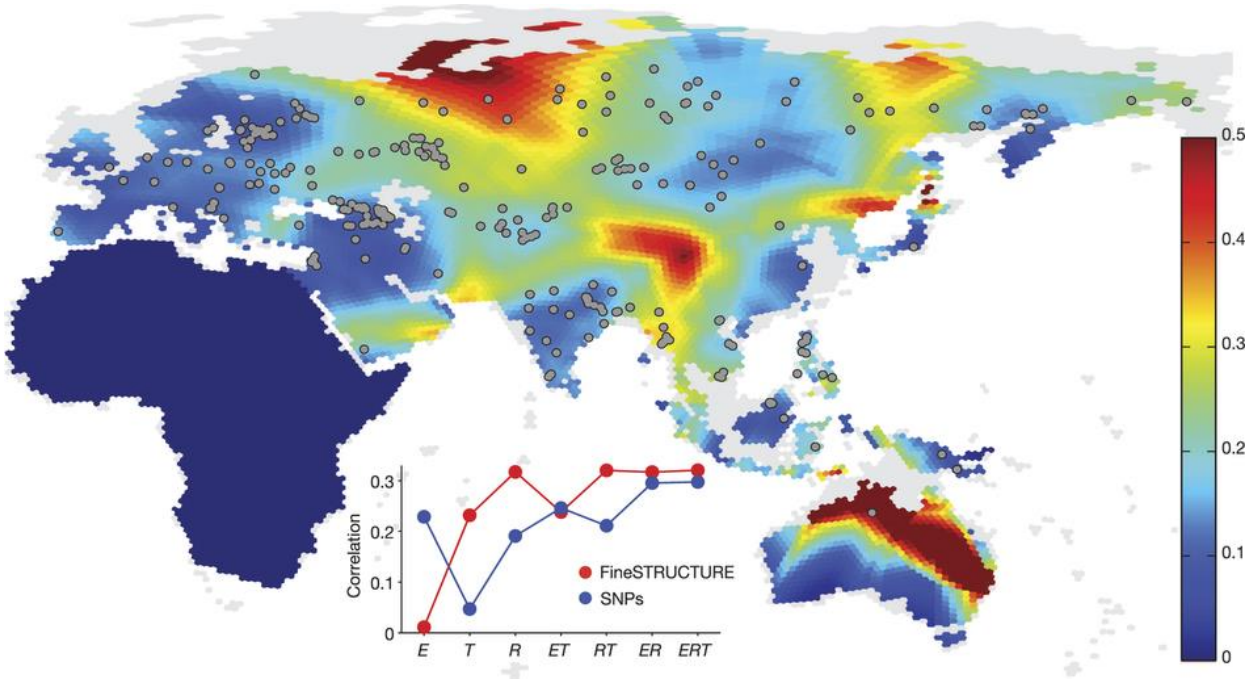
Quantifying relationships among populations: Many populations

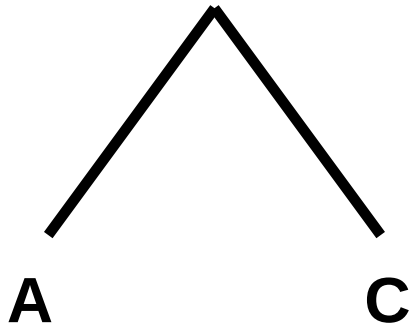


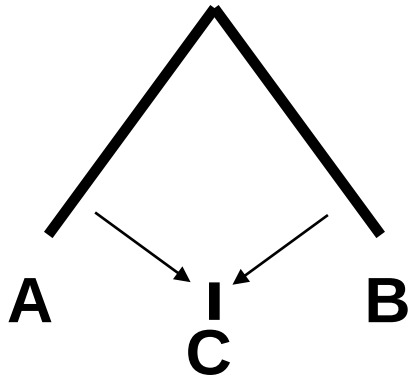
Andrea Manica

UNIVERSITY OF CAMBRIDGE

Outline

- Generalising f_4
- Quantifying distances between populations
- Building trees
- Admixture Graphs
- Some spatial statistics





qpWave: generalising f_4

	0	1	2
L: pop_a, pop_b, pop_c			

	0	1	2
R: pop_d, pop_e, pop_f			

$X(l_i,r_i) = f_4(l_0,l_i; r_0, r_i)$

rank + 1 ≤
n waves

f_4(a,b;d,e)	f_4(a,b;d,f)
f_4(a,b;d,e)	f_4(a,b;d,e)

qpAdm: estimating proportions from multiple sources

T: pop_t

S: pop_a, pop_b, pop_c

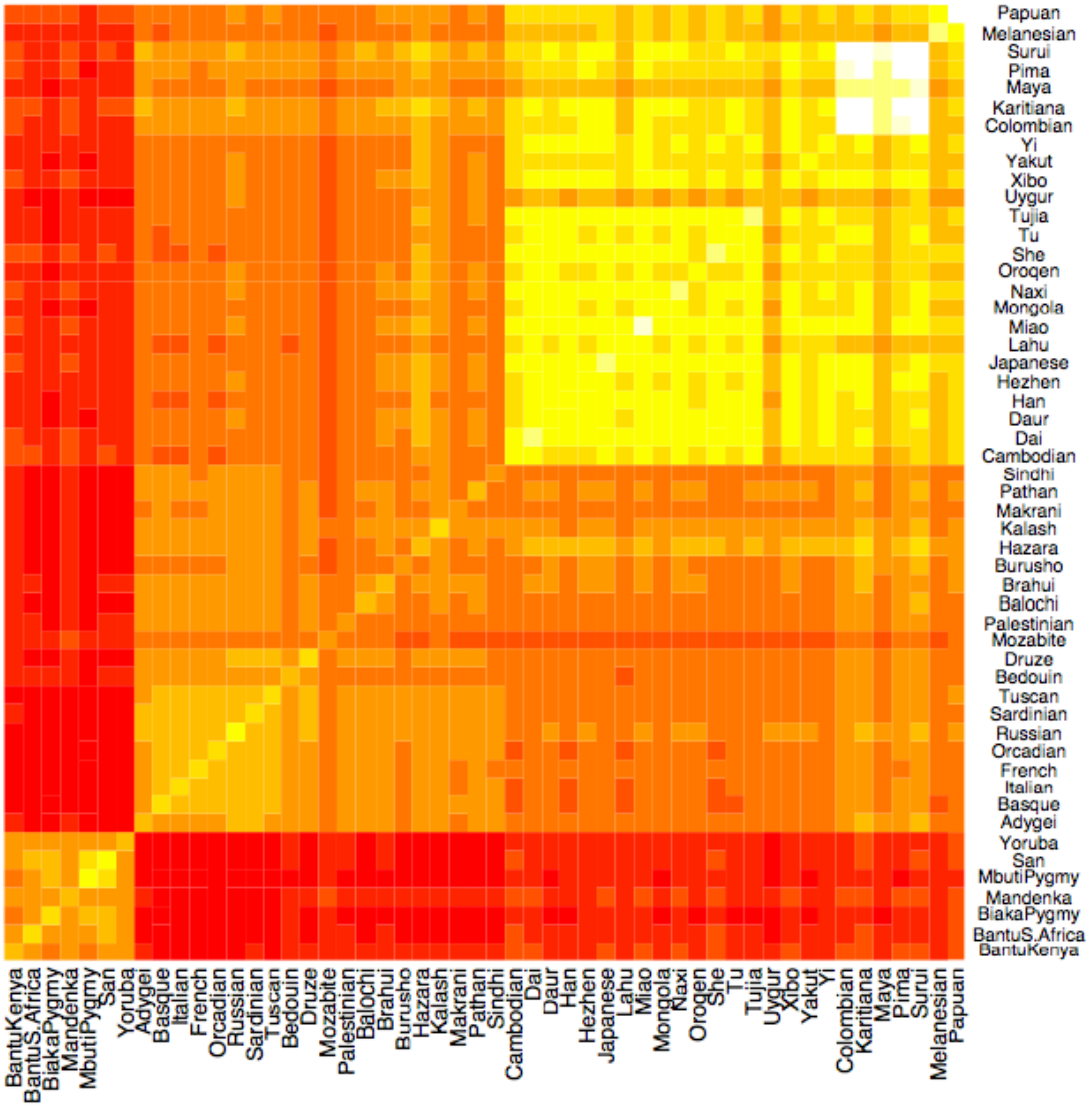
R: pop_d, pop_e, pop_f

$$T = \sum_{i=1}^{n} w_i s_i$$

$$\sum_i w_i f_4(T, s_i, r_1, r_2) = f_4(T, T, r_1, r_2)$$

$$= 0$$

Quantifying distances among populations



Quantifying distances among populations

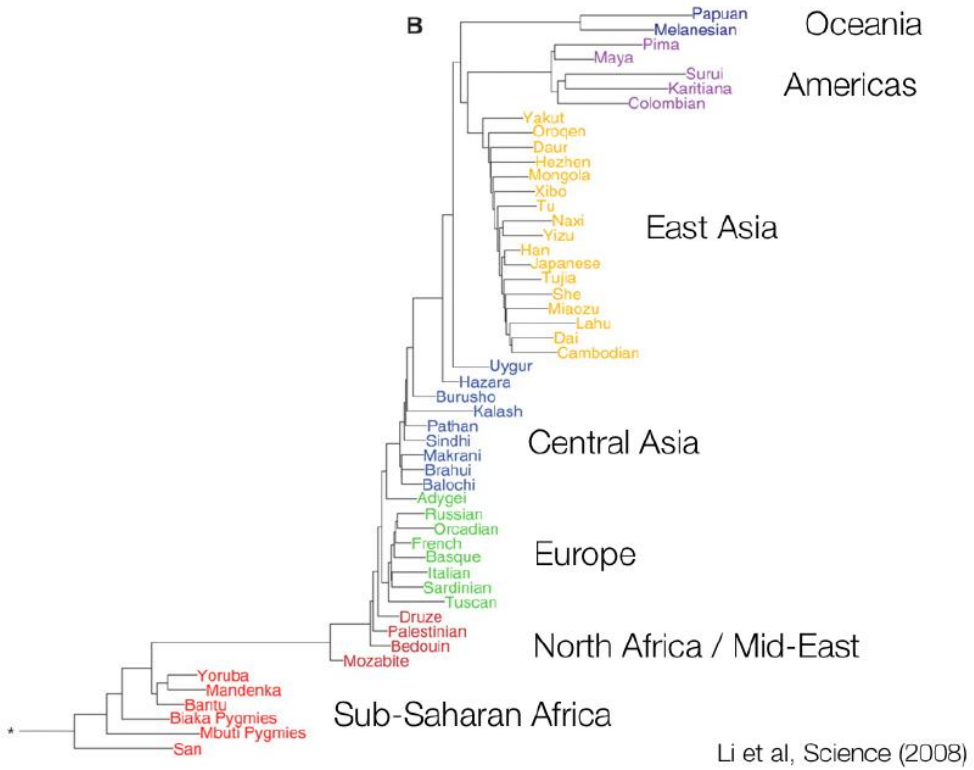
$f_2(A,B) = E[(p_A - p_B)^2]$

$$\hat{F}_2(P_1, P_2) = \pi_{12} - \frac{\pi_{11} + \pi_{22}}{2}.$$

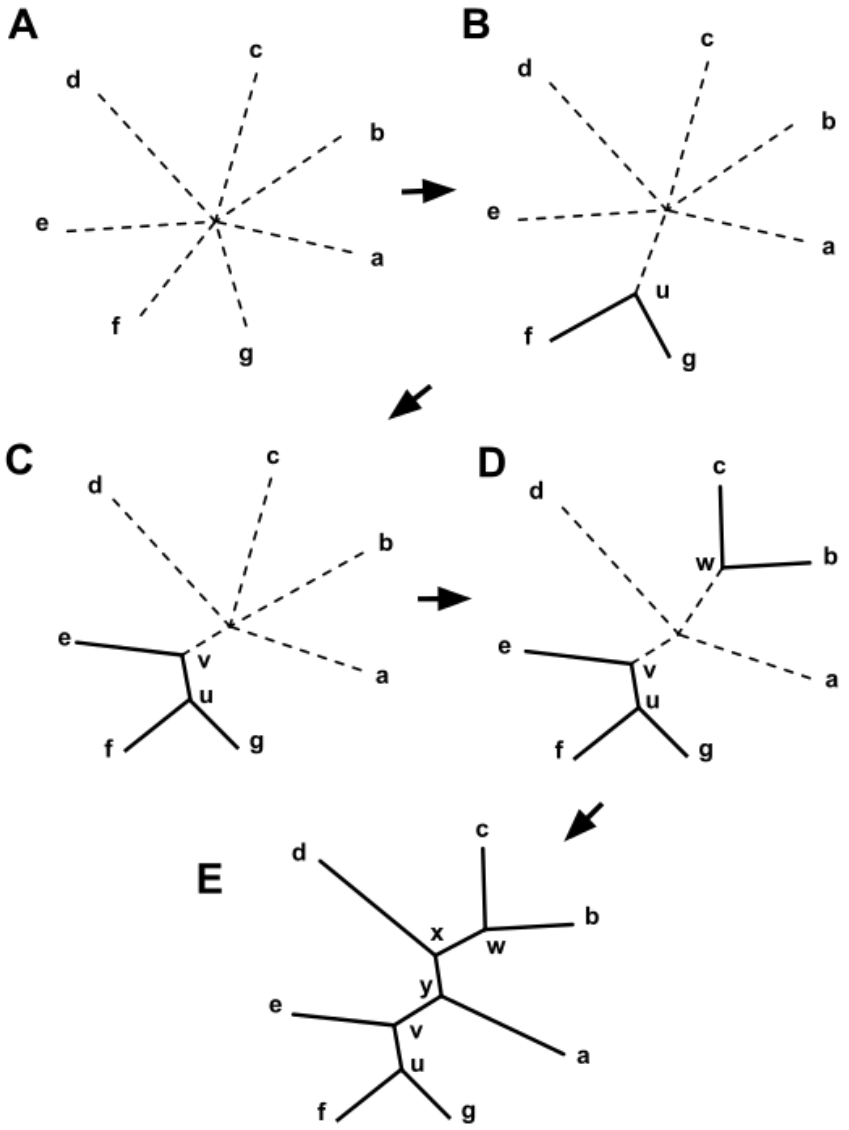
$$F_2(P_1, P_2) = \frac{1}{2} F_{\mathrm{ST}} \mathbb{E} H_{\mathrm{exp}}.$$

$$F_2(P_1, P_2) = 2\mathbb{E}T_{12} - \mathbb{E}T_{11} - \mathbb{E}T_{12}$$

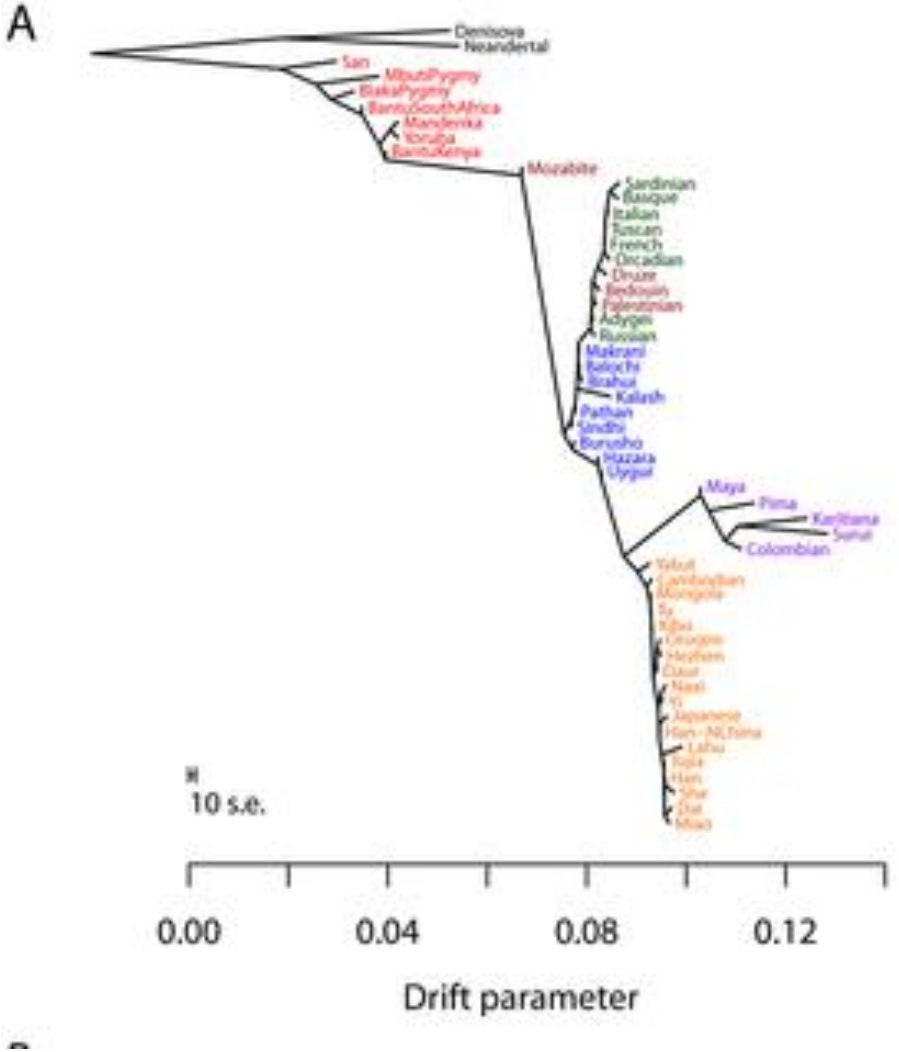
Quantifying distances among populations

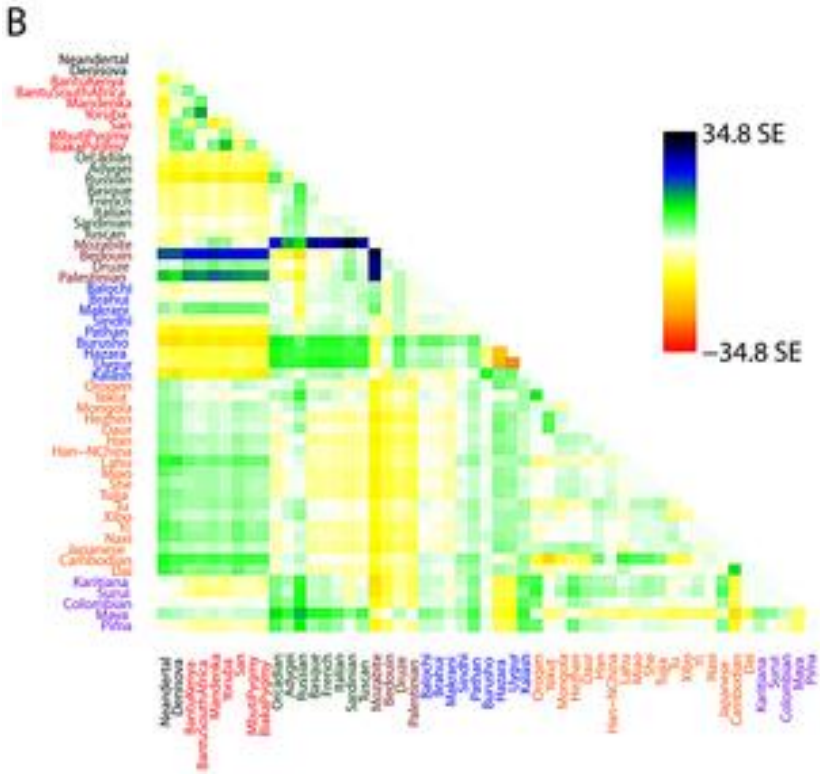


Neighbour Joining trees

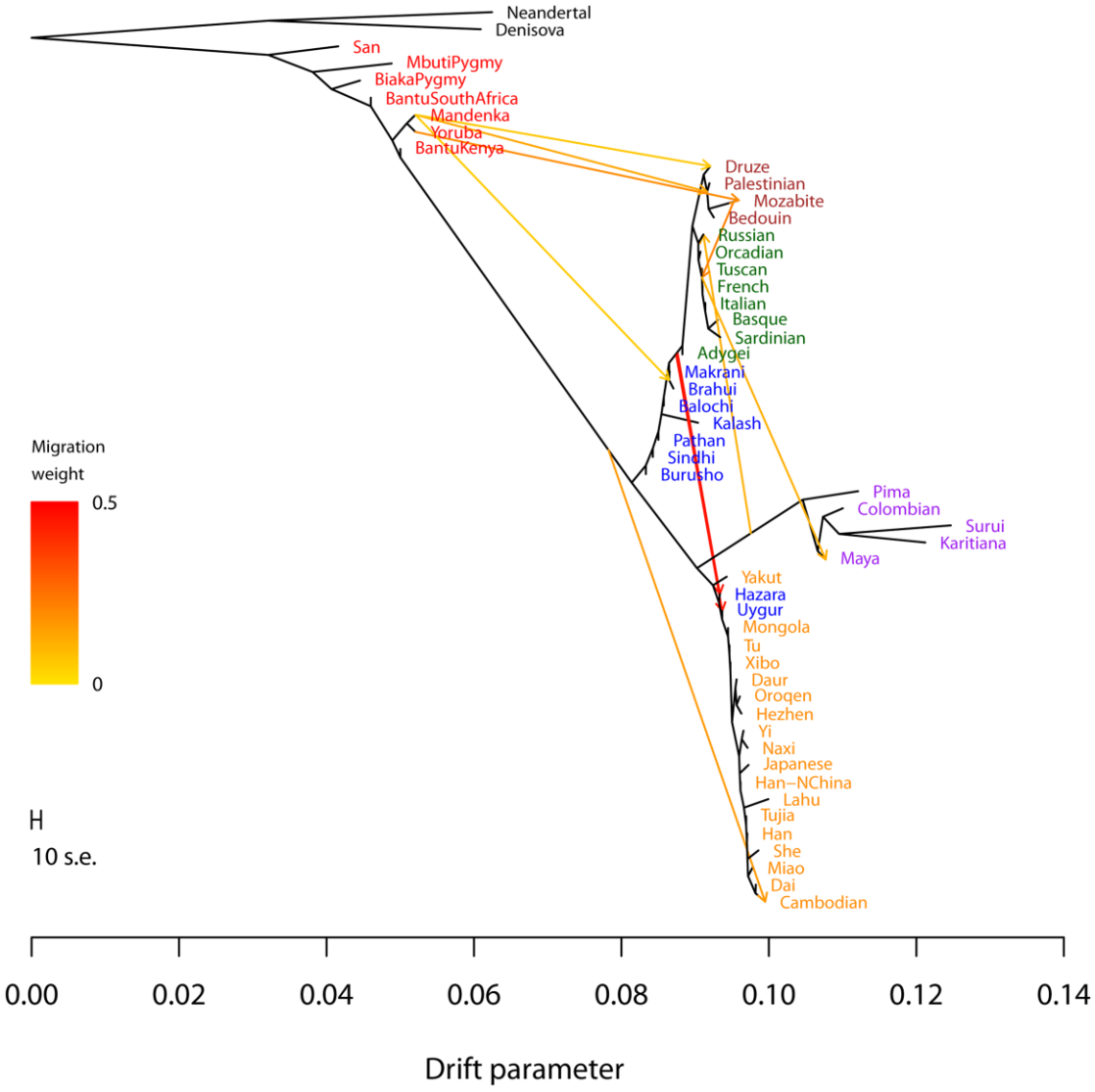


Breaking the tree

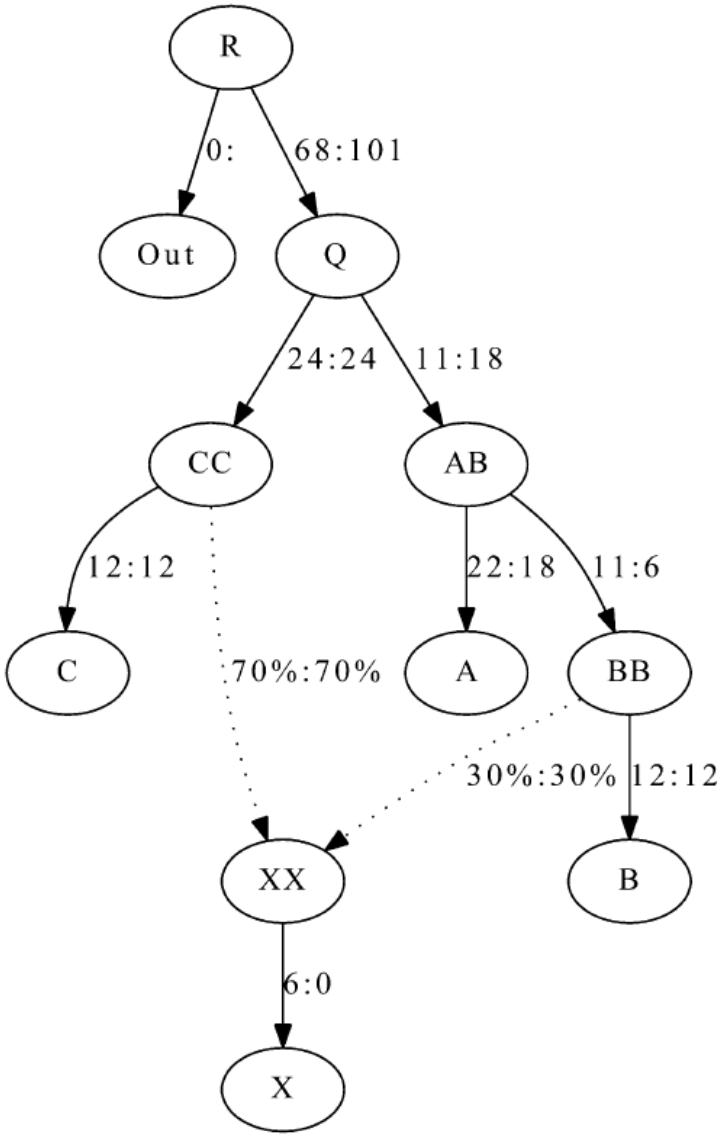




Treemix

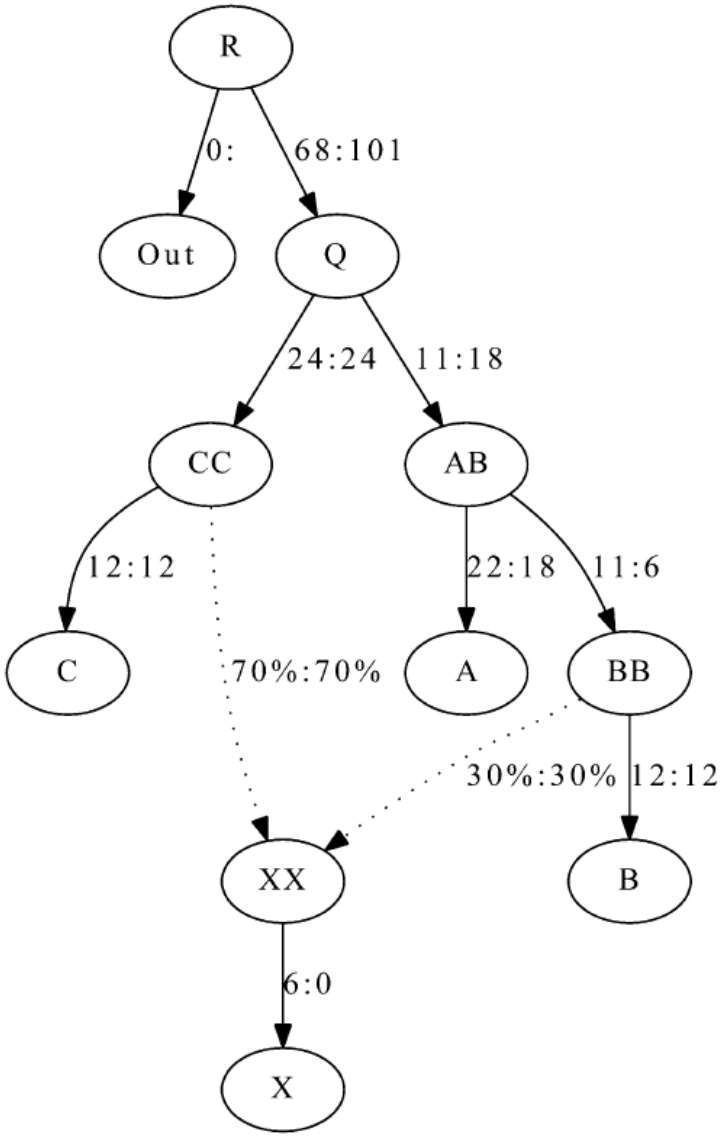

Neandertal
Denisova
San
MbutiPygmy
BiakaPygmy
BantuSouthAfrica
Mandenka
Yoruba
BantuKenya
Druze
Palestinian
Mozabite
Bedouin
Russian
Orcadian
Tuscan
French
Italian
Basque
Sardinian
Adygei
Makrani
Brahui
Balochi
Kalash
Pathan
Sindhi
Burusho
Pima
Colombian
Surui
Karitiana
Maya
Yakut
Hazara
Uygur
Mongola
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

qpGraph



- Fitted on f_2 or f_3
- First fit unadmixed skeleton
- Then test admixture scenarios
- Check for mismatches in predicted vs observed f_3 and f_4
- Not exhaustive, multiple graphs might fit data equally

qpGraph – changing philosophies

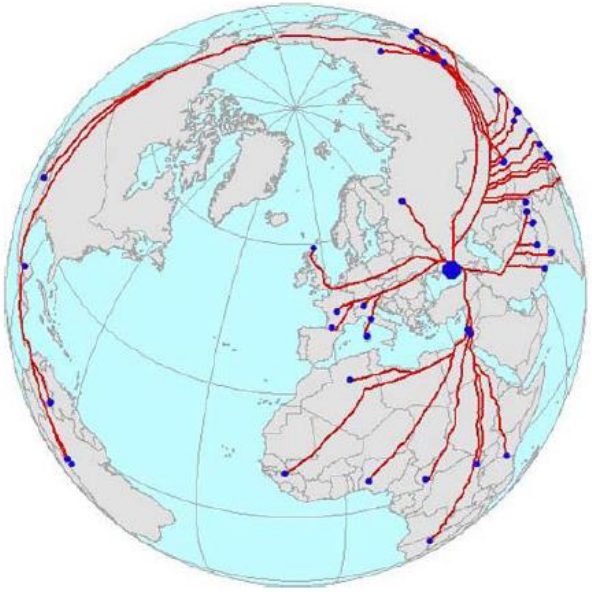


Even for simple scenarios, there are many graph that fit the data

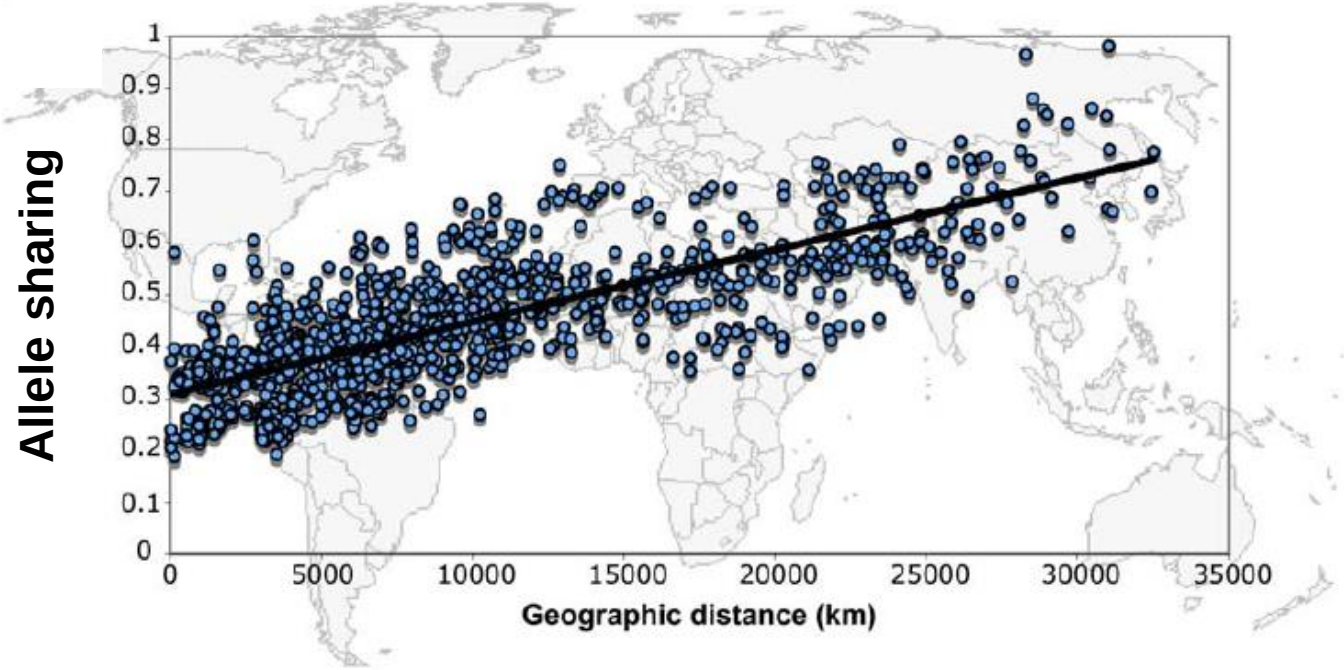
Manual searches are not enough

Extensive searches are needed, and evidence from admixture graphs needs to be complemented with other approaches

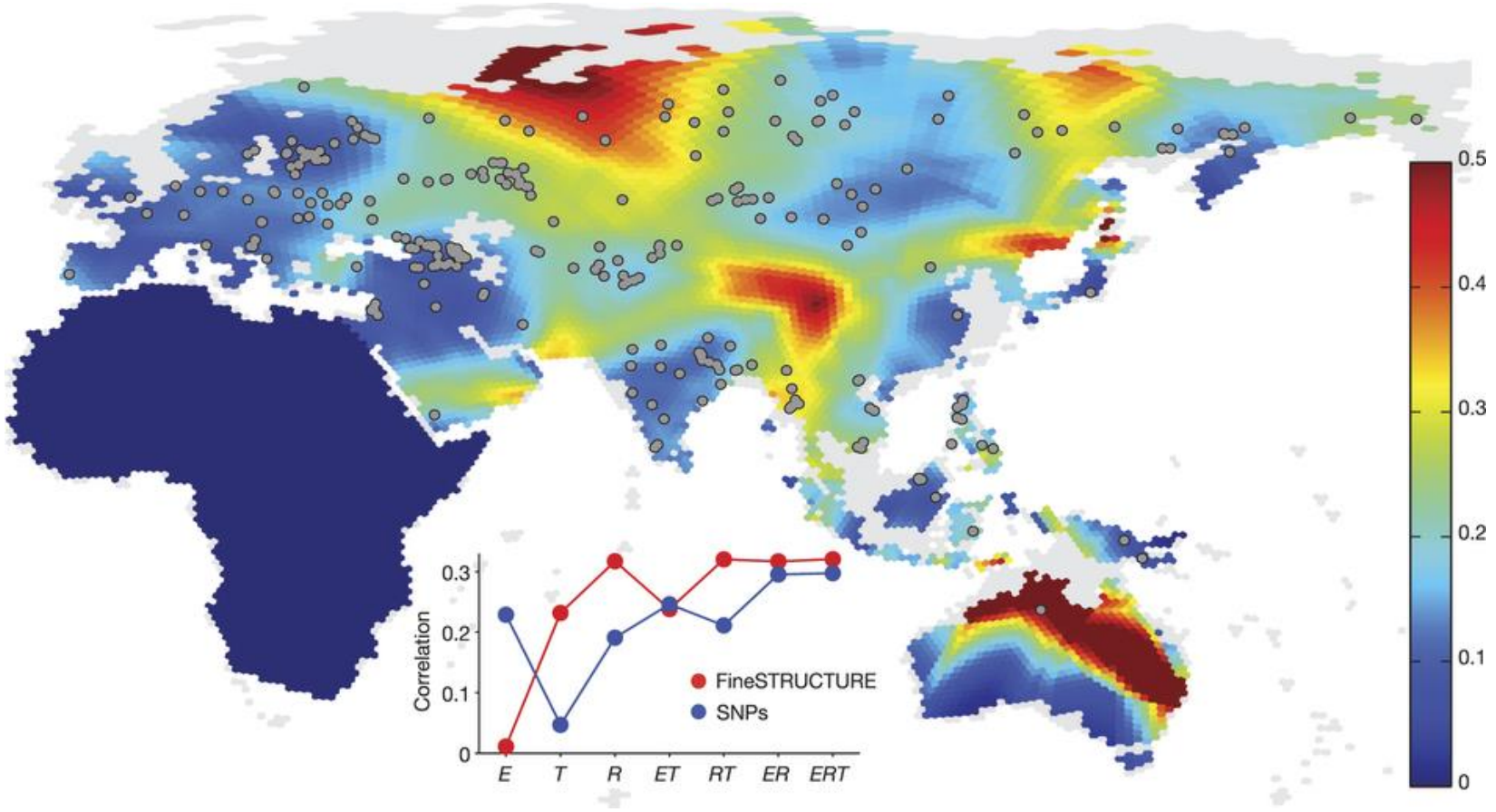
Using space to model many populations



Isolation by distance can explain a lot of differences

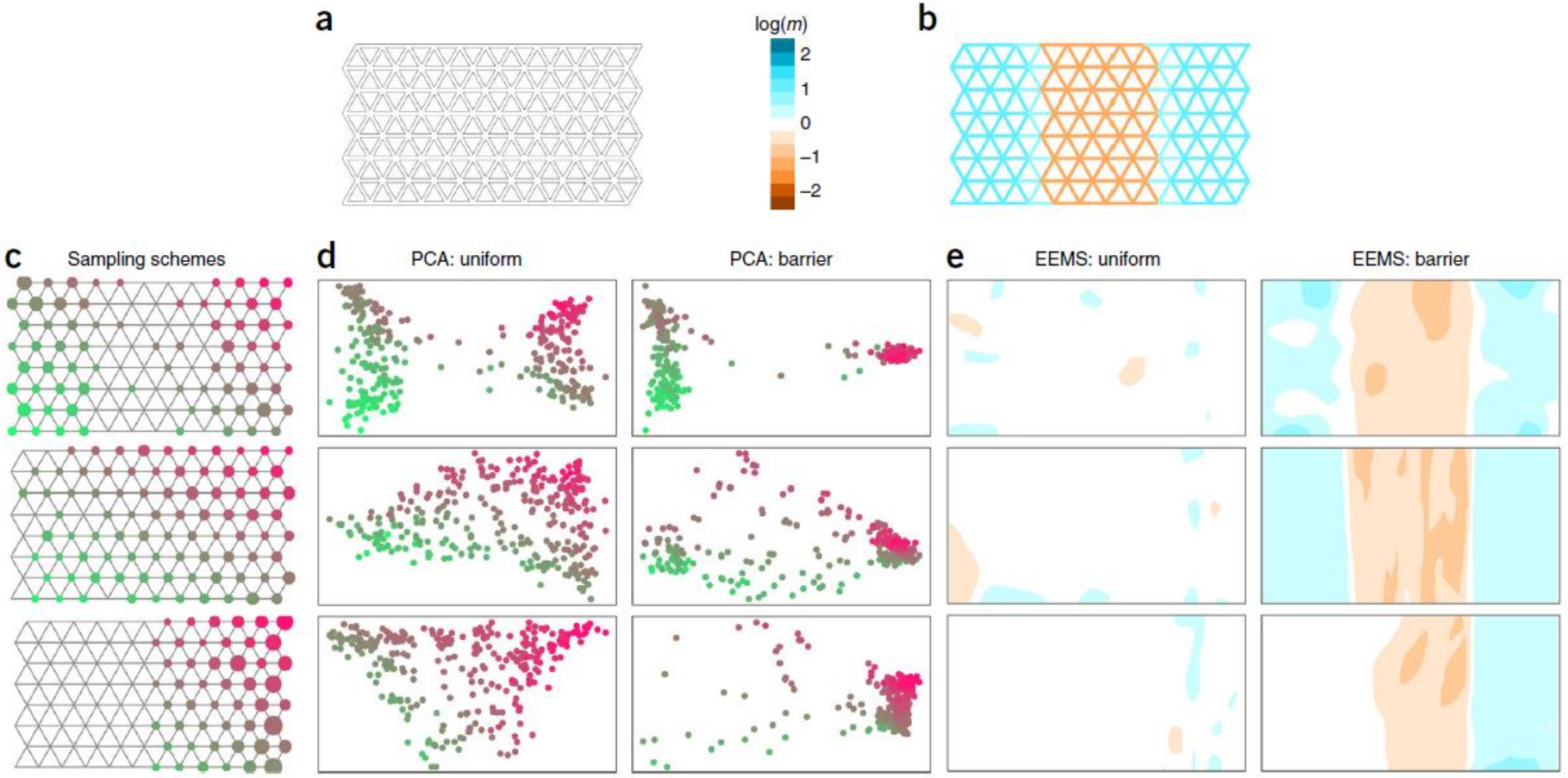


Barriers for human movement

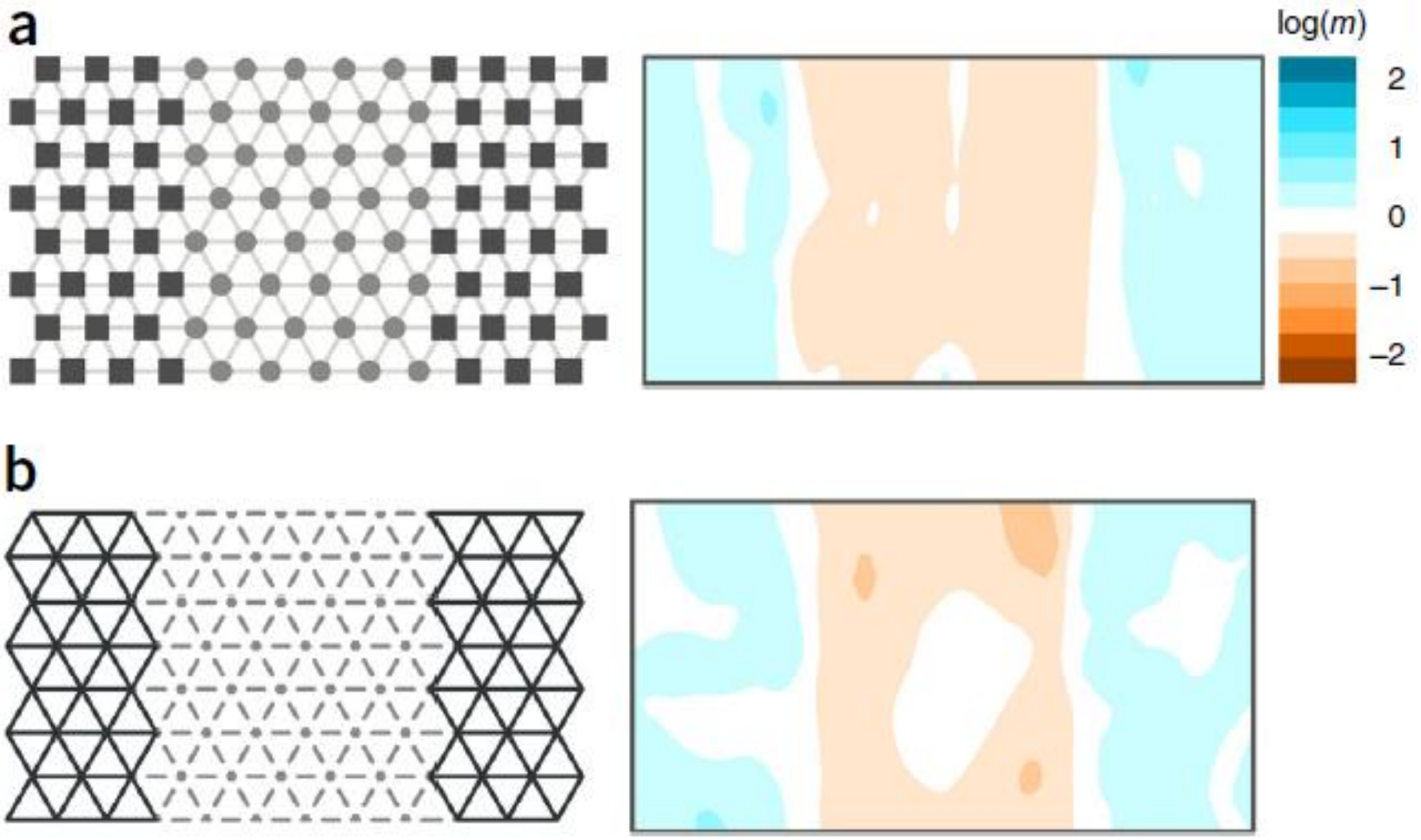


Pagani et al. 2016 Nature

EMMS

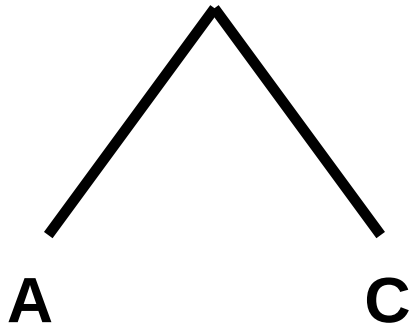

a
log(m)
2
1
0
−1
−2
b
c
Sampling schemes
d
PCA: uniform
PCA: barrier
e
EEMS: uniform
EEMS: barrier

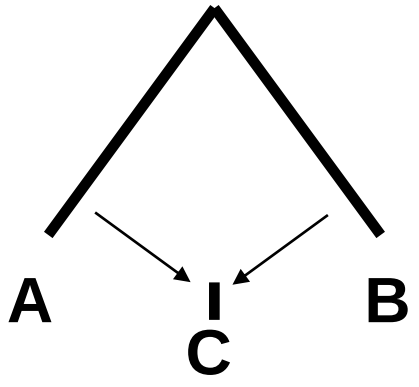
EMMS – migration vs population size



Outline

- Generalising f_4
- Quantifying distances between populations
- Building trees
- Admixture Graphs
- Some spatial statistics





Practical

- Use Admixtools to fit admixture graph
- Human dataset with modern and ancient