On the condition of phylogenetic tree for comparative analysis

Phylogenetic comparative methods incorporate phylogenetic tree to study the relationship among a group of related species. A phylogenetic tree of *n* taxa can be algebraically transformed into an *n* by *n* squared phylogenetic relationship matrix *C* where each element in *C* represent the affinity between extant species *i* and extant species *j*; in other word, equals to the branch length from the root to MRCA(*i*,*j*) (the most recent common ancestor of species *i* and species *j*). Analyses later on to answer the question using comparative analysis heavily rely on the matrix *C*. The condition number of the matrix *C* defined by the ratio of the maximum eigenvalue of *C* to the minimum eigenvalue of C plays an important role for further analysis. While *C* is an ill-conditioned matrix with fairly large value of , results obtained from subsequent analyses may not be reliable. To remediate this issue, we propose several methods to appropriately adjust the tree for obtaining a better condition matrix *C* to enhance the adequacies in the next step analysis.

**Keyword**: phylogenetic comparative analysis, trait evolution, condition number.

*Introduction*

Consider a rather large comparative methods study where the goal is to look at the bound of the reasonable in real data. We can explore the problem from the simulation and from studying the real trees. To do this, we can try to obtain a measure for summarizing some aspects of phylogenies in order to know how this value in real phylogenies stacks up the values obtained from simulated trees. One goal is to investigate the statistical properties of real phylogenies (i.e. trees estimated from molecular data under substitution models) and simulated trees generated in various ways. To our knowledge, prior comparisons have been more verbal and have focused on the *shape* (such as tree balance) of simulated trees and whether they seemed reasonable when compared to the shapes of trees estimated from data. But we haven’t seen any comparison of statistical properties of the trees themselves in the literature and that is why we are investigating and why we needed a reference set of actual chronograms. We will focus more on the algebraic/statistical side where we will look at the condition number of the phylogenetic covariance matrix (Jhwueng 2013), called , for these trees, as well as for tree simulated two different ways. For positive definite matrix, the condition number of a covariance matrix is defined by the ratio of max/min eigenvalues of that matrix:

where and are the maximum and minimum eigenvalues set and satisfies

Basically condition number is a measure of how stable the matrix is for subsequence operations. Small condition numbers are more stable covariance matrices, while larger numbers are less stable. More stable covariance matrices are generally good, as with them there is less error in downstream algebraic operations, using that covariance matrix (or its inverse), such as: data multiplication, projection, linear model prediction, and even simulating data using that covariance matrix. By contrast, large condition numbers mean these operations are unstable and more prone to error propagation.

*Data Collection*

We use the latest updated (Aug. 09, 2016) data in the *Tree of Life* (Maddison and Schulz 2007). Note that wrinkle is that the trees uploaded in *OpenTree* (Hincliff et al. 2015) can sometimes have rounding error: they pass as ultrametric criteria with a tolerance of 0.01 but not all pass at R package: *Ape* (Paradis et al. 2004) ’s

more stringent default tolerance for ultrametric criteria. The code (the functions in [opentree\_chronograms.R](https://github.com/phylotastic/datelife/blob/master/R/opentree_chronograms.R) ) in *phylotastics* (Stoltzfus et al. 2013) will be especially useful to investigate this. We’d have to go back to *OpenTree* to figure out why: could be the issue from *beast* (Drummond et al. 2012) vs. *r8s* (Sanderson 2002) branch lengths, different calibrations, etc.

We first converting those trees that could be converted into chronograms using the method in Boettiger and Land (2012) for interfacing with *TreeBASE* (Sanderson et al. 1994, Piel et al. 2002). However, trees in *TreeBASE* do not have branch lengths so that tree measures could not be evaluated while *OpenTree* still does not have as many as tree as *TreeBASE*, but at least they are labeled as chronograms (even ultrametric trees in *TreeBASE* might not be chronograms: upgma [unweighted pair group method with arithmetic mean], … etc.). We instead access its trees using the *rotl* (Michonneau et al. 2016) package where we can get its cache of 126 chronograms that includes sometimes multiple from one study from *OpenTree* by installing the *datelife* (O’Meara et al. in preparation) and *phylotastic* to pull them in. We also use *rotl* to bring in all open tree trees with branch lengths and then scale them once they are loaded (setting root age to 1). The tree size ranges 6 taxa to 682 taxa with the median size of 72 taxa and mean size of 622 taxa. All the trees are labeled with the citation for the study.

In addition, we try to fix the issue of these trees. In particular, for Tree #93 (Johnson et al. 2011) for a binary state speciation and extinction (BiSSE) model on plants, we are not sure why that tree might be weird, but there could be issues with going from the original file to *OpenTree* to *datelife*: could be worth comparing to the original tree. Note that it can now have more than one tree per study. For example, Hedges et al. (2006) *TimeTree* is has one tree of matched names and a much large tree with all the names from their paper. In such cases, the trees may not be independent. We then check the tree. One thing we found is that there is a tree that has a single branch length as ’*NaN*’.

Now, with respect to our study, the question we have is what to do with it? An obvious thing is to just throw it out, as BM multivariate data cannot be simulated on this tree (which is the start-point of our survey). However, since it passes ultrametric at tolerance 0.01, our suspicion is that a zero-length branch would be a decent substitute. However, we found that putting zero branch length will make worse for further study. Although this issue is limited to just one tree which really is a single-point fix, not a systemic issue, this tree is inferred from Bayesian inference which is not prone to negative branch length. One possibility that leads to this issue might be due to rooting. If we ignore that tree, then it will leave us 126 trees with 512 or fewer taxa in each. That gives us a very nice, and large set of published chronograms with which to work.

*Preliminary analyses*

For case of macro-evolutionary studies, this comes into play during the phylogenetic comparative methods analytics, as many of these methods must obtain the inverse of for use in the calculations. Figure 1 shows that the condition number vs. number of taxa for all 126 chronograms in the database. These fall nicely in a tight increasing pattern with number of taxa (the variance among studies also increases with taxa as one might expect). So this is what real trees look like when viewed by the *stability* of their phylogenetic covariance matrix. Stability decreases as number of taxa increases, but does so somewhat slowly.

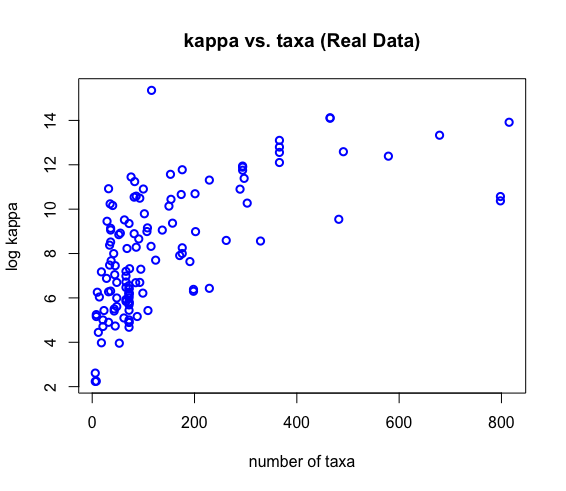


Figure 1: The condition number vs. number of taxa for trees from literature.

Figure 2 shows the conditioned number vs. taxa using tree from simulation. Now there are 3 lines in the plot in Figure 2. These represent the median of 100 runs of simulated phylogenies at different number of taxa. The **blue line** was obtained from random trees using random splits trees method (implemented using *rtree* (Paradis et al. 2004) along with computed BM to make a chronogram), and the **black line** was obtained trees simulated by birth-death trees with a given age on a fixed number of extant taxa (Stadler 2009) using birth rate , death rate , and tree of age 15.35. The **red line** is obtained from trees under a pure-birth process (birth=1, death=0).

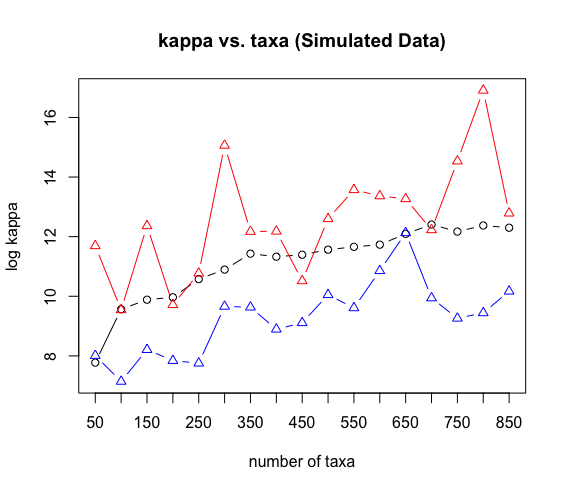


Figure 2: The condition number vs. number of taxa for trees from simulation.

This is particularly curious given that the process we are using to simulate the pure-birth tree makes a lot of intuitive sense biologically, and generally, people in evolutionary biology community seem to think that pure birth and birth-death trees are thus a reasonable way to simulate the process (note: birth-death trees do something similar to the line shown on the plot for pure birth trees, but as shown in Figure 2, we see the contrary where the pure birth tree behave unstably. In other words, compare to Figure 1, we observed that real chronograms are more statistically ‘stable’ than are trees generated from pure-birth simulators (**red line** in Figure 2). That is a very surprising result, and has quite a few consequences; particular for folks using pure birth or birth-death tree as inputs for simulations that then evaluate comparative patterns or even comparative methods themselves. In short, if the trees are unstable statistically, then either the data generated from these simulations could be unstable, the subsequent PCM analyses which use the inverse of C (such as phylogenetic generalized lest square (PGLS)) could be unstable, or both. And these consequences would be expected to be more acute for multivariate data, though it remains under investigation. Further systematic examination of the birth vs. death parameters remains under investigation as well.

So that is the motivational interest of ours in the repository of chronograms. We would like to see how the chronograms behave relative to the simulated trees, in terms of statistical stability. In addition, Jhwueng and O’Meara (2016) developed a PCM for phylogenetic networks, rather than trees, where abnormal likelihoods were observed during the search for the maximum likelihood estimates. They figured it was due to numerical issues in matrix condition. To get around this in various ways, they tried adding variance to the diagonals to make the matrix have better condition, calculate the likelihood, repeat with decreasing additional variance, extrapolate what the likelihood would be with no additional variance, but still having issues to get convergent estimates. Similarly, the weird issue could also rise with OUmva model (Beaulieu et al. 2012) or other OU models (Hansen et al. 2008; Jhwueng and Maroulas 2016) may also come from this.

From above, it is very likely that matrix condition affects a broader swatch of PCMs than anyone has considered. In fact, this issue has not been well recognized within the community, and of the fact that the statistical properties of the covariance matric from a phylogeny or network could lead to downstream issues with PCMs. For empiricists having some sort of checking and then a warning is very useful to alerting folks about the potential issue, and to the fact that results using their tree might be suspect as a consequence. Observing this, this proposal is developed with the goal of finding analytical approaches for empiricists and writing a broader paper alerting the community about it.

*The model for trait evolution*

For species, let be the trait values observed on the tip of the tree. By the evolutionary dynamic under the Ornstein-Uhlenbeck process (Hansen 1997), the joint distribution for random vector of species is a multivariate normal distribution with common mean if not conditioning on the root of the tree (treat as the random), and variance-covariance matrix where is the time that separating species *i* and *j*, and is the time that species *i* and *j* shared a common ancestor. Note that given a tree with know branch lengths, and between any pair extant specie can be completely determined accordingly (Jhwueng 2013). The likelihood function given trait and tree with branch lengths is hence represented as

where is the determinant of and is a vector of 1s.

In particular, when , we have is the phylogenetic relationship matrix and the process for trait evolution is considered as under Brownian motion.

Notice that the best solution for the likelihood function in trait model depends on the covariance matrix itself. To calculate the inverse of the variance-covariance matrix , currently most widely applied method is to use a Moore-Penrose (MP) pseudoinverse (Roger and Charles 1985) that makes the algebra tractable and is a logical place to start. However, there are other methods that may turn out to be better for this situation. Conceptually, what we’re trying to do here is find the best, well-conditioned estimate of our observed covariance matrix, when our observed matrix is *ill conditioned*. In multivariate analysis, there is a huge literature on covariance matrix estimation and what to do with ill-conditioned matrices, MP pseudoinverse is a common approach, but so is finding the nearest valid covariance matrix to our observed matrix. In fact, the *nearPD* function in R implements Higham’s 2002 approach to finding the nearest valid covariance matric, and the inverse of this is not always the same as the pseudoinverse of the original matrix. Hence we think such adjustment under *nearPD* approaches might not be the best.

To solve *ill conditioned* issue, one approach, of course, is just to reject matrices, s, that are poorly conditioned, but the truth estimate can be fell in that region, and users for the packages would be less favorable to hear that their trees can’t be analyzed [though that’s far better than quietly returning a wrong result]. For more complex models it can be that the matrix is poorly conditioned for some parts of parameter space but not for others (trivially, more tip variance can make a matrix well-conditioned). It’s possible that something like species *jackknifing* could work: drop some random taxa, see if the matrix is well conditioned, estimate parameters, and repeats this until the search is done. Crude, but maybe less biased than other approaches.

We propose several possible approaches to remediate the *ill condition* matrix issue from the tree. Our current plan on this is to use some of the following methods to estimate the *best* version of the observed phylogenetic covariance matrix, and then use that for next step analysis. That means doing some comparisons of these methods to see how they perform. One approach would come from the tree-pruning side and the other from the matrix regularization side. We describe three methods

**(A) Shrinkage matrix regularization (B) Dropping shortest tip (C) length stretching** in the following

*(A) Shrinkage matrix regularization*

An approach by Shafer and Strimmer (2005, 2007) was developed for covariance matrices in molecular biology (including some network covariance matrices) and Theiler (2012) improved upon them accordingly for general purposes. Given the tree with the matrix , let and , define the shrinkage matrix estimator of by where. Let , the negative log likelihood function for the shrinkage parameter is

The best shrinkage estimates is to search the optima and the matrix is used for the next step analysis.

We first implemented the method to compare the condition number for the raw tree and shrunk trees. The 100 birth-death trees are simulated using *TreeSim* (Stadler 2009) where each tree is of of extant sampled tips, speciation rate , extinction rate , each tip is included into the final tree with probability 0.5, and the time since origin of the process is . The condition numbers of the variance-covariance matrix for the 100 raw trees and 100 shrunk trees are calculated. The results is plotted in Figure 3 which shows that the shrunk trees have lower condition numbers than the raw trees. This indicates an improvement for the ill-condition issue.

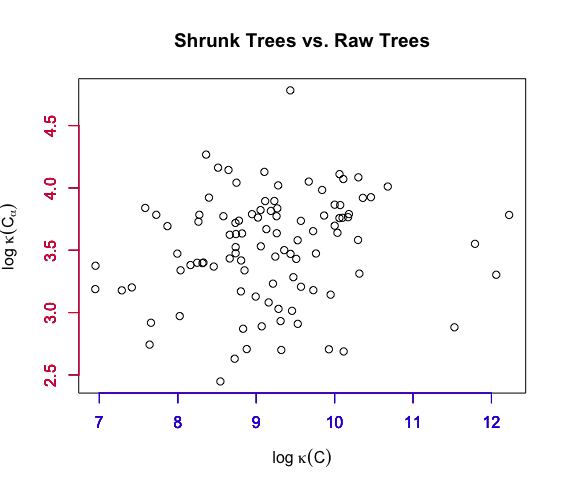


Figure 3: Scatter plot of the condition number of the raw trees

vs. the condition number for the shrunk trees .

We next visualize the phylogenies that result from the estimated covariance matrices, to see how they differ from the observed trees and where. Then we compare aspects of phylogenies that have well-behaved covariance matrices to those that are ill behaved.

Figure 4 compares the topology of the raw tree and the shrunk tree reconstructed using the shrinkage matrix under the unweighted pair group method of arithmetic mean (*upgma*) (Sakal and Michener 1958). Those shrunk tree seems to stretch a lot on the tip lengths.

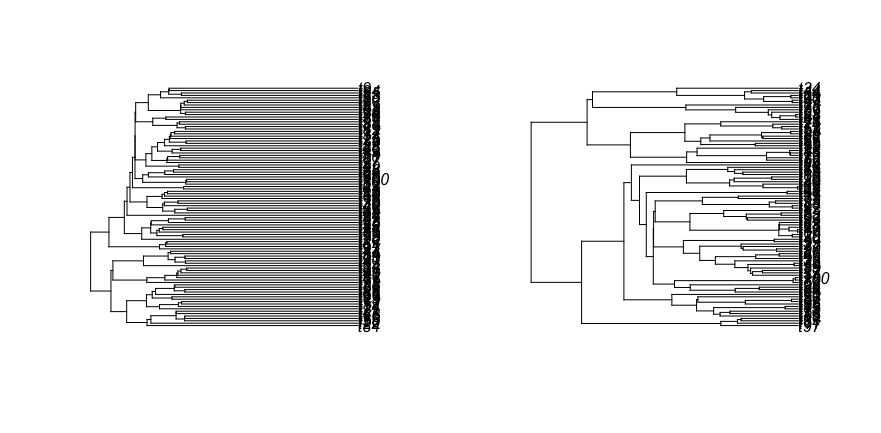


Figure 4: Tree comparison for the shrunk tree (left panel) and the raw tree (right panel)

Next we try to use three different sample sizes (20,100,150) and find that will increase as sample size increases. For one hundred replicates, the average of the shrinkage estimator are 0.02, 0.41, and 0.63 under the sample sizes of 20, 100 and 150, respectively. This shows that the magnitude of lengthen tip for the tree seems increases with the sample size.

*Eigenvalue Shrinkage*

Another common approach is to add a small constant to the eigenvalues of the covariance matrix and re-estimate the covariance matrix from the eigenvectors and adjusted eigenvalues. These are based on the observation that ill-conditioned matrices tend to have too much spread in their eigenvalues: the largest are too large and the smallest are too small. One therefore finds a way of shrinking the variation in these to obtain a more valid covariance matrix. Ledoit and Wolf (2003) outline the conceptual and theoretical issues and provides one implementation. We will implement this method in the future.

*A quick regression analysis for investigating the factor for ill-condition tree*

It might be the case for example that certain parts of the tree, or short inter-nodes, or whatever, are common features of phylogenies with ill-conditioned covariance matrices. That would be very useful to know for empiricists. We guess the shortest terminal branches could be an issue to affect the condition of the tree. But we could have asset of summary states such as (i) minimum branch lengths, (ii) maximum branch lengths, (iii) branch length ratio, (iv) branch length variance, (v) branch length median, (vi) tip lengths, (vii) internal lengths, (viii) minimum tip, (ix) maximum tip, (x) minimum internal length, (xi) maximum internal lengths that could be useful. To verify this, we did a quick analysis by simulating 20,000 trees under uniformly varying number of taxa between size of 10 and 100, birth rate between 0.01 and 0.1, death rate between 0 and , sampling fraction between 0.1 and 1, and tree of height , we calculate the condition

number and compare it to a variety of measures using multiple linear regression. We use R package: *MuMIn* (Bartoñ 2016) to dredge a bunch of models to correlate with various parameters or tree measures. We set the number of variables to 4 and essentially all the weight was on one of 1471 models. The model, where is the number of taxa and is the min tip, accounts for most weight of significance. From previous analysis, it makes sense that higher would contribute to worse (higher) , as does bigger smallest tip branch leading to better (lower) . Currently we are not sure about and , though.

From this simulation, there is something else that is very interesting. We then run all pairwise dot-plots between condition number and the other tree parameters. Basically nothing popped out as having an interesting pattern except one comparison condition number vs. min tip (i.e. vs. ). Figure 5 shows that either removal tends to help, but removing the taxon with the shortest branch helps more. We wonder if this points to a potential solution (or an approximate solution at least): dropping one of the tips with the shortest branch length as a quick run suggests this offers a much bigger improvement than dropping a tip at random.

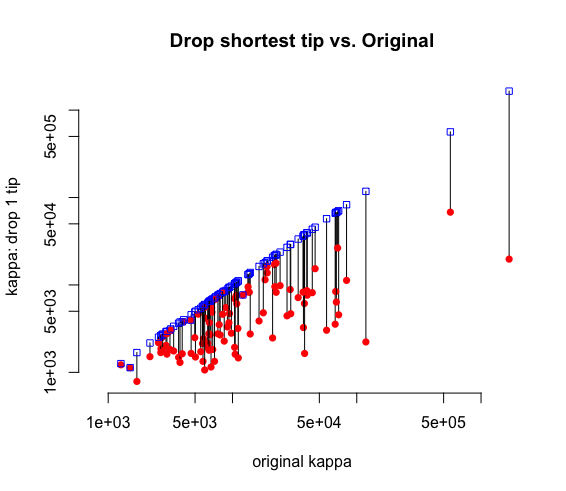


Figure 5: Condition numbers for dropping the shortest tip vs. the raw condition number. The x-axis is

the original kappa, and the lines connect points corresponding to that given tree, but with

one taxon removed: the blue square is removing a taxon at random, and the red dot

removing the taxon with the shortest branch tip length.

In fact, phylogenetic covariance matrices have a nested structure that general covariance matrices might not: it could be that some sort of tree pruning approach taking advantage of the structure allows calculation of the inverse. We find that the shortest tips are actually the smallest eigenvalues of the variance covariance matrix (they are equal). The proof of this property under tree of arbitrary taxa is provided as following. We start by illustrating a 3 taxa example.

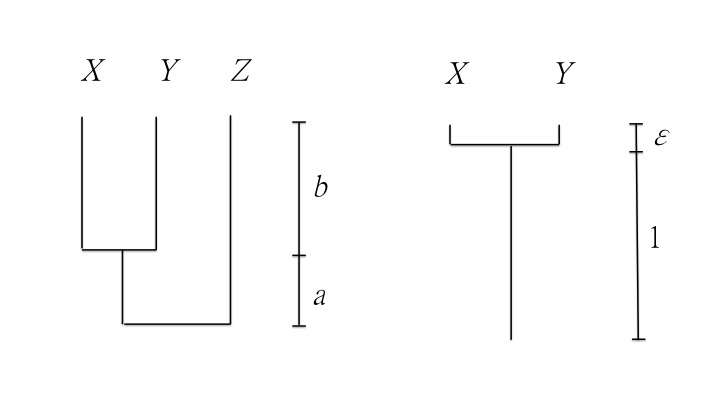


Figure 6: example of tree of three taxa and two taxa.

In Figure 6 (left panel), the tip lengths for species and are , and , respectively. So the shortest tip branch length is . The phylogenetic relationship length matrix for the tree can be represented as following

Claim: is the smallest eigenvalue.

Proof: Let be an eigenvalue for , then we have where is a 3 by 3 identity matrix. Simplifying this equation directly leads to which implies . Then we have 3 eigenvalues . Obviously the smallest eigenvalue is , which is the shortest tip length of the tree.

*General proof: smallest tip is the minimum of the eigenvalues.*

Note that given an ultrametric tree of tips, there exists a unique strictly ultrametric matrix (Nabben and Varga 1994) , which represents the affinity among the group of species.

Again let be the smallest tip length. Then has at least two identical columns as well as two identical rows where is a by identity matrix. This implies that , implies that is an eigenvalue of . The next step is to show that is the smallest eigenvalue for the eigenvalue set of .

*Claim*:

*Proof*: If is an eigenvalue of satisfying , it suffices to show that . Suppose to the contrary that . Consider the matrix , then is still a strictly ultrametric matrix, which is always invertible (see Nabben and Varga 1994, and Corollary 6.2.27 in Horn and Johnson 1985). Then we have and this implies . This consequence indicates that is not an eigenvalue of , which contradicts to the assumption of Therefore, is the smallest eigenvalue of .

The general proof shores up the pruning approach from a theoretical perspective, even if some may object to losing a taxon in their analysis and may lead to a biased estimate.

*Why tiny branch lengths cause ill-condition matrix*

The problem of having ill-conditioned matrix comes from the tiny branch lengths. We illustrate this issue using a simple example of two taxa shown in the Figure 6 (right panel)

The phylogenetic relationship matrix is

.

Note that has two eigenvalues and and the condition number of defined by the ratio of the largest eigenvalues to the smallest eigenvalues is where is the big *O* notation that describes the limit behavior of a function. When we have tiny tip branch (very small ), the value of will be fairly large and this indicates the matrix is more of ill conditioned. For instance, with , while with , we have . The problem becomes serious as is very closed to zero as we have a matrix of two almost identical columns/rows which makes a singular matrix with . In general, for a tree of arbitrary taxa that includes a clade of sub tree described in this case, we will encounter the ill-condition issue. To look into this issue, we can define a measure using the fraction to quantify the magnitude of ill conditioned of the tree where is the tree height, is the smallest tip length and is the branch lengths from the root to the most recent common ancestor of the tips with smallest tip lengths. When approaches to zero, the fraction approaches to zero. In this case, as two columns/rows in the matrix are almost the same, the matrix will suffer an ill condition of order . We use a simulated 100 taxa birth-death tree. For a good tree (), = 1.45, while for the bad tree () , . From here the tree condition can be seen of order .

*(B) Drop the shortest tip can lead a better condition of the*

We will show that the new tree obtained from dropping the shortest tip of the original tree has a better (lower) . Let be the by strictly ultrametric matrix and eigenvalues of . Let be the matrix obtained by dropping the shortest tip from the tree. is still an strictly ultrametic matrix of size by with eigenvalues . By a special case of the Cauchy’s interlacing theorem (Ch. 10.1 in Parlett 1980), we have . The condition number, defined as the ratio of the largest eigenvalue to the smallest eigenvalue are computed as and for and , respectively. From some algebra calculation, we have . This implies that . From above, we can conclude that a lower can be achieved by dropping the smallest tips.

*(C) Method: Length Stretching*

We can obtain a new tree by stretching the branch lengths of the raw tree without changing its topology. The matrix and its condition number for the new tree will change accordingly. For an unltrametric tree, let be the tree height from the root to the tip. Without loss of generality, we can first scale into a unit and

then decompose into components. That is, where represents the length between the *i*th and (*i+1*)th speciation events. For instance, is the length from the root to the first speciation event since the root and is the length for the tip species (with minimum tip length) evolved from its most recent common ancestor. Next, consider the matrix C obtained from the raw tree. Let the -tuple elements be the distinct entries in C satisfying . We can represent the relation between and by the equation .

We use Figure 6 (left panel) as an example. Let and , then the matrix is

=

Observing that , we have , and .

To stretch the lengths but retain the topology of the raw tree, since and we can treat as a -dimensional random vector from a Dirichlet distribution. can be generated by first drawing independent gamma random variables, each with different shape parameters and rate parameter 1 where is an arbitrary but positive constant, then the d-tuple vector is a Dirichlet random vector with , , and concentration parameters . Here the positive constant is an arbitrary scaling variable that always preserves the correct mean. By the property of Dirichlet distribution, we have

= ,

and the mode given by

where

The choice of is thus determined by We can choose a positive integer satisfies

where returns the integer closest but less than the real number

Because the mode of is not equal to the expectation of , such choice of does not guarantee that the distribution of is centered or symmetric around its mean . The mode converges to the expected mean when approach to infinity (i.e. as ). However, although choosing larger helps to center the distribution around , picking too large will cause the samples to be tightly centered around the given estimates . The choice of here is designed to be the minimal needed to prevent the phylogenetic tree from varying too wildly from the given one while still adequately testing robustness.

A less efficient algorithm (Gelman et al. 2003) relies on the univariate marginal and conditional distributions being beta and proceeds as follows. Simulate from a Beta distribution. Then simulate in order, as follows. For , simulate from a Beta distribution, and let . Finally, set .

To test this method, we first simulate a tree, and then calculate its condition number . We then stretch the length of the tree and calculate the corresponding condition number . Figure 7 shows a comparison of the two trees of 10 taxa.

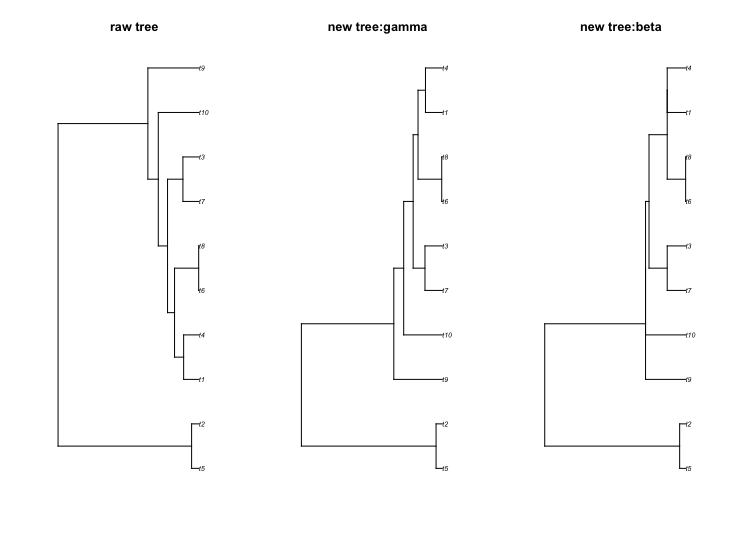


Figure 7: length stretching of tree. The left tree is the raw tree, the middle is the tree stretch under the gamma drawing approach and the right tree is the tree stretch under the beta drawing approach.

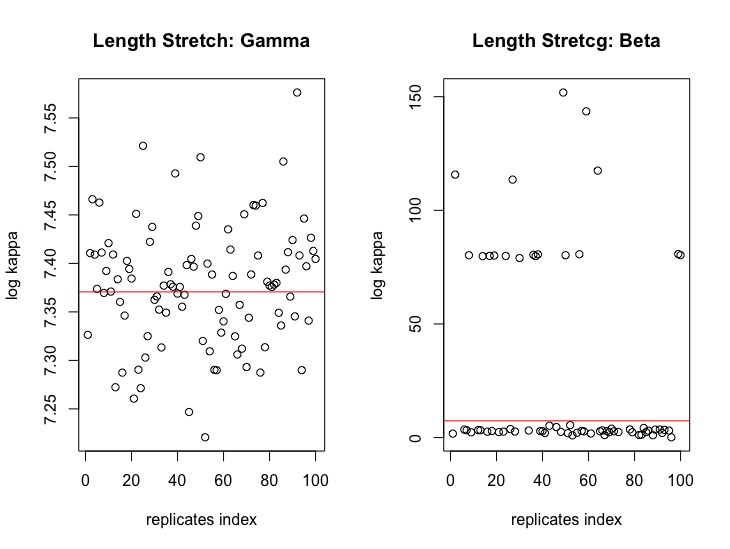


Figure 8: condition number for the stretch trees

We perturb the raw tree 100 times to get 100 new trees where their condition numbers are plotted in the Figure 8. The red line is the condition number of the raw tree. Figure 8 shows that the length stretching method could lead to a better or worse condition number. There is around 50 percent of acceptance for better trees with lower condition number; maybe we can choose a good tree with lower from them.

*Next Step analysis on estimation of the parameter using adjust matrix*

We simulate many trees under a given rate, estimate Brownian Motion rate under a variety of transforms (removing tips, shrinkage method, raw method) for the matrix. We will look at the root mean square error (RMSE) for BM rate under each of these approaches. Currently RMSE between raw method and drop the shortest tip are closed for 50 trees each of 100 taxa, but the shrunk method gives large RMSE in this case. We expect our methods can get convergent and good estimator with higher accuracy due to better condition.

*References*

Barton K (2016) MuMIn: multi-model inference. R package.

Boettiger C and Land DT (2012) Treebase: An R package for discovery, access and manipulation of

online phylogenies, Methods in Ecology and Evolution 3(6) 1060-1066.

Drummond AJ, Suchard MA, Xie D, and Rambaut A (2012) Bayesian phylogenetics with BEAUti and the

BEAST 1.7 Molecular Biology and Evolution 29:1969-1973.

Gelman A, Carlin JB, Stern HS, and Rubin DB (2003) Bayesian Data Analysis (2nd ed.)p582.

Hedges SB, Dudley J and Kumar S (2006) TimeTree: a public knowledge-base of divergence times

among organisms. Bioinformatics 22:2971-2972.

Higham NJ (1988) Computing a nearest symmetric positive semidefinite matrix. Linear algebra and its

application 103: 103-118.

Higham NJ (2002) Computing the nearest correlation matrix- a problem from finance. IMA Journal of

Numerical Analysis 22:329-343.

Hinchliff CE et al. (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life.

Proceedings of the National Academy of Sciences 112(41) 12764-12769.

Jhwueng DC (2013) Assessing the goodness of fit for phylogenetic comparative: a meta analysis and

simulation. PLOS ONE 8(6): e67001.

Johnson MTJ, FitzJohn RG, Smith SD, Rausher MD, and Otto SP (2011) Loss of sexual recombination and

segregation is associated with increased diversification in evening primroses. Evolution 65(11)

3031-3337.

Ledoit O and Wolf M (2004) Honey, I shrunk the sample covariance matrix. Journal of Portfolio

Management 30:110-119.

Liam JR (2012) phytools: an R package for phylogenetic comparative biology (and other things)

Methods Ecol. Evol. 3: 217-223.

Maddison DR and Schulz KS (2007) The Tree of Life Web Project. Internet address http://tolweb.org

Michonneau F, Brown JW and Winter DJ (2016) rotl: an R package to interact with the open tree of life

data. Methods in Ecology and Evolution. doi:10.1111/2-41-210X.12593

Nabben R and Varga RS (1994) A linear algebra proof that the inverse of a strictly ultrametric matrix is

a strictly diagonally dominant Stieltjes matrix. SIAM J. Matrix Anal. Appl. 15(1) 107-113.

O’Meara BC, Estman J, Heath T, Wright A, Klaus S, Chamberlain S, Midford P, Harmon L, Brown J,

Pennell M, Alfaro M. DateLife version 2.2.3 doi:10.5281/zenodo.56803

Paradis E, Claude J and Strimmer K (2004) APE: analysis of phylogenetics and evolution in R language.

Bioinformatics 20:289-290.

Parlett B (1980) The symmetric eigenvalue problem. Prentice-Hall, Englewood Cliffs, N.J.

Piel WH, Donoghue MJ and Sanderson MJ (2002) TreeBASE: a database of phylogenetic knowledge.

Pp.41-47. In: Shimura JK, Wilson L and Gordon D eds. To the interoperate “Catalog of Life” with partners Species 2000 Asia Oceanea. Research Report from the National Institute for Environmental Studies No. 171, Tsukuba, Japan.

Roger AH and Charles RJ (1985) Matrix Analysis, Cambridge University Press. The Press Syndicates if

the University of Cambridge ISBN 0-521-38632-2

Sakal R and Michener C (1958) A statistical method for evaluating systematic relationships. University

of Kansas Science Bulletin 38:1409-1438.

Sanderson MJ, Donoghue MJ, Piel W and Eriksson T. (1994). TreeBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. American Journal of Botany. 81(6) 183

Sanderson MJ (2002) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19(2) 301-302.

Schafer J and Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implication for functional genomics. Statistical Application in Genetics and Molecular Biology 4(1):32

Stadler T (2009) On incomplete sampling under birth-death models and connections to the sampling based coalescent. J. Theo. Biol. 261:58-66.

Stoltzfus A et al. (2013) Phylotastic! Making tree-of-life knowledge accessible, reusable and convenient. BMC Bioinformatics 14:158.