

SISG 2022 - Module 2

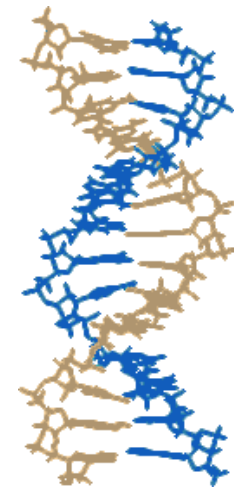
Introduction to Genetics and Genomics

Population Genetics

4pm EDT, Tuesday, July 12th

Joe Lachance and Greg Gibson

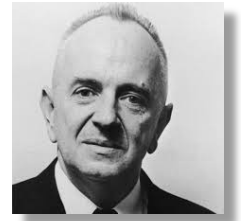
joseph.lachance@biology.gatech.edu



The importance of population genetics

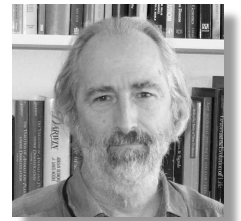
- “Nothing in biology makes sense except in the light of evolution.”

- T. Dobzhansky



- “Nothing in evolution makes sense except in the light of population genetics.”

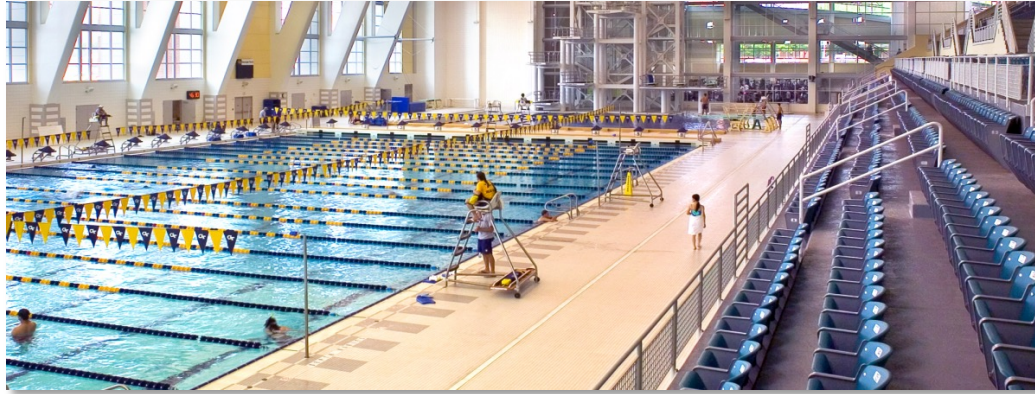
- M. Lynch



- Can you think of any exceptions?



Gene pool metaphor



- Gene pool: aggregate total of all the alleles in a population
- Contributions to next generation's gene pool are weighted by fitness
- Genotypes next generation found by binomial sampling (w/ replacement)

The birth of population genetics (1908)

- The rediscovery of Mendelian genetics in early 20th century was not without some misinterpretations...
 - Brachydactyly is a dominant trait
 - Yule: "75% of people should have short fingers"
 - Punnett: "Something is amiss with that thinking"
- G. H. Hardy correctly inferred the genotype proportions in a randomly mating population, but not without writing:

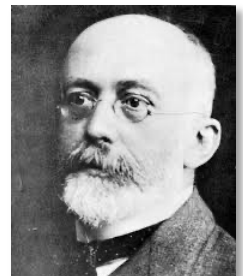
"I would have expected the very simple point which I wish to make to have been familiar to biologists"
- Wilhelm Weinberg, a German physician, independently derived the same result as Hardy



Type E Brachydactyly



G. H. Hardy



Wilhelm Weinberg

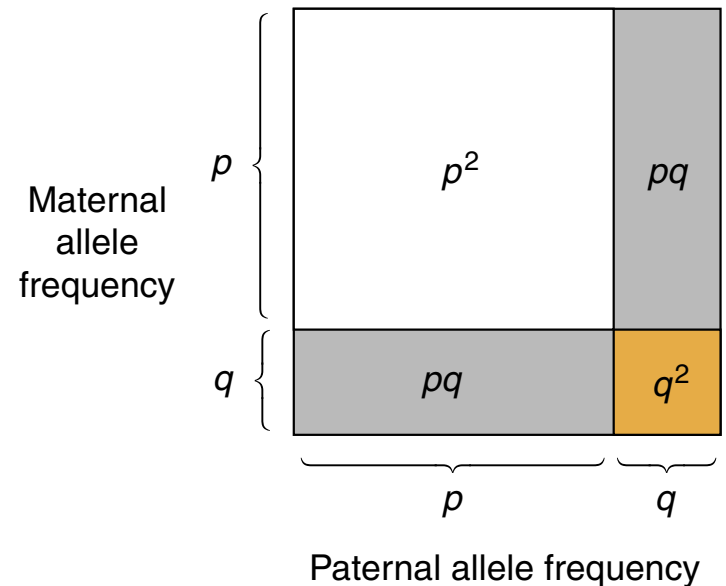
Clearing up some common misconceptions



- Dominant alleles need not be the major (most common) allele
- Higher fitness alleles are not always dominant (and vice versa)
- Higher fitness alleles need not be major alleles

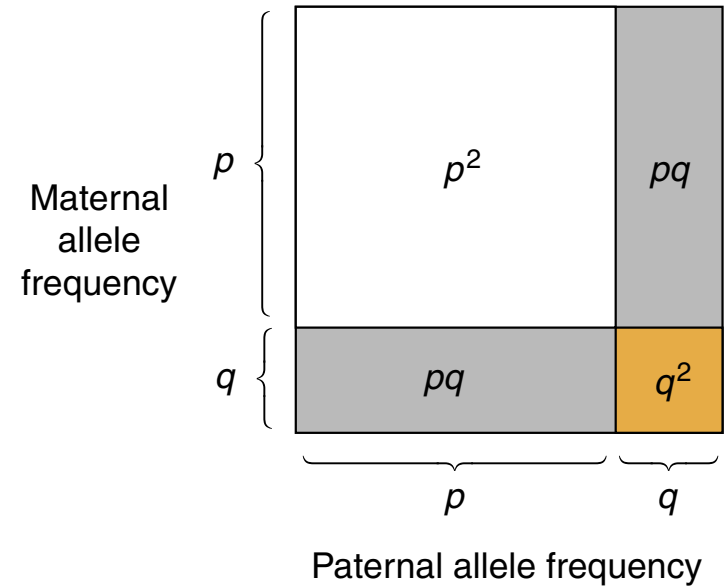
Hardy-Weinberg principle

- Allele frequencies used to calculate genotype frequencies
- Equilibrium reached in a single generation (so long as assumptions hold)
- Allele frequencies
 - Frequency of allele *A*: p
 - Frequency of allele *B*: q
- Genotype frequencies
 - Frequency of *AA* homozygotes: p^2
 - Frequency of *AB* heterozygotes: $2pq$
 - Frequency of *BB* homozygotes: q^2
- $p + q = 1$
- $(p + q)^2 = p^2 + 2pq + q^2 = 1$

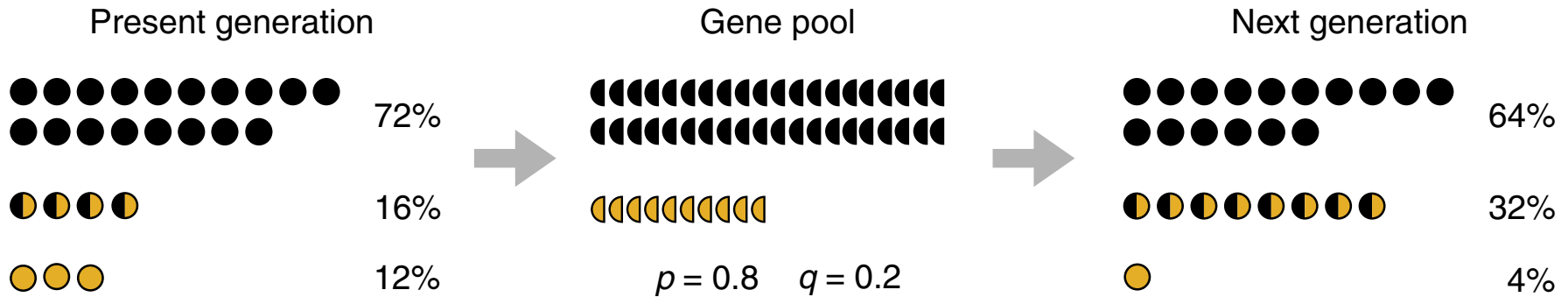


Assumptions of the Hardy-Weinberg principle

- Infinite population size
- No selection
- No mutation
- No migration
- Random mating



Hardy-Weinberg example



- Under Hardy-Weinberg conditions the gene pool remains unchanged
- Subsequent generations will have the same equilibrium genotype frequencies

Testing for departures from HW proportions

- Chi-square test with 1 degree of freedom
- Chi-square (χ^2) > 3.84 indicates statistical significance (p-value < 0.05)
- Example:

Genotype	Observed	Expected	Chi-square
AA	145	131.31	1.426
AB	68	95.37	7.854
BB	31	17.32	10.815
Total	244	244	20.095

$$p = \frac{145 + 68/2}{145 + 68 + 31} = 0.7336$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Using the Hardy Weinberg principle to infer the unseen

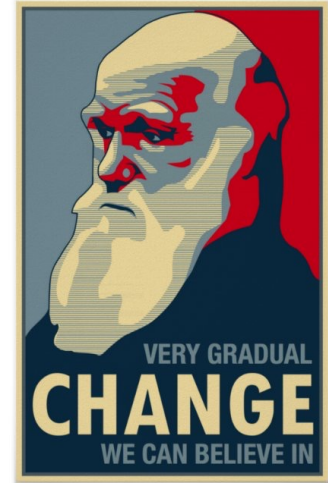


- The wild soybean *Glycine soja*, is polymorphic for flower color, and this trait is controlled by the *W4* locus
- The C allele confers white flowers, and it is recessive to the T allele (which confers purple flowers)
- If 9% of flowers are white....
 - What proportion of all soybean plants do not have a C allele?



Major processes of population genetics

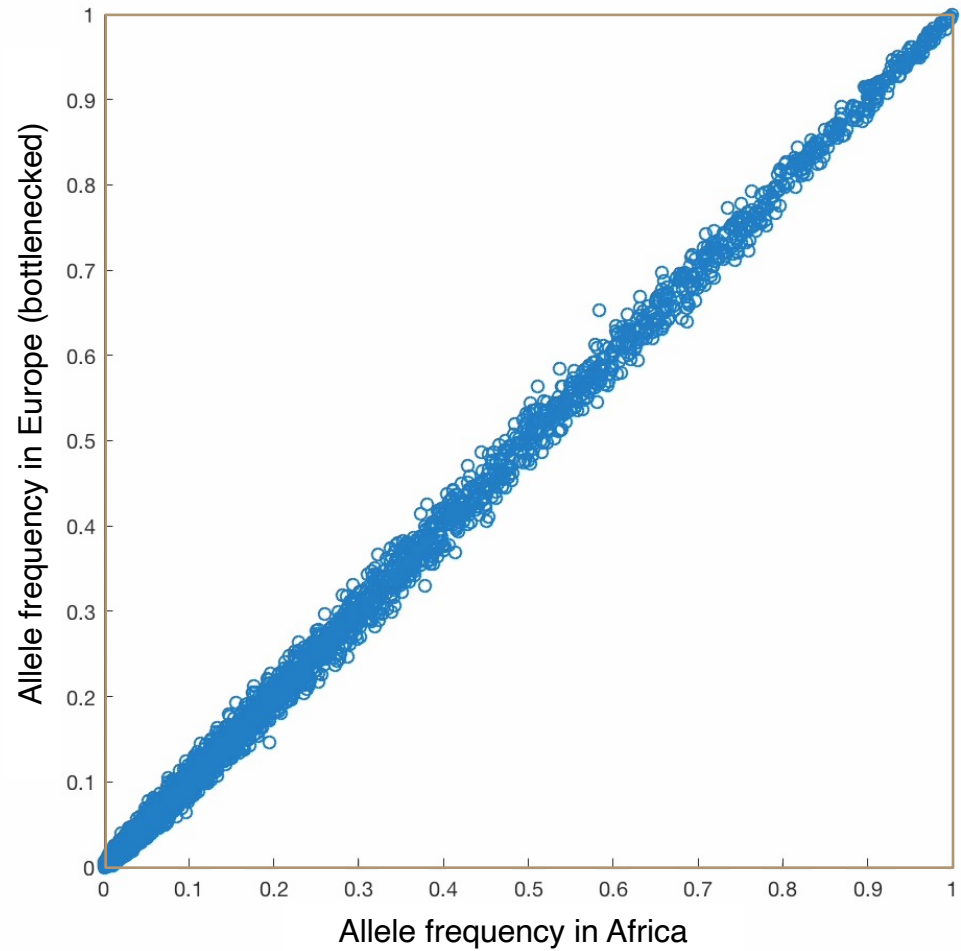
- Genetic drift
 - Natural selection
 - Mutation
 - Migration (gene-flow)
 - Population structure & mating patterns
-
- Each of these processes can lead to departures from Hardy-Weinberg proportions
 - These processes are mechanisms of evolutionary change



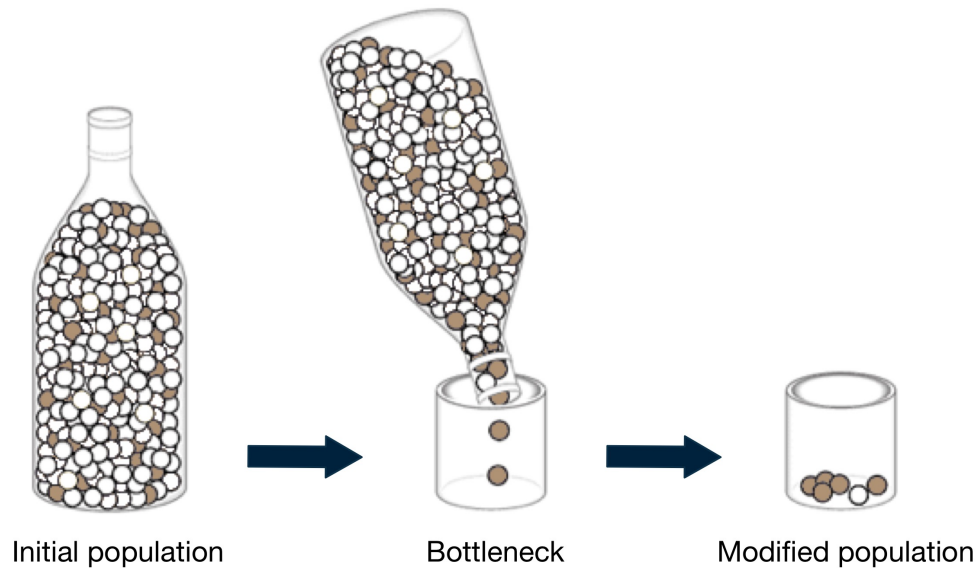
Genetic drift: key points

- Genetic drift is unbiased
- Random fluctuations in allele frequencies are larger in smaller populations
- Drift causes genetic variation to be lost
- Drift causes populations that are initially identical to become different
- An allele can become fixed without the benefit of natural selection

Simulating genetic drift

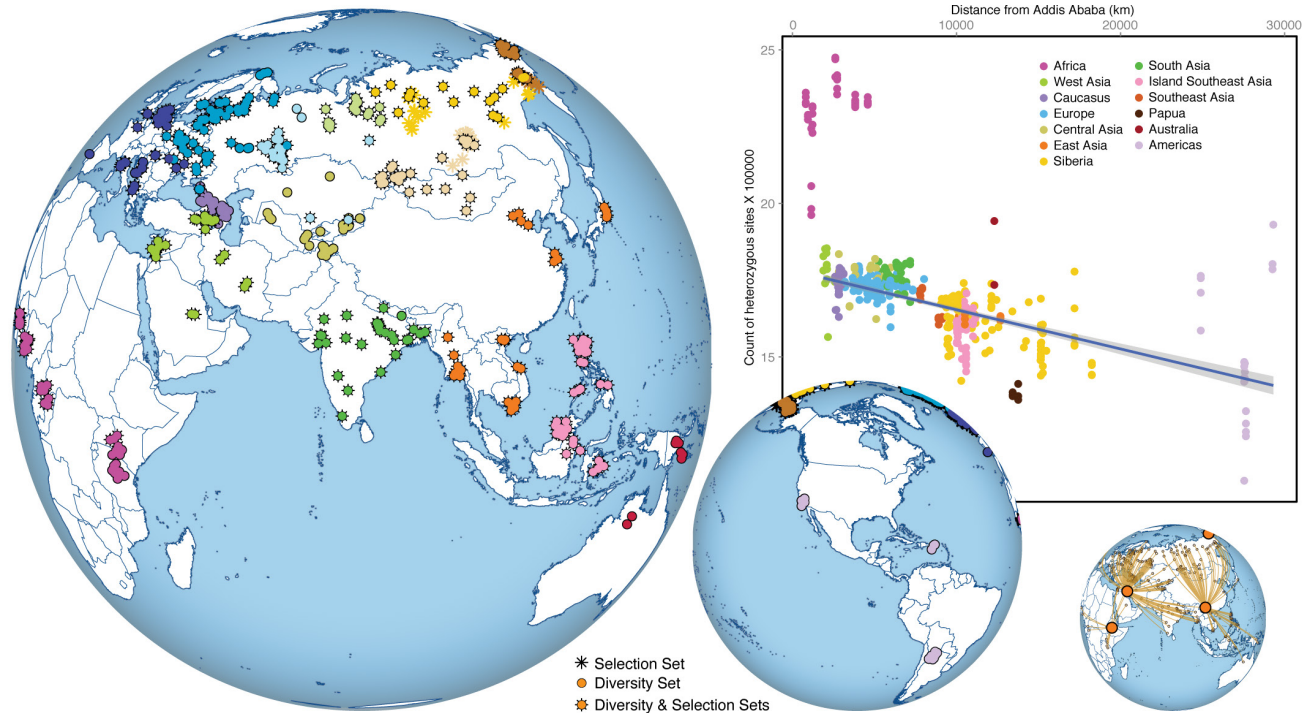


Population bottlenecks and founder effects



- **Population bottleneck:** A sharp reduction in the size of a population
- **Founder effect:** Bottleneck caused by the founding of a new population
- Random chance determines whether an allele increases or decreases in frequency

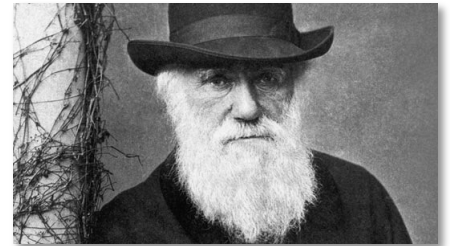
Non-African genomes contain less variation



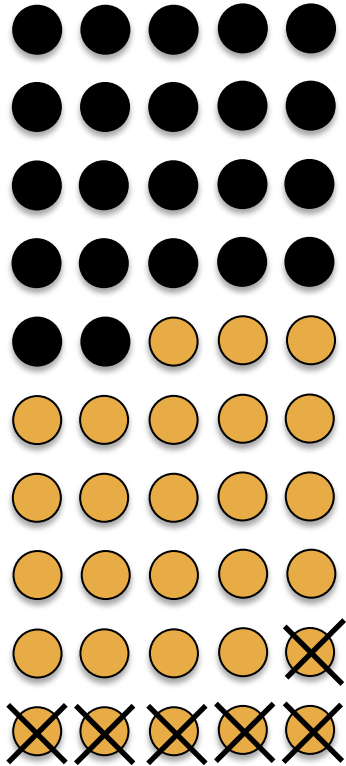
- Serial founder effects reduce the proportion of sites in an individual's genome that are heterozygous

Natural selection

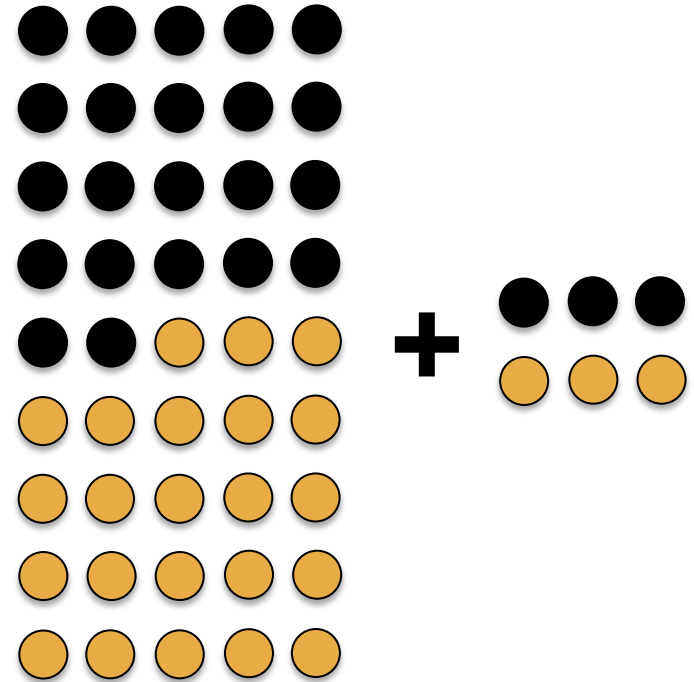
- **Natural selection:** The differential survival and/or reproduction of different genotypes due to unequal fitnesses
- Natural selection is not the same thing as evolution
- Selection coefficient (s)
 - An s of 0.01 indicates a 1% fitness advantage
 - $|s|$ tends to be close to 0
- Operates on short time scales
- The outcome of natural selection depends on fitnesses and initial frequencies
- Probability of fixation: $\sim 2s$
 - Most advantageous mutations are not fixed



Selection changes gene pools



$p = 0.44$



$p = 0.5$

Mathematics of natural selection:

$$\Delta p = p' - p = \frac{pw_A}{\bar{w}} - p$$

- General equation for single generation allele frequency change:

$$\Delta p = \frac{p(w_A - \bar{w})}{\bar{w}}$$

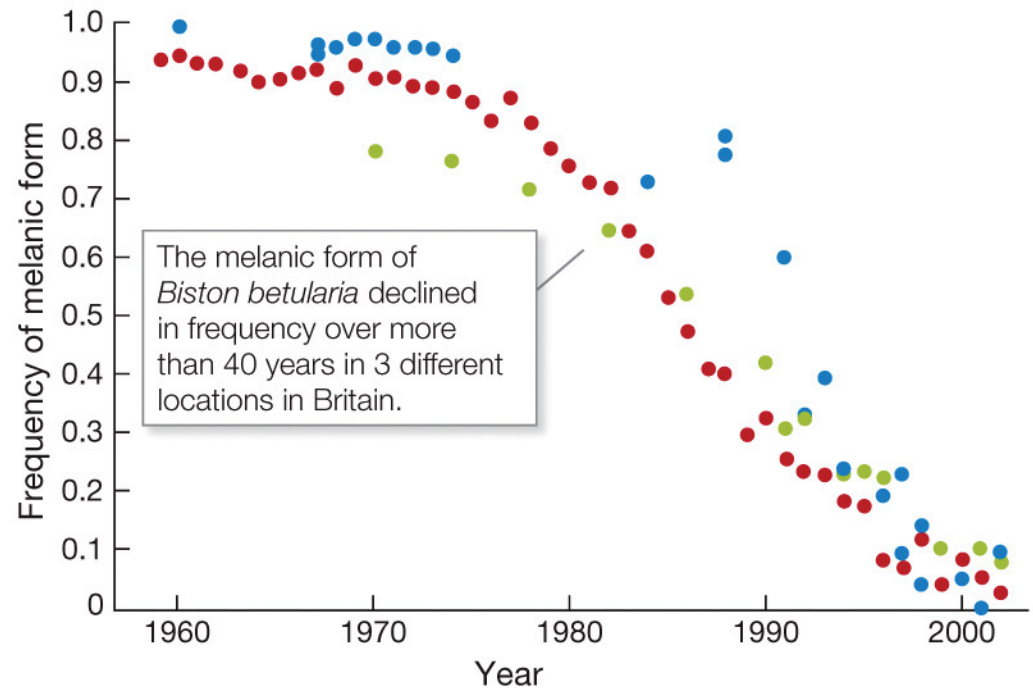
- The response to selection hinges on:
 - Allele frequencies (p)
 - The relative fitness of an allele (w_A)
 - Mean fitness of a population (\bar{w})

Industrial melanism and natural selection

(A)



(B)

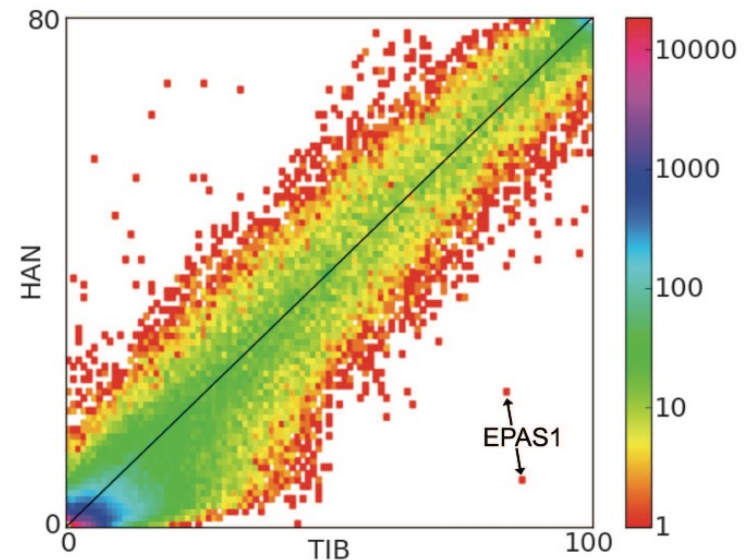


EPAS1 and adaptation to high-altitude

- Reduced $[O_2]$ is a strong selective pressure
- Allele frequencies compared between Tibetans (TIB) and Han Chinese from Beijing (HAN)
- Outlier SNPs are located near *EPAS1*, a hypoxia-induced transcription factor
- The Tibetan *EPAS1* haplotype comes from Denisovans (Huerta-Sanchez et al. 2014)!!!
- Positively selected *EPAS1* haplotype contains a deletion that occurred 12kya (Lou et al. 2015)



Image rights: EasyTourChina



Mutation

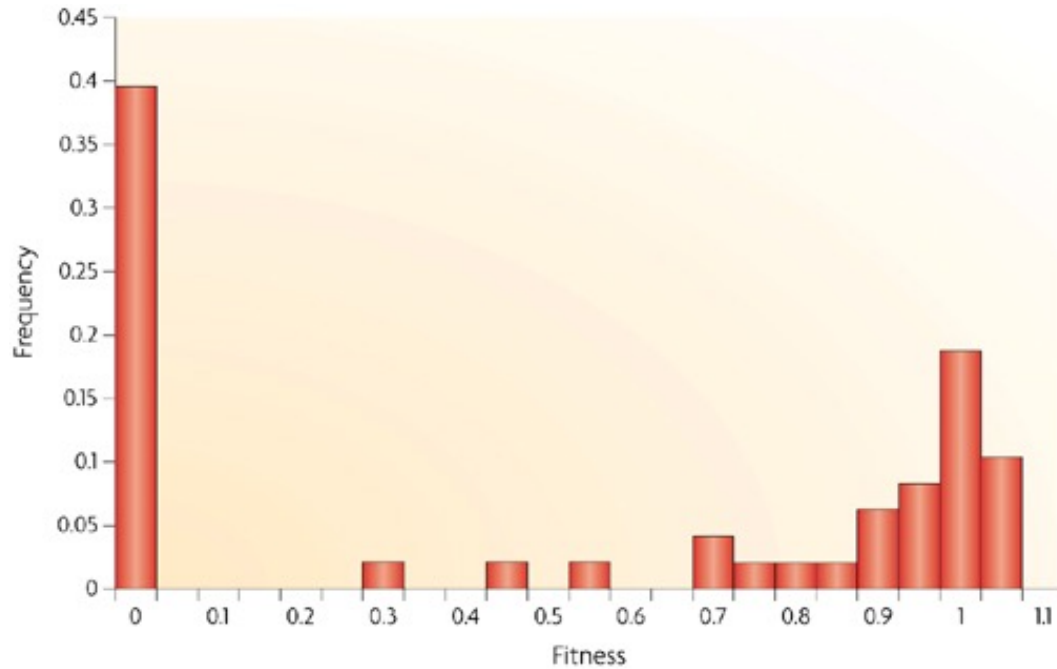
- Human mutation rates
 - $\mu = 2.5 \times 10^{-8} bp^{-1} gen^{-1}$ from comparative genomics (phylogenetic approach)
 - $\mu = 1.2 \times 10^{-8} bp^{-1} gen^{-1}$ from direct sequencing of families
- A “Goldilocks” scenario:
 - Too low a mutation rate and a population will lack genetic diversity
 - Too high of a mutation rate and a population will be unable to purge mutations via natural selection (mutational meltdown, Muller’s ratchet)
- Mutation does not lead to large allele frequency changes in of itself

Mutation example: super-cows?!



- Double-muscling caused by a mutation in the *myostatin* (*GDF8*) gene
- Recurrent mutation in Belgian Blue cattle

Distribution of fitness effects (DFE)

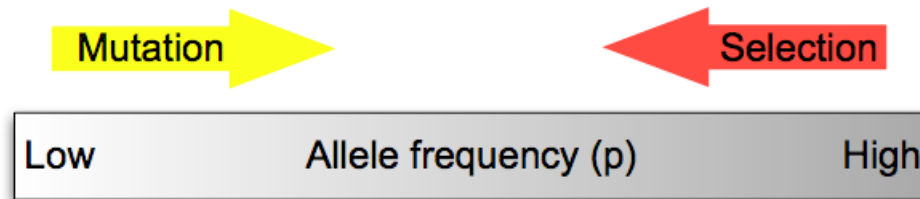


Data source:
Vesicular stomatitis virus

Nature Reviews | Genetics

- Most mutations are deleterious or neutral (they do not increase fitness)
- Mutations of large phenotypic effect are more likely reduce fitness

Mutation-selection balance



- Deleterious alleles are maintained due to a balance between mutation and selection

- Equilibrium allele frequencies: $\hat{p} \approx \sqrt{\frac{\mu}{s}}$ (recessive) $\hat{p} \approx \frac{\mu}{sh}$ (non-recessive)

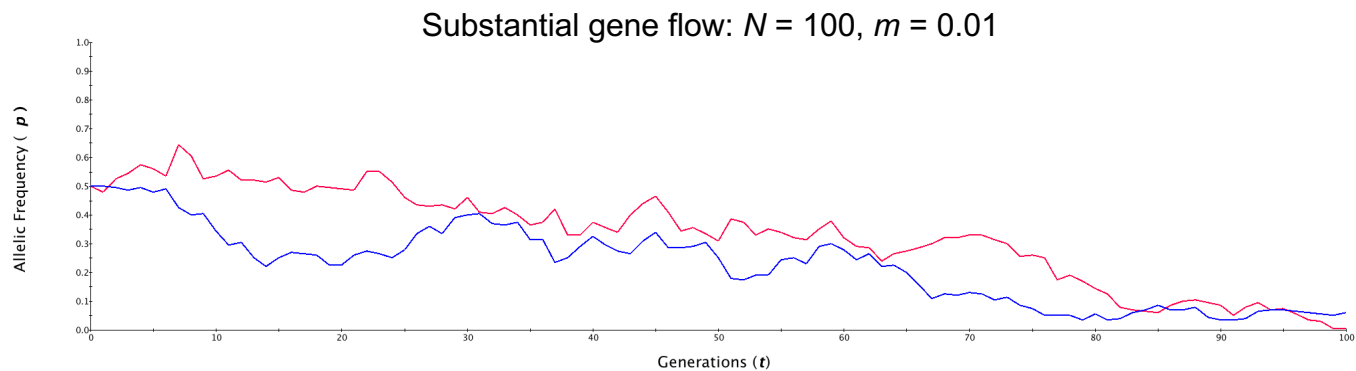
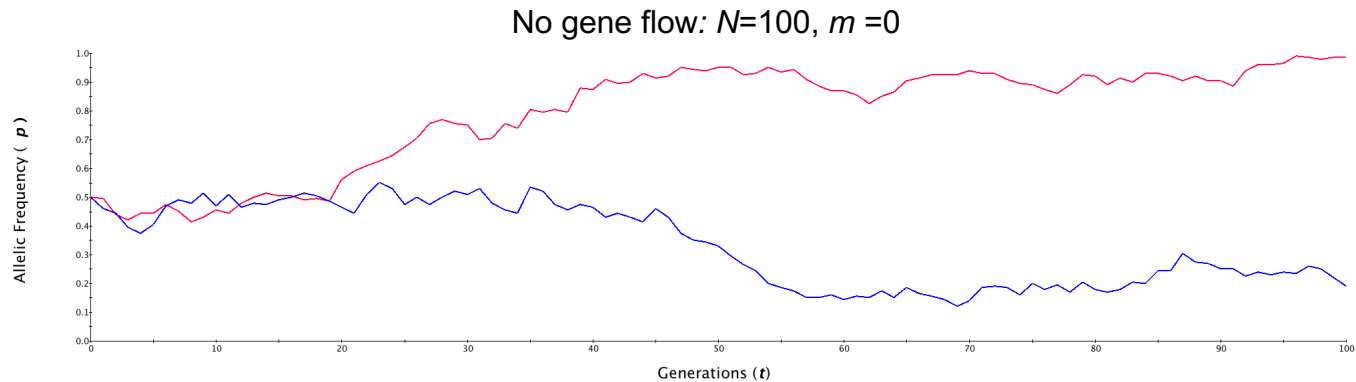
- Implications:
 - Recessive disease alleles can segregate at moderate frequencies
 - Strongly selected disease alleles tend to be rare

Migration

- When population geneticists refer to migration, they mean **gene flow**
- The parameter m equals the proportion of alleles in a population that are from immigrants
- Gene flow homogenizes populations
- Local differentiation occurs when there is < 1 migrant per generation ($Nm < 1$)

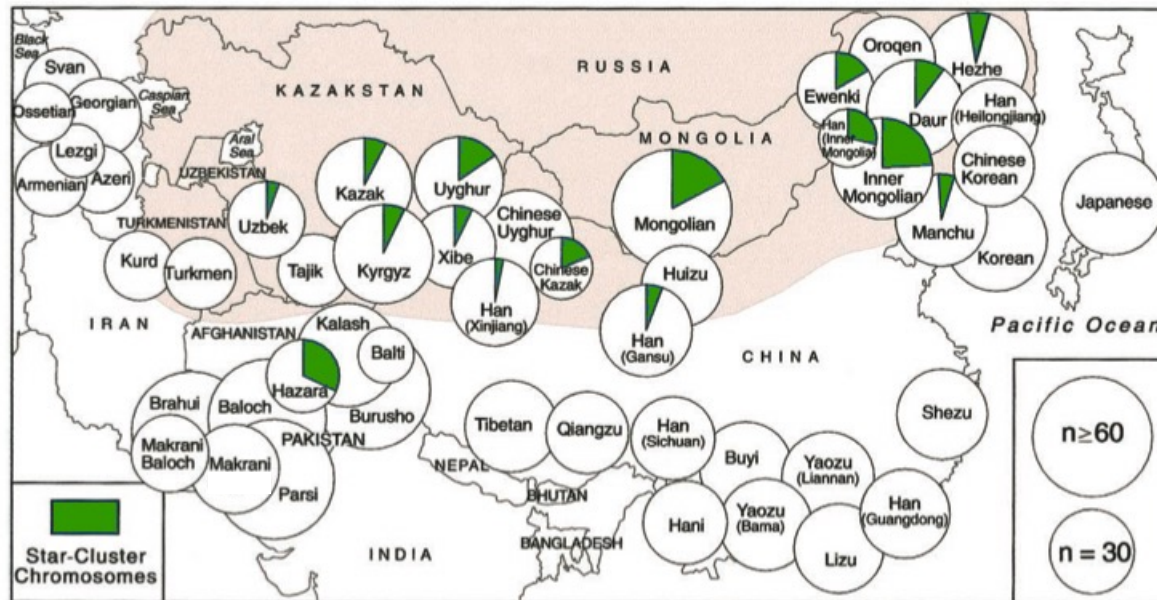


Simulations of migration (and genetic drift)



Migration example

- Geographic proximity results in genetic similarity



- The Y-chromosome legacy of Ghengis Khan
(Zerjal et al. 2003, American Journal of Human Genetics)

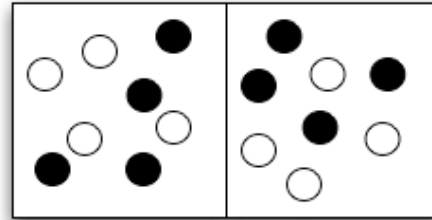
Assortative mating



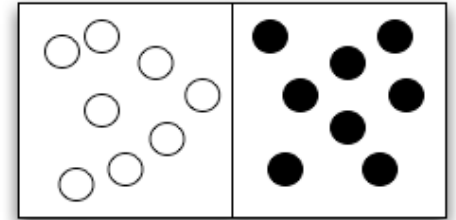
- Positive assortative mating
 - Phenotypically similar individuals prefer to mate with each other
 - Can result in the maintenance of different phenotypes
- Negative assortative mating
 - Phenotypically different individuals prefer to mate with each other
 - Maintains genetic variation despite loss of phenotypic variation

Population structure: F_{ST}

$$F_{ST} = \frac{Var(p)}{\bar{p}(1 - \bar{p})}$$



$$F_{ST} = 0$$

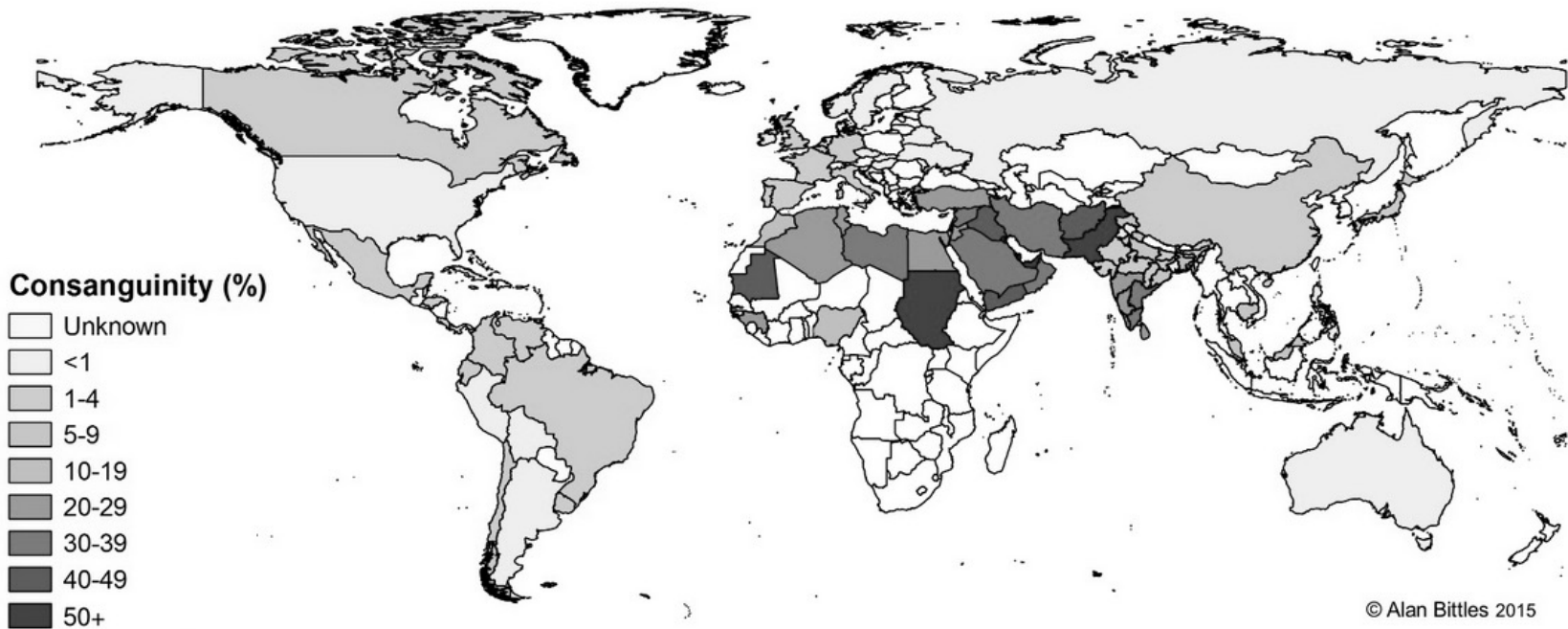


$$F_{ST} = 1$$

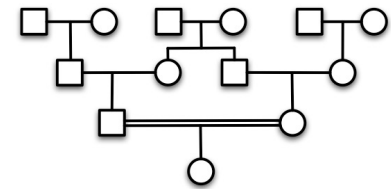
- F_{ST} measures how much genetic variation can be explained by sub-populations within the total population

- F_{ST} between divergent populations increases over time $F_{ST} = 1 - \left(1 - \frac{1}{2N}\right)^t$

Inbreeding is widespread

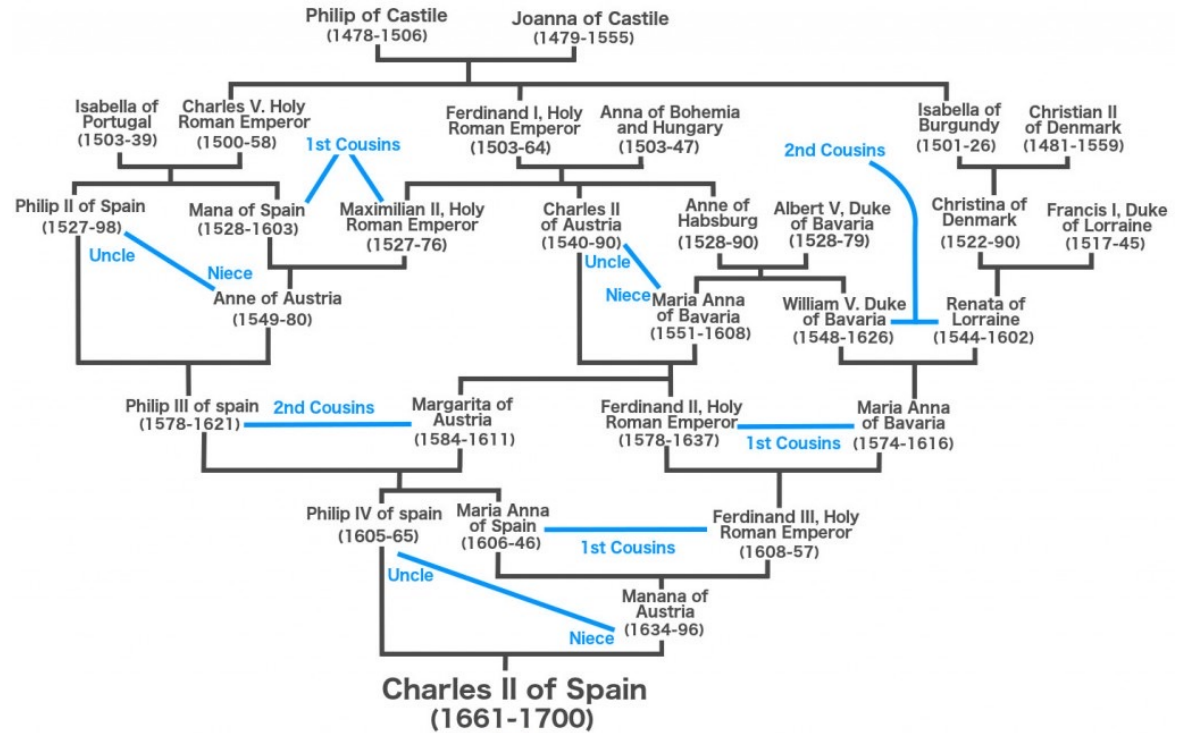


- **Inbreeding:** preferential mating between relatives



- Inbreeding coefficients quantify excess amounts of homozygosity

Effects of inbreeding



- Inbreeding can have negative consequences if disease alleles are recessive

Effects of each major process

	Genetic Drift	Natural Selection	Mutation	Migration	Mating Structure
Time-scale	Medium	Fast	Slow	Medium	Fast
Effect on variation	Reduced	"It depends"	Increased	Homogenized	"It depends"