

README

We updated the human age inference and origination mechanism inference using gene annotations from Ensembl release 115. In addition to a new phylogenetic tree (Fig 1), the same pipeline described in Shao et. al. (2019) is used to infer age and mechanism information.

The genome assemblies used for the whole-genome alignment are listed in Table 1, which were downloaded from NCBI GenBank.

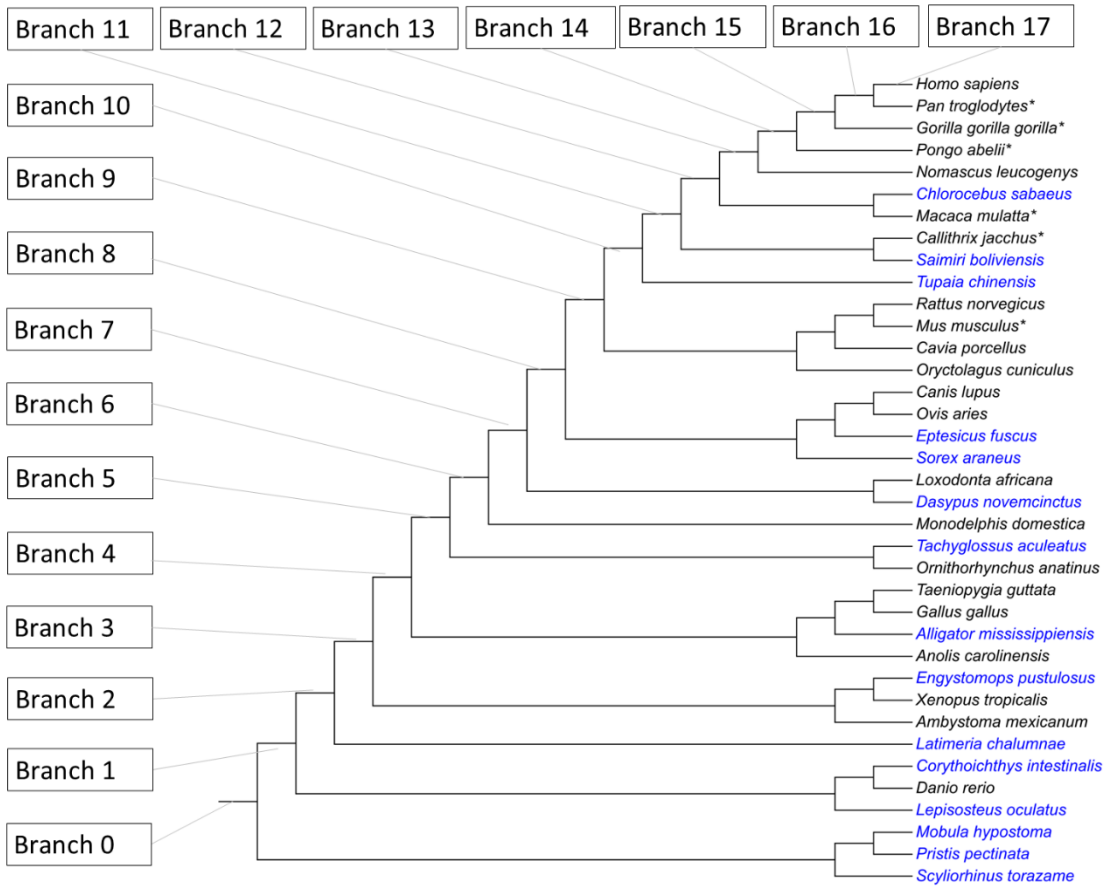


Figure 1. Updated phylogenetic tree

Newly added species were marked in blue. The human assembly remains hg38 (GRCh38.p14); species marked with an asterisk (*) use T2T (telomere-to-telomere) genome assemblies.

Table 1. Genome assemblies used in the new version dataset

Organism Name	Assembly Accession	Assembly Name
Homo sapiens	GCF_000001405.40	GRCh38.p14
Pan troglodytes	GCF_028858775.2	NHGRI_mPanTro3-v2.1_pri
Gorilla gorilla gorilla	GCF_029281585.2	NHGRI_mGorGor1-v2.1_pri
Pongo abelii	GCF_028885655.2	NHGRI_mPonAbe1-v2.1_pri
Nomascus leucogenys	GCF_006542625.1	Asia_NLE_v1
Chlorocebus sabaeus	GCF_047675955.1	mChISab1.0.hap1
Macaca mulatta	GCA_049350105.2	T2T-MMU8v2.0
Callithrix jacchus	GCA_049354665.1	calJac220_pri
Saimiri boliviensis	GCF_048565385.1	mSaiBol1.pri
Tupaia chinensis	GCA_033439345.1	KIZ version 2/TS_2.0
Rattus norvegicus	GCF_036323735.1	GRCr8
Mus musculus	GCA_050437135.1	T2T_mhaESC_v1.0
Cavia porcellus	GCF_034190915.1	mCavPor4.1
Oryctolagus cuniculus	GCF_964237555.1	mOryCun1.1
Canis lupus familiaris	GCF_000002285.5	Dog10K_Boxer_Tasha
Ovis aries	GCF_016772045.2	ARS-UI_Ramb_v3.0
Eptesicus fuscus	GCF_027574615.1	DD_ASM_mEF_20220401
Sorex araneus	GCF_027595985.1	mSorAra2.pri
Loxodonta africana	GCF_030014295.1	mLoxAfr1.hap2
Dasyus novemcinctus	GCF_030445035.2	mDasNov1.1.hap2
Monodelphis domestica	GCF_027887165.1	mMonDom1.pri
Tachyglossus aculeatus	GCF_015852505.1	mTacAcu1.pri
Ornithorhynchus anatinus	GCF_004115215.2	mOrnAna1.pri.v4
Taeniopygia guttata	GCF_048771995.1	bTaeGut7.mat
Gallus gallus	GCF_016699485.2	bGalGal1.mat.broiler.GRCg7b
Alligator mississippiensis	GCF_030867095.1	rAllMis1
Anolis carolinensis	GCF_035594765.1	rAnoCar3.1.pri
Engystomops pustulosus	GCF_040894005.1	aEngPut4.maternal
Xenopus tropicalis	GCF_000004195.4	UCB_Xtro_10.0
Ambystoma mexicanum	GCF_040938575.1	UKY_AmexF1_1
Latimeria chalumnae	GCF_037176945.1	fLatCha1.pri
Corythoichthys intestinalis	GCF_030265065.1	ASM3026506v1
Danio rerio	GCF_049306965.1	GRCz12tu
Lepisosteus oculatus	GCF_040954835.1	fLepOcu1.hap2
Mobula hypostoma	GCF_963921235.1	sMobHyp1.1
Pristis pectinata	GCF_009764475.1	sPriPec2.1.pri
Scyliorhinus torazame	GCF_047496885.1	sScyTor2.1

Origination mechanism classification

For each gene, the origination mechanism was classified into the following categories:

A: candidate *de novo* gene (without identified paralogs)

D / DI: DNA-level duplication / less reliable

R / RI: retroposition (RNA-level duplication) / less reliable

Reference:

Shao, Y., Chen, C., Shen, H., He, B. Z., Yu, D., Jiang, S., Zhao, S., Gao, Z., Zhu, Z., Chen, X., Fu, Y., Chen, H., Gao, G., Long, M., & Zhang, Y. E. (2019). GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Research*, 29(4), 682–696. <https://doi.org/10.1101/gr.238733.118>