Methods

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

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.

Results

Sample assembly details

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.

Supplemental Tables and Figures:

Dataset	Model	Adj. R-squared	slope	P	AIC
BUSCOs	linear (#PIS)	**0.02163**	**-0.0035908**	**<0.0001**	**21179**
	logarithmic (#PIS)	0.02177	-3.6060	<0.0001	21179
	linear (%PIS)	**0.02268**	**24.530**	**<0.0001**	**21177**
	logarithmic (%PIS)	0.01782	7.470	<0.0001	21189
UCE100Flank	**linear (#PIS)**	**0.735**	**-0.599775**	**<0.0001**	**21684**
	logarithmic (#PIS)	0.7001	-23.4444	<0.0001	22041
	linear (%PIS)	**-210.7954**	**0.7493**	**<0.0001**	**21525**
	logarithmic (%PIS)	-23.4846	0.6919	<0.0001	22119
UCE300Flank	linear (#PIS)	0.4308	-0.096988	<0.0001	21345
	logarithmic (#PIS)	**0.6155**	**-20.2605**	**<0.0001**	**20216**
	linear (%PIS)	0.4787	-86.1419	<0.0001	21092
	logarithmic (%PIS)	**0.6373**	**-21.7278**	**<0.0001**	**20047**
UCE1000Flank	linear (#PIS)	0.00254	-0.0015327	<0.0001	18893
	logarithmic (#PIS)	**0.01074**	**-2.7845**	**<0.0001**	**18870**
	linear (%PIS)	**0.02015**	**-12.3668**	**<0.0001**	**18843**
	logarithmic (%PIS)	0.04791	-8.2546	<0.0001	19711
UCE Combined	linear (#PIS)	0.508	-0.0584207	<0.0001	80245
	logarithmic (#PIS)	0.7858	-21.7212	<0.0001	69350
	linear (%PIS)	0.6987	-1213.8019	<0.0001	76049
	logarithmic (%PIS)	**0.8623**	**-23.9949**	**<0.0001**	**74480**

Table S2: 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 with bold text indicate the best model for each paired model test.
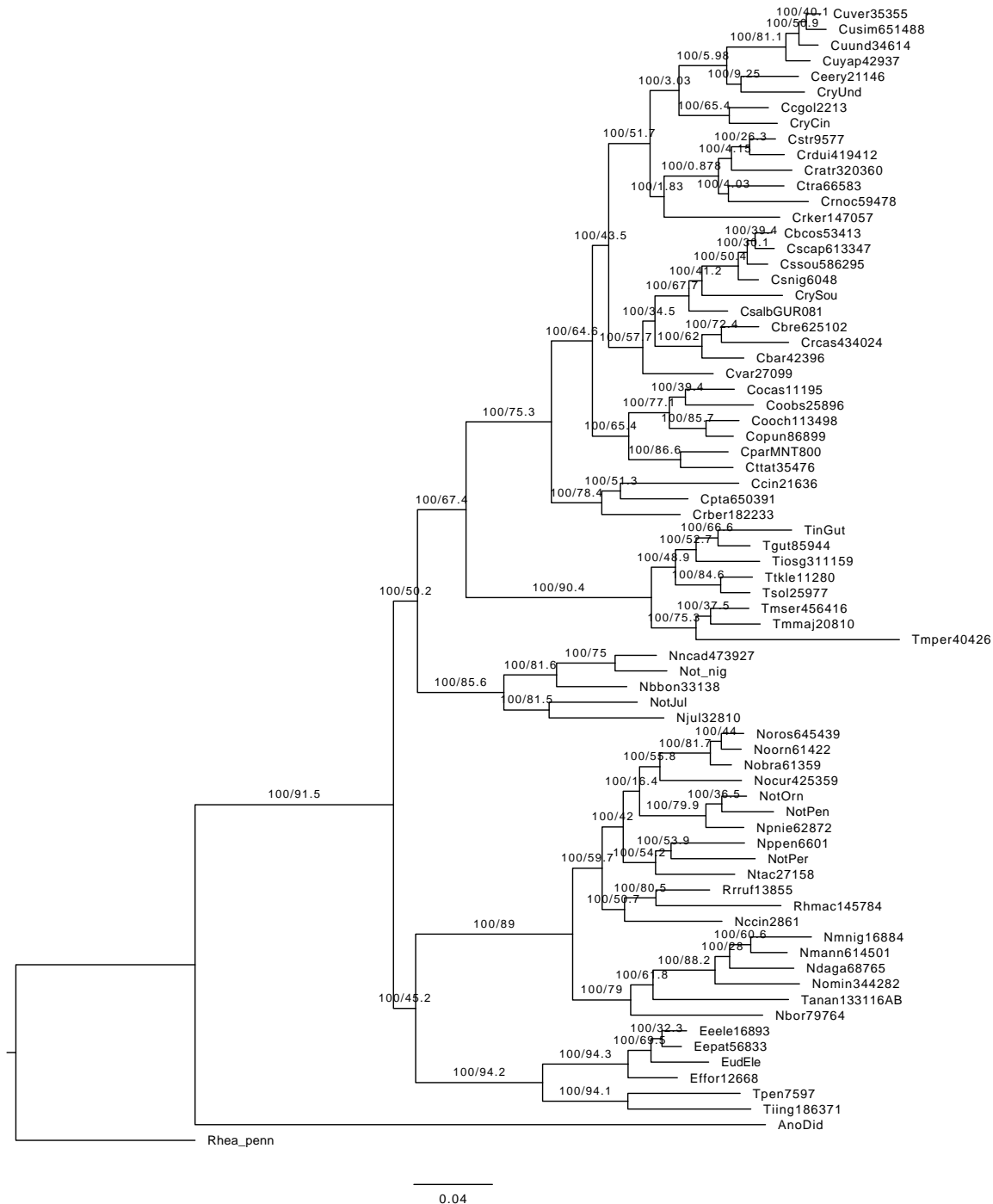

Figure S1: CDS Concatenated phylogeny. Node support is shown at each internal branch, including bootstrap value (first number) and gene concordance factor (second number).
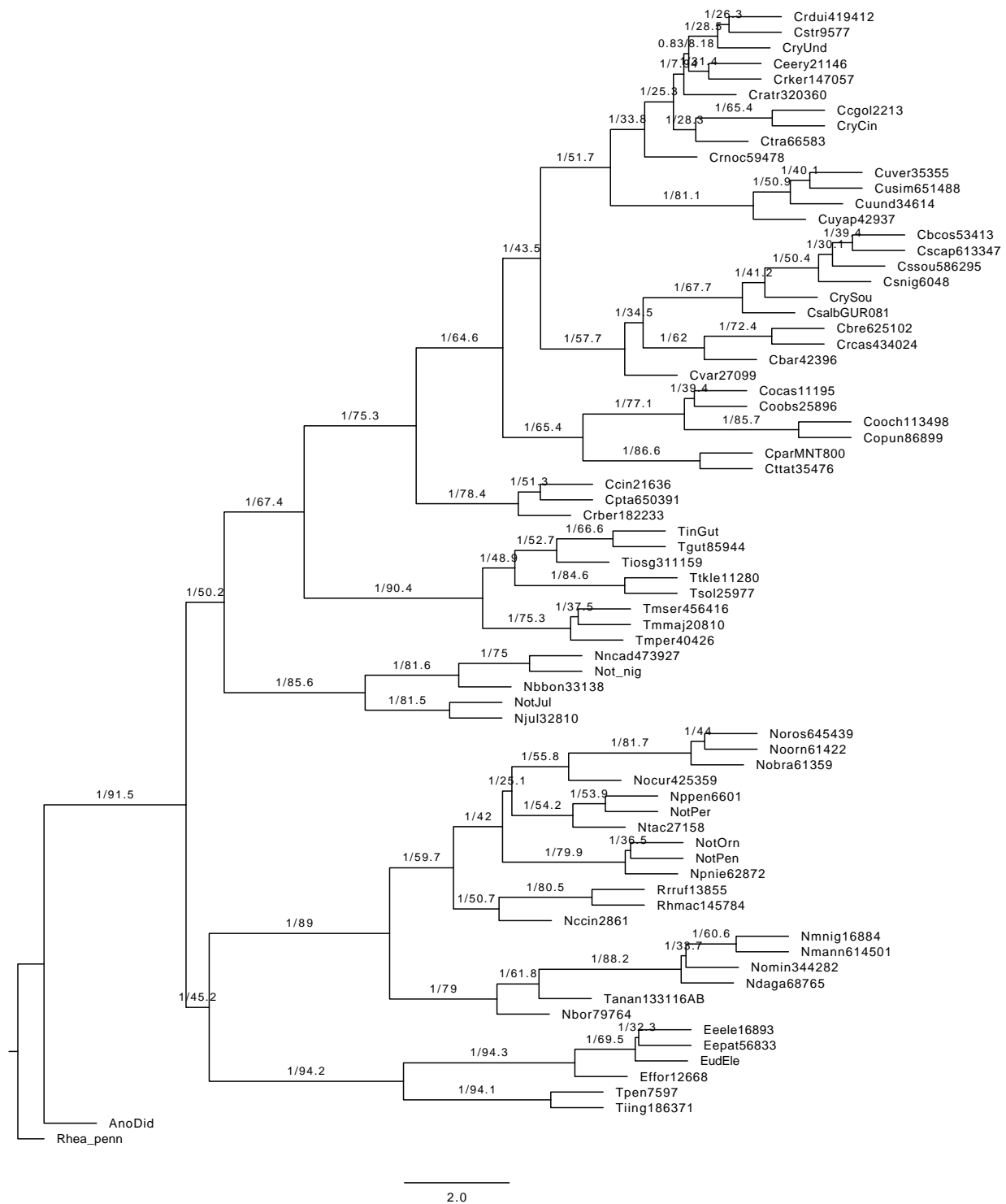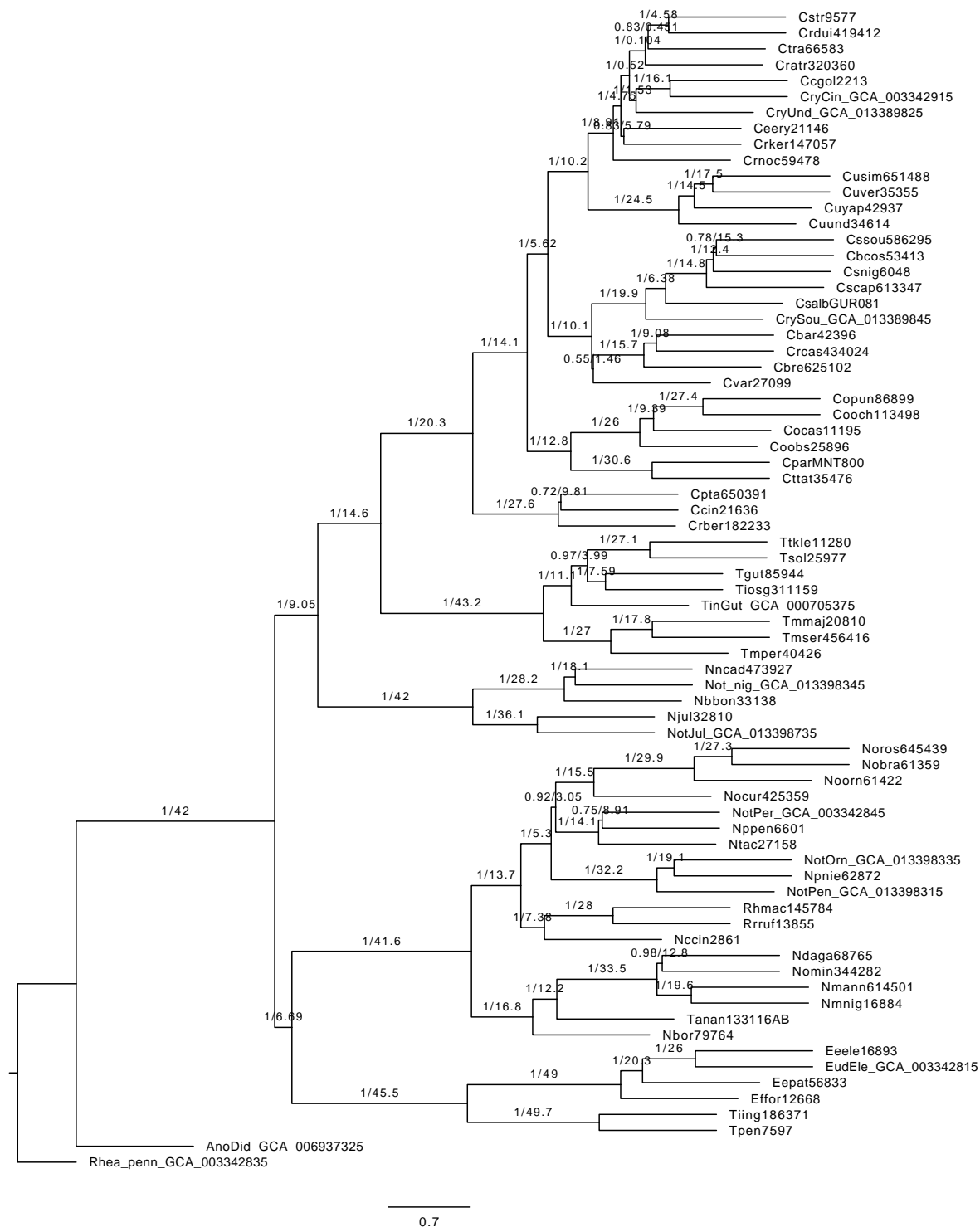
Figure S2: CDS MSC phylogeny. Node support is shown at each internal branch, including posterior probability (first number) and gene concordance factor (second number).
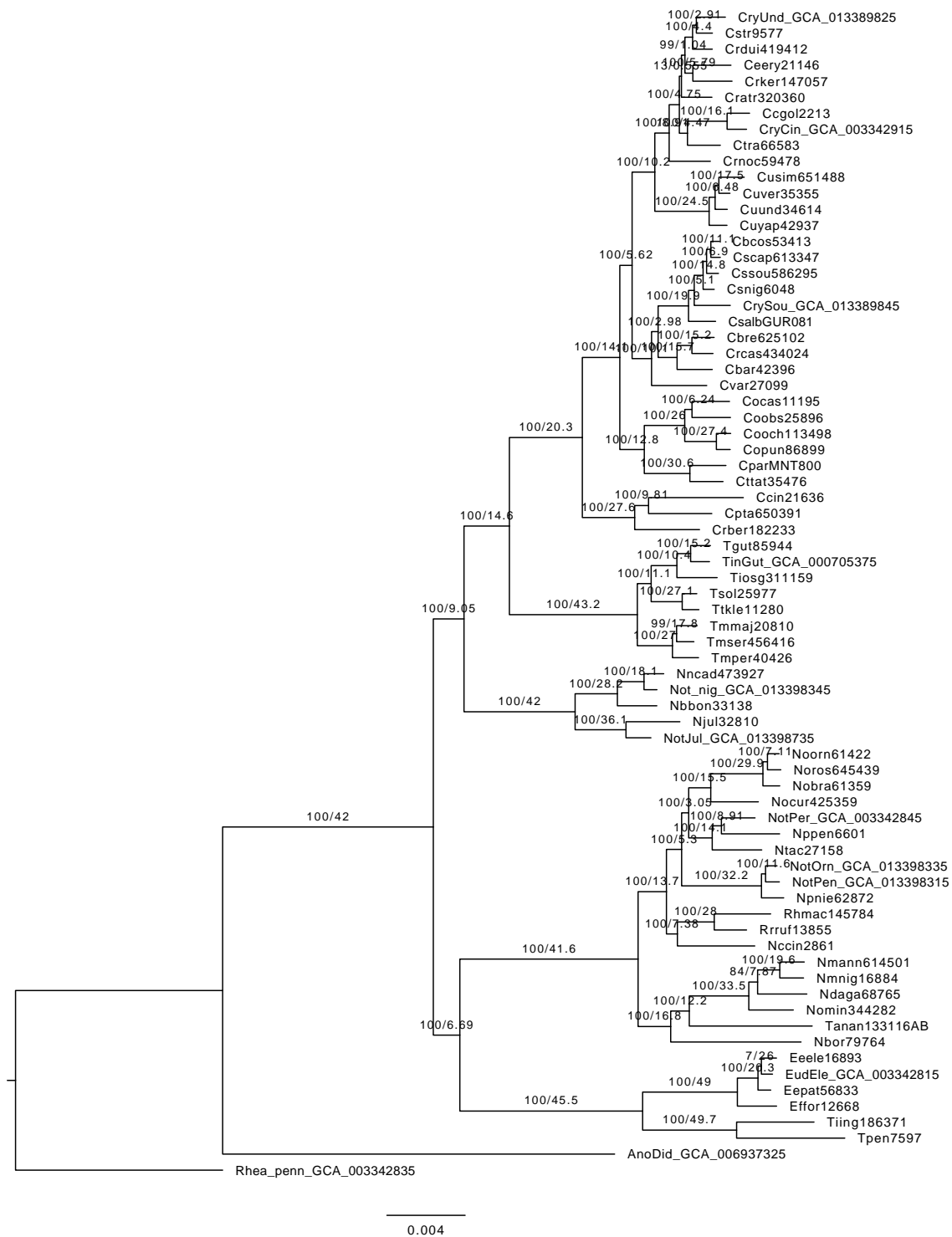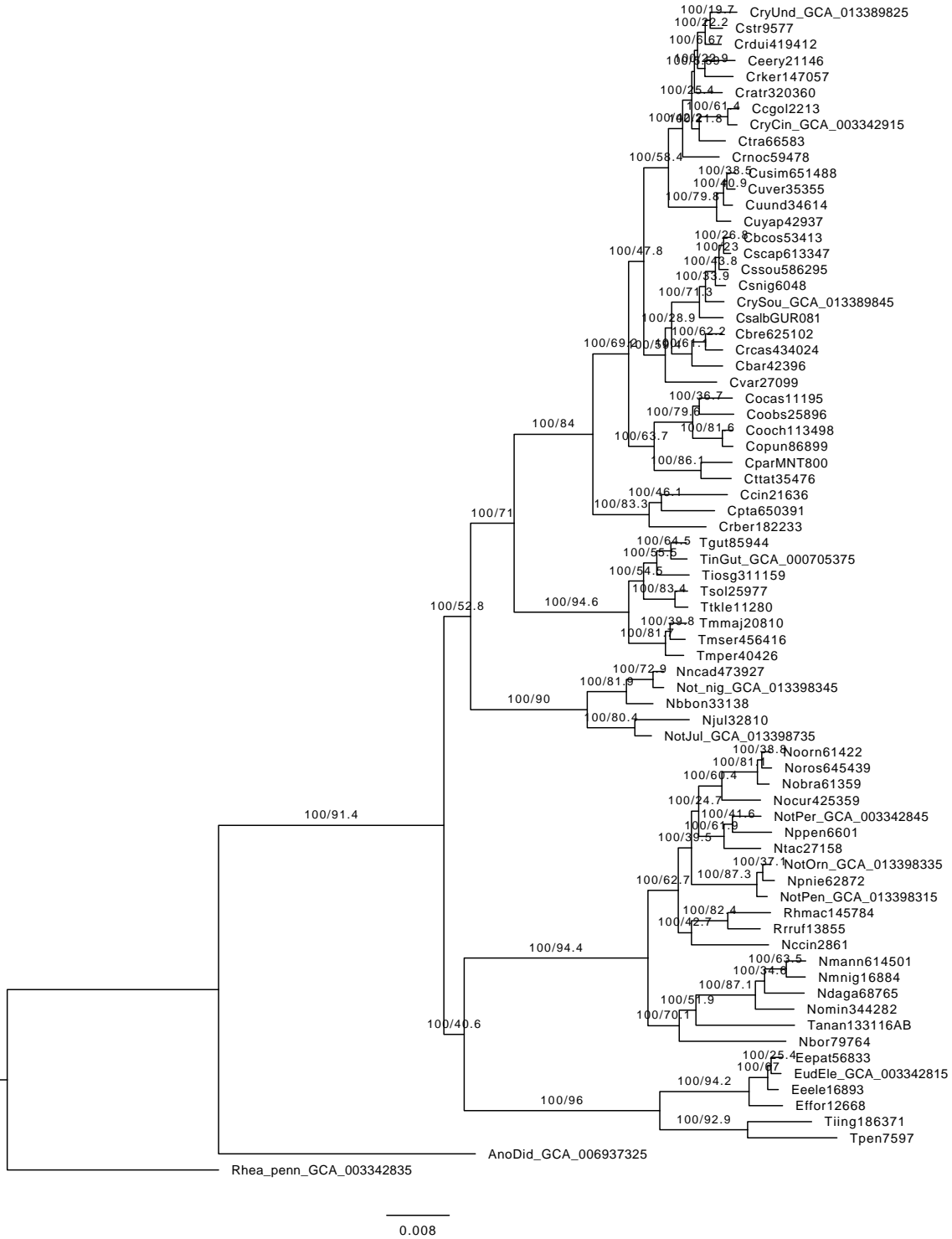
Figure S3: UCE100Flank Concatenated phylogeny. Node support is shown at each internal branch, including bootstrap value (first number) and gene concordance factor (second number).
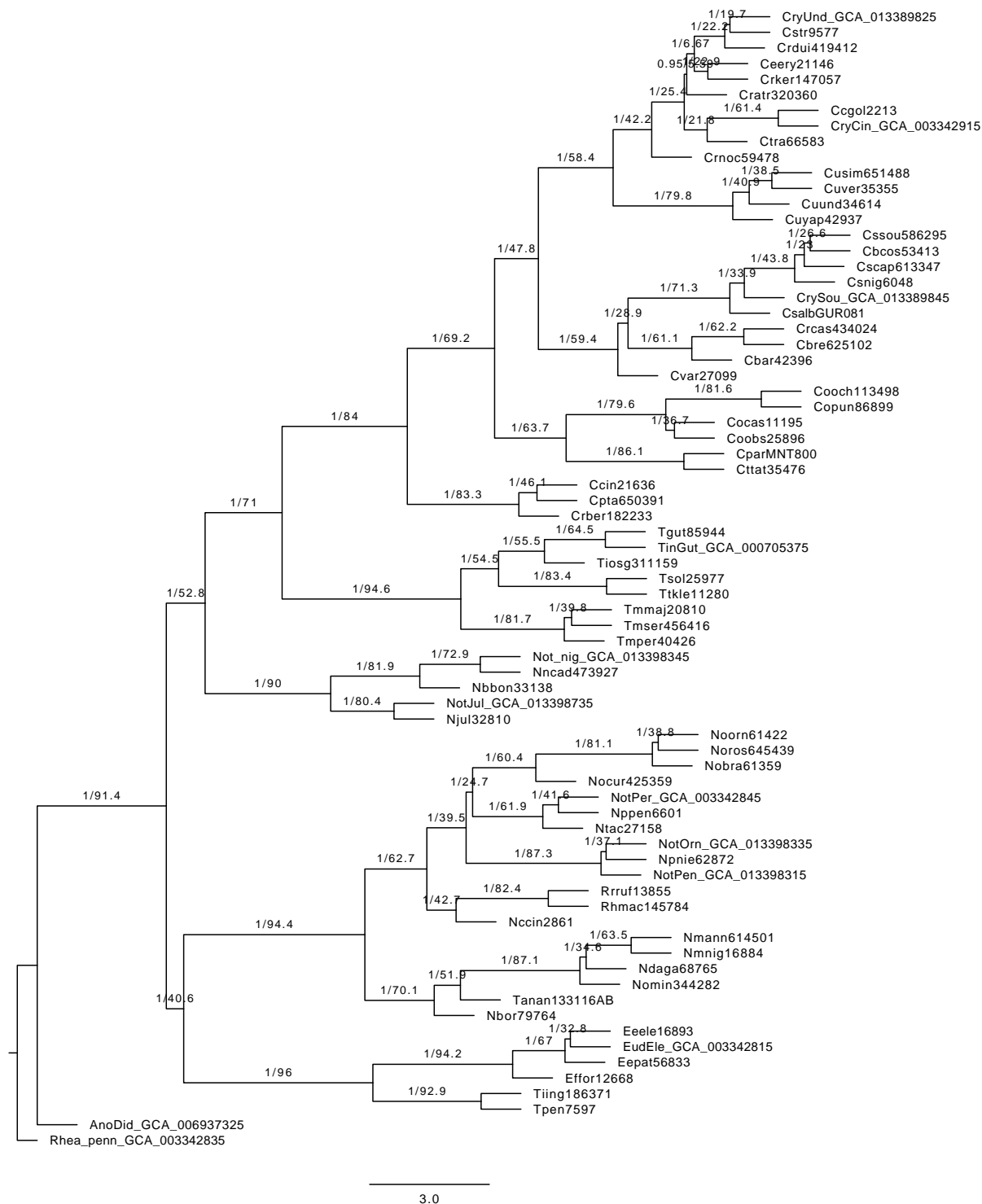
Figure S4: UCE100Flank MSC phylogeny. Node support is shown at each internal branch, including posterior probability (first number) and gene concordance factor (second number).
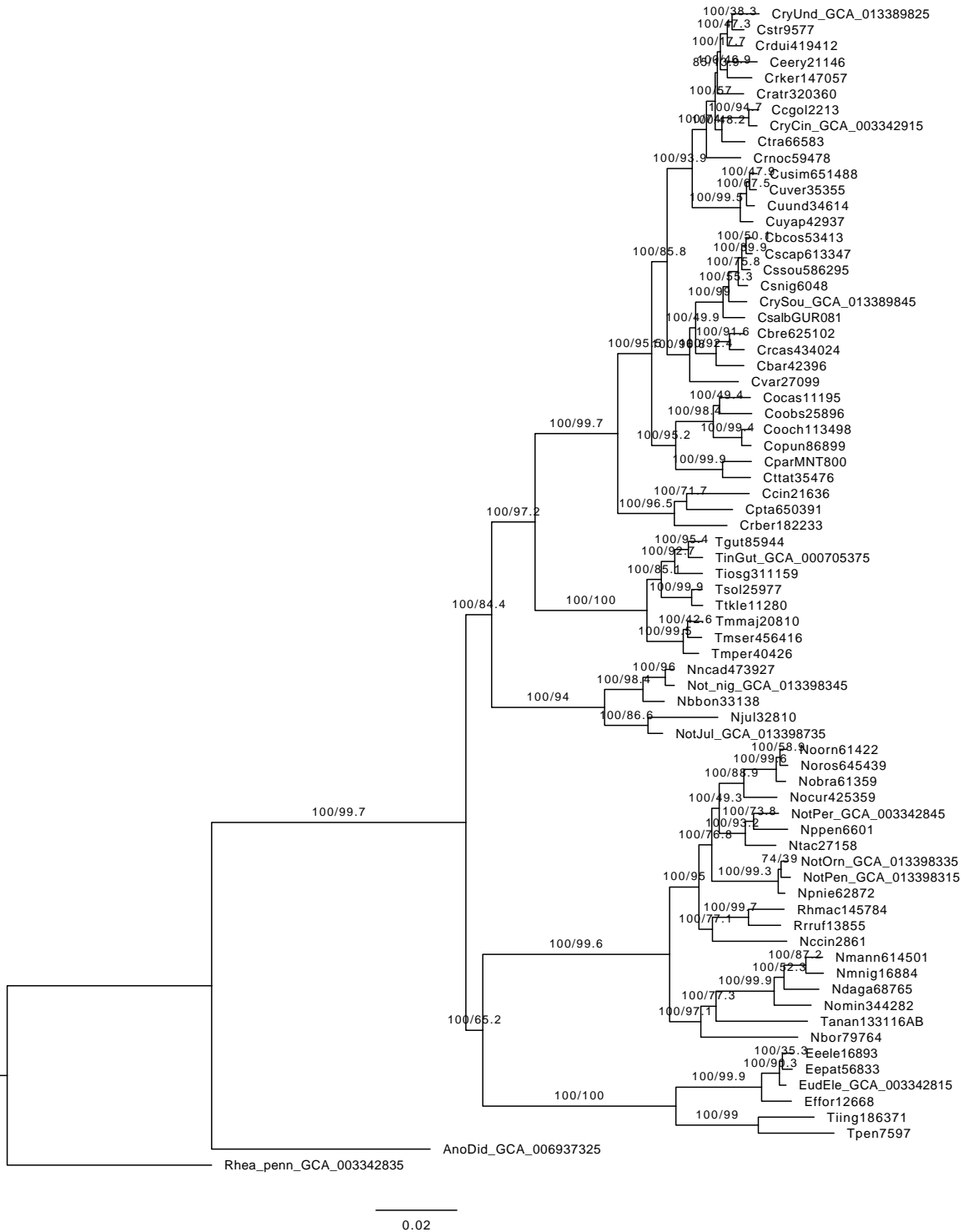
Figure S5: UCE300Flank Concatenated phylogeny. Node support is shown at each internal branch, including bootstrap value (first number) and gene concordance factor (second number).
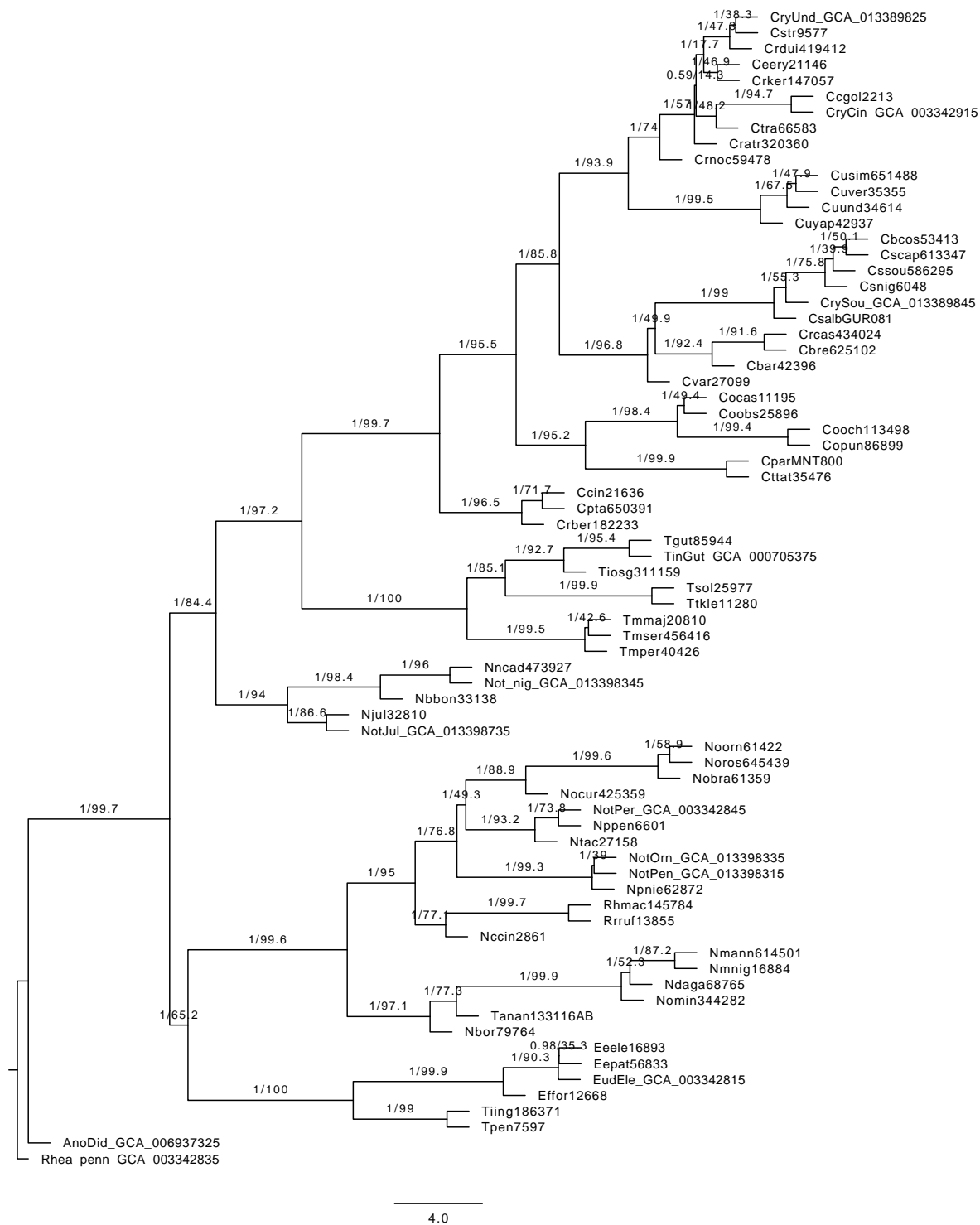
Figure S6: UCE300Flank MSC phylogeny. Node support is shown at each internal branch, including posterior probability (first number) and gene concordance factor (second number).

Figure S7: UCE1000Flank concatenated phylogeny. Node support is shown at each internal branch, including bootstrap value (first number) and gene concordance factor (second number).

Figure S8: UCE1000Flank MSC phylogeny. Node support is shown at each internal branch, including posterior probability (first number) and gene concordance factor (second number).
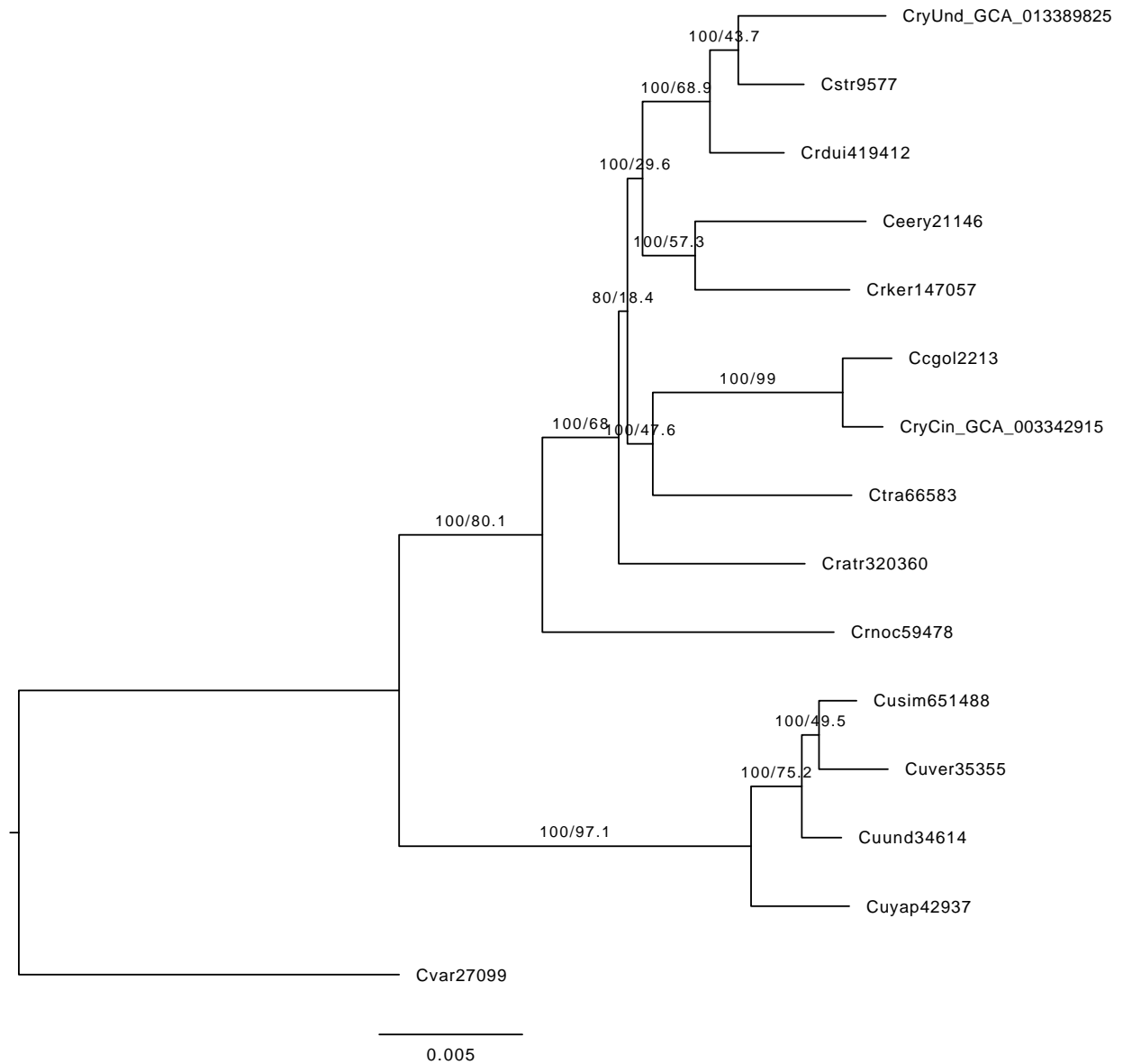
Figure S9: Z-chromosome concatenated phylogeny (*Crypturellus* Clade A only). Node support is shown at each internal branch, including bootstrap value (first number) and gene concordance factor (second number).
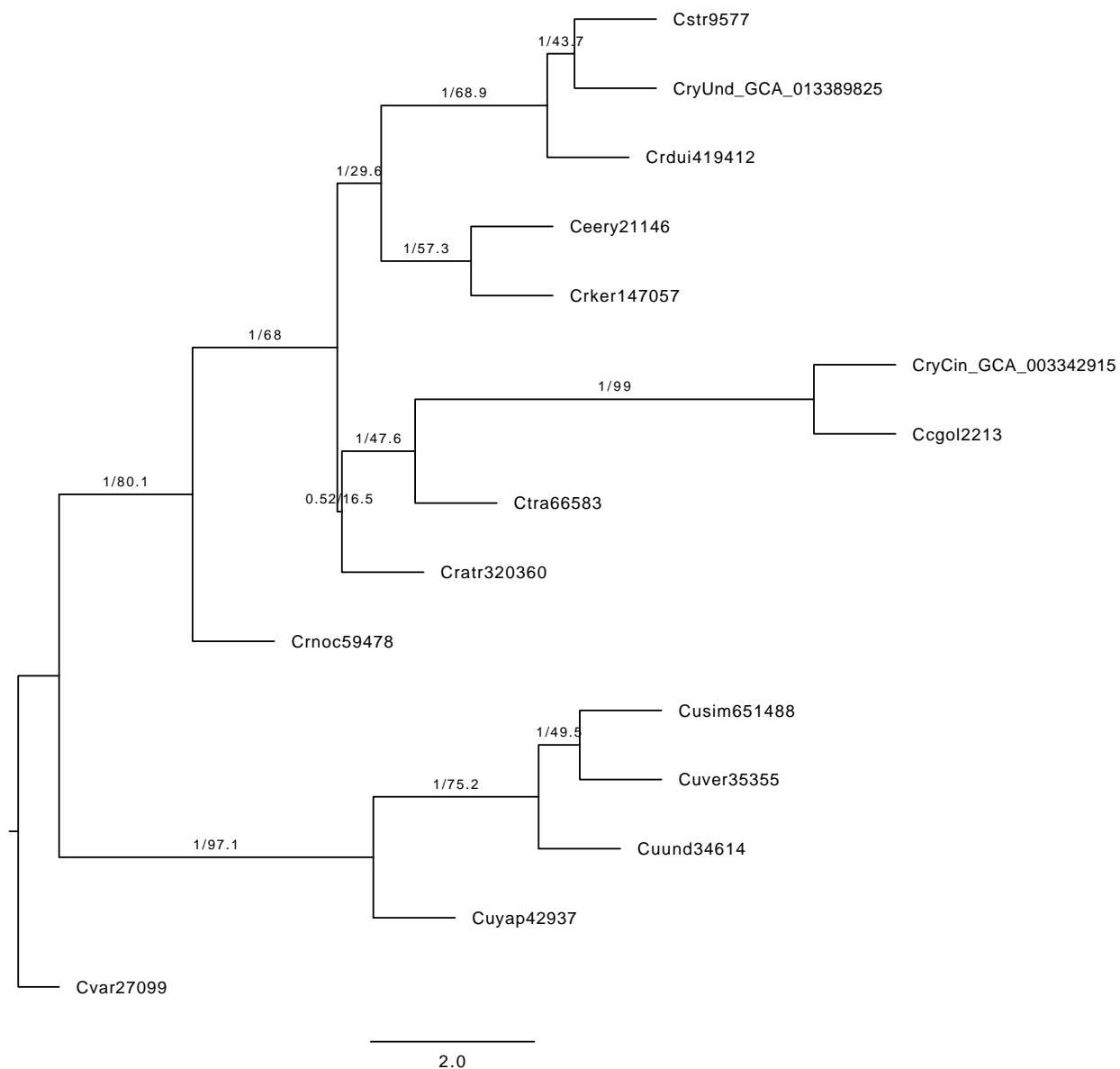
Figure S10: Z-chromosome MSC phylogeny (*Crypturellus* Clade A only). Node support is shown at each internal branch, including posterior probability (first number) and gene concordance factor (second number).
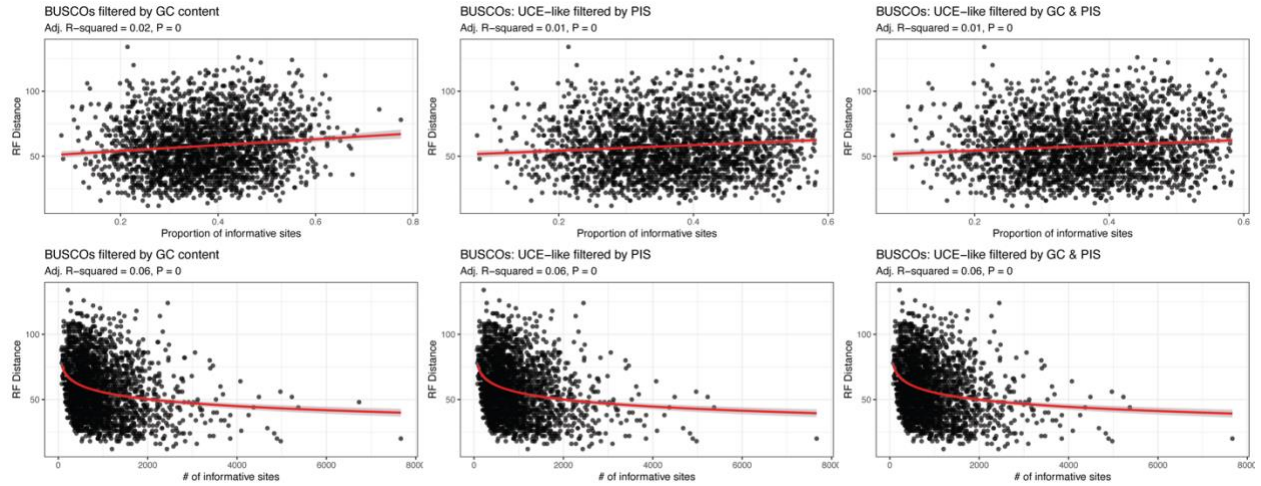
Figure S11: Linear regressions of information content and RF Distance for filtered BUSCOs. The BUSCO dataset was filtered three times. First, the extremes of GC content were removed (BUSCOs filtered by GC content). Second, BUSCO's were filtered by the number and proportion of PIS, wherein BUSCOs with greater than two standard deviations away from the mean number and proportion of PIS were removed. Finally, the two BUSCO filters were combined.
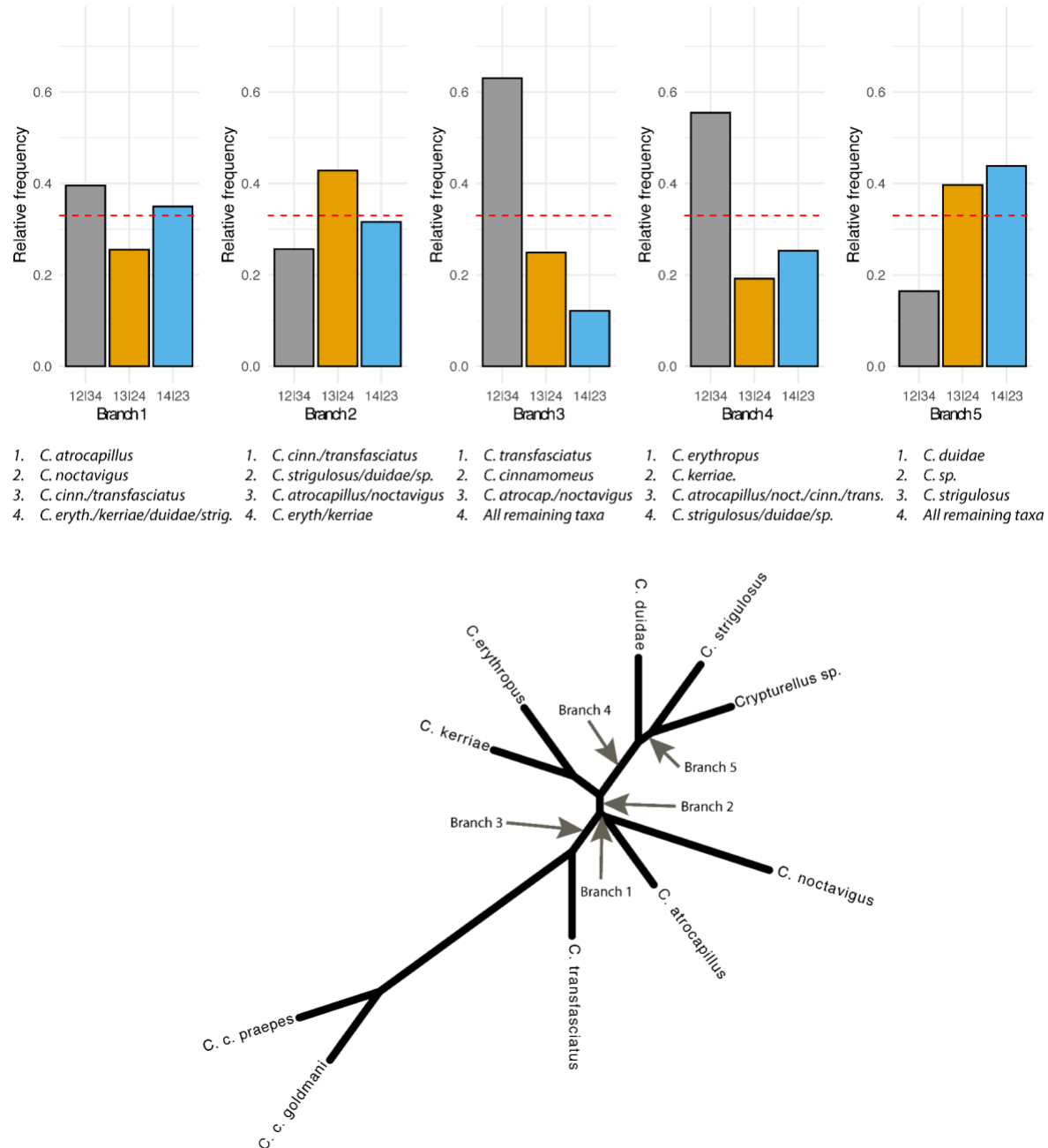
Figure S12: Relative quartet frequencies for five short internodal branches in Clade A based on the UCE1000Flank dataset. Bars depict the relative frequencies of each of three alternative unrooted quartet topologies. The stippled lines indicate the ⅓ threshold for the frequency of gene trees given a multispecies coalescent model, which predicts a single majority quartet topology consistent with the true species tree, and two minority topologies of equivalent frequency below the ⅓ threshold. As quartet frequencies approach the ⅓ threshold, they indicate stronger ILS. Major deviations from the expectations of the MSC model indicate violations of the model may be present, such as gene tree estimation bias or introgression. Branch numbers correspond to Figures 2 and 7.
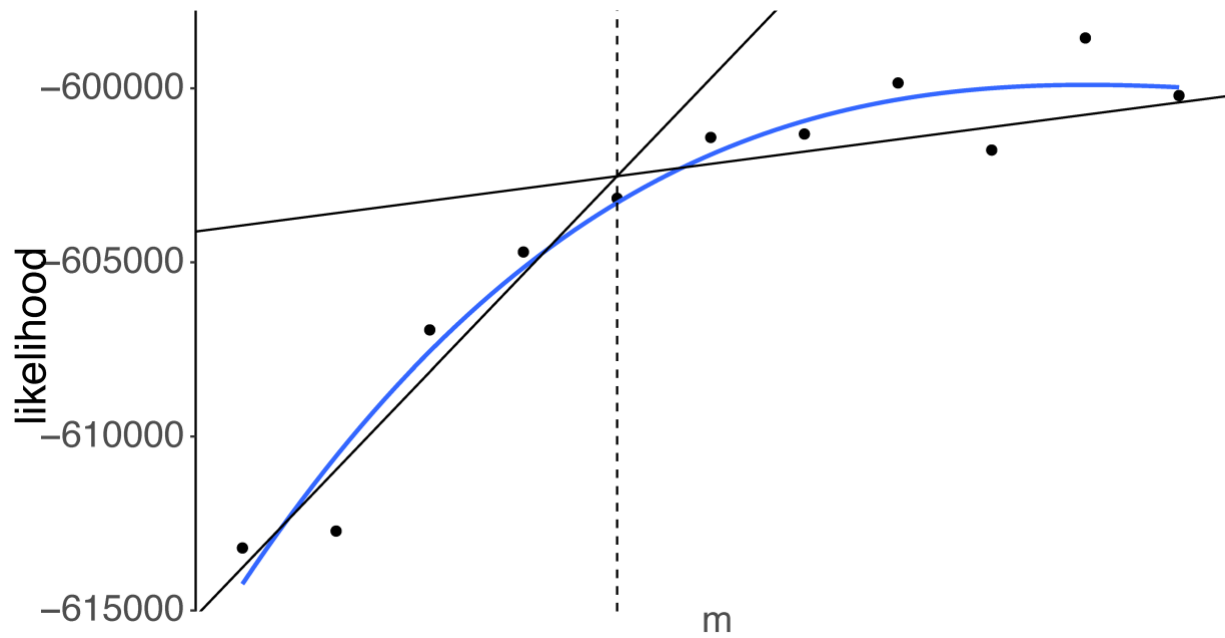
Figure S13: Breakpoint analysis for the best-fit m-value for the introgression analysis.