COMMENTARY

Felsenstein Phylogenetic Likelihood

David Posada[1,2,3] · **Keith A. Crandall**[4,5]

Abstract

In 1981, the *Journal of Molecular Evolution* (JME) published an article entitled "*Evolutionary trees from DNA sequences: A maximum likelihood approach*" by Joseph (Joe) Felsenstein (J Mol Evol 17:368–376, 1981). This groundbreaking work laid the foundation for the emerging field of statistical phylogenetics, providing a tractable way of finding maximum likelihood (ML) estimates of evolutionary trees from DNA sequence data. This paper is the second most cited (more than 9000 citations) in JME after Kimura's (J Mol Evol 16:111–120, 1980) seminal paper on a model of nucleotide substitution (with nearly 20,000 citations). On the occasion of the 50th anniversary of JME, we elaborate on the significance of Felsenstein's ML approach to estimating phylogenetic trees.

Keywords Phylogeny · Maximum likelihood · Models of nucleotide substitution · Evolution

Molecular Phylogenetics in the 80's

Before delving into the substance of this seminal paper, it is important to understand the historical context in which it was written. "In the development of scientific methodology there is no new thing under the sun. Every 'new' idea is like another node in a spreading network" (Edwards 2009). In the

✉ David Posada
  dposada@uvigo.es

✉ Keith A. Crandall
  kcrandall@gwu.edu

1 CINBIO, Universidade de Vigo, 36310 Vigo, Spain

2 Department of Biochemistry, Genetics, and Immunology, Universidade de Vigo, 36310 Vigo, Spain

3 Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Vigo, Spain

4 Computational Biology Institute and Milken Institute School of Public Health, The George Washington University, Washington, DC 20052, USA

5 Department of Biostatistics & Bioinformatics, Milken Institute School of Public Health, The George Washington University, Washington, DC 20052, USA

early 1960s, AWF Edwards and LL Cavalli-Sforza were set on a clear mission to develop modern statistical approaches for reconstructing phylogenetic trees from genetic data, using computers. Over a very short period of time, they were able to develop three methods: least-squares (LS)–see also Fitch and Margoliash (1967), maximum parsimony (MP), and ML (Edwards and Cavalli-Sforza 1963a, b, 1964, 1965). This was before DNA or protein sequences were even available, and their focus was on the use of blood group allele frequency data to recover the history of human populations; for an account of this historical period, see Edwards (2009).

Remarkably, while the LS and MP approaches rapidly became quite popular, during the 1970s the method of ML was pretty much ignored by most researchers. The reasons for this were several, but the most obvious was the simplicity of the former methods, a clear advantage at a time when computers were still very slow and uncommon. Since Edwards and Cavalli-Sforza, the development of ML in phylogenetics was clearly championed by Joe Felsenstein (but see Neyman 1971; Kashyap and Subas 1974), starting with his Ph.D. thesis (Felsenstein 1968)—a current read of which tells us how much of a visionary he was in many regards. In 1973, Felsenstein already simplified the model for quantitative characters proposed by Edwards and Cavalli-Sforza in 1964 to make it tractable (Felsenstein 1973a). Also in 1973, he set the basis for his 1981 JME publication, proposing an algorithm for computing the likelihood of a tree for discrete characters advancing the "pruning" technique (see below;

also present in the other 1973 paper) and proposing new probabilistic models of change (see also Jukes and Cantor 1969; Felsenstein 1973b). During this period, Felsenstein also showed that MP could be inconsistent under certain realistic scenarios (Felsenstein 1978) where "long-branch attraction" is a concern—the so-called "Felsenstein zone" (Huelsenbeck and Hillis 1993).

The major breakthroughs in DNA sequencing technologies took place in the second part of the 70′s (Maxam and Gilbert 1977; Sanger et al. 1977), and during the 80′s an explosion of DNA sequences started to revolutionize the incipient field of molecular phylogenetics. In particular, Felsenstein pointed out that parsimony methods implicitly assume that change is improbable a priori (Felsenstein 1973b, 1979). These accumulating DNA sequence data were clearly indicating that this assumption was not correct and therefore, the need for a tractable likelihood method was palpable. Felsenstein's publication in 1981 was exceptionally timely, given the deluge of DNA sequence data that continues to this day.

A Maximum Likelihood Approach for DNA Sequences

The main goal of Felsenstein's 1981 JME article was to show how to efficiently calculate the probability of a set of aligned nucleotide sequences given a phylogenetic tree. In doing this, and in only nine pages, Felsenstein made several significant contributions, namely: (i) a probabilistic model of nucleotide substitution, (ii) an algorithm to optimize branch lengths, (iii) an algorithm to search for the most likely tree, (iv) a computer program to implement these calculations, (v) likelihood ratio tests to compare phylogenetic hypotheses, and (vi) the first empirical ML tree obtained from (ribo) nucleotide sequences.

The Tree Likelihood

The likelihood of a hypothesis (H) given some data (D) is P(D|H), the conditional probability of observing D given that H is correct (Edwards 1972). In phylogenetics, the tree likelihood is the probability of a sequence alignment (S) given the tree (topology and branch lengths) (T) together with a model of nucleotide substitution (M), that is, P(S|T,M). To facilitate the computation of the tree likelihood, Felsenstein assumed that sites change independently from each other and across different branches, two premises that persist in most phylogenetic methods today, despite being unrealistic.

For a given site, the likelihood of the tree can then be computed simply as the product of the probabilities of change/no change in each branch, times the prior probabilities of each

of the four DNA bases. The problem is that this approach implies, for a rooted tree, the sum of 4^{n-1} products of a number of terms equal to the total number of branches plus one, which can be a lot (for a rooted tree with n sequences at the tips, there are 4 possible nucleotides at each of the $n–1$ interior nodes, and $2n–2$ branches). For example, for only 12 sequences we would need to sum already $4^{11} = 4{,}194{,}304$ products of 23 terms. Therefore, Felsenstein proposed to conduct this computation in terms of conditional likelihoods, starting from the tips and moving towards the root, in a well-known movement called postorder traversal. This "pruning" algorithm had been in fact already proposed by Felsenstein himself for a more general case (Felsenstein 1973b), and in turn, it was based on the "peeling algorithm" for computing likelihoods on human pedigrees (Hilden 1970; Elston and Stewart 1971; Heuch and Li 1972). It is worth mentioning that Felsenstein already devised in his Ph.D. thesis in 1968 a special case of the pruning algorithm for continuous characters changing by Brownian Motion.

The F81 Substitution Model

To compute the tree likelihood, it is necessary to calculate the probability of changing from one nucleotide to another along a branch of a given length in time units. Generally, we do not know the absolute times, so usually, the time unit is arbitrarily set as the expected time required for a single change. In this way, branch lengths are conveniently scaled in expected nucleotide substitutions per site. Here, Felsenstein proposed a simple reversible Markov process similar to the one previously proposed by Kaplan and Langley (1979) in JME for restriction sites. A Markov process implies that the probability of one nucleotide changing to another does not depend on its previous states. Reversible means that this process will look the same backward or forward in time. The process is stationary as the probabilities of change among nucleotides are constant. Felsenstein assumed that these change probabilities only depended on the frequency of the target nucleotide. This model of nucleotide substitution is known as the F81 model.

Finding the Maximum Likelihood Tree

Note that this model does not assume a molecular clock (i.e., a constant mutation rate) nor does it require the sequences to be contemporaneous. In a short appendix, Felsenstein proved what he called the "Pulley principle" by which the likelihood of a tree is not affected by the position of the root. This implies that under this model, the tree does not contain information about the location of the root, which is very convenient for computational purposes. In particular, it allows one to optimize the likelihood of each branch iteratively in a way that maximizing the likelihood of the tree for
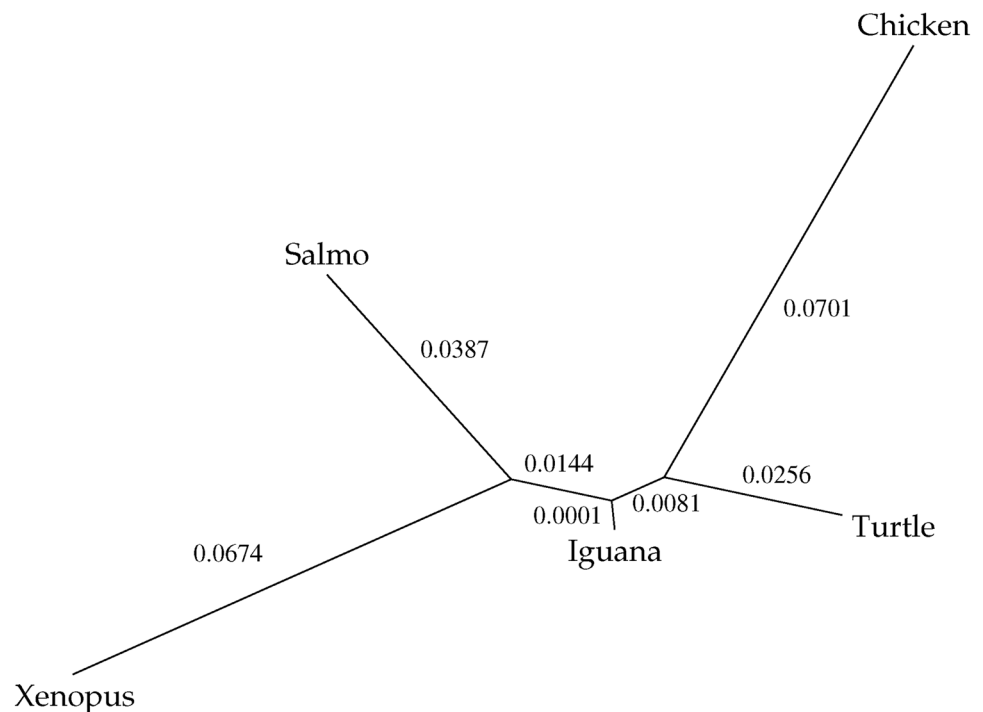
a given topology becomes much more feasible, as he showed in the paper.

Indeed, once we know how to maximize the likelihood for a given tree topology, we still need to find the best tree across all possible topologies. For this Felsenstein proposed to obtain the maximum likelihood tree using a random step-wise addition algorithm (Eck and Dayhoff 1966; Kluge and Farris 1969), by which sequences are added one by one, always looking for the placement with the highest likelihood and performing local rearrangements between additions to try to improve the likelihood score. Indeed, Felsenstein acknowledged this procedure can result in local optima, and recommended repeating this process multiple times but changing the order in which sequences are added.

A Computer Program: *dnaml*

In the article, Felsenstein also made an important announcement, the availability of a computer program (*dnaml*) for optimizing the branch lengths, as part of a package of programs for numerical analysis of evolutionary trees, the great PHYLIP package (https://evolution.genetics.washington.edu/phylip.html). PHYLIP was first released in October 1980 and has helped tens of thousands of researchers across the globe. In version 1.7 (December 1981), that differed only a bit from the first version (1.0), *dnaml* was a program written in Pascal, with only 782 lines of code and able to deal by default at most with 15 sequences 60-bp long (although the users could change these limits when recompiling the program).

More Ideas

The article also hinted at possible extensions of the methodology, like the incorporation of uncertainty in the data –interpreting sequencing chromatograms was not always straightforward–, and a way of dealing with rate heterogeneity among sites. Moreover, the possibility of conducting a likelihood ratio test of the molecular clock, or to test alternative trees by evaluating the uncertainty in the estimation of specific branch lengths, was also mentioned. Indeed, all these ideas would be exploited in subsequent years by different researchers.

The First Maximum Likelihood Tree from (RNA) Sequences

The paper also offered the first application of the ML method to a set of aligned nucleotide sequences (5S and 5.8S RNA) obtained by Erdmann (1982), a process that would be repeated a million times after that. The resulting tree with trout, frog, turtle, iguana and chicken is shown in Fig. 1. Probably this is one of the first phylogenetic estimates in which the author is able to explain with a sound statistical basis that this result is not very reliable.

Fig. 1 Reestimation (by Joe Felsenstein) of the first published maximum likelihood tree obtained from a sequence alignment and published in his 1981 paper. The ML tree in the 1981 paper was computed incorrectly due to a bug in the code that computed branch lengths, and that even affected the tree topology. The bug was found a year or two later and corrected in the PHYLIP code (Felsenstein pers. comm.). Note that the Iguana branch has been artificially lengthened so that its branch length can be drawn next to it

Uses of the Phylogenetic Likelihood

The maximum likelihood framework developed by Felsenstein has been tremendously influential, providing a powerful methodology for phylogenetic inference that has been exploited by many empirical and theoretical researchers to this day (e.g., Ji et al. 2020). Felsenstein's 1981 article has

been cited by thousands of researchers around the world in over 9,000 publications (Fig. 2). As a side note, this is not Felsenstein's most cited work, which is instead his bootstrap paper in the journal *Evolution* (Felsenstein 1985), with more than 32,000 citations.

The reasons for this popularity are straightforward. The method of ML is a standard in statistics that brought a wealth of statistical theory to the field. Its application

Sum of Times Cited per Year
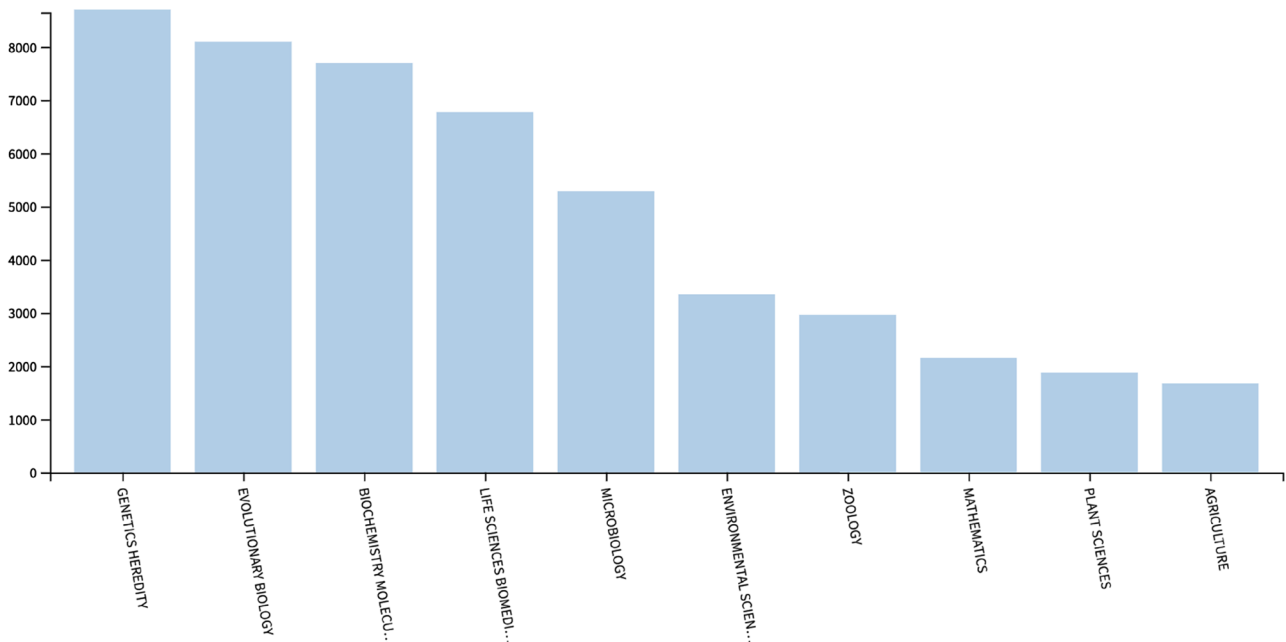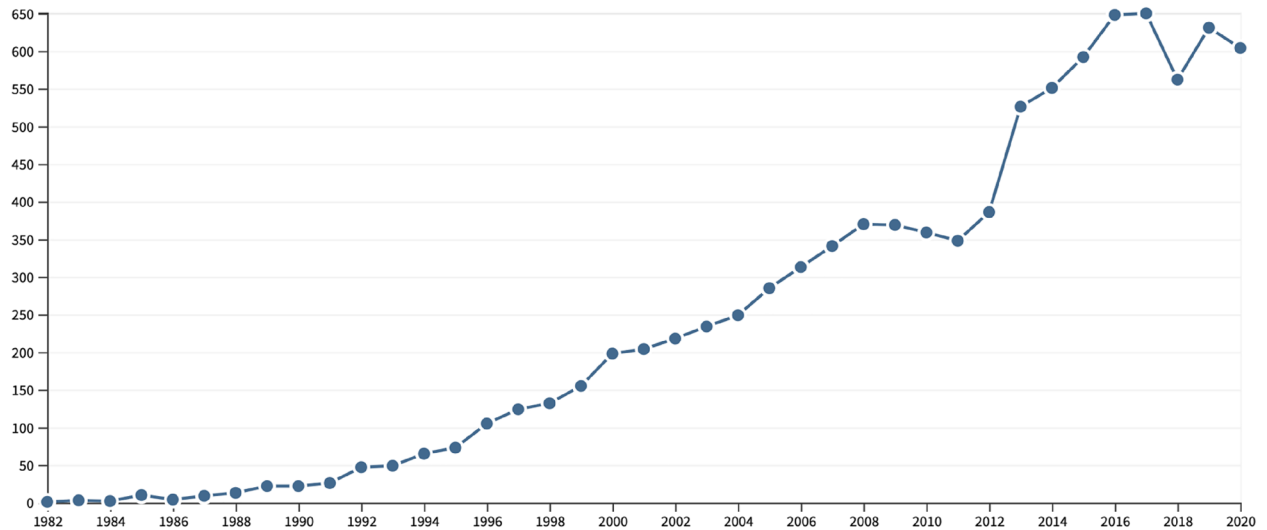
Fig. 2 Bibliographic impact of Felsenstein (1981). **a** Number of citations per year. **b** Distribution of citing articles across scientific fields (nearly 260,000 citing articles—those articles citing one or more of the articles citing Felsenstein's paper (and the paper itself). The total

number of citations is over 9000. The h-index of this list is 245. Data obtained from the Web of Science (webofknowledge.com) in December 2020

to phylogenetics permitted the use of complex models of evolution, including the ability to estimate model parameters and so make inferences about the process of evolution, providing the means to compare competing trees and models (Whelan et al. 2001). Below we describe some of the research prompted by Felsenstein's JME paper. Indeed, such a list cannot, and was not intended to, be exhaustive.

Molecular Systematics

The method of ML was swiftly adopted by many researchers, most rapidly by Hasegawa and collaborators, who used it to infer the tree of Hominoidea (Hasegawa and Yano 1984), and of eukaryotes (Hasegawa et al. 1985a). During the explosion of molecular phylogenetics in the 1990s, ML trees played a fundamental role in the field, but not without competing strategies, and often with accompanying philosophical debates. The method of ML was thoroughly scrutinized from multiple angles and benchmarked against other approaches (Saitou 1988; Saitou and Imanishi 1989; Goldman 1990; Fukami-Kobayashi and Tateno 1991; Hasegawa et al. 1991; Tateno et al. 1994; Kuhner and Felsenstein 1994; Yang 1994c, 1996; Gaut and Lewis 1995; Huelsenbeck 1995a, b). During this time, there was a fierce debate between supporters of MP and those of ML (Farris 1983; Felsenstein and Sober 1986; Sober 1991; Whiting 1998; Huelsenbeck 1998).

ML in phylogenetics has stood the test of time. It, along with Bayesian approaches, are very popular in phylogenetics today. The broad use of ML in phylogenetics has been greatly facilitated by novel software implementations such as RAxML (Stamatakis 2014, 2015), RAxML-NG (Kozlov et al. 2019), and IQTree (Nguyen et al. 2015; Minh et al. 2020)—all of which enable likelihood calculations on very large datasets.

Remarkably, in the current COVID-19 pandemic, ML phylogenetic inference is playing a fundamental role in understanding the origins, diversification and spread of SARS-CoV-2 (e.g., Fauver et al. 2020; Lam et al. 2020; Gonzalez-Reiche et al. 2020; Worobey et al. 2020; Boni et al. 2020).

Models of Molecular Evolution

Models of nucleotide substitution are a central part of the ML estimate of phylogeny as these models define the transitional probabilities from one nucleotide to another for the likelihood calculation. These models are not only integral to phylogeny estimation but are central to testing hypotheses related to tree topology, natural selection, and rates of evolution (see below). Jukes and Cantor (1969) (JC69) proposed the first such stochastic model of DNA substitution which assumed that all nucleotide substitutions occur at equal rates, implying that each nucleotide is equally likely to be a replacement at a given position in an alignment. This model formalized the molecular clock hypothesis put forth by Zuckerkandl (this journal's founder) and Pauling (Zuckerkandl and Pauling 1965) suggesting that the rate of evolution of a given protein (later DNA) is constant over time and across evolutionary lineages (Morgan 1998). Kimura introduced what became a series of extensions building on the JC69 model, introducing the Kimura 2-Parameter (K2P or K80) model (1980) that recognized the empirical result that transitions (changes in nucleotides within purines or pyrimidines resulting, therefore, in a similar biochemical shape) occurred at different frequencies than transversions (changes from a purine to pyrimidine or vice versa). The model presented by Felsenstein (F81) recognized that nucleotide frequencies are often divergent from the implicit assumption of equal across the four nucleotides. For example, in early sequence analyses of mitochondrial DNA, insects in particular showed very high frequencies of A's and T's relative to C's and G's (Jermiin and Crozier 1994). Thus, in the Felsenstein model, the rate of nucleotide substitution depends only on the equilibrium frequency of that nucleotide (Felsenstein 1981). Additional model parameters would be added to accommodate combining nucleotide frequency differences and differences in transitions and transversions (HKY; (Hasegawa et al. 1985b)), differences within transitions and transversions, etc., until eventually the development of the General Time Reversible model or GTR, allowing a unique parameter for each nucleotide substitution (Lanave et al. 1984; Tavaré 1986).

With the GTR model established for transition rates across nucleotides, researchers turned to other aspects of biology to expand models of DNA evolution, based on further empirical observations of model inadequacy relative to the accumulating DNA sequence data (Barry and Hartigan 1987). Yang (1993) introduced a ML approach, building on Felsenstein's work, that allows for substitution rates to vary across sites by implementing a gamma distribution of rate variation. An alternative to modeling rate variation to relax the independent and identically distributed assumption, was to model sites that appear to be invariant in a given alignment versus those that are variable (Waddell and Steel 1997), even when relaxing assumptions of stationarity, reversibility, and homogeneity (Jayaswal et al. 2007). More recently, Lie Markov models have been proposed to average over non-homogeneous Markov processes along the phylogeny (Sumner et al. 2012; Woodhams et al. 2015). Indeed, substitution rates can also change through time due to variable selective pressures resulting from changes at other sites (Fitch and Markowitz 1970), and different covarion models have been proposed to deal with these types of situations (Miyamoto and Fitch 1995; Galtier 2001; Penny et al. 2001; Huelsenbeck 2002; Wang et al. 2007).

With an eye towards DNA sequence alignment, in particular, Thorne et al. (1991, 1992) set the foundation for alignment algorithms to have a stronger statistical footing by integrating DNA sequence models of evolution. Integral to applications of models of evolution to sequence alignment, is a model of insertion/deletion or indels. Their approach builds on the more general articulation of sequence alignment by Smith et al. (1981) by integrating a new model of indel evolution within the context of the previously established models of nucleotide substitution. Note also the work by Sankoff and collaborators (Sankoff et al. 1973; Sankoff and Rousseau 1975) which first made the point that sequence alignment and phylogenetic estimation of phylogenies should not be treated as separate inferences.

Finally, to further integrate the biological reality of protein coding sequences, Goldman and Yang (1994) developed a codon-based model to specifically account for codon usage bias. This model, and a similar model proposed by Muse and Gaut (1994), allowed for formal tests of natural selection at the nucleotide level within a ML framework by distinguishing between synonymous (silent) and nonsynonymous (replacement) substitutions. At the same time, taking into account nucleotide position within a codon allows for more biological realism by acknowledging the lack of independence of sites within a codon triplet, the difference in rates of substitution across the three sites within a codon due to the degeneracy of the genetic code, and the generally higher frequency of silent substitutions relative to replacements.

In addition to site non-independence due to their occurrence within a codon, structural constraints may inflict non-independence among sites within the same neighborhood in a DNA sequence. Muse (1995), and Schoniger and von Haeseler (1994) published the first substitution models with structural constraints, a line of work extended by many others (e.g., Thorne et al. 1996; Moshe and Pupko 2019; Glaser et al. 2003; Robinson et al. 2003; Arenas et al. 2013). Also, Jensen and Pedersen developed a dependent-rates model within a maximum-likelihood framework to accommodate sequences with overlapping reading frames (Pedersen and Jensen 2001) and context dependent rates of evolution (Jensen and Pedersen 2000). Bases themselves can be heterogeneous across a DNA sequence and models have been developed to account for this base compositional heterogeneity (Churchill 1989; Galtier and Gouy 1995, 1998). Similarly, the substitution rates of bases themselves can vary and this variation can also be incorporated into an overall model of evolution (Yang 1993, 1994a; Gu et al. 1995; Felsenstein and Churchill 1996).

Hypothesis Testing

One of the great advantages of ML (for example, over parsimony or distance-based approaches) is that it allows for the easy formulation and testing of phylogenetic hypotheses through the use of likelihood ratio tests (LRTs) (Huelsenbeck and Crandall 1997). Such tests have been developed for a variety of aspects surrounding molecular evolution, including tree topology tests, tests of phylogenetic signal, tests of alternative models of evolution, tests for divergence rate heterogeneity, and tests of natural selection. We address each of these categories in turn, noting that all stem from the initial formulation of the ML estimate of phylogeny (Felsenstein 1981) where Felsenstein includes a specific section on 'Hypothesis Testing' detailing thoughts on testing the constancy of the rate of substitution and alternative tree topologies.

While Felsenstein's empirical example of ML used one of the few multispecies alignments available (chickens, iguanas, trout, frogs, and turtles) (Fig. 1), many subsequent extensions, especially in the models of evolution, often focused on the relationships of humans to our nearest relatives, specifically the human, chimp, gorilla, orangutan question (Hasegawa et al. 1985b; Kishino and Hasegawa 1989) with implications for human origins and comparative genomics. The outstanding question at the time was in determining the sister relationship of humans and all three alternative hypotheses (gorilla, chimpanzee, and orangutan) were proposed based on different data and analyses. Felsenstein, through ML, provided a framework to test both underlying assumptions of the likelihood calculations (the model) as well as alternative hypotheses about relationships (the tree).

Testing Tree Topologies

As Felsenstein set up the application of ML to phylogeny, the phylogeny (plus the model of substitution) is the hypothesis. As an hypothesis, alternatives are eminently testable by calculating likelihood scores of competing hypotheses and comparing the difference of the log likelihoods of the null versus the alternative and comparing it to a Chi-square distribution. A common question in systematics is the monophyly of taxonomic groups, that is the members of a clade all share a most recent common ancestor. Monophyly is the principle upon which modern taxonomy rests and it lends itself to statistical evaluation (Rosenberg 2007). Huelsenbeck et al. (1996) described the first explicit test of monophyly within a phylogenetic context. The likelihood ratio test, as originally conceived (Edwards 1972), is a goodness of fit test for a model with a constraint (null) tested against a model without the constraint (alternative). In the phylogenetic case of monophyly, monophyly is the constrained tree which is tested against an alternative using the standard log likelihood ratio test, comparing to a chi-square distribution with p—q degrees of freedom, where p is the number of parameters under the alternative hypothesis and q is the number of parameters under the null hypothesis. Because

alternative phylogenetic hypotheses are not necessarily nested and it is often difficult to determine the appropriate degrees of freedom to conduct such a test, Huelsenbeck et al. (1996) propose a simulation approach to determine significance. Indeed, other LRTs were proposed through the years to answer different questions (Huelsenbeck and Crandall 1997), for example to detect conflicting phylogenetic signal (Huelsenbeck and Bull 1996) or to identify host-parasite cospeciation (Huelsenbeck et al. 1997). An alternative approach to testing tree topologies using the variance in likelihood scores was proposed by Kishino and Hasegawa (KH Test) (1989). This test has been criticized as it was designed to compare two topologies but is often used to test many topologies which leads to overconfidence in the wrong tree (Shimodaira and Hasegawa 1999; Goldman et al. 2000) and adjustments have been proposed to eliminate this bias (Shimodaira 2002).

Testing the Substitution Model

As explained above, the calculation of the tree likelihood implies a specific model of DNA substitution. At the same time, the use of different substitution models can change under different circumstances impacting the tree likelihood and therefore also the optimal tree, including the inferences derived from it (Kelsey et al. 1999; Kelchner and Thomas 2007; Ripplinger and Sullivan 2008; Arbiza et al. 2011; Hoff et al. 2016). After the Felsenstein 1981 paper, it took some time until statistical model adequacy and model selection was proposed (Goldman 1993a, b; Yang et al. 1994; Yang 1994b; Rzhetsky and Nei 1995), but soon after a number of methodological implementations prompted a lot of interest in this area (e.g., Posada and Crandall 2001; Posada 2001; Suchard et al. 2001; Minin et al. 2003; Posada and Buckley 2004; Sullivan et al. 2005; Sullivan and Joyce 2005) which continues to the present (Kalyaanamoorthy et al. 2017; Lefort et al. 2017; Abadi et al. 2019, 2020; Morel et al. 2019; Darriba et al. 2020). These approaches all share at their core a concept of comparing likelihood scores of different models of nucleotide substitution, given a tree topology. Variations occur in how these scores are compared (e.g., LRTs, Bayesian information criteria, Akaike information criteria, etc.).

Testing for Natural Selection

Golding and Felsenstein (1990) took advantage of the LRT to detect the impact of deleterious selection on alternative tree topologies for an explicit test for selection. With the number of advances in models of nucleotide substitution, tests for selection moved from alternative trees to identifying individual sites under selection. Specifically, codon-based models of evolution (Goldman and Yang 1994; Muse and Gaut 1994) (see above) allowed the distinction between synonymous and nonsynonymous substitutions. Using this distinction, approaches to detecting selection test the expectation that purifying selection leads to a higher rate of synonymous substitutions compared to nonsynonymous, neutral evolution should be reflected by equal rates of synonymous and nonsynonymous substitutions, and diversifying selection should lead to more nonsynonymous substitutions. This molecular evolutionary dogma misses much of natural selection as do summary statistics approaches that ignore the evolutionary history (Crandall et al. 1999). Capitalizing on the ML framework and the codon models of evolution, a number of tests have now been proposed for testing for natural selection from DNA sequences across sites and along lineages (Yang 1993; Yang et al. 2000; Yang and Nielsen 2002; Zhang et al. 2005). Likelihood ratio tests have also been proposed to test for heterogeneity among regions within a nucleotide sequence (Gaut and Weir 1994).

Estimating Divergence Times

Zuckerkandl and Pauling (1962) made the observation that the number of amino acid changes across proteins, haemoglobins in particular, seemed to occur at a constant rate across the evolution of a group. They proposed that the number of amino acid replacements correlated with divergence times based on fossil calibrations, leading to the concept of a molecular clock (Bromham and Penny 2003). In Felsenstein's JME paper, he takes advantage of the ML calculations to propose a test for the molecular clock (rate constancy). The test requires the calculation of a likelihood score for a tree with the constraint of all the tips being contemporaneous and is compared to the alternative without this constraint. These nested hypotheses are neatly compared using the maximum-likelihood ratio test with n-2 degrees of freedom where n is the number of tips in the tree as that is the difference in free parameters between the constrained and unconstrained tree. This brings up an important point relative to the topology tests described above in that the likelihood calculation involves both branch length and topology (together, the tree). Thus, alternatives can be significantly different just in branch length with the same topology, as in the case of the molecular clock test.

Following Felsenstein's articulation of the constancy of rates test using LRTs, a number of tests were subsequently developed based on this foundational framework for differences in rates across lineages (Muse and Weir 1992; Thorne et al. 1998). Such tests are effectively used for identifying the impacts of natural selection on particular lineages and particular genes (e.g., Gaut et al. 1992). Another application of the molecular clock is to estimate divergence times of specific lineages/clades of genes or organisms. Such tests explicitly assume a molecular clock. However, after tests of the molecular clock were implemented, it became apparent

that a molecular clock was often rejected for various data sets. As a consequence, researchers developed approaches for relaxing the molecular clock assumption when calculating divergence times to provide better estimates (Huelsenbeck et al. 2000; Kishino et al. 2001) through explicit models of rate evolution (Aris-Brosou and Yang 2002).

Inferring Ancestral Sequences

Felsenstein's algorithm involves the computation of the partial likelihoods for the different nucleotides at the internal nodes of the tree, and these in turn can immediately be used to estimate the ML ancestral DNA (or protein) sequences (Yang et al. 1995; Koshi and Goldstein 1996), and in general of any discrete state (Pagel 1999). Such approaches have been especially important in testing hypotheses of protein evolution, structure, and function (Harms and Thornton 2010; Gumulya and Gillam 2017).

Bayesian Inference

In Bayesian inference, the posterior probability of a given hypothesis is computed according to Bayes' theorem, in which the prior probability of a given hypothesis is updated with the likelihood of the data given that hypothesis. Not surprisingly, Felsenstein had already briefly discussed in his Ph.D. thesis (1968) how Bayesian inference could be used in phylogenetics. In the late 90′s, the ability to calculate the likelihood from trees and modern statistical techniques like Markov Chain Monte Carlo sampling made possible the Bayesian inference of phylogenies (Rannala and Yang 1996; Mau and Newton 1997; Yang and Rannala 1997; Larget and Simon 1999).

There have been discussions about the relative merits of the likelihood and Bayesian approaches (e.g., Svennblad et al. 2006), particularly regarding confidence measures (bootstrap values vs. posterior probabilities) (Douady et al. 2003). In our opinion, both are close allies that provide a strong statistical, model-based framework for phylogenetic inference that has proven generally superior to non-model strategies.

Software Implementations of the Phylogenetic Likelihood Function

As mentioned above, Felsenstein also introduced in his 1981 paper a foundational program (*dnaml*) for obtaining ML estimates of the branch lengths, included in his phylogenetic package PHYLIP (https://evolution.genetics.washington.edu/phylip.html). Since then, multiple computer programs were written to estimate ML trees and evolutionary parameters from them, several of which became extremely popular, like PAUP* (Swofford 1993), fastDNAmL (Olsen et al. 1994), PAML (Yang 1997), RAxML (Stamatakis et al. 2002, 2005), PhyML (Guindon and Gascuel 2003), or IQtree (Nguyen et al. 2015), among others. These programs, starting with *dnaml*, have been central for the growth and development of the field of molecular phylogenetics. As a side note, Felsenstein maintained (until 2013) an incredible community resource for phylogenetic software that is still available today: https://evolution.genetics.washington.edu/phylip/software.html. While not totally up to date, this is still a great one-stop shopping for a wide diversity of software applications in a variety of aspects of phylogenetics.

Moreover, prompted by the need to analyze increasingly large data sets, more efficient algorithms (Kosakovsky Pond and Muse 2004; Stamatakis and Ott 2008; Sumner and Charleston 2010; Kobert et al. 2017; Ji et al. 2020) and different High-Performance Computing (HPC) solutions were also proposed for faster phylogenetic likelihood calculations, taking advantage of multi-core computers and cluster environments. Thus, phylogenetic likelihood libraries like BEAGLE (Ayres et al. 2019) or PLL (Flouri et al. 2015) have been developed, and different devices and architectures have been explored including Graphical Processing Units (GPUs) (Suchard and Rambaut 2009), Field Programmable Gate Array (FPGAs) (Mak and Lam 2004a, b; Alachiotis et al. 2009; Zierke and Bakos 2010) and other many-core accelerators (Kozlov et al. 2014).

The Future of Statistical Phylogenetics

As discussed in the previous section, Felsenstein's paper inspired a plethora of applications that have defined the field of phylogenetics. Several of the assumptions he made were relaxed through the years, but others (e.g., independence among sites) have endured, suggesting perhaps that there is no clear benefit in redefining them. While multiple models have been developed for the estimation of parameters of biological interests upon trees, the models and algorithms behind the calculation of the tree likelihood itself have not evolved that much (but see Ji et al. 2020 for recently proposed advances).

In the last decade, high-throughput sequencing techniques have changed many fields of biology, including phylogenetics, facilitating the accumulation of massive datasets with thousands of taxa and hundreds of thousands of sites. Many phylogenomic analyses now leverage the 'Felsenstein phylogenetic likelihood' as part of more or less complex models dealing with multiple phylogenetic layers beyond gene trees (i.e., locus trees and species trees). Perhaps, the most important challenge in phylogenomics is how to deal with these enormous data sets while taking advantage of the powerful statistical framework initiated by Edwards, Cavalli-Sforza and Felsenstein and extended in different directions

by many others. As just described, the computation of the phylogenetic likelihood function is now much faster due to the implementation of modern computer science and statistical techniques, indeed favored by the availability of much more powerful computers. Clearly, future advances in computation will facilitate the calculation of the tree likelihood with more data in less time.

Paradigms like probabilistic programming (Fourment and Darling 2019; Ronquist et al. 2020) and statistical procedures like Variational Bayesian phylogenetic inference (Dang and Kishino 2019) promise to help the development of biologically-realistic phylogenetic models that can be efficiently computed. Hopefully, in the future we will see in JME some of the most exciting developments in statistical phylogenetics, honoring the ground-breaking, game-changing article that Joe Felsenstein published in 1981 in this journal.

References

Abadi S, Azouri D, Pupko T, Mayrose I (2019) Model selection may not be a mandatory step for phylogeny reconstruction. Nat Commun 10:934

Abadi S, Avram O, Rosset S et al (2020) ModelTeller: model selection for optimal phylogenetic reconstruction using machine learning. Mol Biol Evol. https://doi.org/10.1093/molbev/msaa154

Alachiotis N, Sotiriades E, Dollas A, Stamatakis A (2009) Exploring FPGAs for accelerating the phylogenetic likelihood function. In: 2009 IEEE International Symposium on Parallel Distributed Processing. pp. 1–8

Arbiza L, Patricio M, Dopazo H, Posada D (2011) Genome-wide heterogeneity of nucleotide substitution model fit. Genome Biol Evol 3:896–908

Arenas M, Dos Santos HG, Posada D, Bastolla U (2013) Protein evolution along phylogenetic histories under structurally constrained substitution models. Bioinformatics 29:3020–3028

Aris-Brosou S, Yang Z (2002) Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. Syst Biol 51:703–714

Ayres DL, Cummings MP, Baele G et al (2019) BEAGLE 3: improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. Syst Biol 68:1052–1061

Barry D, Hartigan JA (1987) Statistical analysis of hominoid molecular evolution. Stat Sci 2:191–207

Boni MF, Lemey P, Jiang X et al (2020) Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol 5:1408–1417

Bromham L, Penny D (2003) The modern molecular clock. Nat Rev Genet 4:216–224

Churchill GA (1989) Stochastic models for heterogeneous DNA sequences. Bull Math Biol 51:79–94

Crandall KA, Kelsey CR, Imamichi H et al (1999) Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. Mol Biol Evol 16:372–382

Dang T, Kishino H (2019) Stochastic variational inference for bayesian phylogenetics: a case of CAT model. Mol Biol Evol 36:825–833

Darriba D, Posada D, Kozlov AM et al (2020) ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol 37:291–294

Douady CJ, Delsuc F, Boucher Y et al (2003) Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol Biol Evol 20:248–254

Eck RV, Dayhoff MO (1966) Atlas of protein sequence and structure, V. 3–5. National Biomedical Research Foundation

Edwards AWF (1972) Likelihood. Cambridge University Press, Cambridge, England

Edwards AWF (2009) Statistical methods for evolutionary trees. Genetics 183:5–12

Edwards AWF, Cavalli-Sforza LL (1963a) A method for cluster analysis. In: Preprints of the 5th International Biometrics Conference

Edwards AWF, Cavalli-Sforza LL (1963b) The reconstruction of evolution. Ann Hum Genet 27:104–105

Edwards AWF, Cavalli-Sforza LL (1964) Reconstruction of evolutionary trees. In: Heywood WH, McNeill J (eds) Phenetic and phylogenetic classification. Systematics Association Publication, London, pp 67–76

Edwards AWF, Cavalli-Sforza LL (1965) A method for cluster analysis. Biometrics 21:362–375

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

Erdmann VA (1982) Collection of published 5S and 5.8S RNA sequences and their precursors. Nucleic Acids Res 10:r93-115

Farris J (1983) The logical basis of phylogenetic analysis. In: Platnick NIFVA (ed) Advances in cladistics II. Columbia University Press, New York, pp 7–36

Fauver JR, Petrone ME, Hodcroft EB et al (2020) Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. Cell 181:990-996.e5

Felsenstein J (1968) Statistical inference and the estimation of phylogenies. University of Chicago, Chicago

Felsenstein J (1973) Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. Syst Biol 22:240–249

Felsenstein J (1973) Maximum-likelihood estimation of evolutionary trees from continuous characters. Am J Hum Genet 25:471–492

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401

Felsenstein J (1979) Alternative methods of phylogenetic inference and their interrelationship. Syst Biol 28:49

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Felsenstein J, Churchill GA (1996) A hidden markov model approach to variation among sites in rate of evolution. Mol Biol Evol 13:93–104

Felsenstein J, Sober E (1986) Parsimony and likelihood: an exchange. Syst Zool 35:617

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Fitch WM, Markowitz E (1970) An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem Genet 4:579–593

Flouri T, Izquierdo-Carrasco F, Darriba D et al (2015) The phylogenetic likelihood library. Syst Biol 64:356–362

Fourment M, Darling AE (2019) Evaluating probabilistic programming and fast variational Bayesian inference in phylogenetics. PeerJ 7:e8272

Fukami-Kobayashi K, Tateno Y (1991) Robustness of maximum likelihood tree estimation against different patterns of base substitutions. J Mol Evol 32:79–91

Galtier N (2001) Maximum-likelihood phylogenetic analysis under a covarion-like model. Mol Biol Evol 18:866–873

Galtier N, Gouy M (1995) Inferring phylogenies from DNA sequences of unequal base compositions. Proc Natl Acad Sci USA 92:11317–11321

Galtier N, Gouy M (1998) Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol Biol Evol 15:871–879

Gaut BS, Lewis PO (1995) Success of maximum likelihood phylogeny inference in the four-taxon case. Mol Biol Evol 12:152–162

Gaut BS, Weir BS (1994) Detecting substitution-rate heterogeneity among regions of a nucleotide sequence. Mol Biol Evol 11:620–629

Gaut BS, Muse SV, Clark WD, Clegg MT (1992) Relative rates of nucleotide substitution at the rbcL locus of monocotyledonous plants. J Mol Evol 35:292–303

Glaser F, Pupko T, Paz I et al (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 19:163–164

Golding B, Felsenstein J (1990) A maximum likelihood approach to the detection of selection from a phylogeny. J Mol Evol 31:511–523

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 Biol 39:345–361

Goldman N (1993a) Statistical tests of models of DNA substitution. J Mol Evol 36:182–198

Goldman N (1993b) Simple diagnostic statistical tests of models for DNA substitution. J Mol Evol 37:182–198

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. Syst Biol 49:652–670

Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ et al (2020) Introductions and early spread of SARS-CoV-2 in the New York City area. Science 369:297–301

Gu X, Fu YX, Li WH (1995) Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol Biol Evol 12:546–557

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Gumulya Y, Gillam EMJ (2017) Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the "retro"approach to protein engineering. Biochem J 474:1–19

Harms MJ, Thornton JW (2010) Analyzing protein structure and function using ancestral gene reconstruction. Curr Opin Struct Biol 20:360–366

Hasegawa M, Yano T-A (1984) Phylogeny and classification of hominoidea as inferred from DNA sequence data. Proc Jpn Acad Ser B 60:389–392

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174

Hasegawa M, Iida Y, Yano T et al (1985) Phylogenetic relationships among eukaryotic kingdoms inferred from ribosomal RNA sequences. J Mol Evol 22:32–38

Hasegawa M, Kishino H, Saitou N (1991) On the maximum likelihood method in molecular phylogenetics. J Mol Evol 32:443–445

Heuch I, Li FHF (1972) PEDIG-a computer program for calculation of genotype probabilities using phenotype information. Clin Genet 3:501–504

Hilden J (1970) GEN EX-an algebraic approach to pedigree probability calculus. Clin Genet 1:319–348

Hoff M, Orf S, Riehm B et al (2016) Does the choice of nucleotide substitution models matter topologically? BMC Bioinform 17:143

Huelsenbeck JP (1995a) Performance of phylogenetic methods in simulation. Syst Biol 44:17–48

Huelsenbeck JP (1995b) The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. Mol Biol Evol 12:843–849

Huelsenbeck JP (1998) Systematic bias in phylogenetic analysis: is the strepsiptera problem solved? Syst Biol 47:519–537

Huelsenbeck JP (2002) Testing a covariotide model of DNA substitution. Mol Biol Evol 19:698–707

Huelsenbeck JP, Bull JJ (1996) A likelihood ratio test to detect conflicting phylogenetic signal. Syst Biol 45:92–98

Huelsenbeck JP, Crandall KA (1997) Phylogeny estimation and hypothesis testing using maximum likelihood. Annu Rev Ecol Syst 28:437–466

Huelsenbeck JP, Hillis DM (1993) Success of phylogenetic methods in the four-taxon case. Syst Biol 42:247–264

Huelsenbeck JP, Hillis DM, Nielsen R (1996) A likelihood-ratio test of monophyly. Syst Biol 45:546–558

Huelsenbeck JP, Rannala B, Yang Z (1997) Statistical tests of host-parasite cospeciation. Evolution 51:410–419

Huelsenbeck JP, Larget B, Swofford D (2000) A compound poisson process for relaxing the molecular clock. Genetics 154:1879–1892

Jayaswal V, Robinson J, Jermiin L (2007) Estimation of phylogeny and invariant sites under the general Markov model of nucleotide sequence evolution. Syst Biol 56:155–162

Jensen JL, Pedersen A-MK (2000) Probabilistic models of DNA sequence evolution with context dependent rates of substitution. Adv Appl Probab 32:499–517

Jermiin LS, Crozier RH (1994) The cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger*: sequence divergence in Hymenoptera may be associated with nucleotide content. J Mol Evol 38:282–294

Ji X, Zhang Z, Holbrook A et al (2020) Gradients do grow on trees: a linear-time 5 (N)-dimensional gradient for statistical phylogenetics. Mol Biol Evol. https://doi.org/10.1093/molbev/msaa130

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Kalyaanamoorthy S, Minh BQ, Wong TKF et al (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589

Kaplan N, Langley CH (1979) A new estimate of sequence divergence of mitochondrial DNA using restriction endonuclease mappings. J Mol Evol 13:295–304

Kashyap RL, Subas S (1974) Statistical estimation of parameters in a phylogenetic tree using a dynamic model of the substitutional process. J Theor Biol 47:75–101

Kelchner SA, Thomas MA (2007) Model use in phylogenetics: nine key questions. Trends Ecol Evol 22:87–94

Kelsey CR, Crandall KA, Voevodin AF (1999) Different models, different trees: the geographic origin of PTLV-I. Mol Phylogenet Evol 13:336–347

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol 29:170–179

Kishino H, Thorne JL, Bruno WJ (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol Biol Evol 18:352–361

Kluge AG, Farris JS (1969) Quantitative phyletics and the evolution of Anurans. Syst Biol 18:1–32

Kobert K, Stamatakis A, Flouri T (2017) Efficient detection of repeating sites to accelerate phylogenetic likelihood calculations. Syst Biol 66:205–217

Kosakovsky Pond SL, Muse SV (2004) Column sorting: rapid calculation of the phylogenetic likelihood function. Syst Biol 53:685–692

Koshi JM, Goldstein RA (1996) Probabilistic reconstruction of ancestral protein sequences. J Mol Evol 42:313–320

Kozlov AM, Goll C, Stamatakis A (2014) Efficient Computation of the phylogenetic likelihood function on the intel MIC architecture. In: 2014 IEEE International Parallel Distributed Processing Symposium Workshops. pp. 518–527

Kozlov AM, Darriba D, Flouri T et al (2019) RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455

Kuhner MK, Felsenstein J (1994) A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol Biol Evol 11:459–468

Lam TT-Y, Jia N, Zhang Y-W et al (2020) Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature 583:282–285

Lanave C, Preparata G, Saccone C, Serio G (1984) A new method for calculating evolutionary substitution rates. J Mol Evol 20:86–93

Larget B, Simon DL (1999) Markov chasin monte carlo algorithms for the bayesian analysis of phylogenetic trees. Mol Biol Evol 16:750–759

Lefort V, Longueville J-E, Gascuel O (2017) SMS: smart model selection in PhyML. Mol Biol Evol 34:2422–2424

Mak TST, Lam KP (2004a) On computing maximum likelihood phylogeny using FPGA. Field programmable logic and application. Springer, Berlin, Heidelberg, p. 1188

Mak TST, Lam KP (2004b) Embedded computation of maximum-likelihood phylogeny inference using platform FPGA. In: Proceedings. 2004 IEEE Computational Systems Bioinformatics Conference, 2004. CSB 2004. pp. 512–514

Mau B, Newton MA (1997) Phylogenetic inference for binary data on dendograms using markov chain Monte Carlo. J Comput Graph Stat 6:122

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci USA 74:560–564

Minh BQ, Schmidt HA, Chernomor O et al (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534

Minin V, Abdo Z, Joyce P, Sullivan J (2003) Performance-based selection of likelihood models for phylogeny estimation. Syst Biol 52:674–683

Miyamoto MM, Fitch WM (1995) Testing the covarion hypothesis of molecular evolution. Mol Biol Evol 12:503–513

Morel B, Kozlov AM, Stamatakis A (2019) ParGenes: a tool for massively parallel model selection and phylogenetic tree inference on thousands of genes. Bioinformatics 35:1771–1773

Morgan GJ (1998) Emile Zuckerkandl, Linus Pauling, and the molecular evolutionary clock, 1959–1965. J Hist Biol 31:155–178

Moshe A, Pupko T (2019) Ancestral sequence reconstruction: accounting for structural information by averaging over replacement matrices. Bioinformatics 35:2562–2568

Muse SV (1995) Evolutionary analyses of DNA sequences subject to constraints of secondary structure. Genetics 139:1429–1439

Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol Biol Evol 11:715–724

Muse SV, Weir BS (1992) Testing for equality of evolutionary rates. Genetics 132:269–276

Neyman J (1971) Molecular studies of evolution: a source of novel statistical problems. In: Gupta SS, Yackel J (eds) Statistical decision theory and related topics. Academic Press, New York, pp. 1–27

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274

Olsen GJ, Matsuda H, Hagstrom R, Overbeek R (1994) fastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput Appl Biosci 10:41–48

Pagel M (1999) The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Syst Biol 48:612–622

Pedersen AM, Jensen JL (2001) A dependent-rates model and an MCMC-based methodology for the maximum-likelihood analysis of sequences with overlapping reading frames. Mol Biol Evol 18:763–776

Penny D, McComish BJ, Charleston MA, Hendy MD (2001) Mathematical elegance with biochemical realism: the covarion model of molecular evolution. J Mol Evol 53:711–723

Posada D (2001) The effect of branch length variation on the selection of models of molecular evolution. J Mol Evol 52:434–444

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol 53:793–808

Posada D, Crandall KA (2001) Selecting the best-fit model of nucleotide substitution. Syst Biol 50:580–601

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

Ripplinger J, Sullivan J (2008) Does choice in model selection affect maximum likelihood analysis? Syst Biol 57:76–85

Robinson DM, Jones DT, Kishino H et al (2003) Protein evolution with dependence among codons due to tertiary structure. Mol Biol Evol 20:1692–1704

Ronquist F, Kudlicka J, Senderov V et al (2020) Universal probabilistic programming offers a powerful approach to statistical phylogenetics. bioRxiv. https://doi.org/10.1101/2020.06.16.154443

Rosenberg NA (2007) Statistical tests for taxonomic distinctiveness from observations of monophyly. Evolution 61:317–323

Rzhetsky A, Nei M (1995) Tests of applicability of several substitution models for DNA sequence data. Mol Biol Evol 12:131–151

Saitou N (1988) Property and efficiency of the maximum likelihood method for molecular phylogeny. J Mol Evol 27:261–273

Saitou N, Imanishi T (1989) Relative Efficiencies of the fitch-margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. Mol Biol Evol 6:514–514

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sankoff D, Rousseau P (1975) Locating the vertices of a steiner tree in an arbitrary metric space. Math Program 9:240–246

Sankoff D, Morel C, Cedergren RJ (1973) Evolution of 5S RNA and the non-randomness of base replacement. Nat New Biol 245:232–234

Schöniger M, von Haeseler A (1994) A stochastic model for the evolution of autocorrelated DNA sequences. Mol Phylogenet Evol 3:240–247

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. Syst Biol 51:492–508

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114–1114

Smith TF, Waterman MS, Fitch WM (1981) Comparative biosequence metrics. J Mol Evol 18:38–46

Sober E (1991) Reconstructing the past: parsimony, evolution, and inference. MIT Press, Cambridge

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313

Stamatakis A (2015) Using RAxML to Infer phylogenies. Curr Protoc Bioinform 51:6.14.1-6.14.14

Stamatakis A, Ott M (2008) Efficient computation of the phylogenetic likelihood function on multi-gene alignments and multi-core architectures. Philos Trans R Soc B 363:3977–3984

Stamatakis AP, Ludwig T, Meier H, Wolf MJ (2002) AxML: a fast program for sequential and parallel phylogenetic tree calculations based on the maximum likelihood method. Proc IEEE Comput Soc Bioinform Conf 1:21–28

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

Suchard MA, Rambaut A (2009) Many-core algorithms for statistical phylogenetics. Bioinformatics 25:1370–1376

Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time markov chain evolutionary models. Mol Biol Evol 18:1001–1013

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

Sullivan J, Abdo Z, Joyce P, Swofford DL (2005) Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation. Mol Biol Evol 22:1386–1392

Sumner JG, Charleston MA (2010) Phylogenetic estimation with partial likelihood tensors. J Theor Biol 262:413–424

Sumner JG, Fernández-Sánchez J, Jarvis PD (2012) Lie markov models. J Theor Biol 298:16–31

Svennblad B, Erixon P, Oxelman B, Britton T (2006) Fundamental differences between the methods of maximum likelihood and maximum posterior probability in phylogenetics. Syst Biol 55:116–121

Swofford DL (1993) PAUP*. Phylogenetic analysis using parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

Tateno Y, Takezaki N, Nei M (1994) Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. Mol Biol Evol 11:261–277

Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. Lect Math Life Sci 17:57–86

Thorne JL, Kishino H, Felsenstein J (1991) An evolutionary model for maximum likelihood alignment of DNA sequences. J Mol Evol 33:114–124

Thorne JL, Kishino H, Felsenstein J (1992) Inching toward reality: an improved likelihood model of sequence evolution. J Mol Evol 34:3–16

Thorne JL, Goldman N, Jones DT (1996) Combining protein evolution and secondary structure. Mol Biol Evol 13:666–673

Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol 15:1647–1657

Waddell PJ, Steel MA (1997) General time-reversible distances with unequal rates across sites: mixing gamma and inverse Gaussian distributions with invariant sites. Mol Phylogenet Evol 8:398–414

Wang H-C, Spencer M, Susko E, Roger AJ (2007) Testing for covarion-like evolution in protein sequences. Mol Biol Evol 24:294–305

Whelan S, Liò P, Goldman N (2001) Molecular phylogenetics: state-of-the-art methods for looking into the past. Trends Genet 17:262–272

Whiting MF (1998) Long-branch distraction and the strepsiptera. Syst Biol 47:134–137

Woodhams MD, Fernández-Sánchez J, Sumner JG (2015) A new hierarchy of phylogenetic models consistent with heterogeneous substitution rates. Syst Biol 64:638–650

Worobey M, Pekar J, Larsen BB et al (2020) The emergence of SARS-CoV-2 in Europe and North America. Science 370:564–570

Yang Z (1993) Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol Biol Evol 10:1396–1401

Yang Z (1994) Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods. Syst Biol 43:329–342

Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 39:306–314

Yang Z (1994) Estimating the pattern of nucleotide substitution. J Mol Evol 39:105–111

Yang Z (1996) Phylogenetic analysis using parsimony and likelihood methods. J Mol Evol 42:294–307

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556

Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–917

Yang Z, Rannala B (1997) Bayesian phylogenetic inference using DNA sequences: a markov chain monte carlo method. Mol Biol Evol 14:717–724

Yang Z, Goldman N, Friday A (1994) Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. Mol Biol Evol 11:316–324

Yang Z, Kumar S, Nei M (1995) A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141:1641–1650

Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 22:2472–2479

Zierke S, Bakos JD (2010) FPGA acceleration of the phylogenetic likelihood function for Bayesian MCMC inference methods. BMC Bioinformatics 11:184

Zuckerkandl E, Pauling L (1962) Molecular disease, evolution, and genetic heterogeneity. In: Pullman B, Kasha M (eds) Horizons in biochemistry. Academic Press, New York, pp 189–225

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) Evolving genes and proteins. Academic Press, New York, pp 97–166

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.