

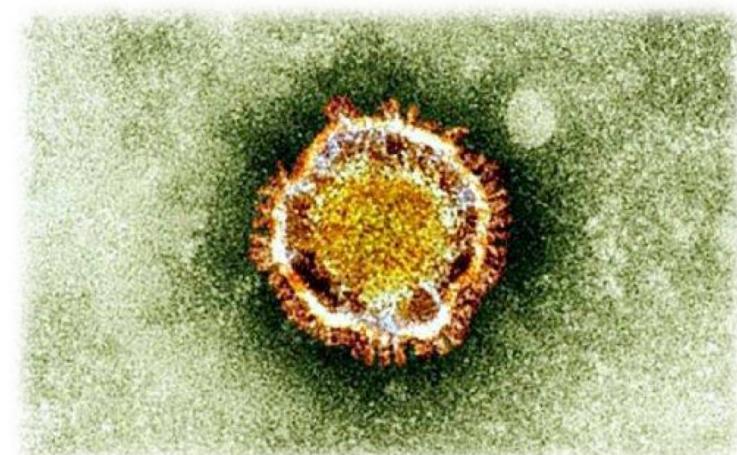
Short Course on Infectious Disease Modeling (2019, SPH Fudan):
Phylogenetics Practical | 14 Aug 2019

Case Study: Middle East Respiratory Syndrome Coronavirus

Tommy Lam

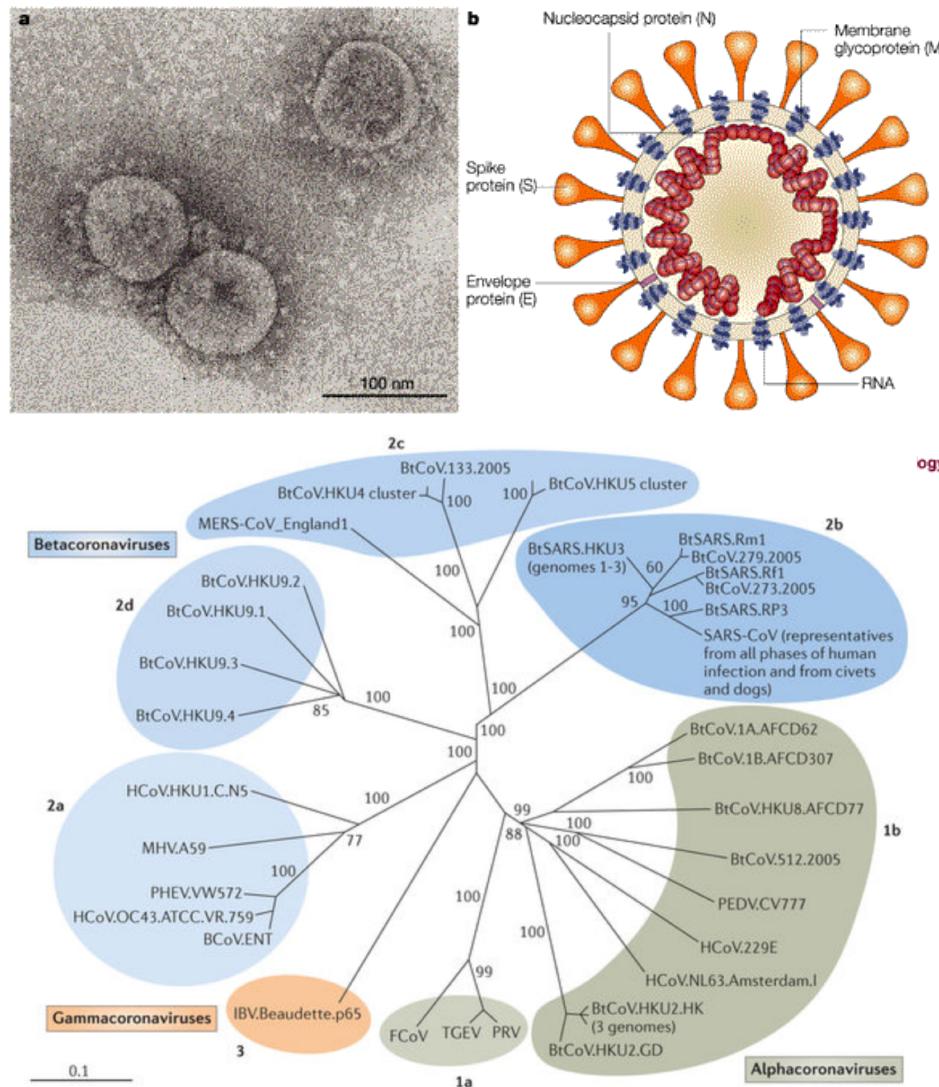
ttylam@hku.hk

State Key Laboratory of Emerging Infectious Diseases
School of Public Health
The University of Hong Kong

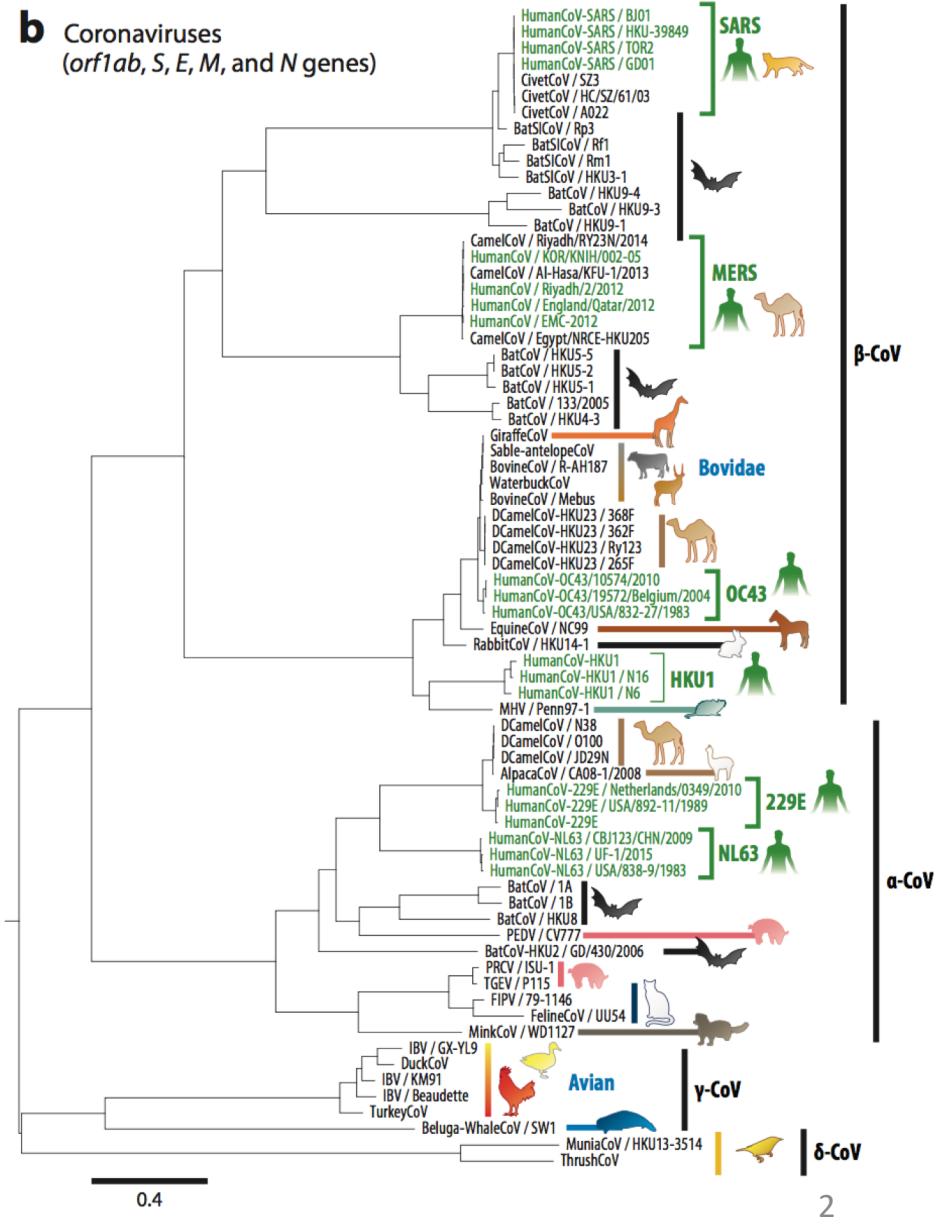


Coronavirus

- +ssRNA virus
- Group 1-4: alpha, beta, gamma, delta
- Many different host species

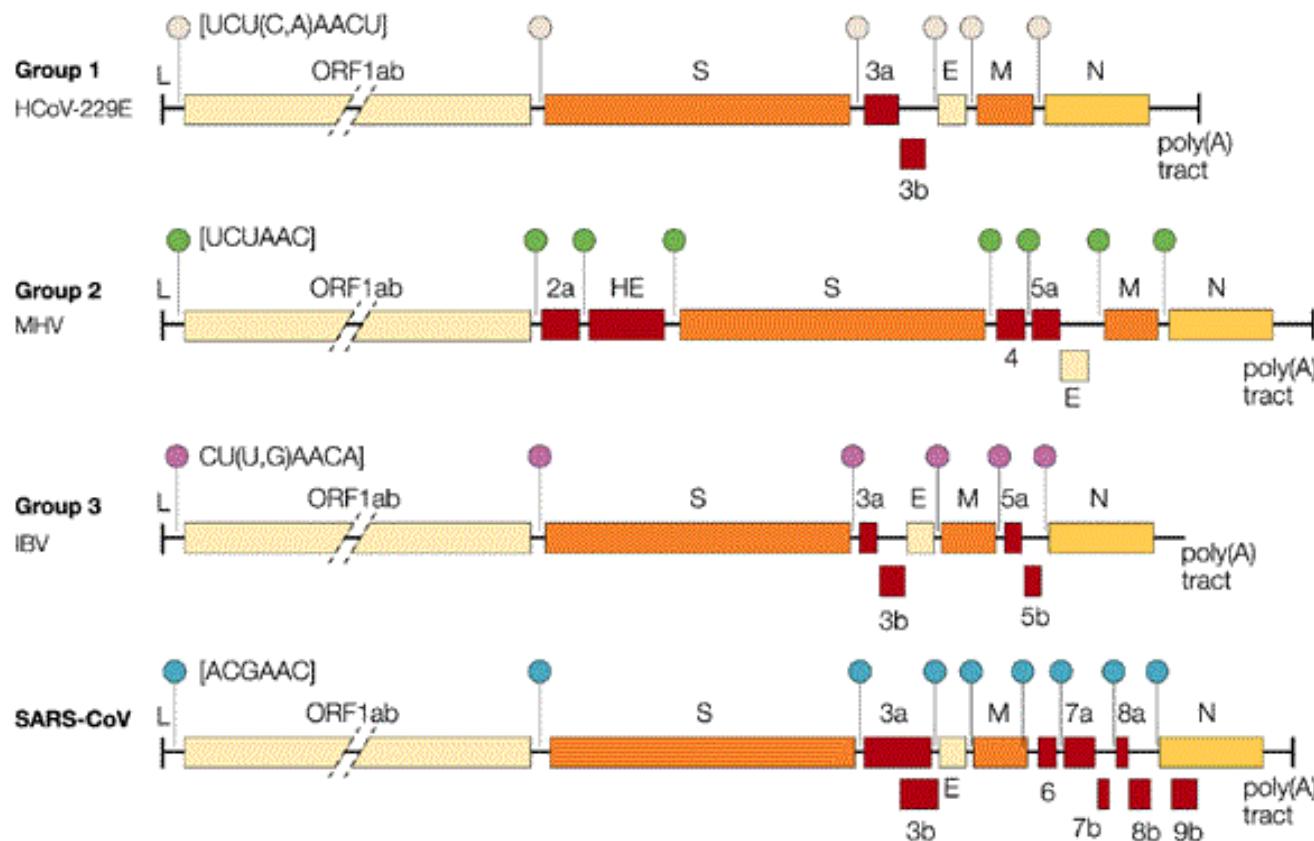


Lam et al. 2016 Annu Rev Genomics Hum Genet doi:10.1146/annurev-genom-083115-022628.



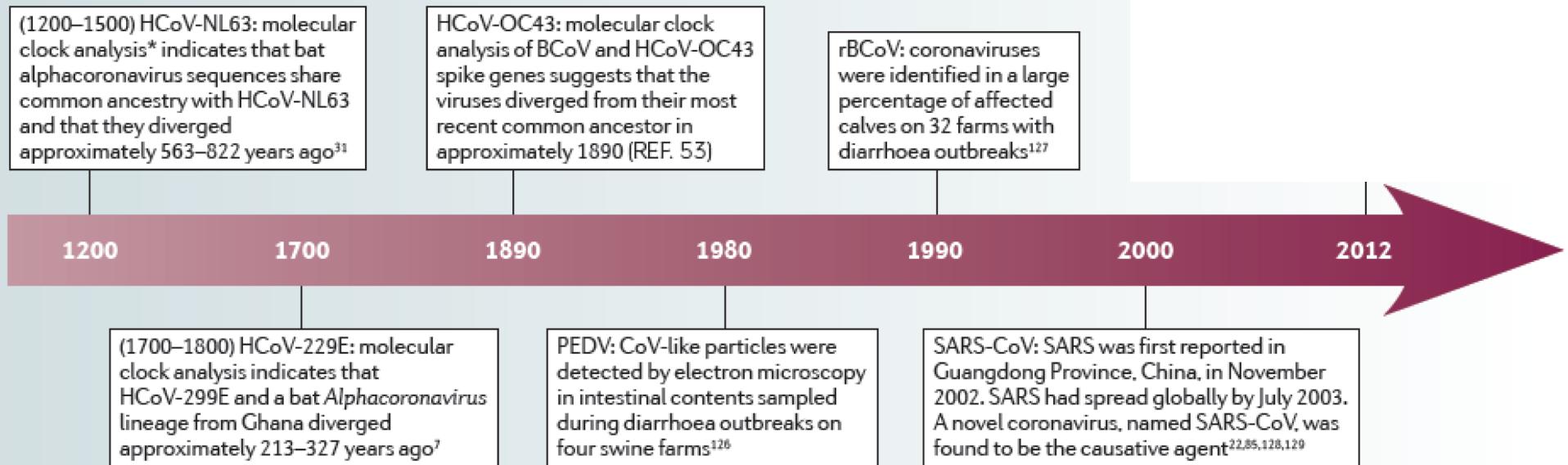
Coronavirus Genome

Common genes: ORF1ab, S, E, M, N



History of Coronavirus Emergence

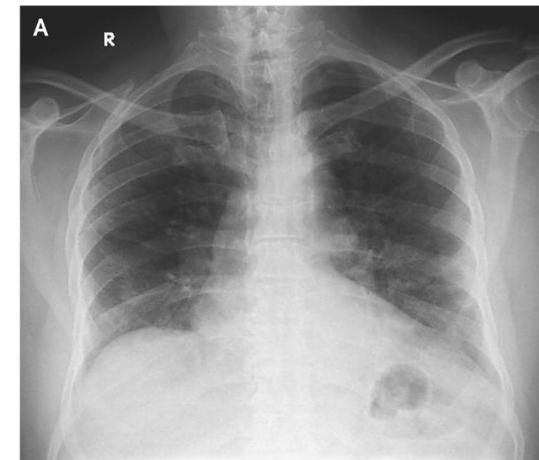
Timeline | Timeline of coronavirus emergence events



BCoV, bat coronavirus; HCoV, human coronavirus; MERS-CoV, Middle East respiratory syndrome coronavirus; PEDV, porcine epidemic diarrhea virus; rBCoV, respiratory bat coronavirus; SARS, severe acute respiratory syndrome; SARS-CoV, severe acute respiratory syndrome coronavirus. *Molecular clock analysis estimates the time at which two gene sequences diverged from each other based on known factors, such as mutation rate.

June 2012

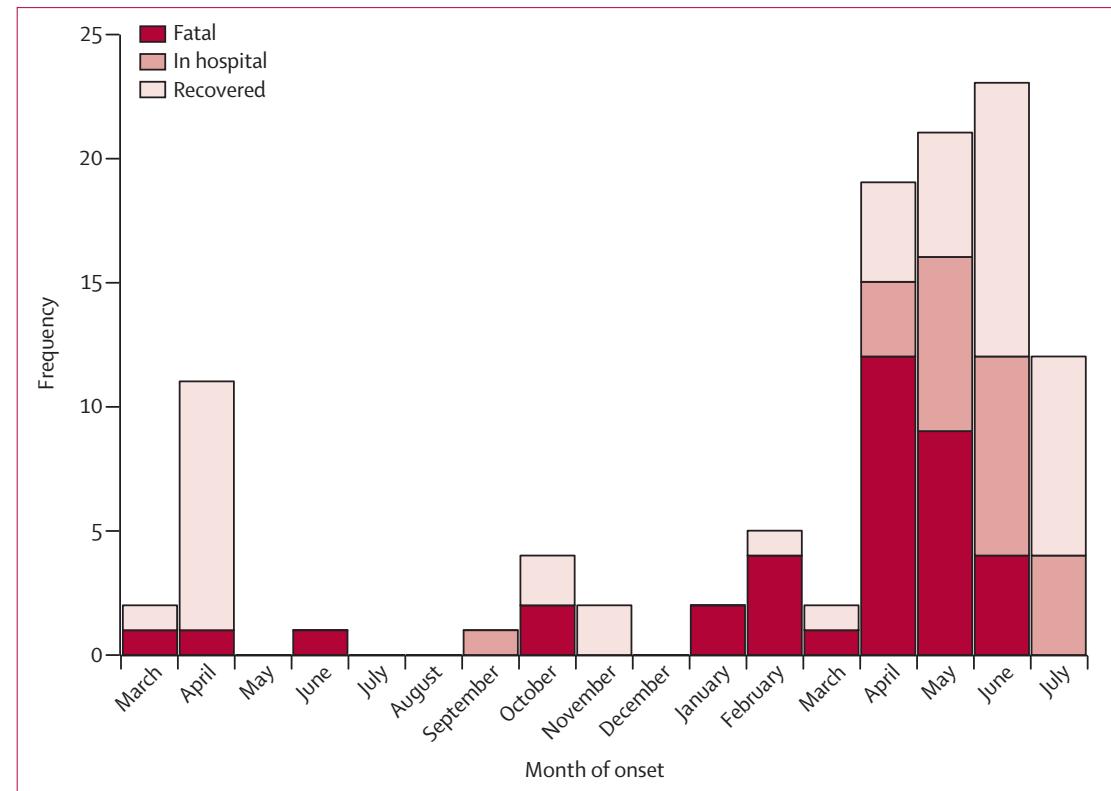
- 60 yo Saudi Arabia man
- Admitted to hospital in Jeddah
13 June 2012
- Fatal pneumonia and renal failure
- Sputum samples obtained
- Detection of the pan-coronavirus PCR
- PCR fragment sequenced



MERS-CoV

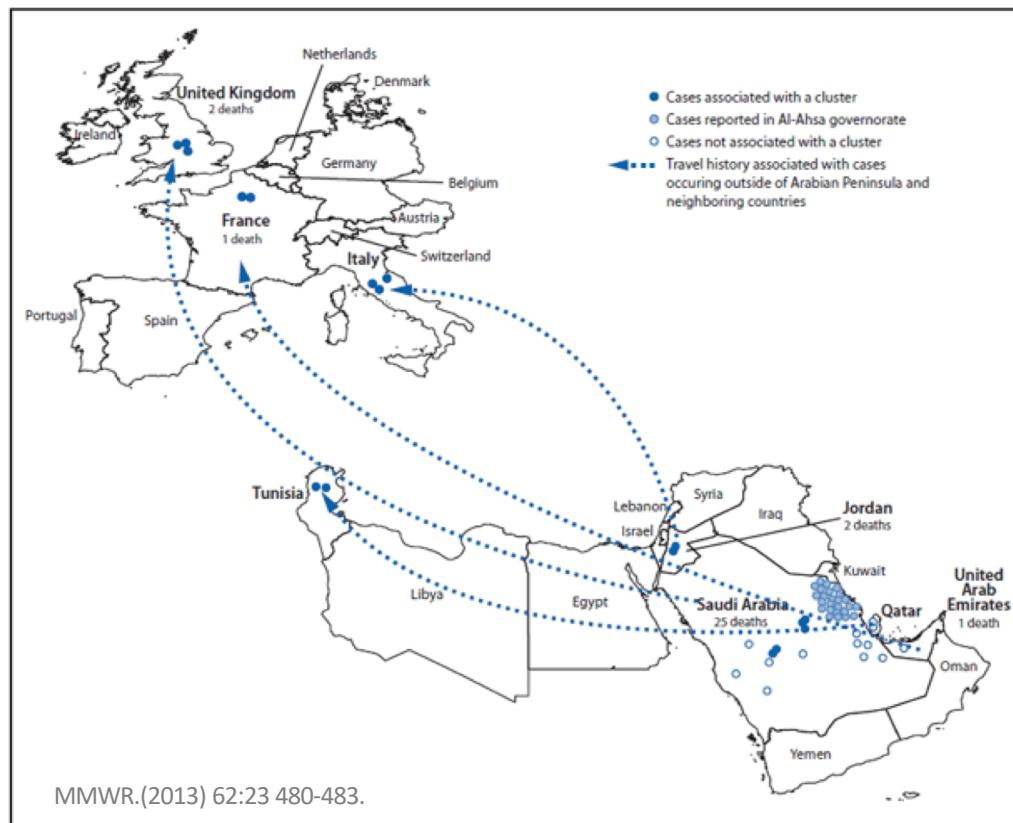
- Middle East Respiratory Syndrome Coronavirus
- Officially announced in September 2012
- Archived samples trace back to March
- **2,040 cases**
- **712 deaths**

(as of 19 July)



Spread of MERS-CoV

- Mainly affect Arabian Peninsula countries including Saudi Arabia, Qatar, Jordan, etc.
- Also exported to Europe, United States, South Korea, China etc.
- Now **27** counties have reported MERS.



MERS-CoV in Camels

- Found viral RNA in dromedary camels
- Very genetically close (~99.97%, full genome) to MERS-CoV from human

Lancet Infectious Diseases, Volume 14, Issue 2, Pages 140 - 145, February 2014

Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation

Bart L Haagmans*, Said H S Al Dhahiry*, Chantal B E M Reusken*, V Stalin Raj*, Monica Galiano, Richard Myers, Gert-Jan Godeke, Marcel Jonges, Elmoumasher Farag, Ayman Diab, Hazem Ghobashy, Farhoud Alhajri, Mohamed Al-Thani, Salih A Al-Marri, Hamad E Al Romaihi, Abdullatif Al Khal, Alison Bermingham, Albert D M E Osterhaus, Mohd M AlHajri, Marion P G Koopmans

Summary

Background Middle East respiratory syndrome coronavirus (MERS-CoV) causes severe lower respiratory tract infection in people. Previous studies suggested dromedary camels were a reservoir for this virus. We tested for the presence of MERS-CoV in dromedary camels from a farm in Qatar linked to two human cases of the infection in October, 2013.

Methods We took nose swabs, rectal swabs, and blood samples from all camels on the Qatari farm. We tested swabs with RT-PCR, with amplification targeting the *E* gene (up*E*), nucleocapsid (*N*) gene, and open reading frame (ORF) 1a. PCR positive samples were tested by different MERS-CoV specific PCRs and obtained sequences were used for phylogenetic analysis together with sequences from the linked human cases and other human cases. We tested serum samples from the camels for IgG immunofluorescence assay, protein microarray, and virus neutralisation assay.

Findings We obtained samples from 14 camels on Oct 17, 2013. We detected MERS-CoV in nose swabs from three camels by three independent RT-PCRs and sequencing. The nucleotide sequence of an ORF1a fragment (940 nucleotides) and a 4·2 kb concatenated fragment were very similar to the MERS-CoV from two human cases on the same farm and a MERS-CoV isolate from Hafr-Al-Batin. Eight additional camel nose swabs were positive on one or more RT-PCRs, but could not be confirmed by sequencing. All camels had MERS-CoV spike-binding antibodies that correlated well with the presence of neutralising antibodies to MERS-CoV.

Interpretation Our study provides virological confirmation of MERS-CoV in camels and suggests a recent outbreak affecting both human beings and camels. We cannot conclude whether the people on the farm were infected by the camels or vice versa, or if a third source was responsible.

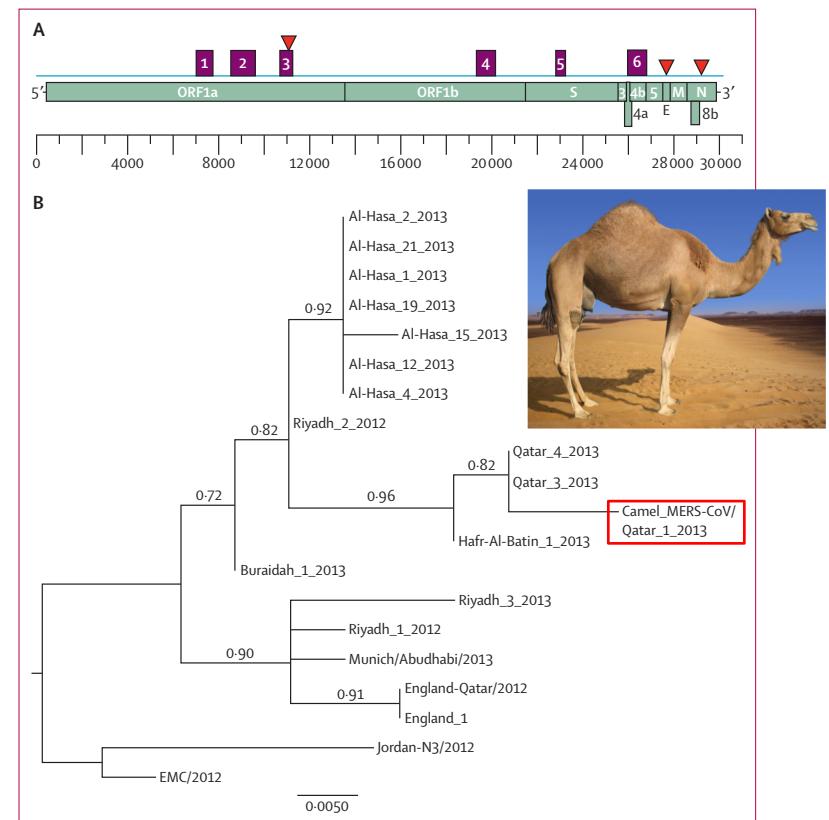


Figure 1: Characterisation of dromedary camel Middle East respiratory syndrome coronavirus (MERS-CoV) genome sequences

(A) Nucleotide sequence fragments obtained from the nose swab of camel 5 that cover different parts of the MERS-CoV genome; purple boxes (1–6) show different fragments obtained with primers specific to MERS-CoV; red triangles show the position of the real-time TaqMan probes targeting different regions of the MERS-CoV genome. (B) Nucleotide sequences of representative MERS-CoVs' concatenated 4·2 kb sequences were analysed and a phylogenetic tree was constructed by the PhyML method; values at the branches show the result of the approximate likelihood-ratio, with values less than 0·70 not depicted.

MERS-CoV in Camels

- Found neutralizing antibodies widely present in dromedary camels in Arabian peninsula and North Africa.

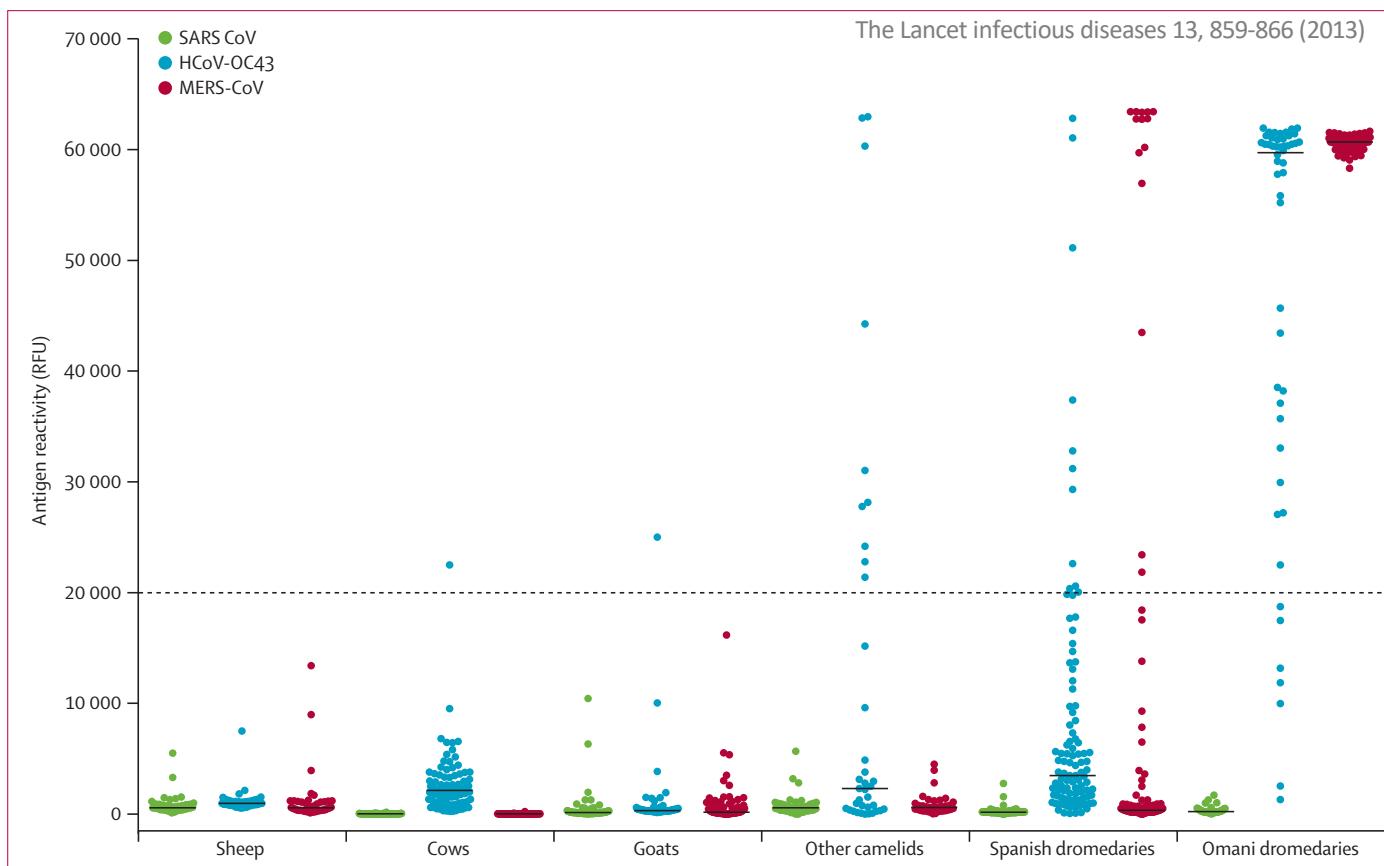


Figure 1: Reactivity of livestock sera with three coronavirus S1 antigens

Fluorescent intensities per antigen at a serum dilution of 1/20. Black lines indicate median. Dashed line is cutoff of the assay. RFU=relative fluorescence units.

SARS-CoV=severe acute respiratory syndrome coronavirus. HCoV=human coronavirus. MERS-CoV=Middle East respiratory syndrome coronavirus.

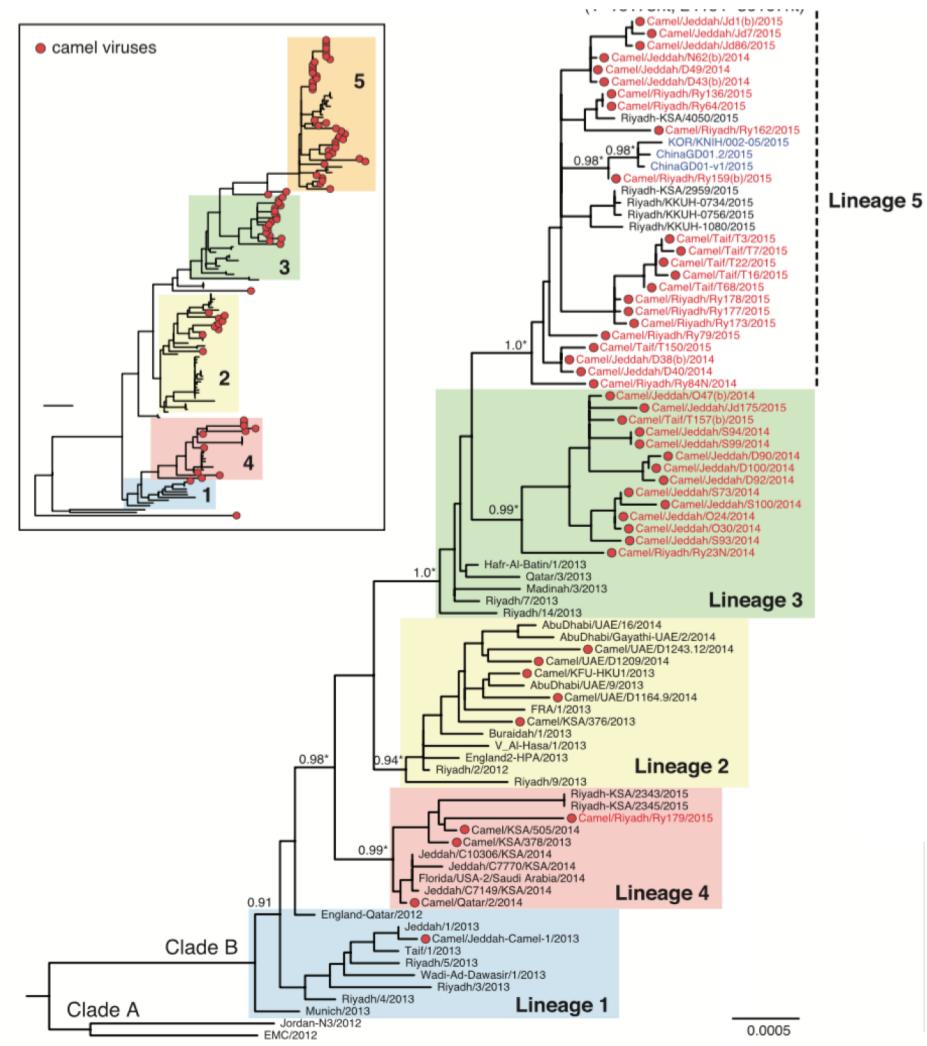
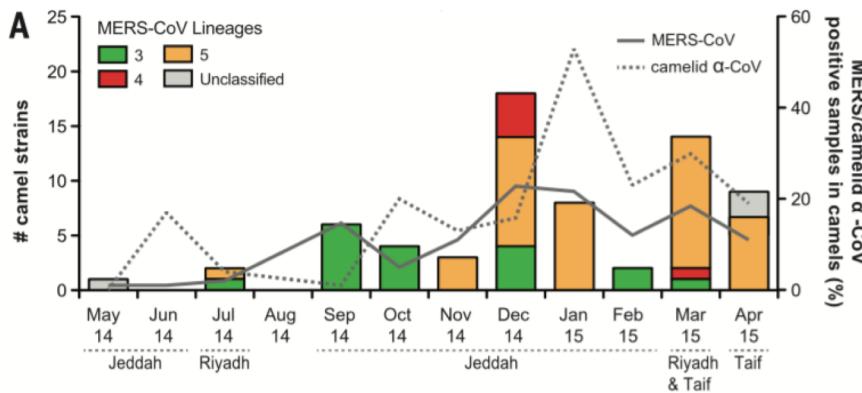
MERS-CoV in Camels

- Wide and year-round circulation in dromedary camels in Saudi Arabia

VIROLOGY Sabir JS, Lam TT, et al. Science. 2016;351(6268):81-4.

Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia

Jamal S. M. Sabir,^{1,*} Tommy T.-Y. Lam,^{2,3,4,*} Mohamed M. M. Ahmed,^{1,6*} Lifeng Li,^{3,4*} Yongyi Shen,^{3,4} Salah E. M. Abo-Aba,^{1,7} Muhammad I. Qureshi,¹ Mohamed Abu-Zeid,^{1,7} Yu Zhang,^{2,3,4} Mohammad A. Khiyami,⁸ Njud S. Alharbi,¹ Nahid H. Hajrah,¹ Meshaal J. Sabir,¹ Mohammed H. Z. Mutwakil,¹ Saleh A. Kabli,¹ Faten A. S. Alsulaimany,¹ Abdullah Y. Obaid,⁹ Boping Zhou,² David K. Smith,⁴ Edward C. Holmes,⁵ Huachen Zhu,^{2,3,4,†} Yi Guan^{1,2,3,4,†}



MERS-CoV Transmissions

Time of Most Recent Common Ancestor

- The tMRCA represents the time of the ancestral MERS-CoV population initially started to jump to humans

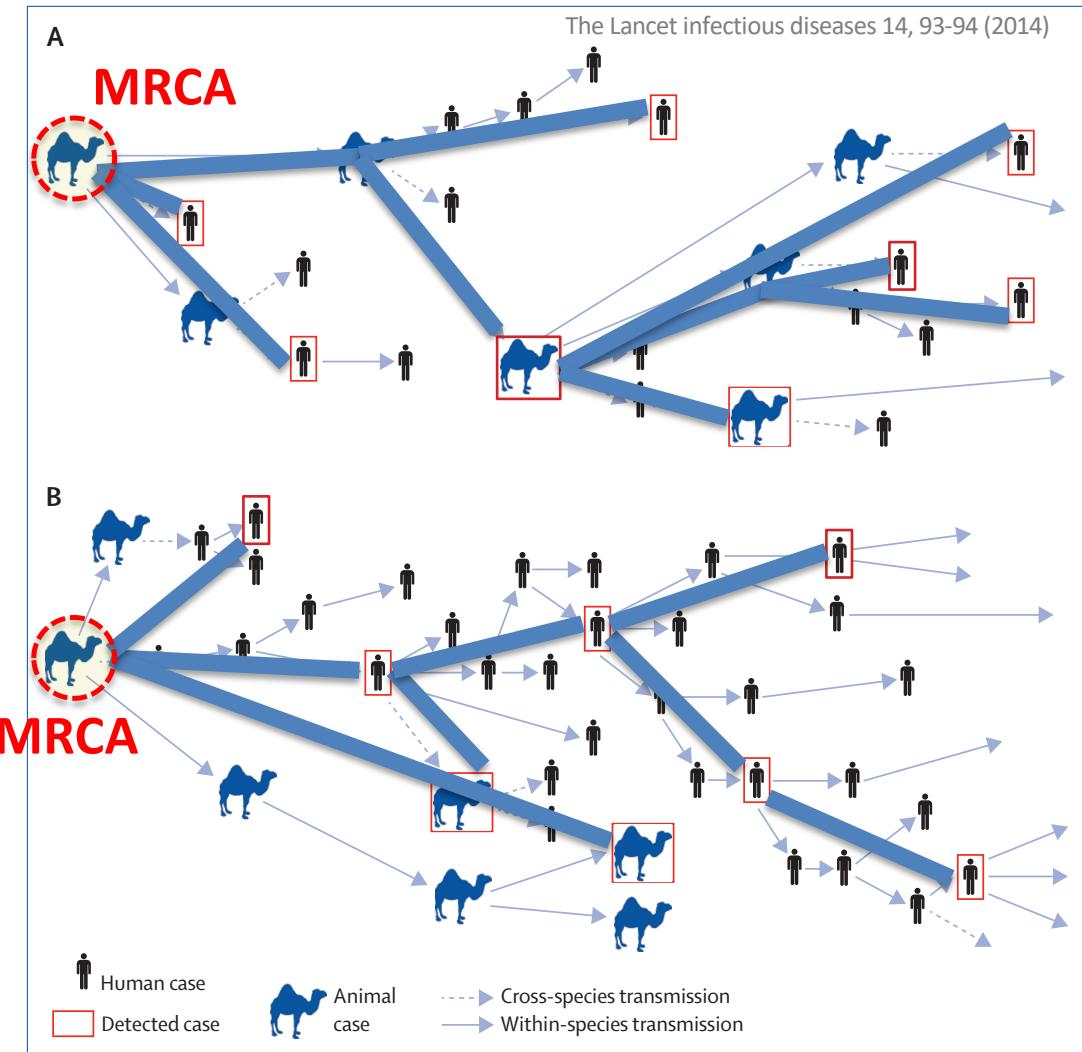


Figure: Possible MERS-CoV transmission scenarios

(A) Self-sustaining transmission in animals, causing spillover infections in people but no self-sustaining human-to-human transmission; control of the animal epidemic might eliminate the virus from human beings in this scenario. (B) Zoonotic exposure triggers a self-sustaining human-to-human epidemic, meaning 11 animal-targeted controls will have only a limited effect on cases in people.

Importance of MERS-CoV tMRCA

- Tracing the factors that have facilitated the zoonotic transmissions to humans
 - Ecological factor
 - Population of animals hosting the virus expanded
 - Environmental factor
 - Climates that favor the transmission to human
 - Genetic factor
 - Some mutations occurred in the virus that become more infectious to human.

Task

- Estimate the
 - Evolutionary history
(phylogeny)
 - Time of the ancestral population
(with molecular clock model)
- of the MERS-CoV from their **genetic sequences & their isolation dates.**

Software and data set required

- Dropbox:
<https://www.dropbox.com/sh/lxbrd0d51d0dzge/AAAfooTJUc5s9MuiH3xVV1BKa>
- Software
 - BEAST - run MCMC
 - Tracer - check MCMC results
 - FigTree - view time-scaled tree

* Note: Java (JRE/JDK) has to be installed first. To check, try open Tracer, if a decent UI window pop up (e.g. the figure on the right), you're safe. Otherwise, if a black/warning window popup or nothing happened, try to install Java from <http://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html>

Or download from **Dropbox**:

 - jre-8u111-macosx-x64.dmg (for Mac)
 - jre-8u111-windows-x64.exe (for Windows 64bit)
 - jre-8u111-windows-i586.exe (for Windows 32bit)
- Data set
 - Input sequence data: “MERSCoV_fullgenome_30.fas”
 - Pre-run result files:
 - Strict.zip (storing MCMC results using strict molecular clock)
 - Relaxed.zip (storing MCMC results using relaxed molecular clock)

Data set

- File name: **MERSCoV_fullgenome_30.fas**
- ‘FASTA’ file format
 - >Seq1_2013.3
ATCGTGTATGTGACATCGATCG
(Virus name) (Isolation date)
 - >Seq2_2013.8
ATCAGCTAAGTCAGCGGCCG
- A set of 30 genome sequences of human and camel MERS-CoVs.
- ~30,000 nucleotides each

Bayesian framework

- A Bayesian framework for **integrated** evolutionary inference was developed (Genetics. 2002 Jul;161(3):1307-20)

- It is implemented in a software called “**BEAST**” (<http://code.google.com/p/beast-mcmc>)

No need to write codes!! 😊



- Phylogeny(g), population dynamics(N_e), evolutionary rate(μ) and substitution model(Q) are jointly estimated from the molecular sequences (D).

Bayesian Inference of N_e, μ, g, Q

Probability (density) of what we don't know given what we do know.

$$P(g, \mu, N_e, Q | D) = \frac{1}{Z} \frac{P(D | g, \mu, Q) f_G(g | N_e) f_m(\mu) f_N(N_e) f_Q(Q)}{\text{Unknown normalizing constant}}$$

Likelihood function

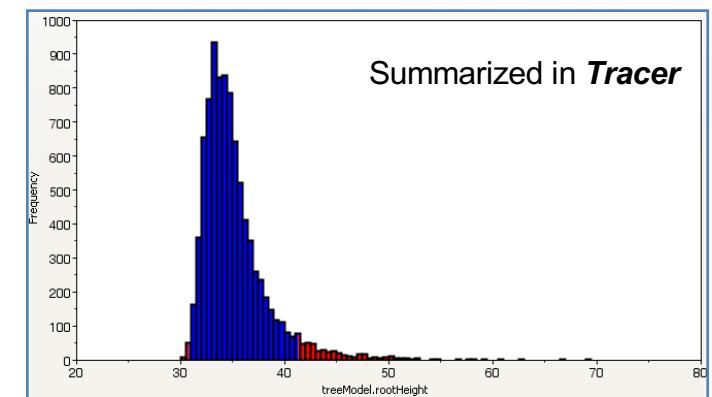
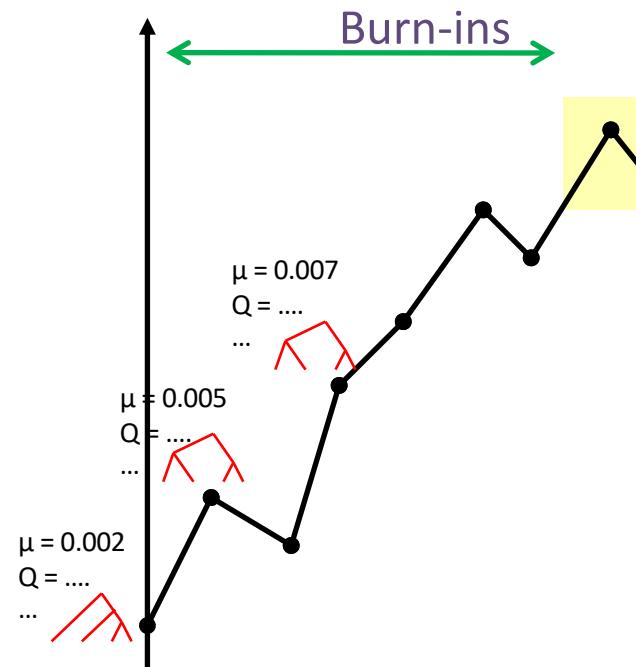
tree prior

other priors

Bayesian Markov Chain Monte Carlo

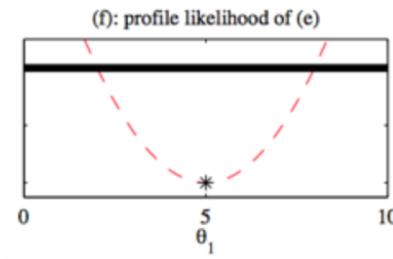
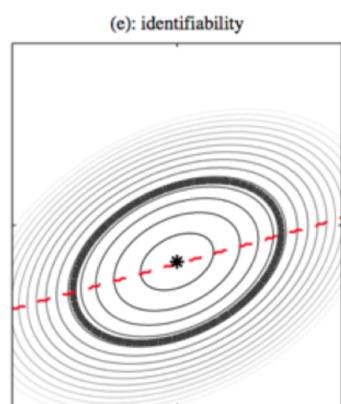
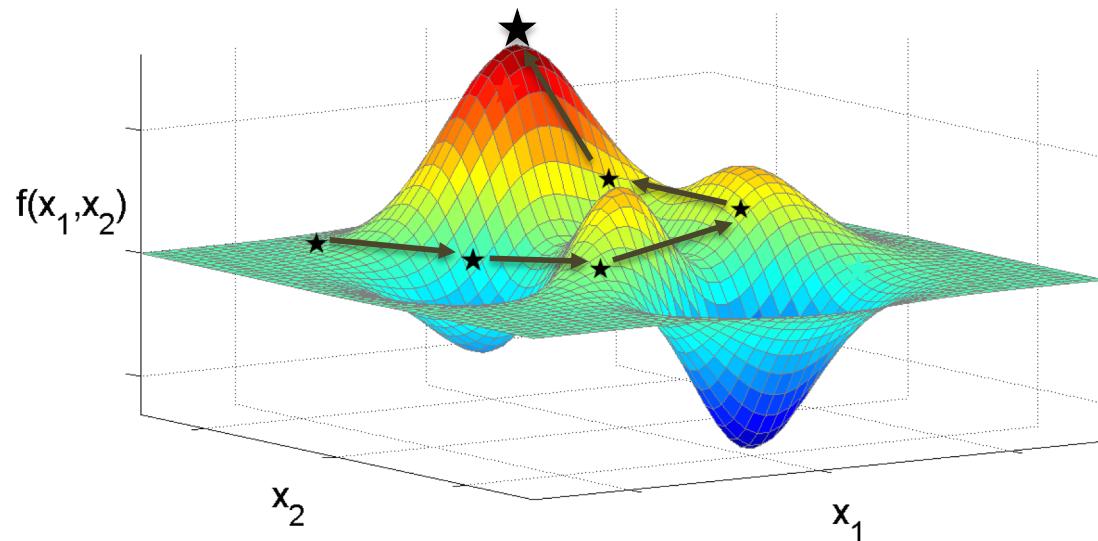
- Markov Chain Monte Carlo method to sample probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution.
 - E.g. Metropolis–Hastings algorithm make proposal and rejection of new samples.

$P(\mathbf{Ne}, \mu, g, Q | D)$

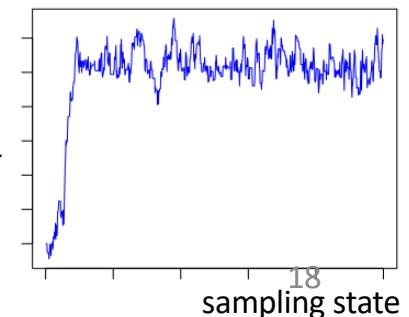
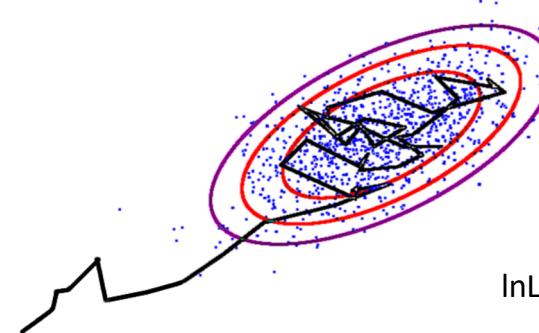
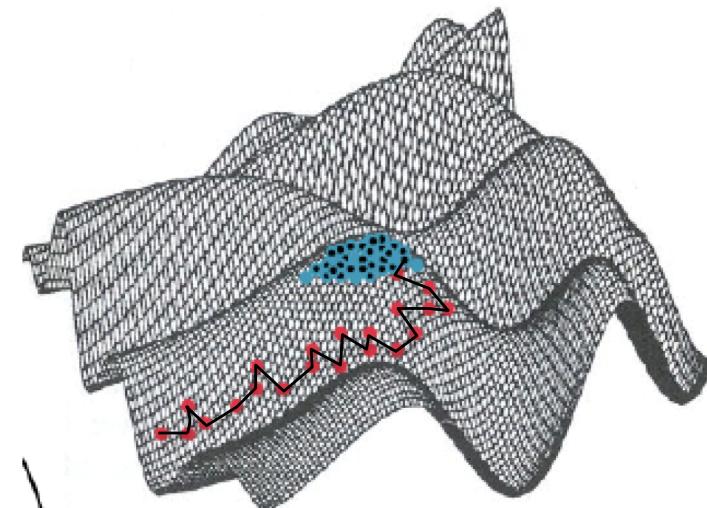


Maximum likelihood (ML) search and Bayesian Markov Chain Monte Carlo (BMCMC) sampling

- ML

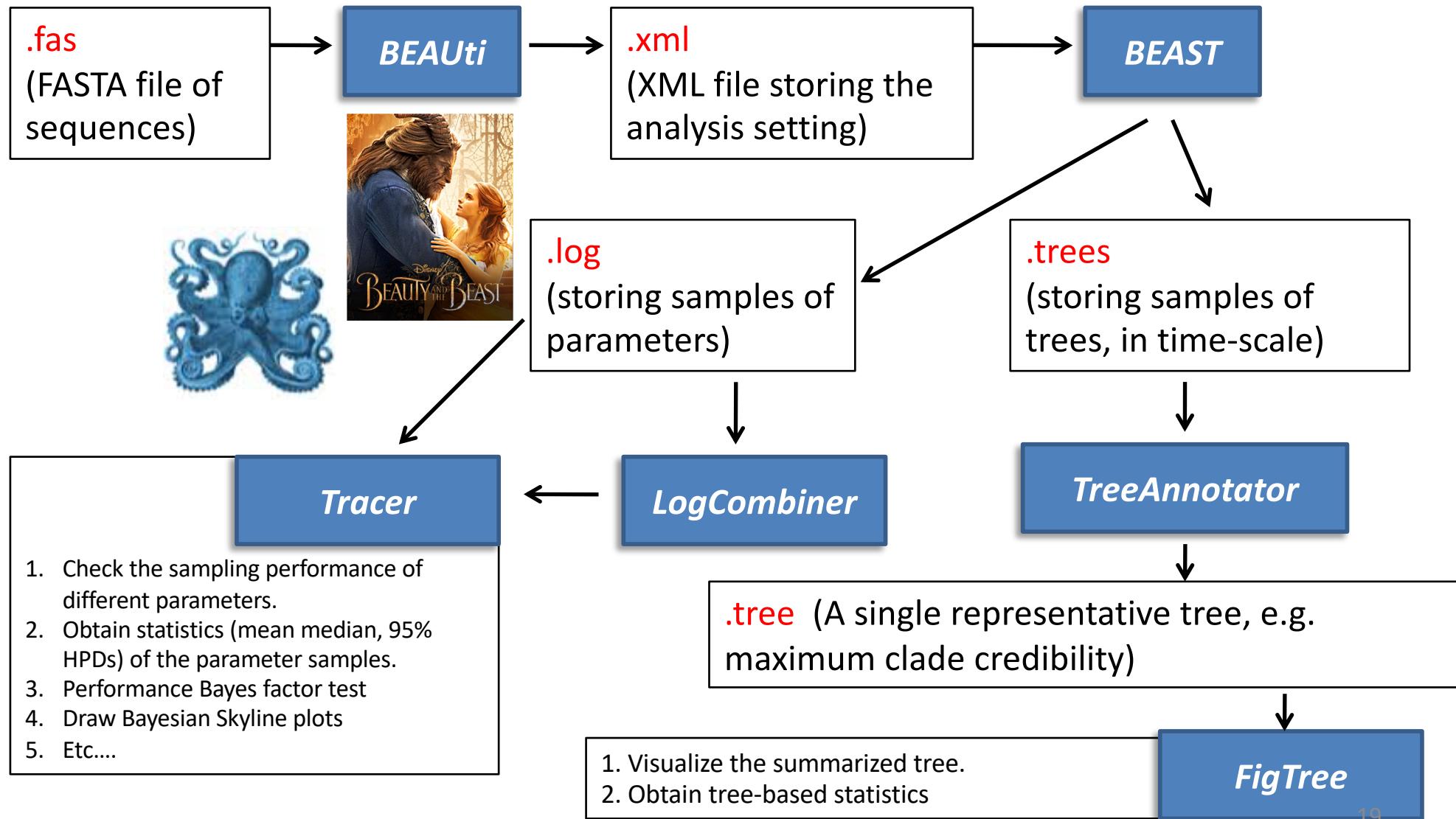


- BMCMC



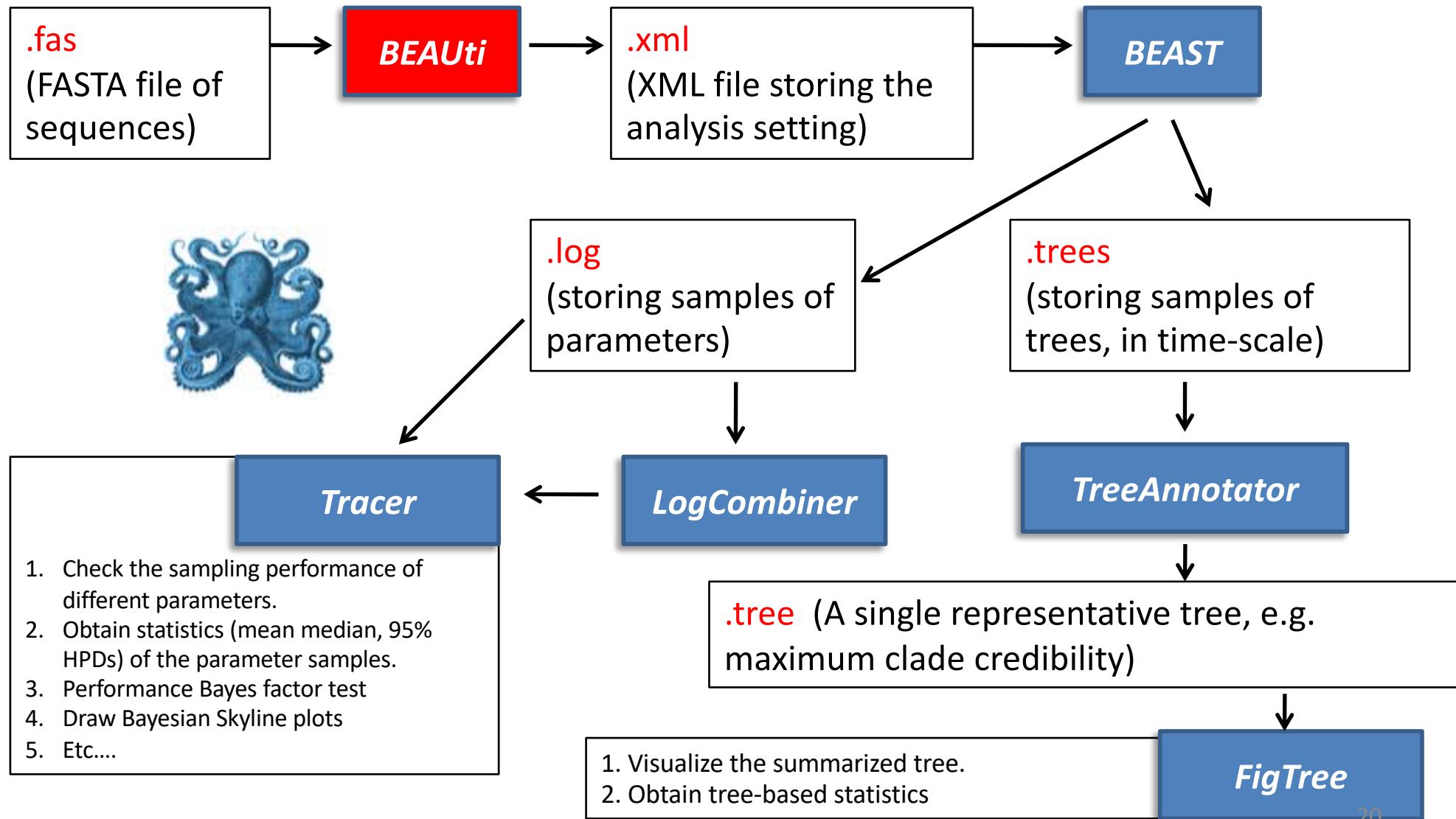
Workflow of the BEAST Analysis

Version 1.8.0



Workflow of the BEAST Analysis

Version 1.8.0



BEAUi

- Create an **XML run file** containing the analysis settings
 - Load the **sequence** file.
 - Set '**dates**' of the sequences.
 - Set '**Substitution Model**' – HKY, GTR, etc.
 - Set '**Clock Model**' – Strict clock, UCLD relaxed clock, UCED relaxed clock, etc.
 - Set '**Population Model**' – Bayesian Skyline or Skyride, etc.
 - Set '**Priors**' distribution of the parameters.
 - Set the **length** of the MCMC.
 - Make other settings...

BEAUti

BEAUti

Partitions Taxa Tips Traits Sites Clocks Trees States Priors Operators MCMC

Use species tree ancestral reconstruction (*BEAST) Heled & Drummond 2010

Unlink Subst. Models Link Subst. Models Unlink Clock Models Link Clock Models Unlink Trees Link Trees

Partition Name	File Name	Taxa	Sites	Data Type	Site Model	Clock Model	Partition Tree
MERSCoV_fullgen...	MERSCoV_fullgen...	30	29333	nucleotide	MERSCoV_fullge...	MERSCoV_fullge...	MERSCoV_fullge...

Sequence data loaded

+ - View Partition ... Create partition from trait ...

Data: 30 taxa, 1 partition; Fix clock rate to 1.0 in nucleotide_group;

Generate BEAST File...

BEAUti

Guess the dates
from your
sequence
names/taxon
labels

BEAUti

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

Guess date values for all taxa

The date is given by a numerical field in the taxon label that is:

- Defined just by its order
- Defined by a prefix and its order
- Defined by regular expression (REGEX)
- Parse as a number

Order: last

Prefix:

Add the following value to each: 1900.0

...unless less than: 15.0

...in which case add: 2000.0

Date format: yyyy-MM-dd

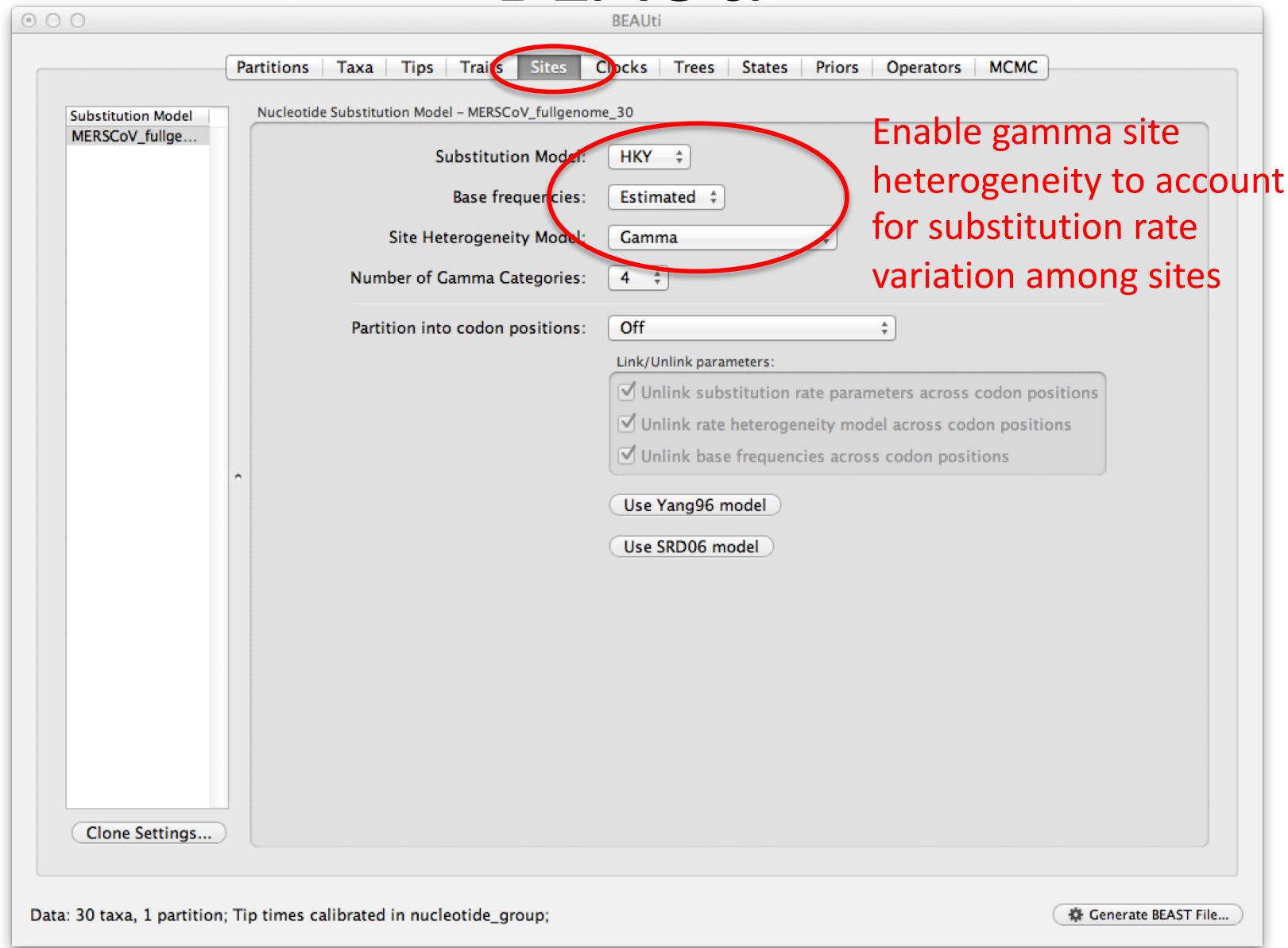
Tip date sampling: Off

Data: 30 taxa, 1 partition; Fix clock rate to 1.0 in nucleotide_group;

Generate BEAST File...

Cancel OK

BEAUti



BEAUti

The screenshot shows the BEAUti software interface. The title bar reads "BEAUti". Below it is a navigation bar with tabs: Partitions, Taxa, Tips, Traits, Sites, **Clocks**, Trees, States, Priors, Operators, and MCMC. The "Clocks" tab is highlighted with a red oval. Below the tabs is a section titled "Clock Model:" containing a table. The table has columns: Name, Model, Estimate, Rate, and Group. A single row is visible: "MERSCoV_fullgenome_30" under "Name", "Strict clock" under "Model", "1.0" under "Rate", and "nucleotide_group" under "Group". The "Estimate" column contains a dropdown menu with a checked checkbox. To the right of the table, the text "Choose your clock model here:" is displayed in red. Below this, four clock models are listed in red: **Strict clock** – same substitution rate in the entire tree; **Lognormal relaxed clock** – rates varied in lognormal distribution among branches; **Exponential relaxed clock** – rates varied in exponential distribution among branches; and **Random local clock** – rates varied sharply in certain lineages. At the bottom of the main window, there is a "Clock Model Group:" section with a table and a "+" button.

Data: 30 taxa, 1 partition; Tip times calibrated in nucleotide_group;

Generate BEAST File...

25

BEAUTi

The screenshot shows the BEAUTi software interface with the 'Trees' tab selected (highlighted with a red circle). A red box highlights the 'Tree Prior' dropdown menu, which is set to 'Coalescent: GMRF Bayesian Skyride'. A red bracket on the right side of the slide points to this same dropdown menu.

Bayesian Skyline or Skyride model – relatively flexible priori on the population dynamics

Trees
MERSCoV_fullgenome_30

Tree prior shared by all tree models

Tree Prior: Coalescent: GMRF Bayesian Skyride

Smoothing: Time-aware

For the Skyride, tree model/tree prior combination not implemented by BEAST. The Skyride is only available for a single model partition in this release. Please try the Skygrid or link all tree models.

Citation:
Minin VN, Bloomquist EW, Suchard MA (2008) Mol Biol Evol 25, 1459–1471
[Skyride Coalescent].
Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Genetics 161, 1307–1320 [Serially Sampled Data].

Tree Model – MERSCoV_fullgenome_30

Random starting tree
 UPGMA starting tree
 User-specified starting tree

Select user-specified tree: no tree loaded

Export format for tree: Newick

Import user-specified starting trees from **NEXUS** format data files using the 'Import Data' menu option.
Trees must be rooted and strictly bifurcating (binary).

(You can choose others, e.g. exponential growth, if you know/assume how it grew)

Data: 30 taxa, 1 partition; Tip times calibrated in nucleotide_group;

Generate BEAST File...

BEAUti

Priors for model parameters and statistics:

Parameter	Prior	Bound	Description
kappa	* LogNormal [1, 1.25], initial=2	[0, ∞)	HKY transition-transversion parameter
freqencies	* Uniform [0, 1], initial=0.25	[0, 1]	base frequencies
alpha	* Exponential [0.5], initial=0.5	[0, ∞)	gamma shape parameter
clock.rate	? Not yet specified, initial=1	[0, ∞)	substitution rate
treeModel.rootHeight	? Using Tree Prior in [2.07, ∞)	[2.07, ∞)	root height of the tree
skyride.precision	* Gamma [0.001, 1000], initial=1	[0, ∞)	GMRF Bayesian skyride precision

Prior for Parameter clock.rate

Select prior distribution for clock.rate

Prior Distribution: Uniform

Initial value: 0.001

Upper: 1.0

Lower: 0.0

Cancel OK

Link parameters into a phylogenetic hierarchical model

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

Data: 30 taxa, 1 partition; Tip times calibrated in nucleotide_group;

Generate BEAST File...

Set the clock.rate prior to a *uniform distribution*

Change the initial value rate value to 0.001 (according to mBio 5 (2014))

RNA virus has clock rate around 10^{-3} to 10^{-4} subs/site/year.

BEAUti

BEAUti

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: MERSCoV_fullgenome_30_strict_hkyg_bsr

Give a more informative file name

Add .txt suffix

Log file name: MERSCoV_fullgenome_30_strict_hkyg_bsr.log

Trees file name: MERSCoV_fullgenome_30_strict_hkyg_bsr.trees

Create tree log file with branch length in substitutions:

Substitutions trees file name:

Create operator analysis file:

Operator analysis file name: MERSCoV_fullgenome_30_strict_hkyg_bsr.ops

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) which performs an additional analysis after the standard MCMC chain has finished.

Perform marginal likelihood estimation (MLE) using path sampling/stepping-stone sampling

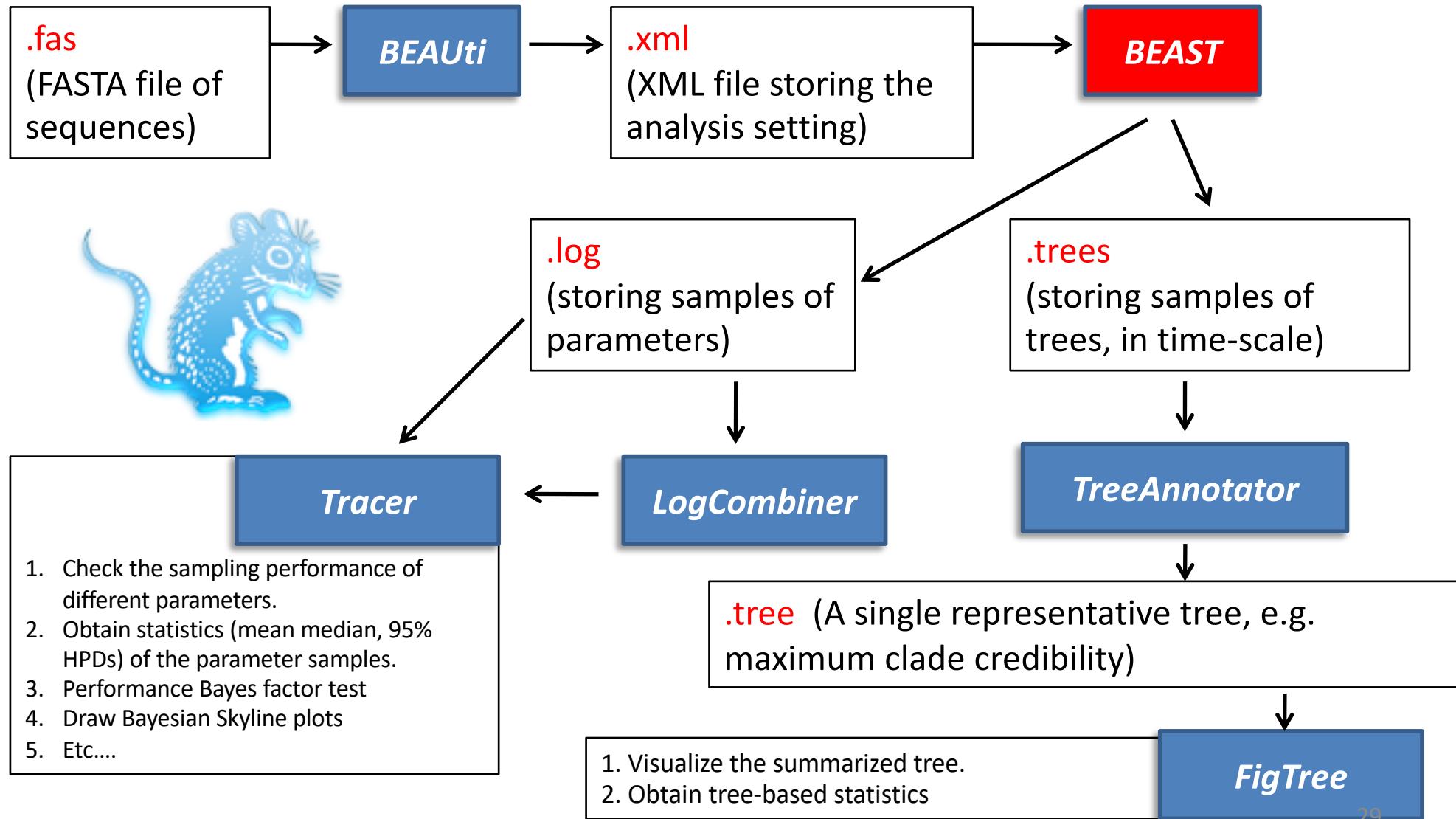
Settings

Data: 30 taxa, 1 partition; Tip times calibrated in nucleotide_group;

Generate BEAST File...

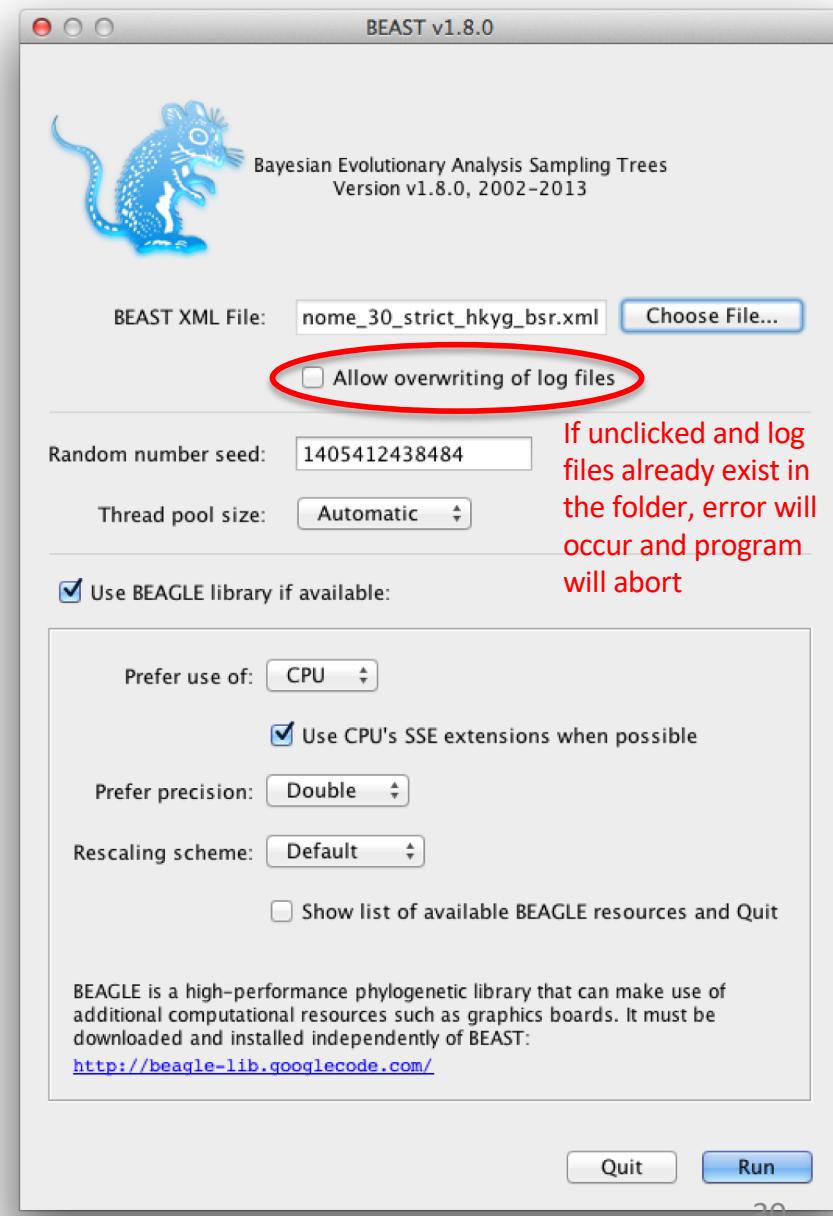
Workflow of the BEAST Analysis

Version 1.8.0



BEAST

- Run with the **XML run file** created by BEAUTi.
 - Set seed for your random generator
 - Set usage of CPU or GPU for computation
 - etc...

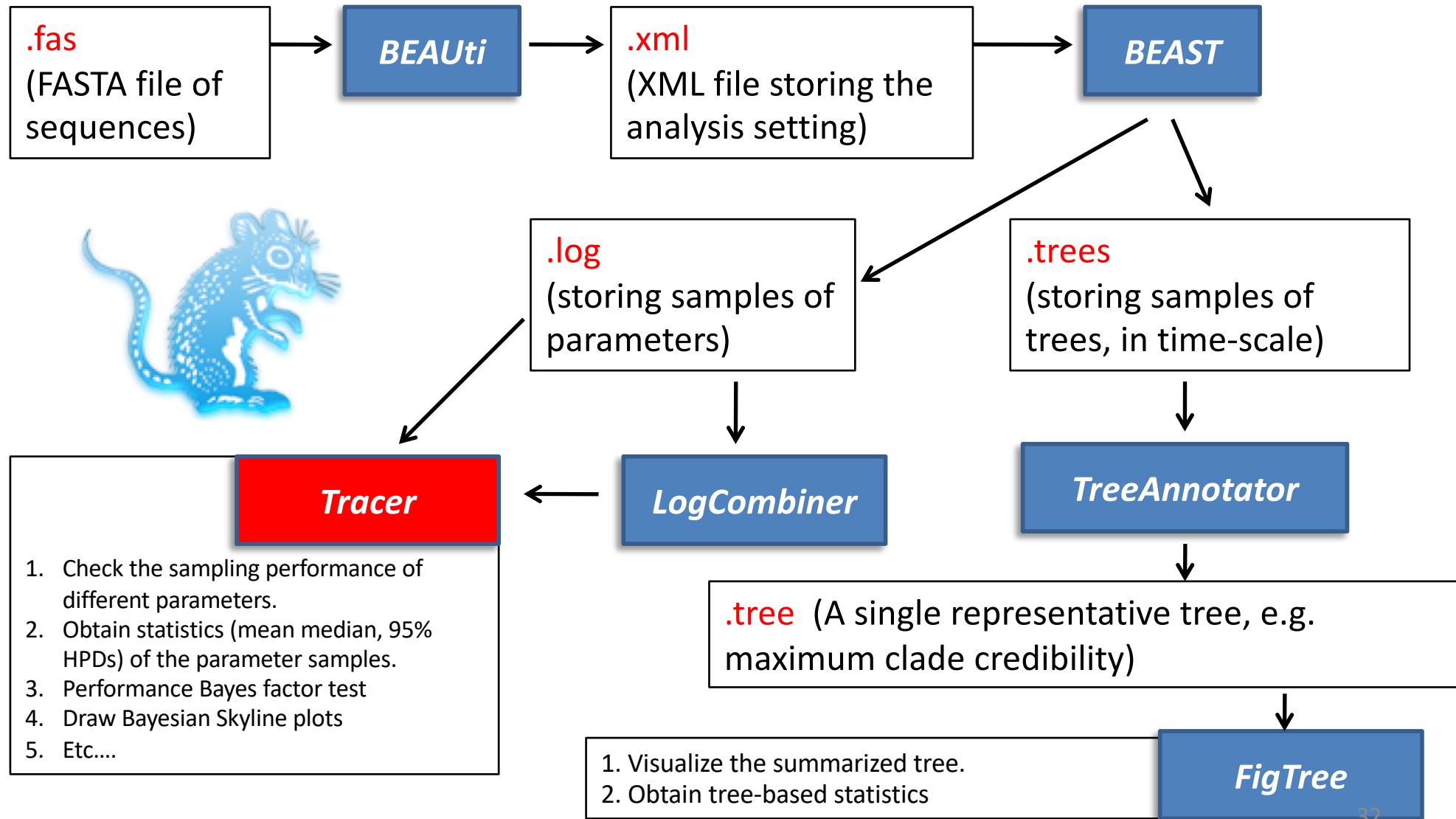


BEAST

Running Wait for 15-20 mins

Workflow of the BEAST Analysis

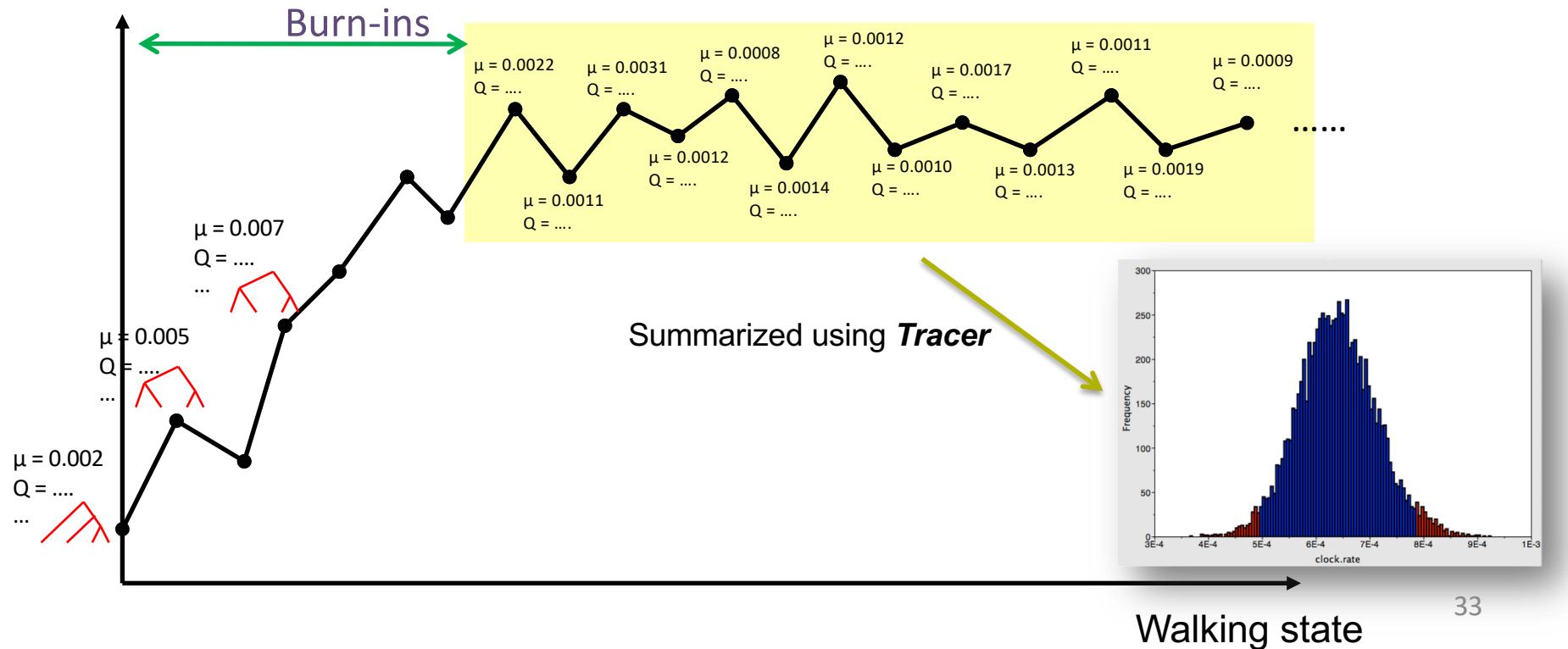
Version 1.8.0



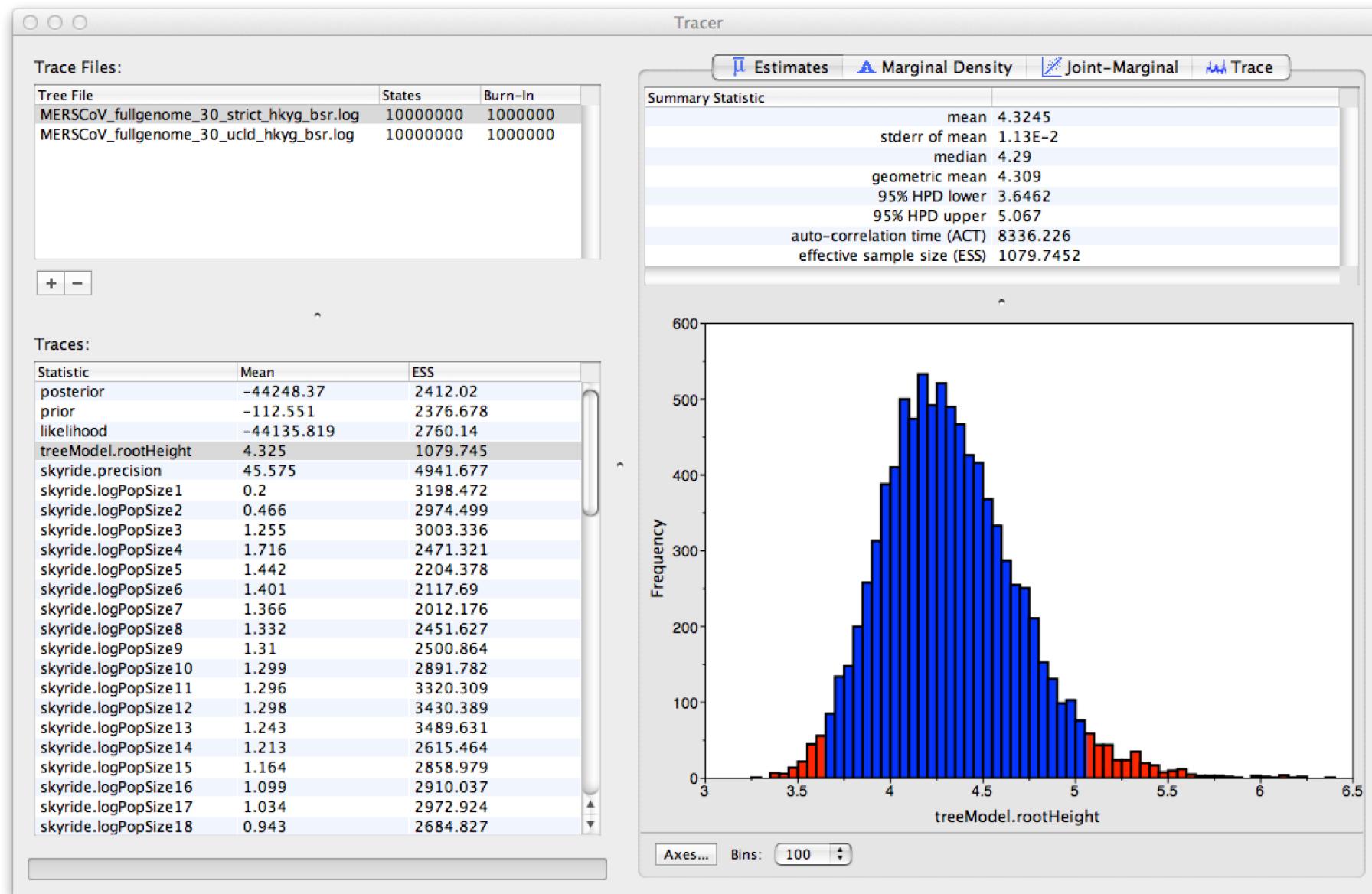
Tracer

- Summary statistics and distributions of the posterior samples of parameters stored in “.log” file.

$P(\mathbf{Ne}, \mu, g, Q | D)$

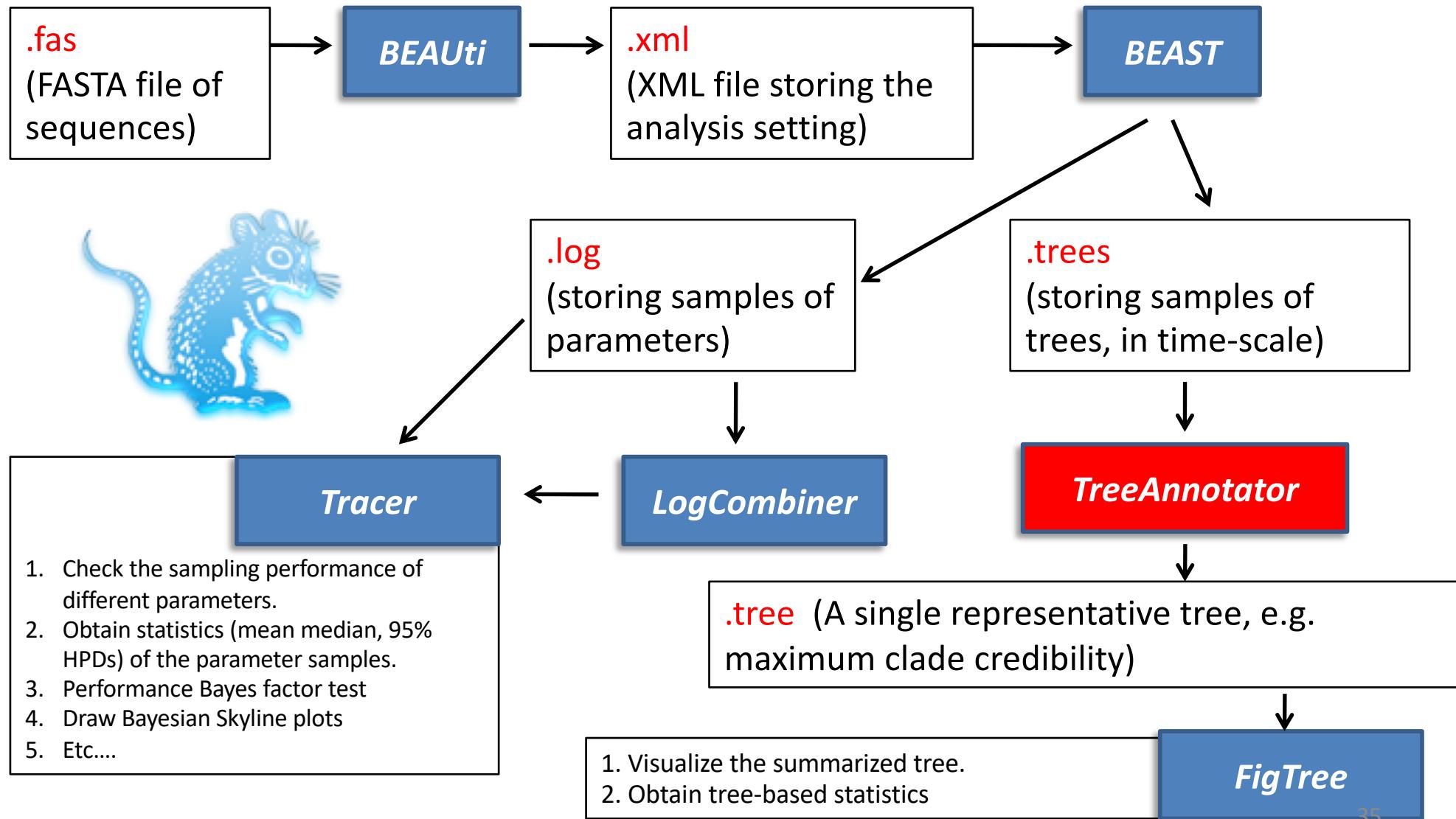


Tracer (check the parameters)



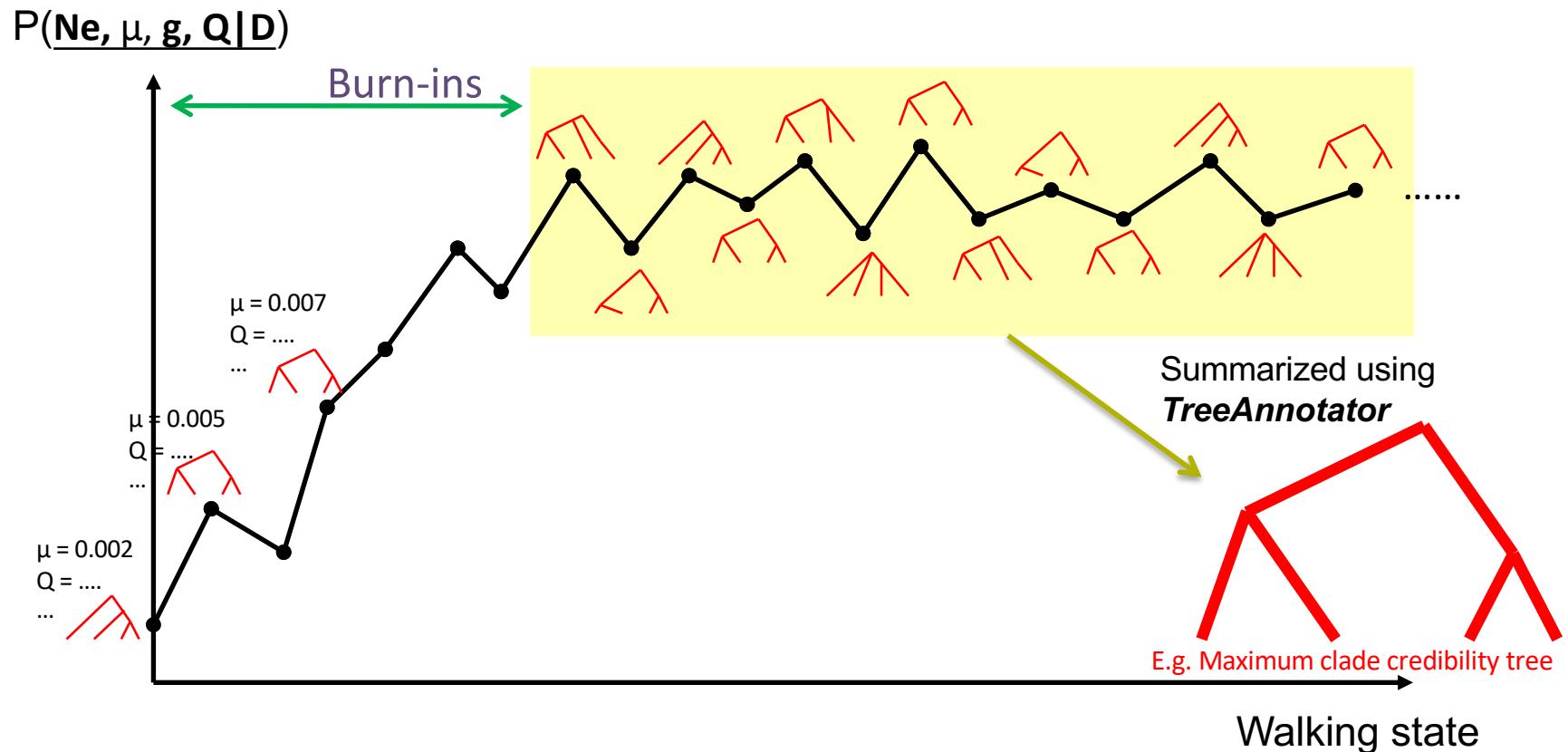
Workflow of the BEAST Analysis

Version 1.8.0

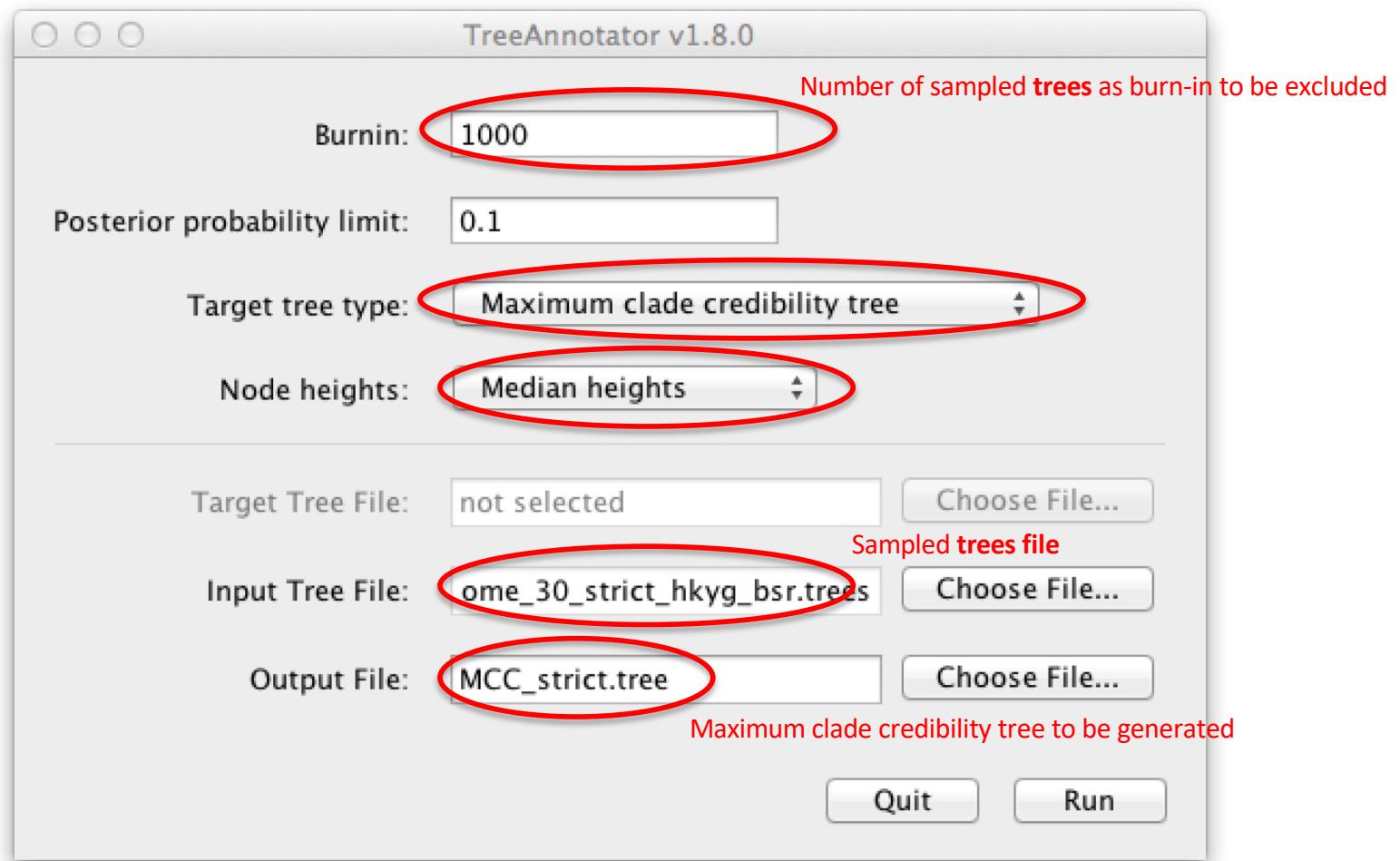


TreeAnnotator

- Summarize a majority-rule tree from your sampled posterior trees stored in “.trees” file.
 - E.g. Maximum clade credibility tree – the tree that combines the most common clades

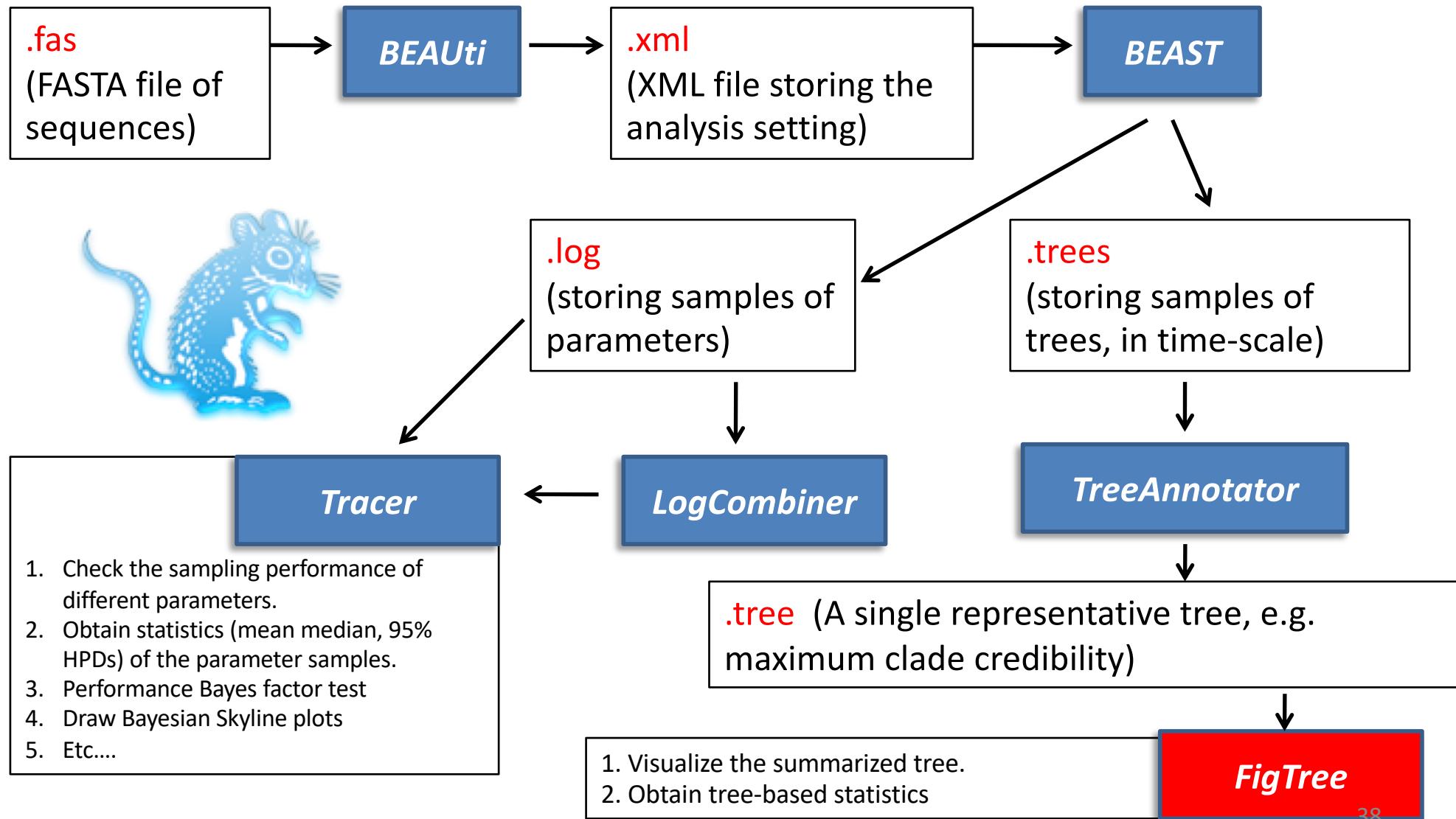


TreeAnnotator

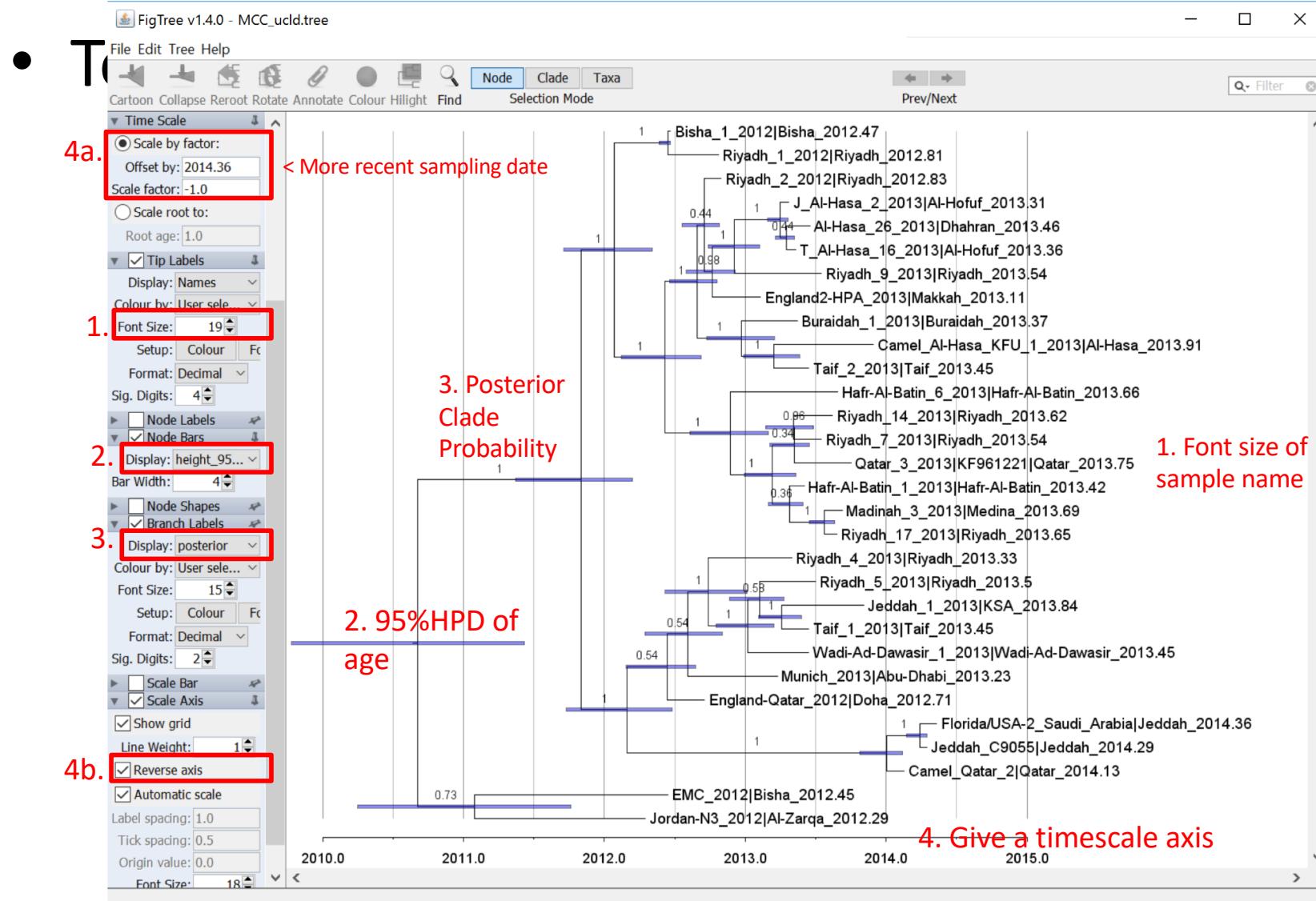


Workflow of the BEAST Analysis

Version 1.8.0



FigTree



tMRCA of MERS-CoV

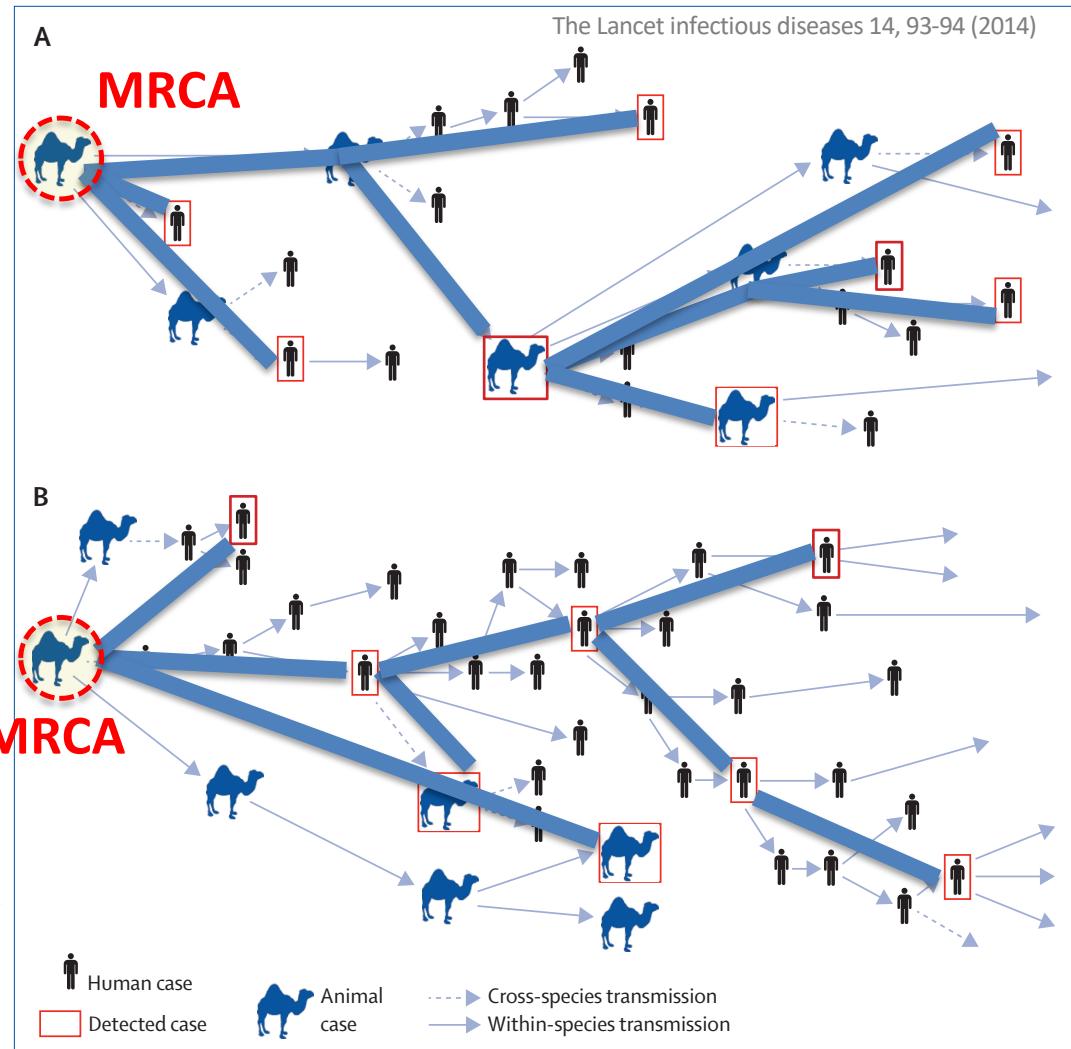
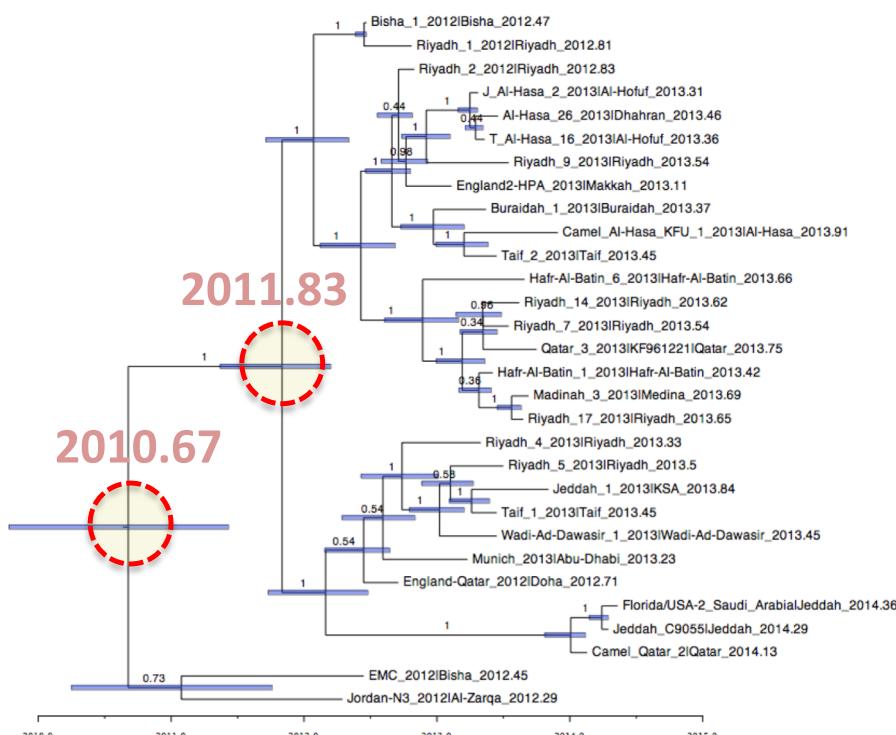


Figure: Possible MERS-CoV transmission scenarios

(A) Self-sustaining transmission in animals, causing spillover infections in people but no self-sustaining human-to-human transmission; control of the animal epidemic might eliminate the virus from human beings in this scenario. (B) Zoonotic exposure triggers a self-sustaining human-to-human epidemic, meaning animal-targeted controls will have only a limited effect on cases in people.⁴⁰

Summary

- Perform **Bayesian** BMCMC phylogenetic inference
- Aware of **estimate uncertainty**
- Compare model fit by **Bayes Factor test**
- Obtain and view **time-scaled tree** using **molecular clock**
- Interpretation of the **tMRCA**

References

- *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing.* 2nd Ed. Philippe Lemey, Marco Salemi, Anne-Mieke Vandamme.
- *Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections.* Lam TTY, Hon CC, Tang JW. 2010. Critical Reviews in Clinical Laboratory Sciences. 47(1):5-49.
- *BEAST.* Andrew Rambaut. beast.bio.ed.ac.uk/
- *Relaxed Phylogenetics and Dating with Confidence.* Alexei J Drummond , Simon Y. W Ho, Matthew J Phillips, Andrew Rambaut. 2006. PLoS Biol 4(5): e88.
- *Skyride through a Rough Skyline: Bayesian Coalescent-Based Inference of Population Dynamics.* Vladimir N. Minin, Erik W. Bloomquist, Marc A. Suchard; Smooth Mol Biol Evol 2008; 25 (7): 1459-1471.