



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

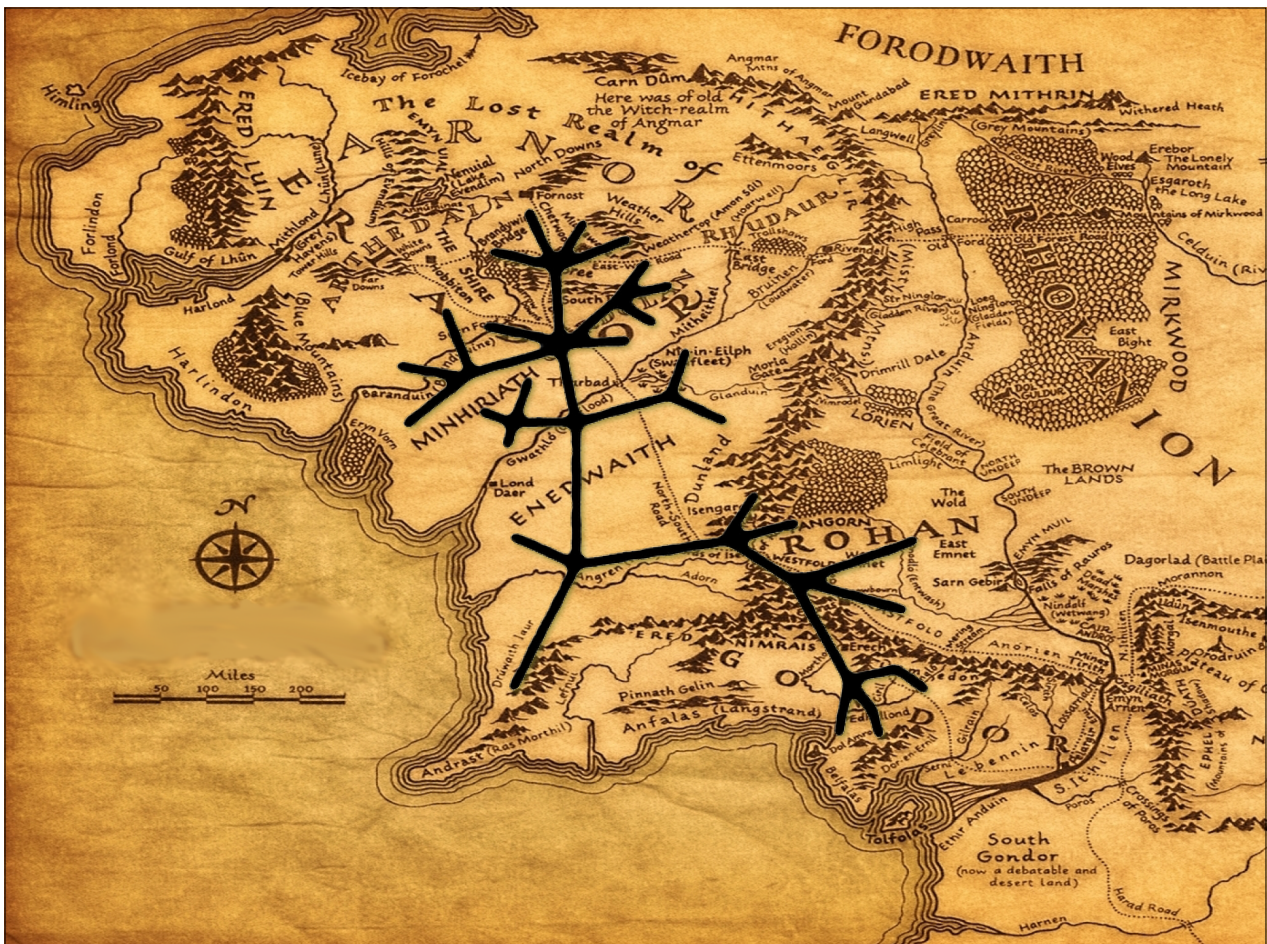


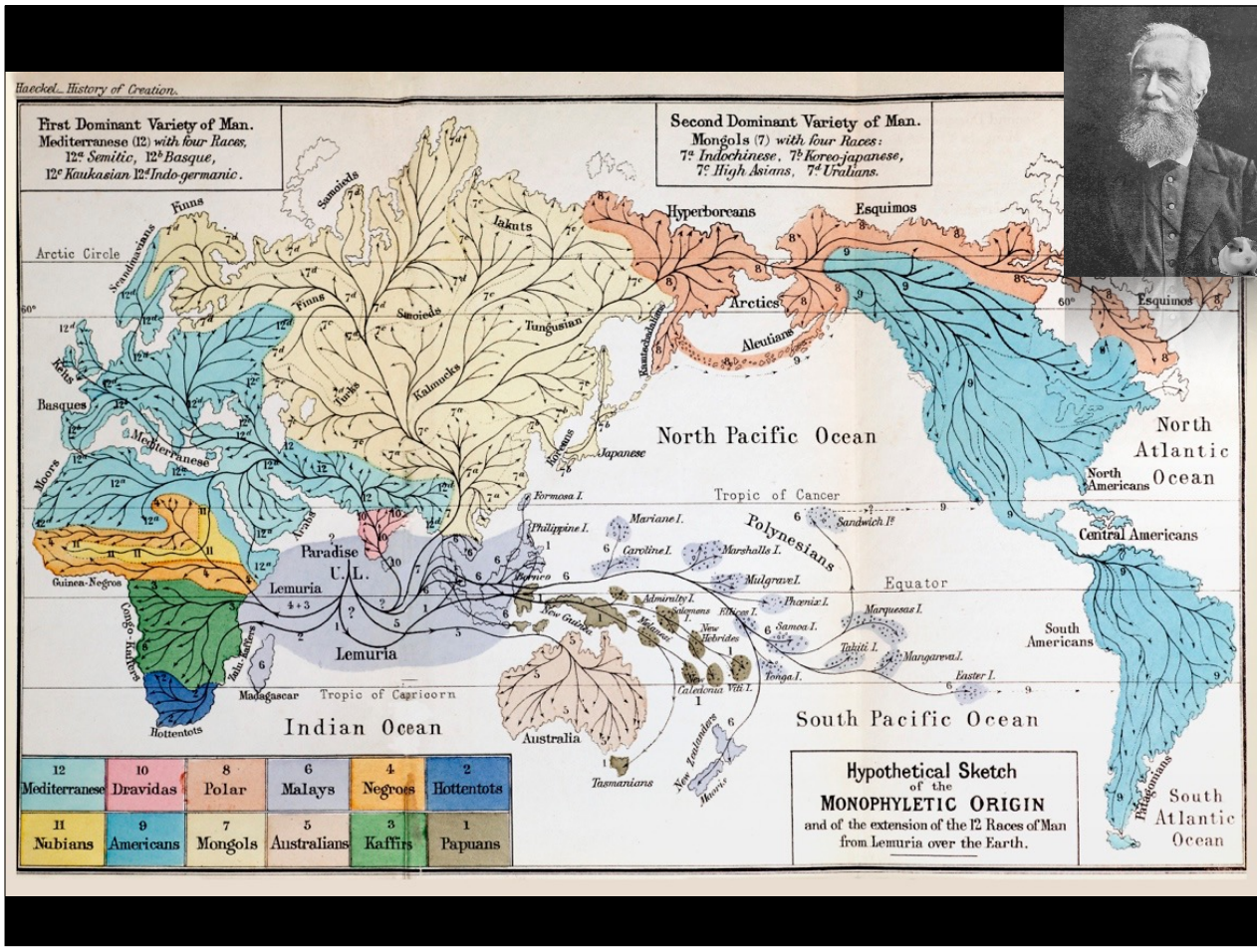
# Phylogenetic diffusion models

Philippe Lemey<sup>1</sup>, Guy Baele<sup>1</sup> & Marc Suchard<sup>2</sup>

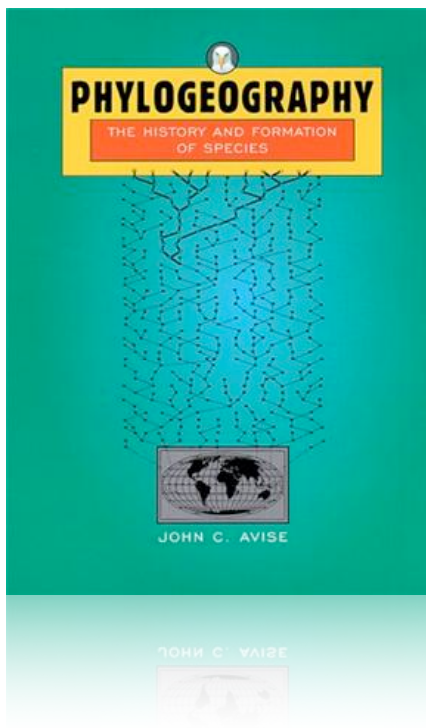
1. Rega Institute, Department of Microbiology and Immunology, K.U. Leuven, Belgium.
2. Departments of Biomathematics and Human Genetics, David Geffen School of Medicine at UCLA. Department of Biostatistics, UCLA School of Public Health

SISIMD, July 20-22, 2016



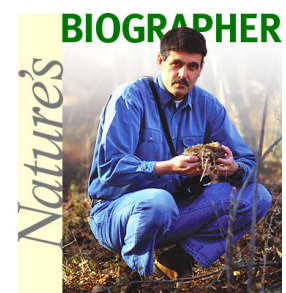


# Phylogeography

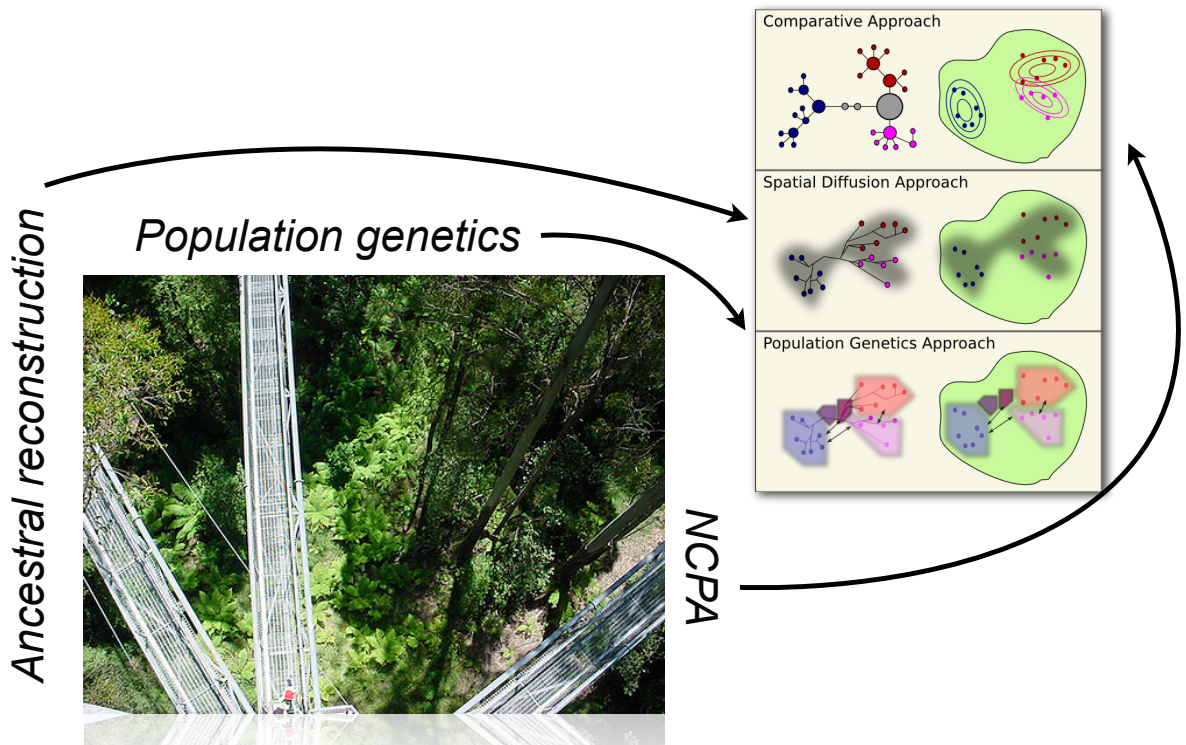


“a field of study concerned with the principles and processes governing the geographic distribution of genealogical lineages, especially those within and among closely related species.”

Avise, 2000

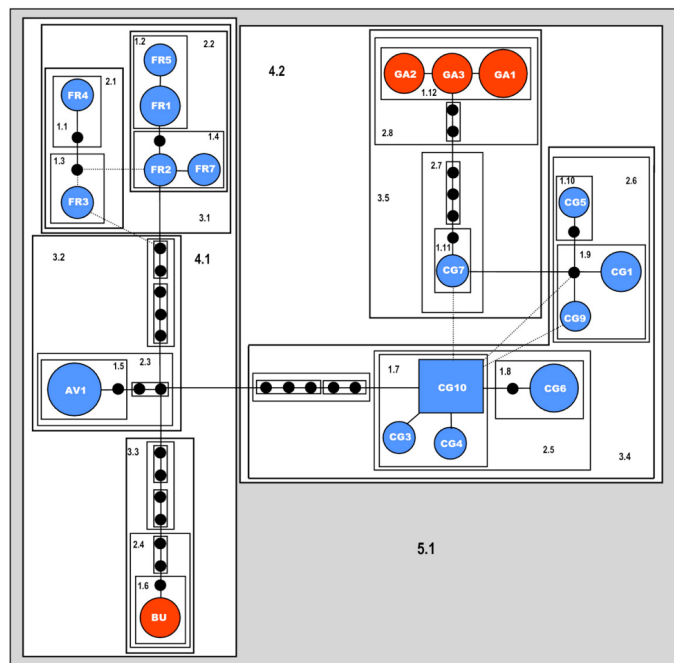
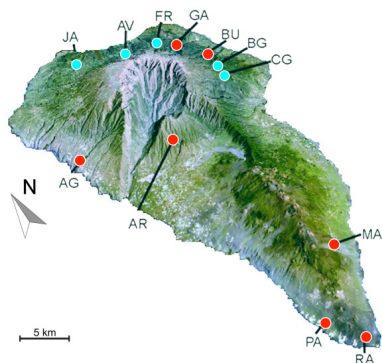


# Phylogeography: three roads diverged?



# Phylogeographic inference (road I)

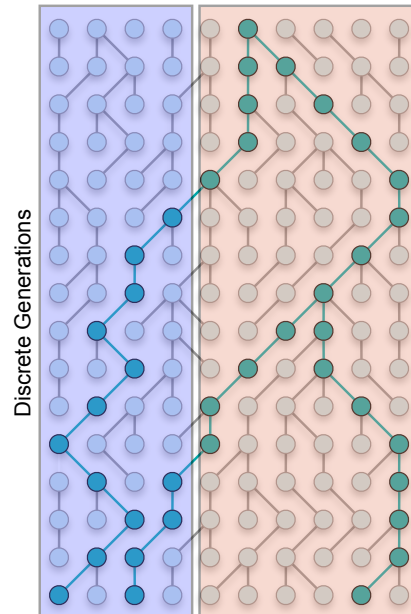
Nested clade phylogeographic analysis (NCPA)



# 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>

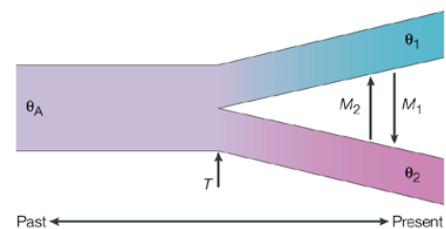


**MIGRATE-N**

estimation of population sizes and gene flow using the coalescent



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



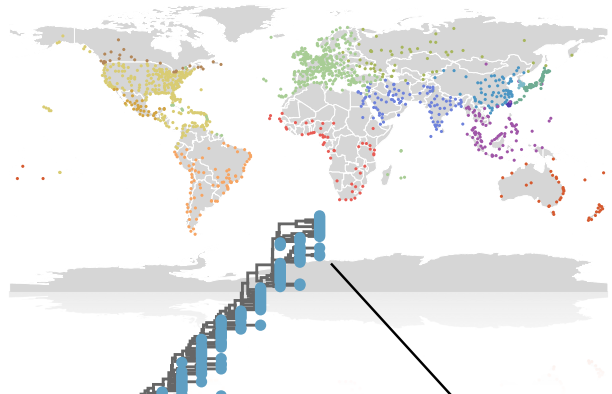
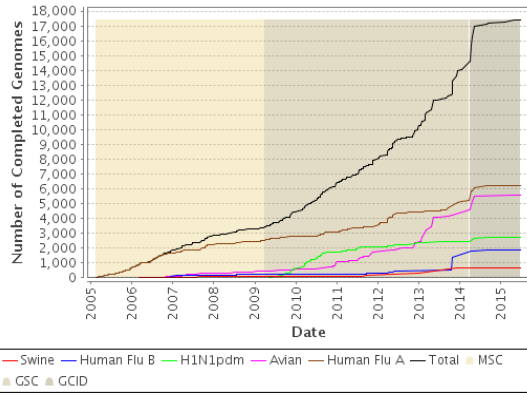
<http://genfaculty.rutgers.edu/hey/software>

**MDIV:** [http://www.biom.cornell.edu/Homepages/Rasmus\\_Nielsen/files.htm](http://www.biom.cornell.edu/Homepages/Rasmus_Nielsen/files.htm)

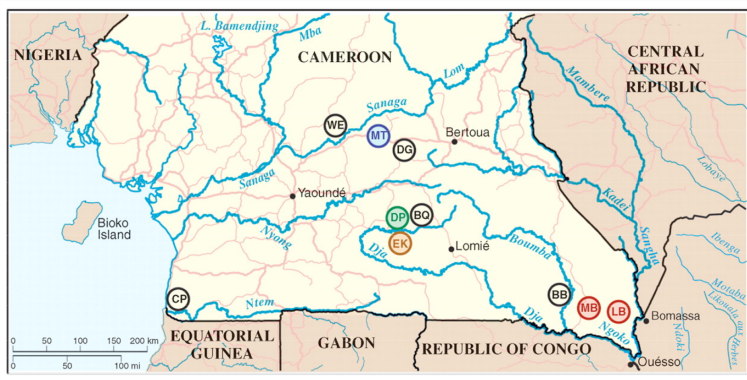
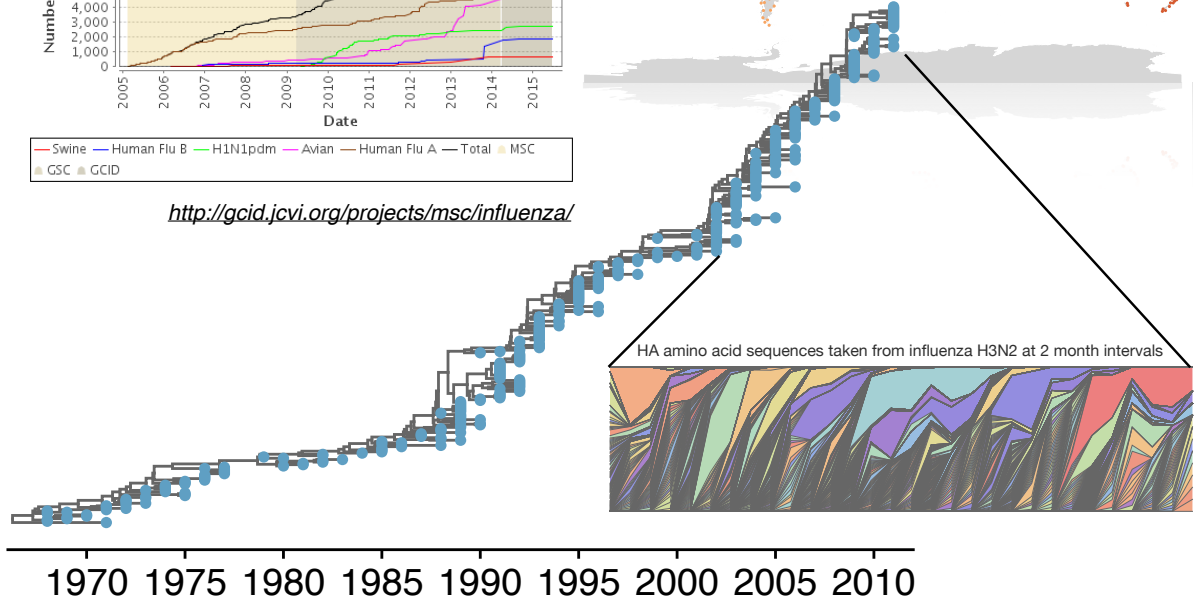
**Batwing:** <http://www.maths.abdn.ac.uk/~ijw/downloads/download.htm>

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

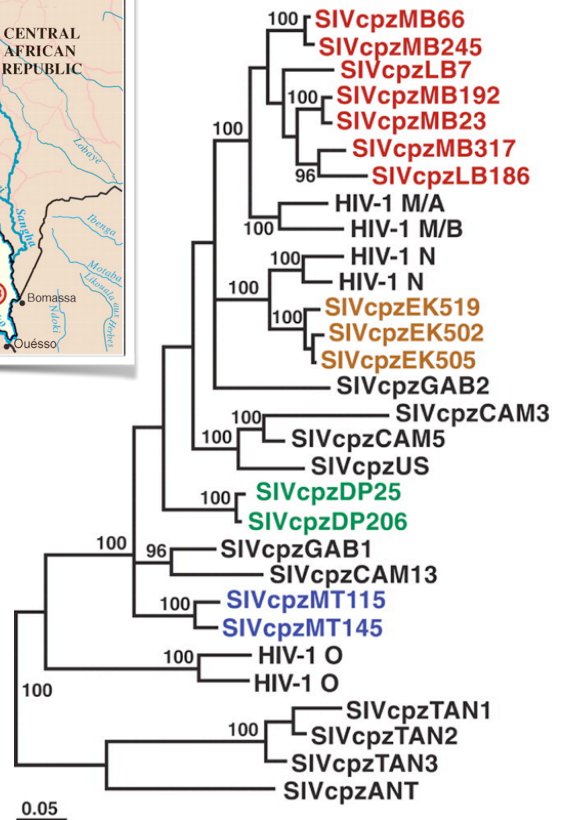
### Sequencing Production as of 2015-07-01



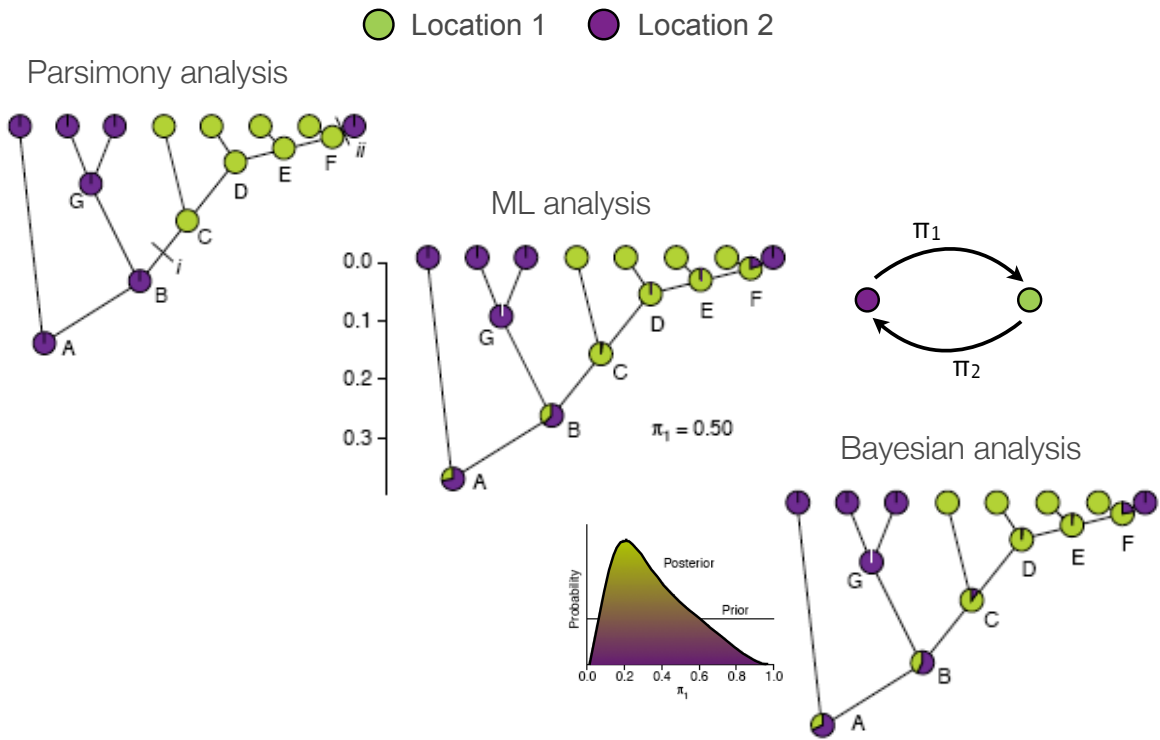
<http://gcid.jcvi.org/projects/msc/influenza/>



Keele et al., 2006, Science



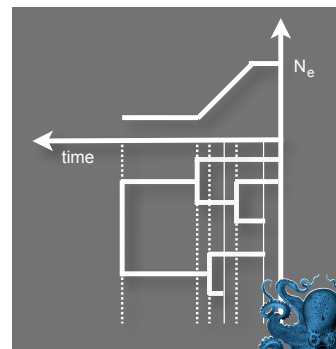
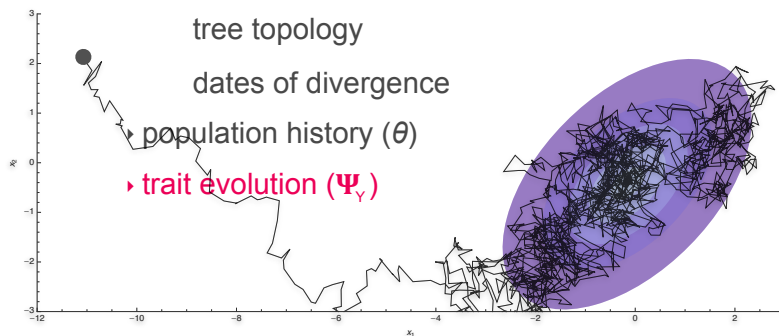
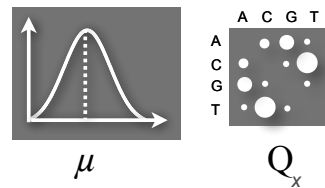
# Inferring discrete ancestral state locations



# Bayesian Evolutionary Analysis Sampling Trees

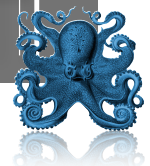
- Given sequence *and trait* data ( $X, Y$ ) that is temporally spaced estimate true values of:

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



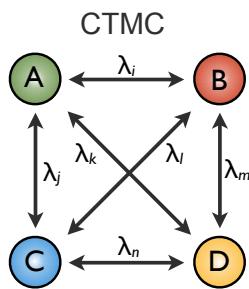
- Bayesian inference

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

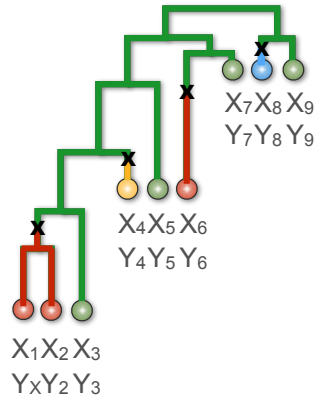


# Phylogenetic diffusion models

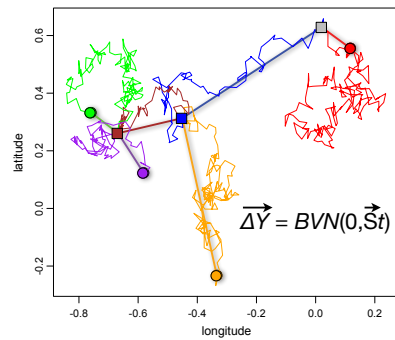
discrete



Lemey et al., PLoS Comp Bio, 2009



continuous

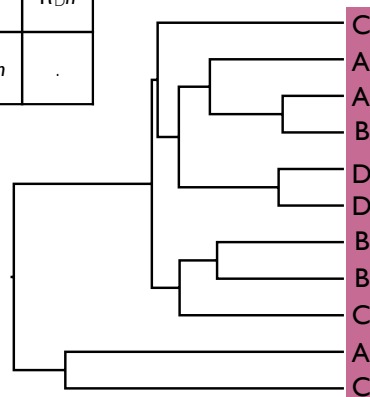
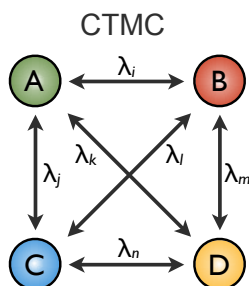


Lemey et al., MBE, 2010

# Phylogenetic diffusion models

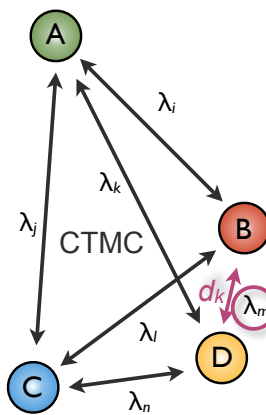
● Discrete Model:

	A	B	C	D
A	.	$\pi_{Bj}$	$\pi_{Cj}$	$\pi_{Dk}$
B	$\pi_{Ai}$	.	$\pi_{Cl}$	$\pi_{Dm}$
C	$\pi_{Aj}$	$\pi_{Bl}$	.	$\pi_{Dn}$
D	$\pi_{Ak}$	$\pi_{Bm}$	$\pi_{Cn}$	.



# Phylogenetic diffusion models

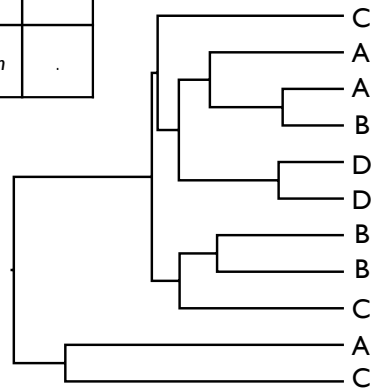
• priors



gamma( $\mu, \sigma$ )

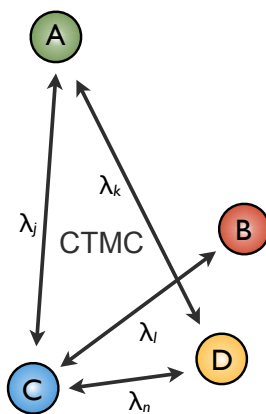
$\mu \sim 1/d_k$   
 $\sigma \sim 1/d_k$

	A	B	C	D
A	.	$\pi_{Bi}$	$\pi_{Cj}$	$\pi_{Dk}$
B	$\pi_{Ai}$	.	$\pi_{Cl}$	$\pi_{Dm}$
C	$\pi_{Aj}$	$\pi_{Bl}$	.	$\pi_{Dn}$
D	$\pi_{Ak}$	$\pi_{Bm}$	$\pi_{Cn}$	.

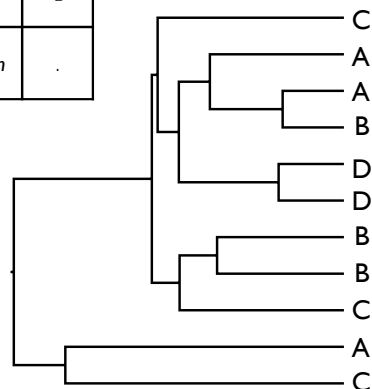


# Phylogenetic diffusion models

• Do we need all those parameters?



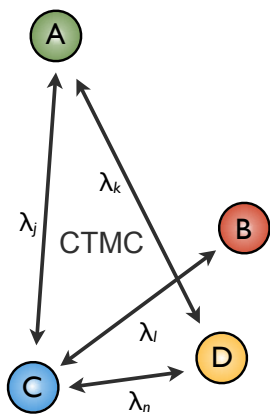
	A	B	C	D
A	.	<del><math>\pi_{Bi}</math></del>	$\pi_{Cj}$	$\pi_{Dk}$
B	<del><math>\pi_{Ai}</math></del>	.	$\pi_{Cl}$	<del><math>\pi_{Dm}</math></del>
C	$\pi_{Aj}$	$\pi_{Bl}$	.	$\pi_{Dn}$
D	$\pi_{Ak}$	<del><math>\pi_{Bm}</math></del>	$\pi_{Cn}$	.





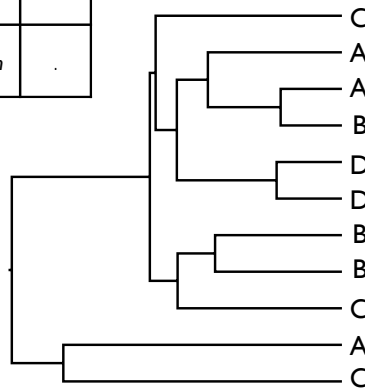
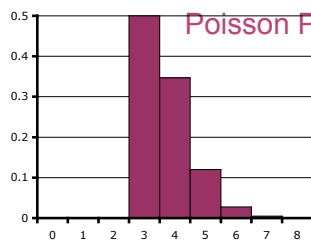
# Phylogenetic diffusion models

- Do we need all those parameters?



	A	B	C	D
A	.	$\pi_{B A} \lambda_j$	$\pi_{C A} \lambda_k$	$\pi_{D A} \lambda_l$
B	$\pi_{A B} \lambda_j$	.	$\pi_{C B} \lambda_l$	$\pi_{D B} \lambda_m$
C	$\pi_{A C} \lambda_j$	$\pi_{B C} \lambda_l$	.	$\pi_{D C} \lambda_n$
D	$\pi_{A D} \lambda_k$	$\pi_{B D} \lambda_m$	$\pi_{C D} \lambda_n$	.

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



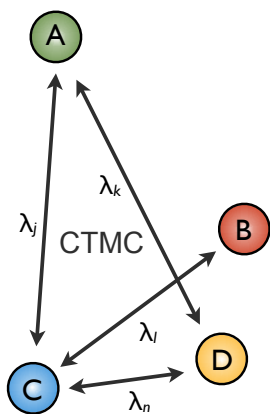
offset =  $K - 1$   
 mean =  $\log(2)$

$$\sum_j I_j$$

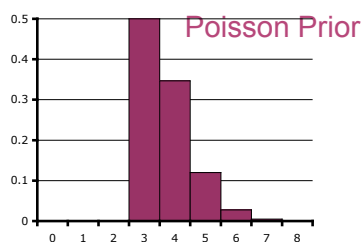
# Phylogenetic diffusion models

- 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'

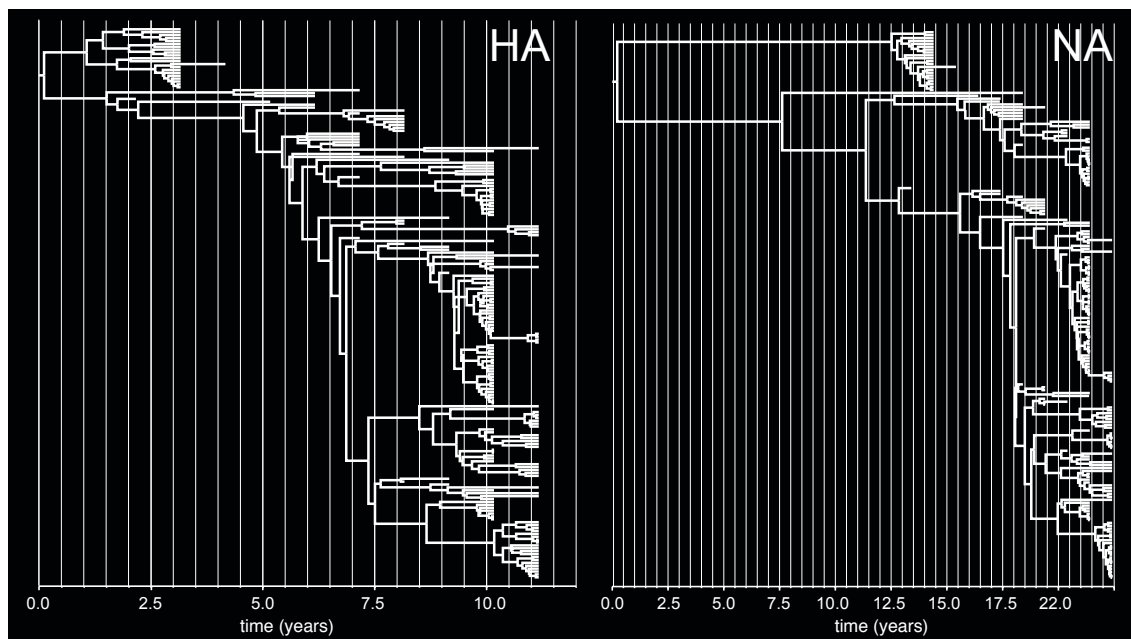
---

- 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



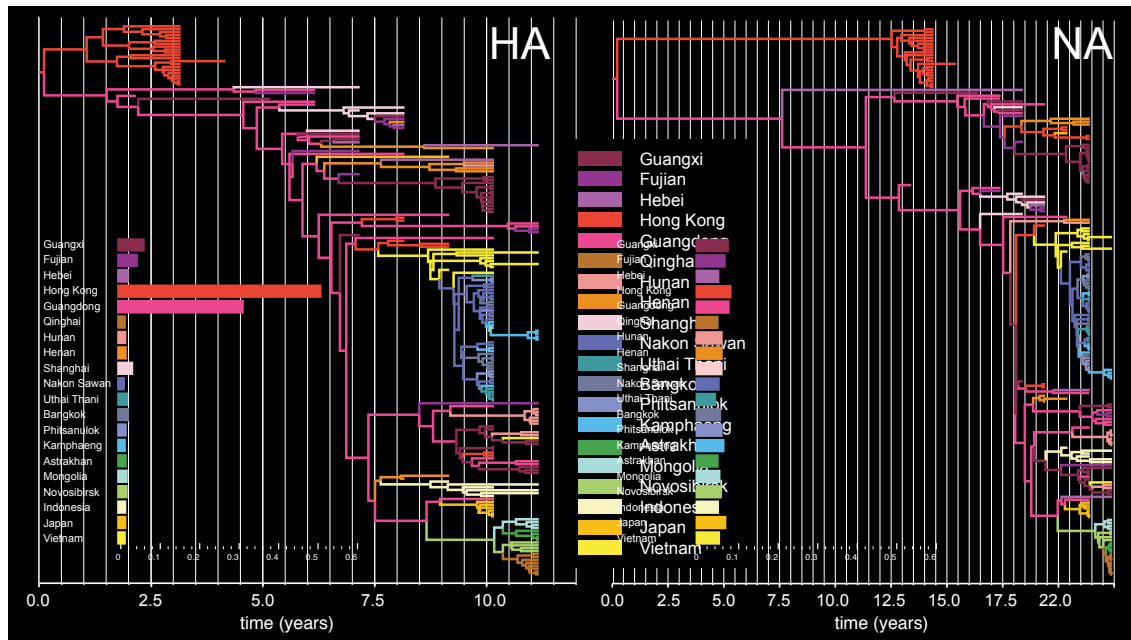
## Influenza A H5N1

---

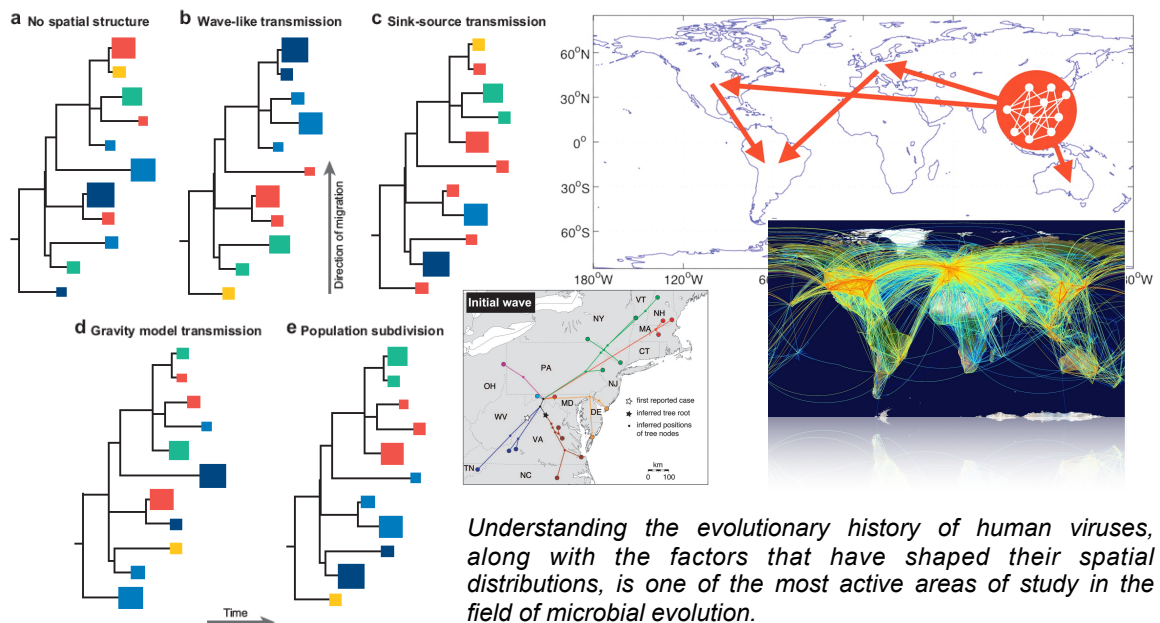


Wallace et al., PNAS, 2007

# Influenza A H5N1: discrete model



# Phylogeographic patterns in RNA viruses

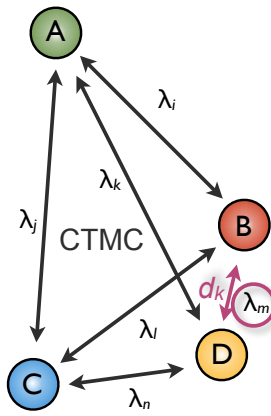


*Understanding the evolutionary history of human viruses, along with the factors that have shaped their spatial distributions, is one of the most active areas of study in the field of microbial evolution.*

*Eddie Holmes, Ann Rev Microbiol 2008*

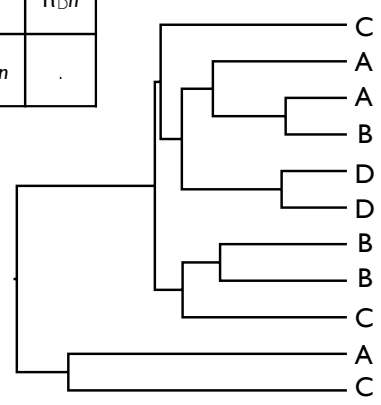
# Phylogeographic hypothesis testing?

• priors



	A	B	C	D
A	.	$\pi_{Bj}$	$\pi_{Cj}$	$\pi_{Dk}$
B	$\pi_{Ai}$	.	$\pi_{Cl}$	$\pi_{Dm}$
C	$\pi_{Aj}$	$\pi_{Bl}$	.	$\pi_{Dn}$
D	$\pi_{Ak}$	$\pi_{Bm}$	$\pi_{Cn}$	.

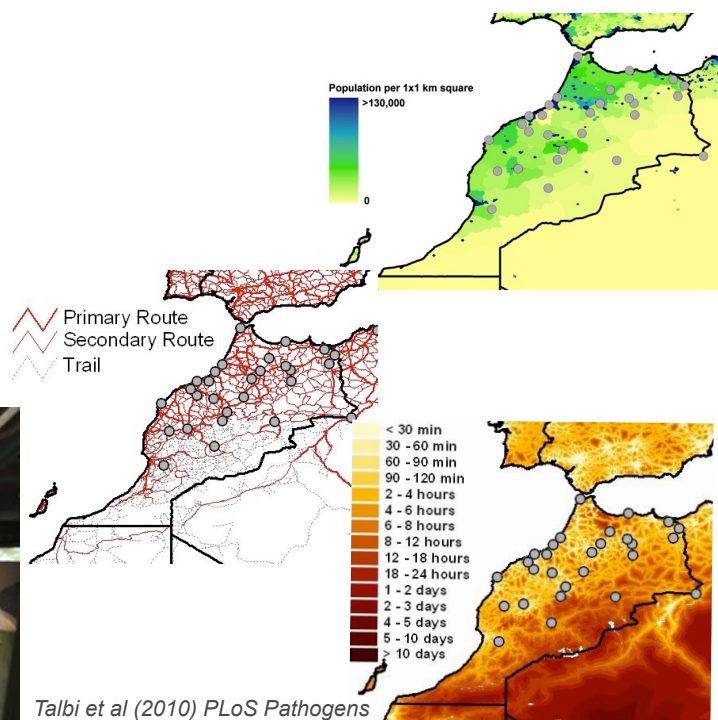
$\text{gamma}(\mu, \sigma)$   
 $\mu \sim 1/d_k$   
 $\sigma \sim 1/d_k$



# Predictors of dog rabies diffusion in Morocco

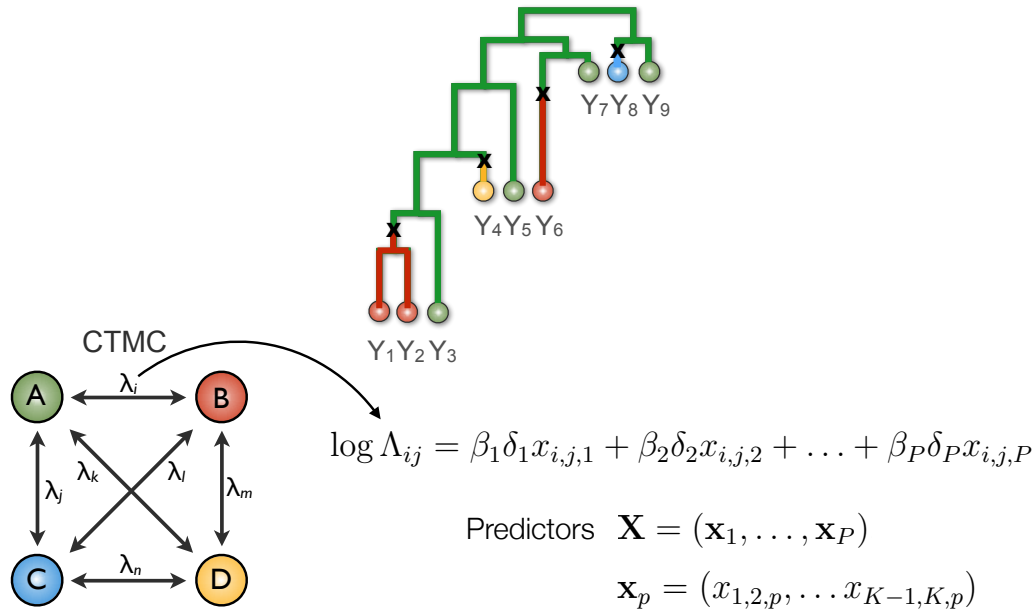
Predictor	Ln Marginal likelihood
Equal rates	-320
Distance	-299.8
Population sizes	-381
Gravity model	-388.6
Population surface	-335.9
Road distances	-298.5

Accessibility



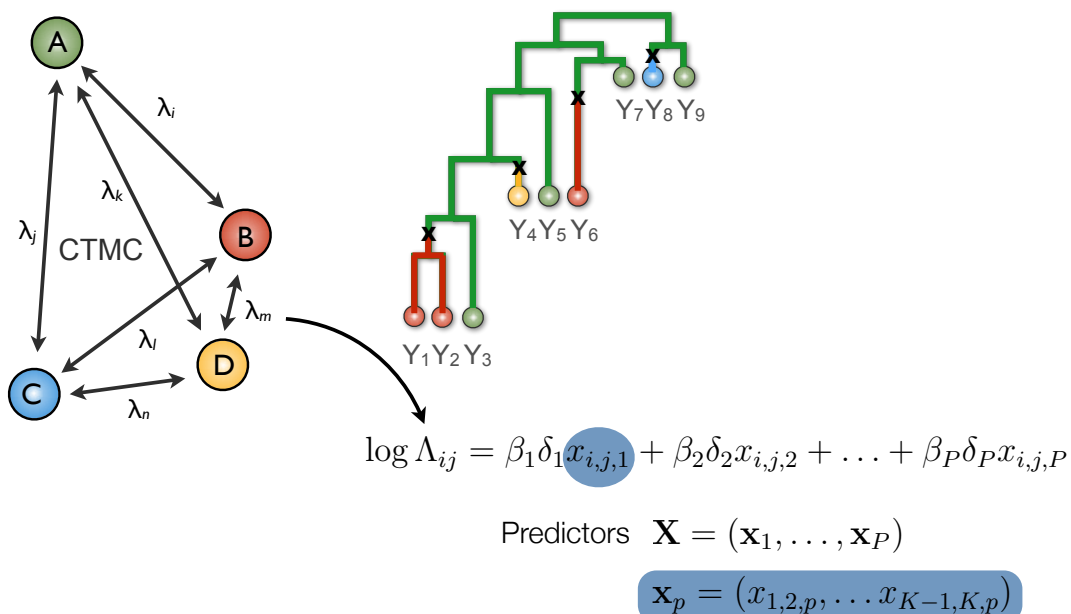
Talbi et al (2010) PLoS Pathogens

## Predictors of phylogenetic diffusion



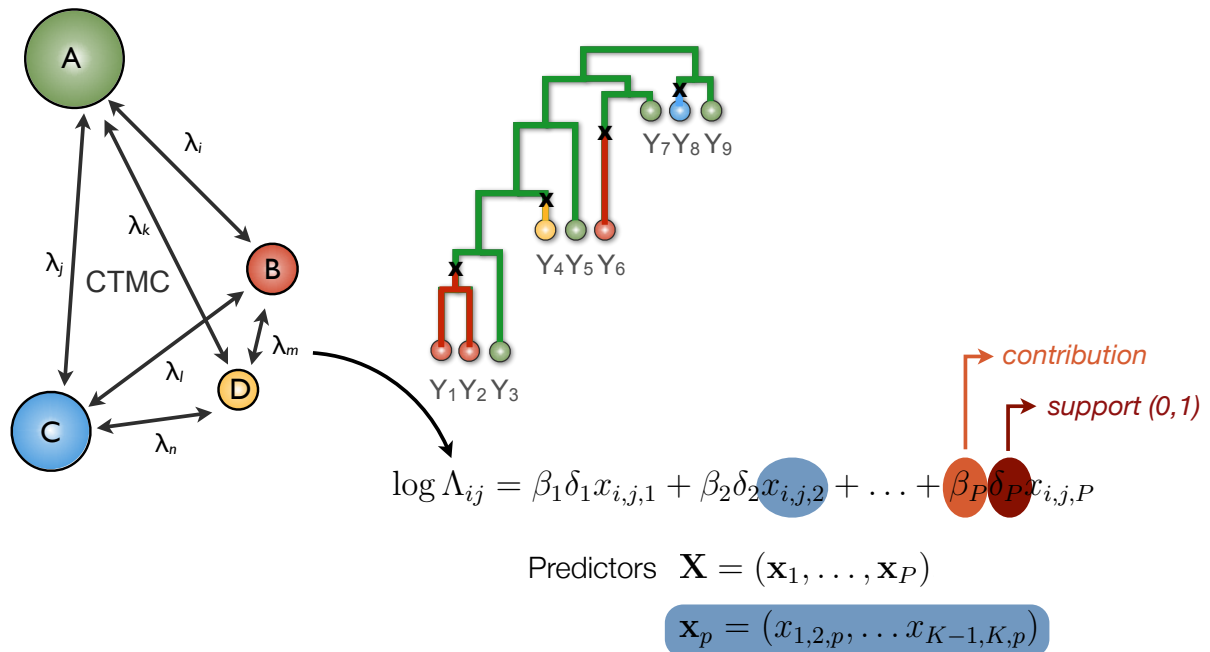
Lemey et al., PLoS Path, 2014

## Predictors of phylogenetic diffusion



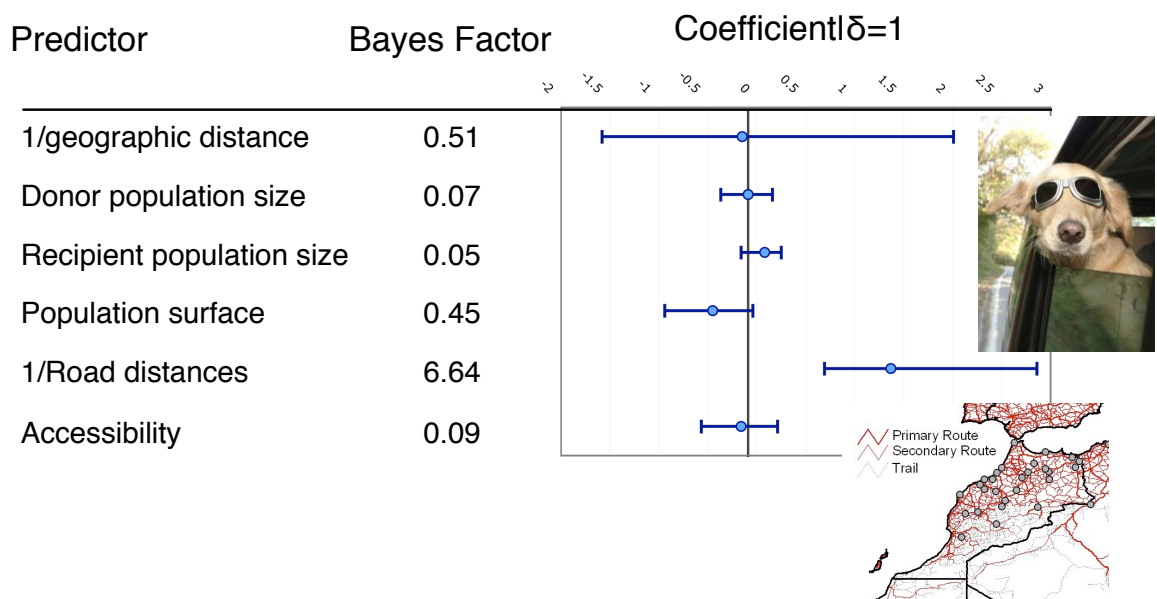
Lemey et al., PLoS Path, 2014

# Predictors of phylogenetic diffusion

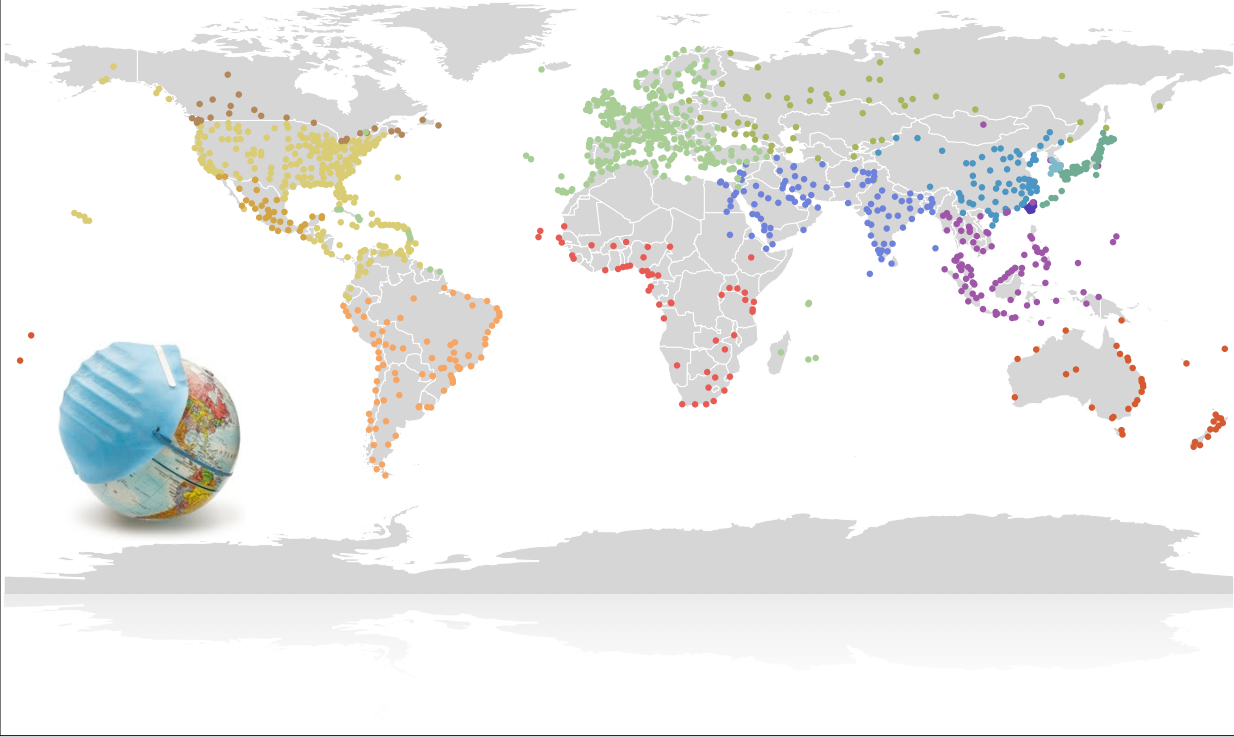


Lemey et al., PLoS Path, 2014

# 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

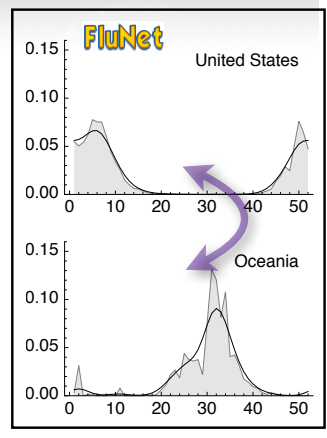
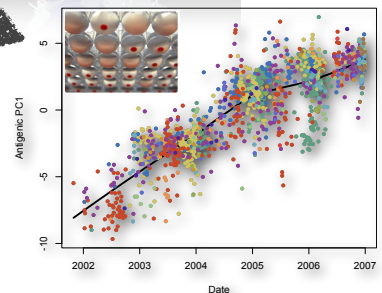
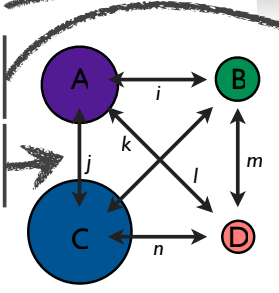
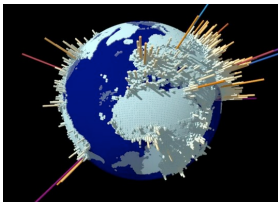
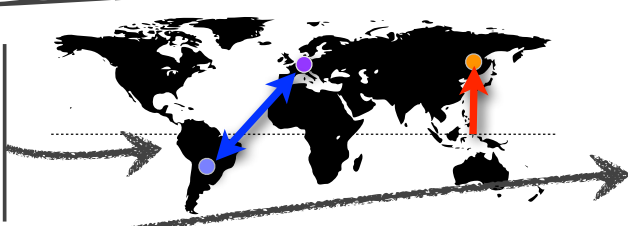
H3 incidence overlap

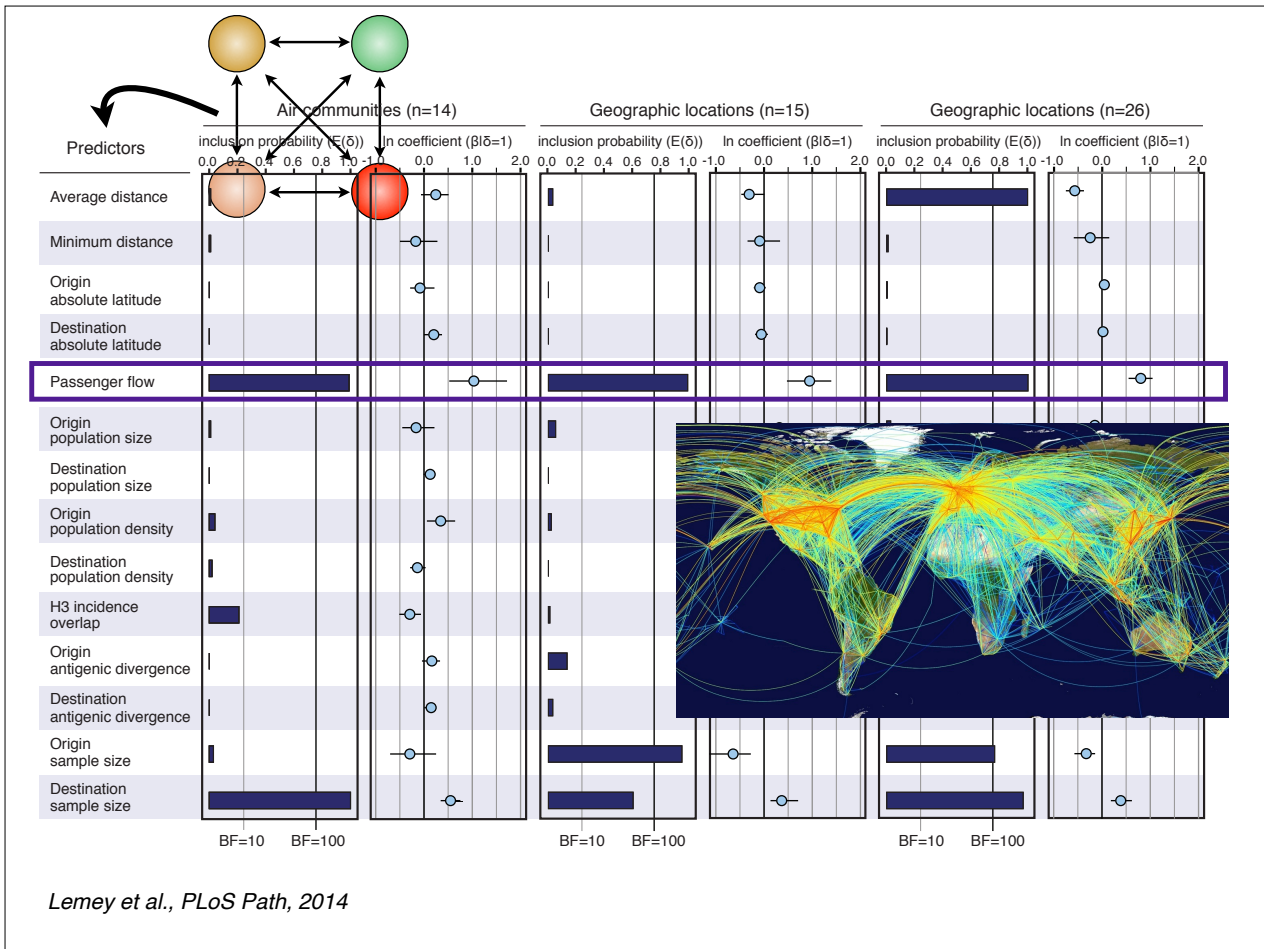
Origin antigenic divergence

Destination antigenic divergence

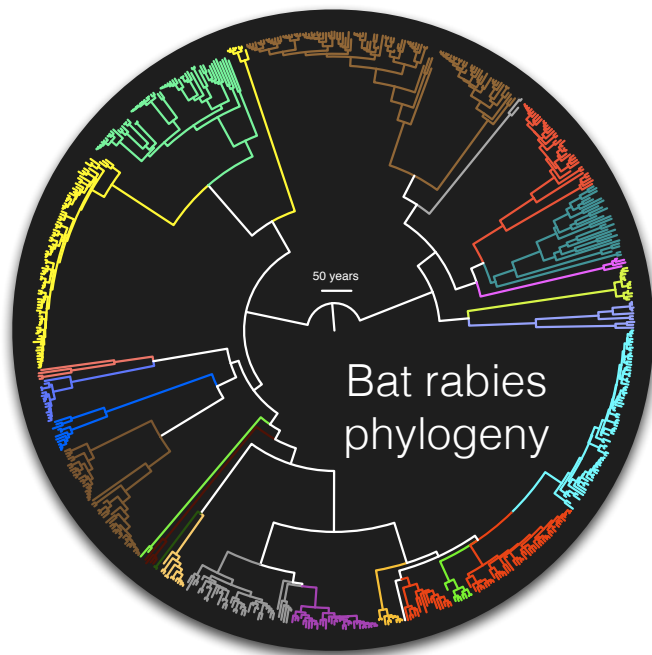
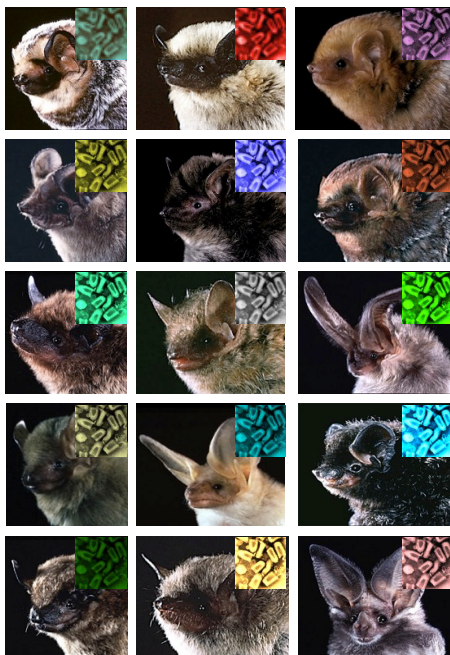
Origin sample size

Destination sample size





## Uncovering cross-species dynamics bat rabies

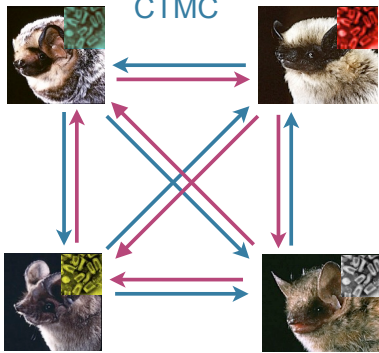


Courtesy of D. Streicker



# Uncovering bat rabies transmission dynamics

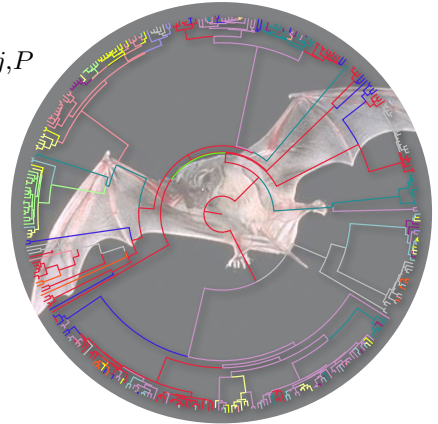
$$\log \Lambda_{ij} = \beta_1 \delta_1 x_{i,j,1} + \beta_2 \delta_2 x_{i,j,2} + \dots + \beta_P \delta_P x_{i,j,P}$$



contribution  
inclusion probability

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

- host divergence
- range overlap
- roost structure overlap
- $\Delta$  wing aspect ratio
- $\Delta$  wing loading
- $\Delta$  body size



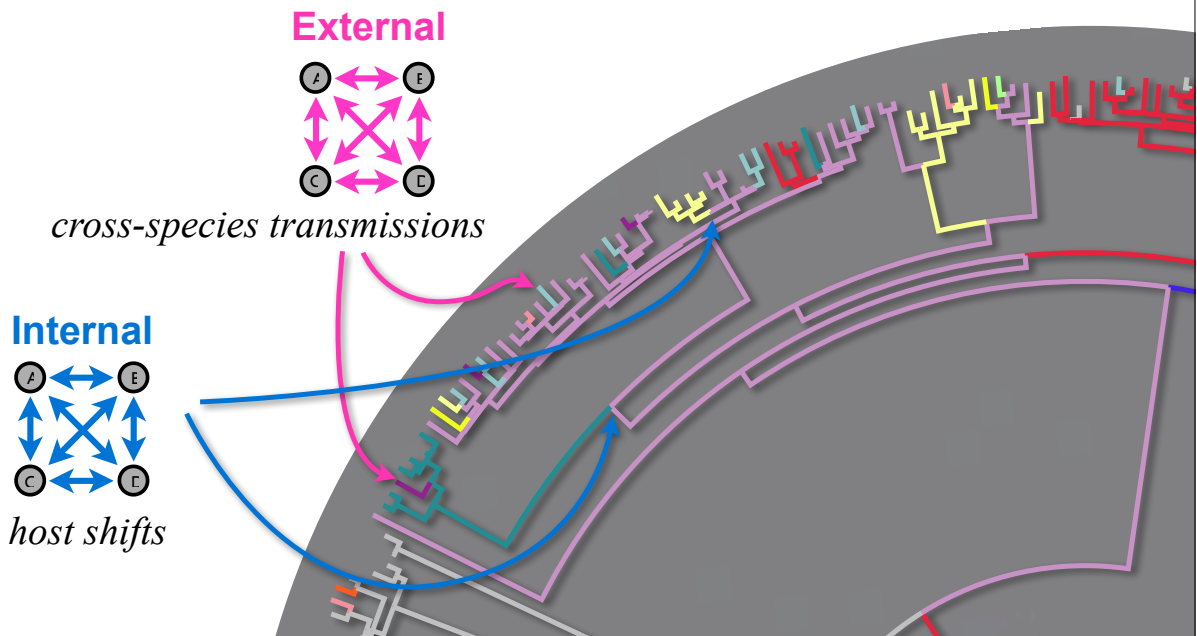
- 372 partial N genes (1997-2005)
- 17 different bat hosts
- 14 US states

Streicker et al., Science, 2010.

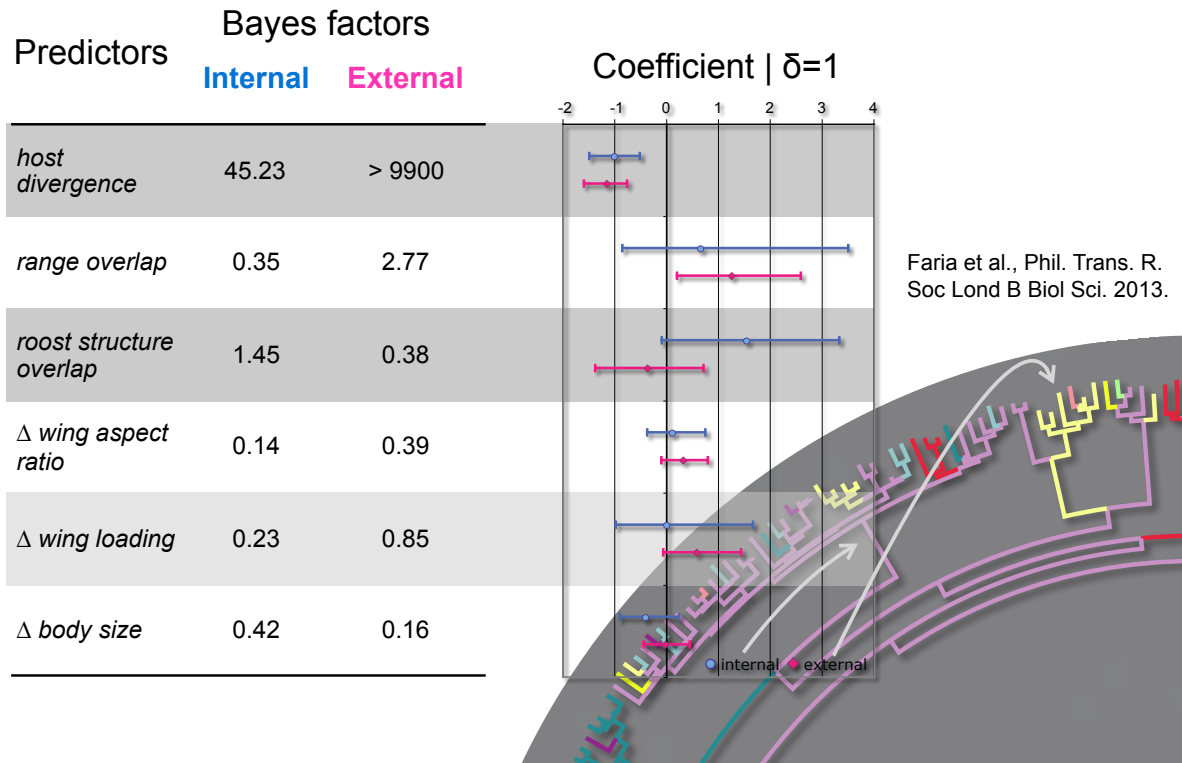
Faria et al., Phil. Trans. R. Soc Lond B Biol Sci. 2013.

# Uncovering bat rabies transmission dynamics

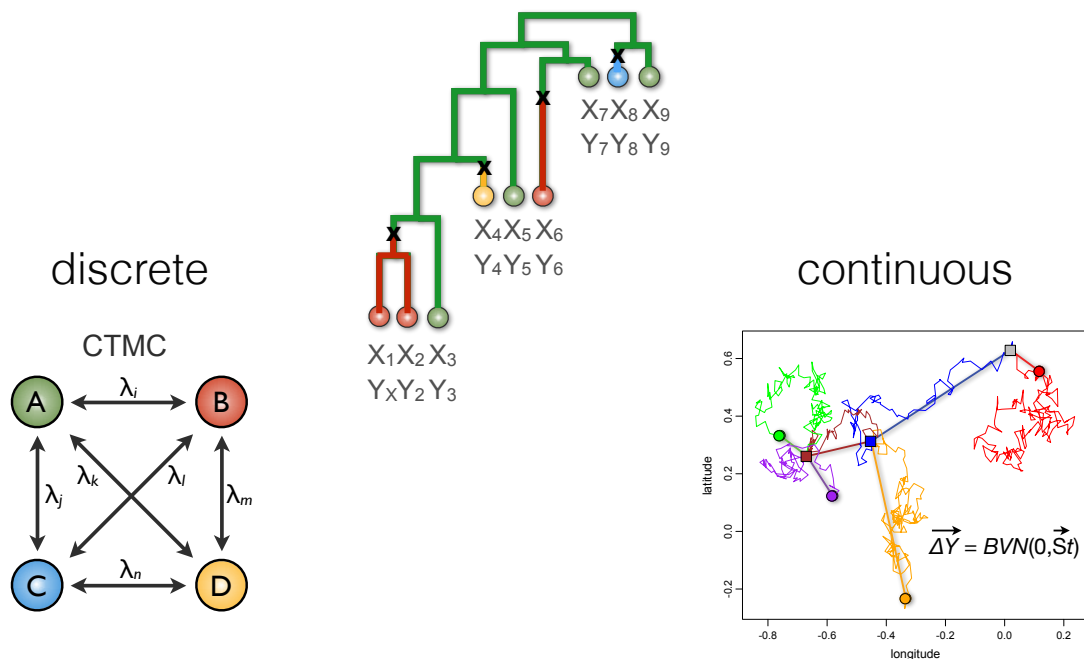
Faria et al., Phil. Trans. R. Soc Lond B Biol Sci. 2013.



# Uncovering bat rabies transmission dynamics



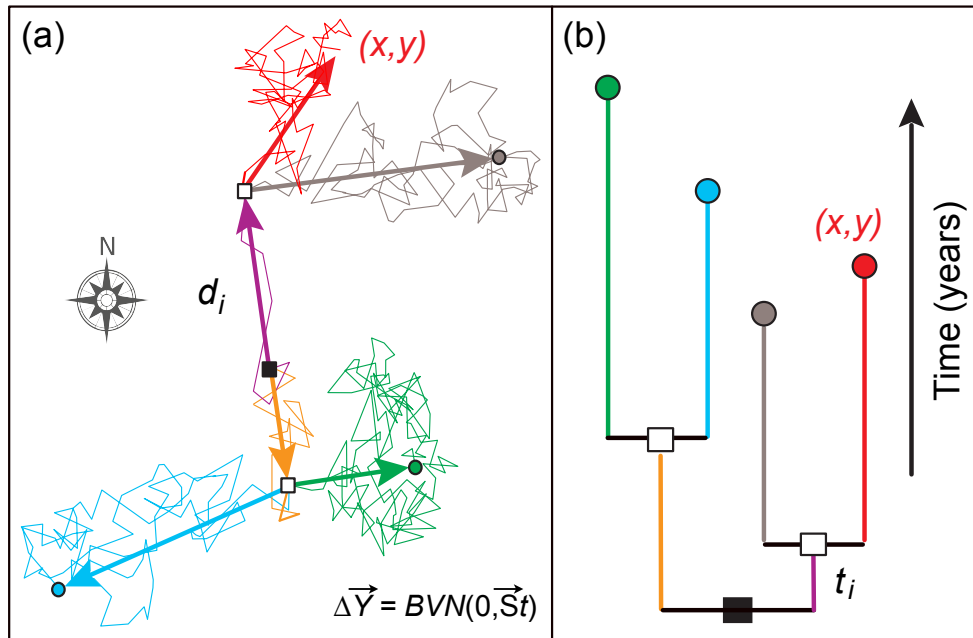
# Phylogenetic diffusion models



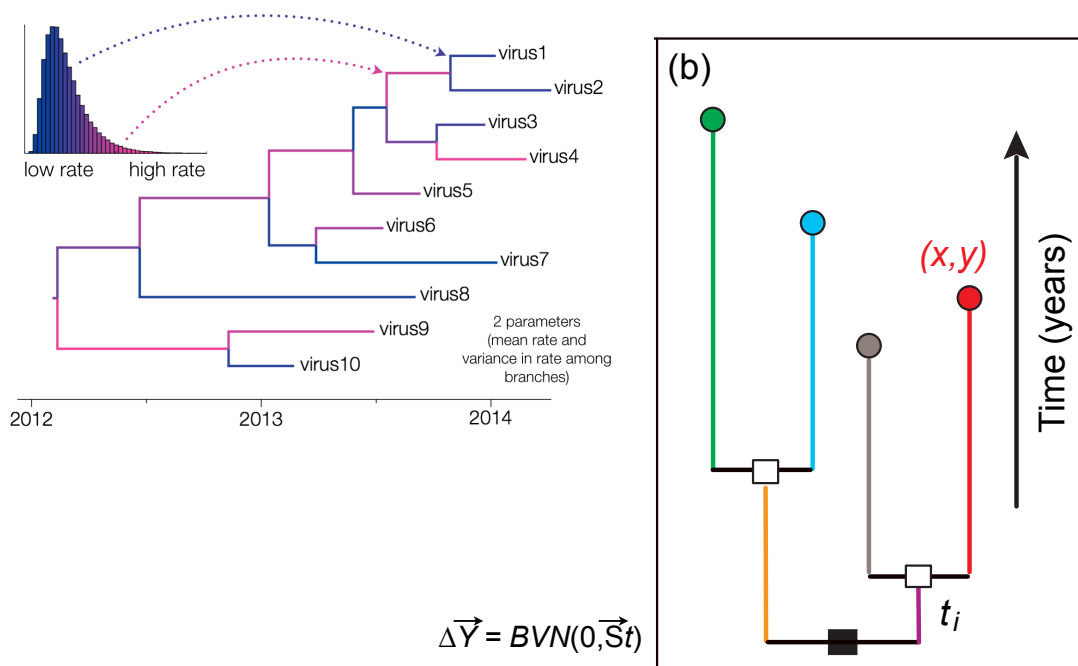
Lemey et al., PLoS Comp Bio, 2009

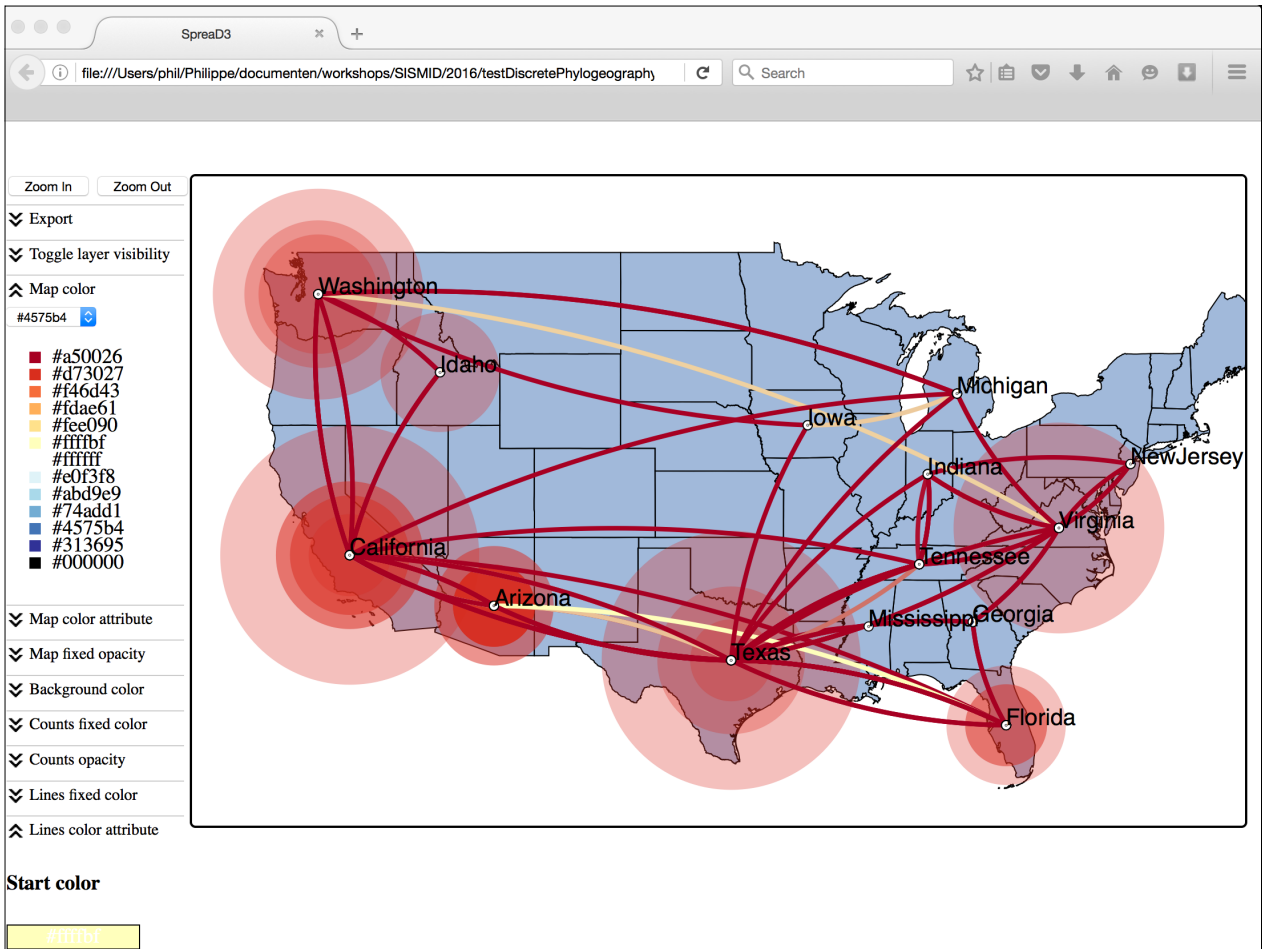
Lemey et al., MBE, 2010

# Continuous phylogenetic diffusion

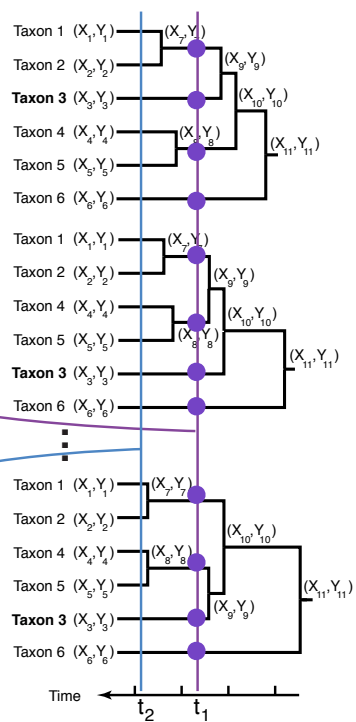
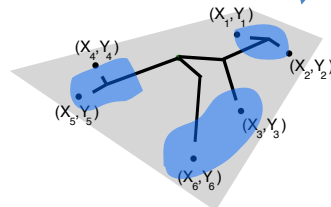
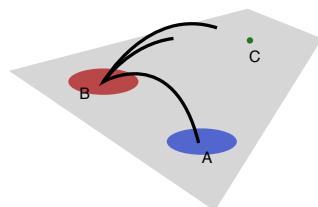
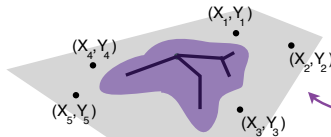
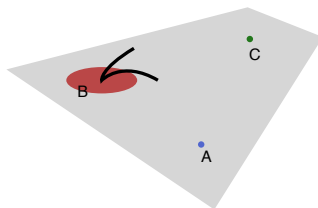
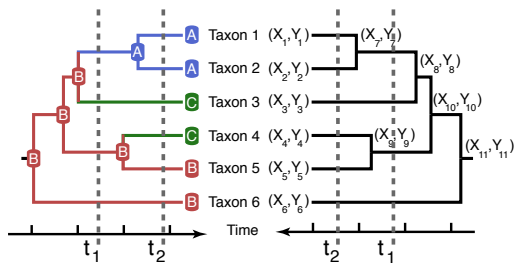


# Relaxed random walk

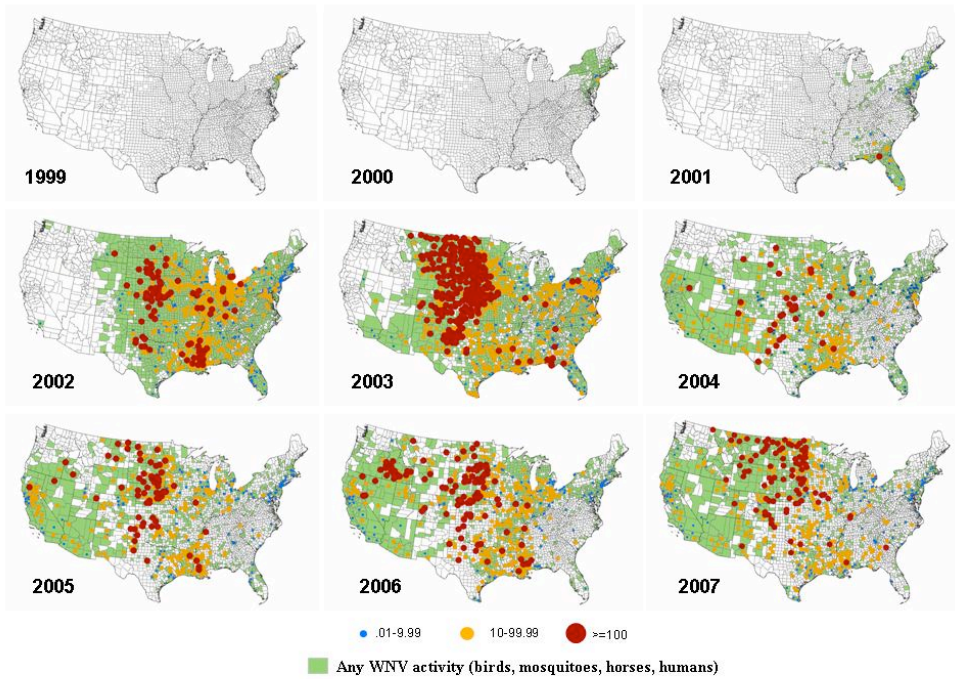


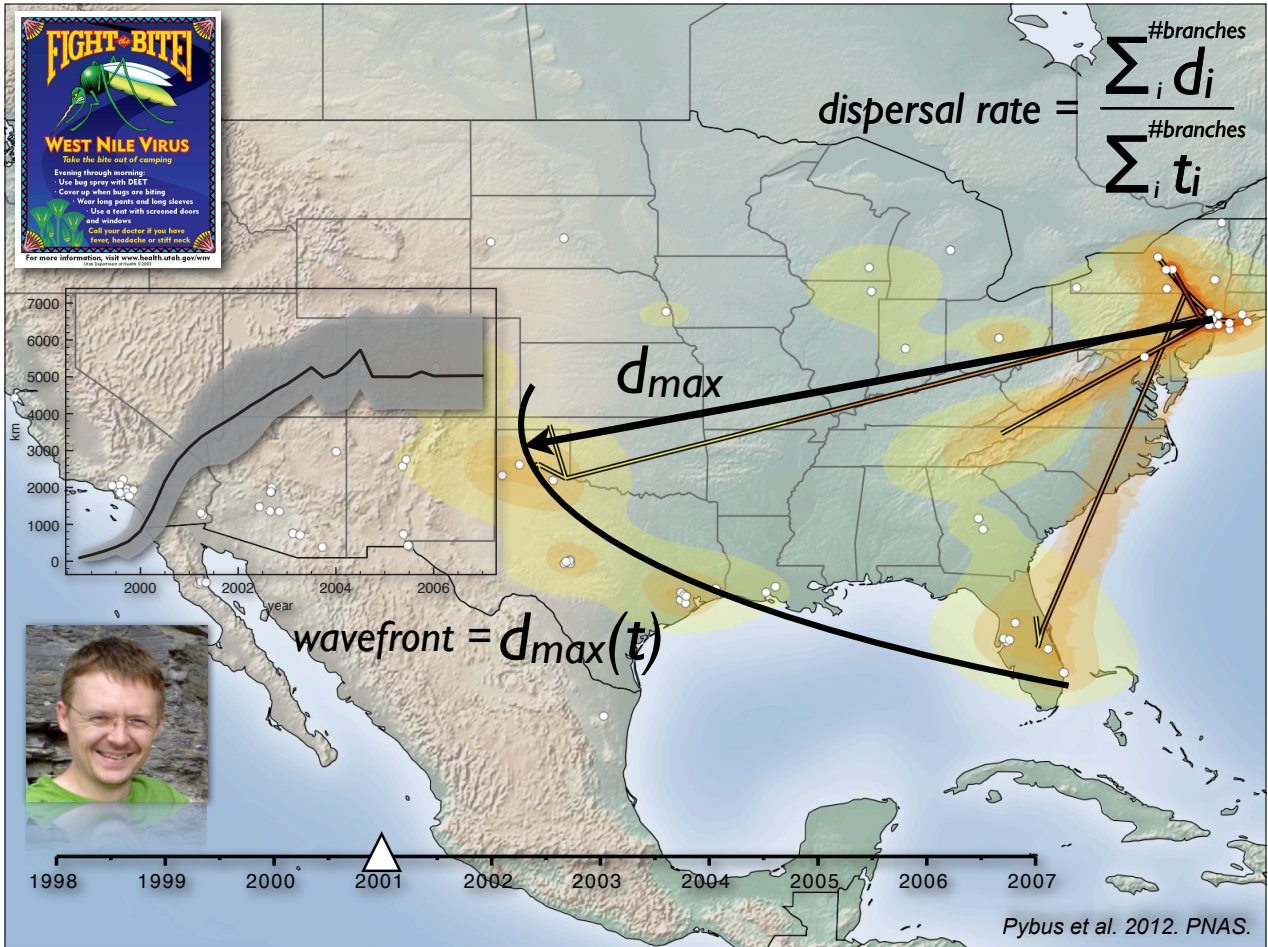


# visualization

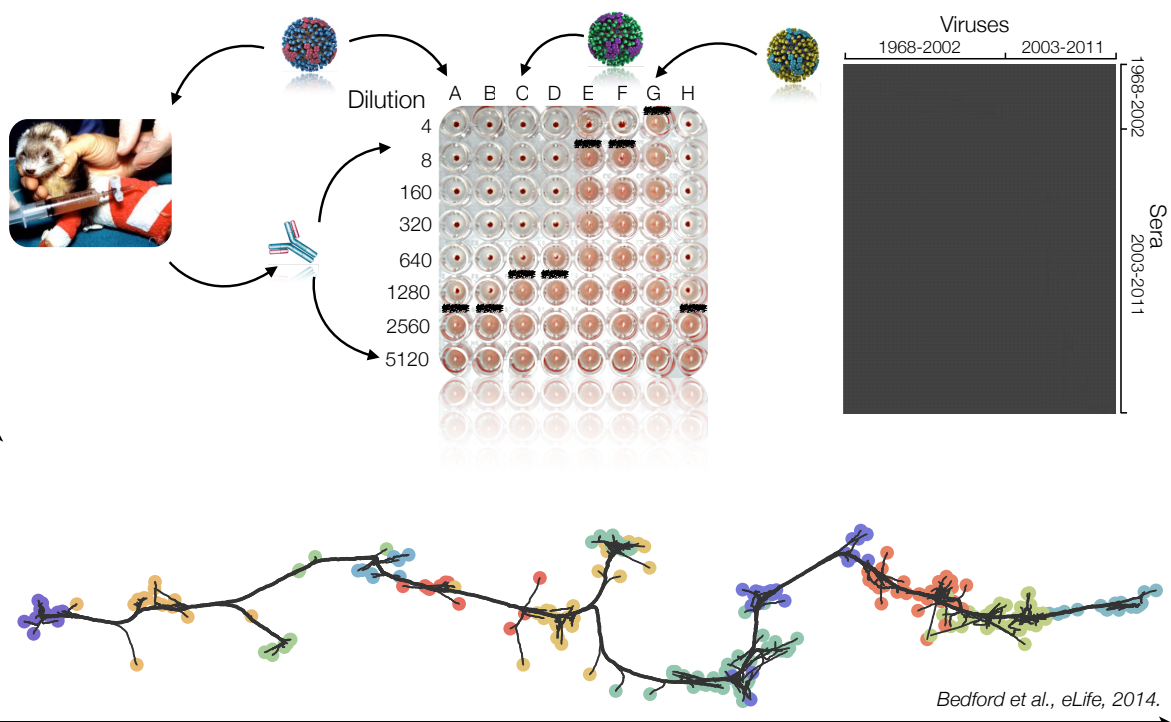


# West Nile virus invasion in North America

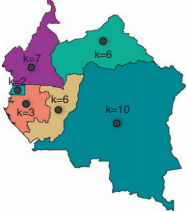
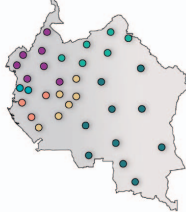


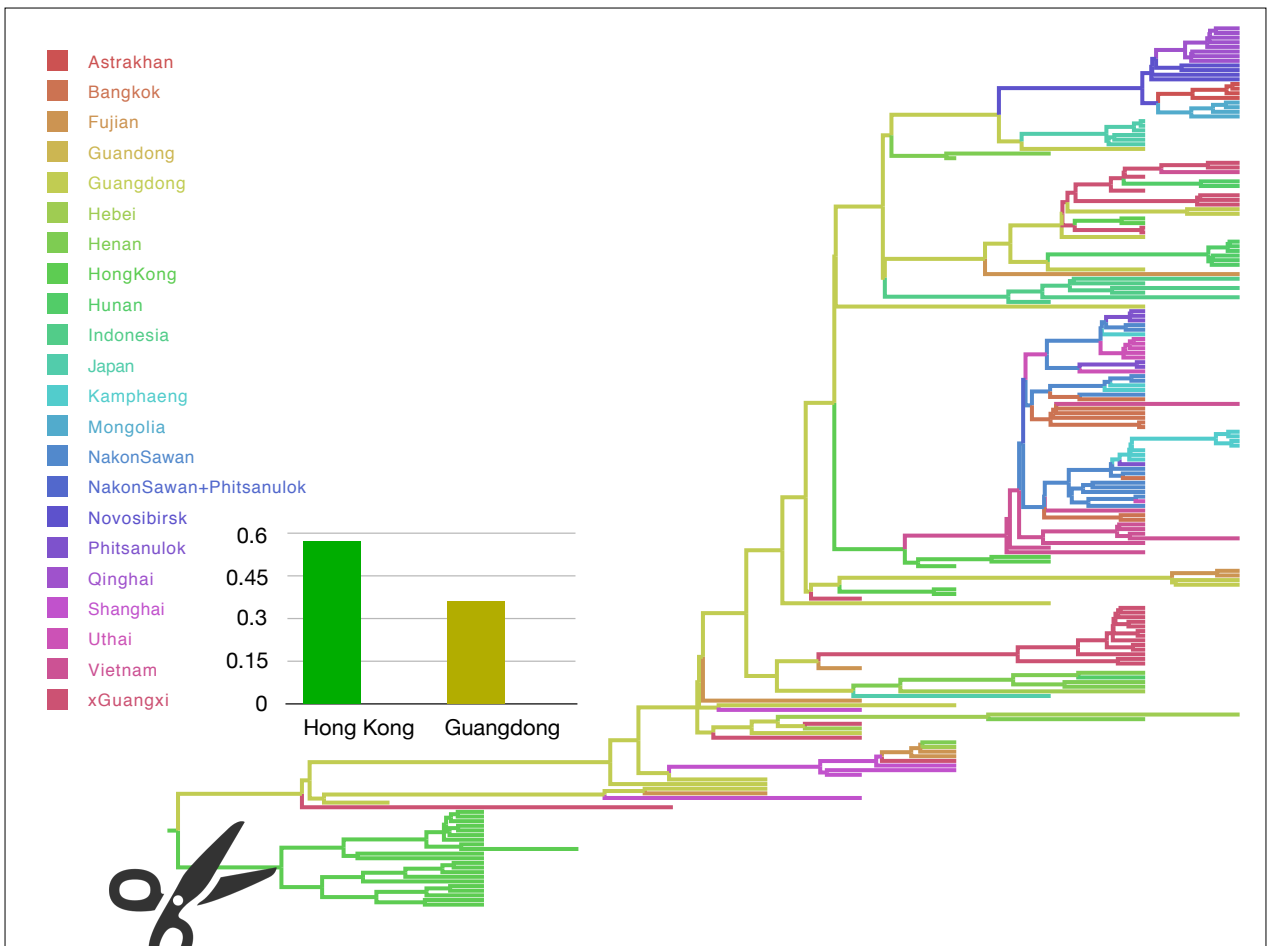


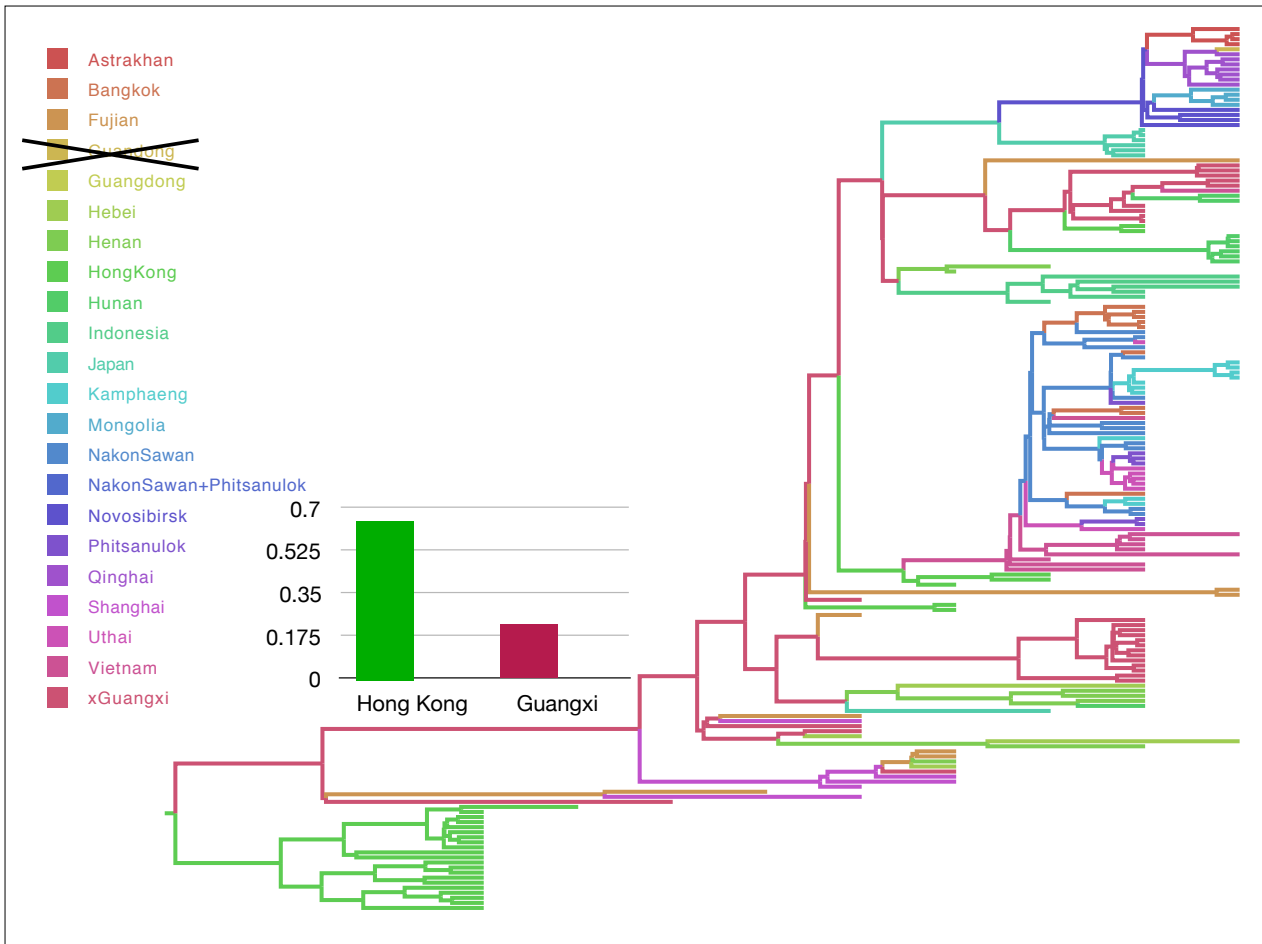
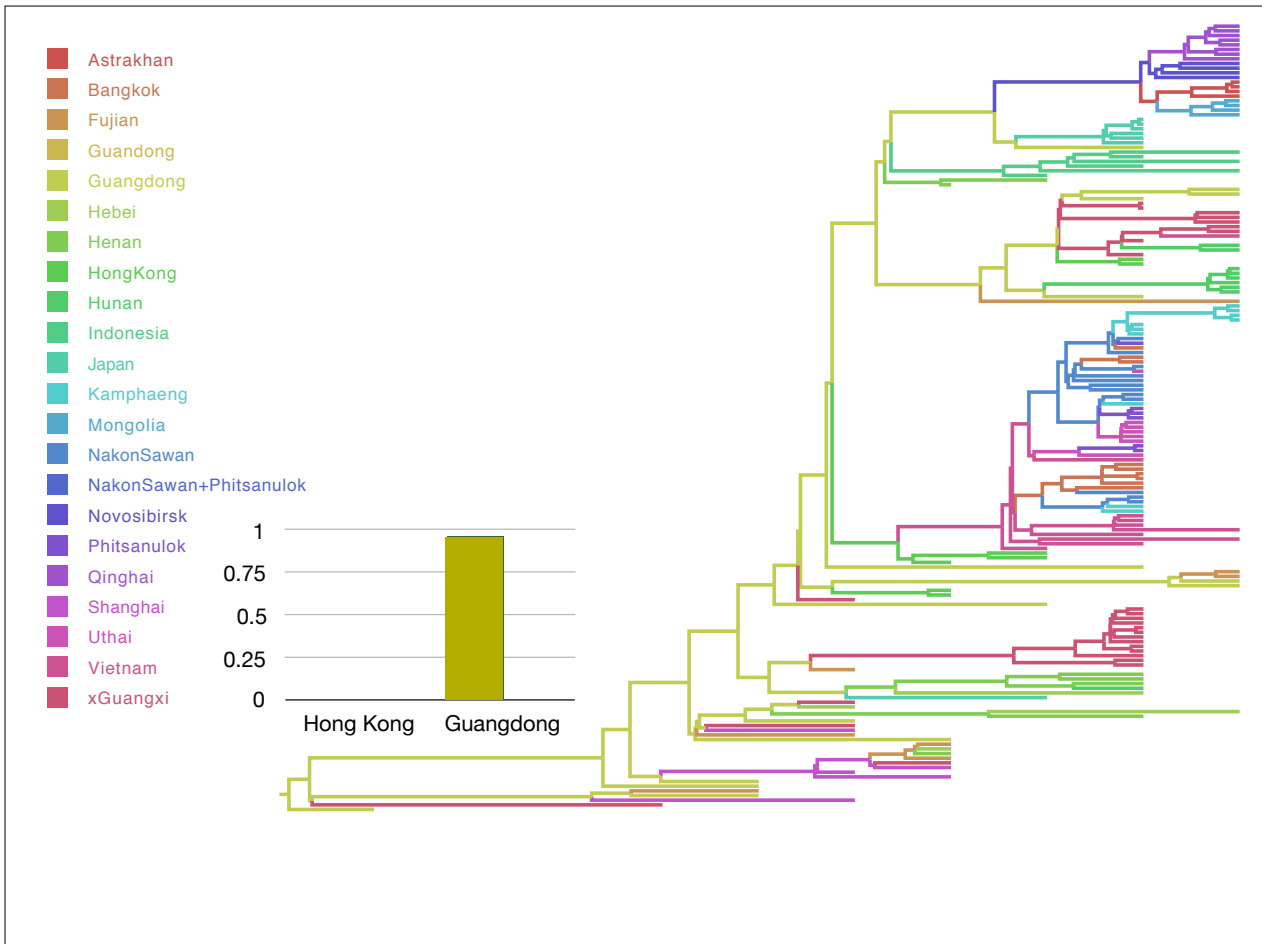
## Phenotypic evolution: antigenic flux



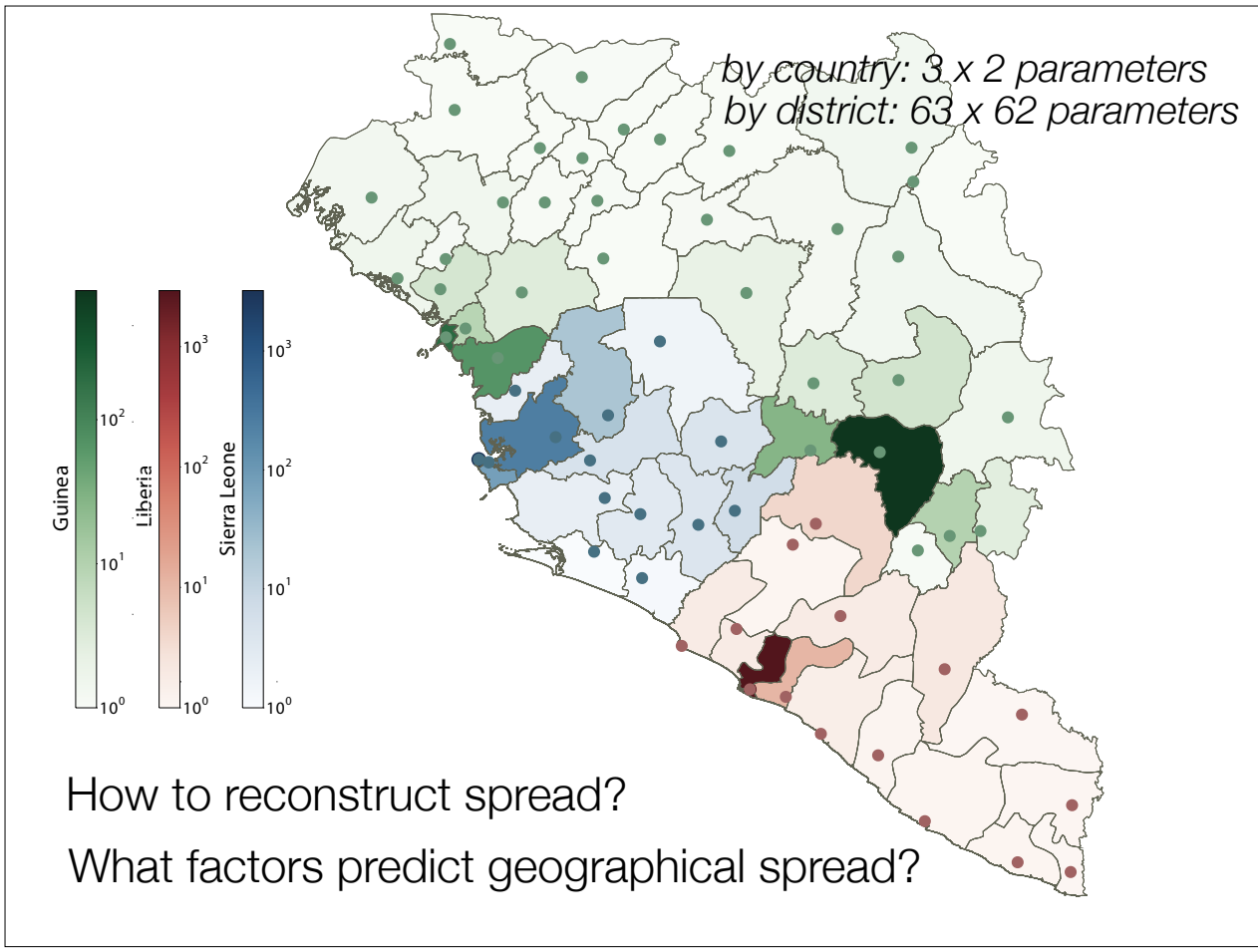
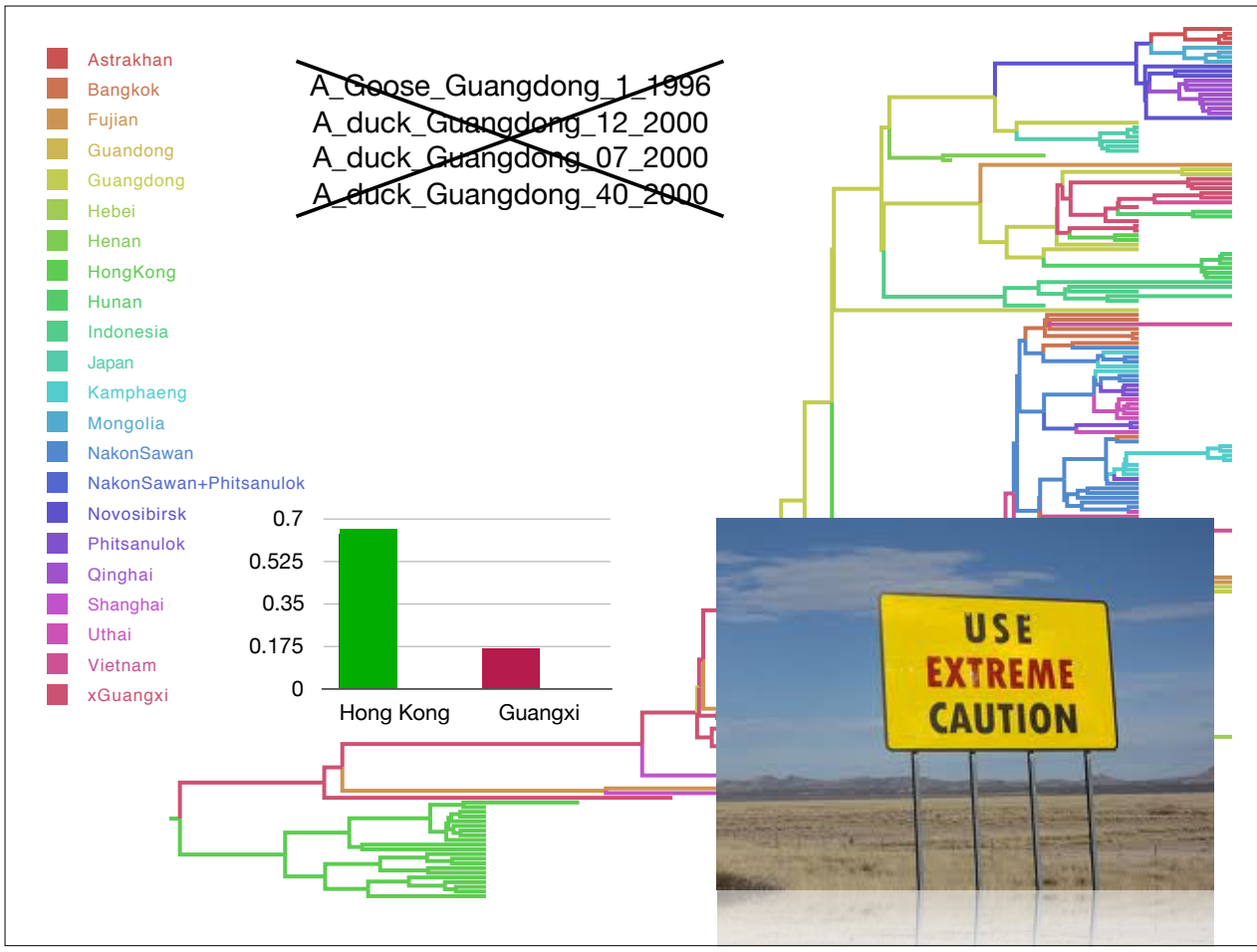
# Phylogeography: discrete or continuous?

	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

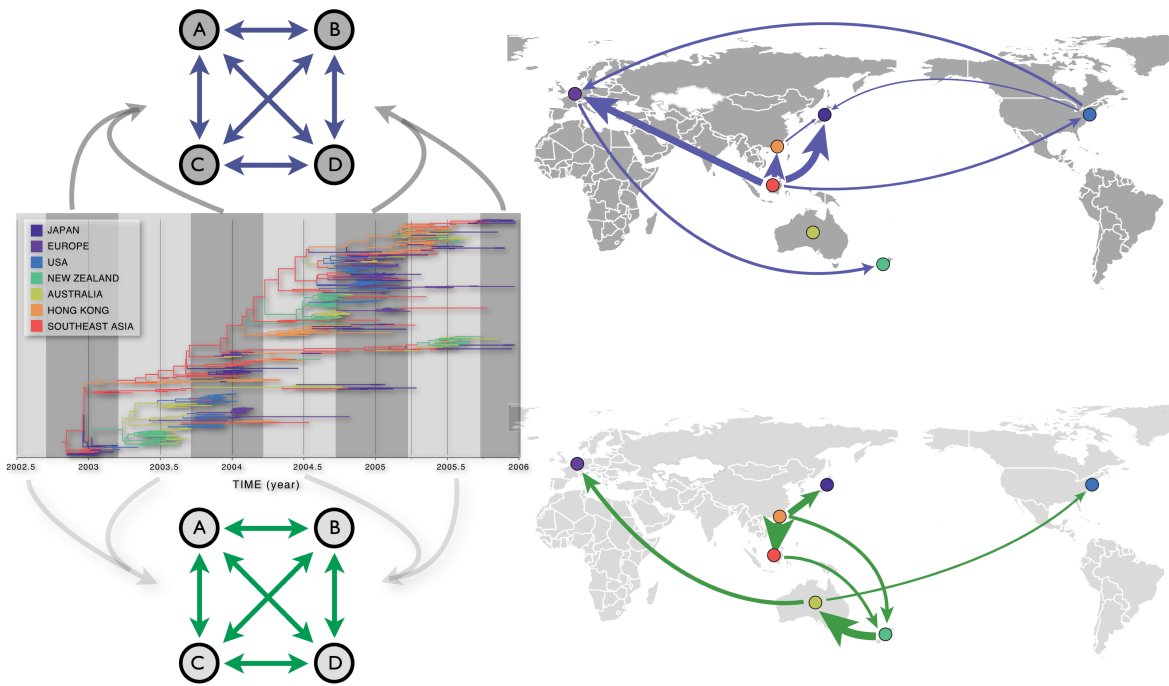








# Temporal heterogeneity

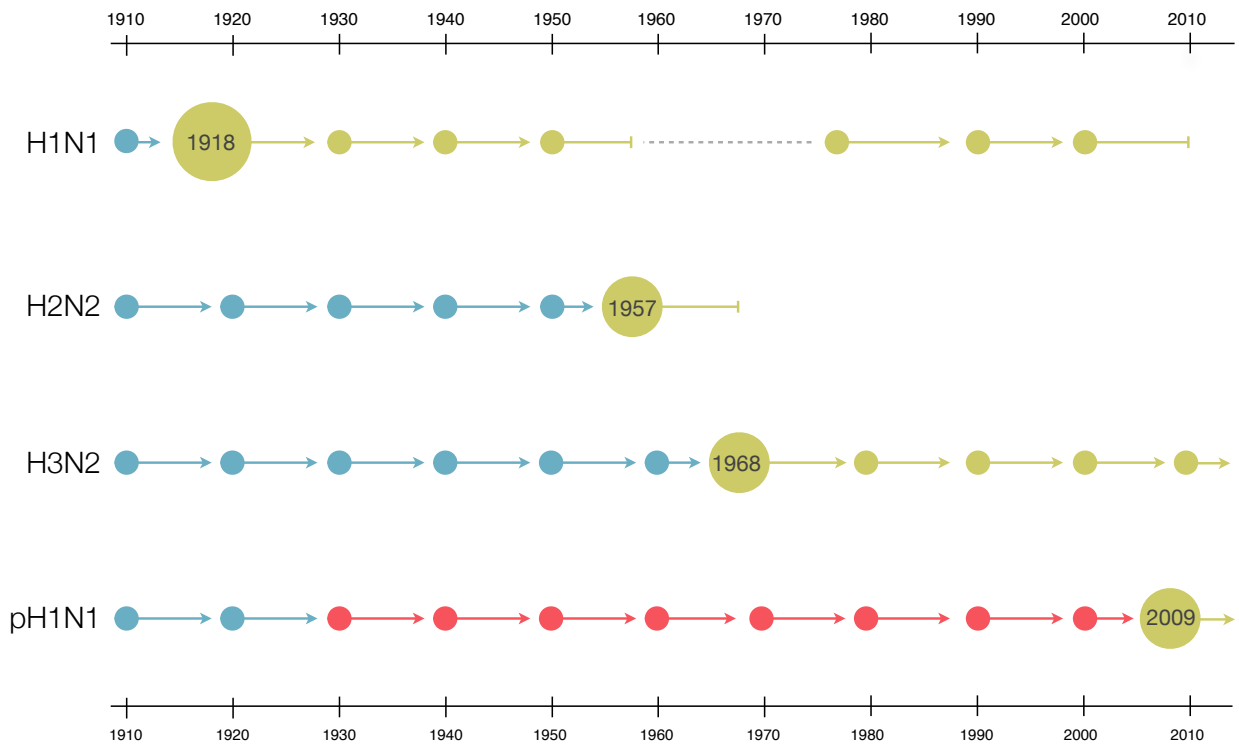


Bielejec et al (2014) Sys Bio

# Influenza A pandemics caused by host switch events

Avian  
Swine  
Human

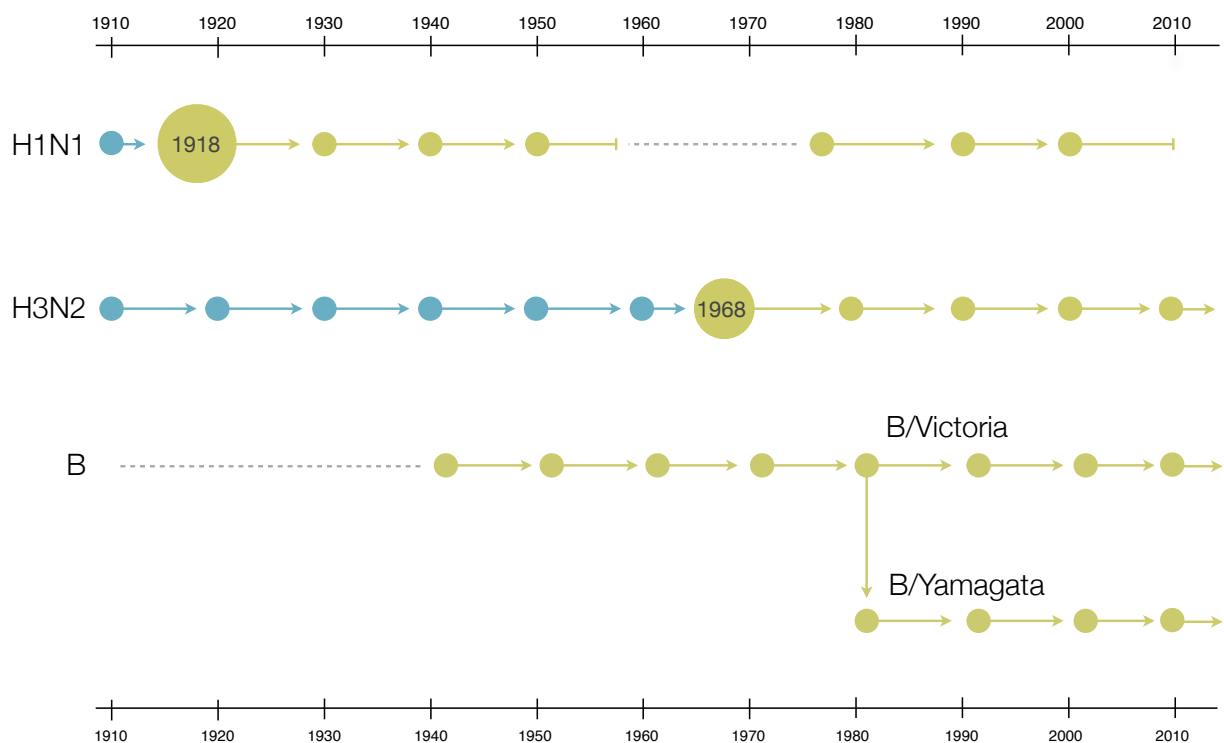
Hemagglutinin (HA) lineage



# Influenza B does not have pandemic potential

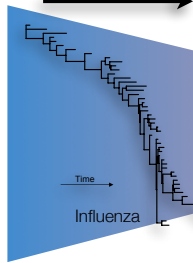
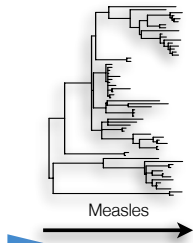
Avian  
Swine  
Human

Hemagglutinin (HA) lineage

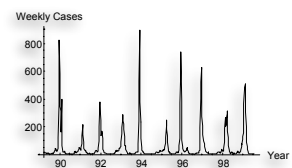
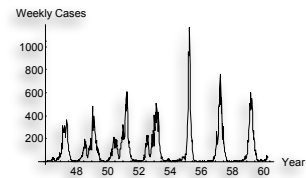


# Phylodynamics™

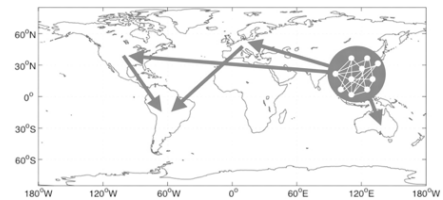
GENETIC DIVERSITY  
(phylogenetics &  
molecular evolution)



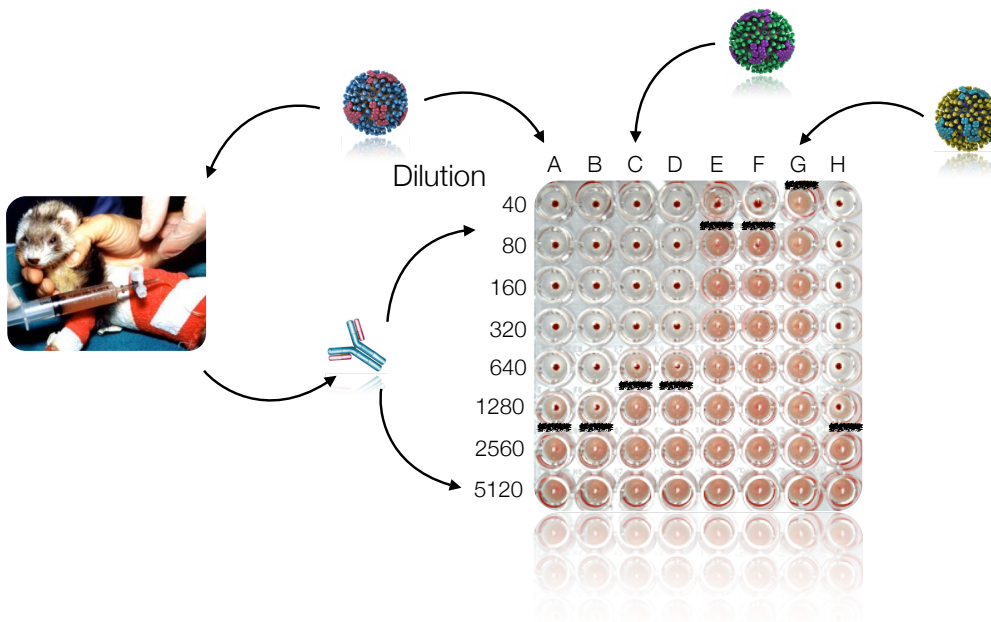
EPIDEMIC DYNAMICS  
(mathematical epidemiology)



NATURAL SELECTION  
(population genetics &  
immunology)



# Influenza hemagglutination inhibition (HI) assay



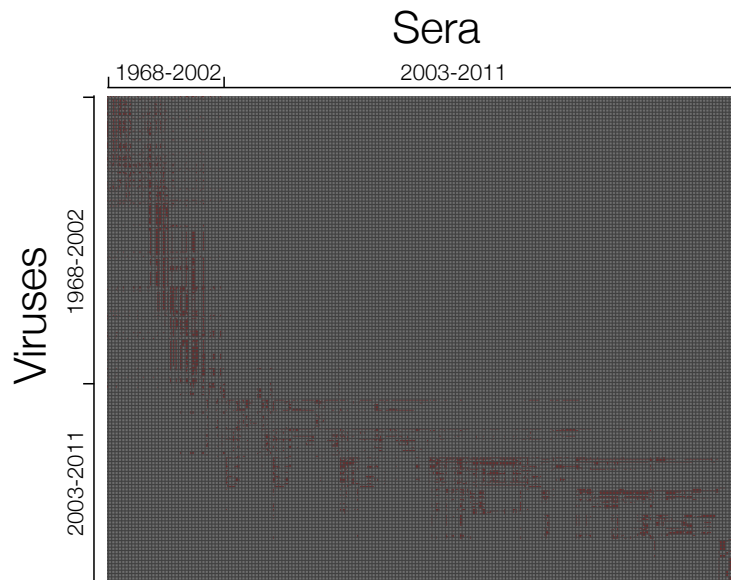
## Data takes the form of a table of cross-reactivities

Influenza A (H3N2) from 1968 to 2011:

340 virus isolates  
438 ferret antisera  
338 strains  
7232 HI titers

Challenges:

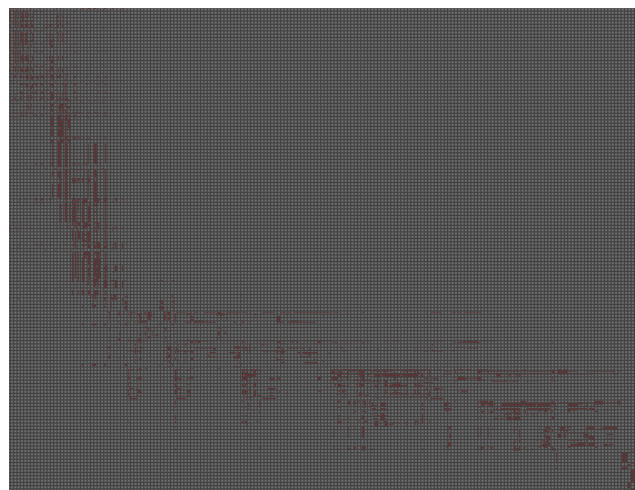
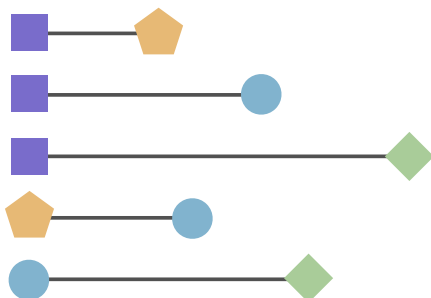
Sparse  
Censored  
Noisy  
High-dimensional



## Antigenic Cartography

Developed by Derek Smith and colleagues

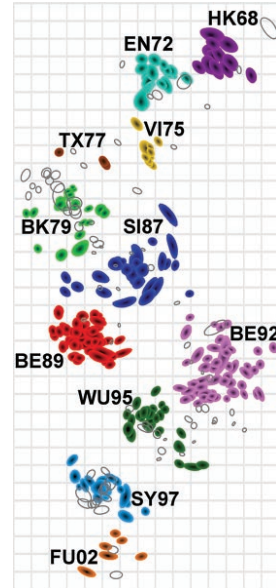
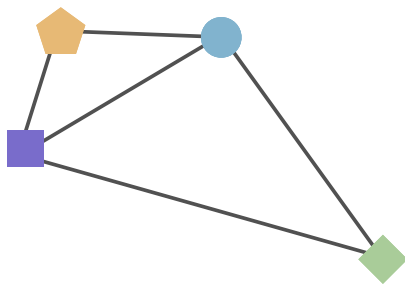
Uses multidimensional scaling (MDS) to position viruses in 2D space such that the distances in this space best fit the HI assay titres.



# Antigenic Cartography

Developed by Derek Smith and colleagues

Uses multidimensional scaling (MDS) to position viruses in 2D space such that the distances in this space best fit the HI assay titres.



Smith et al. 2004. Mapping the antigenic and genetic evolution of the influenza virus. Science.

## multidimension scaling (MDS) and Bayesian MDS

$d_{ij}$  = observed HI distance between virus  $i$  and sera  $j$

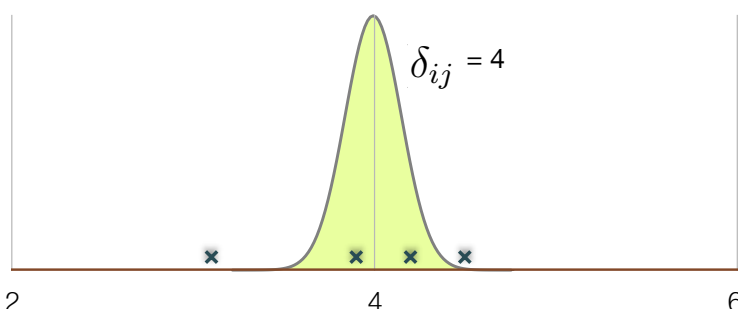
$\delta_{ij}$  = map distance between virus  $i$  at  $X_i$  and sera  $j$  at  $Y_j$

The goal of **MDS** is to find locations that minimize the difference between HI distance and map distance:

$$\sum_{(i,j) \in \mathcal{I}} (\delta_{ij} - d_{ij})^2$$

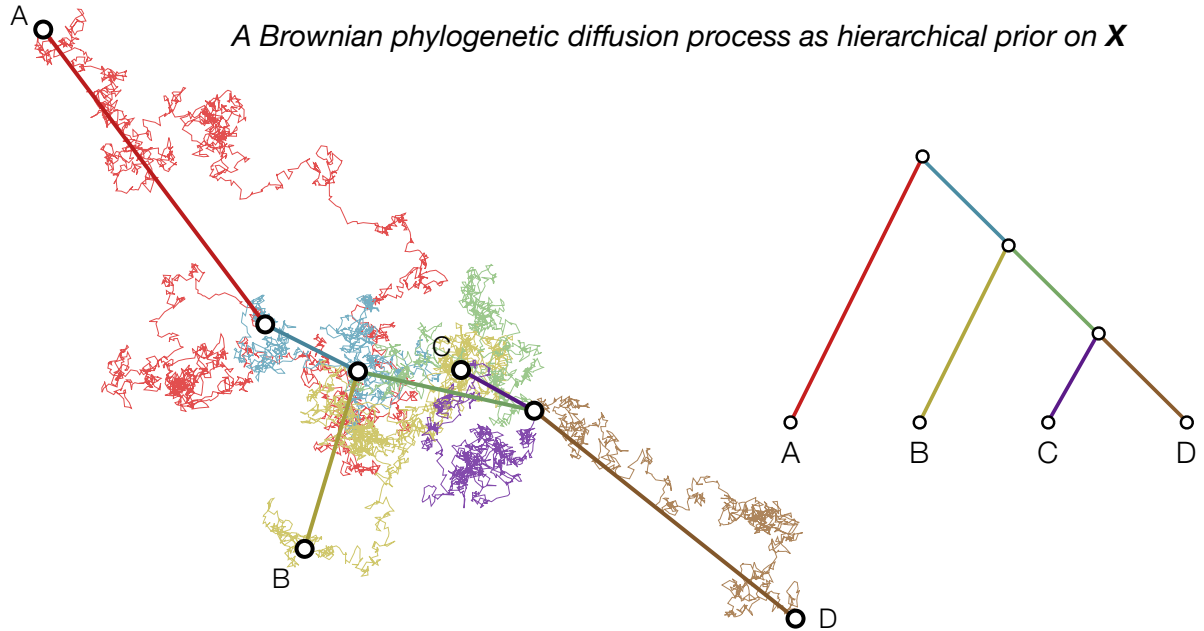
In **Bayesian MDS**, probabilistically reformulated as:

$$d_{ij} \sim \text{Normal}(\delta_{ij}, \sigma)$$



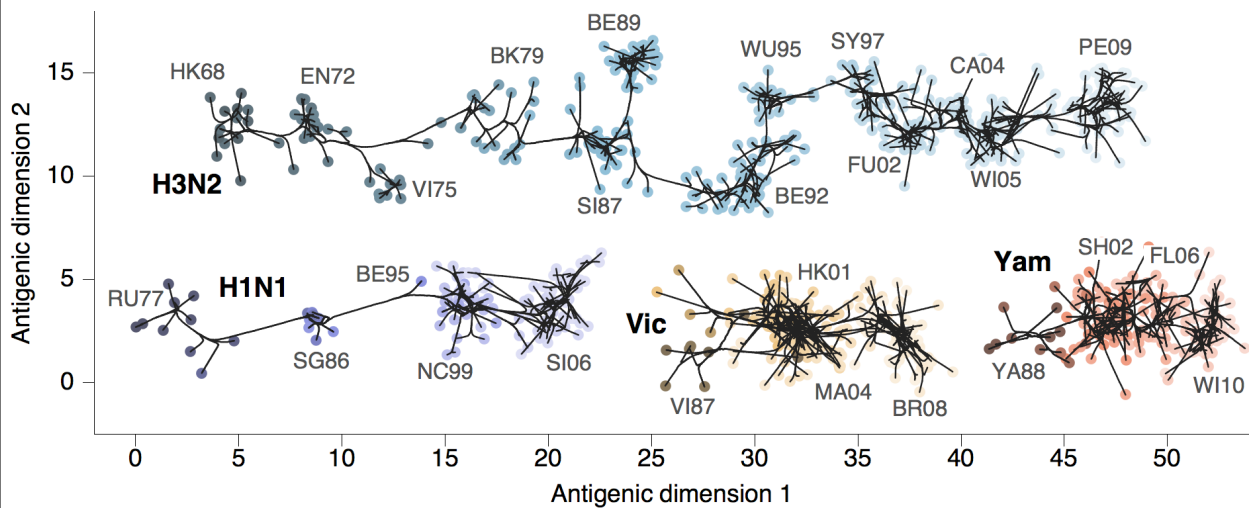
Oh and Raftery. 2001. Bayesian multidimensional scaling and choice of dimension. J Am Stat Assoc.

# Connecting antigenic to genetic evolution

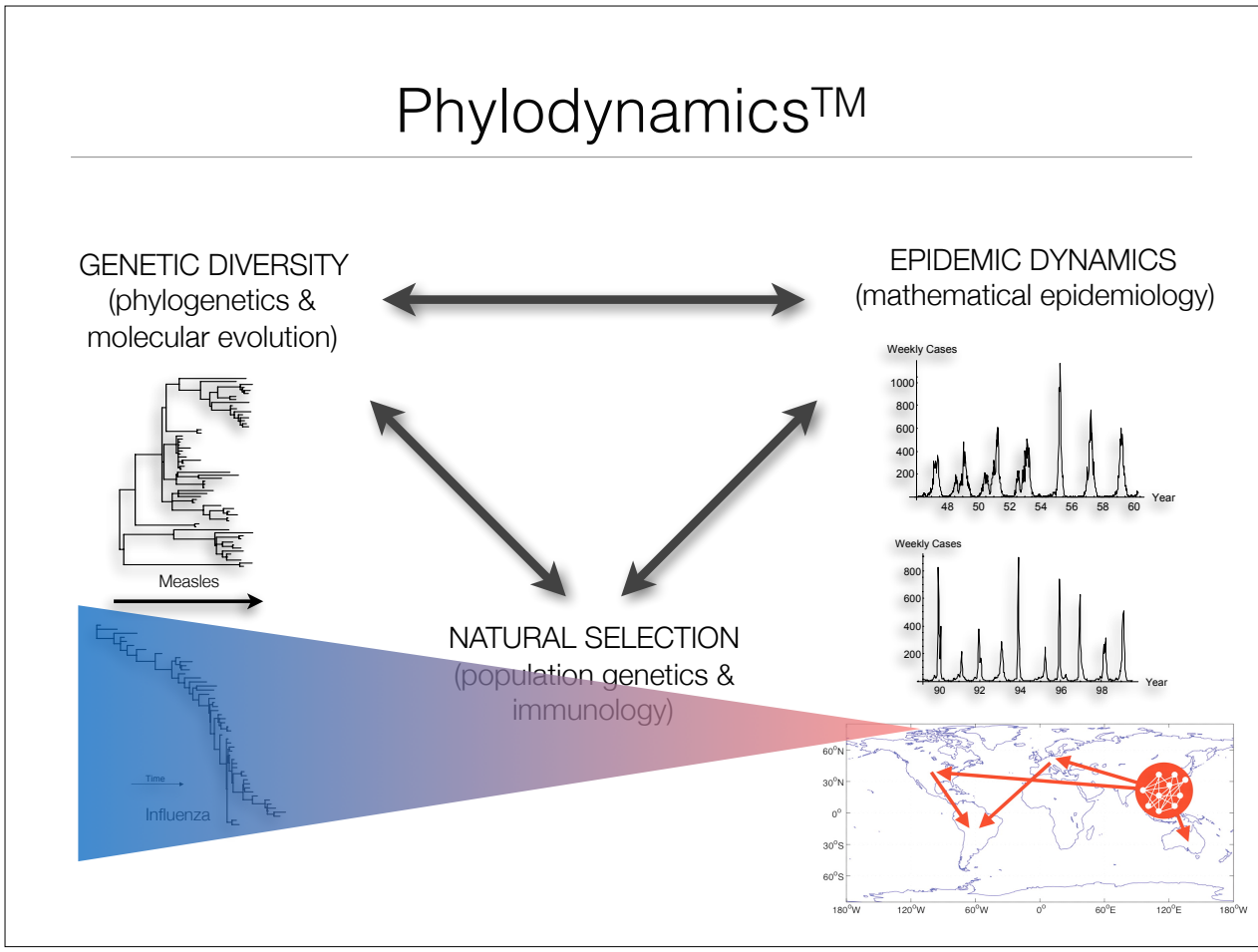
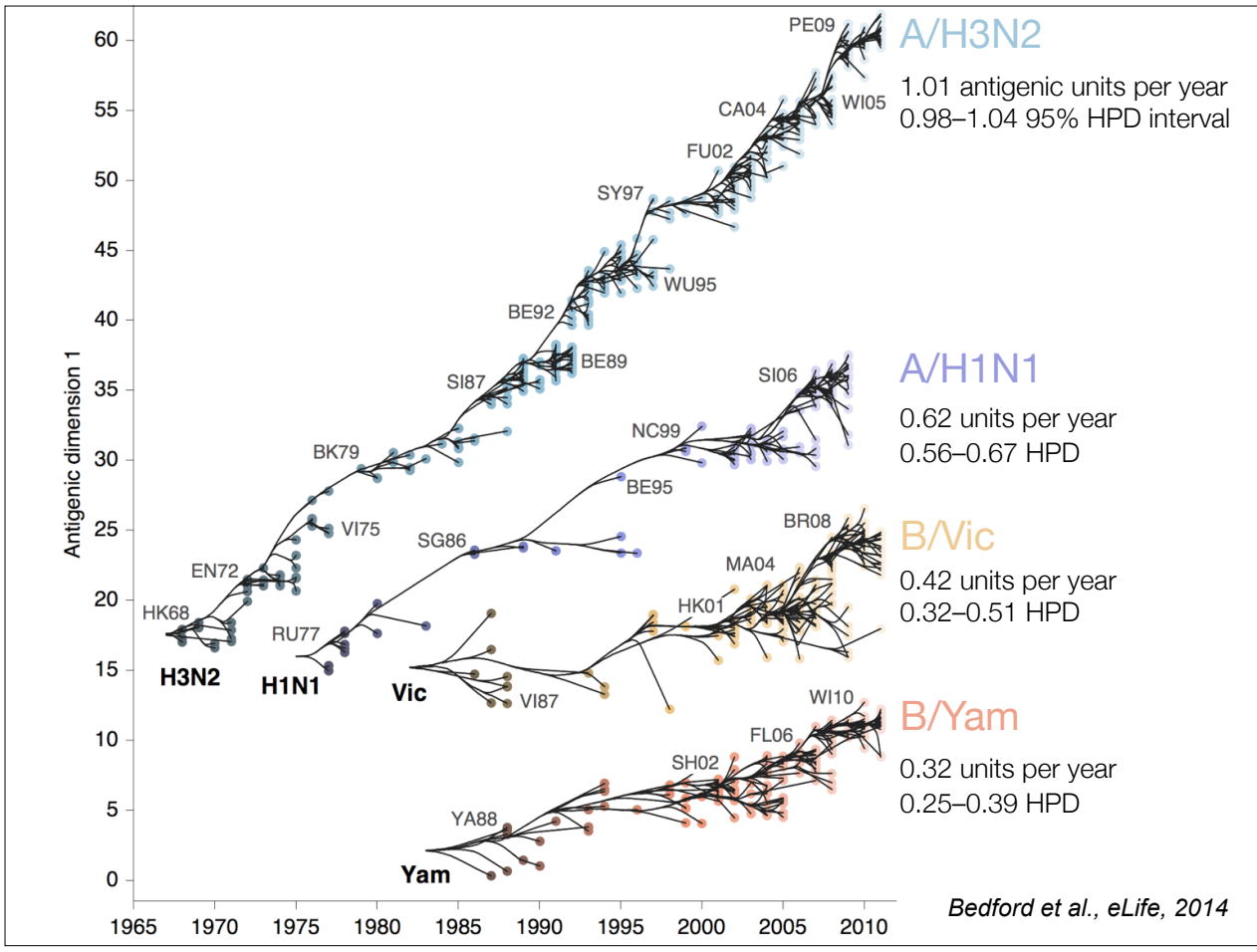


Lemey et al., MBE, 2010

# Comparison of antigenic drift across influenza lineages



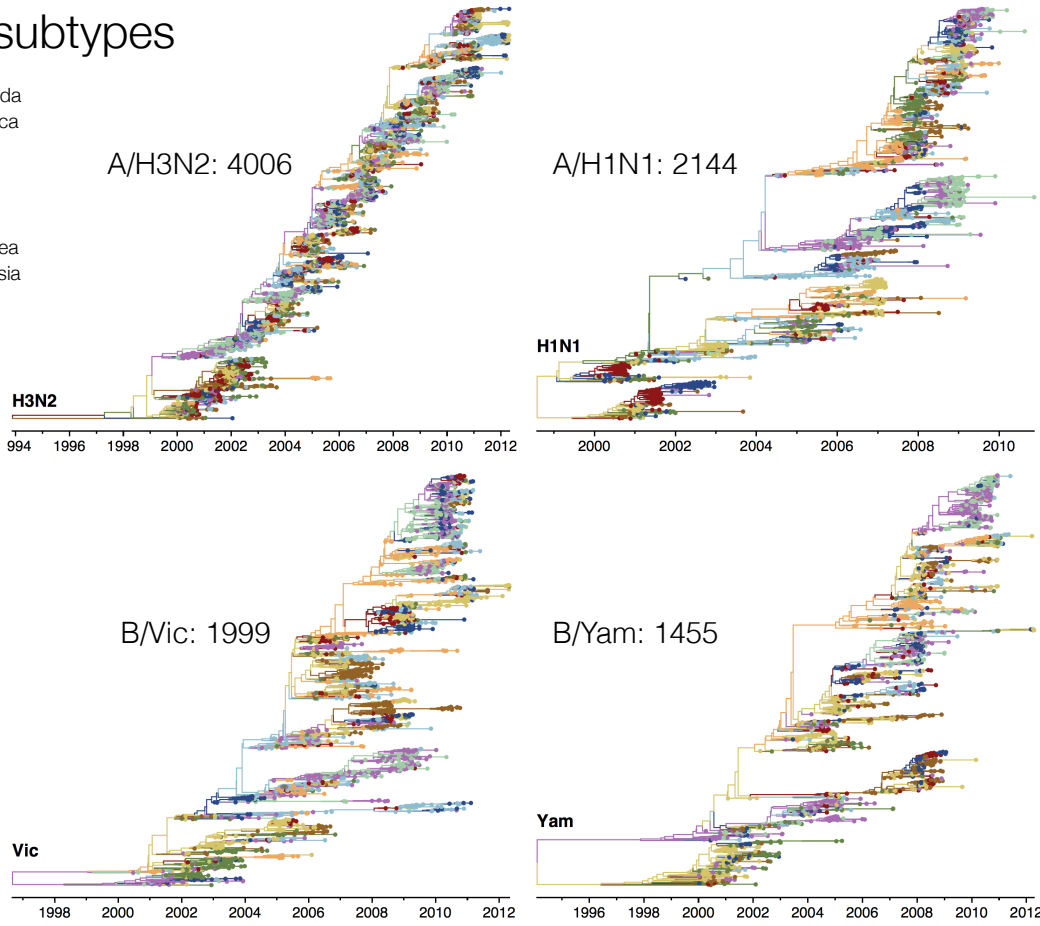
Bedford et al., eLife, 2014





# Across subtypes

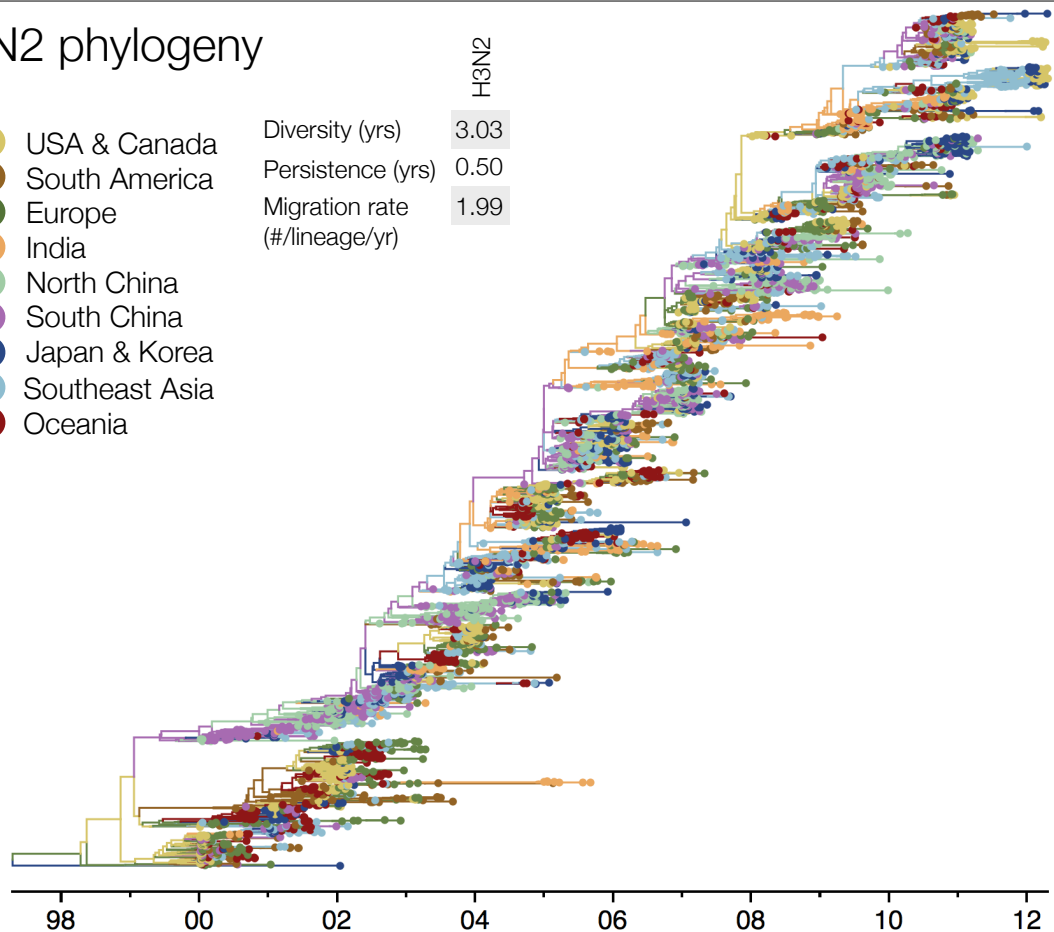
- USA & Canada
- South America
- Europe
- India
- North China
- South China
- Japan & Korea
- Southeast Asia
- Oceania



# H3N2 phylogeny

- USA & Canada
- South America
- Europe
- India
- North China
- South China
- Japan & Korea
- Southeast Asia
- Oceania

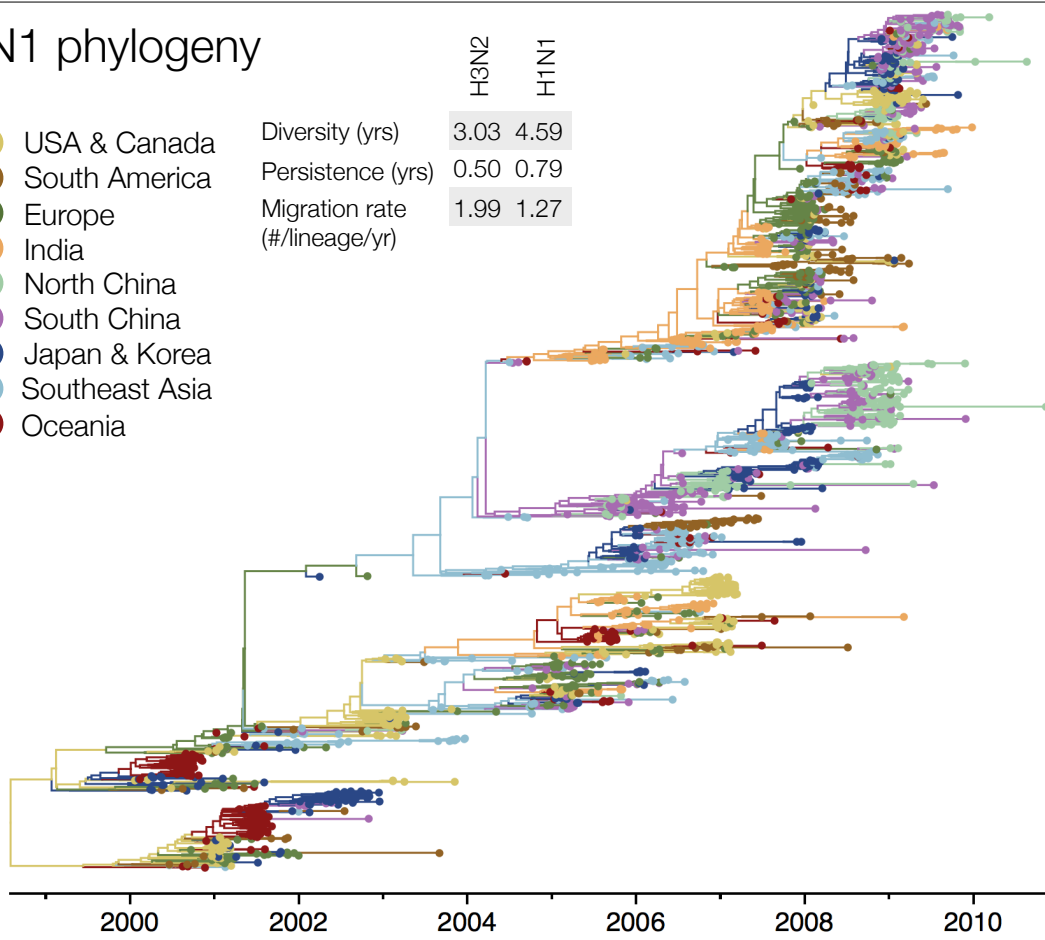
Diversity (yrs)	3.03
Persistence (yrs)	0.50
Migration rate (#/lineage/yr)	1.99



## H1N1 phylogeny

- USA & Canada
- South America
- Europe
- India
- North China
- South China
- Japan & Korea
- Southeast Asia
- Oceania

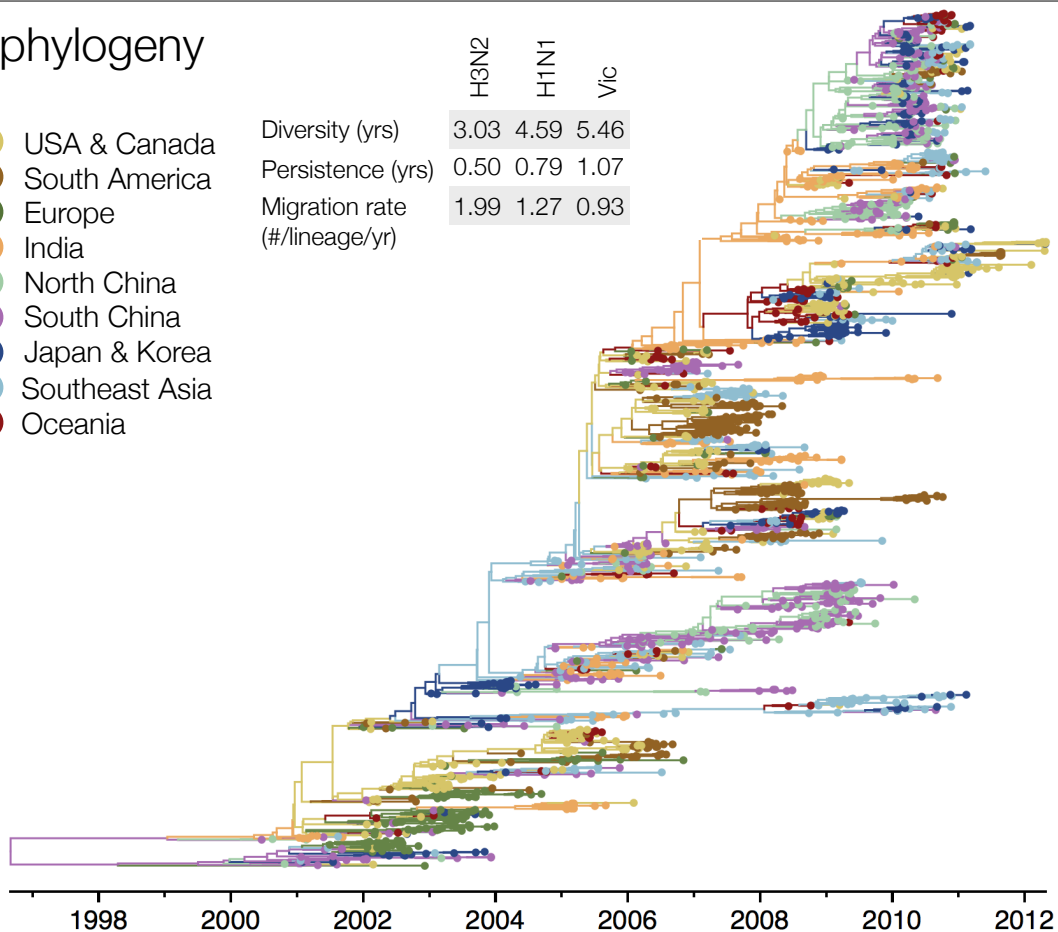
	H3N2	H1N1
Diversity (yrs)	3.03	4.59
Persistence (yrs)	0.50	0.79
Migration rate (#/lineage/yr)	1.99	1.27



## Vic phylogeny

- USA & Canada
- South America
- Europe
- India
- North China
- South China
- Japan & Korea
- Southeast Asia
- Oceania

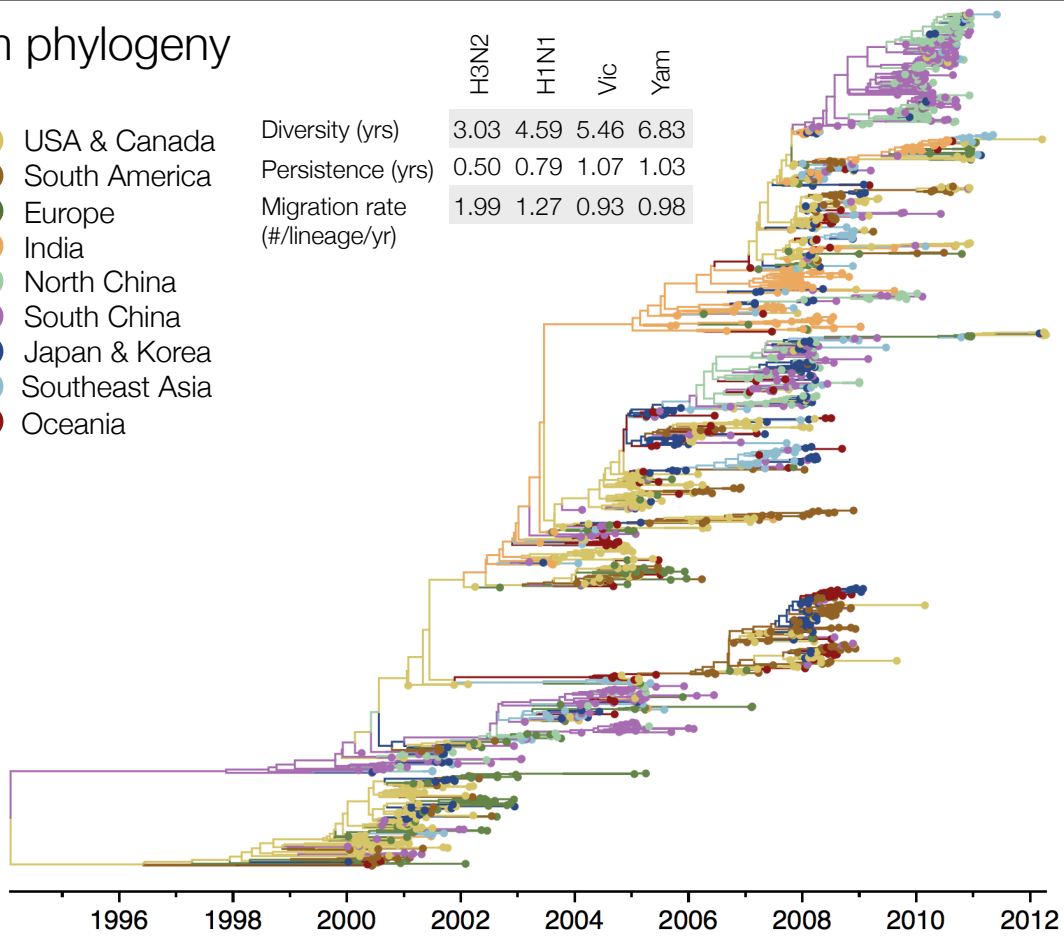
	H3N2	H1N1	Vic
Diversity (yrs)	3.03	4.59	5.46
Persistence (yrs)	0.50	0.79	1.07
Migration rate (#/lineage/yr)	1.99	1.27	0.93



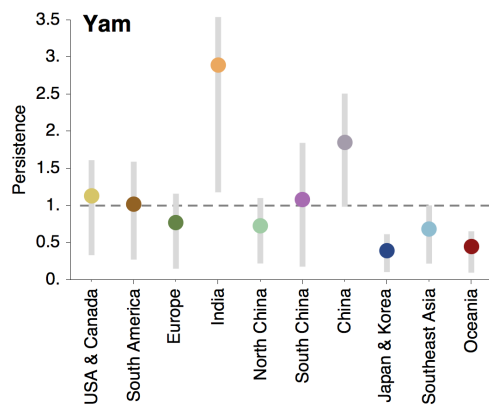
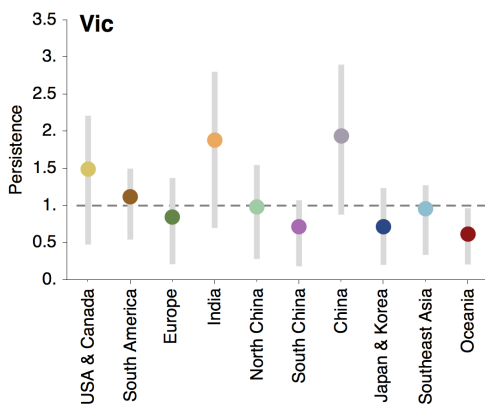
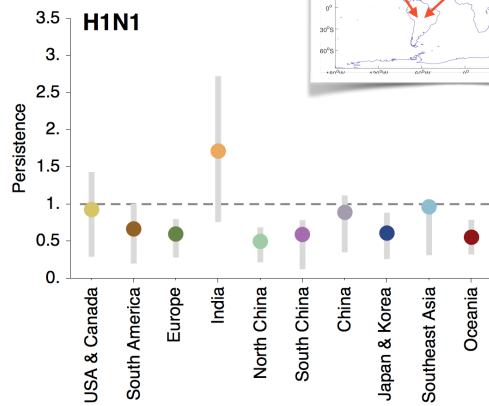
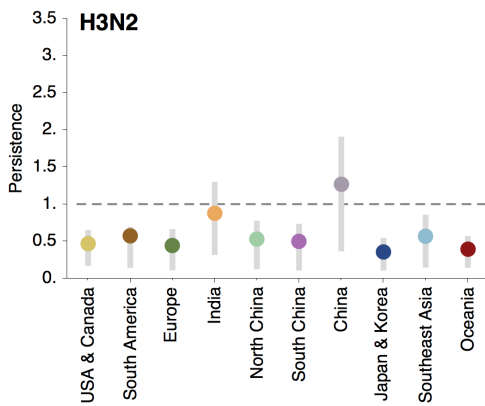
# Yam phylogeny

- USA & Canada
- South America
- Europe
- India
- North China
- South China
- Japan & Korea
- Southeast Asia
- Oceania

	H3N2	H1N1	Vic	Yam
Diversity (yrs)	3.03	4.59	5.46	6.83
Persistence (yrs)	0.50	0.79	1.07	1.03
Migration rate (#/lineage/yr)	1.99	1.27	0.93	0.98

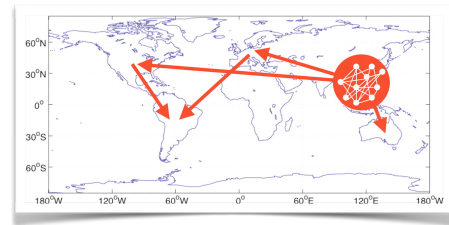
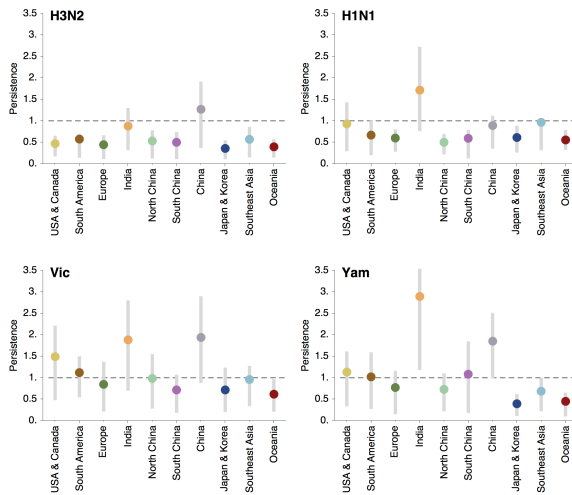
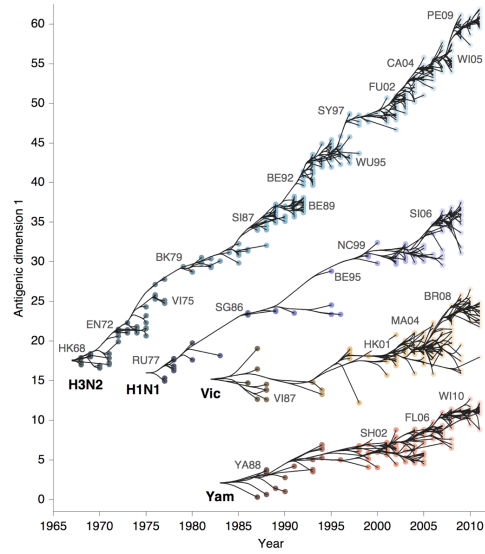


# Regional persistence



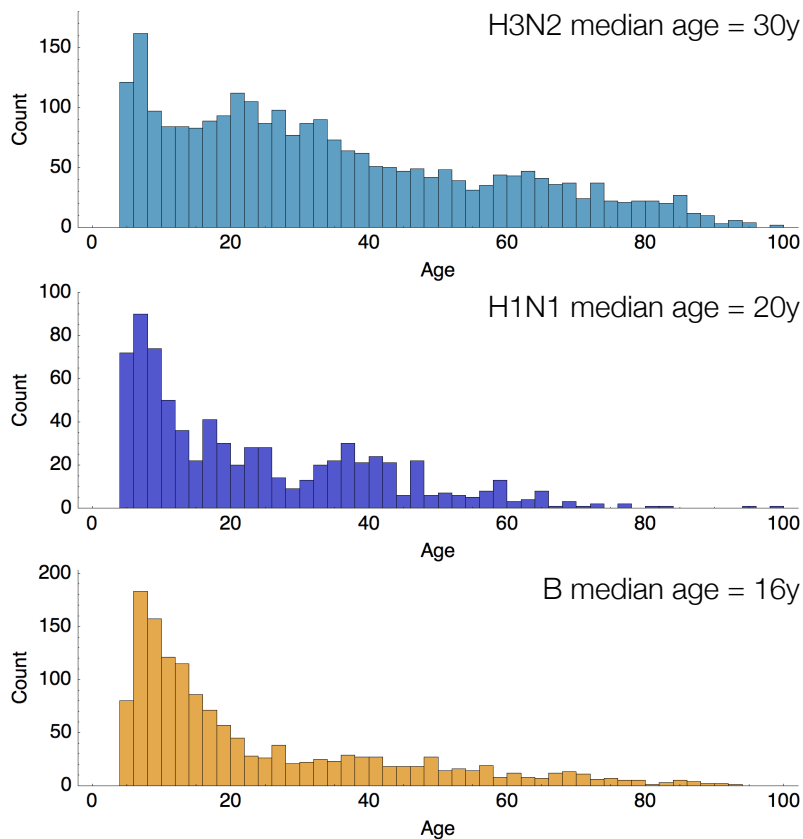
# Summary

	H3N2	H1N1	Vic	Yam
Diversity (yrs)	3.03	4.59	5.46	6.83
Persistence (yrs)	0.50	0.79	1.07	1.03
Migration rate (# lineage/yr)	1.99	1.27	0.93	0.98
Antigenic drift (antigenic units per yr)	1.01	0.62	0.42	0.32

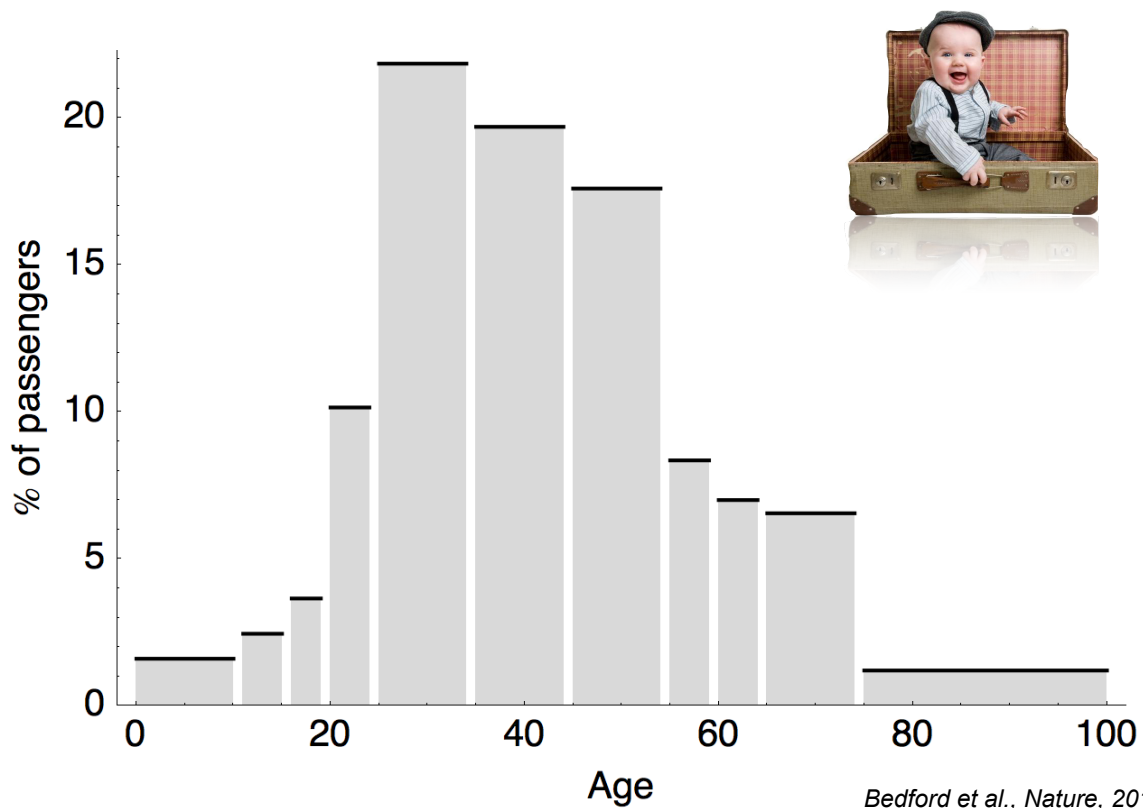


Bedford et al., Nature, 2015

# Age distribution across viruses

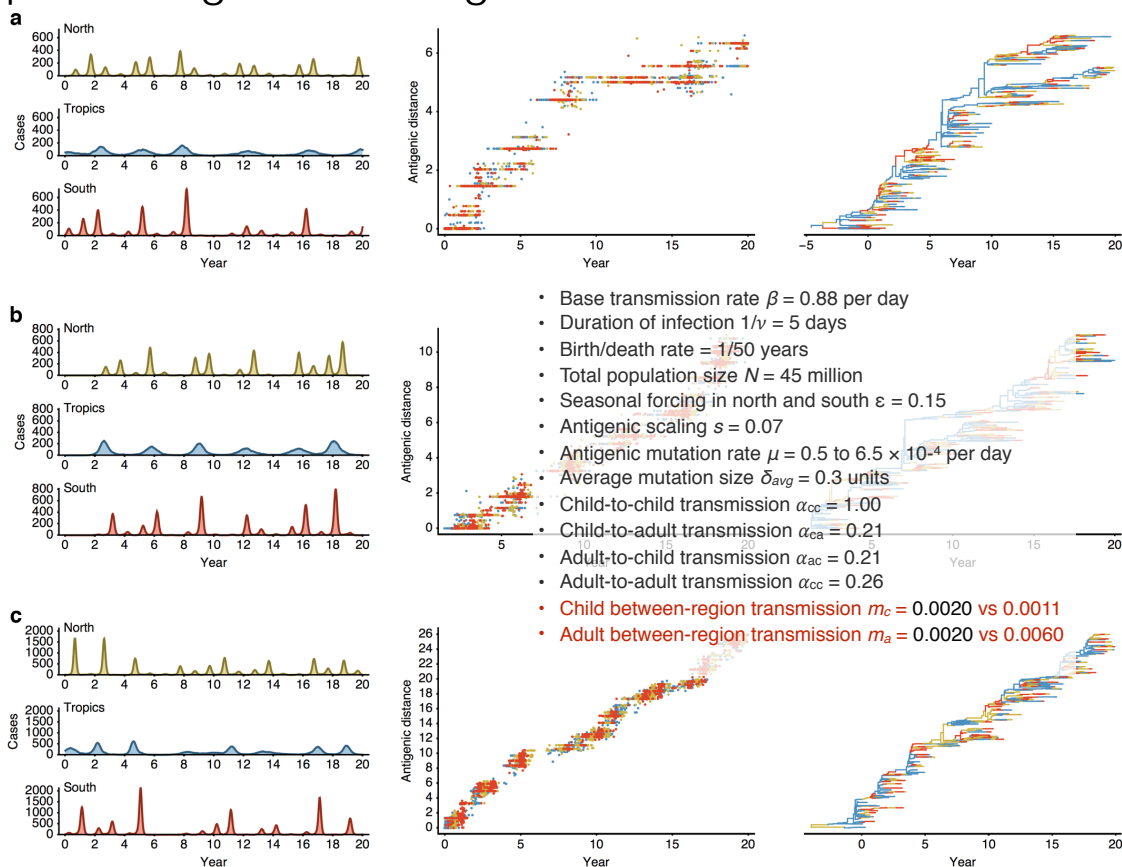


# Air travel differences between adults and children

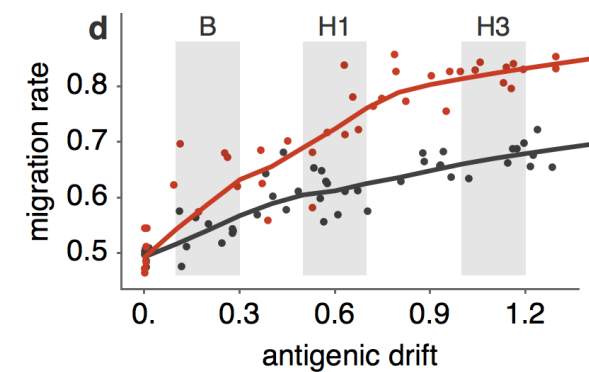
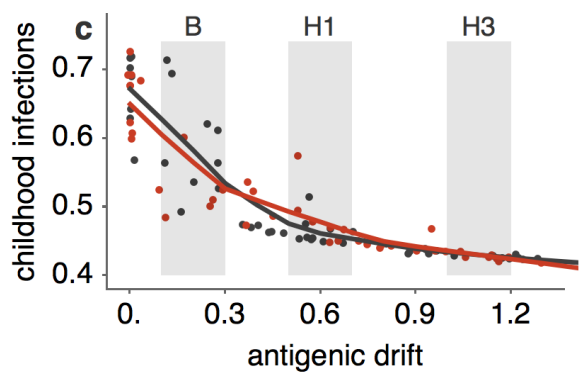
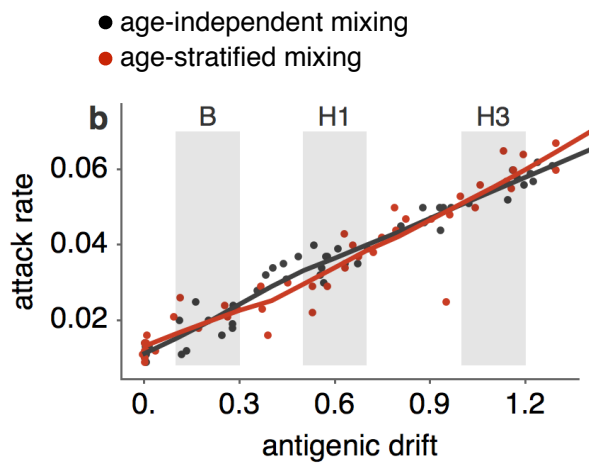
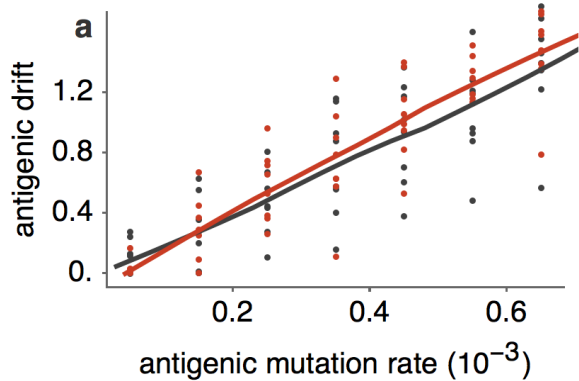


# Epidemiological modeling

*Bedford et al., BMC Biology, 2012*



# Epidemiological results



*Interaction between virus evolution, epidemiology and human behaviour drives migration rate differences*

*evade immunity more efficiently*

*be a virus, see the world*

*more of the*