Novel Information Theory-Based Measures for Quantifying Incongruence among Phylogenetic Trees

Leonidas Salichos,[1] Alexandros Stamatakis,[2,3] and Antonis Rokas*[,1,4]

[1]Department of Biological Sciences, Vanderbilt University

[2]The Exelixis Lab, Scientific Computing Group, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

[3]Institute for Theoretical Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany

[4]Department of Biomedical Informatics, Vanderbilt University Medical Center

***Corresponding author:** E-mail: antonis.rokas@vanderbilt.edu.

Associate editor: Todd Oakley

Abstract

Phylogenies inferred from different data matrices often conflict with each other necessitating the development of measures that quantify this incongruence. Here, we introduce novel measures that use information theory to quantify the degree of conflict or incongruence among all nontrivial bipartitions present in a set of trees. The first measure, internode certainty (IC), calculates the degree of certainty for a given internode by considering the frequency of the bipartition defined by the internode (internal branch) in a given set of trees jointly with that of the most prevalent conflicting bipartition in the same tree set. The second measure, IC All (ICA), calculates the degree of certainty for a given internode by considering the frequency of the bipartition defined by the internode in a given set of trees in conjunction with that of all conflicting bipartitions in the same underlying tree set. Finally, the tree certainty (TC) and TC All (TCA) measures are the sum of IC and ICA values across all internodes of a phylogeny, respectively. IC, ICA, TC, and TCA can be calculated from different types of data that contain nontrivial bipartitions, including from bootstrap replicate trees to gene trees or individual characters. Given a set of phylogenetic trees, the IC and ICA values of a given internode reflect its specific degree of incongruence, and the TC and TCA values describe the global degree of incongruence between trees in the set. All four measures are implemented and freely available in version 8.0.0 and subsequent versions of the widely used program RAxML.

Key words: **internode certainty, bipartition, split, clade support, rare genomic changes, RAxML.**

Introduction

Phylogenetic trees constructed from different genes frequently contradict each other, giving rise to incongruence (Rokas et al. 2003; Rokas and Chatzimanolis 2008). For example, several recent studies examining hundreds of genes in fungi (Hess and Goldman 2011; Salichos and Rokas 2013), plants (Zhong et al. 2013), and mammals (Song et al. 2012) found that the vast majority of gene trees are not topologically congruent either with each other or with the species phylogeny. This incongruence can be due to analytical factors stemming from either inadequate sample sizes (Bull et al. 1993; Cummings et al. 1995; Rokas et al. 2003) or the misfit between data and evolutionary models (Swofford et al. 1996; Kumar et al. 2012) or due to biological factors, such as horizontal gene transfer, lineage sorting, introgression, and hybridization (Pamilo and Nei 1988; Maddison 1997; Slowinski and Page 1999; Degnan and Rosenberg 2009).

Although the challenge of detecting and appropriately handling incongruence has vexed systematists for decades (Bull et al. 1993; Huelsenbeck et al. 1996; Cunningham 1997), the recent realization that a large number of gene trees will typically disagree with the species phylogeny has highlighted the importance and value of measures that capture and quantify incongruence (Salichos and Rokas 2013). Incongruence tests can be broadly classified (Planet 2006) into tests that assess incongruence between characters (Wilson 1965; Le Quesne 1969; Templeton 1983; Kishino and Hasegawa 1989; Faith 1991; Farris et al. 1994; Baker and DeSalle 1997; Shimodaira and Hasegawa 1999; Goldman et al. 2000) and tests that assess incongruence between trees (Rodrigo et al. 1993; Thorley and Wilkinson 1999; Thorley and Page 2000). Note that both character-based and tree-based incongruence tests rely on phylogenetic trees; however, in character-based tests, the assessment of incongruence is focused on the differences between how the distinct data sets fit the trees, whereas in tree-based tests, the assessment of incongruence focuses on the difference between the trees (Planet 2006). For example, the character-based measure developed by Shimodaira and Hasegawa (1999) relies on bootstrap resampling of characters to identify whether any one or more of a set of trees best explains the data, whereas Rodrigo's topology-based measure relies on the distribution of tree distances among bootstrap replicate trees to examine the degree of incongruence between sets of characters (Rodrigo et al. 1993). Although several of these measures are extremely useful in practice, they frequently lack generality because they depend on a particular optimality

Article

criterion (Templeton 1983; Farris et al. 1994; Baker and DeSalle 1997) or clade support measure (Rodrigo et al. 1993; Shimodaira and Hasegawa 1999).

A particularly interesting group of tree-based methods for handling incongruence and summarizing conflict are consensus methods (Bryant 2003). Because each internode (or internal branch) in a phylogenetic tree represents a bipartition that separates two sets of taxa (e.g., fig. 1 shows a bipartition *a, b, c, d, e | f, g, h, i, j* that divides the internode between nodes 1 and 5 into taxon sets {*a, b, c, d, e*} and {*f, g, h, i, j*}), a set of trees can be effectively summarized into a consensus tree that depicts only those bipartitions that are "representative" of the set. For example, the majority-rule consensus (MRC) approach (Bryant 2003) calculates the shared bipartitions across all trees in a set and displays only those shared by the majority of trees. Consequently, each internode in the MRC tree has a value that corresponds to either the number or percentage of individual phylogenetic trees that contain the bipartitions created by splitting up the tree at this internode.

Although consensus methods have been extremely useful and very popular in summarizing agreement and incongruence, they do not provide information on the next most prevalent conflicting bipartition, or more generally, on the distribution of conflicting bipartitions. For example, when an MRC tree reports that 51 out of 100 phylogenetic trees contain a specific bipartition, whether the second most prevalent yet conflicting bipartition is supported by the remaining 49 phylogenetic trees or by only five of these is not known. Information about the distribution of conflicting bipartitions, however, can be informative because the first type of conflict in the previous example (51% vs. 49%) shows that both bipartitions receive almost identical support, whereas the second type (51% vs. 5%) suggests that the first bipartition represents the sole strongly supported bipartition. Although phylogenetic inference programs typically report the distribution of bipartitions from a set of trees, including those that do not appear in the MRC tree (Felsenstein 1993; Swofford 2002), and several methods have been developed to visualize the phylogenetic conflict on each internode (Lento et al. 1995; Huson and Bryant 2006; Huson et al. 2010), measures that also incorporate conflicting bipartitions to quantify incongruence have so far been lacking.

We introduce four related measures that, given a set of trees or characters defining bipartitions, can be used to quantify the degree of incongruence for a given internode, or for an entire tree. The quantification of incongruence or conflict in all four measures is based on Shannon's entropy, a common uncertainty measure for a random variable (Shannon 1948). The first two measures, internode certainty (IC) and IC All (ICA), quantify the degree of certainty for each individual internode by considering the two most prevalent conflicting bipartitions (IC) or all most prevalent conflicting bipartitions (ICA), by providing the log magnitude of their difference. The other two measures, tree certainty (TC) and TC All (TCA), are the sums of IC and ICA values, respectively, over all internodes in a phylogeny. In this study, we present the theory of the four measures and illustrate by example how they can be applied to different types of data and biological questions. Finally, we describe how they have been implemented in the widely used program RAxML.

Four Novel Measures That Use Information Theory to Quantify Incongruence

Phylogenetic trees that represent evolutionary relationships among different genes or taxa are acyclic connected graphs that consist of nodes connected by edges or branches. Each internal branch (or internode) in a phylogenetic tree can also be represented as a bipartition or split that divides the taxa into two disjoint partitions (fig. 1). Therefore, any measure that quantifies internode support will also represent the support for the given bipartition. By considering each internode as a bipartition, any unrooted fully bifurcating phylogenetic tree with k taxa will contain $k-3$ nontrivial bipartitions

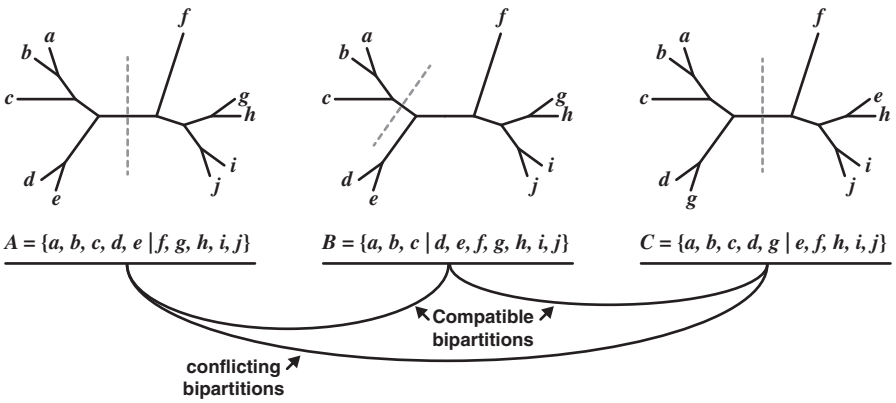

Fig. 1. Compatible and conflicting bipartitions. Bipartition $A = \{a, b, c, d, e \mid f, g, h, i, j\}$ is composed of the partitions $A_1 = \{a, b, c, d, e\}$ and $A_2 = \{f, g, h, i, j\}$, where *a, b, c, d, e, f, g, h, i,* and *j* are taxa. Bipartition $B = \{a, b, c \mid d, e, f, g, h, i, j\}$ is composed of the partitions $B_1 = \{a, b, c\}$ and $B_2 = \{d, e, f, g, h, i, j\}$, and bipartition $C = \{a, b, c, d, g \mid e, f, h, i, j\}$ is composed of the partitions $C_1 = \{a, b, c, d, g\}$ and $C_2 = \{e, f, h, i, j\}$. Bipartitions *A* and *B* are compatible because one of the intersections of their bipartition pairs ($A_2 \cap B_1$) is empty. Bipartitions *B* and *C* are compatible for the same reason ($B_1 \cap C_2$ is empty). In contrast, bipartition *C* conflicts or is incompatible with bipartition *A* because none of the four intersections ($A_1 \cap C_1$, $A_1 \cap C_2$, $A_2 \cap C_1$, $A_2 \cap C_2$) is empty.

(i.e., $k-3$ bipartitions, each of which divides the $k = m + n$ taxa in the tree into two partitions of m and n taxa, respectively, where $m \geq 2$ and $n \geq 2$). If two phylogenetic trees with the same number of taxa k are topologically identical, then the total number of unique nontrivial bipartitions is still only $k-3$ because the union of the set of bipartitions induced by this second tree with the set of bipartitions induced by the first shows that there are no unique nontrivial bipartitions that are only present in one tree but absent from the other. In contrast, if two phylogenetic trees are incongruent, then the set of phylogenetic trees will contain more than $k-3$ bipartitions, where each of the additional bipartitions represent bipartitions that conflict with one or more of the $k-3$ bipartitions.

Compatible and Conflicting Bipartitions

Two bipartitions $A = X_1 \mid X_2$ and $B = Y_1 \mid Y_2$ from the same taxon set are "compatible" if and only if at least one of the intersections of the four bipartition pairs ($X_1 \cap Y_1$, $X_1 \cap Y_2$, $X_2 \cap Y_1$, $X_2 \cap Y_2$) is empty (Bryant 2003; Huson et al. 2010). If this condition is not met, then the bipartitions are said to be "incompatible or incongruent" or to "conflict" with one another.

Example. Let us consider the bipartition $A = \{a, b, c, d, e \mid f, g, h, i, j\}$, composed of the partitions $A_1 = \{a, b, c, d, e\}$ and $A_2 = \{f, g, h, i, j\}$, where a, b, c, d, e, f, g, h, i, and j are taxon names. Let us also consider a second bipartition from the same set of taxa $B = \{a, b, c \mid d, e, f, g, h, i, j\}$, composed of the partitions $B_1 = \{a, b, c\}$ and $B_2 = \{d, e, f, g, h, i, j\}$ (fig. 1). Bipartition B does not conflict with bipartition A because $A_2 \cap B_1$ is empty. In contrast, bipartition $C = \{a, b, c, d, g \mid e, f, h, i, j\}$, composed of the partitions $C_1 = \{a, b, c, d, g\}$ and $C_2 = \{e, f, h, i, j\}$, conflicts or is incompatible with bipartition A because none of the four intersections ($A_1 \cap C_1$, $A_1 \cap C_2$, $A_2 \cap C_1$, $A_2 \cap C_2$) is empty (fig. 1).

Shannon's Entropy and *IC*

Shannon's entropy measures the amount of uncertainty in random variables (Shannon 1948). For two equally probable events, for example, "head or tails" in a fair coin toss, the amount of uncertainty is equal to 1. However, if the coin is not fair, the uncertainty of the outcome decreases proportionally to the coin's "unfairness." In general, for a random variable X with a set of n possible values $\{X_1, X_2, \ldots, X_n\}$, Shannon's entropy $H(X)$ is defined as

$$H(X) = -\sum_{n=1}^{n} P(X_n) \log[P(X_n)],$$

where $P(X_n)$ is the probability of outcome X_n. In its simplest form, if variable X consists of only two possible outcomes X_1 and X_2, Shannon's entropy is equal to

$$H(X) = -\sum_{n=1}^{2} P(X_n) \log_2[P(X_n)].$$

In phylogenetics, let us consider variable $H(X)$ as the entropy that measures the amount of uncertainty of support for a given internode with the set of possible values being the values of the two most prevalent conflicting bipartitions ($n = 2$) for that internode (i.e., $X = \{X_1, X_2\}$), with X_1 being the frequency of support for the bipartition that defines the internode. For these two bipartitions X_1 and X_2, we define $H(X)$ as the internode uncertainty:

$$\begin{aligned}\text{Internode uncertainty} &= -\sum_{n=1}^{2} P(X_n) \log_2[P(X_n)] \\ &= -\{P(X_1) \log_2[P(X_1)] + P(X_2) \log_2[P(X_2)]\},\end{aligned}$$

where $P(X_1) = X_1/(X_1 + X_2)$, $P(X_2) = X_2/(X_1 + X_2)$, and $P(X_1) + P(X_2) = 1$.

Because internode support measures typically quantify the degree of support for a given internode, rather than the lack thereof, we reverse the sign of the equation and add $\log_2(n)$ to it so that the measure corresponds to "certainty" rather than "uncertainty." Thus, we define IC as

$$\begin{aligned}\text{IC} &= \log_2(n) + \sum_{n=1}^{2} P(X_n) \log_2[P(X_n)] \\ &= 1 + P(X_1) \log_2[P(X_1)] + P(X_2) \log_2[P(X_2)],\end{aligned}$$

where $P(X_1) = X_1/(X_1 + X_2)$, $P(X_2) = X_2/(X_1 + X_2)$, and $P(X_1) + P(X_2) = 1$.

For a given internode, IC values correspond to the magnitude of conflict between the bipartition that defines the internode and the most prevalent conflicting bipartition in the given tree set. For example, IC values at or close to 1 indicate the absence of conflict for the bipartition defined by a given internode, whereas IC values at or close to 0 indicate equal support for both bipartitions and hence maximum conflict.

So far, we have assumed that the frequency of the bipartition that defines the internode is equal or higher than the frequency of the most prevalent bipartition, that is, $P(X_1) \geq P(X_2)$. However, in some cases, it may happen that we need to calculate the IC of an internode that was included in the consensus tree (depending on the type of consensus tree constructed, see below), whose bipartition frequency is actually smaller than the frequency of a conflicting bipartition, that is, $P(X_1) \leq P(X_2)$. To distinguish between cases where $P(X_1) \geq P(X_2)$ from cases where $P(X_1) \leq P(X_2)$, we reverse the sign of the IC value for all cases where $P(X_1) \leq P(X_2)$. Thus, negative IC values indicate that the internode of interest conflicts with a bipartition that has a higher frequency, and IC values at or close to -1 indicate an almost complete absence of support for the bipartition defined by the given internode and an almost absolute support for the conflicting bipartition. The behavior of the IC measure for a range of different values of X_1 and X_2 is shown in figure 2.

Examples. Let us consider a set of 100 gene trees from which 62 gene trees support bipartition X_1 (which appears on the MRC tree), and 6 gene trees support the conflicting bipartition X_2 (which does not appear on the MRC tree). In this case,

$$P(X_1) = X_1/(X_1 + X_2) = 62/(62 + 6) = 0.91, \text{ and}$$
$$P(X_2) = X_2/(X_1 + X_2) = 6/(62 + 6) = 0.09.$$

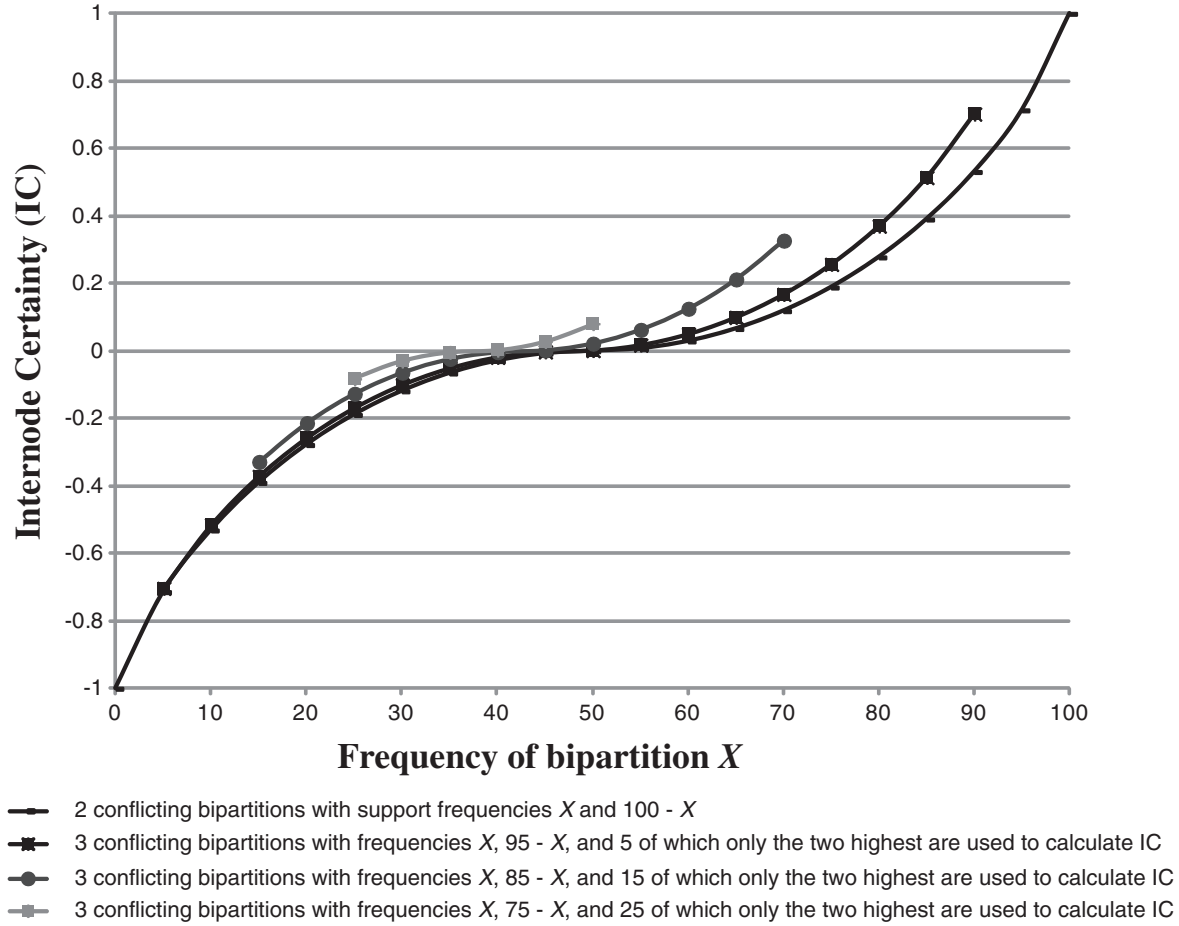

FIG. 2. Visualizing IC for the two most prevalent conflicting bipartitions of a given internode. The default curve represents the case of only two conflicting bipartitions for one internode (only two partitions: {*X*, 100–*X*}). Out of 100 total trees, when 60 trees recover the first bipartition, the remaining 40 will support the second and conflicting bipartition. In the presence of three conflicting bipartitions for a given internode (e.g., {65, 30, 5}), when the two most prevalent bipartitions are considered, the percentage of trees supporting the first bipartition is equal to 65/(65 + 30), whereas the percentage of trees supporting the second conflicting bipartition is equal to 30/(65 + 30). The reason that we do not include the number of trees containing the third bipartition is that we want IC to measure the magnitude of certainty conveyed by the two most prevalent bipartitions. This way, IC will be zero when the two most prevalent conflicting bipartitions have equal frequencies.

Thus,

$$\begin{aligned} \text{IC} &= 1 + P(X_1)\log_2[P(X_1)] + P(X_2)\log_2[P(X_2)] \\ &= 1 + 0.91 \times \log_2(0.91) + 0.09 \times \log_2(0.09) = 0.57. \end{aligned}$$

If $X_1 = 52$ gene trees and the conflicting bipartition $X_2 = 29$ gene trees, then

$$P(X_1) = X_1/(X_1 + X_2) = 52/(52 + 29) = 0.64, \text{ and}$$
$$P(X_2) = X_2/(X_1 + X_2) = 29/(52 + 29) = 0.36.$$

Thus,

$$\begin{aligned} \text{IC} &= 1 + P(X_1)\log_2[P(X_1)] + P(X_2)\log_2[P(X_2)] \\ &= 1 + 0.64 \times \log_2(0.64) + 0.36 \times \log_2(0.36) = 0.06. \end{aligned}$$

Finally, if an internode is defined by a bipartition X_1 supported by five gene trees and the conflicting bipartition X_2 is support by 55 gene trees, then

$$P(X_1) = X_1/(X_1 + X_2) = 5/(5 + 55) = 0.08, \text{ and}$$
$$P(X_2) = X_2/(X_1 + X_2) = 55/(5 + 55) = 0.92.$$

Thus,

$$\begin{aligned} \text{IC} &= 1 + P(X_1)\log_2[P(X_1)] + P(X_2)\log_2[P(X_2)] \\ &= 1 + 0.08 \times \log_2(0.08) + 0.92 \times \log_2(0.92) = 0.59. \end{aligned}$$

However, because $P(X_1) \leq P(X_2)$, the sign of the IC value is set to -0.59.

Extending IC to Include All Prevalent Conflicting Bipartitions

The IC can be extended to consider all n prevalent conflicting bipartitions for a given internode, that is ($X = \{X_1, X_2, \ldots, X_n\}$). This measure, which we name ICA, can be calculated using

$$\begin{aligned} \text{ICA} = \log_n(n) &+ P(X_1)\log_n[P(X_1)] + P(X_1)\log_n[P(X_1)] \\ &+ \ldots + P(X_n)\log_n[P(X_n)], \end{aligned}$$

where $P(X_1) = X_1/(X_1 + X_2 + \ldots + X_n)$, $P(X_2) = X_2/(X_1 + X_2 + \ldots + X_n), \ldots,$ $P(X_n) = X_n/(X_1 + X_2 + \ldots + X_n)$ and $P(X_1) + P(X_2) + \ldots + P(X_n) = 1$.

Because the number of bipartitions that conflict with a given internode in large phylogenetic tree sets can be high, as well as because conflicting bipartitions whose frequency is very low have little impact on the certainty value of a given internode, we restrict the ICA to consider only bipartitions whose frequency is $\geq 5\%$ because this represents a reasonable trade-off between speed and accuracy. To distinguish between cases where $P(X_1)$ is greater than or equal to each single one of the frequencies for all conflicting bipartitions from cases where $P(X_1)$ is lower than one or more conflicting bipartitions, we reverse the sign of the ICA for all cases where $P(X_1)$ is lower. Thus, ICA values at or near 1 indicate the absence of any conflict for the bipartition defined by a given internode, whereas ICA values at or near 0 indicate that one or more conflicting bipartitions have almost equal support. Negative ICA values indicate that the internode of interest conflicts with one or more bipartitions that exhibit a higher frequency, and ICA values at or near -1 indicate the absence of support for the bipartition defined by a given internode. The behavior of the ICA measure for a range of different values of $X_1, X_2, \ldots, X_n$ is shown in figure 3.

Examples. Let us consider a set of 100 gene trees, from which 80 gene trees support bipartition X_1, 6 gene trees support the conflicting bipartition X_2, and 5 gene trees support the conflicting bipartition X_3. In this case,

$$P(X_1) = X_1/(X_1 + X_2 + X_3) = 80/(80 + 6 + 5) = 0.88,$$
$$P(X_2) = X_2/(X_1 + X_2 + X_3) = 6/(80 + 6 + 5) = 0.07,$$
and
$$P(X_3) = X_3/(X_1 + X_2 + X_3) = 5/(80 + 6 + 5) = 0.05.$$

Thus,

$$\begin{aligned}\text{ICA} = {} & 1 + P(X_1)\log_3[P(X_1)] + P(X_2)\log_3[P(X_2)] \\ & + P(X_3)\log_3[P(X_3)] = 1 + 0.88 \times \log_3(0.88) \\ & + 0.07 \times \log_3(0.07) + 0.05 \times \log_3(0.05) = 0.59.\end{aligned}$$

If $X_1 = 52$ gene trees and the conflicting bipartitions $X_2 = 29$ gene trees and $X_3 = 19$ gene trees, then

$$P(X_1) = X_1/(X_1 + X_2 + X_3) = 52/(52 + 29 + 19) = 0.52,$$
$$P(X_2) = X_2/(X_1 + X_2 + X_3) = 29/(52 + 29 + 19) = 0.29,$$
and
$$P(X_3) = X_3/(X_1 + X_2 + X_3) = 19/(52 + 29 + 19) = 0.19.$$

Thus,

$$\begin{aligned}\text{ICA} = {} & 1 + P(X_1)\log_3[P(X_1)] + P(X_2)\log_3[P(X_2)] \\ & + P(X_3)\log_3[P(X_3)] = 1 + 0.52 \times \log_3(0.52) \\ & + 0.29 \times \log_3(0.29) + 0.19 \times \log_3(0.19) = 0.08\end{aligned}$$

Finally, if $X_1 = 5$ gene trees and the conflicting bipartitions $X_2 = 15$ gene trees and $X_3 = 11$ gene trees, then

$$P(X_1) = X_1/(X_1 + X_2 + X_3) = 5/(5 + 15 + 11) = 0.16,$$
$$P(X_2) = X_2/(X_1 + X_2 + X_3) = 15/(5 + 15 + 11) = 0.48,$$
and
$$P(X_3) = X_3/(X_1 + X_2 + X_3) = 11/(5 + 15 + 11) = 0.36.$$

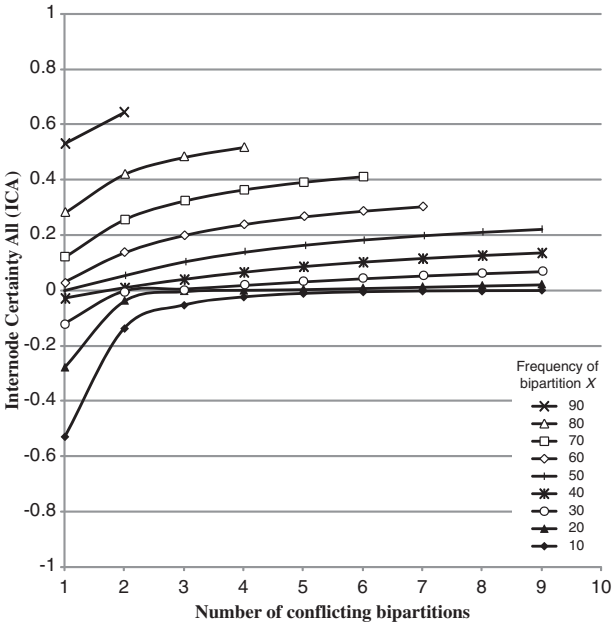

FIG. 3. Visualizing ICA for all the most prevalent conflicting bipartitions of a given internode. For simplicity, calculations were performed using a two-variable system ($X : Y \ldots Y$) with the number of conflicting bipartitions increasing. For example, the open triangle line on the graph illustrates the behavior of ICA when the frequency of the most strongly supported bipartition for a given internode is 80, with the remaining 20% equally divided among all conflicting bipartitions (e.g., if there is one conflicting bipartition, it will have a frequency of 20%, and if there are two conflicting bipartitions, each one will have a frequency of 10%, etc.).

Thus,

$$\begin{aligned}\text{ICA} = {} & 1 + P(X_1)\log_3[P(X_1)] + P(X_2)\log_3[P(X_2)] \\ & + P(X_3)\log_3[P(X_3)] = 1 + 0.16 \times \log_3(0.16) \\ & + 0.48 \times \log_3(0.48) + 0.36 \times \log_3(0.36) = 0.08.\end{aligned}$$

However, because $P(X_1) \leq P(X_2)$ and $P(X_1) \leq P(X_3)$, the sign of the ICA value is reversed to -0.08.

Tree Certainty

Given that empirical examinations of the support frequencies of internodes in a phylogeny suggest that they are generally independent from each other (Salichos and Rokas 2013), it is reasonable to assume that the mutual information or dependence between internodes in a phylogenetic tree is very small. Thus, the sum of all IC or ICA values across a phylogeny can be used to quantify changes in the degree of incongruence produced by the phylogenetic analysis of a given data set when analyzed with a variety of protocols or methods. Thus, for the complete set of $k-3$ internodes (internal branches) in a phylogeny, where k is the number of taxa, we define the TC as

$$\text{TC} = \sum_{i=1}^{i=k-3} \text{IC}_i$$

and TCA as

$$\mathrm{TCA} = \sum_{i=1}^{i=k-3} \mathrm{ICA}_i.$$

The maximum TC or TCA value is equal to $k-3$ and indicates a comprehensive absence of conflict in the phylogeny. When comparing phylogenies with different taxon numbers, a normalized value of TC or TCA can also be obtained by dividing the TC value by $k-3$, the number of internodes in the phylogeny.

Applications of IC, ICA, TC, and TCA

All four measures can be used to quantify incongruence on any data set that contains bipartitions, including from bootstrap replicate trees, gene trees, or individual characters (e.g., from morphology, from large-scale and rare genomic changes, or from individual sites in a sequence alignment). To demonstrate the utility of the four measures, we discuss three commonly used data types here, where one can deploy IC, ICA, TC, and TCA to quantify incongruence.

IC, ICA, TC, and TCA Can Quantify Incongruence in Sets of Trees

The most straightforward use of the four measures is for quantifying incongruence on a set of trees (fig. 4); often, this set is composed of the gene trees obtained from analysis of several different genes collected from the same set of taxa. In this case, calculation of the four measures will be based on the frequency values of the bipartitions present in the entire set of gene trees; note that the frequency value of a bipartition known as gene support frequency (GSF) reflects the percentage of gene trees that contain the bipartition (Gadagkar et al. 2005). When quantifying incongruence in a set of gene trees, the IC and ICA values of a given internode will reflect the degree of incongruence for that internode in the set of gene trees, and the TC and TCA values will reflect the degree of incongruence between the individual gene trees across the entire phylogeny. When applied to a data set of 1,070 gene trees from 23 taxa, the IC and ICA values revealed high levels of incongruence in several internodes of the extended MRC phylogeny and enabled us to distinguish between internodes that have similar GSF values but very different degrees of conflict (fig. 4*D*). Specifically, the placement of *Saccharomyces bayanus* and of *Zygosaccharomyces rouxii* received 52% and 62% GSF, whereas their IC values were 0.05 and 0.59 and their ICA values were 0.14 and 0.47, respectively (fig. 4*D*). This marked difference between the GSF and the IC/ICA values of the two internodes is a result of the absence of well-supported bipartitions that conflict with the placement of *Z. rouxii* and the presence of well-supported bipartitions that conflict with the placement of *S. bayanus* (Yu et al. 2012; Salichos and Rokas 2013).

When analyzing phylogenetic trees from a single gene or set of genes (multiple genes in supermatrix), it is standard practice to calculate the robustness of support for each internode of the gene tree via bootstrapping (Felsenstein 1985; Soltis and Soltis 2003). One can thus use the set of bootstrap replicate trees for a given gene to calculate IC, ICA, TC, and TCA. In this case, calculation of the measures will be based on the frequency values of the bipartitions present in the entire set of bootstrap replicate trees, which are better known as bootstrap support (BS) values. When quantifying incongruence in a set of bootstrap replicate trees from a single gene, the IC and ICA values of a given internode will reflect the degree of incongruence for that internode in the set of bootstrap replicate trees, and the TC and TCA values will reflect the degree of incongruence between the individual bootstrap replicate trees across the entire gene phylogeny. For example, in our recent study (Salichos and Rokas 2013), we ranked 1,070 genes from 23 yeast species based on their TC value as calculated from each gene's bootstrap trees. Interestingly, concatenation analysis of the 131 genes with the highest TC placed *Candida glabrata* in a position that is also supported by several distinct rare genomic changes (Scannell et al. 2006), a result that contradicts both the analysis of all 1,070 genes as well as previously published phylogenomic analyses (Hittinger et al. 2004; Rokas and Carroll 2005; Fitzpatrick et al. 2006; Jeffroy et al. 2006; Hess and Goldman 2011).

IC, ICA, TC, and TCA Can Quantify Incongruence in Sets of Bipartitions

The four measures can also be calculated from a set of partially resolved trees or even directly from bipartitions (fig. 4*B* and *C*). For example, the bipartitions present in each gene tree rarely receive equal support; the bootstrap consensus tree of virtually every gene shows that certain internodes receive higher BS or IC/ICA values, indicating that the degree of congruence of phylogenetic signals as well as the degree of "noise" from a given gene differs widely across internodes. Thus, it may frequently be desirable to use only genes' highly supported bipartitions in the inference of consensus phylogenies (one can easily select the highly supported bipartitions in the bootstrap consensus tree of a given gene by "collapsing" all internodes with BS values below a certain threshold using software such as the CONSENSE program in the PHYLIP package—[Felsenstein 1993]). In this case, calculation of the four measures will be exclusively based on the frequency values of those bipartitions that received high support (e.g., high BS) or present low conflict in the entire set of gene bootstrap consensus trees. Thus, the IC and ICA values of a given internode in the consensus tree will reflect the degree of incongruence for that internode among only the group of highly supported bipartitions present in the set of gene trees, whereas the TC and TCA values will reflect the degree of incongruence between highly supported bipartitions across the entire phylogeny. Note that the use of IC or ICA overcomes potential issues when only a small number of highly supported bipartitions are associated with a given internode by measuring the degree of incongruence independently of the number of bipartitions taken into consideration. For example, both the IC and the ICA value for the sister group *S. cerevisiae* and *S. paradoxus* calculated from an analysis of 1,070 gene trees from 23 yeast taxa is 0.56 (fig. 4*D*). In contrast, both the IC and ICA values

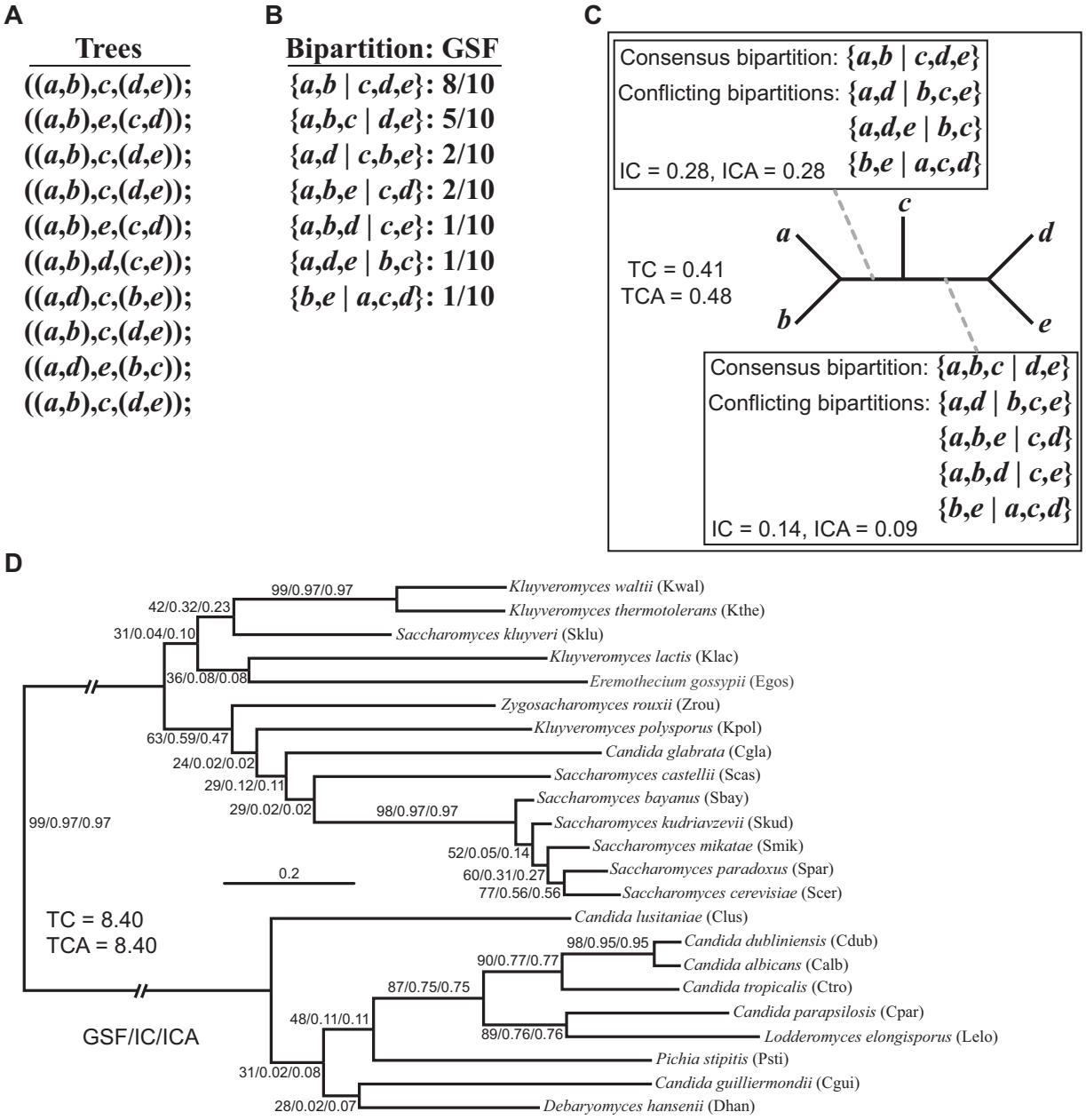

Fig. 4. IC, ICA, TC, and TCA can quantify incongruence in any set of trees or bipartitions. Given a set of trees (*A*) that defines a set of bipartitions (*B*), one can use the four measures to quantify incongruence (*C*). For example, examination of 1,070 gene trees revealed the presence of extensive incongruence in a phylogeny of 23 yeast taxa (*D*) (values near internodes correspond to GSF/IC/ICA values).

calculated using only those bipartitions that received ≥80% BS in individual gene analyses of the same 1,070 genes are 0.85, suggesting that most of the observed incongruence in the resolution of this internode stems from conflict among weakly supported bipartitions.

IC, ICA, TC, and TCA Can Quantify Incongruence in Sets of Individual Characters

Because the four measures can be applied to any data set that contains taxon bipartitions, one can extend their use to quantifying the level of phylogenetic conflict on any character in which the distribution of character states is such that it splits the taxon set into two nontrivial bipartitions (fig. 5). Assuming a character with two states 0 and 1 from a set of $k = m + n$ taxa, where $m \geq 2$ and $n \geq 2$, any site with a character state distribution of $(0_1 \ldots 0_m, 1_1 \ldots 1_n)$ corresponds to the bipartition {*m* taxa}/{*n* taxa}. Thus, one can use IC or ICA to quantify the degree of incongruence for a given bipartition defined by a character across a set of characters by considering the number of characters supporting that bipartition jointly with the number of characters supporting the most prevalent bipartition that conflicts with it (IC) or jointly with the numbers of characters supporting all most prevalent bipartitions that conflict with it (ICA). Note that, much like GSF reflects the frequency of bipartitions in a set of trees, the frequency value of a bipartition defined by a character reflects the percentage of characters that support the bipartition,

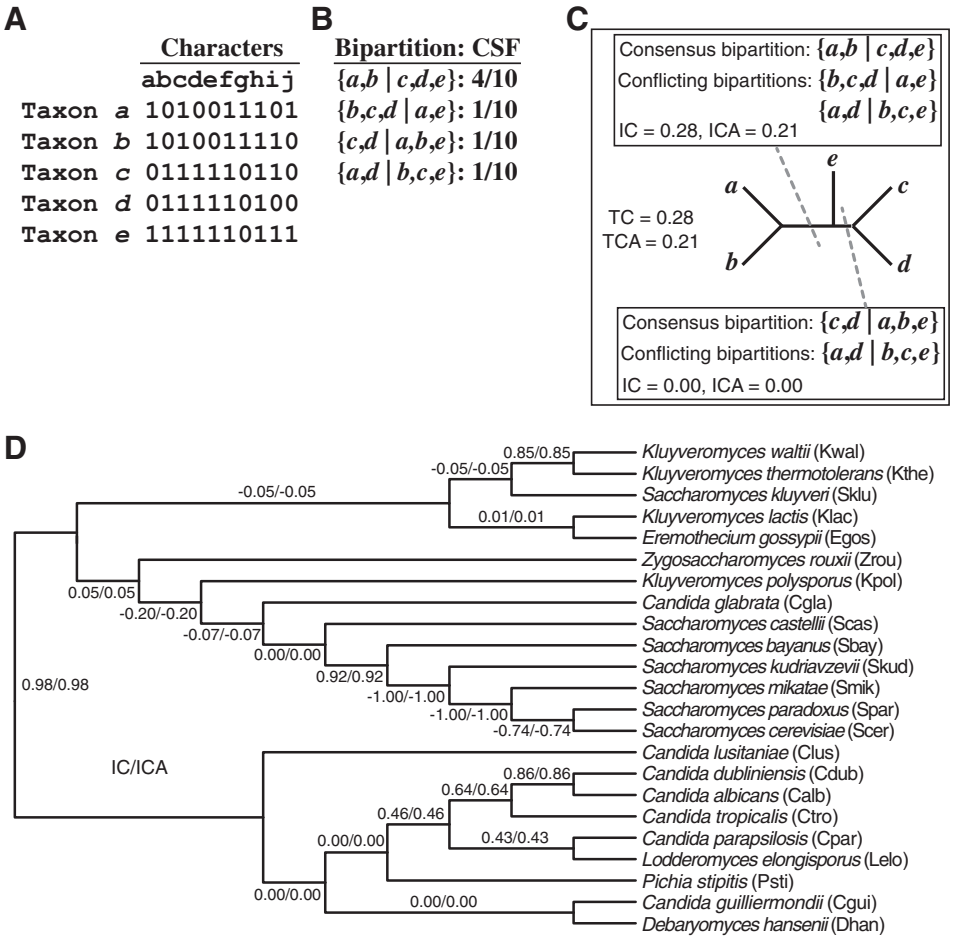

Fig. 5. IC, ICA, TC, and TCA can quantify incongruence in any set of characters that define bipartitions. Given a set of characters (*A*) that defines a set of bipartitions (*B*), one can use the four measures to quantify incongruence (*C*). For example, examination of 20,289 sites that contain single radical substitutions (defined as substitutions with a blosum62 matrix score ≤ -3) from the data set of 1,070 genes from 23 yeast taxa showed that the bipartitions defined by such sites not only lacked information about several internodes of the yeast phylogeny but also displayed considerable levels of incongruence (*D*).

which we denote as character support frequency. Examples of characters that can be used to define bipartitions include rare genomic changes (Rokas and Holland 2000), indels (Belinky et al. 2010), sites that contain a single substitution between amino acids that differ radically in their physicochemical properties (Rogozin et al. 2007), binary morphological characters, as well as any other binary characters. For example, analysis of 20,289 sites that contain single radical substitutions (defined as substitutions with a blosum62 matrix score ≤ -3), from the data set of 1,070 genes from 23 yeast taxa, also known as RGC_CAMs (Rogozin et al. 2007), showed that the bipartitions defined by such sites were more incongruent than the bipartitions present in the 1,070 gene trees.

Using TC and TCA to Evaluate the Impact of Different Practices in Data Analysis

Summing the IC or ICA values across all internodes of a phylogeny amounts to the phylogeny's TC or TCA, respectively. One useful application of the TC and TCA measures is for comparing the relative impact of different analytical practices on incongruence. For example, one could calculate the TC and TCA values of the extended MRC phylogeny constructed from the gene trees estimated from analysis of 100 genes with only those sites that do not contain missing data and compare it with the TC/TCA measured from the eMRC phylogeny constructed from analysis of the same 100 genes in which only sites with more than 50% data missing are excluded. In this case, the practice with the highest TC/TCA value will be that one that displays the lowest degree of incongruence among the 100 gene trees. In contrast, a high decrease in TC/TCA may indicate that a particular data-filtering approach increases incongruence across the phylogeny. For example, examination of the TC of the trees from the 100 slowest evolving genes in a data matrix composed of 1,070 genes from 23 yeast taxa showed that they had a substantially lower TC than the TC calculated by considering all 1,070 gene trees (Salichos and Rokas 2013).

Calculating IC, ICA, TC, and TCA Using the RAxML Software

We implemented the score calculations of the four measures in RAxML (Stamatakis 2006; version 8.0.0, available via

https://github.com/stamatak/standard-RAxML, last accessed January 31, 2014), taking advantage of already available efficient data structures for performing calculations on bipartitions (Aberer et al. 2010). For a full description of the commands for calculation of the four measures and an example, please see the dedicated manual (supplementary text file S1, Supplementary Material online), the new RAxML manual (http://sco.h-its.org/exelixis/resource/download/NewManual.pdf, last accessed January 31, 2014) and test data set (supplementary data files S1 and S2, Supplementary Material online). Given a set of gene trees, RAxML can directly calculate an MRC as well as an eMRC tree on this set that is annotated by the respective IC and ICA values. The particularly compute-intensive inference of eMRC trees (finding the optimal eMRC tree is, in fact, nondeterministic polynomial-time hard [NP-hard; Phillips and Warnow 1996]) relies on the fast parallel implementation presented in Aberer et al. (2010). It can also compute stricter MRC trees with arbitrary threshold settings that range between 51% and 99%. Furthermore, we have implemented an option that allows for drawing IC scores onto a given, strictly bifurcating reference tree (e.g., the best-known ML tree).

Note that the IC and ICA values are represented as branch labels, because, as is the case for BS values, information associated to bipartitions of a tree always refers to its internodes (internal branches) and not its nodes. Each tree viewer (e.g., Dendroscope [Huson and Scornavacca 2012]) that can properly parse the Newick tree format is able to display these branch labels. The rationale for not providing IC values as node labels is that some tree viewers may not properly rotate the node labels when the user reroots the tree, leading to an erroneous internal branch-to-IC-value association.

When calculating the IC and ICA values on extended MRC trees or onto a given reference tree, it may occur that the bipartition that has been included in the tree has lower support than one or more conflicting bipartitions (see also above). In this case, RAxML will display a warning to the user and annotate the internode with a negative IC value. Note that this is not only a theoretical possibility when using extended MRC trees but a frequent observation for bipartitions that have low frequency in a gene tree set or that have low BS in a set of bootstrap replicate trees.

RAxML also calculates the TC and TCA values as well as their relative values that are normalized by the maximum possible TC/TCA values for a given phylogeny. Finally, we have implemented a verbose output option that allows users to further scrutinize particularly interesting conflicting bipartitions. In verbose mode, RAxML will generate two types of output files: one set of files containing the bipartition included in the MRC tree and its corresponding conflicting bipartitions in Newick format and an output file listing all bipartitions (included and conflicting) with their IC and ICA values in a PHYLIP-like format.

Discussion

To tackle gene incongruence, phylogeneticists often resort to creating concatenated data matrices composed of tens or hundreds of genes (Rokas et al. 2003; Rokas et al. 2005; Dunn et al. 2008; Philippe et al. 2009; Regier et al. 2010). Because the vast majority of concatenation studies assesses robustness in inference using bootstrapping, an extremely useful measure of robustness of inference when data are limited (Felsenstein 1985) but one that in the presence of large amounts of data will nearly always result in 100% support (Rokas and Carroll 2006; Kumar et al. 2012; Salichos and Rokas 2013), numerous concatenation studies purport to have resolved long-standing phylogenetic problems. However, different phylogenomic studies focused on the same internodes sometimes provide contradicting, but equally robustly supported, answers (Dunn et al. 2008; Philippe et al. 2009; Kocot et al. 2011; Smith et al. 2011), suggesting that incongruence is not ameliorated, but rather masked, by these practices. Consequently, accurate phylogenetic inference requires not only large amounts of data and absolute BS but also demonstration that the data do not contain substantial amounts of conflicting phylogenetic signal (Salichos and Rokas 2013). Thus, accurate inference requires methods that identify and quantify conflicts in phylogenetic signal.

To quantify the degree of incongruence present in phylogenomic data matrices, we developed two novel measures, IC and ICA, which quantify the degree of conflict on each specific internode of a phylogeny and two novel measures, TC and TCA, which quantify the degree of conflict for the whole tree. All four measures can be used for a wide variety of different phylogenetic markers, from individual characters to gene trees to genomic characters (figs. 4 and 5), and are meant to provide simple, fast, and intuitive measurements that identify the presence of incongruence in a phylogenomic data matrix rather than to elucidate the root cause(s) of the observed incongruence. Even though the absolute values of our measures are not aimed to provide statistical significance, the degree of certainty calculated derives from the amount of information on each internode. For example, in the case of IC, the degree of certainty corresponds to the ratio between the most prevalent and the next most prevalent, but conflicting, bipartition (fig. 2). If the most prevalent bipartition is supported by 95% of the data and the next most prevalent conflicting bipartition is supported by the remaining 5%, then the value of the IC measure will be approximately 0.71, whereas if the two most prevalent conflicting bipartitions have the same frequency of support, then IC will equal zero.

Compared with the very popular incongruence length difference test (Farris et al. 1994), our measures can easily be applied to the study of a single internode or the whole tree, to study one or many data partitions, and are not dependent on a particular optimality criterion. Compared with topology constraint tests, such as the Kishino–Hasegawa (KH) test (Kishino and Hasegawa 1989), the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa 1999), and the approximately unbiased (AU) test (Shimodaira and Hasegawa 2001), there is no need for a priori tree selection, and multiple internodes can be examined simultaneously very quickly. The price of this speed and flexibility, however, is that our tests are not designed to test specific phylogenetic hypotheses or

provide estimates of statistical significance; in many ways, our measures are designed to quickly identify incongruence in phylogenomic data matrices, enabling users to further explore its causes using more custom methods.

Our IC, ICA, TC, and TCA measures do not distinguish whether a low degree of certainty is the result of strong conflicts in phylogenetic signal or random noise due to the absence of any signal. In other words, incongruence between trees does not necessarily indicate conflicting support, because incongruent trees are also the null expectation when a data matrix contains no phylogenetic signal (although differences between IC and ICA values may alert for the presence of more than two signals). In such cases, users are advised to examine whether the tree distance distribution of observed trees deviates significantly from randomness by using a tree distance method (Hess and Goldman 2011; Salichos and Rokas 2013), such as the Robinson–Foulds tree distance (Robinson and Foulds 1981), before inferring that the low degree of certainty in a data matrix is the result of strong conflicts in phylogenetic signal. Other alternatives include employing the more computationally intensive topology constraint KH, SH, or AU tests (Kishino and Hasegawa 1989; Shimodaira and Hasegawa 1999; Shimodaira and Hasegawa 2001).

One potential drawback when applying the IC, ICA, TC, and TCA measures is that their values may not be representative when small numbers of characters or gene trees are used. Although this is a general problem that influences all measures, including BS and GSF, our measures are likely most informative when applied to large amounts of data (e.g., hundreds of characters or dozens of genes or hundreds of bootstrap replicates). Our TC and TCA measures also assume that the support frequencies of internodes in a phylogeny are independent from each other. Even though this is an approximation, previous results suggest that the application of a variety of standard practices aimed at reducing incongruence, such as removal of unstable or fast-evolving taxa, do not affect IC and ICA values across the entire phylogeny; rather, their effects are largely localized on one particular internode (Salichos and Rokas 2013). It should be noted that such a focus on a single internode or a small, local neighborhood of an internode represents a common approximation in phylogenetics and is frequently used to design search heuristics or statistical tests such as approximate likelihood-ratio test (aLRT; Anisimova and Gascuel 2006).

Finally, IC, ICA, TC, and TCA measures, as currently implemented in RAxML, cannot be applied on data sets with missing data (e.g., when some genes are missing from certain taxa), because dealing with trees that only contain a subset of the overall taxon set is computationally substantially more challenging and requires appropriate adaptation and/or extension of supertree methods. Hence, the solution to this problem is not straightforward, but we hope to address this challenging issue in the near future.

Supplementary Material

Supplementary text file S1 and data files S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Christoph Hahn for testing early RAxML implementations of these measures and for constructive feedback. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University. This work was supported by the National Science Foundation (DEB-0844968 to A.R.) and by institutional funding from the Heidelberg Institute for Theoretical Studies (to A.S.).

References

Aberer AA, Pattengale ND, Stamatakis A. 2010. Parallelized phylogenetic post-analysis on multi-core architectures. *J Comp Sci.* 1:107–114.

Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol.* 55: 539–552.

Baker RH, DeSalle R. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. *Syst Biol.* 46:654–673.

Belinky F, Cohen O, Huchon D. 2010. Large-scale parsimony analysis of metazoan indels in protein-coding genes. *Mol Biol Evol.* 27: 441–451.

Bryant D. 2003. A classification of consensus methods for phylogenetics. In: Janowitz M, Lapointe F-J, McMorris FR, Mirkin B, Roberts FS, editors. Bioconsensus, DIMACS. AMS, p .163–184.

Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ. 1993. Partitioning and combining data in phylogenetic analysis. *Syst Biol.* 42:384–397.

Cummings MP, Otto SP, Wakeley J. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. *Mol Biol Evol.* 12: 814–822.

Cunningham CW. 1997. Can three incongruence tests predict when data should be combined? *Mol Biol Evol.* 14:733–740.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol.* 24: 332–340.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.

Faith DP. 1991. Cladistic permutation tests for monophyly and nonmonophyly. *Syst Zool.* 40:366–375.

Farris JS, Kallersjo M, Kluge AG, Bult C. 1994. Testing significance of incongruence. *Cladistics* 10:315–319.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791.

Felsenstein J. 1993. PHYLIP (Phylogeny Inference Package). Distributed by the author. Seattle (WA): Department of Genetics, University of Washington.

Fitzpatrick DA, Logue ME, Stajich JE, Butler G. 2006. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol Biol.* 6:99.

Gadagkar SR, Rosenberg MS, Kumar S. 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *J Exp Zool B Mol Dev Evol.* 304:64–74.

Goldman N, Anderson JP, Rodrigo AG. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst Biol.* 49:652–670.

Hess J, Goldman N. 2011. Addressing inter-gene heterogeneity in maximum likelihood phylogenomic analysis: yeasts revisited. *PLoS One* 6: e22783.

Hittinger CT, Rokas A, Carroll SB. 2004. Parallel inactivation of multiple *GAL* pathway genes and ecological diversification in yeasts. *Proc Natl Acad Sci U S A.* 101:14144–14149.

Huelsenbeck JP, Bull JJ, Cunningham CW. 1996. Combining data in phylogenetic analysis. *Trends Ecol Evol.* 11:152–158.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 23:254–267.

Huson DH, Rupp R, Scornavacca C. 2010. Phylogenetic networks: concepts, algorithms and applications. New York: Cambridge University Press.
Huson DH, Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol.* 61:1061–1067.
Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? *Trends Genet.* 22:225–231.
Kishino H, Hasegawa M. 1989. Evaluation of the maximum-likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J Mol Evol.* 29: 170–179.
Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander C, Moroz LL, Lieb B, et al. 2011. Phylogenomics reveals deep molluscan relationships. *Nature* 477:452–456.
Kumar S, Filipski AJ, Battistuzzi FU, Kosakovsky Pond SL, Tamura K. 2012. Statistics and truth in phylogenomics. *Mol Biol Evol.* 29: 457–472.
Lento GM, Hickson RE, Chambers GK, Penny D. 1995. Use of spectral analysis to test hypotheses on the origin of pinnipeds. *Mol Biol Evol.* 12:28–52.
Le Quesne WJ. 1969. A method of selection of characters in numerical taxonomy. *Syst Zool.* 18:201–205.
Maddison WP. 1997. Gene trees in species trees. *Syst Biol.* 46:523–536.
Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. *Mol Biol Evol.* 5:568–583.
Philippe H, Derelle R, Lopez P, Pick K, Borchiellini C, Boury-Esnault N, Vacelet J, Renard E, Houliston E, Queinnec E, et al. 2009. Phylogenomics revives traditional views on deep animal relationships. *Curr Biol.* 19:706–712.
Phillips C, Warnow TJ. 1996. The asymmetric median tree—a new model for building consensus trees. *Discrete Appl Math.* 71:311–335.
Planet PJ. 2006. Tree disagreement: measuring and testing incongruence in phylogenies. *J Biomed Inform.* 39:86–102.
Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463: 1079–1083.
Robinson DR, Foulds LR. 1981. Comparison of phylogenetic trees. *Math Biosci.* 53:131–147.
Rodrigo AG, Kelly-Borges M, Bergquist PG, Bergquist PL. 1993. A randomisation test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. *New Zeal J Bot.* 31: 257–268.
Rogozin IB, Wolf YI, Carmel L, Koonin EV. 2007. Ecdysozoan clade rejected by genome-wide analysis of rare amino acid replacements. *Mol Biol Evol.* 24:1080–1090.
Rokas A, Carroll SB. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Mol Biol Evol.* 22:1337–1344.
Rokas A, Carroll SB. 2006. Bushes in the tree of life. *PLoS Biol.* 4:e352.
Rokas A, Chatzimanolis S. 2008. From gene-scale to genome-scale phylogenetics: the data flood in, but the challenges remain. *Methods Mol Biol.* 422:1–12.
Rokas A, Holland PWH. 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol.* 15:454–459.
Rokas A, Kruger D, Carroll SB. 2005. Animal evolution and the molecular signature of radiations compressed in time. *Science* 310: 1933–1938.
Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425:798–804.
Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497:327–331.
Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440:341–345.
Shannon CE. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27:379–423.
Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol.* 16: 1114–1116.
Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247.
Slowinski JB, Page RDM. 1999. How should species phylogenies be inferred from sequence data? *Syst Biol.* 48:814–825.
Smith SA, Wilson NG, Goetz FE, Feehery C, Andrade SC, Rouse GW, Giribet G, Dunn CW. 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* 480:364–367.
Soltis PS, Soltis DE. 2003. Applying the bootstrap in phylogeny reconstruction. *Stat Sci.* 18:256–267.
Song S, Liu L, Edwards SV, Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc Natl Acad Sci U S A.* 109: 14942–14947.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sunderland (MA): Sinauer.
Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK, editors. Molecular systematics. Sunderland (MA): Sinauer. p. 407–514.
Templeton AR. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. *Evolution* 37:221–244.
Thorley JL, Page RDM. 2000. RadCon: phylogenetic tree comparison and consensus. *Bioinformatics* 16:486–487.
Thorley JL, Wilkinson M. 1999. Testing the phylogenetic stability of early tetrapods. *J Theor Biol.* 200:343–344.
Wilson EO. 1965. A consistency test for phylogenies based on contemporaneous species. *Syst Zool.* 14:214–220.
Yu Y, Degnan JH, Nakhleh L. 2012. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genet.* 8:e1002660.
Zhong B, Liu L, Yan Z, Penny D. 2013. Origin of land plants using the multispecies coalescent model. *Trends Plant Sci.* 18:492–495.