Admixture graphs

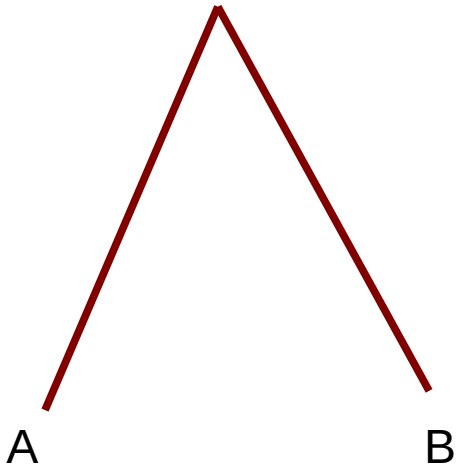
Shyam Gopalakrishnan

Aug 8th 2019

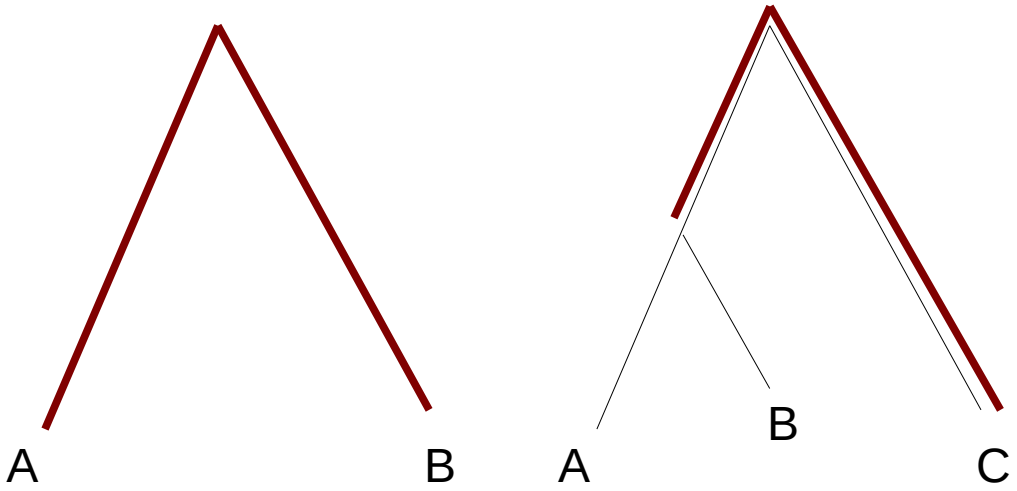
Overview

- Recap of our morning
- More than 4 populations, what do I do?
  - Correlations based methods
  - F stats based
- What if a tree is not enough?
  - Incorporating migrations

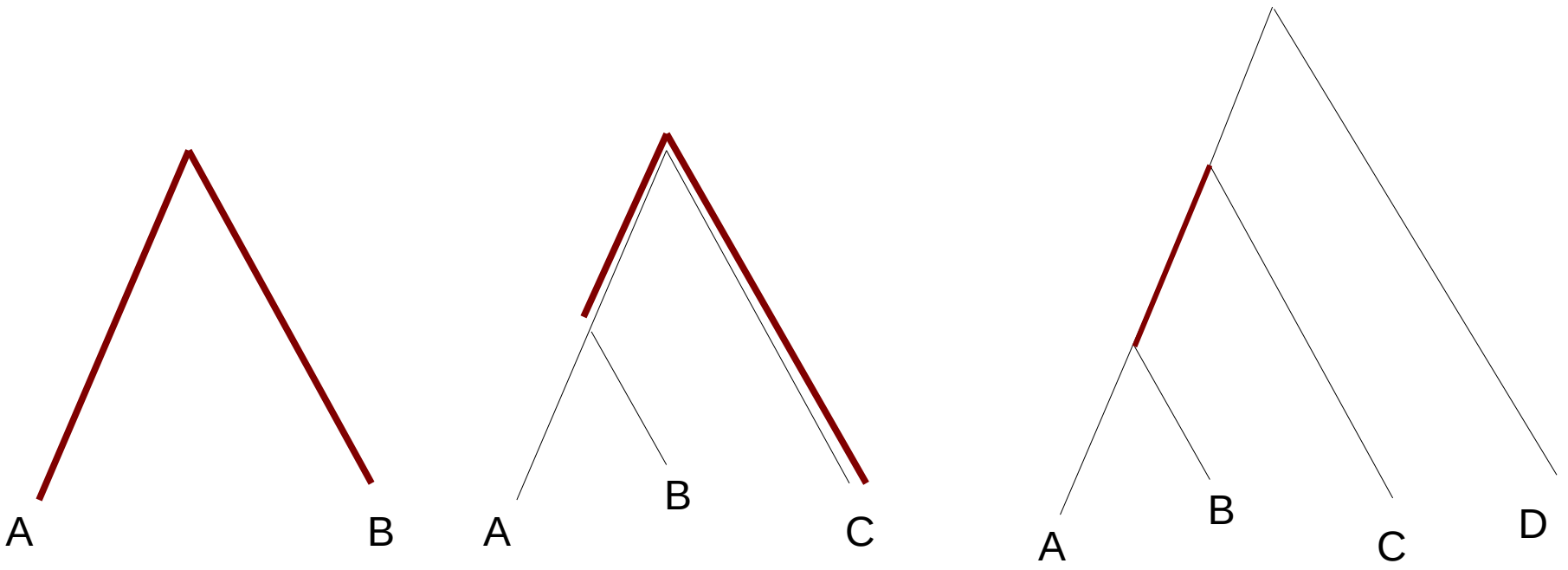
Recap: F - stats



Recap: F - stats

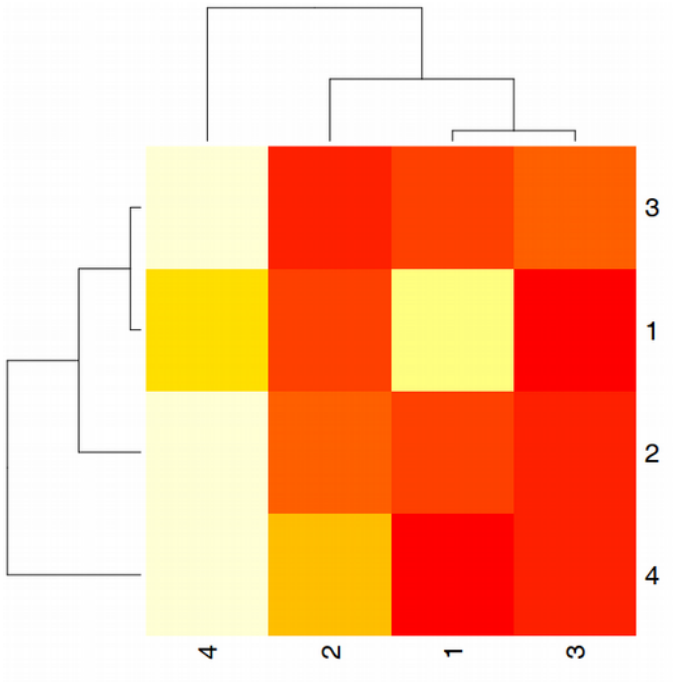


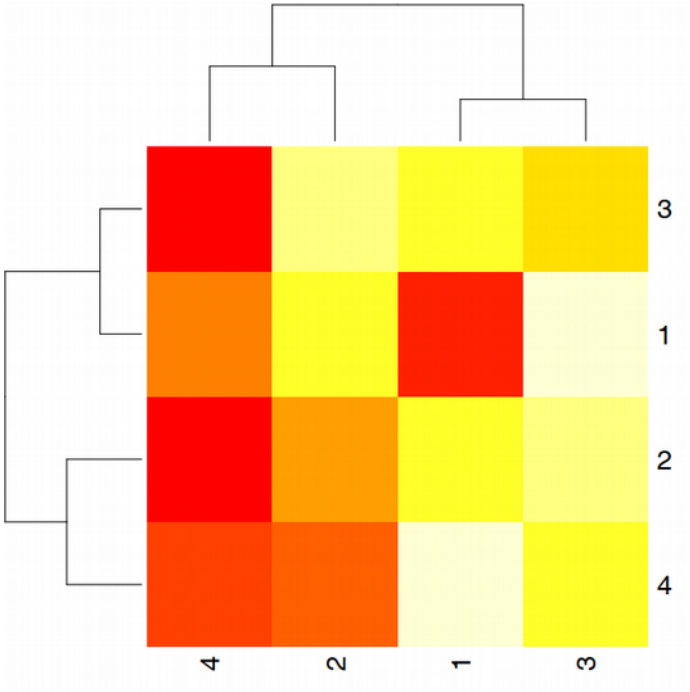
Recap: F - stats



Similarity vs dissimilarity?

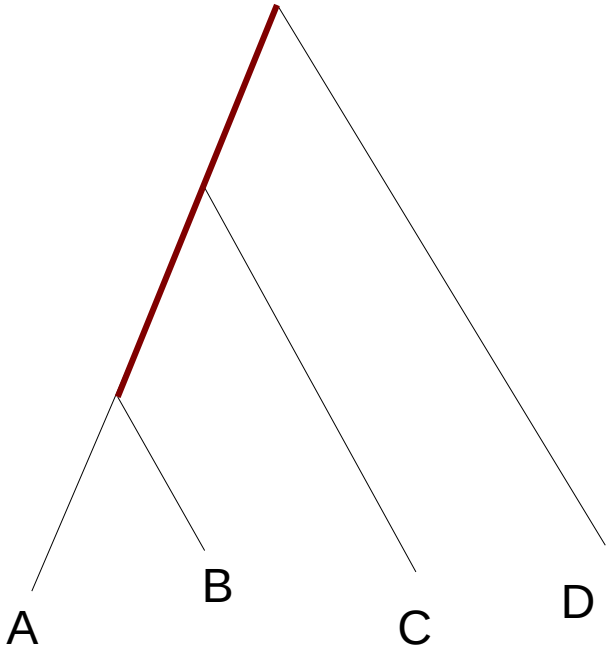
- Correlations or distances?





Correlation based methods

- Trees induce correlation in allele frequencies



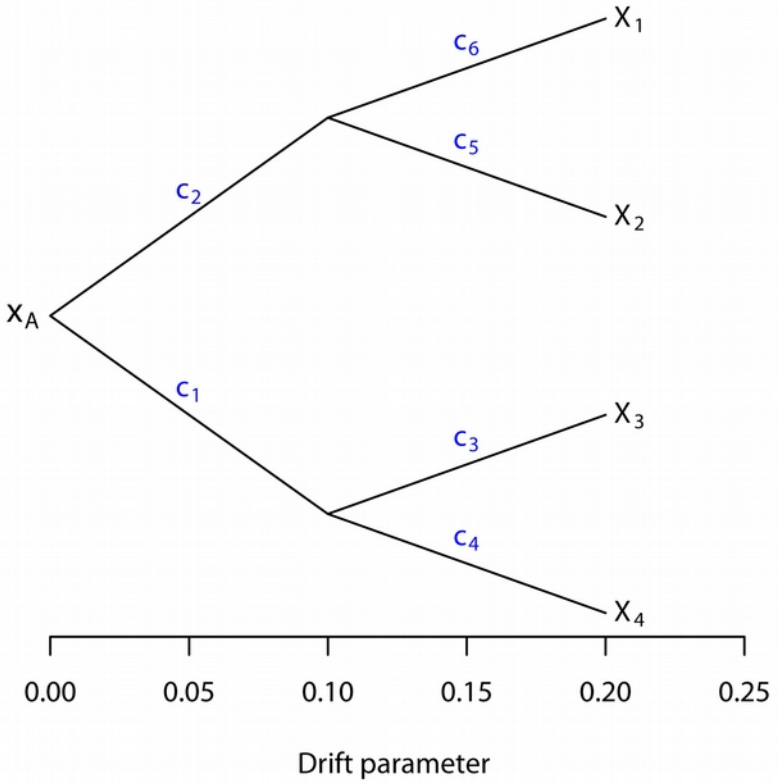
Correlation of AF between A and B:
Shared branch lengths!

Using correlations to construct trees

- TreeMix
  - Use correlations between all pairs of populations to construct tree
  - Use difference between estimated correlation and observed correlation to add migration edges.

TreeMix

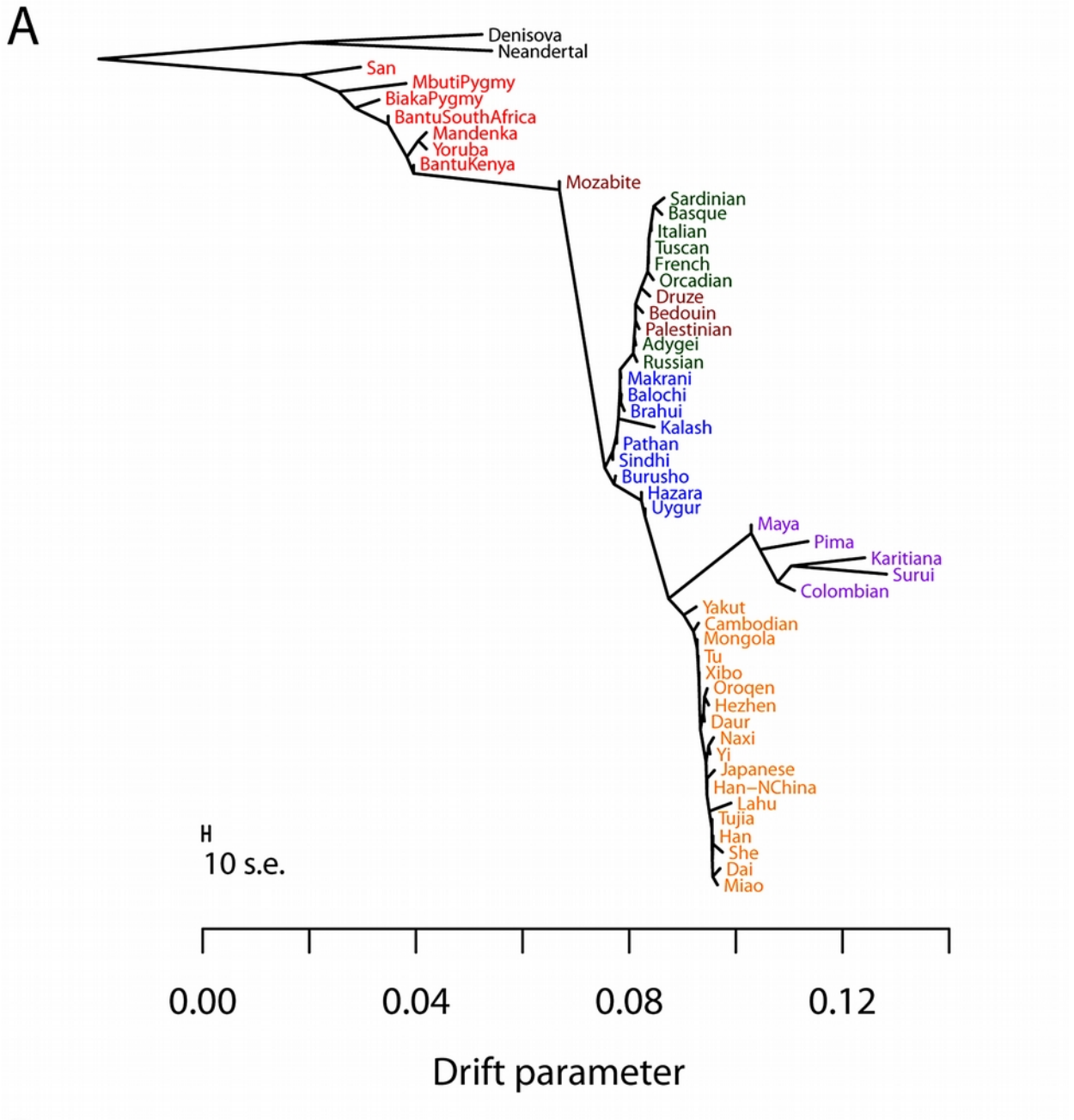
A. Example tree



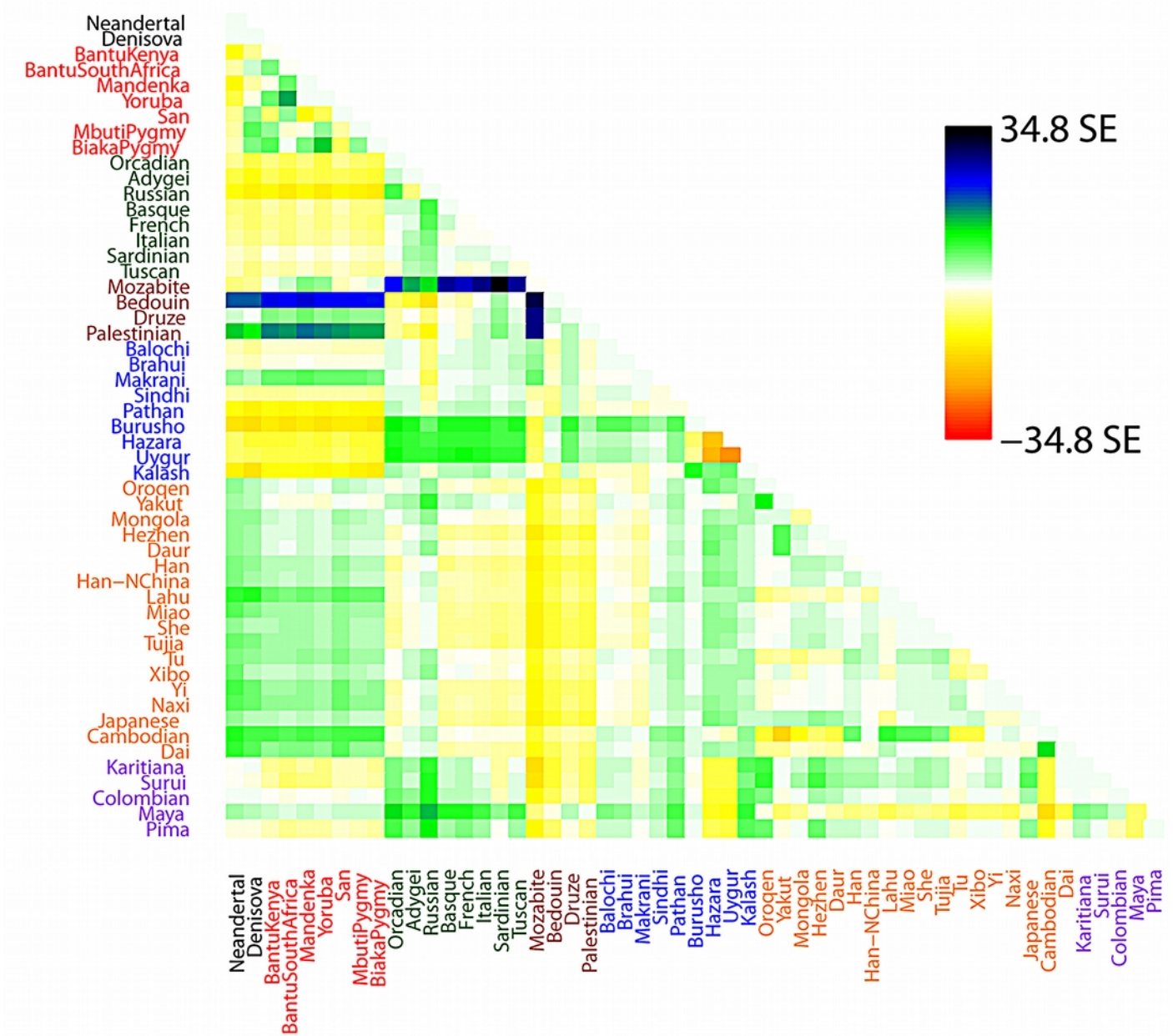
B. Covariance matrix for tree in A.

	X_1	X_2	X_3	X_4
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	$c_1 + c_4$

TreeMix on humans



TreeMix on humans

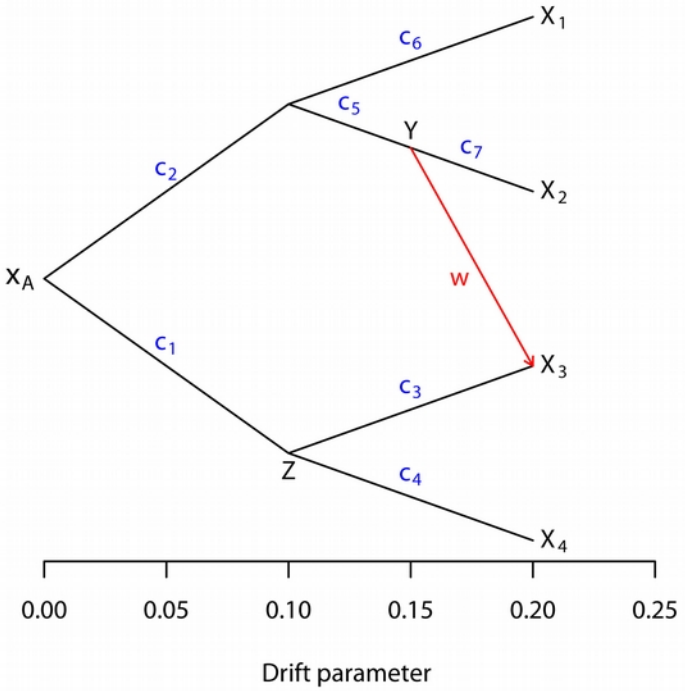


When a Tree is not enough

- Admixture events make the relationship between populations not be a tree anymore – remember the F3 statistic?
- We need to add admixture events to our trees

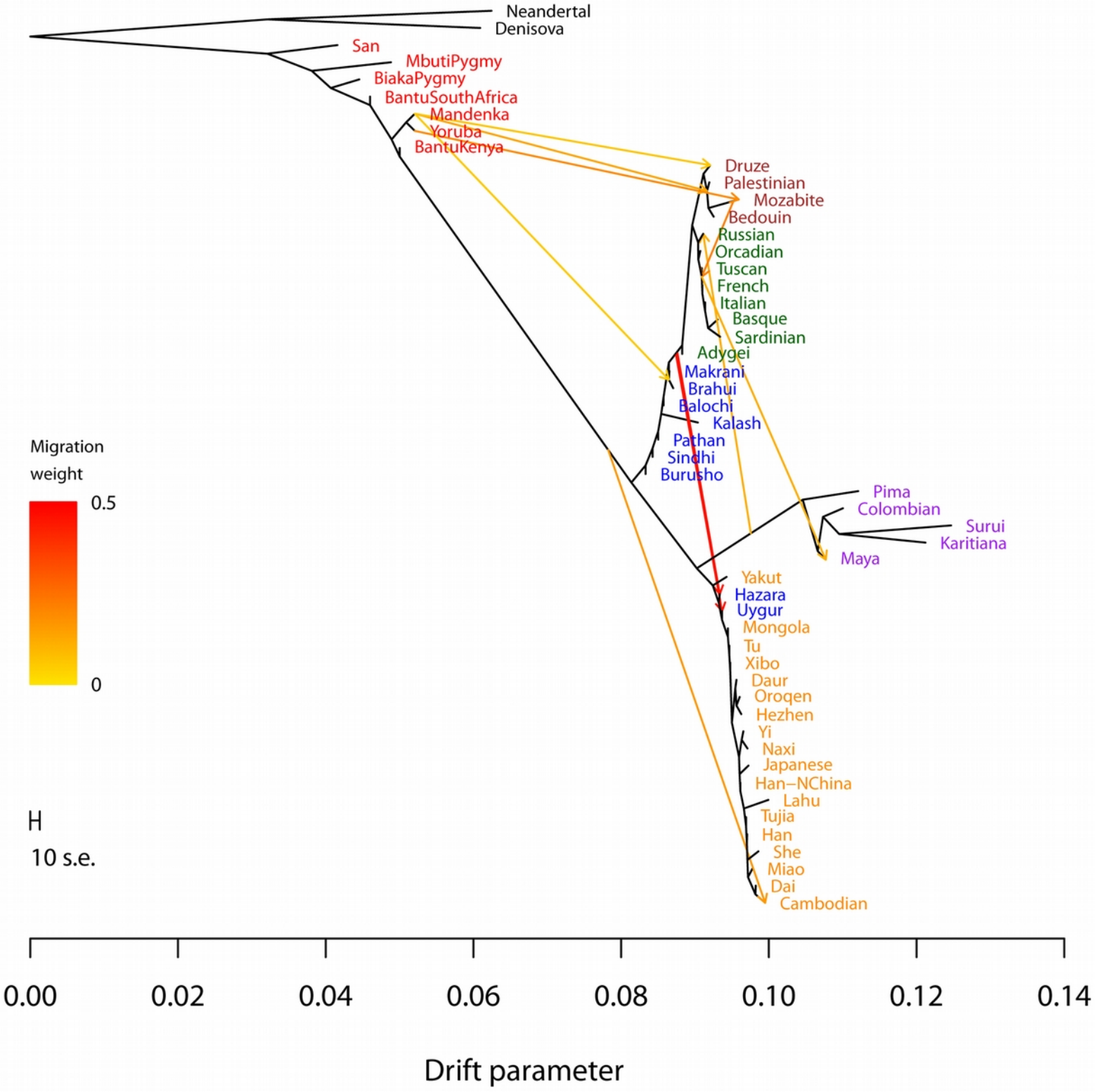
TreeMix with admixture

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	wc_2	0
X_2	c_2	$c_2 + c_5 + c_7$	$w(c_2 + c_5)$	0
X_3	wc_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	$(1-w)c_1$	$c_1 + c_4$


Neandertal
Denisova
San
MbutiPygmy
BiakaPygmy
BantuSouthAfrica
Mandenka
Yoruba
BantuKenya
Druze
Palestinian
Mozabite
Bedouin
Russian
Orcadian
Tuscan
French
Italian
Basque
Sardinian
Adygei
Makrani
Brahui
Balochi
Kalash
Pathan
Sindhi
Burusho
Pima
Colombian
Surui
Karitiana
Maya
Yakut
Hazara
Uygur
Mongola
Tu
Xibo
Daur
Oroqen
Hezhen
Yi
Naxi
Japanese
Han−NChina
Lahu
Tujia
Han
She
Miao
Dai
Cambodian
Migration
weight
0.5
0
H
10 s.e.
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Drift parameter

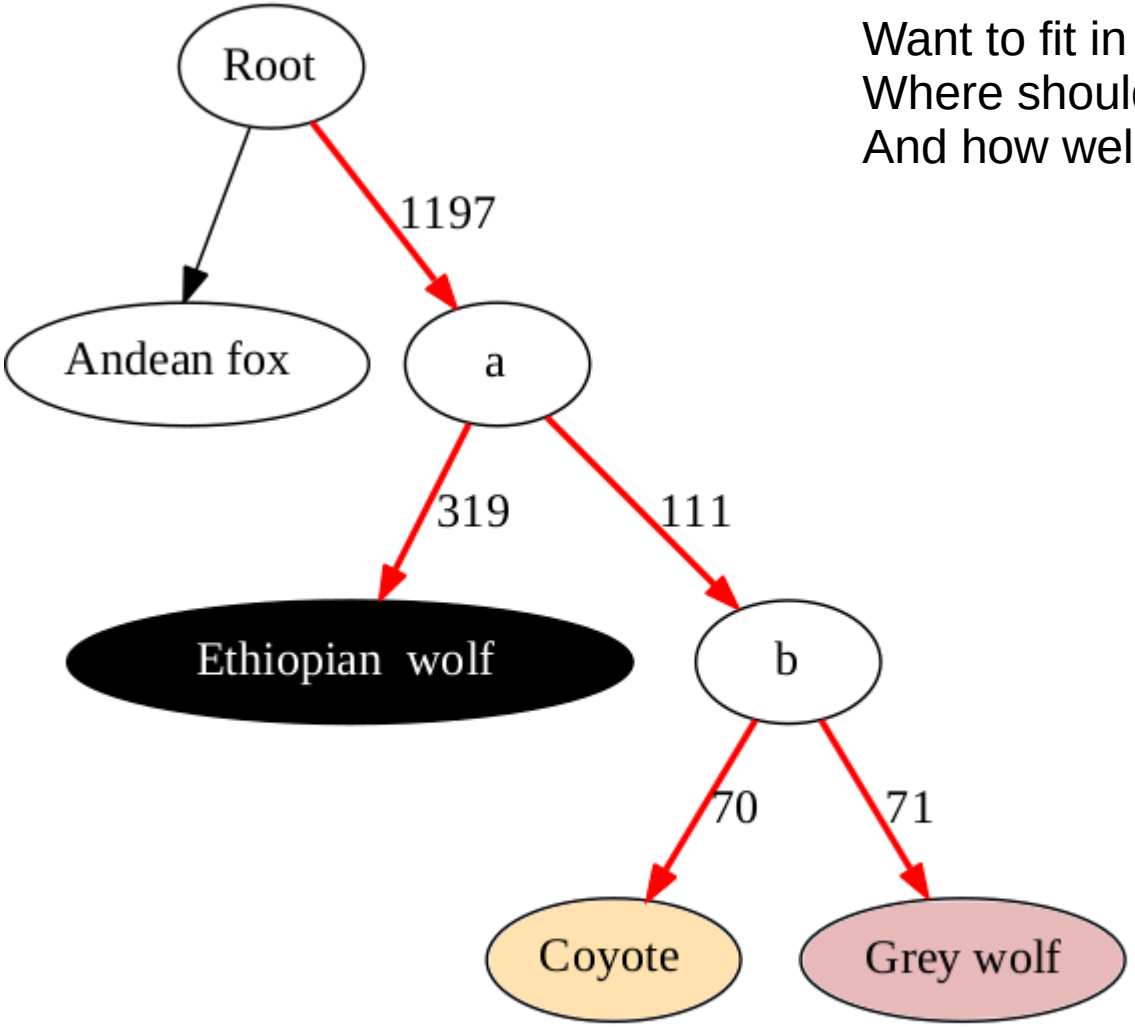
How can we use F-stats to make trees and graphs?

- Since a tree induces a correlations structure, it also induces a dissimilarity measure.
  - If we have all pairwise F2 statistics, can we build a first-guess tree?
  - Using the tree, can we make estimate what the F3 and F4 statistics should be?
  - How can we estimate admixture?

Admixture graphs: qpGraph

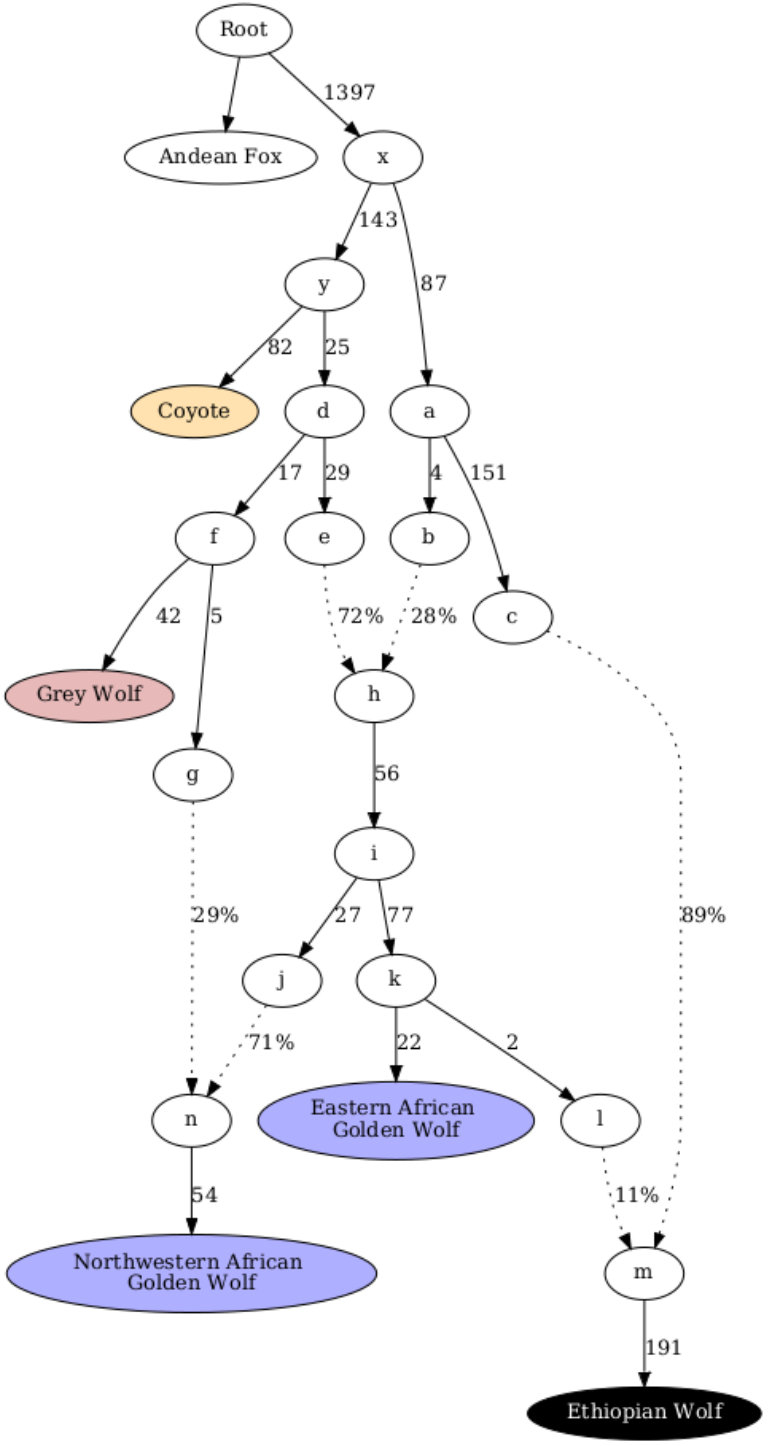
- Similar in idea to TreeMix, but less automated
  - Estimate a tree (or a graph) first
  - Figure out the F-statistics induced by this tree
  - Find out where the induced F-statistics differ from the observed F-statistics
  - Add migration edges to account for these mismatches

qpGraph on Canids



Want to fit in African golden wolves?
Where should we place them?
And how well do they fit?

qpGraph on Canids

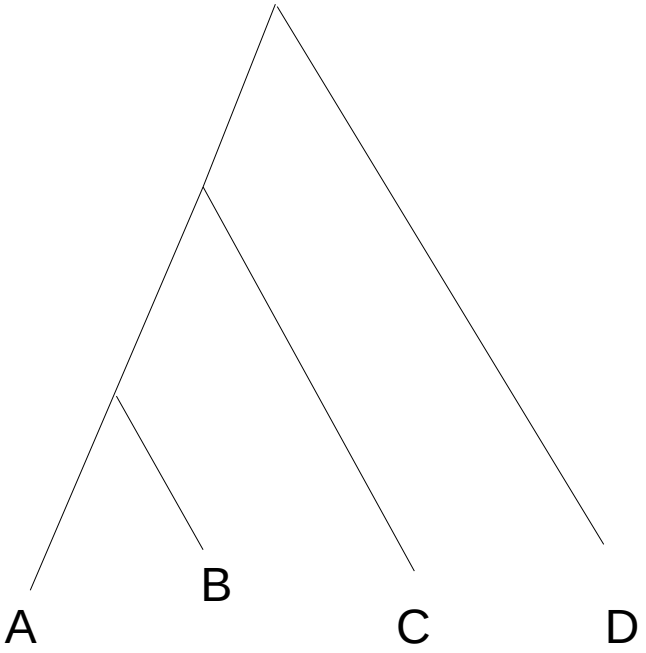


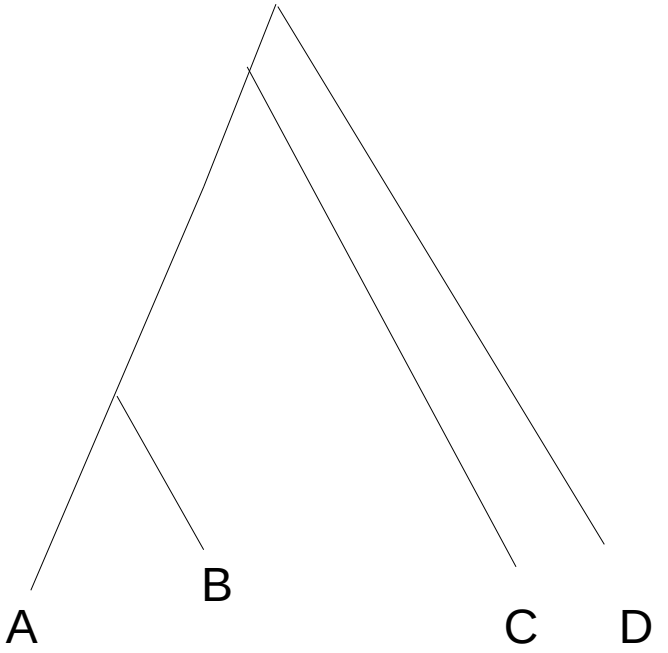
What more can we do with admixture graphs?

- Model complicated histories
- Remember Trondheim?
  - If we can model the genetic compostions of multiple populations, can we improve our selection analysis?

Selection on admixture graphs

- Consider a simple tree
  - Are there parts of the genome where the tree has wacky branch lengths?



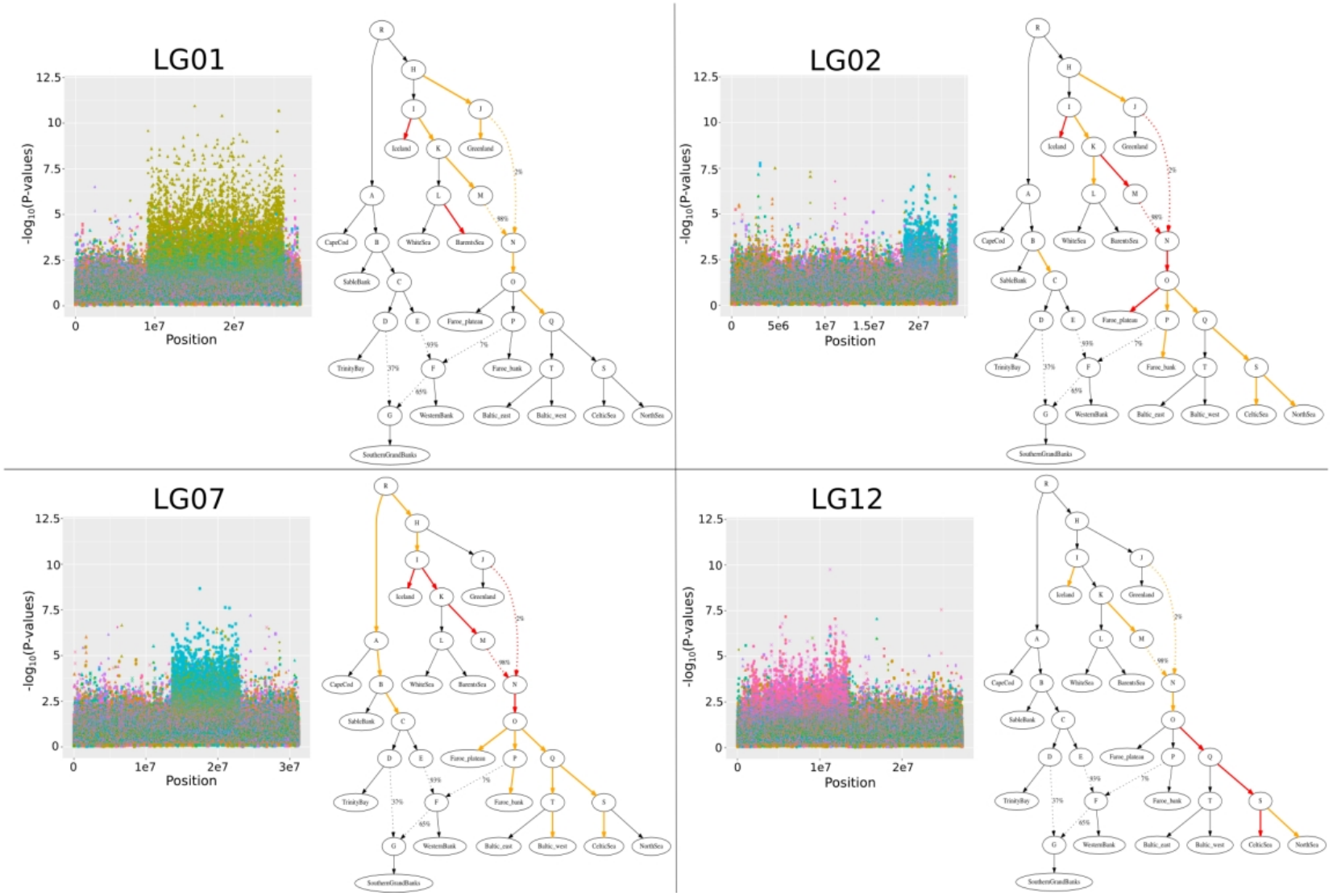


Can we do this with arbitrarily complex histories?

- GroSS (Graph-aware Retrieval of Selective Sweeps) - program to detect selective sweeps on complex histories.

Alba Refoyo-Martinez et al., biorxiv 2018

GRoSS application



Exercise time!!