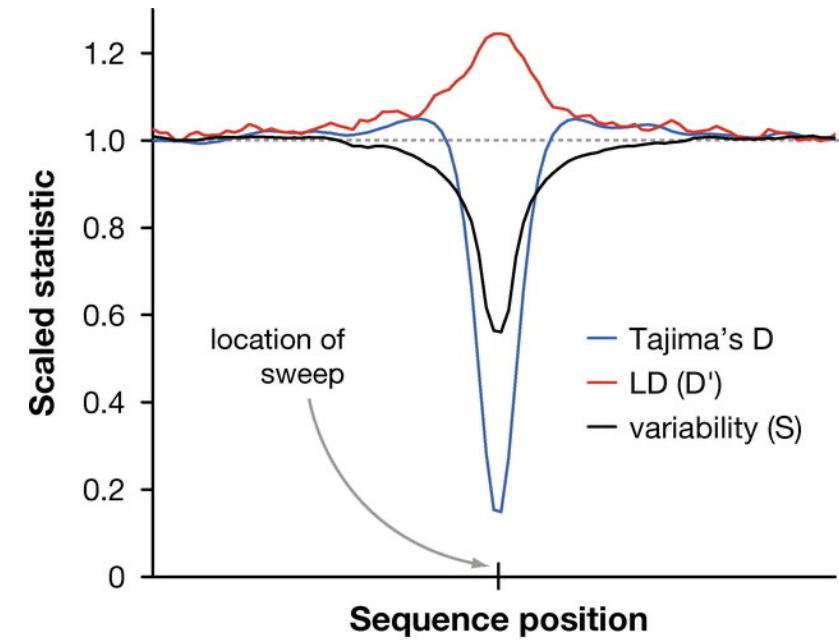
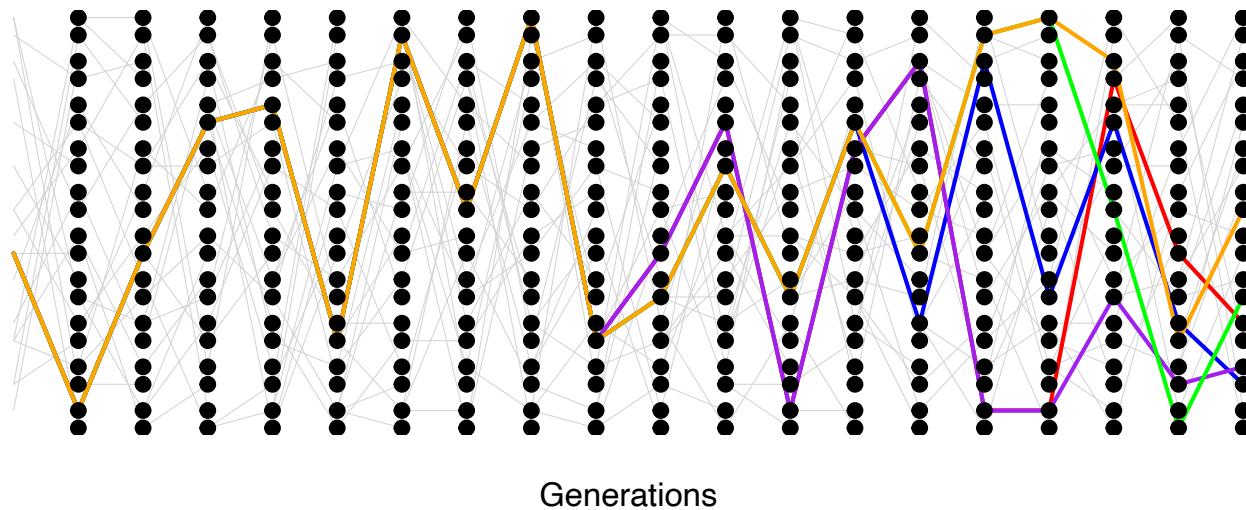


Basics in population genetics and evolution

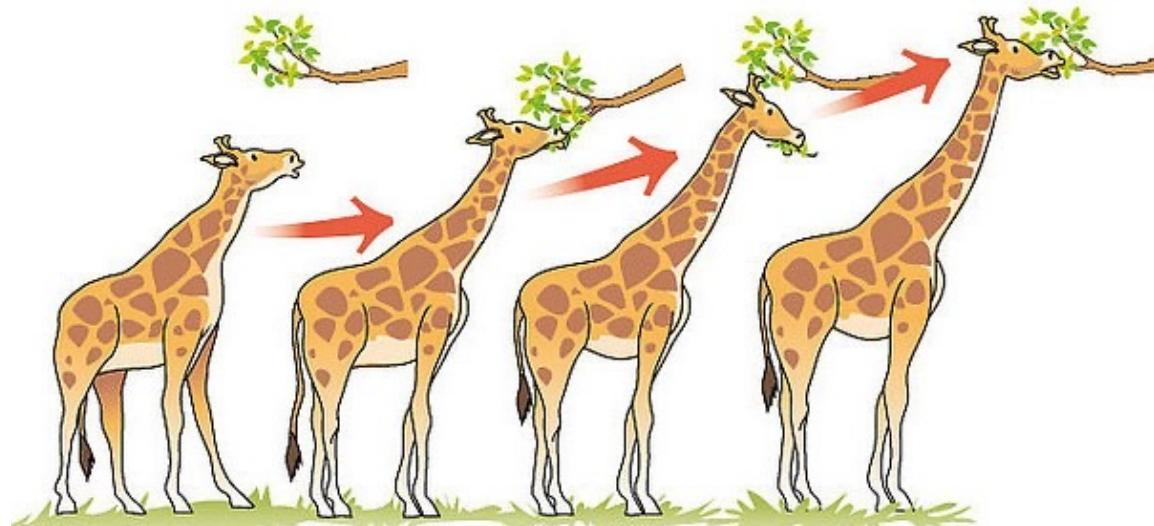
Bioc 201
Reid Brennan



Nielsen R. 2005.
Annu. Rev. Genet. 39:197–218

Lamarck

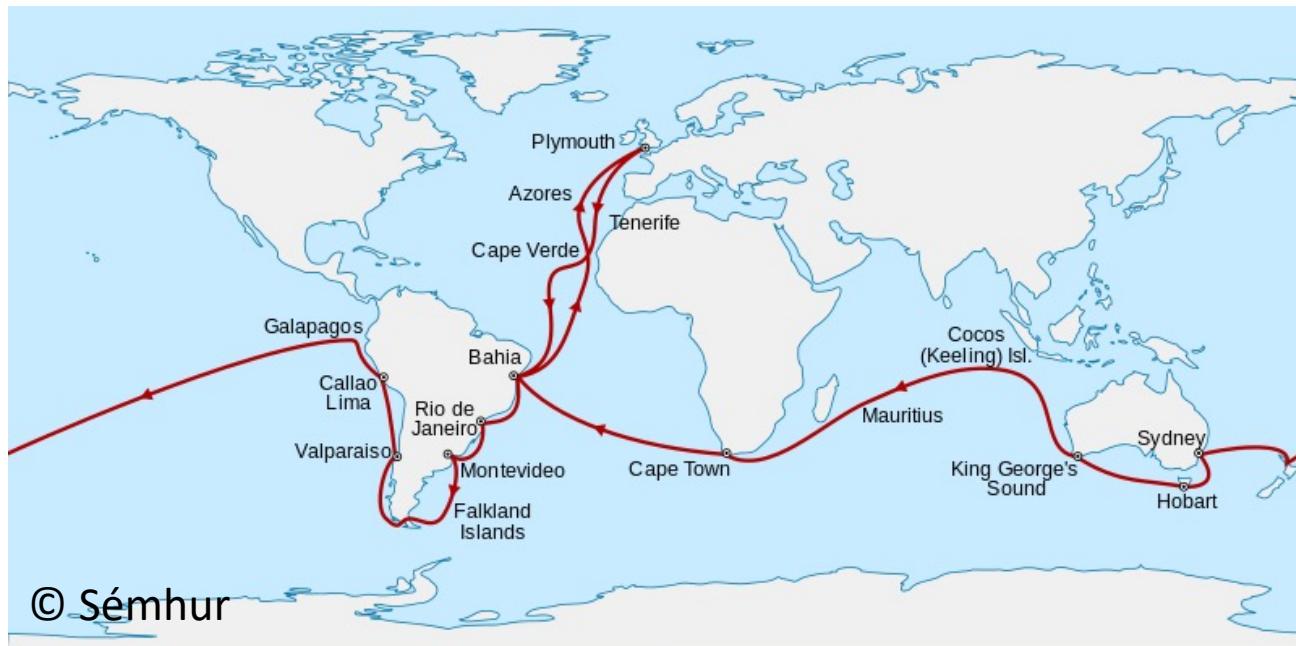
- Knew the earth was old
- 1. Life evolved towards complexity
- 2. Inheritance of acquired characteristics



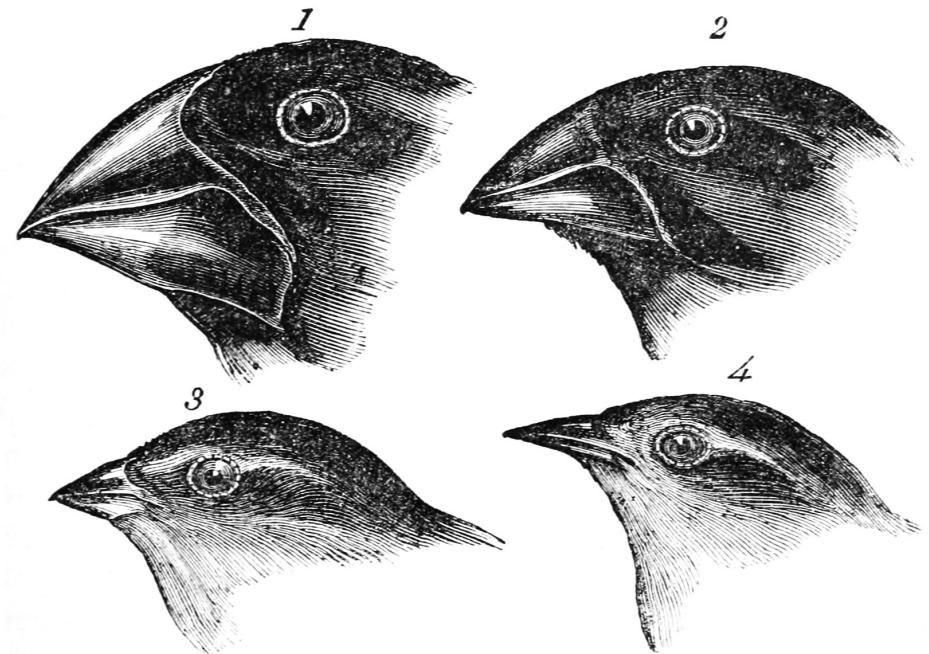
1744-1829

From Sandritaverooka - Eigenes Werk, CC BY-SA 4.0,
<https://commons.wikimedia.org/w/index.php?curid=40890964>

Darwin: 1809-1882



5 year voyage

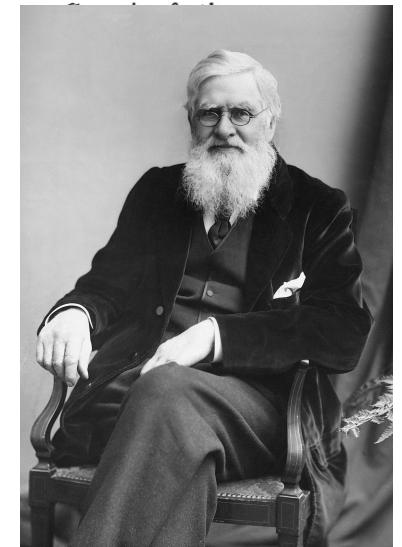
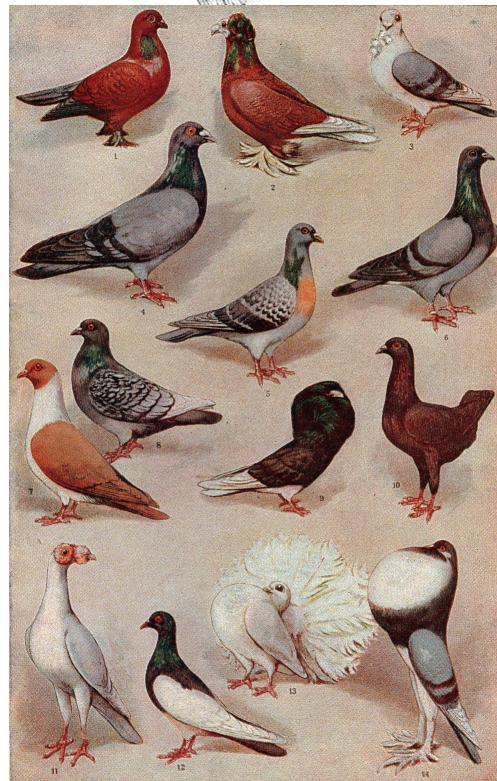
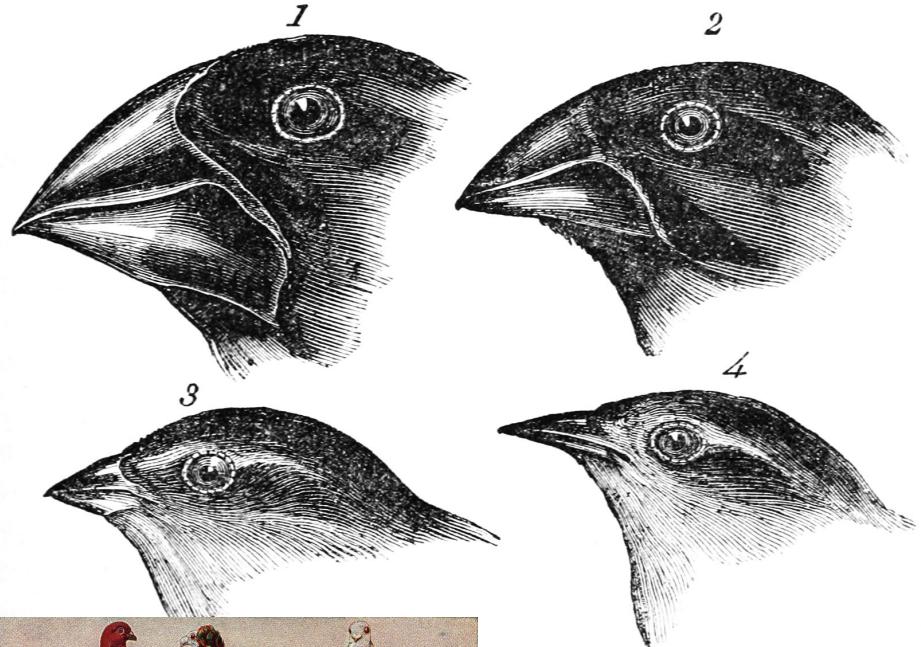


1. *Geospiza magnirostris.*
3. *Geospiza parvula.*

2. *Geospiza fortis.*
4. *Certhidea olivacea.*

Darwin's main insights

- Descent with modification
- Natural selection
 - Different reproductive success between individuals with different traits
- Not all evolution is natural selection
- What Darwin didn't know...
 - Genetics



Alfred Russel Wallace

What is evolution?

- the change over time in the genetic composition of a population. 1951,
Theodosius Dobzhansky

Populations → individuals → genomes

birth – death – migration

Every individual has a unique genome and
every individual contributes to evolution



Population genetics is the study of the genetic composition of natural populations and its evolutionary causes and consequences.

Quantitative genetics is the study of the genetic basis of phenotypic variation and how phenotypic changes evolve over time.

How do the genes and phenotypes of a population change over time?

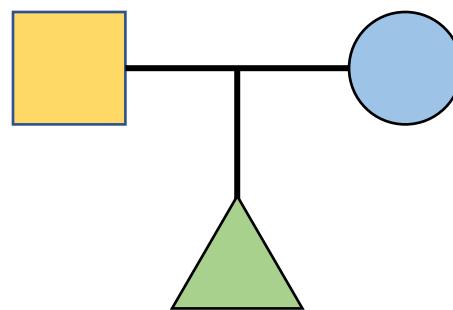
GREGOR MENDEL (1822-1884)



Experiments on Common pea (*Pisum sativum*): 1856 to 1863

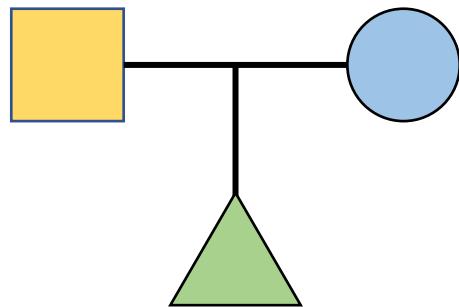


Common thought at the time was blending inheritance

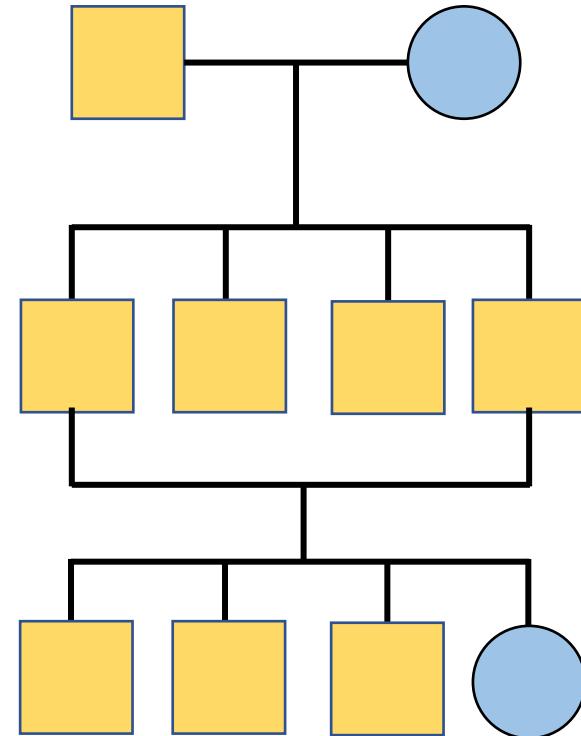


Pea coat colour (actually yellow/green...)

Crosses between 'pure' breeding lines



Parents



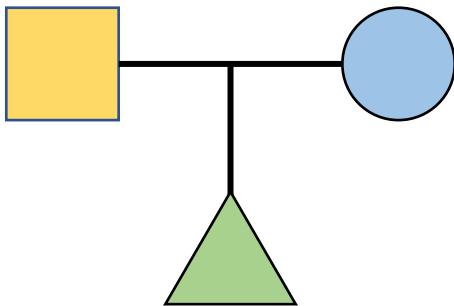
F1

F2

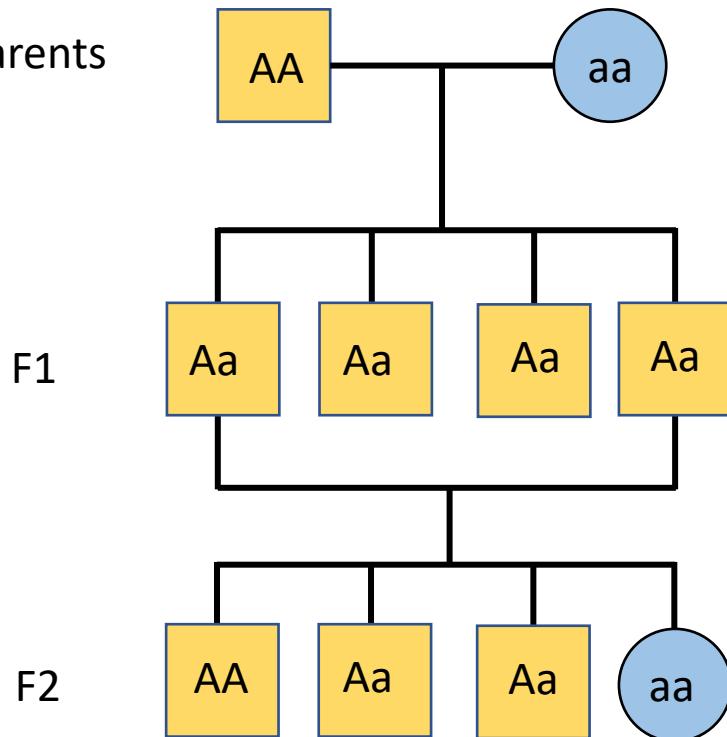
~30,000 plants!

Pea coat colour (actually yellow/green...)

Crosses between 'pure' breeding lines



Parents



F1

F2

independent segregation of alleles (at a single locus)

- *Two members of a gene pair (alleles) segregate separately into gametes so that half of the gametes carry one allele and the other half carry the other allele.*

	A	a
A	AA	Aa
a	Aa	aa

Heterozygote: carries different alleles (A and a)

Homozygote: carries the same two alleles (a and a)

Punnet squares (1905)

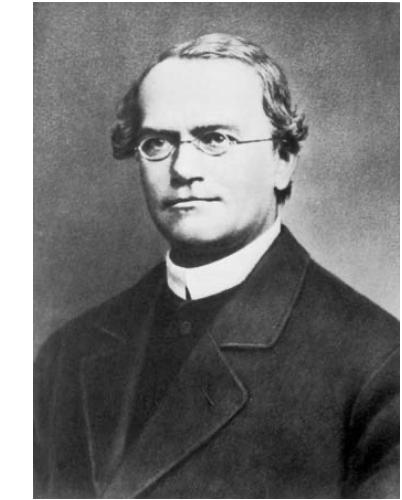
3:1 phenotypes

~30,000 plants!

GREGOR MENDEL



- phenotypes are determined by discrete units (genes) that are inherited intact and unchanged through generations: *particulate inheritance*
- Published in 1866 in the *Proceedings of the Natural Sciences of Brünn* (in German)
- Darwin's contemporary, but unnoticed for 35 years!



Establishes the first and most fundamental prediction of population genetics: ***expected genotypic frequencies***

	A	a
A	AA	Aa
a	Aa	aa

Allele frequencies, genotype frequencies, and HW equilibrium

Definitions....

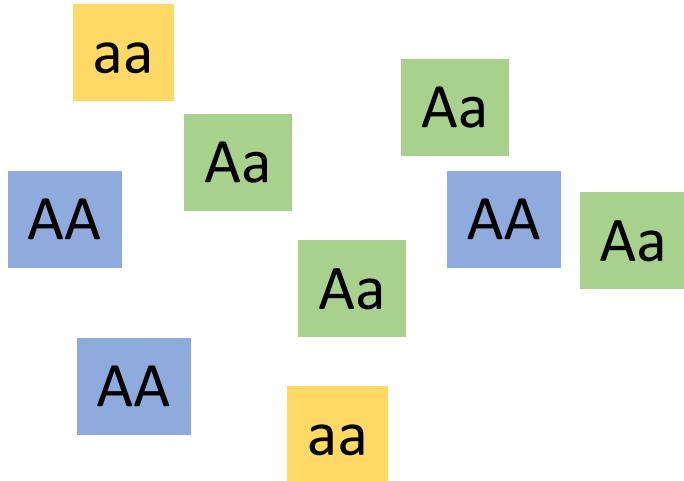
Locus: any location in the genome: gene, single basepair, microsatellite....

Allele: the variants at a locus (polymorphic when more than one)

Genotype: set of alleles possessed by an individual (at one locus or several loci)

Phenotype: an organism's observable attribute (morphological, developmental, biochemical, physiological, behavioral, ...)

Genotype frequency



$$f_{AA} = \frac{N_{AA}}{N}$$

$$f_{Aa} = \frac{N_{Aa}}{N}$$

$$f_{aa} = \frac{N_{aa}}{N}$$

$$N_{AA} = 3$$

$$N_{Aa} = 4$$

$$N_{aa} = 2$$

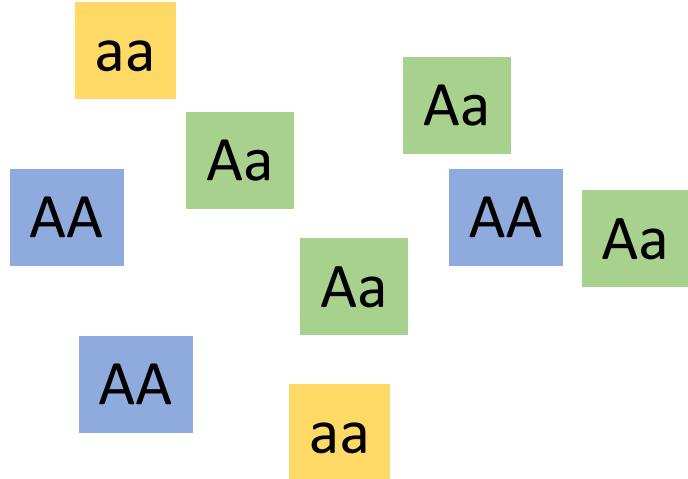
$$N = 9$$

$$f_{AA} = 3/9 = 0.33$$

$$f_{Aa} = 4/9 = 0.44$$

$$f_{aa} = 2/9 = 0.22$$

Allele frequency



$$f_A = p = \frac{2N_{AA} + N_{Aa}}{2N} = f_{AA} + f_{Aa}/2$$

$$f_a = q = 1 - p$$

$$N_{AA} = 3$$

$$N_{Aa} = 4$$

$$N_{aa} = 2$$

$$N = 9$$

$$f_A = 0.33 + 0.44/2 = 0.55$$

$$f_a = 0.22 + 0.44/2 = 0.44$$

$$f_{AA} = 3/9 = 0.33$$

$$f_{Aa} = 4/9 = 0.44$$

$$f_{aa} = 2/9 = 0.22$$

Hardy-Weinberg Equilibrium

- Expected genotype frequencies from allele frequencies
- Need to assume random mating



HARDY-WEINBERG (1908)

- Remember independent segregation...
- To get AA genotype →
- To get aa genotype →
- To get Aa genotype →

Godfrey H. Hardy (1877-1947)
Wilhelm Weinberg (1862-1937)

$$p^2 + 2pq + q^2 = 1$$

AA Aa/aA aa

A null expectation → Deviations = evolution

A single generation of reproduction will result in a population that meets the expected Hardy-Weinberg frequencies, i.e. is at Hardy-Weinberg (HW) equilibrium

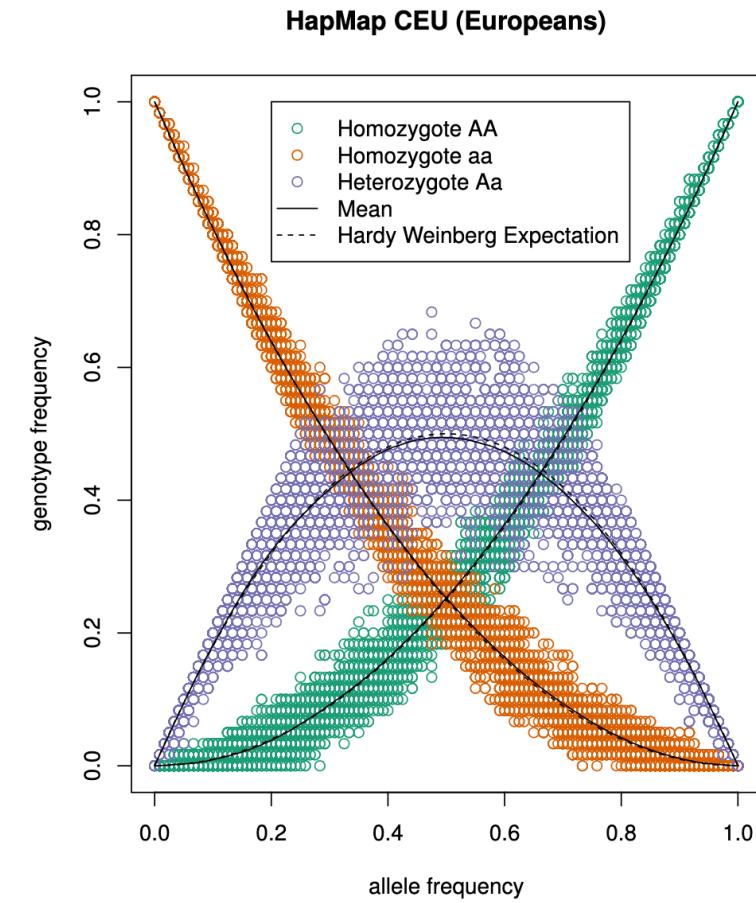
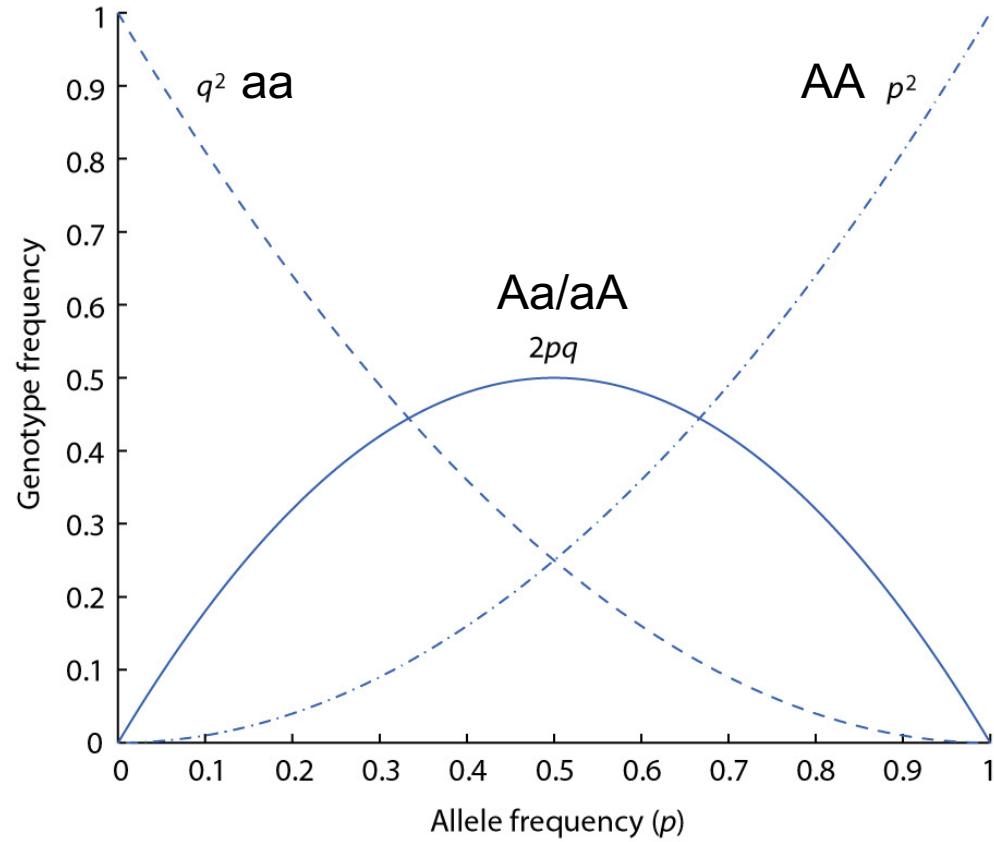
Assuming an “ideal” population, i.e. :

- Diploid organisms
- Sexual reproduction (as opposed to clonal)
- Random mating (as opposed to e.g. assortative) with respect to genotype
- Random union of gametes
- Discrete, non-overlapping generations
- Very large (infinite) population
- No migration
- No population structure
- No natural selection
- Two alleles
- Identical allele frequencies in both sexes

-> Departures from HW equilibrium may indicate:

- Inbreeding
- Assortative mating
- Migration
- Natural selection
- Population structure
- ...

What is the expectation for genotype frequencies?
When do we have the highest heterozygosity?



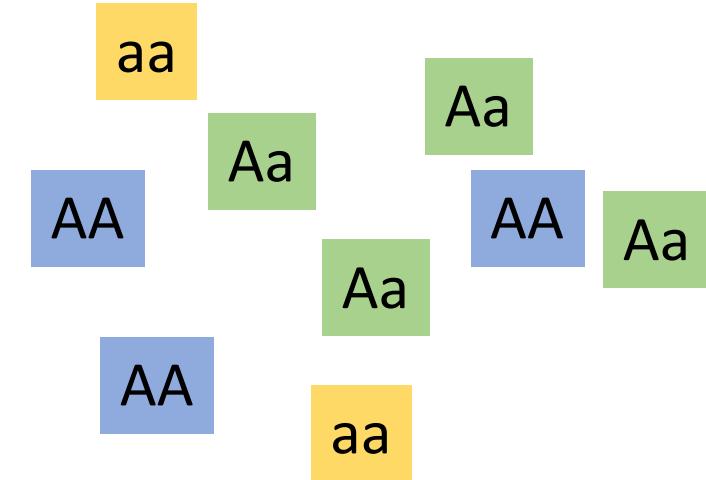
G. coop

Inbreeding coefficient: F_{IS}

Fixation index (F): proportion by which heterozygosity is reduced or increased relative to the heterozygosity of a population at HW equilibrium with the same allele frequencies.

$$F = \frac{H_e - H_o}{H_e} = \frac{2f_A f_a - f_{Aa}}{2f_A f_a}$$

$$f_{Aa} = 2f_A f_a (1 - F)$$



What is F for our population?

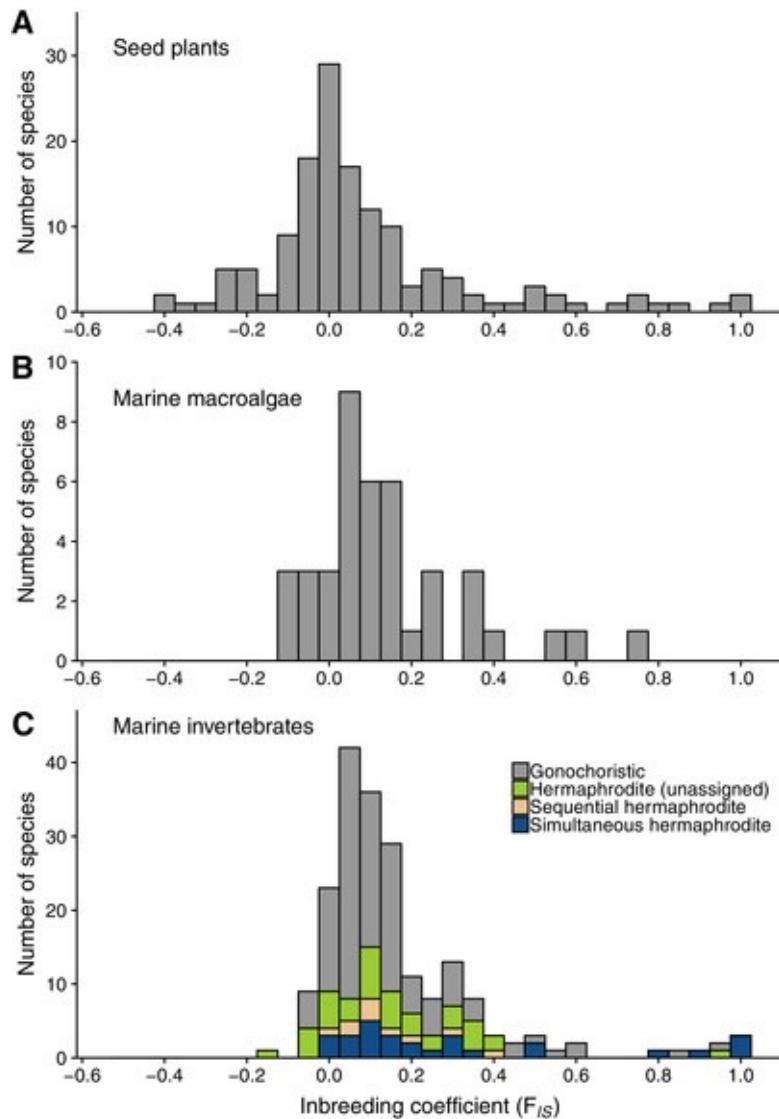
$$\frac{2(0.55)(0.44) - 0.44}{2(0.55)(0.44)}$$

$$f_A = 0.55$$

$$f_a = 0.44$$

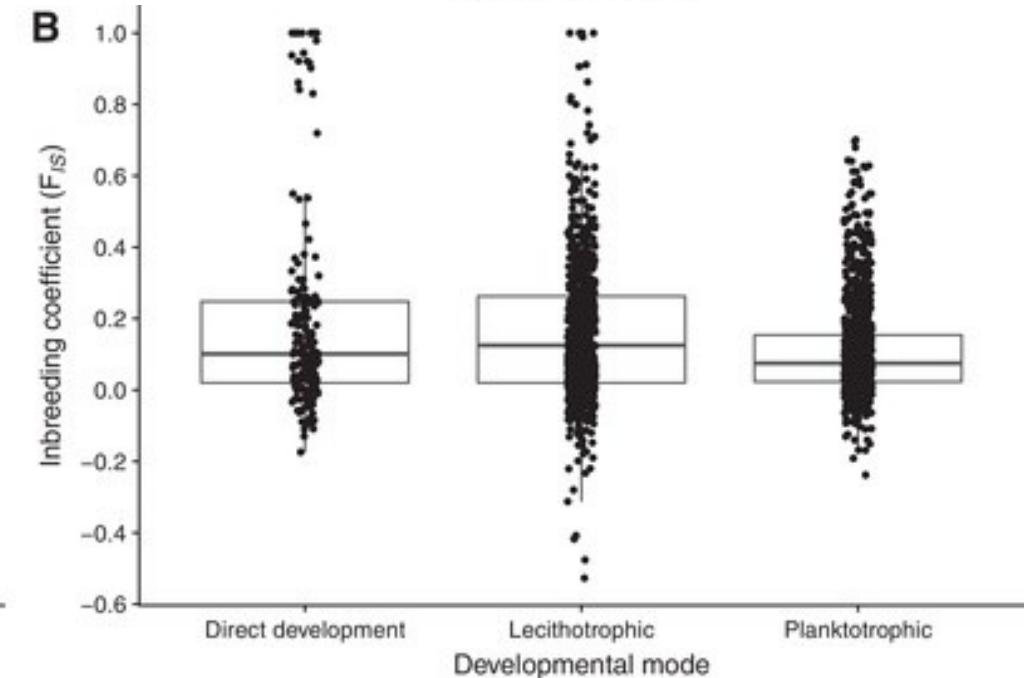
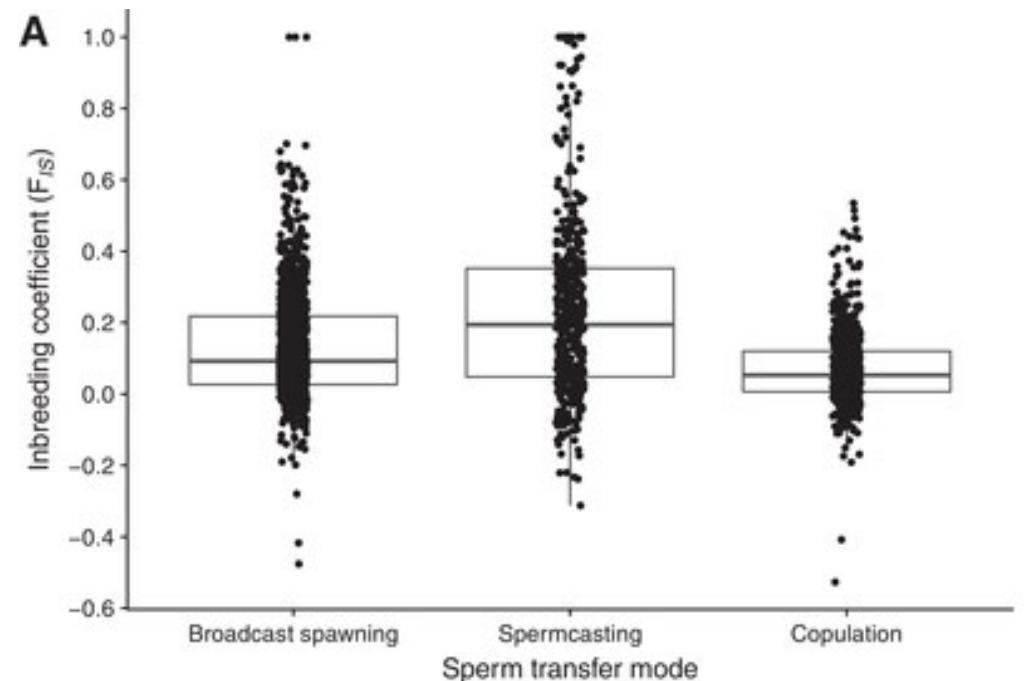
$$f_{Aa} = 0.44$$

Typical F_{IS} values



What influences F_{IS} ?

- Mating strategy
- Dispersal
- Self-fertilization
- Mate choice
- etc



What influences F_{IS} ?

Epiactis lisbethae:

- Internal fertilization
- Brooder
- Crawl away larvae

$$F_{IS} = 0.95$$



Anthopleura elegantissima

- Broadcast spawning
- Planktonic larvae

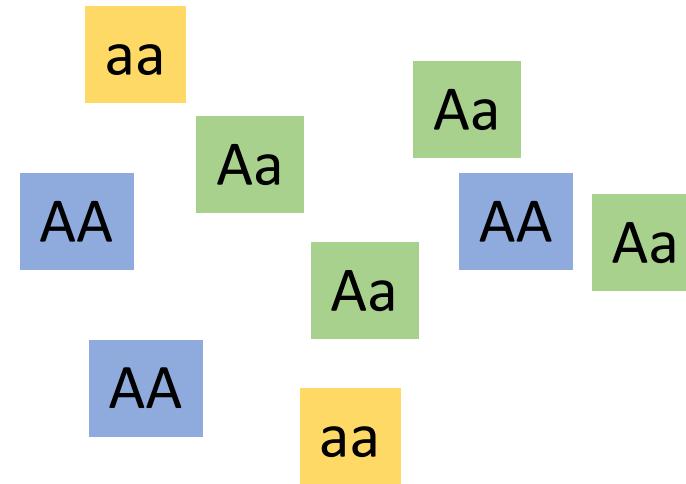
$$F_{IS} = 0.20$$



Drift, mutation, and
selection

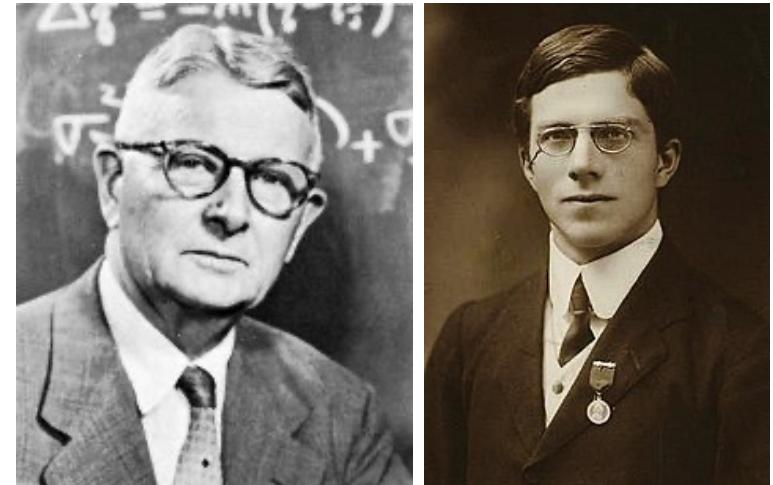
How and why do allele frequencies change through time?

- What contributes to evolution?
- Two main factors:
 - Drift
 - Selection



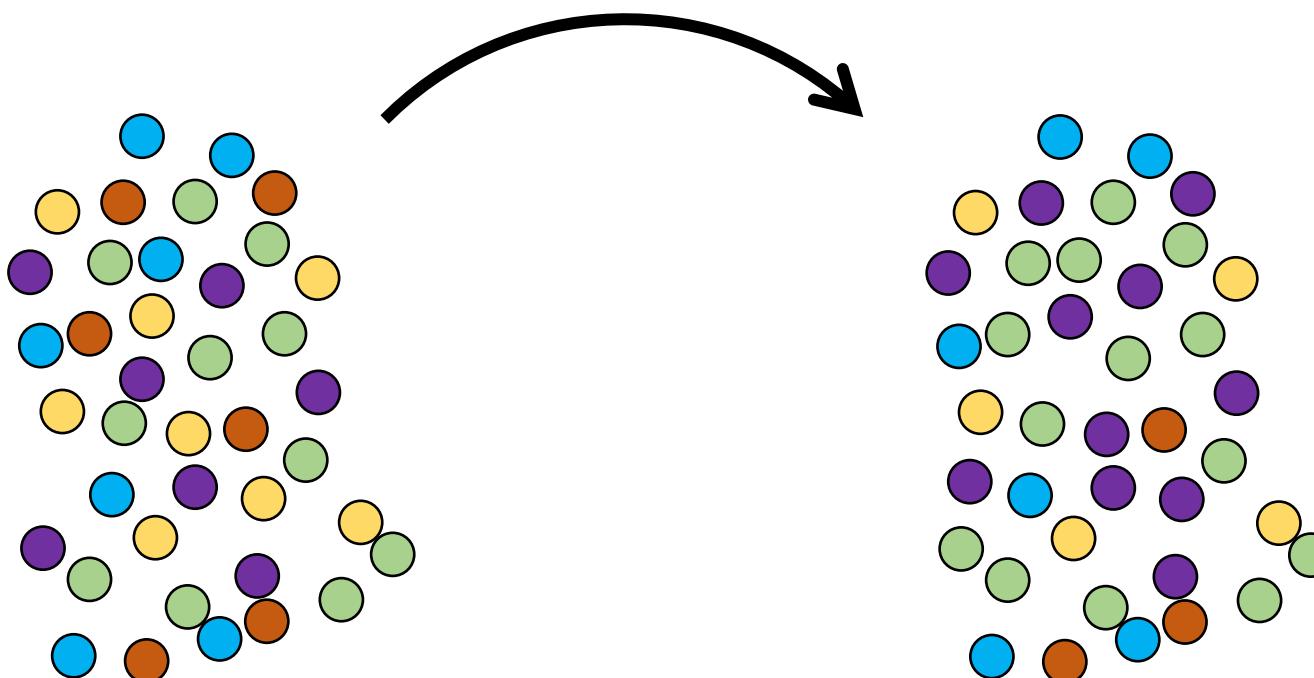
Wright-Fisher Model

- Gene copies are randomly sampled from 1 generation to the next
 - no selection, no mutation, no migration, non-overlapping generation times and random mating
- Think of a bag of marbles



Sewall Wright
1889-1988

Ronald Fisher
1890-1962



Wright-Fisher Model

- Gene copies are randomly sampled from 1 generation to the next
 - What is the expected allele frequency in the next generation?

$$E[f_{A(t+1)}] = \frac{(2N)(f_A)}{2N} = f_A$$

Expected allele frequencies are equal across generations

Why do allele frequencies change?

Wright-Fisher simulation

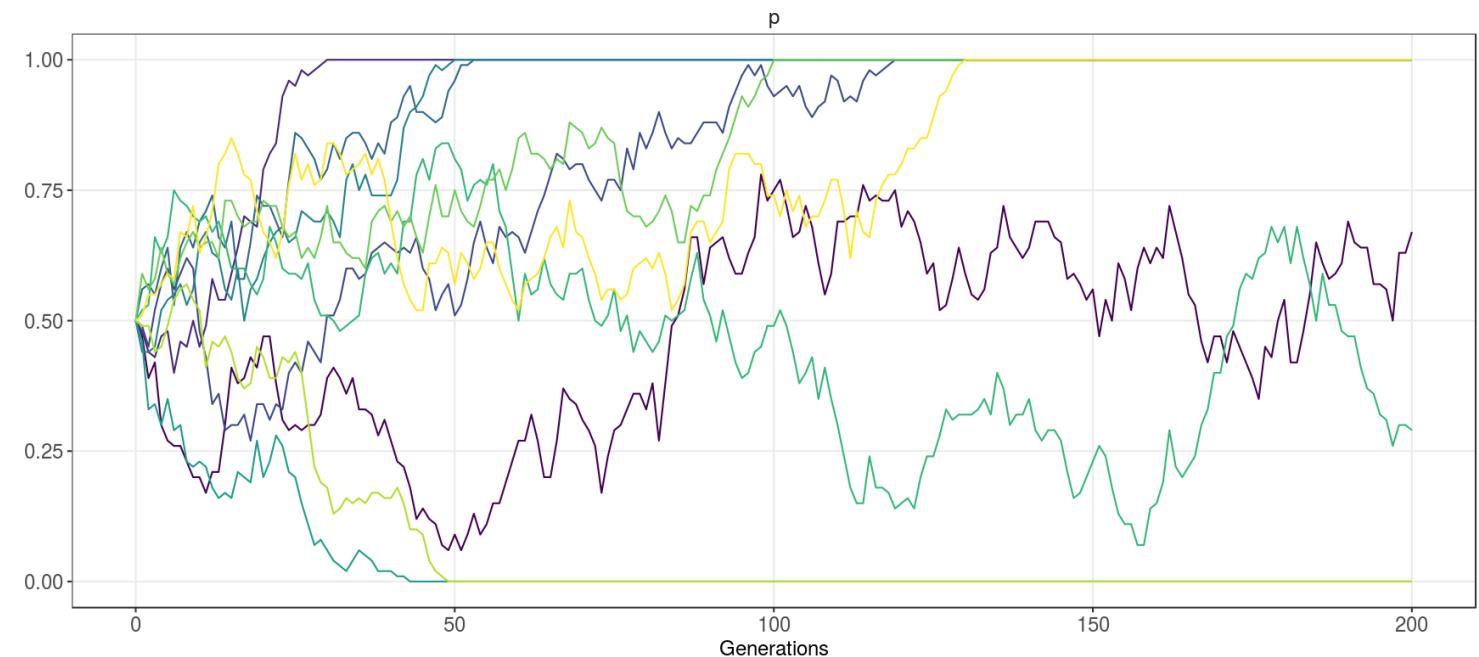
- <https://keholsinger.shinyapps.io/Genetic-Drift/>

Wright-Fisher Model

- Gene copies are randomly sampled from 1 generation to the next

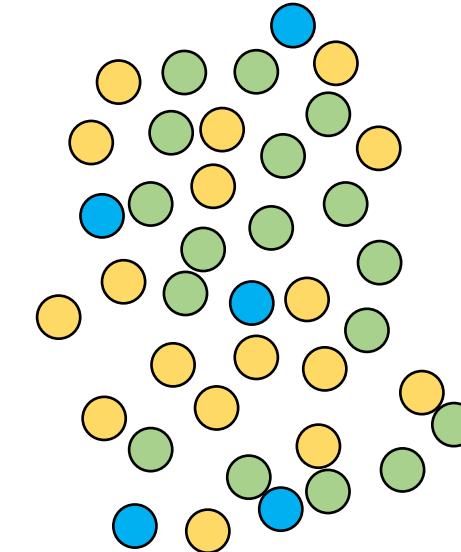
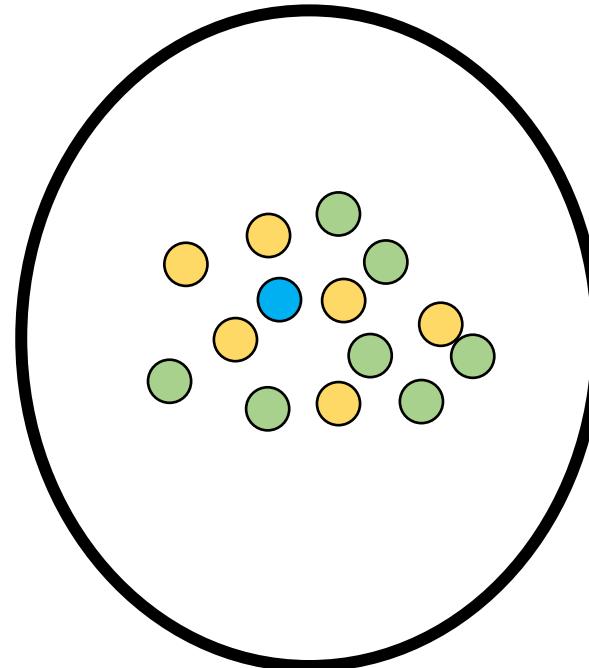
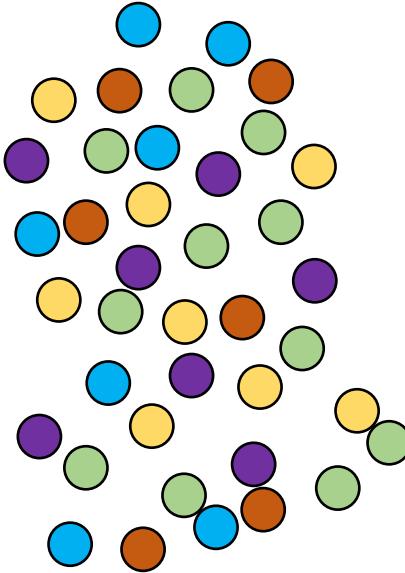
DRIFT

- Think about flipping a coin
- We have finite populations, so alleles are randomly fixed or lost.
- All alleles will eventually be fixed or lost



Effective population size (N_e)

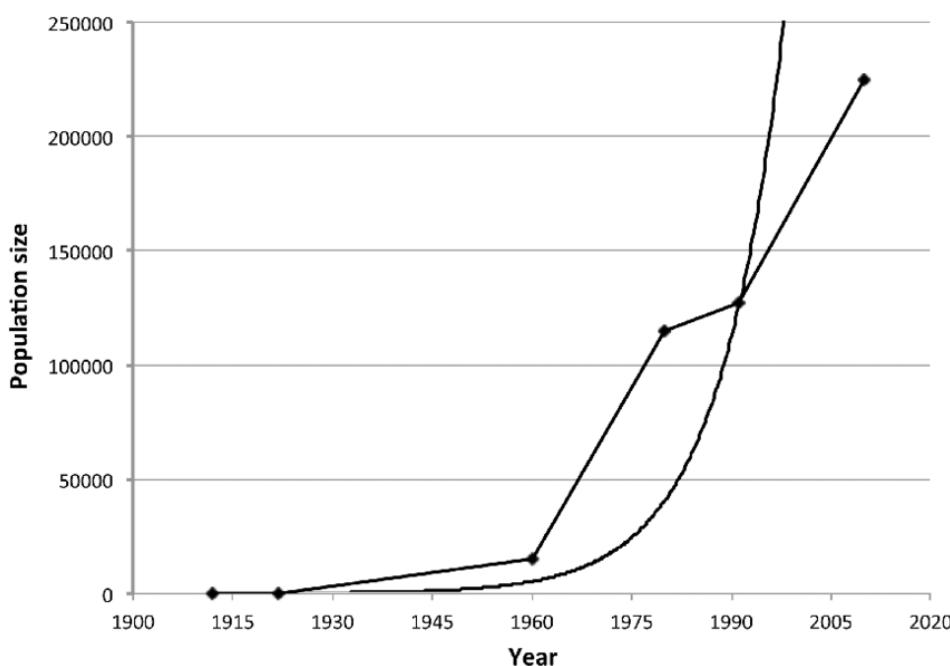
- The number of individuals that would produce the same amount of drift as the observed population
- In contrast to the census size



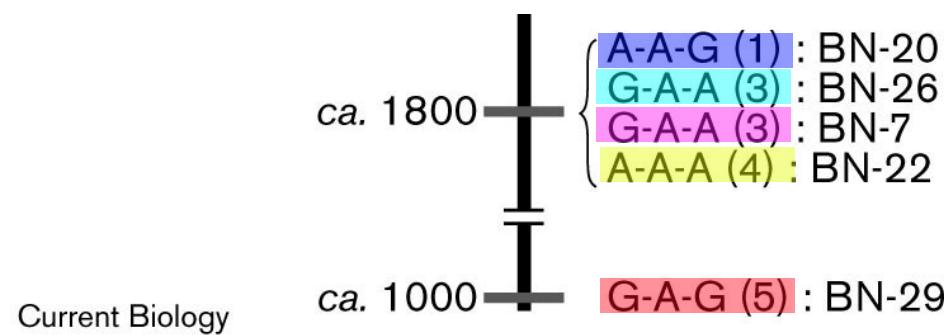
Northern elephant seal



~20 individuals



Abadía-Cardoso et al 2017



An empirical genetic assessment of the severity of the northern elephant seal population bottleneck
Diana S. Weber*, Brent S. Stewart†‡, J. Carlos Garza†§ and Niles Lehman*

Effective population size (N_e)

- The number of individuals that would produce the same amount of drift as the observed population
- In contrast to the census size

Across 3 generations, population size of: 100, 10, 100

What is N_e ?

Harmonic mean

$$\frac{n}{\sum_{i=1}^t \frac{1}{N_i}} = \frac{3}{\left[\frac{1}{100} + \frac{1}{10} + \frac{1}{100} \right]}$$

$$N_e = 25$$

$$\text{Arithmetic mean} = 70$$

Effective population size (N_e)

- What can influence N_e ?
 - Change in population size
 - Variance in reproductive success (common in marine species)
 - Uneven breeding sex ratios (the less frequent sex can be seen as an allelic bottleneck)

$$N_e = 4 \frac{N_m N_f}{N_m + N_f}$$

N_m : number of breeding males

N_f : number of breeding females

All other assumptions of the Wright-Fisher model are met

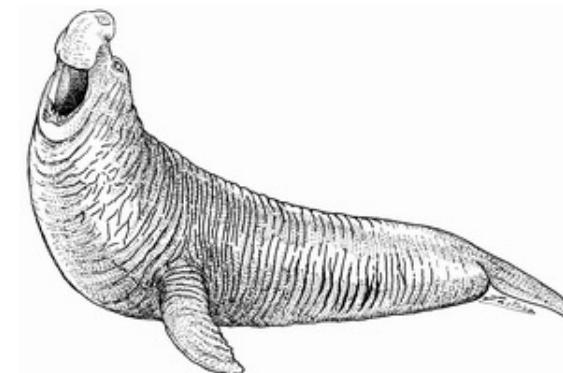
Example: southern elephant seals (*Mirounga leonina*),
Falkland Islands (Fabiani *et al.* 2004)

$$N = 625$$

$$N_m = 21 \text{ (estimated using microsatellite markers)}$$

$$N_f = 550$$

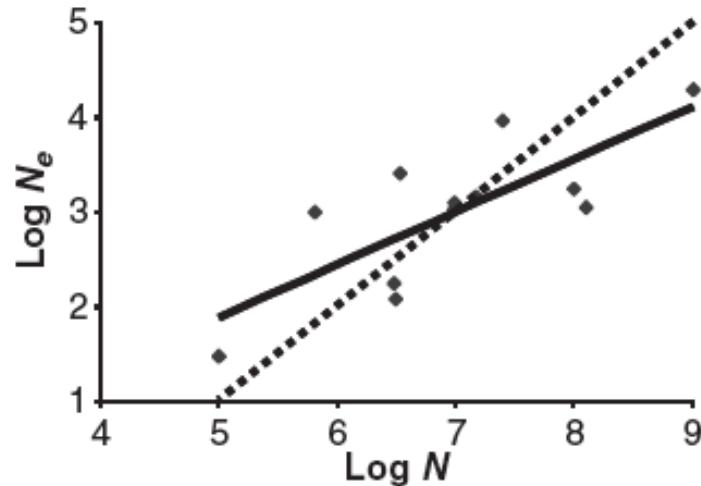
$$N_e = 80.9$$



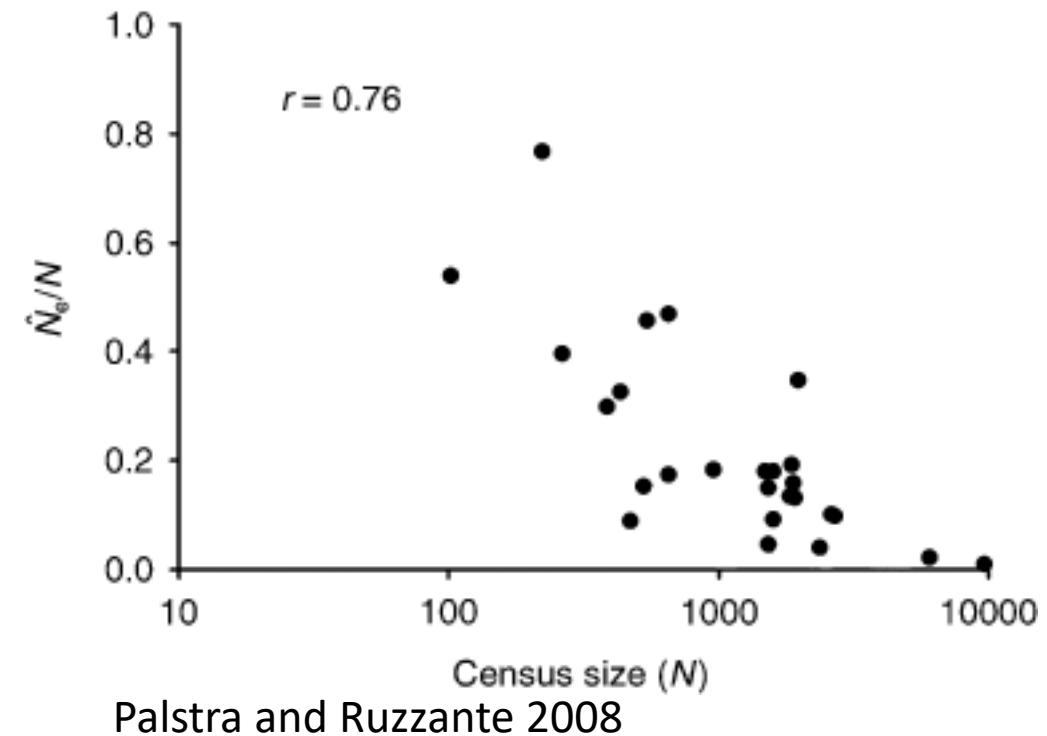
N_e/N estimates in wild populations:

- In fish, 100's to 10,000's typical for N_e
- N_e/N estimates: ~ 0.1 (Frankham 1995)
- N_e/N estimates in *marine* populations
 - 0.01 – 0.000001 (Hauser & Carvalho 2008)

Decreases with increasing population size

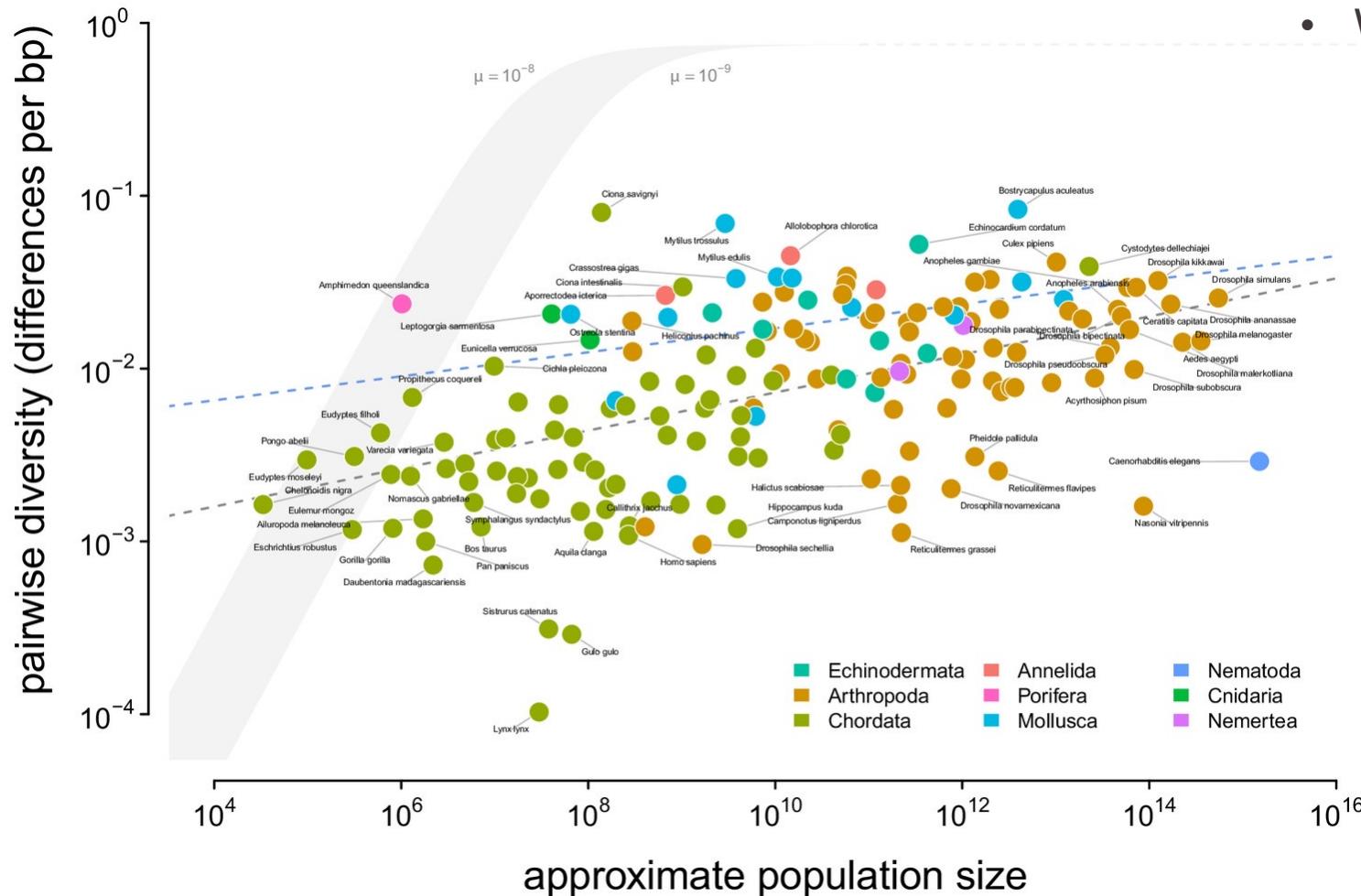


Hauser & Carvalho 2008



Palstra and Ruzzante 2008

N_e /diversity does not scale with population size



- What is N_e for humans?
 - ~3100 – 7500 (Tenesa et al, 2007)
 - Why?

Why are N_e/N estimates so low in marine populations?

- Publication bias?
 - Difficult to obtain finite N_e estimates for large populations
- Fluctuations in population sizes?
 - Yes, but some of the methods used apply to the short-term N_e
- Uneven breeding sex ratios?
 - Yes, but extreme biases in sex ratios would be required to explain the N_e/N ratios,
 - e.g. one of the sexes 97.4% of the spawning population for $N_e/N = 0.1$
- Selection
- “sweepstakes” reproductive systems
 - High fecundity and mortality during the pelagic larval phase where entire families either survive or die (Hedgecock 1994)
 - Lower diversity within and higher differentiation among larval cohorts
 - Larvae or juveniles in recruiting batches tend to be related
 - N_e/N primarily affected by the small proportion of winners,
 - e.g. 0.004 for a pop of 100,000 where 100 produce 1000 offspring, 50,000 two and 49900 zero

Estimating effective population size

Many methods (not reviewed here!)

Three commonly used approaches (Hare *et al.* 2011)

Single time point (contemporary)

- Linkage disequilibrium

- Excess heterozygosity

- Frequency of half and full siblings

Temporal (contemporary)

- Variance N_e

Coalescent (long-term)

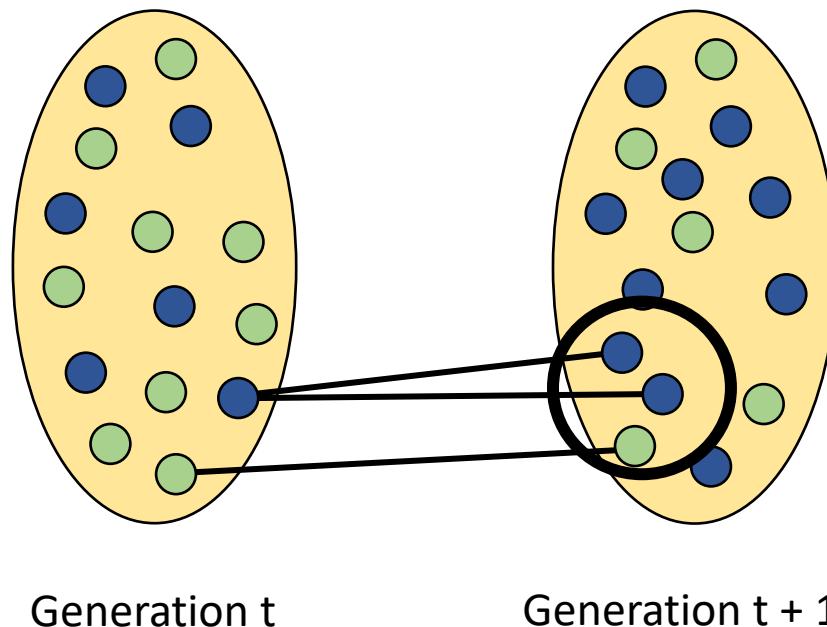
If drift causes all alleles to eventually be fixed or lost, why do we have diversity?

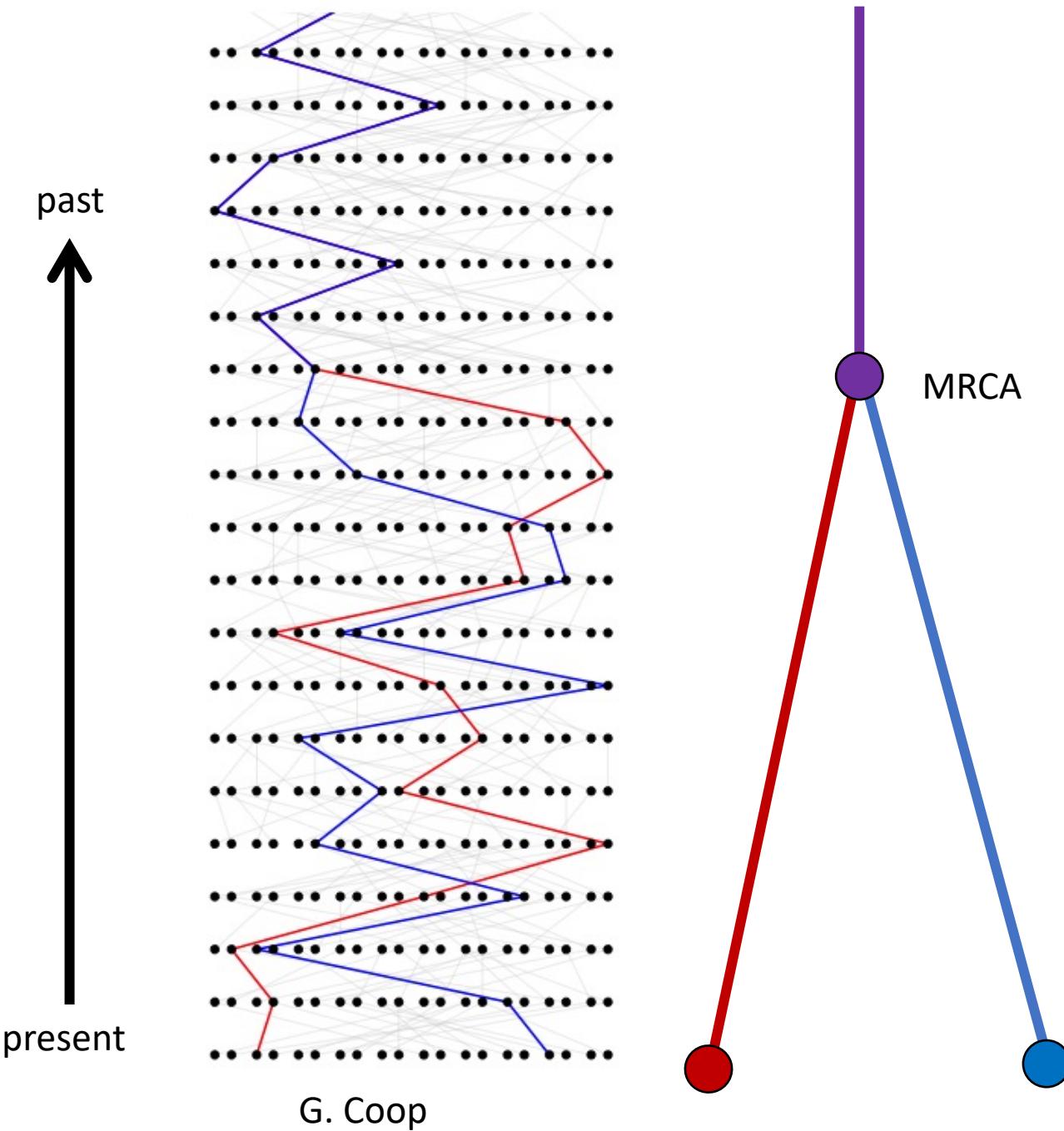
Mutation

- Drift ↓ variation
- Mutation ↑ variation
- Deletions, insertions, inversions, point mutations....
- Mutation rate = μ
 - For example, probability of A → T (or anything)
 - Typically 10^{-7} to 10^{-9} in eukaryotes
- Mutation-drift balance

The coalescent

- Thinking backwards in time
 - 1980's: Primarily, John Kingman, but also others
- When do two gene copies have the same parent?
 - When do they coalesce?
- Useful, because can be directly related to observed data, easy to simulate





What is the probability that two individuals have the same parent in the previous generation?

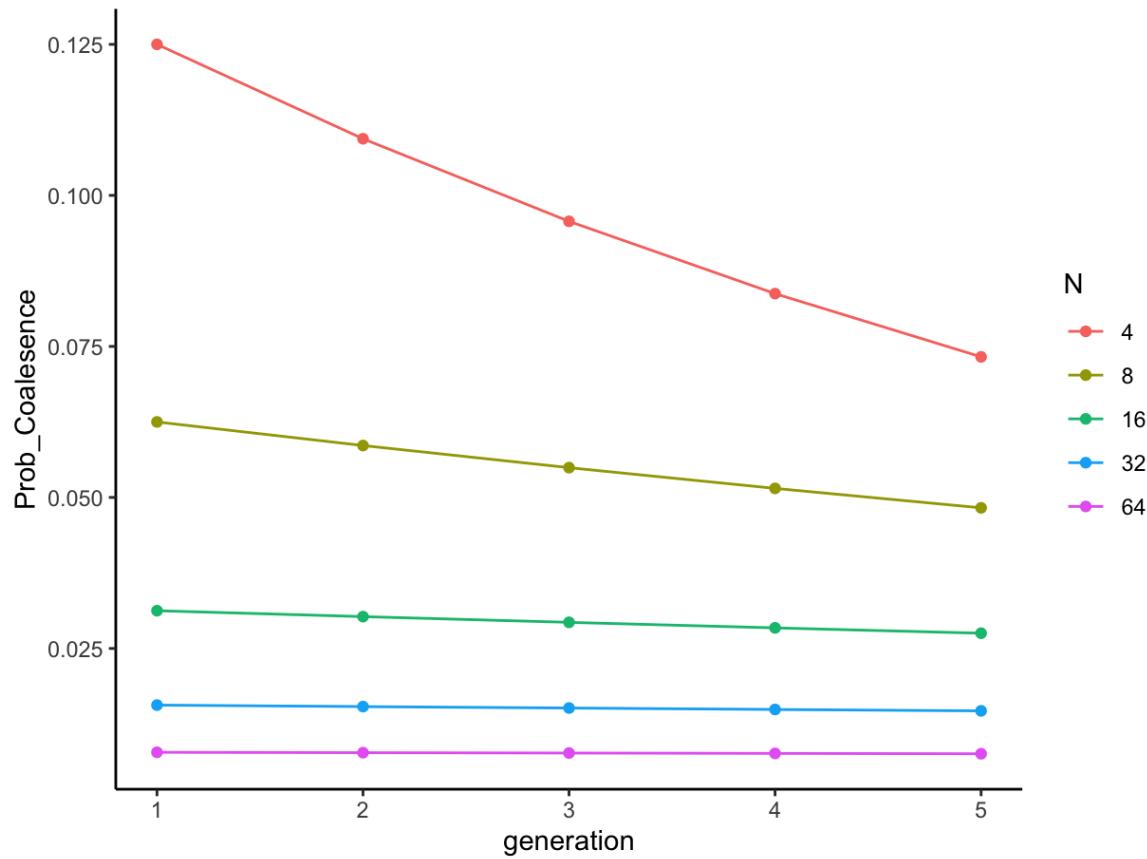
$$\Pr(\text{2 gene copies have same parent in previous gen}) =$$

$$1/(2N)$$

$$\Pr(\text{2 gene copies don't have same parent}) =$$

$$1 - 1/(2N)$$

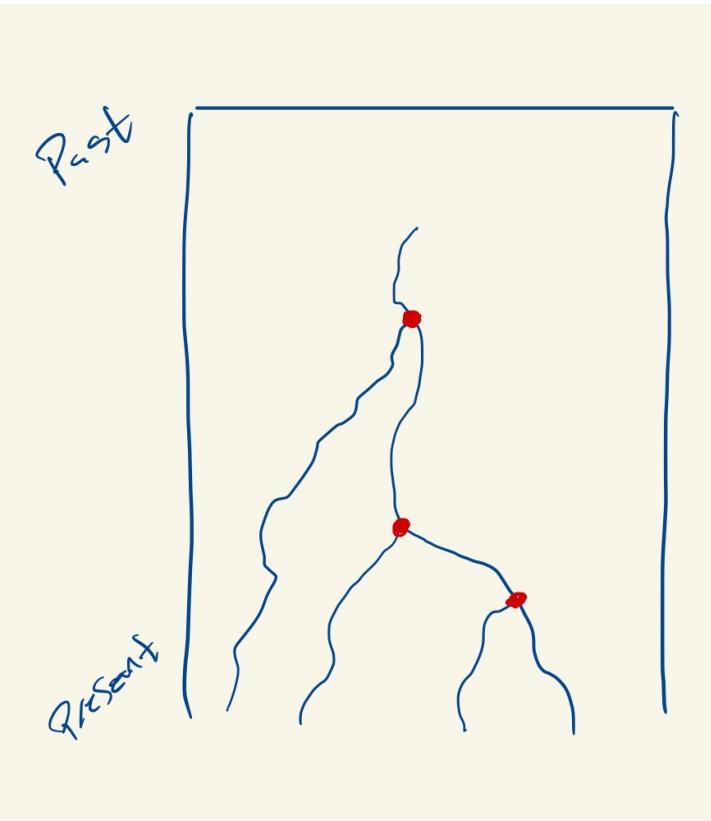
Larger population = longer coalescent time



Mean time to coalescence is $2N$ generations

Large populations have more variation than small ones

Changes in population size have large effect



Coalescence patterns can tell us about demographic history

Genetic variation

$$\theta = 4N_e\mu$$

- The expected amount of variation at a locus
- Links amount of variation to population size
- Large populations have long coalescent times and high variation.

How do we estimate θ ?

- Aka: the expected number of mutations between two DNA sequences?

Genetic variation

- Nucleotide diversity: π (1979)
 - Average number of pairwise differences
 - $\frac{\text{Sum of differences}}{\text{total comparisons}}$
 - Total comparisons: $n(n-1)/2$



Masatoshi Nei



Wen-Hsiung Li

- 2+2+2
- 0+1
- 2
- Total= 9
- Comparisons = $4(3)/2 = 6$
- $\pi = 9/6 = 1.5$

A: 0 0 0
B: 0 1 1
C: 0 1 1
D: 1 0 1

Genetic variation



Watterson's estimator: θ_w (1975)

- The number of segregating sites

$$\theta_w = \frac{s}{\sum_{k=1}^{n-1} \frac{1}{k}}$$

- S = 3
- n = 4
- $1/1 + 1/2 + 1/3 = 1.83$
- $\theta_w = 3/1.83 = 1.64$

Margaret Wu
Not credited until recently

A: 0 0 0
B: 0 1 1
C: 0 1 1
D: 1 0 1

$$\pi = 1.5$$