



Clinical and Epidemiological Virology,
Rega Institute, Department of Microbiology
and Immunology
KU Leuven, Belgium.



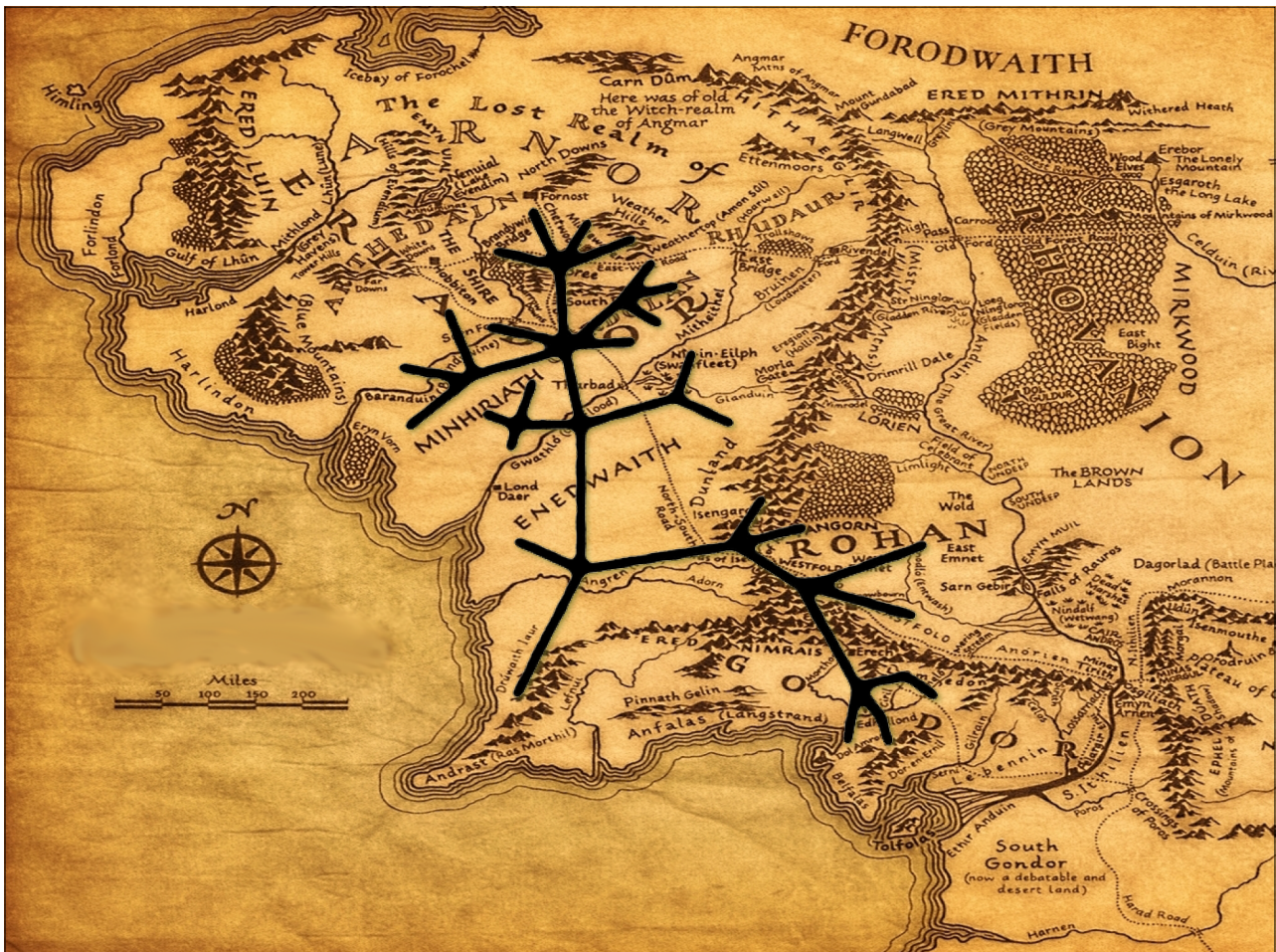
Phylogenetic diffusion models

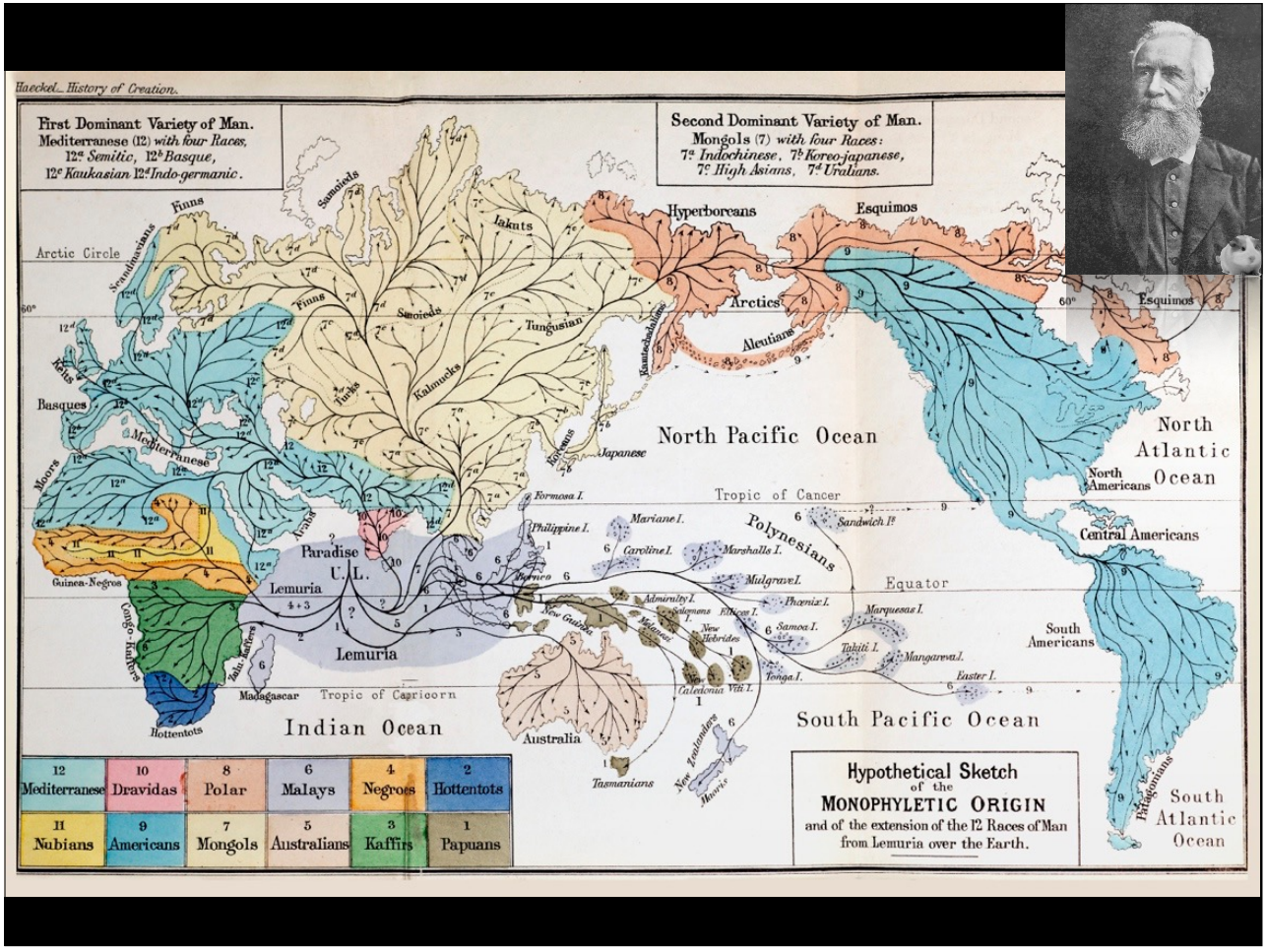
Philippe Lemey¹ and Marc Suchard²

1. Rega Institute, Department of Microbiology and Immunology, K.U. Leuven, Belgium.

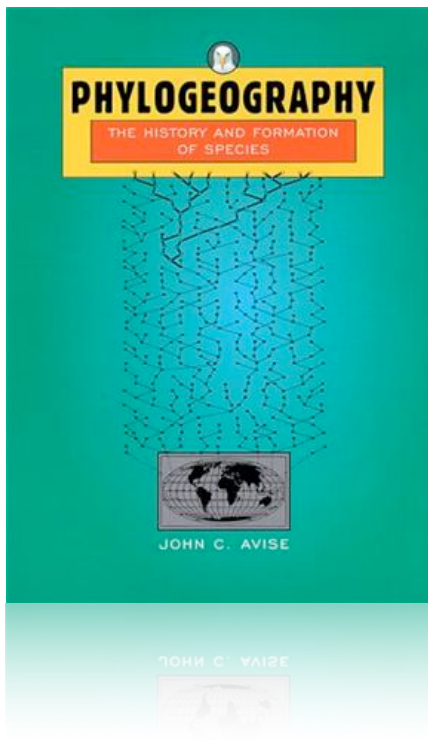
2. Departments of Biomathematics and Human Genetics, David Geffen School of Medicine at UCLA. Department of Biostatistics, UCLA School of Public Health

SISMID, July 19-21, 2017



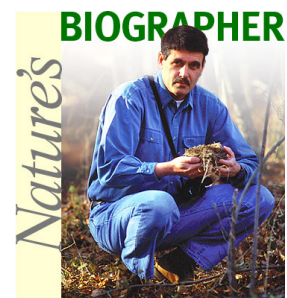


Phylogeography



“a field of study concerned with the principles and processes governing the geographic distribution of genealogical lineages, especially those within and among closely related species.”

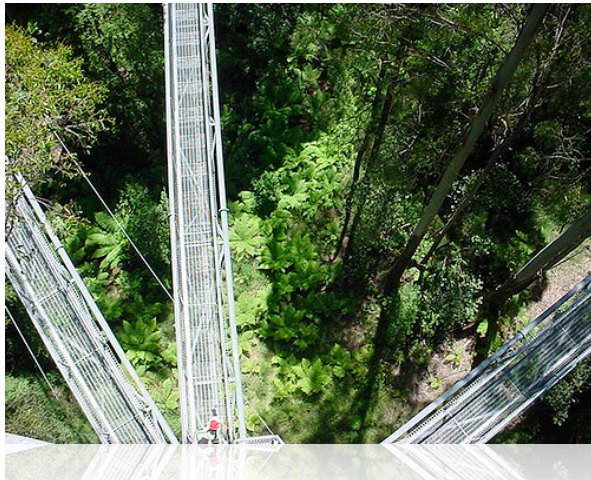
Avise, 2000



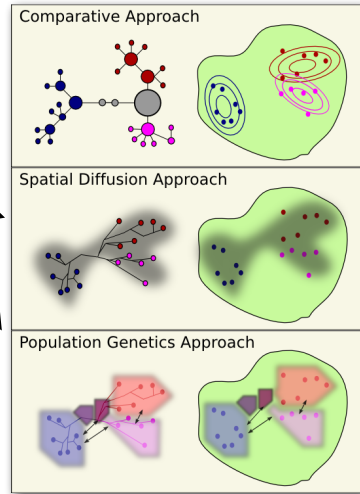
Phylogeography: three roads diverged?

Ancestral reconstruction

Population genetics

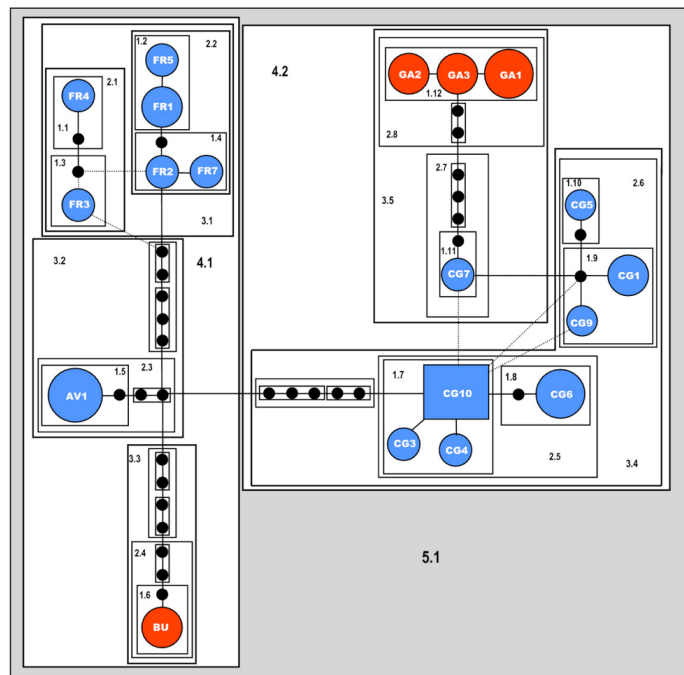
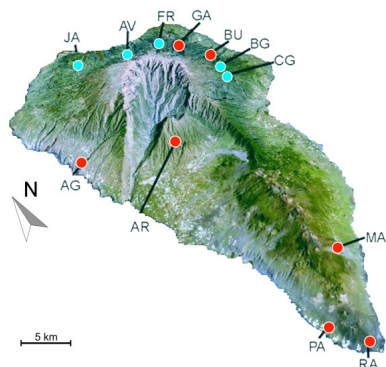


NCPA



Phylogeographic inference (road I)

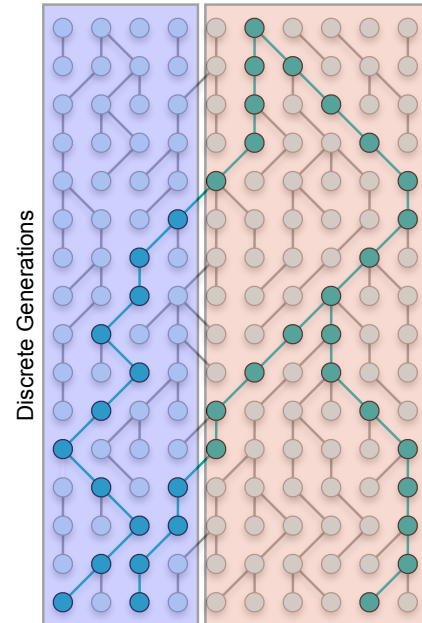
Nested clade phylogeographic analysis (NCPA)



Phylogeographic inference (road II)

Coalescent theory:

- is a statistical framework for the analysis of genetic polymorphism data
- is an extension of classical population-genetics theory and models
- one can estimate time (number of generations) for lineages to coalesce
- many applications (including migration analysis)



Structured coalescent (road II)

LAMARC - Likelihood Analysis with Metropolis Algorithm using Random Coalescence



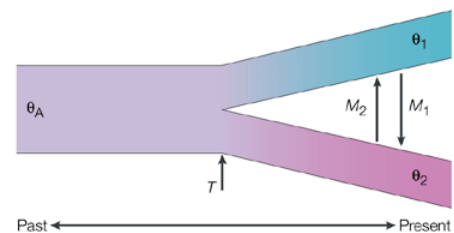
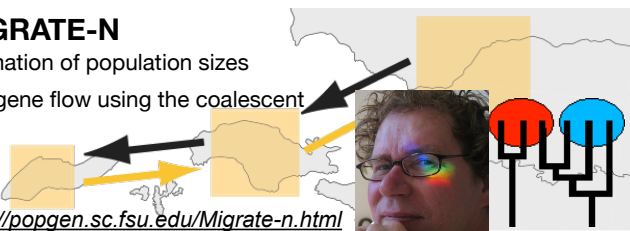
<http://evolution.genetics.washington.edu/lamarc.html>



MIGRATE-N

estimation of population sizes and gene flow using the coalescent

<http://popgen.sc.fsu.edu/Migrate-n.html>



<http://genfaculty.rutgers.edu/hey/software>

MDIV: http://www.biom.cornell.edu/Homepages/Rasmus_Nielsen/files.htm

Batwing: <http://www.maths.abdn.ac.uk/~ijw/downloads/download.htm>

BEAST2: <http://compevol.github.io/MultiTypeTree> / BASTA

Structured coalescent (road II)

New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation

Nicola De Maio, Chieh-Hsi Wu, Kathleen M O'Reilly, Daniel Wilson 

Published: August 12, 2015 • <https://doi.org/10.1371/journal.pgen.1005421>

[Mol Biol Evol.](#) 2016 Aug; 33(8): 2102–2116.

Published online 2016 Apr 9. doi: [10.1093/molbev/msw064](https://doi.org/10.1093/molbev/msw064)

PMCID: PMC4948704

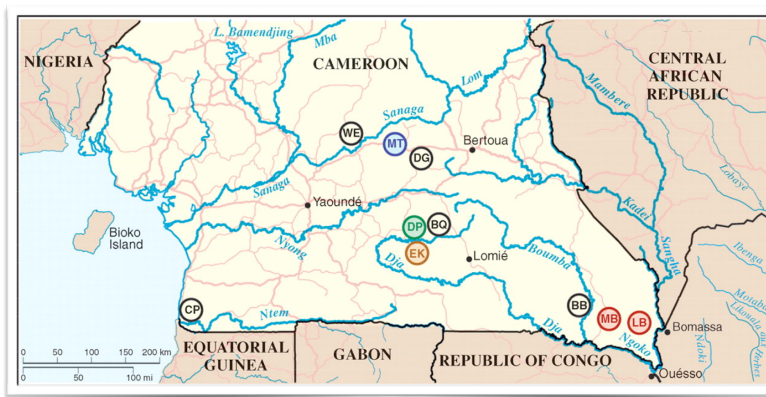
Phylodynamics with Migration: A Computational Framework to Quantify Population Structure from Genomic Data

[Denise Kühnert](#)^{1,1,2,3,4} [Tanja Stadler](#)^{3,4} [Timothy G. Vaughan](#)² and [Alexei J. Drummond](#)²

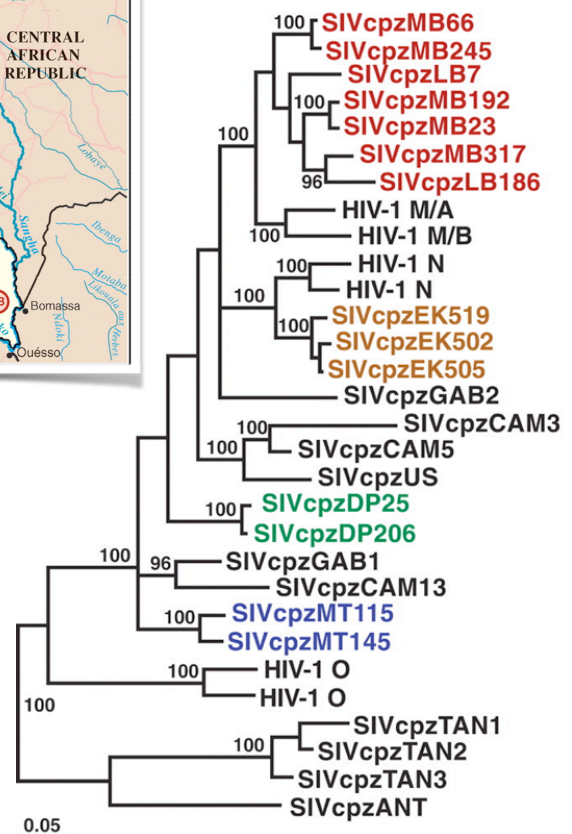
The Structured Coalescent and its Approximations

 Nicola Felix Mueller, David Alan Rasmussen, Tanja Stadler

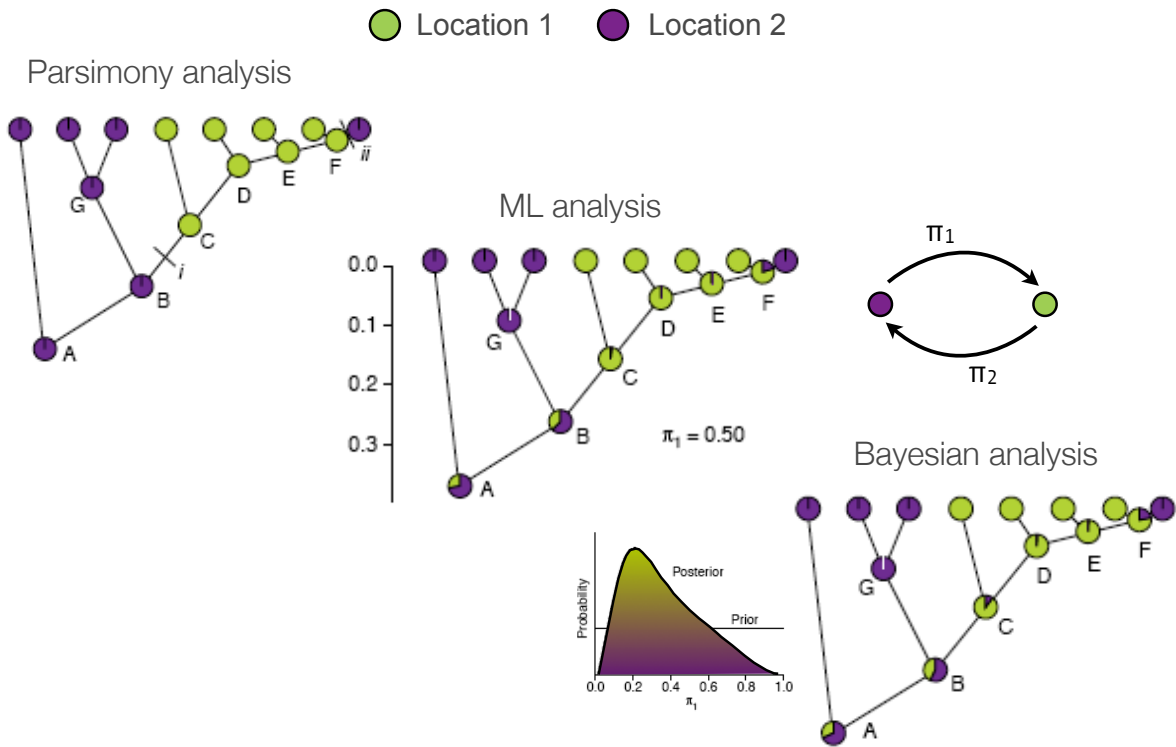
doi: <https://doi.org/10.1101/091058>



Keele et al., 2006, Science



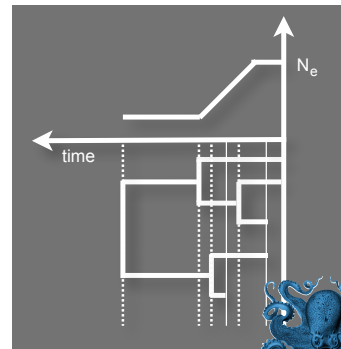
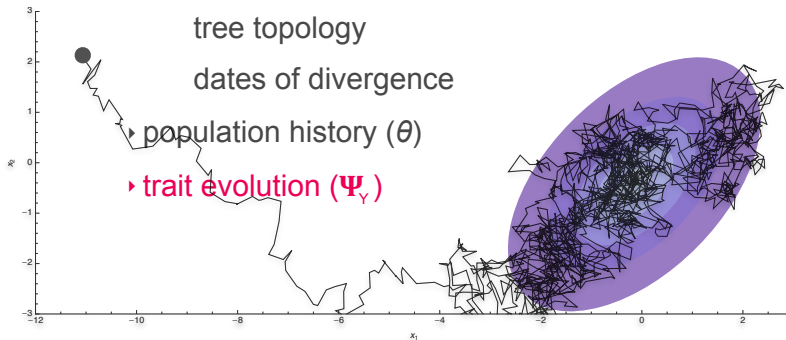
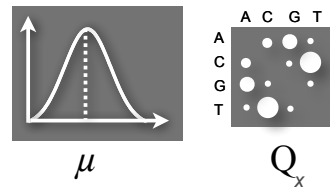
Inferring discrete ancestral state locations



Bayesian Evolutionary Analysis Sampling Trees

- Given sequence *and trait* data (X, Y) that is temporally spaced estimate true values of:

- substitution parameters (μ and Q_x)
- ancestral genealogy ($g = E_g, t_x$)



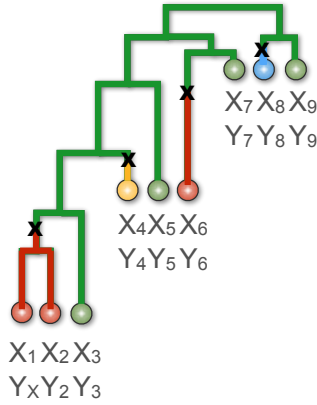
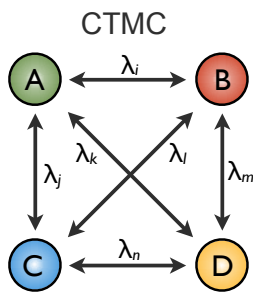
- Bayesian inference

$$P(g, \mu, \theta, Q_x, \Psi_Y | X, Y) = \frac{1}{Z} \Pr\{X|g, \mu, Q_x\} \Pr\{Y|g, \Psi_Y\} f_g(g|\theta) f_\mu(\mu) f_\theta(\theta) f_x(Q_x) f_\Psi(\Psi_Y)$$

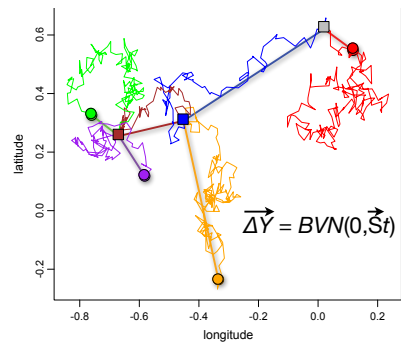


Phylogenetic diffusion models

discrete



continuous



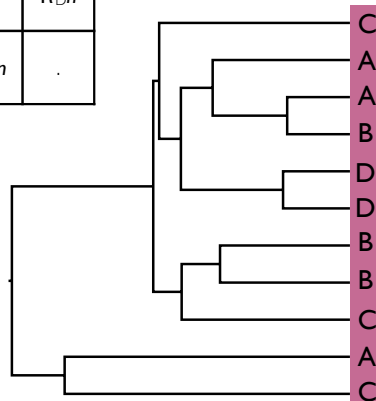
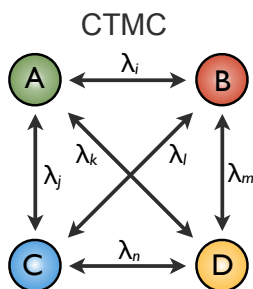
Lemey et al., PLoS Comp Bio, 2009

Lemey et al., MBE, 2010

Phylogenetic diffusion models

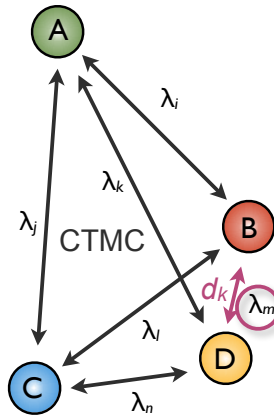
● Discrete Model:

| | A | B | C | D |
|---|-------------|-------------|-------------|-------------|
| A | . | $\pi_{B A}$ | $\pi_{C A}$ | $\pi_{D A}$ |
| B | $\pi_{A B}$ | . | $\pi_{C B}$ | $\pi_{D B}$ |
| C | $\pi_{A C}$ | $\pi_{B C}$ | . | $\pi_{D C}$ |
| D | $\pi_{A D}$ | $\pi_{B D}$ | $\pi_{C D}$ | . |



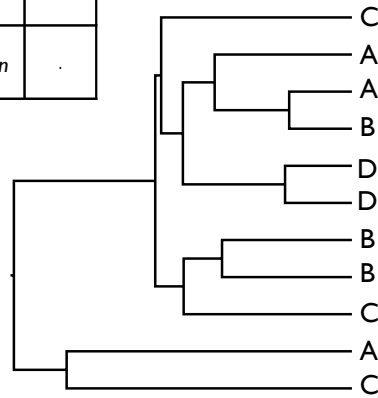
Phylogenetic diffusion models

• priors



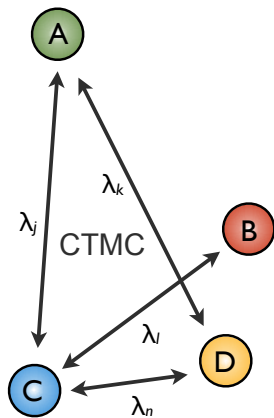
gamma(μ, σ)
 $\mu \sim 1/d_k$
 $\sigma \sim 1/d_k$

| | A | B | C | D |
|---|------------|------------|------------|------------|
| A | . | π_{Bj} | π_{Cj} | π_{Dk} |
| B | π_{Ai} | . | π_{Cl} | π_{Dm} |
| C | π_{Aj} | π_{Bl} | . | π_{Dn} |
| D | π_{Ak} | π_{Bm} | π_{Cn} | . |

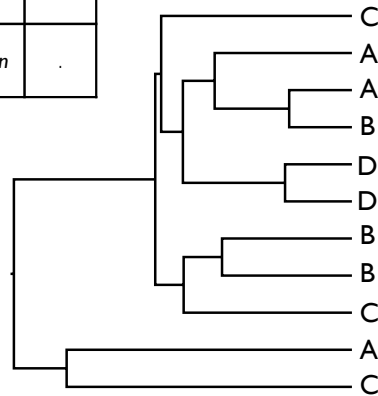


Phylogenetic diffusion models

• Do we need all those parameters?

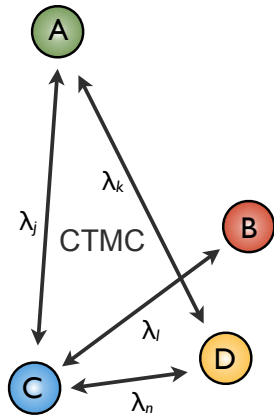


| | A | B | C | D |
|---|----------------------------------|----------------------------------|------------|----------------------------------|
| A | . | π_{Bj} | π_{Cj} | π_{Dk} |
| B | π_{Ai} | . | π_{Cl} | π_{Dm} |
| C | π_{Aj} | π_{Bl} | . | π_{Dn} |
| D | π_{Ak} | π_{Bm} | π_{Cn} | . |



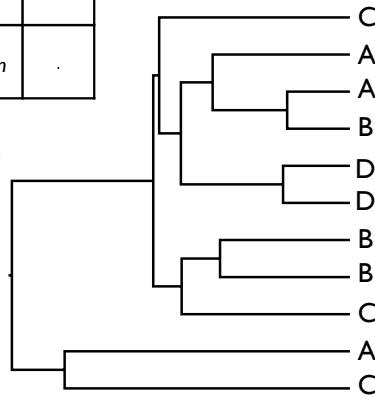
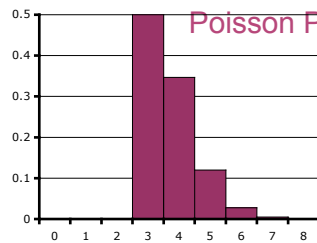
Phylogenetic diffusion models

- Do we need all those parameters?



| | A | B | C | D |
|---|-------------|-------------|-------------|-------------|
| A | . | $\pi_{B A}$ | $\pi_{C A}$ | $\pi_{D A}$ |
| B | $\pi_{A B}$ | . | $\pi_{C B}$ | $\pi_{D B}$ |
| C | $\pi_{A C}$ | $\pi_{B C}$ | . | $\pi_{D C}$ |
| D | $\pi_{A D}$ | $\pi_{B D}$ | $\pi_{C D}$ | . |

Rate Indicators $I_{[0,1]}$



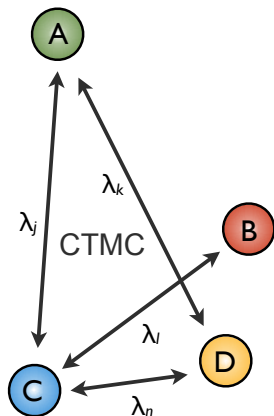
offset = $K - 1$
 mean = $\log(2)$

$$\sum_j I_j$$

Phylogenetic diffusion models

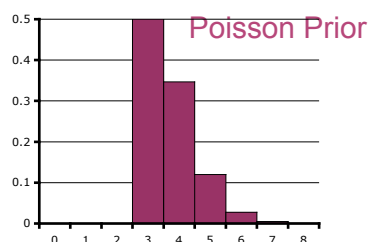
- Bayesian stochastic search variable selection procedure

→ support for a particular rate (connection)?



$$\text{Bayes factor} = \frac{\text{posterior odds}}{\text{prior odds}}$$

$$\frac{\text{Pr}(I=1|D) / (1 - \text{Pr}(I=1|D))}{\text{Pr}(I=1) / (1 - \text{Pr}(I=1))}$$



$$\frac{\text{Poisson offset} + \text{mean}}{K(K-2)/2}$$

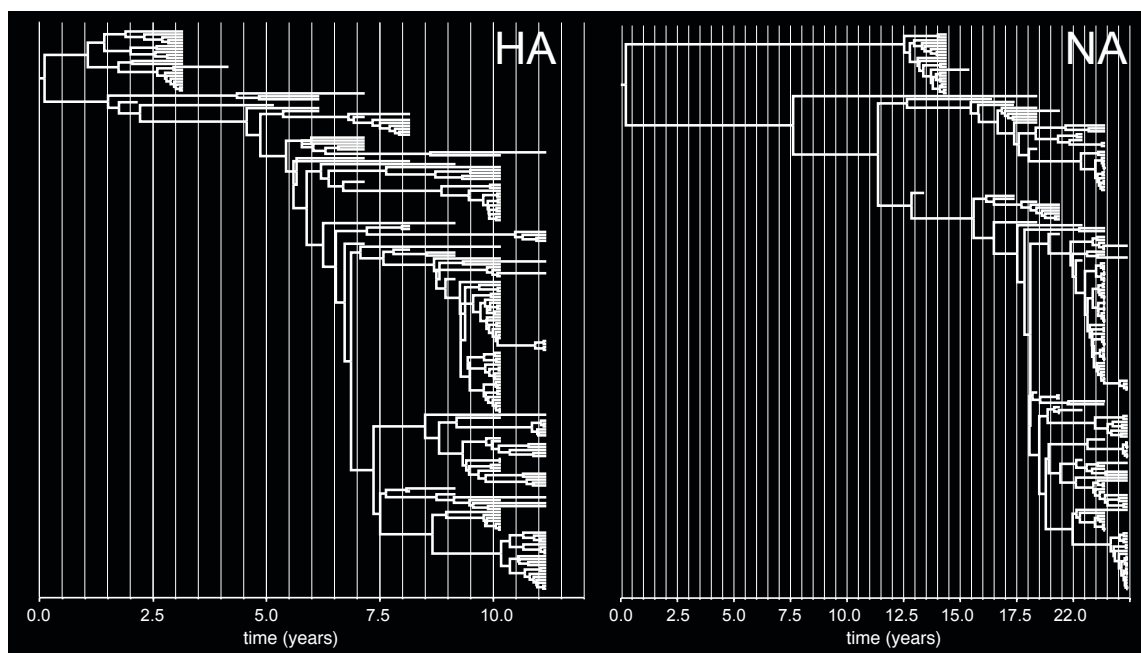
offset = $K - 1$
 mean = $\log(2)$

H5N1 'bird flu'

- Wild fowl act as natural asymptomatic carriers
- first HPAI outbreak in Guangdong, China in 1996
- 'Bird flu' outbreak in Hong Kong in 1997
- the A/goose/Guangdong/1/96 (Gs/GD) virus lineage has become the longest recorded HPAI virus to remain endemic in poultry

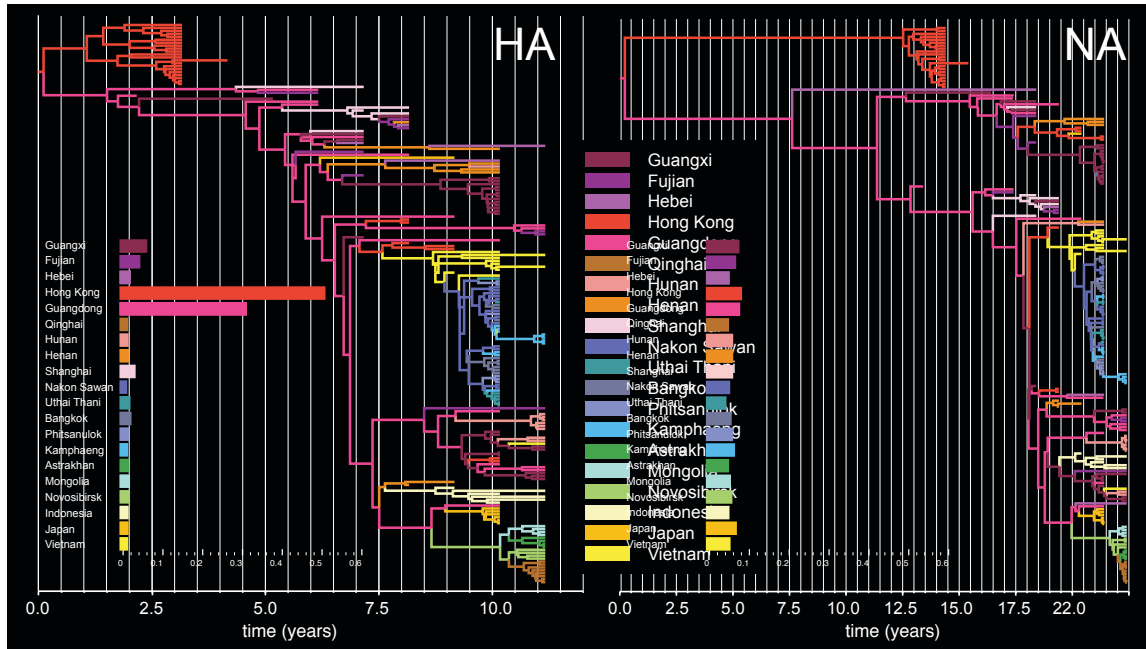


Influenza A H5N1

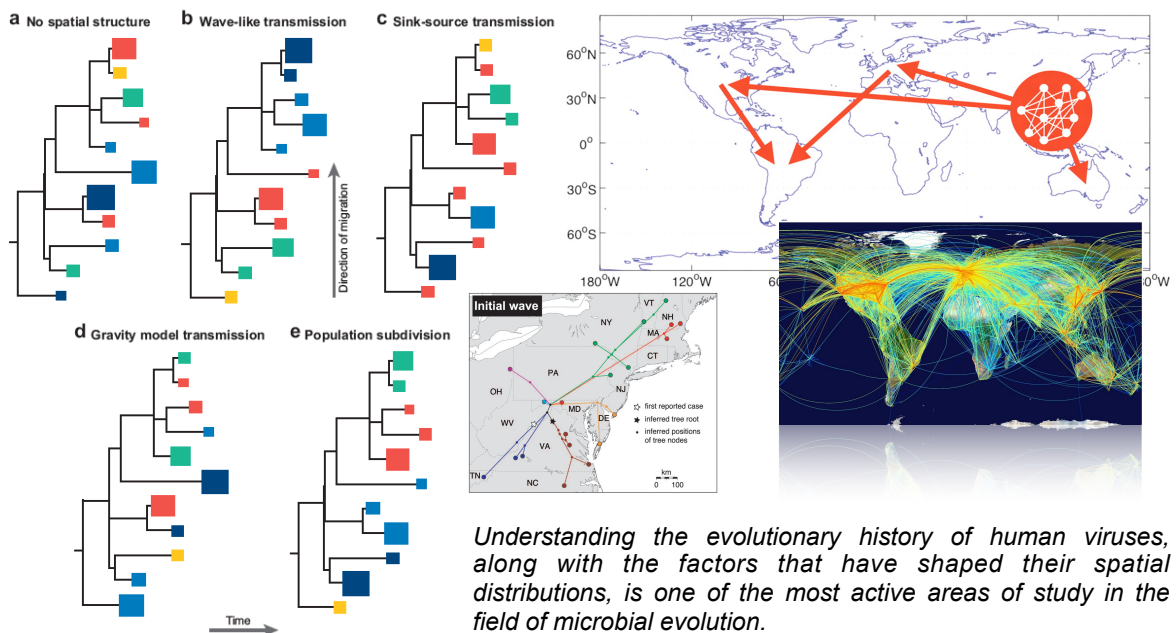


Wallace et al., PNAS, 2007

Influenza A H5N1: discrete model



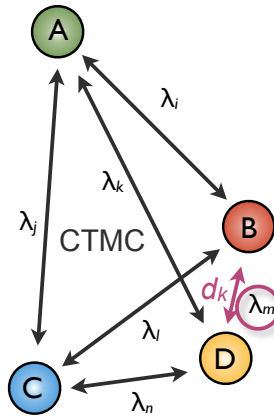
Phylogeographic patterns in RNA viruses



Eddie Holmes, *Ann Rev Microbiol* 2008

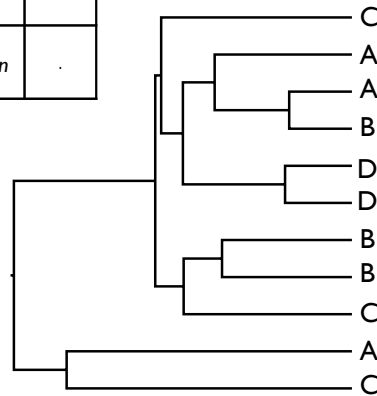
Phylogeographic hypothesis testing?

• priors



gamma(μ, σ)
 $\mu \sim 1/d_k$
 $\sigma \sim 1/d_k$

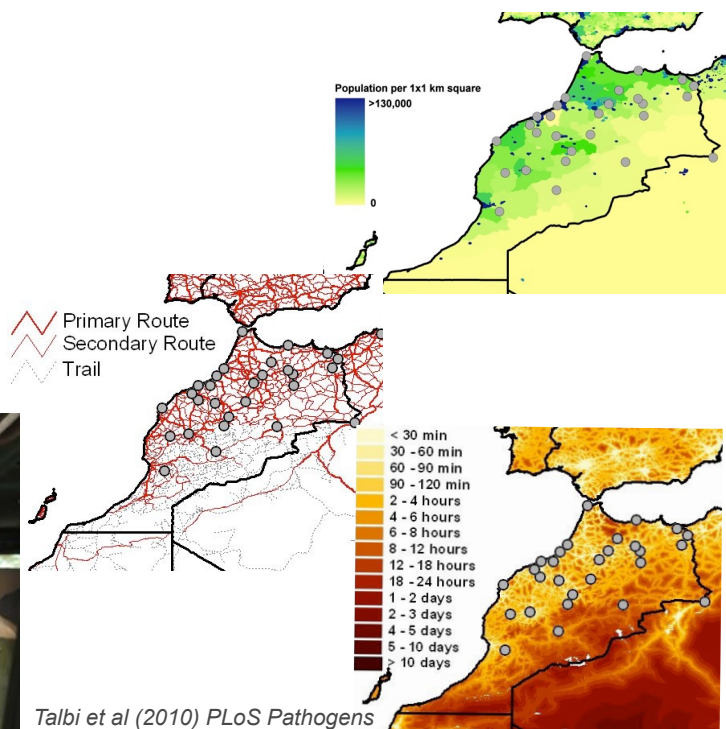
| | A | B | C | D |
|---|------------|------------|------------|------------|
| A | . | π_{Bj} | π_{Cj} | π_{Dk} |
| B | π_{Ai} | . | π_{Cl} | π_{Dm} |
| C | π_{Aj} | π_{Bl} | . | π_{Dn} |
| D | π_{Ak} | π_{Bm} | π_{Cn} | . |



Predictors of dog rabies diffusion in Morocco

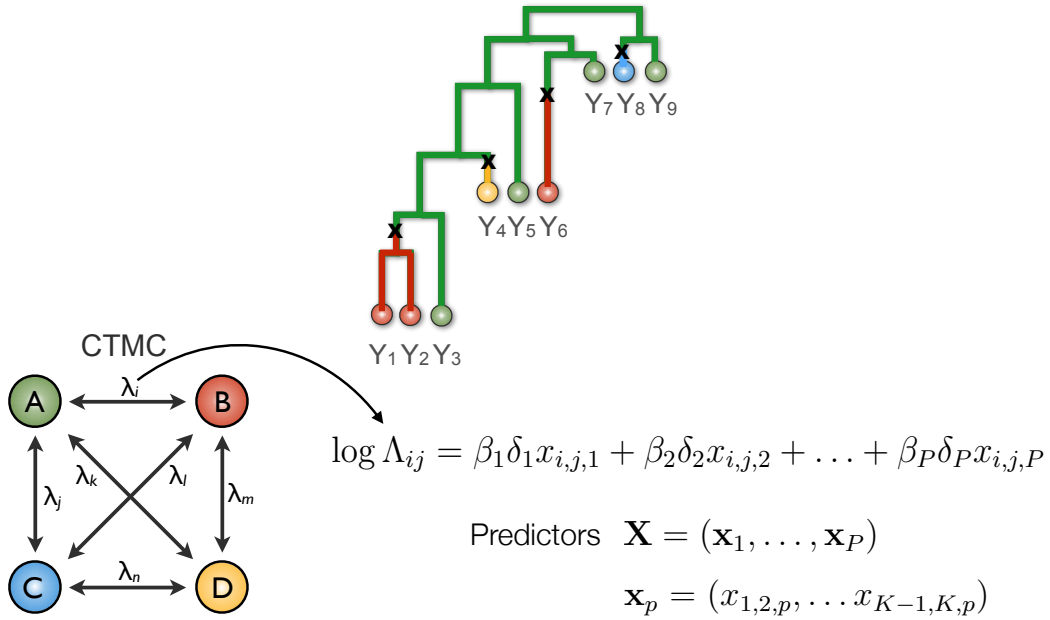
| Predictor | Ln Marginal likelihood |
|--------------------|------------------------|
| Equal rates | -320 |
| Distance | -299.8 |
| Population sizes | -381 |
| Gravity model | -388.6 |
| Population surface | -335.9 |
| Road distances | -298.5 |

Accessibility



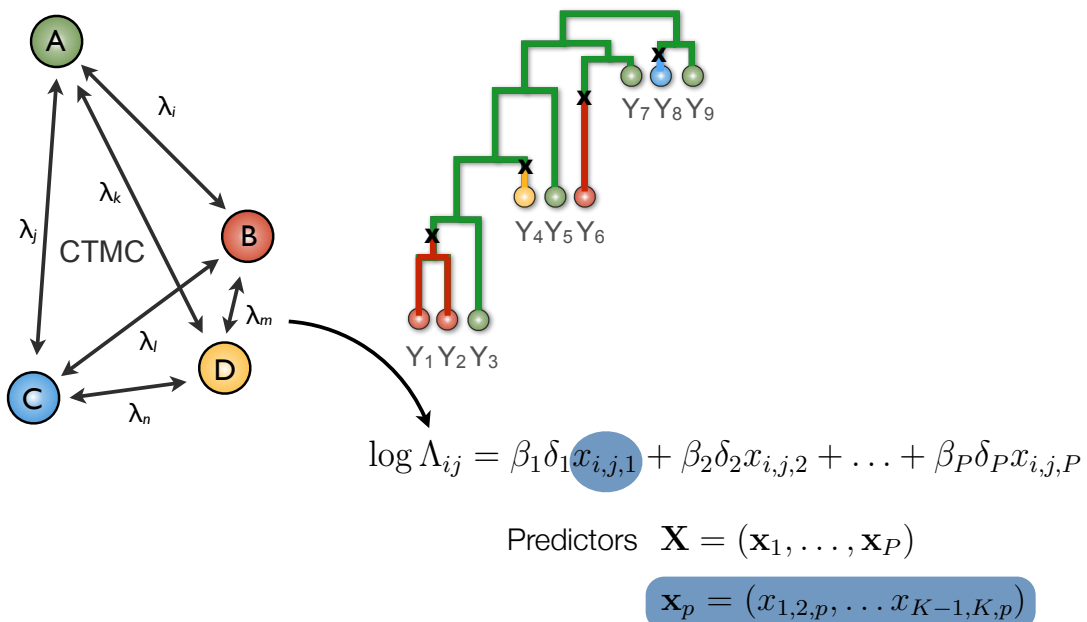
Talbi et al (2010) PLoS Pathogens

Predictors of phylogenetic diffusion



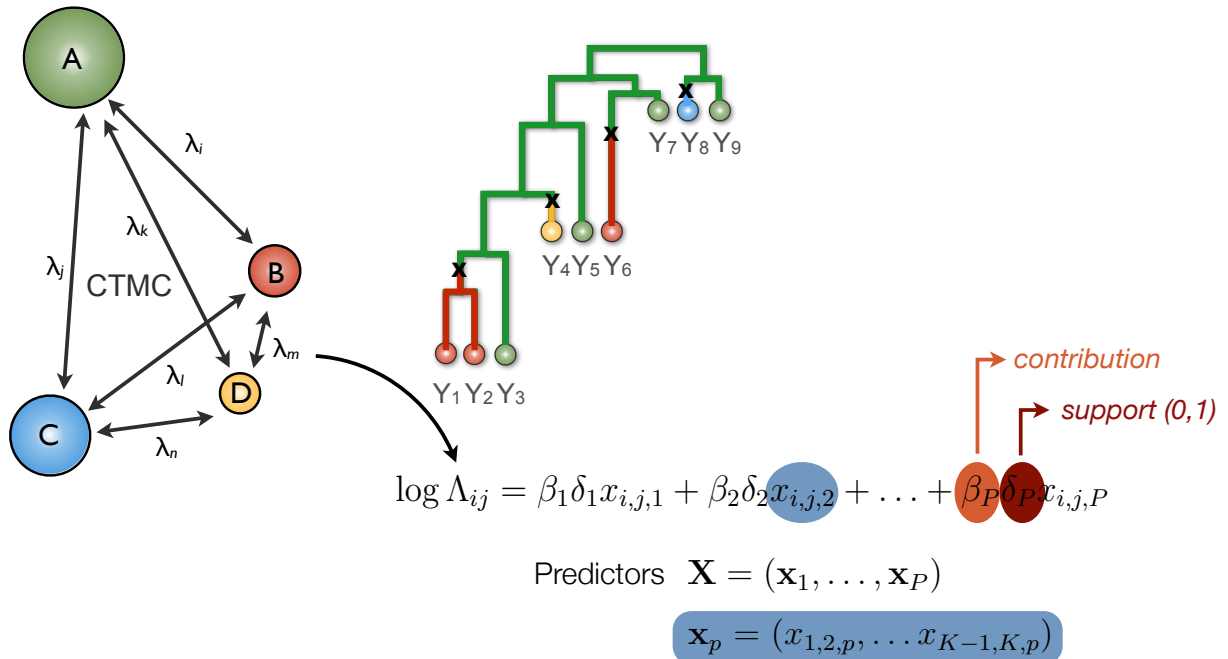
Lemey et al., PLoS Path, 2014

Predictors of phylogenetic diffusion



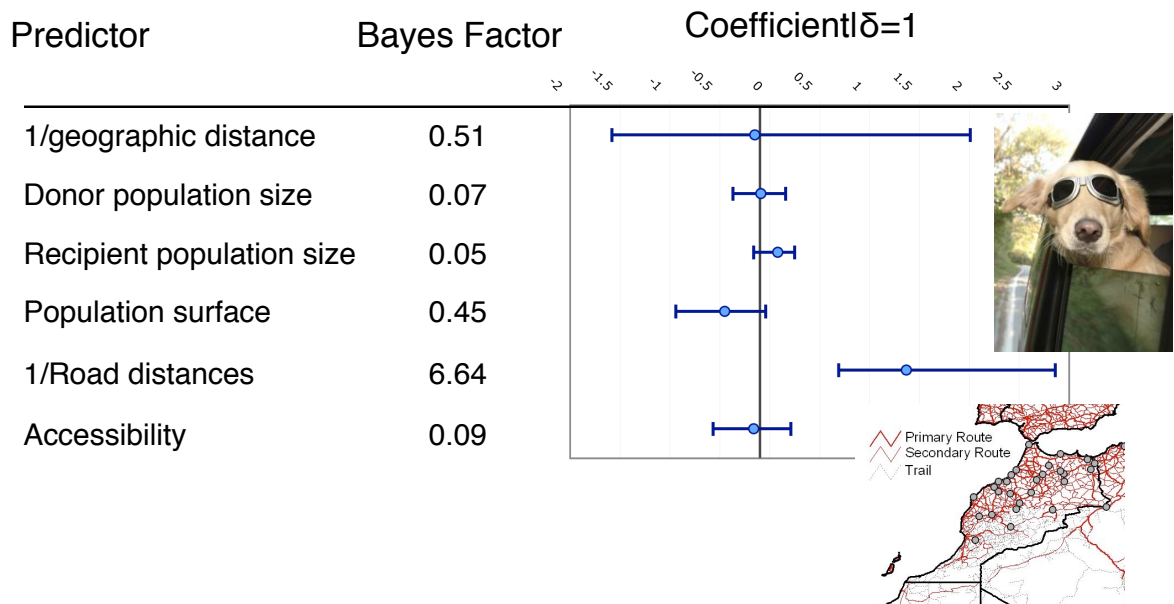
Lemey et al., PLoS Path, 2014

Predictors of phylogenetic diffusion

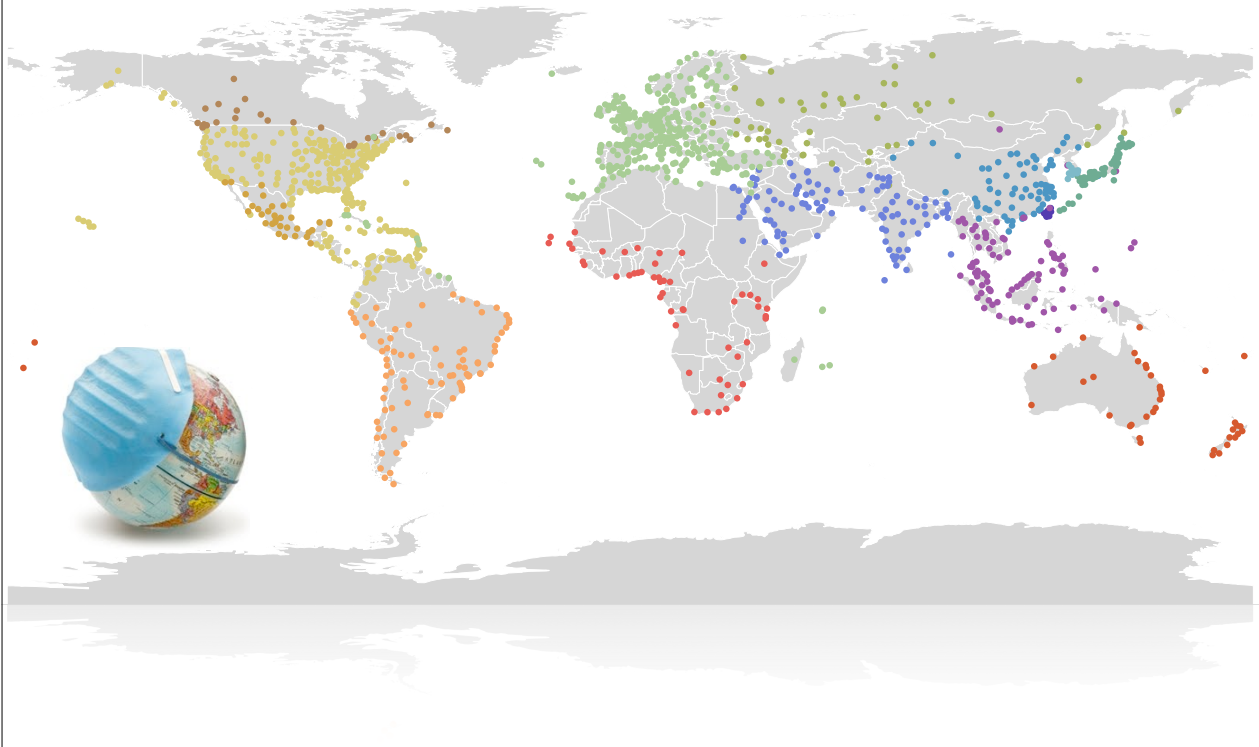


Lemey et al., PLoS Path, 2014

Predictors of dog rabies diffusion in Morocco



The global circulation of human influenza A H3N2



$$\log \lambda_{AB} = \delta_{\log P_1} \beta_{\log P_1} \log P_{1AB} + \delta_{\log P_2} \beta_{\log P_2} \log P_{2AB} + \dots + \delta_{\log P_n} \beta_{\log P_n} \log P_{nAB}$$

Predictors

Average distance

Minimum distance

Origin absolute latitude

Destination absolute latitude

Passenger flow

Origin population size

Destination population size

Origin population density

Destination population density

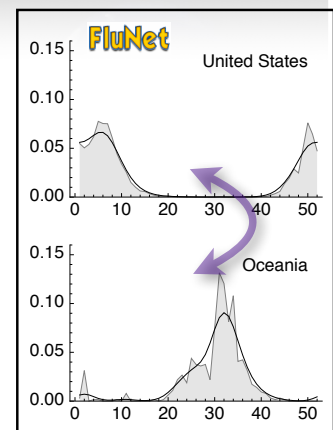
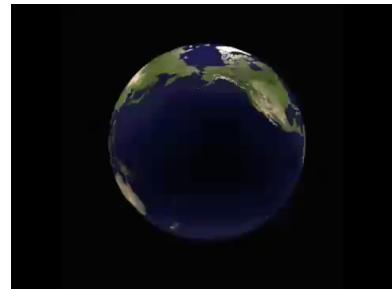
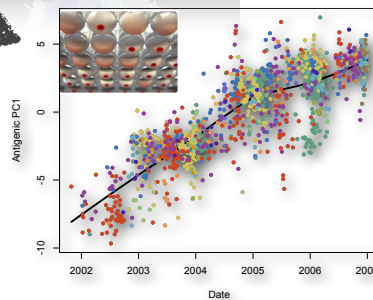
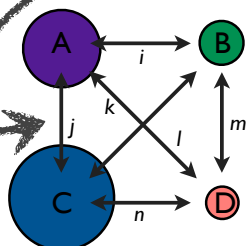
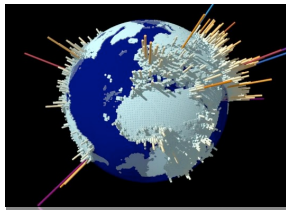
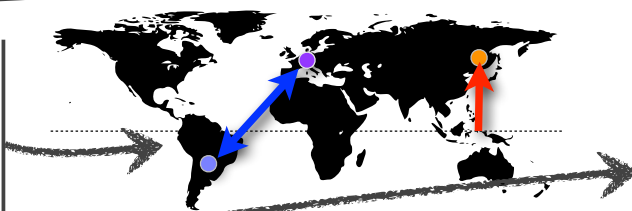
H3 incidence overlap

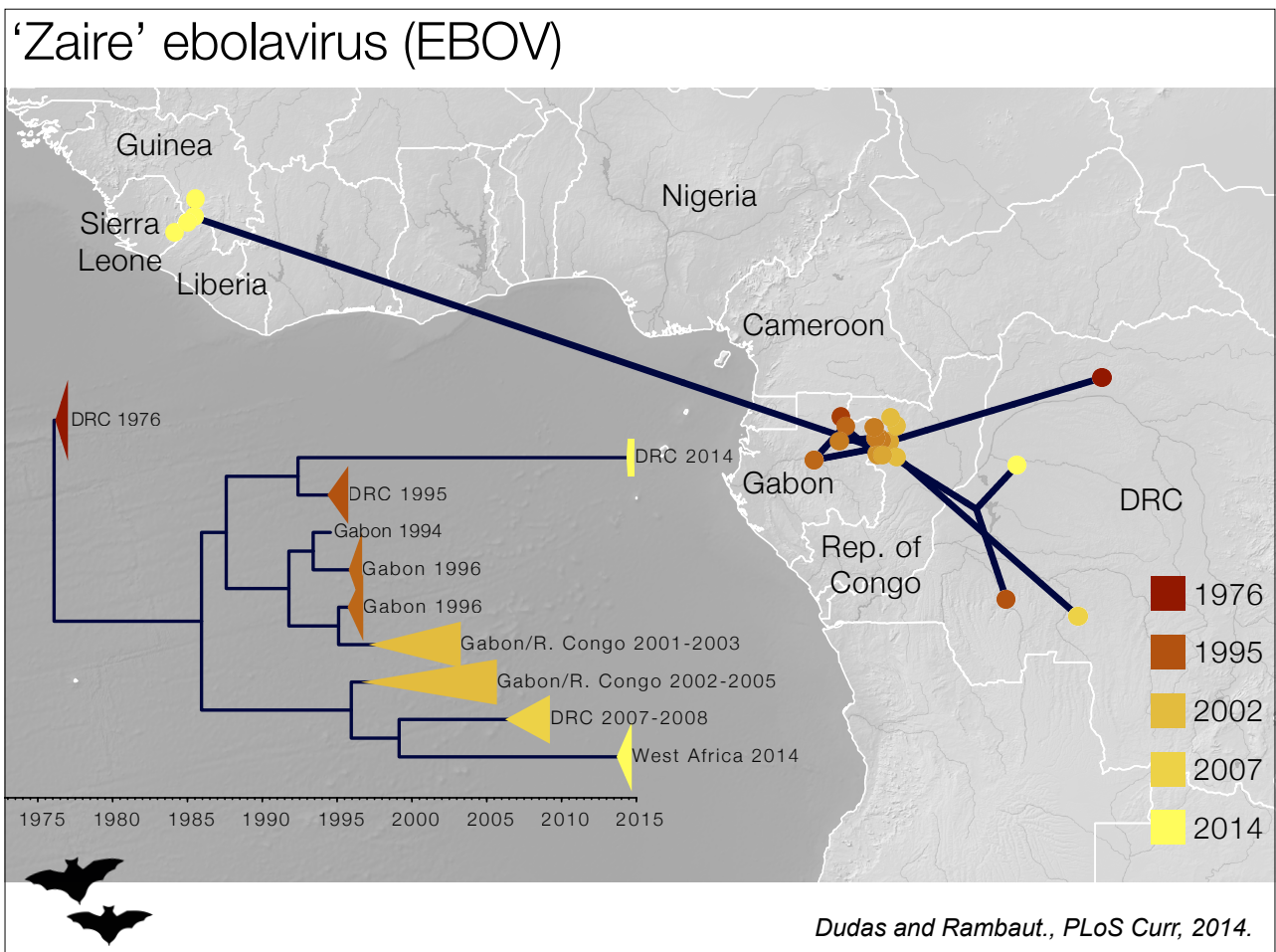
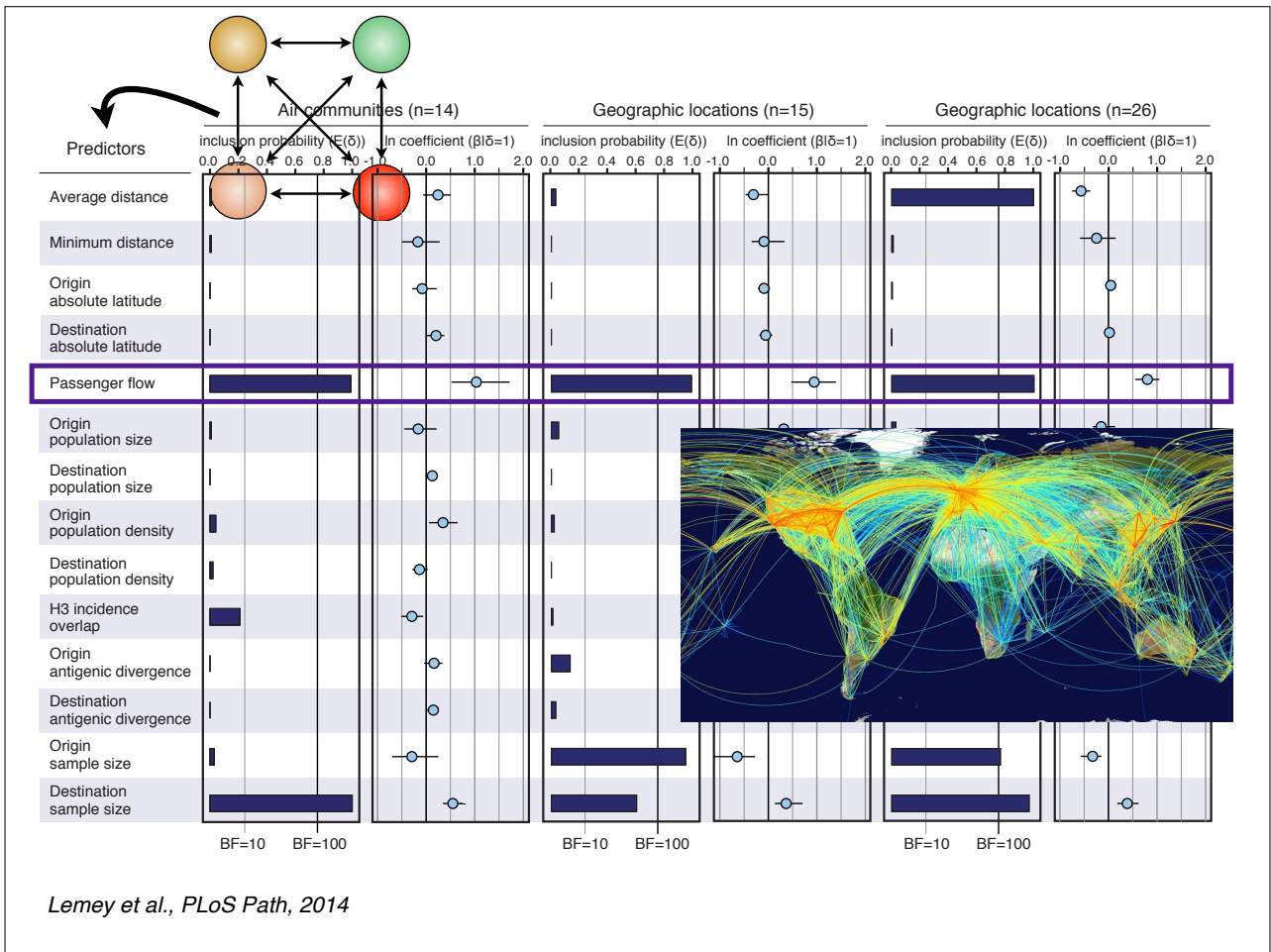
Origin antigenic divergence

Destination antigenic divergence

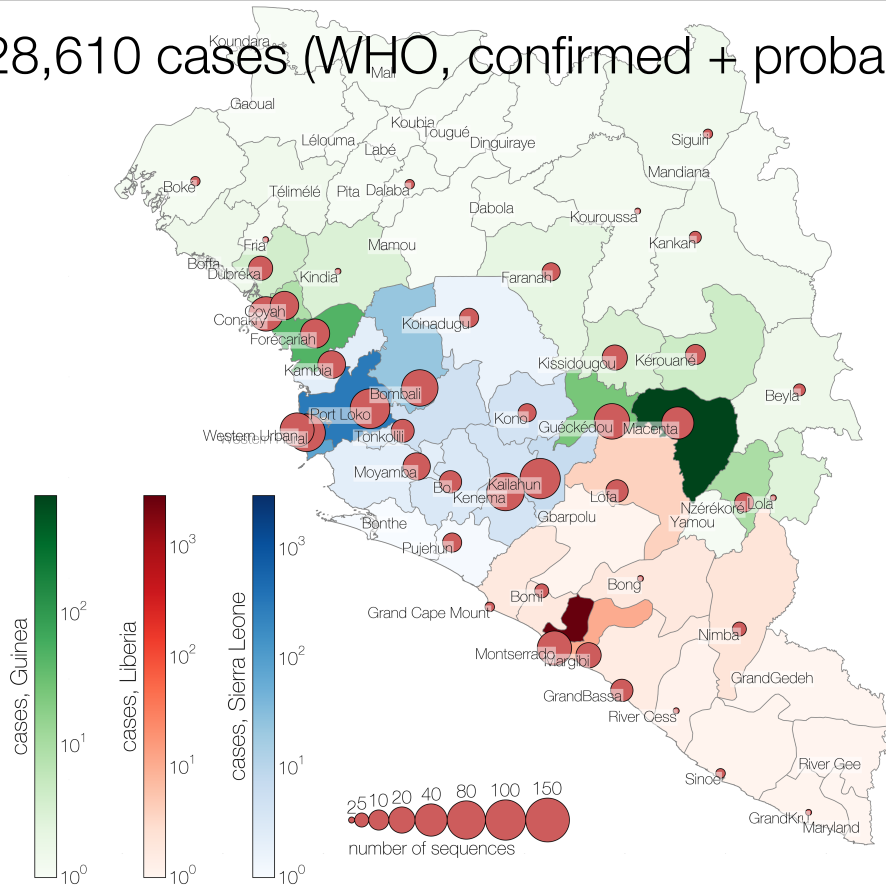
Origin sample size

Destination sample size

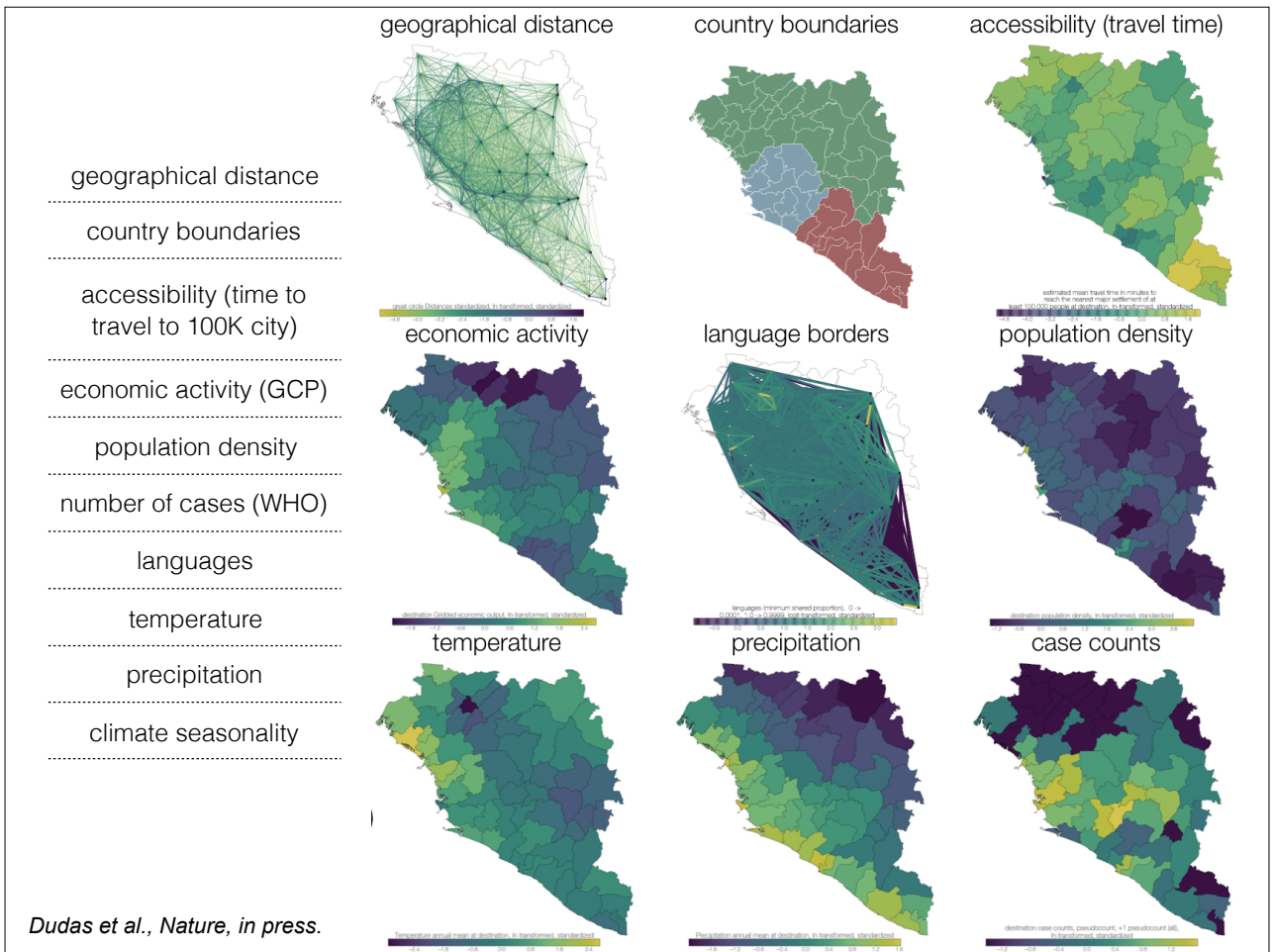




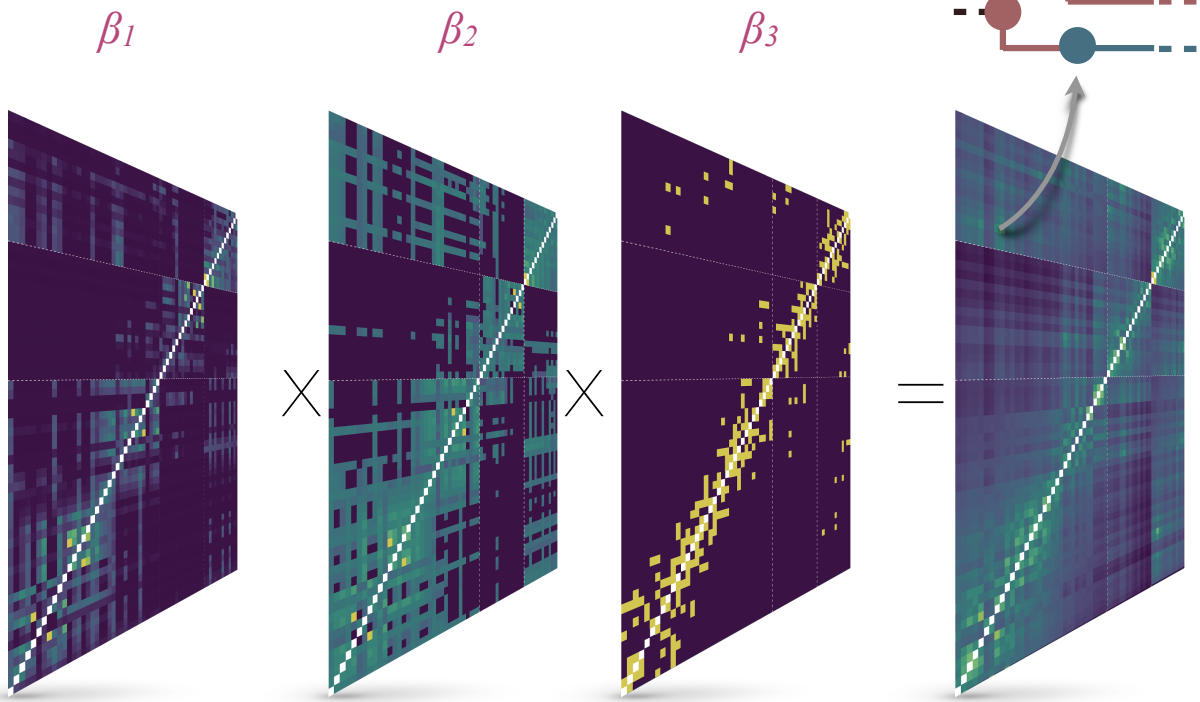
28,610 cases (WHO, confirmed + probable)



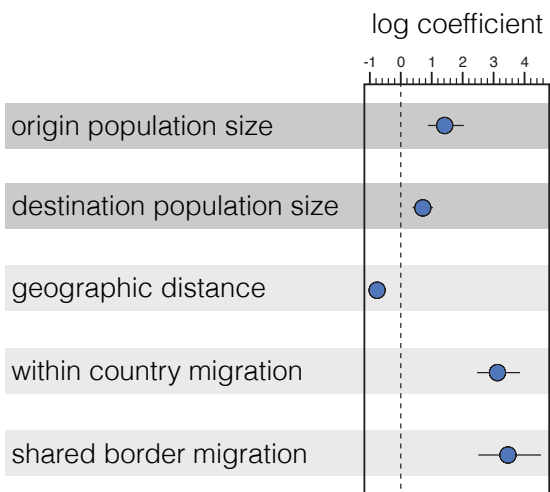
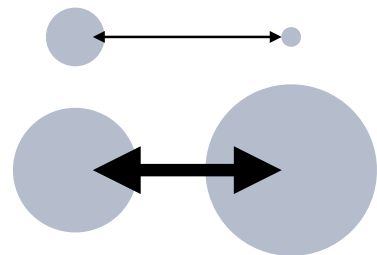
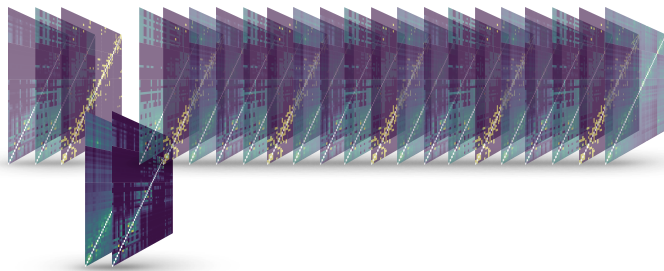
1610 EBOV genomes (>5% of cases)

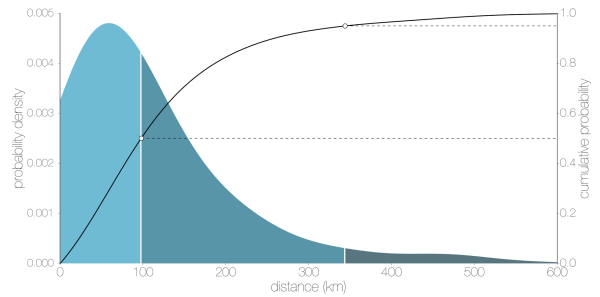
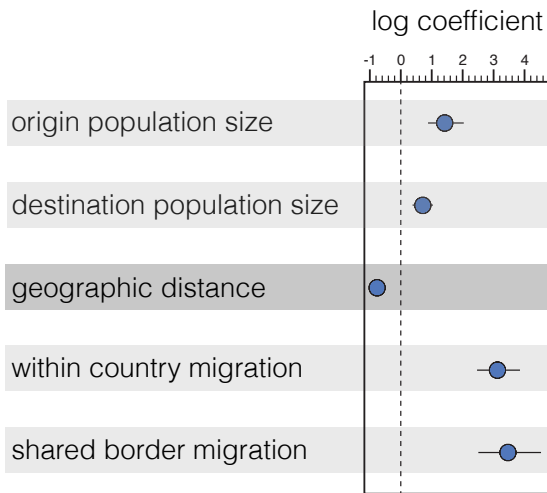
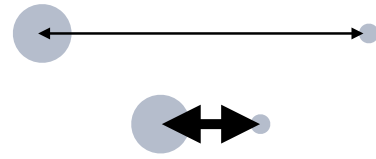
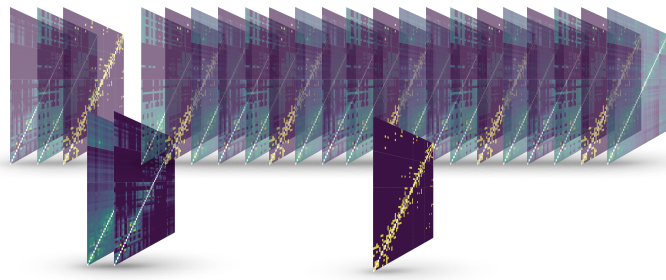


Dudas et al., Nature, in press.



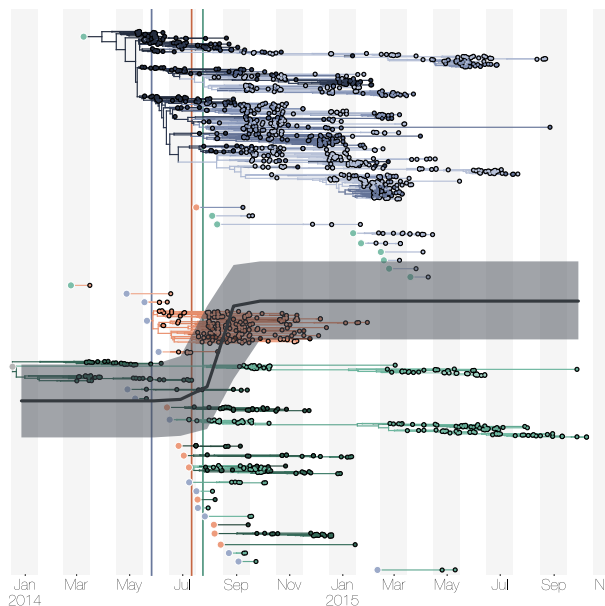
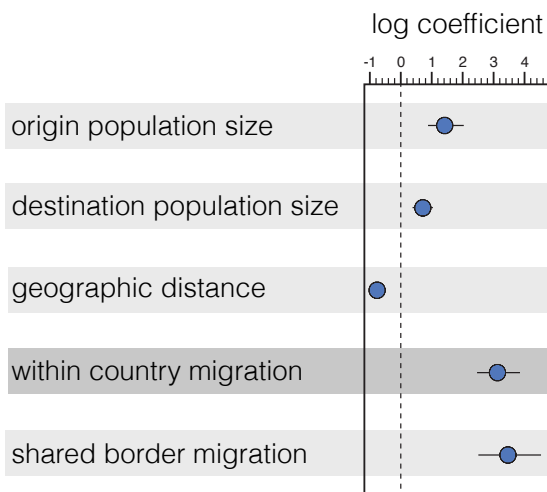
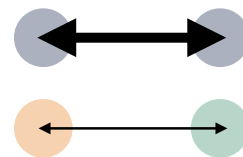
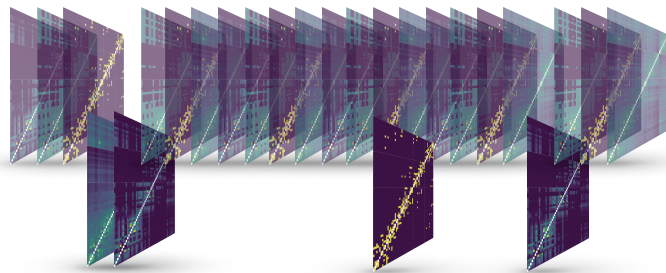
Phylogeographic generalized linear modelling



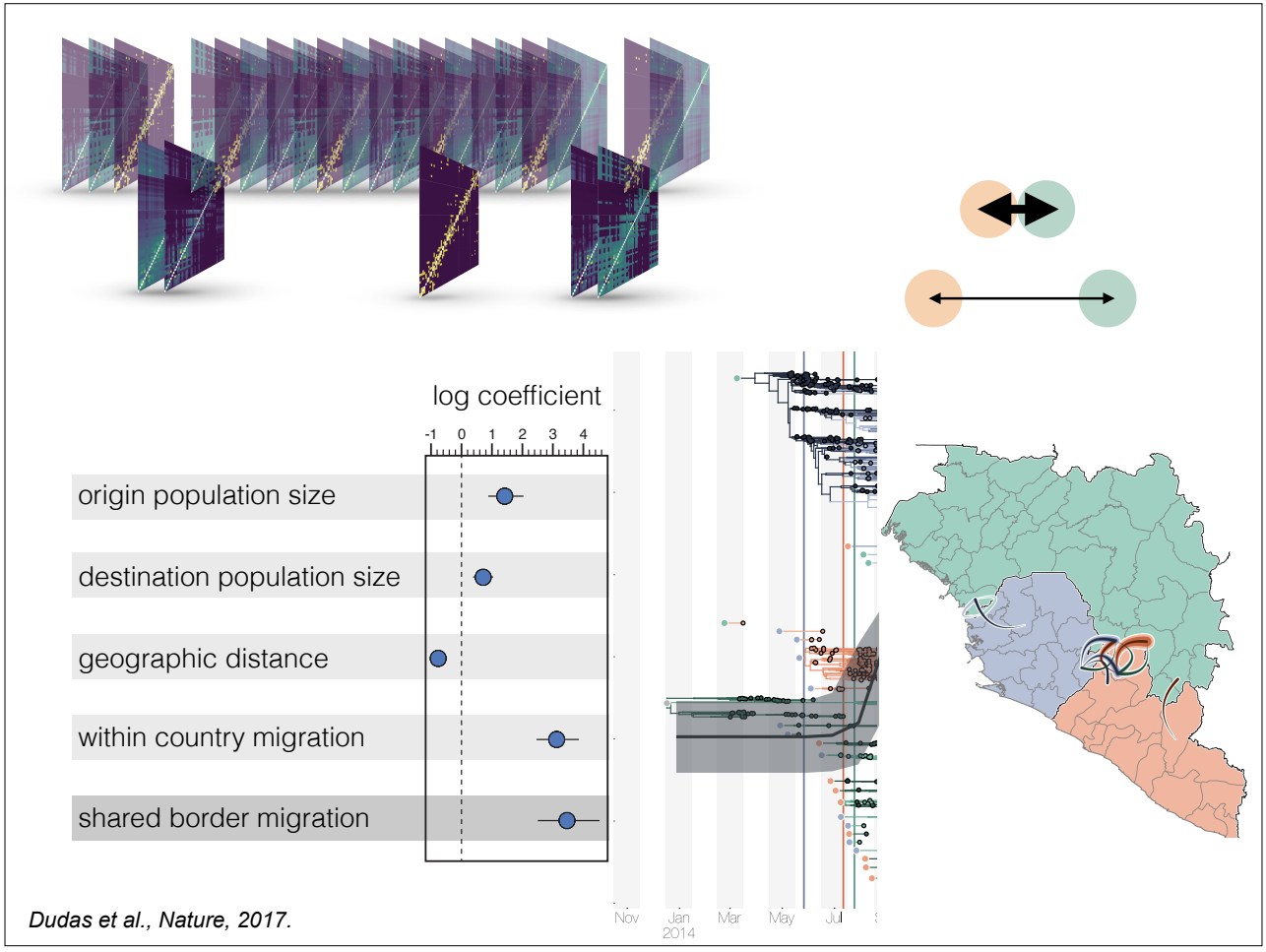


50% of virus migrations are by fewer than 100km
5% are by greater than 340km

Dudas et al., Nature, 2017.

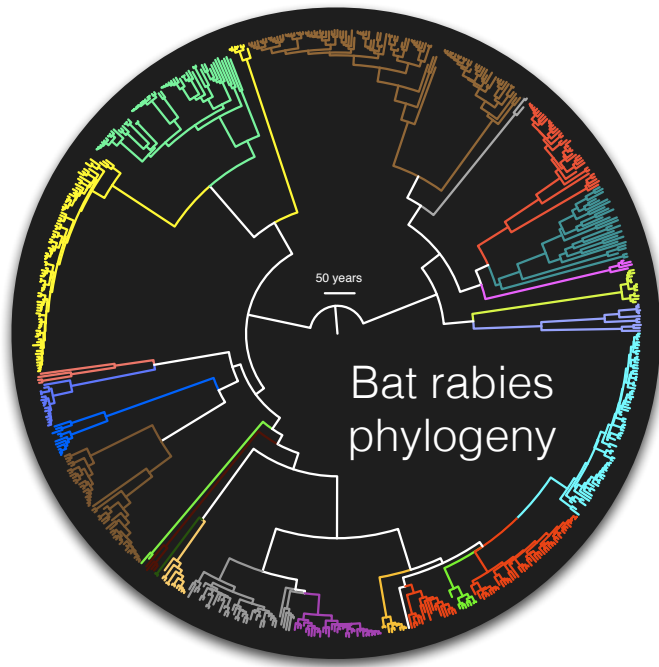
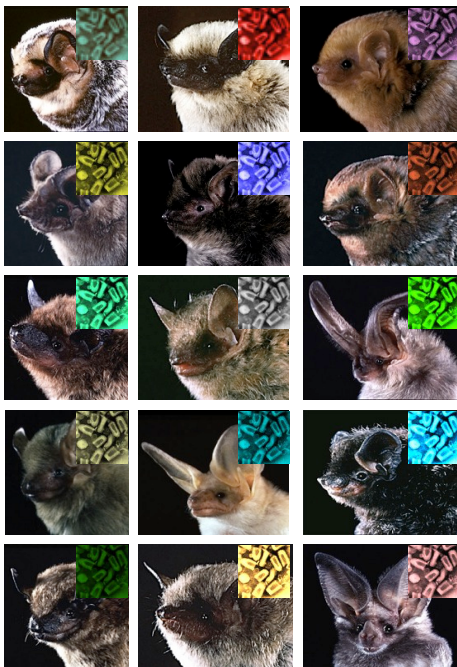


Dudas et al., Nature, 2017.



Dudas et al., Nature, 2017.

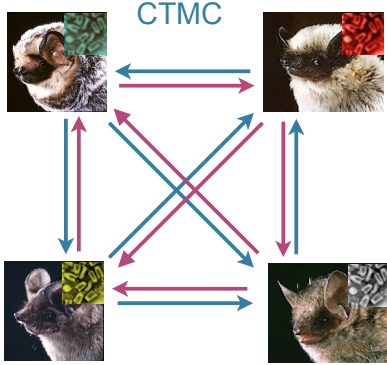
Uncovering cross-species dynamics bat rabies



Courtesy of D. Strecker

Uncovering bat rabies transmission dynamics

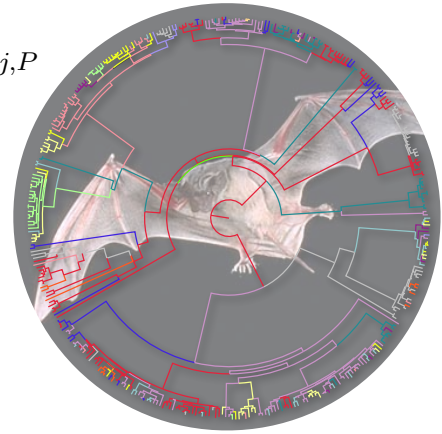
$$\log \Lambda_{ij} = \beta_1 \delta_1 x_{i,j,1} + \beta_2 \delta_2 x_{i,j,2} + \dots + \beta_P \delta_P x_{i,j,P}$$



contribution
inclusion probability

$$\mathbf{X} = (x_1, \dots, x_P)$$

- host divergence
- range overlap
- roost structure overlap
 - Δ wing aspect ratio
 - Δ wing loading
 - Δ body size



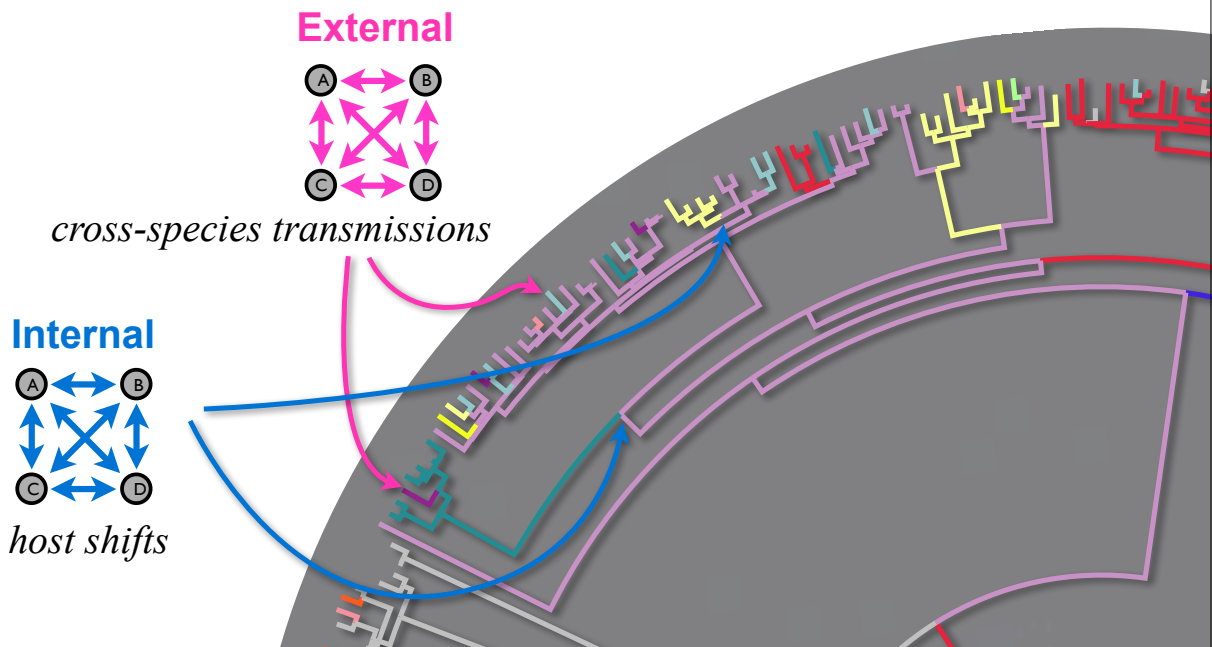
- 372 partial N genes (1997-2005)
- 17 different bat hosts
- 14 US states

Streicker et al., Science, 2010.

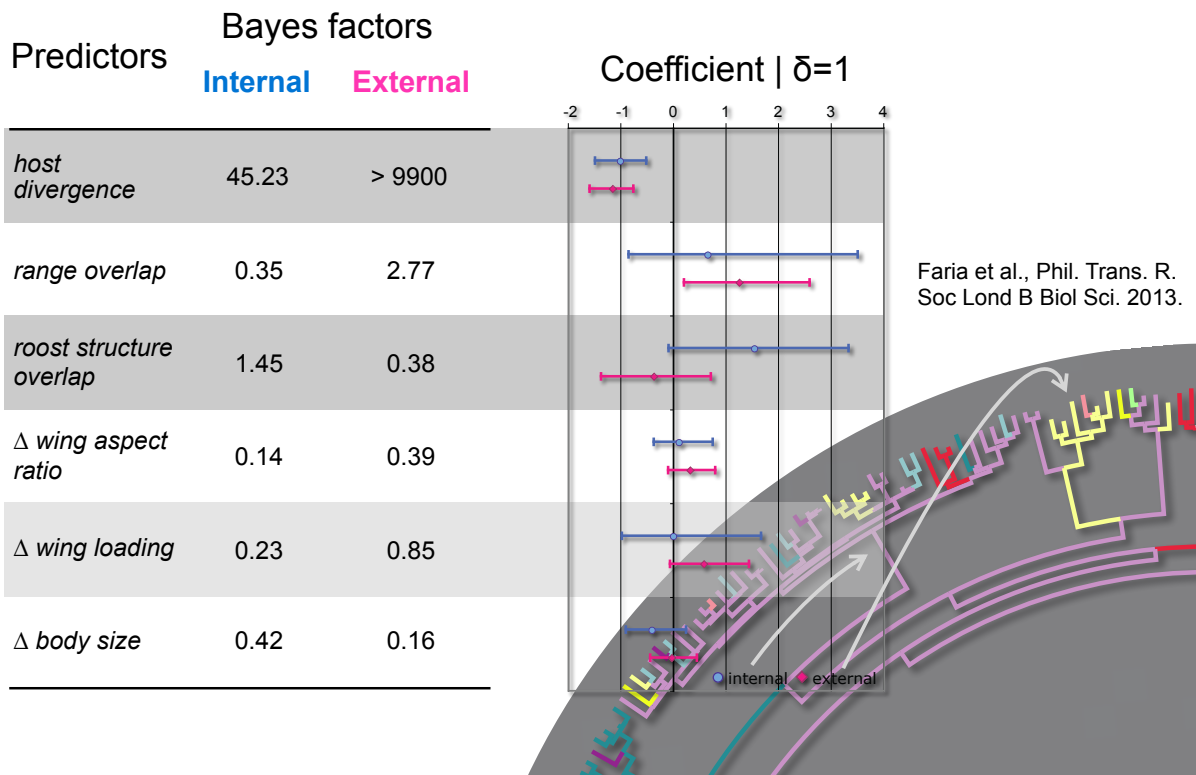
Faria et al., Phil. Trans. R. Soc Lond B Biol Sci. 2013.

Uncovering bat rabies transmission dynamics

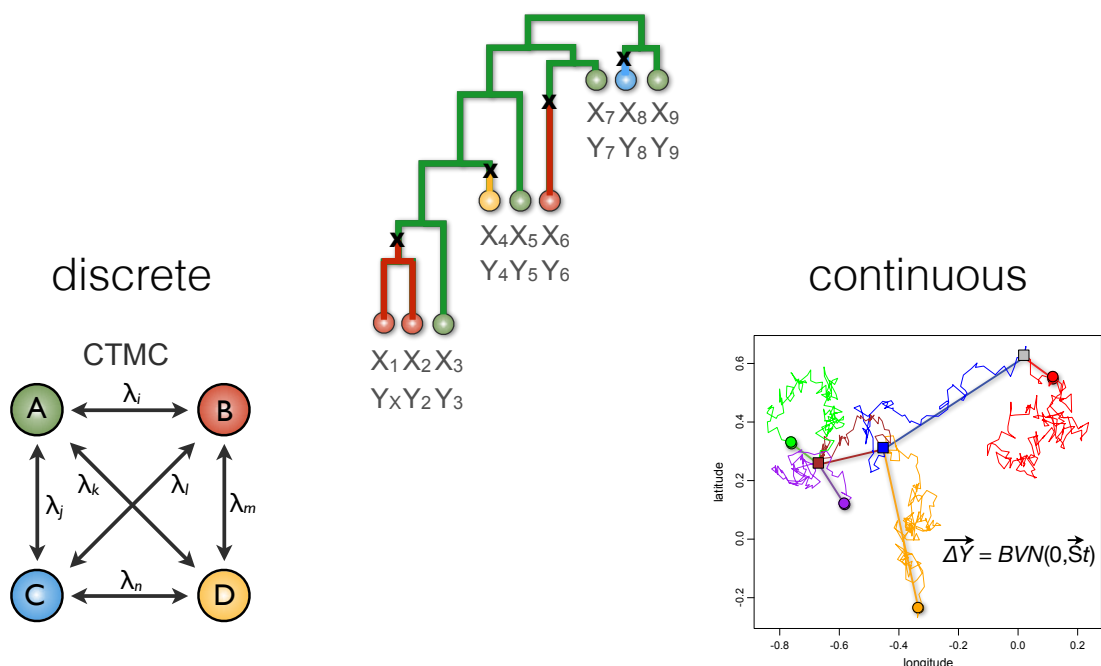
Faria et al., Phil. Trans. R. Soc Lond B Biol Sci. 2013.



Uncovering bat rabies transmission dynamics



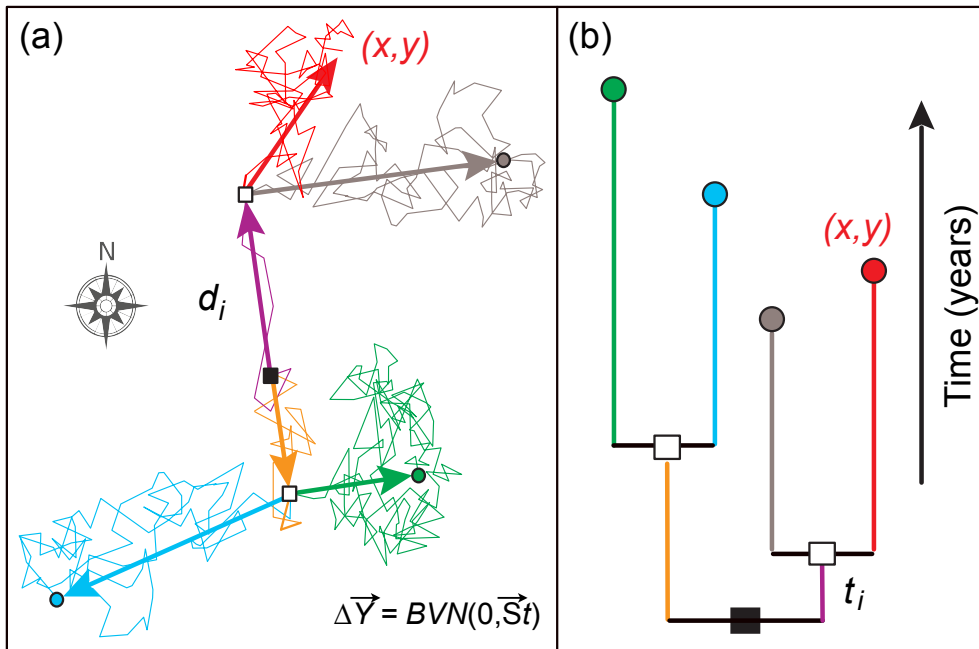
Phylogenetic diffusion models



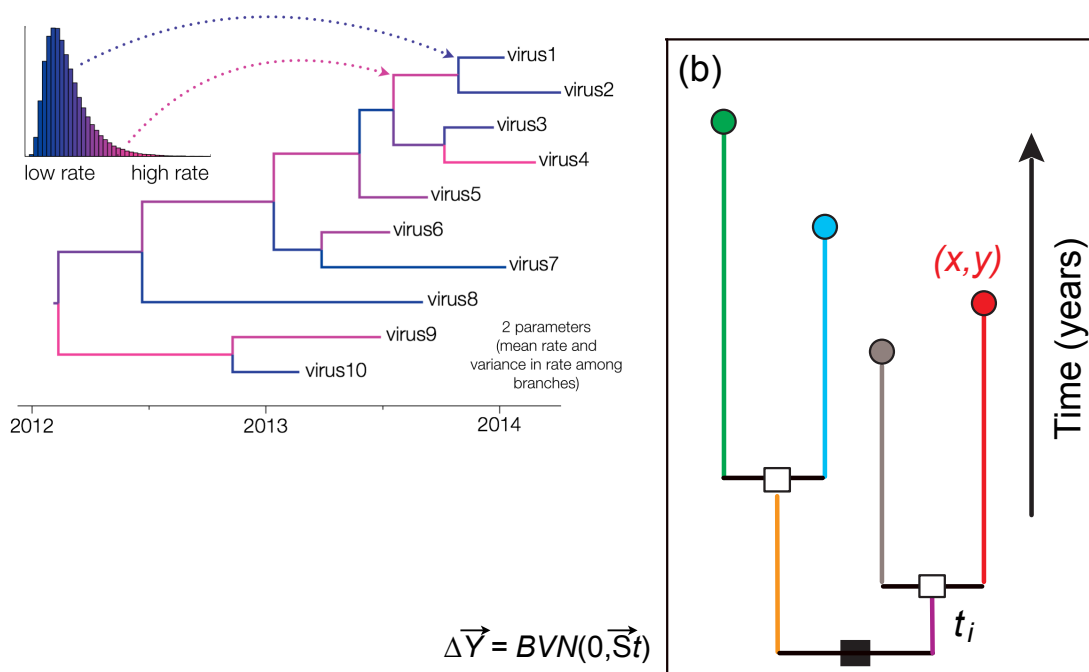
Lemey et al., PLoS Comp Bio, 2009

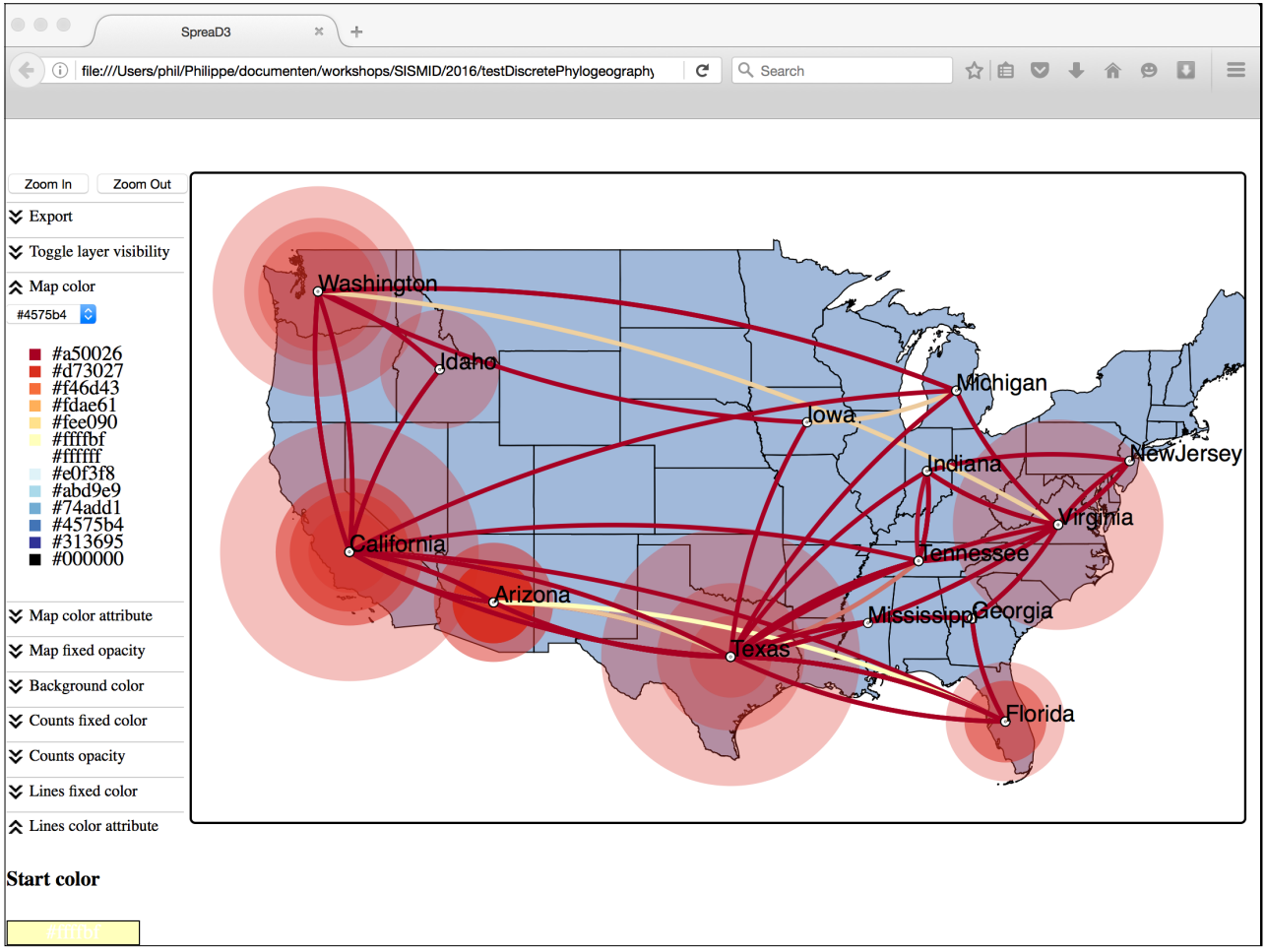
Lemey et al., MBE, 2010

Continuous phylogenetic diffusion

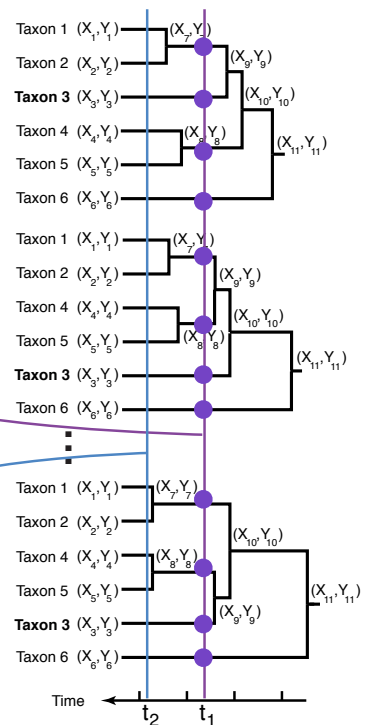
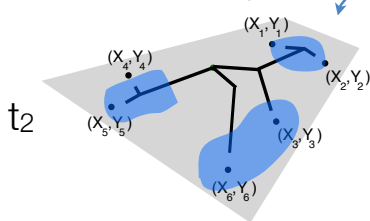
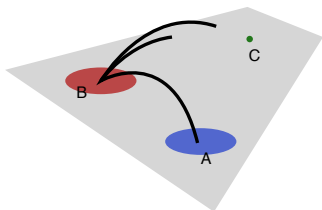
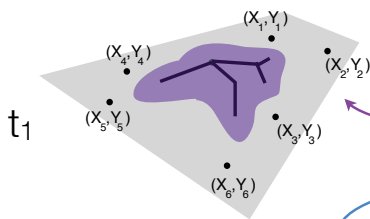
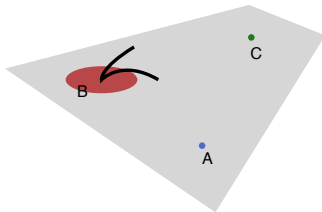
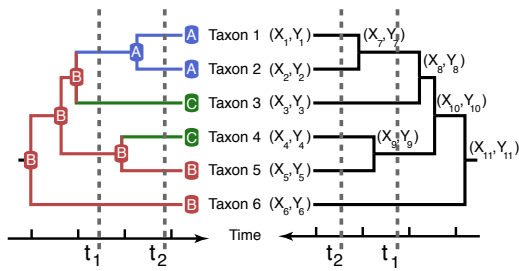


Relaxed random walk

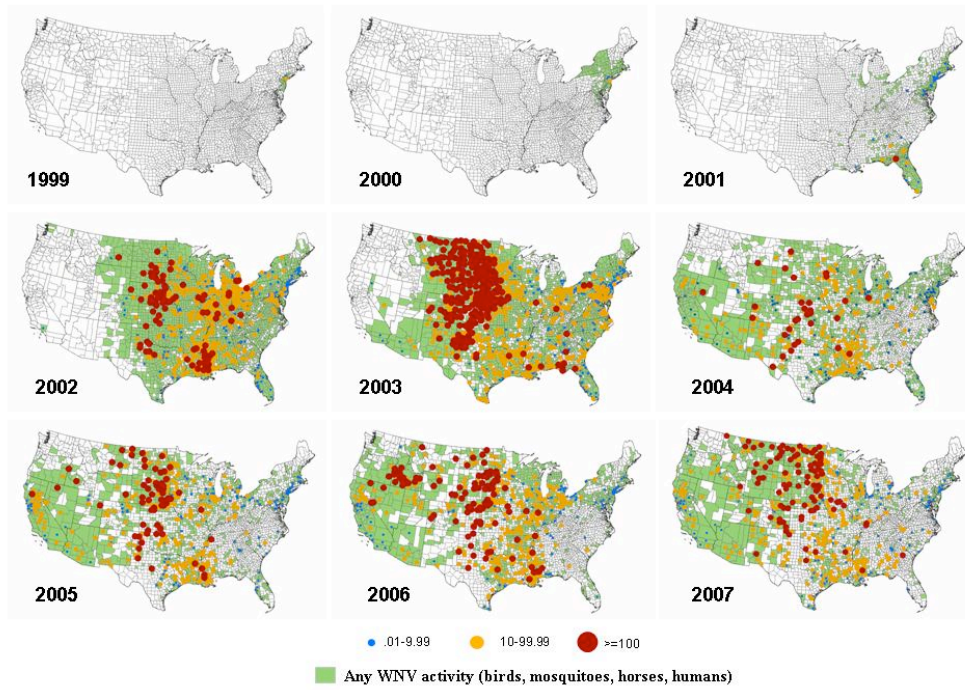


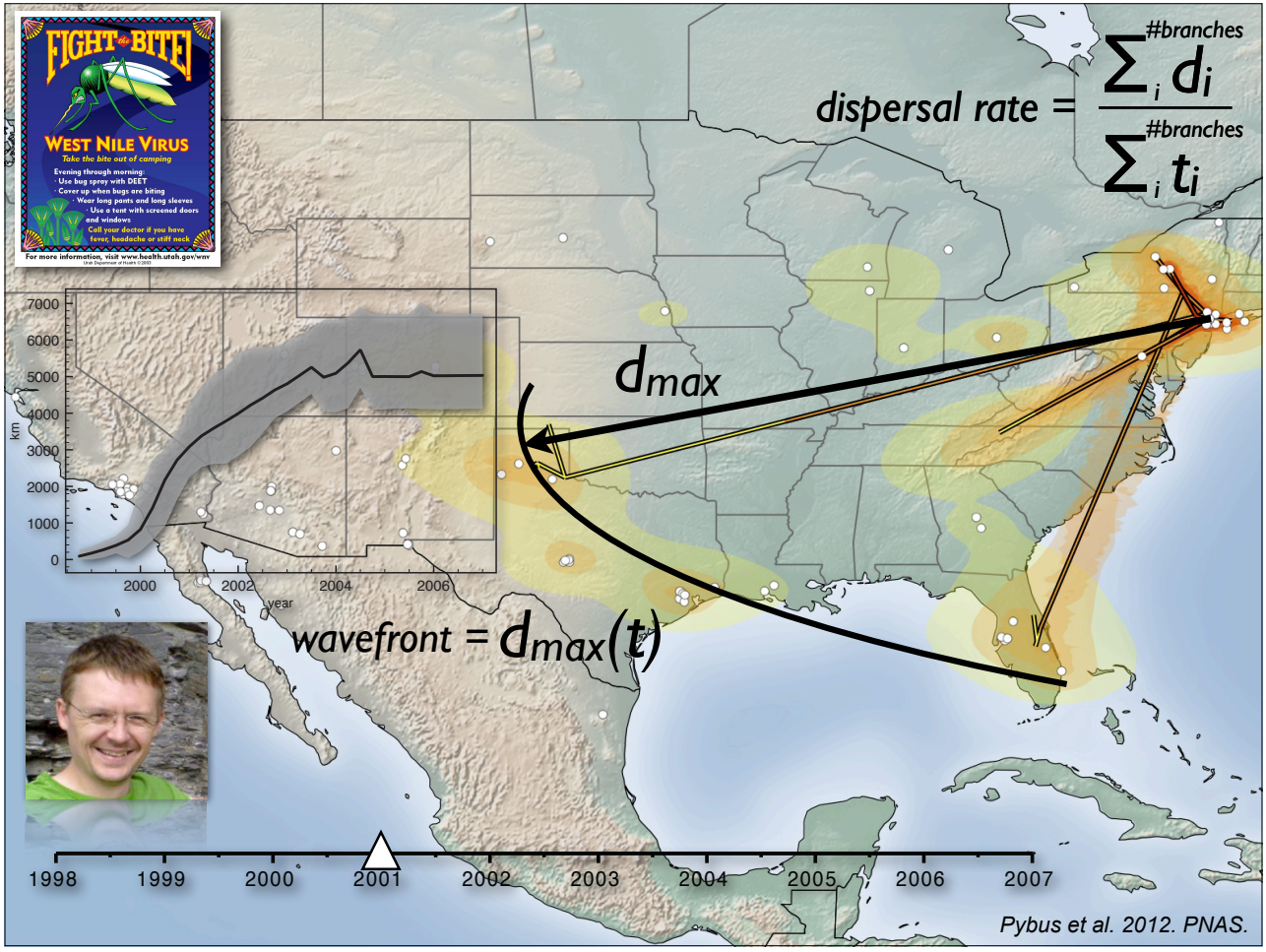


visualization

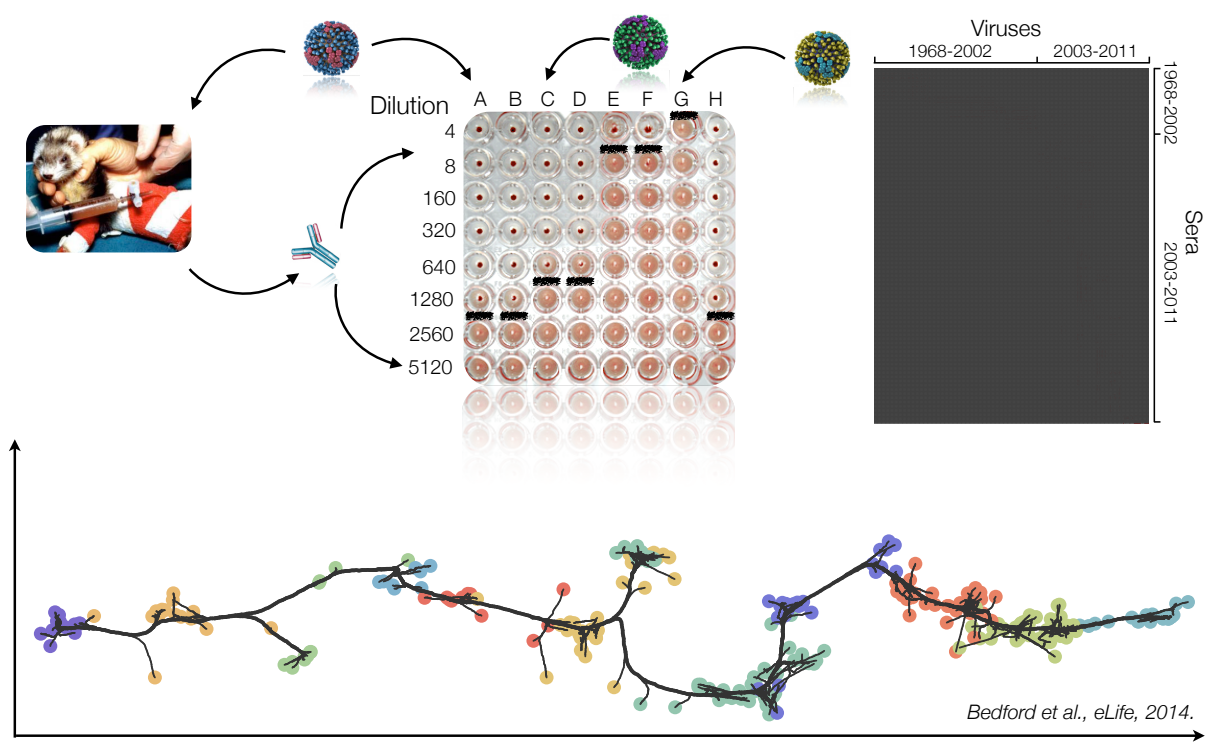


West Nile virus invasion in North America

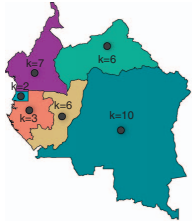
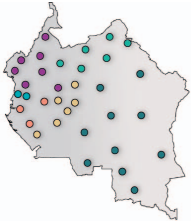


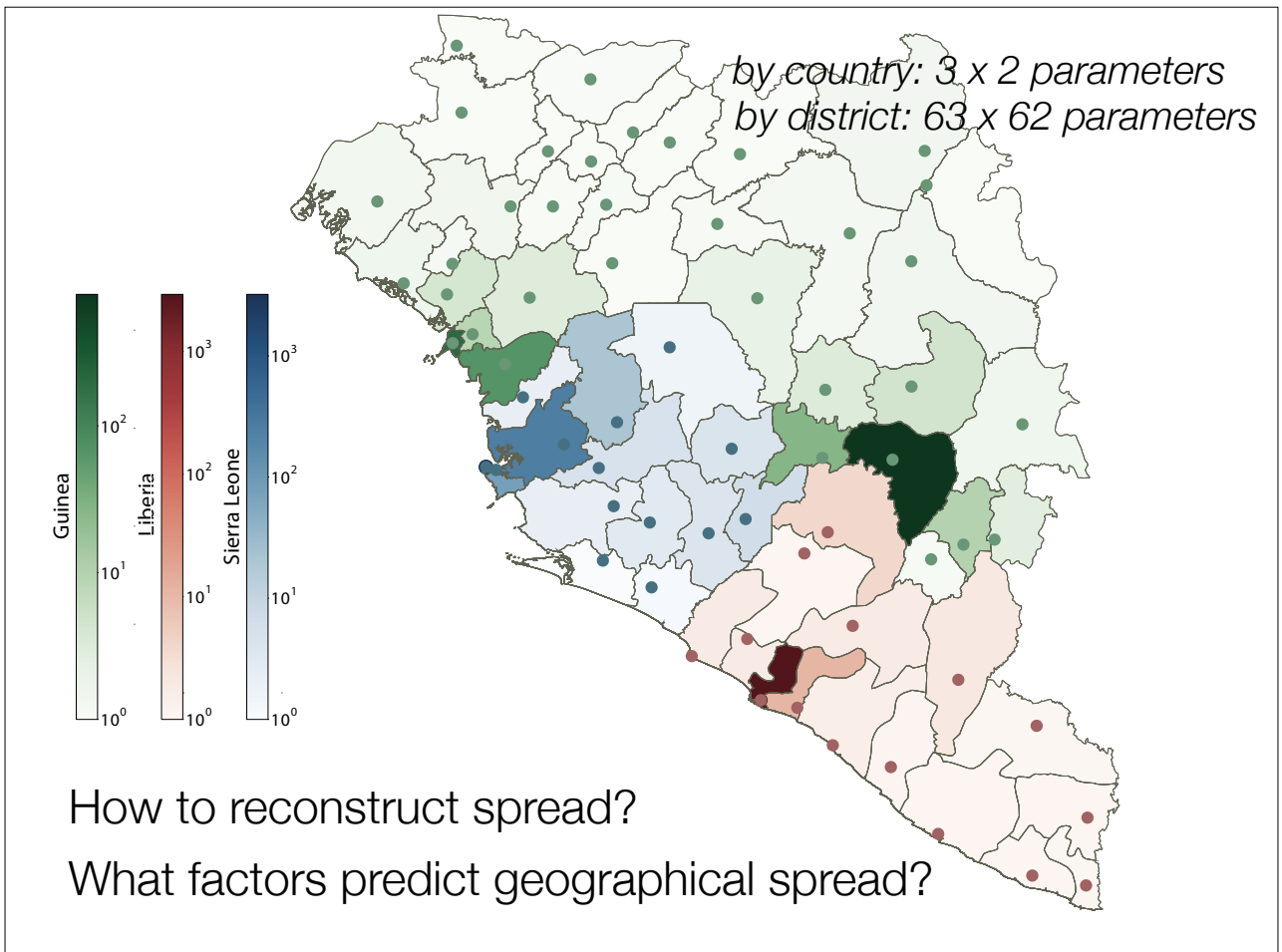


Phenotypic evolution: antigenic flux

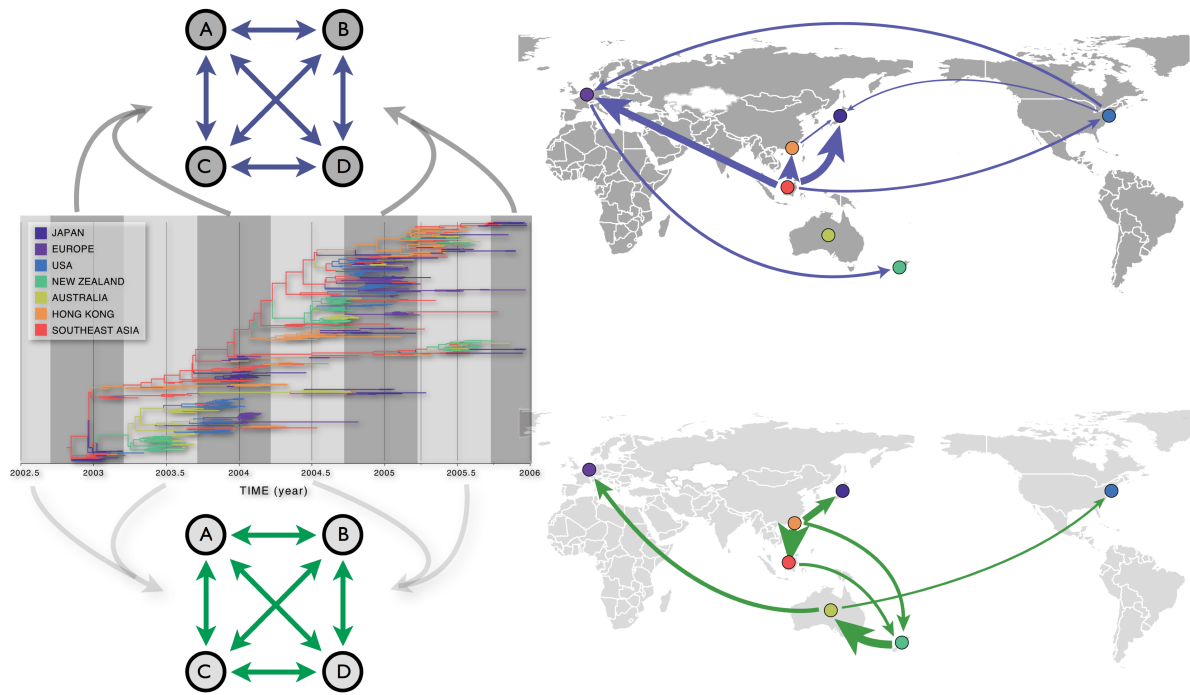


Phylogeography: discrete or continuous?

| | discrete | continuous |
|---------------------|---|---|
| Sampling |  |  |
| BIAS!! | | |
| Ancestral locations | sampling locations | anywhere |
| Geography | complete abstraction flexible | dispersal $\sim f(\text{distance})$ (host mobility dependent) distance = euclidean! |
| hypothesis testing | GLM | Dellicour, BMC Bioinformatics, 2016 |



Temporal heterogeneity



Bielejec et al (2014) Sys Bio