

# Philosophy and Phylogenetics

Joel D. Velasco\*  
Texas Tech University

---

## Abstract

Phylogenetics is the study and reconstruction of evolutionary history and is filled with numerous foundational issues of interest to philosophers. This paper briefly introduces some central concepts in the field, describes some of the main methods for inferring phylogenies, and provides some arguments for the superiority of model-based methods such as Likelihood and Bayesian methods over nonparametric methods such as parsimony. It also raises some underdeveloped issues in the field of interest to philosophers.

---

## 1. Introduction

Biological systematics studies the diversity of life and includes comparative anatomy, taxonomy, and classification. Since we now understand that this diversity is the result of evolutionary processes, systematics also includes the task of reconstructing evolutionary history. Phylogenetics is the study and reconstruction of this history and involves the building of *phylogenies* – genealogical histories usually represented by phylogenetic trees, which depict lineage splits through time. Phylogenetics is central to biology as evidenced by the fact that the *Science Citation Index* contains over 13,000 papers published in 2011 alone containing “phylogeny” (or derivatives like “phylogenetic”) in their abstracts. This paper briefly introduces how phylogenetics is done and shows how its theories and practices are filled with numerous issues of interest to general philosophers of science. I will begin (in Section ) by introducing some relevant terminology and the problem of phylogenetic inference and then (in Section ), introducing the most common methods of inference and engaging in a brief discussion of the strengths and weaknesses of these methods.

## 2. The Central Concepts and Problems in Phylogenetics

### 2.1. THE INPUT: CHARACTER DATA

An example of a phylogenetic question is “Are humans more closely related to chimps or gorillas?” Related questions include “When did humans and chimps last share a common ancestor?” and “What was this common ancestor like?” Biologists draw on various kinds of evidence for reconstructing phylogenies, and the nature of this evidence can vary tremendously. In principle, behavioral, morphological, and molecular traits of living organisms or fossils and even biogeographical and geological data all provide evidence. These are collectively known as “character data”. Mishler (2005) and Harris and Mishler (2009) argue that we can think of phylogenetic inference as divided into two parts: the collection and organization of character data and the construction of phylogenies from these data. While most work on “phylogenetic methods” treats the character data as given, Mishler claims that the construction of the data matrix (where the traits are coded as the same or different) actually constitutes most (or even all) of the important work since the matrix

already builds in presuppositions such as what the relevant characters and traits even are. For two different traits to be coded as different states of “the same character” is roughly to assume that the characters are “homologous”. Although homology can be broadly defined as traits that have evolved from the same ancestral character, exactly what is meant by “homologous” is a central topic in biology (Hall 1994) and has been a concern to philosophers as well (e.g., Griffiths 2006). It is clear that the careful analysis of characters and hypotheses of homology play an essential role in phylogenetics, and there is a need for much more philosophical work in this area.<sup>1</sup> Mishler's (completely standard) way of describing the problem treats the analysis of characters and the construction of a data set as conceptually prior to the application of tree construction methods. But we should not be so quick to accept this. Hull (1970) and many others have argued that our trait observations are theory laden, and the application of phylogenetic methods to character data provides us evidence about our interpretation of those very data in a way similar to what Hennig (1966) called “reciprocal illumination”. While observations of limbs are evidence when inferring the phylogeny of the tetrapods, the phylogeny is itself evidence for which structures in which organisms are modifications of limbs in the first place. Our best understanding of biological taxa, characters, and phylogeny is all the result of an ongoing process of back-and-forth inferences reminiscent of a kind of reflective equilibrium.

## 2.2. THE OUTPUT: A PHYLOGENY

While it is not at all clear what is the best way to infer a phylogeny, it is usually assumed that at least the goal is straightforward: the production of a phylogenetic tree. But this is not straightforward at all. First, it is not clear what a phylogeny is a representation *of* and second, it is not clear *how* this history should be represented. While phylogenetics is sometimes described as reconstructing the “Tree of Life”, there is an ongoing debate about whether such an object even exists.<sup>2</sup> Exactly what a phylogeny represents is a matter of debate and is arguably a context-sensitive matter (Velasco 2012).

Traditionally, a phylogeny represents the history of species and is often defined in this way (Wiley, 1981: 2, Edwards 2009). But this view is hard to reconcile with building phylogenies of populations rather than species (Velasco 2008b, 2013a). In addition, most phylogenies built today are actually molecular phylogenies, which trace the evolutionary history of various biological molecules such as DNA sequences. The relationship between these molecular phylogenies and the phylogenies of groups of organisms is not clear. Some views of species conceptually link phylogenies and species with genetic histories (Baum 2009; Velasco 2010), but on more traditional views of species, gene histories and species histories can differ and a huge area of current research is focused on this “gene tree species tree problem” (Edwards 2009; Haber 2012).

As for how to represent phylogenies, originally, phylogenies were built in service to the goal of classification. Thus, there was a focus on the *topology* of the tree – the evolutionary branching order of the groups. Finding the topology of a tree is equivalent to the problem of finding out which groups are monophyletic (all of the descendants of some common ancestor). The topology certainly is of great importance. But we ought to think of phylogenetic inference more generically as inferring a number of different aspects of evolutionary history. Inferring the tree's topology is intertwined with inferring things such as the rates of change of characters and the length of the branches in time. Reconstructing these aspects of the history is just as important as reconstructing the topology. Further, a phylogenetic tree is a specific kind of graph – one with no reticulation: branches split but never rejoin. But while many of the genealogical patterns in the history of life are tree-like,

many are not. Inferring hybridizations, lateral gene transfers, genetic recombinations, and other such non-tree-like events is part of reconstructing the past and is also properly thought of as part of phylogenetic inference. This is especially clear when we aim to construct networks representing the phylogenetic histories where we do not know ahead of time whether the history is best represented by a tree or not (Morrison 2011). Setting aside these difficult and important preliminary questions, we can move ahead to the second part of Mishler's division of the problem – the construction of a phylogeny from the character data.

### 3. Phylogenetic Inference Methods

In the 1860s, systematists were forced to use their best judgment as to the plausibility of various evolutionary stories about the origins of groups and traits. Quantitative methods began to be developed in the 1960s and have dominated the field ever since. Here, I will introduce three of the most common classes of inference methods: parsimony, maximum likelihood, and Bayesian methods.<sup>3</sup>

#### 3.1. PARSIMONY METHODS

Camin and Sokal (1965) described the first parsimony method (though they thought it unacceptable) and early defenders include Kluge and Farris (1969). There are a number of variants of parsimony methods, but the basic idea assumes that we have a number of characters with discrete states. We assume the branching history is a tree, and the different possible trees are then scored by the minimum number of character state changes required to map the characters onto that tree. The fewer changes required, the better.

To get a feel for how parsimony works, let's examine Fig. 1. Imagine that we have examined organisms from three different species A, B, and C. We are considering three possible tree

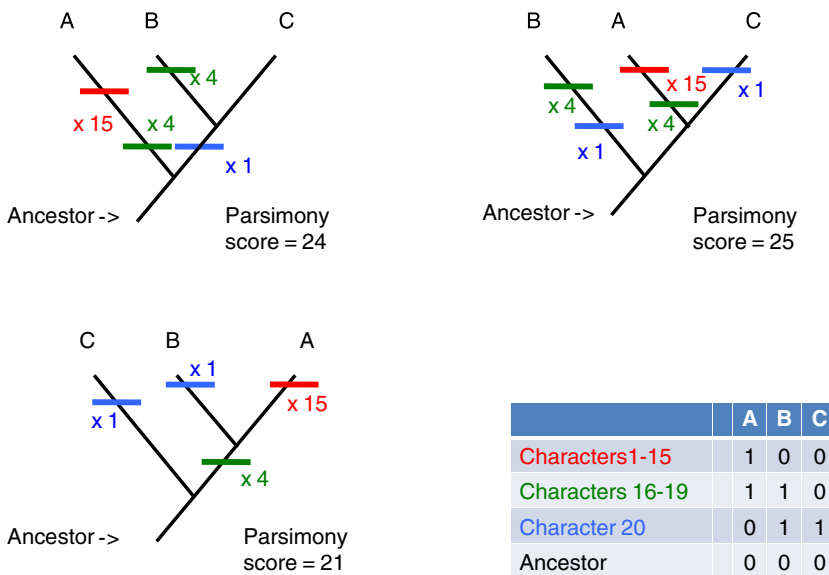


Fig. 1. On these diagrams, time flows from the root of the tree at the bottom to the tips. The colored bars represent the locations for all character state changes on these trees. If there is a change, the change gets inherited up the branch. For example, on the (AB)C tree in the bottom left, if characters 16–19 all changed from 0 to 1 after C split off but before the split leading to B and A, then C would still remain in state 0 while B and A would each be in state 1. Any other way of getting just B and A into state 1 would require more than one change per character.

hypotheses: that A is closer to B than either is to C, that A and C are closest, and that B and C are closest. These are represented as (AB)C, (AC)B, and A(BC).

In Fig. 1, we have evidence from 20 characters coded as 0 or 1 (for absent or present) and assume we know that the ancestral state for all of these characters was absent (0). Characters 1–15 are coded as 100 (meaning that As have this trait while neither Bs nor Cs do), characters 16–19 are 110, and character 20 is 011. Fig. 1 displays one way of getting the minimum number of character state changes possible for each of the three tree hypotheses. According to parsimony, tree (AB)C is to be preferred since it is possible for the data to be explained by 21 changes while the other tree hypotheses would require at minimum 24 and 25 changes. A key lesson to learn from this example is that characters 1–15 are actually “uninformative”. If we removed them from the analysis, the absolute scores of each tree would be reduced by 15, but the difference between each tree would be the same and so importantly, the ordinal ranking of trees, and so the most parsimonious tree, would be the same. While characters 1–15 seem to imply that species B and C are more similar to each other than either is to A (since they are both in state 0 while A is in state 1), they are the same as the ancestral state of the whole group and so no matter where they are placed on the tree, they provide no evidence of how evolution has proceeded. For this reason, grouping by a parsimony method leads to different results than grouping by overall similarity.

### 3.2. LIKELIHOOD METHODS

The method of maximum likelihood estimation (MLE) for statistical inference generally was first carefully developed and examined in Fisher (1922) and was taken by some early phylogeneticists to be the – *in principle* – ideal method for phylogenetic inference. To calculate the likelihood of a tree, we find the probability of the data given that particular tree combined with a particular model of how the data arose. In the case of DNA sequences, we need a model of sequence evolution. This model will have a number of adjustable parameters such as mutation rates. These rates may in turn be constrained by such assumptions as the existence of a molecular clock (that the number of mutations on a branch is proportional to its length in time) or the assumption that a mutation from any base pair to any other is equally probable (this is the assumption of the Jukes–Cantor model). Let  $\theta$  be a set of values for the parameters, which here will include the tree topology, the branch lengths, the parameters in the model of sequence evolution, and so on. The likelihood of a particular  $\theta$  is the probability of the data given that particular set of parameter values:  $P(\text{Data} | \theta)$ . The maximum likelihood estimate is the particular  $\theta$ , which maximizes this value.<sup>4</sup>

Edwards and Cavalli-Sforza (1964) were the first to discuss MLE in the context of phylogenetics. They attempted to infer the relationships between various human populations by using the genetic frequencies of various blood group alleles (for example, they estimated the frequency of the blood-type allele A1 is 29.41% in Eskimos, whereas in the Bantu it is 10.34%, the English, 20.90% and the Koreans, 22.08%). They felt that an MLE approach would be ideal, but as this was computationally prohibitive at the time, other methods were used as approximations. Felsenstein (1981) gave a computationally feasible algorithm for calculating likelihoods with molecular sequence data and with the combined rise of the amount of sequence data and the speed and availability of computers; this method has become the most frequently used in phylogenetics.

### 3.3. BAYESIAN METHODS

Bayesian inference is also a general method for statistical inference. While its history is quite a bit older than likelihood methods, it was only in the mid 1990s that Bayesian inference began

to be used in phylogenetics. In a Bayesian inference, the goal is to infer the probabilities of particular parameter values such as the tree topology or branch lengths. According to Bayes' theorem,  $P(\theta|\text{Data}) = \frac{P(\text{Data}|\theta) \times P(\theta)}{P(\text{Data})}$ .  $P(\text{Data} | \theta)$  is the likelihood while  $P(\theta)$  refers to the prior probability of the set of parameter values. If we are interested in just one parameter (say the tree topology), then we want  $P(\text{tree}|\text{Data})$ , and the other parameters become nuisance parameters. The Bayesian method for dealing with these nuisance parameters is to integrate them out by taking a weighted average over all of the different possible values of these parameters.<sup>5</sup>

Bayesian inference requires filling in prior probabilities – not just for the trees, but for the nuisance parameters as well (that is how they get integrated out). Once we have these values, the inference is computationally demanding (analytic solutions are actually impossible for most cases), but conceptually, the inference is clear and appears justified. But the problem of what to assign for priors is a difficult one.<sup>6</sup> The actual calculation of the posterior distribution requires summing over all possible values for the hundreds or even thousands of parameters for each of the nearly innumerable possible trees (for 50 taxa there are over  $10^{76}$  possible trees), which creates obvious computational difficulties.<sup>7</sup> A discussion of the mathematical and computational issues involved with Bayesian inference is beyond the scope of this paper. But I do want to point out that there is a need for philosophical work on scientific methodology where computational limits, speed, and ease of implementation are serious theoretical issues as well as practical problems. This is the situation that Bayesian phylogeneticists find themselves in.

### 3.4. WHICH METHODS TO USE?

We have just canvassed three broad phylogenetic inference strategies. Which should we use? As a matter of sociological fact, systematists will typically use multiple methods and simply publish the results of all of them. When parsimony and likelihood methods agree on a tree, this is intuitively extra evidence that the results are correct (and when they disagree, this signals that we should look more carefully at the details). But philosophical justifications of this practice are lacking.

If we treat phylogenetic inference as an instance of the general problem of statistical inference, then we inherit standard statistical methods as well as the well-known general debates. The justification for likelihood methods is straightforward. Maximum likelihood estimation is a standard statistical method with a number of virtuous properties. It is statistically consistent – that is, given a sufficiently accurate model, the MLE converges to the true value of the parameters. It is also maximally efficient – it has the smallest possible variance around the true parameter value as the size of the data set increases (Fisher 1922). The justification for Bayesian methods is also straightforward in the same sense. Bayesian inference is also a general method that can be defended on generic grounds. Debates about likelihood vs. Bayesian methods are an instance of the general debate between frequentists and Bayesians in statistics (e.g., Howson and Urbach 2006). Specific instances of this debate in phylogenetics include things such as how to interpret probabilities that appear in phylogenetic statements (Haber 2005) and whether prior probabilities can be justified (Velasco 2008a).

Both MLE and Bayesian inference are model-based methods. An important part of being a justified statistical inference from a model is that your model is itself justified in some way. In the case of DNA sequences, we have a fairly good understanding of some of the basic principles behind molecular evolution; but we do not know the full details. Usually our lack

of knowledge manifests in cases where we know what kinds of things can happen (say a genetic rearrangement), but we do not have a detailed quantitative understanding of the probability of these things happening in various circumstances. How bad is it to use imperfect models? We don't know – and central questions such as what it means for a method to be reliable are up for debate. One kind of reliability depends on the particular robustness properties of the models we use. Statistical methods have well-developed ways of checking for robustness. For example, we can slightly alter the model and measure the effects this has on our inferences; in the case of likelihood inferences, we can do a “bootstrap” procedure, effectively altering the character input slightly. In Bayesian inference, to check robustness, we can alter priors or alter models or even do various kinds of averaging procedure over priors (hyperpriors) or over models (model averaging). These, and closely related methods are part of the large area of statistical research known as model selection. For a review of model selection as applied to phylogenetics, see Sullivan and Joyce (2005).

### 3.5. PARSIMONY VS. STATISTICS

Defenders of parsimony will sometimes argue that parsimony is the logical basis of phylogenetics and (or) that it is a straight application of general principles of scientific inference such as falsificationism, the hypothetico-deductive method, Ockham's Razor, or some kind of principle of explanatory power (Wiley 1981; Farris 1983; Kluge 2005). There are serious problems with all such justifications (Sober 1988). But this does not mean that parsimony is not justified. It may, for example, be justified by reference to some other method such as likelihood. It is known that parsimony will yield identical results to likelihood methods in a range of cases, but it is known to come apart in a range of others (Sober 2004). It is sometimes thought that the justification for parsimony stems from an assumption that evolution is parsimonious, but this is not how defenders see it. What exactly is assumed by parsimony is unclear (Sober 2005), but it would be wrong to say that just because parsimony isn't an explicit model-based method that it is not based on any implicit model either.

While parsimony methods may often deliver the correct answer in the sense of the correct tree topology, there are several reasons why explicit model based methods are superior. Model based methods build in assumptions, but these assumptions are flexible and can be directly tested. For example, parsimony standardly assumes that each character state change counts equally. But in many cases, we know that this does not accurately measure the evidential impact of different changes. Some changes are far more probable than others. The number of limbs that an animal has is a far better guide to genealogical relationships than the number of teeth, which is itself a better guide than the number of hairs. In the case of molecular data, we know that some kinds of mutations are more probable than others. Model based methods easily accommodate such knowledge. Parsimony can take this kind of information into account as well with a weighting scheme, though this seems to count against some of the arguments that have been given for Parsimony methods. In fact, unweighted parsimony *is* a weighting scheme – it simply weights all possible changes equally. Seen in this light, equal weights require just as much justification any other particular weighting scheme, and it is not obvious how to justify such schemes.

The relative inflexibility of the built-in assumptions of parsimony methods lead to other problems. Holder and Lewis (2003) point out that there are often a number of different possible scenarios that could have led to the same tree. But whether there is 1 path or 20 to a certain tree, the parsimony score of the tree remains the same. Parsimony also treats different traits as evolving independently. With model-based methods, models can be created



to correct for these issues. In principle, there is very little restriction on the kinds of models that are possible (though they may become non-identifiable if there are too many parameters). Parsimony methods build in certain assumptions that simply cannot be altered and so are not flexible enough to capture these, and other specific evidential worries.

The most common criticism of parsimony is due to the phenomena of long-branch attraction (Felsenstein 1978). Parsimony assumes that the same change counts the same regardless of which branch it occurs on. But we should obviously expect more changes on longer branches so the evidential value of such changes should be less. This phenomenon leads to the problem that parsimony can be statistically inconsistent – that is, the method can converge on a false answer in specific cases. Imagine that there are two non-sister taxa where the rates of change on the branches leading to these taxa have been much higher than on other branches. Sheer chance alone will guarantee that some of these changes are convergences and parsimony will count these convergent traits as evidence that these two branches are closely related (as it should). However, if the branches are long enough relative to the other branches, then we would actually expect there to be more of these coincidentally convergent matches than genuine homologous matches with the other branches that they are actually closer to. Thus, the two long branches will tend to be “pulled together” in that we infer they are closely related. As we increase the amount of data, this pull will get stronger, and thus, parsimony will converge on the wrong answer. In cases where we expect the branches to be long (say a phylogeny of distinct bacterial phyla), chance variation alone will virtually guarantee that some branches are much longer than others, and the problem will arise.

Parsimony as a phylogenetic method has statistical properties. We can measure its robustness in terms of accuracy across various kinds of cases, we can test its statistical properties (like consistency and efficiency), and we can directly compare it to model based methods such as maximum likelihood. Just because parsimony does not use explicit parametric models of evolution doesn't mean that it is somehow a priori a more general (or better!) method than explicit model based accounts (contra Farris 1983; Kluge 2005, and others). In fact, as I have argued, by being inflexible in a certain way, it makes matters worse.

#### 4. Conclusion

We have canvassed just a few of the philosophical issues that arise in phylogenetic inference. There are of course numerous issues worthy of study that we did not delve into such as understanding the connection between phylogenetics and other fields such as taxonomy and classification (e.g., Rieppel and Kearney 2007). In this piece, I hope to have shown that traditional problems in the philosophy of science relating to confirmation and evidence arise in the particular case of phylogenetics, and so philosophers are well posed to make contributions, but also that by examining phylogenetics on its own terms, we can make contributions to a discipline of enormous scientific importance while at the same time, gaining some perspective and insight into these broader issues or more general concern.

#### *Short Biography*

Joel Velasco's research focuses on questions in biological systematics such as the nature of species and the tree of life as well as in phylogenetic inference. He has authored papers in these areas for *Biology and Philosophy*, *The British Journal for the Philosophy of Science*, *Philosophy of Science*, *Studies in History and Philosophy of Biological and Biomedical Sciences*, and elsewhere. As of the fall of 2013, he is an Assistant Professor at Texas Tech

University. He received his PhD in 2008 from the University of Wisconsin–Madison, and since then has been a Mellon Fellow at Stanford University, a Visiting Professor at Cornell University, and a Post-Doctoral Instructor at the California Institute of Technology.

### Acknowledgements

I would like to thank Matt Barker, Kenny Easwaran, Marc Ereshefsky, Laura Franklin–Hall, Luke Glynn, Matt Haber, Casey Helgeson, Grant Ramsey, Matt Slater, and Elliott Sober as well as an anonymous referee for providing helpful comments on drafts of this paper.

### Notes

\* Correspondence: Philosophy Department, Texas Tech University, Box 43092, Lubbock, TX 79409, USA. Email: joel@joelvelasco.net.

<sup>1</sup> But see, e.g., Richards (2003), Griffiths (2006), Rieppel and Kearney (2007), and Ereshefsky (2009, 2012).

<sup>2</sup> For detailed arguments, see the special issues on the tree of life in *Biology and Philosophy* (2010) and *Biology Direct* (2011). For brief overviews, see O'Malley et al. (2010) and Velasco (2013b).

<sup>3</sup> Helpful introductory reading on phylogenetic methods includes Swofford *et al.* (1996), Huelsenbeck and Crandall (1997), Holder and Lewis (2003), and Felsenstein (2004). For a philosophical take, see Haber (2009)

<sup>4</sup> Since we are often especially interested in the tree topology, we sometimes say that  $P(\text{Data}|\theta)$  is the likelihood of the tree which is part of  $\theta$ . This is technically correct if we fix the value of the nuisance parameters at their MLE values.

<sup>5</sup> In the frequentist MLE method, for each tree, nuisance parameters are set at the value that would maximize the probability of the data on that particular tree. Suboptimal parameter values are irrelevant on this view, but they factor into the likelihood on the Bayesian view. Thus, confusingly, the likelihood of the tree used in Bayes' theorem is not the same as the tree's likelihood score used for maximum likelihood inferences.

<sup>6</sup> Pickett and Randle (2005) claim that it is impossible to assign priors in a consistent way. While this particular argument is based on a mistake (Velasco 2008a), the problem remains a difficult one. Velasco (2008a) advocates priors on tree topologies based on the Yule birth–death process, but leaves open the more difficult problem of priors on branch lengths or model parameters. Alfaro and Holder (2006) attempt to address some of these issues.

<sup>7</sup> Markov Chain Monte Carlo methods (MCMC) attempt to solve this problem with simulation and sampling. While MCMC methods had been used for several decades to solve other problems in Bayesian statistics, they were introduced by example into the phylogenetics literature with Yang and Rannala (1997) and in more detail in Larget and Simon (1999).

### Works Cited

- Alfaro, Michael E. and Mark T. Holder. 'The Posterior and the Prior in Bayesian Phylogenetics.' *Annual Review of Ecology and Systematics* 37 (2006): 19–42.
- Baum, David A. 'Species as Ranked Taxa.' *Systematic Biology* 58 (2009): 74–86.
- Camin, J. H. and R.R. Sokal. 'A Method for Deducing Branching Sequences in Phylogeny.' *Evolution* 19 (1965): 311–326.
- Edwards, A. W. F. and L. L. Cavalli-Sforza. 'Reconstruction of Evolutionary Trees.' *Phenetic and Phylogenetic Classification*. Eds. V. H. Heywood and J. McNeill. London: Systematics Association Publ. No. 6, 1964. 67–76.
- Edwards, Scott V. 'Is a New and General Theory of Molecular Systematics Emerging?' *Evolution* 63.1 (2009): 1–19.
- Ereshefsky, Marc. 'Homology: Integrating Phylogeny and Development.' *Biological Theory* 4.3 (2009): 225–229.
- . 'Homology thinking.' *Biology and Philosophy* 27.3 (2012): 381–400.
- Farris, James S. 'The Logical Basis of Phylogenetic Analysis.' *Advances in Cladistics*, vol. 2. Eds. N. I. Platnick and V. A. Funk. New York: Columbia University Press, 1983. 7–36.
- Felsenstein, Joseph. 'Cases in which Parsimony or Compatibility Methods will be Positively Misleading.' *Systematic Zoology* 27 (1978): 401–10.
- Felsenstein, Joseph. 'Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach.' *Journal of Molecular Evolution* 17.6 (1981): 368–376.
- . *Inferring Phylogenies*. Sunderland, Mass: Sinauer Associates, 2004.
- Fisher, Ronald A. 'On the Mathematical Foundations of Theoretical Statistics.' *Philosophical Transactions of the Royal Society of London, A* 222 (1922): 309–368.



- Griffiths, Paul. 'Function, Homology, and Character Individuation.' *Philosophy of Science* 73 (2006): 1–25.
- Haber, Matthew H. 'Multilevel Lineages and Multidimensional Trees: The Levels of Lineage and Phylogeny Reconstruction.' *Philosophy of Science* 79 (2012): 609–623.
- . 'Phylogenetic Inference.' *Blackwell Companion to the Philosophy of History and Historiography*. Ed. Avi Tucker. Boston: Wiley-Blackwell. 2009. 231–242.
- . 'Probability and Systematics: Possibility, Probability, and Phylogenetic Inference.' *Systematic Biology* 54 (2005): 831–841.
- Hall, Brian K., ed. *Homology: The Basis of Comparative Biology*. San Diego: Academic Press, 1994.
- Harris, Eric S. J. and Brent D. Mishler. 'The Delimitation of Phylogenetic Characters.' *Biological Theory* 4.3 (2009): 230–234.
- Hennig, Willi. *Phylogenetic Systematics*. Chicago: University of Chicago Press, 1966.
- Holder, Mark and Paul O. Lewis. 'Phylogeny Estimation: Traditional and Bayesian Approaches.' *Nature Reviews. Genetics* 4 (2003): 275–284.
- Howson, Colin and Peter Urbach. 'Scientific Reasoning.' *The Bayesian Approach (3rd ed.)*. Chicago: Open Court, 2006.
- Huelsenbeck, John P. and Keith A. Crandall. 'Phylogeny Estimation and Hypothesis Testing Using Maximum Likelihood.' *Annual Review of Ecology and Systematics* 28 (1997): 437–466.
- Hull, David L. 'Contemporary Systematic Philosophies.' *Annual Review of Ecology and Systematics* 1 (1970): 19–54.
- Kluge, Arnold G. 'What is the Rationale for 'Ockham's Razor' (a.k.a. Parsimony) in Phylogenetic Inference?' *Parsimony, Phylogeny and Genomics*. Ed. V. A. Albert. New York: Oxford University Press, 2005. 15–42.
- Kluge, Arnold G. and James S. Farris. 'Quantitative Phyletics and the Evolution of Anurans.' *Systematic Zoology* 18 (1969): 1–32.
- Larget, Bret and Simon, Donald L. 'Markov Chain Monte Carlo Algorithms for the Bayesian Analysis of Phylogenetic Trees.' *Molecular Biology and Evolution* 16.6 (1999): 750–759.
- Mishler, Brent D. 'The Logic of the Data Matrix in Phylogenetic Analysis.' *Parsimony, Phylogeny and Genomics*. Ed. V.A. Albert. New York: Oxford University Press, 2005. 57–70.
- Morrison, David A. *Introduction to Phylogenetic Networks*. Uppsala, Sweden: RJR Productions, 2011.
- O'Malley, Maureen A., William Martin, and John Dupré. 'The Tree of Life: Introduction to an Evolutionary Debate.' *Biology and Philosophy* 25 (2010): 441–453.
- Pickett, Kurt M. and Christopher P. Randle. 'Strange Bayes Indeed: Uniform Topological Priors Imply Non-Uniform Clade Priors.' *Molecular Phylogenetics and Evolution* 34.1 (2005): 203–211.
- Richards, Richard. 'Character Individuation in Phylogenetic Inference.' *Philosophy of Science* 70 (2003): 264–279.
- Rieppel, Olivier and Maureen Kearney. 'The Poverty of Taxonomic Characters.' *Biology and Philosophy* 22 (2007): 95–113.
- . 'The Contest between Likelihood and Parsimony.' *Systematic Biology* 53 (2004): 644–653.
- . 'Parsimony and its Presuppositions.' *Parsimony, Phylogeny and Genomics*. Ed. V. A. Albert. New York: Oxford University Press, 2005. 43–53.
- Sober, Elliott. *Reconstructing the Past: Parsimony, Evolution, and Inference*. Cambridge, Massachusetts: MIT Press, 1988.
- Sullivan, Jack and Paul Joyce. 'Model Selection in Phylogenetics.' *Annual Review of Ecology, Evolution, and Systematics* 36 (2005): 445–466.
- Swofford, David L., Gary J. Olsen, Peter J. Waddell, and David M. Hillis. 'Chapter 11: Phylogenetic Inference.' *Molecular Systematics*. Eds. David. M. Hillis, Craig Moritz, and Barbara K. Mable. Sunderland, MA: Sinauer, 1996. 407–514.
- Velasco, Joel D. 'The Future of Systematics: Tree-Thinking Without the Tree.' *Philosophy of Science* 79 (2012): 624–636.
- . 'Phylogeny as Population History.' *Philosophy and Theory in Biology*, (2013a): 5:e402. DOI: <http://dx.doi.org/10.3998/ptb.6959004.0005.002>.
- . 'The Prior Probabilities of Phylogenetic Trees.' *Biology and Philosophy* 23 (2008a): 455–473.
- . 'Species Concepts should Not Conflict with Evolutionary History, but Often Do.' *Studies in History and Philosophy of Biological and Biomedical Sciences* 39 (2008b): 407–14.
- . 'Species, Genes, and the Tree of Life.' *The British Journal for the Philosophy of Science* 61 (2010): 599–619.
- . 'The Tree of Life.' *The Cambridge Encyclopedia of Darwin and Evolutionary Thought*. Ed. Michael Ruse. New York: Cambridge University Press, (2013b): 340–345.
- Wiley, E. O. *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. New York: Wiley-Interscience, 1981.
- Yang, Ziheng and Bruce Rannala. 'Bayesian Phylogenetic Inference Using DNA Sequences: A Markov Chain Monte Carlo Method.' *Molecular Biology and Evolution* 14.7. (1997): 717–724.