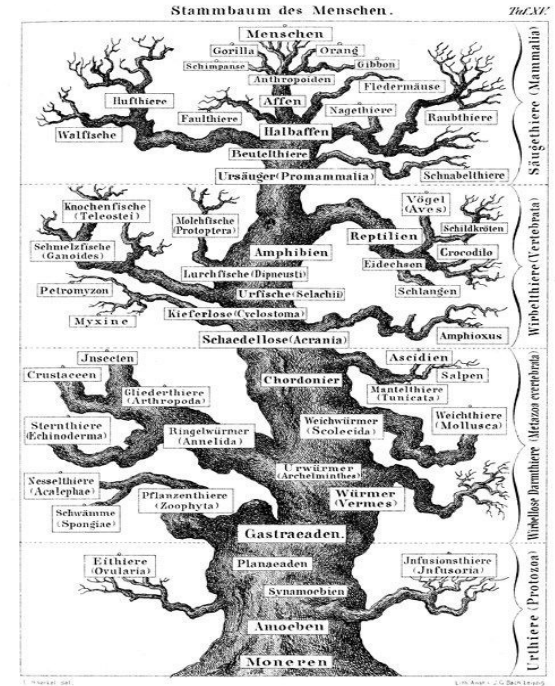
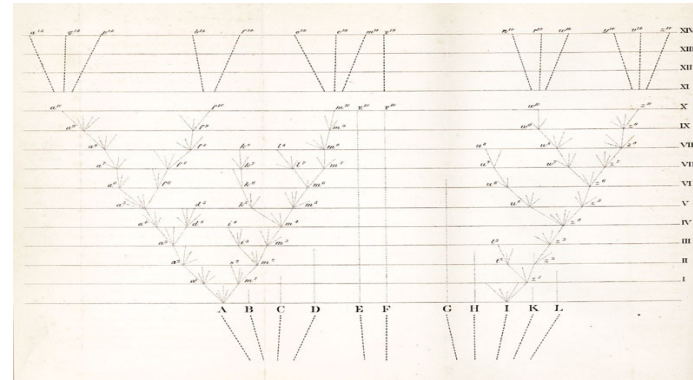
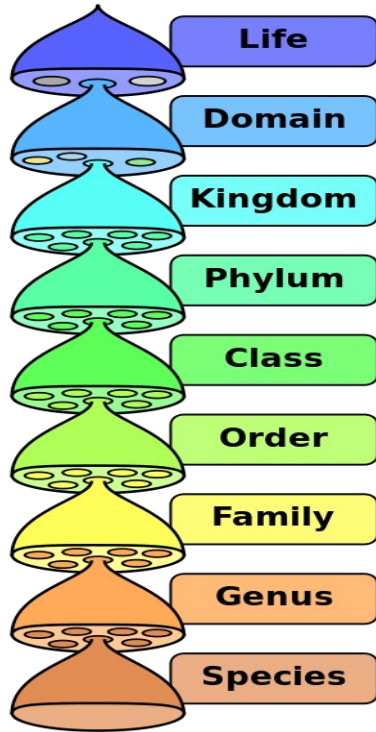


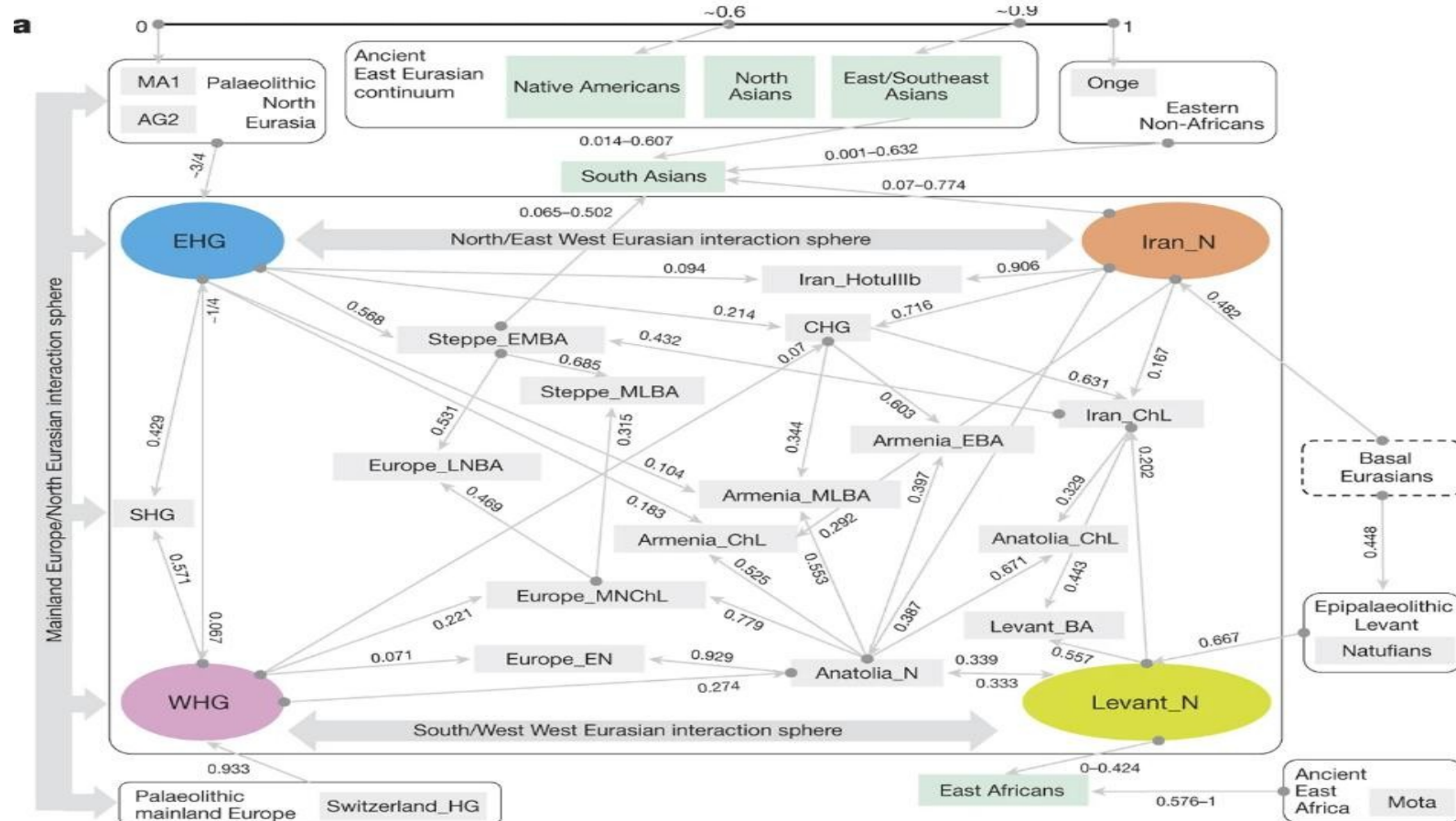
# F-statistics and Population Structure

Benjamin Peter, MPI for Evolutionary Anthropology

# Trees are foundational to Systematics and Evolution



# Or really complicated

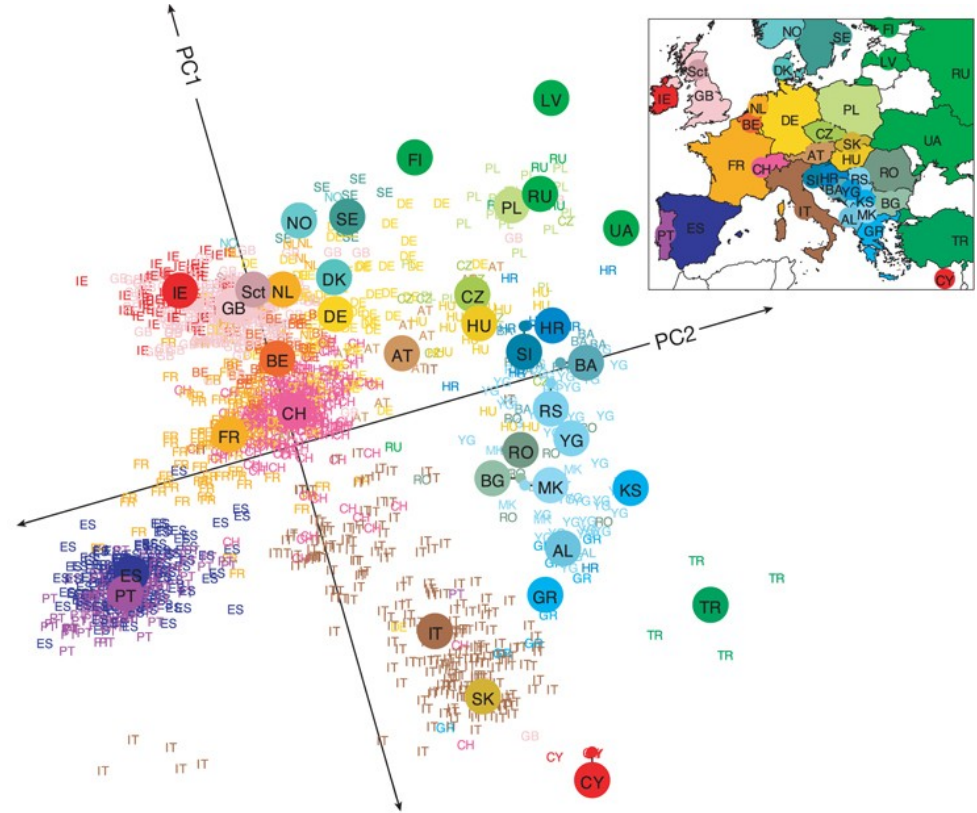


Vol 456 | 6 November 2008 | doi:10.1038/nature07331

# LETTERS

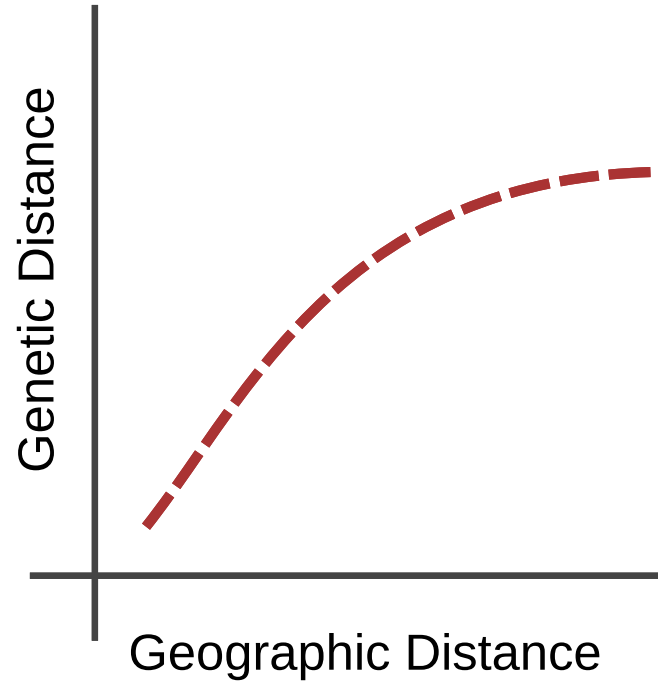
## Genes mirror geography within Europe

John Novembre<sup>1,2</sup>, Toby Johnson<sup>4,5,6</sup>, Katarzyna Bryc<sup>7</sup>, Zoltán Kutalik<sup>4,6</sup>, Adam R. Boyko<sup>7</sup>, Adam Auton<sup>7</sup>, Amit Indap<sup>7</sup>, Karen S. King<sup>8</sup>, Sven Bergmann<sup>4,6</sup>, Matthew R. Nelson<sup>8</sup>, Matthew Stephens<sup>2,3</sup> & Carlos D. Bustamante<sup>7</sup>

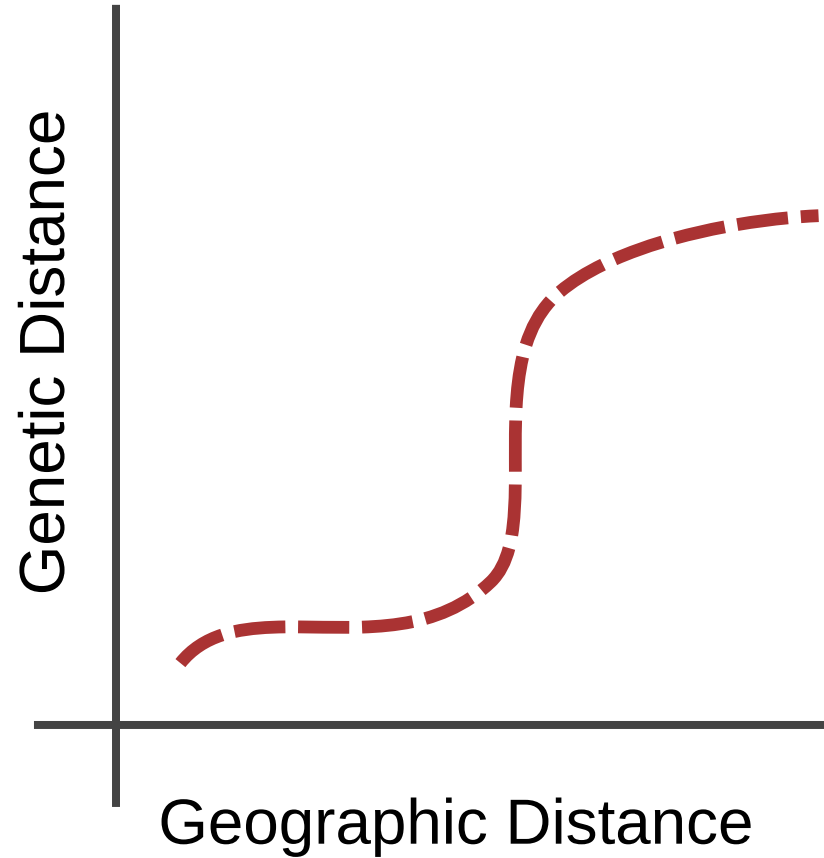


Novembre et al. 2008, Nature

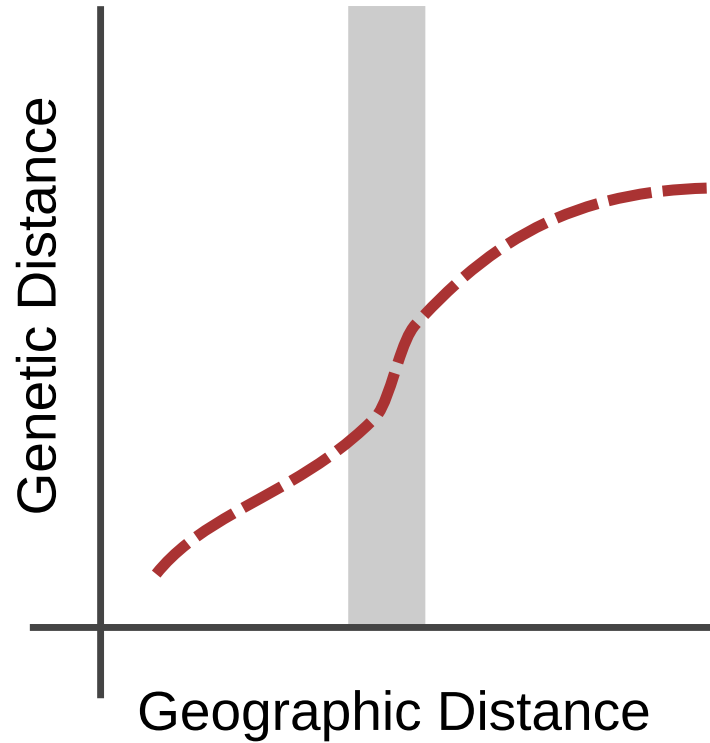
# Continuous structure



# Discrete structure



# Inferring discontinuities



# Where do individuals move?

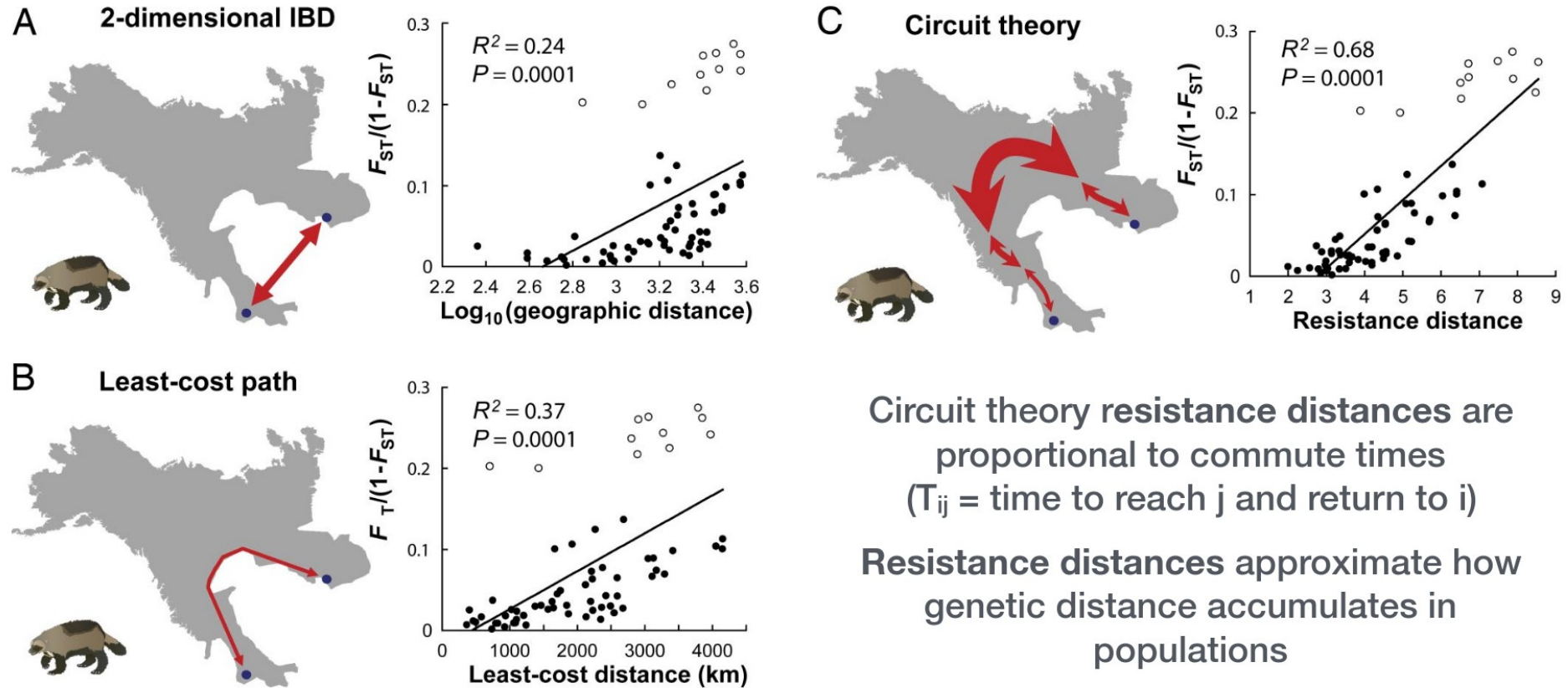
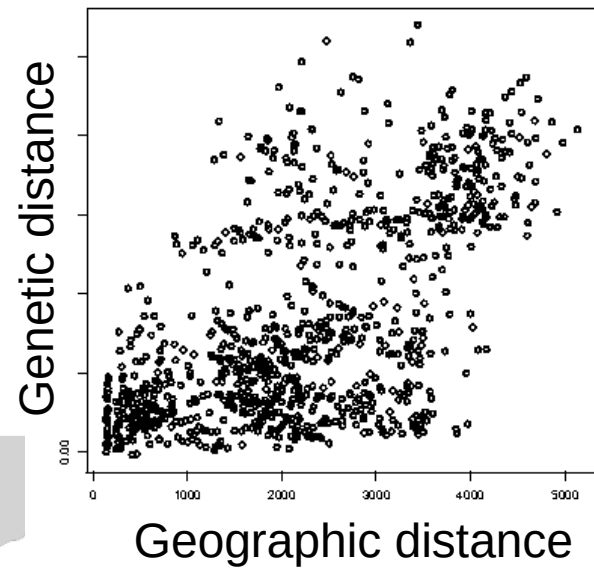
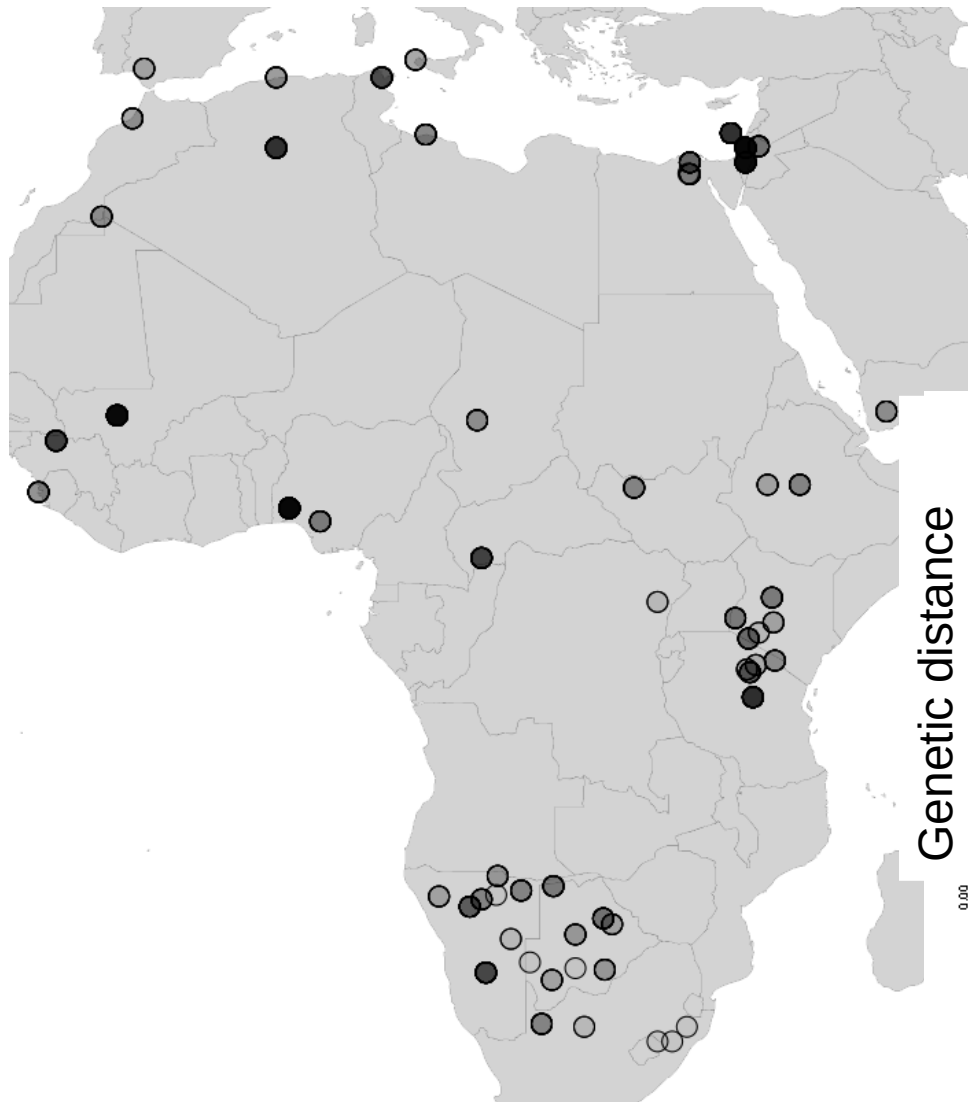
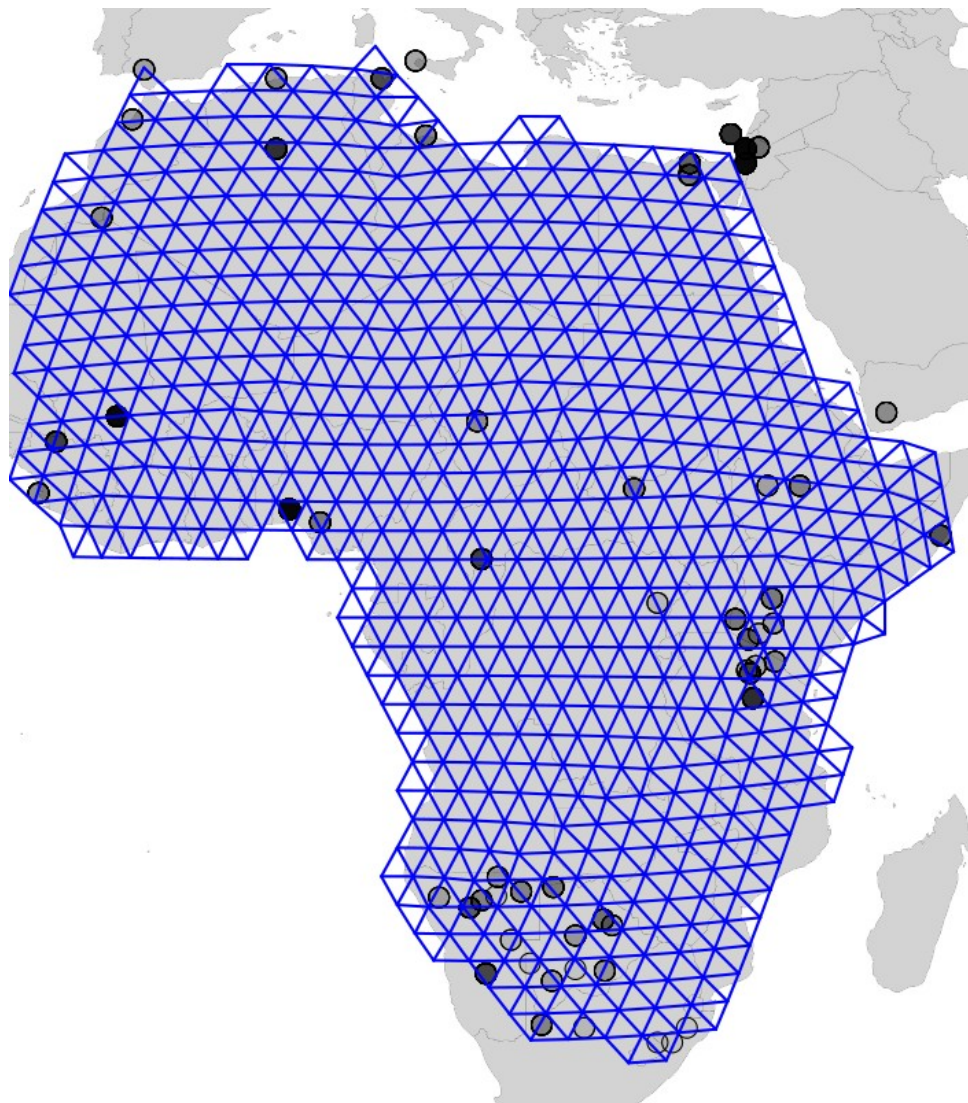


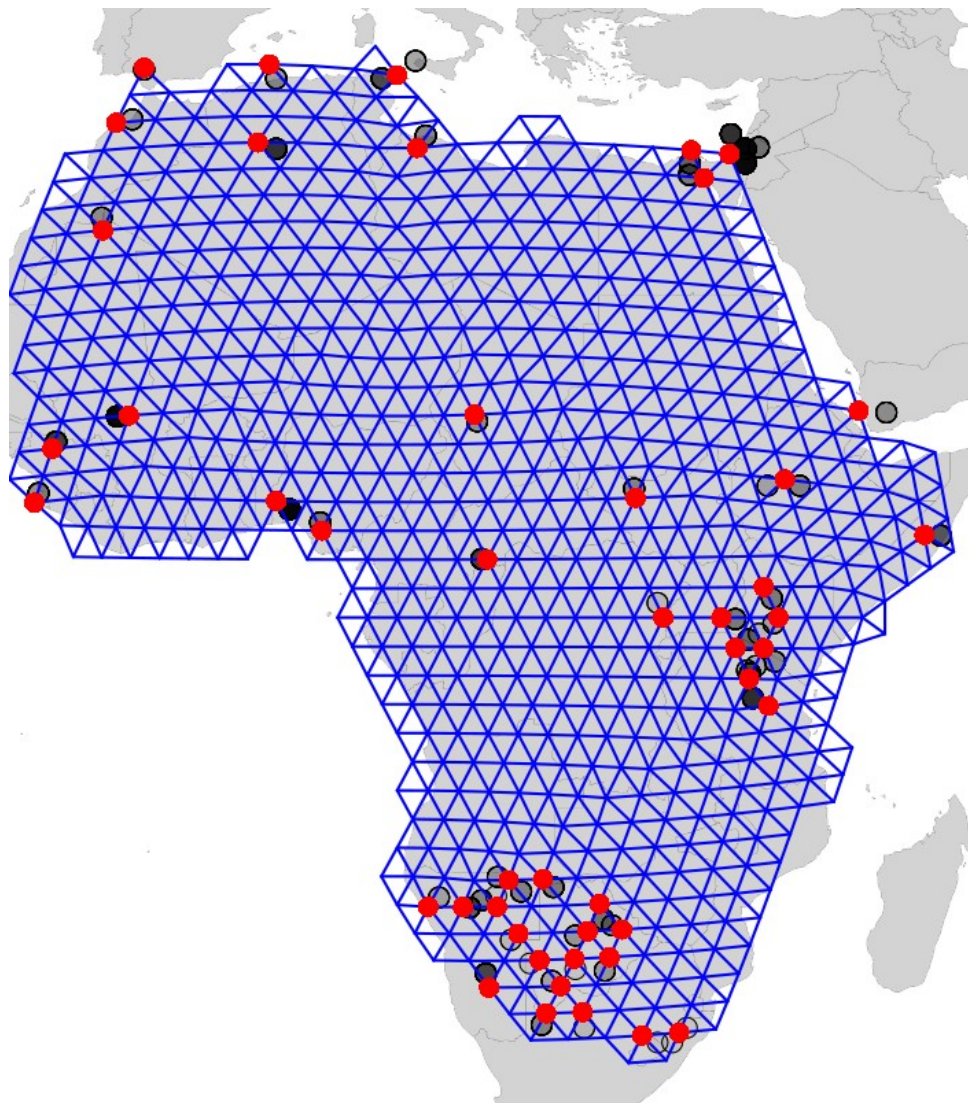
Figure: McRae & Beier 2007



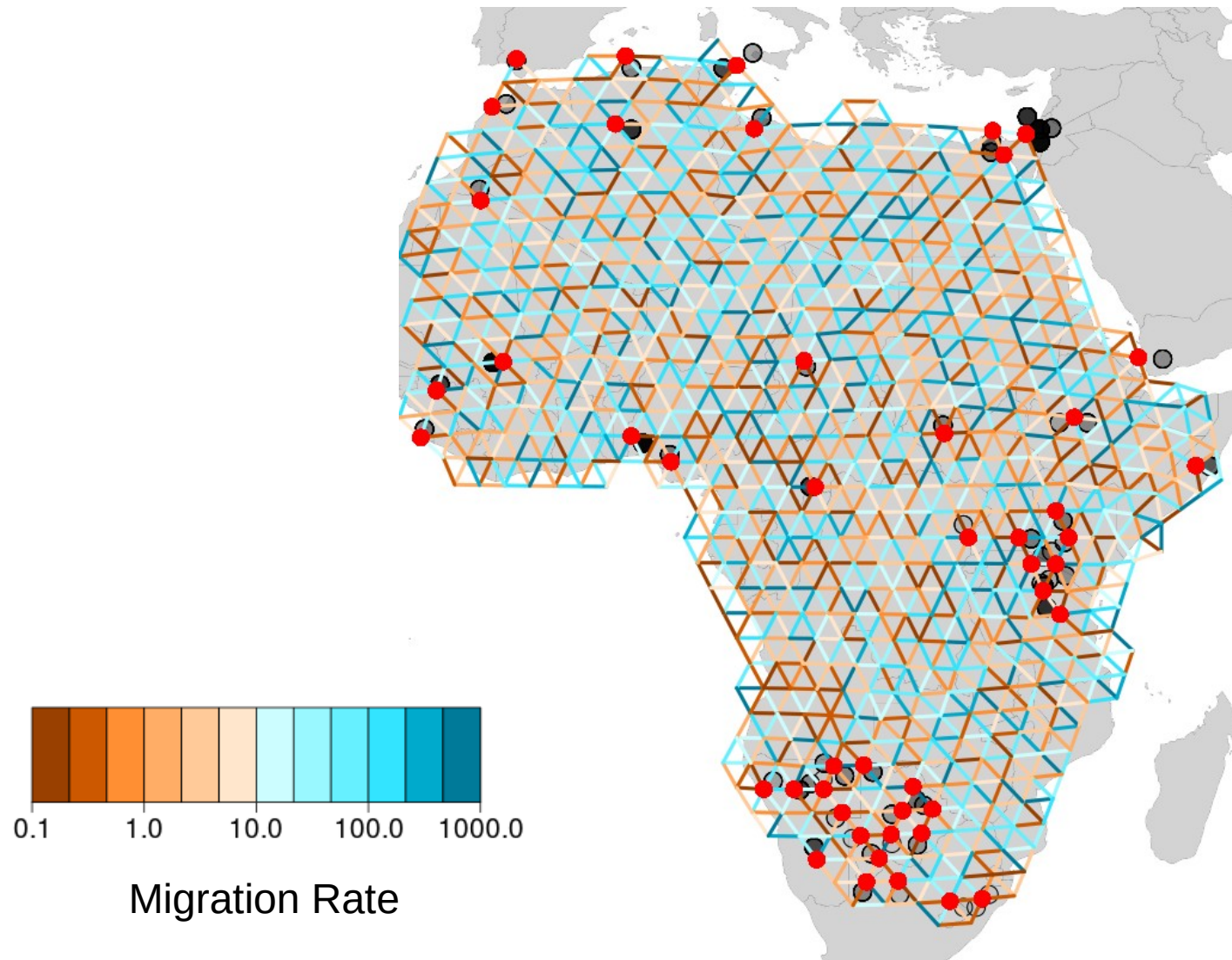


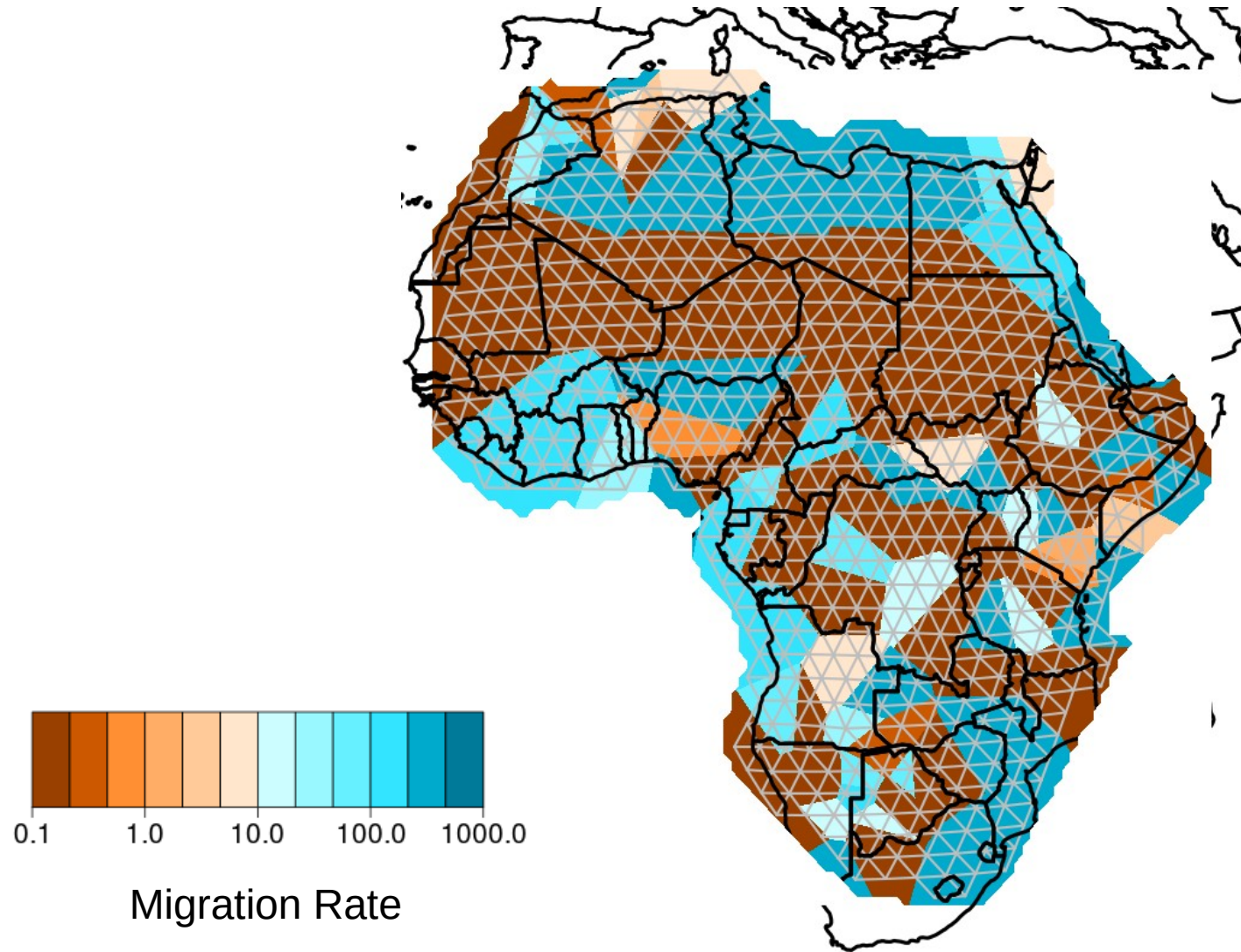


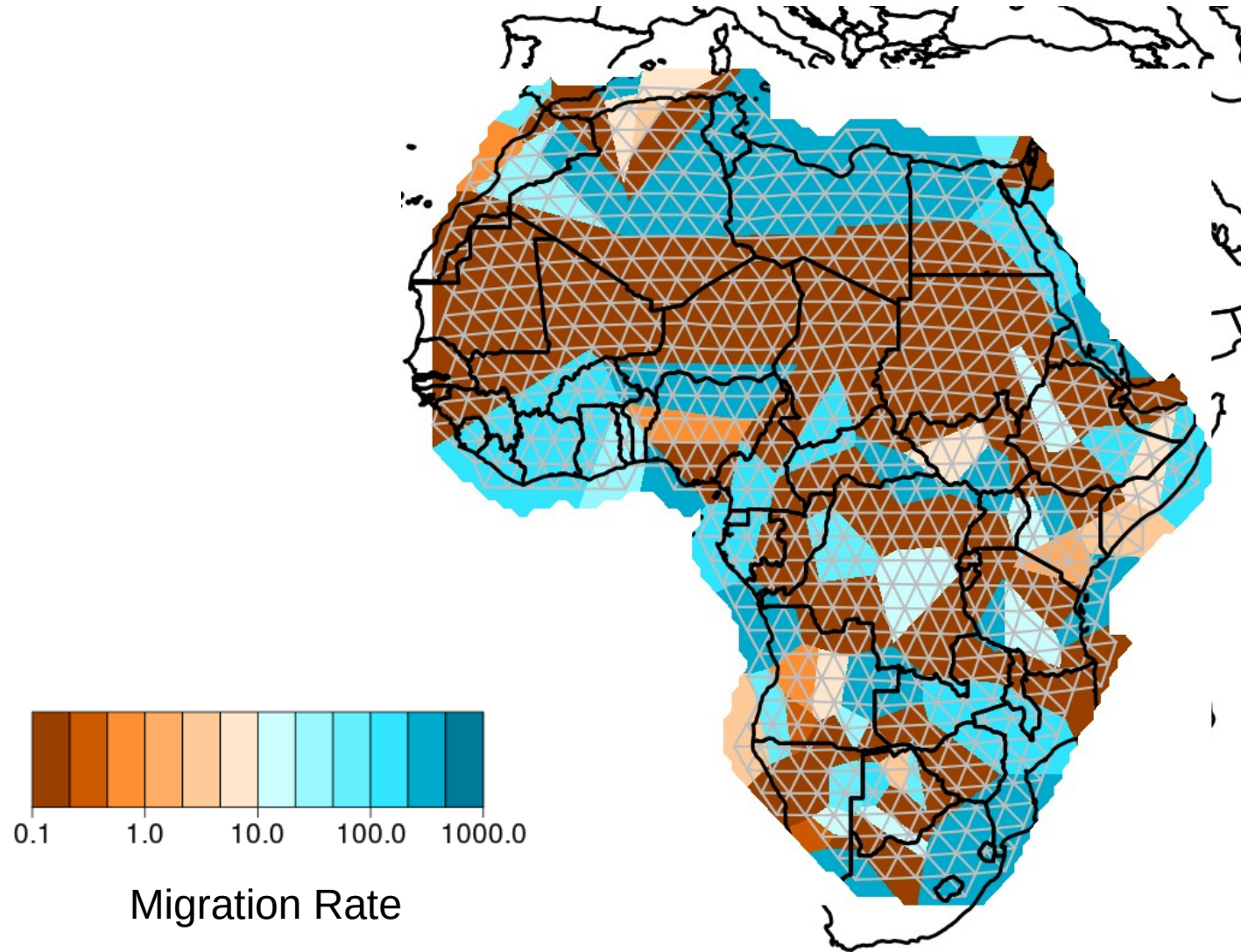




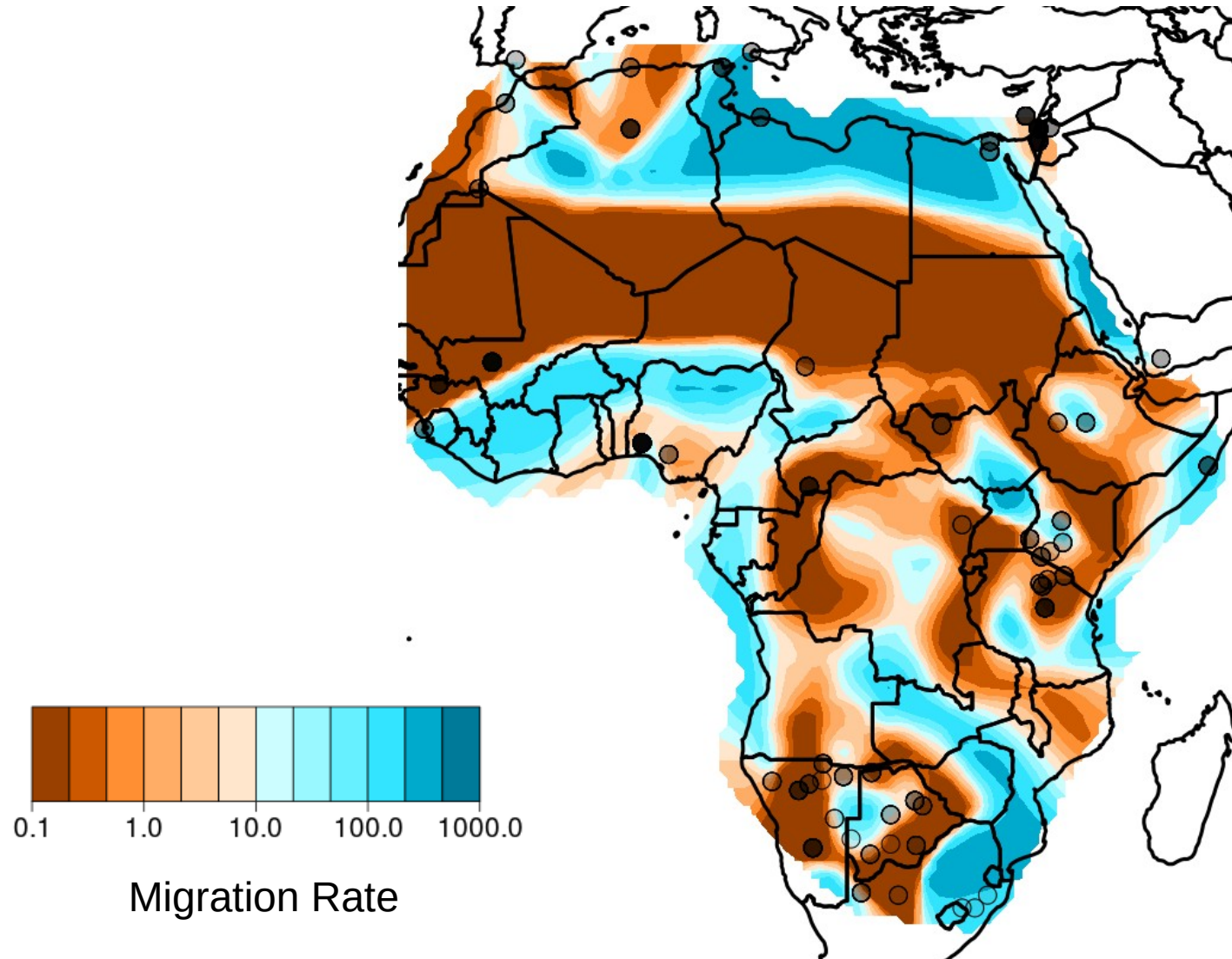




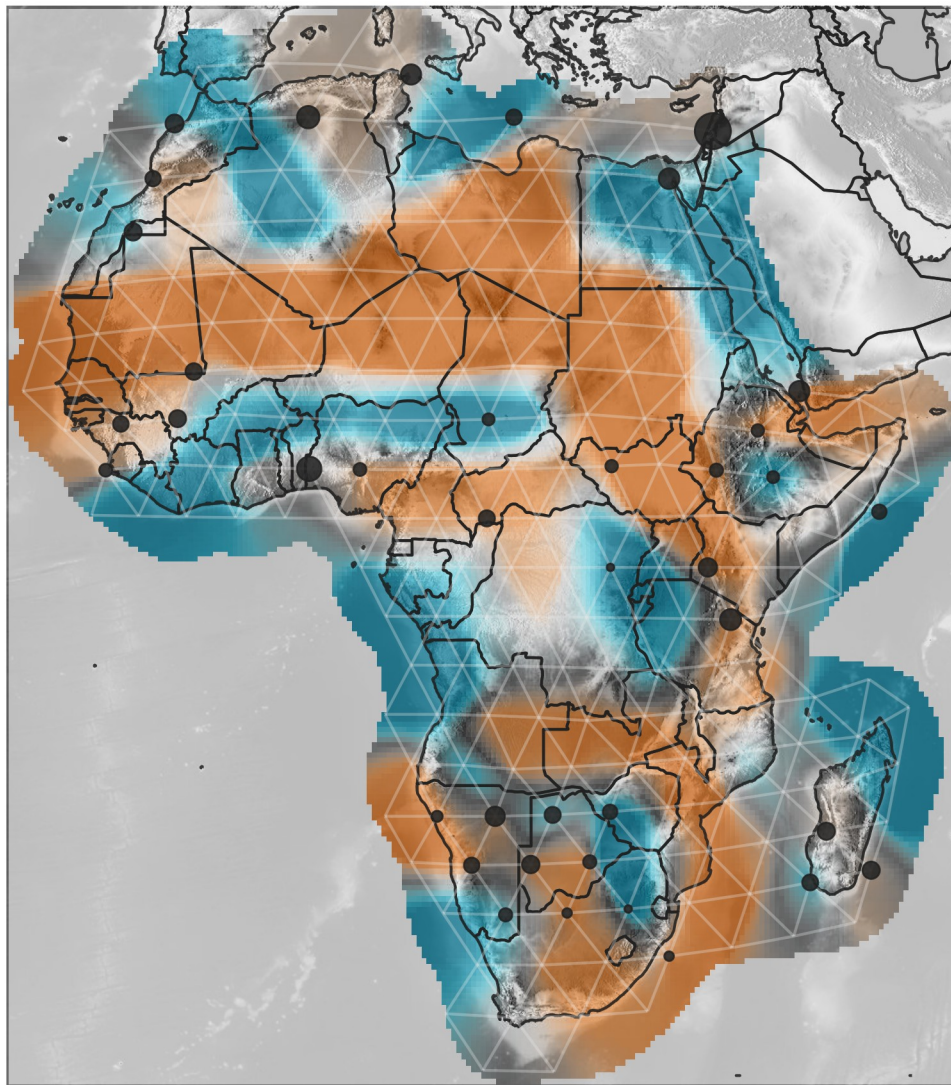




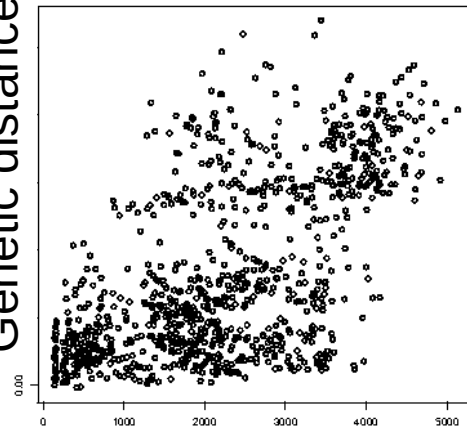






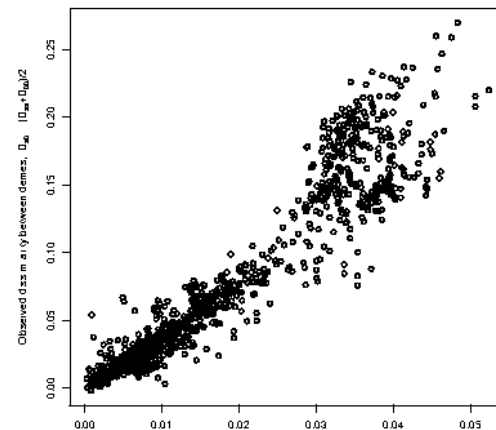


Genetic distance

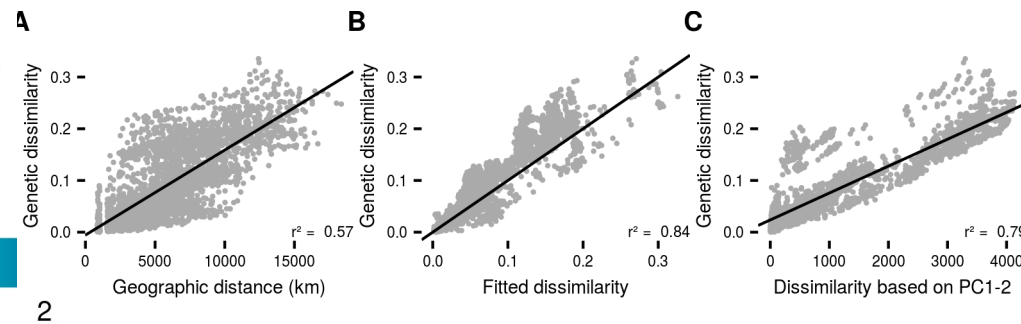
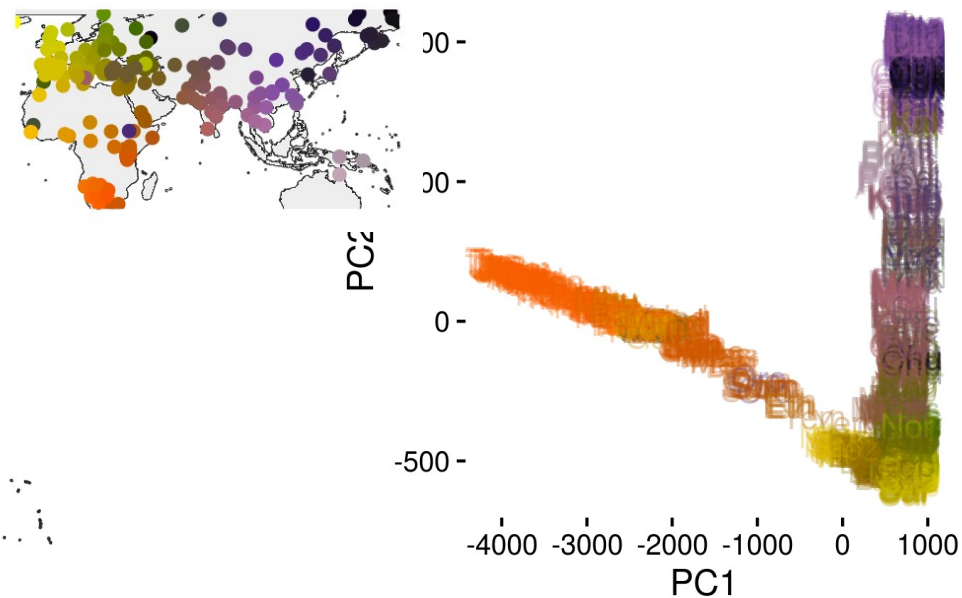
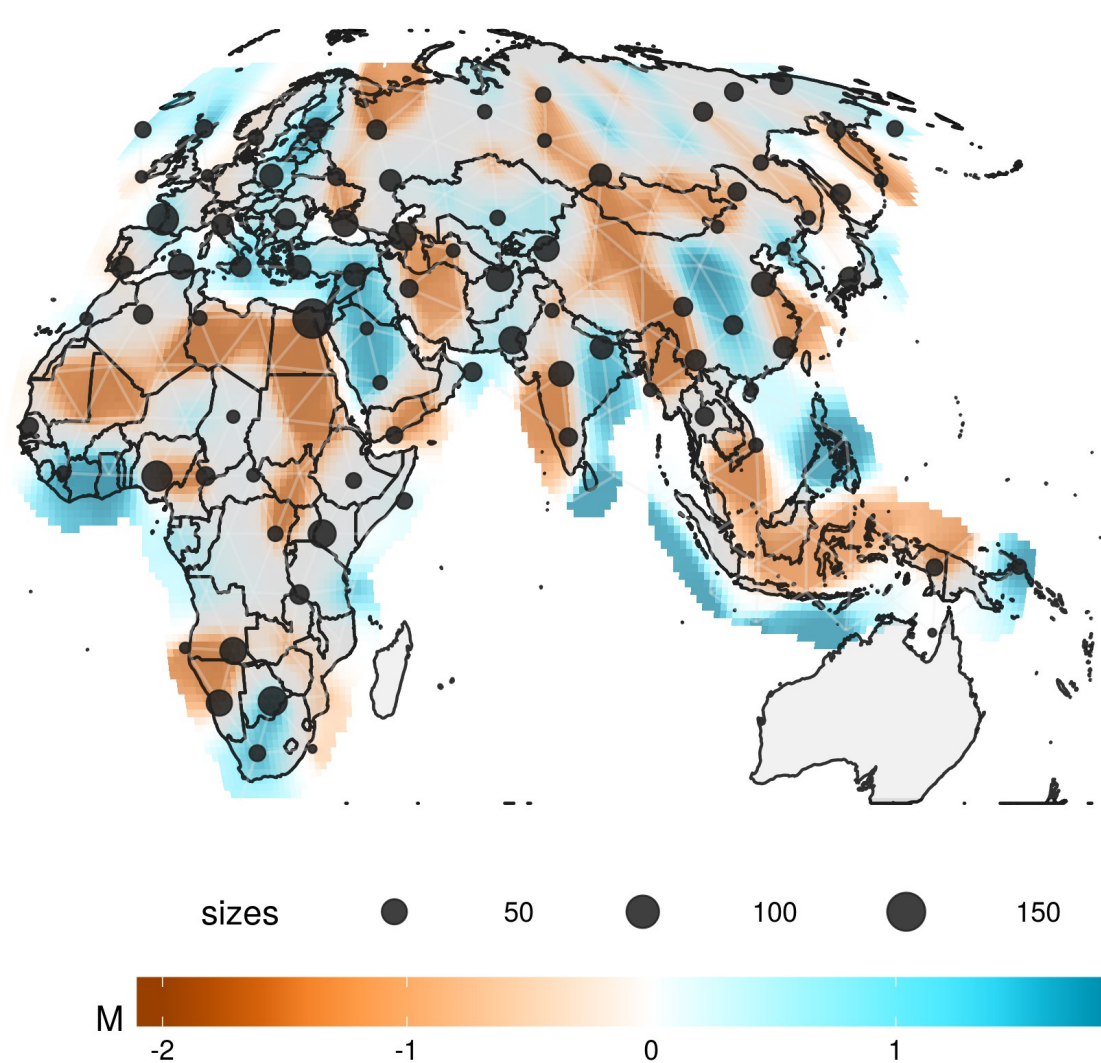


Geographic distance

Gray means a or b has a single individual sampled from



Fitted circuit distance



# Models are starting points!

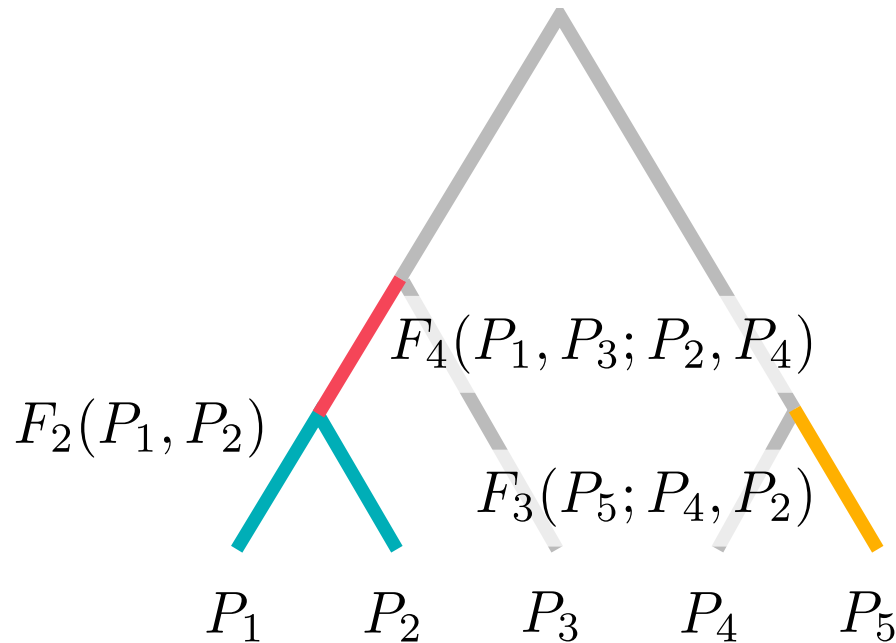
## **Discrete pops**

- Trees
- Clustering
- Admixture graphs

## **Continuous pops**

- PCA
- Spatial methods

# Recap: F-stats as branches on tree



- $F_2$  is pairwise distance
- $F_3$  is external branch
- $F_4$  is internal branch

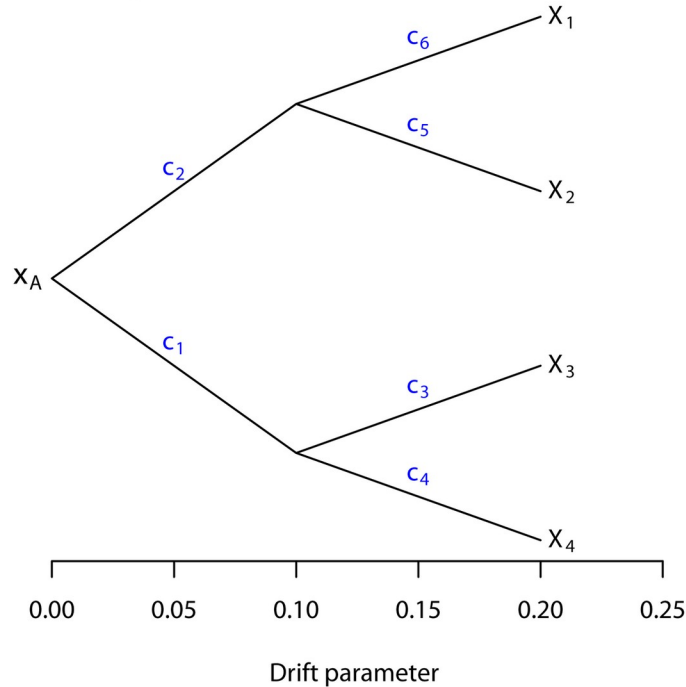
Simple hypothesis tests  
involving 2,3 or four pops

# Estimating Trees from F-stats

- Calculate all F2-statistics
- Treat them as distances
- Estimate tree (fitting covariance, or standard algorithms such as neighbor-joining)

# Covariance of a tree

A. Example tree

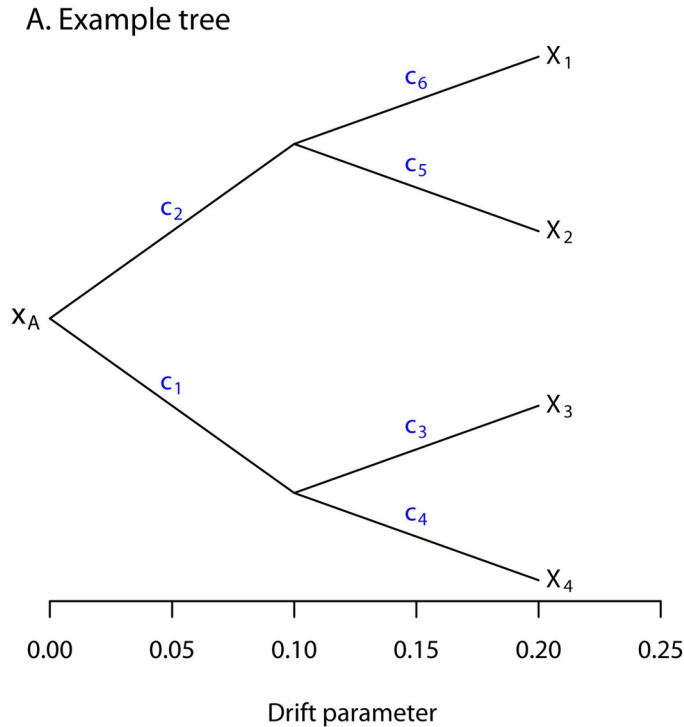


B. Covariance matrix for tree in A.

$X_1$	$c_2 + c_6$	$c_2$	0	0
$X_2$	$c_2$	$c_2 + c_5$	0	0
$X_3$	0	0	$c_1 + c_3$	$c_1$
$X_4$	0	0	$c_1$	$c_1 + c_4$
	$X_1$	$X_2$	$X_3$	$X_4$

$$F2(X, Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$$

# Distance matrix of a tree

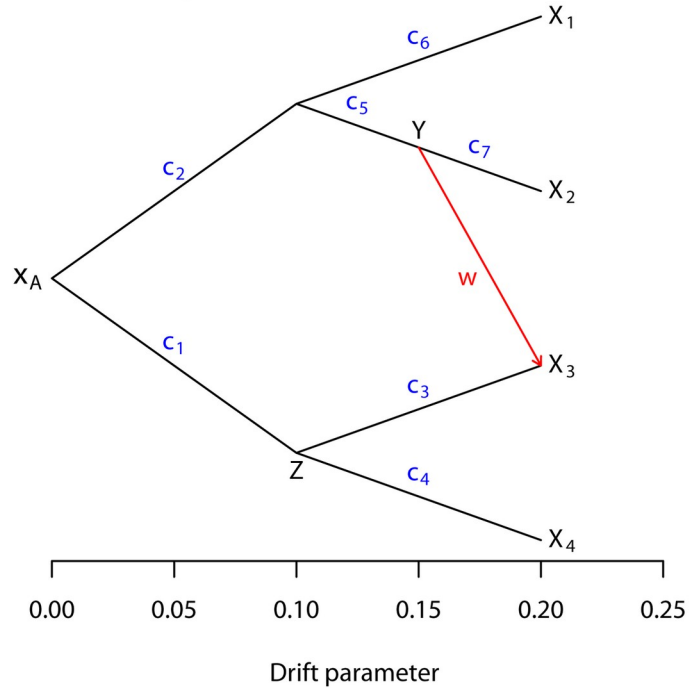


	$X_1$	$X_2$	$X_3$	$X_4$
$X_1$				
$X_2$				
$X_3$				
$X_4$				

$$F2(X, Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$$

# Covariance of Admixture Graph

C. Example graph



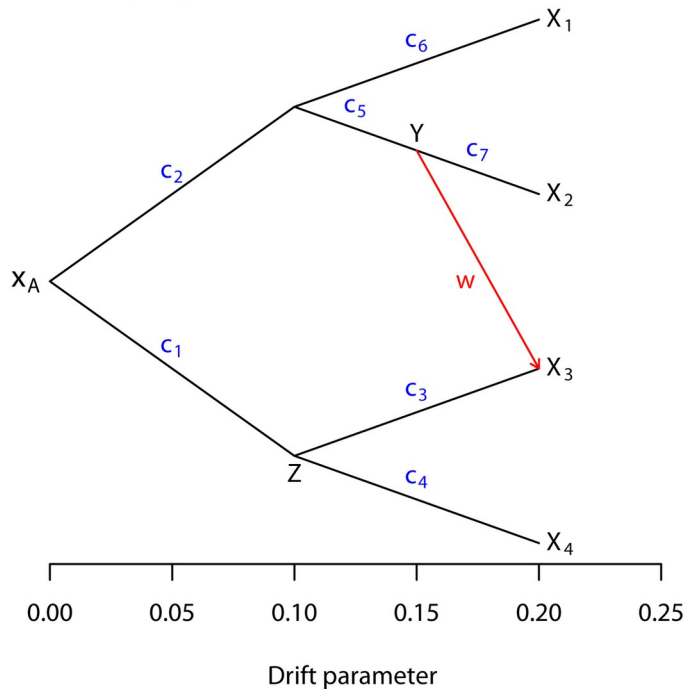
D. Covariance matrix for graph in C.

	$X_1$	$X_2$	$X_3$	$X_4$
$X_1$	$c_2 + c_6$	$c_2$	$w c_2$	0
$X_2$	$c_2$	$c_2 + c_5 + c_7$	$w(c_2 + c_5)$	0
$X_3$	$w c_2$	$w(c_2 + c_5)$	$w^2(c_2 + c_5) + (1-w)^2(c_1 + c_3)$	$(1-w)c_1$
$X_4$	0	0	$(1-w)c_1$	$c_1 + c_4$
	$X_1$	$X_2$	$X_3$	$X_4$



# Distance matrix of Admixture Graph

C. Example graph



	$X_1$	$X_2$	$X_3$	$X_4$
$X_1$				
$X_2$				
$X_3$				
$X_4$				

# Applications

- treemix
- qpgraph (admixtools)



BROWSE PUBLISH ABOUT

## PLOS GENETICS


OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

### Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data

Joseph K. Pickrell , Jonathan K. Pritchard 

Published: November 15, 2012 • <https://doi.org/10.1371/journal.pgen.1002967>

admixtools 2.0.0  Get started Reference Articles

## ADMIXTOOLS 2 Tutorial

Robert Maier

2021-02-23

Source: vignettes/admixtools.Rmd

This tutorial gives an overview of the basic workflow for computing  $f$ -statistics, and for using *qpWave*, *qpAdm*, and *qpGraph*. Documentation for each *ADMIXTOOLS 2* function can be found under *Reference*, and more detailed information about specific topics under *Articles*.

For the examples here and on the other pages, the following R packages need to be loaded.

```
library(admixtools)
library(tidyverse)
```

While this website focuses on the R command line interface of *ADMIXTOOLS 2*, there is also a browser application that can be launched from R with the following command:

```
run_shiny_admixtools()
```

