

ESSAY

The promise and pitfalls of synteny in phylogenomics

Jacob L. Steenwyk^{1,2*}, Nicole King^{1,2}

1 Howard Hughes Medical Institute, University of California, Berkeley, California, United States of America, **2** Department of Molecular and Cell Biology, University of California, Berkeley, California, United States of America

* jsteenwyk@berkeley.edu

Abstract

Reconstructing the tree of life remains a central goal in biology. Early methods, which relied on small numbers of morphological or genetic characters, often yielded conflicting evolutionary histories, undermining confidence in the results. Investigations based on phylogenomics, which use hundreds to thousands of loci for phylogenetic inquiry, have provided a clearer picture of life's history, but certain branches remain problematic. To resolve difficult nodes on the tree of life, 2 recent studies tested the utility of synteny, the conserved collinearity of orthologous genetic loci in 2 or more organisms, for phylogenetics. Synteny exhibits compelling phylogenomic potential while also raising new challenges. This Essay identifies and discusses specific opportunities and challenges that bear on the value of synteny data and other rare genomic changes for phylogenomic studies. Synteny-based analyses of highly contiguous genome assemblies mark a new chapter in the phylogenomic era and the quest to reconstruct the tree of life.



OPEN ACCESS

Citation: Steenwyk JL, King N (2024) The promise and pitfalls of synteny in phylogenomics. *PLoS Biol* 22(5): e3002632. <https://doi.org/10.1371/journal.pbio.3002632>

Published: May 20, 2024

Copyright: © 2024 Steenwyk, King. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: JLS is a Howard Hughes Medical Institute Awardee of the Life Sciences Research Foundation. Research in the King lab is supported by the Howard Hughes Medical Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: JLS is an advisor for ForensisGroup Inc.

Introduction

Arguably, the most ambitious goal in phylogenetics is to reconstruct the entire tree of life. To build phylogenetic trees, diverse data types have been used, and our understanding of the tree of life has undergone significant transformations with each methodological advance.

Early studies relied on aligning single or few loci to reconstruct evolutionary histories [1], but analyses of different loci often yielded phylogenies with conflicting or poorly supported topologies [2,3] (Fig 1A–1C). For example, analyses of different loci have suggested different relationships among humans, bonobos, and chimps, and among sponges, ctenophores, and bilaterians [4,5]. Numerous processes can contribute to loci with evolutionary histories that appear distinct from those of the organisms in which they are found [3,6,7], including horizontal gene transfer, convergent evolution, and incomplete lineage sorting.

The advent of cost-effective whole genome sequencing has paved the way for the phylogenomics era, in which hundreds to thousands of orthologous loci are analyzed in a total evidence approach [10,11]. The promise of phylogenomics has been that the increase in sequence data might allow phylogenetic signal to outcompete noise. Indeed, phylogenomics has successfully been used to delineate previously problematic branches within the tree of life, for

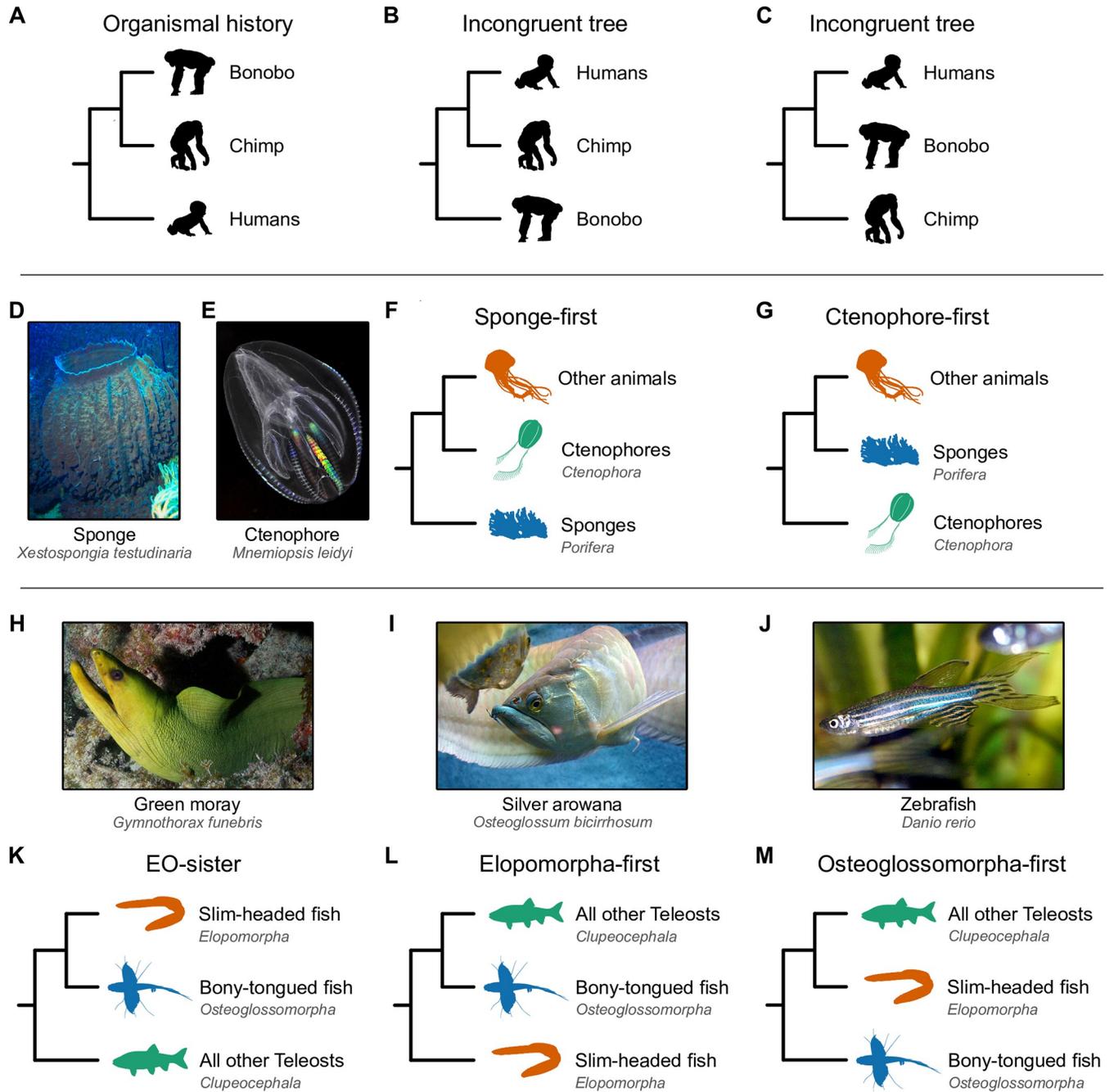


Fig 1. Depictions of incongruence and alternate hypotheses for primates, the base of the animal tree, and teleost fish phylogenies. (A) Example of tree incongruence. The weight of evidence strongly supports a sister relationship between bonobos and chimps, to the exclusion of humans. (B, C) Phylogenies that are incongruent would suggest a sister relationship between humans and chimps (B) or humans and bonobos (C). (D-G) The debate concerning early animal evolution has largely focused on whether sponges (D) or ctenophores (E) diverged first from all other animals: the sponge-first (F) and ctenophore-first (G) hypotheses, respectively. (H-M) Among teleost fish, the debate centers on the relationships among 3 major lineages—the Elopomorpha (mostly slim-headed fish; H), Osteoglossomorpha (mostly bony-tongued fish; I), and Clupeocephala (all other teleost fish; J). The Eloposteoglossomorpha (EO-sister) hypothesis (K) suggests a sister relationship between slim-headed and bony-tongued fish, whereas the Elopomorpha-first (L) and Osteoglossomorpha-first (M) hypotheses suggest that slim-headed fish or bony-tongued fish, respectively, diverged before the other lineages split from one another. Recent studies that employed synteny as a phylogenomic marker supported the ctenophore-first (G) and EO-sister (K) hypotheses [8,9]. All images were obtained from the Wikimedia Commons (<https://commons.wikimedia.org>) or PhyloPic (<https://www.phylopic.org>) and are dedicated to the public domain; all credit goes to their respective contributors.

<https://doi.org/10.1371/journal.pbio.3002632.g001>

example, the monophyletic grouping of nematodes and arthropods within Ecdysozoa [12,13], the placement of turtles as sister to archosaurs (crocodiles and birds) [14], and the placement of eukaryotes within Archaea [15]. These successes have positioned phylogenomics as the current standard for reconstructing most evolutionary histories. However, many branches in the tree of life remain unresolved, including those that concern key evolutionary episodes.

To address unresolved branches, phylogeneticists have sought to identify new genomic characters that accurately reflect evolutionary history, in part because they are unlikely to evolve independently in unrelated groups of organisms [16–18]. To this end, 2 recent studies [8,9] have tested the utility of gene synteny as a character for phylogenetics (Box 1). In this Essay, we review the challenges that inspired these studies, evaluate the current utility of gene synteny as a character for phylogenetics, and offer a roadmap for future use of gene synteny to reconstruct the tree of life.

Glossary

Horizontal gene transfer

Exchange of genetic material between organisms through non-reproductive mechanisms

Convergent evolution

The independent evolution of similar features in unrelated species

Incomplete lineage sorting

The retention and random sorting of ancestral polymorphisms, which can cause phylogenies based on these polymorphisms to, at times, differ from the organismal history

Rare genomic changes

Polymorphisms—indels, transposon integrations, changes in gene order, gene duplications, and others—excluding substitutions

Synten

The conservation of the same order of loci on chromosomes from different species

Orthology inference

The process of determining which genes in different species are orthologs, meaning they diverged due to a speciation event

Microsynten

Conservation of small blocks of genes (typically only a handful) that are found in the same order within the genome

Macrosynten

Large-scale conservation of blocks of genes (hundreds to thousands or more) on chromosomes between species

Reciprocal best BLAST hits

A method used to find orthologous genes, in which 2 genes from different species are each other's best match in a BLAST search

Acrocentric chromosomes

Chromosomes with a centromere near one end, resulting in 1 very short and 1 very long arm

Robertsonian translocation

A chromosomal rearrangement in which 2 acrocentric chromosomes have fused to form a single chromosome

Taxon sampling

The selection of taxa for a phylogenetic study

Maximum likelihood framework

A statistical approach used to infer evolutionary trees by finding the tree topology with the best probability given the underlying data and a model of sequence evolution

Long-branch attraction

An error in phylogenetic inference wherein lineages on long branches (i.e., having many substitutions per site in a data matrix) are incorrectly inferred to be closely related

Tandem duplication

A type of mutation in which a region of a chromosome is duplicated and the copies remain adjacent to each other

Syntenic coverage

The percentage of the full genome that contains syntenic blocks that are conserved in comparator genomes. Determined by taking the sum length of syntenic blocks divided by genome size

Pangenomes

The entire set of genes present within all strains of a species, not just those in a single reference genome

Treeness

A signal-to-noise measure based on the proportion of branch lengths observed among internal branches compared to internal and terminal branches

Rogue taxa

Taxa with placements that are unstable across a set of trees

Ohnologs

Genes duplicated through a whole-genome duplication event

Tangled branches in the tree of life

There are many unresolved branches in the tree of life. Here, we focus on 2 major challenges: how to root the tree of animals, and how major clades of teleost fish, a group encompassing nearly half of all vertebrates, are related. These evolutionary questions exemplify how genome-

scale data analyses can yield incongruent phylogenies and undermine our ability to fully reconstruct the tree of life.

The controversy surrounding the root of the animal tree was somewhat unexpected, as morphological comparisons had, for decades, consistently favored the placement of sponges (Fig 1D), not ctenophores (Fig 1E), as the earliest-branching lineage [19,20]; this hypothesis garnered nearly universal support during the single-locus era of phylogenetics [19,21–23] (Fig 1F). The dawn of phylogenomics, however, changed the situation. A 2008 study based on 150 genes from 77 taxa, including 2 sponges and 2 ctenophores, provided the first support for placing ctenophores at the root of the animal tree (Fig 1G) [24]. Then, in 2009, the sponge-first hypothesis was supported by a study using 128 genes and 55 taxa, including 9 sponges and 3 ctenophores [25]. Since then, investigations powered by ever larger datasets (including dozens of ctenophores and sponges) and analyzed using the latest methods in phylogenomics have provided compelling and contradictory evidence for the 2 competing hypotheses [5,25–30].

Early branching patterns in the teleost fish phylogeny are also intensely debated. Teleosts encompass 3 major clades: Elopomorpha (mostly slim-headed fish like bonefish, eels, and skip-jacks; Fig 1H); Osteoglossomorpha (mostly bony-tongued fish like elephantnose fish, double-sash butterflyfish, and mormyrids; Fig 1I); and Clupeocephala (the remaining extant teleosts like pufferfish and sticklebacks; Fig 1J). Phylogenetics of some single-locus data suggested a sister relationship between Elopomorpha and Osteoglossomorpha—the Eloposteoglossocephala (EO-sister) hypothesis—in which the slim-headed and bony-tongued fish are thought to form a sister clade relative to all other teleosts [31]. However, all possible topologies (Fig 1K–1M) have received support in the phylogenomic era. Challenged by a history of conflict, some have suggested that the base of the teleost fish phylogeny is one of the most important unresolved questions in ray-finned fish evolution [32].

Rare genomic changes as phylogenomic markers

Amid these and other ongoing debates, the value of alternative phylogenetic markers, such as rare genomic changes, has been explored [33]. Rare genomic changes are an independent source of phylogenetic information compared to primary sequence data and can complement sequence data or be used to evaluate alternative phylogenetic scenarios when sequence data are inconclusive [33]. The phylogenetic distributions of some rare genomic changes, including insertions and deletions, gene duplications and losses, and alternative genetic codes, often mirror the inferred evolutionary relationships among major vertebrate, insect, fungal, and related lineages [34–37].

The earliest studies underscoring the promise of rare genomic changes for phylogenetics were conducted before widely available whole-genome sequences. In studies conducted in the 1930s, Sturtevant and Dobzhansky reconstructed phylogenetic relationships among populations of *Drosophila pseudoobscura* by analyzing chromosomal inversions detected in the polytene chromosomes of salivary glands [38,39]. These observations led Sturtevant and Dobzhansky to suggest that comparing "different gene arrangements in the same chromosome may, in certain cases, throw light on the historical relationships of these structures, and consequently on the history of the species as a whole." Supporting this hypothesis, Hampton Carson conducted a similar analysis in 1983 to reconstruct the evolutionary relationships among Hawaiian *Drosophila* [40].

Several other cases of rare genomic changes recapitulating phylogeny have been identified. Copy number variants (duplicated or deleted loci), gene presence–absence polymorphisms, and transposable element insertions and deletions can mirror population structure and deeper-scale evolutionary relationships [41–49]. For example, lineage-specific gene

duplication and loss events have been detected in humans [50] and in lineages of the bipolar budding yeast *Hanseniaspora* [37]. Genetic recoding of CUG to alanine and serine, rather than leucine, occurred in a monophyletic lineage of yeast [51]. Among more ancient divergences, the root of the angiosperm phylogeny has been successfully examined using duplication patterns of phytochrome genes [52,53].

Nonetheless, rare genomic changes are not irreproachable for phylogenetic inference. For example, rare genomic changes can evolve convergently. Losses of gene duplicates have occurred repeatedly in flatworms [54] and genetic recoding of the CUG codon from leucine to serine in Saccharomycotina fungi occurred on 2 occasions independently [55]. Convergence has also been observed among structural genomic features. For example, distributions of mitochondrial genome size, structure, and content have converged among Placozoa, chytrid fungi, and choanoflagellates [56], leading briefly to the inference that Placozoa diverged from all other animals first—a hypothesis largely refuted by phylogenomic analyses of nuclear genes [24–30,57]. Even in closely related species of walnuts, phylogenies inferred from large amounts of local gene-order data, DNA sequence alignments, and gene-family content, yield differing tree topologies [58].

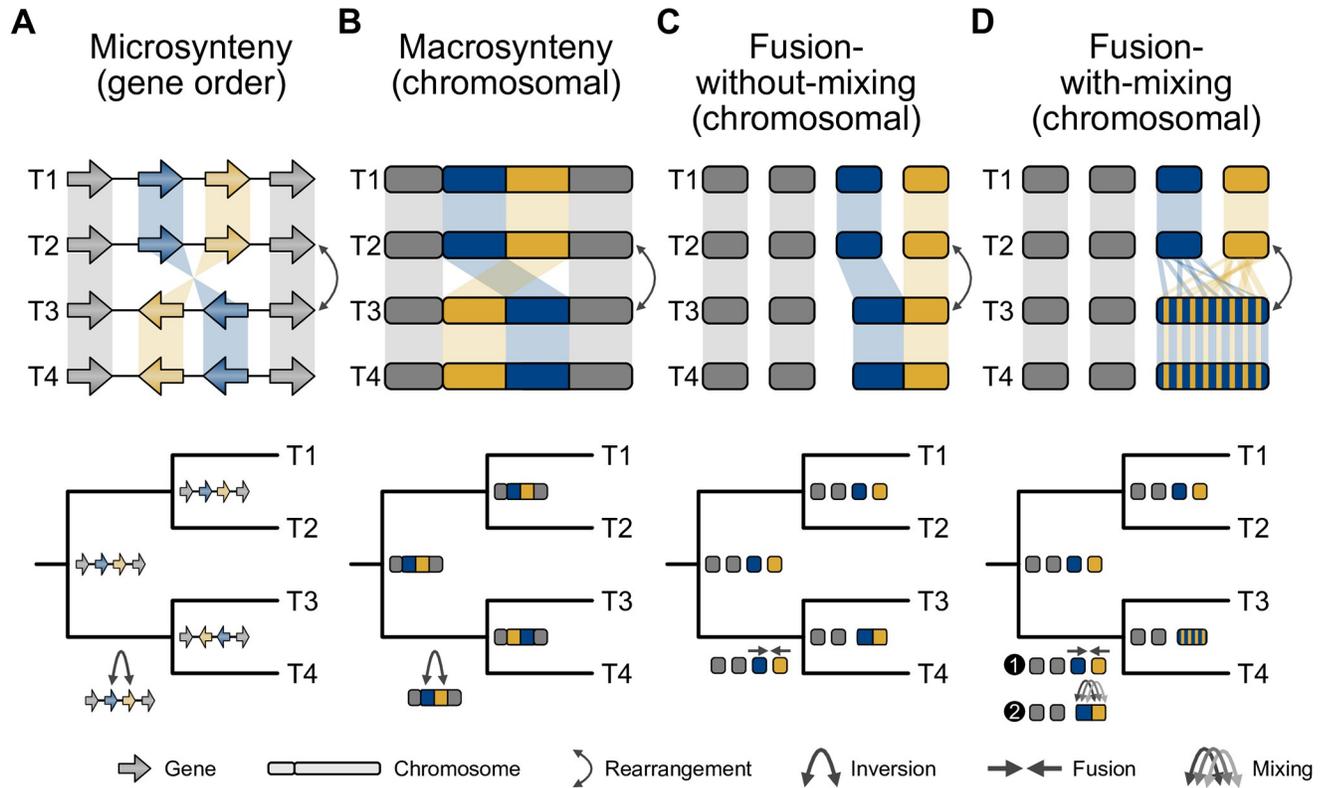
Thus, the utility of rare genomic changes has been mixed. Several examples demonstrate that rare genomic changes can recapitulate evolutionary history, while others contradict generally accepted evolutionary relationships established using other data types. Determining when and what rare genomic changes should be used has been hindered by the sparsity of methods for detecting rare genomic changes and algorithms for analyzing their informativeness.

Synteny emerges in the phylogenomic era

As abundant genome assemblies have become available and algorithm development has followed suit, the field of phylogenomics has become primed to revisit the value of rare genomic changes—specifically synteny—for phylogenetic inference. User-friendly software has enabled the detection of collinear DNA sequences in genomes from related organisms [59–64], thereby streamlining robust orthology inference [10] and analyses of changes in microsynteny and macrosynteny (Fig 2A and 2B). Shared rearrangements in gene order would be predicted to indicate a common evolutionary history, so long as convergence is not at play.

A major molecular mechanism driving syntenic variation is unequal homologous recombination [65]. Genomes with multiple copies of similar sequences, such as transposable elements in plant genomes, can be particularly prone to unequal homologous recombination [66]. Similarly, recombination between highly similar but nonallelic sequences (nonhomologous recombination) can result in major mutational events, such as recurrent deletions or duplications [67]. Other error-prone DNA repair mechanisms—including nonhomologous end joining—can also result in syntenic changes [68]. Whether a recombination event results in a microsyntenic or macrosyntenic change depends on the spacing between recombinant regions.

Saccharomycotina yeast have been a model lineage for developing and testing phylogenetic methods [69–71]. Comparison of the relationships among shared syntenic blocks in Saccharomycotina yeast with an evolutionary history previously inferred using concatenated multiple sequence alignments revealed that nearly 99% of microsyntenic blocks were more likely to be shared among closely related species than expected by random chance [72], reinforcing the notion that synteny can reflect phylogeny [73]. Subsequent developments in software and bioinformatic pipelines, vetted through simulations and examinations of empirical data, have facilitated the inference of organismal histories based on syntenic blocks [74–76]. Although promising, these studies primarily focused on establishing the utility of synteny through proof-of-principle approaches (i.e., reevaluating well-established relationships or using



E Transitions among syntenic states

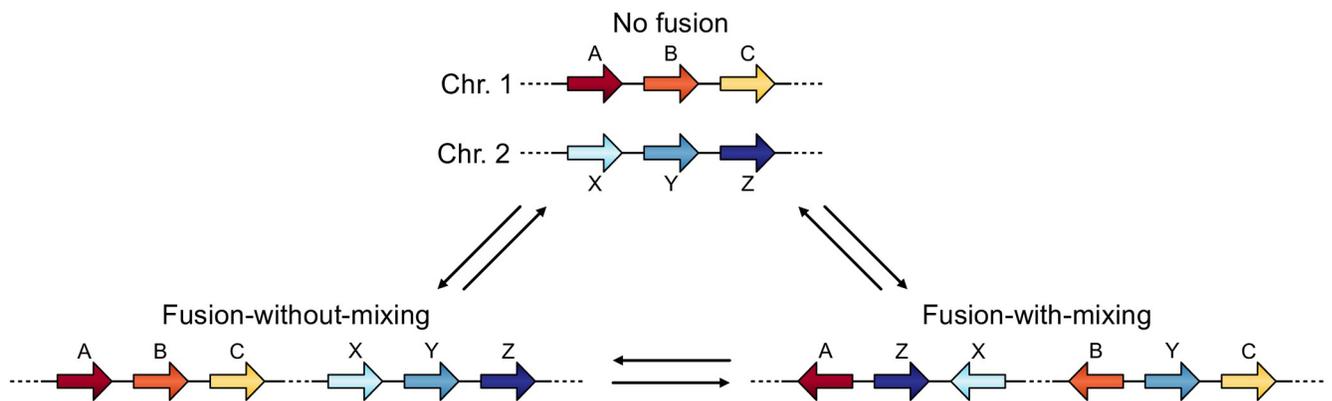


Fig 2. Data types for sequence-based phylogenetics. Consider the relationships among 4 taxa (represented as T1, T2, T3, and T4), wherein the pairs T1 and T2, and T3 and T4 are sister to one another. Changes in genome architecture can be examined at the scale of microsynteny (short stretches of orthologous loci; **A**) or macrosynteny (long stretches of orthologous loci; **B**). Changes in synteny can be described by different processes, such as fusion events without-mixing (**C**) and with-mixing (**D**). (**A**) In the case of microsynteny, evidence of an inversion may occur between the blue and orange loci (bottom), which happened in the ancestor of T3 and T4. (**B**) The same phenomenon can happen in the case of macrosynteny. (**C**) Fusion-without-mixing events between 2 chromosomes may also reflect phylogeny. In this case, a fusion event may have occurred in the ancestor between T3 and T4 (bottom). (**D**) Fusion-with-mixing can also be used to reconstruct phylogeny. Note, the evolutionary scenarios at the bottom of panels **A-D** depict only the most likely of many possible scenarios. (**E**) Fusion-with-mixing events may occur in 2 steps. First, there is a fusion event, then rearrangements occur, scrambling the order of genes that once were encoded on separate chromosomes. As a result, the probability of going from a “no fusion” to “fusion-without-mixing” state (and vice versa), and going from a “fusion-without-mixing” state to a “fusion-with-mixing” state, is relatively higher than going from a “fusion-with-mixing” to a “fusion-without-mixing” state. Transitioning directly from a “no fusion” to a “fusion-with-mixing” state is highly unlikely and may require an intermediate “fusion-without-mixing” state. Transition probabilities may vary depending on the underlying genome biology of the organism, the size of the syntenic region, and other parameters.

<https://doi.org/10.1371/journal.pbio.3002632.g002>

simulated scenarios). Applying these methods to address challenging tree of life debates has been a more recent development.

Synteny brings fresh perspectives to the tree of life

Synteny and the root of the animal tree. A recent reconstruction of ancient gene linkages has brought new data to bear on the sponge-first versus ctenophore-first debate at the base of the animal tree of life [9] (Fig 1F and 1G). This study relied on a new ensemble of genome assemblies from select sponges, ctenophores, bilaterians, cnidarians, and 3 outgroup taxa—a choanoflagellate (*Salpingoeca rosetta*), a filasterean (*Capsaspora owczarzaki*), and an ichthyosporean (*Creolimax fragrantissima*). Although detecting synteny among these genomes was complicated by the accumulation of chromosomal rearrangements across deep time, comparative analyses identified syntenic blocks conserved between outgroup and animal taxa using 3-way or 4-way reciprocal best BLAST hits; 29 and 20 different syntenic blocks were shared between animals and the filasterean or choanoflagellate, respectively. Notably, all 20 syntenic regions identified in the choanoflagellate were also present in the filasterean.

The inferred evolutionary changes to otherwise conserved syntenic blocks were placed in 1 of 3 categories based on outgroup taxa—no fusion, fusion-without-mixing, and fusion-with-mixing (Fig 2C and 2D)—which were then encoded and utilized in a phylogenetic framework. “No fusion” referred to syntenic blocks that remain on separate chromosomes. For example, imagine that an ancestral organism contains genes A, B, and C on 1 chromosome and genes X, Y, and Z on another (Fig 2E). If these blocks are on separate chromosomes (chromosomes 1 and 2) in 2 descendent organisms, there was “no fusion.” In the case of “fusion-without-mixing,” syntenic blocks A and B now coexist on the same chromosome in a descendent genome compared to the ancestor. This phenomenon is relatively well documented among acrocentric chromosomes in humans, which can fuse via a Robertsonian translocation [77]. Finally, “fusion-with-mixing” refers to a rearrangement pattern involving multiple steps between the ancestral genome and the descendent genome; first, chromosomal fusion, followed by one or more rearrangements that cause the syntenic blocks to interweave. For example, a single chromosome might contain a contiguous stretch of DNA encoding genes A, Z, X, B, Y, and C, in that order.

For reconstructing the animal tree of life, the codified matrix of fusion events was then used for phylogenetic inference. The transition probability of changing from a fusion-with-mixing state to another state (i.e., fusion or fission state) was inferred to be unlikely (Fig 2E). Bayesian analysis of this data matrix supported the ctenophore-first hypothesis, as did direct examination of fusions analyzed using parsimony [9]. Specifically, the ctenophore-first hypothesis was supported by 7 fusion events shared by bilaterians, cnidarians, and sponges, but that were missing from extant ctenophores and outgroup taxa. Four of these events occurred with mixing; under the sponge-first hypothesis, convergent fusions-with-mixing or precise reversions are required to explain these data. Thus, the absence of these fusions from ctenophores and outgroup taxa (except variation in region 7) was interpreted as evidence that ctenophores diverged from all other animals before the fusion and mixing events (Fig 3A). Region 7 may have independently undergone fusion and mixing events in the Filasterean lineage. An alternative but less likely scenario is that region 7 was already in a “mixed” state in the ancestor of all sampled taxa and subsequently underwent demixing and defusion events, followed by a complex pattern of fusion and mixing events.

Nonetheless, other findings from the synteny analysis contradict well-established evolutionary relationships. For example, despite phylogenomic analyses robustly supporting choanoflagellates as the closest living relatives of animals [78–82], animals shared more syntenic blocks

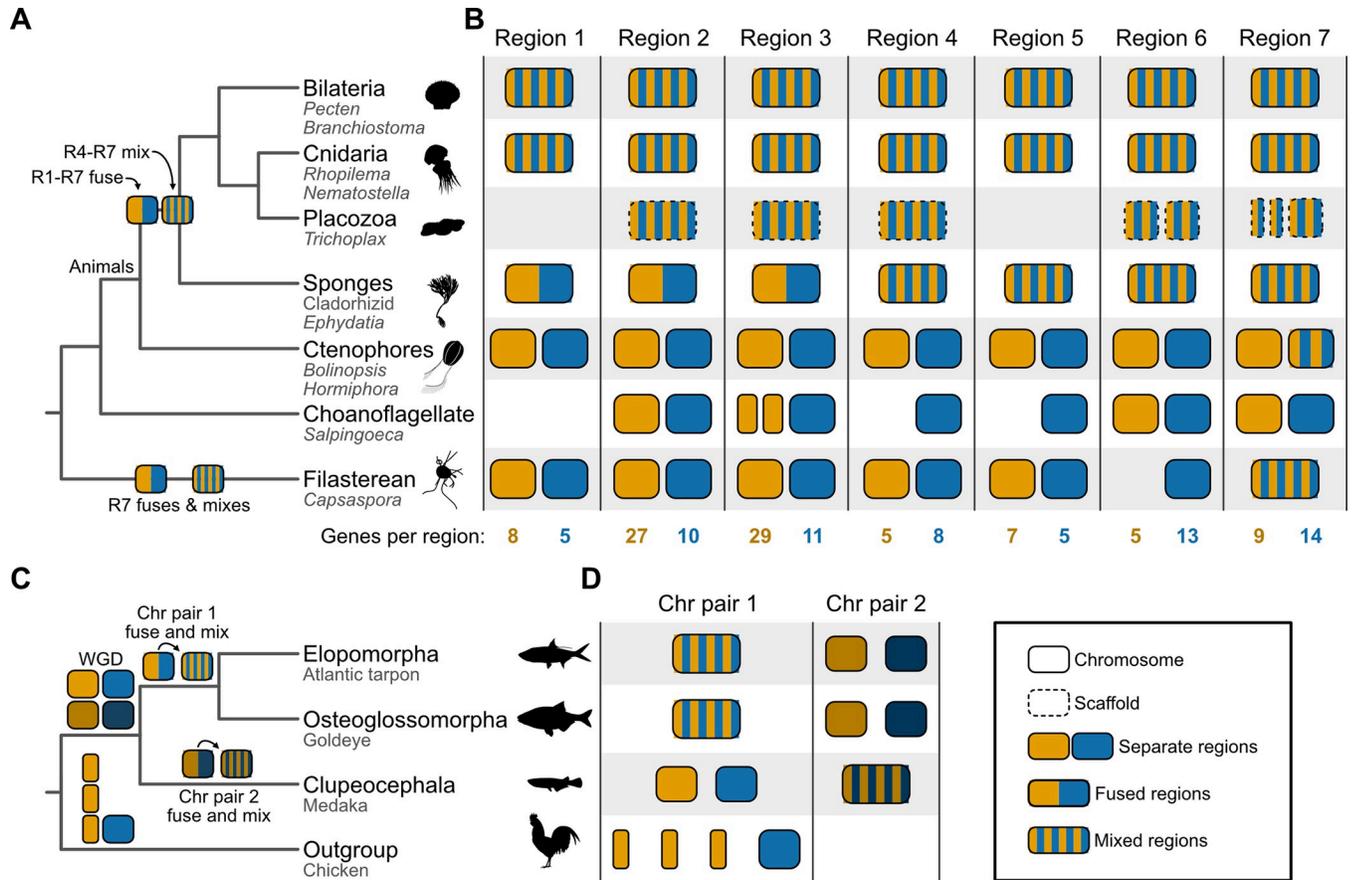


Fig 3. Summary depictions of synteny supporting the ctenophore-first and EO-sister hypotheses. (A) Inferred phylogeny of animal and outgroup taxa used to examine the root of the animal tree. Under the ctenophore-first hypothesis, regions 1–7 each resulted from fusion events between 2 distinct chromosomes. The syntenic block depicted in orange for region 3 underwent a fission event in the choanoflagellate lineage, resulting in 2 chromosomes. Regions 4–7 underwent subsequent mixing events. Underneath each higher-order lineage name, the names of representatives used in the study [9] are listed. For example, among Bilateria, species from the genera *Pecten* and *Branchiostoma* were included in the study. Note, only fusion and mixing events relevant to rooting the animal tree are depicted. (B) Patterns of synteny in 7 different regions most parsimoniously support the ctenophore-first hypothesis. Examination of these regions indicates that all underwent fusion events and 4 also underwent mixing events. Each region is abbreviated as “R” along the phylogeny (for example, R1 refers to region 1). The number of genes in each syntenic region is listed at the bottom of the panel. (C) Inferred phylogeny of the 3 teleost fish groups, including an outgroup taxon (the chicken). Cartoon summary drawings of chromosomes are included for representative species. Common names of these species are provided below the taxonomic names. Highly contiguous genome assemblies facilitated the detection of chromosome fusing and mixing events after a whole genome duplication event. Chr, chromosome. (D) Chromosomes observed in extant species are depicted as cartoon summaries. Duplicated chromosomes from a whole genome duplication event are darkened. Silhouette images were obtained from PhyloPic (<https://www.phylopic.org>) and are dedicated to the public domain; all credit goes to their respective contributors.

<https://doi.org/10.1371/journal.pbio.3002632.g003>

with the filasterean than with the choanoflagellate (29 syntenic blocks compared to 20). There are also more unique syntenic blocks shared between the filasterean and animals than with the choanoflagellate (9 syntenic blocks compared to 2). The incongruence between the pattern of synteny conservation and prior findings from phylogenomics either suggests a previously undetected close evolutionary relationship between filastereans and animals or, more likely, a lineage-specific loss of synteny in choanoflagellates.

Indeed, some choanoflagellates have undergone unique, accelerated genome evolution. Specifically, the choanoflagellate *S. rosetta* (used in [9]) has experienced rapid gene family evolution compared with other choanoflagellates, resulting in a reduced gene repertoire relative to that of the last common ancestor of animals and choanoflagellates [83]. Accordingly, *S. rosetta*

may not be the best representative of choanoflagellates for phylogenetics, highlighting the importance of expanded taxon sampling.

Similarly, unbiased phylogenetic analysis of fusion states did not recover the monophyly of Porifera, which contradicts more recent phylogenomic studies supporting the monophyly of the lineage [24,25,84]. Although some analyses support paraphyly among Porifera [85,86], the exemplar sponges in the study [9] belong to the class Demospongiae, which most analyses support as a monophyletic clade [87]. These observations call for caution in using syntenic blocks, especially when synteny has been lost.

Synteny and the evolutionary relationships among major groups of teleost fish. Early branching patterns in the teleost fish phylogeny were also recently reexamined [8] using a combination of expanded taxon sampling and analysis of syntenic blocks. Synteny was detected using the position of orthologous genes along chromosomes for every pairwise comparison of species. Phylogenetic analyses of the resulting macrosynteny and microsynteny data (Fig 2A and 2B)—wherein lack of syntenic conservation was used to measure distance—supported the EO-sister hypothesis. Using macrosynteny, nearly 20% of breakpoints supported the EO-sister hypothesis, and using microsynteny data, the sister relationship between these lineages received full bootstrap support. Evidence of a single chromosome fusion event unique to slim-headed and bony-tongued fish and another unique to other teleosts corroborated the EO-sister hypothesis; specifically, after a whole genome duplication event along the stem lineage of teleosts, 1 chromosome pair fused among slim-headed and bony-tongued fish, whereas the other chromosome pair fused and mixed among other teleosts (Fig 3C and 3D).

In addition to synteny-based analyses, standard phylogenomic approaches based on sequence data were employed. Phylogenomic analyses and distributions of support frequencies based on analyses of single genes supported the EO-sister hypothesis (Fig 1K) [8]. Interestingly, this finding was not supported by previous studies examining single-gene support frequencies and ultraconserved elements under a maximum likelihood framework [88,89]. Thus, with this expanded set of taxa, the EO-sister hypothesis is supported by synteny analysis as well as by gene sequence concatenation and coalescence, pointing to the influence of expanded taxon sampling.

Analyzing data from more taxa generally improves phylogenetic inference, particularly among close relatives of phylogenetically unstable taxa [3,90,91]. For example, when represented by a single taxon, the placement of the Saccharomycotina family Ascoideaceae conflicted between 2 phylogenomic studies that likely did not suffer from insufficient locus sampling [92,93]. However, expanded sampling of genomes from 3 Ascoideaceae and close relatives robustly supported 1 hypothesis [94]. Additional analyses suggested that increased taxon sampling resulted in improved model fit and greater phylogenetic stability of focal lineages. These studies demonstrate how additional taxon sampling can improve phylogenetic inference. Moreover, the benefits of high-quality, chromosome-scale genome assemblies are multi-fold. For example, standard phylogenomic analyses have benefitted from synteny data to improve orthology predictions, and multiple data types, such as patterns of macrosynteny and microsynteny, provide additional lines of evidence for phylogenomic inquiry [95].

Toward high-quality synteny-based tree of life reconstructions

As highly contiguous genome assemblies become more commonplace, our understanding of synteny as a phylogenomic marker will mature. Here, we provide a roadmap of research opportunities and identify challenges that will shape the use of synteny as a phylogenomic character (Fig 4A).

Considerations for inferring synteny-based phylogenies.

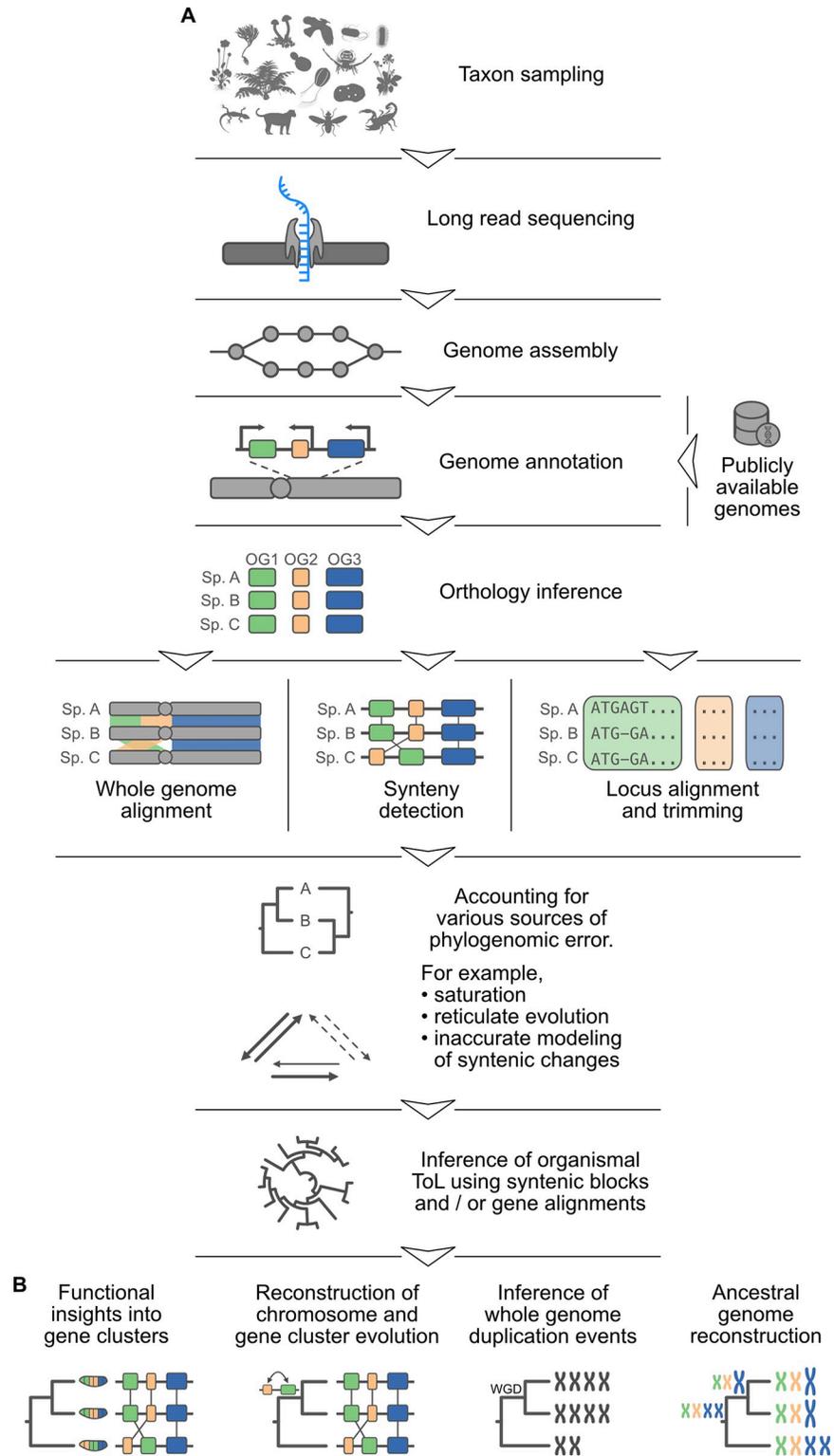


Fig 4. A roadmap of challenges and opportunities for syntenic phylogenomics. (A) A high-level summary of steps toward best practices in syntenic phylogenomics. Limitations in resource availability (computational power and researcher time) dictate that each project begins with a selection of taxa that are most relevant to the phylogenetic question at hand. For those taxa that lack high-quality genome assemblies, it will be necessary to sequence each genome (using long-read sequencing technology) and assemble the reads. In other cases, previously sequenced and

assembled genomes may be publicly available. In either case, the next step is to annotate the genes in all selected genomes using a single high-quality annotation method. Comparisons among the gene complements of each organism should then be used to identify gene orthologs (orthologous loci are depicted in green, yellow, and blue). Orthologs can then be used in whole genome alignment and synteny detection. In addition, alignments of orthologs can be trimmed, assembled into multiple sequence alignments, and used for traditional phylogenomics. After accounting for various sources of error, synteny blocks and multiple sequence alignments can be used to infer the topology of the tree of life. Note that obstacles in one step may be overcome by backtracking in the roadmap; for example, insufficient genome assembly completeness may benefit from additional genome sequencing. (B) Synteny data and organismal histories can be used for numerous research opportunities, including a better understanding of gene cluster function and evolution, reconstructing chromosome evolution, and inferring whole genome duplication events and ancestral genomes. For functional insights into gene clusters, fly embryos are depicted alongside gene clusters indicating how gene cluster organization may influence fly development. Silhouette images were obtained from PhyloPic (<https://www.phylopic.org>) and are dedicated to the public domain. Additional icons were obtained from bioicons (<https://bioicons.com>) and are available according to the CC-BY 4.0 license. Credit for silhouette images and icons goes to their respective contributors.

<https://doi.org/10.1371/journal.pbio.3002632.g004>

Taxon sampling/selection. Taxon sampling influences numerous downstream steps, such as orthology inference. Generally, the more taxa sampled, the better [3,90]. Selection of outgroup taxa can also influence phylogenomic inference; for example, the root of the animal tree is heavily influenced by the taxa selected [29]. Therefore, outgroup taxa should be carefully selected. Fortunately, there are a growing number of chromosome-level or highly contiguous genome assemblies that are publicly available for downloading and analysis. However, representatives from undersampled lineages may require genome sequencing. Thus, taxon sampling should be guided by the phylogenetic question at hand. For example, determining evolutionary relationships among vertebrates does not require taxon sampling among fungi; in fact, poor taxon sampling of distantly related taxa may introduce long branches and contribute to long-branch attraction artifacts [96,97].

Long-read sequencing and chromosomal conformation analyses. Much like traditional phylogenomics using collections of multiple sequence alignments, synteny-based phylogenomics starts with data acquisition. However, unlike multiple sequence alignment-based phylogenomics, high-quality genomes (ideally assembled accurately from telomere-to-telomere on all chromosomes) are necessary. The state of the art for genome assembly requires long-read sequencing (e.g., using Oxford Nanopore or PacBio) [98,99], which, in turn, requires acquisition of high-molecular weight DNA from each organism to be sequenced. For more complex genomes, chromosomal interactions detected from Hi-C analyses will help provide additional lines of evidence for subsequent steps, namely, genome assembly [100].

Genome assembly. With long-read sequences and chromosomal conformation data in hand, the next step for synteny-based phylogenomics is to generate an accurate and precise genome for each species to be analyzed. Poor genome assembly quality can be a source of error when detecting synteny [101] and, in turn, introduce errors in synteny-based phylogenomics. While there is no broadly accepted definition of a “high-quality” assembly, researchers should consider 3 important metrics: completeness, contiguity, and accuracy. Completeness can be assessed by comparing inferred gene content with expectations from transcriptome sequences and the presence/absence of nearly-universal single-copy orthologs [102]. Incomplete genomes may be difficult to incorporate into synteny-based phylogenomics and may necessitate further efforts to improve the original genome assembly. When highly contiguous genomes are difficult to achieve, macrosyntenic blocks that are broken up across several scaffolds should be removed from the data matrix. Alternatively, microsynteny may be more appropriate to use because they are more likely to be preserved, even in a discontinuous genome assembly. Examining assembly accuracy is difficult without physical mapping data from, for example, fluorescence in situ hybridization or optical maps [103]. However, these

data can be useful, not only to validate, but also to improve genome assembly quality, even helping achieve near-complete genomes [103]. Of note, other measures of assembly quality, such as degree of contamination, should be taken into account, particularly when loss of synteny is inferred.

Genome annotation. To detect syntenic blocks across the resulting set of genomes, the relative positions of orthologous genes are often used [72,76]. Thus, phylogeneticists must predict gene boundaries accurately to prevent, for example, erroneously combining 2 genes into a single gene model or missing genes entirely (Fig 4A). Many phylogenomic studies rely on the outputs of genomes annotated using different methods, but recent studies have shown that the outputs of different gene annotation methods can vary substantially [104]. A troubling result of comparing genomes annotated using different annotation methods is the artifactual inflation of the number of unique or lineage-specific genes [104]. Therefore, a single high-quality annotation method trained on the individual organism, or methods that combine the results from multiple gene annotation algorithms, like EVIDENCEModeler [105], may prove helpful. Moreover, incorporating transcriptomic reads will help refine and provide evidence for gene boundary predictions [106].

Orthology inference. The resulting gene predictions are subsequently used to infer orthologous relationships among genes (Fig 4A). Orthology relationships are inferred using all-versus-all sequence similarity information [107]. Researchers face several challenges during orthology inference, stemming from both analytical and biological sources of error [3,108]. Analytical errors may stem from genes that are absent from annotation predictions but that are genuinely encoded in the organism's genome. Other sources of incongruence between the evolutionary history of loci and the species may stem from complex evolutionary histories, such as gene duplication and loss, convergence, or saturation [3,109].

Alternatively, whole-genome alignment methods, like Progressive Cactus and SibekliaZ [110,111], may overcome potential errors stemming from gene annotation errors. One major innovation offered by Progressive Cactus is that it allows reference-free multiple genome alignment (ameliorating reference-based bias) and detecting multicopy orthology relationships, rather than only single-copy orthology [111]. Furthermore, Progressive Cactus can also handle large datasets, such as 600 or more animal genomes.

Establishing best practices in synteny detection

Typically, the distributions of gene orthologs along chromosomes in different species are used to detect potential syntenic blocks. Therefore, differences in the quality of ortholog prediction and in the density of syntenic orthologs detected should profoundly shape the accuracy of syntenic block detection. Both factors—accuracy of ortholog detection and density of syntenic orthologs—will likely drop off when comparing genomes separated by long evolutionary time scales.

Care must be taken, therefore, in the selection of software and analysis parameters [101]. Two key parameters are the minimum number and density of genes necessary to define orthologous syntenic blocks. Higher thresholds are expected to result in more conservative estimates of syntenic blocks (i.e., fewer false positives), but at the cost of potentially having a smaller number of syntenic blocks to analyze. Several software packages facilitate synteny detection, including MCScanX, SynChro, and syntenet [61–63]. Notably, each employs different methodology; for example, SynChro identifies pairwise syntenies using reciprocal best BLAST hits of protein sequence similarity, whereas MCScanX detects synteny blocks across 2 or more genomes [61,62]. MCScanX also provides additional utilities to further classify syntenic blocks based on putative evolutionary origins, such as those originating from whole genome

duplication events or tandem duplication. Although these algorithms vary in efficacy, genome discontinuity appears to be a major driver of error, underscoring the importance of obtaining highly contiguous genome assemblies [101].

To determine how much of the genome is captured during synteny detection, syntenic coverage can be calculated [101]. Syntenic coverage may differ between genomes due to biological phenomena such as genome size, content variation, or analytical factors that can relax the definition of a syntenic block; thus, it will be important to report syntenic coverage for individual genomes as well as summary statistics across them. Ideally, syntenic coverage will be high and cover nearly the entire genome for closely related organisms. However, syntenic coverage may be reduced depending on the threshold applied for detecting synteny, the rate of evolution among chromosomes, the rate of evolution of local gene order, and the evolutionary distance between species analyzed.

Accounting for sources of phylogenomic error/noise

Diverse factors can lead to erroneous species tree inference. Although these are well studied in analyses of multiple sequence alignments [3,108,112], they are underexplored for synteny-based phylogenomics. Here, we discuss potential sources of error/noise for synteny analysis and methods for taking them into account.

Saturation. In nucleotide and amino acid sequence evolution, when multiple, unobservable substitutions occur, the precise stepwise evolutionary history is difficult to trace; this phenomenon is described as “saturation.” Saturation may also occur during synteny evolution, whereby multiple sequential rearrangements may interfere with tracing the step-wise evolution of syntenic blocks. To overcome saturation, one solution may be to purge data matrices of rapidly evolving syntenic blocks, wherein the evolutionary history may be harder to trace.

Incomplete lineage sorting. The random sorting of ancestral polymorphisms can lead to genealogies that differ from the species tree, especially during rapid radiation events [6,113]. Incomplete lineage sorting among structural variants may also be a source of synteny-based phylogenomic noise. Incomplete lineage sorting among gene trees is particularly prevalent during radiation events and in large populations [113,114]. Given that genome rearrangement can occur rapidly in a population [115,116], it raises the possibility that some structural variants may coalesce before a speciation event, i.e., be subject to incomplete lineage sorting. Determining the prevalence (if any) of incomplete lineage sorting among structural variants will elucidate if incomplete lineage sorting is a source of incongruence.

Reticulate evolution. Reticulate evolution refers to nonvertical inheritance of loci, resulting in loci with an evolutionary history that deviates from a strictly bifurcating tree model, such as horizontal gene transfer and introgression/hybridization [117–119]. This issue will have varying influences across different lineages; for example, horizontal gene transfer occurs more frequently among Bacteria and Archaea than many eukaryotic lineages [120,121]. Similarly, hybridization is common among plant lineages [122–124] and has also been observed in other lineages, including animals and fungi [118,125–127].

The nonvertical acquisition of loci may interfere with the detection of otherwise conserved syntenic regions [128]. In the case of horizontal gene transfer, synteny analysis would suggest an erroneous phylogenetic placement of a lineage; for example, synteny analysis of the horizontally acquired bacterial siderophore gene cluster in yeast [129] would suggest a close affinity between yeast and Bacteria, a hypothesis that is incontrovertibly refuted. Loci with signatures of horizontal gene transfer can be pruned from a data matrix. However, in some cases, horizontally acquired loci that undergo vertical inheritance may be helpful markers for synteny-based phylogenomics [130].

Modeling syntenic changes. In standard molecular phylogenetics, substitution models approximate the evolutionary process of transitions between character states. These models vary in complexity and ability to capture biological reality [131–133]. Analogous substitution models for syntenic data have yet, to our knowledge, to be developed. However, structural variants can segregate among human populations [44] and recent developments of reference-free pangenomes may help facilitate their detection and illuminate their evolutionary dynamics [134], paving the way for creating models that capture exchange rates between syntenic states. The empirical determination of best practices for model selection will be important for future studies. Assuming overfitting is not an issue, highly parameterized models may be appropriate for synteny-based tree inference.

Other potential sources of error. Several other sources of error may come into play. For example, although few examples of convergent evolution in genome structure are known [135–137], they nonetheless demonstrate how independent rearrangements that result in the same structure could contribute to noise in synteny-based phylogenomics. Specifically, the currently accepted evolutionary relationships among the major rodent clades of Hystricomorpha (e.g., capybaras and naked-mole rats), Sciuromorpha (e.g., squirrels and marmots), and Myomorpha (e.g., rats and mice) indicate that Hystricomorpha diverged first and that Sciuromorpha and Myomorpha are sister lineages [136]. However, independent splitting events in the ortholog of human 3p21.31 in the Hystricomorpha (e.g., capybaras) and Sciuromorpha (e.g., squirrels) lineages would incorrectly suggest a sister relationship between each lineage [136]. Other sources of error may include an underpowered number of syntenic blocks and intraspecies heterogeneity in karyotype and chromosome structure due to, for example, Robertsonian translocations and copy number variants [77,115].

For phylogenomic analyses based on collections of multiple sequence alignments, researchers have demonstrated that not all loci have equal phylogenetic information. For example, genes displaying a clock-like pattern of evolution have often been favored for divergence-time analysis [138–140]. Measures have been developed to quantify the information encoded in multiple sequence alignments and phylogenetic trees inferred from them. Fortunately, some methods may be easily adapted to synteny data. For example, treeness [141] may help identify syntenic blocks with robust phylogenetic signal. Similarly, rogue taxa can be pruned from a data matrix [91]. Developing methods to measure the phylogenetic informativeness of different syntenic blocks will help increase signal-to-noise ratios among datasets and aid in refining their usage and interpretation within phylogenomic analyses.

Research opportunities using synteny data and species trees

Developing best practices for accurate synteny-based phylogenomics will help address current gaps in our understanding of genome evolution. For example, not only will synteny-based phylogenomics offer a new perspective for tree of life reconstructions, but the underlying synteny data may help provide functional insights into gene clusters (Fig 4B). Synteny-based phylogenomics will also help trace the evolution of chromosomes and gene clusters along phylogenies. Such reconstructions will help identify whole genome duplication events, which have been of longstanding interest to biologists because they provide fodder for molecular innovation, such as functional divergence of the resultant ohnologs [142,143].

Synteny-based phylogenomics may also facilitate ancestral genome reconstruction, potentially enabling (near) reference-level assemblies given sufficient sequenced and assembled genomes from extant species (Fig 4B). Accurate reconstructions of ancestral genomes, coupled with ancient DNA sequencing, may help resurrect the genomes of extinct lineages. More

broadly, a complete understanding of synteny evolution across time and species will contribute to a unified theory of genome architecture evolution.

While these opportunities present only a few exciting research prospects, phylogeneticists must first prioritize evaluating the efficacy of synteny-based phylogenomics for reconstructing ancient and recent divergences, spanning species and populations.

Conclusions

Improvements in genome sequencing, assembly, and annotation have revolutionized the quest to reconstruct the tree of life. With cutting-edge technologies and algorithms that enable the inference of highly contiguous genomes, synteny has reemerged as a powerful character for tree of life inquiries. Two studies tackling longstanding debates in animal phylogeny serve as notable case studies for demonstrating the potential utility and caveats of using synteny to reconstruct life's history [8,9]. These studies mark a new chapter, in which synteny-based phylogenomics promises to bring fresh insights, albeit after a series of technical challenges have been overcome. Tackling these challenges head-on will help shape best practices and deepen our understanding of synteny-based phylogenomics.

It is unlikely that Sturtevant and Dobzhansky, pioneers of their time in the 1930s, could have foreseen the far-reaching implications of their work on synteny as a phylogenetic marker. Nonetheless, their efforts have laid the groundwork for discoveries that continue to unfold today, nearly a century later, as technological advances enable the realization of their ambition. Uniting phylogenomics with comparisons of genome architecture in a whole-evidence approach promises to illuminate the detailed topology of the tree of life.

Acknowledgments

JLS thanks Antonis Rokas, Xing-Xing Shen, and Yuanning Li for fruitful discussions about phylogenomics over the years. In particular, JLS thanks Dr. Rokas for teaching him much of what he knows about phylogenomics. JLS and NK thank Thibaut Brunet, Maxwell C Coyle, and Xing-Xing Shen for reading the manuscript and providing helpful comments and suggestions prior to submission.

Author Contributions

Conceptualization: Jacob L. Steenwyk.

Funding acquisition: Jacob L. Steenwyk.

Supervision: Nicole King.

Visualization: Jacob L. Steenwyk, Nicole King.

Writing – original draft: Jacob L. Steenwyk.

Writing – review & editing: Jacob L. Steenwyk, Nicole King.

References

1. Fitch WM, Margoliash E. Construction of Phylogenetic Trees: A method based on mutation distances as estimated from cytochrome *c* sequences is of general applicability. *Science*. 1967; 155:279–284. <https://doi.org/10.1126/science.155.3760.279> PMID: 5334057
2. Haggerty LS, Martin FJ, Fitzpatrick DA, McInerney JO. Gene and genome trees conflict at many levels. *Philos Trans R Soc Lond B*. 2009; 364:2209–2219. <https://doi.org/10.1098/rstb.2009.0042> PMID: 19571241
3. Steenwyk JL, Li Y, Zhou X, Shen X-X, Rokas A. Incongruence in the phylogenomics era. *Nat Rev Genet*. 2023. <https://doi.org/10.1038/s41576-023-00620-x> PMID: 37369847

4. Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, et al. The bonobo genome compared with the chimpanzee and human genomes. *Nature*. 2012; 486:527–531. <https://doi.org/10.1038/nature11128> PMID: 22722832
5. King N, Rokas A. Embracing Uncertainty in Reconstructing Early Animal Evolution. *Curr Biol*. 2017; 27:R1081–R1088. <https://doi.org/10.1016/j.cub.2017.08.054> PMID: 29017048
6. Feng S, Bai M, Rivas-González I, Li C, Liu S, Tong Y, et al. Incomplete lineage sorting and phenotypic evolution in marsupials. *Cell*. 2022; 185:1646–1660.e18. <https://doi.org/10.1016/j.cell.2022.03.034> PMID: 35447073
7. Scornavacca C, Galtier N. Incomplete Lineage Sorting in Mammalian Phylogenomics. *Syst Biol*. 2016; syw082. <https://doi.org/10.1093/sysbio/syw082> PMID: 28173480
8. Parey E, Louis A, Montfort J, Bouchez O, Roques C, lampietro C, et al. Genome structures resolve the early diversification of teleost fishes. *Science*. 2023; 379:572–575. <https://doi.org/10.1126/science.abq4257> PMID: 36758078
9. Schultz DT, Haddock SHD, Bredeson JV, Green RE, Simakov O, Rokhsar DS. Ancient gene linkages support ctenophores as sister to other animals. *Nature*. 2023 [cited 21 May 2023]. <https://doi.org/10.1038/s41586-023-05936-6> PMID: 37198475
10. Rokas A, Williams BL, King N, Carroll SB. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*. 2003; 425:798–804. <https://doi.org/10.1038/nature02053> PMID: 14574403
11. Kapli P, Yang Z, Telford MJ. Phylogenetic tree building in the genomic age. *Nat Rev Genet* 2020; 21:428–444. <https://doi.org/10.1038/s41576-020-0233-0> PMID: 32424311
12. Philippe H, Lartillot N, Brinkmann H. Multigene Analyses of Bilaterian Animals Corroborate the Monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol*. 2005; 22:1246–1253. <https://doi.org/10.1093/molbev/msi111> PMID: 15703236
13. Giribet G, Edgecombe GD. Current Understanding of Ecdysozoa and its Internal Phylogenetic Relationships. *Integr Comp Biol*. 2017; 57:455–466. <https://doi.org/10.1093/icb/ix072> PMID: 28957525
14. Crotty SM, Minh BQ, Bean NG, Holland BR, Tuke J, Jermini LS, et al. GHOST: Recovering Historical Signal from Heterotachously Evolved Sequence Alignments. Smith S, editor. *Syst Biol*. 2019; syz051. <https://doi.org/10.1093/sysbio/syz051> PMID: 31364711
15. Williams TA, Cox CJ, Foster PG, Szöllösi GJ, Embley TM. Phylogenomics provides robust support for a two-domains tree of life. *Nat Ecol Evol*. 2019; 4:138–147. <https://doi.org/10.1038/s41559-019-1040-x> PMID: 31819234
16. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*. 2014; 346:1320–1331. <https://doi.org/10.1126/science.1253451> PMID: 25504713
17. Choi B, Crisp MD, Cook LG, Meusemann K, Edwards RD, Toon A, et al. Identifying genetic markers for a range of phylogenetic utility—From species to family level. Brewer MS, editor. *PLoS ONE*. 2019; 14:e0218995. <https://doi.org/10.1371/journal.pone.0218995> PMID: 31369563
18. Debray K, Marie-Magdelaine J, Ruttink T, Clotault J, Foucher F, Malécot V. Identification and assessment of variable single-copy orthologous (SCO) nuclear loci for low-level phylogenomics: a case study in the genus *Rosa* (Rosaceae). *BMC Evol Biol*. 2019; 19:152. <https://doi.org/10.1186/s12862-019-1479-z> PMID: 31340752
19. Wainright PO, Hinkle G, Sogin ML, Stickel SK. Monophyletic Origins of the Metazoa: an Evolutionary Link with Fungi. *Science*. 1993; 260:340–342. <https://doi.org/10.1126/science.8469985> PMID: 8469985
20. Brusca RC, Brusca GJ. *Invertebrates*. Sinauer Associates Incorporated; 2002.
21. Collins AG. Evaluating multiple alternative hypotheses for the origin of Bilateria: An analysis of 18S rRNA molecular evidence. *Proc Natl Acad Sci U S A*. 1998; 95:15458–15463. <https://doi.org/10.1073/pnas.95.26.15458> PMID: 9860990
22. Medina M, Collins AG, Silberman JD, Sogin ML. Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. *Proc Natl Acad Sci U S A*. 2001; 98:9707–9712. <https://doi.org/10.1073/pnas.171316998> PMID: 11504944
23. Podar M, Haddock SHD, Sogin ML, Harbison GR. A Molecular Phylogenetic Framework for the Phylum Ctenophora Using 18S rRNA Genes. *Mol Phylogenet Evol*. 2001; 21:218–230. <https://doi.org/10.1006/mpev.2001.1036> PMID: 11697917
24. Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, et al. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*. 2008; 452:745–749. <https://doi.org/10.1038/nature06614> PMID: 18322464

25. Philippe H, Derelle R, Lopez P, Pick K, Borchellini C, Boury-Esnault N, et al. Phylogenomics Revives Traditional Views on Deep Animal Relationships. *Curr Biol*. 2009; 19:706–712. <https://doi.org/10.1016/j.cub.2009.02.052> PMID: 19345102
26. Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, et al. A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Curr Biol*. 2017; 27:958–967. <https://doi.org/10.1016/j.cub.2017.02.031> PMID: 28318975
27. Shen X-X, Hittinger CT, Rokas A. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat Ecol Evol*. 2017; 1:0126. <https://doi.org/10.1038/s41559-017-0126-28812701> PMID: 28812701
28. Whelan NV, Kocot KM, Moroz TP, Mukherjee K, Williams P, Paulay G, et al. Ctenophore relationships and their placement as the sister group to all other animals. *Nat Ecol Evol*. 2017; 1:1737–1746. <https://doi.org/10.1038/s41559-017-0331-3> PMID: 28993654
29. Li Y, Shen X-X, Evans B, Dunn CW, Rokas A. Rooting the Animal Tree of Life. Tamura K, editor. *Mol Biol Evol*. 2021; 38:4322–4333. <https://doi.org/10.1093/molbev/msab170> PMID: 34097041
30. Whelan NV, Halanych KM. Available data do not rule out Ctenophora as the sister group to all other Metazoa. *Nat Commun*. 2023; 14:711. <https://doi.org/10.1038/s41467-023-36151-6> PMID: 36765046
31. Le HLV, Lecointre G, Perasso R. A 28S rRNA-Based Phylogeny of the Gnathostomes: First Steps in the Analysis of Conflict and Congruence with Morphologically Based Cladograms. *Mol Phylogenet Evol*. 1993; 2:31–51. <https://doi.org/10.1006/mpev.1993.1005> PMID: 8081546
32. Dornburg A, Near TJ. The emerging phylogenetic perspective on the evolution of actinopterygian fishes. *Annu Rev Ecol Evol Syst*. 2021; 52:427–452.
33. Rokas A, Holland PWH. Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol*. 2000; 15:454–459. [https://doi.org/10.1016/s0169-5347\(00\)01967-4](https://doi.org/10.1016/s0169-5347(00)01967-4) PMID: 11050348
34. Castresana J, Feldmaier-Fuchs G, Yokobori S, Satoh N, Pääbo S. The Mitochondrial Genome of the Hemichordate *Balanoglossus carnosus* and the Evolution of Deuterostome Mitochondria. *Genetics*. 1998; 150:1115–1123. <https://doi.org/10.1093/genetics/150.3.1115> PMID: 9799263
35. Venkatesh B, Ning Y, Brenner S. Late changes in spliceosomal introns define clades in vertebrate evolution. *Proc Natl Acad Sci U S A*. 1999; 96:10267–10271. <https://doi.org/10.1073/pnas.96.18.10267> PMID: 10468597
36. Rokas A, Kathirithamby J, Holland PWH. Intron insertion as a phylogenetic character: the engrailed homeobox of Strepsiptera does not indicate affinity with Diptera. *Insect Mol Biol*. 1999; 8:527–530. <https://doi.org/10.1046/j.1365-2583.1999.00149.x> PMID: 10620047
37. Steenwyk JL, Oplente DA, Kominek J, Shen X-X, Zhou X, Labella AL, et al. Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. Kamoun S, editor. *PLoS Biol*. 2019; 17:e3000255. <https://doi.org/10.1371/journal.pbio.3000255> PMID: 31112549
38. Sturtevant AH, Dobzhansky T. Inversions in the Third Chromosome of Wild Races of *Drosophila Pseudoobscura*, and Their Use in the Study of the History of the Species. *Proc Natl Acad Sci U S A*. 1936; 22:448–450. <https://doi.org/10.1073/pnas.22.7.448> PMID: 16577723
39. Dobzhansky T, Sturtevant AH. Inversions in the chromosomes of *Drosophila pseudoobscura*. *Genetics*. 1938; 23:28–64. <https://doi.org/10.1093/genetics/23.1.28> PMID: 17246876
40. Carson HL. Chromosomal sequences and interisland colonizations in hawaiian *Drosophila*. *Genetics*. 1983; 103:465–482. <https://doi.org/10.1093/genetics/103.3.465> PMID: 17246115
41. Steenwyk JL, Soghigian JS, Perfect JR, Gibbons JG. Copy number variation contributes to cryptic genetic variation in outbreak lineages of *Cryptococcus gattii* from the North American Pacific Northwest. *BMC Genomics*. 2016; 17:700. <https://doi.org/10.1186/s12864-016-3044-0> PMID: 27590805
42. Lee Y-L, Bosse M, Mullaart E, Groenen MAM, Veerkamp RF, Bouwman AC. Functional and population genetic features of copy number variations in two dairy cattle populations. *BMC Genomics*. 2020; 21:89. <https://doi.org/10.1186/s12864-020-6496-1> PMID: 31992181
43. Brown KH, Dobrinski KP, Lee AS, Gokcumen O, Mills RE, Shi X, et al. Extensive genetic diversity and substructuring among zebrafish strains revealed through copy number variant analysis. *Proc Natl Acad Sci U S A*. 2012; 109:529–534. <https://doi.org/10.1073/pnas.1112163109> PMID: 22203992
44. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015; 526:75–81. <https://doi.org/10.1038/nature15394> PMID: 26432246
45. Doronina L, Reising O, Clawson H, Ray DA, Schmitz J. True Homoplasmy of Retrotransposon Insertions in Primates. Susko E, editor. *Systematic Biology* 2019; 68:482–493. <https://doi.org/10.1093/sysbio/syy076> PMID: 30445649
46. Cloutier A, Sackton TB, Grayson P, Clamp M, Baker AJ, Edwards SV. Whole-Genome Analyses Resolve the Phylogeny of Flightless Birds (Palaeognathae) in the Presence of an Empirical Anomaly

- Zone. Faircloth B, editor. Syst Biol. 2019; 68:937–955. <https://doi.org/10.1093/sysbio/syz019> PMID: 31135914
47. Murphy WJ, Foley NM, Bredemeyer KR, Gatesy J, Springer MS. Phylogenomics and the Genetic Architecture of the Placental Mammal Radiation. *Annu Rev Anim Biosci.* 2021; 9:29–53. <https://doi.org/10.1146/annurev-animal-061220-023149> PMID: 33228377
 48. Takahashi K, Terai Y, Nishida M, Okada N. A novel family of short interspersed repetitive elements (SINEs) from cichlids: the patterns of insertion of SINEs at orthologous loci support the proposed monophyly of four major groups of cichlid fishes in Lake Tanganyika. *Mol Biol Evol.* 1998; 15:391–407. <https://doi.org/10.1093/oxfordjournals.molbev.a025936> PMID: 9549090
 49. Takahashi K, Nishida M, Yuma M, Okada N. Retroposition of the AFC Family of SINEs (Short Interspersed Repetitive Elements) Before and During the Adaptive Radiation of Cichlid Fishes in Lake Malawi and Related Inferences About Phylogeny. *J Mol Evol.* 2001; 53:496–507. <https://doi.org/10.1007/s002390010240> PMID: 11675610
 50. Fortna A, Kim Y, MacLaren E, Marshall K, Hahn G, Meltesen L, et al. Lineage-Specific Gene Duplication and Loss in Human and Great Ape Evolution. Chris Tyler-Smith, editor. *PLoS Biol.* 2004; 2:e207. <https://doi.org/10.1371/journal.pbio.0020207> PMID: 15252450
 51. Mühlhausen S, Schmitt HD, Pan K-T, Plessmann U, Urlaub H, Hurst LD, et al. Endogenous Stochastic Decoding of the CUG Codon by Competing Ser- and Leu-tRNAs in *Ascoidea asiatica*. *Curr Biol.* 2018; 28:2046–2057.e5. <https://doi.org/10.1016/j.cub.2018.04.085> PMID: 29910077
 52. Mathews S, Donoghue MJ. The Root of Angiosperm Phylogeny Inferred from Duplicate Phytochrome Genes. *Science.* 1999; 286:947–950. <https://doi.org/10.1126/science.286.5441.947> PMID: 10542147
 53. One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature.* 2019; 574:679–685. <https://doi.org/10.1038/s41586-019-1693-2> PMID: 31645766
 54. Martín-Durán JM, Ryan JF, Vellutini BC, Pang K, Hejnal A. Increased taxon sampling reveals thousands of hidden orthologs in flatworms. *Genome Res.* 2017; 27:1263–1272. <https://doi.org/10.1101/gr.216226.116> PMID: 28400424
 55. Krassowski T, Coughlan AY, Shen X-X, Zhou X, Kominek J, Opulente DA, et al. Evolutionary instability of CUG-Leu in the genetic code of budding yeasts. *Nat Commun.* 2018; 9:1887. <https://doi.org/10.1038/s41467-018-04374-7> PMID: 29760453
 56. Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, et al. Mitochondrial genome of *Trichoplax adhaerens* supports Placozoa as the basal lower metazoan phylum. *Proc Natl Acad Sci U S A.* 2006; 103:8751–8756. <https://doi.org/10.1073/pnas.0602076103> PMID: 16731622
 57. Laumer CE, Gruber-Vodicka H, Hadfield MG, Pearse VB, Riesgo A, Marioni JC, et al. Support for a clade of Placozoa and Cnidaria in genes with minimal compositional bias. *eLife.* 2018; 7:e36278. <https://doi.org/10.7554/eLife.36278> PMID: 30373720
 58. Ding Y-M, Pang X-X, Cao Y, Zhang W-P, Renner SS, Zhang D-Y, et al. Genome structure-based Juglandaceae phylogenies contradict alignment-based phylogenies and substitution rates vary with DNA repair genes. *Nat Commun.* 2023; 14:617. <https://doi.org/10.1038/s41467-023-36247-z> PMID: 36739280
 59. Haas BJ, Delcher AL, Wortman JR, Salzberg SL. DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics.* 2004; 20:3643–3646. <https://doi.org/10.1093/bioinformatics/bth397> PMID: 15247098
 60. Proost S, Fostier J, De Witte D, Dhoedt B, Demeester P, Van de Peer Y, et al. i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* 2012; 40:e11–e11. <https://doi.org/10.1093/nar/gkr955> PMID: 22102584
 61. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012; 40:e49–e49. <https://doi.org/10.1093/nar/gkr1293> PMID: 22217600
 62. Drillon G, Carbone A, Fischer G. SynChro: a fast and easy tool to reconstruct and visualize synteny blocks along eukaryotic chromosomes. *PLoS ONE.* 2014; 9:e92621. <https://doi.org/10.1371/journal.pone.0092621> PMID: 24651407
 63. Almeida-Silva F, Zhao T, Ullrich KK, Schranz ME, Van De Peer Y. syntenet: an R/Bioconductor package for the inference and analysis of synteny networks. Martelli PL, editor. *Bioinformatics.* 2023;39:btac806. <https://doi.org/10.1093/bioinformatics/btac806> PMID: 36539202
 64. Mackintosh A, De La Rosa PMG, Martin SH, Lohse K, Laetsch DR. Inferring inter-chromosomal rearrangements and ancestral linkage groups from synteny. *Evol Biol.* 2023. <https://doi.org/10.1101/2023.09.17.558111>

65. Robberecht C, Voet T, Esteki MZ, Nowakowska BA, Vermeesch JR. Nonallelic homologous recombination between retrotransposable elements is a driver of de novo unbalanced translocations. *Genome Res.* 2013; 23:411–418. <https://doi.org/10.1101/gr.145631.112> PMID: 23212949
66. Ma J, Bennetzen JL. Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice. *Proc Natl Acad Sci U S A.* 2006; 103:383–388. <https://doi.org/10.1073/pnas.0509810102> PMID: 16381819
67. Liu P, Lacaria M, Zhang F, Withers M, Hastings PJ, Lupski JR. Frequency of Nonallelic Homologous Recombination Is Correlated with Length of Homology: Evidence that Ectopic Synapsis Precedes Ectopic Crossing-Over. *Am J Hum Genet.* 2011; 89:580–588. <https://doi.org/10.1016/j.ajhg.2011.09.009> PMID: 21981782
68. Ferguson S, Jones A, Murray K, Schwessinger B, Borevitz JO. Interspecies genome divergence is predominantly due to frequent small scale rearrangements in Eucalyptus. *Mol Ecol.* 2023; 32:1271–1287. <https://doi.org/10.1111/mec.16608> PMID: 35810343
69. Salichos L, Rokas A. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature.* 2013; 497:327–331. <https://doi.org/10.1038/nature12130> PMID: 23657258
70. Steenwyk JL, Goltz DC, Buida TJ, Li Y, Shen X-X, Rokas A. OrthoSNAP: A tree splitting and pruning algorithm for retrieving single-copy orthologs from gene family trees. Hejnal A, editor. *PLoS Biol.* 2022; 20:e3001827. <https://doi.org/10.1371/journal.pbio.3001827> PMID: 36228036
71. Steenwyk JL, Buida TJ, Li Y, Shen X-X, Rokas A. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. Hejnal A, editor. *PLoS Biol.* 2020; 18:e3001007. <https://doi.org/10.1371/journal.pbio.3001007> PMID: 33264284
72. Li Y, Liu H, Steenwyk JL, LaBella AL, Harrison M-C, Groenewald M, et al. Contrasting modes of macro and microsynteny evolution in a eukaryotic subphylum. *Curr Biol.* 2022; S0960982222016700. <https://doi.org/10.1016/j.cub.2022.10.025> PMID: 36334587
73. Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet.* 2005; 6:361–375. <https://doi.org/10.1038/nrg1603> PMID: 15861208
74. Zheng C, Sankoff D. Gene order in rosid phylogeny, inferred from pairwise syntenies among extant genomes. *BMC Bioinformatics.* 2012; 13:S9. <https://doi.org/10.1186/1471-2105-13-S10-S9> PMID: 22759433
75. Drillon G, Champeimont R, Oteri F, Fischer G, Carbone A. Phylogenetic Reconstruction Based on Synteny Block and Gene Adjacencies. Battistuzzi FU, editor. *Mol Biol Evol.* 2020; 37:2747–2762. <https://doi.org/10.1093/molbev/msaa114> PMID: 32384156
76. Zhao T, Zwaenepoel A, Xue J-Y, Kao S-M, Li Z, Schranz ME, et al. Whole-genome microsynteny-based phylogeny of angiosperms. *Nat Commun.* 2021; 12:3498. <https://doi.org/10.1038/s41467-021-23665-0> PMID: 34108452
77. Therman E, Susman B, Denniston C. The nonrandom participation of human acrocentric chromosomes in Robertsonian translocations. *Ann Hum Genet.* 1989; 53:49–65. <https://doi.org/10.1111/j.1469-1809.1989.tb01121.x> PMID: 2658738
78. Fairclough SR, Chen Z, Kramer E, Zeng Q, Young S, Robertson HM, et al. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol.* 2013; 14:R15. <https://doi.org/10.1186/gb-2013-14-2-r15> PMID: 23419129
79. King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, et al. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature.* 2008; 451:783–788. <https://doi.org/10.1038/nature06617> PMID: 18273011
80. Ocaña-Pallarès E, Williams TA, López-Escardó D, Arroyo AS, Pathmanathan JS, Baptiste E, et al. Divergent genomic trajectories predate the origin of animals and fungi. *Nature.* 2022; 609:747–753. <https://doi.org/10.1038/s41586-022-05110-4> PMID: 36002568
81. Torruella G, Derelle R, Paps J, Lang BF, Roger AJ, Shalchian-Tabrizi K, et al. Phylogenetic Relationships within the Opisthokonta Based on Phylogenomic Analyses of Conserved Single-Copy Protein Domains. *Mol Biol Evol.* 2012; 29:531–544. <https://doi.org/10.1093/molbev/msr185> PMID: 21771718
82. Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF. A Phylogenomic Investigation into the Origin of Metazoa. *Mol Biol Evol.* 2008; 25:664–672. <https://doi.org/10.1093/molbev/msn006> PMID: 18184723
83. Richter DJ, Fozouni P, Eisen MB, King N. Gene family innovation, conservation and loss on the animal stem lineage. *eLife.* 2018; 7:e34226. <https://doi.org/10.7554/eLife.34226> PMID: 29848444
84. Whelan NV, Kocot KM, Moroz LL, Halanych KM. Error, signal, and the placement of Ctenophora sister to all other animals. *Proc Natl Acad Sci U S A.* 2015; 112:5773–5778. <https://doi.org/10.1073/pnas.1503453112> PMID: 25902535

85. Sperling EA, Pisani D, Peterson KJ. Poriferan paraphyly and its implications for Precambrian palaeobiology. London: Geological Society, Special Publications; 2007. pp. 355–368. <https://doi.org/10.1144/SP286.25>
86. Borchiellini C, Manuel M, Alivon E, Boury-Esnault N, Vacelet J, Le Parco Y. Sponge paraphyly and the origin of Metazoa: Sponge paraphyly. *J Evol Biol.* 2001; 14:171–179. <https://doi.org/10.1046/j.1420-9101.2001.00244.x> PMID: 29280585
87. Kenny NJ, Francis WR, Rivera-Vicéns RE, Juravel K, De Mendoza A, Díez-Vives C, et al. Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge *Ephydatia muelleri*. *Nat Commun.* 2020; 11:3676. <https://doi.org/10.1038/s41467-020-17397-w> PMID: 32719321
88. Hughes LC, Ortí G, Huang Y, Sun Y, Baldwin CC, Thompson AW, et al. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc Natl Acad Sci U S A.* 2018; 115:6249–6254. <https://doi.org/10.1073/pnas.1719358115> PMID: 29760103
89. Faircloth BC, Sorenson L, Santini F, Alfaro ME. A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). Moreau CS, editor. *PLoS ONE.* 2013; 8:e65923. <https://doi.org/10.1371/journal.pone.0065923> PMID: 23824177
90. Pollock DD, Zwickl DJ, McGuire JA, Hillis DM. Increased Taxon Sampling Is Advantageous for Phylogenetic Inference. Crandall K, editor. *Syst Biol.* 2002; 51:664–671. <https://doi.org/10.1080/10635150290102357> PMID: 12228008
91. Aberer AJ, Krompass D, Stamatakis A. Pruning Rogue Taxa Improves Phylogenetic Accuracy: An Efficient Algorithm and Webservice. *Syst Biol.* 2013; 62:162–166. <https://doi.org/10.1093/sysbio/sys078> PMID: 22962004
92. Shen X-X, Zhou X, Kominek J, Kurtzman CP, Hittinger CT, Rokas A. Reconstructing the Backbone of the Saccharomycotina Yeast Phylogeny Using Genome-Scale Data. *G3.* 2016; 6:3927–3939. <https://doi.org/10.1534/g3.116.034744> PMID: 27672114
93. Riley R, Haridas S, Wolfe KH, Lopes MR, Hittinger CT, Göker M, et al. Comparative genomics of biotechnologically important yeasts. *Proc Natl Acad Sci U S A.* 2016; 113:9882–9887. <https://doi.org/10.1073/pnas.1603941113> PMID: 27535936
94. Shen X-X, Oplente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, et al. Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell.* 2018; 175:1533–1545.e20. <https://doi.org/10.1016/j.cell.2018.10.023> PMID: 30415838
95. Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature.* 2006; 440:341–345. <https://doi.org/10.1038/nature04562> PMID: 16541074
96. Pisani D, Pett W, Dohrmann M, Feuda R, Rota-Stabelli O, Philippe H, et al. Genomic data do not support comb jellies as the sister group to all other animals. *Proc Natl Acad Sci U S A.* 2015; 112:15402–15407. <https://doi.org/10.1073/pnas.1518127112> PMID: 26621703
97. Brinkmann H, Van Der Giezen M, Zhou Y, De Raucourt GP, Philippe H. An Empirical Assessment of Long-Branch Attraction Artefacts in Deep Eukaryotic Phylogenomics. Hedin M, editor. *Syst Biol.* 2005; 54:743–757. <https://doi.org/10.1080/10635150500234609> PMID: 16243762
98. Marx V. Method of the year: long-read sequencing. *Nat Methods.* 2023; 20:6–11. <https://doi.org/10.1038/s41592-022-01730-w> PMID: 36635542
99. Giani AM, Gallo GR, Gianfranceschi L, Formenti G. Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput Struct Biotechnol J.* 2020; 18:9–19. <https://doi.org/10.1016/j.csbj.2019.11.002> PMID: 31890139
100. Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods.* 2012; 58:268–276. <https://doi.org/10.1016/j.ymeth.2012.05.001> PMID: 22652625
101. Liu D, Hunt M, Tsai IJ. Inferring synteny between genome assemblies: a systematic evaluation. *BMC Bioinformatics.* 2018; 19:26. <https://doi.org/10.1186/s12859-018-2026-4> PMID: 29382321
102. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Kloutchnikov G, et al. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol.* 2018; 35:543–548. <https://doi.org/10.1093/molbev/msx319> PMID: 29220515
103. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature.* 2021; 592:737–746. <https://doi.org/10.1038/s41586-021-03451-0> PMID: 33911273
104. Weisman CM, Murray AW, Eddy SR. Mixing genome annotation methods in a comparative analysis inflates the apparent number of lineage-specific genes. *Curr Biol.* 2022; 32:2632–2639.e2. <https://doi.org/10.1016/j.cub.2022.04.085> PMID: 35588743

105. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 2008; 9:R7. <https://doi.org/10.1186/gb-2008-9-1-r7> PMID: 18190707
106. Rosato M, Hoelscher B, Lin Z, Agwu C, Xu F. Transcriptome analysis provides genome annotation and expression profiles in the central nervous system of *Lymnaea stagnalis* at different ages. *BMC Genomics.* 2021; 22:637. <https://doi.org/10.1186/s12864-021-07946-y> PMID: 34479505
107. Fernández R, Gabaldón T, Dessimoz C. Orthology: definitions, inference, and impact on species phylogeny inference. 2019 [cited 25 May 2023]. <https://doi.org/10.48550/ARXIV.1903.04530>
108. Philippe H, Vienne DMD, Ranwez V, Roure B, Baurain D, Delsuc F. Pitfalls in supermatrix phylogenomics. *Eur J Taxon.* 2017 [cited 5 Dec 2023]. <https://doi.org/10.5852/ejt.2017.283>
109. Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics.* 2012; 28:i409–i415. <https://doi.org/10.1093/bioinformatics/bts386> PMID: 22962460
110. Minkin I, Medvedev P. Scalable multiple whole-genome alignment and locally collinear block construction with SibeliaZ. *Nat Commun.* 2020; 11:6327. <https://doi.org/10.1038/s41467-020-19777-8> PMID: 33303762
111. Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature.* 2020; 587:246–251. <https://doi.org/10.1038/s41586-020-2871-y> PMID: 33177663
112. Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, et al. Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. Penny D, editor. *PLoS Biol.* 2011; 9:e1000602. <https://doi.org/10.1371/journal.pbio.1000602> PMID: 21423652
113. Maddison WP, Knowles LL. Inferring Phylogeny Despite Incomplete Lineage Sorting. Collins T, editor. *Syst Biol.* 2006; 55:21–30. <https://doi.org/10.1080/10635150500354928> PMID: 16507521
114. Avise JC, Robinson TJ. Hemipecty: A New Term in the Lexicon of Phylogenetics. Kubatko L, editor. *Syst Biol.* 2008; 57:503–507. <https://doi.org/10.1080/10635150802164587> PMID: 18570042
115. Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun.* 2017; 8:14061. <https://doi.org/10.1038/ncomms14061> PMID: 28117401
116. Steenwyk J, Rokas A. Extensive Copy Number Variation in Fermentation-Related Genes Among *Saccharomyces cerevisiae* Wine Strains. *G3.* 2017; 7:1475–1485. <https://doi.org/10.1534/g3.117.040105> PMID: 28292787
117. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, et al. Hybridization and speciation. *J Evol Biol.* 2013; 26:229–246. <https://doi.org/10.1111/j.1420-9101.2012.02599.x> PMID: 23323997
118. Irisarri I, Singh P, Koblmüller S, Torres-Dowdall J, Henning F, Franchini P, et al. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat Commun.* 2018; 9:3159. <https://doi.org/10.1038/s41467-018-05479-9> PMID: 30089797
119. Bjornson S, Upham N, Verbruggen H, Steenwyk J. Phylogenomic Inference, Divergence-Time Calibration, and Methods for Characterizing Reticulate Evolution. *Biol Life Sci.* 2023. <https://doi.org/10.20944/preprints202309.0905.v1>
120. Arnold BJ, Huang I-T, Hanage WP. Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol.* 2022; 20:206–218. <https://doi.org/10.1038/s41579-021-00650-4> PMID: 34773098
121. Gophna U, Altman-Price N. Horizontal Gene Transfer in Archaea—From Mechanisms to Genome Evolution. *Annu Rev Microbiol.* 2022; 76:481–502. <https://doi.org/10.1146/annurev-micro-040820-124627> PMID: 35667126
122. Buck R, Ortega-Del Vecchyo D, Gehring C, Michelson R, Flores-Rentería D, Klein B, et al. Sequential hybridization may have facilitated ecological transitions in the Southwestern pinyon pine syngameon. *New Phytol.* 2023; 237:2435–2449. <https://doi.org/10.1111/nph.18543> PMID: 36251538
123. Goulet BE, Roda F, Hopkins R. Hybridization in Plants: Old Ideas, New Techniques. *Plant Physiol.* 2017; 173:65–78. <https://doi.org/10.1104/pp.16.01340> PMID: 27895205
124. Rieseberg LH, Kim S-C, Randell RA, Whitney KD, Gross BL, Lexer C, et al. Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica.* 2007; 129:149–165. <https://doi.org/10.1007/s10709-006-9011-y> PMID: 16955330
125. Steenwyk JL, Lind AL, Ries LNA, dos Reis TF, Silva LP, Almeida F, et al. Pathogenic Allodiploid Hybrids of *Aspergillus Fungi*. *Curr Biol.* 2020; 30:2495–2507.e7. <https://doi.org/10.1016/j.cub.2020.04.071> PMID: 32502407
126. Marcet-Houben M, Gabaldón T. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. Hurst LD, editor. *PLoS Biol.* 2015; 13:e1002220. <https://doi.org/10.1371/journal.pbio.1002220> PMID: 26252497

127. Adavoudi R, Pilot M. Consequences of Hybridization in Mammals: A Systematic Review. *Genes*. 2021; 13:50. <https://doi.org/10.3390/genes13010050> PMID: 35052393
128. League GP, Slot JC, Rokas A. The *ASP3* locus in *Saccharomyces cerevisiae* originated by horizontal gene transfer from *Wickerhamomyces*. *FEMS Yeast Res*. 2012; 12:859–863. <https://doi.org/10.1111/j.1567-1364.2012.00828.x> PMID: 22776361
129. Kominek J, Doering DT, Opulente DA, Shen X-X, Zhou X, DeVirgilio J, et al. Eukaryotic Acquisition of a Bacterial Operon. *Cell*. 2019; 176:1356–1366.e10. <https://doi.org/10.1016/j.cell.2019.01.034> PMID: 30799038
130. Davín AA, Tannier E, Williams TA, Boussau B, Daubin V, Szöllösi GJ. Gene transfers can date the tree of life. *Nat Ecol Evol*. 2018; 2:904–909. <https://doi.org/10.1038/s41559-018-0525-3> PMID: 29610471
131. Jukes TH, Cantor CR. Evolution of Protein Molecules. *Mammalian Protein Metabolism*. Elsevier; 1969. pp. 21–132. <https://doi.org/10.1016/B978-1-4832-3211-9.50009-7>
132. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect Math Life Sci (Am Math Soc)*. 1986; 17:57–86.
133. Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F. Heterotachy and long-branch attraction in phylogenetics. *BMC Evol Biol*. 2005; 5:50. <https://doi.org/10.1186/1471-2148-5-50> PMID: 16209710
134. Liao W-W, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, et al. A draft human pangenome reference. *Nature*. 2023; 617:312–324. <https://doi.org/10.1038/s41586-023-05896-x> PMID: 37165242
135. Svedberg J, Hosseini S, Chen J, Vogan AA, Mozgova I, Hennig L, et al. Convergent evolution of complex genomic rearrangements in two fungal meiotic drive elements. *Nat Commun*. 2018; 9:4242. <https://doi.org/10.1038/s41467-018-06562-x> PMID: 30315196
136. Jain Y, Chandradoss KR, A. V. A, Bhattacharya J, Lal M, Bagadia M, et al. Convergent evolution of a genomic rearrangement may explain cancer resistance in hystrico- and sciuromorpha rodents. *NPJ Aging Mech Dis*. 2021; 7:20. <https://doi.org/10.1038/s41514-021-00072-9> PMID: 34471123
137. Mezzasalma M, Streicher JW, Guarino FM, Jones MEH, Loader SP, Odierna G, et al. Microchromosome fusions underpin convergent evolution of chameleon karyotypes. *Gaitan-Espitia JD, Chapman T, editors. Evolution*. 2023; 77:1930–1944. <https://doi.org/10.1093/evolut/qpad097> PMID: 37288542
138. Steenwyk JL, Shen X-X, Lind AL, Goldman GH, Rokas A. A Robust Phylogenomic Time Tree for Biotechnologically and Medically Important Fungi in the Genera *Aspergillus* and *Penicillium*. *Boyle JP, editor. MBio*. 2019; 10:e00925–19. <https://doi.org/10.1128/mBio.00925-19> PMID: 31289177
139. Liu L, Zhang J, Rheindt FE, Lei F, Qu Y, Wang Y, et al. Genomic evidence reveals a radiation of placental mammals uninterrupted by the KPg boundary. *Proc Natl Acad Sci U S A*. 2017; 114. <https://doi.org/10.1073/pnas.1616744114> PMID: 28808022
140. Smith SA, Brown JW, Walker JF. So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *Escriba H, editor. PLoS ONE*. 2018; 13:e0197433. <https://doi.org/10.1371/journal.pone.0197433> PMID: 29772020
141. Phillips MJ, Penny D. The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol Phylogenet Evol*. 2003; 28:171–185. [https://doi.org/10.1016/s1055-7903\(03\)00057-5](https://doi.org/10.1016/s1055-7903(03)00057-5) PMID: 12878457
142. Ortiz-Merino RA, Kuanyshev N, Braun-Galleani S, Byrne KP, Porro D, Branduardi P, et al. Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. *Hurst L, editor. PLoS Biol*. 2017; 15:e2002128. <https://doi.org/10.1371/journal.pbio.2002128> PMID: 28510588
143. Clark JW, Donoghue PCJ. Whole-Genome Duplication and Plant Macroevolution. *Trends Plant Sci*. 2018; 23:933–945. <https://doi.org/10.1016/j.tplants.2018.07.006> PMID: 30122372