Supplementary Materials for "Whole-genome phylogenomics illuminates the tempo of tinamou (Aves: Tinamidae) diversification and reduced introgression on the Z-chromosome"

Authors:

Lukas J. Musher[1], Therese A. Catanach[1], Thomas Valqui[2], Robb T. Brumfield[3], Alexandre Aleixo[4], Kevin P. Johnson[5], Jason D. Weckstein[1,6]

Author affiliations:

[1] The Academy of Natural Sciences of Drexel University, Department of Ornithology Philadelphia, PA, 19103, USA.

[2] Facultad de Ciencias Forestales, Universidad Nacional Agraria La Molina, Lima, Peru and CORBIDI-Centro de Ornitología y Biodiversidad, Lima, Peru.

[3] Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA, 70803, USA.

[4] Instituto Tecnológico Vale - ITV, Belém, Brazil.

[5] Illinois Natural History Survey, Prairie Research Institute, University of Illinois Urbana-Champaign, Champaign, IL, USA.

[6] Department of Biodiversity, Earth, and Environmental Sciences, Drexel University, Philadelphia, PA, 19103, USA.

Overview

Here we present supplementary Methods, Results, Discussion, Figures and Tables.

Methods

DNA extraction, library preparation, and sequencing

For fresh tissues, we extracted genomic DNA using the MagAttract High Molecular Weight kit from Qiagen (Valencia, California). Toe pad extractions of historical study skins was carried out in a dedicated historical DNA extraction laboratory at the Academy of Natural Sciences of Drexel University to reduce contamination risk. Toe pad samples were first washed in a brief bath of EtOH to help remove superficial contaminants, and then soaked in H_2O for three hours to hydrate the desiccated flesh for DNA lysis. Samples were then digested using 180 µL ATL and 30 µL proteinase K for each sample and incubated at 56º C overnight. DNA isolation was then performed using the QiaQuick spin columns and extraction kit from Qiagen (Valencia, CA).

Shotgun sequencing libraries were prepared for each extract using the Hyper library construction kit (Kapa Biosystems). These libraries were sequenced using 150 bp paired-end reads on an S4 lane of an Illumina NovaSeq 6000. These libraries were pooled and tagged with unique dual-end adaptors, and pooling consisted of between 13 and 18 samples per lane aimed to achieve at least 30X coverage of the nuclear genome. Adapters were trimmed and demultiplexed using bcl2fastq v.2.20. We deposited raw reads in the NCBI SRA (Table S1).

After mapping reads to reference genomes using BWA (see main text), we used '-doFasta 2' flag in ANGSD to ensure that the consensus nucleotide was written for each polymorphic site.

In an earlier draft of this manuscript we used CryUnd as reference for several samples (Musher et al. 2024). During revisions we found that samples mapped to this genome were more likely to cluster with each other than with other more likely relatives in certain phylogenetic

analyses. This led us to remap these samples to a different reference, *C. strigulosus*, in the current version.

Assessing genome completeness

To assess genome quality, completeness, and redundancy for each assembly, we used Benchmarking Universal Single Copy Orthologs (BUSCO) version 5.3.0 (Simão et al. 2015). BUSCO searches genome assemblies and identifies genes present in single copy using a database of known single-copy orthologs from a clade-specific database of genes. We used the 'aves_odb10' lineage dataset, which utilizes 8,338 genes in the chicken genome. We used the '--augustus' flag to obtain nucleotide sequences for genes. This setting uses augustus version 3.2 (Hoff et al. 2019) to annotate each assembly, and outputs full nucleotide gene sequences for all complete single-copy orthologs in fasta format. This step was necessary to obtain our coding gene dataset for phylogenomic analysis. We also used samtools to estimate mean and standard deviation of sequence coverage for each genome.

Alignment trimming

An earlier version of this study found that BUSCOs with more Parsimony Informative Sites did not significantly improve gene tree accuracy, implying a strong signal of systematic error in this dataset. We examined a handful of BUSCO alignments by eye and found long sequences unique to a small number of individuals in each alignment. We suspected that these were driven by misannotation in the BUSCO pipeline, and thus trimmed all loci using TrimAL to remove sites in each alignment with missing data for more than two samples. This seemed to alleviate the problem as we found a tighter correlation between the number of parsimony informative sites and RF distance between gene and species trees after trimming.

Phyluce harvesting of UCEs from whole genomes

To generate Z-linked and autosomal datasets, we first used the Phyluce scripts to harvest all UCEs from each pseudo-chromosome-level genome. We then made a list of UCE loci mapping to the Z-chromosome by using "grep" in UNIX to extract lines in the '.lastz' file containing UCEs on the Z-chromosome in each genome and "grep -v" to exclude lines on the Z-chromosome (autosomal dataset). This generated new '.lastz' files for each sample which were then used in the Phyluce pipeline.

Filtering of BUSCO's to assess robustness of linear/log-linear regression models

We explored the linear models relating alignment informativeness (proportion and number of PIS) and RF distance between gene and species trees (see main text). To examine if differences between UCE and BUSCO results could be driven by systematic error, we explored these models again after filtering the BUSCO for GC content variation (by removing genes with GC content variation greater than two standard deviations from the mean) and for BUSCOs resembling UCE datasets (removing genes greater than two standard deviations above the mean proportion of PIS in the UCE1000Flank dataset and below the mean of the UCE100Flank dataset).

To assess whether the linear relationships of BUSCOs were inherent to the data type rather than a pattern driven by outlier loci, we filtered the BUSCO dataset and reran linear regression models on filtered subsets. We applied three filters. First, we filtered by GC content, removing genes that fell more than two standard deviations away from the mean BUSCO GC content. Second, we filtered for UCE-like genes. Specifically, we removed BUSCO loci falling two standard deviations away from the mean number and proportion of parsimony informative sites (PIS) of all UCE datasets combined. The idea here is to attempt to replicate a dataset that looks like the right-most plots in Figure 5 of the main text (in terms of information content). Finally, we combined both previous filters to make a strictly filtered dataset.

Estimating the tempo of tinamou diversification

First, we filtered the autosomal dataset to include only a single member of each taxon (monotypic species or subspecies) and created a new 100% complete autosomal dataset. Next, we generated new gene trees for these alignments and adapted scripts from a prior study (Musher et al. 2019) to quantify the RF-distance between gene and species tree and a measure of clock-likeness for each alignment in the autosomal UCE dataset. Specifically, we estimated the likelihood ratio of clocklike to non-clocklike models for each gene; lower ratios indicate molecular evolution that is more clocklike. The protocol we followed was identical to that in Musher et al. (2019). After estimating likelihood ratios and RF-distances for all alignments and gene trees, we selected only those loci with RF distances below the mean distance, and with likelihood ratios that were two standard deviations below the mean ratio. This allowed us to include a relatively small number (22 in total) of genes with relatively low variance in molecular substitution rates among branches, but were relatively informative for phylogenetic reconstruction (Figure S26).

To estimate the timeframe of tinamou diversification, we used a fossilized birth-death model with six crown Tinamidae tip-fossil calibrations using an optimized relaxed clock and fossilized birth-death model in Beast v2.7.7 with the Sampled Ancestors and Optimized Relaxed Clock packages (Drummond et al. 2006; Gavryushkina et al. 2014; Bouckaert et al. 2019; Zhang and Drummond 2020; Douglas et al. 2021), and conditioning on the root and rho parameters. We then used the filtered autosomal UCE1000 alignments described above, generating lines of question marks in each alignment for six fossil taxa (see Table S2). We applied tip dates measured as years before the present using initial dates listed in Table S2, and uniform priors on tip ages using the upper and lower bounds of their estimated ages from the fossil record. We linked all clock and tree models but partitioned site models by locus (22 in all), with a GTR + G model for each partition. Default priors and settings were used except where specified in Table S3. We ran the MCMC for 150,000,000 generations with sampling

every 50,000 generations. The tree was summarized as a maximum clade credibility tree with median values for node heights, and a burn-in of 20% in TreeAnnotator v2.7.7 (Bouckaert et al. 2019). Convergence was assessed using Tracer v1.7.1 (Rambaut et al. 2018).

We also modeled speciation and extinction rates using the R-package 'TESS' (Höhna et al. 2016). TESS uses a rjMCMC algorithm to estimate the timing of speciation and extinction rate shifts under an episodic birth-death model. To do so, we pruned fossil tips and outgroups out of the beast tree, and ran the MCMC for 10^6 generations, assuming 85% complet sampling for the clade.

Results

Assembly metrics, completeness, and ortholog metrics

We successfully assembled 68 of 69 newly sequenced genomes and extracted BUSCO and UCE targets from these assemblies plus 10 publicly available tinamou and 2 publicly available outgroup whole genome assemblies (Table S1). Genome wide sequence coverage varied within and among samples, but was generally high; the mean of average coverage across samples was 40.41x (standard deviation = 13.55x). Most genomes were also relatively complete, containing an average of 89.69% (standard deviation = 10.88%) of 8,338 complete single-copy BUSCO genes.

One sample (*C. soui poliocephalus* ANSP 187533) had relatively low coverage (mean coverage = 10.42x, standard deviation = 76.95x) and thus read-mapping resulted in a highly incomplete genome (Table S1). This sample was dropped from all subsequent analyses. Another genome, downloaded from NCBI (GCA 013389825) was labeled as *C. undulatus*, but did not cluster with other members of that species, instead clustering with *C. strigulosus* (most trees). We could not confirm the species identity of this sample because the voucher was unavailable or lost, but photographs of the headless voucher are consistent with *C. strigulosus*. Moreover, another downloaded genome, GCA 013398335, was listed as *Nothoprocta ornata* on

NCBI but after including newly sequenced material, this *N. ornata* clustered with *N. pentlandii* samples from Peru. After examining the voucher specimen (CORBIDI 168605), the sample was indeed confirmed as *N. pentlandii oustaleti* and not *N*. *ornata* as indicated in NCBI.

Breakpoint analysis to find the optimal number of reticulate nodes

The slope of the segmented model was significant and tightly fitting for both autosomal UCEs (Breakpoint = 3, Adj. $R^2 = 0.97$, $P < 0.0001$) and Z-linked UCEs (Breakpoint = 1, Adj. $R^2 = 0.98$, $P < 0.0001$).

Discussion

Comparisons of estimation error between data types and the benefits of whole genome phylogenomics

We measured rates of gene tree estimation error using RF distances between gene trees and the inferred species tree, under the assumption that increases in RF distance were indicative of higher rates of phylogenetic error. Although gene histories are not expected to be congruent with the species tree in many cases (Tajima 1983; Pamilo and Nei 1988; Maddison 1997), the assumption that RF distances are good indicators of error rates is valid because different datasets contained different means and variances in RF distances among gene trees. Although the R^2 was small for the UCE1000 dataset, the overall effect was most evident when the three UCE datasets were combined into a single analysis, which showed asymptotic convergence on relatively low mean and variance RF distances after reaching about 500 PIS (Figure 5). The asymptotic shape of this relationship implies that as PIS are added to an alignment, the resulting gene trees converge on a level of heterogeneity that is representative of or approaching real biological signal (i.e., ILS and/or introgression) rather than methodological or data-driven artifacts. However, it is difficult to separate this hypothesis from one where creasting larger sequences may also lead to the inclusion of multiple c-genes, which could then drive

convergence in gene tree topologies by mimicking the effects of concatenation. Although some level of gene tree heterogeneity is expected in *all* phylogenomic datasets, even if one assumes the impossible, where empirical gene tree estimation error is absent (Maddison 1997; Gatesy and Springer 2014; Edwards et al. 2016), it is reasonable to assume that differences in the mean and variance in RF distances among datasets are good proxies for differences in gene tree estimation error. Nevertheless, data biases and stochastic error had relatively little impact on our overall results, except that MSC results were most impacted by stochastic error; UCE100 and UCE300 MSC trees were quite different from other topologies.

Supplemental Tables and Figures:

Fossil	Stratigraphic level	Estimated age (Ma)	Phylogenetic position justification	Prior Details	Citations
Diogenornis fragillis	Itaboraian-Casamayoran	53–48.6 Initial tip date: NA	Multiple synapomorphies with Rheidae place this fossil in Rheiformes	Lognormal MRCA constraint on age of "Higher ratites" Offset=55, M=7, S=1	Alvarenga 1983, Almeida et al. 2022
Undescribed fossil Macn-sc-3610	Monte observacion, Santa Cruz Formation	18–15.2 Initial tip date: 16.6	Phylogenetic analysis of morphological data placed this fossil as sister to crown *Crypturellus*	Constrained monophyly of *Crypturellus* + stem fossils, thus allowing this fossil to freely vary within stem and crown *Crypturellus*	Bertelli and Chiappe 2005, Bertelli et al. 2014, Almeida et al. 2022
Undescribed fossil Macn-sc-3613	Monte observacion, Santa Cruz Formation	18–15.2 Initial tip date: 16.6	Phylogenetic analysis of morphological data placed this fossil as sister to crown *Crypturellus*	Constrained monophyly of *Crypturellus* + stem fossils, thus allowing this fossil to freely vary within stem and crown *Crypturellus*	Bertelli and Chiappe 2005, Bertelli et al. 2014, Almeida et al. 2022
Crypturellus reai	Cañadón de las Vacas, Santa Cruz Formation	17.5–16.3 Initial tip date: 16.9	Previous phylogenetic analysis was equivocal, placing this specimen either in stem or crown *Crypturellus*	Constrained monophyly of *Crypturellus* + stem fossils, thus allowing this fossil to freely vary within stem and crown *Crypturellus*	Bertelli et al. 2014, Almeida et al. 2022
"Eudromia" sp.	Cerro Azul Formation at Salinas Grandes	7.2–6 Initial tip date: 6.6	Previous phylogenetic analysis placed this fossil as sister to *Eudromia* + *Tinamotis*	Constrained phylogenetic position of "*Eudromia* sp." thus constraining the position sister to *Tinamotis*+*Eudromia*	Bertelli et al. 2014, Almeida et al. 2022
Eudromia olsoni	Farola Monte Hermoso Montehermosan	5–4.5 Initial tip date: 4.75	Previous phylogenetic analysis placed this fossil in the genus *Eudromia*	Constrained monophyly of extant *Eudromia* + fossil allowing the fossil to freely vary within crown *Eudromia*	Bertelli et al. 2014, Almeida et al. 2022

Nothura parvula	Chapadmalal Formation	5–4.5 Initial tip date: 4.75	Previous phylogenetic analysis was equivocal, placing this species either in *Nothura* or as sister to *Nothura* + *Nothoprocta* + *Rhynchotus*. Lack of synapomorphies with *Nothoprocta* + *Rhynchotus* suggest it is a stem or crown *Nothura.*	Constrained monophyly of *Nothoprocta* + *Rhynchotus*, thus allowing this fossil to freely vary within stem and crown *Nothura*, and stem *Nothoprocta* + *Rhynchotus*	Bertelli et al. 2014; Almeida et al. 2022

Table S2: A list of all tip fossils used in the Beast analysis along with justification for their constrained positions, prior distributions and citations.

Parameter Name	**Prior**	**Details**
ORCRates	Log Normal	Initial=0.5, M=1.0, S=0.2, Offset=0.0
ORCSigma	Exponential	Initial=0.2, M=0.3337,Offset=0.0
ORCucldMean	Exponential	Initial=0.001, M=10.0, Offset=0.0
rhoFBD	Uniform	Initial=0.75, Lower=0.5, Upper=1.0, Offset=0.0
SamplingProportionFBD	Exponential	Initial=0.25, M=0.25, Offset=0.0
MRCA	No prior	Constrained monophyly for several clades to improve accuracy of the phylogenetic results given the relatively small number of loci used. These are marked with white circles in Figure 1 of the main text.

Table S3: Non-default priors used in the beast analysis

Dataset	**Number of loci**	**Total concatenated length**
UCE100 complete	2,712	879,985
UCE100 75% complete	4,631	1,482,732
UCE300 complete	2,702	1,864,316
UCE300 75% complete	4,621	3,144,625
UCE1000 complete	2,628	4,643,253
UCE1000 75% complete	4,514	7,846,363
UCE1000 autosomes 75% complete	4,244	7,384,066

UCE1000 Z-chromosome 75% complete	304	677,088
BUSCOs complete	2,274	3,103,102
BUSCOs 75% complete	7,414	10,083,122

Table S4: Number of alignments and total concatenated length for each dataset.

Dataset	Model	Adj. R-squared	slope	P	AIC
BUSCOs	linear (#PIS)	0.3387	-0.0305809	<0.0001	19291
	logarithmic (#PIS)	0.497	-20.6024	<0.0001	18669
	linear (%PIS)	0.001863	-7.36	0.0396	20228
	logarithmic (%PIS)	0.01044	-6.684	<0.0001	20208
UCE100Flank	linear (#PIS)	0.7797	-0.759178	<0.0001	19897
	logarithmic (#PIS)	0.6933	-25.8442	<0.0001	20794
	linear (%PIS)	0.7747	-237.6429	<0.0001	19959
	logarithmic (%PIS)	0.6933	-25.8442	<0.0001	20843
UCE300Flank	linear (#PIS)	0.636	-0.254330	<0.0001	21092
	logarithmic (#PIS)	0.7253	-37.7134	<0.0001	20331
	linear (%PIS)	0.6039	-160.9276	<0.0001	21320
	logarithmic (%PIS)	0.705	-35.7484	<0.0001	20523
UCE1000Flank	linear (#PIS)	0.24	-0.028939	<0.0001	18683
	logarithmic (#PIS)	0.2895	-18.5317	<0.0001	18506
	linear (%PIS)	0.1312	-39.8117	<0.0001	19035
	logarithmic (%PIS)	0.2287	-9.0358	<0.0001	18927
UCE Combined	linear (#PIS)	0.6814	-0.108918	<0.0001	24575

	logarithmic (#PIS)	0.9197	-28.6239	<0.0001	20828
	linear (%PIS)	0.7765	-248.5049	<0.0001	23611
	logarithmic (%PIS)	0.791	-48.0636	<0.0001	23428

Table S5: Model tests of log versus linear fits for all generalized linear models. This model test was done twice for each dataset, once using the number of parsimony informative sites per locus (#PIS) and once using the percentage of parsimony informative sites per locus (%PIS). Rows highlighted in gray indicate the best model for each paired model test.

Table S6: Clade ages from this and other studies (supplemental file only, <u>not shown</u>)

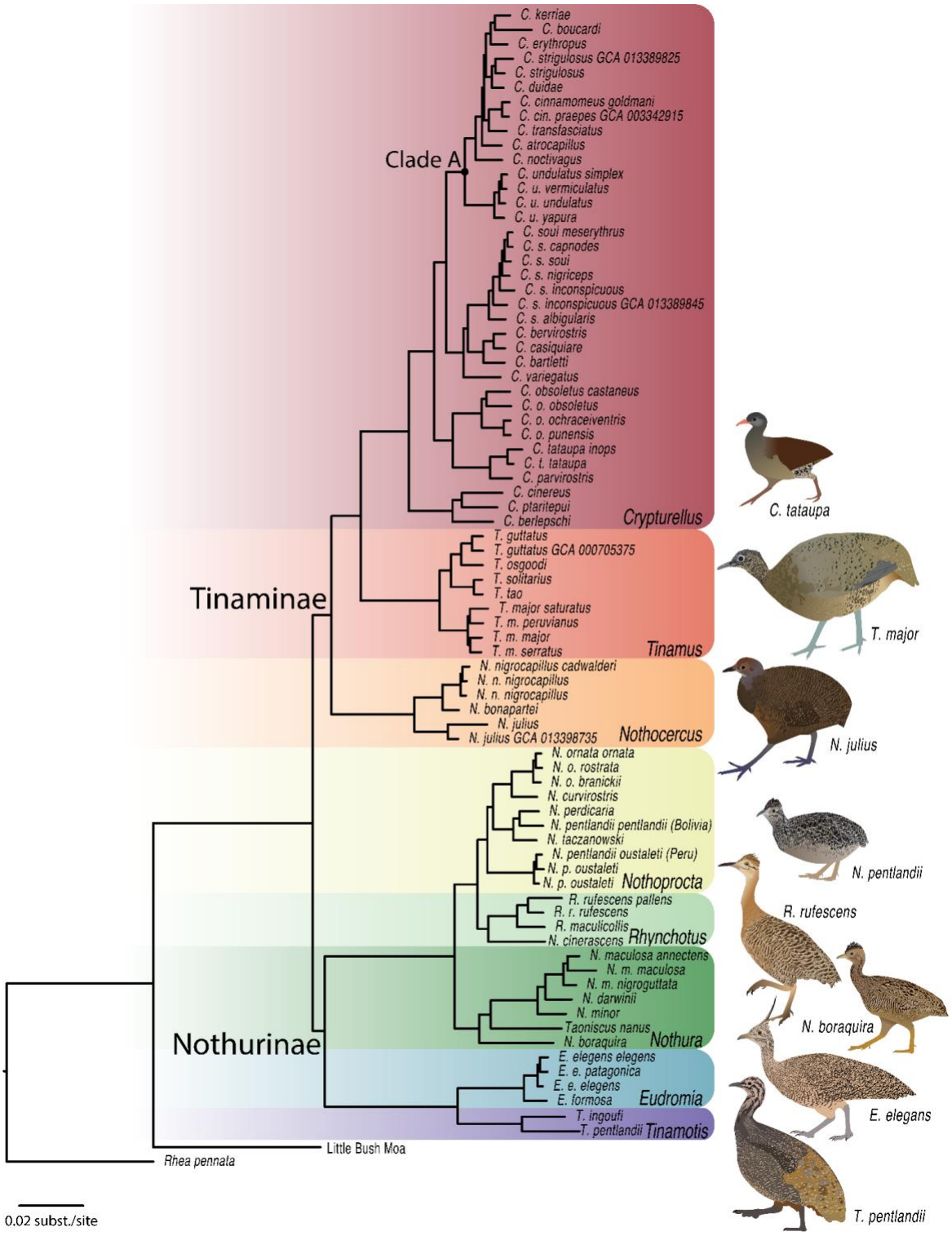


Figure S1: Concatenated Phylogeny of tinamous based on UCEs with 1,000 bp of flanking sequence. All nodes have bootstrap value of 100% except where noted. Note relatively short internodal branch lengths in Clade A. Tinamou illustrations by TAC.

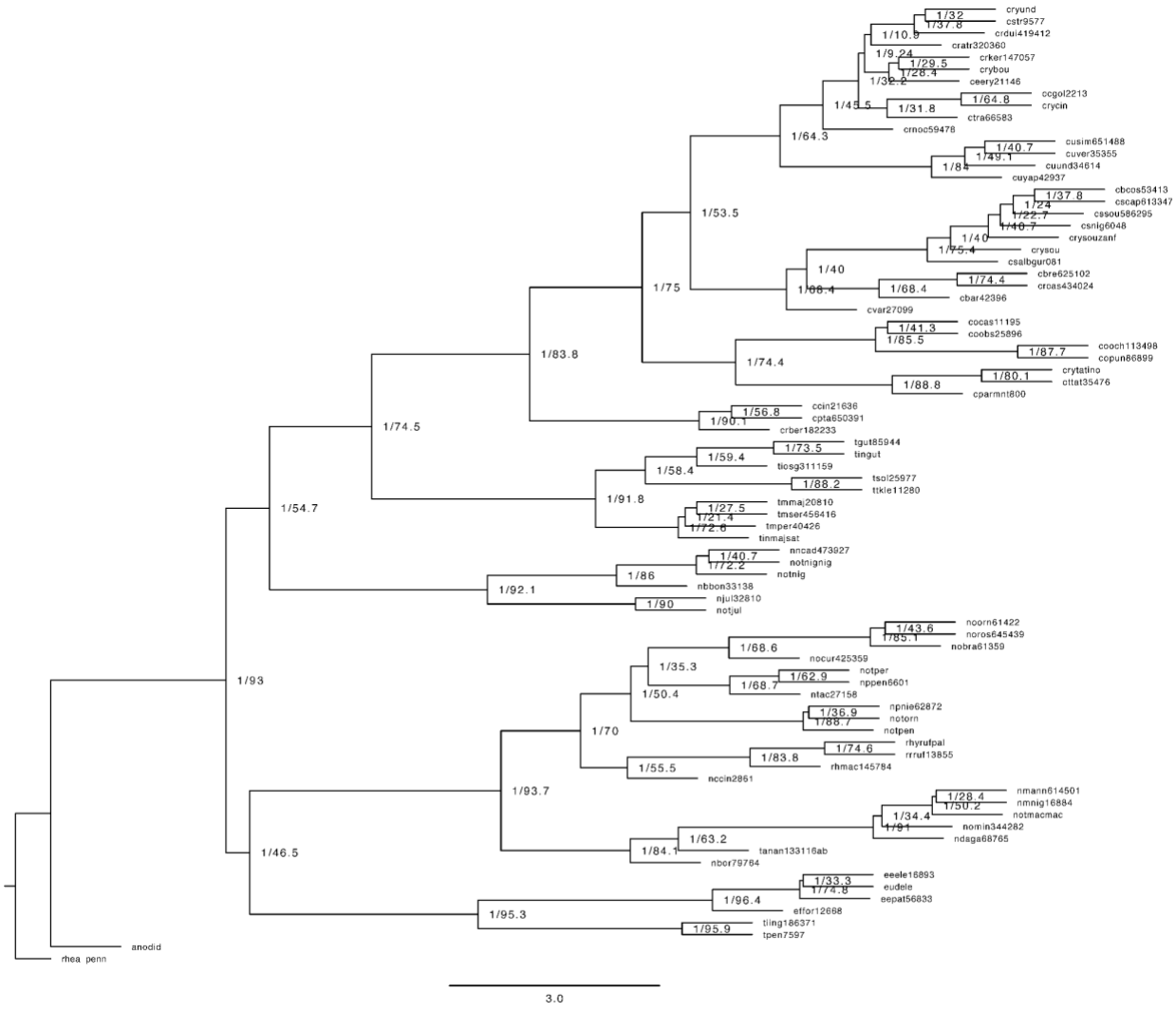


Figure S2: BUSCOs 75% complete astral phylogeny.

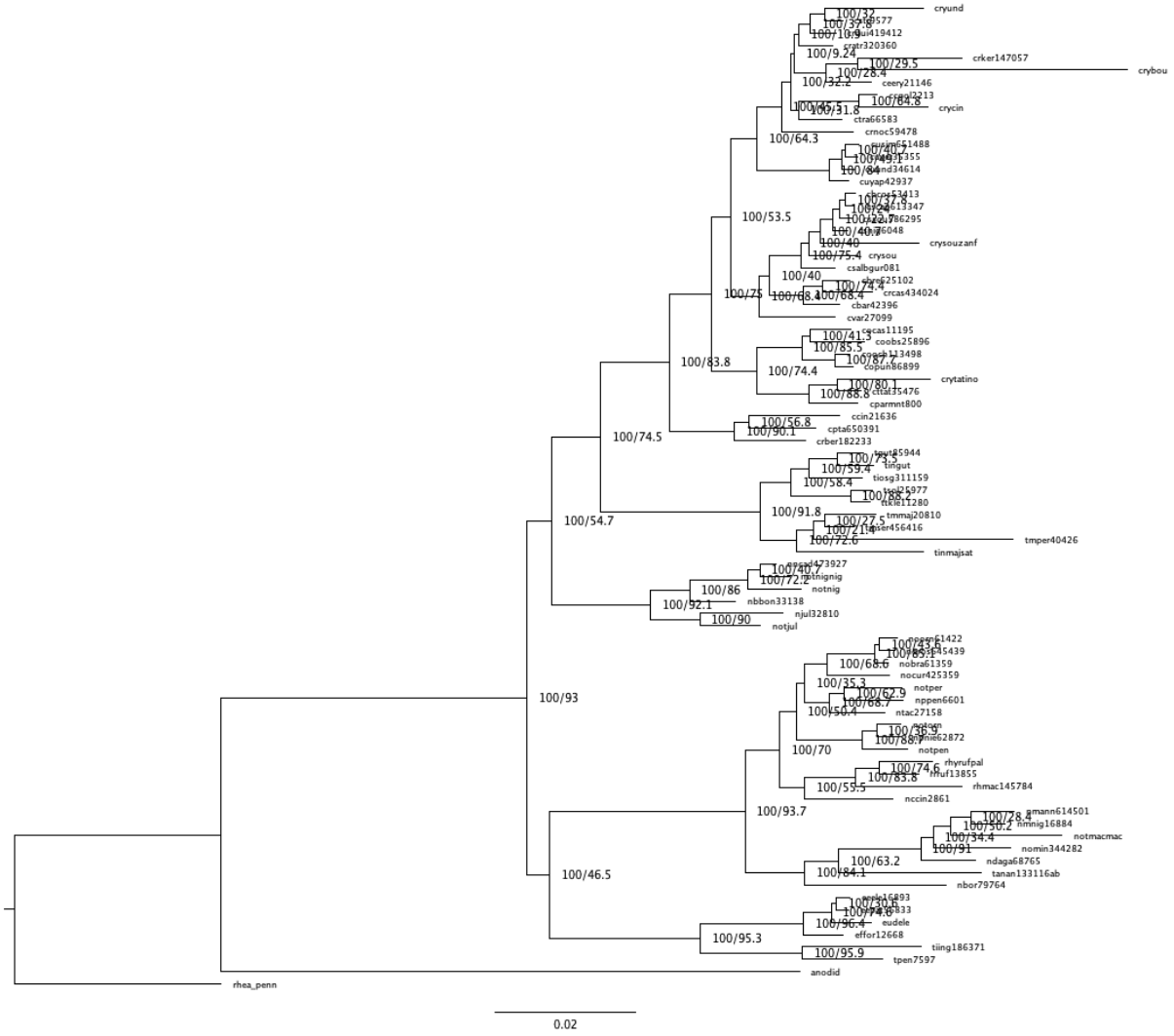


Figure S3: BUSCOs 75% iqtree phylogeny

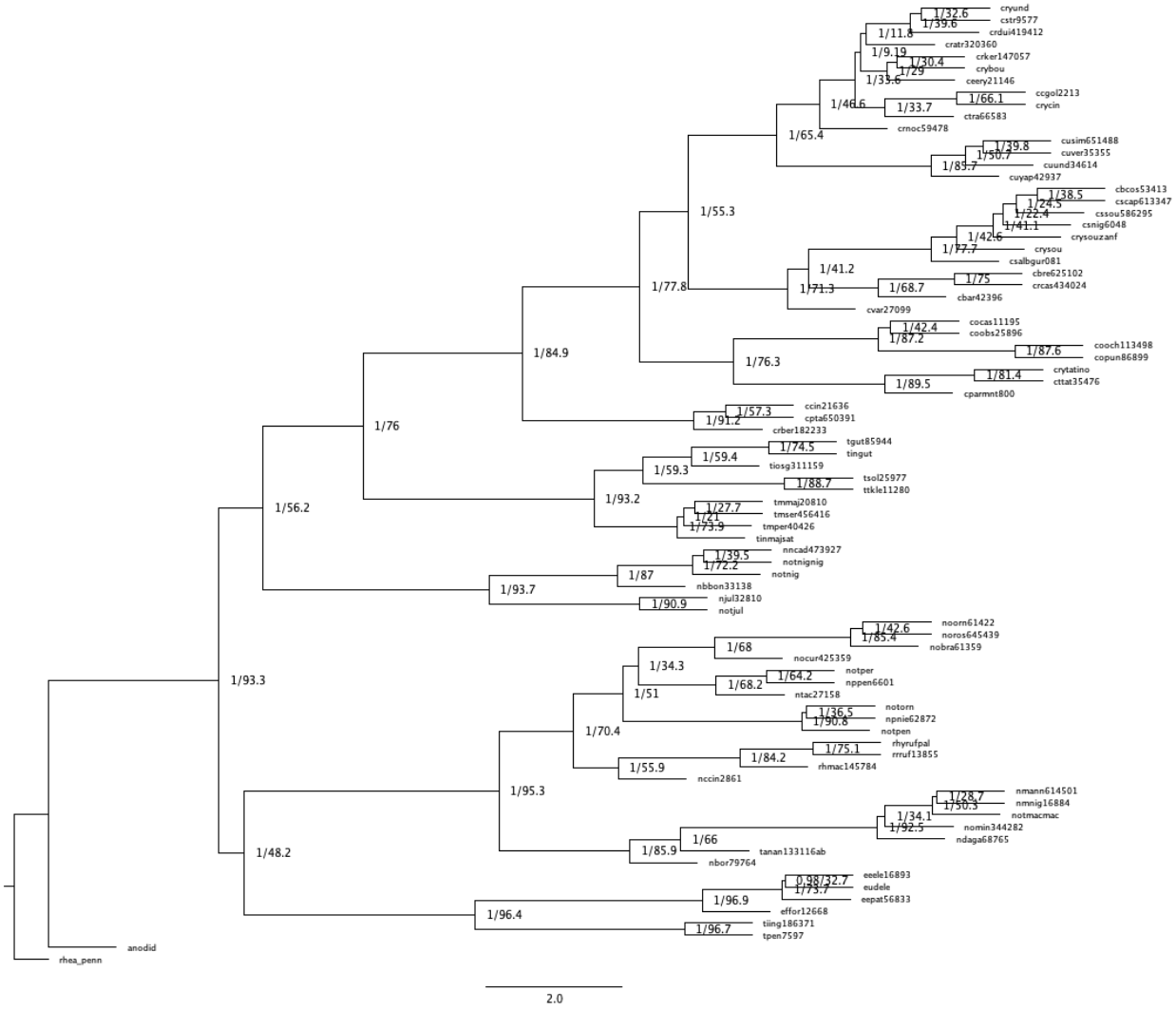


Figure S4: BUSCOs 100% astral phylogeny

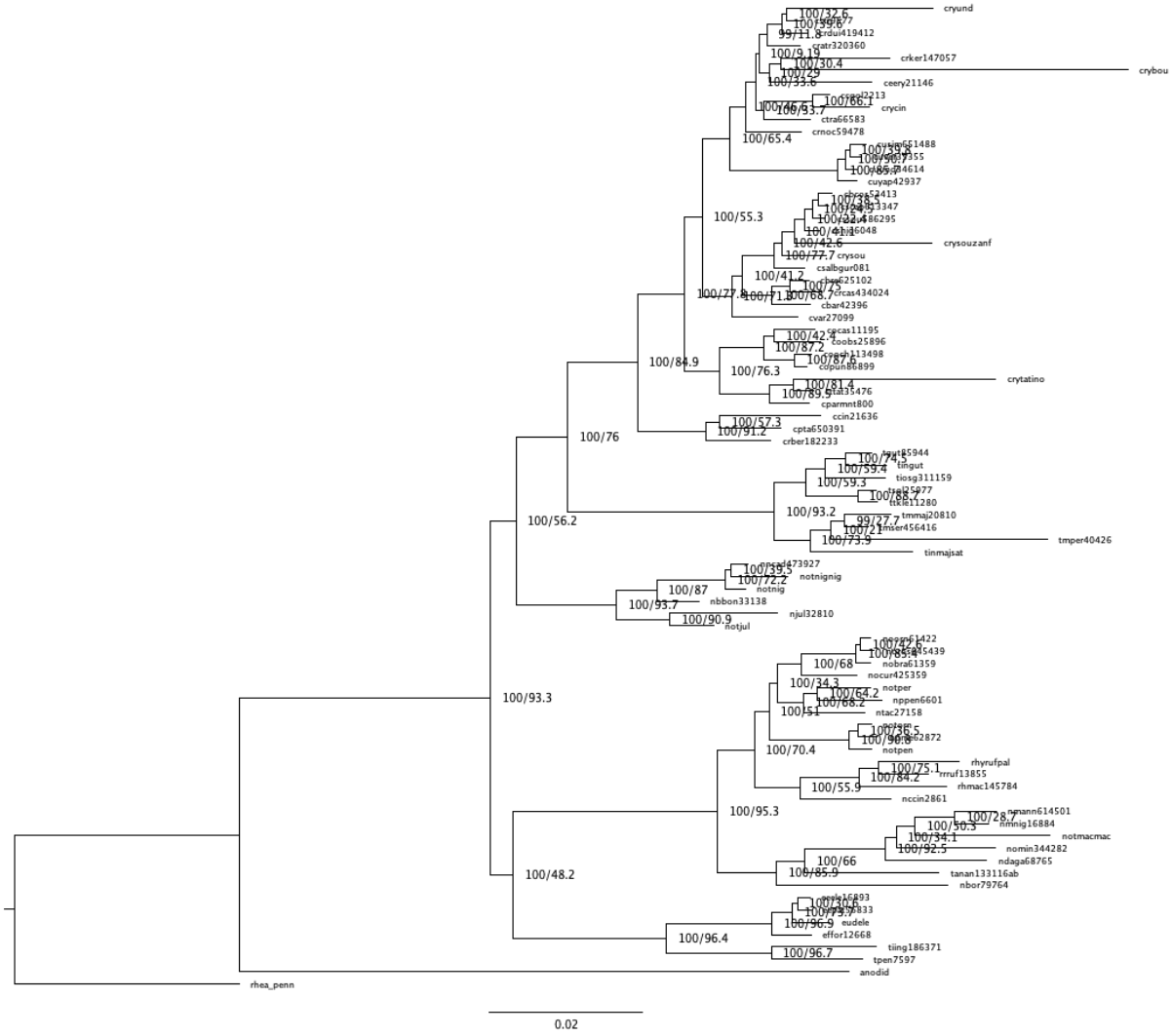


Figure S5: BUSCOs 100% iqtree phylogeny

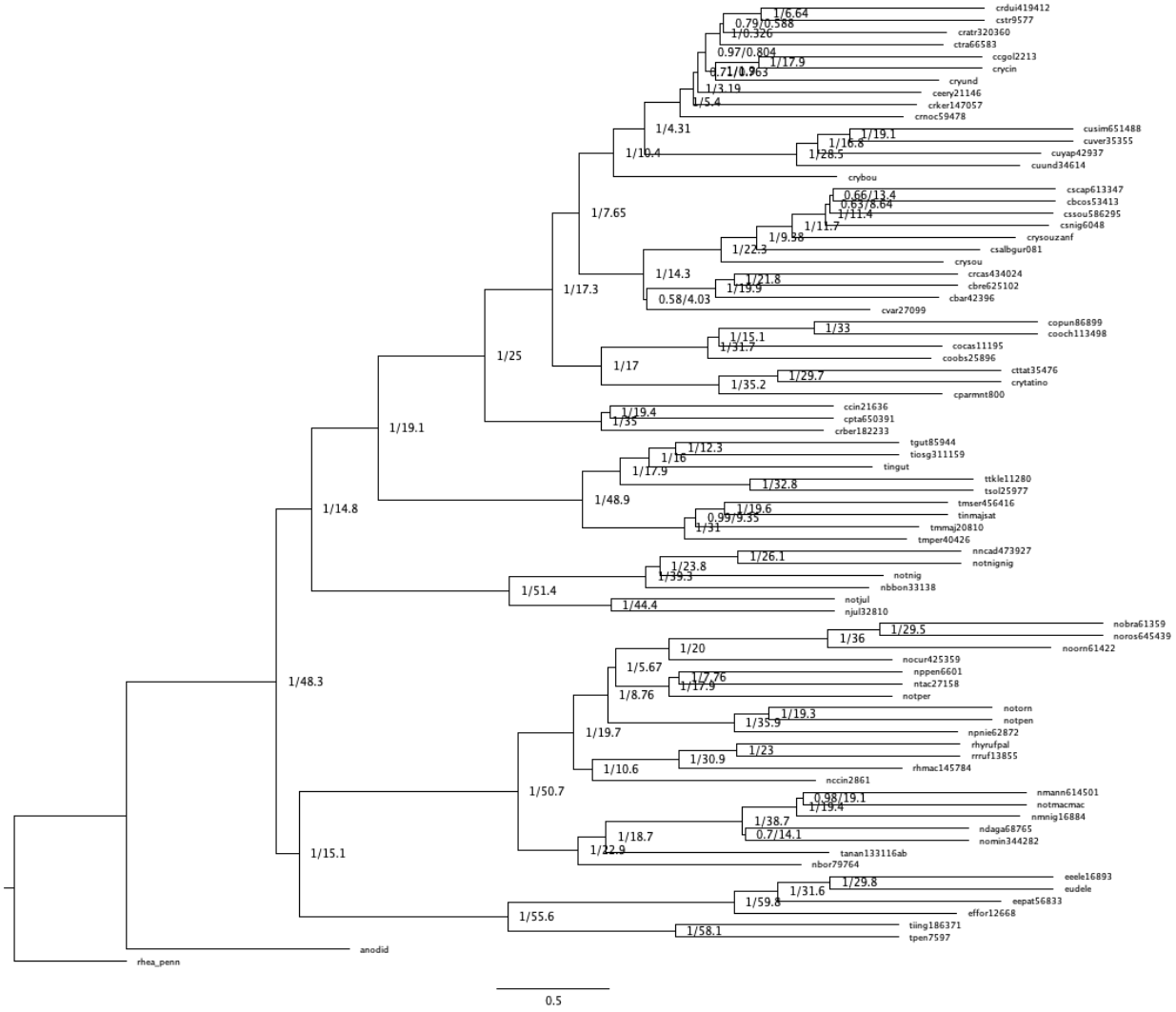


Figure S6: UCE100 75% astral phylogeny

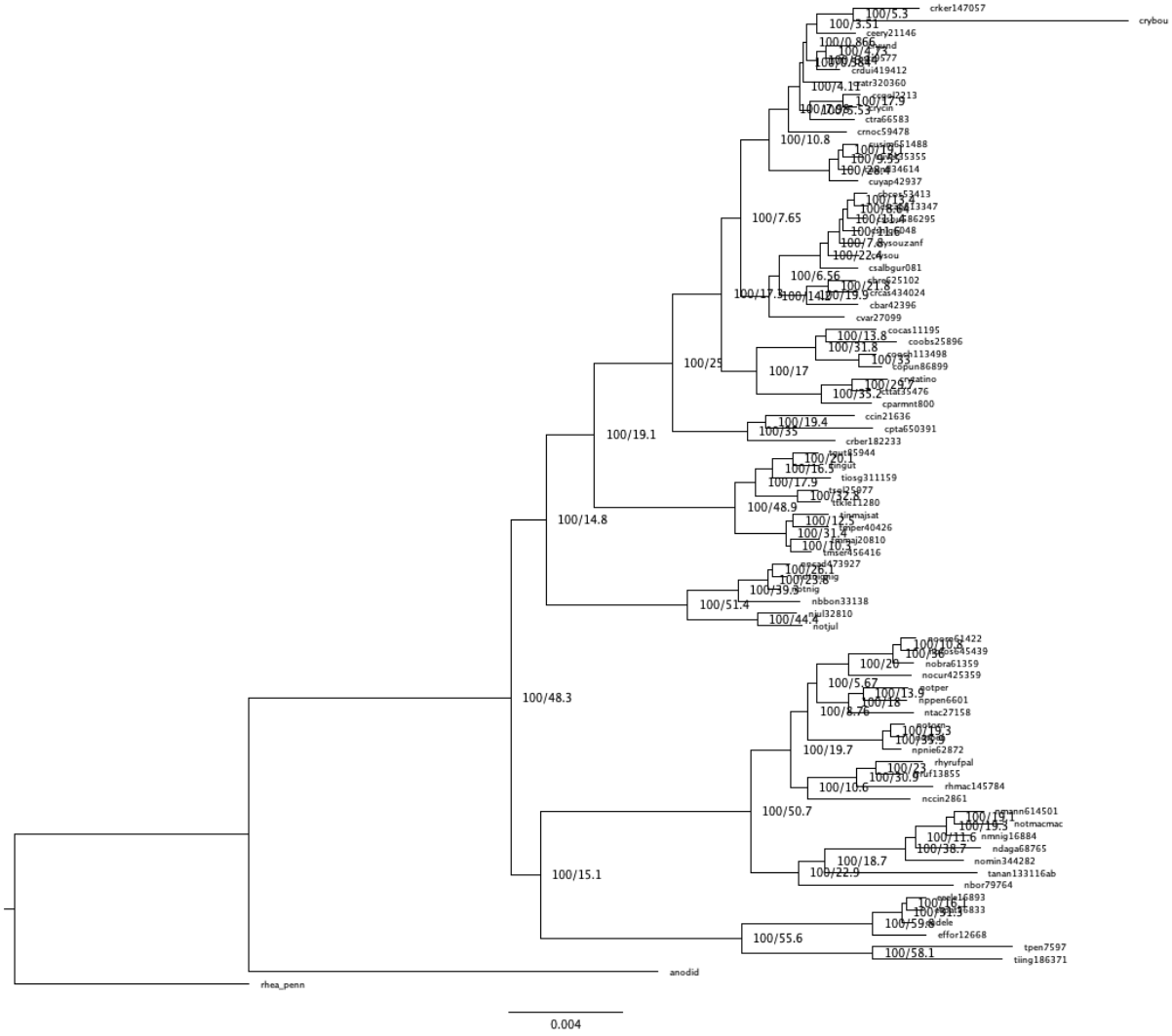


Figure S7: UCE100 75% iqtree phylogeny

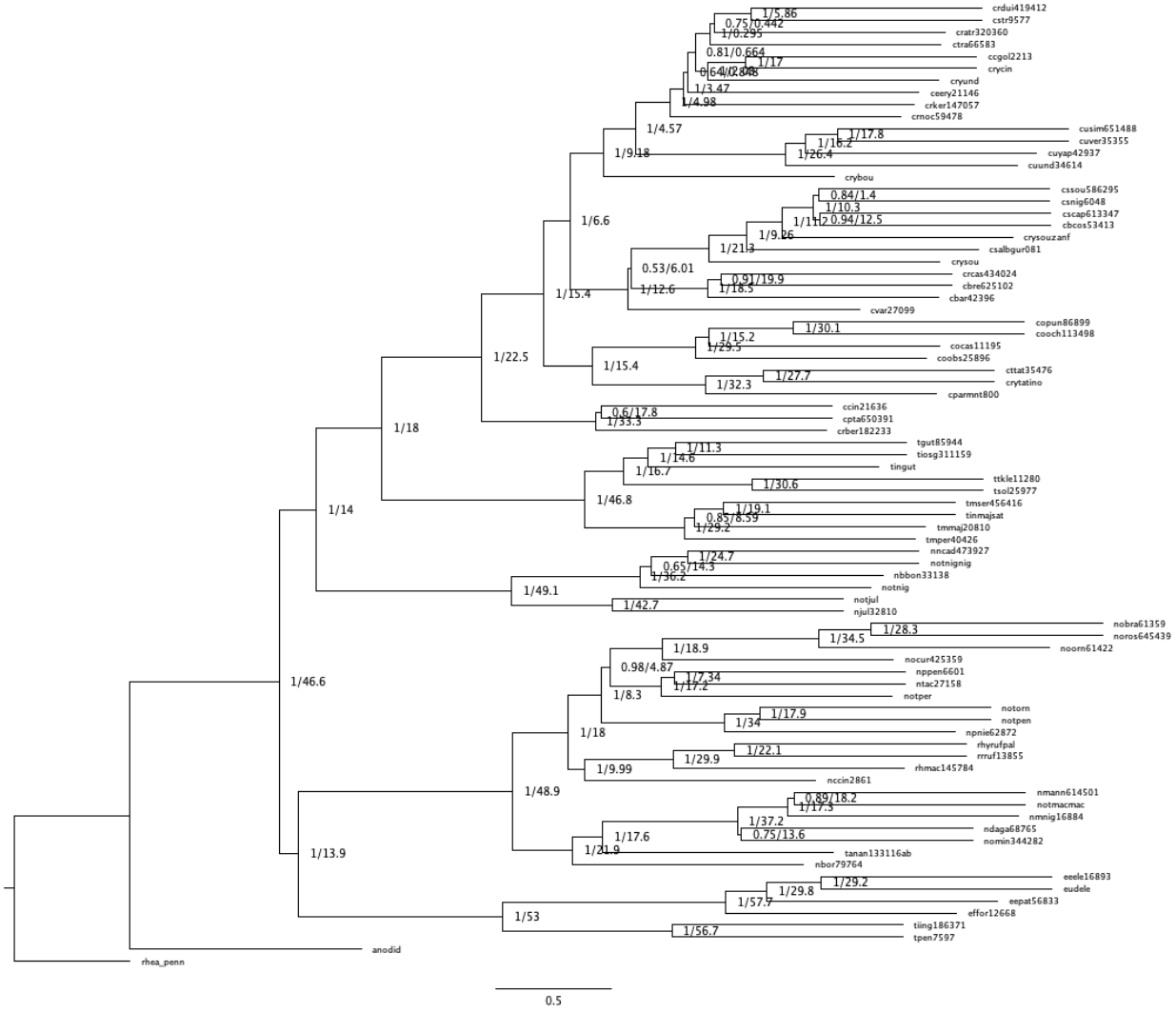


Figure S8: UCE100 100% astral phylogeny

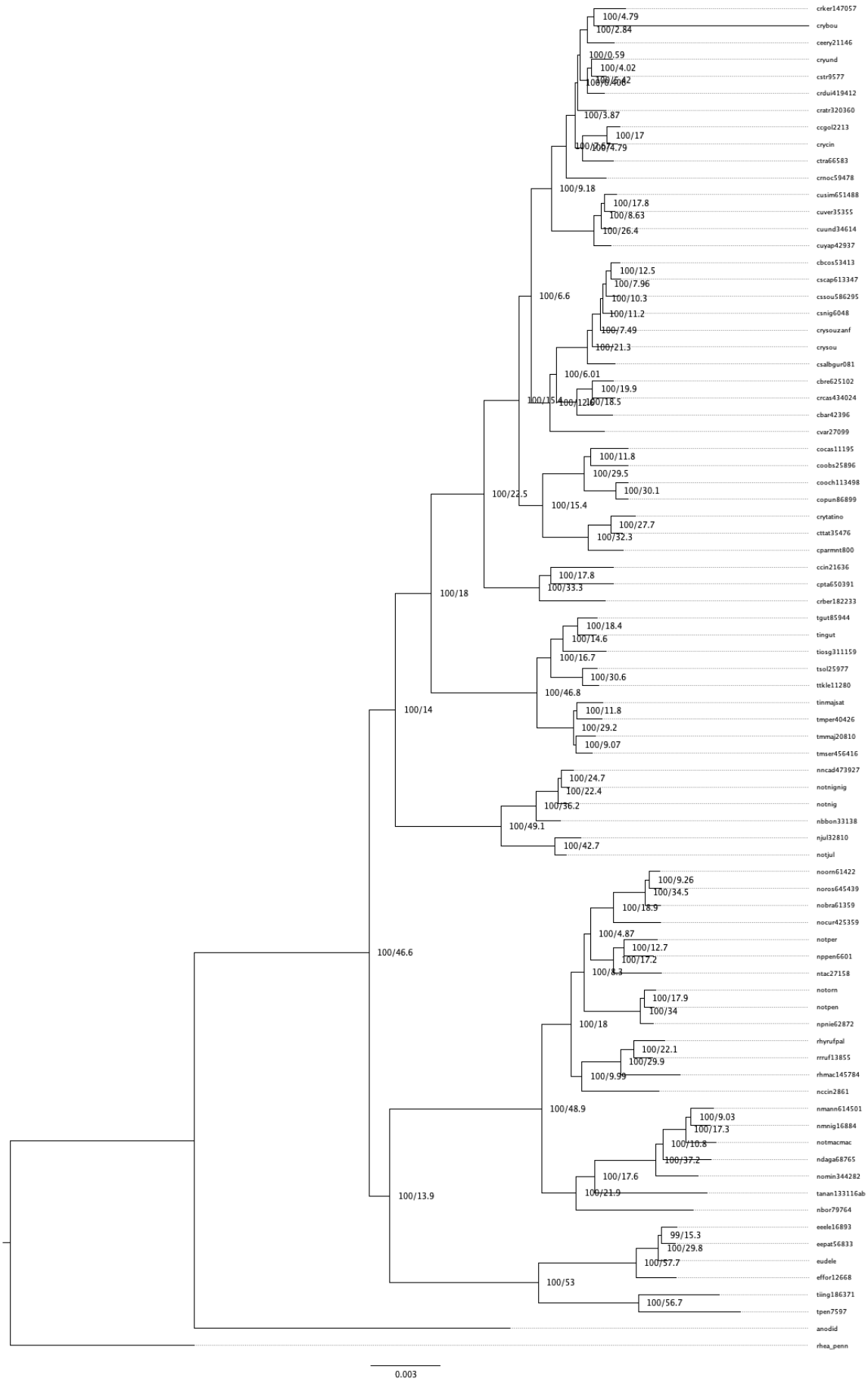

crker147057
crybou
ceery21146
cryund
cstr9577
crdui419412
cratr320360
ccgol2213
crycin
ctra66583
crnoc59478
cusim651488
cuver35355
cuund34614
cuyap42937
cbcos53413
cscap613347
cssou586295
csnig6048
crysouzanf
crysou
csalbgur081
cbre625102
crcas434024
cbar42396
cvar27099
cocas11195
coobs25896
cooch113498
copun86899
crytatino
cttat35476
cparmnt800
ccin21636
cpta650391
crber182233
tgut85944
tingut
tiosg311159
tsol25977
ttkle11280
tinmajsat
tmper40426
tmmaj20810
tmser456416
nncad473927
notnignig
notnig
nbbon33138
njul32810
notjul
noorn61422
noros645439
nobra61359
nocur425359
notper
nppen6601
ntac27158
notorn
notpen
npnie62872
rhyrufpal
rrruf13855
rhmac145784
nccin2861
nmann614501
nmnig16884
notmacmac
ndaga68765
nomin344282
tanan133116ab
nbor79764
eeele16893
eepat56833
eudele
effor12668
tiing186371
tpen7597
anodid
rhea_penn
100/4.79
100/2.84
100/0.59
100/4.02
100/3.87
100/17
100/9.18
100/17.8
100/8.63
100/26.4
100/12.5
100/7.96
100/6.6
100/10.3
100/11.2
100/7.49
100/21.3
100/6.01
100/19.9
100/18.5
100/11.8
100/29.5
100/30.1
100/22.5
100/15.4
100/27.7
100/32.3
100/17.8
100/33.3
100/18
100/18.4
100/14.6
100/16.7
100/30.6
100/46.8
100/14
100/11.8
100/29.2
100/9.07
100/24.7
100/22.4
100/36.2
100/49.1
100/42.7
100/9.26
100/34.5
100/4.87
100/12.7
100/17.2
100/46.6
100/17.9
100/34
100/18
100/22.1
100/29.9
100/9.99
100/48.9
100/9.03
100/17.3
100/10.8
100/37.2
100/17.6
100/21.9
100/13.9
99/15.3
100/29.8
100/57.7
100/53
100/56.7
0.003

Figure S9: UCE100 100% iqtree phylogeny

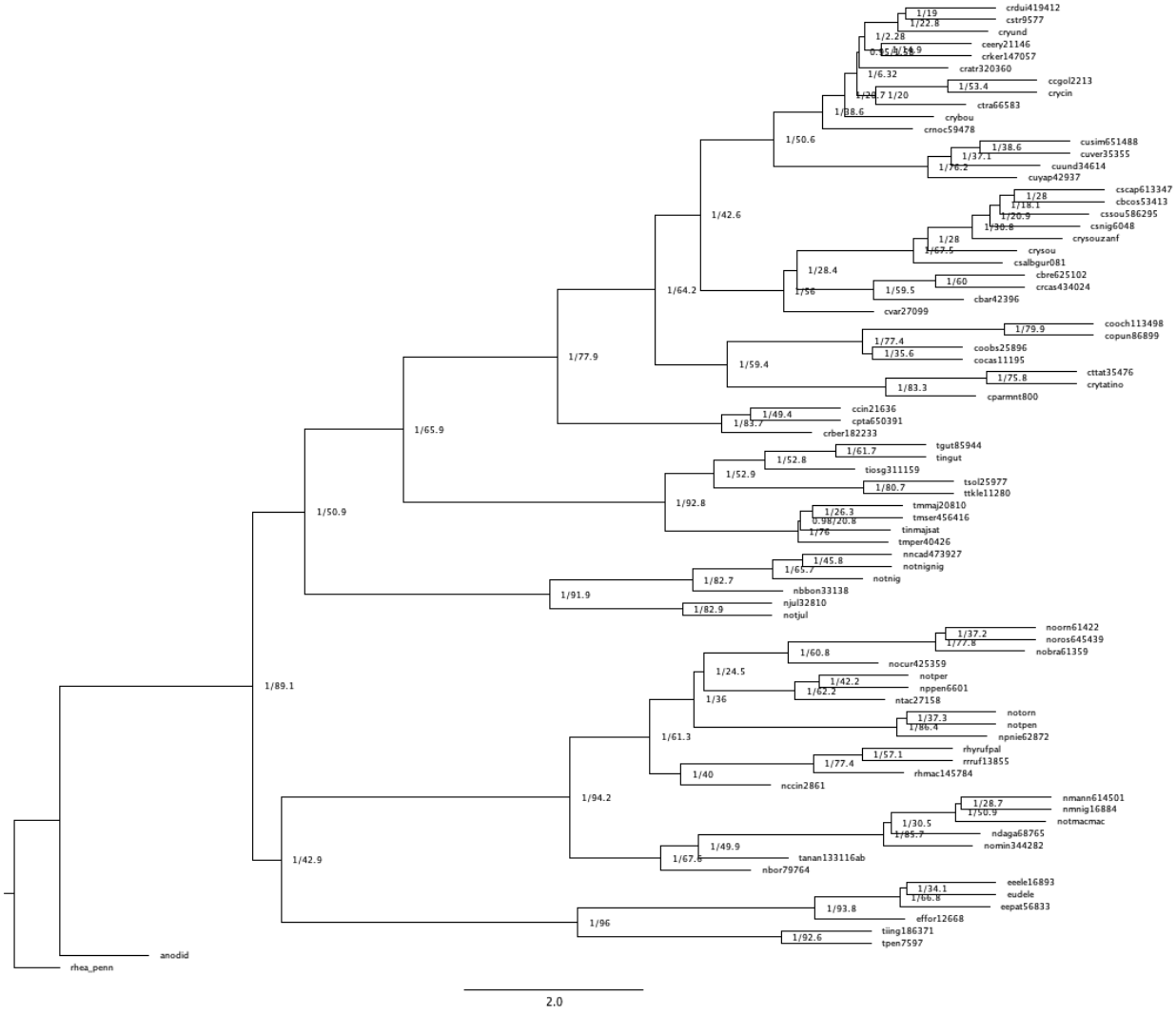


Figure S10: UCE300 75% ast phylogeny

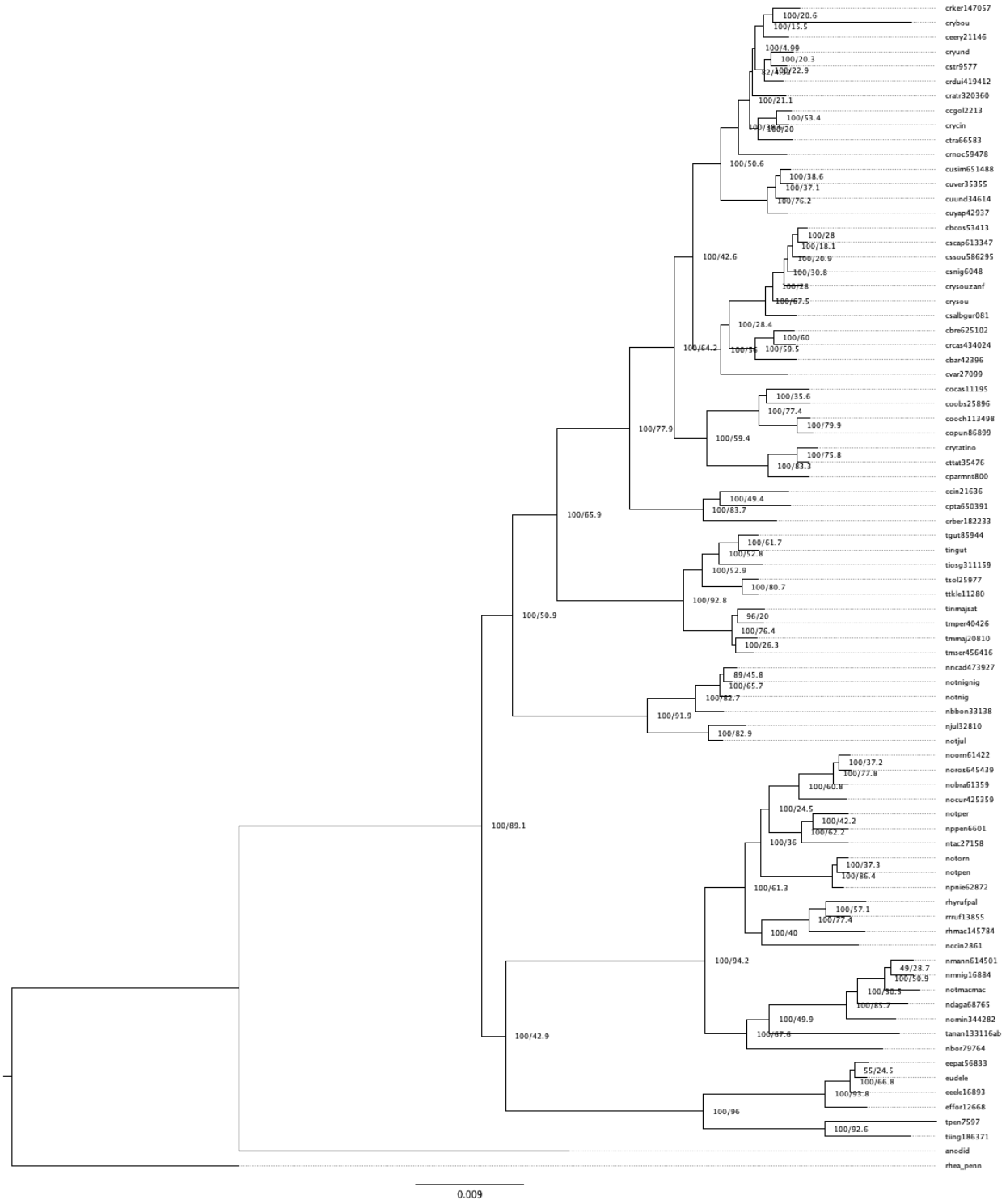


Figure S11: UCE300 75% iqtree phylogeny

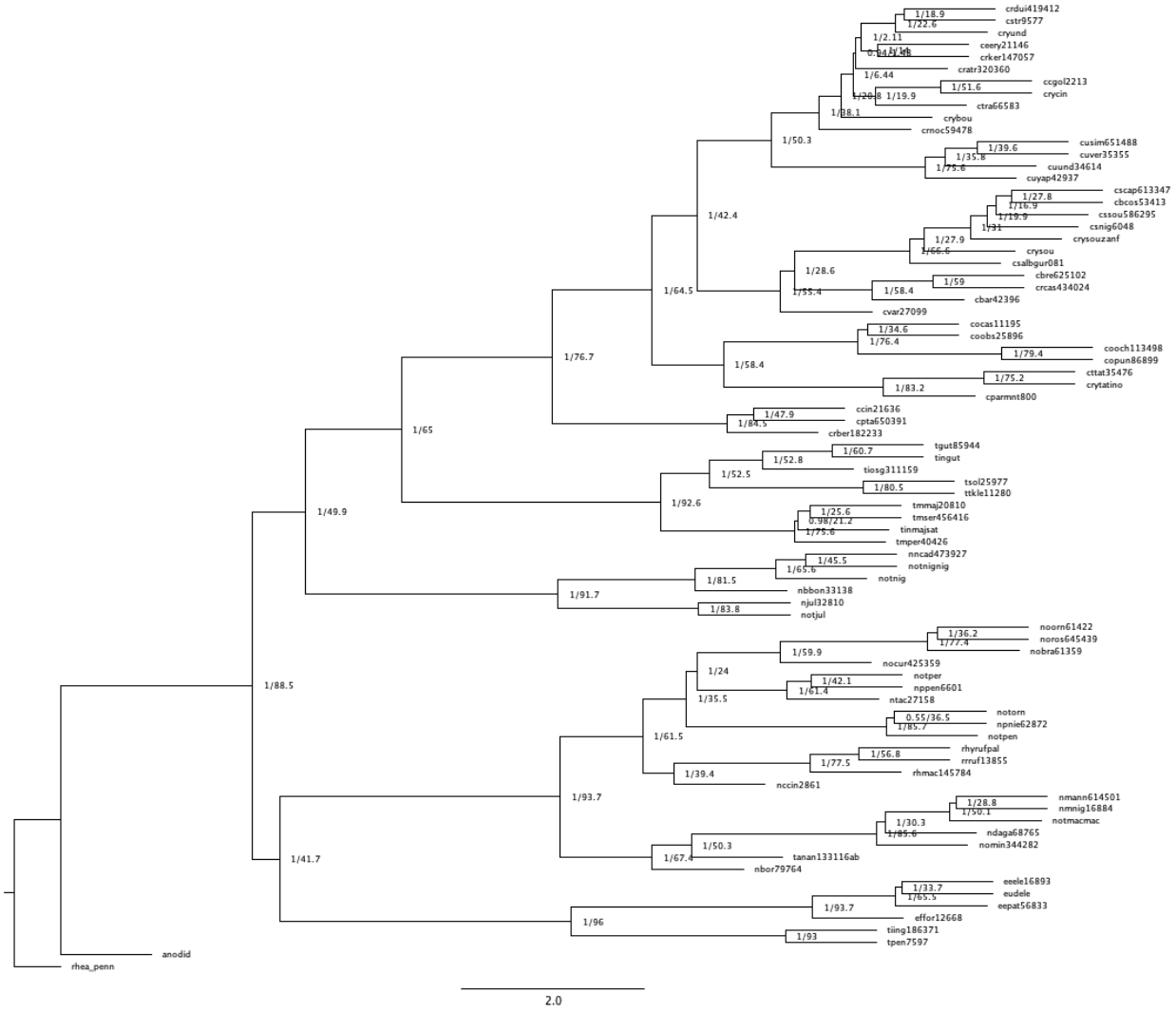


Figure S12: UCE300 100% astral phylogeny

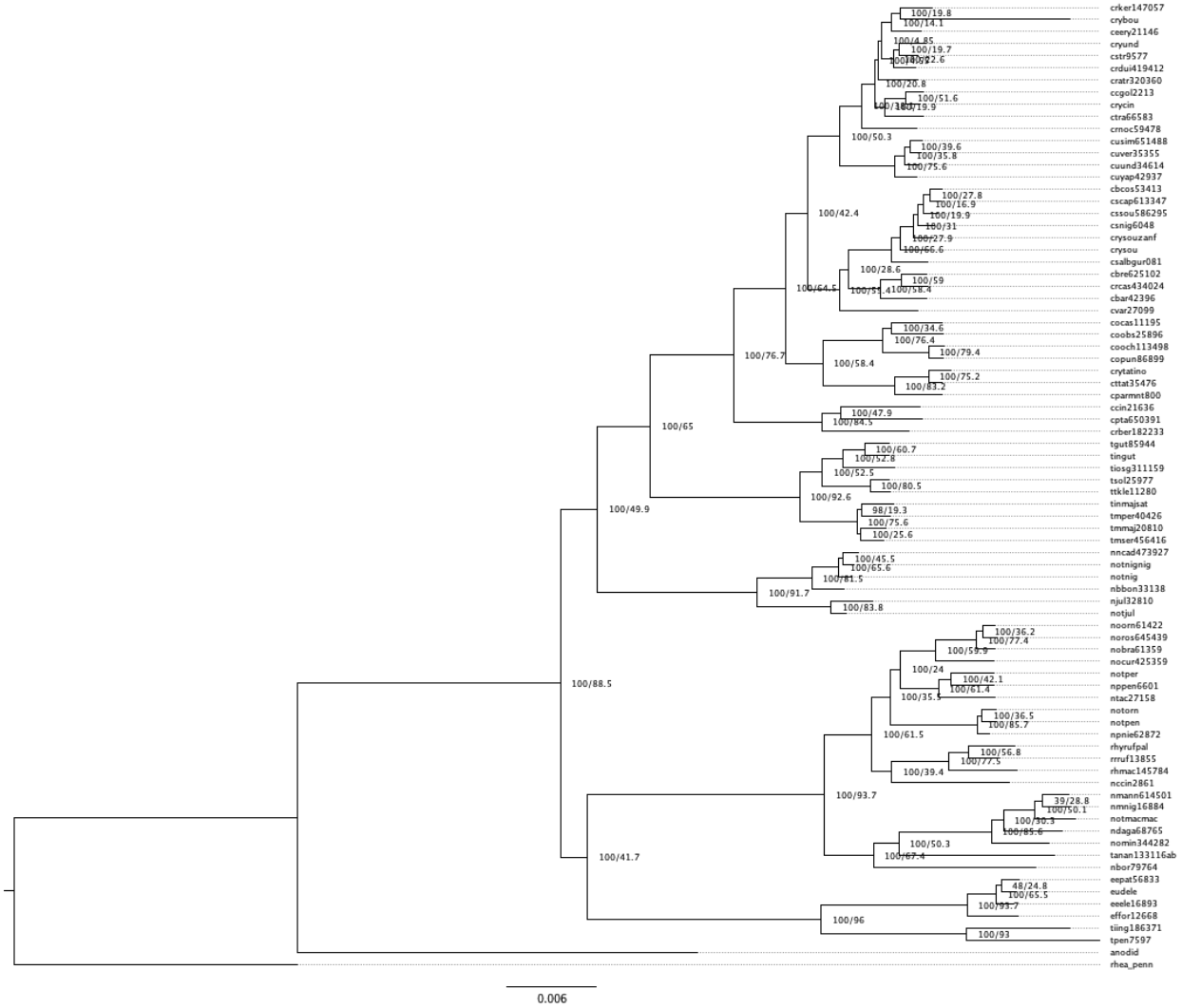


Figure S13: UCE300 100% iqtree phylogeny

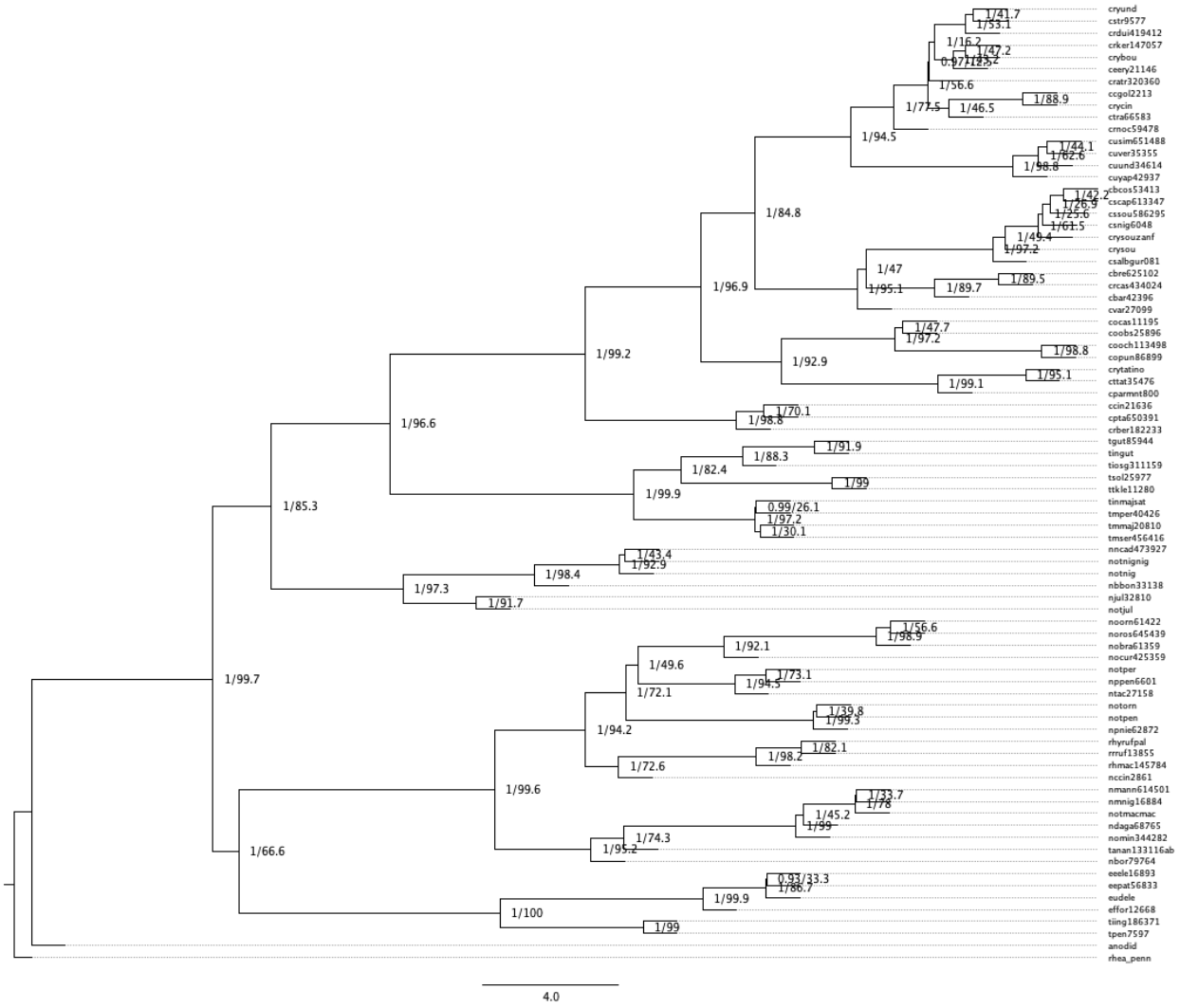


Figure S14: UCE1000 autosomes 75% astral phylogeny

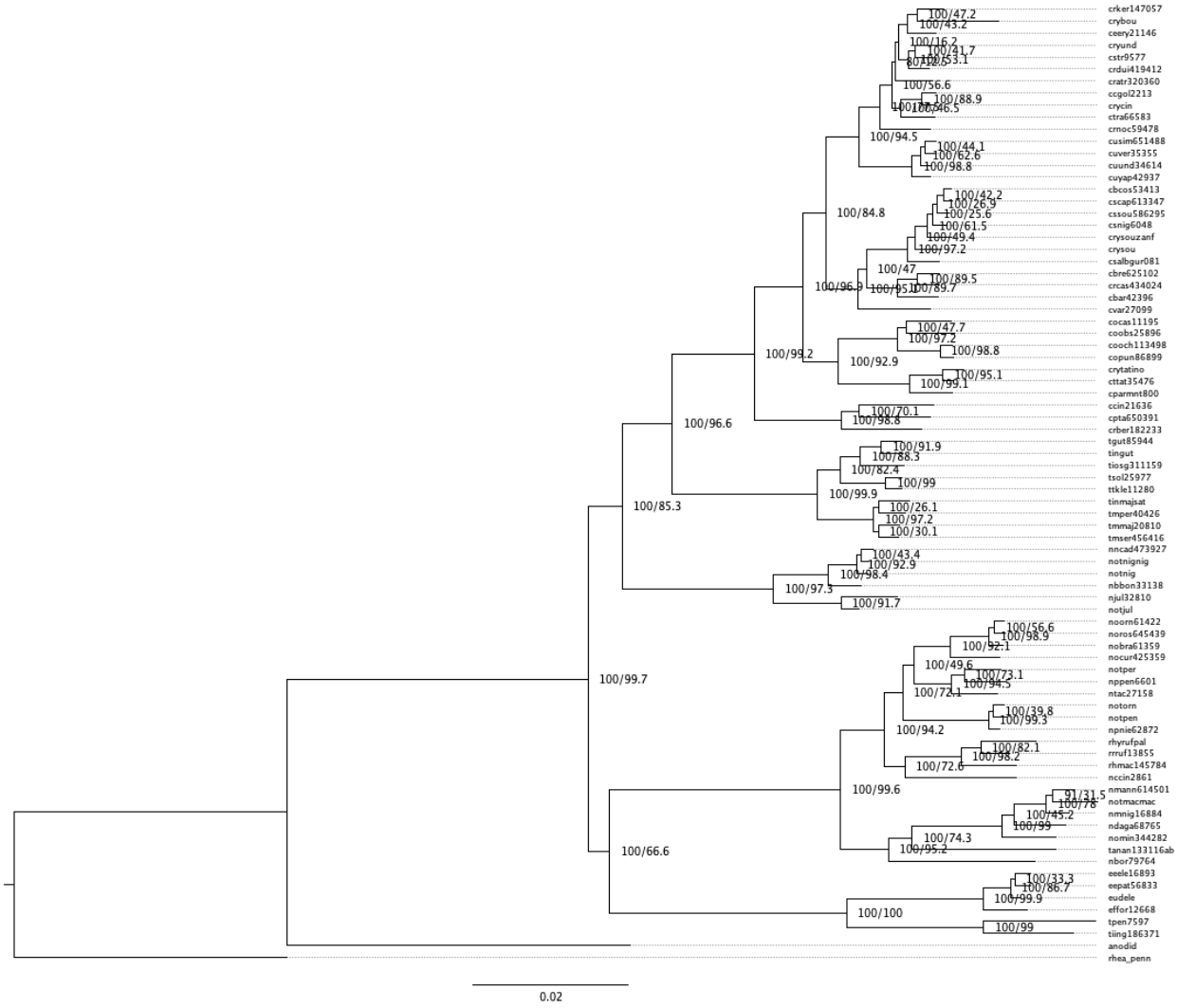


Figure S15: UCE1000 autosomes 75% iqtree phylogeny

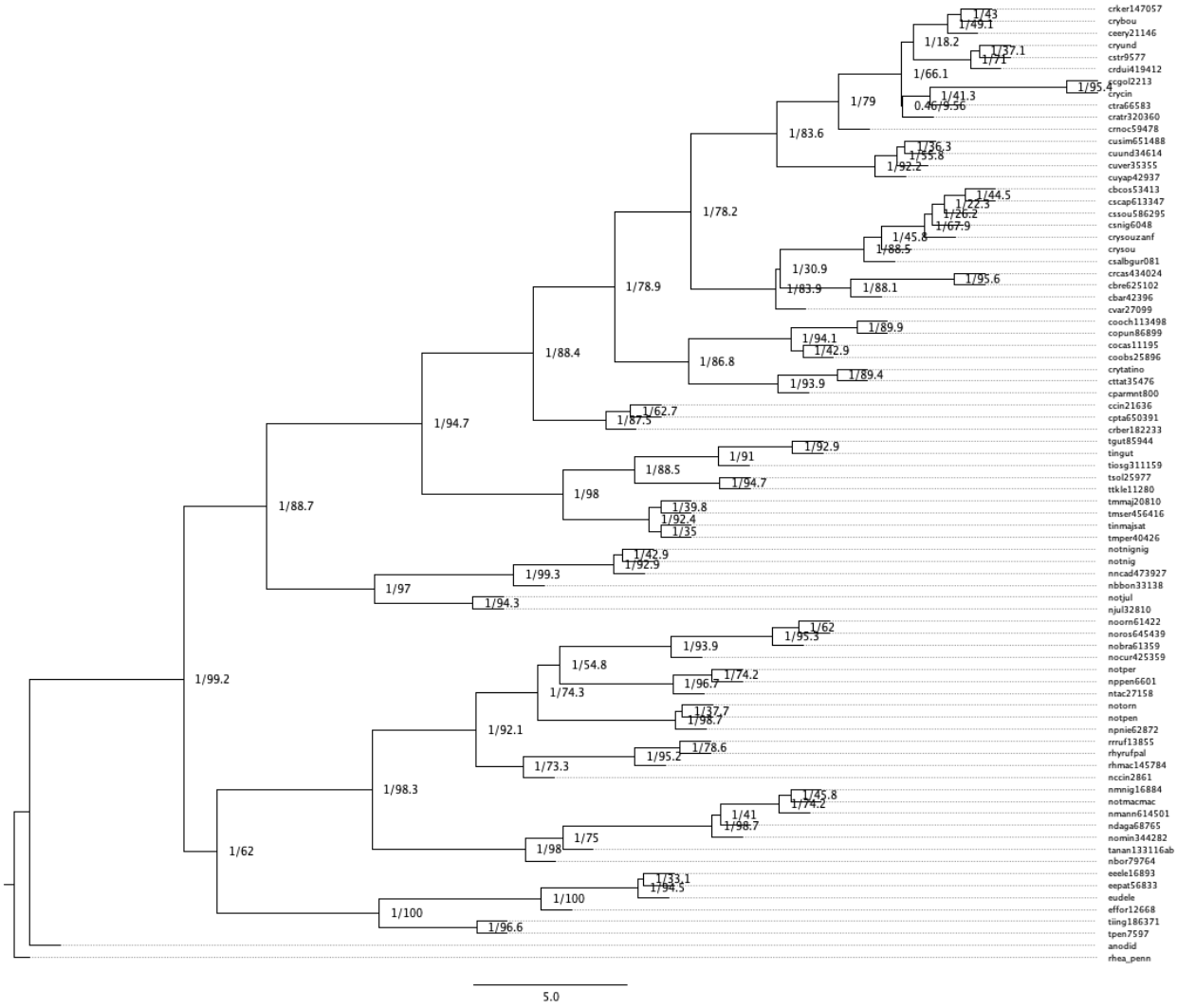


Figure S16: UCE1000 Z chromosome 75% astral phylogeny

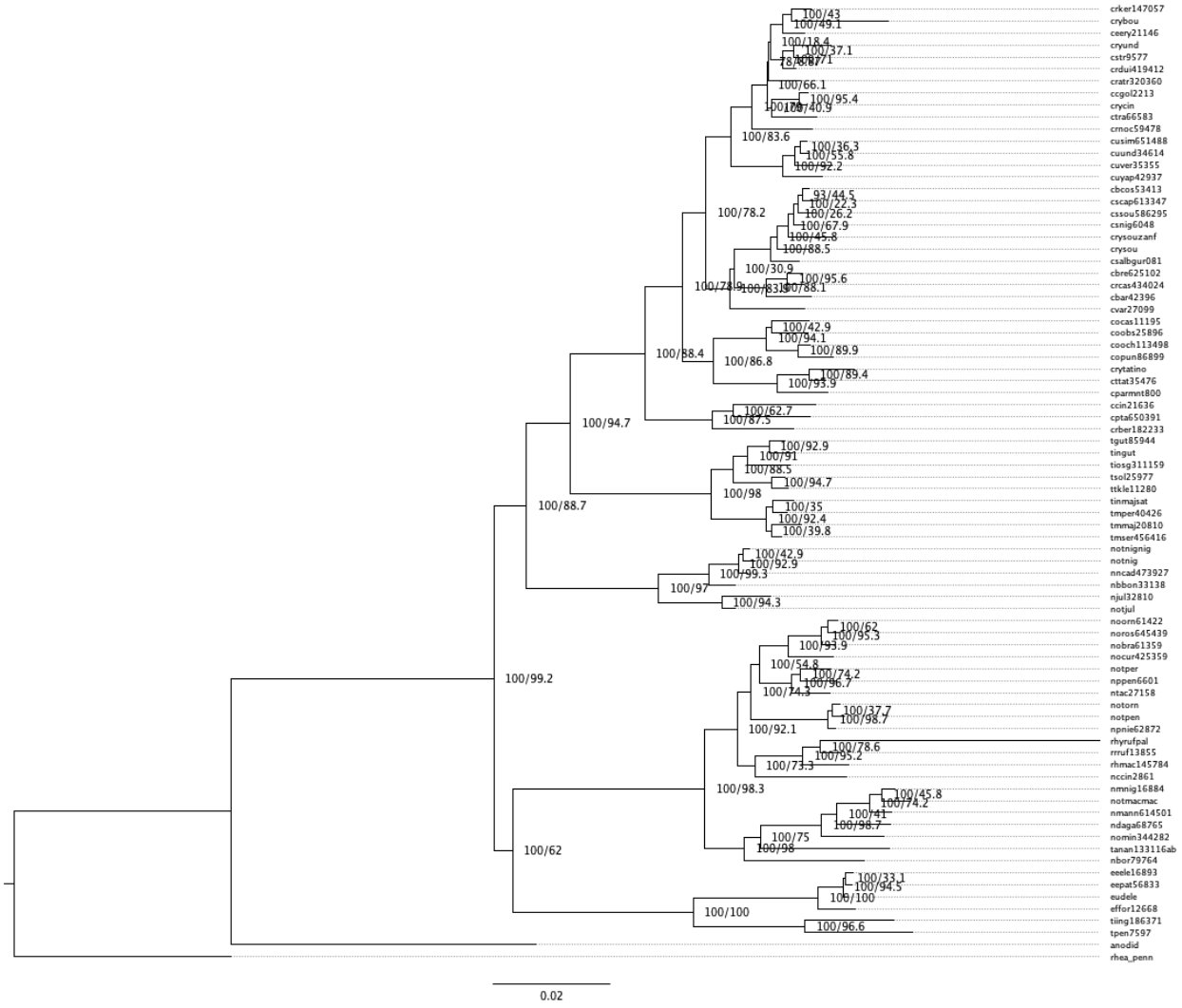


Figure S17: UCE1000 Z chromosome 75% iqtree phylogeny

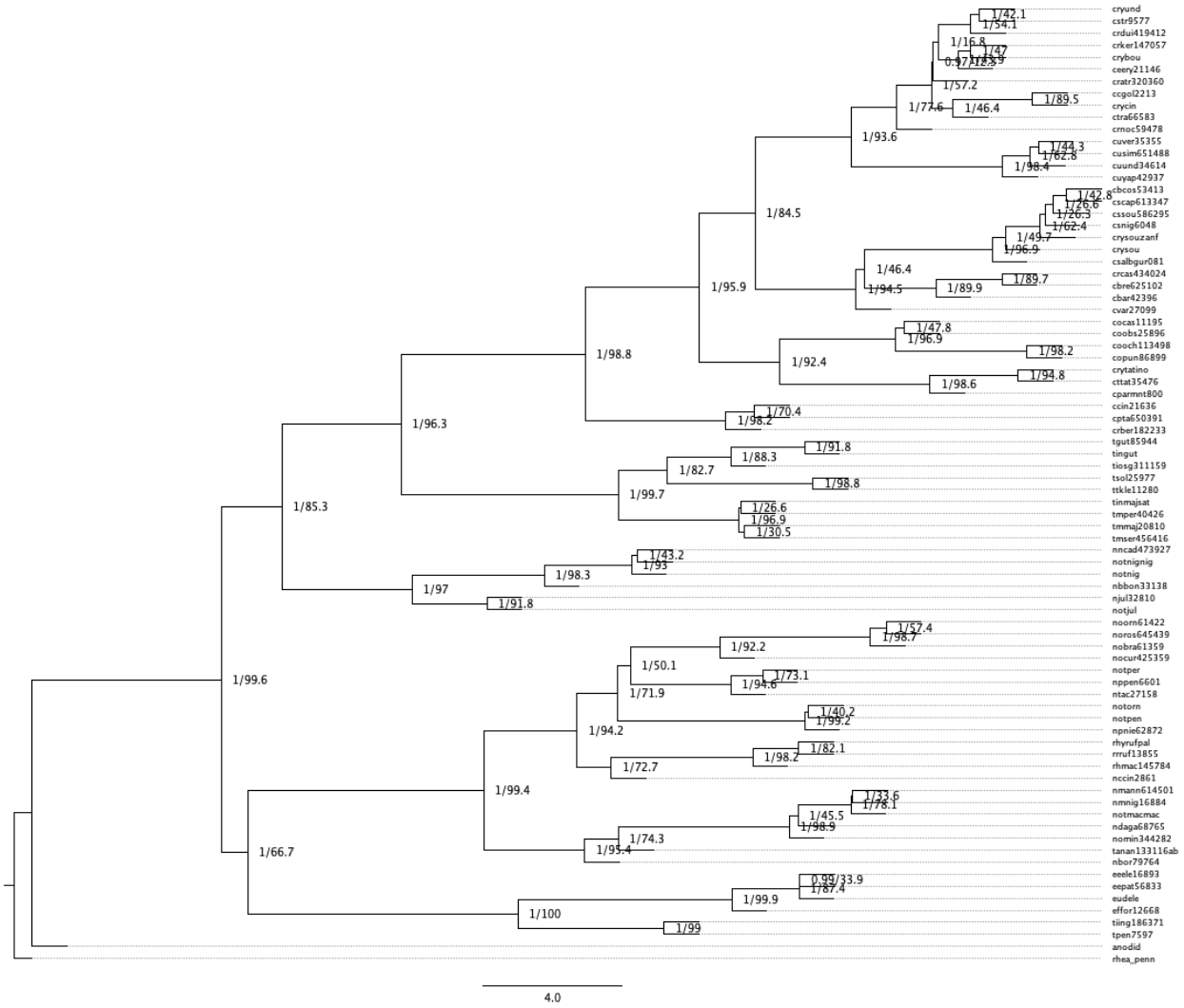


Figure S18: UCE1000 75% whole dataset astral phylogeny

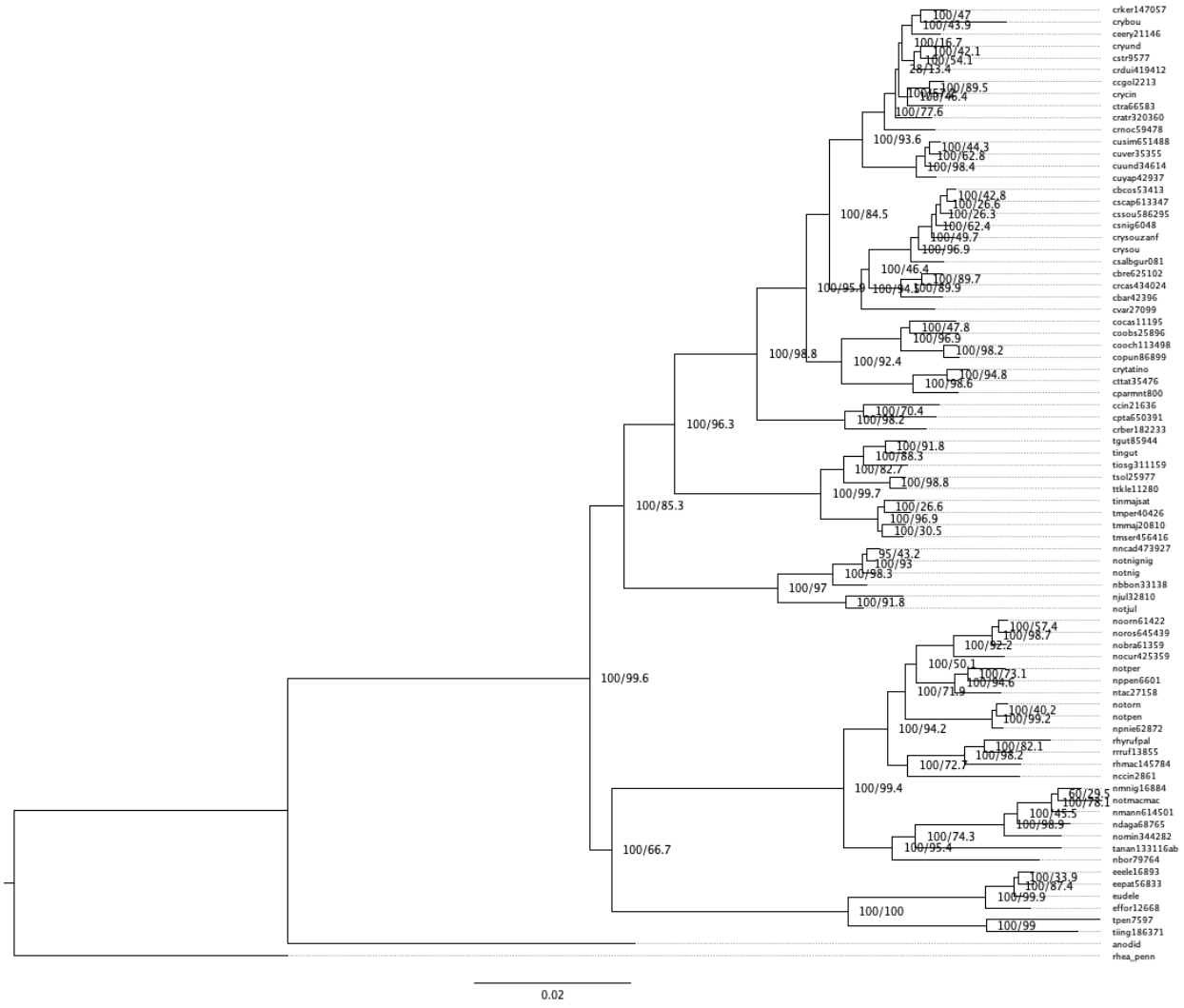


Figure S19: UCE1000 75% whole dataset iqtree phylogeny

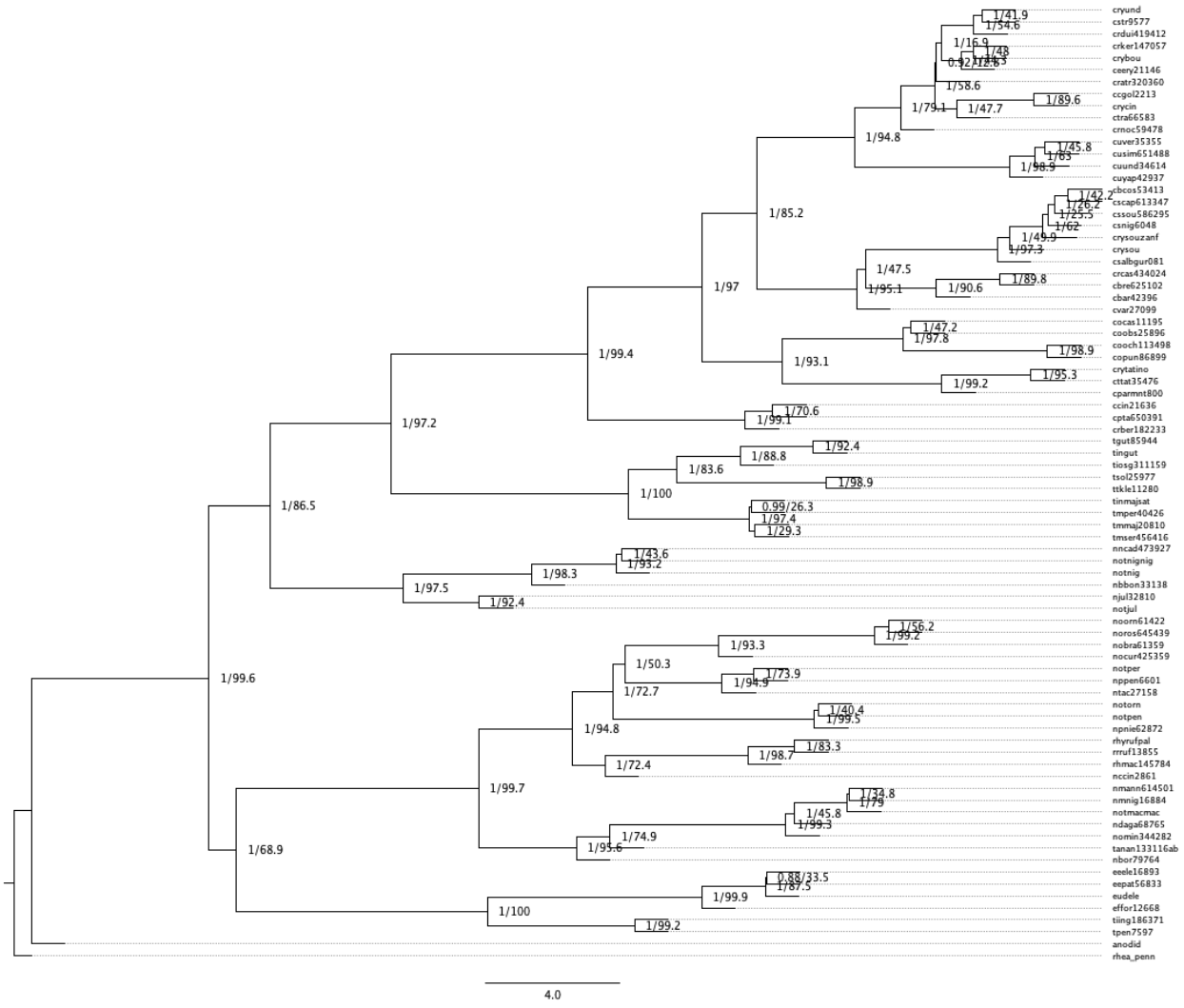


Figure S20: UCE1000 100% whole dataset astral phylogeny

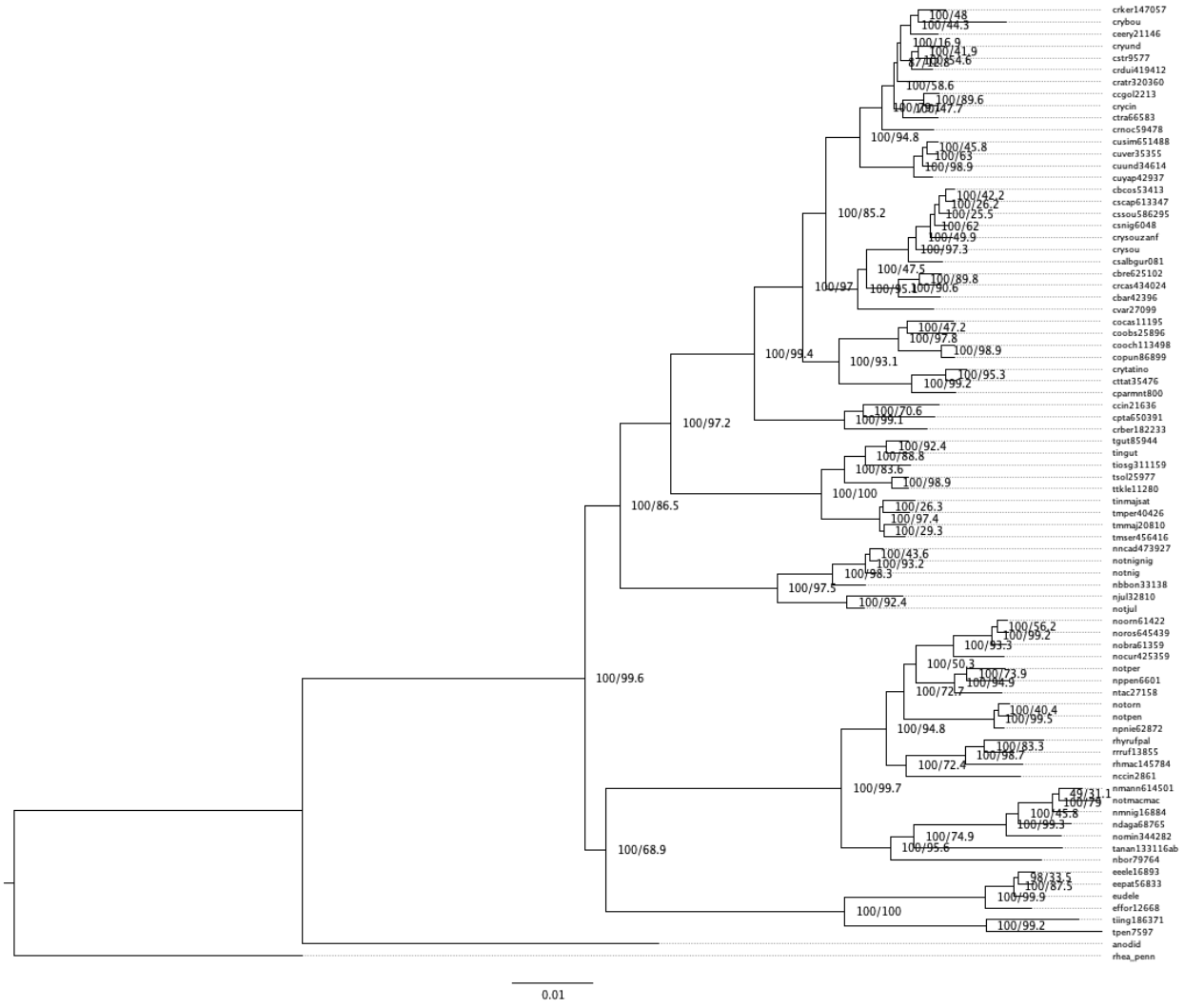


Figure S21: UCE1000 100% whole dataset iqtree phylogeny

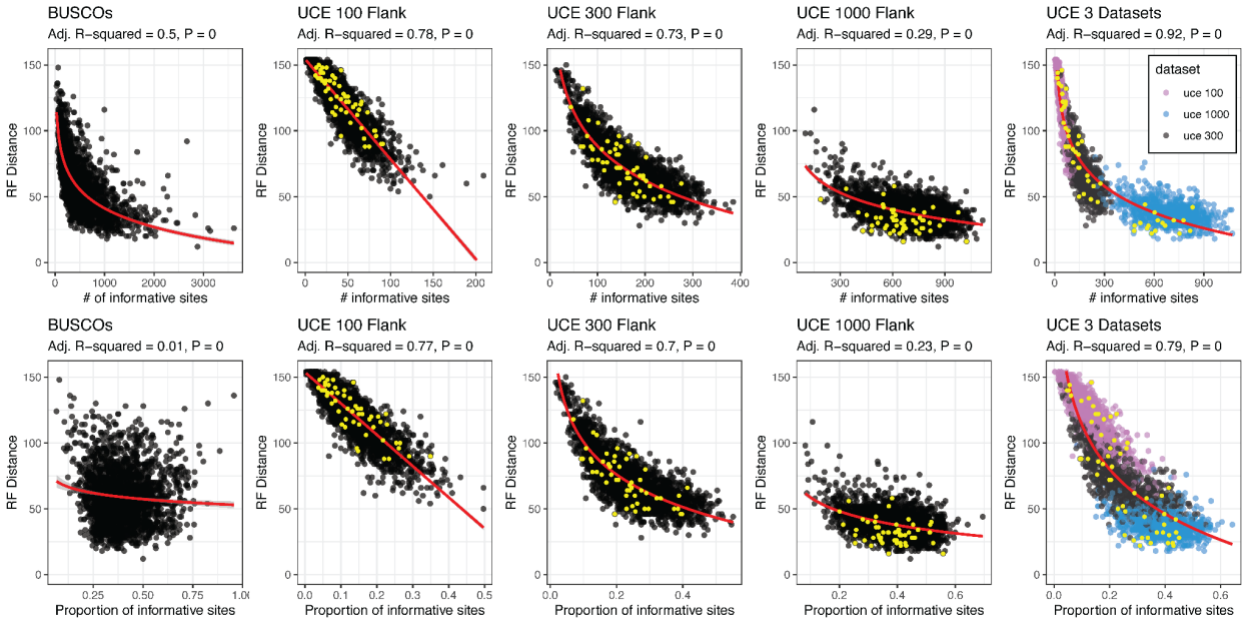


Figure S22: Covariation between information content and gene tree estimation error. Linear regression models exploring the relationship between information content and gene tree heterogeneity. The negative correlation in most instances suggests that alignments with less information content result in increased gene tree estimation error. UCEs that map to the Z-chromosome in each dataset are shown in yellow.

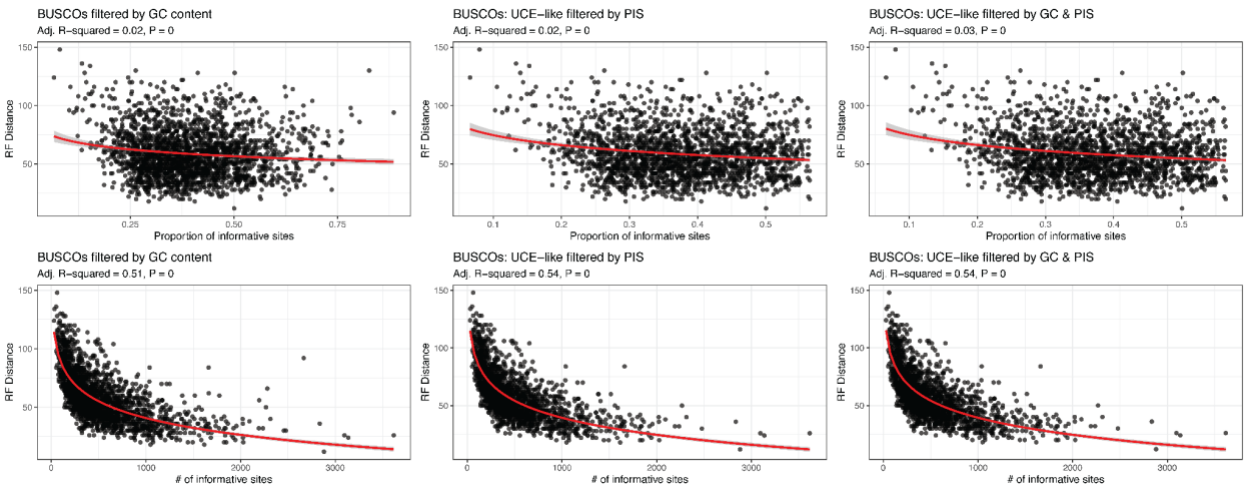


Figure S23: Linear regressions of information content and RF Distance for filtered BUSCOs. The BUSCO dataset was filtered three times. First, the extremes of GC content were

removed (BUSCOs filtered by GC content). Second, BUSCO's were filtered by the number and proportion of PIS, wherein BUSCOs with greater than two standard deviations away from the mean number and proportion of PIS were removed. Finally, the two BUSCO filters were combined.

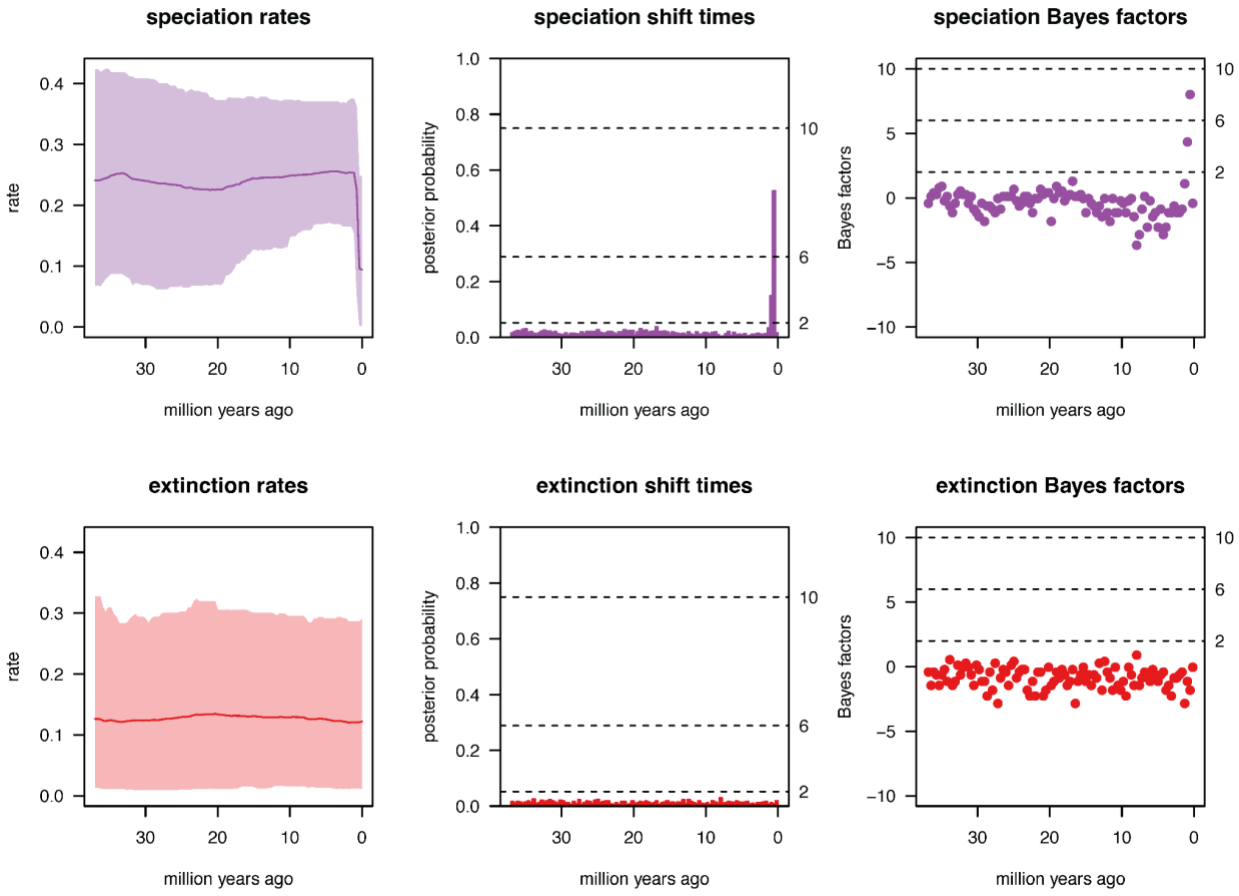


Figure S24: Speciation (purple) and extinction (red) times for Tinamidae along with posterior probabilities and Bayes factors for the presence of rate shifts.

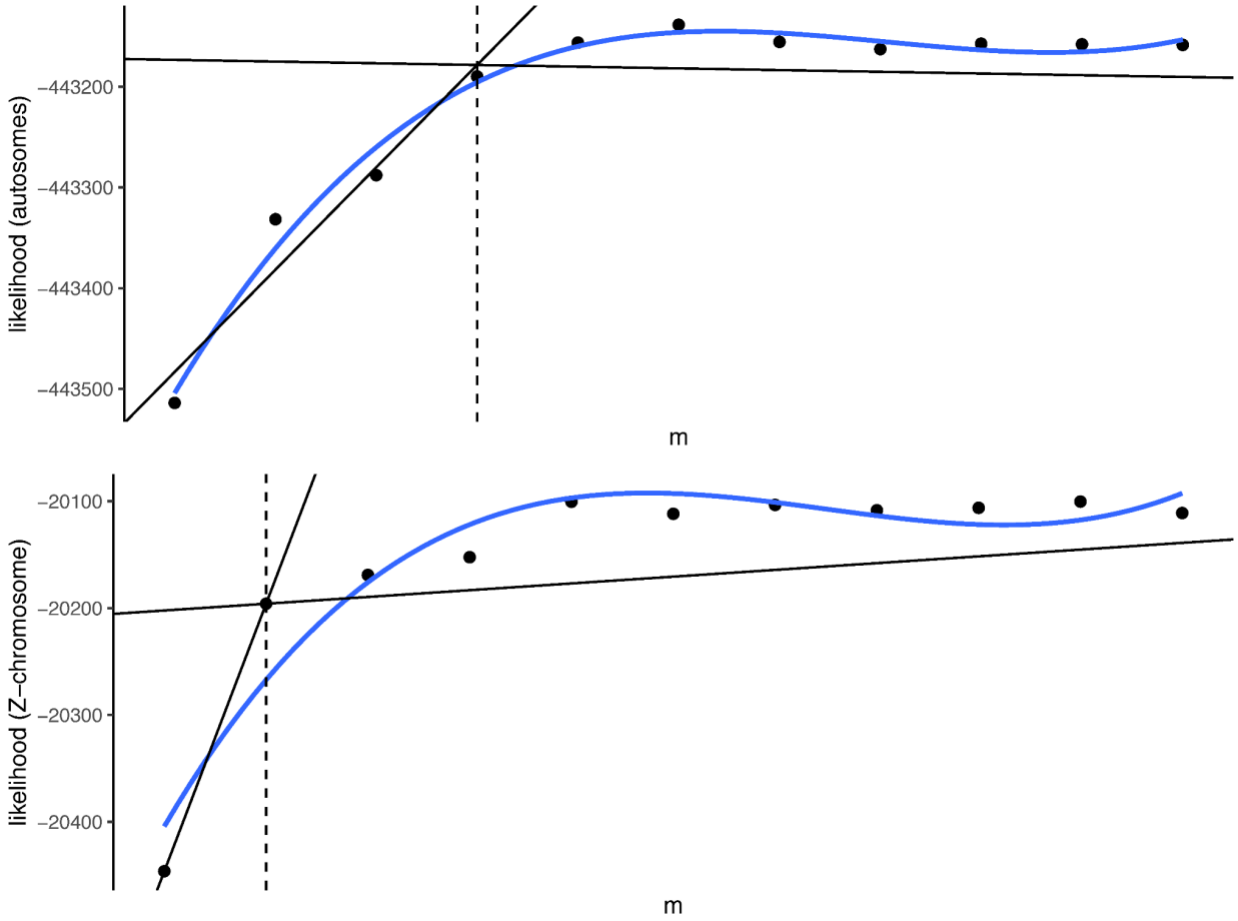
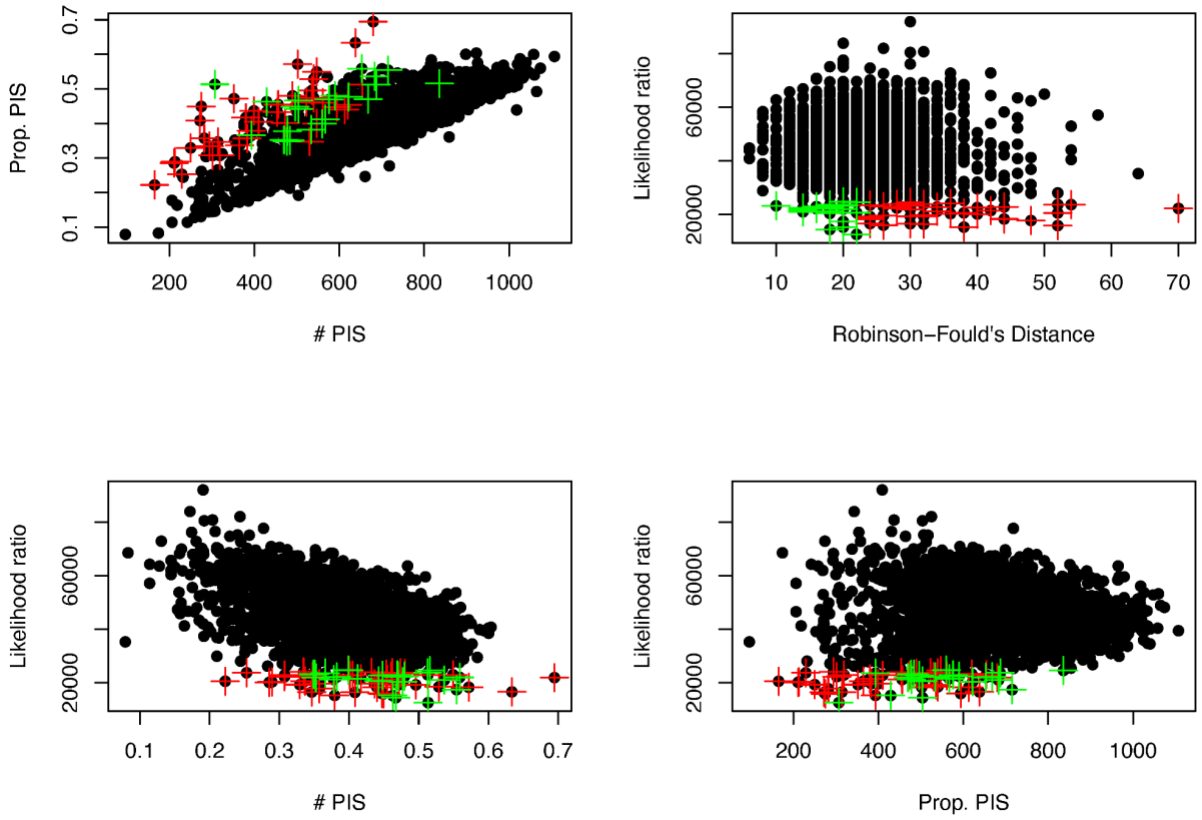


Figure S25: Breakpoint analysis for the best-fit m-value for the introgression analysis.



FigureS26: Results of likelihood ratio and RF filtering for divergence dating analysis. The plots show values for all autosomal genes before (black points) and after (red and gree crosses) filtering. Red crosses represent the alignment/gene trees two standard deviations below the mean likelihood ratio, and green crosses show those that are also below the mean for RF distance and were retained for use in divergence dating.

References:

Bouckaert R., Vaughan T.G., Barido-Sottani J., Duchêne S., Fourment M., Gavryushkina A., Heled J., Jones G., Kühnert D., De Maio N., Matschiner M., Mendes F.K., Müller N.F., Ogilvie H.A., du Plessis L., Popinga A., Rambaut A., Rasmussen D., Siveroni I., Suchard M.A., Wu C.-H., Xie D., Zhang C., Stadler T., Drummond A.J. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15:e1006650.

Douglas J., Zhang R., Bouckaert R. 2021. Adaptive dating and fast proposals: Revisiting the phylogenetic relaxed clock model. PLoS Comput. Biol. 17:e1008322.

Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.

Edwards S.V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R., Leaché A.D., Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94:447–462.

Gatesy J., Springer M.S. 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. 80:231–266.

Gavryushkina A., Welch D., Stadler T., Drummond A. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. PLoS Comput. Biol. 10:e1003919.

Hoff K.J., Lomsadze A., Borodovsky M., Stanke M. 2019. Whole-Genome Annotation with BRAKER. Methods Mol. Biol. 1962:65–95.

Höhna S., May M.R., Moore B.R. 2016. TESS: an R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. Bioinformatics. 32:789–791.

Maddison W.P. 1997. Gene Trees in Species Trees. Syst. Biol. 46:523–536.

Musher L.J., Catanach T.A., Valqui T., Brumfield R.T., Aleixo A., Johnson K.P., Weckstein J. 2024. Whole-genome phylogenomics of the tinamous (Aves: Tinamidae): comparing gene tree estimation error between BUSCOs and UCEs illuminates rapid divergence with introgression. bioRxiv.

Musher L.J., Ferreira M., Auerbach A.L., Cracraft J. 2019. Why is Amazonia a “source”of biodiversity? Climate-mediated dispersal and synchronous speciation across the Andes in an avian group (Tityrinae). Proc. Roy. Soc. B.

Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. 67:901–904.

Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. 2015. BUSCO:

assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31:3210–3212.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics. 105:437–460.

Zhang R., Drummond A. 2020. Improving the performance of Bayesian phylogenetic inference under relaxed clock models. BMC Evol. Biol. 20:54.