BIO-373

Genetics & Genomics

Population genetics

Sebastian M. Waszak
School of Life Sciences
Laboratory of Computational Neuro-Oncology

POPULATION GENETICS

1. Genetic variation in populations
2. Neutral theory and Hardy-Weinberg law
3. Natural selection
4. Mutations
5. Migrations
6. Genetic drift
7. Non-random mating

1. Genetic variation in populations

Population genetics

- Study of genetic variation in populations, and how it changes over time
- Merge between the scientific fields of **genetics** and **evolution**:
  - Mathematical principles developed by Wright, Fisher et Haldane in the early 20st century
  - More recently, demonstration of validity through biochemical and molecular approaches

Population

- A group of individuals belonging to same species, living in the same geographic area, actually (or potentially) interbreed

- In evolution, one should distinguish between:
  - **Microevolution**
    - Evolutionary change within populations of a species
  - **Macroevolution**
    - Evolutionary change leading to emergence of new species

Genetic variation at population scale

- **Gene pool** of a population = genetic information carried by the members of population (all existing alleles)

- Most populations have a large gene pool, so multiple alleles and high heterozygosity

- Exception: threatened populations heading toward extinction, with reduced gene pool

Detection of genetic variation in a population

- Historically seen using **artificial selection**
  - If genetic variation does exist, then phenotype changes over generations
  - Domestic dog as an example:
    - genetic and archeological evidence indicates domestication of dogs took place at least 15,000 years ago
    - Progressive selection of desired traits, based on genetic variation that was present in wild wolves

- Now done by **comparison of DNA sequences**

2. Neutral theory and Hardy–Weinberg law

Neutral theory

- The **neutral theory of molecular evolution**, proposed by Kimura in 1968, postulates that most evolutionary changes and polymorphisms within species are caused by random genetic drift of mutant alleles, and not by natural selection

Hardy–Weinberg law

- Mathematical model describing the theoretical **relationship between allelic and genotypic frequencies** in a population
- Assumes an equilibrium (i.e. **Hardy-Weinberg equilibrium**) of allelic and genotypic frequencies from one generation to the next ⇔ the frequency of alleles in gene pool does not change over time
- Only realized in an **“ideal” population**
  - Infinitely large population (sampling from a large gamete pool)
  - Same allele frequencies in eggs and sperm
  - Random mating
  - Not subject to evolutionary forces
    - No mutation
    - No migration
    - No selection

Hardy–Weinberg Equilibrium (HWE) law

- HWE is binomial sampling of alleles under random mating
- For two alleles **A** and **a** present at the same locus, the frequency of the genotypes **AA**, **Aa,** and **aa** after one generation of random mating can be calculated as follows:

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

Where ***p*** is equal to allele ***A*** frequency and ***q*** is equal to allele ***a*** frequency

Derivation of the HWE equation

- Random union of gametes drawn independently from a large and common gamete pool (same ***p*** in eggs and sperm)
- Let ***p***=*Pr(**A**)* and ***q***=*Pr*(***a***)=1-***p*** in gametes
- Let **X**=number of **A** alleles in a zygote (two Bernoulli trials)
- Probability of getting ***k*** copies of allele ***A*** under random mating:

$$X \sim \text{Binomial}(2, p)$$

$$\rightarrow \ \Pr(X = k) = \binom{2}{k} p^k q^{2-k}$$

  - $\mathrm{k} = 2: \binom{2}{2} p^2 q^{2-2} = \binom{2}{2} p^2 = p^2 \ (\mathrm{AA})$
  - $\mathrm{k} = 1: \binom{2}{1} p^1 q^{2-1} = \binom{2}{1} pq = 2pq \ (Aa)$
  - $\mathrm{k} = 0: \binom{2}{0} p^0 q^{2-0} = \binom{2}{0} q^2 = q^2 \ (aa)$

$$\boxed{p^2 + 2pq + q^2 = (p + q)^2 = 1}$$

Derivation of genotype frequencies in zygotes from allele frequencies in gametes

		Sperm	
		fr(A) = p	**fr(a) = q**
Eggs	**fr(A) = p**	**fr(AA) = p^2**	**fr(Aa) = pq**
	fr(a) = q	**fr(aA) = qp**	**fr(aa) = q^2**

		Sperm	
		fr(A) = 0.7	**fr(a) = 0.3**
Eggs	**fr(A) = 0.7**	**fr(AA) = 0.7×0.7 = 0.49**	**fr(Aa) = 0.7×0.3 = 0.21**
	fr(a) = 0.3	**fr(aA) = 0.3×0.7 = 0.21**	**fr(aa) = 0.3×0.3 = 0.09**

9% of zygotes
have an *aa* genotype

Usefulness of HWE law

- Explains how genetic variability can be maintained in a population
- Explains why dominant traits do not increase from one generation to next
- Allows the calculation of the frequencies of other genotypes from the known frequency of one genotype
- Allows the identification and quantification of the forces involved in evolution, by specifying the conditions of stability

Application to human population

- Concrete example of measurement of allelic frequencies and questions asked in population genetics, using the ***CCR5*** **gene**
  - Encodes CCR5, the main HIV-1 co-receptor expressed at the surface of CD4+ T cells
  - 32 bp deletion = **CCR5Δ32** ➔ non-functional receptor

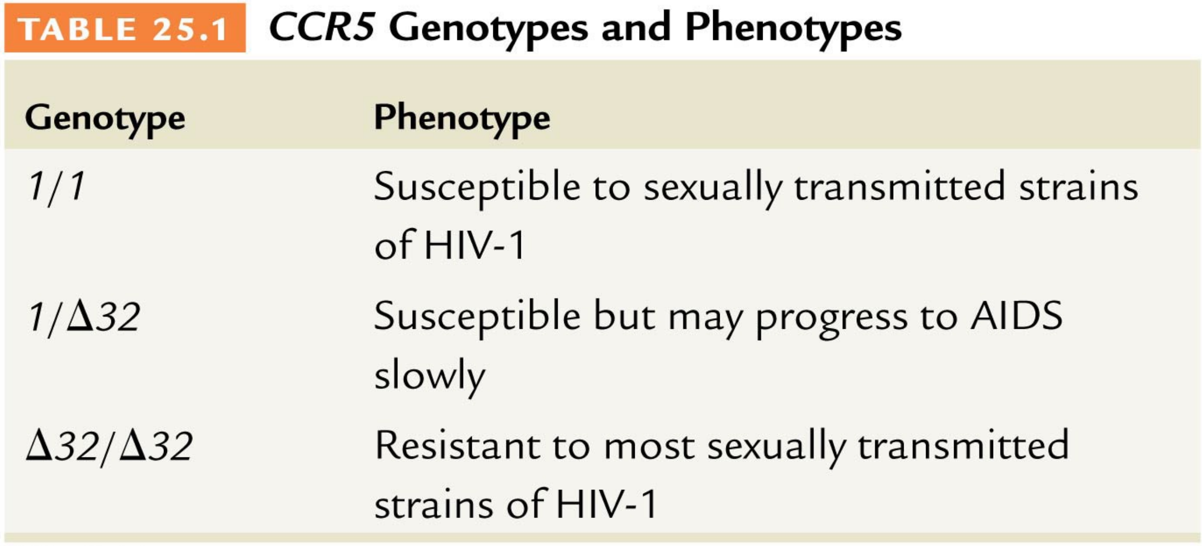
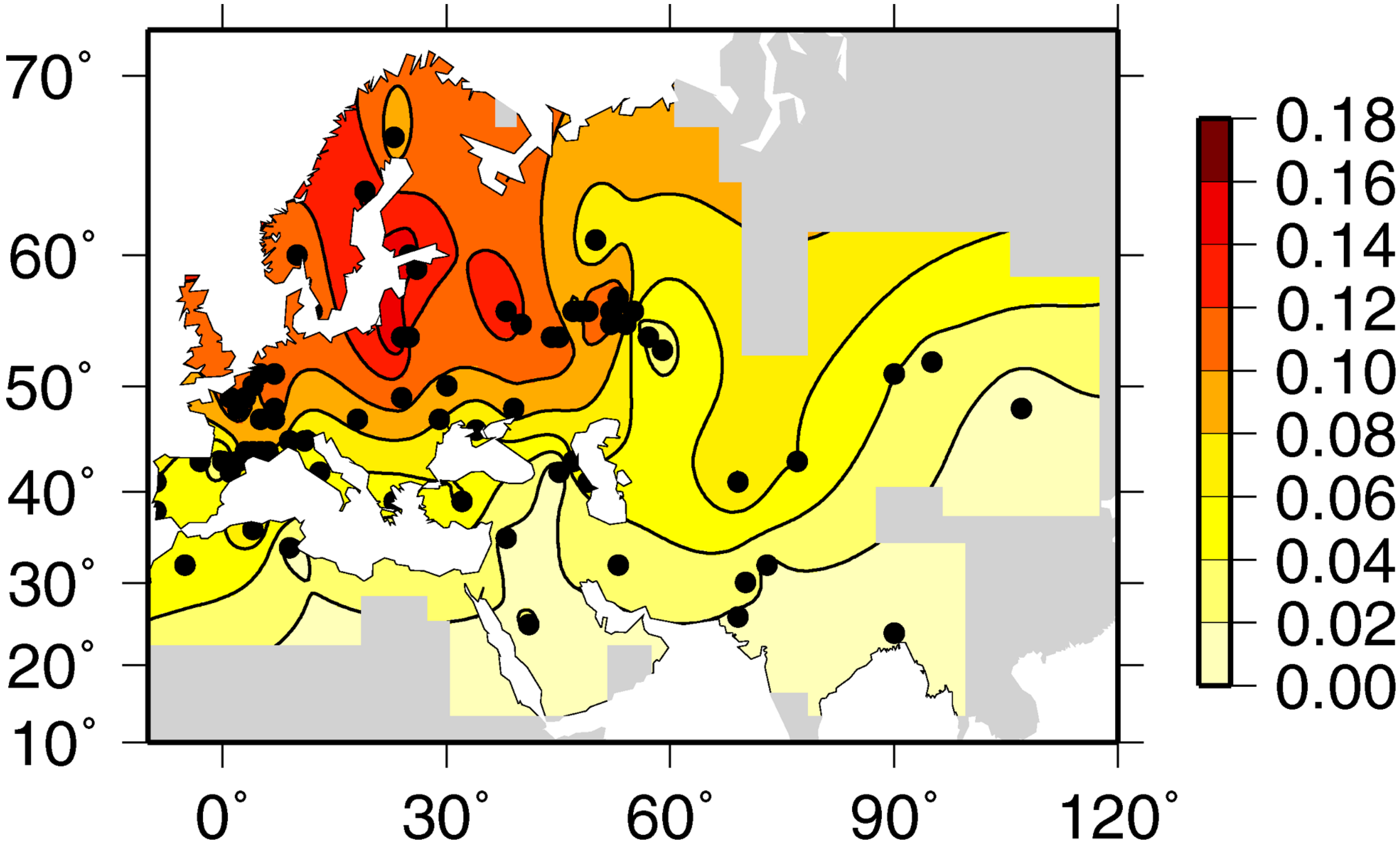


TABLE 25.1 ***CCR5*** **Genotypes and Phenotypes**

Genotype	Phenotype
1/1	Susceptible to sexually transmitted strains of HIV-1
1/Δ32	Susceptible but may progress to AIDS slowly
Δ32/Δ32	Resistant to most sexually transmitted strains of HIV-1

Geographical distribution of CCR5Δ32 allele



Determination of allele frequencies

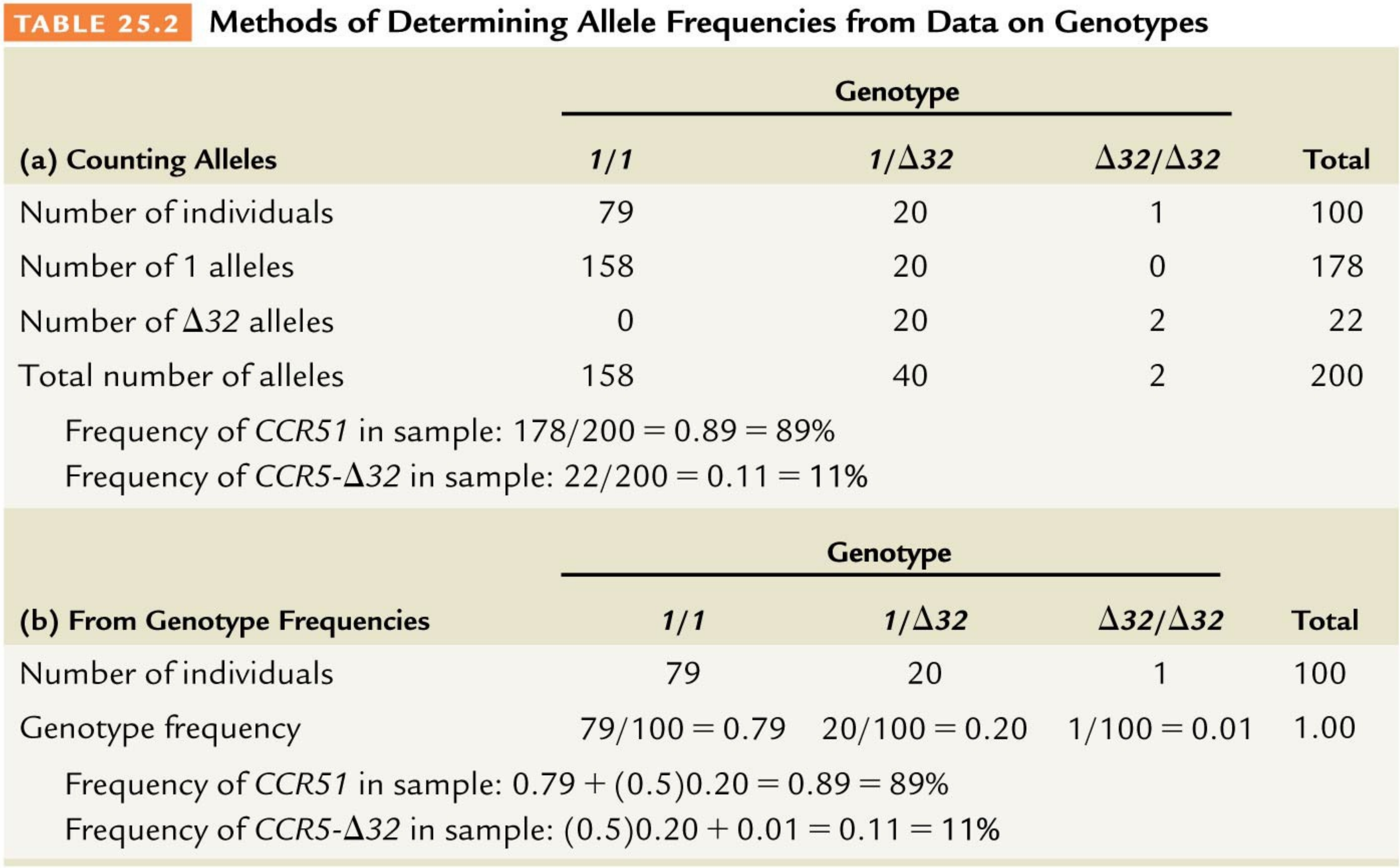


TABLE 25.2 Methods of Determining Allele Frequencies from Data on Genotypes

	Genotype			
(a) Counting Alleles	***1/1***	***1/Δ32***	***Δ32/Δ32***	**Total**
Number of individuals	79	20	1	100
Number of 1 alleles	158	20	0	178
Number of *Δ32* alleles	0	20	2	22
Total number of alleles	158	40	2	200

Frequency of *CCR51* in sample: 178/200 = 0.89 = 89%

Frequency of *CCR5-Δ32* in sample: 22/200 = 0.11 = 11%

	Genotype			
(b) From Genotype Frequencies	***1/1***	***1/Δ32***	***Δ32/Δ32***	**Total**
Number of individuals	79	20	1	100
Genotype frequency	79/100 = 0.79	20/100 = 0.20	1/100 = 0.01	1.00

Frequency of *CCR51* in sample: 0.79 + (0.5)0.20 = 0.89 = 89%

Frequency of *CCR5-Δ32* in sample: (0.5)0.20 + 0.01 = 0.11 = 11%

In the population in Hardy-Weinberg equilibrium?

Observed genotype frequencies	Derived allele frequencies
1/1 = 79%	1 = 89%
1/Δ32 = 20%	Δ32 = 11%
Δ32/Δ32 = 1%	

- From these allele frequencies, Hardy-Weinberg law is used to see if the population is at equilibrium
  - Expected frequency 1/1: $p^2 = (0.89)^2 = 79\%$
  - Expected frequency 1/Δ32: $2pq = 2(0.89)(0.11) = 20\%$
  - Expected frequency Δ32Δ32: $q^2 = (0.11)^2 = 1\%$
- Expected genotype frequencies are close to what is observed ➔ population at equilibrium and no strong selection

Calculating frequencies for multiple alleles

- If there are more than 2 alleles, additional variables can be added to the Hardy–Weinberg equation
- Example: with three alleles $p + q + r = 1$, frequencies of genotypes given by multinomial expansion:

$$(\boldsymbol{p} + \boldsymbol{q} + \boldsymbol{r})^2 = \boldsymbol{p}^2 + \boldsymbol{q}^2 + \boldsymbol{r}^2 + \mathbf{2}\boldsymbol{pq} + \mathbf{2}\boldsymbol{pr} + \mathbf{2}\boldsymbol{qr} = \mathbf{1}$$

Genotype frequency for multiple alleles

- Example of calculation for ABO blood type

TABLE 25.3 Calculating Genotype Frequencies for Multiple Alleles in a Hardy–Weinberg Population Where the Frequency of Allele $I^A = 0.38$, Allele $I^B = 0.11$, and Allele $i = 0.51$

Genotype frequency for multiple alleles

- Example of calculation for ABO blood type

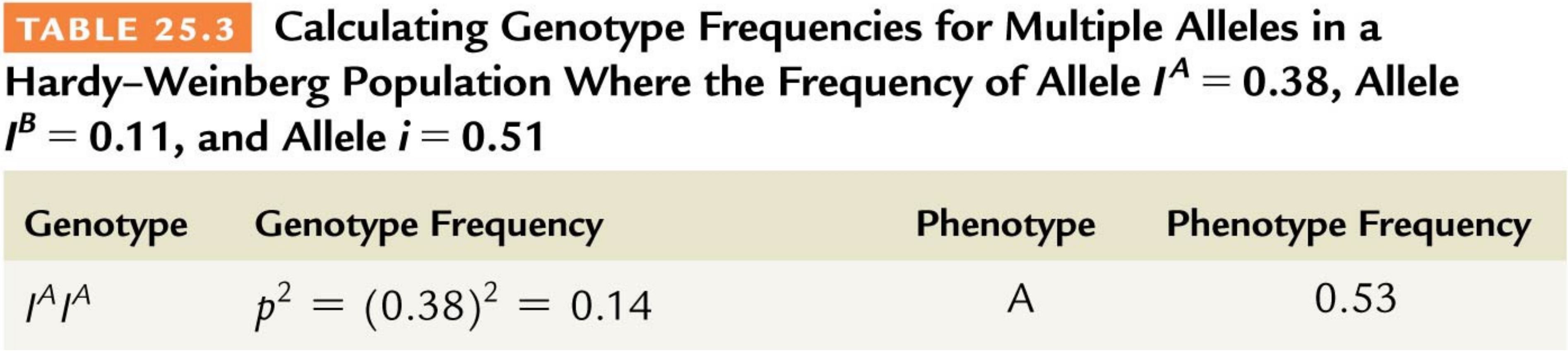


TABLE 25.3 **Calculating Genotype Frequencies for Multiple Alleles in a Hardy–Weinberg Population Where the Frequency of Allele $I^A = 0.38$, Allele $I^B = 0.11$, and Allele $i = 0.51$**

Genotype	Genotype Frequency	Phenotype	Phenotype Frequency
I^AI^A	$p^2 = (0.38)^2 = 0.14$	A	0.53

Genotype frequency for multiple alleles

- Example of calculation for ABO blood type

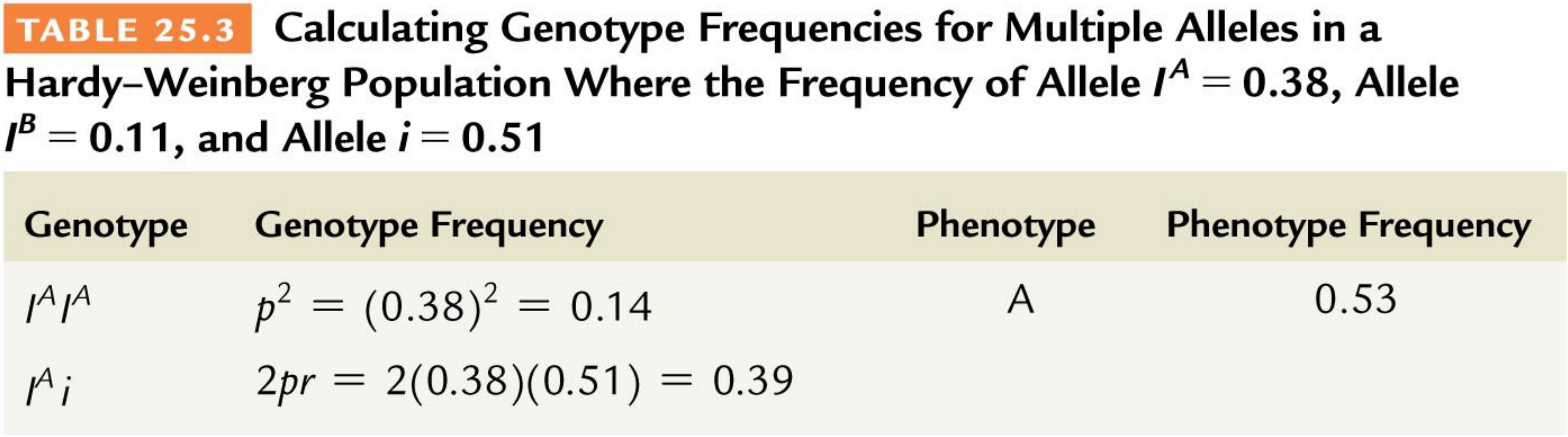


TABLE 25.3 **Calculating Genotype Frequencies for Multiple Alleles in a Hardy–Weinberg Population Where the Frequency of Allele $I^A = 0.38$, Allele $I^B = 0.11$, and Allele $i = 0.51$**

Genotype	Genotype Frequency	Phenotype	Phenotype Frequency
$I^A I^A$	$p^2 = (0.38)^2 = 0.14$	A	0.53
$I^A i$	$2pr = 2(0.38)(0.51) = 0.39$		

Genotype frequency for multiple alleles

- Example of calculation for ABO blood type

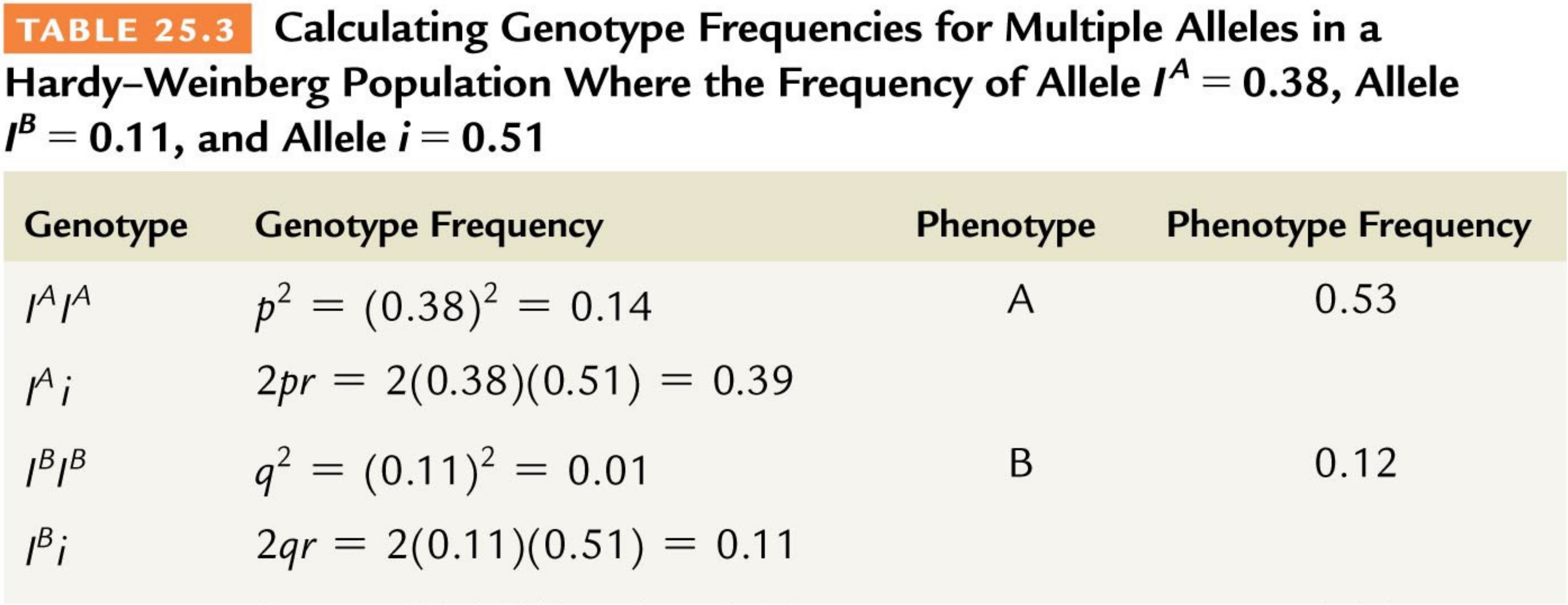


TABLE 25.3 **Calculating Genotype Frequencies for Multiple Alleles in a Hardy–Weinberg Population Where the Frequency of Allele $I^A = 0.38$, Allele $I^B = 0.11$, and Allele $i = 0.51$**

Genotype	Genotype Frequency	Phenotype	Phenotype Frequency
$I^A I^A$	$p^2 = (0.38)^2 = 0.14$	A	0.53
$I^A i$	$2pr = 2(0.38)(0.51) = 0.39$		
$I^B I^B$	$q^2 = (0.11)^2 = 0.01$	B	0.12
$I^B i$	$2qr = 2(0.11)(0.51) = 0.11$		

Genotype frequency for multiple alleles

- Example of calculation for ABO blood type

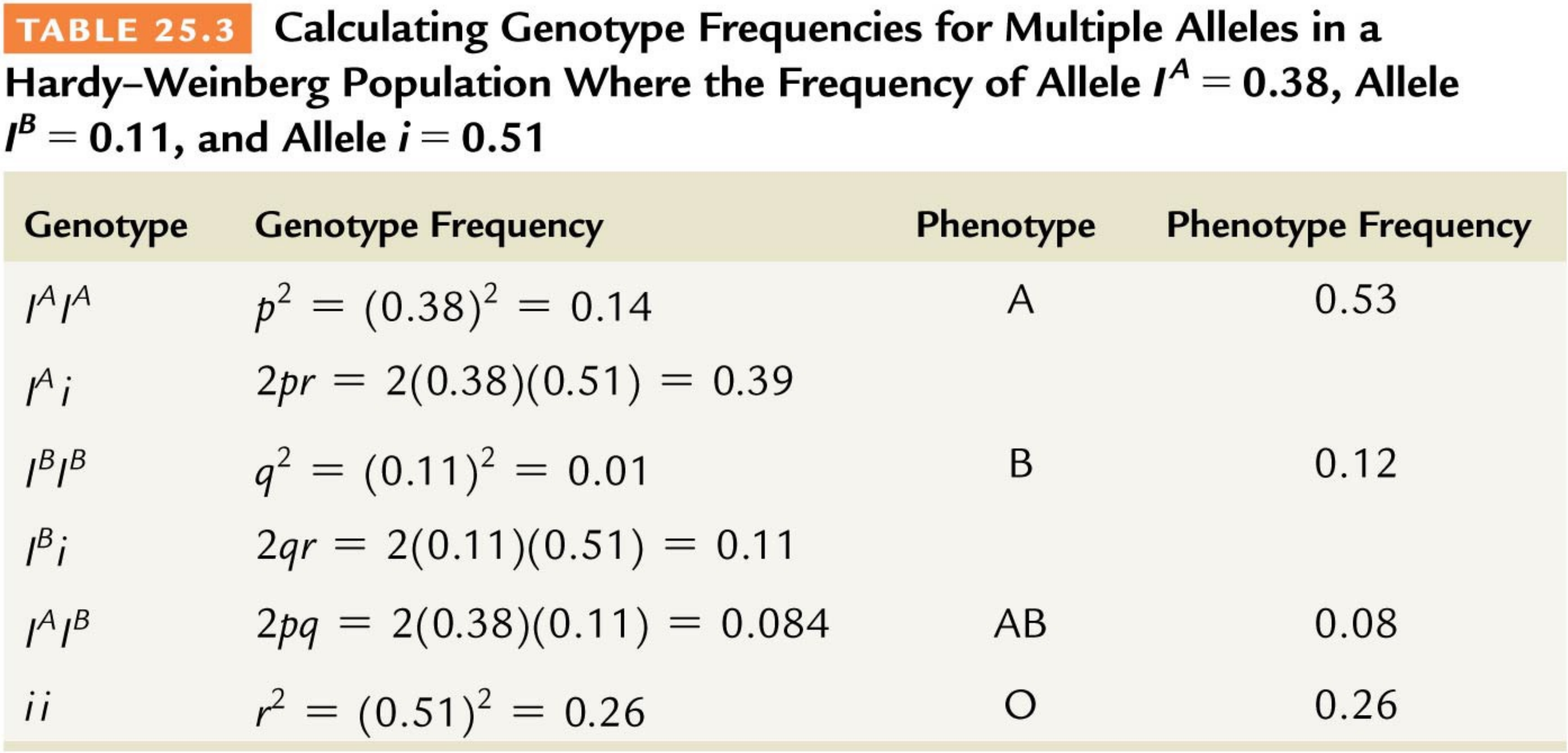


TABLE 25.3 Calculating Genotype Frequencies for Multiple Alleles in a Hardy–Weinberg Population Where the Frequency of Allele $I^A = 0.38$, Allele $I^B = 0.11$, and Allele $i = 0.51$

Genotype	Genotype Frequency	Phenotype	Phenotype Frequency
I^AI^A	$p^2 = (0.38)^2 = 0.14$	A	0.53
I^Ai	$2pr = 2(0.38)(0.51) = 0.39$		
I^BI^B	$q^2 = (0.11)^2 = 0.01$	B	0.12
I^Bi	$2qr = 2(0.11)(0.51) = 0.11$		
I^AI^B	$2pq = 2(0.38)(0.11) = 0.084$	AB	0.08
ii	$r^2 = (0.51)^2 = 0.26$	O	0.26

Frequencies of X-linked traits

- The Hardy–Weinberg equation can also be used to calculate allele and genotype frequencies for **X-linked traits**
  - Frequency of X-linked allele in gene pool = frequency of males expressing X-linked trait = q
  - Frequency of females expressing the trait, i.e. with the allele present on both X chromosomes = q^2

Frequencies of X-linked traits

- Daltonism, present in **8%** of men ➔ means that 8% of X chromosomes in the population carry the causal allele
- Allele frequency
  - $q = 0.08$
- Frequency of women with daltonism
  - $q^2 = 0.08 \times 0.08 = 0.0064 =$ **0.64%**
- Frequency of women carrying the allele
  - $2pq = 2 \times 0.92 \times 0.08 = 0.147 =$ **14.7%**

Heterozygote frequencies

- Hardy–Weinberg law allows estimation of frequency of **heterozygotes** in population
- For **recessive diseases**, this allows the calculation of the frequency of healthy carriers
- First, the frequency of a recessive phenotype can be determined by counting individuals in the population
- Example: **cystic fibrosis** (autosomal recessive)
  - Incidence of 1/2500 = 0.0004 in Europe ($q^2 = 0.0004$)
  - Frequency of recessive allele in population

$$q = \sqrt{q^2} = \sqrt{0.0004} = 0.02$$

  - Frequency of heterozygous individuals:
    Since $p + q = 1$, $p = 1 - 0.02 = 0.98$
    $2pq = 2 \times 0.98 \times 0.02 = 0.039$, or **~4%**

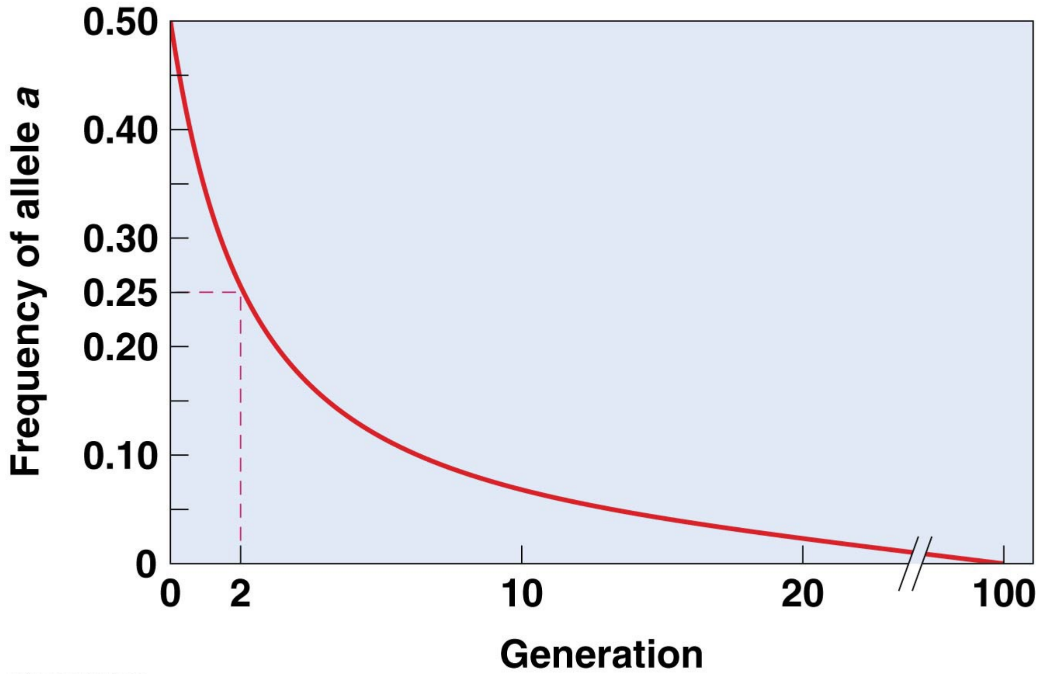
3. Natural Selection

Natural selection

- **Natural selection**
  - Difference in survival and/or reproduction rates between various genotypes
  - Discovered independently by Charles Darwin and Alfred Russel Wallace
  - Major force driving:
    - **Allele frequency change**
    - **Evolutionary changes**
- **Fitness (*w*)** = contribution of a genotype to the future generations

Fitness (*w*)

- Example: for a **lethal recessive allele *a***, with $w_{AA} = 1$, $w_{Aa} = 1$, $w_{aa} = 0$
- The frequency of the allele will decrease very quickly at first (divided by 2 after 2 generations, then again by 2 after 6 generations)
- Then much slower decrease, because most alleles are carried by heterozygous individuals



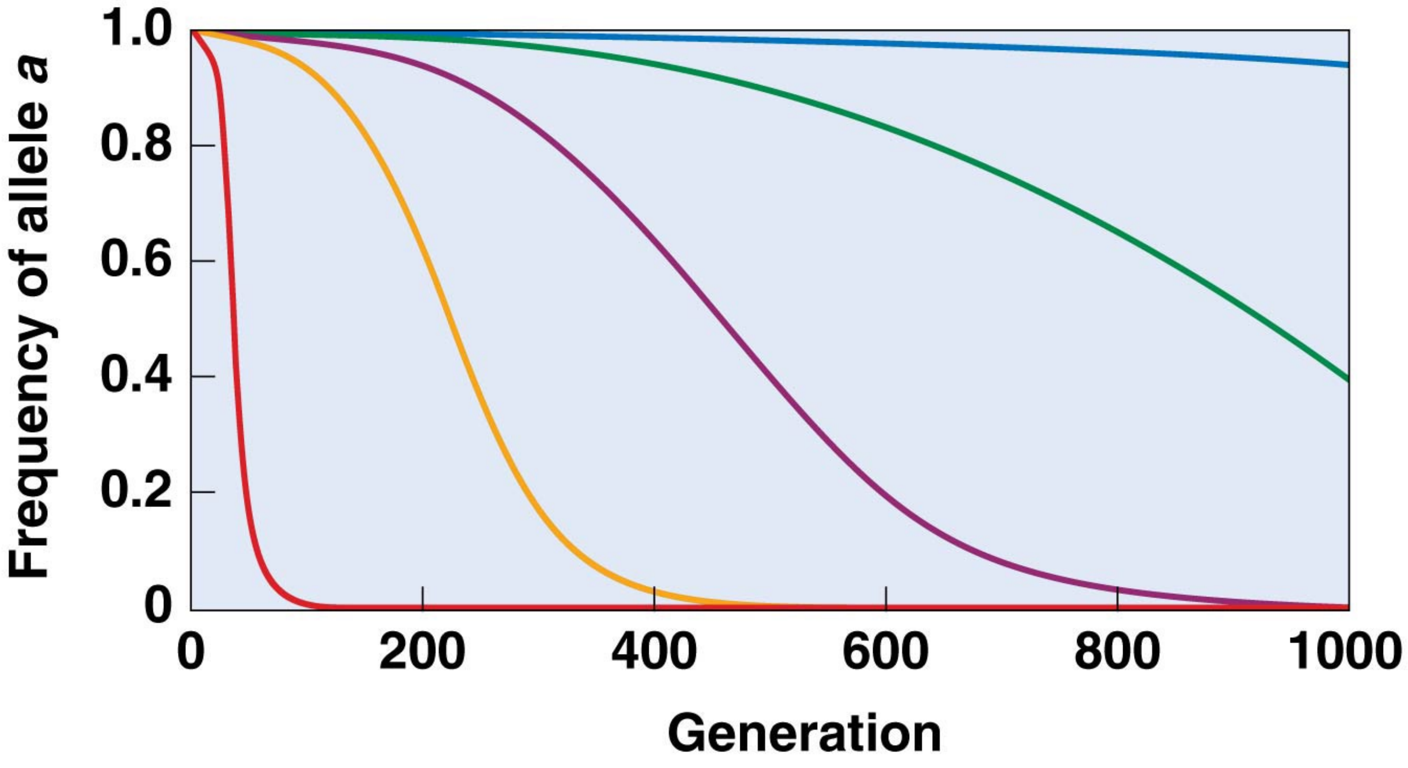
Decrease in allele frequency in defined as:

$$q_g = \frac{q_0}{1 + g q_0}$$

Where q_g is the allele frequency at generation g, q_o the initial frequency, and g the number of generations

Degrees of selection

- For non-lethal allele, the degree (intensity) of selection varies considerably
- Example for moderately deleterious alleles:



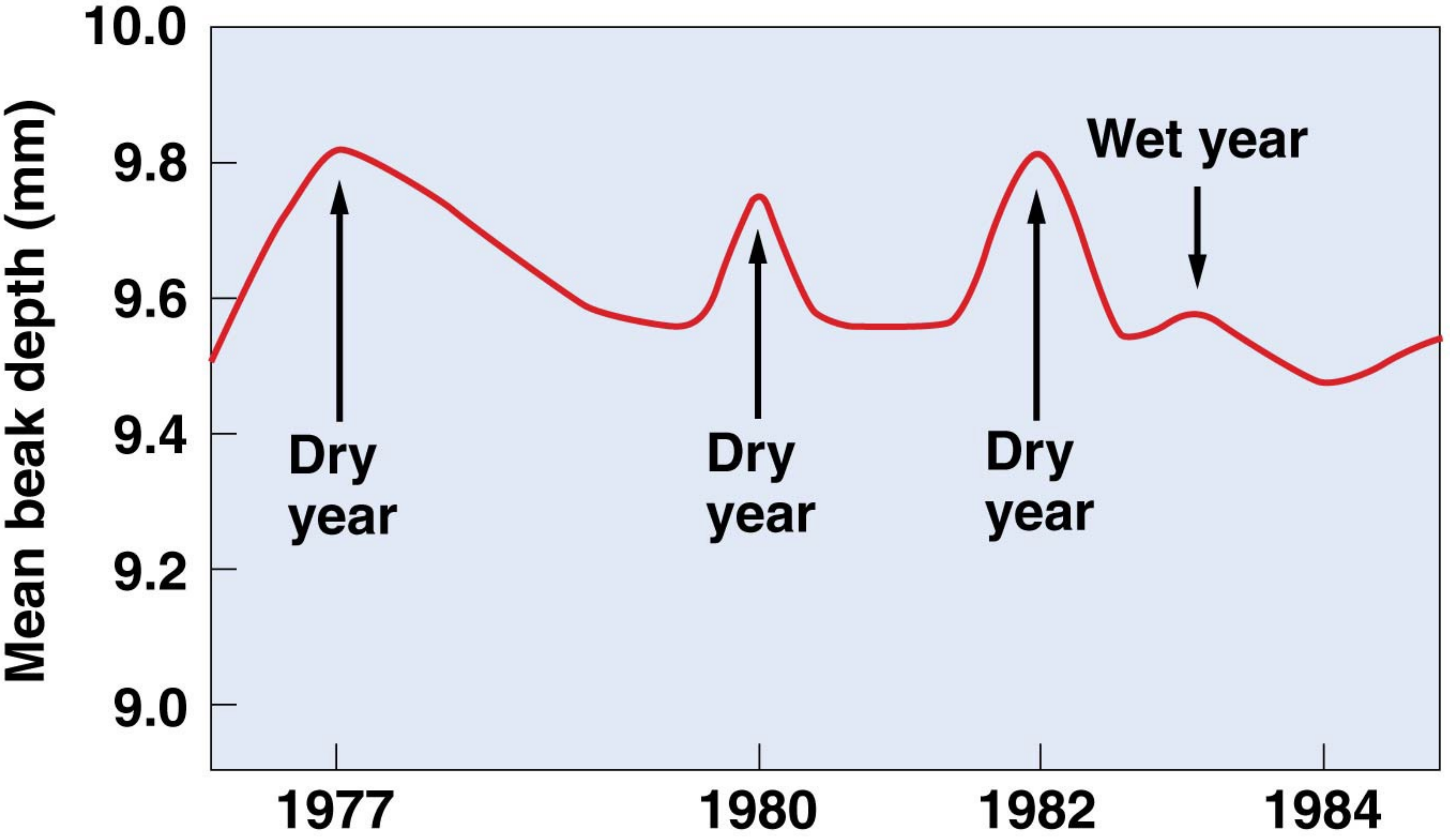
Selection of quantitative phenotypes

- Most phenotypes are quantitative and are controlled by complex combinations of genotypes + environmental influences
- Selection for these traits is classified as
  - **Directional**
  - **Stabilizing**
  - **Disruptive (diversifying)**

Directional selection

- Phenotypes at one end of spectrum become selected for or against
- Displace population mean
- Usually results from a rapid change in environment

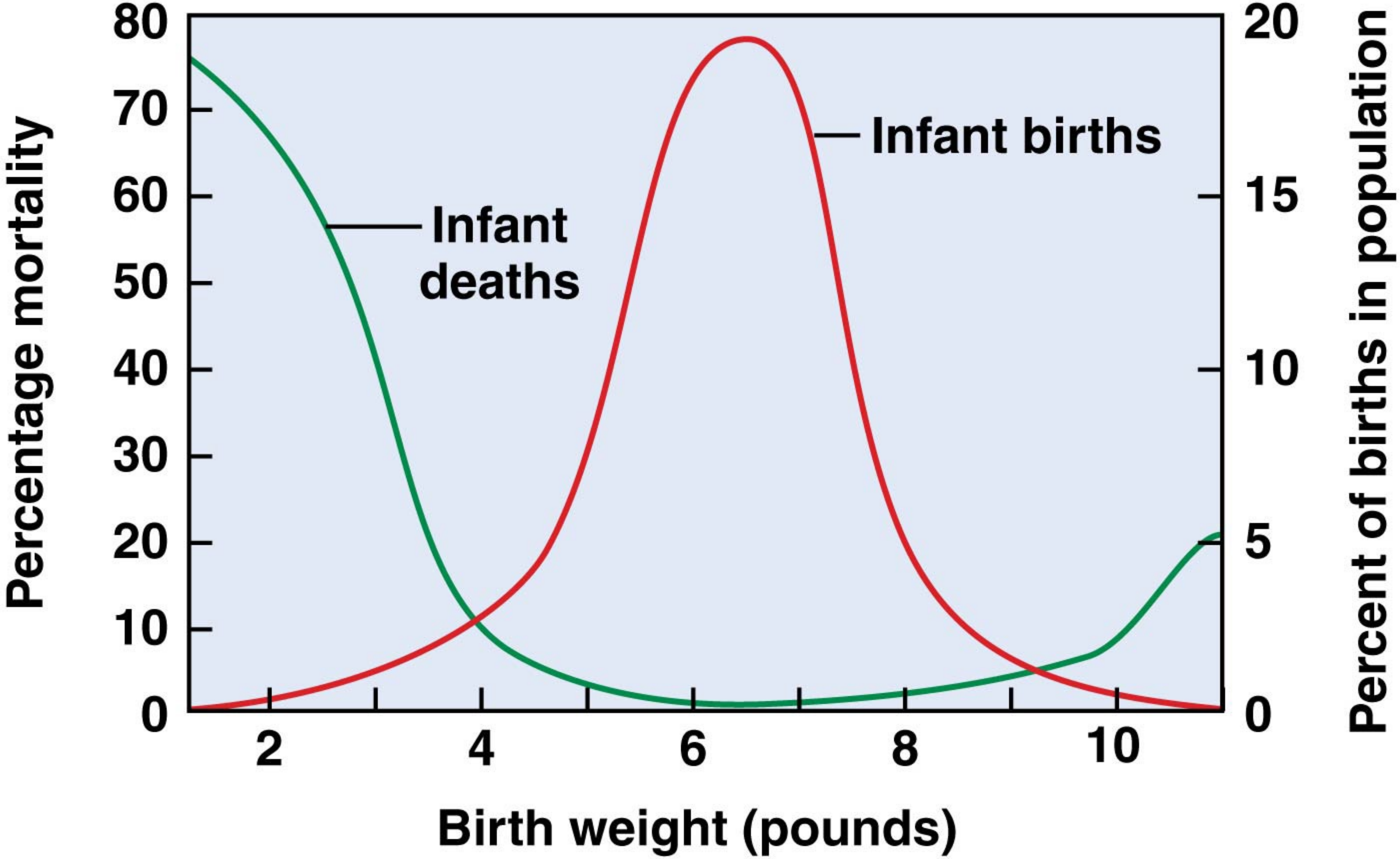
- Example: Beak size in finches during dry years increased due to strong selection (capacity to find food in restricted conditions)


Mean beak depth (mm)
10.0
9.8
9.6
9.4
9.2
9.0
1977
1980
1982
1984
Dry year
Dry year
Dry year
Wet year

Stabilizing selection

- Intermediate types are favored
- both extreme phenotypes are selected against
- Reduces population variance over time but not the mean

- Example: human birth weight and survival at 1 month

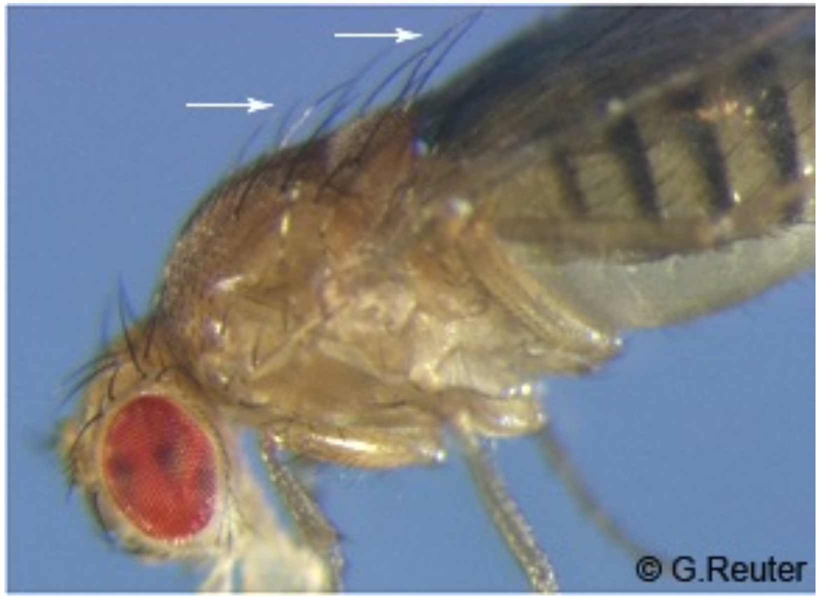


Percentage mortality
80
70
60
50
40
30
20
10
0
Infant deaths
Infant births
20
15
10
5
0
Percent of births in population
2
4
6
8
10
Birth weight (pounds)

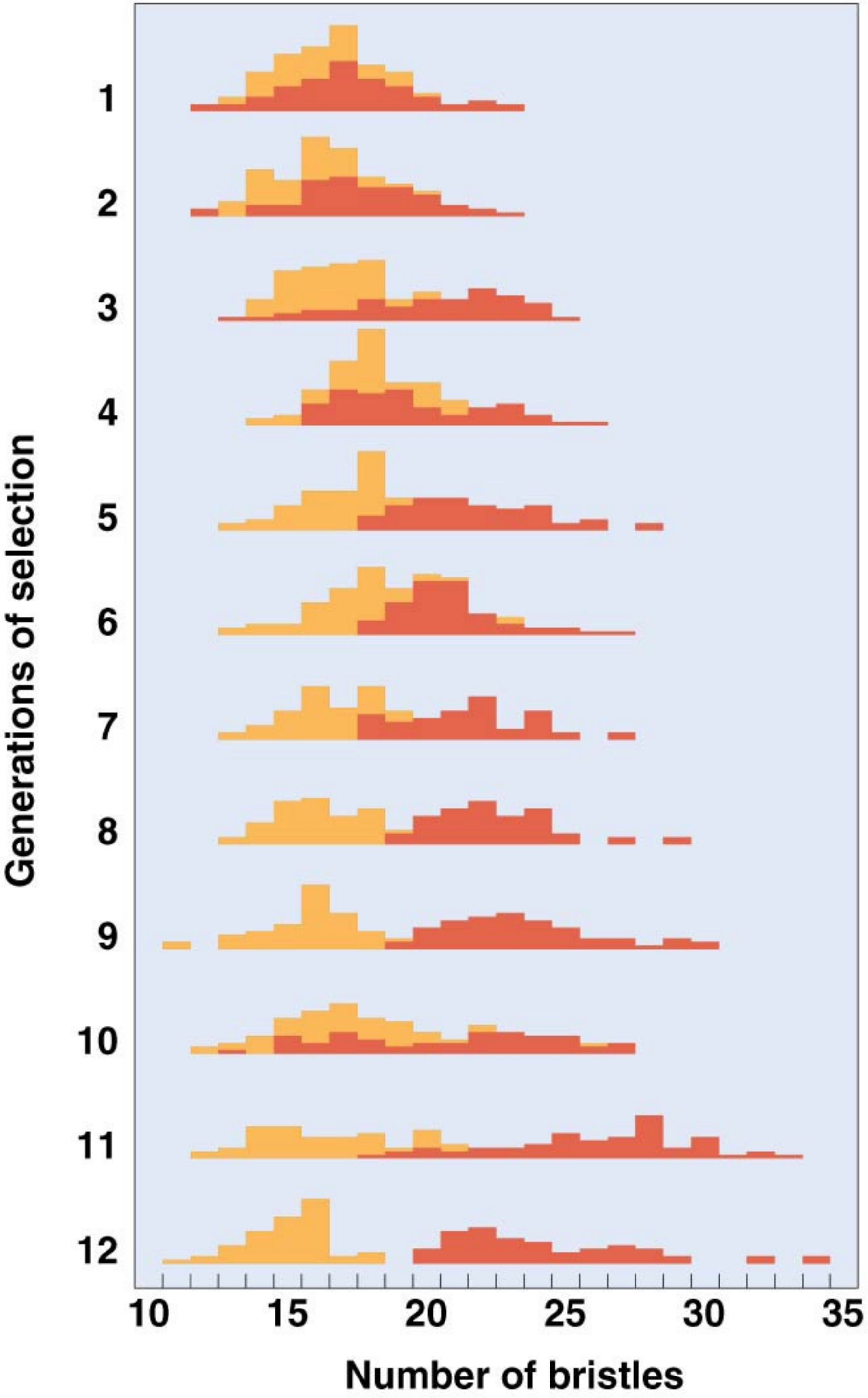
Disruptive (or diversifying) selection

- Both phenotypic extremes are selected for
- Results in population with increasingly bimodal distribution for trait

- Example: selection for low and high bristle number in *Drosophila* population



© G.Reuter



Generations of selection
1
2
3
4
5
6
7
8
9
10
11
12
10
15
20
25
30
35
Number of bristles

4. Mutations

Mutation

- Within a population, the gene pool is reshuffled each generation
- However, **mutation** is the only process that creates new alleles in gene pool
- Mutation rate in humans:

0.5×10^{-9} / bp / year

Mutation

- Very important in medical genetics, because the apparition of a mutation (“*de novo* mutation”) can have a high impact for a specific patient
- Limited importance in population genetics, because mutation has little influence on allelic frequencies at population level
  - Hundreds or thousands of generations are necessary for a new mutation to significantly increase in frequency
  - Unless positive selection is very strong

Mutation

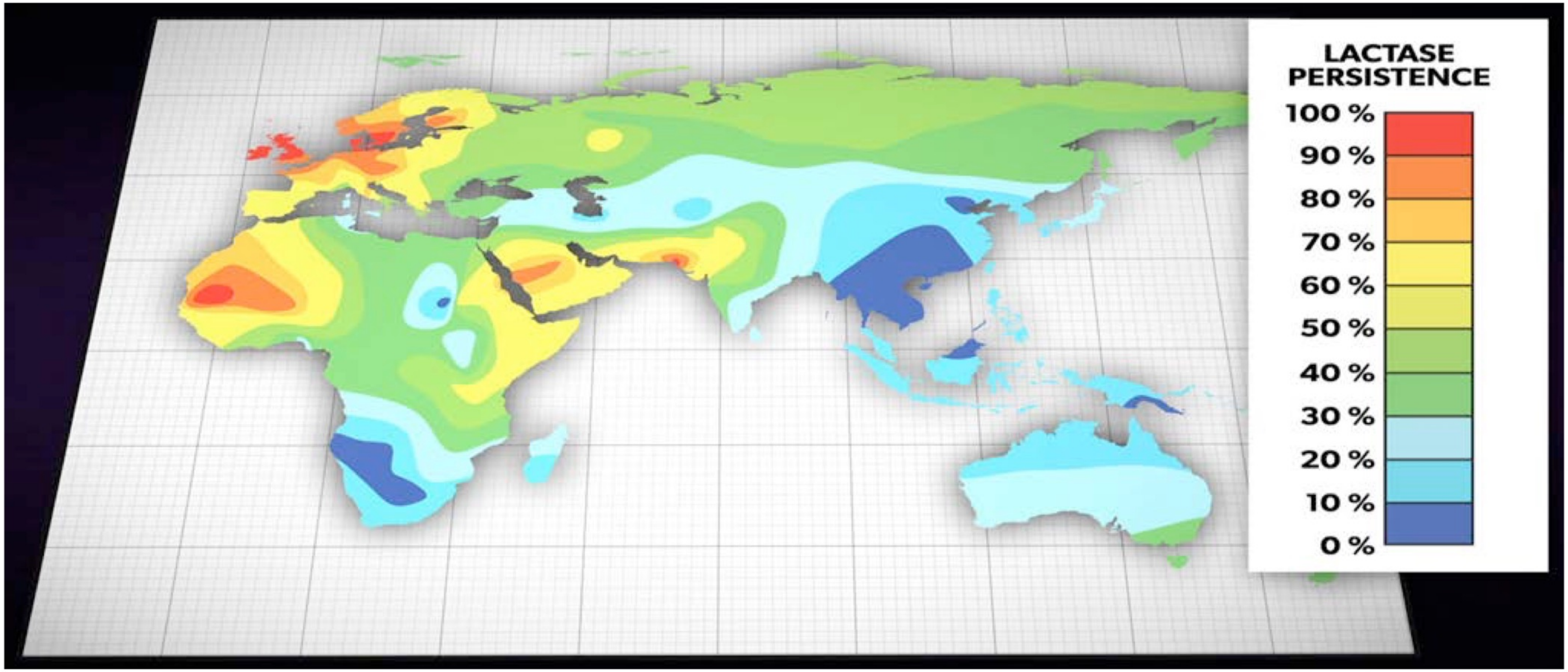


Figure 1. Global Distribution of Lactase Persistence. Worldwide, only about 35 percent of adults can digest lactose, and most are concentrated in particular geographic regions or "hot spots": northern Europe, parts of east and west Africa, the Middle East, and South Asia. (*Source:* Adapted from Curry, Andrew. "The Milk Revolution." *Nature* 500 (2013): 20–22. doi:10.1038/500020a. For an animated version of the map, see the HHMI film *Got Lactase? The Co-evolution of Genes and Culture*, http://www.hhmi.org/biointeractive/making-fittest-got-lactase-co-evolution-genes-and-culture, time stamp 4:12–4:26 minutes.)

5. Migration and gene flow

Migration

- Species might divide into populations that are **separated geographically**
- Allele frequencies in sub-populations may differ over time
- **Migration** occurs when individuals move between sub-populations with different allele frequencies

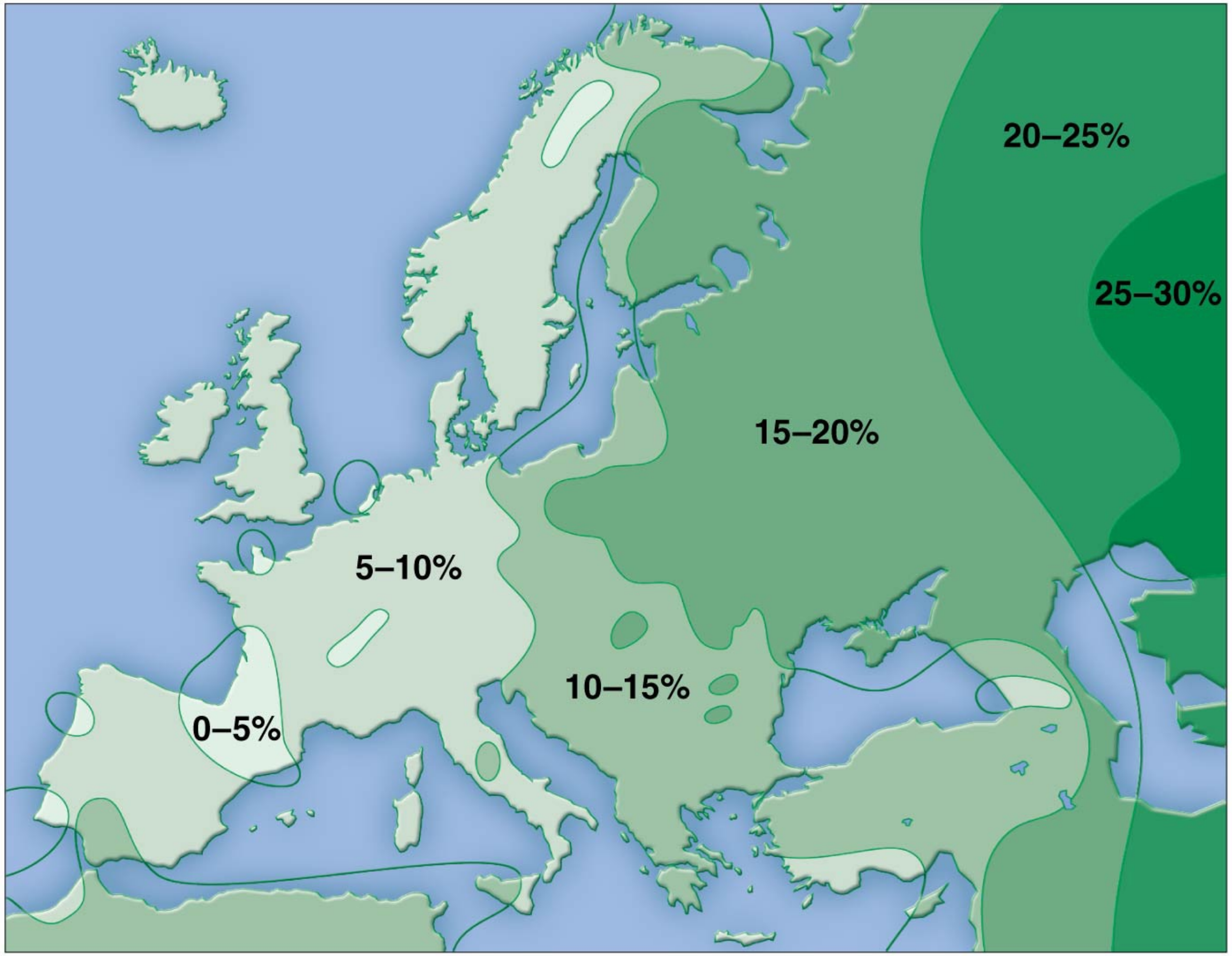
Migration

- Change in **A** allele frequency in an isolated population (p_i) after arrival of a number m of migrants from another population (p_c):

$$p_i' = (1-m)p_i + mp_c$$

- Migration can contribute significantly to allele frequency variation in a population, especially if:
  - number of migrants (m) is large
  - difference in allele frequencies is high ($p_i >> p_c$ ou $p_i << p_c$)

Frequency gradient of B allele (ABO locus), resulting from Mongols migrations between years 500 and 1000



6. Genetic drift

Genetic drift

- **Random fluctuations in allele frequencies** in a (finite) population
  - Not all alleles are transmitted to the next generation due to limited population size
  - The degree of fluctuation increases as population size decreases
- Can also result from severe reduction in gene pool:
  - **Founder effect:** population originates from small number of individuals
  - **Genetic bottleneck:** large population undergoes drastic but temporary reduction in numbers

7. Nonrandom mating

Nonrandom mating

- **Positive assortive mating**: Similar genotypes more likely to mate than dissimilar ones
- **Negative assortive mating**: Dissimilar genotypes more likely to mate than similar ones
- **Inbreeding**: Mating individuals are related
  - For a given allele, inbreeding **increases the proportion of homozygotes** in population
  - Completely inbred population consists only of homozygotes

Coefficient of inbreeding (*F*)

- Quantifies probability that two alleles of a given individual are identical **because they are descended from single copy of allele in ancestor**
  - $F = 1$: All individuals in a population are homozygous; both alleles come from same ancestral copy
  - $F = 0$: No individual has two alleles derived from a common ancestral copy