The meaning and measure of concordance factors in phylogenomics

Robert Lanfear[1,*] and Matthew W. Hahn[2,3,*]

[1]Ecology & Evolution, Research School of Biology, Australian National University, Canberra Australia; [2]Department of Biology, Indiana University, Bloomington, Indiana, USA; [3]Department of Computer Science, Indiana University, Bloomington, Indiana, USA

*Corresponding authors: E-mail: rob.lanfear@anu.edu.au; mwh@indiana.edu

Abstract

As phylogenomic datasets have grown in size, researchers have developed new ways to measure biological variation and to assess statistical support. Larger datasets have many more sites and many more loci, and therefore less sampling variance. While this means that we can more accurately measure the mean signal in these datasets. the lower sampling variance is often reflected in widely used measures of branch support— such as the bootstrap and posterior probability—being uniformly high, limiting their utility. Larger datasets have also revealed a large amount of biological variation in the topologies found across individual loci, such that the single species tree inferred by most phylogenetic methods represents a limited summary of the data. In contrast to measures of statistical support, the degree of underlying topological variation among sites or loci should be approximately constant regardless of the size of the dataset. “Concordance factors” and similar statistics have therefore become increasingly important tools in phylogenetics. In this review, we explain why concordance factors should be thought of as descriptors of topological variation, rather than as measures of statistical support, and argue that they provide important information not contained in measures of support. We review a growing suite of statistics derived from various ways of measuring concordance, comparing them in a common framework that reveals their interrelationships. We also discuss how measures of topological variation might change in the future as we move beyond estimating a single “tree of life” towards estimating the myriad evolutionary histories underlying genomic variation.

Introduction

As recently as a decade ago, the molecular datasets commonly used in phylogenetics were quite small, consisting of perhaps a handful of loci. Although one of the main goals of phylogenetics was and remains estimating a species tree, the limited available data meant there was little point in trying to examine variation in phylogenetic signal among the loci that constituted a dataset. Instead, most concerns revolved around statistical uncertainty, such that most effort was expended trying to quantify the statistical support for each branch in the species tree (Simon 2022). The most commonly used methods to evaluate statistical support are the bootstrap (Felsenstein 1985) and posterior probability (Rannala and Yang 1996); both were intended to evaluate the reliability of inferred trees in an era when such trees were built by sampling small numbers of informative substitutions from a small number of loci. The support levels output by these methods can provide estimates of statistical confidence on each branch of a specific tree, given an alignment and a model of sequence evolution.

More data hopefully means more accurate inferences, at least when the correct models are used—changing the model of sequence evolution can sometimes drastically change support levels (Stefanović et al 2004; Kumar et al. 2012; Shen et al. 2017). Regardless of their dependence on a particular model of sequence evolution, bootstrap and posterior probability support values are almost universally reported as measures of confidence in branches of phylogenetic trees. However, genome-scale data has greatly reduced the sampling variance in typical phylogenetic datasets. While some support methods have been extended for datasets with huge numbers of sites and taxa (Lemoine et al. 2018; Lutterop et al. 2022; Stamatakis et al. 2008; Minh et al. 2013, Hoang et al. 2018), in general the consequence of larger datasets has been to lower sampling variance. This lower variance means that branch support measures will be almost always uniformly high, limiting their utility (Thomson and Brown 2022).

Although sampling variance has been reduced in so-called “phylogenomic” datasets, there has been increased recognition of biological variance in such data (Maddison 1997). Biological variation arises because individual loci do not have to share the same topology with either the species tree or with each other. We refer to gene trees that match the species tree as concordant, and those that do not as discordant. (Note that a “gene tree” can refer to the topology at any locus, not just in protein-coding genes.) Discordance can be due to both technical and biological causes. Technical error, such as model misspecification, misalignment, or simply a limited amount of information, can result in both systematic and stochastic gene tree inference error. The biological processes that drive discordance include incomplete lineage sorting (ILS), introgression, and horizontal gene transfer (Degnan and Rosenberg 2009). Duplication and loss can also generate gene trees that differ from species trees, though the resulting discordance may be due to a combination of both biological processes and technical error (i.e. the misidentification of orthologous sequences).

Importantly, biological discordance contains information about evolutionary processes, and therefore represents a rich source of data (Bravo et al. 2019). In this review, we discuss the various ways that concordance and discordance can be measured and quantified in

phylogenetics. We show how several popular ways to quantify topological heterogeneity are related via a shared set of simpler measures, and explain how each can be interpreted. While the maximization of concordance can be an optimality criterion for choosing a species tree topology within a particular dataset, we stress that concordance is not a measure of statistical support. Biological concordance and discordance should not change with varying amounts of data from the same set of taxa, as they represent measures of statistical variation, not statistical confidence.

Concordance and discordance

To introduce concordance and discordance, it can help to start with a simple example. Imagine a dataset of three gene trees sampled from four clades (*A*, *B*, *C*, and *D*), one of which, *D*, is the outgroup; let us also assume that we have inferred the topology of the species tree relating these clades (Figure 1a). In this example, the three gene trees have three different topologies—one that is concordant with the species tree (Gene tree 1, Figure 1b) and two that are discordant (Gene trees 2 and 3, Figure 1b). Although gene trees can vary in their branch lengths, we do not consider this source variation in our labelling of concordant and discordant trees: only the hierarchical sets of relationships (sometimes referred to as “bipartitions”or “splits”) are considered. Indeed, gene tree branch lengths are expected to differ from species tree branch lengths even when the trees are concordant (Edwards and Beerli 2000), making branch lengths an impractical measure of concordance.

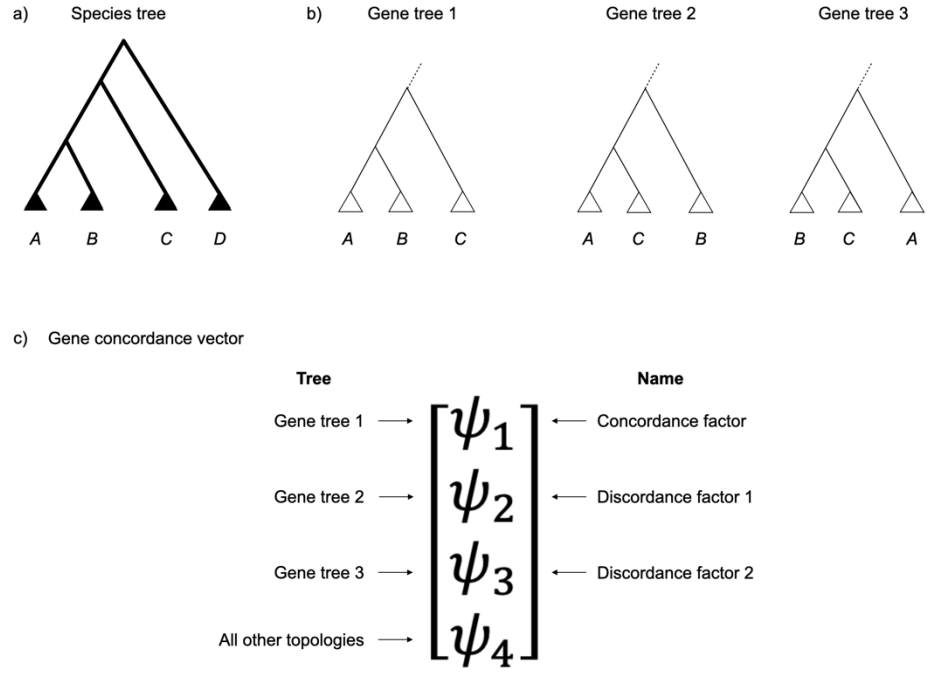

Figure 1. Concordance and discordance. (a) A species tree of four clades of organisms, *A*, *B*, *C*, and *D*. Clade *D* is the outgroup. (b) three possible gene trees derived from the species tree in (a). Gene tree 1 is concordant with the species tree, while gene trees 2 and 3 are discordant. (c) The gene concordance vector, ψ, describes the number of gene trees, or sometimes the proportion of gene trees, that fall into each of four categories. ψ_1 is the number of gene trees that are concordant with the species tree (commonly called the "concordance factor"). ψ_2 and ψ_3 are the number of gene trees that match discordant tree 2 and 3 in (b), ordered with the largest count first. ψ_4 is the number of gene trees that are discordant with the species tree but match neither gene tree 2 or gene tree 3 (for example, any gene tree in which clades *A*, *B*, or *C* are not monophyletic).

For a rooted three-taxon tree, or an unrooted four-taxon tree, there are only three possible topologies, all of which are shown in Figure 1b. As a result, the trees in Figure 1 can only differ in a very limited number of ways, as long as clades A, B, C, and D are always monophyletic. Trees with more taxa have many more possible topologies. For a rooted four-taxon tree there are 15 possible topologies, with the number of topologies growing super-exponentially with the number of taxa (see Table 3.1 in Felsenstein 2004). One implication of this huge number of possible topologies is that, given enough taxa, any given gene tree is likely to be discordant with the species tree on at least one branch. Indeed, in several large datasets (in terms of number of gene trees and number of taxa) there are *no* gene trees that are completely concordant with the species tree (e.g. Jarvis et al. 2014; Pease et al. 2016). Partly because of this, researchers rarely focus on the overall concordance of gene trees with the *entire* species tree. Instead, it is much more informative to focus on a specific internal branch of the species tree, and to ask what proportion of gene trees are concordant with that branch. In other words, most researchers seek to estimate a concordance factor for every branch in the species tree.

The biological (and technical) factors that drive concordance and discordance can vary through time and across the phylogeny, such that the unique combination of processes that underlie the evolutionary history of any clade will produce a particular distribution of gene trees. In order to study these processes, we would therefore like a measure of genealogical heterogeneity that tells us about concordance and discordance on specific branches of the species tree. Concordance factors (CFs) have become a widely used metric to describe this heterogeneity (Baum 2007). (These are also sometimes called gene support frequencies [GSFs]; Gadagkar et al. 2005.) The concordance factor was originally defined as "the proportion of the genome for which a given clade is true" (Baum 2007). Here, the genome is imagined to be made up of many independent loci, each of which has a gene tree topology affected by ILS, introgression, and horizontal gene transfer. This definition also assumes that the species tree is an accurate reflection of species relationships, and that gene trees that match it are "true"; later we relax this assumption, but for now we will accept this phrasing. The concordance factor of a branch or clade in the species tree is therefore a biological parameter that we should be able to estimate by inferring gene trees at many loci. In contrast, discordance factors (DFs) describe the fraction of the genome for which the given clade is *not* true. Because there are multiple ways for a gene tree topology to not match a branch in the species tree, discordant gene trees can be subdivided into several different biologically relevant groups, each represented by its own discordance factor (note that some papers refer to all concordance and discordance measures

collectively as concordance factors; e.g. Allman et al. 2022). Together, concordance and discordance factors provide useful information for understanding evolutionary histories and for testing evolutionary hypotheses; here we summarise them in a single vector called the 'concordance vector' (Figure 1C), which we introduce below.

In the rest of the paper, we review the meaning of concordance factors, how they can be estimated, and how these different estimates can be interpreted. We start by distinguishing concordance factors from measures of statistical support. We then introduce the concordance vector—a set of four concordance and discordance factors for a given branch of the species tree that usefully summarizes topological heterogeneity. Following this, we introduce and compare the different methods that have been developed for estimating concordance and discordance factors from empirical data. We describe how several popular ways to quantify topological heterogeneity are related via this shared set of simple measures, and explain how each can be interpreted. We conclude with suggestions for future directions.

The concordance factor is not a measure of statistical support

A seemingly common misconception is to treat concordance factors as measures of statistical support. This is not correct: concordance factors are (estimates of) biological parameters, not measures of statistical support (Baum 2007). Measures of support such as bootstrap proportions and posterior probabilities are estimates of our confidence that a branch exists, given some assumptions about the data and the models being used (Ané et al. 2007). For consistent statistical methods, these types of support measures will always increase towards their maximum possible value as we add more data to the analysis (cf. Kumar et al. 2012). The same is not true of concordance factors. As we add more data to an analysis, estimates of concordance factors will become more precise, but will not approach any limiting value. This is demonstrated for two empirical datasets in Figure 2—as we add more data to the analyses, measures of statistical support (here, the UltraFast Bootstrap [Hoang et al. 2017] and the ASTRAL posterior probability [Sayyari and Mirarab 2016], both shown in greyscale with square points) tend to increase towards their maximum value, while measures of concordance (shown in colour with circular points) have higher *variance* at lower sample sizes, but quickly stabilise to quite consistent values as sample size increases.

To see what determines concordance factors (and discordance factors), consider the simplest species tree that can have discordance (Figure 1a, imagining a single representative is sampled from clades *A*-*D*). Under a model of incomplete lineage sorting, the length of the internal branch of the species tree, T, determines the degree of concordance and discordance. (Branch lengths here are measured in "coalescent" units, such that $T=t/2N$, where t is the number of generations and N is the effective population size.) If ILS is the only process acting, the probability of sampling a locus with a concordant gene tree is (Hudson 1983):

$$P(concordance) = 1 - {}^{2}/_{3}\, e^{-T} \quad (1)$$

Conversely, the probability of sampling a locus with either one of the two possible discordant gene tree topologies is:

$$P(discordance) = {}^{1}/_{3}\, e^{-T} \quad (2)$$

Both of the two possible discordant topologies have the same probability, and so are always expected to have the same frequency under ILS alone for this species tree.
Here, equation 1 is equivalent to the expected concordance factor—the "proportion of the genome for which a given clade is true"—if we imagine that we have randomly sampled unlinked loci from across the genome. Given these definitions, expected concordance will be highest (approaching 1) with long branch lengths (large *T*), and will be lowest (approaching 1/3) with very short branch lengths (very small *T*). Note that while concordance factors can never be greater than 1, for trees with more than three taxa they can be less than 1/3 because more than two discordant topologies are possible (more on this below).

Most importantly, we can see from this formulation that the expected degree of concordance does not change with the amount of data we use to estimate it. If *T*=0.01, then the concordance factor will always be ≈0.34 for the simple tree in this example—barely any excess concordant gene trees relative to either of the other two discordant gene trees (each 0.33). This expectation is generated by the evolutionary process, not by the sampling of data. As a result, the numerical value of a concordance factor provides little information about the probability that a branch is true, because true branches in a species tree can have almost any value of a concordance factor.

However, it is clearly the case that our confidence in this species tree (as measured by statistical support) will increase with increasing amounts of data (see the examples in Figure 2). If we sampled only 10 loci, we might not even get the concordant tree as the most common topology. With increasing amounts of data (and appropriate models), both the bootstrap proportion and posterior probability of this branch will increase towards 100% (Figure 2). From a statistical perspective, the larger sample size increases our confidence that the concordant tree is the most common. While this is the behaviour we want in a measure of statistical support, it is precisely this observation that has led some authors to question the utility of measures of statistical support in modern phylogenomics (Kumar et al. 2012; Thomson and Brown 2022): our datasets our now so large that almost all such measures reach their maximum value on every branch. Despite this, concordance factors do provide some information on the amount of data that will be needed to have high statistical support: a branch with a low concordance factor will require commensurately more data before measures of statistical support approach 100%. Another useful way to think about the same idea is to consider that measures of support are often equivalent to asking how sure you can be that that the split in the species tree has a concordance factor that is higher than any split that conflicts with it (given a particular sample of genes and/or sites).

What is the biological importance of the distinction between concordance factors and either bootstrap support or posterior probabilities? One useful analogy (suggested to us by Cecile Ané) is with the statistical concepts of standard deviation and standard error. The concordance factor is similar to the standard deviation: it tells you about the spread of values in your data, regardless of how many datapoints you have. In contrast, measures of support are similar to standard errors: they tell you how confident you can be in your estimate of the mean of the data. Extending this analogy to a phylogenetic dataset, just because we are very sure of the topology of a species tree does not mean that all loci must follow this topology. For some questions we might need to know the species tree with high confidence, while for others it will be most important to understand the underlying variability in the gene trees (Hahn and Nakhleh 2016; Bravo et al. 2019).

There are also important practical implications of distinguishing between measures of concordance and measures of support. One relevant scenario arises when choosing an appropriate outgroup for phylogenetic analyses. Outgroups can be used for multiple purposes, including rooting ingroup relationships and polarizing the direction of evolutionary changes. Importantly, the main criterion for choosing outgroup lineages is that the taxa chosen always be sister to all ingroup lineages (i.e. “outside” the ingroup), as their relationships with all other species are not being assessed. Unfortunately, support measures are not always good measures of this property: we may be very sure that a lineage is sister to our ingroup in the species tree, even while an appreciable fraction of gene trees from this lineage lie within the ingroup clade. An effective outgroup should be an outgroup on every gene tree; i.e. the concordance factor for an outgroup should be 100%. Choosing outgroups without this property can mislead phylogenetic inferences (e.g. among the platyrrhine monkeys; Schrago and Seuánez 2019; Vanderpool et al. 2020). Outgroups should therefore be chosen considering levels of concordance, rather than levels of support.

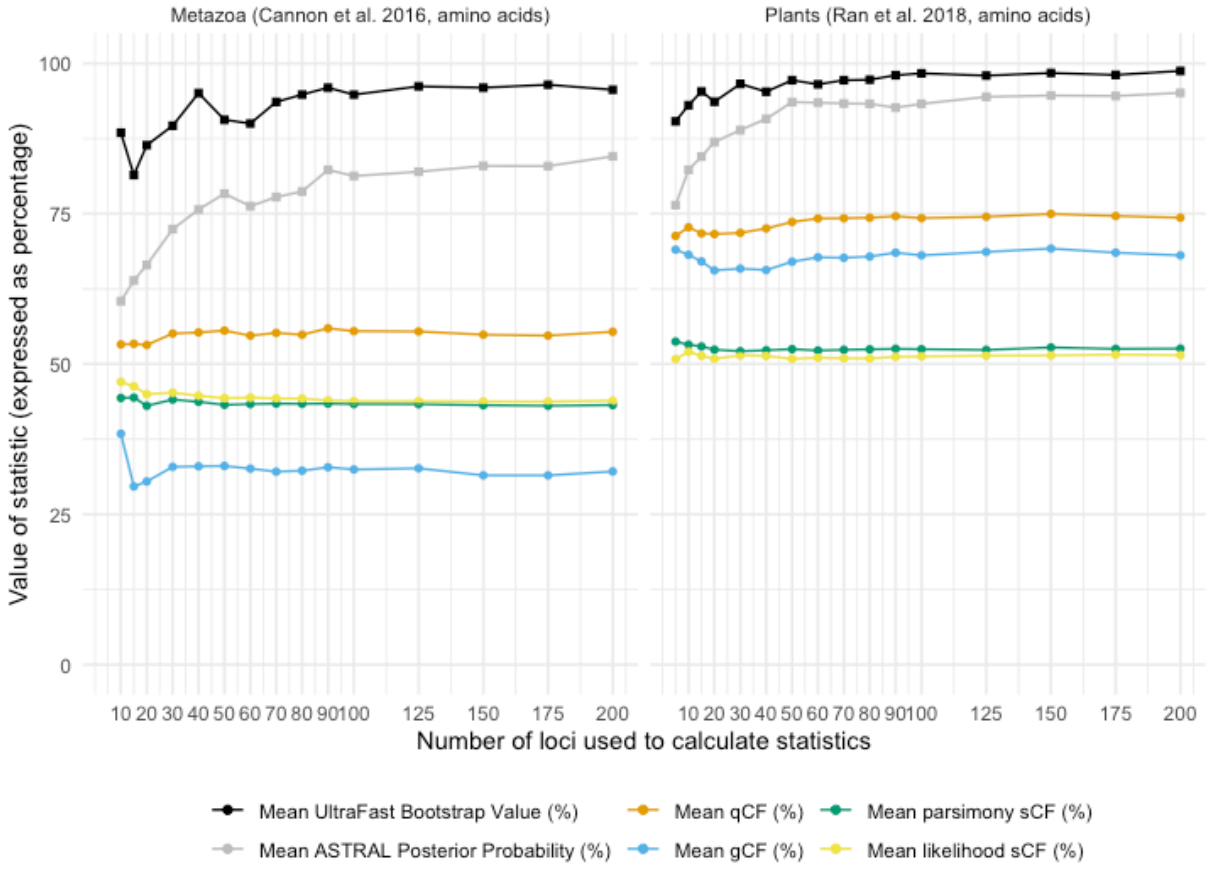

Figure 2. Phylogenetic statistical support (greyscale, square points) compared to measures of phylogenetic statistical variation (coloured, circular points) as dataset size increases. In two example datasets (Cannon et al. 2016; Ran et al. 2018) we calculated the mean values of two measures of statistical support: ultrafast bootstraps (Hoang et al. 2017) and ASTRAL posterior probabilities (Sayyari and Mirarab 2016), as well as four measures of concordance: gene concordance factor (gCF; Minh et al. 2020), quartet concordance factor (qCF; Mirarab et al. 2014) , and the site concordance factor (sCF), calculated with parsimony (Minh et al. 2020) and likelihood (Mo et al. 2023). All measures of concordance assume that the tree calculated from the 200-locus dataset was correct. The figure shows that as more loci are used to calculate the statistics, the average of the measures of statistical support tends to increase towards 100%, but the average of the measures of concordance tends to stabilise after being estimated inaccurately with small numbers of loci.

The concordance vector: a simple way to summarise concordance and discordance

Before we discuss how concordance and discordance factors are estimated, we first provide a simple framework to facilitate synthesis across authors, papers, and methodologies with many different terminologies: the concordance vector. The concordance vector is a simple summary that describes fundamental aspects of concordance and discordance for a branch. Each internal branch of a species tree will have its own concordance vector.

The challenge that the concordance vector seeks to address is to provide a compact and meaningful summary of both the concordance and discordance factors associated with a single branch. This can be difficult because, although the concordance factor is a single proportion, discordance factors can be much more complex. There can be a vast number of ways that gene tree can be discordant with a species tree, and in principle we could calculate a discordance factor for each of them. However, this would involve enumerating all possible tree topologies discordant with the clade of interest, and calculating the expected proportion of the genome matching that topology for each. This approach is neither practical nor particularly helpful: the number of discordant topologies will differ for different branches on the species tree, will be astronomical for many of them, and under most evolutionary scenarios the expected discordance factor for many or most topologies will be effectively zero.

The concordance vector, ψ, addresses this problem by summarising the concordance and discordance factors of a clade into four proportions that sum to 1. The first entry in the concordance vector, ψ_1, is simply the concordance factor; the remaining three entries (ψ_2, ψ_3, and ψ_4) summarise the discordance factors. The second and third entries in the concordance vector, ψ_2 and ψ_3, correspond to the two alternative topologies obtained by swapping around the relationships of the groups *A*, *B*, *C*, and *D* in Figure 1 (equivalent to nearest-neighbor interchanges in phylogenetics). These alternative topologies are biologically important because in many scenarios, such as incomplete lineage sorting, we expect them to be the topologies associated with the two highest discordance factors. Since there is no clear objective way to distinguish between the two topologies that represent ψ_2 and ψ_3, we simply denote ψ_2 to be the larger and ψ_3 the smaller of the two, as this simplifies the description and comparison of some interpretations and derivations of concordance factors (see below). The third discordance factor, ψ_4 is the sum of all other discordance factors, i.e. those associated with all discordant topologies that are not represented by ψ_2 and ψ_3. Since the sum of the concordance vector must equal 1, ψ_4 can be calculated by simply subtracting the rest of the concordance vector from 1.

The concordance vector helps to reveal the relationships between different methods of estimating and interpreting concordance and discordance factors. For example, certain approaches to estimating concordance factors—like the site and quartet concordance factors—assume that ψ_4 is zero, while others can be prone to overestimating or underestimating ψ_4 depending on the properties of the data. Some approaches to testing hypotheses about evolution compare values in the concordance vector; for example, one can test for deviations from a model of incomplete lineage sorting by testing the expectation that ψ_2 and ψ_3 should be equal. And some measures of node support, like Internode Certainty (Salichos and Rokas 2013), can be thought of as asking whether ψ_1 is larger than ψ_2 (see below). We next discuss these and other cases.

Estimating concordance factors

Any given set of biological processes associated with a group of evolving lineages is associated with an *expected* concordance vector for each internal branch of the species tree. In this section we discuss different approaches to estimating concordance vectors from empirical data. Thinking of empirical concordance factors as estimates of the true values helps to clarify how and why they differ from each other, while also retaining the original definition of a concordance factor as a biological parameter describing the “proportion of a genome for which [a given clade] is true” (Baum 2007). Despite this, we note that it is common for researchers to discuss concordance factors as summary statistics similar to various earlier notions of consensus and congruence (e.g. Adams 1972; Hillis 1987; Carpenter 1988). We discuss cases where this can be particularly useful below.

We cover three approaches to estimating concordance vectors: using genes, using quartets, and using sites. We follow recent convention by prefixing estimated concordance factors to indicate the input data for each, as this helps to distinguish them both from each other and from their expected values (i.e. the true but unknown concordance vectors). We denote concordance factors estimated from genes as gene concordance factors (gCFs), from quartets as quartet concordance factors (qCFs), and from sites as site concordance factors (sCFs). We name each associated concordance vector in the same way, for example gene concordance vector (gCV).

Gene concordance factors (gCFs)

Gene concordance factors (gCFs), first defined in 2007 (Ané et al. 2007; Baum 2007), seek to estimate the true concordance vector from a large collection of gene trees, themselves estimated from alignments of independent genes (here defined as unlinked loci without internal recombination). The simplest and most naïve way to calculate the gene concordance vector of a branch of interest is to first obtain a single topology for each gene tree (for example, using maximum likelihood), and then to count the proportion of these trees associated with the topologies assigned to ψ_1 to ψ_4 (e.g. the estimate of ψ_1 would simply be the proportion of gene trees that contain the branch of interest). A problem arises when some gene trees have missing taxa, but this can be solved by considering only those gene trees that *could* have contained the branch of interest, known as *decisive* gene trees (Minh et al. 2020). The resulting proportions can be considered estimates of the true entries in the concordance vector, whose accuracy and precision will depend on the total number of decisive gene trees for each branch and on the accuracy and precision with which each of the underlying gene trees is estimated, which are discussed further below.

Quartet concordance factors (qCFs)

Quartet concordance factors (qCFs) can refer to any approach in which each alignment of an independent locus is summarised not by a single tree (as in the gCF), but by a collection of sub-trees of four taxa (quartets). qCFs became popular alongside the program ASTRAL (Mirarab et al. 2014), as they are a standard output of this tool (though they are referred to as quartet "scores” or "frequencies" there). Since this time, they have been more widely adopted to quantify and explore conflicting signal in multi-locus datasets (e.g. Sayyari and Mirarab 2016;

Solís-Lemus and Ané 2016; Zhou et al. 2020; Rhodes et al. 2021; Allman et al. 2022). Calculating quartet concordance factors usually involves two steps. The first step is to estimate a set of quartets for each locus. This can be done by subsampling the alignment of that locus into all (or a large number of) possible groups of four taxa, and then estimating the unrooted quartet trees directly from the sequence data. However, it is more common to first estimate a single gene tree of all taxa for each locus (as for gCFs above), and to then extract unrooted quartets from that gene tree (i.e. the quartets relevant to the branch of interest); this approach provides more accurate quartets. The second step in estimating qCFs is to count the proportion of relevant quartets associated with the topologies assigned to ψ_1 to ψ_3 for the branch of interest. Because unrooted quartets only have four taxa and three possible topologies, they can only display internal branches that match ψ_1 to ψ_3 (i.e. ψ_4 is always zero for this measure). In other words, it is impossible for a quartet to display a branch that is not either in the species tree, or that represents one of the two splits that could be induced by ILS occurring on that branch.

Site concordance factors (sCFs)

Site concordance factors (sCFs) were first introduced in 2020 (Minh et al. 2020), using a parsimony-based approach, and later updated to use maximum likelihood (Mo et al. 2023). Most sites (for example, constant sites) contain no information about any branch in the tree. For that reason, the sCF focuses on *decisive* sites—those that contain information about the branch of interest. Similarly to qCFs, sCFs use a quartet of states at a single site to determine the implied topology. The site concordance vector is estimated by first sampling quartets (which is done slightly differently in the two sCF methods), then by counting the proportion of decisive sites in the sample that match ψ_1, ψ_2, or ψ_3. As with the qCF, ψ_4 is always zero for sCFs because it is impossible for a single decisive site for a branch to display any internal branch other than those associated with ψ_1, ψ_2, or ψ_3.

Understanding concordance factors

All of the measures of concordance and discordance described above seek to estimate entries of the concordance vector, but each comes with its own set of advantages and disadvantages. They also often estimate slightly different things and are used differently by downstream methods, so it is important to know how they differ. Below we discuss the key issues associated with each quantity.

Gene concordance factors

Gene concordance factors offer the fullest view of genealogical variation, but also come with the greatest caveats. In terms of information provided, not only are gene concordance factors the only approach that can estimate ψ_4, but they also allow us to expand the gene concordance vector beyond four entries. Recall that ψ_4 is associated with all topologies not accounted for by ψ_1, ψ_2, and ψ_3 (Figure 1). While the smaller vector used here allows us to have a common vocabulary for all the different approaches to quantifying concordance and discordance, there

can be quite a lot of information hidden within ψ_4, information that is only available when estimating full topologies from genes.

As an example of the sort of information found in ψ_4, Salichos and Rokas (2013; see also Salichos et al. 2014) introduced a measure called “internode certainty" based on the gene concordance vector. While we are not so sure that this statistic measures certainty of any kind, it does measure the magnitude of conflict among gene trees. In the simpler version of this statistic (denoted “IC”), for a single branch we calculate the degree of conflict between the two most common splits—often, but not always, ψ_1 and ψ_2—using an entropy-based measure. In the fuller version (called internode certainty all, or “ICA"), we calculate the degree of conflict for a branch among the n most common splits using an entropy-based measure. If n is greater than 3, then we must expand our concordance vector to be of length n (since ψ_4 typically refers to many more than a single split). We may be able to glean quite a lot from the frequency of gene trees beyond the most common three, but unfortunately ICA is one of the only methods we know of that uses this information. One reason for this may be that there are few programs that output these frequencies in a usable format, making the data inaccessible to most researchers.

The biggest limitation of gene concordance factors is gene tree estimation error. Estimation error is unavoidable with the limited phylogenetic information available in single-locus alignments. The requirement that loci be non-recombining often means that alignments will be even shorter than a whole gene, or that single-gene alignments will mistakenly contain recombination events, misleading tree inference in complex ways. Gene tree estimation error will cause trees to be assigned to the wrong entry in the concordance vector. Consider an extreme case of no biological discordance: small amounts of estimation error will decrease ψ_1 and increase ψ_2 and ψ_3, leading to an overestimate of the amount of discordance. Even in cases with biological discordance, small amounts of error will cause ψ_1, ψ_2, and ψ_3 to become more similar to one another. However, as the degree of error (as measured per-tree) increases, more and more estimated gene trees will not match any of these tree topologies, and ψ_4 will increase. With large amounts of gene tree estimation error, almost all of the gene trees will fall into ψ_4, regardless of the true level of biological discordance. When this occurs—such as in cases with high sequence divergence and short alignments—alternative approaches may be needed to estimate the entries in the gene concordance vector (e.g. Rosenzweig et al. 2022). Such topological errors may be responsible for some studies in which no single gene tree matches the species tree (e.g. Jarvis et al. 2014).

It is also possible for the gene concordance factor (i.e. ψ_1) to be overestimated using gCFs. This is most likely to occur when individual gene tree alignments span recombination breakpoints (an issue sometimes called "concatalescence"; Gatesy and Springer 2014). In this case, although multiple different topologies may be represented among the constituent loci, the resulting inferred gene tree from the combined alignment will reflect the majority signal in the data. Because ψ_1 is typically associated with the majority signal in the data, the result will often be an overestimate of ψ_1 and a concomitant underestimate of ψ_2, ψ_3, and ψ_4 (e.g. Mendes et al. 2019). In the extreme, we could estimate a “gene tree” from an entire chromosome or genome, and the resulting topology would most likely reflect the most common gene tree (but not

always—see Kubatko and Degnan 2007; Mendes and Hahn 2018). The effect of the two biases discussed here on gCFs will depend on the properties of each dataset and the exact methods used to estimate the gene trees. Regardless, gene concordance vectors should be interpreted with both issues in mind.

Quartet concordance factors

For both computational expediency and biological interpretability, full gene trees are often downsampled into quartets of taxa. Sampling a quartet usually means choosing four independent tips from a larger tree—for instance, we could sample hypothetical species *a*, *b*, *c*, and *d* from clades *A*, *B*, *C*, and *D* in Figure 1a. As mentioned before, for an unrooted quartet around a single branch of interest there are only three possible topologies, so we only have entries for ψ_1, ψ_2, and ψ_3 in the quartet concordance vector.

The more limited resolution of quartets (i.e. the assumption that ψ_4 is always zero) can be seen as both a strength and a weakness. The strength of this method is that each quartet contributes to an informative entry in a concordance vector, even if the full gene tree that it is a part of does not. For instance, if gene tree error caused even one of the clades in Figure 1 (*A*, *B*, *C*, or *D*) to be non-monophyletic, then the full gene tree would be placed in ψ_4; this would be true even if a single lineage was placed in the wrong clade (in this sense, so called 'rogue taxa' may have a large influence on gCFs and gCVs). In contrast, there would still be many informative quartets that we could sample from such a gene tree, even quartets that contain the single misplaced lineage. In this sense, small amounts of gene tree error are much more easily dealt with by counting quartet frequencies. This same feature could also be a weakness of using quartets: by assuming that ψ_4 is zero, ψ_1, ψ_2, and ψ_3 will be biased upwards when the true ψ_4 Is greater than zero. In the extreme case of high levels of per-gene error—when gene concordance measures might have little or no evidence for trees matching ψ_1-ψ_3—quartet calculations will be forced to populate these entries in the vector.

By far the biggest advantage of quartets is that many types of operations can be done on them easily and quickly. Quartet-based methods have become the dominant approach for inferring species trees, especially using the program ASTRAL (Mirarab et al. 2014; Mirarab and Warnow 2015; Zhang et al. 2018a). There is a rich history of methods for constructing species trees from constituent quartet trees, which are often called "puzzling", "amalgamation," or "assembly" methods (e.g. Strimmer and von Haeseler 1996; Bryant and Steel 2001; Snir and Rao 2010). The conceptual leap between these older methods and newer methods was largely driven by two advances. First, the growing size of datasets meant that instead of a handful of quartets from loci with nonoverlapping sets of taxa, genome-scale data provided many quartets estimated from each of thousands of loci containing mostly the same taxa. Second, it was recognized that using unrooted quartets provided accurate estimates of the species tree even in cases where there was discordance due to ILS—i.e. these methods are statistically consistent under the multispecies coalescent model. Any method for counting and combining quartets accurately should have this property, because the unrooted quartet topology (or rooted triplet topology) matching the species tree is always the most frequent under ILS alone (Hudson 1983;

Allman et al. 2011). (Rooted triplets work just as well as unrooted quartets in inferring species trees, but methods employing them are used less frequently [e.g. DeGiorgio and Degnan 2010; Liu et al. 2010].) For similar reasons, quartets sampled from reconstructed gene trees have become the currency of multiple methods that aim to infer introgression between species as violations of the ILS-only model (e.g. Huson et al. 2005; Solís-Lemus and Ané 2016).

In ASTRAL, quartet trees are sampled many times for each gene tree, with quartets across gene trees all counted together. Because tips are largely chosen at random, many quartets are not providing information about only a single internal branch of the species tree, but rather a span of branches. For this and similar reasons, many quartets from the same gene tree will not be independent of one another; any method that counts quartet frequencies must therefore take this non-independence into account. As with any method for inferring species trees, quartet methods can provide support measures—the confidence one should have in the inferred branch of the tree. ASTRAL uses local posterior probabilities based on the qCV (Sayyari and Mirarab 2016). In this case (and similar ones) the support metric is assessing our confidence that ψ_1 is the most frequent topology, which is obviously related to its magnitude relative to ψ_2 and ψ_3, as well as to the number of independent quartets we have sampled. Confidence measures for internal branches can also be calculated using the bootstrap with gene tree- and quartet-based methods: one simply has to construct many bootstrapped samples of the set of individual gene trees, reconstructing the species tree from each sampled set of trees to assess confidence.

Site concordance factors

Breaking gene trees down into quartet trees obviates some of the problems due to gene tree estimation error, but not all (Molloy and Warnow 2018). One must still infer a full gene tree from a short alignment, which is always prone to error. Therefore, as an alternative approach, one can calculate site concordance factors (Minh et al. 2020; Mo et al. 2023). sCFs were explicitly developed to estimate concordance and discordance without the need to divide alignments into short, non-recombining loci. As described earlier, they can be easily calculated from a long, concatenated alignment.

Site concordance factors may differ from gCFs and qCFs for both practical and theoretical reasons. There are a number of practical factors that affect sCFs, with complex interactions among these factors making it hard to predict whether they will be consistently higher or lower than the other measures on any particular branch of a tree or in any particular dataset (e.g. Figure 2). sCFs are most similar to qCFs, in that they assume ψ_4 is zero. As a consequence, sCFs may be higher than gCFs because ψ_1, ψ_2, and ψ_3 will all be upwardly biased. However, both gCFs and qCFs may be higher than sCFs if they are calculated from loci that contain multiple tree topologies, leading to an overestimate of ψ_1 (see above). Most importantly, multiple substitutions at a single site and variation in nucleotide substitution rates across a tree can cause sCFs to overestimate the amount of discordance (Kück et al. 2022; Mo et al. 2023). The likelihood version of sCFs (Mo et al. 2023) attempts to minimize this problem, though care must still be taken when using a single site to reveal an underlying tree topology.

Aside from practical considerations, there is an important conceptual reason why sCFs may differ from other measures: they are measuring a slightly different quantity. Both gCFs and qCFs are directly estimating the quantity expressed in equation 1: the proportion of the genome having a particular tree topology. In contrast, sCFs are measuring the proportion of sites supporting a particular tree topology, which is a function of both gene tree frequency (the quantity measured by gCFs and qCFs) and the length of the relevant branch in each gene tree. This measure will be very close to, but slightly different from, the parameter estimated by gCFs and qCFs. If the expected value of those measures is given by equation 1, the expected value of sCFs (assuming that sites are unlinked) is:

$$P(concordant\ site) = 1 - \frac{2}{3+3Te^{-T}} \qquad (3)$$

Another way to think about this is that the length of the branch of interest in the species tree impacts the sCF twice—once by determining the frequency of the gene tree topologies, and then again by determining the branch lengths of those topologies. Because of this, and as can be seen in Figure 3, sCFs are expected to go up slightly faster than either gCFs or qCFs as a function of the length of the internal branch of the species tree. This implies that sCFs are expected to be slightly higher than the two other methods in the absence of estimation error. However, the expected values are close enough that we predict that practical considerations in estimation (such as gene tree estimation error) will be the deciding factor in which is higher or lower, and this seems to be borne out by the examples in Figure 2.

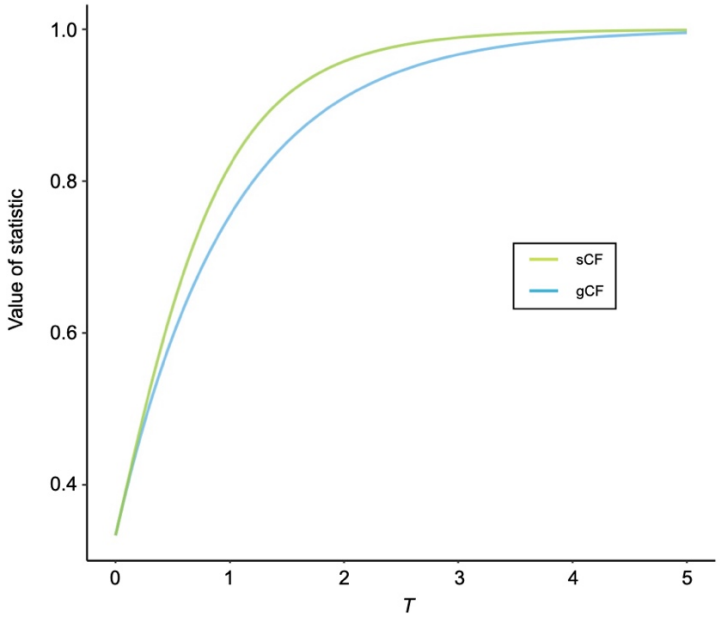

Figure 3. The difference between expected site concordance factors (sCF) and gene concordance factors (gCF) as the branch length of the species tree changes. Here we plot the expected values of gCFs (equation 1; blue) and sCFs (equation 3; yellow-green) as a function of the internal branch length, *T*. The figure shows that sCFs go up faster than gCFs, as they are affected both by the number of concordant gene trees and the length of the internal branch of such trees.

Although sCFs were introduced as a way to measure concordance and discordance on a species tree, the various numbers underlying these calculations have been used in multiple preceding applications. These applications ask about the underlying species tree topology, as well as deviations from this history due to introgression. The most widely used method employing site-based quartets to infer a species tree is SVDquartets (Chifman and Kubatko 2014; Swofford and Kubatko 2023). SVDQuartets considers all quartet site-patterns—i.e. not just decisive sites, and separate patterns for each combination of nucleotides—placing them in one of three matrices for each branch. The three matrices represent the three possible splits corresponding to ψ_1, ψ_2, and ψ_3. If singular value decomposition (SVD) is carried out on each matrix, it can be shown that the one with the lowest SVD score will match the species tree. If such calculations are carried out for many sampled quartets of tip species, the resulting set of highest-scoring quartet trees can be input into an assembly method (as described above for other quartet methods), giving a full species tree as output (see a similar site-pattern method in Zhang et al. 2023).

Possibly the most widely known use of the site concordance vector is the ABBA-BABA (“D”) test for introgression (Green et al. 2010; Patterson et al. 2012). This test employs only the counts of sites in entries ψ_2 and ψ_3: given the species tree in Figure 1a, if we denote ancestral states with “A” and derived states with “B” (and always assign the outgroup to be “A”) then the decisive site-pattern in ψ_2 can be written as “BABA” (because species *A* and *C* will share the derived state) and the one in ψ_3 can be written as “ABBA” (Figure 1b). As mentioned above, in the absence of introgression (and ancestral population structure; Slatkin and Pollack 2008; Durand et al. 2011), we expect ψ_2 and ψ_3 to be equal. The ABBA-BABA test asks whether this is true across a genome, taking the non-independence of nearby sites into account via block bootstrapping of genomic windows. Similarly, the Δ statistic of Huson et al. (2005) tests for the equality of ψ_2 and ψ_3 using quartet trees, with a similar interpretation of introgression if they are not equal (see Vanderpool et al. 2020 and Suvorov et al. 2022 for applications to whole genomes).

The future of concordance factors

Concordance factors and concordance vectors are already very useful summaries of biological variation, but there are many ways they could be improved, and many new areas in which they could be applied to better understand complex evolutionary histories.

The simplest, and perhaps most useful, modification to current practice would be to routinely provide confidence intervals for all entries of the concordance vector. Currently, concordance factors and other entries of the concordance vector are presented as point estimates on every branch of the tree, largely because this is the output given by popular software for calculating concordance factors like ASTRAL and IQ-TREE (Mirarab et al. 2014; Minh et al. 2020). However, point estimates can sometimes be misleading, as their interpretation can vary dramatically depending on the confidence intervals around them. This problem is compounded by the fact that the sample size for concordance factors can vary from branch to branch in a tree: it can depend on taxon sampling for gCFs and qCFs, and on the number of informative

sites per branch for sCFs. To give an extreme example: a gCF of 50% may indicate substantial underlying variation in gene trees when the sample size is large (e.g. with 1000 gene trees, the 95% confidence intervals on the gCF would be 47% and 53%), but may contain relatively little useful information if the sample size is small (e.g. with 4 gene trees the 95% confidence intervals are 0% and 100%). Calculating confidence intervals on all entries of the concordance vector is very simple, and could be done via the bootstrap or by using a Dirichlet-multinomial distribution (Gelman et al. 2013). Displaying confidence intervals alongside concordance factors (and other entries of the concordance vector) would help biologists to better interpret the underlying biological variation, particularly when considering biological hypotheses about the causes of this variation.

One of the biggest challenges to estimating concordance factors is gene tree estimation error. Gene tree estimation error affects both gene and quartet concordance factors, though to differing degrees (see above). As a result, it affects any methods that rely on these estimates. For example, because gene tree estimation error leads to overestimates of discordance, it will lead to underestimates of branch lengths based on the concordance vector (e.g. those calculated in ASTRAL). Many approaches have been developed to mitigate gene-tree estimation error and its effects on concordance factors (Larget et al. 2010; Boussau et al. 2013; Zhang and Mirarab 2022). One additional option may be to move beyond a binary view of concordance—i.e. that a gene tree is either concordant or discordant with a branch of interest—and instead to incorporate the *degree* of discordance demonstrated by a gene tree. Such a measure could be achieved in many ways, for example by measuring the difference in likelihoods when a gene tree is constrained to contain a branch of interest, or by calculating how much a gene tree would have to be altered to recover the branch of interest (e.g. using the Transfer Bootstrap Expectation; Lemoine et al. 2018). Regardless, accounting for and/or mitigating gene tree estimation error when calculating concordance factors remains a largely open problem.

Concordance factors could also be extended in multiple ways. For instance, the current discussion has assumed a bifurcating species tree and single-copy gene trees, though more and more datasets may extend beyond both of these constraints. In terms of the species tree, it is now possible to infer species networks from many datasets (Wen et al. 2018; Zhang et al. 2018b). A question then arises: how best to represent and measure concordance with a network? One simple approach might be to consider all of the relationships shown in the network as concordant entries in the concordance vector; e.g. both ψ_1 and ψ_2 could be concordance factors if both appear in a network. However, we suspect that there will also be other ways to summarize concordance between gene trees and species networks. Similarly, species tree topologies can now be accurately reconstructed using gene trees containing paralogs (see Smith and Hahn 2021 for a recent review). Using gene trees with multiple tips from the same species requires new ways to quantify concordance: gCFs and sCFs cannot yet be calculated for such trees, although qCFs can (e.g. Smith et al. 2022). New approaches for calculating concordance factors from all of these non-standard data types will be necessary in the near future.

Finally, visualising concordance and discordance remains challenging. The simplest and most commonly used approach is to represent the data using a single best-estimate of a binary tree (usually the species tree), and then to label each branch with entries of the concordance vector (most often by displaying only the first entry, the concordance factor). While this representation can give some indication of the scale of discordance, it does not represent the discordant relationships themselves. The latter can be achieved using networks to represent the discordance (e.g. Huson 1998; Huson and Scornavacca 2012), overlaying the topology of each gene tree onto the species tree itself (e.g. Bouckaert 2010; Schliep 2011), or by colouring alignments based on the inferred topology of each region (e.g. Fontaine et al. 2015; Edelman et al. 2019). However, none of these approaches is ideal, and each has its own limitations. New approaches that enable fast and clear ways to visualise and to query discordant relationships would help researchers to quickly understand and interrogate their phylogenomic datasets.

Concluding remarks

Phylogenomic datasets are rapidly approaching complete sampling, i.e. entire genomes sequenced and assembled for every tip of the tree being estimated. Largely because of this, researchers will continue to move beyond single representations of the relationships among taxa (e.g. species trees and species networks), and will increasingly focus on estimating and interpreting the complex set of relationships underlying all sites of the sampled genomes. Concordance factors are a useful tool for summarising and interpreting this variation, ones that will be particularly useful for bridging the gap between species trees and underlying genomic variation. We envision their ever-widening use and further development for the foreseeable future.

Acknowledgements

We thank Megan Smith, Yu Mo, and Minh Bui for helpful conversations and assistance. This work was supported by National Science Foundation grant DEB-1936187 to M.W.H. and Australian Research Council grant DP200103151 to M.W.H. and R.L.

References

Adams EN, III. 1972. Consensus techniques and the comparison of taxonomic trees. Systematic Zoology 21:390-397.

Allman ES, Degnan JH, Rhodes JA. 2011. Identifying the rooted species tree from the distribution of unrooted gene trees under the coalescent. Journal of Mathematical Biology 62:833-862.

Allman ES, Mitchell JD, Rhodes JA. 2022. Gene tree discord, simplex plots, and statistical tests under the coalescent. Systematic Biology 71:929-942.

Ané C, Larget B, Baum DA, Smith SD, Rokas A. 2007. Bayesian estimation of concordance among gene trees. Molecular Biology and Evolution 24:412-426.
Baum DA. 2007. Concordance trees, concordance factors, and the exploration of reticulate genealogy. Taxon 56:417-426.
Bouckaert RR. 2010. DensiTree: Making sense of sets of phylogenetic trees. Bioinformatics 26:1372-1373.
Boussau B, Szöllosi GJ, Duret L, Gouy M, Tannier E, Daubin V. 2013. Genome-scale coestimation of species and gene trees. Genome Research 23:323-330.
Bravo GA, Antonelli A, Bacon CD, Bartoszek K, Blom MP, Huynh S, Jones G, Knowles LL, Lamichhaney S, Marcussen T, et al. 2019. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. PeerJ 7:e6399.
Bryant D, Steel M. 2001. Constructing optimal trees from quartets. Journal of Algorithms 38:237-259.
Cannon, J. T., B. C. Vellutini, J. Smith, F. Ronquist, U. Jondelius, and A. Hejnol. 2016. Xenacoelomorpha is the sister group to Nephrozoa. Nature 530:89-93.
Carpenter JM. 1988. Choosing among multiple equally parsimonious cladograms. Cladistics 4:291-296.
Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30:3317-3324.
DeGiorgio M, Degnan JH. 2010. Fast and consistent estimation of species trees using supermatrix rooted triples. Molecular Biology and Evolution 27:552-569.
Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution 24:332-340.
Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. Molecular Biology and Evolution 28:2239-2252.
Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. Science 366:594-599.
Edwards SV, Beerli P. 2000. Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54:1839-1854.
Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.
Felsenstein J. 2004. Inferring phylogenies. Sunderland, MA: Sinauer Associates.
Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, Jiang X, Hall AB, Catteruccia F, Kakani E, et al. 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science 347:1258524.
Gadagkar SR, Rosenberg MS, Kumar S. 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 304:64-74.
Gatesy J, Springer MS. 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Molecular Phylogenetics and Evolution 80:231-266.
Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2013. Bayesian Data Analysis. New York: Chapman and Hall/CRC.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, et al. 2010. A draft sequence of the Neandertal genome. Science 328:710-722.
Hahn MW, Nakhleh L. 2016. Irrational exuberance for resolved species trees. Evolution 70:7-17.
Hillis DM. 1987. Molecular versus morphological approaches to systematics. Annual Review of Ecology and Systematics 18:23-42.
Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2017. UFBoot2: Improving the ultrafast bootstrap approximation. Molecular Biology and Evolution 35:518-522.
Hudson RR. 1983. Testing the constant-rate neutral allele model with protein sequence data. Evolution 37:203-217.
Huson DH. 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14:68-73.
Huson DH, Klopper T, Lockhart PJ, Steel MA. 2005. Reconstruction of reticulate networks from gene trees. Research in Computational Molecular Biology, Proceedings 3500:233-249.
Huson DH, Scornavacca C. 2012. Dendroscope 3: An interactive tool for tooted phylogenetic trees and networks. Systematic Biology 61:1061-1067.
Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346:1320-1331.
Kubatko LS, Degnan JH. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology 56:17-24.
Kück P, Romahn J, Meusemann K. 2022. Pitfalls of the site-concordance factor (sCF) as measure of phylogenetic branch support. NAR Genomics and Bioinformatics 4.
Kumar S, Filipski AJ, Battistuzzi FU, Kosakovsky Pond SL, Tamura K. 2012. Statistics and truth in phylogenomics. Molecular Biology and Evolution 29:457-472.
Larget BR, Kotha SK, Dewey CN, Ané C. 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26:2910-2911.
Lemoine F, Domelevo Entfellner JB, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, Gascuel O. 2018. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature 556:452-456.
Liu L, Yu L, Edwards SV. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evolutionary Biology 10:302.
Lutteropp S, Kozlov AM, Stamatakis A. 2020. A fast and memory-efficient implementation of the transfer bootstrap. Bioinformatics 36:2280-2281.
Maddison WP. 1997. Gene trees in species trees. Systematic Biology 46:523-536.
Mendes FK, Hahn MW. 2018. Why concatenation fails near the anomaly zone. Systematic Biology 67:158-169.
Mendes FK, Livera AP, Hahn MW. 2019. The perils of intralocus recombination for inferences of molecular convergence. Philosophical Transactions of the Royal Society B 374:20180244.
Minh BQ, Hahn MW, Lanfear R. 2020. New methods to calculate concordance factors for phylogenomic datasets. Molecular Biology and Evolution 37:2727-2733.
Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Molecular Biology and Evolution 30:1188-1195.
Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30:i541-i548.

Mirarab S, Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31:i44-i52.
Mo YK, Lanfear R, Hahn MW, Minh BQ. 2023. Updated site concordance factors minimize effects of homoplasy and taxon sampling. Bioinformatics 39:btac741.
Molloy EK, Warnow T. 2017. To include or not to include: The impact of gene filtering on species tree estimation methods. Systematic Biology 67:285-303.
Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. Genetics 192:1065-1093.
Pease JB, Haak DC, Hahn MW, Moyle LC. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biology 14:e1002379.
Ran, J.-H., T.-T. Shen, M.-M. Wang, and X.-Q. Wang. 2018. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. Proceedings of the Royal Society B: Biological Sciences 285:20181012.
Rannala B, Yang Z. 1996. Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. Journal of Molecular Evolution 43:304-311.
Rhodes JA, Baños H, Mitchell JD, Allman ES. 2021. MSCquartets 1.0: quartet methods for species trees and networks under the multispecies coalescent model in R. Bioinformatics 37:1766-1768.
Rosenzweig BK, Kern AD, Hahn MW. 2022. Accurate detection of incomplete lineage sorting via supervised machine learning. bioRxiv:2022.2011.2009.515828.
Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497:327-331.
Salichos L, Stamatakis A, Rokas A. 2014. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. Molecular Biology and Evolution 31:1261-1271.
Sayyari E, Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Molecular Biology and Evolution 33:1654-1668.
Schliep KP. 2010. phangorn: Phylogenetic analysis in R. Bioinformatics 27:592-593.
Schrago CG, Seuánez HN. 2019. Large ancestral effective population size explains the difficult phylogenetic placement of owl monkeys. American Journal of Primatology 81:e22955.
Shen X-X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nature Ecology & Evolution 1:0126.
Simon C. 2020. An evolving view of phylogenetic support. Systematic Biology 71:921-928.
Slatkin M, Pollack JL. 2008. Subdivision in an ancestral species creates asymmetry in gene trees. Molecular Biology and Evolution 25:2241-2246.
Smith ML, Hahn MW. 2021. New approaches for inferring phylogenies in the presence of paralogs. Trends in Genetics 37:156-169.
Smith ML, Vanderpool D, Hahn MW. 2022. Using all gene families vastly expands data available for phylogenomic inference. Molecular Biology and Evolution 39:msac112.
Snir S, Rao S. 2010. Quartets MaxCut: A divide and conquer quartets algorithm. IEEE/ACM Trans. Comput. Biol. Bioinformatics 7:704–718.
Solís-Lemus C, Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLOS Genetics 12:e1005896.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. Systematic Biology 57:758-771.
Stefanović S, Rice DW, Palmer JD. 2004. Long branch attraction, taxon sampling, and the earliest angiosperms: Amborella or monocots? BMC Evolutionary Biology 4:35.
Strimmer K, von Haeseler A. 1996. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. Molecular Biology and Evolution 13:964-964.
Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell PJ, Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. Current Biology 32:111-123.e115.
Swofford DL, Kubatko LS. 2023. Species tree estimation using site pattern frequencies. In: Kubatko L, Knowles LL, editors. Species Tree Inference. Princeton, NJ: Princeton University Press.
Thomson RC, Brown JM. 2022. On the need for new measures of phylogenomic support. Systematic Biology 71:917-920.
Vanderpool D, Minh BQ, Lanfear R, Hughes D, Murali S, Harris RA, Raveendran M, Muzny DM, Hibbins MS, Williamson RJ, et al. 2020. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. PLoS Biology 18:e3000954.
Wen D, Yu Y, Zhu J, Nakhleh L. 2018. Inferring phylogenetic networks using PhyloNet. Systematic Biology 67:735-740.
Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018a. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:153.
Zhang C, Ogilvie HA, Drummond AJ, Stadler T. 2018b. Bayesian inference of species networks from multilocus sequence data. Molecular Biology and Evolution 35:504-517.
Zhang C, Mirarab S. 2022. Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. Molecular Biology and Evolution 39.
Zhang C, Nielsen R, Mirarab S. 2023. CASTER: Direct species tree inference from whole-genome alignments. bioRxiv:2023.2010.2004.560884.
Zhou X, Lutteropp S, Czech L, Stamatakis A, Looz MV, Rokas A. 2020. Quartet-based computations of internode certainty provide robust measures of phylogenetic incongruence. Systematic Biology 69:308-324.