UC Irvine

UC Irvine Previously Published Works

Title

Forty Years of Inferential Methods in the Journals of the Society for Molecular Biology and Evolution.

Permalink

https://escholarship.org/uc/item/9tp7k7g3

Journal

Molecular Biology and Evolution (MBE), 41(1)

Authors

Russo, Claudia

Eyre-Walker, Adam

Katz, Laura

et al.

Publication Date

2024-01-03

DOI

10.1093/molbev/msad264

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Forty Years of Inferential Methods in the Journals of the Society for Molecular Biology and Evolution

Claudia A.M. Russo [1,*] Adam Eyre-Walker [2] Laura A. Katz [3] and Brandon S. Gaut [4]

[1]Departamento de Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
[2]School of Life Sciences, University of Sussex, Brighton, UK
[3]Department of Biological Sciences, Smith College, Northampton, MA, USA
[4]School of Biological Sciences, University of California, Irvine, CA, USA

***Corresponding author:** E-mail: claudiaamrusso@gmail.com.

Associate editor: Heather Rowe

Abstract

We are launching a series to celebrate the 40th anniversary of the first issue of *Molecular Biology and Evolution*. In 2024, we will publish virtual issues containing selected papers published in the Society for Molecular Biology and Evolution journals, *Molecular Biology and Evolution* and *Genome Biology and Evolution*. Each virtual issue will be accompanied by a perspective that highlights the historic and contemporary contributions of our journals to a specific topic in molecular evolution. This perspective, the first in the series, presents an account of the broad array of methods that have been published in the Society for Molecular Biology and Evolution journals, including methods to infer phylogenies, to test hypotheses in a phylogenetic framework, and to infer population genetic processes. We also mention many of the software implementations that make methods tractable for empiricists. In short, the Society for Molecular Biology and Evolution community has much to celebrate after four decades of publishing high-quality science including numerous important inferential methods.

Key words: **Society for Molecular Biology and Evolution, anniversary celebration, inferential methods, software packages, SMBE journals.**

Forty Years of the Society for Molecular Biology and Evolution Journals

Molecular Biology and Evolution (*MBE*) has reached its fortieth anniversary. The journal, and ultimately the Society for Molecular Biology and Evolution (SMBE), traces its origin to a June 1982 meeting held at the State University of New York at Stony Brook. Following a symposium on the "Evolution of Genes and Proteins," the participants gathered to discuss an idea for a new journal. They saw a need for a journal that bridged the gap between molecular and evolutionary biology, provided a high-quality forum for publications at an accessible cost, and was governed by the scientific community. The first issue of *MBE* was published only 18 months later, in December 1983.

The success of *MBE* contributed to the formation of the SMBE in 1992. SMBE has grown as a society and supported the founding of a second journal, *Genome Biology and Evolution* (*GBE*), in 2009 (Fig. 1). Originally designed to focus on emerging genome-scale data, *GBE* was among the first society-owned open-access journals. Recently, *MBE*, following *GBE*'s lead, has also become open access. In keeping with the goals of the founding group, the cost of publishing in the SMBE journals remains modest, and both journals have an inclusive waiver policy with a commitment to publish high-quality science regardless of an author's ability to pay publication costs. The proceeds from SMBE publishing support other SMBE activities, including the annual meeting, IDEA initiatives, satellite meetings, and various awards and fellowships.

To mark our 40-year anniversary, the SMBE journals are launching a collaborative year-long celebration. Throughout 2024, *MBE* and *GBE* will publish perspectives centered on topics of both contemporary and historical significance within our field. Each perspective will be complemented by virtual issues housing selected publications from both journals that highlight work relevant to the monthly topic. In addition to their availability on journal websites, the virtual issues can be accessed from the newly launched website that represents the SMBE family of journals (www.academic.oup.com/smbejournals). We hope that the perspectives and virtual issues will encourage our community to celebrate the lasting impacts of journals built by, and for, the community.

The monthly topics in the anniversary series will be as varied as the papers in the journals, including topics like testing for selection, human diversification, the mechanisms and consequences of recombination, sex chromosomes, and microbial diversity. For this first installment, we have created virtual issues and this perspective to

Perspective

Received: November 08, 2023. **Accepted:** November 27, 2023
© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

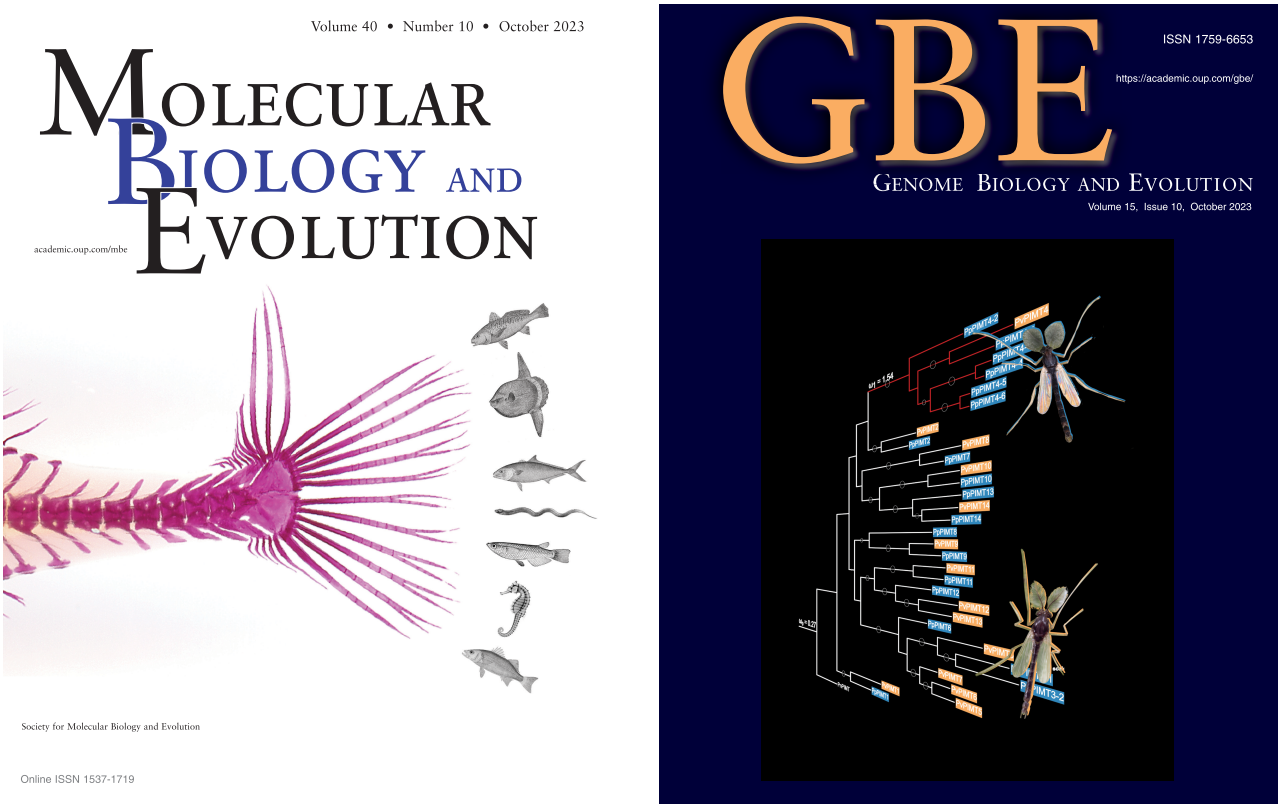

Fig. 1. Recent covers from the two SMBE journals, MBE (left) and GBE (right).

highlight papers on inferential methods in evolutionary biology. Over the last four decades, our journals have played a central role in developing methods for the study of molecular evolution, resulting in a rich literature of highly cited studies. In fact, the literature is so vast that we cannot hope to mention all the impactful papers here. Instead, we focus on a few key research areas (e.g. phylogenetic inference, hypothesis testing in a phylogenetic framework, and population genetic analysis). We hope that this perspective highlights the variety of methodological improvements that have been published in the SMBE journals and illustrates the success of the founders' goal to bridge gaps between molecular and evolutionary biology (Fig. 2).

An Overview of Methods Published in SMBE Journals

Phylogenetic Inference

The SMBE journals have played a pivotal role in the development and application of molecular phylogenetic methods. One example is the neighbor-joining algorithm (Saitou and Nei 1987), which remains the most cited single publication in the SMBE journals with more than 50,000 citations. Neighbor-joining has had a lasting influence because it is easily implemented, accurate if distances are unbiased (Rzhetsky and Nei 1993), and lightning fast, even for very large numbers of sequences (Huelsenbeck 1995).

Other tree-building methods presented in our journals include BIONJ (Gascuel 1997), FastTree (Price et al. 2009), minimum-evolution (Rzhetsky and Nei 1992), weighted neighbor-joining (Bruno et al. 2000), generalized neighbor-joining (Pearson et al. 1999), NJML (Ota and Li 2000), and quartet puzzling (Strimmer and von Haeseler 1996). An alternative to these bifurcating tree-building methods is family-joining, which is for taxa that have been sampled over evolutionary time when a strictly bifurcating tree may not be appropriate (Kalaghatgi et al. 2016). Family-joining is a distance method in which operational taxonomic units (OTUs), such as viruses, can be placed not only at external branches but also at internal vertices. Another set of papers have focused on network construction, which can be more relevant for building intraspecific trees with low genetic distances among OTUs. Network approaches published in our journals include a combinatorial method (Huson and Scornavacca 2011), median-joining (Bandelt et al. 1999), neighbor-net (Bryant and Moulton 2004), and the Bayesian inference for species' network (Zhang et al. 2018).

SMBE journals have played a leading role in adopting Bayesian approaches for constructing phylogenies. For example, the work of Yang and Rannala (1997) was among the first to implement the Markov chain Monte Carlo (MCMC) algorithms for Bayesian phylogenetics. This was a significant step forward and merits recognition, because MCMC has become the go-to approach to approximate the posterior distribution of evolutionary parameters.

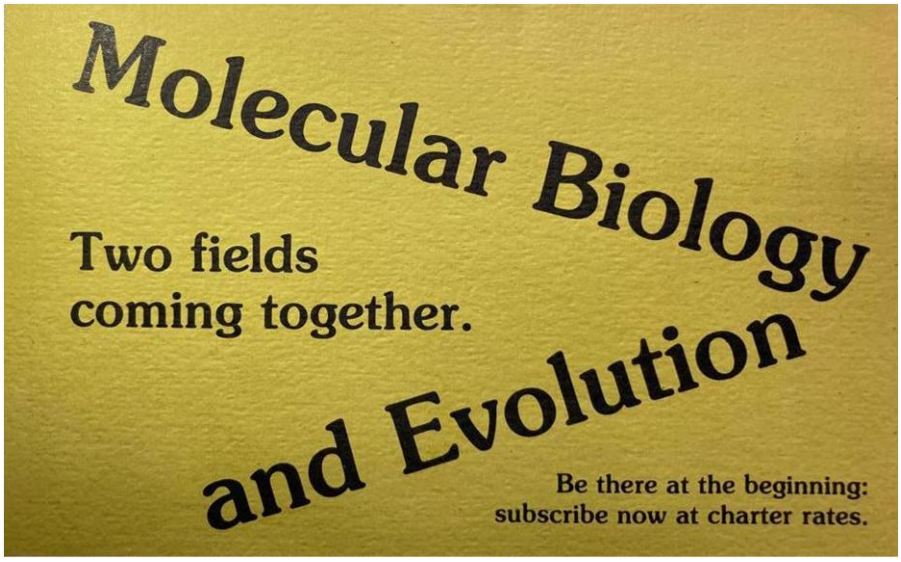

Fig. 2. The first issue of *MBE* included a flier that advertised the intention to merge the fields of molecular biology with evolutionary biology.

Many tree-building methods use a concatenated (multigene) alignment to build a single phylogenetic tree. For the same set of OTUs, however, individual genes may yield different topologies due to distinct evolutionary histories such as gene duplications/loss, lateral gene transfers, or the presence of genetic polymorphism in the ancestral species. Pamilo and Nei (1988) provided the theoretical foundation to phylogenetic reconstructions based on multispecies coalescence models. Their method provides analytical solutions for the probability of topological matches between gene and species trees. In this approach, a phylogenetic tree is first built for each available marker, and then individual gene trees are summarized into a single species tree, analyzing conflicts between gene trees. Similar methods have been published across our journals, including extensions of coalescent approaches (Ané et al. 2007; Heled and Drummond 2010) and methods suitable to large data sets (Ullah et al. 2015).

A newer method for testing incongruences between gene trees is PhylteR, which is useful for phylogenomic data sets. In this case, distance matrices built from each marker are compared in order to detect (and possibly remove) outlier sequences that are the OTUs that do not follow the general pattern (Compte et al. 2023). In 2012, Bryant et al. introduced a new method, for biallelic and nonlinked markers, to build a species tree considering all possible gene trees (Bryant et al. 2012). In this method, rather than integrating over all possible genealogies to determine the probability of data given the species tree, the authors developed a pruning algorithm for analytical calculation of this value, bypassing the construction of individual gene trees.

It might be useful to delimit species using molecular markers, particularly if the taxonomic group presents few conspicuous morphological characteristics. Using the multispecies coalescent model in a Bayesian framework, Yang and Rannala (2014) developed a method to infer species delimitation and to estimate the species phylogeny at the same time. This new method differed from previous algorithms that required a guiding phylogeny.

Studies published in our journals have also advanced our ability to evaluate the robustness of phylogenetic tree clades for large data sets, where algorithm speed is critical. These methods include UFBoot, an ultrafast version of the bootstrap test of branch support (Minh et al. 2013; Hoang et al. 2018) and a coalescent-based method that quickly evaluates branch support from quartet frequencies (Sayyari and Mirarab 2016). Earlier studies have provided important insights about reliability estimates; for example, Hedges (1992) estimated the number of bootstrap replications needed for a precise estimate of the bootstrap support value.

Phylogenies cannot be produced without quality alignments, and the SMBE journals have contributed to this important step, publishing versions of the popular MAFFT (Katoh and Standley 2013) and GBlocks (Castresana 2000) software. MAFFT allows rapid generation of multisequence alignments under a variety of parameterizations, while GBlocks evaluates alignment quality, identifies the most conservative (and presumably trustworthy) blocks of alignment, and allows the user to discard more variable sections with poorer alignments.

Most methods of phylogenetic inference require nucleotide substitution models. But how accurate and general are these models, and which one best suits specific data sets? The availability of numerous nucleotide substitution models creates the challenge of model choice because the wrong model can affect both phylogenetic inference and other statistical analyses. Our journals have led the field in proposing and evaluating nucleotide substitution models, ranging from the first model published in *MBE* (Tajima and Nei 1984) to refinements that improve their generality (Tamura and Nei 1993). The journals have also published several methods to test the fit of evolutionary models, such as PartitionFinder (Lanfear et al. 2012; Lanfear et al. 2017), SMS (Lefort et al. 2017),

ModelTest-NG (Darriba et al. 2020), jModelTest (Posada 2008), and additional tests that take codon position into account (Shapiro et al. 2006). Huelsenbeck et al. (2004) expanded the pool of models for comparison using Bayes factors and a reversible jump MCMC; their approach included the possibility to test among nonnested models (Huelsenbeck et al. 2004). Using empirical data, it is a complex task to ascertain the correct model and it is useful to know if the tree-building method is robust against model violations. To assess the scale and impact of model violations for phylogenetic inference, Naser-Khdour et al. (2019) implemented a maximal matched-pairs test of homogeneity.

Finally, rooting an evolutionary tree is a critical step in phylogenetic inference, as only rooted trees can be directly related to divergence times. Lake et al. (2007) developed the top-down indel rooting method that uses nonubiquitous genes to root the tree of life. Unlike previous methods, this rooting method uses information not only on indel gains and losses but also on gene gains and losses (Lake et al. 2007).

Hypothesis Testing in a Phylogenetic Framework

Phylogenies allow us to address questions about species relationships, but they also form the basis for downstream analyses. Some fundamental analyses use phylogenies to visualize data (phenotypic, experimental, clinical, etc.) onto a phylogenetic framework. For cases like these, Yu et al. (2018) developed an R package named ggtree to map and visualize data.

Other applications require additional inference. One useful goal is to produce a timetree that places divergence times on interior nodes. The construction of timetrees has a lengthy history in the SMBE journals. For example, Takezaki et al. (1995) developed a method based on the assumption of a strict molecular clock. Their application used a single calibration time point and allowed users to remove lineages that deviated significantly from the rate constancy assumption. This linearized tree method differs from the current relaxed-clock methods that are currently popular. For example, Yang and Rannala (2005) developed a Bayesian MCMC algorithm that allows for multiple and softbound calibration priors for estimating divergence times, contrasting with the single or hardbound time priors of earlier methods. New methods can even accommodate phylogenetic uncertainty (Baele et al. 2012), and most of them are able to estimate divergence times when evolutionary rates vary across lineages (Sanderson 1997; Rambaut and Brooman 1998; Thorne et al. 1998; Sanderson 2002).

Many of these methods employ Bayesian approaches, which can be computationally intensive and prohibitive with large data sets. RelTime is a faster algorithm for estimating timetrees. Although originally published elsewhere (Tamura et al. 2012), recent advances and evaluations of RelTime have been published in *MBE*, including the theoretical foundation of the method (Tamura et al. 2018) and evaluations based on simulated (Filipski et al. 2014) and empirical (Mello et al. 2017) data sets.

We have also published a helpful online resource called TimeTree of Life (Kumar et al. 2017). This project assembles information on the timescale of life, in a similar way to the Tree of Life project (Maddison et al. 2007), but focuses on divergence times rather than clades. This resource includes searchable divergence time data from ~4,000 studies across ~148,000 species and is now in its fifth release (Kumar et al. 2022). Apart from time estimates, TimeTree also allows users to explore timetrees that are extracted, for a given taxonomic group, from the global timetree.

Another major theme, which is closely related to estimating and testing molecular clocks, is estimating rates of nonsynonymous (d_N) and synonymous (d_S) evolution and their ratio (d_N/d_S) along evolutionary branches of a tree. The use of d_N/d_S as a signal of adaptive evolution was fueled in part by advances like the publication of a simple distance method that estimates d_N and d_S between pairs of OTUs (Nei and Gojobori 1986). The later publication of codon-based models of nucleotide substitution (Goldman and Yang 1994; Muse and Gaut 1994) laid the groundwork for maximum likelihood hypothesis testing of d_N/d_S on evolutionary trees, including the capability to test for adaptive evolution on individual branches and on specific codons (Yang and Nielsen 2000, 2002).

Population Genetic Analysis

Population genetics has been a prominent topic in SMBE journals since their inception. In fact, the first issue featured a paper by Motoo Kimura, which extended his "neutral theory of molecular evolution and polymorphism" to estimate the fraction of selectively neutral alleles among new mutations from what became known as the site frequency spectrum (Kimura 1983). This method presaged those that estimate the distribution of fitness effects (DFE) and the rate of adaptive evolution (Eyre-Walker and Keightley 2009; Huang et al. 2021). These and similar methods have spawned numerous empirical studies (e.g. Slotte et al. 2010; Gossmann et al. 2012; Tsagkogeorga et al. 2012; Campos et al. 2018) that have provided insights into both the evolutionary processes and the variation of DFEs across populations and species.

Another major theme has been analyzing the effect of evolutionary forces, such as gene flow and drift, on rates of population divergence. Before these forces can be characterized, however, it may be important to first identify separate populations. To this end, Hudson et al. (1992) published a widely used test for detecting geographic subdivision among populations. Once populations are identified, one can model the process of divergence in the face of gene flow between populations. Also, Hey (2010) contributed seminal work to this topic by extending isolation–migration models to multiple populations, which also provided a means to estimate divergence times and migration rates across populations. Related

innovations include the ABBA–BABA statistics (Durand et al. 2011) and their extensions (Martin and Amos 2021). These statistics can detect the signatures of introgression between populations and, in some cases, can also infer the direction of historical introgression events (Martin and Amos 2021).

Population divergence is affected by genetic drift, which is a function of demographic history, another recurring focus of our journals. The inference of demography has undergone a revolution with the introduction of methods that estimate the chronological history of effective population size (N_e) without a predefined demographic model. These methods rely on the temporal rate of coalescence in a genealogy, and they typically produce a plot of estimated N_e over time—i.e. a "skyline plot." Several refinements and improvements to this approach have been published in SMBE journals, including generalizing skyline plots for cases with low divergence (Strimmer and Pybus 2001), estimating plots directly from sequence data rather than from an inferred genealogy (Drummond et al. 2005), and improving temporal smoothing and inference (Minin et al. 2008).

We have touched on many themes in population genetics, but some might argue that *the* major theme is detecting adaptive evolution. This topic is so broad, with such a rich history in our journals, that one of the perspectives in the 40th anniversary series will focus on detecting deviations from neutrality. In this context, we want to draw attention to three basic points. First, the historical effects of demographic changes often complicate the inference of selection, because these processes can produce similar diversity patterns (Johri et al. 2022). Substantive efforts have been invested into controlling for demographic history prior to inferring selection using both empirical (Tenaillon et al. 2004; Stajich and Hahn 2005) and modeling approaches. An example of the latter is *dadi*, which uses allele frequency information to infer demographic history (Gutenkunst et al. 2009) but can incorporate selection on single sites and predict the joint distribution of selected alleles among populations. An update in *MBE* improves the performance of *dadi* and applies it to more than three populations (Gutenkunst 2021).

Second, our community has shown strong interests in detecting and understanding the behavior of selective sweeps. For example, SweeD adapted the CLR (Composite Likelihood Ratio) test (Nielsen et al. 2005) to a high-performance computing environment (Pavlidis et al. 2013). The CLR test is most effective at detecting hard sweeps caused by a single, new adaptive mutation. However, both theory (Pennings and Hermisson 2006) and empirical data (Schrider and Kern 2017) suggest that soft sweeps—i.e. sweeps that result from multiple (often competing) alleles—are likely to be common in nature. Haplotype-based methods have been particularly helpful for detecting soft sweeps. Excoffier and Slatkin (1995) produced key early work in this area by presenting methods to estimate haplotype frequencies; Ferrer-Admetlla et al. (2014) introduced a summary statistic based on haplotype frequencies to test for selection, with improvements in statistical power compared to previous methods; and Harris and DeGiorgio (2020) published an approach that is useful for detecting both hard and soft sweeps from the haplotype frequency spectrum. More recent publications include a haplotype-based method (Flex-sweep) that utilizes convolutional neural networks (Lauterbur et al. 2023) and an efficient haplotype-based approach suitable for large data sets (Kirsch-Gerweck et al. 2023).

Many methods for detecting both soft and hard sweeps focus on single loci or genomic regions, but adaptation is often polygenic (Pritchard et al. 2010). The detection of polygenic adaptation requires different approaches, and one valuable approach for detecting potential polygenic adaptation was published in *MBE*. Frichot et al. (2013) applied a "latent factor mixed model" (LFMM) to identify alleles (i) with frequency distributions that are not easily explained by population structure and (ii) that are associated with other variables, such as bioclimatic measures. Alleles that fit both of these criteria are candidates for contributing to local adaptation that may be polygenic. A subsequent publication updated the LFMM approach (Caye et al. 2019).

Finally, the prevalence and role of balancing selection in the evolutionary process remains poorly understood. However, many of the methods that have been used to detect balancing selection have been published in SMBE journals (Hunter-Zinck and Clark 2015; Siewert and Voight 2017; Bitarello et al. 2018; Cheng and DiGiorgio 2019).

Software Packages

In addition to publishing methods, the SMBE journals have published numerous popular software packages that make methods accessible to our community. Many of these packages embed features like alignment, model choice, phylogenetic inference, and associated downstream analyses. MEGA is the most prominent of these packages. Although MEGA was first announced elsewhere (Kumar et al. 1994), several versions have been published in *MBE* (e.g. Tamura et al. 2007; Kumar et al. 2018; Tamura et al. 2021). These more recent versions have widely increased its scope, including more statistical analyses and various tree-building and timetree methods. Taken together, the collection of MEGA papers has been cited more than any other set of publications in our journals, which is a strong testament to both their scientific contributions and their practical importance. The fast RelTime analyses have been implemented in MEGA (Tamura et al. 2021), allowing users to evaluate the robustness of time estimates by, for example, comparing divergence times using distinct calibration sets. Three protocol papers, which detail pipelines for empirical analyses, have been published that focus on analyses using MEGA. These protocols guide MEGA users for building phylogenetic trees (Hall 2013), performing bootstrap tests of branch support (Russo and Selvatti 2018), and estimating divergence times (Mello 2018).

Our journals have published other prominent software packages like DAMBE (Xia 2013), PAML (Yang 2007), IQ-TREE (Nguyen et al. 2015; Minh et al. 2020), and SEAVIEW (Gouy et al. 2010). DAMBE covers most of the major steps of phylogenetic inference, such as sequence alignment, model selection, and tree building, as well as analyses like codon bias detection and inferring the isoelectric point of a particular enzyme in a solution. SEAVIEW is useful for aligning sequences, to inspect and edit alignments, and to concatenate individual blocks of alignment. It also performs some phylogenetic inference, although it is not its major focus. PAML features a large number of evolutionary models that can be used to compare and test phylogenetic trees and to test alternative biological hypotheses. Importantly, PAML can also be used to reconstruct ancestral gene and protein sequences. Many nonsynonymous (d_N) and synonymous (d_S) evolution and similar methods are also available in PAML (Yang 2007) and HYPHY (Pond et al. 2005); these programs are largely responsible for the burst of interest in investigating adaptive molecular evolution in coding sequences during the 2000s. *MBE* has published recent updates to both programs (Xu and Yang 2013; Pond et al. 2020) as well as FUBAR, an approach that can rapidly detect positive and purifying natural selection with large data sets (Murrel et al. 2013). A useful PAML protocol to estimate synonymous and nonsynonymous distances and to detect positive selection was recently published in *MBE* (Álvarez-Carretero et al. 2023).

The IQ-TREE package is another popular and user-friendly program. It includes a search algorithm that greatly improves exploration of tree space, yielding ML trees with higher likelihoods (Nguyen et al. 2015). In the latest release, search algorithms are able to use multicore CPUs and a parallel MPI (Message Passing Interface) system to speed analyses (Minh et al. 2020). The latest version incorporates over 200 time-reversible evolutionary models for DNA, protein, codon, binary, and multistate morphological data, as well as the ultrafast bootstrap and ModelFinder algorithms.

For timetree analysis, a popular program is BEAST, which can use either strict- or relaxed-clock models in a Bayesian framework. The species tree inference using biallelic markers, such as SNPs and AFLPs, was also implemented in BEAST (Bryant et al. 2012). BEAST has now been updated elsewhere (Suchard et al. 2018), but an earlier version was published in SMBE journals (Drummond et al. 2012). We have also published BEAST-related protocols for estimation of past population dynamics (Hill and Beale 2019) and for phylogeographic inference (Dellicour et al. 2021).

The latest update of the BUSCO method (Simão et al. 2015) was published in *MBE* (Manni et al. 2021). The primary purpose of BUSCO is to provide quality control on new genome assemblies by assessing the complement of near-universal single-copy orthologs within an assembly. BUSCO works with taxon-specific databases of single-copy orthologs, and the latest versions include greatly expanded representation of eukaryotic, prokaryotic, and viral genes. The new version of BUSCO also enables automatic database selection based on the phylogenetic insertion of input sequences.

With respect to population genetic analyses, the SMBE journals have published packages such as DNASP (Rozas et al. 2017), POPGENOME (Pfeifer et al. 2014), FUBAR (Murrell et al. 2013), and SLiM3 (Haller and Messer 2019a). The latest version of DNASP is configured for large data sets and is particularly suitable for genomic partitioning data such as RADseq. The POPGENOME package uses the R environment to process genome-scale data; it offers a range of population genetic analyses, including neutrality tests, population differentiation analysis, and recombination and disequilibrium detection tests. Finally, the SLiM forward simulation software is a popular and invaluable tool for population genetic analyses; *MBE* has published updates to this software (Haller and Messer 2017, 2019a) and a step-by-step protocol for new users (Haller and Messer 2019b).

It is important to emphasize that these software packages do not simply provide access to various methods and algorithms; new versions often introduce new and more accurate methods, enabling users to pursue their own analytical designs. Furthermore, the packages are freely available, usually multiplatform, and often feature detailed manuals with user-friendly graphical interfaces. In many cases, they can run on regular desktop computers without large memory or disk space requirements. These factors are aligned with SMBE journals' policies and traditions and could explain their enormous success and critical role in bringing new members to our molecular evolution community.

The Next 40 Years

The SMBE journals are thriving, publishing over 500 articles a year that merge the approaches of molecular biology, computational biology, statistics, genomics, and evolutionary theory. They have grown to encompass new fields and technologies associated with functional and evolutionary genomics, and they continue to be ranked among top evolutionary biology and genetics journals. Moreover, the SMBE journals remain dedicated to ensuring that data and tools are fully accessible from the point of publication onward, exemplified by *GBE*'s initiative to manually verify data availability statements in all accepted manuscripts.

This perspective, along with the accompanying virtual issue, should convince readers that authors publishing in SMBE journals have had a major impact on the methods and software used for evolutionary inference. Given the success of the last 40 years, we cannot help but wonder what the next 40 will bring. Although we are well into the "post-genomics era," the pace of methodological advances for interpreting genomic data has not slowed. Some of these advances are fueled by the relative ease of procuring new data, contributing to ever larger data sets. Other innovations are fueled by new data types—e.g. single cell expression, chromatin and epigenomic data, improved biochemical structures, 3D genome topologies, and long-read assemblies.

As data sets expand, computational and statistical approaches will continue to evolve. Machine learning

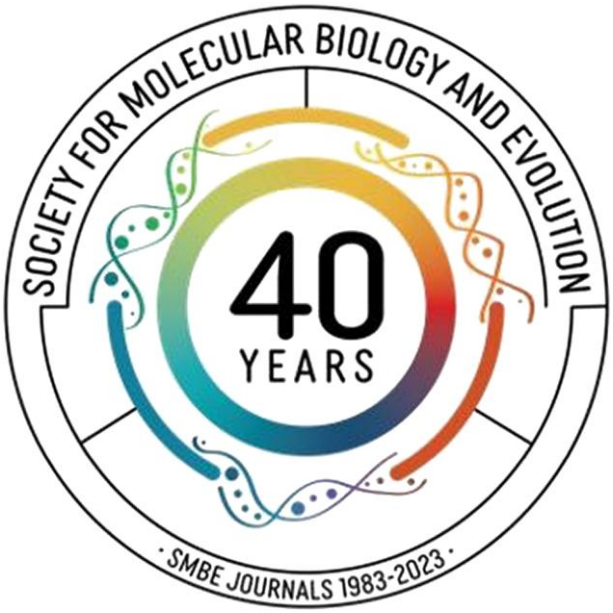

Fig. 3. A logo to celebrate the 40th anniversary, which was designed by SMBE member Ana Carolina Martins Junqueira from the Federal University of Rio de Janeiro, Brazil.

methods are increasingly applied to numerous problems in our field, including phylogenetic inference (Azouri et al. 2021), coevolutionary rates of branches on a phylogeny (Tamura et al. 2021), model selection for phylogenetic inference (Abadi et al. 2020), detection of selective sweeps (Lauterbur et al. 2023), and other population genetic inferences (Flagel et al. 2019). Likewise, the broader field of artificial intelligence (AI) is already having substantive impacts, both formally and informally. Researchers are, for example, using ChatGPT and other AI platforms for editing, coding, and preliminary data analyses. The application of these methods across distributed computing platforms will yield a new generation of methods for the curation, analysis, and interpretation of large data sets. We want to be at the forefront of these developments over the next four decades, just as our journals have contributed to revolutions in genomic, genetic, and structural analyses over the last four.

We conclude this perspective by inviting you to celebrate the 40th anniversary (Fig. 3) by perusing the upcoming virtual issues and accompanying perspectives. But you need no invitation, because the SMBE journals are your journals. They were established by our burgeoning community in 1983, and they continue to be managed by the community and for the benefit of the community. They rely on your expert opinions as authors, reviewers, and editors. If you have authored, reviewed, edited, or read papers in *GBE* and *MBE*, then this is your celebration!

References

Abadi S, Avram O, Rosset S, Pupko T, Mayrose I. ModelTeller: model selection for optimal phylogenetic reconstruction using machine learning. *Mol Biol Evol*. 2020:**37**(11):3338–3352. https://doi.org/10.1093/molbev/msaa154.

Álvarez-Carretero S, Kapli P, Yang Z. Beginner's guide on the use of PAML to detect positive selection. *Mol Biol Evol*. 2023:**40**(4):msad041. https://doi.org/10.1093/molbev/msad041.

Ané C, Larget B, Baum DA, Smith SD, Rokas A. Bayesian estimation of concordance among gene trees. *Mol Biol Evol*. 2007:**24**(2):412–426. https://doi.org/10.1093/molbev/msl170.

Azouri D, Abadi S, Mansour Y, Mayrose I, Pupko T. Harnessing machine learning to guide phylogenetic-tree search algorithms. *Nat Commun*. 2021:**12**(1):1983. https://doi.org/10.1038/s41467-021-22073-8.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol*. 2012:**29**(9):2157–2167. https://doi.org/10.1093/molbev/mss084.

Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999:**16**(1):37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036.

Bitarello BD, Filippo C, Teixeira JC, Schmidt JM, Kleinert P, Meyer D, Andrés AM. Signatures of long-term balancing selection in human genomes. *Gen Biol Evol*. 2018:**10**(3):939–955. https://doi.org/10.1093/gbe/evy054.

Bruno WJ, Socci ND, Halpern AL. Weighted neighbor-joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol Biol Evol*. 2000:**17**(1):189–197. https://doi.org/10.1093/oxfordjournals.molbev.a026231.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhry A. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol Biol Evol*. 2012:**29**(8):1917–1932. https://doi.org/10.1093/molbev/mss086.

Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol*. 2004:**21**(2):255–265. https://doi.org/10.1093/molbev/msh018.

Campos JL, Johnston KJA, Charlesworth B. The effects of sex-biased gene expression and X-linkage on rates of sequence evolution in *Drosophila*. *Mol Biol Evol*. 2018:**35**(3):655–665. https://doi.org/10.1093/molbev/msx317.

Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 2000:**17**(4):540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Caye K, Jumentier B, Lepeule J, François O. LFMM 2: fast and accurate inference of gene-environment associations in genome-wide studies. *Mol Biol Evol*. 2019:**36**(4):852–860. https://doi.org/10.1093/molbev/msz008.

Cheng X, DiGiorgio M. Detection of shared balancing selection in the absence of trans-species polymorphism. *Mol Biol Evol*. 2019:**36**(1):177–199. https://doi.org/10.1093/molbev/msy202.

Comte A, Tricou T, Tannier E, Joseph J, Siberchicot A, Penel S, Allio R, Delsuc F, Dray S, de Vienne DM. Phylter: efficient identification of outlier sequences in phylogenomic datasets. *Mol Biol Evol*. 2023:**40**(11):msad234. https://doi.org/10.1093/molbev/msad234.

Darriba D, Posada D, Kozlov AM M, Stamatakis A, Morel B, Flouri T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol Biol Evol*. 2020:**37**(1):291–294. https://doi.org/10.1093/molbev/msz189.

Dellicour S, Gill MS, Faria NR, Rambaut A, Pybus OG, Suchard MA, Lemey P. Relax, keep walking—a practical guide to continuous phylogeographic inference with BEAST. *Mol Biol Evol*. 2021:**38**(8):3486–3493. https://doi.org/10.1093/molbev/msab031.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol*. 2005:**22**(5):1185–1192. https://doi.org/10.1093/molbev/msi103.

Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012:**29**(8):1969–1973. https://doi.org/10.1093/molbev/mss075.

Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. *Mol Biol Evol*. 2011:**28**(8):2239–2252. https://doi.org/10.1093/molbev/msr048.

Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol*. 1995:**12**(5):921–927. https://doi.org/10.1093/oxfordjournals.molbev.a040269.

Eyre-Walker A, Keightley PD. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol*. 2009:**26**(9):2097–2108. https://doi.org/10.1093/molbev/msp119.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol Evol*. 2014:**31**(5):1275–1291. https://doi.org/10.1093/molbev/msu077.

Filipski A, Murillo O, Freydenzon A, Tamura K, Kumar S. Prospects for building large timetrees using molecular data with incomplete gene coverage among species. *Mol Biol Evol*. 2014:**31**(9):2542–2550. https://doi.org/10.1093/molbev/msu200.

Flagel L, Brandvain Y, Schrider DR. The unreasonable effectiveness of convolutional neural networks in population genetic inference. *Mol Biol Evol*. 2019:**36**(2):220–238. https://doi.org/10.1093/molbev/msy224.

Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol*. 2013:**30**(7):1687–1699. https://doi.org/10.1093/molbev/mst063.

Gascuel O. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol*. 1997:**14**(7):685–695. https://doi.org/10.1093/oxfordjournals.molbev.a025808.

Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. 1994:**11**(5):725–736. https://doi.org/10.1093/oxfordjournals.molbev.a040153.

Gossmann TI, Keightley PD, Eyre-Walker A. The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Gen Biol Evol*. 2012:**4**(5):658–667. https://doi.org/10.1093/gbe/evs027.

Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 2010:**27**(2):221–224. https://doi.org/10.1093/molbev/msp259.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*. 2009:**5**(10):e1000695. https://doi.org/10.1371/journal.pgen.1000695.

Gutenkunst RN. Dadi.CUDA: accelerating population genetics inference with graphics processing units. *Mol Biol Evol*. 2021:**38**(5):2177–2178. https://doi.org/10.1093/molbev/msaa305.

Hall BG. Building phylogenetic trees from molecular data using MEGA. *Mol Biol Evol*. 2013:**30**(5):1229–1235. https://doi.org/10.1093/molbev/mst012.

Haller BC, Messer PW. SLiM 2: flexible, interactive forward genetic simulations. *Mol Biol Evol*. 2017:**34**(1):230–240. https://doi.org/10.1093/molbev/msw211.

Haller BC, Messer PW. SLiM 3: forward genetic simulations beyond the Wright–Fisher model. *Mol Biol Evol*. 2019a:**36**(3):632–637. https://doi.org/10.1093/molbev/msy228.

Haller BC, Messer PW. Evolutionary modeling in SLiM 3 for beginners. *Mol Biol Evol*. 2019b:**36**(5):1101–1109. https://doi.org/10.1093/molbev/msy237.

Harris AM, DeGiorgio M. A likelihood approach for uncovering selective sweep signatures from haplotype data. *Mol Biol Evol*. 2020:
37(10):3023–3046. https://doi.org/10.1093/molbev/msaa115.

Hedges SB. The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Mol Biol Evol*. 1992:**9**(2):366–369. https://doi.org/10.1093/oxfordjournals.molbev.a040725.

Heled J, Drummond AJ. Bayesian inference of species trees from multilocus data. *Mol Biol Evol*. 2010:**27**(3):570–580. https://doi.org/10.1093/molbev/msp274.

Hey J. Isolation with migration models for more than two populations. *Mol Biol Evol*. 2010:**27**(4):905–920. https://doi.org/10.1093/molbev/msp296.

Hill V, Baele G. Bayesian estimation of past population dynamics in BEAST 1.10 using the skygrid coalescent model. *Mol Biol Evol*. 2019:**36**(11):2620–2628. https://doi.org/10.1093/molbev/msz172.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018:**35**(2):518–522. https://doi.org/10.1093/molbev/msx281.

Huang X, Fortier AL, Coffman AJ, Struck TJ, Irby MN, James JE, León-Burguete JE, Ragsdale AP, Gutenkunst RN. Inferring genome-wide correlations of mutation fitness effects between populations. *Mol Biol Evol*. 2021:**38**(10):4588–4602. https://doi.org/10.1093/molbev/msab162.

Hudson RR, Boos DD, Kaplan NL. A statistical test for detecting geographic subdivision. *Mol Biol Evol*. 1992:**9**(1):138–151. https://doi.org/10.1093/oxfordjournals.molbev.a040703.

Huelsenbeck JP. The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Mol Biol Evol*. 1995:**12**(5):843–849. https://doi.org/10.1093/oxfordjournals.molbev.a040261.

Huelsenbeck JP, Larget B, Alfaro ME. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol Biol Evol*. 2004:**21**(6):1123–1133. https://doi.org/10.1093/molbev/msh123.

Hunter-Zinck H, Clark A. Aberrant time to most common ancestor as a signature of natural selection. *Mol Biol Evol*. 2015:**32**(10):2784–2797. https://doi.org/10.1093/molbev/msv142.

Huson DH, Scornavacca C. A survey of combinatorial methods for phylogenetic networks. *Gen Biol Evol*. 2011:**3**(1):23–35. https://doi.org/10.1093/gbe/evq077.

Johri P, Eyre-Walker A, Gutenkunst RN, Lohmueller KE, Jensen JD. On the prospect of achieving accurate joint estimation of selection with population history. *Gen Biol Evol*. 2022:**14**(7):evac088. https://doi.org/10.1093/gbe/evac088.

Kalaghatgi P, Pfeifer N, Lengauer T. Family-Joining: a fast distance-based method for constructing generally labeled trees. *Mol Biol Evol*. 2016:**33**(10):2720–2734. https://doi.org/10.1093/molbev/msw123.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013:**30**(4):772–780. https://doi.org/10.1093/molbev/mst010.

Kimura M. Rare variant alleles in the light of the neutral theory. *Mol Biol Evol*. 1983:**1**(1):84–93. https://doi.org/10.1093/oxfordjournals.molbev.a040305.

Kirsch-Gerweck B, Bohnenkämper L, Henrichs MT, Alanko JN, Bannai H, Cazaux B, Peterlongo P, Burger J, Stoye J, Diekmann Y. HaploBlocks: efficient detection of positive selection in large population genomic datasets. *Mol Biol Evol*. 2023:**40**(3):msad027. https://doi.org/10.1093/molbev/msad027.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018:**35**(6):1547–1549. https://doi.org/10.1093/molbev/msy096.

Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. 2017:**34**(7):1812–1819. https://doi.org/10.1093/molbev/msx116.

Kumar S, Suleski M, Craig JE, Kasprowicz AE, Sanderford M, Stecher G, Hedges SB. TimeTree 5: an expanded resource for species divergence times. *Mol Biol Evol*. 2022:**39**(8):msac174. https://doi.org/10.1093/molbev/msac174.

Kumar S, Tamura K, Nei M. MEGA: molecular evolutionary genetics analysis software for microcomputers. *Comput Appl Biosci*. 1994:**10**(2):189–191. https://doi.org/10.1093/bioinformatics/10.2.189.

Lake JA, Herbold CW, Rivera MC, Servin JA, Skophammer RG. Rooting the tree of life using nonubiquitous genes. *Mol Biol Evol*. 2007:**24**(1):130–136. https://doi.org/10.1093/molbev/msl140.

Lanfear R, Calcott B, Ho SY, Guindon S. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol*. 2012:**29**(6):1695–1701. https://doi.org/10.1093/molbev/mss020.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott C. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol*. 2017:**34**(3):772–773. https://doi.org/10.1093/molbev/msw260.

Lauterbur ME, Munch K, Enard D. Versatile detection of diverse selective sweeps with Flex-sweep. *Mol Biol Evol*. 2023:**40**(6): msad139. https://doi.org/10.1093/molbev/msad139.

Lefort V, Longueville J-E, Gascuel O. SMS: smart model selection in PhyML. *Mol Biol Evol*. 2017:**34**(9):2422–2424. https://doi.org/10.1093/molbev/msx149.

Maddison DR, Schulz K-S, Maddison WP. The tree of life project. *Zootaxa* 2007:**1668**(1):19–40. https://doi.org/10.11646/zootaxa.1668.1.4.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021:**38**(10): 4647–4654. https://doi.org/10.1093/molbev/msab199.

Martin SH, Amos W. Signatures of introgression across the allele frequency spectrum. *Mol Biol Evol*. 2021:**38**(2):716–726. https://doi.org/10.1093/molbev/msaa239.

Mello B. Estimating timetrees with MEGA and the TimeTree resource. *Mol Biol Evol*. 2018:**35**(9):2334–2342. https://doi.org/10.1093/molbev/msy133.

Mello B, Tao Q, Tamura K, Kumar S. Fast and accurate estimates of divergence times from big data. *Mol Biol Evol*. 2017:**34**(1):45–50. https://doi.org/10.1093/molbev/msw247.

Minh BQ, Nguyen AT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol*. 2013:**30**(5):1188–1195. https://doi.org/10.1093/molbev/mst024.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020:**37**(5):1530–1534. https://doi.org/10.1093/molbev/msaa015.

Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol*. 2008:**25**(7):1459–1471. https://doi.org/10.1093/molbev/msn090.

Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. *Mol Biol Evol*. 2013:**30**(5): 1196–1205. https://doi.org/10.1093/molbev/mst030.

Muse SV, Gaut BS. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol*. 1994:**11**(5): 715–724. https://doi.org/10.1093/oxfordjournals.molbev.a040152.

Naser-Khdour S, Minh BQ, Zhang W, Stone AE, Lanfear R. The prevalence and impact of model violations in phylogenetic analysis. *Gen Biol Evol*. 2019:**11**(12):3341–3352. https://doi.org/10.1093/gbe/evz193.

Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 1986:**3**(5):418–426. https://doi.org/10.1093/oxfordjournals.molbev.a040410.

Nguyen L-T, Schmidt HA, von Haesler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015:**32**(1):268–274. https://doi.org/10.1093/molbev/msu300.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. *Genome Res*. 2005:**15**(11):1566–1575. https://doi.org/10.1101/gr.4252305.

Ota S, Li W-H. NJML: a hybrid algorithm for the neighbor-joining and maximum-likelihood methods. *Mol Biol Evol*. 2000:**17**(9): 1401–1409. https://doi.org/10.1093/oxfordjournals.molbev.a026423.

Pamilo P, Nei M. Relationships between gene trees and species trees. *Mol Biol Evol*. 1988:**5**(5):568–583. https://doi.org/10.1093/oxfordjournals.molbev.a040517.

Pavlidis P, Živković D, Stamatakis A, Alachiotis N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol*. 2013:**30**(9):2224–2234. https://doi.org/10.1093/molbev/mst112.

Pearson WR, Robins G, Zhang T. Generalized neighbor-joining: more reliable phylogenetic tree reconstruction. *Mol Biol Evol*. 1999: **16**(6):806–816. https://doi.org/10.1093/oxfordjournals.molbev.a026165.

Pennings PS, Hermisson J. Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Mol Biol Evol*. 2006:**23**(5):1076–1084. https://doi.org/10.1093/molbev/msj117.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol*. 2014:**31**(7):1929–1936. https://doi.org/10.1093/molbev/msu136.

Pond KSL, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 2005:**21**(5):676–679. https://doi.org/10.1093/bioinformatics/bti079.

Pond KSL, Poon AFY, Velazques R, Weaver S, Hepler NL, Murrell B, Shank SD, Rife Magalis B, Bouvier D, Nekrutenko A, *et al.* HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol*. 2020:**37**(1):295–299. https://doi.org/10.1093/molbev/msz197.

Posada D. jModelTest: phylogenetic model averaging. *Mol Biol Evol*. 2008:**25**(7):1253–1256. https://doi.org/10.1093/molbev/msn083.

Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*. 2009:**26**(7):1641–1650. https://doi.org/10.1093/molbev/msp077.

Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol*. 2010:**20**(4):R208–R215. https://doi.org/10.1016/j.cub.2009.11.055.

Rambaut A, Brooman L. Estimating divergence dates from molecular sequences. *Mol Biol Evol*. 1998:**15**(4):442–448. https://doi.org/10.1093/oxfordjournals.molbev.a025940.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017:**34**(12):3299–3302. https://doi.org/10.1093/molbev/msx248.

Russo CAM, Selvatti AP. Bootstrap and rogue identification tests for phylogenetic analyses. *Mol Biol Evol*. 2018:**35**(9):2327–2333. https://doi.org/10.1093/molbev/msy118.

Rzhetsky A, Nei M. A simple method for estimating and testing minimum-evolution trees. *Mol Biol Evol*. 1992:**9**(5):945. https://doi.org/10.1093/oxfordjournals.molbev.a040771.

Rzhetsky A, Nei M. Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol Biol Evol*. 1993:**10**(5):1073–1095. https://doi.org/10.1093/oxfordjournals.molbev.a040056.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987:**4**(4): 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

Sanderson MJ. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol Biol Evol*. 1997:**14**(12):1218–1231. https://doi.org/10.1093/oxfordjournals.molbev.a025731.

Sanderson MJ. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol*. 2002:**19**(1):101–109. https://doi.org/10.1093/oxfordjournals.molbev.a003974.

Sayyari E, Mirarab S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol Biol Evol*. 2016:**33**(7):1654–1668. https://doi.org/10.1093/molbev/msw079.

Schrider DR, Kern AD. Soft sweeps are the dominant mode of adaptation in the human genome. *Mol Biol Evol*. 2017:**34**(8):1863–1877. https://doi.org/10.1093/molbev/msx154.

Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol*. 2006:**23**(1):7–9. https://doi.org/10.1093/molbev/msj021.

Siewert KM, Voight BF. Detecting long-term balancing selection using allele frequency correlation. *Mol Biol Evol*. 2017:**34**(11):2996–3005. https://doi.org/10.1093/molbev/msx209.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015:**31**(19):3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Slotte T, Foxe JP, Hazzouri KM, Wright SI. Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. *Mol Biol Evol*. 2010:**27**(8):1813–1821. https://doi.org/10.1093/molbev/msq062.

Stajich JE, Hahn MW. Disentangling the effects of demography and selection in human history. *Mol Biol Evol*. 2005:**22**(1):63–73. https://doi.org/10.1093/molbev/msh252.

Strimmer K, Pybus OG. Exploring the demographic history of DNA sequences using the generalized skyline plot. *Mol Biol Evol*. 2001:**18**(12):2298–2305. https://doi.org/10.1093/oxfordjournals.molbev.a003776.

Strimmer K, von Haeseler A. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol Biol Evol*. 1996:**13**(7):964–969. https://doi.org/10.1093/oxfordjournals.molbev.a025664.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. 2018:**4**(1):vey016. https://doi.org/10.1093/ve/vey016.

Tajima F, Nei M. Estimation of evolutionary distance between nucleotide sequences. *Mol Biol Evol*. 1984:**1**(3):269–285. https://doi.org/10.1093/oxfordjournals.molbev.a040317.

Takezaki N, Rzhetsky A, Nei M. Phylogenetic test of the molecular clock and linearized trees. *Mol Biol Evol*. 1995:**12**(5):823–833. https://doi.org/10.1093/oxfordjournals.molbev.a040259.

Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci USA*. 2012:**109**(47):19333–19338. https://doi.org/10.1073/pnas.1213199109.

Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007:**24**(8):1596–1599. https://doi.org/10.1093/molbev/msm092.

Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*. 1993:**10**(3):512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023.

Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. 2021:**38**(7):3022–3027. https://doi.org/10.1093/molbev/msab120.

Tamura K, Tao Q, Kumar S. Theoretical foundation of the RelTime method for estimating divergence times from variable evolutionary rates. *Mol Biol Evol*. 2018:**35**(7):1770–1782. https://doi.org/10.1093/molbev/msy044.

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. Selection versus demography: a multilocus investigation of the domestication process in maize. *Mol Biol Evol*. 2004:**21**(7):1214–1225. https://doi.org/10.1093/molbev/msh102.

Thorne JL, Kishino H, Painter IS. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol*. 1998:**15**(12):1647–1657. https://doi.org/10.1093/oxfordjournals.molbev.a025892.

Tsagkogeorga G, Cahais V, Galtier N. The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis*. *Gen Biol Evol*. 2012:**4**(8):852–861. https://doi.org/10.1093/gbe/evs054.

Ullah I, Parviainen P, Lagergren J. Species tree inference using a mixture model. *Mol Biol Evol*. 2015:**32**(9):2469–2482. https://doi.org/10.1093/molbev/msv115.

Xia X. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol*. 2013:**30**(7):1720–1728. https://doi.org/10.1093/molbev/mst064.

Xu B, Yang Z. PAMLX: a graphical user interface for PAML. *Mol Biol Evol*. 2013:**30**(12):2723–2724. https://doi.org/10.1093/molbev/mst179.

Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 2000:**17**(1):32–43. https://doi.org/10.1093/oxfordjournals.molbev.a026236.

Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*. 2002:**19**(6):908–917. https://doi.org/10.1093/oxfordjournals.molbev.a004148.

Yang Z, Rannala B. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Mol Biol Evol*. 1997:**14**(7):717–724. https://doi.org/10.1093/oxfordjournals.molbev.a025811.

Yang Z, Rannala B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol*. 2005:**23**(1):212–226. https://doi.org/10.1093/molbev/msj024.

Yang Z, Rannala B. Unguided species delimitation using DNA sequence data from multiple loci. *Mol Biol Evol*. 2014:**31**(12):3125–3135. https://doi.org/10.1093/molbev/msu279.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007:**24**(8):1586–1591. https://doi.org/10.1093/molbev/msm088.

Yu G, Lam TT-Y, Zhu H, Guan Y. Two methods for mapping and visualizing associated data on phylogeny using Ggtree. *Mol Biol Evol*. 2018:**35**(12):3041–3043. https://doi.org/10.1093/molbev/msy194.

Zhang C, Ogilvie HA, Drummond AJ, Stadler T. Bayesian inference of species networks from multilocus sequence data. *Mol Biol Evol*. 2018:**35**(2):504–517. https://doi.org/10.1093/molbev/msx307.