

Population structure

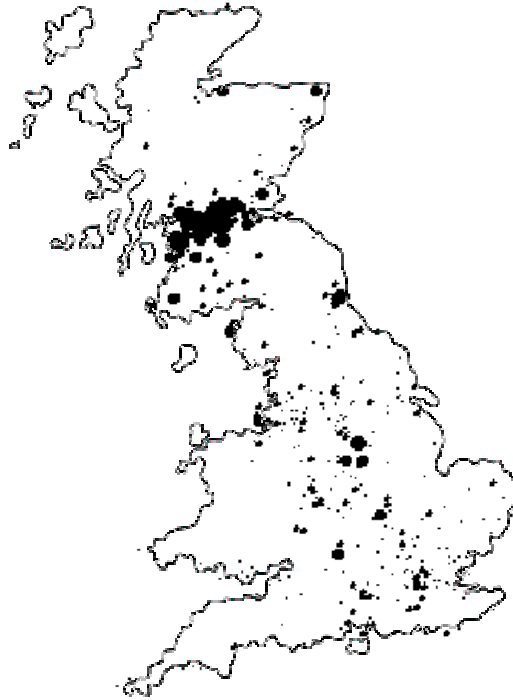
- The evolutionary significance of structure
- Detecting and describing structure
 - Wright's F statistics
- Implications for genetic variability
 - Inbreeding effects of structure
 - The Wahlund effect
 - Drift and founder effects
- Island models of population structure
 - Identity by descent
 - Diffusion methods
 - The coalescent with structure
- Selection in subdivided populations
 - Location adaptation
 - Clines
 - Wright's Shifting-Balance theory

Population structure

Distribution of surname

Hannah

- Non-random location
- Non-random mating

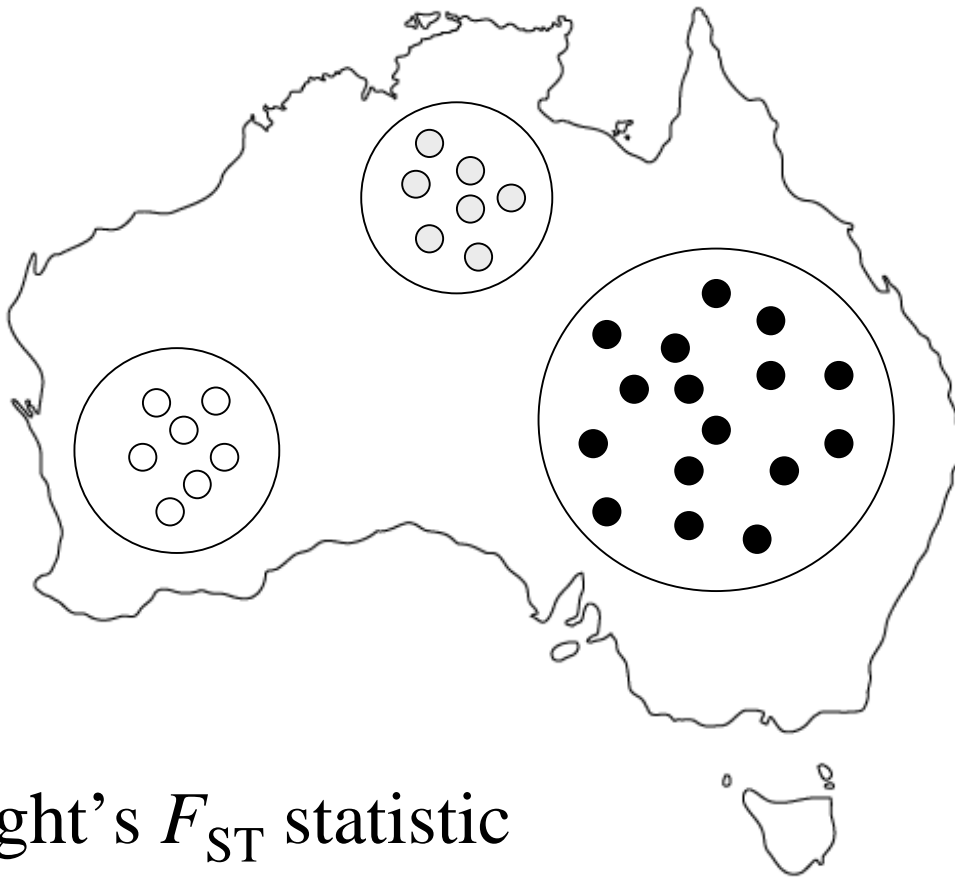


Goodacre and Sykes

Genetic and phenotypic divergence due to

Chance
Selection
Selection plus chance

Detecting and describing genetic structure



Wright's F_{ST} statistic

Heterozygosity over
all populations

$$= \frac{H_T - \overline{H_S}}{H_T}$$

Average heterozygosity
within subpopulations

Testing by permutation



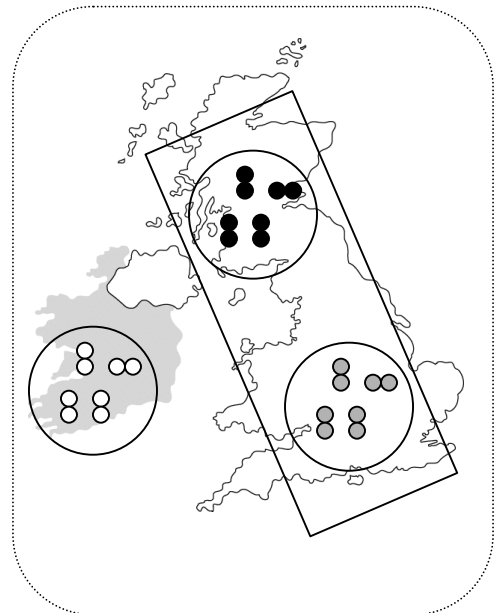
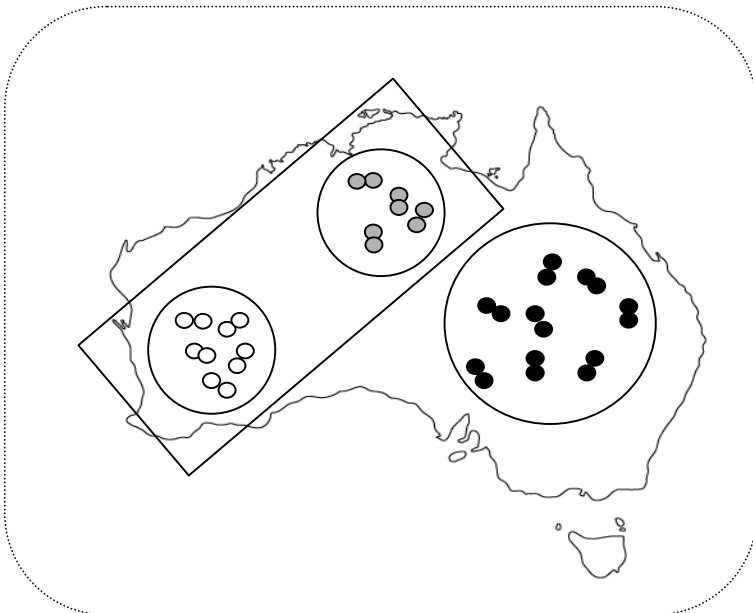
The hierarchical nature of F statistics

- F statistics can be used to contrast structure at different levels

e.g.
$$F_{IS} = \frac{\overline{H}_S - \overline{H}_I}{\overline{H}_S}$$

← Average within-individual heterozygosity

⇒ measure of inbreeding



$$H_{Individual} < H_{Subpopulation} < H_{Population} < H_{Region} < H_{Total}$$

F_{ST} in natural populations

Allozymes

Organism	H_T	\bar{H}_S	F_{ST}
Human (major races)	0.130	0.121	0.069
Human (Yanomama)	0.039	0.036	0.077
House mouse	0.097	0.086	0.113
Jumping rodent	0.037	0.012	0.676

Nei (1975)

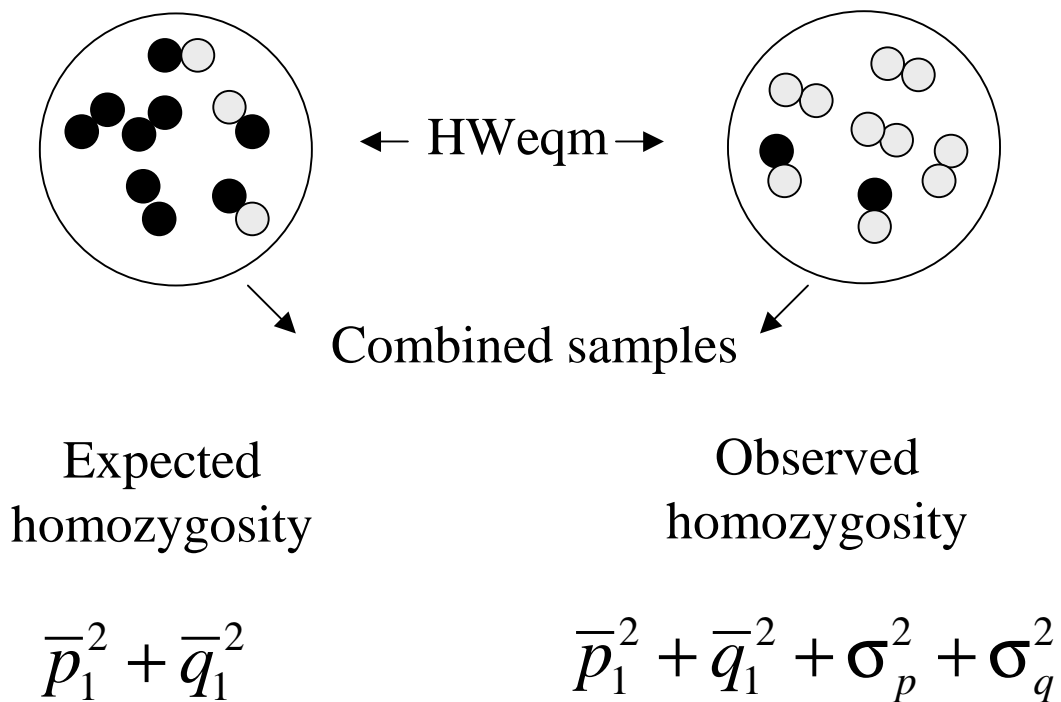
SNPs

Organism	H_T	\bar{H}_S	F_{ST}
Human (major races)	0.195	0.201	0.067
<i>Drosophila melanogaster</i> ^a	0.0154	0.0151	0.023

^aBased on pairwise diversity

The inbreeding effect of population structure

- Differences in allele frequency between populations lead to an excess of homozygotes

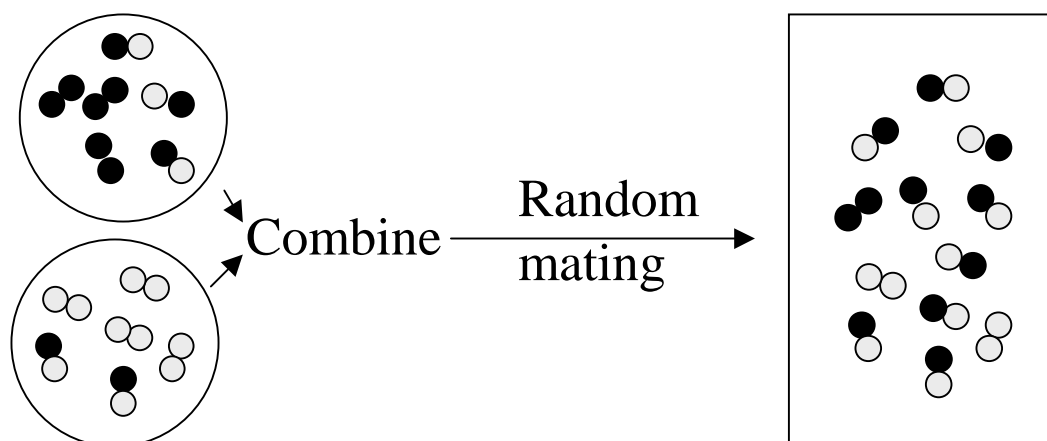


Heterozygosity = 1 - Homozygosity

$$\Rightarrow F_{ST} = \frac{\bar{F}_S - F_T}{1 - F_T} = \frac{\sigma_p^2 + \sigma_q^2}{1 - \bar{p}^2 - \bar{q}^2}$$

The Wahlund effect

- Increase in heterozygosity following mixing of isolated populations



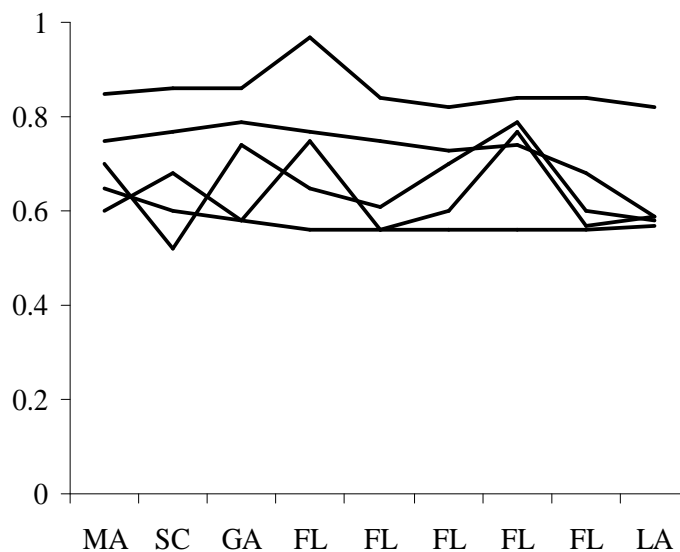
- Medical implications for disease incidence in admixed populations
 - Recessive disease reduced by mixing

Disease	High risk population	Disease allele frequency
Cystic fibrosis	Caucasians	0.022
Albinism	Hopi	0.07
Tay-Sachs disease	Ashkenazi Jews	0.013

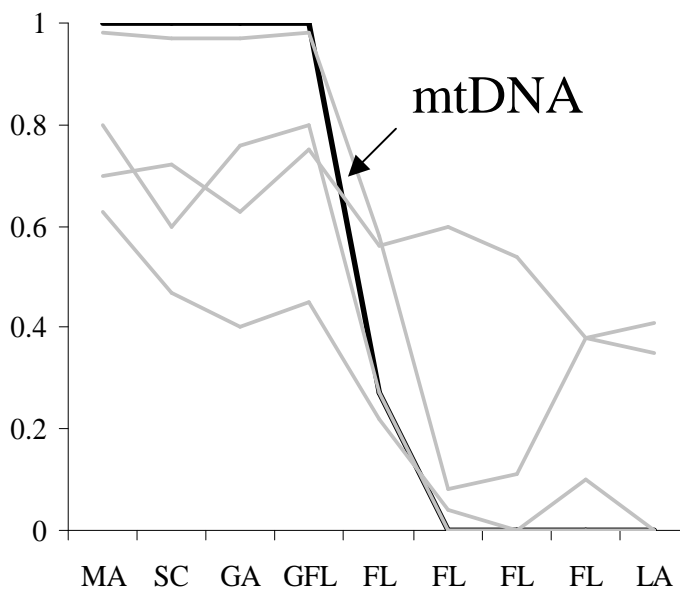
Differences between allozymes and DNA?

- American oysters (*Crassostrea virginica*)

Allozymes



DNA



Avise (1994)

Differences between allozymes?

Locus	F_{ST}	Checkersport butterfly
<i>pgm</i>	0.028	<i>Euphydryas editha</i>
<i>pgi</i>	0.052	
<i>got</i>	0.017	McKechnie et al. 1975
<i>ak</i>	0.062	
<i>bdh</i>	0.034	
α - <i>gpdh</i>	0.027	
<i>to</i>	0.035	
<i>hk</i>	0.291	← Unusually high differentiation

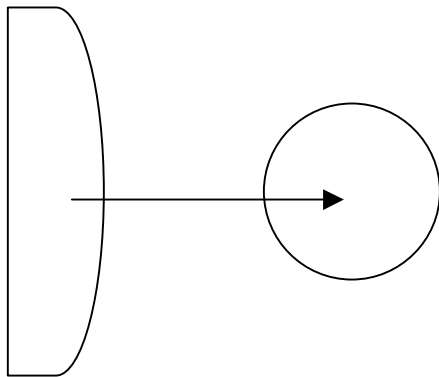
Problems with F_{ST}

- Arbitrary *a priori* choice of structure to test
- High sampling variance when polymorphism low
- Throws away much information

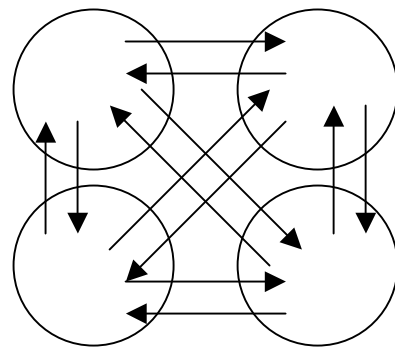
Population genetics models of structure

- Quantify relationship between genetic drift, selection and population differentiation

Island model



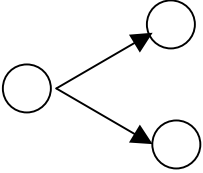
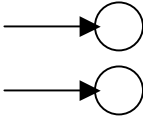
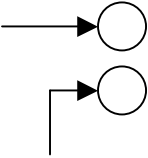
n-island model



- Assumptions

- Infinite mainland population (island)
- Equal population size (n-island)
- Constant population size
- Proportion m of population replaced migrants each generation
- Symmetric migration (n-island)

Identity by descent in the island model

Event	Same parent	Different parents	Migration
			
Identity	1	f_{t-1}	0
Probability	$1/2N_e$	$1 - 1/2N_e - 2m$	$2m$

At equilibrium

$$f = \frac{1}{1 + 4N_e m}$$

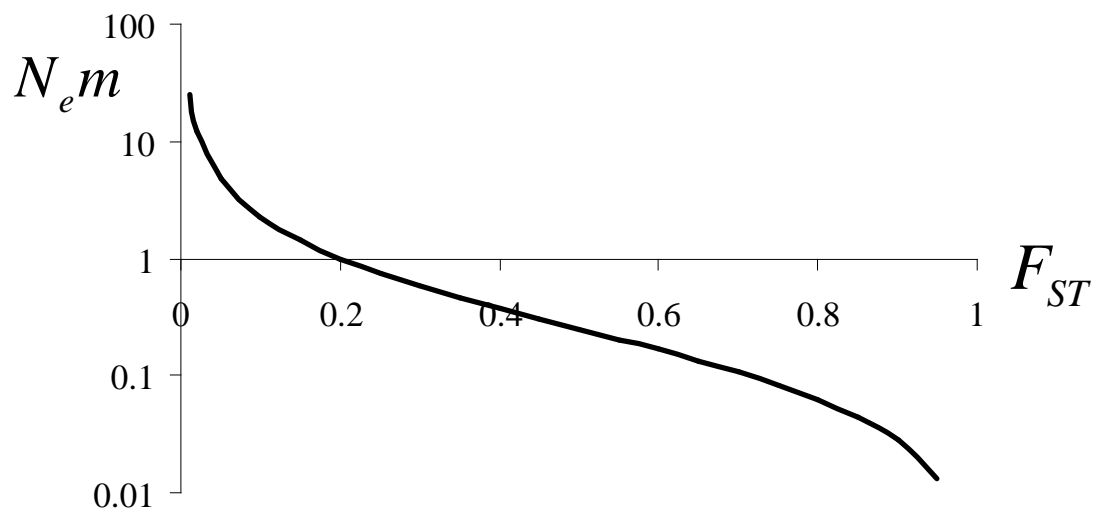
$4N_e m = 2 \times$ Number of migrants per generation

⇒ Only a few migrants each generation are required to prevent a build up of identity within the island population

Relationship between F_{ST} and migration rate

$$E[F_{ST}] \approx \frac{1}{1 + 4N_e m}$$

- Can estimate scaled migration rate from estimated F_{ST} (assuming equilibrium, etc.)



E.g. in humans, $F_{ST} \approx 0.067$

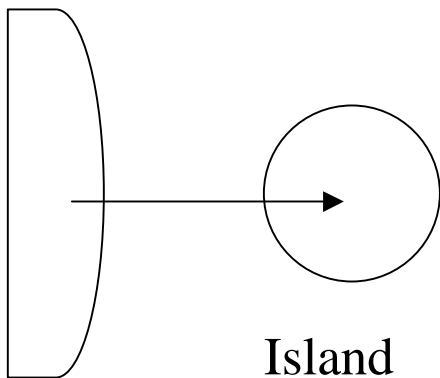
$\implies N_e m \approx 3.5$

NB: This is NOT a good estimator – do not trust the answer!

Wright's diffusion model for allele frequencies with migration

Mainland

frequency = x_m



Island

frequency = x

Deterministic

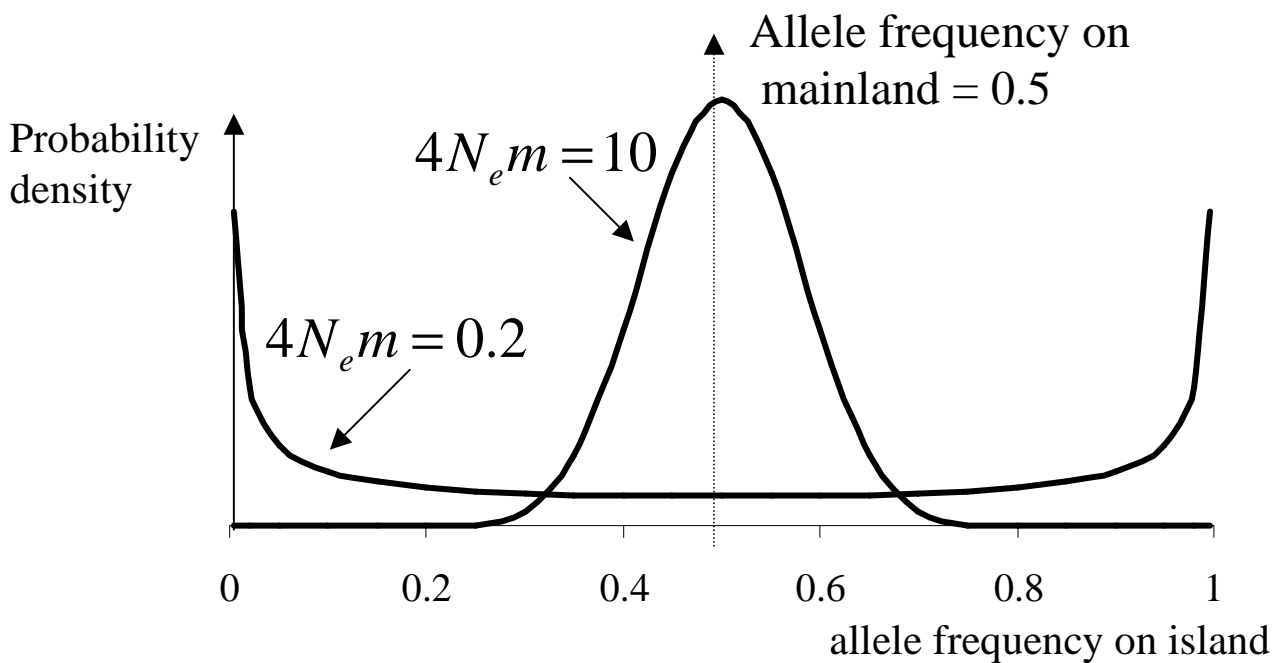


$$M_{\delta x} = m(x_m - x)$$

$$V_{\delta x} = \frac{x(1-x)}{2N_e}$$

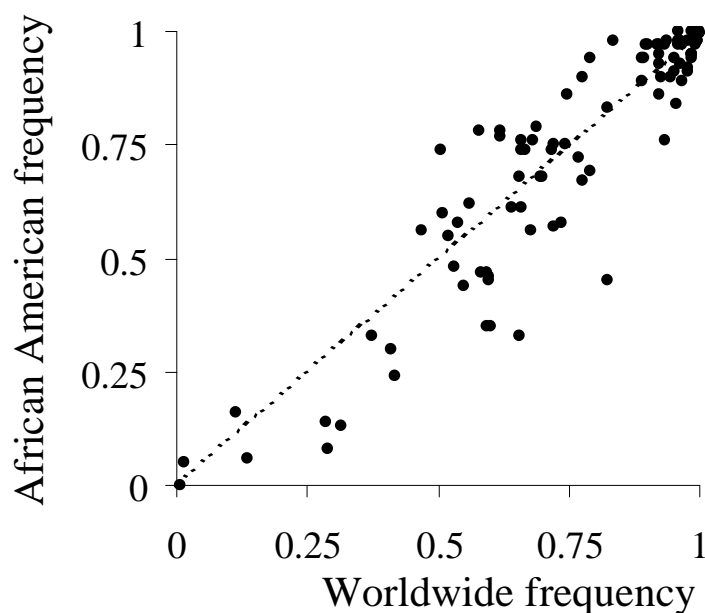
Drift

Wright (1951)

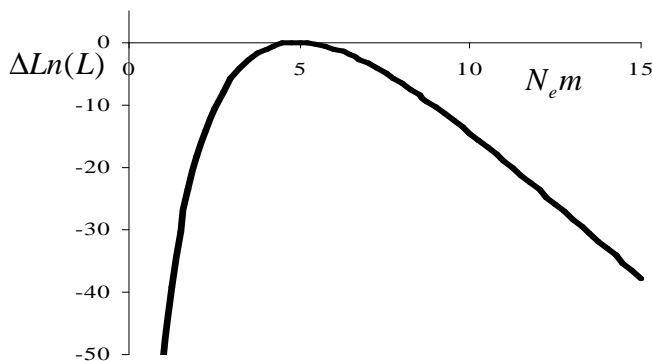


Example: SNP frequencies in African Americans

- Goddard et al. (2000)
 - 114 SNPs in 33 genes
 - 190 African Americans sampled



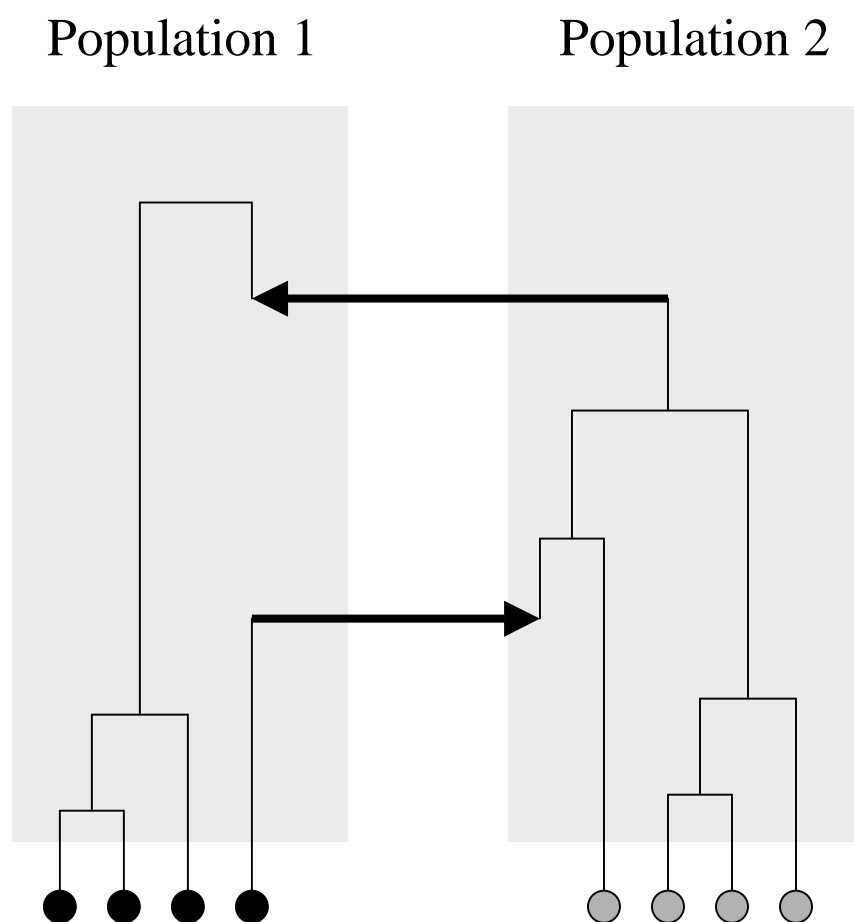
- Likelihood estimation of $N_e m$ from sample
 - assume independence between SNPs



$$\widehat{N_e m} = 5.0$$

The coalescent in structured populations

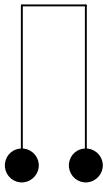
- Two-island model



$$\Pr\{\text{coalescence}\} = \frac{n_i(n_i - 1)}{4N_e}$$

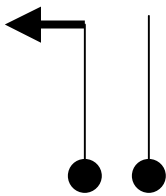
$$\Pr\{\text{migration}\} = n_i m$$

The time to coalescence for two sequences sampled from the same population



Pr{1st event is a coalescence}

$$\frac{1/2N_e}{1/2N_e + 2m} = \frac{1}{1 + 4N_e m}$$



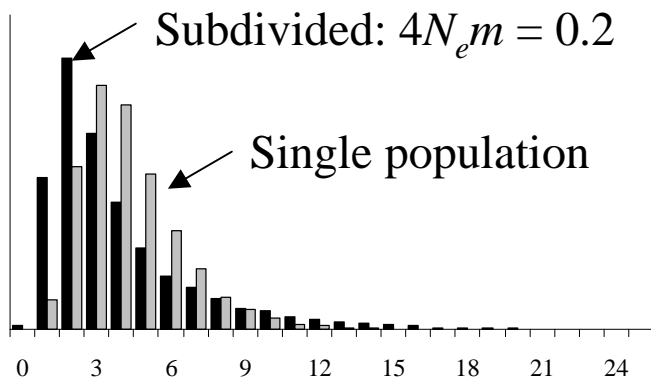
Pr{1st event is a migration}

$$\frac{2m}{1/2N_e + 2m} = \frac{4N_e m}{1 + 4N_e m}$$

Expected time to coalescence = $4N_e$

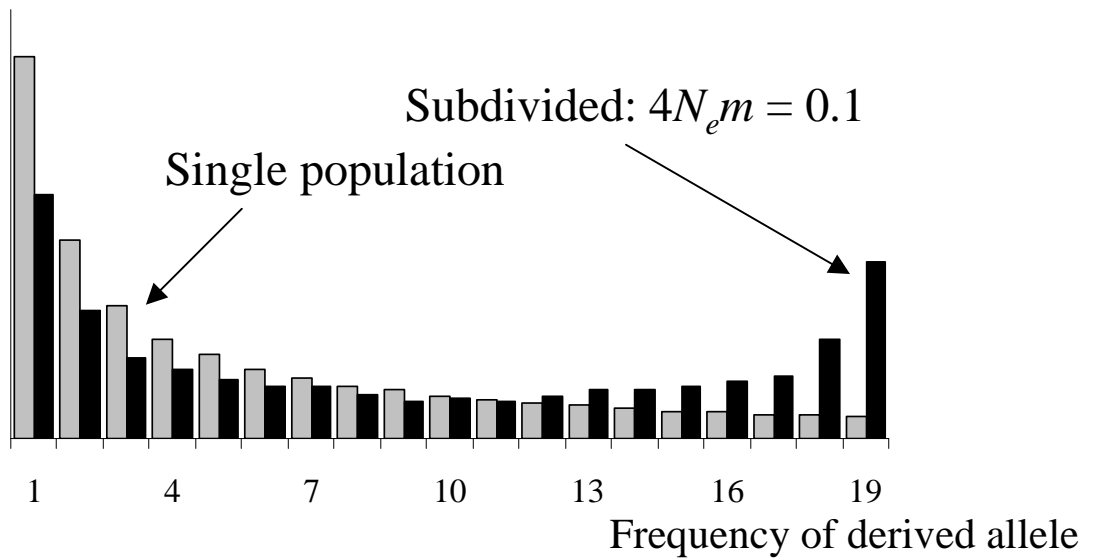
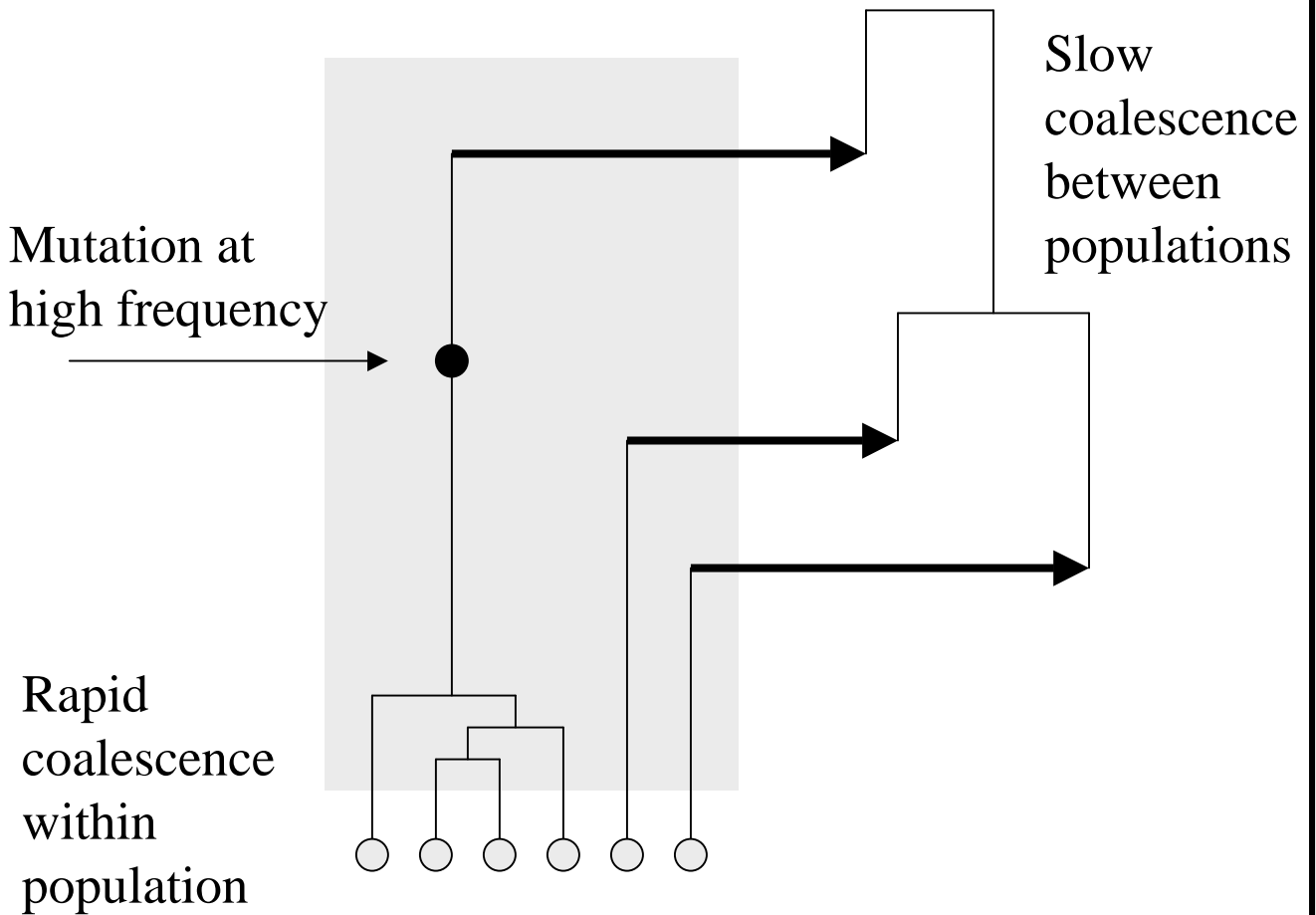
⇒ For expected pairwise diversity (within population) $\begin{matrix} \bigcirc N_e \\ \bigcirc N_e \end{matrix} \equiv \bigcirc 2N_e$

BUT



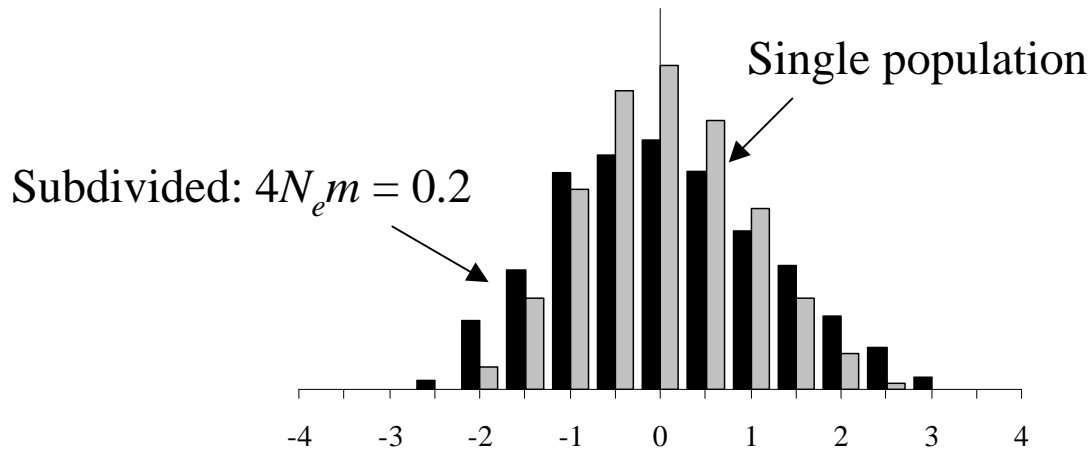
Variance affected by population structure

Effect on allele frequency spectrum

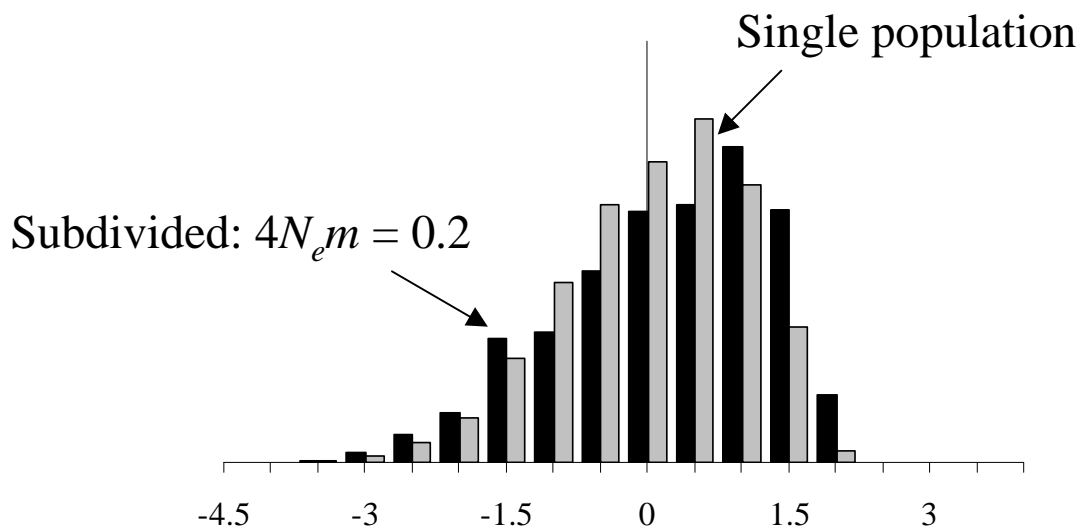


Effect on neutrality statistics within populations

- Tajima's D statistic



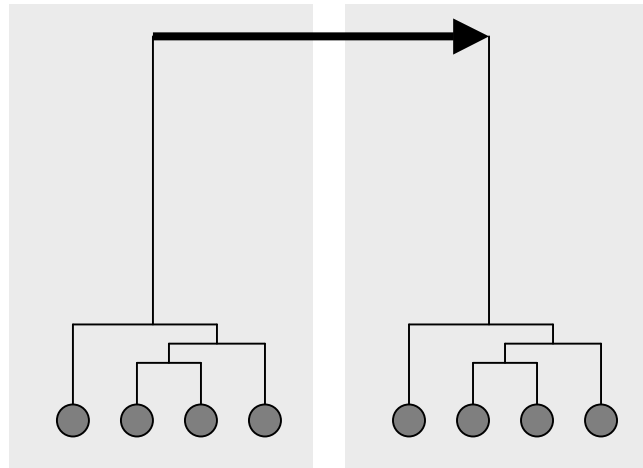
- Fu and Li D statistic



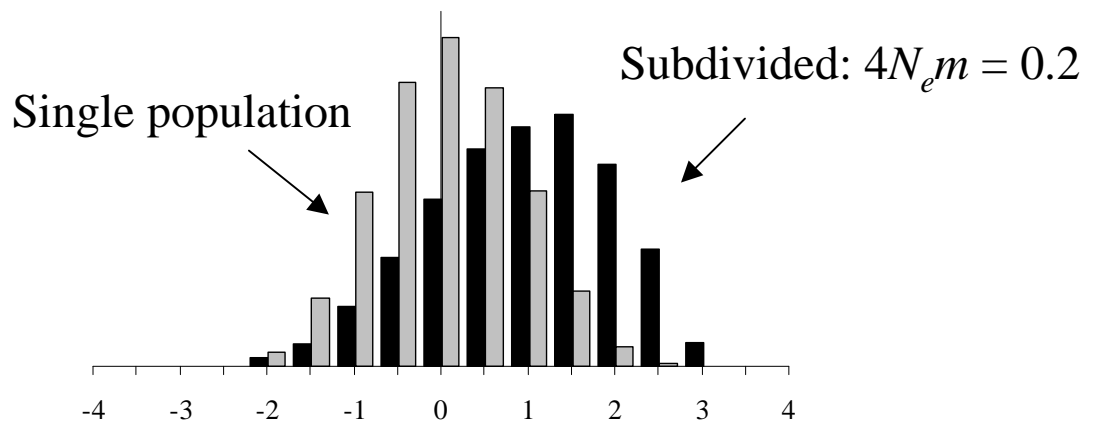
⇒ Main effect is to increase the variance

Other statistics (e.g. Fay and Wu, 2000) more sensitive

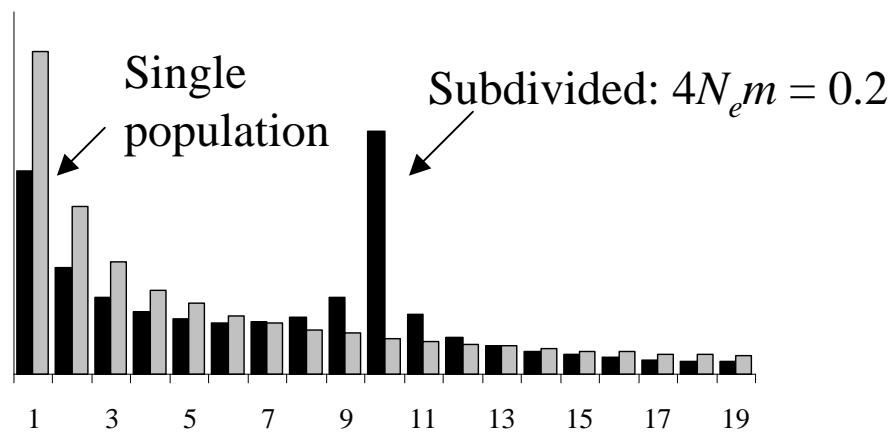
Effect on polymorphism between populations



- Tajima's D statistic



- Frequency distribution

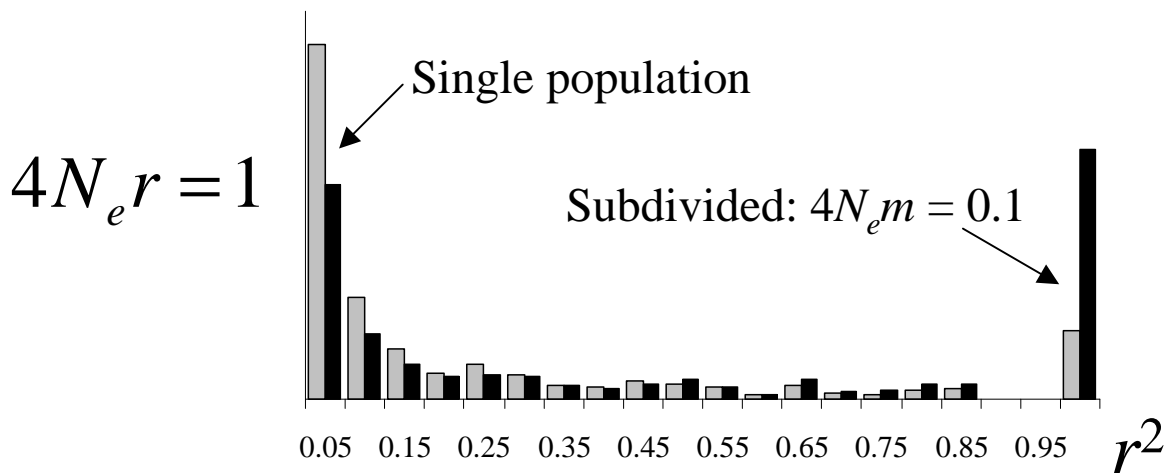


Effect on linkage disequilibrium

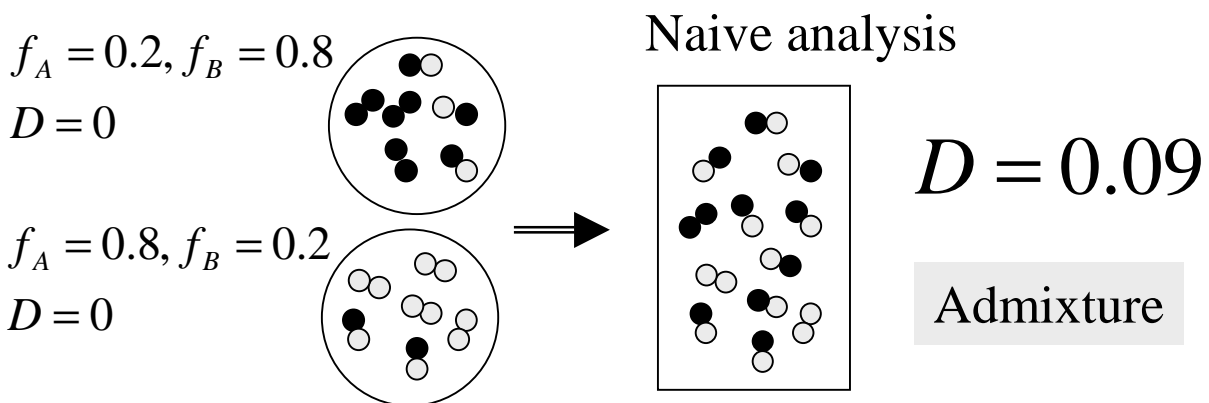
- Linkage disequilibrium measures correlations between alleles at different loci

$$D = f_{AB} - f_A f_B$$

- Population structure increases linkage disequilibrium between linked loci

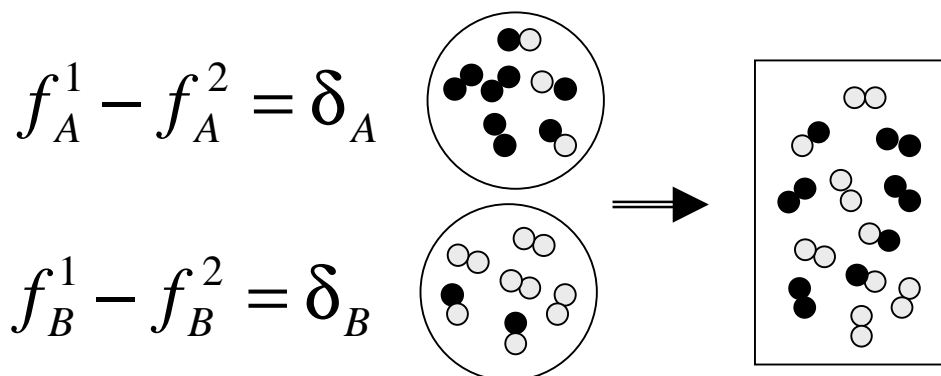


- Population structure creates linkage disequilibrium between unlinked loci in different populations



Admixture dynamics

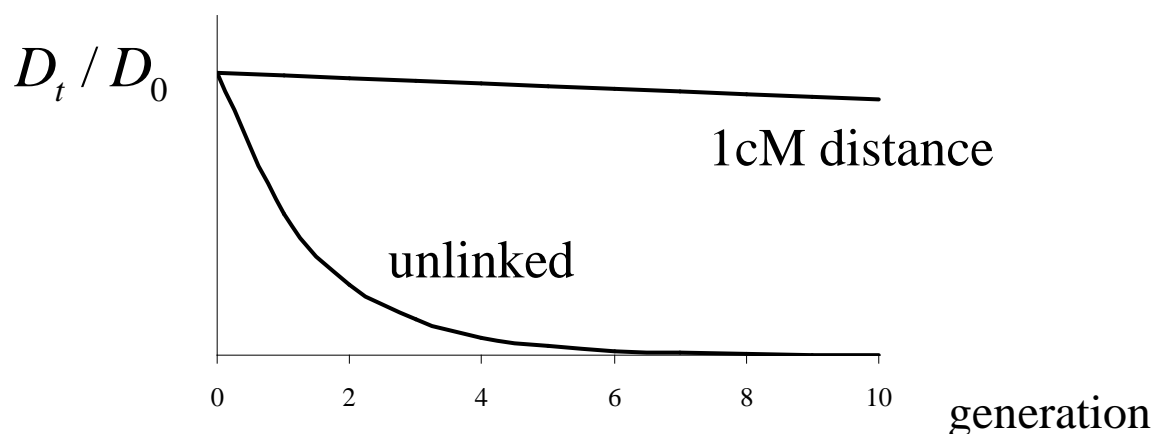
- Combination of two previously separated populations



- Over time random mating returns population to equilibrium

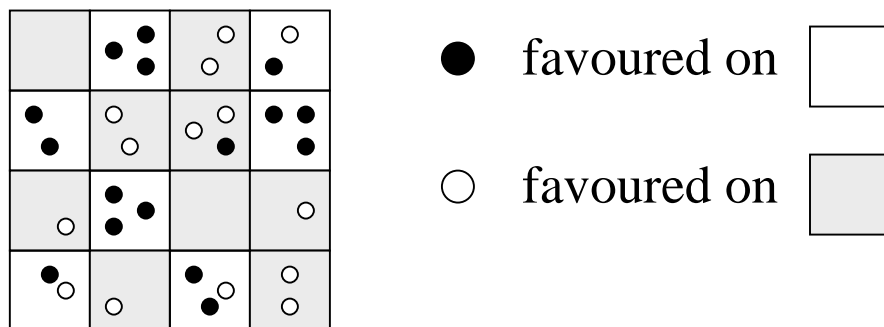
$$D_0 = \frac{1}{4} \delta_A \delta_B \quad D_t = D_0 (1 - r)^t$$

- Disequilibrium between unlinked loci can persist for several generations, while Hardy-Weinberg equilibrium is achieved instantly



Selection in a subdivided population

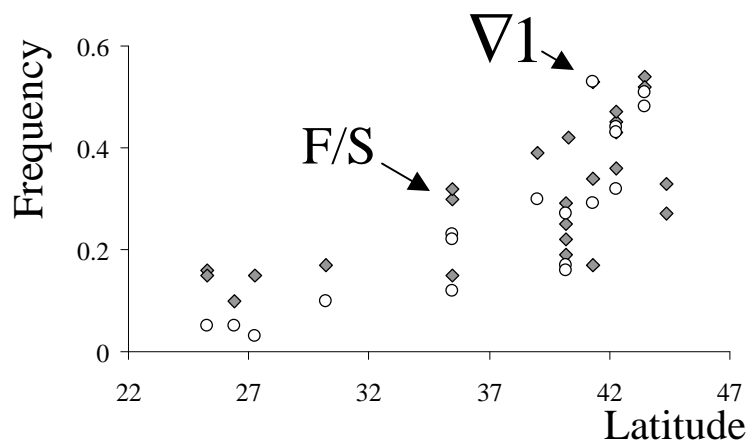
- Maruyama (1970)
 - The fixation probability of an unconditionally beneficial mutation is unaffected by population structure ($P_{fix} \approx 2s$)
- Levene (1953)
 - Environmental heterogeneity can maintain genetic polymorphism



- BUT
 - If migration high, selection has to be strong and finely balanced to habitat frequencies to maintain polymorphism
- Low migration rates can promote local adaptation
 - Heavy metal tolerance in plants
 - Melanism in the peppered moth
 - Milk tolerance in humans

Selection at different scales

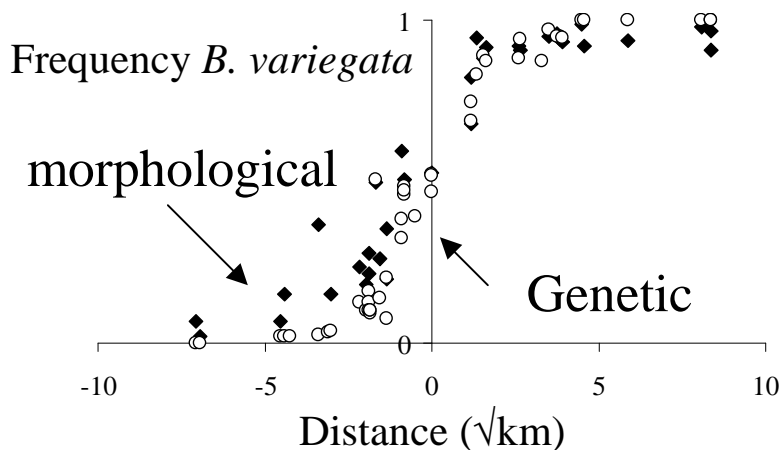
- Evidence for local adaptation from gradients in allele frequency : clines
- Continental clines in *Adh* activity and allozyme variation in *Drosophila*



⇒ Driven by scale of environmental heterogeneity

Berry & Kreitman (1993)

- Clines in genetic and morphological characters in the toad *Bombina*

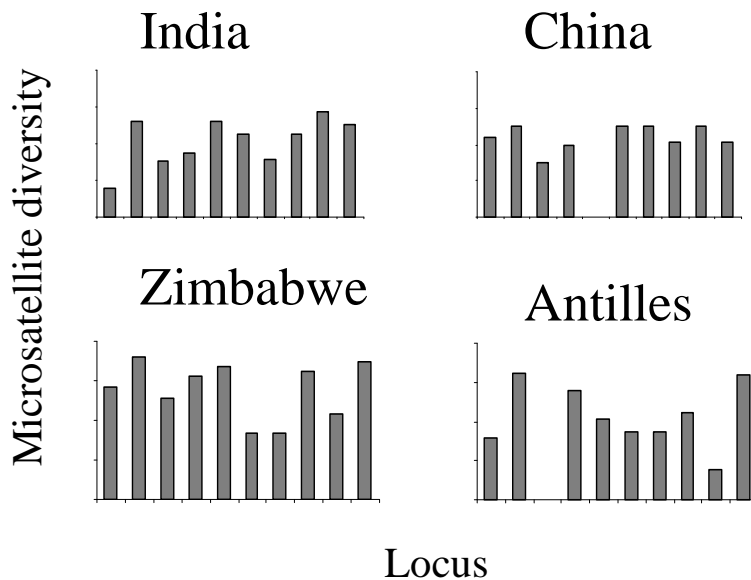


⇒ Balance between selection against hybrids and migration, following secondary contact

Szymura & Barton (1991)

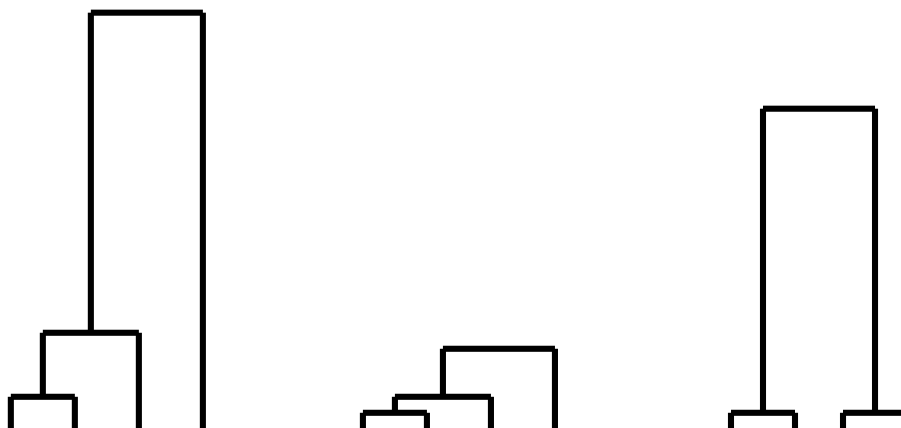
Indirect evidence for local adaptation?

- Local hitch-hiking?



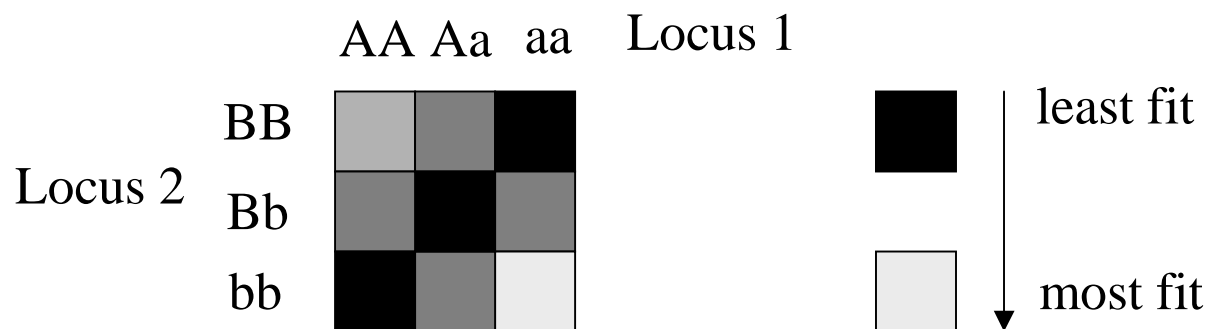
Schlötterer *et al.* (1997)

- But the structured coalescent also leads to variation in coalescence times

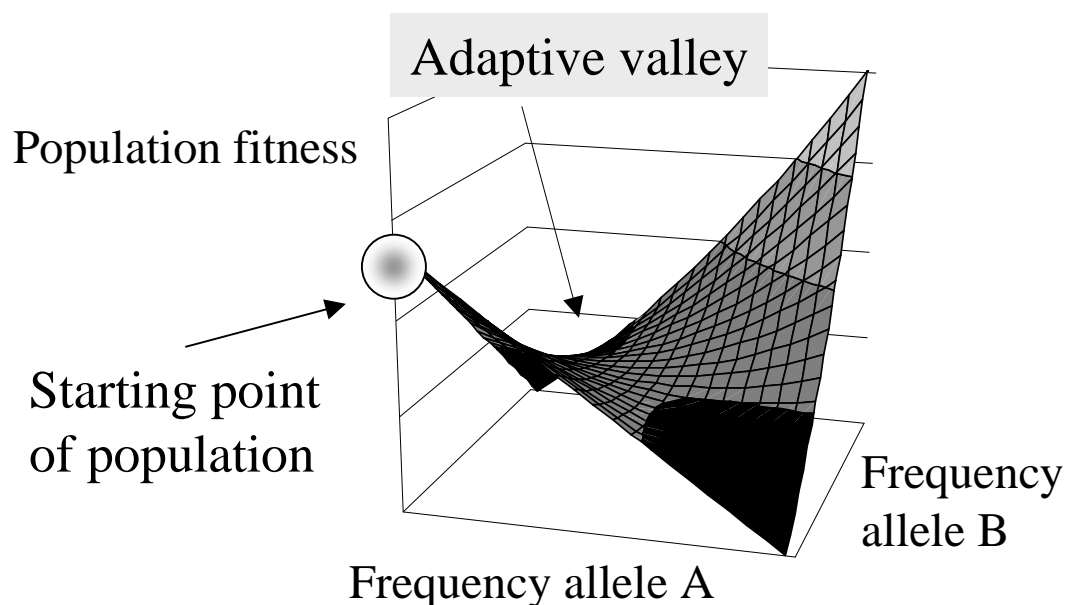


The interaction between selection, gene flow and genetic drift

- Wright's Shifting Balance theory
- Epistasis between alleles at different loci

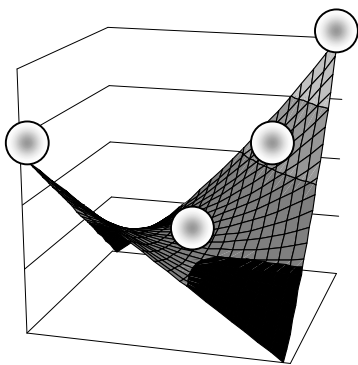


- The adaptive landscape
 - Epistasis creates adaptive valleys between peaks of fitness



The Shifting Balance theory

- Drift allows population to cross adaptive valley due to stochastic processes in finite populations



⇒ Subpopulations are natural experiments, allowing species to evolve across complex adaptive landscapes

- Evidence for widespread epistasis?
 - F2 hybrid breakdown
 - Coadapted gene complexes
- Theoretical issues
 - Very difficult for a population that has crossed a valley to spread throughout rest of population
 - The interaction between epistatic selection and genetic drift may be important in reproductive isolation
 - e.g. recessive epistatic interactions important in Haldane's rule of unisexual hybrid sterility

Future directions

- Theoretical and statistical issues
 - Methods for discriminating between local adaptation and chance effects of coalescence in a structured population
 - The relationship between population structure and linkage disequilibrium
 - Selection on polygenic traits in subdivided populations
- Empirical challenges
 - Describing patterns of gene diversity at many loci across genomes (from an well-chosen sample)
 - Comparing differentiation for different types of mutation (e.g. silent v replacement)
 - Mapping genetic variation to phenotypic variation