

Diagnosing Discordance: Signal in Data, Conflict in Paradigms

Aleta Quinn*

I analyze recent debates between proponents of concatenation versus coalescence in phylogenetic inference. I argue that concatenation is the latest manifestation of a paradigm weaving through phylogenetics that has focused on a successive series of models thought to be justified by the “principle of total evidence.” I analyze the principle of total evidence as the main philosophical strand linking parsimony versus likelihood (1980s), character congruence versus consensus trees (1990s), and concatenation versus coalescence (2000–10s). My hope is to provide a foothold for philosophers to engage with contemporary phylogenetics, in the face of the discipline’s bewildering and rapidly expanding array of computational models. The basic idea of total evidence—include all data that is relevant to an analysis, that has signal with respect to the problem at hand—is extremely attractive at an intuitive level. However, the general intuition is less clear in the case that all relevant data are included in the overall study, but no single method employs the total dataset in one inferential step. Moreover, simulation studies demonstrate that there are cases in which excluding some data, even when that data provides signal, leads to a better result by a particular method. Each of these points is explored through analysis of the historical and contemporary debates.

Keywords

phylogenetics • multi-species coalescent • concatenation • total evidence • parsimony • cladist

Part of the special issue Species in the Age of Discordance, guest-edited by Matthew H. Haber and Daniel J. Molter.

*Department of Politics and Philosophy, University of Idaho, Department of Vertebrate Zoology, National Museum of Natural History, 875 Perimeter Drive MS 3165, Moscow ID 83844, USA, aquinn@uidaho.edu

Received 1 December 2017; Revised 17 February 2019; Accepted 8 March 2019
doi:10.3998/ptpbio.16039257.0011.017



1 Introduction

Sterner and Lidgard (2018) urge that philosophers of phylogenetics move beyond the “systematics wars”, referring to the 1960s–80s debates between numerical taxonomists, evolutionary taxonomists, and phylogenetic systematists. Indeed, philosophers would do well to move beyond those wars, and to focus even more recently than the parsimony versus likelihood debates of the 1980s–90s. In this paper I use integrated historical–philosophical analysis of those debates to clarify a contemporary dispute between proponents of coalescence-based methods (Edwards et al. 2016; Liu et al. 2015; Liu and Edwards 2015) and proponents of concatenation (Gatesy and Springer 2014; Gatesy et al. 2016; Gatesy et al., forthcoming; Simmons and Gatesy 2015; Springer and Gatesy 2016). My intent is to illuminate the present state of the field of phylogenetics by tracing the use of one particular philosophical argument, “total evidence”, through several distinct scientific debates.

I analyze the principle of total evidence as the main philosophical strand linking parsimony versus likelihood (Section 2: Kluge), character congruence versus consensus trees (3: Nixon and Carpenter), and concatenation versus coalescence (4: Gatesy). The basic idea of total evidence—include all data that is relevant to an analysis, that has signal with respect to the problem at hand—is extremely attractive at an intuitive level. However, the general intuition is less clear in the case that all relevant data are included in the overall study, even if no single method employs the total dataset in one inferential step.

Moreover, simulation studies demonstrate that there are cases in which excluding some data, even when that data provides signal, leads to a better result by a particular method. For example Talavera and Castresana (2007) found that omitting uncertain sections of gene alignments can lead to better inference of phylogenetic trees. These uncertain sequences, when analyzed independently, converged on the correct phylogenetic tree when given enough time and data. But omitting the sequences led each of parsimony, maximum likelihood, and minimum evolution approaches to converge on the correct result more efficiently.¹ Talavera and Castresana suggest that this result is explained by the claim that signal-to-noise ratio can be more important than the amount of signal itself. In other cases, omitting data may help because error in the data (the hypothesis that conflicting data are homoplasious and thus not reliable evidence) is conflated with discordance in real entities (disagreement between gene trees, disagreement between gene trees and species trees, or evolution within gene trees generating long branch attraction).

Most recently, critics of coalescence-based phylogenetic inference have appealed to the principle of total evidence as reason to distrust coalescence methods as currently implemented. Coalescence-based phylogenetic inference is indeed a newly emerging paradigm, as has been suggested by Edwards (2009), both for inferring species and phylogeny. Non-coalescence-based methods may be positively misleading (statistically inconsistent) in the case that there is discordance between gene trees and species trees (Kubatko and Degnan 2007; Roch and Steel 2015). In contrast, coalescence-based methods use the signal provided by discordant gene trees to infer species trees (Liu et al. 2015) or to estimate species boundaries (Yang and Rannala 2010; Yang 2015). Though these particular methods of implementing the multi-species coalescent model have been challenged, most critics acknowledge that methods are needed to handle the challenge of discordance between gene trees, and between gene trees and species trees (Warnow 2015). Indeed, the challenge is perhaps best described as reconceptualizing the empirical phe-

1. The authors note two situations in which ML does not appear to perform better after problematic sequences are omitted (572). They suggest (rather unconvincingly) that the particular method of screening sequences would perform better if more realistic models of evolution were applied to the situations in question. This issue is not relevant to my arguments, which only require that omitting data in some situations (not all) leads to a better result.

nomenon of discordance: it can be a problem for some methods, but it can also be a source of signal to be exploited by other methods. Objections made on the basis of total evidence target the fact that coalescence methods partition the available evidence in several ways, as will be seen below (4: Gatesy).

Edwards et al. (2016) objected to the view that coalescence-based approaches are engaged in a “battle for supremacy” with concatenation methods in contemporary phylogenetics. As I will demonstrate, the apparent hostility between proponents of concatenation versus coalescence traces to the earlier phylogenetic debates. Concatenation is the latest manifestation of a paradigm weaving through systematics that has focused on a successive series of models thought to be justified by the principle of total evidence. I conclude (5: Cladists) with remarks on some asymmetries between the total evidence versus coalescence paradigms.

That coalescence-based methods are a new paradigm does not imply that older methods of studying phylogeny are obsolete or inferior. Just as developmental morphology remains critical to phylogenetic inference and, more broadly, to the synthesis of systematics and evolutionary theory, parsimony remains an invaluable approach in many situations. The theoretical and pragmatic limits of chosen methods should be understood in phylogenetic studies, as in any science.

I have included a glossary with some phylogenetic terminology. I hope that this may prove useful to readers of this special issue, and I welcome comments—philosophers well know how difficult it can be to define basic terms.

2 Kluge on Total Evidence

Several researchers have remarked on the uses of Popper by competing sides in the parsimony versus maximum likelihood debates of the 1980s–90s (Hull 1983, 1999; Rieppel 2008). Kluge (Kluge 1989) has also drawn on Whewell, (Kluge 1989), Sober (1988), (Kluge and Wolf 1993), and Carnap (1950) to develop his own philosophical arguments regarding the principle of total evidence.

The principle might be interpreted very broadly as the conviction that all available evidence should be considered when evaluating a hypothesis or theory. This general principle would need refinement in light of the Duhem-Quine thesis that whether an observation counts as evidence is theory dependent (Quine 1951). Carnap’s (1950) intent was to develop an account of the confirmation relation within the context of a theory understood as a set of sentences (or perhaps more strictly, propositions). Hempel (1965) extended the principle to the conditions for rationality of belief. Kluge interpreted the principle in the context of cladistic theory both more narrowly, as applying to a single inferential step, and more broadly, as a definitive claim about the nature of science.

The narrow component of Kluge’s interpretation was to apply the principle to the inference of cladograms. Cladograms express claims about the branching pattern of descent between lineages. Kluge argued that to make and defend hypotheses about the pattern, one must use all available data in a single inferential step. Kluge further argued that the procedure of maximizing parsimony was the only correct method available to implement this principle. Then available alternatives included maximum likelihood analysis, which can support results that do not match the most parsimonious solution(s), and using partitioned data sets to analyze distinct parts of the tree in succession. Under the latter procedure, distinct parts of the tree are combined to produce an overall tree, a procedure Kluge referred to as taxonomic congruence.

The broader component of Kluge’s interpretation was that alternative methods for inferring cladograms must necessarily be weaker than parsimony analysis of combined data sets. Alternative methods reflect “questionable scientific philosophy” (Kluge and Wolf 1993, 190):

For example, we believe the methodological strategies of taxonomic congruence and the use of suboptimal hypotheses involve many decisions arrived at without reason, and it will not be possible to argue that cladistics is coherent and logically consistent if those practices become a routine part of phylogenetic systematics. (195–6).

In this framework, “suboptimal hypotheses” are those that are less parsimonious than the optimal solution that is arrived at via parsimony analysis of the total combined data set. Kluge argued that any partitioning of the data set violates the principle of total evidence and therefore requires justification; and that such justification is lacking. Proponents of partitioning might point out that all data is used in the course of the overall analysis, even if no single inferential step uses all data (Miyamoto and Fitch 1995). While this might satisfy a general prescription to use all available evidence, data partitioning does not satisfy Kluge’s strict interpretation of the principle of total evidence. In turn, according to Kluge, methods that partition data cannot be responsible science.

Kluge later (1997) argued that parsimony is to be preferred over maximum likelihood because only results inferred through parsimony can be satisfactorily supported using Popper’s criterion of degree of corroboration. However, Kevin de Queiroz and Steven Poe (2001) showed a way in which Popper’s degree of corroboration corresponds to likelihood. Haber (2005) demonstrated further problems with Siddall and Kluge’s (1997) arguments from falsificationism and probability theory.

Kluge’s arguments did little to convince proponents of alternative methods of phylogenetic inference, but have been cited widely by proponents of parsimony.

3 Nixon and Carpenter on Character Congruence

Nixon and Carpenter picked up the thread of Kluge’s total evidence arguments but offered a new term for the critical issue, “simultaneous analysis” (Nixon and Carpenter 1996): the single inferential step that produces a tree hypothesis must use all the character data simultaneously. Nixon and Carpenter’s target was any procedure that assigns different weights to character data.

Arguments about how to weight characters—how much evidential significance to assign to each character relative to others—are at least as old as Adanson (1763). Indeed, some numerical taxonomists cited Adanson as intellectual forebear for their own programme to measure and weight as many characters as possible in completely “objective” fashion (Sneath 2015). Numerical taxonomists and their cladist opponents alike tended to gloss over the fact that all characters are implicitly weighted by the process of assigning evidential status to some observation in the first place. The same tooth could be measured once, twice, twenty, or one hundred times, or modeled as a nearest fit function from which parameters could be abstracted, or described in terms of development or material composition, etc. Discussion of objective character selection sometimes took on a very loose epistemological and metaphysical tone: “Certainly there is a boundary between the mind and the reality of any boundary” (Nixon and Carpenter 1996, 225).

The issue of character selection also cropped up in discussion of whether chosen characters are independent. In some cases the dependence is clear (the volume of the tooth is not independent of its width; the length of each tooth, considered for all teeth, may not be independent of the length of the palate). The situation is far less clear when both morphological and genetic data are included in a single analysis, sometimes alongside behavioral characters and other sources of evidence entirely (e.g. dating information from the fossil record, biogeographical information). With respect to observed characters of organisms, the worry was that scoring separately

characters that are in fact interdependent leads to arbitrarily assigning greater weight to some data than to others. Nixon and Carpenter (1996) remarked that ideally each character should be weighted in terms of the probability that it represents a synapomorphy independent of other hypothesized synapomorphies. Unfortunately, it is extremely unclear precisely how to do this in practice. Uncertainty about character independence is one justification for the practice of partitioning data according to the source or type of data used (Bull et al. 1993; Huelsenbeck et al. 1994; Huelsenbeck, Bull, and Cunningham 1996; William and Ballard 1996).

Nonetheless, proponents of total evidence continued to urge that there are “compelling philosophical reasons for combining all relevant character evidence in simultaneous analysis (Miyamoto, 1985; Kluge, 1989, 1997; Brower et al., 1996; Nixon and Carpenter, 1996; DeSalle and Brower, 1997; Farris, 1997; Siddall, 1997; Siddall and Kluge, 1997)” (Gatesy, O’Grady, and Baker 1999).

Nixon and Carpenter (1996) claimed that the argument from “explanatory power” was decisive. Their analysis appears to rely on the premise that phylogenetic inference is necessarily abductive (see Fitzhugh 2006, Quinn 2016, and Sober 1991 for distinct approaches to phylogenetic inference as abductive). Nixon and Carpenter apparently take it that because phylogenetics inference is abductive, explanatory power is the most essential criterion to evaluate phylogenetic hypotheses. Explanatory power, they claim, can only be maximized by simultaneous analysis of all data. By “explanatory power” is meant the link between the phylogenetic hypothesis and the data, since it is by virtue of the phylogenetic hypothesis that individual characters are interpreted as synapomorphies (or dismissed as homoplasy). Any analysis of a partition of the data set can only yield a hypothesis that explains less of the overall data, when joined to hypotheses generated by other partitions, since the combined hypothesis will produce a phylogenetic tree that identifies fewer characters as synapomorphies.²

There are several problems with this argument. Overall the problem is that this approach prioritizes explanatory fit to such an extent as to ignore the possibility that some, possibly very much, of the data is misleading. Felsenstein (1978) demonstrated cases in which parsimony analysis is statistically inconsistent, that is, that the analysis yields the wrong answer with increasing confidence as more data is added. Felsenstein’s critics argued (and continue to argue) that any method of analysis will be statistically inconsistent in certain situations. This observation does not mean that Felsenstein’s demonstration can be dismissed entirely, however (see the recent (2016) editorial in *Cladistics* for such a dismissal). The critical point is that any procedure of maximizing explanatory fit will be vulnerable to the problem of adding misleading data. Maximizing explanatory fit therefore requires some thought. Attempts to analyze the signal in partitions of the data set are one approach to investigating sample error. This is a philosophical justification for using consensus trees in some situations (Hillis 1987).

The point that maximizing explanatory fit requires careful thought may seem obvious to philosophers for other reasons. From Duhem (1906) onward philosophers of science have discussed diverse epistemic virtues as criteria for theory choice. Explanatory fit is often identified as one such epistemic virtue. Maximizing explanatory fit at the expense of all other virtues, such as simplicity, scope, or fecundity would lead to problems.

There are also several clues that Nixon and Carpenter favored simultaneous analysis not for general philosophical reasons but because they endorsed the superiority of parsimony over other approaches. Certainly Nixon and Carpenter believed that parsimony was itself superior to other approaches because of general philosophical reasons, but that claim requires independent support. It is not my purpose here to evaluate such arguments (a task which would require at

2. The only exception is the case in which the partitioned data sets all happen to produce an identical phylogenetic tree; in this case partitioned analysis does just as well (and only as well) as simultaneous analysis.

least a separate paper, and which has been the subject of several books). Rather, I want to point out the fact that the argument about simultaneous analysis versus consensus approaches was explicitly identified as the debate about parsimony:

We agree that the simultaneous analysis of combined data is the best approach to phylogenetic inference: it is the one that best applies parsimony. (Nixon and Carpenter 1996, 221)

Another way of stating this is that simultaneous analysis of combined data better maximizes cladistic parsimony than separate analyses, hence is to be preferred. (237)

Nixon and Carpenter (1996) also claim that their opponents see the debate in these terms:

The position that datasets should be analysed separately is clearly based on a rejection of the principle of parsimony in cladistics. (237)

Nixon and Carpenter supported this claim by demonstrating that one form of partitioning data effectively removes discordant character signal from the analysis. Nixon and Carpenter in turn equate this with abandoning parsimony:

Viewed in terms of a single dataset with homoplasy, the approach of analysing character cliques separately, while consistent with the “class” definitions of [Alan] de Queiroz et al., Huelsenbeck et al., and other proponents of character segregation, can be seen as merely a way of removing the parsimony criterion from the analysis. (237)

The particular method of analyzing characters that Nixon and Carpenter discuss, taken from Alan de Queiroz et al. (1995), does effectively eliminate parsimony as the criterion for adjudicating conflicting signal. However, this is not the only method for partitioning data, and it is not the only way of interpreting the results of partition analysis. As well, it is not the case that proponents of partitioning datasets were motivated to eliminate parsimony. The results of partition analyses can be interpreted in diverse ways, including some that are consistent with a parsimony framework. While Nixon and Carpenter saw the debate about simultaneous analysis as equivalent to the debate about using parsimony, not all of their opponents agreed.

4 Gatesy on Hidden Support

Gatesy et al. (1999) developed a concept called “hidden support” that they claimed emerges from the philosophical concept of total evidence. The general approach matches Kluge (section 2) and Nixon and Carpenter (section 3): under the general philosophical principle that more evidence is better, Gatesy et al. argued to include all characters in the critical part of phylogenetic inference. Gatesy et al.’s arguments occurred in the context of debate between proponents of concatenation versus congruence, and later, concatenation versus coalescence. Concatenation is essentially synonymous with Nixon and Carpenter’s simultaneous analysis, while congruence refers to methods that analyze data subsets or portions of trees independently of the overall dataset. Coalescence refers to a specific set of methods that use congruence and that have been proposed to represent a new paradigm in phylogenetic analysis (Edwards 2009).

Though they intended their argument to apply to all phylogenetic analyses, Gatesy, O'Grady, and Baker (1999) defined "hidden support" very specifically in a way that only makes sense within a parsimony framework. The idea is that combining data sets enables subsets of the data to provide signal that is lost when an individual subset is considered alone. Specifically, a data set of characters may yield an unresolved quartet with respect to four taxa (OTU's) with branch support of zero, when that data set is considered on its own. When combined with other data sets, that same initial data set can provide information about a clade within the unresolved quartet. This is because the additional data sets can provide information about the rooting of the quartet, and once this rooting is hypothesized, signal emerges from the initial data set. This is because the rooting information can interpret character states as derived or ancestral. The signal that emerges from the data subset when combined with all data, versus the signal that was present in the data subset considered on its own, is called hidden support.

There is no clear analogue for this concept in a likelihood framework. In a parsimony framework, the hypothesized tree diagram and the hypothesized evidentiary status of characters are mutually confirmatory. The philosophical directive to include all possible evidence might count as a reason to combine all data sets when running a parsimony-based analysis, in order to maximize the overall degree of confirmation (and the possibility of emergent disconfirmation). However, the philosophical directive does not mark a difference between concatenation and congruence approaches independently of the parsimony framework. Thus the measure may be useful in parsimony analyses, but it provides no justification for preferring concatenation over congruence approaches in general.

Contrary to my claim in the above paragraph, Gatesy and Barker (2005) claimed that the concept of hidden support does indeed make sense in a likelihood framework. The relevant definitions were provided by Lee and Hugall (2003) within a strictly model-based approach. Lee and Hugall described likelihood support for a particular clade (part of the tree) as the difference in log likelihood scores between optimal topologies that included and excluded that clade. Partitioned likelihood support can be used to summarize the contributions of different data sets to likelihood support at a node in a combined analysis of several data sets (Lee and Hugall 2003). By logical extension, hidden likelihood support for a clade supported by a combined analysis of multiple data sets would be the likelihood support for that clade in combined analysis, minus the sum of likelihood support scores, positive or negative, in separate analyses of the individual data sets.

While the mathematical demonstrations are valid, the concept of hidden likelihood support does not correspond to any real biological phenomenon. Likelihood methods work by optimizing evolutionary parameters in combination with an optimized tree. Optimizing subsets of the data set would lead to different values for the evolutionary parameters. There is nothing wrong with the assumption that evolutionary parameters have different values at different parts of the tree; indeed, modern likelihood methods employ that assumption. The problem is that comparing signal from the same parts of the tree is not legitimate when different evolutionary parameters are posited on those very parts of the tree.

To take a simple example (see Figure 1), suppose that tree 1 is poorly supported in a likelihood analysis that included the set of data $[a, b, c, d]$. Tree 2, which is topologically identical to tree 1, is highly supported in a likelihood analysis that included the set of data $[a, b, c, d, e, f, g, h]$. Posit that the trees share the same character assignments, and that characters $e, f, g,$ and h occur on the long branch at the right of tree 2. In this case, the trees are identical under a parsimony framework. Hidden support for the two nodes on the left of the tree has emerged after the addition of characters that do not, in themselves, provide signal for those nodes. However, the trees are not identical in a likelihood framework. Different values have been assigned to parameters

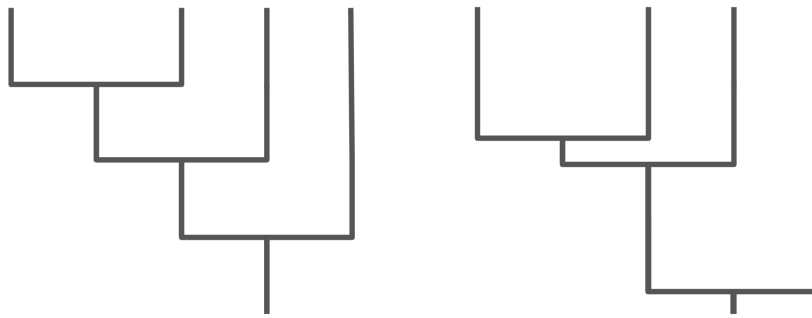


Figure 1: At left, tree 1; at right, tree 2. Explanation in text.

in the distinct trees. For example, the long branches in tree 2 may indicate the passage of a lot of time, and parameters may indicate that the rate of character change was slow during this time. In contrast, tree 1 posits approximately equal time passed between speciation events, and the parameters for character change will reflect this hypothesis.

In a likelihood framework, the two trees in this example represent different hypotheses about the taxa in question. Comparing the support scores across the distinct hypotheses simply is not comparing support scores for the same hypothesis given different sets of data.³

In an empirical study, Lee and Hugall (2003) noted that hidden support might be evident in their simultaneous maximum likelihood (ML) analysis of four genes from cetartiodactyl mammals. However, they could not confirm the presence of hidden support because branch lengths for different genes were not optimized identically in both separate and combined analyses. As predicted, models that use different sets of data optimize parameter values differently.

It is also the case that in the presence of incomplete lineage sorting (Degnan and Rosenberg 2009), subsets of data are expected to show the “wrong” signal with respect to the overall species tree. This is because some characters are indeed inherited in an order that does not match the pattern in which organismal populations split. The phenomenon is known as species-tree gene-tree discordance (see elsewhere in this volume). There is no way for the calculation of “hidden support” to distinguish between (1) correct signal being drawn from subsets that yield incorrect signal on their own, versus (2) incorrect signal being drawn from subsets that yield correct signal on their own. Thus, concatenation approaches may conflate error in the data with discordance of real entities.

Coalescence methods account for the possibility of gene-tree species-tree discordance. Indeed, coalescence methods use the predicted frequency of distinct gene tree topologies to optimize species trees. Springer and Gatesy (2016) continue to argue that hidden support provides a reason to favor concatenation versus congruence approaches, including coalescence approaches. However, I have argued that Gatesy’s hidden support arguments only make sense in a parsimony framework, and only yield the correct degree of support in the special case that distinct likelihood models assign identical values to evolutionary parameters. Gatesy and his co-authors have not resolved the theoretical problem that concatenation may conflate error with discordance.

3. It is possible that the branch lengths in each distinct likelihood analysis would turn out to be identical. In this special case, hidden support would make sense in the likelihood framework. This special case corresponds to empirical situations in which parsimony and likelihood analyses yield the same results. Cases in which branch lengths differ across analyses will be far more common.

5 Cladists on Parsimony

I have elsewhere commented on Goldman's (1990) proof that parsimony can be interpreted as a special case of the likelihood model. Here I comment on a separate proof, by Tuffley and Steel (1997), that parsimony amounts to a special case of maximum likelihood. This result explains why the arguments in favor of total evidence (section 2), simultaneous analysis (section 3), and concatenation (section 4) make sense only in a parsimony framework. The arguments only make sense in a likelihood framework in the special case that likelihood results exactly match parsimony results. This may in fact happen in some particular cases of evolution. However, in these cases, the parsimony calculation encodes a strange assumption about how evolution must work. To see this, we can examine Tuffley and Steel's (1997) proof.

Tuffley and Steel begin from a likelihood model. The probability that the data matrix obtains on the particular tree T , given the model of evolution expressed by p , is the product of the probability distributions calculated for each data point given the specified tree and evolutionary model:

$$P(X|T, p) = \prod_{i=1}^k P(x_i|T, p) \quad (1)$$

This model assumes a common mechanism for molecular change—that is, that the probability of mutation is the same at all sites. Other models assign different probabilities for different sites, for example by assigning higher probabilities to transitions versus transversions (Kimura 1980), or by altering the probabilities on the basis of relative frequency of bases (Felsenstein 1981). The Generalized Time-Reversible (GTR) model includes parameters for each substitution type plus base frequency parameters (Tavarè 1986); $GTR + I$ includes a parameter for invariant sites; and $GTR + I + \Gamma$ assigns a gamma distribution of the proportion of substitution rates across sites (Sullivan and Swofford 1997; Nguyen, Von Haeseler, and Minh 2018). The general form of a likelihood model can be expressed:

$$P(X|T, Y) = \prod_{i=1}^c P(x_i|T, y_i) \quad (2)$$

Wherein Y is the model of evolution, and y_i is the particular assignment of parameters within Y at the branch under consideration.

However, one could devise a model that assigns a separate (independent) parameter to the probability of state change of each character on every branch. In that case,

$$P(X|T, p_i) = \prod_{i=1}^k P(x_i|T, p_i) \quad (3)$$

Tuffley and Steel (1997) demonstrate that this function will be maximized precisely when the likelihood tree matches the tree produced by maximum parsimony.

This no common mechanism model is over-parameterized in such a way as to prioritize fit at the expense of simplicity (parsimony in the general sense). This mathematical demonstration reflects the above (page 5) comment that the principle of total evidence prioritizes explanatory fit at the expense of all other considerations that might favor a hypothesis. Here, a resulting problem is that the overfit model does not fit well with biological theory. Our best current biological theories indicate that the mechanism(s) of molecular change are not independent across base substitution type, let alone across every change on every branch of the tree. For example, we know that rates of molecular evolution are slow within the crocodylian clade. Rates vary across crocodile, alligator, gharial, caiman, and tomistoma species, and across loci within

individual genomes. However, the mean rate of base substitution is markedly lower in these species versus other Tetrapods (Green et al. 2014). The no common mechanism model throws out this information.

The over-parameterization is problematic in that it violates the epistemic value of simplicity, and also because it weakens the available evidence by treating individual characters as unlinked. These problems are more salient than the basic observation that all models require assumptions that may (and sometimes are known to) violate empirical data.

Some researchers have argued that all models require assumptions and that maximum likelihood and parsimony are on equal footing with respect to statistical consistency (Editors 2016). Researchers who favor parsimony claim that maximum likelihood requires making substantially more and stronger assumptions, in order to assign values to parameters within process models. The Tuffley and Steel proof may be interpreted as demonstrating that parsimony effectively requires at least as many process assumptions as maximum likelihood. Moreover, the character of the assumptions is more problematic because they weaken the evidence and violate the epistemic value of simplicity.

The same general arguments apply to ongoing debates about concatenation versus coalescence. The “first principles” espoused by the competing sides are informative. Proponents of concatenation urge care and criticize investigator errors, some of which they take to be glaring instances of sloppiness. Coalescence proponents urge the use of theoretically-informed methodology and genome-scale data sets. Coalescence-based methods involve more inferential steps, but have been implemented in computer programs, and so multi-species coalescent (MSC)-based methods are readily available to many labs and students. Pressures to use the greatest quantity of data together with the newest methods of analysis can result in researchers ignoring model assumptions and limitations. To the extent that MSC-based methods are more complex, there are more ways to use them wrongly. The possibility of investigator error when using MSC-based methods must be weighed against the theoretical limitations of concatenation methods.

In any case it is often more valuable to assess paradigms over the span of decades or longer. In the present paper I have traced the total evidence argument from Kluge’s defense of parsimony to Gatesy’s critique of coalescence. The continuous evolution of this argument, together with sociological factors (Quinn 2017), are indicators that proponents of parsimony, simultaneous analysis of character congruence, and concatenation share a set of commitments not shared by other phylogeneticists. It may be helpful to consider those who deploy total evidence arguments as members of a shared paradigm.

This total evidence paradigm has made a series of empirical assumptions in order to justify its established methodological commitments. Parsimony and character congruence proponents did not assume that homoplasy is rare, but they did assume that long branch attraction (Felsenstein 1978) is non-existent (Farris 1983), rare, or at least not problematic in practice. Concatenation proponents do not deny that gene tree discordance can occur, but they do claim that the processes leading to discordance are rare and not problematic in practice (Gatesy and Springer 2014).

It would be misleading to describe parsimony as a paradigm in decline. Rather, the parsimony/concatenation paradigm is a strand weaving in and out of systematics, making various contributions of general use, but also digging its heels in at critical moments. Both proponents and critics of the parsimony-congruence-concatenation paradigm have produced valuable empirical and theoretical insights in recent decades.

Glossary

Anomaly Zone	With respect to the MSC, empirical situations are called “in the Anomaly Zone” when the majority of gene trees do not match the topology of the species tree.
Clade	A group of organisms that includes a common ancestor together with all of its descendants. Clade is often used synonymously with “monophyletic group”, and defined as a group of species that includes an ancestral species and all of its descendants.
Coalescence	The point in a tree where gene lineages coalesce; i.e. the most recent common ancestor of gene lineages. “Coalescence methods” use the multi-species coalescent model to infer species trees.
Gene tree	The branching pattern of descent of an individual gene within organismal populations over time.
Lateral gene transfer	Transfer of genetic material between organisms through means other than direct parent-offspring descent (for example, the transfer of a gene from one lineage to another via a viral intermediate). Synonymous with horizontal gene transfer (because the transfer happens horizontally i.e. laterally between branches on a tree).
Maximum likelihood	A statistical method for finding the best fit of parameters to a model, given a set of data. In the case of the phylogeny problem, the parameters include the branching topology of a tree, the estimated amount of change along the branches of the tree, and the parameters of the assumed model of evolution; the data are some observations (such as DNA sequences) among a set of samples. The basic procedure is to adjust the free parameters of the phylogeny to find the solution that leads to the highest probability of producing the observed data.
MSC	The multi-species coalescent, a model of gene inheritance in populations of organisms across time that takes the coalescence of individual gene trees within populations into account.
Parsimony	The principle that the simplest hypothesis that explains a set of observations should be preferred over more complex hypotheses. Applied to the phylogeny problem, the parsimonious tree is the tree that minimizes the number of required evolutionary changes among the observed character states in the samples that are related on the tree.

Phylogeny	The branching pattern of descent. When referring to the branching pattern of descent of species, “phylogenetic tree” is synonymous with “species tree”. The concept is also applied to the branching pattern of descent at other biological levels, notably to hypothesize gene phylogenies.
Semaphorant	An individual organism considered at an individual point in time. According to Hennig (1966), the most conceptually basic unit of phylogenetic systematics.
Species delimitation	The research task of generating and testing hypotheses about how many distinct species are present in a group of organisms. Computational methods for evaluating gene flow and population structure are frequently used to support claims about how many species are present in the sampled population(s).
Species tree	The branching pattern of descent of organismal populations over time. “Species tree methods” (here used synonymously with “coalescence methods”) hypothesize species trees, recognizing the distinction between species trees and gene trees. This has been enabled by the MSC.
Supertree methods	Methods of phylogenetic inference that assemble a tree based on information about sub-components of the tree.
Taxonomic congruence methods	Methods that evaluate the level of agreement among phylogenetic trees or other systems for classifying biological organisms.

Acknowledgments

I thank Matt Haber and an anonymous reviewer for comments on the ms, and David Hillis and Sam Sweet for comments on the glossary. I thank James Foster, Scott Lidgard, Kevin de Queiroz, Olivier Rieppel, and Jack Sullivan for discussion. Mike Braun and Noor White organize the PhyloPizza series at the National Museum of Natural History, which enabled me to hear from speakers engaged in these debates. I thank Matt Haber for inviting me to present this research at the conference Species in the Age of Discordance and at the 2017 ISHPSSB meetings, supported by Haber’s NSF grant 1557117, “Evolution and the Levels of Lineage”. Parts of this research were funded by a Fellowship from the Notre Dame Institute for Advanced Studies and a Fellowship from the California Institute of Technology.

Literature cited

- Adanson, Michel. 1763. *Familles Des Planters Par M. Adanson*. chez Vincent.
- Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford, and P. J. Waddell. 1993. “Partitioning and Combining Data in Phylogenetic Analysis.” *Systematic Biology* 42 (3): 384–397. doi:10.1093/sysbio/42.3.384.
- Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago University Press.
- De Queiroz, A., M. J. Donoghue, and J. Kim. 1995. “Separate versus Combined Analysis of Phylogenetic Evidence.” *Annual Review of Ecology and Systematics* 26:657–681. doi:10.1146/annurev.es.26.110195.003301.

- De Queiroz, Kevin, and Steven Poe. 2001. "Philosophy and Phylogenetic Inference: A Comparison of Likelihood and Parsimony Methods in the Context of Karl Popper's Writings on Corroboration." *Systematic Biology* 50 (3): 305–321. doi:10.1080/10635150118268.
- Degnan, James H., and Noah A. Rosenberg. 2009. "Gene Tree Discordance, Phylogenetic Inference and the Multispecies Coalescent." *Trends in Ecology and Evolution* 24 (6): 332–340. doi:10.1016/j.tree.2009.01.009.
- Duhem, Pierre. 1906. *La Théorie Physique: Son Object et Sa Structure*. Chevalier et Rivière.
- Editors. 2016. "Editorial." *Cladistics* 32 (1): 1.
- Edwards, Scott V. 2009. "Is a New and General Theory of Molecular Systematics Emerging?" *Evolution* 63 (1): 1–19. doi:10.1111/j.1558-5646.2008.00549.x.
- Edwards, Scott V., Zhenxiang Xi, Axel Janke, Brant C. Faircloth, John McCormack, Travis C. Glenn, Bojian Zhong, et al. 2016. "Implementing and Testing the Multispecies Coalescent Model a Valuable Paradigm for Phylogenomics." *Molecular Phylogenetics and Evolution* 94 (Part A): 447–462. doi:10.1016/j.ympev.2015.10.027.
- Farris, J. S. 1983. "The Logical Basis of Phylogenetic Analysis." In *Advances in Cladistics*, edited by N. Platnick and V. Funk, 7–36. 2. Columbia University Press.
- Felsenstein, Joseph. 1978. "Cases in Which Parsimony or Compatibility Methods Will Be Positively Misleading." *Systematic Zoology* 27 (4): 401–410. doi:10.2307/2412923.
- Felsenstein, Joseph. 1981. "Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach." *Journal of Molecular Evolution* 17 (6): 368–376. doi:10.1007/bf01734359.
- Fitzhugh, Kirk. 2006. *The Abduction of Phylogenetic Hypotheses*. Magnolia Press.
- Gatesy, John, and R. H. Baker. 2005. "Hidden Likelihood Support in Genomic Data: Can Forty-Five Wrongs Make a Right?" *Systematic Biology* 54 (3): 483–492. doi:10.1080/10635150590945368.
- Gatesy, John, Robert W. Meredith, Jan E. Janeka, Mark P. Simmons, William J. Murphy, and Mark S. Springer. 2016. "Resolution of a Concatenation/Coalescence Kerfuffle: Partitioned Coalescence Support and a Robust Family-Level Tree for Mammalia." *Cladistics* 33 (3): 295–332. doi:10.1111/cla.12170.
- Gatesy, John, Patrick O'Grady, and Richard H. Baker. 1999. "Corroboration among Data Sets in Simultaneous Analysis: Hidden Support for Phylogenetic Relationships among Higher Level Artiodactyl Taxa." *Cladistics* 15 (3): 271–313. doi:10.1111/j.1096-0031.1999.tb00268.x.
- Gatesy, John, Daniel Sloan, Jessica Warren, Richard H. Baker, Mark P. Simmons, and Mark Springer. Forthcoming. "Partitioned Coalescence Support Reveals Biases in Species-Tree Methods and Detects Gene Trees That Determine Phylogenomic Conflicts." *bioRxiv*, no. 461699. doi:10.1101/461699.
- Gatesy, John, and Mark S. Springer. 2014. "Phylogenetic Analysis at Deep Timescales: Unreliable Gene Trees, Bypassed Hidden Support, and the Coalescence/Concatenation Conundrum." *Molecular Phylogenetics and Evolution* 80:231–266. doi:10.1016/j.ympev.2014.08.013.
- Goldman, Nick. 1990. "Maximum Likelihood Inference of Phylogenetic Trees, With Special Reference to a Poisson Process Model of DNA Substitution and to Parsimony Analyses." *Systematic Zoology* 39 (4): 345–361. doi:10.2307/2992355.
- Green, Richard E., Edward L. Braun, Joel Armstrong, Dent Earl, Ngan Nguyen, Glenn Hickey, Michael W. Vandeweyer, et al. 2014. "Three Crocodylian Genomes Reveal Ancestral Patterns of Evolution Among Archosaurs." *Science* 346 (6215): 1254449.

- Haber, Matt. 2005. "On Probability and Systematics: Possibility, Probability, and Phylogenetic Inference." *Systematic Biology* 54 (5): 831–841. doi:10.1080/106351591007444.
- Hempel, Carl G. 1965. *Aspects of Scientific Explanation*. The Free Press.
- Hennig, Willi. 1966. *Phylogenetic Systematics*. University of Illinois Press, Urbana IL.
- Hillis, David M. 1987. "Molecular Versus Morphological Approaches to Systematics." *Annual Review of Ecology and Systematics* 18:23–42. doi:10.1146/annurev.ecolsys.18.1.23.
- Huelsenbeck, John P., J. J. Bull, and C. W. Cunningham. 1996. "Combining Data in Phylogenetic Analyses." *Trends in Ecology and Evolution* 11 (152–158).
- Huelsenbeck, John P., D. L. Swofford, C. W. Cunningham, J. J. Bull, and P. J. Waddell. 1994. "Is Character Weighting a Panacea for the Problem of Data Heterogeneity in Phylogenetic Analysis?" *Systematic Biology* 43:288–291. doi:10.2307/2413470.
- Hull, David. 1983. "Karl Popper and Plato's Metaphor." Edited by V. Platnick N.; Funk. *Advances in Cladistics*, no. 2: 177–189.
- Hull, David. 1999. "The Use and Abuse of Sir Karl Popper." *Biology and Philosophy* 14 (4): 481–504. doi:10.1023/a:1006554919188.
- Kimura, M. 1980. "A Simple Method for Estimating Evolutionary Rate of Base Substitutions." *Journal of Molecular Evolution* 16:111–120. doi:10.1007/bf01731581.
- Kluge, Arnold G. 1989. "A Concern for Evidence and a Phylogenetic Hypothesis of Relationships Among Epicrates (Boidae, Serpentes)." *Systematic Biology* 38 (1): 7–25. doi:10.2307/2992432.
- Kluge, Arnold G. 1997. "Testability and the Refutation and Corroboration of Cladistic Hypotheses." *Cladistics* 13 (1): 81–96. doi:10.1006/clad.1997.0033.
- Kluge, Arnold G., and J. A. Wolf. 1993. "Cladistics: What's in a Word?" *Cladistics* 9 (2): 183–199. doi:10.1111/j.1096-0031.1993.tb00217.x.
- Kubatko, Laura Salter, and J. H. Degnan. 2007. "Inconsistency of Phylogenetic Estimates From Concatenated Data Under Coalescence." *Systematic Biology* 56 (1): 17–24. doi:10.1080/10635150601146041.
- Lee, M. S. Y., and A. F. Hugall. 2003. "Partitioned Likelihood Support and the Evaluation of Data Set Conflict." *Systematic Biology* 52 (1): 15–22. doi:10.1080/10635150309349.
- Liu, Liang, and Scott V. Edwards. 2015. "Comment on 'Statistical Binning Enables an Accurate Coalescent-Based Estimation of the Avian Tree'." *Science* 350 (6257): 171–171. doi:10.1126/science.aaa7343.
- Liu, Liang, Zhenxiang Xi, Shaoyuan Wu, Charles C. Davis, and Scott V. Edwards. 2015. "Estimating Phylogenetic Trees From Genome-Scale Data." *Annals of the New York Academy of Sciences* 1360 (1): 36–53. doi:10.1111/nyas.12747.
- Miyamoto, Michael M., and Walter M. Fitch. 1995. "Testing Species Phylogenies and Phylogenetic Methods With Congruence." *Systematic Biology* 44 (64–76). doi:10.1093/sysbio/44.1.64.
- Nguyen, Lam-Tung, Arndt Von Haeseler, and Bui Quang Minh. 2018. "Complex Models of Sequence Evolution Require Accurate Estimators as Exemplified With the Invariable Site Plus Gamma Model." *Systematic Biology* 67 (3): 552–558. doi:10.1101/185652.
- Nixon, Kevin C., and James M. Carpenter. 1996. "On Simultaneous Analysis." *Cladistics* 12 (3): 221–241. doi:10.1111/j.1096-0031.1996.tb00010.x.

- Quine, W. V. O. 1951. "Two Dogmas of Empiricism." *The Philosophical Review* 60 (1): 20–43.
- Quinn, Aleta. 2016. "Phylogenetic Inference to the Best Explanation and the Bad Lot Argument." *Synthese* 193 (9): 3025–3039. doi:10.1007/s11229-015-0908-9.
- Quinn, Aleta. 2017. "When Is a Cladist Not a Cladist?" *Biology and Philosophy* 32 (4): 581–598. doi:10.1007/s10539-017-9577-z.
- Rieppel, Olivier. 2008. "Re-Writing Popper's Philosophy of Science for Systematics." *History and Philosophy of the Life Sciences* 30 (3/4): 293–316.
- Roch, S., and M. Steel. 2015. "Likelihood-Baese Tree Reconstruction on a Concatenation of Aligned Sequence Data Sets Can Be Statistically Inconsistent." *Theoretical Population Biology* 100:56–62. doi:10.3410/f.725294372.793534069.
- Siddall, Mark. 1997. "Prior Agreement: Arbitration or Arbitrary?" *Systematic Biology* 46:765–769. doi:10.2307/2413505.
- Simmons, Mark P., and John Gatesy. 2015. "Coalescence vs. Concatenation: Sophisticated Analyses vs. First Principles Applied to Rooting the Angiosperms." *Molecular Phylogenetics and Evolution* 91:98–122. doi:10.1016/j.ympev.2015.05.011.
- Sneath, P. H. A. 2015. "Mathematics and Classification From Adanson to the Present," edited by G. H. Lawrence. Hunt Institute for Botanical Documentation.
- Sober, Elliott. 1988. *Reconstructing the Past: Parsimony, Evolution, and Inference*. 1st ed. MIT Press.
- Sober, Elliott. 1991. *Reconstructing the Past: Parsimony, Evolution, and Inference*. 2nd ed. MIT Press.
- Springer, Mark S., and John Gatesy. 2016. "The Gene Tree Delusion." *Molecular Phylogenetics and Evolution* 94:1–33. doi:10.1016/j.ympev.2015.07.018.
- Sterner, Beckett, and Scott Lidgard. 2018. "Moving Past the Systematics Wars." *Journal of the History of Biology* 51 (1): 31–67. doi:10.1007/s10739-017-9471-1.
- Sullivan, J., and D. L. Swofford. 1997. "Are Guinea Pigs Rodents? The Importance of Adequate Models in Molecular Phylogenetics." *Journal of Mammalian Evolution* 4 (2): 77–86.
- Talavera, G., and J. Castresana. 2007. "Improvement of Phylogenies After Removing Divergent and Ambiguously Aligned Blocks From Protein Sequence Alignments." *Systematic Biology* 56 (4): 564–577. doi:10.1080/10635150701472164.
- Tavarè, Simon. 1986. "Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences." *Lectures on Mathematics in the Life Sciences* 17 (2): 57–86.
- Tuffley, Chris, and Mike Steel. 1997. "Links Between Maximum Likelihood and Maximum Parsimony Under a Simple Model of Site Substitution." *Bulletin of Mathematical Biology* 59 (3): 581–607. doi:10.1007/bf02459467.
- Warnow, Tandy. 2015. "Concatenation Analyses in the Presence of Incomplete Lineage Sorting." *PLoS Currents*. doi:10.1371/currents.tol.8d41ac0f13d1abedf4c4a59f5d17b1f7.
- William, J., and O. Ballard. 1996. "Combining Data in Phylogenetic Analyses." *Trends in Ecology and Evolution* 11:334. doi:10.1016/0169-5347(96)81133-5.
- Yang, Ziheng. 2015. "The BPP Program for Species Tree Estimation and Species Delimitation." *Current Zoology* 61 (5): 854–865. doi:10.1093/czoolo/61.5.854.
- Yang, Ziheng, and Bruce Rannala. 2010. "Bayesian Species Delimitation Using Multilocus Sequence Data." *Proceedings of the National Academy of Sciences* 107 (20): 9264–9269.

© 2019 Author(s)

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits anyone to download, copy, distribute, display, or adapt the text without asking for permission, provided that the creator(s) are given full credit.

ISSN 2475-3025