COMMENTARY

Felsenstein Phylogenetic Likelihood

David Posada1,2,3 · Keith A. Crandall4,5

Received: 28 September 2020 / Accepted: 7 December 2020 / Published online: 13 January 2021 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

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 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/phylib.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

---

**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 covariation 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, hemoglobins 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 contemporary 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.

Acknowledgements First and foremost, we want to acknowledge Joe Felsenstein for his paramount contributions to the field of statistical phylogenetics. Secondly, we want to thank JME and David Liberles for the opportunity to write this perspective. DP is supported by the European Research Council (Grant ERC-617457-PHYLOCANCER) and by the Spanish Ministry of Economy and Competitiveness—MINECO (Grant BFU2015-63774-P awarded to D.P.). D.P. receives further support from Xunta de Galicia. KAC is supported by the National Institutes of Health Grant Number UL1TR000075 and the National Science Foundation Grant Number DEB-2028280. We thank two anonymous reviewers, the Crandall Lab Group, Jeff Thorne, and, especially, Joe Felsenstein for helpful comments to improve our manuscript.

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, Dapero P, 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) Phyentic 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 (1972) 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
