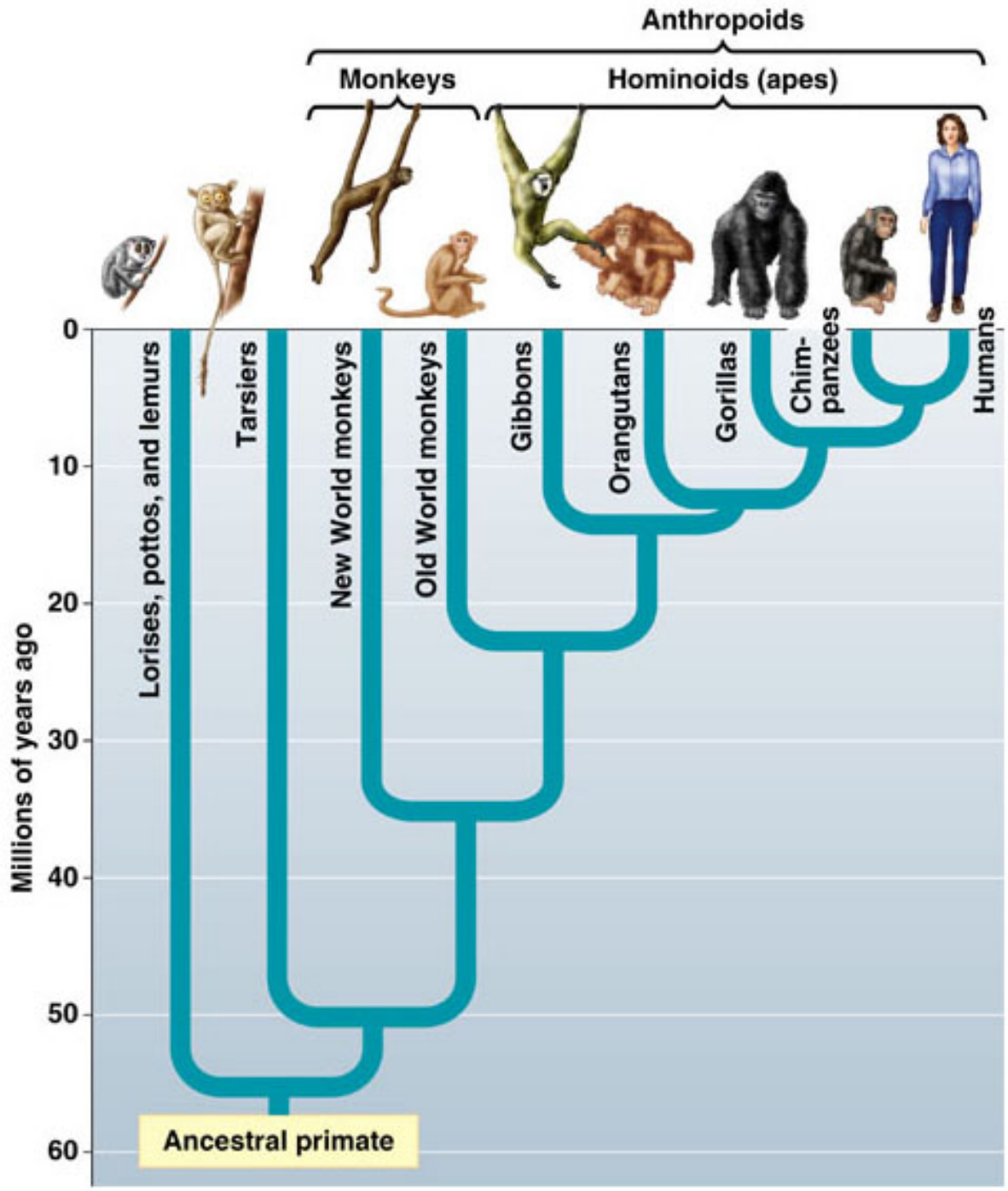
Key concepts in this section

- What is molecular epidemiology and differences Vs epidemiology
- What are phylogenies or phylogenetic trees?
  - Terminology such as extant, ancestral, branch point, branch length
- Why build phylogenetic trees?
  - Algorithms to build phylogenetic trees
    - Distance-based methods
    - Parsimony methods
      - Minimize the number of changes
    - Probabilistic methods
      - Tree that best explains the data using probabilistic models
- Viral infections – Viral epidemics
  - origin, expansion, geographic distribution, evolution

Origin of homo sapiens



Why phylogenetic trees?

- Inform multiple sequence alignments
- Identify signatures of conservation of sequences
- Understand how organisms are related
  - Do humans and chimpanzees share a more recent common ancestor than do humans and gorillas ?
- Ask how closely organisms are related
  - Humans and chimpanzees shard a common ancestor 5 mya
- How specific functions/traits have evolved
  - What made us human?

Σημασία στοίχισης αλληλουχιών

- Προκειμένου να εκτιμήσουμε σχέσεις μεταξύ του ίδιου ή διαφορετικών οργανισμών, από αλληλουχίες DNA ή/και πρωτεϊνών, θα πρέπει να υπολογίσουμε το βαθμό ετερογένειας *(heterogeneity)* μεταξύ τους.

- Προκειμένου να το επιτύχουμε αυτό, θα πρέπει να συγκρίνουμε περιοχές DNA που αντιστοιχούν σε ομόλογες *(homologous)* περιοχές.

- Για παράδειγμα, για να μελετήσουμε την προέλευση του homo sapiens sapiens με διάφορα πιθηκοειδή συγκρίνουμε τα μορφολογικά του χαρακτηριστικά (π.χ μέγεθος εγκεφάλου, θώρακας, ανατομία πέλματος κλπ) με τα αντίστοιχα από τα διαφορετικά ανθρωποειδή (π.χ χιμπαντζής, γορίλας κλπ).

Παράδειγμα (Ι)

- Κατ’ αντιστοιχία με τα μορφολογικά χαρακτηριστικά προκειμένου να συγκρίνουμε αλληλουχίες DNA ή πρωτεϊνών από τον ίδιο ή διαφορετικούς οργανισμούς, θα πρέπει να συγκρίνουμε ομόλογες περιοχές DNA ή πρωτεϊνών, ή με άλλα λόγια περιοχές που κωδικοποιούν για παρόμοιες λειτουργίες.

- Παράδειγμα:

• Ἄνθρωπος	1. PHYLOGENY 2. POSTGRADUATE
• Πίθηκος	1. PHILOGENY 2. POSTGRADIATE
• Κοτόπουλο	1. PHELOSINE 2. POSTPROJECTS

Παράδειγμα (II)

- Για να μελετήσουμε την εξελικτική σχέση μεταξύ των οργανισμών με διαθέσιμες αλληλουχίες θα πρέπει να συγκρίνουμε το DNA τους από ομόλογες περιοχές.

- παράδειγμα

• Ἄνθρωπος	1. PHYLOGENY 2. POSTGRADUATE
• Πίθηκος	1. PHILOGENY 2. POSTGRADIATE
• Πουλερικά	1. PHELOSENE 2. POSTPROJECTS
	** ** ** **** * *

Παράδειγμα (III)

- Και όχι μεταξύ διαφορετικών περιοχών, από τις οποίες δεν θα μπορούσε να βγει κανένα συμπέρασμα

- παράδειγμα

• Ἄνθρωπος	1. PHYLOGENY
• Πίθηκος	2. POSTGRADIATE
• Πουλερικά	1. PHELOSENE

Η στοίχιση προηγείται πάντα της ανάλυσης

- Οι αλληλουχίες DNA ή πρωτεϊνών αποτελούνται από σειρές χαρακτήρων, συνεπώς, προκειμένου να μπορούμε να τις συγκρίνουμε θα πρέπει να βρούμε τις ομόλογες περιοχές για κάθε μια από τις αλληλουχίες και να τις στοιχίσουμε *(align)* τη μια κάτω από την άλλη.

- Η διαδικασία της εύρεσης και στοίχισης των ομόλογων περιοχών χαρακτηρίζεται ως στοίχιση–συστοιχία αλληλουχιών *(alignment).*

- Προκειμένου να πραγματοποιήσουμε οποιουδήποτε είδους φυλογενετικές αναλύσεις θα πρέπει οπωσδήποτε, πριν, να στοιχίσουμε τις αλληλουχίες DNA ή πρωτεϊνών, που πρόκειται να αναλυθούν/μελετηθούν.

Στοίχιση αλληλουχιών *(Alignment)*

- Κατά τη στοίχιση αλληλουχιών μπορεί να παρατηρούνται αντικαταστάσεις, ενθέσεις ή και διαγραφές μεταξύ των διαφορετικών αλληλουχιών.

- παράδειγμα

```
1  ACACT-GATCGA
2  ATACT-CATCGA
3  ACACTCGATCGA
4  ACACT-GAT-GA
     * ***  ** **
```

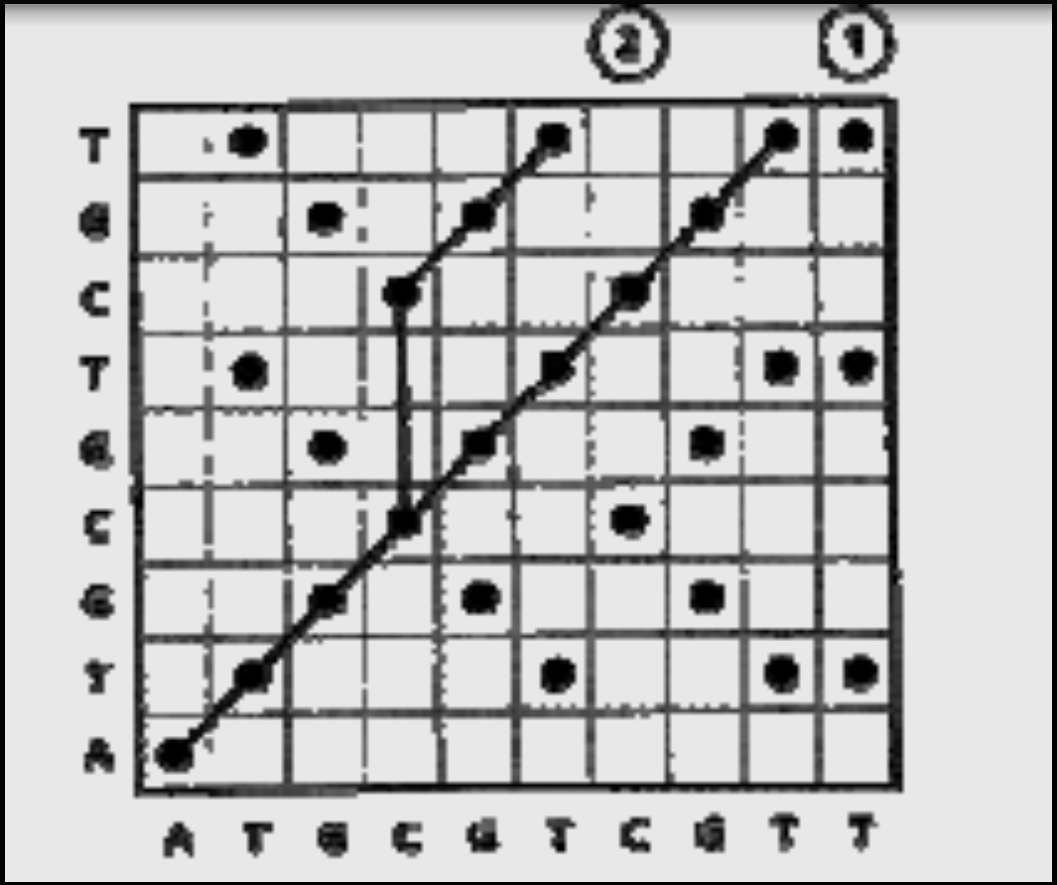
Αλγόριθμοι στοίχισης

- Για να γίνει στοίχιση των ομόλογων περιοχών DNA ή πρωτεινών, έχουν αναπτυχθεί αλγόριθμοι που βασίζονται στο κριτήριο βέλτιστης στοίχισης.
- Συγκεκριμένα υπολογίζεται μια ποσότητα *(score)* της συστοιχίας, την οποία ο αλγόριθμος προσπαθεί να βελτιστοποιήσει

- Για παράδειγμα ας υποθέσουμε 2 αλληλουχίες:

```
Seq1  ATGCGTCGTT
Seq2  ATGCGTCGT
```

Διάγραμμα κουκίδων των αλληλουχιών




```
Seq1    ATGCGTCGTT
        ||||||||||
Seq2    ATGCGTCGT
```

- Από το διάγραμμα προκύπτει

Εναλλακτική στοίχιση

- Παρότι η παραπάνω στοίχιση φαίνεται ιδανική, υπάρχουν και άλλες εναλλακτικές στοιχίσεις, όπως φαίνεται τόσο από το διάγραμμα κουκίδων. Συγκεκριμένα η εναλλακτική “διαδρομή” 2 αντιστοιχεί στη συστοιχία:

```
Seq1    ATG--CGTCGTT
        |||  |||
Seq2    ATGCGTCGT
```

- Σε αυτήν την περίπτωση το πρόγραμμα έχει εισάγει κενά *(gaps)* στην αλληλουχία 1 που αντιστοιχεί στην κάθετη γραμμή-διακλάδωση στο διάγραμμα κουκίδων.

Επιλογή στοίχισης - κόστος

- Γενικά είναι δυνατόν να στοιχίσουμε οποιασδήποτε αλληλουχίες DNA αφού εισάγουμε κενά *(gaps)* και αντικαταστάσεις *(substitutions)* σε διαφορετικά σημεία της συστοιχίας *(alignment)*

- Μετρώντας το σύνολο των παραπάνω (κενών και αντικαταστάσεων) είναι δυνατό να υπολογίσουμε ένα μέτρο του κόστους *(score)* μιας συγκεκριμένης στοίχισης.

- Το κόστος καθορίζεται από τη σχέση:

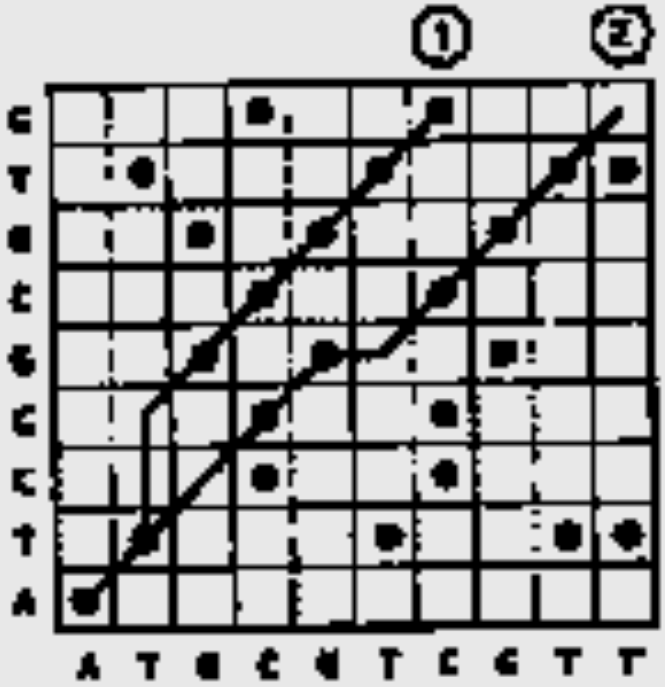
$$D = s + wg$$

S = αριθμός αντικαταστάσεων, g = το συνολικό μήκος των κενών και w = το κόστος των κενών *(gap penalty)*

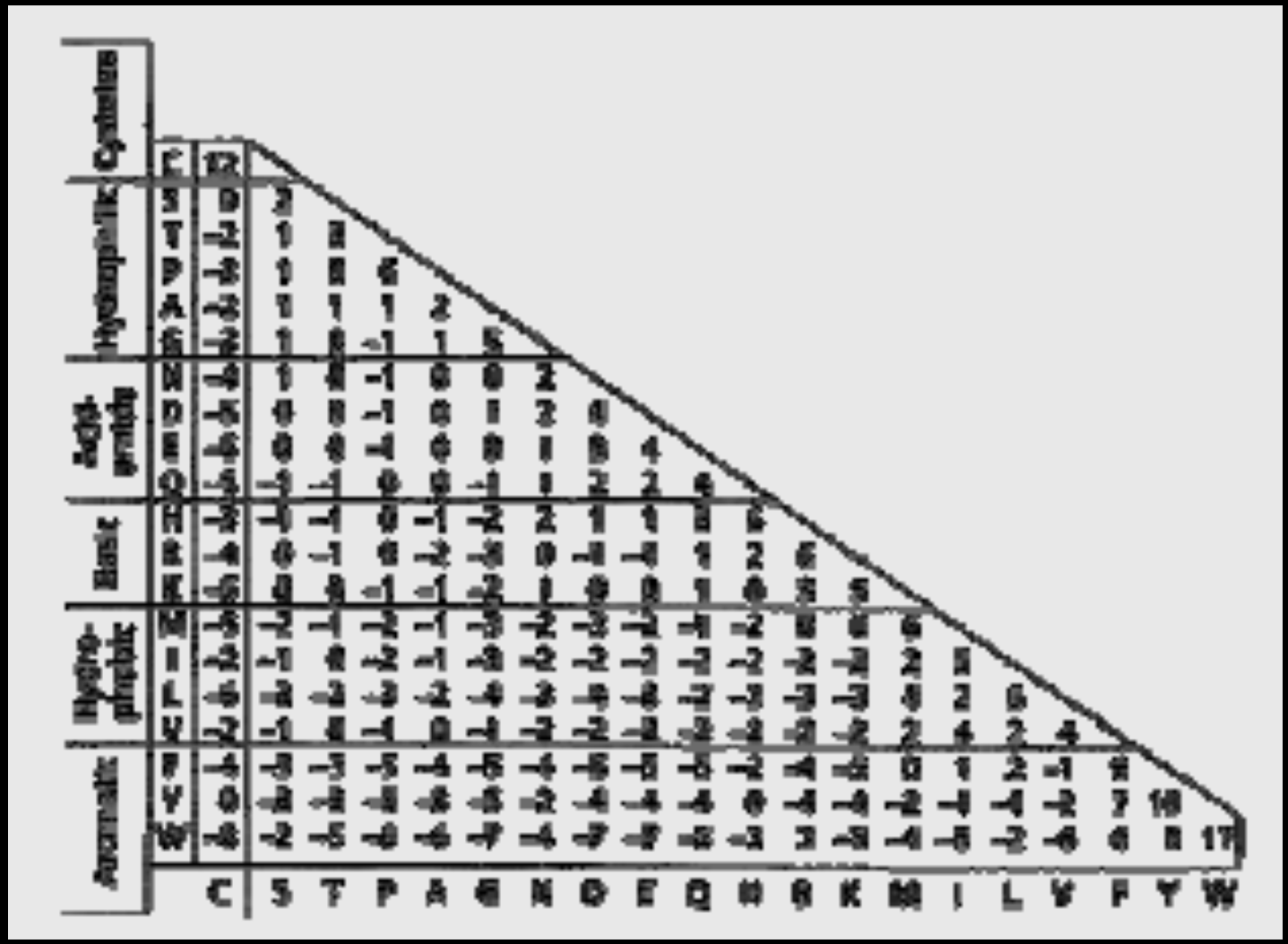
Υπολογισμός κόστους στοίχισης

```
• Seq1    AT--GCGTCGTT
•         ||  |||||
• Seq2    ATCCGCGTC

• Seq1    ATGCGTCGTT
•         || || |||
• Seq2    ATCCG-CGTC
```



Στοίχιση πρωτεϊνών



Στοίχιση πολλαπλών αλληλουχιών

- Στη φυλογενετική ανάλυση, όμως, ο ελάχιστος αριθμός αλληλουχιών που χρησιμοποιείται είναι 4 και ως εκ τούτου σπάνια στοιχίζουμε μόνο 2 αλληλουχίες.

- Οι αλγόριθμοι που χρησιμοποιούνται έχουν ως βάση κυρίως το δυναμικό προγραμματισμό *(dynamic programming)* γενετικούς αλγορίθμους *(genetic algorithms),* κρυμμένα Μαρκοβιανά Μοντέλα *(Hidden Markov Models, HMMs)* ή προοδευτικούς αλγορίθμους *(progressive algorithms).*

- Ο αλγόριθμος προοδευτικής στοίχισης *(progressive algorithms)* αποτελεί την πιο ευρέως διαδεδομένη μέθοδο στοίχισης αλληλουχιών *(align sequences).*

Παράδειγμα προοδευτικής στοίχισης

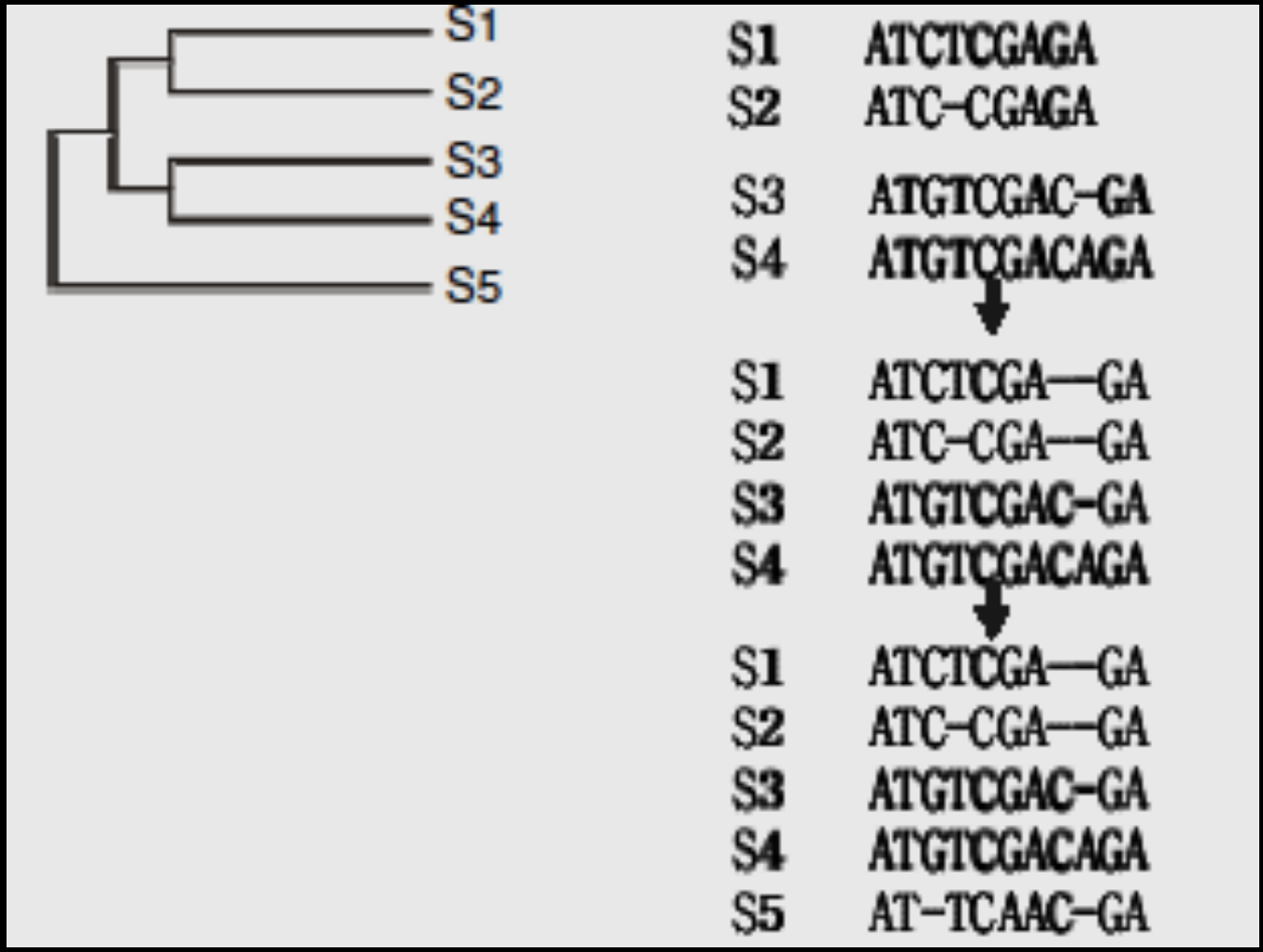
Παράδειγμα

S1	ATCTCGAGA
S2	ATCCGAGA
S3	ATGTCGACGA
S4	ATGTCGACAGA
S5	ATTCAACGA

Παράδειγμα προοδευτικής στοίχισης

	S1	S2	S3	S4	S5
S1	-				
S2	0.11	-			
S3	0.20	0.30	-		
S4	0.27	0.36	0.09	-	
S5	0.30	0.33	0.20	0.27	-

Παράδειγμα προοδευτικής στοίχισης



Αλγόριθμοι κατασκευής φυλογενετικών δένδρων
Μέθοδοι Φυλογένειας

Distance based methods

Neighbor-joining

- Minimizes distance between nearest neighbors

Parsimony methods

Maximum parsimony

- •Minimizes total evolutionary change

Probabilistic methods

Maximum likelihood – MrBayes (Bayesian inference)

- Maximizes likelihood of observed data

Comparison of Methods

Neighbor-joining	Maximum parsimony	Maximum likelihood
Uses only pairwise distances	Uses only shared derived characters	Uses all data
Minimizes distance between nearest neighbors	Minimizes total distance	Maximizes tree likelihood given specific parameter values
Very fast	Slow	***Very*** slow
Easily trapped in local optima	Assumptions fail when evolution is rapid	Highly dependent on assumed evolution model
Good for generating tentative tree, or choosing among multiple trees	Best option when tractable (<30 taxa, homoplasy rare)	Good for very small data sets and for testing trees built using other methods

Which procedure should we use?
Which one is the best?

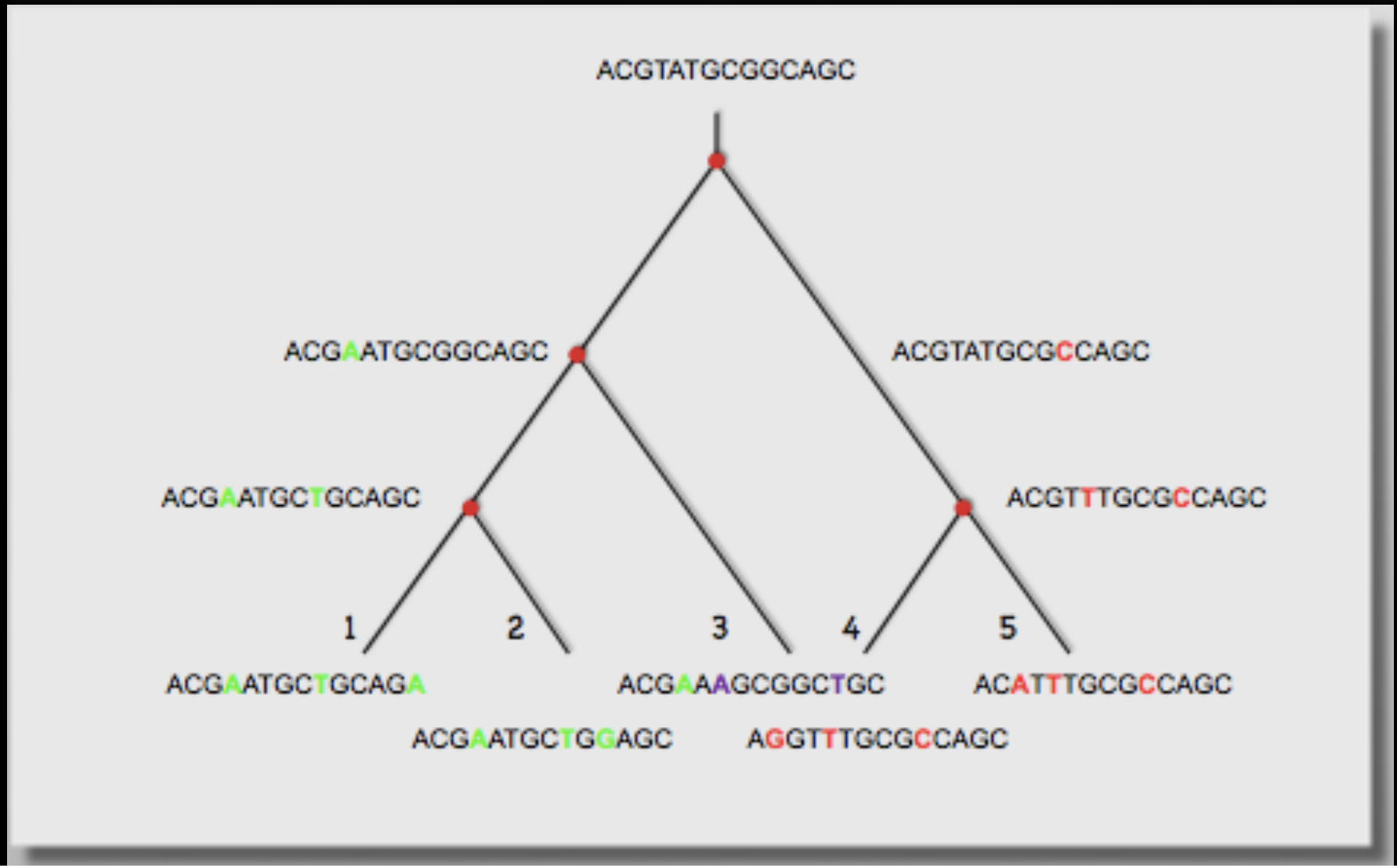
Neighbor-joining

Maximum parsimony ? Maximum likelihood

All that we can!

- Each method has its own strengths
- Use multiple methods for cross-validation
- In some cases, none of the three gives the correct phylogeny!

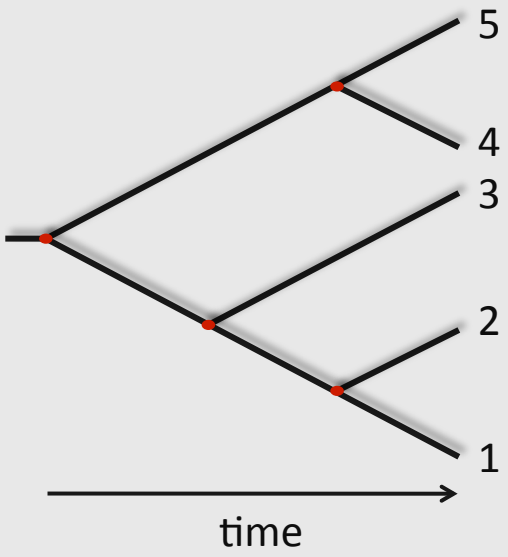
DNA data and phylogenies



Phylogeny assessment

```
1    ACGAATGCTGCAGA
2    ACGAATGCTGGAGC
3    ACGAAAGCGGCTGC
4    AGGTTTGCGCCAGC
5    ACATTTGCGCCAGC
```

	1	2	3	4	5
1	-	2	4	6	6
2		-	4	6	6
3			-	6	6
4				-	2
5					-



Tree building methods

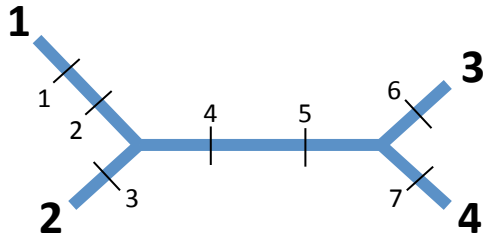
Sequences → Distances

sites

sequence	1	2	3	4	5	6	7
1	T	T	A	T	T	A	A
2	A	A	T	T	T	A	A
3	A	A	A	A	A	T	A
4	A	A	A	A	A	A	T

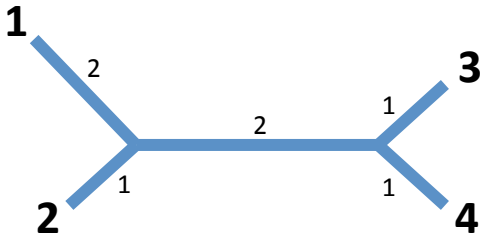
sequence	1	2	3
2	3		
3	5	4	
4	5	4	2

sequences



Discrete method
e.g. Parsimony tree

The 7 substitutions are placed on the 5 branches



Distance method
e.g. Neighbor-joining tree

The 7 substitutions are apportioned over the 5 branches

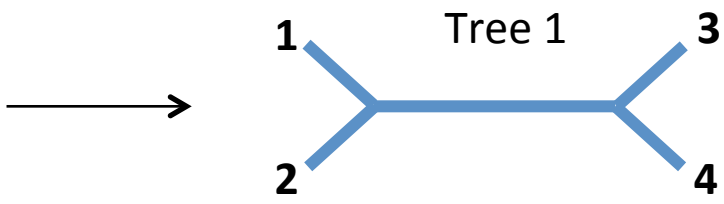
Sum of the branch length is the same in both trees: 7
But in the parsimony tree we see which site contributes to the length of each branch

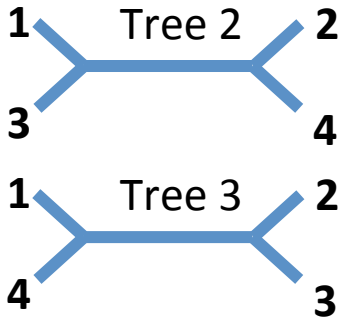
Maximum parsimony

Method:

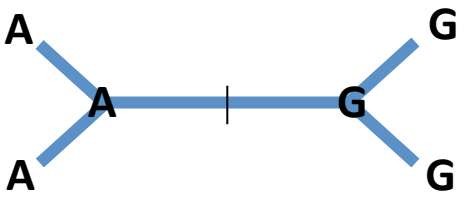
1. Make all possible trees
2. Score them by the total number of character state changes required (= "evolutionary steps") to explain distribution of each character
3. Pick the tree that infers the least steps/number of changes = the most parsimonious

1 ATATT
2 ATCGT
3 GCAGT
4 GCCGT

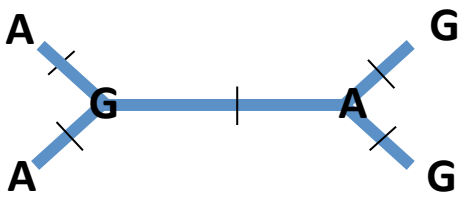




For the first site:
(two examples of what the ancestral version (internal node) could have looked like)



Only one change required
= more parsimonious



Five change required

http://tree.bio.ed.ac.uk/software/

molecular evolution, phylogenetics and epidemiology | Home | Research | People | Software | Publications

BEAST

BEAST is a cross-platform program for Bayesian MCMC analysis of molecular sequences.

Current version: v1.8.2

Description:

BEAST, Bayesian Evolutionary Analysis Sampling Trees, is a cross-platform program for Bayesian analysis of molecular sequences using MCMC. The program is orientated towards (strict and relaxed) molecular clock analyses...

Further information and downloads...

FigTree

Produce high-quality figures of phylogenetic trees

Current version: v1.4.2

Description:

FigTree is designed as a graphical viewer of phylogenetic trees and as a program for producing publication-ready figures. As with most of my programs, it was written for my own needs so may not be as polished and feature-complete as a commercial program

Software Development

Latest Software:

FigTree
Produce high-quality figures of phylogenetic trees
Latest version: *v1.4.2*

Path-O-Gen
Explore temporal signal and clocklikeness in trees
Latest version: *v1.4*

TreeStat

https://www.phylo.org/portal2/tools.action

Folders

Total Storage: 1 GB

- HIV CRF02 compare
- Greek comperative analysis
- HBV
- HCV NS1 1a
  - Data (7)
  - Tasks (11)
  - Genotyping process
    - Data (1)
    - Tasks (1)

Phylogeny / Alignment Tools

BEAST2 on XSEDE (2.3.0) - Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE

BEAST on XSEDE (1.8.0; 1.8.1; 1.8.2) - Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE

Clearcut (1.0.9) - Fast Implementation of Relaxed Neighbor Joining

ClustalW (1.82) - Create Multiple Alignments from Sequences

Consense (Phylip 3.66) - Find A Consensus Tree

DPPDIV on XSEDE (1.0) - Estimating species divergence times and lineage-specific substitution rates on a fixed topology run on XSEDE

FastTreeMP on XSEDE (2.1.8) - Fast (Approximate) Maximum Likelihood tree construction - run on XSEDE

GARLI 2.01 on XSEDE (2.01) - Genetic Algorithm for Rapid Likelihood Inference - run on XSEDE.

GARLI.conf Creator (2.0) - Creates a Garli.conf file for up to five partitions

jModelTest2 on XSEDE (2.1.6) - Statistical selection of best-fit models of nucleotide substitution, run on XSEDE

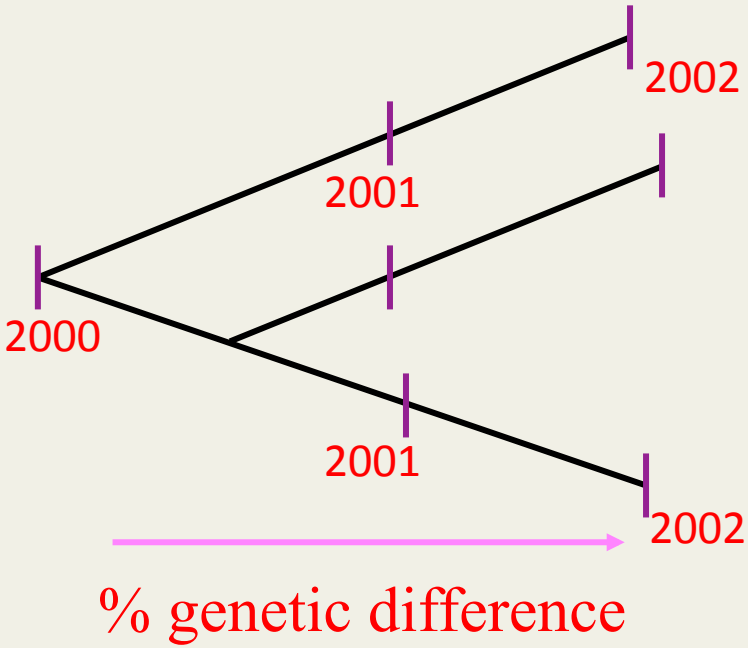
MAFFT on XSEDE (7.187) - Multiple alignment program for amino acid or nucleotide sequences; parallel version

Migrate-N on XSEDE (3.6.11) - Estimation of Population Sizes and Gene Flow using the Coalescent

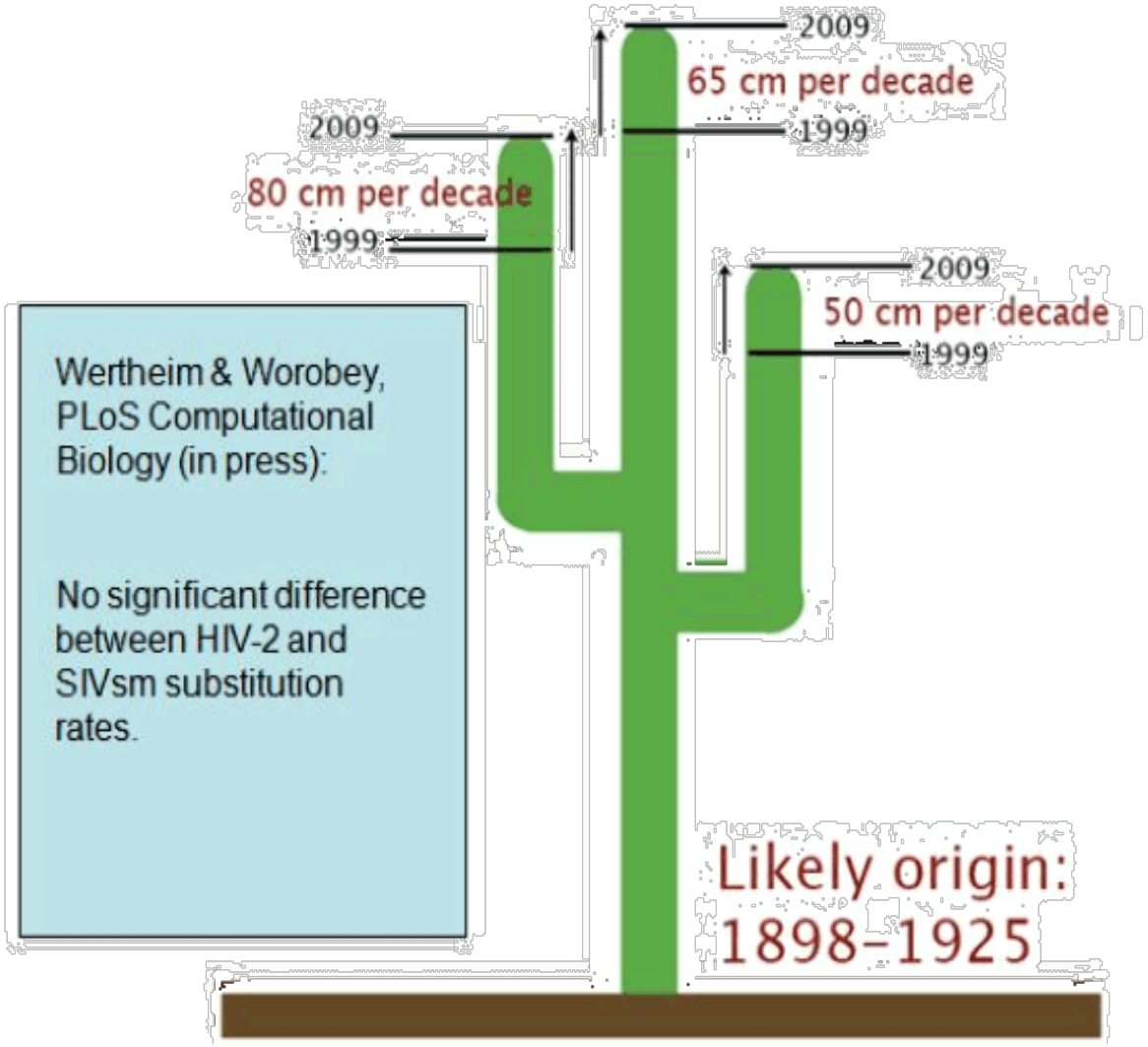
Molecular Clock

- Mutation/substitution occurs as a random (Poisson) process. If mutations accumulate at a constant rate over time and across all branches, the phylogeny is said to obey a

  molecular clock.

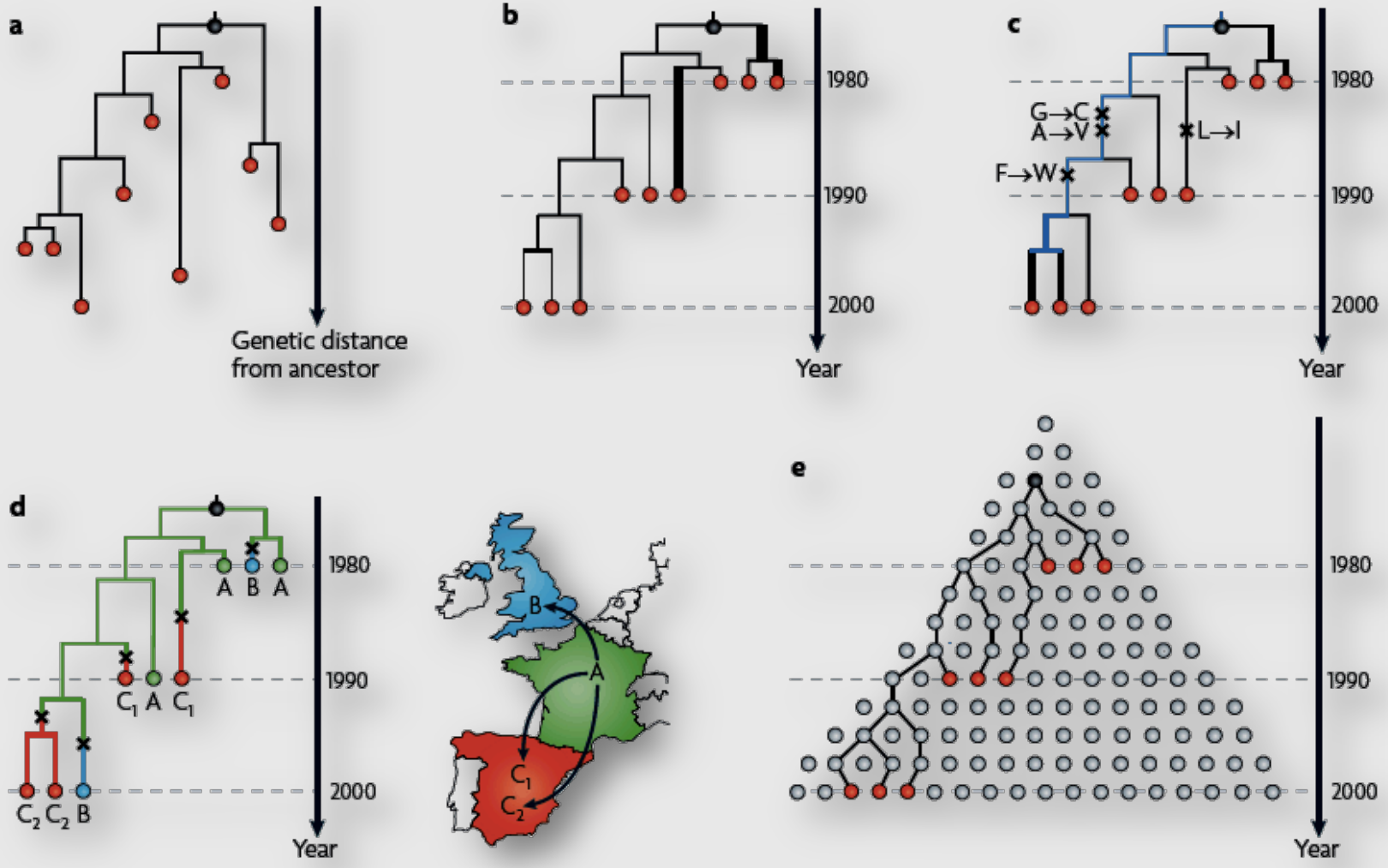


Molecular Clock

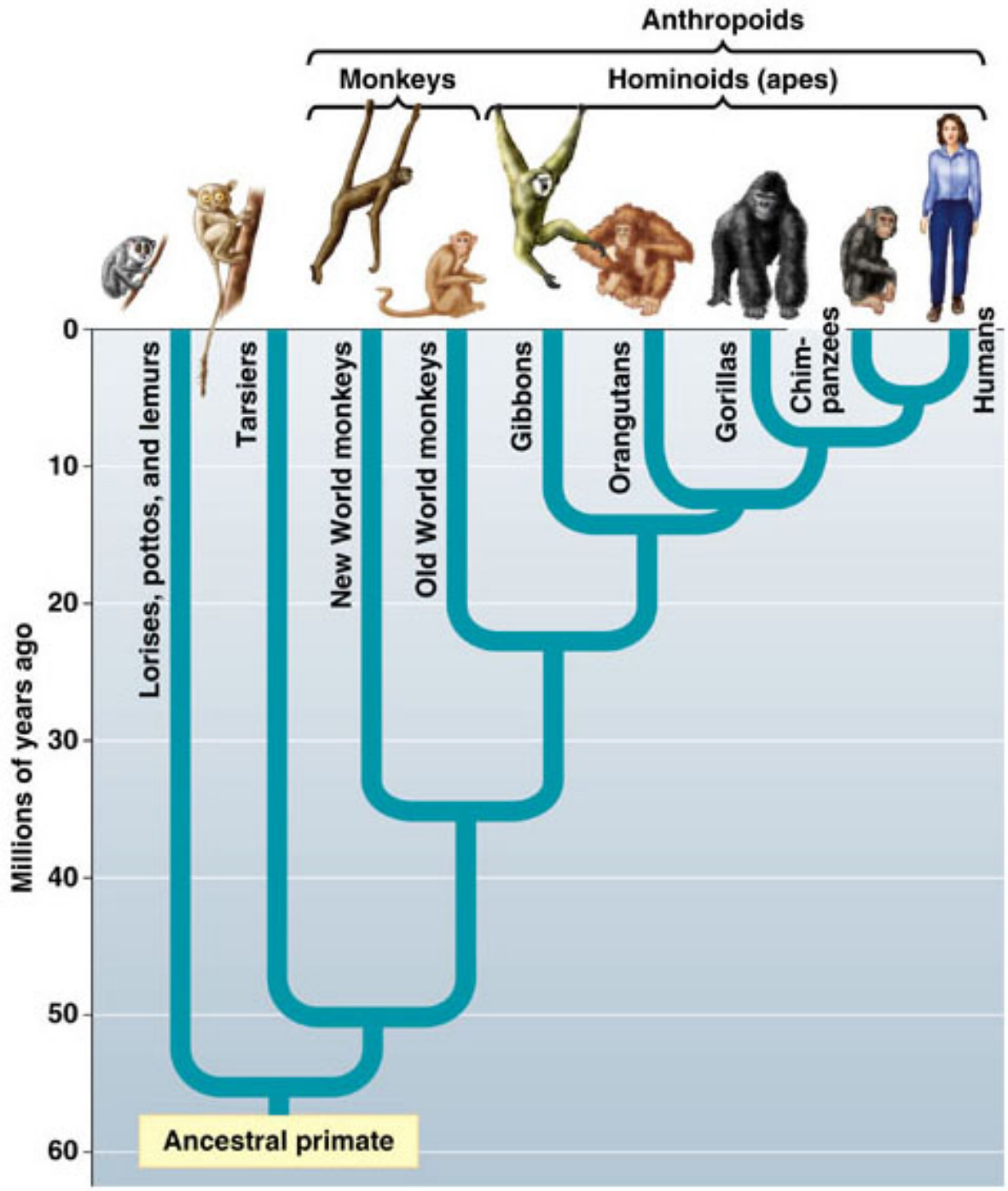


Worobey CROI 2009

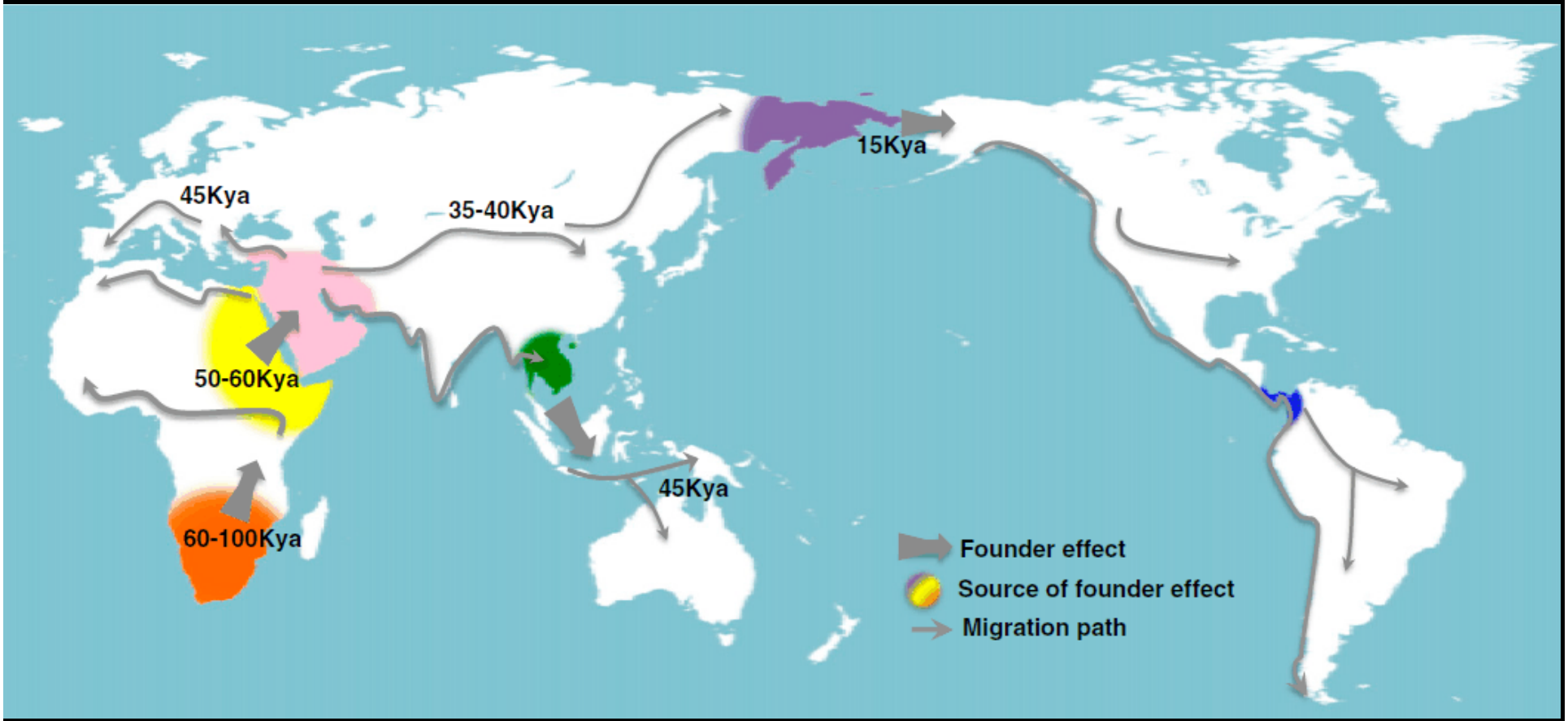
Phylodynamics



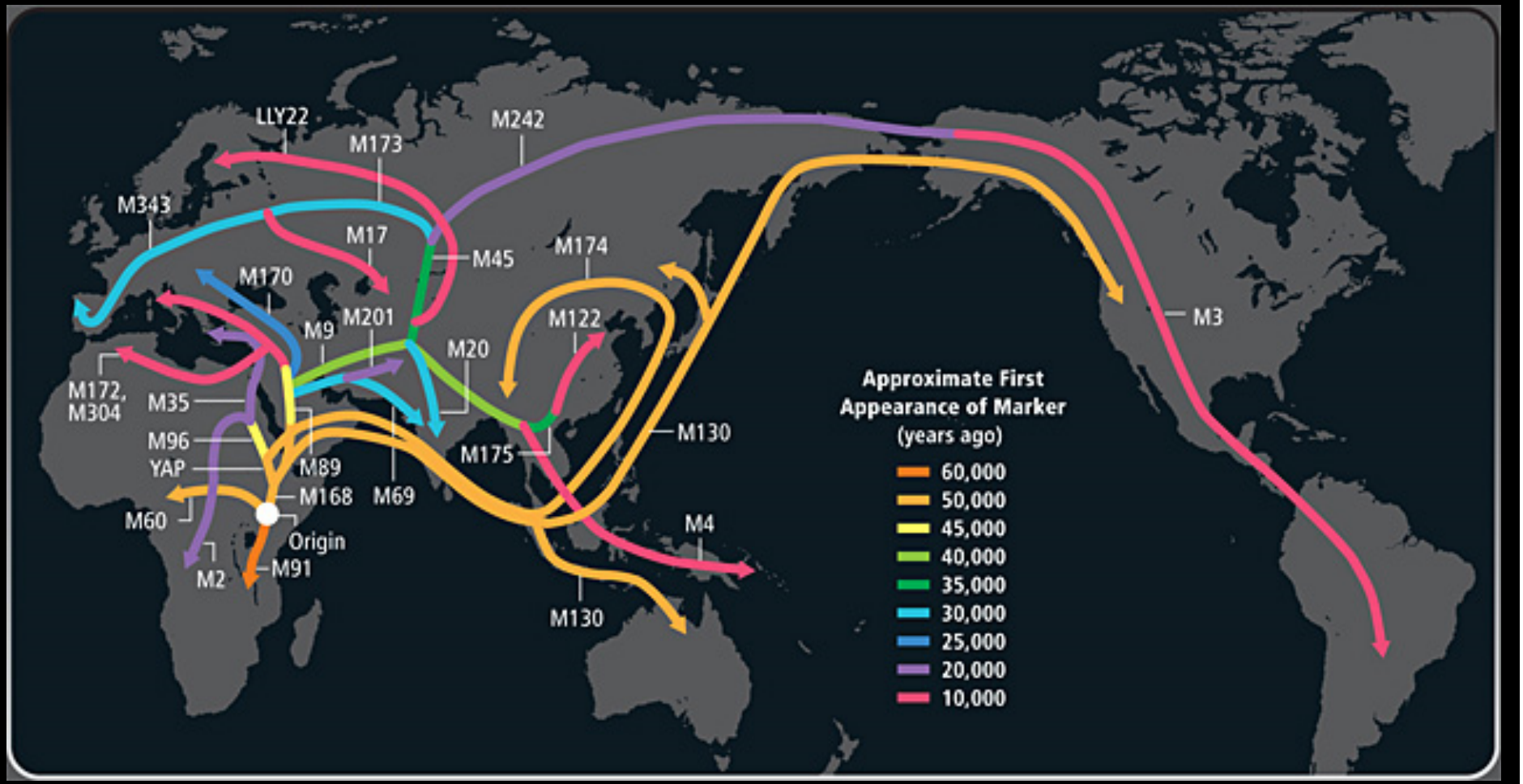
Origin of homo sapiens



Migration of modern humans



Migration of modern humans



Key concepts in this section

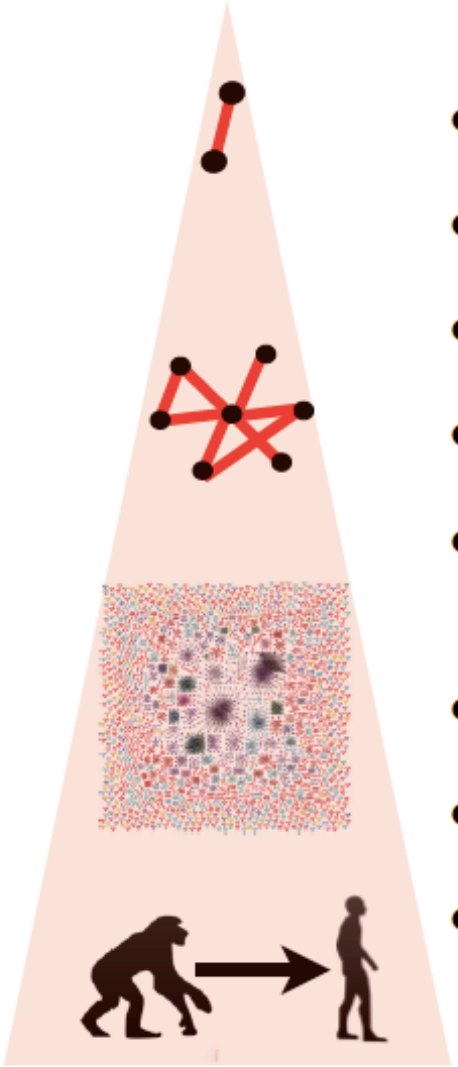
- What is molecular epidemiology and differences Vs epidemiology
- What are phylogenies or phylogenetic trees?
  - Terminology such as extant, ancestral, branch point, branch length
- Why build phylogenetic trees?
  - Algorithms to build phylogenetic trees
    - Distance-based methods
    - Parsimony methods
      - Minimize the number of changes
    - Probabilistic methods
      - Tree that best explains the data using probabilistic models
- Viral infections – Viral epidemics
  - origin, expansion, geographic distribution, evolution

Molecular epidemiology of viral pathogens

- Advances in sequencing technology - Decrease of costs
- DNA sequences data
- Advances in methodologies for genetic analyses and Computation (HPC)

- Genetic heterogeneity
- R_0
- Most rapidly evolving species/pathogens
- Availability of HIV sequences and epidemiological data

Insights gained by molecular epidemiology

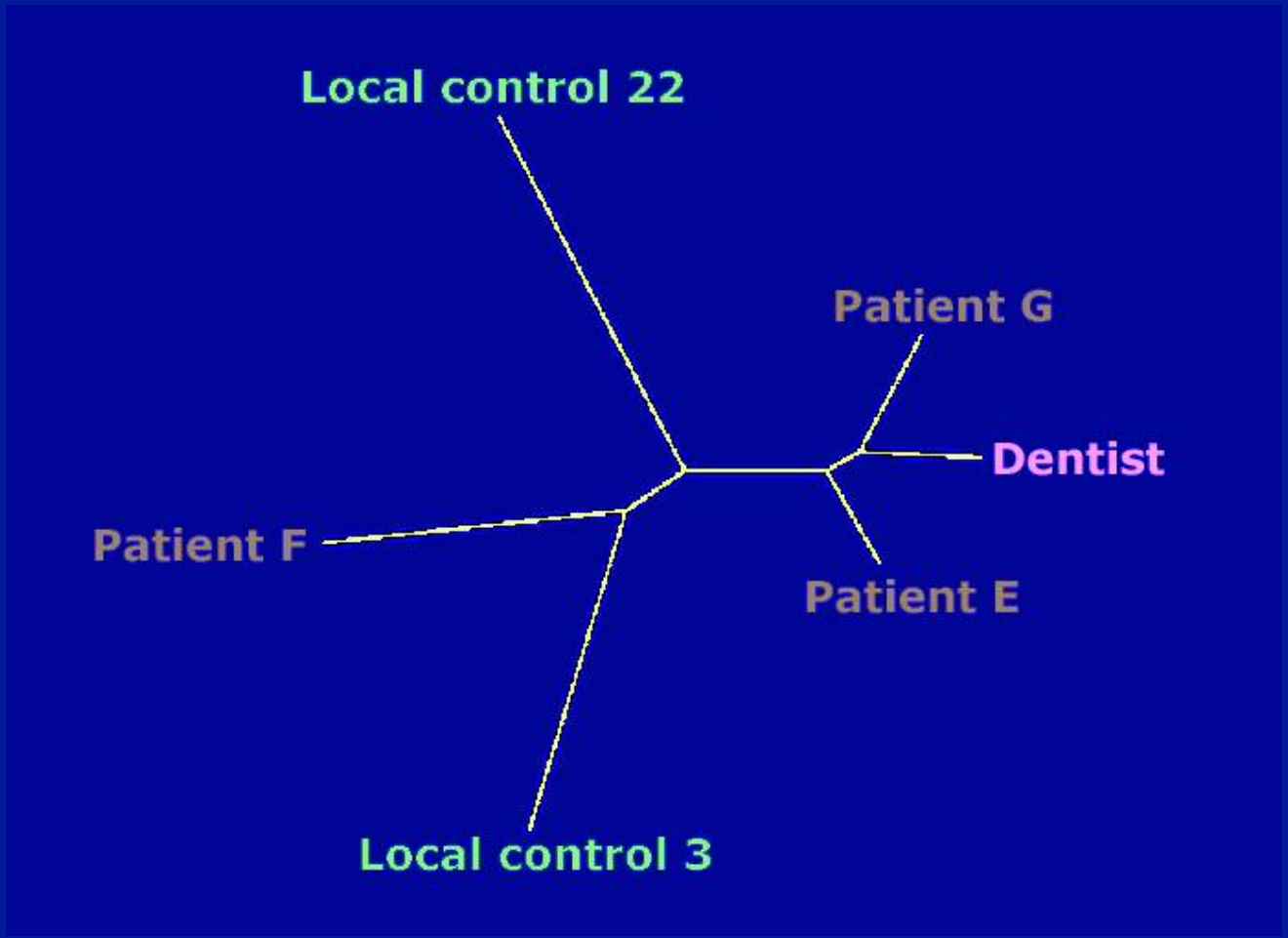


- Forensics
- Shape of contact network
- Pathogenesis/Heritability
- Spread of Resistant HIV
- Interaction between risk groups
- Assessing Interventions
- Geographical Structure
- Zoonotic transmission & early spread of HIV

HIV-1 forensic: example 1

- ✓ 1990 case: Did a patient's HIV infection result from an invasive dental procedure performed by an HIV(+) dentist?
- ✓ HIV evolves so fast that transmission patterns can be reconstructed from viral sequence (molecular forensics).
- ✓ Compared viral sequence from the dentist, three of his HIV+ patients, and two HIV+ local controls.

HIV-1 forensic: Florida Dentist case



Who was the painter?

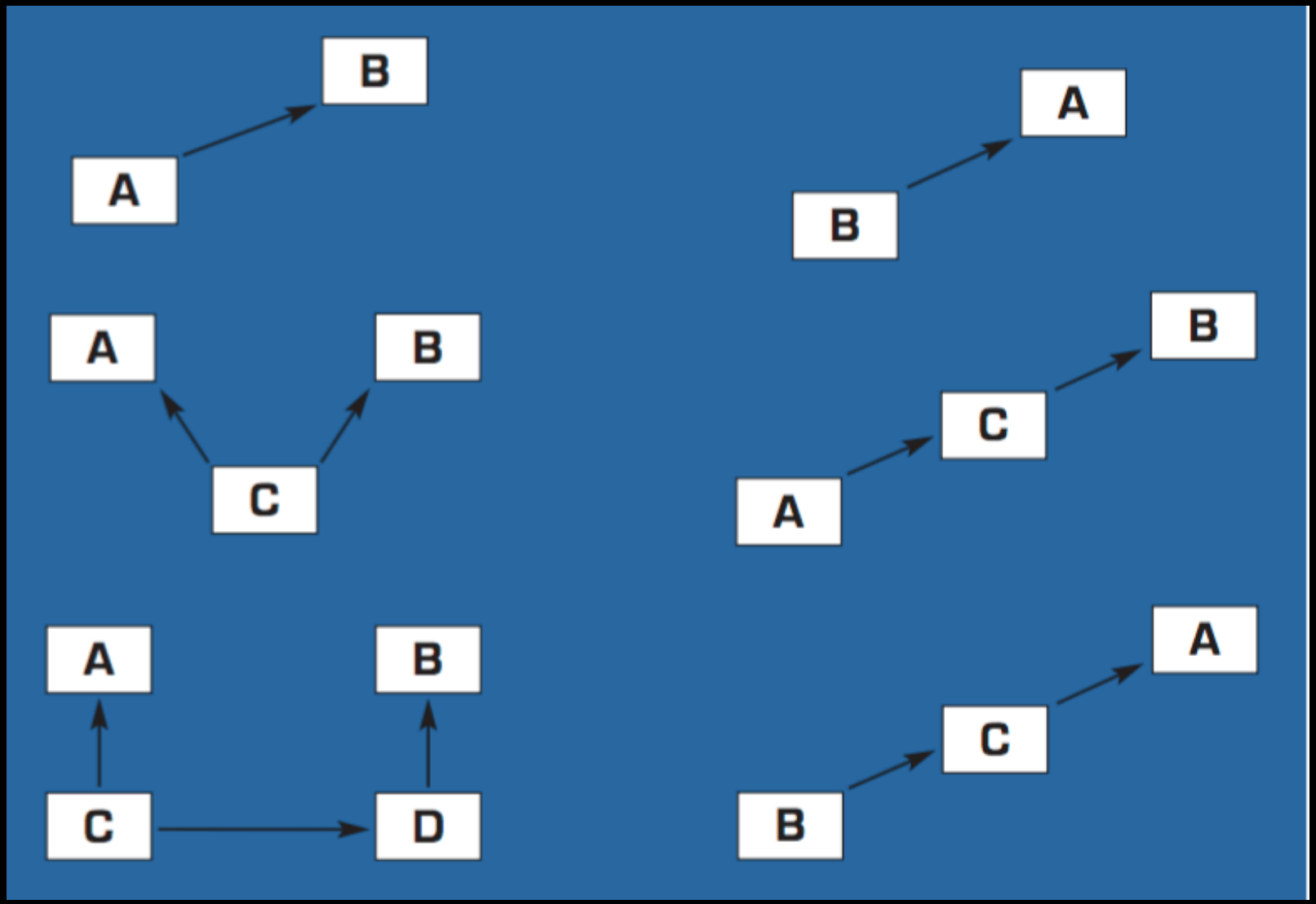
A. Michelangelo
B. Leonardo da Vinci
C. Marcel Duchamp
D. Frida Kahlo

Who was the painter?

A. Michelangelo
B. Leonardo da Vinci
C. Marcel Duchamp
D. Frida Kahlo

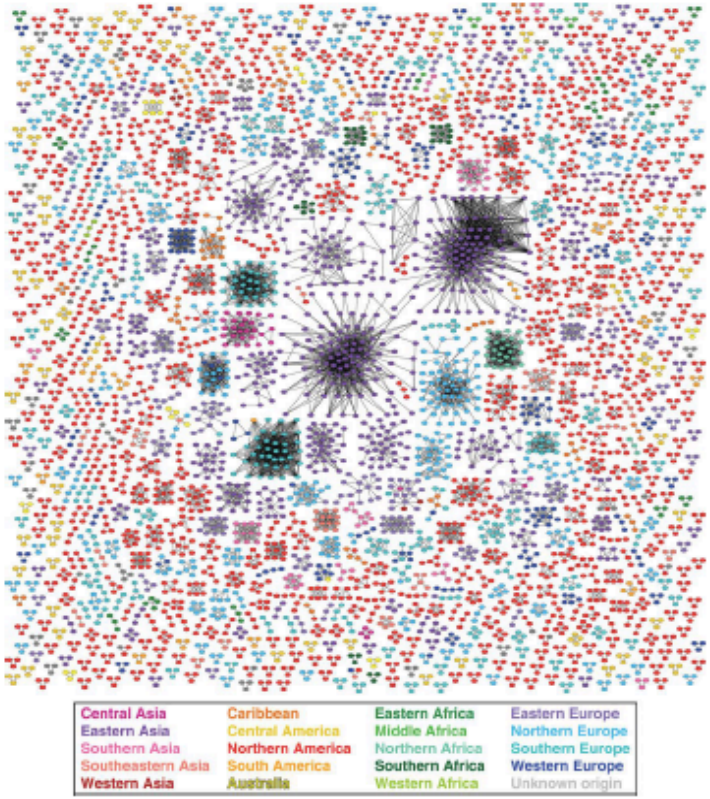
1919, L.H.O.O.Q

HIV-1 forensic: Florida Dentist case



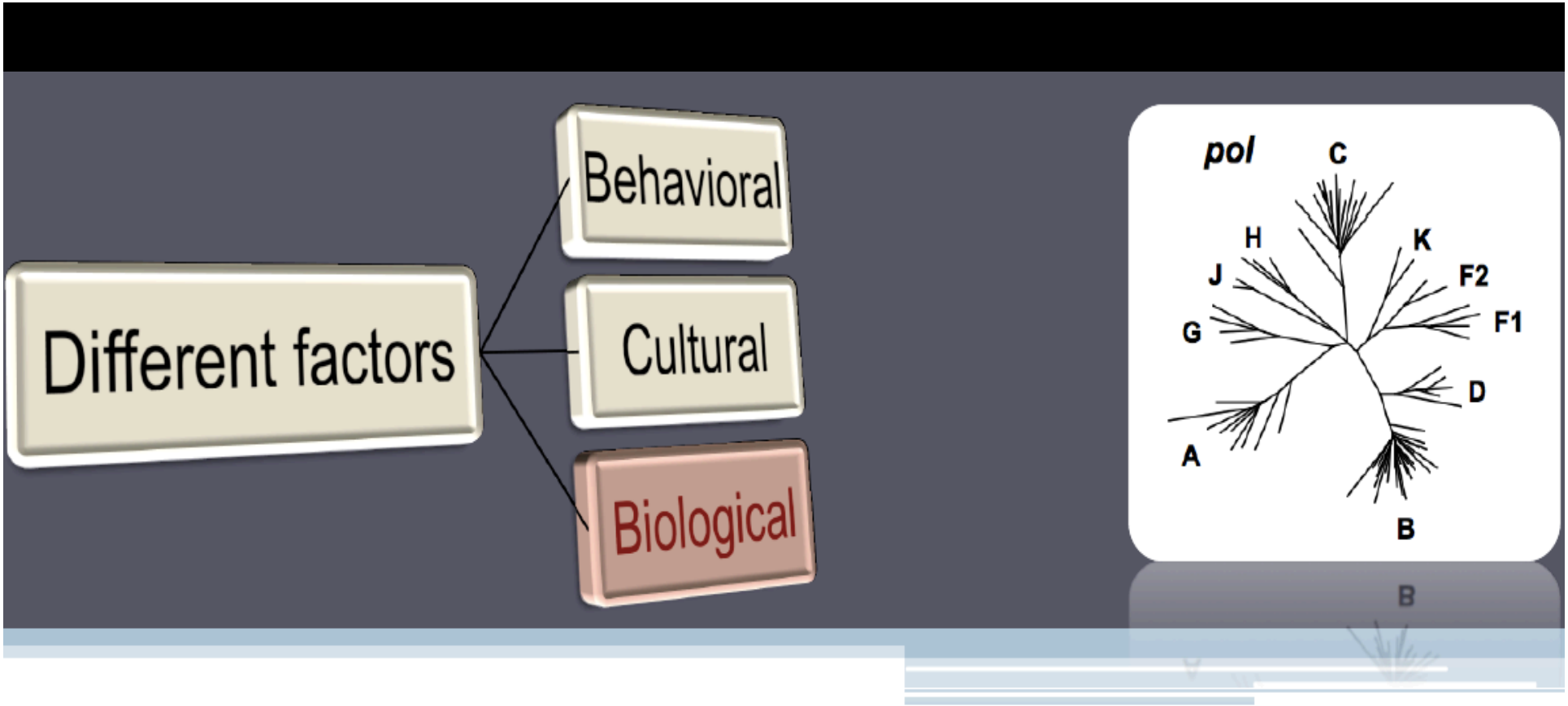
Bernard EJ, Azad Y, Vandamme AM, Weait M, Geretti AM. HIV Med. 2007 Sep;8(6).

HIV transmission networks



Wertheim J O et al. J Infect Dis. 2014;209:304-313

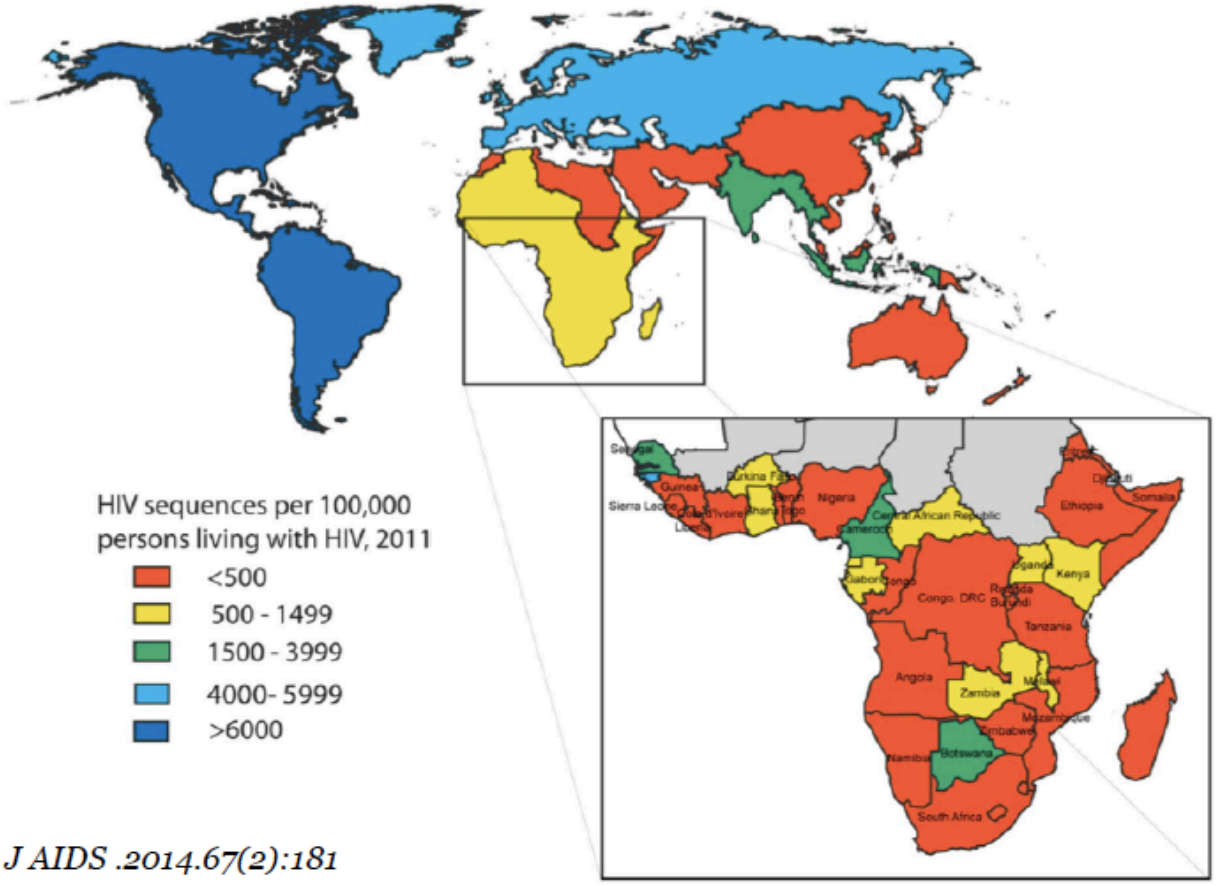
- HIV Transmission not random but shaped by
  - Geography
  - Uneven Distribution of Risk Behavior (super-spreader)
  - Transmission Modes (IDU/MSM/HETs)
  - Stage of Infection
- Importance for Public Health
  - Determining the good Geographical scale for Interventions/Trials
  - Targeting Prevention
  - Predicting the Success of Prevention Measures (Test & Treat)



The characterization of local HIV epidemics is needed to evaluate the impact, cost-effectiveness and sustainability of prevention strategies

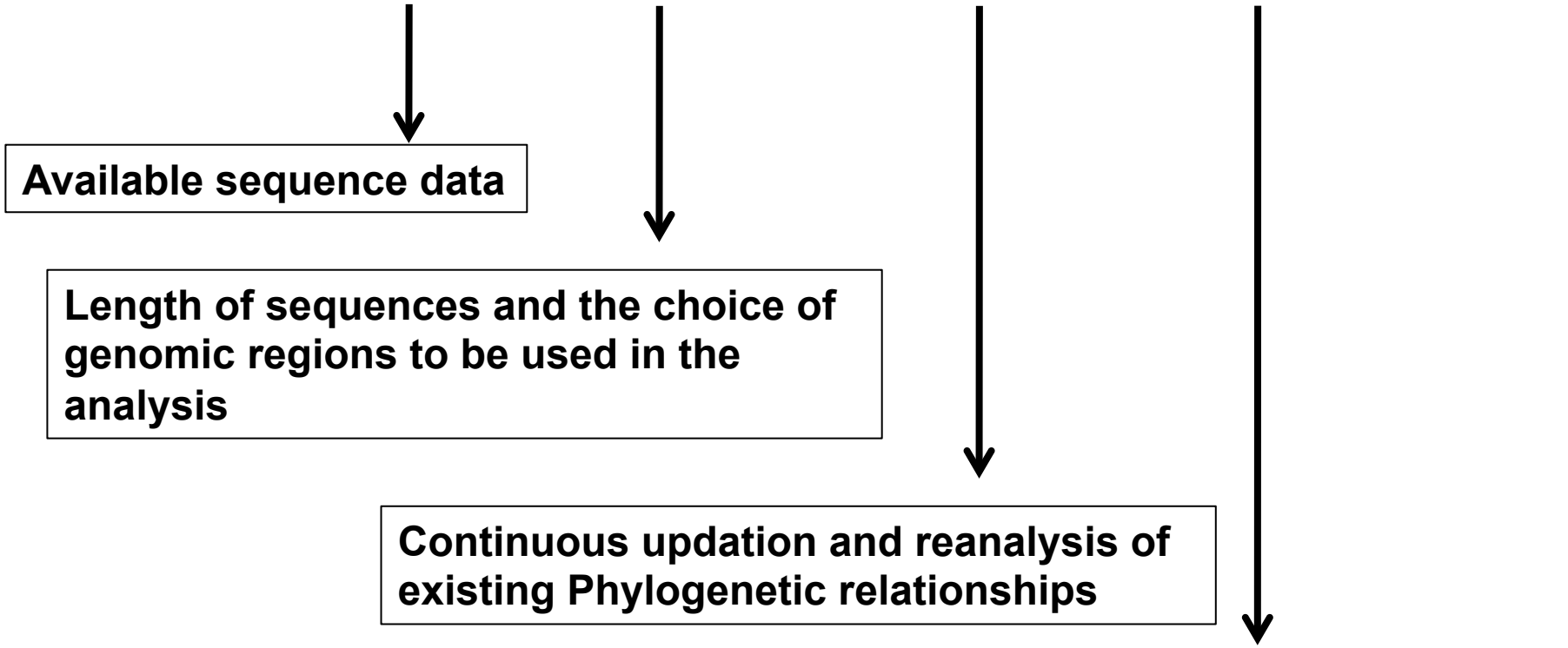
DeGruttola V et al. CID. 2010

Fewer sequences are available from some regions of Asia and Africa

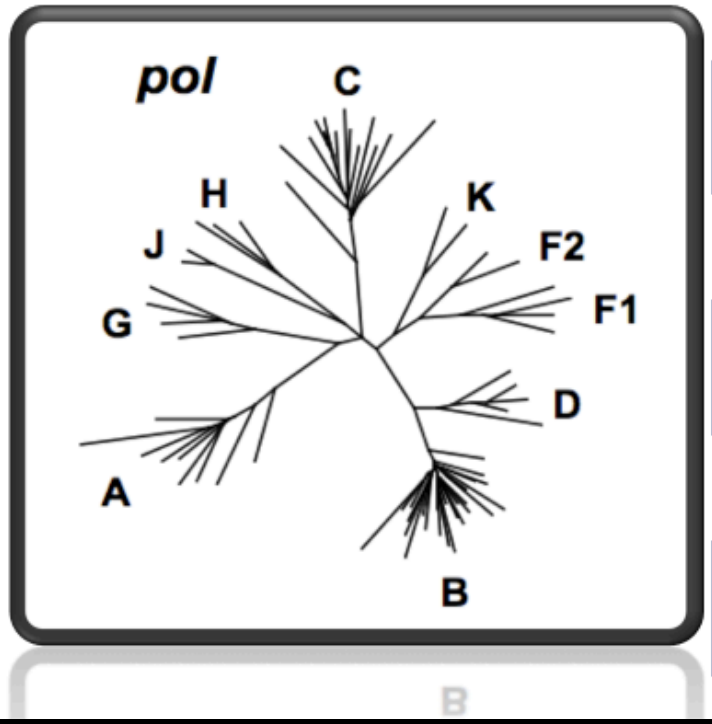


Dennis et al. J AIDS .2014.67(2):181

Limitations of Molecular Phylogenetics



Phylogenetics in **Epidemiology**



Phylogeography

Phylodynamics

Molecular epidemiology

UNIVERSITY OF LIVERPOOL | INSTITUTE OF INFECTION AND GLOBAL HEALTH

PARTNER

Research

Original Investigation

Sexual Activity Without Condoms and Risk of HIV Transmission in Serodifferent Couples When the HIV-Positive Partner Is Using Suppressive Antiretroviral Therapy

Alison J. Rodger, MD; Valentina Cambiano, PhD; Tina Bruun, RN; Pietro Vernazza, MD; Simon Collins; Jan van Lunzen, PhD; Giulio Maria Corbelli; Vicente Estrada, MD; Anna Maria Geretti, PhD; Apostolos Beloukas, PhD; David Asboe, FRCP; Pompeyo Viciana, MD; Félix Gutiérrez, MD; Bonaventura Clotet, PhD; Christian Pradier, MD; Jan Gerstoft, MD; Rainer Weber, MD; Katarina Westling, MD; Gilles Wandeler, MD; Jan M. Prins, PhD; Armin Rieger, MD; Marcel Stoeckle, MD; Tim Kümmerle, PhD; Teresa Bini, MD; Adriana Ammassari, MD; Richard Gilson, MD; Ivanka Krznaric, PhD; Matti Ristola, PhD; Robert Zangerle, MD; Pia Handberg, RN; Antonio Antela, PhD; Sris Allan, FRCP; Andrew N. Phillips, PhD; Jens Lundgren, MD; for the PARTNER Study Group

Sequencing and Phylogenetic analyses of the PARTNER samples

Background

- A key factor in assessing the effectiveness of ART as a prevention strategy is the **absolute risk of HIV transmission through condomless sex (CL)** for a person on ART with undetectable plasma VL (suppressive ART)
- There are however a number of gaps in currently available evidence
- There is no direct evidence at all for anal sex in men who have sex with men
- In transmission studies in HT couples most CYFU in context of reported consistent condom use

Rodger et al. Antivir Ther. 2013;18(3):285-7

PARTNER Study

The PARTNER study was an observational multi-centre study of HIV serodifferent couples (MSM and HT) in which the positive partner is on ART in 75 European clinical sites

Primary Aim

To follow serodifferent partnerships that have penetrative sex without using condoms where the HIV-positive partner is on ART with a plasma HIV-1 RNA load <200 copies/ mL in order to study risk of HIV transmission

Study Procedures

- Eligible CYFU (couple-years of follow-up):
  - had condomless sex during the time period
  - there was no reported PEP or PrEP use
  - plasma HIV-1 RNA load <200 copies/mL within 12 months
  - follow-up occurred before 31st May 2014 (i.e. censoring date)

- Overall by 31st May 2014, 888 (548 HT and 340 MSM) couples contributed 1238 eligible CYFU

HIV transmissions

- A total of 11 of the originally HIV negative partners acquired HIV during eligible follow up.
- Of the 11 people who became infected, 10 were MSM, 1 was heterosexual, 8 (73%) of these reported that they had recent condomless sex with others apart from their study partner.

- Viral sequences were recovered successfully from all couples, comprising **22/22 (100%) subjects** for *pol* and **20/22 (91%) subjects** for *env*.

- Samples collected from the two partners of each couple were median 0 months' apart (IQR 0.0-5.9).

"Transmission" PARTNER Couples

- 11 putative transmission events
- 27 specimens examined
  - N= Initially negative partner
  - P= Initially positive partner
- Sanger (population) sequencing of plasma HIV-1 RNA or cell-associated HIV-1 DNA using high-fidelity polymerase
- **HIV-1 *pol*** (aa1-414; HXB2 2253-3495nt) and ***env*** (aa43-708; HXB2 6351-8348nt) sequences

Couple #	Partner	Sample Type
1, 7, 10	N	Plasma
	P	PBMC
	P	Plasma
2-6, 8	N	Plasma
	P	PBMC
9	N	PBMC
	N	PBMC
	N	PBMC
	P	PBMC
11	N	Plasma
	P	Plasma

Geretti AM et al, J Antimicrob Chemother 2014,69:1090-1097.

Phylogenetic analysis

- Substitution model: JModelTest v2
  - General Time Reversible (GTR) nucleotide substitution model
  - 012212+I+G+F sub-model
- **Maximum likelihood (ML)** and **Bayesian Markov Chain Monte-Carlo (MCMC)** inferences were determined with RAxML-HCP2 v8 and Mr Bayes v3.2.6, respectively
- Controls: i) the 10 closest GenBank sequences, ii) replicate partners' sequences (obtained from different sampling points, different specimen types, or repeat testing of the same sample), and iii) sequences from confirmed HIV- transmission pairs
- Criteria for linking infections was monophyletic clustering with high support (e.g bootstrap value ≥0.90 (ML) or a posterior probability ≥0.95 (MCMC)), and a pairwise genetic distance of ≤0.015 nucleotide substitutions per *pol* site

Beloukas A et al, Virus Research 2012

#	Partner	Sample Type	Pol
1	N	Plasma	B
	P	PBMC	B
	P	Plasma	B
7	N	Plasma	B
	P	PBMC	B
	P	Plasma	B
10	N	Plasma	B
	P	PBMC	B
	P	Plasma	-

#	Partner	Sample Type	Pol
2-4, 8	N	Plasma	B
	P	PBMC	B
5	N	Plasma	A1
	P	PBMC	B
6	N	Plasma	CRF14
	P	PBMC	B
9	N	PBMC	B
	N	PBMC	B
	N	PBMC	B
	P	PBMC	B
11	N	Plasma	B
	P	Plasma	B

Summary:

- *Pol* sequences from 26/27 specimens and 11/11 couples
- *Env* sequences from 22/27 specimens and 9/11 couples (82%)

Phylogenetic tree of *pol* sequences from nine couples with subtype B infection

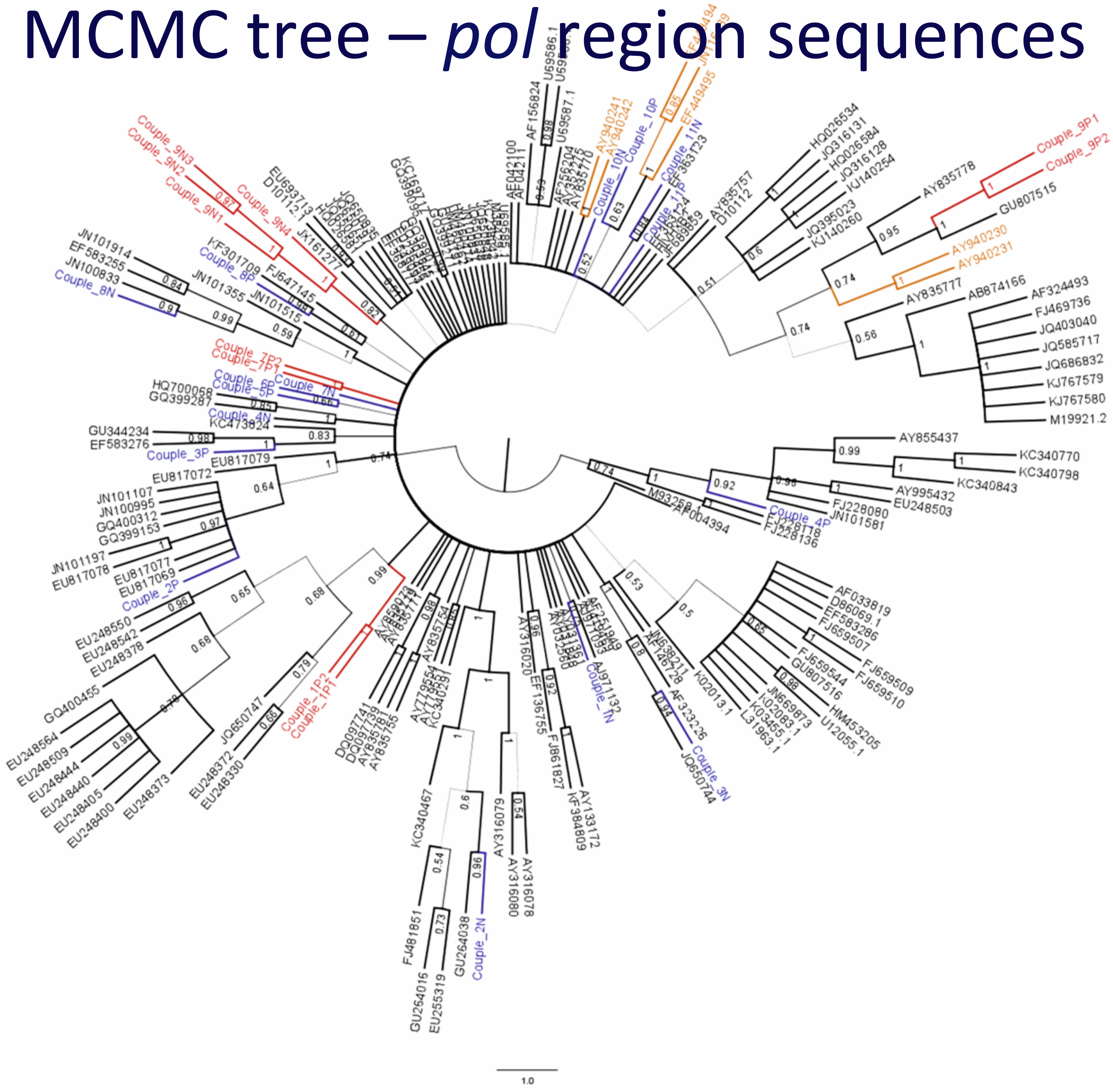
- The Positive partner was always infected with HIV-1 subtype B and in 2/11 cases the Negative partner found to have been infected by a different subtype/CRF
- None of the partners' sequences clustered together, with consistent results observed across analyses.
- The partners' sequences showed pairwise genetic distances consistently >0.040.
- The control sequences always clustered together (ML bootstrap >98% and MCMC pp = 1.00 and their median genetic distance was found 0.004.

Genetic Distances

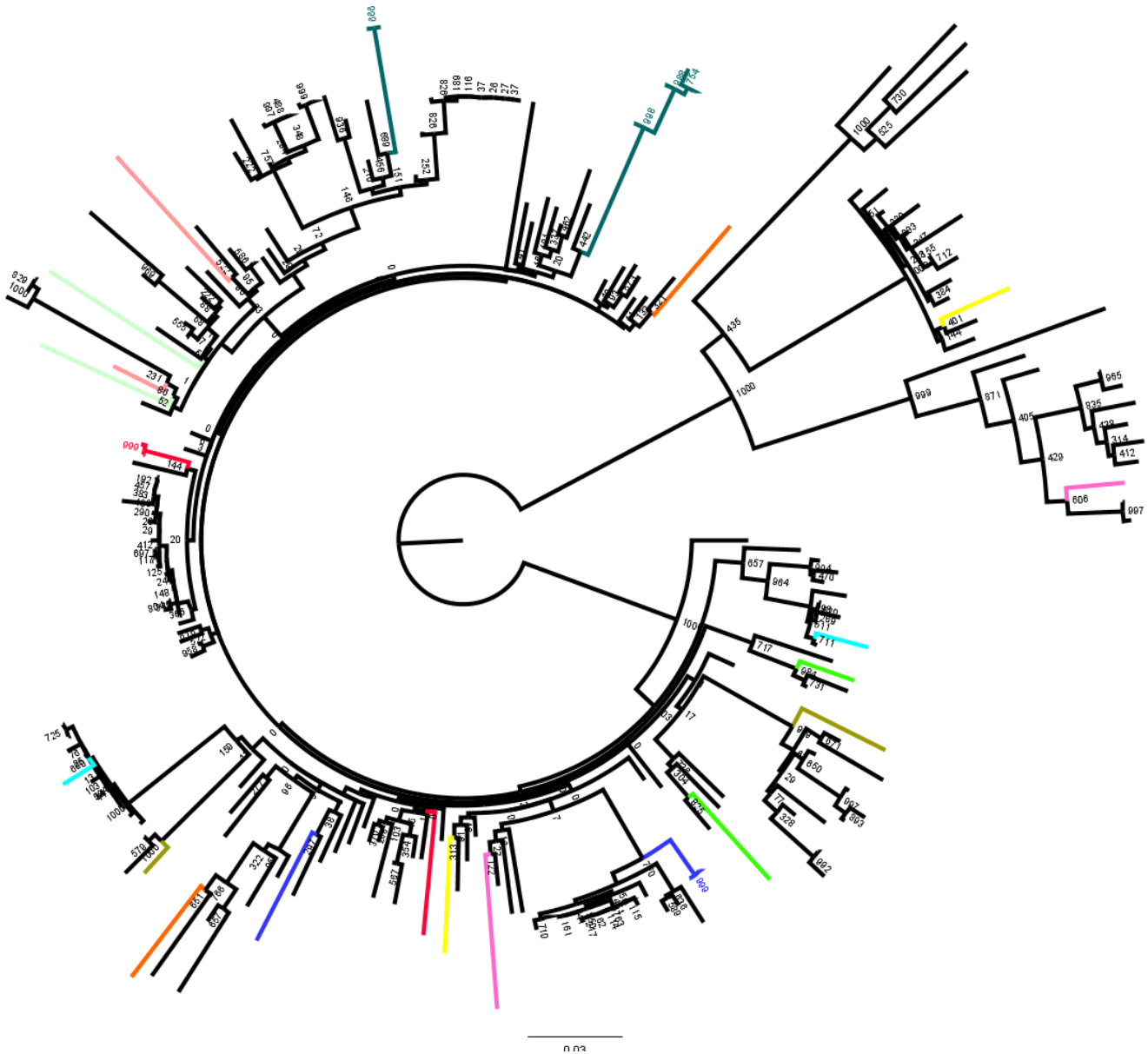
- Screening criterion for potential clustering set as a genetic distance **≤4.0% (0.040)** for the pol site and a branch support of ≥90% (bootstrap value) or ≥0.95 (posterior probability)

Sequences	Median (IQR)
Partner *pol*	0.067 (0.042, 0.078)
Partner total	0.08 (0.068, 0.121)
Partner control *pol*	0.001 (0.000, 0.003)
Partner control *total*	0.003 (0.000, 0.003)
Greek control pol	0.008 (0.006, 0.009)

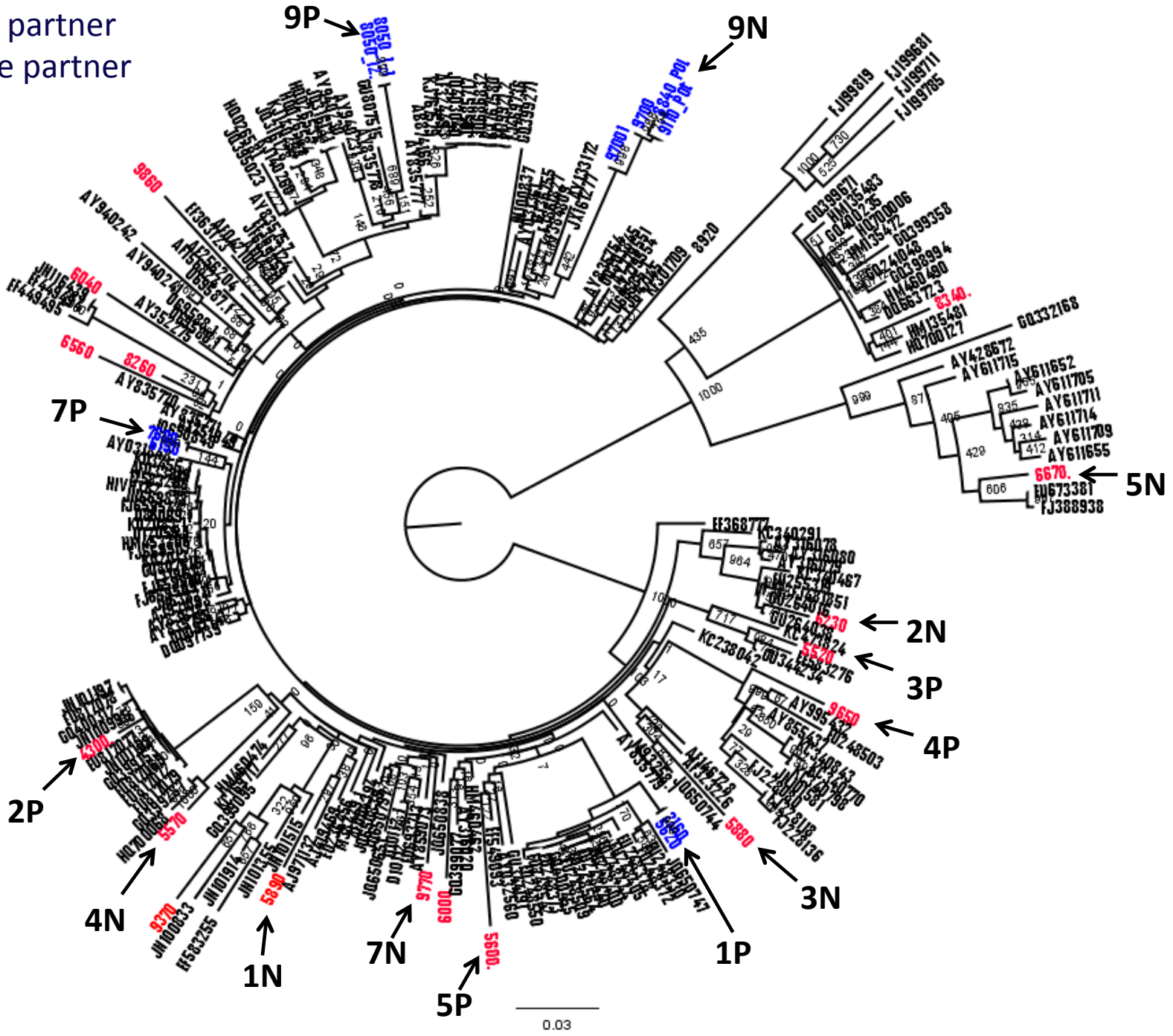
MCMC tree – *pol* region sequences



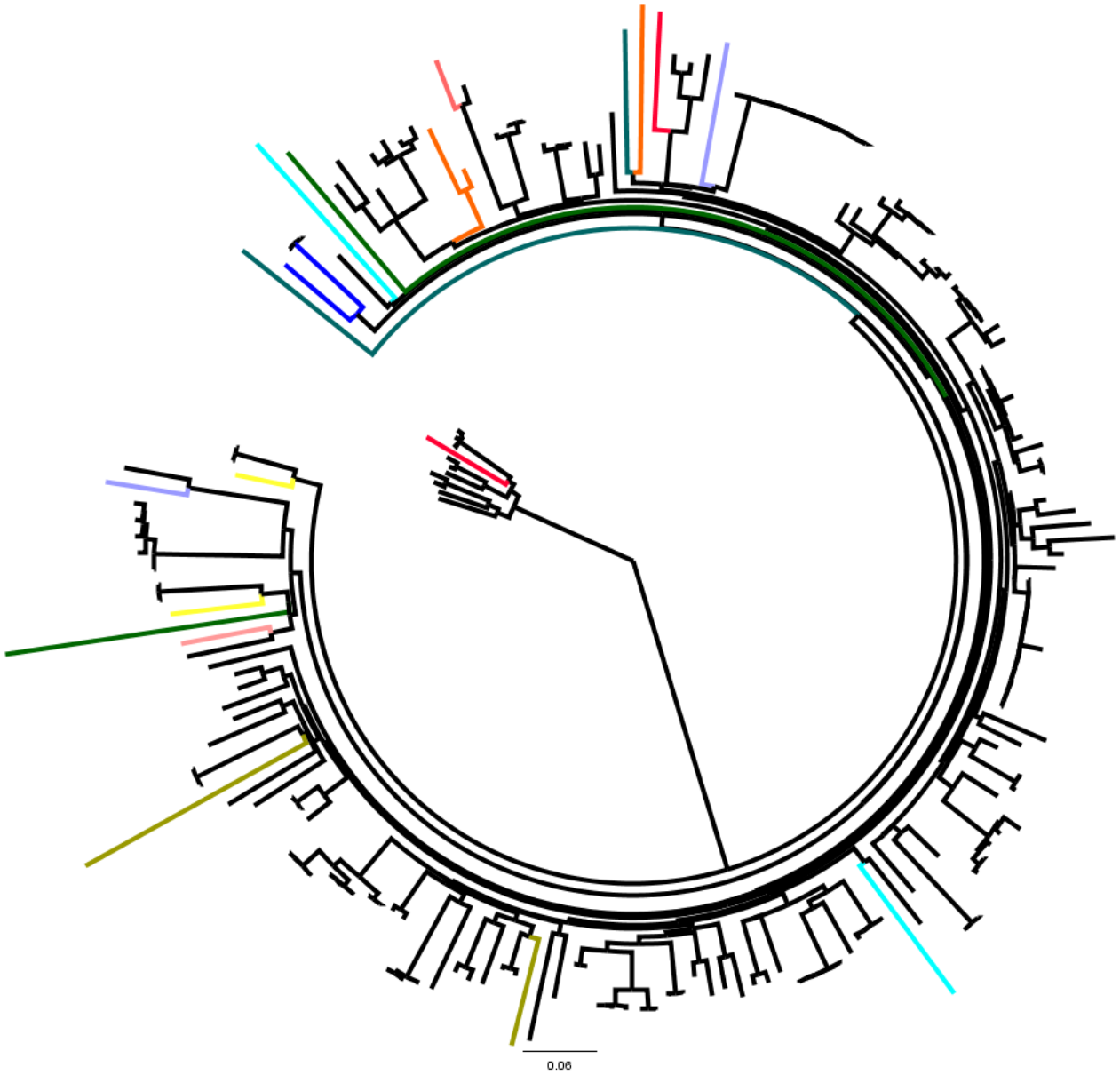
ML tree – *pol* region sequences



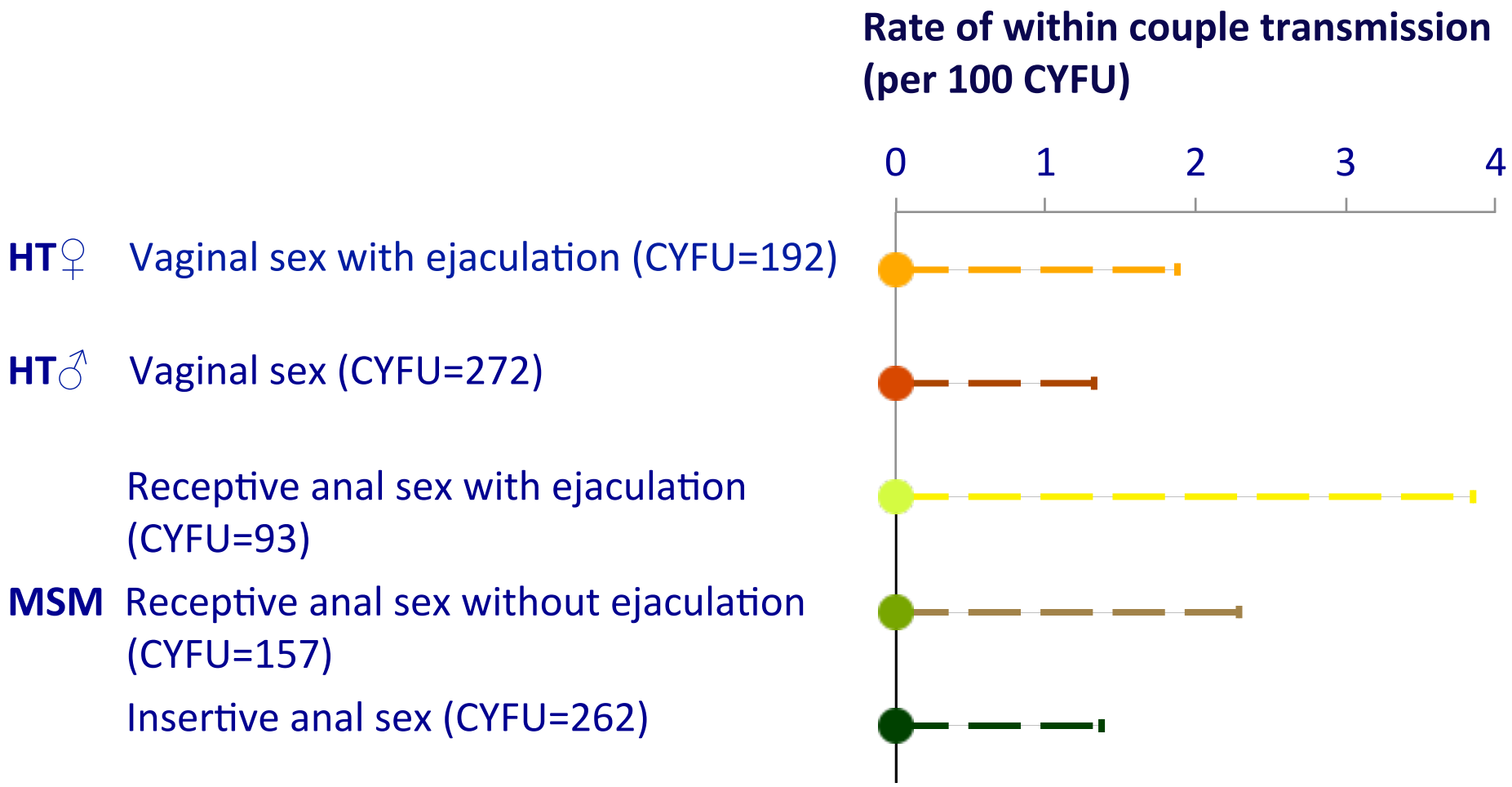
P: positive partner
N: negative partner



ML tree – *env* region sequences



Rate of HIV transmission according to sexual behaviour reported by the negative partner



Conclusions

- Phylogenetic analysis of the Sanger *pol* and *env* sequences of 11 putative transmission couples do not indicate that the couples have linked viruses
- Among serodifferent HT and MSM couples in which the HIV-positive partner was using suppressive ART and despite a significant number of condomless sex acts over 1238 CYFU, there were **zero documented cases of within-couple HIV transmission** (upper 95% CI 0.30/100 CYFU)
- This provides the first estimate of HIV transmission risk for MSM **through condomless anal sex** with suppressed plasma HIV VL.
- Additional longer-term follow-up in MSM is necessary to provide more precise estimates of risk to inform policy and also individual choice on condom use

Παράδειγμα 2

Διερεύνηση της ετερογένειας μιας επιδημίας και συσχετισμός με κλινικά και επιδημιολογικά δεδομένα

HIV-1 Local Transmission Networks (LTNs) in 9 European Countries and Canada:
Association with demographic and clinical factors

Sampling countries

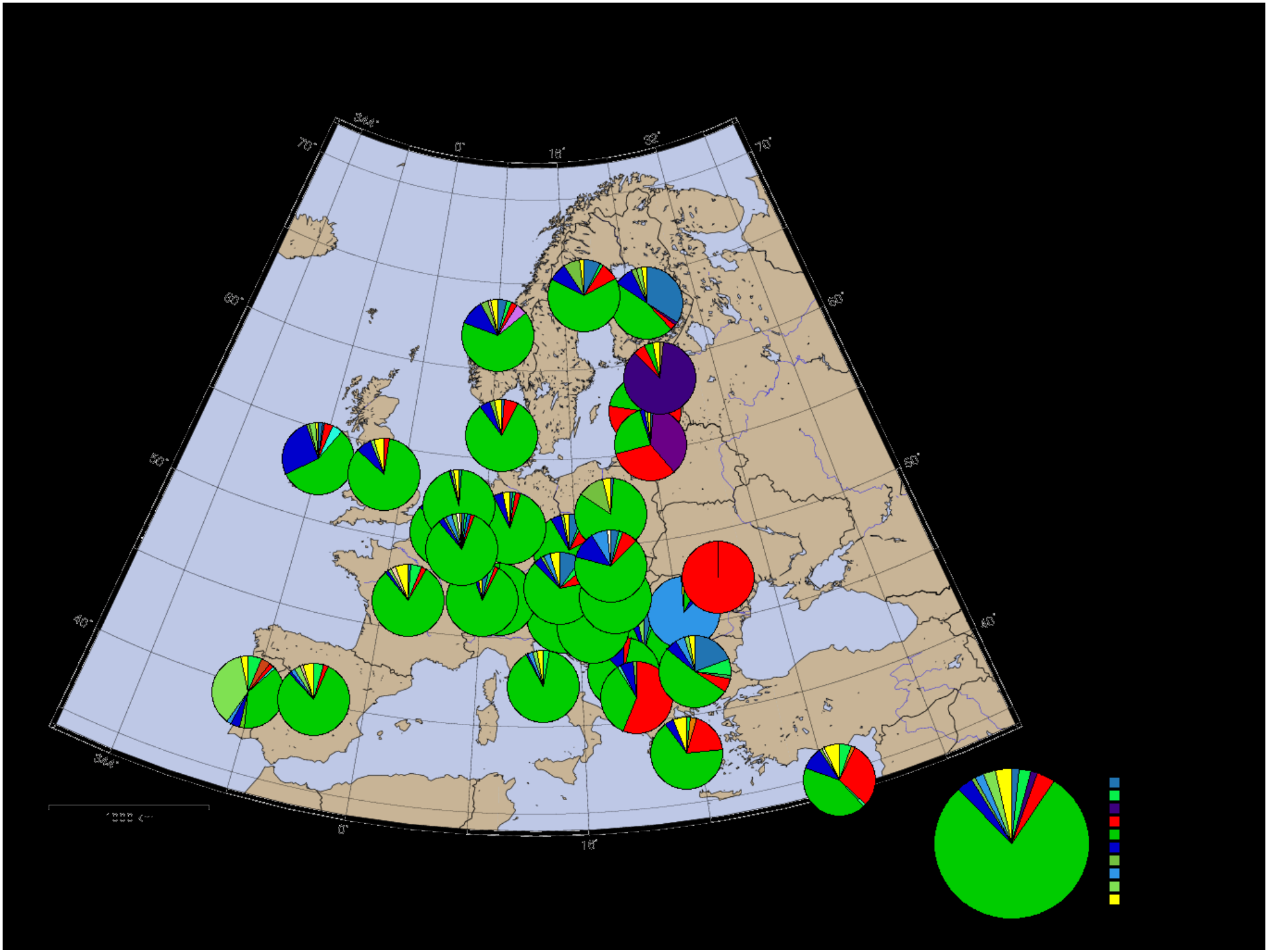
9 European countries:

Spain, United Kingdom, Germany, Austria, Italy, France, Norway, the Netherlands and Greece

Canada

Sequences per country / cohort

	SC Unknown n=5,905 N (%)	SC known n=3,050 N (%)	Overall N=8,955 N (%)
Country			
Canada	913 (15.5)	28 (0.9)	941 (10.5)
France	6 (0.1)	17 (0.6)	23 (0.3)
Germany	628 (10.6)	914 (30.0)	1542 (17.2)
Greece	0 (0.0)	35 (1.1)	35 (0.4)
Italy	719 (12.2)	378 (12.4)	1097 (12.3)
Netherlands	0 (0.0)	58 (1.9)	58 (0.6)
Norway	436 (7.4)	189 (6.2)	625 (7.0)
UK	394 (6.7)	1165 (38.2)	1559 (17.4)
Austria	1097 (18.6)	0 (0.0)	1097 (12.3)
Spain	1712 (29.0)	266 (8.7)	1978 (22.1)



Demographic and clinical characteristics (I)

	SC Unknown n=5,905	SC known n=3,050	Overall N=8,955
	N (%)	N (%)	N (%)
Sex			
Male	4202 (71.2)	2757 (90.4)	6959 (77.7)
Female	1158 (19.6)	293 (9.6)	1451 (16.2)
Unknown	545 (9.2)	0 (0.0)	545 (6.1)
Risk group			
MSM	2568 (47.8)	2412 (79.1)	4980 (59.2)
IDU	729 (13.6)	209 (6.9)	938 (11.1)
MSW	1706 (31.8)	381 (12.5)	2087 (24.8)
Haemophiliacs	12 (0.2)	0 (0.0)	12 (0.1)
Other-Unknown	353 (6.6)	48 (1.6)	401 (4.8)
Ethnic/racial group			
White	2299 (38.9)	1554 (51.0)	3853 (43.0)
Black	375 (6.4)	74 (2.4)	449 (5.0)
Other	371 (6.3)	166 (5.4)	537 (6.0)
Unknown	2860 (48.4)	1256 (41.2)	4116 (46.0)

Demographic and clinical characteristics (II)

	SC Unknown n=5,905 N (%)	SC known n=3,050 N (%)	Overall N=8,955 N (%)
Sampling date			
1987-02	1208 (20.5)	727 (23.8)	1935 (21.6)
2003-06	1368 (23.2)	870 (28.5)	2238 (25.0)
2007-08	1683 (28.5)	514 (16.9)	2197 (24.5)
2009-11	1633 (27.7)	327 (10.7)	1960 (21.9)
Not Available	13 (0.2)	612 (20.1)	625 (7.0)
SC date			
1981-96		767 (25.1)	767 (8.6)
1997-03		870 (28.5)	870 (9.7)
2004-06		671 (22.0)	671 (7.5)
2007-11		742 (24.3)	742 (8.3)

Proportion % in LTN by country and subtype

	B	C	A	CRF02	CRF01	Other (D,F,G)	Overall	
Country	**%**	**%**	**%**	**%**	**%**	**%**	**%**	**N**
Canada	47.43	24.00	18.24	30.43	36.36	21.74	40.70	941
France	29.41						21.74	23
Germany	62.83	29.79	20.00	12.90	17.50	12.50	58.43	1542
Greece	23.81			16.67			20.00	35
Italy	16.11	18.18		57.14		43.24	17.14	1097
Netherlands	12.07						12.07	58
Norway	43.54	27.27	23.58	14.29	28.89	40.00	36.16	625
UK	40.14		6.90	13.33		25.00	37.46	1559
Austria	42.03	25.35	18.64	28.05	16.22	49.30	37.37	1097
Spain	51.03	42.70	43.75	54.55		22.22	50.00	1978
Total	**44.40**	**28.43**	**18.71**	**26.15**	**18.75**	**35.85**	**41.32**	**8955**

Multivariate logistic regression for probability belonging in a LTN

Factor	OR	95% C.I.	p
Sex (*vs.* male)			
Female	0.66	(0.56, 0.78)	<0.001
Unknown	0.44	(0.04, 5.10)	0.514
Subtype (*vs.* B)			
CRF02_AG	0.70	(0.53, 0.94)	0.016
C	0.51	(0.38, 0.69)	<0.001
A	0.65	(0.48, 0.89)	0.007
CRF01_AE	0.36	(0.24, 0.54)	<0.001
Other (D,F,G)	1.04	(0.76, 1.42)	0.814
Country (*vs.* Spain)			
Canada	1.30	(0.98, 1.73)	0.069
France	0.33	(0.09, 1.23)	0.098
Germany	1.43	(1.20, 1.69)	<0.001
Greece	0.24	(0.10, 0.58)	0.002
Italy	0.55	(0.41, 0.75)	<0.001
Netherlands	0.28	(0.12, 0.65)	0.003
Norway	0.97	(0.71, 1.32)	0.846
UK	0.72	(0.52, 0.99)	0.042
Austria	0.97	(0.74, 1.27)	0.823

Factor	OR	95% C.I.	p
Sampling date (*vs.* 1987-02)			
2003-06	2.01	(1.67, 2.43)	<0.001
2007-08	2.38	(1.95, 2.91)	<0.001
2009-11	2.61	(2.12, 3.19)	<0.001
SC known			
Yes/No	1.44	(1.23, 1.69)	<0.001
Risk group (*vs.* MSM)			
IDU	0.62	(0.52, 0.74)	<0.001
MSW	0.69	(0.59, 0.80)	<0.001
Haemo	0.27	(0.06, 1.27)	0.097
Other-NK	0.55	(0.42, 0.72)	<0.001
Ethnic group			
Black	0.44	(0.32, 0.62)	<0.001
Other	0.70	(0.55, 0.88)	0.002
Unknown	0.91	(0.72, 1.17)	0.467
ART Naïve			
Yes/No	1.19	(1.04, 1.35)	0.010
Age (10 yrs)			
	0.79	(0.75, 0.84)	<0.001

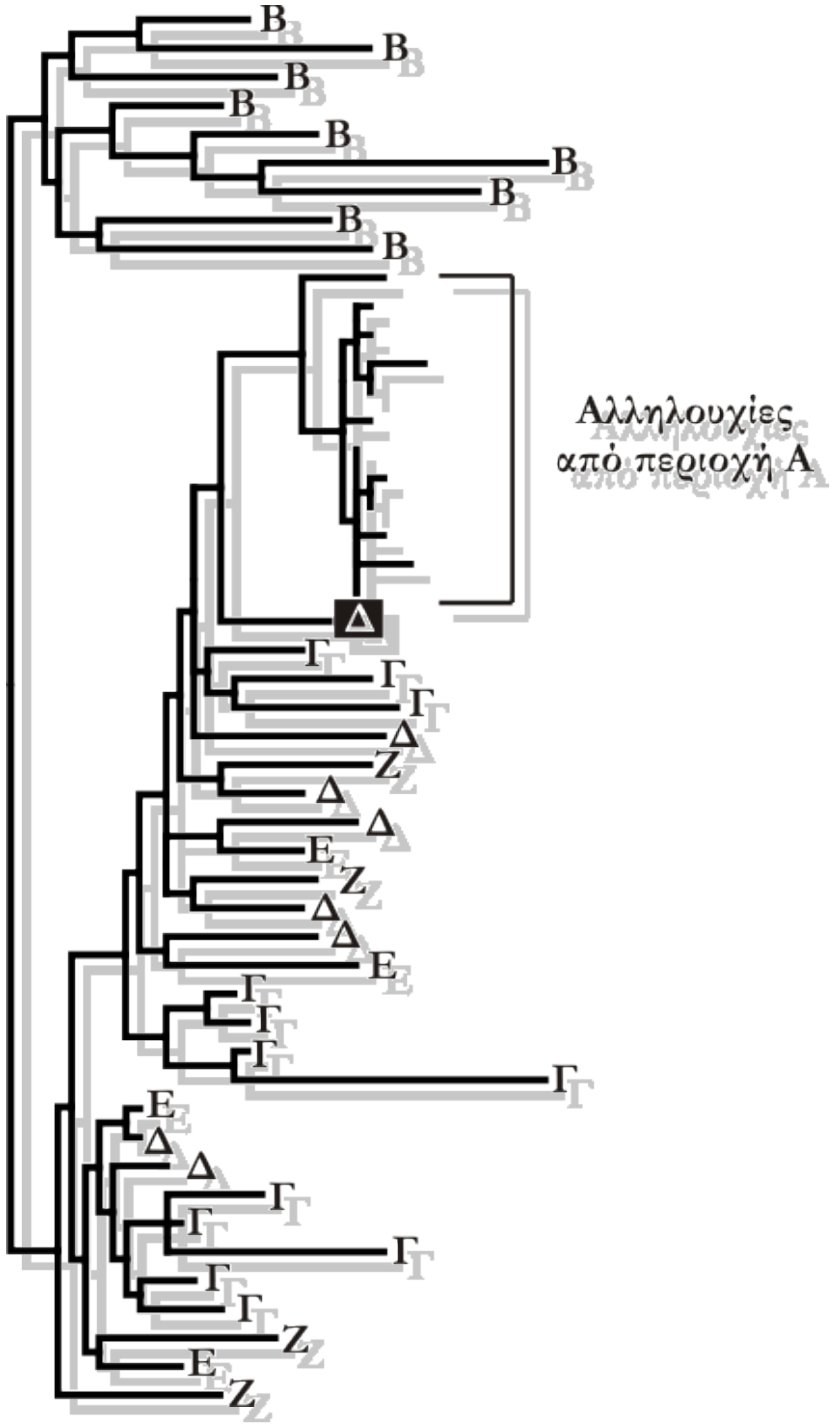
Παράδειγμα 3

Διερεύνηση του τρόπου διασποράς μια επιδημίας

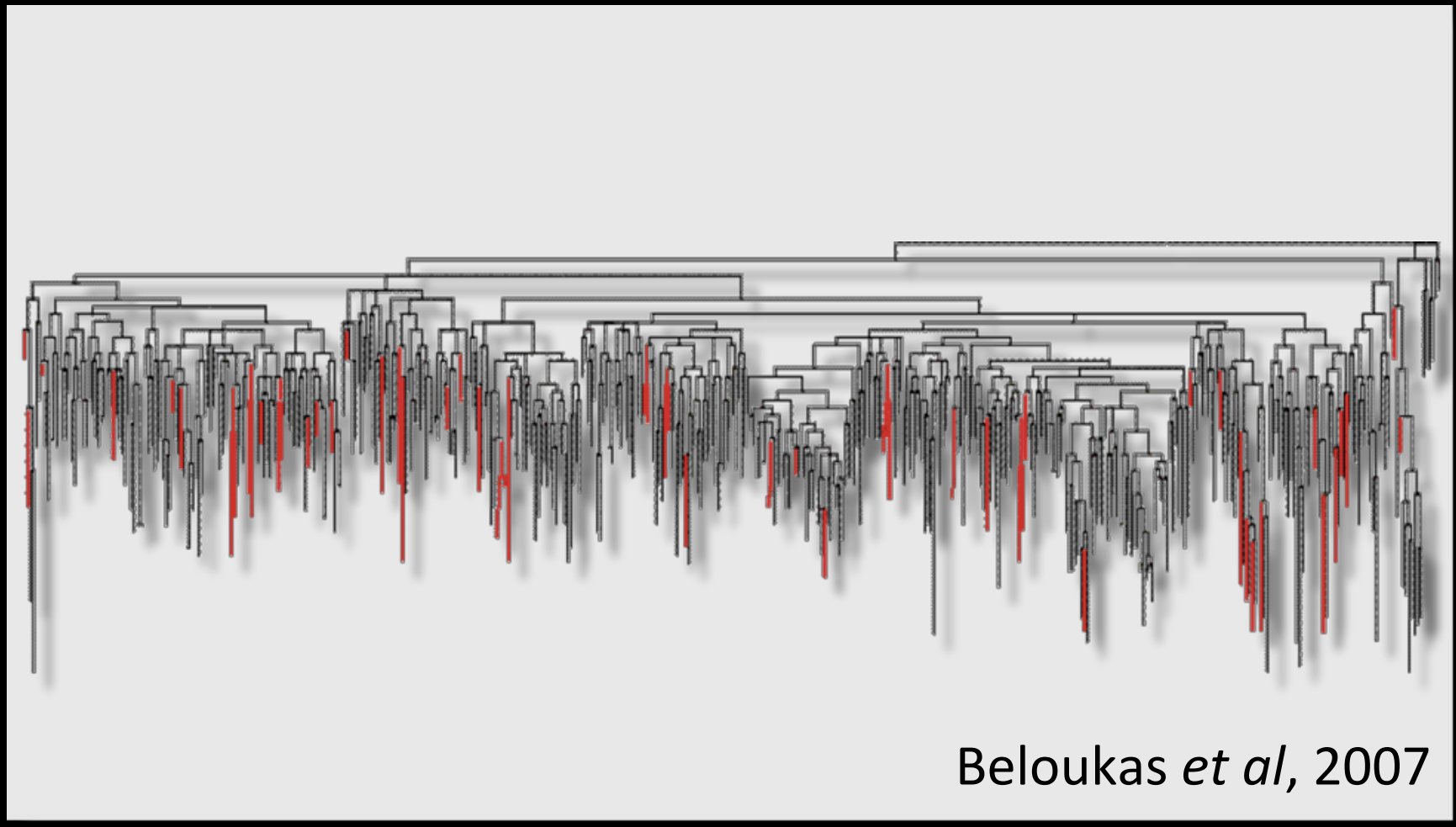
Διερεύνηση τρόπου διασποράς και γεωγραφικής προέλευσης και μιας επιδημίας

- ✓ Ο τρόπος εισαγωγής μιας επιδημίας σε μια γεωγραφική περιοχή, μπορεί να διερευνηθεί συγκρίνοντας το γενετικό υλικό του ιού που έχει απομονωθεί από άτομα σε συγκεκριμένη γεωγραφική περιοχή, με το αντίστοιχο από υποδόχα από διαφορετικές περιοχές

- ✓ Η επιδημία μπορεί να έχει προκληθεί από μια ή περισσότερες πηγές

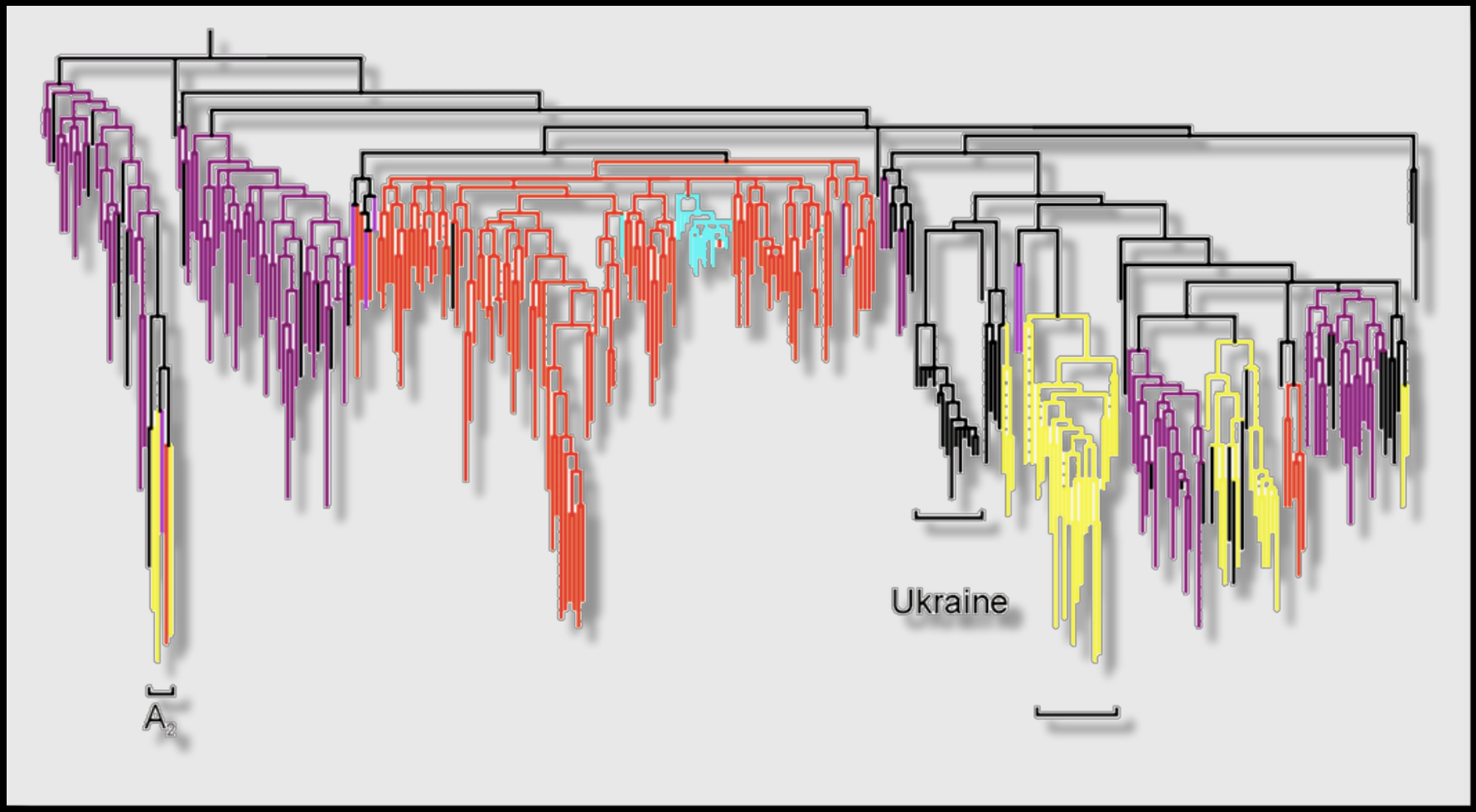

Β
Β
Β
Β
Β
Β
Β
Β
Β
Αλληλουχίες
από περιοχή Α
Δ
Γ
Γ
Γ
Δ
Ζ
Δ
Δ
Ε
Ζ
Δ
Δ
Ε
Γ
Γ
Γ
Γ
Ε
Δ
Δ
Γ
Γ
Γ
Γ
Γ
Ζ
Ε
Ζ

Επιδημία HIV-1 υπότυπος C



Beloukas *et al*, 2007

Μονοφυλετική επιδημία του HIV-1 υπότυπου Α στην Ελλάδα

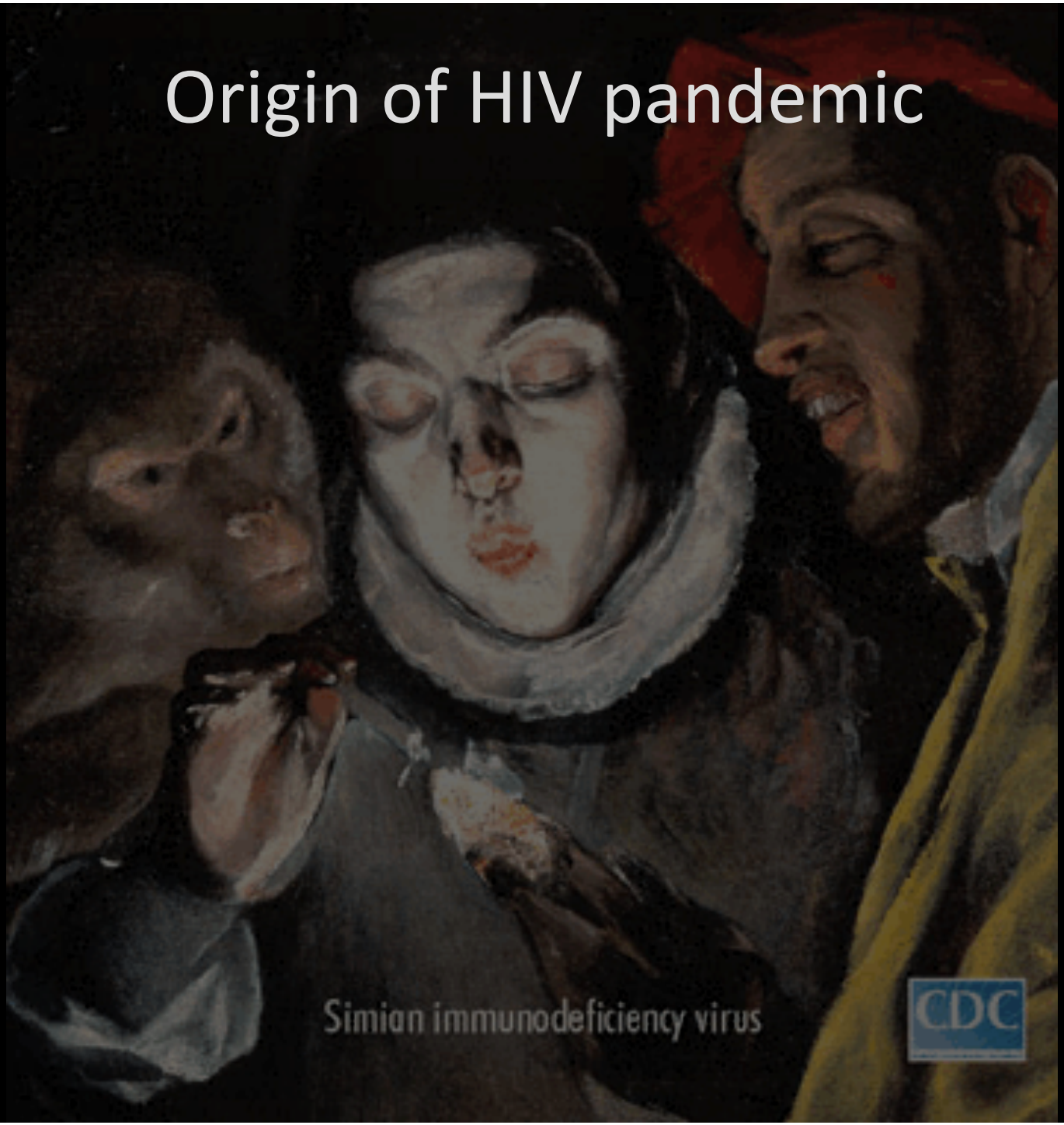


Beloukas *et al*, 2007

Παράδειγμα 3

Διερεύνηση της αρχής και του τρόπου εξέλιξης μιας επιδημίας

Origin of HIV pandemic



Evolution of HIV

HIV belongs to lentivirus subgroup of retroviruses.

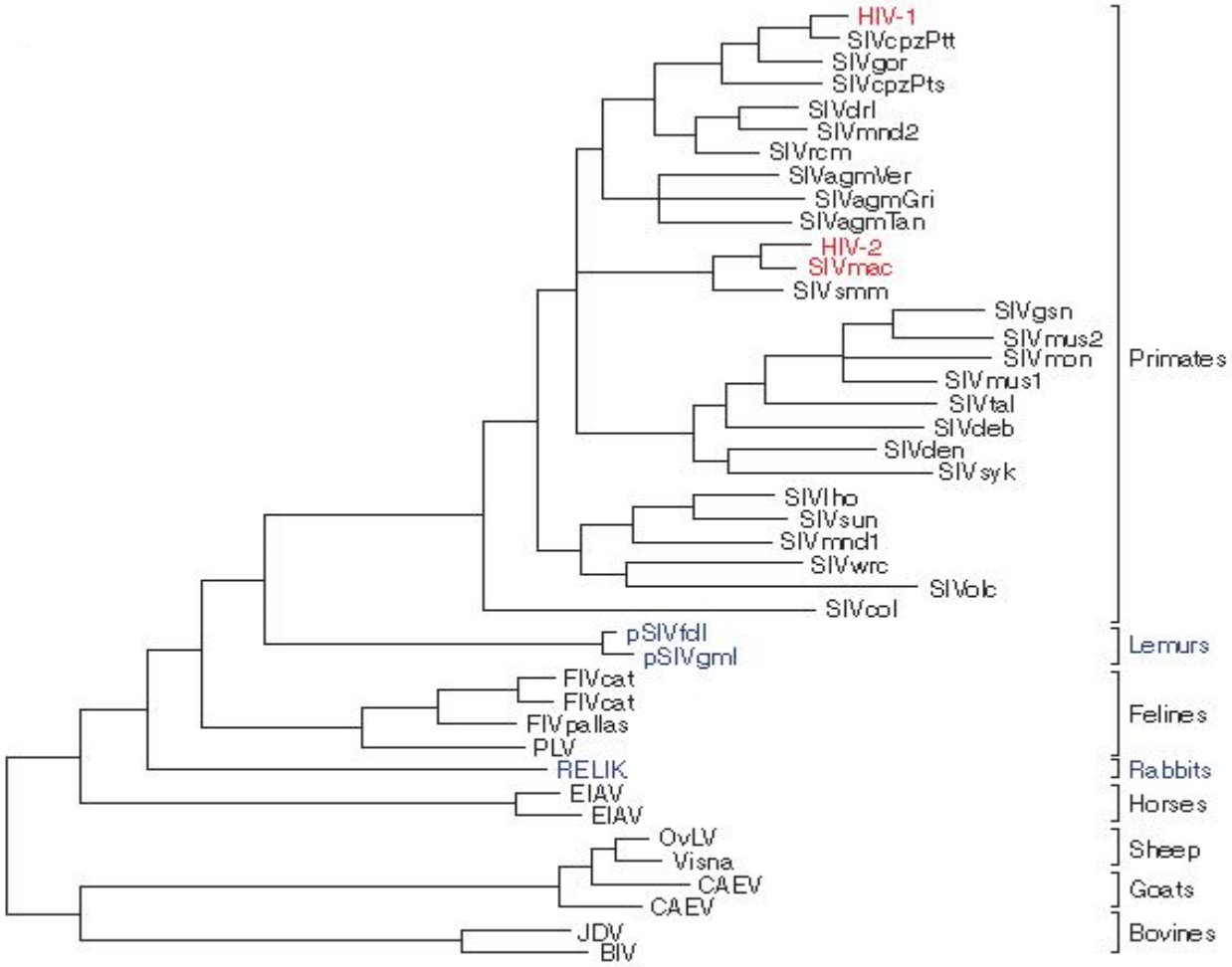
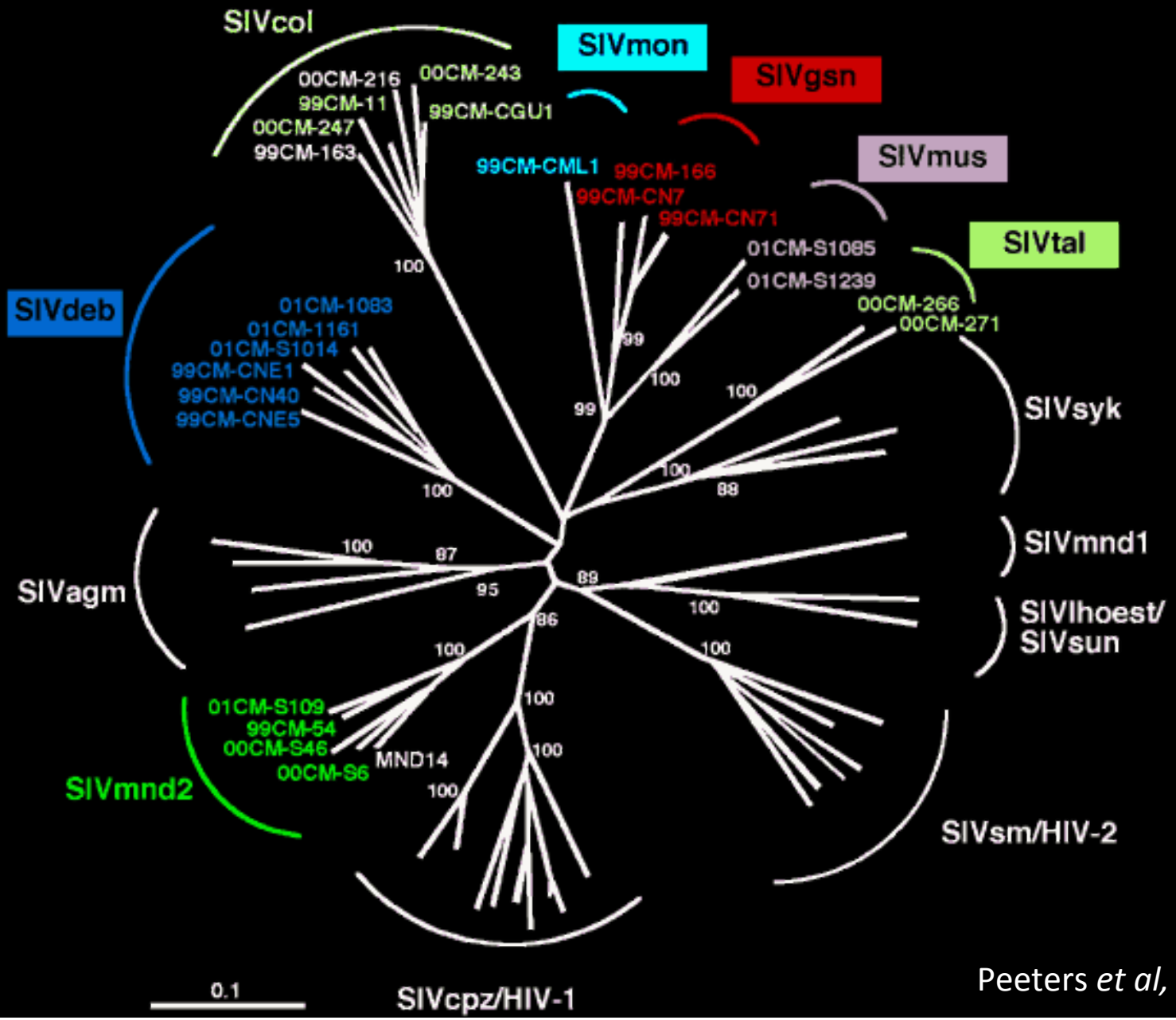


Figure 2. Phylogeny of lentiviruses. The evolutionary relationships among Pol sequences (~ 770 amino acids) derived from various mammalian lentiviruses; host species are indicated at the right. Exogenous viruses are depicted in black, with HIV-1, HIV-2, and SIVmac highlighted in red; endogenous viruses are shown in purple. The phylogenetic tree was estimated using maximum likelihood methods (Guindon and Gascuel 2003). The scale bar represents 0.10 amino acid replacements per site.

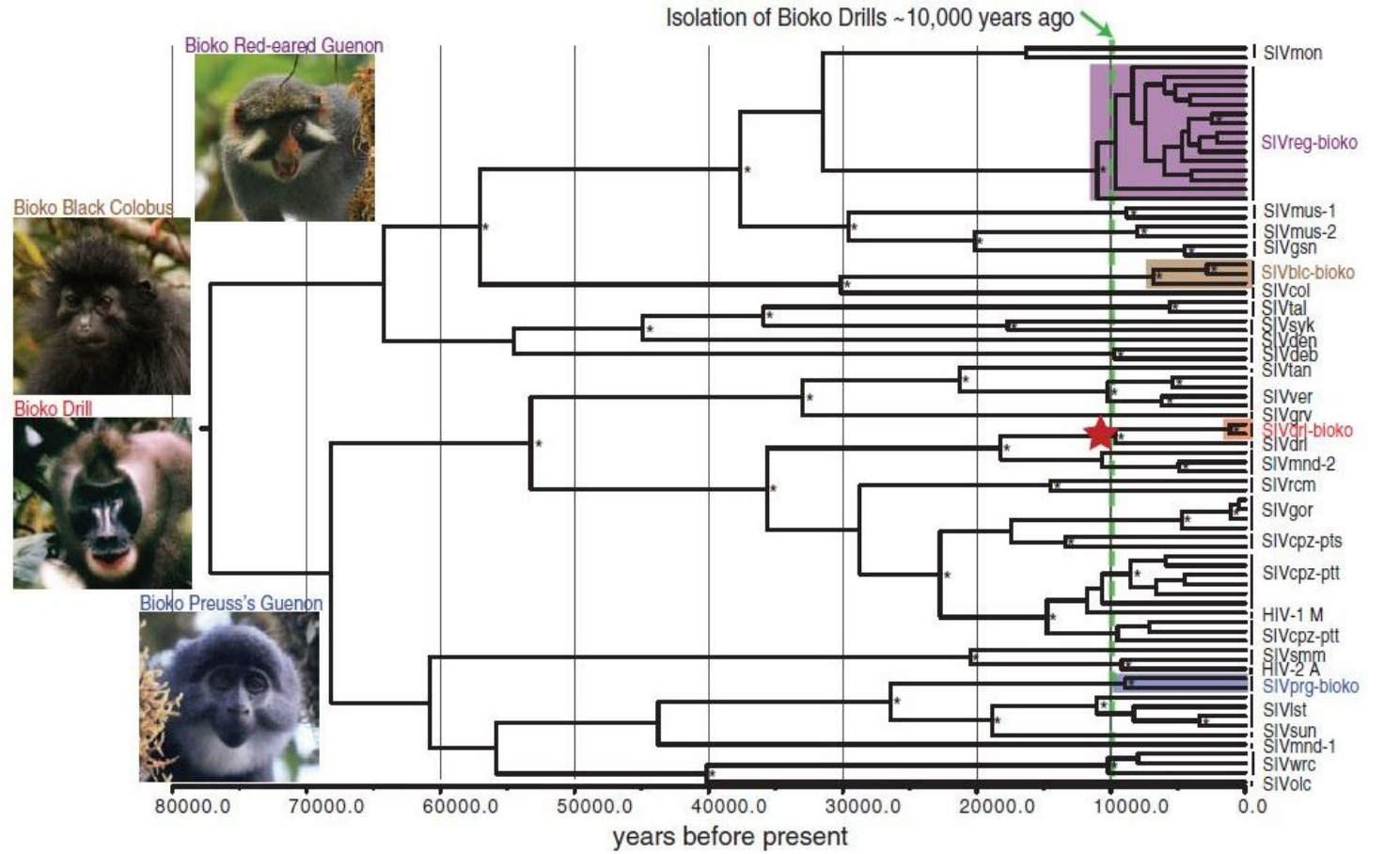
Sharp and Hahn 2011

SIVs from different primate species



Peeters *et al, 2002*

Dating SIVs origin



➢Worobey et al., 2010, taking SIV samples from a wider regions, substantiated the earlier conclusion and found the infection to be present at least 30,000 years ago.

Origin of AIDS pandemic Virus

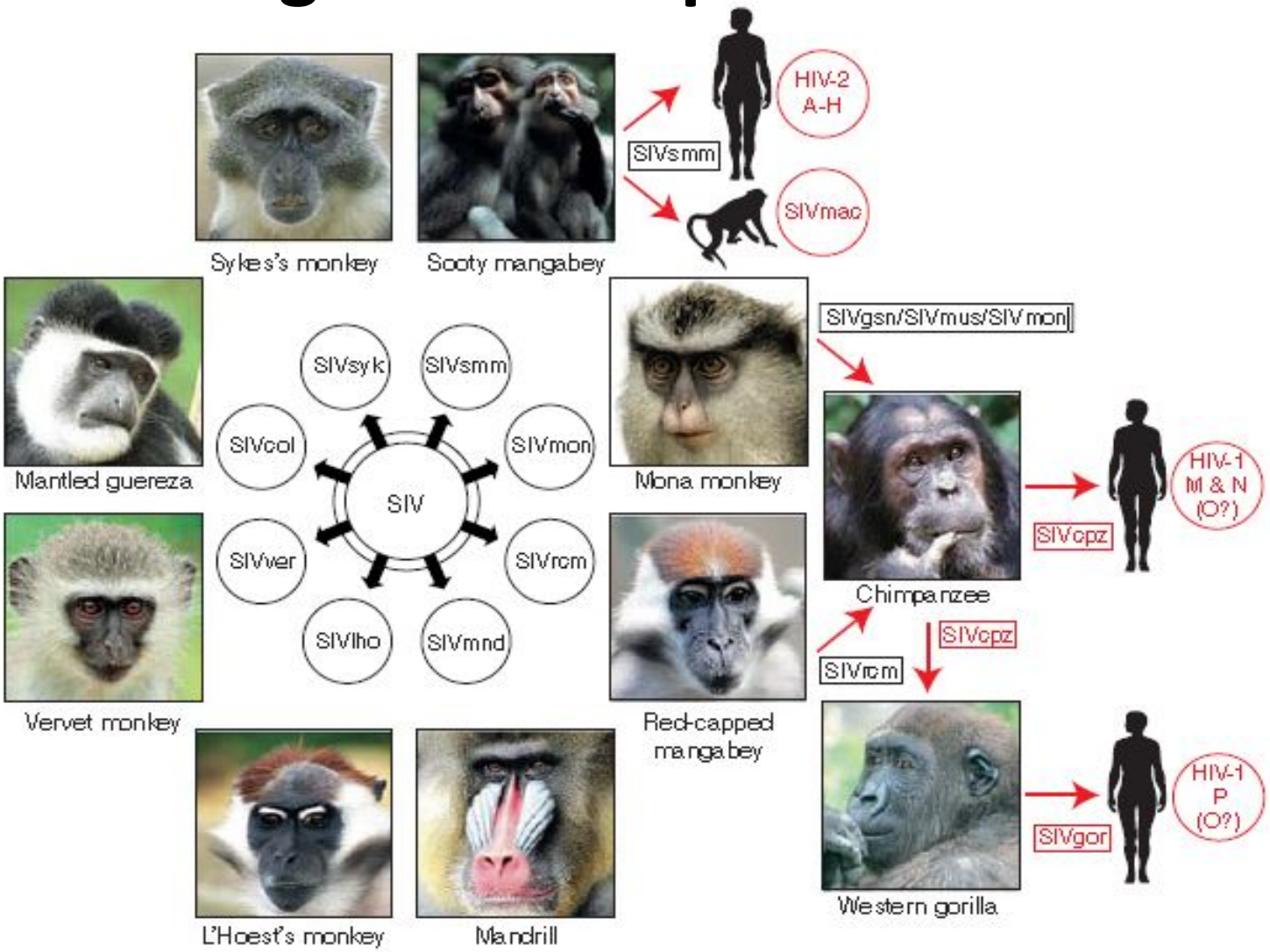


Figure 1. Origins of human AIDS viruses. Old World monkeys are naturally infected with more than 40 different lentiviruses, termed simian immunodeficiency viruses (SIVs) with a suffix to denote their primate species of origin (e.g., SIVsmm from sooty mangabeys). Several of these SIVs have crossed the species barrier to great apes and humans, generating new pathogens (see text for details). Known examples of cross-species transmissions, as well as the resulting viruses, are highlighted in red.

Sharp and Hahn 2011

Sooty mangabeys

Chimpanzee *(Pan troglodytes troglodytes)*

Origin of HIV-1

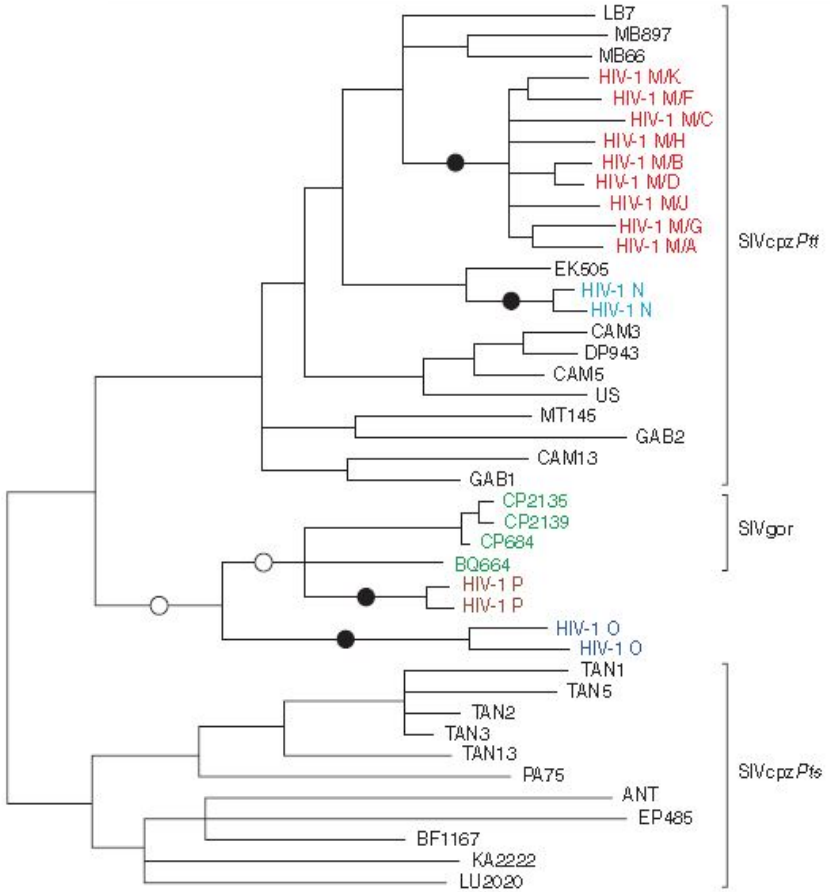


Figure 4. HIV-1 origins. The phylogenetic relationships of representative SIVcpz, HIV-1, and SIVgor strains a shown for a region of the viral *pol* gene (HIV-1/HXB2 coordinates 3887 4778). SIVcpz and SIVgor sequenc are shown in black and green, respectively. The four groups of HIV-1, each of which represents an independe cross-species transmission, are shown in different colors. Black circles indicate the four branches where cro species transmission-to-humans has occurred. White circles indicate two possible alternative branches on whi chimpanzee-to-gorilla transmission occurred. Brackets at the right denote SIVcpz from *P. t. troglodytes* (SIVcpz*Ptt*) and *P. t. schweinfurthii* (SIVcpz*Pts*), respectively.

Origin of HIV-2

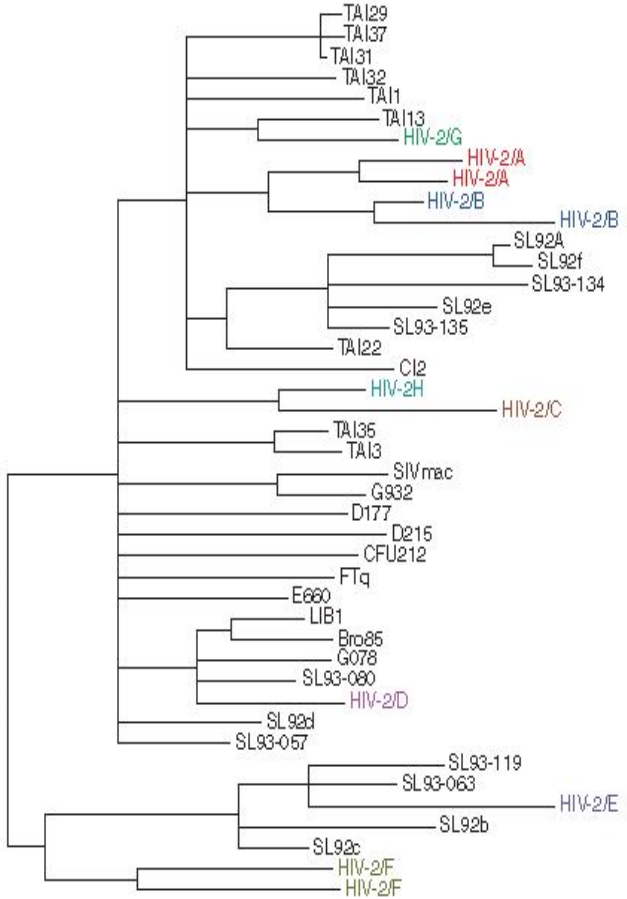


Figure . HIV-2 origins. The phylogenetic relationships of representative SIVsmm and HIV-2 strains are shown for a region of the viral *gag* gene (SIVmac239 coordinates 1191 1921). SIVsmm and SIVmac are shown in black; the eight groups of HIV-2, each of which represents an independent cross-species transmission, are shown in different colors. The phylogenetic tree was estimated using maximum likelihood methods (Guindon and Gascuel 2003). The scale bar represents 0.05 nucleotide substitutions per site.

Sharp & Hahn 2011