Historical Essay

Syst. Biol. 0(0):1–8, 2020

DOI:10.1093/sysbio/syaa068

An Evolving View of Phylogenetic Support

CHRIS SIMON*

Department of Ecology and Evolutionary Biology, 75 N. Eagleville Road, University of Connecticut, Storrs, CT 06268-3043, USA
**Correspondence to be sent to: Department of Ecology and Evolutionary Biology, 75 N. Eagleville Road, University of Connecticut, Storrs, CT 06268-3043, USA; E-mail: chris.simon@uconn.edu.*

Received 14 February 2020; reviews returned 4 August 2020; accepted 15 August 2020
Associate Editor: Robert Lanfear

Abstract.—If all nucleotide sites evolved at the same rate within molecules and throughout the history of lineages, if all nucleotides were in equal proportion, if any nucleotide or amino acid evolved to any other with equal probability, if all taxa could be sampled, if diversification happened at well-spaced intervals, and if all gene segments had the same history, then tree building would be easy. But of course, none of those conditions are true. Hence, the need for evaluating the information content and accuracy of phylogenetic trees. The symposium for which this historical essay and presentation were developed focused on the importance of phylogenetic support, specifically branch support for individual clades. Here, I present a timeline and review significant events in the history of systematics that set the stage for the development of the sophisticated measures of branch support and examinations of the information content of data highlighted in this symposium. [Bayes factors; bootstrap; branch support; concordance factors; internode certainty; posterior probabilities; spectral analysis; transfer bootstrap expectation.]

Trees are meaningless without branch supports, much like means are useless without variances. Branch supports—sometimes called nodal supports—measure how much of the information in the data set used to create a particular tree supports each internal branch in that tree. Branch support increases with increasing numbers of characters and branches with 100% support are not guaranteed to be correct (Pamilo and Nei 1988). But even when using large genomic data sets with tens of thousands of informative characters, some clades can possess ambiguous (or low) support values leaving relationships uncertain (e.g., Dietrich et al. 2017; Skinner et al. 2019).

Branch support varies across data sets and across clades in a single tree. This can be caused by random error related to the number of informative sites in the data set (e.g., Swofford et al. 2001); taxon sampling (e.g., Hedtke et al. 2006); systematic error/model misspecification (e.g., Sullivan and Joyce 2005); and/or short internal branches caused by slow substitution rates or short periods of time between speciation events (Maddison and Knowles 2006). Short internal branches increase the probability of incomplete lineage sorting (ILS). ILS occurs when ancestral polymorphisms are inherited during speciation events and then one of the alleles at the polymorphic locus is randomly fixed in the population. ILS can result in multiple histories within and among genes (Pamilo and Nei 1988; Maddison 1997). Short times between splitting events also increase the probability of speciation with gene flow or hybridization (e.g., Sullivan et al. 2014). Both ILS and hybridization reduce branch support strength due to conflicting signal. Finally, short branches increase the probability of the juxtaposition of long and short branches in trees that makes phylogenetic resolution extremely difficult (Hendy and Penny 1989; Swofford et al. 2001). The bottom line is that because evolutionary processes are so complicated, meaningful branch supports are difficult to devise and interpret.

Bootstrap Resampling: Definitions, Critiques, and Corrections

Branch supports have been calculated in many ways throughout the history of molecular systematics. Figure 1 juxtaposes a timeline for major events in phylogenetic systematics with a timeline for the development of different methods for understanding branch support. The earliest branch supports were resampling techniques. Statistically, resampling techniques originated with Fisher (1935) who, as described by Rodgers (1999), recognized the value of empirically generated sampling distributions for estimating the approximate variance of a statistic when the sampling distribution is unknown. Jackknife resampling techniques were developed by Quenouille (1949) and Tukey (1958). Efron (1979) modified jackknife resampling to require replacement of observations during each subsample such that some were sampled twice, some never. In this procedure, called bootstrap resampling, data are sampled randomly until the fabricated data set or pseudosample is the same size as the original. The jackknife was introduced into phylogenetics (for genetic distances) by Mueller and Ayala (1982) and the bootstrap by Felsenstein (1983, 1985) and Penny and Hendy (1985). For phylogenetic bootstraps, a tree is created for each replicate and the bootstrap proportion is the percent of pseudosamples in which the original branch is supported (Felsenstein 1985; Swofford et al. 1996). Bootstrapping requires that the original data sample be sufficiently large to capture

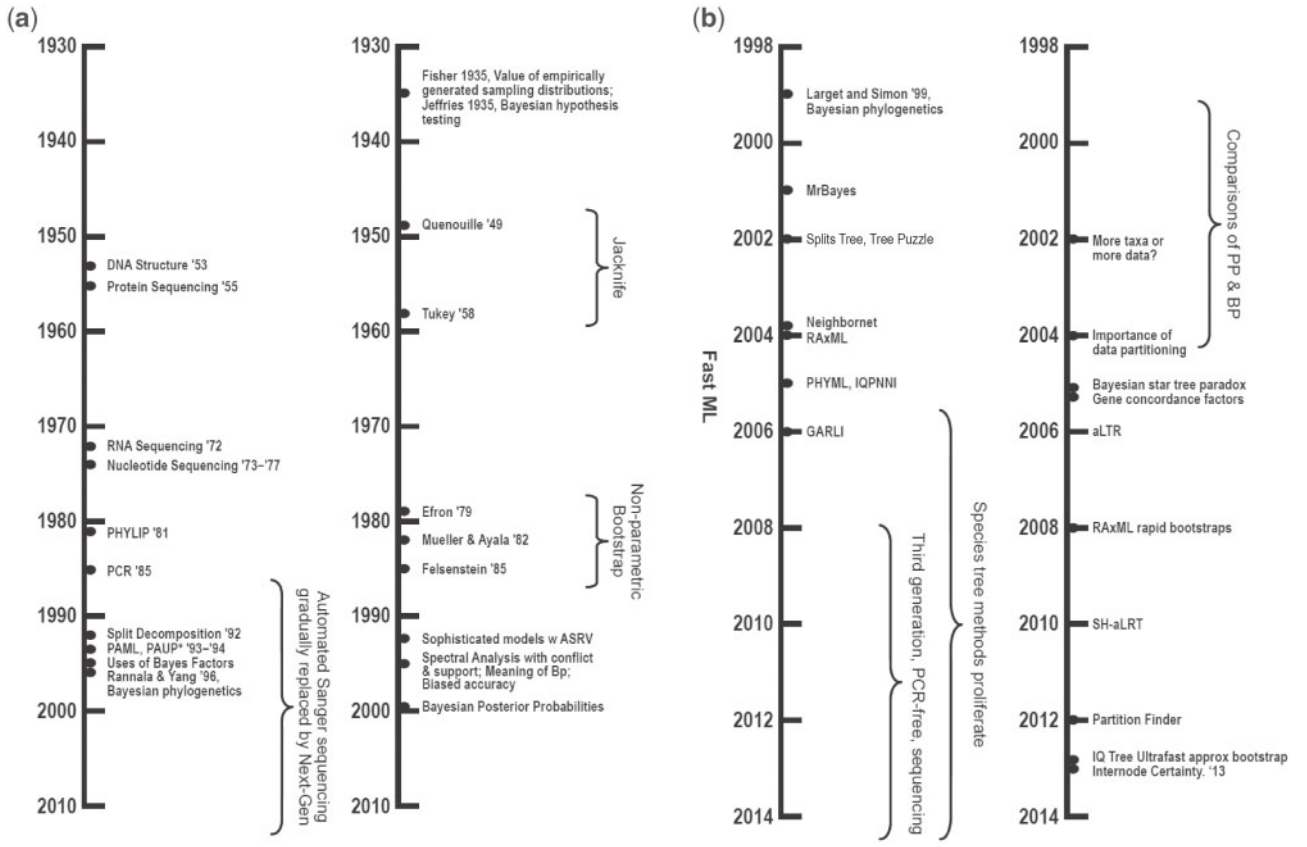

FIGURE 1. Time line of significant events in the history of phylogenetic support (right) compared to relevant events in the history of molecular systematics (left). a) 1930–2010; b) 1998–2014. PHYLIP = "Phylogeny Inference Package"; PCR = Polymerase Chain Reaction; PAML = "Phylogenetic Analysis using Maximum Likelihood" program; PAUP* = "Phylogenetic Analysis Using Parsimony" and * = "other methods" program; ASRV = Among-Site Rate Variation; Bp = Bootstrap Proportions; RAxML = "Randomized Axelerated Maximum Likelihood" program; MrBayes = program for Bayesian phylogenetic analysis; Splits Tree = program for tree building via split decomposition; IQPNNI = "Important Quartets" maximum likelihood tree building with quartet puzzling and "Nearest Neighbor Interchange" program; GARLI = "Genetic Algorithm for Rapid Likelihood Inference" program; aLRT = Approximate Likelihood Ratio Test; SH-aLRT = Shimodaira-Hasegawa-like approximate Likelihood Ratio Test. See text for details and citations.

relevant features of the unknown distribution from which the data are sampled.

During the 1990s a series of papers discussed the meaning of bootstrap support (also known as bootstrap proportions), its biased nature, and corrections for that bias (Zharkikh and Li 1992a,b; Sanderson 1995; Sanderson et al. 2000). Arguments broke out (reviewed in Sanderson 1995) that focused on whether bootstraps represented confidence intervals (they do not) and whether bootstrap proportions can be used to provide evidential support (they can). Bootstraps were shown to correlate well with other measures of branch support such as spectral analysis (discussed below). Hillis and Bull (1993) used both simulated data and well-supported phylogenies to examine the accuracy of bootstraps (defined as the probability that a result represents the true phylogeny). They found that, under their simulation conditons, low bootstrap support overestimated accuracy while high bootstrap support underestimated accuracy. They pointed out that bootstraps are influenced by the number of characters in the data set, the number of taxa, and the location of internal branches. Their "cut-off" value of 70% bootstrap support has been taken as a rule of thumb but of course is specific to the examples they examined.

A number of phylogenetic statisticians attempted definitions of bootstrap support. Hillis and Bull (1993, p. 183), called it a "measure of accuracy" or "the probability that a specified group is contained in the true tree." Felsenstein and Kishino (1993) in their rebuttal to Hillis and Bull (1993)—repeating an idea put forward by Felsenstein (1985)—stated that accuracy was not the intended interpretation of bootstrap proportion (BP) but rather that 1 - BP is a *P* value for the test of the null hypothesis that a split is not present. Berry and Gascuel (1996) argued that the 50% bootstrap consensus trees represent the phylogenetic estimate with the lowest combined probability of Type I (including incorrect groups) and Type II (excluding correct groups) error. Swofford et al. (1996, p. 509) described bootstraps as a measure of support that reflects "the frequency in which a group appears in replicate trees" rather than a statistical statement that would require that the node of interest be specified in advance. Efron et al. (1996) argued

that the *P* value interpretation was approximately correct (with increasing sequence length) and later, Susko (2009) demonstrated that BP was not approximately a *P* value but that it was conservative under the hypothesis that if a split is not present, BP $\geq$ 95% should occur less than 5% of the time. All of this presumes, a fixed split specified in advance. Page and Holmes (1998, p. 222) described bootstrap support as a measure of "precision not accuracy" because branches with 100% support could be wrong. Holmes (2003, p. 244) described bootstraps as, "a measure of the robustness of the results of a phylogenetic analysis to small changes in the data." Bootstrap supports were widely applied to distance tree-building analyses as well as maximum likelihood and parsimony.

Long-branched rogue taxa and properties of DNA data can influence bootstrap support. Page (1996) and Wilkinson (1996) demonstrated that if one taxon is unstable due to homoplasy and jumps around the tree, then bootstraps will be lowered in that region of the tree. Page and Holmes (1998) noted that if only one site supports a node, bootstraps could be spuriously low. Sanderson pointed out that the bootstrap measurement assumes that sites are independent (not correlated) and identically distributed (no among site rate or pattern variation), which they never are (Simon et al. 1996, 2006). Of course, analytical methods can violate assumptions to some degree before they become completely invalid (Penny et al. 1992).

Meanwhile, Back in the Lab

At the same time that Mueller and Ayala (1982) and Felsenstein (1985) were introducing bootstraps to phylogenetic analysis, polymerase chain reaction (PCR) was invented (Saiki et al. 1985; Mullis and Faloona 1987) making it easy to sequence DNA. In the following year, the first PAUP manual (Swofford 1985) was released, and the 2 years after that MacClade version 1.0 and 2.0 (Maddison 1997; Maddison and Maddison 1987). Automated, second-generation sequencing based on PCR was developed in the late 1980s and was increasingly deployed in individual laboratories throughout the 1990s. More data meant higher branch supports! More taxa raised or lowered branch supports depending on their placement (Sanderson 1995). By 1993, early unpublished versions of Ziheng Yang's PAML and Dave Swofford's PAUP* were making sophisticated models available that included among site rate variation (ASRV). Frati et al. (1997) demonstrated that ASRV was the most important model parameter to accommodate in maximum likelihood analyses and Buckley et al. (2001) demonstrated that the manner in which ASRV was accommodated made a big difference. However, in the 1990s and early 2000s, maximum likelihood programs were frustratingly slow when more than ten taxa were analyzed with realistic models, so it was time consuming to calculate branch supports.

Spectral Analysis: Conflict, Support, and Predictors of Bootstraps

Spectral analysis was adapted for phylogenetics in the early 1990s (Hendy and Penny 1993; Penny et al. 1993). It was performed using the program HadTree. This program uses the Hadamard matrix/Hadamard conjugation (a discrete Fourier transform and mathematical trick for speeding matrix manipulations) and Split Decomposition (Bandelt and Dress 1992) or the Closest Tree method (Penny et al. 1993) to build a phylogeny. Like maximum likelihood, HadTree did not lose information in the original sequence data during transformation or optimization. Spectral analysis could be applied to parsimony, corrected parsimony, likelihood, or distance analyses and implemented with four models: Cavender 78, JC 69 K2P 80, and K3P 81. An application of this method (Lento et al. 1995) is illustrated in Figure 2. Lento et al. (1995) created a tree to examine pinniped evolution. The Spectra (the frequency of each split) were plotted as a bar graph (Fig. 2) where support equals the frequency of that split in the data and conflict equals the sum of contradictory splits. Because all splits cannot be represented simultaneously on a single tree, their tree contains only the mutually compatible splits that have the most support. A very nice feature of the method is the ability to visualize both support and conflict for multiple alternative hypotheses of relationships on one graph including those splits not represented on the tree. They also calculated "predictors of bootstrap" (PB values) using their spectral signals of support and conflict and showed that PB values were correlated with bootstrap support values. Unfortunately, this method was not widely applied due to the lack of a user-friendly program. The introduction of the user-friendly program SplitsTree (Huson 1998) later facilitated the use of split decomposition and the plotting of splits graphs with useful box-like visualizations of branch conflicts based on distance data but did not include HadTree or Lento Plots.

Bayesian Phylogenetics and Posterior Probabilities

In the mid-1990s, in the field of statistics, Kass and Raftery (1995) discussed the uses of the Bayes Factor for hypothesis testing in genetics, ecology, and other fields setting the stage for the introduction of Bayesian phylogenetics. Rannala and Yang (1996) introduced to phylogenetics the idea of using Bayesian analysis and posterior probabilities (PP) to estimate phylogenetic trees and branch supports, respectively. Thorne et al. (1998) introduced the first Bayesian relaxed-clock dating analysis 2 years after that. The Bayesian approach to phylogenetics was developed further by Larget and Simon (1999) and widespread adoption was facilitated by the user-friendly program, MrBayes (Huelsenbeck and Ronquist 2001). Huelsenbeck et al. (2001) highlighted the benefits of Bayesian phylogenetic analysis in an article in Science. In 2007, Drummond and Rambaut (2007) introduced the BEAST package

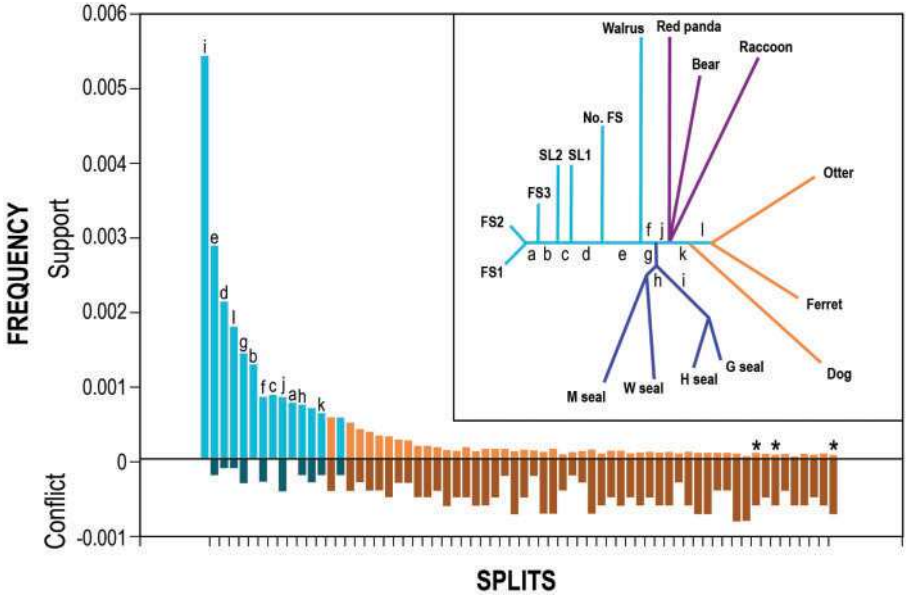

FIGURE 2. Lento plots of conflict and support based on spectral analysis redrawn from Lento et al. (1995). Support above the zero line and conflict below in darker shades. Splits on the tree are labeled with lower case letters and correspond to the blue bars on the graph. The unlabeled two blue bars show support for the resolution of the bear, raccoon, red panda trichotomy after the application of the LogDet correction. The brown splits are not contained on the preferred tree. Trivial splits are not shown. Asterisks indicate splits representing an alternative hypothesis. FS = fur seal; SL = sea lion; No. FS = Northern fur seal; M = monk; W = Weddell; H = Harbor; G = Gray. Color version can be found online.

which was the first user-friendly Bayesian program for dating phylogenetic trees using relaxed-clock methods (Drummond and Rambaut 2007).

The systematics community was extremely receptive to the Bayesian approach because trees and supports were seemingly much faster to calculate than in maximum likelihood analyses and PP were higher numbers than bootstrap supports and looked better in publications! Furthermore, PP was straightforward to interpret because they were calculated as the percentage of all the trees in the Markov Chain Monte Carlo (MCMC) chain after burn-in in which a clade was found. This is equivalent to the probability that a particular node is found in the true tree given the data, the model, and the priors including branch lengths and topology. The speed seemed too good to be true, and it was. It turned out that at first, users were not running the chains long enough to attain convergence (Nylander et al. 2004). Once that was realized, Bayesian analysis was much slower. Another obstacle to its initial adoption was the assertion by Suzuki et al. (2002) that Bayesian phylogenetics was flawed because it produced, in simulation, high PP for nonexistent clades. Also worrying, they said, was that high PP were seen in real data at nodes where maximum likelihood bootstraps were low. Fortunately, Lewis et al. (2005) were quick to point out that this "star tree paradox" resulted from a restrictive model that precluded assigning any PP mass to a tree topology containing one or more polytomies. They developed a reversible-jump MCMC method that introduced a topology prior that allowed both fully resolved as well as polytomous tree topologies. Other authors found that support could be anomalously large for trees with long branches together even when the generating tree had a polytomy at that node (star tree). True trees with long branches that are connected by small middle edge-lengths will result in similar estimation behavior to a star tree (Steel and Matsen 2007; Yang 2007; Susko 2008, 2015). Steel and Matsen (2007) proved mathematically that when the number of sites approaches infinity in a star tree, the PP of any one resolution of a three-way split does not converge to one-third for each topology but that there is a distinct probability that PP will be large. Yang (2007) obtained the same result using Laplacian expansion. He found PP for polytomies to be very sensitive to the prior. Similar problems with polytomies or near polytomies in likelihood analyses were explored by Simmons and Norton (2014) who also present, in their introduction, an excellent mini-review of problems caused by different implementations and interpretations of the bootstrap.

The appearance of Bayesian PP with their higher numerical values compared to bootstrap support immediately stimulated a series of comparison papers employing simulations and empirical data. Findings included the demonstration that PP was sensitive to violations of the model (Buckley 2002; Erixon et al. 2003; Huelsenbeck and Rannala 2004), that modest overparameterization was not a problem for PP in that it produced only a slightly increased variance (Cunningham et al. 1998), and that underparameterization can inflate PP (Erixon et al. 2003; Huelsenbeck and Rannala 2004; Lemmon and Moriarty 2004). Users should not have been surprised that PP were correlated, but different, from bootstrap support because the two values were measuring different things.

The PP represents the actual probability that a given node is found in the true tree, given the data, the model, and the priors. Bootstrap supports are more nebulous and interpreted differently by numerous statisticians, as described earlier.

Even More Data

From 1987 to 2014 automation of the Sanger method had enabled high throughput sequencing, necessitating methods that could work for even more characters and taxa, mostly from mitochondrial DNA at first (Kocher et al. 1989; Simon et al. 1994) and later for nuclear DNA. Beginning in the late 1990s massively parallel "next-generation" or "second-generation" sequencing (e.g., 454, Illumina, Ion Torrent) gradually replaced automated Sanger (Shendure et al. 2017). As more DNA sequence data became available it was quickly apparent that data partitioning was of the utmost importance for model fitting (Castoe et al. 2004; Nylander et al. 2004; Brandley et al. 2005; Simon et al. 2006). Partition finder (Lanfear et al. 2012, 2016) made data easier to handle by condensing model-fitted-partitioned data into a smaller number of partitions each of which shared similar properties. Arguments over whether it was better to add longer DNA sequences or more taxa (Hillis et al. 2003; Rosenberg and Kumar 2003) gradually faded from memory. Single-molecule-real-time, "third generation" PCR-less DNA sequencing technology diversified between 2009 and 2014 and continues to improve today resulting in more and cheaper DNA data (Eid et al. 2009; Kulski 2016; Shendure et al. 2017). Bioinformatics rather than wet-lab technology has become the limiting step in the generation of DNA sequence data.

Speed

The realization that Bayesian phylogenetics and PP were not the solution to all problems, caused renewed interest in maximum likelihood methods and improvement in their speed helped increase their popularity as more data accumulated. Schmidt et al. (2002) released Tree Puzzle, followed quickly by IQP-NNI (Vinh and von Haeseler 2004) a maximum likelihood method that was based on quartets and used parallel computing. Other fast ML programs such as RAxML (Stamatakis et al. 2005a,b), PHYML (Hordijk 2005), and GARLI (Zwickl 2006), followed quickly.

Rapid Branch Supports

Aided by continually increasing computer speed, phylogenetic methods developers produced techniques to address the increasing numbers of taxa and larger amounts of sequence data. In 2008, RAxML rapid bootstraps (Stamatakis et al. 2008) allowed the calculation of approximate bootstrap support in a much shorter time frame but these were later criticized for providing inflated branch supports (Simmons and Norton 2013, 2014) especially in cases where branches are very short or missing data are distributed nonrandomly (Lemmon et al. 2009). Other fast branch support techniques such as the parametric approximate likelihood ratio test (Anisimova and Gascuel 2006), and the nonparametric Shimodaira Hasegawa-like approximate likelihood ratio test (Guindon et al. 2010), proved to be accurate and relatively robust to small to moderate model violations (Anisimova et al. 2011). Ultrafast bootstrap approximation (UF-Boot in IQTree) was released in 2013 (Minh et al. 2013) and an improvement was announced 5 years later (Hoang et al. 2018) that minimized the chances of artifacts caused by severe model violations and polytomies/rapid radiations (as pointed out by Simmons and Norton 2014). Lemoine et al. (2018) defined a "transfer bootstrap expectation" (TBE) to address the problem that in phylogenies with very large numbers of taxa, deep branches are poorly supported because it is harder for resampled clades to be identical to the original tree. Minor rearrangements in large clades are inevitable. TBE takes into account the number of branches that differ between two clades in each resampled tree to give a more relaxed support measure for trees with large numbers of taxa that is easy to compute, has higher supports than normal phylogenetic bootstraps, and has a low number of falsely supported branches. TBE is a measure of branch repeatability. It is now implemented in RAxML-NG (Lutteropp et al. 2020) as well as in the original web tool 'BoosterWeb' (http://booster.c3bi.pasteur.fr) Nick Goldman (personal communication) points out that TBE must be used with care due to sampling issues. For example, if many closely related taxa are added to the tree TBE values will increase across the entire tree. This is because the measure is based on counting the number of sequences sampled, not taking into account their variation. The TBE is the proportion of taxa that on average over bootstrap replicates remain in the smaller of the two clades defined by a split.

Species Trees

Second- and third-generation DNA sequencing methods allowed the amount of nuclear DNA data to catch-up with and then surpass the abundant organelle sequence data that had been made possible by PCR and Sanger sequencing. Multiple, potentially independent, genes facilitated the search for species histories as opposed to gene histories. Coalescent methods proliferated starting in 2005 and continue to improve today with consideration of both ILS and hybridization (e.g., Nakhleh 2013; Chifman and Kubatko 2014; Ogilvie et al. 2016; Allman et al. 2019), providing an alternative to concatenation of genes in super-matrices (Edwards et al. 2016; Springer and Gatesy 2016).

Conflict versus Support Revisited in Species Trees

Because species trees contain multiple gene trees, an obvious measure of branch support was the proportion

of inferred single-locus trees that contain a particular branch of interest; this measure is called the gene concordance factor (gCF) (Gadagkar et al. 2005; Ane et al. 2007; Baum 2007). Minh et al. (2020) modify this statistic to account for variable taxon sampling across gene trees and introduce a new "site concordance factor" (sCF) that, as the name suggests, measures the proportion of sites supporting a particular branch in the reference tree. The sCF is related to spectral analysis (Hendy and Penny 1993) but differs in that it averages site supports for individual splits over a large set of repeatedly subsampled quartets. Minh et al. (2020) demonstrate that concordance factor values are not necessarily correlated with standard bootstrap support and suggest that calculating both can be useful.

Salichos and Rokas (2013) used the gCF, which they renamed "Gene Support Frequency" (GSF), to calculate an "Internode Certainty" (IC) statistic that compared the GSF of the best-supported split to the GSF of the best-represented conflicting or competing split. Ties for the best-represented competing split are broken arbitrarily. They annotated each node of the tree with values for both support (GSF) and conflict (IC), as did Lento et al. (1995) for concatenated DNA sequence data. Quartet-based internode statistics were developed further by Salichos et al. (2014), Zhou et al. (2020), and Pease et al. (2018) and are discussed in this symposium.

Salichos and Rokas (2013) also listed "standard" methods that help remove the conflict between gene trees. These included: 1) removing sites containing gaps; 2) removing 'rogue' genes that produced alignments of bad quality; 3) removing unstable and quickly evolving species (long branches); 4) using only genes that recover a particular internode widely regarded as certain or well established, from prior data; 5) using only slowly evolving genes; and 6) using conserved amino acid substitutions or indels (as shared blocks). They did not mention removing sites with extremely high substitution rates (Cummins and McInerney 2011), or genes that show extreme heterogeneity in nucleotide bias across taxa, but these strategies have proved effective in other cases (Collins et al. 2005; Skinner et al. 2019). Lewis et al. (2016) review methods for assessing phylogenetic information content and develop an entropy-based method for Bayesian phylogenetics.

New Measures of Branch Support

This symposium highlights new measures of branch support and methods of understanding the information in DNA sequence data that have grown out of the rich history of branch support thinking reviewed here, new work on Bayes factors (Brown and Thomson 2017), gene genealogy interrogation (Arcila et al. 2017), identification of outlier genes (Walker et al. 2018), or outlier sites that drive phylogenetic relationships (Shen et al. 2017), site and gCFs (Minh et al. 2020), and improvements to species networks given hybridization (Allman et al. 2019; Mitchell et al. 2019). It is important to attack the problem of branch support in multiple ways and the extensions of the above studies, presented in this symposium, combine to do just that.

Funding

This work was supported by National Science Foundation (DEB 1655891).

Acknowledgments

I thank Jeremy Brown, Bob Thompson, and Rob Lanfear for organizing the SSB Evolution 2019 symposium, "New Measures of Phylogenetic Support for the Genomic Era" and for inviting me to participate. I thank the following people for helpful comments on the manuscript: Jack Sullivan, Eric Gordon, Rob Lanfear, Mark Simmons, and an anonymous reviewer. Thanks to Paul Lewis for comments on the star-tree paradox and Rob Lanfear and Nick Goldman for comments related to the interpretation of the transfer bootstrap expectation. Thanks to NSF grant DEB 1655891 for supporting this project. Finally, I thank all of the phylogenetics methods developers and sequencing technologists who make our empirical work possible.

References

Allman E.S., Banos H., Rhodes J.A. 2019. NANUQ: a method for inferring species networks from gene trees under the coalescent model. Algorithms Mol. Biol. 14:24 (25 pages).

Ane C., Larget B., Baum D., Smith S., Rokas A. 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24:412–426.

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

Anisimova M., Gil M., Dufayard J.F., Dessimoz C., Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60:685–699.

Arcila D., Ortí G., Vari R., Armbruster J.W., Stiassny M.L.J., Ko K.D., Sabaj M.H., Lundberg J., Revell L.J., Betancur-R R. 2017. Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. Nat. Ecol. Evol. 1:20 (10 pages).

Bandelt H.-J., Dress A. 1992. A canonical decomposition theory for metrics on a finite set. Adv. Math. 92:47–105.

Baum D.A. 2007. Concordance trees, concordance factors, and the exploration of reticulate genealogy. Taxon 56:417–426.

Berry V., Gascuel O. 1996. On the interpretation of bootstrap trees: appropriate threshold of clade selection and induced gain. Mol. Biol. Evol. 13:999–1011.

Brandley M.C., Schmitz A., Reeder T.W. 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54:373–390.

Brown J.M., Thomson R.C. 2017. Bayes factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses. Syst. Biol. 66:517–530.

Buckley T.R. 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 51: 509–523.

Buckley T.R., Simon C., Chambers G.K. 2001. Exploring among-site rate variation models in a maximum likelihood framework using empirical data: effects of model assumptions on estimates of topology, branch lengths, and bootstrap support. Syst. Biol. 50:67–86.

Castoe T., Doan T., Parkinson C. 2004. Data partitions and complex models in Bayesian analysis: the phylogeny of Gymnophthalmid lizards. Syst. Biol. 53:448–469.

Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30:3317–3324.

Collins T.M., Fedrigo O., Naylor G.J. 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. Syst. Biol. 54:493–500.

Cummins C.A., McInerney J.O. 2011. A method for inferring the rate of evolution of homologous characters that can potentially

improve phylogenetic inference, resolve deep divergence and correct systematic biases. Syst. Biol. 60:833–844.
Cunningham C.W., Zhu H., Hillis D.M. 1998. Best-fit maximum likelihood models for phylogenetic inference: empirical tests with known phylogenies. Evolution 52:978–987.
Dietrich C.H., Allen J.M., Lemmon A.R., Lemmon E.M., Takiya D.M., Evangelista O., Walden K.K.O., Grady P.G.S., Johnson K.P. 2017. Anchored hybrid enrichment-based phylogenomics of leafhoppers and treehoppers (Hemiptera: Cicadomorpha: Membracoidea). Insect Syst. Divers. 1:57–72.
Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:0699–0710.
Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214 (8 pages).
Edwards S.V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94:447–462.
Efron B. 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7:1–26.
Efron B., Halloran E., Holmes S. 1996. Bootstrap confidence levels for phylogenetic trees. Proc. Natl. Acad. Sci. USA 93:13429–13434.
Eid J., Fehr A., Gray J., Luong K., Lyle J., Otto G., Peluso P., Rank D., Baybayan P., Bettman B., Bibillo A., Bjornson K., Chaudhuri B., Christians F., Cicero R., Clark S., Dalal R., Dewinter A., Dixon J., Foquet M., Gaertner A., Hardenbol P., Heiner C., Hester K., Holden D., Kearns G., Kong X., Kuse R., Lacroix Y., Lin S., Lundquist P., Ma C., Marks P., Maxham M., Murphy D., Park I., Pham T., Phillips M., Roy J., Sebra R., Shen G., Sorenson J., Tomaney A., Travers K., Trulson M., Vieceli J., Wegener J., Wu D., Yang A., Zaccarin D., Zhao P., Zhong F., Korlach J., Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. Science 323:133–138.
Erixon P., Svennblad B., Britton T., Oxelman B. 2003. The reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. 52:665–673.
Felsenstein J. 1983. Statistical inference of phylogenies. J. R. Stat. Soc. A 146:246–272.
Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.
Fisher R.A. 1935. The design of experiments. Edingurgh: Oliver and Boyd.
Frati F., Simon C., Sullivan J., Swofford D.L. 1997. Evolution of the mitochondrial cytochroma oxidase II gene in Collembola. J. Mol. Evol. 44:145–158.
Gadagkar S.R., Rosenberg M.S., Kumar S. 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J. Exp. Zool B 304B:64–74.
Guindon S., Dufayard J.-F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.
Hedtke S.M., Townsend T.M., Hillis D.M., Collins T. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst. Biol. 55:522–529.
Hendy M.D., Penny D.1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38:297–309.
Hendy M.D., Penny D. 1993. Spectral analysis of phylogenetic data. J. Classif. 10:5–24.
Hillis D.M., Bull J.J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42:182–192.
Hillis D.M., Pollock D., McGuire J.A., Zwickl D.J. 2003. Is sparse taxon sampling a problem for phylogenetic inference? Syst. Biol. 52:124–126 (8 pages).
Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. 2018. Ufboot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35:518–522.
Holmes S. 2003. Bootstrapping phylogenetic trees: theory and methods. Stat. Sci. 18:241–255.
Hordijk W., Gascuel O. 2005. Improving the efficiency of SPR improves in phylogenetic tree search methods based on maximum likelihood. Bioinformatics 21:4338–4347.
Huelsenbeck J.P., Rannala B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst. Biol. 53:904–913.
Huelsenbeck J.P., Ronquist F., Nielsen R., Bollback J.P. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314.
Huson D.H. 1998. SplitsTree: a program for analyzing and visualizing evolutionary data. Bioinformatics 14:68–73.
Kass R.E., Raftery A.E. 1995. Bayes factors. J. Am. Stat. Assoc. 90:773–795.
Kocher T.D., Thomas W.K., Meyer A., Edwards S.V., Paabo S., Villablanca F.X., Wilson A.C. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86:6196–6200.
Kulski J.K. 2016. Next-generation sequencing — an overview of the history, tools, and "omic" applications. In: Chapter 1 in Next Generation Sequencing - Advances, Applications and Challenges. Intech Open. doi: 10.5772/61964.
Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.
Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34:772–773.
Larget B., Simon D.L. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16:750–759.
Lemmon A.R., Brown J.M., Stanger-Hall K., Lemmon E.M. 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. Syst. Biol. 58:130–145.
Lemmon A.R., Moriarty E.C. 2004. The importance of proper model assumption in Bayesian phylogenetics. Syst. Biol. 53:265–277.
Lemoine F., Domelevo Entfellner J.B., Wilkinson E., Correia D., Dávila Felipe M., De Oliveira T., Gascuel O. 2018. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature 556:452–456.
Lento G.M., Hickson R.E., Chambers G.K., Penny D. 1995. Use of spectral analysis to test hypotheses on the origin of pinnipeds. Mol. Biol. Evol. 12:28–52.
Lewis P.O., Chen M.H., Kuo L., Lewis L.A., Fucikova K., Neupane S., Wang Y.B., Shi D. 2016. Estimating Bayesian phylogenetic information content. Syst Biol 65:1009–1023.
Lewis P.O., Holder M.T., Holsinger K.E. 2005. Polytomies and Bayesian phylogenetic inference. Syst. Biol. 54:241–253.
Lutteropp S., Kozlov A.M., Stamatakis A. 2020. A fast and memory-efficient implementation of the transfer bootstrap. Bioinformatics 36:2280–2281.
Maddison, W.P., Knowles, L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21–30.
Maddison W.P. 1986. MacClade version 1.0, an interactive computer program for reconstructing phylogeny and analyzing character evolution. Distributed by the authors.
Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.
Maddison W.P., Maddison D.R. 1987. MacClade version 2.1., an interactive computer program for reconstructing phylogeny and analyzing character evolution. Distributed by the authors.
Minh B.Q., Hahn M.W., Lanfear R. 2020. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37:2727–2733.
Minh B.Q., Nguyen M.A.T., von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30:1188–1195.
Mitchell J.D., Allman E.S., Rhodes J.A. 2019. Hypothesis testing near singularities and boundaries. Electron. J. Stat. 13:2150–2193.
Mueller L.D., Ayala F.J. 1982. Estimation and interpretation of genetic distance in empirical studies. Genet. Res. 40:127–137.
Mullis K.B., Faloona F.A. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Meth. Enzymol. 155:335–350.
Nakhleh L. 2013. Computational approaches to species phylogeny inference and gene tree reconciliation. Trends Ecol. Evol. 28:719–728.

Nylander J., Ronquist F., Huelsenbeck J., Nieves-Aldrey J. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53:47–67.

Nylander J.A.A., Wilgenbusch J.C., Warren D.L., Swofford D.L. 2007. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24:581–583.

Ogilvie H.A., Heled J., Xie D., Drummond A.J. 2016. Computational performance and statistical accuracy of *BEAST and comparisons with other methods. Syst. Biol. 65:381–396.

Page R.D.M. 1996. On consensus, confidence and "total evidence". Cladistics 12:83–92.

Page R.D.M., Holmes E.C.. 1998. Molecular evolution: a phylogenetic approach. Oxford: Blackwell Science Limited.

Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

Pease J.B., Brown J.W., Walker J.F., Hinchliff C.E., Smith S.A. 2018. Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. Am. J. Bot. 105:385–403.

Penny D., Hendy M.D. 1985. Testing methods of evolutionary tree construction. Cladistics 1:266–278.

Penny D., Hendy M.D., Steel M.A. 1992. Progress with methods for constructing evolutionary trees. Trends Ecol. Evol. 7:73–79.

Penny D., Watson E.E., Hickson R.E., Lockhart P.J. 1993. Recent progress with methods for evolutionary trees. N.Z.J. Bot. 31:275–278.

Quenouille M. 1949. Approximate tests of correlation in time series. J. R. Stat. Soc. B 11:18–84.

Rannala B., Yang Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43:304–311.

Rodgers J.L. 1999. The bootstrap, the jackknife, and the randomization test: a sampling taxonomy. Multivar. Behav. Res. 34:441–456.

Rosenberg M.S., Kumar S. 2003. Taxon sampling, bioinformatics and phylogenomics. Syst. Biol. 52:119–123.

Saiki R.K., Scharf S., Faloona F., Mullis K.B., Horn G.T., Erlich H.A., Arnheim N. 1985. Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350–1354.

Salichos L., Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497:327–331.

Salichos L., Stamatakis A., Rokas A. 2014. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. Mol. Biol. Evol. 31:1261–1271.

Sanderson M. 1995. Objections to bootstrapping phylogenies: a critique. Syst. Biol. 44:299–320.

Sanderson M.J., Wojciechowski M.F., Olmstead R. 2000. Improved bootstrap confidence limits in large-scale phylogenies, with an example from neo-Astragalus (Leguminosae). Syst. Biol. 49: 671–685.

Schmidt H.A., Strimmer K., Vingron M., von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504.

Shen X.-X., Hittinger C.T., Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat. Ecol. Evol. 1:126 (10 pages).

Shendure J., Balasubramanian S., Church G.M., Gilbert W., Rogers J., Schloss J.A., Waterston R.H. 2017. DNA sequencing at 40: past, present and future. Nature 550:345–353.

Simmons M.P., Norton A.P. 2013. Quantification and relative severity of inflated branch-support values generated by alternative methods: an empirical example. Mol. Phylogenet. Evol. 67:277–296.

Simmons M.P., Norton A.P. 2014. Divergent maximum-likelihood-branch-support values for polytomies. Mol. Phylogenet. Evol. 73:87–96.

Simon C., Buckley T.R., Frati F., Stewart J.B., Beckenbach A.T. 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 37:545–579.

Simon C., Frati F., Beckenbach A., Crespi B., Liu H., Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved PCR primers. Ann. Entomol. Soc. Am. 87:651–701.

Simon C., Nigro L., Sullivan J., Holsinger K., Martin A., Grapputo A., Franke A., McIntosh C. 1996. Large differences in substitutional pattern and evolutionary rate of 12S ribosomal RNA genes. Mol. Biol. Evol. 13:923–932.

Skinner R.K., Dietrich C.H., Walden K.K.O., Gordon E., Sweet A.D., Podsiadlowski L., Petersen M., Simon C., Takiya D.M., Johnson K.P. 2019. Phylogenomics of Auchenorrhyncha (Insecta: Hemiptera) using transcriptomes: examining controversial relationships via degeneracy coding and interrogation of gene conflict. Syst. Entomol. 45:85–113.

Springer M.S., Gatesy J. 2016. The gene tree delusion. Mol. Phylogenet. Evol. 94:1–33.

Stamatakis A., Hoover P., Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57:758–771.

Stamatakis A., Ludwig T., Meier H., RAxML-III. 2005a. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21:456–463.

Stamatakis A., Ott M., Ludwig T. 2005b. RAxML-OMP: an efficient program for phylogenetic inference on SMPs. Lect. Notes Comput. Sci. 3606:288–302.

Steel M., Matsen F.A. 2007. The Bayesian "star paradox" persists for long finite sequences. Mol. Biol. Evol. 24:1075–1079.

Sullivan J., Demboski J.R.,. Bell K.C, Hird S., Sarver B., Reid N., Good J.M. 2014. Divergence with gene flow within the recent chipmunk radiation (Tamias). Heredity 113:185–194.

Sullivan J., Joyce P. 2005. Model selection in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 36:445–466.

Susko E. 2008. On the distributions of bootstrap support and posterior distributions for a star tree. Syst. Biol. 57:602–612.

Susko E. 2009. Bootstrap support is not first-order correct. Syst. Biol. 58:211–223.

Susko E. 2015. Bayesian long branch attraction bias and corrections. Syst. Biol. 64:243–255.

Suzuki Y., Glazko G.V., Nei M. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acad. Sci. USA 99:16138–16143.

Swofford D.L. 1985. PAUP: phylogenetic analysis using parsimony. User's manual. Illinois: Illinois Natural History Survey Champaign.

Swofford D.L., Olsen G.P., Waddell P.J., Hillis D.M. 1996. Phylogenetic inference. In: Hillis D.M., Moritz C., Mable B.K., editors. Molecular systematics. Sunderland, MA: Sinauer. p. 407–514.

Swofford D.L., Waddell P.J., Huelsenbeck J.P., Foster P.G., Lewis P.O., Rogers J.S. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst. Biol. 50:525–539.

Thorne J.L., Kishino H., Painter I.S. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.

Tukey J.W. 1958. Bias and confidence in not quite large samples (abstract). Ann. Math. Stat. 29:614.

Vinh L.S., von Haeseler A. 2004. IQPNNI: moving fast through tree space and stopping in time. Mol. Biol. Evol. 21:1565—1571.

Walker J.F., Brown J.W., Smith S.A. 2018. Analyzing contentious relationships and outlier genes in phylogenomics. Syst. Biol. 67:916–924.

Wilkinson M. 1996. Majority-rule reduced consensus trees and their use in bootstrapping. Mol. Biol. Evol. 13:437–444.

Yang Z. 2007. Fair-balance paradox, star-tree paradox, and Bayesian phylogenetics. Mol. Biol. Evol. 24:1639–1655.

Zharkikh A., Li W.-H. 1992a. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. 1. Four taxa with a molecular clock. Mol. Biol. Evol. 9:1119–1147.

Zharkikh A., Li W.-H. 1992b. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. II. Four taxa without a molecular clock. J. Mol. Evol. 35:356–366.

Zhou, X., Lutteropp, S., Czech, L., Stamatakis, A., Looz, M.V., Rokas, A. 2020. Quartet-based computations of internode certainty provide robust measures of phylogenetic incongruence. Syst. Biol. 69:308–324.

Zwickl D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. [PhD], Austin (TX): The University of Texas.