A new phylogenetic protocol: Dealing with model misspecification and confirmation bias in molecular phylogenetics

Lars S Jermiin[1,2,3,4,*], Renee A Catullo[1,2,5], Barbara R Holland[6]

[1] CSIRO Land & Water, Canberra, ACT 2601, Australia

[2] Research School of Biology, Australian National University, Canberra, ACT 2601, Australia

[3] School of Biology & Environment Science, University College Dublin, Belfield, Dublin 4, Ireland

[4] Earth Institute, University College Dublin, Belfield, Dublin 4, Ireland

[5] School of Science and Health & Hawkesbury Institute of the Environment, Western Sydney University, Penrith, NSW, Australia

[6] School of Natural Sciences, University of Tasmania, TAS 7001, Australia

[*]Correspondence should be addressed to L.S.J. (lars.jermiin@anu.edu.au)

Dedication: To the memory of Rossiter H. Crozier (1943-2009), an evolutionary biologist, who with his generosity and inquisitiveness inspired many students and scientists, in Australia and abroad.

Keywords: Phylogenetic protocol; Multiple sequence alignment; Phylogenetic trees; Model selection; Model misspecification; Confirmation bias; Parametric bootstrap; Goodness-of-fit.

Molecular phylogenetics plays a key role in comparative genomics and has an increasingly-significant impact on science, industry, government, public health, and society. We report that the current phylogenetic protocol is missing two critical steps, and that their absence allows model misspecification and confirmation bias to unduly influence the phylogenetic estimates. Based on the potential offered by well-established but under-used procedures (i.e., assessment of phylogenetic assumptions and test of goodness-of-fit), we introduce a new phylogenetic protocol that will reduce confirmation bias and increase the accuracy of phylogenetic estimates.

Molecular phylogenetics plays a pivotal role in the analysis of genomic data and has already had a significant, wide-reaching impact in science, industry, government, public health, and society (**Table 1**). Although the science and methodology behind applied phylogenetics is increasingly well understood within parts of the scientific community[42], there is still a worryingly large body of research where the phylogenetic component was done with little attention to the consequences of a statistical misfit between the phylogenetic data and the assumptions embedded in the phylogenetic methods.

One reason for this is that molecular phylogenetics relies extensively on mathematics, statistics, and computer science, and many users of phylogenetic methods find the relevant subsections of these disciplines challenging to comprehend. A second reason is that methods and software often are chosen because they are easy to use and comprehend, or simply already popular, rather than because they are the most appropriate for the scientific questions and phylogenetic data at hand. A third reason is that much of the phylogenetic research done so far has relied on phylogenetic protocols[43-48], which have evolved to become a standard to which it seems sensible to adhere. Although these protocols vary, they have, at their core, a common set of sensible features that are linked in a seemingly logical manner (see below).

Here we report that, although the current phylogenetic protocol has many useful features, it

is missing two crucial components whereby the quality of fit between the data and models applied is assessed. This means that using the phylogenetic protocol in its current form may lead to biased conclusions. We suggest a modification to the protocol that will make it more robust and reliable.

The current phylogenetic protocol

Phylogenetic analysis of alignments of nucleotides or amino acids usually follows a protocol like that in **Figure 1**. Initially, the phylogenetic data are chosen on the basis of the assumption that the data will allow the researchers to solve their particular scientific questions. This choice of sequence data is often based on prior knowledge, developed locally or gleaned from the literature, and recommendations. Then, a multiple sequence alignment (MSA) method is chosen, often on the basis of prior experience with a specific method. The sequences are then aligned—the aim is to obtain an MSA, wherein homologous characters (i.e., nucleotides or amino acids) are identified and aligned. In practice, it is often necessary to insert gaps between the characters in some of the sequences to obtain an optimal MSA, and, in some cases, there may be sections of the MSA that cannot be aligned reliably.

Then follows the task of selecting the sites that will be used in the phylogenetic analysis. The rationale behind doing so is to maximize the signal-to-noise ratio in the MSA. By deleting poorly-aligned and highly-variable sections of the MSA, which are thought to create noise due to the difficulty of establishing homology, it is hoped that the resulting sub-MSA retains a strong historical signal[49] that will allow users to obtain a well-supported and well-resolved phylogeny. The choice of sites to retain is made by visual inspection of the MSA or by using purpose-built software[50-54]. The automated ways of filtering MSAs have been questioned[55].

Having obtained a sub-MSA, the next step in the protocol is to select a phylogenetic method for analysis of the data. Importantly, this means that it is assumed that the sequences have diverged along the edges of a single bifurcating tree (the tree-likeness assumption) and that the evolutionary processes operating at the variable sites are independent and identically-

distributed processes (the IID assumption). If model-based molecular phylogenetic methods are chosen, it is also assumed that the evolutionary processes operating at the variable sites can be modelled accurately by time-reversible Markov models, and that the processes were stationary, reversible and homogeneous[56-58] over time (the assumption of SRH conditions). In practice, the choice is one between methods assuming that the underlying evolutionary process can be modelled using a Markov model of nucleotide or amino-acid substitutions (i.e., distance methods[59-64], likelihood methods[59, 61, 63-69], Bayesian methods[70-75]), or using non-parametric phylogenetic methods (i.e., parsimony methods[59, 61, 63, 64, 76-78]). In reality, most researchers analyze their data by using different model-based phylogenetic methods, and reports that only use parsimony methods are increasingly rare. Depending on the chosen phylogenetic method, researchers may have to select a suitable model of sequence evolution (i.e., a substitution model and a rate-across-sites model) to apply to the sub-MSA. This choice is often made by using model-selection methods[6, 79-90].

Having chosen the phylogenetic method and, in relevant cases, a suitable model of sequence evolution, the next step involves obtaining accurate estimates of the tree and evolutionary processes that led to the data. There is a plethora of programs that implement phylogenetic methods[59-78]. Depending on the methods chosen, users often also obtain the nonparametric bootstrap probability[91] or clade credibility[92] to measure support for individual divergence events in the phylogeny. These estimates are often thought of as measures of the accuracy of the phylogenetic estimate or the confidence we might have in the inferred divergence events. Doing so might be unwise because they are only measures of consistency[93] (e.g., a phylogenetic estimate may consistently point to an incorrect tree).

Having inferred the phylogeny, with or without bootstrap probability or clade credibility for all of the internal branches (edges), the final step in the protocol is to interpret the result. Under some conditions—most commonly the inclusion of out-group sequences—the tree can be drawn and interpreted as a rooted phylogeny, in which case the order of divergence events and the lengths of the individual edges may be used to infer, for example, the tempo

and mode of evolution of the data. The inferred phylogeny often confirms an earlier-reported or assumed evolutionary relationship. Often, too, there are surprises that are difficult to understand and explain. If the phylogenetic estimate is convincing and newsworthy, the discoveries may be reported, for example, through papers in peer-reviewed journals.

On the other hand, if the surprises are too numerous or do not appear credible, the researchers will begin the task of finding out what may have 'gone wrong' during the phylogenetic analysis. This process is illustrated as dashed feedback loops in **Figure 1**. The researchers may examine the data using alternative methods: use other Markov models, employ different phylogenetic methods, use a different sub-MSA, align the sequences differently, use a different alignment method, or use another data set. Given enough patience, the researches may reach a conclusion about the data, and they may decide to publish their results.

Problems with the current phylogenetic protocol

Although the current phylogenetic protocol has led to many fine scientific discoveries, it also has left many scientists with strong doubts about or, alternatively, unduly strong confidence in the estimates. The literature is rife with examples where analyses of the same data have led to disagreements among experts about what is the 'right' phylogeny (cf. e.g., [94-96]). Such disagreements can be confusing, especially for non-experts and the public. To avoid this, it is necessary to understand the challenges that applied phylogenetic research still faces.

While it is clear that the right data are needed to answer the scientific question at hand, making that choice is not always as trivial as it might seem. In some cases, the sequences may have evolved too slowly and/or be too short, in which case there may not be enough information in the data, or they have evolved so fast that the historical signal has largely been lost[97]. In rarely-reported cases, the data are not what they purport to be[98].

Next, there is no consensus on what constitutes an optimal MSA. Clearly, what is required is an accurate MSA where every site is a correct homology statement. Currently, however, there is no automatic procedure for homology assessment, so to infer an accurate MSA is still more an art than science[99], putting the whole phylogenetic protocol into jeopardy. One way to mitigate this problem is to rely on simulation-based comparisons and reviews of MSA methods[99-107], but they appear to have had less impact than deserved.

The choice to remove poorly-aligned or highly-variable sites is confounded by the fact that the different MSA methods frequently return different MSAs, implying different homology statements—they cannot all be right. However, the choice of what sites to retain depends not only on the MSA method used but also on how difficult it is to identify the sites (e.g., it is impractical to visually inspect and edit MSAs with over ~50 sequences and ~400 sites). In the past, expert knowledge about the data was often applied (e.g., structural information about the gene or gene product), but automated methods[50-54] are now typically used. However, these methods often produce different sub-MSAs from the same MSA, leaving confusion and doubt.

The choice of what phylogenetic method to use for the data is rated by many as the most challenging one to make (e.g., because the assumptions of each method are often poorly understood), and it is often solved by using several phylogenetic methods. If these methods return the same phylogenetic tree, many authors feel confident that they have inferred the 'true' phylogeny and would go on to report their discoveries. However, while this approach may have led to correct trees, it is perhaps more due to luck than to scientific rigor that the right tree was identified. This is because every phylogenetic method is based on assumptions (see above), and if these assumptions are not violated too strongly by the data, the true tree has a high probability of being found. On the other hand, if these violations are strong, there is currently no way of knowing whether the true tree was found. Indeed, strong violation of phylogenetic assumptions could lead to similar but nevertheless wrong trees being inferred using different phylogenetic methods[49, 108].

Over the last two decades, the choice of a suitable model of sequence evolution has often been made by using purpose-built model-selection methods[6, 79-90]. Assuming a tree, these methods step through an array of predefined models, evaluating each of them, one by one, until the list of models is exhausted. This is sensible if the true or most appropriate model is included in the set of predefined models. On the other hand, if this model is not included in the set of predefined models, the popular model-selection methods may never be able to return an accurate estimate. They will return an optimal estimate, but it will be conditional on the models considered. Importantly, most popular model-selection methods only consider time-reversible Markov models. If the data have evolved on a single tree but under more complex conditions, then there is no way that a simple, time-reversible Markov model is sufficient to approximate the evolutionary processes across all edges of the tree[110]. Hence, it is worrying that researchers still ignore or dismiss the implication of compositional heterogeneity across sequences[108]: it implies that the evolutionary process for a set of sites has changed over time (e.g., the third position of codons has evolved under different evolutionary processes across time, requiring multiple models of sequence evolution for these data). This implication must be taken seriously when data are analyzed phylogenetically; typically, it is not.

The choice of phylogenetic program is often driven more by prior experiences and transaction costs (i.e., the time it takes to become a confident and competent user of the software) rather than by a profound understanding of the strengths, limitations, and weaknesses of the available software. This may not substantially impact the accuracy of the phylogenetic estimate, as long as the data are consistent with the phylogenetic assumptions of the methods and the methods thoroughly search tree space and model space.

Finally, once a well-supported phylogenetic estimate has been obtained, a researcher's prior expectations are likely to influence whether the results are considered both reliable and newsworthy. In some cases, where information on the phylogeny is known (e.g., serially-

sampled viral genomes), not meeting the prior expectations may signal an issue with the phylogenetic analysis. However, if a researcher's expectations are confirmed by the phylogenetic estimates, it is more likely that a report will be written without a thorough assessment of what might have gone wrong during the analysis. This tendency to let prior expectations influence the interpretation of phylogenetic estimates is called confirmation bias. Confirmation bias is not discussed in phylogenetics, even though it is recognized as a critical factor in other disciplines (e.g., psychology and social science[111]), so it is timely that the phylogenetic community takes onboard the serious implications of this.

The new phylogenetic protocol

Although the current phylogenetic protocol has many shortcomings, it has many desirable attributes, including that it is easy to follow and implement as a pipeline. However, to mitigate its limitations, it is necessary to redesign the protocol to accommodate well-established but largely-ignored procedures and new feedback loops.

Figure 2 shows our proposal for new phylogenetic protocol. It shares many features found in the current protocol (e.g., the first four steps), but the fifth step (assess phylogenetic assumptions) is novel. Because all phylogenetic methods are based on assumptions, it is sensible to validate these assumptions at this point in the protocol. Since many phylogenetic methods assume that the data (e.g., different genes) have evolved over the same tree, and that the chosen data partitions have evolved independently under the same time-reversible Markovian conditions, it is wise to survey the sub-MSA for evidence that the sequences actually have evolved under these conditions. If the data violate the phylogenetic assumptions of some methods, then it would be wise to avoid these phylogenetic methods and to employ other such methods. Alternatively, it may be worth following the relevant feedback loops in **Figure 2**—perhaps something led to a biased sub-MSA? The relevance and benefits of this step are illustrated using a case study (**Box 1**), which focuses on determining whether a data set is consistent with the phylogenetic assumption of evolution under time-

reversible conditions. Assessments of other phylogenetic assumptions require other types of tests and surveys (it is beyond the scope of this review to discuss these here).

Next follows the choice of phylogenetic method, but this choice is now made on the basis of the previous step, rather than cultural or computational reasons. If the sequences have evolved on a single tree under time-reversible Markovian conditions, there is a large set of phylogenetic methods to choose from[59-78]. On the other hand, if these data have evolved under more complex Markovian conditions, the number of suitable phylogenetic methods is frustratingly limited[5, 57, 112-136], and most of these methods are aimed at finding the optimal model of sequence evolution for a given tree rather than finding the optimal tree. Users of phylogenetic methods therefore are sometimes confronted by a dilemma: Do they abandon their data set because it has evolved under non-time-reversible condition and because there are no phylogenetic methods for such data, or do they take the risk and use robust time-reversible phylogenetic methods? Fortunately, there is a way around this dilemma (see below).

Having inferred the phylogeny using model-based phylogenetic methods, it is possible to test the fit between tree, model and data (step 10 of the new protocol). A suitable test of goodness-of fit was proposed in 1993[137] (**Fig. 4**). In brief, using the inferred optimal tree (with edge lengths included), it is possible to simulate data sets under the null model (i.e., the inferred optimal model of sequence evolution with its parameter values included). This is called a parametric bootstrap. Given this tree and this model of sequence evolution, several sequence-generating programs[5, 126, 138, 139] facilitate procurement of pseudo-data. Having generated, say, $n = 1{,}000$ pseudo-data, the next step involves finding the difference (δ) between the unconstrained (i.e., without assuming a tree and a model) and constrained (i.e., assuming a tree and a model) log-likelihoods (i.e., $\delta = lnL(\mathbf{D}) - lnL(\mathbf{D}|T, M)$, where $\mathbf{D}$ is the data, T is the tree, and M is the model of sequence evolution). If the estimate of δ is greater for the real data than for the pseudo-data, then that result reveals a poor fit between tree, model, and data[109]. The approach described here works well for likelihood-

based phylogenetic analysis and a similar approach is available for Bayesian-based phylogenetic analysis[140]. Parametric bootstrapping is computationally expensive and time-consuming, so it should only be done if the data appears to meet the assumptions of phylogenetic method. The advantages of using such a goodness-of-fit test is that it allows users to determine if the lack of fit is large enough to not be due to chance. It does not say anything about whether or not the lack of fit matters. If the fit is poor, then the relevant feedback loops should be followed (**Fig. 2**)—perhaps a biasing factor was missed? If the phylogenetic tree and model of sequence evolution are found to fit the data, then that implies that these estimates represent a plausible explanation of the data. It is these estimates that should be reported, but only as one plausible explanation, not as the only possible explanation. This is because there may be other plausible explanations of the data that never were considered during the analysis. [739 words]

The future: Areas in most need of methodological research

Adherence to the new phylogenetic protocol would undoubtedly lead to improved accuracy of phylogenetic estimates and a reduction of confirmation bias. The advantage of the fifth step in the new phylogenetic protocol (i.e., assess phylogenetic assumptions) is that users are able to decide how to do the most computationally-intensive parts of the phylogenetic study without wasting valuable time on, for example, a high-performance computer centre. Model selection, phylogenetic analysis, and parametric bootstrapping are computationally-intensive and time-consuming, and there is a need for new, computationally efficient strategies that can be used to analyse sequences that have evolved under complex phylogenetic conditions.

The advantage of the tenth step in the new phylogenetic protocol (i.e., test goodness-of-fit) is its ability to answer whether an inferred phylogeny explains the data well, or not. In so doing, this step tackles the issue of confirmation bias front on. Clearly, without information gleaned from the fifth step, the parametric bootstrap might return an unwanted answer

(i.e., the inferred tree and model of sequence evolution does not fit the data well), so to avoid such disappointments it is better to embrace the new phylogenetic protocol in full.

Results emerging from studies that rely on the new phylogenetic protocol might well call into question published phylogenetic research, but there is also a chance that research might gain stronger support. This is good for everyone concerned, especially since it will become easier to defend the notion that the research was done without prejudice or preference for a particular result. Objectivity should be restored in phylogenetics—it is no longer reasonable to defend phylogenetic results on the basis that they were obtained using the best available tools; if these tools do not model the evolutionary processes accurately, then that should be reported rather than be hidden away. This is critical as it increases transparency and aids other researchers to understand the nature of the challenges encountered.

Notwithstanding the likely benefits offered by the new phylogenetic protocol and the methods supporting it, it would be unwise to assume that further development of phylogenetic methods will no longer be needed. On the contrary, there is a lot of evidence that method development will be needed in different areas:

- *MSA Methods* — There is a dire need for MSA methods that are accurate in the sense of homology statements. Likewise, there is a great need for methods that allow users (*i*) to determine how accurate different MSA methods are and (*ii*) to select MSA methods that are most suitable for the data at hand.
- *Methods for Masking MSAs* — Assuming an MSA has been inferred, there is a need for a set of strategies that can be used to identify and distinguish between poorly-aligned and highly-variable regions of MSA. Well aligned but highly-variable regions of MSAs may be more informative than poorly-aligned regions of such MSAs, so to delete them may be unwise.
- *Model-selection Methods* — Model selection is important when parametric phylogenetic methods are used. However, the model-selection methods currently

employed may not be accurate, especially for sequences that have evolved under complex conditions (e.g., heterotachous, covarion, or non-time-reversible conditions). Critically, the evolutionary process may have be considered an evolving entity in its own right.

- *Phylogenetic Methods* — While there is a plethora of accurate phylogenetic methods for analysis of data that have evolved under time-reversible Markovian conditions, there is a dearth of accurate phylogenetic methods suitable for analysis of data that have evolved under complex conditions. Added to this challenge are methods that accurately consider incomplete lineage sorting of genetic markers and the special conditions associated with the analysis of SNP data.
- *Goodness-of-fit Tests* — Although suitable goodness-of-fit tests are available, there is not only a need for a wider understanding of the merits of these tests, but also of how they can be tailored to suit different requirements. In particular, there is a need for programs that can generate pseudo data under extremely complex evolutionary conditions. Some programs are available[5, 126], but they only cater for a limited set of conditions.
- *Analysis of Residuals* — Although goodness-of-fit tests can tell you whether or not the lack of fit observed is potentially due to chance, they do not answer the more useful question of whether or not that lack of fit matters or how the lack of fit arises[141, 142]. For this reason, residual diagnostic tools that can inform the user about the way in which their model fails to fit the data would be very useful.

In summary, while calls for better phylogenetic methods and more careful considerations of the data have occurred[110], we believe there is a need for a comprehensive overhaul of the current phylogenetic protocol. The proposed new phylogenetic protocol is unlikely to be the final product; rather, it is probably a first, but important step towards a more scientifically sound phylogenetic protocol, which not only will lead to more accurate phylogenetic estimates and but also to a reduction in the likelihood of confirmation bias.

Conclusions

The Holy Grail in molecular phylogenetics is clearly being able to obtain accurate, reproducible, transparent, and trustworthy phylogenetic estimates from the phylogenetic data. We are not there yet, but encouraging progress is being made in not only in the design of the phylogenetic protocol but also in phylogenetic methodology based on the likelihood and Bayesian optimality criteria.

Notwithstanding this progress, a quantum shift in attitudes and habits will be needed within the phylogenetic community—it is no longer sufficient to obtain an optimal phylogenetic estimate. The fit between trees, models, and data must be evaluated before the phylogenetic estimates can be considered newsworthy. We owe it to the community and wider public to be as rigorous as we can—the attitude “She’ll be alright, mate” is no longer appropriate in this discipline.

ACKNOWLEDGEMENTS

L.S.J. thanks the University College Dublin for its generous hospitality. We thank D. Higgins, A. Locatelli, and K. H. Wolfe for their constructive feedback.

AUTHOR CONTRIBUTIONS

L.S.J. conceived the new phylogenetic protocol. R.A.C., B.R.H., and L.S.J. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

BOX 1 — CASE STUDY

To illustrate the relevance and benefits of the fifth step in the new phylogenetic protocol, we examined the phylogenetic data used to infer the evolution of insects[3]. The tetrahedral plots in **Figure 3a-3c** reveal that the nucleotide composition at the three codon positions is heterogeneous, implying that the evolutionary processes that operated at these positions are unlikely to have been time-reversible. However, the plots are deceptive because the presence of constant sites (i.e., sites with the same nucleotide or amino acid) in the data can mask how compositionally dissimilar the sequences actually are. To learn how to resolve this issue, it is necessary to focus on the evolution of two sequences on a tree (**Fig. 3d**) and the corresponding divergence matrix at time 0 (**Fig. 3e**) and at time t (**Fig. 3f**). At time 0, the two sequences are beginning to diverge from one another, so the off-diagonal elements of the divergence matrix are all zero. Later, the divergence matrix may look like that in **Figure 3f**. All the off-diagonal elements are now greater than zero, and the so-called matching off-diagonal elements of the divergence matrix might differ (i.e., $x_{ij} \neq x_{ji}$). The degree of divergence between the two sequences can be inferred by comparing the off-diagonal elements to the diagonal elements, while the degree of difference between the two evolutionary processes can be inferred by comparing the above-diagonal elements to the below-diagonal elements. If the two evolutionary processes were the same, the matching off-diagonal elements in **Figure 3f** would be similar. A lack of symmetry (i.e., $x_{ij} \neq x_{ji}$) implies that the evolutionary processes along the two descendant lineages may be different. A matched-pairs test of symmetry[143] can be used to determine whether this observed deviation from symmetry is statistically significant. **Figures 3g-3i** show the distributions of the observed and expected p values from these tests for the data assessed in **Figures 3a-3c**. Because the dots in these plots do not fall along the diagonal line in the plots (showing that a lack of symmetry is not statistically significant), there is an overwhelming evidence that the evolutionary processes at these positions cannot have been time-reversible. The same is the case for the corresponding amino acid alignment (not shown). Consequently, it would be unwise to assume that the data evolved under time-reversible conditions. A far more complex evolutionary process is likely to explain the data, so the time-reversible

phylogenetic methods used by Misof et al.[3] were clearly not suitable for analysis of these data. However, such methods were not available at the time, and that is still the case!

TABLE 1.

Examples of phylogenetic research, divided into high-level areas, based in impact or relevance.

Area	Examples of phylogenetic research
Science	Provide accurate estimates of the evolution of, for example, species[1-4]
	Provide accurate estimates of evolutionary processes at the molecular level[5, 6]
	Addressing macroevolutionary questions pertaining to birds and mammals[7, 8]
	Understand biogeographic patterns and diversity[9, 10]
	Reconstruction of ancestral states[11, 12]
Industry	Facilitate the design and engineering of novel enzymes[13] and drugs[14-16]
Government	Reveal likely sources and dispersal routes of agricultural pests and pathogens[17-21]
	Assign conservation priorities to species or biogeographic regions based on estimates of genetic diversity[22-24]
Public health	Reveal the origin and spread of human pathogens[25-28]
	Predict the evolution of human influenza A[29]
	Reveal the origin and evolution of cancers[30, 31]
Society	Reveal the evolution of human language[32, 33]
	Map the relationship among ancient texts[34], tales[35], and music[36]
	Reveal evolution of humans since their divergence from other primates[37-41]

Figure 1. The current phylogenetic protocol. Solid arrows show the order of actions normally taken during a phylogenetic analysis. Dashed arrows show feedback loops often employed in phylogenetic research. For details, see the main text.

Figure 2. A new phylogenetic protocol. Solid arrows show the order of actions normally taken during a phylogenetic analysis. Dashed arrows show feedback loops often employed in phylogenetic research. For details, see the main text.

Figure 3. Illustrating the merits of the fifth step in the new phylogenetic protocol. Panels **a** - **c** show the average nucleotide composition at first, second, and third codon position of the sequences (144 sequences, 413,459 codons) originally examined by Misof et al.[3]. Each dot represents the nucleotide composition of a single sequence. The spread of dots in panels **a** - **c** reveals compositional heterogeneity at first, second, and third codon position, indicating these data violate important assumptions underlying most phylogenetic methods. The tetrahedral plots were generated using SeqVis[144]. Panel **d** shows a two-tipped tree with an ancestral sequence and two diverged sequences. Panels **e** and **f** show the divergence matrix for these sequences at time 0 and time t. Each number in a cell of a divergence matrix corresponds to the number of sites with nucleotide i in one sequence and nucleotide j in the other. Panels **h** - **i** shows the PP plots for the data already analysed in panels **a** - **c**. A total of 10,296 tests were done for each of the three codon positions using Homo 1.3 (http://www.csiro.au/homo/).

Figure 4. Diagram showing the parametric bootstrap procedure that may be used to conduct a suitable goodness-of-fit test. The procedure includes three steps. For details, see the main text.

References

1. dos Reis, M. et al. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. B* **279**, 3491-3500 (2012).
2. Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E. & Burleigh, J.G. From algae to angiosperms—inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. *BMC Evol. Biol.* **14**, 26 (2014).
3. Misof, B. et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763-767 (2014).
4. Prum, R.O. et al. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569-U247 (2015).
5. Jayaswal, V., Wong, T.K.F., Robinson, J., Poladian, L. & Jermiin, L.S. Mixture models of nucleotide sequence evolution that account for heterogeneity in the substitution process across sites and across lineages. *Syst. Biol.* **63**, 726-742 (2014).
6. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.F.K., Von Haeseler, A. & Jermiin, L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Meth.* **14**, 587-589 (2017).
7. Penny, D. & Phillips, M.J. The rise of birds and mammals: are microevolutionary processes sufficient for macroevolution. *Trends Ecol. Evol.* **19**, 516-522 (2004).
8. Meredith, R.W. et al. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521-524 (2011).
9. Knapp, M. et al. Relaxed molecular clock provides evidence for long-distance dispersal of Nothofagus (southern beech). *PLoS Biol.* **3**, 38-43 (2005).
10. Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K. & Mooers, A.O. The global diversity of birds in space and time. *Nature* **491**, 444-448 (2012).
11. Marazzi, B. et al. Locating evolutionary precursors on a phylogenetic tree. *Evolution* **66**, 3918-3930 (2012).
12. Pagel, M., Meade, A. & Barker, D. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53**, 673-684 (2004).
13. Wilding, M. et al. Reverse engineering: transaminase biocatalyst development using ancestral sequence reconstruction. *Green Chem.* **19**, 5375-5380 (2017).
14. Searls, D.B. Pharmacophylogenomics: Genes, evolution and drug targets. *Nat. Rev. Drug Discov.* **2**, 613-623 (2003).
15. Goodfellow, M. & Fiedler, H.P. A guide to successful bioprospecting: informed by actinobacterial systematics. *Antonie Van Leeuwenhoek* **98**, 119-142 (2010).
16. Wright, G.D. & Poinar, H. Antibiotic resistance is ancient: implications for drug discovery. *Trends Microbiol.* **20**, 157-159 (2012).
17. Boykin, L.M., Armstrong, K.F., Kubatko, L.S. & De Barro, P.J. Species delimitation and global biosecurity. *Evol. Bioinform.* **8**, 1-37 (2012).
18. Hosokawa, T., Nikoh, N. & Fukatsu, T. Fine-Scale Geographical Origin of an Insect Pest Invading North America. *PLoS One* **9**, 5 (2014).
19. Yasaka, R. et al. Phylodynamic evidence of the migration of turnip mosaic potyvirus from Europe to Australia and New Zealand. *J. Gen. Virol.* **96**, 701-713 (2015).
20. Tay, W.T. et al. Mitochondrial DNA and trade data support multiple origins of Helicoverpa armigera (Lepidoptera, Noctuidae) in Brazil. *Scientific Rep.* **7**, 45302 (2017).

21. Anderson, C.J. et al. Hybridization and gene flow in the mega-pest lineage of moth, Helicoverpa. *Proc. Natl. Acad. Sci. USA* **115**, 5034-5039 (2018).
22. Gonzalez-Orozco, C.E. et al. Phylogenetic approaches reveal biodiversity threats under climate change. *Nat. Clim. Chang.* **6**, 1110-+ (2016).
23. Rosauer, D.F. et al. Phylogeography, hotspots and conservation priorities: an example from the Top End of Australia. *Biol Conserv* **204**, 83-93 (2016).
24. Tucker, C.M. et al. A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biol Rev* **92**, 698-715 (2017).
25. Andersen, K.G. et al. Clinical sequencing uncovers origins and evolution of Lassa Virus. *Cell* **162**, 738-750 (2015).
26. Holmes, E.C., Dudas, G., Rambaut, A. & Andersen, K.G. The evolution of Ebola virus: Insights from the 2013-2016 epidemic. *Nature* **538**, 193-200 (2016).
27. Lanciotti, R.S., Lambert, A.J., Holodniy, M., Saavedra, S. & Signor, L.D.C. Phylogeny of Zika Virus in Western Hemisphere, 2015. *Emerg. Infect. Dis.* **22**, 933-935 (2016).
28. Lessler, J. et al. Assessing the global threat from Zika virus. *Science* **353**, 10 (2016).
29. Bush, R.M., Bender, C.A., Subbarao, K., Cox, N.J. & Fitch, W.M. Predicting the evolution of human influenza A. *Science* **286**, 1921-1925 (1999).
30. Alves, J.M., Prieto, T. & Posada, D. Multiregional tumor trees are not phylogenies. *Trends Cancer* **3**, 546-550 (2017).
31. Schwartz, R. & Schaffer, A.A. The evolution of tumour phylogenetics: principles and practice. *Nat. Rev. Genet.* **18**, 213-229 (2017).
32. Pagel, M. Human language as a culturally transmitted replicator. *Nat. Rev. Genet.* **10**, 405-415 (2009).
33. Bouckaert, R. et al. Mapping the origins and expansion of the Indo-European language family. *Science* **337**, 957-960 (2012).
34. Barbrook, A.C., Howe, C.J., Blake, N. & Robinson, P. The phylogeny of The Canterbury Tales. *Nature* **394**, 839-839 (1998).
35. Tehrani, J.J. The phylogeny of Little Red Riding Hood. *PLoS One* **8**, 11 (2013).
36. Windram, H.F., Charlston, T. & Howe, C.J. A phylogenetic analysis of Orlando Gibbons's Prelude in G. *Early Music* **42**, 515-+ (2014).
37. Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708-713 (2000).
38. Ke, Y.H. et al. African origin of modern humans in East Asia: A tale of 12,000 Y chromosomes. *Science* **292**, 1151-1153 (2001).
39. Schraiber, J.G. & Akey, J.M. Methods and models for unravelling human evolutionary history. *Nat. Rev. Genet.* **16**, 727-740 (2015).
40. Posth, C. et al. Pleistocene mitochondrial genomes suggest a single major dispersal of Non-Africans and a late glacial population turnover in Europe. *Curr. Biol.* **26**, 827-833 (2016).
41. Nielsen, R. et al. Tracing the peopling of the world through genomics. *Nature* **541**, 302-310 (2017).
42. Yang, Z.H. & Rannala, B. Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* **13**, 303-314 (2012).
43. Harrison, C.J. & Langdale, J.A. A step by step guide to phylogeny reconstruction. *Plant J* **45**, 561-572 (2006).
44. Hunt, T. & Vogler, A.P. A protocol for large-scale rRNA sequence analysis: Towards a detailed phylogeny of Coleoptera. *Mol. Phylogenet. Evol.* **47**, 289-301 (2008).

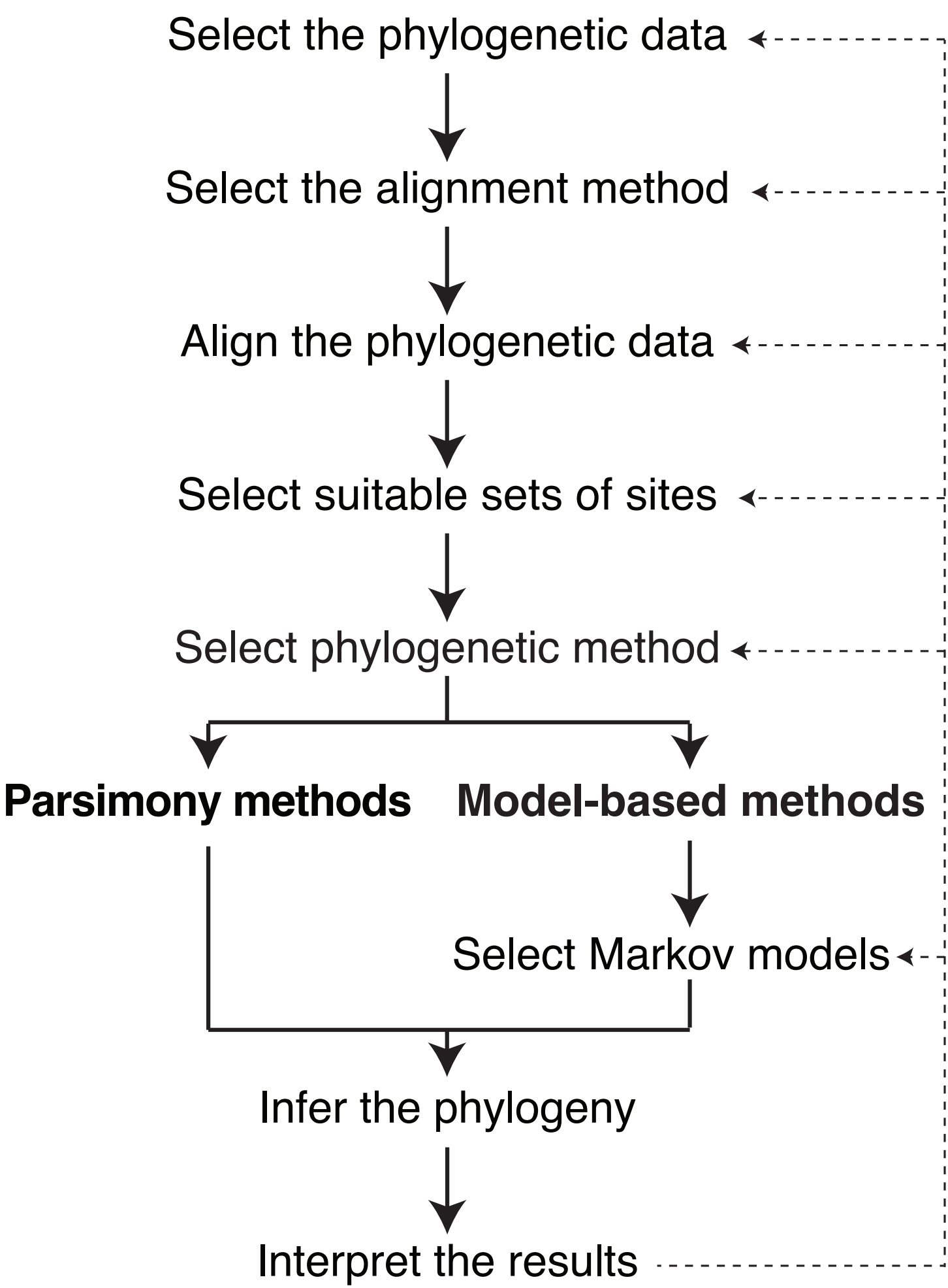
45. Hall, B.G. Building phylogenetic trees from molecular data with MEGA. *Mol. Biol. Evol.* **30**, 1229-1235 (2013).
46. Lemmon, E.M. & Lemmon, A.R. High-throughput genomic data in systematics and phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* **44**, 99-121 (2013).
47. O'Halloran, D. A oractical guide to phylogenetics for nonexperts. *J. Vis. Exp.*, 14 (2014).
48. Wilding, M., Nachtschatt, M., Speight, R. & Scott, C. An improved and general streamlined phylogenetic protocol applied to the fatty acid desaturase family. *Mol. Phylogenet. Evol.* **115**, 50-57 (2017).
49. Ho, S.Y.W. & Jermiin, L.S. Tracing the decay of the historical signal in biological sequence data. *Syst. Biol.* **53**, 623-637 (2004).
50. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564-577 (2007).
51. Misof, B. & Misof, K. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Syst. Biol.* **58**, 21-34 (2009).
52. Kück, P. et al. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front. Zool.* **7**, 10 (2010).
53. Penn, O., Privman, E., Landan, G., Graur, D. & Pupko, T. An alignment confidence score capturing robustness to guide tree uncertainty. *Mol. Biol. Evol.* **27**, 1759-1767 (2010).
54. Wu, M.T., Chatterji, S. & Eisen, J.A. Accounting for alignment uncertainty in phylogenomics. *PLoS One* **7**, e30288 (2012).
55. Tan, G. et al. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst. Biol.* **64**, 778-791 (2015).
56. Bryant, D., Galtier, N. & Poursat, M.-A. in Mathematics of Evolution and Phylogeny. (ed. O. Gascuel) 33-62 (Oxford University Press, Oxford; 2005).
57. Jayaswal, V., Jermiin, L.S. & Robinson, J. Estimation of phylogeny using a general Markov model. *Evol. Bioinform.* **1**, 62-80 (2005).
58. Ababneh, F., Jermiin, L.S. & Robinson, J. Generation of the exact distribution and simulation of matched nucleotide sequences on a phylogenetic tree. *J. Math. Model. Algor.* **5**, 291-308 (2006).
59. Swofford, D.L., Edn. 4 (Sinauer Associates, Sunderland, Massachusetts.; 2003).
60. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289-290 (2004).
61. Felsenstein, J., Edn. 3.6 (Distributed by the author, Seattle; 2005).
62. Popescu, A.A., Huber, K.T. & Paradis, E. ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* **28**, 1536-1537 (2012).
63. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870-1874 (2016).
64. Xia, X.H. DAMBE6: New tools for microbial genomics, phylogenetics, and molecular evolution. *J. Hered.* **108**, 431-437 (2017).
65. Knight, R. et al. PyCogent: a toolkit for making sense from sequence. *Genome Biol.* **8**, 16 (2007).

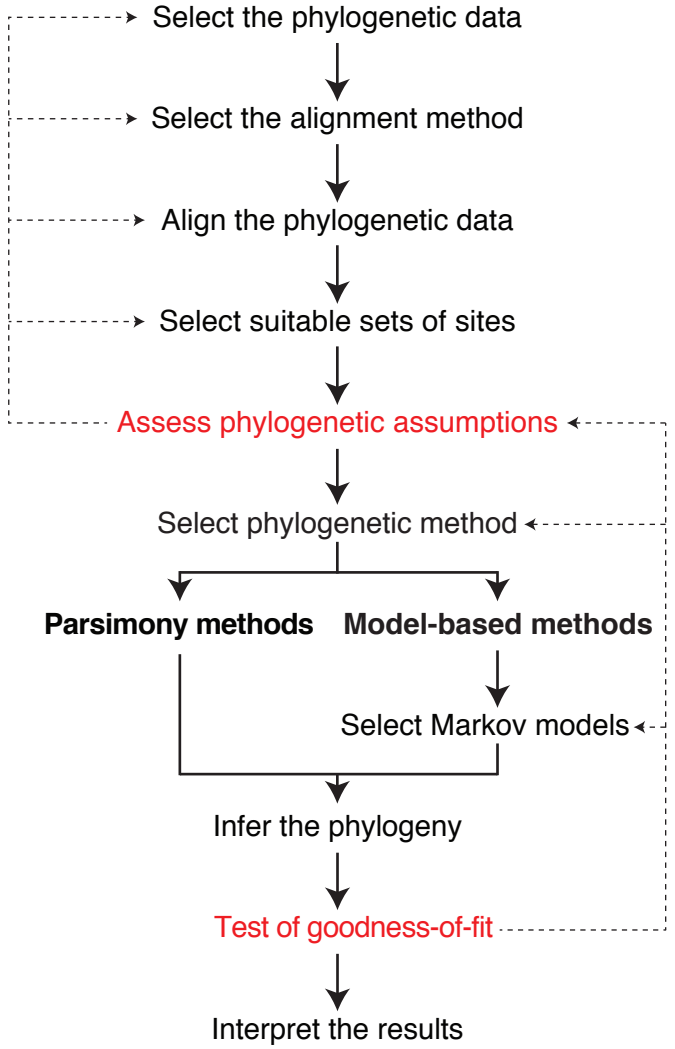
66. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307-321 (2010).
67. Bazinet, A.L., Zwickl, D.J. & Cummings, M.P. A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. *Syst. Biol.* **63**, 812-818 (2014).
68. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
69. Nguyen, L.-T., Schmidt, H.A., Von Haeseler, A. & Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268-274 (2015).
70. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286-2288 (2009).
71. Ronquist, F. et al. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539-542 (2012).
72. Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: Phylogenetic Reconstruction with Infinite Mixtures of Profiles in a Parallel Environment. *Syst. Biol.* **62**, 611-615 (2013).
73. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comp. Biol.* **10**, 6 (2014).
74. Höhna, S. et al. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* **65**, 726-736 (2016).
75. Ogilvie, H.A., Heled, J., Xie, D. & Drummond, A.J. Computational performance and statistical accuracy of *BEAST and comparisons with other methods. *Syst. Biol.* **65**, 381-396 (2016).
76. Goloboff, P.A., Farris, J.S. & Nixon, K.C. TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774-786 (2008).
77. Goloboff, P.A. & Catalano, S.A. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* **32**, 221-238 (2016).
78. White, W.T.J. & Holland, B.R. Faster exact maximum parsimony search with XMP. *Bioinformatics* **27**, 1359-1367 (2011).
79. Posada, D. & Crandall, K.A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818 (1998).
80. Chiotis, M., Jermiin, L.S. & Crozier, R.H. A molecular framework for the phylogeny of the ant subfamily Dolichoderinae. *Mol. Phylogenet. Evol.* **17**, 108-116 (2000).
81. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-2105 (2005).
82. Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J. & McInerney, J.O. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* **6**, 29 (2006).
83. Posada, D. ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online. *Nucl. Acid. Res.* **34**, W700-W703 (2006).
84. Posada, D. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253-1256 (2008).
85. Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164-1165 (2011).

86. Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nature Meth.* **9**, 772 (2012).
87. Lanfear, R., Calcott, B., Ho, S.Y.W. & Guindon, S. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29**, 1695-1701 (2012).
88. Santorum, J.M., Darriba, D., Taboada, G.L. & Posada, D. jmodeltest.org: selection of nucleotide substitution models on the cloud. *Bioinformatics* **30**, 1310-1311 (2014).
89. Whelan, S., Allen, J.E., Blackburne, B.P. & Talavera, D. ModelOMatic: Fast and automated model selection between RY, nucleotide, amino acid, and codon substitution models. *Syst. Biol.* **64**, 42-55 (2015).
90. Lefort, V., Longueville, J.E. & Gascuel, O. Replace!!! SMS: Smart model selection in PhyML. *Mol. Biol. Evol.* **34**, 2422-2424 (2017).
91. Minh, B.Q., Nguyen, M.A.T. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188-1195 (2013).
92. Larget, B. & Simon, D. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* **16**, 750-759 (1999).
93. Jermiin, L.S., Poladian, L. & Charleston, M.A. Evolution - Is the "Big Bang" in animal evolution real? *Science* **310**, 1910-1911 (2005).
94. Goremykin, V.V. et al. The evolutionary root of flowering plants. *Syst. Biol.* **62**, 50-61 (2013).
95. Drew, B.T. et al. Another look at the root of the angiosperms reveals a familiar tale. *Syst. Biol.* **63**, 368-382 (2014).
96. Goremykin, V.V., Nikiforova, S.V., Cavalieri, D., Pindo, M. & Lockhart, P. The root of flowering plants and total evidence. *Syst. Biol.* **64**, 879-891 (2015).
97. Rokas, A., Krüger, D. & Carroll, S.B. Animal evolution and the molecular signature of radiations compressed in time. *Science* **310**, 1933-1938 (2005).
98. Catullo, R.A. & Oakeshott, J.G. Problems with data quality in the reconstruction of evolutionary relationships in the Drosophila melanogaster species group: Comments on Yang et al. (2012). *Mol. Phylogenet. Evol.* **78**, 275-276 (2014).
99. Morrison, D.A. Is sequence alignment an art or a science? *Syst. Bot.* **40**, 14-26 (2015).
100. Morrison, D.A. Multiple sequence alignment for phylogenetic purposes. *Aust. Syst. Bot.* **19**, 479-539 (2006).
101. Golubchik, T., Wise, M.J., Easteal, S. & Jermiin, L.S. Mind the gaps: Evidence of bias in estimates of multiple sequence alignments. *Mol. Biol. Evol.* **24**, 2433-2442 (2007).
102. Morrison, D.A. A framework for phylogenetic sequence alignment. *Plant Syst. Evol.* **282**, 127-149 (2009).
103. Morrison, D.A. Why would phylogeneticists ignore computerized sequence alignment? *Syst. Biol.* **58**, 150-158 (2009).
104. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
105. Thompson, J.D., Linard, B., Lecompte, O. & Poch, O. A comprehensive benchmark study of multiple sequence alignment methods: Current challenges and future perspectives. *PloS One* **6**, 14 (2011).
106. Chatzou, M. et al. Multiple sequence alignment modeling: methods and applications. *Brief. Bioinf.* **17**, 1009-1023 (2016).
107. Chowdhury, B. & Garai, G. A review on multiple sequence alignment from the perspective of genetic algorithm. *Genomics* **109**, 419-431 (2017).

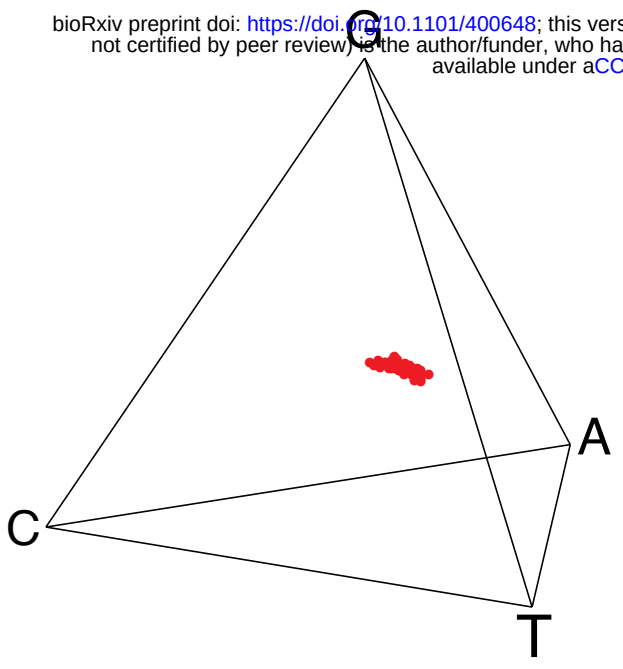
108. Jermiin, L.S., Ho, S.Y.W., Ababneh, F., Robinson, J. & Larkum, A.D.W. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Syst. Biol.* **53**, 638-643 (2004).
109. Jermiin, L.S., Jayaswal, V., Ababneh, F.M. & Robinson, J. in Bioinformatics: Volume 1: Data, Sequence Analysis, and Evolution. (ed. J. Keith) 379-420 (Humana Press, Totowa, NJ; 2017).
110. Cooper, E.D. Overly simplistic substitution models obscure green plant phylogeny. *Trends Plant Sci.* **19**, 576-582 (2014).
111. Winking, J. Exploring the great schism in the Social Sciences: Confirmation bias and the interpretation of results relating to biological influences on human behavior and psychology. *Evol. Psychol.* **16**, 10 (2018).
112. Barry, D. & Hartigan, J.A. Statistical analysis of hominoid molecular evolution. *Stat. Sci.* **2**, 191-210 (1987).
113. Reeves, J. Heterogeneity in the substitution process of amino acid sites of proteins coded for by the mitochondrial DNA. *J. Mol. Evol.* **35**, 17-31 (1992).
114. Steel, M.A., Lockhart, P.J. & Penny, D. Confidence in evolutionary trees from biological sequence data. *Nature* **364**, 440-442 (1993).
115. Lake, J.A. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Proc. Natl. Acad. Sci. USA* **91**, 1455-1459 (1994).
116. Lockhart, P.J., Steel, M.A., Hendy, M.D. & Penny, D. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11**, 605-612 (1994).
117. Steel, M.A. Recovering a tree from the leaf colourations it generates under a Markov model. *Appl. Math. Lett.* **7**, 19-23 (1994).
118. Galtier, N. & Gouy, M. Inferring phylogenies from DNA sequences of unequal base compositions. *Proc. Natl. Acad. Sci. USA* **92**, 11317-11321 (1995).
119. Steel, M.A., Lockhart, P.J. & Penny, D. A frequency-dependent significance test for parsimony. *Mol. Phylogenet. Evol.* **4**, 64-71 (1995).
120. Yang, Z. & Roberts, D. On the use of nucleic acid sequences to infer early branches in the tree of life. *Mol. Biol. Evol.* **12**, 451-458 (1995).
121. Gu, X. & Li, W.-H. Bias-corrected paralinear and logdet distances and tests of molecular clocks and phylogenies under nonstationary nucleotide frequencies. *Mol. Biol. Evol.* **13**, 1375-1383 (1996).
122. Galtier, N. & Gouy, M. Inferring pattern and process: maximum-likelihood implementation of a nonhomogenous model of DNA sequence evolution for phylogenetic analysis. *Mol. Biol. Evol.* **15**, 871-879 (1998).
123. Gu, X. & Li, W.-H. Estimation of evolutionary distances under stationary and nonstationary models of nucleotide substitution. *Proc. Natl. Acad. Sci. USA* **95**, 5899-5905 (1998).
124. Galtier, N., Tourasse, N. & Gouy, M. A nonhyperthermophilic common ancestor to extant life forms. *Science* **283**, 220-221 (1999).
125. Tamura, K. & Kumar, S. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. *Mol. Biol. Evol.* **19**, 1727-1736 (2002).
126. Foster, P.G. Modelling compositional heterogeneity. *Syst. Biol.* **53**, 485-495 (2004).
127. Thollesson, M. LDDist: a Perl module for calculating LogDet pair-wise distances for protein and nucleotide sequences. *Bioinformatics* **20**, 416-418 (2004).

128. Blanquart, S. & Lartillot, N. A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Mol. Biol. Evol.* **23**, 2058-2071 (2006).
129. Jayaswal, V., Robinson, J. & Jermiin, L.S. Estimation of phylogeny and invariant sites under the General Markov model of nucleotide sequence evolution. *Syst. Biol.* **56**, 155-162 (2007).
130. Blanquart, S. & Lartillot, N. A site- and time-heterogeneous model of amino acid replacement. *Mol. Biol. Evol.* **25**, 842-858 (2008).
131. Dutheil, J. & Boussau, B. Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. *BMC Evol. Biol.* **8**, 255 (2008).
132. Jayaswal, V., Jermiin, L.S., Poladian, L. & Robinson, J. Two stationary, non-homogeneous Markov models of nucleotide sequence evolution. *Syst. Biol.* **60**, 74-86 (2011).
133. Jayaswal, V., Ababneh, F., Jermiin, L.S. & Robinson, J. Reducing model complexity when the evolutionary process over an edge is modeled as a homogeneous Markov process. *Mol. Biol. Evol.* **28**, 3045-3059 (2011).
134. Dutheil, J.Y. et al. Efficient selection of branch-specific models of sequence evolution. *Mol. Biol. Evol.* **29**, 1861-1874 (2012).
135. Zou, L.W., Susko, E., Field, C. & Roger, A.J. Fitting nonstationary general-time-reversible models to obtain edge-lengths and frequencies for the Barry-Hartigan model. *Syst. Biol.* **61**, 927-940 (2012).
136. Groussin, M., Boussau, B. & Gouy, M. A branch-heterogeneous model of protein evolution for efficient inference of ancestral sequences. *Syst. Biol.* **62**, 523-538 (2013).
137. Goldman, N. Statistical tests of models of DNA substitution. *J. Mol. Evol.* **36**, 182-198 (1993).
138. Rambaut, A. & Grassly, N.C. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* **13**, 235-238 (1997).
139. Fletcher, W. & Yang, Z.H. INDELible: a flexible simulator of biological sequence evolution. *Mol. Biol. Evol.* **26**, 1879-1888 (2009).
140. Bollback, J.P. Bayesian model adequacy and choice in phylogenetics. *Mol. Biol. Evol.* **19**, 1171-1180 (2002).
141. Kumar, S., Filipski, A.J., Battistuzzi, F.U., Pond, S.L.K. & Tamura, K. Statistics and truth in phylogenomics. *Mol. Biol. Evol.* **29**, 457-472 (2012).
142. Holland, B.R. The rise of statistical phylogenetics. *Aust. N. Zea. J. Stat.* **55**, 205-220 (2013).
143. Ababneh, F., Jermiin, L.S., Ma, C. & Robinson, J. Matched-pairs tests of homogeneity with applications to homologous nucleotide sequences. *Bioinformatics* **22**, 1225-1231 (2006).
144. Ho, J.W.K. et al. SeqVis: Visualization of compositional heterogeneity in large alignments of nucleotides. *Bioinformatics* **22**, 2162-2163 (2006).

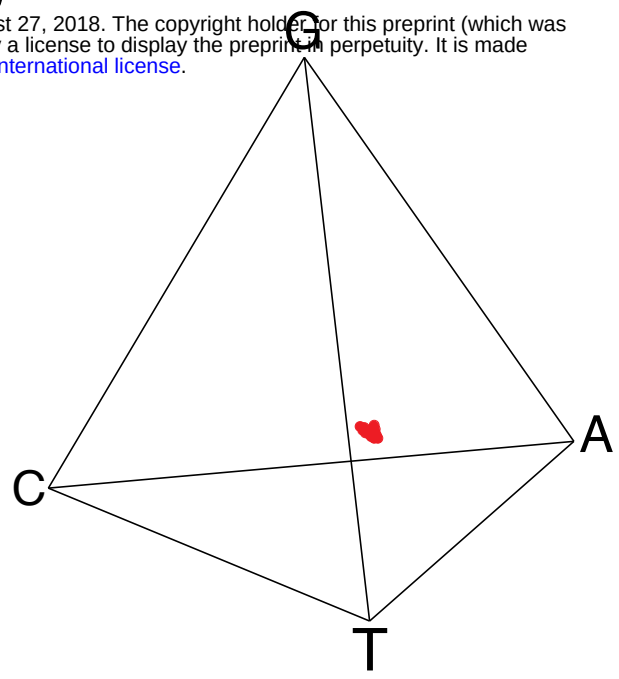

Select the phylogenetic data
Select the alignment method
Align the phylogenetic data
Select suitable sets of sites
Select phylogenetic method
Parsimony methods
Model-based methods
Select Markov models
Infer the phylogeny
Interpret the results


Select the phylogenetic data
Select the alignment method
Align the phylogenetic data
Select suitable sets of sites
Assess phylogenetic assumptions
Select phylogenetic method
Parsimony methods
Model-based methods
Select Markov models
Infer the phylogeny
Test of goodness-of-fit
Interpret the results

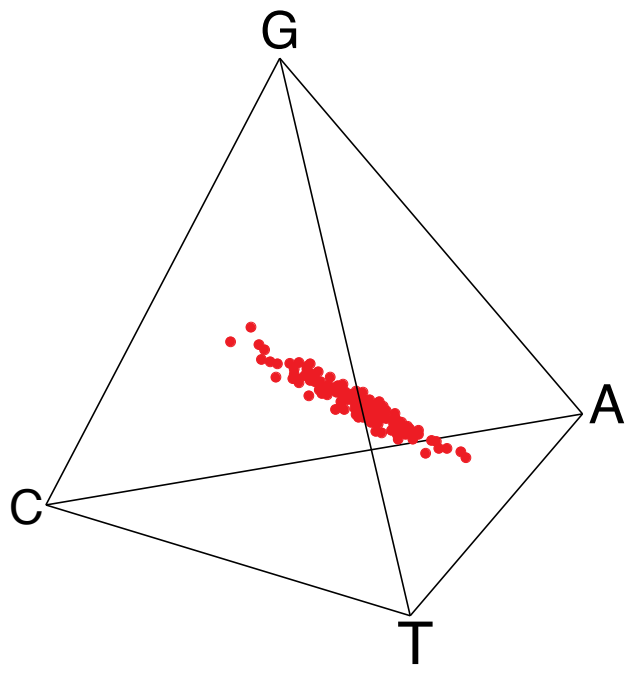
a



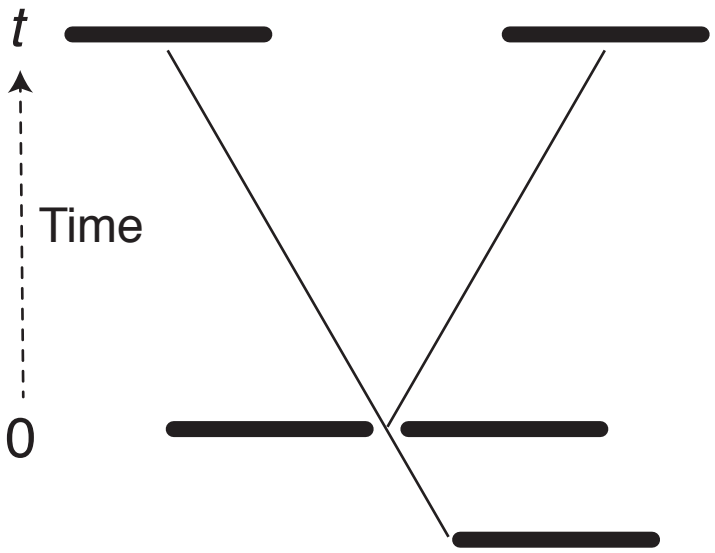
b



c



d



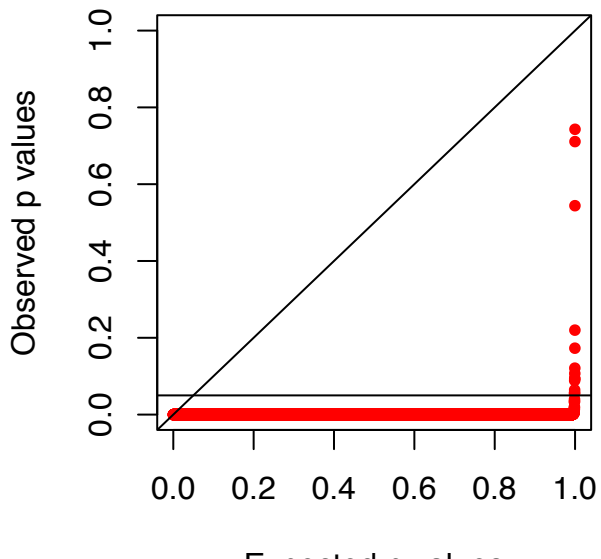
e

Time = 0	A	C	G	T
A	294	0	0	0
C	0	374	0	0
G	0	0	829	0
T	0	0	0	655

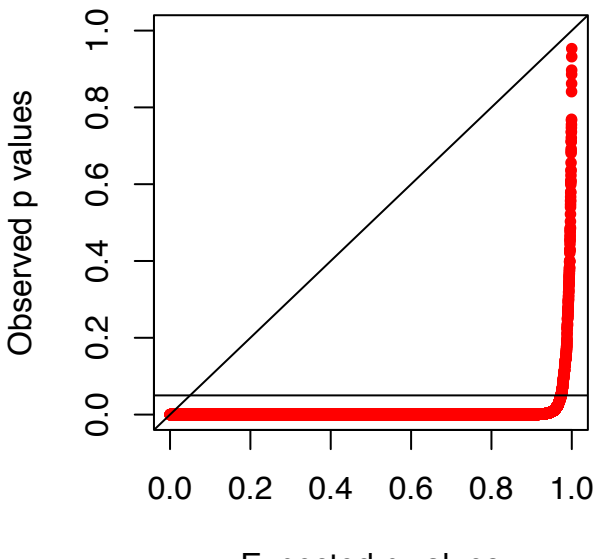
f

Time = *t*	A	C	G	T
A	139	68	68	19
C	13	152	23	12
G	10	31	134	9
T	39	59	57	167

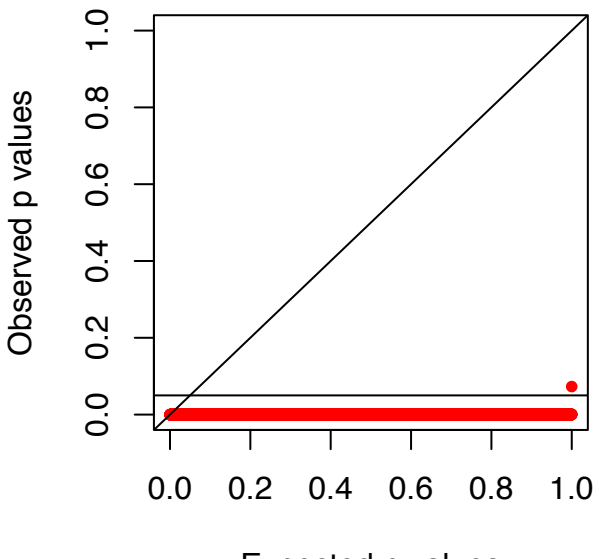
g

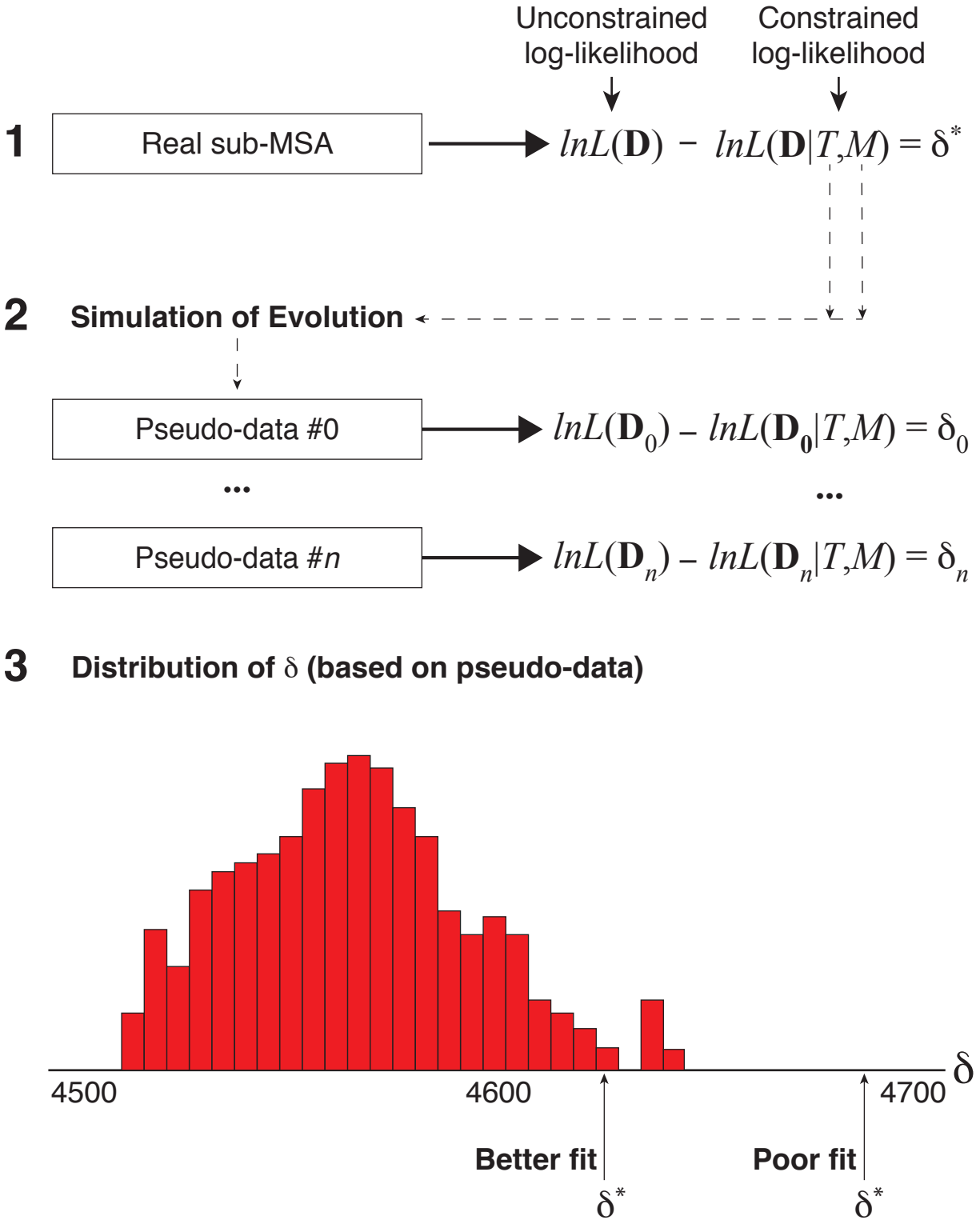


h



i




Unconstrained log-likelihood
Constrained log-likelihood
1
Real sub-MSA
$lnL(\mathbf{D}) - lnL(\mathbf{D}|T,M) = \delta^*$
2
Simulation of Evolution
Pseudo-data #0
$lnL(\mathbf{D}_0) - lnL(\mathbf{D}_0|T,M) = \delta_0$
...
...
Pseudo-data #n
$lnL(\mathbf{D}_n) - lnL(\mathbf{D}_n|T,M) = \delta_n$
3
Distribution of δ (based on pseudo-data)
4500
4600
4700
δ
Better fit
Poor fit
δ^*
δ^*