 1. *[Koch, 2014] Using the notation of Definition 1 the entropy, negentropy and mutual information posses the below properties and relationships between them.*

1. *The negentropy* $\mathcal{J} \geq 0$ *and* $\mathcal{J}(f) = 0$ *iff* f *is Gaussian.*

2. *The mutual information* $\mathcal{I} \geq 0$ *and* $\mathcal{I} = 0$ *iff* $f = \prod_{j=1}^{d} f_j$.

3. *If* f *is Gaussian then it has entropy*

$$\mathcal{H}(f) = \frac{1}{2}\left(n\left(1 + \log\left(2\pi\right)\right) + \log\det\mathbf{V}\right). \tag{5}$$

4. *If* $\mathbf{V}$ *is invertible, then*

$$\mathcal{I}(f) = \mathcal{J}(f) - \sum_{j=1}^{n} \mathcal{J}(f_j) + \frac{1}{2}\log\left(\frac{\prod_{j=1}^{n}\sigma_j^2}{\det\mathbf{V}}\right) \tag{6}$$

where σ_j^2 *are the diagonal elements of* $\mathbf{V}$ *— the marginal variances. If* f *is Gaussian this simplifies to*

$$\mathcal{I}(f) = \frac{1}{2}\log\left(\frac{\prod_{j=1}^{n}\sigma_j^2}{\det\mathbf{V}}\right). \tag{7}$$

It would be tempting to propose mutual information effective sample size as something like

$$\left(1 - \frac{\mathcal{I}(f)}{\sum_{j=1}^{n}\mathcal{H}_j}\right) n$$

but this will not work as $\mathcal{H}_j$ can easily be negative. We therefore have to find some other way of using the entropy. Lin et al. [2007] used a similar motivation to define an effective sample size in order to obtain correct standard errors for parameter estimates. Theirs was a Bayesian setting and they define the effective sample size as a minimizer of a relative entropy. This relative entropy is between the posterior parameter distribution under the true model and the the posterior parameter distribution under the effective sample. However their approach does not allow for fractional sample sizes and could require in the phylogenetic case optimizing over the power set of species. Therefore I propose to define the mutual information ESS (miESS) as

$$n_e^{MI} = 1 + \frac{1}{e(\mathcal{I}(f))}(n-1), \tag{8}$$

where $e(\cdot)$ is a strictly increasing function such that $e(0) = 1$ and $e(\infty) = \infty$. One example of such a function is the logarithm of $\mathcal{I}(f)$ increased by $\exp(1)$, considered in this work. I

choose this function as compared to other formulae, e.g. $\exp(\cdot)$, it resulted in phylogenetic ESSs similar to those defined by the two other formulae. However this formula for $e(\cdot)$ should only be treated as a temporary definition. Further work is needed to appropriately define it so that e.g. in the case of normal processes (like BM or OU ones) it agrees with the rESS.

The ESS defined as such has the desirable properties of being between 1 and n. In the Gaussian case will equal

$$n_e^{MI} = 1 + \log\left(\exp(1) + \frac{1}{2}\left(\sum_{j=1}^{n} \log \sigma_j^2 - \log\det \mathbf{V}\right)\right)^{-1}(n-1). \tag{9}$$

It is important to notice that the three proposed concepts of effective sample sizes are not compatible with each other. Firstly the mESS is meant to quantify only information about the expected value of the sample, not about independent signal. The motivations behind miESS and rESS are the same but it remains for a further study to define an appropriate transformation $e(\cdot)$ that will make miESS equal to rESS in the normal process case. In Sections 3, 4, 5 and 6 I study their behaviour for simulated and real data.

2.1 Multivariate extension

All of the above three definitions assumed that the each of the sample points is univariate. However methods for studying multiple co–evolving traits on the phylogeny are being developed [see e.g. Bartoszek et al., 2012, Beaulieu et al., 2012, Clavel et al., Hansen et al., 2008] and all these ESS concepts are immediately generalizable for these models. Assume now that we have a d dimensional trait. This means that we have for each of our n points a d dimensional observation and our sample is of size $d \cdot n$ correlated points instead of n and $\mathbf{V} \in \mathbb{R}^{nd\times nd}$ instead of $\mathbb{R}^{n\times n}$. This means that for model selection purposes we can use the above described procedures replacing n with $d \cdot n$ inside all formulae, as most software packages do.

The miESS and rESS can be elegantly generalized to quantify how many d–dimensional observations we have effectively, i.e. how many effectively independent species do we have amongst our n species, regardless of the dimensionality of each species. Notice that Eq. (8) does not depend on the dimensionality of the species hence it can be used nearly without change

$$n_e^{MI} = 1 + \frac{1}{e(\mathcal{I}(f))}(n-1). \tag{10}$$

The only difference is that here $\mathcal{H}_j$ is the entropy not of a univariate random variable but of the d_j–dimensional random vector of the j–th species. In the Gaussian case this formula becomes

$$n_e^{MI} = 1 + \log\left(\exp(1) + \frac{1}{2}\left(\sum_{j=1}^{n} \log\det \mathbf{V}_j - \log\det \mathbf{V}\right)\right)^{-1}(n-1) \tag{11}$$

where $\mathbf{V}_j$ is the j–th d_j–dimensional diagonal block of $\mathbf{V}$, i.e. the marginal covariance matrix of the j–th d_j–dimensional observation.

In a similar fashion we can adapt the n_e^R to count the number of effective species in the multitrait case. We sum the conditional total variances i.e.,

$$n_e^R = \sum_{i=1}^{n} \det\left(\mathbf{I} - \mathbf{V}_{i,i}^{-1}\mathbf{V}_{i,-i}\mathbf{V}_{-i,-i}^{-1}\mathbf{V}_{-i,i}\right), \tag{12}$$

where $\mathbf{I}$ is the unit matrix of dimension equalling the number of traits. Notice that in no case is it required that all species are of the same dimension. This allows for proper handling of missing data and no species needs to be removed.

3 Phylogenetic effective sample size

Effective sample size calculation is very important in the phylogenetic context but it seems to have received little attention. Actually it seems that the attitude is that phylogenetic comparative methods have taken care of the problem of inflated sample size in statistical analysis when analyzing inter–species data. This is true with the caveat the it has taken care of only some of the problems. We obtain the correct likelihood value, may in principle obtain correct confidence intervals and p–values. However when the problem actually depends on the sample size then these methods still need to be further developed. Effective (and not observed) sample sizes are important when quantifying the biodiversity of a clade to e.g. develop conservation strategies or when doing model selection.

It would seem desirable to be able to read–off the effective sample size from the phylogeny and base any further estimate on this value of n_e. In fact this seems to be the postulated approach by Nunn [Ch. 11 2011] that one should use the tree's phylogenetic diversity to obtain the amount of information in the sample. Nunn [2011] does not formulate it exactly in this way but this is how mathematically it should be understood. In Section 5 on phylogenetic diversity and conservation I discuss this in detail. However using phylogenetic diversity to obtain an effective sample size for a trait (or suite of them) will be akin to assuming Brownian motion (neutral drift) model of evolution. Phylogenetic diversity is the sum of all branch lengths on a tree and this is proportional to the sum of the variances of independent changes on the tree.

However as Hansen and Orzack [2005] pointed out Brownian change is not appropriate for traits under stabilizing selection. I discussed earlier that all considered definitions of effective sample size will depend on $\mathbf{V}$, the between–species covariance matrix, and how it differs from a diagonal matrix. Therefore we need to calculate n_e based on $\mathbf{V}$ and not just the phylogeny. The between species covariance matrix depends not only the phylogeny but also on the model of evolution. We denote by $\mathbf{T} = [t_{ij}]_{1\leq i,j,\leq n}$ the matrix of speciation times, where t_{ij} is the speciation time of species i and j and t_i the time of species i (these will be all equal to the tree height if the tree is ultrametric). Bartoszek et al. [2012] report the form of $\mathbf{V}$ for various models of evolution.

- Unconstrained evolutionary model — univariate Brownian motion defined by the SDE $\mathrm{d}X_t = \sigma \mathrm{d}B_t$, we have

$$\mathbf{V} = \sigma^2 \mathbf{T}. \tag{13}$$

- Constrained evolutionary model — univariate Ornstein–Uhlenbeck process defined by the SDE $\mathrm{d}X_t = -\alpha(X_t - \theta_t)\mathrm{d}t + \sigma \mathrm{d}B_t$, we have

$$\mathbf{V}[i,j] = \frac{\sigma^2}{2\alpha}\left(e^{-\alpha(t_i+t_j-2t_{i,j})} - e^{-\alpha(t_i+t_j)}\right). \tag{14}$$

- Multitrait unconstrained evolutionary model — multivariate Brownian motion defined by the SDE $\mathrm{d}\vec{X}_t = \mathbf{\Sigma}\mathrm{d}\vec{B}_t$, we have

$$\mathbf{V} = \mathbf{T} \otimes \left(\mathbf{\Sigma}\mathbf{\Sigma}^T\right) \tag{15}$$

where $\otimes$ is the Kronecker product.

- Multitrait constrained evolutionary model, traits adapting to constrained traits — multivariate Ornstein–Uhlenbeck process defined by the SDE $\mathrm{d}\vec{X}_t = -\mathbf{A}(\vec{X}_t - \vec{\theta}_t)\mathrm{d}t + \mathbf{\Sigma}\mathrm{d}\vec{B}_t$, we have

$$\begin{aligned}
\mathbf{V}_{ij} &= e^{-\mathbf{A}(t_i - t_{i,j})} \int_0^{t_{i,j}} e^{-\mathbf{A}v}\mathbf{\Sigma}\mathbf{\Sigma}^T e^{-\mathbf{A}^T v}\mathrm{d}v e^{-\mathbf{A}^T(t_j - t_{i,j})} \\
&= \mathbf{P}e^{-\mathbf{\Lambda}(t_i - t_{i,j})}\left(\left[\frac{1}{\lambda_k+\lambda_r}\left(1 - e^{-(\lambda_k+\lambda_r)t_{i,j}}\right)\right]_{1\le r,k\le d} \odot \mathbf{P}^{-1}\mathbf{\Sigma}\mathbf{\Sigma}^T\mathbf{P}^{-T}\right)e^{-\mathbf{\Lambda}(t_i - t_{i,j})}\mathbf{P}^T,
\end{aligned} \tag{16}$$

where $\odot$ is the Hadamard product, $\mathbf{P}$, $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ are the eigenvectors and eigenvalues of $\mathbf{A}$ and $\mathbf{V}_{ij}$ is the block i, j of dimension $d \times d$ of $\mathbf{V}$, i.e. the intersection of the rows $((i-1)d, \ldots, id)$ and columns $((j-1)d, \ldots, jd)$.

- Multitrait constrained evolutionary model, traits adapting to unconstrained traits — multivariate Ornstein–Uhlenbeck process defined by the SDE system

$$\begin{aligned}
\mathrm{d}\vec{Y}_t &= -\mathbf{A}\left(\vec{Y}_t - \left(\vec{\theta}_t + \mathbf{B}\vec{X}_t\right)\right)\mathrm{d}t + \mathbf{\Sigma}_y \mathrm{d}\vec{B}^y_t \\
\mathrm{d}\vec{X}_t &= \mathbf{\Sigma}_x \mathrm{d}\vec{B}^x_t,
\end{aligned}$$

we have

$$\mathbf{V}_{ij} = \left[\begin{array}{l|l} e^{-\mathbf{A}(t_i-t_{i,j})}\left(\int_0^{t_{i,j}} e^{-\mathbf{A}v}\mathbf{\Sigma}_y\mathbf{\Sigma}_y^T e^{-\mathbf{A}^T v}\mathrm{d}v + \int_0^{t_{i,j}} e^{-\mathbf{A}v}\mathbf{B}\mathbf{\Sigma}_x\mathbf{\Sigma}_x^T\mathbf{B}^T e^{-\mathbf{A}^T v}\mathrm{d}v\right)e^{-\mathbf{A}^T(t_j-t_{i,j})} & t_{i,j}\mathbf{B}\mathbf{\Sigma}_x\mathbf{\Sigma}_x^T \\ +e^{-\mathbf{A}(t_i-t_{i,j})}(\mathbf{I}-e^{-\mathbf{A}t_{i,j}})\mathbf{A}^{-1}\mathbf{B}\mathbf{\Sigma}_x\mathbf{\Sigma}_x^T\mathbf{B}^T & -e^{-\mathbf{A}(t_i-t_{i,j})}(\mathbf{I}-e^{-\mathbf{A}t_{i,j}})\mathbf{A}^{-1}\mathbf{B}\mathbf{\Sigma}_x\mathbf{\Sigma}_x^T \\ +\mathbf{B}\mathbf{\Sigma}_x\mathbf{\Sigma}_x^T\mathbf{B}^T\mathbf{A}^{-T}(\mathbf{I}-e^{-\mathbf{A}^T t_{i,j}})e^{-\mathbf{A}^T(t_j-t_{i,j})} & \\ +t_{i,j}\mathbf{B}\mathbf{\Sigma}_x\mathbf{\Sigma}_x^T\mathbf{B}^T & \\ \hline t_{i,j}\mathbf{\Sigma}_x\mathbf{\Sigma}_x^T\mathbf{B}^T & \\ -\mathbf{\Sigma}_x\mathbf{\Sigma}_x^T\mathbf{B}^T\mathbf{A}^{-T}(\mathbf{I}-e^{-\mathbf{A}^T t_{i,j}})e^{-\mathbf{A}^T(t_j-t_{i,j})} & t_{i,j}\mathbf{\Sigma}_x\mathbf{\Sigma}_x^T \end{array}\right] \tag{17}$$

where $\mathbf{I}$ is the identity matrix of dimensions $d \times d$.

This means that before reporting an effective sample size for a clade one has to estimate the parameters of the evolutionary model.

Given a phylogenetic tree and model of evolution we can easily calculate the effective sample size by plugging in the appropriate formula. Below I present the values of the different definitions of ESS for the Brownian motion model of evolution as the others would be too lengthy to be readable. We assume the tree is ultrametric with height T.

$$\begin{aligned} n_e^{MI} &= 1 + \left(\sqrt{\frac{\det \mathbf{T}}{T^n}}\right)(n-1) \\ n_e^{R} &= \sum_{i=1}^{n}\left(1 - \tfrac{1}{T}\mathbf{T}_{i,-i}\mathbf{T}^{-1}_{-i,-i}\mathbf{T}_{-i,i}\right) \\ n_e^{E} &= T\vec{1}^T\mathbf{T}^{-1}\vec{1}. \end{aligned} \tag{18}$$

I illustrate the above formulae in Figs. 2. I also include the effective sample sizes for Ornstein–Uhlenbeck models. The considered evolutionary scenarios are a Brownian motion and Ornstein–Uhlenbeck process. We fix the initial state $X_0 = 0$ and $\sigma^2 = 1$. For the Ornstein–Uhlenbeck process we also fix the optimum $\theta = 1$. We vary the adaptation rate $\alpha = 0, 0.25, 0.5, 1$. We consider three binary phylogenetic tree setups (see Fig. 1). Two are deterministic trees a completely unbalanced tree, a completely balanced tree (number of tips is a power of two). The third type is a random one — a conditioned on the number of tip species Yule (pure birth) tree [Bartoszek and Sagitov, 2015b, Gernhard, 2008a,b, Sagitov and Bartoszek, 2012]. The rate of speciation is taken at $\lambda = 1$. I take the number of tip species to be from 5 to 200. Of course in the balanced tree only those that are powers of two are allowed, hence there is significantly fewer trees. Each point is the average over 1000 simulations.

To make the simulations comparable the heights of the two deterministic tree types were scaled to $\log n$, the expected height of the Yule tree. Also for these topologies randomness was added by drawing the length of the root branch from the exponential with rate 1 distribution.

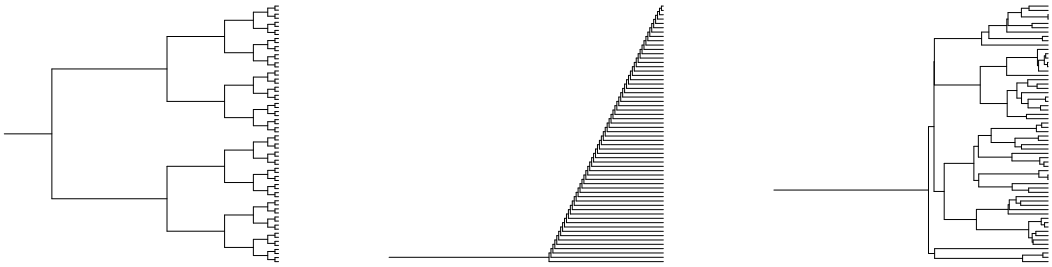

Figure 1: Different binary phylogenetic tree setups used in the simulation studies. Left: fully balanced tree, centre fully unbalanced tree, right: single realization of a pure birth tree. The balanced tree has 64 tips, the other two 60.

In the case of the OU model it allows the process to approach stationarity/stasis before speciation starts to take effect.

4 Phylogenetic information criteria

My main motivation for studying the effective sample size in the phylogenetic context is obtaining correct values of information criteria that depend on sample size. Information criteria are necessary for e.g. finding the best evolutionary model, testing evolutionary hypotheses, distinguishing between competing phylogenies [Bartoszek and Lió, 2014] or regime layouts [Butler and King, 2004]. If the evolutionary models/hypotheses are nested then models can be compared by a likelihood–ratio test — does the increase in the number of parameters significantly improve the model fit. Alternatively when the models are not nested the Akaike information criteria that penalizes for the number of extra parameters

$$\mathrm{AIC} = 2k - 2\log\mathcal{L}$$

was proposed. In the above k is the number of parameters and $\mathcal{L}$ the likelihood. The model with the lower AIC value is the better one. However both the χ^2 distribution of the likelihood ratio test and the AIC are asymptotic approaches. They will be correct when the sample size is infinite (or large in practice). In phylogenetic comparative studies the number of species is usually small. Therefore two alternative criteria that correct for small sample size were proposed to the phylogenetic comparative method community [Hansen et al., 2008]

$$\mathrm{AIC}_c = \mathrm{AIC} + \frac{2k(k+1)}{n-k-1}$$

and the Bayesian (or Schwarz) information criterion [Butler and King, 2004]

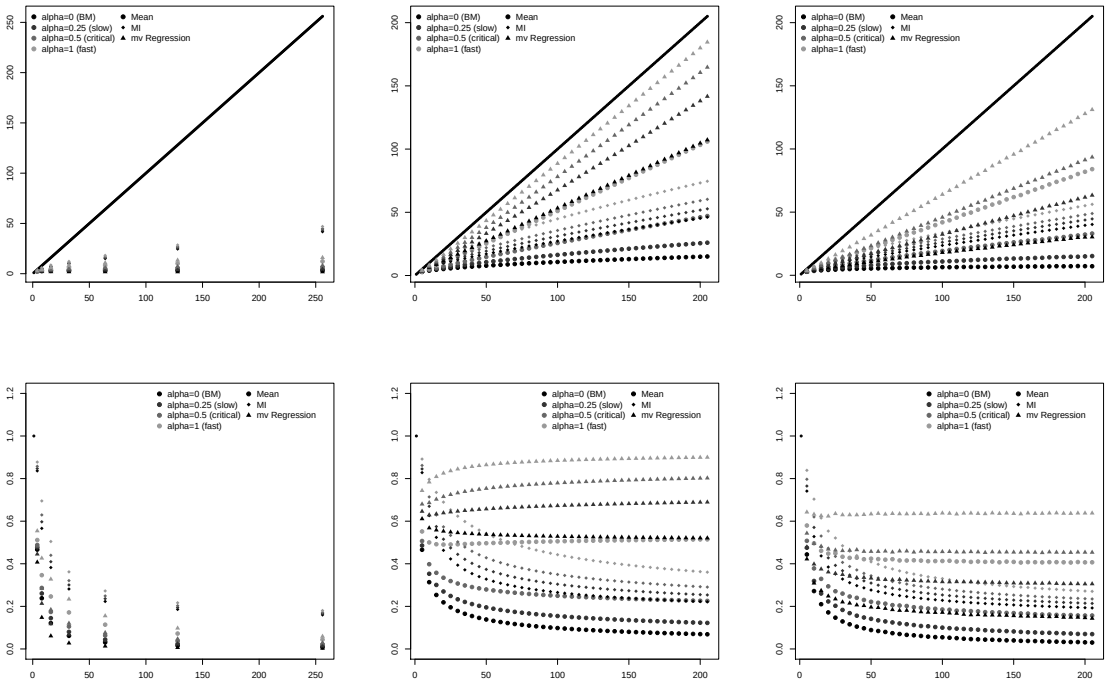

Figure 2: Phylogenetic effective sample sizes for different types of trees and evolutionary processes. First column: balanced tree, second column: left unbalanced tree, third column: average of 1000 pure–birth Yule trees ($\lambda = 1$). The balanced trees and unbalanced trees were generated using the function stree() of the R [R Core Team, 2013] ape package [Paradis, 2012], the Yule trees by the TreeSim R package [Stadler, 2009, 2011]. First row: phylogenetic effective sample sizes, n_e second row: phylogenetic effective sample size factors, p. The parameters of the processes are Brownian motion ($X_0 = 0$, $\sigma^2 = 1$), second row: Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$), third row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$), fourth row: Ornstein–Uhlenbeck process ($\alpha = 1$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$). The straight black line is the observed number of taxa n.

$$\mathrm{BIC} = -2\log\mathcal{L} + k\log n.$$

Of these two the AIC_c seems to be the more used one (but AIC is also very popular). However if one goes back to the derivation of the AIC_c [Hurvich and Tsai, 1989] and BIC [Schwarz, 1978] one can see that the n observations are assumed independent. Therefore a phylogenetic comparative model violates this assumption, in the best case by inflating the sample size, which corresponds to not penalizing enough for additional parameters. This is similar with the BIC. However in their original paper Hurvich and Tsai [1989] derive the same AIC_c formula for autoregressive models so this warrants further study in the phylogenetic setting where the covariance structure is hierarchical.

Therefore using the number of species (unless the phylogeny is a star) results in a risk of overfitting for small phylogenies or those with most speciation events near the tips. I therefore propose to take into account the effective number of species during the model selection procedure. The newest version of mvSLOUCH (available from `http://cran.r-project.org/web/packages/mvSLOUCH/index.html`) does exactly this. It does automatic model selection if one treats n as the true sample size and also if one corrects for the dependencies and uses effective sample size (with the user allowed to indicate which definition of pESS to use, by default the regression ESS). Importantly mvSLOUCH allows for an arbitrary pattern of missing data — no observation is removed and the likelihood is based on all provided information.

To see how much of a difference this makes I performed a simulation study under various evolutionary scenarios. Under each scenario I simulate data $N = 1000$ times and from this obtain histograms of the AIC_c values under the true model and an alternative using both the number of species and the effective sample size, Figs. S.1—S.8 in the supplementary material. I also plot in Fig. 3 how the average value of the small sample size correction changes under the different evolutionary models and effective sample size value. We consider the same evolutionary scenarios as in Fig. 2 and observe that for large α identifiability of the true model is easier. The histograms of the AIC_c are shown for small ($n = 30$) and large ($n = 205$) phylogenies. We can see that in the large phylogeny case all definitions of sample size result in the same distribution of AIC_c. However for the small phylogeny the mean and regression ESSs, n_e^R and n_e^E, seem to be more effective with the balanced phylogeny and fast adaptation. Also it is visible that distinguishing different adapting OU models from each other and the BM one can be difficult. This difficulty, especially with smaller αs, is to be expected as the slowly adapting processes can take a lot of time to reach stationarity and loose ancestral signal [Adamczak and Miłoś, 2014, 2015, Ané et al., 2014, Bartoszek and Sagitov, 2015b]. In fact our simulations confirm in this respect Cressler et al. [2015]'s recent study — "Selection opportunity (*i.e.* α) is substantially more difficult to estimate accurately: ... relative errors exceeding 100% are common, even when the correct model has been selected." This is especially true for small n and α [Fig. 6 of Cressler et al., 2015]. This indicates that significantly larger sample sizes would be needed to identify slowly adapting models. Figure 3 also tells us that even with smaller sized phylogenies all pESS definitions should result in similar AIC_c values. This observed agreement suggests that the likelihood

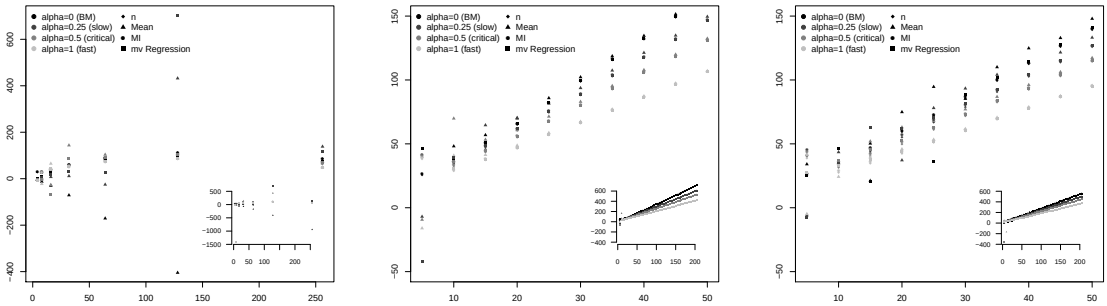

Figure 3: AIC_c effective sample size correction for different types of trees and evolutionary processes. Left: balanced tree, centre: left unbalanced tree, right: average of 1000 pure–birth Yule trees ($\lambda = 1$). The balanced trees and unbalanced trees were generated using the function stree() of the R ape package, the Yule trees by the TreeSim R package. The parameters of the processes are Brownian motion ($X_0 = 0$, $\sigma^2 = 1$), second row: Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$), third row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$), fourth row: Ornstein–Uhlenbeck process ($\alpha = 1$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$).

dominates the AIC_c, which is not surprising as the data is simulated under the BM or OU models. A similar consistency is observed when working with real data (Section 6). The situation is different for the fully balanced tree which holds the most dependencies between the species. In this case probably a much larger tree would be needed to obtain stability.

It is also worth noticing that the AIC_c values can be heavy tailed especially in the Yule tree case and n_e^E. This is visible in the supplementary histograms by peaks at the ends of the x–axis. Ané [2008] noticed that for a Brownian motion models of evolution effective sample sizes can be very small. Garland, T., Jr. et al. [1993]'s mammal phylogeny had $n_e^E = 6$ with 49 tip species. My simulations give very similar numbers (Fig. 2). A Yule tree of 50 tips has $\mathrm{E}\left[n_e^E\right] = 5.391$, $\mathrm{E}\left[n_e^{MI}\right] = 14.574$ and $\mathrm{E}\left[n_e^R\right] = 10.668$, a fully unbalanced tree with 50 tips has $\mathrm{E}\left[n_e^E\right] = 7.781$, $\mathrm{E}\left[n_e^{MI}\right] = 17.06$ and $\mathrm{E}\left[n_e^R\right] = 27.349$ and a fully balanced tree of 64 tips has $\mathrm{E}\left[n_e^E\right] = 2.909$, $\mathrm{E}\left[n_e^{MI}\right] = 9.729$ and $\mathrm{E}\left[n_e^R\right] = 1.854$. The very low amount of independent information is evident. In most cases the mean effective sample size is the lowest because it measures the information that the sample contains on the mean value. In the BM case this is the ancestral state and there is very little information on it. The other pESSs look more holistically at what dependencies are in the data and hence are larger. If we move to more and more adaptive OU models (increase α) then all, but especially n_e^E, increase. The mean ESS is nearly always the smallest. However if adaptation is fast and terminal branches are long (i.e. the contemporary sample is nearly independent) then it can also be nearly n (see Tab. 1).

5 Phenotypic diversity and conservation

An important application of phylogenetic methods is to quantify the biodiversity of a group of species. Phylogenetic methods allow one to formulate definitions of species that are useful from an evolutionary point of view [Ch. 11 Nunn, 2011]. I will not be concerned with a definition of a species but assume that some phylogeny relating predefined taxonomic units is available. The impact of a phylogenetic definition of species was investigated by Agapow et al. [2004] and they noticed that this caused an average increase of the number of species by about 49% when compared to alternative definitions. Naturally this means that a lot of species were "split", resulting in species with smaller populations and geographic ranges. In turn as these are variables contributing to classifying a species as endangered, it may lead to more labelled as such. Therefore Agapow et al. [2004] postulated quantifying conservation value using alternative variables, one of which was trait diversity.

Faith [1992] suggested to quantify biodiversity through phylogenetic information. The main idea is that one should concentrate on feature diversity — how diverse are organisms. This is of course something difficult to quantify, we do not even know of all the variables to measure. Crozier [1997] pointed out that one of the aims of conservation is to "maximize the preserved information of the planet's biota best in terms of genetic information". He then points out that phylogenetic based measures which include branch lengths will be better indicators than just counting the number of species. Therefore as a proxy Nunn [2011] following Faith [1992] proposes [but also refers the reader to Faith, 1994, 2002, Crozier, 1997, Purvis et al., 2005] to quantify feature diversity with phylogenetic diversity (PD) — the sum of branch lengths of a tree/clade. The extinction of a clade (or species) is therefore equivalent to subtracting the amount of branch lengths particular to this clade. Phylogenetic diversity has a rich literature [e.g. Crawford and Suchard, 2013, Mooers et al., 2012, Stadler and Steel, 2012] and therefore it is possible to make quantitative predictions about diversity loss/retention under different models of tree growth, extinction and conservation.

From a mathematical perspective this means that we are quantifying the amount of feature diversity as the amount of accumulated variance under the assumption that evolution follows Brownian motion. One may say that this is sensible as an overall feature variable describing a species will be the sum of of effects of many traits. This then one could argue will be evolving as a Brownian drift.

An alternative approach that could be used to quantify the biodiversity (or feature diversity) of a clade of n species is the effective amount of species in this clade n_e. This is done in a straightforward way. We prune the phylogeny to the subtree which contains only this clade, and use the methods described in the work to obtain n_e for this subtree. This approach could be more appropriate for various reasons. For example it could turn out that the traits important from a conservation point of view are quantified by another process e.g. Ornstein–Uhlenbeck. In this case the changes along disjoint parts of the phylogeny might not be independent and the variance is not a linear function of time. However even if we have Brownian evolution then just summing up variances (and forgetting about covariances) might not be satisfactory.

To illustrate this I will first use an example from linguistics [where phylogenetic methods

have wide applications as discussed by Nunn, 2011]. This is not meant to be an authoritative statement in the field but rather an illustration that should be much more intuitive as the complexity is lesser while there are significant similarities with species conservation. With the set of languages one may associate a tree structure describing their evolutionary relationships similar as in the species case. A network would of course be more appropriate as there is constant flow between most of them but we remain in the "tree world". Imagine a clade of languages that had a very large (as large as needed for the argument) very recent radiation but also has a very long root branch. Under the phylogenetic diversity definition of feature diversity we would say that this clade of languages is diverse. Loosing say half of the clade would lead to a significant drop in the phylogenetic diversity implying a significant loss of feature diversity. Counting language loss would give even more horrific results. However this situation is not exactly like this. Since the radiation was very recent we would expect that all the languages to be very similar. Hence the effective sample size $n_e \approx 1$ even though n can be as large as we want (compare in the balanced long root branch sample sizes in Tab. 1. This is obvious due to the recent divergence we cannot expect much new variation and change [we are not considering processes with jumps Bartoszek, 2014, Bokma, 2002, 2008] between the different tips. If the evolution of the languages would be following an Ornstein–Uhlenbeck process then n_e could be even smaller. Therefore even if half the languages were lost one would not suffer any significant decrease in the sample size. The information content would be essentially the same. Of course in reality this would be different as the loss of cultural diversity associated with the radiation that is not necessarily reflected in the differences between languages could be enormous. But this boils down to what one is modelling with mathematics and how one defines what is information.

In the species context we have a similar situation. Phylogenetic diversity measures can be overturned if one uses the diversity of a (suite of) trait(s) as a proxy for biodiversity. In a very wide and recent shallow radiation the diversity of a trait can be very small while the sum of branch lengths can be large. On the other hand if we have trees with few very old tips then they may have much lower phylogenetic diversity. However they might have diverged so far back in time and accumulated so much change in their phenotype (without speciating) that the loss of even one tip results in a much more significant loss of phenotypic innovation than even of most of a recent shallow radiation. This is intuitively obvious — in a recent shallow radiation the majority of information about all the species is coded inside essentially all of the species. It suffices for only one to survive for most of the information to be retained. But making the radiation wider and wider one can imagine increasing the measure of phylogenetic diversity as much as desired. Of course loosing tips in this case is equivalent to loosing small innovations that set the species apart. This is naturally a value in itself but the majority of information is stored in any individual tip. However in the many old species case every single species is a distinct entity not containing much information about the rest. Hence any loss of a single species leads to an irreplaceable loss of diversity [Nee and May, 1997], while the phylogenetic diversity measure might not pick this up. Nunn [2011, p. 319] points out that we are losing biological and cultural diversity at a faster rate than ever before. Therefore it is important to quantify how much of what we loose.

Essentially this is as phylogenetic diversity does not account for the topology of the tree. It just considers total branch length — naturally different topologies can result in the same value of this. This includes trees that have very long tip branches — a nearly independent sample and those that have multiple recent radiations — a nearly dependent sample. Hence I point out that if one uses the effective sample size as an alternative indicator of diversity one might not fall into these pitfalls. One will be able to quantify how many independent species one really has in a clade. Or another way of phrasing it is that one is trying to quantify evolutionary distinctiveness [Avise, 2005]. This is inline with what Brooks et al. [2005] write, that what really matters is how much unique evolutionary history a species represents. They consider that the value of a clade should be measured by the amount of unique species it has. But of course this comes back to the problem of what is a species and one needs to be able to compare diversity between older, younger clades. The usage of n_e for a clade should allow exactly for this.

Looking at it in yet another way I am changing the currency of biodiversity. Instead of the standard currencies "species" or PD I use diversity in traits. In other words I sum up innovations particular to species (partition the variance) and this allows one to identify "innovative" clades which contain a lot of information. As Agapow [2005] points out very often it is easier to motivate and "find money for conservation of charismatic and easily recognizable organisms." In the public opinion species are recognizable by specific phenotypes. Methodologies identifying clades with high diversity of phenotypes could help in arguing for these clades. This could be one approach towards species–free methods postulated by Agapow [2005]. Mine is not completely species–free of course it still includes the phylogeny. The tips of the tree are pre–defined by experts taxonomic units according to some definition of species. However it is not a only–species methodology as e.g. counting species would be. It includes evolutionary process information that takes into account the topology of the tree — how much of one species is there in another. Also Agapow [2005] discusses that the problem with species methodologies is that depending on the definition of species we can get wildly different counts. This choice has direct financial and conservation implications. And as Isaac and Purvis [2004] point out that correct identification of species numbers is important for understanding the diversity of our world.

Therefore if one misidentifies a species, problems could occur — the species count will be wrong and hence the phylogenetic diversity. It will be based on too few or too many branches. And what if one missed a particular subpopulation that had something very special attached to it? Can one still include its diversity even though it does not appear on the phylogeny? The pESS can precisely do this through integrating data from different sources. Assigning the effective clade size that takes into account the phylogeny and trait variance and between–species covariance should allow one to strike a balance between expert knowledge concerning species and uncertainty attached to correct demarcation. For example evolutionary models can be easily extended to include intra–species variance, often called "measurement error" [e.g. Felsenstein, 2008, Hansen and Bartoszek, 2012, Rohlfs et al., 2013, to name a few]. Mathematically these methods boil down to adding to the matrix $\mathbf{V}$ a matrix $\mathbf{M}$ which is the intra–species variance ("measurement error"). Then this new covariance matrix $\mathbf{V}+\mathbf{M}$

can be treated as the old $\mathbf{V}$ to obtain a value of effective species number. The intra–species variance can be a representation of our uncertainty about species demarcation and be used to correct for species miscall. If a species has many subpopulations that are very diverse representing a species by only its mean over all (measured) individuals will not be the best option. Including the variability of the trait inside the species can partially alleviate the need to know the correct species structure. It can be thought of averaging over all possible species demarcations that we are not sure about.

Mooers et al. [2005] discuss that one can look at this from an ethics point of view should all species be considered equal and protected in the same way or should one protect the features of evolution that are of some value for us. Then phylogenetic diversity is a measure that quantifies a particular feature of evolution. What I propose in my work is quantifying a different feature of evolution. What sets it apart from PD is that it requires the researcher to define traits — exactly what features of evolution are valuable. To illustrate this Nee and May [1997] point out that the loss of *Homo sapiens* would result in a loss of a tiny fraction of evolutionary history when one uses a measure that takes into account only the tree. If we would choose a trait associated with e.g. civilization achievements and then calculate the ESS of the human lineage (1 by definition) and non–human clades we would obtain a completely different result.

In a way one could say that this is merely replacing counting species with counting the effective number of species. However the difference is in how we count. Counting just the number of species means enumerating taxonomic units according to some definition. Counting the effective number of species in the way I propose is really saying how much biodiversity we have in a clade where biodiversity is represented by some (suite of) trait(s). This measure can also be though of as calculating how much innovation we have in the clade. Of course my approach shifts the responsibility to the biologist to identify what traits are important.

Jetz and Freckleton [2015] have very recently published an analysis that is distantly related to what I discuss. They notice that on many species we have too little data to say if they are endangered or not. On the one hand this would mean that we could assume that all data–deficient species are endangered but as Agapow [2005] pointed out this would be far too costly. On the other hand Jetz and Freckleton [2015] point out that Butchart and Bird [2010] observed that data–deficient birds are at no greater extinction risk than assessed birds. This suggests that one could use, as Jetz and Freckleton [2015] do e.g. body–mass, to predict threat status/threat probability. Of course as species are dependent in such an analysis the phylogeny needs to be accounted for. Such an approach has the drawback as Jetz and Freckleton [2015] discuss that a logistic regression, i.e. threatened/not threatened, will require a large dataset. Therefore it might be possible, but this of course requires further development and linkeage with phylogeographical models, that effective clade size could also be a proxy for threat status. In addition Jetz and Freckleton [2015] point out that many species have missing measurements on phenotypes. The evolutionary models used to obtain pESS can handle this in a natural manner. Therefore there is no need to remove a species from an analysis even it has missing data.

In Tabs. 1 and 2 I present situations where the pESS approaches produce results which are in agreement and disagreement with phylogenetic diversity. I considered a number of different phylogenies (see Fig. 4), with recent shallow radiations, with long tip branches, short tip branches and Yule trees. Two considered types are geometrically or harmonically increasing or decreasing branch lengths. These are balanced trees. In the geometric case each level's branch is half of or twice as (decrease or increase) the previous level's one. In the harmonic case the branch length of the i–th level (counting from the root — decreasing or from the tips — increasing) is $1/i$ of the tree's height. On top of these trees I considered the BM process and the OU process with different parameter values. All trees have an expected height of $\log n$. In deterministic trees (balanced and unbalanced, i.e. non–Yule) some randomness to the topology is added by a root branch of exponential with rate 1 length. This is so that the models are more comparable — that some variance is attached to the trait evolution and the OU model is allowed to approach stationarity/stasis before speciation effects begin. For each setup 1000 simulations were made.

The first thing that can strike us in Tabs. 1 and 2 is that PD can be identical despite very different topologies, dependencies and tip species numbers. For example the Yule and unbalanced trees have nearly identical PDs for $n = 16$ while the pESSs suggest that there is a difference between their information content. On the other hand when $n = 125$ there is a large difference between the PDs while not that much in the pESSs.

If we compare the balanced short terminal tree with $n = 128$ and the $n = 16$ balanced harmonic/geometric increase trees then they have nearly identical PDs. Their pESSs are also similar but they explain what is going on, in the first case we have many very similar species in the second a few very distinct ones. This is exactly the situation as I discussed previously — in the latter situation the loss of a species means loosing a completely separate entity, in the former all species contain significant information about all the others.

This is even more evident when considering relative PDs and pESSs, i.e. PD$/n$, n_e/n, (Tab. 2). In this case the PD also does not explain the dependency structure induced by the different topologies. In the first example above (unbalanced and Yule) the relative regression ESS seems stable (similar growth with α) when comparing the small and large phylogenies (both Yule and unbalanced). It clearly shows that there is more independence in the unbalanced tree — as expected there are more long terminal branches. The relative PD does not distinguish between the small Yule and unbalanced phylogenies, and the large Yule phylogeny, while $n_e^R/n = 0.322, 0.588, 0.168$ for small Yule BM, small unbalanced BM and large Yule BM respectively. The regression ESS clearly shows how the tree influences the dependency structure of the tips. Unfortunately the mutual information and mean ones do not describe these dependencies so clearly. However in their case this is explainable — the mean measures only information on the expected value and the MI one needs further refinement with respect to the $e(\cdot)$ transformation.

The general pattern from Tabs. 1 and 2 is that if there is a lot of independence then PD will be large. But as said this does not capture everything. For example I look in more detail at the balanced long terminal, harmonic and geometric increase topologies. The PD measures (absolute and relative) do not distinguish between them. However in Fig. 4 we can

see that there is a substantial difference between the long terminal one and the harmonic and geometric increases. The long terminal sample should essentially be independent while the other two should exhibit dependencies. The n_e^R describes this pattern perfectly. On the long terminal tree all process generate a nearly independent sample with this measure. For the other two the process has to evolve quickly to loose ancestral dependencies. But on the other hand by the PD measure the long terminal branch tree carries less independence (diversity) than the harmonic and geometric increase trees. Furthermore it is interesting to notice that the growth of the relative pESSs with α is similar for all pESS definitions. The geometric increase has larger pESSs due to the longer terminal branches.

Model	n [1]	E [PD] [2]	E $\left[n_e^{MI}\right]$ [3]	E $\left[n_e^{E}\right]$ [4]	E $\left[n_e^{R}\right]$ [5]
Yule					
BM [6]	15	15.102	7.119	3.938	4.827
slow OU [7]	15	15.102	7.744	4.661	6.301
medium OU [8]	15	15.102	8.404	5.562	7.691
fast OU [9]	15	15.102	9.611	7.507	9.778
unbalanced					
BM	15	15.461	8.335	4.55	8.816
slow OU	15	15.461	9.045	5.208	9.875
medium OU	15	15.461	9.808	6.006	10.84
fast OU	15	15.461	11.292	7.885	12.328
balanced					
BM	16	8.584	6.735	2.824	1.914
slow OU	16	8.584	7.146	3.172	2.772
medium OU	16	8.584	7.598	3.608	3.768
fast OU	16	8.584	8.568	4.698	5.914
balanced short terminal					
BM	16	10.829	5.892	2.689	1
slow OU	16	10.829	6.2	3.159	1
medium OU	16	10.829	6.497	3.762	1
fast OU	16	10.829	7.003	5.078	1.675
balanced long terminal					
BM	16	44.023	15.991	15.238	15.994
slow OU	16	44.023	15.998	15.637	15.999
medium OU	16	44.023	16	15.852	16
fast OU	16	44.023	16	15.982	16
balanced harmonic decrease					
BM	16	19.908	7.292	3.053	3.3
slow OU	16	19.908	8.578	4.13	6.228
medium OU	16	19.908	10.107	5.694	9.21

[1] number of tips [2] expected phylogenetic diversity [3] expected mutual information effective sample size [4] expected mean effective sample size [5] expected regression effective sample size [6] Brownian motion $\alpha = 0$ [7] Ornstein–Uhlenbeck $\alpha = 0.25$ [8] Ornstein–Uhlenbeck $\alpha = 0.5$ [9] Ornstein–Uhlenbeck $\alpha = 1$

fast OU	16	19.908	12.896	9.158	13.155
balanced harmonic increase					
BM	16	39.702	11.242	6.25	11.196
slow OU	16	39.702	13.921	9.187	14.313
medium OU	16	39.702	15.454	12.325	15.596
fast OU	16	39.702	15.986	15.379	15.987
balanced geometric decrease					
BM	16	15.171	6.735	2.824	1.914
slow OU	16	15.171	7.54	3.549	3.638
medium OU	16	15.171	8.445	4.55	5.647
fast OU	16	15.171	10.247	6.827	9.22
balanced geometric increase					
BM	16	38.661	12.072	7.5	12.227
slow OU	16	38.661	14.335	10.198	14.667
medium OU	16	38.661	15.544	12.836	15.662
fast OU	16	38.661	15.983	15.418	15.987
balanced long root branch					
BM	16	3.598	7.045	2.824	1.914
slow OU	16	3.598	7.116	2.868	2.025
medium OU	16	3.598	7.189	2.914	2.141
fast OU	16	3.598	7.34	3.011	2.382
Yule					
BM	125	124.839	27.745	6.636	21.033
slow OU	125	124.839	30.748	12.229	39.868
medium OU	125	124.839	33.829	22.896	57.462
fast OU	125	124.839	39.212	51.921	79.882
unbalanced					
BM	125	244.124	32	11.905	66.217
slow OU	125	244.124	36.518	18.825	85.055
medium OU	125	244.124	41.85	30.96	98.566
fast OU	125	244.124	52.44	64.035	111.347
balanced					
BM	128	28.19	24.599	2.977	1.810
slow OU	128	28.19	25.746	4.073	4.04
medium OU	128	28.19	26.798	5.765	6.981
fast OU	128	28.19	28.526	10.189	13.393
balanced short terminal					
BM	128	36.485	24.776	2.983	2.25
slow OU	128	36.485	26.3	4.478	5.893
medium OU	128	36.485	27.662	6.886	10.654
fast OU	128	36.485	29.87	12.816	20.46
balanced long terminal					

BM	128	615.155	124.514	89.930	127.719
slow OU	128	615.155	127.759	116.316	127.980
medium OU	128	615.155	127.993	125.927	127.999
fast OU	128	615.155	128	127.967	128
balanced harmonic decrease					
BM	128	178.258	27.567	3.745	11.906
slow OU	128	178.258	35.106	13.792	49.818
medium OU	128	178.258	44.221	36.794	81.893
fast OU	128	178.258	67.237	78.888	113.052
balanced harmonic increase					
BM	128	683.621	39.878	14.424	72.631
slow OU	128	683.621	101.417	81.108	124.878
medium OU	128	683.621	127.303	123.606	127.94
fast OU	128	683.621	128	127.939	128
balanced geometric decrease					
BM	128	54.926	24.599	2.977	1.81
slow OU	128	54.926	26.783	5.735	6.932
medium OU	128	54.926	28.502	10.116	13.293
fast OU	128	54.926	31.144	19.056	25.482
balanced geometric increase					
BM	128	658.375	48.870	36.286	93.029
slow OU	128	658.375	106.934	95.213	125.675
medium OU	128	658.375	127.340	123.876	127.943
fast OU	128	658.375	128	127.939	128
balanced long root branch					
BM	128	7.259	24.599	2.977	1.81
slow OU	128	7.637	24.768	3.072	2.015
medium OU	128	7.637	24.902	3.175	2.233
fast OU	128	7.637	25.171	3.404	2.712

Table 1: Comparison of phylogenetic diversity with the proposed pESS definitions for different evolutionary models and topologies. The values are means from a 1000 simulations. The value of 1 for the regression ESS indicates that the calculated value was below 1 and hence the rounding up.

Model	n	$\mathrm{E}[\mathrm{PD}/n]$	$\mathrm{E}\left[n_e^{MI}/n\right]$	$\mathrm{E}\left[n_e^{E}/n\right]$	$\mathrm{E}\left[n_e^{R}/n\right]$
Yule					
BM	15	1.007	0.475	0.263	0.322
slow OU	15	1.007	0.516	0.311	0.42
medium OU	15	1.007	0.56	0.371	0.513

fast OU	15	1.007	0.641	0.5	0.652
unbalanced					
BM	15	1.031	0.556	0.303	0.588
slow OU	15	1.031	0.603	0.347	0.658
medium OU	15	1.031	0.654	0.4	0.723
fast OU	15	1.031	0.753	0.526	0.822
balanced					
BM	16	0.537	0.421	0.177	0.12
slow OU	16	0.537	0.447	0.198	0.173
medium OU	16	0.537	0.475	0.226	0.235
fast OU	16	0.537	0.535	0.294	0.37
balanced short terminal					
BM	16	0.677	0.368	0.168	0.063
slow OU	16	0.677	0.388	0.197	0.063
medium OU	16	0.677	0.406	0.235	0.063
fast OU	16	0.677	0.438	0.317	0.105
balanced long terminal					
BM	16	2.751	0.999	0.952	1
slow OU	16	2.751	1	0.977	1
medium OU	16	2.751	1	0.991	1
fast OU	16	2.751	1	0.999	1
balanced harmonic decrease					
BM	16	1.244	0.456	0.191	0.206
slow OU	16	1.244	0.536	0.258	0.389
medium OU	16	1.244	0.632	0.356	0.576
fast OU	16	1.244	0.806	0.572	0.822
balanced harmonic increase					
BM	16	2.481	0.703	0.391	0.7
slow OU	16	2.481	0.87	0.574	0.895
medium OU	16	2.481	0.966	0.77	0.975
fast OU	16	2.481	0.999	0.961	0.999
balanced geometric decrease					
BM	16	0.948	0.421	0.177	0.12
slow OU	16	0.948	0.471	0.222	0.227
medium OU	16	0.948	0.528	0.284	0.353
fast OU	16	0.948	0.64	0.427	0.576
balanced geometric increase					
BM	16	2.416	0.754	0.469	0.764
slow OU	16	2.416	0.896	0.637	0.917
medium OU	16	2.416	0.972	0.802	0.979
fast OU	16	2.416	0.999	0.964	0.999
balanced long root branch					

BM	16	0.225	0.44	0.177	0.12
slow OU	16	0.225	0.445	0.179	0.127
medium OU	16	0.225	0.449	0.182	0.134
fast OU	16	0.225	0.459	0.188	0.149
Yule					
BM	125	0.999	0.222	0.053	0.168
slow OU	125	0.999	0.246	0.098	0.319
medium OU	125	0.999	0.271	0.183	0.46
fast OU	125	0.999	0.314	0.415	0.639
unbalanced					
BM	125	1.953	0.256	0.095	0.53
slow OU	125	1.953	0.292	0.151	0.68
medium OU	125	1.953	0.335	0.248	0.789
fast OU	125	1.953	0.42	0.512	0.891
balanced					
BM	128	0.22	0.192	0.023	0.014
slow OU	128	0.22	0.201	0.032	0.032
medium OU	128	0.22	0.209	0.045	0.055
fast OU	128	0.22	0.223	0.08	0.105
balanced short terminal					
BM	128	0.285	0.194	0.023	0.018
slow OU	128	0.285	0.205	0.035	0.046
medium OU	128	0.285	0.216	0.054	0.083
fast OU	128	0.285	0.233	0.1	0.16
balanced long terminal					
BM	128	4.806	0.973	0.703	0.998
slow OU	128	4.806	0.998	0.909	1
medium OU	128	4.806	1	0.984	1
fast OU	128	4.806	1	1	1
balanced harmonic decrease					
BM	128	1.393	0.215	0.029	0.093
slow OU	128	1.393	0.274	0.108	0.389
medium OU	128	1.393	0.345	0.287	0.64
fast OU	128	1.393	0.525	0.616	0.883
balanced harmonic increase					
BM	128	5.341	0.312	0.113	0.567
slow OU	128	5.341	0.792	0.634	0.976
medium OU	128	5.341	0.995	0.966	1
fast OU	128	5.341	128	1	1
balanced geometric decrease					
BM	128	0.429	0.192	0.023	0.014
slow OU	128	0.429	0.209	0.045	0.054

medium OU	128	0.429	0.223	0.079	0.104
fast OU	128	0.429	0.243	0.149	0.199
balanced geometric increase					
BM	128	5.144	0.382	0.283	0.727
slow OU	128	5.144	0.835	0.744	0.982
medium OU	128	5.144	0.995	0.968	1
fast OU	128	5.144	1	1	1
balanced long root branch					
BM	128	0.057	0.192	0.023	0.014
slow OU	128	0.06	0.194	0.024	0.016
medium OU	128	0.06	0.195	0.025	0.017
fast OU	128	0.06	0.197	0.027	0.021

Table 2: Comparison of relative phylogenetic diversity with the proposed relative pESSs. The values are means from the same 1000 simulations from Tab. 1.

6 pESS in biological data sets

Using the new version of mvSLOUCH I analyzed a number of datasets to see what effects using different definitions of pESS would have on inference. The results of this analysis are presented in Tab. 3. All trees were rescaled to a height of $\log(n-1)$ to be comparable with other results here. I take the -1 as there is no root branch in these trees. In the analysis, OU processes, were assumed to have a single constant optimum over the phylogeny.

It can be seen that all definitions of pESS lead to the same conclusions except for the mESS. Using it can lead to at first sight dramatically different conclusions — an Ornstein–Uhlenbeck process with disruptive selection (i.e. $\alpha < 0$). However when looking into the estimate of α in all cases it was negative but very close to 0 — hence resembling a Brownian motion. Also mESS is not, as explained in the beginning, designed to measure how much independent signal there is in the data. It measures how much information there is in the data to make inference about the mean value parameters. The quantification of the independent signal depends rather on the covariance between the data points — hence the regression and mutual information ESSs seem to make more sense. A reader might ask how is it possible that a more complex model (nOU) is chosen when the mESS is significantly smaller than n. But the mESS for BM models in these situations is even lower hence the disruptive OU model is favoured. However the α parameter is estimated at the magnitude of -10^9 so effectively this is a Brownian motion. A similar phenomena can be observed in the *Anolis* SSD analysis. With the mean ESS the more complex OU is chosen as $n_e^{\mathrm{E}} \approx n$. This choice is made as under the BM model, $n_e^{\mathrm{E}} \approx 5.199$.

From these results one can draw the conclusion that even with noisy "real–world" data the likelihood should still be expected to dominate. However the mESS will not be a fortunate choice to use especially if the data seem to follow a BM. There is a very good explanation

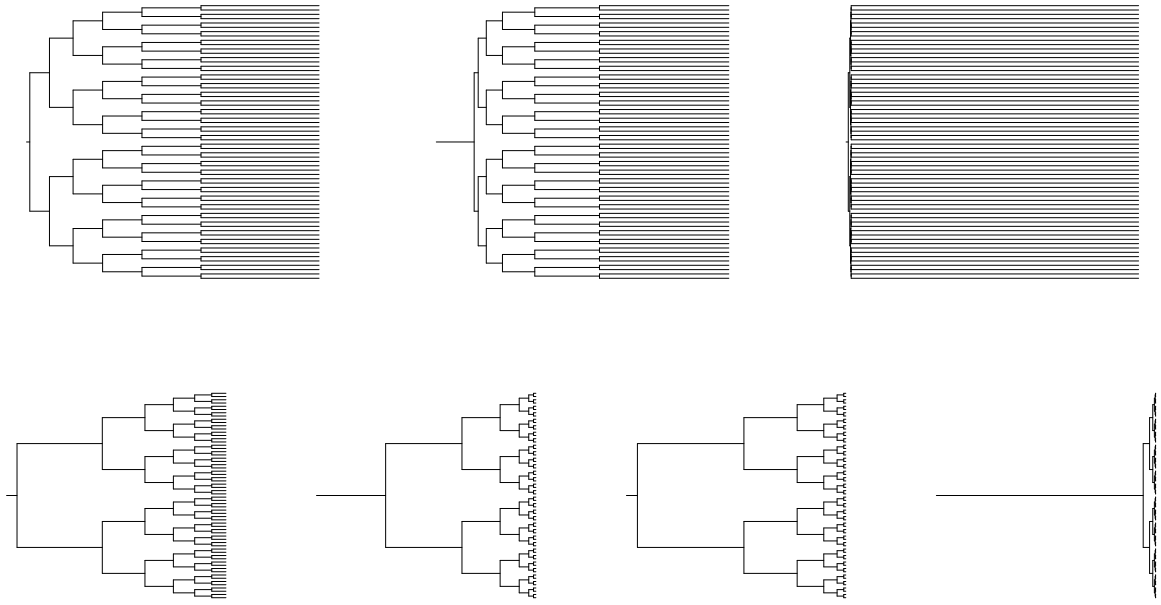

Figure 4: Balanced phylogenies used in pESS for biodiversity simulations. Top row, left to right: branches increasing towards root in a harmonic fashion, branches increasing towards root in a geometric fashion, terminal branches are 99% of tree height. Bottom row, left to right: branches increasing towards root in a harmonic fashion, branches increasing towards root in a geometric fashion, terminal branches are 1% of tree height, root branch is 95% of tree height. The number of tips is 64. For other types of phylogenies see Fig. 1.

for this. Under the phylogenetic BM model inference about the ancestral state are next to impossible from only the contemporary sample. Due to the noise level one cannot obtain consistent estimators of it [Ané, 2008, Bartoszek and Sagitov, 2015a, Sagitov and Bartoszek, 2012]. As the mESS measures the amount of information available to estimate mean parameters and the ancestral state equals the mean in the BM model then n_e^{E} will be small. Hence AIC_c will be high in this case and this model will not be favoured. However with other definitions of pESS Brownian motion is not discriminated in this way. When the true model is the OU one the mESS does not seem to lead to wrong conclusions. This is as in the OU model there is a lot of information about θ [Bartoszek and Sagitov, 2015b] which describes the mean value.

If we look at the turtles and primates results then we can again see that the PD does not tell the full story of diversity. Both have similar relative (and absolute) PDs but their n_e^R are very different. The primates body size follows a Brownian motion and the phylogeny highly correlates contemporary species. The turtles' body size on the other hand follows an OU process and there is much more independence in the data set. This is despite the fact

that when investigating the phylogenies the primates' one has clades diverging further back in the past.

Data set	n	PD$/n$	n_e^{MI}/n	n_e^{E}/n	n_e^{R}/n
Mantellidae male snout–vent length [1]	40 BM[2]	1.364	0.411 BM	0.097 nOU[3]	0.512 BM
Mantellidae Range [4]	40 OU[5]	1.364	1 OU	1 OU	1 OU
Chaerophyllum fruit length [6]	33 OU	0.419	0.634 OU	0.097 nOU	0.834 OU
Carnivores body size [7]	16 BM	1.104	0.712 BM	0.362 BM	0.751 BM
Salamanders body size [12]	197 OU	0.869	0.227 OU	0.059 OU	0.372 OU
Turtles body size [12]	226 OU	0.544	0.228 OU	0.233 OU	0.553 OU
Primates body size [12]	233 BM	0.513	0.181 BM	0.025 BM	0.081 BM
Darwin's finches wing length [12]	13 BM	0.703	0.922 OU	0.292 nOU	0.936 OU
Darwin's finches tarsus length [12]	13 BM	0.703	0.504 BM	0.292 nOU	0.407 BM
Darwin's finches culmen length [12]	13 OU	0.703	0.952 OU	0.293 nOU	0.96 OU
Darwin's finches beak diameter [12]	13 BM	0.703	0.504 BM	0.292 nOU	0.682 OU
Darwin's finches gonys width [12]	13 BM	0.703	0.753 OU	0.292 nOU	0.782 OU
Ducks brightness [13]	38 OU	0.863	0.461 OU	0.592 OU	0.677 OU
Ducks hue [14]	38 OU	0.863	0.492 OU	0.639 OU	0.718 OU
Ducks spacing [14]	38 OU	0.863	0.692 OU	0.84 OU	0.854 OU
Anolis SSD [15]	23 BM	0.97	0.61 BM	1 OU	0.707 BM

Table 3: Results of analysis on real data with different definitions of pESS. In the situation where the OU model with disruptive selection ($\alpha < 0$) the value of α was tiny, about 10^{-9}. Hence these dynamics on the scale of the phylogeny are indistinguishable from a BM.

7 Conclusions

In this study I approached the question of quantifying the amount of independent signal in a phylogenetic data set. I used proposed two definitions of an effective sample size and compared it to the one considered by Ané [2008]. My work is mainly heuristic — to see how do these proposed definitions behave on real and simulated datasets.

The most important goal of my paper is — does it make sense to use information criteria for model selection with data phylogenetically correlated. The most popular information criterion, Akaike's, is an asymptotic one with infinite sample size. Because phylogenetic samples are usually small this was not satisfactory — e.g. more realistic but parameter richer models are rejected in favour of simpler ones. Therefore small sample size corrected criteria were implemented, e.g. the considered here AIC_c (BIC an alternative one). However

[1] [Pabijan et al., 2012] [2] Brownian motion [3] Ornstein–Uhlenbeck with $\alpha < 0$ [5] Ornstein–Uhlenbeck with $\alpha > 0$ [6] [Piwczyński et al., 2015] [7] [Harmon et al., 2008] [13] [Eliason et al., 2014] [15] [Butler and King, 2004]

these corrections were derived under the assumption of independence. One of the aims of this paper is to propose a formula that allows for replacing the sample size with the amount of independent observations and then see if this changes the models indicated by the criterion. In most cases it seems that the likelihood part of information criterion dominates and all definitions of pESS lead to similar conclusions especially with many tip species. This means that for model selection, dependencies in the data do not cause serious problems. However for small phylogenies it seems reasonable to compare the conclusions from different pESS definitions (Tab. 3 Darwin's finches OU conclusion for n_e^R and BM for n).

The second goal of the paper is to quantify the amount of dependency in a phylogenetic sample and to understand patterns associated with it. Obtaining the pESS of clades can indicate clades where more sampling or research effort is needed. For example is a low pESS due to there being really few species or should we expect more or possibly a reclassification of species is needed. Of course all of this is with respect to a specific trait(s). This specificity allows for identification of interesting clades. Considering a trait like body size we obtain the distribution of relative (this is important for comparability between clades) pESSs across a set of clades. In the next step this allows one to identify outlier clades — extremely high or low pESSs for further research. Low relative pESSs could indicate recent radiations or other factors not allowing different species to evolve independently. High relative pESS, especially close to 1, would mean that the species are under completely independent evolutionary pressures. Phylogenetic ESSs of a clade can indicate undersampling of species. If we have high relative pESS with a low absolute number of species then perhaps the very recently evolved species are missing. This can be helpful to indicate where biologists and taxonomists should put efforts to fill in the gaps [Isaac and Purvis, 2004].

A possibly appealing application of this measurement of independence is the quantification of biodiversity. The most commonly used evolutionary measurement tool is phylogenetic diversity — the sum of branch lengths. It seems however that this number does not say much (even when scaled by the number of tip species) about the "value" of an individual species and comparison between clades is difficult (very different ones can have identical values, cf. Tabs. 1 and 2 long terminal with geometric and harmonic increases, or Tab. 3 primates and turtles). Therefore to give the "value" of a single species I propose to use the relative pESS (i.e. n_e/n). If this is low then the loss of a single species does not result in much biodiversity loss — as the other species contain information on it. On the other hand loosing a species when the number is close to 1 results in the loss of a unique entity.

The pESS approach also forces one to define biodiversity in terms of a specific trait — the one described by the stochastic process. This has the advantage of precision — biodiversity is expressed by the variability of specific entities directly linked to species. In a sense the pESS links the concept of a species as both a pattern and process [Lidén and Oxelman, 1989]. The process is the evolving trait, an entity that can be directly observed and measured. The pattern are the pre–identified entities on the phylogenetic tree. On the other hand it has the disadvantage of being specific — one looks only at one (or a couple if it is a suite of traits) dimension of the species.

Quantifying the number of species by the pESS of a clade has the advantage of being

objective and not subject to potentially arbitrary calls. Not splitting a group is compensated by intra–species variability which can be accommodated by the pESS concept. The need to identify exceptional lineages and possibly novel traits associated with them is underlined by Beaulieu and O'Meara [2015]. They discuss this in the context of clade specific increased/decreased speciation rates. The phylogenetic effective sample size allows for direct comparison between clades with respect to traits, e.g. ones suspected/known of contributing to speciation. Outlier values of pESSs will indicate "interesting" groups of species. Such a methodology combines data from multiple sources, morphological (the traits) and genetic (the phylogeny) — a truly multi–omics approach. With the availability of more and more data from diverse sources mathematical methods that integrate them are being developed more and more in the evolutionary biology world [e.g. Bartoszek and Lió, 2014, Solís-Lemus et al., 2014].

Martins and Hansen [1996] point out that one should expect comparative data sets to contain phylogenetic correlations. It is their absence that should be proved. This is a difficult problem in general — to prove dependence or independence. One way would be to use information criteria but it is not clear how many degrees of freedom does the tree have. The relative pESS is an alternative way of showing that phylogenetic correlations are not important. If it is close to 1 then the data set is essentially independent.

Maddison and FitzJohn [2015] regret the lack of a method to quantify the number of pseudoreplicates in a phylogenetically correlated dataset. They point out that the case of discrete traits is even more complicated in this respect. In them it is the unobserved number of independent origins of a trait that matters. Power and p–values, unless one derives model specific tests or uses simulation methods, of e.g. association tests should depend on this number and not on the observed number of species. However as this number is unknown there is "no quantitative correction to apply to these methods" [Maddison and FitzJohn, 2015], e.g. a contingency table test. The concept of the pESS is what Maddison and FitzJohn [2015] seem to be looking for but I considered it here in the continuous trait case. Further work is needed to carry the ideas over to the discrete case. However there is a potential heuristic way of applying the pESS to categorical traits. If one is able to identify continuous traits that are reasonably related to the discrete one and their pESSs are similar then their average can be used as a plug–in for the pESS of the discrete trait in a further downstream analysis/test i.e. an estimator of the number of shifts. The fact that these pESSs are correlated, the traits are dependent through the categorical one and probably between themselves, is actually an advantage. We want the pESS to be nearly the same for each trait and their similarity would indicate sensibility of the described "proxy" approach. If they are dissimilar then this indicates the need for further investigation, especially choice of traits. This is of course only a suggestion that might alleviate the problem and it needs further study alongside the development of models where continuous and categorical traits jointly co–evolve.

The phylogenetic ESS definitions are also interesting from a statistical point of view. The mean ESS measures the amount of information on the mean value and hence often results in a small pESS. This is especially in the BM case where there is limited information

on the ancestral state. From all the simulations presented it seems that the regression ESS seems to capture what was aimed at the beginning of the study — the amount of independent observations in the data. This is not surprising as by construction it adds up the variance of the independent residuals. Both of these definitions can be used for non–normal processes but we should not expect the regression ESS to be so effective. Rather it would only measure the amount of linearly independent observations. In a general case I suggest the mutual information ESS but here work still needs to be done on defining an appropriate $e(\cdot)$ transformation in order for $n_e^{MI} \in [1, n]$ to be in agreement with n_e^R for normal samples.

Acknowledgments

I would like to thank Chad Eliason for providing me with the ducks data and Marcin Piwczyński for the Mantellidae and *Chaerophyllum* data. I would like to thank Bengt Oxelman and Tanja Stadler for encouragement and numerous comments on this work.

References

R. Adamczak and P. Miłoś. U–statistics of Ornstein–Uhlenbeck branching particle system. *J. Th. Probab.*, 27(4):1071–1111, 2014.

R. Adamczak and P. Miłoś. CLT for Ornstein–Uhlenbeck branching particle system. *Elect. J. Probab.*, 20(42):1–35, 2015.

P. Agapow. Species: demarcation and diversity. In A. Purvis, J. L. Gittleman, and T. Brooks, editors, *Phylogeny and Conservation*, pages 57–75. Cambridge University Press, Cambridge, 2005.

P. Agapow, K. A. Crandall, J. L. Gittleman, G. M. Mace, J. C. Marshall, and A. Purvis. The impact of species concept on biodiversity studies. *Quart. Rev. Biol.*, 79(2):161–179, 2004.

C. Ané. Analysis of comparative data with hierarchical autocorrelation. *Ann. Appl. Stat*, 2 (3):1078–1102, 2008.

C. Ané, L. S. T. Ho, and S. Roch. Phase transition on the convergence rate of parameter estimation under an Ornstein–Uhlenbeck diffusion on a tree. *ArXiv e-prints*, 2014.

J. C. Avise. Phylogenetic units and currencies above an below the species level. In A. Purvis, J. L. Gittleman, and T. Brooks, editors, *Phylogeny and Conservation*, pages 76–100. Cambridge University Press, Cambridge, 2005.

K. Bartoszek. Quantifying the effects of anagenetic and cladogenetic evolution. *Math. Biosci.*, 254:42–57, 2014.

K. Bartoszek and P. Lió. A novel algorithm to reconstruct phylogenies using gene sequences and expression data. In *Int. Proc. Chem. Biol. Env. Eng.; Env. Energy and Biotech. III*, volume 70, pages 8–12, 2014.

K. Bartoszek and S. Sagitov. A consistent estimator of the evolutionary rate. *J. Theor. Biol.*, 371:69–78, 2015a.

K. Bartoszek and S. Sagitov. Phylogenetic confidence intervals for the optimal trait value. *J. App. Prob.*, 52(4), 2015b.

K. Bartoszek, J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen. A phylogenetic comparative method for studying multivariate adaptation. *J. Theor. Biol.*, 314:204–215, 2012.

J. M. Beaulieu, D.-C. Jhwueng, C. Boettiger, and B. C. O'Meara. Modeling stabilizing selection: Expanding the Ornstein–Uhlenbeck model of adaptive evolution. *Evolution*, 66: 2369–2389, 2012.

J.M. Beaulieu and B.C. O'Meara. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *bioRxiv e-prints*, 2015. doi: `http://dx.doi.org/10.1101/016386`.

F. Bokma. Detection of punctuated equilibrium from molecular phylogenies. *J. Evol. Biol.*, 15:1048–1056, 2002.

F. Bokma. Detection of "punctuated equilibrium" by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. *Evolution*, 62(11):2718–2726, 2008.

T. M. Brooks, J. D. Pilgrim, A. S. L. Rodrigues, and G. A. B. Da Fonseca. Conservation status and geographic distribution of avian evolutionary history. In A. Purvis, J. L. Gittleman, and T. Brooks, editors, *Phylogeny and Conservation*, pages 267–294. Cambridge University Press, Cambridge, 2005.

S. H. M. Butchart and J. P. Bird. Data deficient birds on the IUCN Red List: what we don't know and why does it matter? *Biol. Conserv.*, 143:239–247, 2010.

M. A. Butler and A. A. King. Phylogenetic comparative analysis: a modelling approach for adaptive evolution. *Am. Nat.*, 164(6):683–695, 2004.

J. Clavel, G. Escarguel, and G. Merceron. mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data. *Meth. Ecol. Evol.*, . in press.

F. W. Crawford and M. A. Suchard. Diversity, disparity, and evolutionary rate estimation for unresolved Yule trees. *Syst. Biol.*, 62(3):439–455, 2013.

C. E. Cressler, M. A. Butler, and A. A. King. Detecting adaptive evolution in phylogenetic comparative analysis using the Ornstein–Uhlenbeck model. *Syst. Biol.*, 2015.

R. H. Crozier. Preserving the information content of species: Genetic diversity, phylogeny and concervation worth. *Annu. Rev. Ecol. Syst.*, 28:243–268, 1997.

C. M. Eliason, R. Maia, and M. D. Shawkey. Modular color evolution facilitated by a complex nanostructure in birds. *Evolution*, 69(2):357–367, 2014.

C. Faes, G. Molenberghs, M. Aerts, G. Verbeke, and M. G. Kenward. The effective sample size and an alternative small–sample degrees–of–freedom method. *Am. Stat.*, 63(4):389–399, 2009.

D. P. Faith. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.*, 61:1–10, 1992.

D. P. Faith. Phylogenetic pattern and the quantification of organismal biodiversity. *Philos. Trans. Roy. Soc. B*, 345:45–58, 1994.

D. P. Faith. Quantifying biodiversity: A phylogenetic perspective. *Conserv. Biol.*, 16:248–252, 2002.

J. Felsenstein. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *Am. Nat.*, 171:713–725, 2008.

Garland, T., Jr., A. W. Dickerman, C. M. Janis, and J. A. Jones. Phylogenetic analysis of covariance by computer simulation. *Syst. Biol.*, 42:265–292, 1993.

T. Gernhard. The conditioned reconstructed process. *J. Theor. Biol.*, 253:769–778, 2008a.

T. Gernhard. New analytic results for speciation times in neutral models. *B. Math. Biol.*, 70:1082–1097, 2008b.

T. F. Hansen and K. Bartoszek. Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Syst. Biol.*, 61(3):413–425, 2012.

T. F. Hansen and S. H. Orzack. Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. *Evolution*, 59(10): 2063–2072, 2005.

T. F. Hansen, J. Pienaar, and S. H. Orzack. A comparative method for studying adaptation to a randomly evolving environment. *Evolution*, 62:1965–1977, 2008.

L. J. Harmon, J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. GEIGER: investigating evolutionary radiations. *Bioinformatics*, 24:129–131, 2008.

C. M. Hurvich and C. L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.

T. Ingram and D. L. Mahler. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein–Uhlenbeck models with stepwise Akaike Information Criterion. *Meth. Ecol. Evol.*, 4(5):416–425, 2013.

N. J. B. Isaac and A. Purvis. The 'species problem' and testing macroevolutionary hypotheses. *Diversity Distrib.*, 10:275–281, 2004.

W. Jetz and R. P. Freckleton. Towards a general framework for predicting threat status of data–deficient species from phylogenetic, spatial and environmental information. *Phil. Trans. R. Soc. B.*, 370:20140016, 2015.

I. Koch. *Analysis of Multivariate and High–Dimensional Data*. Cambridge University Press, Cambridge, 2014.

M. Lidén and B. Oxelman. Species—pattern or process? *Taxon*, 38(2):228–232, 1989.

X. Lin, M. Steel, J. Pittman, and B. Clarke. Information conversion, effective samples, and parameter size. *IEEE. Trans. Inf. Theory*, 53(12):4438–4456, 2007.

W. P. Maddison and R. G. FitzJohn. The unsolved challenge to phylogenetic correlation tests for categorical characters. *Syst. Biol.*, 64(1):127–136, 2015.

E. P. Martins and T. F. Hansen. The statistical analysis of interspecific data: A review and evaluation of phylogenetic comparative methods. In E. P. Martins, editor, *Phylogenies and the Comparative Method in Animal Behaviour*, pages 22–75. Oxford University Press, 1996.

A. Mooers, O. Gascuel, T. Stadler, H. Li, and M. Steel. Branch lengths on birth-death trees and the expected loss of phylogenetic diversity. *Syst. Biol.*, 61(2):195–203, 2012.

A. Ø. Mooers, S. B. Heard, and E. Chrostowski. Evolutionary heritage as a metric for conservation. In A. Purvis, J. L. Gittleman, and T. Brooks, editors, *Phylogeny and Conservation*, pages 120–138. Cambridge University Press, Cambridge, 2005.

S. Nee and R. M. May. Extinction and loss of evolutionary history. *Science*, 278:692–694, 1997.

C. L. Nunn. *The Comparative Approach in Evolutionary Anthropology and Biology*. The University of Chicago Press, Chicago, 2011.

M. Pabijan, K. C. Wollenberg, and M. Vences. Small body size increases the regional differentiation of populations of tropical mantellid frogs (Anura: Mantellidae). *J. Evol. Biol.*, 25:2310–2324, 2012.

M. D. Pagel. Seeking the evolutionary regression coefficent: An analysis of what compaarative methods measure. *J. Theor. Biol.*, 164:191–205, 1993.

E. Paradis. *Analysis of Phylogenetics and Evolution with R*. Springer, New York, 2012.

M. Piwczyński, R. Puchałka, and K. Spalik. The infrageneric taxonomy of Chaerophyllum (Apiaceae) revisited: new evidence from nrDNA ITS sequences and fruit anatomy. *Bot. J. Linn. Soc.*, 178:298–313, 2015.

A. Purvis, J. L. Gittleman, and T. Brooks, editors. *Phylogeny and Conservation*. Cambridge University Press, Cambridge, 2005.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org`.

R. V. Rohlfs, P. Harrigan, and R. Nielsen. Modeling gene expression evolution with an extended Ornstein–Uhlenbeck process accounting for within-species variation. *Mol. Biol. Evol.*, 31(1):201–211, 2013.

S. Sagitov and K. Bartoszek. Interspecies correlation for neutrally evolving traits. *J. Theor. Biol.*, 309:11–19, 2012.

G. Schwarz. Estimating the dimension of a model. *Ann. Stat.*, 5(2):461–461, 1978.

R. J. Smith. Degrees of freedom in interspecific allometry: An adjustment for the effects of phylogenetic constraint. *Am. J. Phys. Anthropol.*, 93:95–107, 1994.

C. Solís-Lemus, L. L. Knowles, and C. Ané. Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, 69(2):492–507, 2014.

T. Stadler. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *J. Theor. Biol.*, 261(1):58–68, 2009.

T. Stadler. Simulating trees with a fixed number of extant species. *Syst. Biol.*, 60(5):676–684, 2011.

T. Stadler and M. Steel. Distribution of branch lengths and phylogenetic diversity under homogeneous speciation models. *J. Theor. Biol.*, 297:33–40, 2012.

J. C. Uyeda and L. J. Harmon. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Syst. Biol.*, 63(6): 902–918, 2014.

Phylogenetic effective sample size
Supplementary histograms

Krzysztof Bartoszek

Department of Mathematics, Uppsala University, 751 06 Uppsala, Sweden.
bartoszekkj@gmail.com

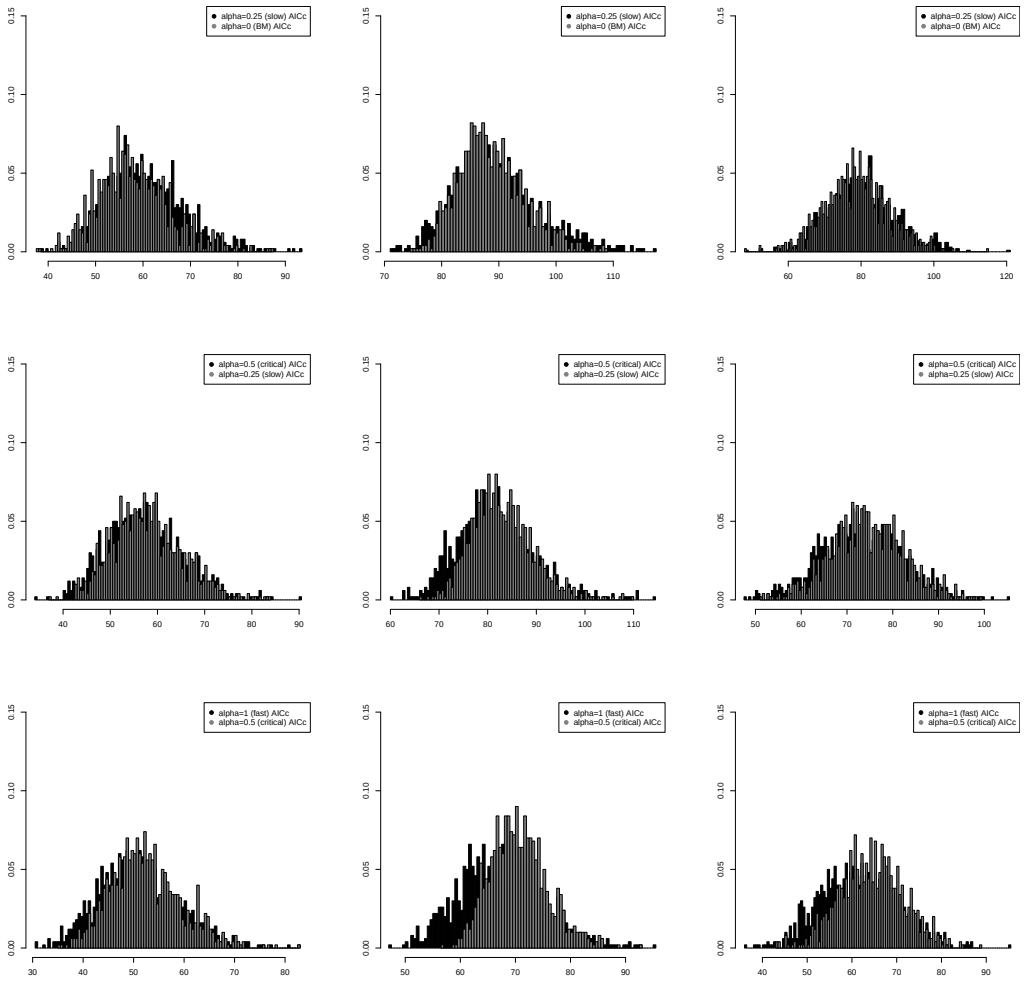

Figure S.1: Histograms of AIC_c values with n_e^{MI} effective sample size correction for different types of trees and evolutionary processes. The sample sizes are $n = 30$ (left unbalanced tree and Yule) and $n = 32$ (balanced tree). First column: balanced tree, second column: left unbalanced tree, third column: 1000 pure–birth Yule trees ($\lambda = 1$). The balanced trees and unbalanced trees were generated using the function stree() of the R ape package, the Yule trees by the TreeSim R package. First row: Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Brownian motion ($X_0 = 0$, $\sigma^2 = 1$ gray alternative model), second row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model), third row: Ornstein–Uhlenbeck process ($\alpha = 1$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), fourth row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model). We simulate data under both the true and alternative evolutionary models 1000 times and then calculate AIC_c values for each simulated pair.

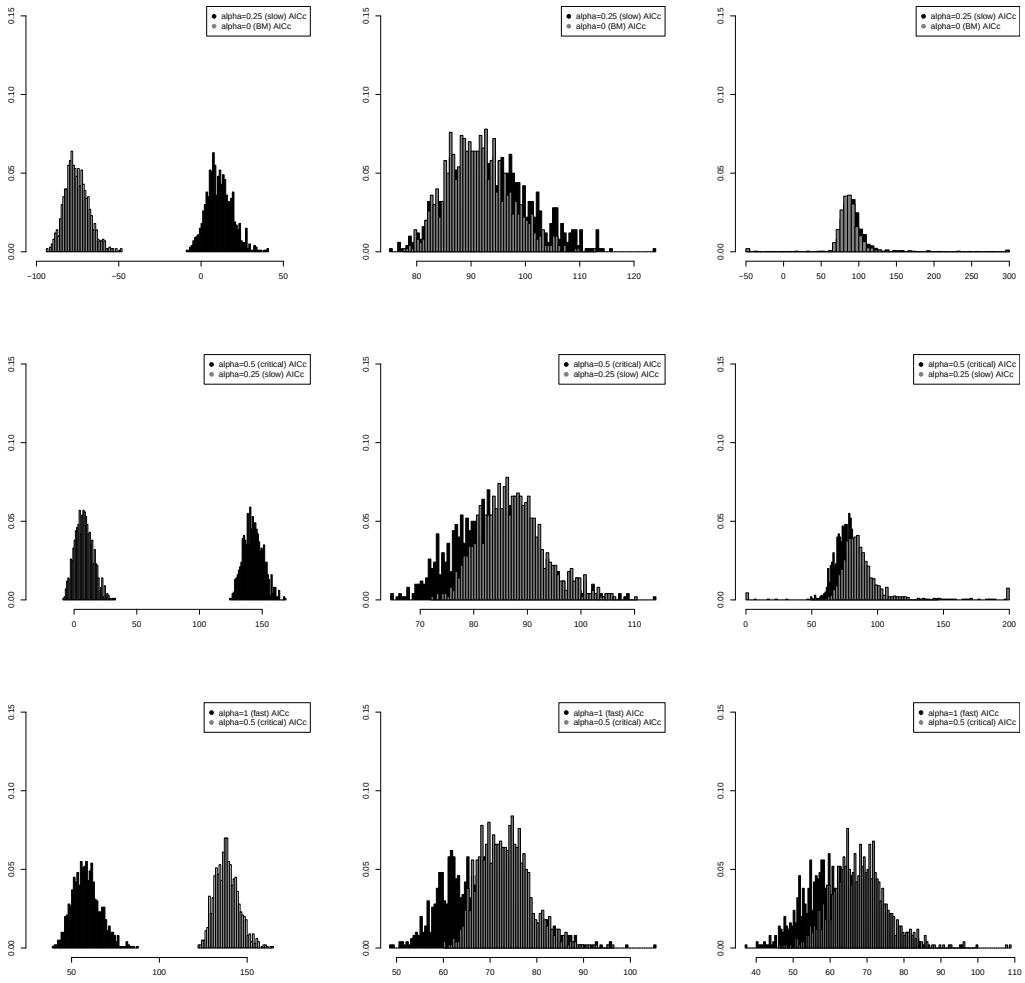

Figure S.2: Histograms of AIC_c values with n_e^E effective sample size correction for different types of trees and evolutionary processes. The sample sizes are $n = 30$ (left unbalanced tree and Yule) and $n = 32$ (balanced tree). First column: balanced tree, second column: left unbalanced tree, third column: 1000 pure–birth Yule trees ($\lambda = 1$). The balanced trees and unbalanced trees were generated using the function stree() of the R ape package, the Yule trees by the TreeSim R package. First row: Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Brownian motion ($X_0 = 0$, $\sigma^2 = 1$ gray alternative model), second row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model), third row: Ornstein–Uhlenbeck process ($\alpha = 1$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), fourth row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model). We simulate data under both the true and alternative evolutionary models 1000 times and then calculate AIC_c values for each simulated pair.

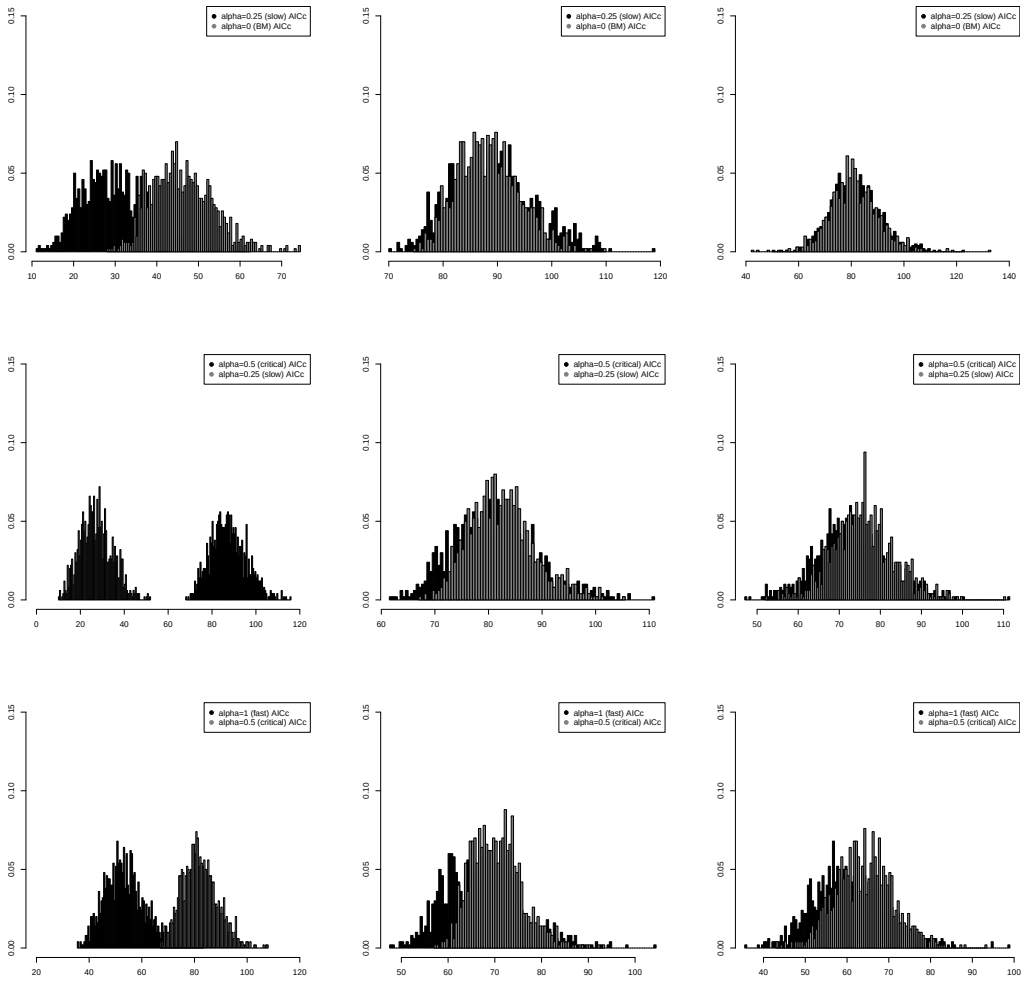

Figure S.3: Histograms of AIC_c values with n_e^R effective sample size correction for different types of trees and evolutionary processes. The sample sizes are $n = 30$ (left unbalanced tree and Yule) and $n = 32$ (balanced tree). First column: balanced tree, second column: left unbalanced tree, third column: 1000 pure–birth Yule trees ($\lambda = 1$). The balanced trees and unbalanced trees were generated using the function stree() of the R ape package, the Yule trees by the TreeSim R package. First row: Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Brownian motion ($X_0 = 0$, $\sigma^2 = 1$ gray alternative model), second row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model), third row: Ornstein–Uhlenbeck process ($\alpha = 1$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), fourth row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model). We simulate data under both the true and alternative evolutionary models 1000 times and then calculate AIC_c values for each simulated pair.

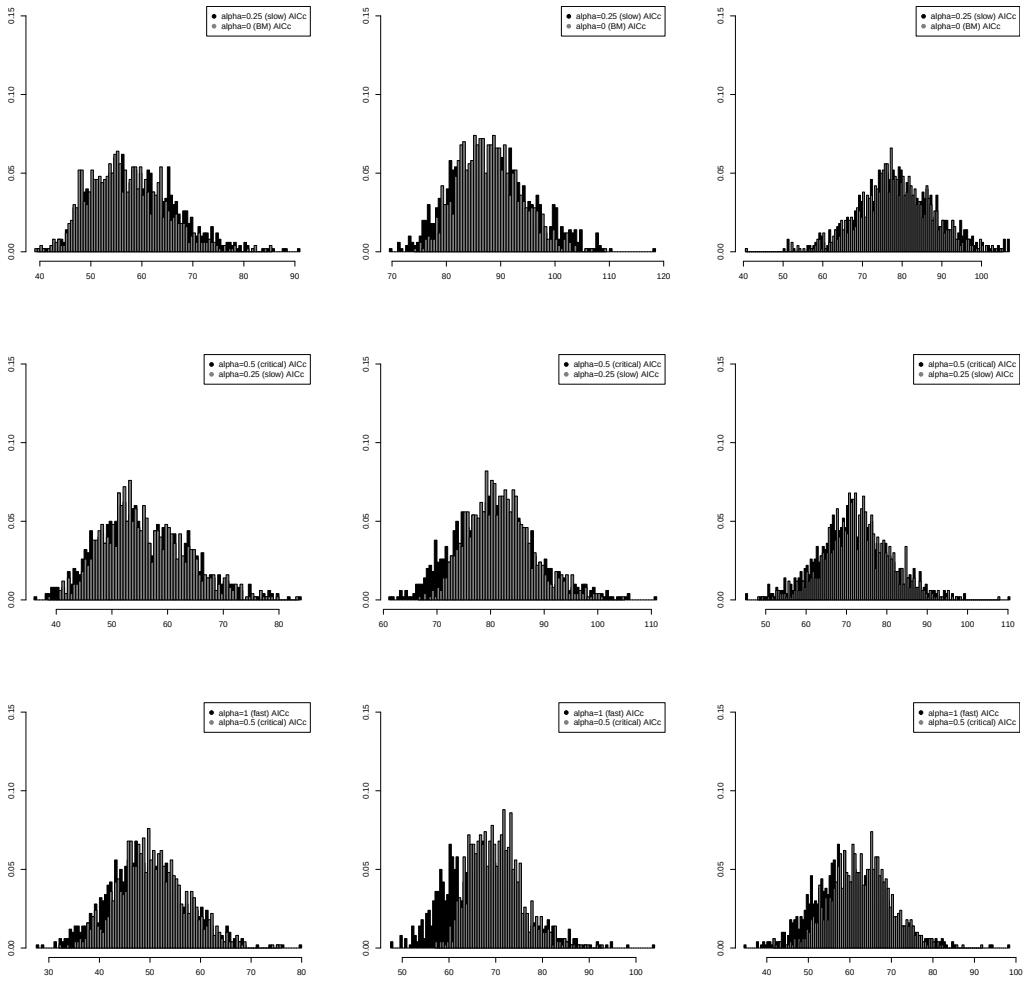

Figure S.4: Histograms of AIC_c values with no effective sample size correction for different types of trees and evolutionary processes. The sample sizes are $n = 30$ (left unbalanced tree and Yule) and $n = 32$ (balanced tree). First column: balanced tree, second column: left unbalanced tree, third column: 1000 pure–birth Yule trees ($\lambda = 1$). The balanced trees and unbalanced trees were generated using the function stree() of the R ape package, the Yule trees by the TreeSim R package. First row: Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Brownian motion ($X_0 = 0$, $\sigma^2 = 1$ gray alternative model), second row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model), third row: Ornstein–Uhlenbeck process ($\alpha = 1$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), fourth row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model). We simulate data under both the true and alternative evolutionary models 1000 times and then calculate AIC_c values for each simulated pair.

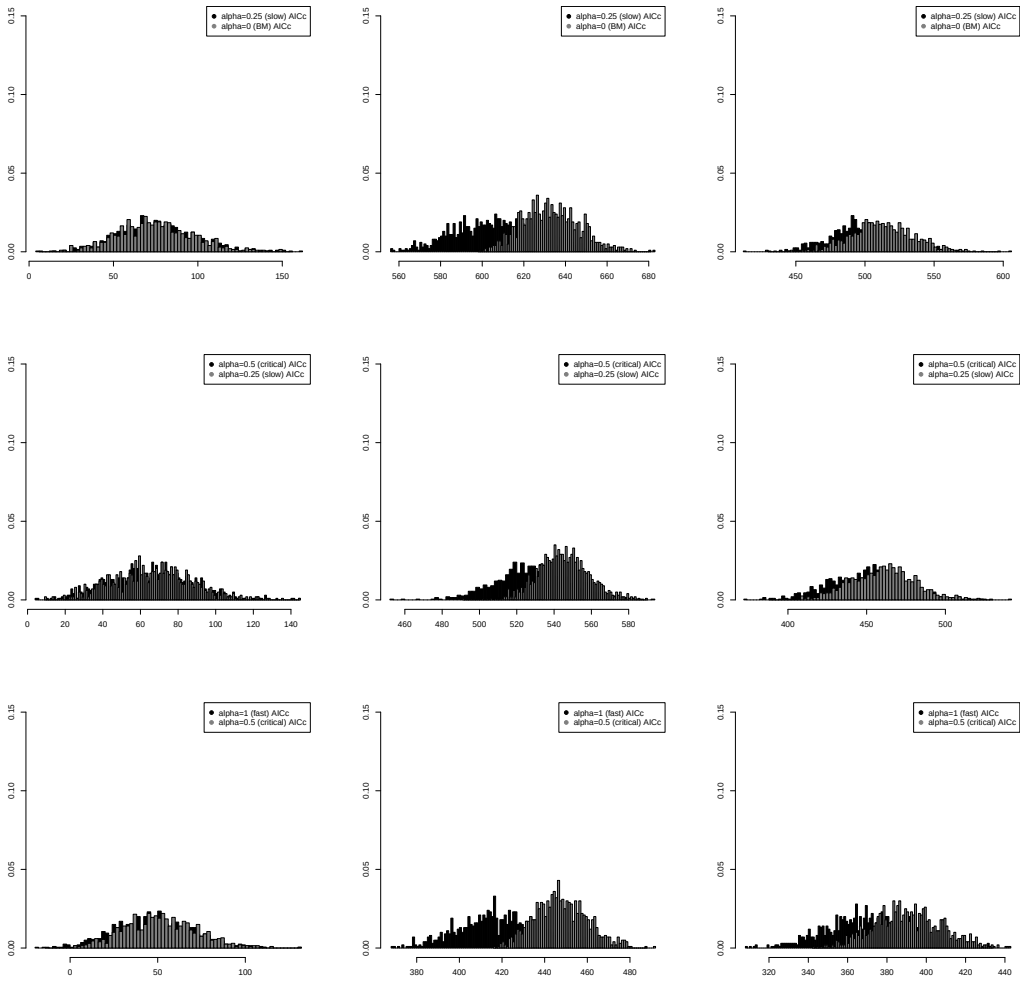

Figure S.5: Histograms of AIC_c values with n_e^{MI} effective sample size correction for different types of trees and evolutionary processes. The sample sizes are $n = 205$ (left unbalanced tree and Yule) and $n = 256$ (balanced tree). First column: balanced tree, second column: left unbalanced tree, third column: 1000 pure–birth Yule trees ($\lambda = 1$). The balanced trees and unbalanced trees were generated using the function stree() of the R ape package, the Yule trees by the TreeSim R package. First row: Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Brownian motion ($X_0 = 0$, $\sigma^2 = 1$ gray alternative model), second row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model), third row: Ornstein–Uhlenbeck process ($\alpha = 1$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), fourth row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model). We simulate data under both the true and alternative evolutionary models 1000 times and then calculate AIC_c values for each simulated pair.

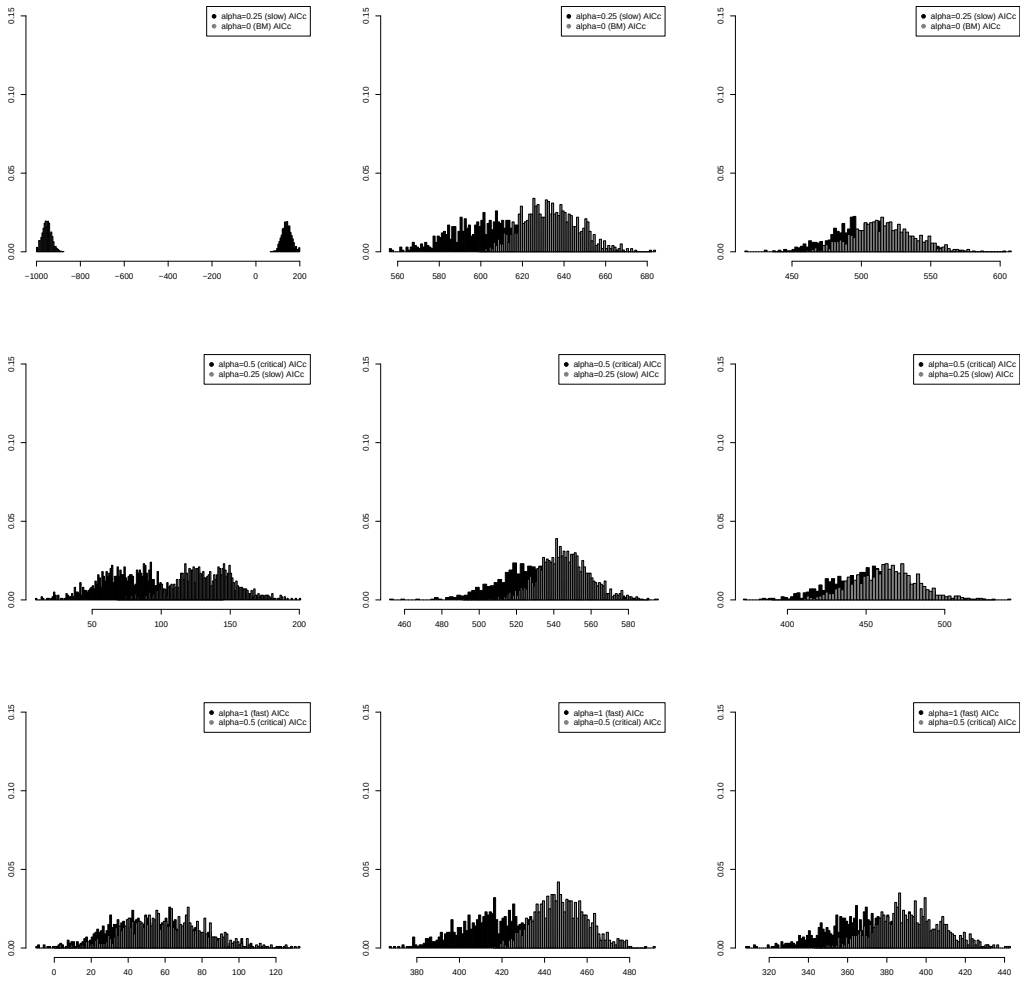

Figure S.6: Histograms of AIC_c values with n_e^E effective sample size correction for different types of trees and evolutionary processes. The sample sizes are $n = 205$ (left unbalanced tree and Yule) and $n = 256$ (balanced tree). First column: balanced tree, second column: left unbalanced tree, third column: 1000 pure–birth Yule trees ($\lambda = 1$). The balanced trees and unbalanced trees were generated using the function stree() of the R ape package, the Yule trees by the TreeSim R package. First row: Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Brownian motion ($X_0 = 0$, $\sigma^2 = 1$ gray alternative model), second row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model), third row: Ornstein–Uhlenbeck process ($\alpha = 1$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), fourth row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model). We simulate data under both the true and alternative evolutionary models 1000 times and then calculate AIC_c values for each simulated pair.

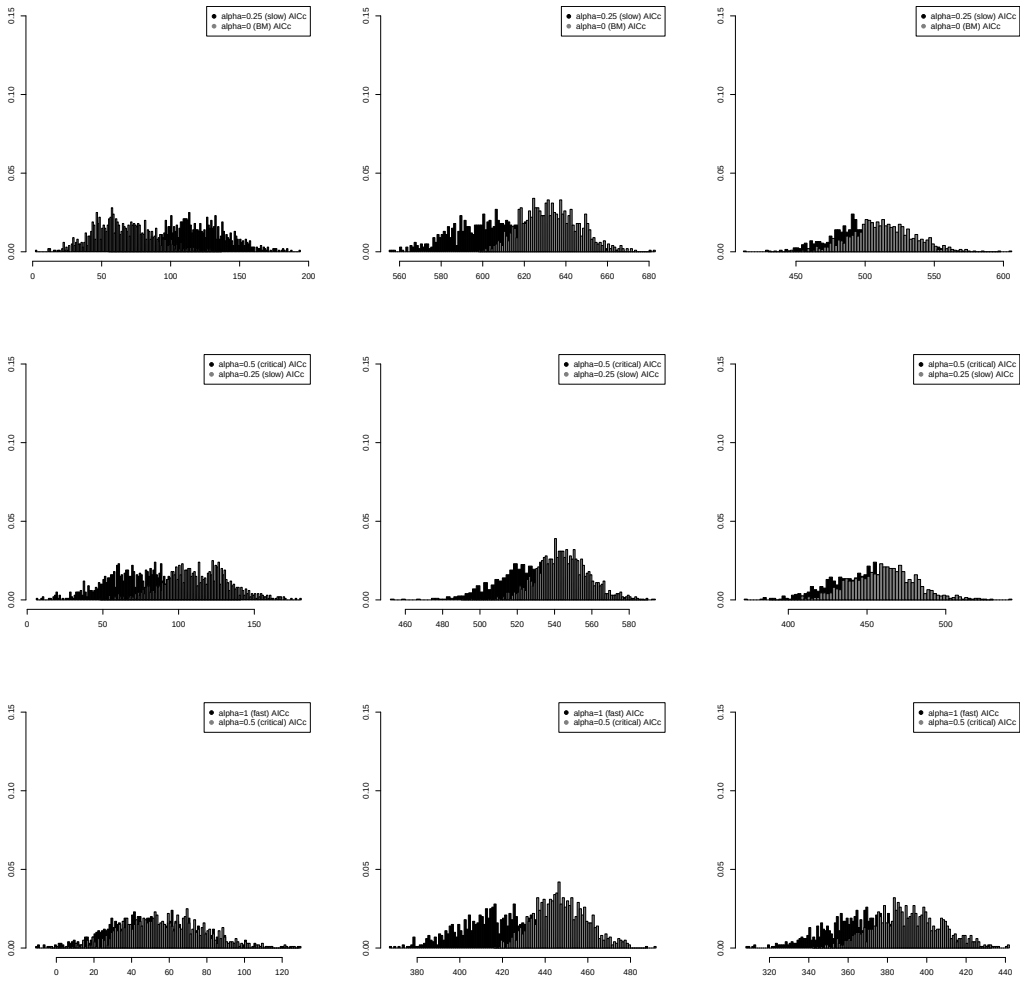

Figure S.7: Histograms of AIC_c values with n_e^R effective sample size correction for different types of trees and evolutionary processes. The sample sizes are $n = 205$ (left unbalanced tree and Yule) and $n = 256$ (balanced tree). First column: balanced tree, second column: left unbalanced tree, third column: 1000 pure–birth Yule trees ($\lambda = 1$). The balanced trees and unbalanced trees were generated using the function stree() of the R ape package, the Yule trees by the TreeSim R package. First row: Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Brownian motion ($X_0 = 0$, $\sigma^2 = 1$ gray alternative model), second row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model), third row: Ornstein–Uhlenbeck process ($\alpha = 1$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), fourth row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model). We simulate data under both the true and alternative evolutionary models 1000 times and then calculate AIC_c values for each simulated pair.

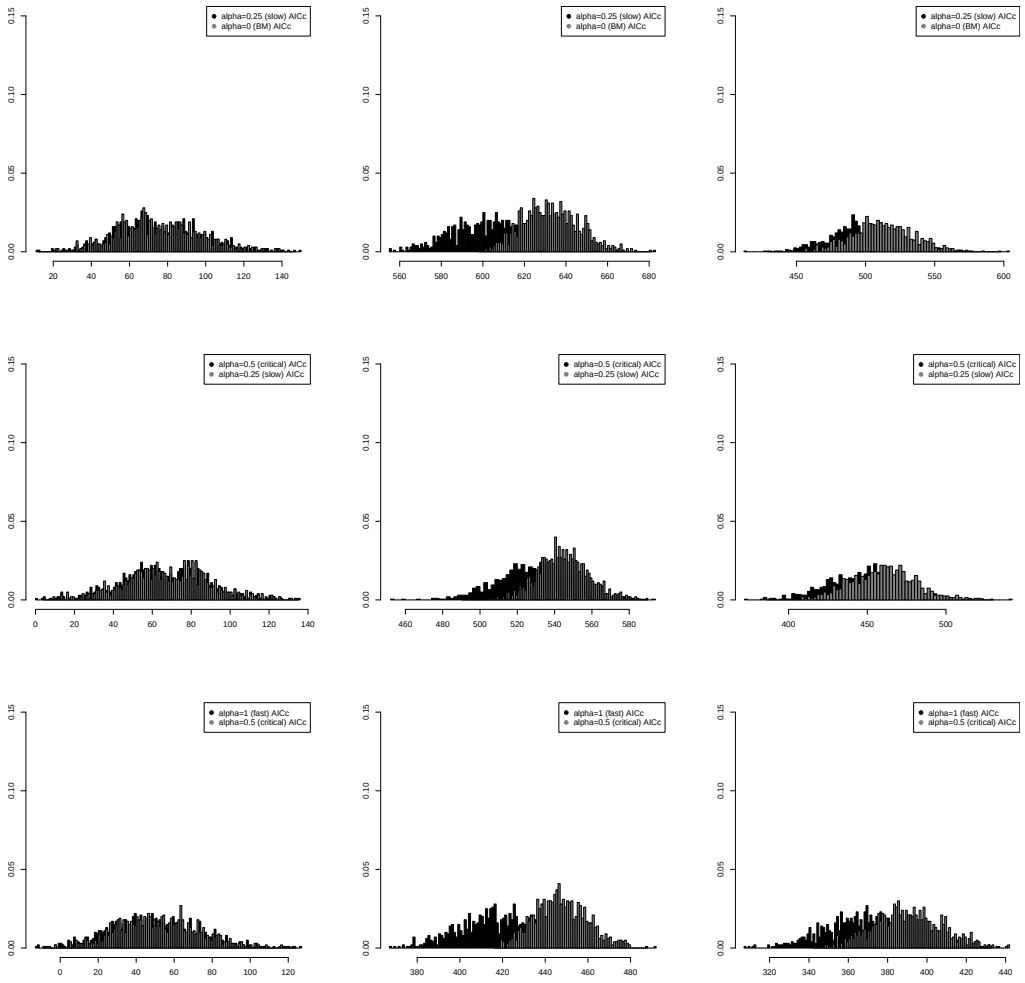

Figure S.8: Histograms of AIC_c values with no effective sample size correction for different types of trees and evolutionary processes. The sample sizes are $n = 205$ (left unbalanced tree and Yule) and $n = 256$ (balanced tree). First column: balanced tree, second column: left unbalanced tree, third column: 1000 pure–birth Yule trees ($\lambda = 1$). The balanced trees and unbalanced trees were generated using the function stree() of the R ape package, the Yule trees by the TreeSim R package. First row: Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Brownian motion ($X_0 = 0$, $\sigma^2 = 1$ gray alternative model), second row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), Ornstein–Uhlenbeck process ($\alpha = 0.25$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model), third row: Ornstein–Uhlenbeck process ($\alpha = 1$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ black true model), fourth row: Ornstein–Uhlenbeck process ($\alpha = 0.5$, $\sigma^2 = 1$, $X_0 = 0$, $\theta = 0$ gray alternative model). We simulate data under both the true and alternative evolutionary models 1000 times and then calculate AIC_c values for each simulated pair.