Version dated: April 19, 2018

RH: Characters correlation

Influence of different modes of morphological character correlation on phylogenetic tree inference

Thomas Guillerme[1,2,*], and Martin D. Brazeau[2,3]

[1] *School of Biological Sciences, University of Queensland, St. Lucia, Queensland, Australia.*

[2] *Imperial College London, Silwood Park Campus, Department of Life Sciences, Buckhurst Road, Ascot SL5 7PY, United Kingdom.*

[3] *Department of Earth Sciences, Natural History Museum, Cromwell Road, London, SW75BD, United Kingdom.*

***Corresponding author.** *guillert@tcd.ie*

Abstract

Phylogenetic analysis algorithms require the assumption of character independence - a condition generally acknowledged to be violated by morphological data. Correlation between characters can originate from intra-organismal features, shared phylogenetic history or forced by particular character-state coding schemes. Although the two first sources can be investigated by biologists *a posteriori* and the third one can be avoided *a priori* with good practices, phylogenetic software do not distinguish between any of them.

In this study, we propose a new metric of raw character difference as a proxy for character correlation. Using thorough simulations, we test the effect of increasing or decreasing character differences on tree topology. Overall, we found an expected positive effect of reducing character correlations on recovering the correct topology. However, this effect is less important for matrices with a small number of taxa (25 in our simulations) where reducing character correlation is not more effective than randomly drawing characters. Furthermore, in bigger matrices (350 characters), there is a strong effect of the inference method with Bayesian trees being consistently less affected by character correlation than maximum parsimony trees.

These results suggest that ignoring the problem of character correlation or independence can often impact topology in phylogenetic analysis. However, encouragingly, they also suggest that, unless correlation is actively maximised or minimised, probabilistic methods can easily accommodate for a random correlation between characters.

(Keywords: Character difference, correlation, topology, Bayesian, maximum parsimony)

Introduction

The last two decades have witnessed a "resurgence" of interest in the use of morphological character data in phylogenetic studies. This owes in large part to the use of fossils to undertake at least partial reconstructions of phylogenetic trees, especially where ancestral states reconstructions or absolute calibrations of divergence times are necessary. Morphological character data are often considered inferior to molecular sequence data, but are often the only source of phylogenetic data for extinct species. While there is a general appreciation of the limits of morphological data, they are frequently dismissed without any empirical investigations into their statistical properties. As morphological data are likely to continue to play an extensive role in phylogenetic analysis, it is essential to understand the circumstances under which morphological data might be expected to "misbehave". This opens up possibilities for predicting problematic datasets and possibly proposing new confidence measures in phylogenetic datasets.

The non-independence of large numbers of morphological characters is often cited in anticipation of problems with morphological data. The assumption of character independence is central to phylogenetic inference methods such as maximum likelihood and maximum parsimony (e.g. Joysey and Friday, 1982; Felsenstein, 1985; Lewis, 2001; Felsenstein, 2004). Especially for discrete morphological data, this assumption of independence is probably violated frequently due to the very nature of phylogenetic data: correlations are expected to occur (to some degree) when characters are depending on each other. Before discussing character correlation further, it is important to understand that it may manifest itself in at least three distinct ways:

- **Intra-organismal dependence:** this is the result of an intrinsic biological link between two characters through development, pleiotropy, and/or biological function. For example the lower and upper molar characters in mammals generally occlude one another. Therefore, one character describing a feature of a lower molar will be expected to be complemented by the surface of the occluding upper molar. Characters of the occlusal surface of two opposing molars will be expected to directly covary. Pleiotropy also results in covariation between different aspects of phenotype. From a phylogenetic perspective, it can be especially pernicious because the relationship between the traits in question may have no obvious link from a morphological or functional comparison alone. Intra-organismal links can be the targets of comparative developmental biology (Goswami and David Polly, 2010; Kelly and Sears, 2010; Stoessel et al., 2013; Goswami et al., 2014) or functional investigations.

- **Evolutionary dependence:** this is the result of sets of characters co-evolving due to selection, likely related to functional links between two traits that help serve an overall lifestyle trait. Unlike the case of intra-organismal dependence, there need not be an intrinsic constraint that causes these traits to covary. For example, in vertebrates, axial elongation can be correlated to limb reduction with snake-like bodies evolving multiple times in numerous tetrapod lineages. This is thought to correspond to adaptations for fossoriality or aquatic lifestyles. Such covariances are generally studied in the context of a given phylogeny, often one derived from molecular data with the morphological traits of interest mapped on it. Many methods have been developed to study these correlations, especially since they can provide us with a lot of information on how specific groups acquired specific characteristics (Russell Lande, 1983; Maddison, 1990; Pagel, 1994; Mark Pagel, 2006; Grabowski and Porto, 2016). However, these methods do not give us a means to objectively control correlations that might adversely affect phylogenetic inference.

- **Coding dependence:** this is the results of researcher methodology for defining or/and coding discrete morphological characters (Brazeau, 2011; Simões et al., 2017). Coding dependence manifests itself in several ways, particularly in coding redundant information. For instance, coding for the same absence in different characters creates state transformations associated with the loss or gain of a particular character. This occurs when a number of multistate characters include two variable feature states (e.g. large, small; red, blue etc.) in conjunction with absence. It is worth noting, however, that these correlations could also be due to the nature of the available data, especially in palaeontology. For example, when only one fragmentary molar is available to describe a specimen, researchers have to “extract” as much phylogenetic information from the available data as possible, potentially inducing correlations. This coding dependency is linked to hierarchical dependency between characters (Wilkinson, 1995; Brazeau et al., 2017). Finally this can also be due to a bias in the amount of characters available. For example, in skulls, because of their complexity, there is a high likelihood of inducing correlation (by effectively reducing structural complexity to discrete characters).

Of course, the three sources of dependence have an interaction: characters describing the left and right lower/upper molars will have induced dependence due to the modularity of the molars, their shared history and the duplicated coding. Logical dependence, however, is easily distinguished prior to phylogenetic inference, while the two other ones (intra-organismal and evolutionary) are much harder. However, the development of algorithms and software has not yet caught up with the need to deal with these interdependencies (De Laet, 2015; Brazeau et al., 2017). Intra-organismal dependence requires more detailed, often extremely time-consuming studies (and possibly beyond the limits of available technology). Even after all of the effort is expended, the results might then only be known for a single (model) species. Evolutionary dependence itself requires the resolution of a phylogenetic tree, and is best determined by independent character sets. This is frequently accomplished by mapping morphological traits on molecular phylogenetic trees.

These sources of dependence between characters are well studied in biology. Biological and evolutionary dependences are inherent parts to Evo-Devo and macroevolutionary studies and best practices to avoid coding-induced dependences are commonly known and applied. However, eventually, all these characters, whether they are independent or not are analysed through phylogenetic inferences software that are blind to these distinctions. If fact, what the software are confronted with is a two dimensional matrix problem that renders the morphological subtleties described opaque. This introduces a new, less studied, source of character dependence: **Correlation between characters detected by the software**: this is the result how software actually interprets the differences between characters. The vast majority of phylogenetic software ignores both the character's definition and the different states signification (simply treating them as different or similar tokens). Therefore a great number of characters and - traditionally - a few number of tokens can easily lead to dependence between characters. For example, if we consider the following matrix containing four cetartiodactyls - say a pig (e.g. *Sus*), a deer (*Cervus*), a hippo (*Hippopotamus*) and a whale (*Balaenoptera*) - and four binary characters - say (**C1**: presence (1) or absence (0) of an astragalus; **C2**: presence (0) or absence (1) of baleen; **C3**: presence (0) or absence (1) of a left astragalus with a double pulley; **C4**: presence (0) or absence (1) of a right astragalus with a double pulley:

In the example in Table 1, the characters **C1** and **C2** are the most likely to be truly independent; characters **C3** and **C4** suffer from a a coding induced dependency; characters **C1** and **C3**/**C4** have an evolutionary induce dependency and again, characters **C2** and **C3**/**C4** are likely to be independent. Yet a phylogenetic software will treat all these four characters in exactly the same way: only the sheer difference between the character states tokens will be used in order to infer the tree. Some

	C1	C2	C3	C4
Sus	1	1	1	1
Cervus	1	1	1	1
Hippopotamus	1	1	1	1
Balaenoptera	0	0	0	0

Table 1: Example of a matrix with software induced character correlation. **C1**: presence (1) or absence (0) of an astragal; **C2**: presence (0) or absence (1) of baleens; **C3**: presence (0) or absence (1) of a left astragalus with a double pulley; **C4**: presence (0) or absence (1) of a right astragal with a double pulley.

characters will therefore be expected to covary in non-phylogenetic way, and that this phenomenon can reasonably be expected to mislead phylogenetic analysis. Yet the question has never been explored through a thorough simulation framework (although it has been tackled empirically for morphological data Dávalos et al. 2014 or molecular data Zou and Zhang 2016).

How does these correlation really affect topology? We expect matrices with a high level of correlation to recover precise but inaccurate topologies but will matrices with low level of correlation (i.e. with high levels of homoplasy) actual cancel out the effects of correlation? Here we formally assess the effect of discrete character's correlation using simulated data. We propose a new distance metric to measure the difference between characters (as a proxy for these three sources of correlation as interpreted by the software) and a protocol to modify discrete morphological matrices to increase/decrease the overall differences or similarities between characters. We found that overall, there is a detectable effect of character correlation on topology where an increase in character dependence results in a decrease in the ability to recover the correct topology. These results, however, vary greatly in magnitude depending on the size of matrix and the inference method used.

Methods

To assess the effects of character correlation on the accuracy of phylogenetic inference we generated a series of matrices exhibiting different levels of correlation between some characters (Fig.1 - note that each step is described in more details below):

1. **Simulating matrices**: we simulated discrete morphological matrices with 25, 75 and 150 taxa for 100, 350 and 1000 characters, hereafter called the “normal” matrices. This step resulted in 9 matrices.

2. **Modifying matrices**: we changed the “normal” matrices by modifying the characters in order to maximise or minimise characters differences (hereafter called respectively “maximised” and “minimised” matrices) by removing respectively the least different or most different characters and replacing them randomly by the remaining characters. Our protocol for measuring character difference is detailed below.

We also randomly duplicated characters from the "normal" matrices without biasing towards maximising or minimising character differences to create randomised matrices (hereafter called the "randomised" matrices - equivalent to a null expectancy). This step resulted in 36 matrices.

3. **Inferring topologies**: we inferred the topologies from the "normal", "maximised", "minimised" and "randomised" matrices using both maximum parsimony and Bayesian inference. Hereafter, the resultant topologies are called the "normal", "maximised", "minimised" and "randomised" trees). This step resulted in 72 topologies.

4. **Comparing topologies**: finally, we compared the "normal" to the "maximised", "minimised" and "randomised" trees to measure the effect of character correlation on topology.

Each step was replicated 35 times and are described below in more detail, along with our proposed definition for measuring the difference between characters.

Measuring differences between characters

To measure the effect of character correlation as interpreted by the phylogenetic software, we define characters as being entirely correlated if they give the same phylogenetic information. In order to measure this, we propose a new distance metric to measure the difference between two characters:

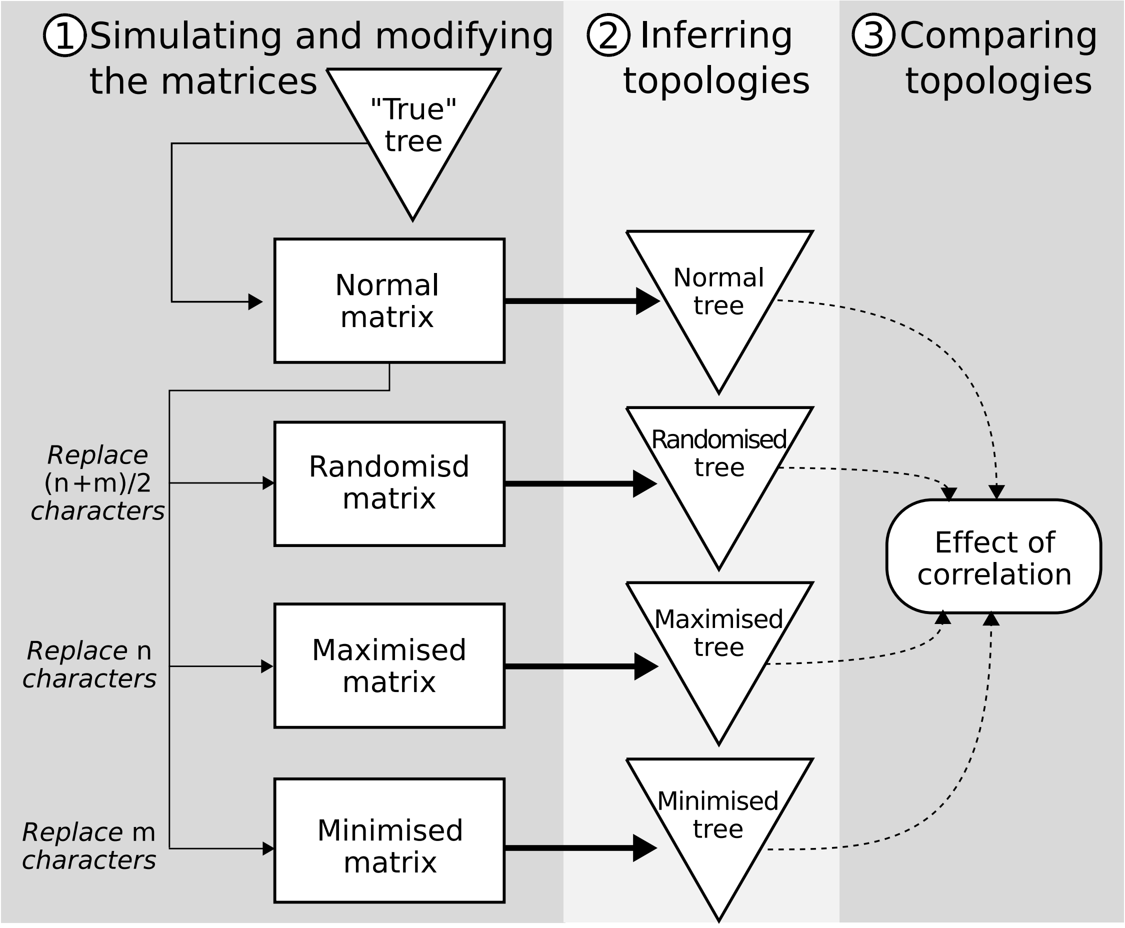

Figure 1: Outline of the simulation protocol: the first step includes both the simulation and the modification of the matrices (thin solid lines); the second step includes tree inference using MP and BPP methods (thick solid lines); the third step includes comparing the resulting tree topologies (dashed lines). n and m corresponds to the number of characters with a character difference < 0.25 and > 0.75 respectively.

Character Difference (CD).—

$$CD_{(x,y)} = 1 - 2\left(\left|\frac{\sum_i^n |x_i - y_i|}{n} - \frac{1}{2}\right|\right) \tag{1}$$

Where n is the number of taxa with comparable characters x, y and x_i, y_i are each character's state for the i^{th} taxon. CD is a continuous distance metric bounded between 0 and 1 (see the mathematical demonstration in the supplementary material 1). Since we are considering differences as being only Fitch-like (non-additive) and unweighted, we calculated the difference between character states in a qualitative way. Two same character states tokens have a difference of zero and two different ones have a difference of one (e.g. $0 - 0 = 0$ or $1 - 8 = 1$). Additionally, we only consider differences for taxa with shared information (i.e. a Gower distance; Gower, 1971).

We standardised each character by arbitrarily modifying their character state tokens (or symbols) by order of appearance. In other words, we replaced all the occurrences of the first token to be 1, the second to be 2, etc. This procedure was used to treat all the characters are unordered with no assumption on the meaning of the character state (e.g. in a binary character 0 is not necessary ancestral to 1). It also greatly improved the speed of our algorithm implementation to compare the characters. This way, a character `A = {2,2,3,0,0,3}` for six taxa would be standardised as `A' = {1,1,2,3,3,2}` (following the *xyz* notation in Felsenstein, 2004, p.13). Note that in terms of phylogenetic signal, both `A` and `A'` are exactly identical (forming three distinct splits in the tree inference process).

When the character difference is null (0) it means that characters convey the same phylogenetic signal (i.e. characters are entirely correlated). When the character difference is maximal (1) it means it conveys the greatest difference in phylogenetic signal (i.e. characters are uncorrelated). It is important to stress that a character difference of 0 (i.e. the same phylogenetic signal) does not mean the opposite of 1 (i.e. *not* the opposite phylogenetic signal but the most different number of implied splits) . For example with three characters `A = {0,1,1,1}`, `B = {1,0,0,0}` and `C = {0,1,2,3}`, $CD_{(A,B)} = 0$ and $CD_{(A,C)} = 1$. Because the character is continuous and bounded between $(0, 1)$, it can be interpreted as the probability of two characters leading to a different set of splits (i.e. a different phylogenetic signal).

Simulating discrete morphological matrices

To simulate the matrices we applied a protocol very similar to Guillerme and Cooper (2016b). First, we generate random birth-death trees with the birth (λ) and death (μ) parameters sampled from a uniform $(0, 1)$ distribution maintaining $\lambda > \mu$ using the `diversitree` `R` package (v0.9-8; FitzJohn, 2012) and saving the tree after reaching either 25, 75 or 150 taxa. For each tree, we arbitrarily set the outgroup to be the first taxon (alphabetically) thus effectively rooting the trees on this taxon. These trees are hereafter called the “true” trees (see distinction below). We then simulated discrete morphological characters on the topology of these trees using the either of the two following models:

- The morphological HKY-binary model (O’Reilly et al., 2016) which is an HKY model (Hasegawa et al., 1985) with a random states frequency (sampled from a

Dirichlet distribution $Dir(1,1,1,1)$) and using a transition/transvertion rate of 2 (Douady et al., 2003) but where the purines (A,G) were changed into state 0 and the pyrimidines (C,T) in state 1. This model has the advantage of not favouring Bayesian inference (since it doesn't use an M*k* model; O'Reilly et al., 2016, ; see discussion) but the downside of it is it can only generate binary state characters (or 4 states; Puttick et al., 2017).

- To generate more than binary states characters, we used the M*k* model (Lewis, 2001). We draw the number of character states with a probability of 0.85 for binary characters and 0.15 for three state characters (Guillerme and Cooper, 2016b; Zou and Zhang, 2016). This model assumes a equal transition rate between character states which might seem overly simplistic, excluding other observed transition patterns (e.g. Dollo characters; Dollo, 1893; Wright et al., 2015). Recently however, Wright et al. (2016) have shown that an equal rate transition is still the most present in empirical data.

For each character, both models (morphological HKY-binary or M*k*) where chosen randomly and run with an overall evolutionary rate drawn from a gamma distribution ($\beta = 100$ and $\alpha = 5$). These low evolutionary rate values allowed reduction in the number of homoplasic character changes, thus reinforcing the phylogenetic information in the matrices. We re-simulated every invariant characters to obtain a matrix with no invariant characters in order to better approximate real morphological data matrices. To ensure that our simulations were reflecting realistic observed parameters, we only selected matrices with Consistency Indices (CI) superior to 0.26 (O'Reilly et al., 2016).

For each tree with 25, 75 or 150 taxa we generated matrices with 100, 350 and 1000 characters following O'Reilly et al. (2016). The matrices were generated using the `dispRity R` package (Guillerme, 2016). To estimate the variance of our simulations and assess the effect of our random parameters, we repeated this step 35 times resulting in 315 "normal" morphological matrices.

Modifying the matrices

We calculated the pairwise character differences for each generated matrix using the `dispRity R` package (Guillerme, 2016). We then modified the matrices to either maximise or minimise the pairwise character differences for each matrix using three different algorithms. For maximising the pairwise differences between characters, we selected the characters that were the most similar to all the others (i.e. with an average character difference < 0.25) and replaced them randomly by any of the remaining characters. This operation increased the overall pairwise character difference in the matrix thus making the characters more dissimilar. Conversely, for minimising the pairwise character differences, we selected the most dissimilar characters (i.e. with an average character difference < 0.75) and randomly replaced them with the remaining ones. Finally, because this operation effectively changes the weight of characters (i.e. giving the characters < 0.25 or > 0.75 a weight of 0 and giving the randomly selected remaining characters a weight of $+1$), we randomly replaced the average number of characters replaced in the character maximisation and minimisation by any other characters as a randomised expectation modification (i.e. randomly weighting characters). Each of the three matrices are effectively a bootstrap pseudo-replication of the "normal" matrix with the "randomised" one being a random one and the "maximised" and "minimised" being conditional bootstraps. This step resulted in a total of 1260 matrices (hereafter called the "normal", "maximised", "minimised" and "randomised" matrices - see Fig. 2 for an illustration). The algorithms for the three modifications are available on GitHub (`https://github.com/TGuillerme/CharactersCorrelation`)

Inferring topologies

We inferred the topologies with both BPP and MP using MrBayes (v3.2.6; Ronquist et al., 2012) and PAUP* (v4.0a151; Swofford, 2001) respectively. For both methods, we used the arbitrarily chosen outgroup in the simulations to root our trees. The maximum parsimony inference was run using a heuristic search with random sequence addition replicate 100 times with a limit of 5×10^6 rearrangements per replicates (`hsearch addseq=random nreps=100 rearrlimit=5000000 limitperrep=yes`).

Bayesian inference was run using an M*k* model with ascertainment bias and four discrete gamma rate categories (M*kv* 4Γ - `lset nst=1 rates=gamma Ngammacat=4`) with an variable rate prior an exponential (0.5) shape (`prset ratepr=variable Shapepr=Exponential(0.5)`). We ran two runs of 6 chains each (2 hot, 4 cold) for a maximum of 1×10^9 generations with a sampling every 200 generations. We automatically stopped the MCMC when the average standard deviation of split

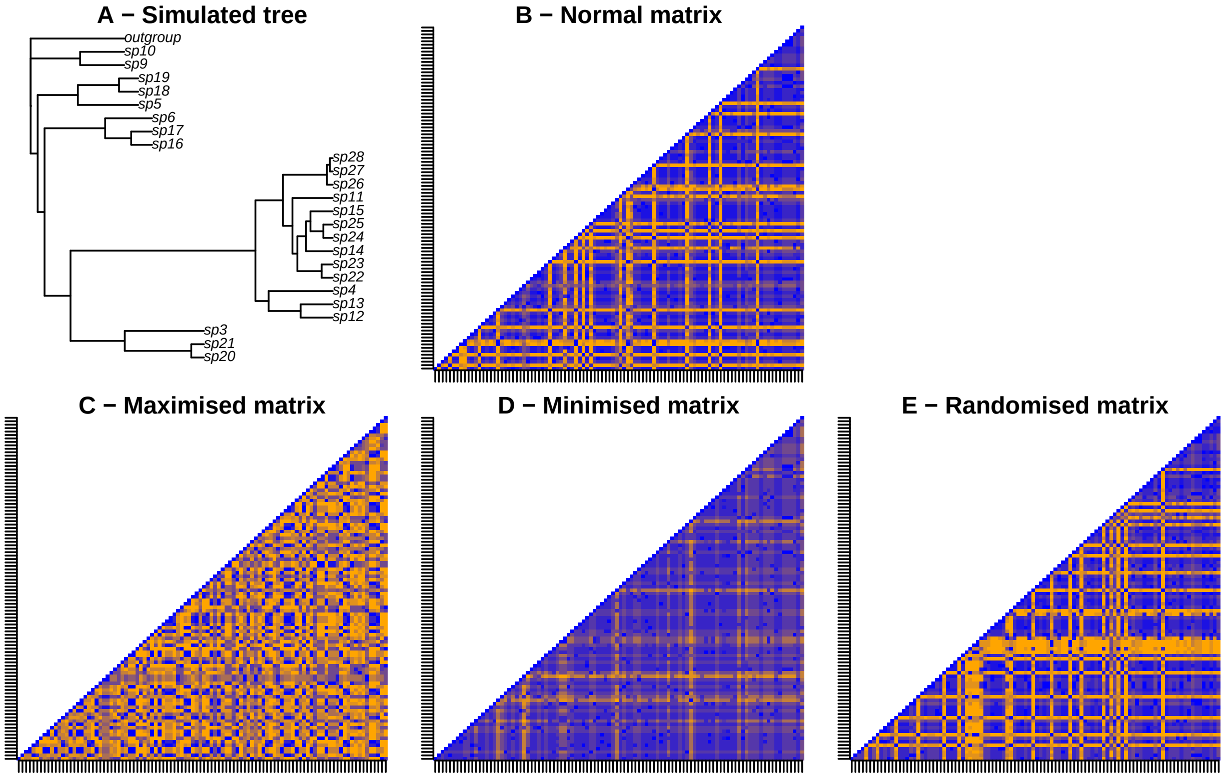

Figure 2: Example illustration of the protocol for modifying matrices. The matrices represent the pairwise character differences for 100 characters. Blue colours correspond to low character differences and orange colours correspond to high character differences. **A** - a random Birth-Death tree is simulated and used for generating the "normal" matrix (**B**), characters in this matrix are then removed or duplicated to favour maximised (**C**), minimised (**D**) or randomise character difference (**E**). The differences between the characters is low in **C** (minimised compared to **A**) implying a high correlation between the characters. Conversely, the character differences is high in **D** (maximised compared to **A**) implying a low correlation between the characters.

frequencies (ASDSF) between both runs fell below 0.01 (with a diagnosis every 1×10^4 generations - `mcmc nruns=2 Nchains=6 ngen=1000000000 samplefreq=200 printfreq=2000 diagnfreq=10000 Stoprule=YES stopval=0.01 mcmcdiagn=YES`). Due to cluster hardware requirements an to save some time, when chains didn't converged and the runs exceeded 5GB each, we aborted the MCMC and computed the consensus tree from the unconverged chains. In practice, these few MCMC got stuck at an ASDSF around (but not below) 0.01.

A strict majority rule tree was then calculated for both Bayesian an maximum parsimony trees. For the Bayesian consensus trees, the 25% first trees of the posterior tree distribution were excluded as a burnin. The 2880 tree inferences took around one CPU century on the Imperial College High Performance Computing Service (2-3GHz clock rate; ICHPC, 2011).

Comparing topologies

We compared the topologies using the same approach as in Guillerme and Cooper (2016b): we measured both the Robinson-Foulds distance (Robinson and Foulds, 1981) and the triplets distance (Dobson, 1975) between the trees inferred from the "maximised", "minimised" and "randomised" matrices and the tree inferred from the "normal" matrix. We explored the effect of character difference on recovering the "normal" topology by comparing the "maximised", "minimised" and "randomised" trees to the "normal" tree (Figs 3 and 4 and supplementary materials 3 Figs 1 and 2). Note that we are not comparing the trees to the "true" tree used to simulate the matrices. First, in biology, this tree is always unknown. Second, our objective is to measure the direct effect of character correlation approximated by the difference in topology between the "normal", "maximised" and "minimised" trees. When measuring the difference between these trees and the "true" tree, we would also confound the effect of simulating a birth-death tree and simulating a discrete morphological matrices from it.

The metric scores where calculated using the `TreeCmp` javascript (Bogdanowicz et al., 2012). The measurements where then standardised using the Normalised Tree Similarity metric (NTS; i.e. centering the metrics scores using the mean metric score for 1000 pairwise comparisons between random trees with n taxa; Bogdanowicz et al., 2012; Guillerme and Cooper, 2016b). When the normalised metric has a score of one it means both trees are identical, when it has a score of zero it means the trees are no more different than expected by chance and when it has a score < 0 the trees are more different than expected by chance. The normalised score for both metrics thus reflects two distinct aspects of tree topology: (1) the Normalised Robinson-Foulds (NTS_{RF}) Similarity reflects the conservation of clades (i.e. a score close to 1 indicates that most clades are identical in both trees); and (2) the Normalised Triplets Similarity (NTS_{Tr}) reflects the position of taxa (i.e. a score close to 1 indicates that most taxa have the same neighbours in both trees).

Because both NTS_{RF} and NTS_{Tr} metrics are bounded at one. The residuals of any model based on the NTS scores were not normal thus preventing the use of parametric tests for comparisons (see online material `https://rawgit.com/TGuillerme/CharactersCorrelation/master/Analysis/02-EffectCorrelationFullResults.html`). Similarly, a non-parametric Wilcoxon rank test (Hollander et al., 2013) would be biased in its p-value calculation due to the presence of equal values in the NTS distributions (e.g. when multiple trees are equal to the "normal" tree). Therefore, we used a combination of the Wilcoxon rank test with a Bonferonni-Holm corrections (to ensure our significant results were robust to Type I error rate inflation; Holm, 1979) and a simple non-parametric metric for measuring the probability of overlap between two distributions, the Bhattacharyya Coefficient (BC; Bhattacharyya, 1943; Guillerme and Cooper, 2016b; Soto et al., 2016). Thus, additionally to the Wilcoxon test results, we considered distribution to be significantly similar if they had an overlap probability > 0.95 and different if they had an overlap probability > 0.05. Comparisons falling between these range can not be designated as strictly similar/different but can still be ranked (e.g. for three distributions A, B, C, if $BC_{(A,B)} = 0.15$ and $BC_{(A,C)} = 0.65$, we cannot consider either distribution significantly different or similar but B still has a lower probability of being similar to A than C).

The resulting full simulation was 3.5TB big so is not shared here (though the parameters are). However, the resulting consensus trees on which the topological differences are calculated are available at `https://figshare.com/s/7a8fde8eaa39a3d3cf56`.

Results

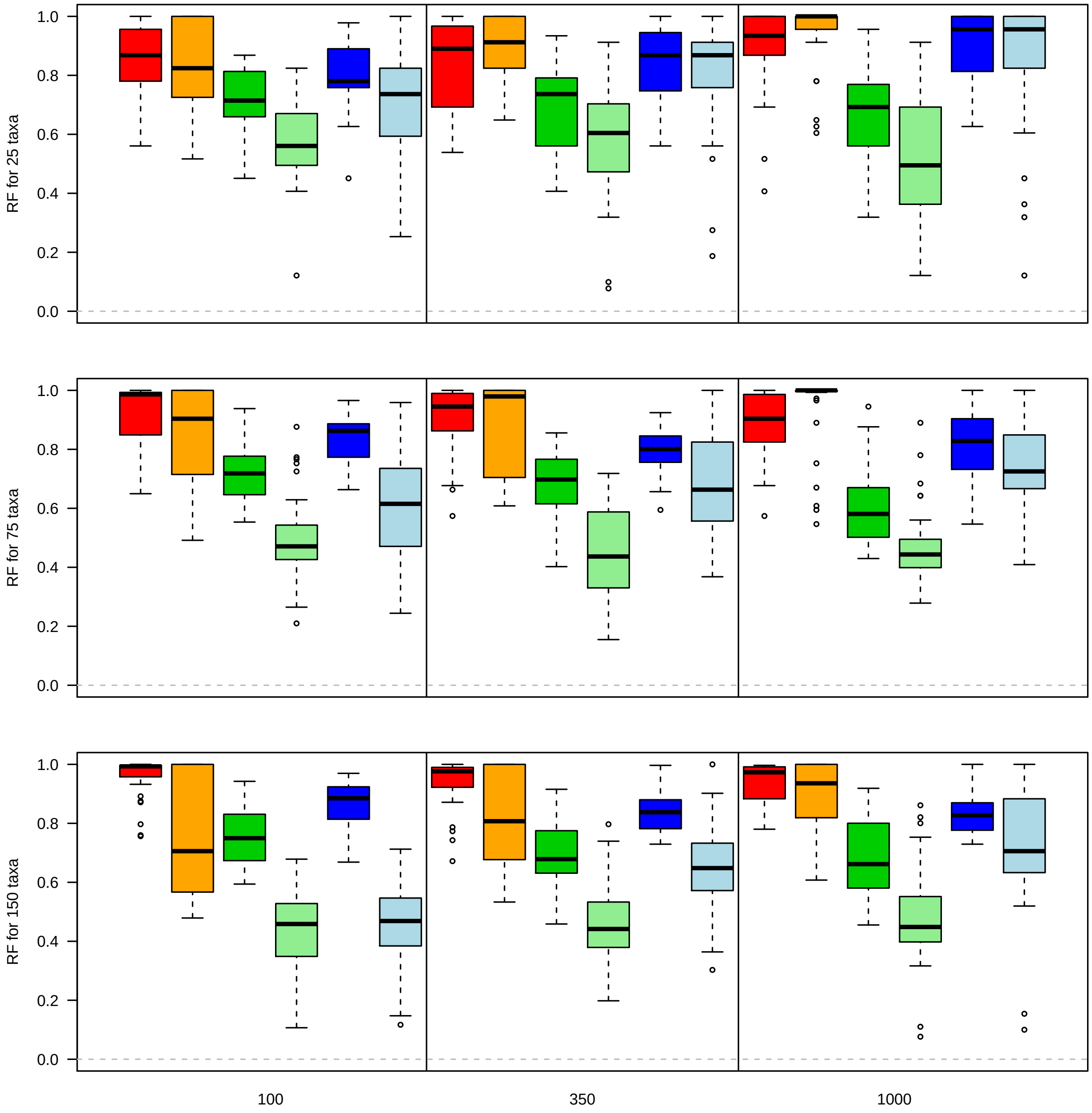

Figure 3: Effect of character difference on recovering the "normal" topology. The y axis represents the Normalised Tree Similarity using Robinson-Fould distance for matrices with 25, 75 and 150 taxa from top to bottom respectively. The x axis represents the different character difference scenarios and tree inference method with the "maximised" character difference in Bayesian (red) and under maximum parsimony (orange), the "minimised" character difference in Bayesian (dark green) and under maximum parsimony (light green) and the "randomised" character difference in Bayesian (dark blue) and under maximum parsimony (light blue) for matrices of 100, 350 and 1000 characters in the panels from left to right.

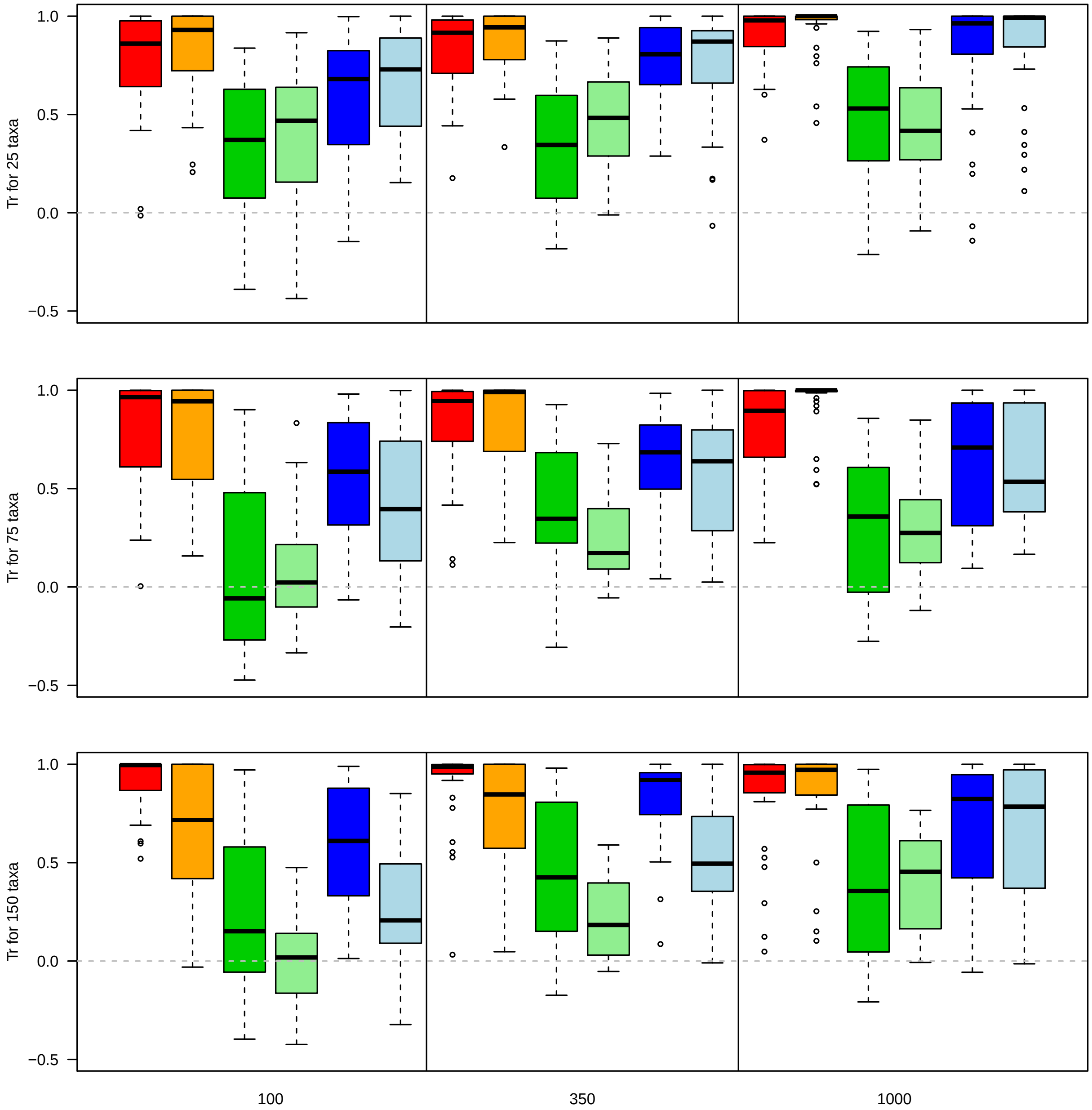

Figure 4: Effect of character difference on recovering the "normal" topology. The axis are identical to figure 3 but y axis represents the Normalised Tree Similarity using Triplets distance.

Effect of character differences on topology

The overall amount of character difference in a matrix has an effect of the ability to recover the correct topology when maximising character difference leading to the smallest loss in phylogenetic information (median NTS_{RF} = 0.956 and median NTS_{Tr} = 0.839) followed by simply randomising the characters (median NTS_{RF} = 0.762 and median NTS_{Tr} = 0.628) and minimising the character difference (median NTS_{RF} = 0.605 and median NTS_{Tr} = 0.303 - see supplementary material 3 for the full summary statistics). There is a significant difference between all scenarios (maximising, minimising and randomising) with the highest probability of overlap being between maximising and randomising the character difference (Bhattacharrya Coefficient of 0.873 for the NTS_{RF} and 0.908 for the NTS_{Tr} - Table 2) and the lowest probability between maximising and minimising the character difference (Bhattacharrya coefficient of 0.573 for the NTS_{RF} and 0.614 for the NTS_{Tr} - Table 2)

Number of characters.— This effect of the character difference is not dependent on the number of characters when looking at clade conservation (i.e. NTS_{RF}). The median NTS_{RF} was similar for 100, 350 and 1000 characters (0.730, 0.745, 0.767 respectively - see supplementary materials 3) with a significant difference only between 100 and 1000 and 350 and 1000 characters (Table 3). The number of characters affects the character difference more in terms of taxon placement for a low number of characters (median NTS_{Tr} for 100, 350 and 1000 characters equals 0.544, 0.693, 0.799 respectively - see supplementary materials 3) with a significant difference between 100 and 350 or 1000

metric	test	bhatt.coeff	statistic	p.value
RF	maxi:mini	0.573	356436.000	**0**
	maxi:rand	0.873	287225.000	**0**
	mini:rand	0.856	95841.500	**0**
Tr	maxi:mini	0.614	358800.000	**0**
	maxi:rand	0.908	288223.500	**0**
	mini:rand	0.858	98507.500	**0**

Table 2: Difference between the pooled scenarios. Bhatt.coeff is the Bhattacharrya Coefficient (probability of overlap), the statistic and the p.value are from a non-parametric wilcoxon test (with Bonferonni-Holm correction)

characters (Table 3). However, these differences have to be contrasted by a very high probability of overlap between each number of characters and metrics (Bhattacharrya Coefficient always > 0.95) suggesting that the significant effects of the number of characters still leads to really similar distributions.

Number of taxa.—

Similar to the effect of number of characters on character difference, the number of taxa seems to have only a marginal effect. A low number of taxa (25) resulted in significant differences with both 75 or 150 taxa in both NTS_{RF} and NTS_{Tr} but no differences between 75 and 150 taxa (medians for 25, 75 and 150 taxa equals 0.802, 0.76, 0.763 NTS_{RF} and 0.758, 0.588 and 0.615 NTS_{Tr} respectively - Table 4 and see

metric	test	bhatt.coeff	statistic	p.value
RF	c100:c350	**0.99**	190357.500	1
	c100:c1000	**0.98**	174085.500	**0.001**
	c350:c1000	**0.984**	180460.000	**0.032**
Tr	c100:c350	**0.961**	166609.500	**0**
	c100:c1000	**0.956**	151389.500	**0**
	c350:c1000	**0.981**	178793.500	**0.014**

Table 3: Difference between the pooled number of characters. Bhatt.coeff is the Bhattacharrya Coefficient (probability of overlap), the statistic and the p.value are from a non-parametric wilcoxon test (with Bonferonni-Holm correciton)

supplementary materials 3). Again, however, the significant differences have to be contrasted with still high probabilities of overlaps for each NTS_{RF} and NTS_{Tr} distributions for every number of taxa (Table 4).

Effect of character differences on the inference method

Regarding the inference method, there is a significant difference in clade conservation between Bayesian and maximum parsimony (Table 5 - median NTS_{RF} of 0.828 and 0.679 respectively) but not in terms of individual taxon placements (Table 5 - median NTS_{Tr} of 0.738 and 0.601 respectively).

Combined effects of taxa, characters and correlation on topology

metric	test	bhatt.coeff	statistic	p.value
RF	t25:t75	**0.976**	218421.000	**0.012**
	t25:t150	**0.988**	220529.000	**0.004**
	t75:t150	**0.99**	201037.000	1
Tr	t25:t75	**0.976**	233282.000	**0**
	t25:t150	**0.978**	227288.000	**0**
	t75:t150	**0.992**	194201.000	1

Table 4: Difference between the pooled number of taxa. Bhatt.coeff is the Bhattacharrya Coefficient (probability of overlap), the statistic and the p.value are from a non-parametric wilcoxon test (with Bonferonni-Holm correciton)

When looking at the combined effect of each parameter, the "maximised" and "minimised" scenarios are always significantly different with no high probability of overlap for both NTS_{RF} and NTS_{Tr} (Wilcoxon rank test p.value < 0.05 and Bhattacharrya Coefficient < 0.95 - see supplementary material 3). The same differences are observed when comparing the "maximised" scenario against the "randomised" one expect for: (1) the Bayesian inference with 25 taxa (with 100, 350 and 1000 characters) and with 75 taxa with 1000 characters for both NTS_{RF} and NTS_{Tr}; and (2) the maximum parsimony for 25 taxa (with 350 and 1000) characters for both NTS_{RF} and NTS_{Tr} and 75 taxa with 100 characters for NTS_{Tr}. Identically, there was always a significant difference between the "minimised" scenario and the "randomised" one was

metric	test	bhatt.coeff	statistic	p.value
RF	bayesian:parsimony	0.891	579437.500	**0**
Tr	bayesian:parsimony	**0.984**	470621.500	0.084

Table 5: Difference between the pooled methods. Bhatt.coeff is the Bhattacharrya Coefficient (probability of overlap), the statistic and the p.value are from a non-parametric wilcoxon test (with Bonferonni-Holm correciton)

expect for the matrix of 150 taxa and 100 characters under maximum parsimony for NTS_{RF} and the matrix of 150 and 1000 characters under Bayesian inference for NTS_{Tr}. The full list of comparisons and summary statistics are available in the supplementary materials 3.

Discussion

Effect of character differences on topology

As expected, there is a significant effect of the character difference in the ability to recover the correct topology. The character difference metric can be seen as the inverse of character correlation (see Methods): a high character difference approximates a low level of character correlation and vice versa. When characters are correlated, one could expect the matrices to convey a strong (but potentially misleading) phylogenetic signal since every character agrees with each other and conversely, when characters are uncorrelated, one could expect them to convey a weaker phylogenetic signal with a high amount of homoplasy. Intuitively, this would lead the "minimised" character difference scenario to lead to incorrect but consistent trees, the "maximised" scenario to lead to poorly resolved once (really homoplasic trees) and the "randomised" scenario to perform the best at recovering the correct topology. Although the expected results appear to be true for a low character difference scenario, increasing the character difference surprisingly improves the ability to recover the "normal" topology both in terms of clade conservation (NTS_{RF}) and taxa placement (NTS_{Tr}) for both inference methods (especially in bigger matrices; Figs 3 and 4). Furthermore, the trees generated by the "minimised" scenario do not appear better resolved (towards any topology) than the other scenarios (see Supplementary material 3, Figs 3, 4 and 5).

Number of characters and taxa.— Because of the nature of our simulation protocol, one could expect that the effect of character correlation would have increased with the number of characters (i.e. the more characters available, the more characters are modified in each scenario). Similarly, one could expect the number of taxa to have an effect of the raw ability to recover the "normal" topology (i.e. the more taxa, the more likely taxa are misplaced by chance).

Although we measured a significant difference between "small" and larger matrices (both in terms of number of taxa and characters; Tables 3 and 4), these differences have to be contrasted with the probability of overlap between the results distributions that are always really high between every matrices sizes (Bhattacharrya

Coefficients > 0.965 for both the different number of characters and taxa). This suggest that the effect of character correlation on recovering the right topology is independent of the size of the matrix when pooling the data. For the number of characters, this suggests that the overall character difference metric is a good proxy for character correlation as it is independent of the number of characters analysed. Similarly, using the a Normalised Tree Similarity metric (NTS) accounts for the fact that topological difference is affected by the sheer number of taxa considered (i.e. we corrected for the expected difference when comparing two random trees with the same number of taxa).

Effect of character differences on the inference method

When considering the pooled effect of the tree inference method, we only detected a significant difference between the Bayesian and the maximum parsimony trees in terms of clade conservation but none in terms of taxa placement (both using a Wilcoxon test and the Bhattacharrya Coefficient; Table 5). The difference in the ability of each method to recover the "correct" topology has been heavily discussed in the last five years with some indications that Bayesian inference will outperform parsimony when analysing discrete morphological characters alone (Wright and Hillis 2014; O'Reilly et al. 2016; Puttick et al. 2017; although some critics have raised issues with these investigations Spencer and Wilberg 2013; Goloboff et al. 2017). In this study, it is possible that our simulation protocol for generating the characters (favouring slightly more M*k*-based characters rather than HKY ones) could slightly favour Bayesian inference over maximum parsimony, however, our protocol for selecting matrices (i.e. those with in a

$CI < 0.26$ in a quick preliminary parsimony search; O'Reilly et al., 2016) could also favour maximum parsimony analysis. It was however not the purpose of this study to compare the overall performance of both methods but rather to measure the effect of character correlation on each of those methods separately.

The differences in performance of the two methods observed here could be due to the inherent mechanisms of each method. For any given topology T that was obtained from the "normal" matrix and a matrix with high homoplasy, both methods will generate score differently: (1) in parsimony, the topology will probably be given a bad optimality score (on that implies many changes along the tree) and the optimality criterion (favouring the minimum score) will likely discard the tree. The tree search will thus likely result in a topology island that will not contain the given topology T. (2) in Bayesian inference, the topology will also be given a bad optimality score (i.e. low likelihood) although the high homoplasy can be accommodated in the tree through high evolution rates or/and long branches. The rate and the branch length being two parameters among others, the optimality score (the likelihood) will change less drastically than for using parsimony. Furthermore, in Bayesian inference, a reasonable difference in optimality between two topologies (the acceptance) will not necessarily mean that the given topology T will be discarded. This difference in both mechanisms could explain why, on average, Bayesian Inference seems better to recover the "normal" topology than maximum parsimony.b

Distinction between different character correlations

Here we mention three different types of character correlations but evolutionary biologists are mainly interested in the intra-organismal and evolutionary correlations (e.g. in evo-devo Goswami and Janis 2006; or in macroevolution FitzJohn et al. 2014). These two types of correlations can only be studied *a posteriori* with a phylogenetic hypothesis and should not used *a priori* as a criterion to select characters. In other words, intra-organisaml and evolutionary correlation should be studied based on an underlying phylogenetic framework making the correlation induced by data collection (i.e. coding correlation) the only type of correlation that can affect the phylogeny *a priori*. This dichotomy thus creates a trade of between: (1) coding fewer characters (stochastically reducing *a priori* correlation) but making the *a posteriori* correlation more dependent on the coding; and (2) coding more characters (increasing *a priori* dependence) but allowing the *a posteriori* correlation being less dependent on the coding correlations.

It is important to note that the two other sources of character correlation could also be present in our simulations although they were not explicitly modelled: (1) evolutionary correlation is implied by simulating the characters based on Birth-Death trees; and (2) intra-organismal correlation could also be present in the matrices for those characters randomly simulated but sharing similar evolutionary simulation regimes (i.e. creating “modules” of characters). However, the effect of these sources of correlation was out of the scope of this study and would have required *a posteriori* changes to the matrices which are - when using empirical data - at best bad practice and at worth dishonest.

Limitations

First, simulating evolutionary history is complex. Not only because the models we're using to infer phylogenies are ever improving (e.g. Heath et al., 2014; Wright et al., 2016) but also because generalising morphological evolution across vastly different organisms is probably impossible (see constrasted discussions from Goloboff et al., 2018; O'Reilly et al., 2018). However, we do not compare the "maximised", "minimised" and "randomised" to the "true" tree but rather to the "normal" tree. This allows us to reduce the caveats from our simulations on the effect of character correlation since we only compare the simulation end products to each other (the outputs) rather than to the simulation inputs.

Second, measuring and modifying character correlation is difficult. In our simulation protocol we chose to create simulation by duplicating characters in a matrix to maximise or minimise correlation. In biology, this correlation arises from either intra-organismal or evolutionary mechanisms. This could lead to correlations between characters to be more present in some parts of the trees that other (e.g. in the case of inapplicable data Brazeau et al., 2017). However, because of the number of characters, it is actually complex to actually measure their correlation in a biological sense and is still actively discussed in the literature (Russell Lande, 1983; Maddison, 1990; Pagel, 1994; Mark Pagel, 2006; Goswami and Janis, 2006; Goswami and David Polly, 2010; Goswami et al., 2014; Grabowski and Porto, 2016). Additionally, as discussed in the introduction, character correlation can also simply arise by chance due to the discrete coding scheme (i.e. some sets of characters can be highly correlated but effectively describe independent information). Therefore, we made the choice to simplify our simulations by generating character correlation as a stochastic process rather than a biological one.

Third, comparing phylogenetic inference methods is not trivial. As mentioned above, both maximum parsimony and Bayesian inference, although aiming (and often achieving) to infer evolutionary history only have similar outputs and vastly differ in how optimality is measured. But there are also difficulties in summarising both methods with consensus trees OReilly and Donoghue (2017). However, we want to point out again that here we're no comparing the methods to each other *per se* but rather how they both, individually, react to an increase or decrease of correlated characters.

Potential applications

Effectively, our simulation protocol bootstraps our data "with bias". In the "randomised" scenarios the data is simply randomly bootstrapped simply we randomly remove and resample characters (i.e. giving the weight of 0 to some and > 1 to other). However, in the "minimised" and "maximised" scenario, the bootstrapping we remove the characters with the lowest/highest overall character difference. For example, in the "maximised" scenario, we randomly remove some characters that are strongly correlated with other and randomly resample from the left characters.

It is noteworthy to point that in rather small matrices (25 × 100), there was no significant difference in terms of recovering the right topology when maximising or randomising the character differences. Since many discrete morphological matrices are of similar size (Guillerme and Cooper, 2016a) a simple bootstrap re-sampling (i.e. the equivalent of randomising the character differences in our analysis) will be sufficient to obtain a robust topology (*cf.* actively collecting different characters). In matrices with more taxa, however, the "maximised" scenario resulted in better topological recovery than any other scenarios. Applying this kind of bootstraps that maximises character difference by biasing the random sampling could thus results in better resolved trees.

Conclusion

Correlation between characters can be induced through three main phenomena: intra-organismal relationships, selection-driven covariation or biases in coding the characters yet only the latter can be improved upon to investigate phylogenetic relationships. Useful best practices guidelines (e.g. Brazeau, 2011; Simões et al., 2017) and algorithms for dealing with different types of character correlations (e.g. for characters hierarchy **?**Brazeau et al., 2017) already exist. However, with the regain of popularity in discrete morphological data and the expansion of dataset size (e.g. Ni et al., 2013; O'Leary et al., 2013, with more than 1000 characters each), we can expect the correlation between characters to increase stochastically. Moreover, because phylogenetic inference software are unable to *a priori* differentiate these difference correlations, it is important to understand to what extant topologies can be induced by such bias.

We found that character differences as a proxy for character correlation have a strong effect on recovering the "normal" topology: when character correlation was high (low character differences), the topology was always the furthest away from the "normal" topology. Conversely, when correlation between characters was low, the topology was always the closest to the "normal" topology. These results seem independent on the size of the matrix (number of taxa and/or characters) but can be influenced by the phylogenetic inference method used with Bayesian inference faring better in terms of clade conservation, especially in larger matrices.

However, in modest size matrices (25 taxa; 100 to 350 characters), the effect of actively choosing to minimise character correlation was not more significant than simply bootstrapping the matrix, suggesting that character correlation is more a problem in large discrete morphological matrices. For such matrices, minimising the character correlation (resampling characters $< 25\%$ different) or maximising it ($> 75\%$) respectively significantly decreased and increased correct topological recovery compared to randomly resample matrices.

Data availability, repeatability and reproducibility

The consensus trees are available on figshare at `https://figshare.com/s/7a8fde8eaa39a3d3cf56`. The simulations are fully replicable following the explanations at `https://github.com/TGuillerme/CharactersCorrelation`. The post-simulation analysis, tables and figures (reported in this manuscript) are fully reproducible see (`https://github.com/TGuillerme/CharactersCorrelation`).

Funding

European Research Council under the European Unions Seventh Framework Programme (FP/20072013)/ERC Grant Agreement number 311092.

Acknowledgements

Calculations where done using the Imperial College London Cluster Services (doi: 10.14469/hpc/2232). We thank Alberto Pascual Garcia for input in the design of the simulations protocol and Guillermo Herraiz Yebes for helping with the CD metric proof.

*

References

Bhattacharyya, A. 1943. On a measure of divergence between two statistical populations defined by their probability distributions. Bulletin of the Calcutta Mathematical Society 35:99–109.

Bogdanowicz, D., K. Giaro, and B. Wróbel. 2012. TreeCmp: Comparison of trees in polynomial time. Evolutionary Bioinformatics 8:475–487.

Brazeau, M. D. 2011. Problematic character coding methods in morphology and their effects. Biological Journal of the Linnean Society 104:489–498.

Brazeau, M. D., T. Guillerme, and M. R. Smith. 2017. Morphological phylogenetic analysis with inapplicable data. bioRχiv .

Dávalos, L. M., P. M. Velazco, O. M. Warsi, P. D. Smits, and N. B. Simmons. 2014. Integrating incomplete fossils by isolating conflicting signal in saturated and non-independent morphological characters. Systematic Biology 63:582–600.

De Laet, J. 2015. Parsimony analysis of unaligned sequence data: maximization of homology and minimization of homoplasy, not minimization of operationally defined total cost or minimization of equally weighted transformations. Cladistics 31:550–567.

Dobson, A. J. 1975. Comparing the shapes of trees vol. 452 of *Lecture Notes in Mathematics* Pages 95–100. Springer Berlin Heidelberg.

Dollo, L. 1893. Les lois de l'évolution. Bull Soc Belge Geol Pal Hydr Page 164166.

Douady, C., F. Delsuc, Y. Boucher, W. Doolittle, and E. Douzery. 2003. Comparison of bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Molecular Biology and Evolution 20:248–254.

Felsenstein, J. 1985. Phylogenies and the comparative method. The American Naturalist 125:1–15.

Felsenstein, J. 2004. Inferring phylogenies vol. 2. Sinauer Associates Sunderland.

FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution 3:1084–1092.

FitzJohn, R. G., M. W. Pennell, A. E. Zanne, P. F. Stevens, D. C. Tank, and W. K. Cornwell. 2014. How much of the world is woody? Journal of Ecology 102:1266–1272.

Goloboff, P. A., A. T. Galvis, J. S. Arias, and A. Smith. 2018. Parsimony and modelbased phylogenetic methods for morphological data: comments on o'reilly etal. Palaeontology 0.

Goloboff, P. A., A. Torres, and J. S. Arias. 2017. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. Cladistics .

Goswami, A. and P. David Polly. 2010. The influence of character correlations of phylogenetic analyses: a case study of the carnivoran cranium. Pages 141–164 *in* Carnivoran Evolution: New Views on Phylogeny, Form, and Function. (A. Goswami and A. Friscia, eds.). Cambridge University Press, Cambridge.

Goswami, A. and C. Janis. 2006. Morphological integration in the carnivoran skull. Evolution 60:169–183.

Goswami, A., J. Smaers, C. Soligo, and P. Polly. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. Phil. Trans. R. Soc. B 369:20130254.

Gower, J. C. 1971. A general coefficient of similarity and some of its properties. Biometrics 27:857–871.

Grabowski, M. and A. Porto. 2016. How many more? sample size determination in studies of morphological integration and evolvability. Methods in Ecology and Evolution Pages n/a–n/a.

Guillerme, T. 2016. dispRity: v0.2.

Guillerme, T. and N. Cooper. 2016a. Assessment of available anatomical characters for linking living mammals to fossil taxa in phylogenetic analyses. Biology letters 12:20151003.

Guillerme, T. and N. Cooper. 2016b. Effects of missing data on topological inference using a total evidence approach. Molecular Phylogenetics and Evolution 94, Part A:146 – 158.

Hasegawa, M., H. Kishino, and T. A. Yano. 1985. Dating of the human ape splitting by a molecular clock of mitochondrial-DNA. Journal of Molecular Evolution 22:160–174.

Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Sciences 111:E2957–E2966.

Hollander, M., D. A. Wolfe, and E. Chicken. 2013. Nonparametric statistical methods. John Wiley & Sons.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics Pages 65–70.

ICHPC. 2011. Imperial college high performance computing service. `http://www.imperial.ac.uk/admin-services/ict/self-service/research-support/hpc/`.

Joysey, K. A. and A. E. Friday. 1982. Problems of phylogenetic reconstruction. London etc.:[Systematics Assn. Spec. vol. No. 21.] Academic Press 576:561.

Kelly, E. M. and K. E. Sears. 2010. Reduced phenotypic covariation in marsupial limbs and the implications for mammalian evolution. Biological Journal of the Linnean Society 102:22–36.

Lewis, P. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic Biology 50:913–925.

Maddison, W. P. 1990. A method for testing the correlated evolution of two binary characters: Are gains or losses concentrated on certain branches of a phylogenetic tree? Evolution 44:539–557.

Mark Pagel, A. E. B. J. C. E. J. B. L., Andrew Meade. 2006. Bayesian analysis of correlated evolution of discrete characters by reversiblejump markov chain monte carlo. The American Naturalist 167:808–825.

Ni, X., D. Gebo, M. Dagosto, J. Meng, P. Tafforeau, J. Flynn, and K. Beard. 2013. The oldest known primate skeleton and early haplorhine evolution. Nature 498:60–64.

O'Leary, M. A., J. I. Bloch, J. J. Flynn, T. J. Gaudin, A. Giallombardo, N. P. Giannini, S. L. Goldberg, B. P. Kraatz, Z.-X. Luo, J. Meng, X. Ni, M. J. Novacek, F. A. Perini, Z. S. Randall, G. W. Rougier, E. J. Sargis, M. T. Silcox, N. B. Simmons, M. Spaulding, P. M. Velazco, M. Weksler, J. R. Wible, and A. L. Cirranello. 2013. The placental mammal ancestor and the post-K-Pg radiation of placentals. Science 339:662–667.

O'Reilly, J. E., M. N. Puttick, L. Parry, A. R. Tanner, J. E. Tarver, J. Fleming, D. Pisani, and P. C. J. Donoghue. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. Biology Letters 12.

O'Reilly, J. E., M. N. Puttick, D. Pisani, P. C. J. Donoghue, and A. Smith. 2018. Empirical realism of simulated data is more important than the model used to generate it: a reply to goloboff etal. Palaeontology 0.

OReilly, J. E. and P. C. Donoghue. 2017. The efficacy of consensus tree methods for summarising phylogenetic relationships from a posterior sample of trees estimated from morphological data. Systematic biology .

Pagel, M. 1994. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. Proceedings of the Royal Society of London B: Biological Sciences 255:37–45.

Puttick, M. N., J. E. O'Reilly, A. R. Tanner, J. F. Fleming, J. Clark, L. Holloway, J. Lozano-Fernandez, L. A. Parry, J. E. Tarver, D. Pisani, et al. 2017. Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. Proceedings of the Royal Society B 284:20162290.

Robinson, D. F. and L. R. Foulds. 1981. Comparison of phylogenetic trees. Mathematical Biosciences 53:131–147.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539–42.

Russell Lande, S. J. A. 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.

Simões, T. R., M. W. Caldwell, A. Palci, and R. L. Nydam. 2017. Giant taxon-character matrices: quality of character constructions remains critical regardless of size. Cladistics 33:198–219.

Soto, D. X., J. Benito, E. Gacia, E. García-Berthou, and J. Catalan. 2016. Trace metal accumulation as complementary dietary information for the isotopic analysis of complex food webs. Methods in Ecology and Evolution 7:910–918.

Spencer, M. R. and E. W. Wilberg. 2013. Efficacy or convenience? Model-based approaches to phylogeny estimation using morphological data. Cladistics 29:663–671.

Stoessel, A., B. M. Kilbourne, and M. S. Fischer. 2013. Morphological integration versus ecological plasticity in the avian pelvic limb skeleton. Journal of Morphology 274:483–495.

Swofford, D. L. 2001. Paup*: Phylogenetic analysis using parsimony (and other methods) 4.0. b5 .

Wilkinson, M. 1995. Coping with abundant missing entries in phylogenetic inference using parsimony. Syst. Biol. 44:501–514.

Wright, A. M. and D. M. Hillis. 2014. Bayesian analysis using a simple Likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. PLoS ONE 9:e109210.

Wright, A. M., G. T. Lloyd, and D. M. Hillis. 2016. Modeling character change heterogeneity in phylogenetic analyses of morphology through the use of priors. Systematic Biology 65:602–611.

Wright, A. M., K. M. Lyons, M. C. Brandley, and D. M. Hillis. 2015. Which came first: the lizard or the egg? robustness in phylogenetic reconstruction of ancestral states. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 324:504–516.

Zou, Z. and J. Zhang. 2016. Morphological and molecular convergences in mammalian phylogenetics. Nature Communications 7:12758 EP –.