PLOS BIOLOGY

PERSPECTIVE

The dawn of relaxed phylogenetics

Jacob L. Steenwyk[1]*, **Antonis Rokas**[2,3]*

1 Howards Hughes Medical Institute and the Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California, United States of America, **2** Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, United States of America, **3** Vanderbilt Evolutionary Studies Initiative, Vanderbilt University, Nashville, Tennessee, United States of America
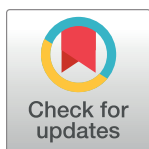
* jlsteenwyk@berkeley.edu (JLS); antonis.rokas@vanderbilt.edu (AR)

Tracing the history of evolution across time is a primary goal of evolutionary biology. The 2006 publication of a landmark study on relaxed phylogenetics in PLOS Biology *enabled biologists to shed light on evolution's tempo and shaped the future of evolutionary studies.*

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: JLS is a scientific consultant for Latch AI Inc. JLS is a scientific advisor for WittGen Biotechnologies. AR is a scientific consultant for LifeMine Therapeutics, Inc.

This article is part of the *PLOS Biology* 20th Anniversary Collection.

Since Emile Zuckerkandl and Linus Pauling proposed their hypothesis of a "molecular evolutionary clock" in their seminal 1965 paper titled "Evolutionary Divergence and Convergence in Proteins," biologists have been fascinated by the prospect of adding a temporal dimension to their inferences of evolutionary relationships of genes and organisms. Early methods for divergence time estimation implemented a "global" model of clock-like evolution, which assumed a constant rate of sequence evolution along a fixed and presumed-to-be-correct phylogeny (Fig 1) [1]. However, it is now well known that rates of sequence evolution differ across lineages (e.g., rates of molecular evolution in organisms with shorter generation times, such as microbes, are much faster than those with longer generation times, such as mammals) violating a fundamental assumption in the global clock model. This variation in evolutionary rate can cause severe problems in molecular evolutionary dating and the inference of phylogenetic relationships [2]. Consequently, many statistical phylogenetic methods implement evolutionary models in which each branch has an independent rate of molecular evolution [3]. While these models revolutionized phylogenetic inference, they cannot disentangle evolutionary rate and time; a branch in a phylogenetic tree may be long (i.e., have a high number of substitutions) because the gene/organism in question is evolving at a fast pace or because it has been accumulating substitutions for a long time or both.

Given that rates of evolution vary, how can we then date phylogenies using molecular data? Early efforts to solve the challenge featured "local" molecular clocks that enabled users to define lineages that experienced different rates of sequence evolution (Fig 1) [4]. However, local clock models still require a known phylogeny and a priori knowledge of which lineages differ in their evolutionary rates. Autocorrelated relaxed clock models do not require a priori knowledge, implementing a model in which the rate of evolution can vary across branches in a phylogeny by positing that closely related species have similar rates and distantly related species may have different rates (Fig 1) [5]. However, autocorrelation of evolutionary rates is itself
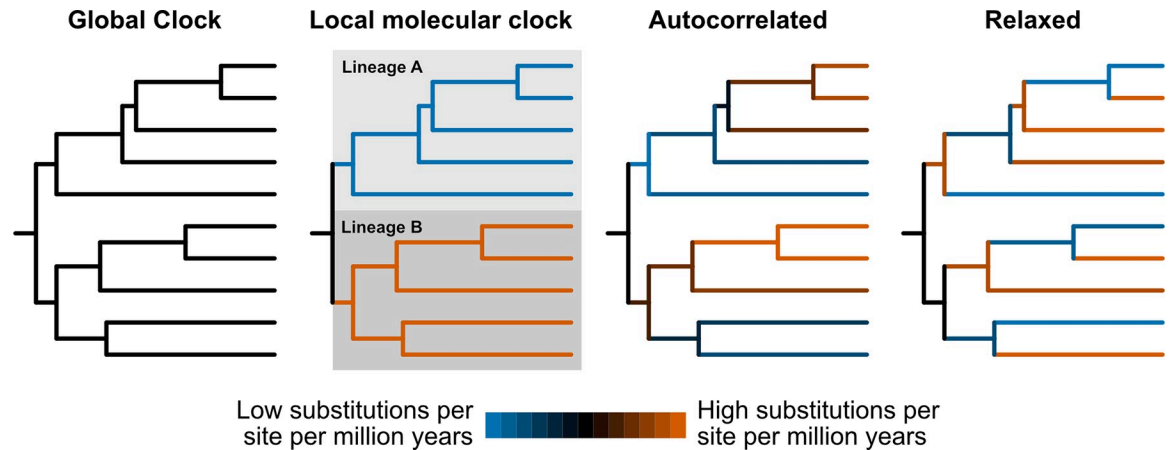
Fig 1. Cartoon representations of different clock models. Global clocks impose the same substitution rate across the phylogeny. Local clocks use the same substitution rate for user-defined lineages. Autocorrelated clocks assume that substitution rates gradually change across speciation events resulting in closely related lineages having similar rates. In relaxed clocks—the critical component of relaxed phylogenetics—the substitution rate of each branch is independent of other branches.

https://doi.org/10.1371/journal.pbio.3001998.g001

an assumption, is challenging to detect, and may vary depending on the breadth of taxon sampling and the evolutionary depth represented in the dataset. Moreover, constraining the phylogenetic tree can be problematic if branches are poorly supported. Lastly, converting relative divergence times to absolute times requires fixed calibration points; for example, fossils can provide a minimum age constraint but are susceptible to uncertainties such as placement on the phylogeny and absolute age.

In a landmark 2006 study in *PLOS Biology*, Drummond and colleagues developed a Bayesian Markov Chain Monte Carlo method to co-estimate phylogenies and divergence times that overcame the hurdles associated with earlier methods (Fig 1) [6]. Key features of this novel "relaxed" approach include allowing each branch in a phylogeny to have a different evolutionary rate (also known as the uncorrelated relaxed clock model), co-estimating substitution and relaxed clock parameters (e.g., substitution rates and time intervals), and implementing priors that took into account calibration uncertainties. These methods were incorporated in the foundational software BEAST [7].

Drummond and colleagues [6] accurately co-estimated phylogeny and divergence times from simulated and empirical data from diverse lineages to demonstrate the efficacy of relaxed phylogenetics. These analyses yielded numerous insights, such as the highly clock-like pattern of evolution among marsupial mammals. Divergence time estimation among marsupials also benefitted from using probabilistic priors, rather than fixed values, for calibration, thereby accounting for uncertainty in the fossil record. The relaxed phylogenetics approach also enabled measuring the clock-likeness of genes, revealing that clock-like evolution is relatively rare and that most genes exhibit non-clock-like rates of sequence evolution.

Relaxed phylogenetics ushered phylogeny estimation into the era where co-estimating phylogeny and divergence times could be done accurately and efficiently. Notable successes, which leverage present-day and ancient samples, include tracing the epidemiology of pathogenic viruses involved in notorious pandemics, like HIV [8], unraveling the timing of the evolution of charismatic megafauna, such as bison [9], and charting the course of recent human evolution [10]. Examining the evolutionary history of HIV shed light on its origin and transmission, facilitating the estimation of effective reproductive numbers in different geographic regions. For example, the relaxed phylogenetics approach revealed that the HIV-1 subtype B in

the United States of America was likely introduced from the Caribbean in the early 1970s and underwent subsequent rapid expansion such that by the late 1970s the epidemic had already spread and diversified across the US [8]. Examination of bison evolution also benefitted from other features in BEAST, such as co-estimating demographic parameters (e.g., population growth and size) and divergence times [9]. Evolutionary relationships among Neanderthals, Denisovans, and modern humans were inferred using BEAST, helping elucidate human evolution [10], a pioneering achievement recently recognized in Svante Pääbo's 2022 Nobel Prize in Physiology or Medicine. These three examples—chosen to highlight the power of relaxed phylogenetics—represent only a tiny sample of the seminal evolutionary insights that continue to be made using this method.

Future iterations of the relaxed phylogenetics approach will hopefully speed up computation, enabling use among increasingly large phylogenomic data matrices and reducing the environmental cost of extensive computation. Also, clarifying circumstances when the accuracy and precision of phylogenetic inference would improve from the use of a relaxed molecular clock approach may help guide experimental design and reduce computational burdens if models with greater complexity are not needed. Machine learning, other forms of artificial intelligence, and advances in computer science and engineering hold promise to overcome these issues.

The Swiss army knife-like properties of the BEAST software have also empowered evolutionary biologists and served as a pillar in the community. The success of BEAST is in part because the software is user-friendly, well-developed, has extensive and clear documentation, and is continually being updated. Companion software such as BEAUti, Tracer, and FigTree, among others, augment the BEAST ethos. These rich resources have enabled generations of scientists to conduct evolutionary inferences and help unravel the tempo and mode of biological evolution.

In the early days of the molecular evolutionary clock, model assumptions constrained the accuracy of divergence time estimates. By introducing relaxed phylogenetics, the 2006 *PLoS Biology* article by Drummond and colleagues enabled the widespread adoption of a biologically more realistic and accurate approach for estimating the tempo and mode of genetic evolution, facilitating impactful—including Nobel-worthy—discoveries.

References

1. Aris-Brosou S, Yang Z. Effects of Models of Rate Evolution on Estimation of Divergence Dates with Special Reference to the Metazoan 18S Ribosomal RNA Phylogeny. Huelsenbeck J, editor. Syst Biol. 2002; 51:703–714. https://doi.org/10.1080/10635150290102375 PMID: 12396585

2. Yoder AD, Yang Z. Estimation of Primate Speciation Dates Using Local Molecular Clocks. Mol Biol Evol. 2000; 17:1081–1090. https://doi.org/10.1093/oxfordjournals.molbev.a026389 PMID: 10889221

3. Felsenstein J. Inferring phylogenies. Sinauer Associates, Sunderland, MA; 2004.

4. Sanderson MJ. A Nonparametric Approach to Estimating Divergence Times in the Absence of Rate Constancy. Mol Biol Evol. 1997; 14:1218–1231. https://doi.org/10.1093/oxfordjournals.molbev.a025731

5. Thorne JL, Kishino H, Painter IS. Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol. 1998; 15:1647–1657. https://doi.org/10.1093/oxfordjournals.molbev.a025892 PMID: 9866200

6. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed Phylogenetics and Dating with Confidence. Penny D, editor. PLoS Biol. 2006; 4:e88. https://doi.org/10.1371/journal.pbio.0040088 PMID: 16683862

7. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007; 7:214. https://doi.org/10.1186/1471-2148-7-214 PMID: 17996036

8. Worobey M, Watts TD, McKay RA, Suchard MA, Granade T, Teuwen DE, et al. 1970s and 'Patient 0' HIV-1 genomes illuminate early HIV/AIDS history in North America. Nature. 2016; 539:98–101. https://doi.org/10.1038/nature19827 PMID: 27783600

9. Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, et al. Rise and Fall of the Beringian Steppe Bison. Science. 2004; 306:1561–1565. https://doi.org/10.1126/science.1101074 PMID: 15567864

10. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010; 468:1053–1060. https://doi.org/10.1038/nature09710 PMID: 21179161