

Population genomics: Background, tools and programming

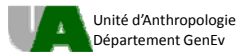


## Coalescent simulations in time and space

April 1<sup>st</sup>, 2019 – Procida

**Mathias Currat**

*Department of Genetics and Evolution – Anthropology Unit  
Institute for Genetics and Genomics in Geneva (IGE3)  
University of Geneva, Switzerland*



### Outline

---

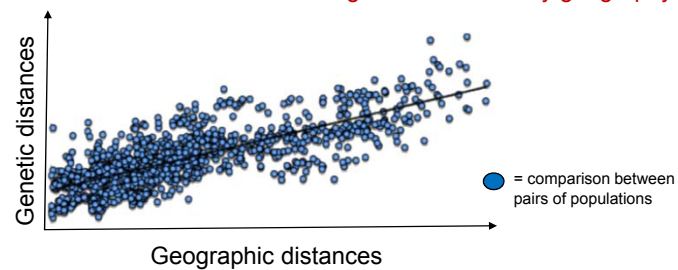
1. Geography and Genetic Diversity
2. Spatially Explicit Simulations
3. Coalescent in Space and Time
4. Testing Population Continuity through Time
5. Practicals

## Genetic consequences of past migrations



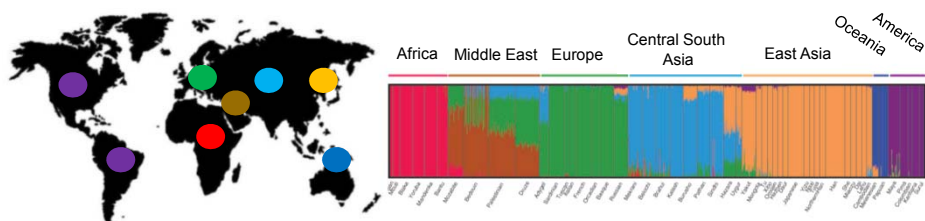
## Geographic structure of genetic diversity

General observation: continuous genetic variation by geography



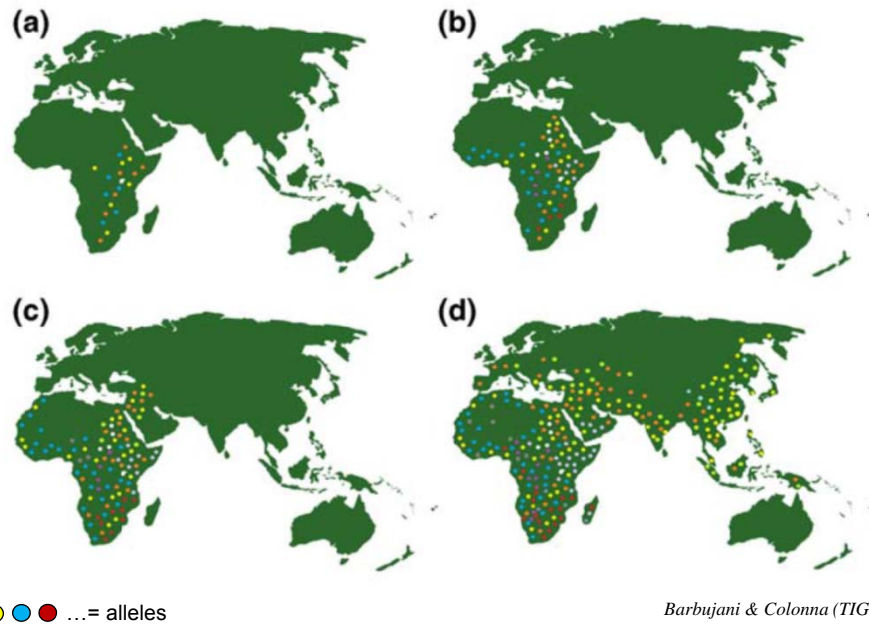
Manica et al. (Hum Genet 2005)

... but genetic discontinuities due to geographic and/or cultural barriers, etc....



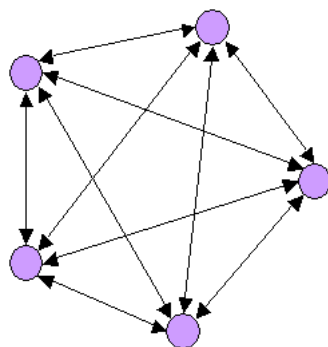
Li et al. (Science 2008)

## Past migrations and isolation by distance



## Models of gene flow between populations

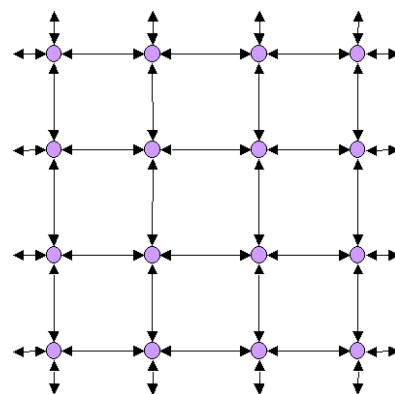
Wright's island model, 1943



→ FastSimCoal  
(this morning)

Kimura's stepping-stone models, 1964

2 dimensions



→ SPLATCHE  
(now)

## Spatially explicit computer simulation approach

### Spatially-explicit simulation program

SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity

2004 Version 1.0 - Currat, Ray, Excoffier (*Mol Ecol Res* 2004)

2010 Version 2.0 - Ray, Currat, Foll, Excoffier (*Bioinformatics* 2010)

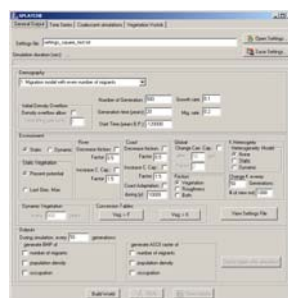
2019 Version 3.0 - Currat et al. Submitted



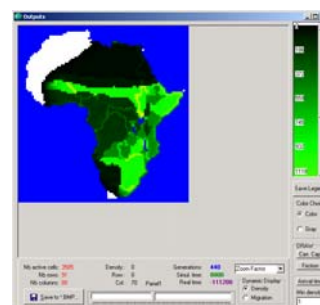
Nicolas Ray  
enviroSpace  
University of Geneva



Laurent Excoffier  
CMPG lab  
University of Berne



Freely available at  
[www.splatche.com](http://www.splatche.com)



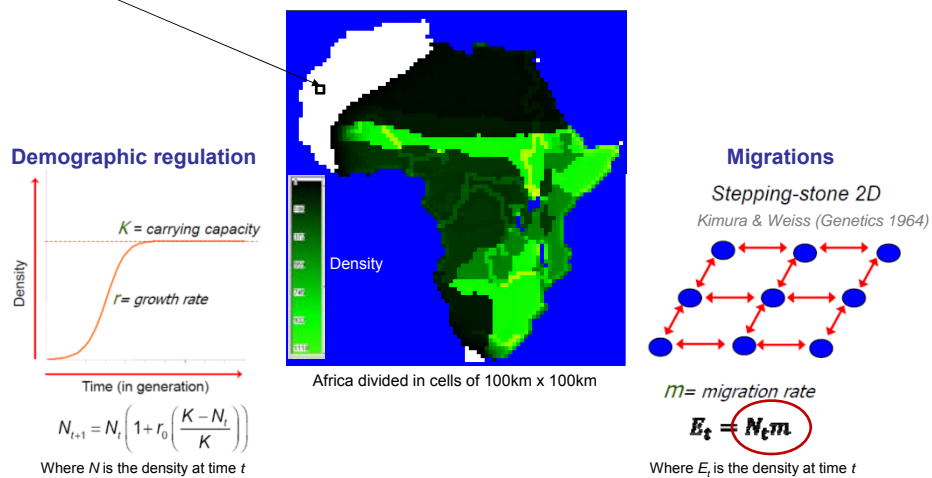
## SPLATCHE runs in two phases

1- Forward Demographic simulation

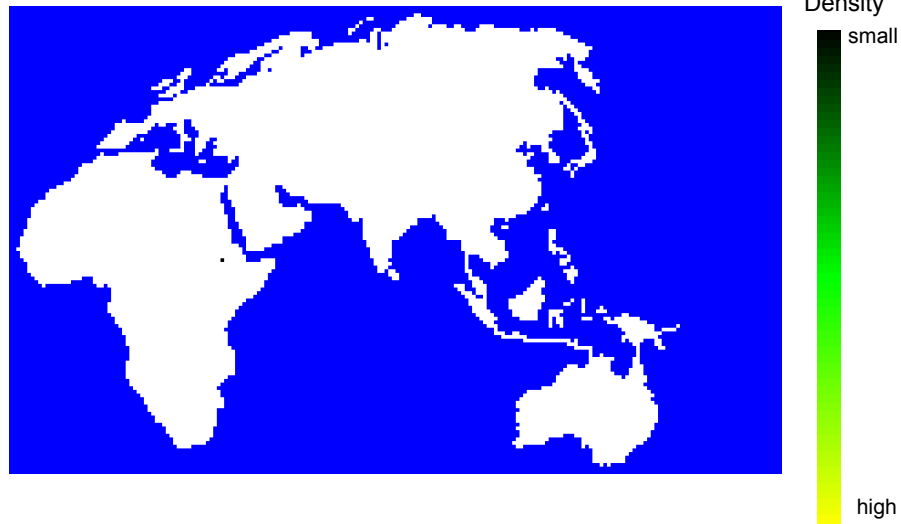
2- Backward Genetic simulation

## Demographic simulation

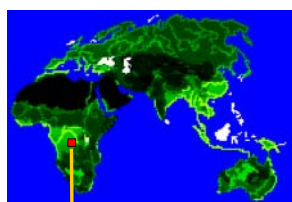
Simulation of populations during  $t$  generations in a virtual area made up of cells which represent sub-populations (demes)



## Example of demographic simulation

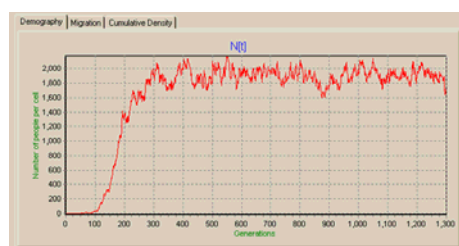


## Demographic database

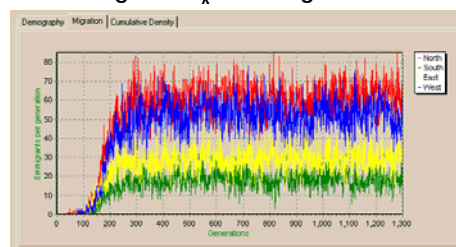


Demographic  
database

Density  $N$  at each generation



Immigrants  $I_k$  at each generation



---

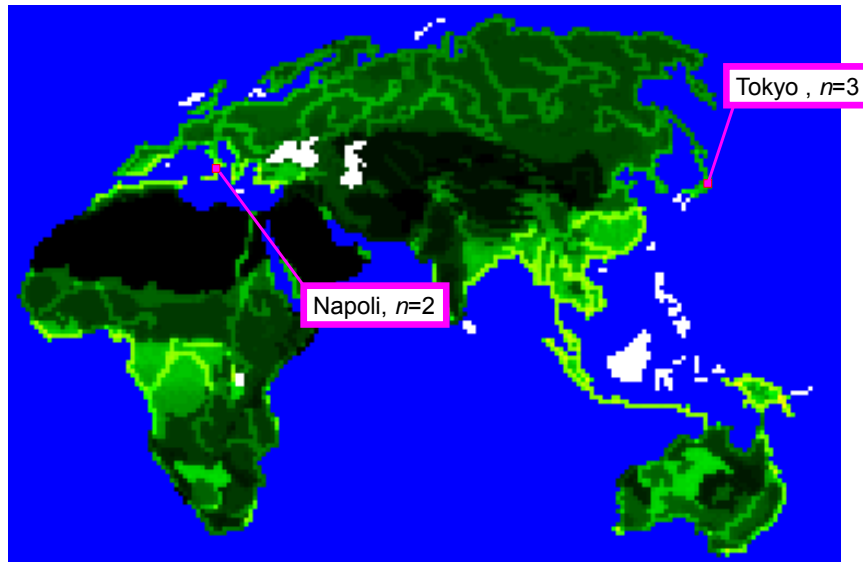
Coalescent in space and time

SPLATCHE runs in two phases

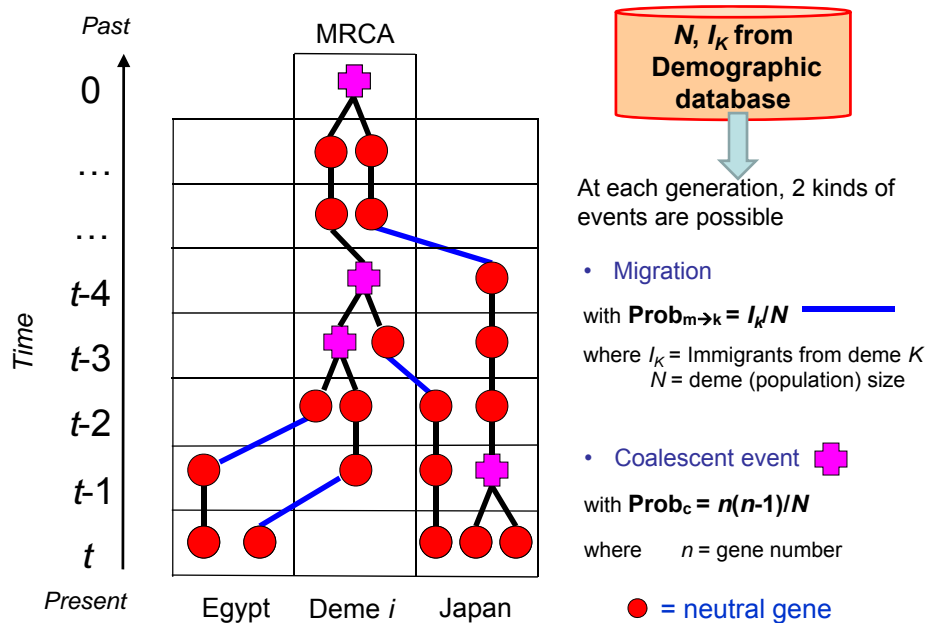
1- Forward Demographic simulation

2- Backward Genetic simulation

## Genetic samples

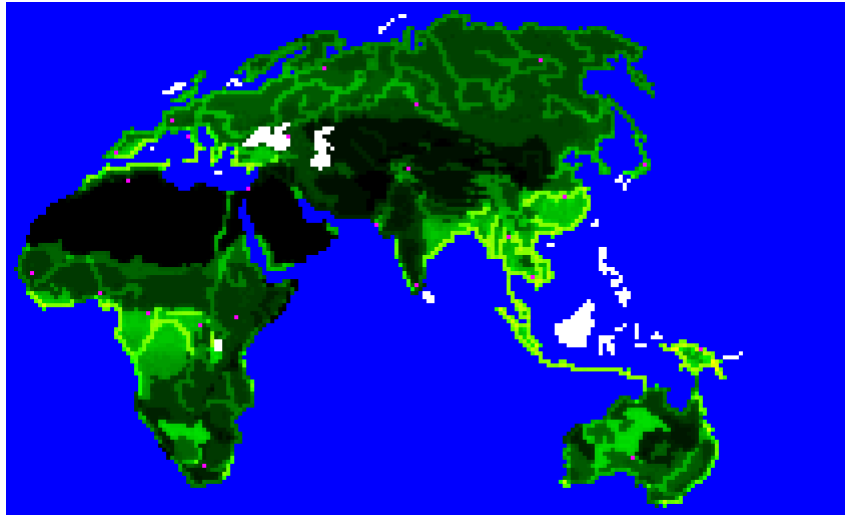


## Coalescent reconstruction



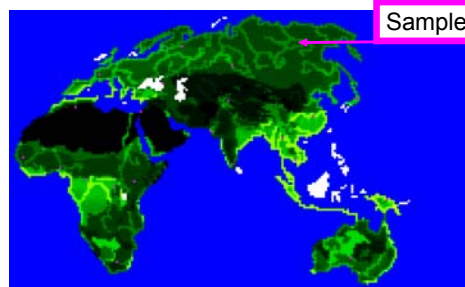


## Genetic simulation



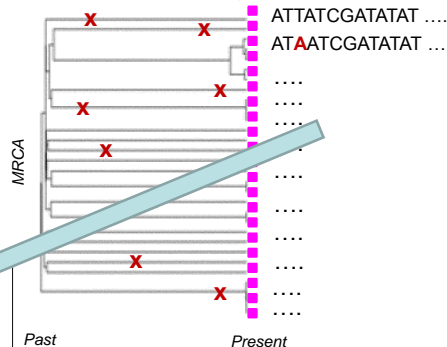
Backward in time

## Simulation of genetic diversity



Sampled genes

Genetic diversity  
 $\mu$  = mutation rate



[Profile]		Title="A series of simulated samples"	
		NbSamples=1	
		GenotypicData=1	
		DataType=DNA	
[Data]		[[Samples]]	
		SampleName="Sample 1"	
		SampleSize=25	
		SampleData= {	
1_1	1	TATTCTAATTCAGCTTCTGAACGTAAGG	
		TAGTAGTCTGCATAGCGGCGTTGTGCGA	
1_2	1	TAGTCGTCTGCGTATTGGGGTTGTGCAG	
		TAGTCGTCTGCGTATTGGGGTTGTGCAG	
1_3	1	TATGCTAATTCAGCTTCTGATCGTAAGG	
		TAGTCGTCTGCATAGTGGCGTTGTGCGA	
1_4	1	AATGCTAATTCAGCTTCTGATCGTAAGG	

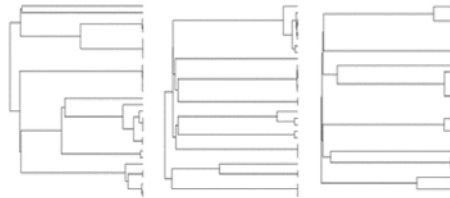


computation of summary  
statistics using ARLEQUIN

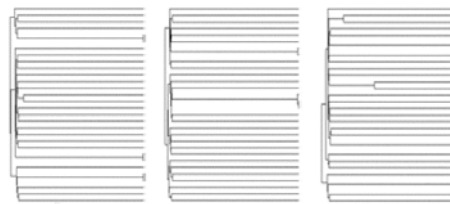
Excoffier & Lischer (Mol Ecol Res 2010)

## A stochastic process

**Small size**  
Expanding  
population

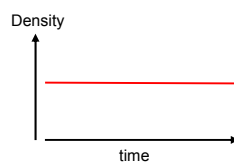


**Large size**  
Expanding  
Population



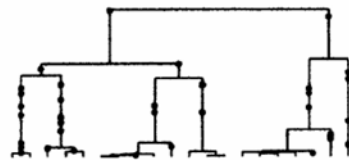
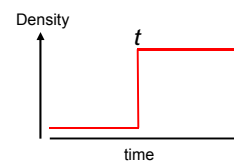
## Translation of demography to genetics

Population with constant size

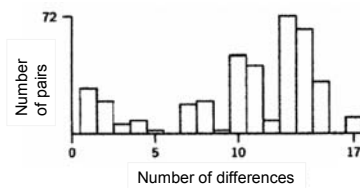
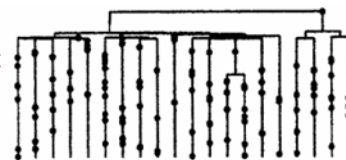


Demographic  
scenarios

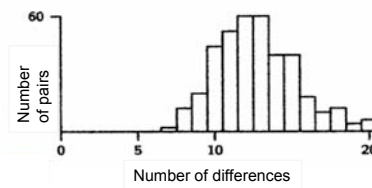
Population after a demographic increase



Coalescent  
trees

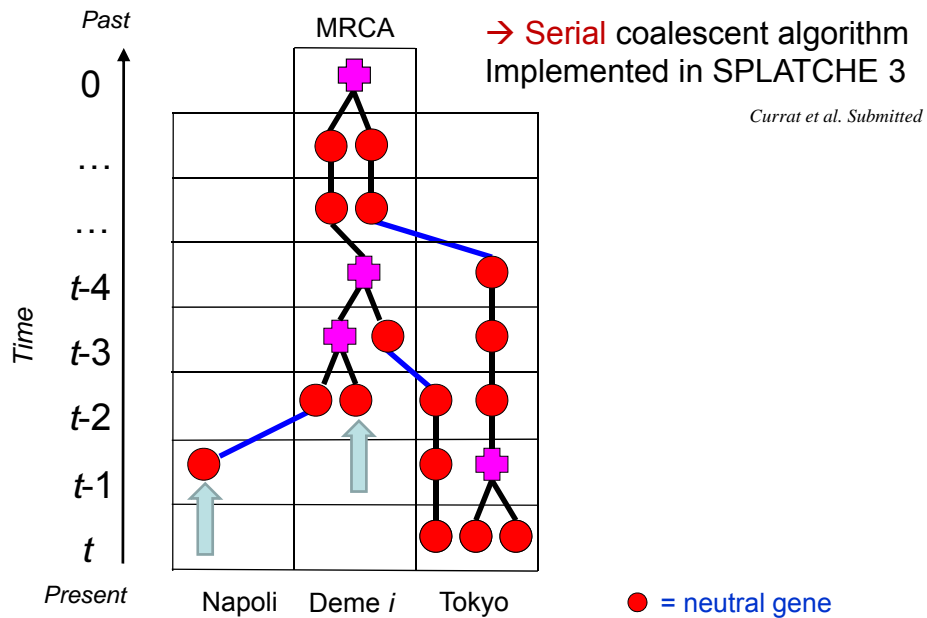


Summary  
statistics



*Modified from Harpending et al. (PNAS 1998)*

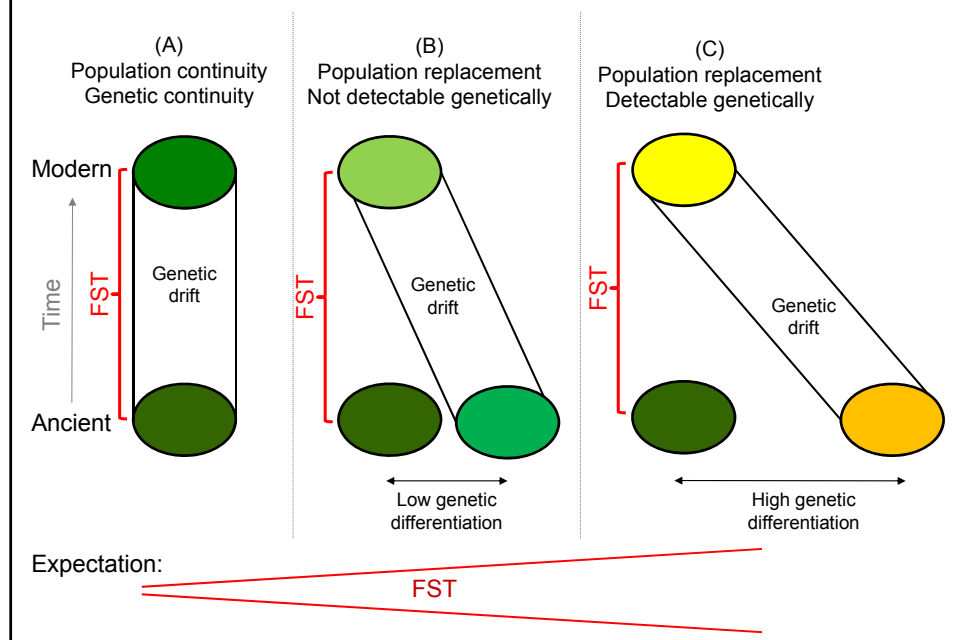
## Serial coalescent reconstruction



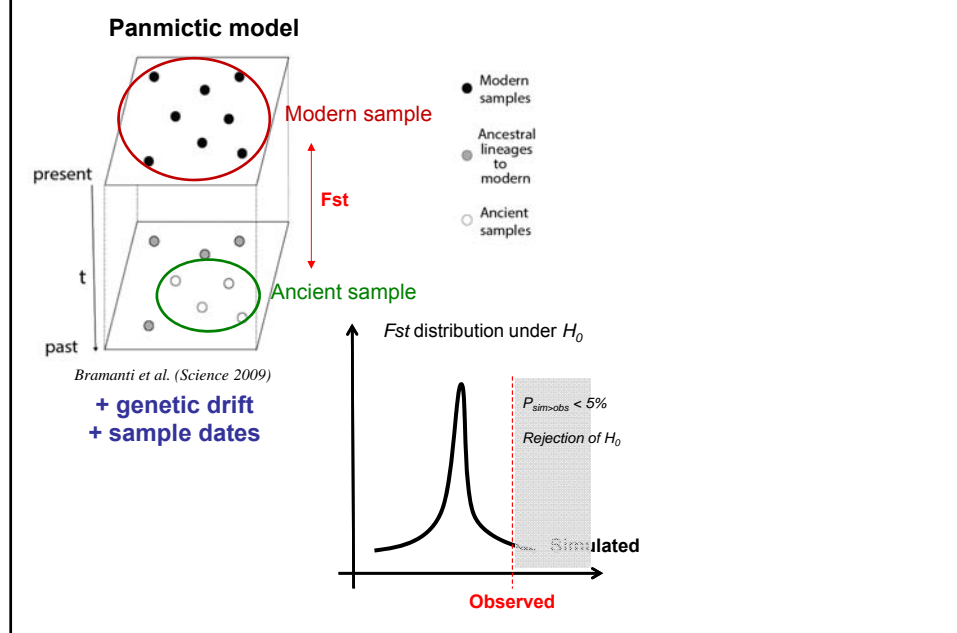
Jérémy Rio and Nuno Silva

## Population continuity test

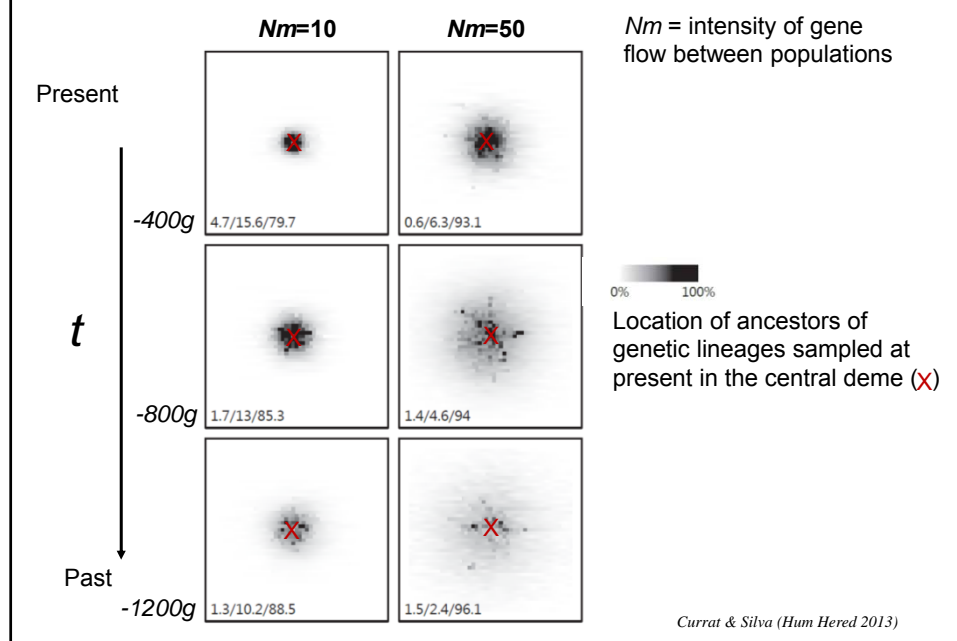
## Definition of population continuity



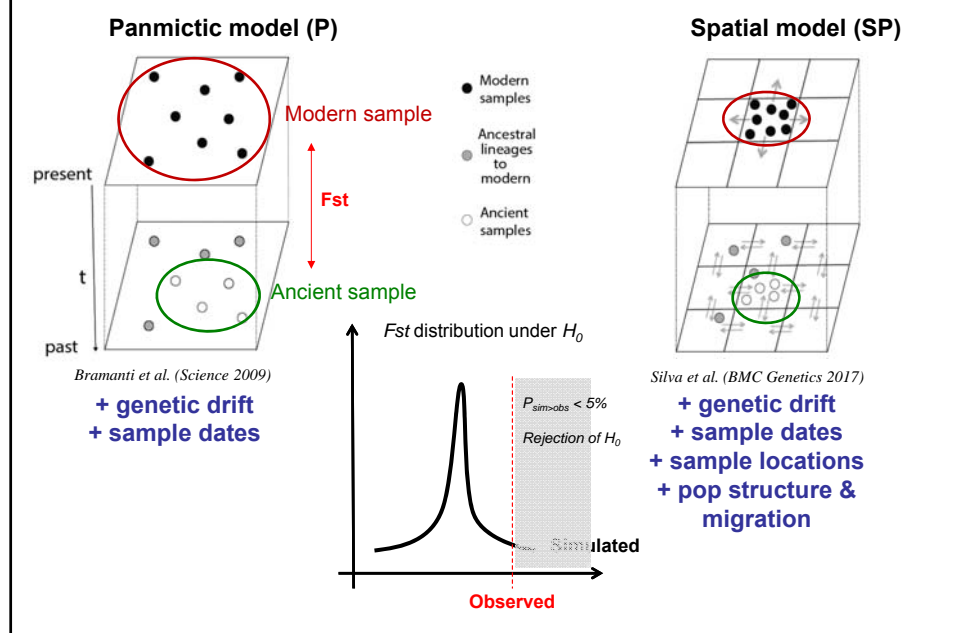
## $H_0$ = Continuous population



## Following ancestors of modern lineages



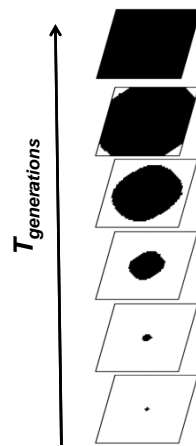
## $H_0$ = Continuous population



## Spatial test of population continuity

### Population expansion

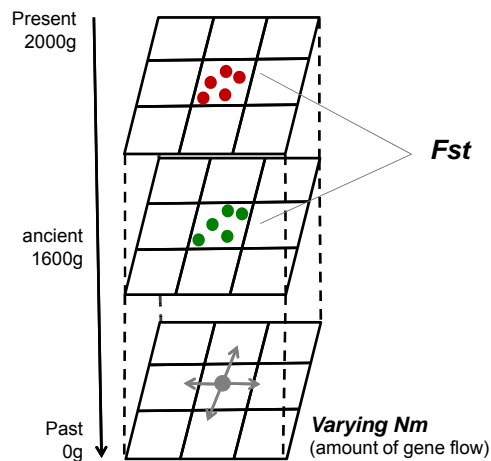
Square World, 50x50 demes



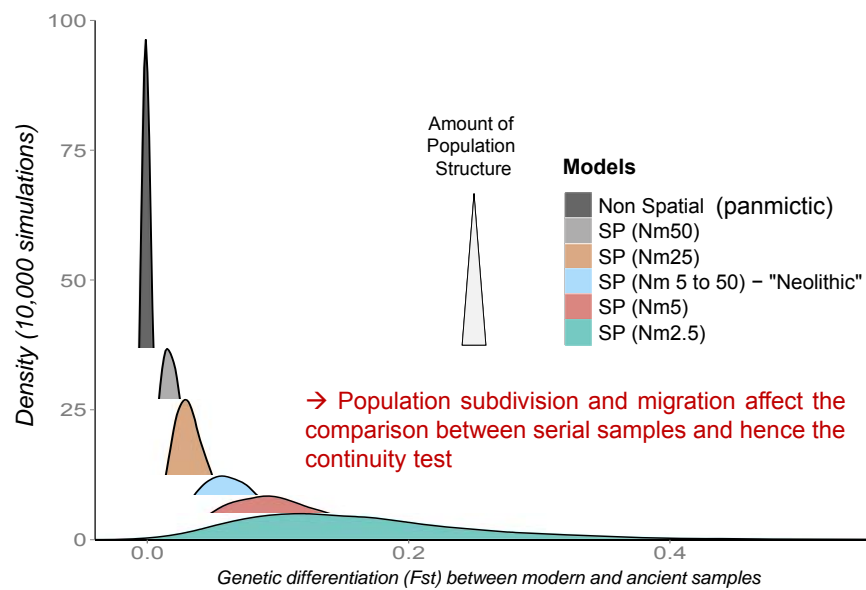
### Genetic sampling

DNA 400bp – 30 modern

30 ancient



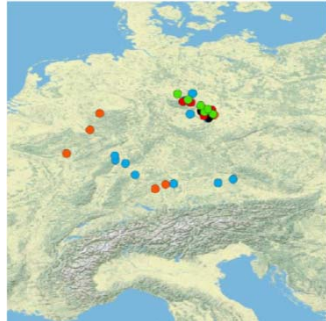
## FST distributions for various continuity models



Silva et al. (BMC Genetics 2017)

## Application to central Europe mtDNA

Mitochondrial sequences of 348bp

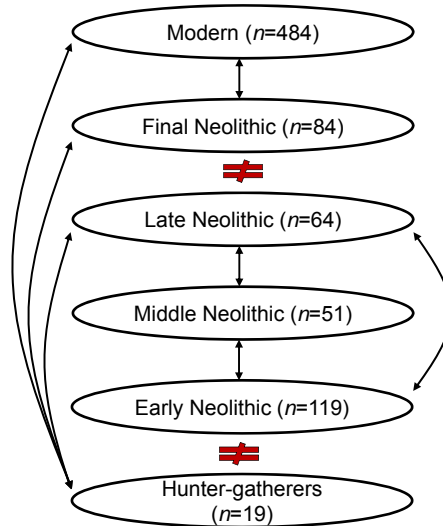


Hunter-Gatherers (orange) Early Neolithic (blue), Middle Neolithic (green), Late Neolithic (red), and Final Neolithic (black) lineages

Bramanti et al 2009, Fu et al 2013, Bollongino et al 2013, Lazaridis et al 2013, Haak et al 2005, Balanovsky et al 2010, Brotherton et al 2013, Brandt et al 2013, S. Kreuzer personal communication



Simulated map



Susanne Kreuzer

Silva, Kreuzer et al. (in prep)

## Application to genome wide data

Data

~200,000 SNP



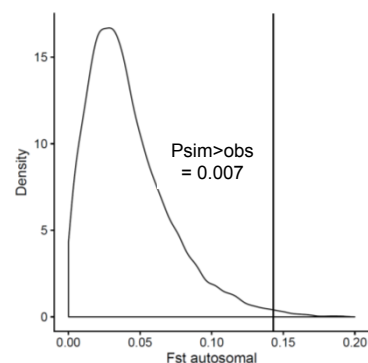
Loschbour  
Luxembourg  
6147-6047 cal BCE  
Mesolithic

Stuttgart  
Germany  
5500-4800 cal BCE  
LBK

Lazaridis et al. (Nature 2014)

Simulation

10,000 SNP



Distribution of simulated Fst under the hypothesis of population continuity. The bar shows the observed value.

Silva, Rio et al (Evol App 2018)

→ Rejection of full continuity during the Neolithic transition in central Europe using two ancient genomes

## Limitations

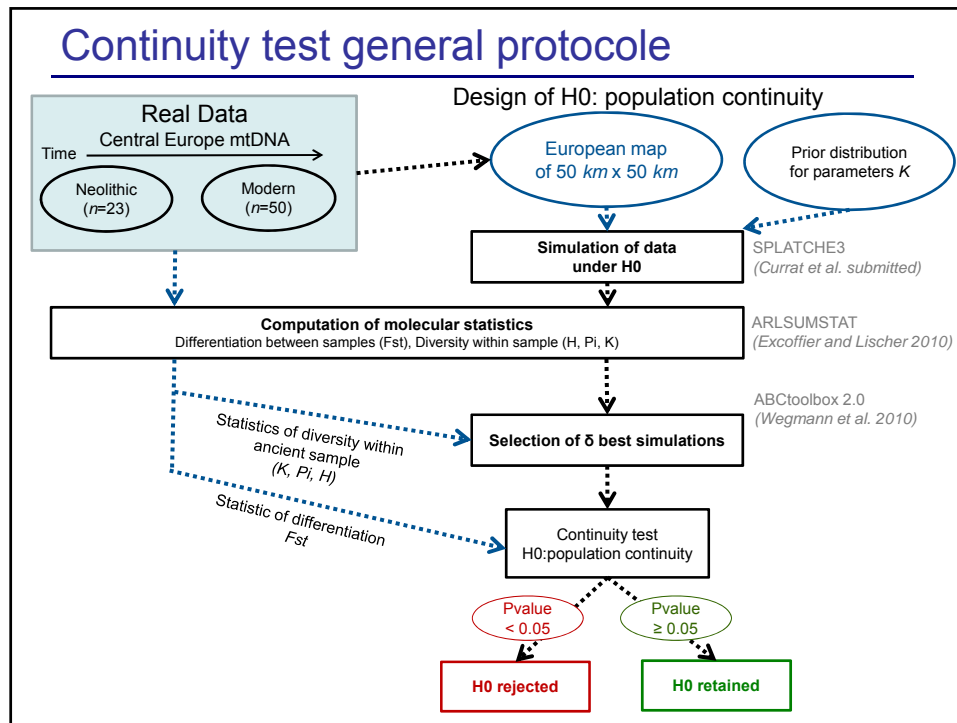
---

- Comparison by pairs of sampling times
- Need of enough overlapping positions (sufficient genomic coverage)
- Estimation of  $Nm$  is important for the accuracy of the continuity test
- Simulation of numerous loci requires an extensive computational time

---

## Practicals





## Practicals

**STEP 1:** GRAPHICAL SIMULATION OF POPULATION CONTINUITY (SPLATCHE3 GUI)

**STEP 2:** GRAPHICAL SIMULATION OF GENETIC DIVERSITY (SPLATCHE3 GUI)

**STEP 3:** SIMULATION WITH THE CONSOLE VERSION (SPLATCHE3)

**STEP 4:** MERGE ANCIENT SAMPLES

**STEP 5:** DRAW NEOLITHIC POPULATION SIZE FROM PRIOR DISTRIBUTION (ABCtoolbox)

**STEP 6:** CHOICE OF THE BEST SET OF SIMULATIONS WITH ABC (ABCtoolbox)

**STEP 7:** TEST OF POPULATION CONTINUITY