

Statistical consistency and phylogenetic inference: a brief review

Andrew V. Z. Brower*

Evolution and Ecology Group, Department of Biology, Middle Tennessee State University, Murfreesboro, TN 37132, USA

Accepted 27 July 2017

Abstract

The claim that parsimony can be statistically inconsistent remains the chief criticism of the cladistic approach, and also the main justification for alternative model-based approaches such as maximum likelihood and Bayesian inference. Despite its refutation in the 1980s, this persistent myth of parsimony's Achilles' heel is entrenched in the primary literature, and has metastasized into textbooks, as well. Here, I review historical controversies, and offer three short arguments as to why statistical consistency is not only irrelevant to systematics, but to empirical science in general.

© The Willi Hennig Society 2017.

“One does not “justify” a method by showing that there is an extremely special case in which it does its work well; nor does one “refute” a method by showing that there is another special case in which it makes a hash of things.”

(Sober, 1988, p. 76)

History

In 1978, Felsenstein (p. 402, in a paper whose short title was “misleading parsimony methods”) asserted:

“Phylogenies constructed by the proper maximum likelihood method typically have the property of *consistency*. A statistical estimation method has the property of consistency when the estimate of a quantity is certain to converge to its true value as more and more data are accumulated...parsimony methods... do not possess the property of consistency in all cases.”

As Farris (1983, p. 15) said, “other things being equal, it would be desirable if an estimate would converge to the correct value as the amount of evidence increased,” and no one disputes that Felsenstein's artfully qualified (“proper,” “typically,” “in all cases”) assertions are untrue—of course parsimony could give a wrong answer, and of course an estimation using a model that is known to be true will converge on “the correct

value”. Rather, what is at issue is whether the property of consistency is empirically relevant to phylogenetic inference:

“Consistency is a logical relationship between an estimation method and a probability model. In the hypothetical case imagined by Felsenstein, his method would have obtained the right answer, but whether the method would work in practice depends on whether the model is accurate. If it is not, then the consistency of the estimator under the model implies nothing about the accuracy of the inferred tree.”

(Farris, 1983, p. 16)

Although Farris's (1983) criticisms of Felsenstein's consistency argument frequently have been misconstrued as an objection to Felsenstein's particular model (cf. Yang, 2014), Farris's point in this statement was that although “likelihood” as a general framework may produce consistent results in the abstract, whether or not a particular maximum-likelihood (ML) analysis of empirical data results in the true tree is entirely dependent on the accuracy of its particular model of character evolution, relative to the way evolution actually went (see also Farris, 1986). Parsimony can be inconsistent because any method can be inconsistent under the wrong circumstances. Inconsistency is thus not a potential shortcoming unique to parsimony.

In the early 1990s, however, there began to appear in the literature a number of less circumspect and

*Corresponding author.

E-mail address: abrower@mtsu.edu

more sweeping amplifications of Felsenstein's innuendos about parsimony's supposed flaw:

"...parsimony, when based on discrete characters, can fail, under certain conditions, to be a consistent estimator of the phylogeny." (DeBry, 1992)

"Parsimony has furthermore been shown to be statistically inconsistent (to arrive at the wrong answer at statistical significance) in a wide range of relatively realistic schemes." (Sidow, 1994)

"The principal objection to parsimony is that under some models of evolution it is not consistent, that is, even if we add more data it is possible to obtain the wrong tree." (Page and Holmes, 1998, p. 191)

Advocates of model-based approaches sowed further doubt by contriving simulations that showed various additional circumstances under which parsimony might behave inconsistently (cf. Zharkikh and Li, 1993; Kuhner and Felsenstein, 1994; Takezaki and Nei, 1994; Kim, 1996; Huelsenbeck and Lander, 2003).

In 2003, Goloboff suggested that the tide of parsimony-bashing on the grounds of consistency was ebbing, due to a variety of demonstrations that ML could also be inconsistent. If invocation of consistency was indeed diminishing, it is likely because it was being supplanted by references to a putative causal mechanism—"long branch attraction" that resulted in inconsistent hypotheses (cf. Hendy and Penny, 1989; Bergsten, 2005). Yet Felsenstein (2004, p. 121) cited consistency in dismissing parsimony out of hand:

"The inconsistency of parsimony has been the strongest challenge to its use. It becomes difficult to argue that parsimony methods have logical and philosophical priority, if one accepts that consistency is a highly desirable property."

Many of the above citations are 20+ years old, and it may seem like this controversy ought to be ancient history. But the invocation of inconsistency as a problem for phylogenetic analyses persists in the literature (Lartillot, 2015; McTavish et al., 2015; Kjer et al., 2016), and as the social media outcry over the January 2016 editorial in *Cladistics* (The Editors, 2016) showed, there remains a great deal of animus and misunderstanding about the interplay between science and philosophy regarding phylogenetic inference (cf. Simon, 2016). Furthermore, the droning of this anti-cladistic mantra has been so steady and prolonged that it has become established as dogma in recent textbooks:

"...long branches attract. Under these conditions, maximum parsimony is likely to yield the wrong phylogenetic tree." By contrast, "Maximum likelihood analysis takes into account the differences in branch length...and is more likely to find the correct phylogeny." (Futuyma, 2013, p. 35)

"...parsimony can be inconsistent over some portions of the parameter space; in particular, the method tends to suffer from the problem of long branch attraction." (Yang, 2014, p. 170)

"...parsimony approaches are not without problems. The worst of these problems is that parsimony is not a *consistent estimator*; that is, an estimation procedure that, given enough data, will ensure that we get the right answer. Thus, if we use parsimony to reconstruct a phylogeny, it is possible for us to get the wrong tree, no matter how much data we have available." (Bergstrom and Dugatkin, 2016, p. 153)

"...maximum parsimony methods may perform poorly whenever some branches on the tree are much longer than others, because parsimony will tend to cluster long branches together." (Graur, 2016, p. 211)

"While parsimony often gives the correct result, it does a poor job of estimating the phylogeny in some situations, in particular when there is a lot of homoplasy on the phylogeny and when evolutionary rates vary among branches of the tree. Those problems motivate methods that use statistical methods as likelihood, which is explained in the appendix." (Futuyma and Kirkpatrick, 2017, p. 411)

Despite this disparagement of parsimony, when these books endeavour to explain phylogenetic inference, they invariably do so with cladistic terminology, such as synapomorphy vs. symplesiomorphy, and employ parsimony as their heuristic optimality criterion. Superficial explanations of ML and Bayesian approaches are relegated to sidebars or appendices (sometimes even nonexistent ones; cf. Bergstrom and Dugatkin, 2016; p. 150). Implying that parsimony is now viewed as obsolete, Baum and Smith (2013, p. 207) said, "...the scientific community generally expects researchers to use maximum-likelihood or Bayesian methods when analysing molecular data. Nonetheless, because parsimony is effective for many data sets, is less computationally demanding, and is easier to understand, it is still widely used in educational contexts and for preliminary data exploration."

It is time, once and for all, to lay this biased nonsense to rest. If there is a reason to prefer model-based phylogenetic methods over parsimony, it has nothing to do with statistical consistency, for reasons explained below.

How big is the Felsenstein zone?

"...it is difficult to decide where the cutoff between reality and fantasy exists." (Huelsenbeck, 1995, p. 22)

Most of the evidence fuelling debates over whether or not phylogenetic methods are consistent has involved studies in which the data are simulated according to a known model, and then the various methods are tested to see if they can recover the "true tree". Huelsenbeck and Hillis (1993) named the part of conceptual tree space in which methods are inconsistent the Felsenstein zone. Advocates of models like to highlight the success of "more realistic" models versus parsimony over a broader portion of this space, whereas advocates of parsimony observe that the absolute size of

the Felsenstein zone appears to be small (Albert et al., 1993). (For in-depth discussion of the current state of affairs in model-based versus cladistic simulations, see Goloboff et al. (2017).)

Regrettably, none of that has much to do with assessing the reliability of phylogenetic inference of relationships among actual taxa. As Felsenstein (1983, p. 188) said, “The difficulty we face is that we know too little to specify a realistic model of evolutionary change. Even if we could do so, it would not be mathematically tractable. In this sense the advocacy of total realism is a counsel of despair.” What was true in 1983 remains true today. Thus, although Kim’s (1996, p. 46) observation that “the relevant question is not whether an estimator is consistent or inconsistent, but how common the conditions are under which a phylogenetic estimator is inconsistent” may be answerable within the make-believe, Platonic realm of model comparisons, in the world of actual phylogenetic data, whose particular evolutionary paths are shrouded by historical contingency, the answer is always: “there is no way to tell” (cf. Farris, 1999). Conversely, when we infer an empirical tree for some group, we can know neither that our optimality criterion behaved consistently, nor that the tree selected by our optimality criterion is correct. Brower (2016) has explored the logical contradictions of the realist approach to phylogenetic inference.

In the absence of verifiability, another, perhaps more empirically pertinent way to compare the success of alternative methods is to assess how often they give different answers for a given set of real phylogenetic data. In particular, if “realistic” models are thought to perform better than parsimony (cf. Huelsenbeck et al., 2011), then we might expect phylogenetic results obtained from both methods applied to the same data to differ in cases where parsimony is suspected to behave inconsistently. Rindal and Brower (2011) assessed this question in a meta-analysis of 1000 empirical articles in *Molecular Phylogenetics and Evolution*. Of 411 studies employing both parsimony and a model-based approach, only three had topological differences which their authors deemed significant. Smith (2013) suggested that more differences might be found in studies of more inclusive taxa, but Brower and Rindal (2013) showed that even those did not differ significantly due to the alternative analytical approaches they employed.

Of course getting the same topology twice does not corroborate the result *per se*, nor does it suggest that the tree is more likely to be “true” (*contra* Kim, 1993). But if results are the same from the data analysed by parsimony and by a model intended to correct for characters that mislead parsimony, then either those unruly characters do not exist in that dataset, or they have also misled the model-based analysis. Whichever

is true, this empirical meta-analysis showed that the effective size differential of the Felsenstein zone between parsimony and model-based methods appears to be inconsequential. Thus, claims that models outperform parsimony are unsubstantiated by patterns of relationships among actual taxa, arguably the most relevant criterion for comparing the success of alternative methods.

Empirical indeterminacy and the necessary insufficiency of models

“As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.”
(Einstein, 1922, p. 1)

Statistical simulations are conducted in a contrived closed system in which parameters are defined *a priori* and results may be tested against known values for accuracy. Only under such circumstances is it possible to ask how often a given method recovers the “true tree.” However, when performing empirical studies, because the true history of the taxa is unknown, there is no way to assess accuracy relative to the actual pattern of historical diversification. Farris recognized this long ago, (1986, p. 14) noting “(c)riticisms by Felsenstein that parsimony analysis may be statistically inconsistent lead to an equivalent criticism of all statistical methods, and so are of no value for evaluating phylogenetic methods”. Similar caveats were echoed by: Sober (1988, p. 171), “If the assumed model is in fact false, there is no guarantee that the method will converge on the truth when applied to the real world”; Hillis (1995, p. 4), “All methods are consistent when their assumptions (explicit and implicit) are met, and all methods are inconsistent when these assumptions are violated sufficiently”; Swofford et al. (1996, p. 427, footnote), “a maximum likelihood method will guarantee consistency only if evolution proceeds according to the assumed model”; and even Felsenstein himself (2004, p. 272), “likelihood is usually consistent if we use the correct model in our analysis. When we use the wrong model, there are few guarantees”. The fundamental problem here is that there is no empirical means to determine whether the model accurately reflects the process by which the characters really evolved. So, to paraphrase Popper (1983, p. 79), even if the model is true, there is no way to know that it is true.

A proxy criterion for evaluating models in an absolute sense is to compare the fit of the data to a variety of alternatives, and select the one with the highest likelihood. Of course, the more parameters one adds to a model, the better a fit to the data it is likely to achieve, and to balance this infinite regress of complexity, parameter-rich models are penalized formulaically by

various “information criteria”. But comparing relative fits of data to models begs the question of how well the model fits the data from an absolute perspective. Gatesy (2007, p. 509) offered the following pungent simile:

“Given the simplicity of most models, it is possible that model selection in modern systematics is analogous to an overweight man shopping in the petites department of a women’s clothing store. A particular garment might fit the portly man best, but this does not imply a good overall fit.”

This metaphor, just as comparative model fitting itself, is an instance of “the principle of the drunkard’s search” (Kaplan, 1964), in which the drunkard hunts for a lost wallet under a streetlight because that is where the light is, rather than where the wallet is. The rationale that “we seek models that are good approximations of reality given that reality itself is unknowable” (Kelchner and Thomas, 2007, p. 87)—constraining one’s search for truth to the realm of tractable model space—is hardly a “realist” approach. There is no “accuracy” if you are aiming at a nonexistent target (see further discussion of these issues in Brower, 2016). A troubling implication of this disconnect is that models selected under such circumstances are quite unlikely to be statistically consistent. And even if they were, how would you know?

Induction and parsimony

“If no other coins are available for testing, consistent inference is not possible.”

(Goldman, 1990, p. 348)

Many authors have noted that the history of life is a singular event (Wenzel and Carpenter, 1994; Siddall and Kluge, 1997; Kluge, 2005; Brower, 2010). If that is true, it seems that Goldman’s coin-flipping critique, above, also applies to phylogenetic inference. In order to be amenable to statistical testing, various unrealistic, or at least not demonstrably realistic, assumptions about the evidence must be made. The notion of statistical consistency as applied to phylogenetic data makes the standard statistical assumptions that characters are independent and drawn from an identical distribution (IID). Siddall and Kluge (1997), Siddall (1998) and Farris (1999) attacked Felsenstein’s (1973) reliance on Wald’s (1949) demonstration that likelihood is consistent, and in response Felsenstein (2004, p. 271) offered a “proof” that as far as I can tell assumes *a priori* what it is intended to prove: “Let us assume that characters evolve independently according to the same Markov process.” Searching under the streetlight again.

Pickett et al. (2005) claimed not only that phylogenetic data often are not homogeneous (IID), but also

that when they are not, parsimony can get back on track to consistency by the addition of new data that exhibit heterotachy (within site rate heterogeneity among branches). Although that observation may be valid, it seems to cede the ground to its opponents that statistical consistency is a criterion we ought to be concerned about (not to mention that in order to determine what was consistent and what was not, Pickett et al. needed an *a priori* idea of the truth—in that case, their test criterion was monophyly of the vespid wasp genus *Apoica*). Generally speaking, in order to recognize “long-branch attraction” as a phenomenon of potential concern in the first place, one must have an *a priori* idea that the relationship implied by the joining of those branches is not correct. Finally, the broader validity of Pickett et al.’s rescue by heterotachy observation is unknowable: as Sober (1993) argued, results of experimental phylogenetic studies cannot be extrapolated into general principles.

Whether or not phylogenetic data are IID, statistical models assume that they are. This is a uniformitarian, inductive parsimony assumption that “the future resembles the past” or, more aptly for phylogenetic inference, the past resembles the present. When numbers are used to quantify empirical entities and phenomena in the physical world, they become prone to error. This is not simply statistical error, like the standard deviation around a mean, but systematic error, if the assumptions upon which the means of quantification and the attributes of the things being quantified do not align. If, for example, the probability of a coin toss landing heads is not fixed and might vary arbitrarily through time, then no number of tosses will yield an accurate estimate of the probability of getting heads into the future. The “law of large numbers,” the statistical premise that as the sample size increases, the estimate approaches with greater precision and accuracy the population parameter, plainly rests upon the assumption that the future will rigidly resemble the past. Sober (2015, p. 285) made a similar point:

“...maximum likelihood is statistically consistent...The pertinent Humean point about this claim concerning statistical consistency is that the claim’s justification depends upon assumptions that go beyond the observations you have made to date. You are assuming that the coin is an *i.i.d.* system - that past, present and *future* tosses are independent of each other and there is a single probability of heads that applies to past, present and future tosses.”

Or, as Hume (1748) would have said, the claim of statistical consistency is inductive and cannot be substantiated in any empirical instance, let alone justified as a general principle. Thus, it plainly holds no force whatsoever as a criticism of cladistic parsimony or any other phylogenetic method.

Conclusion

This essay aims to debunk a persistent myth about a critical failing of parsimony in phylogenetic inference as a rationale for model-based approaches. The object is not to show that parsimony is superior, but merely to remove an arrow from the quiver of its critics. Goloboff et al. (2017) have noted that the cutting edge of the controversy over competing methods has moved on, that many recent models may themselves be inconsistent estimators, about which their authors no longer seem to care. That literature is dense, dispersed and technically complex, and fundamental assumptions of a given approach may be buried in the “flute music”. Of course, quantitative complexity provides no more a guarantee of realism than does statistical consistency. There is no escape from philosophy when we are asking questions like “what is real?”, even in regard to seemingly simple problems like preferring one phylogenetic hypothesis over another. As Goloboff et al. (2017) observed, “considering that methods can be justified only with statistical principles is itself a ‘philosophical’ position”. The very modest, and hardly original point here, which I hope is expressed in a manner intelligible to systematists, teachers and students, is that the conventional wisdom that potential for statistical inconsistency offers a reason to reject parsimony is false.

Acknowledgments

I thank Ivonne Garzón-Orduña and Maureen O’Leary for discussions about these ideas, and Pablo Goloboff and an anonymous reviewer for valuable suggestions on the manuscript. Support for this effort was provided by a collaborative grant, *Dimensions US-Biota-São Paulo: Assembly and evolution of the Amazon biota and its environment: an integrated approach*, supported by the US National Science Foundation (NSF DEB 1241056), National Aeronautics and Space Administration (NASA), and the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP Grant 2012/50260-6).

References

- Albert, V.A., Chase, M.W., Mishler, B.D., 1993. Character-state weighting for cladistic analysis of protein-coding DNA sequences. *Ann. Mo. Bot. Gard.* 80, 752–766.
- Baum, D.A., Smith, S.D., 2013. *Tree Thinking: An Introduction to Phylogenetic Biology*. Roberts & Co., Greenwood Village, CO.
- Bergsten, J., 2005. A review of long-branch attraction. *Cladistics* 21, 163–193.
- Bergstrom, C.T., Dugatkin, L.A., 2016. *Evolution*, 2nd edn. W. W. Norton, New York, NY.
- Brower, A.V.Z., 2010. Stability, replication, pseudoreplication and support. *Cladistics* 26, 112–113.
- Brower, A.V.Z., 2016. Are we all cladists? In: Williams, D.M., Schmitt, M., Wheeler, Q.D. (Eds.), *The Future of Phylogenetic Systematics: The Legacy of Willi Hennig*. Cambridge University Press, Cambridge, pp. 88–114.
- Brower, A.V.Z., Rindal, E., 2013. Reality check: a reply to Smith. *Cladistics* 29, 464–465.
- DeBry, R.W., 1992. The consistency of several phylogeny-inference methods under varying evolutionary rates. *Mol. Biol. Evol.* 9, 537–551.
- Einstein, A., 1922. Geometry and experience. In: Einstein, A (Ed.), *Sidelights on Relativity*. Methuen, London, pp. 15–28.
- Farris, J.S., 1983. The logical basis of phylogenetic analysis. In: Platnick, N.I., Funk, V.A. (Eds.), *Advances in Cladistics*. Columbia University Press, New York, NY, pp. 7–36.
- Farris, J.S., 1986. On the boundaries of phylogenetic systematics. *Cladistics* 2, 14–27.
- Farris, J.S., 1999. Likelihood and inconsistency. *Cladistics* 15, 199–204.
- Felsenstein, J., 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst. Zool.* 22, 240–249.
- Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27, 401–410.
- Felsenstein, J., 1983. Parsimony in systematics: biological and statistical issues. *Annu. Rev. Ecol. Syst.* 14, 313–333.
- Felsenstein, J., 2004. *Inferring Phylogenies*. Sinauer Associates, Sunderland, MA.
- Futuyma, D.J., 2013. *Evolution*, 3rd edn. Sinauer Associates, Sunderland, MA.
- Futuyma, D.J., Kirkpatrick, M., 2017. *Evolution*, 4th edn. Sinauer Associates, Sunderland, MA.
- Gatesy, J., 2007. A tenth crucial question regarding model use in phylogenetics. *Trends Ecol. Evol.* 22, 509–510.
- Goldman, N., 1990. Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson process model of DNA substitution and to parsimony analyses. *Syst. Zool.* 39, 345–361.
- Goloboff, P.A., 2003. Parsimony, likelihood, and simplicity. *Cladistics* 19, 91–103.
- Goloboff, P.A., Torres, A., Arias, J.S., 2017. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics*. (early view, <https://doi.org/10.1111/cla.12205>).
- Graur, D., 2016. *Molecular and Genome Evolution*. Sinauer Associates, Sunderland, MA.
- Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38, 297–309.
- Hillis, D.M., 1995. Approaches for assessing phylogenetic accuracy. *Syst. Biol.* 44, 3–16.
- Huelsenbeck, J.P., 1995. Performance of phylogenetic methods in simulation. *Syst. Biol.* 44, 17–48.
- Huelsenbeck, J.P., Hillis, D.M., 1993. Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* 42, 247–264.
- Huelsenbeck, J.P., Lander, K.M., 2003. Frequent inconsistency of parsimony under a simple model of cladogenesis. *Syst. Biol.* 52, 641–648.
- Huelsenbeck, J.P., Alfaro, M.E., Suchard, M.A., 2011. Biologically inspired phylogenetic models strongly outperform the no common mechanism model. *Syst. Biol.* 60, 225–233.
- Hume, D., 1748. *An Inquiry Concerning Human Understanding*. Bobbs-Merrill, Indianapolis, IN (1955 edn.).
- Kaplan, A., 1964. *The Conduct of Inquiry: Methodology for Behavioral Science*. Chandler Publishing Co., San Francisco, CA.
- Kelchner, S.A., Thomas, M.A., 2007. Model use in phylogenetics: nine key questions. *Trends Ecol. Evol.* 22, 87–94.
- Kim, J., 1993. Improving the accuracy of phylogenetic estimation by combining different methods. *Syst. Biol.* 42, 331–340.
- Kim, J., 1996. General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. *Syst. Biol.* 45, 363–374.
- Kjer, K.M., Simon, C., Yavorskaya, M., Beutel, R.G., 2016. Progress, pitfalls and parallel universes: a history of insect phylogenetics. *J. R. Soc. Interface* 13, 20160363.

- Kluge, A.G., 2005. What is the rationale for ‘Ockham’s razor’ (a.k.a. parsimony) in phylogenetic inference? In: Albert, V. A. (Ed.) *Parsimony, Phylogeny, and Genomics*. Oxford University Press, Oxford, pp. 15–42.
- Kuhner, M.K., Felsenstein, J., 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* 11, 459–468.
- Lartillot, N., 2015. Probabilistic models of eukaryotic evolution: time for integration. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20140338.
- McTavish, E.J., Steel, M.A., Holder, M.T., 2015. Twisted trees and inconsistency of tree estimation when gaps are treated as missing data - the impact of model mis-specification in distance corrections. *Mol. Phylogenet. Evol.* 93, 289–295.
- Page, R.D.M., Holmes, E.C., 1998. *Molecular Evolution: A Phylogenetic Approach*. Blackwell Science, Oxford.
- Pickett, K.M., Tolman, G.L., Wheeler, W.C., Wenzel, J.W., 2005. Parsimony overcomes statistical inconsistency with the addition of more data from the same gene. *Cladistics* 21, 438–445.
- Popper, K.R., 1983. *Realism and the Aim of Science*. Routledge, New York, NY.
- Rindal, E., Brower, A.V.Z., 2011. Do model-based phylogenetic analyses perform better than parsimony? A test with empirical data *Cladistics* 27, 331–334.
- Siddall, M.E., 1998. Success of parsimony in the four-taxon case: long branch repulsion by likelihood in the Farris Zone. *Cladistics* 14, 209–220.
- Siddall, M.E., Kluge, A.G., 1997. Probabilism and phylogenetic inference. *Cladistics* 13, 313–336.
- Sidow, A., 1994. Parsimony or statistics? *Nature* 367, 26.
- Simon, M., 2016. Twitter nerd-fight reveals a long, bizarre scientific feud. *Wired* <https://www.wired.com/2016/02/twitter-nerd-fight-reveals-a-long-bizarre-scientific-feud/>.
- Smith, M.R., 2013. Likelihood and parsimony diverge at high taxonomic levels. *Cladistics* 29, 463.
- Sober, E., 1988. *Reconstructing the Past: Parsimony, Evolution and Inference*. MIT Press, Cambridge, MA.
- Sober, E., 1993. Experimental tests of phylogenetic inference methods. *Syst. Biol.* 42, 85–89.
- Sober, E., 2015. *Ockham’s Razors: A User’s Manual*. Cambridge University Press, Cambridge, UK.
- Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Mable, B.K., Moritz, C. (Eds.), *Molecular Systematics*. Sinauer Associates, Sunderland, MA, pp. 407–514.
- Takezaki, N., Nei, M., 1994. Inconsistency of the maximum parsimony method when the rate of nucleotide substitution is constant. *J. Mol. Evol.* 39, 210–218.
- The Editors, 2016. Editorial. *Cladistics* 32, 1.
- Wald, A., 1949. Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.* 20, 595–601.
- Wenzel, J.W., Carpenter, J.M., 1994. Comparing methods: adaptive traits and tests of adaptation. In: Eggleton, P., Vane-Wright, R.I. (Eds.), *Phylogenetics and Ecology*. Academic Press, London, pp. 79–101.
- Yang, Z., 2014. *Molecular Evolution: A Statistical Approach*. Oxford University Press, Oxford, UK.
- Zharkikh, A., Li, W.H., 1993. Inconsistency of the maximum-parsimony method: the case of five taxa with a molecular clock. *Syst. Biol.* 42, 113–125.