Journal of Heredity, 2023, **114**, 1–13
https://doi.org/10.1093/jhered/esac061
Advance access publication 18 February 2023
Invited Review

American Genetic Association

OXFORD

Invited Review

Estimating phylogenies from genomes: A beginners review of commonly used genomic data in vertebrate phylogenomics

Javan K. Carter[1,2], Rebecca T. Kimball[3], Erik R. Funk[1,4], Nolan C. Kane[1], Drew R. Schield[1], Garth M. Spellman[5], Rebecca J. Safran[1]

[1]Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, CO, United States,
[2]Genomics, Bioinformatics, and Translational Research Center, Research Triangle Institute International, RTP, NC, United States,
[3]Department of Biology, University of Florida, Gainesville, FL, United States,
[4]Department of Conservation Genetics, San Diego Zoo Wildlife Alliance, Escondido, CA, United States,
[5]Department of Zoology, Denver Museum of Nature and Science, Denver, CO, United States

Address correspondence to J.K. Carter at the address above, or e-mail: Javan.carter@colorado.edu.

Corresponding Editor: Mark Springer

Abstract

Despite the increasing feasibility of sequencing whole genomes from diverse taxa, a persistent problem in phylogenomics is the selection of appropriate genetic markers or loci for a given taxonomic group or research question. In this review, we aim to streamline the decision-making process when selecting specific markers to use in phylogenomic studies by introducing commonly used types of genomic markers, their evolutionary characteristics, and their associated uses in phylogenomics. Specifically, we review the utilities of ultraconserved elements (including flanking regions), anchored hybrid enrichment loci, conserved nonexonic elements, untranslated regions, introns, exons, mitochondrial DNA, single nucleotide polymorphisms, and anonymous regions (nonspecific regions that are evenly or randomly distributed across the genome). These various genomic elements and regions differ in their substitution rates, likelihood of neutrality or of being strongly linked to loci under selection, and mode of inheritance, each of which are important considerations in phylogenomic reconstruction. These features may give each type of marker important advantages and disadvantages depending on the biological question, number of taxa sampled, evolutionary timescale, cost effectiveness, and analytical methods used. We provide a concise outline as a resource to efficiently consider key aspects of each type of genetic marker. There are many factors to consider when designing phylogenomic studies, and this review may serve as a primer when weighing options between multiple potential phylogenomic markers.

Key words: genomic marker types, genomics, phylogenomics, reduced representation

Introduction

The ability to acquire sequence data on a genomic scale has revolutionized biology, including the establishment and growth of new fields of study related to bioinformatics, genomics, and transcriptomics. This change has also resulted in the ongoing shift from phylogenetic inference relying on 1 or several genetic loci from organisms to phylogenomic studies harnessing the power of genome-wide data.

The transition from phylogenetics to phylogenomics has been facilitated by the decreased financial cost of acquiring large sequencing datasets. Yet, due to the computational demands involved in analyzing genome-scale data or the inability to gain sequence coverage across the entire genome, researchers typically target or subsample a subset of the total available genomic information to generate a reduced representation phylogenomic marker set, particularly for highly dimensional and complex analyses. Reduced representation markers used in phylogenomics often constitute less than 5% of the genome, but nonetheless facilitate high-resolution inferences from numerous snapshots of the genome, making them suitable for constructing phylogenomic hypotheses for diverse taxa. Many types of reduced representation, structural, and method-based genomic data are used in practice, yet to our knowledge there is not a thorough synthesis on when and how different marker types are best applied to phylogenomic analyses. This potentially complicates the design and planning stages of phylogenomics projects, especially for researchers with limited experience in genomics, bioinformatics, and molecular biology.

Phylogenomics focuses largely on analyzing evolutionary histories to reconstruct relationships between taxa. These evolutionary relationships can range in taxonomic hierarchy from species level to kingdoms. Unfortunately, there are several practical limitations (e.g. exorbitant computational

Received July 22, 2022; Accepted October 26, 2022

time and failure to achieve convergence in Bayesian analyses) and modeling issues (e.g. difficulty aligning highly variable regions and high heterogeneity across gene trees) that make genome-wide phylogenies intractable (Philippe *et al.* 2005; Young and Gillung 2020; Zhang and Lai 2020). Specific reduced representation genomic marker types that are widely used in practice can be generated using methods falling into 3 broad categories: target capture, transcriptomics, and restriction site associated DNA sequencing (RADseq). In this review, we address the various uses of data produced primarily by these reduced representation sequencing methods but will also include structural marker types and method-based marker types as well, with references to example studies that highlight their utility in phylogenomics.

A major goal of this review is to streamline the decision-making process for choosing a marker set for phylogenomic studies by introducing various types of data, their evolutionary characteristics, and their different utilities in resolving older/deep time versus recent relationships (Table 1). The suitable divergence time for each marker type (deep = order or higher, moderate = family and genus, shallow = species and subspecies) is relative to the study system being examined and biological question being investigated, therefore, in the context of the review, should only be used as a reference. Specifically, we review the use of ultraconserved elements (UCEs; including flanking regions), anchored hybrid enrichment (AHE) loci, conserved nonexonic elements (CNEEs), untranslated regions (UTRs), introns, exons, mitochondrial DNA (mtDNA), single nucleotide polymorphisms (SNPs), and anonymous genomic regions. The latter (anonymous genomic regions) refers to randomly or systematically subsampling regions of the genome where otherwise genome-wide data are available. Although target capture methods UCE, AHE, and CNEE are in fact methods and not a genomic marker type, we will refer to them as marker types because of the types of loci typically targeted using these methods. Our overview of these methods and associated types of phylogenomic markers is designed to be useful to anyone interested in potential marker types and techniques for genome-scale estimation of phylogenies and will be especially useful to those planning empirical phylogenomic studies of their own. Readers who are new to the field of evolutionary genomics and phylogenomics may also appreciate the accessibility of this review for applying information from different genomic regions to evolutionary hypothesis testing. While some of these marker types may become replaced as genomic and computational resources continue to improve, we also highlight several methods that are emerging as new common practice, due to a combination of reliability, low cost, and ease of use.

Overview of types of markers used in phylogenomics

Target capture approaches (UCEs, AHE, and CNEEs)

UCEs, CNEEs, and AHE elements are very similar to one another in goal and theory. Each approach targets evolutionarily conserved loci across the genome and are each especially useful for phylogenomic studies focused on deeper timescales. Differences between these markers lie in the specific methodologies used to extract targeted loci, associated rates of evolution, and accordingly the level of conservation among loci across taxa.

Ultraconserved elements

Key features

The use of UCEs in phylogenomic analysis was first developed and presented in Faircloth *et al.* (2012), who described these markers as "molecular fossils" due to their extremely slow rates of evolution and conservation across deeply divergent taxa (Fig. 1; Bejerano *et al.* 2004; Stephen *et al.* 2008; Faircloth *et al.* 2012; Andersen *et al.* 2019). Additionally, UCEs are not typically associated with paralogous genes or retrotransposons (Faircloth *et al.* 2012; McCormack *et al.* 2012; Harvey *et al.* 2016) and so generally represent single-copy loci, meeting the assumptions of most phylogenetic models. While the central regions of UCEs are highly conserved (referred to as *core UCEs*), the flanking regions of these elements (i.e. *flanking UCEs*) contain greater genetic variation that is informative for phylogenomic inference (Fig. 1). Core UCEs are typically 100 to 150 bp in length while flanking UCEs extend from 400 to 1200 bp up- and downstream of the core UCE (Bejerano *et al.* 2004; Stephen *et al.* 2008; Faircloth *et al.* 2012).

Most current studies use a probe set to anchor to core UCEs in order to capture, enrich, and analyze the flanking UCE regions (Fig. 1 and see Dapprich *et al.* 2016) for further explanation on the capture and enrichment process). The terms "probes" and "baits" can be seen in the literature used interchangeably. Probes/baits are custom RNA sequences that bind to complementary DNA strands which are then "captured" and amplified via PCR then sequenced (Andermann *et al.* 2020). The conservation of core UCEs across taxa makes the downstream task of aligning homologous loci much simpler, allowing for broad comparisons of taxa with conserved sets of loci. Because UCE markers are scattered across the genome, analyses using these markers benefit from genome-wide representation of variation and are not tied to single regions of the genome that may have unusual evolutionary histories or evolutionary/substitution rates compared with some marker types (e.g. mitochondrial genes). Because of these properties, UCEs are valuable genomic tools for phylogenomics, and assembly does not require a reference genome. An additional limitation of UCEs is that they are not well suited for use with extremely recently diverged radiations for which UCEs will likely bear limited phylogenetic signal.

Effective uses of UCEs

UCEs are versatile markers for phylogenomic analyses. Faircloth *et al.* (2012) both developed and originally demonstrated examples in which they test UCE use for a deep time phylogenomic hypothesis for species (primarily fish) spanning multiple families and orders. While segments of these sequences are highly conserved and mainly useful for deep time taxonomic comparisons, UCE flanking regions also allow for the study of more shallow divergence times (e.g. subspecies and species level; Smith *et al.* 2014; Harvey and Brumfield 2015; Harvey *et al.* 2016; Mason *et al.* 2018; Winker *et al.* 2018). For example, Mason *et al.* (2018) constructed a RAxML phylogeny of the subspecies using 4,000 UCEs loci in a phylogenomic study of neotropical birds from Central and South America, the white collared seedeaters (*Sporophila torqueola*).

While UCEs have become a commonly used tool in phylogenomic studies due to their utility for estimating both deep and shallow time phylogenomic relationships, a

Table 1. Overview of the key aspects of the marker types explained in the review.

Marker types	Specimen samples needed/minimum quality needed	Evolutionary history estimate (deep-moderate-shallow)	Types of genetic and genomic method needed/reference genome requirement	Relative cost based on genomic method needed (see previous column)	Reference papers/example papers
UCEs = ultraconserved element flanking regions	Historic specimens, fresh tissue, blood samples/low quality	Deep to shallow time estimates	Target method, WGS data with computational target method/reference genome needed if designing probes	Low to moderate	(Bejerano *et al.* 2004; Faircloth *et al.* 2012, 2015; McCormack, Faircloth *et al.* 2012; McCormack *et al.* 2012, 2013; Crawford *et al.* 2015; Harvey and Brumfield 2015; Faircloth 2016; Harvey *et al.* 2016; Mason *et al.* 2018; Winker *et al.* 2018; Andersen *et al.* 2019)
AHE = anchor hybrid enrichment	Historic specimens, fresh tissue, blood samples/low quality	Deep to shallow time estimates	Target method, WGS data with computational target method/reference genome needed if designing probes	Low to moderate	(Prum *et al.* 2015; Hamilton *et al.* 2016; Young *et al.* 2016; Maddison *et al.* 2017; Espeland *et al.* 2018; Godwin *et al.* 2018; Haddad *et al.* 2018; Pepper *et al.* 2018; Shin *et al.* 2018; Braun and Kimball 2021)
CNEEs = conserved nonexonic elements	Historic specimens, fresh tissue, blood tissue/moderate quality	Moderate to deep time estimates	Target method or WGS data (with computational target method (preferred)/reference genome preferred	Low to moderate	(Visel *et al.* 2008; Lowe et al. 2011, 2015; Kvon *et al.* 2016; Leal and Cohn 2016; Edwards *et al.* 2017)
Exon	Historic specimens, fresh tissue, blood sample, RNA samples/low to high quality	Mostly deep time estimates	WGS data, RNA/transcriptomic data, exonic target capture/reference genome preferred	Moderate to high	(Bi *et al.* 2012; Ilves and López-Fernández 2014; Jarvis *et al.* 2014; Bragg *et al.* 2016; Hugall et al. 2016; Portik *et al.* 2016; O'Hara *et al.* 2017; Karin *et al.* 2020)
Introns	Historic specimens, fresh tissue, blood sample/moderate quality	Deep to moderate time estimates	WGS data/transcriptomic data, intronic target capture/reference genome preferred	Low to moderate	(Lessa 1992; Mk et al. 2004; Creer 2007; Matthee *et al.* 2007; Chojnowski *et al.* 2008; Hackett *et al.* 2008; Salicini *et al.* 2011; Jarvis *et al.* 2014; Foley *et al.* 2015; Chen *et al.* 2017)
UTRs = untranslated regions	Fresh tissue, blood sample, RNA samples/high quality	Deep or shallow time estimates	WGS, RNA/transcriptomic data/reference genome preferred	Moderate to high	(Stebbins-Boaz and Richter 1997; Conne *et al.* 2000; Murphy *et al.* 2004; Bonilla *et al.* 2010; Irisarri and Meyer 2016; Kuhl *et al.* 2020; Xiong *et al.* 2018; Wang *et al.* 2019)
Mitochondrial DNA	Historic specimens, fresh tissue, blood samples/low quality	Shallow time estimates	mtDNA primers, WGS/no reference genome required	Low	(Brown *et al.* 1979; Moore 1995; Boore and Brown 1998; Boore 1999; Phillips and Penny 2003; Rubinoff and Holland 2005; Peng *et al.* 2006; Avise 2012; Wallace and Chalkia 2013)
Anonymous loci/regions (RADseq)	Historic specimens (not including RADseq), blood sample, tissue sample/moderate quality	Moderate to shallow time estimates; ultimately depends on the primers used	PCR primers, WGS/no reference genome required for PCR primers	Low	(Miller *et al.* 2007; Baird *et al.* 2008; Davey and Blaxter 2010; Peterson et al. 2012; Harrison *et al.* 2014; Allman et al. 2016; Andrews *et al.* 2016; Harvey *et al.* 2016; McKenzie and Eaton 2020)
SNPs = single nucleotide polymorphisms	Historic specimens, fresh tissue, blood sample, RNA samples/low quality	Deep to shallow time estimates	Any genomic data prep/reference genome required	Low to high—costs ARE applicable, but variable depending on data	(Brumfield *et al.* 2003; Morin *et al.* 2004; Baird *et al.* 2008; Hohenlohe *et al.* 2011; Li *et al.* 2012; McGill *et al.* 2013; Leaché et al. 2015; Leaché and Oaks 2017; Vachaspati and Warnow 2018; Wang et al. 2020)

Specimen samples needed/minimum quality needed (column 2): types or quality of samples recommended in order to yield the most sequence data range. Quality is based on the degradation of the DNA. Evolutionary history estimate (column 3): relative divergence time each marker type is more suitable for (deep = order or higher, moderate = family and genus, shallow = species and subspecies). The suitable divergent time is relative to the study system being examined and biological question being investigated therefore, in this context, should only be used as a reference. All citations for Table 1 can be found in the Supplementary Material for Table 1 file in the supplement section.

limitation to their use is the cost of, and effort associated with developing baits for generating UCE sequencing libraries. Probe sets now exist for many broad taxonomic groups (i.e. amniotes, fishes, insects) and can be purchased through bioscience companies (i.e. Arbor Biosciences); it is also possible to freely obtain files that has probe positions that can

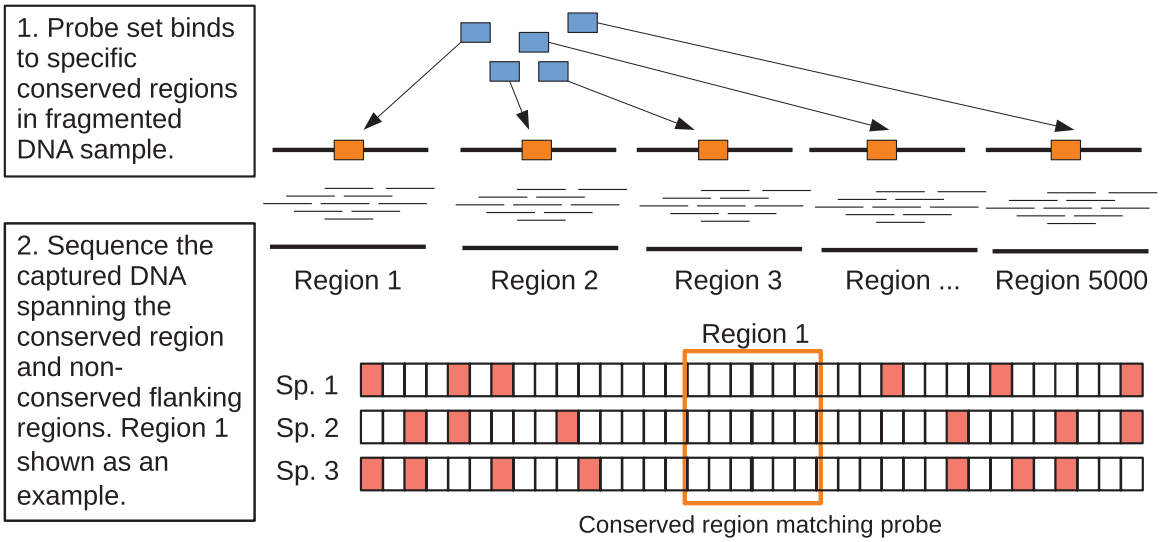

Fig. 1. Chronological order of how target capture marker types such as UCEs and AHEs are collected (while AHEs use tiled baits, we are showing a single bait for clarity). 1) The blue squares represent the probe set that binds to a conserved region represented by an orange square. 2) Sp. is an abbreviation for "species" and each box represents an individual nucleotide of a sequence. The open orange box represents the conserved region of the UCE or AHE and the figure is showcasing the variability of the flanking regions of UCEs. Variant sites, red boxes, are typically located away from the conserved region, which contains mostly invariant sites, white boxes. This cartoon includes a reduced number of nucleotides for clarity.

be used for more organisms with more conserved genomic architectures. As more reference genome assemblies from diverse taxa or nonmodel organisms become available for bait development, the utility of UCEs is also likely to increase.

Lastly, missing data can be an issue with UCEs but is also a challenge for almost all marker types and/or datasets in genetic and genomic studies because of the reduction of phylogenetic accuracy and/or poorly resolved tree. Missing data can be observed from "1) Stochasticity inherent in collecting data across thousands of loci, where not all loci are detected in all genomic libraries; 2) variable sequence yield among sample libraries leading to missing data across alignments; and 3) biological processes including insertions, deletions, and other chromosomal changes" (Hosner *et al.* 2016). There are ways to reduce missing data but typically at the cost of site quantity, financial expenses, and computational power (Philippe *et al.* 2004; Hosner *et al.* 2016; Streicher *et al.* 2016). For more information on missing data please consider reading the following literature (Philippe *et al.* 2004; Wiens and Morrill 2011; Hosner *et al.* 2016; Streicher *et al.* 2016).

Anchored hybrid enrichment

Key features

AHE is an approach to capture homologous regions of the genome from potentially hundreds of taxa (Fig. 1). This capture method is very similar in principle to UCEs (Lemmon *et al.* 2012), though AHE utilizes multiple baits per locus to facilitate more robust sequence capture spanning the locus and often results in fewer loci than UCEs. AHE uses highly conserved genomic regions that are longer than core UCE regions and flanking areas as anchors. Similar to flanking UCEs, the flanking regions of AHE anchors can be readily used in phylogenomic analyses because they harbor a higher frequency of genetic variants than the core or anchor region of AHE loci. AHE uses a tiled bait approach to maximize the length of the target loci to increase the accuracy of captioning homologous loci observed across a wide range of species (Lemmon *et al.* 2012). This was initially done using a set of probes that were assembled from 5 animals representing major vertebrate groups: *Homo sapiens* (mammals), *Gallus gallus domesticus* (birds), *Anolis carolinensis* (squamates), *Xenopus tropicalis* (amphibians), and *Danio rerio* (fish). The initial study produced 512 unique loci based on 3 key parameters. First, the core anchor (240 base pairs) had to be genomically unique among all 5 species. Second, the flanking regions around the anchor (700 base pairs up- and downstream of the core anchor) could not contain any repeat elements. Lastly, the probe regions could not have a high number of indels. The total amount of the genome represented by the 512 loci was approximately 122,800 base pairs. In Prum *et al* (2015), as is generally true of AHE studies, most AHEs appear to be affiliated with conserved exons (Braun and Kimball 2021). This is distinct from vertebrate UCEs, which mainly fall within noncoding regions. Since the initial study, AHE loci have been developed for a wide range of taxa.

Effective uses of AHEs

The original goal of the AHE approach was to aid in the development of deep time species trees for vertebrate groups (Lemmon *et al.* 2012). There have since been numerous studies that have used AHEs to generate species trees for shallow and deep time estimates (e.g. Eytan *et al.* 2015; Hamilton *et al.* 2016; Singhal *et al.* 2017; Pepper *et al.* 2018) along with several studies that produced phylogenomic estimates for invertebrates (Eytan *et al.* 2015; Hamilton *et al.* 2016; Young *et al.* 2016; Maddison *et al.* 2017; Singhal *et al.* 2017; Espeland *et al.* 2018; Godwin *et al.* 2018; Haddad *et al.* 2018; Pepper *et al.* 2018; Zhang *et al.* 2019). These and other examples illustrate that AHE loci are highly useful for inferring phylogenies of diverse taxonomic groups (Hamilton *et al.* 2016).

Several advantages associated with using UCEs are also applicable to AHE loci. One of which includes enabling the user to ask questions regarding deep time phylogenomics while reducing the burden of financial cost. Another advantage is related to their robustness within nonmodel systems.

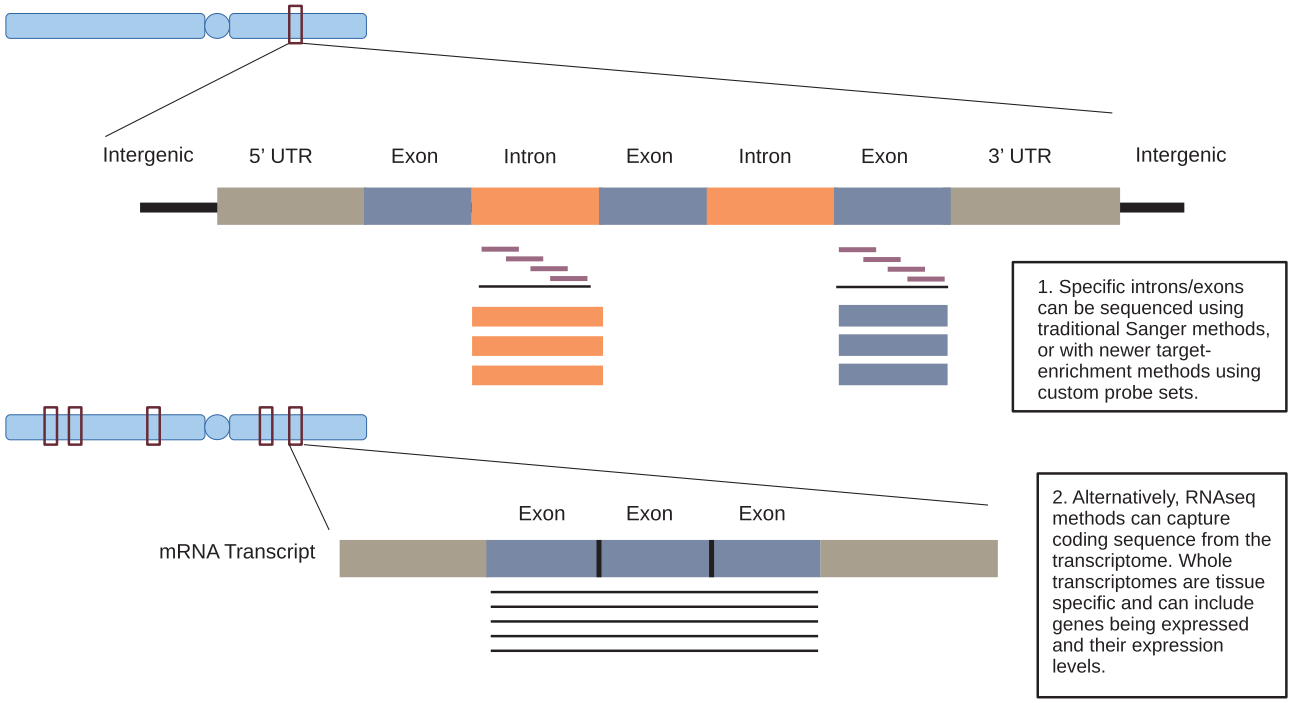

Fig. 2. Graphical representation of exonic, intronic, and UTR target capture methods. 1) The blue figure represents a generic chromosome. The open purple box is enlarging an intragenic region at random and showcasing the normal orientation of UTR, exons, and introns in comparison to intergenic regions. A simplistic representation of probes tiled across targeted loci is represented by the purple bands. Target enrichment methods such as this can be applied to UTR, exons, introns, and mtDNA. 2) Represents graphical ration of RNA-seq or any transcriptomic method approach (e.g. UTRs and exons). Open purple boxes represent coding regions and stacked black lines represent sequence reads.

The conserved nature of AHEs and UCEs across such a wide range of disparately related taxa makes it simpler to use and acquire easily aligned data than when searching for loci using genome scans or other methods (Faircloth *et al.* 2012; McCormack *et al.* 2012; Crawford *et al.* 2015; Hamilton *et al.* 2016; Shin *et al.* 2018).

Conserved nonexonic elements

Key features

CNEEs are the most recently developed marker type for phylogenomics reviewed here, and although very similar to UCEs, are not commonly obtained using target capture approaches. While CNEEs were first described in 2008 (Visel *et al.* 2008), they were first used in a phylogenomic context in 2015 (Lowe *et al.* 2015). These markers are derived from regulatory regions of the genome, have a slower substitution rate than their associated protein-coding regions on average, and can be found in a diversity of vertebrates (Edwards *et al.* 2017). These elements are associated with the recruitment of transcription factors which manipulate the expression of genes in close proximity (Kvon *et al.* 2016; Leal and Cohn 2016). CNEEs are very similar to conserved noncoding elements (CNEs) described in (Marcovitz *et al.* 2016). The primary difference between CNEs and CNEEs is the absence of exon sequences in CNEEs (Edwards *et al.* 2017). CNEEs have overlap with approximately 50% of the vertebrate UCEs markers (Edwards *et al.* 2017), so care needs to be taken to avoid nonindependence if both marker types are used. CNEEs were originally extracted computationally using a hidden Markov model search for locations with slow evolutionary rates across the genome. Extraction of CNEEs from Whole Genome Sequence (WGS) data is recommended but not required since due to evolutionary turnover in regulatory sequences among vertebrates, homologous CNEEs may not be available for all vertebrate taxa. The appearance and disappearance of CNEEs throughout the tree of life are being investigated to better understand regulatory aspects of phenotypic traits (Lowe et al. 2011, 2015; Edwards *et al.* 2017).

Current and most effective way to use CNEEs

CNEEs are noncoding elements which have a more neutral (slower) evolutionary rate compared with coding regions. The conserved nature of CNEEs results in fewer nucleotide substitutions, so these may be most appropriate for higher level phylogenomic studies in vertebrates. Organisms with shallow evolutionary histories may share identical homologous CNEE regions and thus, these markers may not be suitable for phylogenetic inference in these cases.

Intragenic regions (exons, introns, and UTRs)

Exons

Key features

Exons have retained value in the transition from first-generation sequencing phylogenetics to phylogenomics due to their direct association to proteins and gene function. Exons are the protein-coding segments of a gene that, when combined with other exons, will determine the protein that is produced. Exons make up, on average, a very small portion of the genome (Fig. 2). Exons are often under purifying selection; this reduces the accumulation of mutations at these loci. Thus, exons are more conserved than introns, and tend to have little length variation across taxa, requiring less computationally intense pipelines for alignment and analysis than other genetic

structural variants (i.e. INDELS). A 5% to 40% higher constraint can be seen between 5′ and 3′UTR (16.4%, 13.3%), intergenic (6.8%), and intronic (2.2%) regions to exonic regions (coding position sites CDS1 65.5%, CDS2 70.8%, CDS3 24.6%) (Pollard *et al.* 2010).

Effective uses of exons

Exonic data have a long history of being used in deep time phylogenetic and phylogenomic studies. Moderate to deep time phylogenomic signals can be observed and interpreted with this marker type across the tree of life (George *et al.* 2011; Bi *et al.* 2012; Ilves and López-Fernández 2014; Bragg *et al.* 2016; Portik *et al.* 2016; Teasdale *et al.* 2016; O'Hara *et al.* 2017; Scornavacca and Galtier 2017; Jiang *et al.* 2019). The ability to cost-effectively sequence transcriptomes means that exons can be readily obtained for organisms in which it is possible to obtain and store tissues for mRNA extraction, without need for a reference genome. Alternatively, exons can be obtained using probe sets (see AHE above). For example, a new exonic computational probe set, called Rapidly Evolving Long Exon Capture (RELEC), was designed for phylogenomic studies. This focuses on longer and more rapidly evolving exons (Karin *et al.* 2020) and provides a strong phylogenomic signal (Karin *et al.* 2020).

Introns

Key features

Introns were once believed to hold little to no value in biology because they are noncoding regions of genes. However, it is now known that some introns hold various functions, including gene expression/regulation, alternative splicing for generating several types of protein for a single gene, and mRNA transportation control (Cenik *et al.* 2011; Rearick *et al.* 2011; Bicknell *et al.* 2012). While it was initially assumed that introns were selectively neutral, introns may be under selection due to these various functions. Since the early to mid 2000s, this marker type has become widely used in phylogenomics.

Effective uses of introns

The length of introns varies from tens to many thousands of base pairs in size and also range in the quantity of introns per gene (i.e. for humans and most primates it averages roughly 7 per gene) (Sakharkar *et al.* 2004; Creer 2007). Because introns are located within genes (Fig. 2), they will always be under some degree of purifying selection (Pollard *et al.* 2010). This means the intronic segments will in part be conserved and show a slower substitution rate than intergenic regions (in mammal it is show that intergenic regions are slightly more conserved than intronic regions; Pollard *et al.* 2010), making them robust when looking at a moderate to deep evolutionary relationships (Matthee *et al.* 2007; Chojnowski *et al.* 2008; Hackett *et al.* 2008; Salicini *et al.* 2011; Foley *et al.* 2015; Jarvis *et al.* 2015; Chen *et al.* 2017). Probably the most notable, recent intronic phylogenomic study was a reconstruction of the Neoaves phylogeny with 48 species representing all of the orders (Jarvis *et al.* 2014). When comparing exons, UCEs, and introns, introns were found to be associated with producing the most robust and well-resolved phylogenomic hypothesis when confronted with ancient rapid radiations that contributed to the high incomplete lineage sorting originally seen in the group (Jarvis *et al.* 2014). It is important to note that when using intronic or exonic marker types, these areas of the genome show considerable evolutionary rate variation from one gene to another. The use of introns has decreased with the transition from phylogenetics to phylogenomics, though the ability to extract introns from WGS data is ongoing.

Untranslated regions

Key features

Unlike UCEs and CNEEs, untranslated regions (otherwise known as UTRs) can be upstream or downstream regulatory elements and are associated with mRNA stability, mRNA localization, and protein–protein interactions and have been used in phylogenetics since the late 1990s (Stebbins-Boaz and Richter 1997; Conne *et al.* 2000; Kuhl et al. 2020). Structurally, 3′UTRs represent the noncoding downstream end of mRNA while the 5′UTR resides upstream of the coding region (Fig. 2). Apart from functionality, 3′UTR and 5′UTR differ in size, 3′UTR averaging 3 times longer in nucleotide base pairs than 5′UTR, making for an easier segment to align and thus easier for phylogenomic analyses. 3′UTRs can best be acquired from transcriptomic data or WGS (Fig. 2). Unfortunately, depending on the tissue type for RNA analysis, the difficulty of acquiring transcriptomic data can be moderately challenging under some conditions due to the speed at which RNA naturally degrades.

Effective uses of UTRs

Since 3′UTRs are associated with post-transcriptional regulation and are directly influencers of gene expression, they are often under positive or purifying selection. The variation in this marker type is associated with nucleotide substitutions and length differentiation (Xiong *et al.* 2018; Wang *et al.* 2019). These marker types are typically more variable than exons, and are commonly used for taxa exhibiting shallow to moderate degrees of divergence (Murphy *et al.* 2004; Bonilla *et al.* 2010; Xiong *et al.* 2018; Kuhl *et al.* 2020). The lack of empirical studies using 3′UTRs could be related to the difficulty in collecting or extracting UTR sequences using traditional PCR approaches relative to other less expensive and easier methods (Bonilla *et al.* 2010). Although the use of 3′UTRs has decreased with the transition from phylogenetics to phylogenomics in the past decade, the ability to extract UTRs from transcriptomic data may increase their use over time. Also, UTRs vary in length among taxa, leading to challenges with alignment and coding of insertion and deletion events (indels). Lastly, any recent large-scale duplications or changes in ploidy further complicates the phylogenomic resolution if using UTRs, similar to most marker types in this review (Irisarri and Meyer 2016).

Mitochondrial DNA

Key features

What makes mtDNA unique is that it is maternally inherited and forms a single haplotype that undergoes little or no recombination (depending on taxa). Unlike basic features of nuclear DNA, which vary drastically among organisms from size to genomic architecture to genetic content, mtDNA remains very consistent in size throughout most of the animal kingdom, ranging from 14 to 20 kilobases, depending on the number of noncoding regions. It typically consists of 37 genes (Boore 1999), including 13 protein-coding genes, tRNAs, and

rRNAs. Mitochondrial genome size is more variable in some taxa (e.g. mammals), but gene content is conserved to a large degree over even very broad scales (Janouškovec *et al.* 2017). Although mtDNA genomes maintain a highly conserved architectural structure, in metazoans the evolutionary rate in terms of nucleotide substitutions is one of the fastest among all of the marker types in this review (Brown *et al.* 1979; Saccone *et al.* 2006).

Effective uses of mtDNA

Prior to the 2000s, mtDNA was the dominant marker type for phylogenetic inference because of its size, ease of data collection, and very low recombination rate (Boore and Brown 1998; Li *et al.* 2001; Herrnstadt *et al.* 2002; Gibson *et al.* 2005; Macaulay *et al.* 2005; Minegishi *et al.* 2005; Peng *et al.* 2006; Cameron et al. 2007, 2008; Fenn *et al.* 2008; Meredith *et al.* 2011). During the early to mid 2000s, studies documenting incongruent tree topologies with nuclear markers questioned the accuracy of mtDNA for inferring species trees (Hurst and Jiggins 2005; RubCoff and Holland 2005; Milián-García *et al.* 2020). This was concordant with a shift in phylogenetics away from individual gene trees (the evolutionary history of a single gene or locus) toward species trees (the inferred evolutionary history between organisms) (Avise *et al.* 1983; Maddison 1997; Funk and Omland 2003; Avise 2012). Due to its high nucleotide substitution rate (Boore 1999; Saccone *et al.* 2006; Jiang *et al.* 2016; Zhu *et al.* 2018), mtDNA is still often used at the population and species level, and as the DNA barcoding locus in most animals. However, few vertebrate studies solely use mtDNA possibly due to several reasons. First, although it has been possible to sequence whole mitochondrial genomes for over 2 decades, it only makes up 20 kb of molecular information, a miniscule fraction of the available molecular information from animals. Second, the evolutionary history reflected in mtDNA reveals is limited to tracking maternal inheritance. This gives an incomplete and potentially biased representation of the relationship among species (Hurst and Jiggins 2005; Rubinoff and Holland 2005; Balloux 2010). Third, mtDNA is a single locus because it lacks recombination for most organisms (though this does allow estimation of a single gene tree with higher accuracy than may be true for many nuclear gene trees). This reduces its ability to represent the overall evolutionary history of the group in question.

What genomic data are needed to obtain mtDNA?

MtDNA is among the most cost-effective marker types one can acquire primarily for 3 reasons. First, due to its long history within phylogenetics, an ample number of resources have already been developed to effectively target mtDNA regions that are favorable within phylogenetics (i.e. genetic primers that can be used for PCR). In addition to collecting new mtDNA sequences, mtDNA has been extracted, collected, and submitted to several different global databases where they are readily available for thousands to tens of thousands of species (e.g. NCBI). Lastly, because of the large abundance of mitochondria found within most cells, lower coverage genome sequencing can still provide high coverage of the mitochondrion (Reich *et al.* 2010). This may mean that mtDNA data will often be present in target capture studies (e.g. UCE and AHE), particularly those using mitochondrially enriched tissues, unless highly stringent washing procedures are employed. Tissues like muscle, liver, and brain almost always contain ample mtDNA, whereas DNA extracted from blood rarely does except in taxa with nucleated blood cells like birds (Shuster *et al.* 1988). Even degraded tissues are known to sometimes yield a sufficient amount of mtDNA to assemble much or all of the mitogenome.

Anonymous regions (RADseq, sliding windows, and rare genomic change)

Anonymous regions are areas of the genome that are not characterized by position or biological functionality, but rather, homologous sequences thought to be orthologous across taxa (Harrison *et al.* 2014; Allman et al. 2016; McKenzie and Eaton 2020). Examples of these include RADseq data, sliding windows genomic screening, and rare genomic changes (RGCs).

RADseq

Key features

Some once considered RADseq as the "most important scientific breakthrough" of the 2010s decade (Andrews *et al.* 2016) because of its revolutionary approach to collecting hundreds to thousands of genomic regions for the fraction of normal sequencing cost (Miller *et al.* 2007; Baird *et al.* 2008; Davey and Blaxter 2010; Andrews *et al.* 2016). RADseq data, including genotyping-by-sequencing approaches, use restriction enzymes and NGS to sequence "random" homologous sites. It has become standard to extract SNPs from the RADseq outputs. Because RADseq outputs vary where the enzymes cut in the genome and could contain partial genes or incomplete regulatory regions, RADseq sites may not be completely neutral. RADseq also cannot be used if inquiry on specific regions of the genome is of interest, since the regions targeted depend upon restriction sites, rather than function. Another issue with using RADseq data is in its incompatibility with other RADseq datasets. Because of its seemingly random location selection and read length, combining different empirical studies that use RADseq is extremely difficult and generally not advised.

Effective use of RADseq

As whole-genome sequencing is becoming cheaper and gene tree/species tree programs that can handle more data and more complex systematic predictions become available, the advantages of RADseq may become less important. Phylogenetically, RADseq data are more suited for population-, subspecies-, and species-level divergence between taxa and not favored for more deep time divergences (e.g. genus, family, or order).

Sliding windows

Key features

The use of genomic sliding windows is a technique that involves moving along the genome using a WGS (usually) dataset and building a phylogeny for sequences of uniform size, typically 5 to 100 kb, depending on how much data is needed or desired. The 2 main advantages of this technique are the ability to reduce the data size, which in turn reduces the computational power needed for analyses, and the ability to obtain nonbiased sampling evenly throughout the genome. This technique, like all anonymous region techniques or approaches, prevents one from knowing the biological

characteristic of the genomic regions, although filters could be established that might focus on or exclude certain regions, e.g. by eliminating coding regions, from consideration if the goal is to focus on neutral sites. Trade-offs exist when choosing the window size and distance between windows to include sufficient phylogenomic information while minimizing computation time.

Rare genomics changes

Key features

RGCs refer to mutations whose infrequent occurrence means these should essentially be "perfect" characters—that is, characters that exhibit no (or very low) homoplasy that can be used as phylogenetic markers (e.g. Rokas and Holland 2000). Commonly used RGC in phylogenetic studies includes insertions and deletions (indels), particularly those involving movement, duplication, or loss of transposable elements (TEs) (Nikaido *et al.* 1999; Springer *et al.* 2020). For a given TE, the probability the same TE will insert at the same location in the genome independently in different species is hypothesized to be extremely low, as is precise deletion of a TE insertion. So, the presence of a particular TE insertion in multiple taxa should be a synapomorphy that unites those taxa. Although TEs exhibit low homoplasy, there does appear to be some potential for homoplasy (Han *et al.* 2011). In addition to TEs, there are many other types of mutations that have also been suggested to be RGCs. These include other types of insertions, such as insertions of mtDNA into the nuclear genome (Liang *et al.* 2018), genomic rearrangements, including organelle gene order (Tyagi *et al.* 2020), inversions or microinversions (Braun *et al.* 2011), microRNAs (Field *et al.* 2014), and more typical insertion/deletion events, particularly in noncoding regions (e.g. Houde *et al.* 2019).

Effective uses of RGCs

Most RGCs can be identified from whole-genome sequences, though TE insertions have been targeted without whole-genome sequencing in some studies (e.g. Shimamura *et al.* 1997). Since RGCs occur infrequently, approaches that sample small portions of the genome (e.g. reduced representation methods such as UCEs, AHEs, and RADseq) may sample too few of these events to provide much phylogenetic data. Some inversions, such as those involving large portions of a Maney and Goodson (2011), may also occur so rarely as to be uninformative. However, when whole genomes are available, looking for RGCs may provide additional data. Although RGCs may be perfect (or near perfect) characters, they are perfect with respect to the evolutionary history of the genomic region they represent. Due to incomplete lineage sorting, some regions of the genome will have evolutionary histories that are distinct from that of the species as a whole, leading to an RGC that may appear to exhibit homoplasy relative to the species tree, but where the RGC is actually matching the evolutionary history for its region (Avise and Robinson 2008). Thus, RGCs may often be excellent for defining gene trees (or bipartitions) that can be used to infer species trees (Houde *et al.* 2019; Springer *et al.* 2020).

Single nucleotide polymorphism

Key features

After the conclusion of the Human Genome Project (HGP) in 2003, a greater focus was placed on analyzing genetic variation across the genome by identifying SNPs. What makes SNPs significantly different from other genomic marker types in this review is that SNPs are only the variant sites in the genetic material between 2 or more subjects and are often analyzed without information on position and surrounding sequence. SNPs are the result of point mutations at the base-pair level and are abundant on a genome-wide scale in coding and noncoding regions. This holds true for all animals and most taxonomic levels, which allows for a useful genomic marker type for studying molecular phylogenomic relationships (Rokas and Holland 2000). This marker type is favorable for examining population demographics, adaptation, quantitative genetics, phylogeography, genome evolution, and phylogenomics (Brumfield *et al.* 2003; Morin *et al.* 2004). Since SNP analyses can be applied to various marker types, they can be exacted from any dataset (such as those collected using approaches described in this review, as well as from whole-genome comparisons).

Effective uses of SNPs

Because SNPs can be a subset of various reduced representation marker types, the range of phylogenomic inferences that can be examined is directly influenced by the genomic marker type from which they were obtained. SNP data matrices can be used in various ways, but some options are more controversial than others. A concatenated matrix (super matrix) is a standard method within phylogenomics that combines all of the SNPs from each sample. Although commonly used, this approach assumes all of the SNPs share the same coalescent history (Edwards et al. 2016a, b; Leaché and Oaks 2017) and is susceptible to the same biases as other concatenation approaches. For these reasons, many phylogenomicist argue against the super matrix approach in favor of the multilocus coalescent model approach that maintain the SNPs as distinct loci and generate species trees from gene trees (Edwards et al. 2016a, b; Leaché and Oaks 2017). Other common methods for SNP analyses bypass gene trees while still incorporating the multispecies coalescent into species tree building programs such as SNAPP, SVDquartet, SVDquest, and PoMo (Bryant *et al.* 2012; Chifman and Kubatko 2014; De Maio *et al.* 2015; Vachaspati and Warnow 2018). These programs are convenient as they allow reduction in computational steps while allowing for a model-based coalescent analysis. Like marker choice, appropriate method choice can heavily depend on the biological questions and the timescales being examined.

Lastly, 2 common concerns associated with SNP data are ascertainment bias (a deviation of statistics from theoretical expectation from bias nonrandom sampling) (Lachance and Tishkoff 2013; McGill *et al.* 2013) and missing data (missing of data from different samples or different alleles that complicates statistical analysis) (Li *et al.* 2012; McGill *et al.* 2013). Ascertainment bias can be reduced with programs designed to accommodate it (e.g. IQ-TREE and RAxML) (Stamatakis 2014; Nguyen *et al.* 2015) whereas, missing data effects can be reduced by selecting and preserving genetic material properly, correctly preforming its extraction and library preps, and the appropriately filtering after sequencing used.

What genomic data are needed to obtain SNPs?

SNPs can be extracted from any marker type but most commonly they are obtained through RADseq, transcriptomes, and WGS studies (Baird *et al.* 2008; Hohenlohe et al. 2010,

2011, 2012; Narum *et al.* 2013; Wagner *et al.* 2013; Harvey *et al.* 2016; Leaché and Oaks 2017). For marker types that would otherwise be too large in size to be analyzed with available computational power (e.g. WGS), the use of SNPs offers an extreme reduction in dataset size while maximizing the number of variant sites that are critical for phylogenomic inference. There is often a level of difficulty involved in computationally collecting and constructing a SNP database, calling variants, and filtering for high-quality SNPs versus low-quality or sequencing errors. The common approach of blindly filtering out low-frequency alleles (rare SNPs) can create a biased dataset; one must be cautious and mindful when filtering SNP data to prevent removing important biological information (McGill *et al.* 2013). When using SNPs in any capacity, the quality of the reference genome is of great importance. Some marker types that are typically used for SNP analysis such as RADseq data do not require a reference genome, though it may still be beneficial to include one. A long-read assembled reference genome with large contigs and a small number of gaps can greatly improve SNP calls depending on the taxonomic level and scale of your project.

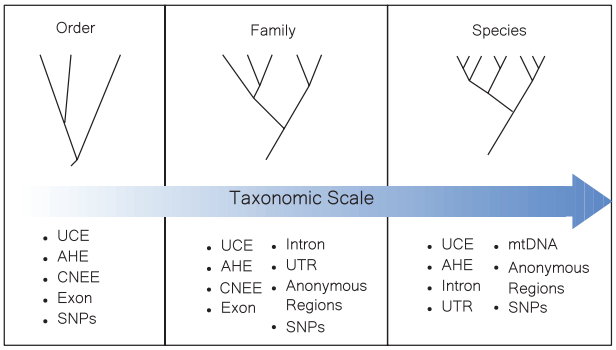

Fig. 3. The various marker types are not equal in their ability to reconstruct reliable and well supported trees at all phylogenetic scales. Here, we provide advice regarding which type may be most appropriate at various phylogenetic scales. Each time point has a list of corresponding marker types that could be used to potentially maximize the biological functionality and evolutionary changes the marker type represents. The complied lists are based mostly on general substitution rates and past studies that have used such marker types.

Conclusion

Just as Sanger sequencing revolutionized the field of phylogenetics, the development of Next Generation Sequencing technologies has dramatically changed our ability to address complex questions in evolutionary biology by providing cost-effective means to generate genome-scale datasets. Genomic data are mosaic and fluid in relationship to substitution rate, which makes the idea of choosing <1% to 5% of the entire genome for phylogenomic inference difficult. We recommend the following criteria for choosing marker types for a specific phylogenomic study: 1) Financial barriers are important to consider in any study. We advocate choosing a method or marker type that is cost effective yet relevant for phylogenomic inference given the taxa under study (Table 1). For example, enrichment methods typically increase the cost of library preps, but do not require as much expensive sequencing, which would be advantageous for organisms with moderate to large genomes (over a Gb in size). When working with a species with a small genome (500 MB or less), it may be more cost effective to sequence the whole genome than to use an enrichment method, even if only a portion of the genome will be used, particularly as costs for library preparation continue to drop. 2) Computational power is also an important consideration for all phylogenomic analyses. For example, analyzing all the exonic or intronic regions of an organism by running a gene tree to species tree pipeline on a laptop or insufficient server could take months or never finish reach convergence. The more loci and/or taxa being analyzed at once, the more computation time will be required for the server in question. Adjusting the number of loci targeted, filtering loci to include only the most informative sequences, or other approaches may be used if further reduction in dataset size is critical. 3) The taxonomic scale of a particular study will help guide the selection of genomic markers; focus on genomic marker types that will complement the divergence scale of the taxa and project (Fig. 3 and Table 1). Using UCEs, AHEs, SNPs (depending on how they were obtained), or exons may be good choices for very deep divergences, with introns and UTRs also sufficient at moderate divergences. If working with a group that has undergone a recent and rapid radiation, SNPs, mtDNA, or noncoding anonymous regions are likely to be the most informative. 4) The availability of published data is key to help with genomic marker choice. For some marker types, data are already readily available. We suggest scanning data archives such as the "National Center of Biotechnology Information (NCBI)," "UK BioBank," or "Ensembl genome database" to determine what data already exist and can be complemented by new data collection. Utilizing the extensive data already freely available online can be a no-cost option to expand a project. 5) The type of tissue needed for a particular genomic marker may limit the choice of certain marker types (i.e. WGS works best with high-quality tissue samples whereas target capture can work with lower-quality tissue samples). Here, constraints will vary during the data collection stage of the project depending on resources and availability of genetic material. 6) Understanding phylogenetic informativity prior to selection of marker type is important. For example, using UCE datasets, Jarvis et al. used 2,509 loci but only 1,062 loci were shown to be informative and Hosner et al. used 462 loci but it was shown that 37 loci were phylogenetically informative (Jarvis *et al.* 2014; Hosner *et al.* 2016). Multiple factors could have influenced the signal difference observed between these studies (i.e. timescale, probe set, data processing pipeline, Quality Control of data). More sites do not always equal higher signal and *only* focusing on marker type selection does not always equal similar signal of previously studied systems.

Pairing several different genomic marker types is an encouraged practice within phylogenomics because it may allow improved recovery of relationships across a variety of different evolutionary depths and may allow identification of or ameliorate biases that may be present in some datasets (e.g. avian exons; Jarvis *et al.* 2014; Kimball and Braun 2021). When pairing different marker types in an empirical study each marker type can be analyzed independently and later, once phylogenetic trees are constructed, can be used to compare topologies. Complementary genomic marker type pairs may include, for example, UCEs and AHEs. As a result, each may contain unique phylogenetic signal (and could increase sites of informativity) and produce discordant topologies (Degnan and Rosenberg 2006). There are several processes

that can lead to discordance or incongruent trees; incomplete linage sorting, technical artifacts, gene loss or duplication, and introgression can all play a role making the loci of different marker types appear to be incongruent with one another (Martin *et al.* 2017; Martin and Höhna 2018). Discordance among genomic marker types may help better understand the evolutionary history of independent features of the genome that are affected by evolution in different ways from one another.

As technology advances, phylogenomics will continue to adapt. Technological and theoretical advances mean that some methods may be very short lived, while others have been used for decades in this field and may remain relevant. We provided recommendations based upon which set of marker types best suit the taxa under exploration, and the biological questions being asked (Fig. 3 and Table 1). We hope this review helps the novice entering the field of phylogenomics by better acclimating them to the various marker types available and help them in their journey of adding to the scientific community.

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Acknowledgments

We are grateful for Dr. Scott Edwards of Harvard University, and his contribution during the planning and reviewing stages of this article. We are also grateful for Drs. Stacey D. Smith and Samuel M. Flaxman of University of Colorado, Boulder, for their contribution during earlier stages of this project. We are also grateful for Research Triangle Institute International for funding and support. Lastly, we are grateful to Brittaney Buchanan for reviewing the manuscript during the later stages of the project.

Authors' contribution

Javan K. Carter: conceptualization, data curation, investigation, review and editing, writing-original draft, project administration. Rebecca T. Kimball: conceptualization, investigation, review and editing. Erik R. Funk: visualization, review and editing. Nolan C. Kane and Drew R. Schield: conceptualization, review and editing. Garth M. Spellman: review and editing. Rebecca J. Safran: conceptualization, data curation, review and editing, project administration.

Data availability

No empirical data were collected or analyzed during the invited technical review.

References

Allman ES, Kubatko LS, Rhodes JA. Split scores: a tool to quantify phylogenetic signal in genome-scale data. *System. Biol.* 2016; 66(4):620–636.

Andermann T, Torres Jiménez MF, Matos-Maraví P, Batista R, Blanco-Pastor JL, Gustafsson ALS, Kistler L, Liberal IM, Oxelman B, Bacon CD, et al. A guide to carrying out a phylogenomic target sequence capture project. *Front Genet*. 2020;10:1407. doi:10.3389/fgene.2019.01407

Andersen MJ, McCullough JM, Friedman NR, Peterson AT, Moyle RG, Joseph L, Nyári ÁS. Ultraconserved elements resolve genus-level relationships in a major Australasian bird radiation (Aves: Meliphagidae). *Emu*. 2019;119(3):218–232.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*. 2016;17(2):81–92.

Avise JC. *Molecular markers, natural history and evolution*. Springer Science & Business Media; 2012.

Avise JC, Robinson TJ. Hemiplasy: a new term in the lexicon of phylogenetics. *Syst Biol*. 2008;57(3):503–507.

Avise JC, Shapira JF, Daniel SW, Aquadro CF, Lansman RA. Mitochondrial DNA differentiation during the speciation process in *Peromyscus*. *Mol Biol Evol*. 1983;1(1):38–56.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*. 2008;3(10):e3376.

Balloux F. The worm in the fruit of the mitochondrial DNA tree. *Heredity*. 2010;104(5):Article 5.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. Ultraconserved elements in the human genome. *Science*. 2004;304(5675):1321–1325.

Bi K, Vanderpool D, Singhal S, Linderoth T, Moritz C, Good JM. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics*. 2012;13(1):403.

Bicknell AA, Cenik C, Chua HN, Roth FP, Moore MJ. Introns in UTRs: why we should stop ignoring them. *Bioessays*. 2012;34(12):1025–1034.

Bonilla AJ, Braun EL, Kimball RT. Comparative molecular evolution and phylogenetic utility of 3′-UTRs and introns in Galliformes. *Mol Phylogenet Evol*. 2010;56(2):536–542.

Boore JL. Animal mitochondrial genomes. *Nucleic Acids Res*. 1999;27(8):1767–1780.

Boore JL, Brown WM. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr Opin Genet Dev*. 1998;8(6):668–674.

Bragg JG, Potter S, Bi K, Moritz C. Exon capture phylogenomics: efficacy across scales of divergence. *Mol Ecol Resour*. 2016;16(5):1059–1068.

Braun EL, Kimball RT. Data types and the phylogeny of Neoaves. *Birds*. 2021;2(1):Article 1.

Braun EL, Kimball RT, Han K-L, Iuhasz-Velez NR, Bonilla AJ, Chojnowski JL, Smith JV, Bowie RC, Braun MJ, Hackett SJ, et al. Homoplastic microinversions and the avian tree of life. *BMC Evol Biol*. 2011;11(1):141.

Brown WM, George M, Wilson AC. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA*. 1979;76(4):1967–1971.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV. The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol Evol*. 2003;18(5):249–256.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol Biol Evol*. 2012;29(8):1917–1932.

Cameron SL, Dowton M, Castro LR, Ruberu K, Whiting MF, Austin AD, Diement K, Stevens J. Mitochondrial genome organization and phylogeny of two vespid wasps. *Genome*. 2008;51(10):800–808.

Cameron SL, Lambkin CL, Barker SC, Whiting MF. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. *Syst Entomol*. 2007;32(1):40–59.

Cenik C, Chua HN, Zhang H, Tarnawsky SP, Akef A, Derti A, Tasan M, Moore MJ, Palazzo AF, Roth FP. Genome analysis reveals interplay between 5′UTR introns and nuclear mRNA export for secretory and mitochondrial genes. *PLoS Genet*. 2011;7(4):e1001366.

Chen M-Y, Liang D, Zhang P. Phylogenomic resolution of the phylogeny of Laurasiatherian mammals: exploring phylogenetic signals

within coding and noncoding sequences. *Genome Biol Evol*. 2017;9(8):1998–2012.
Chifman J, Kubatko L. Quartet inference from SNP data under the coalescent model. *Bioinformatics*. 2014;30(23):3317–3324.
Chojnowski JL, Kimball RT, Braun EL. Introns outperform exons in analyses of basal avian phylogeny using clathrin heavy chain genes. *Gene*. 2008;410(1):89–96.
Conne B, Stutz A, Vassalli JD. The 3′ untranslated region of messenger RNA: a molecular 'hotspot' for pathology? *Nat Med*. 2000;6(6):Article 6.
Crawford NG, Parham JF, Sellas AB, Faircloth BC, Glenn TC, Papenfuss TJ, Henderson JB, Hansen MH, Simison WB. A phylogenomic analysis of turtles. *Mol Phylogenet Evol*. 2015;83:250–257.
Creer S. Choosing and using introns in molecular phylogenetics. *Evol Bioinform Online*. 2007;3:99–108.
Dapprich J, Ferriola D, Mackiewicz K, Clark PM, Rappaport E, D'Arcy M, Sasson A, Gai X, Schug J, Kaestner KH, et al. The next generation of target capture technologies—large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity. *BMC Genomics*. 2016;17(1):486.
Davey JW, Blaxter ML. RADSeq: next-generation population genetics. *Brief Funct Genomics*. 2010;9(5–6):416–423.
De Maio N, Schrempf D, Kosiol C. PoMo: an allele frequency-based approach for species tree estimation. *Syst Biol*. 2015;64(6):1018–1031.
Degnan JH, Rosenberg NA. Discordance of species trees with their most likely gene trees. *PLoS Genet*. 2006;2(5):e68.
Edwards SV, Cloutier A, Baker AJ. Conserved nonexonic elements: a novel class of marker for phylogenomics. *Syst Biol*. 2017;66(6):1028–1044.
Edwards SV, Potter S, Schmitt CJ, Bragg JG, Moritz C. Reticulation, divergence, and the phylogeography–phylogenetics continuum. *Proc Natl Acad Sci USA*. 2016a;113(29):8025–8032.
Edwards SV, Xi Z, Janke A, Faircloth BC, McCormack JE, Glenn TC, Zhong B, Wu S, Lemmon EM, Lemmon AR, et al. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Mol Phylogenet Evol*. 2016b;94:447–462.
Espeland M, Breinholt J, Willmott KR, Warren AD, Vila R, Toussaint EFA, Maunsell SC, Aduse-Poku K, Talavera G, Eastwood, R, et al. A comprehensive and dated phylogenomic analysis of butterflies. *Curr Biol*. 2018;28(5):770–778.e5.
Eytan RI, Evans BR, Dornburg A, Lemmon AR, Lemmon EM, Wainwright PC, Near TJ. Are 100 enough? Inferring acanthomorph teleost phylogeny using Anchored Hybrid Enrichment. *BMC Evol Biol*. 2015;15(1):113.
Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst Biol*. 2012;61(5):717–726.
Fenn JD, Song H, Cameron SL, Whiting MF. A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. *Mol Phylogenet Evol*. 2008;49(1):59–68.
Field DJ, Gauthier JA, King BL, Pisani D, Lyson TR, Peterson KJ. Toward consilience in reptile phylogeny: miRNAs support an archosaur, not lepidosaur, affinity for turtles. *Evol Dev*. 2014;16(4):189–196.
Foley NM, Thong VD, Soisook P, Goodman SM, Armstrong KN, Jacobs DS, Puechmaille SJ, Teeling EC. How and why overcome the impediments to resolution: lessons from rhinolophid and hipposiderid bats. *Mol Biol Evol*. 2015;32(2):313–333.
Funk DJ, Omland KE. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Evol Syst*. 2003;34(1):397–423.
George RD, McVicker G, Diederich R, Ng SB, MacKenzie AP, Swanson WJ, Shendure J, Thomas JH. Trans genomic capture and sequencing of primate exomes reveals new targets of positive selection. *Genome Res*. 2011;21(10):1686–1694.
Gibson A, Gowri-Shankar V, Higgs PG, Rattray M. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. *Mol Biol Evol*. 2005;22(2):251–264.
Godwin RL, Opatova V, Garrison NL, Hamilton CA, Bond JE. Phylogeny of a cosmopolitan family of morphologically conserved trapdoor spiders (Mygalomorphae, Ctenizidae) using Anchored Hybrid Enrichment, with a description of the family, Halonoproctidae Pocock 1901. *Mol Phylogenet Evol*. 2018;126:303–313.
Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han K-L, Harshman J, et al. A phylogenomic study of birds reveals their evolutionary history. *Science*. 2008;320(5884):1763–1768.
Haddad S, Shin S, Lemmon AR, Lemmon EM, Svacha P, Farrell B, Ślipiński A, Windsor D, Mckenna DD. Anchored hybrid enrichment provides new insights into the phylogeny and evolution of longhorned beetles (Cerambycidae). *Syst Entomol*. 2018;43(1):68–89.
Hamilton CA, Lemmon AR, Lemmon EM, Bond JE. Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evol Biol*. 2016;16(1):212.
Han K-L, Braun EL, Kimball RT, Reddy S, Bowie RCK, Braun MJ, Chojnowski JL, Hackett SJ, Harshman J, Huddleston CJ, et al. Are transposable element insertions homoplasy free? An examination using the avian tree of life. *Syst Biol*. 2011;60(3):375–386.
Harrison PW, Jordan GE, Montgomery SH. SWAMP: sliding window alignment masker for PAML. *Evol Bioinform*. 2014;10. doi:10.4137/EBO.S18193
Harvey MG, Brumfield RT. Genomic variation in a widespread Neotropical bird (Xenops minutus) reveals divergence, population expansion, and gene flow. *Molecular Phylogenetics Evol*. 2015;83:305–316.
Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Syst Biol*. 2016;65(5):910–924.
Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, et al. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European Haplogroups. *Am J Hum Genet*. 2002;70(5):1152–1171.
Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Mol Ecol Resour*. 2011;11(s1):117–122.
Hohenlohe PA, Bassham S, Currey M, Cresko WA. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos Trans R Soc B Biol Sci*. 2012;367(1587):395–408.
Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*. 2010;6(2):e1000862.
Hosner PA, Faircloth BC, Glenn TC, Braun EL, Kimball RT. Avoiding missing data biases in phylogenomic inference: an empirical study in the Landfowl (Aves: Galliformes). *Mol Biol Evol*. 2016;33(4):1110–1125.
Houde P, Braun EL, Narula N, Minjares U, Mirarab S. Phylogenetic signal of indels and the Neoavian radiation. *Diversity*. 2019;11(7):Article 7.
Hurst GDD, Jiggins FM. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc R Soc B Biol Sci*. 2005;272(1572):1525–1534.
Ilves KL, López-Fernández H. A targeted next-generation sequencing toolkit for exon-based cichlid phylogenomics. *Mol Ecol Resour*. 2014;14(4):802–811.
Irisarri I, Meyer A. The identification of the closest living relative(s) of tetrapods: phylogenomic lessons for resolving short ancient internodes. *Syst Biol*. 2016;65(6):1057–1075.

Janouškovec J, Tikhonenkov DV, Burki F, Howe AT, Rohwer FL, Mylnikov AP, Keeling PJ. A new lineage of eukaryotes illuminates early mitochondrial genome reduction. *Curr Biol*. 2017;27(23):3717–3724.e5.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C. Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*. 2014;346(6215):1320–1331.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C. Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al.; Consortium, T. A. P. Phylogenomic analyses data of the avian phylogenomics project. *GigaScience*. 2015;4(1). doi:10.1186/s13742-014-0038-1

Jiang F, Pan X, Li X, Yu Y, Zhang J, Jiang H, Dou L, Zhu S. The first complete mitochondrial genome of *Dacus longicornis* (Diptera: Tephritidae) using next-generation sequencing and mitochondrial genome phylogeny of Dacini tribe. *Sci Rep*. 2016;6(1):Article 1.

Jiang J, Yuan H, Zheng X, Wang Q, Kuang T, Li J, Liu J, Song S, Wang W, Cheng F, et al. Gene markers for exon capture and phylogenomics in ray-finned fishes. *Ecol Evol*. 2019;9(7):3973–3983.

Karin BR, Gamble T, Jackman TR. Optimizing phylogenomics with rapidly evolving long exons: comparison with anchored hybrid enrichment and ultraconserved elements. *Mol Biol Evol*. 2020;37(3):904–922.

Kuhl H, Frankl-Vilches C, Bakker A, Mayr G, Nikolaus G, Boerno ST, Klages S, Timmermann B, Gahr M. An unbiased molecular approach using 3′UTRs resolves the avian family-level tree of life. *Mol Biol Evol*. 2020. doi:10.1093/molbev/msaa191

Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, Tissières V, Pickle CS, Plajzer-Frick I, Lee EA, et al. Progressive loss of function in a limb enhancer during snake evolution. *Cell*. 2016;167(3):633–642.e11.

Lachance J, Tishkoff SA. SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *BioEssays*. 2013;35(9):780–786.

Leaché AD, Oaks JR. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annu Rev Ecol Evol Syst*. 2017;48(1):69–84.

Leal F, Cohn MJ. Loss and re-emergence of legs in snakes by modular evolution of sonic hedgehog and HOXD enhancers. *Curr Biol*. 2016;26(21):2966–2973.

Lemmon AR, Emme SA, Lemmon EM. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst Biol*. 2012;61(5):727–744.

Li M, Badger JH, Chen X, Kwong S, Kearney P, Zhang H. An information-based sequence distance and its application to whole mitochondrial genome phylogeny. *Bioinformatics*. 2001;17(2):149–154.

Li Z, Gopal V, Li X, Davis JM, Casella G. Simultaneous SNP identification in association studies with missing data. *Ann Appl Stat*. 2012;6(2):432–456.

Liang B, Wang N, Li N, Kimball RT, Braun EL. Comparative genomics reveals a burst of homoplasy-free numt insertions. *Mol Biol Evol*. 2018;35(8):2060–2064.

Lowe CB, Clarke JA, Baker AJ, Haussler D, Edwards SV. Feather development genes and associated regulatory innovation predate the origin of Dinosauria. *Mol Biol Evol*. 2015;32(1):23–28.

Lowe CB, Kellis M, Siepel A, Raney BJ, Clamp M, Salama SR, Kingsley DM, Lindblad-Toh K, Haussler D. Three periods of regulatory innovation during vertebrate evolution. *Science*. 2011;333(6045):1019–1024.

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, et al. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science*. 2005;308(5724):1034–1036.

Maddison WP. Gene trees in species trees. *Syst Biol*. 1997;46(3):523–536.

Maddison WP, Evans SC, Hamilton CA, Bond JE, Lemmon AR, Lemmon EM. A genome-wide phylogeny of jumping spiders (Araneae, Salticidae), using anchored hybrid enrichment. *ZooKeys*. 2017;695:89–101.

Maney DL, Goodson JL. Chapter 5—Neurogenomic mechanisms of aggression in songbirds. In: Huber R, Bannasch DL, Brennan P, editors. *Advances in genetics*, Vol. 75. Academic Press; 2011. p. 83–119.

Marcovitz A, Jia R, Bejerano G. "Reverse genomics" predicts function of human conserved noncoding elements. *Mol Biol Evol*. 2016;33(5):1358–1369.

Martin CH, Höhna S. New evidence for the recent divergence of Devil's Hole pupfish and the plausibility of elevated mutation rates in endangered taxa. *Mol Ecol*. 2018;27(4):831–838.

Martin CH, Höhna S, Crawford JE, Turner BJ, Richards EJ, Simons LH. The complex effects of demographic history on the estimation of substitution rate: concatenated gene analysis results in no more than twofold overestimation. *Proc R Soc B Biol Sci*. 2017;284(1860):20170537.

Mason NA, Olvera-Vital A, Lovette IJ, Navarro-Sigüenza AG. Hidden endemism, deep polyphyly, and repeated dispersal across the Isthmus of Tehuantepec: diversification of the White-collared Seed-eater complex (Thraupidae: *Sporophila torqueola*). *Ecol Evol*. 2018;8(3):1867–1881.

Matthee CA, Eick G, Willows-Munro S, Montgelard C, Pardini AT, Robinson TJ. Indel evolution of mammalian introns and the utility of non-coding nuclear markers in eutherian phylogenetics. *Mol Phylogenet Evol*. 2007;42(3):827–837.

McCormack JE, Maley JM, Hird SM, Derryberry EP, Graves GR, Brumfield RT. Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Mol Phylogenet Evol*. 2012;62(1):397–406.

McGill JR, Walkup EA, Kuhner MK. Correcting coalescent analyses for panel-based SNP ascertainment. *Genetics*. 2013;193(4):1185–1196.

McKenzie PF, Eaton DAR. The multispecies coalescent in space and time, bioRxiv, doi:10.1101/2020.08.02.233395, 2020, preprint: not peer reviewed.

Meredith RW, Hekkala ER, Amato G, Gatesy J. A phylogenetic hypothesis for *Crocodylus* (Crocodylia) based on mitochondrial DNA: evidence for a trans-Atlantic voyage from Africa to the New World. *Mol Phylogenet Evol*. 2011;60(1):183–191.

Milián-García Y, Amato G, Gatesy J, Hekkala E, Rossi N, Russello M. Phylogenomics reveals novel relationships among Neotropical crocodiles (*Crocodylus* spp.). *Mol Phylogenet Evol*. 2020;152:106924.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res*. 2007;17(2):240–248.

Minegishi Y, Aoyama J, Inoue JG, Miya M, Nishida M, Tsukamoto K. Molecular phylogeny and evolution of the freshwater eels genus *Anguilla* based on the whole mitochondrial genome sequences. *Mol Phylogenet Evol*. 2005;34(1):134–146.

Morin PA, Luikart G, Wayne RK; The SNP Workshop Group. SNPs in ecology, evolution and conservation. *Trends Ecol Evol*. 2004;19(4):208–216.

Murphy WJ, Pevzner PA, O'Brien SJ. Mammalian phylogenomics comes of age. *Trends Genet*. 2004;20(12):631–639.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. Genotyping-by-sequencing in ecological and conservation genomics. *Mol Ecol*. 2013;22(11):2841–2847.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–274.

Nikaido M, Rooney AP, Okada N. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: hippopotamuses are the closest extant relatives of whales. *Proc Natl Acad Sci USA*. 1999;96(18):10261–10266.

O'Hara TD, Hugall AF, Thuy B, Stöhr S, Martynov AV. Restructuring higher taxonomy using broad-scale phylogenomics: the living Ophiuroidea. *Mol Phylogenet Evol*. 2017;107:415–430.

Peng Z, Wang J, He S. The complete mitochondrial genome of the helmet catfish *Cranoglanis bouderius* (Siluriformes: Cranoglanididae)

and the phylogeny of otophysan fishes. *Gene*. 2006;376(2):290–297.
Pepper M, Sumner J, Brennan IG, Hodges K, Lemmon AR, Lemmon EM, Peterson G, Rabosky DL, Schwarzkopf L, Scott IAW, et al. Speciation in the mountains and dispersal by rivers: molecular phylogeny of *Eulamprus* water skinks and the biogeography of Eastern Australia. *J Biogeogr*. 2018;45(9):2040–2052.
Philippe H, Delsuc F, Brinkmann H, Lartillot N. Phylogenomics. *Annu Rev Ecol Evol Syst*. 2005;36(1):541–562.
Philippe H, Snell EA, Bapteste E, Lopez P, Holland PWH, Casane D. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol Biol Evol*. 2004;21(9):1740–1752.
Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. 2010;20(1):110–121.
Portik DM, Smith LL, Bi K. An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). *Mol Ecol Resour*. 2016;16(5):1069–1083.
Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*. 2015;526(7574):Article 7574.
Rearick D, Prakash A, McSweeny A, Shepard SS, Fedorova L, Fedorov A. Critical association of ncRNA with introns. *Nucleic Acids Res*. 2011;39(6):2357–2366.
Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*. 2010;468(7327):Article 7327.
Rokas A, Holland PWH. Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol*. 2000;15(11):454–459.
Rubinoff D, Holland BS. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Syst Biol*. 2005;54(6):952–961.
Saccone C, Lanave C, De Grassi A. Metazoan OXPHOS gene families: evolutionary forces at the level of mitochondrial and nuclear genomes. *Biochim Biophys Acta Bioenerg*. 2006;1757(9):1171–1178.
Sakharkar MK, Chow VT, Kangueane P. Distributions of exons and introns in the human genome. *In Silico Biol*. 2004;4(4):387–393. https://pubmed.ncbi.nlm.nih.gov/15217358/
Salicini I, Ibáñez C, Juste J. Multilocus phylogeny and species delimitation within the Natterer's bat species complex in the Western Palearctic. *Mol Phylogenet Evol*. 2011;61(3):888–898.
Scornavacca C, Galtier N. 2017. Incomplete lineage sorting in mammalian phylogenomics. *Syst Biol*. 2017;66(1):112–120.
Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, Kishiro T, Goto M, Munechika I, Okada N. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature*. 1997;388(6643):Article 6643.
Shin S, Clarke DJ, Lemmon AR, Moriarty Lemmon E, Aitken AL, Haddad S, Farrell BD, Marvaldi AE, Oberprieler RG, McKenna DD. Phylogenomic data yield new and robust insights into the phylogeny and evolution of weevils. *Mol Biol Evol*. 2018;35(4):823–836.
Shuster RC, Rubenstein AJ, Wallace DC. Mitochondrial DNA in anucleate human blood cells. *Biochem Biophys Res Commun*. 1988;155(3):1360–1365.
Singhal S, Grundler M, Colli G, Rabosky DL. Squamate Conserved Loci (SqCL): a unified set of conserved loci for phylogenomics and population genetics of squamate reptiles. *Mol Ecol Resour*. 2017;17(6):e12–e24.
Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Syst Biol*. 2014;63(1):83–95.
Springer MS, Molloy EK, Sloan DB, Simmons MP, Gatesy J. ILS-aware analysis of low-homoplasy retroelement insertions: inference of species trees and introgression using quartets. *J Hered*. 2020;111(2):147–168.
Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–1313.
Stebbins-Boaz B, Richter JD. Translational control during early development. *Crit Rev Eukaryot Gene Expr*. 1997;7(1–2). doi:10.1615/CritRevEukarGeneExpr.v7.i1-2.50
Stephen S, Pheasant M, Makunin IV, Mattick JS. Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. *Mol Biol Evol*. 2008;25(2):402–408.
Streicher JW, Schulte JA, II, Wiens JJ. How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in Iguanian lizards. *Syst Biol*. 2016;65(1):128–145.
Teasdale LC, Köhler F, Murray KD, O'Hara T, Moussalli A. Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture. *Mol Ecol Resour*. 2016;16(5):1107–1123.
Tyagi K, Chakraborty R, Cameron SL, Sweet AD, Chandra K, Kumar V. Rearrangement and evolution of mitochondrial genomes in Thysanoptera (Insecta). *Sci Rep*. 2020;10(1):Article 1.
Vachaspati P, Warnow T. SVDquest: improving SVDquartets species tree estimation using exact optimization within a constrained search space. *Mol Phylogenet Evol*. 2018;124:122–136.
Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Afzal V, Rubin EM, Pennacchio LA. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet*. 2008;40(2):Article 2.
Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S, Greuter L, Sivasundar A, Seehausen O. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Mol Ecol*. 2013;22(3):787–798.
Wang W, Fang D, Gan J, Shi Y, Tang H, Wang H, Fu M, Yi J. Evolutionary and functional implications of 3′ untranslated region length of mRNAs by comprehensive investigation among four taxonomically diverse metazoan species. *Genes Genomics*. 2019;41(7):747–755.
Wiens JJ, Morrill MC. Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Syst Biol*. 2011;60(5):719–731.
Winker K, Glenn TC, Faircloth BC. Ultraconserved elements (UCEs) illuminate the population genomics of a recent, high-latitude avian speciation event. *PeerJ*. 2018;6:e5735.
Xiong P, Hulsey CD, Meyer A, Franchini P. Evolutionary divergence of 3′ UTRs in cichlid fishes. *BMC Genomics*. 2018;19(1):433.
Young AD, Gillung JP. Phylogenomics—principles, opportunities and pitfalls of big-data phylogenetics. *Syst Entomol*. 2020;45(2):225–247.
Young AD, Lemmon AR, Skevington JH, Mengual X, Ståhls G, Reemer M, Jordaens K, Kelso S, Lemmon EM, Hauser M, et al. Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae). *BMC Evol Biol*. 2016;16(1):143.
Zhang Y, Deng S, Liang D, Zhang P. Sequence capture across large phylogenetic scales by using pooled PCR-generated baits: a case study of Lepidoptera. *Mol Ecol Resour*. 2019;19(4):1037–1051.
Zhang J, Lai J. Phylogenomic approaches in systematic studies. *Zool. System*. 2020;45(3):151–162
Zhu K, Lü Z, Liu L, Gong L, Liu B. The complete mitochondrial genome of *Trachidermus fasciatus* (Scorpaeniformes: Cottidae) and phylogenetic studies of Cottidae. *Mitochondrial DNA B*. 2018;3(1):301–302.