

Calibrating the Retaria molecular clock

Case study:

Genetic and morphological divergence in the
warm-water planktonic foraminifera genus
Globigerinoides



Calibrating the Retaria molecular clock



Raphael Morard

rmorard@marum.de

Senior scientist at the Center for
Marine Environmental Sciences,
Bremen, Germany



Miguel M. Sandin

miguel.mendez@ibe.upf.csic.es

Postdoctoral fellow at the Institute of
Evolutionary Biology (CSIC-UPF),
Barcelona, Spain

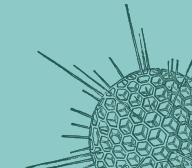


Mattia Greco

mgreco@icm.csic.es

Postdoctoral fellow at the Institute of
Marine Science (CSIC), Barcelona,
Spain

- Introduction to molecular phylogenetics and tree thinking
- Introduction to the molecular clock concept and best practices calibrating Retaria
- Case study: Calibrating *Globigerinoides*



Manuscript overview

This study investigates how genetic and morphological diversity arise within the planktonic foraminifera genus *Globigerinoides*

→ genetically distinct species can appear morphologically similar and vice versa

GOALS:

→ to clarify species boundaries

→ to explore the drivers of evolutionary divergence

PLOS ONE

RESEARCH ARTICLE

Genetic and morphological divergence in the warm-water planktonic foraminifera genus *Globigerinoides*

Raphaël Morard^{1,*}, Angelina Füllberg¹, Geert-Jan A. Brummer^{2,3}, Mattia Greco⁴, Lukas Jonkers⁵, André Wizemann⁶, Agnes K. M. Weiner^{7,8}, Kate Darling^{6,9}, Michael Siccha¹, Ronan Ledevin¹⁰, Hiroshi Kitazato¹⁰, Thibault de Garidel-Thoron¹⁰, Colombian de Vargas^{1,11}, Michal Kučera¹

¹ MARUM Center for Marine Environmental Sciences, University of Bremen, Leobener Strasse, Bremen, Germany, ² NIOZ Royal Netherlands Institute for Sea Research, Department of Ocean Systems, and Utrecht University, Den Burg, University of, The Netherlands, ³ Vrije Universiteit Amsterdam, Department of Earth Sciences, Faculty of Science, Amsterdam, The Netherlands, ⁴ Laboratoire de Troposphère et de Climatologie, Paris, France, ⁵ Department of Biological Sciences, Smith College, Northampton, Massachusetts, United States of America, ⁶ School of GeoSciences, University of Edinburgh, Edinburgh, Scotland, United Kingdom, ⁷ School of Geography and Sustainable Development, University of St Andrews, St Andrews, Scotland, United Kingdom, ⁸ Biology Department, University of, Aix-Marseille Saint Jean, Marseille, France, ⁹ Japan Agency for Marine Earth Sciences and Technology (JAMSTEC), Yokosuka, Kanagawa, Japan, ¹⁰ Aix-Marseille Université, CNRS, IRD Collège de France, INRA CEREGE, Aix-en-Provence, France, ¹¹ Sorbonne Université, CNRS, Station Biologique de Roscoff, UMR 7144, ECOMAP, Roscoff, France, ¹² Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022 Tara GOSEE, Paris, France

* morard@marum.de

OPEN ACCESS

Citation: Morard R, Füllberg A, Brummer G-JA, Greco M, Jonkers L, Wizemann A, et al. (2019) Genetic and morphological divergence in the warm-water planktonic foraminifera genus *Globigerinoides*. PLoS ONE 14(12): e0225246. <https://doi.org/10.1371/journal.pone.0225246>

Editor: Fabrizio Frontalini, Università degli Studi Urbino Carlo Bo, ITALY

Received: June 23, 2019

Accepted: October 31, 2019

Published: December 5, 2019

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0225246>

Copyright: © 2019 Morard et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All newly generated Sanger sequences are accessible on NCBI under the accession numbers MN383323–MN384218.

Abstract

The planktonic foraminifera genus *Globigerinoides* provides a prime example of a species-rich genus in which genetic and morphological divergence are uncorrelated. To shed light on the evolutionary processes that lead to the present-day diversity of *Globigerinoides*, we investigated the genetic, ecological and morphological divergence of its constituent species. We assembled a global collection of single-cell barcode sequences and show that the genus consists of eight distinct genetic types organized in five extant morphospecies. Based on morphological evidence, we reassess the species *Globoturborotalita tenella* to *Globigerinoides* and amend *Globigerinoides ruber* by formally proposing two new subspecies, *G. ruber albus* n.subsp. and *G. ruber ruber* in order to express their subspecies level distinction and to replace the informal *G. ruber "white"* and *G. ruber "pink"*, respectively. The genetic types within *G. ruber* and *Globigerinoides elongatus* show a combination of endemism and coexistence, with little evidence for ecological differentiation. CT-scanning and ontogeny analysis reveal that the diagnostic differences in adult morphologies could be explained by alterations of the ontogenetic trajectories towards final (reproductive) size. This indicates that heterochrony may have caused the observed decoupling between genetic and morphological diversification within the genus. We find little evidence for environmental forcing of either the genetic or the morphological diversification, which allude to biotic interactions such as symbiosis, as the driver of speciation in *Globigerinoides*.

PLOS ONE | <https://doi.org/10.1371/journal.pone.0225246> December 5, 2019

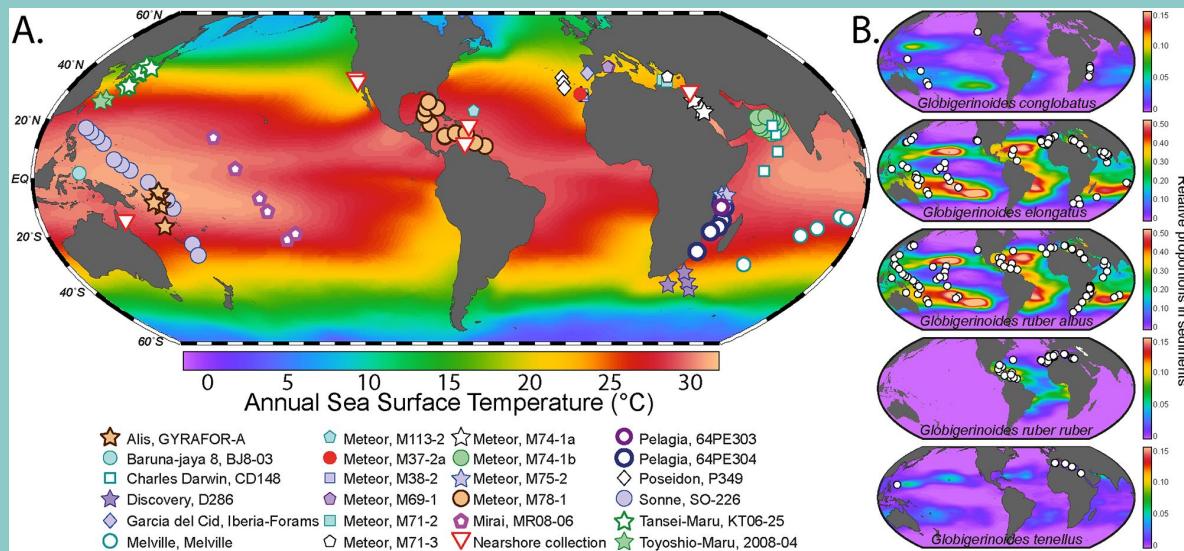
1 / 30

Manuscript overview

DATASET

Produced SSU rDNA sequences for species *G. rubescens*, *G. ruber*, *G. conglobatus*, *G. elongatus* and *G. tenellus*

Collected more sequences from **NCBI** and **PFR²**

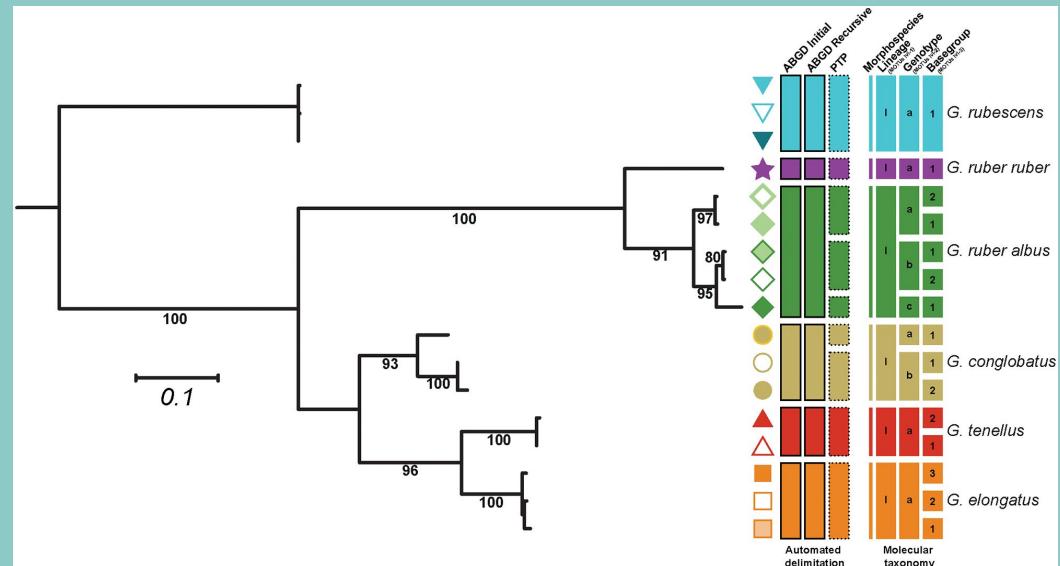


Manuscript overview

DIVERSITY *Globigerinoides*

G. rubescens= root

Genus comprises 8 distinct genotypes within 5 morphospecies

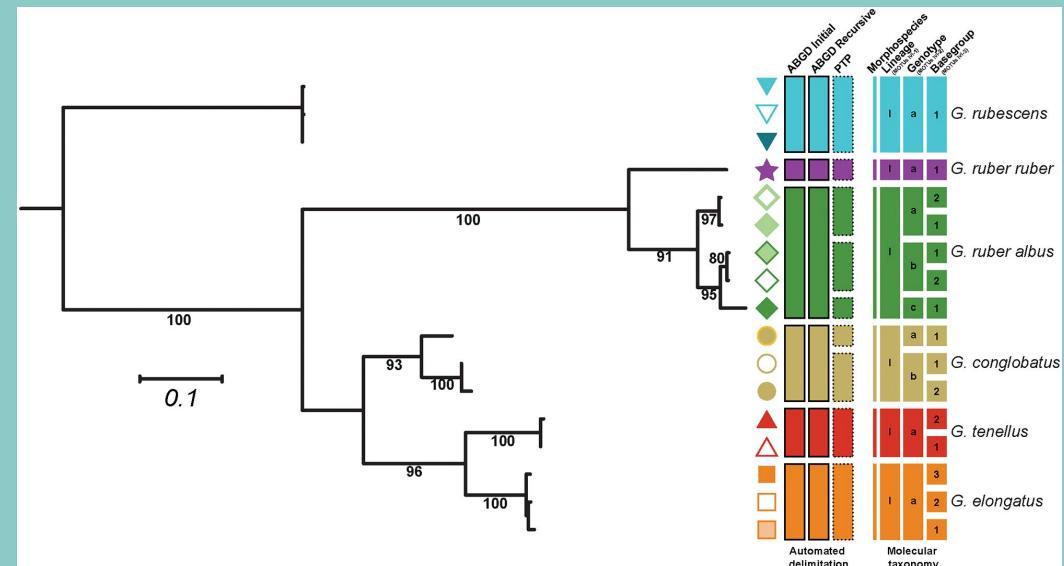
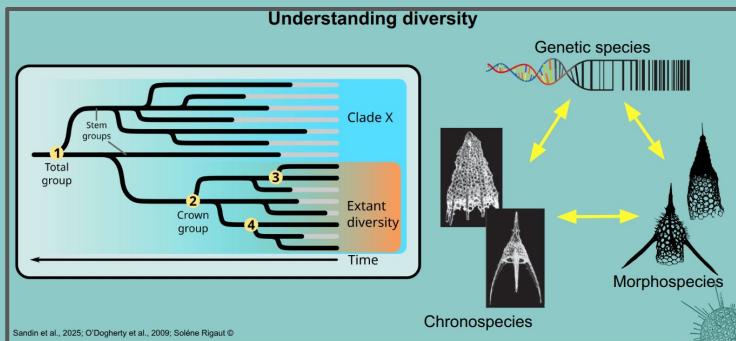


Manuscript overview

DIVERSITY *Globigerinoides*

G. rubescens= root

Genus comprises 8 distinct genotypes within 5 morphospecies



Manuscript overview

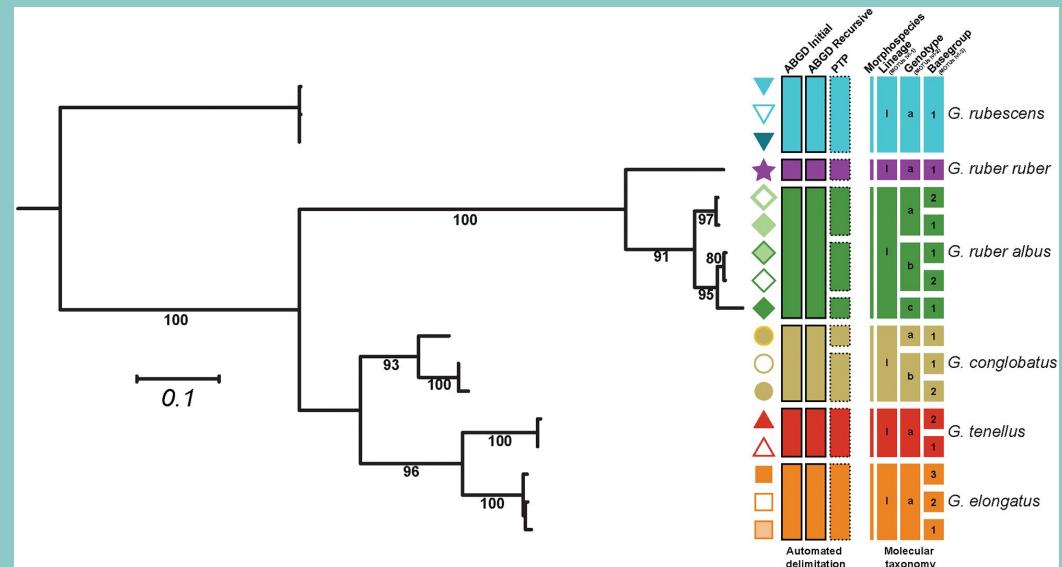
DIVERSITY *Globigerinoides*

G. rubescens= root

Genus comprises 8 distinct genotypes within 5 morphospecies



**WHEN DID
DIVERSIFICATION
OCCUR?**



Molecular clock

Phylogeny and molecular clock

To reconstruct the evolutionary history of the genus *Globigerinoides*, we applied a molecular clock estimation using the same alignment as for the maximum likelihood tree inference ([Fig 2](#)). We used the divergence between *G. rubescens* and the genus *Globigerinoides* (23.8 Ma [[48](#)]), the First Appearance Datum (FAD) of *G. conglobatus* (8–8.6 Ma) and *G. tenellus* (2.5 Ma), which are known from the fossil record [[49](#)], as minimum ages to constrain the phylogeny. We used a relaxed clock model implemented in BEAST v.1.8.4 [[50](#)]. Model parameters were set using BEAUti v1.8.4. The distribution of the fixed node age prior was considered normal and the speciation rate was assumed constant under the Yule-Process. The GTR (Generalised Time Reversible) model was selected as substitution model and an UPGMA (Unweighted Pair Group method with arithmetic mean) tree was calculated as starting tree. Markov-Chain-Monte Carlo (MCMC) analyses were conducted for 10,000,000 generations, with a burn-in of 1000 generations and saving each 1000th generation. The maximum clade credibility tree with median node heights was calculated in TREEAnnotator from the BEAST package, with a burn-in of 100 trees and a posterior probability limit of 0. The resulting tree was then visualized in FigTree v. 1.3.1 [[51](#)] and is shown in [Fig 7](#).

Molecular clock

PROGRAMS USED:

Bayesian Evolutionary Analysis Sampling Trees

-A collection of programs for performing Bayesian

MCMC analysis of molecular sequences -

- BEAUti - Creating XML input files
- BEAST - MCMC analysis of molecular sequences
- Tracer - Viewing MCMC output
- LogCombiner - Combining output files
- TreeAnnotator - Generate the consensus tree
- FigTree - Drawing a tree

Programs distributed as part of the core BEAST package



BEAUti | Bayesian Evolutionary Analysis Utility. This program is used to import data, design the analysis, and generate the BEAST control file.



TreeAnnotator | This is a post-analysis program that will produce a summary tree from the output of BEAST.



LogCombiner | This is a utility program that will combine log files from different runs and reduce the sampling frequency (thin them).

These programs are distributed independently but are important part of the BEAST toolkit.



Tracer | This is a graphical program for exploring the output of BEAST, diagnosing problems, and summarizing the results.



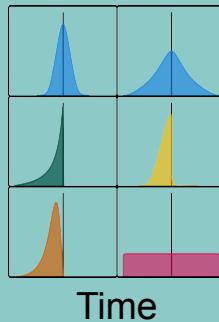
FigTree | This is a graphical program for viewing trees, displaying summary information from TreeAnnotator and generating publication-ready figures.

Molecular clock

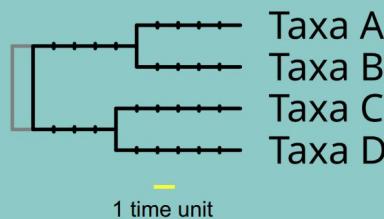
WORKFLOW:

1. Input: Data + Model

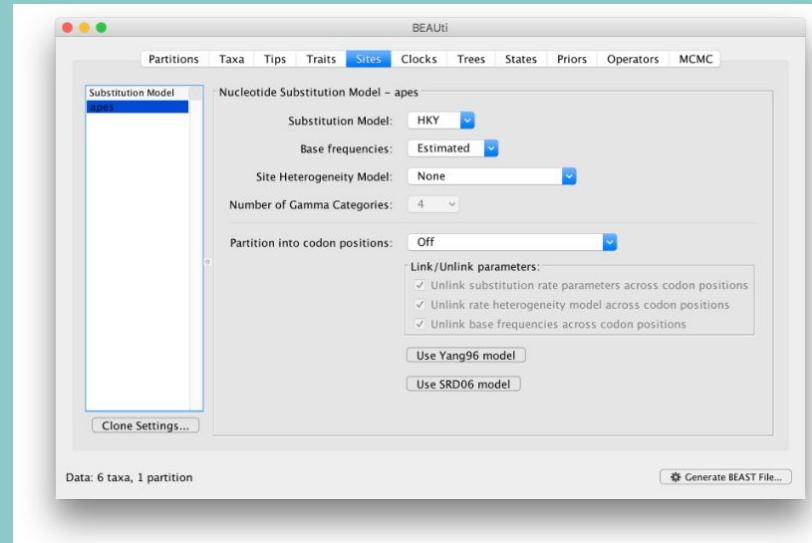
- A Sequence alignment
- A BEAST XML config file (defines:
 - Substitution model
 - Clock model (strict, relaxed, etc.)
 - Tree prior
 - Priors on parameters
 - MCMC settings



Time



XML file → specially formatted text file
that contains all the information needed
to run a Bayesian phylogenetic analysis
using BEAST.



https://beast.community/constructing_models

Molecular clock

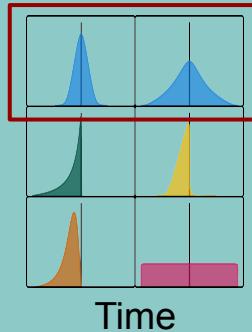
WORKFLOW:

1. Input: Data + Model

- A DNA alignment
- A BEAST XML config file (defines:
 - Substitution model ([GTR](#))
 - Clock model ([relaxed](#))
 - Tree prior ([Yule +UPGMA](#))
 - Priors on parameters ([FAD](#))
 - MCMC settings
 - i. 10M generations
 - ii. burn-in 1k generations

Phylogeny and molecular clock

To reconstruct the evolutionary history of the genus *Globigerinoides*, we applied a molecular clock estimation using the same alignment as for the maximum likelihood tree inference (Fig 2). We used the divergence between *G. rubescens* and the genus *Globigerinoides* (23.8 Ma [48]), the First Appearance Datum (FAD) of *G. conglobatus* (8–8.6 Ma) and *G. tenellus* (2.5 Ma), which are known from the fossil record [49], as minimum ages to constrain the phylogeny. We used a relaxed clock model implemented in BEAST v.1.8.4 [50]. Model parameters were set using BEAUTi v1.8.4. The distribution of the fixed node age prior was considered normal and the speciation rate was assumed constant under the Yule-Process. The GTR (Generalised Time Reversible) model was selected as substitution model and an UPGMA (Unweighted Pair Group method with arithmetic mean) tree was calculated as starting tree. Markov-Chain-Monte Carlo (MCMC) analyses were conducted for 10,000,000 generations, with a burn-in of 1000 generations and saving each 1000th generation. The maximum clade credibility tree with median node heights was calculated in TREEAnnotator from the BEAST package, with a burn-in of 100 trees and a posterior probability limit of 0. The resulting tree was then visualized in FigTree v. 1.3.1 [51] and is shown in Fig 7.



Molecular clock

WORKFLOW:

2. Run BEAST

This launches the **MCMC process**, which:

- Iteratively proposes new trees, rates, times, etc.
- Accepts/rejects them based on posterior probability
- Stores accepted samples every N steps

#	BEAST v1.X				
	# Generated Sun Jul 23 17:42:37 BST 2017 [seed=1500828054875]				
state	Posterior	Prior	Likelihood	rootAge	
0	-6812.7534	-481.7973	-6330.9561	102.559	-
10000	-2910.2955	-246.4611	-2663.8344	13.7383	-
20000	-2082.7664	-234.0125	-1848.7539	0.12112	-
30000	-2049.3031	-230.0213	-1819.2818	5.84813E-2	-
40000	-2043.1005	-226.3604	-1816.7401	6.21508E-2	-
50000	-2042.1593	-226.1915	-1815.9678	5.79924E-2	-
.					
.					
.					

https://beast.community/first_tutorial

Molecular clock

WORKFLOW:

3. Outputs: Posterior Samples

BEAST writes two key files:

- *.log file → Posterior samples of parameters (rates, dates, likelihood, priors, etc.)
- *.trees file → Posterior samples of full trees with divergence times

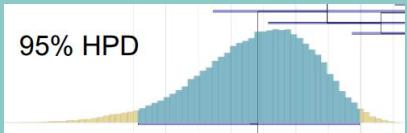
Molecular clock

WORKFLOW:

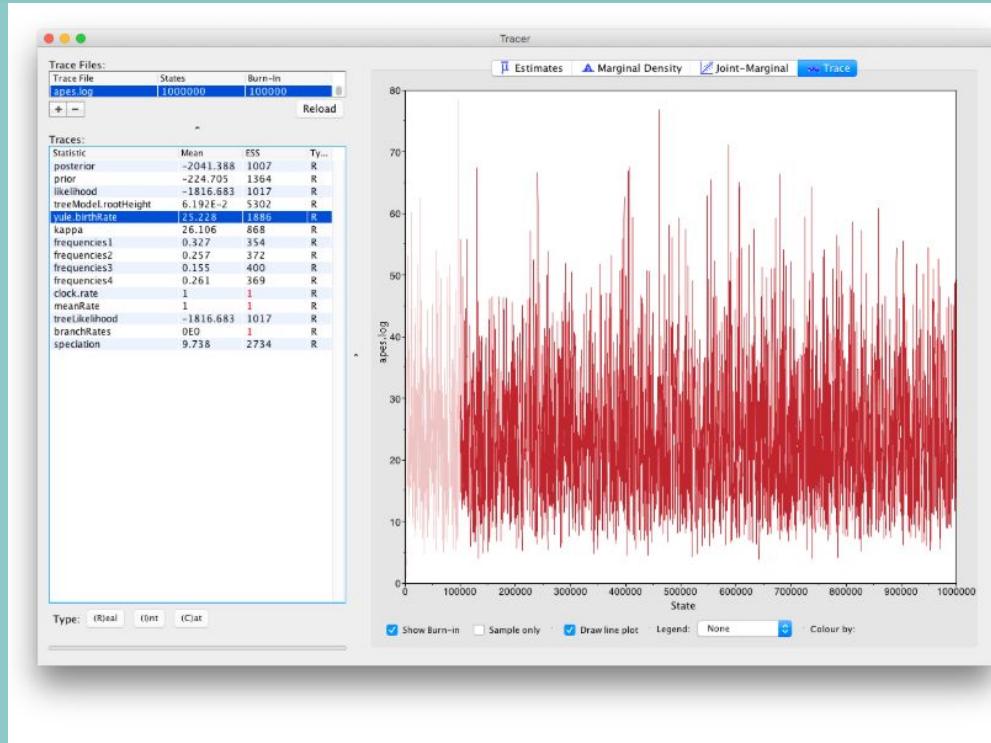
4. Post-processing: Analyze the Posterior

- **Tracer** → to plot trace plots, histograms, and get **mean**, **median**, **95% HPD**, and **ESS** values.

HPD Highest Posterior Density interval



ESS Estimated Sample Size

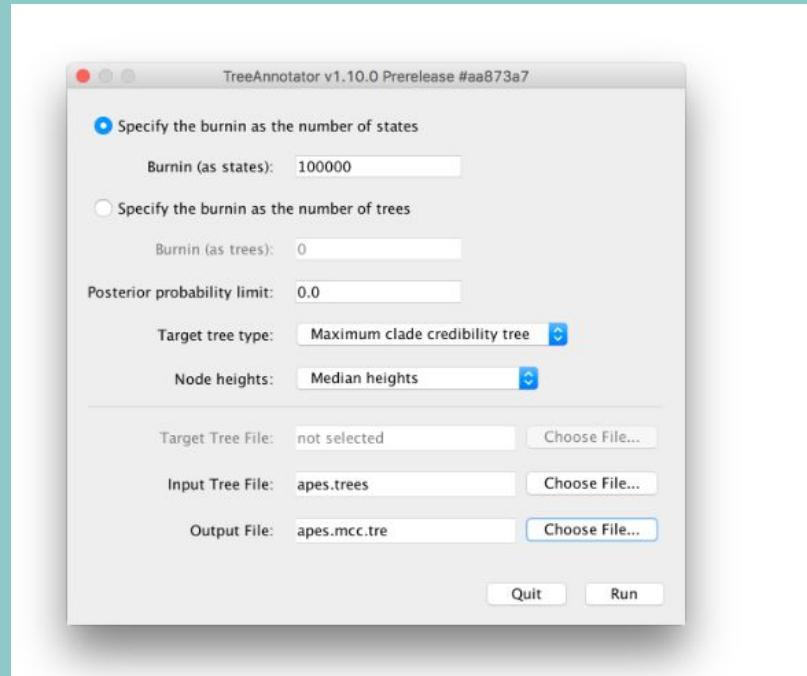


Molecular clock

WORKFLOW:

4. Post-processing: Analyze the Posterior

- **Tracer** → to plot trace plots, histograms, and get **mean, median, 95% HPD, and ESS** values.
- **TreeAnnotator** → to summarize the **posterior tree distribution** into a **maximum clade credibility (MCC) tree** with 95% HPD intervals on node ages.



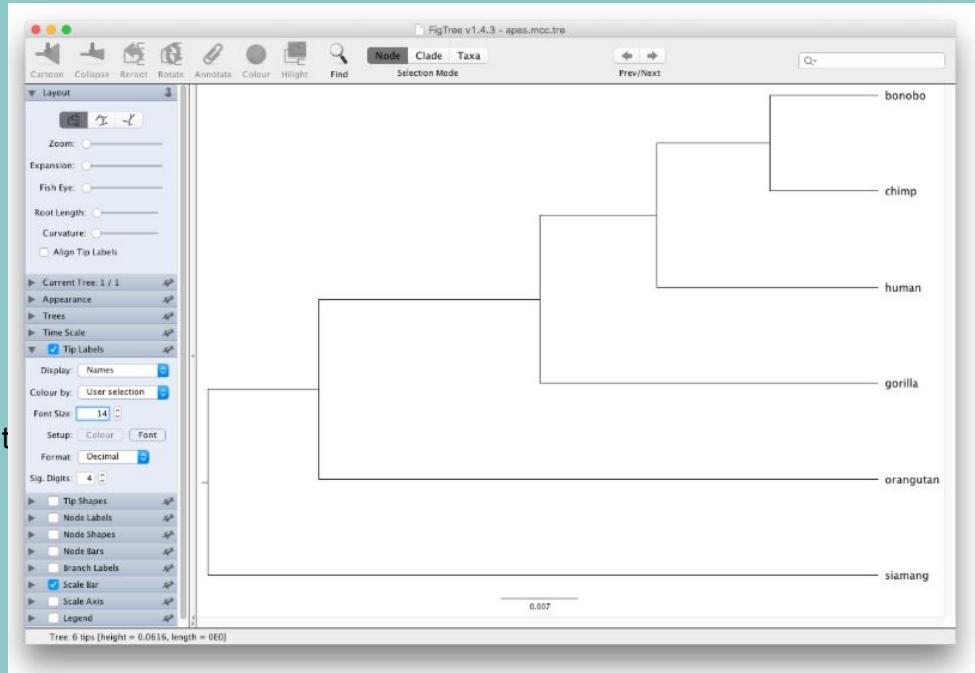
https://beast.community/second_tutorial

Molecular clock

WORKFLOW:

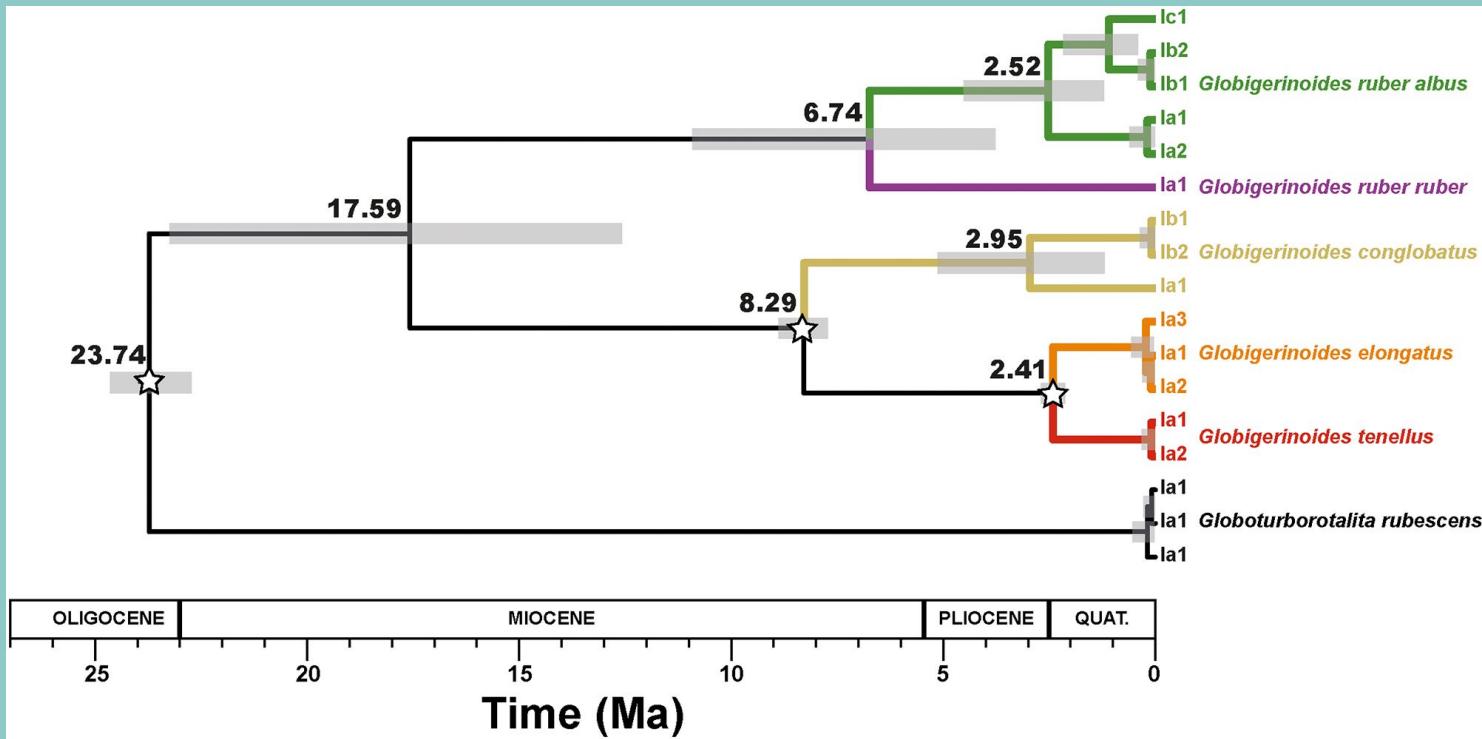
4. Post-processing: Analyze the Posterior

- **Tracer** → to plot trace plots, histograms, and get **mean**, **median**, **95% HPD**, and **ESS** values.
- **TreeAnnotator** → to summarize the **posterior tree distribution** into a **maximum clade credibility (MCC)** tree with 95% HPD intervals on node ages.
- **FigTree** → to visualize the tree with posterior node height and supports.



https://beast.community/second_tutorial

Molecular clock



The Life of Retaria Seminar Series

A meeting to foster exchange among the Retaria community