



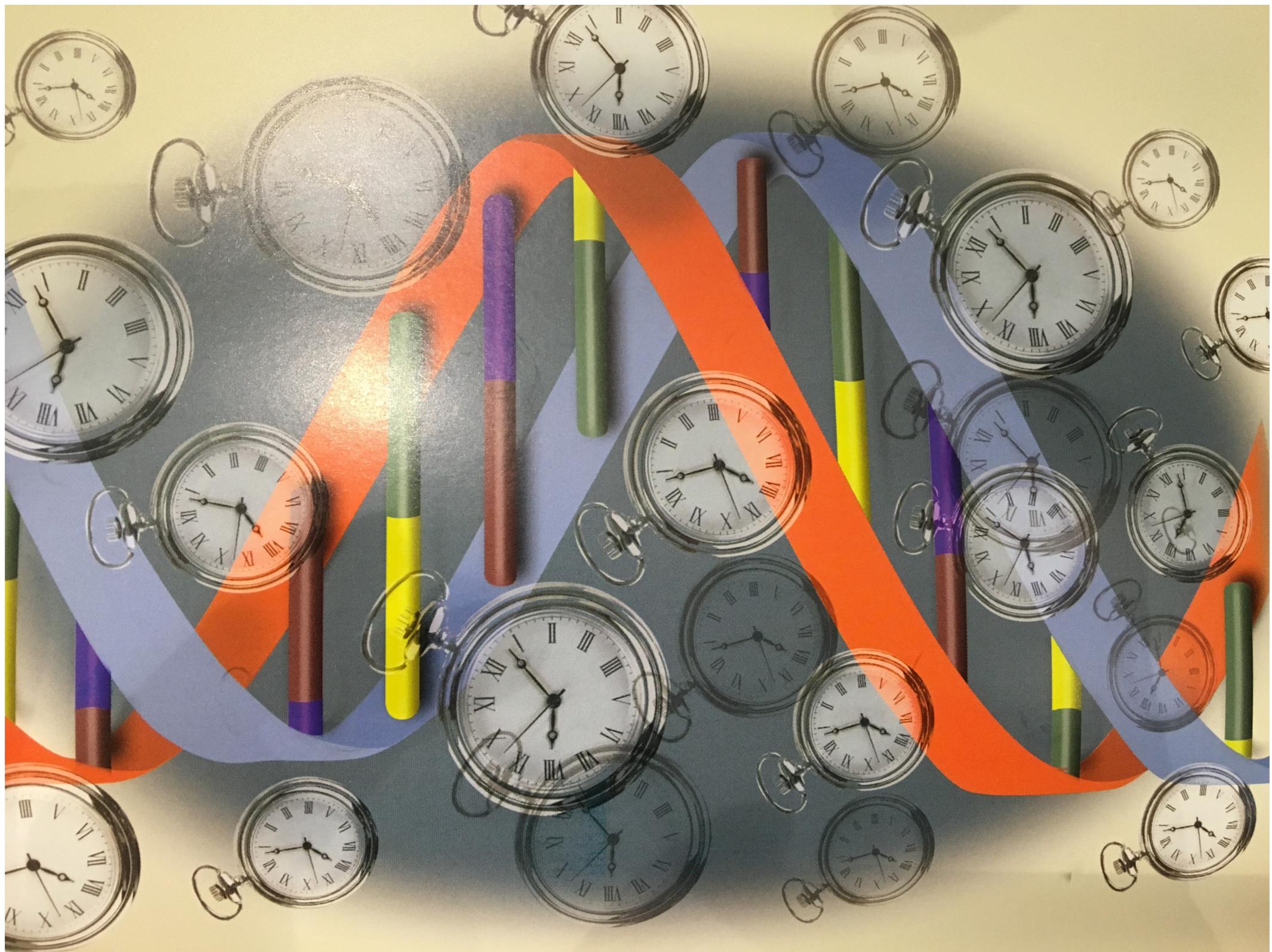
Phylodynamic inference from sequence data for epidemiology of infectious disease

“Shinny” Pimplapas Leekitcharoenphon
Research Group for Genomic Epidemiology
DTU Food

$$f(x+\Delta x) = \sum_{i=0}^{\infty} \frac{(\Delta x)^i}{i!} f^{(i)}(x)$$
$$\int_a^b \Theta + \Omega \int \delta e^{i\pi} =$$
$$\sqrt{17} \sum \chi^2 > \Sigma!$$

$\infty = \{2.7182818284$

BEAST



BEAST

- BEAST - Bayesian Evolutionary Analysis Sampling Trees
- BEAST is a program to reconstruct and date phylogenetic tree
- When building a tree, every reconstruction methods is outputting branch length (bl) that are a function of the rate of mutation (μ) and the time of evolution (t)

$$bl = \mu * t$$

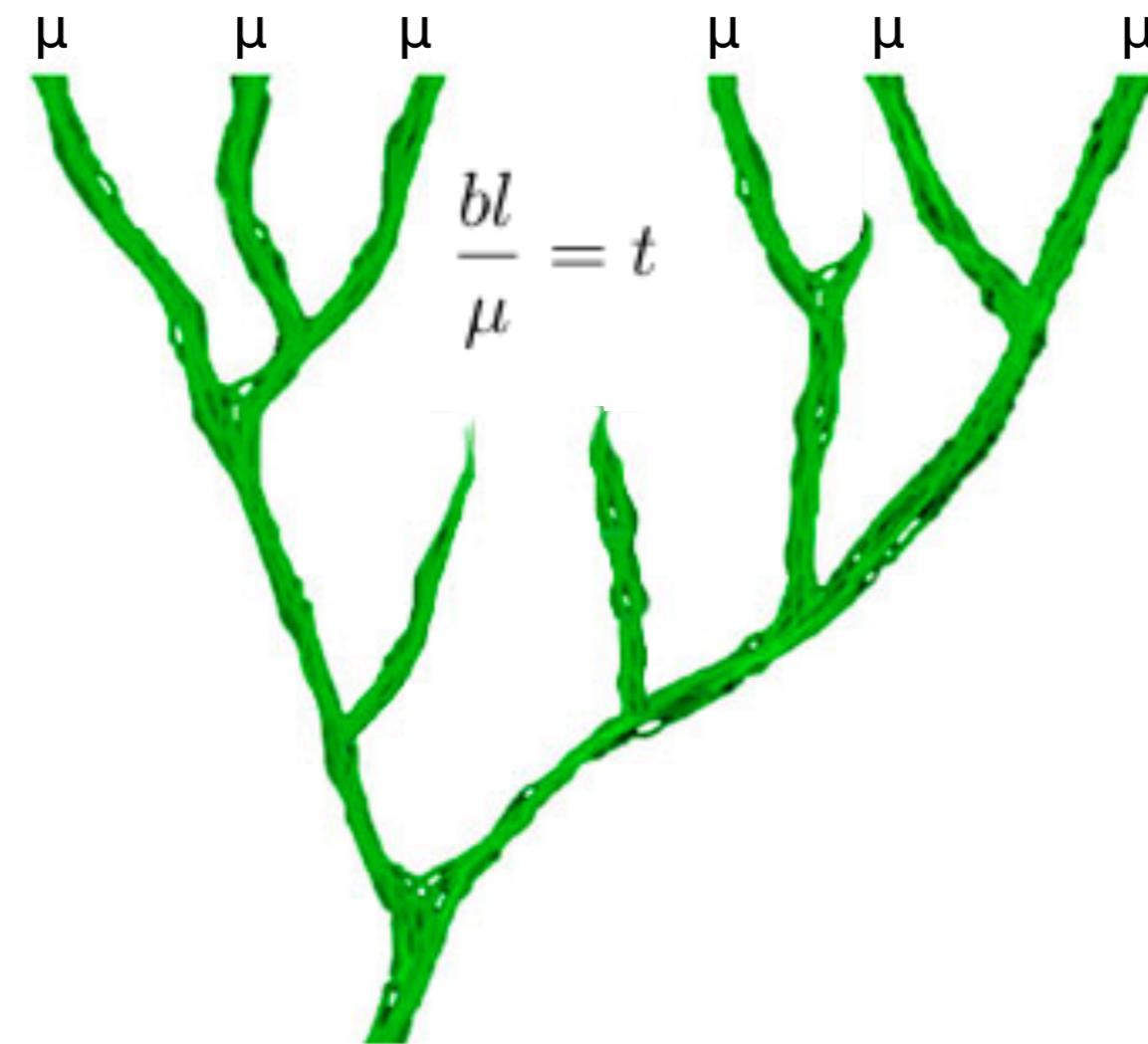
- If we want to estimate the divergence time t of each node, we need to separate the two parameters from each branch length.
- To do so, we need to model how μ might vary between every branches in the tree. By knowing μ , it is easy to get t

$$\frac{bl}{\mu} = t$$

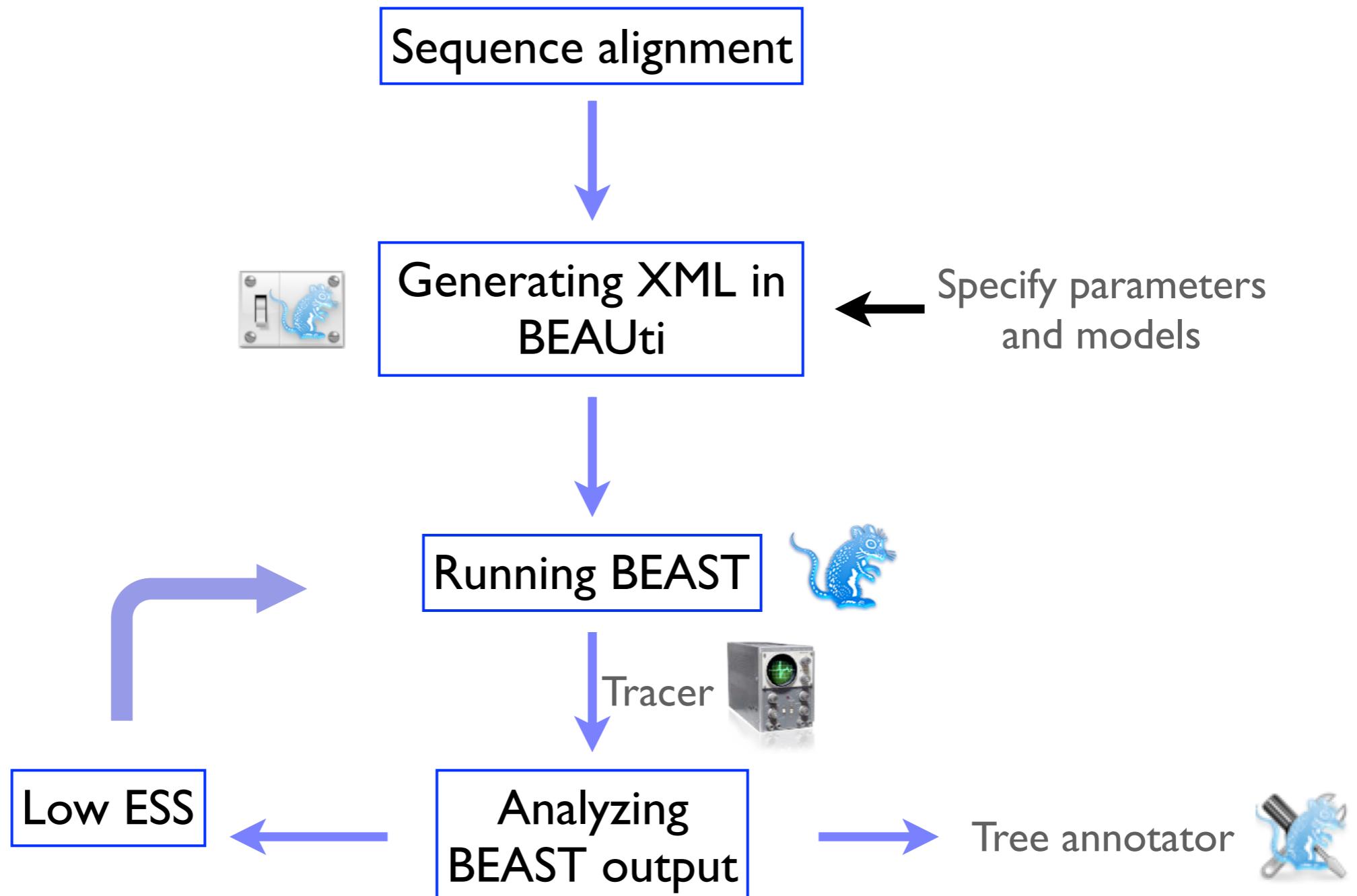
- How do we model μ ? There are a few possibilities:

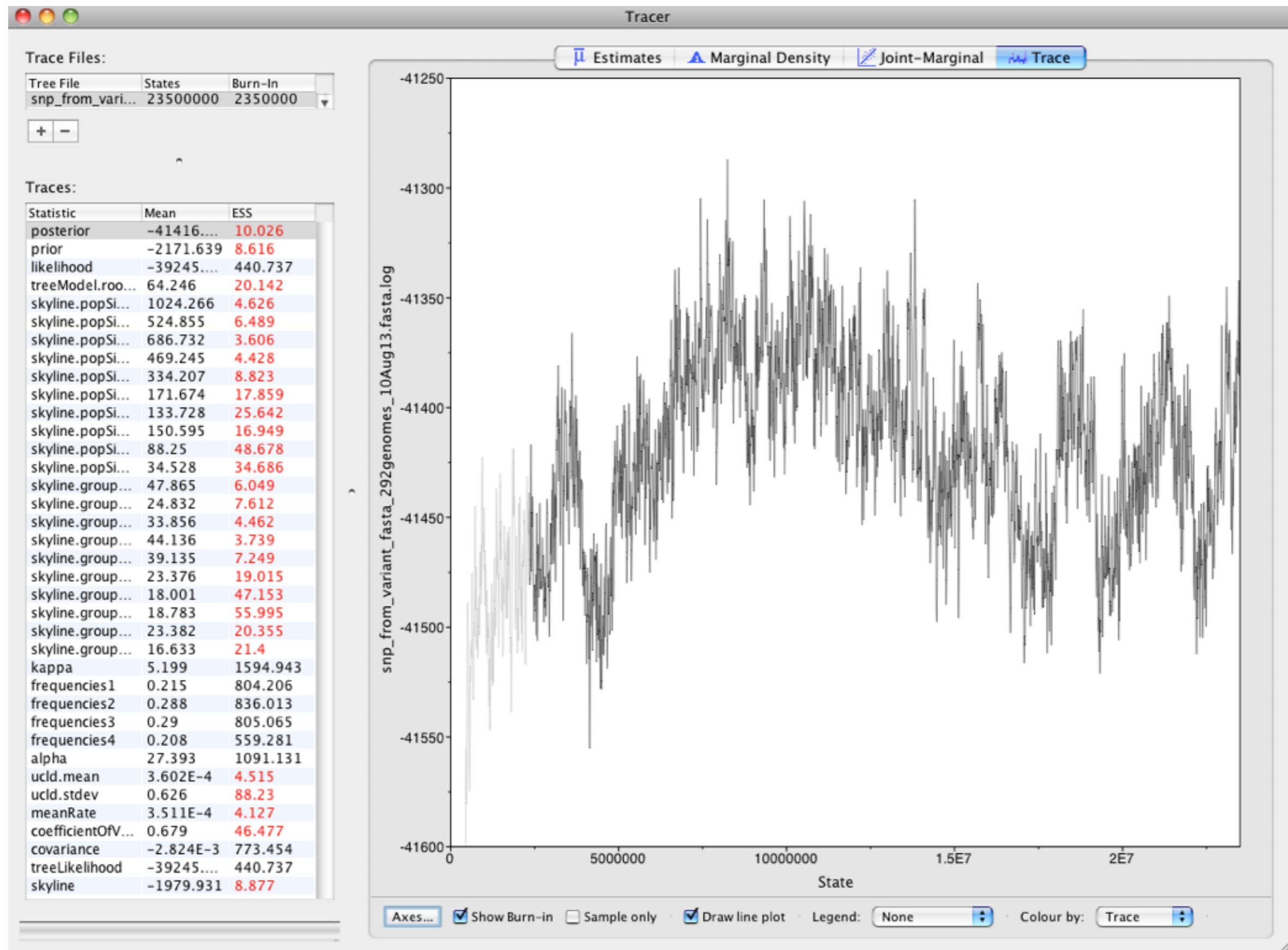
BEAST

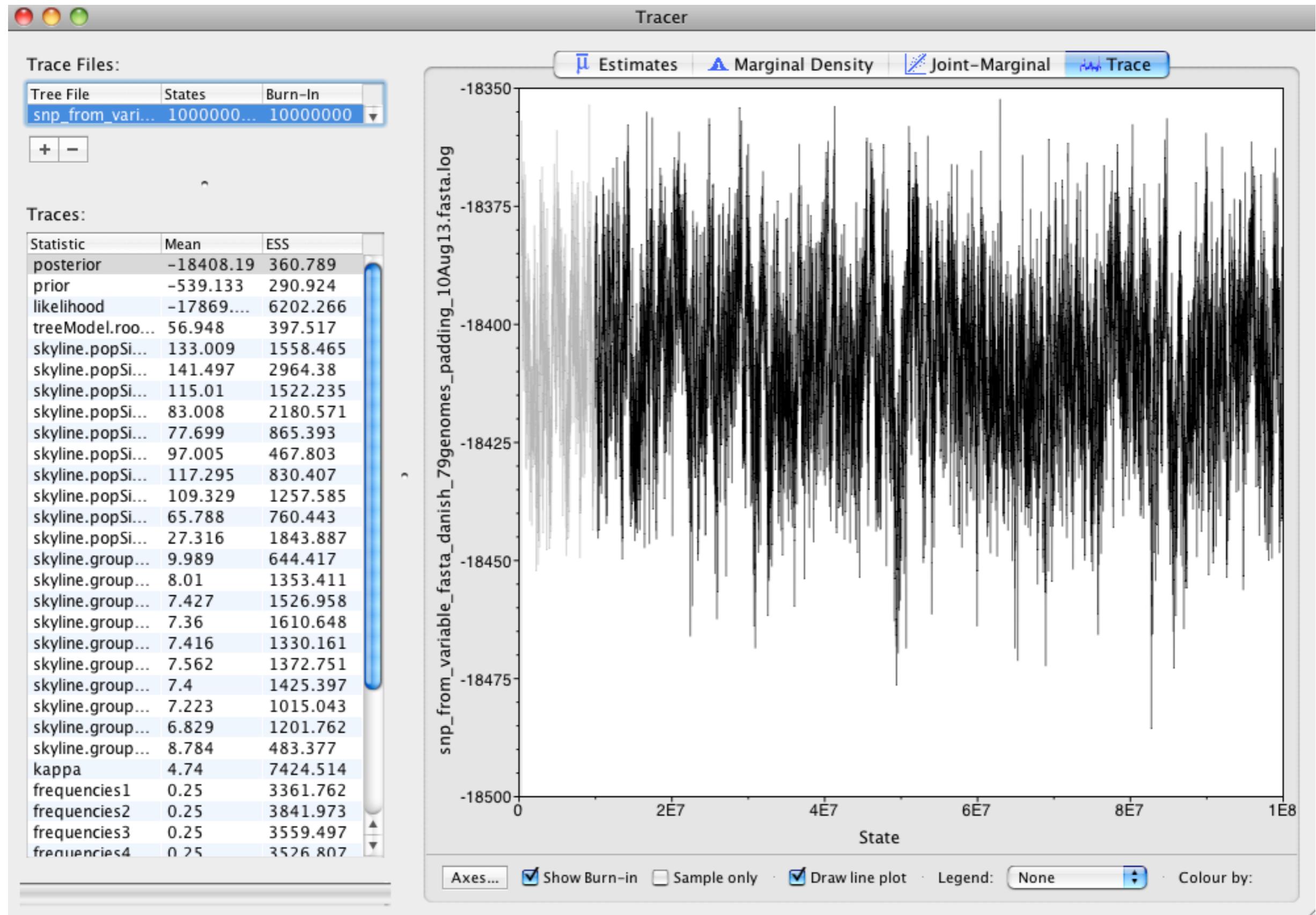
- BEAST - Bayesian Evolutionary Analysis Sampling Trees
- BEAST is a program to reconstruct and date phylogenetic tree



BEAST workflow







- beast01 : Relaxed uncorrelated lognormal clock & Coalescent Exponential Growth
 beast02 : Relaxed uncorrelated lognormal clock & Coalescent Bayesian Skyline
 beast03 : Relaxed uncorrelated lognormal clock & Coalescent Constant Size
 beast04 : Relaxed uncorrelated lognormal clock & Coalescent Logistic Growth
 beast05 : Relaxed uncorrelated lognormal clock & Coalescent Expansion Growth
 beast06 : Relaxed uncorrelated lognormal clock & Coalescent GMRF Bayesian Skyride
 beast07 : Strict clock & Coalescent Exponential Growth
 beast08 : Strict clock & Coalescent Bayesian Skyline

Bayes Factors

Show: Bayes Factors

trace: likelihood, smoothed estimate (S.E. estimated using 1000 bootstrap replicates)

	Trace	In P(model...)	S.E.	snp_from_...	snp_from_...	snp_from_...	snp_from_...	snp_from_...	snp_from_variable...	snp_from_varia...	
beast01	snp...	-17889.0...	+/- 0.331	-	0.078	1.324	1.706	1.788	0.179	568794645.008	3258276.575
beast02	snp...	-17886.5...	+/- 0.374	12.79	-	16.932	21.817	22.868	2.294	7275138938.868	41674820.591
beast03	snp...	-17889.3...	+/- 0.347	0.755	0.059	-	1.289	1.351	0.135	429665139.359	2461288.746
beast04	snp...	-17889.5...	+/- 0.349	0.586	0.046	0.776	-	1.048	0.105	333458607.58	1910180.378
beast05	snp...	-17889.6...	+/- 0.306	0.559	0.044	0.74	0.954	-	0.1	318137738.994	1822416.494
beast06	snp...	-17887.3...	+/- 0.334	5.576	0.436	7.382	9.512	9.97	-	3171732981.954	18168918.023
beast07	snp...	-17909.2...	+/- 0.347	0	0	0	0	0	-	-	0.006
beast08	snp...	-17904.0...	+/- 0.273	0	0	0	0	0	174.569	-	

Marginal likelihood estimated using the method Newton & Raftery (ref)with the modifications proposed by Suchard et al (2001, MBE 18: 1001–1013)

Global genomic epidemiology of *Salmonella* Typhimurium DT104

Leekitcharoenphon P, et al. Global Genomic Epidemiology of *Salmonella enterica* Serovar Typhimurium DT104. Appl Environ Microbiol. 2016 Apr 4;82(8):2516-26. doi: 10.1128/AEM.03821-15. PMID: 26944846

$$f(x+\Delta x) = \sum_{i=0}^{\infty} \frac{(\Delta x)^i}{i!} f^{(i)}(x)$$
$$\int_a^b \Theta + \Omega \int \delta e^{i\pi} =$$
$$\frac{\sqrt{17}}{\infty} = \{2.7182818284\}$$
$$\sum \gg,$$

S. Typhimurium DT104

- During the last three decades, *S. Typhimurium* phage type DT104 emerged as the most important phage type and one of the best-studied because of its rapid global dissemination [Lan R, *et al.* Infect Genet Evol. 2009] [Helms M, *et al.* Emerg Infect Dis. 2005]
- DT104 has a multiple antimicrobial resistance pattern to ampicillin, chloramphenicol, streptomycin, sulphonamide and tetracycline (ACSSuT) [Mulvey MR, *et al.* Microbes Infect. 2006]
- Previous epidemics with MDR phage types of *S. Typhimurium*, such as DTs 29, 204, 193 and 204c, were mostly restricted to cattle [Threlfall EJ. J Antimicrob Chemother. 2000]
- DT104 spread among all domestic animals including cattle, poultry, pigs and sheep [Threlfall EJ. J Antimicrob Chemother. 2000]

History of DT104

MDR DT104 rapidly emerged globally in the 1990s and became the most prevalent reported phage type isolated from humans and animals in many countries

[Helms M, et al. Emerg Infect Dis. 2005]

[Threlfall EJ. J Antimicrob Chemother. 2000]

The first report of isolates from agricultural animals was in the UK in 1988

and in the US in 1990

[Threlfall EJ, et al. Vet Rec. 1994]

[Hancock D, et al. Emerging Diseases of Animals, 2000]

It emerged as multidrug-resistant (MDR) DT104 in the early 1980s in humans and birds from the UK

[Threlfall EJ. J Antimicrob Chemother. 2000]

Susceptible DT104 was first reported in 1960s in humans

[Threlfall EJ. J Antimicrob Chemother. 2000]

A decline in MDR DT104 has been reported in the last decade

1960

1970

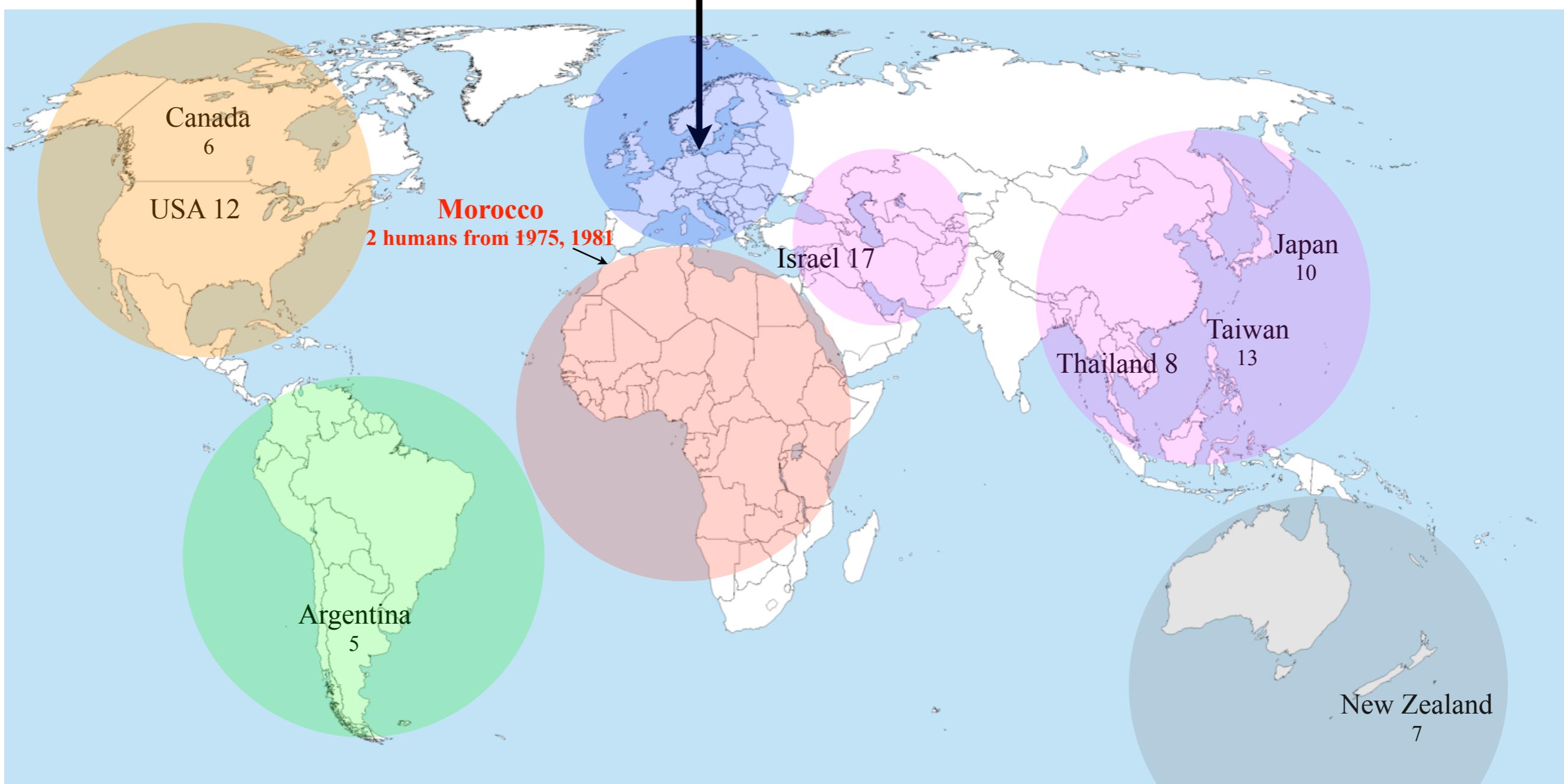
1980

1990

2000

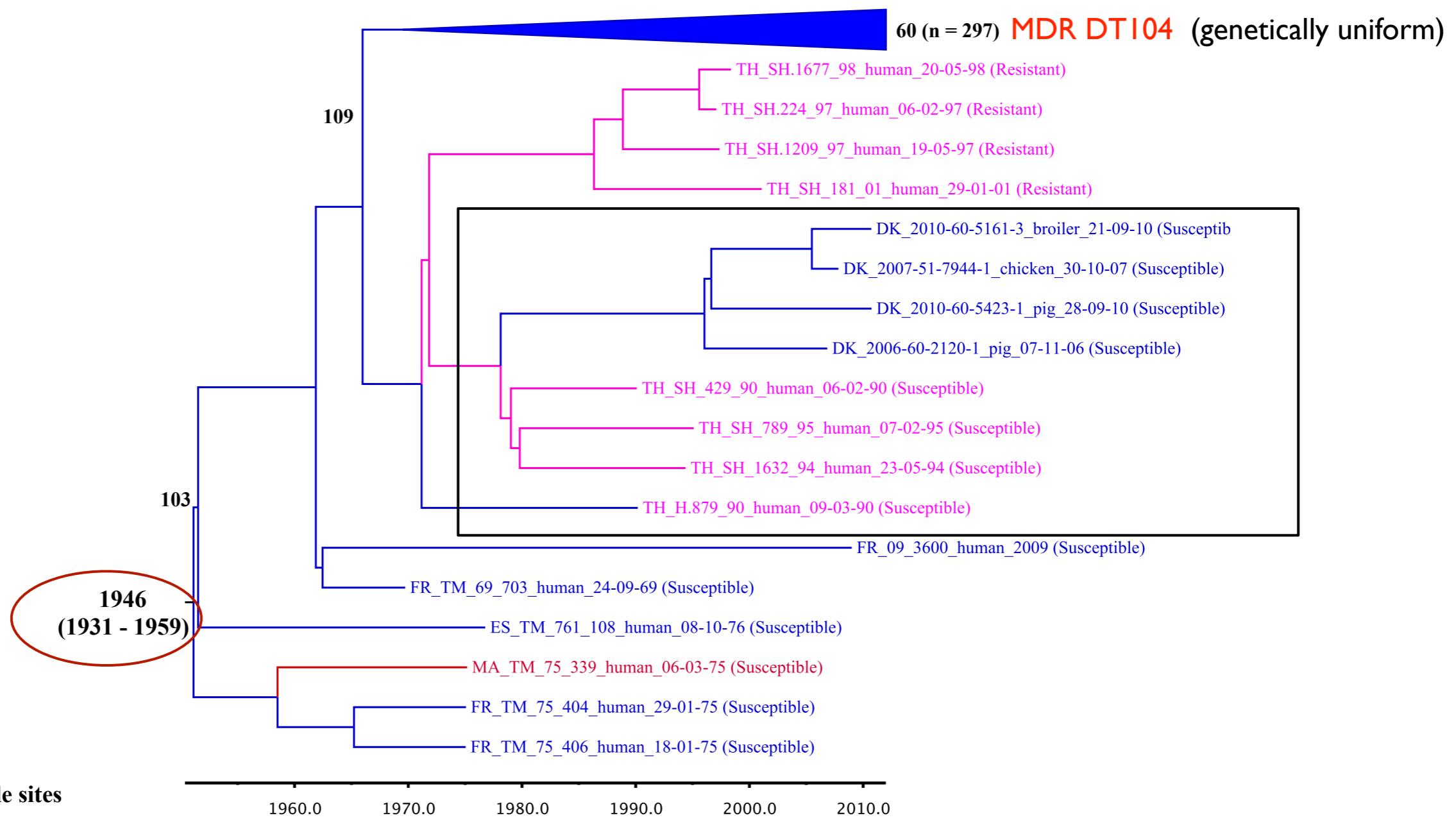
2010

Ireland 10	Scotland 14	Denmark 79
Spain	Austria 30	Netherlands 22
1 human from 1976	Germany 27	Switzerland 8
France 9	Luxembourg 13	Czech Republic 9
1 human from 1969	Poland 13	

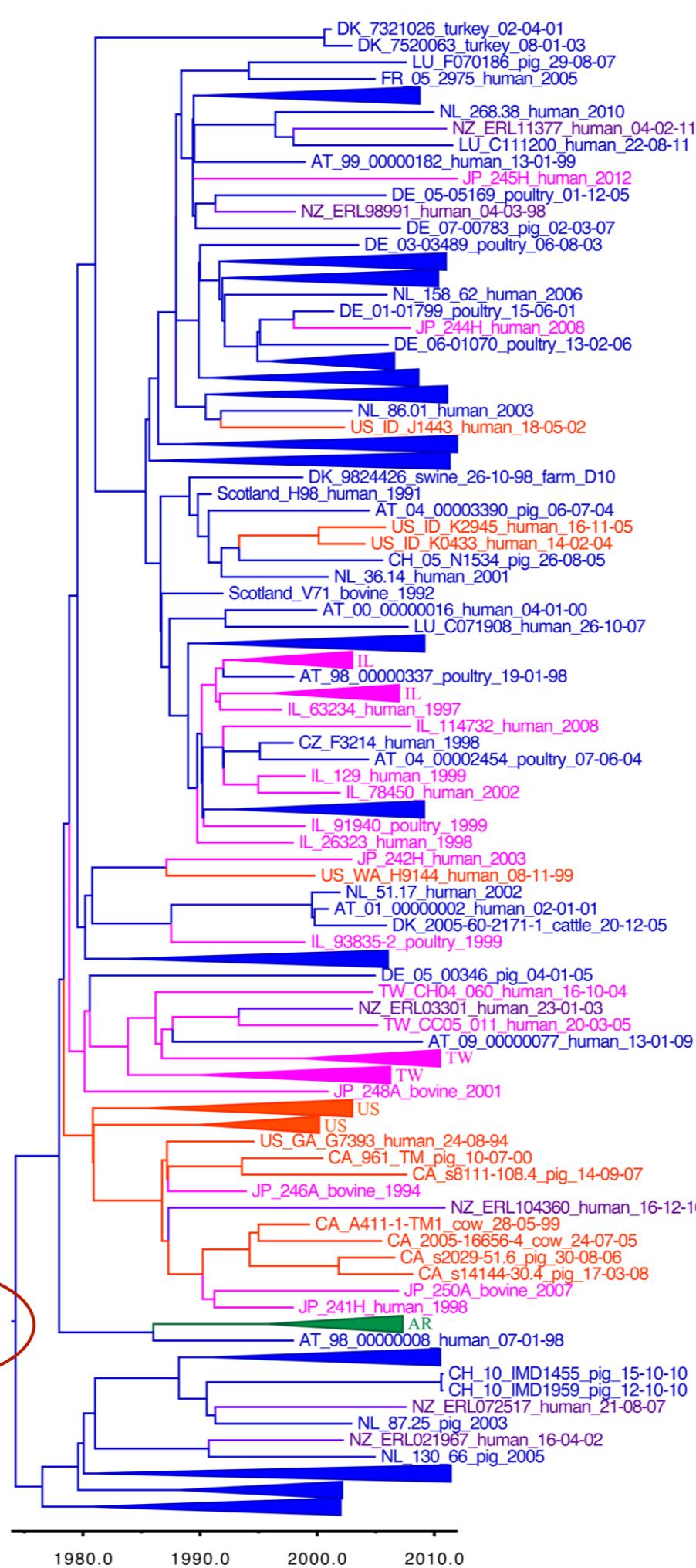


Global phylogeny of DT104

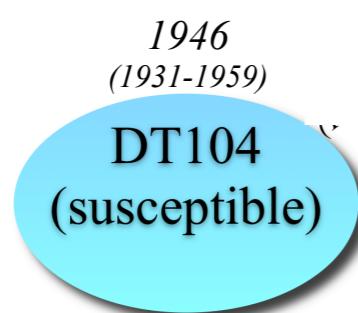
- Europe
- North America
- South America
- Asia
- Australia



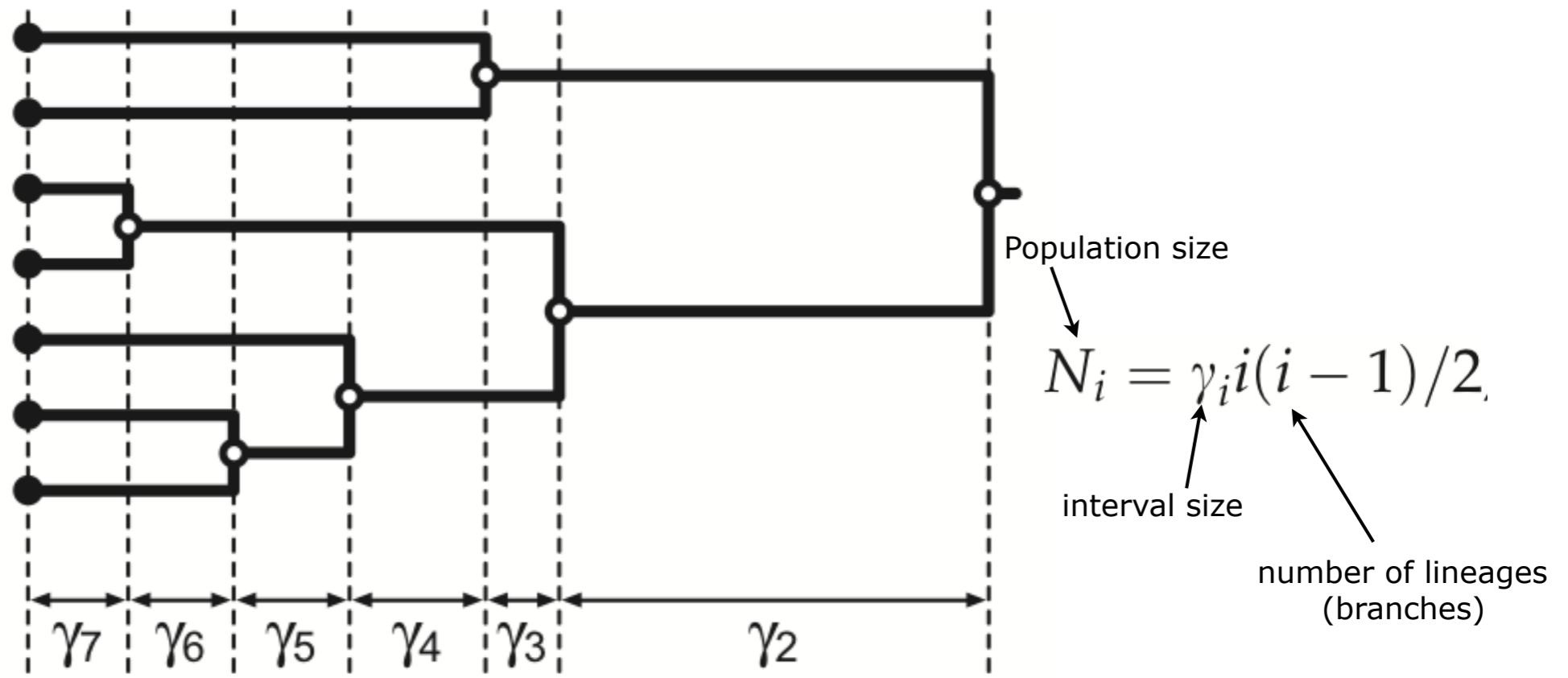
Temporal phylogenetic tree of MDR DT104

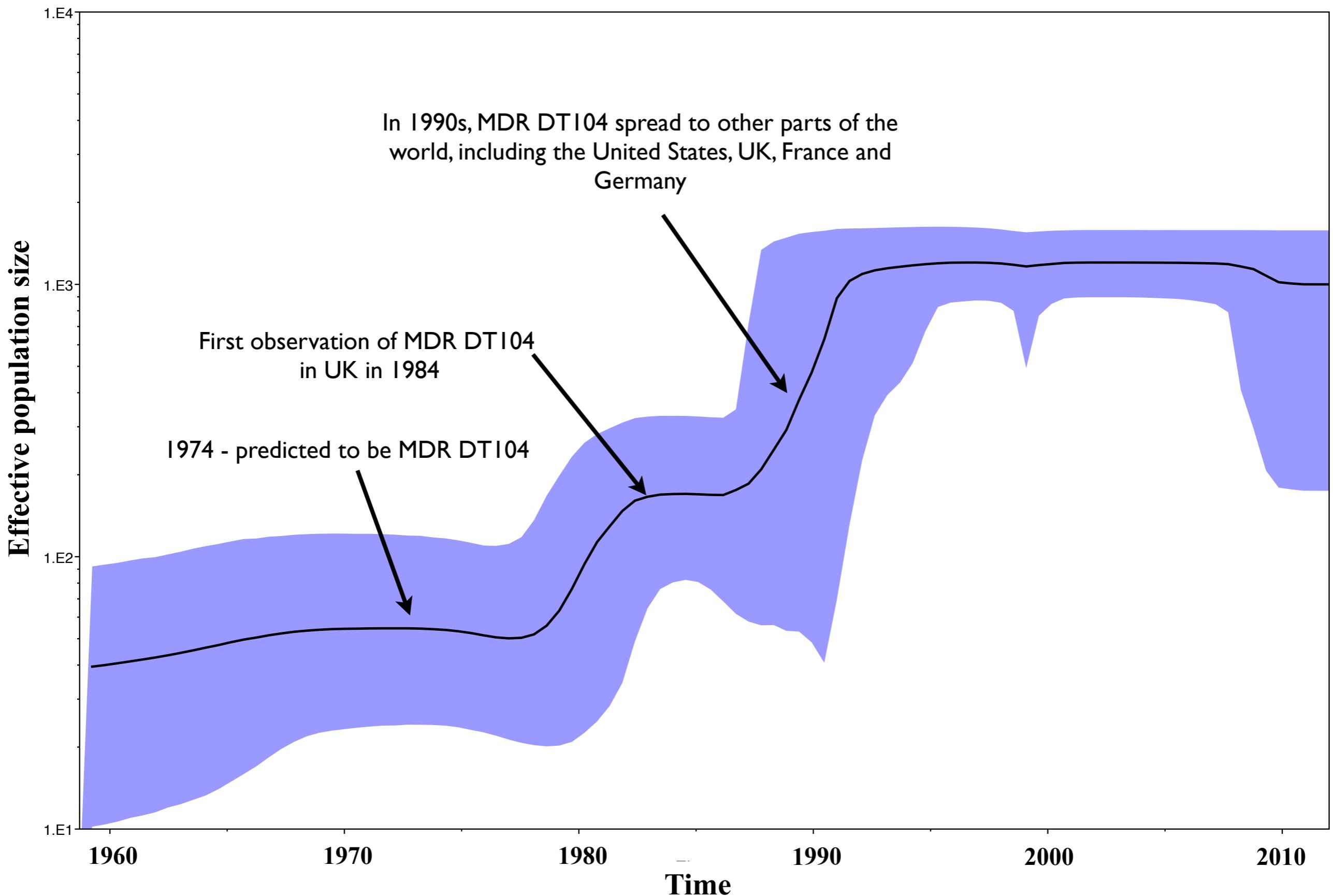


Evolution and global dissemination of DT104



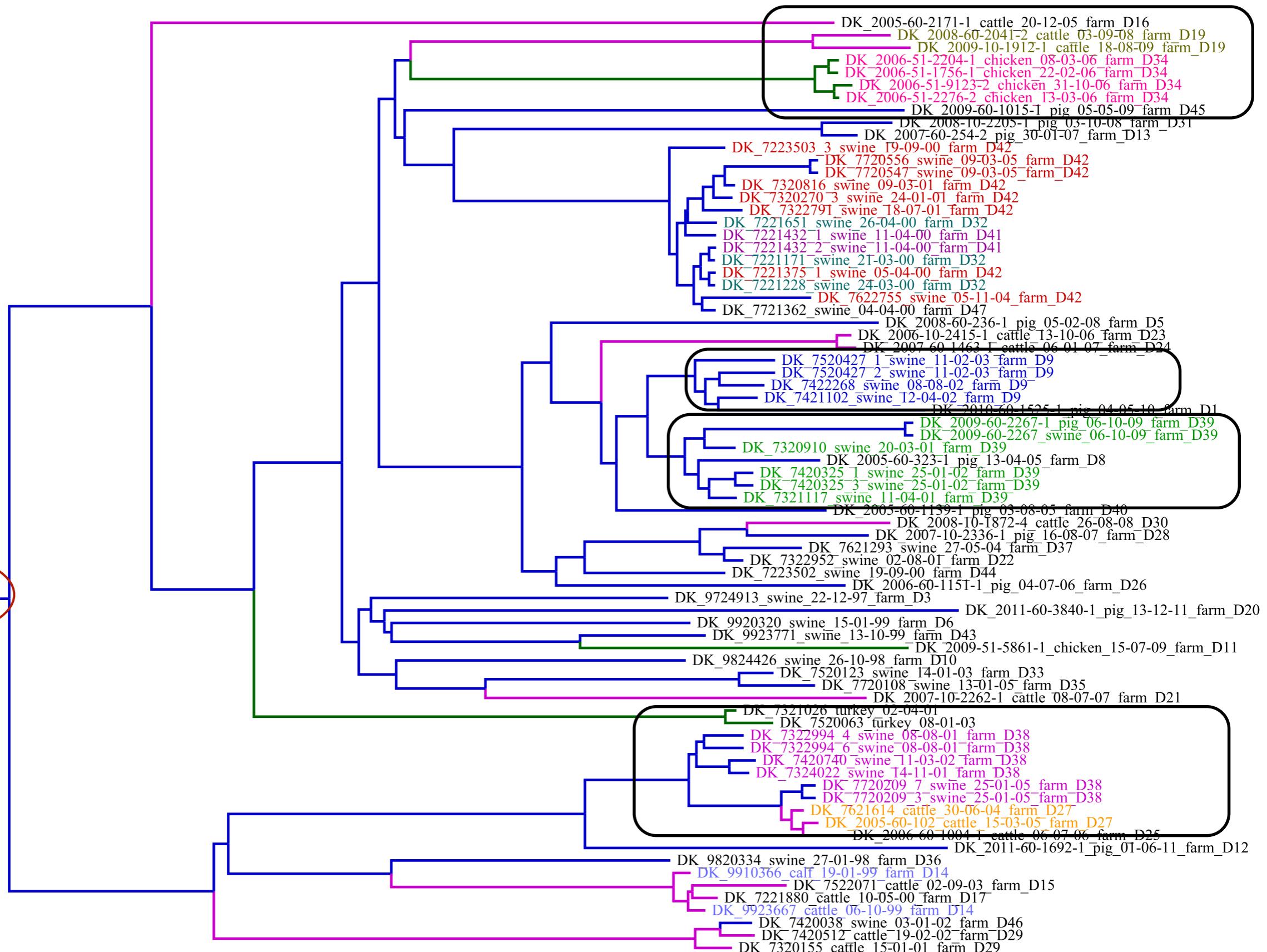
Demographic history using Bayesian skyline plot





Local phylogeny of DT104 (Denmark)

- Swine
- Cattle
- Poultry



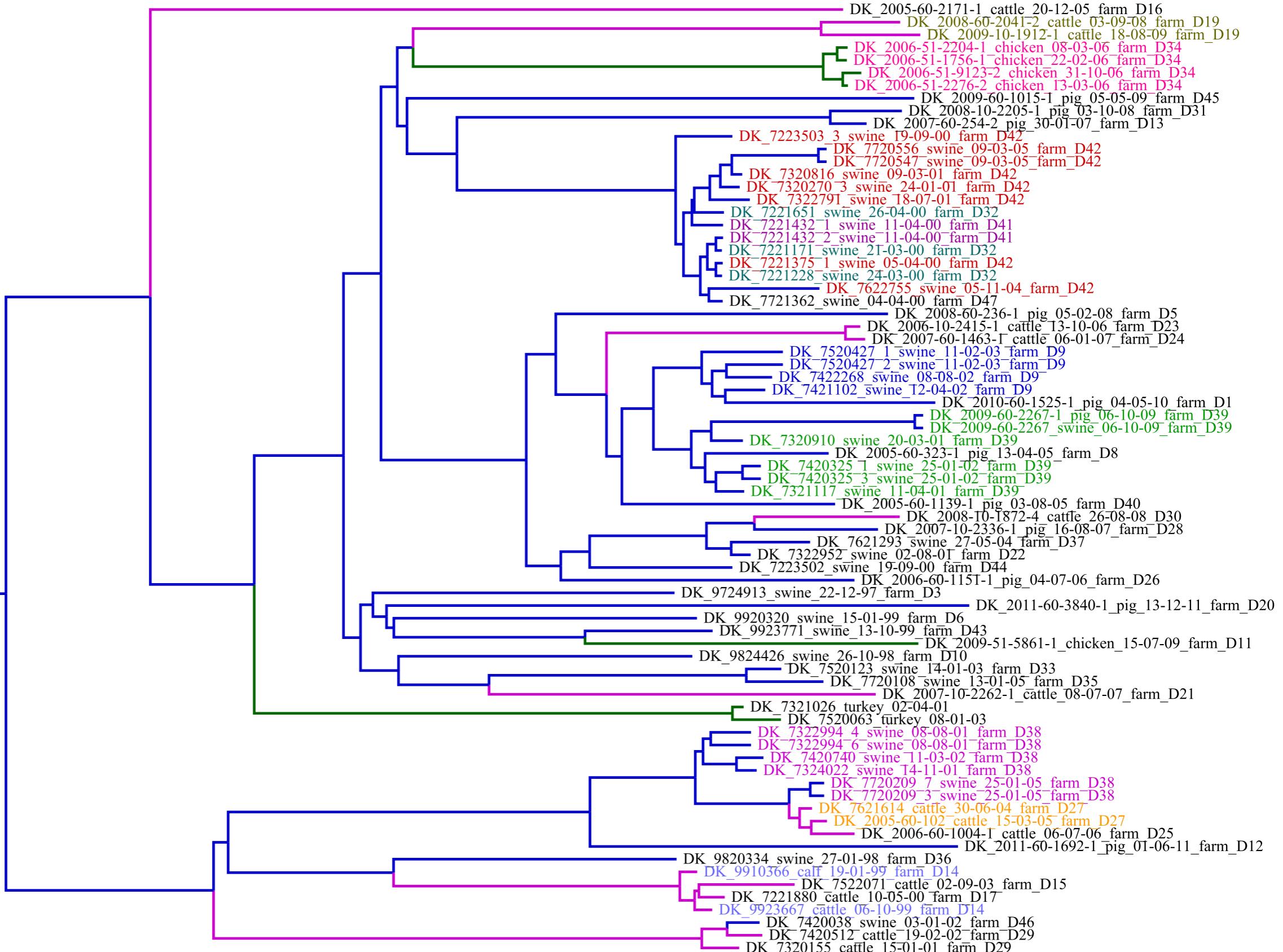
Local phylogeny of DT104 (Denmark)

Swine

Cattle

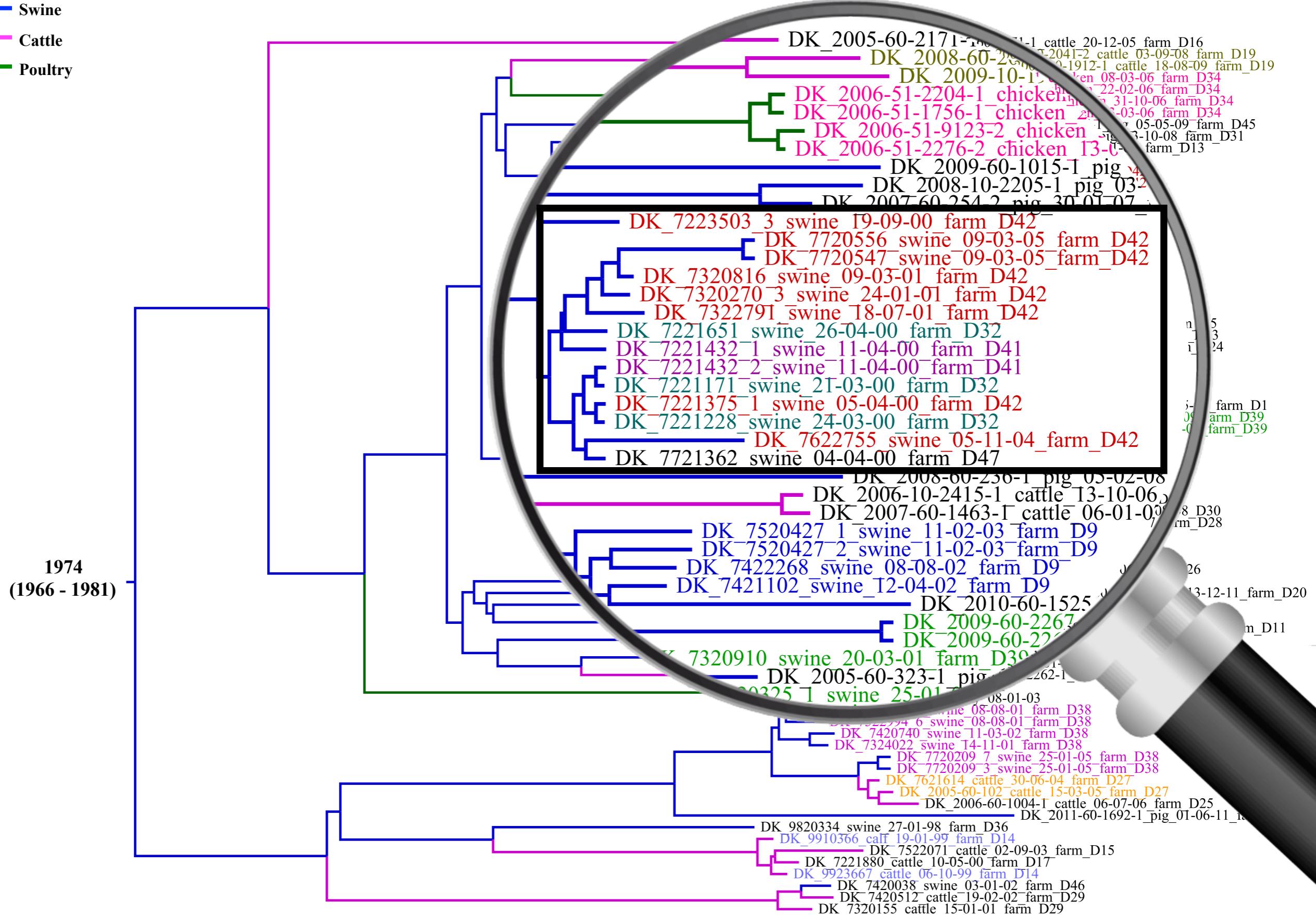
Poultry

1974
(1966 - 1981)



Local phylogeny of DT104 (Denmark)

- Swine
- Cattle
- Poultry



Search



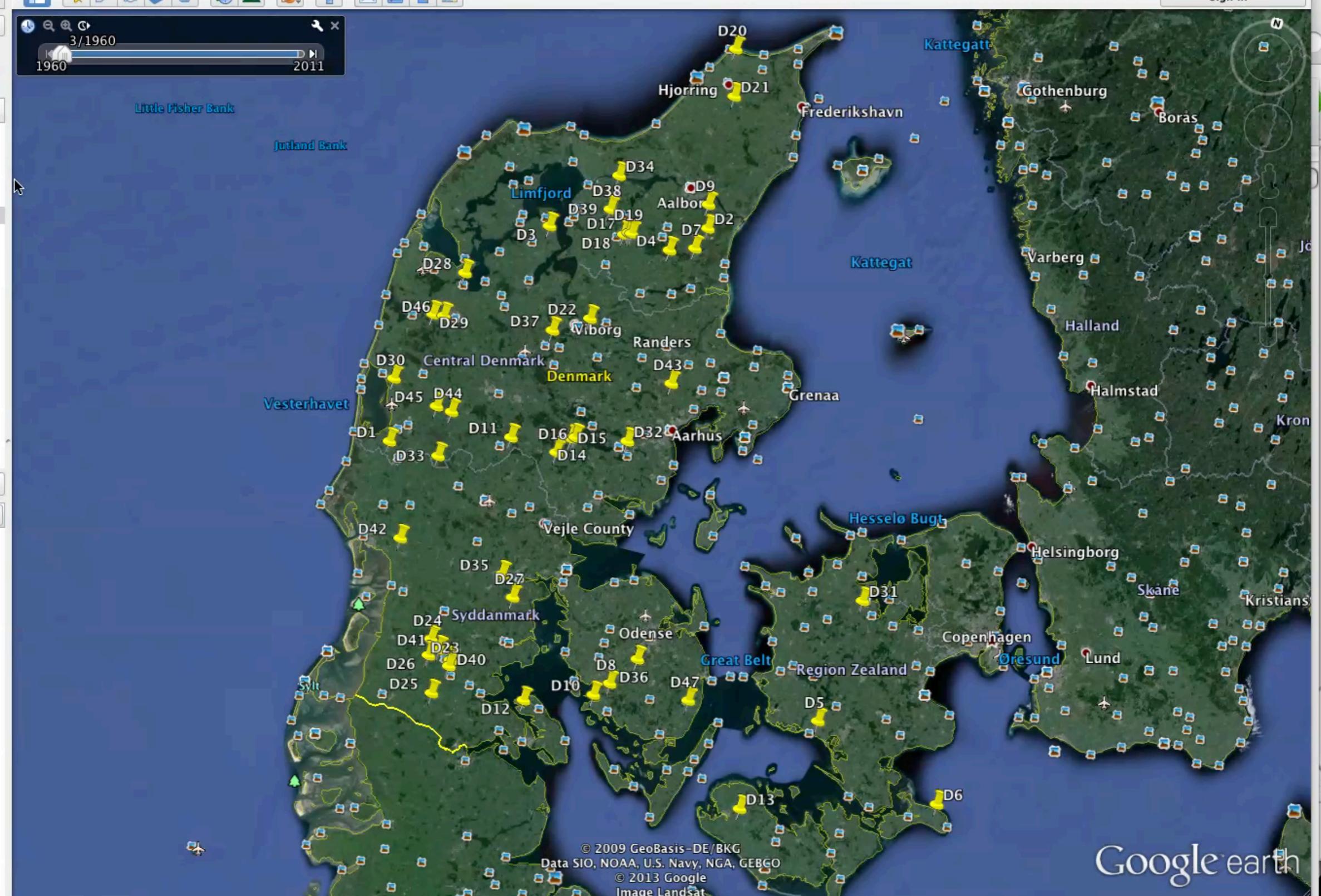
Get Directions History

our
Buildings
ed
es
able_fasta_danish...

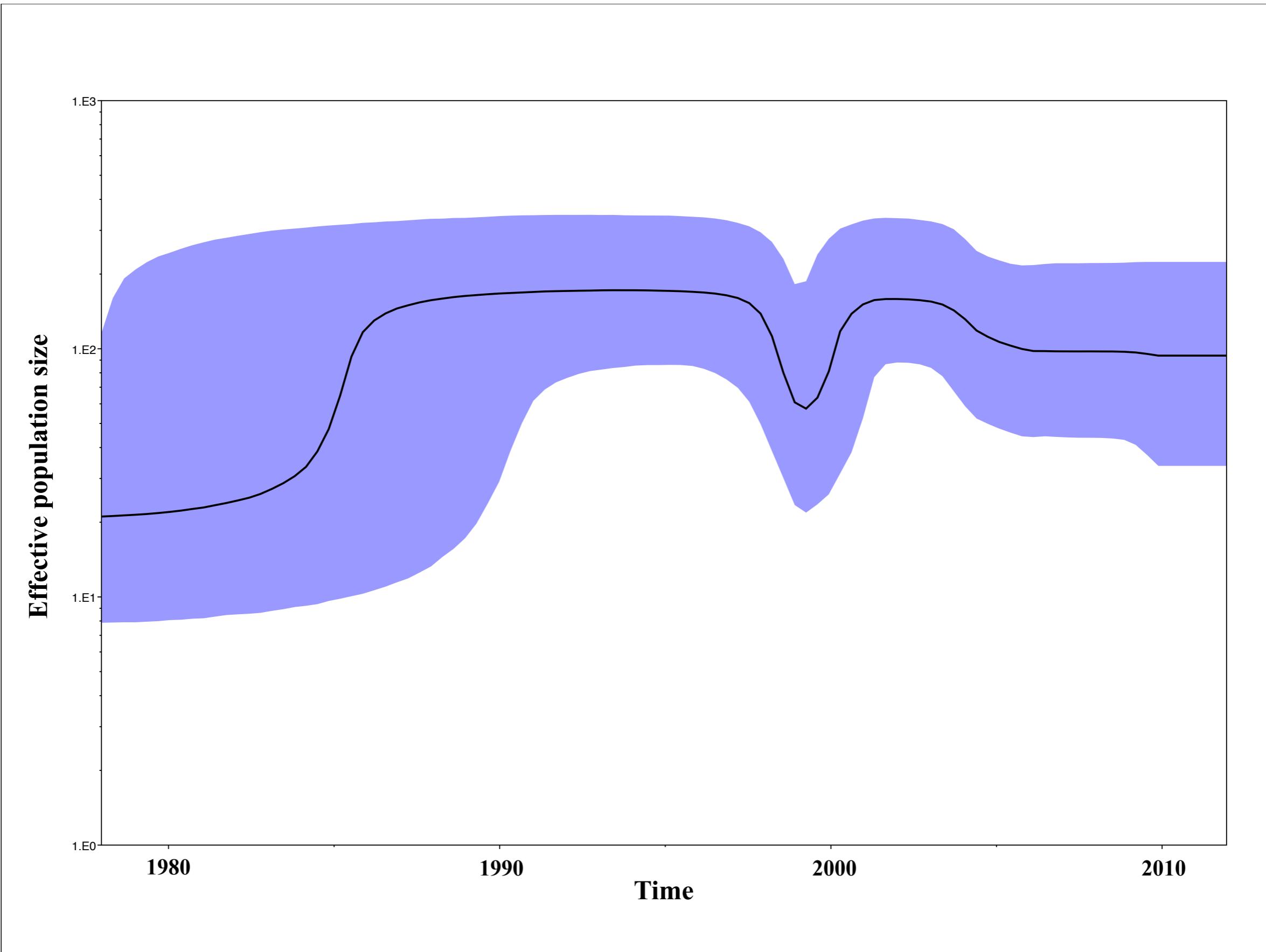
Earth Gallery >>

use
labels

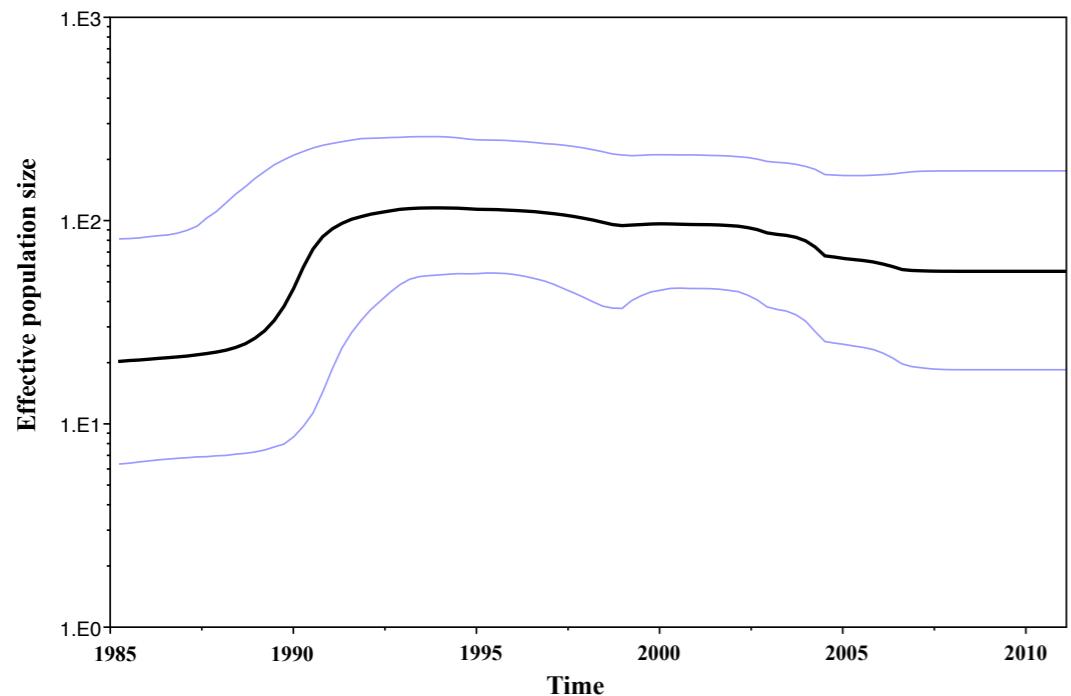
ness



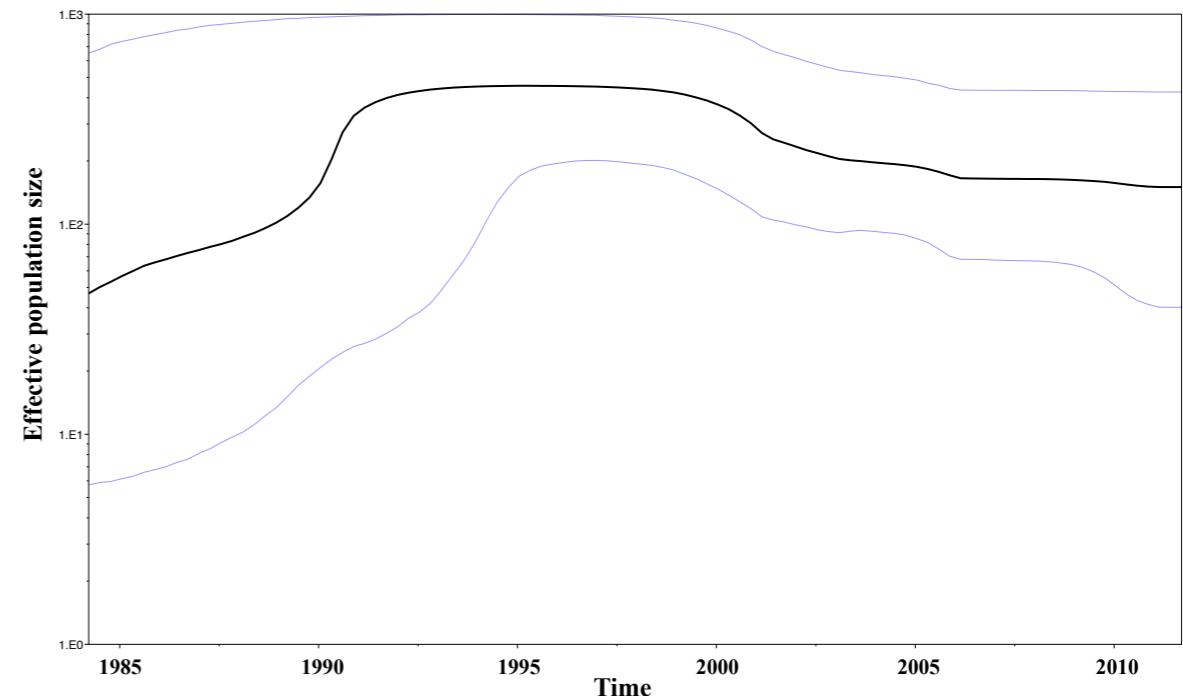
Demographic history of Danish MDR DT104



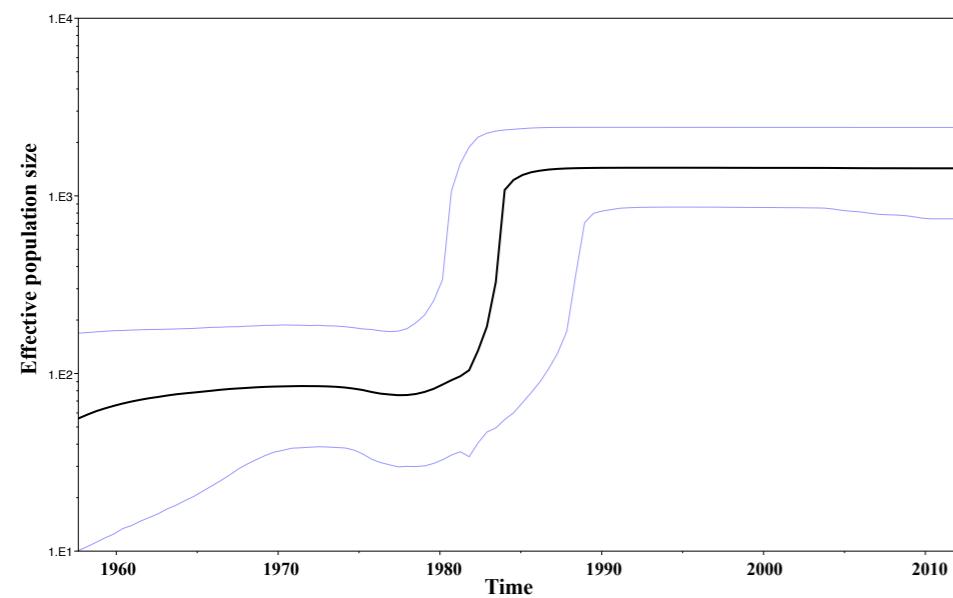
Cattle



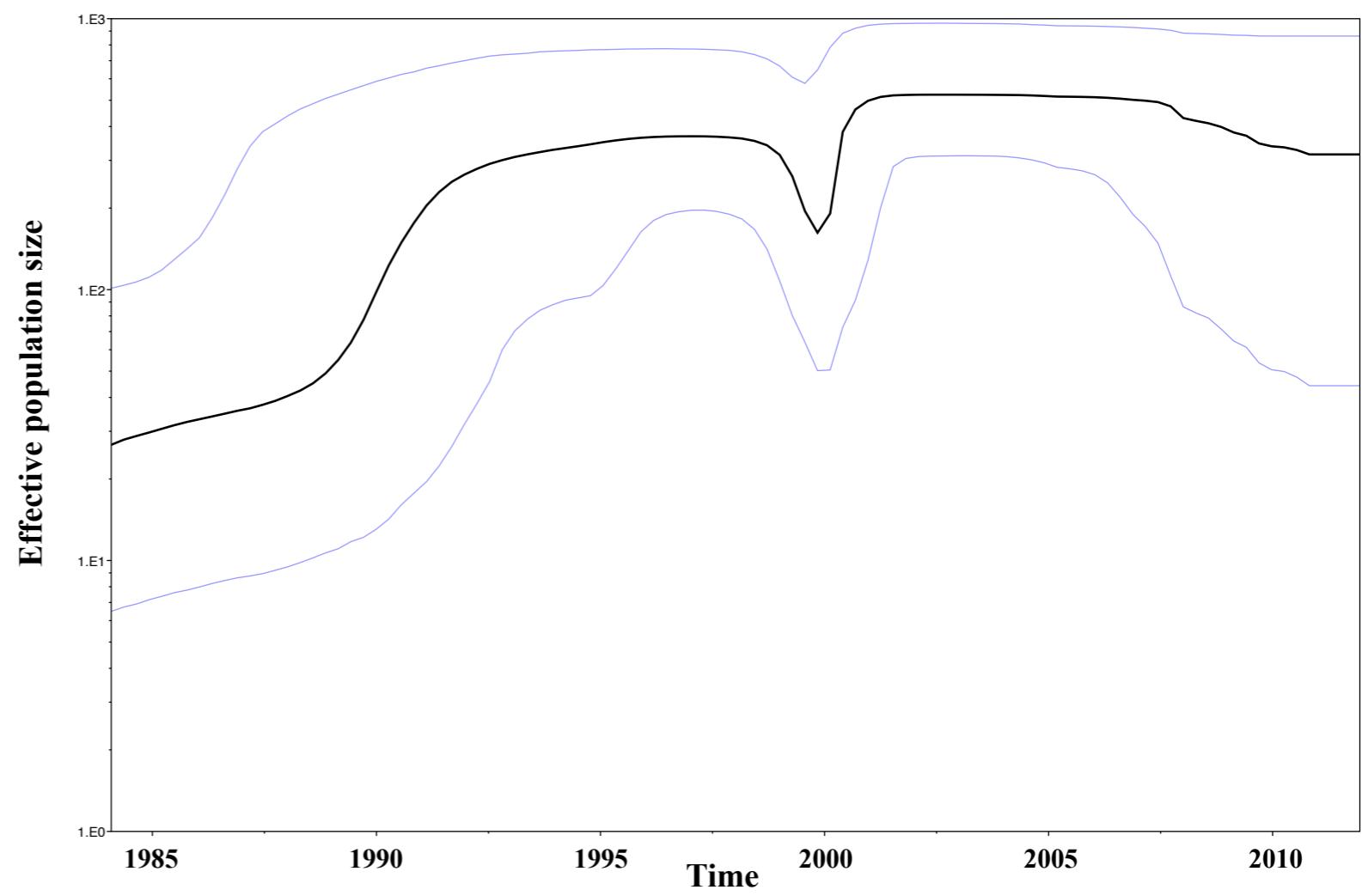
Poultry



Human



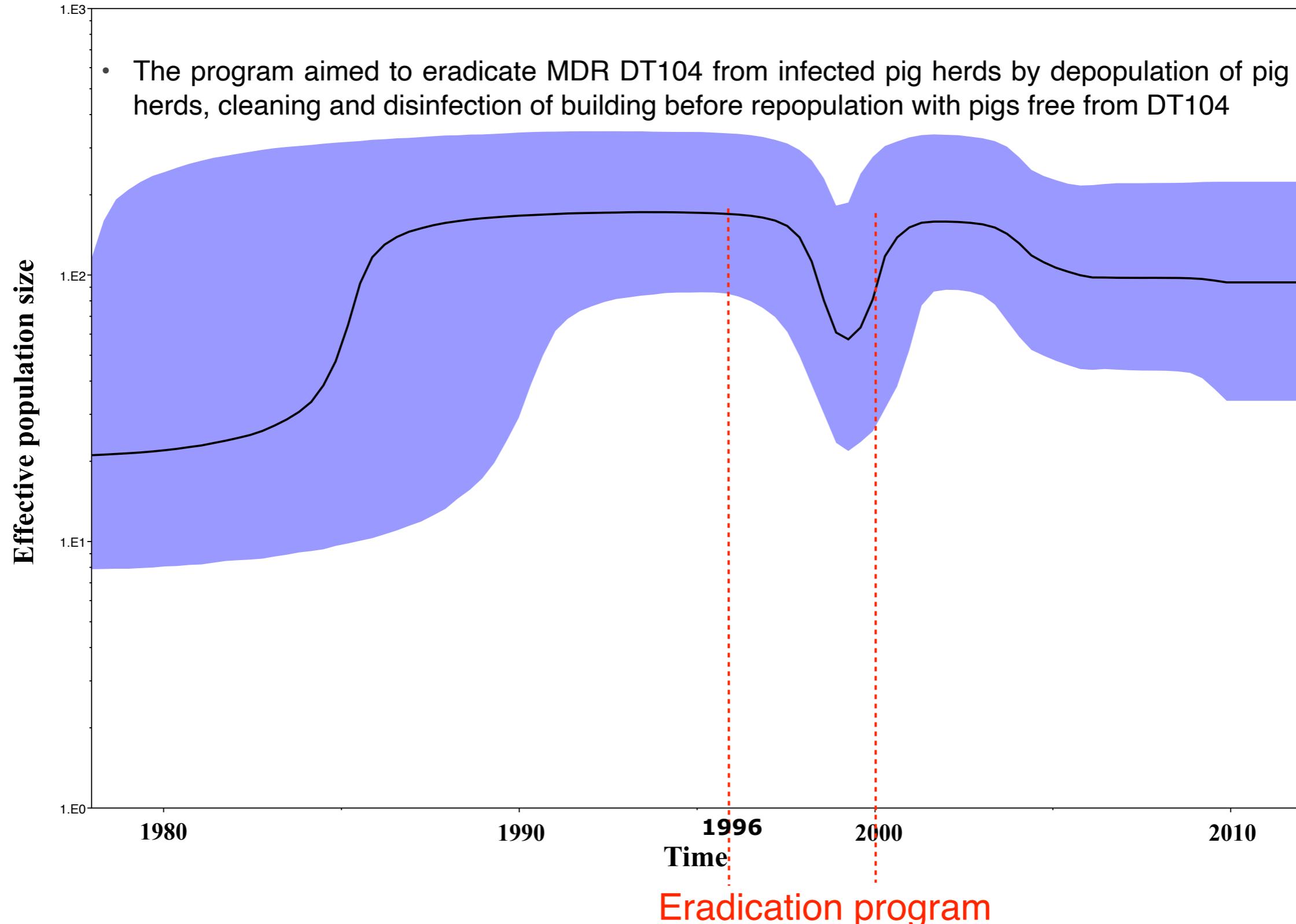
Swine

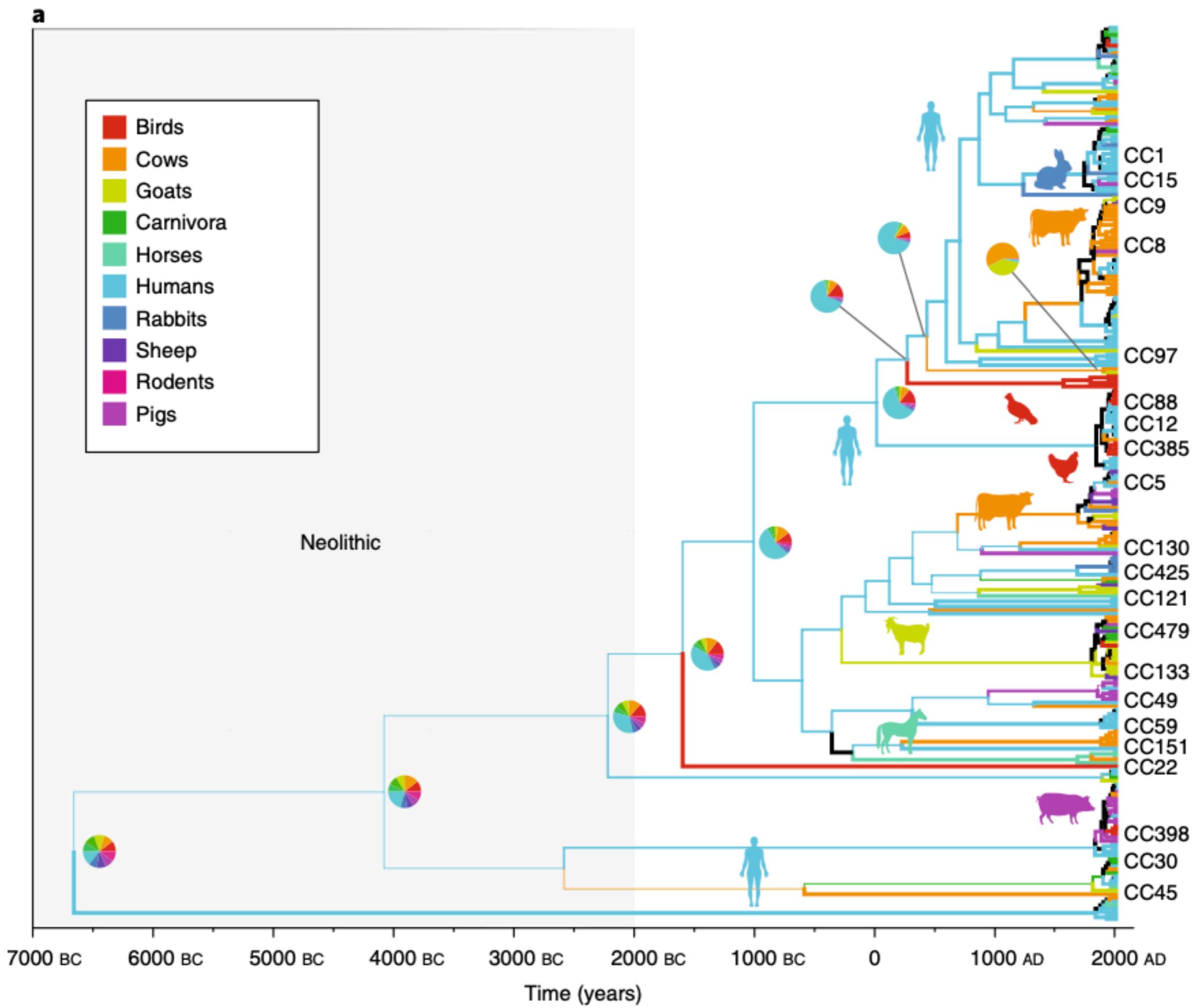


Demographic history of Danish MDR DT104

- The evidence of the success of the eradication program implemented by the Federation of Danish Pig Producers and Slaughterhouse in collaboration with the Danish Veterinary Service

- The program aimed to eradicate MDR DT104 from infected pig herds by depopulation of pig herds, cleaning and disinfection of building before repopulation with pigs free from DT104





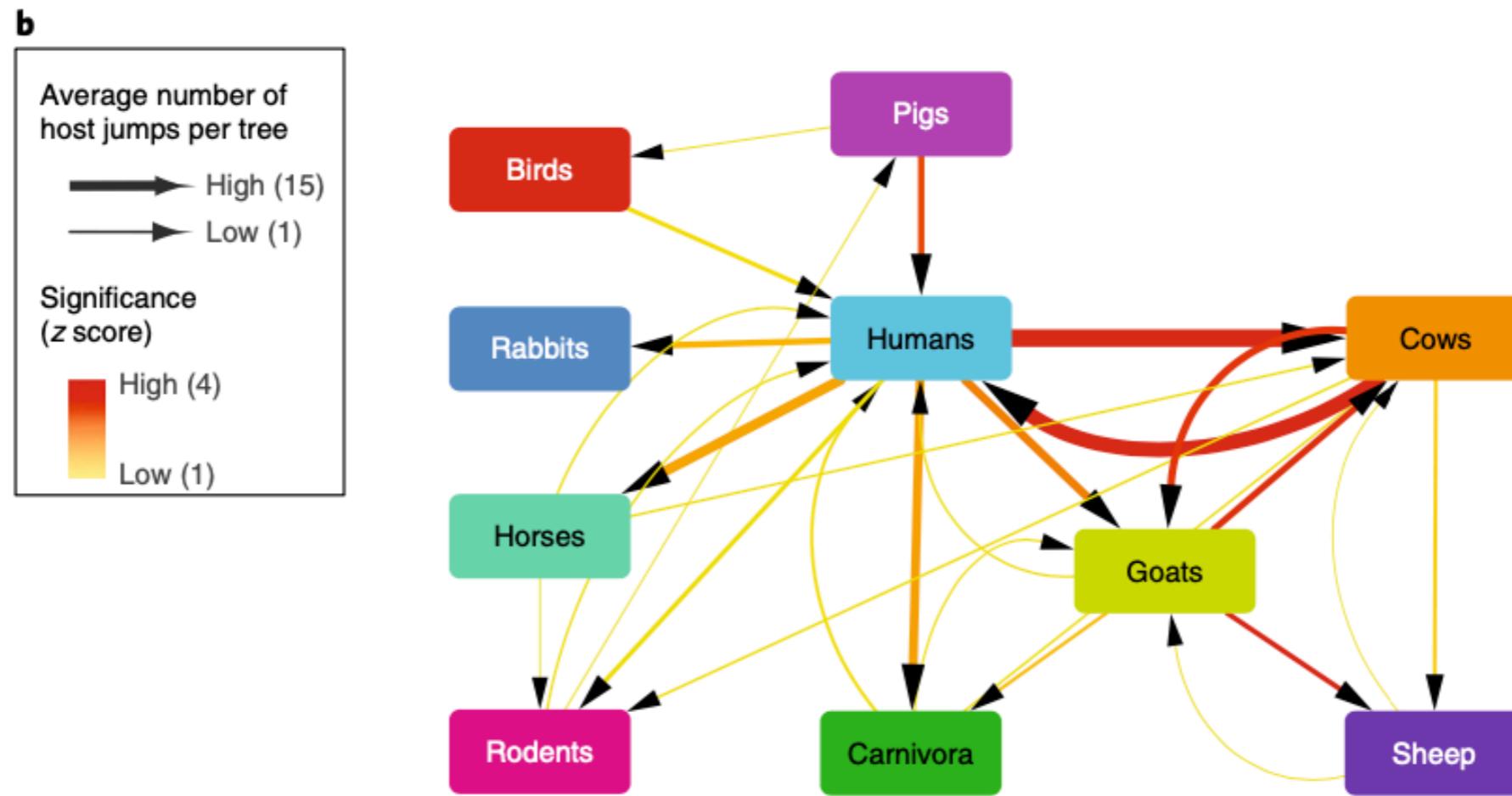


Fig. 2 | *S. aureus* has undergone extensive ancient and recent host-switching events, with humans acting as a major hub. **a**, Time-scaled phylogeny of a subsample of the *S. aureus* sequences. Clonal complexes are labelled, and branches are coloured according to the host-species group. Pie charts indicate the relative probability of host origin at the ancestral nodes, and line thickness corresponds to the probability of the majority host (see Supplementary Figs. 6–10 for all subsamples). **b**, Quantification of the number of host-switching events, illustrated as a host-transition count network based on BEAST Markov Jumps models averaged over all subsamples of the data. The line width represents the Markov Jump count per tree averaged over all subsamples (Supplementary Figs. 4 and 5), and line colour represents the significance compared with permuted label analysis (z score). Only transitions with higher counts compared with models with permuted host labels are shown (z score ≥ 0.5).

BEAST exercise

- **Exercise:** The swine-origin influenza A outbreak

- Make xml file using BEAUTi (on your own computer) to generate xml file for BEAST using [S-OIV.nex](#) as an input file (chain length of **10,000,000** and sample every [Log every] **1000** generations). SOIV.nex contains an alignment of 50 genomes (concatenated segments), 13109 nucleotides in length. [Link to download BEAST 2 including BEAUTi](#). [Link to tutorial](#) to help setting up substitution model, molecular clock model, tree prior and priors.

```
/home/projects/course_23262/course/week07/beast/S-OIV_v2/S-OIV.xml
```

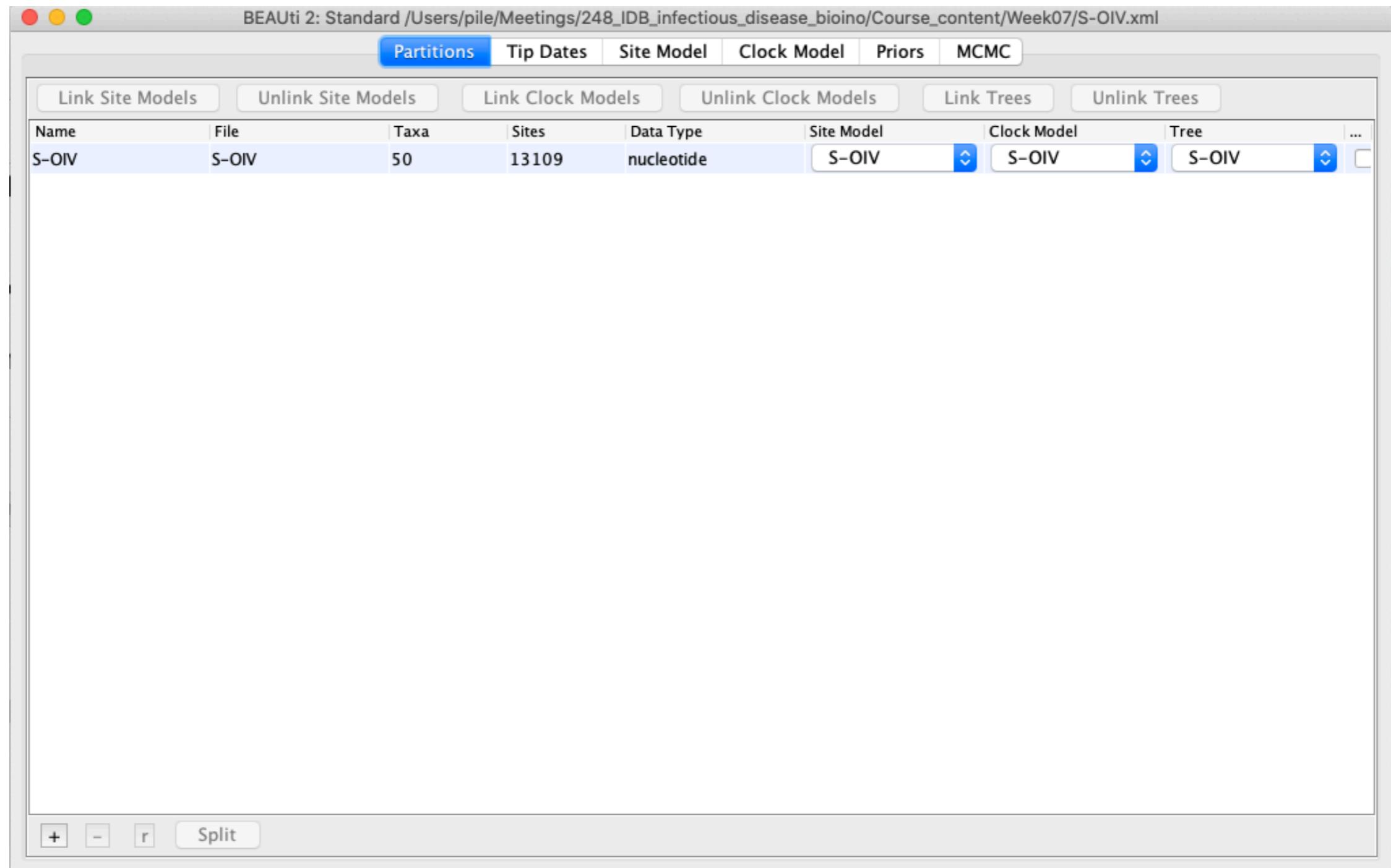
- Run BEAST simulation (on Computerome)
- Check the quality of BEAST runs using Tracer (on your own computer using output file with extension .log). [Link to download Tracer](#).
- Summarise the trees using Treeannotator (on Computerome using output file with extension .trees). Treeannotator is included in BEAST package. Use 10% burnin (with a chain of 10,000,000 steps, sampling every 1000 steps, there are 10000 trees in the file. To obtain the 10% burnin, set the burnin to 1000)

```
/home/projects/course_23262/course/week07/beast/S-OIV_v2/S-OIV.trees
```

- Visualise the output tree in FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>)

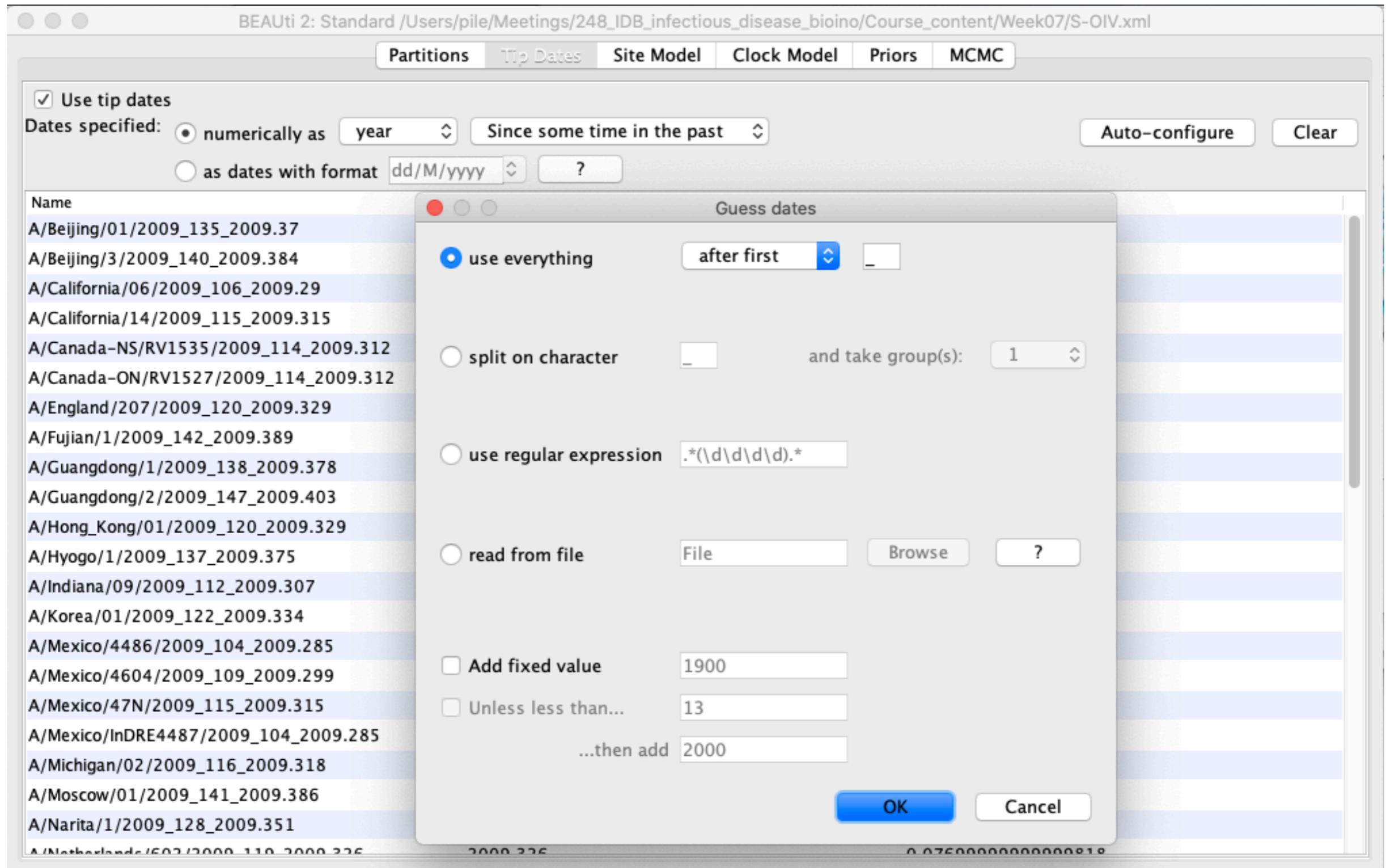
Set up and guest date in BEAUTi

Open nex file in BEAUTi



Set up and guess date in BEAUTi

Guess dates



Set up and guess date in BEAUTi

Guess dates

BEAUTi 2: Standard /Users/pile/Meetings/248_IDB_infectious_disease_bioinfo/Course_content/Week07/S-OIV.xml

Partitions Tip Dates Site Model Clock Model Priors MCMC

Use tip dates

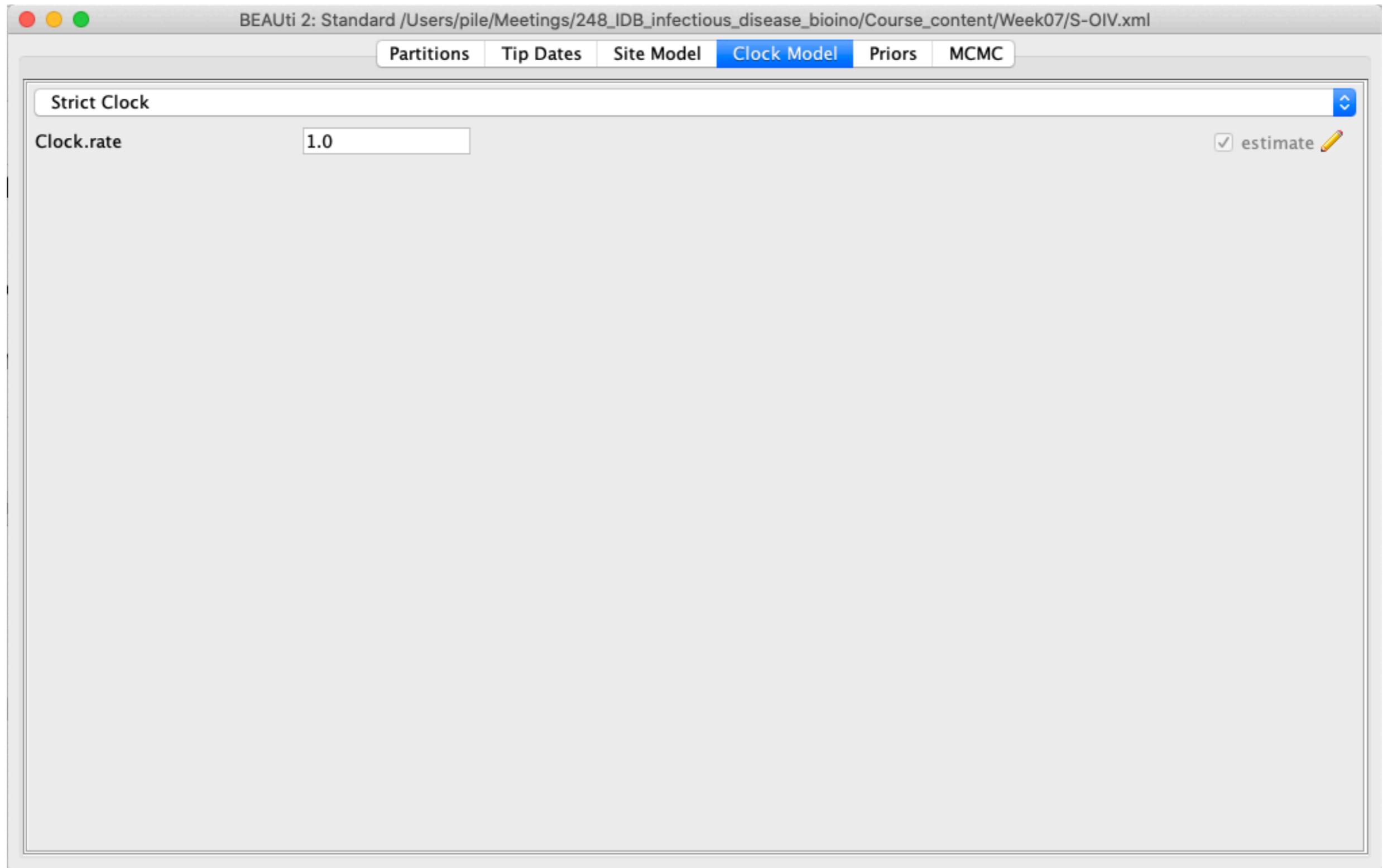
Dates specified: numerically as year Since some time in the past

as dates with format dd/M/yyyy ?

Name	Date (raw value)	Height
A/Beijing/01/2009_135_2009.37	2009.37	0.03300000000012915
A/Beijing/3/2009_140_2009.384	2009.384	0.01900000000005457
A/California/06/2009_106_2009.29	2009.29	0.1130000000005639
A/California/14/2009_115_2009.315	2009.315	0.0879999999996544
A/Canada-NS/RV1535/2009_114_2009.312	2009.312	0.09100000000012187
A/Canada-ON/RV1527/2009_114_2009.312	2009.312	0.09100000000012187
A/England/207/2009_120_2009.329	2009.329	0.0740000000006912
A/Fujian/1/2009_142_2009.389	2009.389	0.014000000000123691
A/Guangdong/1/2009_138_2009.378	2009.378	0.02500000000009095
A/Guangdong/2/2009_147_2009.403	2009.403	0.0
A/Hong_Kong/01/2009_120_2009.329	2009.329	0.0740000000006912
A/Hyogo/1/2009_137_2009.375	2009.375	0.0280000000002001
A/Indiana/09/2009_112_2009.307	2009.307	0.0960000000000364
A/Korea/01/2009_122_2009.334	2009.334	0.0689999999995998
A/Mexico/4486/2009_104_2009.285	2009.285	0.1179999999993815
A/Mexico/4604/2009_109_2009.299	2009.299	0.1040000000004184
A/Mexico/47N/2009_115_2009.315	2009.315	0.0879999999996544
A/Mexico/InDRE4487/2009_104_2009.285	2009.285	0.1179999999993815
A/Michigan/02/2009_116_2009.318	2009.318	0.0850000000003638
A/Moscow/01/2009_141_2009.386	2009.386	0.0170000000005275
A/Narita/1/2009_128_2009.351	2009.351	0.0519999999990723
A/Netherlands/1607/2009_110_2009.226	2009.226	0.076000000000819

Set up and guest date in BEAUTi

Choose “Clock Model”



Set up and guest date in BEAUTi

Choose “Priors”

The screenshot shows the BEAUTi software interface with the title bar "BEAUTi 2: Standard /Users/pile/Meetings/248_IDB_infectious_disease_bioinfo/Course_content/Week07/S-OIV.xml". The "Priors" tab is selected in the top navigation bar. The main area displays various parameters and their prior distributions:

- Tree.t:S-OIV**: Coalescent Exponential Population
- clockRate.c:S-OIV**: Uniform, initial = [1.0] $[-\infty, \infty]$, substitution rate of partition c:S-OIV
- ePopSize.t:S-OIV**: 1/X, initial = [0.3] $[-\infty, \infty]$, Coalescent population size parameter of partition t:S-OIV
- freqParameter.s:S-OIV**: Uniform, initial = [0.25] [0.0,1.0]
- growthRate.t:S-OIV**: Laplace Distribution, initial = [3.0E-4] $[-\infty, \infty]$, Coalescent growth rate of partition t:S-OIV
- kappa.s:S-OIV**: Log Normal, initial = [2.0] [0.0, ∞], HKY transition-transversion parameter of partition s:S-OIV

A "+ Add Prior" button is located at the bottom right of the parameter list.

Set up and guest date in BEAUTi

Set “Chain Length” and “Log Every” AND save the file as ‘.xml’

BEAUTi 2: Standard /Users/pile/Meetings/248_IDB_infectious_disease_bioinfo/Course_content/Week07/S-OIV.xml

Partitions Tip Dates Site Model Clock Model Priors MCMC

Chain Length 10000000

Store Every -1

Pre Burnin 0

Num Initialization Attempts 10

▼ tracelog

File Name S-OIV.log

Log Every 1000

Mode autodetect

Sort smart

Sanitise Headers

posterior
likelihood
prior
treeLikelihood.S-OIV
TreeHeight.t:S-OIV
clockRate.c:S-OIV
kappa.s:S-OIV
CoalescentExponential.t:S-OIV
ePopSize.t:S-OIV
growthRate.t:S-OIV
freqParameter.s:S-OIV

► screenlog
► treelog.t:S-OIV

Running BEAST on Computerome

- Look at the file beast.sh on DTU Learn as an example to run beast and treennotator
- run BEAST

```
/home/projects/course_23262/course/week07/beast/beast_v2/bin/beast /home/projects/course_23262/course/week07/beast/S-0IV_v2/S-0IV.xml
```

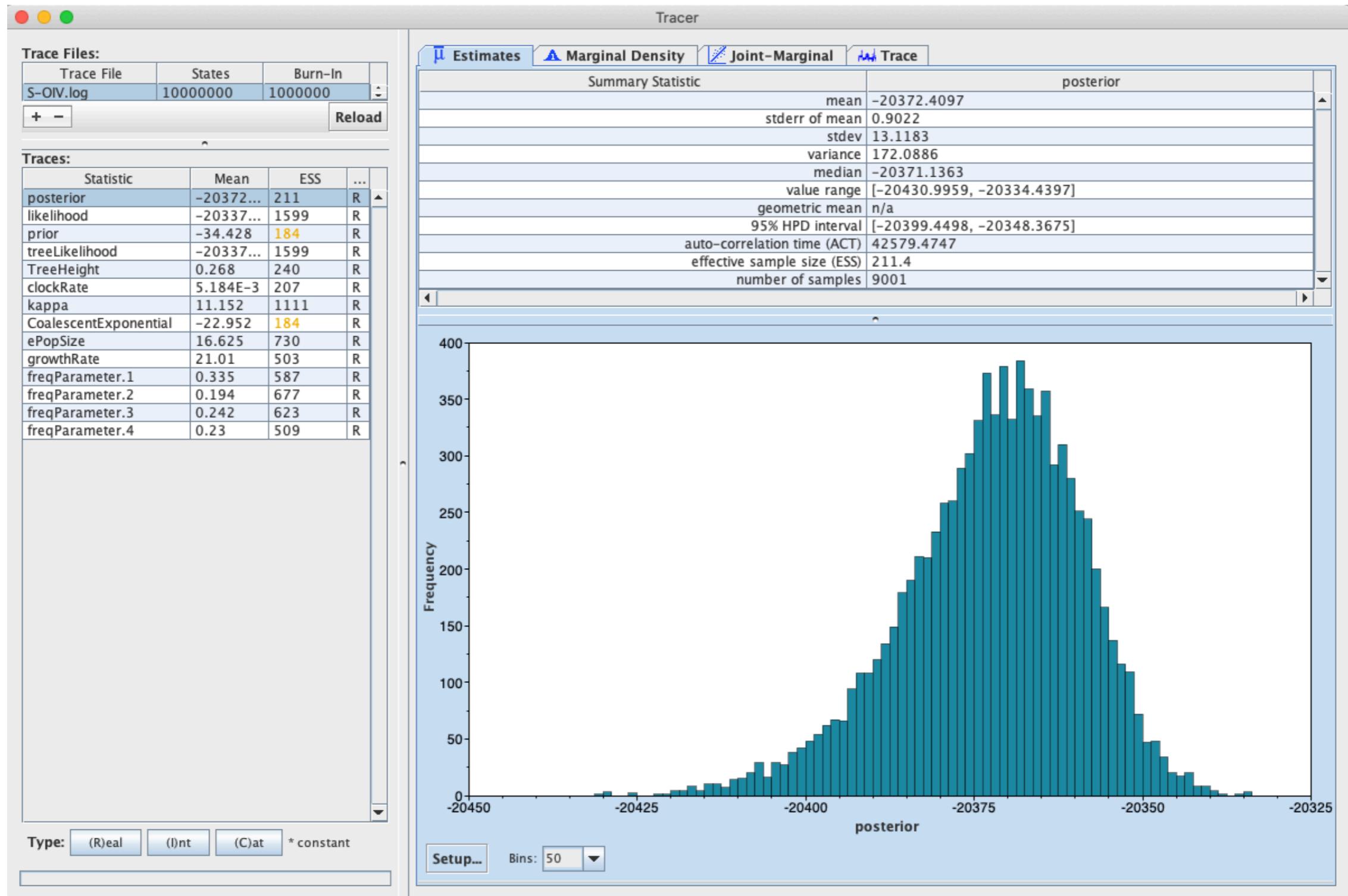
- run treeannotator

```
/home/projects/course_23262/course/week07/beast/beast_v2/bin/treeannotator  
-burnin 10 -heights mean /home/projects/course_23262/course/week07/beast/S-  
0IV_v2/S-0IV.trees S-0IV.tree
```

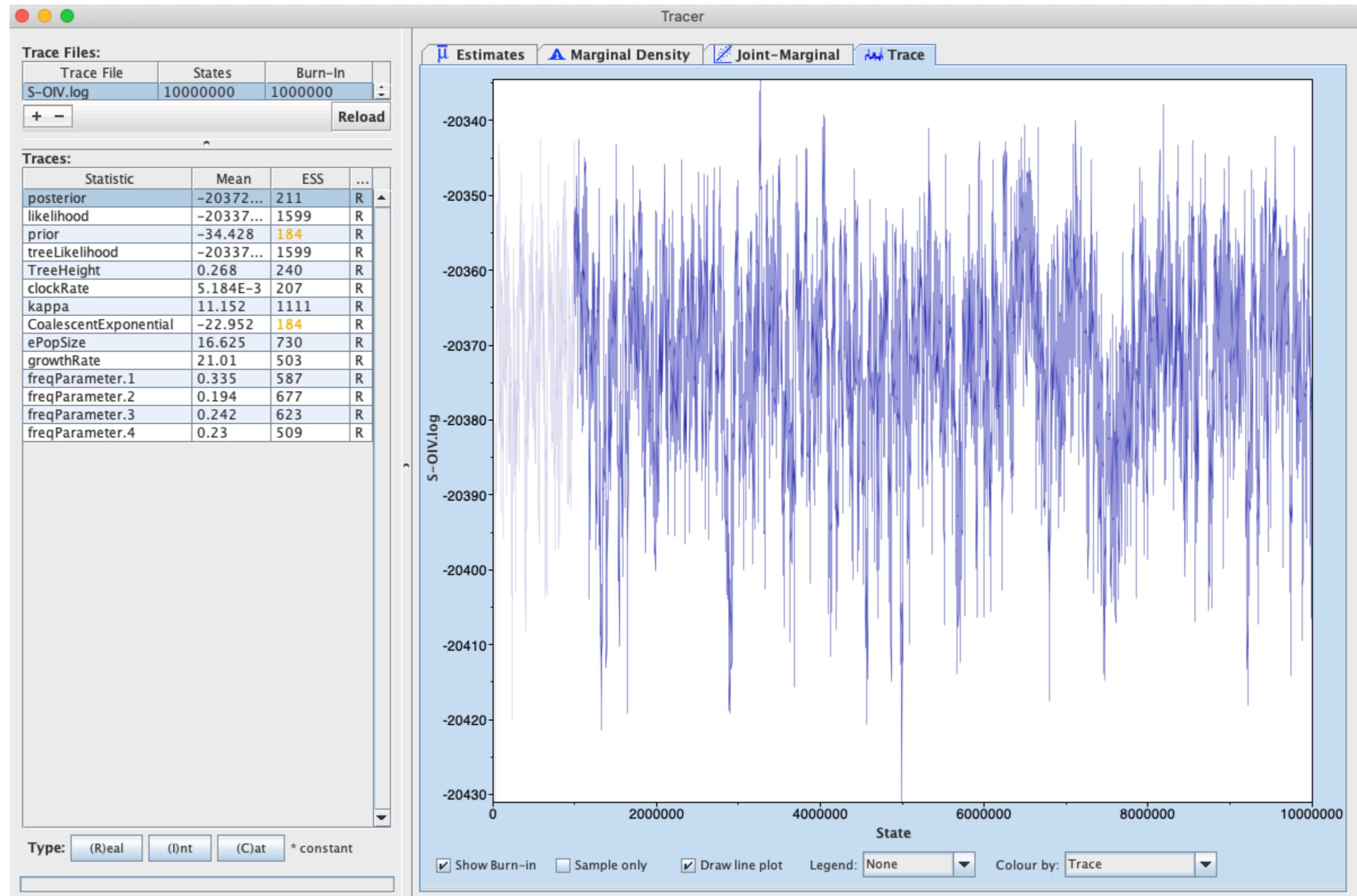
Check BEAST output in Tracer

Open .log file in Tracer

```
/home/projects/course_23262/course/week07/beast/S-OIV_v2/S-OIV.log
```

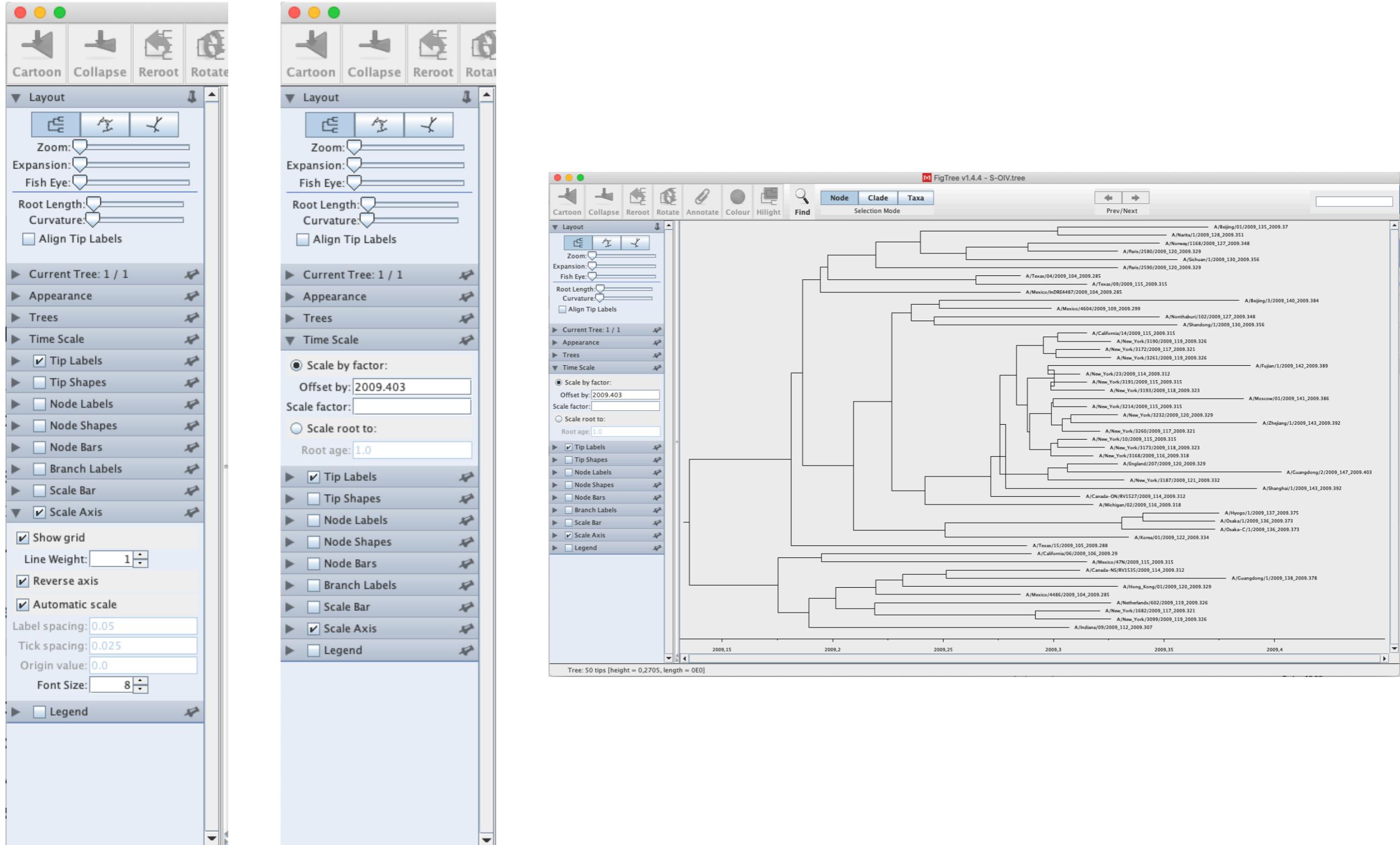


Check BEAST output in Tracer



Visualize BEAST tree using FigTree

Open S-OIV.tree using FigTree (the S-OIV.tree can be found on DTU Learn)



Other things to consider

- Recombination
 - Recombinant regions were excluded using Gubbins v2.4.1
 - Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015 Feb 18;43(3):e15.
 - <https://github.com/nickjcroucher/gubbins>
- Robustness of chosen molecular clock
 - To test the robustness of the molecular clock signal, ten further beast runs with randomized tip dates were generated using the same model
 - Rieux A, Khatchikian CE. Tipdatingbeast: an R package to assist the implementation of phylogenetic tip-dating tests using beast. Mol Ecol Resour. 2017 Jul;17(4):608-613.
 - <https://cran.r-project.org/web/packages/TipDatingBeast/index.html>