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

Introduction to molecular epidemiology and infectious disease phylodynamics

Philippe Lemey[1] and Marc Suchard[2]

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 18-20, 2018

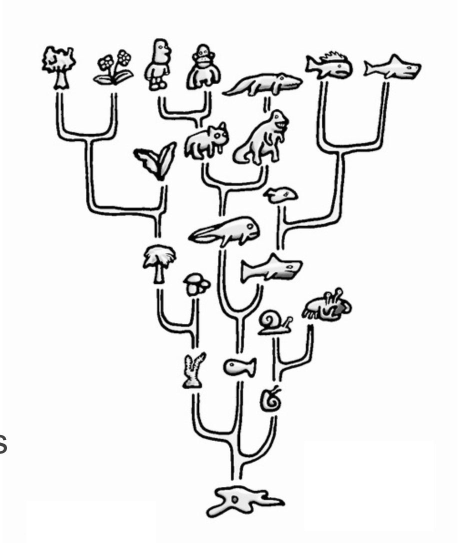
This course (SISMID module 12)

- Wednesday, July 18
  - ➡ Introduction
  - ➡ Alignment, substitution models and phylogenetic inference
- Thursday, July 19
  - ➡ Phylogenetic inference practical
  - ➡ Bayesian phylogenetics
  - ➡ Molecular clocks and model testing
  - ➡ BEAST practical
- Friday, July 20
  - ➡ Viral epidemiology and the coalescent
  - ➡ BEAST practical
  - ➡ Phylogeography
  - ➡ BEAST practical
- Bonus
  - ➡ Phylo-Alignment
  - ➡ Recombination
  - ➡ Robust Counting
  - ➡ Antigenic cartography

(We are here to cater for your needs!)

Molecular evolution and phylogenetics

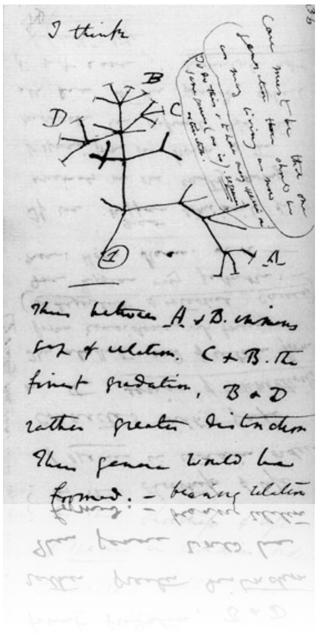
- biological **sequences** (DNA, RNA, protein) contain information about the processes and events that formed them
- this information is often **scrambled**, **fragmentary**, **hidden**, or **lost** completely
- our aim is to use **mathematical models** to recover and decipher this information
- The central concept is a **phylogeny**: a diagram depicting the ancestral relationships among characters or genetic sequences



```
HIV-1 (UK)   ATC---TGCTAAAGCATATGACACAGAGGTACATAATGTTT
HIV-1 (USA)  ATCGGATGCTAGAGCTTATGATACAGAGGTACA---TGTTT
```

Phylogenetics

- Darwin, 1837



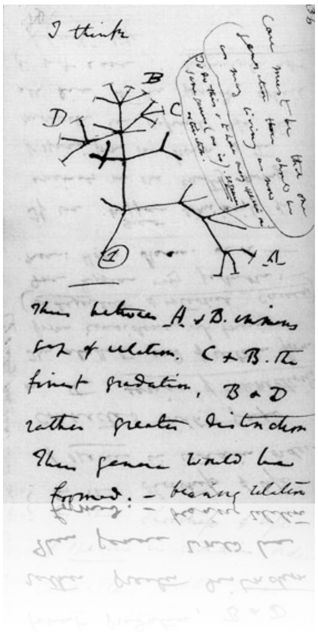
I think

Then between A & B. immense gap of relation. C & B. the finest gradation, B & D rather greater distinction. Thus genera would be formed. — bearing relation

- Haeckel, 1866

Phylogenetics

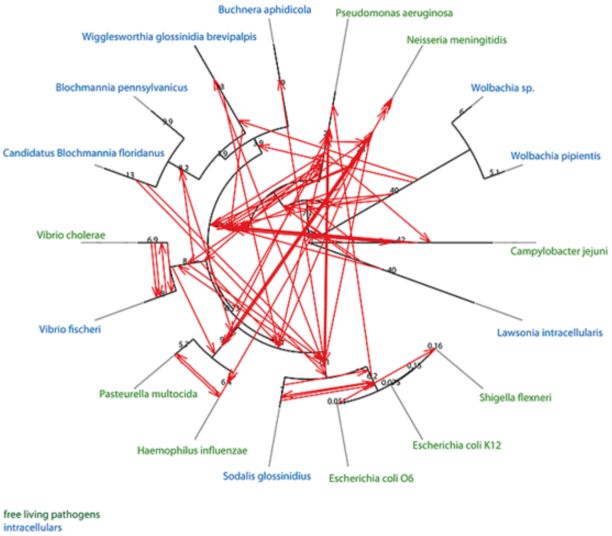
- Darwin, 1837



I think

Case must be that one generation then should be as many living as now. To do this & to have many species in same genus (as is) requires extinction.

Then between A & B. immense gap of relation. C & B. the finest gradation, B & D rather greater distinction. Thus genera would be formed. — bearing relation



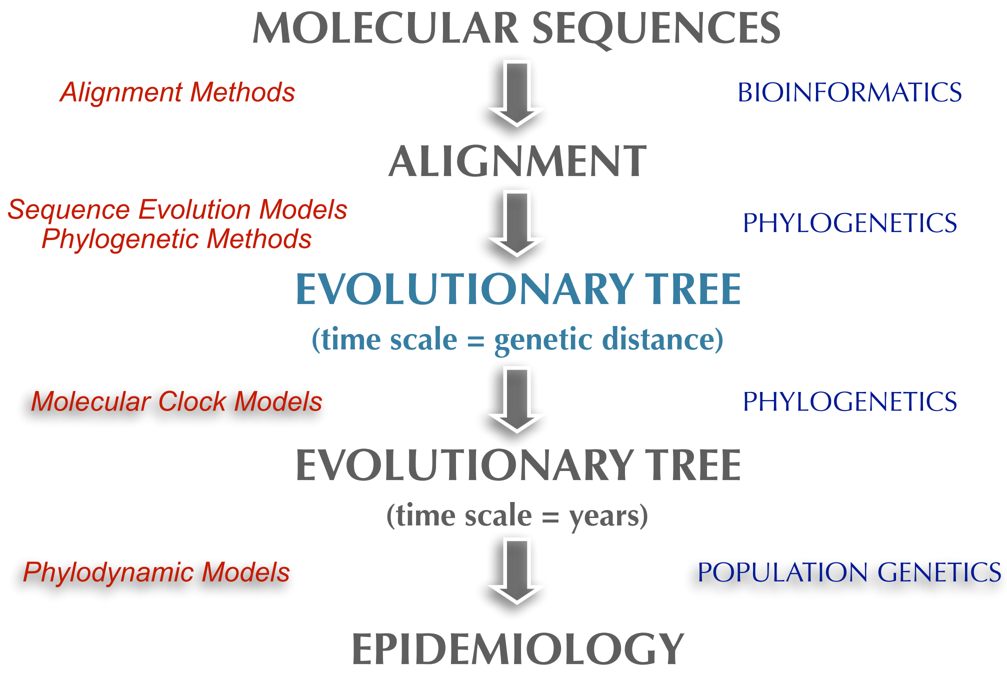
Genome distance tree with lateral gene transfers

Information in (viral) molecular sequences

- Genetic distances among strains
- Phylogeny
  - ➡ subtyping/classification
  - ➡ identification of transmission clusters
  - ➡ association with risk factors / traits
  - ➡ forensics
- Dates of historical events
- Evolutionary processes
  - ➡ recombination
  - ➡ natural selection
- Epidemiological processes
  - ➡ transmission rates
  - ➡ movement among locations
- Phenotypic trait evolution?

```
HIV-1 (UK)   ATC---TGCTAAAGCATATGACACAGAGGTACATAATGTTT
HIV-1 (USA)  ATCGGATGCTAGAGCTTATGATACAGAGGTACA---TGTTT
```

Our goal



MOLECULAR SEQUENCES

Alignment Methods — BIOINFORMATICS

ALIGNMENT

Sequence Evolution Models
Phylogenetic Methods — PHYLOGENETICS

EVOLUTIONARY TREE
(time scale = genetic distance)

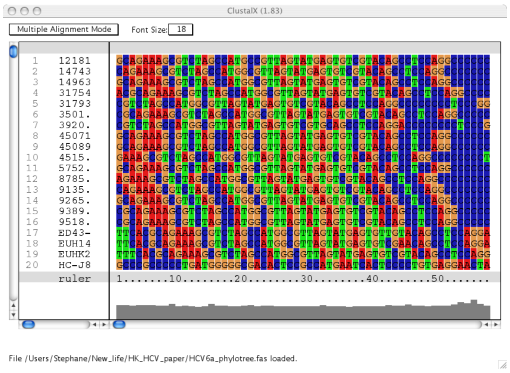
Molecular Clock Models — PHYLOGENETICS

EVOLUTIONARY TREE
(time scale = years)

Phylodynamic Models — POPULATION GENETICS

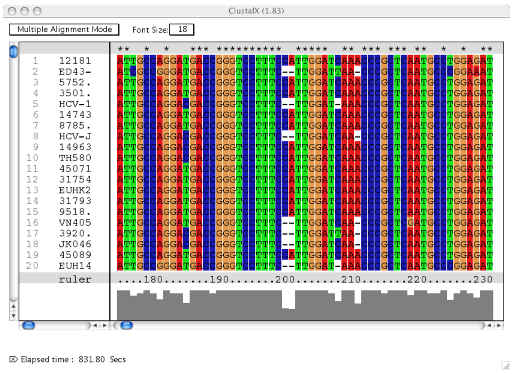
EPIDEMIOLOGY

Sequence alignment



```
 1   12181   GCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCCC
 2   14743   CAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCCCC
 3   14963   GCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCCC
 4   31754   ACGCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCC
 5   31793   CGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCCCCCTCCCGG
 6   3501.   CGCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCC
 7   3920.   CGTCTAGCCATGGCGTTAGTATGAGTGTCGTGCAGCCTCCAGGACCCCCCCCTCCCG
 8   45071   GCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCCC
 9   45089   GCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCCC
10   4515.   GAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCCCCT
11   5752.   GCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCCC
12   8785.   AGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCCCCC
13   9135.   CAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCCCC
14   9265.   GCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCCC
15   9389.   CGCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCC
16   9518.   CGCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGGCCCCC
17   ED43-   TTCACGCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTTGTACAGCCTCCAGGA
18   EUH14   TTCACGCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGAACAGCCTCCAGGA
19   EUHK2   TTTCACGCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTACAGCCTCCAGG
20   HC-J8   GCCCGCCCCCTGATGGGGGCGACACTCCGCCATGAATCACTCCCCTGTGAGGAACTA
     ruler   1.......10........20........30........40........50.......
```

File /Users/Stephane/New_life/HK_HCV_paper/HCV6a_phylotree.fas loaded.



```
             **   *  *   ***  **********   *****   **  ***  ***  **   *   *   **
 1   12181   ATTGCCAGGATGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
 2   ED43-   ATCGCCGGGATGACCGGGTCCTTTC--TTGGATTAA-CCCGCTCAATGCCCGGAAAT
 3   5752.   ATTGCCAGGATGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
 4   3501.   ATTGCCAGGATGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
 5   HCV-1   ATTGCCAGGACGACCGGGTCCTTTC--TTGGAT-AAACCCGCTCAATGCCTGGAGAT
 6   14743   ATTGCCAGGATGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
 7   8785.   ATTGCCAGGATGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
 8   HCV-J   ATTGCCAGGACGACCGGGTCCTTTC--TTGGATCAA-CCCGCTCAATGCCTGGAGAT
 9   14963   ATTGCCAGGACGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
10   TH580   ATTGCCAGGACGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
11   45071   ATTGCCAGGATGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
12   31754   ATTGCCAGGATGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
13   EUHK2   ATTGCCAGGATGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
14   31793   ATTGCCAGGATGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
15   9518.   ATTGCCAGGATGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
16   VN405   ATTGCCAGGATGACCGGGTCCTTTC--TTGGATCAA-CCCGCTCGATGCCTGGAGAT
17   3920.   ATTGCCAGGACGACCGGGTCCTTTC--TTGGATTAA-CCCGCTCAATGCCTGGAGAT
18   JK046   ATTGCCAGGACGACCGGGTCCTTTC--TTGGATCAA-CCCGCTCAATGCCTGGAGAT
19   45089   ATTGCCAGGATGACCGGGTCCTTTCCATTGGATCAAACCCGCTCAATGCCTGGAGAT
20   EUH14   ATTGCCGGGATGACCGGGTCCTTTC--TTGGAT-AAACCCGCTCAATGCCCGGAGAT
     ruler   ....180.......190.......200.......210.......220.......230
```

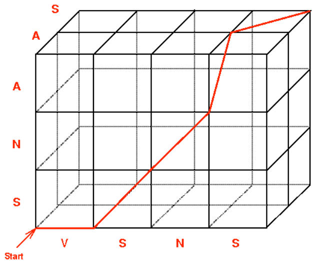
⌧ Elapsed time : 831.80 Secs

Sequence X

Sequence Y

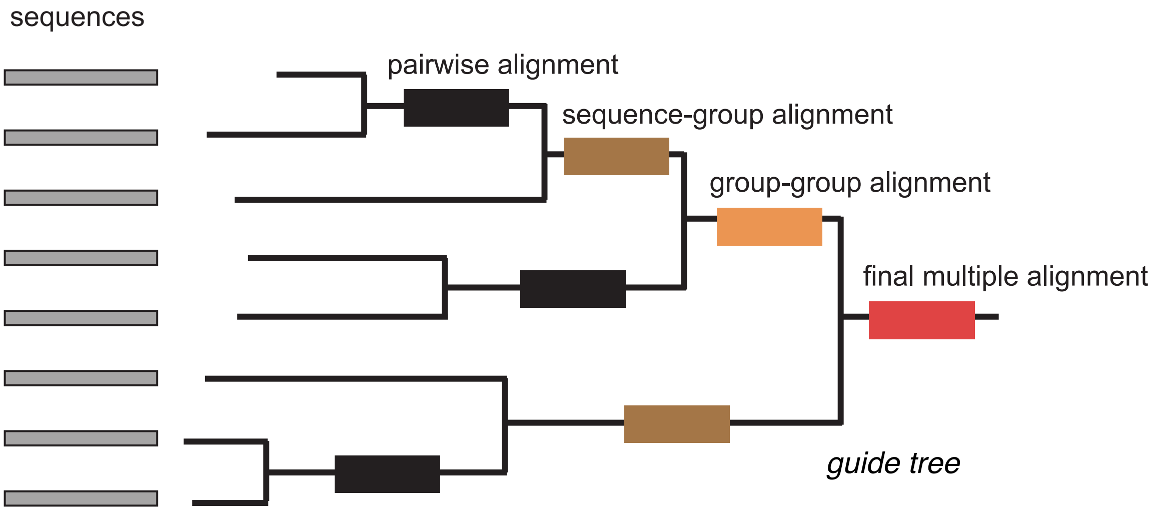
j \ i		1	2	3	4	5	6	7	8
		*	G	R	Q	T	A	G	L
1	*	0	-8	-10	-12	-14	-16	-18	-20
2	G	-8	6	-2	-4	-6	-8	-10	-12
3	T	-10	-2	5	-3	1	-6	-10	-11
4	A	-12	-4	-3	4	-3	5	-3	-11
5	Y	-14	-6	-5	-4	2	-3	2	-4
6	D	-16	-8	-7	-5	-3	0	-4	-2
7	L	-18	-10	-9	-9	-6	-4	-4	0

```
GRQTA-GL
G--TAYDL
```



```
V S N _ S
_ S N A _
_ _ _ A S
```

Progressive alignment



http://www.kuleuven.be/aidslab/phylogenybook/Table3.1.html

Genetic distances

```
SIVcpz     ATGGGTGCGA GAGCGTCAGT TCTAACAGGG GGAAAATTAG ATCGCTGGGA
HIV-1      ATGGGTGCGA GAGCGTCAGT ATTAAGCGGG GGAGAATTAG ATCGATGGGA

SIVcpz     AAAAGTTCGG CTTAGGCCCG GGGGAAGAAA AAGATATATG ATGAAACATT
HIV-1      AAAAATTCGG TTAAGGCCAG GGGGAAAGAA AAAATATAAA TTAAAACATA

SIVcpz     TAGTATGGGC AAGCAGGGAG CTGGAAAGAT TCGCATGTGA CCCCGGGCTA
HIV-1      TAGTATGGGC AAGCAGGGAG CTAGAACGAT TCGCAGTTAA TCCTGGCCTG

SIVcpz     ATGGAAAGTA AGGAAGGATG TACTAAATTG TTACAACAAT TAGAGCCAGC
HIV-1      TTAGAAACAT CAGAAGGCTG TAGACAAATA CTGGGACAGC TACAACCATC

SIVcpz     TCTCAAAACA GGCTCAGAAG GACTGCGGTC CTTGTTTAAC ACTCTGGCAG
HIV-1      CCTTCAGACA GGATCAGAAG AACTTAGATC ATTATATAAT ACAGTAGCAA

SIVcpz     TACTGTGGTG CATACATAGT GACATCACTG TAGAAGACAC ACAGAAAGCT
HIV-1      CCCTCTATTG TGTGCATCAA AGGATAGAGA TAAAAGACAC CAAGGAAGCT

SIVcpz     CTAGAACAGC TAAAGCGGCA TCATGGAGAA CAACAGAGCA AAACTGAAAG
HIV-1      TTAGACAAGA TAGAG--GAA ------GAGCA AAACAAAAGT AA---GAAAA

SIVcpz     TAACTCAGGA AGCCGTGAAG GGGGAGCCAG TCAAGGCGCT AGTGCCTCTG
HIV-1      AAGCACAGCA AGC------AG CAGCTGACA- -CAGGACAC- AG--CAGC--

SIVcpz     CTGGCATTAG TGGAAATTAC
HIV-1      CAGG--TCAG CCAAAATTAC
```

chimpanzee SIV vs HIV-1 envelope gene

Not all mutations are equally likely

- some point substitutions are more likely to occur than others:

  transitions are more likely than transversions

  - *transitions*:

    purine↔purine or

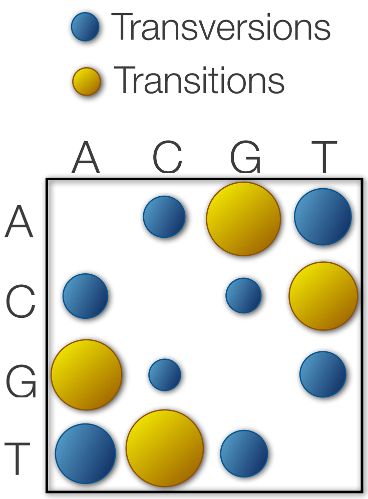
    pyrimidine↔pyrimidine

    A↔G C↔T

  - *transversions*:

    A↔C A↔T

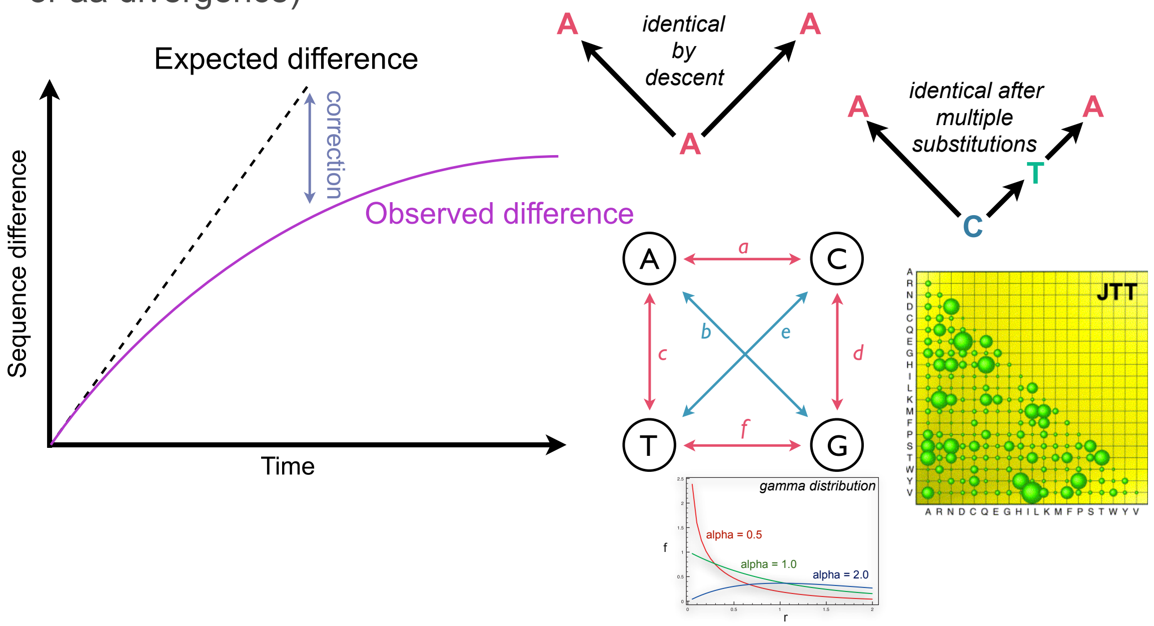
    G↔C G↔T



Unambiguous changes on most parsimonious tree of Ciliate SSUrDNA

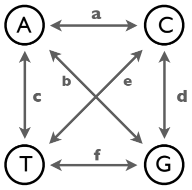
Substitution models

- During evolution, 'multiple hits' can occur at a single position: the evolutionary distance is almost always larger than the dissimilarity (% nt or aa divergence)



Nucleotide substitution models

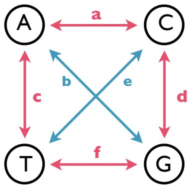
Simplest (few parameters)



1. Base frequencies are equal and all substitutions are equally likely (Jukes-Cantor)

(a=b=c=d=e=f)

↓



2. Base frequencies are equal but transitions and transversions occur at different rates (Kimura 2-parameter)

(a=c=d=f, b=e)

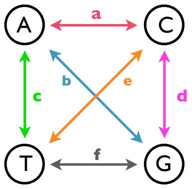
↓

3. Unequal base frequencies and transitions and transversions occur at different rates (Hasegawa-Kishino-Yano)

(a=c=d=f, b=e)

↓

Most complex (many parameters)



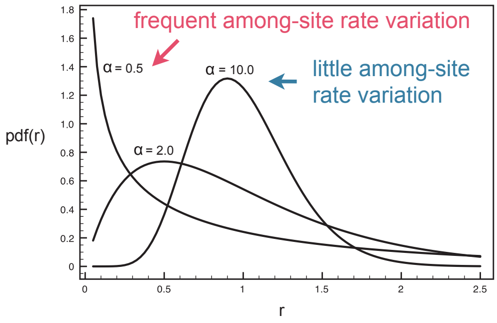
4. Unequal base frequencies and all substitution types occur at different rates (General Reversible Model)

(a, b, c, d, e, f)

Does this matter?

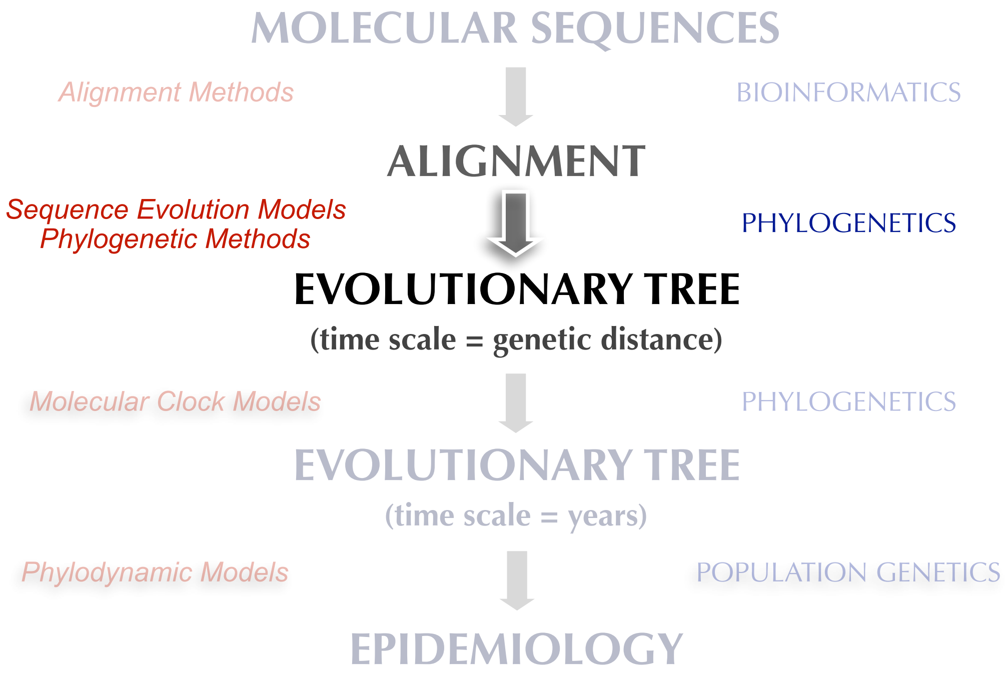
Estimated genetic distances between SIVcpz and HIVlai, under different substitution models:

Observed % mismatches	= 0.406
JC (Jukes-Cantor)	= 0.586
HKY (Hasegawa-Kishino-Yano)	= 0.611
GTR (General Time Reversible)	= 0.620
GTR + gamma	= 1.017



Gene	α
Prolactin	1.37
Albumin	1.05
C-myc	0.47
Ctyochrome β (mtDNA)	0.44
Insulin	0.40
D-loop (mtDNA)	0.17
12S rRNA (mtDNA)	0.16

Phylogenetic reconstruction



MOLECULAR SEQUENCES

Alignment Methods

BIOINFORMATICS

ALIGNMENT

Sequence Evolution Models
Phylogenetic Methods

PHYLOGENETICS

EVOLUTIONARY TREE

(time scale = genetic distance)

Molecular Clock Models

PHYLOGENETICS

EVOLUTIONARY TREE

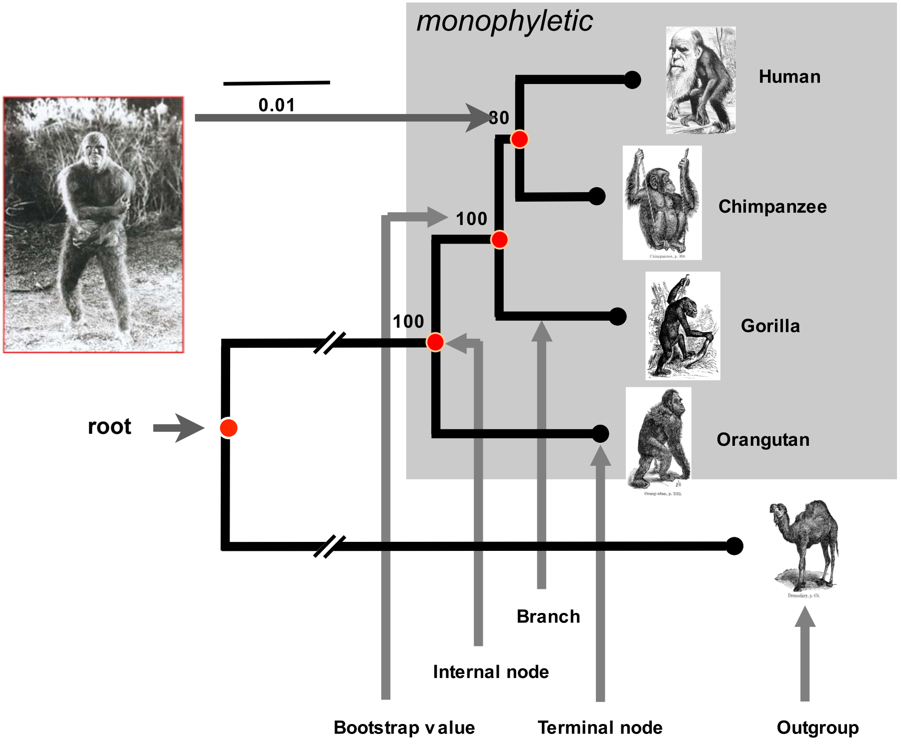
(time scale = years)

Phylodynamic Models

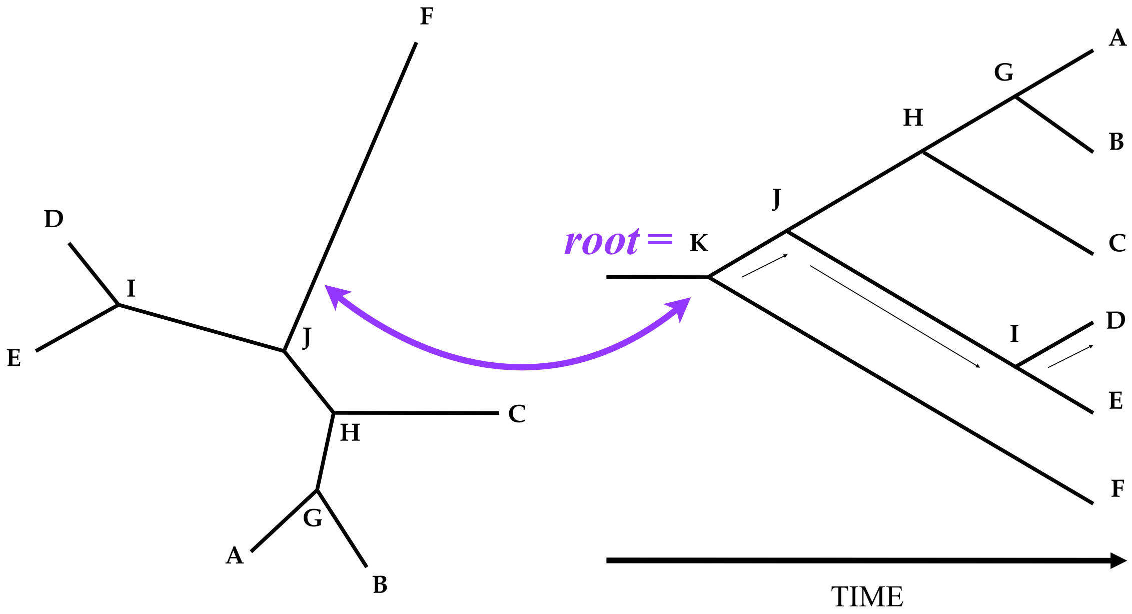
POPULATION GENETICS

EPIDEMIOLOGY

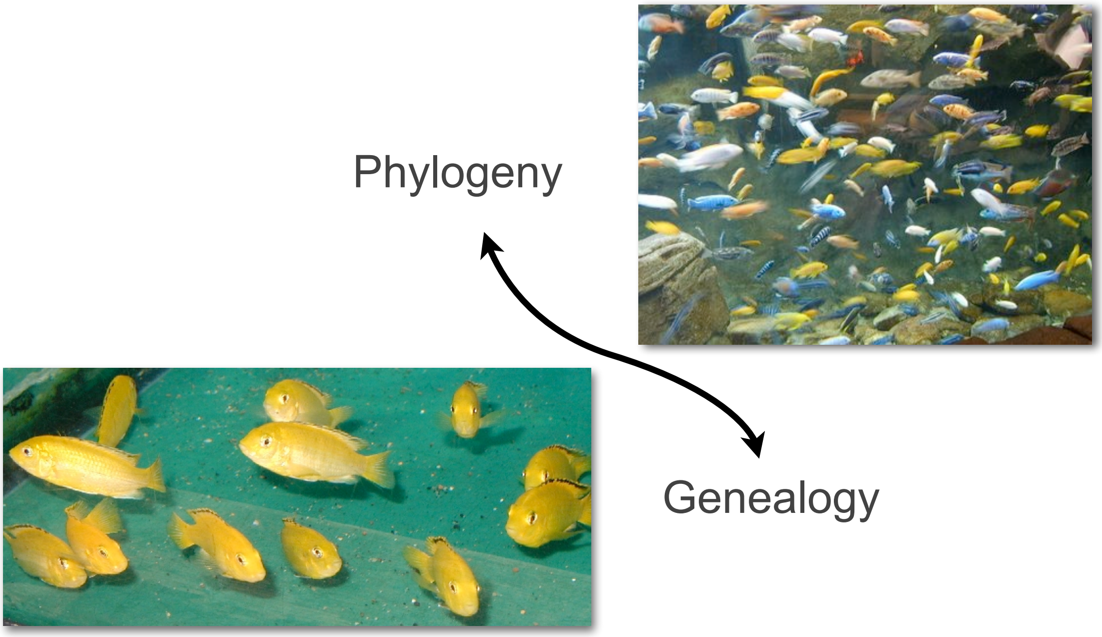
What is a tree?



Tree terminology: unrooted and rooted



Tree Terminology



Phylogenetic reconstruction

- **CLUSTERING APPROACHES:** These begin with a genetic distance between each pair of sequences. A ‘clustering algorithm’ then transforms the genetic distances into a tree.
  - e.g. UPGMA, Neighbour-Joining
  - Simple, faster.
  - No measure of how good the estimated tree is (non-statistical)
- **OPTIMALITY METHODS**: These define a score for each possible tree. ‘Search algorithms’ are then used to find the tree with the highest score.
  - e.g. Parsimony, Maximum Likelihood (& Bayesian Inference)
  - More complex, slower. Search may not locate the ‘best’ tree.
  - Quality of each tree can be directly compared (statistical)

Phylogenetic reconstruction

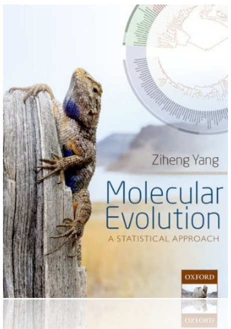
- For n taxa, there are:

  $(2n-3)!/[(2^{n-2})*(n-2)!]$

  rooted, binary trees

# taxa	# trees	
4	15	enumerable by hand
5	105	enumerable by hand on a rainy day
6	945	enumerable by computer
7	10395	still searchable very quickly on computer
8	135135	a bit more than the number of hairs on your head
9	2027025	population of Glasgow
10	34459425	≈ upper limit for exhaustive searching; about the number of possible combinations of numbers in the National Lottery
20	8.20×10^{21}	≈ upper limit for branch-and-bound searching
48	3.21×10^{70}	≈ the number of particles in the universe
136	2.11×10^{267}	=number of trees to choose from in the “Out of Africa” data (Vigilant et al., 1991)

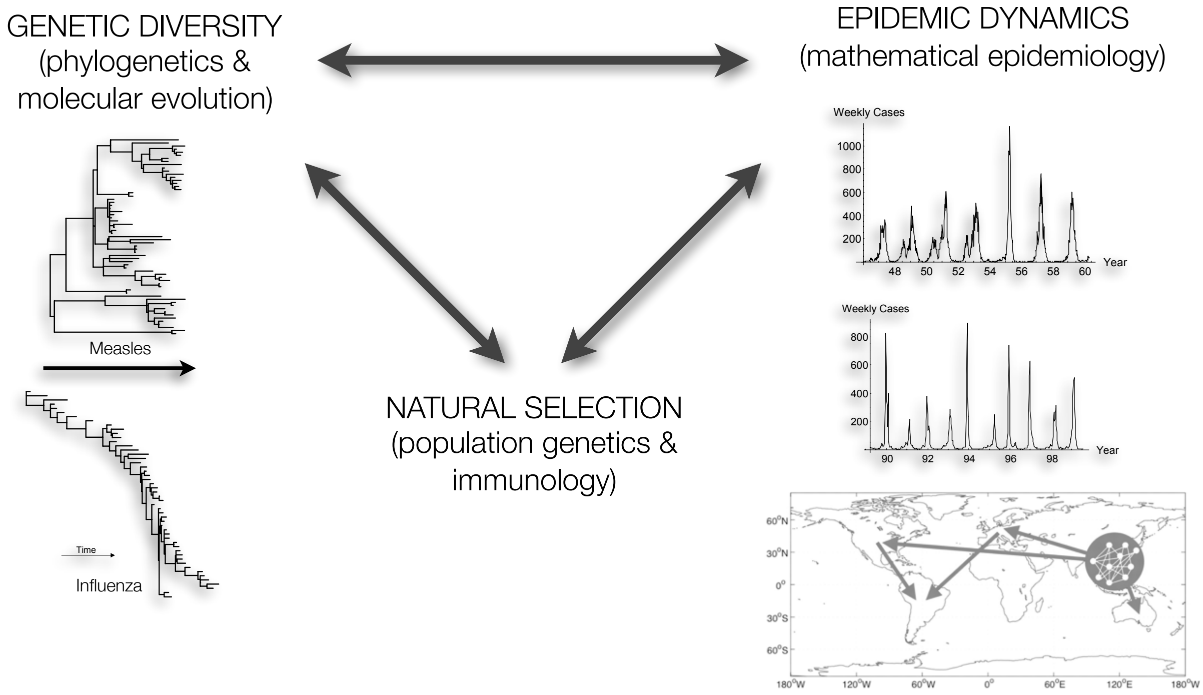
Phylogenetic inference: books



- Yang Z. (2003). *Computational Molecular Evolution*. Oxford University Press
- Nei M & Kumar S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press.
- Page RDM & Holmes EC. (1998). *Molecular Evolution: A Phylogenetic Approach*. Blackwell Science Ltd, Oxford.
- Yang Z (2014) Molecular Evolution: A Statistical Approach
- Bayesian Phylogenetics: Methods, Algorithms, and Applications. Chen M-H, Kuo L. and Lewis PO. Chapman & Hall/CRC.
- Lemey P, Salemi M & Vandamme A-M. (2009). *The Phylogenetic Handbook, 2nd Edition*. Cambridge University Press.
- Felsenstein J. (2003). *Inferring phylogenies*. Sinauer Associates

Computer Software: http://evolution.genetics.washington.edu/phylip/software.html

Phylodynamics™

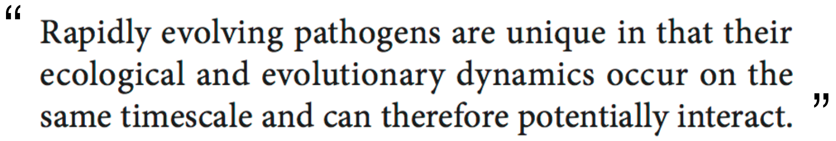


GENETIC DIVERSITY
(phylogenetics & molecular evolution)

EPIDEMIC DYNAMICS
(mathematical epidemiology)

NATURAL SELECTION
(population genetics & immunology)

Unifying principle



> “ Rapidly evolving pathogens are unique in that their ecological and evolutionary dynamics occur on the same timescale and can therefore potentially interact. ”

Pybus & Rambaut (2009) Nat. Rev. Genetics 10:540-50

Fundamental Phylodynamic Questions

- How genetically diverse is a pathogen population?
- How do pathogen genomes change through time?
- How does pathogen genetic diversity vary through time and space?
- What are the effects of pathogen genetic diversity on virulence, transmissibility, resistance to treatment, etc.

Specific questions

- When did a epidemic start?
- Where did it come from?
- How fast is it transmitting?
- In what direction is it spreading?
- Are hosts X, Y & Z epidemiologically linked?
- Of how many strains is the epidemic composed?
- Are strains associated with particular transmission routes?
- What adaptations has it accrued?

Fundamental Phylodynamic Questions

- How genetically diverse is a pathogen population?
- How do pathogen genomes change through time?
- How does pathogen genetic diversity vary through time and space?
- What processes and/or events determine these changes?
- What are the effects of pathogen genetic diversity on virulence, transmissibility, resistance to treatment, etc.

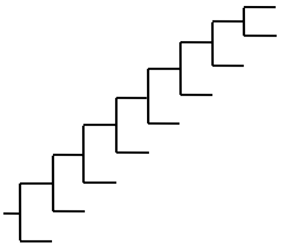
Measuring sequence diversity

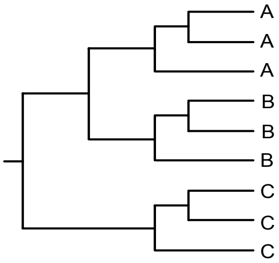
- Not as straightforward as you might think...
- Are your pathogen sequences all sampled at the same time?

  If sequences not sampled over time it's difficult to separate the effects of diversity and divergence on genetic diversity.

- Are you measuring sample diversity or population diversity?

  The former is simply a summary of your data, the latter is an inference about the population you have sampled. Sequences should be sampled randomly to estimate the latter.





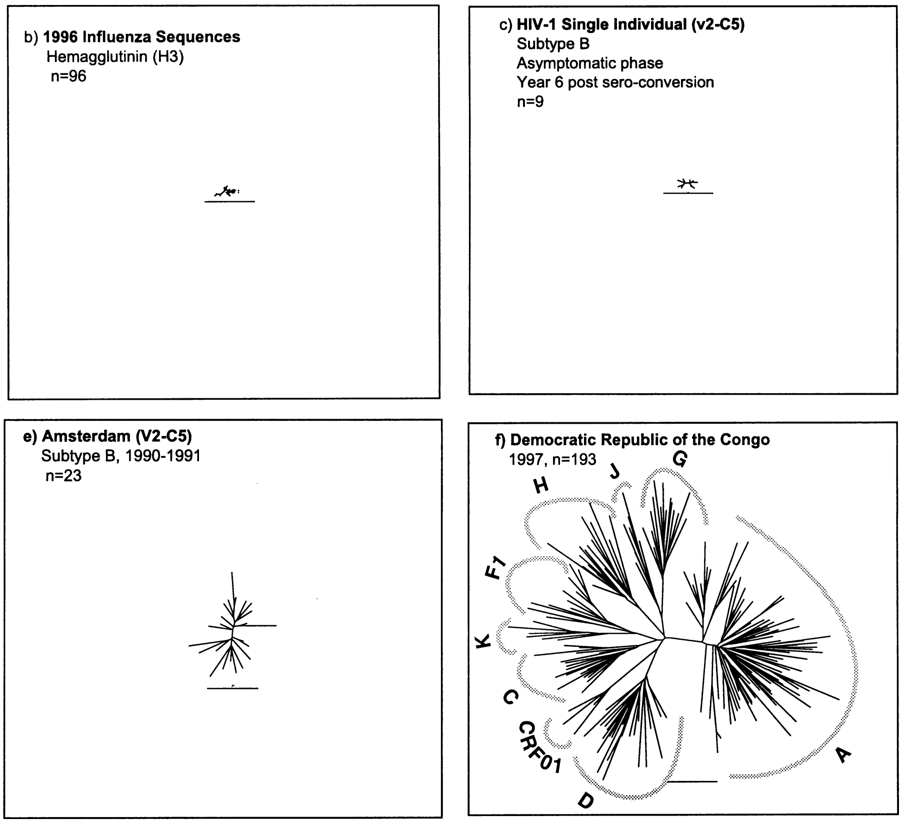
Measuring sequence diversity

- Are you studying an inter-host or intra-host population?

  For the former, each sequence represents a different infection. For the latter, each sequence represents a different virion within an infected individual. The measure of diversity must be interpreted accordingly.

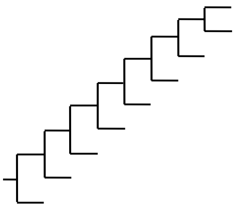
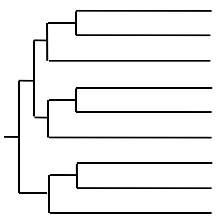
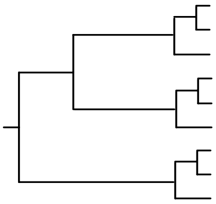
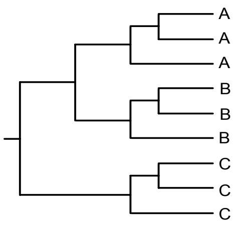
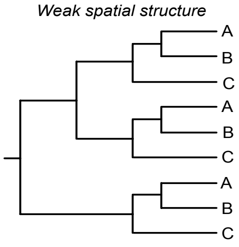
- How do we deal with intra-host diversity when studying the inter-host level?
- Intra-host diversity is low for most acute infections (e.g. influenza) but can be high for chronic infections (e.g. HIV).

Example: diversity of HIV-1 versus influenza


b) 1996 Influenza Sequences
Hemagglutinin (H3)
n=96
c) HIV-1 Single Individual (v2-C5)
Subtype B
Asymptomatic phase
Year 6 post sero-conversion
n=9
e) Amsterdam (V2-C5)
Subtype B, 1990-1991
n=23
f) Democratic Republic of the Congo
1997, n=193
H
J
G
F1
K
C
CRF01
D
A

Scale bar represents a genetic distance of 0.1 substitutions per site.
Korber et al. 2001. British Medical Bulletin 58:19-42

Phylodynamic Patterns

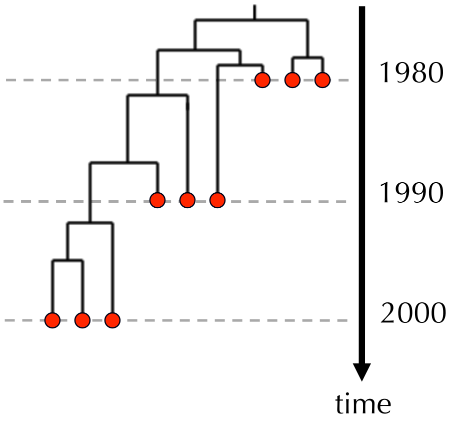
	Continual Immune Selection	**Weak/No Immune Selection**	
		Population dynamics	Spatial dynamics
Idealised Phylogeny Shapes		*Population growth*  *Population decline* 	*Strong spatial structure*  *Weak spatial structure* 
Examples	Human influenza A within-host HIV	among-host HIV among-host HCV	Measles Rabies, Dengue

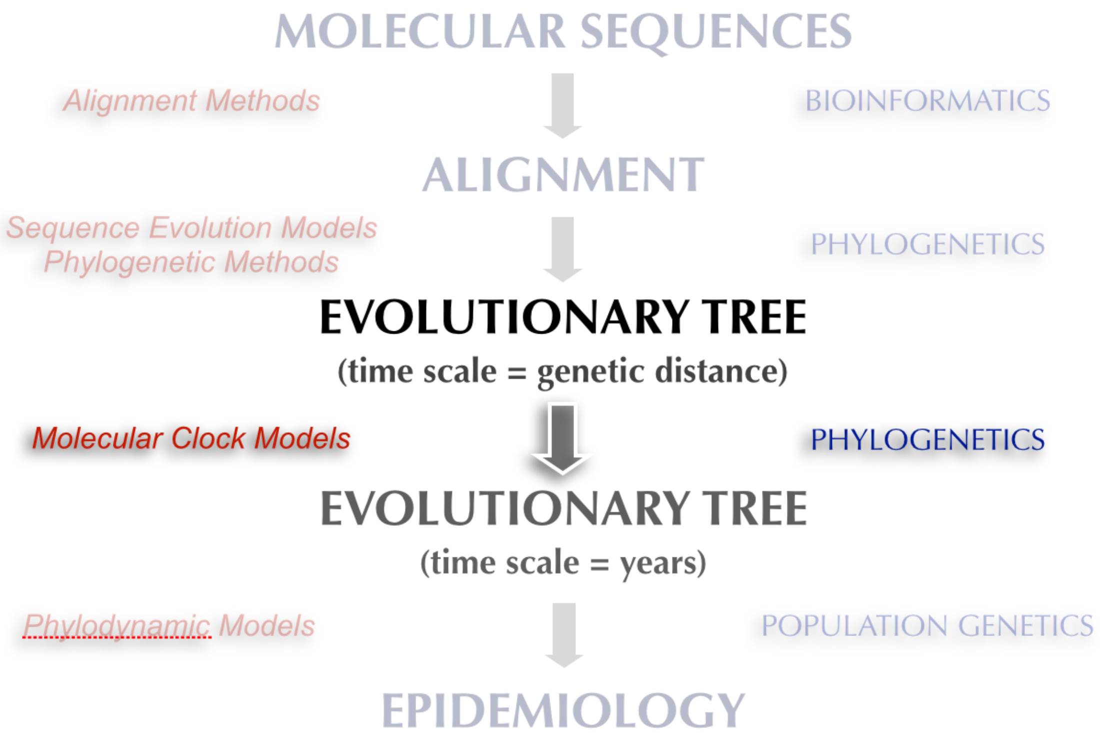
Fundamental Phylodynamic Questions

- How genetically diverse is a pathogen population?
- How do pathogen genomes change through time?
- How does pathogen genetic diversity vary through time and space?
- What processes and/or events determine these changes?
- What are the effects of pathogen genetic diversity on virulence, transmissibility, resistance to treatment, etc.

‘Phylodynamic’ Data

- Pathogen genomes are sampled at different points in time and from different locations.
- Hence transmission history is estimated on a real time-scale (e.g. years).





MOLECULAR SEQUENCES

Alignment Methods

BIOINFORMATICS

ALIGNMENT

Sequence Evolution Models
Phylogenetic Methods

PHYLOGENETICS

EVOLUTIONARY TREE
(time scale = genetic distance)

Molecular Clock Models

PHYLOGENETICS

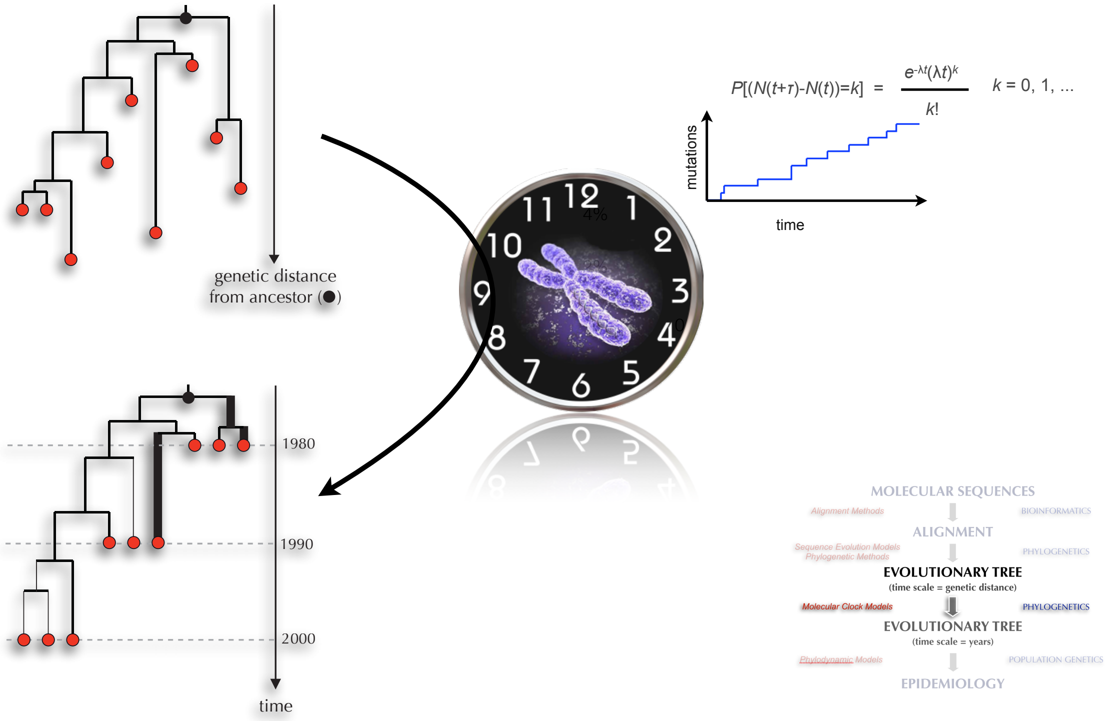
EVOLUTIONARY TREE
(time scale = years)

Phylodynamic Models

POPULATION GENETICS

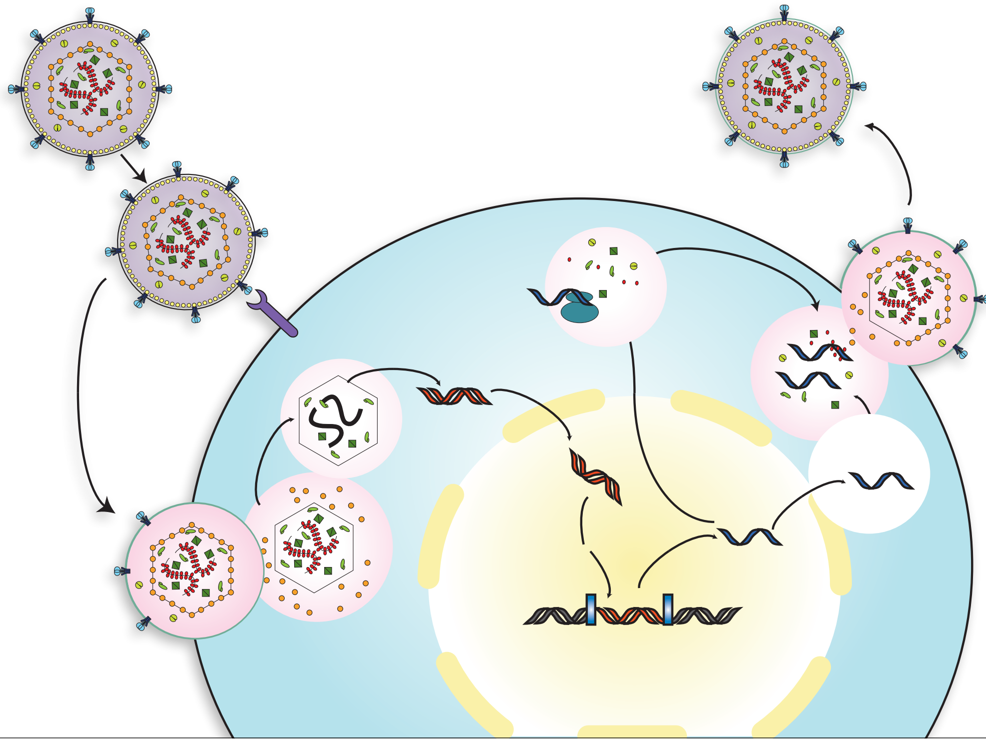
EPIDEMIOLOGY

Molecular clocks

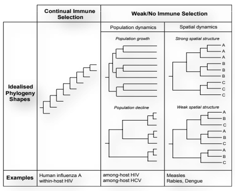


$$P[(N(t+\tau)-N(t))=k] = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \qquad k = 0, 1, \ldots$$

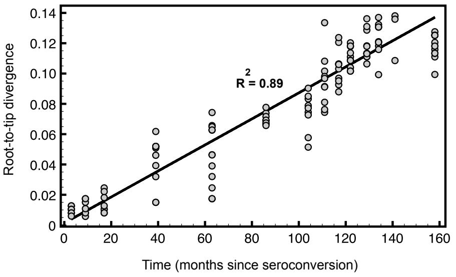
HIV: the ultimate evolver

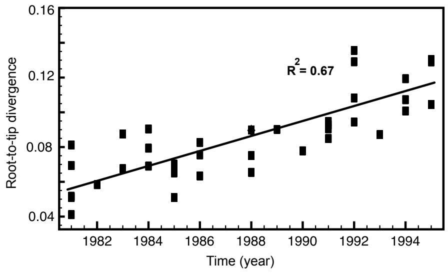


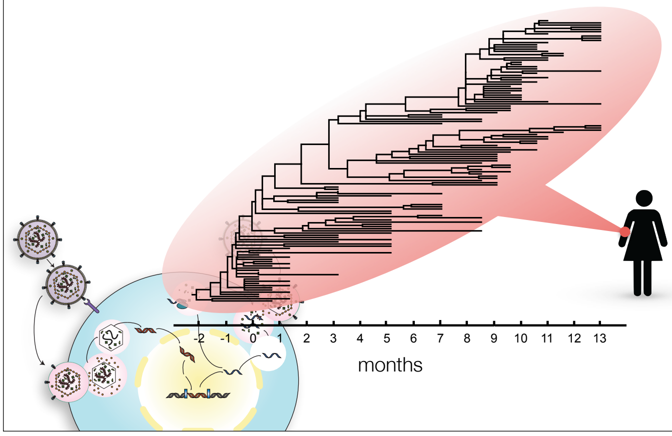
measurable evolution of HIV-1

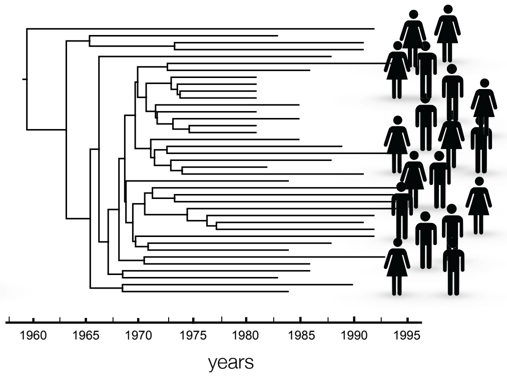


	Continual Immune Selection	Weak/No Immune Selection	
		Population dynamics	Spatial dynamics
Idealised Phylogeny Shapes		*Population growth* *Population decline*	*Strong spatial structure* *Weak spatial structure*
Examples	Human influenza A within-host HIV	among-host HIV among-host HCV	Measles Rabies, Dengue



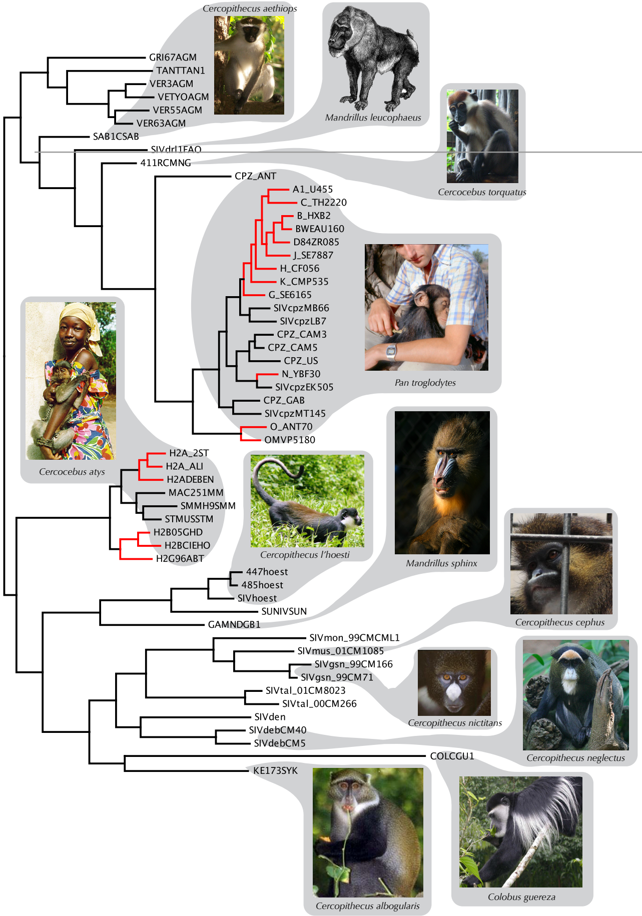


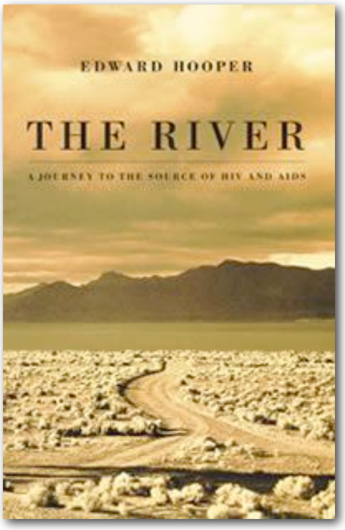




Lemey et al 2006 AIDS Rev

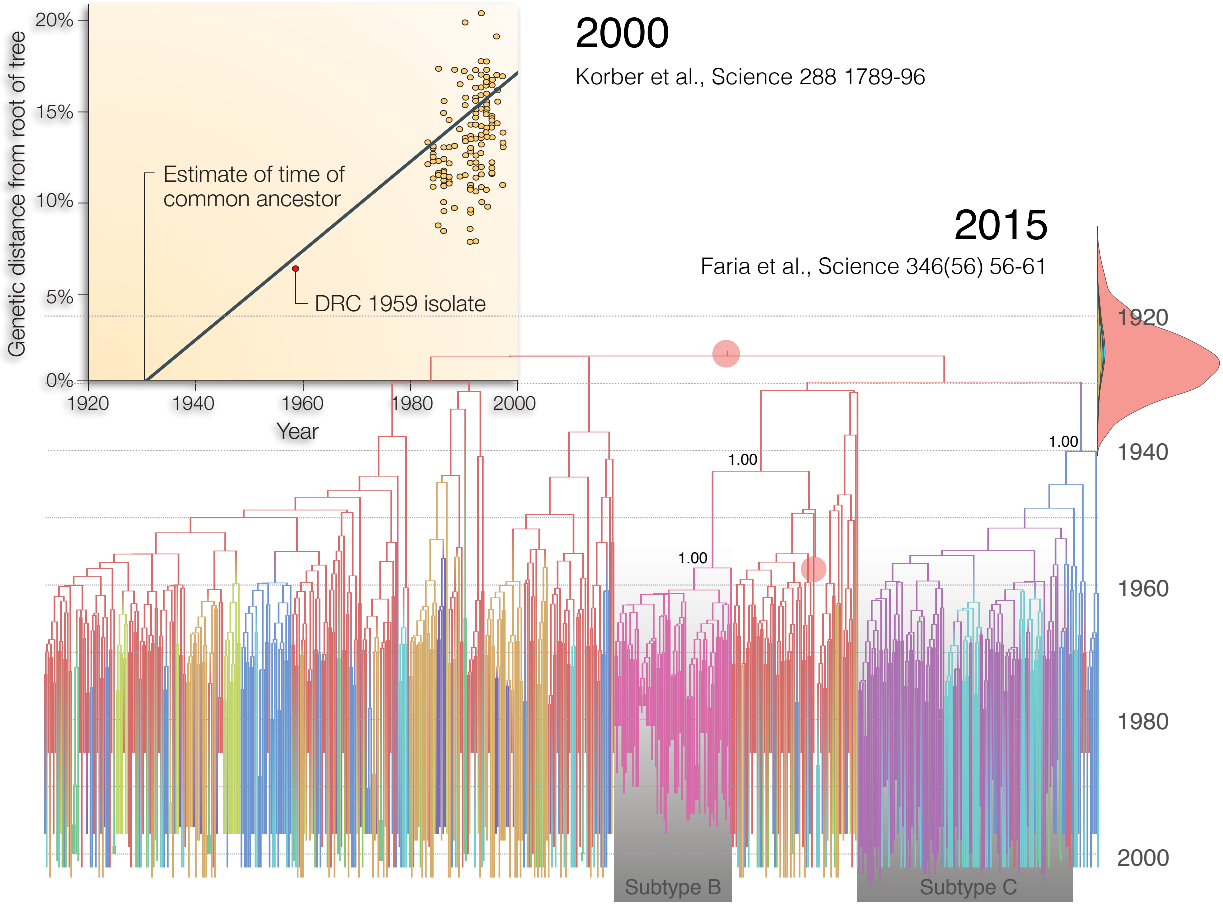
The origin of HIV-1




EDWARD HOOPER

THE RIVER

A JOURNEY TO THE SOURCE OF HIV AND AIDS



2000

Korber et al., Science 288 1789-96

2015

Faria et al., Science 346(56) 56-61

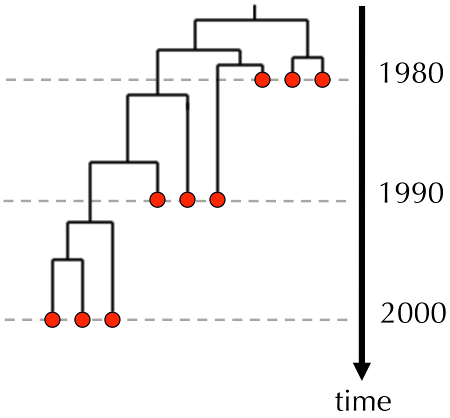
‘Phylodynamic’ Data

- Pathogen genomes are sampled at different points in time and from different locations.
- Hence transmission history is estimated on a real time-scale (e.g. years).
- The ability to genetically distinguish sequences sampled at different times depends on:

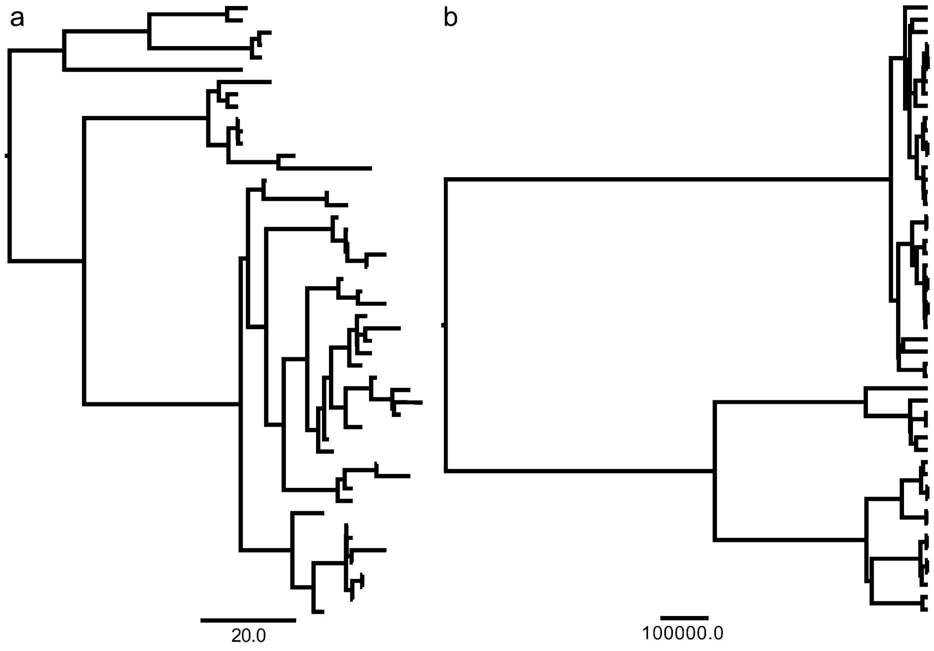
  (i) the rate of evolution of the gene/genome that is obtained

  (ii) the length of time between samples

  (iii) the sequence length of the gene/genome that is obtained



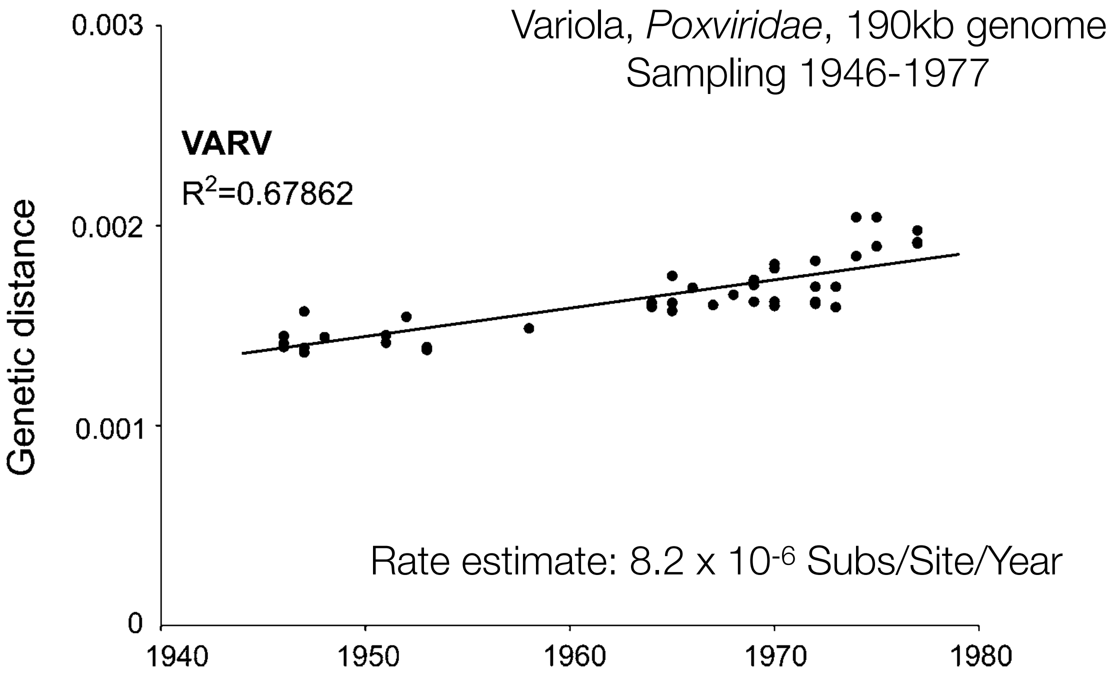
‘Phylodynamic’ Data



10^{-4} subs/site/year

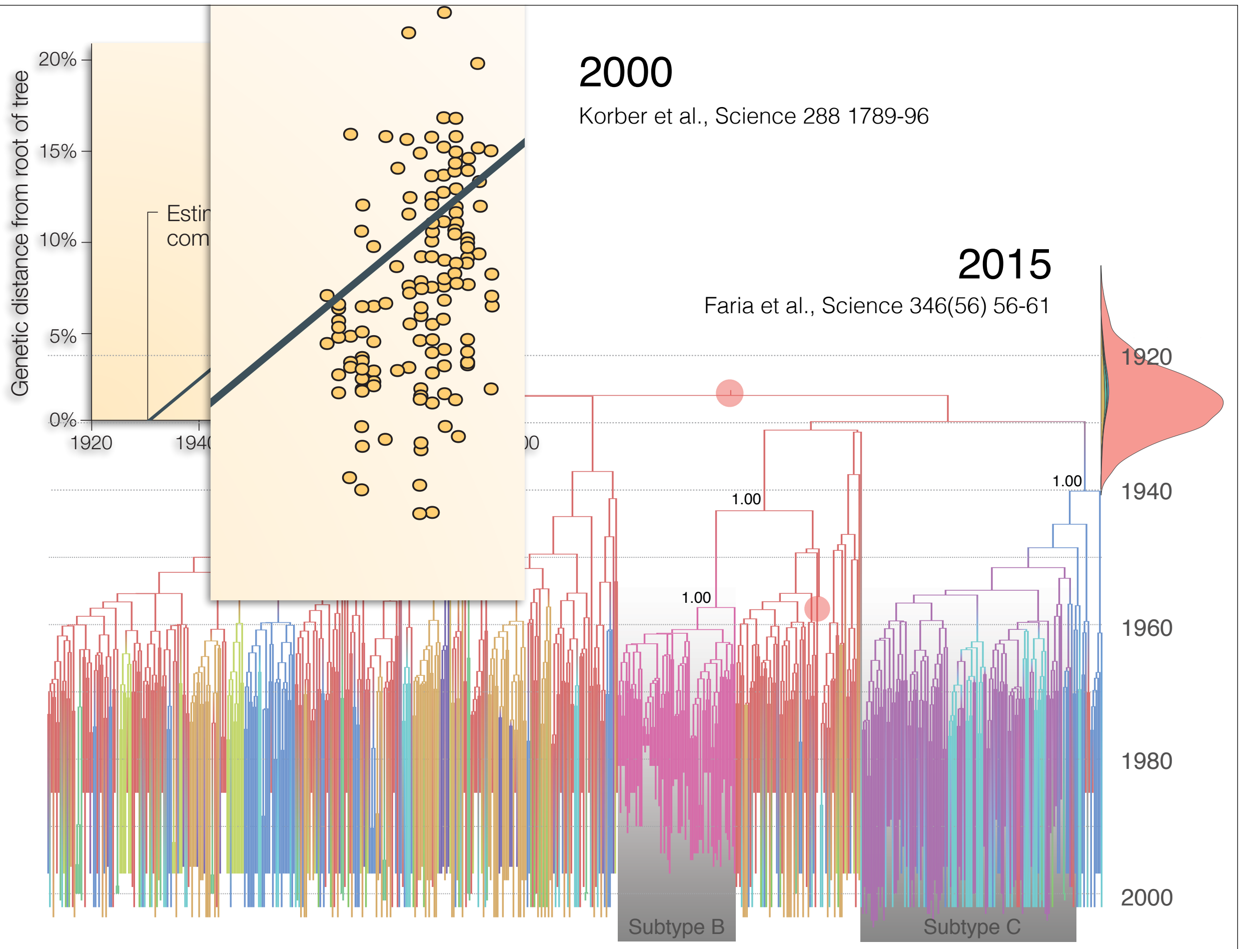
10^{-8} subs/site/year

A DNA virus (smallpox)



Variola, *Poxviridae*, 190kb genome
Sampling 1946-1977

Rate estimate: 8.2 x 10^{-6} Subs/Site/Year

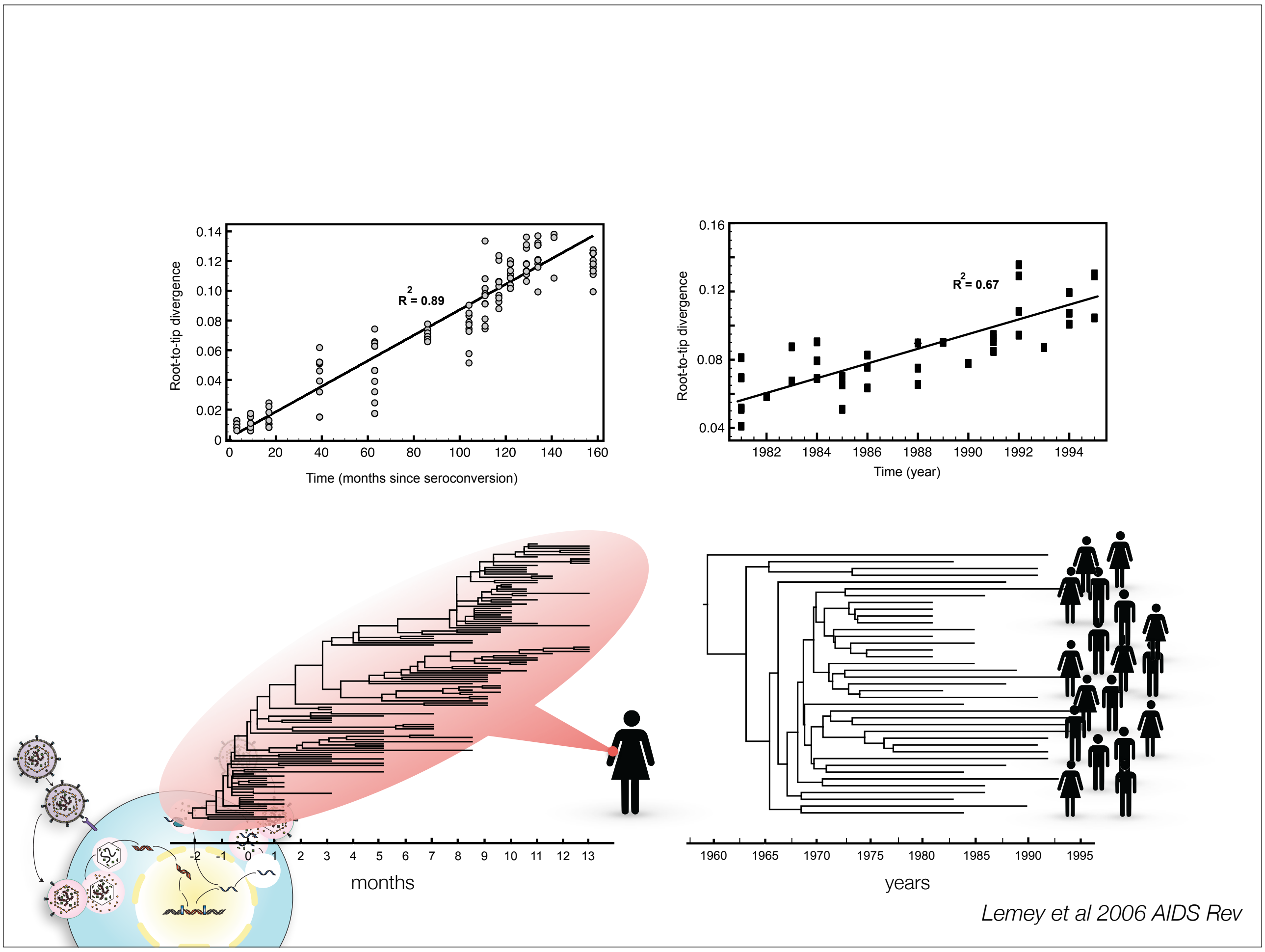


2000

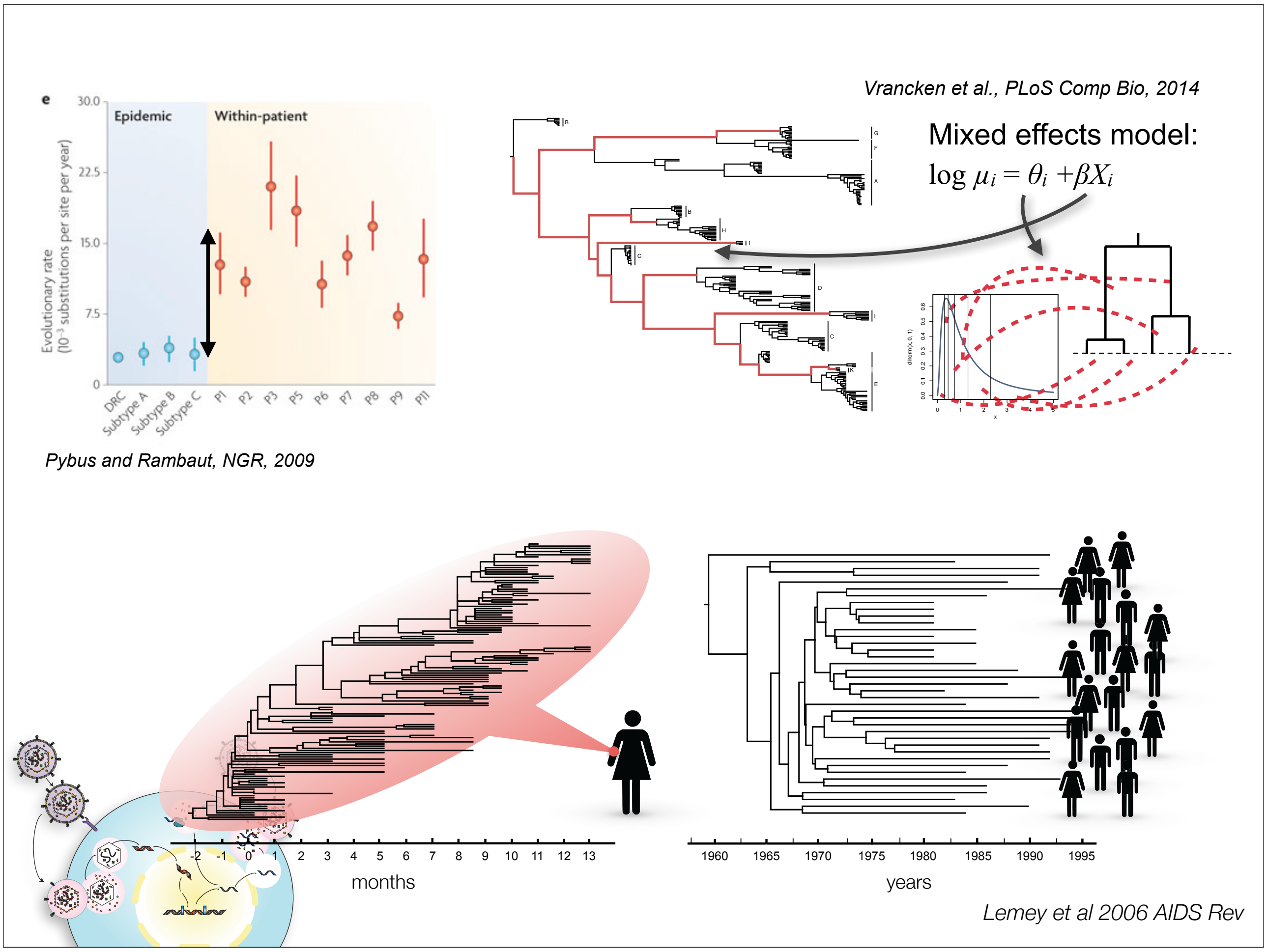
Korber et al., Science 288 1789-96

2015

Faria et al., Science 346(56) 56-61



Lemey et al 2006 AIDS Rev



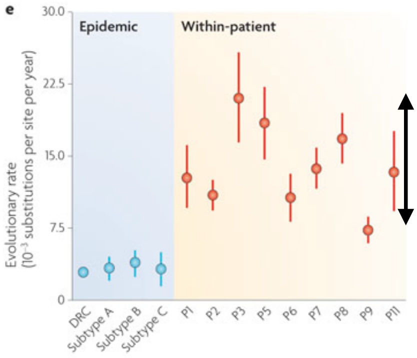
Vrancken et al., PLoS Comp Bio, 2014

Mixed effects model:

$\log \mu_i = \theta_i + \beta X_i$

Pybus and Rambaut, NGR, 2009

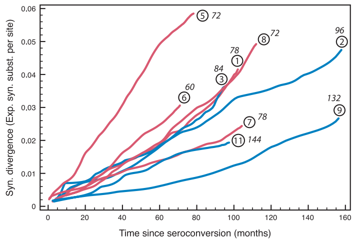
Lemey et al 2006 AIDS Rev



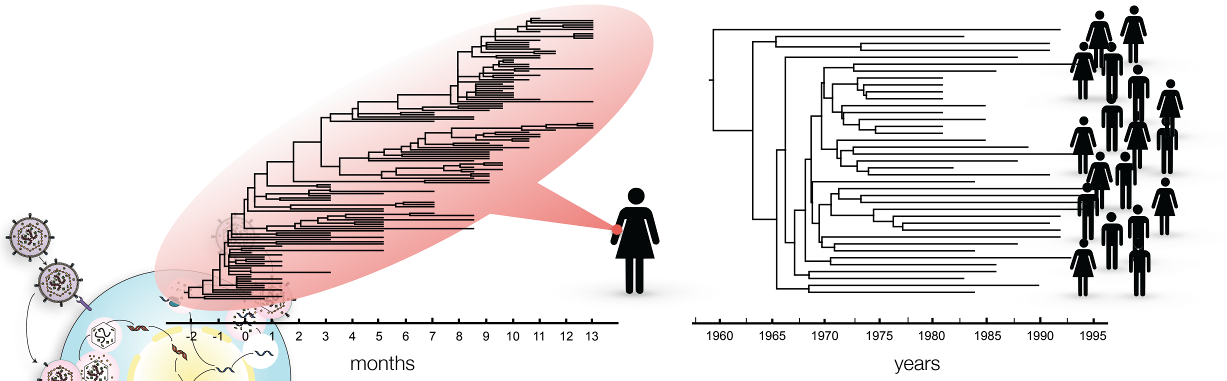
Pybus and Rambaut, NGR, 2009

Edo-Matas et al., Mol Biol Evol, 2011

$$\log\theta_i = \beta_0 + \delta_{\mathrm{LTNP}}\beta_{\mathrm{LTNP}}\mathrm{LTNP}_i + \delta_{\Delta 32}\beta_{\Delta 32}\Delta 32_i + \varepsilon_i$$

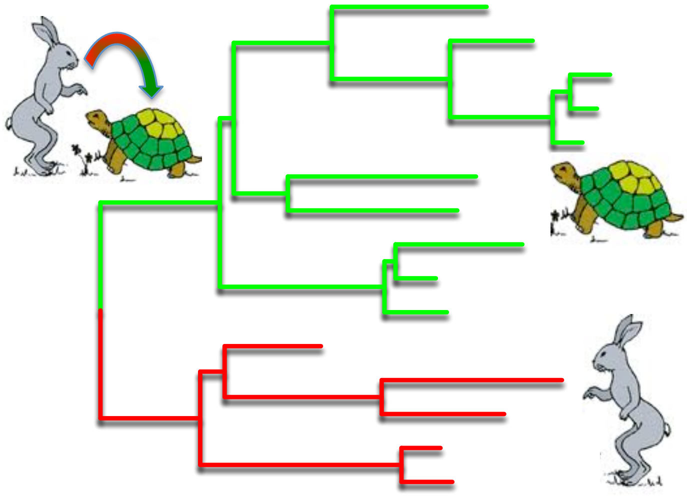


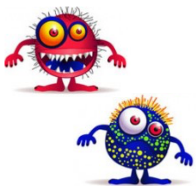
Lemey et al., PLoS Comp Bio, 2007



Lemey et al 2006 AIDS Rev

What drives the tempo of pathogen evolution?



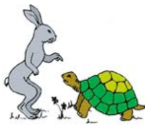


Pathogen factors

Mutation rate

Life cycle/replication dynamics

Host factors

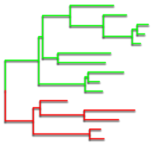


Life history

Seasonality

Metabolic rate etc.

Historical factors

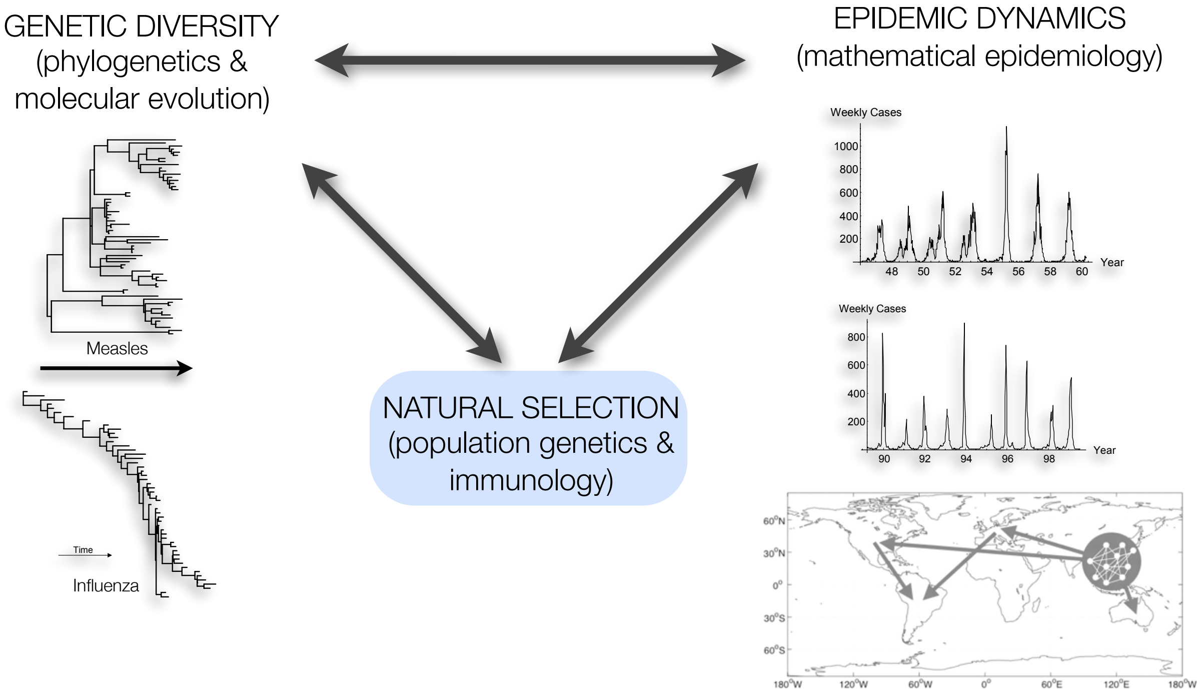


Pathogen phylogeny

Fundamental Phylodynamic Questions

- How genetically diverse is a pathogen population?
- How do pathogen genomes change through time?
- How does pathogen genetic diversity vary through time and space?
- What processes and/or events determine these changes?
- What are the effects of pathogen genetic diversity on virulence, transmissibility, resistance to treatment, etc.

Phylodynamics™



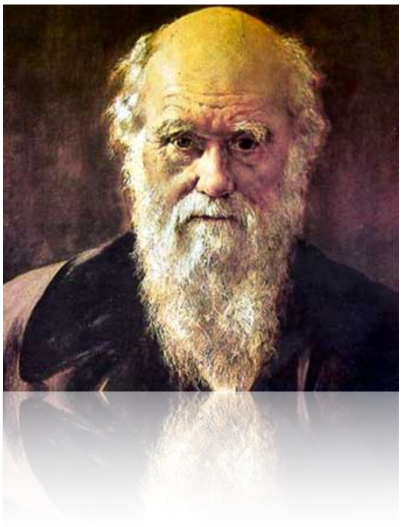
GENETIC DIVERSITY (phylogenetics & molecular evolution)

EPIDEMIC DYNAMICS (mathematical epidemiology)

NATURAL SELECTION (population genetics & immunology)

Evolutionary processes: natural selection

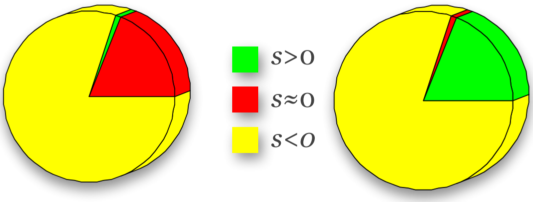
- “the preservation of favourable variations and the rejection of injurious variations, i call natural selection. variations neither useful nor injurious would not be affected by natural selection, and would be left a fluctuating element”
  - darwin, the origin of species



Evolutionary processes: natural selection

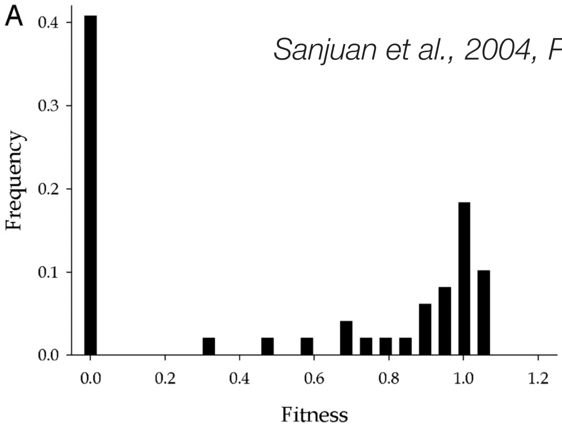
most fixed mutations are neutral

neutralist model
motoo kimura



most fixed mutations are advantageous

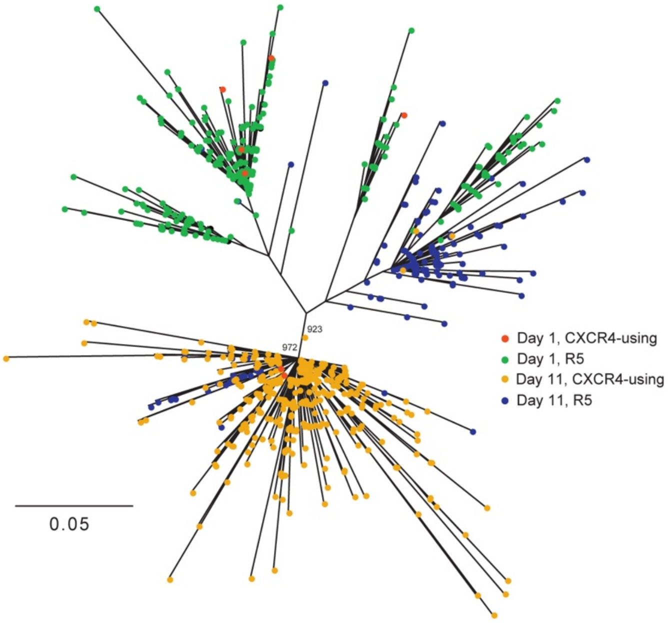
selectionist model
john gillespie



Sanjuan et al., 2004, PNAS

Evolutionary processes: natural selection

- Immune escape (antibodies*, T-cells*, innate immune responses)
- Antiviral drug resistance
- Vaccine escape mutations
- Cell & tissue tropism
- Inter-host viral transmission (i.e. for viral emergence)

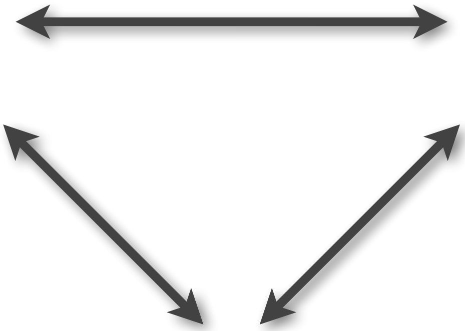


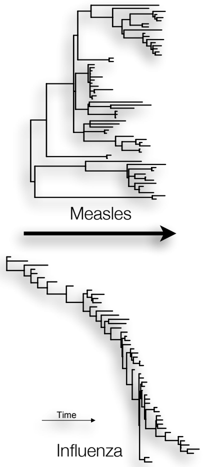
module 15: Pathogen evolution, selection and immunology

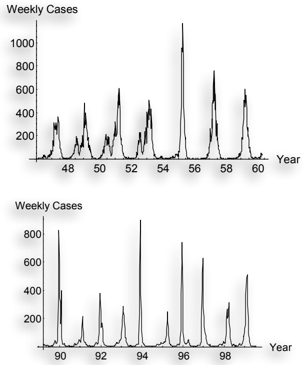
Phylodynamics™

GENETIC DIVERSITY (phylogenetics & molecular evolution)

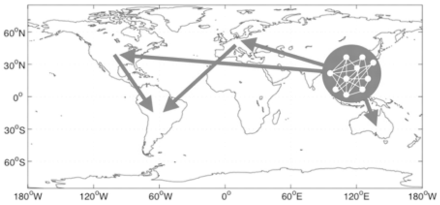
EPIDEMIC DYNAMICS (mathematical epidemiology)

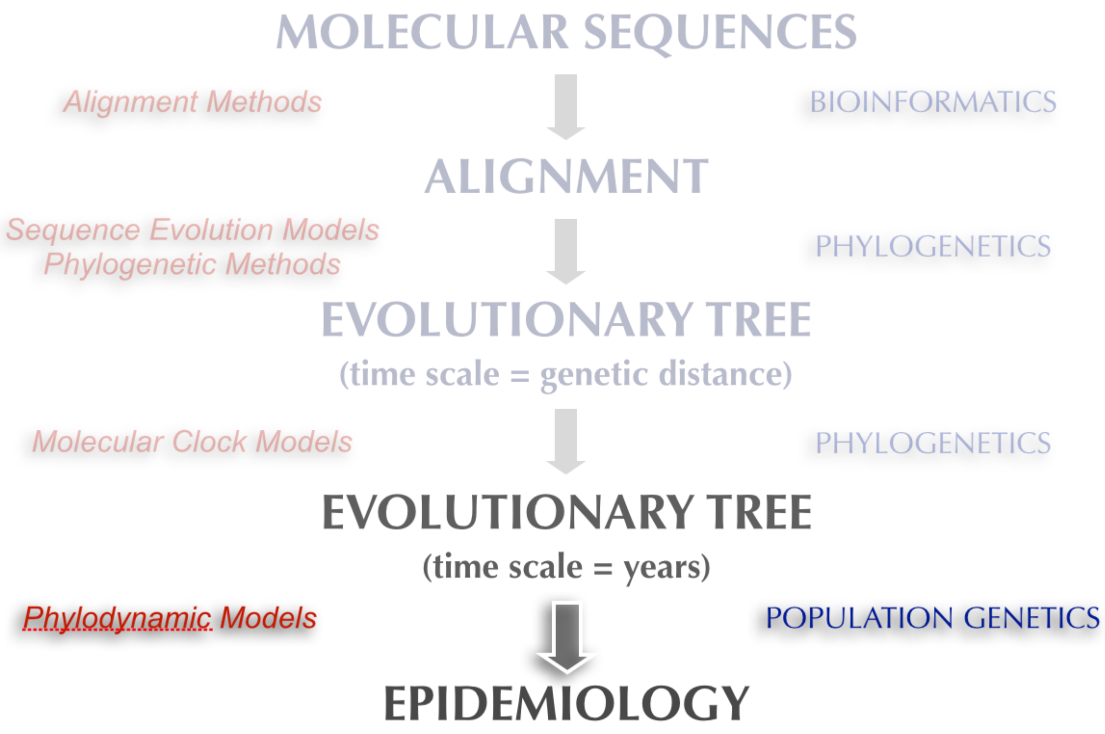






NATURAL SELECTION (population genetics & immunology)





MOLECULAR SEQUENCES

Alignment Methods

BIOINFORMATICS

ALIGNMENT

Sequence Evolution Models
Phylogenetic Methods

PHYLOGENETICS

EVOLUTIONARY TREE
(time scale = genetic distance)

Molecular Clock Models

PHYLOGENETICS

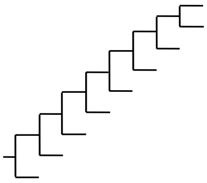
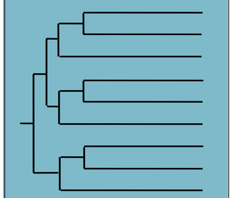
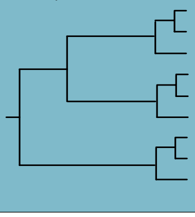
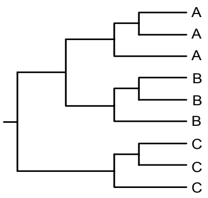
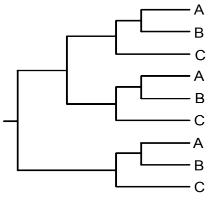
EVOLUTIONARY TREE
(time scale = years)

Phylodynamic Models

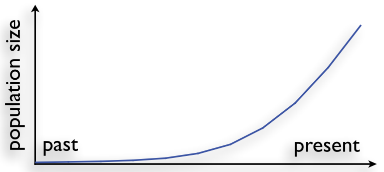
POPULATION GENETICS

EPIDEMIOLOGY

Phylodynamic Patterns

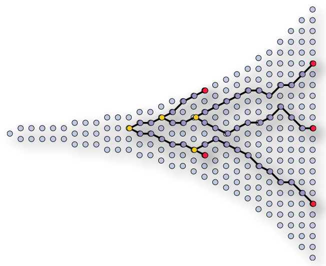
	Continual Immune Selection	**Weak/No Immune Selection**	
		Population dynamics	Spatial dynamics
Idealised Phylogeny Shapes		*Population growth*  *Population decline* 	*Strong spatial structure*  *Weak spatial structure* 
Examples	Human influenza A within-host HIV	among-host HIV among-host HCV	Measles Rabies, Dengue

Demography and coalescent theory



- The rate at which lineages 'coalesce' depends on population size and population structure.

Kingman JFC (1982) Journal of Applied Probability 19A:27–43



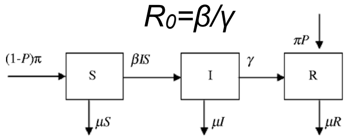
- Population dynamics can be reconstructed using parametric or flexible nonparametric models (the 'skyline or skyride plot' method)

Pybus et al. (2000) Genetics 155:1429-37

Drummond, Rambaut, Shapiro & Pybus (2005) Mol Biol Evol 22:1185-92

Minin, Bloomquist and Suchard (2008) Mol Biol Evol 25:1459-71

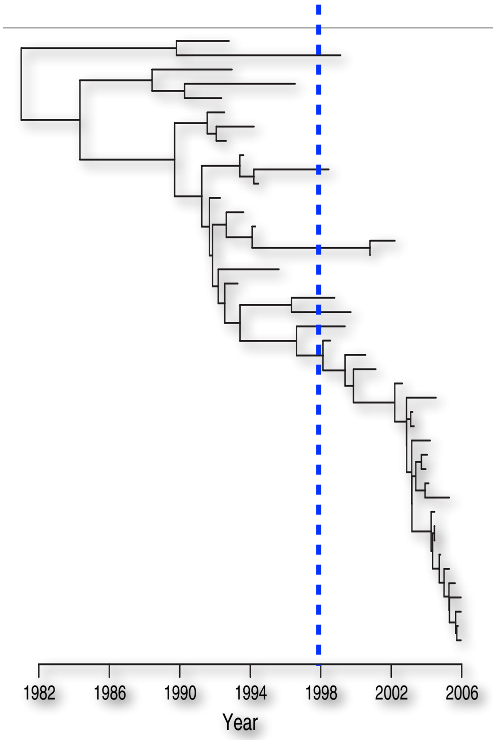
$R_0=\beta/\gamma$

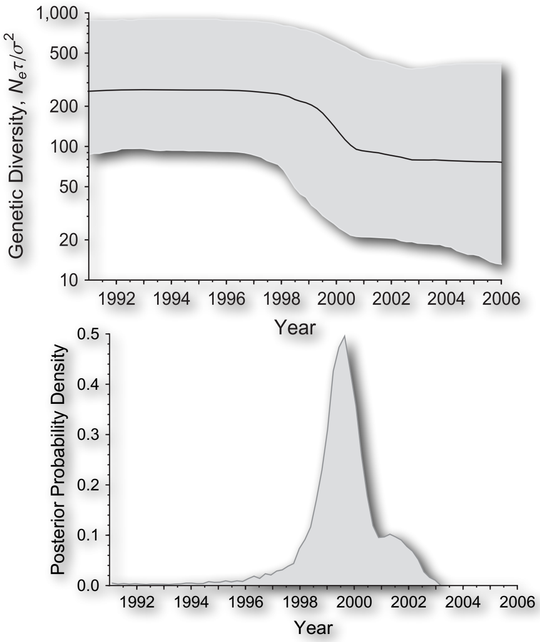


- Birth-death models can also be used as the tree-generative model and just like coalescent models they can be parametrized in terms of compartmental epidemic models.

Stadler et al. (2012) MBE 29:347-357

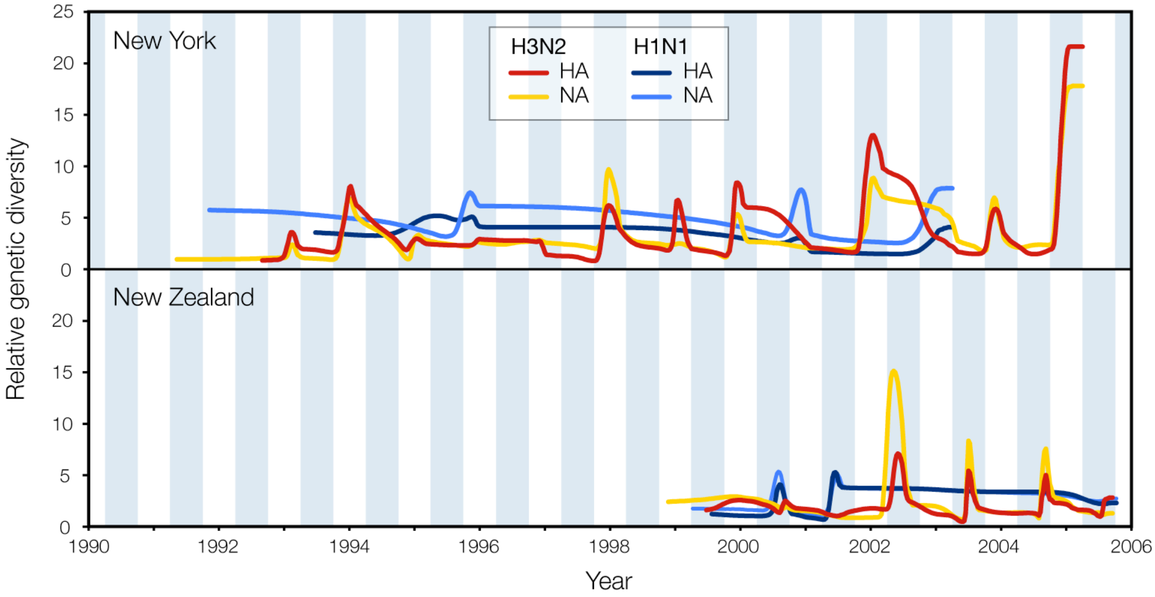
HBV Vaccination in Amsterdam





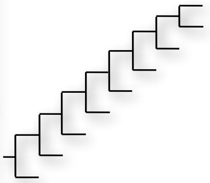
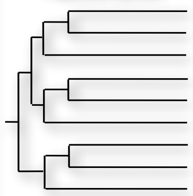
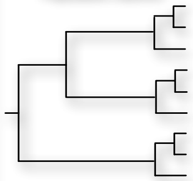
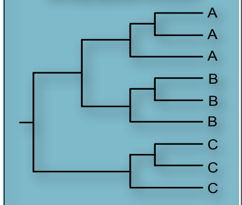
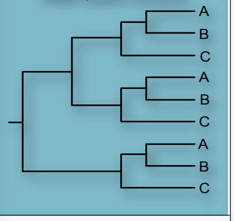
van Ballegooijen et al. 2009. Am. J. Epidemiol. 170:1455-63

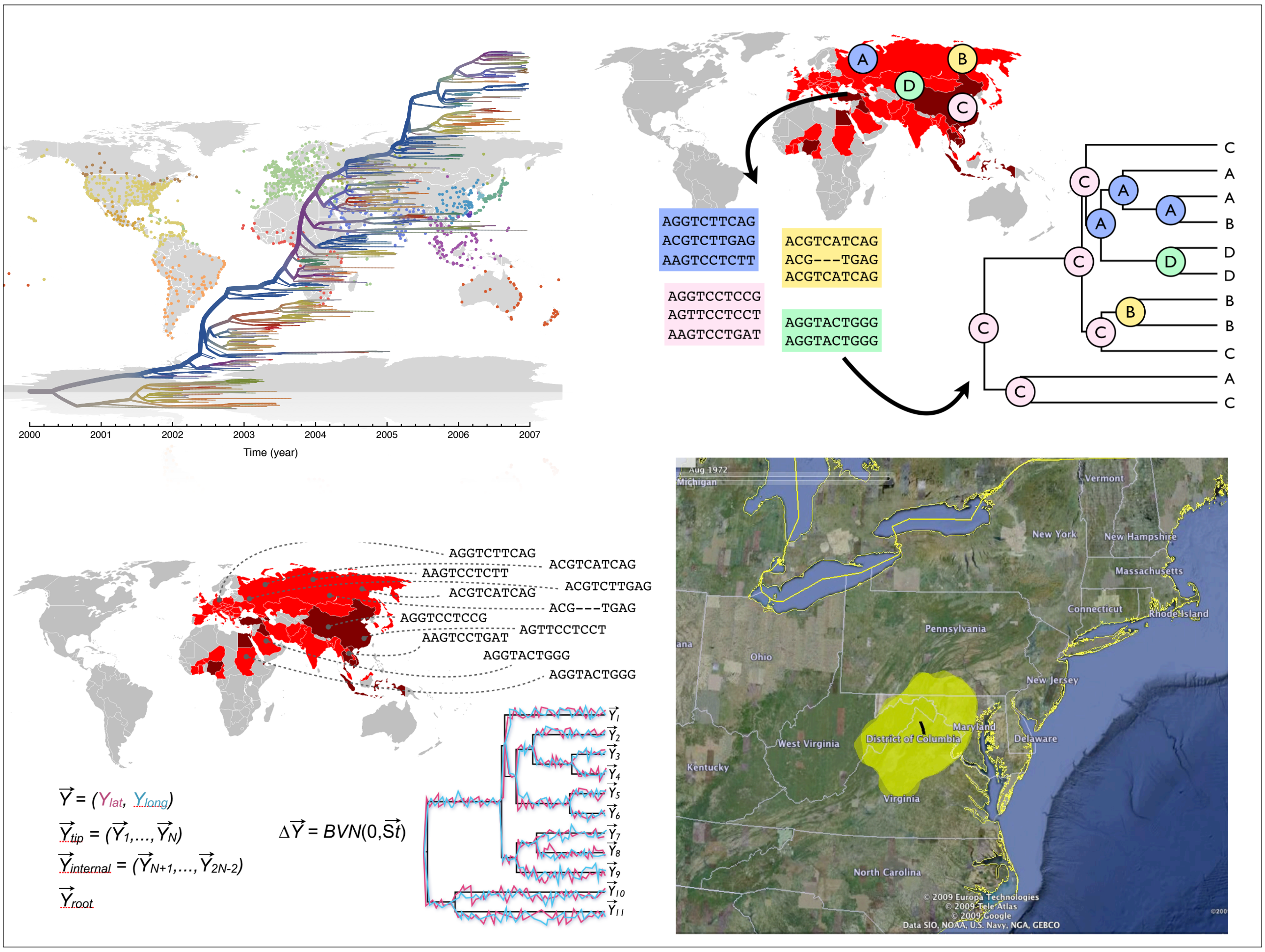
Influenza H3N2 epidemic dynamics

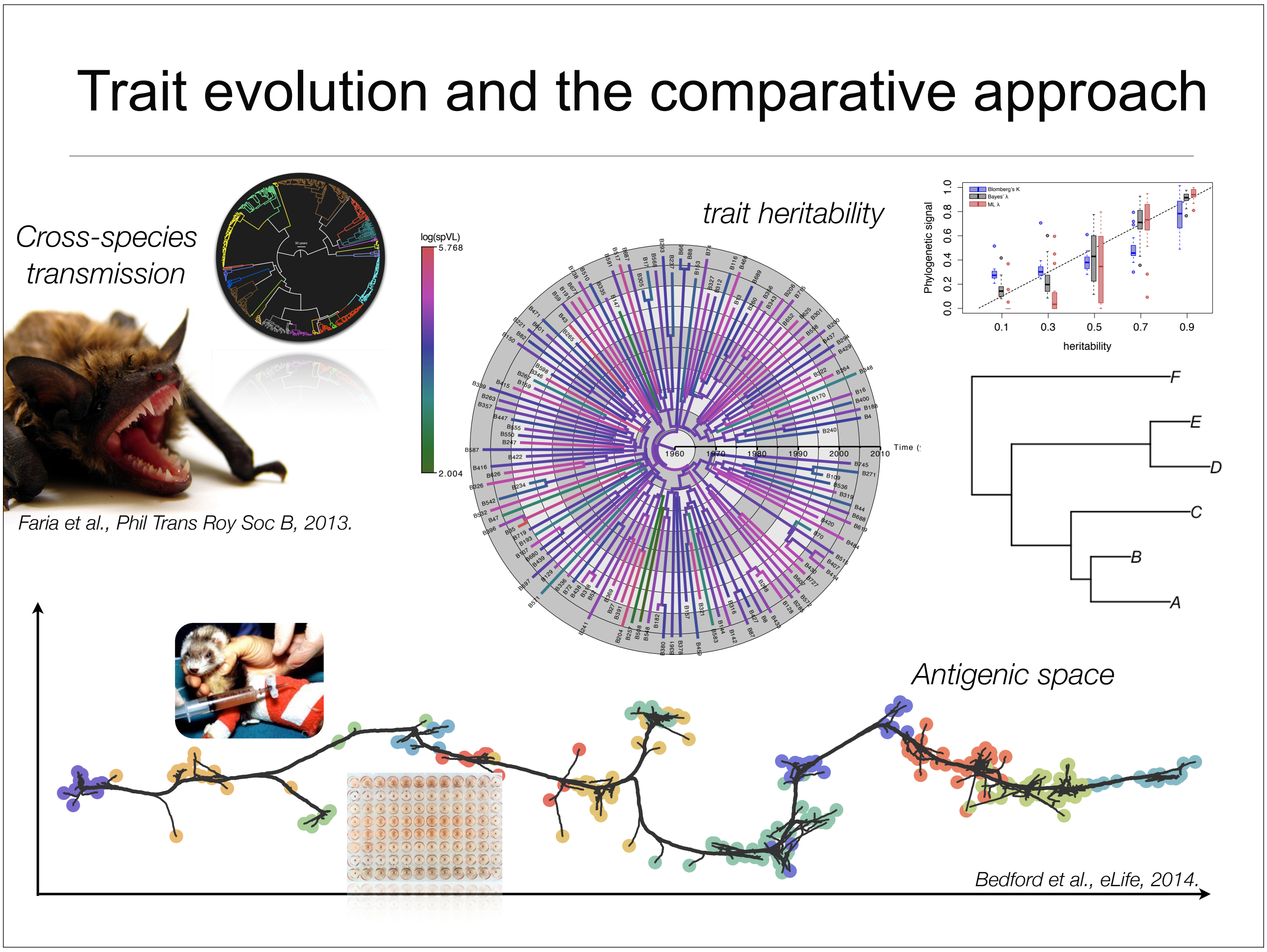


Rambaut et al. 2009. Nature

PhyloGEOdynamic Patterns

	Continual Immune Selection	Weak/No Immune Selection	
		Population dynamics	Spatial dynamics
Idealised Phylogeny Shapes		*Population growth*  *Population decline* 	*Strong spatial structure*  *Weak spatial structure* 
Examples	Human influenza A within-host HIV	among-host HIV among-host HCV	Measles Rabies, Dengue





Trait evolution and the comparative approach

Cross-species transmission

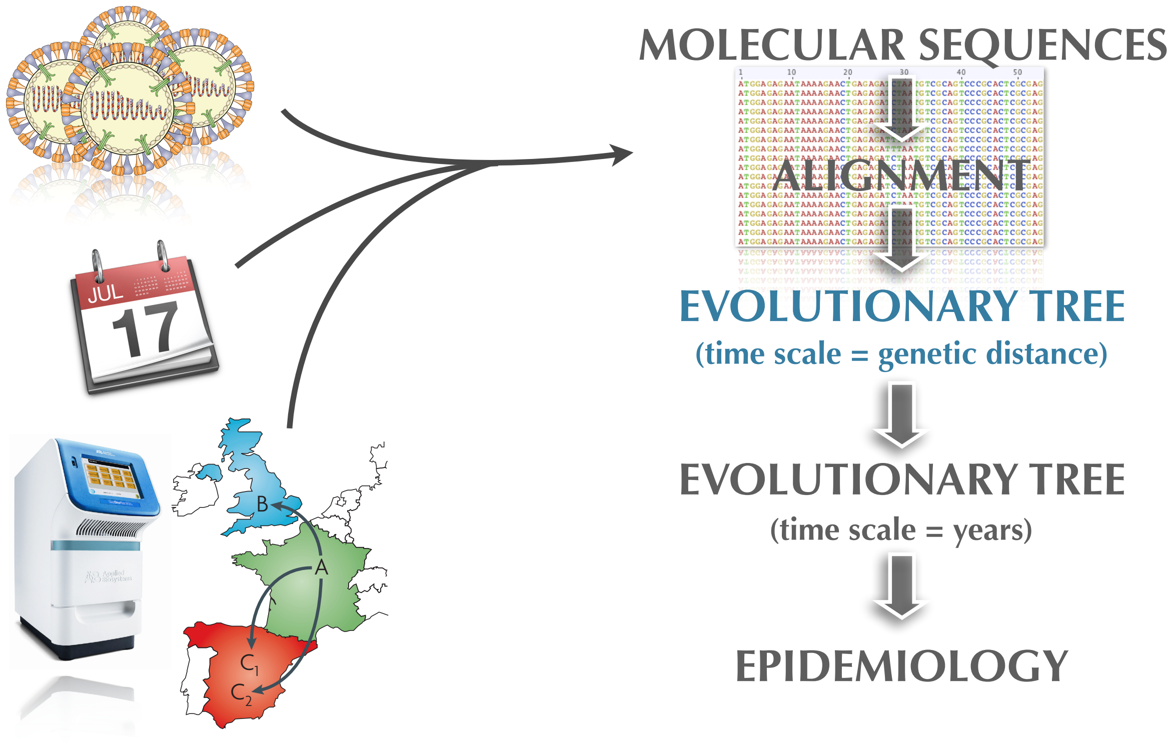
Faria et al., Phil Trans Roy Soc B, 2013.

trait heritability

Antigenic space

Bedford et al., eLife, 2014.

Bayesian Evolutionary Analysis Sampling Trees



MOLECULAR SEQUENCES

ALIGNMENT

EVOLUTIONARY TREE

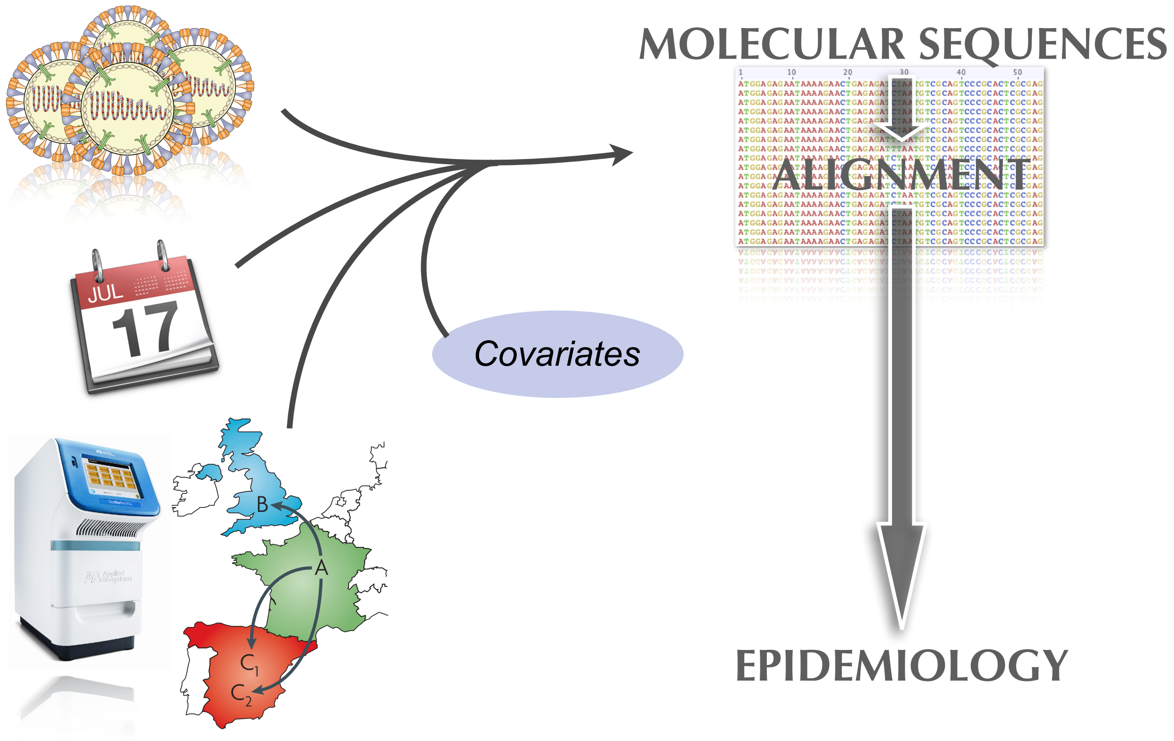
(time scale = genetic distance)

EVOLUTIONARY TREE

(time scale = years)

EPIDEMIOLOGY

Bayesian Evolutionary Analysis Sampling Trees



MOLECULAR SEQUENCES

ALIGNMENT

Covariates

EPIDEMIOLOGY