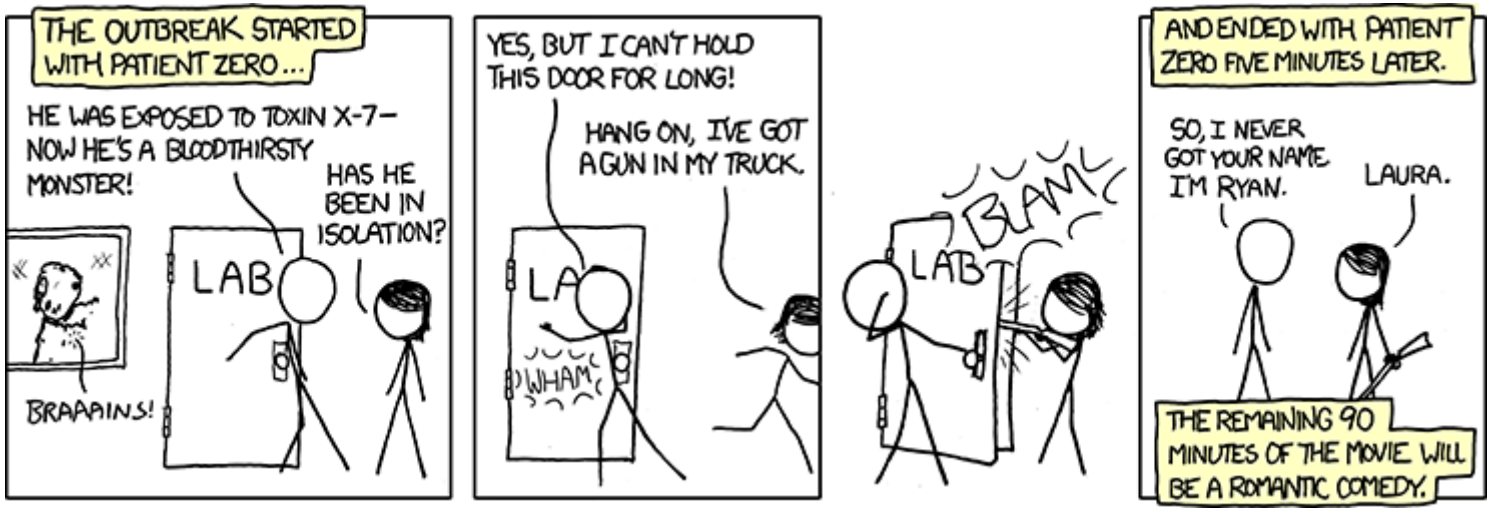
MIMM4750G

SIR and the BEAST



Models of epidemics

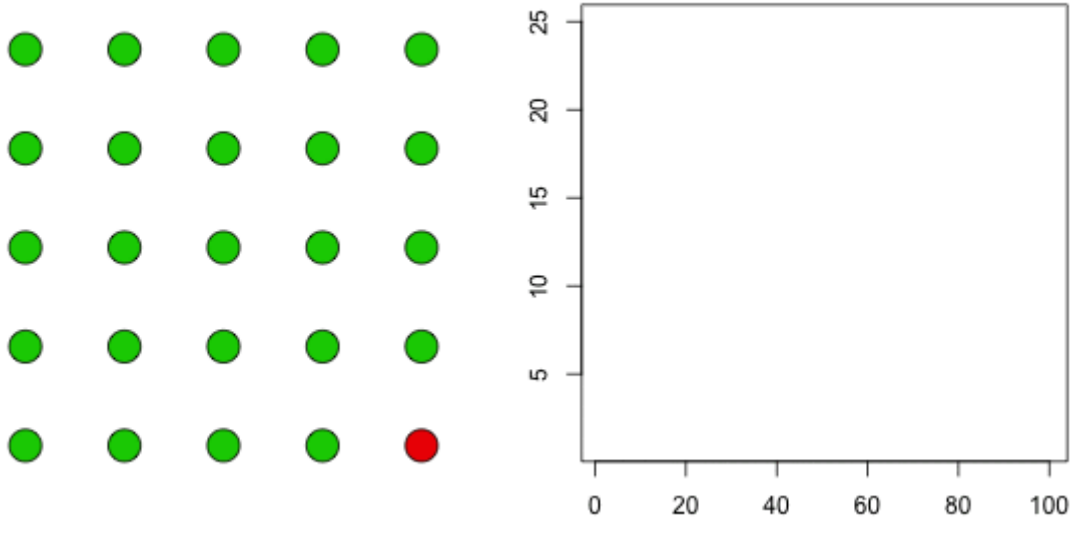
- Suppose a single infected individual enters a population.
- Let I be the number of infected people.
- Let S be the number of uninfected people susceptible to infection.
- Let β be the rate that a susceptible person who comes in contact with an infected person becomes infected.
- The rate of new infections is $\beta SI = \beta(N - I)I$.

The dumb assumptions

1. The population is "well mixed". Any one person is equally likely to come into contact with any other person.
2. The transmission rate β is the same for everybody, *forever*.
3. The size of the total population does not change.

SI dynamics

- This model is known as the susceptible-infected (SI) model.
- I increases at the fastest rate when $S = I$, *i.e.*, $I = N/2$.



Compartmental models

- The SI model belongs to a class of epidemic models known as compartmental models.
- A compartment is a sub-population with distinct numbers and rate parameters.
- These models are generally represented by systems of differential equations.
- For example, the SI model can be written as:

$$\frac{dS}{dt} = -\beta SI, \ \frac{dI}{dt} = \beta SI$$

Compartmental models

- SIS: infected individuals recover and become susceptible again.
- SIR: adds a "removed" compartment where infected individuals eventually *recover* or *die*.
- SEIR: individuals become *exposed* but not immediately infectious.
- and many others!

R_0

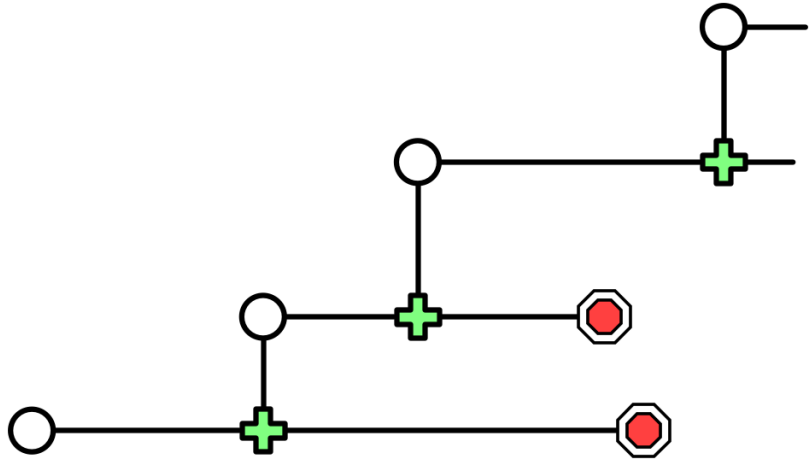
- "R naught" (basic reproduction number) is a fundamental parameter of epidemiology.
- The expected number of secondary cases from the index case.
- If $R_0 < 1$, then we expect the infection to die out. If $R_0 > 1$, then we expect it to spread.
- For the SIR model (with death rate γ),

$$R_0 = \frac{\beta N}{\gamma}$$

- Increases with transmission rate (β), and declines with death rate γ.

Birth-death models

- The transmission of a pathogen to a new host is a *lineage birth*.
- The removal of an infected individual by recovery or death is a *lineage death*.
- If we follow these lineages forward in time, we can generate a tree.



Coalescent and birth-death

- Both are models of how trees are shaped by population-level processes.

Coalescent	Birth-death
Reverse time	**Forward time**
Assumes $n << N$	**Robust to large n**
Sampling times are data	**Sampling must be modeled**
No selection	**Can handle selection naturally**
Better for a single population	**Better for pathogens?**

- Both models are *priors* - if you have enough data, it shouldn't really matter which one you use.

BEAST

- Bayesian Evolutionary Analysis by Sampling Trees
- Primarily developed by Alexei Drummond and Andrew Rambaut in the Java programming language.
- BEAST has become one of the most influential software packages in infectious disease research in the last decade.
- Over 10,000 citations, including over 40 Nature papers and over 20 Science papers since 2007.

BEAST

- Can use the coalescent or birth-death model as a prior over trees.
- Rearranges and rescales parts of the tree to perform a *random walk* over the posterior distribution of trees.
- At the same time, samples parameters of a model of evolution that is needed to reconstruct the tree from sequences.
- Outputs trace logs of posterior, likelihood, prior and model parameters — trees are written to a separate log.

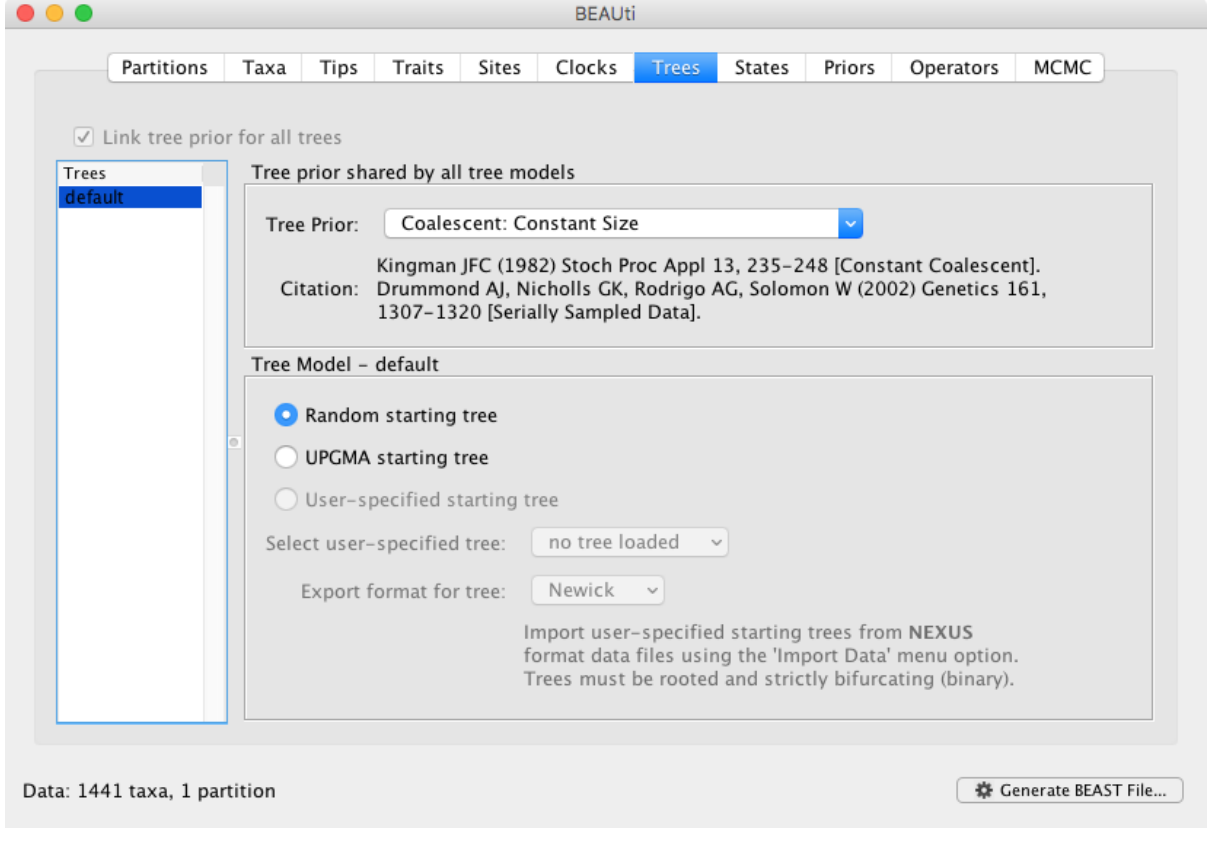
XML

- A BEAST analysis is set up in an XML file
- eXtensible Markup Language

```
<?xml version="1.0"?>
<beast>
    <alignment>
        <sequence dataType="nucleotide">
            <taxon id="Brazi82"/>
            ATGATCGTAGTATCGTAGCTCGGTTTTTACGATCGGAC
        </sequence>
        <sequence dataType="nucleotide">
            <taxon id="ElSal83"/>
            ATGATCGTAGTATCGTAGCTCGGTTTTTACGATCGGAC
        </sequence>
    </alignment>
</beast>
```

Generating XML

- Most users set up an analysis with the GUI program BEAUti (*Bayesian Evolutionary Analysis Utility*) rather than directly edit XML.



Running BEAST

- BEAST now requires BEAGLE to run.
- "Broad-platform Evolutionary Analysis General Likelihood Evaluator".
- BEAGLE is a library — a collection of functions that can be called from other programs.
- Using BEAGLE can make BEAST about 5-10 times faster!
- (It used to be possible to run BEAST *without* BEAGLE...)

Running BEAST

- BEAST can take a very long time!
- Since we are trying to explore an enormous model space, we usually run chains of 10^8 steps or more!
- BEAST tries to estimate how long it takes to perform 1 million steps.

```
# BEAST v1.10.5 Prerelease #23570d1
# Generated Sun Mar 03 13:42:57 EST 2019 [seed=1551638575176]
# benchmark2.xml
state    Posterior        Prior          Likelihood      rootHeight
0    -568698.9433    -16.9645       -568681.9788    0.20000          1.0
10000    -215056.1166    59.1339         -215115.2505    0.26252
20000    -202604.8926    72.8223         -202677.7148    0.23844
30000    -196143.5414    72.3079         -196215.8493    0.27209
40000    -193720.6293    59.7853         -193780.4145    0.39626
50000    -193206.4970    56.6362         -193263.1333    0.45645
```

Ebolavirus outbreak in West Africa

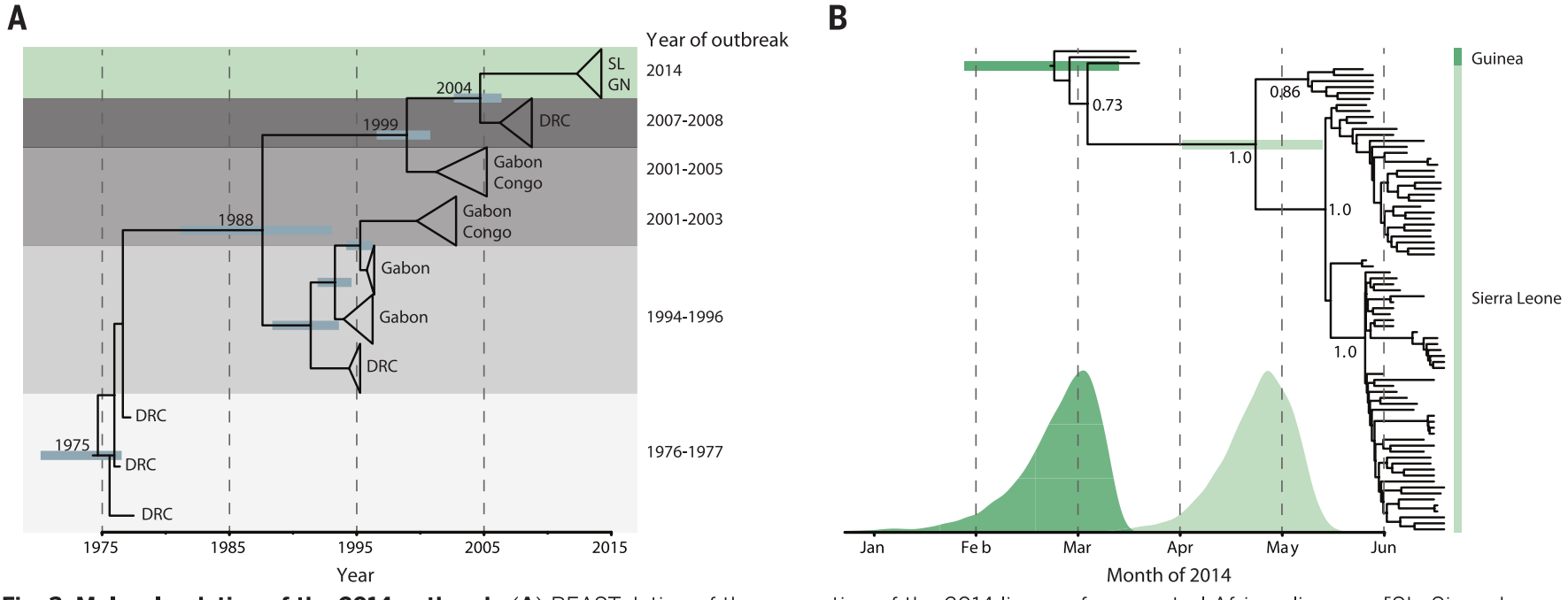
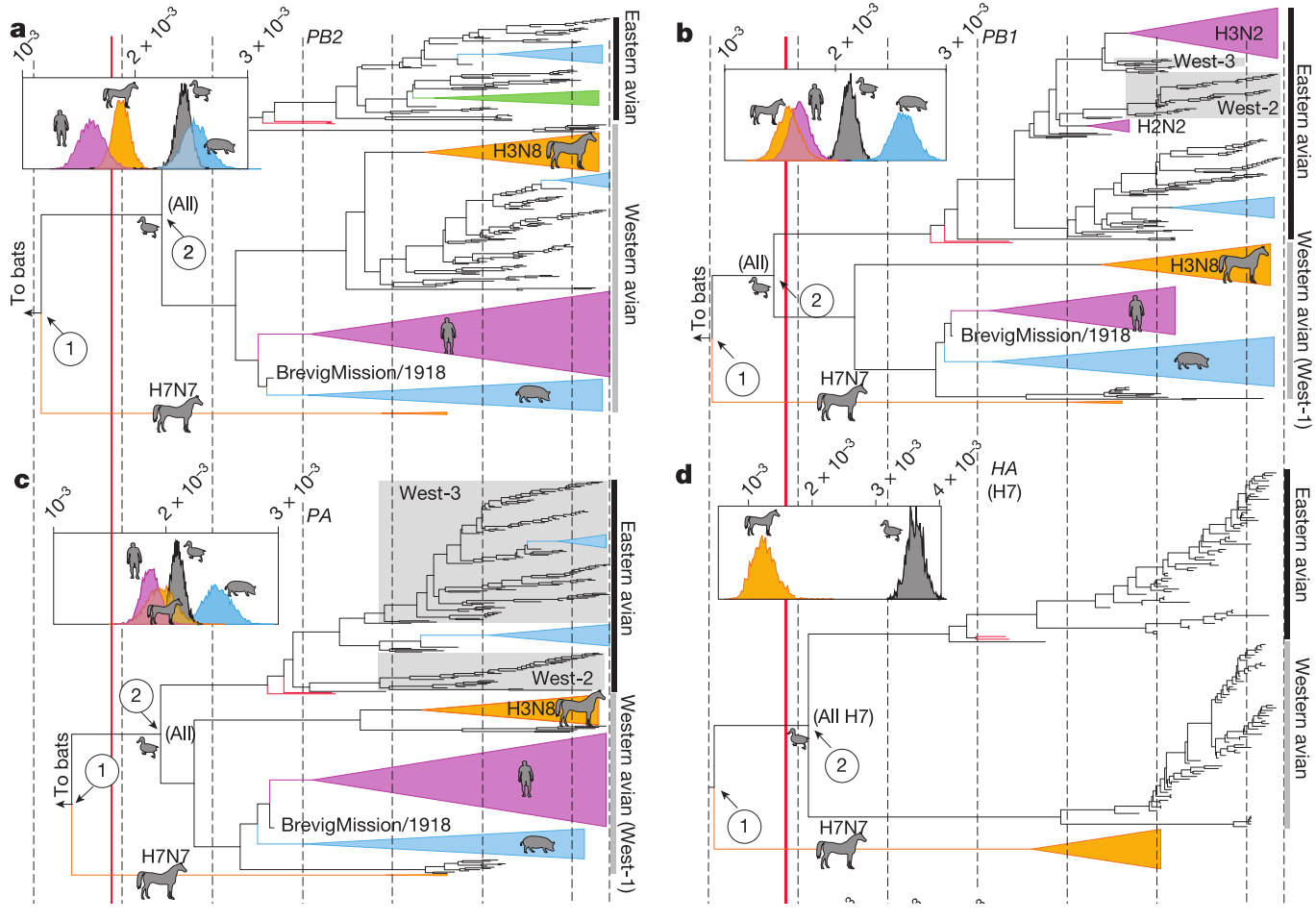


Fig. 3. Molecular dating of the 2014 outbreak. (**A**) BEAST dating of the separation of the 2014 lineage from central African lineages [SL, Sierra Leone; GN, Guinea; DRC, Democratic Republic of Congo; time of most recent common ancestor (tMRCA), September 2004; 95% highest posterior density (HPD), October 2002 to May 2006]. (**B**) BEAST dating of the tMRCA of the 2014 West African outbreak (23 February; 95% HPD, 27 January to 14 March) and the tMRCA of the Sierra Leone lineages (23 April; 95% HPD, 2 April to 13 May). Probability distributions for both 2014 divergence events are overlaid below. Posterior support for major nodes is shown.

SK Gire *et al.* 2014. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 345: 1369-1372.

Origin of modern avian influenza virus



M Worobey, G-Z Han and A Rambaut, 2014. A synchronized global sweep of the internal genes of modern avian influenza virus. Nature 508: 254-257.

Neolithic expansion of human tuberculosis

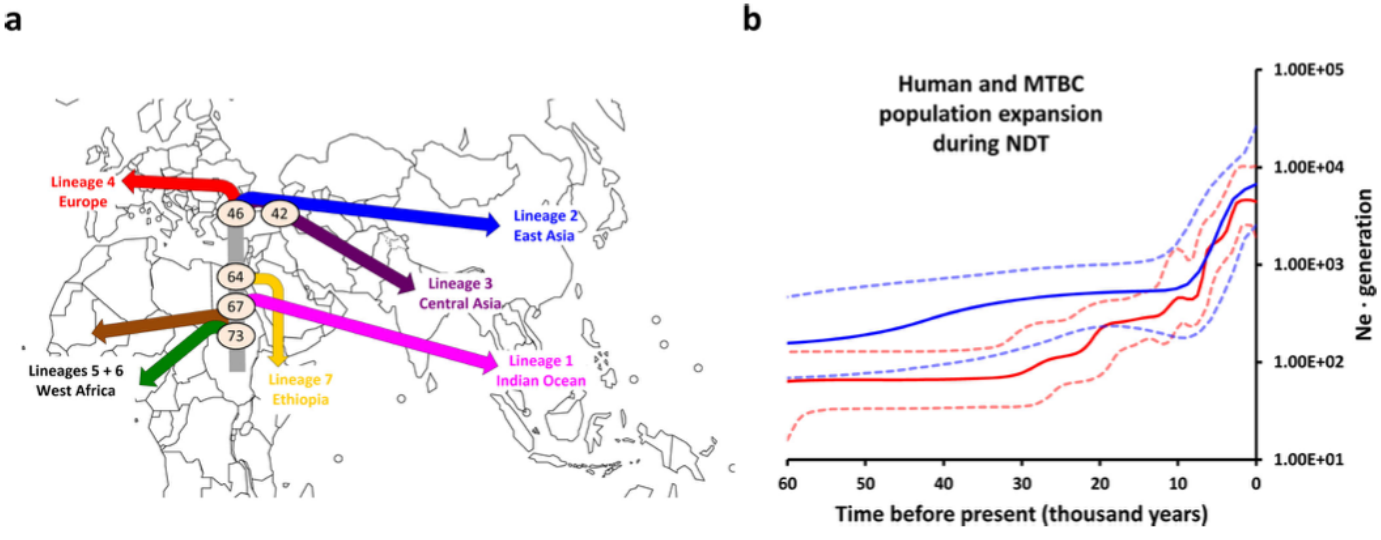


Figure 2. Out-of-Africa and Neolithic expansion of MTBC

A, Map summarizing the results of the phylogeographic and dating analyses of MTBC. The colour codes used for lineages are according to Fig. 1A. Major splits are annotated with the median value (in kya) of the dating of the relevant node. Lineage 7 (yellow) has so far been isolated exclusively from patients with known country of origin in the Horn of Africa[14]. Lineage 7 diverged subsequent to the proposed Out-of-Africa migration of MTBC; it may have arisen amongst a human population that remained in Africa, or a population that returned to Africa. **B,** Bayesian skyline plots illustrating changes in population diversity of MTBC (red line) and humans based on mitochondrial DNA (blue lines) during the last 60 ky. Dashed lines represent the 95% highest probability density (HPD) intervals for the estimated population sizes.

I Comas *et al.* 2013. Out-of-Africa migration and Neolithic co-expansion of *Mycobacterium tuberculosis* with modern humans. Nature Genetics 45: 1176-1182.