The Meaning and Measure of Concordance Factors in Phylogenomics

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

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

***Corresponding authors:** E-mails: rob.lanfear@anu.edu.au; mwh@iu.edu.

Associate editor: Koichiro Tamura

Abstract

As phylogenomic datasets have grown in size, researchers have developed new ways to measure biological variation and to assess statistical support for specific branches. Larger datasets have more sites and loci and therefore less sampling variance. While we can more accurately measure the mean signal in these datasets, lower sampling variance is often reflected in uniformly high measures of branch support—such as the bootstrap and posterior probability—limiting their utility. Larger datasets have also revealed substantial 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 for many purposes. In contrast to measures of statistical support, the degree of underlying topological variation among loci should be approximately constant regardless of the size of the dataset. "Concordance factors" (CFs) and similar statistics have therefore become increasingly important tools in phylogenetics. In this review, we explain why CFs should be thought of as descriptors of topological variation rather than as measures of statistical support, and argue that they provide important information about the predictive power of the species tree not contained in measures of support. We review a growing suite of statistics for measuring concordance, compare them in a common framework that reveals their interrelationships, and demonstrate how to calculate them using an example from birds. We also discuss how measures of topological variation might change in the future as we move beyond estimating a single "tree of life" toward estimating the myriad evolutionary histories underlying genomic variation.

Key words: **concordance, phylogenomics, coalescent.**

Introduction

As recently as a decade ago, the molecular datasets commonly used in phylogenetics were quite small, consisting of perhaps a handful of loci. The limited available data meant that there was little point in trying to examine variation in phylogenetic signal among the loci that constituted a dataset. Instead, most concerns revolved around quantifying 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 are measures of sampling variance, intended to evaluate the reliability of trees inferred from small numbers of informative substitutions and 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 mean 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 have greatly reduced the sampling variance in typical phylogenetic datasets. While some support methods have been extended to be faster for datasets with huge numbers of sites or to measure slightly different quantities for datasets with huge numbers of taxa (Stamatakis et al. 2008; Minh et al. 2013; Hoang et al. 2017; Lemoine et al. 2018; Lutterop et al. 2020), in general the consequence of larger datasets has been to lower sampling variance. This lower variance means that branch support measures are almost always uniformly high, even when the inferred branches are

Received: December 22, 2023. **Revised:** September 25, 2024. **Accepted:** October 04, 2024

wrong (Gadagkar et al. 2005). This limits the utility of branch support measures (Thomson and Brown 2022).

Although sampling variance has been reduced in so-called "phylogenomic" datasets, there has been increased recognition of biological variation and technical error in such data, which has led to calls to quantify and report this variation alongside support measures (Maddison 1997; Gadagkar et al. 2005). Biological variation arises because individual loci do not have to share the same topology with either the species tree or with each other. The biological processes that drive this variation 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). Regardless of the cause, biological variation contains information about evolutionary processes and therefore represents a rich source of data (Bravo et al. 2019). Technical error (such as model misspecification or misalignment), or simply a limited amount of information, can result in both systematic and stochastic gene tree inference error. Combined with biological variation, these errors mean that many—sometimes most—gene trees in a large phylogenomic dataset will not match the species tree. We refer to these gene trees as discordant and to those that match the species tree as concordant. (Note that a "gene tree" can refer to the topology at any locus, not just in protein-coding genes.)

It is particularly important to consider concordance and discordance when interpreting the evolutionary history of genes and traits. This is because, thanks to discordance, neither changes in genes nor changes in phenotypic traits will necessarily follow the species tree (Hahn and Nakhleh 2016). Therefore, when gene trees do not match species trees, standard phylogenetic methods that rely on the species tree to infer the number of times a trait has evolved, the direction of evolution, the rate of evolution, and the timing of trait transitions can be misled (Guerrero and Hahn 2018; Mendes et al. 2018; Hibbins et al. 2020, 2023; Adams et al. 2024; Schraiber et al. 2024). There are a growing number of biological examples of traits controlled by variation at discordant loci, such that forcing these traits to be analyzed on the species tree will generate false patterns of homoplasy (e.g. Fontaine et al. 2015; Lamichhaney et al. 2016; Li et al. 2016; Pease et al. 2016; Han et al. 2017; Palesch et al. 2018; Wu et al. 2018; Hibbins et al. 2020; Urban et al. 2021; Feng et al. 2022). Similar problems due to discordance can arise when estimating branch lengths or testing for positive selection in DNA sequences (Good et al. 2013; Mendes and Hahn 2016; Carruthers et al. 2022; Yan et al. 2023), when estimating the number of gene duplications and losses (Hahn 2007; Neafsey et al. 2015; Mishra et al. 2023), or when studying patterns of amino acid convergence through time (Mendes et al. 2016; Mendes et al. 2019; Corbett-Detig et al. 2020). Discordance between gene trees and the true underlying species tree can also result in incorrect estimation of the species tree itself when using concatenation methods (Kubatko and Degnan 2007; Liu and Edwards 2009; Mendes and Hahn 2018). All of these incorrect inferences arise because of the assumption that gene trees match each other and the species tree.

In this sense, concordance describes the extent to which the species tree is *predictive* of the history of any given gene or trait for a clade, and therefore the extent to which analyses based on the species tree will be correct (noting, of course, that accurately estimating concordance can be difficult; see below). Even when the inferred species tree is the correct description of population splitting through time, there are now many phylogenomic examples in which the majority of gene trees do not match the species tree across multiple branches (e.g. Jarvis et al. 2014; Fontaine et al. 2015; Pease et al. 2016; Copetti et al. 2017; Wu et al. 2018; Edelman et al. 2019; Thomas et al. 2023; Larson et al. 2024). In these cases, it might actually be surprising to see a gene or a trait whose evolutionary history is best explained by the species tree. Concordance factors (CFs) thus add important context to any estimate of a species tree, because they allow us to identify which branches of the species tree are most relevant to predicting the history of the genome at any given locus and which are not.

At least partly because of their important role in evolutionary inference, recent years have seen a rapid proliferation in ways to estimate and interpret concordance and discordance in phylogenomic datasets. In this review, we introduce these methods and their interpretation, highlighting how several of them are related via a shared set of simpler measures and explaining 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. We also stress that 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 (Fig. 1a). In this example, the three gene trees have three different topologies—one that is concordant with the species tree (gene tree 1, Fig. 1b) and two that are discordant (gene trees 2 and 3, Fig. 1b). Although gene trees can vary in their branch lengths, we do not consider this source of variation in our labeling 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

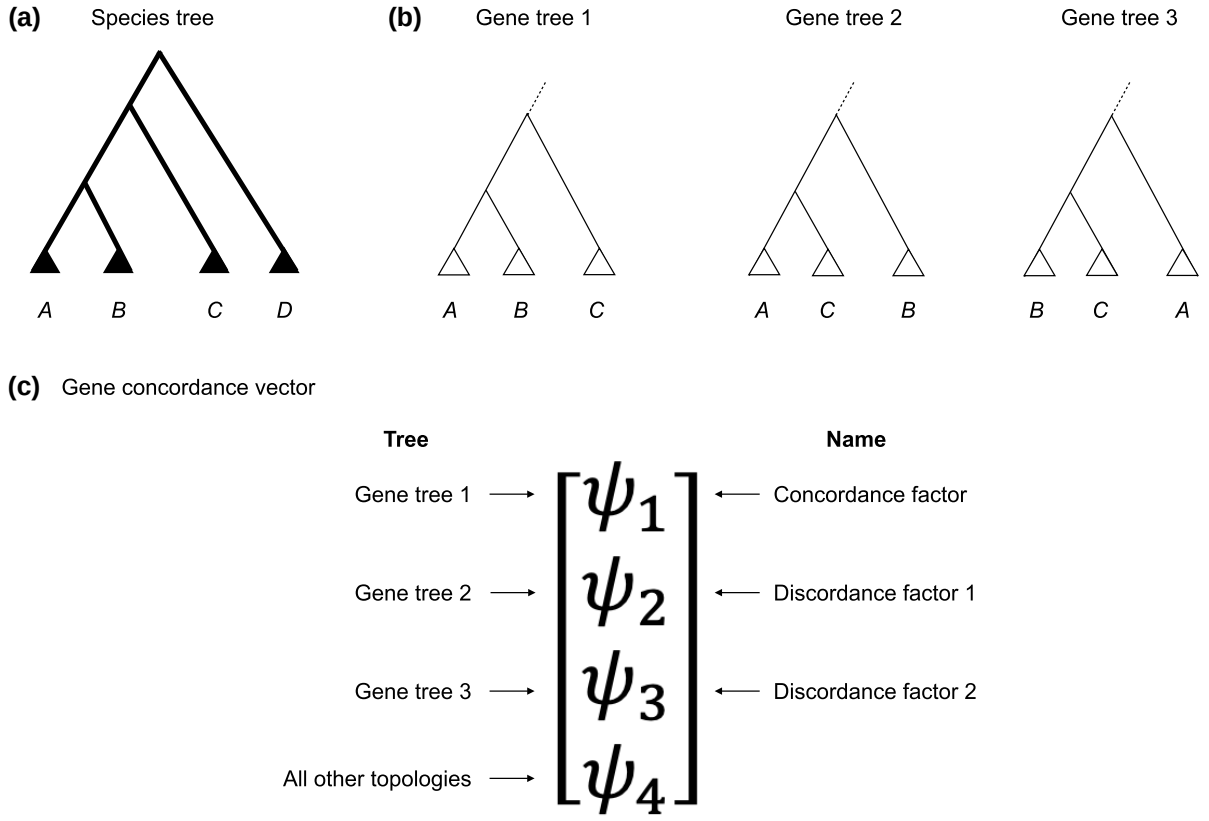

Fig. 1. Concordance and discordance. a) A species tree of four clades of organisms, *A*, *B*, *C*, and *D*. Clade *D* is the outgroup. b) Three possible gene trees were derived from the species tree in a). Gene tree 1 is concordant with the species tree, while gene trees 2 and 3 are discordant. c) The gCV, 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 "CF"). ψ_2 and ψ_3 are the number of gene trees that match discordant trees 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 nor gene tree 3 (for example, any gene tree in which clades *A*, *B*, or *C* are not monophyletic).

2000), making branch lengths a less useful measure of concordance.

For a rooted three-taxon tree, or an unrooted four-taxon tree, there are only three possible topologies, all of which are shown in Fig. 1b. As a result, the trees in Fig. 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 double-factorially 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; Sun et al. 2021; Larson et al. 2024). 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 proportions of gene trees are concordant with that branch. In other words, most researchers seek to estimate a CF 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. CFs have become a widely used metric to describe this heterogeneity (Baum 2007). (These are also sometimes called gene support frequencies; Gadagkar et al. 2005.) The CF 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 makes clear why CFs describe the predictive power of the species tree: they directly tell us what fraction of the genome follows the species tree and therefore is predicted by it. The 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 CF 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 DF (note that some papers refer to all concordance and discordance measures collectively as CFs; e.g. Allman et al. 2022). Together, CFs and DFs provide useful information for understanding evolutionary histories and for testing evolutionary hypotheses; here, we summarize them in a single vector called the concordance vector (Fig. 1c), which we introduce below.

In the rest of the paper, we review the meaning of CFs, how they can be estimated, and how these different estimates can be interpreted; we highlight these varied interpretations using an example from birds. We start by distinguishing CFs 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 different methods that have been developed for estimating CFs and DFs 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. Using data from 363 bird genomes (Stiller et al. 2024), we show how to calculate and compare these different measures of concordance and discordance. We conclude with suggestions for future directions.

The Concordance Factor Is Not a Measure of Statistical Support

A seemingly common misconception is to treat CFs as measures of statistical support. This is not correct: CFs are (estimates of) biological parameters, i.e. the proportion of the genome for which a given clade is true, not measures of statistical support for a clade (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 toward their maximum possible value as we add more data to the analysis (cf. Kumar et al. 2012). The same is not true for CFs. As we add more data to an analysis, estimates of CFs will become more precise but will not approach any limiting value. This is demonstrated for two empirical datasets in Fig. 2—as we add more data to the analyses, measures of statistical support (here, the UltraFast Bootstrap [Hoang et al. 2017] and the ASTRAL local posterior probability [Sayyari and Mirarab 2016], both shown in gray scale with square points) tend to increase toward their maximum value, while measures of concordance (shown in color with circular points) have higher variance at lower sample sizes but quickly stabilize to consistent values as sample size increases. This distinction highlights the fact that we should not think of any particular value of concordance as "good" or "bad". CFs are estimates of biological parameters, so it makes as little sense to attach judgment to a particular value of a CF as it would attach judgement to a particular value of a variance or a standard deviation.

To see what determines CFs (and DFs), consider the simplest species tree that can have discordance (Fig. 1a, imagining a single representative is sampled from clades *A to D*). Under a model of ILS, the length of the single 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, then the probability of sampling a locus with a concordant gene tree is (Hudson 1983):

$$P(\text{concordance}) = 1 - \frac{2}{3}e^{-T}. \quad (1)$$

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

$$P(\text{discordance}) = \frac{1}{3}e^{-T}. \quad (2)$$

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 CF—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). As an example, after one coalescent unit ($T = 1$), ~75% of trees will be concordant and 25% will be discordant. Note that while CFs can never be >1, for trees with more than three taxa they can be <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 CF 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 CF provides little information about

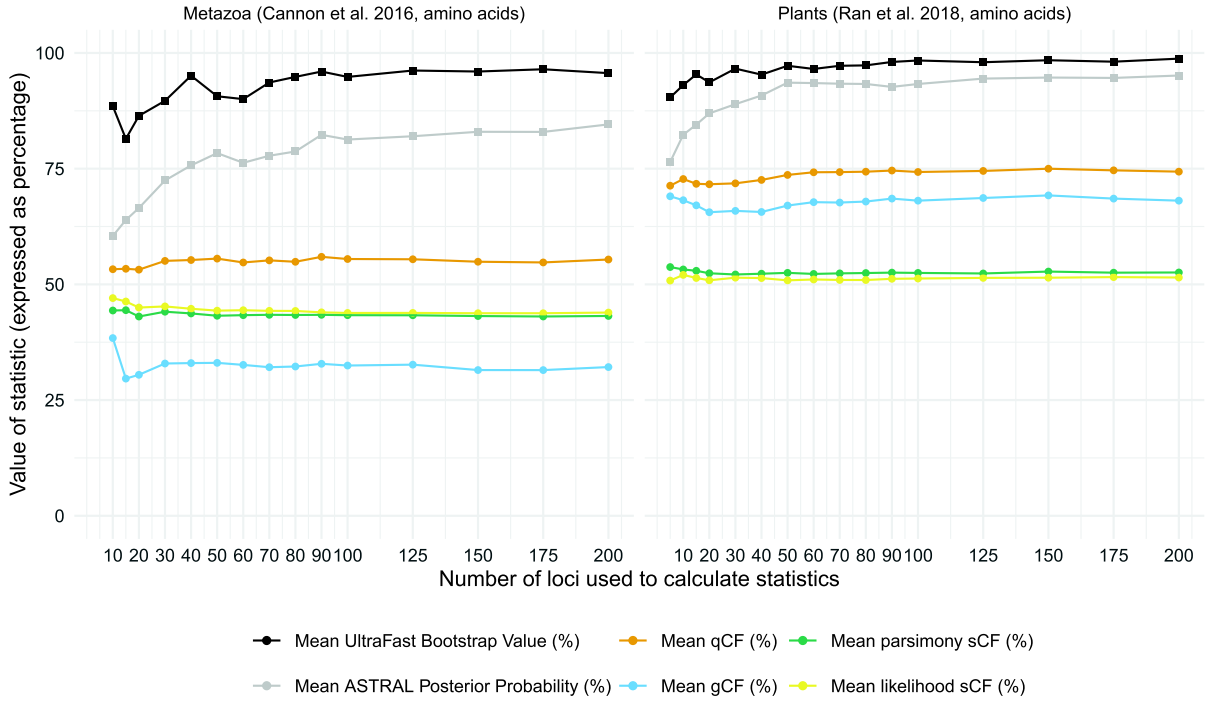

Fig. 2. Phylogenetic statistical support compared with measures of phylogenetic statistical variation as dataset size increases. In two example datasets (Cannon et al. 2016; Ran et al. 2018), we calculated the mean values of two measures of statistical support: UltraFast Bootstraps (Hoang et al. 2017) and ASTRAL local posterior probabilities (Sayyari and Mirarab 2016), as well as four measures of CFs: gCF (Minh et al. 2020), qCF (Mirarab et al. 2014), and the sCF calculated with parsimony (Minh et al. 2020) and likelihood (Mo et al. 2023), assuming 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 (grayscale lines, square points) tends to increase toward 100%, but the average of the measures of concordance (colored lines, circular points) tends to stabilize after being estimated inaccurately with small numbers of loci. Even with just 200 loci, most bootstrap and posterior probability values become very high (86.7% of bootstrap values and 69.3% of posterior probability values were >95% for the metazoa dataset; 94.3% of bootstrap values and 91.4% of posterior probability values were >95% for the plants dataset, respectively).

the probability that a branch is true, because true branches in a species tree can have almost any value of a CF.

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 examples in Fig. 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 toward 100% (Fig. 2). From a statistical perspective, the larger sample size increases our confidence that the concordant tree is the most common. While this is the behavior 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: our datasets are now so large that almost all such measures reach their maximum value on every branch (Kumar et al. 2012; Thomson and Brown 2022). Despite this, CFs do provide some information on the amount of data that will be needed to have high statistical support: a branch with a low CF will require commensurately more data before the 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 the split in the species tree has a CF 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 CFs 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 CF 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, as the example from birds described below will show. This is also what we mean by the predictive power of the species tree: 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, particularly when

the species tree is relatively uninformative about the majority of genes and traits (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 will 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 lies within the ingroup clade. An effective outgroup should be an outgroup on *every* gene tree, i.e. the CF 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.

The Concordance Vector: A Simple Way to Summarize Concordance and Discordance

Before we discuss how CF and DF 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 (Fig. 1c), with each internal branch of a species tree having its own concordance vector. Similar ideas to the concordance vector have been presented in numerous previous studies (Mirarab et al. 2014; Salichos et al. 2014; Minh et al. 2020) but never been formalized. By formalizing it here, we hope to facilitate future research by providing a common vocabulary.

The challenge that the concordance vector seeks to address is to provide a compact and meaningful summary of both the CF and DF associated with a single branch. This can be difficult because although the CF is a single proportion, DFs can be much more complex. There can be a vast number of ways that gene trees can be discordant with a species tree, and in principle, we could calculate a DF 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 and will be astronomical for many of them, and under most evolutionary scenarios, the expected DF for many or most topologies will be effectively zero.

The concordance vector, ψ, addresses this problem by summarizing the CF and DFs of a clade into four proportions that sum to 1. The first entry in the concordance vector, ψ_1, is simply the CF; the remaining three entries (ψ_2, ψ_3, and ψ_4) summarize the DFs. 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 Fig. 1 (equivalent to nearest-neighbor interchanges in phylogenetics). These alternative topologies are biologically important because in many scenarios, such as ILS, we expect them to be the topologies associated with the two highest DFs. 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 be the smaller of the two, as this simplifies the description and comparison of some interpretations and derivations of CFs (see below). The third DF, ψ_4, is the sum of all other DFs, 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 be equal to 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 CFs and DFs. For example, certain approaches to estimating CFs—like the site concordance factor (sCF) and quartet concordance factor (qCF)— 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 ILS 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 how to estimate various CFs.

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 CFs 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 CF 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 CFs 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 CFs to indicate the input data for each, as this helps to distinguish them from each other and from their expected values (i.e. the true but unknown concordance vectors). We denote CFs estimated from genes as gene CFs (gCFs), from quartets as qCFs, and from sites as sCFs. We name each associated concordance vector in the same way, for example gene concordance vector (gCV).

gCFs

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 gCV 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). This approach, while popular, remains fundamentally limited by the issue of gene tree estimation error. This was recognized and addressed by a Bayesian method called BUCKy (Ané et al. 2007; Larget et al. 2010), which allows the topology of each gene tree to be informed by the topologies of the other gene trees, and for gCFs and their confidence intervals to be calculated. However, this rather elegant solution has fallen out of favor, as Bayesian methods do not scale well to today's phylogenomic datasets. Another 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.

qCFs

qCFs can refer to any approach in which each alignment of an independent locus is summarized not by a single tree (as in the gCF) but by a collection of subtrees 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," "support," or "frequencies" there). They have since been more widely adopted to quantify and explore conflicting signal in multilocus 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 qCFs 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 then to extract unrooted quartets from that gene tree (i.e. the quartets relevant to the branch of interest); this approach likely 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 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 around a branch of interest to display a branch that is not either the branch of interest or that represents one of the two splits that could be induced by ILS occurring on that branch.

sCFs

sCFs were first introduced using a parsimony-based approach (Minh et al. 2020) and later updated to use maximum likelihood (Mo et al. 2023). They estimate CFs by counting proportions of individual site patterns. 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, and simply count the proportion of decisive site patterns that agree with the trees represented by ψ_1, ψ_2, or ψ_3 for the branch of interest. Similar 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 different in the two sCF methods) and 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, followed by an example highlighting many of these issues using data from a recent study of birds.

gCFs

gCFs offer not only the fullest view of genealogical variation but also come with the most caveats. In terms of information provided, not only are gCFs the only approach

that can estimate ψ_4, but they also allow us to expand the gCV beyond four entries. Recall that ψ_4 is associated with all topologies not accounted for by ψ_1, ψ_2, and ψ_3 (Fig. 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 a version of the gCV, which is calculated using splits rather than decisive gene trees. 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 of this statistic (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 $n > 3$, then we must expand our concordance vector to be of length n (since ψ_4 typically refers to many more than a single split). In essence, both measures of internode certainty tell us how likely it is that the CF of the branch of interest is higher than the next most common split we have considered. We may be able to glean quite a lot of useful information from the frequency of gene trees beyond the most common three, but unfortunately ICA is one of the only methods we know 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 gCFs is gene tree estimation error. Estimation error is unavoidable with the limited phylogenetic information available in single-locus alignments. The requirement that loci be nonrecombining 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 gCV (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) and for some of the very low gCF values shown below in the bird tree.

Broadly speaking, there are three methods in common usage for calculating gCFs from sets of gene trees, and these differ substantially in the extent to which they are affected by and/or address gene tree estimation error. BUCKy takes as input a posterior distribution of trees for each gene and then uses a form of statistical shrinkage to update the distribution of each gene tree using information from the other gene trees (Ané et al. 2007; Larget et al. 2010). This is highly effective in reducing the effects of gene tree estimation error, and the Bayesian framework also provides confidence intervals on all estimated CFs. However, the method does not scale well to large datasets and requires the user to determine the strength of statistical shrinkage *a priori*. Other widely used methods treat gene trees as entirely independent, without attempting to correct for gene tree estimation error. For example, in order for a gene tree to contribute to ψ_1, ψ_2, or ψ_3, the gene concordance method implemented in IQ-TREE2 (Minh et al. 2020) requires all four clades around the branch of interest to be monophyletic (i.e. clades *A*, *B*, *C*, and *D* must all be monophyletic in Fig. 1). This is a restrictive requirement that causes any gene tree in which *A*, *B*, *C* or *D* is not monophyletic to fall into ψ_4. In the presence of gene tree estimation error, this can lead to substantial underestimates of ψ_1 to ψ_3, likely explaining the large discrepancy between gCFs and qCFs in many of the clades in the bird example below. A less restrictive approach is to use bipartitions, such that a tree displaying the bipartition separating the taxa in *A* and *B* from those in *C* and *D* contributes to ψ_1 regardless of whether *A*, *B*, *C*, and *D* are all monophyletic. This approach is less impacted by gene tree estimation error and is used to calculate the internode certainty measure described above (Salichos et al. 2014), as well as discordance in a recent paper examining relationships among angiosperms (Zuntini et al. 2024). However, using splits rather than decisive gene trees also means that we no longer have straightforward theoretical expectations for ψ_1, ψ_2, and ψ_3.

While gene tree estimation error tends to bias the gCF downwards, it is also possible for the gCF (i.e. ψ_1) to be overestimated. 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 of signal in the data. Because ψ_1 is typically associated with the majority of 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, the approach used to define the boundaries of nonrecombining loci, and the methods used to estimate the gene trees. Regardless, gCVs should be interpreted with both underestimation and overestimation issues in mind.

qCFs

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 Fig. 1a. As mentioned before, for an unrooted quartet around a single branch of interest, there are only three possible topologies, so we only have nonzero entries for ψ_1, ψ_2, and ψ_3 in the quartet concordance vector (qCV).

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 Fig. 1 (*A*, *B*, *C*, or *D*) to be nonmonophyletic, then the method used to calculate gCFs in IQ-TREE2 would place the full gene tree in ψ_4 (Minh et al. 2020); 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. Thus, 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 upward when the true $\psi_4 > 0$. 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 to ψ_3—quartet calculations will be forced to populate these entries in the vector (see the bird example below). In addition, it is worth noting that since qCFs are ultimately based on estimated gene trees, many issues affecting gCFs (such as the misleading effects of concatalescence) will also affect qCFs.

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 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 nonindependence 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–based 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. Bootstrap methods are therefore not limited to concatenated nucleotide or amino acid alignments.

sCFs

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 gene tree from a short alignment, which is always prone to error. Therefore, as an alternative approach, one can calculate

sCFs (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, nonrecombining loci. As described earlier, they can be easily calculated from a long, concatenated alignment.

sCFs can 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. Fig. 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 due to concatenation (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, as some amount of homoplasy is unavoidable.

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:

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

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 Fig. 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 other two 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 Fig. 2.

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 an input into an assembly method (as described above for other quartet methods), giving a full species tree as an output (see a similar site pattern method in the study by 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 Fig. 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" (Fig. 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 nonindependence of nearby sites into account via block bootstrapping of genomic windows. Similarly, the Δ statistic in the study by 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).

An Example from Birds

To highlight both the calculation and interpretation of CFs and concordance vectors, we use a recent example from the genomes of 363 bird species (Stiller et al. 2024). The original paper does a very good job of inferring a species tree using multiple methods and multiple types of markers; we therefore do not reexamine species relationships here. Instead, we focus on the branches supporting six major clades of birds, particularly how measures of support differ from measures of concordance and how different measures of concordance can differ from each other. Here in the main text we present results based on gene trees inferred in the study by Stiller

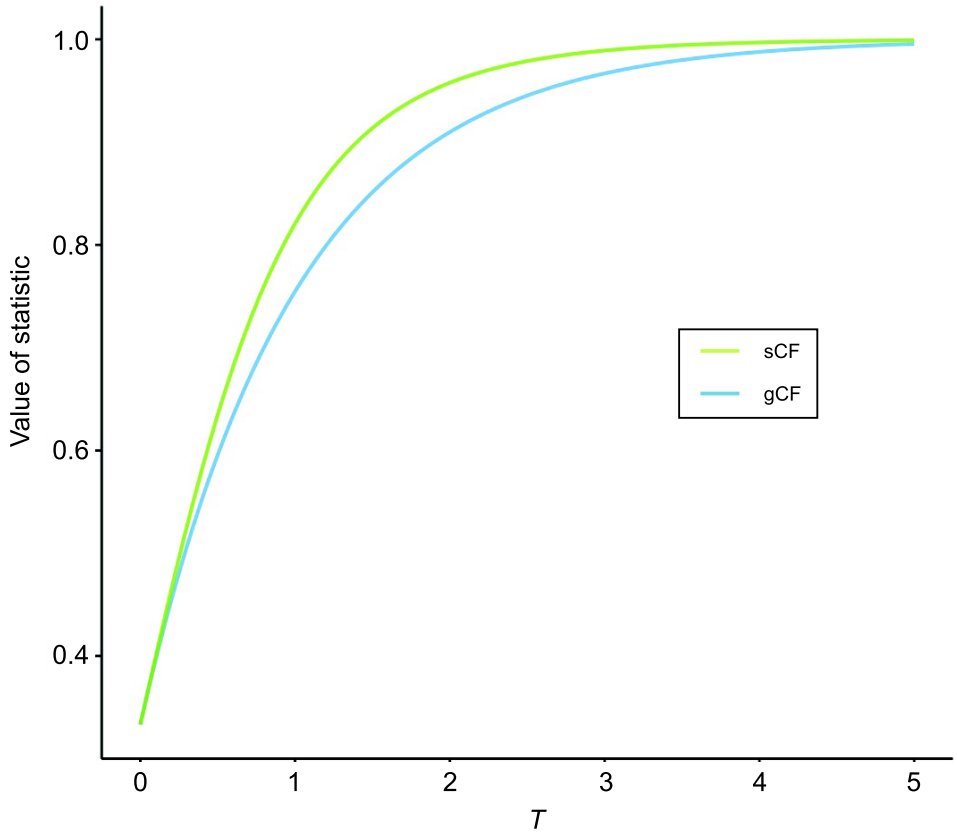

Fig. 3. The difference between expected sCFs and gCFs as the branch length of the species tree changes. Here, we plot the expected values of gCFs (equation 1; blue) and sCFs (equation 3; yellow-green) as a function of the internal branch length, *T*, for a species tree resembling the one shown in Fig.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

et al. (2024); in an online tutorial (http://www.iqtree.org/doc/recipes/concordance-vector), we show how to go from alignments to concordance vectors for the same data.

Figure 4a shows the relationships among six major clades of birds, as well as the branch lengths and posterior probabilities for the branches subtending each of them. All six of these clades have the maximum possible statistical support (a posterior probability of 1.0), but branch lengths vary from 0.01 coalescent units to 8.71. Given the species tree and the unambiguous support for each of these clades, one might erroneously conclude that all six clades are highly predictive of the evolutionary history of the genes and traits in this group. However, CFs reveal dramatic differences in the predictive power of each clade. For example, the branch leading to the Palaeognathae has very high CFs (e.g. gCF and qCF of 100% and sCF of 75.7%), while the branch leading to the Elementaves (a novel clade defined in Stiller et al. 2024) has strikingly low CFs (gCF = 0.1%, qCF = 33.9%, and sCF = 35.0%). Thus, despite both branches receiving the maximum possible value of statistical support, the estimated proportion of the genome that matches each clade differs dramatically. Indeed, the Elementaves clade was recovered in just 43 of the 63,000 gene trees estimated in the original study.

The difference observed here between support and concordance is important, because it informs debates about the extent to which each clade matters in terms of understanding the history of genes and traits in birds. For example, CFs tell us that we should expect to see many phenotypic traits common to the Palaeognathae, but that this is likely to be far less common for the Elementaves. In fact, there may not even be a single morphological character uniquely common to all members of the Elementaves. Note that none of this means that Elementaves is not a "true" clade—it is simply that the bird radiation occurred so rapidly that few genes support even the true history. The topology supporting the Elementaves still has more support than the alternatives (Fig. 4b; recall that ψ_4 in the gCV is likely made up of a huge number of topologies that do not contain the Elementaves clade).

The example clades used here also serve to emphasize how different methods for calculating CFs can provide different results. One obvious difference is between the gCF and the other two measures, especially when concordance is low. Because qCFs and sCFs define ψ_4 to be 0, the lowest possible value of ψ_1 should be 33.3 for these measures (under most circumstances). In contrast, gCFs can be arbitrarily low. As can be seen in Fig. 4b, in three clades (Telluraves, Elementaves, and Columbaves) ψ_4 has by far the largest value

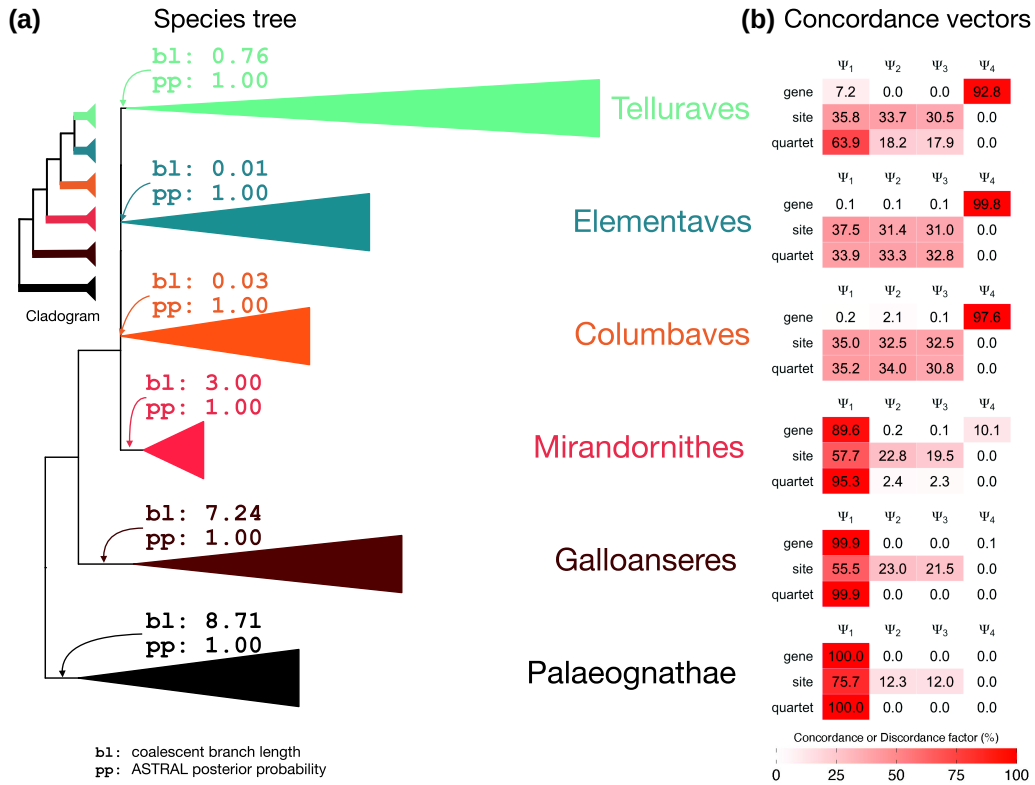

Fig. 4. Concordance vectors for six major clades of birds show dramatic differences in concordance despite maximal statistical support. a) The phylogeny shows six major clades of birds identified in a recent phylogenomic study (Stiller et al. 2024). Each clade is named and shown in a color that matches those used in the study by Stiller et al. 2024. The stem branch length of each clade (measured in coalescent units) and the posterior probability of each branch are shown. The right-most extent of each clade corresponds to the maximum root-to-tip distance of the tips in a clade (using coalescent branch lengths from ASTRAL). The inset shows a cladogram that clarifies the topology of the inferred species tree. b) Concordance vectors for each clade reveal dramatic differences in the CF and DF of each clade. Opposite each clade name is a matrix of the gene, site, and quartet concordance vectors. The numbers in each cell are percentages, with higher percentages colored in darker shades of red.

in the gCV—this means that most genes do not have a topology matching ψ_1, ψ_2, or ψ_3. Given that neighboring branches are also very short, the most likely source of the additional topologies is the nonmonophyly of the subclades that define each of the three branches in the species tree. This nonmonophyly is also the probable cause of another interesting pattern observed in the Columbaves: ψ_2 is greater than either ψ_1 or ψ_3. The explanation for this pattern was suggested to us by Megan Smith, who demonstrated that unequal rates of nonmonophyly in clades *A* and *B* (assuming a species tree similar to the one in Fig. 1) will lead to unequal numbers of nondecisive trees in the two, which will in turn lead to biased values of ψ_2 or ψ_3. In other words, the loss of decisive trees, rather than any biological process, can lead to unequal counts of minor topologies in the gCV. In this case, methods that are unaffected by nonmonophyly—namely qCF or sCF—will be much more accurate. Indeed, the lack of bias toward ψ_2 in the site and quartet concordance vectors in Fig. 4 strongly suggest that the high ψ_2 seen in the gCV results from bias, not from biology.

Another obvious pattern among the clades examined here is the consistently lower values of sCFs when concordance is high. In the Mirandornithes, Galloanseres, and Palaeognathae, ψ_1 is the highest value in the concordance vector by far, but it is appreciably lower using sCFs than the other two methods. The likely reason for this pattern is homoplasy at individual sites, especially given that the branches we are interested in span periods ~65 Ma (Stiller et al. 2024). Overall, it seems that quartet concordance vectors may offer the clearest, least-biased view of concordance and discordance in this dataset, although it should be kept in mind that the calculation of the quartet concordance vector assumes that ψ_4 is zero. In addition, qCFs are the basis for inferred branch lengths in ASTRAL used here, so these values will be most highly correlated with branch lengths inferred in coalescent units.

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, CFs 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 CFs 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 CFs 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 1,000 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% CIs 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). In the online tutorial that accompanies this paper, we have implemented an approach to calculate 95% confidence intervals on all entries in the concordance vector by bootstrapping the number of genes, sites, or quartets associated with each vector (http://www.iqtree.org/doc/recipes/concordance-vector). We hope that displaying confidence intervals alongside CFs (and other entries of the concordance vector) will help biologists to better interpret the underlying biological variation, particularly when considering biological hypotheses about the causes of such variation.

One of the biggest challenges to estimating CFs is gene tree estimation error. Gene tree estimation error affects both gCF and qCF, 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 CFs (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 CFs in large datasets remains a largely open problem (for small datasets, BuCKy is an excellent solution; Larget et al. 2010).

CFs 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 CFs 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 (Smith and Hahn 2021). 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 CFs from all of these nonstandard data types will be necessary in the near future.

Finally, visualizing 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 CF). 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 coloring 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 visualize 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. CFs are a useful tool for summarizing 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.

Acknowledgments

We thank Megan Smith, Yu Mo, Richard Wang, and Minh Bui for helpful conversations and assistance, as well as

Brandon Gaut and two reviewers for constructive comments.

Funding

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.

Conflict of Interest: The authors declare no competing interests.

Data availability

No new data produced.

References

Adams EN III. Consensus techniques and the comparison of taxonomic trees. Syst Zool. 1972:**21**(4):390–397. https://doi.org/10.2307/2412432.

Adams R, Lozano JR, Duncan M, Green J, Assis R, DeGiorgio M. A tale of too many trees: a conundrum for phylogenetic regression. bioRxiv 580530. https://doi.org/10.1101/2024.02.16.580530, 20 February 2024, preprint: not peer reviewed.

Allman ES, Degnan JH, Rhodes JA. Identifying the rooted species tree from the distribution of unrooted gene trees under the coalescent. J Math Biol. 2011:**62**(6):833–862. https://doi.org/10.1007/s00285-010-0355-7.

Allman ES, Mitchell JD, Rhodes JA. Gene tree discord, simplex plots, and statistical tests under the coalescent. Syst Biol. 2022:**71**(4):929–942. https://doi.org/10.1093/sysbio/syab008.

Ané C, Larget B, Baum DA, Smith SD, Rokas A. Bayesian estimation of concordance among gene trees. Mol Biol Evol. 2007:**24**(2):412–426. https://doi.org/10.1093/molbev/msl170.

Baum DA. Concordance trees, concordance factors, and the exploration of reticulate genealogy. Taxon. 2007:**56**(2):417–426. https://doi.org/10.1002/tax.562013.

Bouckaert RR. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics. 2010:**26**(10):1372–1373. https://doi.org/10.1093/bioinformatics/btq110.

Boussau B, Szöllosi GJ, Duret L, Gouy M, Tannier E, Daubin V. Genome-scale coestimation of species and gene trees. Genome Res. 2013:**23**(2):323–330. https://doi.org/10.1101/gr.141978.112.

Bravo GA, Antonelli A, Bacon CD, Bartoszek K, Blom MPK, Huynh S, Jones G, Knowles LL, Lamichhaney S, Marcussen T, et al. Embracing heterogeneity: coalescing the tree of life and the future of phylogenomics. PeerJ. 2019:**7**:e6399. https://doi.org/10.7717/peerj.6399.

Bryant D, Steel M. Constructing optimal trees from quartets. J. Algorithms. 2001:**38**(1):237–259. https://doi.org/10.1006/jagm.2000.1133.

Cannon JT, Vellutini BC, Smith J III, Ronquist F, Jondelius U, Hejnol A. Xenacoelomorpha is the sister group to Nephrozoa. Nature. 2016:**530**(7588):89–93. https://doi.org/10.1038/nature16520.

Carpenter JM. Choosing among multiple equally parsimonious cladograms. Cladistics. 1988:**4**(3):291–296. https://doi.org/10.1111/j.1096-0031.1988.tb00476.x.

Carruthers T, Sun M, Baker WJ, Smith SA, de Vos JM, Eiserhardt WL. The implications of incongruence between gene tree and species tree topologies for divergence time estimation. Syst Biol. 2022:**71**(5):1124–1146. https://doi.org/10.1093/sysbio/syac012.

Chifman J, Kubatko L. Quartet inference from SNP data under the coalescent model. Bioinformatics. 2014:**30**(23):3317–3324. https://doi.org/10.1093/bioinformatics/btu530.

Copetti D, Búrquez A, Bustamante E, Charboneau JLM, Childs KL, Eguiarte LE, Lee S, Liu TL, McMahon MM, Whiteman NK, et al. Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. Proc Natl Acad Sci U S A. 2017:**114**(45):12003–12008. https://doi.org/10.1073/pnas.1706367114.

Corbett-Detig RB, Russell SL, Nielsen R, Losos J. Phenotypic convergence is not mirrored at the protein level in a lizard adaptive radiation. Mol Biol Evol. 2020:**37**(6):1604–1614. https://doi.org/10.1093/molbev/msaa028.

DeGiorgio M, Degnan JH. Fast and consistent estimation of species trees using supermatrix rooted triples. Mol Biol Evol. 2010:**27**(3):552–569. https://doi.org/10.1093/molbev/msp250.

Degnan JH, Rosenberg NA. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol Evol. 2009:**24**(6):332–340. https://doi.org/10.1016/j.tree.2009.01.009.

Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. Mol Biol Evol. 2011:**28**(8):2239–2252. https://doi.org/10.1093/molbev/msr048.

Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, et al. Genomic architecture and introgression shape a butterfly radiation. Science. 2019:**366**(6465):594–599. https://doi.org/10.1126/science.aaw2090.

Edwards SV, Beerli P. Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution. 2000:**54**:1839–1854. https://doi.org/10.1111/j.0014-3820.2000.tb01231.x.

Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985:**39**(4):783–791. https://doi.org/10.2307/2408678.

Felsenstein J. Inferring phylogenies. Sunderland (MA): Sinauer Associates; 2004.

Feng S, Bai M, Rivas-González I, Li C, Liu S, Tong Y, Yang H, Chen G, Xie D, Sears KE, et al. Incomplete lineage sorting and phenotypic evolution in marsupials. Cell. 2022:**185**(10):1646–1660.e18. https://doi.org/10.1016/j.cell.2022.03.034.

Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, Jiang X, Hall AB, Catteruccia F, Kakani E,et al. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science. 2015:**347**(6217):1258524. https://doi.org/10.1126/science.1258524.

Gadagkar SR, Rosenberg MS, Kumar S. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J Exp Zool B Mol Dev Evol. 2005:**304**(1):64–74. https://doi.org/10.1002/jez.b.21026.

Gatesy J, Springer MS. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol Phylogenet Evol. 2014:**80**:231–266. https://doi.org/10.1016/j.ympev.2014.08.013.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. New York (NY): Chapman and Hall/CRC; 2013.

Good JM, Wiebe V, Albert FW, Burbano HA, Kircher M, Green RE, Halbwax M, André C, Atencia R, Fischer A, et al. Comparative population genomics of the ejaculate in humans and the great apes. Mol Biol Evol. 2013:**30**(4):964–976. https://doi.org/10.1093/molbev/mst005.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, et al. A draft sequence of the Neandertal genome. Science. 2010:**328**(5979):710–722. https://doi.org/10.1126/science.1188021.

Guerrero RF, Hahn MW. Quantifying the risk of hemiplasy in phylogenetic inference. Proc Natl Acad Sci U S A. 2018:**115**(50):12787–12792. https://doi.org/10.1073/pnas.1811268115.

Hahn MW. Bias in phylogenetic tree reconciliation methods: implications for vertebrate genome evolution. Genome Biol. 2007:**8**(7):R141. https://doi.org/10.1186/gb-2007-8-7-r141.

Hahn MW, Nakhleh L. Irrational exuberance for resolved species trees. Evolution. 2016:**70**(1):7–17. https://doi.org/10.1111/evo.12832.

Han F, Lamichhaney S, Grant BR, Grant PR, Andersson L, Webster MT. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. Genome Res. 2017:**27**(6):1–12. https://doi.org/10.1101/gr.212522.116.

Hibbins MS, Breithaupt LC, Hahn MW. Phylogenomic comparative methods: accurate evolutionary inferences in the presence of gene tree discordance. Proc Natl Acad Sci U S A. 2023:**120**(22):e2220389120. https://doi.org/10.1073/pnas.2220389120.

Hibbins MS, Gibson MJ, Hahn MW. Determining the probability of hemiplasy in the presence of incomplete lineage sorting and introgression. Elife. 2020:**9**:e63753. https://doi.org/10.7554/eLife.63753.

Hillis DM. Molecular versus morphological approaches to systematics. Annu Rev Ecol Syst. 1987:**18**(1):23–42. https://doi.org/10.1146/annurev.es.18.110187.000323.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2017:**35**(2):518–522. https://doi.org/10.1093/molbev/msx281.

Hudson RR. Testing the constant-rate neutral allele model with protein sequence data. Evolution. 1983:**37**(1):203–217. https://doi.org/10.2307/2408186.

Huson DH. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics. 1998:**14**(1):68–73. https://doi.org/10.1093/bioinformatics/14.1.68.

Huson DH, Klöpper T, Lockhart PJ, Steel MA. 2005. Reconstruction of reticulate networks from gene trees. In: Miyano S, Mesirov J, Kasif S, Istrail S, Pevzner PA, Waterman M, editors. Research in computational molecular biology. Vol. 3500. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. p. 233–249. http://link.springer.com/10.1007/11415770_18

Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol. 2012:**61**(6):1061–1067. https://doi.org/10.1093/sysbio/sys062.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 2014:**346**(6215):1320–1331. https://doi.org/10.1126/science.1253451.

Kubatko LS, Degnan JH. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst Biol. 2007:**56**(1):17–24. https://doi.org/10.1080/10635150601146041.

Kück P, Romahn J, Meusemann K. Pitfalls of the site-concordance factor (sCF) as measure of phylogenetic branch support. NAR Genom Bioinform. 2022:**4**(3):lqac064. https://doi.org/10.1093/nargab/lqac064.

Kumar S, Filipski AJ, Battistuzzi FU, Kosakovsky Pond SL, Tamura K. Statistics and truth in phylogenomics. Mol Biol Evol. 2012:**29**(2):457–472. https://doi.org/10.1093/molbev/msr202.

Lamichhaney S, Han F, Berglund J, Wang C, Almén MS, Webster MT, Grant BR, Grant PR, Andersson L. A beak size locus in Darwin's finches facilitated character displacement during a drought. Science. 2016:**352**(6284):470–474. https://doi.org/10.1126/science.aad8786.

Larget BR, Kotha SK, Dewey CN, Ané C. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics. 2010:**26**(22):2910–2911. https://doi.org/10.1093/bioinformatics/btq539.

Larson DA, Staton ME, Kapoor B, Islam-Faridi N, Zhebentyayeva T, Fan S, Stork J, Thomas A, Ahmed AS, Stanton EC, et al. A haplotype-resolved reference genome of *Quercus alba* sheds light on the evolutionary history of oaks. bioRxiv 579671. https://doi.org/10.1101/2024.02.13.579671, 14 February 2024, preprint: not peer reviewed.

Lemoine F, Domelevo Entfellner JB, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, Gascuel O. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature. 2018:**556**(7702):452–456. https://doi.org/10.1038/s41586-018-0043-0.

Li Y-Y, Liu Z, Qi F-Y, Zhou X, Shi P. Functional effects of a retained ancestral polymorphism in *Prestin*. Mol Biol Evol. 2016:**34**(1):88–92. https://doi.org/10.1093/molbev/msw222.

Liu L, Edwards SV. Phylogenetic analysis in the anomaly zone. Syst Biol. 2009:**58**(4):452–460. https://doi.org/10.1093/sysbio/syp034.

Liu L, Yu L, Edwards SV. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol Biol. 2010:**10**(1):302. https://doi.org/10.1186/1471-2148-10-302.

Lutteropp S, Kozlov AM, Stamatakis A. A fast and memory-efficient implementation of the transfer bootstrap. Bioinformatics. 2020:**36**(7):2280–2281. https://doi.org/10.1093/bioinformatics/btz874.

Maddison WP. Gene trees in species trees. Syst Biol. 1997:**46**(3):523–536. https://doi.org/10.1093/sysbio/46.3.523.

Mendes FK, Fuentes-González JA, Schraiber JG, Hahn MW. A multispecies coalescent model for quantitative traits. Elife. 2018:**7**:e36482. https://doi.org/10.7554/eLife.36482.

Mendes FK, Hahn MW. Gene tree discordance causes apparent substitution rate variation. Syst Biol. 2016:**65**(4):711–721. https://doi.org/10.1093/sysbio/syw018.

Mendes FK, Hahn MW. Why concatenation fails near the anomaly zone. Syst Biol. 2018:**67**(1):158–169. https://doi.org/10.1093/sysbio/syx063.

Mendes FK, Hahn Y, Hahn MW. Gene tree discordance can generate patterns of diminishing convergence over time. Mol Biol Evol. 2016:**33**(12):3299–3307. https://doi.org/10.1093/molbev/msw197.

Mendes FK, Livera AP, Hahn MW. The perils of intralocus recombination for inferences of molecular convergence. Philos Trans R Soc Lond B Biol Sci. 2019:**374**(1777):20180244. https://doi.org/10.1098/rstb.2018.0244.

Minh BQ, Hahn MW, Lanfear R. New methods to calculate concordance factors for phylogenomic datasets. Mol Biol Evol. 2020:**37**(9):2727–2733. https://doi.org/10.1093/molbev/msaa106.

Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013:**30**(5):1188–1195. https://doi.org/10.1093/molbev/mst024.

Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics. 2014:**30**(17):i541–i548. https://doi.org/10.1093/bioinformatics/btu462.

Mirarab S, Warnow T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 2015:**31**(12):i44–i52. https://doi.org/10.1093/bioinformatics/btv234.

Mishra S, Smith ML, Hahn MW. reconcILS: a gene tree-species tree reconciliation algorithm that allows for incomplete lineage sorting. bioRxiv 565544. https://doi.org/10.1101/2023.11.03.565544, 5 November 2023, preprint: not peer reviewed.

Mo YK, Lanfear R, Hahn MW, Minh BQ. Updated site concordance factors minimize effects of homoplasy and taxon sampling. Bioinformatics. 2023:**39**(1):btac741. https://doi.org/10.1093/bioinformatics/btac741.

Molloy EK, Warnow T. To include or not to include: the impact of gene filtering on species tree estimation methods. Syst Biol. 2018:**67**(2):285–303. https://doi.org/10.1093/sysbio/syx077.

Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, Amon J, Arcà B, Arensburger P, Artemov G, et al. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. Science. 2015:**347**(6217):1258522. https://doi.org/10.1126/science.1258522.

Palesch D, Bosinger SE, Tharp GK, Vanderford TH, Paiardini M, Chahroudi A, Johnson ZP, Kirchhoff F, Hahn BH, Norgren RB, et al. Sooty mangabey genome sequence provides insight into AIDS resistance in a natural SIV host. Nature. 2018:**553**(7686):77–81. https://doi.org/10.1038/nature25140.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. Ancient admixture in human history. Genetics. 2012:**192**(3):1065–1093. https://doi.org/10.1534/genetics.112.145037.

Pease JB, Haak DC, Hahn MW, Moyle LC. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biol. 2016:**14**(2):e1002379. https://doi.org/10.1371/journal.pbio.1002379.

Ran J-H, Shen T-T, Wang M-M, Wang X-Q. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. Proc Biol Sci. 2018:**285**(1881):20181012. https://doi.org/10.1098/rspb.2018.1012.

Rannala B, Yang Z. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J Mol Evol. 1996:**43**(3):304–311. https://doi.org/10.1007/BF02338839.

Rhodes JA, Baños H, Mitchell JD, Allman ES. MSCquartets 1.0: quartet methods for species trees and networks under the multispecies coalescent model in R. Bioinformatics. 2021:**37**(12):1766–1768. https://doi.org/10.1093/bioinformatics/btaa868.

Rosenzweig BK, Kern AD, Hahn MW. Accurate detection of incomplete lineage sorting via supervised machine learning. bioRxiv 515828. https://doi.org/10.1101/2022.11.09.515828, 10 November 2022, preprint: not peer reviewed.

Salichos L, Rokas A. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature. 2013:**497**(7449):327–331. https://doi.org/10.1038/nature12130.

Salichos L, Stamatakis A, Rokas A. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. Mol Biol Evol. 2014:**31**(5):1261–1271. https://doi.org/10.1093/molbev/msu061.

Sayyari E, Mirarab S. Fast coalescent-based computation of local branch support from quartet frequencies. Mol Biol Evol. 2016:**33**(7):1654–1668. https://doi.org/10.1093/molbev/msw079.

Schliep KP. Phangorn: phylogenetic analysis in R. Bioinformatics. 2011:**27**(4):592–593. https://doi.org/10.1093/bioinformatics/btq706.

Schrago CG, Seuánez HN. Large ancestral effective population size explains the difficult phylogenetic placement of owl monkeys. Am J Primatol. 2019:**81**(3):e22955. https://doi.org/10.1002/ajp.22955.

Schraiber JG, Edge MD, Pennell M. Unifying approaches from statistical genetics and phylogenetics for mapping phenotypes in structured populations. PLOS Biology. 2024:**22**(10):e3002847.

Shen X-X, Hittinger CT, Rokas A. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat Ecol Evol. 2017:**1**(5):0126. https://doi.org/10.1038/s41559-017-0126.

Simon C. An evolving view of phylogenetic support. Syst Biol. 2022:**71**(4):921–928. https://doi.org/10.1093/sysbio/syaa068.

Slatkin M, Pollack JL. Subdivision in an ancestral species creates asymmetry in gene trees. Mol Biol Evol. 2008:**25**(10):2241–2246. https://doi.org/10.1093/molbev/msn172.

Smith ML, Hahn MW. New approaches for inferring phylogenies in the presence of paralogs. Trends Genet. 2021:**37**(2):156–169. https://doi.org/10.1016/j.tig.2020.08.012.

Smith ML, Vanderpool D, Hahn MW. Using all gene families vastly expands data available for phylogenomic inference. Mol Biol Evol. 2022:**39**(6):msac112. https://doi.org/10.1093/molbev/msac112.

Snir S, Rao S. Quartets MaxCut: a divide and conquer quartets algorithm. IEEE/ACM Trans Comput Biol Bioinform. 2010:**7**(4):704–718. https://doi.org/10.1109/TCBB.2008.133.

Solís-Lemus C, Ané C. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLoS Genet. 2016:**12**(3):e1005896. https://doi.org/10.1371/journal.pgen.1005896.

Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web servers. Syst Biol. 2008:**57**(5):758–771. https://doi.org/10.1080/10635150802429642.

Stefanović S, Rice DW, Palmer JD. Long branch attraction, taxon sampling, and the earliest angiosperms: Amborella or monocots? BMC Evol Biol. 2004:**4**(1):35. https://doi.org/10.1186/1471-2148-4-35.

Stiller J, Feng S, Chowdhury A-A, Rivas-González I, Duchêne DA, Fang Q, Deng Y, Kozlov A, Stamatakis A, Claramunt S, et al. Complexity of avian evolution revealed by family-level genomes. Nature. 2024:**629**(8013):851–860. https://doi.org/10.1038/s41586-024-07323-1.

Strimmer K, von Haeseler A. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol Biol Evol. 1996:**13**(7):964–964. https://doi.org/10.1093/oxfordjournals.molbev.a025664.

Sun C, Huang J, Wang Y, Zhao X, Su L, Thomas GWC, Zhao M, Zhang X, Jungreis I, Kellis M, et al. Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits. Mol Biol Evol. 2021:**38**(2):486–501. https://doi.org/10.1093/molbev/msaa240.

Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell P, Lang M, Courtier-Orgogozo V, et al. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. Curr Biol. 2022:**32**(1):111–123.e5. https://doi.org/10.1016/j.cub.2021.10.052.

Swofford DL, Kubatko LS. Species tree estimation using site pattern frequencies. In: Kubatko L, Knowles LL, editors. Species tree inference. Princeton (NJ): Princeton University Press; 2023. p. 68–88.

Thomas GWC, Hughes JJ, Kumon T, Berv JS, Erik Nordgren C, Lampson M, Levine M, Searle JB, Good JM. The genomic landscape, causes, and consequences of extensive phylogenomic discordance in old world mice and rats. bioRxiv 555178. https://doi.org/10.1101/2023.08.28.555178, 29 August 2023, preprint: not peer reviewed.

Thomson RC, Brown JM. On the need for new measures of phylogenomic support. Syst Biol. 2022:**71**(4):917–920. https://doi.org/10.1093/sysbio/syac002.

Urban S, Nater A, Meyer A, Kratochwil CF. Different sources of allelic variation drove repeated color pattern divergence in cichlid fishes. Mol Biol Evol. 2021:**38**(2):465–477. https://doi.org/10.1093/molbev/msaa237.

Vanderpool D, Minh BQ, Lanfear R, Hughes D, Murali S, Harris RA, Raveendran M, Muzny DM, Hibbins MS, Williamson RJ, et al. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. PLoS Biol. 2020:**18**(12):e3000954. https://doi.org/10.1371/journal.pbio.3000954.

Wen D, Yu Y, Zhu J, Nakhleh L. Inferring phylogenetic networks using PhyloNet. Syst Biol. 2018:**67**(4):735–740. https://doi.org/10.1093/sysbio/syy015.

Wu M, Kostyun JL, Hahn MW, Moyle LC. Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. Mol Ecol. 2018:**27**(16):3301–3316. https://doi.org/10.1111/mec.14780.

Yan H, Hu Z, Thomas GWC, Edwards SV, Sackton TB, Liu JS. PhyloAcc-GT: a Bayesian method for inferring patterns of substitution rate shifts on targeted lineages accounting for gene tree discordance. Mol Biol Evol. 2023:**40**(9):msad195. https://doi.org/10.1093/molbev/msad195.

Zhang C, Mirarab S. Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. Mol Biol Evol. 2022:**39**(12):msac215. https://doi.org/10.1093/molbev/msac215.

Zhang C, Nielsen R, Mirarab S. CASTER: direct species tree inference from whole-genome alignments. bioRxiv 560884. https://doi.org/10.1101/2023.10.04.560884, 6 October 2023, preprint: not peer reviewed.

Zhang C, Ogilvie HA, Drummond AJ, Stadler T. Bayesian inference of species networks from multilocus sequence data. Mol Biol Evol. 2018b:**35**(2):504–517. https://doi.org/10.1093/molbev/msx307.

Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 2018a:**19**(S6):153. https://doi.org/10.1186/s12859-018-2129-y.

Zhou X, Lutteropp S, Czech L, Stamatakis A, Looz MV, Rokas A. Quartet-based computations of internode certainty provide robust measures of phylogenetic incongruence. Syst Biol. 2020:**69**(2):308–324. https://doi.org/10.1093/sysbio/syz058.

Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, Epitawalage N, Françoso E, Gallego-Paramo B, McGinnie C, et al. Phylogenomics and the rise of the angiosperms. Nature. 2024:**629**(8013):843–850. https://doi.org/10.1038/s41586-024-07324-0.