Selection

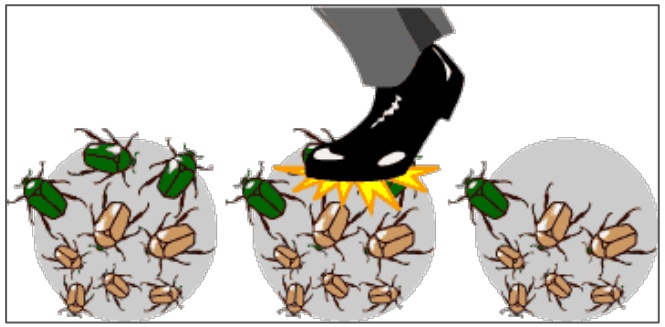
Hancock February 22

Evolutionary processes

Factors that alter allele frequencies from one generation to the next

- Genetic drift
- Gene flow
- Selection

Neutral processes



evolution.berkeley.edu/evolibrary/article/evo_24

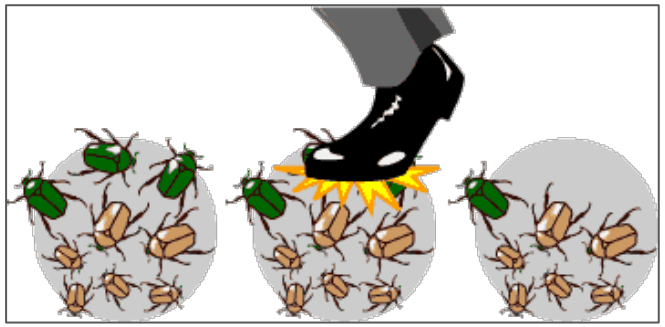
Evolutionary processes

Factors that alter allele frequencies from one generation to the next

- Genetic drift
- Gene flow

Neutral processes

- Selection



evolution.berkeley.edu/evolibrary/article/evo_24

Natural selection requires:

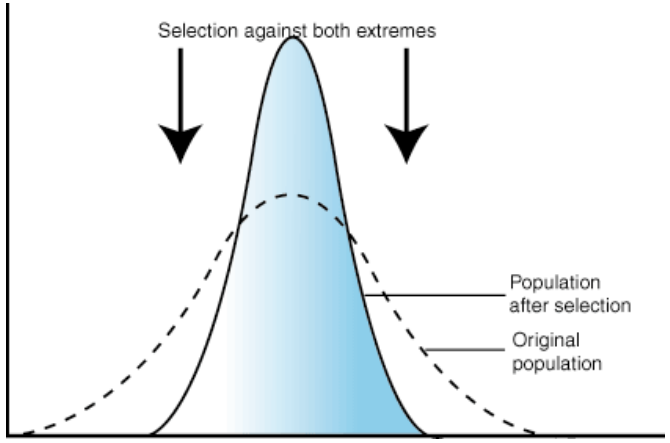
- A trait that varies in the population
- The trait must impact fitness (survival or reproductive success)
- The traits must be genetically encoded (at least in part)

Types of selection

- **Stabilizing selection:** Purifying selection, selection that favors stasis and maintenance of the current trait value
  - Selection acts against new mutations that would pull the trait away from its optimal value
- **Directional selection**: selection that favors alleles that either increase or decrease a trait value
  - Truncating selection is an extreme case, where only those with extreme trait values reproduce
- **Balancing selection**: selection coefficient is condition-dependent
  - Heterozygote advantage (overdominant selection)
  - Disruptive selection (selection favoring extreme values)
  - Negative frequency dependent selection (selection for less common trait)

Selection can affect traits in different ways

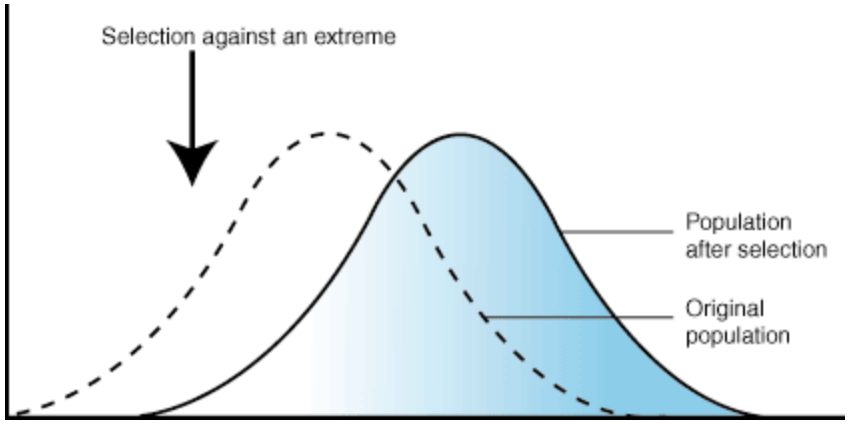
Stabilizing selection



Selection that removes variation from the population

Selection maintains the status quo
The most common form of selection

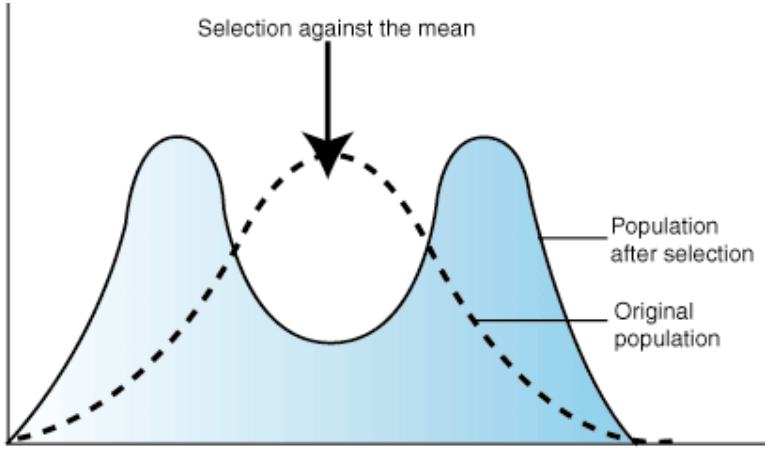
Directional selection



Selection that changes the mean trait value in the population

Type of selection responsible for adaptation to a novel environment

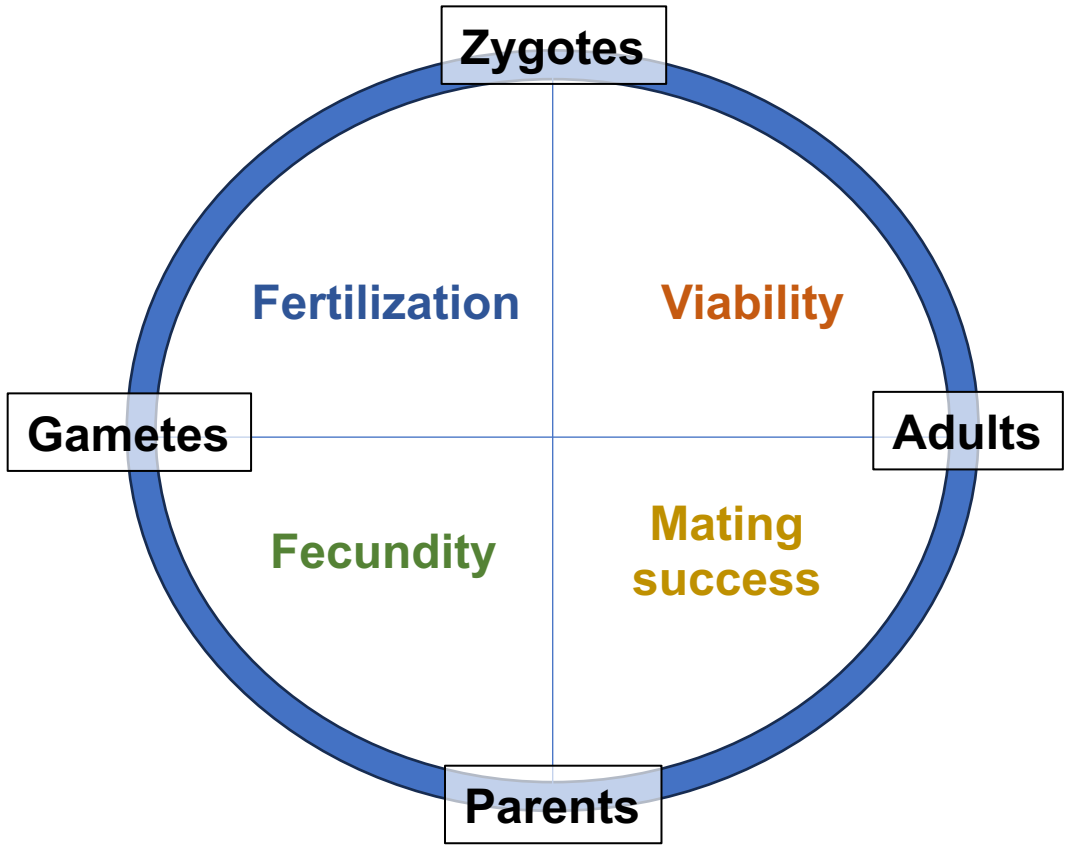
Disruptive selection



Selection that maintains variation in the population

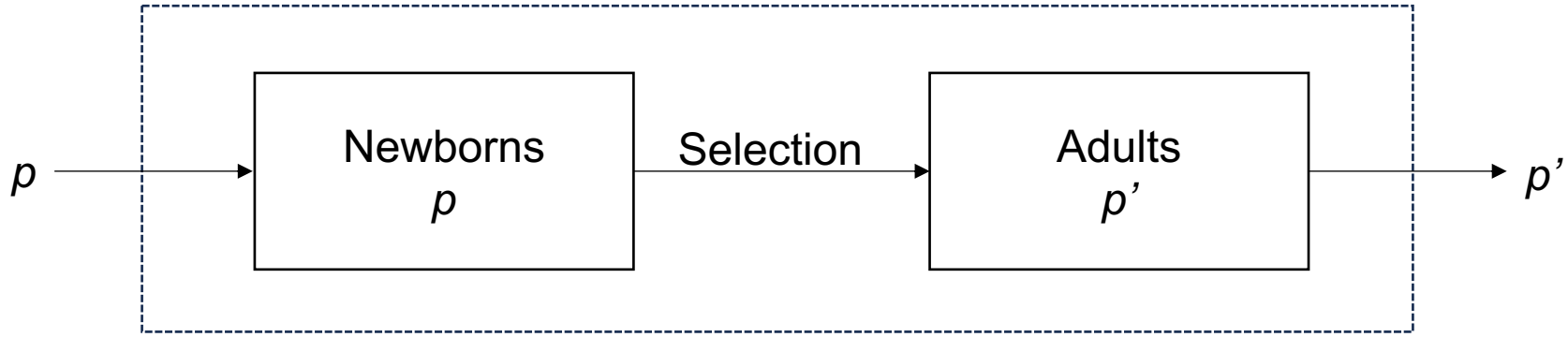
Selection that occurs when there are differences in pressures across time or space

Fitness is the currency of selection



There are many potential components of fitness that may act at different points in the life cycle

The simple life cycle used in the fundamental model



Gillespie, figure 3.2

Selection at a single locus

The model:

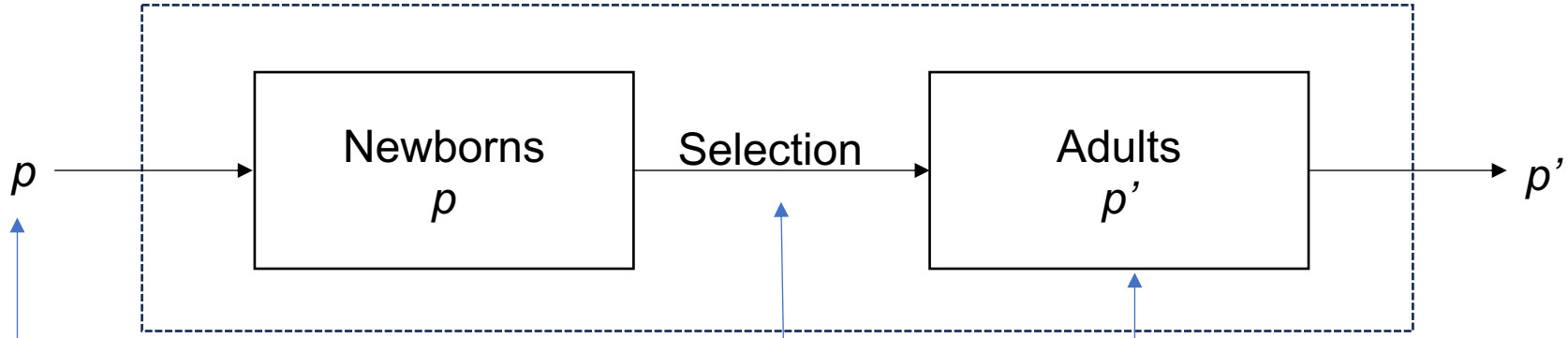
- One locus, two alleles, $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$
- Frequency of $A_1 = p$ and frequency of $A_2 = q = 1\text{-}p$
- W_{12} = *fitness of* A_1A_2 = the probability that an individual of genotype A_1A_2 survives to reproduce
- Before selection, the frequencies of A_1 and A_2 are in HWE
- The contribution of each genotype to the gene pool of the next generation is proportional to the product of its frequency and fitness

See "Not quite enough selection (or maybe a bit too much)" in course material for more information

The simple life cycle used in the fundamental model

One locus, two alleles, A_1 and A_2

The contribution of each genotype to the gene pool of the next generation is proportional to the product of its frequency and fitness



Starting frequency of $A_1 = p$ and frequency of $A_2 = q = 1\text{-}p$

Before selection, the frequencies of A_1 and A_2 are in HWE

One generation

Allele frequency after selection is proportional to newborn frequency x viability

W_{ij} = *fitness of* A_1A_2 = the probability that an individual of genotype A_1A_2 survives to reproduce

Gillespie, figure 3.2

Genotypic fitnesses and frequencies

2 allele model; p = 1-q

Genotype	A_1A_1	A_1A_2	A_2A_2
Frequency before selection	p^2	$2pq$	q^2
Fitness	W_{11}	W_{12}	W_{22}
Frequency after selection	$\frac{p^2W_{11}}{\bar{W}}$	$\frac{2pqW_{12}}{\bar{W}}$	$\frac{q^2W_{22}}{\bar{W}}$

$$\bar{W} = \rho^2 W_{11} + 2pqW_{12} + q^2W_{22}$$

$\bar{W}$ = mean fitness, the weighted average of fitness for each genotype

Genotypic fitnesses and frequencies

Genotype	A_1A_1	A_1A_2	A_2A_2
Frequency before selection	p^2	$2pq$	q^2
Fitness	W_{11}	W_{12}	W_{22}
Frequency after selection	$\frac{p^2W_{11}}{\bar{W}}$	$\frac{2pqW_{12}}{\bar{W}}$	$\frac{q^2W_{22}}{\bar{W}}$

← These are genotype frequencies so they sum to 1

$$\bar{W} = p^2W_{11} + 2pqW_{12} + q^2W_{22}$$

Genotypic fitnesses and frequencies

Genotype	A_1A_1		A_1A_2		A_2A_2	
Frequency before selection	p^2	+	$2pq$	+	q^2	= 1
Fitness	W_{11}		W_{12}		W_{22}	
Frequency after selection	$\frac{p^2W_{11}}{\bar{W}}$	+	$\frac{2pqW_{12}}{\bar{W}}$	+	$\frac{q^2W_{22}}{\bar{W}}$	= 1

$$\bar{W} = p^2W_{11} + 2pqW_{12} + q^2W_{22}$$

How frequencies change over time

$$p' = \frac{p^2 W_{11}}{\bar{W}} + \frac{pqW_{12}}{\bar{W}}$$

The frequency in the next generation, p', is the sum of the contributions from A_1A_1 and A_1A_2 genotypes

$$p' = \frac{p^2 W_{11}}{\bar{W}} + \frac{pqW_{12}}{\bar{W}} = \frac{p(pW_{11} + qW_{12})}{\bar{W}}$$

$$\bar{W}_1 = pW_{11} + qW_{12}$$

This term is referred to as the marginal fitness of A_1

$$\bar{W}_2 = pW_{12} + qW_{22}$$

And the marginal fitness of A_2

The frequency of the A_1 allele after selection can be written as:

$$p' = \frac{p\bar{W}_1}{\bar{W}}$$

How frequencies change over time

The frequency of A_1 after selection can be written as: $p' = \frac{p\bar{W}_1}{\bar{W}}$

To get the change in *p*, subtract *p* from *p'*:

$$\Delta p = p' - p = \frac{p\bar{W}_1}{\bar{W}} - p = \frac{p(\bar{W}_1 - \bar{W})}{\bar{W}}$$

The A_1 allele's frequency will increase or decline depending on whether its fitness is greater than or less than the mean fitness

With substitution $\bar{W} = p\bar{W}_1 + q\bar{W}_2$ and rearrangement:

$$\Delta p = \frac{p\bar{W}_1}{\bar{W}} - p = \frac{p(\bar{W}_1 - \bar{W})}{\bar{W}} = \boxed{pq\,\frac{\bar{W}_1 - \bar{W}_2}{\bar{W}}}$$

The frequency of allele A_1 depends on the difference between its fitness and the fitness of allele A_2

Fitnesses can also be expressed relative to the fitness of the A_1 homozygote

Genotype	A_1A_1	A_1A_2	A_2A_2
fitness	W_{11}	W_{12}	W_{22}
Relative fitness	1	$\frac{W_{12}}{W_{11}}$	$\frac{W_{22}}{W_{11}}$

Fitnesses can also be expressed relative to the fitness of the A_1 homozygote

Genotype	A_1A_1	A_1A_2	A_2A_2
fitness	W_{11}	W_{12}	W_{22}
Relative fitness	1	$\frac{W_{12}}{W_{11}}$	$\frac{W_{22}}{W_{11}}$
Alternative notation for relative fitness	1	1-*hs*	1-*s*

where *h* is the heterozygous effect.

Fitnesses can also be expressed relative to the fitness of the A_1 homozygote

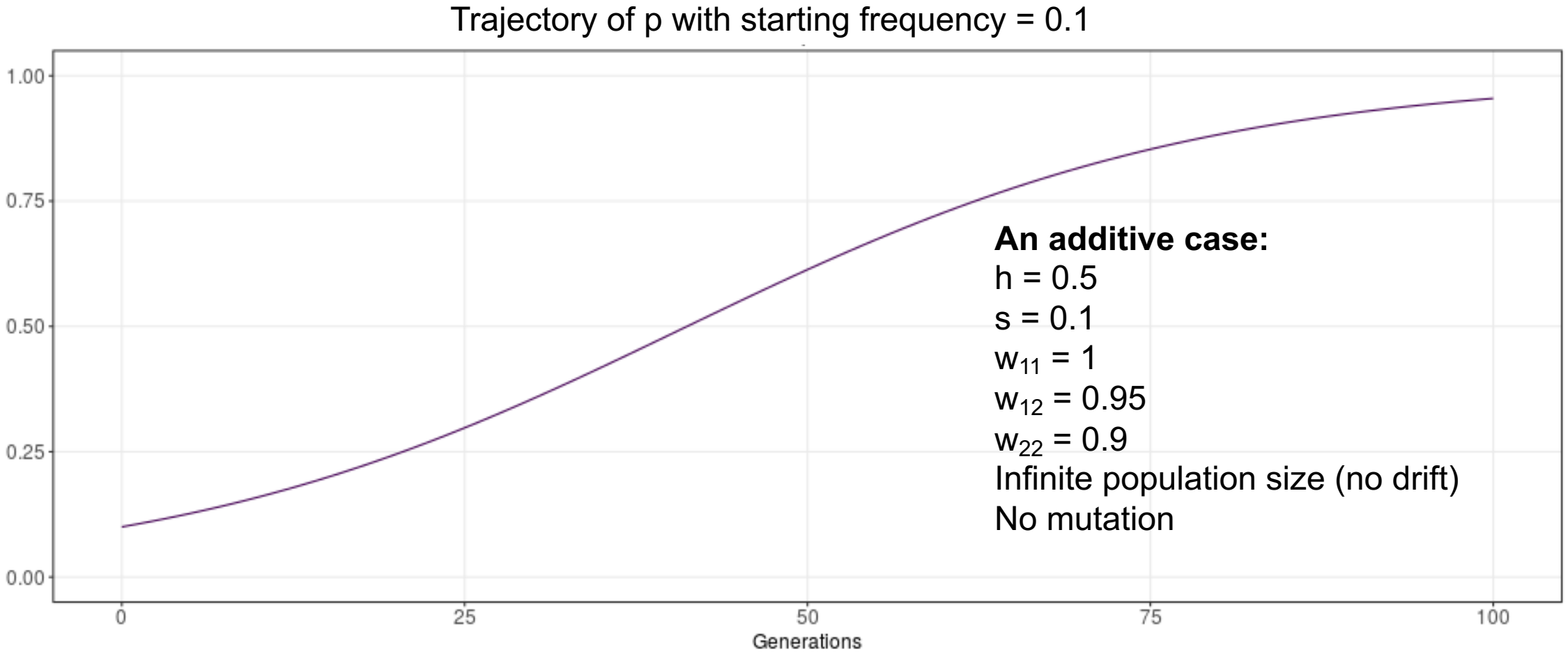
Genotype	A_1A_1	A_1A_2	A_2A_2
fitness	W_{11}	W_{12}	W_{22}
Relative fitness	1	$\frac{W_{12}}{W_{11}}$	$\frac{W_{22}}{W_{11}}$
Alternative notation for relative fitness	1	1-*hs*	1-*s*

where *h* is the heterozygous effect.

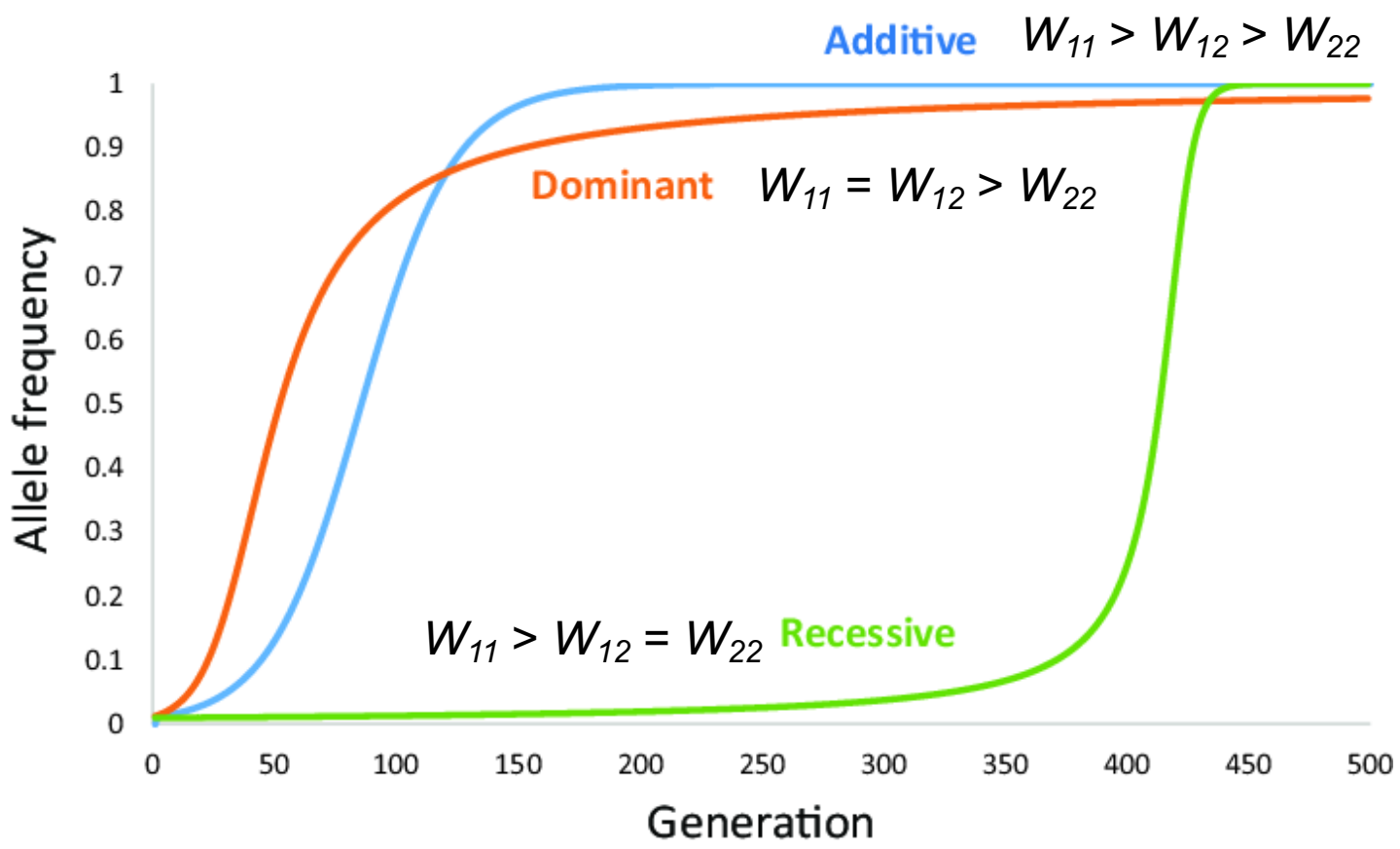
***h* is a measure of dominance.**

$h = 0$	A_1 dominant, A_2 recessive
$h = 1$	A_2 dominant, A_1 recessive
$0 < h < 1$	Incomplete dominance
$h < 0$	Overdominance (heterozygote advantage)
$h > 1$	Underdominance

An allele frequency trajectory under directional selection



Allele frequency trajectories under different degrees of dominance

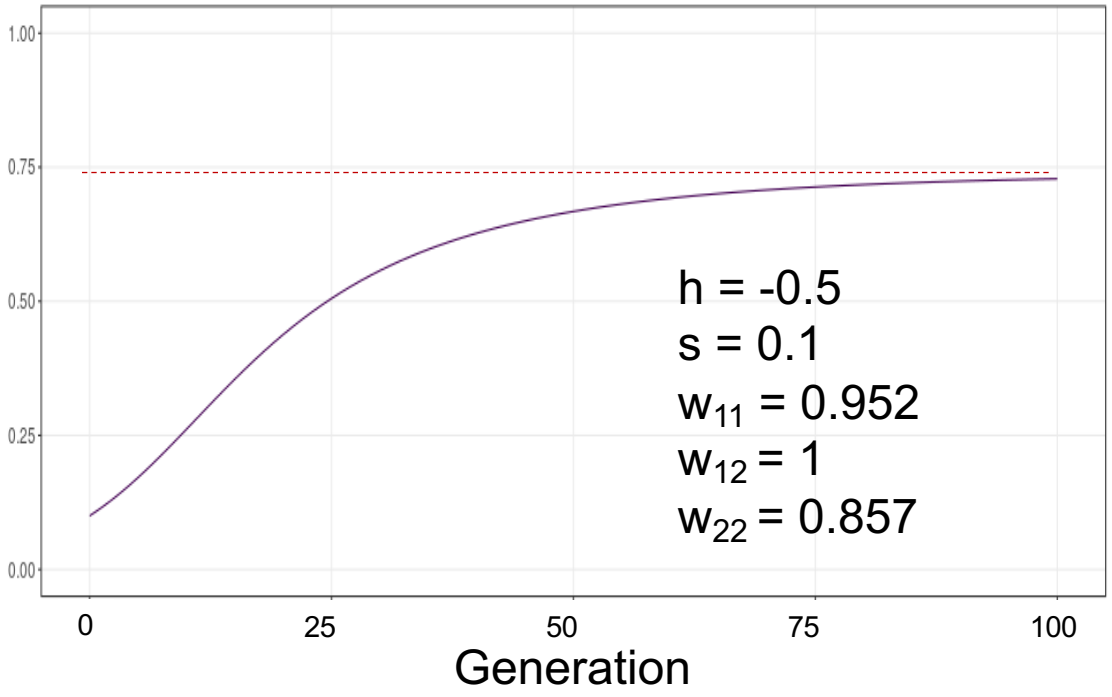


Some patterns that emerge:

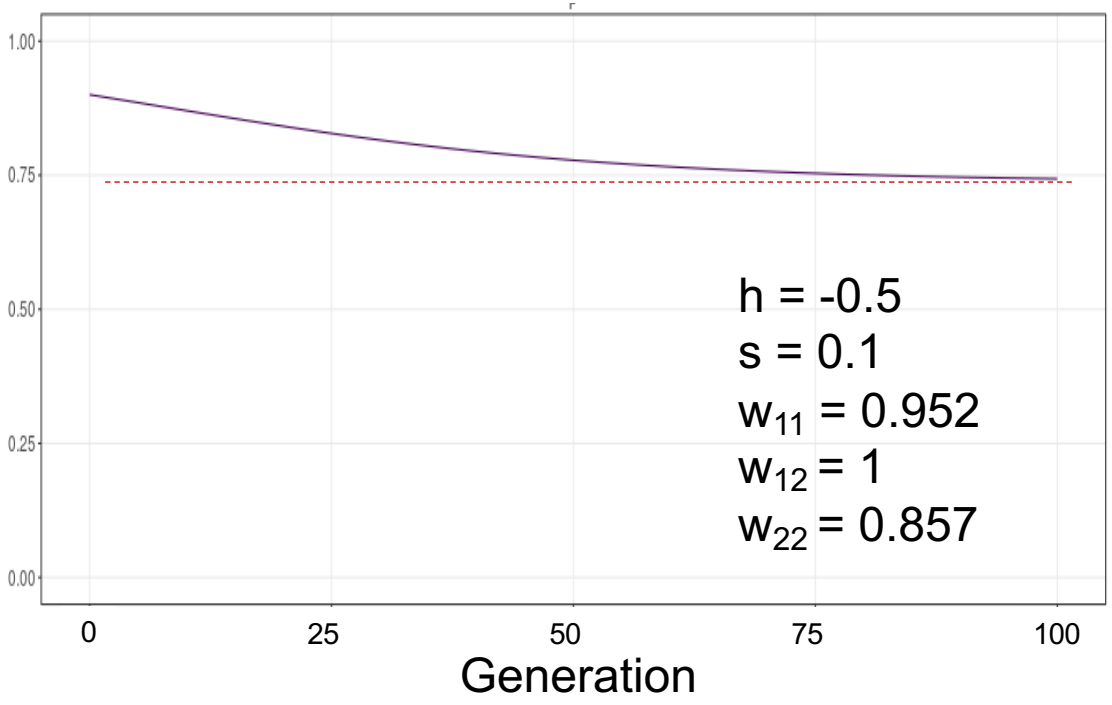
- The response to selection is slow when either allele is rare
- Selection is especially slow when a beneficial recessive allele is rare
- Selection is especially slow when a new recessive allele is advantageous

Overdominance (Heterozygote advantage)

Trajectory of p with starting frequency = 0.1



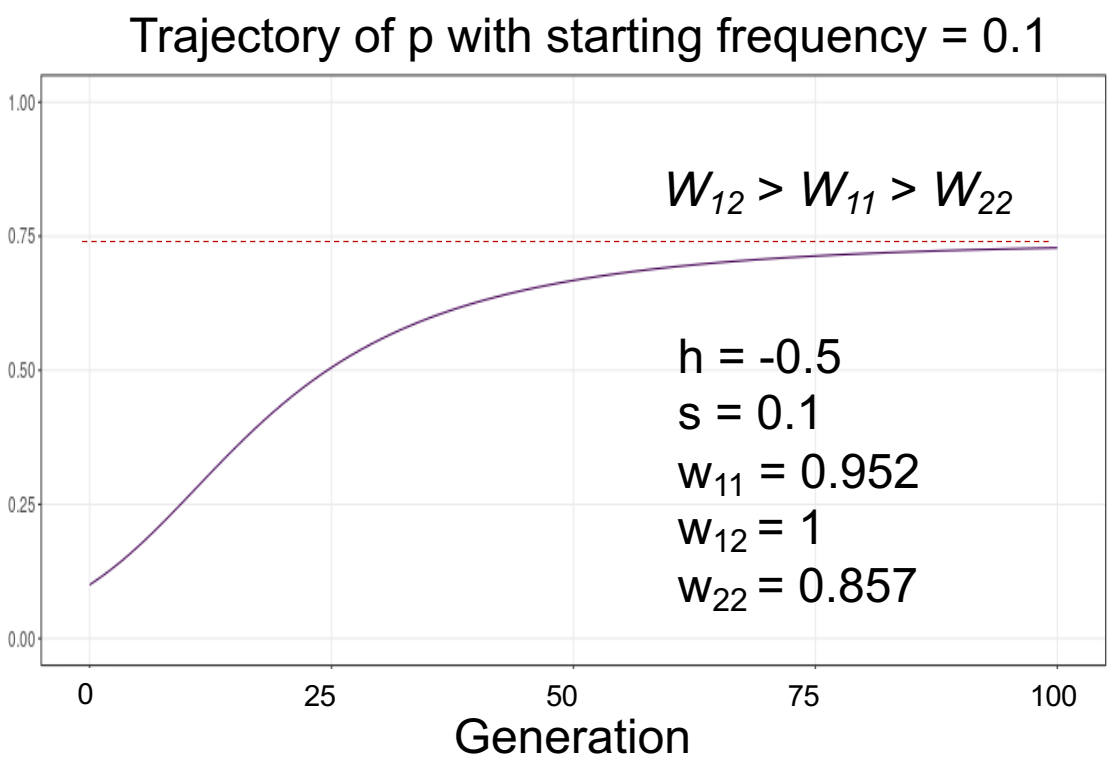
Trajectory of p with starting frequency = 0.9

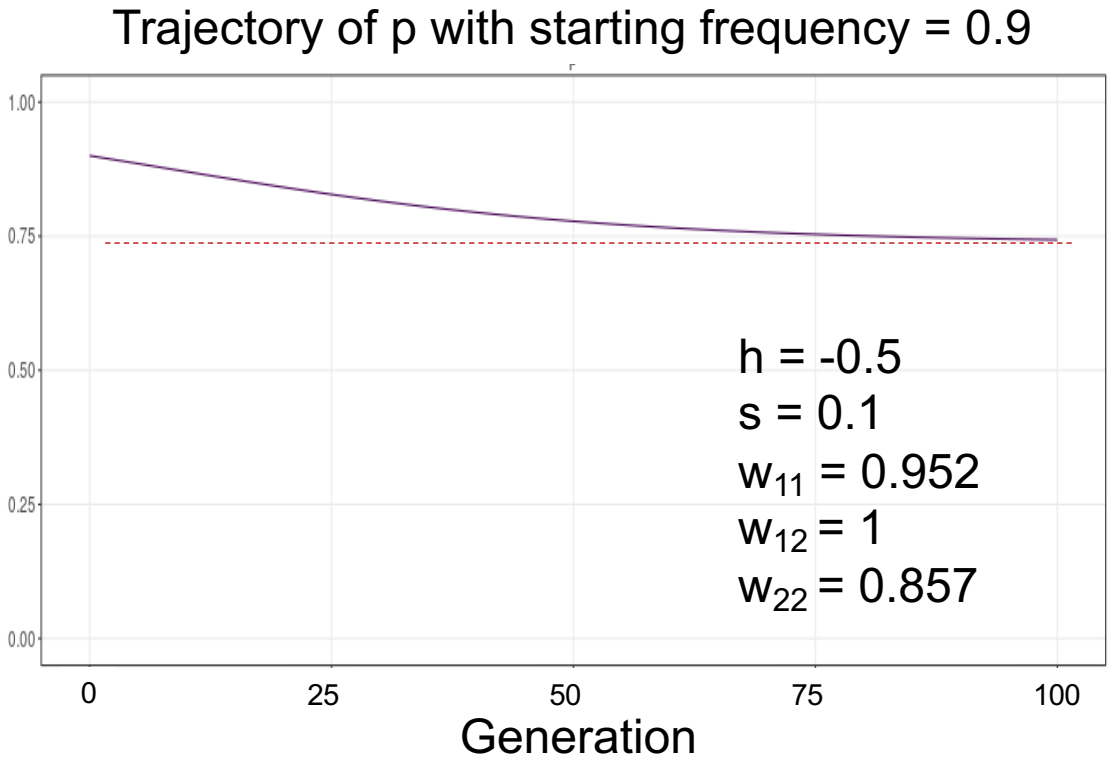


With s = 0.1, equilibrium is reached at: $\hat{p} = \frac{h-1}{2h-1} = \frac{-0.5-1}{-1-1} = \frac{1.5}{2} = 0.75$

Gillespie eq. 3.2 and 3.4

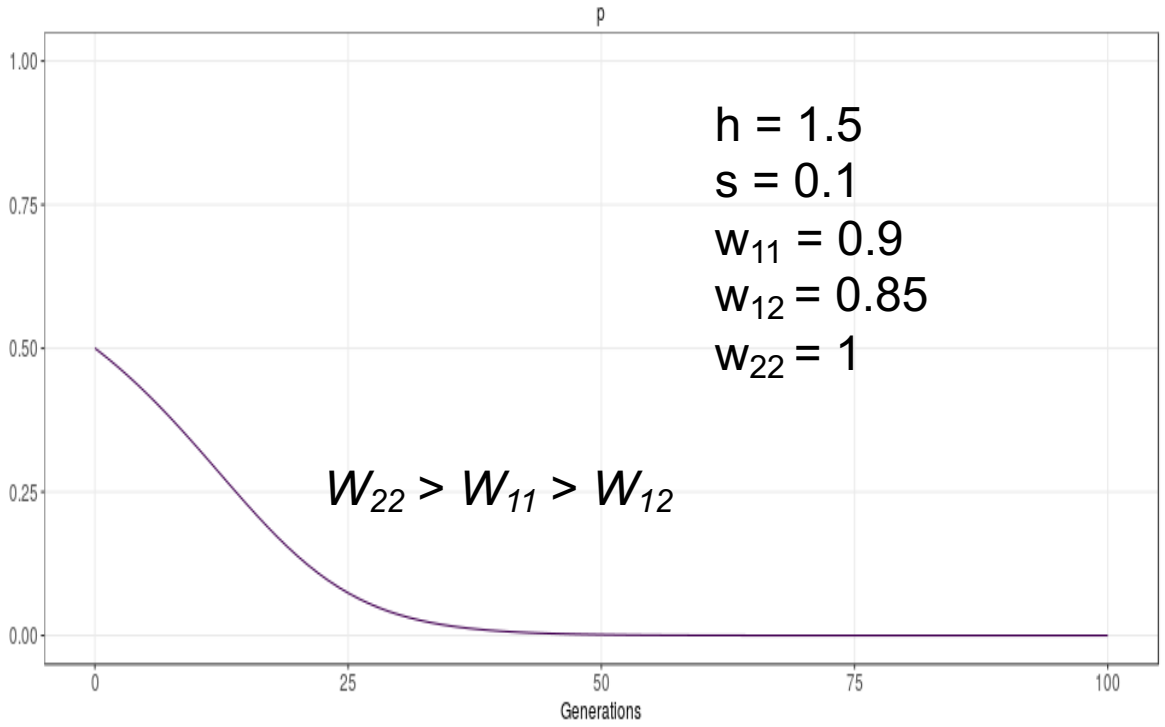
Overdominance (Heterozygote advantage)

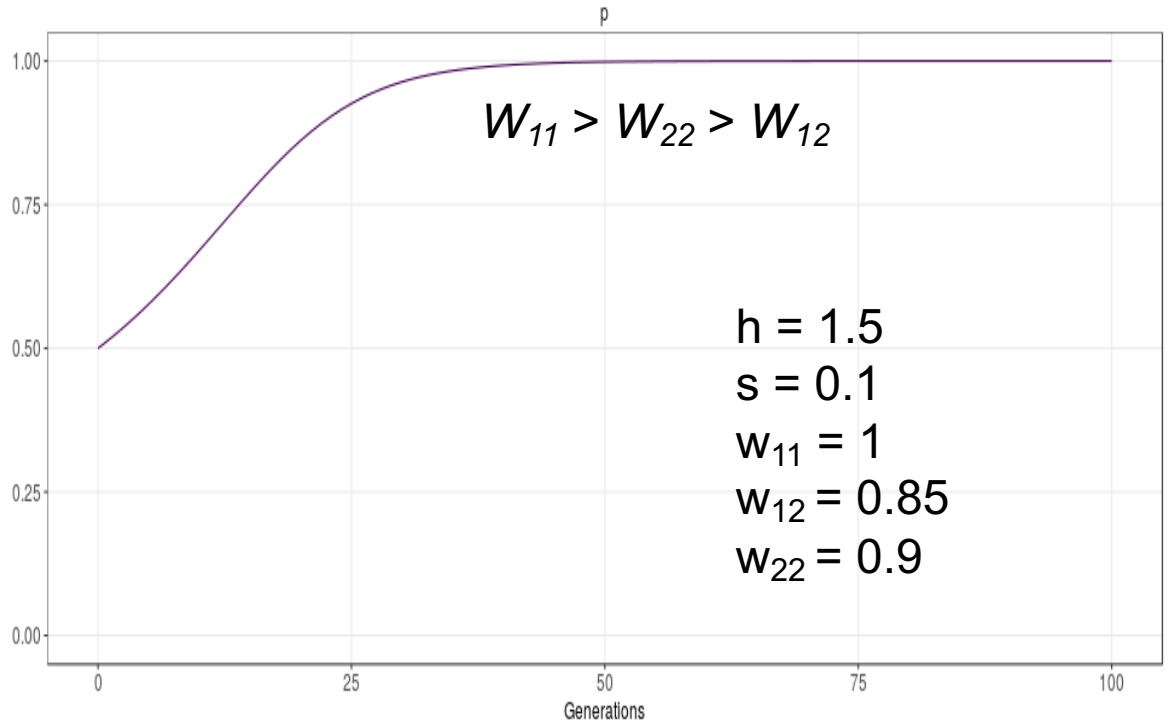




- A rare allele spreads if its heterozygote is fitter than the homozygote
- If the heterozygote has the highest fitness, then A_1 evolves toward an intermediate equilibrium
- When the heterozygote has the highest fitness, the allele with the highest homozygote fitness is most common allele at equilibrium

Underdominance: when the heterozygote has the lowest fitness





If the heterozygote has the lowest fitness then there are two stable equilibria, one where A_1 is fixed and the other where A_1 is lost

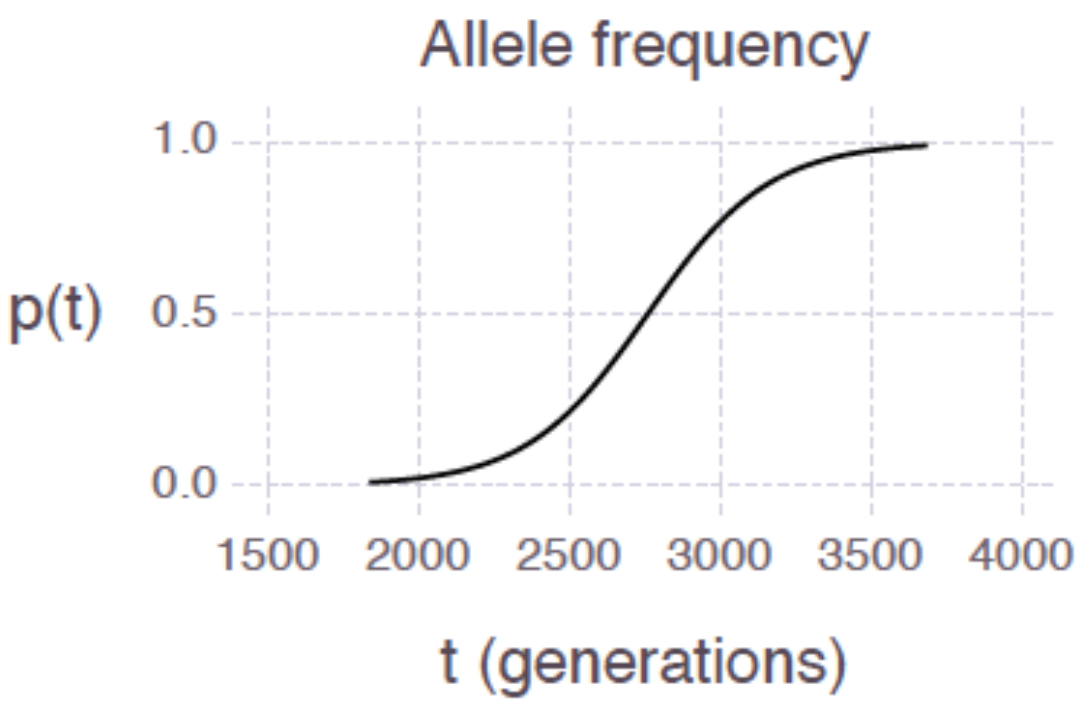
Rate of change in frequency of A_1

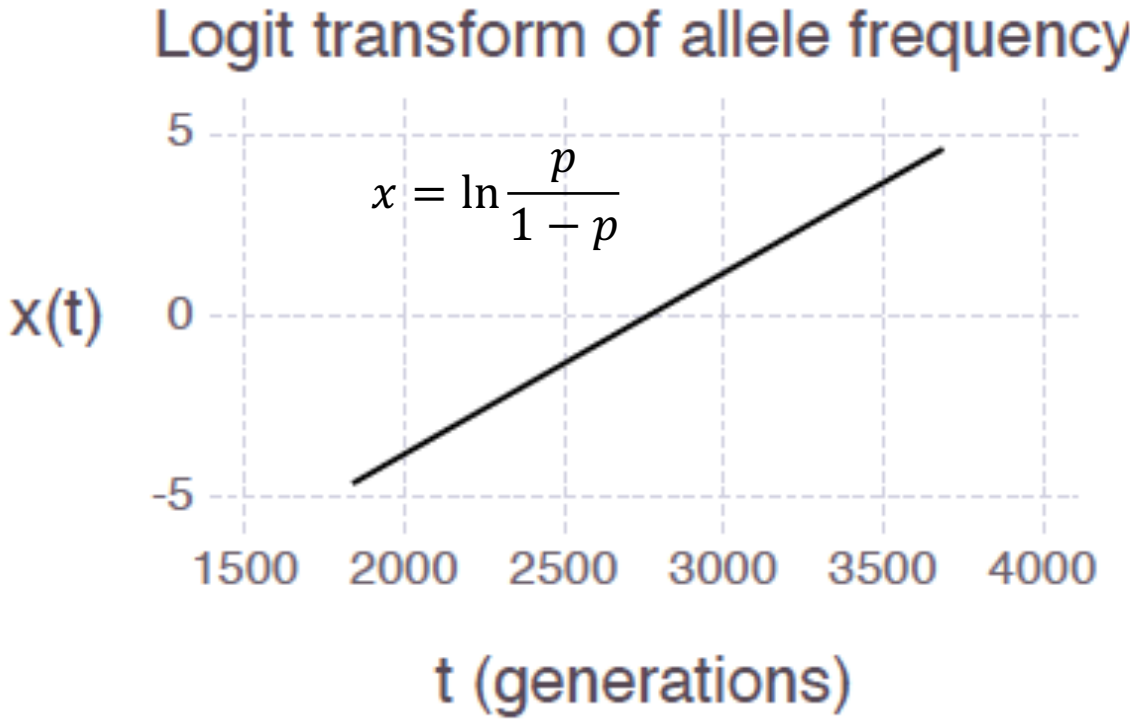
In terms of genotypic fitness: $\Delta p = \frac{pq}{\bar{W}} [p(W_{11} - W_{12}) + q(W_{12} - W_{22})]$

In terms of genic fitness: $\Delta p = \frac{pq(W_1 - W_2)}{\bar{W}}$

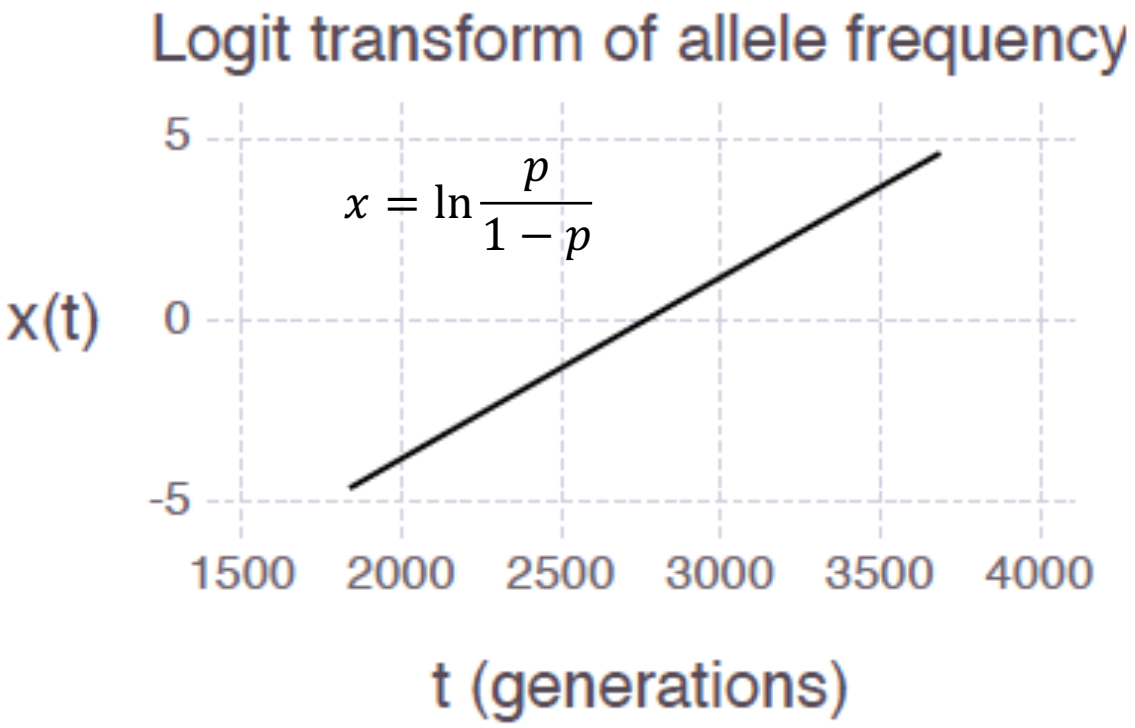
- Higher heterozygosity enables faster spread
- A rare allele spreads if its heterozygote is fitter than the homozygote
- If the heterozygote has the highest fitness, then A_1 evolves toward an intermediate equilibrium
- When the heterozygote has the highest fitness, the allele with the highest homozygote fitness is most common allele at equilibrium

Time required for a beneficial allele to sweep to fixation





Time required for a beneficial allele to sweep to fixation under an additive model



Slope: $x(t) = x_0 + \frac{ts}{2}$

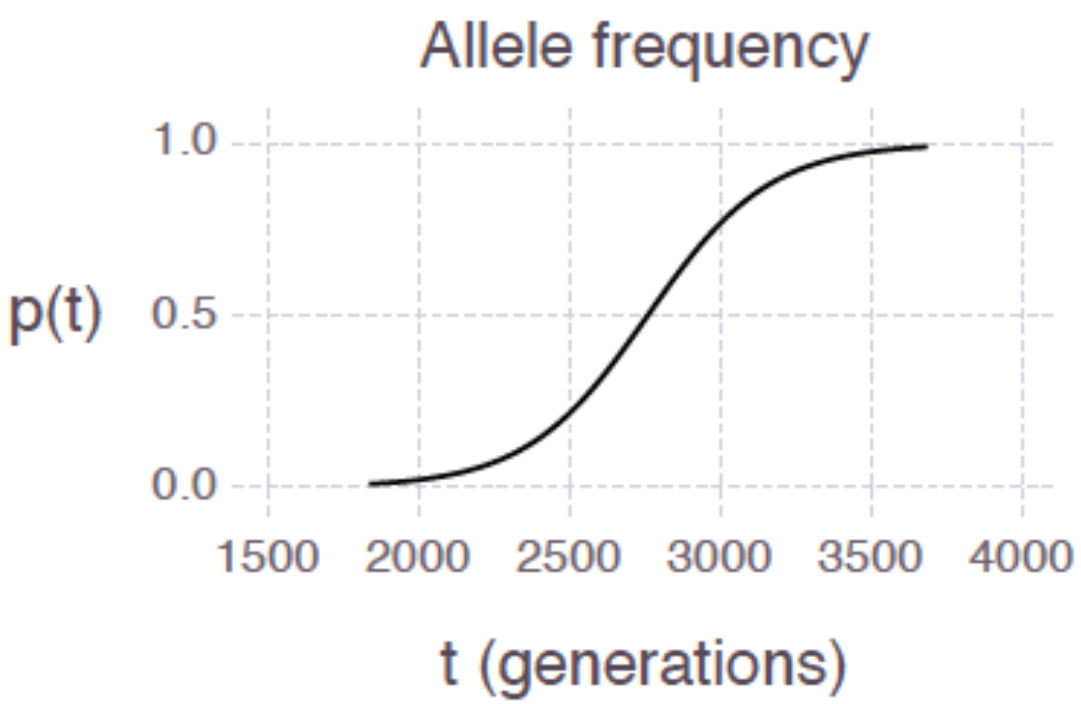
Solve for *t*: $t = \frac{2}{s}(x(t) - x_0)$

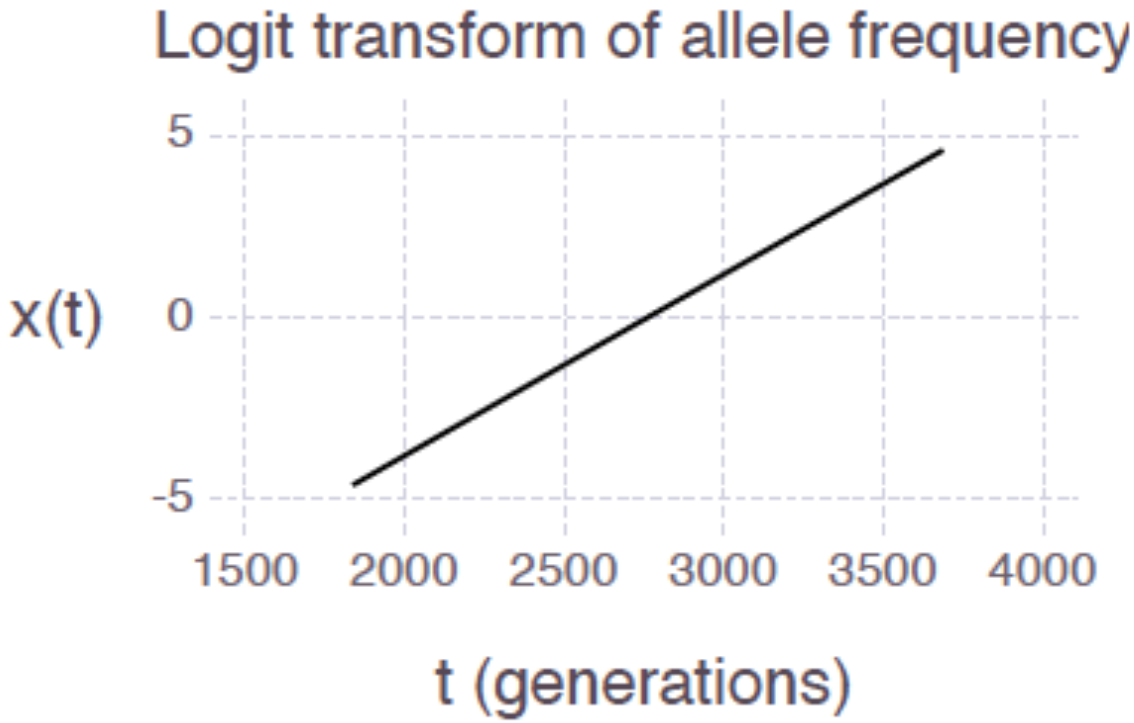
$= \frac{2}{s} \ln \frac{p_t q_0}{q_t p_0}$

How long for an advantageous allele to increase from 0.01 to 0.99?

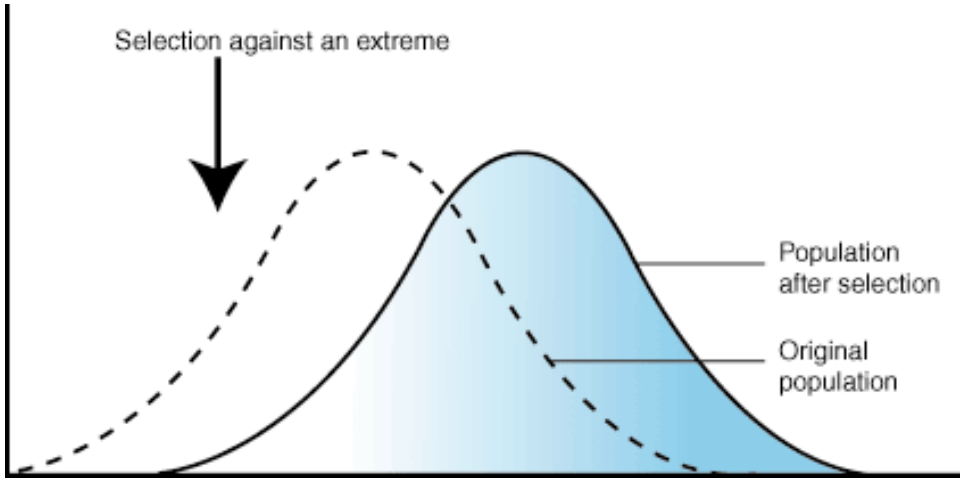
$$t = \frac{2}{s} \ln \frac{p_t q_0}{q_t p_0} = \frac{2}{s} \ln \frac{0.99 * 0.99}{0.01 * 0.01} \approx \frac{18}{s}$$

Time required for a beneficial allele to sweep to fixation





In the absence of dominance, it takes roughly 18/s generations for an advantageous allele to increase from 1% to 99% frequency



Some examples: Directional selection in nature

Industrial revolution and melanism in British peppered moths

Industrial melanism in moths represents rapid evolution in response to environmental change

Ford 1975

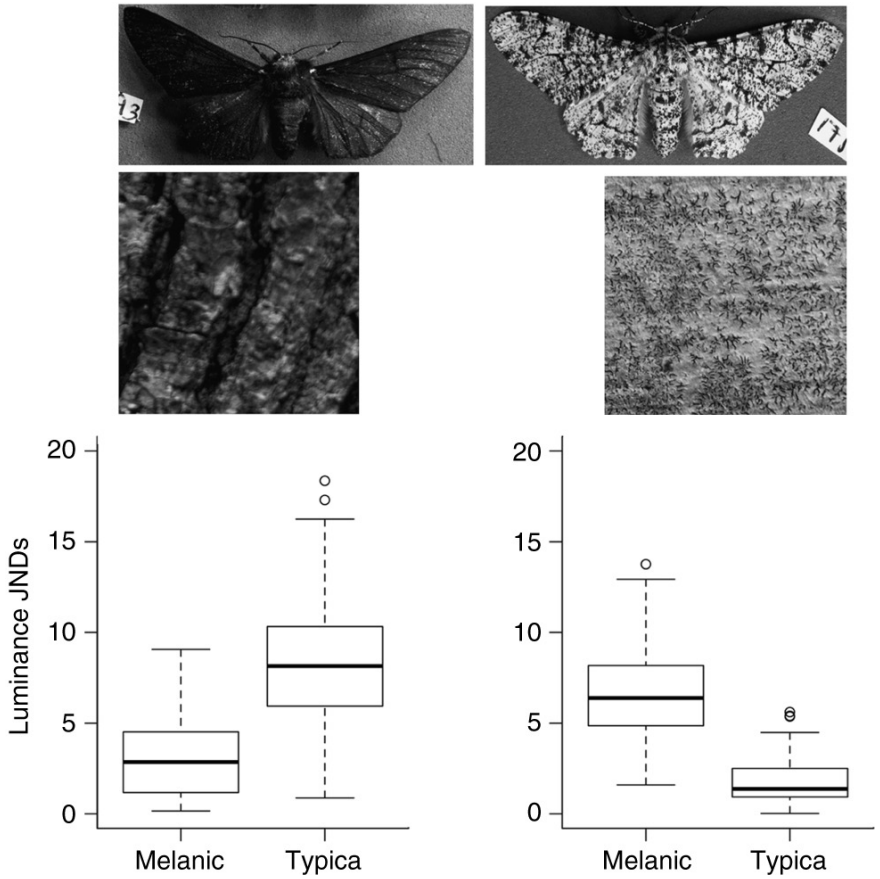
Industrial melanism: A classic example of selection

- The ancestral form of the peppered moth (*typica*) is white with dark speckles but, after the Industrial Revolution, a darker form (*carbonaria*) became more common due to natural selection in sooty environments
- The *carbonaria* form displaced the light-colored moths in the polluted woodlands of Europe, reportedly due to selective predation by birds on the lighter moths when located on sooty trees (Cook et al. 2012, Luiggi 2012)
- As pollution levels decreased, the *carbonaria* form appears to be less common due to selection for light-colored moths

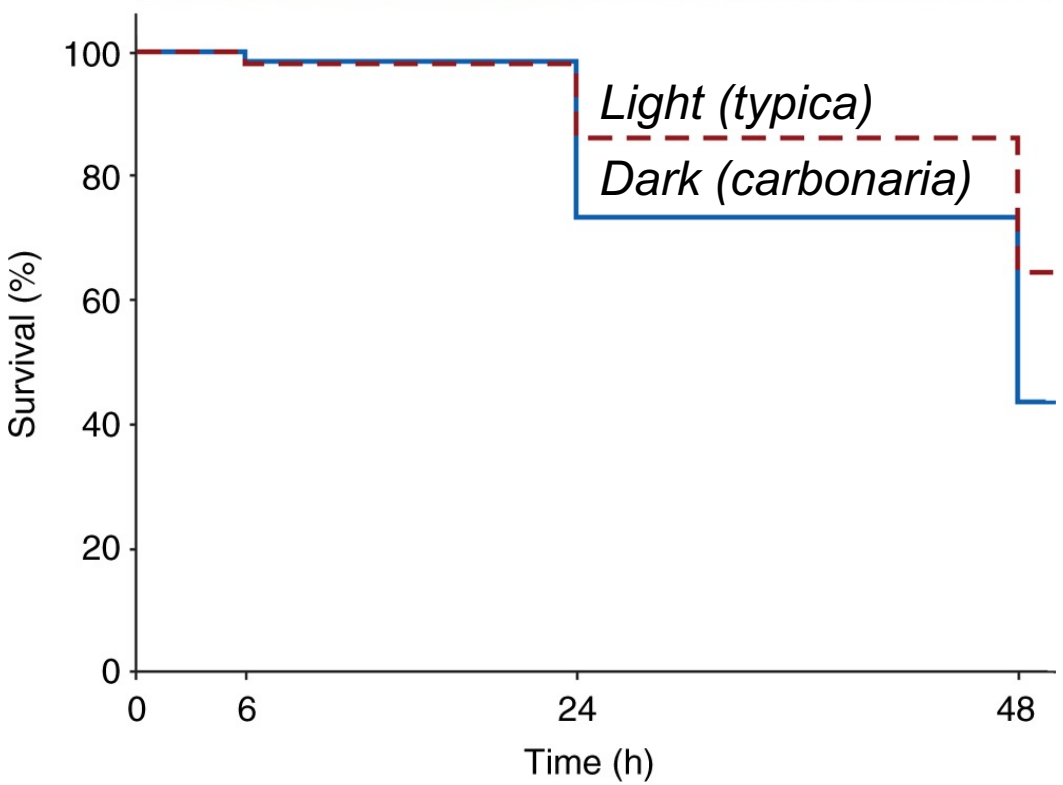
From Hoy et al., Insect Molecular Genetics

Evidence that current selection against dark peppered moth morphs are due to avian predation

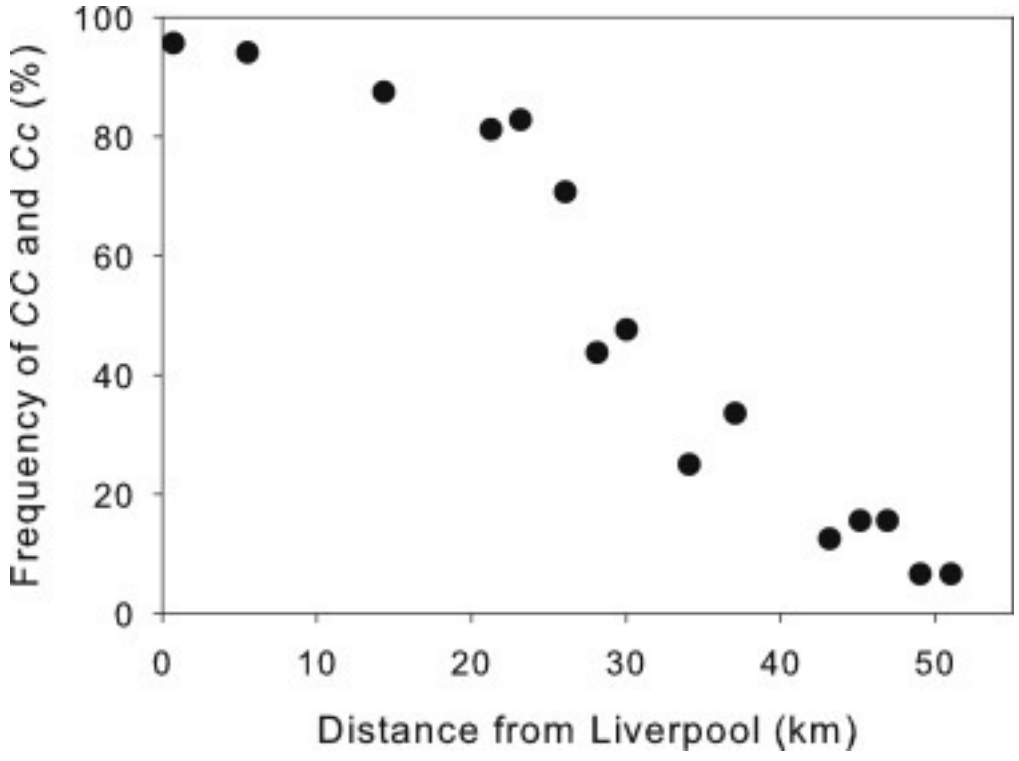
Visibility of moths based on an avian model



"Survival" of artificial moth targets with predation



A genetic cline in allele frequency associated with wing pigmentation



- It was hypothesized that the pigmentation would decrease with distance from polluted cities
- A SNP in the cortex gene is associated with wing pigmentation and distance from Liverpool, England
- More recently, it was discovered that the causal variant for pigmentation variation is a transposable element insertion in the *Cortex* gene

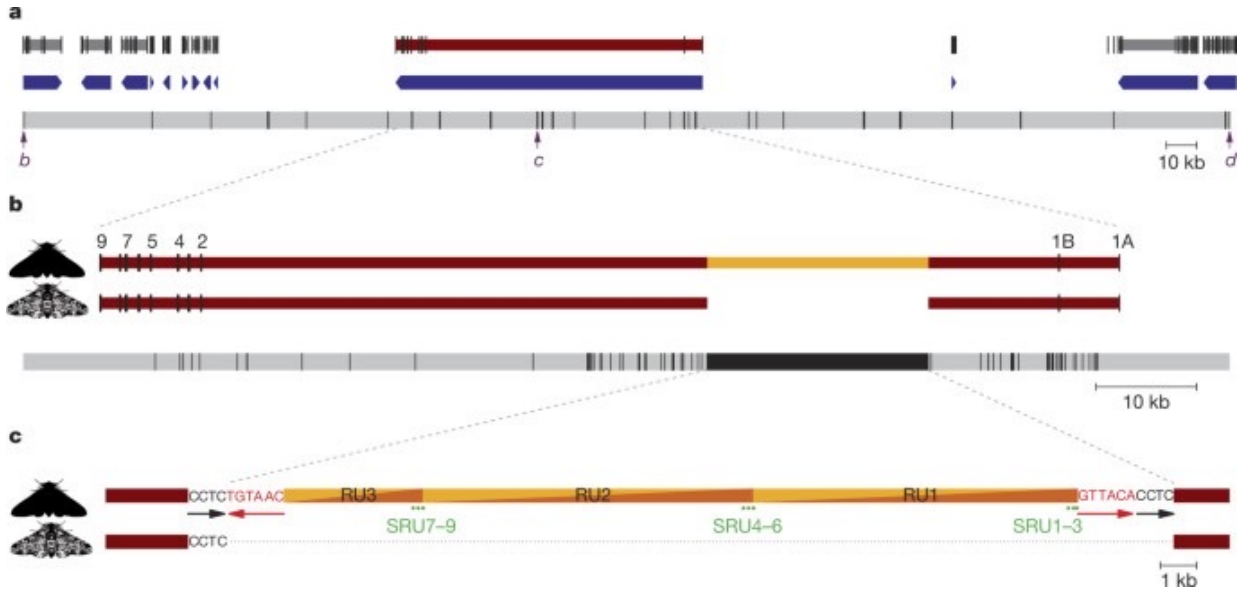
The genetic basis for the melanism variant

The black form of the moth was known to be due to a single-locus dominant allele, but the biochemical basis of this phenotype remained unknown until van't Hof et al. (2011, 2016) mapped the melanism to a 200-kb region of a chromosome and further narrowed it to a transposable element insertion into the *cortex* gene

Large-scale candidate region

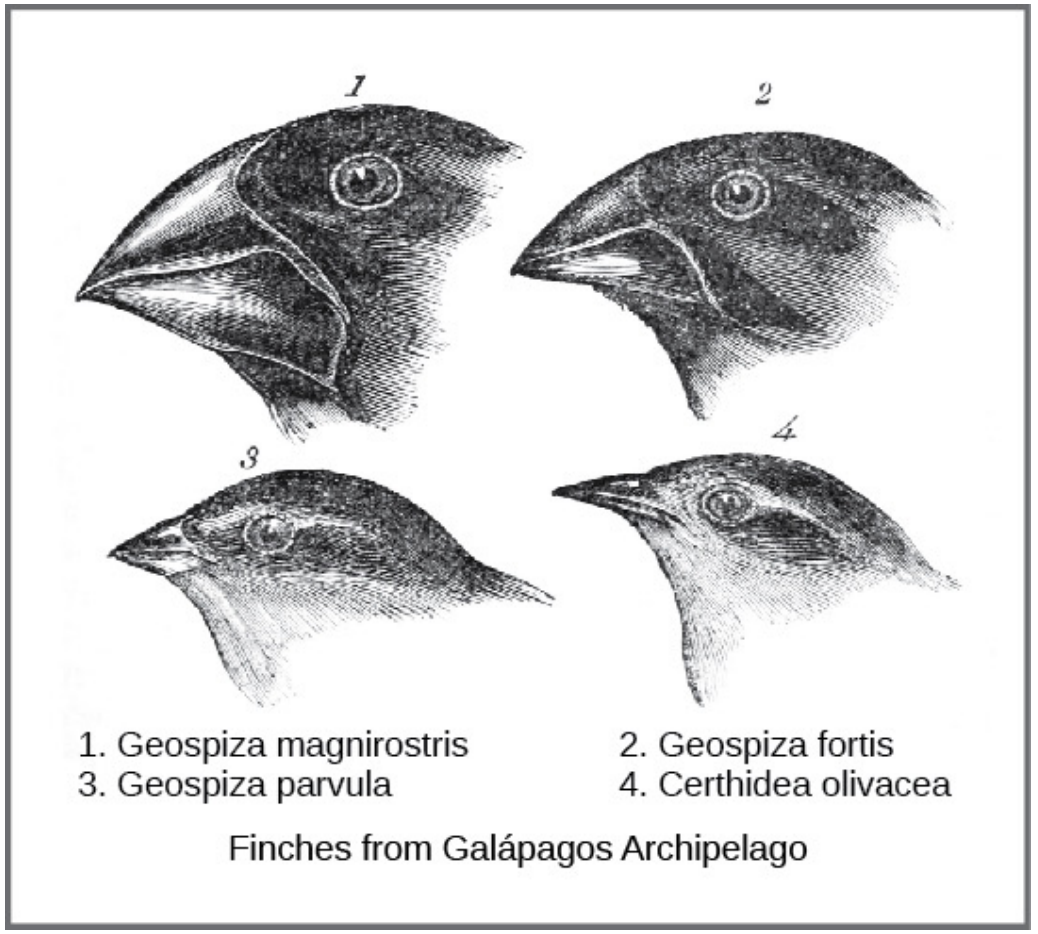
Candidate variants

TE insertion, which they confirm has a phenotypic effect

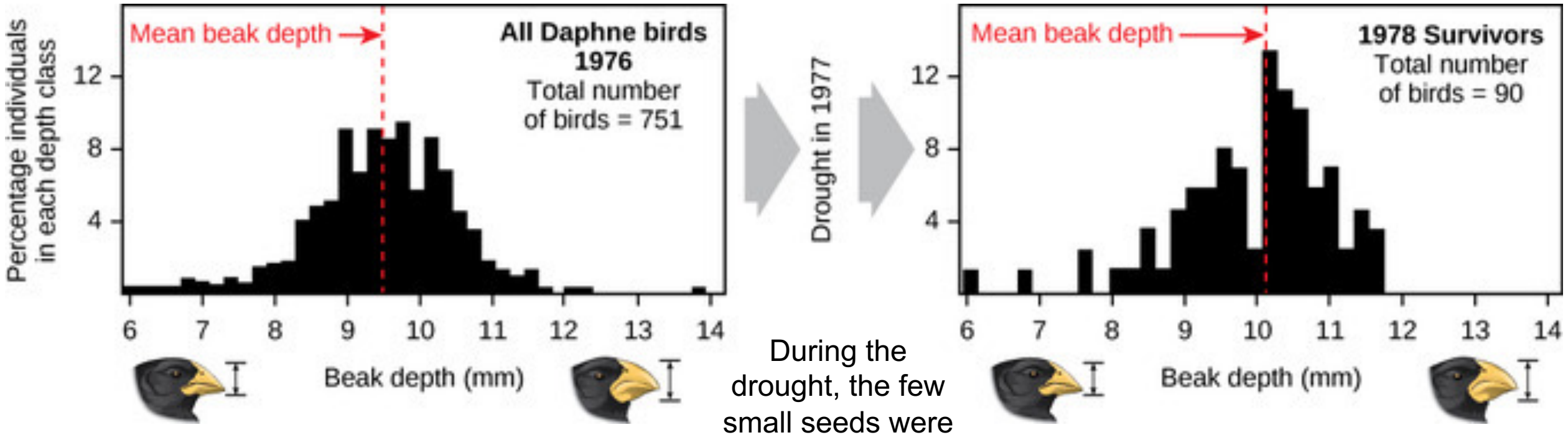


Van't Hof et al., Nature, 2016

Beak morphology in Darwin's finches: a classic example of adaptive radiation



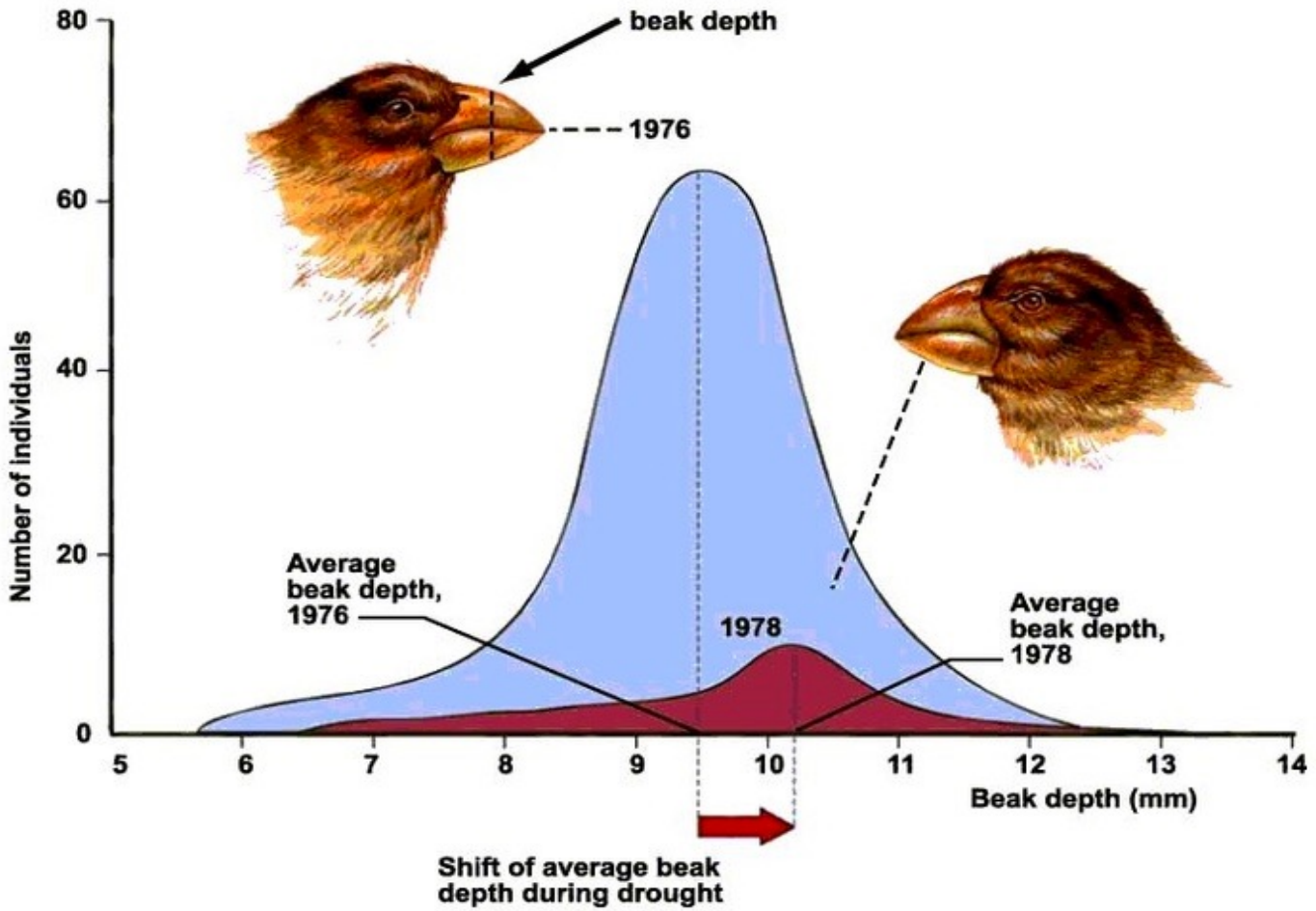
Beak morphology is a complex trait that shifts with changing environment



During the drought, the few small seeds were eaten, only large seeds left. Birds with bigger beaks could eat those

bio.libretexts.org

Beak morphology is a complex trait that shifts with changing environment



Adaptation to dietary shift: lactase persistence in Europeans

Simoons hypothesized that the distribution of pastoralism could explain the striking differences in lactase persistence among populations

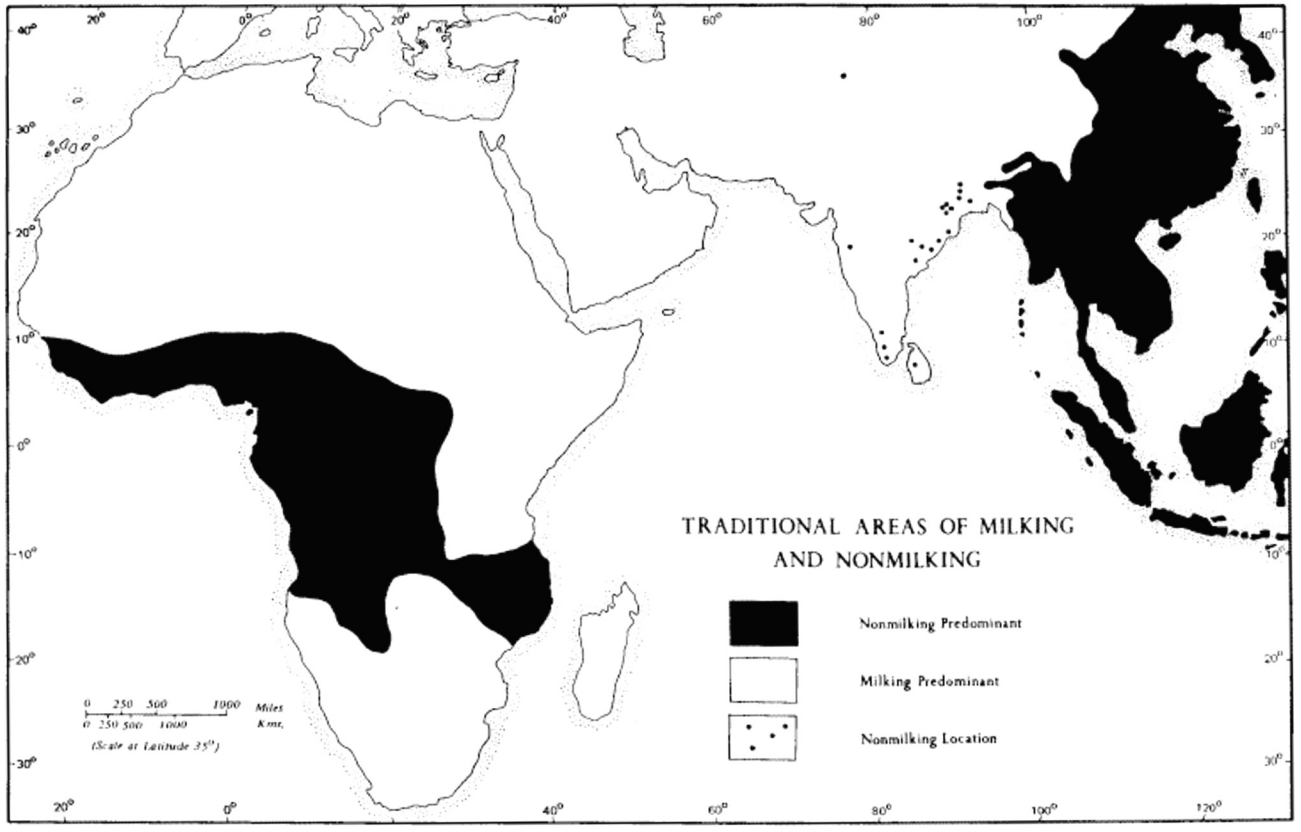
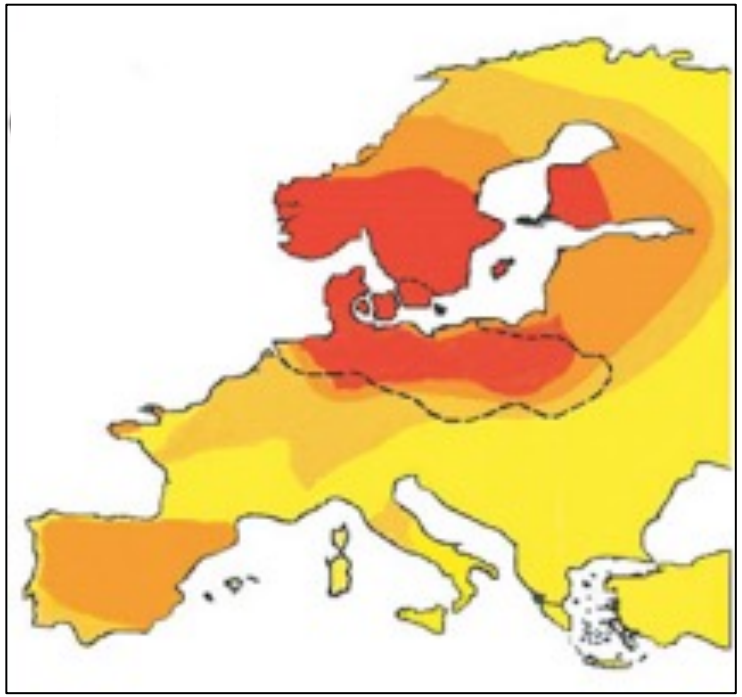


Fig 1. Traditional areas of milking and nonmilking.

Simoons, 1970

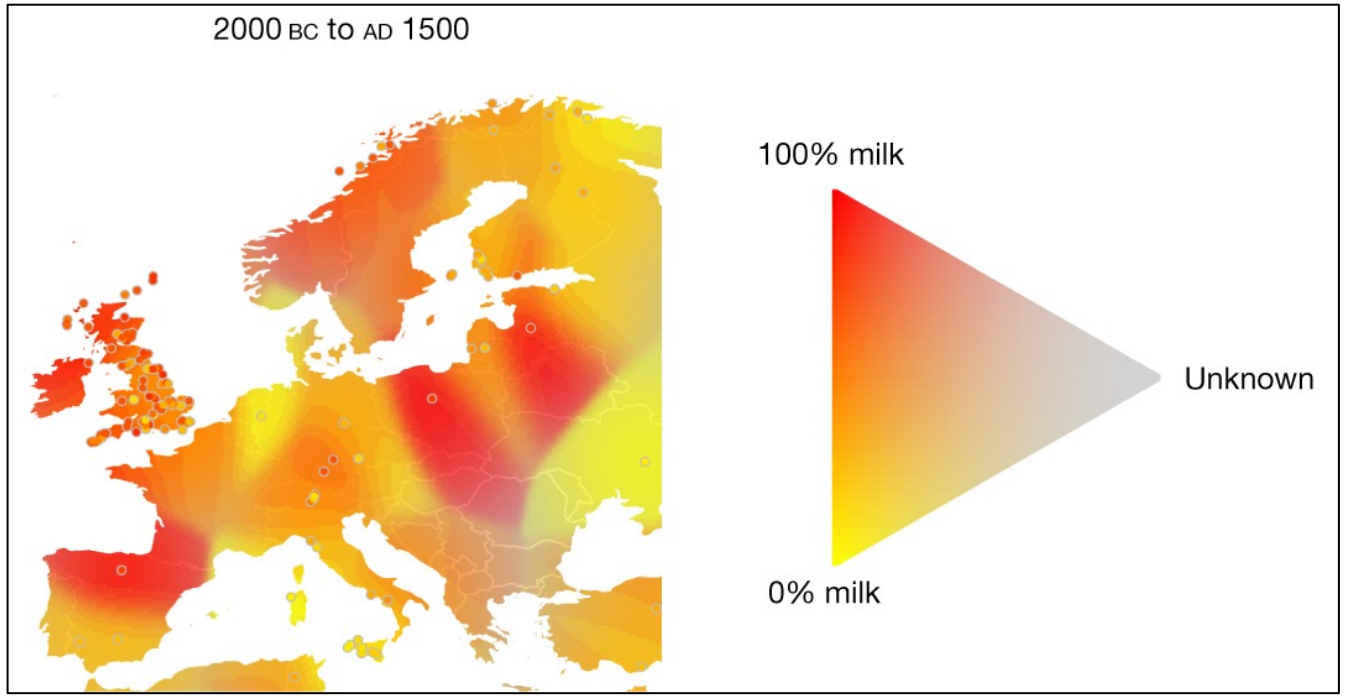
Lactase persistence in Europeans

Frequency of lactase persistence in Europe



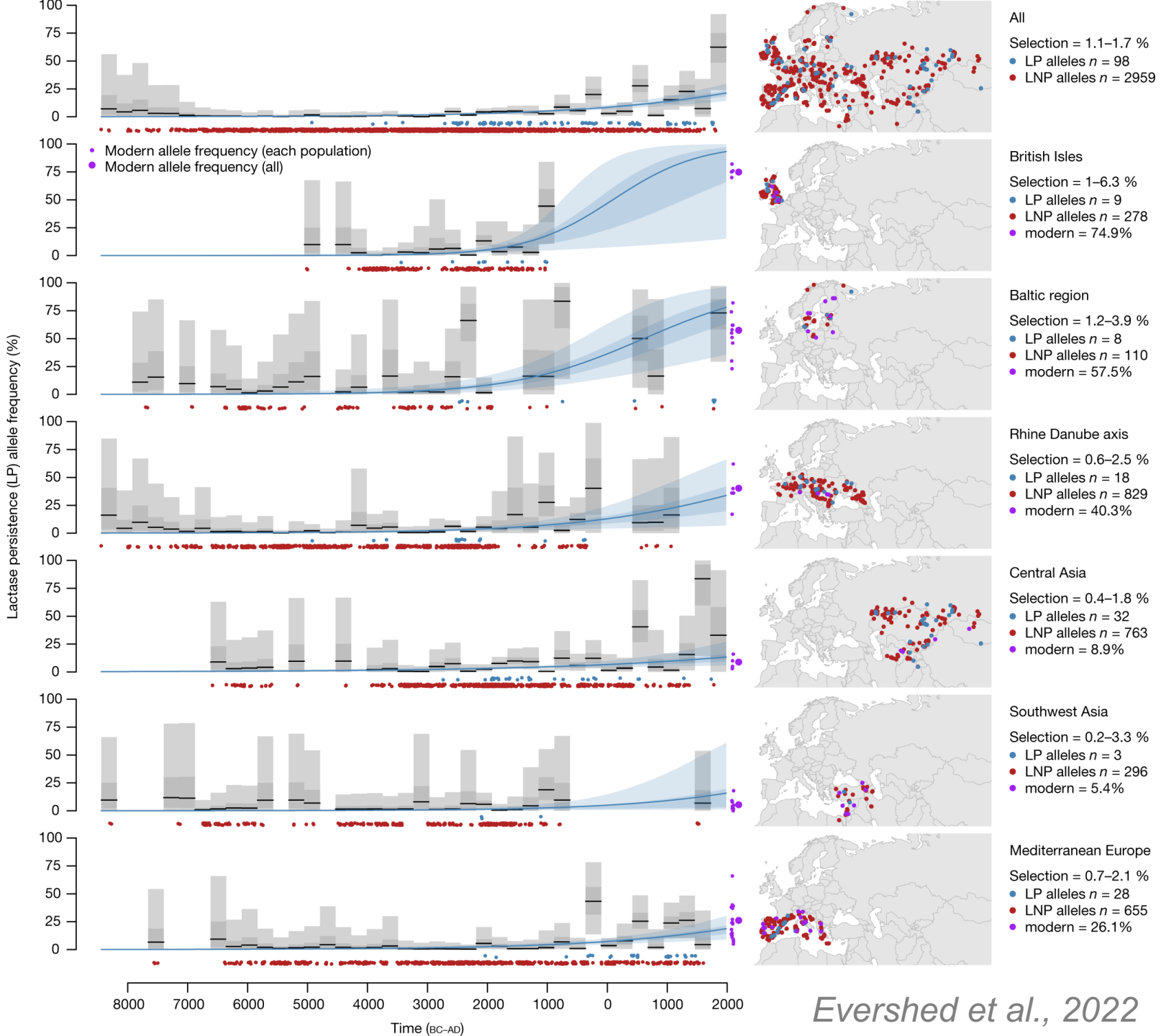
Beja-Pereira et al., 2003

Dairy fat residues at archeaological sites



Evershed et al., 2022

Inferred LCT allele frequency increase over time across populations



Evershed et al., 2022

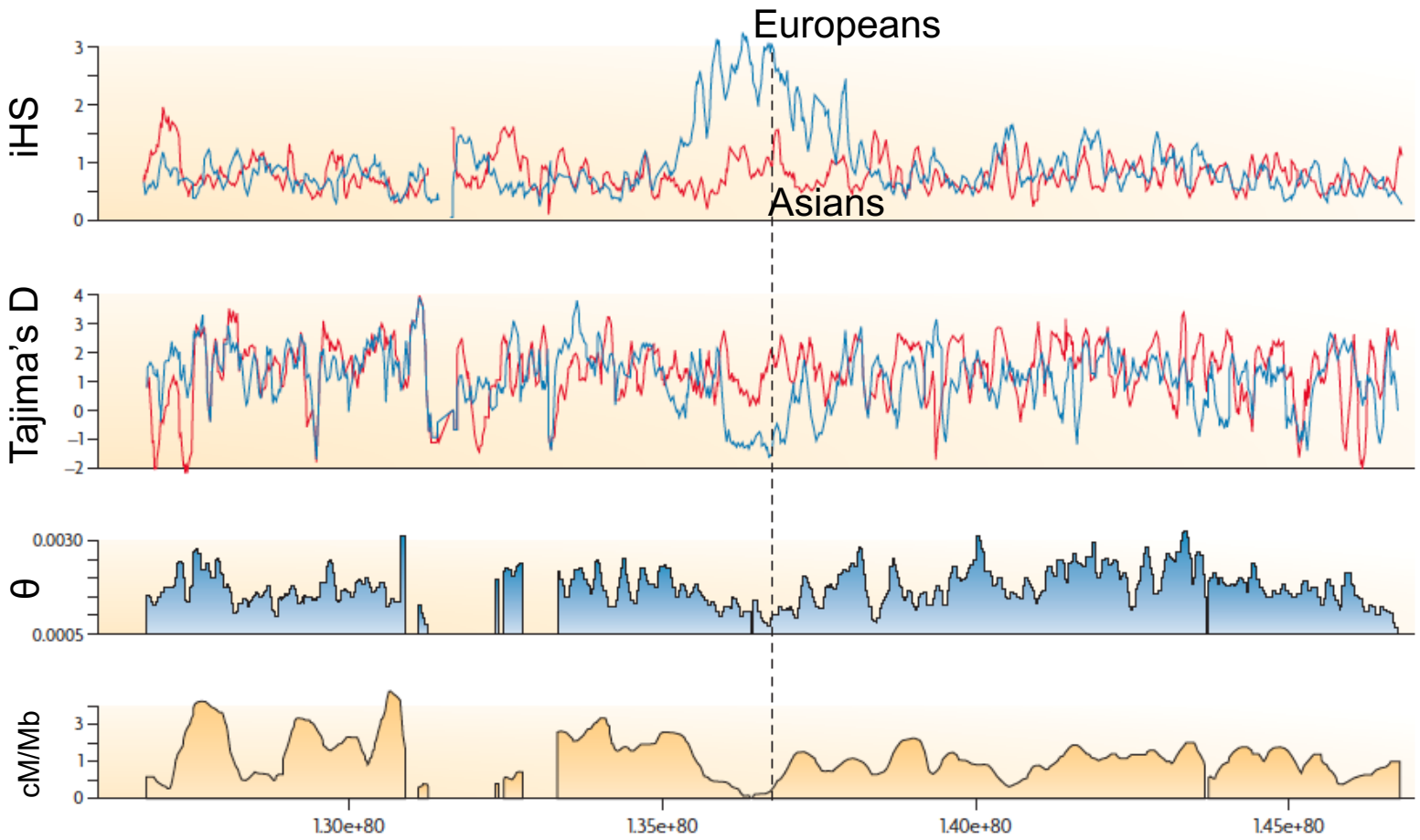
Adaptation to dietary shift: selection signatures at the lactase gene region in humans

Haplotype homozygosity

Scaled difference between θ_π and θ_W

diversity

recombination



Nielsen et al., 2007

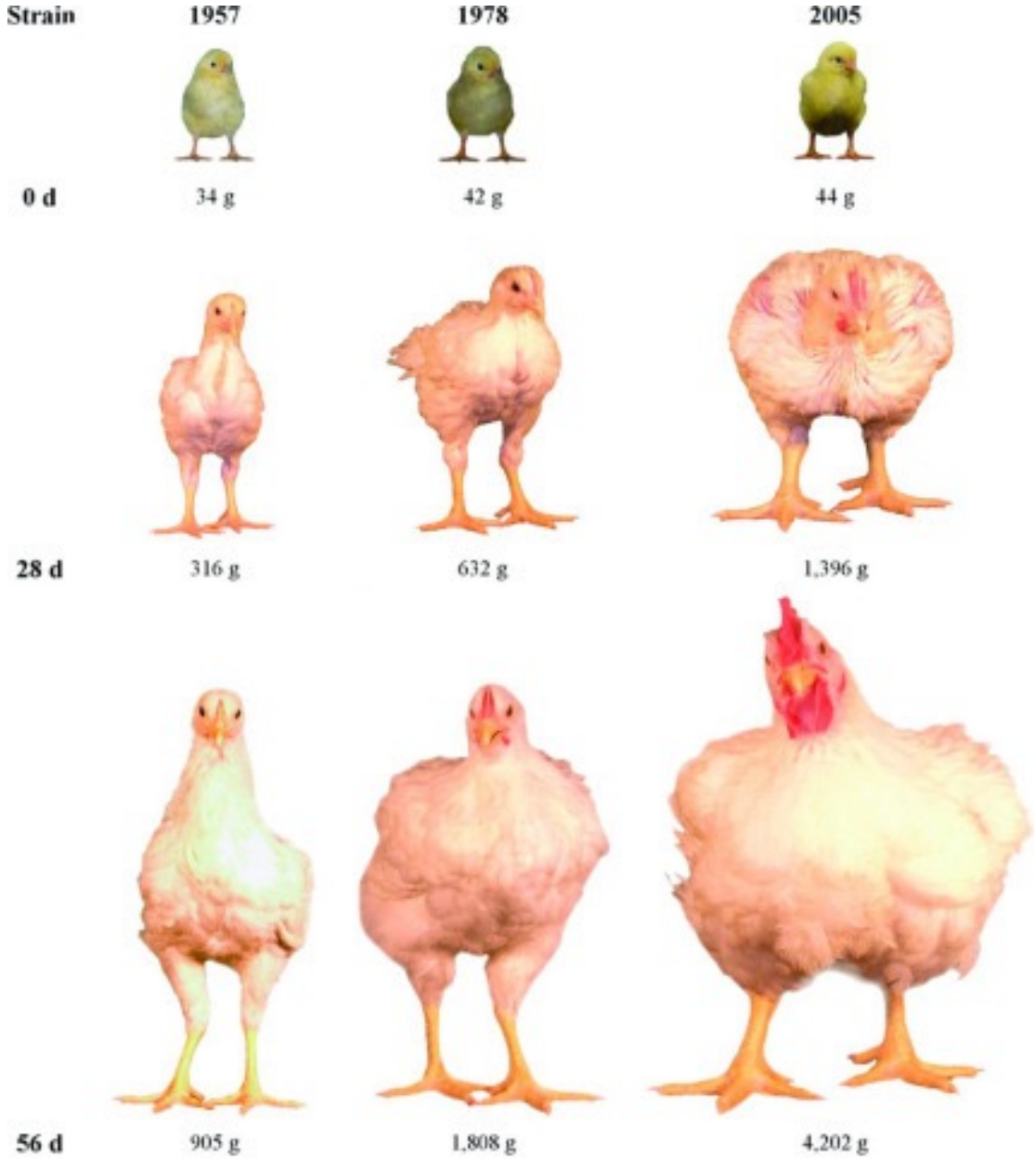
Artificial selection during domestication

Artificial selection is selection by humans, in contrast to selection in the wild

Nearly all chicken we eat comes from the “Cobb 500”

But there is a lot of variation among chicken breeds

Artificial selection on three strains of chicken shown increasing size over time

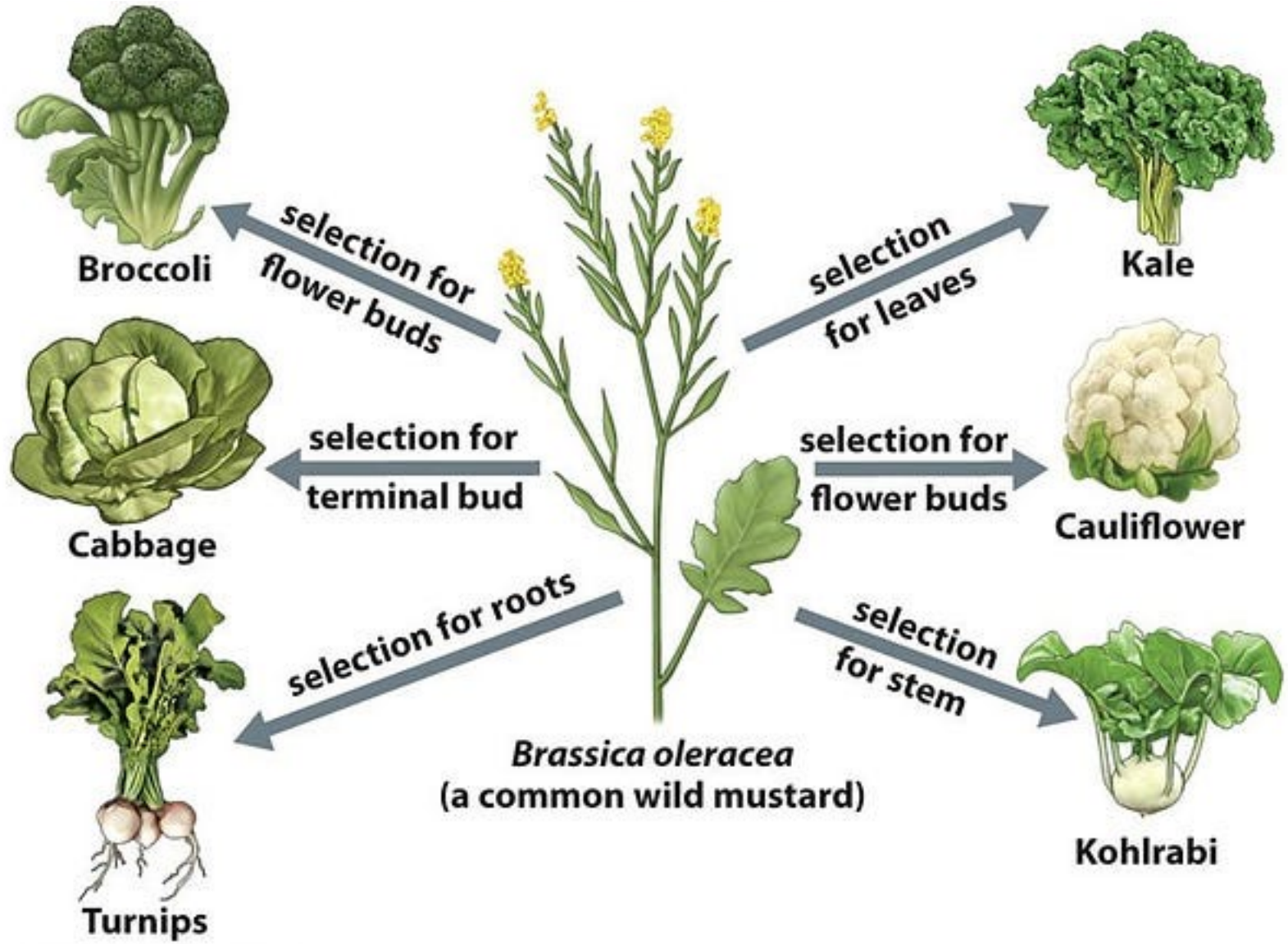


Zuidhof et al., 2014; https://www.sciencedirect.com/science/article/pii/S0032579119385505#fig1

What do these plants have in common?

All these plants are varieties of cabbage: kohlrabi, cauliflower, Romanesco cabbage, Brussels sprouts, Savoy sprouts, red head cabbage and broccoli. Photo: Coyau, Wikimedia Commons

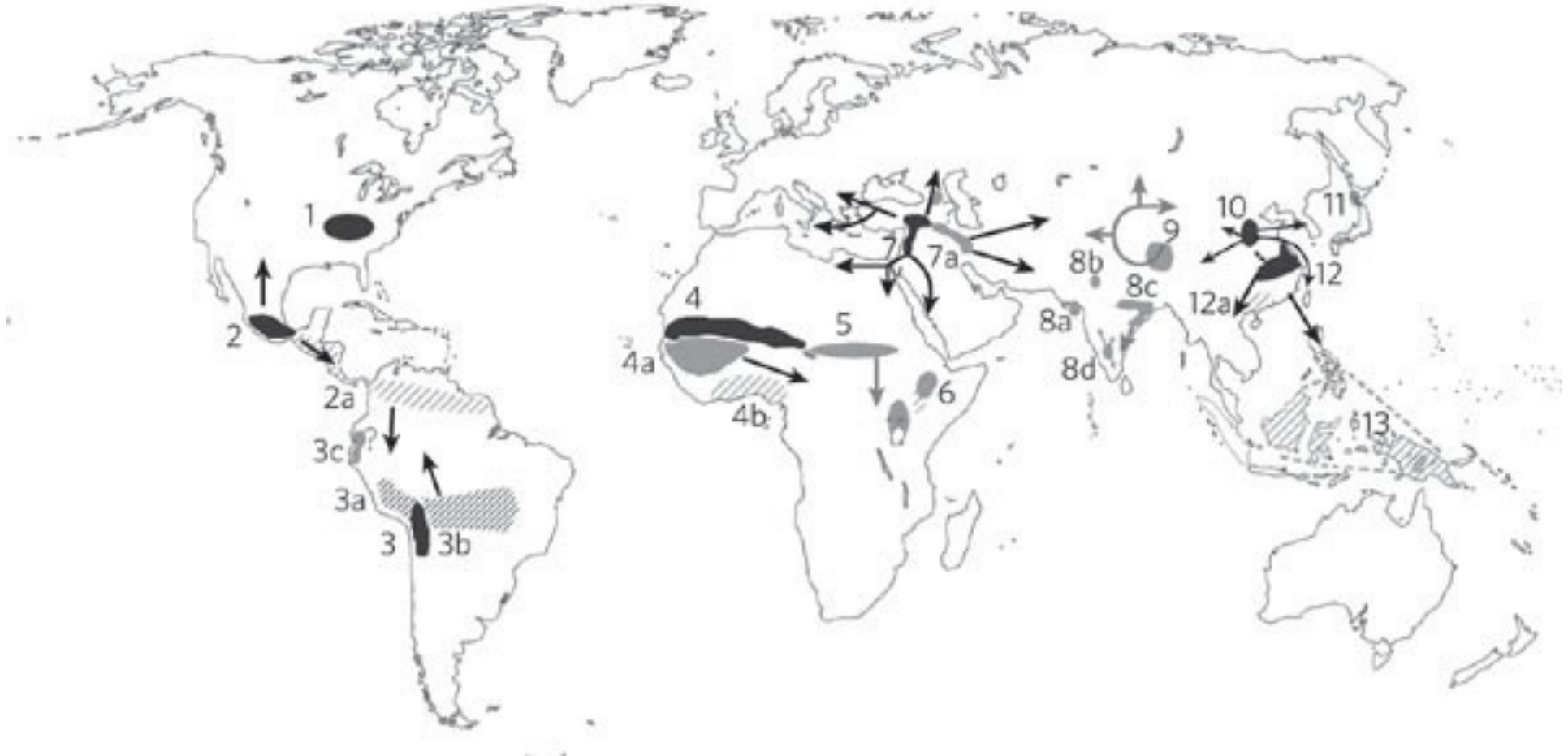
Diverse crops are derived from wild cabbage (Brassica oleracea)



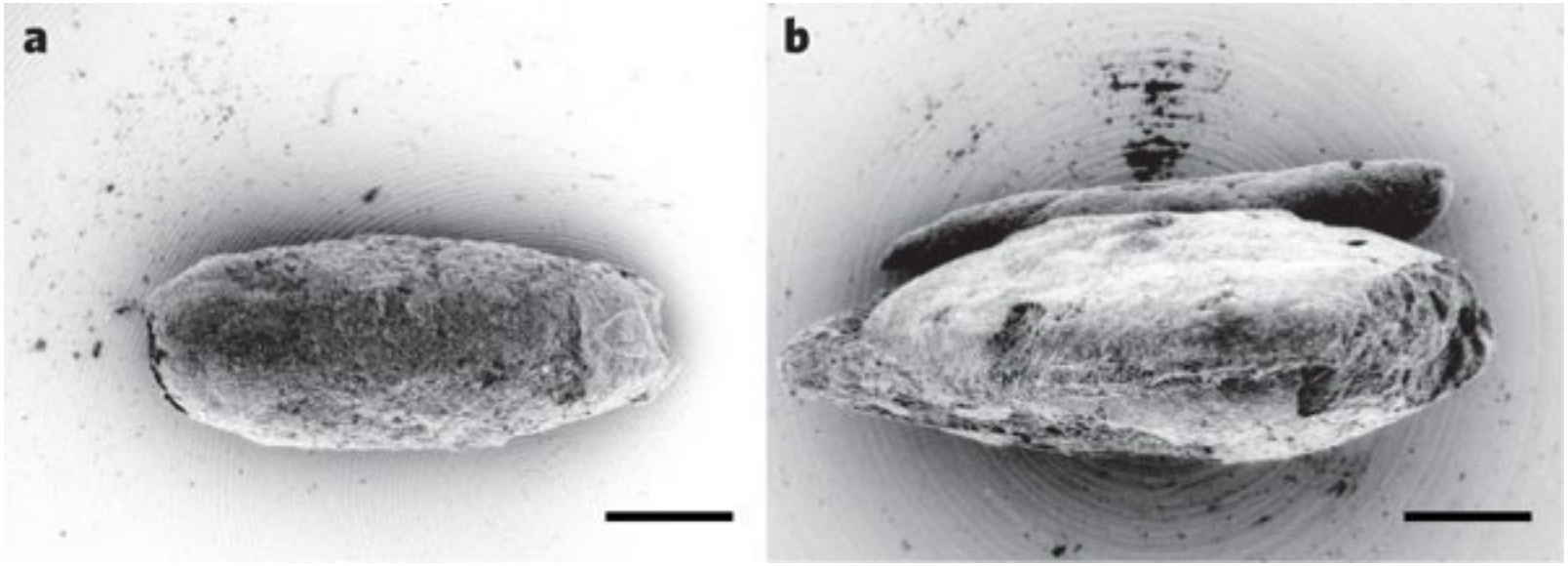
Discover Biology, 5/e Figure 1.10

Centers of plant domestication

Plants were domesticated independently in different regions worldwide, centered around the equator

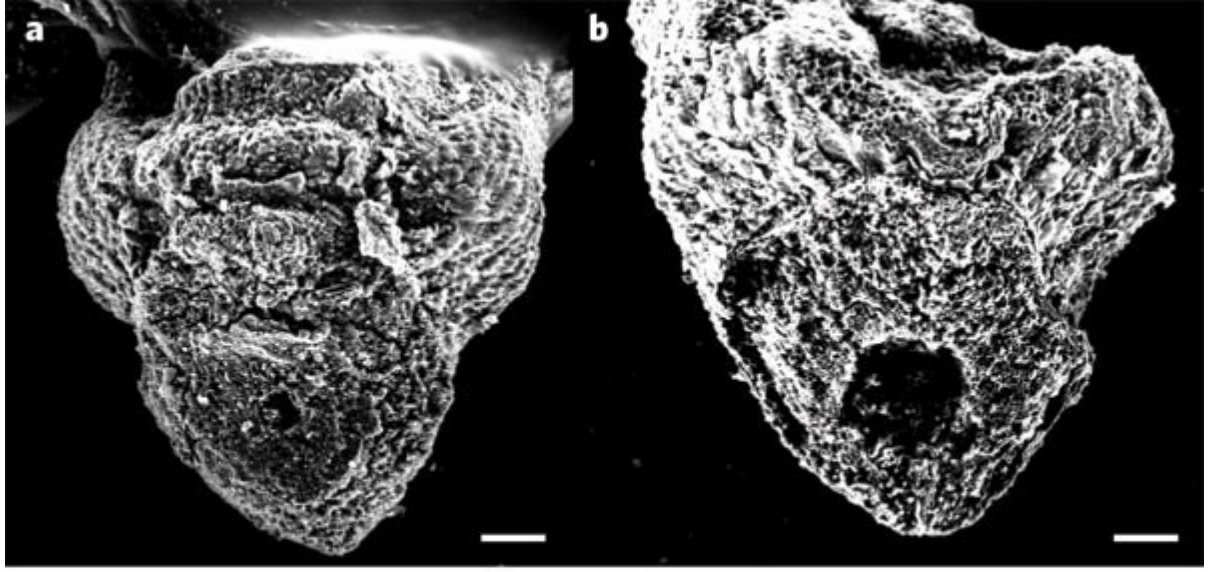


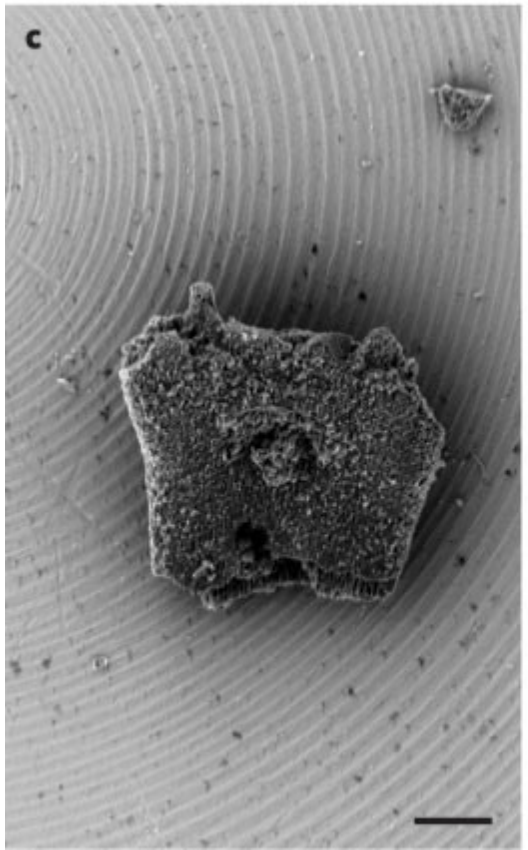
Evidence of increasing grain size from the archaeological record

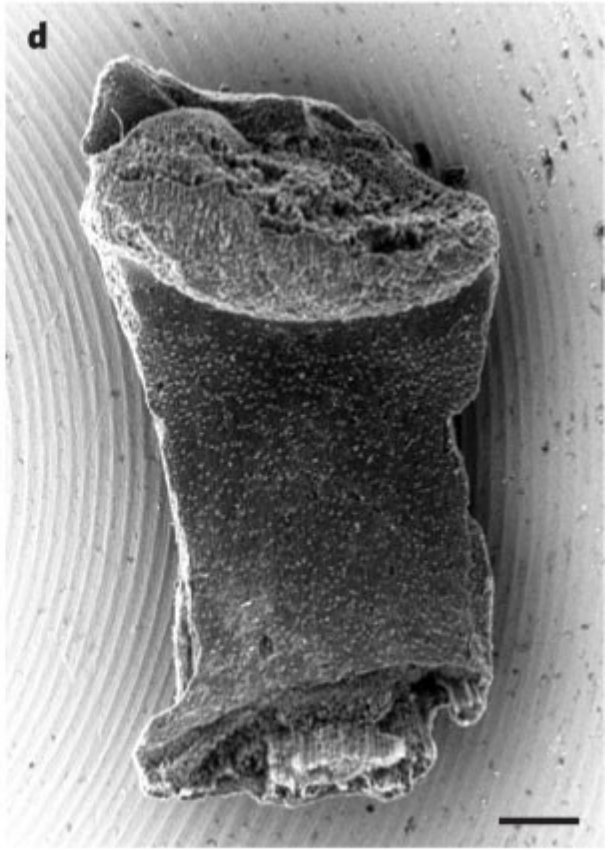


Purugganan and Fuller, 2009

Evolution of non-shattering seeds from the archaeological record

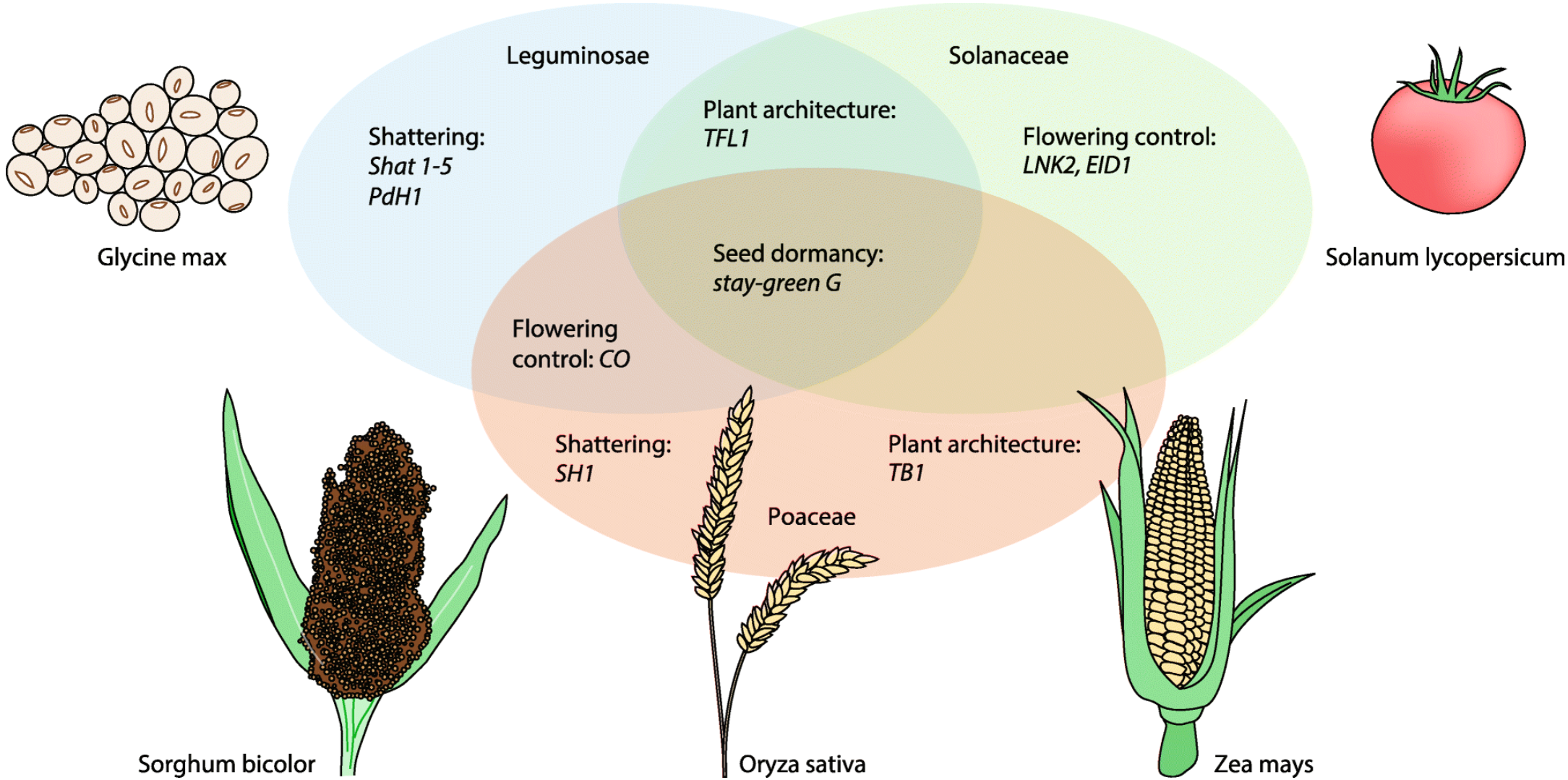




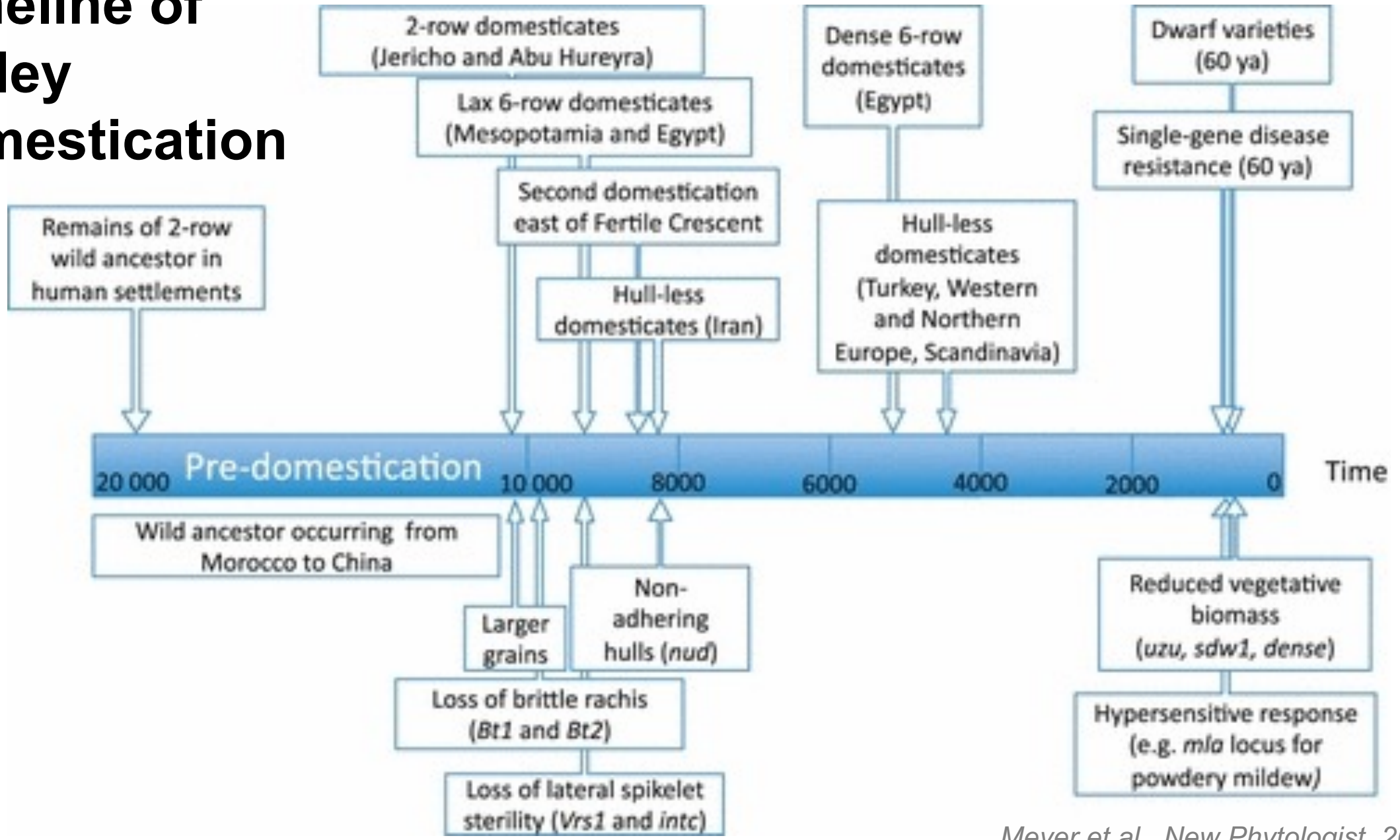


Purugganan and Fuller, 2009

Traits and genes involved in plant domestication

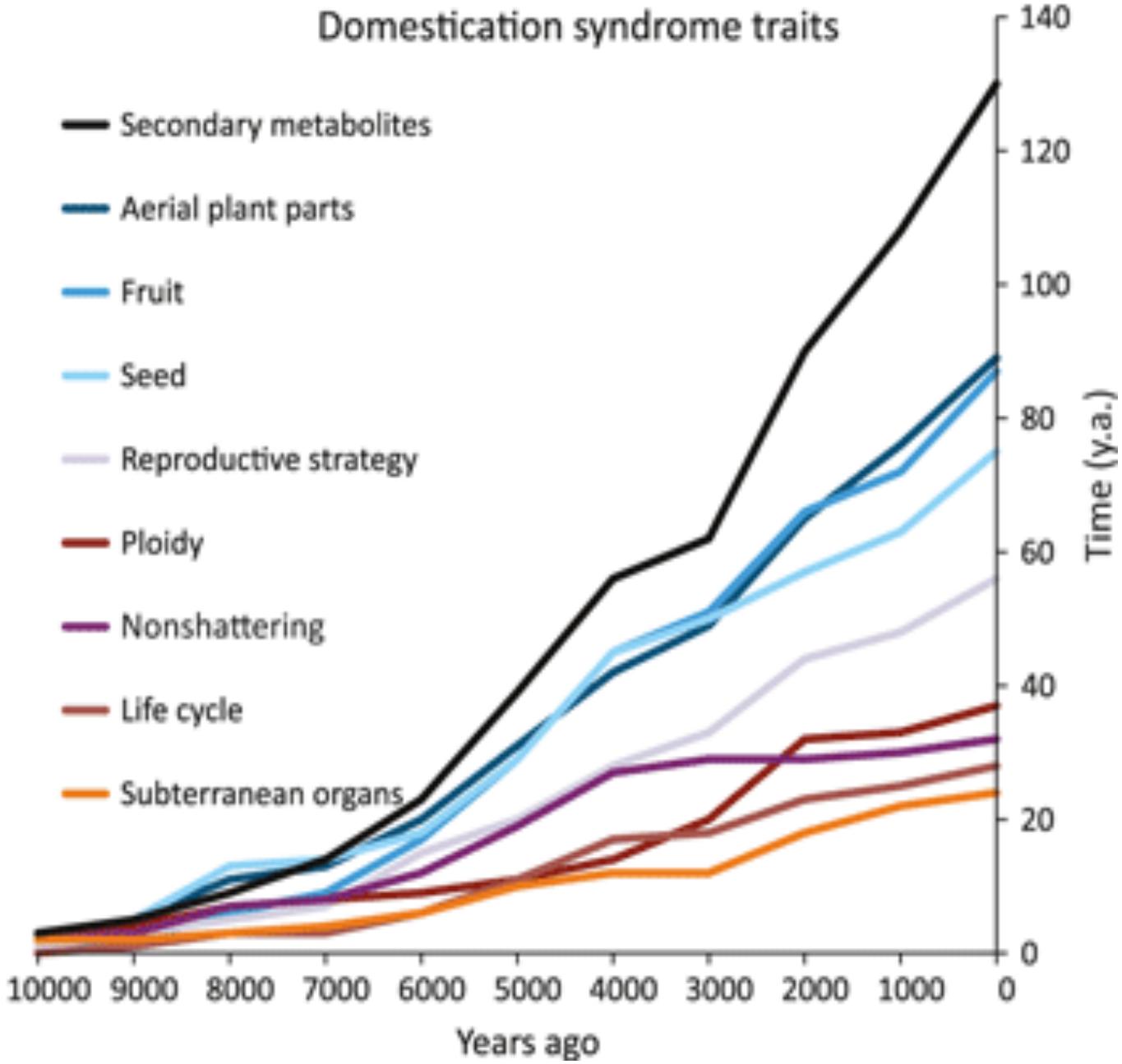


Timeline of barley domestication



Meyer et al., New Phytologist, 2012

Artificial selection on many traits over thousands of years led to current domesticated plant varieties



Meyer et al., New Phytologist, 2012

Example: Experimental evolution in microbial populations

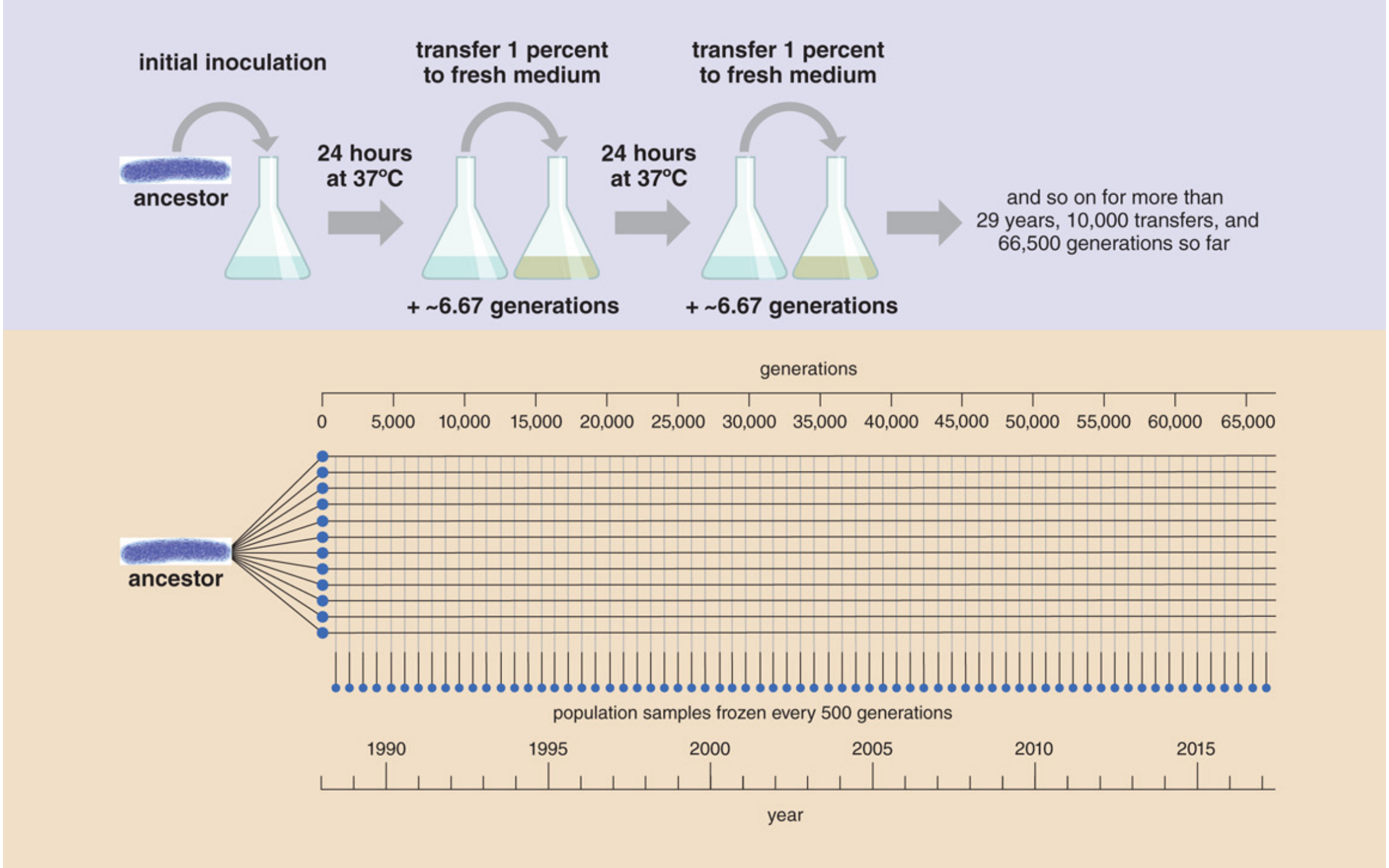
Experimental evolution in *E. coli* - LTEE

Since microbial populations can have rapid generation times, we can observe evolutionary dynamics over many generations

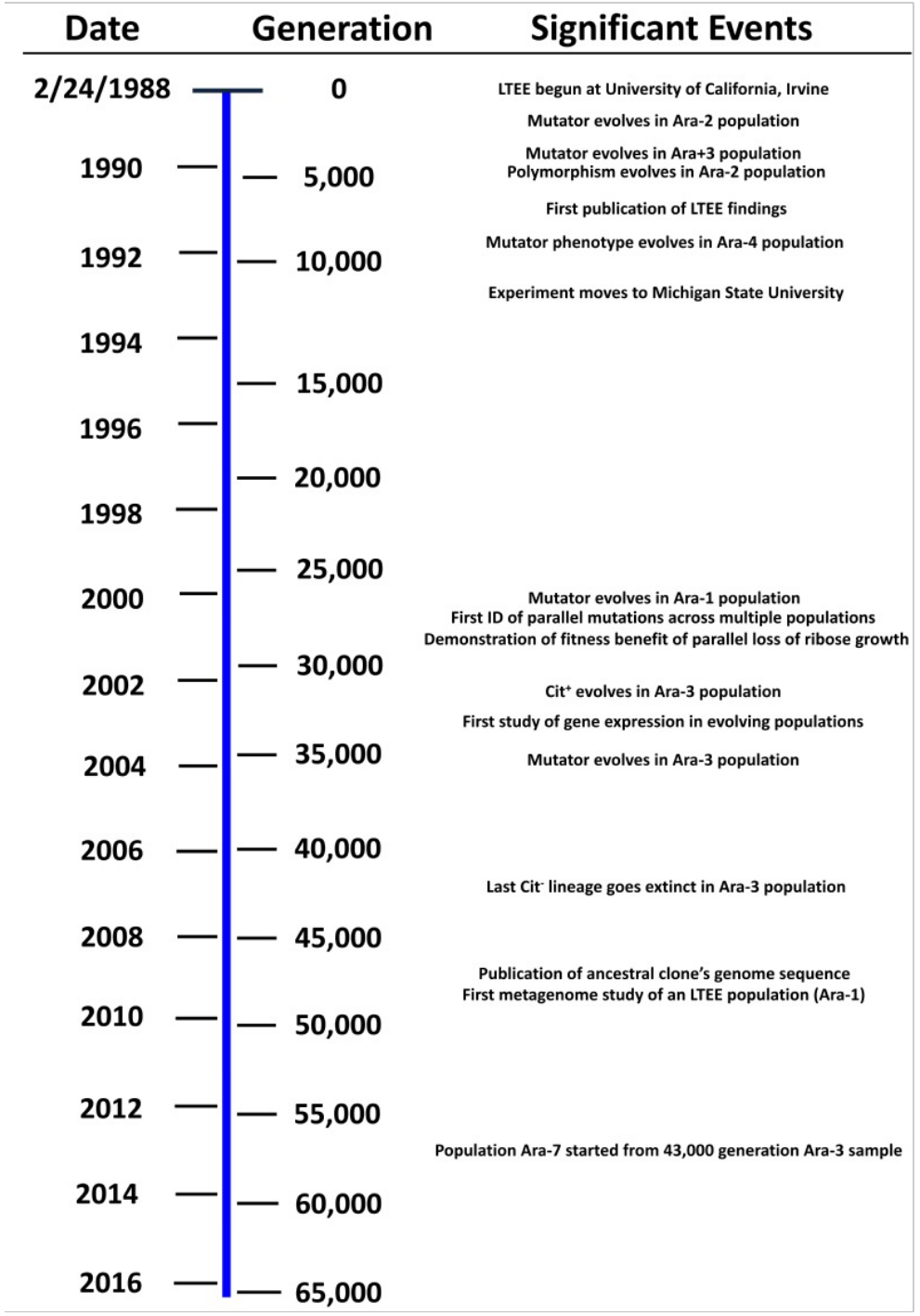
- Experiment has run since 1988
- Daily transfers of 12 replicate E. coli populations with approx. 2000 generations/year
- Ara+ and Ara- strains were grown under glucose limitation on arabinose sugar and strains were competed
- Differences in relative fitness could be due to differences in lag phase, growth rate, and/or survival at stationary phase

"Evolutionary biologists say I'm asking the right questions but studying the wrong organism, and microbiologists tell me I'm studying the right organism but asking the wrong questions."

Fox and Lensky, PLOS Bio, 2015: 10.1371/journal.pbio.1002185


initial inoculation
transfer 1 percent to fresh medium
transfer 1 percent to fresh medium
ancestor
24 hours at 37°C
24 hours at 37°C
and so on for more than 29 years, 10,000 transfers, and 66,500 generations so far
+ ~6.67 generations
+ ~6.67 generations
generations
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
45,000
50,000
55,000
60,000
65,000
ancestor
population samples frozen every 500 generations
1990
1995
2000
2005
2010
2015
year

Timeline of the long term evolution experiment (LTEE)



https://en.wikipedia.org/wiki/E._coli_long-term_evolution_experiment#/media/File:LTEE_Timeline_as_of_May_28,_2016.png

Experimental evolution in *E. coli* - LTEE

"I was pretty sure we would see the rate of fitness improvement decelerate over time, and it has; and I was also pretty sure we'd see a quasi-step-like dynamic to the early fitness increases, and we did. Nonetheless, these analyses have yielded surprises as well, including evidence that fitness can increase indefinitely, and essentially without limit, even in a constant environment"

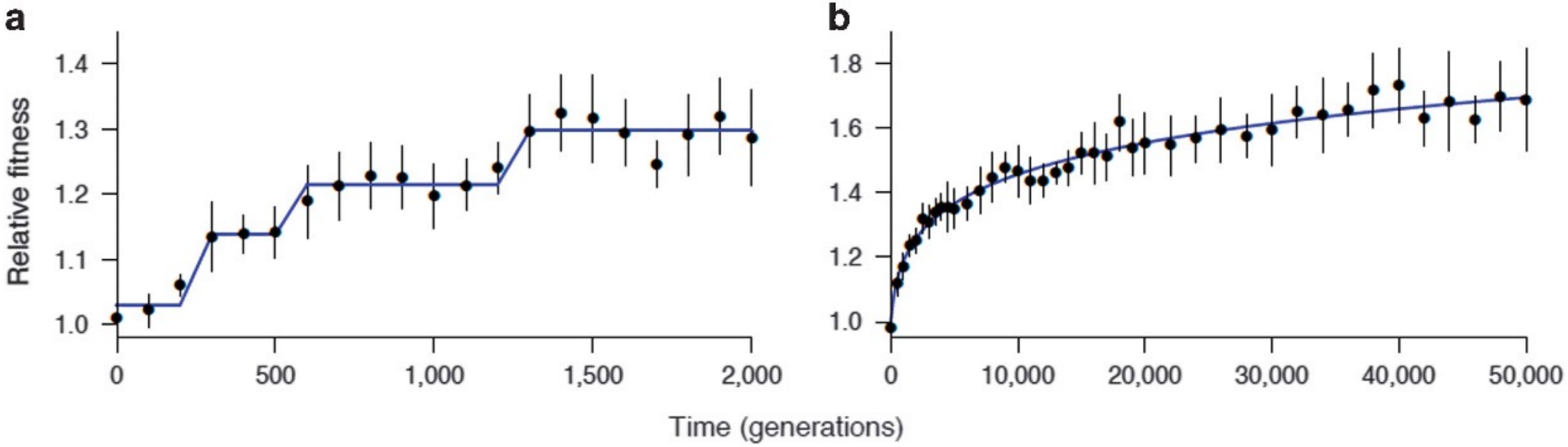
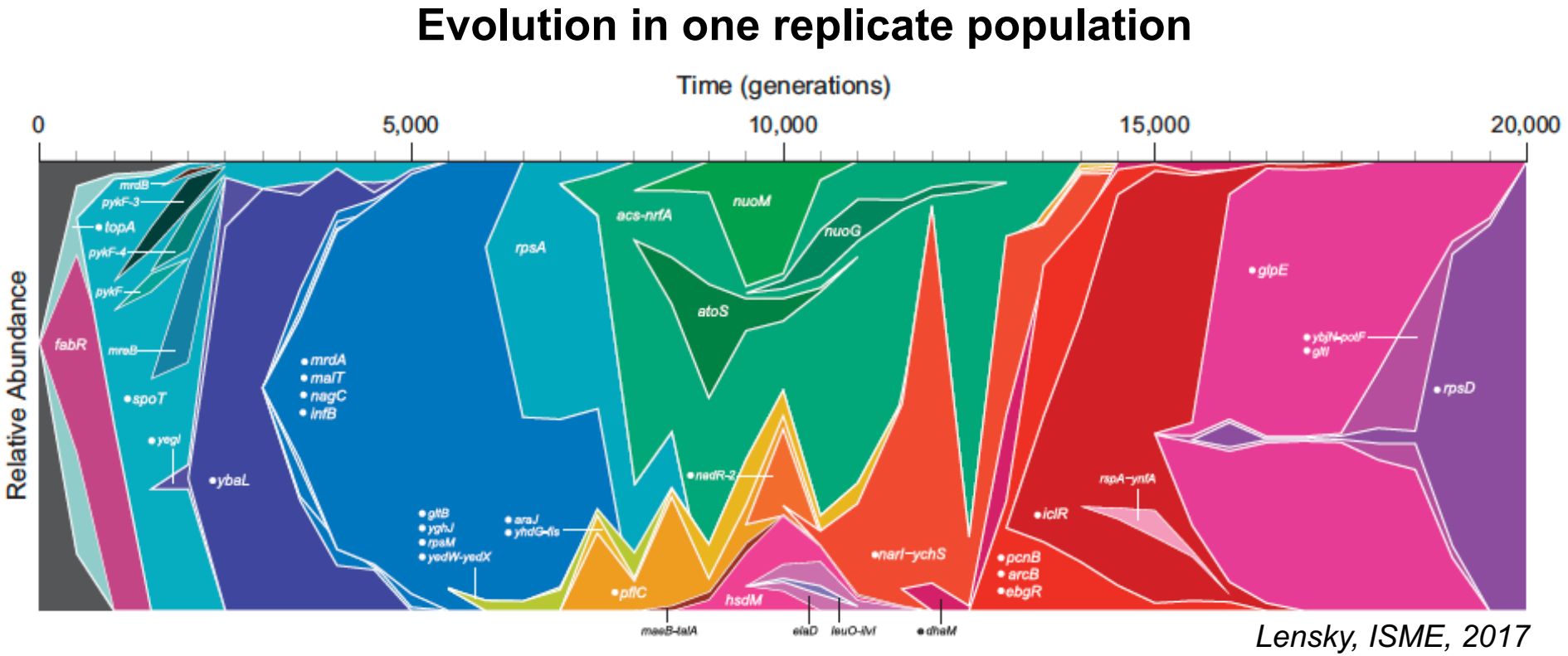


Figure 1 Fitness trajectories of evolving E. coli populations. (a) Fitness trajectory for one population, Ara–1, relative to its ancestor over the first 2000 generations of the LTEE. Error bars are 95% confidence intervals based on replicated assays. The line segments show the fit of a step model to the data. Modified from Lenski and Travisano (1994). (b) Trajectory for the grand-mean fitness across the LTEE populations over 50 000 generations. Error bars are 95% confidence limits based on replicate populations. The curve shows the fit of a power-law model. Modified from Wiser et al. (2013).

Lensky, ISME, 2017 and Fox and Lensky 2015

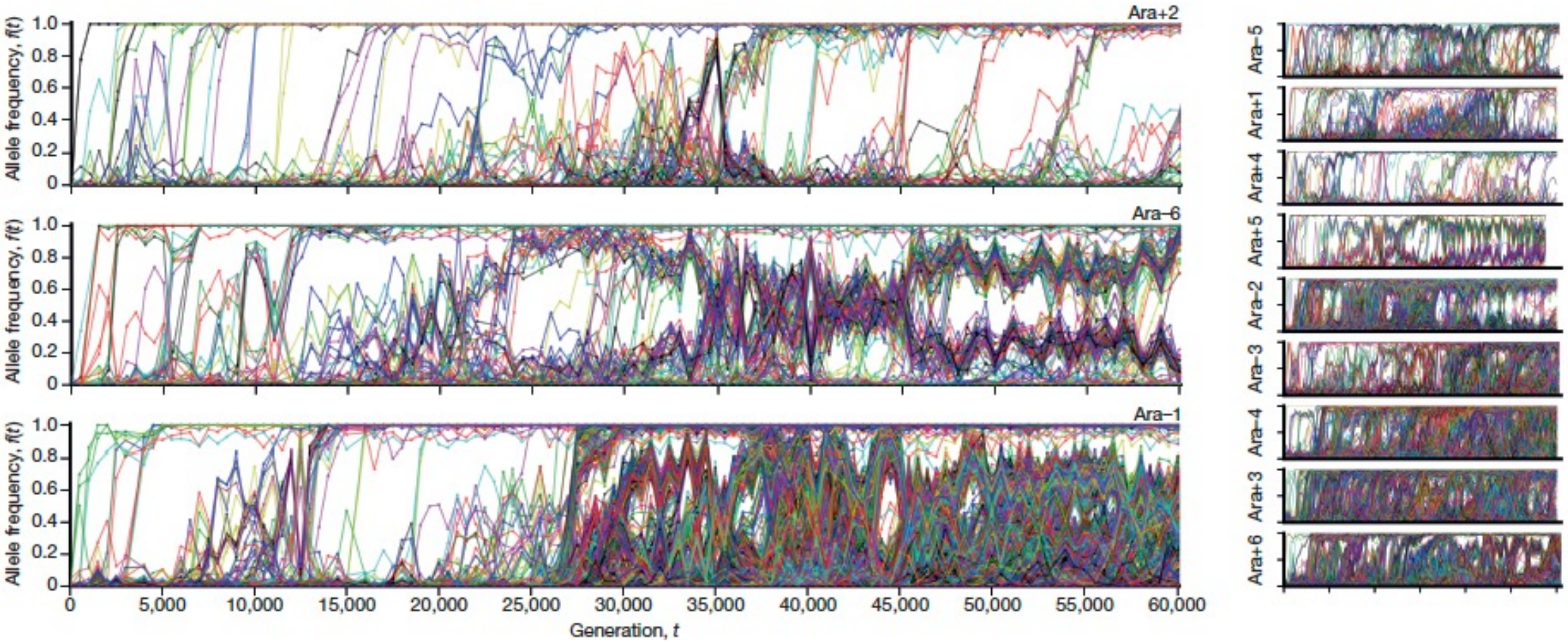
Once sequencing was possible, individual mutations could be studied



Lensky, ISME, 2017

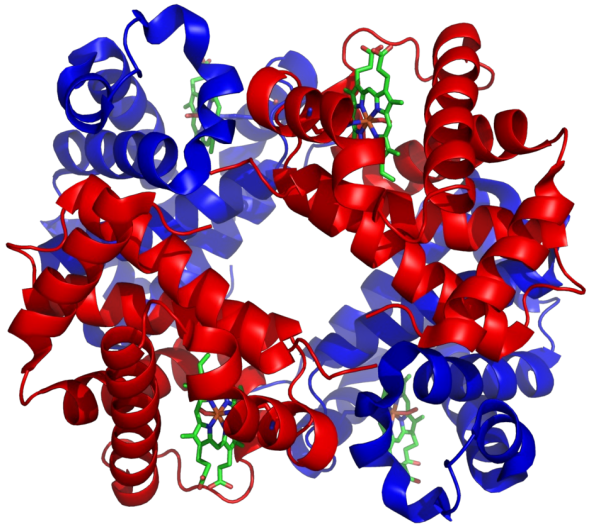
- Beneficial mutations arose, sometimes in parallel across replicates
- Sometimes initially beneficial mutations were outcompeted
- Some mutations were maintained over long time scales due to negative frequency-dependent interactions

Allele frequency dynamics over 60,000 generations



Example: Balancing selection at the beta globin locus

Beta globin is a subunit of hemoglobin

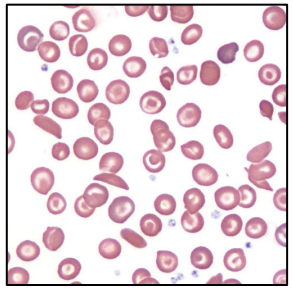


Structure of human hemoglobin. **α** and **β** globin subunits are in red and blue, respectively, and the iron-containing heme groups in green.
From PDB: 1GZX *Proteopedia Hemoglobin*

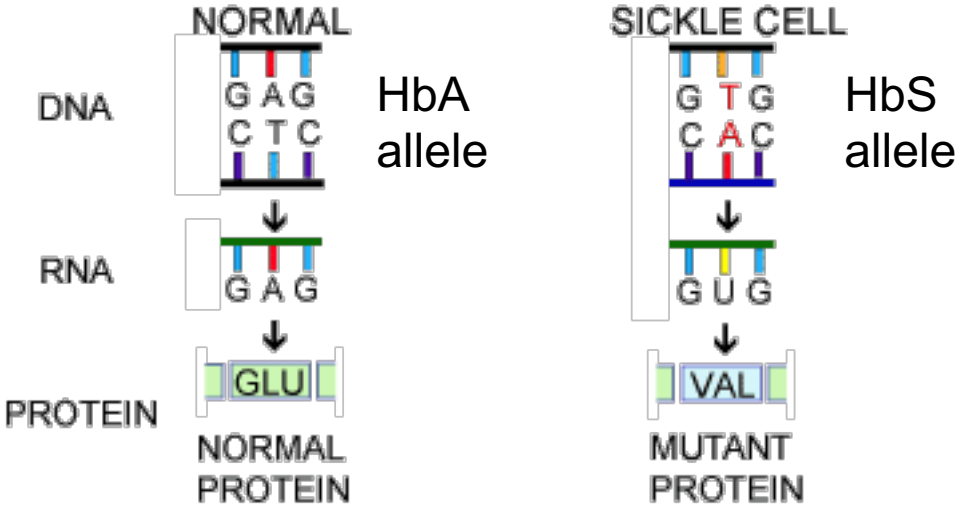
- Hemoglobin is a protein in red blood cells that carries oxygen
- Mutations in the beta globin subunit of hemoglobin can cause beta thalassemias, hemoglobin C, hemoglobin E and Sickle cell disease
- High frequencies of these diseases, especially in malaria-endemic regions, led evolutionary geneticists to hypothesize that disease-causing polymorphisms might protect carriers against malaria

The oldest discovered and best-understood case of balancing selection

- The genetic mutation in the beta globin gene that produces sickle hemoglobin (HbS) causes severe vascular complications that can lead to early death in individuals who are homozygous (SS) for the mutation
- However, in its heterozygous form (AS), the variant partially protects against severe malaria (*P. falciparum)*
- Malaria-infected individuals with the AS genotype have a 50-90% reduction in parasite density – this can result in a huge fitness advantage for carriers of the heterozygote!



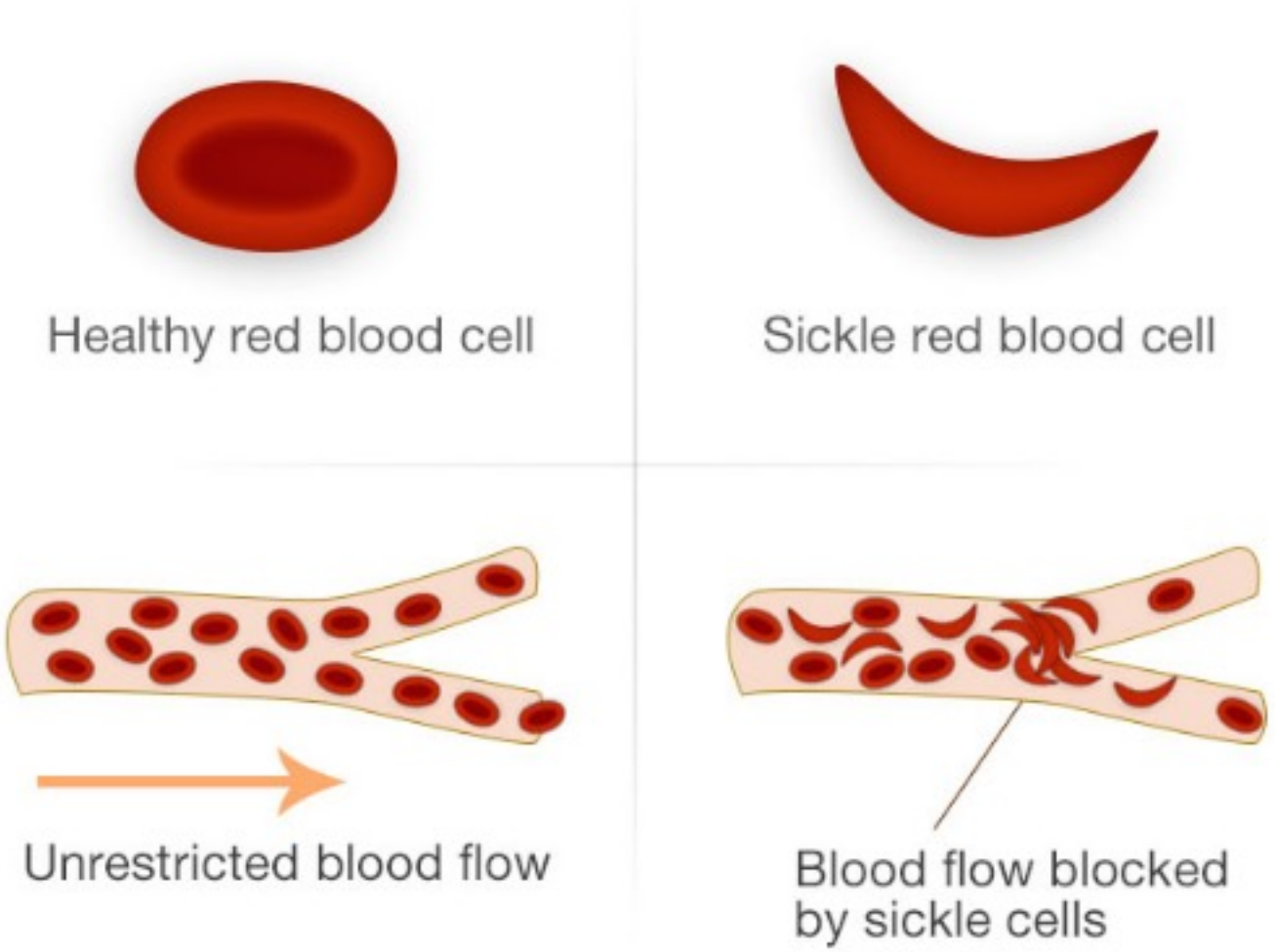
Beta globin locus and sickle cell anemia: A case of overdominance



- Sickle cell disease is caused by a SNP variant that encodes a NS protein polymorphism in the beta globin gene
- This variant causes red blood cells to be sickle-shaped

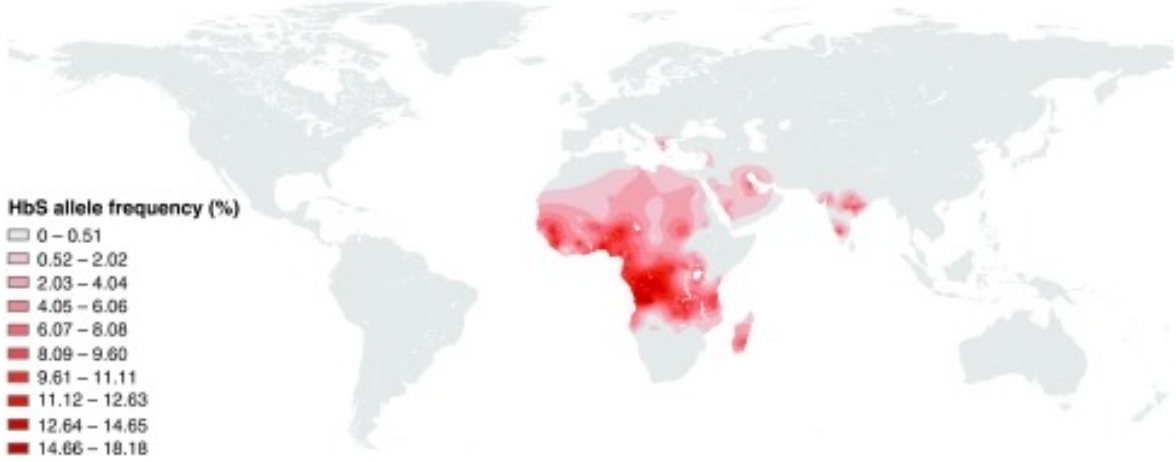
https://evolution.berkeley.edu/a-case-study-sickle-cell-anemia/

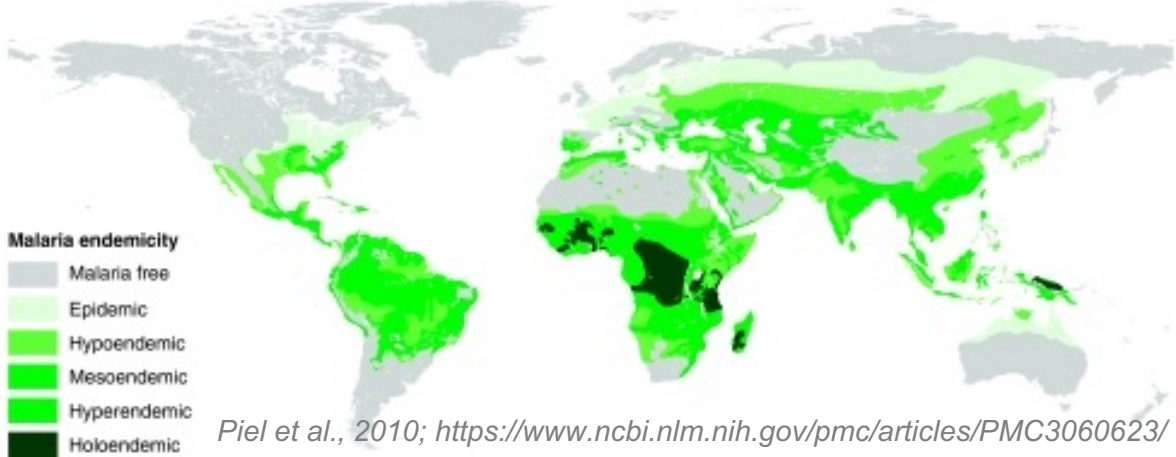
Homoygotes for the HbS allele have severe anemia



- Sickle-shaped cells are shorter-lived and less effective at carrying oxygen and tend to clump more easily, blocking blood flow
- Homozygotes for the HbS allele have severe health problems that can lead to death

Global distribution of the HbS allele compared with the distribution of endemic malaria

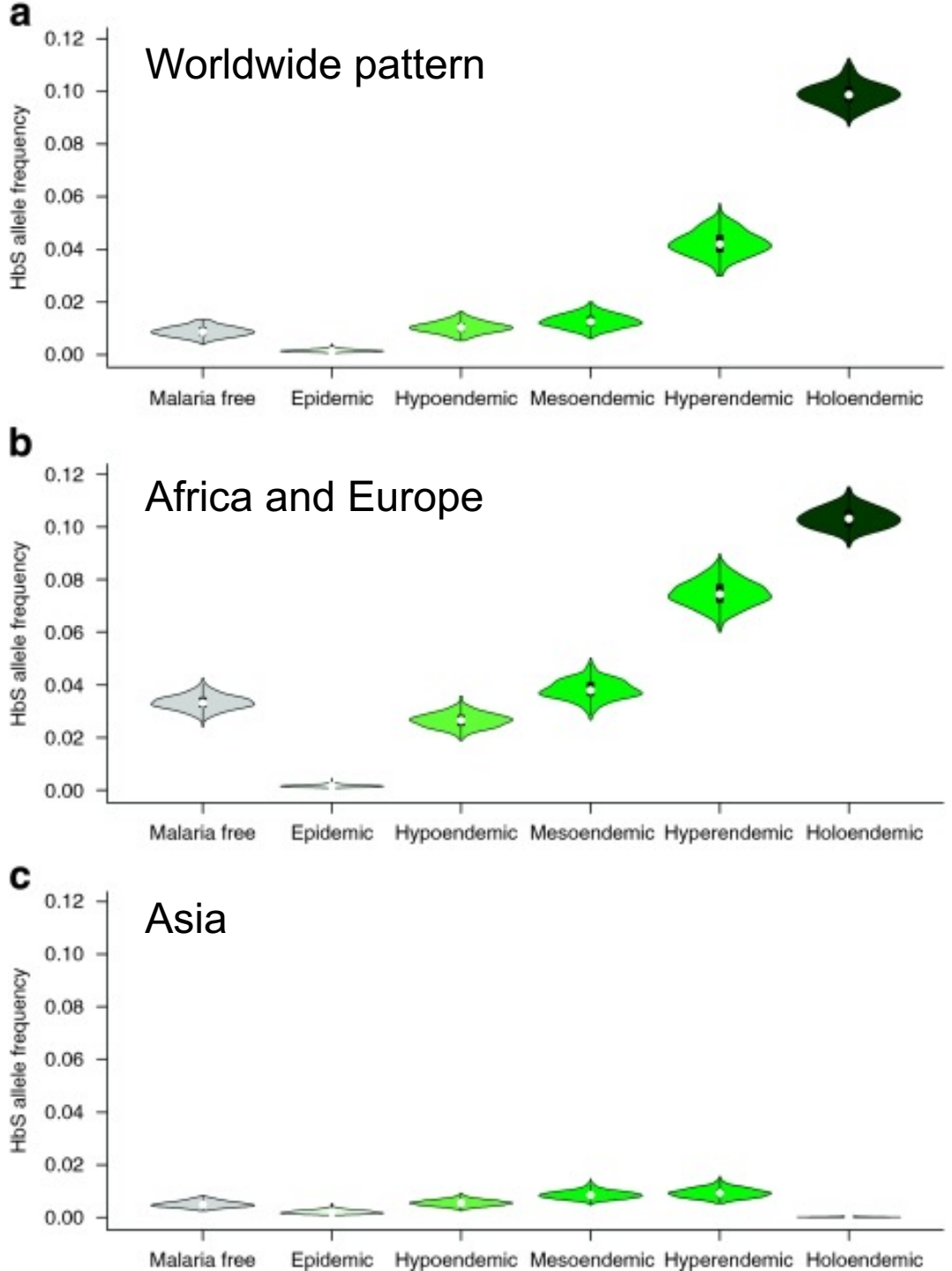




Piel et al., 2010; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3060623/

HbS allele frequency increases with higher levels of malaria endemicity

The pattern is driven mainly by variation in Africa



Summary

- Selection is one of the factors that can change allele frequencies
- Some major types of selection are directional selection, overdominant selection and purifying selection
- In the simple one locus, two allele model of selection, equilibria occur where change of $p = 0$
- There are always equilibria at $p = 0$ and $p = 1$. Other stable equilibria are possible where peaks occur in the graph of mean fitness against p
- When the heterozygote has intermediate fitness, selection favors the allele with higher fitness
- When the heterozygote has the highest fitness, the system evolves towards an equilibrium where p has intermediate frequency
- Examples of gene variants implicated in selection in wild populations, domestication, and lab-based evolution experiments have been identified