Type: Article (Discoveries)

A test statistic to quantify treelikeness in phylogenetics

Caitlin Cherryh[1,*], Bui Quang Minh[1,2], Rob Lanfear[1]

[1] Division of Ecology and Evolution, Research School of Biology, Australian National University, Canberra, Australian Capital Territory, Australia

[2] Research School of Computer Science, Australian National University, Canberra, Australian Capital Territory, Australia

***Author for Correspondence:** email: caitlin.cherryh@anu.edu.au

Key words: phylogenetic inference, model violation, systematic bias

Abstract

Most phylogenetic analyses assume that the evolutionary history of an alignment (either that of a single locus, or of multiple concatenated loci) can be described by a single bifurcating tree, the so-called the treelikeness assumption. Treelikeness can be violated by biological events such as recombination, introgression, or incomplete lineage sorting, and by systematic errors in phylogenetic analyses. The incorrect assumption of treelikeness may then mislead phylogenetic inferences. To quantify and test for treelikeness in alignments, we develop a test statistic which we call the tree proportion. This statistic quantifies the proportion of the edge weights in a phylogenetic network that are represented in a bifurcating phylogenetic tree of the same alignment. We extend this statistic to a statistical test of treelikeness using a parametric bootstrap. We use extensive simulations to compare tree proportion to a range of related approaches. We show that tree proportion successfully identifies non-treelikeness in a wide range of simulation scenarios, and discuss its strengths and weaknesses compared to other approaches. The power of the tree-proportion test to reject non-treelike alignments can be lower than some other approaches, but these approaches tend to be limited in their scope and/or the ease with which they can be interpreted. Our recommendation is to test treelikeness of sequence alignments with both tree proportion and mosaic methods such as 3Seq. The scripts necessary to replicate this study are available at https://github.com/caitlinch/treelikeness

Introduction

A phylogenetic tree is a representation of the relationships between species or individuals. Many estimates of phylogenetic trees implicitly assume that sites in a sequence alignment share the same evolutionary history and conform to a single bifurcating tree. This assumption is called treelikeness. This concept was first introduced by Dress (1984) and was first used to assess how well data fit a tree by Eigen *et al.* (1988). Perfectly treelike alignments are likely to be rare not only due to noise such as sequencing or alignment error but also because biological processes like incomplete lineage sorting (ILS), recombination, or introgression mean even very short alignments may have an evolutionary history that cannot be represented by a single bifurcating phylogeny (Mallet *et al.* 2016; Mendes *et al.* 2019; Scornavacca and Galtier 2017). Although the treelikeness assumption is almost

42 universally made in phylogenetic analyses, it remains rare to test the validity of this
43 assumption. If treelikeness is incorrectly assumed, phylogenetic inferences may be misled
44 (Brown and Thomson 2018), so it is important to test whether the treelikeness assumption
45 holds prior to estimating a phylogenetic tree. Doing so may assist phylogeneticisits in
46 choosing the best combination of data and inference method with which to infer their
47 phylogeny.
48
49 Most estimates of phylogenetic trees assume that the data are treelike at some level.
50 Concatenation methods (also known as supermatrix methods) assume that all loci in an
51 alignment share a single evolutionary history. This approach has been criticised as the
52 histories of individual loci may vary dramatically, potentially resulting in incorrect
53 phylogenetic inferences (Shi and Yang 2018; Weisrock *et al.* 2012; Wielstra *et al.* 2014; Wu
54 *et al.* 2018; Zhao *et al.* 2016). Concatenating alignments from loci with different
55 evolutionary histories clearly violates the treelikeness assumption. Coalescent methods
56 improve on the concatenation method by explicitly incorporating non-treelikeness due to
57 ILS. To do so, coalescent methods allow each locus to have a separate tree topology and use
58 the distribution of these topologies to then infer the species tree under a model of ILS.
59 However, these methods still assume that the alignment used to produce each single-locus
60 tree is treelike. Recent work shows that different exons in the same gene often have
61 different evolutionary histories, meaning that in many cases the alignments used to
62 estimate single-locus trees may not be treelike (Mendes *et al.* 2019; Scornavacca and Galtier
63 2017). As a result, both concatenation and coalescent tree estimation methods may be
64 vulnerable to errors introduced by violation of the treelikeness assumption.
65
66 Previous studies have proposed a range of approaches to measuring certain aspects of non-
67 treelikeness. Goldman (1993) developed the first general test for model adequacy in
68 evolutionary models, which simultaneously assess all assumptions of the evolutionary
69 model, including the treelikeness assumption. Unfortunately, as all parameters of the model
70 are tested simultaneously, the treelikeness of the alignment cannot be extracted from the
71 results of this test. Additionally, several methods for visualizing treelikeness have been
72 suggested. Likelihood mapping (Strimmer and von Haeseler 1997) is primarily used to
73 visualise and estimate the phylogenetic signal of an alignment (Baric *et al.* 2003; Salzburger

et al. 2002; Steiner and Dreyer 2003) but has also been used to assess whether an alignment had a treelike structure (Nadan *et al.* 2003). The δ plot method (Holland *et al.* 2002) allows assessment of the treelikeness of an alignment using a mathematical approach based on assessing the treelikeness of all possible quartets of taxa in the alignment. The mean δ_q value from the δ plot method has been used to draw inferences about the overall treelikeness of an alignment, although the interpretation of this value varies: values of 0.11 (Kozak *et al.* 2015) and 0.28 (Short *et al.* 2014) have been suggested to indicate significant non-bifurcating signal, and a value of 0.18 was suggested to indicate that an alignment was treelike (Grimm and Renner 2013).

Phylogenetic split networks have also been used to visualise the treelikeness of an alignment. A split network generalises a phylogenetic tree by representing incompatible phylogenetic signals present in the sequence alignment as additional edges (Huson *et al.* 2010). Compared to a tree, a phylogenetic split network includes more information about the relationships between taxa as it includes conflicting phylogenetic signals and alternate histories (Bryant and Moulton 2004). The treelikeness of alignments can be visually determined based on the number, size and position of parallelograms within the network (Bryant and Moulton 2004; Kennedy *et al.* 2005; Kück *et al.* 2010). Phylogenetic networks provide a very useful visual tool for assessing treelikeness, but they can be somewhat difficult to interpret and there currently exists no framework for comparing the treelikeness of different alignments using networks. Ideally, a test statistic should quantify treelikeness in a way that is comparable across alignments and allow biologists to make informed decisions about which data and methods to use for inferring evolutionary histories.

Here, we introduce the tree proportion for quantifying the treelikeness of a multiple sequence alignment. The tree proportion of an alignment describes the proportion of a phylogenetic network of that alignment that can be represented by a single bifurcating tree. Specifically, it is the proportion of non-trivial split weights of an inferred network that are contained in a bifurcating phylogenetic tree of the same alignment. A split is trivial if one side of the bipartition contains only one taxon, i.e. terminal branches on a phylogeny represent trivial splits, because they are contained in all trees and networks and thus provide no information about treelikeness. Tree proportion ranges from 0 to 1, where a

score of 0 indicates that none of the non-trivial splits in the network are represented in the tree. A score of 1 indicates that all of the non-trivial splits in the network are represented in the tree (i.e. the alignment is perfectly treelike). More generally, the better that bifurcating phylogenetic tree represents an alignment, the closer the tree proportion will be to 1.

In addition to providing an intuitive measure of treelikeness, we describe how the tree proportion can be used to ask whether the assumption of treelikeness can be rejected for any given alignment. To do this, we use a parametric bootstrap to simulate treelike datasets with model parameters estimated from the original alignment, and then ask whether the observed tree proportion is surprisingly small relative to the tree proportions observed from the (truly treelike) simulated datasets. This allows us to generate a p-value for the test statistic, where a p-value < 0.05 indicates that the assumption of treelikeness can be rejected for the alignment in question. Finally, using introgression as a framework for simulating alignments of varying treelikeness, we demonstrate how the tree proportion can be used to quantify and test for treelikeness, and we compare its performance to previously suggested methods for estimating treelikeness, as well as certain measures that have been suggested specifically for testing introgression.

New approaches

Tree proportion

Tree proportion is defined as follows. For a split network denoted by (S, λ) where S is the set of non-trivial splits and λ is a split weight function, and a phylogenetic tree (T) the tree proportion is calculated as:

$$Tree\ proportion = \frac{\sum_{\sigma \in S \cap T} \lambda(\sigma)}{\sum_{\sigma \in S} \lambda(\sigma)} \quad (1)$$

In other words, tree proportion is the proportion of the total weight of non-trivial splits in the network that are represented by the tree. Figure 1 illustrates how to calculate the tree proportion for a simple five-taxon split network and tree.

Calculating tree proportion requires both a bifurcating phylogenetic tree and a phylogenetic split network estimated from the same alignment. In principle the split network and the

bifurcating phylogenetic tree could be inferred with any method. Indeed, the maximum tree proportion for any given split network can be calculated simply from using the maximum spanning tree of the network as the bifurcating tree. In this study, we used Maximum Likelihood to estimate bifurcating trees for two reasons: (1) Maximum Likelihood is one of the most commonly used methods for tree inference, and (2) Maximum Likelihood naturally allows us to extend our approach to include a parametric bootstrap test because it co-estimates the bifurcating tree and the parameters of a model of molecular evolution. We used NeighborNet (Bryant and Moulton 2004) to estimate the split network. NeighborNet is a distance based agglomerative method for generating replicable and statistically consistent split networks. NeighborNet measures conflict rather than evolutionary history, so the resulting network represents conflicting signals within the alignment.

To calculate the tree proportion of an alignment, we first estimated a NeighborNet network in SplitsTree v4.14.6 (Huson and Bryant 2006). Next, we estimated a maximum likelihood tree for the same alignment using IQ-Tree v2.0 with ModelFinder (Kalyaanamoorthy *et al.* 2017; Minh *et al.* 2020). Finally, we calculated tree proportion in R using code available from https://github.com/caitlinch/treelikeness.

Parametric bootstrap

Because we do not assume any prior distribution of tree proportion, we rely on a parametric bootstrap procedure to determine whether the tree proportion is significantly lower than would be expected for truly treelike alignments as follows. For a given alignment D, we reconstruct a maximum likelihood tree T_{ML} with the best-fit substitution model M. From T_{ML} and M we simulate n (n=199 by default) alignments $D_1, \ldots, D_n$. From each D_i we reconstruct a maximum likelihood tree T_i and a NeighborNet network S_i and use T_i and S_i to calculate the tree proportion TP_i. We calculate the statistics $TP_1, \ldots, TP_n$. The p-value is then computed as the fraction of TP_i greater than or equal to TP of the original alignment.

Results

Decreased treelikeness due to increasing proportions of introgressed DNA

Of the six test statistics for treelikeness we compared, tree proportion (R^2 = 0.863) and mean δ_q (R^2 = 0.818) showed the strongest correlations with the proportion of introgressed DNA (Figure 2). Both tree proportion and mean δ_q were strongly correlated with the proportion of introgressed DNA whether the simulated introgression was reciprocal or non-reciprocal (Supplementary Figure 1: Tree proportion R^2 = 0.863 and 0.704; mean δ_q R^2 = 0.818 and 0.902). Tree proportion was strongly correlated with the proportion of introgressed DNA regardless of the simulated tree depth (Supplementary Figure 2, all R^2 > 0.692). Mean δ_q showed strong correlations on tree depths up to 0.5 (Supplementary Figure 2, all R^2 > 0.705) but a much weaker correlation when the simulated tree depth was 1.0 (Supplementary Figure 2, R^2 = 0.344). The strength of the correlations between the proportion of introgressed DNA and the other four test statistics was highly variable and never higher than 0.511 under any simulation conditions (Figure 2, Supplementary Figures 1 and 2).

The ability of each test to statistically reject treelikeness under simulated introgression events varied greatly (Figure 3). PHI and 3SEQ had the highest power to reject non-treelike alignments, and the tests successfully detect 100% of introgression events after the proportion of introgressed DNA reached 0.2 and 0.1 respectively (Figure 3). Mean δ_q and tree proportion have intermediate results. At a proportion of introgressed DNA of 0.5, these tests detect 98% and 99% respectively of the alignments containing introgression (Figure 3). Proportion of resolved quartets and mode δ_q failed as statistical tests and did not successfully detect introgression events for any tree depth or event type (Figure 3, Supplementary Figures 3 and 4). Test results were very similar regardless of whether introgression was simulated as a reciprocal or a non-reciprocal event (Supplementary Figure 3). All test statistics except proportion of resolved quartets and mode δ_q have acceptable false positive rates (i.e., a significant result in approximately 5% of tests when there is no introgression in the simulation, shown as 0 on the x-axis of each panel in Figure 3). Both 3SEQ and PHI have less power to detect introgression at lower tree depths. For a tree depth of 0.05 substitutions per site and at a proportion of introgressed DNA of 0.5, PHI and 3SEQ correctly identify 96% and 95% of introgression events respectively (Supplementary Figure 4). Conversely, tree proportion has less power to detect introgression at higher tree depths. For a tree depth of 1 substitution per site at a proportion of introgressed DNA of 0.5, tree

198 proportion correctly identifies just over a third (35%) of introgression events
199 (Supplementary Figure 4).
200
201 ## Decreased treelikeness due to increasing number of introgression events
202 Of the six test statistics we compared, only 3SEQ and tree proportion revealed clear
203 decreases in treelikeness as the number of introgression events increased (Figure 4).
204 Encouragingly, tree proportion responded similarly for all tree depths and whether or not
205 the simulated recombination was reciprocal or non-reciprocal (Supplementary Figures 5 and
206 6). 3SEQ test statistic values were similar across reciprocal and non-reciprocal
207 recombination events (Supplementary Figure 5), but the range of values is dependent on
208 tree depth (Supplementary Figure 6). The PHI test statistic only identified nonreciprocal
209 introgression events (Supplementary Figure 5), and was more strongly correlated to the
210 number of introgression events at higher tree depths (Supplementary Figure 6). The
211 proportion of resolved quartets, mean δ_q and mode δ_q test statistic values showed at best
212 weak correlations with the number of introgression events, regardless of the simulation
213 conditions (Figure 4; Supplementary Figures 5 and 6).
214
215 Three test statistics performed well as statistical tests to reject treelikeness in the presence
216 of multiple introgression events (Figure 5). 3SEQ, tree proportion and PHI had the highest
217 power to reject non-treelike alignments, as the tests successfully detected 100% of
218 alignments containing introgression after 1, 2, and 3 events respectively (Figure 5). These
219 results were similar for nonreciprocal and reciprocal events (Supplementary Figure 7). The
220 best performing test statistic was tree proportion, which behaved similarly for all event
221 types and tree depths (Supplementary Figures 7 and 8). 3SEQ performed well at tree depth
222 of 0.5 substitutions per site (Figure 5) but showed very poor performance at the lowest tree
223 depth of 0.05 substitutions per site (Supplementary Figure 8). PHI also performed well at
224 tree depth of 0.5 substitutions per site (Figure 5), but its performance dropped below that
225 of tree proportion at low tree depths (Supplementary Figure 8). The other three statistics
226 (proportion of resolved quartets, mean δ_q and mode δ_q) were unable to reject treelikeness
227 for the simulated introgressed alignments at any tree depth or event type (Figure 5,
228 Supplementary Figures 7 and 8). Only PHI and tree proportion had acceptable false positive

rates (i.e., a significant result in approximately 5% of tests when there is no introgression in the simulation, shown as 0 on the x-axis of each panel in Figure 5).

Discussion

In this study, we introduce the tree proportion as a way of measuring treelikeness, and testing (with a parametric bootstrap) whether a single bifurcating phylogenetic tree is sufficient to explain the evolutionary history of an alignment. Importantly for a proposed measure of treelikeness, tree proportion values are easy to interpret: a value of 1 corresponds to a perfectly treelike alignment (i.e. one whose evolutionary history can be perfectly explained by a single bifurcating phylogenetic tree), and as treelikeness reduces the tree proportion will decrease towards zero. These properties mean that the tree proportion can be used to directly compare the treelikeness of different alignments.

We use a suite of simulations to compare the tree proportion to five other tests, both as a measure of treelikeness and as a statistical test that can be used to reject treelikeness for a given alignment. Our results show that the tree proportion is a very useful measure of treelikeness: under a huge range of simulation conditions tree proportion consistently declines in concert with declines in the treelikeness of the alignment. When used as a statistical test to ask whether an alignment can reject treelikeness, the six test statistics we compare have varied success at detecting multiple causes of decreased treelikeness, and no one test performed the best across all simulation conditions. Tree proportion, 3SEQ (Lam *et al.* 2018) and PHI (Bruen *et al.* 2006) performed well as statistical tests rejecting treelikeness under simulated introgression events, and we found that at least one of these tests detected almost every simulated recombinant alignment in the majority of cases. It is perhaps unsurprising that PHI and 3SEQ performed well: both tests are designed to detect recombinant sequences of exactly the type we simulated. However, introgression is just one example of a biological event that reduces the treelikeness of an alignment, and our simulations show that both PHI and 3SEQ performed poorly in certain simulation conditions. As a result, we suggest that statistical tests of treelikeness for empirical alignments would be best served by combining a phylogenetic approach such as tree proportion with a mosaic-based test such as 3SEQ. The tree proportion test works in a wide variety of

conditions, and produces an easy-to-interpret test statistic. However, in many conditions (such as if the non-treelikeness is caused by introgression and the tree depth is above a certain threshold) 3SEQ has much more power to detect non-treelikeness. Using both tests therefore provides the most generality and power across all possible causes of non-treelikeness that may impact phylogenetic analyses.

Tree proportion joins a growing group of tests for absolute model adequacy. Penny *et al.* (1992) wrote that a fundamental criterion for a scientific method is that the data must be able to reject the model, a requirement that is rarely met in phylogenetics (Brown and Thomson 2018). Most phylogenetic analyses proceed with only relative tests of model adequacy (such as ModelFinder (Kalyaanamoorthy *et al.* 2017) which selects the best model from a pre-defined set of a models) or no test (Cui *et al.* 2013; de Souza *et al.* 2018; Grismer *et al.* 2018; Grybchuk *et al.* 2018; Kang *et al.* 2014; Lei and Dong 2016; Pearce *et al.* 2017; Tay *et al.* 2017). As model violation is widespread across phylogenetic datasets (Naser-Khdour *et al.* 2019), phylogenetic analyses may benefit if absolute tests for model adequacy are performed prior to tree estimation. Tree proportion builds on the absolute test for model accuracy developed by Goldman (1993). While Goldman's test encompasses all assumptions of the tree and model which are used to calculate the likelihood in phylogenetic analyses, the tree proportion assesses a subset of model assumptions, and asks specifically to what extent a single bifurcating tree is adequate for explaining the evolutionary history of a given alignment.

The PHI test has been widely used to detect recombination in phylogenetic alignments (Cabanne *et al.* 2008; Croll and Sanders 2009; Croucher *et al.* 2015; D'Horta *et al.* 2011; Faria *et al.* 2016; Harris *et al.* 2012; Joly and Bruneau 2006; Ogura *et al.* 2009; Pinho *et al.* 2008; Tian *et al.* 2012; Weinert *et al.* 2009). The PHI test performed well in our simulations and detected almost all introgression events under all simulation conditions. Similarly, Bruen *et al.* (2006) and Haubold *et al.* (2013) found the PHI test accurately detected recombination in simulated coalescent data and in empirical datasets including bacteria, fungi and virus DNA, and animal mtDNA. However, the PHI test performs poorly when sequence diversity is low (less than 10%), when alignments are short and/or when the number of taxa is low (less than 10) (Bruen *et al.* 2006; White *et al.* 2013; White and

Gemmell 2009). In previous simulations where the power of the PHI test was low, the sequence diversity ranged from 0.01 to 1.25 × 10^{-3} (Bruen *et al.* 2006; White *et al.* 2013; White and Gemmell 2009). The diversity in our simulated alignments was much higher than this, and therefore our simulations contained sufficient informative sites and incompatibilities for the test to perform well. The PHI test is a powerful and accurate test for recombination when the sequence diversity and number of taxa are sufficiently large, and is conservative and results in false negatives when they are not.

Likelihood mapping and δ plots allow visualisation of the phylogenetic content of a sequence alignment, but we show here that they do not perform well as test statistics for treelikeness. The mean δ value has been used to quantify the treelikeness of an alignment, with values from 0 to 0.2 generally interpreted as highly treelike (Coiro and Barone Lumaga 2018; Dashper *et al.* 2017; Grimm and Renner 2013; Kozak *et al.* 2015; Meier-Kolthoff and Göker 2019; Short *et al.* 2014; Stanborough *et al.* 2018). However, the mean δ responds inconsistently to causes of decreased treelikeness, so a low mean δ value does not necessarily indicate that an alignment is treelike (e.g., the mean δ value was 0.048 for one of our simulations in which the alignment contained 8 introgression events). Similarly, the proportion of resolved quartets from likelihood mapping has been interpreted as an indicator of the treelikeness of an alignment, with values from 0.8 and up interpreted as treelike (Buesa *et al.* 2002; Elena *et al.* 2001; Li *et al.* 2017; Morgan *et al.* 2014; Nadan *et al.* 2003; Pitra *et al.* 2002; Salemi *et al.* 2000; Shi *et al.* 2012; Verbruggen and Theriot 2008). Likelihood mapping displays the phylogenetic content of an alignment by plotting the treelikeness of individual quartets. However, our results show that this correlates very poorly with the overall treelikeness of an alignment (e.g., the proportion of resolved quartets was 1.0 for some of our simulations in which the alignment contained 8 introgression events). The proportion of resolved quartets for our simulations is high – the minimum value was 70% (Supplementary Figure 2), but the majority of simulations had a proportion of resolved quartets above 85%, despite most of them containing significant non-treelikeness such as multiple introgression events involving a large fraction of the alignment. The reason that the proportion of resolved quartets remains high in these simulations is that each introgression event involves only a few taxa, meaning only a small

proportion of quartets are affected and the contribution to the proportion of resolved quartets is low. As a result of these limitations, we recommend using these methods for visual assessment of phylogenetic information, but not for quantifying or testing for treelikeness.

While the parametric bootstrap approach we propose here is designed to isolate non-treelikeness from other signals in the data, this separation will not always be perfect. A statistically significant result from the parametric bootstrap indicates a significant difference in test statistic values between the original alignment and the bootstrap replicates. In our simulations, the only difference between the original alignment and the bootstrap replicates was that the former included introgression events. The same will not be true for empirical alignments, as the models we use for empirical data are gross oversimplifications of the true underlying process (e.g. (Song *et al.* 2010) (Lemmon and Moriarty 2004)). As a result, the parametric bootstrap may return a significant result if other types of model violation lead to differences in the treelikeness between the empirical and the bootstrap-replicate alignment. Given that model violation is widespread and common within phylogenetic datasets (Naser-Khdour *et al.* 2019), we suggest that a significant result from the parametric bootstrap should be interpreted as likely, but not certain, to be caused by non-treelikeness in the empirical alignment. Regardless of the cause, a significant result from the parametric bootstrap should be cause for concern, and perhaps warrant further investigation of the offending alignment.

Materials and Methods

Simulation approach

Introgression is a potential source of non-treelikeness that is known to mislead phylogenetic inferences (Posada and Crandall 2002; Wiens 1998). However, it is hard to account for with current methods. Introgression provides a framework wherein different types and amounts of non-treelikeness can be simulated on a linear scale (Posada and Crandall 2002). In this study, we use introgression as a framework for simulating non-treelike alignments in order to compare new and existing measures and tests for treelikeness. Here, we compare tree proportion with two tests for introgression: the Pairwise Homoplasy Index (PHI) (Bruen *et*

354 *al.* 2006) and 3SEQ (Boni *et al.* 2007; Lam *et al.* 2018). We also applied two existing methods
355 that have been used to test for treelikeness in previous studies: δ plotting (Holland *et al.*
356 2002) and likelihood mapping (Strimmer and von Haeseler 1997).
357
358 PHI measures the minimum number of convergent mutations on any tree to describe the
359 genealogy of a pair of sites. PHI is a widely used test for recombination or introgression and
360 was previously found to outperform other similar tests (Bruen *et al.* 2006). We calculated
361 the PHI value and p-value for the each alignment using PhiPack (Bruen 2005).
362
363 The second test, 3SEQ, attempts to calculate the number and location of introgression
364 events for a given alignment by testing each triplet of sequences using a hypergeometric
365 random walk to determine if one sequence is the child of the other two (Boni *et al.* 2007;
366 Lam *et al.* 2018). 3SEQ has been shown to perform well in simulations (Boni *et al.* 2007; Lam
367 *et al.* 2018). We used the 3SEQ implementation (Lam *et al.* 2018) to calculate the number of
368 recombinant triplets, number of recombinant sequences and p-value for each alignment.
369
370 We also applied both the δ plotting (Holland *et al.* 2002) and likelihood mapping (Strimmer
371 and von Haeseler 1997) methods, which have been previously used to estimate treelikeness
372 of an alignment. To obtain a test statistic for the δ plotting method, we applied the
373 delta.plot function in the R package ape v5.4-1 (Paradis *et al.* 2004) to the distance matrix
374 for each alignment and calculated the mean and mode δ_q value. The mean or mode δ_q will
375 be between 0 and 1, where larger values are less treelike. We used the likelihood mapping
376 implementation in IQ-Tree (Minh *et al.* 2020) with the number of quartets to sample set to
377 25 times the number of taxa, and took the test statistic to be the number of fully-resolved
378 quartets. This test statistic is a proportion, with a value of 1 indicating that every quartet
379 sampled was treelike. The test statistic value decreases as the quartets become less treelike.
380 The p-values for both the δ plotting and likelihood mapping methods were calculated using
381 a parametric bootstrap.
382
383 ## Simulating multiple sequence alignments with introgression

384 To simulate alignments with introgression, we extended the two-tree simulation approach
385 described in Posada and Crandall (2002) and shown in Figure 6. This method uses forward
386 time phylogenetic simulations to simulate alignments which mimic those that would be
387 produced by introgression events and allows for control over the placement and timing and
388 of introgression events. In principle, one alignment can be simulated along a tree and then a
389 portion of DNA from one species replaced by a portion of DNA from a second species. Here,
390 we achieve the same results by simulating DNA along two trees and then concatenating the
391 two sequences to mimic the result of introgression. This approach provides a simple and
392 flexible framework for simulating introgression on multiple sequence alignments (Posada
393 and Crandall 2002).
394
395 We used this framework to simulate datasets under two scenarios of varying treelikeness:
396 increasing proportion of introgressed DNA and increasing number of introgression events.
397 Firstly, we investigated the effect of increasing the proportion of introgressed DNA (i.e.
398 proportion of final tree) by simulating sequence alignments on a balanced 8-taxon tree from
399 0 – 50% introgressed DNA sequence in 1% intervals. 10 replicates were conducted for each
400 set of simulation parameters. We applied the following test statistics to each alignment:
401 tree proportion, δ plots, likelihood mapping, PHI test and 3SEQ. Due to the high
402 computational expense of the parametric bootstrap, our tree proportion test was calculated
403 only for the proportion of introgressed DNA from 0% – 50% in 10% intervals (6 intervals
404 total). Secondly, we investigated the effect of increasing the number of introgression events
405 by simulating 0 to 8 introgression events on a 32-taxon balanced tree. A 32-taxon balanced
406 tree has 8 balanced subtrees, each consisting of two clades with two species each. Each
407 introgression event takes place within one subtree, allowing from 0 to 8 simultaneous
408 events. For this set of simulations, we fixed the proportion of introgressed DNA at 50%. We
409 performed 100 replicates of each set of simulation parameters. Five test statistics were
410 applied to each alignment as above. The tree proportion test was only performed for the
411 first ten replicates of each set of simulation parameters due to the high computational load.
412
413 Other simulation parameters were as follows. All simulations were repeated for reciprocal
414 and non-reciprocal introgression events. A non-reciprocal event, in which DNA is
415 introgressed unidirectionally from one lineage into another, is shown in Figure 6. In a

reciprocal event, there is a bidirectional exchange of genetic material between two species. For all simulations, we fixed the sequences length to 1300 base pairs (the average length of a transcript in eukaryotes, from Xu *et al.* (2006)), and the model of substitution to the Jukes-Cantor model (Jukes and Cantor 1969). We simulated four substitution rates (in substitutions per site), 0.05, 0.1, 0.5 and 1, to simulate varying rates of molecular evolution across simulations. The total number of simulated alignments was 4080 for the first set of simulations and 6800 for the second set.

The scripts for this analysis were written in R v3.6.3 (R Core Team 2020) using the packages ape v5.4-1 (Paradis *et al.* 2004), ggplot2 v3.3.2 (Wickham 2016), phangorn v2.5.5 (Schliep 2011), phytools v0.7-20 (Revell 2012) ,seqinr v3.6.1 (Charif and Lobry 2007) , stringr v1.4.0 (Wickham 2019) and TreeSim v2.4 (Stadler 2017). Code to replicate all simulations is available from https://github.com/caitlinch/treelikeness. Results from the simulations containing test statistic and statistical test results for all six tests are available in the article and in its online supplementary material.

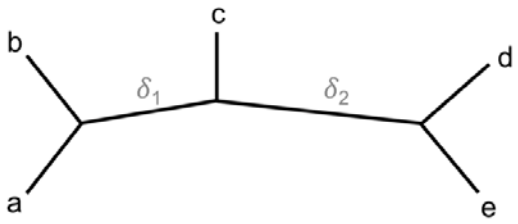
Acknowledgements

The authors would like to thank Barbara Holland, Lindell Bromham, David Gordon and Rod Peakall for their comments and advice. This work was supported by the Australian Research Council grant no. DP-200103151 to R.L. and B.Q.M.

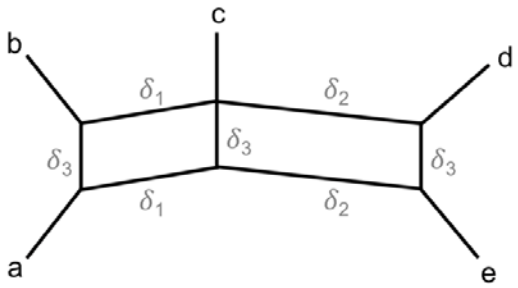
436 **Figures**

437 Figure 1

(a)



(b)



(c)

Split	Bipartition	Weight (λ)
δ_1	ab | cde	0.4
δ_2	abc | de	0.6
δ_3	ae | bcd	0.2

(d)

$$\textit{Tree proportion } (TP) = \frac{\lambda(\delta_1) + \lambda(\delta_2)}{\lambda(\delta_1) + \lambda(\delta_2) + \lambda(\delta_3)}$$

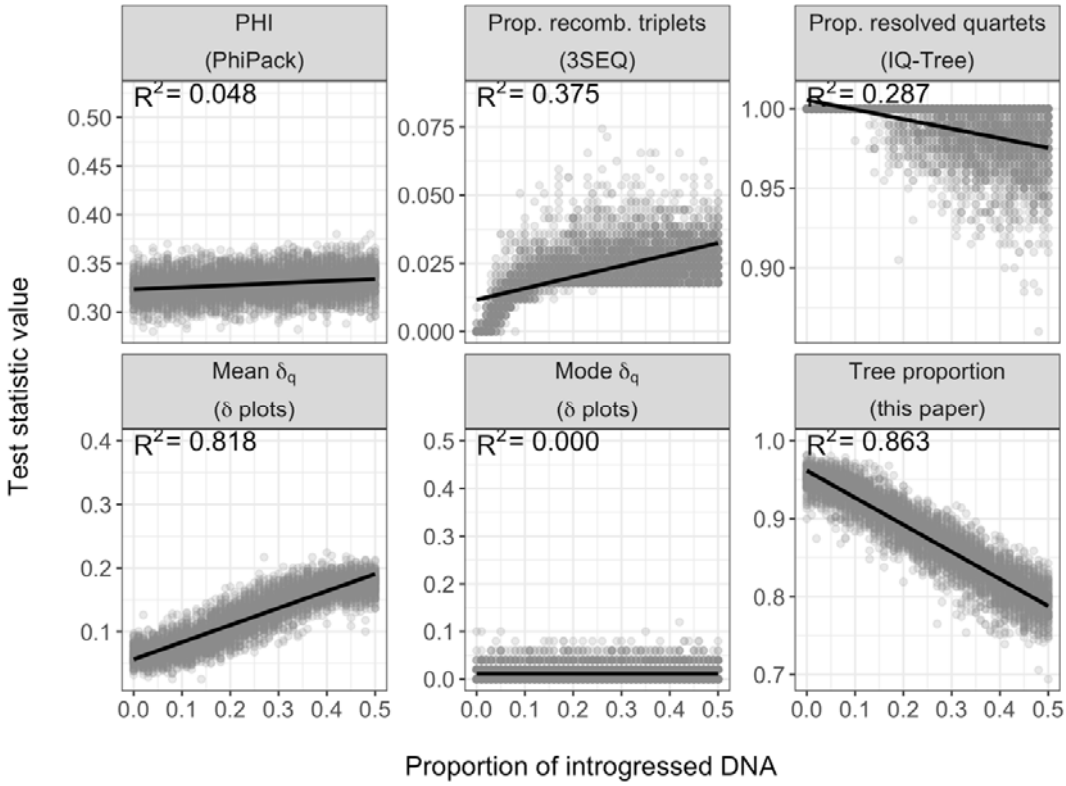
$$TP = \frac{0.4 + 0.6}{0.4 + 0.6 + 0.2}$$

$$TP = 0.83$$

438

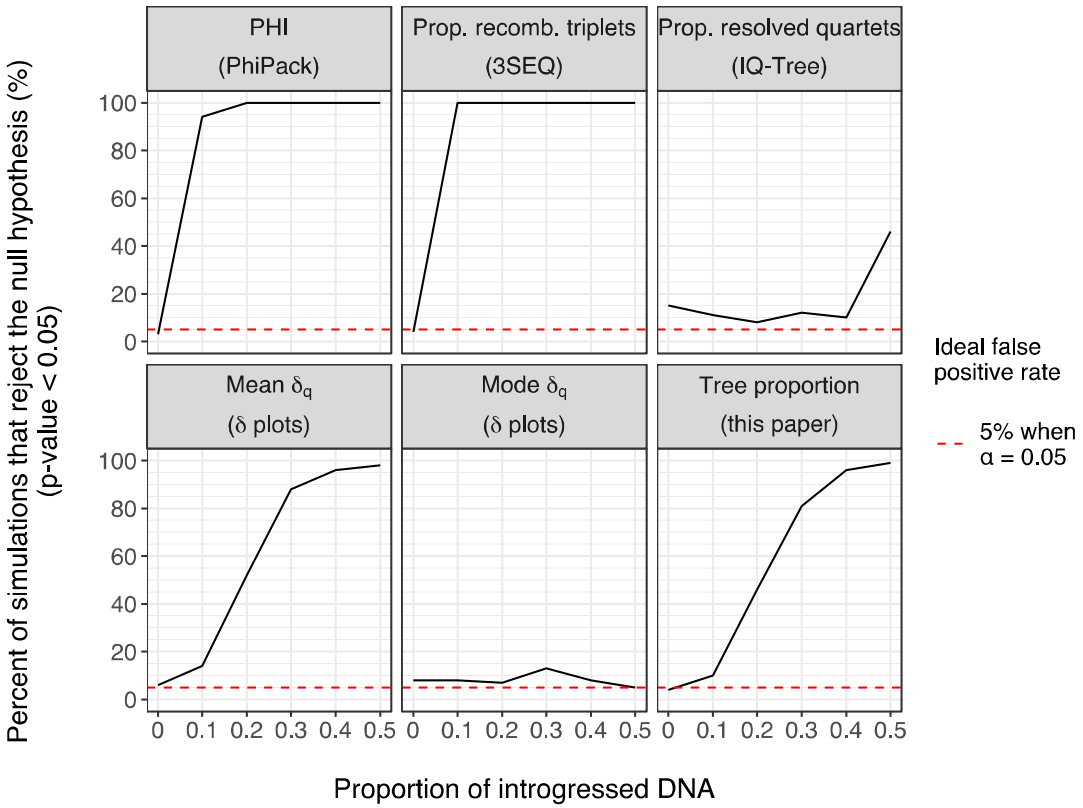
439

440 Figure 2



441
442

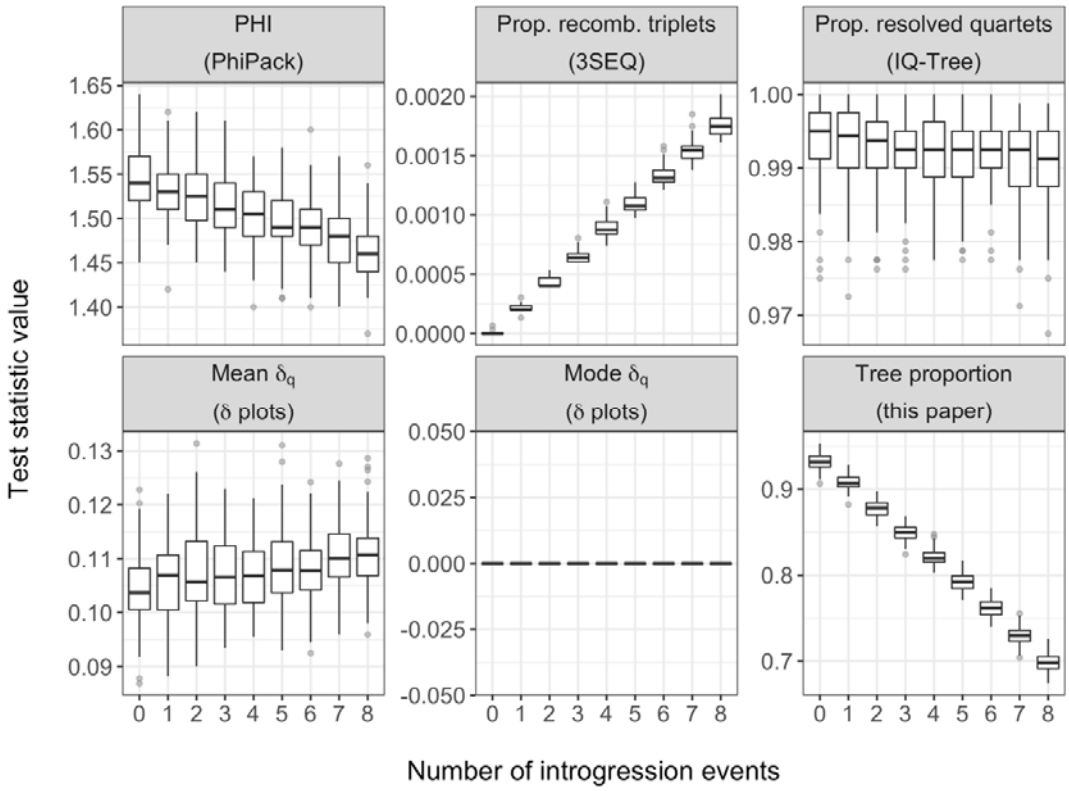
443 Figure 3



444

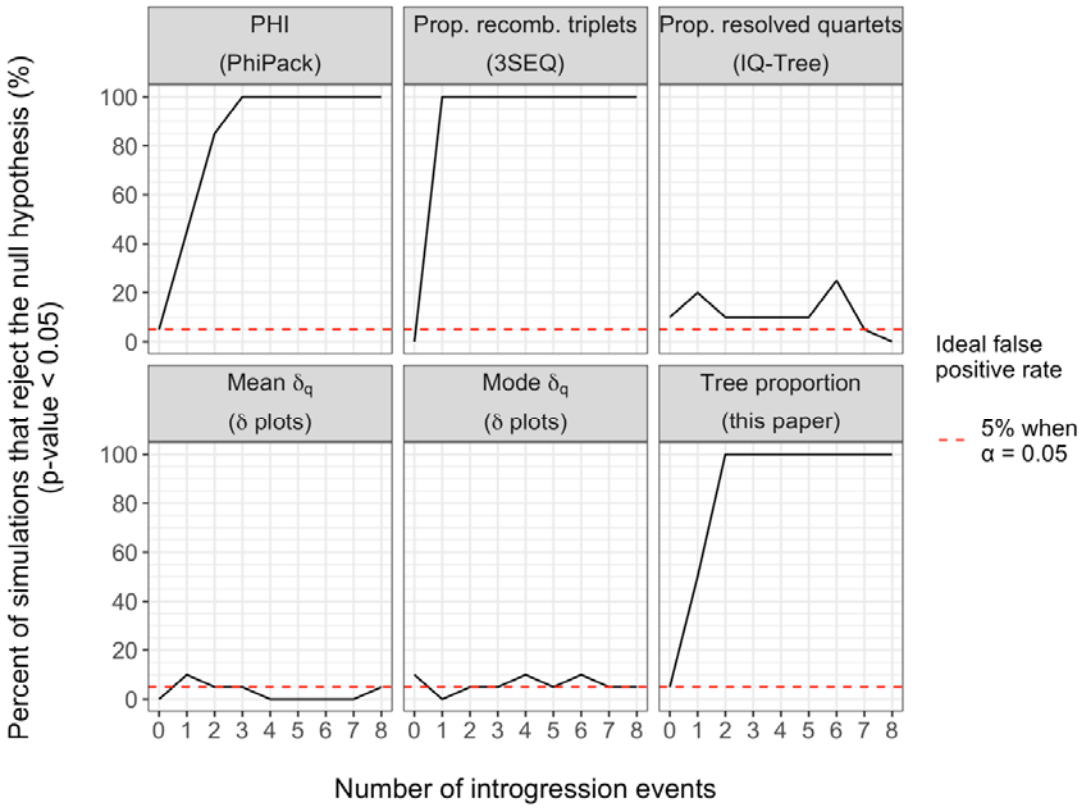
445

446 Figure 4



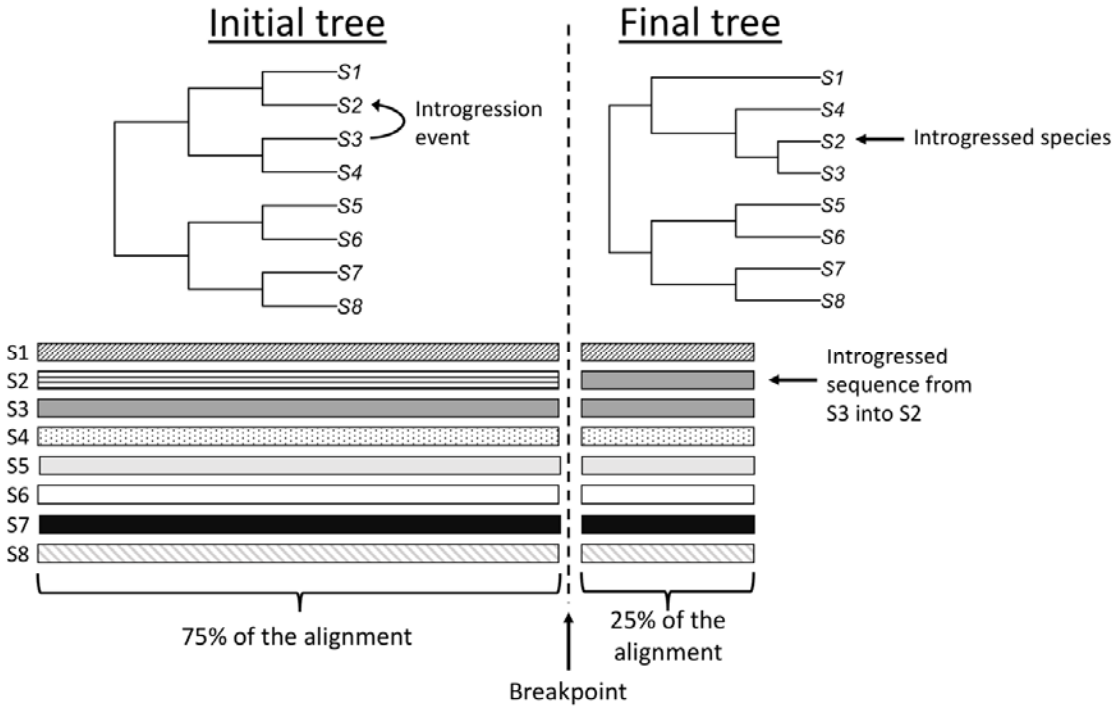
447
448

449 Figure 5



450
451

452 Figure 6



453

454

455 ## Figure Legends

456 **Figure 1. a.** a simple phylogenetic tree for five taxa. Non-trivial splits are labelled. **b.** A split
457 phylogenetic network for the same five taxa. Non-trivial splits are labelled. **c.** Table showing
458 the bipartition of taxa and weight for each non-trivial split. **d.** Sample calculation of the tree
459 proportion for the alignment with tree **a** and network **b**.

460 **Figure 2:** Test statistic values for increasing proportions of introgressed DNA from 0 to 0.5 in
461 0.01 increments with one close, non-reciprocal introgression events and tree depth of 0.5 (n
462 = 100). Each point represents the test statistic value for a single simulated alignment. Prop.
463 recomb. triplets is the proportion of recombinant triples. Prop. resolved quartets is the
464 proportion of resolved quartets.

465 **Figure 3:** Percentage of simulated alignments that reject the null hypothesis as the
466 proportion of introgressed DNA increases for all six test statistics. Simulated alignments had
467 a single close, non-reciprocal introgression event and tree depth was 0.5. Each line
468 represents the number of alignments (out of 100 replicates) that reject the null hypothesis

of treelikeness. Prop. recomb. triplets is the proportion of recombinant triples. Prop. resolved quartets is the proportion of resolved quartets.

Figure 4: Test statistic values for number of non-reciprocal introgression events and tree depth of 0.5 (n = 100). Each box shows the distribution of test statistic values at a certain number of introgression events. The whiskers extend to the closest observed value no more than 1.5 times the interquartile range away from the box. Points represent outliers. Prop. recomb. triplets is the proportion of recombinant triples. Prop. resolved quartets is the proportion of resolved quartets.

Figure 5: Percentage of simulated alignments that reject the null hypothesis as the number of non-reciprocal introgression events increases for all six test statistics. Simulated alignments had a tree depth of 0.5. Each line represents the number of alignments (out of 100 replicates) that reject the null hypothesis of treelikeness. The red dotted line represents the ideal false positive rate of 5% when $\alpha = 0.05$. Prop. recomb. triplets is the proportion of recombinant triples. Prop. resolved quartets is the proportion of resolved quartets.

Figure 6: Simulation of a mosaic multiple sequence alignment with an introgression event, adapted from Posada and Crandall (2002). The initial sequence (n% of the final sequence, here n = 75%) is simulated along a balanced 8-taxon tree. The introgression event shown here consists of genetic material from S3 overwriting the original sequence in S2 (as shown by the arrows). In the final tree, the introgression event has occurred, moving the position of S2 in the phylogeny. 25% ((1 - n) %) of the alignment is then simulated along this tree. Two trees are needed to explain the evolutionary history of this alignment, violating the treelikeness assumption.

References

Baric S, Salzburger W, Sturmbauer C. 2003. Phylogeography and evolution of the Tanganyikan Cichlid genus *Tropheus* based upon mitochondrial DNA Sequences. *J. Mol. Evol.* 56(1):54-68.

Boni MF, Posada D, Feldman MW. 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 176(2):1035-1047.

Brown JM, Thomson RC. 2018. Evaluating model performance in evolutionary biology. *Annu. Rev. Ecol. Evol. Syst.* 49(1):95-114.

Bruen T. 2005. PhiPack. Online, available at https://www.maths.otago.ac.nz/~dbryant/software.html.

501 Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the
502 presence of recombination. *Genetics* 172(4):2665.
503 Bryant D, Moulton V. 2004. Neighbor-Net: an agglomerative method for the construction of
504 phylogenetic networks. *Mol. Bio. Evol.* 21(2):255-265.
505 Buesa J, Collado B, López-Andújar P, Abu-Mallouh R, Rodríguez Díaz J, García Díaz A, Prat J,
506 Guix S, Llovet T, Prats G et al. . 2002. Molecular epidemiology of caliciviruses causing
507 outbreaks and sporadic cases of acute gastroenteritis in Spain. *J. Clin. Microbiol.*
508 40(8):2854-2859.
509 Cabanne GS, d'Horta FM, Sari EHR, Santos FR, Miyaki CY. 2008. Nuclear and mitochondrial
510 phylogeography of the Atlantic forest endemic *Xiphorhynchusfuscus* (Aves:
511 Dendrocolaptidae): biogeography and systematics implications. *Mol. Phylogenet.*
512 *Evol.* 49(3):760-773.
513 Charif D, Lobry JR. 2007. SeqinR 1.0-2: a contributed package to the R project for statistical
514 computing devoted to biological sequences retrieval and analysis. In: Bastolla U,
515 Porto M, Roman HE, Vendruscolo M, editors. Structural approaches to sequence
516 evolution: Molecules, networks, populations. New York: Springer Verlag. p. 207-232.
517 Coiro M, Barone Lumaga MR. 2018. Disentangling historical signal and pollinator selection
518 on the micromorphology of flowers: an example from the floral epidermis of the
519 Nymphaeaceae. *Plant Biol.* 20(5):902-915.
520 Croll D, Sanders IR. 2009. Recombination in *Glomus intraradices*, a supposed ancient asexual
521 arbuscular mycorrhizal fungus. *BMC Evol. Biol.* 9(1):13.
522 Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR.
523 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole
524 genome sequences using Gubbins. *Nucleic Acids Res.* 43(3):e15.
525 Cui R, Schumer M, Kruesi K, Walter R, Andolfatto P, Rosenthal GG. 2013. Phylogenomics
526 reveals extensive reticulate evolution in *Xiphophorus* fishes. *Evolution* 67(8):2166-
527 2179.
528 D'Horta FM, Cabanne GS, Meyer D, Miyaki CY. 2011. The genetic effects of Late Quaternary
529 climatic changes over a tropical latitudinal gradient: diversification of an Atlantic
530 Forest passerine. *Mol. Ecol.* 20(9):1923-1935.
531 Dashper SG, Mitchell HL, Seers CA, Gladman SL, Seemann T, Bulach DM, Chandry PS, Cross
532 KJ, Cleal SM, Reynolds EC. 2017. *Porphyromonas gingivalis* uses specific domain
533 rearrangements and allelic exchange to generate diversity in surface virulence
534 factors. *Front. Microbiol.* 8:48.
535 de Souza WM, Fumagalli MJ, de Araujo J, Sabino-Santos G, Maia FGM, Romeiro MF, Modha
536 S, Nardi MS, Queiroz LH, Durigon EL et al. . 2018. Discovery of novel anelloviruses in
537 small mammals expands the host range and diversity of the *Anelloviridae*. *Virology*
538 514:9-17.
539 Dress AWM. 1984. Trees, tight extensions of metric spaces, and the cohomological
540 dimension of certain groups: a note on combinatorial properties of metric spaces
541 *Adv. Math.* 53:321-402.
542 Eigen M, Winkler-Oswatitsch R, Dress A. 1988. Statistical geometry in sequence space: a
543 method of quantitative comparative sequence analysis. *Proc. Natl. Acad. Sci. U.S.A.*
544 85(16):5913.
545 Elena SF, Dopazo J, de la Peña M, Flores R, Diener TO, Moya A. 2001. Phylogenetic analysis
546 of viroid and viroid-like satellite RNAs from plants: a reassessment. *J. Mol. Evol.*
547 53(2):155-159.

548 Faria NR, Azevedo RdSdS, Kraemer MUG, Souza R, Cunha MS, Hill SC, Thézé J, Bonsall MB,
549 Bowden TA, Rissanen I et al. . 2016. Zika virus in the Americas: early epidemiological
550 and genetic findings. *Science* 352(6283):345-349.
551 Goldman N. 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* 36(2):182-
552 198.
553 Grimm GW, Renner SS. 2013. Harvesting Betulaceae sequences from GenBank to generate a
554 new chronogram for the family. *Bot. J. Linn. Soc.* 172(4):465-477.
555 Grismer LL, Wood JRPL, Quah ESH, Anuar S, Ngadi EB, Izam NAM, Ahmad N. 2018.
556 Systematics, ecomorphology, cryptic speciation and biogeography of the lizard genus
557 *Tytthoscincus* Linkem, Diesmos & Brown (Squamata: Scincidae) from the sky-island
558 archipelago of Peninsular Malaysia. *Zool. J. Linn. Soc.* 183(3):635-671.
559 Grybchuk D, Akopyants NS, Kostygov AY, Konovalovas A, Lye L-F, Dobson DE, Zangger H,
560 Fasel N, Butenko A, Frolov AO et al. . 2018. Viral discovery and diversity in
561 trypanosomatid protozoa with a focus on relatives of the human parasite
562 *Leishmania*. *Proc. Natl. Acad. Sci. U.S.A.* 115(3):E506.
563 Harris SR, Clarke IN, Seth-Smith HMB, Solomon AW, Cutcliffe LT, Marsh P, Skilton RJ, Holland
564 MJ, Mabey D, Peeling RW et al. . 2012. Whole-genome analysis of diverse *Chlamydia*
565 *trachomatis* strains identifies phylogenetic relationships masked by current clinical
566 typing. *Nat. Genet.* 44:413.
567 Haubold B, Krause L, Horn T, Pfaffelhuber P. 2013. An alignment-free test for recombination.
568 *Bioinformatics* 29(24):3121-3127.
569 Holland BR, Huber KT, Dress A, Moulton V. 2002. δ plots: a tool for analyzing phylogenetic
570 distance data. *Mol. Biol. Evol.* 19(12):2051-2059.
571 Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies.
572 *Mol. Bio. Evol.* 23(2):254-267. Software available from www.splitstree.org.
573 Huson DH, Rupp R, Scornavacca C. 2010. Phylogenetic networks. Cambridge: Cambridge
574 University Press.
575 Joly S, Bruneau A. 2006. Incorporating allelic variation for reconstructing the evolutionary
576 history of organisms from multiple genes: an example from *Rosa* in North America.
577 *Syst. Biol.* 55(4):623-636.
578 Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, editor.
579 Mammalian Protein Metabolism. New York: Academic Press. p. 21-132.
580 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder:
581 fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14(6):587-
582 589.
583 Kang YJ, Kim SK, Kim MY, Lestari P, Kim KH, Ha B-K, Jun TH, Hwang WJ, Lee T, Lee J et al. .
584 2014. Genome sequence of mungbean and insights into evolution within *Vigna*
585 species. *Nat. Commun.* 5:5443.
586 Kennedy M, Holland BR, Gray RD, Spencer HG. 2005. Untangling long branches: identifying
587 conflicting phylogenetic signals using spectral analysis, Neighbor-Net, and consensus
588 networks. *Syst. Biol.* 54(4):620-633.
589 Kozak KM, Wahlberg N, Neild AFE, Dasmahapatra KK, Mallet J, Jiggins CD. 2015. Multilocus
590 species trees show the recent adaptive radiation of the mimetic *Heliconius*
591 butterflies. *Syst. Biol.* 64(3):505-524.
592 Kück P, Meusemann K, Dambach J, Thormann B, von Reumont BM, Wägele JW, Misof B.
593 2010. Parametric and non-parametric masking of randomness in sequence
594 alignments can be improved and leads to better resolved trees. *Front. Zool.* 7(1):10.

595 Lam HM, Ratmann O, Boni MF. 2018. Improved algorithmic complexity for the 3SEQ
596 recombination detection algorithm. *Mol. Biol. Evol.* 35(1):247-251.
597 Lei M, Dong D. 2016. Phylogenomic analyses of bat subordinal relationships based on
598 transcriptome data. *Sci. Rep.* 6:27726.
599 Lemmon AR, Moriarty EC. 2004. The importance of proper model assumption in Bayesian
600 phylogenetics. *Syst. Biol.* 53(2):265-277.
601 Li X, Liu H, Liu L, Feng Y, Kalish ML, Ho SYW, Shao Y. 2017. Tracing the epidemic history of
602 HIV-1 CRF01_AE clusters using near-complete genome sequences. *Sci. Rep.*
603 7(1):4024.
604 Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species? *Bioessays* 38(2):140-
605 149.
606 Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for
607 state-of-the-art genome-based taxonomy. *Nat. Commun.* 10(1):2182.
608 Mendes F, K., Livera A, P., Hahn M, W. 2019. The perils of intralocus recombination for
609 inferences of molecular convergence. *Philos. Trans. R. Soc. B.* 374(1777):20180244.
610 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear
611 R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in
612 the genomic era. *Mol. Biol. Evol.* 37(5):1530-1534.
613 Morgan CC, Creevey CJ, O'Connell MJ. 2014. Mitochondrial data are not suitable for
614 resolving placental mammal phylogeny. *Mamm. Genome.* 25(11):636-647.
615 Nadan S, Walter JE, Grabow WOK, Mitchell DK, Taylor MB. 2003. Molecular characterization
616 of astroviruses by reverse transcriptase PCR and sequence analysis: comparison of
617 clinical and environmental isolates from South Africa. *Appl. Environ. Microb.*
618 69(2):747.
619 Naser-Khdour S, Minh BQ, Zhang W, Stone EA, Lanfear R. 2019. The prevalence and impact
620 of model violations in phylogenetic analysis. *Genome Biol. Evol.* 11(12):3341-3352.
621 Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K,
622 Kurokawa K et al. . 2009. Comparative genomics reveal the mechanism of the
623 parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli. Proc.*
624 *Natl. Acad. Sci. U.S.A.* 106(42):17939.
625 Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R
626 language. *Bioinformatics* 20(2):289-290.
627 Pearce SL, Clarke DF, East PD, Elfekih S, Gordon KHJ, Jermiin LS, McGaughran A, Oakeshott
628 JG, Papanikolaou A, Perera OP et al. . 2017. Genomic innovations, transcriptional
629 plasticity and gene loss underlying the evolution and divergence of two highly
630 polyphagous and invasive *Helicoverpa* pest species. *BMC Biol.* 15(1):63.
631 Penny D, Hendy MD, Steel MA. 1992. Progress with methods for constructing evolutionary
632 trees. *Trends Ecol. Evol.* 7(3):73-79.
633 Pinho C, Harris DJ, Ferrand N. 2008. Non-equilibrium estimates of gene flow inferred from
634 nuclear genealogies suggest that Iberian and North African wall lizards (*Podarcis*
635 spp.) are an assemblage of incipient species. *BMC Evol. Biol.* 8(1):63.
636 Pitra C, Lieckfeldt D, Frahnert S, Fickel J. 2002. Phylogenetic relationships and ancestral
637 areas of the bustards (Gruiformes: Otididae), inferred from mitochondrial DNA and
638 nuclear intron sequences. *Mol. Phylogenet. Evol.* 23(1):63-74.
639 Posada D, Crandall KA. 2002. The effect of recombination on the accuracy of phylogeny
640 estimation. *J. Mol. Evol.* 54(3):396-402.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Revell LJ. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3:217-223.

Salemi M, Desmyter J, Vandamme AM. 2000. Tempo and mode of human and simian T-lymphotropic virus (HTLV/STLV) evolution revealed by analyses of full-genome sequences. *Mol. Biol. Evol.* 17(3):374-386.

Salzburger W, Martens J, Sturmbauer C. 2002. Paraphyly of the Blue Tit (*Parus caeruleus*) suggested from cytochrome b sequences. *Mol. Phylogenet. Evol.* 24(1):19-25.

Schliep KP. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27(4):592-592.

Scornavacca C, Galtier N. 2017. Incomplete lineage sorting in mammalian phylogenomics. *Syst. Biol.* 66(1):112-120.

Shi CM, Yang Z. 2018. Coalescent-based analyses of genomic sequence data provide a robust resolution of phylogenetic relationships among major groups of gibbons. *Mol. Biol. Evol.* 35(1):159-179.

Shi W, Freitas IT, Zhu C, Zheng W, Hall WW, Higgins DG. 2012. Recombination in hepatitis C virus: identification of four novel naturally occurring inter-subtype recombinants. *PloS one* 7(7):e41997-e41997.

Short DPG, O'Donnell K, Geiser DM. 2014. Clonality, recombination, and hybridization in the plumbing-inhabiting human pathogen *Fusarium keratoplasticum* inferred from multilocus sequence typing. *BMC Evolutionary Biology* 14(1):91.

Song H, Sheffield NC, Cameron SL, Miller KB, Whiting MF. 2010. When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. *Syst. Entomol.* 35(3):429-448.

Stadler T. 2017. TreeSim: simulating phylogenetic trees. R package version 2.4. http://CRAN.R-project.org/package=TreeSim

Stanborough T, Fegan N, Powell SM, Singh T, Tamplin M, Chandry PS. 2018. Genomic and metabolic characterization of spoilage-associated *Pseudomonas* species. *Int. J. Food Microbiol.* 268:61-72.

Steiner G, Dreyer H. 2003. Molecular phylogeny of Scaphopoda (Mollusca) inferred from 18S rDNA sequences: support for a Scaphopoda–Cephalopoda clade. *Zool. Scr.* 32(4):343-356.

Strimmer K, von Haeseler A. 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. U.S.A.* 94(13):6815-6819.

Tay WT, Walsh TK, Downes S, Anderson C, Jermiin LS, Wong TKF, Piper MC, Chang ES, Macedo IB, Czepak C et al. . 2017. Mitochondrial DNA and trade data support multiple origins of *Helicoverpa armigera* (Lepidoptera, Noctuidae) in Brazil. *Sci. Rep.* 7:45302.

Tian CF, Zhou YJ, Zhang YM, Li QQ, Zhang YZ, Li DF, Wang S, Wang J, Gilbert LB, Li YR et al. . 2012. Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. *Proc. Natl. Acad. Sci. U.S.A.* 109(22):8629.

Verbruggen H, Theriot EC. 2008. Building trees of algae: some advances in phylogenetic and evolutionary analysis. *Eur. J. Phycol.* 43(3):229-252.

Weinert LA, Werren JH, Aebi A, Stone GN, Jiggins FM. 2009. Evolution and diversity of *Rickettsia* bacteria. *BMC Biol.* 7(1):6.

688 Weisrock DW, Smith SD, Chan LM, Biebouw K, Kappeler PM, Yoder AD. 2012. Concatenation
689 and concordance in the reconstruction of mouse lemur phylogeny: an empirical
690 demonstration of the effect of allele sampling in phylogenetics. *Mol. Biol. Evol.*
691 29(6):1615-1630.
692 White DJ, Bryant D, Gemmell NJ. 2013. How good are indirect tests at detecting
693 recombination in human mtDNA? *G3 (Bethesda)* 3(7):1095-1104.
694 White DJ, Gemmell NJ. 2009. Can indirect tests detect a known recombination event in
695 human mtDNA? *Mol. Biol. Evol.* 26(7):1435-1439.
696 Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.
697 Wickham H. 2019. stringr: Simple, Consistent Wrappers for Common String Operations. R
698 package version 1.4.0. https://CRAN.R-project.org/package=stringr.
699 Wielstra B, Arntzen JW, van der Gaag KJ, Pabijan M, Babik W. 2014. Data concatenation,
700 Bayesian concordance and coalescent-based analyses of the species tree for the
701 rapid radiation of *Triturus* newts. *PLOS ONE* 9(10):e111011.
702 Wiens JJ. 1998. Combining data sets with different phylogenetic histories. *Syst. Biol.*
703 47(4):568-581.
704 Wu M, Kostyun JL, Hahn MW, Moyle LC. 2018. Dissecting the basis of novel trait evolution in
705 a radiation with widespread phylogenetic discordance. *Mol. Ecol.* 27(16):3301-3316.
706 Xu L, Chen H, Hu X, Zhang R, Zhang Z, Luo ZW. 2006. Average gene length is highly
707 conserved in prokaryotes and eukaryotes and diverges only between the two
708 kingdoms. *Mol. Biol. Evol.* 23(6):1107-1108.
709 Zhao L, Li X, Zhang N, Zhang S-D, Yi T-S, Ma H, Guo Z-H, Li D-Z. 2016. Phylogenomic analyses
710 of large-scale nuclear genes provide new insights into the evolutionary relationships
711 within the rosids. *Mol. Phylogenet. Evol.* 105:166-176.
712