

# Resolving the phylogenetic position of Darwin's extinct ground sloth (*Mylodon darwini*) using mitogenomic and nuclear exon data

Frédéric Delsuc, Melanie Kuch, Gillian C. Gibb, Jonathan Hughes, Paul Szpak, John Southon, Jacob Enk, Ana T. Duggan' and Hendrik N. Poinar

## Supplementary Material

**Table S1:** Taxon sampling and sequence accession numbers.

**Table S2:** Detailed results of the PartitionFinder analysis for the mitogenomic dataset.

**Table S3:** Detailed results of the PartitionFinder analysis for the nuclear dataset.

**Table S4:** Detailed comparisons between our *Mylodon* mitogenome and the one produced by Slater et al. (2016).

**Figure S1:** mapDamage DNA fragmentation and nucleotide mis-incorporation plots.

**Figure S2:** mapDamage DNA fragment length distributions.

**Figure S3:** Bayesian mitogenomic tree under the CAT-GTR+G<sub>4</sub> mixture model (PhyloBayes).

**Figure S4:** Bayesian mitogenomic tree under a mixed model (MrBayes).

**Figure S5:** Maximum likelihood mitogenomic tree under a GTR+G+I mixed model (RAxML).

**Figure S6:** Bayesian nuclear tree under the CAT-GTR+G<sub>4</sub> mixture model (PhyloBayes).

**Figure S7:** Bayesian nuclear tree under a mixed model (MrBayes).

**Figure S8:** Maximum likelihood nuclear tree under a GTR+G mixed model (RAxML).

**Figure S9:** Bayesian mitogenomic chronogram under the CAT-GTR+G<sub>4</sub> mixture model and an autocorrelated lognormal model of clock relaxation (PhyloBayes).

**Figure S10:** Bayesian nuclear chronogram under the CAT-GTR+G<sub>4</sub> mixture model and an autocorrelated lognormal model of clock relaxation (PhyloBayes).

**Figure S11:** Comparison of the two *Mylodon* mitogenomes aligned with the *Choloepus didactylus* reference mitogenome.

**Figure S12:** Maximum likelihood tree of sloth mitogenomes under a single GTR+G model (RAxML).

**Table S1.** Detailed taxonomic sampling and corresponding sequence accession numbers of the mitochondrial and nuclear datasets. Newly acquired sequences from this study are indicated in bold.

<b>Species</b>	<b>Common name</b>	<b>Mitogenomes</b>	<b>ADORA3</b>	<b>APOB</b>	<b>BCHE</b>	<b>BRCA1</b>	<b>BRCA2</b>	<b>RAG2</b>	<b>TTN</b>
<i>*Bradypus pygmaeus</i>	Pygmy sloth	KT818523	<b>LT852562</b>	<b>LT852583</b>	<b>LT852604</b>	<b>LT852627</b>	<b>LT852645</b>	<b>LT852662</b>	<b>LT852684</b>
<i>*Bradypus torquatus</i>	Maned sloth	KT818524	<b>LT852563</b>	<b>LT852584</b>	<b>LT852605</b>	<b>LT852628</b>	<b>LT852646</b>	<b>LT852663</b>	<b>LT852685</b>
<i>*Bradypus tridactylus</i>	Pale-throated three-toed sloth	KT818525	<b>LT852564</b>	<b>LT852585</b>	<b>LT852606</b>	<b>LT852629</b>	<b>LT852647</b>	<b>LT852664</b>	<b>LT852686</b>
	Brown-throated three-toed sloth	NC_006923	<b>LT852565</b>	<b>LT852586</b>	<b>LT852607</b>	<b>LT852630</b>	<b>LT852648</b>	<b>LT852665</b>	<b>LT852687</b>
<i>*Choloepus didactylus</i>	Southern two-toed sloth	NC_006924	<b>LT852570</b>	<b>LT852591</b>	<b>LT852613</b>	<b>LT852634</b>	<b>LT852651</b>	<b>LT852671</b>	<b>LT852693</b>
	Hoffmann's two- toed sloth	KT818538	<b>LT852571</b>	<b>LT852592</b>	<b>LT852614</b>	<b>LT852635</b>	<b>LT852652</b>	<b>LT852672</b>	<b>LT852694</b>
<i>*Mylodon darwini</i>	Darwin's ground sloth	<b>MF061314</b>	<b>LT852575</b>	<b>LT852596</b>	<b>LT852619</b>	<b>LT852639</b>	<b>LT852656</b>	<b>LT852676</b>	<b>LT852698</b>
<i>*Cyclopes didactylus</i>	Pygmy anteater	KT818539	<b>LT852572</b>	<b>LT852593</b>	<b>LT852615</b>	<b>LT852636</b>	<b>LT852653</b>	<b>LT852673</b>	<b>LT852695</b>
<i>*Myrmecophaga tridactyla</i>	Giant anteater	KT818549	<b>LT852576</b>	<b>LT852597</b>	<b>LT852620</b>	<b>LT852640</b>	<b>LT852657</b>	<b>LT852677</b>	<b>LT852699</b>
<i>*Tamandua mexicana</i>	Northern tamandua	KT818551	<b>LT852578</b>	<b>LT852599</b>	<b>LT852622</b>	<b>LT852641</b>	<b>LT852659</b>	<b>LT852679</b>	<b>LT852701</b>
<i>*Tamandua tetradactyla</i>	Southern tamandua	NC_004032	<b>LT852579</b>	<b>LT852600</b>	<b>LT852623</b>	<b>LT852642</b>	<b>LT852660</b>	<b>LT852680</b>	<b>LT852702</b>
<i>*Dasypus kappleri</i>	Greater long-nosed armadillo	KT818541	<b>LT852574</b>	<b>LT852595</b>	<b>LT852617</b>	<b>LT852638</b>	<b>LT852655</b>	<b>LT852675</b>	<b>LT852697</b>
<i>*Dasypus hybridus</i>	Southern long- nosed armadillo	KT818540	<b>LT852573</b>	<b>LT852594</b>	<b>LT852616</b>	<b>LT852637</b>	<b>LT852654</b>	<b>LT852674</b>	<b>LT852696</b>
<i>Dasypus septemcinctus</i>	Seven-banded armadillo	KT818546	-	-	-	-	-	-	-
<i>Dasypus novemcinctus</i> US	Nine-banded armadillo	NC_001821	XM_004473 603	JN413957	XM_0125197 96	AF484222	XM_012519 803	XM_004483 174	XM_0125 18926
<i>Dasypus novemcinctus</i> FG	Nine-banded armadillo	KT818542	-	-	-	-	-	-	-
<i>Dasypus pilosus</i>	Hairy long-nosed armadillo	KT818544	-	-	-	-	-	-	-

<i>Dasybus sabanicola</i>	Northern long-nosed armadillo	KT818545	-	-	-	-	-	-	-
<i>Dasybus yepesi</i>	Yunga's long-nosed armadillo	KT818547	-	-	-	-	-	-	-
* <i>Chaetophractus vellerosus</i>	Screaming hairy armadillo	KT818533	<b>LT852568</b>	<b>LT852589</b>	<b>LT852611</b>	<b>LT852632</b>	NA	<b>LT852669</b>	<b>LT852691</b>
* <i>Chaetophractus villosus</i>	Large hairy armadillo	KT818535	<b>LT852569</b>	<b>LT852590</b>	<b>LT852612</b>	<b>LT852633</b>	NA	<b>LT852670</b>	<b>LT852692</b>
* <i>Euphractus sexcinctus</i>	Six-banded armadillo	KT818548	AY011193	NA	<b>LT852618</b>	AF484224	NA	AY011922	NA
* <i>Zaedyus pichiy</i>	Pichi	KT818555	<b>LT852582</b>	<b>LT852603</b>	<b>LT852626</b>	<b>LT852644</b>	NA	<b>LT852683</b>	<b>LT852705</b>
* <i>Calyptophractus retusus</i>	Greater fairy armadillo	KT818532	<b>LT852567</b>	<b>LT852588</b>	<b>LT852610</b>	FR821709	<b>LT852650</b>	<b>LT852668</b>	<b>LT852690</b>
<i>Chlamyphorus truncatus</i>	Pink fairy armadillo	KT818536	-	-	-	FR821708	-	-	-
* <i>Priodontes maximus</i>	Giant armadillo	KT818550	<b>LT852577</b>	<b>LT852598</b>	<b>LT852621</b>	AF484225	<b>LT852658</b>	<b>LT852678</b>	<b>LT852700</b>
* <i>Tolypeutes matacus</i>	Southern three-banded armadillo	KT818553	<b>LT852580</b>	<b>LT852601</b>	<b>LT852624</b>	AF484227	<b>LT852661</b>	<b>LT852681</b>	<b>LT852703</b>
* <i>Tolypeutes tricinctus</i>	Brazilian three-banded armadillo	KT818554	<b>LT852581</b>	<b>LT852602</b>	<b>LT852625</b>	<b>LT852643</b>	NA	<b>LT852682</b>	<b>LT852704</b>
* <i>Cabassous chacoensis</i>	Chacoan naked-tailed armadillo	KT818528	NA	NA	<b>LT852608</b>	<b>LT852631</b>	NA	<b>LT852666</b>	<b>LT852688</b>
* <i>Cabassous unicinctus</i>	Southern naked-tailed armadillo	KT818531	<b>LT852566</b>	<b>LT852587</b>	<b>LT852609</b>	AF484228	<b>LT852649</b>	<b>LT852667</b>	<b>LT852689</b>
<i>Cabassous centralis</i>	Northern naked-tailed armadillo	KT818527	-	-	-	-	-	-	-
<i>Cabassous tatouay</i>	Greater naked-tailed armadillo	KT818529	-	-	-	-	-	-	-
<i>Dugong dugon</i>	Dugong	NC_003314	JN633415	KR827337	KR827307 XM_0105960	AF284019	JN414242	JN633302	JN632896 XM_0105
<i>Loxodonta africana</i>	African elephant	NC_000934	AY011204	JN413956	53 XM_0079487	AF284021	JN414245	AY011932	86379
<i>Orycteropus afer</i>	Aardvark	NC_002078	AY011206	JN413950	19	AF284030	JN414239	AY011935	JN632894

Notes -- \*: Libraries selected for nuclear capture; NA: Not available because capture failed to provide enough coverage.

**Table S2:** Detailed results of the PartitionFinder analysis for the mitogenomic dataset.

Subset	Best Model	Subset Partitions
1	GTR+I+G	ATP6_p1, ATP8_p1, ATP8_p2, ATP8_p3, CYTB_p1, ND1_p1, ND2_p1, ND3_p1, ND4L_p1, ND4_p1, ND5_p1, ND6_p2, ND6_p3, rRNA_12S, rRNA_16S, tRNAs
2	GTR+I+G	ATP6_p2, COX1_p1, COX1_p2, COX2_p1, COX2_p2, COX3_p1, COX3_p2, CYTB_p2, ND1_p2, ND2_p2, ND3_p2, ND4L_p2, ND4_p2, ND5_p2
3	GTR+I+G	ATP6_p3, COX1_p3, COX2_p3, COX3_p3, CYTB_p3, ND1_p3, ND2_p3, ND3_p3, ND4L_p3, ND4_p3, ND5_p3, ND6_p1

### Settings used

alignment : ./mylo\_35mitos\_concat\_seqCat.phylip  
branchlengths : unlinked  
models : GTR+I+G, GTR+G  
model\_selection : bic  
search : greedy

### Best partitioning scheme

Scheme Name : step\_39  
Scheme lnL : -169349.5486  
Scheme BIC : 340923.742014  
Number of params : 231  
Number of sites : 15222  
Number of subsets : 3

### RaxML-style partition definitions

DNA, p1 = 1-681\3, 682-876\3, 683-876\3, 684-876\3, 3883-5022\3, 5023-5976\3, 5977-7020\3, 7021-7365\3, 7366-7662\3, 7663-9039\3, 9040-10842\3, 10844-11367\3, 10845-11367\3, 11368-12305, 12306-13766, 13767-15222

DNA, p2 = 2-681\3, 877-2415\3, 878-2415\3, 2416-3099\3, 2417-3099\3, 3100-3882\3, 3101-3882\3, 3884-5022\3, 5024-5976\3, 5978-7020\3, 7022-7365\3, 7367-7662\3, 7664-9039\3, 9041-10842\3

DNA, p3 = 3-681\3, 879-2415\3, 2418-3099\3, 3102-3882\3, 3885-5022\3, 5025-5976\3, 5979-7020\3, 7023-7365\3, 7368-7662\3, 7665-9039\3, 9042-10842\3, 10843-11367\3

**Table S3:** Detailed results of the PartitionFinder analysis for the seven nuclear exons dataset.

Subset	Best Model	Subset Partitions
1	GTR+G	ADORA3_pos1, ADORA3_pos2, APOB_pos2
2	GTR+G	ADORA3_pos3, APOB_pos3, RAG2_pos3
3	GTR+G	APOB_pos1, BRCA1_pos1
4	GTR+G	BCHE_pos1, RAG2_pos1, TTN_pos1
5	GTR+I+G	BCHE_pos2, RAG2_pos2, TTN_pos2
6	GTR+G	BCHE_pos3, BRCA2_pos3
7	GTR+G	BRCA1_pos2, BRCA2_pos2
8	GTR+G	BRCA1_pos3, TTN_pos3
9	GTR+G	BRCA2_pos1

#### Settings used

alignment : ./Mylo\_7exons\_seqCat.phylip  
branchlengths : linked  
models : GTR+I+G, GTR+G  
model\_selection : bic  
search : greedy

#### Best partitioning scheme

Scheme Name : step\_12  
Scheme lnL : -61830.71542  
Scheme BIC : 125038.535538  
Number of params : 143  
Number of sites : 15216  
Number of subsets : 9

#### RaxML-style partition definitions

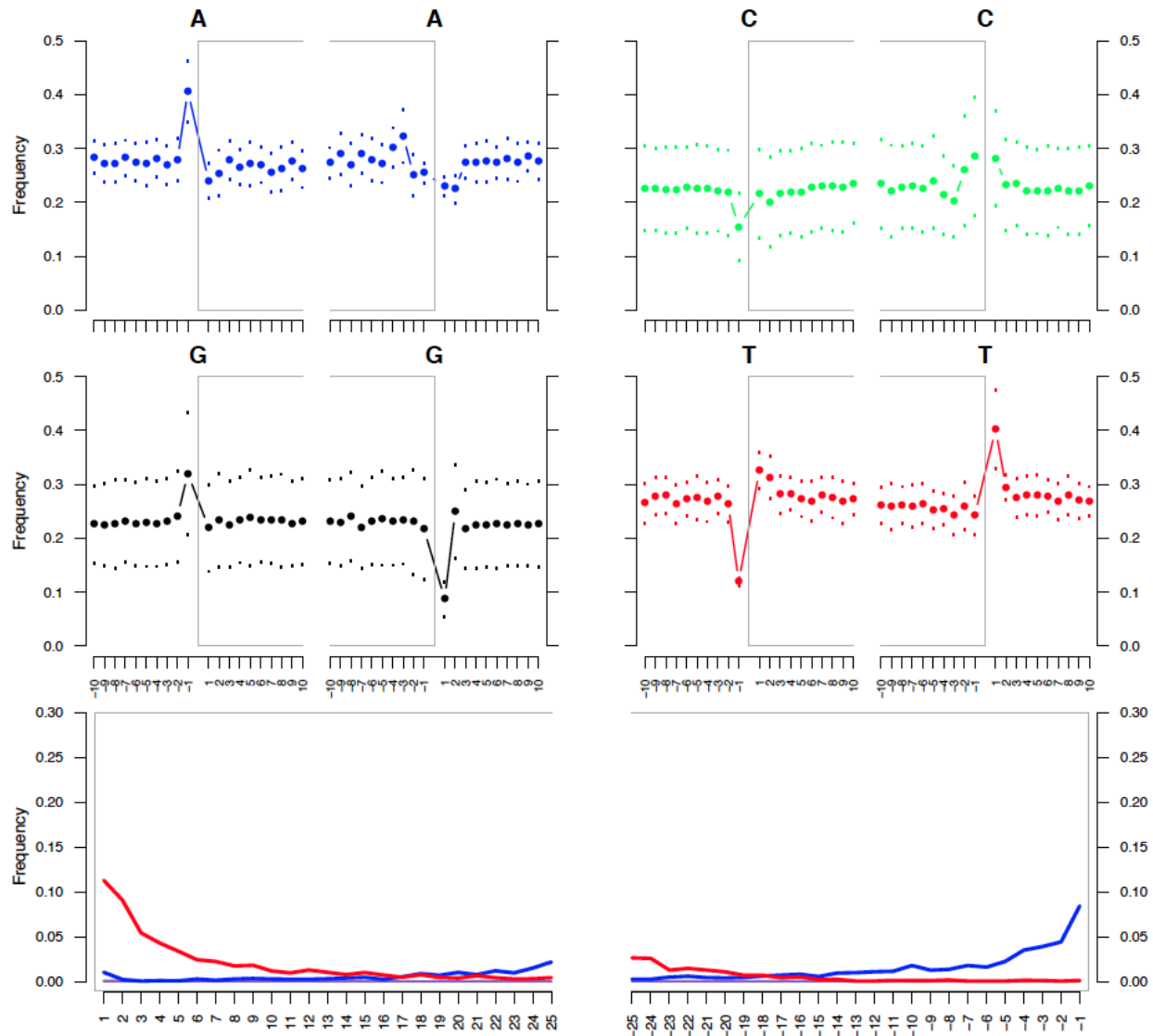
DNA, p1 = 1-312\3, 2-312\3, 314-2724\3  
DNA, p2 = 3-312\3, 315-2724\3, 10374-10797\3  
DNA, p3 = 313-2724\3, 3706-6483\3  
DNA, p4 = 2725-3705\3, 10372-10797\3, 10798-15216\3  
DNA, p5 = 2726-3705\3, 10373-10797\3, 10799-15216\3  
DNA, p6 = 2727-3705\3, 6486-10371\3  
DNA, p7 = 3707-6483\3, 6485-10371\3  
DNA, p8 = 3708-6483\3, 10800-15216\3  
DNA, p9 = 6484-10371\3

**Table S4:** Detailed comparisons between our *Mylodon* mitogenome and the one produced by Slater et al. (2016).

<b>Annotated regions</b>	<b>Length (bp)</b>	<b>Identity (%)</b>	<b>Problems in Slater et al. (2016) sequence (KR336794)</b>
tRNA-Phe	70	100	None
12S rRNA	963	99.2	1 divergent region (12 bp) including a 4 bp insertion
tRNA-Val	68	100	None
16S rRNA	1575	98.7	1 substitution + 1 divergent region (26 bp ) + 3 Ns
tRNA-Leu	75	100	None
ND1	957	88.7	4 substitutions + 3 divergent regions (26 bp, 14bp, 19bp) including missing start codon + 60 Ns (7 stop codons)
tRNA-Ile	69	100	None
tRNA-Gln	73	100	None
tRNA-Met	69	100	None
ND2	1044	91.0	2 substitutions + 4 divergent regions (88 bp, 41 bp, 51 bp, 6 bp) + 3 deletions + 1 insertion + 4 Ns (8 stop codons)
tRNA-Trp	69	100	None
tRNA-Ala	69	100	None
tRNA-Asn	73	100	None
Origin of L-strand replication	34	100	None
tRNA-Cys	67	100	None
tRNA-Tyr	66	100	None
COX1	1542	90.5	3 substitutions + 43 Ns
tRNA-Ser	69	100	None
tRNA-Asp	68	97.1	1 substitution + 1 insertion
COX2	684	94.2	21 substitutions + 2 divergent regions (10 bp, 19 bp)
tRNA-Lys	69	100	None
ATP8	204	52.5	3 substitutions + 94 gaps (1 stop codon)
ATP6	681	41.8	13 substitutions + 3 divergent regions (39 bp, 29 bp, 34 bp) + 175 Ns + 259 gaps (1 stop codon)
COX3	784	96.4	2 substitutions + 1 divergent region (23 bp) + 1 insertion (10 stop codons)
tRNA-Gly	70	100	None
ND3	346	54.3	7 substitutions + 151 Ns
tRNA-Arg	70	100	None
ND4L	297	59.1	1 divergent region (14 bp) + 110 Ns
ND4	1378	85.7	187 Ns + 2 insertions (29 stop codons)
tRNA-His	69	100	None
tRNA-Ser	60	100	None

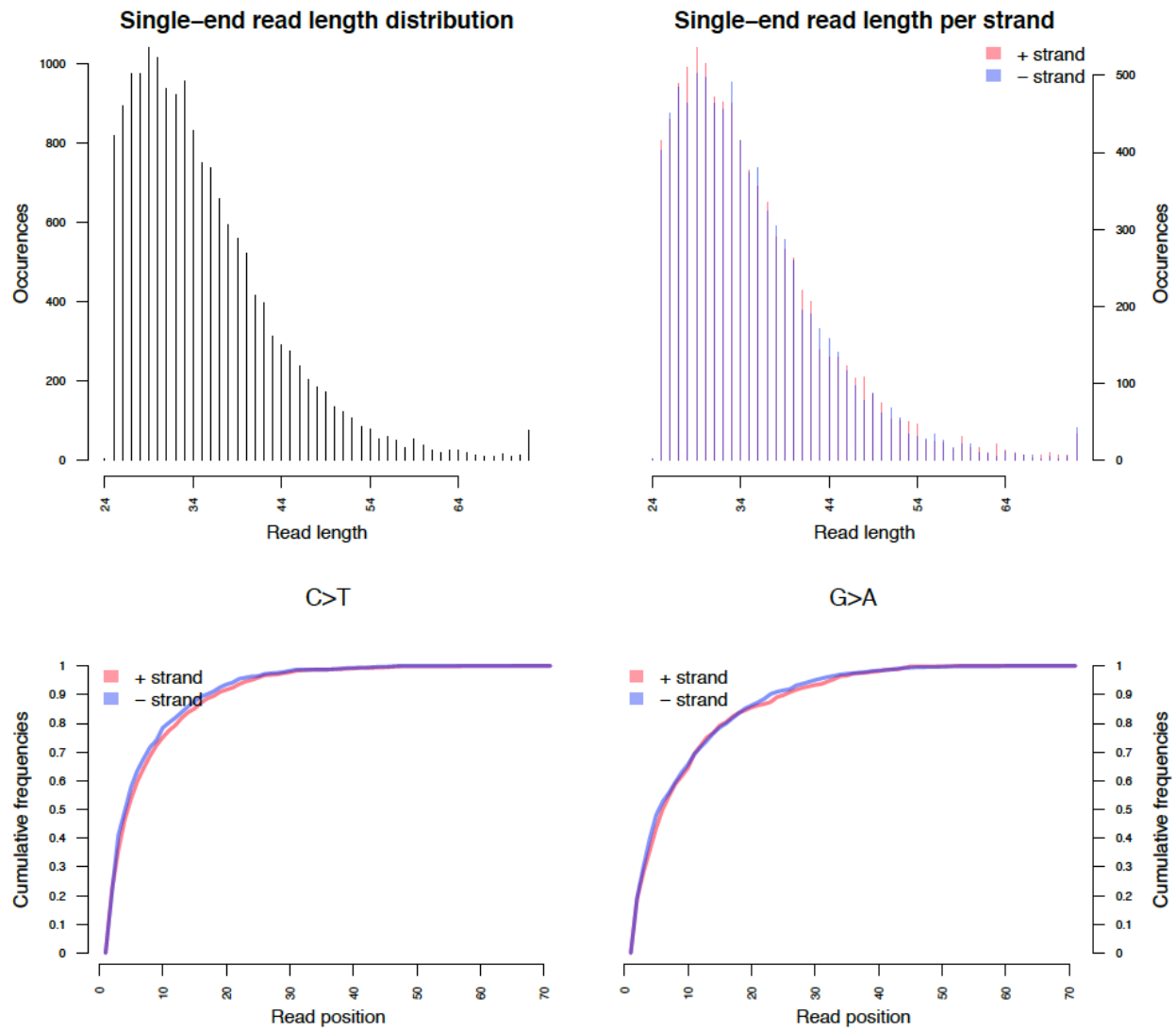
tRNA-Leu	70	100	None
ND5	1824	53.5	13 substitutions + 2 divergent regions (10 bp, 3 bp) + 809 gaps + 16 Ns (9 stop codons)
ND6	528	72.2	6 substitutions + 1 divergent region (13 bp) + 4 gaps + 126 Ns (2 stop codons)
tRNA-Glu	69	100	None
CYTB	1140	98.9	11 substitutions + 2 insertions (27 stop codons)
tRNA-Thr	67	29.4	22 gaps + 26 Ns
tRNA-Pro	70	0	70 Ns
Control Region	993	60.0	48 gaps + 310 Ns

---

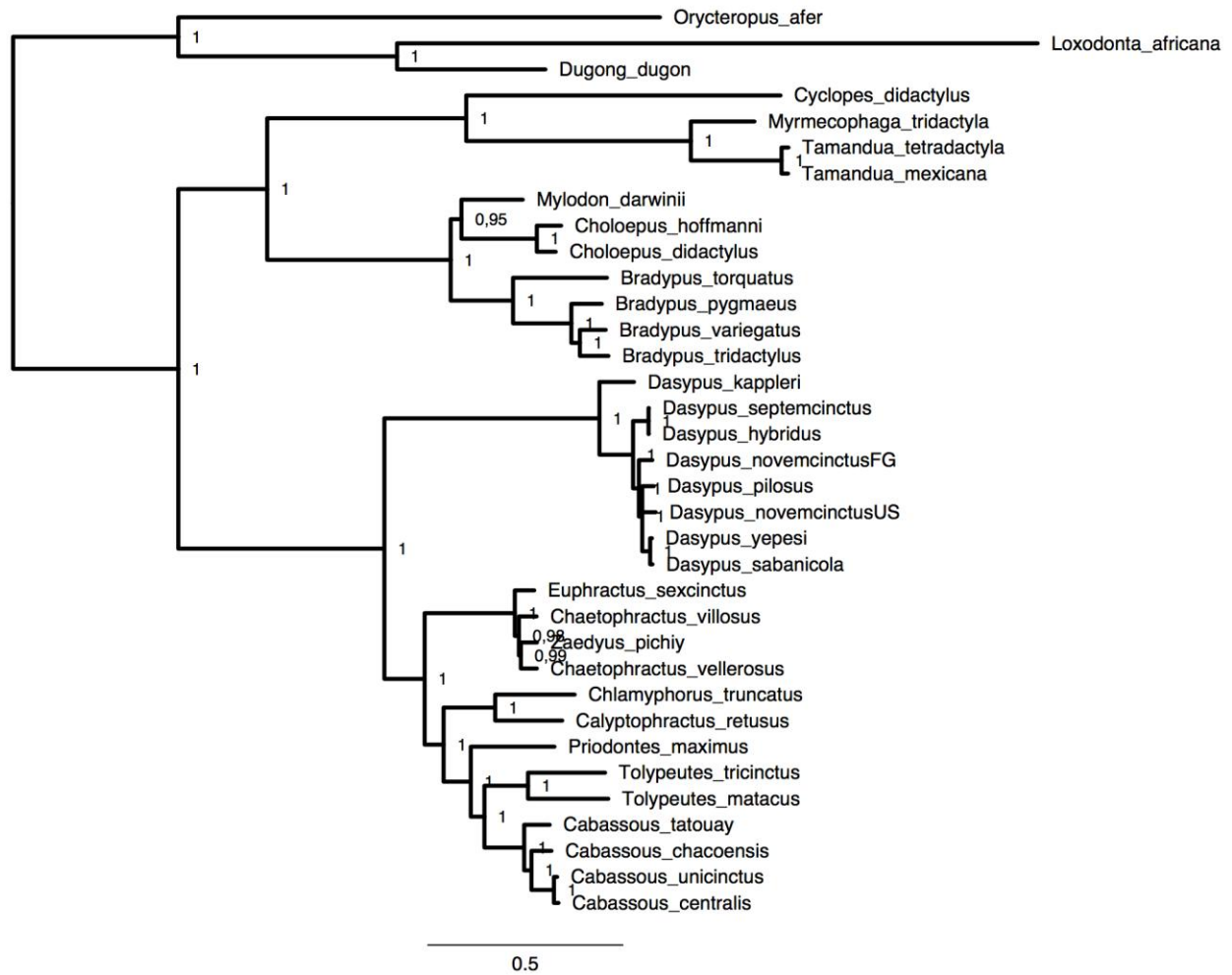


**Figure S1.** mapDamage DNA fragmentation and nucleotide mis-incorporation plots of the 53,550 cleaned shotgun Illumina reads mapping to the *Myiodon darwinii* SP57 mitogenome. 1st and 2nd rows: DNA fragmentation patterns for the 10 bp preceding read starts (positions -1 to -10 on the left panels of each base composition profile) and the 10 bp following read ends (positions 1 to 10 on the right panels of each base composition profile). 3rd row: Nucleotide mis-incorporation patterns for the first 25 nucleotides and the last 25 sequenced, respectively, with C=>T substitution frequencies (red) and G=>A substitution frequencies (blue).

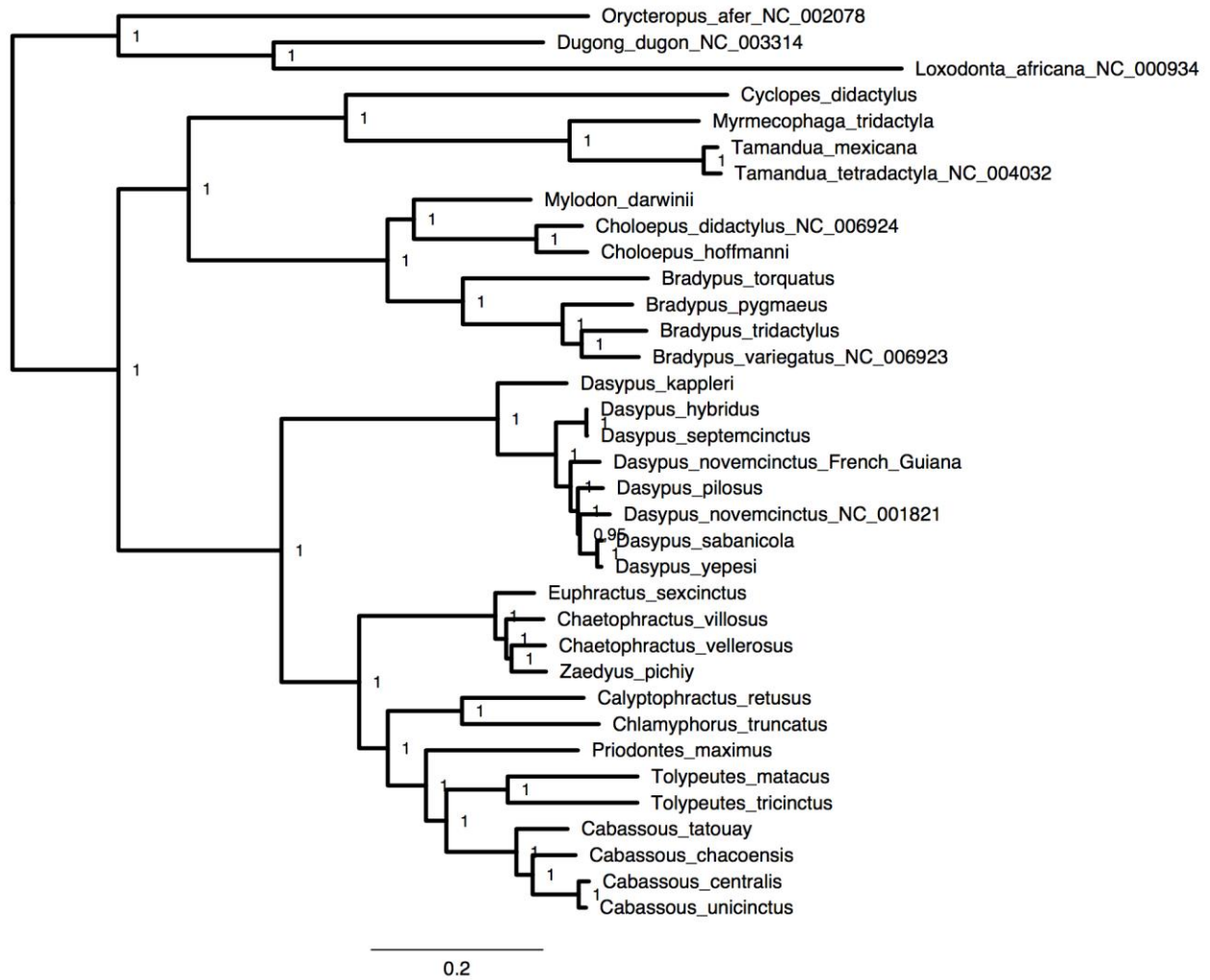




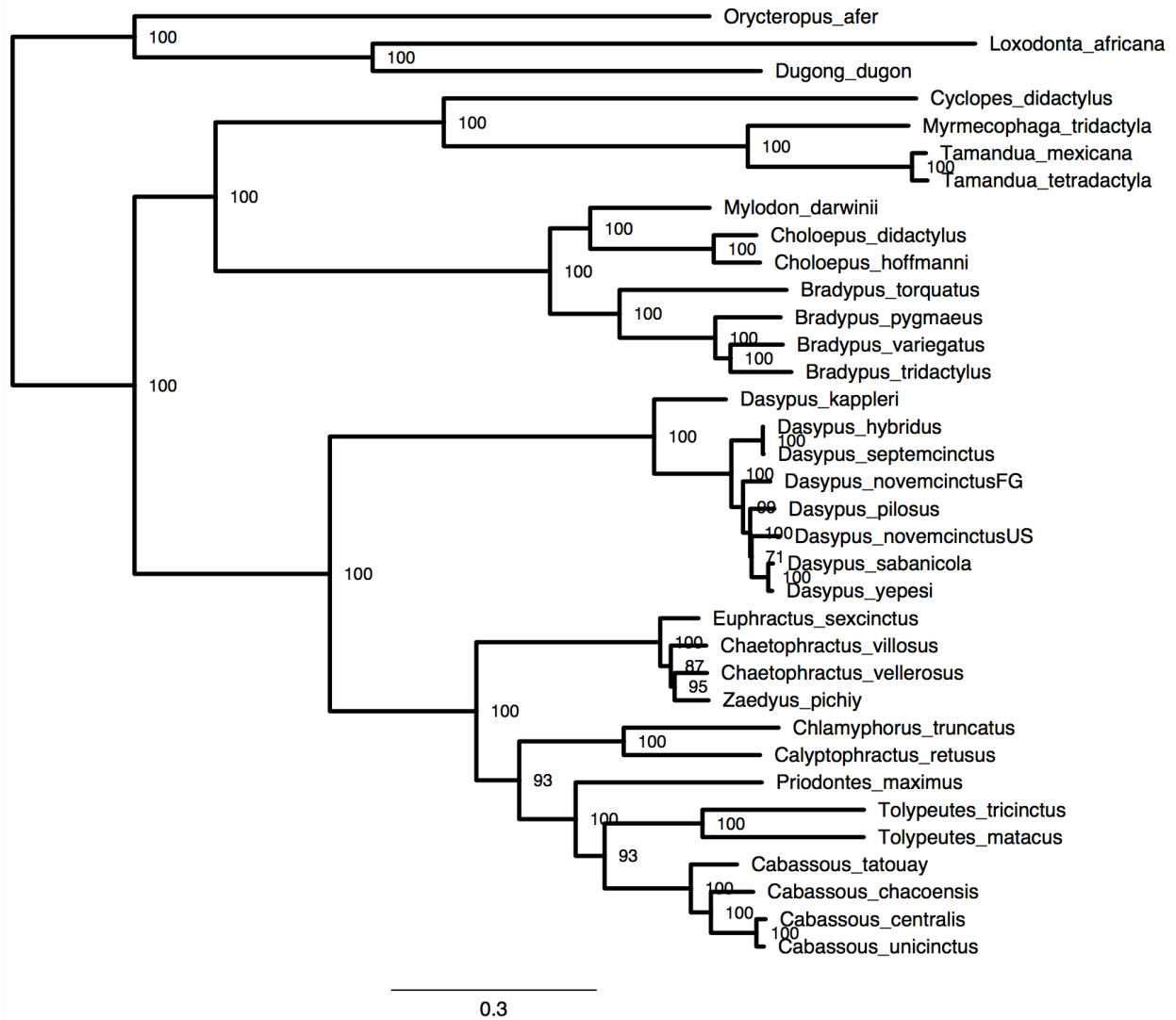
**Figure S2.** mapDamage fragment length distribution of the 53,550 shotgun cleaned Illumina reads mapping to the *Myiodon darwinii* SP57 mitogenome. Mean fragment length was 30 base pairs.



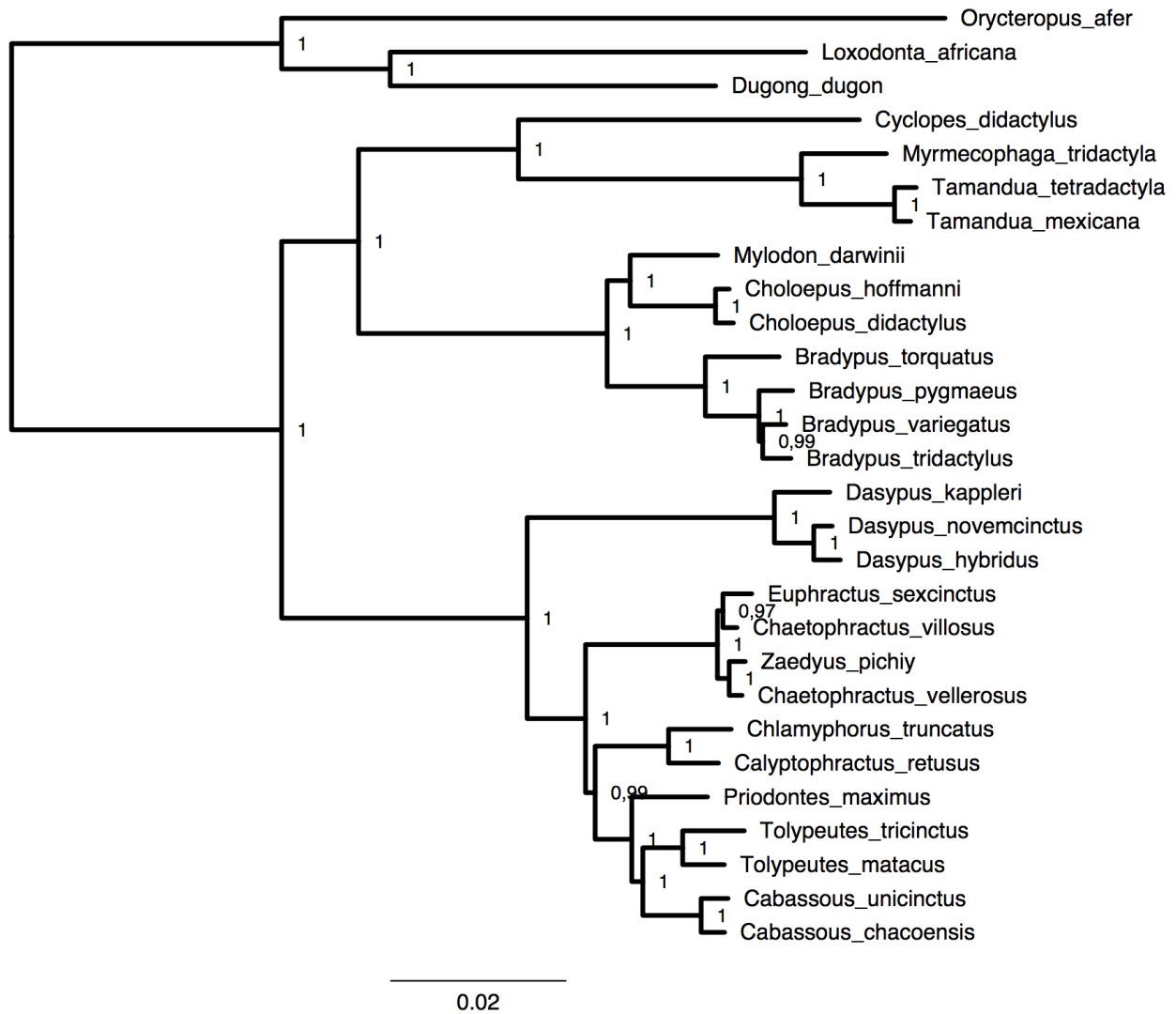
**Figure S3:** Bayesian consensus mitogenomic tree obtained with PhyloBayes under the CAT-GTR+G<sub>4</sub> mixture model. Bayesian posterior probabilities are indicated at nodes. Branch lengths represent mean number of substitutions per site.



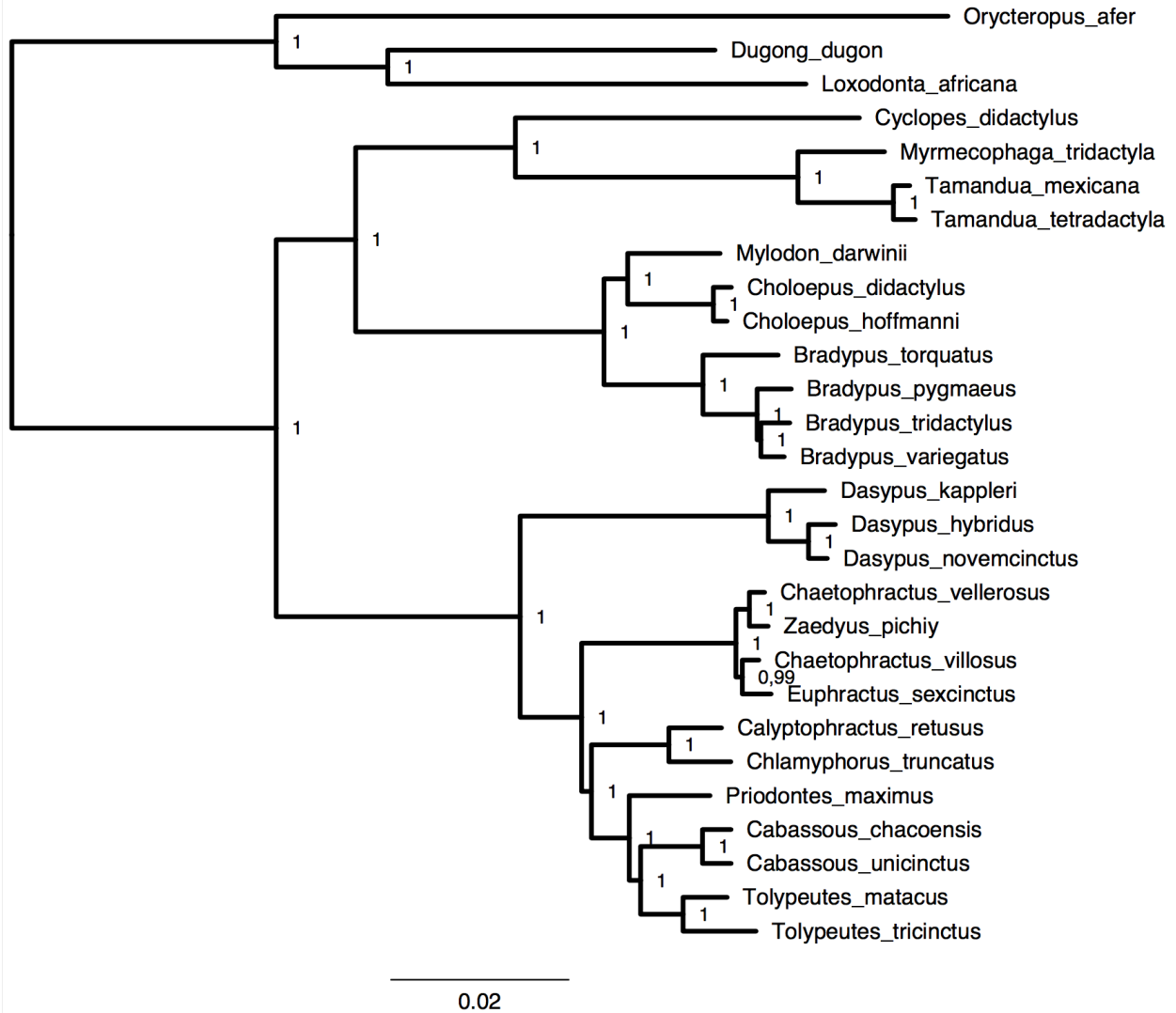
**Figure S4:** Bayesian consensus mitogenomic tree obtained with MrBayes under a mixed model for each best-fitting partition as estimated by PartitionFinder. Bayesian posterior probabilities are indicated at nodes. Branch lengths represent mean number of substitutions per site.



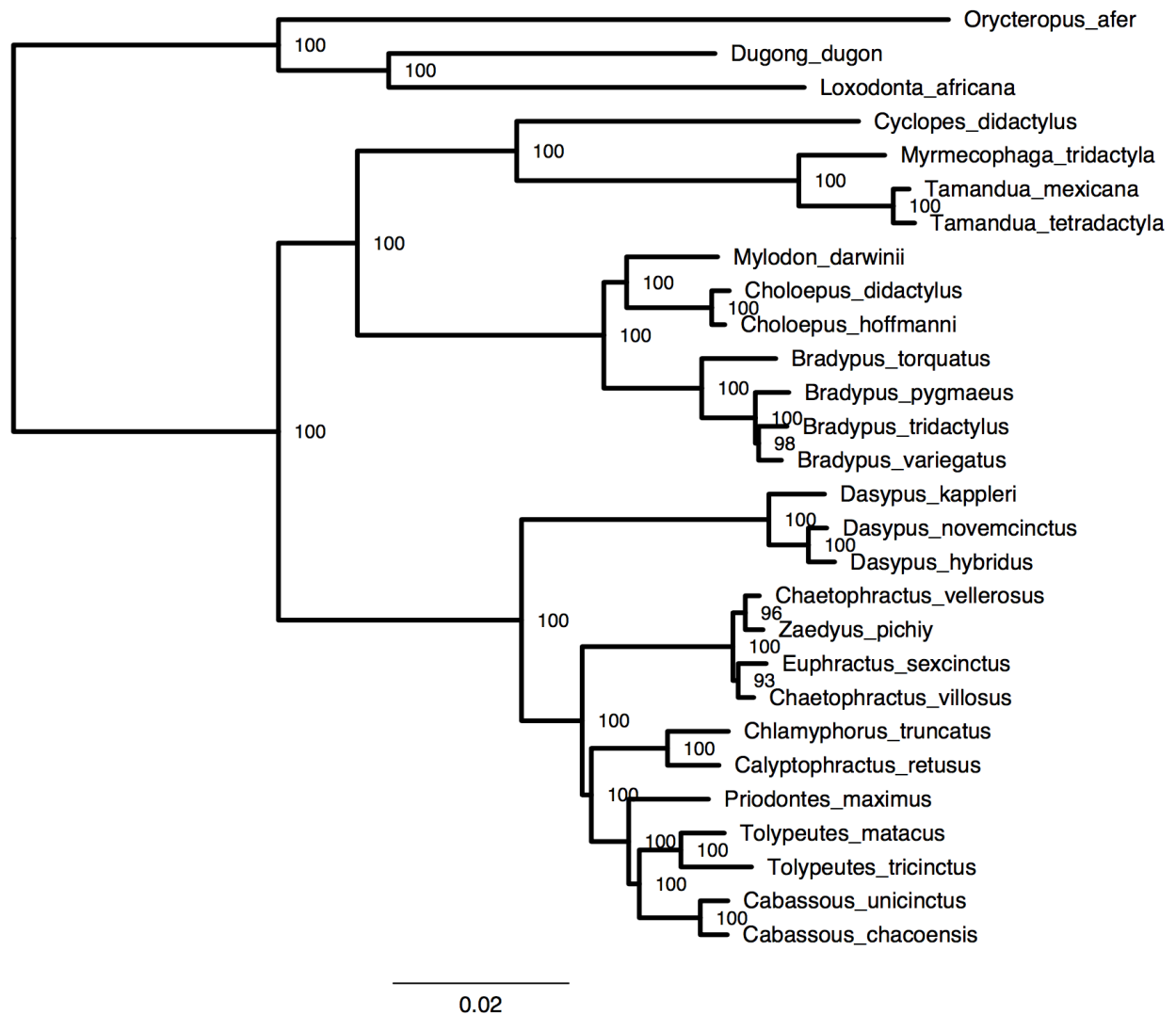
**Figure S5:** Maximum likelihood mitogenomic tree obtained with RAxML under a GTR+G+I mixed model for each best-fitting partition as estimated by PartitionFinder. Bootstrap percentages are indicated at nodes. Branch lengths represent mean number of substitutions per site.



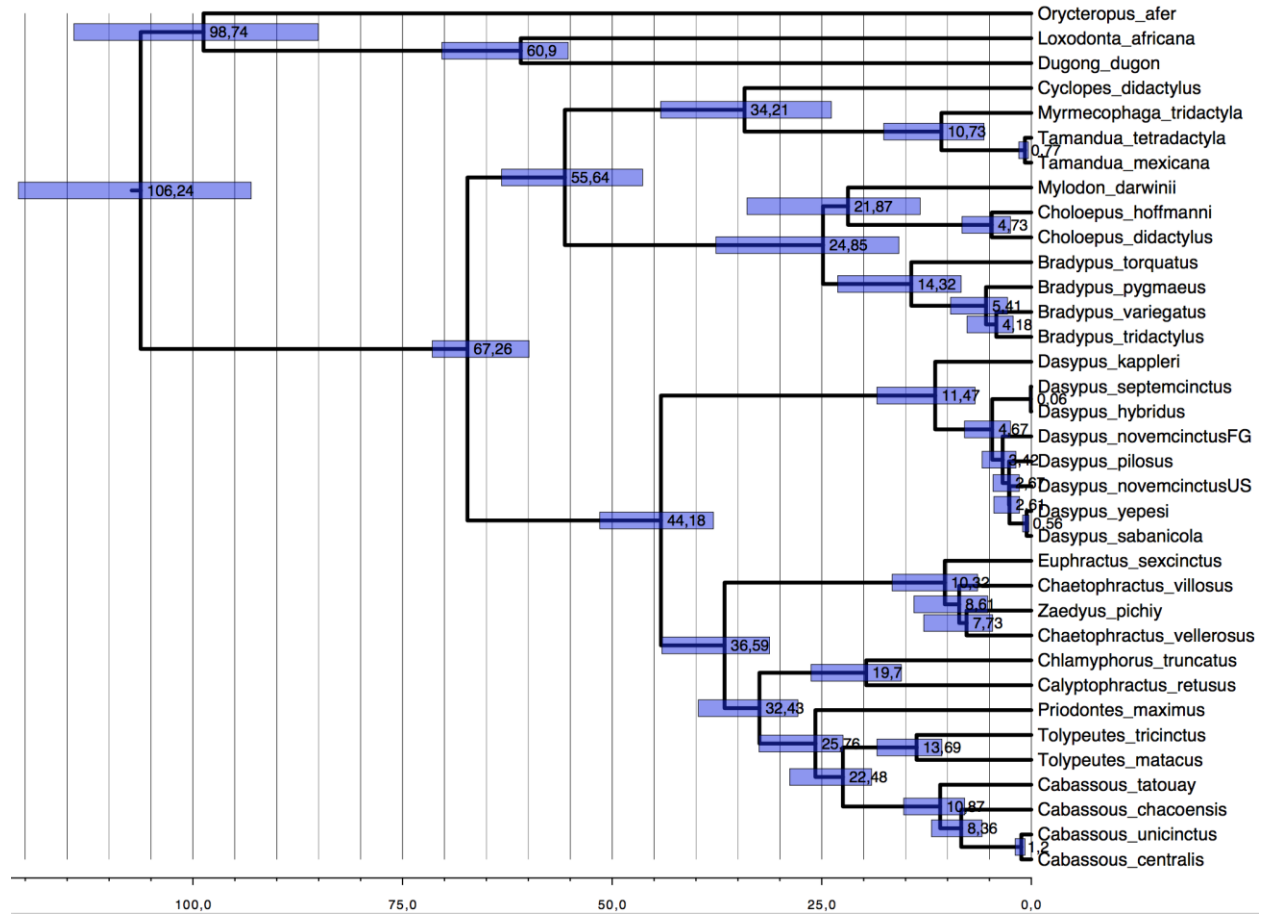
**Figure S6:** Bayesian consensus nuclear tree (concatenation of 7 exons) obtained with PhyloBayes under the CAT-GTR+G<sub>4</sub> mixture model. Bayesian posterior probabilities are indicated at nodes. Branch lengths represent mean number of substitutions per site.



**Figure S7:** Bayesian consensus nuclear tree (concatenation of 7 exons) obtained with MrBayes under a mixed model for each best-fitting partition as estimated by PartitionFinder. Bayesian posterior probabilities are indicated at nodes. Branch lengths represent mean number of substitutions per site.

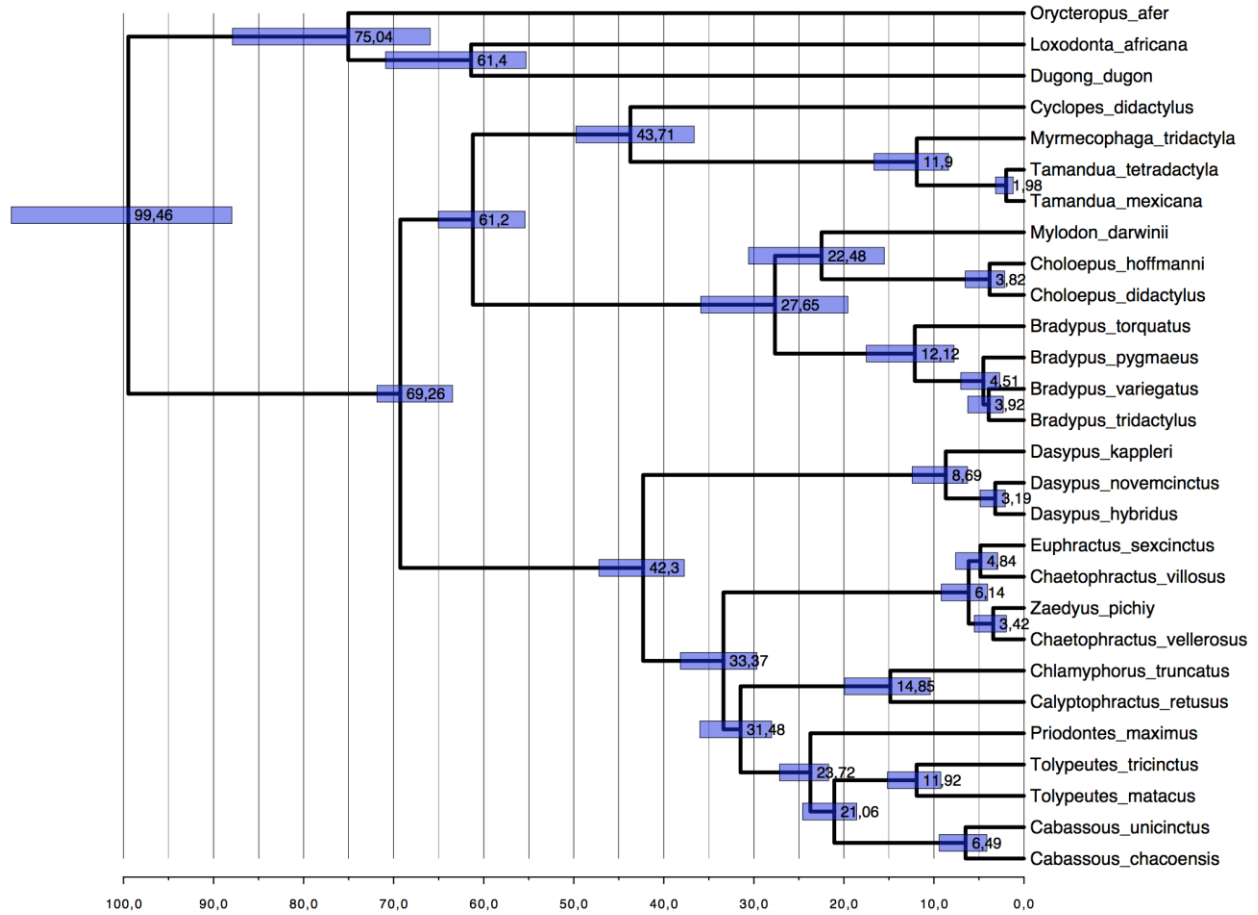


**Figure S8:** Maximum likelihood nuclear tree (concatenation of 7 exons) obtained with RAxML under a GTR+G mixed model for each best-fitting partition as estimated by PartitionFinder. Bootstrap percentages are indicated at nodes. Branch lengths represent mean number of substitutions per site.



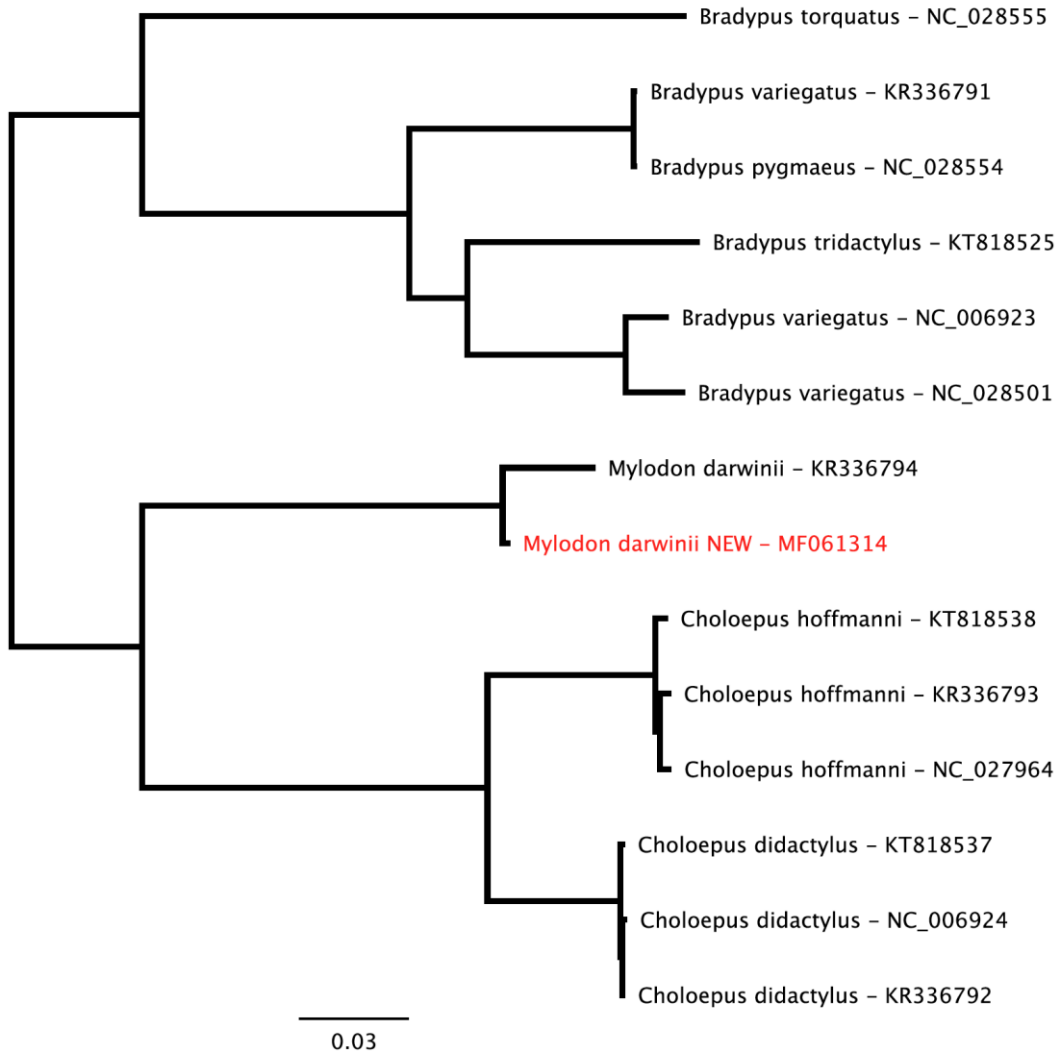
**Figure S9:** Bayesian mitogenomic chronogram obtained with PhyloBayes under the CAT-GTR+G<sub>4</sub> mixture model and an autocorrelated lognormal model of clock relaxation. Timescale is in million years ago.





**Figure S10:** Bayesian nuclear chronogram (concatenation of 7 exons) obtained with PhyloBayes under the CAT-GTR+G<sub>4</sub> mixture model and an autocorrelated lognormal model of clock relaxation. Timescale is in million years ago.

**Figure S11:** Comparison of available *Mylodon* mitogenomes with the *Choloepus didactylus* reference mitogenome.



**Figure S12:** Maximum likelihood tree of available sloth mitogenomes obtained with RAxML under a single GTR+G model. Our new *Mylodon* mitogenome sequence is indicated in red to illustrate its much shorter branch length relative to the one previously obtained by Slater et al. (2016). Branch lengths represent mean number of substitutions per site.