20

Phylogenetic Inference

MATT HABER

Introduction

The task of reconstructing the history of life on earth falls, in part, to systematists. Practitioners of systematics study the historical pattern of evolution among groups of living things, i.e., phylogeny. Phylogeny is not directly observable; instead, it must be inferred from data that are incomplete at best. Systematists face a set of problems that is similar to those confronted by anyone who makes claims about the past: is there sufficient evidence? Given the available evidence, what questions can or should be asked and answered? What methods render the most reliable or justifiable conclusions? What are the limits of these methods?

While the shape of these problems is familiar to philosophers and historians, there are particular problems of inference in a field where evidence can be so scarce or difficult to interpret. We can observe fossils, which may give information about common ancestors of living taxa. However, some ancient forms presumably left no descendants, so systematists do not always know what to look for. This situation is compounded by poor or non-existent fossils of soft body parts, as well as by a spotty fossil record (Vermeij 2006; Sober and Steel 2002). There are, of course, other kinds of clues to phylogenetic relationships. The genetic, morphological, and behavioral characters of living taxa suggest patterns of ancestry. This evidence is sometimes problematic, as it is often difficult to discern whether these characters are similar because of common ancestry or for other reasons (Sober and Steel 2002).

So what can biologists meaningfully say about phylogeny? This is one of the central problems in systematics. Broadly, two different issues have been at the center of recent systematics debates: given epistemic limitations, whether *any* inference of phylogeny may justifiably be drawn; and given an affirmative answer, what methods ought biologists use to justifiably infer phylogenies, and what are the limits of these inferences? Here I will place these debates in their historical, scientific, and conceptual contexts.

There is general agreement in the systematics community that meaningful inferences of phylogeny may be drawn. Contemporary systematists are typically called "*phyloge-netic systematists*" or "*cladists*" to reflect their commitment to the task of reconstruct-ing these relationships (though, as will be seen below, debates over methodological tools have given rise to controversy over the application of these labels). Though the issue

of whether phylogeny reconstruction is a legitimate task for biologists is largely settled, the shape of that debate set the tone for contemporary conceptual debates in systematics. In the next section I describe what was at stake in that debate, and how it left the nature of phylogenetic inference as a central conceptual issue in systematics. With the important issue of *whether* systematics is possible largely settled, there have been several debates centered on *how* to go about it. Phylogeny reconstruction methodology, and the subsequent shape and justification of phylogenetic inference, are still very much live issues. I provide an overview of recent and current disputes in section three.

From Art to Science: An Introduction to Schools of Thought

A demand for rigor: numerical taxonomy

The publication of Linnaeus' Systema Naturae in 1735 may mark the beginning of modern systematics, and the publication of Darwin's On The Origins Of Species obviously had a major impact on biologists. However, it is only rather recently that the broad goal of systematics has been to display the evolutionary relations of taxa using phylogenetic trees (figure 20.1) (Hull 1988). The *cladistic* school arose in the late 1960s/early 1970s and challenged what was known as *numerical taxonomy,* or *phenetics*. *Pheneticists* advocated using similarity algorithms as a method of classification of taxa. This involved quantitatively coding characters of the taxa in question (e.g., morphological traits such as fur length), then using a similarity matrix to cluster those groups that were most similar (figure 20.1b). Notably, pheneticists did *not* think that the evolutionary relationships of these groups could be validly inferred, and strove for what they claimed would be an objective classification based on observable character states:

> Before proceeding, it is necessary that we clearly define our use of the term "numerical taxonomy." We mean by it *the numerical evaluation of the affinity or similarity between taxonomic units and the ordering of these units into taxa on the basis of their affinities.* (Sokal and Sneath 1963: 48, emphasis in original).
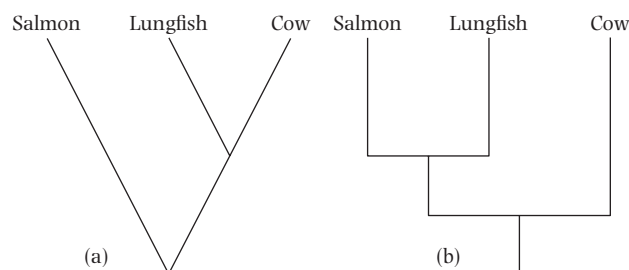

Figure 20.1   A phylogenetic classification (a) groups taxa by evolutionary relatedness, as defined by most recent common ancestor. The result is a cladogram displaying nested sets of clades. Phenetic classifications (b) produce phenograms that group taxa according to similarity indexes, and do not display evolutionary relationships

Numerical taxonomy arose in response to entrenched methods of taxonomy – what came to be known as evolutionary taxonomy. The primary criticism from pheneticists was that taxonomy was being practiced more as art than science. Pheneticists argued that taxonomists ought to strive for a *scientific* approach to taxonomy – namely, using methods that were both *objective* and *repeatable* (Sokal and Sneath 1963: 49). This amounted to a demand for rigor and articulation of method, i.e., a dispute over what could be justified as a scientific explanation or hypothesis.

For advocates of pheneticism, a scientific approach to taxonomy amounted to using transparent, explicit methods that produced repeatable outcomes that could be tested by anyone familiar with the methodology and with access to the data set. This stood in contrast to traditional practice in taxonomy, which pheneticists charged relied primarily on intuition and appeals to authority. Testability, repeatability, and transparency of methodology was tied to a notion of objectivity – in particular, to the notion that overall similarity between organisms is a function of the similarity of the characters being compared between the organisms (Sokal and Sneath 1963: 50, axiom 3). The pheneticist line was that incorporation of more characters would produce better overall similarity measures, and group taxa in the most informative way. (Notice, too, the emphasis placed on characters, as opposed to history. This concerns a related debate over the nature of biological taxa.)

The "objectivity" criterion that anchors the other scientific criteria identified by pheneticists is easily misunderstood. "Objective" was not meant to imply that the affinities between entities were theory free; simply that given some set of parameters and data, a mean similarity could be determined without resort to subjective interference (Sokal and Sneath 1963: 268–9). Whether the notions of "objective" advocated by pheneticists in theory were consistent with phenetic methodology in practice, and whether these notions are coherent is a matter of some controversy (Hull 1970).

Phenetic claims of objectivity should also not be understood as equivalent to staking out a realist position with regard to hypothesized taxa. These are independent issues, though often conflated; pheneticists could just as easily be instrumentalists as realists with regard to proposed taxonomies. That said, there was an underlying assumption in numerical taxonomy that selection and identification of characters was a strictly empirical procedure, and that theory should not enter into such selections. This is resonant in the resistance pheneticists exhibited towards partitioning of data into informative and non-informative, ultimately producing a clash with cladists over what counts as evidence:

> Until and unless methods are developed for objectively assessing and quantifying the phylogenetic significance of character differences or affinities, the consideration of such information is incompatible with our stated aim of objectivity and repeatability for the taxonomic process. (Sokal and Sneath 1963: 55)

Objectivity: the phylogenetic perspective

Phylogeneticists argued that no such objective classification was possible based on the inductive methods of phenetics, and that inferences of evolutionary relationships of taxa could, indeed, be justified. Phylogeneticists followed Willi Hennig in taking the aim of

systematics to be the construction of phylogenetic trees that reflected nested sets of sister-taxa, that is, taxa that are descended from a common ancestor (figure 20.1a) (Hennig 1966). These sets of sister-taxa are *clades* or *monophyletic groups* (i.e., an ancestor and all and only its descendants), hence the term *cladistic analysis*. Phylogenetic analyses produce *cladograms* that display phylogenetic relationships, but do not necessarily convey any objective similarity between the taxa. This reflects the phylogenetic claim that morphological similarity alone does not necessarily correlate to evolutionary relatedness, and that only derived traits were informative in systematic analysis. Not surprisingly, phenetic and phylogenetic analyses of data can produce significantly different results (figure 20.1).

Indeed, cladist criticisms of phenetics call into question the entire notion of objective similarity in evolutionary analysis (Hull 1970). In its place, cladists argued that in a phylogenetic context relevant similarity should be understood in terms of shared, derived history. This argument amounts to the claim that absent some theoretically justifiable notion of relevance, similarity is an empty claim (Griffiths 1974; see also Goodman 1972). Specifying phylogenetic context as a theoretical basis for relevance provides content to claims of similarity. This is an instance of a more general issue concerning the nature of how similarity relations hold between theoretical models and designated systems (Giere 1988; Teller 2001; Callender and Cohen 2006).

The pheneticists' notion of objectivity was also called into question in a closely related debate over the nature of higher taxa. Many cladists embraced the thesis that taxa (including species) are historical individuals/systems, as opposed to classes (Ghiselin 1974; Griffiths 1974; Hull 1976). Committing to this tenet recasts much of the debate, *e.g.*, over what constitutes a "biologically interesting group" – characters or ancestry (Griesemer, 2000; Hamilton and Haber 2006; Ghiselin 2007).

Cladists rejected the notion that objective similarity was something that could be discovered using quantitative analysis, instead advocating a systematic analysis justified by appeal to theoretically driven notions of relevance. Nonetheless, cladists *embraced* the notions of repeatability and testability as criteria of scientific methodology that must be met to produce satisfactory hypotheses of phylogeny. These criteria concern which techniques are suitable for inferring phylogeny; hence the shift in systematics from debates over *whether* to infer phylogeny to debates concerning *how* to infer phylogeny.

How to Infer Phylogeny, Or, Why Some Cladists Aren't "Cladists"

The phylogenetic technique of choice among cladists was broadly known as parsimony (Hull 1988; Sober 1988). Though there are some variants of parsimony, for the purpose at hand these can be ignored. Which phylogenetic tree is most parsimonious depends, of course, on what is getting counted. Proponents of parsimony analysis count evolutionary events, i.e., hypotheses of the evolution of a trait. The most parsimonious tree is that one that requires the least number of evolutionary events yet is consistent with the observed data – *viz.*, the distribution of characters across groups. Phylogenetic trees could, however, be ranked for parsimony based on other features. This is just to say that simplicity is not simply read directly off of these phylogenetic models, but is itself a claim in need of justification or explanation. This is reminiscent

of modern formulations of the problem of induction (Goodman 1955). All this leaves aside another, more fundamental, issue: the nature of parsimony in science. (See Sober 1988 for a discussion of the role parsimony plays in scientific reasoning, and Skipper (2002) and Plutynski (2005) for an example of how different positions in that debate might be applied to cases in biology.)

Many leading cladistic theorists initially justified parsimony techniques by explicitly appealing to Karl Popper's falsificationism (Wiley 1975; Eldredge and Cracraft 1980; Farris 1983). Briefly, falsificationism is a philosophical thesis about the scientific method. On Popper's account, scientific theories and hypotheses should not be evaluated on the basis of confirmation (or verification), but rather of falsification (Popper 1959). Experiments producing confirming evidence of a theory are of little to no value; only experiments that test scientific theories by seeking falsifying evidence offer evaluative information conforming to the scientific method. For Popper, the best theories are bold hypotheses that have passed rigorous attempts at falsification, and are, thus, corroborated.

Popper presented falsificationism as the scientific method that successfully solves the problem of induction. Cladists pressed falsificationism into service to solve the problem of phylogenetic inference, while simultaneously satisfying the criteria of scientific methodology inherited from prior debates in systematics. Phenetic methods were decried as confirmationist, while parsimony analysis was framed in a falsificationist framework (Wiley 1975; Farris 1983). The most parsimonious phylogenetic tree was said to be the most corroborated phylogenetic hypothesis, and a bold hypothesis of the evolutionary relationships between the taxa being studied. This phylogenetic tree was subject to being falsified if the discovery or addition of new characters revealed a more parsimonious phylogenetic tree.

The reconstruction of phylogeny using parsimony involves the construction of a tree-like (i.e., bifurcating) model to represent a section of the actual phylogenetic pattern of a historical lineage (e.g., new branches represent a speciation event). Note the importance of the distinction between the "true tree" and a "phylogenetic tree." The "true tree" is the actual historical lineage of life, (or part of that actual historical lineage). The "phylogenetic tree," on the other hand, is a historiographic model that systematists treat as a hypothesis about the structure of the actual lineage (Haber 2005). Parsimony provides a way to both construct and evaluate different hypotheses of a section of the actual historical lineage.

In the late 1970s, phylogeneticist Joseph Felsenstein discovered that lineages of a certain shape were subject to a systematic error in parsimony analysis (Felsenstein 1978, 1979, 1981, 2004). One example of this problem is called long-branch attraction (after the shape of the lineage described by Felsenstein; figure 20.2). In phylogenetic trees, branches connect nodes to other nodes or taxa. The length of these branches can be used to represent the amount of evolutionary change along a branch. In long-branch attraction, the taxa at the ends of long branches of lineages are mistakenly grouped together by parsimony instead of with the groups with which they actually share a more recent common ancestor (figure 20.2b).

Long-branch attraction is not merely a problem at an operational level, but also presents a challenge to the falsificationist underpinnings of parsimony. As more data (in the form of characters) are discovered or added to analyses, the previously most
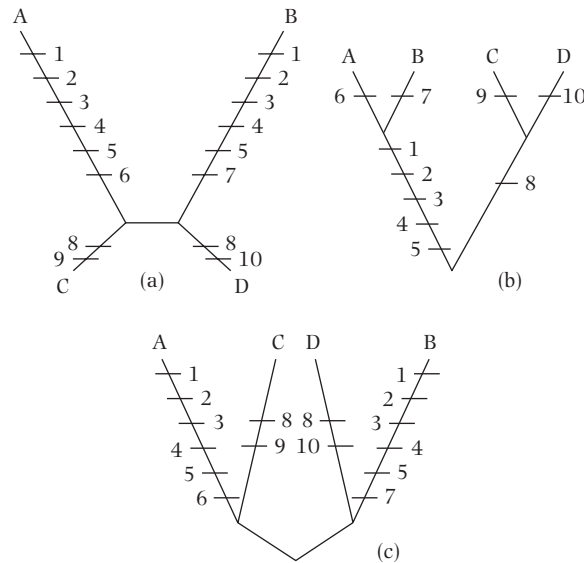
MATT HABER

Figure 20.2 (a) represents the actual pattern of evolution (i.e., the actual historical lineage of A, B, C, and D), with the branch lengths indicating amount of evolutionary change. This tree is unrooted, i.e., it is not time directional. The phylogenetic trees at (b) and (c) are rooted, with extant taxa at the labeled tips. (b) is a parsimony analysis of the data from (a), whereas (c) is an ML analysis of the same data. Note that the evolutionary relationships displayed in the parsimony phylogenetic tree are nor isomorphic with the actual evolutionary relationships. The long-branch error occurs in (b) in hypotheses based on the number of proposed evolutionary events needed to account for all the characters exhibited by the taxa. For example, hypothesis (b) would receive a parsimony score of 10, whereas (c) would receive a parsimony score of 16

parsimonious phylogenetic trees are subject to being falsified in favor of more parsimonious phylogenetic trees. These new phylogenetic trees, then, are held up as bold hypotheses and either corroborated or rejected in favor of ever more parsimonious trees that are subject to being tested. Felsenstein showed that as more characters are added, parsimony techniques become *more subject* to making a long-branch attraction error. This is because as more characters are added, it raises the possibility that there will be a long branch in the tree. In other words, parsimony techniques are more prone to rejecting hypotheses that correctly capture actual phylogenetic relations while corroborating less accurate but more parsimonious phylogenetic hypotheses – seemingly producing a systematic inferential error!

To avoid long-branch attraction, Felsenstein proposed using maximum-likelihood (ML) techniques to formulate phylogenetic trees (Felsenstein 1981). ML picks out the phylogenetic tree that has the highest likelihood value (conditional on the data); i.e., the phylogenetic hypothesis that confers the highest probability on the data is the phylogenetic tree given the highest ML value. Simulation studies have shown that ML methods are not subject to long-branch attraction (*fig. 2c*) (Huelsenbeck and Hillis 1993; Hillis et al. 1994; Swofford et al. 1996; Huelsenbeck and Rannala 1997).

Many cladistic theorists did not welcome ML methods. A split formed in the systematics community between those that advocated using parsimony techniques exclusively, and those that advocated using statistical techniques such as ML. The former group appropriated the name *Cladists* (whom I shall call capital "C" cladists); the latter group identified themselves as *statistical phylogeneticists*. Statistical phylogeneticists typically consider parsimony methods to be just another statistical method (or, at least, as a method that can be construed as using or derived from statistical methods rather than from a falsificationist framework (Tuffly and Steel 1997; Sober 2004)). Cladists, on the other hand, do not consider statistical techniques to be valid forms of phylogenetic inference, and some question whether statistical methods conform to what they consider the "scientific method," viz., a particular form of falsificationism (Farris 1983; Kluge 1997a, 1997b). The debate is primarily over the justification of techniques and subsequent inferences that may be drawn. Indeed, Cladists have argued that parsimony analysis itself should be rejected as unsubstantiated if formulated using statistical methods:

> If reasoning from unsubstantiated suppositions cannot legitimately question parsimony, then neither can it properly bolster that criterion. The statistical approach to phylogenetic inference was wrong from the start for it rests on the idea that to study phylogeny at all, one must first know in great detail how evolution has proceeded. That cannot very well be the way in which scientific knowledge is obtained. (Farris 1983: 17)

Just as long-branch attraction poses a conceptual challenge to Cladists, the resistance to explicitly statistical techniques is grounded in conceptual concerns. At stake is whether particular methods satisfy some scientific criteria – typically, whether ML techniques satisfy falsificationist criteria or are best understood as embodying a verificationist scientific method (in which case, the argument runs, they ought to be rejected on principle). This dispute has two common fronts: (1) whether statistical techniques can be properly construed in a falsificationist framework; and (2) how to best understand falsificationism (Siddall and Kluge 1997; de Queiroz and Poe 2001; Kluge 2001; de Queiroz and Poe 2003). As a result, literature in systematics can read like Popper studies.

Another question is the extent to which the falsificationism espoused by systematists actually resembles that discussed in the philosophical literature (see Farris 1983; Hull 1983, 1999; Sober 1988). There do seem to be at least some important differences, which may have bearing on the debates over phylogenetic inference (Gillies 1990; Urbach 1991; Siddall and Kluge 1997).

Most contemporary philosophers of science are critical of the idea that falsificationism is the *only* acceptable method of science (Sober 2000). Many question whether it is even a very good thesis about scientific methodology (Howson and Urbach 1993), and most post-Quine and Kuhn philosophers of science have come to reject the thesis that any scientific hypothesis can be falsified in isolation. So it is perhaps somewhat surprising to philosophers of science that very few systematists have argued against the premise that falsificationism is *the* desired scientific methodology (though see Sober and Steel 2002, for an account of parsimony presented to systematists that is openly distanced from falsificationism). This situation is likely to change with the

MATT HABER

emergence of a new group within statistical phylogenetics: *Bayesians* (Huelsenbeck and Ronquist 2001; Alfaro and Holder 2006).

Bayesian phylogeneticists, as the name would imply, use Bayesian phylogenetic techniques to construct phylogenetic trees. The purpose of this analysis is to distribute posterior probabilities (probabilities that take into account evidence and prior knowledge) over a range of possible phylogenetic trees. Bayesian phylogeneticists use what are known as Markov Chain Monte Carlo (MCMC) algorithms to approximate the rational distribution of these posterior probabilities that lead to the selection of which phylogenetic trees (or consensus trees) are most probable (Larget and Simon 1999; Huelsenbeck and Ronquist 2001; Huelsenbeck et al. 2001; Huelsenbeck et al. 2002; Alfaro and Holder 2006). Though non-Bayesian statistical phylogeneticists typically consider Bayesians fellow statistical phylogeneticists, some harbor skepticism towards the Bayesian phylogenetic methods. Indeed, this points to what will likely constitute the most important debates concerning phylogenetic inference: (1) the nature of specification and distribution of relevant "priors" for phylogenetic analysis, i.e. how do biologists evaluate the prior probabilities of phylogenic trees prior to considering new evidence that causes the posterior redistribution of probabilities; and (2) the use of these posterior probabilities as justification of phylogenetic inference (Alfaro and Holder 2006). Another problem that Bayesian statistical phylogeny shares with philosophical discussion of causation in general is the assumption of Markov conditions; namely, that "A variable represented by a node in the Bayesian network is independent of all variables represented by its non-descendent nodes in the Bayesian network, conditional on all variables represented by its parents nodes" (Bovens and Hartmann 2003 69). This assumption has often failed to hold in the history of life because of hybridization, the merger of two species, and symbiotic relations between two species that then become a single organism. Life, in other words, is causally "incestuous." Accounting for hybridization is a challenge for all methods of phylogenetic analysis. Though each method has particular strategies for accommodating this, typically it simply gets ignored. That is, phylogenetic models, like scientific models more generally, include simplifying idealizations, one of which is the assumption that branches split, but do not coalesce. Once again, philosophers will recognize these issues as familiar, and will have the opportunity to constructively contribute to how these debates unfold in systematics.

Summary and Synthesis

Systematists widely agree that reconstruction of phylogeny is a central task in systematics. As a result, justification of phylogenetic inference is a central conceptual issue for systematists. Broadly, three major schools have emerged in modern systematics, each tackling the problem of phylogenetic inference differently: Cladists, statistical phylogeneticists, and Bayesians. Cladists espouse using parsimony analysis on falsificationist grounds, looking to Popper for a solution to the problem of inference. Statistical phylogeneticists advocate using ML methods, relying on classical statistical techniques (which may or may not be grounded in falsificationism) and simulation studies.

Bayesians aim to incorporate inference more directly into their phylogenetic hypotheses, explicitly displaying prior assumptions and posterior probabilities.

One of the central problems to be tackled is how to adjudicate among the products of the various methods in phylogenetics. These methods reflect different philosophical stances that may be staked out by scientists concerned with historical events, reflecting a difference in how various research goals and problems are valued and prioritized. What makes a hypothesis about the past best or better? Hypotheses that more closely resemble history as it in fact unfolded, hypotheses that conform to methodological criteria, or hypotheses that present clearly testable explanations of particular historical phenomena?

One possible model for sifting through the issues raised by competing methodologies in phylogenetics is offered by Wilson and Sober's (1989) *competing conceptual frameworks*. Rather than conceive of competing phylogenetic techniques as hypotheses in direct competition, they may be better thought of as competing conceptual frameworks. Whereas competing hypotheses are exclusionary (i.e., if one is right, then all others must be false), competing conceptual frameworks should be understood as competing differently. Competing hypotheses share a vocabulary, i.e., the meaning of theoretical terms is consistent across competing hypotheses. Competing conceptual frameworks, on the other hand, may share *terms*, though these terms may have very different meanings in different frameworks, i.e., there may not be a shared vocabulary. So though at a surface level competing conceptual frameworks may seem exclusionary, they are often, in fact, simply making very different claims about the same system. As a result, the relation between competing conceptual frameworks might not be contrary or contradictory, but complementary.

The notion of competing conceptual frameworks begins to capture the state of the art in systematics, for to truly understand what a phylogenetic hypothesis amounts to, one must have a deep understanding of the nature, scope, and strength of justification that a particular method imparts on that hypothesis. It is for this reason that though two phylogenetic trees may share a topology, if produced from different methods they must be taken as conveying different hypotheses about phylogeny. Though the topology of the trees may be inter-translatable, the same may not be said of the inferences drawn from those trees (*pace* Sober 2004).

Note that in all three phylogenetic schools of thought the hypotheses of phylogeny are conditional claims. Phylogenetic systematists are all too aware of the epistemic limitations they face, hence the heated debates over the proper justification of phylogenetic inference. Philosophers have paid some attention to systematics, but not nearly to the extent that the material – or even the scientists – demands. Systematists have been openly engaging in philosophical discourse in the biological literature, but have very different research interests and agendas than philosophers. There is a rich story here, waiting for philosophical analysis.

Acknowledgment

References

Alfaro, M. E. and Holder, M. T. (2006). "The Posterior and the Prior in Bayesian Phylogenetics," *Ann. Rev. Ecol. Evol. Syst.*, 37, pp. 19–42.

Bovens, Luc and Hartmann, Stephan (2003). *Bayesian Epistemology* (Oxford: Oxford University Press).

Callender, C. and Cohen, J. (2006). "There Is No Special Problem about Scientific Representation," *Theoria*, 55, pp. 7–25.

Cracraft, J. (1974). "Phylogenetic Models and Classification," *Systematic Zoology*, 23, pp. 71–90.

de Queiroz, K. and Poe, S. (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, pp. 305–21.

de Queiroz, K. and Poe, S. (2003). "Failed Refutations: Further Comments on Parsimony and Likelihood Methods and Their Relationship to Popper's Degree of Corroboration," *Systematic Biology*, 52, pp. 322–30.

Darwin, C. (1975) [1859]. *On the Origin of Species* (Cambridge, MA: Harvard University Press).

Eldredge, N. and Cracraft, J. (1980). *Phylogenetic Patterns and the Evolutionary Process* (Columbia University Press: New York).

Farris, J. S. (1983). "The Logical Basis of Phylogenetic Analysis," in N. I. Platnick and V. A. Funk (eds.), *Advances in Cladistics*, vol. 2: *Proceedings of the Second Meeting of the Willi Hennig Society* (Columbia University Press: New York), pp. 7–36.

Felsenstein, J. (1978). "Cases in Which Parsimony or Compatibility Methods Will Be Positively Misleading," *Systematic Zoology*, 27, pp. 401–10.

Felsenstein, J. (1979). "Alternative Methods of Phylogenetic Inference and Their Interrelationships," *Systematic Zoology*, 27, pp. 401–10.

Felsenstein, J. (1981). "Maximum Likelihood Estimation of Evolutionary Trees from Continuous Characters," *American Journal of Human Genetics*, 25, pp. 471–92.

Felsenstein, J. (2004). *Inferring Phylogenies* (Sunderland, MA: Sinauer Associates).

Ghiselin, M. (1974). "A Radical Solution to the Species Problem," *Systematic Zoology*, 23(4), pp. 536–544.

Ghiselin, M. (2007). "Is the Pope a Catholic?" *Biology and Philosophy*, 22(2), pp. 283–91.

Giere, R. (1988). *Explaining Science: A Cognitive Approach* (Chicago, IL: University of Chicago Press: Chicago).

Gillies, D. (1990). "Bayesianism versus Falsificationism," *Ratio* (new series), 3, pp. 82–98.

Goodman, N. (1955). *Fact, Fiction and Forecast* (Cambridge, MA: Harvard University Press).

Goodman, N. (1972). "Seven Strictures on Similarity," in N. Goodman (ed.), *Problems and Projects* (New York: Bobbs-Merrill), pp. 437–47.

Griesemer, J. (2000). "The Units of Evolutionary Transition," *Selection*, 1(3), pp. 67–80.

Griffiths, G. (1974). "On the Foundations of Biological Systematics," *Acta Biotheoretica*, 23(3–4), pp. 85–131.

Haber, M. (2005). "On Probability and Systematics: Possibility, Probability, and Phylogenetic Inference," *Systematic Biology*, 54, pp. 831–41.

Hamilton, A. and Haber, M. (2006). "Clades Are Reproducers," *Biological Theory*, 1(4), pp. 381–91.

Hennig, W. (1979) [1966]. *Phylogenetic Systematics*, trans. D. D. Davis and R Zangerl (Urbana, IL: University of Illinois Press).

Hillis, D. M., Huelsenbeck, J. P., and Cunningham, C. W. (1994). "Application and Accuracy of Molecular Phylogenies," *Science*, 264, pp. 671–7.

Howson, C. and Urbach, P. (1993). *Scientific Reasoning: The Bayesian Approach*, 2nd edn. (Chicago, IL: Open Court).

Huelsenbeck, J. P. and Hillis, D. M. (1993). "Success of Phylogenetic Methods in the Four-Taxon Case," *Systematic Biology*, 42, pp. 247–64.

Huelsenbeck, J. P. and Rannala, B. (1997). "Maximum Likelihood Estimation of Phylogeny Using Stratigraphic Data," *Paleobiology*, 23, pp. 174–80.

Huelsenbeck, J. P. and Ronquist, F. (2001). "MRBAYES: Bayesian Inference of Phylogenetic Trees," *Bioinformatics*, 17, pp. 754–5.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R., and Bollback, J. P. (2001). "Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology," *Science*, 294, pp. 2310–14.

Huelsenbeck, J. P., Larget, B., Miller, R. E., and Ronquist, F. (2002). "Potential Applications and Pitfalls of Bayesian Inference of Phylogeny," *Systematic Biology*, 51, pp. 673–88.

Hull, D. L. (1970). "Contemporary Systematic Philosophies," *Annual Review of Ecology and Systematics*, 1, pp. 19–54.

Hull, D. L. (1976). "Are Species Really Individuals?" *Systematic Zoology*, 25(2), pp. 174–91.

Hull, D. L. (1983). "Karl Popper and Plato's Metaphor," in N. I. Platnick and V. A. Funk (eds.), *Advances in Cladistics*, vol. 2: *Proceedings of the Second Meeting of the Willi Hennig Society* (New York: Columbia University Press, New York), pp. 177–89.

Hull, D. L. (1988). *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science* (Chicago, IL: University of Chicago).

Hull, D. L. (1999). "The Use and Abuse of Sir Karl Popper," *Biology & Philosophy*, 14, pp. 481–504.

Kluge, A. G. (1997a). "Sophisticated Falsification and Research Cycles: Consequences for Differential Character Weighting in Phylogenetic Systematics," *Zoologica Scripta*, 26, pp. 349–60.

Kluge, A. G. (1997b). "Testability and the Refutation and Corroboration of Cladistic Hypotheses," *Cladistics*, 13, pp. 81–96.

Kluge, A. G. (2001). "Philosophical Conjectures and Their Refutation," *Systematic Biology* 50, pp. 322–30.

Larget, B. and Simon, D. (1999). "Markov Chain Monte Carlo Algorithms for the Bayesian Analysis of Phylogenetic Trees," *Molecular Biology and Evolution*, 16, pp. 750–9.

Platnick, N. (1977). "Cladograms, Phylogenetic Trees, and Hypothesis Testing," *Systematic Zoology*, 26, pp. 438–42.

Plutynski, A. (2005). "Parsimony and the Fisher–Wright Debate," *Biology and Philosophy*, 20, pp. 697–713.

Popper, K. (1959). *The Logic of Scientific Discovery* (New York: Basic Books).

Siddall, M. E. and Kluge, A. G. (1997). "Probabilism and Phylogenetic Inference," *Cladistics*, 13, pp. 313–36.

Skipper, R. A. (2002). "The Persistence of the R. A. Fisher-Sewall Wright Controversy," *Biology and Philosophy*, 17, pp. 341–67.

Sober, E. (1983). "Parsimony and Methods in Systematics," in N. I. Platnick, and V. A. Funk (eds.), *Advances in Cladistics*, vol. 2: *Proceedings of the Second Meeting of the Willi Hennig Society* (New York: Columbia University Press), pp. 37–47.

Sober, E. (1988). *Reconstructing the Past: Parsimony, Evolution, and Inference* (Cambridge, MA: MIT Press).

Sober, E. (2000). *Philosophy of Biology*, 2nd edn. (Boulder, CO: Westview Press).

Sober, E. (2004). "The Contest Between Parsimony and Likelihood," *Systematic Biology*, 53, pp. 644–53.

Sober, E. and Steel, M. (2002). "Testing the Hypothesis of Common Ancestry," *Journal of Theoretical Biology*, 218(4), pp. 395–408.

Sokal, R. R. and Sneath, P. H. A. (1963). *Numerical Taxonomy* (San Francisco, CA: W. H. Freeman).

Swofford, D. L., Olsen, G. J., Waddell, P. J., and Hillis, D. M. (1996). Chapter 11: "Phylogenetic Inference," in D. M. Hillis, C. Moritz, and B. Mable (eds.), *Molecular Systematics* (Sunderland, MA: Sinauer), pp. 407–514.

Teller, P. (2001). "Twilight of the Perfect Model Model," *Erkenntnis*, 55, pp. 393–415.

Tuffly, C. and Steel, M. (1997). "Links between Maximum Likelihood and Maximum Parsimony under a Simple Model of Site Substitution," *Bulletin of Mathematical Biology*, 59, pp. 581–607.

Urbach, P. (1991). "Bayesian Methodology: Some Criticisms Answered," *Ratio* (new series), 4, pp. 170–84.

Vermeij, G. (2006). "Historical Contingency and the Purported Uniqueness of Evolutionary Innovations," *Proceedings of the National Academy of Sciences*, 103(6), pp. 1804–9.

Wiley, E. O. (1975). "Karl R. Popper, Systematics and Classification: A Reply to Walter Bock and Other Evolutionary Taxonomists," *Systematic Zoology*, 24, pp. 233–43.

Wilson, D. S. and Sober, E. (1989). "Reviving the Superorganism," *Journal of Theoretical Biology*, 136(3), pp. 337–56.