

Lecture 4

Inbreeding and Population structure

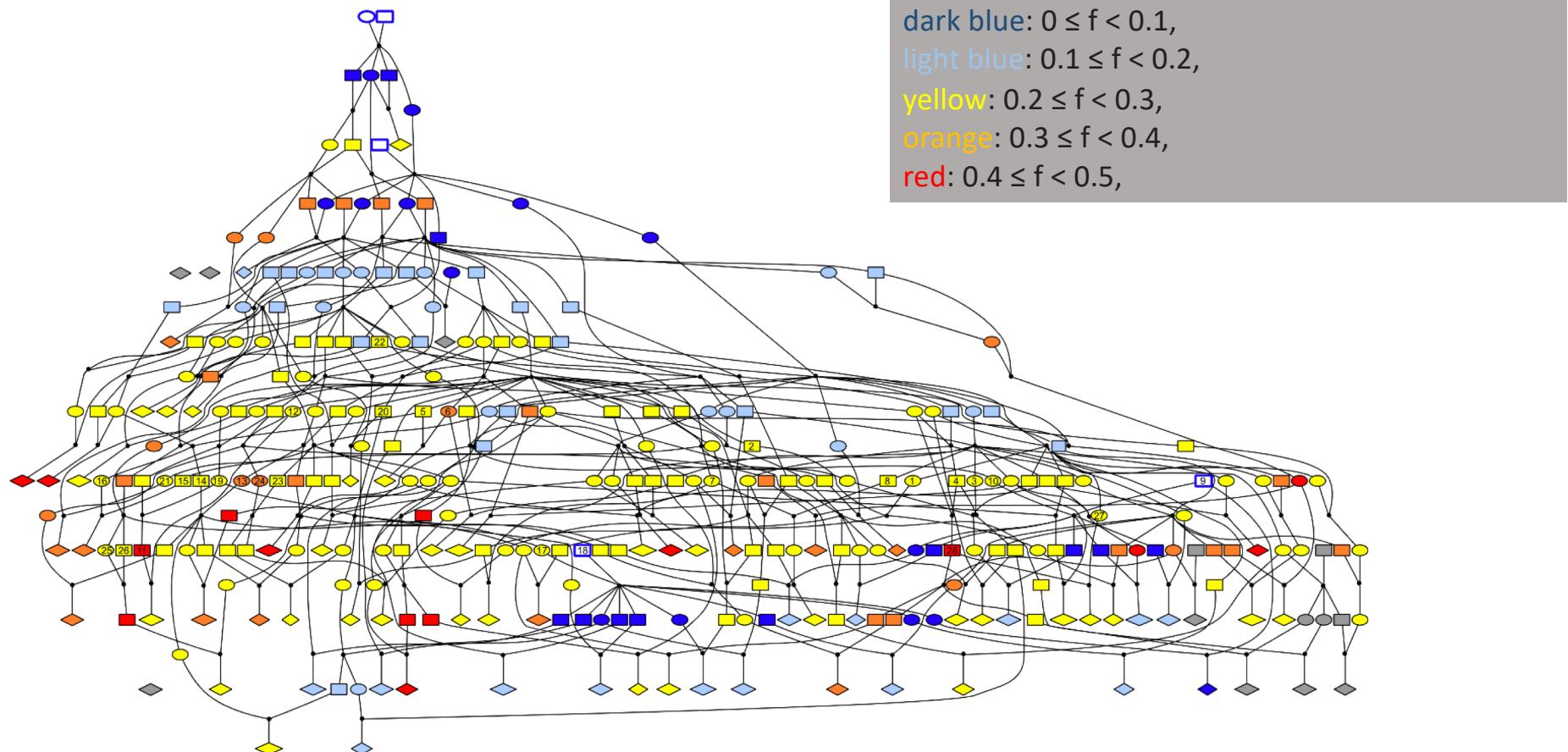
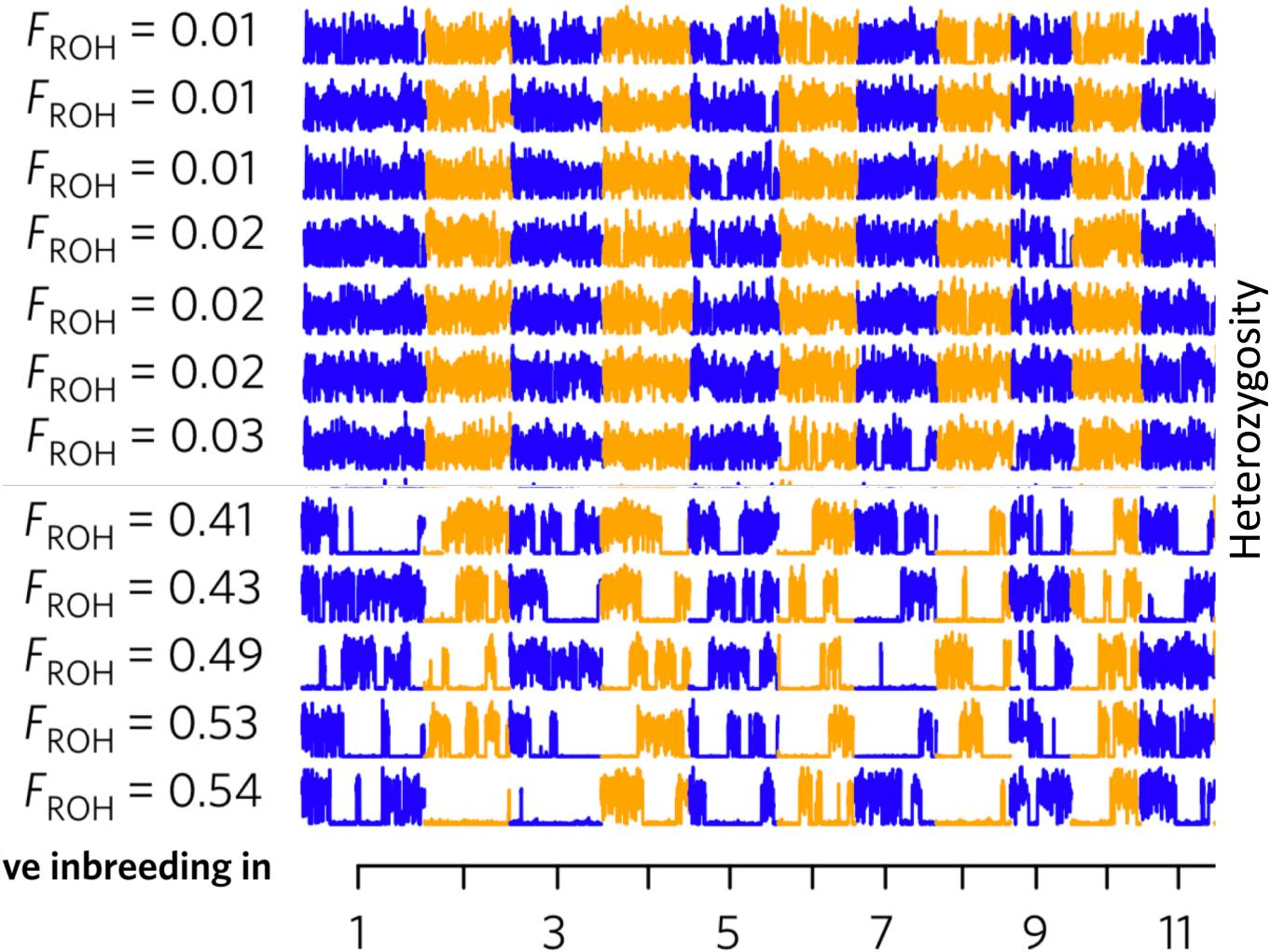


Fig. 1 Pedigree of breeding Scandinavian wolves from 1983 to 2012. Ellipses are females, rectangles are males, and diamonds represent litters where no offspring has yet entered the breeding population. The colours represent the inbreeding coefficient f of the individuals.

Genetic rescue in a severely inbred wolf population

MIKAEL AKESSON,^{*} OLOF LIBERG,^{*} HÅKAN SAND,^{*} PETTER WABAKKEN,[†]
STAFFAN BEN SCH[‡] and ØYSTEIN FLAGSTAD[§]

0.2, yellow: $0.2 \leq f < 0.3$, orange: $0.3 \leq f < 0.4$, red: $0.4 \leq f < 0.5$, grey: offspring $f = 0$, are nonfilled symbols with a blue frame. The pedigree S3, Supporting information) included in the 14 target pairs.



Genomic consequences of intensive inbreeding in an isolated wolf population

Marty Kardos^{1,2}, Mikael Åkesson³, Toby Fountain^①, Øystein Flagstad⁴, Olof Liberg³, Pall Olason⁵, Håkan Sand³, Petter Wabakken⁶, Camilla Wikner^③ and Hans Ellegren^{①*}

Estimating Inbreeding coefficients

Notes: Equations 2.11 on page 23.



Figure 1. Blue- and white-flowered plants of *Linanthus parryae* from the Mojave Desert near Pearblossom, California, USA. D. W. Schemske.

Imagine that we knew the population frequency
Of the blue colour allele to be 0.1.

If we observed that only 10%
of the population are heterozygotes (blue/white allele),
what is our estimate of the inbreeding coefficient?

Population structure

- Inbreeding is just one form of assortative mating.
- A common form of assortative mating:
population structure or population subdivision, where
mating takes place within sub-groups of the whole
population.

F statistics as a measure of population structure

- Wright (1943, 1951) defined F_{XY} as:
- the correlation between random gametes, drawn from the same X, relative to Y.
- Measures the reduction in heterozygosity of X relative to that expected in Y.
- Often in hierarchical manner: F_{IS} , F_{ST} , and F_{IT}
- There are a number of different estimators of F statistics.

We summarize the differences in allele frequencies across populations by F statistics



Figure 1. Blue- and white-flowered plants of *Linanthus parryae* from the Mojave Desert near Pearblossom, California, USA. D. W. Schemske.

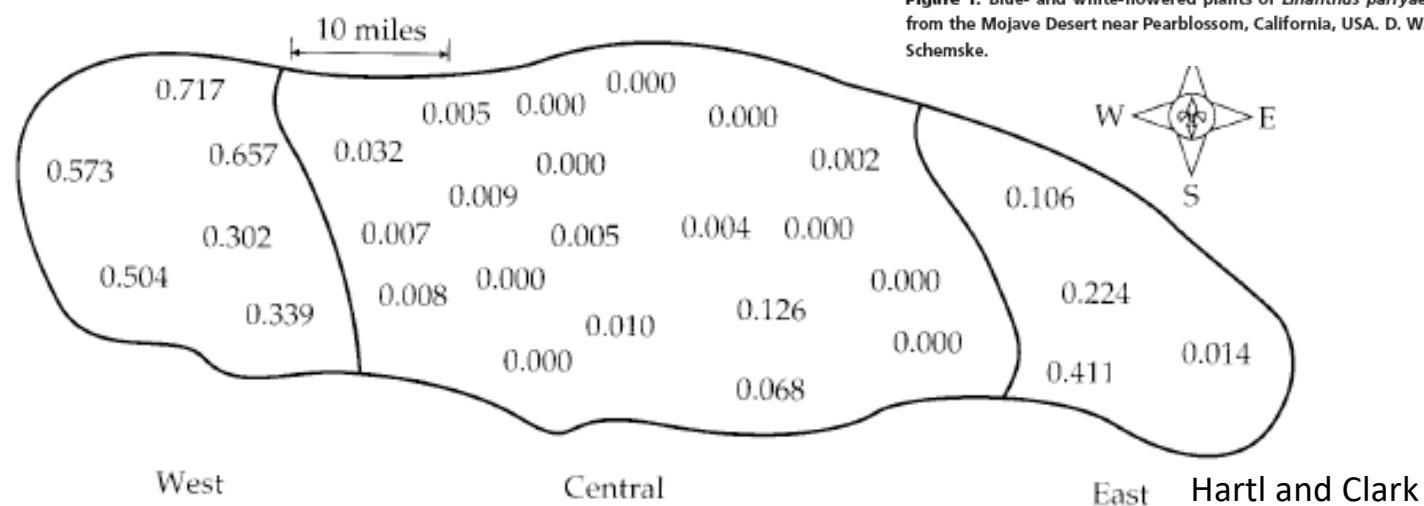


FIGURE 6.13 Estimated frequency of a recessive allele for blue flower color in populations of *Linanthus parryae* in an area of approximately 900 square miles in the Mojave desert. Each allele frequency is based on an examination of approximately 4000 plants over an area of about 30 square miles. (After Wright 1943a.)

Eqn 2.12-2.14 on page 27

Taken from Wright 1943

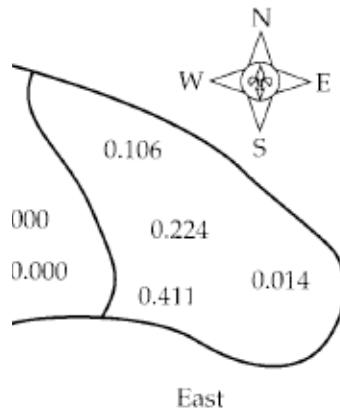


Figure 1. Blue- and white-flowered plants of *Linanthus parryae* from the Mojave Desert near Pearblossom, California, USA. D. W. Schemske.

allele freq	Heterozygosity ($H_S = 2p_S q_S$)	$F_{ST} = 1 - H_S / H_T$
0.106	0.190	0.383
0.224	0.348	-0.132
0.411	0.484	-0.577
0.014	0.028	0.910
Average	0.189	0.262
		0.146

$$H_T = 2p_T(1-p_T) = 2 \times 0.189 \times (1-0.189) = 0.307$$

$$\bar{F}_{ST} = 1 - \bar{H}_S / H_T = 0.146$$

Imagine that in the top population we observe that Individual Heterozygosity is actually 0.1
What Is F_{IS} and F_{IT} ?

- F_{ST} = proportion of the variance due to between-population (rather than within-population) differences.
- Low values of F_{ST} indicate that on average differences are found among individuals from the same population rather than in one population but not the other.

Equation 2.18, page 29

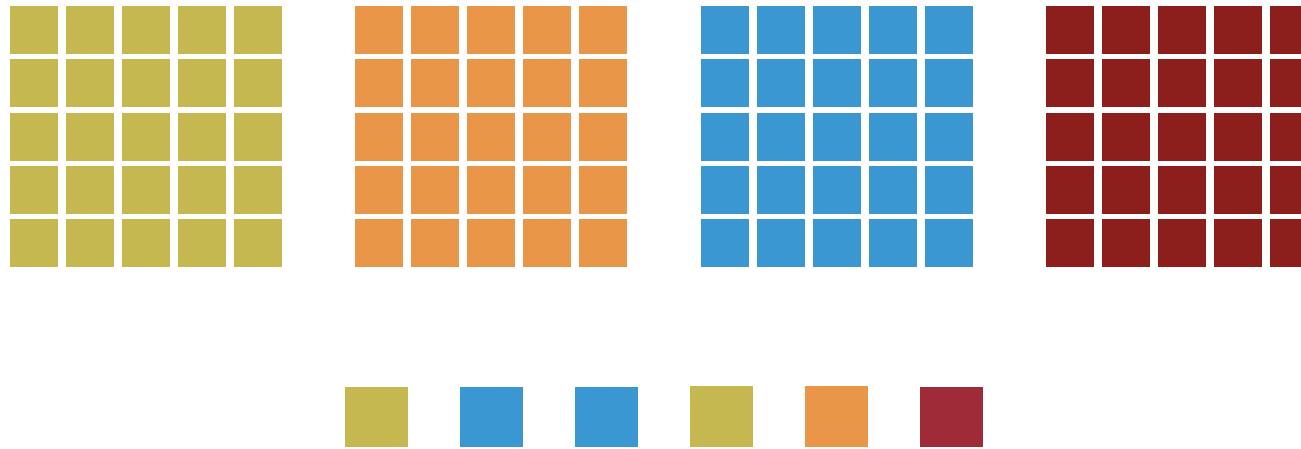
If you look at polymorphic sites in a sample of Maasai and Inuits, most differences are found *among individuals from the same population* rather than in one population but not the other.



In other words, the genomes of these two Inuits is only slightly more (~15%) similar to one another than are the genomes of an Inuit and a Maasai.

Reference Populations

Section 2.3.3 of notes



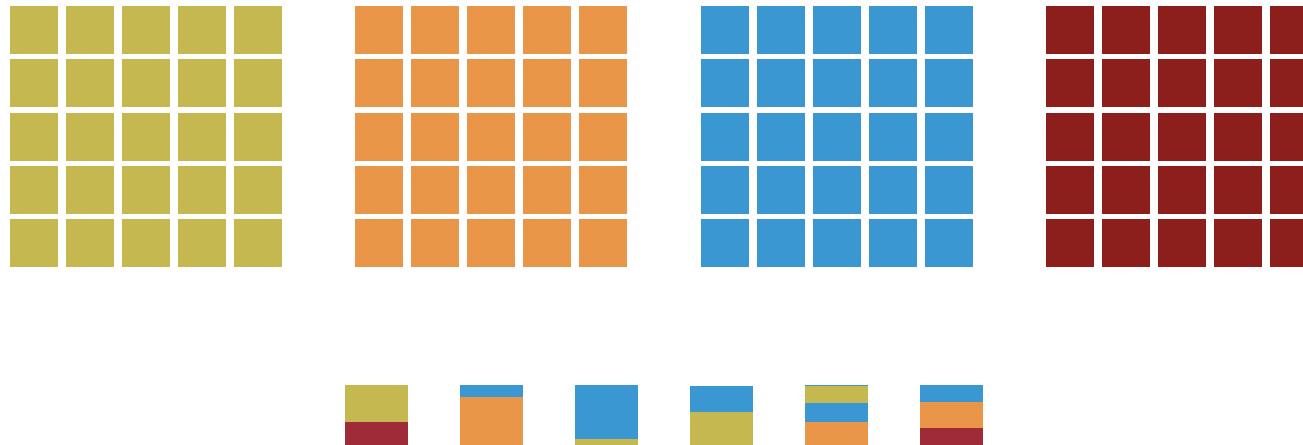
Identify **origins** of individuals
each with a **single** ancestry

Approaches: Based on Hardy-Weinberg genotype proportions

Adapted from John Novembre slide

Section 2.3.3 of notes

Reference Populations (known or unknown)



Identify **ancestry proportions** for individuals with **admixed** ancestry

Approaches: Structure (MCMC, Bayesian)
Or ADMIXTURE (quadratic programming)

Adapted from John Novembre slide

Genetic Structure of Human Populations

Noah A. Rosenberg,^{1*} Jonathan K. Pritchard,² James L. Weber,³
Howard M. Cann,⁴ Kenneth K. Kidd,⁵ Lev A. Zhivotovsky,⁶
Marcus W. Feldman⁷

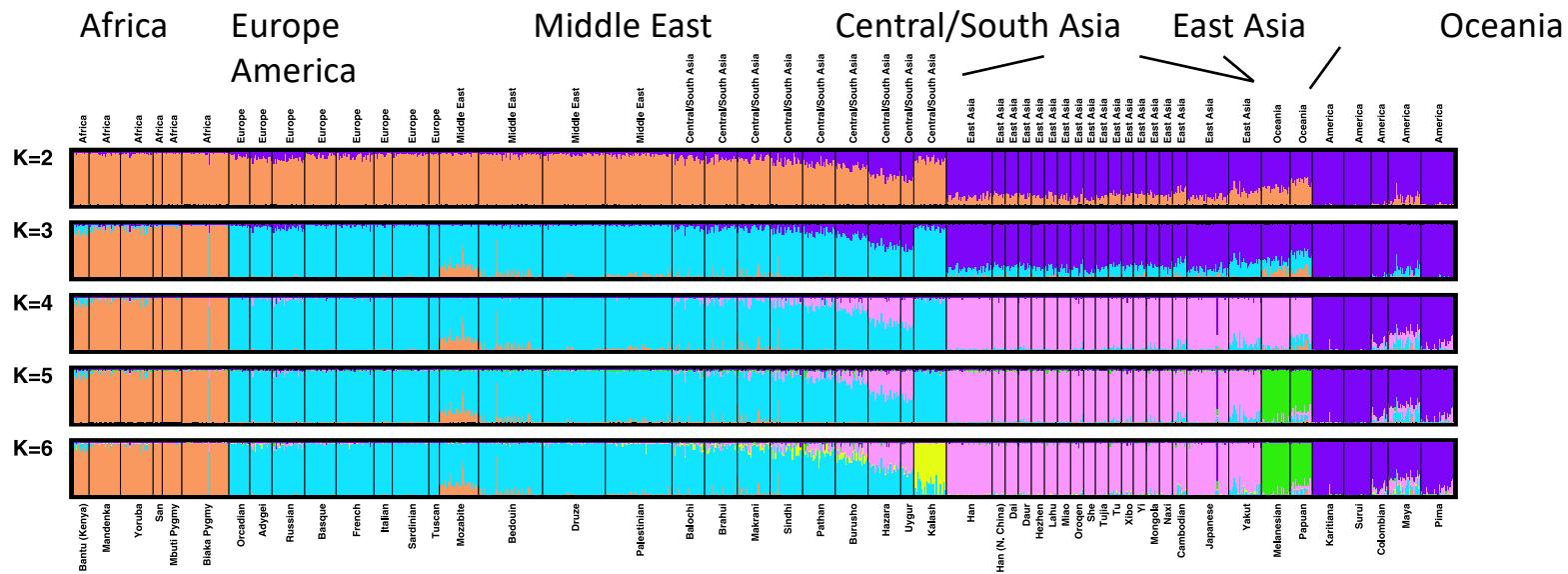
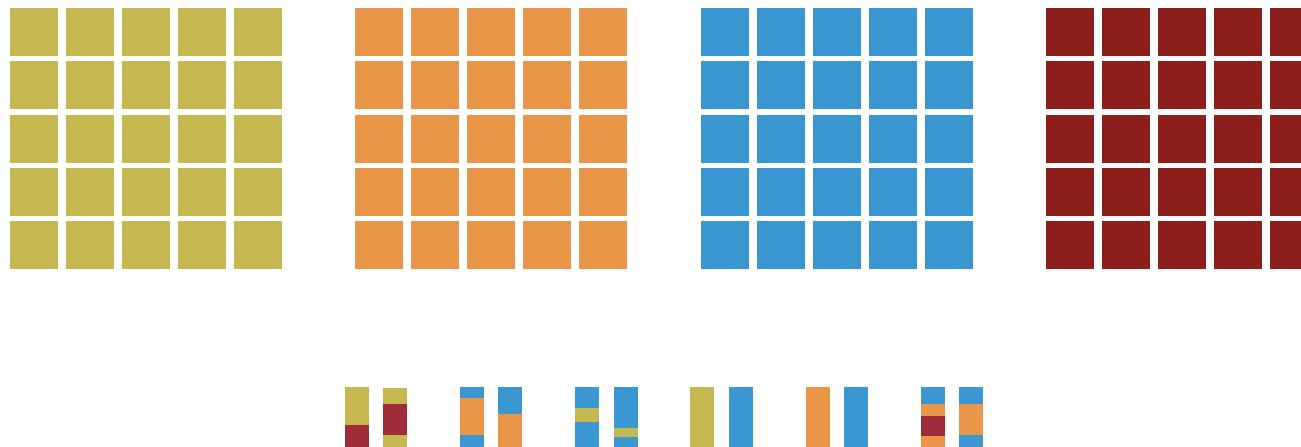


Fig. 1. Estimated population structure. Each individual is represented by a thin vertical line, which is partitioned into K colored segments that represent the individual's estimated membership fractions in K clusters. Black lines separate individuals of different populations. Populations are labeled below the figure, with their regional affiliations above it. Ten structure runs at each

K produced nearly identical individual membership coefficients, having pairwise similarity coefficients above 0.97, with the exceptions of comparisons involving four runs at $K = 3$ that separated East Asia instead of Eurasia, and one run at $K = 6$ that separated Karitiana instead of Kalash. The figure shown for a given K is based on the highest probability run at that K .

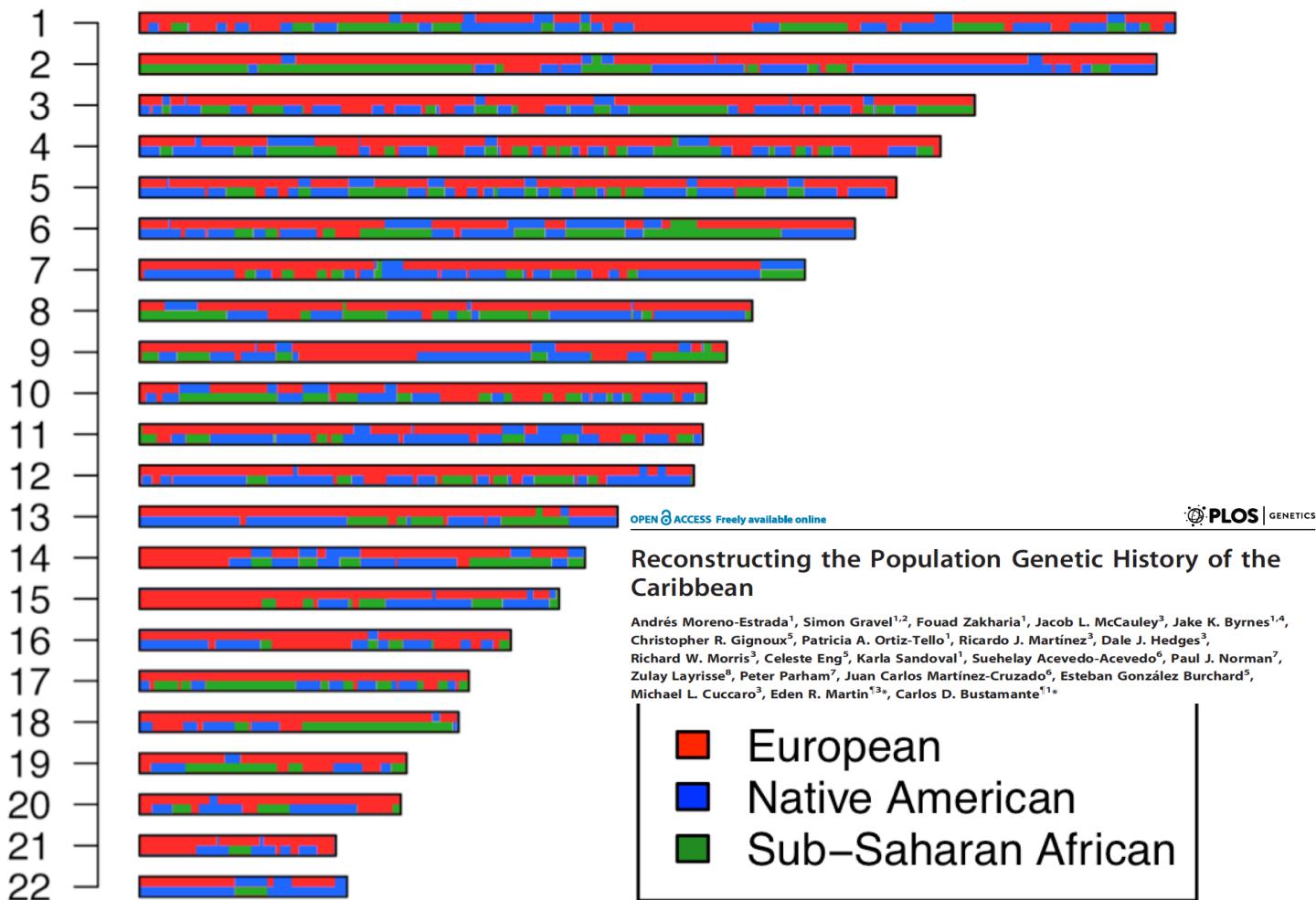
Reference Populations



Identify origins of **chromosomal segments** in individuals of **admixed** ancestry

Approaches: Based on Hidden Markov Models

Local ancestry



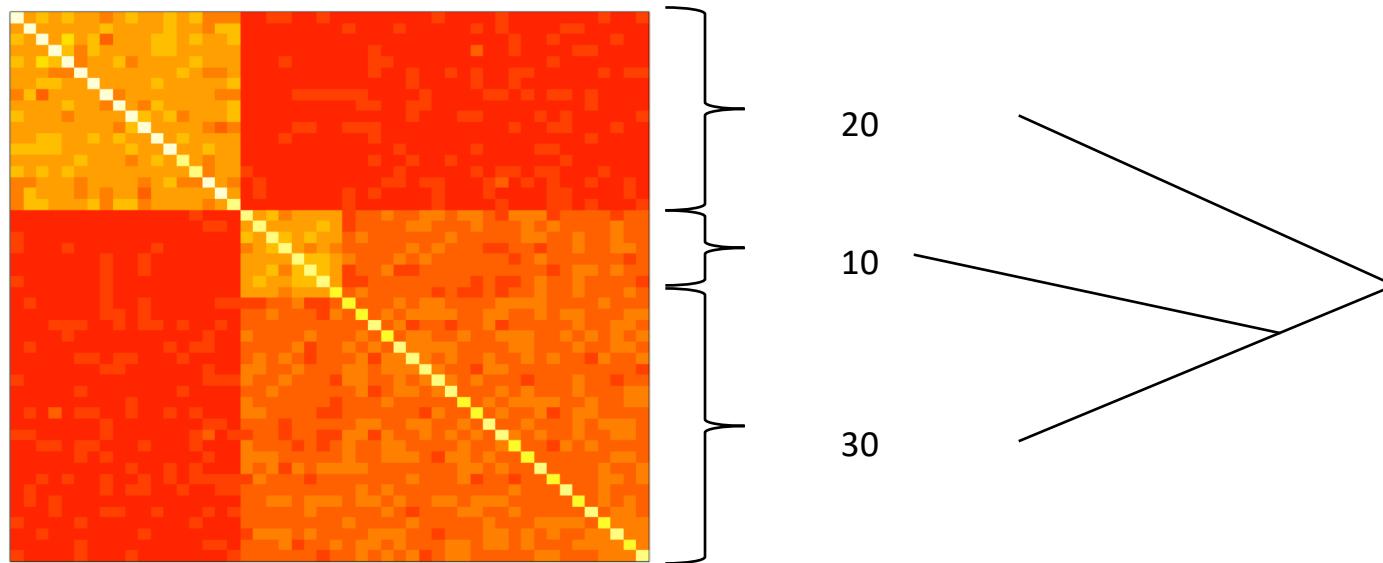
Principal components analysis

Example

Section 2.3.4 of notes

(Simulated data, N=50 individuals, L=1000 SNPs)

Relatedness matrix R



ith and jth entry = average over loci (l) of $(X_{il} - \bar{X}_l)(X_{jl} - \bar{X}_l)$

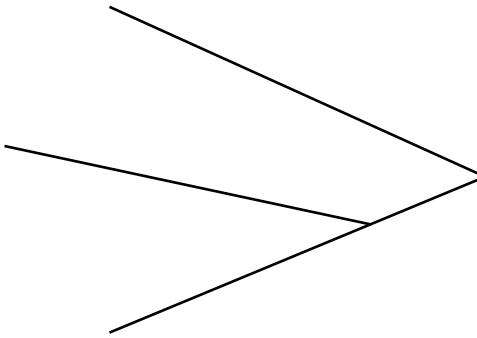
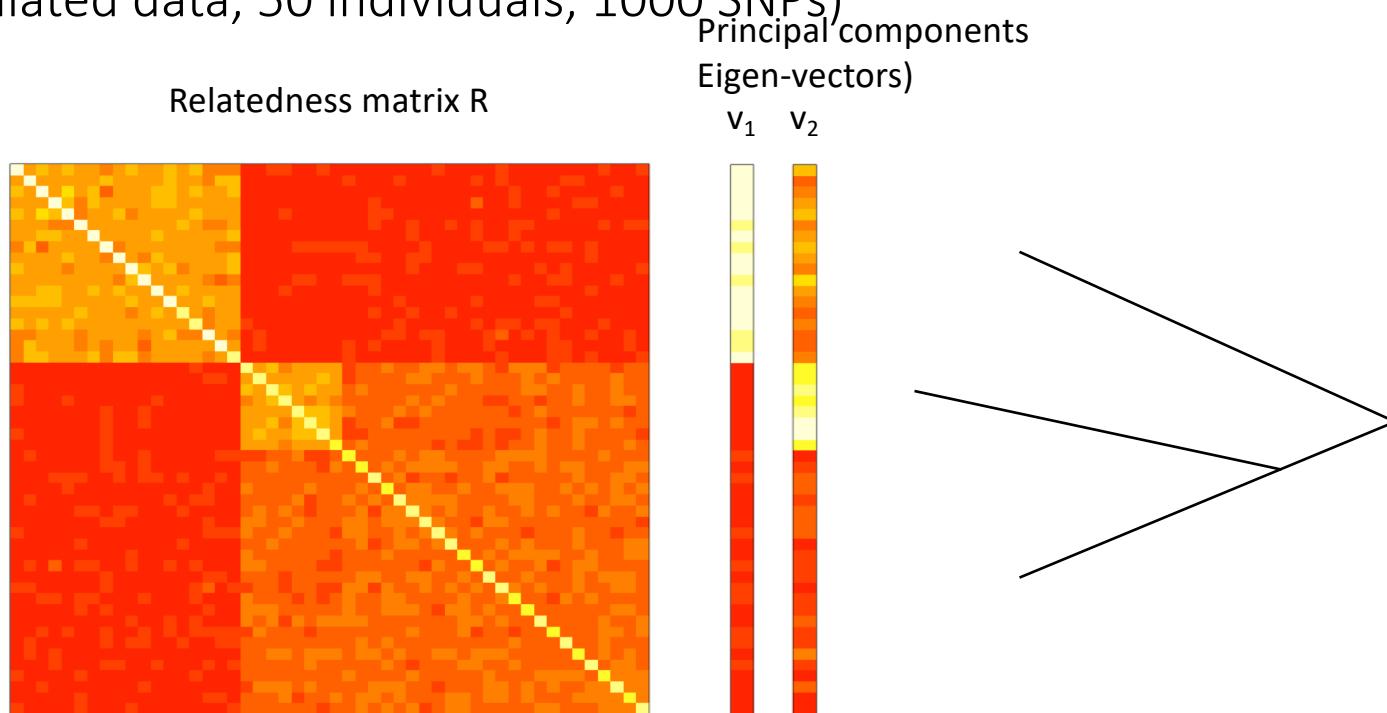
Where X_l is mean freq. of the lth locus.

Modified from slide by Gavin Band

Principal components analysis

Example

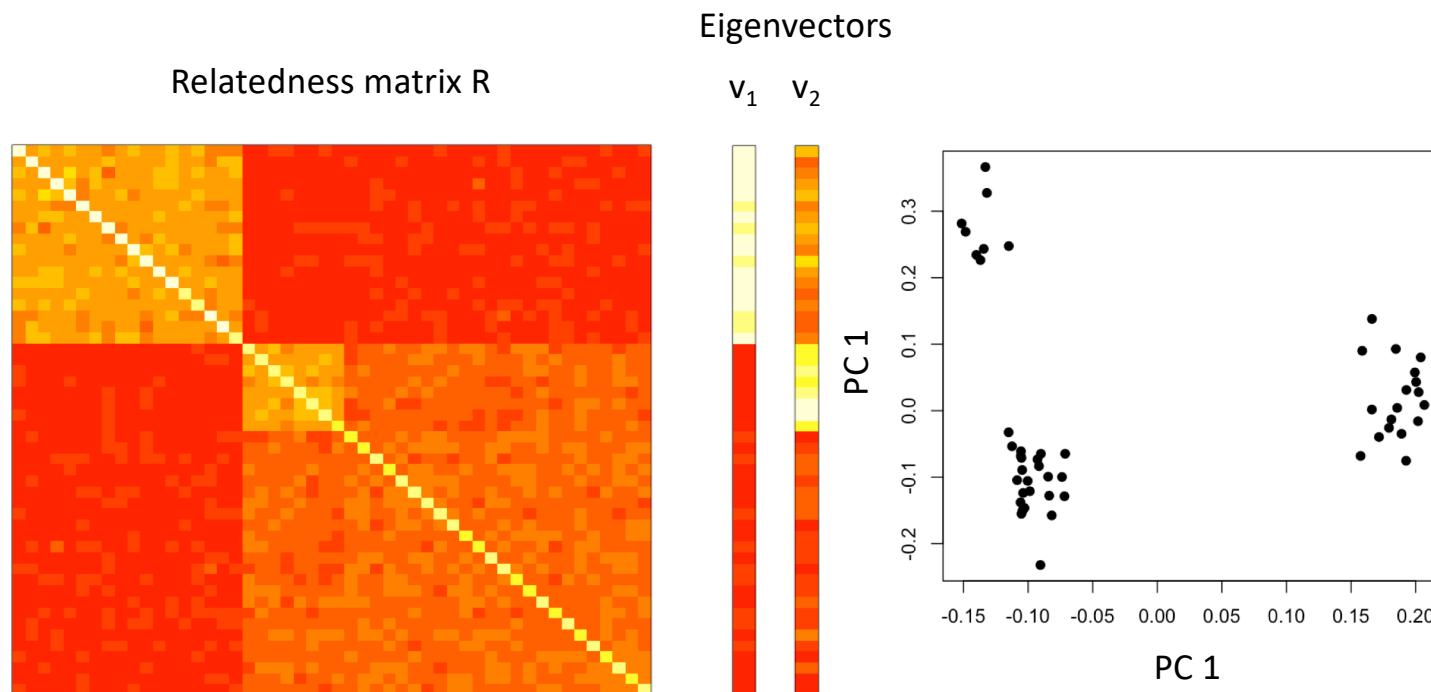
(Simulated data, 50 individuals, 1000 SNPs)



Modified from slide by Gavin Band

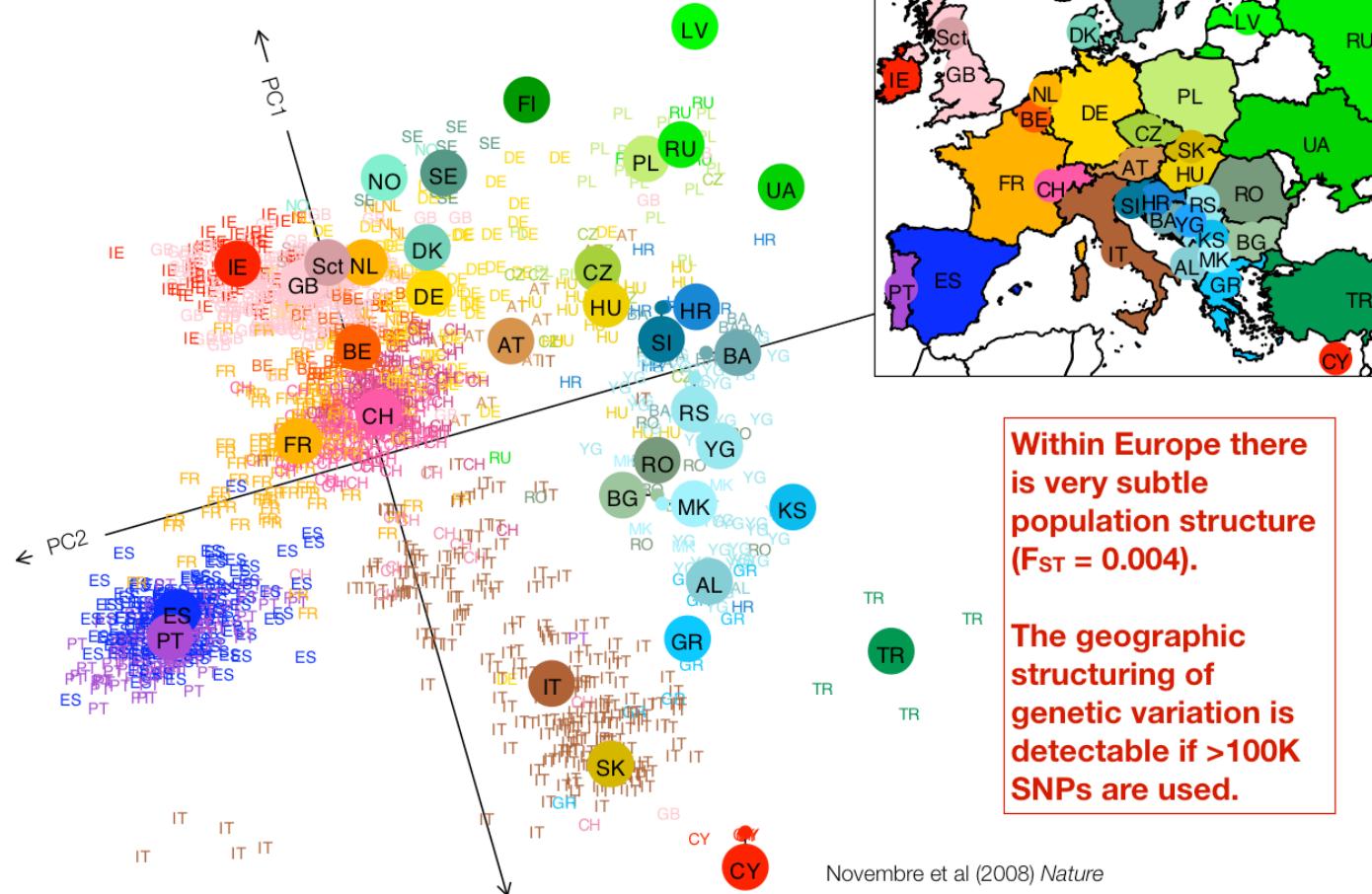
Principal components analysis Example

(Simulated data, 50 individuals, 1000 SNPs)



Modified from slide by Gavin Band

Principal Component Analysis of Europeans



**Within Europe there
is very subtle
population structure
($F_{ST} = 0.004$).**

The geographic structuring of genetic variation is detectable if >100K SNPs are used.

Novembre et al (2008) *Nature*