

Jan 13, 2021

BEAST v1.X tutorial: Mammalian timetree

kawaoso¹¹Soichi Osozawa

1

Works for me

dx.doi.org/10.17504/protocols.io.brwym27w

KawaOso



kawaoso

ABSTRACT

A Bayesian inference (BI) tree is constructed using the software BEAST v.1X (Bayesian Evolutionary Analysis Sampling Trees; Suchard et al., 2018; released on 10th June, 2018), running BEAUti, BEAST, TreeAnnotator, and FigTree, in ascending order. Before operating the BEAST software, the BEAGLE Library must be downloaded. Platform software is BEAST, and the calculation time was more than 2 hours.

DOI

dx.doi.org/10.17504/protocols.io.brwym27w

PROTOCOL CITATION

kawaoso 2021. BEAST v1.X tutorial: Mammalian timetree. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.brwym27w>



LICENSE

————— This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

CREATED

Jan 13, 2021

LAST MODIFIED

Jan 13, 2021

PROTOCOL INTEGER ID

46232

MATERIALS TEXT

References for calibrations

Bajpai, S. and Gingerich, P.D. A new Eocene archaeocete (Mammalia, Cetacea) from India and the time of origin of whales. *Proceedings of the National Academy of Sciences, USA* 95:15464-15468 (1998).

Barry, J.C., Morgan, M.L.E., Flynn, L.J., Pilbeam, D., Behrensmeyer, A.K., Raza, S.M., Khan, I.A., Badgley, C., Hicks, J., and Kelley, J. Faunal and environmental change in the late Miocene Siwaliks of northern Pakistan. *Paleobiology* 28:1-71 (2002).

Benton MJ, Donoghue PCJ, Asher RJ, Friedman MN, Thomas J, Jakob V (2015) Constraints on the timescale of animal evolutionary history. *Palaeontologia Electronica* 18.1.1FC; 1-106 (2009).

Grande L. Paleontology of the Green River Formation with a review of the fish fauna. *The Geological Survey of Wyoming Bulletin* 63: 1-333 (1980).

Johnson NM, Stix J, Tauxe L, Cervený PF, Tahirkheli RAK. Paleomagnetic chronology, fluvial processes, and tectonic implications of the Siwalik deposits near Chinji Village, Pakistan. *Journal of Geology* 93: 27-40 (1985).

Katoh S, Beyene Y, Itaya T, Hyodo H, Hyodo M, Yagi K, Gouzu C, WoldeGabriel G, Hart WK, Ambrose SH, Nakaya, Bernor RL, Boisserie JR, Bibi F, Saegusa H, Sasaki T, Sano K, Asfaw B, Suwa G. New geological and palaeontological age constraint for the gorilla–human lineage split. *Nature* 530: 215-218 (2016).

Luo ZX, Yuan CX, Meng QJ, Ji Q. A Jurassic eutherian mammal and divergence of marsupials and placentals. *Nature* 476:442-445 (2011).

McBrearty S, Jablonski NG. First fossil chimpanzee. *Nature* 437 (2005). doi:10.1038/nature04008

Roberts EM, Stevens NJ, O'Connor PM, Dirks PHGM, Gottfried MD, Clyde WC, Armstrong RA, Kemp AIS Hemming S. Initiation of the western branch of the East African Rift coeval with the eastern branch. *Nature Geoscience* 5 (2010). doi: 10.1038/NGEO1432

Robinson, P., Gunnell, G.F., Walsh, S.L., Clyde, W.C., Storer, J.E., Stucky, R.K., Froehlich, D.J., Ferrusquia-Villafranca, I., and McKenna, M.C. Wasatchian through Duchesnean biochronology, p. 106-155 (2004). In Woodburne, M.O. (ed.), *Late Cretaceous and Cenozoic mammals of North America: Biostratigraphy and geochronology*. Columbia University Press, New York.

Rose, K.D., DeLeon, V.B., Missiaen, P., Rana, R.S., Sahni, A., Singh, L., Smith, T. Early Eocene lagomorph (Mammalia) from Western India and the early diversification of Lagomorpha. *Proceedings of the Royal Society, Series B* 275:1203-1208 (2008).

Smith ME, Singer B, Carroll A. 40 Ar/39 Ar geochronology of the Eocene Green River Formation, Wyoming. *Geological Society of America Bulletin* 115: 549–565 (2003).

Stecher G, Tamura K, Kumar S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution* (2020). doi.org/10.1093/molbev/msz312

Stevens NJ, Seiffert ER, O'Connor PM, Roberts EM, Schmitz MD, Krause C, Gorscak E, Ngasala S, Tobin L, Hieronymus TL, Temu J. Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes. *Nature* 497: 611-614 (2013).

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., Rambaut, A. (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* 4: vey016.

DISCLAIMER:

DISCLAIMER – FOR INFORMATIONAL PURPOSES ONLY; USE AT YOUR OWN RISK

The protocol content here is for informational purposes only and does not constitute legal, medical, clinical, or safety

advice, or otherwise; content added to protocols.io is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with protocols.io, can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

ABSTRACT

A Bayesian inference (BI) tree is constructed using the software BEAST v.1X (Bayesian Evolutionary Analysis Sampling Trees; Suchard et al., 2018; released on 10th June, 2018), running BEAUti, BEAST, TreeAnnotator, and FigTree, in ascending order. Before operating the BEAST software, the BEAGLE Library must be downloaded. Platform software is BEAST, and the calculation time was more than 2 hours.

1 General flow

A Bayesian inference (BI) tree is constructed using the software BEAST v.1X (Bayesian Evolutionary Analysis Sampling Trees; Suchard et al., 2018; released on 10th June, 2018), running BEAUti, BEAST, TreeAnnotator, and FigTree, in ascending order. Before operating the BEAST software, the BEAGLE Library must be downloaded. Platform software is BEAST, and the calculation time was more than 2 hours. 5 hours were needed for the present analysis.

BEAST v1.X has an important function of calibration in the associated software of BEAUti (Bayesian Evolutionary Analysis Utility). The calibration protocol of BEAUti is to input a time of the most recent common ancestor (tMRCA = minimum age) of the ingroup species, either by fossil calibration or by the geological event calibration. Using BEAST v1.X, without fixing maximum age constraint, we can simply build a reliably-dated phylogeny.

Another important function is combined gene analysis, but we applied whole mitochondrial sequence data from GenBank/DBJ. The alignment was by ClustalW in MEGA X (Stecher et al., 2020). Gapped areas of non-protein coding regions were excluded, and the consequent protein coding sequence data have totally 10,112 bp.

Calibrations points are shown on the consequent BI tree, and these dates were input in "Priors" in BEAUti; they are summarized below. See references in our paper.

Calibration point A: Fossil chimpanzee was reported from the East African Rift, and dated by the Ar-Ar method as 0.545 ± 0.003 Ma (McBrearty and Jablonski, 2005).

Calibration point B: Fossil gorilla was recently reported also from the East African Rift, and combined with the Ar-Ar method and magneto-stratigraphy, reliably dated as 8.072 ± 0.8 Ma (Kato et al., 2016).

Benton et al. (2015) reviewed many calibration data available for animals, and the following mammalian calibration points and dates are addressed by their proposal, except for calibration point K that we ourselves offered. Ages are after International Chronostratigraphic Chart by International Commission on Stratigraphy, 2013.

Calibration point C: Crown Hominoidea (86) by Benton et al. (2015). Fossil euprimate was found from Pakistan, and considered to be the Serravallian stage (12.72 ± 1.1 Ma).

Calibration point D: Crown Catarrhini (Cercopithecoidea + Hylobatidae + Hominoidea) (84) by Benton et al. (2015). The Nsungwe Formation, Tanzania, contains these mammalian fossils, and dated by U-Pb method as 24.93 ± 0.49 Ma (Roberts et al., 2010; Stevens et al. 2013).

Calibration point E: Crown primates (81) by Benton et al. (2015). Fossil euprimate was found from Morocco, and considered to be the Thanetian stage (57.6 ± 1.6 Ma).

Calibration point F: Crown Euarchontoglires (74) by Benton et al. (2015). Fossil Paromomys was found from

Montana, USA, and considered to be the Danian stage (63.8 ± 2.2 Ma).

Calibration point G: Crown Eulipotyphla (68) by Benton et al. (2015). Fossil Erinaceidae was found from Montana, USA, and considered to be the Danian stage (63.8 ± 2.2 Ma).

Calibration point H: The common ancestor of leporids (rabbits and hares) and ochotonids (pikas) (76) by Benton et al. (2015). The Indian fossil horizon was considered to be the Yepresian stage (51.96 ± 4.6 Ma) based on the foraminifer zone (Rose et al., 2008).

Calibration point I: Crown Rodentia (81) by Benton et al. (2015). Fossil Paramys was found from Montana, USA, and considered to be the Thanetian stage (57.6 ± 1.6 Ma).

Calibration point J: Intra-Murinae, Divergence of *Mus* from *Rattus* (79) by Benton et al. (2015). The fossil horizon, Pakistan, was stratigraphically, magneto-stratigraphically, oxygen isotopically dated at 10.4 Ma (Barry et al., 2002).

Calibration point K: *Orohippus* was reported from the Green River Formation, USA (Grande, 1980). Ar-Ar dating applied to the silicic tuff within the formation yields ages of 53.5 - 48.5 Ma (weighted average of 51.25 ± 0.31 Ma; Smith et al., 2003). Although *Orohippus* was evolved from equids such as *Eohippus*, we used above date for the calibration. Bat fossils, as well as primate fossils, were also reported from the Green River Formation (Grande, 1980), and bat is included in our timetree (Fig. 1).

Calibration point L: Crown Carnivora (69) by Benton et al. (2015). The fossil was from the Canadian Duchesnian with radiometric age (Robinson et al., 2004) and estimated at 37.8 ± 0.5 Ma.

Calibration point M: Crown Artiodactyla (70) by Benton et al. (2015). The fossil horizon, Pakistan, was correlated to the planktonic foraminifer zone, and 53.25 ± 0.75 Ma (Bajpai and Gingerich, 1998).

Calibration point N: Divergence of Bovinae (cow) and Antilopinae (sheep) (73) by Benton et al. (2015). The fossil horizon in the Siwalik Deposits is the Burdigalian stage (18.205 ± 2.235 Ma) based on the paleomagnetic chronology (Johnson et al., 1985).

Calibration point O: Crown Xenarthra (62) by Benton et al. (2015). Fossil Xenarthra was found from Brazil, and considered to be the Thanetian stage (57.6 ± 1.6 Ma).

Calibration point P: Crown Afrotheria (63) by Benton et al. (2015). Fossil *Eritherium* was found from Morocco, and considered to be the Thanetian stage (57.6 ± 1.6 Ma).

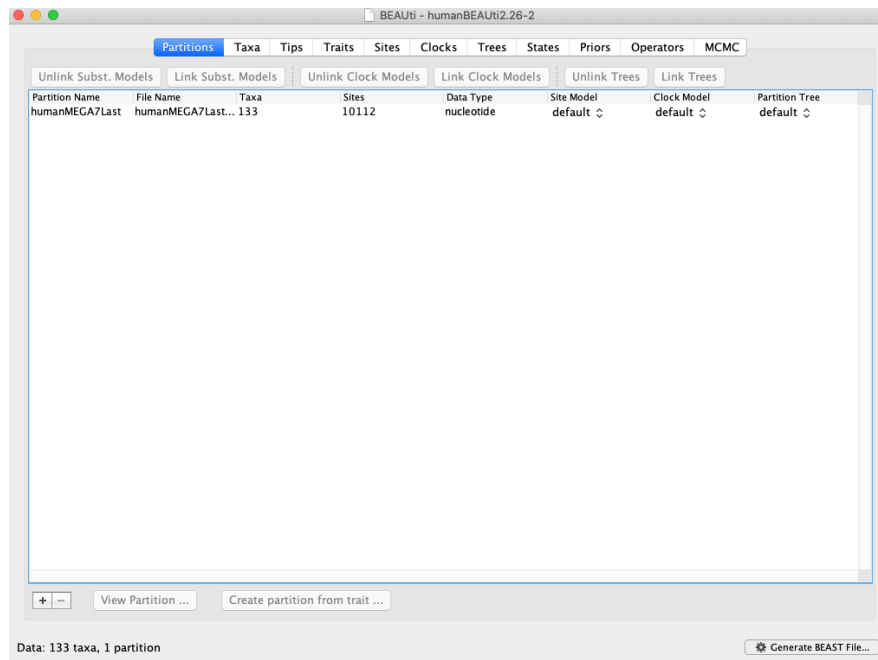
Calibration point Q: Crown Marsupialia (58) by Benton et al. (2015). Fossil *Djarthia*, Australidelphia, was found from Murgon, Australia, and considered to be the Yepresian stage (51.96 ± 4.6 Ma).

Calibration point R: Crown Theria (59) by Benton et al. (2015). *Juramaia sinensis* from Liaoning, Northeastern China, with an Ar-Ar age of 160.7 ± 0.4 Ma corresponding to the Oxfordian stage (Luo et al., 2011).

Calibration point S: Crown Mammalia (57) by Benton et al. (2015). *Ambondro mahabo* from Madagascar is placed within Monotremata, and considered to be the Bathonian stage (162.7 ± 1.1 Ma).

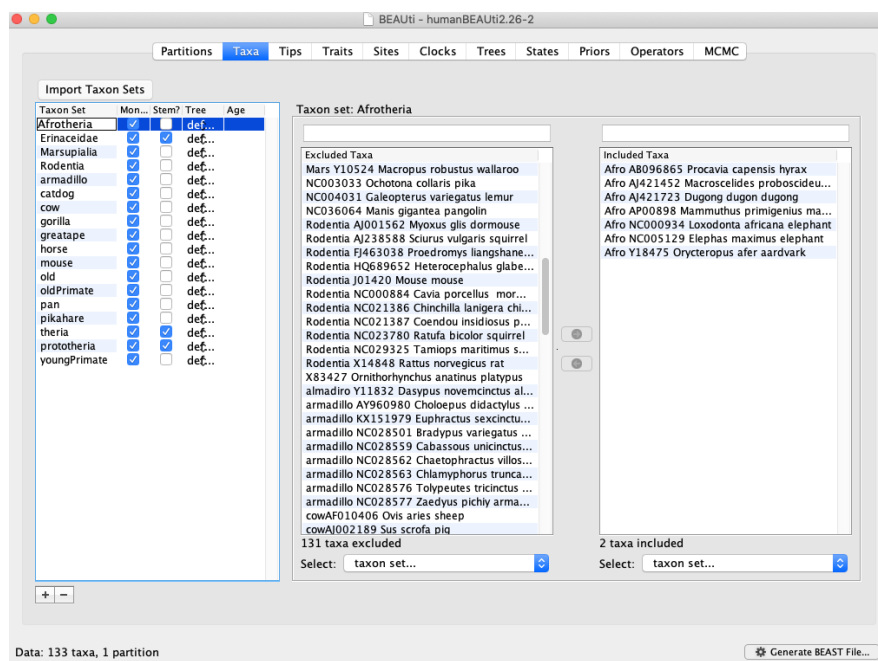
2 Prepare .txt files of whole mitochondrial data.

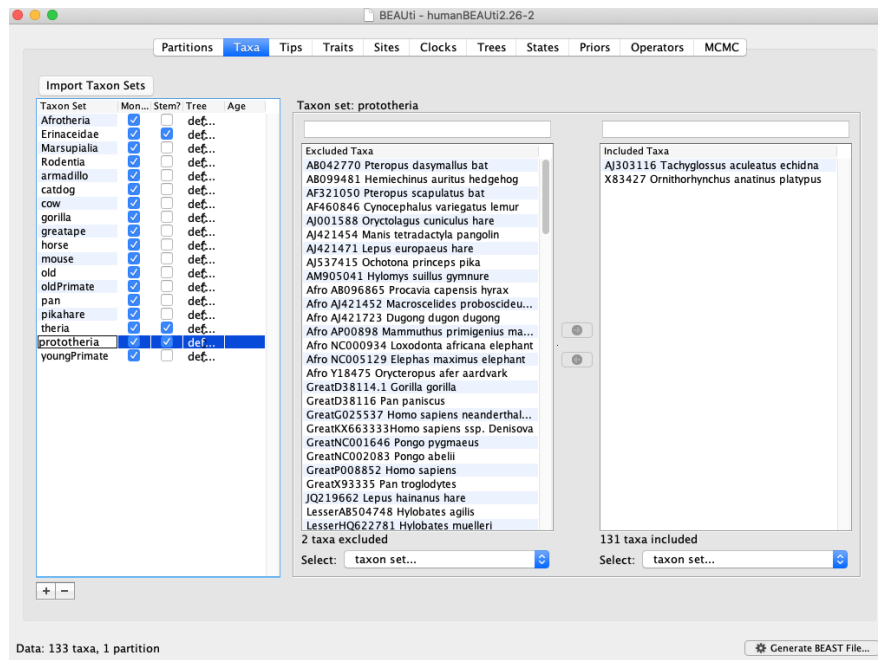
3 Partitions: Loading .text file is by using the plus button.



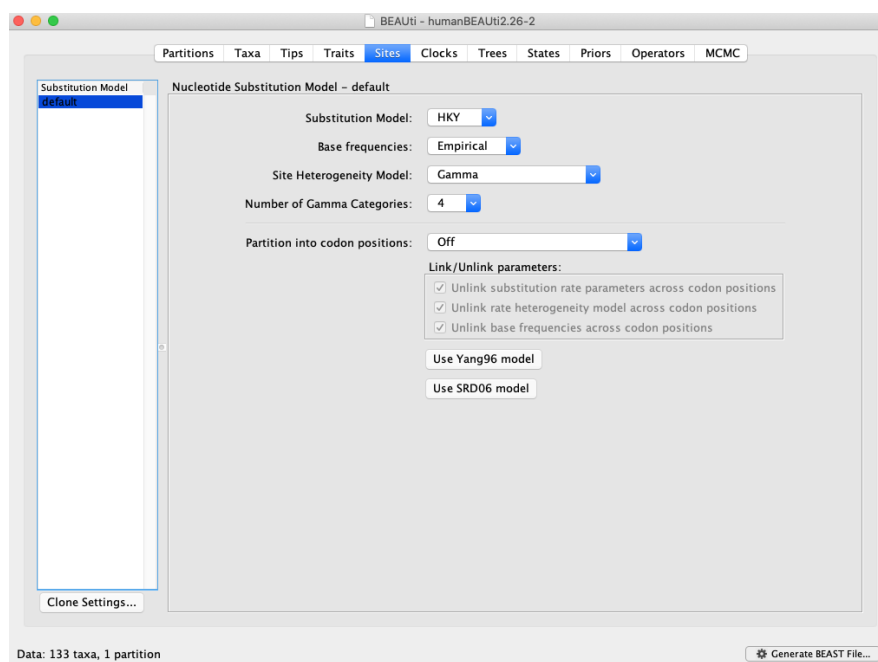
4 Taxa: Loading of taxa as ingroup was by using the plus button.

The left screen: Taxon Set (monophyletic boxes were checked for all in the present case, and stem box were checked in case by case but for Erinaceidae, Theria, and Prototheria), and the right screen: Included Taxa.



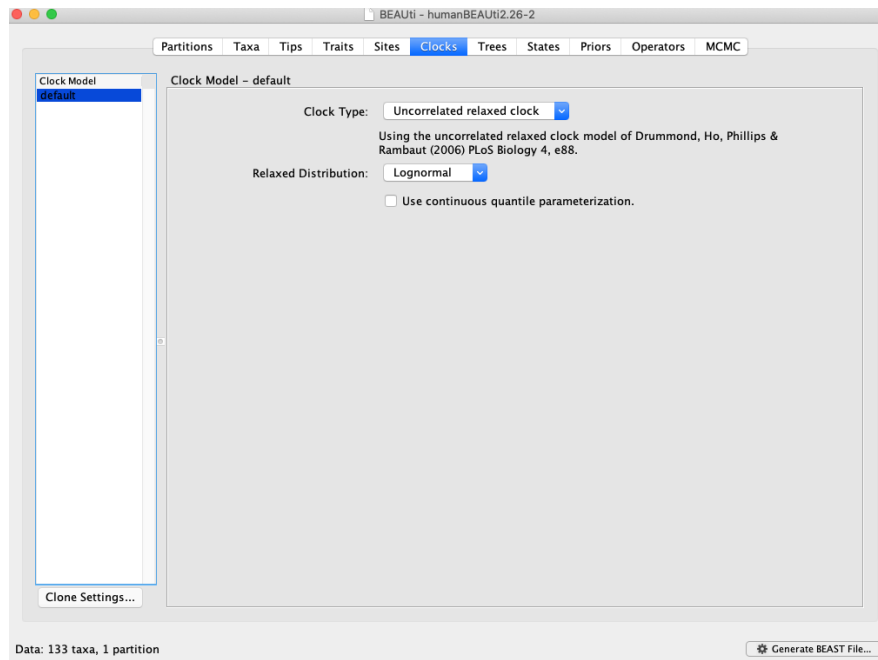


5 Sites: Substitution Model: HKY (Hasegawa, Kishino and Yano) model.

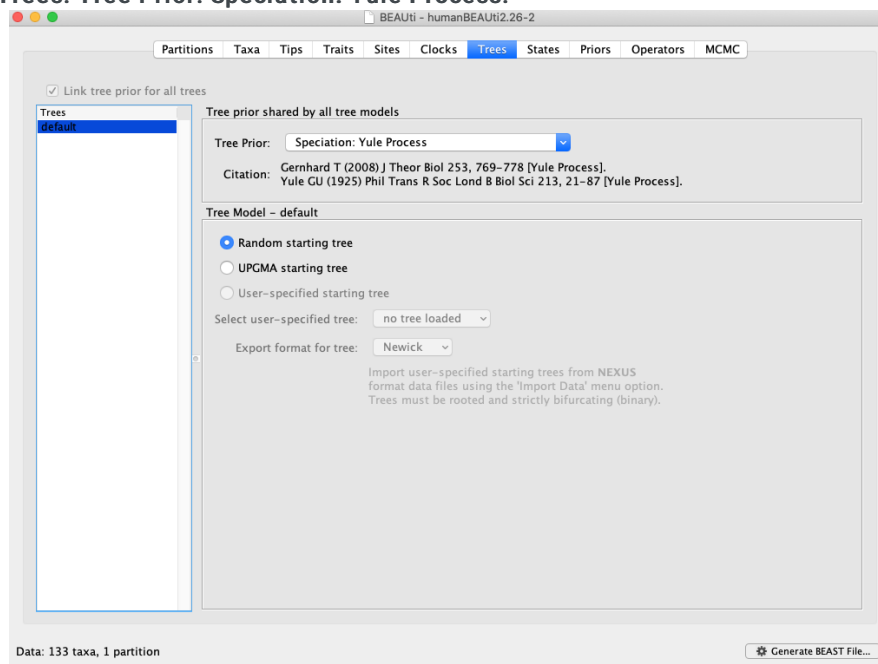


6 Clocks: Clock Type: Uncorrected relaxed clock, Relaxed Distribution: Lognormal.

Uncorrelated relaxed clocks allow each branch of a phylogenetic tree to have its own evolutionary rate under log-normal distribution, and the node rate is the rate median of three branches (Drummond *et al.*, 2006).



7 Trees: Tree Prior: Speciation: Yule Process.



8 Priors: tmrca (time of MRCA) was input from the calibration point date noted above as Prior Distribution: Normal, and the Mean and Standard deviation.

BEAUti - humanBEAUti2.26-2

Partitions Taxa Tips Traits Sites Clocks Trees States Priors Operators MCMC

☐ Use classic priors/operators

Parameter	Prior	Bound	Description
tmrca(Afrotheria)	Normal [57.6, 1.6]	n/a	tmrca statistic for taxon set Afrotheria on tree default
tmrca(Erinaceidae)	Normal [63.8, 2.2]	n/a	tmrca statistic for taxon set Erinaceidae on tree default
tmrca(Marsupialia)	Normal [54.6, 0.05]	n/a	tmrca statistic for taxon set Marsupialia on tree default
tmrca(Rodentia)	Normal [57.6, 1.6]	n/a	tmrca statistic for taxon set Rodentia on tree default
tmrca(armadillo)	Normal [57.6, 1.6]	n/a	tmrca statistic for taxon set armadillo on tree default
tmrca(catdog)	Normal [39.74, 0.07]	n/a	tmrca statistic for taxon set catdog on tree default
tmrca(cow)	Normal [53.25, 0.75]	n/a	tmrca statistic for taxon set cow on tree default
tmrca(gorilla)	Normal [8.072, 0.8]	n/a	tmrca statistic for taxon set gorilla on tree default
tmrca(greatape)	Normal [12.72, 1.1]	n/a	tmrca statistic for taxon set greatape on tree default
tmrca(horse)	Normal [51.9, 4.1]	n/a	tmrca statistic for taxon set horse on tree default
tmrca(mouse)	Normal [10.4, 0.1]	n/a	tmrca statistic for taxon set mouse on tree default
tmrca(old)	Normal [24.93, 0.49]	n/a	tmrca statistic for taxon set old on tree default
tmrca(oldPrimate)	Normal [63.8, 2.2]	n/a	tmrca statistic for taxon set oldPrimate on tree default
tmrca(pan)	Normal [0.545, 0.003]	n/a	tmrca statistic for taxon set pan on tree default
tmrca(pikahare)	Normal [51.9, 4.1]	n/a	tmrca statistic for taxon set pikahare on tree default
tmrca(theria)	Normal [160.7, 0.4]	n/a	tmrca statistic for taxon set theria on tree default
tmrca(prototheria)	Normal [162.7, 1.1]	n/a	tmrca statistic for taxon set prototheria on tree default
tmrca(youngPrimate)	Normal [57.6, 1.6]	n/a	tmrca statistic for taxon set youngPrimate on tree default
kappa	* LogNormal [1, 1.25], initial=2	[0, ∞]	HKY transition-transversion parameter
alpha	* Exponential [0.5], initial=0.5	[0, ∞]	gamma shape parameter
ucld.mean	* Approx. Reference Prior, initial=...	[0, ∞]	uncorrelated lognormal relaxed clock mean
ucld.stdev	* Exponential [0.333333], initial=...	[0, ∞]	uncorrelated lognormal relaxed clock stdev
treeModel.rootHeight	* Using Tree Prior in [0, ∞]	[0, ∞]	root height of the tree
yule.birthRate	* LogNormal [1, 1.5], initial=2	[0, ∞]	Yule speciation process birth rate

Link parameters together Link parameters into a hierarchical model Unlink parameters

* Marked parameters currently have a default prior distribution. You should check that these are appropriate.

Data: 133 taxa, 1 partition Generate BEAST File...

BEAUti - humanBEAUti2.26-2

Partitions Taxa Tips Traits Sites Clocks Trees States Priors Operators MCMC

☐ Use classic priors/operators

Parameter	Prior	Bound	Description
tmrca(Afrotheria)	Normal [57.6, 1.6]	n/a	tmrca st
tmrca(Erinaceidae)	Normal [63.8, 2.2]	n/a	tmrca st
tmrca(Marsupialia)	Normal [54.6, 0.05]	n/a	tmrca st
tmrca(Rodentia)	Normal [57.6, 1.6]	n/a	tmrca st
tmrca(armadillo)	Normal [57.6, 1.6]	n/a	tmrca st
tmrca(catdog)	Normal [39.74, 0.07]	n/a	tmrca st
tmrca(cow)	Normal [53.25, 0.75]	n/a	tmrca st
tmrca(gorilla)	Normal [8.072, 0.8]	n/a	tmrca st
tmrca(greatape)	Normal [12.72, 1.1]	n/a	tmrca st
tmrca(horse)	Normal [51.9, 4.1]	n/a	tmrca st
tmrca(mouse)	Normal [10.4, 0.1]	n/a	tmrca st
tmrca(old)	Normal [24.93, 0.49]	n/a	tmrca st
tmrca(oldPrimate)	Normal [63.8, 2.2]	n/a	tmrca st
tmrca(pan)	Normal [0.545, 0.003]	n/a	tmrca st
tmrca(pikahare)	Normal [51.9, 4.1]	n/a	tmrca st
tmrca(theria)	Normal [160.7, 0.4]	n/a	tmrca st
tmrca(prototheria)	Normal [162.7, 1.1]	n/a	tmrca st
tmrca(youngPrimate)	Normal [57.6, 1.6]	n/a	tmrca st
kappa	* LogNormal [1, 1.25], initial=2	[0, ∞]	HKY trar
alpha	* Exponential [0.5], initial=0.5	[0, ∞]	gamma
ucld.mean	* Approx. Reference Prior, initial=...	[0, ∞]	uncorrel
ucld.stdev	* Exponential [0.333333], initial=...	[0, ∞]	uncorrel
treeModel.rootHeight	* Using Tree Prior in [0, ∞]	[0, ∞]	root heig
yule.birthRate	* LogNormal [1, 1.5], initial=2	[0, ∞]	Yule spe

Link parameters together Link parameters into a hierarchical model Unlink parameters

* Marked parameters currently have a default prior distribution. You should check t

Prior for Parameter tmrca(gorilla)

Select prior distribution for tmrca(gorilla)

Prior Distribution: Normal

Mean: 8.072

Stdev: 0.8

☐ Truncate to:

Upper: +INF ∞

Lower: 0.0 -∞

Quantiles: 2.5%: 6.504
5%: 6.756
Median: 8.072
95%: 9.388
97.5%: 9.64

Cancel OK

Data: 133 taxa, 1 partition Generate BEAST File...

9 MCMC: Length of chain: 10,000,000. Default setting.

BEAUti - humanBEAUti2.26-2

Partitions Taxa Tips Traits Sites Clocks Trees States Priors Operators **MCMC**

Length of chain: 10000000

Echo state to screen every: 1000

Log parameters every: 1000

File name stem: humanMEGA7Last

☐ Add .txt suffix

Log file name: humanMEGA7Last.log

Trees file name: humanMEGA7Last.trees

☐ Create tree log file with branch length in substitutions:

Substitutions trees file name:

☐ Create operator analysis file:

Operator analysis file name:

☐ Sample from prior only - create empty alignment

Select the option below to perform marginal likelihood estimation (MLE) using path sampling (PS) / stepping-stone sampling (SS) or generalized stepping-stone sampling (GSS) which performs an additional analysis after the standard MCMC chain has finished.

Marginal likelihood estimation (MLE): None

Settings

Data: 133 taxa, 1 partition

Generate BEAST File...

10 The consequent tree was drawn by FigTree v1.4.2, for that, the tree files were input into TreeAnnotator.

TreeAnnotator v1.10.4 Prerelease #bc6cbd9

MCMC Output analysis
by
Andrew Rambaut and Alexei J. Drummond
Institute of Evolutionary Biology
University of Edinburgh
a.rambaut@ed.ac.uk
Department of Computer Science
University of Auckland
alex@cs.auckland.ac.nz

Specify the burnin as the number of states

Burnin (as states): 0

Specify the burnin as the number of trees

Burnin (as trees): 0

Posterior probability limit: 0.0

Target tree type: Maximum clade credibility tree

Node heights: Median heights

Target Tree File: not selected Choose File...

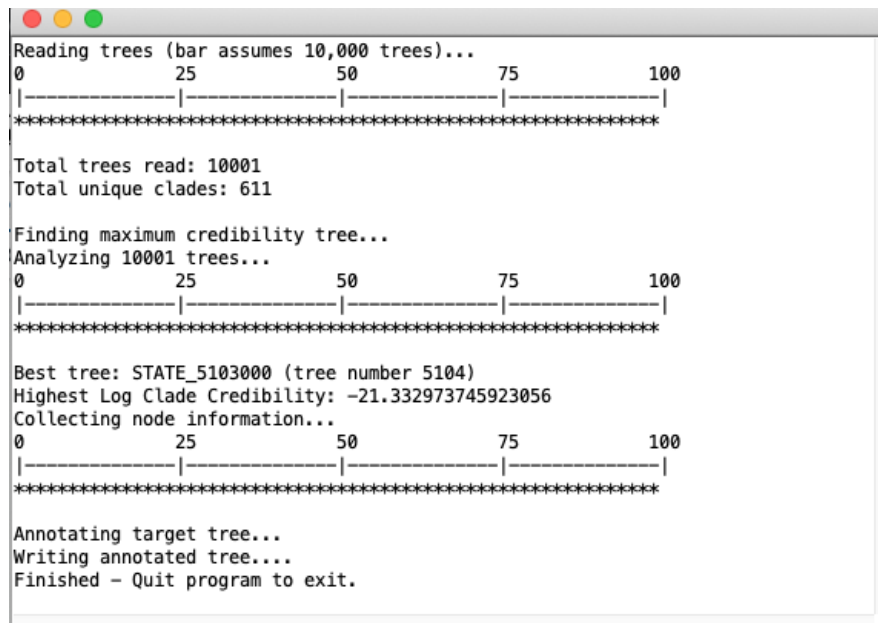
Input Tree File: humanMEGA7Last.trees Choose File...

Output File: humanAnnotator-1-13 Choose File...

Quit Run

Data: 133 taxa, 1 partition

Add figure legend here (optional)



11 FigTree

The 95% highest posterior density for confidence intervals of ages can be output in FigTree, but not shown in BI tree to avoid confusion. In FigTree, posterior probability ("posterior"), posterior age ("Node ages"), and "rate median" (not constant) can be output.



12 Adobe Illustrator