

Review

Not peer-reviewed version

Phylogenomic Inference, Divergence-Time Calibration, and Methods for Characterizing Reticulate Evolution

Saelin Bjornson , Nathan Upham , Heroen Verbruggen , [Jacob Steenwyk](#) *

Posted Date: 14 September 2023

doi: 10.20944/preprints202309.0905.v1

Keywords: phylogenetics; hybridization; introgression; horizontal gene transfer; lateral gene transfer; phylogenetic incongruence; gene-tree-species-tree discordance



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

Phylogenomic Inference, Divergence-Time Calibration, and Methods for Characterizing Reticulate Evolution

Saelin Bjornson ¹, Nathan S. Upham ², Heroen Verbruggen ¹ and Jacob L. Steenwyk ^{3,*}

¹ School of BioSciences, University of Melbourne, Victoria, Australia

² School of Life Science, Arizona State University, Tempe, AZ, USA

³ Howards Hughes Medical Institute and the Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA

* Correspondence: jlsteenwyk@berkeley.edu

Running title: Phylogenomics and reticulate evolution

Abstract: Phylogenomics has enriched our understanding of the Tree of Life. Non-vertical modes of evolution—such as hybridization/introgression and horizontal gene transfer—deviate from a strictly bifurcating tree model, mirroring a network-like or reticulate structure. Here, we present an overview of a phylogenomic workflow for inferring organismal histories, calibrating those histories to evolutionary time, and detecting reticulate evolution. Mitigating analytical sources of error facilitates accurate reconstructions of evolutionary history and, in turn, characterization of non-vertical modes of evolution. Workflows and methods discussed herein may aid in the rigorous inference of organismal histories in geologic time and reticulation, providing a clearer understanding of the evolutionary process.

Keywords: phylogenetics; hybridization; introgression; horizontal gene transfer; lateral gene transfer; phylogenetic incongruence; gene-tree-species-tree discordance

Introduction

Phylogenomics—phylogenetic analysis using genome-scale data—has been used to infer the evolutionary history of diverse lineages across the Tree of Life, including animals, fungi, plants, bacteria, archaea, and viruses (Dunn et al. 2008; Misof et al. 2014; Wickett et al. 2014; Worobey et al. 2016; Simion et al. 2017; Parks et al. 2018; Shen et al. 2018; One Thousand Plant Transcriptomes Initiative 2019; Zhu et al. 2019; Coleman et al. 2021; Galindo et al. 2021; Li et al. 2021; Tahon et al. 2021). These studies have resolved numerous phylogenetic controversies, deepening our understanding of life's history (Capella-Gutiérrez et al. 2012; King and Rokas 2017; Williams et al. 2019; Pipes et al. 2021; Steenwyk et al. 2023b). Phylogenomics has also proven useful for delineating lineage relationships at various taxonomic scales, ranging from species to higher-order relationships (Díaz-Tapia et al. 2017; Muñoz-Gómez et al. 2017; Mateo-Estrada et al. 2019; Bringle et al. 2021; Steenwyk et al. 2022a; Sierra-Patev et al. 2023), and provides the necessary framework for comparative evolutionary genomic studies, such as determining gene duplication and loss events or studying phenotypic innovation (Zhang et al. 2014a; Steenwyk et al. 2019a; Fernández and Gabaldón 2020; Shen et al. 2020; Phillips et al. 2021; Li et al. 2022a; Oplente et al. 2023).

However, errors may be introduced during analytical steps, such as orthology and site-wise homology inference, resulting in inaccurate reconstructed evolutionary histories (Martín-Durán et al. 2017; Ashkenazy et al. 2019; Emms and Kelly 2019; Steenwyk et al. 2023b). Similarly, the inferred timing of ancestor-to-descendent divergence events can be erroneous due to misspecifying molecular clock models, fossil calibrations, or other factors (Ho and Duchêne 2014; Tao et al. 2020; Carruthers and Scotland 2021). Careful consideration of these analytical sources of error during experimental design is crucial for improving the accuracy of phylogenomic inference (Steenwyk et al. 2023b).

Incongruence between the evolutionary histories of single loci and organisms (locus-tree-species-tree incongruence or discordance) can also arise from biological factors (Steenwyk et al. 2023b). These factors include reticulate evolutionary processes like hybridization/introgression, the interbreeding between distinct lineages, which can disrupt inferences of both the timing and pattern of historical divergences (Rieseberg et al. 2007; Racimo et al. 2015; Barley et al. 2018; Gonçalves et al. 2018; Gonçalves and Gonçalves 2019; Steenwyk et al. 2020b; Li et al. 2022b; Suvorov et al. 2022; Tiley et al. 2023). Hybridization/introgression has been documented in plants, algae, fungi, and a variety of animals among other lineages (Rieseberg et al. 2007; Neafsey et al. 2010; Stukenbrock 2016; Edelman et al. 2019; Edger et al. 2019; Sousa et al. 2019; Mixão and Gabaldón 2020; Steenwyk et al. 2020b; Bringloe et al. 2021; Wang et al. 2022). Among humans, loci originating from admixture events between early humans and Neanderthals have been associated with adaptation, phenotypic variation, and disease risk, including for severe COVID-19 (Sankararaman et al. 2016; Simonti et al. 2016; Dannemann and Kelso 2017; Dannemann et al. 2017; Zeberg and Pääbo 2020). Hybridization can also result in allopolyploid wherein the genome of the hybrid organism encodes (nearly) the entire genome of both parents. Allopolyploidy has been observed in numerous plants, fungi, and a few vertebrates (Ozkan et al. 2001; Session et al. 2016; Edger et al. 2019; Steenwyk et al. 2020b; Chen et al. 2022; Session and Rokhsar 2023). Genome evolution in allopolyploids can be rapid, marked by pronounced loss of genetic material (Ozkan et al. 2001), or be relatively stable (Steenwyk et al. 2020b). In either case, introgression/hybridization results in novel combinations of genes and genetic backgrounds that can result in distinct phenotypic profiles (Steenwyk et al. 2020b; Bautista et al. 2021).

Another mode of reticulate evolution, horizontal gene transfer—the transfer of genetic material without sexual reproduction—also causes locus-tree-species-tree incongruence and has been documented in diverse organisms, especially among prokaryotes and archaea (Galtier 2007; Yue et al. 2012; Van Etten and Bhattacharya 2020; Arnold et al. 2022; Gonçalves and Gonçalves 2022; Gophna and Altman-Price 2022; Li et al. 2022b; Steenwyk et al. 2023a). Horizontal gene transfer can be advantageous, endowing recipient organisms with potentially novel functionality (Gonçalves and Gonçalves 2019; Kominek et al. 2019; Li et al. 2022b). In certain cases, complex patterns of horizontal gene transfer or lateral acquisition of entire gene clusters can occur, resulting in new metabolic capabilities such as alcohol fermentation and the biosynthesis of thiamine and siderophores (Gonçalves et al. 2018; Gonçalves and Gonçalves 2019; Kominek et al. 2019). Horizontally acquired genes can also facilitate adaptation to extreme environments. For example, ice-binding proteins originating from bacteria are thought to contribute to algal adaptation to Arctic environments (Dorrell et al. 2023) and mercuric reductase, an enzyme responsible for converting mercury to a less toxic form, was transferred from bacteria to extremophilic algae commonly isolated from environments with a high mercury concentration (Schönknecht et al. 2013). Among protists, approximately 1% of gene repertoires are estimated to have been horizontally acquired (Van Etten and Bhattacharya 2020). These observations emphasize the significance of horizontal gene transfer as a major evolutionary mode.

From data acquisition to divergence time estimation, this review outlines notable steps for phylogenomic inference, discusses potential analytical sources of error, and explores methodologies for detecting reticulate evolution. By rigorously considering sources of error, researchers can disentangle analytical and biological factors contributing to locus-tree-species-tree incongruence, thereby enhancing the accuracy of phylogenomic inference, and facilitating the characterization of reticulate evolutionary processes.

A Workflow for Robust Phylogenomic Inference

Data Acquisition and Preparation

The first step of phylogenomic tree inference involves acquiring high-quality genomic/transcriptomic data from the target taxa (Figure 1A) (Cheon et al. 2020; Kapli et al. 2020). The abundance of available genomic/transcriptomic resources, accessible through online repositories like the National Center for Biotechnology Information (NCBI), enables researchers to address worthwhile questions without generating new data. Nevertheless, expanded taxon sampling through novel 'omic resources can improve phylogenomic inference, help address contentious issues, and

provide a valuable resource to the community (Pollock et al. 2002; Dunn et al. 2008; Wiens and Tiu 2012; Shen et al. 2018; Blaxter et al. 2022; Steenwyk et al. 2023b). To improve taxon sampling, researchers can apply a variety of methods to the taxa of interest, including targeted sequencing, reduced representation sequencing, transcriptomics, or whole-genome sequencing (Dunn et al. 2008; Peterson et al. 2012; One Thousand Plant Transcriptomes Initiative 2019; Hale et al. 2020). Applying standard methods for sequence data processing and quality control, such as removing low-quality data and contaminated sequences, is crucial (Zhou and Rokas 2014). Numerous tools exist for removing low-quality sequencing data, such as Trimmomatic and fastp (Bolger et al. 2014; Chen et al. 2018). Contaminated sequences can be purged from datasets using dinucleotide odds ratios or coverage-versus-lengths plots (Schmieder and Edwards 2011; Douglass et al. 2019).

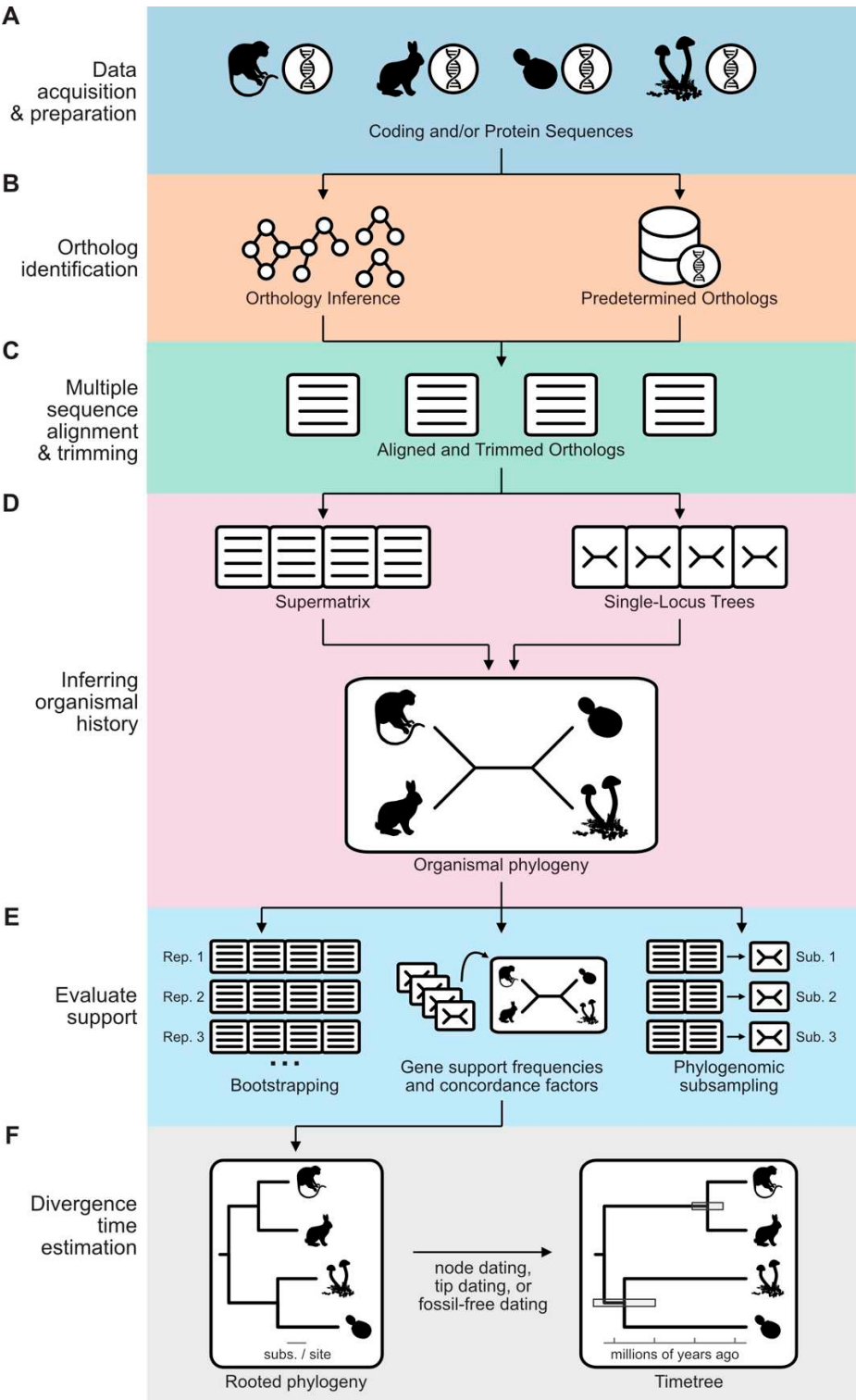


Figure 1. A workflow for phylogenomic inference. (A) The first step in a phylogenomic workflow is data acquisition and preparation. This often entails identifying gene boundaries in genome assemblies or assembling transcripts in transcriptomes. (B) The next step is to identify orthologs using (left) *de novo* approaches—for example, all-by-all sequence similarity calculations followed by graph-based clustering—or (right) from predetermined sets of orthologs. (C) Orthologous groups of genes suitable for phylogenomics (i.e., one-to-one orthologs and SNAP-OGs) are subsequently aligned and trimmed. (D) The resulting multiple sequence alignments can be (left) concatenated into a supermatrix or (right) collections of single-locus phylogenies can be used in a coalescence-based approach. (E) Support for the inferred phylogeny can be evaluated using multiple approaches, such as bootstrap statistics, gene support frequencies / concordance factors, and phylogenomic subsampling. (F) Divergence time estimates can be inferred using node dating, tip dating, or fossil-free analyses. Branch lengths represent substitutions per site in the phylogeny on the left and time on the right. Grey boxes in the timetree represent confidence intervals. Silhouette images were obtained from PhyloPic (<https://www.phylopic.org/>); credit goes to their respective contributors.

For the sake of simplicity, the following sections of this review focus on phylogenomic inference using alignments of protein-coding sequences. However, it is important to note that different data sources require tailored approaches to facilitate the subsequent steps. For instance, when working with genomic data, an additional step of gene boundary predictions is necessary (Stanke et al. 2006; Brůna et al. 2021). Similarly, SNP-based phylogenomics, which is also suitable for detecting hybridization/introgression, or synteny data, an emerging phylogenomic marker of collinearity between two or more genomes (Rokas and Holland 2000; Bringloe et al. 2021; Parey et al. 2023; Schultz et al. 2023; Steenwyk and King 2023), requires unique considerations not covered here.

Gene Orthology Determination

Genes that arise from speciation events in a shared ancestor are termed orthologous genes and serve as the foundation for phylogenomic analysis (Gabaldón and Koonin 2013). Relationships among orthologous genes can be described as one of three categories: one-to-one, one-to-many, and many-to-many (Fernández et al. 2019). Considering two genomes, one-to-one orthologs are encoded in each genome once; one-to-many orthologs are encoded in one genome once and the other multiple times; and many-to-many orthologs refer to a gene with multiple copies in each genome. Although methods have been developed to infer organismal histories using one-to-many and many-to-many orthologs (Zhang et al. 2020; Smith et al. 2022), most phylogenomic analyses rely on one-to-one orthologous genes and will be the focus of the present article. Moreover, single-copy orthologs are often the substrate of many downstream molecular evolution analyses such as measures of selection, relative evolutionary rates, and gene-gene coevolution (Chikina et al. 2016; Kowalczyk et al. 2019; Steenwyk et al. 2021, 2022d; Álvarez-Carretero et al. 2023).

Inferring orthology can be accomplished through two main approaches: global and targeted inference. A commonly utilized framework for global orthology inference involves all-to-all sequence similarity calculations followed by clustering (Figure 1B). Several tools implement this approach, including OrthoFinder, OrthoMCL, or MMseqs2 (Li et al. 2003; Emms and Kelly 2015, 2019; Steinegger and Söding 2018). To perform all-to-all sequence similarity calculations, software like BLAST, DIAMOND, or MMseqs2 can be employed (Camacho et al. 2009; Steinegger and Söding 2018; Buchfink et al. 2023). Sequence similarity scores can be influenced by biases arising from differences in sequence lengths. Correcting this analytical error can enhance the accuracy of orthology inference (Emms and Kelly 2015). Subsequently, graph-based clustering methods, such as the *Markov Cluster Process* (Van Dongen 2008), are applied to categorize genes into distinct orthologous groups of genes. During Markov clustering, the inflation value will impact the granularity of orthologous groups of genes (Brohée and Van Helden 2006). Relaxed inflation values may combine orthologous groups of genes whereas stringent inflation values may oversplit them. Oversplitting can also stem from the inability to detect remote homology among orthologs encoded in distantly related species or rapidly evolving taxa (Weisman et al. 2020). There is no golden rule for an appropriate inflation value; the

best value will likely be dataset-dependent. The resulting clusters of orthologous genes can serve as proxies for gene families.

In phylogenomic inference, single-copy gene families, orthologous genes encoded in each taxon once, are preferred as they (presumably) have not experienced duplication or loss (Li et al. 2017). Organismal histories are often inferred using hundreds to thousands of loci to avoid inadequate sampling of gene loci, a known source of error while inferring organismal histories (Rokas et al. 2003; Steenwyk et al. 2023b). However, as the number of taxa and evolutionary distances between them increases, the number of single-copy orthologous genes tends to decrease (Emms and Kelly 2018). For instance, in a dataset comprising 42 plants with complex evolutionary histories involving dynamic gene gain and loss events (Clark and Donoghue 2018), no single-copy orthologous genes were identified (Emms and Kelly 2018). In such cases, no gene families are fit for standard phylogenomic analyses.

To overcome this challenge, tree decomposition algorithms that partition multi-copy gene families into subgroups of single-copy orthologous genes and pruning algorithms that remove species-specific paralogs (i.e., duplicated genes observed in a single species) can be useful (Kocot et al. 2013; Willson et al. 2022). The software OrthoSNAP combines tree splitting and pruning to identify single-copy orthologs nested within larger gene families (Steenwyk et al. 2022c). This process is analogous to snapping branches in a tree; thus, the resulting subgroups of single-copy orthologs are termed SNAP-OGs (*splitting and pruning*). Single-copy gene families and SNAP-OGs are statistically indistinguishable across diverse measures of information content (Steenwyk et al. 2022c), indicating their suitability for phylogenomic analyses.

While global orthology inference offers valuable insights, it often demands significant computational resources. An alternative approach involves the identification of pre-determined single-copy genes (Figure 1B). One method is the BUSCO pipeline (Benchmarking Universal Single-Copy Ortholog), which identifies near-universally single-copy orthologous genes within a genome or transcriptome (Waterhouse et al. 2018). BUSCO uses predefined near-universally single-copy orthologs from OrthoDB (Kriventseva et al. 2019). This analysis also serves as a useful quality control measure for assessing assembly gene content completeness (Waterhouse et al. 2018). When researchers prefer to use a custom set of predefined single-copy orthologous genes, orthofisher can be used (Steenwyk and Rokas 2021). Additionally, various sequence similarity search tools such as HMMER, BLAST, and DIAMOND can be employed to support the identification of single-copy orthologous genes (Camacho et al. 2009; Eddy 2011; Buchfink et al. 2023). If the putative orthologs identified using this approach are multi-copy, the same splitting and pruning procedure in OrthoSNAP can identify SNAP-OGs (Steenwyk et al. 2022c). If the dataset contains a small number of genes, single-locus phylogenies can also be visually inspected, and paralogous sequences can be manually removed.

Multiple Sequence Alignment and Trimming

Once a curated set of phylogenomic markers has been obtained, the next step is multiple sequence alignment (Figure 1C), which aims to determine the site-wise homology across a group of sequences. Progressive alignment is a widely used strategy for multiple sequence alignment and involves an iterative pairwise alignment approach (Feng and Doolittle 1987). Over time, advancements have incorporated additional biological information to improve multiple sequence alignment. For instance, PRANK utilizes a phylogeny-aware approach, while 3DCoffee incorporates protein structure information (O'Sullivan et al. 2004; Löytynoja and Goldman 2005).

Several databases, such as BALiBASE and PREFAB (Thompson et al. 1999; Edgar 2004), have been developed to evaluate the effectiveness of these approaches. By using gold standard alignments from such databases or alignments generated by simulation software like INDELible (Fletcher and Yang 2009), the accuracy of multiple sequence alignment algorithms can be assessed using metrics like sum-of-pairs (pairwise alignment accuracy) and column scores (column-wise alignment accuracy) (Thompson et al. 1999; Steenwyk et al. 2021). Additionally, secondary structure predictions derived from the aligned sequences can be compared to known secondary structures to evaluate

alignment accuracy (Sievers and Higgins 2020). Despite the establishment of standards, variable outcomes among benchmarking studies indicate that there is currently no universally superior algorithm for multiple sequence alignment (Raghava et al. 2003; Wang et al. 2018b; Sievers and Higgins 2020). Generating study-specific simulated alignments based on expected phylogenetic diversity can help determine which alignment strategy performs well for a given dataset.

Multiple sequence alignments are commonly subjected to trimming, which involves the removal of specific sites or blocks of sites within alignments. Traditionally, sites are trimmed because they are highly variable, which is thought to stem from erroneously inferred site homology, or are devoid of phylogenetic informativeness due to saturation by multiple substitutions (Talavera and Castresana 2007; Criscuolo and Gribaldo 2010). However, comparing multiple sequence alignment trimming algorithms has revealed that these methods often lead to poorer phylogenetic inference for single-gene analyses (Tan et al. 2015). This finding suggests that current trimming methods may inadvertently remove phylogenetically informative sites. In contrast, the software ClipKIT uses an alternative approach where informative sites (e.g., parsimony informative sites) are retained and all others removed, outperforming other algorithms under diverse evolutionary scenarios (Steenwyk et al. 2020a). A secondary advantage of trimming multiple sequence alignments is that shorter alignments require fewer computational resources, an important consideration for environmentally conscious computing practices (Kumar 2022; Steenwyk et al. 2023b). To address other types of errors in multiple sequence alignments, tools such as TAPER, Divviers, or MACSE can be employed (Ranwez et al. 2018; Ali et al. 2019; Zhang et al. 2021).

Model Selection

In molecular phylogenetics, the identification of an optimal model of sequence evolution occurs after alignment trimming and before tree inference. Although substitution models may, at times, not have a large impact on phylogenetic inference, model misspecification is a well-appreciated source of error (Shen et al. 2018; Abadi et al. 2019; Betancur-R. et al. 2019; Cao et al. 2022; Steenwyk et al. 2023b). Accordingly, substitution models have been a major research focus. Two main categories of substitution models exist: site homogeneous and site heterogeneous. Site homogeneous models employ the same parameters of character frequencies and substitution rates across the entire alignment (Galtier and Gouy 1995). Site heterogeneous mixture models allow for different character frequencies at each site (Lartillot et al. 2007; Si Quang et al. 2008). Site homogeneous models may be more prone to model misspecification, which can be overcome by generating alignment-specific time-reversible and time-non-reversible substitution models (Rodríguez-Ezpeleta et al. 2007; Minh et al. 2021). Site heterogeneous models have a larger parameter space and are robust to over-parameterization but require substantial computational resources (Baños et al. 2022). To accelerate computation, site heterogeneous models can be approximated using posterior mean site frequency profiles (Wang et al. 2018a).

Software tools that perform model testing are often integrated into phylogenetic tree inference software. For example, ModelFinder is included in the IQ-TREE toolkit (Kalyanamoorthy et al. 2017; Minh et al. 2020b). These tools employ statistical frameworks—such as likelihood ratio testing and calculations of information criterion—to facilitate assessing model fit (Darriba et al. 2012, 2020; Kalyanamoorthy et al. 2017).

Gene Alignment Concatenation or Coalescence-based Methods for Inferring Organismal History

There are two commonly used approaches for organismal tree inference from genome-scale datasets: gene alignment concatenation (or simply concatenation) and coalescence (Figure 1D) (Rokas et al. 2003; Liu et al. 2009a; Steenwyk et al. 2023b). Each employs different theoretical and statistical frameworks. Concatenation primarily utilizes maximum likelihood or Bayesian statistics. Coalescence relies on the multi-species coalescent model, which accounts for discordance between single loci and species-trees stemming from processes like incomplete lineage sorting.

The concatenation approach combines the aligned and trimmed orthologous genes into a single matrix. SequenceMatrix, FASconCAT-G, and PhyKIT can concatenate multiple sequence alignments

(Vaidya et al. 2011; Kück and Longo 2014; Steenwyk et al. 2021). The resulting supermatrix can then be analyzed using various software that employs maximum likelihood statistical frameworks, such as RAxML-NG and IQ-TREE (Kozlov et al. 2019; Minh et al. 2020b), or Bayesian frameworks, such as MrBayes, BEAST, PhyloBayes, and RevBayes (Huelsenbeck and Ronquist 2001; Ronquist et al. 2012; Höhna et al. 2016; Bouckaert et al. 2019; Lartillot 2020). The RevBayes software allows users to customize how parameters, priors, and data are directly connected in the model, either in the Rev language or in the R programming language (Höhna et al. 2017; Charpentier and Wright 2022). In concatenation approaches, a single substitution model can be applied to the entire alignment, or the supermatrix can be partitioned, with different substitution models applied to each partition (Kainer and Lanfear 2015).

Comparative studies have systematically evaluated the performance of tools implementing different flavors of heuristics for maximum likelihood calculations and tree rearrangement methods, revealing variation in their performance (Zhou et al. 2018). This observation suggests that software choice is an important consideration. More broadly, software choice is encompassed in a newly described source of error: treatment errors, the negative impact of software choice, or data treatment upon stochastic and systematic errors (Steenwyk et al. 2023b).

In coalescent-based methods, two different approaches—one-step and two-step methods—are commonly employed for phylogenomic inference. In one-step approaches, single-locus phylogenies are estimated simultaneously with the species tree. This methodology is implemented in software such as BEST, StarBeast, and bpp (Liu et al. 2008; Yang and Rannala 2010; Douglas et al. 2022). Similarly, SVDquartets derives species trees from aligned, unlinked single nucleotide polymorphism data using the coalescent framework (Chifman and Kubatko 2014). In two-step approaches, as implemented in STAR and ASTRAL (Liu et al. 2009b; Zhang et al. 2018), involve first inferring individual single-locus phylogenies and then constructing a summary tree from the collection of single-locus phylogenies. One-step approaches are computationally expensive and may be difficult to apply to large phylogenomic datasets. Two-step approaches may be susceptible to errors resulting from inaccurate single-locus tree inference (Degnan and Rosenberg 2009). Collapsing poorly supported branches before applying summary tree methods can help mitigate the impact of uncertainty in single-locus phylogenies (Steenwyk et al. 2023b).

Although both concatenation and coalescence methods are widely used in species tree inference, there is currently no definitive guidance on when to prefer one method. However, it has been consistently observed that concatenation and coalescence can yield different topologies and varying levels of gene-wise support (Gatesy et al. 2017; Shen et al. 2018, 2021; Steenwyk et al. 2019b, 2023b; Li et al. 2020). The performance of these methods may vary depending on factors such as the data characteristics—the number and length of sampled loci, completeness of taxon sampling, the extent of reticulate evolution or incomplete lineage sorting among taxa, and the evolutionary diversity of the taxa, for example. A comprehensive study is needed to elucidate evolutionary scenarios where one method outperforms the other.

Examining Bipartition Support

Assessing the confidence in the resulting phylogenetic tree is crucial for identifying unstable bipartitions and gaining insights into lineages with reticulate evolutionary histories (Steenwyk et al. 2023b). Various methods have been developed to evaluate the support for different branches in a phylogeny and identify poorly supported bipartitions.

Bootstrapping is one of the earliest and most widely used methods for evaluating confidence in phylogenetic tree inference (Figure 1E). This statistical procedure involves resampling sites from an alignment with replacement to create multiple replicates of the original data (Felsenstein 1985). Each replicate is then used to infer a phylogenetic tree, and the frequency of observing a particular branch among the bootstrap replicates quantifies the support for that clade. Bootstrapping is typically performed with many replicates, such as 100 or more, to obtain robust support values. However, bootstrapping can be computationally intensive and time-consuming.

As phylogenomic datasets have become larger, alternatives have been introduced to expedite computation. For example, RAxML implements a rapid bootstrap approach (Stamatakis et al. 2008) and IQ-TREE employs an ultrafast bootstrap approximation method (Hoang et al. 2018). The "bag of little bootstraps" can also reduce the computational burden of long alignments by combining resampling with subsampling (Sharma and Kumar 2021). The gradual "transfer" distance method has also been proposed for bootstrap inference in phylogenies with hundreds to thousands of taxa (Lemoine et al. 2018).

Other methods to evaluate confidence use branching patterns observed among single-gene phylogenies. These include single-locus or -site support frequencies (also known as concordance factors) and quartet mapping (Figure 1E), as implemented in IQ-TREE, ASTRAL, and PhyKIT (Zhang et al. 2018; Minh et al. 2020b; Steenwyk et al. 2021). Calculations of internode certainty, an entropy-based measure of branch support, can help identify when alternative topologies are well supported among a collection of single-locus or bootstrap trees (Salichos and Rokas 2013; Salichos et al. 2014; Kobert et al. 2016; Zhou et al. 2020).

Phylogenomic subsampling is another approach for exploring tree space and identifying unstable bipartitions (Figure 1E) (Edwards 2016; Steenwyk et al. 2023b). In phylogenomic subsampling, subsamples of loci in a full phylogenomic data matrix are used to re-infer organismal histories. Bipartitions incongruent between the full and subsampled matrices warrant further investigation and can be considered unstable. Subsampling is typically done by selecting, for example, half of the loci in a full data matrix with the desirable feature associated with phylogenetic signal. These features—such as alignment length, average bipartition support, treeness divided by relative composition variability, and the number of parsimony and variable sites, among others (Phillips and Penny 2003; Shen et al. 2016, 2018; Steenwyk et al. 2019b, 2020a, 2021, 2022b; Mongiardino Koch 2021; McCarthy et al. 2023; Redmond et al. 2023)—aim to capture the information content of a locus.

Time-calibration of inferred phylogenetic divergences

Divergence times among branches in a phylogenomic analysis can be estimated by using fossils, mutation rates, or other temporal evidence to calibrate a molecular clock model (Ho and Phillips 2009; Dos Reis et al. 2016, 2018; Tiley et al. 2020). This procedure converts the relative divergences of molecular substitution rates to absolute time, often in units of thousands or millions of years ago. The resulting time-calibrated phylogenies, which are often referred to as 'timetrees' or 'chronograms,' differ from uncalibrated phylogenies ('phylograms') in that the former is directly comparable to other types of time-scaled data. Timetrees can be used to investigate causal eco-evolutionary dynamics relative to a broad array of independent evidence; for example, past changes in global temperature versus rates of lineage divergence (Oliveros et al. 2019; Schubert et al. 2019; S. Meseguer and Condamine 2020; Feijó et al. 2022); co-diversification among taxa (Sabrina Pankey et al. 2022; Nelsen et al. 2023); and rates of speciation in related clades (Harvey et al. 2020; Upham et al. 2021).

Approaches to estimating divergence times can be divided into node dating, tip dating, and fossil-free dating. Node dating places temporal constraints (i.e., calibrations) on a bifurcating internal node of a phylogeny, whereas tip dating places calibrations on terminal taxa that existed at some time in the past (Ho and Phillips 2009; Heath et al. 2014). The ages of serially sampled taxa—usually fossils or viruses and other microbes (Stadler and Yang 2013)—are the most reliable data for calibrating divergence times in phylogenomic datasets. Fossils and their associated ages (typically a confidence interval for the age of rock formations above and below a fossil discovery) can calibrate divergence times at either nodes or tips, typically using a probability distribution to incorporate age uncertainty (Ho and Phillips 2009; Stadler and Yang 2013). A fossil's phylogenetic position relative to living members of a given clade must be inferred or assumed based on other data for that fossil to serve as a time calibration (Parham et al. 2012). Viruses and other microbes evolve rapidly enough that samples collected in the last few decades offer valuable tip calibrations, analogous to the role of fossils in longer-lived mammals or plants (Volz et al. 2013; Andréoletti et al. 2022). The resulting 'phylodynamic' analyses help expose the population-dynamic processes that generate the

phylogenetic patterns inferred from phylogenomic datasets. Phylodynamics is a promising integration of phylo-centric fields, enabling advances in cell biology, epidemiology, and macroevolution (Stadler et al. 2021; Andréoletti et al. 2022).

Clock models are used to extrapolate species divergence times from calibrated nodes. Strict clock models assume a fixed mutation rate in all branches, which is often violated when comparing more distant relatives (e.g., the 2% per million years rate used for bird mitochondrial genes; (Ho 2007)). However, strict clocks may lack biological realism. To address this, the strict clock assumption can be relaxed, such as in autocorrelated clock models wherein closely related branches have similar mutation rates or, in uncorrelated models, each branch is given an independent rate (Drummond et al. 2006; Lepage et al. 2007; Steenwyk and Rokas 2023). Relaxed clocks allow greater flexibility for handling the inherent molecular-rate variation among lineages, and thus are in wide use today for all types of time-calibration approaches, including fossil dating and multi-species coalescent approaches (Dos Reis et al. 2016, 2018; Flouri et al. 2022). These latter approaches enable the simultaneous estimation of species divergence times and ancestral population sizes using phylogenomic data and can be quite accurate when mutation rates are known from pedigrees (Tiley et al. 2020).

What if no fossils or other serial samples are available for a particular taxon? Two main options exist to calibrate divergences: use a fixed, strict clock model to project estimates back from tips or secondary calibrations. Secondary calibrations use previously estimated (from primary fossil or rate calibrations) divergence times of a sister taxon or outgroup to calibrate an internal or root node in the clade of interest (Shaul and Graur 2002). However, caution is required to avoid specifying overly precise secondary calibrations (Schenk 2016).

Choosing which software to use for divergence-time estimation involves a trade-off between available compute resources and the desired level of biological realism. At one extreme, the most realistic models (e.g., BPP and StarBEAST (Flouri et al. 2018; Douglas et al. 2022)) will perform Bayesian inference to estimate multi-species coalescent parameters across thousands of gene genealogies, considering multiple rate priors, and integrating across both phylogenetic and temporal uncertainty to yield a posterior distribution of time-scaled trees. However, these 'full methods' do not scale to large numbers of taxa or distant relatives (Tiley et al. 2020; Jiao et al. 2021).

At the other extreme, concatenated sequence data are used step-wise to first estimate the phylogenetic tree topology in units of substitutions/site, and then calibrations are applied in a second step of divergence-time estimation. Step-wise methods most commonly use maximum-likelihood (e.g., r8s, treePL, RelTime; (Sanderson 2003; Smith and O'Meara 2012; Tao et al. 2020)), but can also be implemented using Bayesian inference in programs like BEAST or MrBayes, which often requires fixing the tree topology. Midway between these extremes is the use of concatenated sequence data to perform simultaneous estimation of topology and divergence times, generally as implemented in a Bayesian framework (e.g., BEAST, MCMCtree, MrBayes, PhyloBayes, RevBayes). This latter approach has been implemented in large datasets (e.g., 800 taxa by 40,000 sites; (Upham et al. 2019)), and continues to be aided by GPU-based computing libraries (Ayres et al. 2019). Strategies for setting priors can impact divergence time estimation and are an important consideration (Barba-Montoya et al. 2017).

During divergence time estimation, a range of dates are typically plausible under the experimental conditions. Thus, divergence times are depicted as a confidence interval. Divergence times can also be inferred using a bootstrapping approach for intractably large datasets. Overall, the choices of node, tip, or fossil-free dating; strict or relaxed clocks; and concatenation or multi-species coalescent methods depend upon the question of interest, available molecular and morphological data, and prevalence of locus-tree-species-tree incongruence.

Detecting Reticulate Evolution

Reticulate evolutionary processes, such as hybridization/introgression and horizontal gene transfer, result in loci with evolutionary histories that differ from organismal history (Dobzhansky 1982; Abbott et al. 2013; Steenwyk et al. 2023b). There is a spectrum of outcomes for hybridization

ranging from adaptive changes due to ecological selection, or compromised viability or fertility due to hybrid incompatibilities (Racimo et al. 2015; Moran et al. 2021). Similarly, horizontal gene transfer endows recipient organisms with novel genetic material and can be adaptive or not (Schönknecht et al. 2013; Gonçalves and Gonçalves 2019; Arnold et al. 2022; Gophna and Altman-Price 2022; Li et al. 2022b; Dorrell et al. 2023). The impact of horizontal gene transfer is well appreciated among prokaryotes and archaea but also occurs in the eukaryotes (Shen et al. 2018; Gonçalves and Gonçalves 2019; Lartillot 2020; Arnold et al. 2022; Gophna and Altman-Price 2022; Li et al. 2022b).

Hybridization/introgression

Hybridization/introgression can be adaptive or not, either expanding the ecological repertoire of organisms as exemplified by sunflower adaptation to novel environments or leading to the reabsorption of incipient species (Mallet 2005, 2008; Racimo et al. 2015; Buck et al. 2023). Hybrid progeny can have improved growth and reproductive success or conversely be sterile (Zanewich et al. 2018; Qiao et al. 2019; Allen et al. 2020; Adavoudi and Pilot 2021). Hybridization has been observed in microbial pathogens and thus may contribute to higher or lower organismal virulence (Lin et al. 2009; Depotter et al. 2016; Mixão and Gabaldón 2020). Hybridization can also result in neutral or negative fitness outcomes (Allen et al. 2020; Adavoudi and Pilot 2021).

The identification of hybridization/introgression events can be accomplished by utilizing phylogenetic and comparative genomic methodologies (Scannell et al. 2006; Marcet-Houben and Gabaldón 2015; Ortiz-Merino et al. 2017; Mixão and Gabaldón 2020; Steenwyk et al. 2020b, 2023a). Among phylogenetic approaches, it is crucial to discriminate between incongruences among single-locus phylogenies stemming from hybridization instead of incomplete lineage sorting, the random sorting of ancestral alleles that can, at times, result in locus-tree-species-tree incongruence (Yu et al. 2013). To make this distinction, two nearly equally supported topologies among multiple single-locus phylogenies indicate hybridization, especially if hybridization was a recent event. Conversely, in the case of incomplete lineage sorting, the alternative topology is expected to be observed less frequently, especially among more recent divergences (Steenwyk et al. 2019b). The expected degree of incongruence stemming from incomplete lineage sorting can be modeled using the multispecies coalescent model. Deviations from that model may indicate a hybridization event (Degnan and Rosenberg 2009). Calculating internode certainty can also help quantify the degree of support for the two most prevalent topologies (Salichos and Rokas 2013; Kobert et al. 2016).

Many methods have been developed to detect hybridization events more directly using phylogenetic trees (Hibbins and Hahn 2022). One pioneering approach is the utilization of the D-statistic, also known as the ABBA-BABA test, which employs biallelic site patterns within a phylogenetic framework (Figure 2) (Green et al. 2010). Specifically, the ABBA-BABA test examines asymmetric support between ABBA and BABA patterns at biallelic sites, which suggests an introgression/hybridization event; in contrast, equal proportions of ABBA and BABA site patterns would suggest no introgression/hybridization. Leveraging genome-scale data, the ABBA-BABA test accurately quantifies introgression across a wide parameter space (Zheng and Janke 2018).

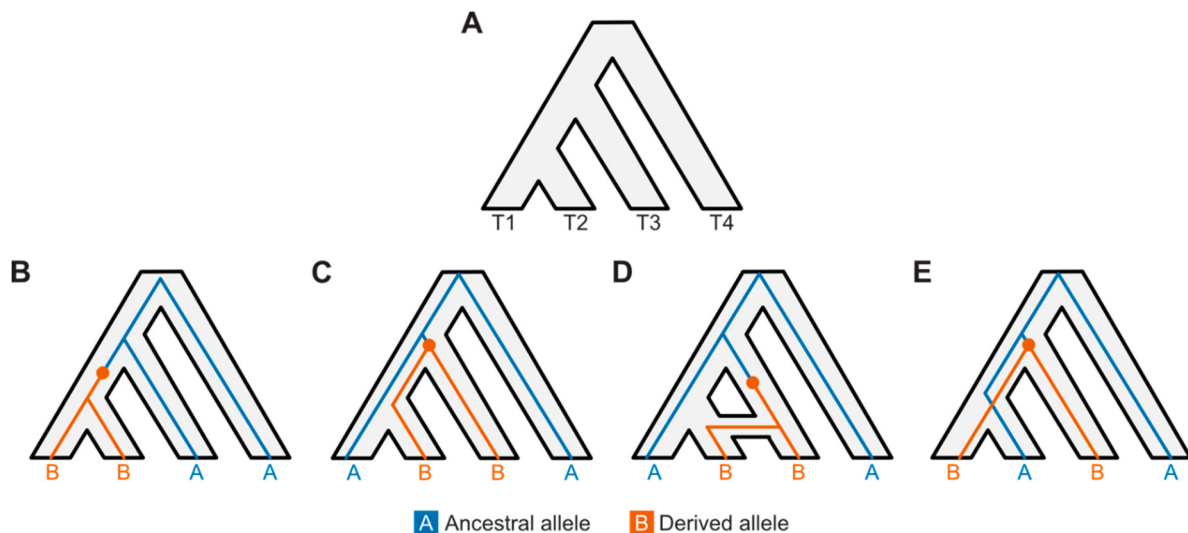


Figure 2. The ABBA-BABA test for detecting introgression/hybridization. To detect introgression/hybridization in the (A) four-taxon case (represented as T1 through T4 where T4 is the outgroup taxa), the D-statistic or ABBA-BABA test can be used. (B) The orange dot represents a mutation from the ancestral allele 'A' (blue) to a derived allele 'B' (orange). The BBAA pattern, which is not directly accounted for in the ABBA-BABA test, is a biallelic site that follows the organismal phylogeny. Asymmetric patterns of ABBA and BABA biallelic site patterns suggest the occurrence of an introgression/hybridization event. The ABBA pattern can arise from (C) incomplete lineage sorting or (D) introgression/hybridization, whereas the (E) BABA pattern can arise from incomplete lineage sorting; thus, unequal frequencies of ABBA and BABA patterns are suggestive of introgression/hybridization.

Discerning the direction of introgression poses a greater challenge (Martin et al. 2015). To address this, the D_P -statistic, an extension of the ABBA-BABA test, incorporates patterns observed across all variable sites, thus enabling the determination of gene flow directionality (Hamlin et al. 2020). In the context of a symmetric five-taxon phylogeny, the D_{FOIL} statistic, another extension of the ABBA-BABA test built on the partitioned D-statistic (Eaton and Ree 2013), explicitly examines all potential introgression events, facilitating precise and sensitive detection of the extent and direction of introgression (Pease and Hahn 2015). Several other methodologies, including the F_4 -ratio, f_d , the f_3 -statistic, and algorithms implemented in the HyDe software, have been developed to identify introgression events (Green et al. 2010; Reich et al. 2012; Martin et al. 2015; Blischak et al. 2018; Jacobs et al. 2018). Note, these methods examine rates of tree incongruences and are distinct from F-statistics, such as F_{st} , which are related to population genetic analysis.

In the context of allopolyploid hybrids, where the genome of the hybrid organism contains (nearly) the complete genetic complement of both parental genomes and, therefore, two or more copies of most genes, supplementary methods play a crucial role in detecting and unraveling the evolution of a hybrid genome. Ancient allopolyploid events can be identified by a burst of gene duplications (Chain et al. 2011; Marcet-Houben and Gabaldón 2015; Session et al. 2016), but would benefit from other lines of evidence such as synteny information. In allopolyploids, it may be helpful to determine the parent of origin for each gene in a hybrid genome. This can be accomplished by evaluating phylogenetic distances between the parental species and the gene copies encoded in the hybrid (Steenwyk et al. 2023a). Additionally, K_s plots, which visualize sequence divergence between loci in the hybrid taxon and their closest homolog in a parental species, prove valuable in assigning parent-of-origin for loci within a hybrid genome (Ortiz-Merino et al. 2017; Steenwyk et al. 2020b).

Horizontal gene transfer

Horizontal gene transfer, or lateral gene transfer, is the acquisition of genetic material from other organisms without sexual reproduction (Mallet 2005). Horizontal gene transfer is considered a major mode of evolution in prokaryotes and archaea wherein the mechanism of transfer—conjugation,

transduction, and transformation—is better understood (Galtier 2007; Arnold et al. 2022; Gonçalves and Gonçalves 2022; Gophna and Altman-Price 2022). In recent years, horizontal gene transfer in eukaryotes is better appreciated (Coelho et al. 2013; Gonçalves et al. 2018; Husnik and McCutcheon 2018; Shen et al. 2018; Zhou et al. 2018; Gonçalves and Gonçalves 2019; Van Etten and Bhattacharya 2020; Irwin et al. 2021). One hypothesized transfer mechanism among microeukaryotes serves as the foundation for the “you are what you eat” hypothesis, which posits that organisms that phagocytose prey may be more susceptible to rare and accidental nucleic acid integration from prey DNA (Sibbald et al. 2020). In support of this hypothesis, recent analyses found that carnivorous mammal genomes contain a greater prevalence of DNA-based transposable elements than do herbivores or omnivores, suggestive of origins via horizontal gene transfer from ingested prey or their viruses (Osmanski et al. 2023).

Horizontal gene transfer is also well appreciated in the context of mitochondrial and plastid organellogenesis. These organelles arose from the ancient assimilation and degradation of endosymbiotic bacteria in eukaryotic cells, during which many endosymbiont genes were lost or transferred to the host nucleus (Cenci et al. 2018; Ponce-Toledo et al. 2018; Sibbald and Archibald 2020). Charting endosymbiotic transfers has aided in estimating the order and mechanisms of higher-order plastid acquisitions and revealed that many plastid-related genes have highly chimeric origins (Stiller et al. 2014; Dorrell et al. 2017; Strassert et al. 2021). These observations may support scenarios of ancient serial plastid replacements before permanent plastid integration (Minge et al. 2010; Ponce-Toledo et al. 2018; Morozov and Galachyants 2019). Transferred genes have also provided evidence of relic plastids, illuminating patterns of plastid gain and loss (Cenci et al. 2018; Gawryluk et al. 2019). Some barriers to horizontal gene transfer are also known, such as multicellularity and incompatible genetic codes (Yue et al. 2012; Shen et al. 2018).

The methods employed for detecting horizontally acquired loci vary in precision and accuracy. Early methods relied on identifying deviations in gene sequence characteristics. In the case of very recent prokaryote-to-eukaryote horizontal gene transfer, detection could be achieved by observing genes that deviate in guanine-cytosine content, intron content, gene order, and codon usage across the host genome (Friedman and Ely 2012; Zhang et al. 2014b; Jaramillo et al. 2015; Gonçalves and Gonçalves 2022). In the phylogenomic era, these methods are often employed to support the identification of horizontal gene transfer events rather than serving as primary detection tools.

Another approach is to calculate the alien index—a score that compares the similarity between sequences within the target group and sequences from outgroup taxa (Gladyshev et al. 2008; Alexander et al. 2016)—of all genes in a host genome. Loci exhibiting alien indices indicative of potential horizontal gene transfer are then selected for further investigation through phylogenetic inference, the gold standard approach for horizontal gene transfer detection. Several software tools have been developed to calculate alien indices or similar metrics for assessing horizontal gene transfer. Examples include AvP, HGTector, and HGTphyloDetect (Zhu et al. 2014; Koutsovoulos et al. 2022; Yuan et al. 2023).

Phylogenetic trees that suggest horizontal gene transfer events are characterized by the confident placement of one or a few sequences within an unexpected taxonomic group (Figure 3). For instance, in the case of prokaryote-to-eukaryote horizontal gene transfer, sequences in a eukaryotic genome may be nested deep within a prokaryotic lineage (Coelho et al. 2013; Gonçalves et al. 2018; Husnik and McCutcheon 2018; Shen et al. 2018; Zhou et al. 2018; Gonçalves and Gonçalves 2019; Kominek et al. 2019; Van Etten and Bhattacharya 2020; Irwin et al. 2021; Li et al. 2022b). The evidence for horizontal gene transfer can be strengthened using topology tests, such as the Kishino-Hasegawa and Shimodaira-Hasegawa tests (Kishino and Hasegawa 1989; Shimodaira and Hasegawa 1999). These tests compare the likelihood of a phylogeny constrained to reflect a vertical evolutionary scenario (the null hypothesis) with the observed topology, reflecting the occurrence of horizontal gene transfer (the alternative hypothesis) (Gonçalves et al. 2018; Shen et al. 2018).

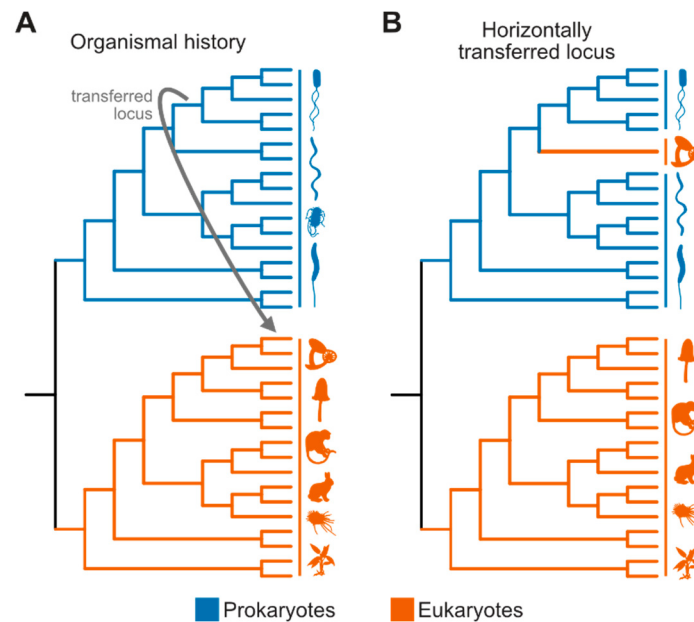


Figure 3. Phylogenetic signatures of a horizontal gene transfer event. To detect horizontal gene transfer, (A) organismal histories are compared to (B) single-locus phylogenies. Horizontal gene transfer is suggested when sequences are placed within an unexpected taxonomic group in single-locus phylogenies. Here, an example of prokaryote-to-eukaryote transfer is depicted wherein an organism from the fungal kingdom (orange) is monophyletic with prokaryotic sequences (blue). The horizontal gene transfer event is depicted as a grey arrow. Silhouette images were obtained from PhyloPic (<https://www.phylopic.org/>); credit goes to their respective contributors.

Although putatively transferred genes should be investigated using molecular phylogenetics, accurate phylogenomic species tree inference may not be crucial since it is widely accepted that the taxa involved are of distinct lineages; for example, fungi and bacteria are well known to comprise different clades. However, when examining horizontal gene transfer among more closely related taxa, robust species tree inference becomes essential (Ropars et al. 2015). The potential for shared ancestral variation or asymmetric gene loss patterns among closely related organisms increases the risk of erroneously inferring horizontal gene transfer events (i.e., type I error / false positive). Horizontally transferred loci may also complicate accurate inference of organismal histories; thus, purging phylogenomic data matrices from potentially horizontally acquired loci may be warranted. Lastly, to rule out putatively horizontally transferred genes stemming from contamination, this analysis is often restricted to well-assembled contigs, devoid of contamination signatures (Shen et al. 2018; Li et al. 2022b).

Phylogenetic Networks

Hybridization/introgression and horizontal gene transfer challenge the strictly bifurcating tree model. Phylogenetic networks can help explore organismal histories in the presence of these factors (Huson 1998; Lutteropp et al. 2022; Steenwyk et al. 2023b), and to large extent can be viewed as complementary to strictly bifurcating phylogenetic analyses (Blair and Ané 2020). Phylogenetic networks can be broadly categorized into reticulate networks, which capture reticulate evolutionary processes (Huson et al. 2005), and splits networks, which represent all splits in a collection of single-locus phylogenies (Huson 1998). Additionally, visualizing the density of phylogenetic trees can reveal common branching patterns across these collections, which can help identify regions of the tree that require further investigation. DensiTree is a tool commonly employed to visualize the density of phylogenetic tree topologies in a posterior distribution of trees (Bouckaert 2010). Gene-support frequencies (or quartet frequencies) and gene-wise phylogenetic signals offer additional insights to scrutinize and assess alternative branching patterns (Shen et al. 2017, 2021; Sayyari and

Mirarab 2018; Steenwyk et al. 2019b; Minh et al. 2020a). The methods discussed herein facilitate rigorous phylogenomic inference and detection of reticulate evolution in organismal histories.

Conclusion

Knowledge of organismal history underpins many evolutionary studies. This article outlines the many branching choices that investigators must make when performing phylogenomic inference, and some of the known pathways toward assembling robust methodological pipelines. The resulting analyses of genome-wide evolutionary signatures can, in turn, be used to enrich our understanding of life's history and the evolutionary process. Ameliorating analytical factors driving apparent incongruence will facilitate more accurate inference of organismal histories. True histories of reticulate evolution will deepen our understanding of the Tree of Life, adding a dimension of 'webiness.' Future computational improvements, reduced cost of high-quality genome sequencing, and algorithmic advances culminating in more chromosome-scale genome assemblies will pave the way for increasing larger datasets and greater elucidation of the Tree of Life. Such advances will also bring new challenges and opportunities. To encourage the continued investigation of biodiversity via phylogenomics, we hope this article provides helpful guidance for scientific investigation and pedagogy alike.

Acknowledgments: We thank Drs. Xing-Xing Shen and Yuaning Li for helpful discussion over the years. J.L.S. thanks Dr. Antonis Rokas, who taught him to embrace phylogenomic incongruence. Chat-GPT was used to edit portions of the manuscript; AI-generated text was further edited.

Funding: J.L.S. is a Howard Hughes Medical Institute Awardee of the Life Sciences Research Foundation. N.S.U. was supported by Arizona State University start-up funds. H.V. received support from the Australian Biological Resources Study (4-G046WSD) and the Australian Research Council (DP200101613).

Competing Interests: J.L.S. is a scientific advisor for WittGen Biotechnologies. J.L.S. is an advisor for ForensisGroup Inc.

References

- Abadi, S., D. Azouri, T. Pupko, and I. Mayrose. 2019. Model selection may not be a mandatory step for phylogeny reconstruction. *Nat Commun* 10:934.
- Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf, and D. Zinner. 2013. Hybridization and speciation. *J. Evol. Biol.* 26:229–246.
- Adavoudi, R., and M. Pilot. 2021. Consequences of Hybridization in Mammals: A Systematic Review. *Genes* 13:50.
- Alexander, W. G., J. H. Wisecaver, A. Rokas, and C. T. Hittinger. 2016. Horizontally acquired genes in early-diverging pathogenic fungi enable the use of host nucleosides and nucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 113:4116–4121.
- Ali, R. H., M. Bogusz, and S. Whelan. 2019. Identifying Clusters of High Confidence Homologies in Multiple Sequence Alignments. *Molecular Biology and Evolution* 36:2340–2351.
- Allen, R., H. Ryan, B. W. Davis, C. King, L. Frantz, E. Irving-Pease, R. Barnett, A. Linderholm, L. Loog, J. Haile, O. Lebrasseur, M. White, A. C. Kitchener, W. J. Murphy, and G. Larson. 2020. A mitochondrial genetic divergence proxy predicts the reproductive compatibility of mammalian hybrids. *Proc. R. Soc. B.* 287:20200690.
- Álvarez-Carretero, S., P. Kapli, and Z. Yang. 2023. Beginner's Guide on the Use of PAML to Detect Positive Selection. *Molecular Biology and Evolution* 40:msad041.
- Andréoletti, J., A. Zwaans, R. C. M. Warnock, G. Aguirre-Fernández, J. Barido-Sottani, A. Gupta, T. Stadler, and M. Manceau. 2022. The Occurrence Birth–Death Process for Combined-Evidence Analysis in Macroevolution and Epidemiology. *Systematic Biology* 71:1440–1452.
- Arnold, B. J., I.-T. Huang, and W. P. Hanage. 2022. Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol* 20:206–218.
- Ashkenazy, H., I. Sela, E. Levy Karin, G. Landan, and T. Pupko. 2019. Multiple Sequence Alignment Averaging Improves Phylogeny Reconstruction. *Systematic Biology* 68:117–130.

- Ayres, D. L., M. P. Cummings, G. Baele, A. E. Darling, P. O. Lewis, D. L. Swofford, J. P. Huelsenbeck, P. Lemey, A. Rambaut, and M. A. Suchard. 2019. BEAGLE 3: Improved Performance, Scaling, and Usability for a High-Performance Computing Library for Statistical Phylogenetics. *Systematic Biology* 68:1052–1061.
- Baños, H., E. Susko, and A. J. Roger. 2022. Is Over-parameterization a Problem for Profile Mixture Models? *Evolutionary Biology*.
- Barba-Montoya, J., M. Dos Reis, and Z. Yang. 2017. Comparison of different strategies for using fossil calibrations to generate the time prior in Bayesian molecular clock dating. *Molecular Phylogenetics and Evolution* 114:386–400.
- Barley, A. J., J. M. Brown, and R. C. Thomson. 2018. Impact of Model Violations on the Inference of Species Boundaries Under the Multispecies Coalescent. *Systematic Biology* 67:269–284.
- Bautista, C., S. Marsit, and C. R. Landry. 2021. Interspecific hybrids show a reduced adaptive potential under DNA damaging conditions. *Evol Appl* 14:758–769.
- Betancur-R., R., D. Arcila, R. P. Vari, L. C. Hughes, C. Oliveira, M. H. Sabaj, and G. Ortí. 2019. Phylogenomic incongruence, hypothesis testing, and taxonomic sampling. *Evolution* 73:329–345. [Society for the Study of Evolution, Wiley].
- Blair, C., and C. Ané. 2020. Phylogenetic Trees and Networks Can Serve as Powerful and Complementary Approaches for Analysis of Genomic Data. *Systematic Biology* 69:593–601.
- Blaxter, M., J. M. Archibald, A. K. Childers, J. A. Coddington, K. A. Crandall, F. Di Palma, R. Durbin, S. V. Edwards, J. A. M. Graves, K. J. Hackett, N. Hall, E. D. Jarvis, R. N. Johnson, E. K. Karlsson, W. J. Kress, S. Kuraku, M. K. N. Lawniczak, K. Lindblad-Toh, J. V. Lopez, N. A. Moran, G. E. Robinson, O. A. Ryder, B. Shapiro, P. S. Soltis, T. Warnow, G. Zhang, and H. A. Lewin. 2022. Why sequence all eukaryotes? *Proc. Natl. Acad. Sci. U.S.A.* 119:e2115636118.
- Blischak, P. D., J. Chifman, A. D. Wolfe, and L. S. Kubatko. 2018. HyDe: a Python package for genome-scale hybridization detection. *Systematic Biology* 67:821–829. Oxford University Press.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Bouckaert, R. R. 2010. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* 26:1372–1373.
- Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. De Maio, M. Matschiner, F. K. Mendes, N. F. Müller, H. A. Ogilvie, L. du Plessis, A. Poppinga, A. Rambaut, D. Rasmussen, I. Siveroni, M. A. Suchard, C.-H. Wu, D. Xie, C. Zhang, T. Stadler, and A. J. Drummond. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 15:e1006650.
- Bringloe, T. T., D. Zaparenkov, S. Starko, W. S. Grant, C. Vieira, H. Kawai, T. Hanyuda, K. Filbee-Dexter, A. Klimova, T. A. Klochova, D. Krause-Jensen, B. Olesen, and H. Verbruggen. 2021. Whole-genome sequencing reveals forgotten lineages and recurrent hybridizations within the kelp genus *Alaria* (Phaeophyceae). *J. Phycol.* 57:1721–1738.
- Brohée, S., and J. Van Helden. 2006. Evaluation of clustering algorithms for protein-protein interaction networks. *BMC Bioinformatics* 7:488.
- Brüna, T., K. J. Hoff, A. Lomsadze, M. Stanke, and M. Borodovsky. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics* 3:lqaa108.
- Buchfink, B., H. Ashkenazy, K. Reuter, J. A. Kennedy, and H.-G. Drost. 2023. Sensitive clustering of protein sequences at tree-of-life scale using DIAMOND DeepClust. *Genomics*.
- Buck, R., D. Ortega-Del Vecchyo, C. Gehring, R. Michelson, D. Flores-Rentería, B. Klein, A. V. Whipple, and L. Flores-Rentería. 2023. Sequential hybridization may have facilitated ecological transitions in the Southwestern pinyon pine syngameon. *New Phytologist* 237:2435–2449.
- Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Cao, Z., M. Li, H. A. Ogilvie, and L. Nakhleh. 2022. The Impact of Model Misspecification on Phylogenetic Network Inference. *Evolutionary Biology*.
- Capella-Gutiérrez, S., M. Marcet-Houben, and T. Gabaldón. 2012. Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. *BMC Biol* 10:47.
- Carruthers, T., and R. W. Scotland. 2021. Uncertainty in Divergence Time Estimation. *Systematic Biology* 70:855–861.
- Cenci, U., S. J. Sibbald, B. A. Curtis, R. Kamikawa, L. Eme, D. Moog, B. Henrissat, E. Maréchal, M. Chabi, C. Djemiel, A. J. Roger, E. Kim, and J. M. Archibald. 2018. Nuclear genome sequence of the plastid-lacking cryptomonad *Goniomonas avonlea* provides insights into the evolution of secondary plastids. *BMC Biol* 16:137.
- Chain, F. J., J. Dushoff, and B. J. Evans. 2011. The odds of duplicate gene persistence after polyploidization. *BMC Genomics* 12:599.
- Chaprentier, C. P., and A. M. Wright. 2022. RevTiculate: An R framework for interaction with REVBayes. *Methods Ecol Evol* 13:1177–1184.

- Chen, L., J. Xu, X. Sun, and P. Xu. 2022. Research advances and future perspectives of genomics and genetic improvement in allotetraploid common carp. *Reviews in Aquaculture* 14:957–978.
- Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890.
- Cheon, S., J. Zhang, and C. Park. 2020. Is Phylotranscriptomics as Reliable as Phylogenomics? *Molecular Biology and Evolution* 37:3672–3683.
- Chifman, J., and L. Kubatko. 2014. Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics* 30:3317–3324.
- Chikina, M., J. D. Robinson, and N. L. Clark. 2016. Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. *Mol Biol Evol* 33:2182–2192.
- Clark, J. W., and P. C. J. Donoghue. 2018. Whole-Genome Duplication and Plant Macroevolution. *Trends in Plant Science* 23:933–945.
- Coelho, M. A., C. Gonçalves, J. P. Sampaio, and P. Gonçalves. 2013. Extensive Intra-Kingdom Horizontal Gene Transfer Converging on a Fungal Fructose Transporter Gene. *PLoS Genet* 9:e1003587.
- Coleman, G. A., A. A. Davin, T. A. Mahendrarajah, L. L. Szánthó, A. Spang, P. Hugenholtz, G. J. Szöllősi, and T. A. Williams. 2021. A rooted phylogeny resolves early bacterial evolution. *Science* 372:eabe0511.
- Criscuolo, A., and S. Gribaldo. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 10:210.
- Dannemann, M., and J. Kelso. 2017. The Contribution of Neanderthals to Phenotypic Variation in Modern Humans. *The American Journal of Human Genetics* 101:578–589.
- Dannemann, M., K. Prüfer, and J. Kelso. 2017. Functional implications of Neandertal introgression in modern humans. *Genome Biol* 18:61.
- Darriba, D., D. Posada, A. M. Kozlov, A. Stamatakis, B. Morel, and T. Flouri. 2020. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution* 37:291–294.
- Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772–772.
- Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24:332–340.
- Depotter, J. R., M. F. Seidl, T. A. Wood, and B. P. Thomma. 2016. Interspecific hybridization impacts host range and pathogenicity of filamentous microbes. *Current Opinion in Microbiology* 32:7–13.
- Díaz-Tapia, P., C. A. Maggs, J. A. West, and H. Verbruggen. 2017. Analysis of chloroplast genomes and a supermatrix inform reclassification of the Rhodomelaceae (Rhodophyta). *J. Phycol.* 53:920–937.
- Dobzhansky, T. 1982. *Genetics and the Origin of Species*. Columbia university press.
- Dorrell, R. G., G. Gile, G. McCallum, R. Méheust, E. P. Baptiste, C. M. Klinger, L. Brillet-Guéguen, K. D. Freeman, D. J. Richter, and C. Bowler. 2017. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *eLife* 6:e23717.
- Dorrell, R. G., A. Kuo, Z. Füßsy, E. H. Richardson, A. Salamov, N. Zarevski, N. J. Freyria, F. M. Ibarbalz, J. Jenkins, J. J. Pierella Karlusich, A. Stecca Steindorff, R. E. Edgar, L. Handley, K. Lail, A. Lipzen, V. Lombard, J. McFarlane, C. Nef, A. M. Novák Vanclová, Y. Peng, C. Plott, M. Potvin, F. R. J. Vieira, K. Barry, C. De Vargas, B. Henrissat, E. Pelletier, J. Schmutz, P. Wincker, J. B. Dacks, C. Bowler, I. V. Grigoriev, and C. Lovejoy. 2023. Convergent evolution and horizontal gene transfer in Arctic Ocean microalgae. *Life Sci. Alliance* 6:e202201833.
- Dos Reis, M., P. C. J. Donoghue, and Z. Yang. 2016. Bayesian molecular clock dating of species divergences in the genomics era. *Nat Rev Genet* 17:71–80.
- Dos Reis, M., G. F. Gunnell, J. Barba-Montoya, A. Wilkins, Z. Yang, and A. D. Yoder. 2018. Using Phylogenomic Data to Explore the Effects of Relaxed Clocks and Calibration Strategies on Divergence Time Estimation: Primates as a Test Case. *Systematic Biology* 67:594–615.
- Douglas, J., C. L. Jiménez-Silva, and R. Bouckaert. 2022. StarBeast3: Adaptive Parallelized Bayesian Inference under the Multispecies Coalescent. *Systematic Biology* 71:901–916.
- Douglass, A. P., C. E. O'Brien, B. Offei, A. Y. Coughlan, R. A. Ortiz-Merino, G. Butler, K. P. Byrne, and K. H. Wolfe. 2019. Coverage-Versus-Length Plots, a Simple Quality Control Step for *de Novo* Yeast Genome Sequence Assemblies. *G3 Genes|Genomes|Genetics* 9:879–887.
- Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol* 4:e88.
- Dunn, C. W., A. Hejnol, D. Q. Matus, K. Pang, W. E. Browne, S. A. Smith, E. Seaver, G. W. Rouse, M. Obst, G. D. Edgecombe, M. V. Sørensen, S. H. D. Haddock, A. Schmidt-Rhaesa, A. Okusu, R. M. Kristensen, W. C. Wheeler, M. Q. Martindale, and G. Giribet. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.
- Eaton, D. A. R., and R. H. Ree. 2013. Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). *Systematic Biology* 62:689–706.
- Eddy, S. R. 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol* 7:e1002195.

- Edelman, N. B., P. B. Frandsen, M. Miyagi, B. Clavijo, J. Davey, R. B. Dikow, G. García-Accinelli, S. M. Van Belleghem, N. Patterson, D. E. Neafsey, R. Challis, S. Kumar, G. R. P. Moreira, C. Salazar, M. Chouteau, B. A. Counterman, R. Papa, M. Blaxter, R. D. Reed, K. K. Dasmahapatra, M. Kronforst, M. Joron, C. D. Jiggins, W. O. McMillan, F. Di Palma, A. J. Blumberg, J. Wakeley, D. Jaffe, and J. Mallet. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science* 366:594–599.
- Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32:1792–1797. Oxford University Press.
- Edger, P. P., T. J. Poorten, R. VanBuren, M. A. Hardigan, M. Colle, M. R. McKain, R. D. Smith, S. J. Teresi, A. D. L. Nelson, C. M. Wai, E. I. Alger, K. A. Bird, A. E. Yocca, N. Pumplin, S. Ou, G. Ben-Zvi, A. Brodt, K. Baruch, T. Swale, L. Shiue, C. B. Acharya, G. S. Cole, J. P. Mower, K. L. Childs, N. Jiang, E. Lyons, M. Freeling, J. R. Puzey, and S. J. Knapp. 2019. Origin and evolution of the octoploid strawberry genome. *Nat Genet* 51:541–547.
- Edwards, S. V. 2016. Phylogenomic subsampling: a brief review. *Zool Scr* 45:63–74.
- Emms, D. M., and S. Kelly. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20:238.
- Emms, D. M., and S. Kelly. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 16:157.
- Emms, D. M., and S. Kelly. 2018. STAG: Species Tree Inference from All Genes. *Evolutionary Biology*.
- Feijó, A., D. Ge, Z. Wen, J. Cheng, L. Xia, B. D. Patterson, and Q. Yang. 2022. Mammalian diversification bursts and biotic turnovers are synchronous with Cenozoic geoclimatic events in Asia. *Proc. Natl. Acad. Sci. U.S.A.* 119:e2207845119.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *evolution* 39:783–791. Blackwell Publishing Inc Malden, USA.
- Feng, D.-F., and R. F. Doolittle. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol* 25:351–360.
- Fernández, R., and T. Gabaldón. 2020. Gene gain and loss across the metazoan tree of life. *Nat Ecol Evol* 4:524–533.
- Fernández, R., T. Gabaldón, and C. Dessimoz. 2019. Orthology: definitions, inference, and impact on species phylogeny inference. , doi: 10.48550/ARXIV.1903.04530. arXiv.
- Fletcher, W., and Z. Yang. 2009. INDELible: A Flexible Simulator of Biological Sequence Evolution. *Molecular Biology and Evolution* 26:1879–1888.
- Flouri, T., J. Huang, X. Jiao, P. Kapli, B. Rannala, and Z. Yang. 2022. Bayesian Phylogenetic Inference using Relaxed-clocks and the Multispecies Coalescent. *Molecular Biology and Evolution* 39:msac161.
- Flouri, T., X. Jiao, B. Rannala, and Z. Yang. 2018. Species Tree Inference with BPP Using Genomic Sequences and the Multispecies Coalescent. *Molecular Biology and Evolution* 35:2585–2593.
- Friedman, R., and B. Ely. 2012. Codon Usage Methods for Horizontal Gene Transfer Detection Generate an Abundance of False Positive and False Negative Results. *Curr Microbiol* 65:639–642.
- Gabaldón, T., and E. V. Koonin. 2013. Functional and evolutionary implications of gene orthology. *Nat Rev Genet* 14:360–366.
- Galindo, L. J., P. López-García, G. Torruella, S. Karpov, and D. Moreira. 2021. Phylogenomics of a new fungal phylum reveals multiple waves of reductive evolution across Holomycota. *Nat Commun* 12:4973.
- Galtier, N. 2007. A Model of Horizontal Gene Transfer and the Bacterial Phylogeny Problem. *Systematic Biology* 56:633–642.
- Galtier, N., and M. Gouy. 1995. Inferring phylogenies from DNA sequences of unequal base compositions. *Proceedings of the National Academy of Sciences* 92:11317–11321. National Acad Sciences.
- Gatesy, J., R. W. Meredith, J. E. Janecka, M. P. Simmons, W. J. Murphy, and M. S. Springer. 2017. Resolution of a concatenation/coalescence kerfuffle: partitioned coalescence support and a robust family-level tree for Mammalia. *Cladistics* 33:295–332.
- Gawryluk, R. M. R., D. V. Tikhonenkov, E. Hehenberger, F. Husnik, A. P. Mylnikov, and P. J. Keeling. 2019. Non-photosynthetic predators are sister to red algae. *Nature* 572:240–243.
- Gladyshev, E. A., M. Meselson, and I. R. Arkhipova. 2008. Massive Horizontal Gene Transfer in Bdelloid Rotifers. *Science* 320:1210–1213.
- Gonçalves, C., and P. Gonçalves. 2019. Multilayered horizontal operon transfers from bacteria reconstruct a thiamine salvage pathway in yeasts. *Proc. Natl. Acad. Sci. U.S.A.* 116:22219–22228.
- Gonçalves, C., J. H. Wisecaver, J. Kominek, M. S. Oom, M. J. Leandro, X.-X. Shen, D. A. Opulente, X. Zhou, D. Peris, C. P. Kurtzman, C. T. Hittinger, A. Rokas, and P. Gonçalves. 2018. Evidence for loss and reacquisition of alcoholic fermentation in a fructophilic yeast lineage. *eLife* 7:e33034.
- Gonçalves, P., and C. Gonçalves. 2022. Horizontal gene transfer in yeasts. *Current Opinion in Genetics & Development* 76:101950.
- Gophna, U., and N. Altman-Price. 2022. Horizontal Gene Transfer in Archaea—From Mechanisms to Genome Evolution. *Annu. Rev. Microbiol.* 76:481–502.

- Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, and M. H.-Y. Fritz. 2010. A draft sequence of the Neandertal genome. *science* 328:710–722. American Association for the Advancement of Science.
- Hale, H., E. M. Gardner, J. Viruel, L. Pokorny, and M. G. Johnson. 2020. Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. *Appl Plant Sci* 8.
- Hamlin, J. A. P., M. S. Hibbins, and L. C. Moyle. 2020. Assessing biological factors affecting postspeciation introgression. *Evolution Letters* 4:137–154.
- Harvey, M. G., G. A. Bravo, S. Claramunt, A. M. Cuervo, G. E. Derryberry, J. Battilana, G. F. Seeholzer, J. S. McKay, B. C. O'Meara, B. C. Faircloth, S. V. Edwards, J. Pérez-Emán, R. G. Moyle, F. H. Sheldon, A. Aleixo, B. T. Smith, R. T. Chesser, L. F. Silveira, J. Cracraft, R. T. Brumfield, and E. P. Derryberry. 2020. The evolution of a tropical biodiversity hotspot. *Science* 370:1343–1348.
- Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proc. Natl. Acad. Sci. U.S.A.* 111.
- Hibbins, M. S., and M. W. Hahn. 2022. Phylogenomic approaches to detecting and characterizing introgression. *Genetics* 220:iyab173.
- Ho, S. Y. M. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. *J Avian Biology* 38:409–414.
- Ho, S. Y. W., and S. Duchêne. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. *Mol Ecol* 23:5947–5965.
- Ho, S. Y. W., and M. J. Phillips. 2009. Accounting for Calibration Uncertainty in Phylogenetic Estimation of Evolutionary Divergence Times. *Systematic Biology* 58:367–380.
- Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35:518–522.
- Höhna, S., M. J. Landis, and T. A. Heath. 2017. Phylogenetic Inference Using RevBayes. *Current Protocols in Bioinformatics* 57.
- Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist. 2016. RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. *Syst Biol* 65:726–736.
- Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
- Husnik, F., and J. P. McCutcheon. 2018. Functional horizontal gene transfer from bacteria to eukaryotes. *Nature Reviews Microbiology* 16:67–79. Nature Publishing Group UK London.
- Huson, D. H. 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14:68–73.
- Huson, D. H., T. Klöpper, P. J. Lockhart, and M. A. Steel. 2005. Reconstruction of Reticulate Networks from Gene Trees. Pp. 233–249 in S. Miyano, J. Mesirov, S. Kasif, S. Istrail, P. A. Pevzner, and M. Waterman, eds. *Research in Computational Molecular Biology*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Irwin, N. A. T., A. A. Pittis, T. A. Richards, and P. J. Keeling. 2021. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nat Microbiol* 7:327–336.
- Jacobs, A., M. Carruthers, R. Eckmann, E. Yohannes, C. E. Adams, J. Behrmann-Godel, and K. R. Elmer. 2018. Rapid niche expansion by selection on functional genomic variation after ecosystem recovery. *Nat Ecol Evol* 3:77–86.
- Jaramillo, V. D. A., S. A. Sukno, and M. R. Thon. 2015. Identification of horizontally transferred genes in the genus *Colletotrichum* reveals a steady tempo of bacterial to fungal gene transfer. *BMC Genomics* 16:2.
- Jiao, X., T. Flouri, and Z. Yang. 2021. Multispecies coalescent and its applications to infer species phylogenies and cross-species gene flow. *National Science Review* 8:nwab127.
- Kainer, D., and R. Lanfear. 2015. The Effects of Partitioning on Phylogenetic Inference. *Molecular Biology and Evolution* 32:1611–1627.
- Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. Von Haeseler, and L. S. Jermin. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589.
- Kapli, P., Z. Yang, and M. J. Telford. 2020. Phylogenetic tree building in the genomic age. *Nat Rev Genet* 21:428–444.
- King, N., and A. Rokas. 2017. Embracing Uncertainty in Reconstructing Early Animal Evolution. *Current Biology* 27:R1081–R1088.
- Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol* 29:170–179.
- Kobert, K., L. Salichos, A. Rokas, and A. Stamatakis. 2016. Computing the Internode Certainty and Related Measures from Partial Gene Trees. *Mol Biol Evol* 33:1606–1617.
- Kocot, K. M., M. R. Citarella, L. L. Moroz, and K. M. Halanych. 2013. PhyloTreePruner: A Phylogenetic Tree-Based Approach for selection of Orthologous sequences for phylogenomics. *Evol Bioinform Online* 9:EBO.S12813.

- Kominek, J., D. T. Doering, D. A. Opulente, X.-X. Shen, X. Zhou, J. DeVirgilio, A. B. Hulfachor, M. Groenewald, M. A. Mcgee, S. D. Karlen, C. P. Kurtzman, A. Rokas, and C. T. Hittinger. 2019. Eukaryotic Acquisition of a Bacterial Operon. *Cell* 176:1356-1366.e10.
- Koutsovoulos, G. D., S. Granjeon Noriot, M. Bailly-Bechet, E. G. J. Danchin, and C. Rancurel. 2022. AvP: A software package for automatic phylogenetic detection of candidate horizontal gene transfers. *PLoS Comput Biol* 18:e1010686.
- Kowalczyk, A., W. K. Meyer, R. Partha, W. Mao, N. L. Clark, and M. Chikina. 2019. RERconverge: an R package for associating evolutionary rates with convergent traits. *Bioinformatics* 35:4815-4817.
- Kozlov, A. M., D. Darriba, T. Flouri, B. Morel, and A. Stamatakis. 2019. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35:4453-4455.
- Kriventseva, E. V., D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, F. A. Simão, and E. M. Zdobnov. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research* 47:D807-D811.
- Kück, P., and G. C. Longo. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool* 11:81.
- Kumar, S. 2022. Embracing Green Computing in Molecular Phylogenetics. *Molecular Biology and Evolution* 39:msac043.
- Lartillot, N. 2020. PhyloBayes: Bayesian Phylogenetics Using Site-heterogeneous Models. P. 1.5:1--1.5:16 in C. Scornavacca, F. Delsuc, and N. Galtier, eds. *Phylogenetics in the Genomic Era*. No commercial publisher | Authors open access book.
- Lartillot, N., H. Brinkmann, and H. Philippe. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol* 7:S4.
- Lemoine, F., J.-B. Domelevo Entfellner, E. Wilkinson, D. Correia, M. Dávila Felipe, T. De Oliveira, and O. Gascuel. 2018. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 556:452-456.
- Lepage, T., D. Bryant, H. Philippe, and N. Lartillot. 2007. A General Comparison of Relaxed Molecular Clock Models. *Molecular Biology and Evolution* 24:2669-2680.
- Li, L., C. J. Stoeckert, and D. S. Roos. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 13:2178-2189.
- Li, Y., K. T. David, X.-X. Shen, J. L. Steenwyk, K. M. Halanych, and A. Rokas. 2020. Feature frequency profile-based phylogenies are inaccurate. *Proc. Natl. Acad. Sci. U.S.A.* 117:31580-31581.
- Li, Y., H. Liu, J. L. Steenwyk, A. L. LaBella, M.-C. Harrison, M. Groenewald, X. Zhou, X.-X. Shen, T. Zhao, C. T. Hittinger, and A. Rokas. 2022a. Contrasting modes of macro and microsynteny evolution in a eukaryotic subphylum. *Current Biology* S0960982222016700.
- Li, Y., Z. Liu, C. Liu, Z. Shi, L. Pang, C. Chen, Y. Chen, R. Pan, W. Zhou, X. Chen, A. Rokas, J. Huang, and X.-X. Shen. 2022b. HGT is widespread in insects and contributes to male courtship in lepidopterans. *Cell* 185:2975-2987.e10.
- Li, Y., J. L. Steenwyk, Y. Chang, Y. Wang, T. Y. James, J. E. Stajich, J. W. Spatafora, M. Groenewald, C. W. Dunn, C. T. Hittinger, X.-X. Shen, and A. Rokas. 2021. A genome-scale phylogeny of the kingdom Fungi. *Current Biology* 31:1653-1665.e5.
- Li, Z., A. R. De La Torre, L. Sterck, F. M. Cánovas, C. Avila, I. Merino, J. A. Cabezas, M. T. Cervera, P. K. Ingvarsson, and Y. Van De Peer. 2017. Single-Copy Genes as Molecular Markers for Phylogenomic Studies in Seed Plants. *Genome Biology and Evolution* 9:1130-1147.
- Lin, X., S. Patel, A. P. Litvintseva, A. Floyd, T. G. Mitchell, and J. Heitman. 2009. Diploids in the *Cryptococcus neoformans* Serotype A Population Homozygous for the α Mating Type Originate via Unisexual Mating. *PLoS Pathog* 5:e1000283.
- Liu, L., D. K. Pearl, R. T. Brumfield, and S. V. Edwards. 2008. Estimating species trees using multiple-allele DNA sequence data. *Evolution* 62:2080-2091. Blackwell Publishing Inc Malden, USA.
- Liu, L., L. Yu, L. Kubatko, D. K. Pearl, and S. V. Edwards. 2009a. Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution* 53:320-328.
- Liu, L., L. Yu, D. K. Pearl, and S. V. Edwards. 2009b. Estimating Species Phylogenies Using Coalescence Times among Sequences. *Systematic Biology* 58:468-477.
- Löytynoja, A., and N. Goldman. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences* 102:10557-10562. National Acad Sciences.
- Lutteropp, S., C. Scornavacca, A. M. Kozlov, B. Morel, and A. Stamatakis. 2022. NetRAX: accurate and fast maximum likelihood phylogenetic network inference. *Bioinformatics* 38:3725-3733.
- Mallet, J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* 20:229-237.
- Mallet, J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Phil. Trans. R. Soc. B* 363:2971-2986.
- Marcet-Houben, M., and T. Gabaldón. 2015. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. *PLoS Biol* 13:e1002220.

- Martin, S. H., J. W. Davey, and C. D. Jiggins. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular biology and evolution* 32:244–257. Oxford University Press.
- Martín-Durán, J. M., J. F. Ryan, B. C. Vellutini, K. Pang, and A. Hejnol. 2017. Increased taxon sampling reveals thousands of hidden orthologs in flatworms. *Genome Res.* 27:1263–1272.
- Mateo-Estrada, V., L. Graña-Miraglia, G. López-Leal, and S. Castillo-Ramírez. 2019. Phylogenomics Reveals Clear Cases of Misclassification and Genus-Wide Phylogenetic Markers for *Acinetobacter*. *Genome Biology and Evolution* 11:2531–2541.
- McCarthy, C. G. P., P. O. Mulhair, K. Siu-Ting, C. J. Creevey, and M. J. O’Connell. 2023. Improving Orthologous Signal and Model Fit in Datasets Addressing the Root of the Animal Phylogeny. *Molecular Biology and Evolution* 40:msac276.
- Minge, M. A., K. Shalchian-Tabrizi, O. K. Tørresen, K. Takishita, I. Probert, Y. Inagaki, D. Klaveness, and K. S. Jakobsen. 2010. A phylogenetic mosaic plastid proteome and unusual plastid-targeting signals in the green-colored dinoflagellate *Lepidodinium chlorophorum*. *BMC Evol Biol* 10:191.
- Minh, B. Q., C. C. Dang, L. S. Vinh, and R. Lanfear. 2021. QMaker: Fast and Accurate Method to Estimate Empirical Models of Protein Evolution. *Systematic Biology* 70:1046–1060.
- Minh, B. Q., M. W. Hahn, and R. Lanfear. 2020a. New Methods to Calculate Concordance Factors for Phylogenomic Datasets. *Molecular Biology and Evolution* 37:2727–2733.
- Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, and R. Lanfear. 2020b. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* 37:1530–1534.
- Misof, B., S. Liu, K. Meusemann, R. S. Peters, A. Donath, C. Mayer, P. B. Frandsen, J. Ware, T. Flouri, R. G. Beutel, O. Niehuis, M. Petersen, F. Izquierdo-Carrasco, T. Wappler, J. Rust, A. J. Aberer, U. Aspöck, H. Aspöck, D. Bartel, A. Blanke, S. Berger, A. Böhm, T. R. Buckley, B. Calcott, J. Chen, F. Friedrich, M. Fukui, M. Fujita, C. Greve, P. Grobe, S. Gu, Y. Huang, L. S. Jermiin, A. Y. Kawahara, L. Krogmann, M. Kubiak, R. Lanfear, H. Letsch, Y. Li, Z. Li, J. Li, H. Lu, R. Machida, Y. Mashimo, P. Kapli, D. D. McKenna, G. Meng, Y. Nakagaki, J. L. Navarrete-Heredia, M. Ott, Y. Ou, G. Pass, L. Podsiadlowski, H. Pohl, B. M. Von Reumont, K. Schütte, K. Sekiya, S. Shimizu, A. Slipinski, A. Stamatakis, W. Song, X. Su, N. U. Szucsich, M. Tan, X. Tan, M. Tang, J. Tang, G. Timelthaler, S. Tomizuka, M. Trautwein, X. Tong, T. Uchifune, M. G. Walz, B. M. Wiegmann, J. Wilbrandt, B. Wipfler, T. K. F. Wong, Q. Wu, G. Wu, Y. Xie, S. Yang, Q. Yang, D. K. Yeates, K. Yoshizawa, Q. Zhang, R. Zhang, W. Zhang, Y. Zhang, J. Zhao, C. Zhou, L. Zhou, T. Ziesmann, S. Zou, Y. Li, X. Xu, Y. Zhang, H. Yang, J. Wang, J. Wang, K. M. Kjer, and X. Zhou. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346:763–767.
- Mixão, V., and T. Gabaldón. 2020. Genomic evidence for a hybrid origin of the yeast opportunistic pathogen *Candida albicans*. *BMC Biol* 18:48.
- Mongiardino Koch, N. 2021. Phylogenomic Subsampling and the Search for Phylogenetically Reliable Loci. *Molecular Biology and Evolution* 38:4025–4038.
- Moran, B. M., C. Payne, Q. Langdon, D. L. Powell, Y. Brandvain, and M. Schumer. 2021. The genomic consequences of hybridization. *eLife* 10:e69016.
- Morozov, A. A., and Y. P. Galachyants. 2019. Diatom genes originating from red and green algae: Implications for the secondary endosymbiosis models. *Marine Genomics* 45:72–78.
- Muñoz-Gómez, S. A., F. G. Mejía-Franco, K. Durnin, M. Colp, C. J. Grisdale, J. M. Archibald, and C. H. Slamovits. 2017. The New Red Algal Subphylum Proteorhodophytina Comprises the Largest and Most Divergent Plastid Genomes Known. *Current Biology* 27:1677–1684.e4.
- Neafsey, D. E., B. M. Barker, T. J. Sharpton, J. E. Stajich, D. J. Park, E. Whiston, C.-Y. Hung, C. McMahan, J. White, S. Sykes, D. Heiman, S. Young, Q. Zeng, A. Abouelleil, L. Aftuck, D. Bessette, A. Brown, M. FitzGerald, A. Lui, J. P. Macdonald, M. Priest, M. J. Orbach, J. N. Galgiani, T. N. Kirkland, G. T. Cole, B. W. Birren, M. R. Henn, J. W. Taylor, and S. D. Rounsley. 2010. Population genomic sequencing of *Coccidioides* fungi reveals recent hybridization and transposon control. *Genome Res.* 20:938–946.
- Nelsen, M. P., C. S. Moreau, C. Kevin Boyce, and R. H. Ree. 2023. Macroecological diversification of ants is linked to angiosperm evolution. *Evolution Letters* 7:79–87.
- Oliveros, C. H., D. J. Field, D. T. Ksepka, F. K. Barker, A. Aleixo, M. J. Andersen, P. Alström, B. W. Benz, E. L. Braun, M. J. Braun, G. A. Bravo, R. T. Brumfield, R. T. Chesser, S. Claramunt, J. Cracraft, A. M. Cuervo, E. P. Derryberry, T. C. Glenn, M. G. Harvey, P. A. Hosner, L. Joseph, R. T. Kimball, A. L. Mack, C. M. Miskelly, A. T. Peterson, M. B. Robbins, F. H. Sheldon, L. F. Silveira, B. T. Smith, N. D. White, R. G. Moyle, and B. C. Faircloth. 2019. Earth history and the passerine superradiation. *Proc. Natl. Acad. Sci. U.S.A.* 116:7916–7925.
- One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574:679–685.
- Opulente, D. A., A. L. LaBella, M.-C. Harrison, J. F. Wolters, C. Liu, Y. Li, J. Kominek, J. L. Steenwyk, H. R. Stoneman, J. VanDenAvond, C. R. Miller, Q. K. Langdon, M. Silva, C. Gonçalves, E. J. Ubbelohde, Y. Li, K. V. Buh, M. Jarzyna, M. A. B. Haase, C. A. Rosa, N. Čadež, D. Libkind, J. H. DeVirgilio, A. B. Hulfachor, C. P. Kurtzman, J. P. Sampaio, P. Gonçalves, X. Zhou, X.-X. Shen, M. Groenewald, A. Rokas, and C. T. Hittinger. 2023.

Genomic and ecological factors shaping specialism and generalism across an entire subphylum. *Evolutionary Biology*.

Ortiz-Merino, R. A., N. Kuanyshev, S. Braun-Galleani, K. P. Byrne, D. Porro, P. Branduardi, and K. H. Wolfe. 2017. Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. *PLoS Biol* 15:e2002128.

Osmanski, A. B., N. S. Paulat, J. Korstian, J. R. Grimshaw, M. Halsey, K. A. M. Sullivan, D. D. Moreno-Santillán, C. Crookshanks, J. Roberts, C. Garcia, M. G. Johnson, L. D. Densmore, R. D. Stevens, Zoonomia Consortium†, J. Rosen, J. M. Storer, R. Hubley, A. F. A. Smit, L. M. Dávalos, E. K. Karlsson, K. Lindblad-Toh, D. A. Ray, G. Andrews, J. C. Armstrong, M. Bianchi, B. W. Birren, K. R. Bredemeyer, A. M. Breit, M. J. Christmas, H. Clawson, J. Damas, F. Di Palma, M. Diekhans, M. X. Dong, E. Eizirik, K. Fan, C. Fanter, N. M. Foley, K. Forsberg-Nilsson, C. J. Garcia, J. Gatesy, S. Gazal, D. P. Genereux, L. Goodman, J. Grimshaw, M. K. Halsey, A. J. Harris, G. Hickey, M. Hiller, A. G. Hindle, R. M. Hubley, G. M. Hughes, J. Johnson, D. Juan, I. M. Kaplow, E. K. Karlsson, K. C. Keough, B. Kirilenko, K.-P. Koepfli, J. M. Korstian, A. Kowalczyk, S. V. Kozyrev, A. J. Lawler, C. Lawless, T. Lehmann, D. L. Levesque, H. A. Lewin, X. Li, A. Lind, K. Lindblad-Toh, A. Mackay-Smith, V. D. Marinescu, T. Marques-Bonet, V. C. Mason, J. R. S. Meadows, W. K. Meyer, J. E. Moore, L. R. Moreira, D. D. Moreno-Santillan, K. M. Morrill, G. Muntané, W. J. Murphy, A. Navarro, M. Nweeia, S. Ortmann, A. Osmanski, B. Paten, N. S. Paulat, A. R. Pfenning, B. N. Phan, K. S. Pollard, H. E. Pratt, D. A. Ray, S. K. Reilly, J. R. Rosen, I. Ruf, L. Ryan, O. A. Ryder, P. C. Sabeti, D. E. Schäffer, A. Serres, B. Shapiro, A. F. A. Smit, M. Springer, C. Srinivasan, C. Steiner, J. M. Storer, K. A. M. Sullivan, P. F. Sullivan, E. Sundström, M. A. Supple, R. Swofford, J.-E. Talbot, E. Teeling, J. Turner-Maier, A. Valenzuela, F. Wagner, O. Wallerman, C. Wang, J. Wang, Z. Weng, A. P. Wilder, M. E. Wirthlin, J. R. Xue, and X. Zhang. 2023. Insights into mammalian TE diversity through the curation of 248 genome assemblies. *Science* 380:eabn1430.

O'Sullivan, O., K. Suhre, C. Abergel, D. G. Higgins, and C. Notredame. 2004. 3DCoffee: combining protein sequences and structures within multiple sequence alignments. *Journal of molecular biology* 340:385–395. Elsevier.

Ozkan, H., A. A. Levy, and M. Feldman. 2001. Allopolyploidy-Induced Rapid Genome Evolution in the Wheat (*Aegilops* – *Triticum*) Group. *Plant Cell* 13:1735–1747.

Parey, E., A. Louis, J. Montfort, O. Bouchez, C. Roques, C. Iampietro, J. Lluch, A. Castinel, C. Donnadieu, T. Desvignes, C. Floi Bucac, E. Jouanno, M. Wen, S. Mejri, R. Dirks, H. Jansen, C. Henkel, W.-J. Chen, M. Zahm, C. Cabau, C. Klopp, A. W. Thompson, M. Robinson-Rechavi, I. Braasch, G. Lecointre, J. Bobe, J. H. Postlethwait, C. Berthelot, H. R. Crollius, and Y. Guiguen. 2023. Genome structures resolve the early diversification of teleost fishes. *Science* 379:572–575.

Parham, J. F., P. C. J. Donoghue, C. J. Bell, T. D. Calway, J. J. Head, P. A. Holroyd, J. G. Inoue, R. B. Irmis, W. G. Joyce, D. T. Ksepka, J. S. L. Patané, N. D. Smith, J. E. Tarver, M. Van Tuinen, Z. Yang, K. D. Angielczyk, J. M. Greenwood, C. A. Hipsley, L. Jacobs, P. J. Makovicky, J. Müller, K. T. Smith, J. M. Theodor, R. C. M. Warnock, and M. J. Benton. 2012. Best Practices for Justifying Fossil Calibrations. *Systematic Biology* 61:346–359.

Parks, D. H., M. Chuvochina, D. W. Waite, C. Rinke, A. Skarszewski, P.-A. Chaumeil, and P. Hugenholtz. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004.

Pease, J. B., and M. W. Hahn. 2015. Detection and Polarization of Introgression in a Five-Taxon Phylogeny. *Systematic Biology* 64:651–662.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE* 7:e37135.

Phillips, M. A., J. L. Steenwyk, X.-X. Shen, and A. Rokas. 2021. Examination of Gene Loss in the DNA Mismatch Repair Pathway and Its Mutational Consequences in a Fungal Phylum. *Genome Biology and Evolution* 13:evab219.

Phillips, M. J., and D. Penny. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. *Molecular Phylogenetics and Evolution* 28:171–185.

Pipes, L., H. Wang, J. P. Huelsenbeck, and R. Nielsen. 2021. Assessing Uncertainty in the Rooting of the SARS-CoV-2 Phylogeny. *Molecular Biology and Evolution* 38:1537–1543.

Pollock, D. D., D. J. Zwickl, J. A. McGuire, and D. M. Hillis. 2002. Increased Taxon Sampling Is Advantageous for Phylogenetic Inference. *Systematic Biology* 51:664–671.

Ponce-Toledo, R. I., D. Moreira, P. López-García, and P. Deschamps. 2018. Secondary Plastids of Euglenids and Chlorarachniophytes Function with a Mix of Genes of Red and Green Algal Ancestry. *Molecular Biology and Evolution* 35:2198–2204.

Qiao, H., W. Liu, Y. Zhang, Y. Zhang, and Q. Q. Li. 2019. Genetic admixture accelerates invasion via provisioning rapid adaptive evolution. *Mol Ecol* 28:4012–4027.

Racimo, F., S. Sankararaman, R. Nielsen, and E. Huerta-Sánchez. 2015. Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics* 16:359–371. Nature Publishing Group UK London.

- Raghava, G., S. M. Searle, P. C. Audley, J. D. Barber, and G. J. Barton. 2003. OXBench: A benchmark for evaluation of protein multiple sequence alignment accuracy. *BMC Bioinformatics* 4:47.
- Ranwez, V., E. J. P. Douzery, C. Cambon, N. Chantret, and F. Delsuc. 2018. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Molecular Biology and Evolution* 35:2582–2584.
- Redmond, A. K., D. Casey, M. K. Gundappa, D. J. Macqueen, and A. McLysaght. 2023. Independent rediploidization masks shared whole genome duplication in the sturgeon-paddlefish ancestor. *Nat Commun* 14:2879.
- Reich, D., N. Patterson, D. Campbell, A. Tandon, S. Mazieres, N. Ray, M. V. Parra, W. Rojas, C. Duque, and N. Mesa. 2012. Reconstructing native American population history. *Nature* 488:370–374. Nature Publishing Group UK London.
- Rieseberg, L. H., S.-C. Kim, R. A. Randell, K. D. Whitney, B. L. Gross, C. Lexer, and K. Clay. 2007. Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* 129:149–165.
- Rodríguez-Ezpeleta, N., H. Brinkmann, B. Roure, N. Lartillot, B. F. Lang, and H. Philippe. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. *Systematic biology* 56:389–399. Oxford University Press.
- Rokas, A., and P. W. H. Holland. 2000. Rare genomic changes as a tool for phylogenetics. *Trends in Ecology & Evolution* 15:454–459.
- Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425:798–804.
- Ronquist, F., M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology* 61:539–542.
- Ropars, J., R. C. Rodríguez de la Vega, M. López-Villavicencio, J. Gouzy, E. Sallet, É. Dumas, S. Lacoste, R. Debuchy, J. Dupont, A. Branca, and T. Giraud. 2015. Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi. *Current Biology* 25:2562–2569.
- S. Meseguer, A., and F. L. Condamine. 2020. Ancient tropical extinctions at high latitudes contributed to the latitudinal diversity gradient*. *Evolution* 74:1966–1987.
- Sabrina Pankey, M., D. C. Plachetzki, K. J. Macartney, M. Gastaldi, M. Slattery, D. J. Gochfeld, and M. P. Lesser. 2022. Cophylogeny and convergence shape holobiont evolution in sponge–microbe symbioses. *Nat Ecol Evol* 6:750–762.
- Salichos, L., and A. Rokas. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497:327–331.
- Salichos, L., A. Stamatakis, and A. Rokas. 2014. Novel Information Theory-Based Measures for Quantifying Incongruence among Phylogenetic Trees. *Molecular Biology and Evolution* 31:1261–1271.
- Sanderson, M. J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19:301–302.
- Sankararaman, S., S. Mallick, N. Patterson, and D. Reich. 2016. The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Current Biology* 26:1241–1247. Elsevier.
- Sayyari, E., and S. Mirarab. 2018. Testing for Polytomies in Phylogenetic Species Trees Using Quartet Frequencies. *Genes* 9:132.
- Scannell, D. R., K. P. Byrne, J. L. Gordon, S. Wong, and K. H. Wolfe. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440:341–345.
- Schenk, J. J. 2016. Consequences of Secondary Calibrations on Divergence Time Estimates. *PLoS ONE* 11:e0148228.
- Schmieder, R., and R. Edwards. 2011. Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. *PLoS ONE* 6:e17288.
- Schönknecht, G., W.-H. Chen, C. M. Ternes, G. G. Barbier, R. P. Shrestha, M. Stanke, A. Bräutigam, B. J. Baker, J. F. Banfield, R. M. Garavito, K. Carr, C. Wilkerson, S. A. Rensing, D. Gagneul, N. E. Dickenson, C. Oesterheld, M. J. Lercher, and A. P. M. Weber. 2013. Gene Transfer from Bacteria and Archaea Facilitated Evolution of an Extremophilic Eukaryote. *Science* 339:1207–1210.
- Schubert, M., T. Marcussen, A. S. Meseguer, and S. Fjellheim. 2019. The grass subfamily Pooideae: Cretaceous–Palaeocene origin and climate-driven Cenozoic diversification. *Global Ecol Biogeogr* 28:12923.
- Schultz, D. T., S. H. D. Haddock, J. V. Bredeson, R. E. Green, O. Simakov, and D. S. Rokhsar. 2023. Ancient gene linkages support ctenophores as sister to other animals. *Nature*, doi: 10.1038/s41586-023-05936-6.
- Session, A. M., and D. S. Rokhsar. 2023. Transposon signatures of allopolyploid genome evolution. *Nat Commun* 14:3180.
- Session, A. M., Y. Uno, T. Kwon, J. A. Chapman, A. Toyoda, S. Takahashi, A. Fukui, A. Hikosaka, A. Suzuki, M. Kondo, S. J. Van Heeringen, I. Quigley, S. Heinz, H. Ogino, H. Ochi, U. Hellsten, J. B. Lyons, O. Simakov, N. Putnam, J. Stites, Y. Kuroki, T. Tanaka, T. Michiue, M. Watanabe, O. Bogdanovic, R. Lister, G. Georgiou, S. S. Paranjpe, I. Van Kruijsbergen, S. Shu, J. Carlson, T. Kinoshita, Y. Ohta, S. Mawaribuchi, J. Jenkins, J. Grimwood,

- J. Schmutz, T. Mitros, S. V. Mozaffari, Y. Suzuki, Y. Haramoto, T. S. Yamamoto, C. Takagi, R. Heald, K. Miller, C. Haudenschild, J. Kitzman, T. Nakayama, Y. Izutsu, J. Robert, J. Fortriede, K. Burns, V. Lotay, K. Karimi, Y. Yasuoka, D. S. Dichmann, M. F. Flajnik, D. W. Houston, J. Shendure, L. DuPasquier, P. D. Vize, A. M. Zorn, M. Ito, E. M. Marcotte, J. B. Wallingford, Y. Ito, M. Asashima, N. Ueno, Y. Matsuda, G. J. C. Veenstra, A. Fujiyama, R. M. Harland, M. Taira, and D. S. Rokhsar. 2016. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* 538:336–343.
- Sharma, S., and S. Kumar. 2021. Fast and accurate bootstrap confidence limits on genome-scale phylogenies using little bootstraps. *Nat Comput Sci* 1:573–577.
- Shaul, S., and D. Graur. 2002. Playing chicken (*Gallus gallus*): methodological inconsistencies of molecular divergence date estimates due to secondary calibration points. *Gene* 300:59–61.
- Shen, X.-X., C. T. Hittinger, and A. Rokas. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat Ecol Evol* 1:0126.
- Shen, X.-X., D. A. Opulente, J. Kominek, X. Zhou, J. L. Steenwyk, K. V. Buh, M. A. B. Haase, J. H. Wisecaver, M. Wang, D. T. Doering, J. T. Boudouris, R. M. Schneider, Q. K. Langdon, M. Ohkuma, R. Endoh, M. Takashima, R. Manabe, N. Čadež, D. Libkind, C. A. Rosa, J. DeVirgilio, A. B. Hulfachor, M. Groenewald, C. P. Kurtzman, C. T. Hittinger, and A. Rokas. 2018. Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell* 175:1533–1545.e20.
- Shen, X.-X., L. Salichos, and A. Rokas. 2016. A Genome-Scale Investigation of How Sequence, Function, and Tree-Based Gene Properties Influence Phylogenetic Inference. *Genome Biol Evol* 8:2565–2580.
- Shen, X.-X., J. L. Steenwyk, A. L. LaBella, D. A. Opulente, X. Zhou, J. Kominek, Y. Li, M. Groenewald, C. T. Hittinger, and A. Rokas. 2020. Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Sci. Adv.* 6:eabd0079.
- Shen, X.-X., J. L. Steenwyk, and A. Rokas. 2021. Dissecting Incongruence between Concatenation- and Quartet-Based Approaches in Phylogenomic Data. *Systematic Biology* 70:997–1014.
- Shimodaira, H., and M. Hasegawa. 1999. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution* 16:1114–1116.
- Si Quang, L., O. Gascuel, and N. Lartillot. 2008. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* 24:2317–2323.
- Sibbald, S. J., and J. M. Archibald. 2020. Genomic Insights into Plastid Evolution. *Genome Biology and Evolution* 12:978–990.
- Sibbald, S. J., L. Eme, J. M. Archibald, and A. J. Roger. 2020. Lateral Gene Transfer Mechanisms and Pan-genomes in Eukaryotes. *Trends in Parasitology* 36:927–941.
- Sierra-Patev, S., B. Min, M. Naranjo-Ortiz, B. Looney, Z. Konkel, J. C. Slot, Y. Sakamoto, J. L. Steenwyk, A. Rokas, J. Carro, S. Camarero, P. Ferreira, G. Molpeceres, F. J. Ruiz-Dueñas, A. Serrano, B. Henrissat, E. Drula, K. W. Hughes, J. L. Mata, N. K. Ishikawa, R. Vargas-Isla, S. Ushijima, C. A. Smith, J. Donoghue, S. Ahrendt, W. Andreopoulos, G. He, K. LaButti, A. Lipzen, V. Ng, R. Riley, L. Sandor, K. Barry, A. T. Martínez, Y. Xiao, J. G. Gibbons, K. Terashima, I. V. Grigoriev, and D. Hibbett. 2023. A global phylogenomic analysis of the shiitake genus *Lentinula*. *Proc. Natl. Acad. Sci. U.S.A.* 120:e2214076120.
- Sievers, F., and D. G. Higgins. 2020. *QuaTest2*: benchmarking multiple sequence alignments using secondary structure prediction. *Bioinformatics* 36:90–95.
- Simion, P., H. Philippe, D. Baurain, M. Jager, D. J. Richter, A. Di Franco, B. Roure, N. Satoh, É. Quéinnec, A. Ereskovsky, P. Lapébie, E. Corre, F. Delsuc, N. King, G. Wörheide, and M. Manuel. 2017. A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Current Biology* 27:958–967.
- Simonti, C. N., B. Vernot, L. Bastarache, E. Bottinger, D. S. Carrell, R. L. Chisholm, D. R. Crosslin, S. J. Hebring, G. P. Jarvik, I. J. Kullo, R. Li, J. Pathak, M. D. Ritchie, D. M. Roden, S. S. Verma, G. Tromp, J. D. Prato, W. S. Bush, J. M. Akey, J. C. Denny, and J. A. Capra. 2016. The phenotypic legacy of admixture between modern humans and Neandertals. *Science* 351:737–741.
- Smith, M. L., D. Vanderpool, and M. W. Hahn. 2022. Using all Gene Families Vastly Expands Data Available for Phylogenomic Inference. *Molecular Biology and Evolution* 39:msac112.
- Smith, S. A., and B. C. O'Meara. 2012. *treePL*: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28:2689–2690.
- Sousa, F., J. Neiva, N. Martins, R. Jacinto, L. Anderson, P. T. Raimondi, E. A. Serrão, and G. A. Pearson. 2019. Increased evolutionary rates and conserved transcriptional response following allopolyploidization in brown algae: GENOME EVOLUTION IN ALLOPOLYPOID ALGAE. *Evolution* 73:59–72.
- Stadler, T., O. G. Pybus, and M. P. H. Stumpf. 2021. Phylodynamics for cell biologists. *Science* 371:eaah6266.
- Stadler, T., and Z. Yang. 2013. Dating Phylogenies with Sequentially Sampled Tips. *Systematic Biology* 62:674–688.
- Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology* 57:758–771.
- Stanke, M., O. Keller, I. Gunduz, A. Hayes, S. Waack, and B. Morgenstern. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research* 34:W435–W439.

- Steenwyk, J., and N. King. 2023. From Genes to Genomes: Opportunities and Challenges for Synteny-based Phylogenies. Preprints, doi: 10.20944/preprints202309.0495.v1. Preprints.
- Steenwyk, J. L., C. Balamurugan, H. A. Raja, C. Gonçalves, N. Li, F. Martin, J. Berman, N. H. Oberlies, J. G. Gibbons, G. H. Goldman, D. M. Geiser, D. S. Hibbett, and A. Rokas. 2022a. Phylogenomics reveals extensive misidentification of fungal strains from the genus *Aspergillus*. *Evolutionary Biology*.
- Steenwyk, J. L., T. J. Buida, C. Gonçalves, D. C. Goltz, G. Morales, M. E. Mead, A. L. LaBella, C. M. Chavez, J. E. Schmitz, M. Hadjifrangiskou, Y. Li, and A. Rokas. 2022b. BioKIT: a versatile toolkit for processing and analyzing diverse types of sequence data. *Genetics* 221:iyac079.
- Steenwyk, J. L., T. J. Buida, A. L. Labella, Y. Li, X.-X. Shen, and A. Rokas. 2021. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics* 37:2325–2331.
- Steenwyk, J. L., T. J. Buida, Y. Li, X.-X. Shen, and A. Rokas. 2020a. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol* 18:e3001007.
- Steenwyk, J. L., D. C. Goltz, T. J. Buida, Y. Li, X.-X. Shen, and A. Rokas. 2022c. OrthoSNAP: A tree splitting and pruning algorithm for retrieving single-copy orthologs from gene family trees. *PLoS Biol* 20:e3001827.
- Steenwyk, J. L., S. L. Knowles, R. Bastos, C. Balamurugan, D. Rinker, M. E. Mead, C. D. Roberts, H. A. Raja, Y. Li, A. C. Colabardini, P. A. Castro, T. F. Reis, D. Canovas, R. L. Sanchez, K. Lagrou, E. Torrado, F. Rodrigues, N. H. Oberlies, X. Zhou, G. Goldman, and A. Rokas. 2023a. Evolutionary origin, population diversity, and diagnostics for a cryptic hybrid pathogen. *Evolutionary Biology*.
- Steenwyk, J. L., Y. Li, X. Zhou, X.-X. Shen, and A. Rokas. 2023b. Incongruence in the phylogenomics era. *Nature Reviews Genetics*, doi: 10.1038/s41576-023-00620-x.
- Steenwyk, J. L., A. L. Lind, L. N. A. Ries, T. F. dos Reis, L. P. Silva, F. Almeida, R. W. Bastos, T. F. de C. Fraga da Silva, V. L. D. Bonato, A. M. Pessoni, F. Rodrigues, H. A. Raja, S. L. Knowles, N. H. Oberlies, K. Lagrou, G. H. Goldman, and A. Rokas. 2020b. Pathogenic Allodiploid Hybrids of *Aspergillus* Fungi. *Current Biology* 30:2495–2507.e7.
- Steenwyk, J. L., D. A. Ofulente, J. Kominek, X.-X. Shen, X. Zhou, A. L. Labella, N. P. Bradley, B. F. Eichman, N. Čadež, D. Libkind, J. DeVirgilio, A. B. Hulfachor, C. P. Kurtzman, C. T. Hittinger, and A. Rokas. 2019a. Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. *PLoS Biol* 17:e3000255.
- Steenwyk, J. L., M. A. Phillips, F. Yang, S. S. Date, T. R. Graham, J. Berman, C. T. Hittinger, and A. Rokas. 2022d. An orthologous gene coevolution network provides insight into eukaryotic cellular and genomic structure and function. *Sci. Adv.* 8:eabn0105.
- Steenwyk, J. L., and A. Rokas. 2021. orthofisher: a broadly applicable tool for automated gene identification and retrieval. *G3 Genes|Genomes|Genetics* 11:jkab250.
- Steenwyk, J. L., and A. Rokas. 2023. The dawn of relaxed phylogenetics. *PLoS Biol* 21:e3001998.
- Steenwyk, J. L., X.-X. Shen, A. L. Lind, G. H. Goldman, and A. Rokas. 2019b. A Robust Phylogenomic Time Tree for Biotechnologically and Medically Important Fungi in the Genera *Aspergillus* and *Penicillium*. *mBio* 10:e00925-19.
- Steinegger, M., and J. Söding. 2018. Clustering huge protein sequence sets in linear time. *Nat Commun* 9:2542.
- Stiller, J. W., J. Schreiber, J. Yue, H. Guo, Q. Ding, and J. Huang. 2014. The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat Commun* 5:5764.
- Strasser, J. F. H., I. Irisarri, T. A. Williams, and F. Burki. 2021. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat Commun* 12:1879.
- Stukenbrock, E. H. 2016. The Role of Hybridization in the Evolution and Emergence of New Fungal Plant Pathogens. *Phytopathology* 106:104–112.
- Suvorov, A., B. Y. Kim, J. Wang, E. E. Armstrong, D. Peede, E. R. R. D'Agostino, D. K. Price, P. J. Waddell, M. Lang, V. Courtier-Orgogozo, J. R. David, D. Petrov, D. R. Matute, D. R. Schrider, and A. A. Comeault. 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Current Biology* 32:111–123.e5.
- Tahon, G., P. Geesink, and T. J. G. Ettema. 2021. Expanding Archaeal Diversity and Phylogeny: Past, Present, and Future. *Annu. Rev. Microbiol.* 75:359–381.
- Talavera, G., and J. Castresana. 2007. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Systematic Biology* 56:564–577.
- Tan, G., M. Muffato, C. Ledergerber, J. Herrero, N. Goldman, M. Gil, and C. Dessimoz. 2015. Current Methods for Automated Filtering of Multiple Sequence Alignments Frequently Worsen Single-Gene Phylogenetic Inference. *Syst Biol* 64:778–791.
- Tao, Q., K. Tamura, and S. Kumar. 2020. Efficient Methods for Dating Evolutionary Divergences. Pp. 197–219 in S. Y. W. Ho, ed. *The Molecular Evolutionary Clock*. Springer International Publishing, Cham.
- Thompson, J. D., F. Plewniak, and O. Poch. 1999. BALI-BASE: a benchmark alignment database for the evaluation of multiple alignment programs. *Bioinformatics (Oxford, England)* 15:87–88.
- Tiley, G. P., T. Flouri, X. Jiao, J. W. Poelstra, B. Xu, T. Zhu, B. Rannala, A. D. Yoder, and Z. Yang. 2023. Estimation of species divergence times in presence of cross-species gene flow. *Systematic Biology* 72:820–836.
- Tiley, G. P., J. W. Poelstra, M. Dos Reis, Z. Yang, and A. D. Yoder. 2020. Molecular Clocks without Rocks: New Solutions for Old Problems. *Trends in Genetics* 36:845–856.

- Upham, N. S., J. A. Esselstyn, and W. Jetz. 2019. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol* 17:e3000494.
- Upham, N. S., J. A. Esselstyn, and W. Jetz. 2021. Molecules and fossils tell distinct yet complementary stories of mammal diversification. *Current Biology* 31:4195–4206.e3.
- Vaidya, G., D. J. Lohman, and R. Meier. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27:171–180.
- Van Dongen, S. 2008. Graph Clustering Via a Discrete Uncoupling Process. *SIAM J. Matrix Anal. & Appl.* 30:121–141.
- Van Etten, J., and D. Bhattacharya. 2020. Horizontal Gene Transfer in Eukaryotes: Not if, but How Much? *Trends in Genetics* 36:915–925.
- Volz, E. M., K. Koelle, and T. Bedford. 2013. Viral Phylodynamics. *PLoS Comput Biol* 9:e1002947.
- Wang, H.-C., B. Q. Minh, E. Susko, and A. J. Roger. 2018a. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Systematic Biology* 67:216–235.
- Wang, M.-S., G. G. R. Murray, D. Mann, P. Groves, A. O. Vershinina, M. A. Supple, J. D. Kapp, R. Corbett-Detig, S. E. Crump, I. Stirling, K. L. Laidre, M. Kunz, L. Dalén, R. E. Green, and B. Shapiro. 2022. A polar bear paleogenome reveals extensive ancient gene flow from polar bears into brown bears. *Nat Ecol Evol* 6:936–944.
- Wang, Y., H. Wu, and Y. Cai. 2018b. A benchmark study of sequence alignment methods for protein clustering. *BMC Bioinformatics* 19:529.
- Waterhouse, R. M., M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva, and E. M. Zdobnov. 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution* 35:543–548.
- Weisman, C. M., A. W. Murray, and S. R. Eddy. 2020. Many, but not all, lineage-specific genes can be explained by homology detection failure. *PLoS Biol* 18:e3000862.
- Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, M. S. Barker, J. G. Burleigh, M. A. Gitzendanner, B. R. Ruhfel, E. Wafula, J. P. Der, S. W. Graham, S. Mathews, M. Melkonian, D. E. Soltis, P. S. Soltis, N. W. Miles, C. J. Rothfels, L. Pokorny, A. J. Shaw, L. DeGironimo, D. W. Stevenson, B. Surek, J. C. Villarreal, B. Roure, H. Philippe, C. W. dePamphilis, T. Chen, M. K. Deyholos, R. S. Baucom, T. M. Kutchan, M. M. Augustin, J. Wang, Y. Zhang, Z. Tian, Z. Yan, X. Wu, X. Sun, G. K.-S. Wong, and J. Leebens-Mack. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U.S.A.* 111.
- Wiens, J. J., and J. Tiu. 2012. Highly Incomplete Taxa Can Rescue Phylogenetic Analyses from the Negative Impacts of Limited Taxon Sampling. *PLoS ONE* 7:e42925.
- Williams, T. A., C. J. Cox, P. G. Foster, G. J. Szöllösi, and T. M. Embley. 2019. Phylogenomics provides robust support for a two-domains tree of life. *Nat Ecol Evol* 4:138–147.
- Willson, J., M. S. Roddur, B. Liu, P. Zaharias, and T. Warnow. 2022. DISCO: Species Tree Inference using Multicopy Gene Family Tree Decomposition. *Systematic Biology* 71:610–629.
- Worobey, M., T. D. Watts, R. A. McKay, M. A. Suchard, T. Granade, D. E. Teuwen, B. A. Koblin, W. Heneine, P. Lemey, and H. W. Jaffe. 2016. 1970s and ‘Patient 0’ HIV-1 genomes illuminate early HIV/AIDS history in North America. *Nature* 539:98–101.
- Yang, Z., and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences* 107:9264–9269. *National Acad Sciences*.
- Yu, Y., R. M. Barnett, and L. Nakhleh. 2013. Parsimonious Inference of Hybridization in the Presence of Incomplete Lineage Sorting. *Systematic Biology* 62:738–751.
- Yuan, L., H. Lu, F. Li, J. Nielsen, and E. J. Kerkhoven. 2023. HGTphyloDetect: facilitating the identification and phylogenetic analysis of horizontal gene transfer. *Briefings in Bioinformatics* 24:bbad035.
- Yue, J., X. Hu, H. Sun, Y. Yang, and J. Huang. 2012. Widespread impact of horizontal gene transfer on plant colonization of land. *Nat Commun* 3:1152.
- Zanewich, K. P., D. W. Pearce, and S. B. Rood. 2018. Heterosis in poplar involves phenotypic stability: cottonwood hybrids outperform their parental species at suboptimal temperatures. *Tree Physiology* 38:789–800.
- Zeberg, H., and S. Pääbo. 2020. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* 587:610–612.
- Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19:153.
- Zhang, C., C. Scornavacca, E. K. Molloy, and S. Mirarab. 2020. ASTRAL-Pro: Quartet-Based Species-Tree Inference despite Paralogy. *Molecular Biology and Evolution* 37:3292–3307.
- Zhang, C., Y. Zhao, E. L. Braun, and S. Mirarab. 2021. TAPER: Pinpointing errors in multiple sequence alignments despite varying rates of evolution. *Methods Ecol Evol* 12:2145–2158.
- Zhang, G., C. Li, Q. Li, B. Li, D. M. Larkin, C. Lee, J. F. Storz, A. Antunes, M. J. Greenwold, R. W. Meredith, A. Ödeen, J. Cui, Q. Zhou, L. Xu, H. Pan, Z. Wang, L. Jin, P. Zhang, H. Hu, W. Yang, J. Hu, J. Xiao, Z. Yang, Y. Liu, Q. Xie, H. Yu, J. Lian, P. Wen, F. Zhang, H. Li, Y. Zeng, Z. Xiong, S. Liu, L. Zhou, Z. Huang, N. An, J. Wang, Q. Zheng, Y. Xiong, G. Wang, B. Wang, J. Wang, Y. Fan, R. R. Da Fonseca, A. Alfaro-Núñez, M. Schubert, L. Orlando,

- T. Mourier, J. T. Howard, G. Ganapathy, A. Pfenning, O. Whitney, M. V. Rivas, E. Hara, J. Smith, M. Farré, J. Narayan, G. Slavov, M. N. Romanov, R. Borges, J. P. Machado, I. Khan, M. S. Springer, J. Gatesy, F. G. Hoffmann, J. C. Opazo, O. Håstad, R. H. Sawyer, H. Kim, K.-W. Kim, H. J. Kim, S. Cho, N. Li, Y. Huang, M. W. Bruford, X. Zhan, A. Dixon, M. F. Bertelsen, E. Derryberry, W. Warren, R. K. Wilson, S. Li, D. A. Ray, R. E. Green, S. J. O'Brien, D. Griffin, W. E. Johnson, D. Haussler, O. A. Ryder, E. Willerslev, G. R. Graves, P. Alström, J. Fjeldså, D. P. Mindell, S. V. Edwards, E. L. Braun, C. Rahbek, D. W. Burt, P. Houde, Y. Zhang, H. Yang, J. Wang, Avian Genome Consortium, E. D. Jarvis, M. T. P. Gilbert, J. Wang, C. Ye, S. Liang, Z. Yan, M. L. Zepeda, P. F. Campos, A. M. V. Velazquez, J. A. Samaniego, M. Avila-Arcos, M. D. Martin, R. Barnett, A. M. Ribeiro, C. V. Mello, P. V. Lovell, D. Almeida, E. Maldonado, J. Pereira, K. Sunagar, S. Philip, M. G. Dominguez-Bello, M. Bunce, D. Lambert, R. T. Brumfield, F. H. Sheldon, E. C. Holmes, P. P. Gardner, T. E. Steeves, P. F. Stadler, S. W. Burge, E. Lyons, J. Smith, F. McCarthy, F. Pitel, D. Rhoads, and D. P. Froman. 2014a. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346:1311–1320.
- Zhang, R., H.-Y. Ou, F. Gao, and H. Luo. 2014b. Identification of Horizontally-transferred Genomic Islands and Genome Segmentation Points by Using the GC Profile Method. *CG* 15:113–121.
- Zheng, Y., and A. Janke. 2018. Gene flow analysis method, the D-statistic, is robust in a wide parameter space. *BMC bioinformatics* 19:1–19. Springer.
- Zhou, X., S. Lutteropp, L. Czech, A. Stamatakis, M. V. Looz, and A. Rokas. 2020. Quartet-Based Computations of Internode Certainty Provide Robust Measures of Phylogenetic Incongruence. *Systematic Biology* 69:308–324.
- Zhou, X., and A. Rokas. 2014. Prevention, diagnosis and treatment of high-throughput sequencing data pathologies. *Mol Ecol* 23:1679–1700.
- Zhou, X., X.-X. Shen, C. T. Hittinger, and A. Rokas. 2018. Evaluating Fast Maximum Likelihood-Based Phylogenetic Programs Using Empirical Phylogenomic Data Sets. *Molecular Biology and Evolution* 35:486–503.
- Zhu, Q., M. Kosoy, and K. Dittmar. 2014. HGTector: an automated method facilitating genome-wide discovery of putative horizontal gene transfers. *BMC Genomics* 15:717.
- Zhu, Q., U. Mai, W. Pfeiffer, S. Janssen, F. Asnicar, J. G. Sanders, P. Belda-Ferre, G. A. Al-Ghalith, E. Kopylova, D. McDonald, T. Kosciółek, J. B. Yin, S. Huang, N. Salam, J.-Y. Jiao, Z. Wu, Z. Z. Xu, K. Cantrell, Y. Yang, E. Sayyari, M. Rabiee, J. T. Morton, S. Podell, D. Knights, W.-J. Li, C. Huttenhower, N. Segata, L. Smarr, S. Mirarab, and R. Knight. 2019. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat Commun* 10:5477.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.