

# Module 4: Introduction to Phylogenetics and Public Health II

Ifeanyi Ezeonwumelu, PhD

PATHOGEN MULTIMICS AND BIOINFORMATICS III

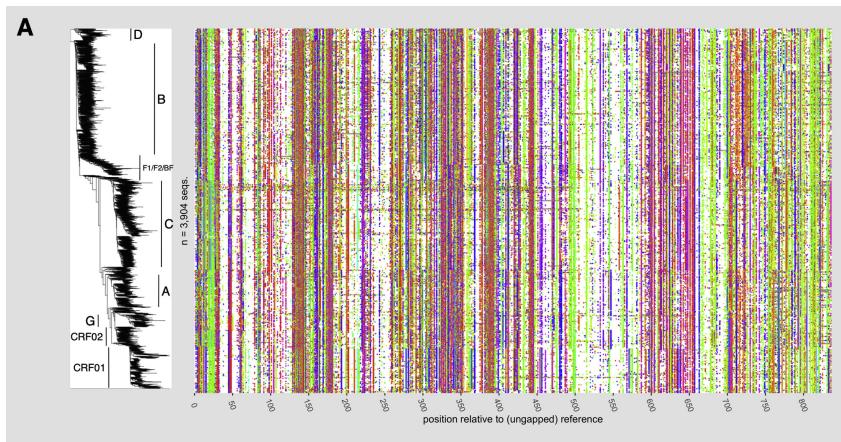
05.07.2023

**GLADSTONE**  
**INSTITUTES**

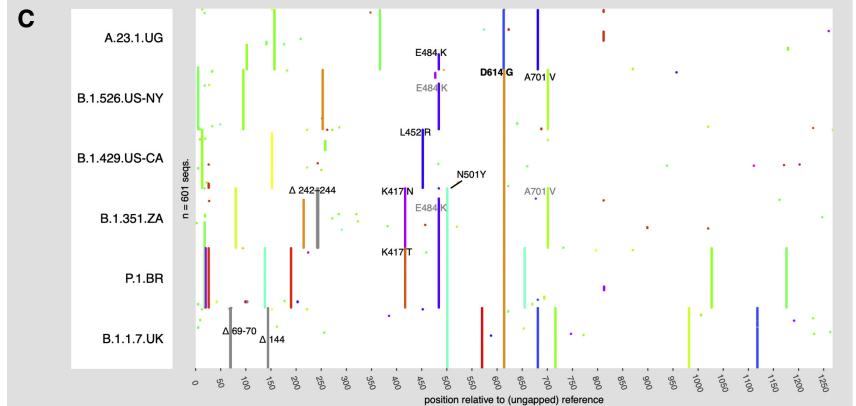
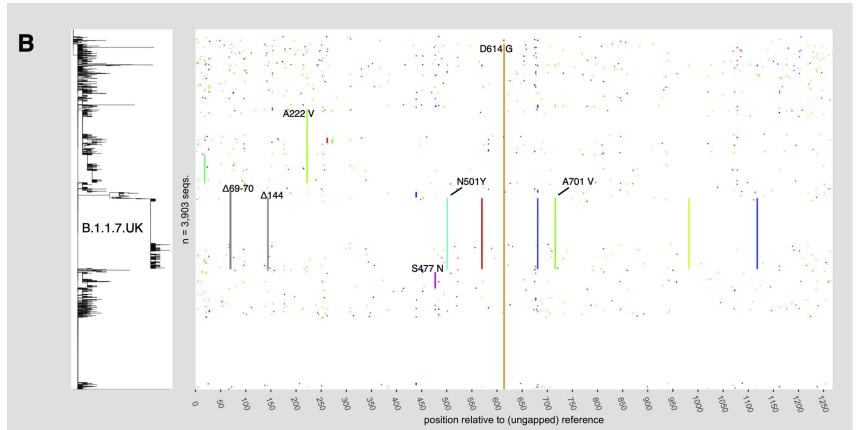


# Viral pandemics: virus evolution matters!

"All viruses are equal, but some viruses are more equal than others"



Variant-visualized amino-acid sequence alignments of HIV-1 Env (A) and SARS-CoV-2 Spike (B and C).



# Evolutionary rates differ across pathogens

**Mutation rate:** average number of errors created in genomes of viral progeny, per base, per replication cycle (mut/nuc/rep).  $1.5 \times 10^{-3} - 10^{-8}$  mut/nuc/rep

**Substitution rate, or *Evolutionary rate*** is the rate at which mutations fix in a population, measured by comparing the genomes of different isolates of a virus collected at several different time points.  $0.01 - 7 \times 10^{-7}$  substitutions per site per year (sub/site/yr).

**Table 1. Estimated short-term and maximum substitution rate SSY according to the PoW model**

Viral group	Type of virus	Short-term substitution rate, $\langle \mu \rangle$	Fastest rate group, $\mu_{\max}$
I	dsDNA virus	$2(0.3 - 16) \times 10^{-5}$	$3(0.6 - 10) \times 10^{-3}$
II	ssDNA virus	$3(1 - 6) \times 10^{-4}$	$2(1 - 3) \times 10^{-2}$
IV	(+)ssRNA virus	$2(1 - 4) \times 10^{-3}$	$4(3 - 6) \times 10^{-2}$
V	(-)ssRNA virus	$1(0.7 - 3) \times 10^{-3}$	$4(3 - 6) \times 10^{-2}$
VI	RT-RNA virus	$1(0.7 - 3) \times 10^{-3}$	$4(3 - 6) \times 10^{-2}$
VII	RT-DNA virus	$5(1 - 20) \times 10^{-5}$	$4(2 - 10) \times 10^{-3}$

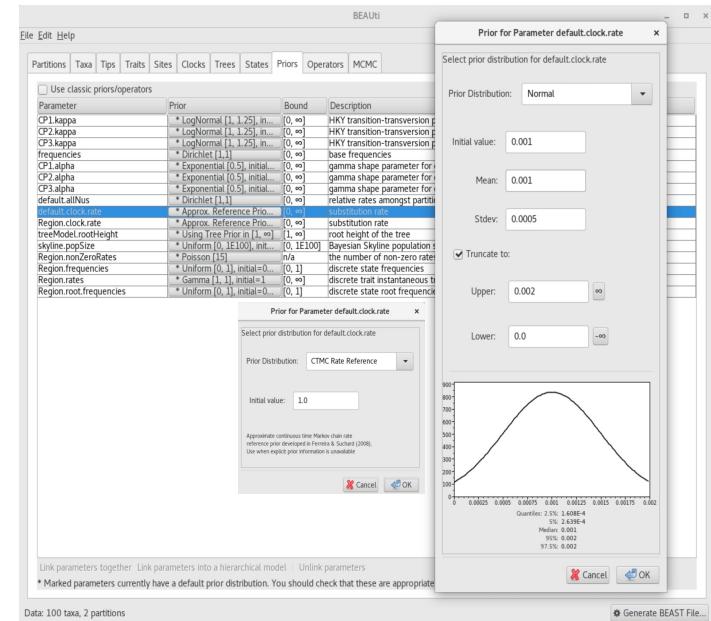
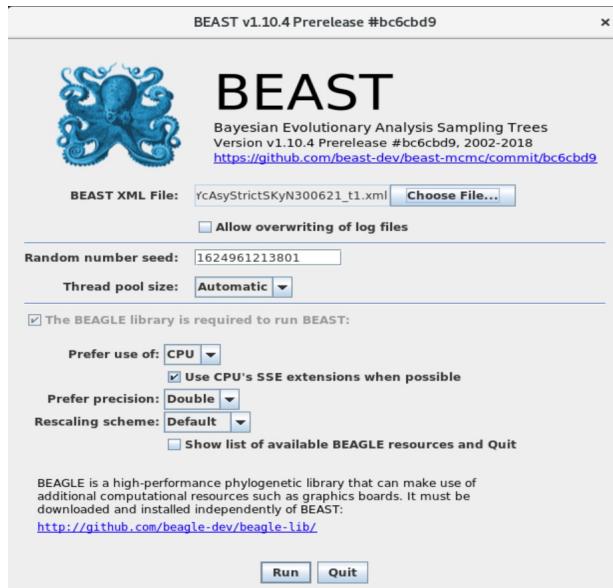
Much higher in RNA viruses (HIV, HCV) than in DNA viruses (HSV, Monkeypox virus)

# BEAUTi and the BEAST



The BEAST package contains a suite of programmes for computing Bayesian Evolutionary Analysis and Trees Sampling.

**BEAUTi | Bayesian Evolutionary Analysis Utility:**  
This program is used to import data (*FASTA* & *NEXUS*), design the analysis, and generate the BEAST control (*XML*) file.



**BEAST | Bayesian Evolutionary Analysis Sampling Trees:**  
This is the main program that takes as input, *XML* files generated by BEAUTi and performs the analysis.

# BEAST: Output Analysis



LogCombiner | This is a utility program that will combine log & tree files from different runs and reduce the sampling frequency (thin them).

TreeAnnotator | This is a post-analysis program that will produce a summary tree from the output of BEAST. Takes as input .tree files to generate a Maximum clade credibility tree.

Not included in the BEAST package but indispensable in BEAST Output Analysis



Tracer:

A graphical program for exploring the output of BEAST, diagnosing problems, and summarizing the results.



FigTree:

A graphical program for viewing trees, displaying summary information from TreeAnnotator and generating publication-ready figures.

# SpreeD3: Time-scaled Phylogreography

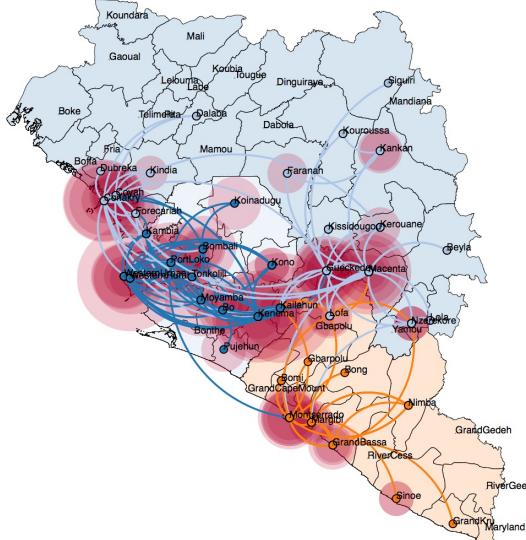


SpreeD3 | (Spatial Phylogenetics Reconstruction of Evolutionary Dynamics using Data-Driven Documents (D3)):

A package for analysis and visualisation of pathogen phylodynamic reconstructions

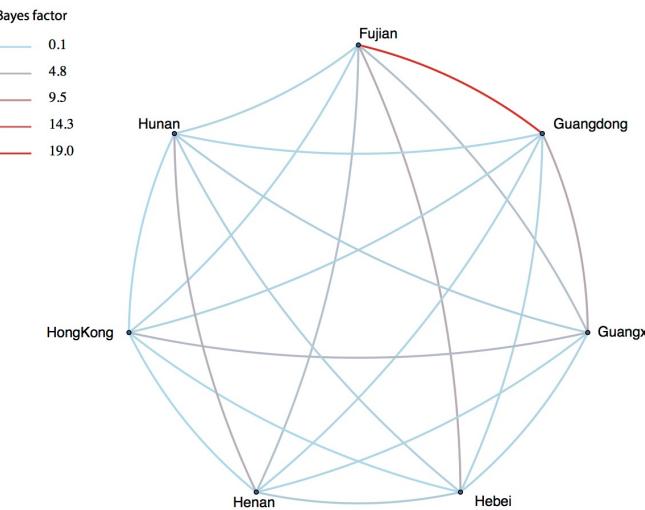
## Input:

- .mcc tree files & region.rates.log files (BEAST-BSSVS)
- .geoJSON file (map locations)
- Geographical coordinates text files



## Output:

- Interactive visualizations of spatio-temporal phylogenies
- Bayes factor analysis for epidemic migrations



# Phylogeography Practice Session:

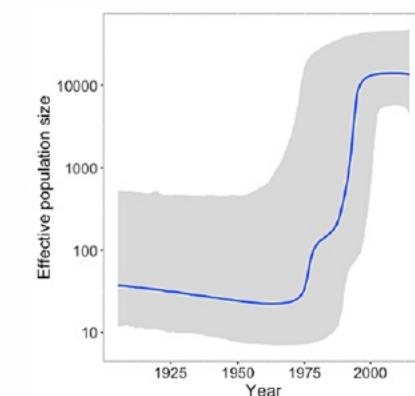
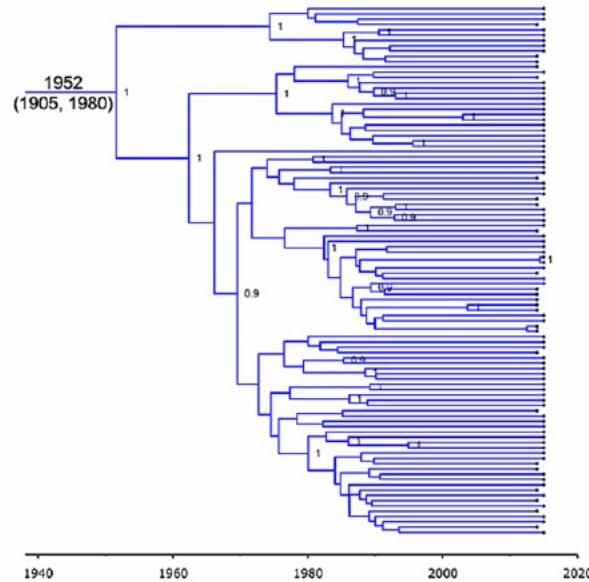
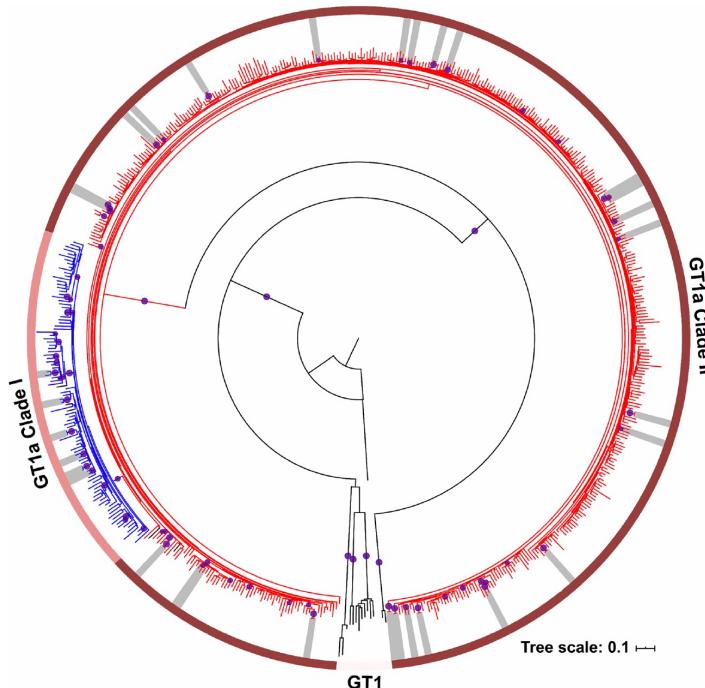
## Case study

SCIENTIFIC  
REPORTS  
nature research

 Check for updates

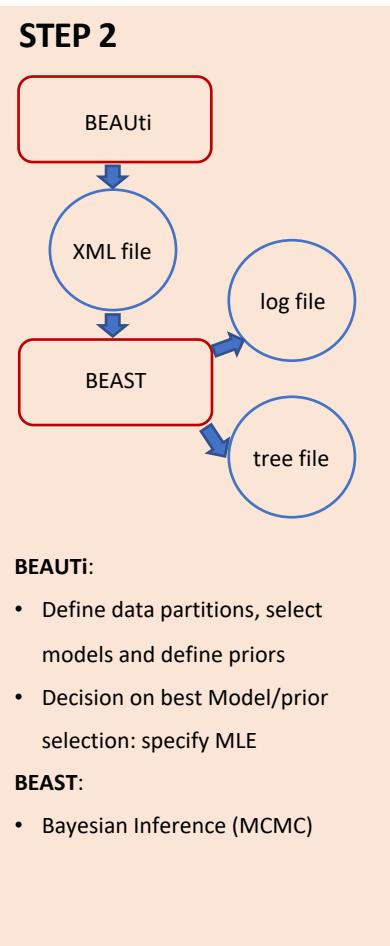
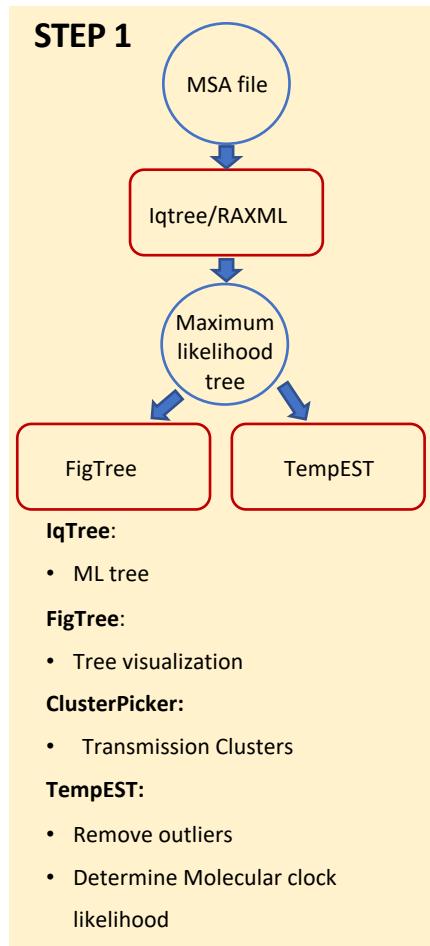
### Epidemic history and baseline resistance to NS5A-specific direct acting drugs of hepatitis C virus in Spain

Claudia Palladino<sup>1,4</sup>, Ifeanyi Jude Ezeonwumelu<sup>1,4</sup>, Irene Mate-Cano<sup>2</sup>,  
Pedro Borrego<sup>1</sup>, Paula Martínez-Román<sup>2</sup>, Sonia Arca-Lafuente<sup>1,2</sup>, Salvador Resino<sup>1,2</sup>,  
Nuno Taveira<sup>1,3</sup> & Verónica Briz<sup>1,2</sup>

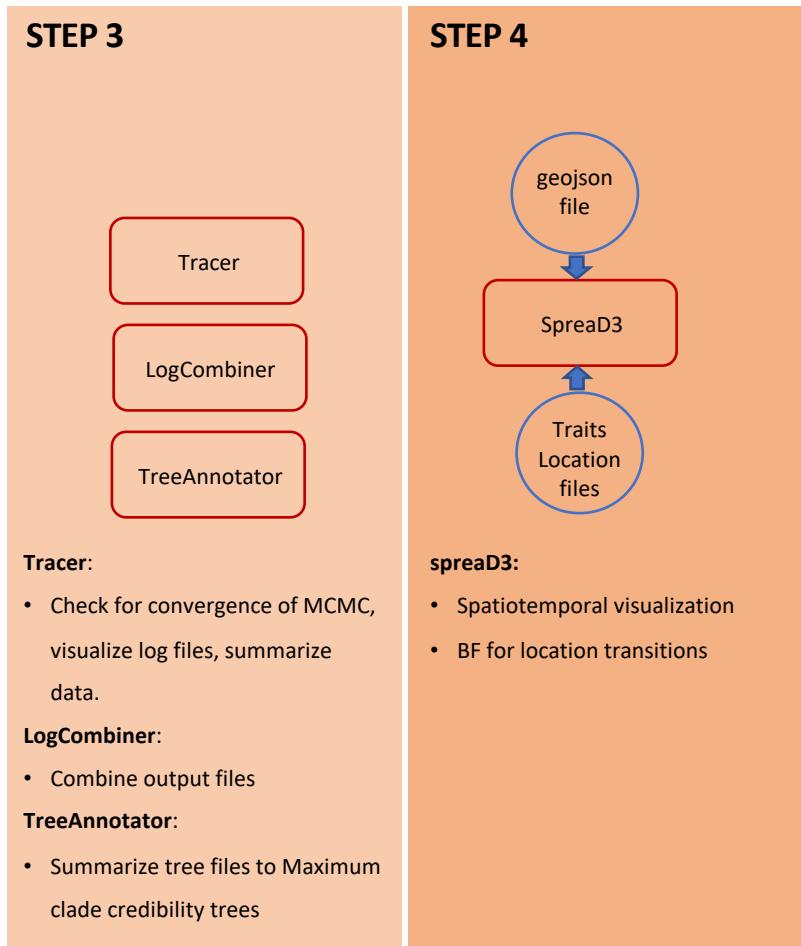


# Phylogeography Practice Session II: OUTLINE

## Session I



## Session II

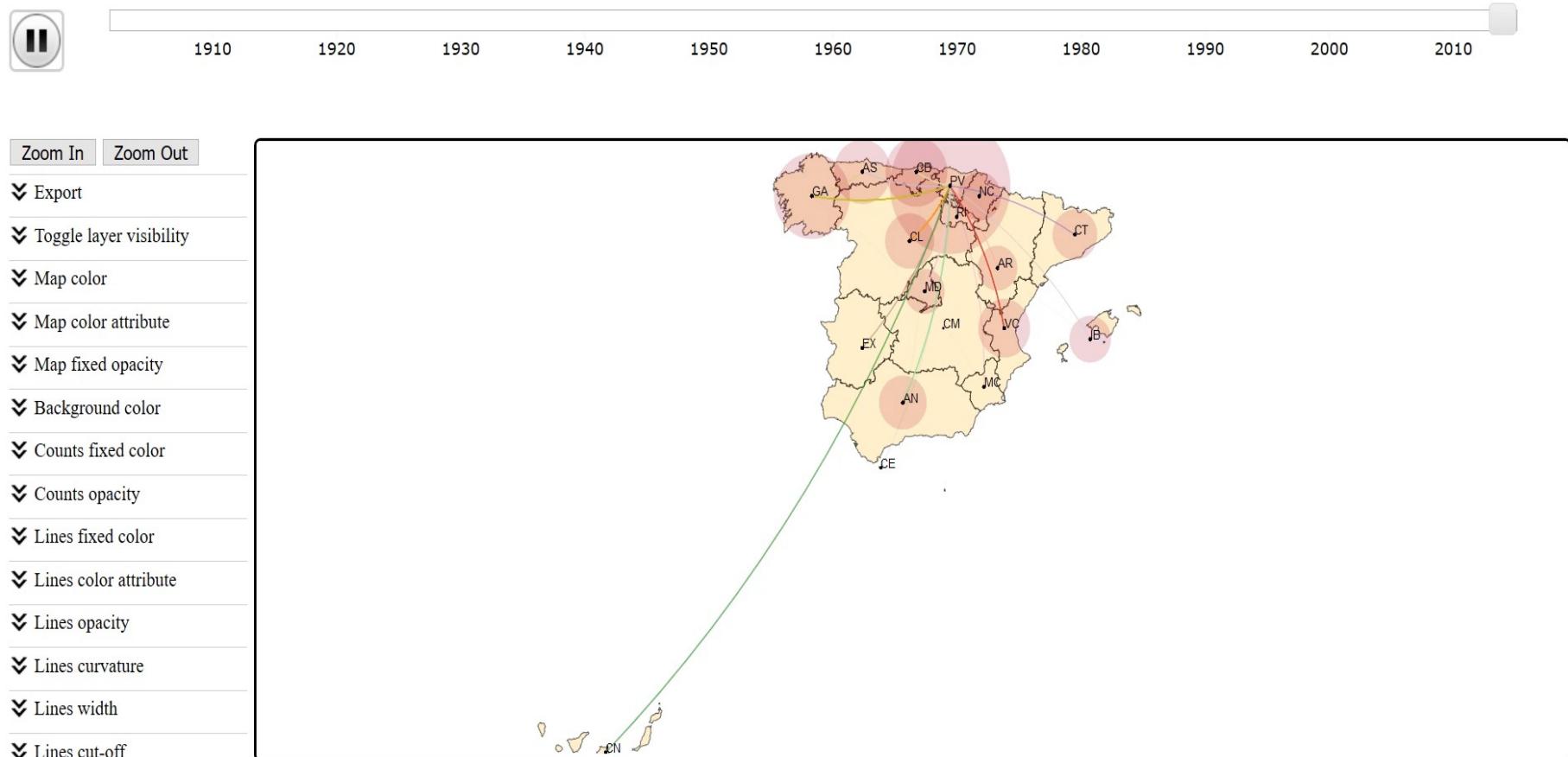


# Phylogeography Practice Session: Case study resolution

HCV GT1a Clade II

Current date: 2013/12/14

www.BANDICAM.com

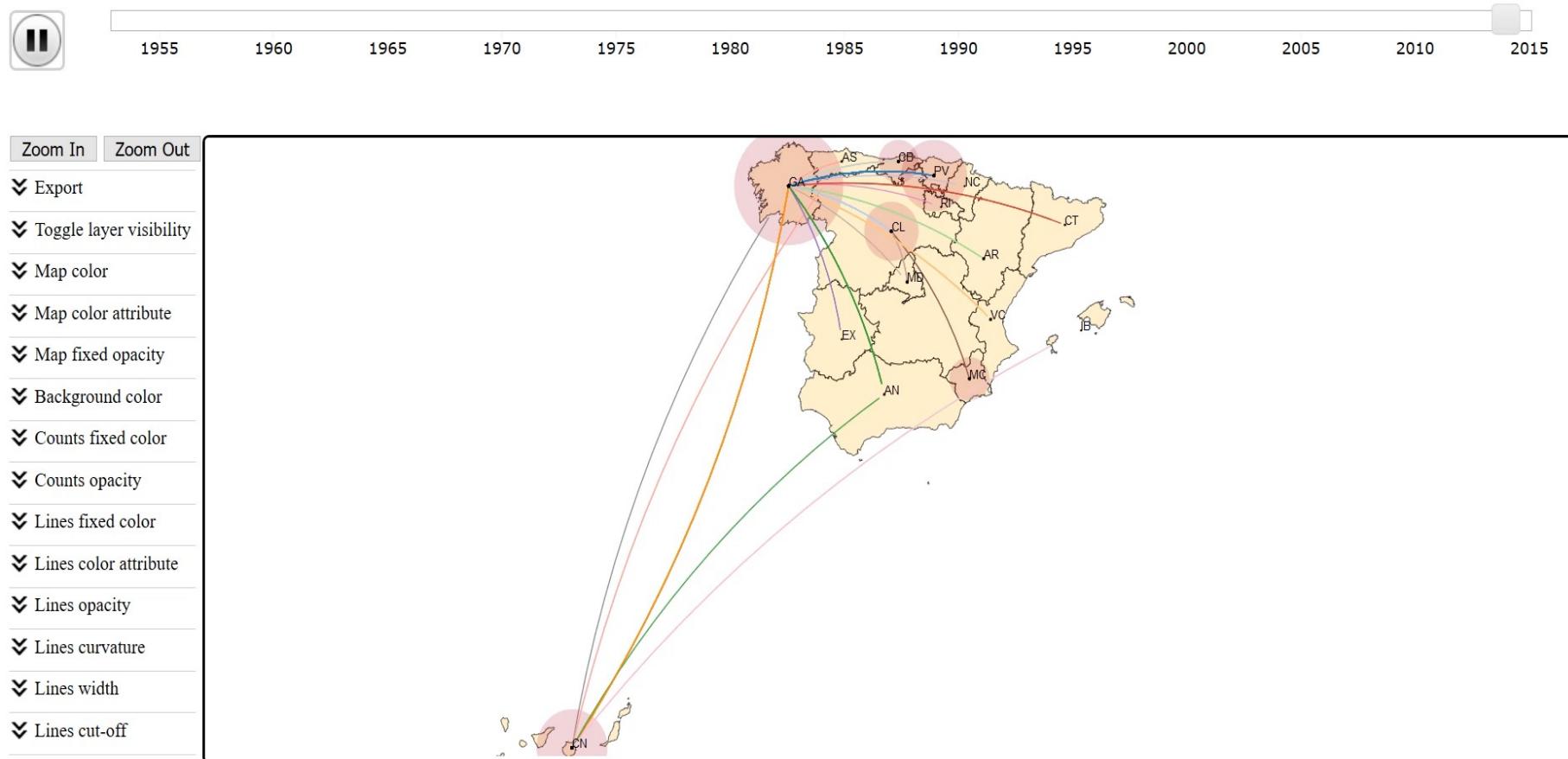


# Phylogeography Practice Session: Case study resolution

## HCV GT1a Clade I

Current date: 2014/06/18

www.BANDICAM.com



# Phylogeography Practice Session:

## Case study resolution

Clade I				Clade II			
Migration		Bayes factor <sup>a</sup>	Posterior probability (%)	Migration		Bayes factor <sup>a</sup>	Posterior probability (%)
From	To			From	To		
Andalusia	Balearic Islands	48.06	77.08	Andalusia	La Rioja	181.91	91.78
	Basque Country	20.69	59.14		Extremadura	3.28	16.76
	Galicia	13.51	48.60	Aragon	Catalonia	8.86	35.21
	Asturias	10.08	41.36	Basque Country	Andalusia	192.13	92.18
	Murcia	7.80	35.31		Madrid	170.24	91.26
Asturias	Madrid	32.03	69.14	Catalonia	Aragon	5.34	24.67
	Canary Islands	24.65	63.30	Extremadura	Canary Islands	3.15	16.19
	Andalusia	21.79	60.39	Galicia	Navarra	4.23	20.60
	Galicia	10.81	43.06	La Rioja	Asturias	4.71	22.41
	Castile Leon	9.45	39.80	Madrid	Galicia	42.61	72.34
	Catalonia	3.425	19.33		La Rioja	3.13	16.11
Cantabria	Extremadura	3.18	18.18	Murcia	Castile Leon	15.94	49.45
Catalonia	Extremadura	3.29	18.73	Navarra	Murcia	8.20	33.48
Galicia	Basque Country	3.36	19.05		Galicia	5.99	26.89
Madrid	Catalonia	41.15	74.22				
	Cantabria	29.61	67.44				
	Valencia	5.95	29.38				
	Murcia	5.22	26.75				
	Castile Leon	4.8	25.19				
	Extremadura	4.06	22.10				

# Phylogeography Practice Session:

## Case study conclusion

Clade II: Older, Basque country

Clade I: Younger, Madrid & Andalucia

