# Supplemental Results

## On UCEs as loci for phylogenetic inference

Ultra-conserved elements (UCEs) have proven increasingly popular and useful as loci for phylogenomic analyses across a range of taxonomic levels and time scales (Crawford et al. 2012; McCormack et al. 2012; Faircloth et al. 2013; McCormack et al. 2013; Jarvis et al. 2014; Blaimer et al. 2015; Meiklejohn et al. 2016; Alexander et al. 2017; Burress et al. 2018; Bossert et al. 2019; Oliveros et al. 2019; Quattrini et al. 2020; Alda et al. 2021) (McClean et al., 2022; Smith et al., 2013). UCEs are typically collected using bait capture and targeted enrichment approaches, harvested from whole genomes as we have done, or a combination of the two. This is driven by both their ready application to non-model organisms and museum specimens (Bi et al. 2013; Faircloth 2017; Zhang et al. 2019), their possession of desirable characters for use in phylogenetic reconstruction (Faircloth et al. 2012; McCormack et al. 2012), and their subsequent ability to provide resolutions to previously intractable nodes (Faircloth et al. 2013; Blaimer et al. 2015; Alda et al. 2021). However, both the function of UCEs and the influence of selection upon them remain poorly understood (Bejerano et al. 2004; Katzman et al. 2007). Rates of molecular evolution may also vary within UCE loci (Tagliacollo and Lanfear 2018) and phylogenetic reconstructions using UCEs (and other phylogenomic markers) can be confounded by gene tree discordance driven by processes such as incomplete lineage sorting and introgression (Jeffroy et al. 2006; Degnan and Rosenberg 2009; Meiklejohn et al. 2016; Chan et al. 2020; Alda et al. 2021).

While UCEs are generally treated as non-coding and independent, they may not be evenly distributed throughout the genome (McCole et al. 2018) and may overlap the same genes (Van Dam et al. 2021). Merging UCEs that overlap genes can increase topological support but has limited impact on the topology recovered (Van Dam et al. 2021). When available, filtering by phylogenetic signal and noise (Gilbert et al. 2018) can also lead to better supported estimates of tree topology, as can allele phasing of SNPs derived from UCE loci (Andermann et al. 2019; McLean et al. 2022). However, as we notice in our reconstructions and as has been observed elsewhere with multiple phylogenomic markers (Reddy et al. 2017; Chan et al. 2020; Minh et al. 2020a; Vanderpool et al. 2020), bootstraps are a poor indicator of support for the correct topology when using phylogenomic data. We estimated high approximated bootstrap values across all nodes in our species tree (Fig. 1A), even for nodes which are deeply suspect (contrast Fig. 1 with Fig. S1) and show elevated gene tree discordance (Figures XX and S2).

UCEs have recently seen increased use in dating analyses and best practices continue to be developed (Blaimer et al. 2015; Branstetter et al. 2017; Bossert et al. 2019; Oliveros et al. 2019; Quattrini et al. 2020). Critically, the scale of UCE datasets can make popular methods of divergence time estimation computationally intractable, while gene tree discordance and substitution rate heterogeneity complicate the selection of appropriate models (Van Dam et al. 2017; Tagliacollo and Lanfear 2018). One strategy to overcome these issues is the selection of a set of loci with model-appropriate properties including clock-like behaviour and low gene tree discordance, as described by the SortaDate software package (Smith et al. 2018). Such gene filtering approaches have been suggested as best practice (Walker et al. 2019) and SortaDate specifically has been used in a range of phylogenomic analyses across multiple taxonomic groups (Lind et al. 2019; Del Cortona et al. 2020; Quattrini et al. 2020; Shee et al. 2020; Koenen et al. 2021), though some researchers have found gene filtering has a limited impact on estimated divergence times but increases the associated variance (Oliveros et al. 2019; McGowen et al. 2020).

Least-squares dating purports vastly reduced computation time with similar accuracy (To et al. 2016) to comparable dating methods implemented in software such as BEAST (Drummond et al. 2006) and r8s (Sanderson 2003). It appears to be robust to topological error (To et al. 2016) and substitution rate heterozygosity between lineages (To et al. 2016; Tong et al. 2018). While developed for dating of rapidly evolving viruses (To et al. 2016), it has seen use in eukaryotes (Yue et al. 2017; Anijalg et al. 2018; Brüniche–Olsen et al. 2018; Tong et al. 2018; Thomas et al. 2019). Given the known rapid evolution of substitution rate in rodents (Douzery et al., 2003), our concerns with tree topology, and its tractability, it is likely appropriate for our dataset of filtered UCE loci.

## Eliminating possible sources of error in reconstruction of the rodent phylogeny

Given the unexpected placement of the two *Apodemus* species and *R. soricoides* recovered in the inferred species level phylogenies, we attempted to systematically account for sources of error in both the data itself and the methodology employed. We first note that the two *Apodemus* species are represented by chromosome level assemblies obtained by different research groups, while for *R. soricoides* we present a novel chromosome level assembly (see Table 1). Their unexpected taxonomic placement would therefore appear not to be an artifact of assembly quality.

The mislabeling of taxa at any step of data processing could lead to mislabeled species trees, leading us to incorrectly conclude the placement of some species. To check for mislabeling, we took a random sample of 100 UCE loci that we extracted from every species and used BLAST (Altschul et al. 1990) to align them to the original whole genomes. In the case of introduced mislabeling, we would expect UCE loci for one species to align best to a different species. However, every BLASTed locus aligned to the expected species. Sequences are also unlikely to be paralogs, as putative paralogs are removed by lastz (Harris 2007) as part of the standard PHYLUCE pipeline (Faircloth 2016).

As noted in Table 1, not all UCE loci were successfully extracted for every taxa ranging between 2223 for *Apodemus speciosus* to the full set of 2645 for *Mus musculus*. While ML reconstruction methods like those employed in IQ-TREE 2 (Minh et al. 2020b) and the gene-tree coalescent methods in ASTRAL-III (Zhang et al. 2018) are generally robust to missing data even with moderate incomplete lineage sorting (Chernomor et al. 2016; Xi et al. 2016), this is dependent on missing data being randomly distributed and sufficiently many genes being sampled (Sanderson et al. 2010; Xi et al. 2016). To assess the impact of missing data, we reran the same concatenated species tree reconstruction in IQ-TREE 2 (Minh et al. 2020b) on a set of 1600 UCE loci that are present in all 18 of our taxa of interest. The topology of the resulting species tree was identical to the one inferred when using all loci, meaning that our results are not biased by missing data.

In order to investigate other sources of error in sequence data that may have been transmitted into the phylogenetic analyses, we used TreeShrink (Mirarab et al. 2014) to identify abnormally long branch lengths among the gene trees and remove erroneous taxa (default parameters). Each taxon was cut from between 5 and 126 gene trees. The filtered gene trees were used to construct a consensus phylogeny with ASTRAL-III (Zhang et al. 2018) as in the Methods and returned the same topology. Further, we used the R (R Core Team 2021) package Rogue (Smith, 2021) to identify rogue taxa – those with a variable position across trees because of conflict among their data that can confound consensus tree construction (Wilkinson 1994; Kearney 2002; Wilkinson 2003) – in our dataset. If the removal of a taxon from a set of inferred trees – such as those obtained from bootstrapping or from sampling a Bayesian posterior probability – leads to increased information content in the resulting consensus tree, it is marked as rogue. When applied to a set of 100 bootstrap trees generated in IQ-TREE 2 (Minh et al. 2020b), Rogue did not detect any rogue taxa. This comports with the high estimated bootstrap support (Fig. 1A) for our tree and supports the notion that bootstrapping methods will fail to evidence the true topology when it is obscured by gene tree discordance (see Discussion).

While Rogue (Smith, 2021) is intended for summarizing sets of trees that have previously been inferred, we also applied it to our estimated UCE trees. In principle, this is equivalent to filtering the data prior to inference by removing taxa that behave in a “rogue-like” fashion. We identified three rogue taxa from UCE trees: *Rhynchomys soricoides*, *Rhabdomys dilectus*, and *Grammomys dolichurus*. We then realigned and filtered our UCE loci without the three putative rogue taxa and repeated our species tree estimations as described in the Methods (Fig. S7). Because *R. soricoides* is a key taxon in our dataset for comparing our inferred groupings with previously inferred trees, its removal makes it difficult to discern the effects of these rogue taxa. These three taxa do not collectively represent the most discordant nodes in our species trees (Figures XX and S1), and whether their identification as rogue taxa indicates a problem with the underlying data or is informative about their biology is unclear.

Next, using IQ-TREE 2 (Minh et al. 2020b), we performed a constrained tree search to force the topology to match the relationships recovered by previous work (Schenk et al. 2013) (Fig. S1). However, the unconstrained tree (Figure 1A) significantly outperformed the constrained tree (Table S1), suggesting that the UCE loci in our dataset genuinely better support this alternative topology.

We also constructed phylogenetic networks (Fig. S6) with the SNaQ estimation method (Solis-Lemus and Ane 2016) within PhyloNetworks (Solis-Lemus et al. 2017; Bastide et al. 2018). Phylogenetic networks allow reticulation points to be formed between two branches, which may represent common important biological events that are known to confound phylogenetic inference, including incomplete lineage sorting and introgression (Maddison 1997; Barton 2001; Mallet 2007; Kubatko 2009; Hibbins and Hahn 2022). The initial network topology with zero hybrid edges matched that of the coalescent species tree (Figure 1B). Adding one hybrid edge produces an overall topology that is instead consistent with the concatenated tree, with the hybrid edge suggesting introgression from *Arvicanthis niloticus* into the other Arvicanthini. Adding a second hybrid edge suggests introgression from the Praomyini into *Mus musculus* and joins *Rhynchomys soricoides* into a clade with *Grammomys dolichurus* and *Rhabdomys dilectus*.

# Other text I wasn’t sure where to put

As our analyses are based on the available whole genomes of murid rodents, we are inevitably limited in our taxon sampling. Coalescent methods such as ASTRAL-III (Zhang et al., 2018) appear to be resilient to analyses on a small number of taxa (Song et al., 2012; Xi et al., 2015), though adding taxa to a dataset is well understood to improve phylogenetic resolution in general (see Bravo et al., 2019 and citations therein for a discussion). In contrast, adding more genomic data rather than taxa can, counterintuitively, lead to increased support for erroneous topologies (Kumar et al., 2012; Roycroft et al., 2019). It remains to be seen therefore the degree to which the topologies we recover are the result of taxon sampling versus underlying properties of the data, although the extent of gene tree discordance we observe remains striking. Ultimately, choice of data and the models used may matter more than taxon sampling (Reddy et al., 2017), though knowing which data to choose a priori remains an unsolved problem.

Phylogenomic datasets represent a wealth of opportunity to better understand taxonomic relationships, but their size and complexity can make them challenging to use and prone to introducing error (Young & Gillung, 2020; Zhang et al, 2019).

Here, we have leveraged the resources of the model organisms the house mouse (*Mus musculus*) and brown rat (*Rattus norvegicus*) along with new genomes from 8 closely related species and eight previously sequenced rodent genomes to understand the systematics of murine rodents and causes and consequences of phylogenetic discordance along the murine genome. Our genomes begin to fill the gap in sampling of murine rodents which, despite their outstanding diversity and speciousness, have relatively few whole genomes sequenced and help to place these important model systems in an evolutionary context and provide us with the resources to study the landscape of phylogenetic discordance along the chromosome.

# Supplemental Figure legends

**Figure S1**: A reference cladogram illustrating the predicted relationships between taxa for taxa included in this study based on previous results in Steppan and Schenk (2017). Lineage names follow the classifications used by Lecompte (2008). Note that the genus *Rhynchomys* is represented by Rhynchomys isarogensis in Steppan and Schenk (2017) while out study uses *Rhynchomys soricoides*. Similarly, Apodemus sylvaticus is absent from the Steppan and Schenk (2017) topology. We manually place it as sister to Apodemus speciosus in this cladogram for comparison.

**Figure S2**:Gene-concordance factors (gCF) and site-concordance factors (sCF) for each branch in the concatenated species tree. Note that the lowest possible value for sCF is approximately 30, while gCF can be as low as 0.

**Figure S3**: Time-calibrated version of the concatenated species tree estimated without putatively rogue taxa (*Rhynchomys soricoides*, *Rhabdomys dilectus*, and *Grammomys dolichurus*). Node age confidence intervals are shown as blue bars. Axis on the bottom represents time in millions of years before present. Calibrations are described in Tables 2 and S2. Node C was used as a fixed calibration,

**Figure S4**: The number of variable and informative sites from 165,409 10kb windows from alignments of seven taxa in the mouse reference (mm10) coordinate system. A variable site is defined as any site with more than one allele. An informative site is defined as any site with at least two alleles present in at least two taxa. Average number of variable sites: 1187.8. Average number of informative sites: 401.0.

**Figure S5**: Distributions of recombination rate across 19 mouse autosomes and the X chromosome estimated in 5Mb windows.

**Figure S6**: Phylogenetic networks for 18 rodent species. Hmax indicates the maximum permitted number of hybrid edges. Hmax of 3 or higher leads to reticulations that conflict with rooting the tree to *Phodopus sungorus* and are therefore excluded. When hmax=0 (-Ploglik = 20422.547) we recover the same topology as the coalescent species tree (Fig. 1B). For hmax=1 (-Ploglik = 20422.547), the overall topology is instead consistent with the concatenated tree but introduces a hybrid edge that suggests a low level of introgression from *Arvicanthis niloticus* into the other Arvicanthini. While the topology is still in conflict with the generally accepted relationships within Muridae ((Steppan and Schenk 2017); Fig. S1), the signal of the relationship between the Arvicanthini is recovered by the hybrid edge. For hmax=2 (-Ploglik = 12726.4) an additional hybrid edge suggesting introgression from the Praomyini into *Mus musculus* is recovered.

**Figure S7**: Species trees generated after the removal of the putative rogue taxa (*Rhynchomys soricoides*, *Rhabdomys dilectus*, and *Grammomys dolichurus*) from sequence data. A) Maximum likelihood tree from concatenation of all UCE alignments. Branch lengths are in units substitutions per site and node labels indicate approximated bootstrap and SH-aLRT values. B) Quartet-based species tree inferred from individual UCE trees. Internal branch lengths are in coalescent units such that shorter branches indicate greater discordance. Branch labels indicate quadripartition support. Final normalized quartet score is 0.92, i.e. 92% of quartet trees induced by the gene trees are represented in the species tree.

XX

# Supplemental Tables

### Table S1: Results of tree topology tests from IQ-TREE2 (Minh et al., 2020) comparing the original species tree shown in Figure 2A (Free) with a second tree of constrained topology (Constrained) such that the relationships between the Otomyini, Arvicanthi, Apodemini, Murini, and Praomyini are forced to match those shown in Supplementary Figure 1. Both trees were generated using the same set of filtered alignments, but best fit models were recalculated using ModelFinder (Kalyaanamoorthy et al., 2017). Tree: Tested topology. logL: Maximum likelihood. deltaL: Difference in logL from maximal logL. bp-RELL: bootstrap proportion using RELL method (Kishino et al. 1990). p-KH: p-value of one sided Kishino-Hasegawa test (1989). p-SH: p-value of Shimodaira-Hasegawa test (2000). c-ELW: Expected Likelihood Weight (Strimmer & Rambaut 2002). p-AU: p-value of approximately unbiased (AU) test (Shimodaira, 2002). Plus signs (+) indicate a topology was not rejected, while minus signs (-) indicate the topology was significantly excluded. Despite the constrained topology matching the expected relationships, there is clearly better support for the alternative topologies in Figure 2 when using UCE loci.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Tree** | **logL** | **deltaL** | **bp-RELL** | **p-KH** | **p-SH** | **c-ELW** | **p-AU** |
| Free | -36730176.13 | 0 | 1 + | 1 + | 1 + | 1 + | 1 + |
| Constrained | -37160286.86 | 4.3011x105 | 0 - | 0 - | 0 - | 0 - | 2.64x10-77 - |

# Table S2: The estimated age of each node in Figure 4. Nodes with an asterisk (\*) indicate those used in calibration, as in Table 2. Node C was used as the fixed calibration (see Methods: Divergence time estimation). Minimum and maximum ages are the lowest and highest values respectively at each node across the three estimated time trees.

|  |  |  |
| --- | --- | --- |
| Node | Estimated Age (mya) | Min/Max Age (mya) |
| A | 24.27 | 19.59/29.64 |
| B | 22.33 | 18.01/28 |
| C\* | 13.08 | 12.1/14.05 |
| D | 11.9 | 10.39/14.05 |
| E | 11.45 | 9.73/13.67 |
| F\* | 5.96 | 5.3/7.2 |
| G | 8.38 | 7.08/10.39 |
| H | 8.28 | 7/10.3 |
| I | 5.64 | 4.14/7.77 |
| J | 7.39 | 6.36/9.07 |
| K | 4.88 | 3.86/6.35 |
| L | 4.05 | 3.06/5.31 |
| M\* | 5.59 | 5.3/7.2 |
| N | 3.55 | 2.68/5.1 |
| O | 1.58 | 1.09/2.32 |
| P | 8.17 | 6.06/10.84 |
| Q\* | 2.4 | 2.4/3.59 |

# Jon: Need clarification

Chart

Description automatically generated

**Figure 3:** Discordance at nodes on the ASTRAL species tree established using the PhyParts package (Smith et al., 2015). The number of gene trees that support the shown species tree at each node is given above each branch and is represented by the blue portion of the pie chart. The number of gene trees that show a supported conflict with the gene tree is below the branch and is represented in pie charts by green (the most common conflicting partition) and red (all other supported conflicts). The grey section of the pie charts show conflicting gene trees with no support.

Diagram, schematic

Description automatically generated

**Supplementary Figure 2:** Maximum likelihood species trees generated in IQ-TREE2 (Minh et al., 2020) from concatenated UCE loci as described in methods. Here, individual taxa with unexpected placement have been removed from the analysis: *Apodemus speciosus* (2A), *Apodemus sylvaticus* (2B), *Rhynchomys soricoides* (2C), and all three of these taxa (2D). Compared to Figure 2 in the manuscript, which was generated on a full set of taxa we see that 2A recovered the same overall topology. In 2B, *Apodemus speciosus* still diverges after *Otomys typus* and *Arvicanthis niloticus*, but *Rhynchomys soricoides* now diverges after *Grammomys dolichurus* and *Rhabdomys dilectus* rather than before. Removing *Rhynchomys soricoides*, as in 2C, causes a dramatic shift in topology, splitting the *Mus* group. In 2D we recover a topology more aligned with expectation (Supplementary Figure 1), though the Arvincanthini remain paraphyletic.

Alda F, Ludt WB, Elias DJ, McMahan CD, Chakrabarty P. 2021. Comparing ultraconserved elements and exons for phylogenomic analyses of middle american cichlids: When data agree to disagree. *Genome Biol Evol*. 13.

Alexander AM, Su YC, Oliveros CH, Olson KV, Travers SL, Brown RM. 2017. Genomic data reveals potential for hybridization, introgression, and incomplete lineage sorting to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. *Evolution*. 71:475-488.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403-410.

Andermann T, Fernandes AM, Olsson U, Topel M, Pfeil B, Oxelman B, Aleixo A, Faircloth BC, Antonelli A. 2019. Allele phasing greatly improves the phylogenetic utility of ultraconserved elements. *Syst Biol*. 68:32-46.

Anijalg P, Ho SYW, Davison J, Keis M, Tammeleht E, Bobowik K, Tumanov IL, Saveljev AP, Lyapunova EA, Vorobiev AA, et al. 2018. Large-scale migrations of brown bears in eurasia and to north america during the late pleistocene. *Journal of Biogeography*. 45:394-405.

Barton NH. 2001. The role of hybridization in evolution. *Mol Ecol*. 10:551-568.

Bastide P, Solis-Lemus C, Kriebel R, William Sparks K, Ane C. 2018. Phylogenetic comparative methods on phylogenetic networks with reticulations. *Syst Biol*. 67:800-820.

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

Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C. 2013. Unlocking the vault: Next-generation museum population genomics. *Mol Ecol*. 22:6018-6032.

Blaimer BB, Brady SG, Schultz TR, Lloyd MW, Fisher BL, Ward PS. 2015. Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: A case study of formicine ants. *BMC Evol Biol*. 15:271.

Bossert S, Murray EA, Almeida EAB, Brady SG, Blaimer BB, Danforth BN. 2019. Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of apidae. *Mol Phylogenet Evol*. 130:121-131.

Branstetter MG, Danforth BN, Pitts JP, Faircloth BC, Ward PS, Buffington ML, Gates MW, Kula RR, Brady SG. 2017. Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. *Current Biology*. 27:1019-1025.

Brüniche–Olsen A, Jones ME, Burridge CP, Murchison EP, Holland BR, Austin JJ. 2018. Ancient DNA tracks the mainland extinction and island survival of the tasmanian devil. *Journal of Biogeography*. 45:963-976.

Burress ED, Alda F, Duarte A, Loureiro M, Armbruster JW, Chakrabarty P. 2018. Phylogenomics of pike cichlids (cichlidae: Crenicichla): The rapid ecological speciation of an incipient species flock. *J Evol Biol*. 31:14-30.

Chan KO, Hutter CR, Wood PL, Jr., Grismer LL, Brown RM. 2020. Target-capture phylogenomics provide insights on gene and species tree discordances in old world treefrogs (anura: Rhacophoridae). *Proc Biol Sci*. 287:20202102.

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol*. 65:997-1008.

Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, Glenn TC. 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biol Lett*. 8:783-786.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol*. 24:332-340.

Del Cortona A, Jackson CJ, Bucchini F, Van Bel M, D'Hondt S, Skaloud P, Delwiche CF, Knoll AH, Raven JA, Verbruggen H, et al. 2020. Neoproterozoic origin and multiple transitions to macroscopic growth in green seaweeds. *Proc Natl Acad Sci U S A*. 117:2551-2559.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 4:e88.

Faircloth BC. 2017. Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution*. 8:1103-1112.

Faircloth BC. 2016. Phyluce is a software package for the analysis of conserved genomic loci. *Bioinformatics*. 32:786-788.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst Biol*. 61:717-726.

Faircloth BC, Sorenson L, Santini F, Alfaro ME. 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (uces). *PLoS One*. 8:e65923.

Gilbert PS, Wu J, Simon MW, Sinsheimer JS, Alfaro ME. 2018. Filtering nucleotide sites by phylogenetic signal to noise ratio increases confidence in the neoaves phylogeny generated from ultraconserved elements. *Mol Phylogenet Evol*. 126:116-128.

Harris RS. 2007. Improved pairwise alignment of genomic DNA. [The Pennsylvania State University.

Hibbins MS, Hahn MW. 2022. Phylogenomic approaches to detecting and characterizing introgression. *Genetics*. 220.

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

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: The beginning of incongruence? *Trends Genet*. 22:225-231.

Katzman S, Kern AD, Bejerano G, Fewell G, Fulton L, Wilson RK, Salama SR, Haussler D. 2007. Human genome ultraconserved elements are ultraselected. *Science*. 317:915.

Kearney M. 2002. Fragmentary taxa, missing data, and ambiguity: Mistaken assumptions and conclusions. *Syst Biol*. 51:369-381.

Koenen EJM, Ojeda DI, Bakker FT, Wieringa JJ, Kidner C, Hardy OJ, Pennington RT, Herendeen PS, Bruneau A, Hughes CE. 2021. The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the cretaceous-paleogene (k-pg) mass extinction event. *Syst Biol*. 70:508-526.

Kubatko LS. 2009. Identifying hybridization events in the presence of coalescence via model selection. *Syst Biol*. 58:478-488.

Lecompte E, Aplin K, Denys C, Catzeflis F, Chades M, Chevret P. 2008. Phylogeny and biogeography of african murinae based on mitochondrial and nuclear gene sequences, with a new tribal classification of the subfamily. *BMC Evol Biol*. 8:199.

Lind AL, Lai YYY, Mostovoy Y, Holloway AK, Iannucci A, Mak ACY, Fondi M, Orlandini V, Eckalbar WL, Milan M, et al. 2019. Genome of the komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards. *Nat Ecol Evol*. 3:1241-1252.

Maddison WP. 1997. Gene trees in species trees. *Systematic Biology*. 46:523-536.

Mallet J. 2007. Hybrid speciation. *Nature*. 446:279-283.

McCole RB, Erceg J, Saylor W, Wu CT. 2018. Ultraconserved elements occupy specific arenas of three-dimensional mammalian genome organization. *Cell Rep*. 24:479-488.

McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res*. 22:746-754.

McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, Brumfield RT. 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS One*. 8:e54848.

McGowen MR, Tsagkogeorga G, Alvarez-Carretero S, Dos Reis M, Struebig M, Deaville R, Jepson PD, Jarman S, Polanowski A, Morin PA, et al. 2020. Phylogenomic resolution of the cetacean tree of life using target sequence capture. *Syst Biol*. 69:479-501.

McLean BS, Bell KC, Cook JA. 2022. Snp-based phylogenomic inference in holarctic ground squirrels (urocitellus). *Mol Phylogenet Evol*. 169:107396.

Meiklejohn KA, Faircloth BC, Glenn TC, Kimball RT, Braun EL. 2016. Analysis of a rapid evolutionary radiation using ultraconserved elements: Evidence for a bias in some multispecies coalescent methods. *Syst Biol*. 65:612-627.

Minh BQ, Hahn MW, Lanfear R. 2020a. New methods to calculate concordance factors for phylogenomic datasets. *Mol Biol Evol*. 37:2727-2733.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020b. Iq-tree 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 37:1530-1534.

Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. Astral: Genome-scale coalescent-based species tree estimation. *Bioinformatics*. 30:i541-548.

Oliveros CH, Field DJ, Ksepka DT, Barker FK, Aleixo A, Andersen MJ, Alstrom P, Benz BW, Braun EL, Braun MJ, et al. 2019. Earth history and the passerine superradiation. *Proc Natl Acad Sci U S A*. 116:7916-7925.

Quattrini AM, Rodriguez E, Faircloth BC, Cowman PF, Brugler MR, Farfan GA, Hellberg ME, Kitahara MV, Morrison CL, Paz-Garcia DA, et al. 2020. Palaeoclimate ocean conditions shaped the evolution of corals and their skeletons through deep time. *Nat Ecol Evol*. 4:1531-1538.

R Core Team. 2021. R: A language and environment for statistical computing. Vienna, Austria.

Reddy S, Kimball RT, Pandey A, Hosner PA, Braun MJ, Hackett SJ, Han KL, Harshman J, Huddleston CJ, Kingston S, et al. 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Syst Biol*. 66:857-879.

Sanderson MJ. 2003. R8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*. 19:301-302.

Sanderson MJ, McMahon MM, Steel M. 2010. Phylogenomics with incomplete taxon coverage: The limits to inference. *BMC Evol Biol*. 10:155.

Schenk JJ, Rowe KC, Steppan SJ. 2013. Ecological opportunity and incumbency in the diversification of repeated continental colonizations by muroid rodents. *Syst Biol*. 62:837-864.

Shee ZQ, Frodin DG, Camara-Leret R, Pokorny L. 2020. Reconstructing the complex evolutionary history of the papuasian schefflera radiation through herbariomics. *Front Plant Sci*. 11:258.

Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *PLoS One*. 13:e0197433.

Solis-Lemus C, Ane C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genet*. 12:e1005896.

Solis-Lemus C, Bastide P, Ane C. 2017. Phylonetworks: A package for phylogenetic networks. *Mol Biol Evol*. 34:3292-3298.

Steppan SJ, Schenk JJ. 2017. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. *PLoS One*. 12:e0183070.

Tagliacollo VA, Lanfear R. 2018. Estimating improved partitioning schemes for ultraconserved elements. *Mol Biol Evol*. 35:1798-1811.

Thomas JE, Carvalho GR, Haile J, Rawlence NJ, Martin MD, Ho SY, Sigfusson A, Josefsson VA, Frederiksen M, Linnebjerg JF, et al. 2019. Demographic reconstruction from ancient DNA supports rapid extinction of the great auk. *Elife*. 8.

To TH, Jung M, Lycett S, Gascuel O. 2016. Fast dating using least-squares criteria and algorithms. *Syst Biol*. 65:82-97.

Tong KJ, Duchene DA, Duchene S, Geoghegan JL, Ho SYW. 2018. A comparison of methods for estimating substitution rates from ancient DNA sequence data. *BMC Evol Biol*. 18:70.

Van Dam MH, Henderson JB, Esposito L, Trautwein M. 2021. Genomic characterization and curation of uces improves species tree reconstruction. *Syst Biol*. 70:307-321.

Van Dam MH, Lam AW, Sagata K, Gewa B, Laufa R, Balke M, Faircloth BC, Riedel A. 2017. Ultraconserved elements (uces) resolve the phylogeny of australasian smurf-weevils. *PLoS One*. 12:e0188044.

Vanderpool D, Minh BQ, Lanfear R, Hughes D, Murali S, Harris RA, Raveendran M, Muzny DM, Hibbins MS, Williamson RJ, et al. 2020. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *PLoS Biol*. 18:e3000954.

Walker JF, Walker-Hale N, Vargas OM, Larson DA, Stull GW. 2019. Characterizing gene tree conflict in plastome-inferred phylogenies. *PeerJ*. 7:e7747.

Wilkinson M. 1994. Common cladistic information and its consensus representation: Reduced adams and reduced cladistic consensus trees and profiles. *Systematic Biology*. 43:343-368.

Wilkinson M. 2003. Missing entries and multiple trees: Instability, relationships, and support in parsimony analysis. *Journal of Vertebrate Paleontology*. 23:311-323.

Xi Z, Liu L, Davis CC. 2016. The impact of missing data on species tree estimation. *Mol Biol Evol*. 33:838-860.

Yue JX, Li J, Aigrain L, Hallin J, Persson K, Oliver K, Bergstrom A, Coupland P, Warringer J, Lagomarsino MC, et al. 2017. Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat Genet*. 49:913-924.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. Astral-iii: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*. 19:153.

Zhang YM, Williams JL, Lucky A. 2019. Understanding uces: A comprehensive primer on using ultraconserved elements for arthropod phylogenomics. *Insect Systematics and Diversity*. 3.