

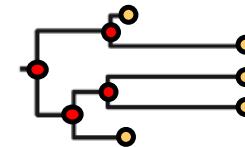
# *Infectious Disease 'Omics*



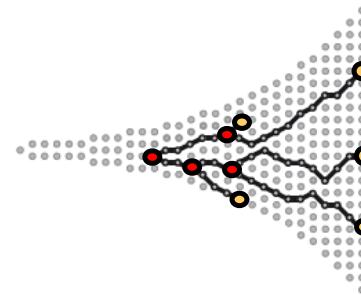
Introduction to phylodynamics and  
populations

# Lecture Overview

1) Introduction

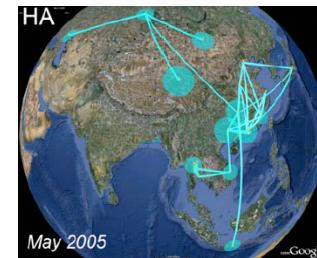


2) Molecular Dating



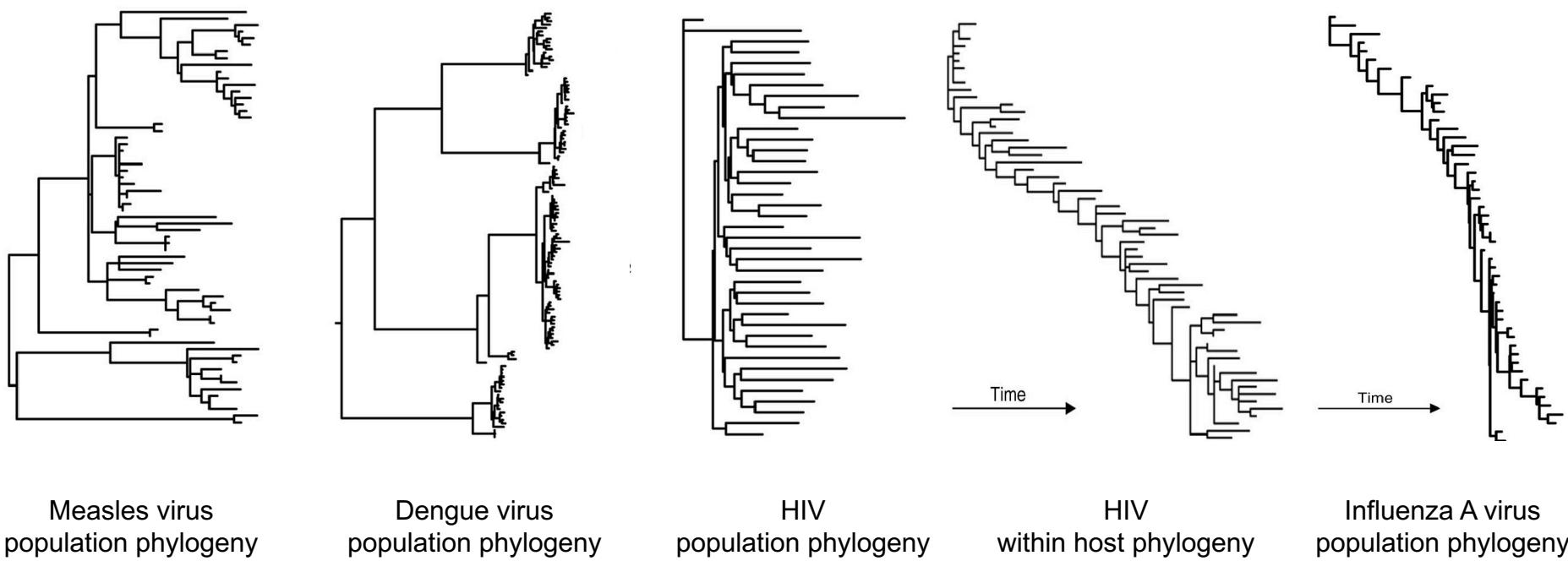
3) Population Dynamics

4) Phylogeography



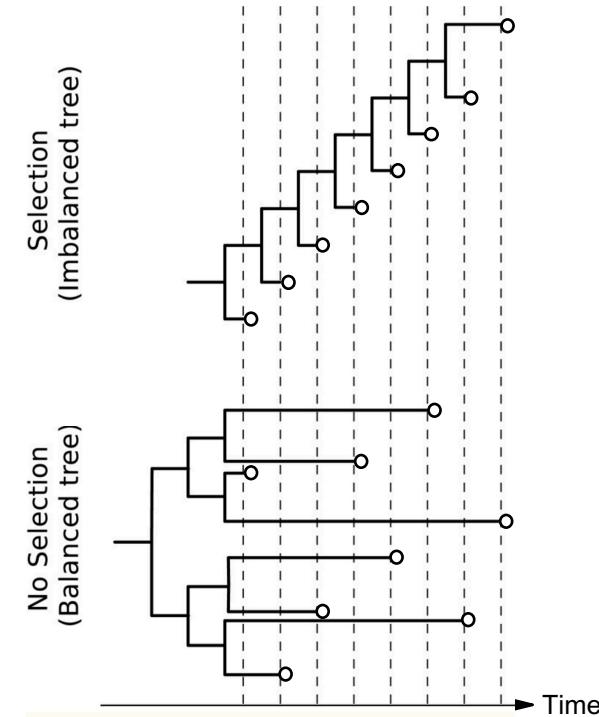
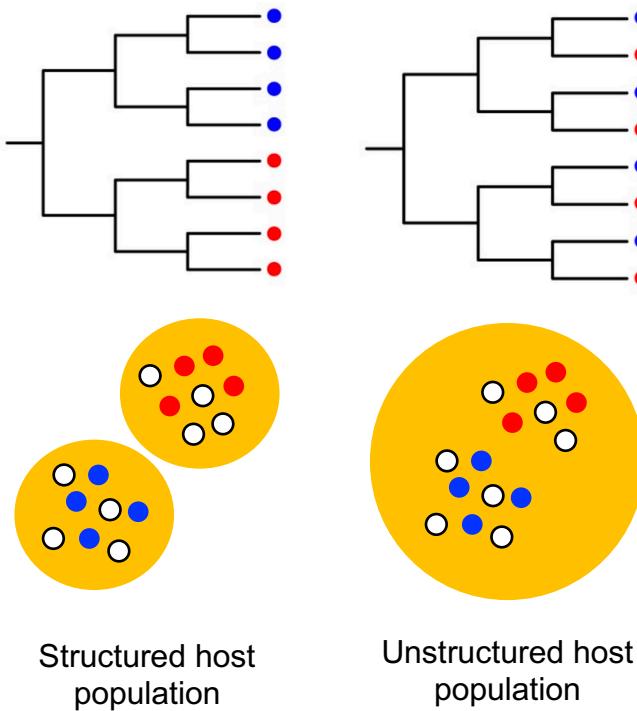
# Introduction

# Unifying epidemiology and evolution



# Unifying epidemiology and evolution

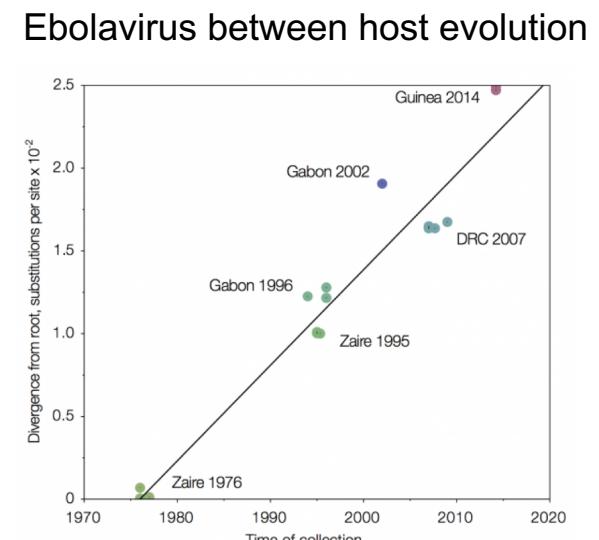
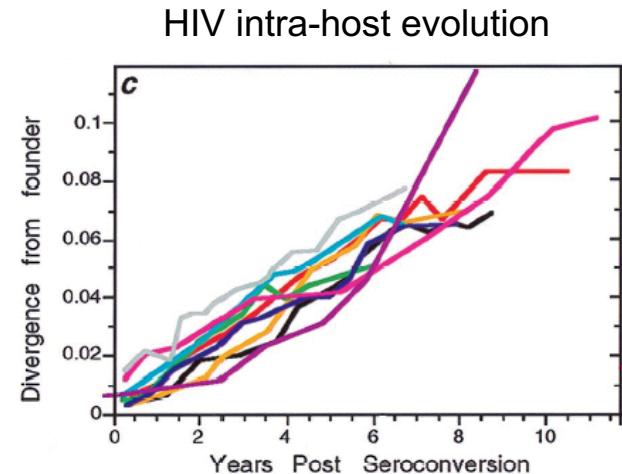
- **Phydynamics:** the study of how epidemiological, immunological and evolutionary processes shape viral phylogenies



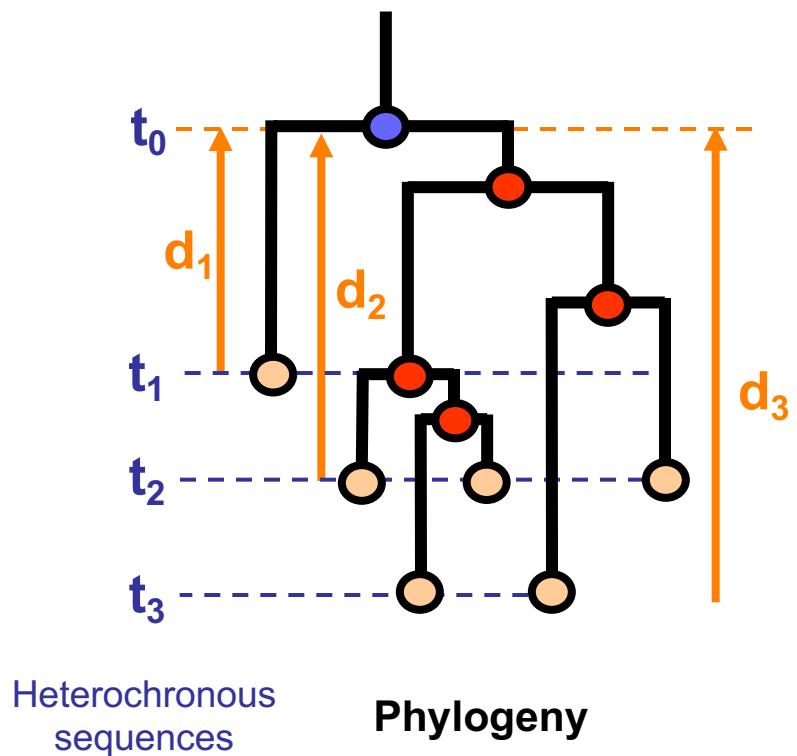
# Molecular Dating

# The Molecular Clock Hypothesis

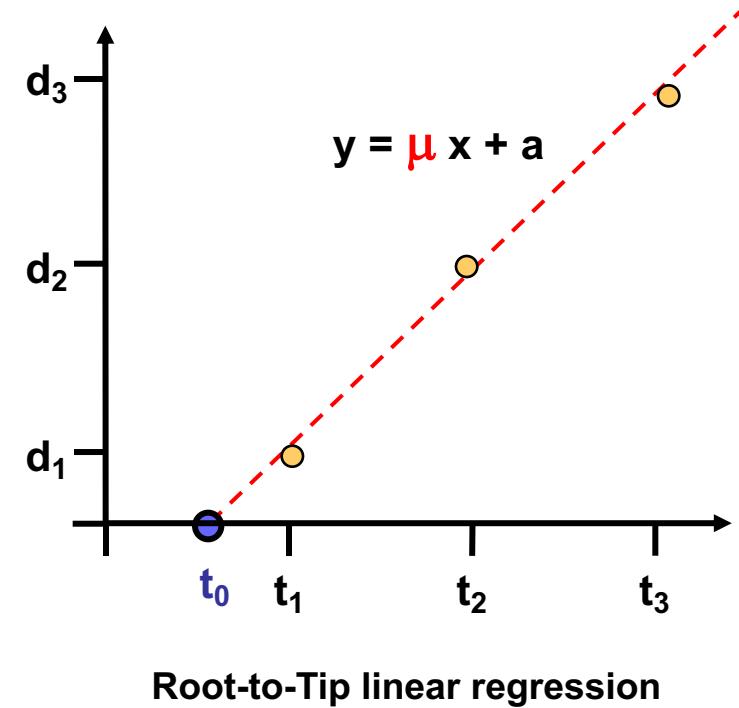
- For a given organism, gene or strain, mutations are fixed at a (nearly) **constant pace**
- Molecular Clock models define a relationship between genetic distance and time:
- Strict clock:** all lineages evolve at the **same rate**
- Relaxed clock:** Allows evolutionary rates to **change** over time and along lineages



# Estimating Substitution Rates

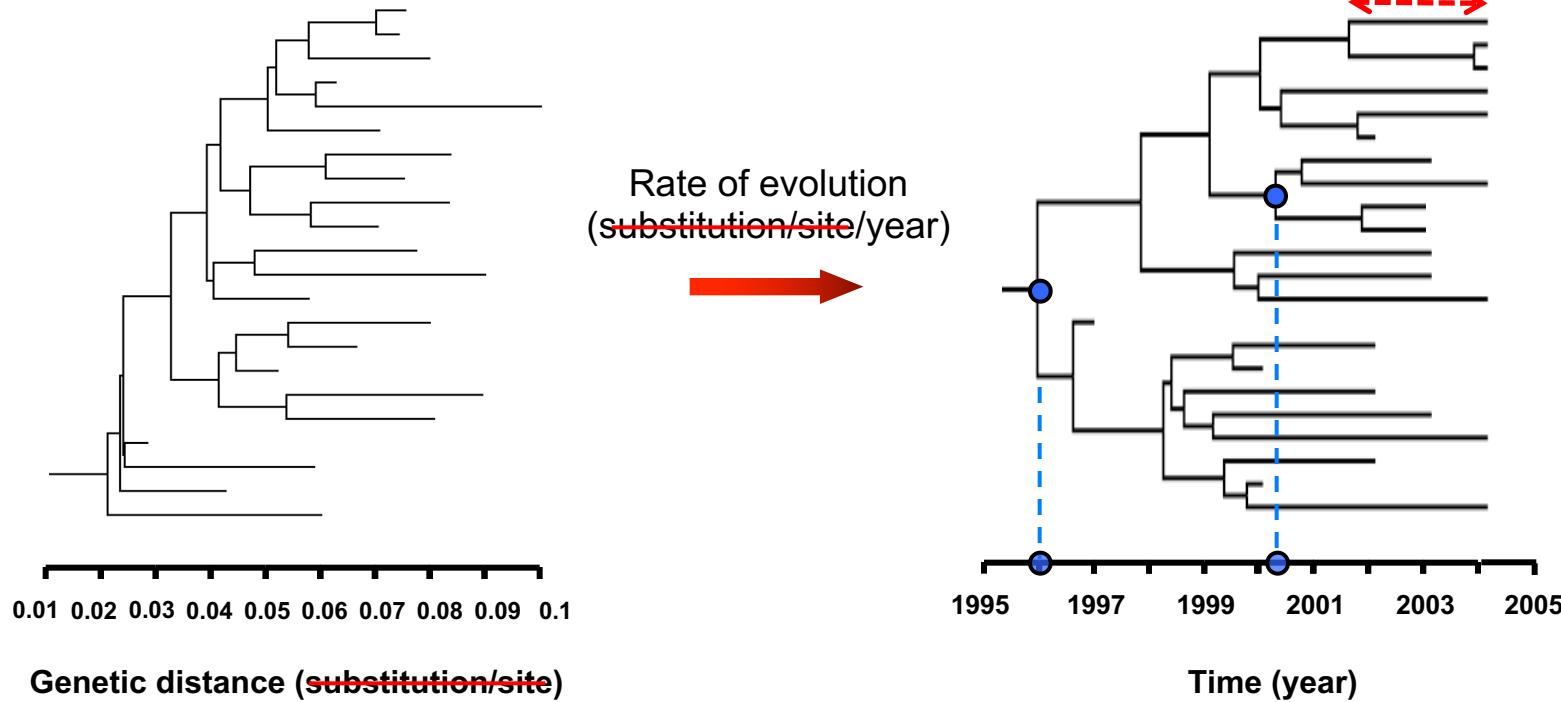


$d_i$  : genetic distance between sequence and root  
 $t_i$  : Sampling time of sequence  $i$



Evolution rate:  $\mu = \Delta d / \Delta t$   
 Time of origin:  $t_0$

# From Genetic Diversity to Time



- Date the **origin** of an epidemic (root) or outbreak (internal node)
- Estimate transmission **intervals** (internal branch lengths)

# Statistical Inference of Evolution Rates

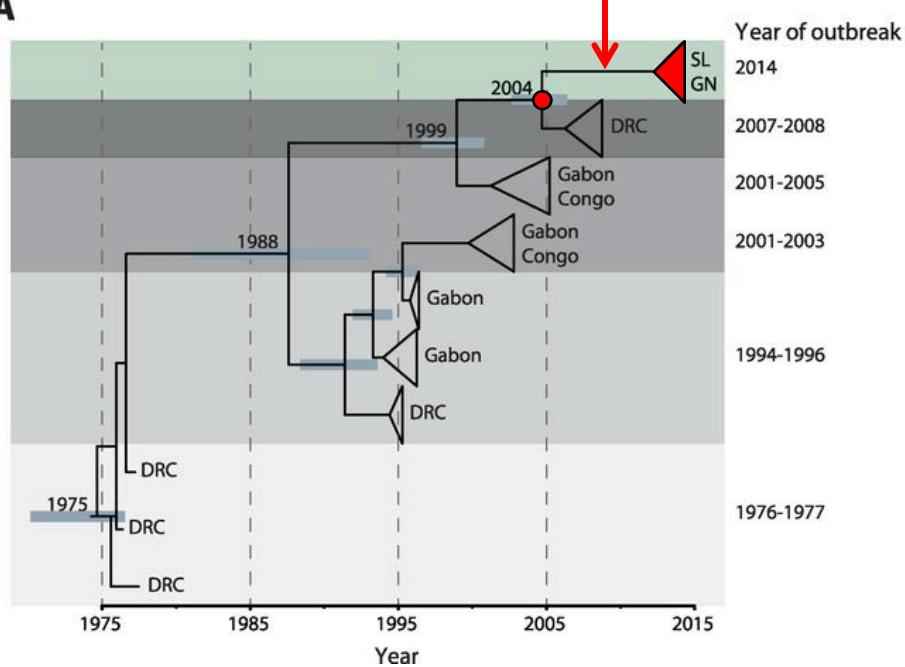
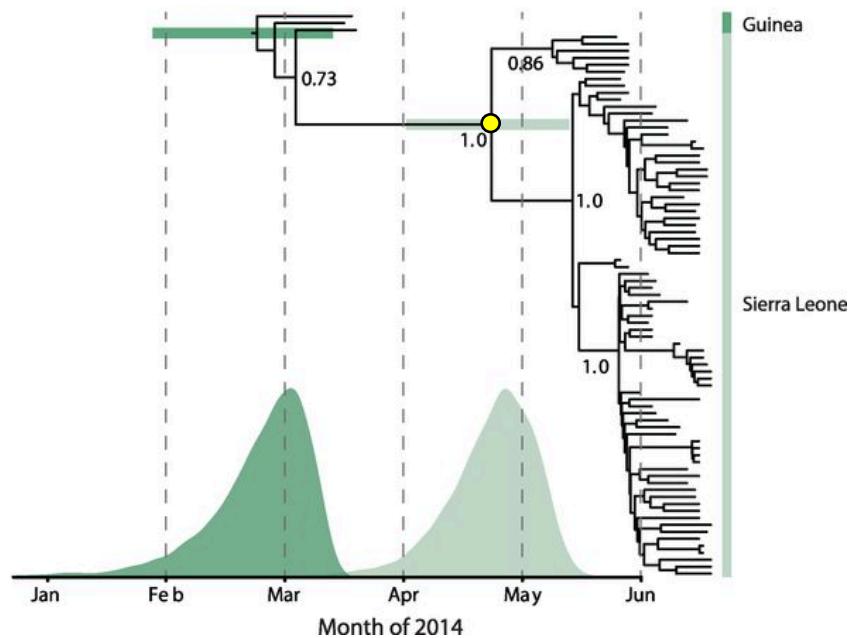
- Root-to-Tip linear regressions rely on two **unrealistic** assumptions:
  - **Constant** evolution rate over time
  - **Independence** of evolutionary histories (violated in internal branches)

⇒ Incorporation of phylogenetic structure & **statistical hypothesis testing**:

- 1) Each tip of the tree has a known time
- 2) Internal nodes are given arbitrary starting times consistent with their order in the tree
- 3) Vector of internal nodes times ( $t_0, t_1, \dots t_n$ ) and **substitution rate**  $\mu$  are optimized until the values providing the **best fit** are found

⇒ **Maximum likelihood** or **Bayesian** statistical framework

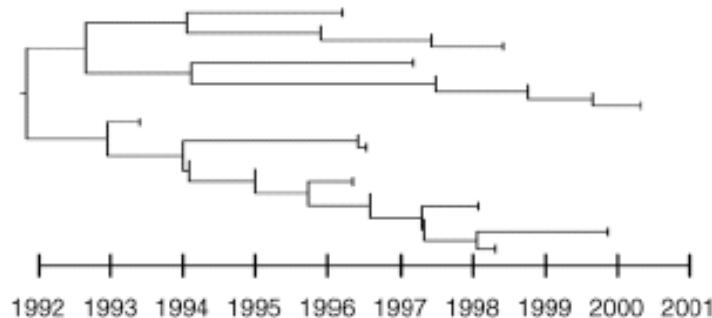
# The Guinea 2014 Ebolavirus Outbreak

**A****B**

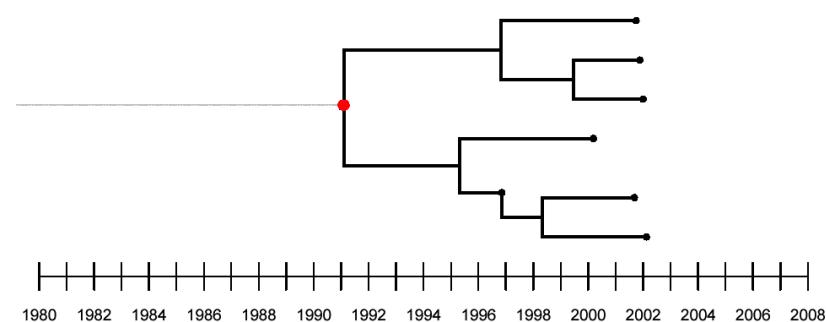
- Single introduction from the natural reservoir ( → in A)
- Divergence from Central African variant late 2004 (● in A), crossed from Guinea to Sierra Leone in May 2014 (● in B)
- No evidence of further zoonotic infections

# HIV Sexual Transmission in the UK

## Sex between men (Subtype B)



## Heterosexual contact (Non-B Subtypes)



Subtype	n Patients	Frequency of linkage (> 10 individuals)	Median transmission interval*	Frequency of transmission within 6 months post-infection
B	2 126	25%	14 months	25%
Non-B	11 071	5%	27 months	2%

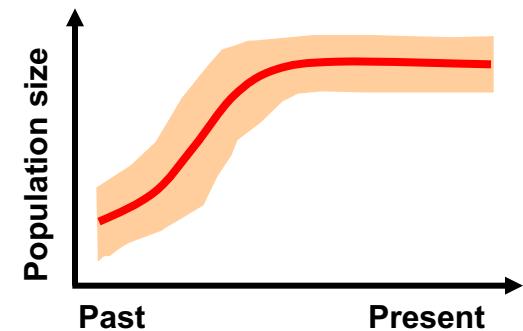
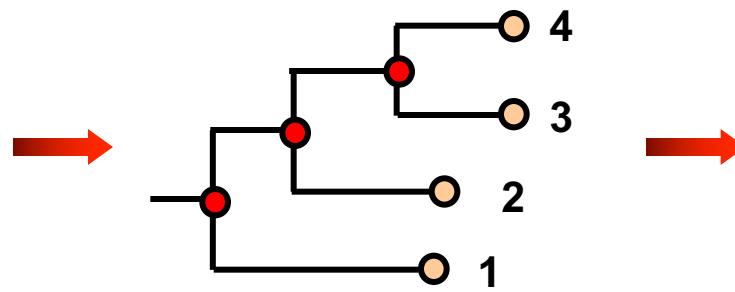
\* Median internode interval within a cluster

# Population Dynamics

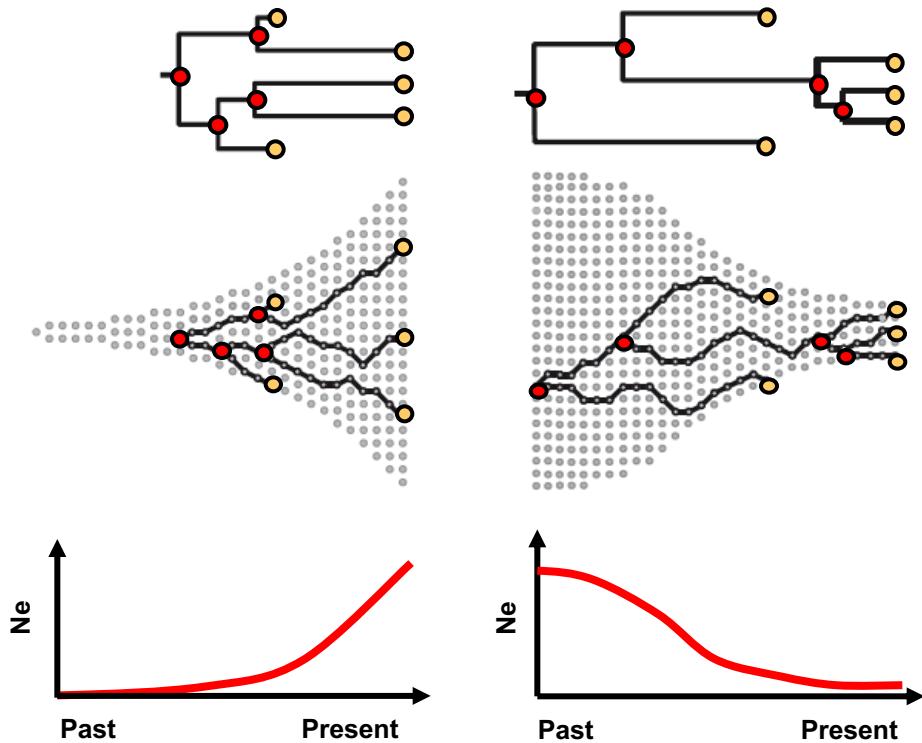
# Phylogenies & Population Dynamics

- Inference of **demographic characteristics** from phylogenies:
  - **Effective population  $N_e$**  (fraction of the total population contributing to the next generation)
  - **Growth rate** of a population
  - **Time of origin** of an epidemic

1 AAAGTTCTCAGACT...  
2 AAAGTTCTCA**C**ACT...  
3 AAT**T**TTCTCAGACT...  
4 AAT**T**TTCTCAGACT...



# The Coalescent Theory

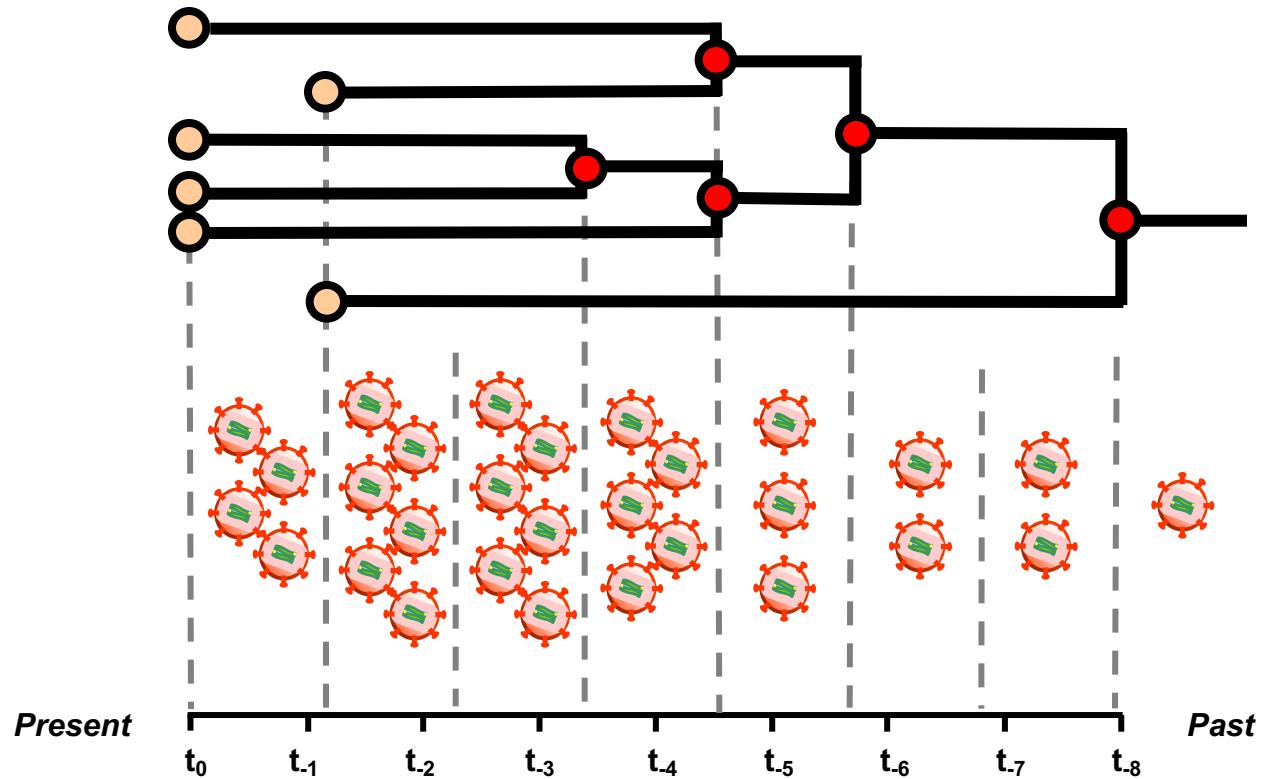


Relationship between genetic diversity and population size:

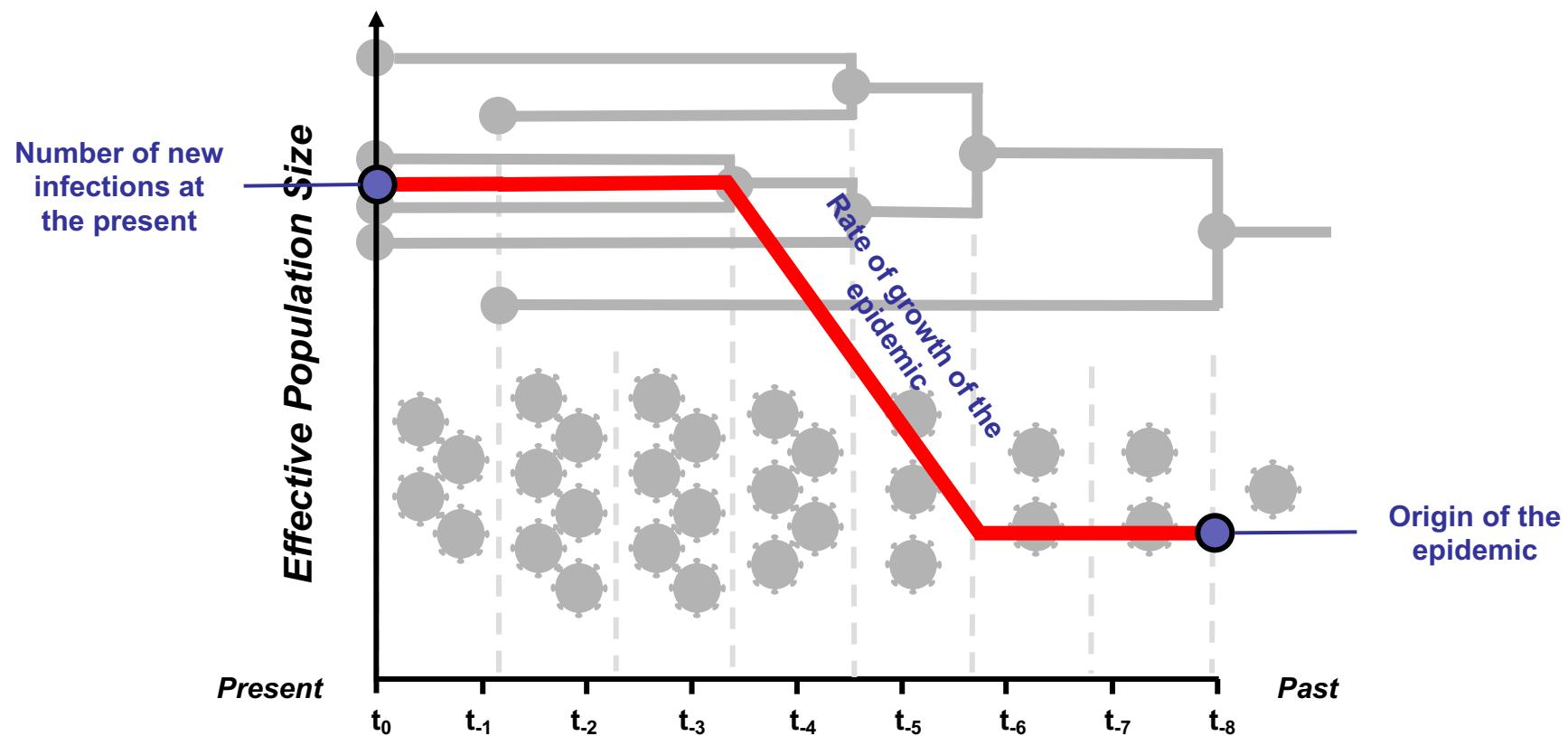
The rate at which lineages ‘coalesce’ depends on **population size** and **population structure**

**Skyline plot:** Effective population time over time

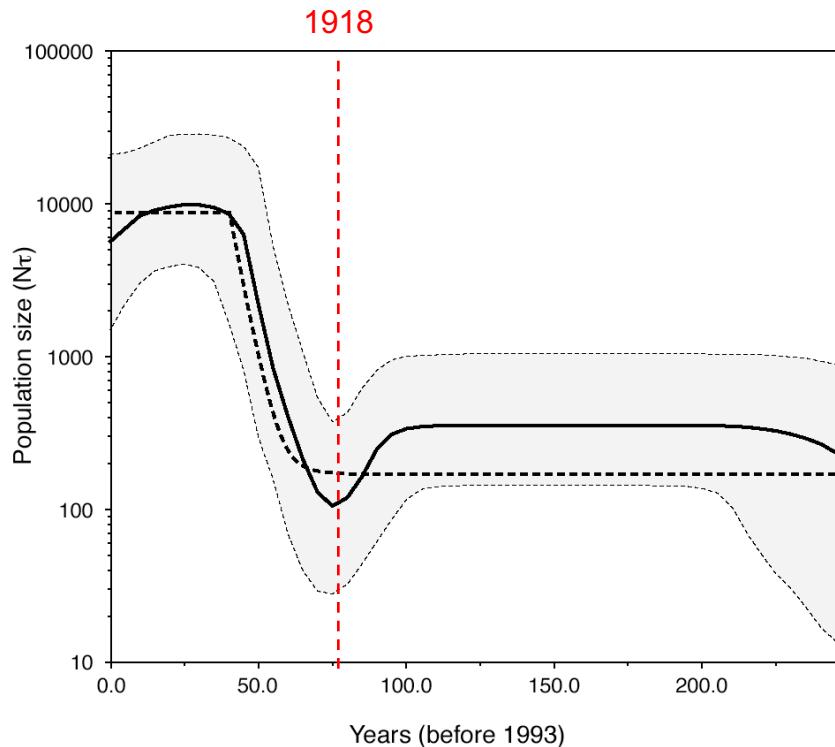
# The Coalescent Theory II



# The Coalescent Theory II

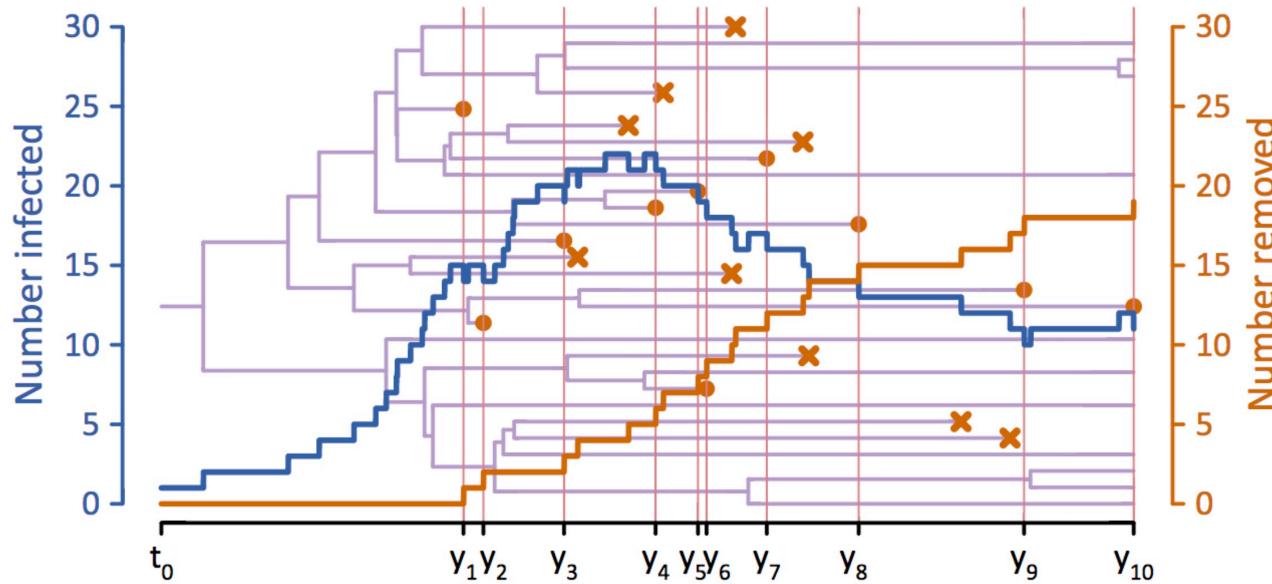


# Hepatitis C virus in Egypt



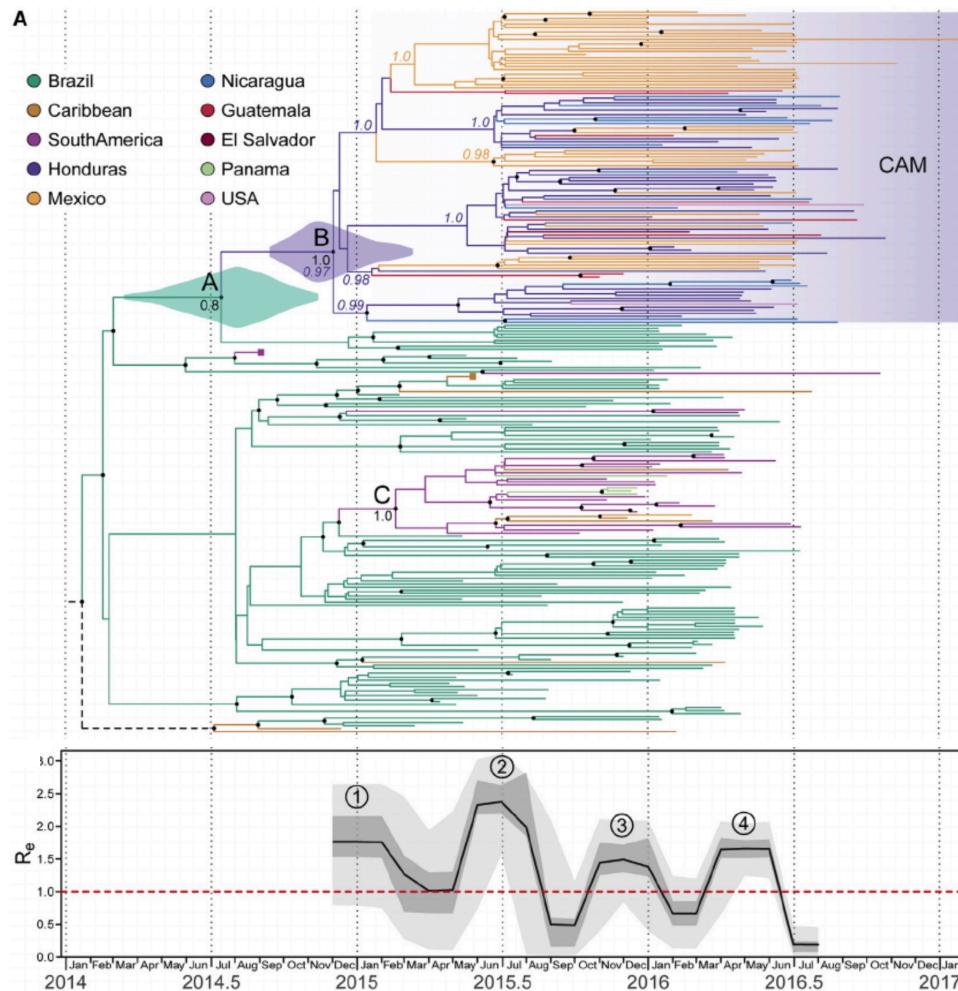
- HCV **demographic history** in Egypt estimated from 63 E1 gene sequences
- Ne dramatically **increased** in the mid 1900s
- Probably caused by **contamination** of injectable antischistosomiasis treatment

# Estimating epidemiological parameters



Number of infectious and removed (i.e., non-infectious) individuals of an epidemic estimated from a viral phylogeny

# Estimating Zika Re in Central America & Mexico



**Zika effective reproductive number (Re) through time, estimated using a birth-death skyline approach**

- (1) ZIKV spread from Honduras to other CA countries
- (2) Rapid radiation of ZIKV lineages
- (3) Period prior to rapid increase in reported ZIKV cases
- (4) Observed epidemic (April–July 2016)

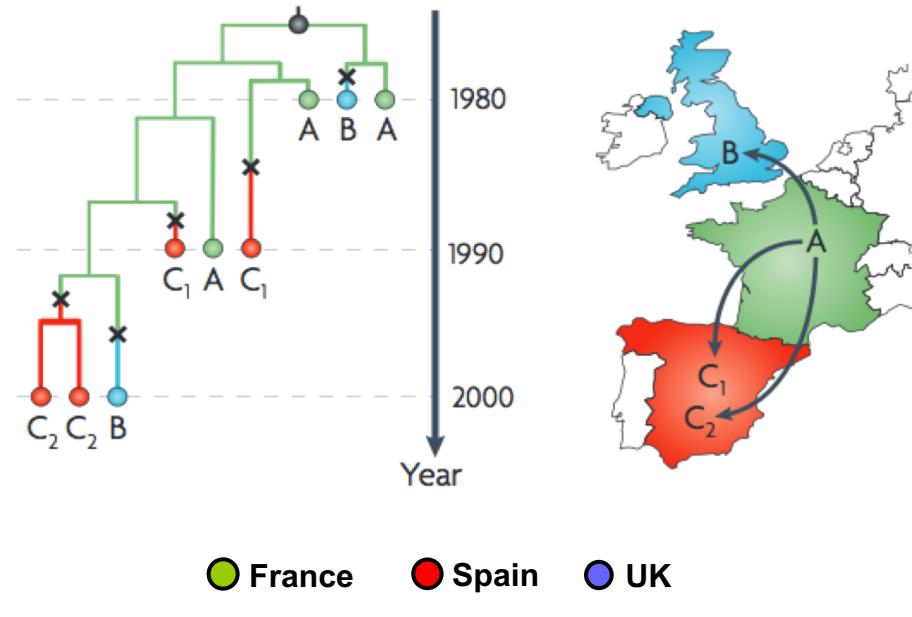
# 5 – Phylogeography

# Reconstructing Geographic Dispersal

- Statistical methods to reconstruct the history of **pathogen spread**:

Geographic location of the pathogen estimated on each branch of a phylogeny

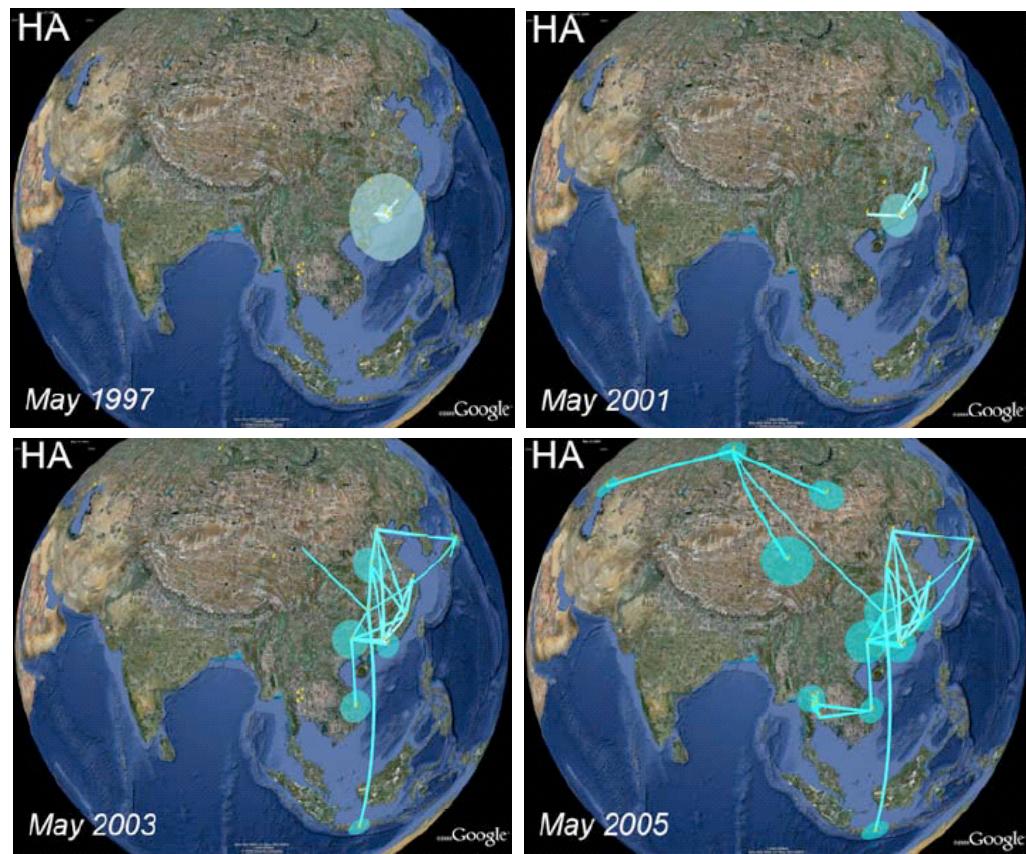
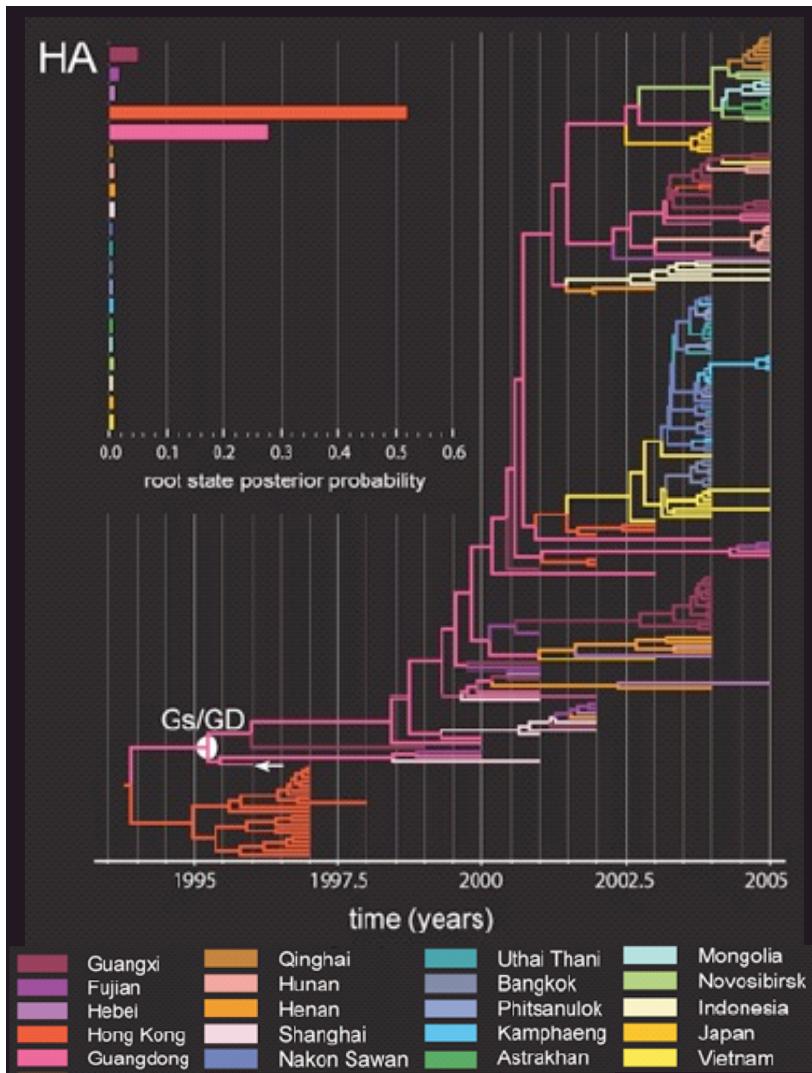
- **Parsimony approach:** Ancestral state reconstruction with minimal set of migration events consistent with the phylogeny (Ockham's razor)



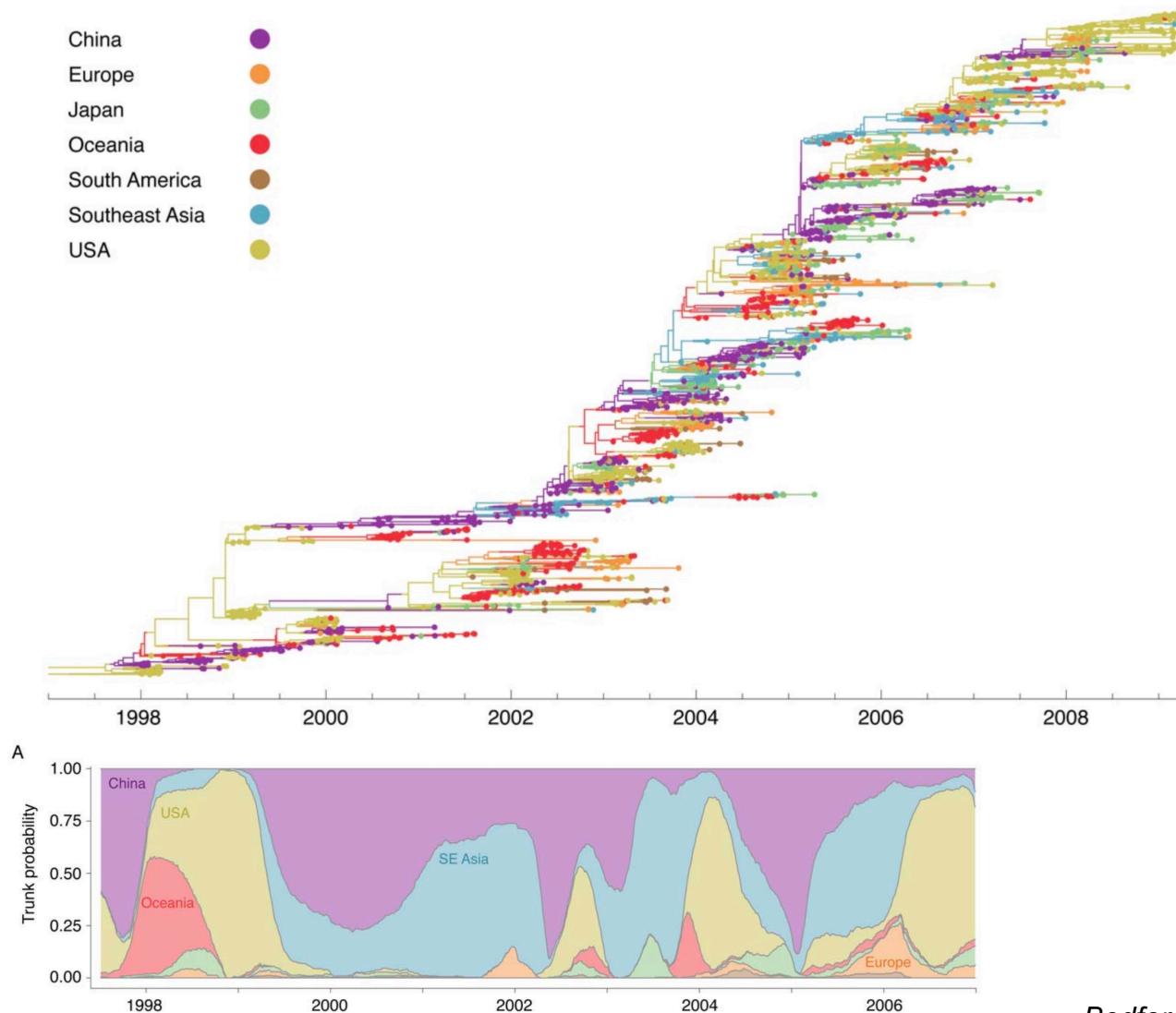
# Reconstructing Geographic Dispersal

- Parsimony dismisses phylogenetic **uncertainty** and **unequal state exchange probability**
- **Bayesian MCMC** approach:
  - **Ancestral reconstruction** of spatial states along a dated phylogeny:
    - ⇒ Spatial mapping in natural time scales
  - Calculate the **probability** of each ancestral state given the observed data
  - Calibrated under strict / relaxed **molecular clock + demographic** models

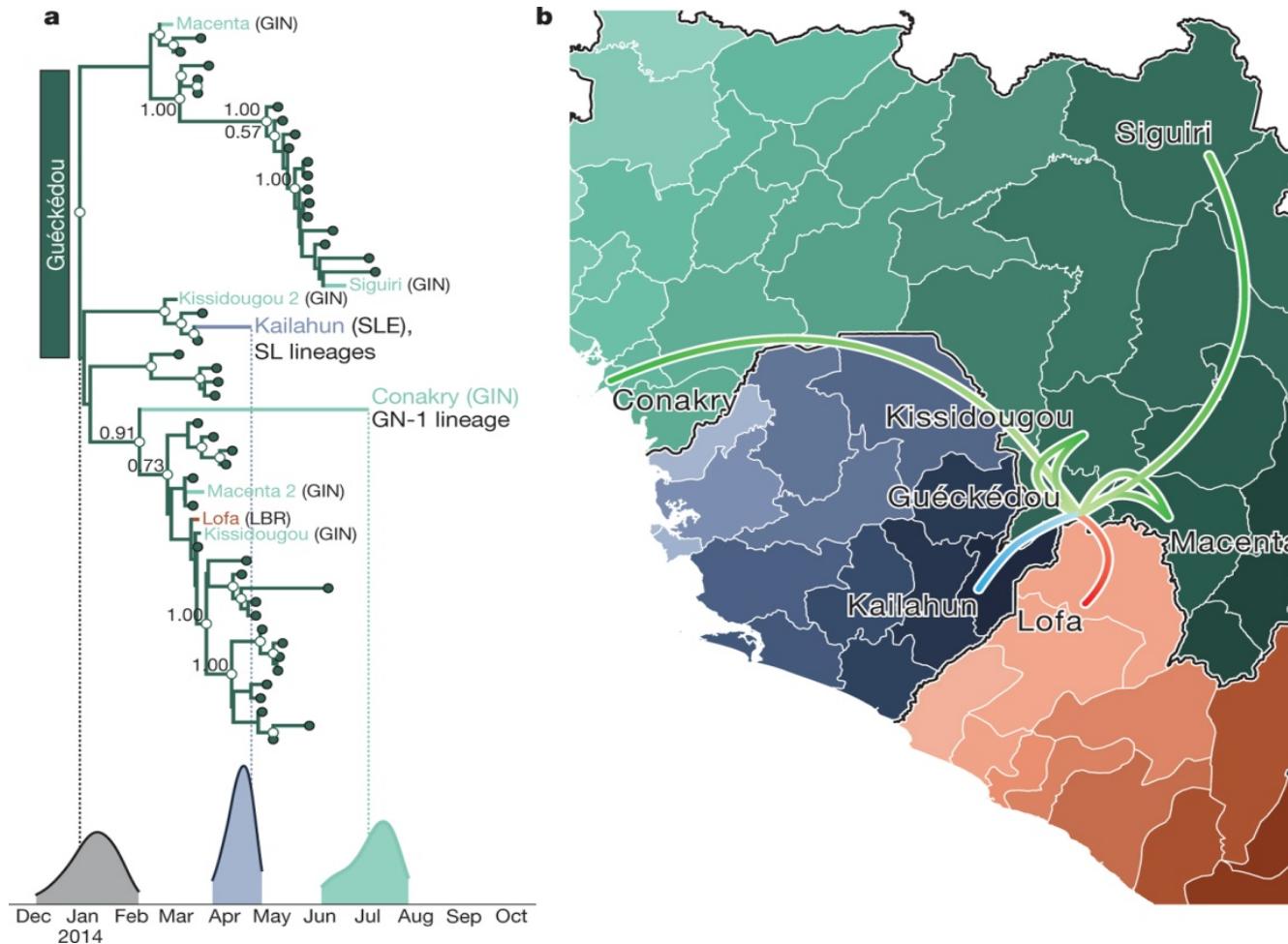
# Influenza A H5N1 Dispersal in South East Asia



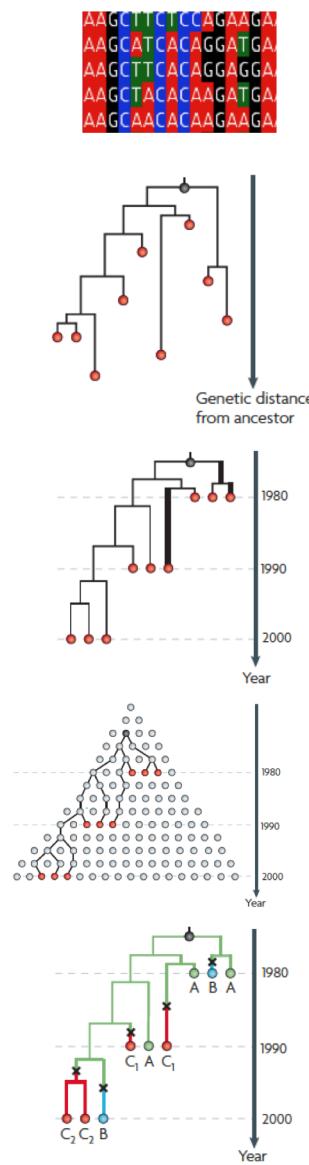
# Global migration of Human Influenza A H3N2



# Early EBOV Dispersal in the 2014 Epidemic



# Summary



Pathogen Gene Sequences

*Model of Nucleotide Substitution*

Evolutionary Tree

*Molecular Clock Model*

Dated Evolutionary Tree

*Demographic Model*

Population Dynamics

*Translocation Model*

Geographic Dispersal

# References

- *Bedford et al.* 2010. Global Migration Dynamics Underlie Evolution and Persistence of Human Influenza A (H3N2). *PLoS Pathog.* 2010 May; 6(5): e1000918.
- *Dudas et al.* 2017. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature.* 2017 Apr 20;544(7650):309-315.
- *Dudas & Rambaut.* 2014. Phylogenetic Analysis of Guinea 2014 EBOV Ebolavirus Outbreak. *PLoS Curr.* 2014 May 2;6.
- *du Plessis & Stadler.* 2015. Getting to the root of epidemic spread with phylodynamic analysis of genomic data. *Trends Microbiol.* 2015 Jul;23(7):383-6.
- *Gire et al.* 2014. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science.* 2014 Sep 12;345(6202):1369-72.
- *Grenfell et al.* 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science.* 2004 Jan 16;303(5656):327-32.
- *Hughes et al.* 2009. Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. *PLoS Pathog.* 2009 Sep;5(9):e1000590.
- *Lemey et al.* 2009. Bayesian phylogeography finds its roots. *PLoS Comput Biol.* 2009 Sep;5(9):e1000520.
- *Lewis et al.* 2008. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med.* 2008 Mar 18;5(3):e50.
- *Pybus & Rambaut.* 2009. Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet.* 2009 Aug;10(8):540-50.
- *Shankarappa.* 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol.* 1999 Dec;73(12):10489-502.
- *Thézé et al.* 2018. Genomic Epidemiology Reconstructs the Introduction and Spread of Zika Virus in Central America and Mexico. *Cell Host Microbe.* 2018 Jun 13;23(6):855-864.e7.
- *Volz et al.* 2013. Viral phylodynamics. *PLoS Comput Biol.* 2013;9(3):e1002947.

Acknowledgements: Stephane Hue