Molecular Biology and Evolution, 2026, 43, 1–14
https://doi.org/10.1093/molbev/msag155
Published: 29 June 2026
Discoveries

Phylogenomic blind spots: the limits of UCE and BUSCO loci in the presence of gene flow

Nicole M. Foley [1,2]*, **William J. Murphy** [1,2]

[1]Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, USA
[2]Center for Comparative Genomics, Texas A&M AgriLife Research, College Station, TX, USA
*Corresponding author: E-mail: nfoley@cvm.tamu.edu.
Associate editor: Mary O'Connell

Abstract

Ultra-conserved elements (UCEs) and BUSCO genes are commonly used markers in reduced-representation phylogenomic studies. They are valued for their evolutionary conservation, ease of alignment, and cost-effectiveness in generating phylogenomic datasets for nonmodel species. Recombination-aware phylogenomic approaches reveal that, with increased historical and recent gene flow, the species tree may be limited to genomic regions with low recombination rates, whereas introgression-associated alleles are most often found in high-recombining regions. In this study, we aimed to determine whether widely used UCE and BUSCO datasets can reliably recover the species tree in mammalian clades where extensive introgression has previously been documented using recombination-aware methods. Our analyses indicate that UCEs and BUSCO loci are not sampled randomly from the genome and are underrepresented in mammalian sex chromosomes and other low-recombining genomic regions. Concatenation and coalescent-based phylogenomic analyses across 12 clades with varying degrees of gene flow showed that UCEs and BUSCO datasets do not recover the true species topology when introgression is frequent. Although neutral loci are generally preferred for phylogenomic analyses, per-base constraint measures estimated genome-wide show that UCE and BUSCO loci originate from genomic regions under very strong selective constraint. Comparisons of branch lengths and node heights from trees based on accelerated, neutral, and conserved PhyloP datasets revealed that those derived from UCE and BUSCO data are compressed relative to trees from neutral regions. We conclude by proposing mitigation strategies to address some of the issues identified in this study, thereby improving the use of UCE-based or other target-enrichment methods in phylogenomics.

Keywords UCEs, BUSCO genes, phylogenomics, introgression, hybridization, post-speciation gene flow

Introduction

Ultra-conserved elements (UCEs) are short, highly conserved DNA sequences (~20 to 200 bp) (Bejerano et al. 2004; Christmas et al. 2023) that have remained virtually unchanged across lineages over deep evolutionary time (Bejerano et al. 2004). Approximately 500 UCE sequences were first identified within alignments of human, rat, mouse, chicken, and fish genomes (Bejerano et al. 2004). The number detected in mammals alone now approaches 4,500 and are distributed throughout the genome. Although the functions of many noncoding UCEs are not well characterized, their conservation over long evolutionary periods suggests that they have evolved under strong selective constraint and likely serve important regulatory roles (Bejerano et al. 2004; Christmas et al. 2023). Indeed, both exonic and nonexonic UCEs are notably enriched for essential biological processes such as RNA processing and transcriptional regulation, highlighting their likely transcriptional significance (Bejerano et al. 2004; Christmas et al. 2023).

Due to their conservation over deep evolutionary time, UCEs have become anchors for cost-effective targeted enrichment of loci in phylogenomic analyses (Faircloth et al. 2012). These methods enable the simultaneous amplification of hundreds of orthologous loci and are particularly useful for species without prior genome data. The conserved nature of the anchoring UCE makes these loci easy to align, and the co-amplified flanking regions provide essential genetic variation for phylogenomics (Faircloth et al. 2012). Since its initial development, this targeted capture approach has gained popularity and widespread use (Fig. 1). UCEs have been successfully used to generate data for phylogenomic studies across mammals, birds, plants, fish, amphibians, reptiles, anthozoans, arthropods, and platyhelminths, among

Received: February 9, 2026. **Revised:** April 28, 2026. **Accepted:** June 3, 2026

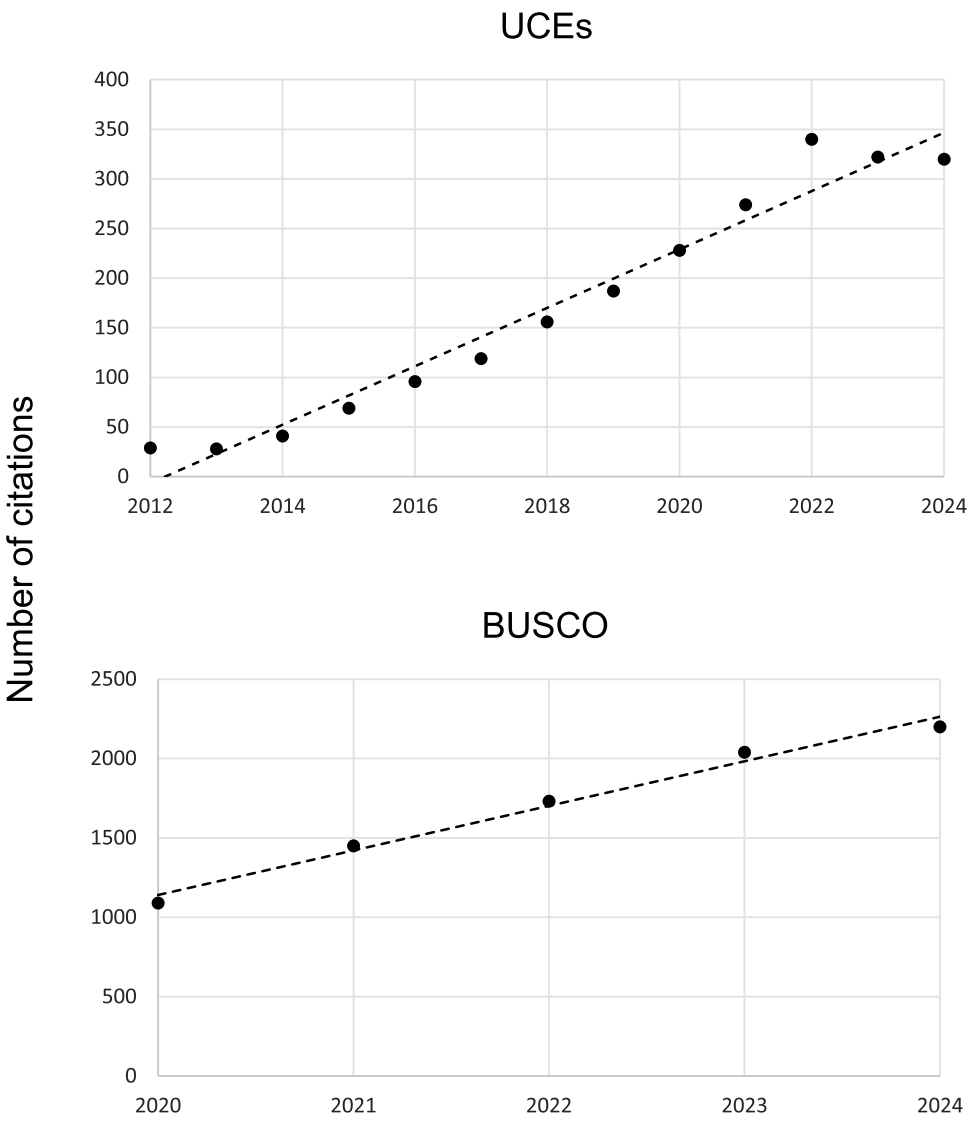

Figure 1 The number of citations per year since the publication of the first paper describing the use of UCEs for phylogenomic analyses (top). The number of citations for the past 5 years for the use of BUSCO loci for phylogenomic analyses (bottom). Citations are derived from Google Scholar searches of "UCE" + "Phylogeny" per year. Similarly, the query was "BUSCO" + "Phylogeny" for BUSCO loci.

others (McCormack et al. 2012, 2013; Quattrini et al. 2018; Streicher et al. 2020; Van Dam et al. 2021; Ebbs et al. 2022; Card et al. 2023; Derkarabetian et al. 2025).

As whole-genome assemblies become increasingly available across many clades of species, identifying orthologous loci for phylogenomic analyses has been a subject of great interest. Because sequence paralogy can confound phylogenomic analyses, many researchers have turned to the Benchmarking Universal Single Copy Ortholog (BUSCO) gene set (Dias et al. 2020; Li et al. 2021; Armstrong et al. 2022; Sahbou et al. 2022; Liu et al. 2024; Chen et al. 2025; Wooldridge et al. 2025; Zhang et al. 2025). These loci comprise orthologs spanning deep evolutionary time and are so conserved that they are widely used to evaluate the completeness of genome assemblies (Simão et al. 2015; Huang and Li 2023).

Although UCEs and BUSCO loci are commonly used by the systematics community to construct phylogenomic datasets, our understanding of genome evolution has changed significantly since their original development and application. Comprehensive whole-genome analyses of phylogenomic signal partitioned by individual windows or loci across entire genomes show that hybridization is far more common than previously believed. Phylogenomic research across diverse vertebrates, arthropods, and plants has demonstrated that when gene flow occurs, much of the genome can be overwritten by phylogenomic signals consistent with introgression (Edelman et al. 2019; Li et al. 2019; Hennelly et al. 2021; Nelson et al. 2021; Feng et al. 2023; Foley et al. 2023; Ottenburghs et al. 2023; Owens et al. 2023; Stubbs et al. 2023). Collectively, these studies have revealed that in the presence of gene flow, (i) the species tree is often not the dominant genome-wide phylogenomic signal, (ii) the distribution of the species tree is nonrandom, and (iii) recombination rate predicts the distribution of the species tree (Burbrink et al. 2025, Foley et al. 2026). Pervasive introgression throughout a clades evolutionary history may restrict the species tree to regions with low recombination, particularly on the X or Z sex chromosomes. As recombination landscapes vary over time and are coupled with hybridization/introgression

histories, genomes can become mosaics of distinct local evolutionary ancestry patterns (Murphy et al. 2021; Harris et al. 2022). The democratic vote approach inherent in many phylogenomic analyses, including UCE-based and BUSCO methods that use concatenation and summary coalescence models, can be confounded by gene flow, often returning trees that reflect past introgression events rather than the true species tree (Fontaine et al. 2015; Edelman et al. 2019; Hibbins and Hahn 2019; Li et al. 2019; Hibbins and Hahn 2022).

Here we evaluated how UCE and BUSCO loci perform under gene flow. Although UCEs are generally thought to be sampled throughout the genome, their original description noted a depletion of these loci on human chr18 and the Y sex chromosome, which may have implications for phylogenomic resolution. Similarly, the genomic positions of UCE and BUSCO loci relative to recombination rate have not been investigated. Neutrally evolving loci are typically preferred for phylogenomic analyses because rate variation among sites, owing to differing selective pressures, can confound results (Yang 1996; Foley et al. 2023). Neutrality is also a key assumption of the multispecies coalescent model, which underpins coalescence-based tree-building approaches (Degnan and Rosenberg 2009). While it is widely understood that UCE and BUSCO loci are highly conserved and subject to purifying selection (Bejerano et al. 2004), we can now use publicly available estimates of per-base evolutionary constraint, PhyloP scores (Sullivan et al. 2023), to place loci derived from UCE-based target enrichment and BUSCO orthology databases in the context of genome-wide constraint and to determine whether any biases are associated with these data types. Finally, we extracted UCE and BUSCO loci from published mammalian whole-genome datasets where gene flow has limited the species tree to low-recombining regions of the genome, to assess whether these widely used phylogenetic markers can recover the species tree in the presence of gene flow.

Results

UCE and BUSCO loci are underrepresented on sex chromosomes

We visualized the genomic locations of UCE loci on the hg38 human genome assembly (Fig. 2a). Our results show that UCE loci are widely distributed throughout the genome but are more densely clustered in regions where loci have been tiled to increase locus length. However, from this genome-wide view, the number of loci on the sex chromosomes (X and Y) appeared reduced. To further examine this observation, we assumed that the number of loci per chromosome should be proportional to chromosome length if loci were randomly distributed. We tested this null hypothesis by fitting a linear model to the ordered distribution of hg38 chromosome lengths, yielding $R^2 = 0.9647$ (Fig. 2b). In contrast, a linear model fitted to the proportion of UCE loci as a function of chromosome length was a poor fit, with $R^2 = 0.0702$ (Fig. 2c). This comparison showed that UCE loci are unlikely to be randomly distributed. Instead, our analysis showed that human chromosomes 2, 10, 14, and 18 contained more UCE loci than predicted by the linear model. Our analysis also confirmed that fewer UCEs were sampled from the human X and Y chromosomes. This is notable given that previous phylogenomic analyses of hybridizing species have shown that phylogenetic branching histories consistent with the species tree are typically enriched in low-recombining regions of the X chromosome, suggesting that UCE loci may yield incorrect topologies in clades with an evolutionary history of introgressive hybridization.

UCE loci occurred at the highest frequency in the highest recombining regions (quintile 5) of the *Bos taurus* and *Giraffa camelopardalis* genomes (Fig. 3a), while UCEs occurred at the lowest frequency in the lowest recombining regions (quintile 1). A similar pattern was observed in the *Felis catus* genome (Fig. 3a). However, loci were more evenly distributed across all five recombination quintiles in the *Ursus arctos* genome (Fig. 3a). Similar variability was observed in the distribution of BUSCO loci, with loci most frequent in high-recombining regions in the *Felis catus* genome but more evenly distributed across recombination quintiles in the *Giraffa camelopardalis* genome (Fig. 3b).

UCE and BUSCO loci do not reliably recover the species tree in clades with histories of extensive introgression

For each clade, the alignment length, number of variable sites, and number of parsimony-informative sites were recorded across all datasets: (i) All UCE loci, (ii) UCE loci from autosomes, and (iii) UCE loci on the X chromosome (Table S2). The relationships recovered at each node for each dataset were recorded and compared with the species tree inferred from the original studies for UCE loci (Fig. 4a; Table S1) and BUSCO loci (Fig. 4b). Clades were assigned to groups following Foley et al. (2026). In Group 2 clades that exhibit moderate discordance due to introgression, the species tree is recovered in all genomic partitions (whole genome, autosomes, X, and high- and low-recombining regions). By contrast, the species tree is restricted to lowest recombining genomic partitions in Group 3 clades which exhibit extensive discordance due to introgression. Group 4 clades exhibit discordance due to incomplete lineage sorting (ILS).

The species tree was recovered across all genomic partitions (All UCEs, autosomal UCEs, X-linked UCEs) in the concatenation-based phylogenomic analysis conducted for group 2 clades, *Martes* and *Sus* (Fig. 4a). For all but one group 3 clade, the topology recovered for the All UCEs and autosomal UCEs partitions did not match the inferred species tree. The only exception was the delphinid clades, where the species tree was recovered for all partitions except the X-linked UCEs (Fig. 4a). The X-linked partition recovered the species tree for only 4 of 8 clades. In group 4 (ILS only clades) the species tree was only recovered by the X-linked UCEs partition (Fig. 4a). Similar results were obtained applying the SVDquartets coalescence-based analysis to the same datasets (Figure S1a). All genomic partitions recovered the species tree for the group 2 *Martes* clade using BUSCO loci (Fig. 4b). For group 3 clades, the species tree was recovered only in the X-linked BUSCO partition for the *Giraffa* clade. No partitioned dataset of BUSCO loci recovered the species tree for the *Panthera* clade. The same topologies were inferred in the coalescence-based analysis of the same datasets (Figure S1b).

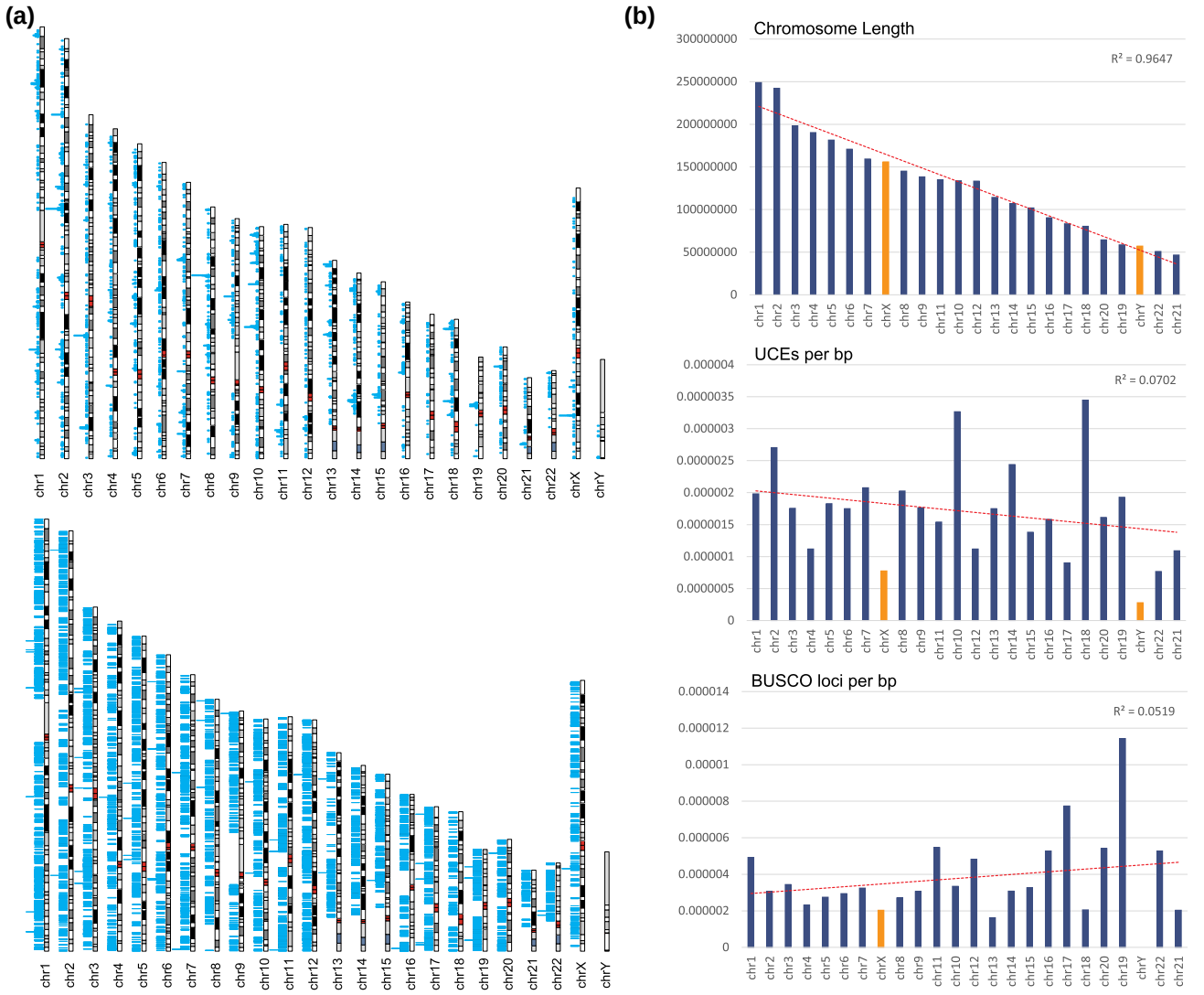

Figure 2 The distribution of a) UCE (top) and BUSCO (bottom) loci across the hg38 human genome assembly. b) Assuming the number of UCE loci per chromosome is proportional to chromosome length, a linear model describes the relationship between the number of UCE loci per chromosome (sex chromosomes are highlighted in orange). The observed data show that neither UCE (middle) nor BUSCO loci (bottom) are uniformly distributed throughout the genome, and the data are poorly described by a linear model. Note that both UCEs and BUSCO loci are particularly underrepresented on sex chromosomes (orange).

Trees derived from UCE and BUSCO loci exhibit branch length distortion relative to neutral loci

We obtained and ranked the per-base averaged PhyloP scores of UCE loci extracted from the Zoonomia alignment of 241 mammals (Fig. 5). The median UCE locus PhyloP score, 3.27, fell within the distribution of values under a significantly strong constraint (PhyloP ≥2.27; FDR 0.5) (Sullivan et al. 2023). For comparison, values derived from first and second codon positions of protein-coding genes had median PhyloP scores of 4.9 and 6.0, respectively. By contrast, third codon positions, which are relatively free to vary due to codon redundancy (Bofkin and Goldman 2007), had a median PhyloP score of 0.68. The median exome score, which should represent a proxy for highly conserved BUSCO loci, is 6.0.

Next, we extracted and plotted branch lengths from phylogenomic analyses of accelerated, neutral, and conserved genomic partitions of the Zoonomia 241 placental mammal alignment (Fig. 6). Branches from the accelerated dataset were longest, those from the conserved dataset were shortest, and neutral branches were intermediate. Differences in branch lengths among the three partitions are most pronounced at intermediate to long branch lengths. Branch lengths for the topology derived from the UCE ML concatenated dataset were most similar to the distribution observed for the conserved dataset (Fig. 6). In addition, Kuhner-Felsenstein's distances, which account for differences in both topology and branch length, indicated that the UCE and BUSCO data were most similar to the conserved PhyloP dataset (Table 1).

Finally, we investigated how selective constraints on loci used in phylogenomic analyses affect interordinal and ordinal node heights (e.g. measurement from tree tip to the node containing

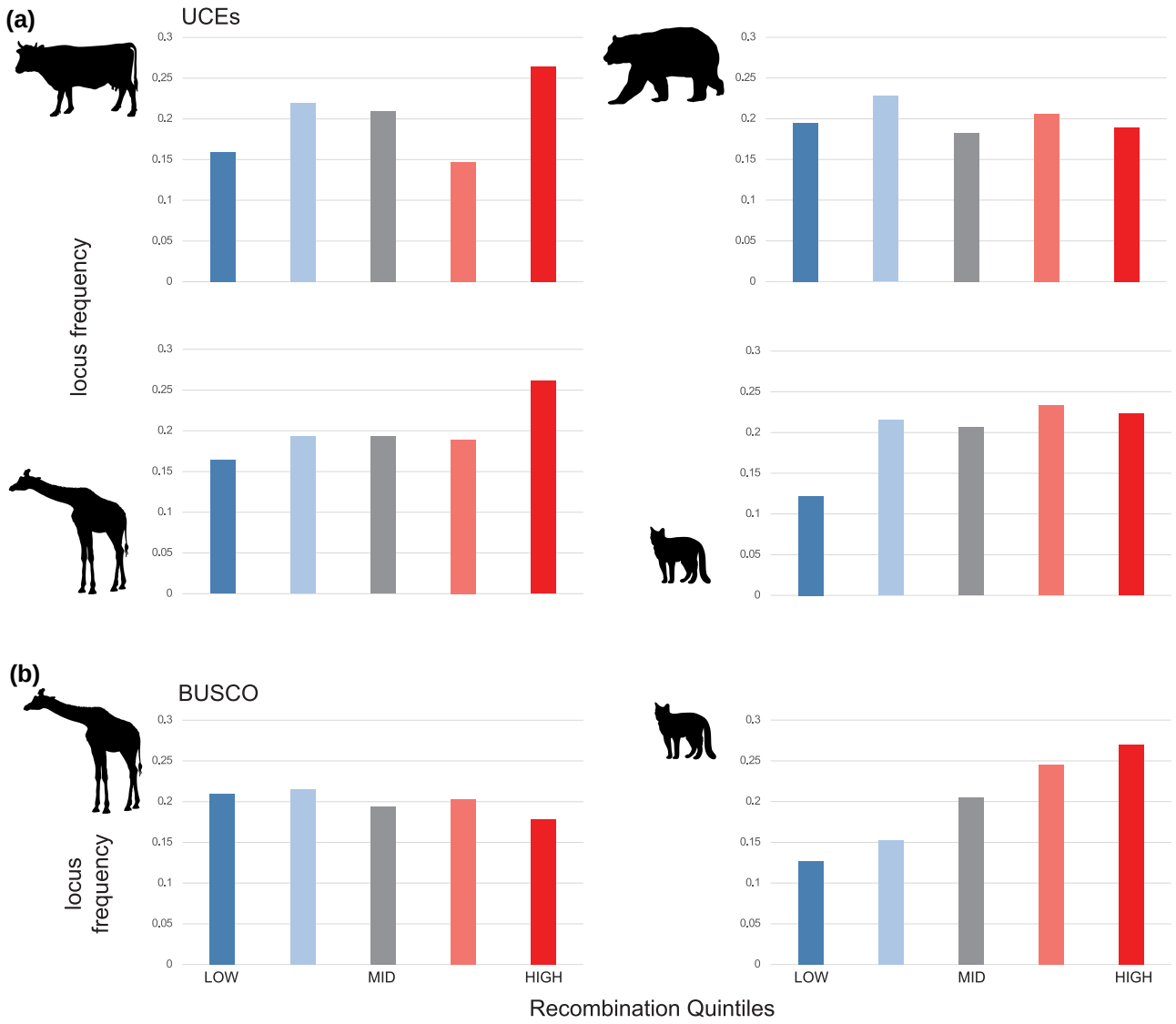

Figure 3 The frequency of UCE loci in genomic regions of high and low recombination a), as defined by whole-genome quintiles. Recombination maps used in this analysis correspond to (left to right) *Bos taurus*, *Ursus arctos*, *Giraffa camelopardalis*, and *Felis catus*. b) The frequency of BUSCO loci in the same recombination regions. All PhyloPic silhouettes are in the public domain under the license https://creativecommons.org/publicdomain/zero/1.0/; *Bos* credit Steven Tarver; *Ursus* credit Tracy Heath; *Giraffa* credit Steven Tarver; *Felis* credit Ferran Sayol.

all bats) in placental mammals (Christmas et al. 2023). Our results show that node heights are systematically compressed in phylogenies derived from UCE, BUSCO, and conserved-site datasets relative to the neutral phylogeny (Fig. 7 and Figure S2). Across all nodes examined, node heights from conserved-, UCE-, and BUSCO-based datasets are compressed relative to those from the neutral dataset.

Discussion

UCE and BUSCO loci have become increasingly popular genomic markers for generating phylogenomic datasets for nonmodel species due to their ease of use, scalability and cost-effective methodology. While datasets generated from these loci have had extraordinary value, expanding our understanding of the Tree of Life, our results indicate that caution is warranted when employing these markers for phylogenomic analyses in clades suspected of post-speciation gene flow. Numerous recent studies across diverse taxa have shown that extensive gene flow produces a nonrandom distribution of phylogenomic signal across the genome, and that the species tree distribution is predicted by recombination rate (Edelman et al. 2019; Li et al. 2019; Foley et al. 2024).

The species tree is typically enriched on the X or Z sex chromosomes due to extended regions of low recombination and their enrichment for loci involved in barriers to gene flow during speciation (Burbrink et al. 2025). Recent studies have identified a 50 Mb X-linked recombination desert that resists gene flow among diverse placental mammals and spans ~100 mya of evolution (Murphy et al. 2021, Foley et al. 2023, 2026). Given its high gene density and low historical recombination rate, this region behaves superficially like a speciation supergene (Thompson and Jiggins 2014). The same region also contains a high density of

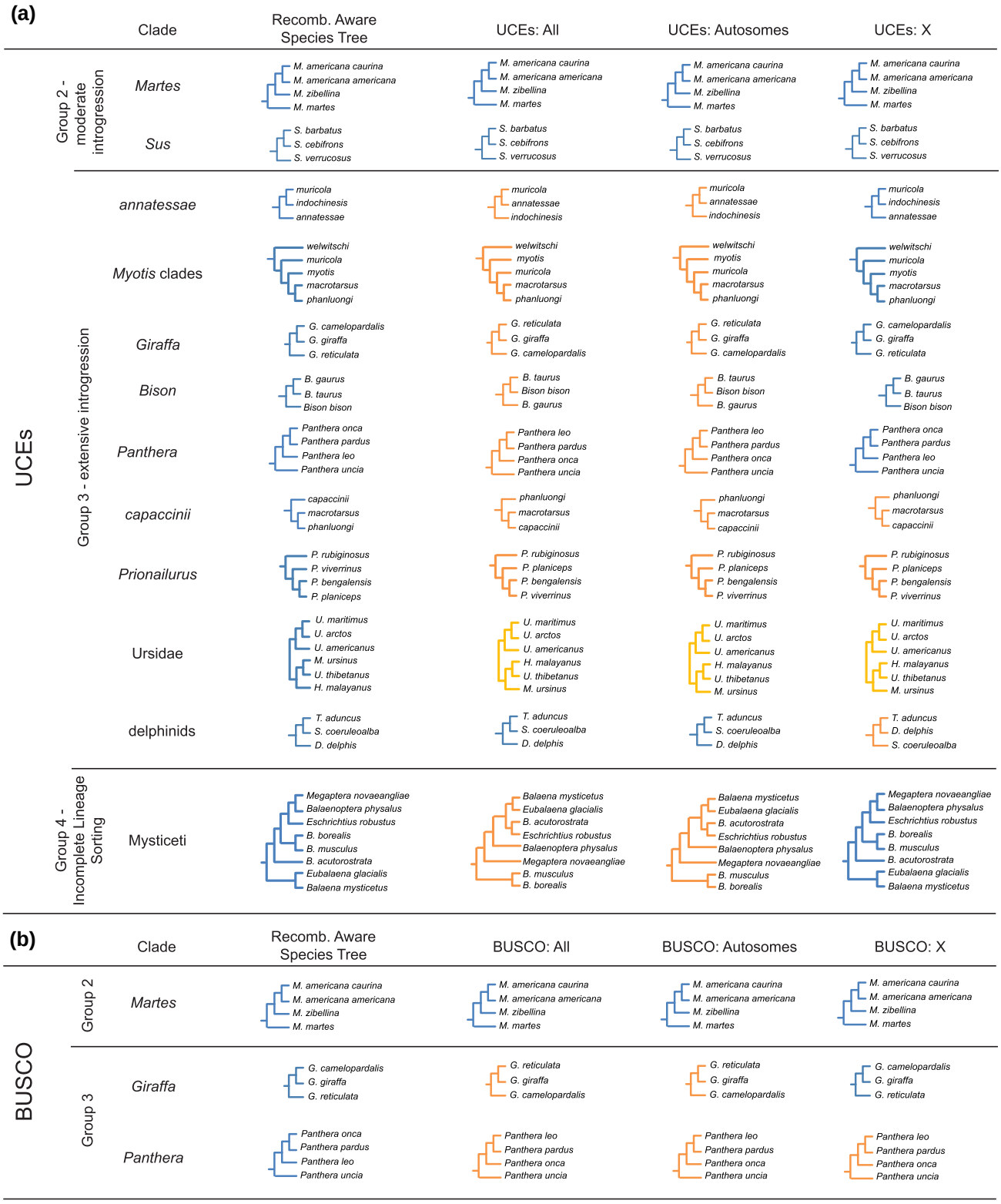

Figure 4 a) Results from IQ-tree2 concatenation-based phylogenomic analysis of UCE datasets. The resulting topologies were compared with the species tree (blue) inferred from previous recombination-aware phylogenomic studies, which showed that these clades evolved with varying degrees of discordance due to introgression and ILS (see Table S1). Three datasets were analyzed for each clade: All UCE loci, Autosomal UCE loci, and X-linked UCE loci. Warm-toned colors indicate where the recovered topology did not match the inferred species tree. b) Results from analysis of BUSCO loci.

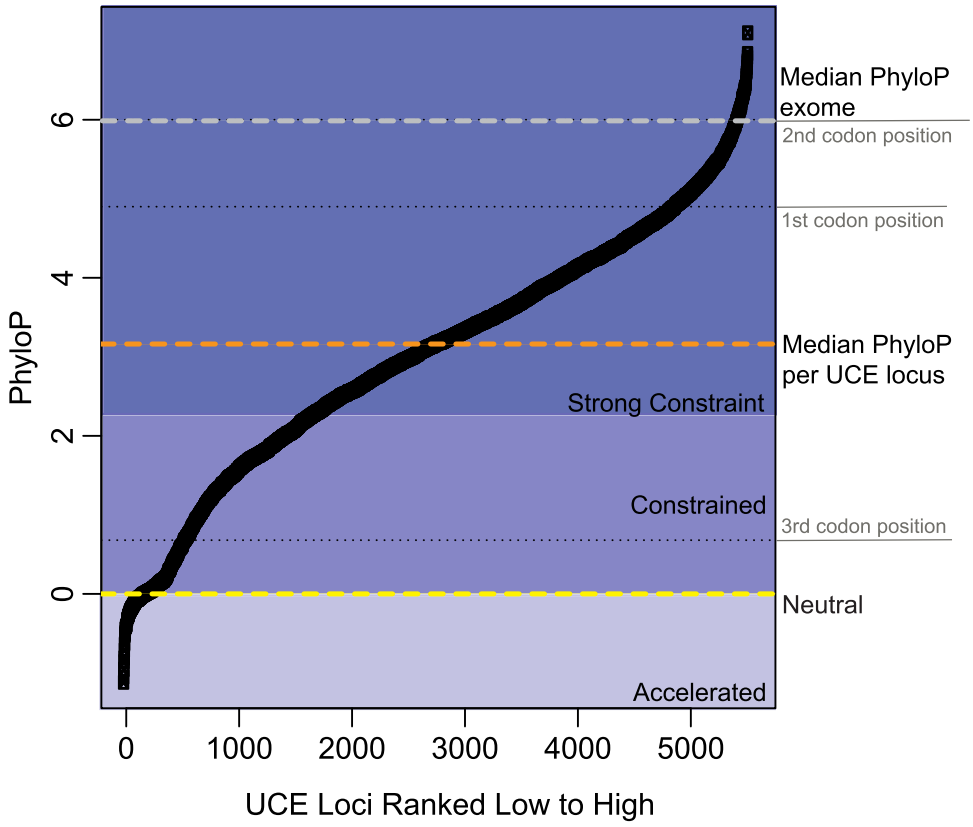

Figure 5 Distribution of median PhyloP scores per UCE locus, across first, second, and third codon positions, and median PhyloP BUSCO/exome scores.

cis-interacting elements involved in X-chromosome inactivation (Deng et al. 2015) and many testis-specific gene clusters, making this region a target for linked selection in the context of reproductive isolation. These architectural and functional genomic features of X and Z chromosomes, coupled with phylogenomic analyses of hybridizing clades supporting their enrichment of the species tree, underscore the necessity of including sex chromosomes in phylogenomic analyses where relevant and available. Importantly, sex chromosomes are undersampled (Fig. 1) for both UCE and BUSCO loci.

Sequence conservation is leveraged in generating phylogenomic datasets using anchor-enriched UCEs or computationally derived BUSCO loci. We have now contextualized these markers in terms of genome-wide constraint. Recent studies estimate that ~10.7% of the human genome is evolving under strong purifying selection (Christmas et al. 2023). To investigate disease-associated changes in the human genome, another study focused on a subset of sites with PhyloP scores ≥2.27 (false discovery rate [FDR] 0.05 threshold), which were considered to be evolving under the strongest measures of constraint, representing just 3.6% of the genome. Our results indicate that the median PhyloP score of UCE and BUSCO loci exceeds this classification and should not be analyzed with coalescent-based approaches that assume neutrality.

Branch lengths are key inputs for many downstream evolutionary studies, including trait-based analyses (Borges et al. 2019), disease transmission dynamics (Guinat et al. 2023), species delimitation (Zhang et al. 2013), measures of phylogenetic diversity (Mooers et al. 2011), and comparative genome analyses (Lin et al. 2019). Our results demonstrate systematic differences in branch lengths among gene trees generated from datasets comprising accelerated, neutral, and conserved sites (Fig. 2b). Branch lengths derived from independent UCE and BUSCO datasets most closely resembled those from the most conserved datasets. Node heights were consistently elevated in the accelerated dataset yet compressed in the UCE and BUSCO datasets relative to the neutral tree. Collectively, our results illustrate a systematic difference in branch lengths linked to data acquisition bias (e.g. neutral versus conserved; UCE or BUSCO datasets). Therefore, we caution against the use of UCE- or BUSCO-derived phylogenies for downstream analyses that rely on branch lengths.

The genome-wide distribution of the species tree is predicted by recombination rate in the presence of post-speciation gene flow (Burbrink et al. 2025). With this in mind, we examined the distribution of UCE and BUSCO loci in previously published whole-genome recombination-aware phylogenomic datasets. In three of the four reference species of the focal species groups (*Bos taurus*, *Giraffa camelopardalis*, and *Felis catus*), where the distribution of UCE loci was examined across the genome, UCE loci occurred more frequently in high-recombining quintiles than in the lowest quintile. In the *Ursus arctos* genome, UCE loci were more evenly distributed across all recombination quintiles. This result suggests that the distribution of UCE loci with respect to local recombination rate has not remained static throughout mammalian chromosome evolution (Fig. 3a). A similar pattern was observed for BUSCO loci, which were most frequently found in the highest recombination quintile in *Felis*

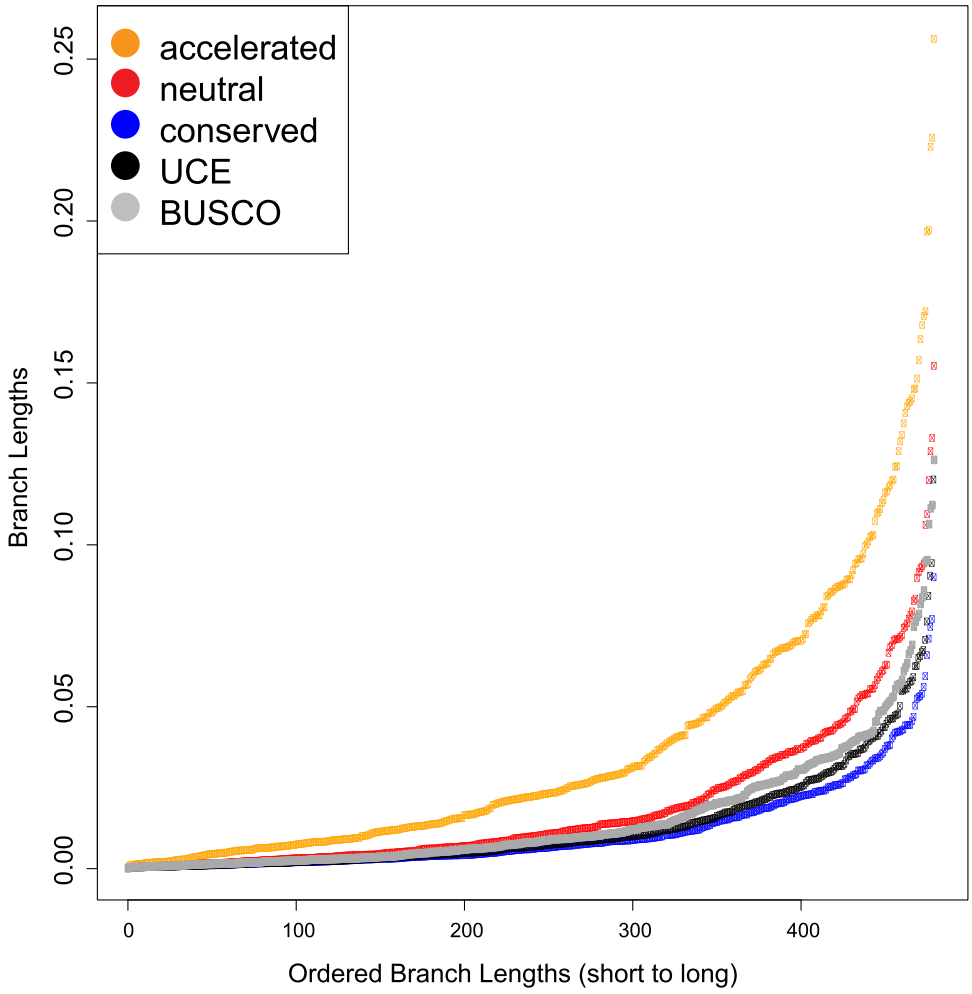

Figure 6 Distribution of branch lengths for the Zoonomia 241 taxon trees, based on datasets comprising accelerated, neutral, and conserved PhyloP scores. Branch lengths for the UCE and BUSCO trees were also plotted and are shown to most closely resemble the conserved branch lengths.

Table 1 Kuhner–felsenstein (KF) tree distance metric comparing topologies derived from accelerated, neutral, and conserved PhyloP scores, as well as from the UCE and BUSCO datasets.

KF branch	Accelerated	Neutral	Conserved	UCEs	BUSCO
Accelerated	…	…	…	…	…
Neutral	0.570	…	…	…	…
Conserved	0.837	0.274	…	…	…
UCEs	0.762	0.200	0.088	…	…
BUSCO	0.681	0.126	0.168	0.104	…

catus but were more evenly distributed across recombination quintiles in *Giraffa camelopardalis* (Fig. 3b). Our results suggest that for the majority of the clades examined here, UCEs effectively generate datasets that recover the dominant phylogenetic signal across the whole genome. However, the presence of extensive post-speciation gene flow, the genome-wide phylogenomic signal is often not reflective of the species tree (Foley et al. 2026).

Because introgressed alleles are more effectively unlinked in high-recombining regions of the genome, these regions are more likely to reflect phylogenomic signal consistent with introgression (Burbrink et al. 2025). With repeated introgression and backcrossing, the genome eventually reflects a mosaic of genomic ancestry, where the evolutionary history of the species becomes mixed with histories of introgression. The species tree becomes restricted to low-recombining regions of the genome, which are depleted of signatures of hybridization due to increased background selection. These characteristics are especially pronounced on the X and Z chromosomes, which are predicted to be enriched for loci associated with reproductive isolation (Burbrink et al. 2025). To test this, we extracted UCE and BUSCO datasets from previously published whole-genome

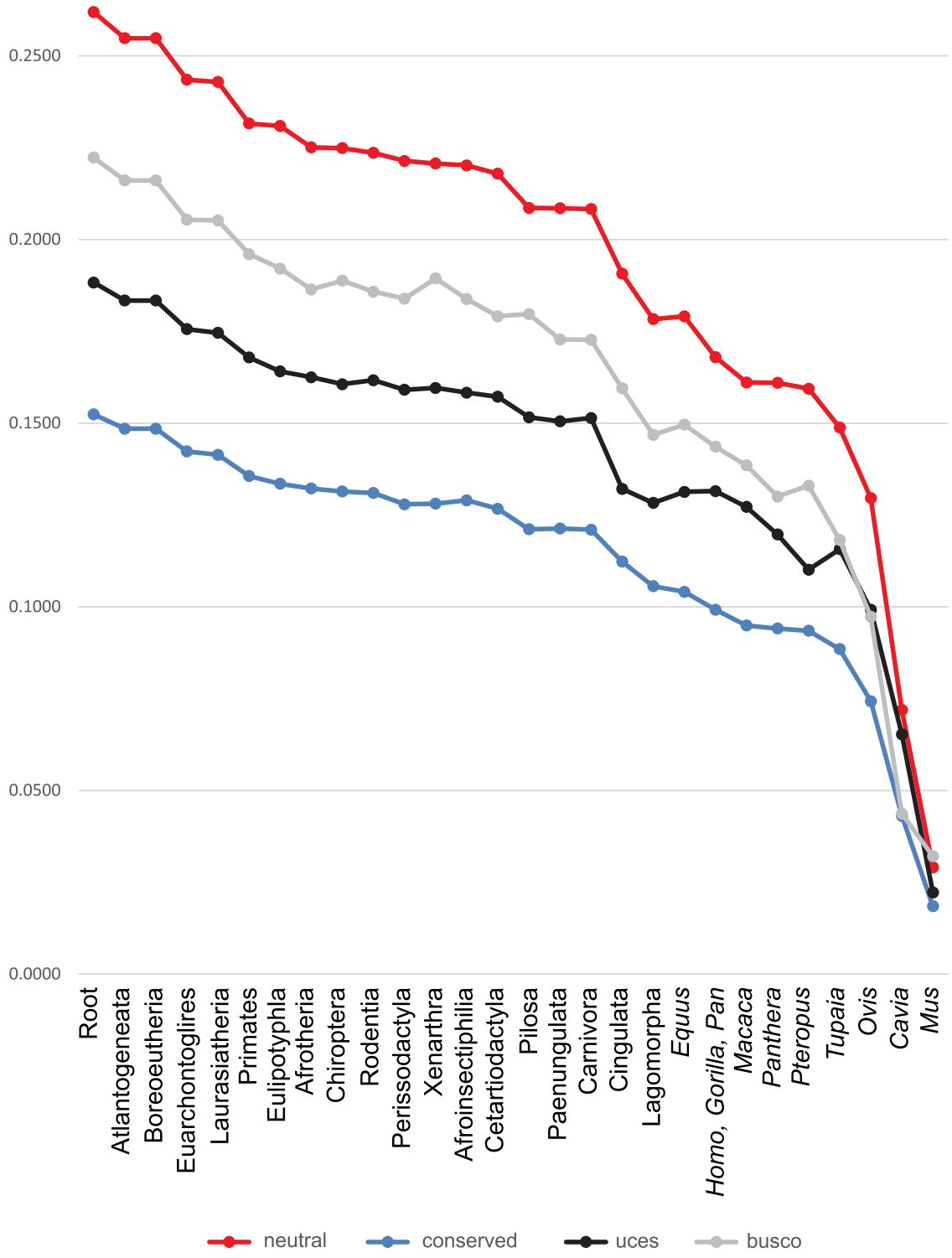

Figure 7 Distribution of selected node heights for the Zoonomia 241 mammal trees, based on datasets comprising neutral and conserved PhyloP scores, UCE, and BUSCO loci (see Figure S2 for all shared nodes).

phylogenomic datasets for mammalian species exhibiting varying degrees of genome-wide introgression. Our results show that even when loci are partitioned into X and autosomes, and even with longer UCEs, the UCE and BUSCO-derived phylogenies tend to recover the topology reflective of the most frequent genome-wide tree (Fig. 4, Figures S1 and S2). Therefore, a sampling strategy focusing solely on the X chromosome is unlikely to ensure recovery of the species tree because of the small dataset size (Table 1).

Collectively, our results caution against using standard UCE target-enrichment approaches and BUSCO loci in phylogenomic analyses of clades that have undergone extensive post-speciation gene flow. Undoubtedly, the development of UCE loci through target enrichment has been a boon for

undersampled clades where preexisting primers or whole-genome data are unavailable. Indeed, the approach remains as time- and cost-efficient as ever and remains critical for the study of museum specimens, where DNA degradation may preclude whole-genome approaches. Instead, we recommend that future phylogenetic studies employing UCEs develop and analyze UCE loci from low-recombining regions or reanalyze the data in a recombination-aware framework (Edelman et al. 2019; Li et al. 2019; Foley et al. 2024; Burbrink et al. 2025). In mammalian or avian clades, it would be most beneficial to target the conserved, low-recombining region on the X chromosome (the XLRD, as defined by Foley et al. 2026) or the Z chromosome (Stiller et al. 2024). For example, the Zoonomia project has reported a total of 4,552 zooUCEs (Christmas et al. 2023), representing an increase of 3,799 loci over the 481 initially described for vertebrates (Bejerano et al. 2004). For taxa where it may be challenging to predict introgression *a priori*, future studies should sample loci from high-, low-, and intermediate-recombining regions in sufficient numbers to assess phylogenomic discordance between these genomic partitions. Mitigating the branch-length bias observed here may be achieved by employing long-read sequencing technologies to increase the length of the flanking sequence captured when using higher-quality DNA (Iyer et al. 2024). In this way, users could increase the number of neutral sites available for analysis and exclude the most conserved portion of the UCE locus core itself, thereby reducing the number of ultra-conserved sites in the final alignment.

Conclusion

Despite their widespread adoption as efficient and cost-effective markers for generating phylogenetic datasets for poorly studied taxa, our results suggest that UCE and BUSCO loci may not perform optimally in clades with abundant post-speciation gene flow. We demonstrated that, in the presence of gene flow, UCE and BUSCO loci frequently return a topology that reflects the dominant introgression signal rather than the species topology. Both sets of loci evolve under strong evolutionary constraint, violating assumptions of many coalescent-based phylogenomic methods. Furthermore, the use of UCE and BUSCO loci introduces an acquisition bias, as evidenced by compressed branch lengths and node heights relative to those derived from neutral data. Their distribution relative to the recombination landscape varies among species. Future UCE/BUSCO-based studies focusing on clades that have evolved with post-speciation gene flow should sample loci from high- and low-recombining regions to assess whether discordance exists between these genomic extremes, which are typically indicative of the interplay between natural selection and recombination rate on the retention of admixed ancestry.

Materials and methods

The genomic context of UCE and BUSCO loci

The 5 kb UCE probe set was downloaded from GitHub (https://github.com/faircloth-lab/uce-probe-sets/uce-5k-probe-set) in FASTA format. The human genome assembly GRCh38.p14 (GCF_000001405.40 or hg38) was used to determine the genomic context of UCE loci and BUSCO genes. The hg38 reference genome was converted from FASTA to 2bit format, and UCE loci were identified in the assembly using the 5 kb probe set as a query with iterative BLAT searches (Kent 2002) with -q=dna -stepSize=5 -repMatch=2253 -minScore=20 -minIdentity=1. An existing script (https://gist.github.com/davetang/7314846) was used to convert the raw BLAT output to a BED file containing the best-scoring BLAT hit for each UCE locus. The location of BUSCO genes in the hg38 human genome assembly was determined using the Mammalia odb10 database in Compleasm (Huang and Li 2023). The resulting annotation was filtered to retain only single-copy BUSCO genes for further analysis. The resulting UCE and BUSCO coordinates were plotted on the hg38 genome assembly using karyoploteR (Gel and Serra 2017) in R to visualize the distribution of UCE loci across the genome. If UCE loci were randomly distributed throughout the genome, the number of loci per chromosome should be proportional to chromosome length. This expectation was represented by fitting a linear model to the distribution of chromosome lengths. To test our hypothesis, we then plotted the number of UCE loci recovered per chromosome and accounted for length variation by dividing the number of UCE loci by chromosome length. A linear model was fitted to this distribution, and the R^2 values of both models were compared.

Next, we investigated the distribution of UCE and BUSCO loci across the genome-wide recombination landscape. Previously published recombination maps for *Bos taurus*, *Giraffa camelopardalis*, *Ursus arctos,* and *Felis catus* (Foley et al. 2026) were ordered from low to high values and assigned to quintiles, with 1 indicating the lowest recombination quintile and 5 the highest. UCE loci were identified in the reference genomes used to generate the corresponding recombination maps, as described above: *Bos taurus* (GCF_002263795.3), *Giraffa camelopardalis* (GCA_017591445.1), *Ursus arctos* (GCF_023065955.1), and *Felis catus* (GCA_000181335.3). The coordinates of single-copy BUSCO loci were identified as before in the *Giraffa camelopardalis* and *Felis catus* reference genomes. The intersect function from bedtools (Quinlan and Hall 2010) was used to determine which recombination quintile each UCE and BUSCO locus overlapped for each species. The frequency of UCE and BUSCO loci in each recombination quintile was calculated.

Do UCE and BUSCO-based phylogenies recover the species tree in the presence of gene flow?

To investigate whether UCE datasets could recover the species tree in the presence of gene flow, we used a suite of previously published phylogenomic datasets (Table S1). These studies used a recombination-aware phylogenomic approach to show that the species tree was enriched in low-recombining parts of the genome and that trees enriched in regions of high recombination were due to introgression (Li et al. 2019; Foley et al. 2024). UCE loci were identified in each reference genome (Table S1) as previously described, except that an additional 200 bp was added on either side of the best BLAT hit to mimic UCE capture

in silico. The decision to select 200 bp on either side was based on the approximate contig length obtained via target enrichment in the following studies (Faircloth et al. 2012, 2013, 2015; Blaimer et al. 2016; Branstetter et al. 2017, 2021). Each UCE locus was filtered to remove any alignment column with >10% missingness. Recent studies have demonstrated that merging co-occurring UCEs to form longer loci resulted in more accurate gene trees (Van Dam et al. 2021). These improvements were thought to be linked to longer loci, which ought to contain more informative sites for analysis. To explore whether the use of longer loci would improve our phylogenomic analyses, we also extracted UCE loci with an additional 800 bp on either side. We conducted this test for a subset of the clades: *Martes*, *Giraffa, Ursidae*, and *Panthera*.

BUSCO loci were identified by running Compleasm with the Mammalian odb10 database against the reference genomes for the *Martes*, *Giraffa*, and *Panthera* clades (Table S1). For each clade, only coding domain sequences (cds) of single-copy BUSCO genes were retained for each analysis. Extracted alignments were concatenated per chromosome using AMAS (Borowiec 2016), and alignment columns were filtered for 10% missingness using trimAl (Capella-Gutiérrez et al. 2009). The EMBOSS function extract_align was used to extract per-cds alignments from per-chromosome alignments (Rice et al. 2000). As such, for each clade, we generated three aligned datasets: (i) including all UCE or BUSCO loci, (ii) autosomal UCE or BUSCO loci, and (iii) X-linked UCE and BUSCO loci. AMAS was used to estimate summary statistics for each dataset and clade (Borowiec 2016).

For each dataset, we used IQtree2 to generate maximum-likelihood (ML) trees under the GTR model of sequence evolution, in conjunction with the GHOST model of rate heterogeneity (Crotty et al. 2020), to account for the apparent increase in heterogeneity due to post-speciation gene flow. Nodal support was evaluated using 10,000 ultrafast bootstrap replicates (Minh et al. 2013; Hoang et al. 2018). To prepare our datasets for a coalescence-based analysis, we extracted SNPs from each dataset using SNP-sites v2.5.1 (Page et al. 2016). Using SVDQuartets, we used the SNP sites to impute quartet trees, which were then aggregated to infer a 50% majority-rule tree for each dataset using all quartets.

Locus selection and branch length distortion

Although UCEs should be highly conserved by definition, per-base estimates of constraint from the 241-species Zoonomia mammal alignment enabled us to precisely estimate constraints acting on these loci and place them in a broader genomic context. To achieve this, publicly available PhyloP scores from the 241 Zoonomia mammal alignment were downloaded from https://zoonomiaproject.org/the-data/. PhyloP scores are used to measure the evolutionary conservation of a site at a given position in the genome. Given a multiple sequence alignment, models are used to compare if substitution rates at a given site are faster or slower compared with a null model, neutral evolution due to drift (Pollard et al. 2010). PhyloP scores are negative log *P*-values arising from comparison to the null model where positive PhyloP scores indicate conservation, zero values represent neutrally evolving sites, and negative numbers indicate accelerated sites (Pollard et al. 2010; Hubisz et al. 2011). PhyloP scores corresponding to the UCE loci were collected using bedtools intersect (Quinlan and Hall 2010), and the average PhyloP score per UCE locus was calculated. The median PhyloP score per locus was then plotted against the genome-wide distribution of PhyloP scores, along with whole-genome median values for 1st, 2nd, and 3rd codon positions, neutral, accelerated, and conserved sites. Similarly, sequences corresponding to the exome of the human genome assembly hg38 were extracted from the Zoonomia dataset, and the median PhyloP value was also plotted.

Previously, it has been shown that the use of neutral, accelerated and conserved PhyloP derived genomic partitions does not improve our ability to infer the species tree in clades where species with extensive discordance due to ILS or hybridization (see Figure S1 in Foley et al. 2023). As such, we set out to determine whether selecting particular sequence classes introduced branch-length bias in phylogenomic analyses, SNP datasets representing extremes of the PhyloP scale (accelerated ≤ -4, neutral $= 0.000$, conserved ≥ 6) were subset from the human hg38-referenced Zoonomia whole-genome FASTA alignment and concatenated (Foley et al. 2023). Invariable sites were removed from the alignment to retain only SNP sites using Geneious Prime (Version 2023.1.2). UCE loci mapped to the hg38 genome assembly were extracted from the Zoonomia Alignment. Single-copy BUSCO orthologs were identified using Compleasm with the more conserved gene set corresponding to the Eukaryota odb10 database. Only coding domain sequences associated with single-copy BUSCO genes were retained for further analysis. UCE and BUSCO datasets were concatenated with AMAS and filtered for 10% missingness using trimAl. Invariant sites in both datasets were removed in Geneious. ML trees were generated using IQtree2 (Minh et al. 2020) for the neutral, accelerated, conserved, UCE, and BUSCO SNP datasets, with 10,000 ultrafast bootstrap replicates (Hoang et al. 2018) and a GTR+I+CAT model of sequence evolution for each dataset. The +ASC option was also specified as part of the model to account for the removal of invariant sites from the datasets. Branch-length differences were investigated using the KF.dist function in phangorn (Schliep 2011) in R. Unlike Robinson-Foulds differences (Robinson and Foulds 1981), which consider only topological differences between two trees, KF.dist, which implements the Kuhner-Felsenstein branch score distance, also accounts for branch-length differences (Kuhner and Felsenstein 1994). Node height information for all nodes shared among the 5 trees was extracted using Phytools (Revell 2012). Nodes corresponding to major higher- and lower-order mammalian lineages were also plotted separately.

Acknowledgments

This research was performed using computational resources housed and managed by the Texas A&M High Performance Research Computing Center. We thank Brant Faircloth and David Swofford for helpful discussions on the topics discussed herein.

Supplementary material

Supplementary material is available at *Molecular Biology and Evolution* online.

Funding

This work was supported by the National Science Foundation (DEB-2150664) by a grant awarded to W.J.M.

Conflicts of interest

None declared.

Data availability

This study was conducted using publicly available data published as part of the original papers cited within the main text and listed in Table S1. Multispecies sequence alignments used as part of this analysis can be found at https://zenodo.org/uploads/19830248. Links to data repositories associated with original papers are listed where available. Zoonomia PhyloP Scores—https://zoonomiaproject.org/the-data/; Fasta-formatted Zoonomia alignments—https://zenodo.org/records/5823345; *Myotis* bat data—https://doi.org/10.5281/zenodo.7044479. Various mammal clade data—https://doi.org/10.5281/zenodo.15131984.

References

Armstrong EE *et al.* A beary good genome: haplotype-resolved, chromosome-level assembly of the brown bear (*Ursus arctos*). *Genome Biol Evol*. 2022:14:evac125. https://doi.org/10.1093/gbe/evac125.

Bejerano G *et al.* Ultraconserved elements in the human genome. *Science*. 2004:304:1321–1325. https://doi.org/10.1126/science.1098119.

Blaimer BB, Lloyd MW, Guillory WX, Brady SG. Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. *PLoS One*. 2016:11:e0161531. https://doi.org/10.1371/journal.pone.0161531.

Bofkin L, Goldman N. Variation in evolutionary processes at different codon positions. *Mol Biol Evol*. 2007:24:513–521. https://doi.org/10.1093/molbev/msl178.

Borges R, Machado JP, Gomes C, Rocha AP, Antunes A. Measuring phylogenetic signal between categorical traits and phylogenies. *Bioinformatics*. 2019:35:1862–1869. https://doi.org/10.1093/bioinformatics/bty800.

Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*. 2016:4:e1660. https://doi.org/10.7717/peerj.1660.

Branstetter MG, Longino JT, Ward PS, Faircloth BC. Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other hymenoptera. *Methods Ecol Evol*. 2017:8:768–776. https://doi.org/10.1111/2041-210X.12742.

Branstetter MG, Müller A, Griswold TL, Orr MC, Zhu CD. Ultraconserved element phylogenomics and biogeography of the agriculturally important mason bee subgenus *Osmia* (*Osmia*). *Syst Entomol*. 2021:46:453–472. https://doi.org/10.1111/syen.12470.

Burbrink FT, DeBaun D, Foley NM, Murphy WJ. Recombination-aware phylogenomics. *Trends Ecol Evol*. 2025:40:900–912. https://doi.org/10.1016/j.tree.2025.06.011.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009:25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Card DC, Jennings WB, Edwards SV. Genome evolution and the future of phylogenomics of non-avian reptiles. *Animals (Basel).* 2023:13:471. https://doi.org/10.3390/ani13030471.

Chen Z *et al.* A genome-based phylogeny for Mollusca is concordant with fossils and morphology. *Science*. 2025:387:1001–1007. https://doi.org/10.1126/science.ads0215.

Christmas MJ *et al.* Evolutionary constraint and innovation across hundreds of placental mammals. *Science*. 2023:380:eabn3943. https://doi.org/10.1126/science.abn3943.

Crotty SM *et al.* GHOST: recovering historical signal from heterotachously evolved sequence alignments. *Syst Biol*. 2020:69:249–264. https://doi.org/10.1093/sysbio/syz051.

Degnan JH, Rosenberg NA. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol*. 2009:24:332–340. https://doi.org/10.1016/j.tree.2009.01.009.

Deng X *et al.* Bipartite structure of the inactive mouse X chromosome. *Genome Biol*. 2015:16:152. https://doi.org/10.1186/s13059-015-0728-8.

Derkarabetian S *et al.* Covering all bases: a universal metazoan UCE probe set to democratize phylogenomics. *Genome Biol Evol*. 2025:17:evaf193. https://doi.org/10.1093/gbe/evaf193.

Dias GR, Dupim EG, Vanderlinde T, Mello B, Carvalho AB. A phylogenomic study of steganinae fruit flies (Diptera: Drosophilidae): strong gene tree heterogeneity and evidence for monophyly. *BMC Evol Biol*. 2020:20:141. https://doi.org/10.1186/s12862-020-01703-7.

Ebbs ET *et al.* Phylogenomics and diversification of the schistosomatidae based on targeted sequence capture of ultraconserved elements. *Pathogens*. 2022:11:769. https://doi.org/10.3390/pathogens11070769.

Edelman NB *et al.* Genomic architecture and introgression shape a butterfly radiation. *Science*. 2019:366:594–599. https://doi.org/10.1126/science.aaw2090.

Faircloth BC *et al.* Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst Biol*. 2012:61:717–726. https://doi.org/10.1093/sysbio/sys004.

Faircloth BC, Branstetter MG, White ND, Brady SG. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among hymenoptera. *Mol Ecol*. 2015:15:489–501. https://doi.org/10.1111/1755-0998.12328.

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). *PLoS One*. 2013:8:e65923. https://doi.org/10.1371/journal.pone.0065923.

Feng C, Wang J, Liston A, Kang M. Recombination variation shapes phylogeny and introgression in wild diploid strawberries. *Mol Biol Evol*. 2023:40:msad049. https://doi.org/10.1093/molbev/msad049.

Foley NM *et al.* A genomic timescale for placental mammal evolution. *Science*. 2023:380:eabl8189. https://doi.org/10.1126/science.abl8189.

Foley NM *et al.* Karyotypic stasis and swarming influenced the evolution of viral tolerance in a species-rich bat radiation.

Cell Genom. 2024:4:100482. https://doi.org/10.1016/j.xgen.2023.100482.
Foley NM *et al.* An ancient recombination desert is a speciation supergene in placental mammals. *Nature*. 2026:649:1228–1236. https://doi.org/10.1038/s41586-025-09740-2.
Fontaine MC *et al.* Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*. 2015:347:1258524. https://doi.org/10.1126/science.1258524.
Gel B, Serra E. Karyoploter: an R/bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*. 2017:33:3088–3090. https://doi.org/10.1093/bioinformatics/btx346.
Guinat C *et al.* Bayesian phylodynamics reveals the transmission dynamics of avian influenza A(H7N9) virus at the human–live bird market interface in China. *Proc Natl Acad Sci U S A.* 2023:120:e2215610120. https://doi.org/10.1073/pnas.2215610120.
Harris AJ, Foley NM, Williams TL, Murphy WJ. Tree house explorer: a novel genome browser for phylogenomics. *Mol Biol Evol*. 2022:39:msac130. https://doi.org/10.1093/molbev/msac130.
Hennelly LM *et al.* Ancient divergence of Indian and Tibetan wolves revealed by recombination-aware phylogenomics. *Mol Ecol*. 2021:30:6687–6700. https://doi.org/10.1111/mec.16127.
Hibbins MS, Hahn MW. The timing and direction of introgression under the multispecies network coalescent. *Genetics*. 2019:211:1059–1073. https://doi.org/10.1534/genetics.118.301831.
Hibbins MS, Hahn MW. Phylogenomic approaches to detecting and characterizing introgression. *Genetics*. 2022:220:iyab173. https://doi.org/10.1093/genetics/iyab173.
Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018:35:518–522. https://doi.org/10.1093/molbev/msx281.
Huang N, Li H. Compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics*. 2023:39:btad595. https://doi.org/10.1093/bioinformatics/btad595.
Hubisz MJ, Pollard KS, Siepel A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinform*. 2011:12:41–51. https://doi.org/10.1093/bib/bbq072.
Iyer SV, Goodwin S, McCombie RW. Leveraging the power of long reads for targeted sequencing. *Genome Res.* 2024:34:1701–1718. https://doi.org/10.1101/gr.279168.124.
Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res*. 2002:12:656–664. https://doi.org/10.1101/gr.229202.
Kuhner MK, Felsenstein J. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol Biol Evol*. 1994:11:459–468. https://doi.org/10.1093/oxfordjournals.molbev.a040126.
Li G, Figueiró HV, Eizirik E, Murphy WJ. Recombination-aware phylogenomics reveals the structured genomic landscape of hybridizing cat species. *Mol Biol Evol*. 2019:36:2111–2126. https://doi.org/10.1093/molbev/msz139.
Li Y *et al.* A genome-scale phylogeny of the kingdom fungi. *Curr Biol*. 2021:31:1653–1665.e5. https://doi.org/10.1016/j.cub.2021.01.074.
Lin G *et al.* Evolutionary rates of bumblebee genomes are faster at lower elevations. *Mol Biol Evol*. 2019:36:1215–1219. https://doi.org/10.1093/molbev/msz057.
Liu H *et al.* A taxon-rich and genome-scale phylogeny of Opisthokonta. *PLoS Biol*. 2024:22:e3002794. https://doi.org/10.1371/journal.pbio.3002794.
McCormack JE *et al.* Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res*. 2012:22:746–754. https://doi.org/10.1101/gr.125864.111.
McCormack JE *et al.* A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS One*. 2013:8:e54848. https://doi.org/10.1371/journal.pone.0054848.
Minh BQ *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020:37:1530–1534. https://doi.org/10.1093/molbev/msaa015.
Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol*. 2013:30:1188–1195. https://doi.org/10.1093/molbev/mst024.
Mooers A, Gascuel O, Stadler T, Li H, Steel M. Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. *Syst Biol*. 2011:61:195–203. https://doi.org/10.1093/sysbio/syr090.
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.
Nelson TC *et al.* Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (*Mimulus* section *Erythranthe*). *PLoS Genet*. 2021:17:e1009095. https://doi.org/10.1371/journal.pgen.1009095.
Ottenburghs J *et al.* Highly differentiated loci resolve phylogenetic relationships in the bean goose complex. *BMC Ecol Evol*. 2023:23:2. https://doi.org/10.1186/s12862-023-02103-3.
Owens GL, Huang K, Todesco M, Rieseberg LH. Re-evaluating homoploid reticulate evolution in *Helianthus* sunflowers. *Mol Biol Evol*. 2023:40:msad013. https://doi.org/10.1093/molbev/msad013.
Page AJ *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom*. 2016:2:e000056. https://doi.org/10.1099/mgen.0.000056.
Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. 2010:20:110–121. https://doi.org/10.1101/gr.097857.109.
Quattrini AM *et al.* Universal target-enrichment baits for anthozoan (cnidaria) phylogenomics: new approaches to long-standing problems. *Mol Ecol Resour*. 2018:18:281–295. https://doi.org/10.1111/1755-0998.12736.
Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010:26:841–842. https://doi.org/10.1093/bioinformatics/btq033.
Revell LJ. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 2012:3:217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x.
Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet*. 2000:16:276–277. https://doi.org/10.1016/S0168-9525(00)02024-2.

Robinson DF, Foulds LR. Comparison of phylogenetic trees. *Math Biosci*. 1981:53:131–147. https://doi.org/10.1016/0025-5564(81)90043-2.

Sahbou A-E, Iraqi D, Mentag R, Khayi S. BuscoPhylo: a webserver for Busco-based phylogenomic analysis for non-specialists. *Sci Rep*. 2022:12:17352. https://doi.org/10.1038/s41598-022-22461-0.

Schliep KP. Phangorn: phylogenetic analysis in R. *Bioinformatics*. 2011:27:592–593. https://doi.org/10.1093/bioinformatics/btq706.

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

Stiller J *et al.* Complexity of avian evolution revealed by family-level genomes. *Nature*. 2024:629:851–860. https://doi.org/10.1038/s41586-024-07323-1.

Streicher JW, Loader SP, Varela-Jaramillo A, Montoya P, de Sá RO. Analysis of ultraconserved elements supports African origins of narrow-mouthed frogs. *Mol Phylogenet Evol*. 2020:146:106771. https://doi.org/10.1016/j.ympev.2020.106771.

Stubbs RL *et al.* Whole-genome analyses disentangle reticulate evolution of primroses in a biodiversity hotspot. *New Phytol*. 2023:237:656–671. https://doi.org/10.1111/nph.18525.

Sullivan PF *et al.* Leveraging base-pair mammalian constraint to understand genetic variation and human disease. *Science*. 2023:380:eabn2937. https://doi.org/10.1126/science.abn2937.

Thompson MJ, Jiggins CD. Supergenes and their role in evolution. *Heredity (Edinb).* 2014:113:1–8. https://doi.org/10.1038/hdy.2014.20.

Van Dam MH, Henderson JB, Esposito L, Trautwein M. Genomic characterization and curation of UCEs improves species tree reconstruction. *Syst Biol*. 2021:70:307–321. https://doi.org/10.1093/sysbio/syaa063.

Wooldridge TB *et al.* Chromosome-scale genomes show rapid diversification and ancient gene flow among bear species. *Genome Biol Evol*. 2025:17:evaf188. https://doi.org/10.1093/gbe/evaf188.

Yang Z. Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol Evol*. 1996:11:367–372. https://doi.org/10.1016/0169-5347(96)10041-0.

Zhang J, Kapli P, Pavlidis P, Stamatakis A. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*. 2013:29:2869–2876. https://doi.org/10.1093/bioinformatics/btt499.

Zhang J, Lin L, Mu Y, Brelsford A, Purcell J. A comparison of phylogenomic inference pipelines for low-coverage whole-genome sequencing in *Formica* ants. *Syst Entomol*. 2025:50:611–629. https://doi.org/10.1111/syen.12670.