

Lecture 8:

Space

Spatial population genetics

landscape genetics

geographical genetics

SPACE IS
THE PLACE



*Annual Review of Ecology, Evolution, and
Systematics*

Spatial Population Genetics: It's About Time

Gideon S. Bradburd¹ and Peter L. Ralph^{2,3}



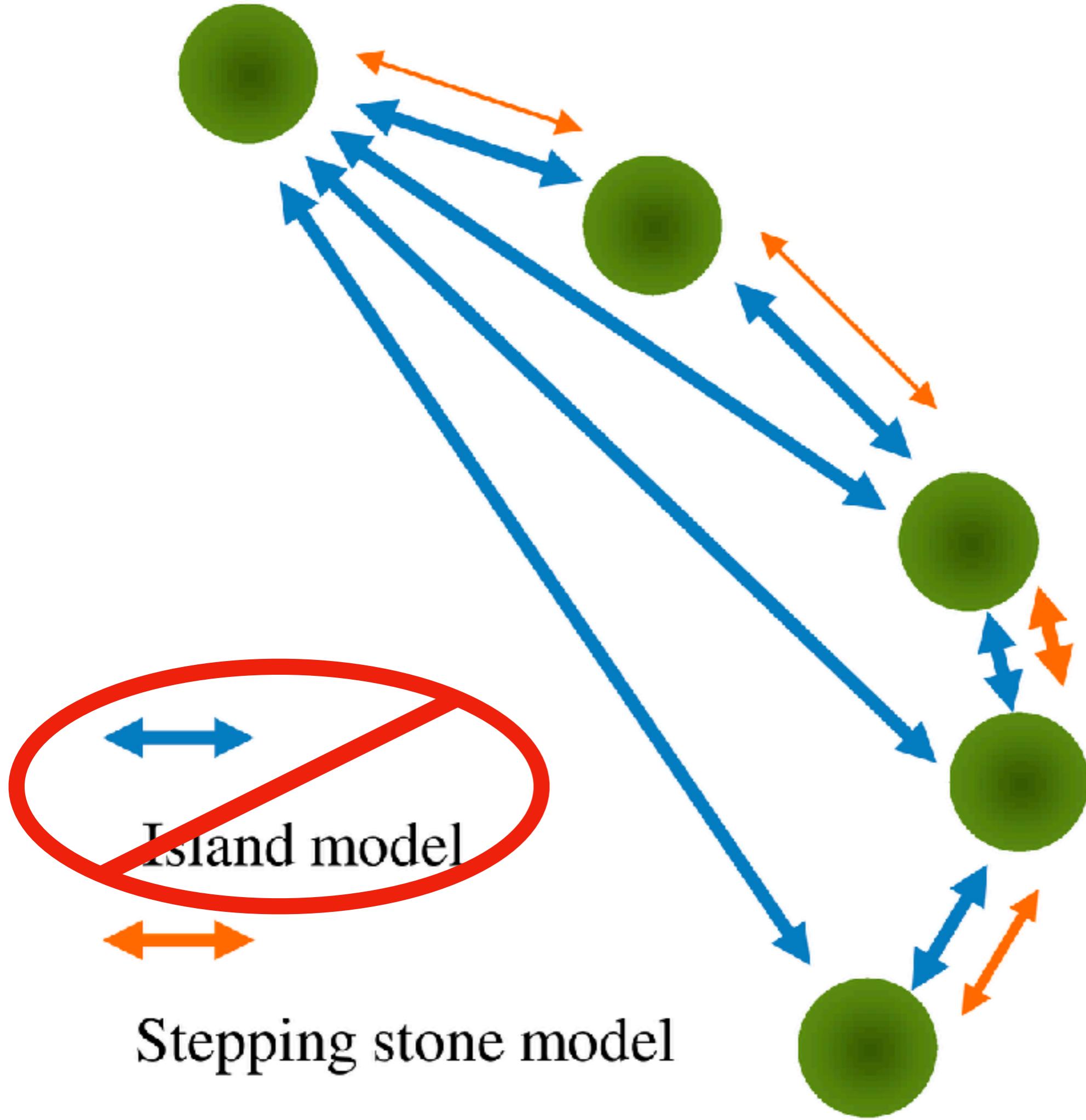
Spatial population genetics

landscape genetics

geographical genetics

SPACE IS
THE PLACE

historical vs ahistorical



Stepping stone model

Dobeš et al. 2017

variance in parent/offspring distance

approaches isolation by distance (IBD) expectations (ie more with distance)

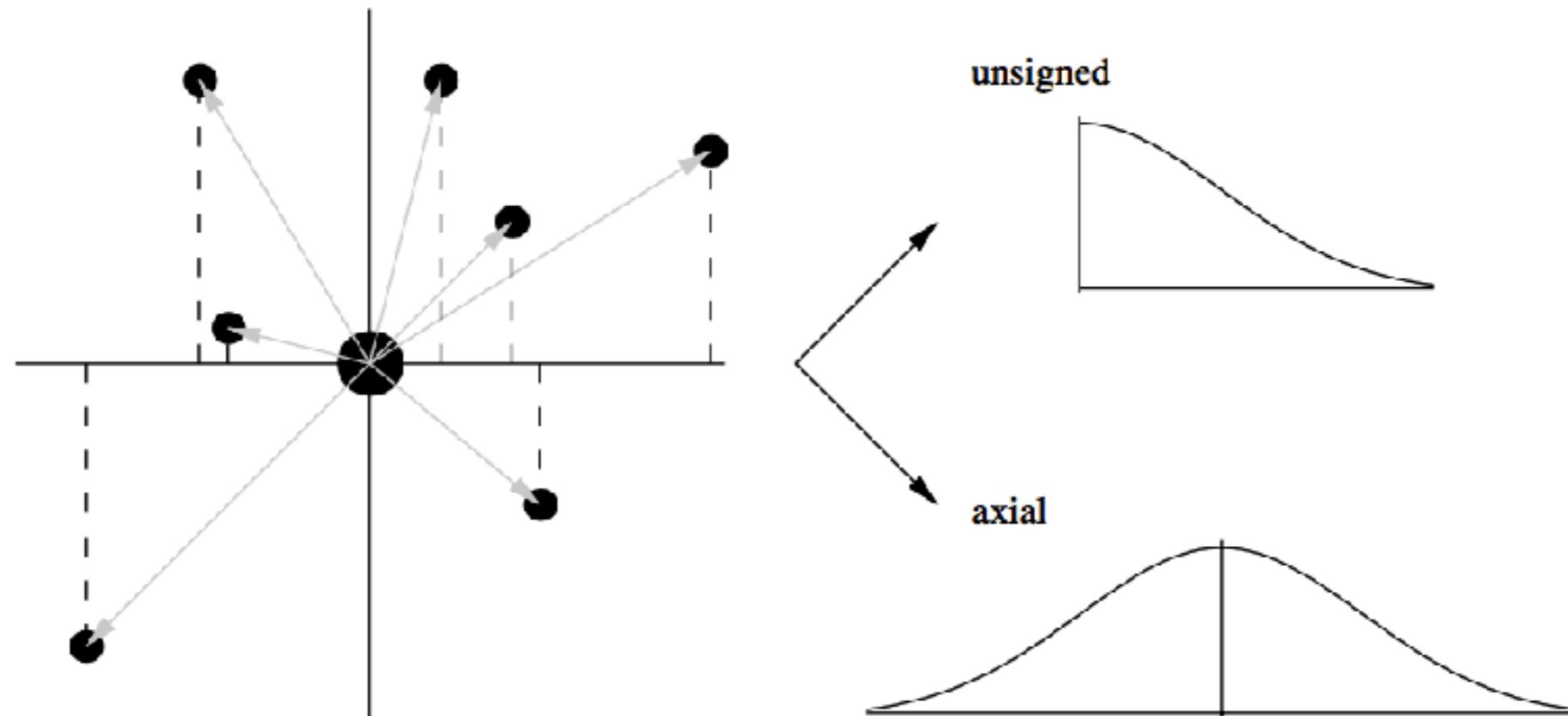
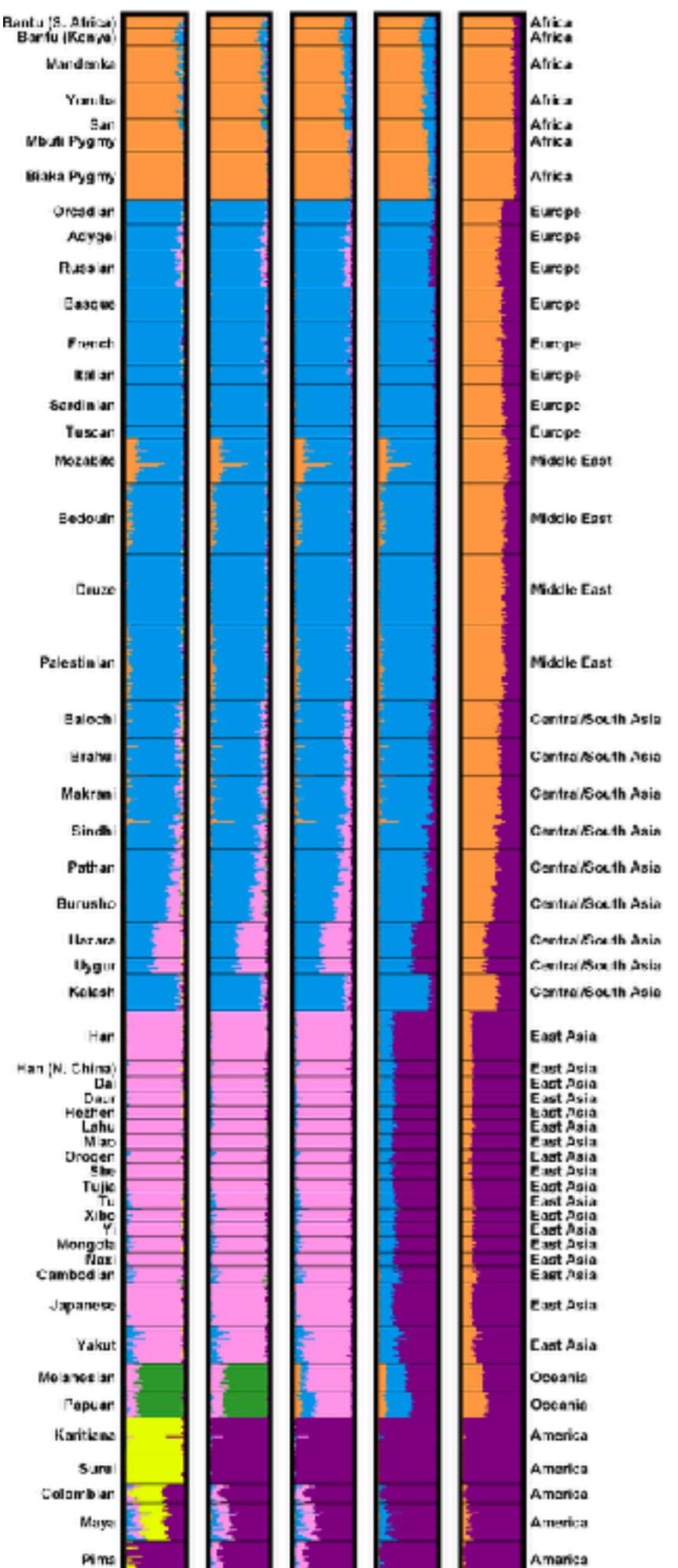


Figure 23.1 σ^2 in two dimensions. One considers the two-dimensional dispersal distances (grey arrows) between one parent (large central dot) and different offspring (or different parent – offspring pairs). The projection of these vectors on two axes yield signed axial distances on each axis. In terms of variance, σ^2 is the variance of the distribution of one such axial dispersal distance (bottom right). This is *not* the variance of the unsigned dispersal distance (top right).



clusters may be from sampling discontinuities rather than true discrete populations

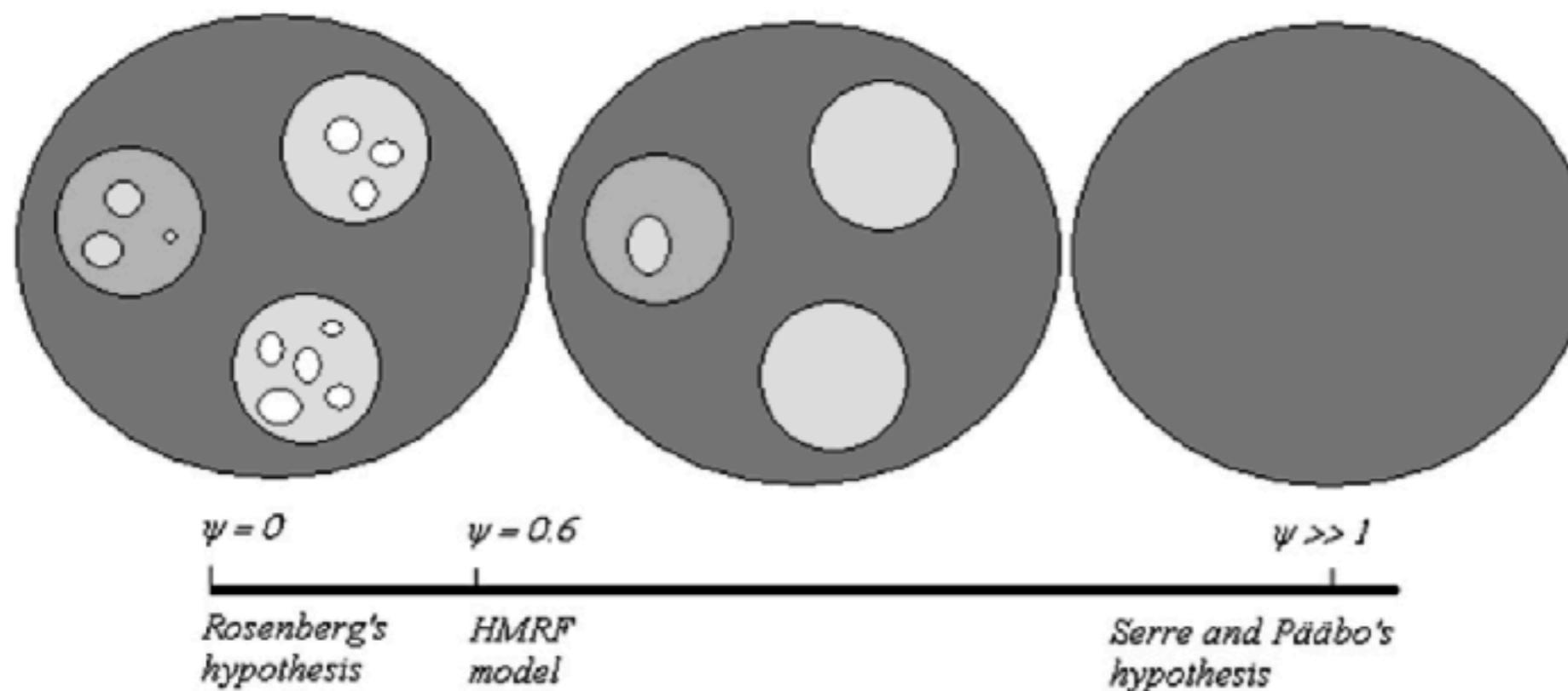
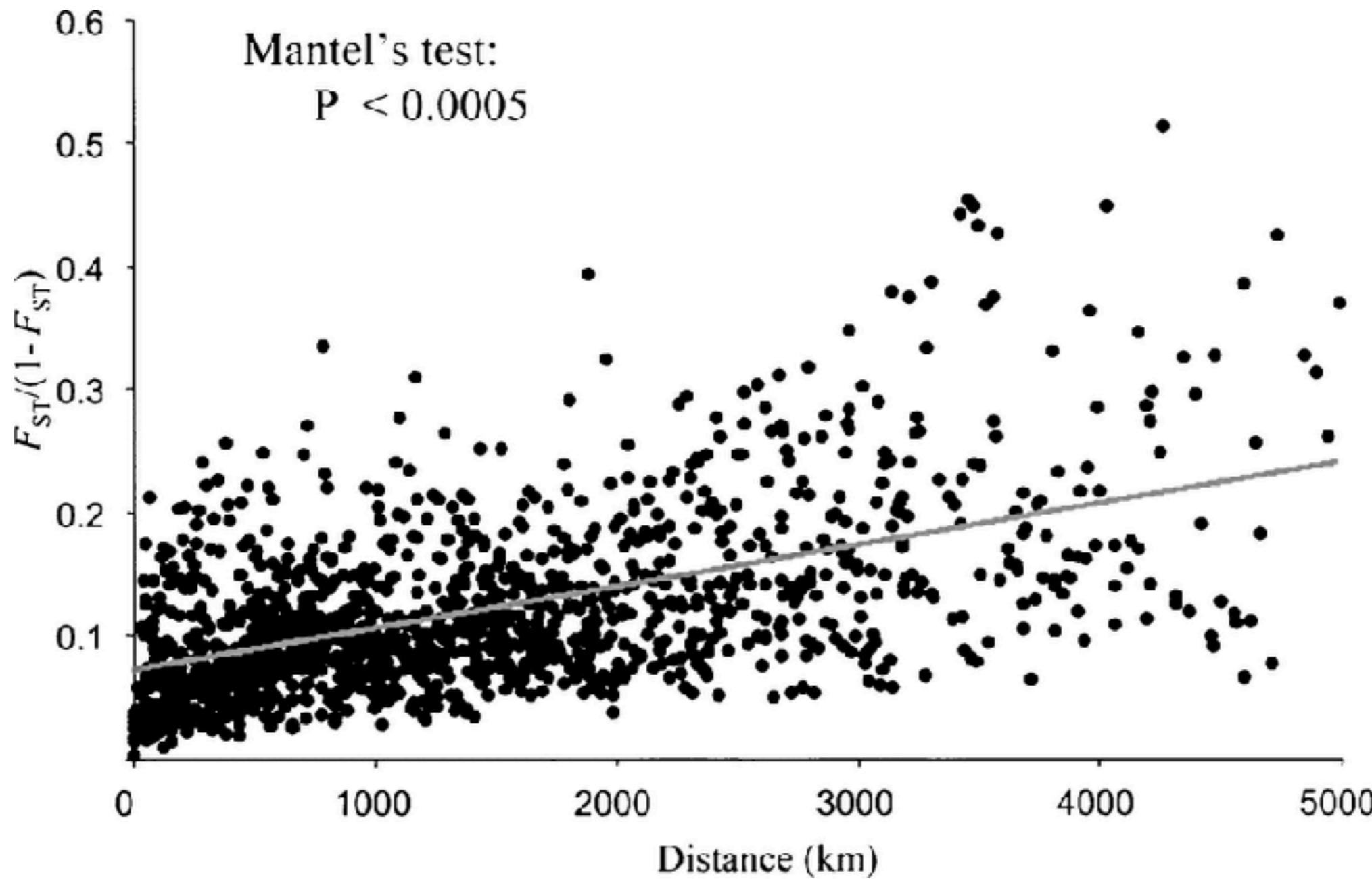


FIGURE 5.—The reconciliation illustrated. At the left of the ψ -axis, a clustering analysis does not account for the spatial continuity of allele frequencies and may detect more clusters than actually exist. At the right, the pure continuity hypothesis assumes no cluster. Here the vision is intermediate, with the main discontinuities confirmed, but some small clusters may be considered nonsignificant.

but using sampling location as prior can lesson this problem

Isolation by Distance ??



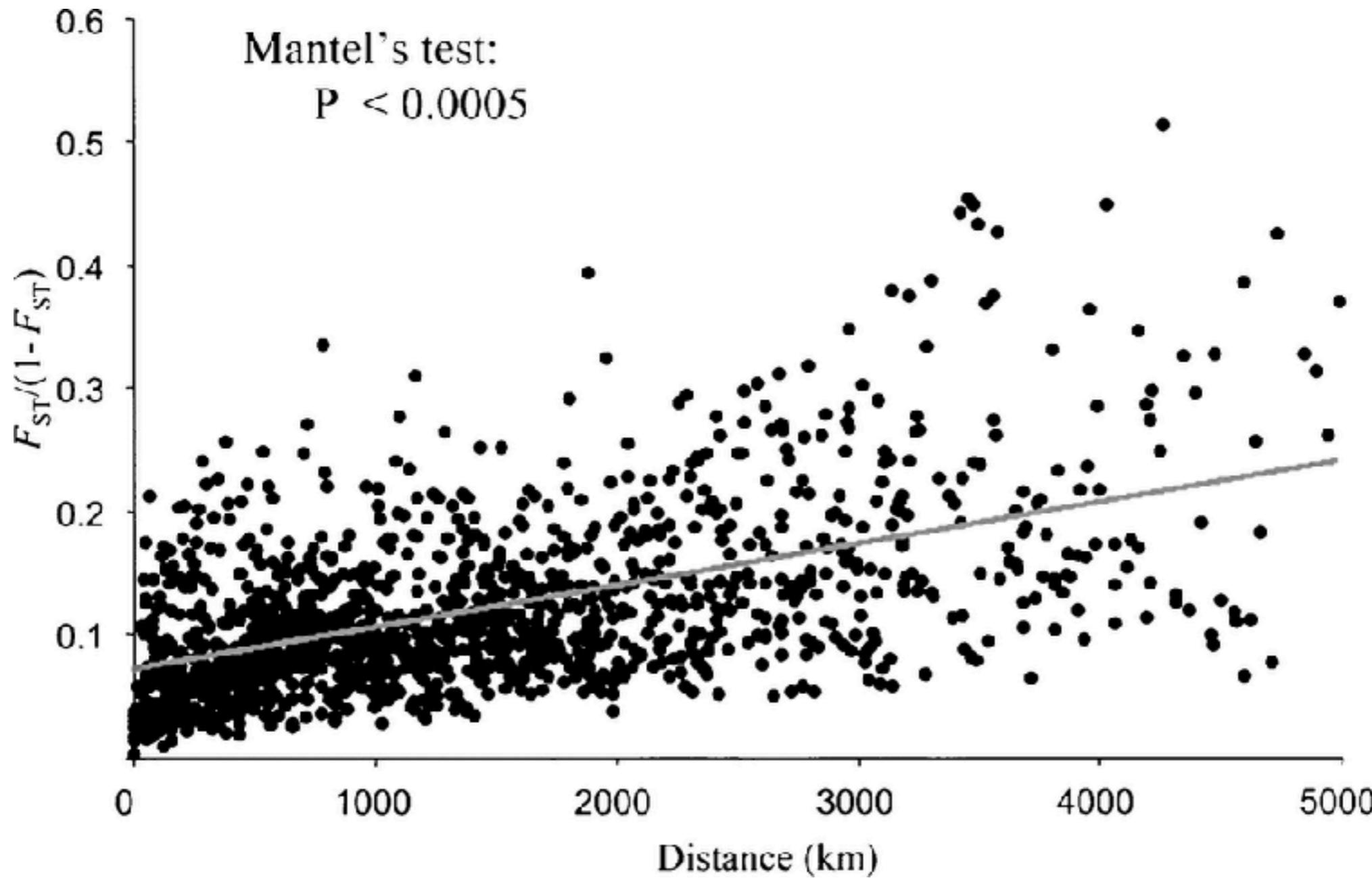
Isolation by Distance ??

**first, you have to assume drift
migration equilibrium! (more later)**



classic testing of Isolation by Distance

- Mental test (permuting locations)



Isolation by Distance ??

**first, you have to assume drift
migration equilibrium!**



Isolation by Distance ??

IBD is *only* expected if there is drift-migration equilibrium

slope = $N_e * \text{dispersal distance}$

no IBD could mean

- inadequate spatial sampling

no IBD could mean non-equilibrium

- range expansion or divergence can disrupt IBD (depends on dispersal potential)

Isolation by Distance ??

also generally

- assumes that migration is symmetrical**
- does not let you estimate how migration differs across the landscape**

But we really want to know:

- how migration differs across the landscape through time**
- how population density differs across space through time**
- how populations shift over space and time**

Principal Components Analysis (PCA)=Reduction of dimensions

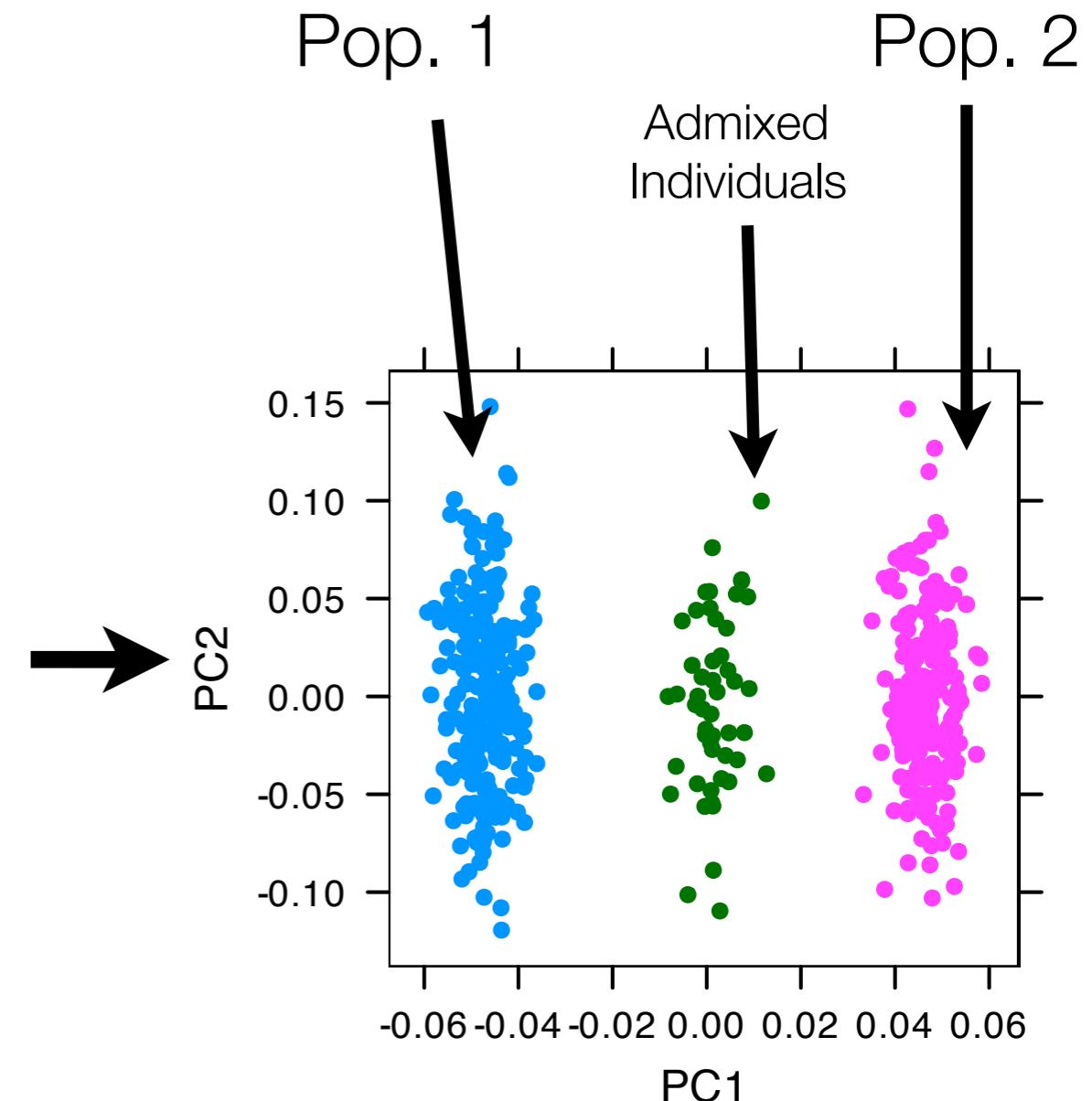
Single Nucleotide
Polymorphisms

Individuals

201101120102200122122101210
10201122020120111212101211
202120220001200021122211121
0011121202212122202122221
0020122202022121122121222
21111002021201122211112212
20102212111101221221110222
122111120212201222212012221
202110120102201222121202211
011101120201100222121212221
011000120011211222122112221
012010121201200222120112222
11101222010120222221211211
110021220212200121122210212
101021220102201222111102220
111211120112101122120102220
212021220222101122221112211
212111221211201222211211222
112100220112101222221002221
212110220112200111022001120
⋮

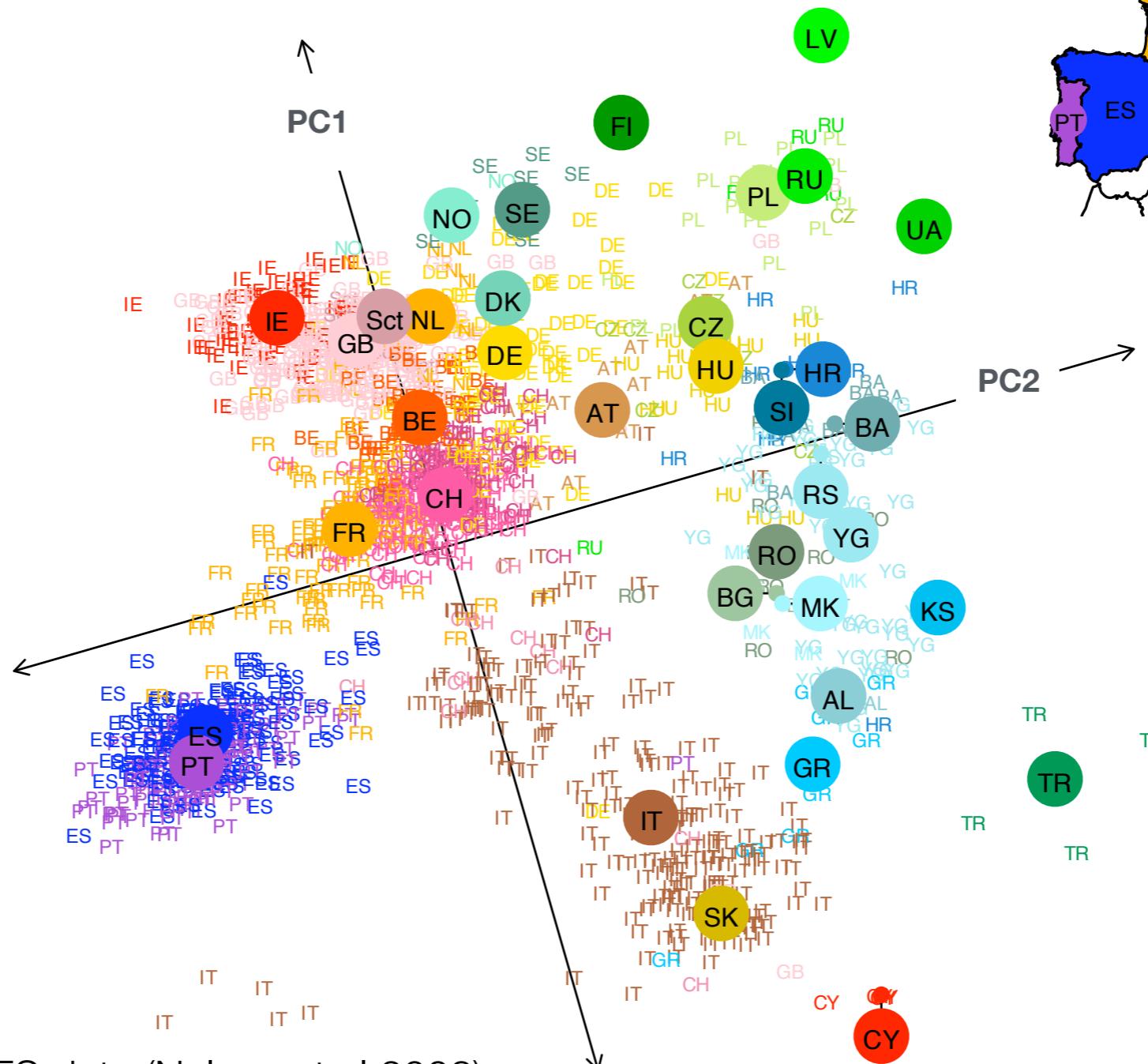
0 = AA
1 = Aa
2 = aa

	PC1	PC2
	0.055	-0.08
	0.042	0.10
	-0.052	-0.05
	0.010	0.05
	0.052	0.08
	0.043	-0.02
	-0.058	0.04
	0.020	-0.03
	-0.030	-0.05
	-0.050	0.07
	0.052	0.08
	0.043	-0.02
	-0.058	-0.04
	-0.04	0.06
	0.020	-0.02
	-0.031	0.05
	-0.053	0.08
	0.014	-0.02
	0.533	-0.04
	-0.044	0.03
⋮	⋮	⋮



For introduction see Patterson et al (PLoS Genetics 2006)

Genes mirror geography in Europe



European subset of POPRES data (Nelson et al 2008)

PC1 = 0.3% of variation ; PC2 = 0.15% of variation

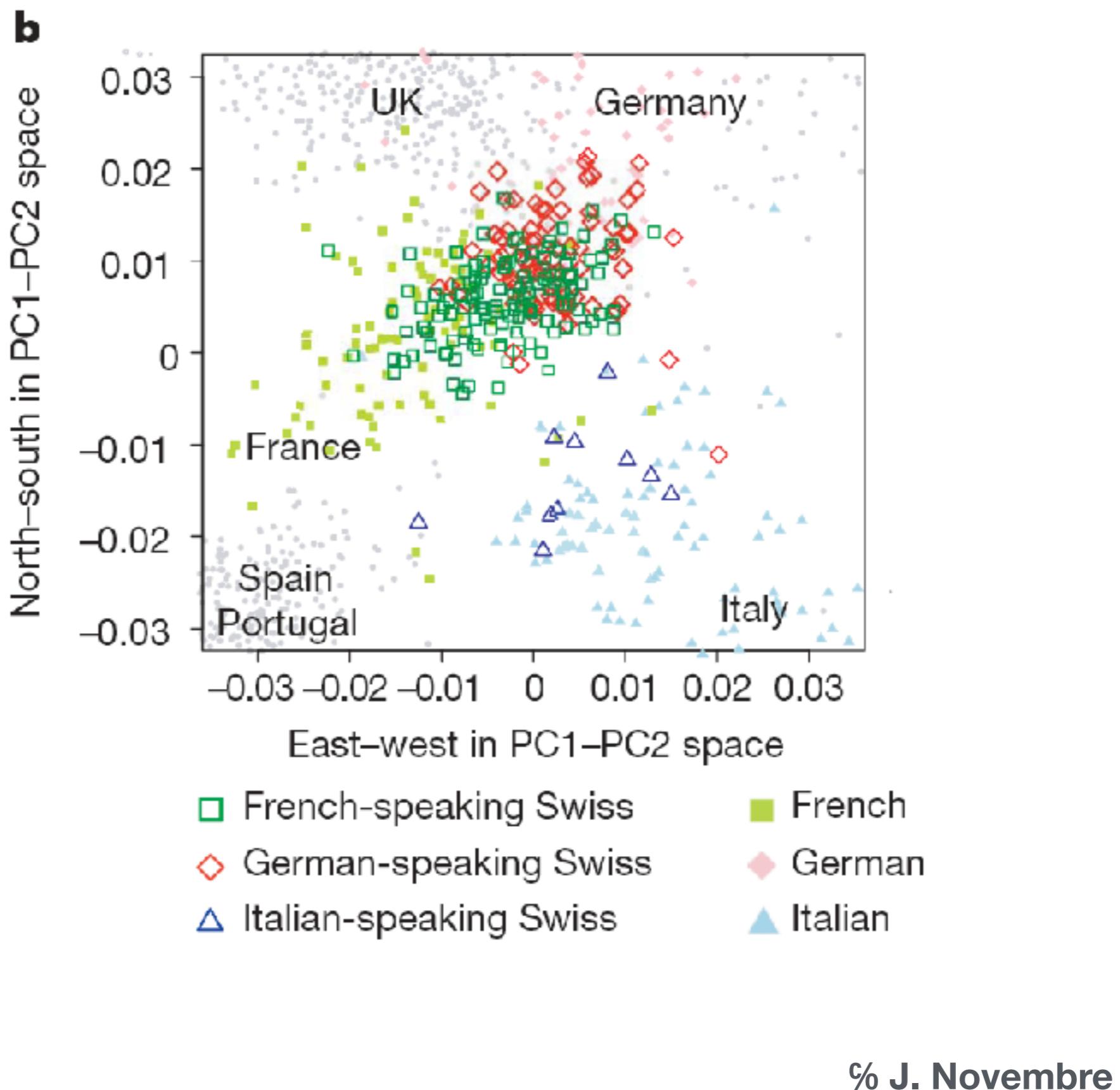
Individuals with grandparents from a single origin and SNPs not in LD

1,387 Europeans, ~138,000 SNPs

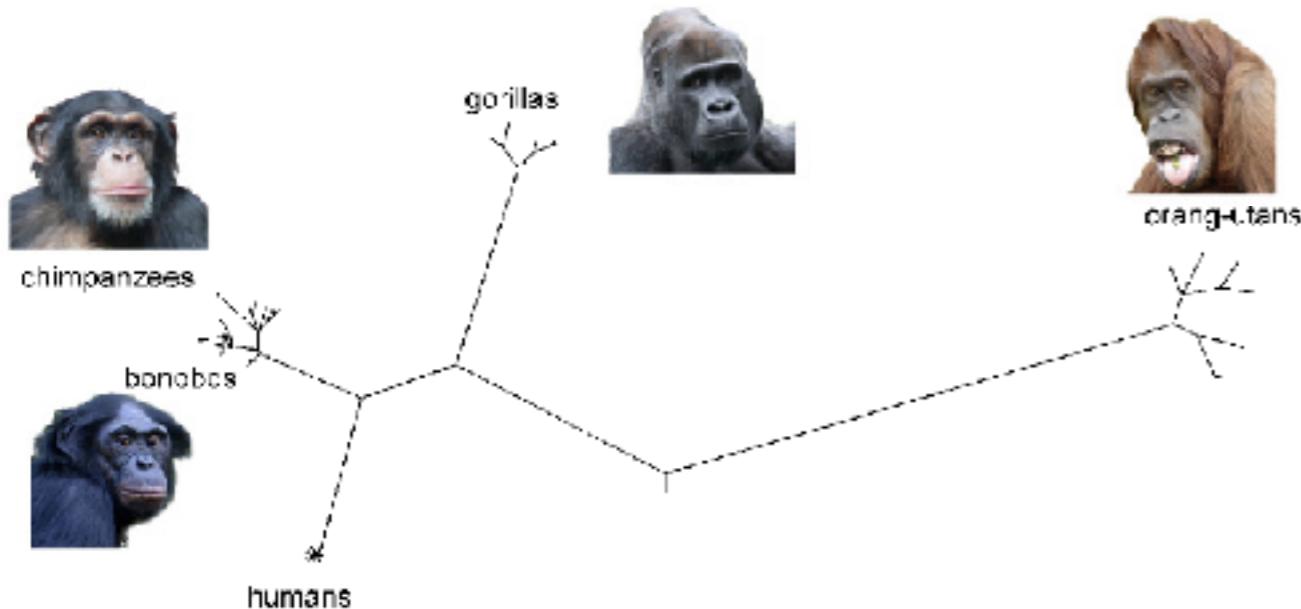
Samples principally from Lausanne & London

Labelled by grandparental ancestry

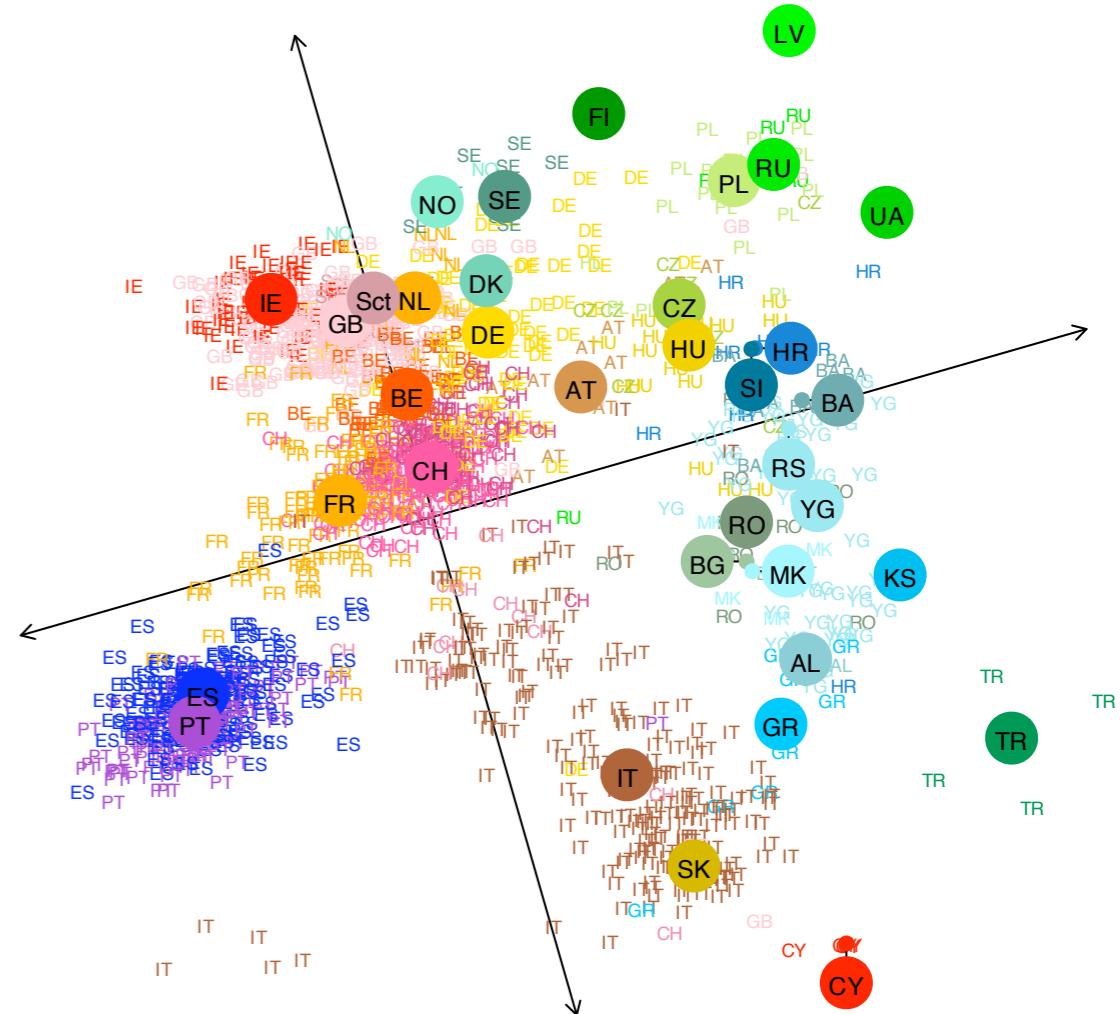
Novembre et al (Nature, 2008)



Both of these are true

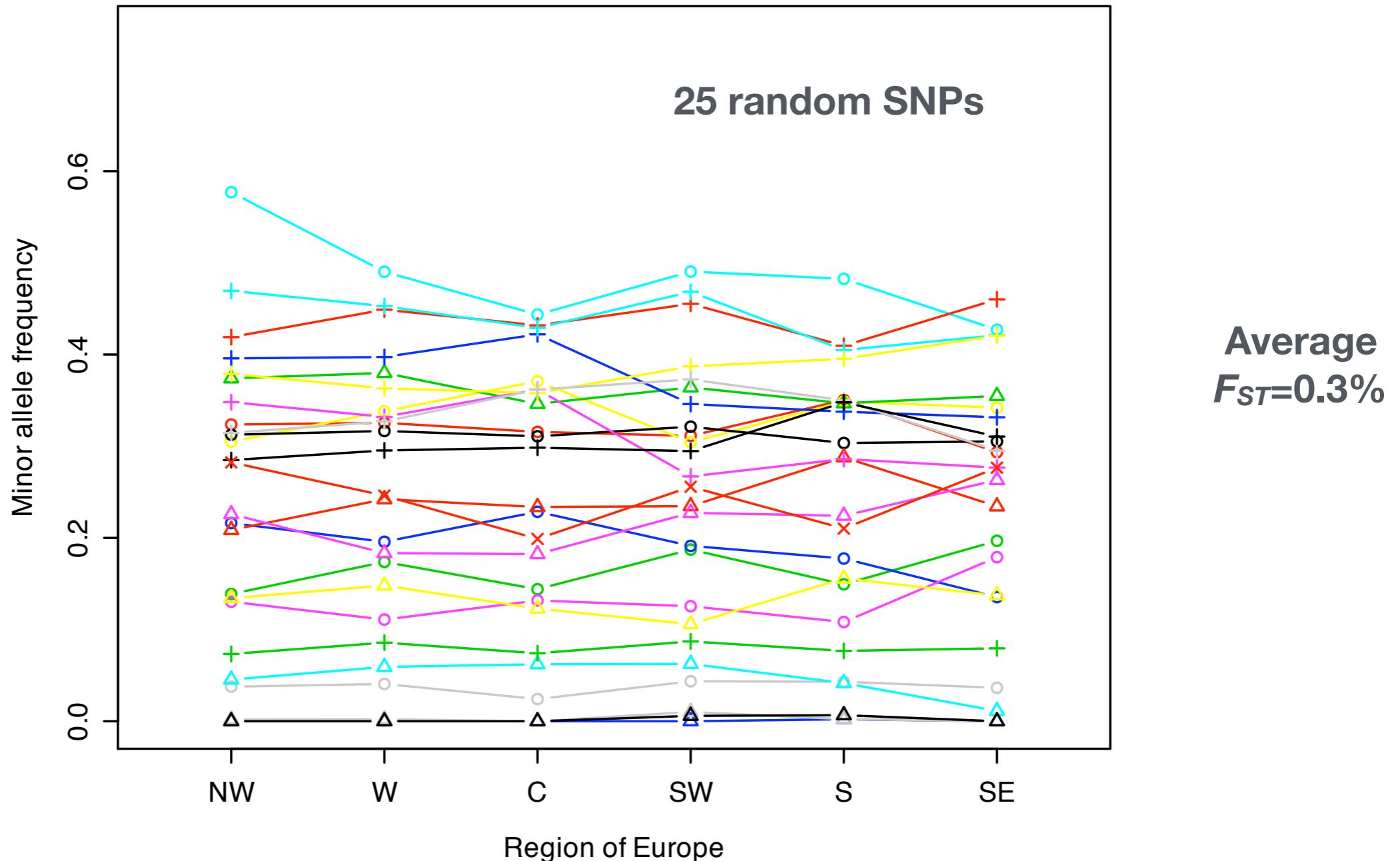


Humans comparatively closely related
Sequence divergence is low: 1 per 1000 bp



Sub-continental origins of individuals
reflected in genetics

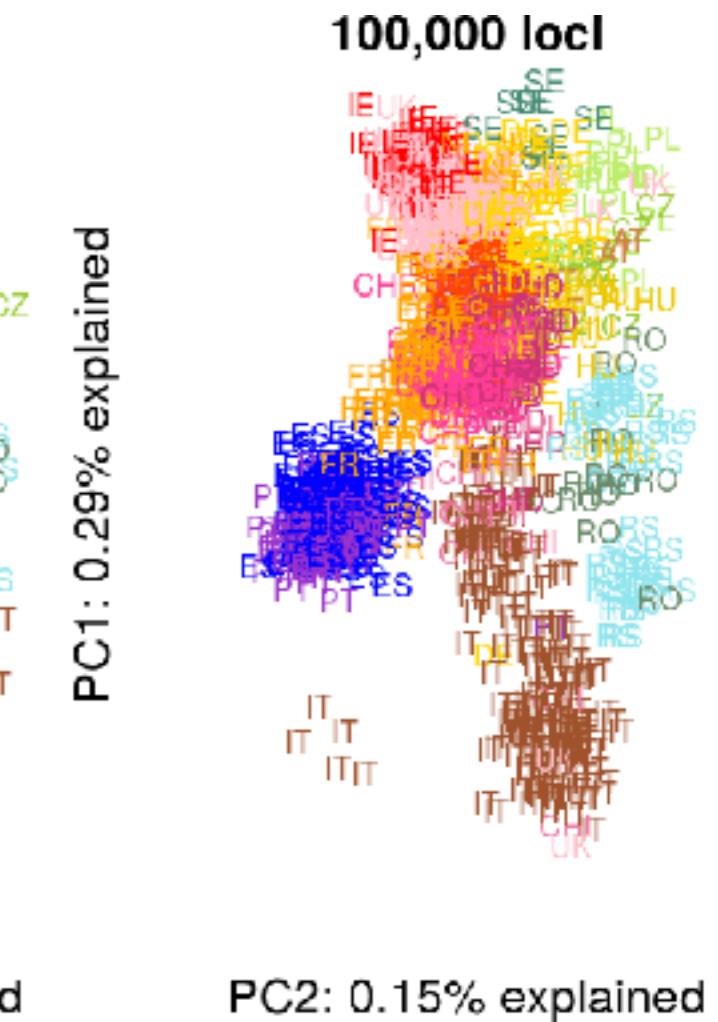
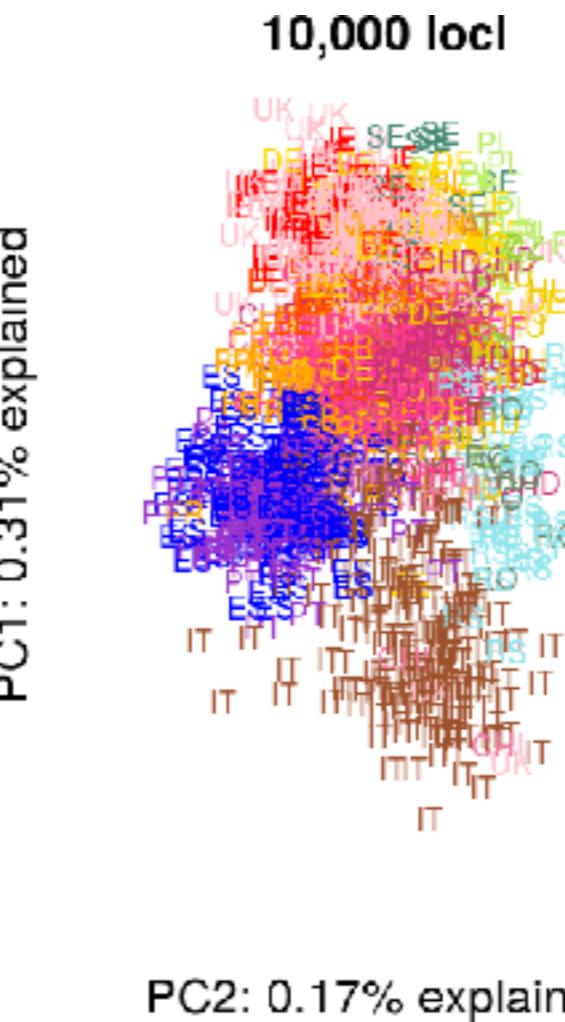
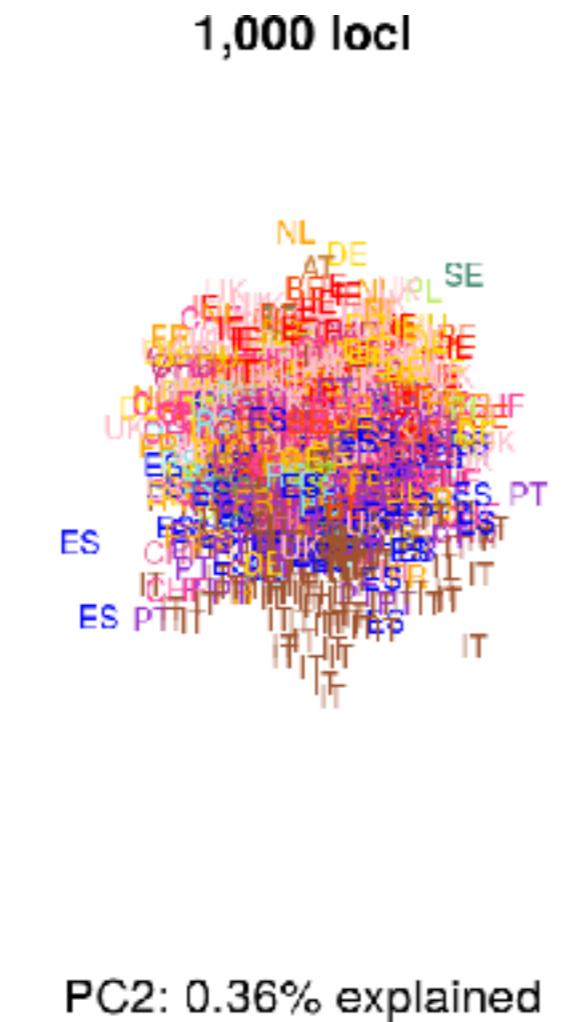
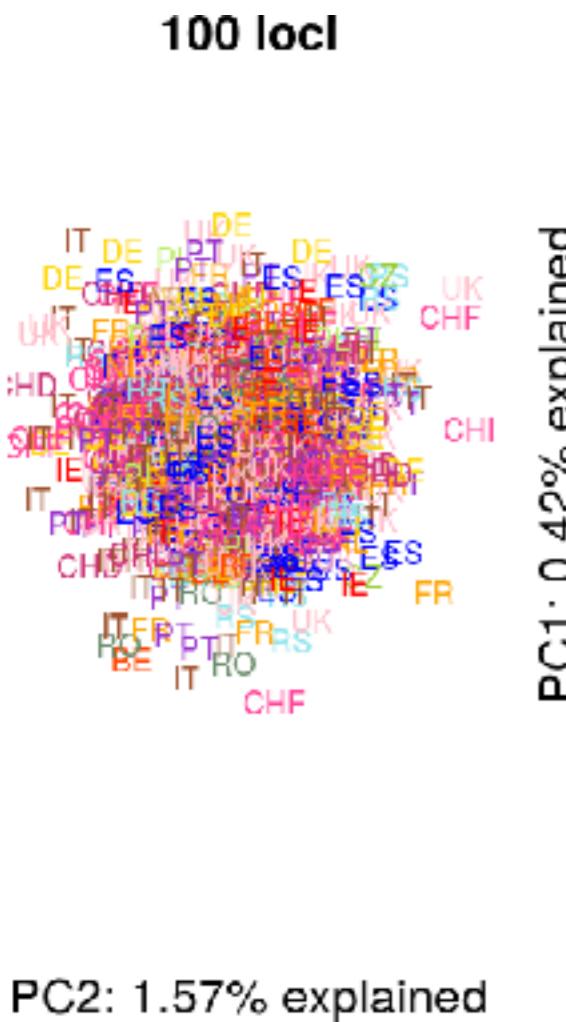
Plotting random variant frequencies across populations



Data from POPRES-Europe (Novembre et al, 2008)

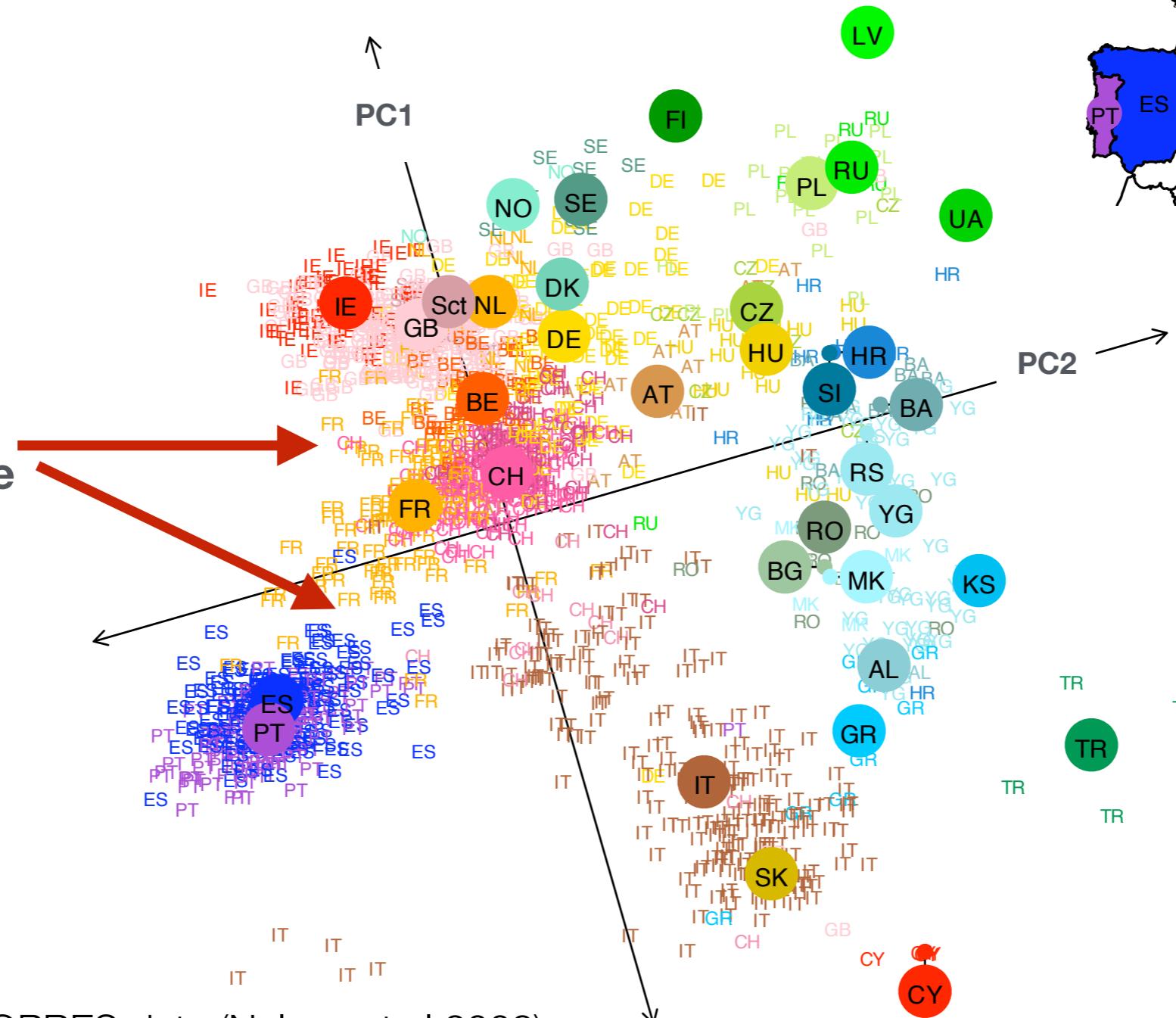
Power of big data: Methods pool weak signals across many variants

PC1: 1.63% explained



But is the mirror warped?

Should one interpret gaps in PCA space as regions of accelerated differentiation (“depressed gene flow”)?



European subset of POPRES data (Nelson et al 2008)

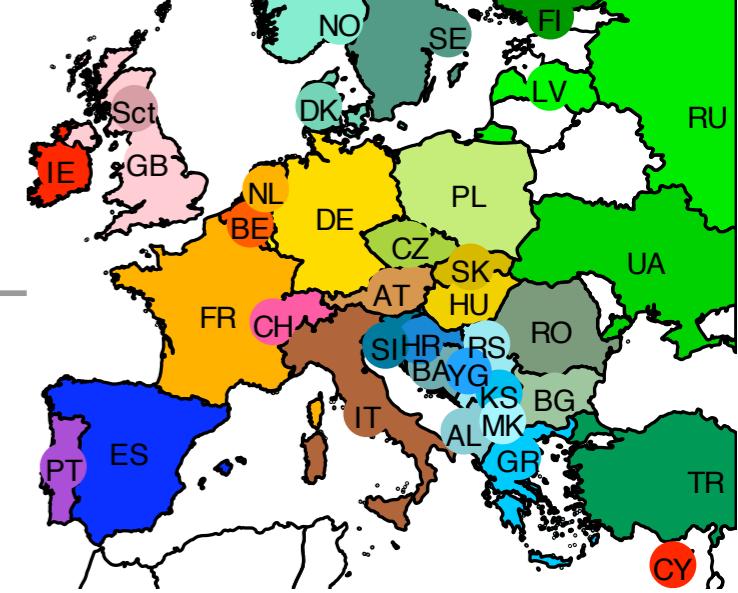
PC1 = 0.3% of variation ; PC2 = 0.15% of variation

Individuals with grandparents from a single origin and SNPs not in LD

1,387 Europeans, ~138,000 SNPs

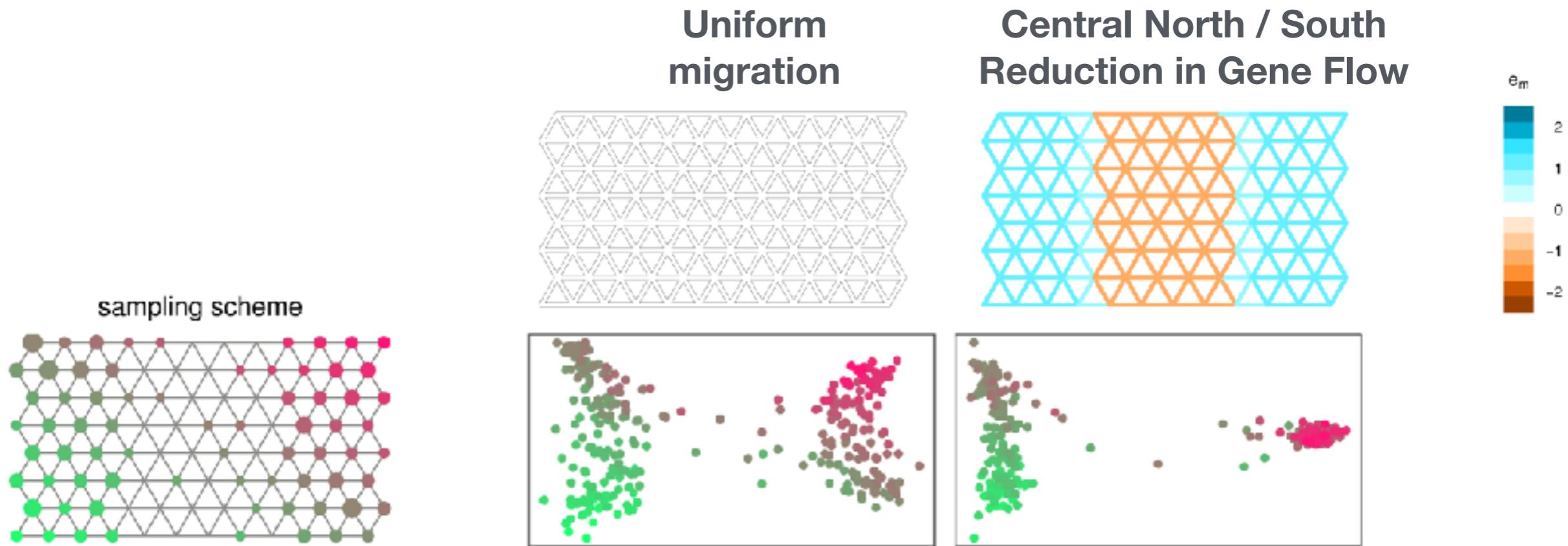
Samples principally from Lausanne & London

Labelled by grandparental ancestry

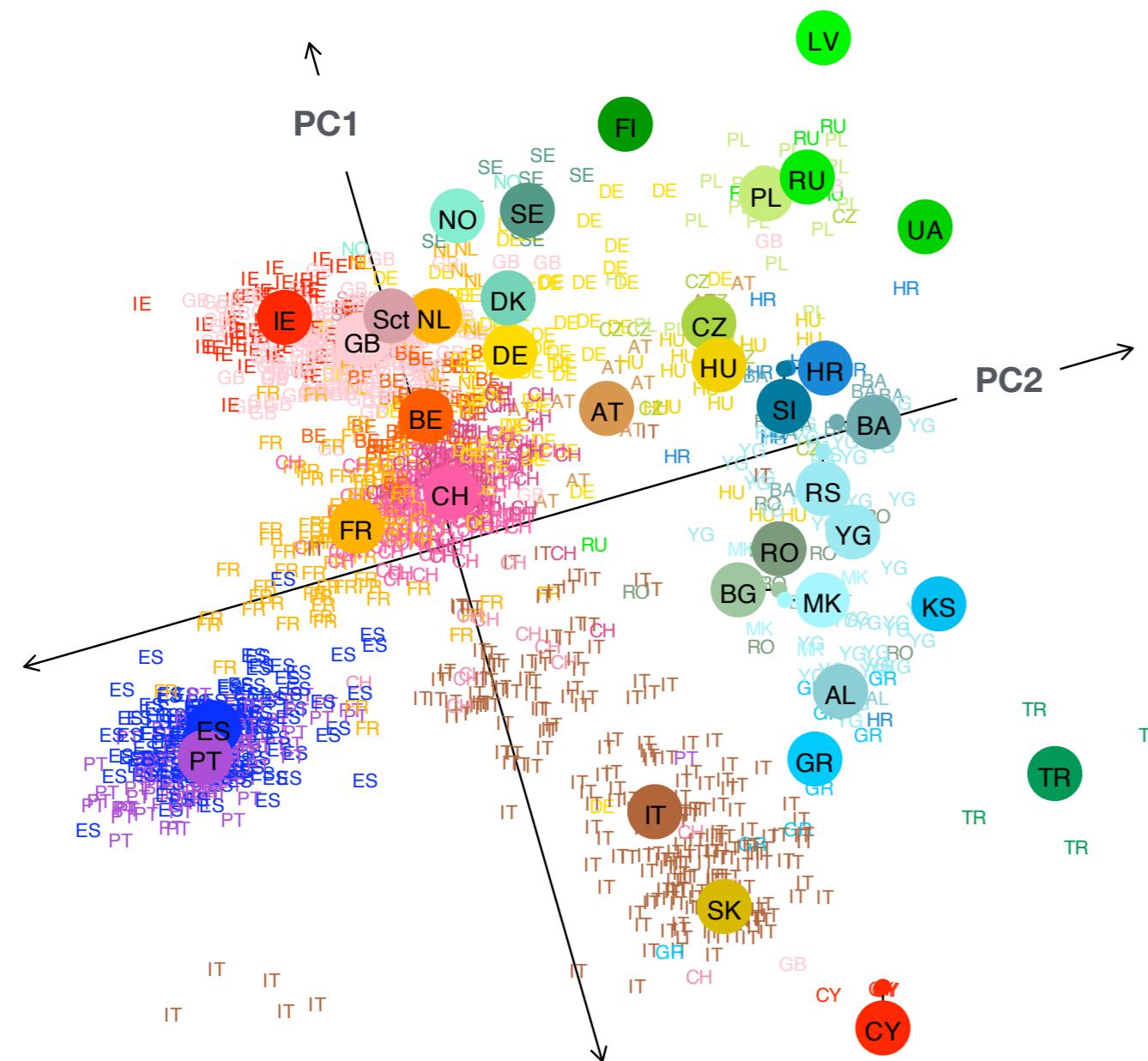


Novembre et al (Nature, 2008)

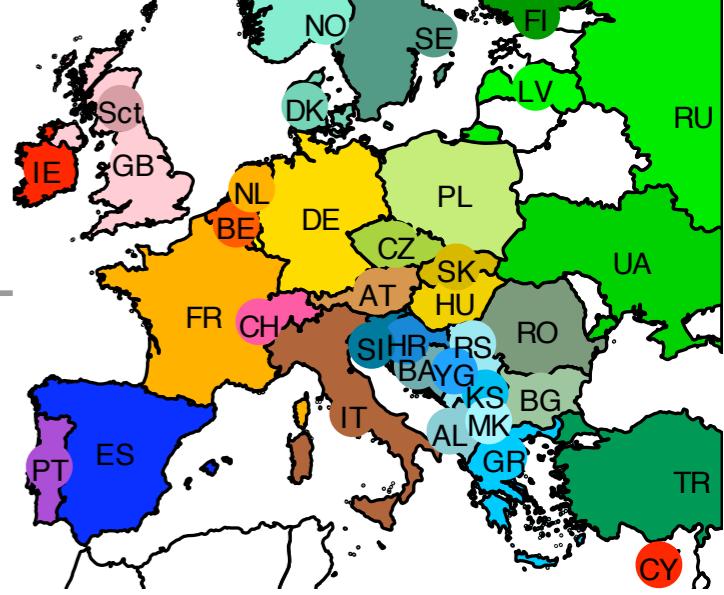
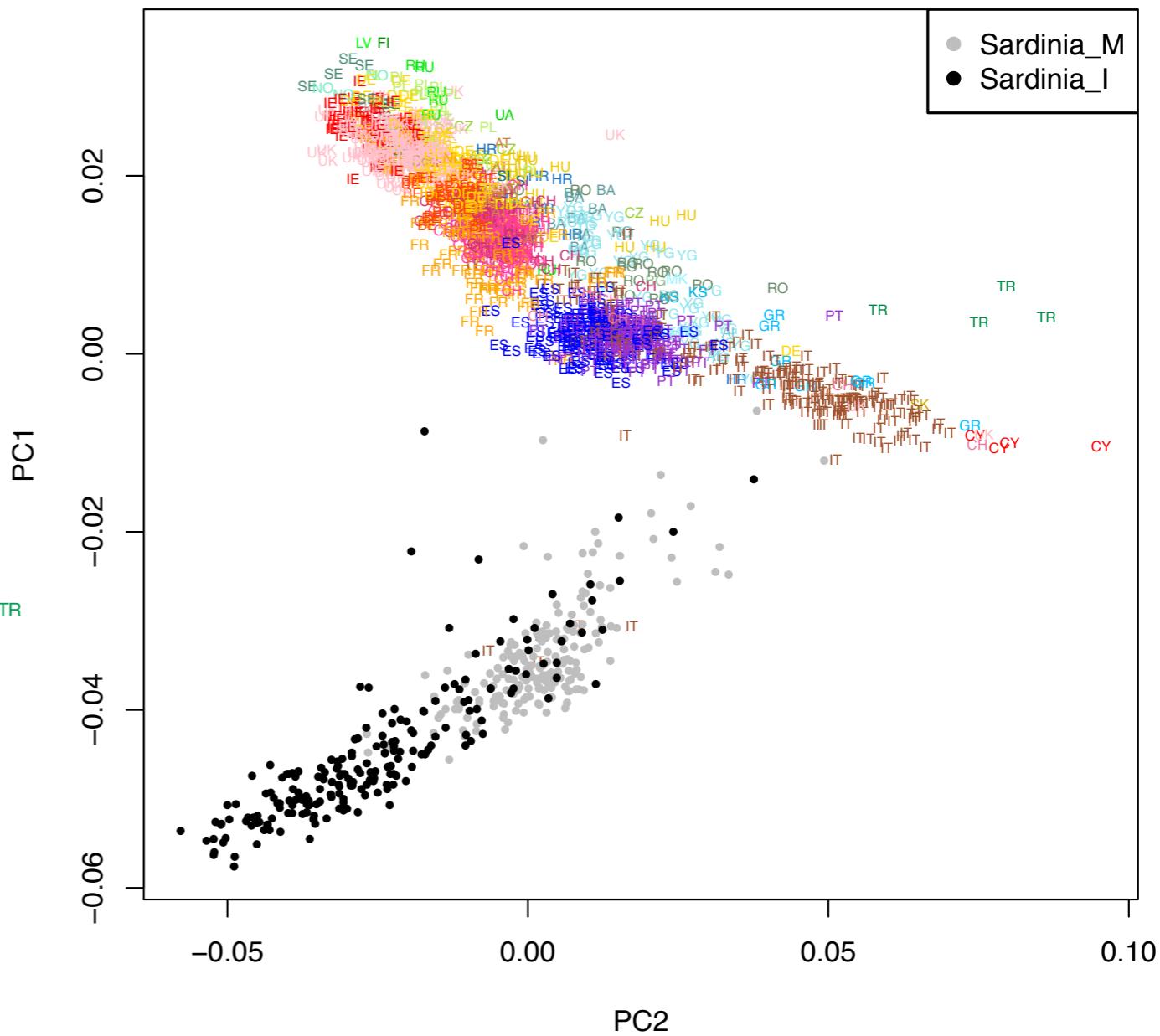
Challenge: The effect of sampling scheme on PCA



Example: Over-sampling Sardinia



1,387 Europeans, ~138,000 SNPs



Estimated effective migration surfaces (EEMS)

Uses geographical information of samples

Adjust for uneven sampling

Yet still provide insight to the genetic relationships among samples



Desi Petkova



John Novembre

Estimated effective migration surfaces (EEMS)

TECHNICAL REPORTS

Visualizing spatial population structure with estimated effective migration surfaces

Desislava Petkova^{1,2}, John Novembre³ & Matthew Stephens^{1,3}



Estimated effective migration surfaces (EEMS)

In a stepping stone model...



Migration rates & deme sizes...



imply

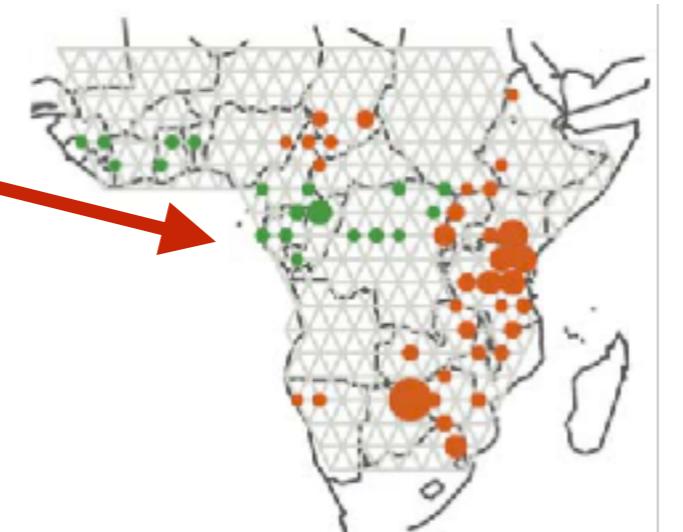
Pairwise coalescent times



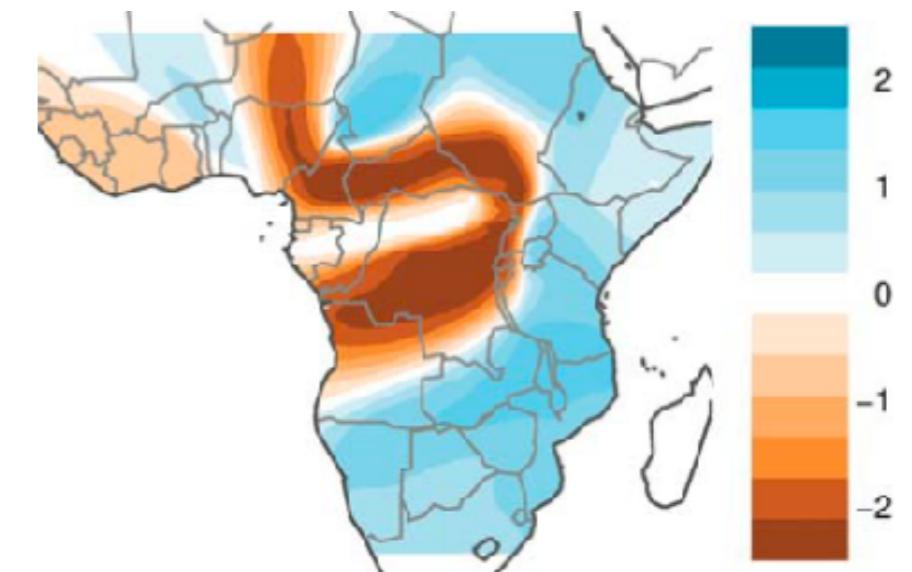
imply

Expected genetic dissimilarity

$$E[D_{ij} | \vec{T}] = \sigma_2(4T_{\delta(i)\delta(j)} - T_{\delta(i)} - T_{\delta(j)})$$



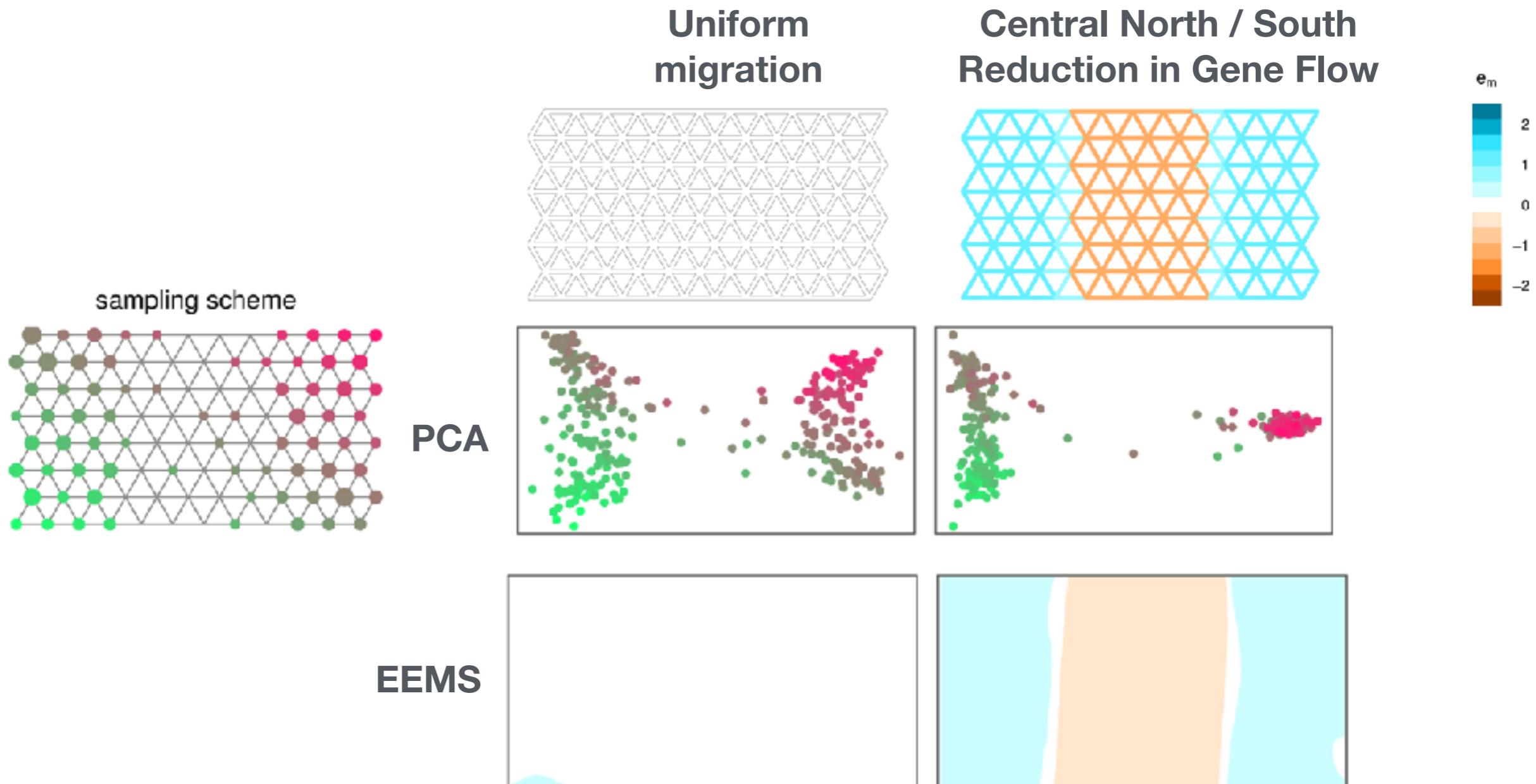
Using Markov Chain Monte Carlo we can obtain...



Smoothed maps of mean posterior migration rates

0 (white) = Average migration
>0 (blue) : Above average
<0 (brown) : Below average

low sensitivity to sampling bias



Questions:

- Given PCA on human data can look so cleanly “geographical” - will EEMS infer a “flat” map with no local changes in effective migration rate? Will it provide a better representation than PCA?

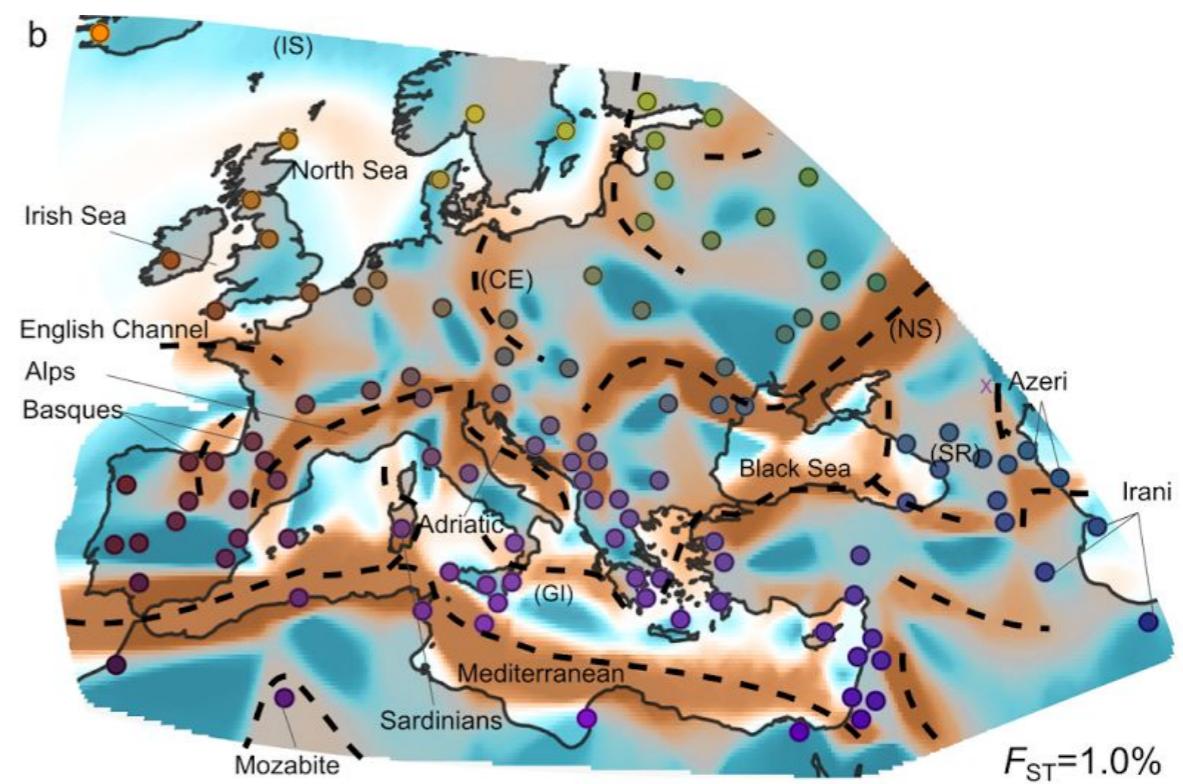


Questions:

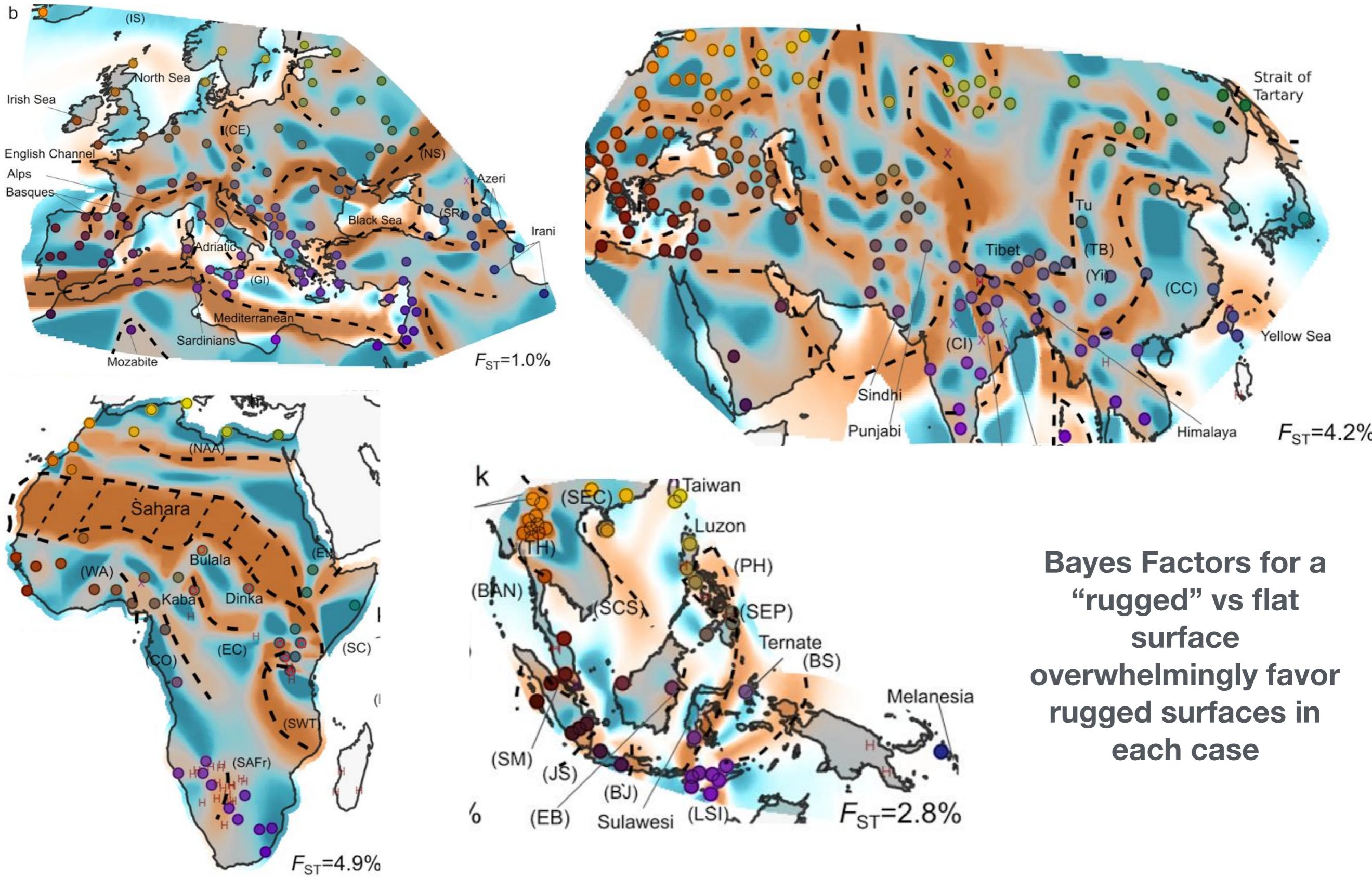
- Given PCA on human data can look so cleanly “geographical” - will EEMS infer a “flat” map with no local changes in effective migration rate? Will it provide a better representation than PCA?



EEMS



EEMS surfaces for human data are “rugged”

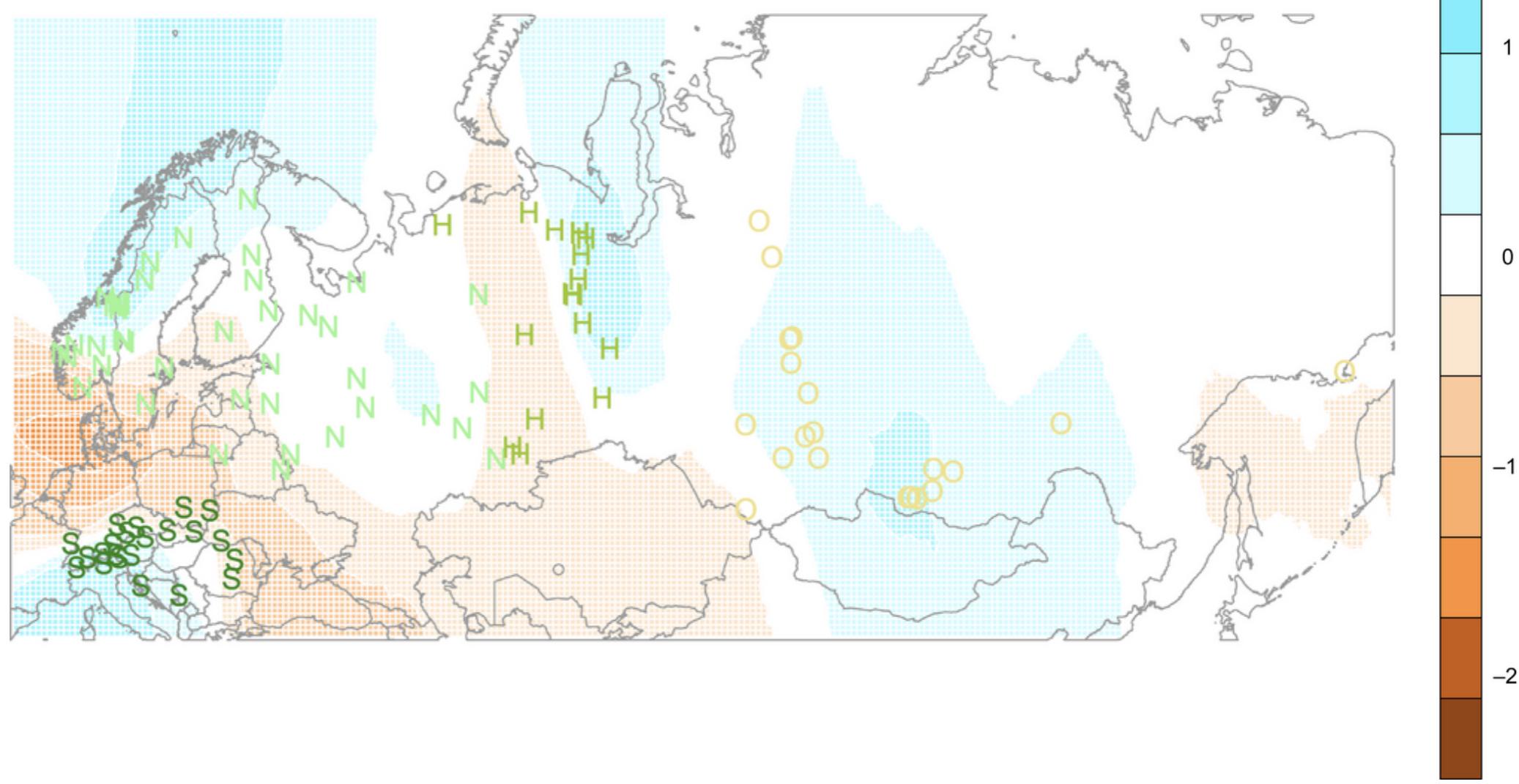


The extent and meaning of hybridization and introgression between Siberian spruce (*Picea abies*) and Norway spruce (*Picea sitchensis*): cryptic refugia as stepping stones to the west?

YOSHIAKI TSUDA,^{*†} JUN CHEN,^{*†} MICHAEL STOCKS,^{*†} THOMAS KÄLLMAN,^{*} JØRN HENRIK SØNSTEBØ,[‡] LAURA PARDOCCI,^{*} VLADIMÍR SEMERIKOV,[§] CHRISTOPH SPERISEN,[¶] DMITRY POLITOV,^{**} TUINA RUNKAINEN,^{††} MINNA VALIRANTA,^{††} GIOVANNI GIUSEPPE VENDRAMIN,[†] MARI METTE TOLLEFSRUD[‡] and MARTIN LASCOUX^{*}



South (S)
North (N)
Hybrid (H)
Obvata (O)

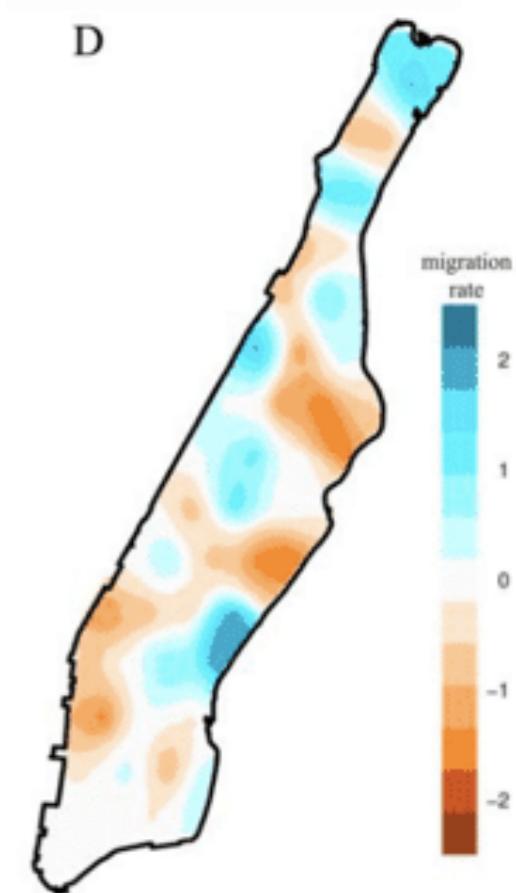
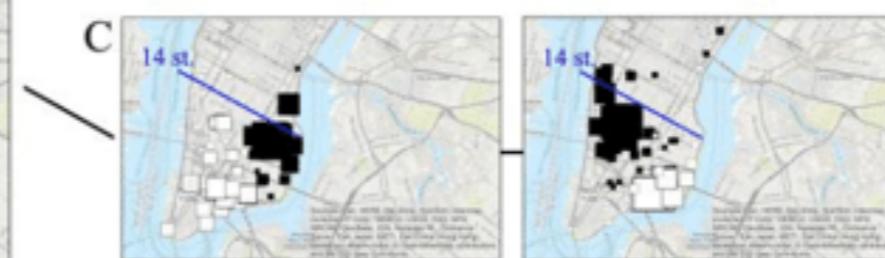
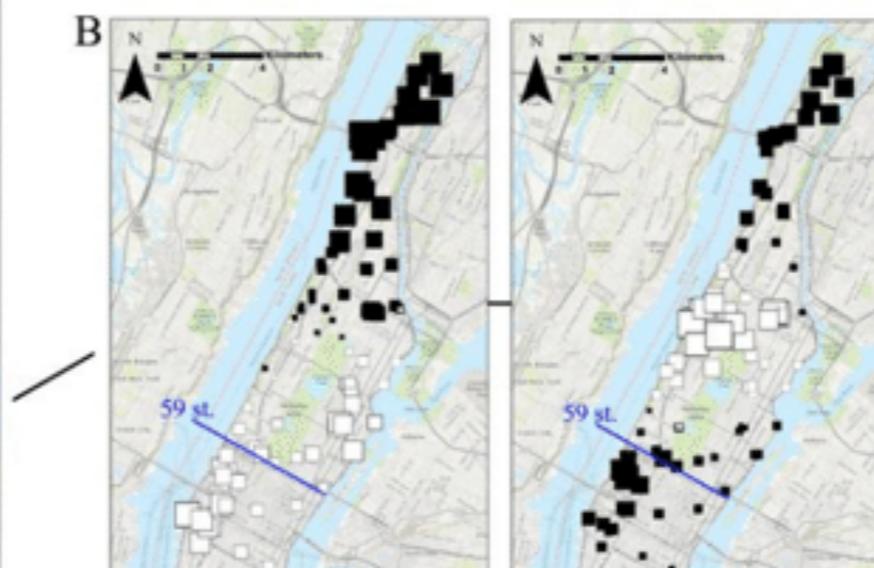
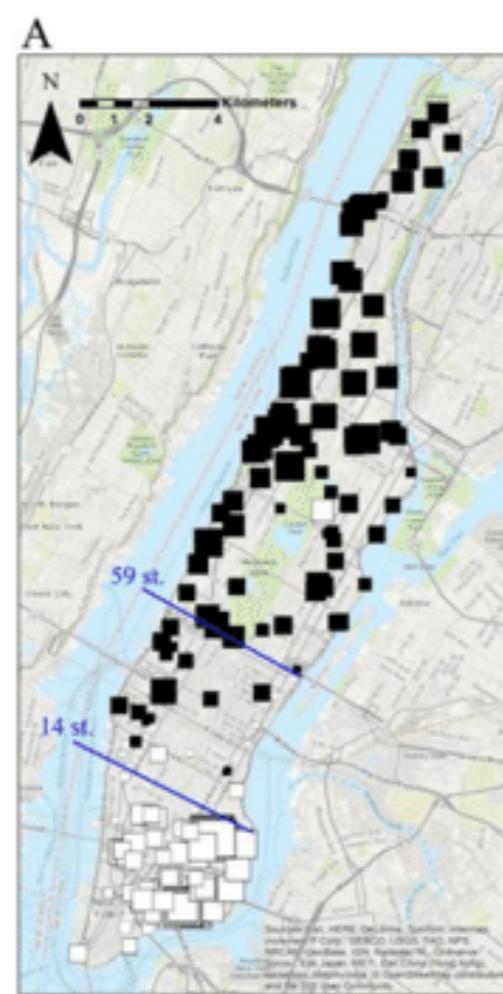
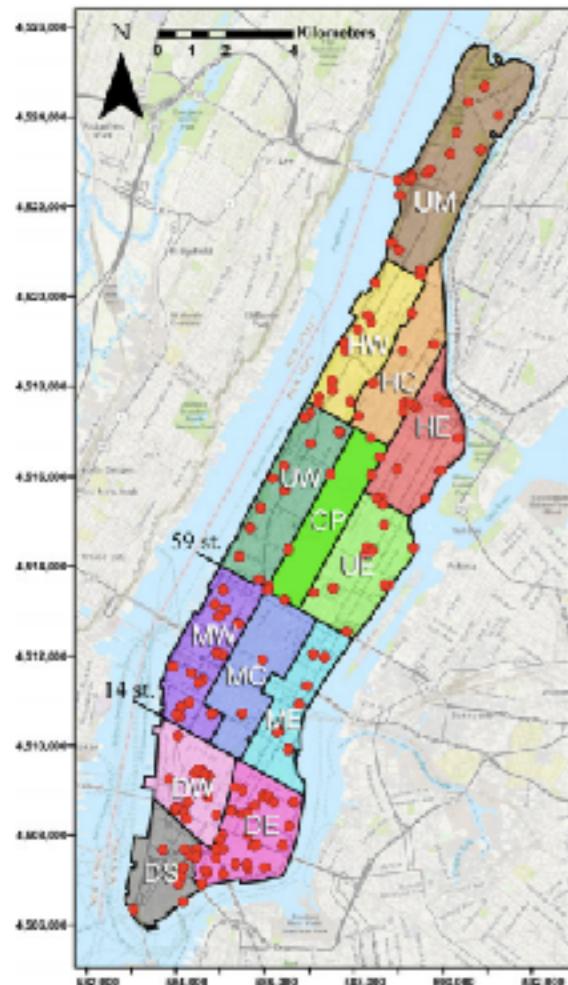


ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Spatial population genomics of the brown rat (*Rattus norvegicus*) in New York City

Matthew Combs¹  | Emily E. Puckett¹ | Jonathan Richardson² | Destiny Mims¹ |
Jason Munshi-South¹ 



Still, there is the assumption of migration being constant through time

Still, there is the assumption of migration being constant through time



BROWSE

PUBLISH

ABOUT

OPEN ACCESS



PEER-REVIEWED

RESEARCH ARTICLE

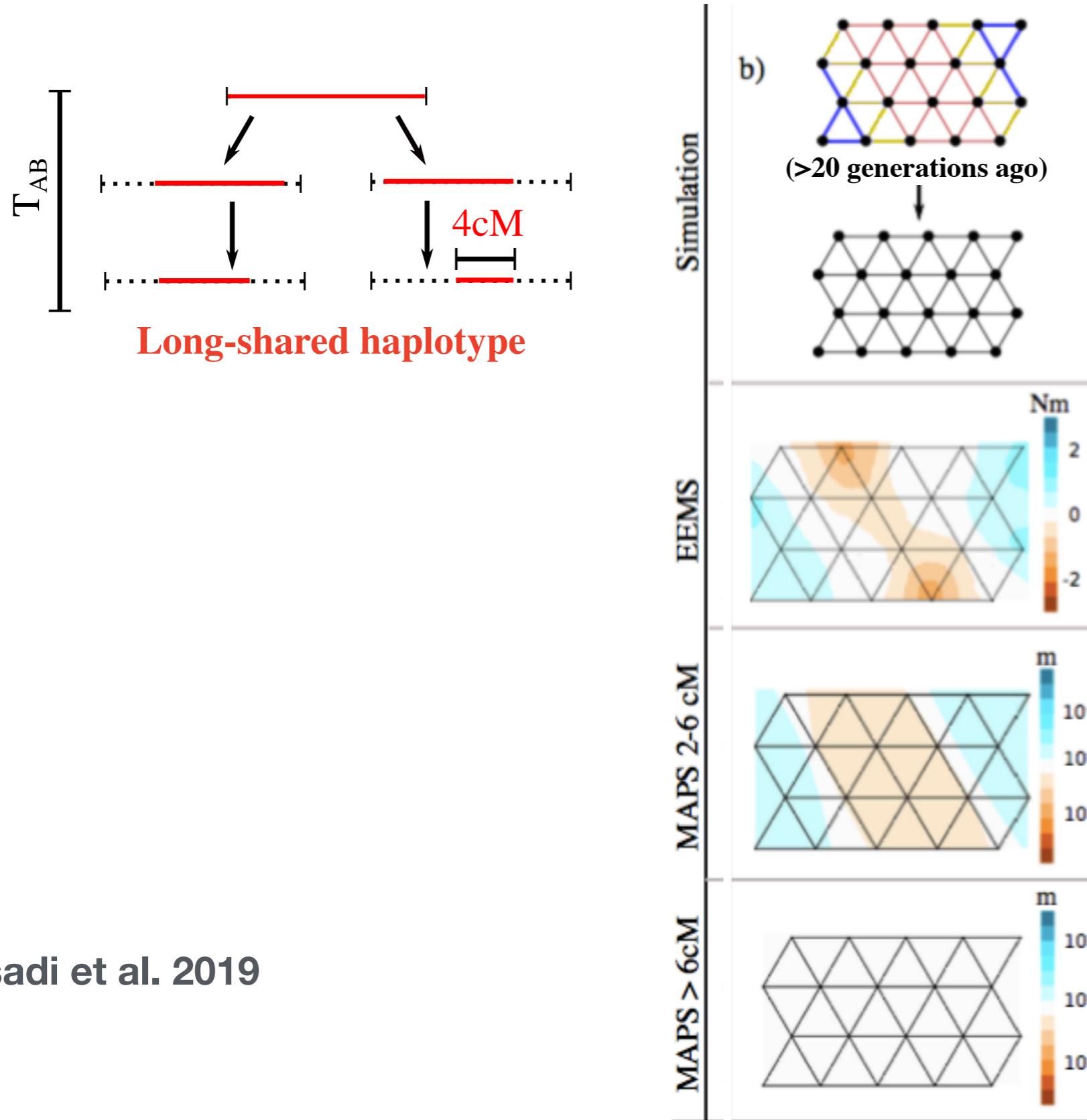
Estimating recent migration and population-size surfaces

Hussein Al-Asadi , Desislava Petkova, Matthew Stephens , John Novembre

Version 1 Uncorrected Proof

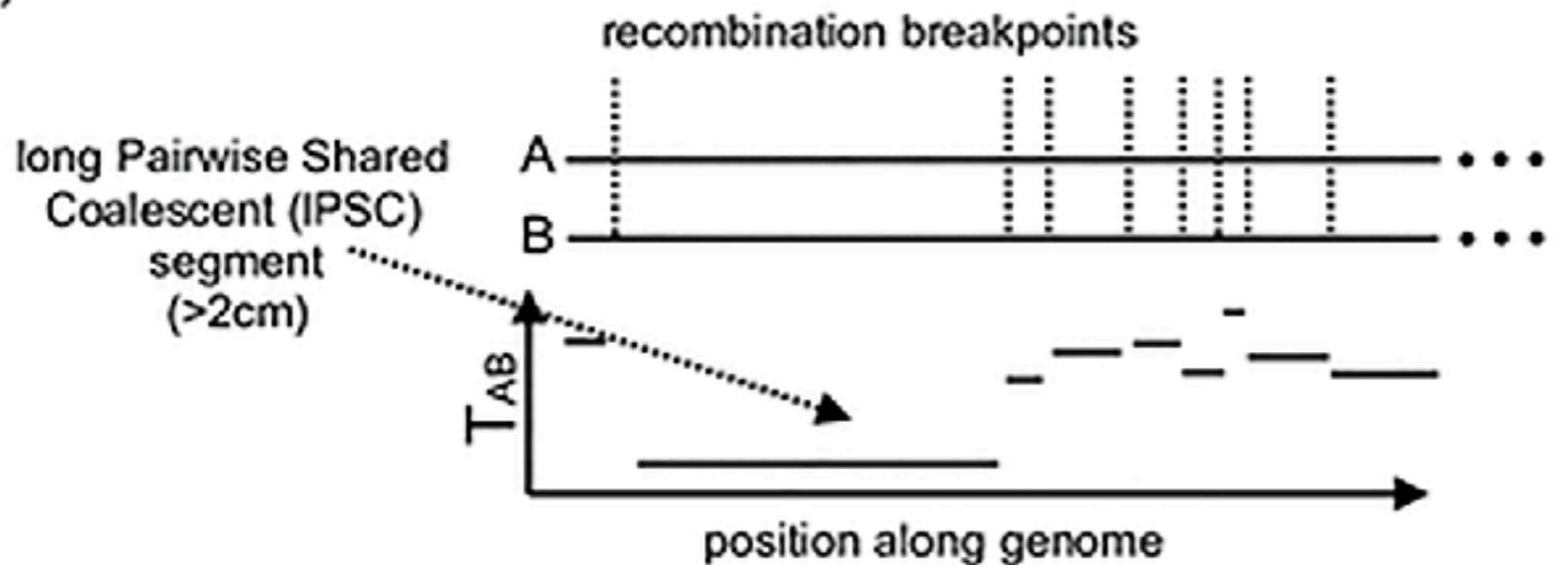
Published: January 14, 2019 • <https://doi.org/10.1371/journal.pgen.1007908>

MAPS: EEMS for long-shared haplotypes (“IBD tracts”)



Additionally:
Using haplotypic
data allows one to
untangle N and m

(a)



(b)

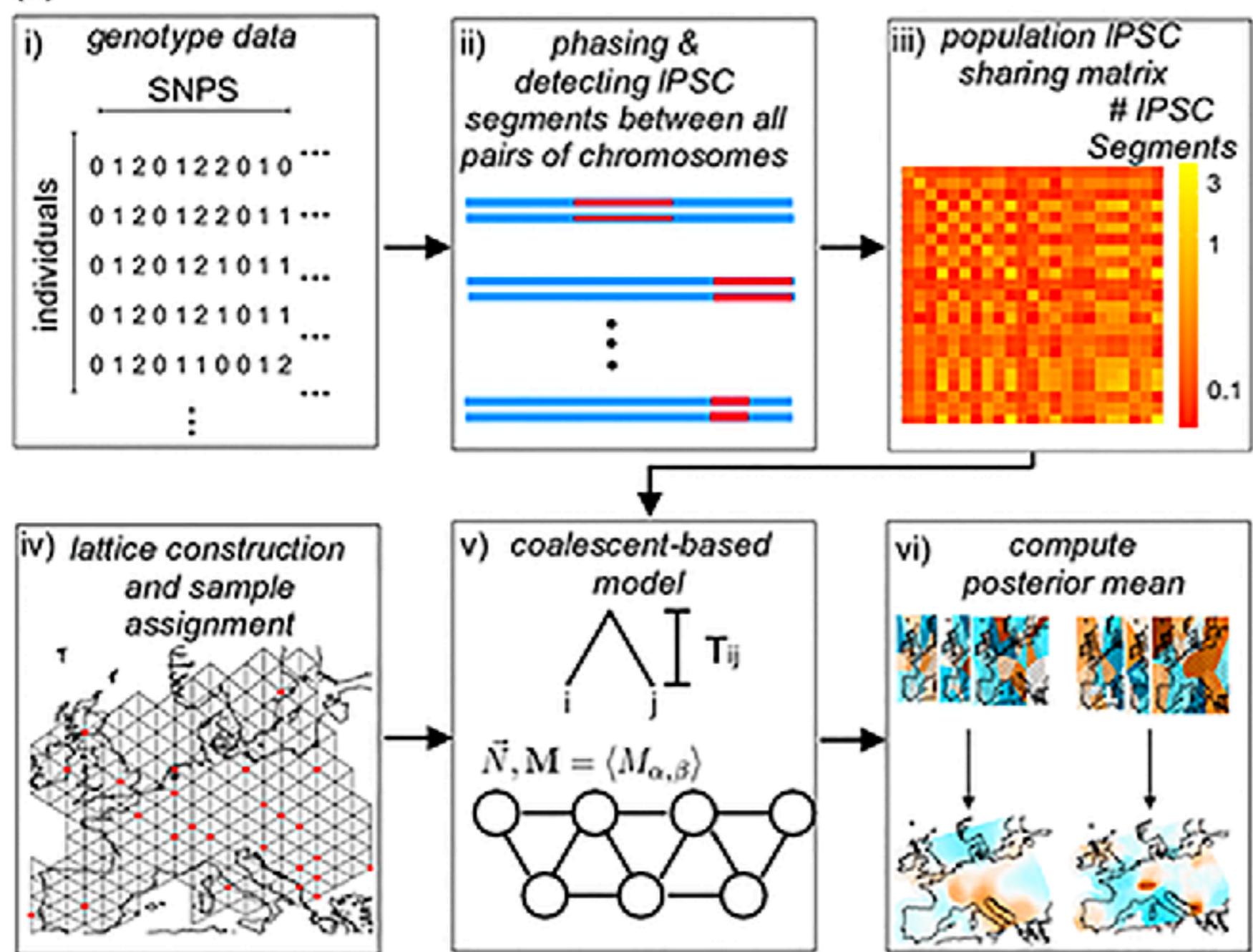


Fig 1. Schematic overview of MAPS. (a) Coalescent times between a pair of haplotypes (A and B) will vary across the genome in discrete segments bordered by recombination breakpoints. On average, longer segments represent shorter pairwise coalescent times (TAB) (b) Flow diagram of MAPS. i) We start with a matrix of called genotypes; ii) IPSC segments between all pairs of chromosomes across the genome are identified from the data using external methods (such as BEAGLE, [27]); iii) IPSC segments between pairs of individuals are aggregated at the levels of pairs of populations; iv) A grid is constructed and individuals are assigned to the most nearby node; v) The probability of the PSC sharing matrix can be computed under a stepping-stone model where each node represents a population and each edge represents symmetric migration; vi) We use an MCMC scheme to sample from the posterior distribution of migration rates and population sizes. The final MAPS output is the mean over these posterior samples, and the averaged rates can be transformed to units of dispersal rate and population density. The diagram does not show a bootstrapping step used to estimate likelihood weights to account for correlations between IPSC segments, see Eq (6) in Methods.

Al-asadi et al. 2019

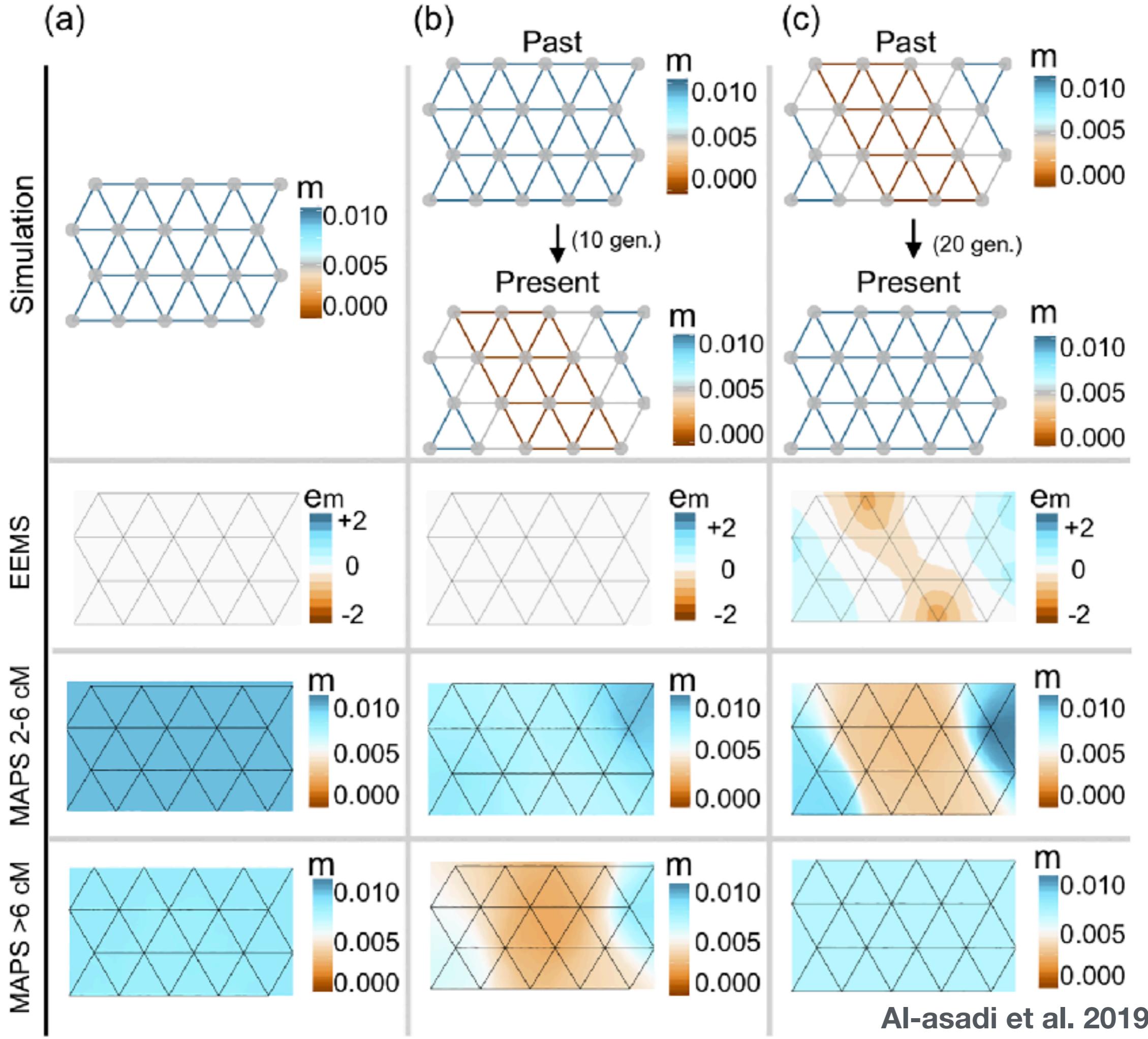
differences between MAPS and EEMS

MAPS likelihood is based on IPSC sharing, rather than a simple average genetic distance across markers.

MAPS can estimate absolute values for the parameters M and effective pop density, whereas EEMS can estimate only “effective” parameters

MAPS needs a recombination map, which acts as an independent clock to calibrate the decay of PSC segments.

MAPS uses a coalescent model, whereas EEMS uses a resistance distance approximation



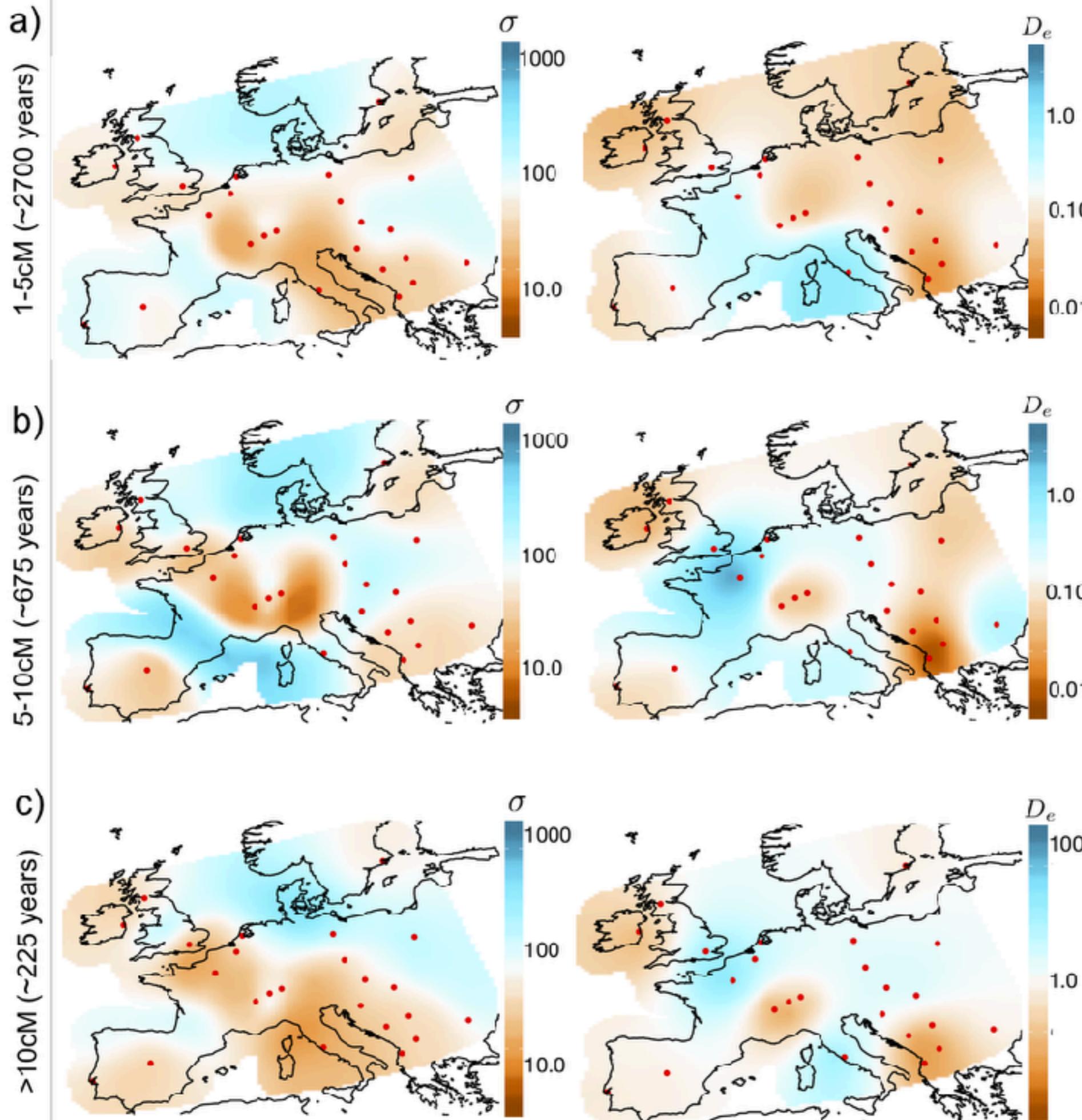
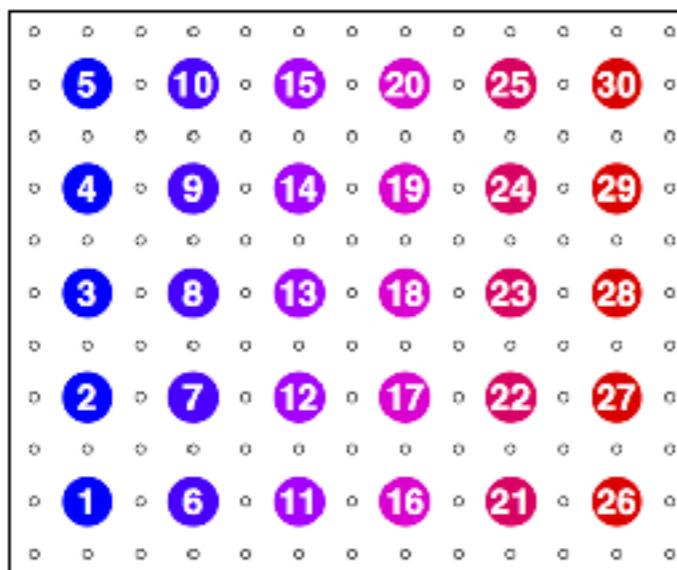


Fig 4. Inferred dispersal surfaces and population density surfaces over time for Europe.

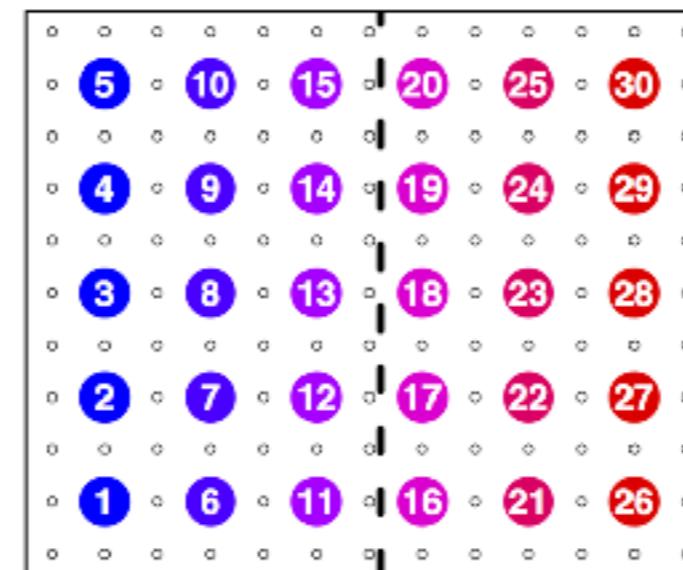
SpaceMix

populations are like individual genotype points of a PCA

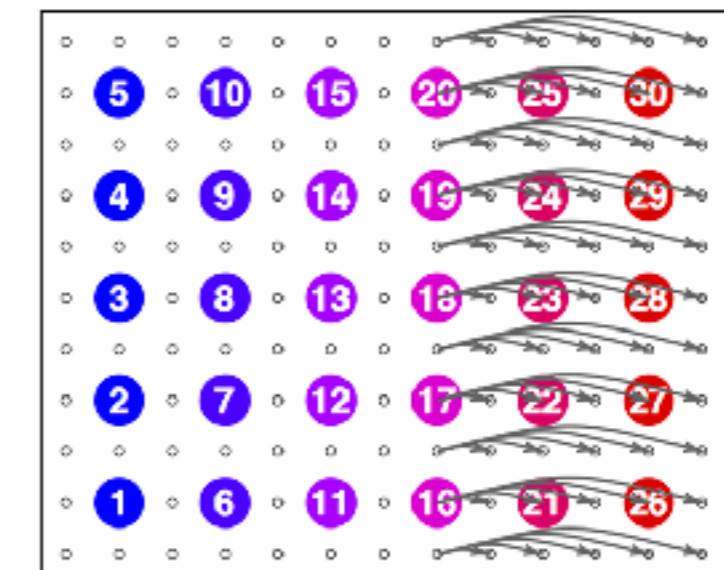
can detect direction



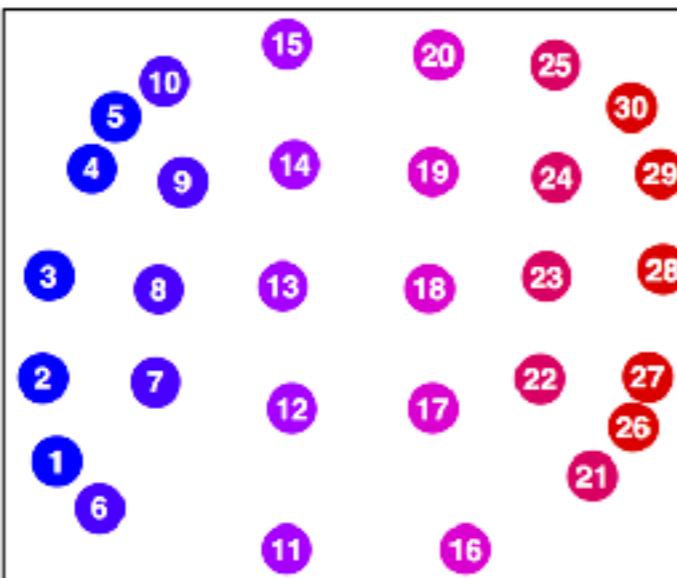
(a) homogeneous scenario



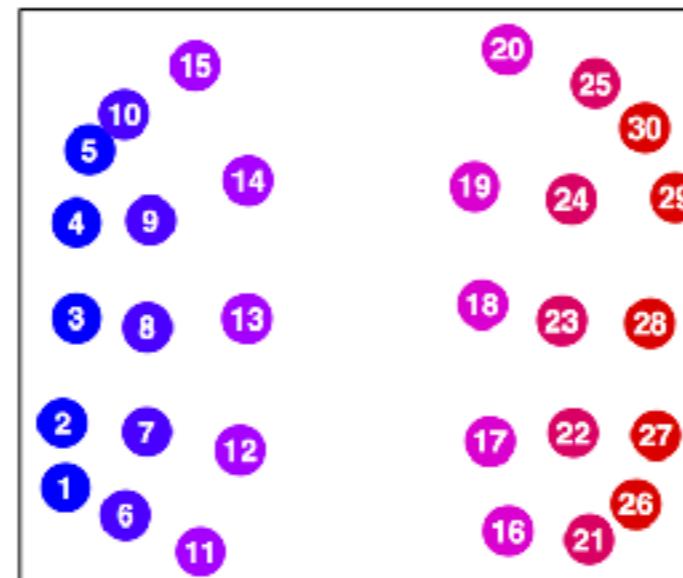
(b) barrier scenario



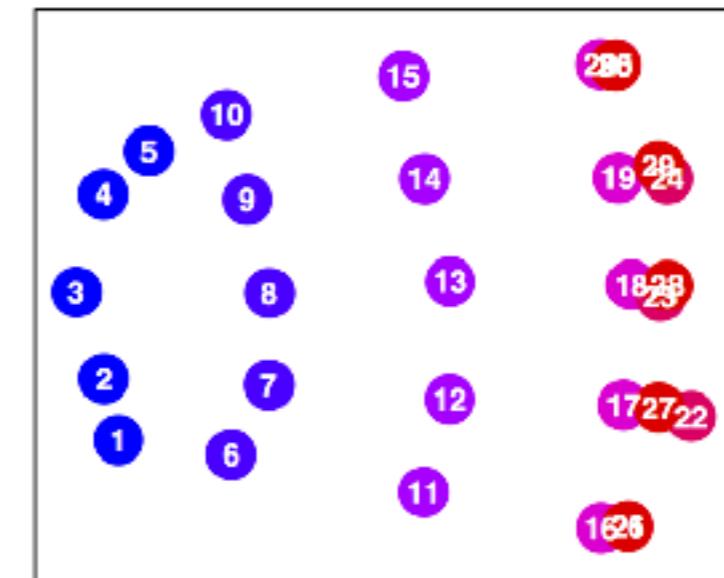
(c) expansion scenario



(d) geogenetic map of 1a



(e) geogenetic map of 1b



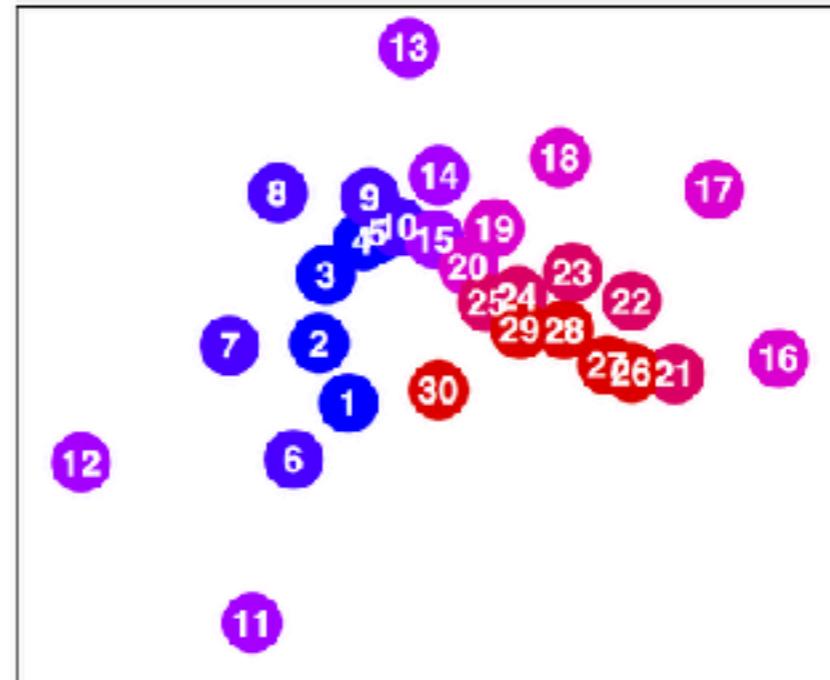
(f) geogenetic map of 1c

SpaceMix

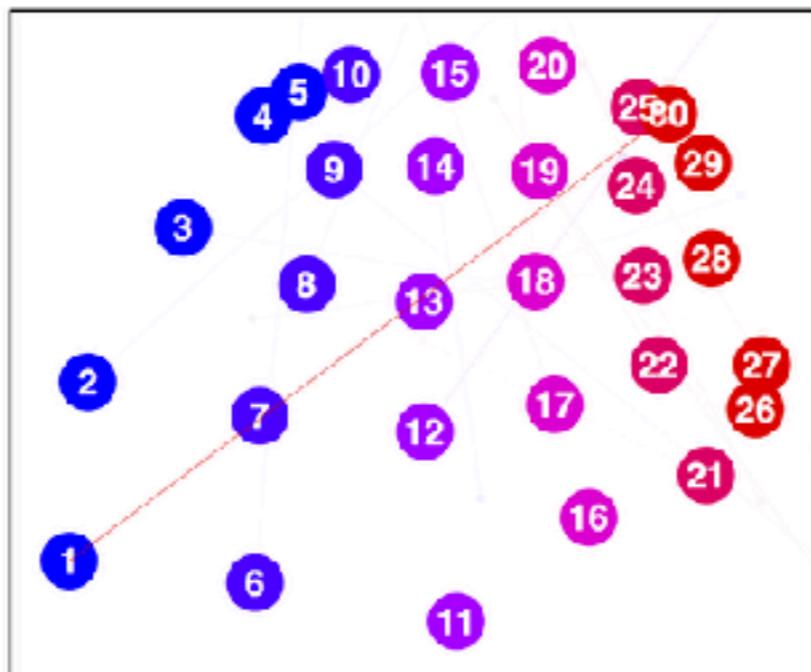
can detect long distance migration



(a) simulated lattice with admixture



(b) geogenetic map without admixture inference



(c) geogenetic map with admixture inference

*Annual Review of Ecology, Evolution, and
Systematics*

Spatial Population Genetics: It's About Time

Gideon S. Bradburd¹ and Peter L. Ralph^{2,3}

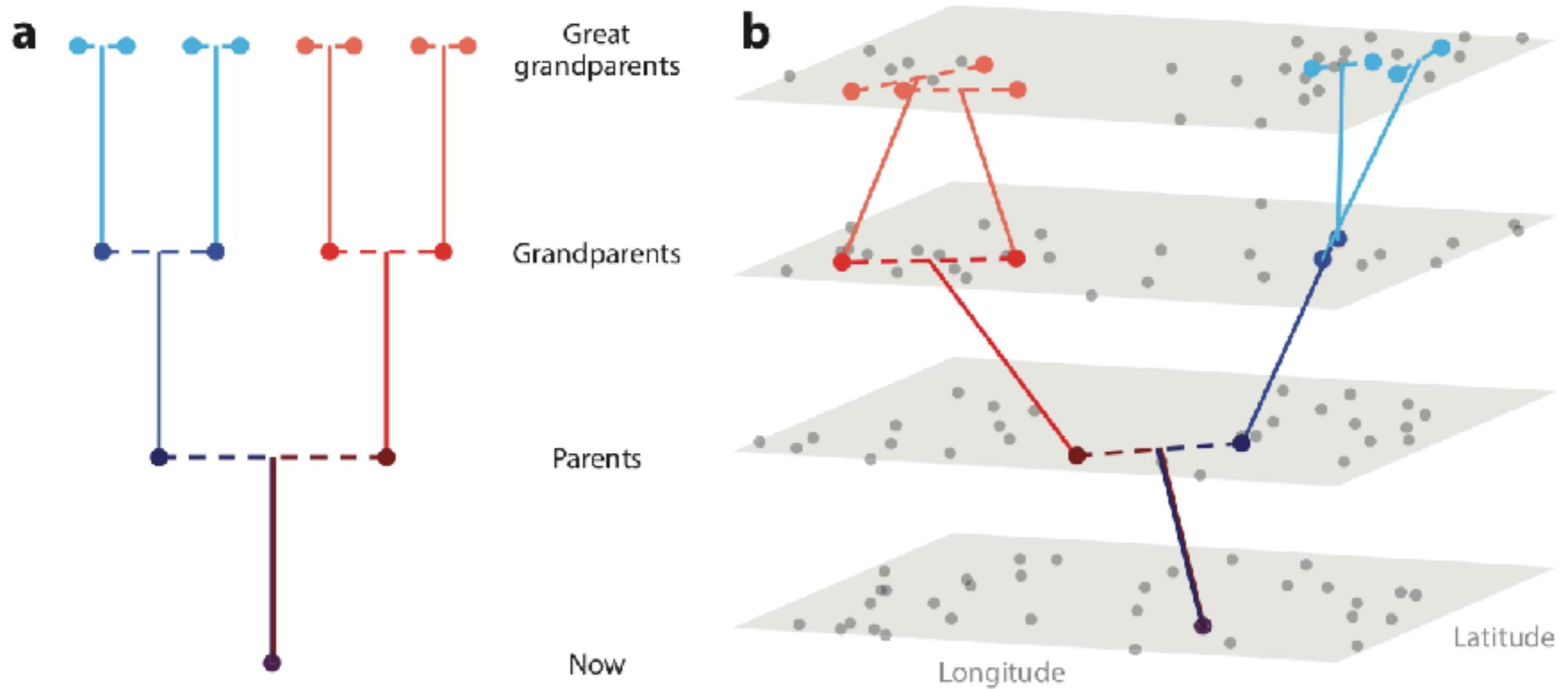
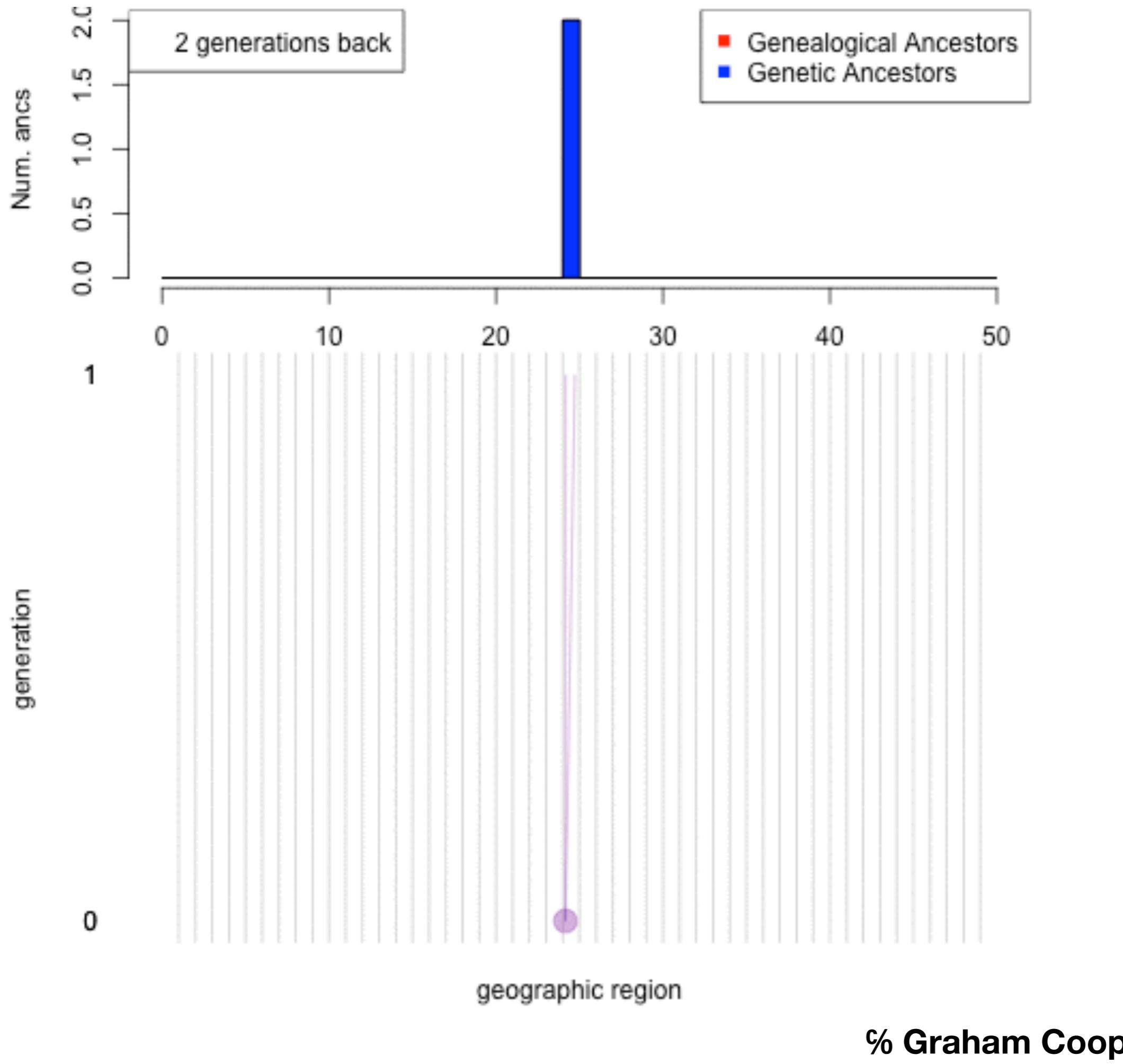
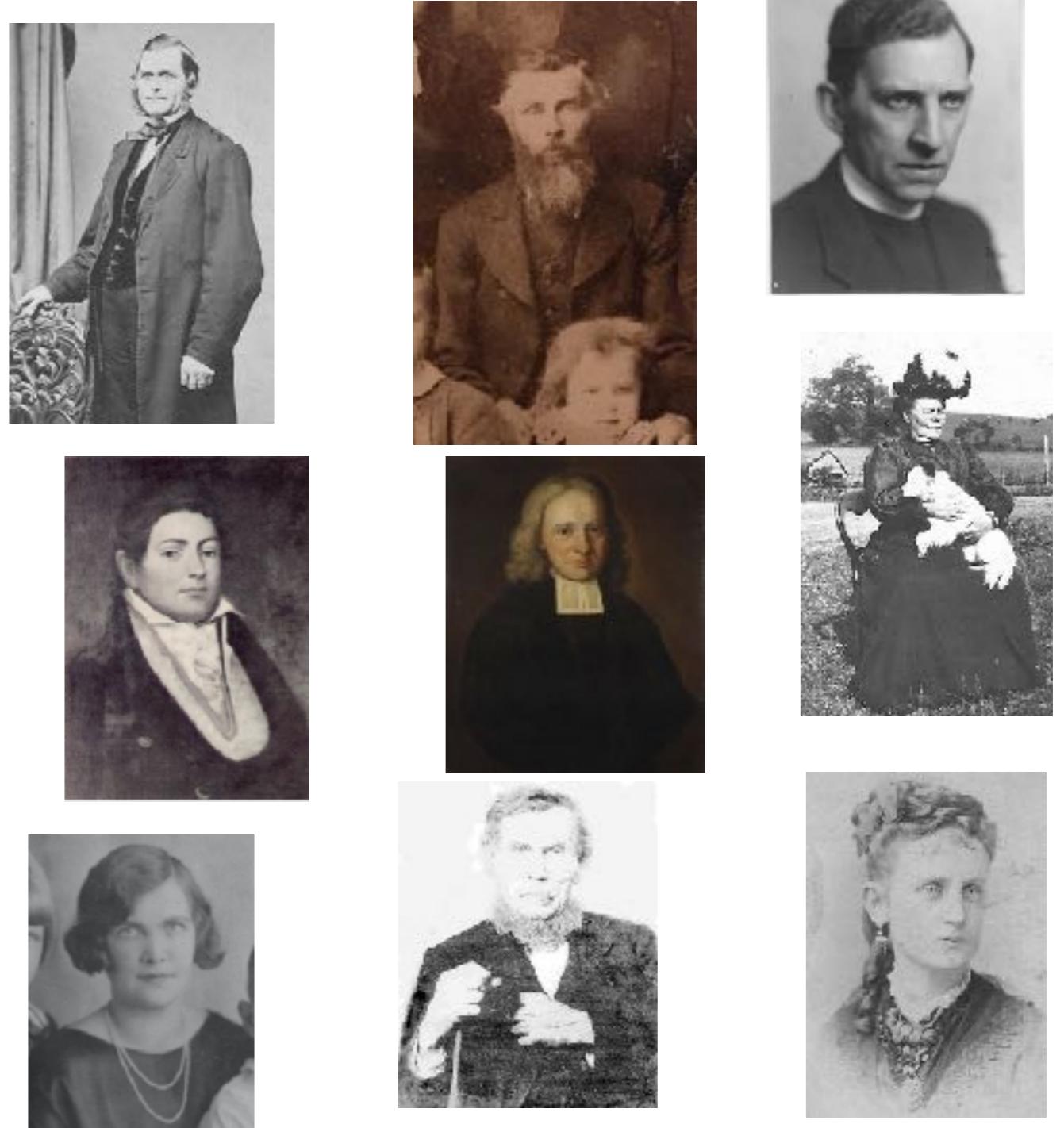
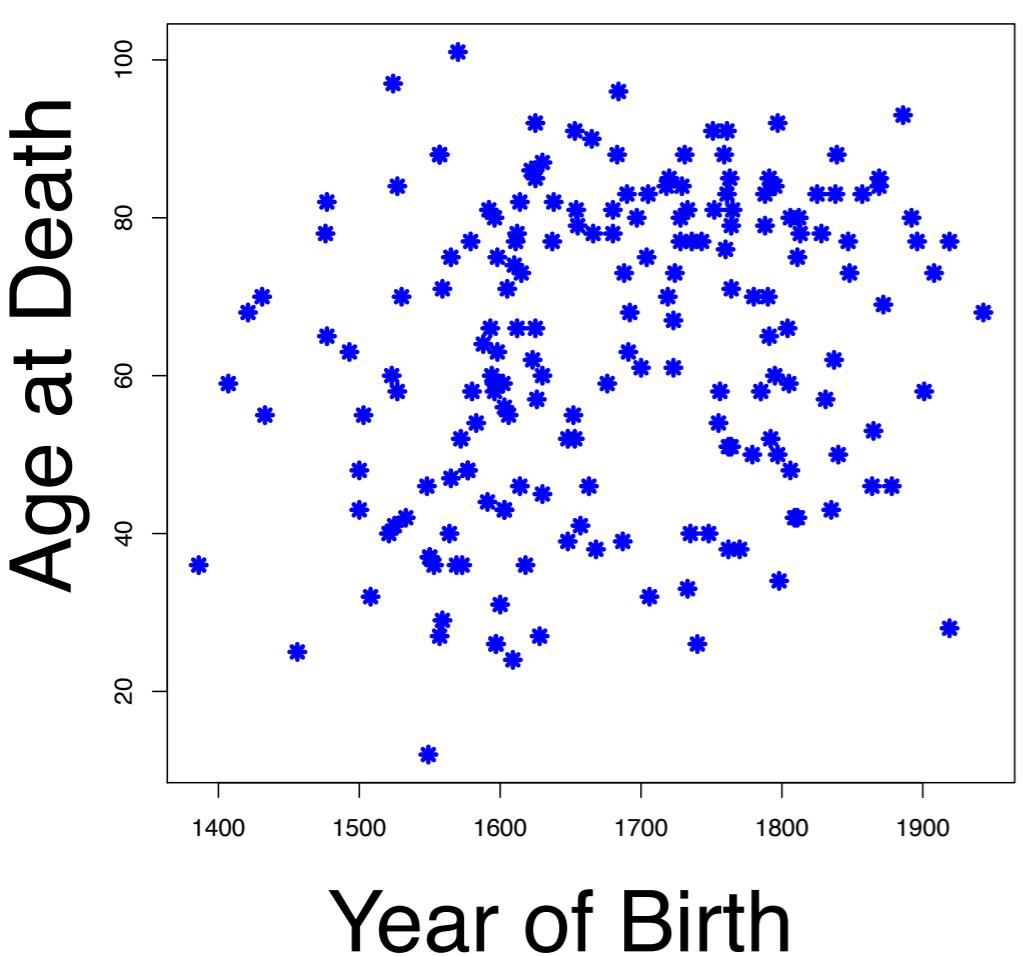


Figure 1

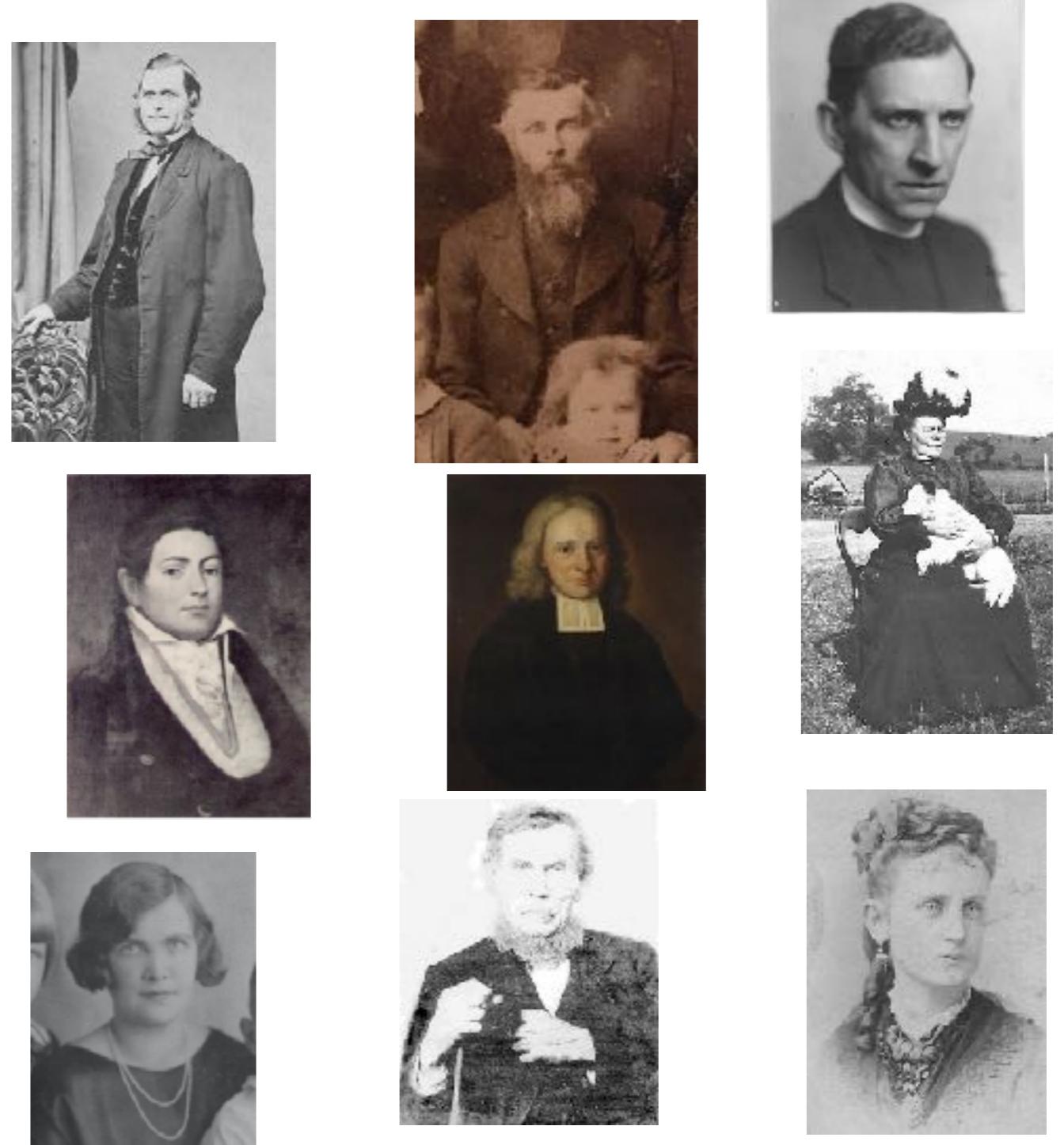
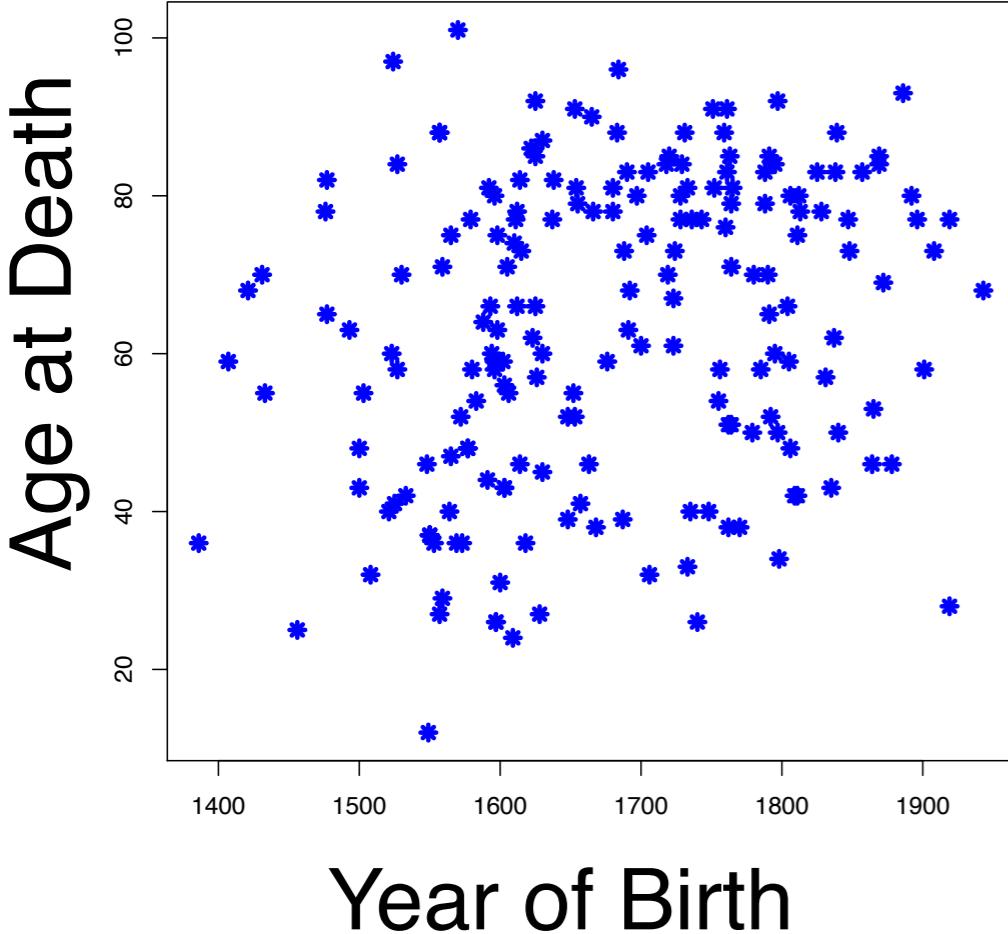
An example pedigree (*a*) of a focal individual sampled in the modern day, placed in its geographic context to make the spatial pedigree (*b*). Dashed lines denote matings, and solid lines denote parentage, with red hues for the maternal ancestors and blue hues for the paternal ancestors. In the spatial pedigree, each plane represents a sampled region in a discrete (nonoverlapping) generation, and each dot shows the birth location of an individual. The pedigree of the focal individual is highlighted back through time and across space.



193 of Hickerson's direct ancestors

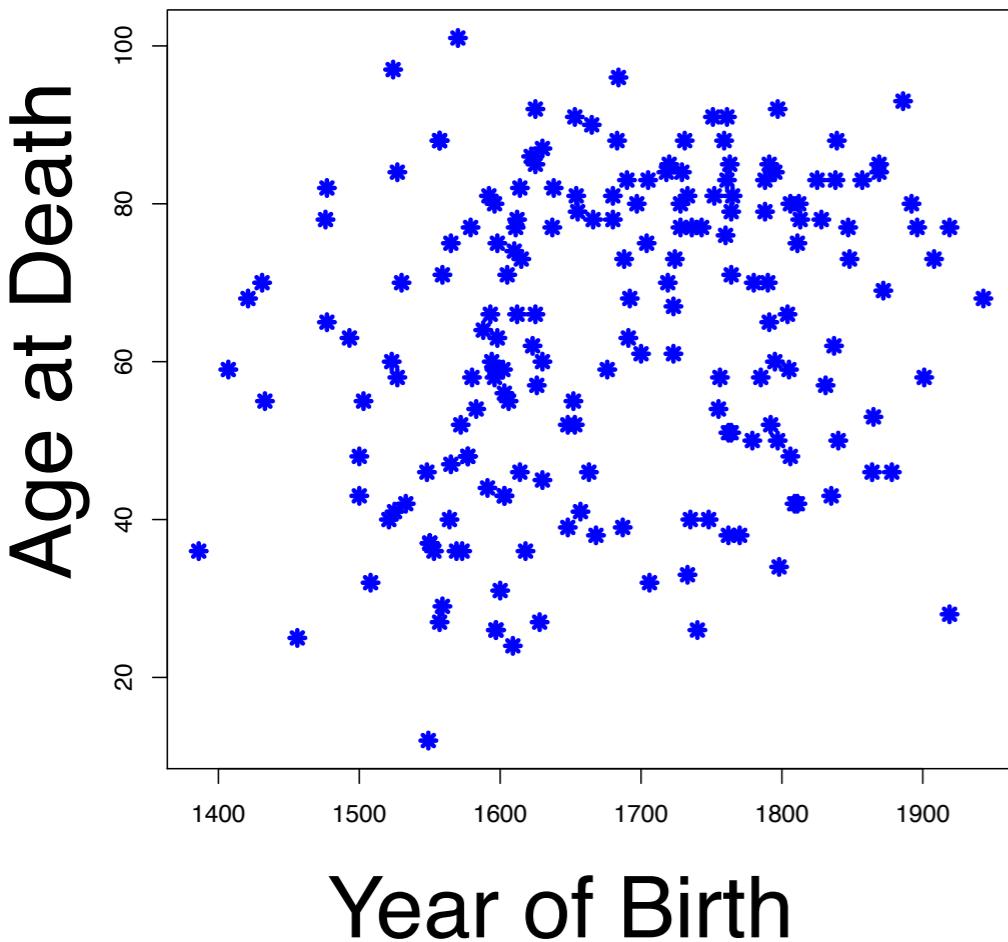


193 of Hickerson's direct ancestors
- did people have shorter lives btw. 1400-1650?)



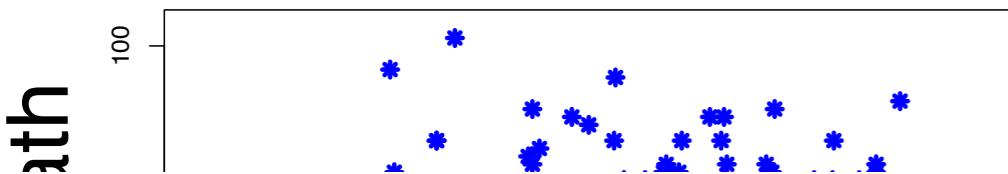
193 of Hickerson's direct ancestors

- did people have shorter lives btw. 1400-1650?)
- how is this a biased sample?



193 of Hickerson's direct ancestors

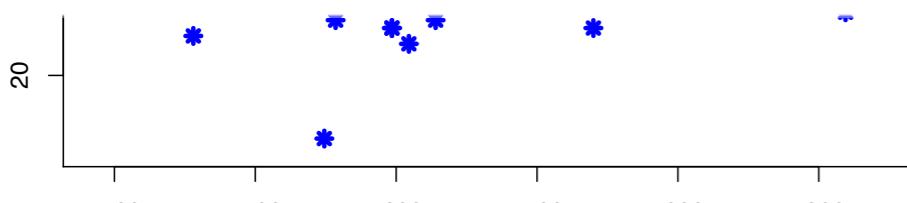
- did people have shorter lives btw. 1400-1650?)
- how is this a biased sample?



2.3. Estimating “Effective” Parameters from the Spatial Pedigree

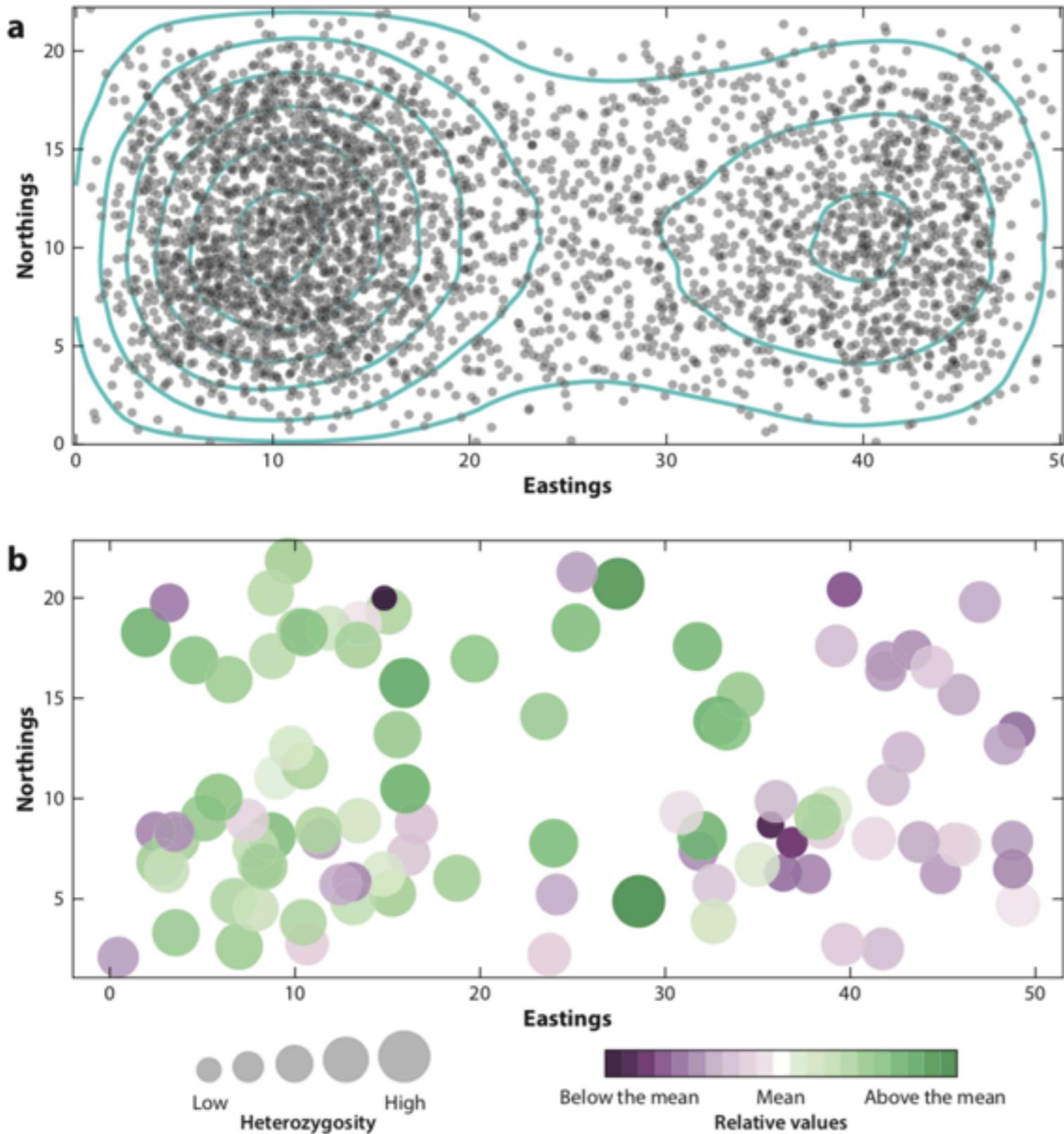
On one hand, every single one of my ancestors going back billions of years has managed to figure it out. On the other hand, that's the mother of all sampling biases.

—Randall Munroe, <https://xkcd.com/674/>



Year of Birth





We want to estimate local densities population size (and how this changed over time)

Figure 2

(a) Map of population density of the simulated population scenario we use for examples throughout. Populations have local density-dependent regulation within two valleys with high local carrying capacity (indicated by contours), separated and surrounded by areas with lower carrying capacity. The western valley has a carrying capacity about twice as large as that of the eastern valley. (b) Heterozygosity of a sample of 100 individuals across the range. Circle sizes are proportional to heterozygosity, and circle color indicates relative

**spatial
ancestors
back in time
(size = % of
relatedness)**

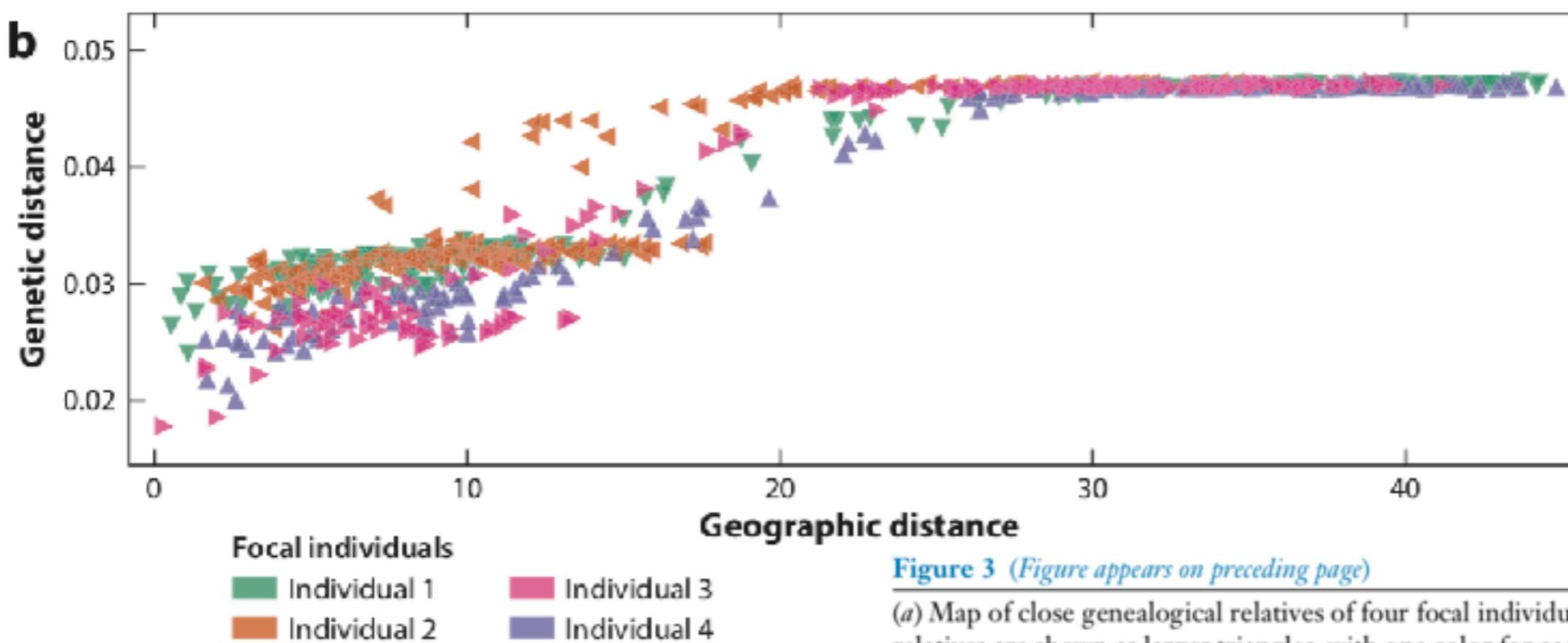
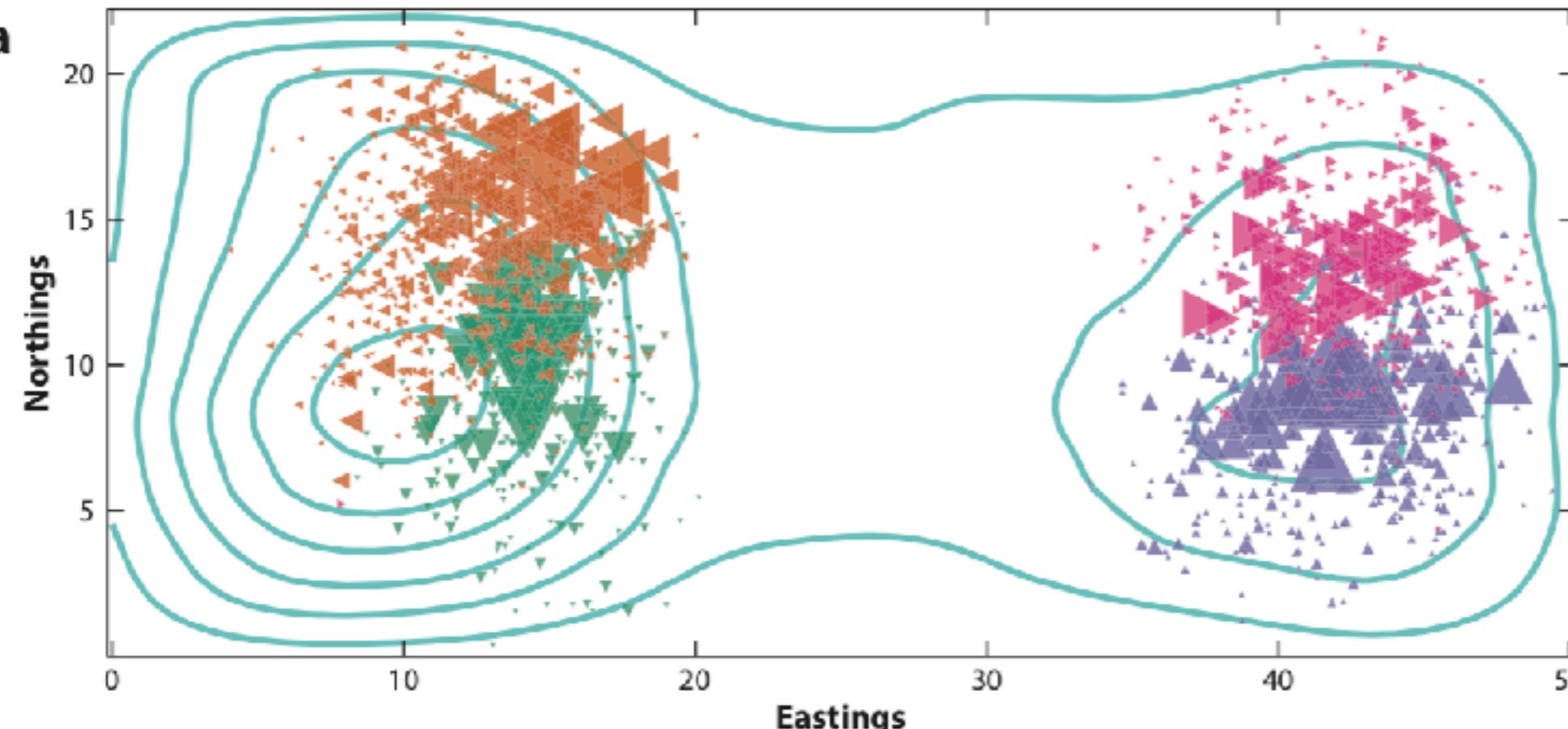


Figure 3 (Figure appears on preceding page)

(a) Map of close genealogical relatives of four focal individuals. Geographic locations of genealogical relatives are shown as larger triangles, with one color for each focal individual, with triangle size proportional to the expected proportion of genome shared (i.e., 2^{-k} , where k is the number of steps in the shortest path through the pedigree to the focal individual). (b) Genetic against geographic distance between each of four focal individuals and 100 other individuals, from the same map, and using the same colors, as in panel a. Genetic distance (calculated as expected sequence divergence) increases with geographic distance within and between each valley at similar rates but levels off within each valley at different values due to different population sizes. (c) Flux across two boundaries on the same continuous map as panel a. All individuals across

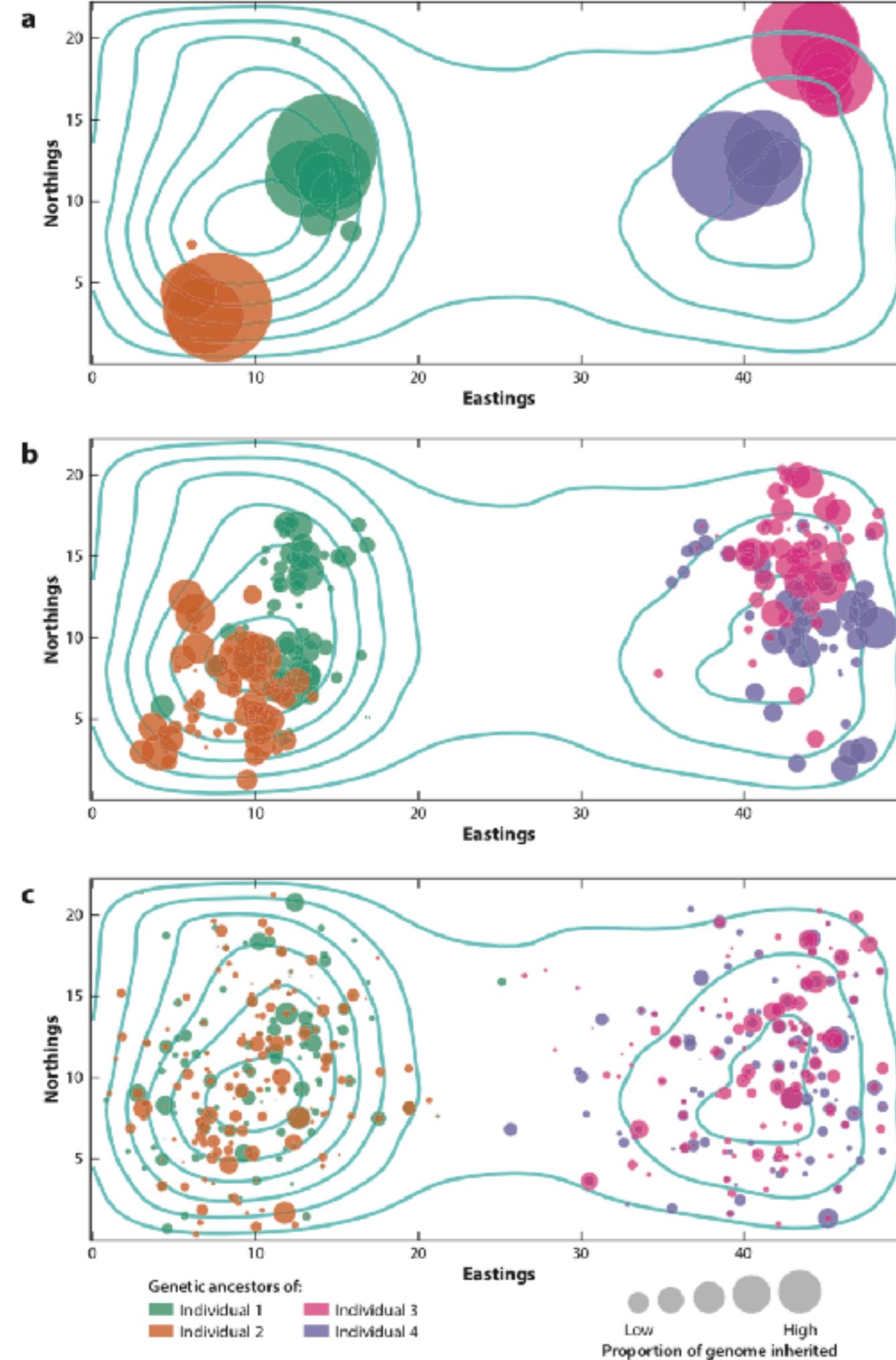
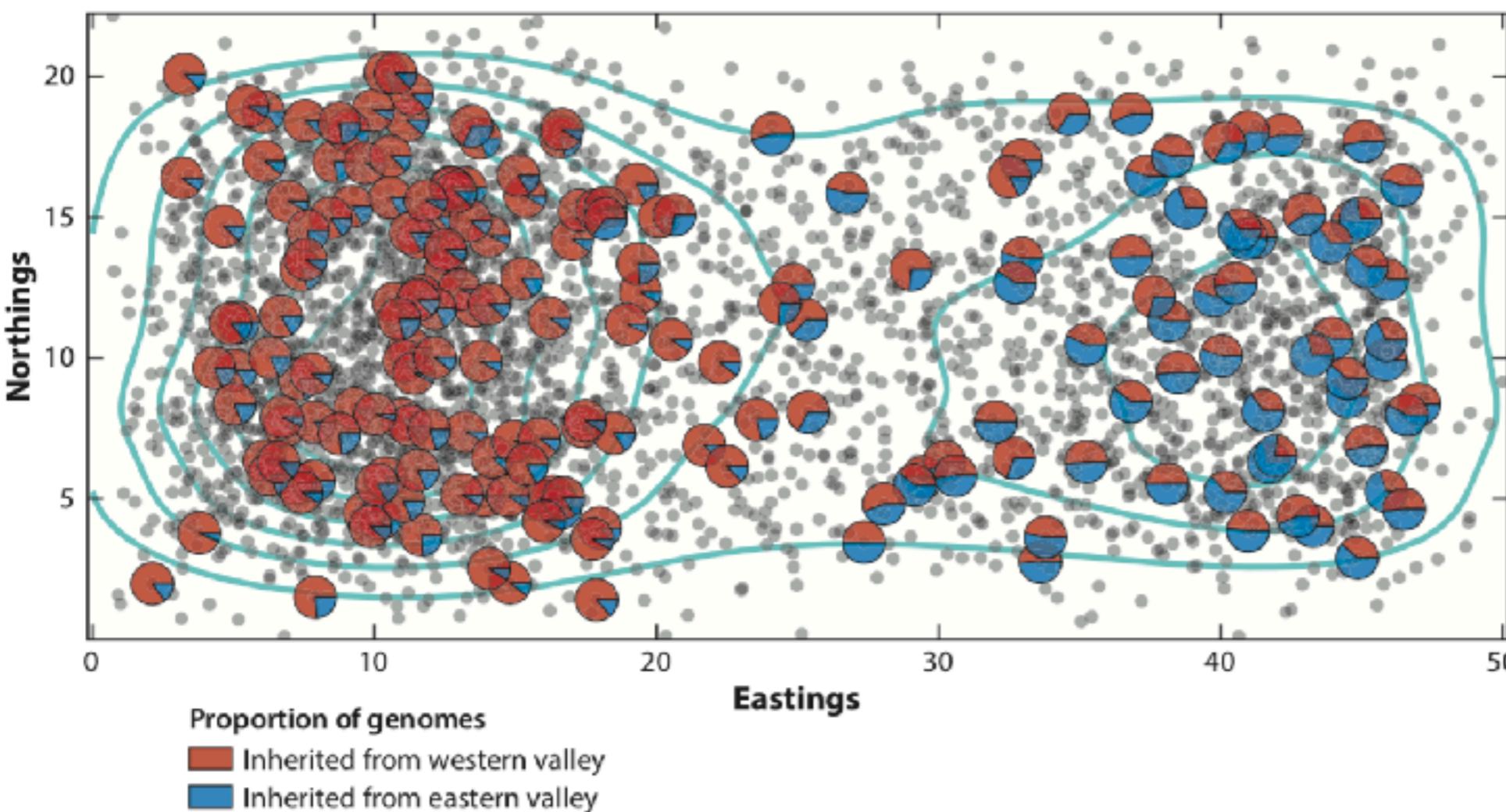
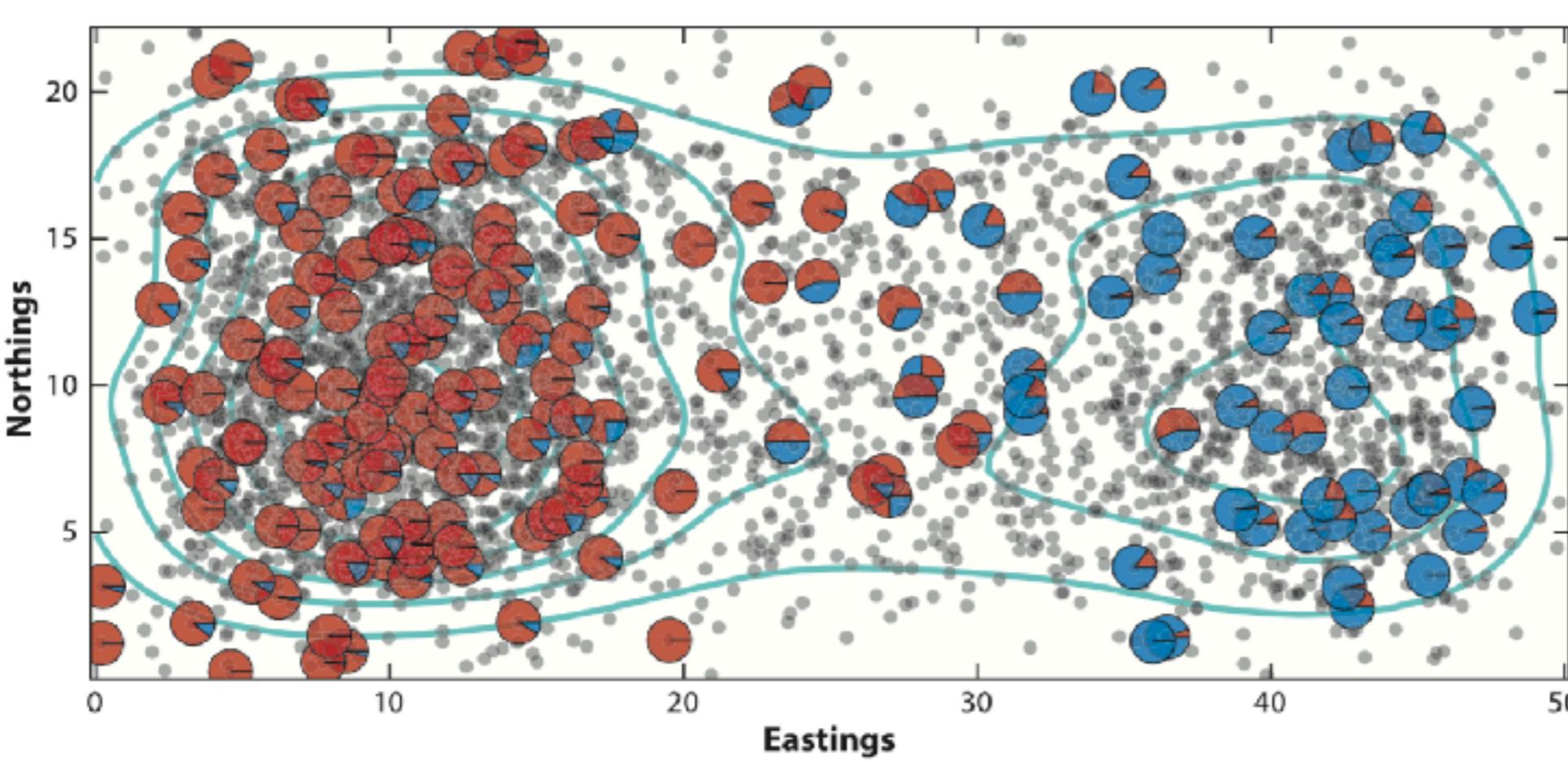


Figure 4

Spatial locations of the ancestors of four individuals at (a) 1, (b) 15, and (c) 75 generations in the past. All individuals from whom the four individuals have inherited genome are denoted by a circle, with one color for each individual, and with circle area proportional to the amount of genome inherited. The landscape is the same as in **Figure 3**. Since the simulation uses overlapping generations, ancestors of different degrees are present at the same time.



Admixture proportions after a recent secondary contact. Each pie shows the proportion of a diploid individual's genomes that inherit from the western (red) and eastern (blue) valleys, respectively. The plots show the geographic distribution of that ancestry (a) 8 generations, and (b) 75 generations after a barrier along the top of the ridge is removed.

