An introduction to population genetics

Date	Topic	
23rd Jan	An introduction to population genetics	GM
30th Jan	Neutral mutations in populations	GM
6th Feb	The coalescent	GM
13th Feb	Natural selection	GM
20th Feb	Human population genetics	MP
27th Feb	Recombination	PF
6th March	Population structure	GM
13th March	Medical applications of population genetics	JP

GM	Gil McVean	MP	Molly Prseworski
PF	Paul Fernhead	JP	Jon Pritchard

Books

Crow JF &Kimura M. 1970. **An introduction to population genetics theory**. Harper and Row, New York.

Gillespie JH. 1998. **Populations genetics: a concise guide**. The Johns Hopkins University Press, Baltimore.

Hartl DL & Clark AG (1989). **Principles of population genetics**. Sinauer Associates, Sunderland, Mass.

The early history of population genetics

Date	Event
1859	Darwin's *Origin of Species*
1856-63	Mendel's experiments on peas
1900	Rediscovery of Mendel's laws
1909	Nilsson-Ehle's experiments on wheat
1912-1920	Pearl, Jennings and Wright's work on inbreeding
1915	Morgan's experiments on *Drosophila*
1918	Fisher's paper on phenotypic correlations between relatives
1918	Sturtevant's artificial selection experiments on *Drosophila*
1930	Fisher's *The Genetical Theory of Natural Selection* (Fundamental theorem)
1931	Wright's *Evolution in Mendelian populations*
1932	Haldane's *The Causes of Evolution*
1955	Kimura diffusion equation solution to the distribution of allele frequencies

Definitions

Gene or locus

Molecular: Open reading frame and associated regulatory elements.

Classical genetic: Chromosomal region to which a phenotypic mutation can be mapped.

Evolutionary: A stretch of hereditary material sufficiently small such that it is not broken up by recombination, and which can be acted on by natural selection (the unit of selection).

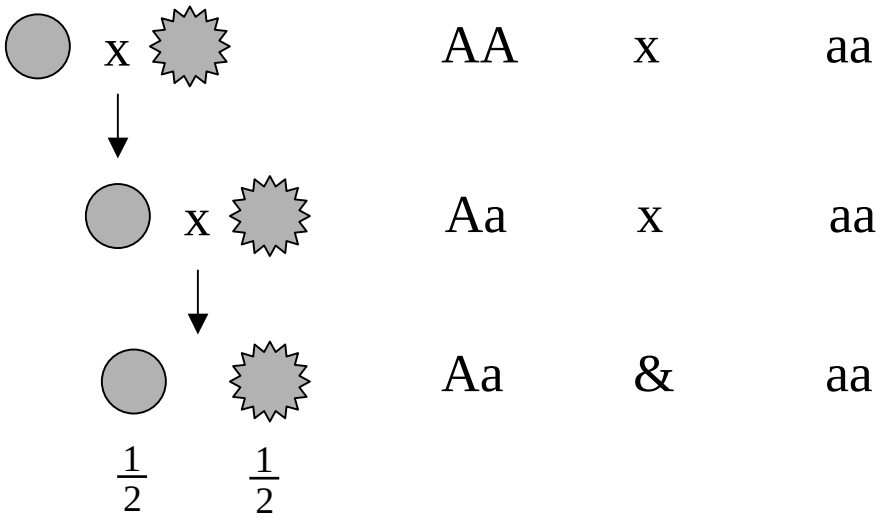
Allele

One of two or more possible forms of gene (locus).

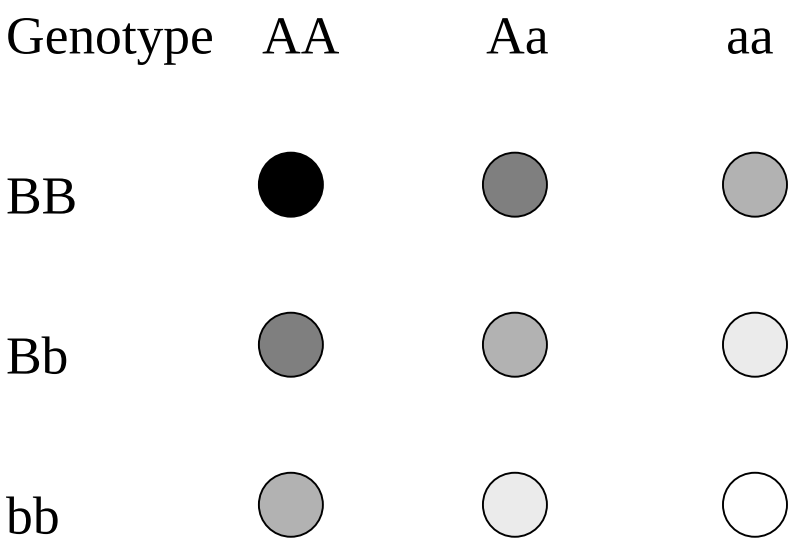
Polymorphism

The presence of multiple forms in natural populations

Mendel's peas

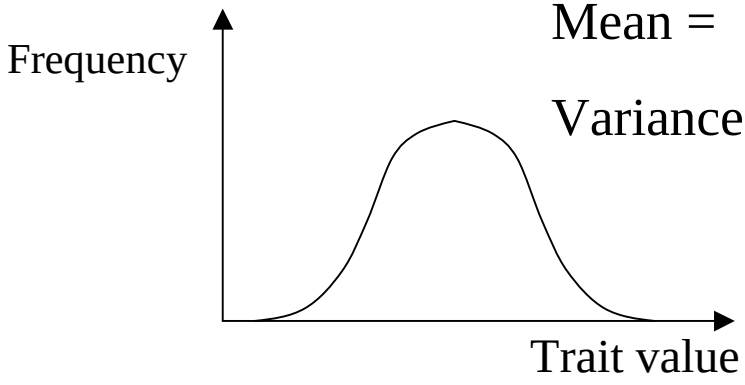


Nilsson Ehle's wheat



Quantitative trait variation

- Three types of quantitative trait
  - Continuous (weight, height, milk yield)
  - Meristic (bristle number in *Drosophila*)
  - Discrete with continuous liability (disease susceptibility)



$$\text{Mean} = \frac{1}{N}\sum_i x_i = \mu$$

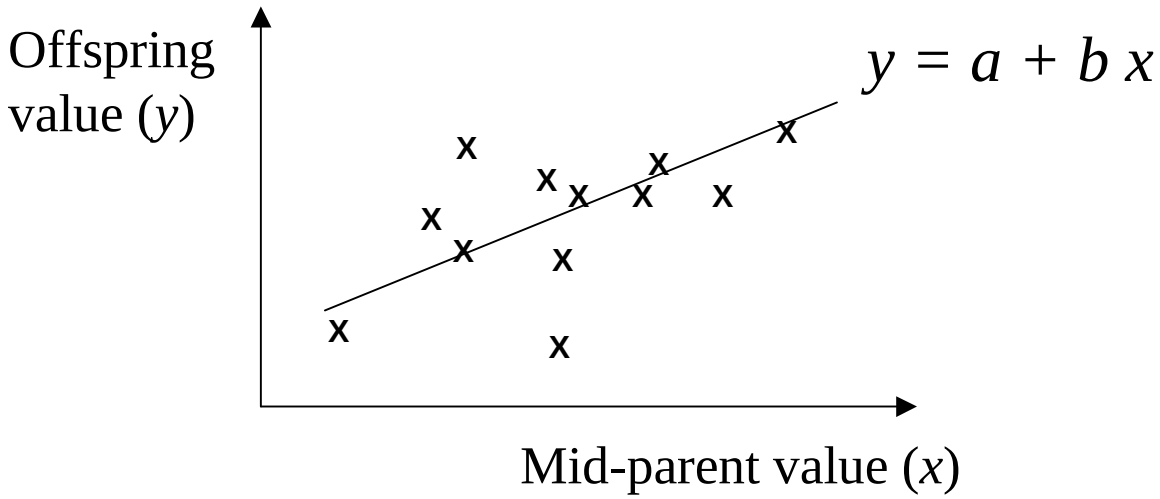
$$\text{Variance} = \frac{1}{N}\sum_i (x_i - \mu)^2 = \sigma^2_P$$

$$\sigma^2_P = \sigma^2_A + \sigma^2_D + \sigma^2_I + \sigma^2_E$$

Phenotypic | Additive genetic | Dominance | Epistatic | Environmental

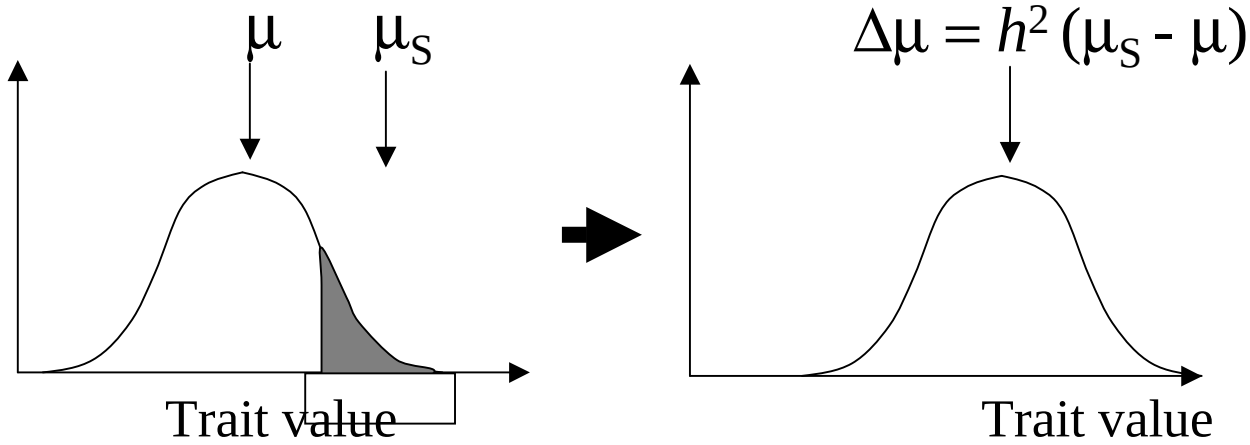
Genetic

Estimating the genetic component of quantitative traits

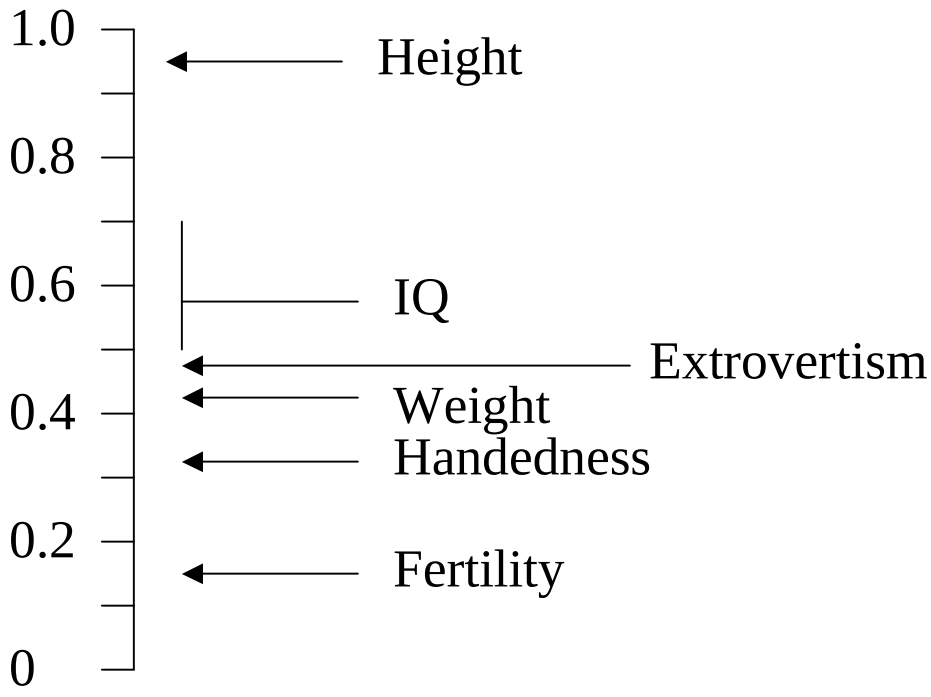


$$b = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)} = h^2 = \frac{\sigma_A^2}{\sigma_P^2}$$

Selection response



Heritabilities of human traits



Twin concordance in human disease

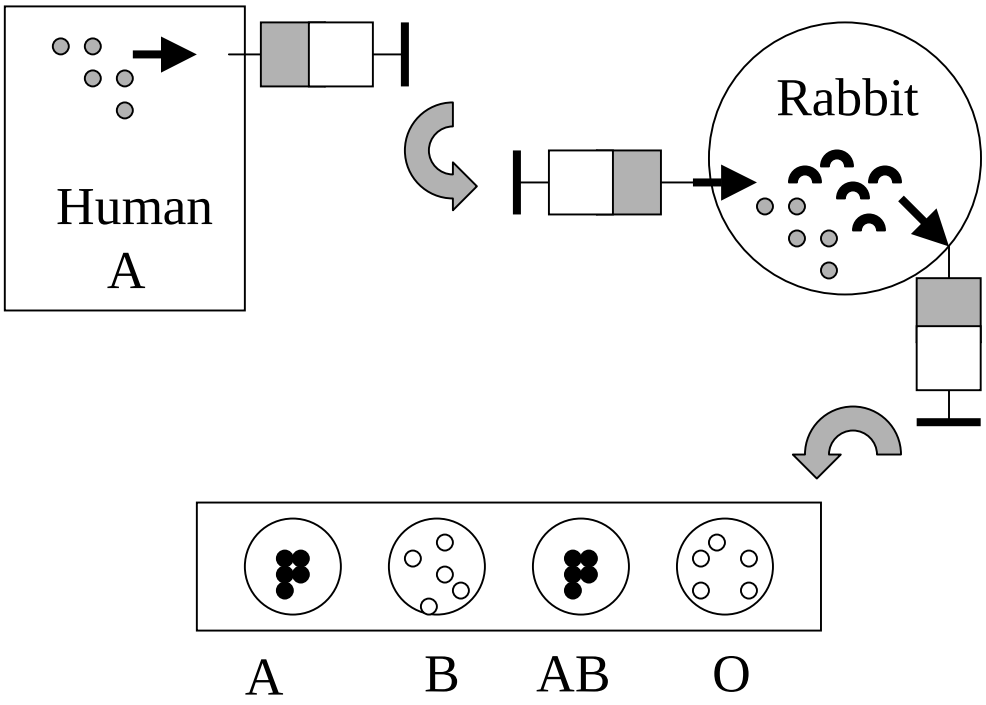
Disease	Concordance DZ	Concordance MZ	Genetic Determinism
Cancer	6.8	2.6	0.23-0.33
Arterial hypertension	25.0	6.6	0.53-0.62
Manic-depressive psychosis	67.0	5.0	1.04-1.05
Tuberculosis	37.2	15.3	0.53-0.65

From Cavalli-Sforza & Bodmer (1971)

Fisher, Haldane, and Wright

- RA Fisher
  - *The Genetical Theory of Natural Selection* (1930)
  - Fisher's fundamental theory
  - Geometric model of adaptation
  - The concept of likelihood in statistical analysis
  - Experimental design
- JBS Haldane
  - *The Causes of Evolution* (1932)
  - Fixation probabilities of advantageous alleles
  - Theory of sex-linked loci
  - Eloquent exponent of the theory of evolution by natural selection
- Sewall Wright
  - Evolution in Mendelian populations (1931)
  - Developed the use of diffusion theory in population genetics
  - Importance of genetic drift
  - Selection at multiple-loci
  - Shifting-balance theory of evolution
  - Four volume *Evolution and the genetics of populations* (1968-1978)

Serological techniques for detecting variation

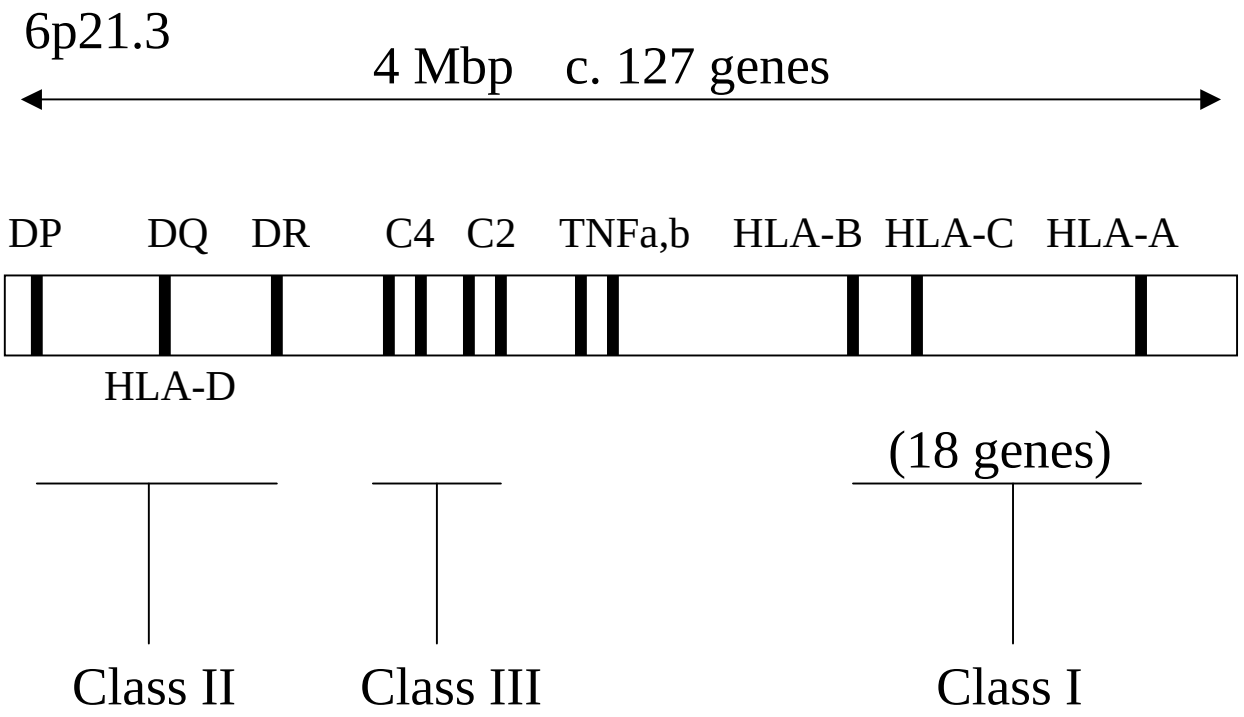


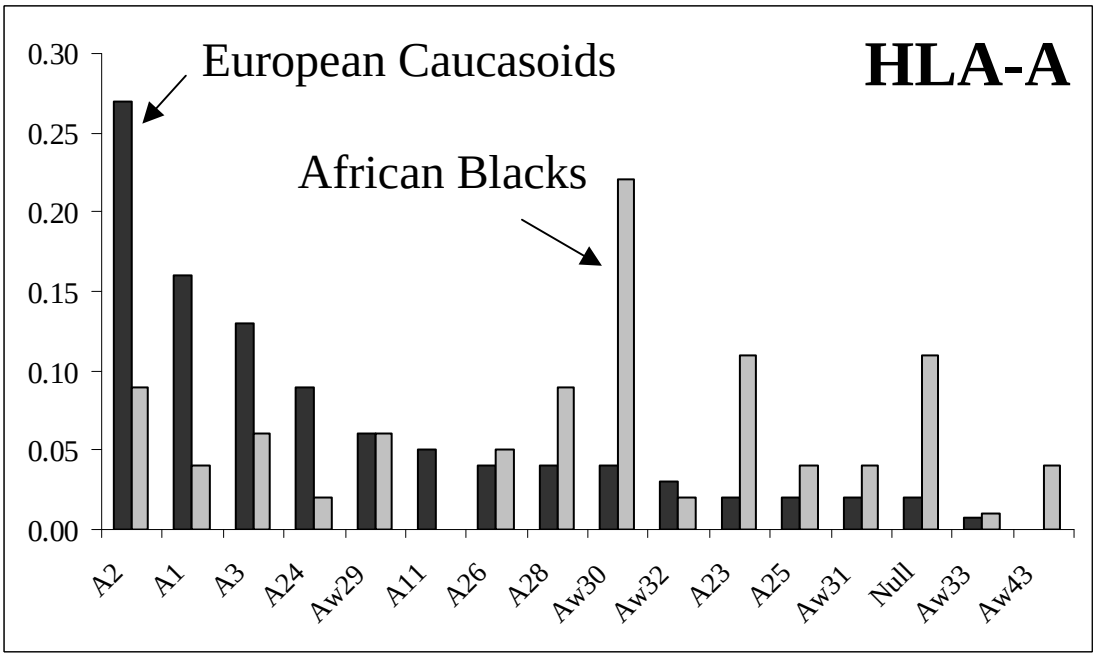
Polymorphic blood groups in the white English population (no. types)

ABO	(4)	Kidd	(3)
Rh	(7)	Dombrock	(2)
MNS	(6)	Auberger	(2)
P	(3)	Xg	(2)
Secretor	(2)	Sd	(2)
Duffy	(3)	Lewis	(2)

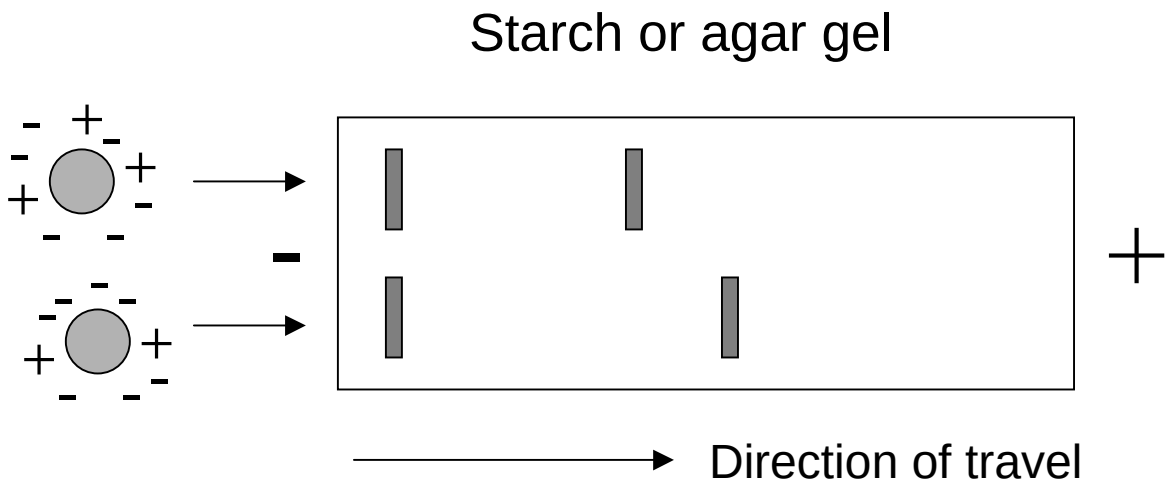
Pr{2 people same blood type} ≈ 3 in 10,000

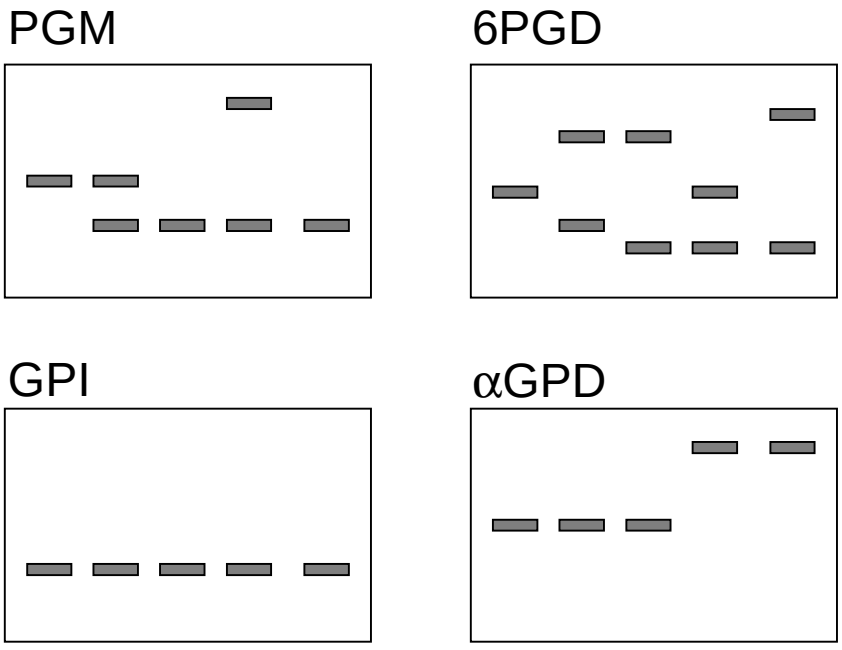
HLA diversity at the MHC locus





Protein electrophoresis

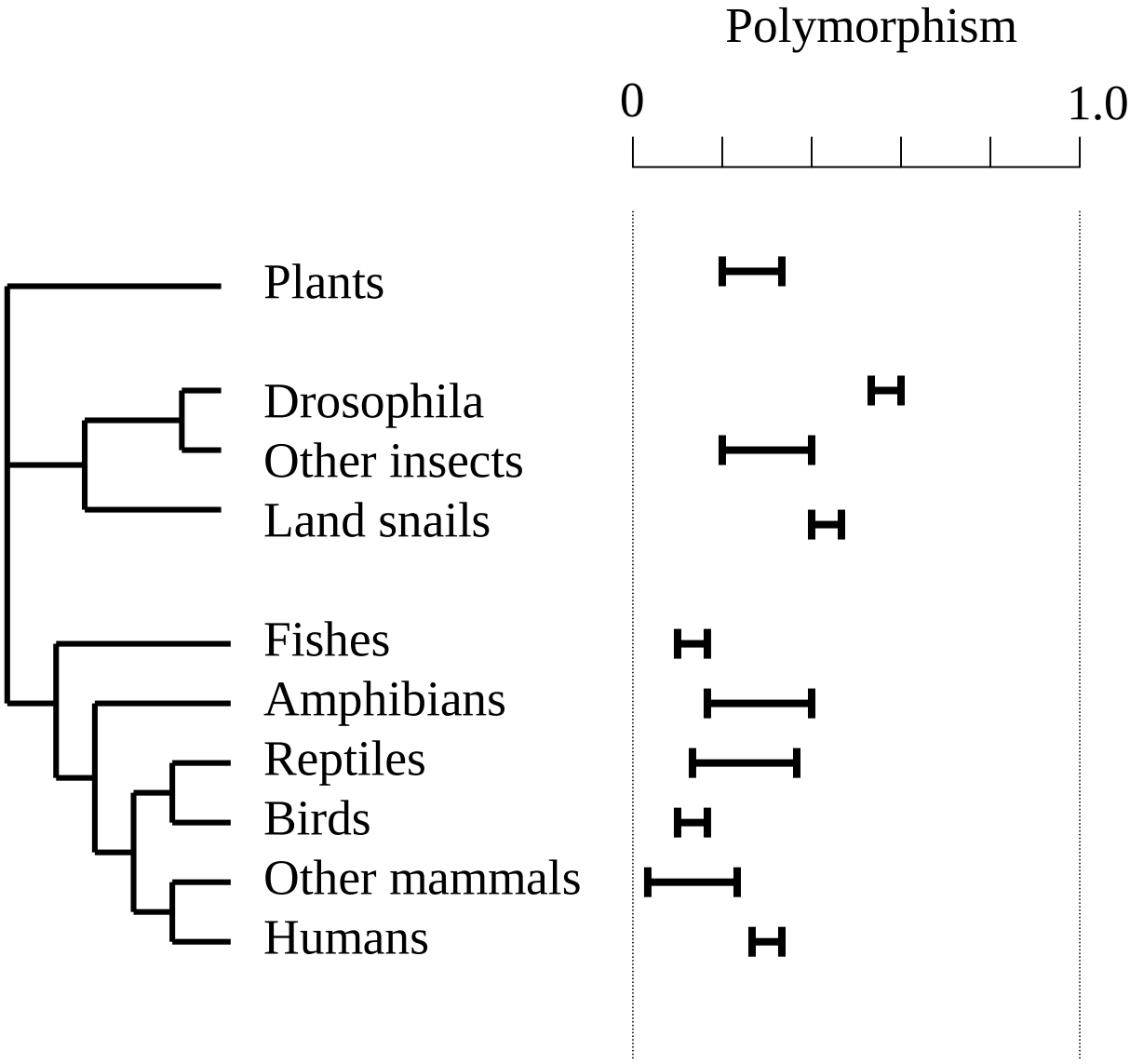




Polymorphism = 0.75

Heterozygosity = 0.30

The phylogenetic distribution of allozyme variation



Humans Polymorphism = 0.31

Heterozygosity = 0.06

Two haploid genomes are expected to differ at c. 6,000 loci

The rise of the neutral theory

- Observations
  - Constancy of rate of molecular evolution (the molecular clock)
  - More important regions of proteins evolve at a slower rate than less important domains
  - High levels of protein polymorphism
  - High rates of molecular evolution (about 1.5×10^{-9} changes per amino acid per year)
- Theoretical considerations
  - Haldane's cost of natural selection
  - Segregation load of balanced polymorphisms

Some population genetic terminology

Population = set of inter-mating/competing individuals

N = Number of individuals in a population

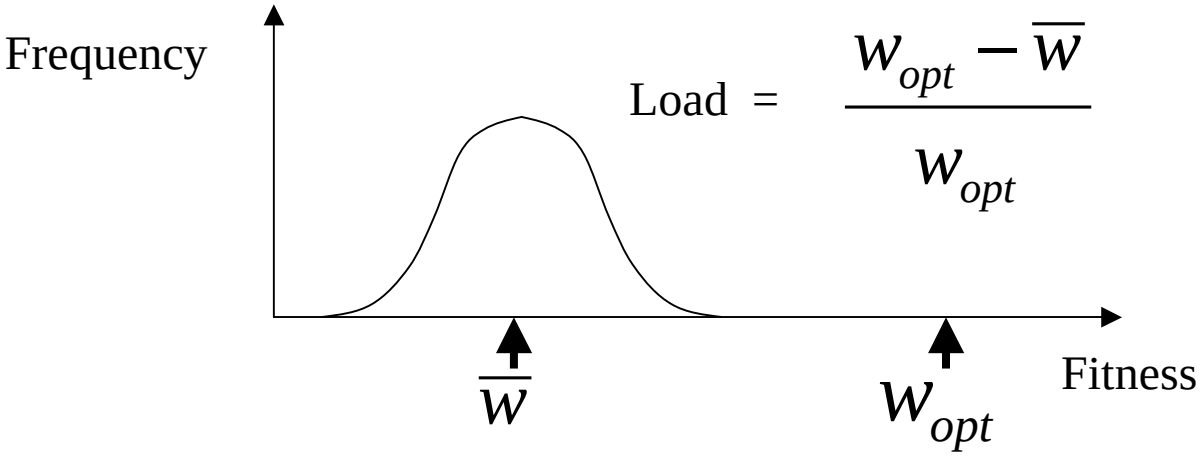
x = allele frequency = $N_{(x)}/N$ as $N \rightarrow \infty$

s = selective advantage

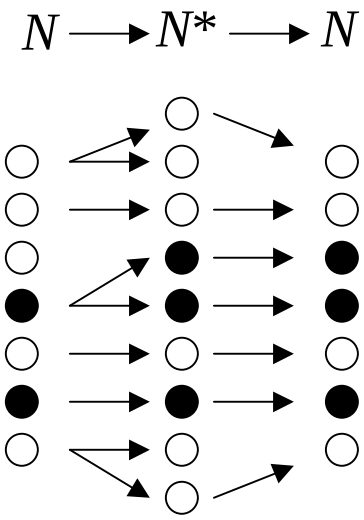
Genetic load

Fitness (w)

= Expected number of offspring given genotype



Haldane's cost of natural selection



$w(\bigcirc) = 1$

$w(\bullet) = 1+s$

Nsx selective deaths occur every generation

To fix ● there must be a total of $4.6N$ selective deaths if it has a 1% advantage

Segregation load due to balanced polymorphisms

Genotype	AA	Aa	aa
Fitness	$1 - s$	1	$1 - s$
Frequency	x^2	$2x(1-x)$	$(1-x)^2$

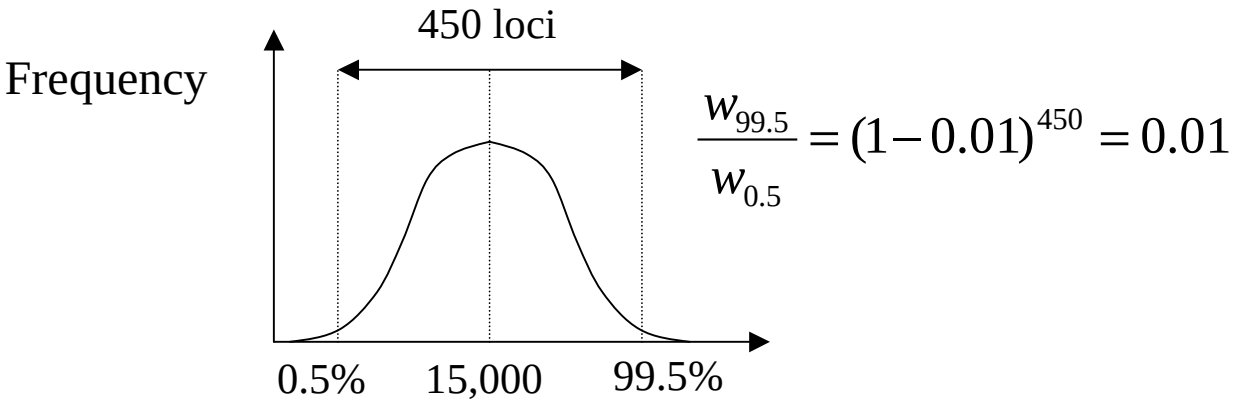
$$\frac{w_{opt} - \overline{w}}{w_{opt}} = 2sx(1-x)$$

$$\text{if } x = 0.5, \quad L = \frac{s}{2}$$

To maintain 30,000 polymorphisms, each of which has a heterozygote advantage of 1% creates a load of

$$L = 1 - 0.995^{30,000} = 1 - 5 \times 10^{-66}$$

Variation



$$\frac{w_{99.5}}{w_{0.5}} = (1-0.01)^{450} = 0.01$$

Features of the neutral theory

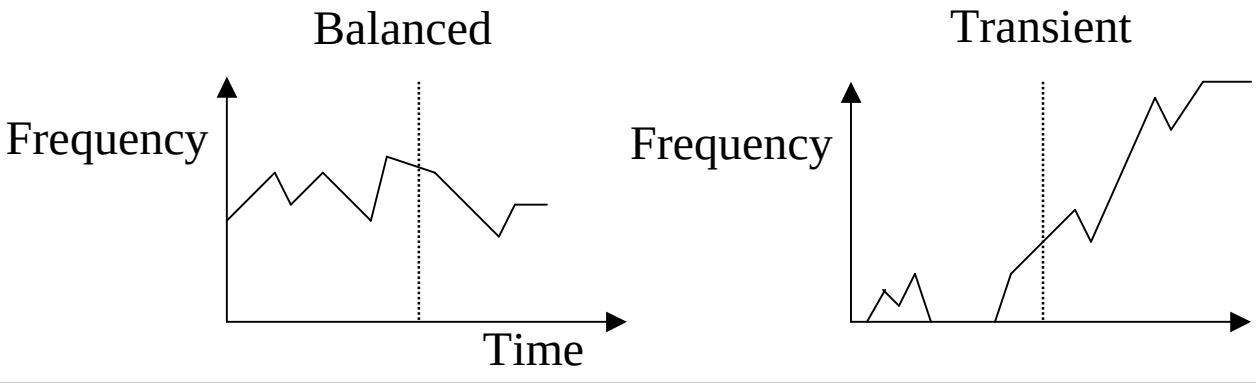
- The majority of changes in proteins and at the level DNA which are fixed between species, or segregate within species, are of no selective importance
- The rate of substitution is equal to the rate of neutral mutation

$$k = f_{neutral}\ \mu$$

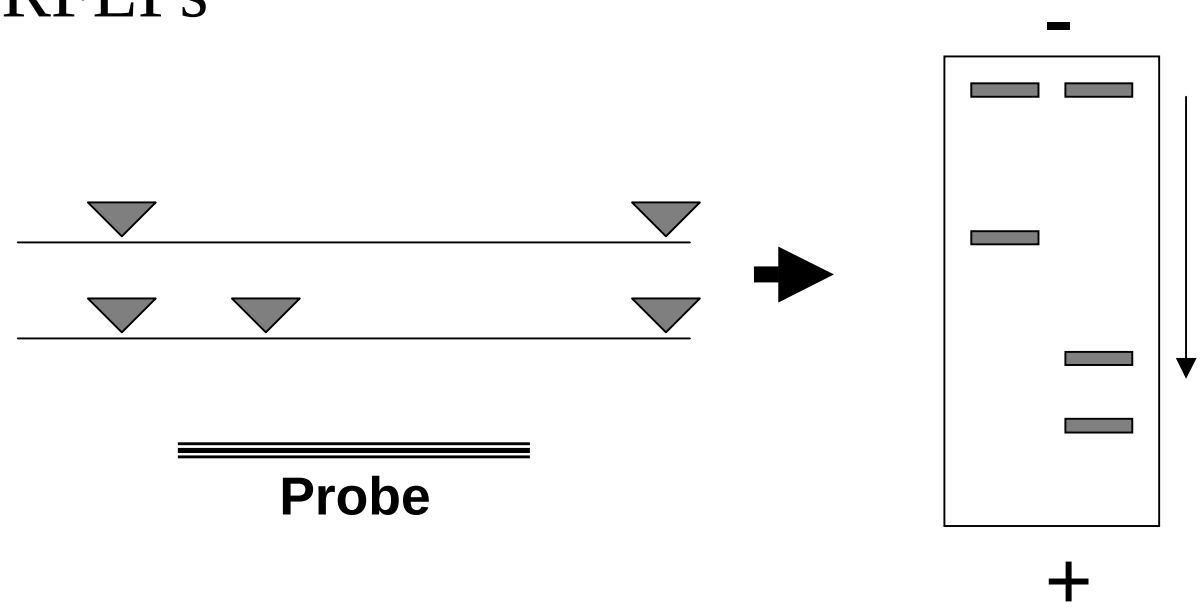
- The level of polymorphism in a population is a function of the effective population size and the neutral mutation rate

$$\pi = \frac{4N_e\mu}{1+4N_e\mu}$$

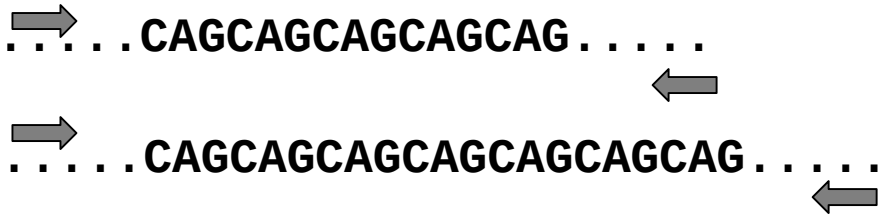
- Polymorphisms are transient rather than balanced



RFLPs



PCR analysis of microsatellites



Full sequence analysis

```
ATGTGAATGCTAATG
...A..T........
.C.A.......G...
.C.......--.G...
...A..T.--.....
```

SNPs

```
ATGTGAATGCTAATG
...A..T........
.C.A.......G...
.C.......--.G...
...A..T.--.....
```

Segregating site

Indel

Statistics of polymorphism

No. segregating sites (S) = 4

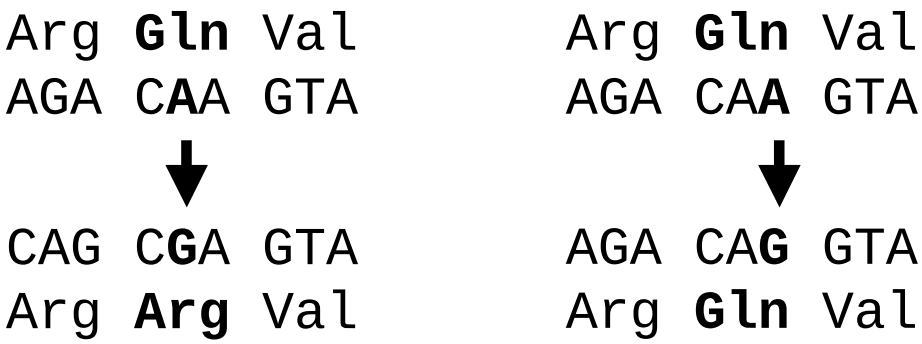
Average pairwise differences (π) = 2.4
= 0.16 per site

Seq	2	3	4	5
1	2	3	2	2
2		3	4	0
3			1	3
4				4

No. haplotypes = 4

Patterns of variation at the DNA level

- Synonymous & nonsynonymous mutations



D. simulans

π_{total} = 0.010 per site
π_{silent} = 0.038
$\pi_{noncoding}$ = 0.023

- Nucleotide variation v. protein variation?

	Humans	*D. melanogaster*
Allozyme	6%	14%
Nucleotide	0.1%	1%

Current issues in population genetics

- Medical applications
  - Disease gene identification by association mapping
  - Understanding genetic basis of quantitative variation
- Statistical issues
  - Methods for detecting natural selection
  - Full likelihood methods for estimating evolutionary parameters from sequence data
  - The design of population genetic experiments
- Theoretical and empirical issues
  - The maintenance of quantitative genetic variation
  - Interactions between alleles at selected loci
  - The molecular clock
  - Reproductive isolation and speciation