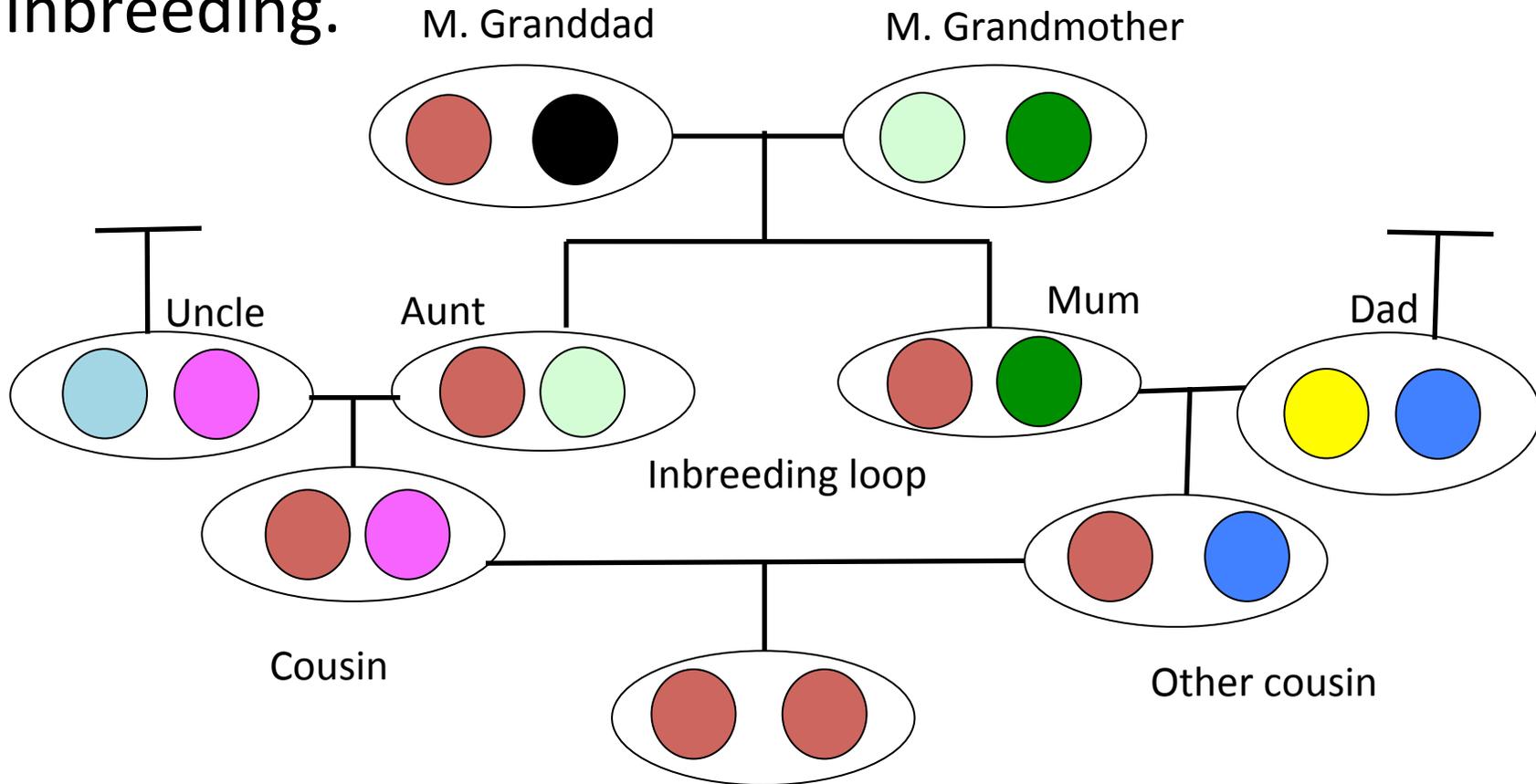


Inbreeding.



What is the probability that the cousins share 0, 1, or 2 alleles identical by descent?

What is probability that the child of first cousins is homozygous by descent?

What's the probability that the child of first cousins is heterozygote?

Notes: Equation 3 & 4, page 9.

Generalized HW

Inbreeding coefficient F =
probability that an individual
inherits two alleles identical by
descent at a locus

AA

$$(1-F)p^2 + Fp$$

Aa

$$(1-F)2p(1-p)$$

aa

$$(1-F)(1-p)^2 + F(1-p)$$

Inbreeding:

Increases proportion of homozygotes

Decreases proportion of heterozygotes

Note no change in
allele frequency.

Only genotype frequency

Estimating Inbreeding coefficients

Notes: Equations 5&6, pages 9-10.



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.

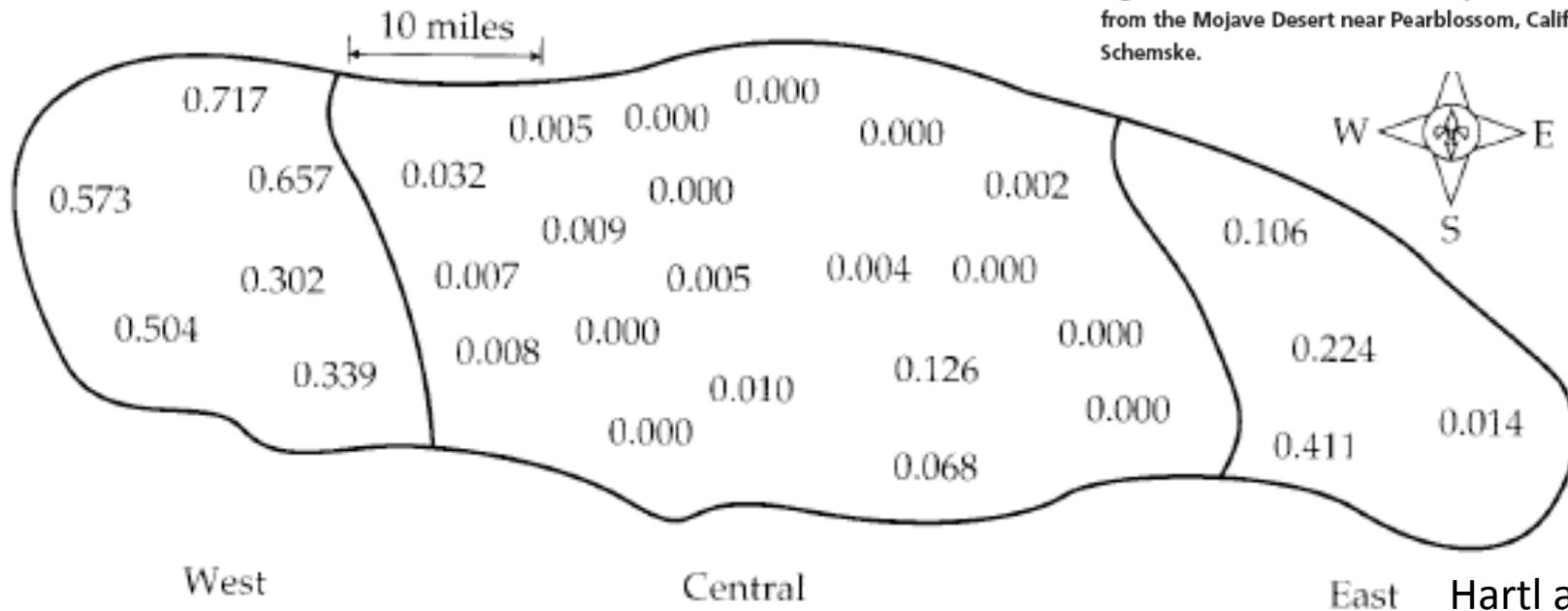
F statistics as a measure of population structure

- 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.



Hartl and Clark

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.)

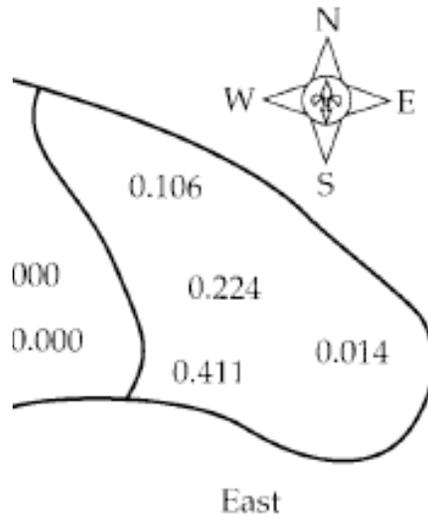


Figure 1. Blue- and white-flowered plants of *Linthanthus parryae* from the Mojave Desert near Pearblossom, California, USA. D. W. Schemske.

Taken from Wright 1943

	allele freq	Heterozygosity ($H_S=2p_Sq_S$)	$F_{ST}= 1-H_S/H_T$
	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} ?

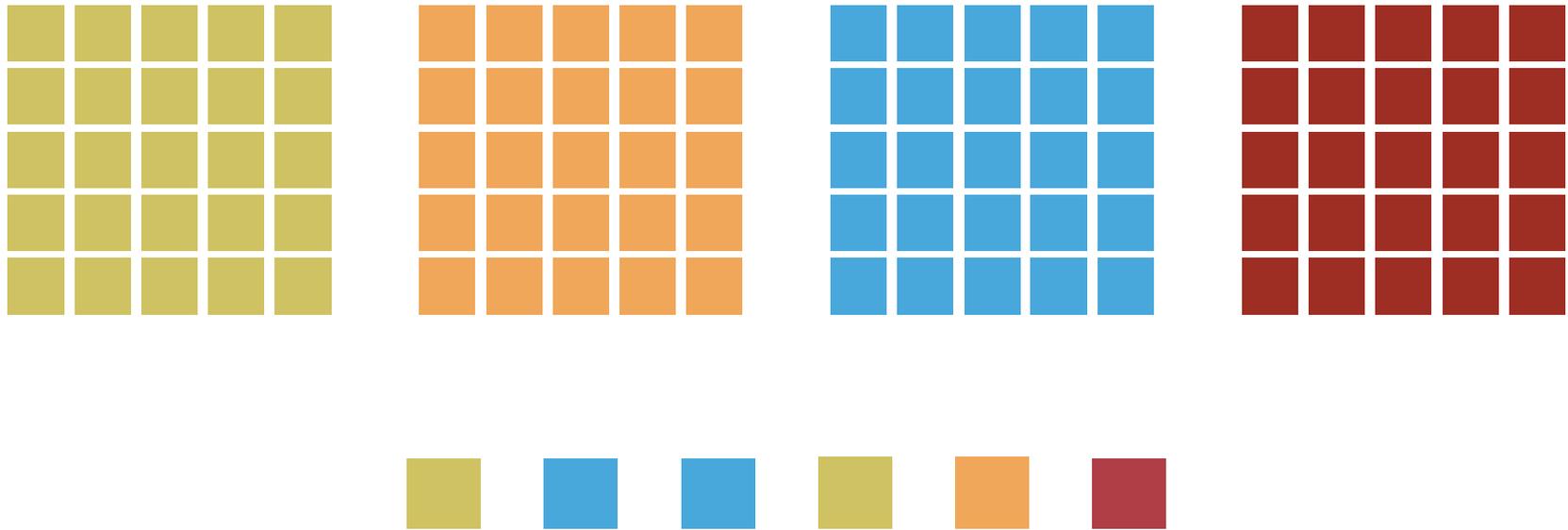
- F_{ST} = proportion of the variance due to between-population (rather than within-population) differences.
- Low values of F_{ST} indicate that on average differences are found among individuals from the same population rather than in one population but not the other.

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.

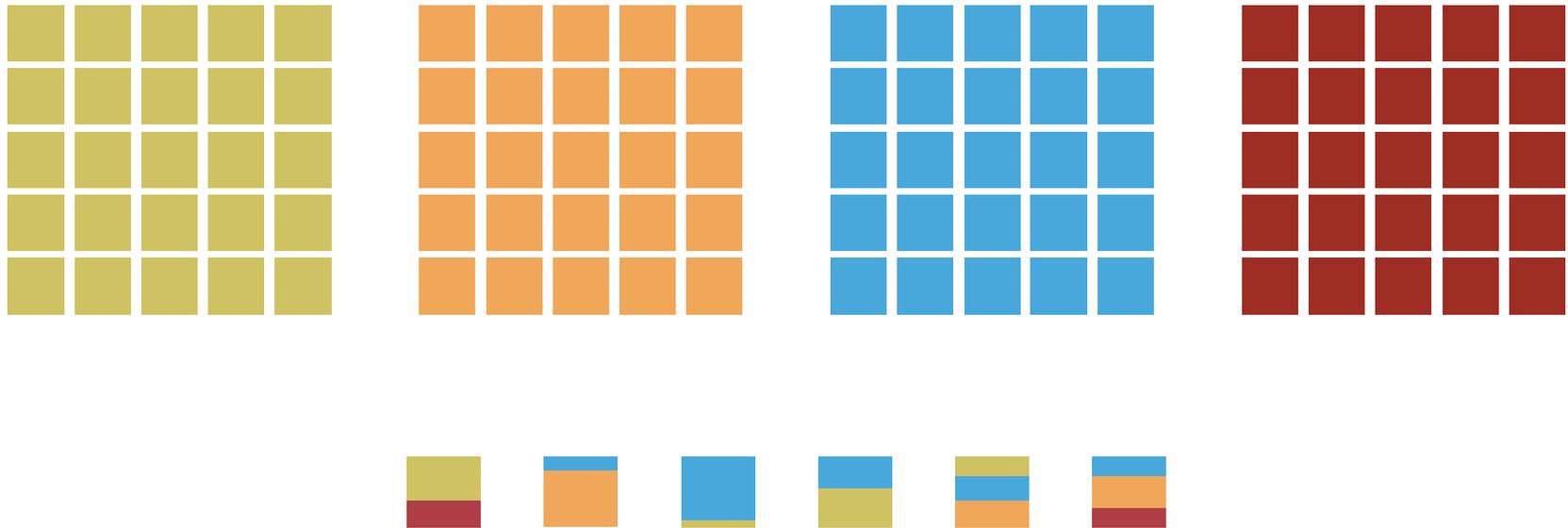
Reference Populations



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

Approaches: Based on Hardy-Weinberg genotype proportions

Reference Populations (known or unknown)



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

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

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

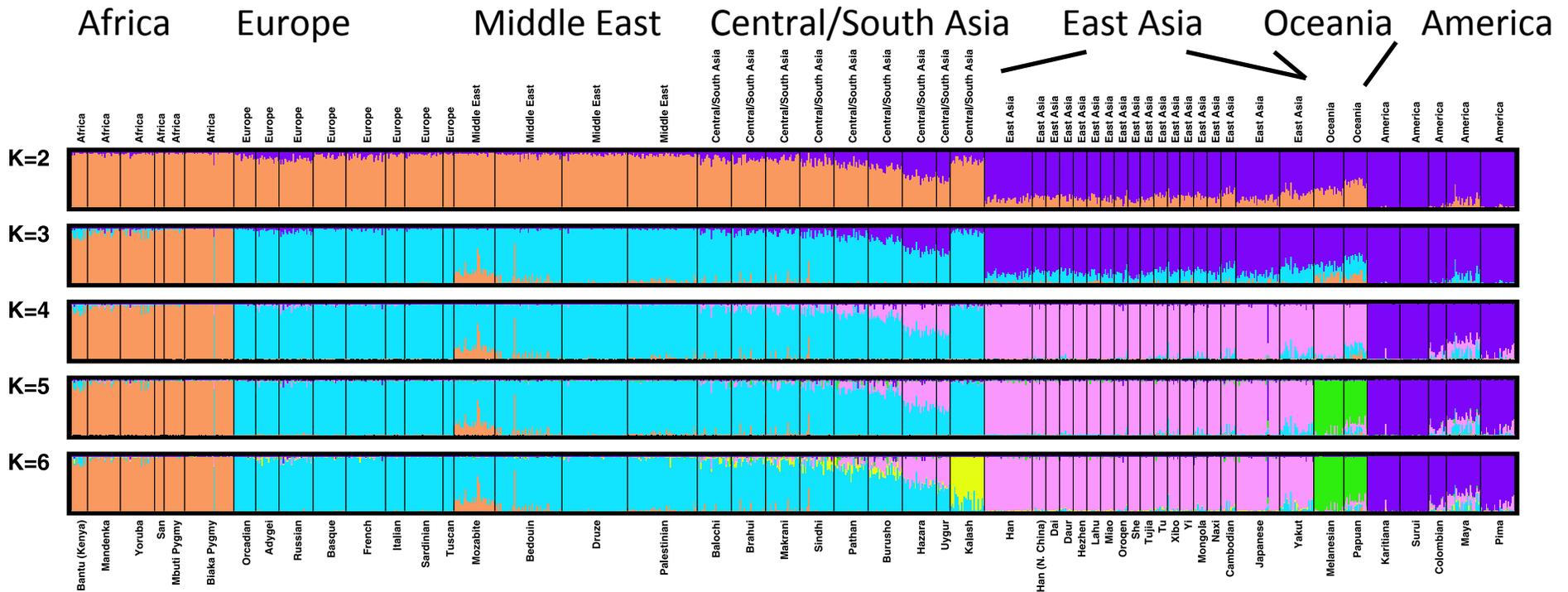


Fig. 1. Estimated population structure. Each individual is represented by a thin vertical line, which is partitioned into K colored segments that represent the individual's estimated membership fractions in K clusters. Black lines separate individuals of different populations. Populations are labeled 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 .