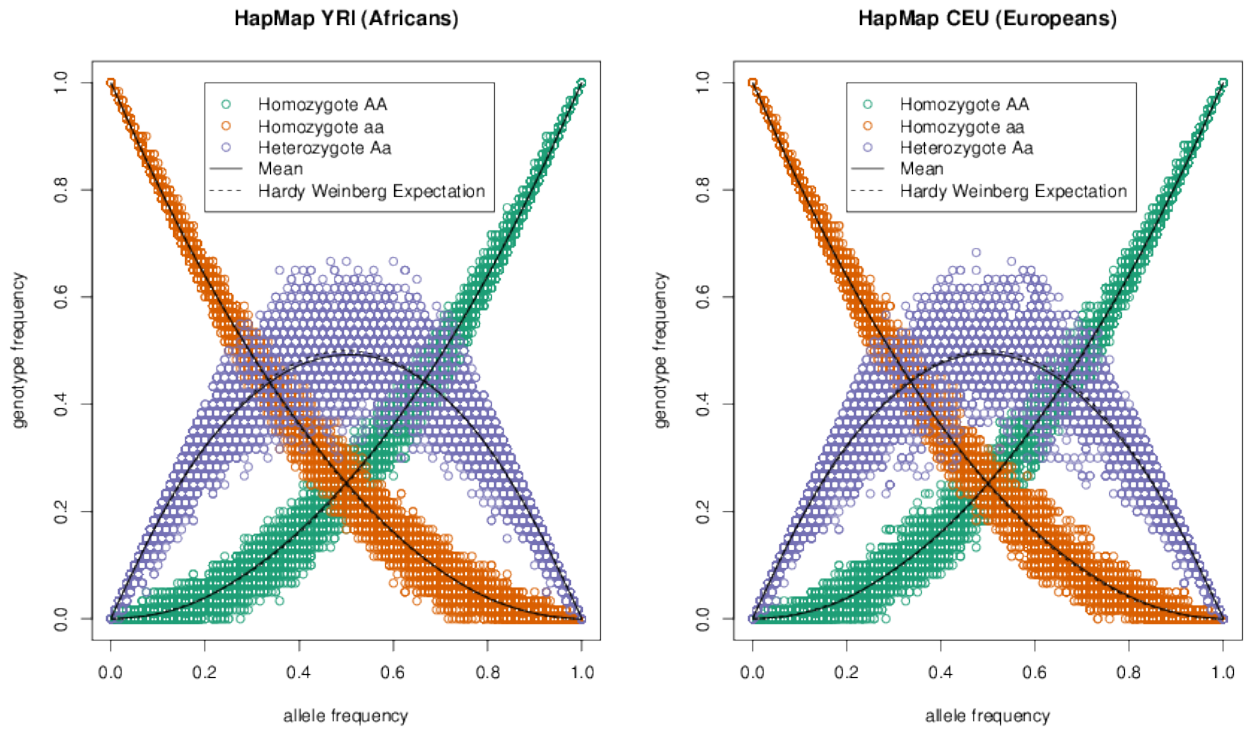
Coop, Chapter 2: Intro-2.1.1

Allele and Genotype Frequencies



Population Genetics is *all* about variation

- To understand evolutionary processes at the population level, we must consider how sequence differs across individuals both within and between populations:


>ATGGAGAACGATGAACTCAGCCCAGAAGCCAGCTAA


>ATGGAGAATGATGAACTCAGCCCAGAAGCCAGCTAA


>ATGGAAAATGATGAACTCAGCCCAGAAGCCAGCTAA


>ATGGAAAACGATGAACTCAGCACAGAAGCCAGCTAA


>ATGGCGAACGATGAACTCAGCACAGAAGCCAGCTAA

Population Genetics is *all* about variation

And there is a lot of vocabulary to keep straight

Terminology


>ATGGAGAACGATGAACTCAGCCCAGAAGCCAGCTAA


>ATGGAGAATGATGAACTCAGCCCAGAAGCCAGCTAA


>ATGGAAAATGATGAACTCAGCCCAGAAGCCAGCTAA


>ATGGAAAACGATGAACTCAGCACAGAAGCCAGCTAA


>ATGGAGAACGATGAACTCAGCACAGAAGCTAGCTAA

Terminology


>ATGGAGAACGATGAACTCAGCCCAGAAGCCAGCTAA N


>ATGGAGAATGATGAACTCAGCCCAGAAGCCAGCTAA

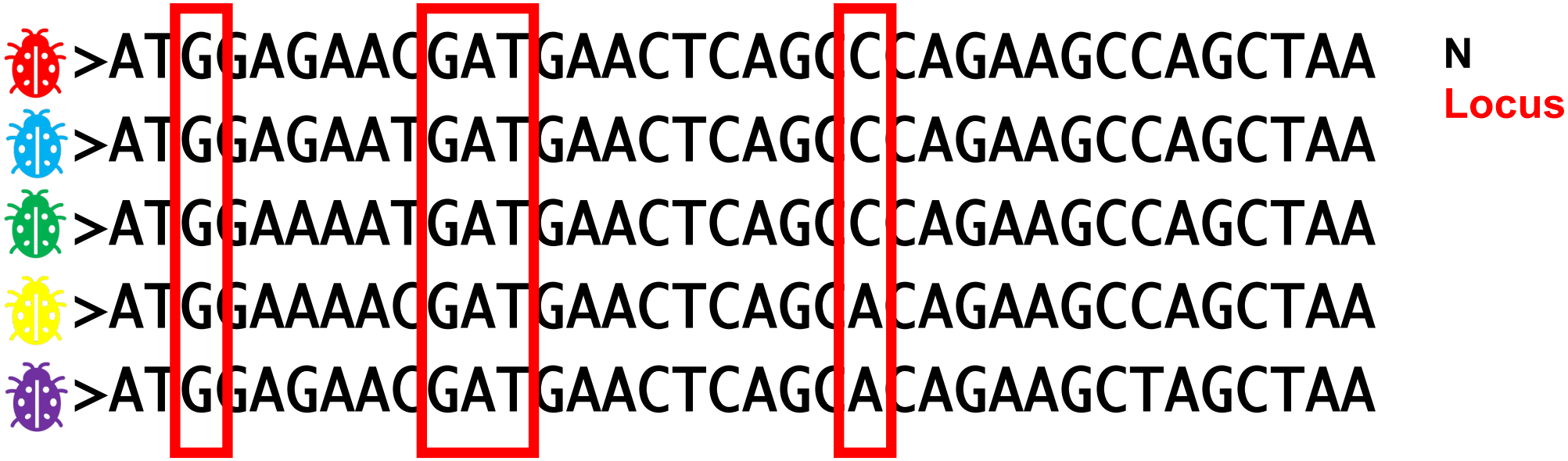

>ATGGAAAATGATGAACTCAGCCCAGAAGCCAGCTAA


>ATGGAAAACGATGAACTCAGCACAGAAGCCAGCTAA


>ATGGAGAACGATGAACTCAGCACAGAAGCTAGCTAA

N: Number of individuals sampled
from our population

Terminology

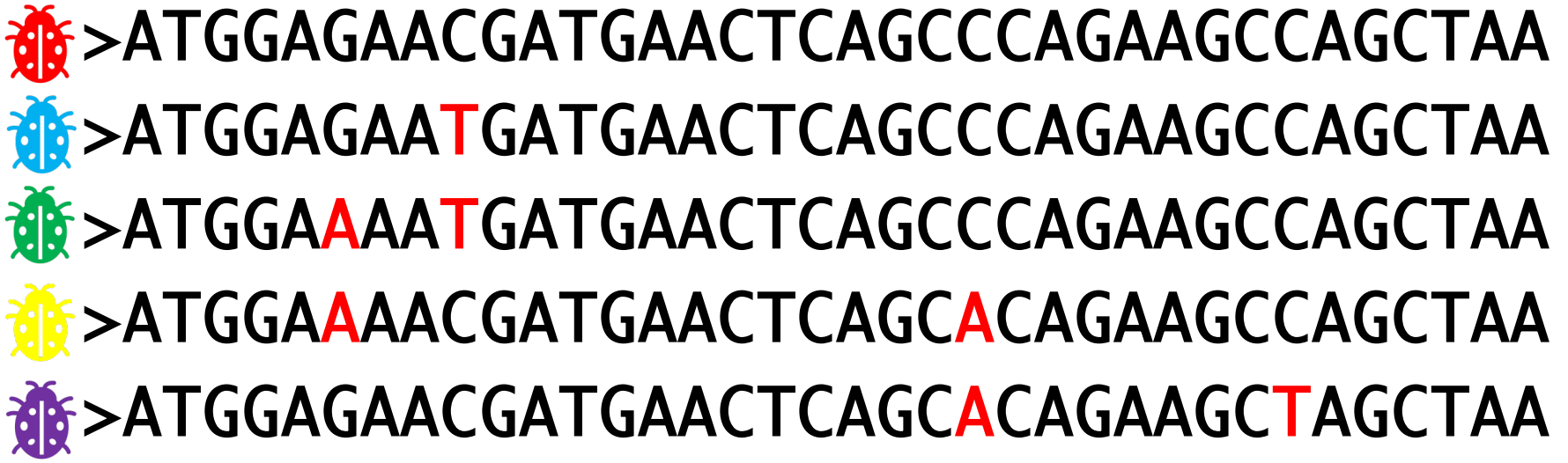


Locus: Any position in a gene or genome (plural: loci)

Terminology

Allele: A genetic variant at a locus

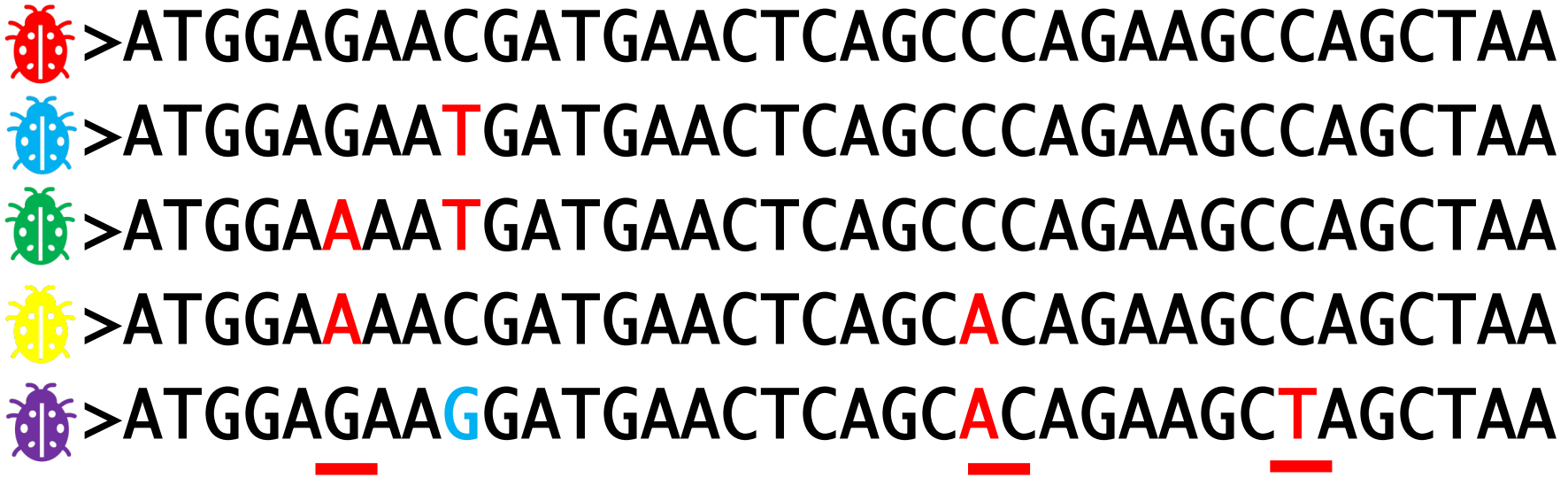
Terminology



N
Locus
Allele
Polymorphic
Segregating Site

Polymorphic: When a locus has more than one (*poly*) allele (*morphs*)
Segregating site: Same as a polymorphic locus

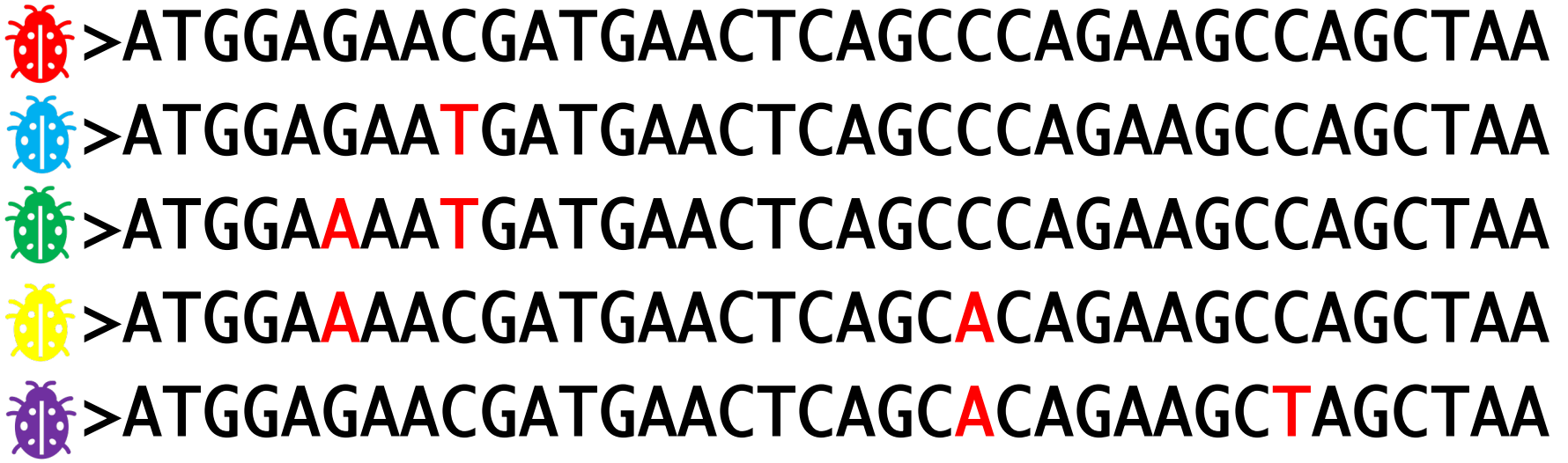
Terminology



N
Locus
Allele
Polymorphic
Segregating Site
Biallelic

Biallelic Site: A locus with two segregating alleles

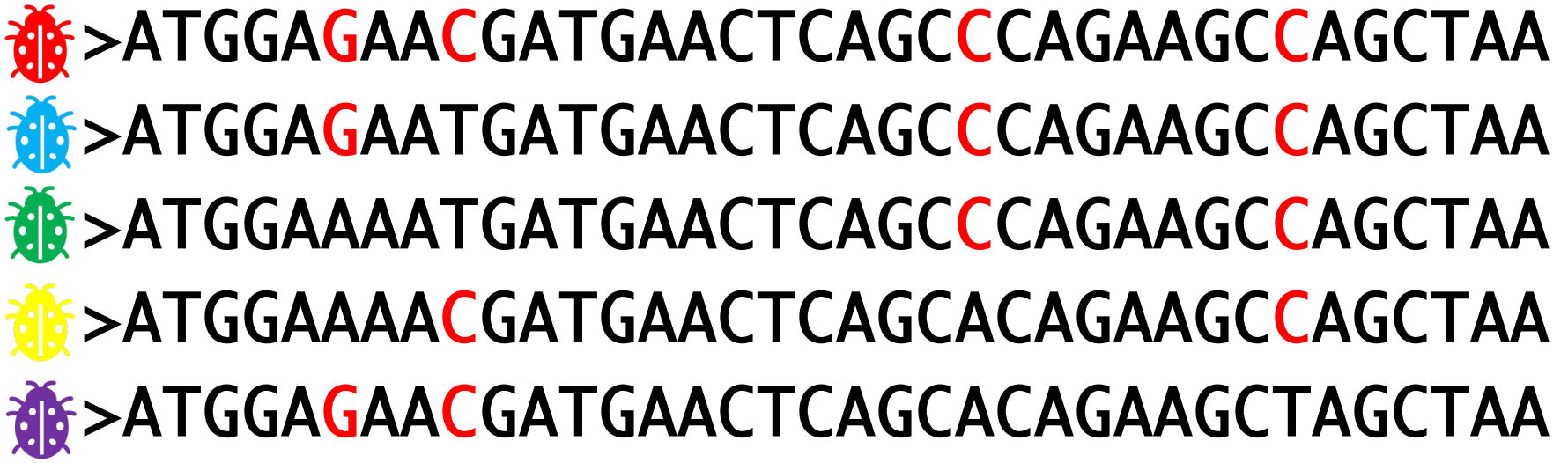
Terminology



N
Locus
Allele
Polymorphic
Segregating Site
Biallelic
Minor Allele

Minor Allele: The least common allele at a biallelic site

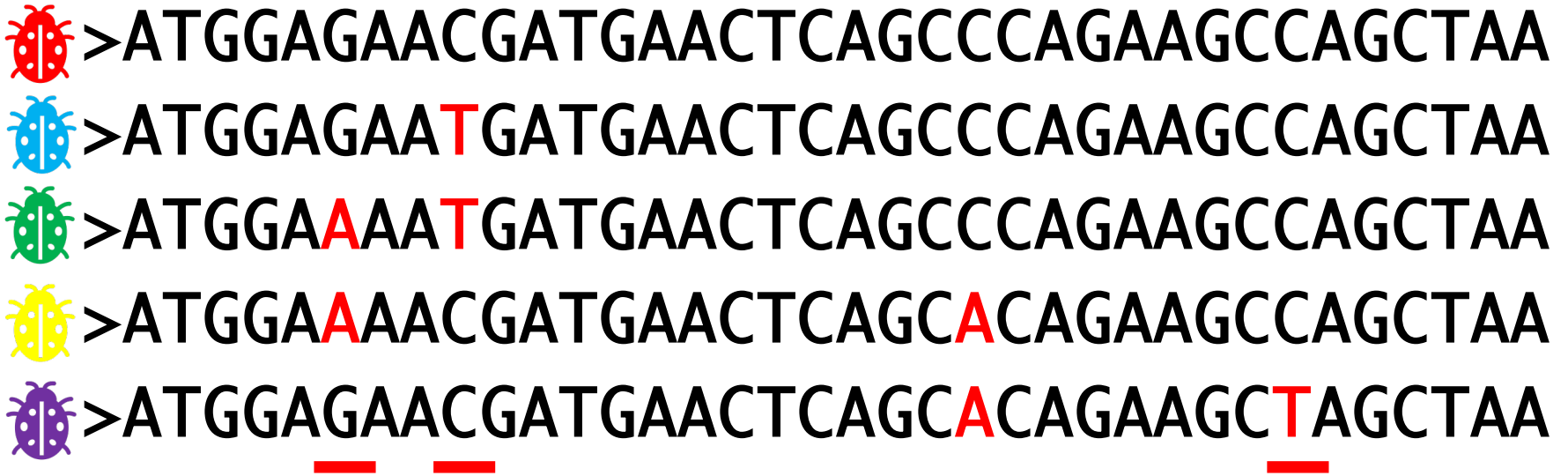
Terminology



N
Locus
Allele
Polymorphic
Segregating Site
Biallelic
Minor Allele
Major Allele

Major Allele: The most common allele at a biallelic site

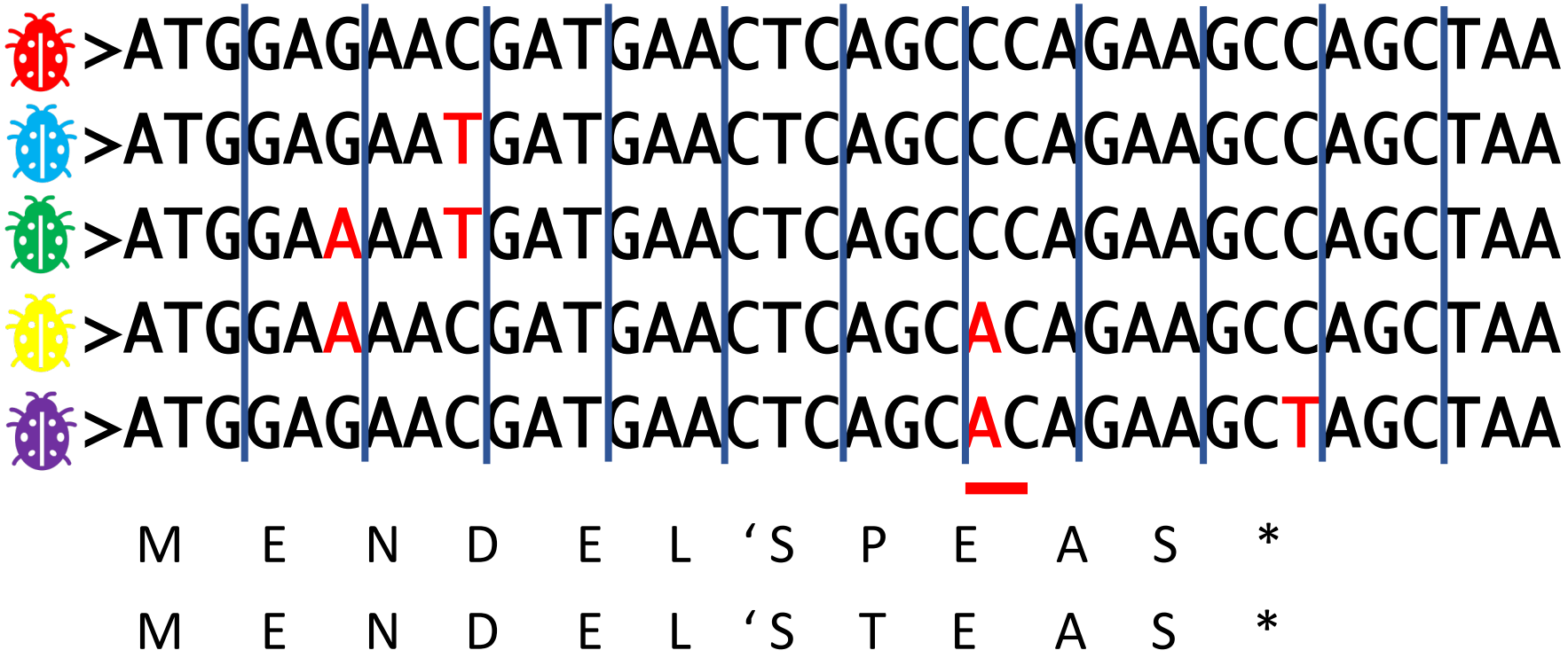
Terminology



N
Locus
Allele
Polymorphic
Segregating Site
Biallelic
Minor Allele
Major Allele
Synonymous

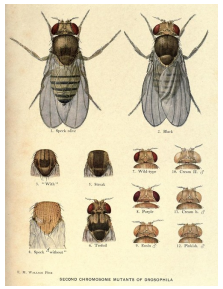
Synonymous: An allele that encodes the same amino acid in a protein

Terminology

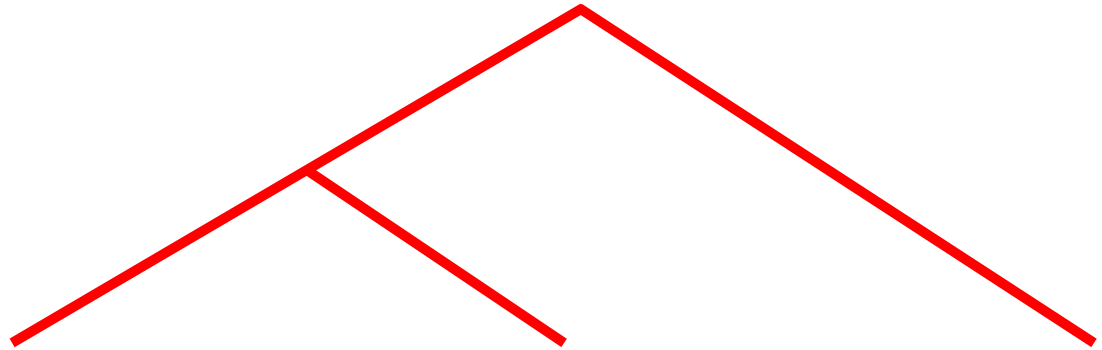


N
Locus
Allele
Polymorphic
Segregating Site
Biallelic
Minor Allele
Major Allele
Synonymous
Nonsynonymous

Nonsynonymous: An allele that encodes a different amino acid in a protein



More Terminology

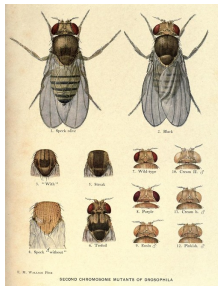


pos.	con.	a	b	c	d	e	f	g	h	i	j	k	l	a	b	c	d	e	f	a	b	c	d	e	f	g	h	i	j	k	l	NS/S
781	G	T	T	T	T	T	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	NS
789	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	C	C	C	C	C	C	C	C	C	S
808	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	NS
816	G	T	T	T	T	-	-	-	-	-	-	-	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	S
834	T	-	-	-	-	-	-	-	-	-	-	-	-	C	C	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	S
859	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	NS
867	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	A	G	G	G	G	G	G	S
870	C	T	T	T	T	T	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S

Position

Major Allele

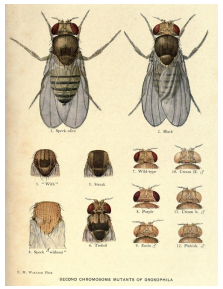
Synonymous/
Nonsynonymous



More Terminology

Polymorphism

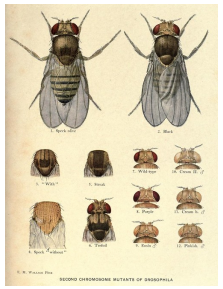
pos.	con.	a	b	c	d	e	f	g	h	i	j	k	l	a	b	c	d	e	f	a	b	c	d	e	f	g	h	i	j	k	l	NS/S
781	G	T	T	T	T	T	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	NS
789	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	C	C	C	C	C	C	C	C	C	S
808	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	NS
816	G	T	T	T	T	-	-	-	-	-	-	-	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	S
834	T	-	-	-	-	-	-	-	-	-	-	-	-	C	C	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	S
859	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	NS
867	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	A	G	G	G	G	G	G	S
870	C	T	T	T	T	T	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S
950	G	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S
974	G	-	-	-	-	-	-	-	-	-	-	-	-	T	-	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	S
983	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	C	C	C	C	C	C	C	C	C	S



More Terminology

Fixed Difference

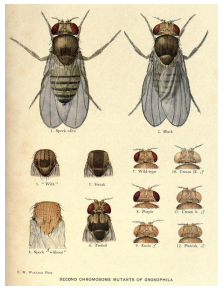
pos.	con.	a	b	c	d	e	f	g	h	i	j	k	l	a	b	c	d	e	f	a	b	c	d	e	f	g	h	i	j	k	l	NS/S
781	G	T	T	T	T	T	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	NS
789	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	C	C	C	C	C	C	C	C	C	S
808	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	NS
816	G	T	T	T	T	-	-	-	-	-	-	-	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	S
834	T	-	-	-	-	-	-	-	-	-	-	-	-	C	C	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	S
859	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	NS
867	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	A	G	G	G	G	G	G	S
870	C	T	T	T	T	T	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S
950	G	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S
974	G	-	-	-	-	-	-	-	-	-	-	-	-	T	-	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	S
983	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	C	C	C	C	C	C	C	C	C	S



More Terminology

Nonsynonymous Fixed Difference

pos.	con.	a	b	c	d	e	f	g	h	i	j	k	l	a	b	c	d	e	f	a	b	c	d	e	f	g	h	i	j	k	l	NS/S
781	G	T	T	T	T	T	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	NS
789	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	C	C	C	C	C	C	C	C	C	S
808	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	NS
816	G	T	T	T	T	-	-	-	-	-	-	-	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	S
834	T	-	-	-	-	-	-	-	-	-	-	-	-	C	C	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	S
859	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	NS
867	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	A	G	G	G	G	G	G	S
870	C	T	T	T	T	T	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S
950	G	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S
974	G	-	-	-	-	-	-	-	-	-	-	-	-	T	-	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	S
983	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	C	C	C	C	C	C	C	C	C	S



More Terminology

D. melanogaster, derived allele

D. simulans, ancestral allele

pos.	con.	a	b	c	d	e	f	g	h	i	j	k	l	a	b	c	d	e	f	a	b	c	d	e	f	g	h	i	j	k	l	NS/S
781	G	T	T	T	T	T	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	NS
789	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	C	C	C	C	C	C	C	C	C	S
808	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	NS
816	G	T	T	T	T	-	-	-	-	-	-	-	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	S
834	T	-	-	-	-	-	-	-	-	-	-	-	-	C	C	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	S
859	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	NS
867	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	A	G	G	G	G	G	G	S
870	C	T	T	T	T	T	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S
950	G	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S
974	G	-	-	-	-	-	-	-	-	-	-	-	-	T	-	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	S
983	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	C	C	C	C	C	C	C	C	C	S

2.1 Allele Frequencies

Allele frequencies are a central unit of population genetic analyses.

An individual's "genotype" is its complement of alleles at a locus

Based on genotypes we can calculate allele frequencies

2.1 Allele Frequencies

N : Total number of individuals in a population

A_1 ,A_2 *:* Two alleles at a biallelic locus

N_{11} : Number of homozygous A_1A_1 individuals

N_{12} : Number of heterozygous A_1A_2 individuals

Genotype frequencies:

$f_{11} = N_{11} / N$ $\qquad$ $f_{12} = N_{12} / \mathrm{N}$

The frequency of allele A_1 :

$$p = \frac{2N_{11} + N_{12}}{2\mathrm{N}} \quad \textbf{and} \quad p = f_{11} + \frac{1}{2} f_{12}$$

The frequency of allele A_2:

$$q = 1 - p$$

2.1.1 Measures of Genetic Variability

- Nucleotide diversity (π): Average number of single nucleotide differences between haplotypes chosen at random from a population

- With 6 samples, there are 15 different possible pairs

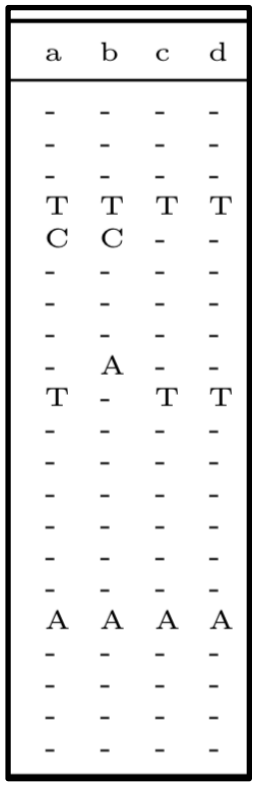
ab = 2 ac = 1 ad = 1 ae = 1 af = 0

bc = 3 bd = 3 be = 3 bf = 2

cd = 0 ce = 0 cf = 1

de = 0 df = 1

ef = 1



a	b	c	d
-	-	-	-
-	-	-	-
-	-	-	-
T	T	T	T
C	C	-	-
-	-	-	-
-	-	-	-
-	-	-	-
-	A	-	-
T	-	T	T
-	-	-	-
-	-	-	-
-	-	-	-
-	-	-	-
-	-	-	-
-	-	-	-
A	A	A	A
-	-	-	-
-	-	-	-
-	-	-	-
-	-	-	-

D. simulans data

- π = 1/15(2+1+1+1+0+3+3+3+2+0+0+1+0+1+1)
- π = 1.26

2.1.1 Measures of Genetic Variability

- π depends on the length of a sequence, so let's instead calculate π per base pair:
  - Total length of *ADH* gene (monomorphic + polymorphic sites) = 397bp
  - π per bp = 1.26/397 = 0.0032

- Interpretation: on average, about a third of a percent (1/300) of sites are polymorphic between two *D. simulans* individuals at the *ADH* locus

- We can also calculate π at only synonymous sites, or only nonsynonymous sites

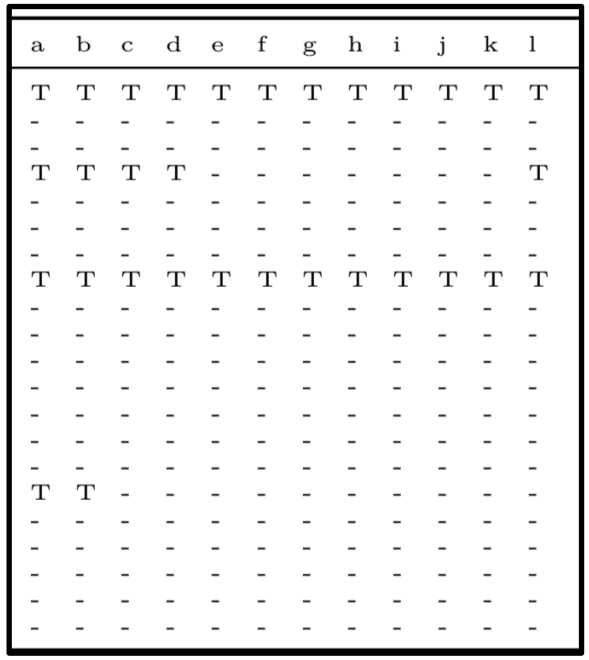
2.1.1 Measures of Genetic Variability

- The number of sites that are segregating within a given locus can also be calculated as a summary of genetic variability

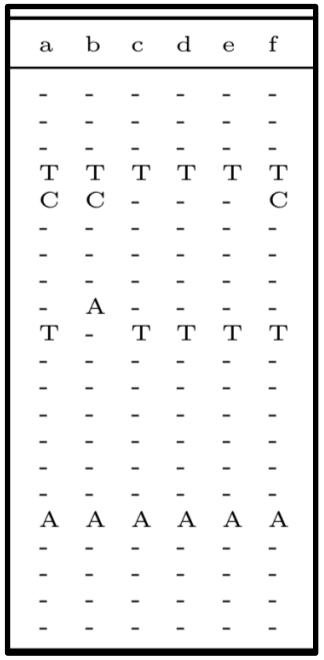
- There are 2 segregating sites in *D. melanogaster* and 3 segregating sites in *D. simulans* at the *ADH* locus

- This depends critically on the number of individuals sequenced

D. melanogaster

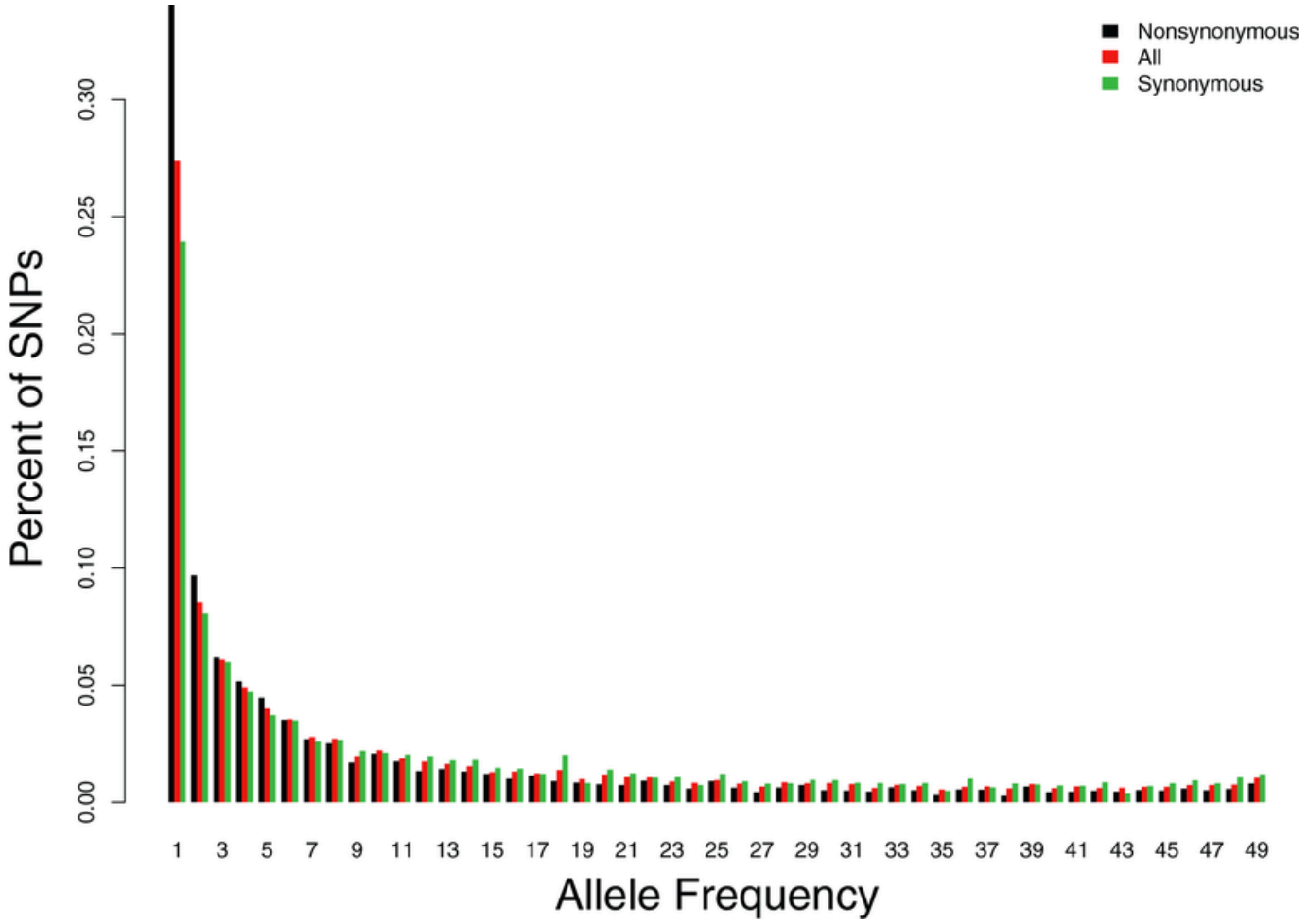


D. simulans



2.1.1 Measures of Genetic Variability

The Site Frequency Spectrum



- Summary of the frequency of minor (folded SFS) or derived (unfolded SFS) alleles across all genotyped sites
- When the minor or derived allele occurs in only 1 of all genotyped individuals, this is called a singleton; when in two individuals, this is a doubleton, etc…

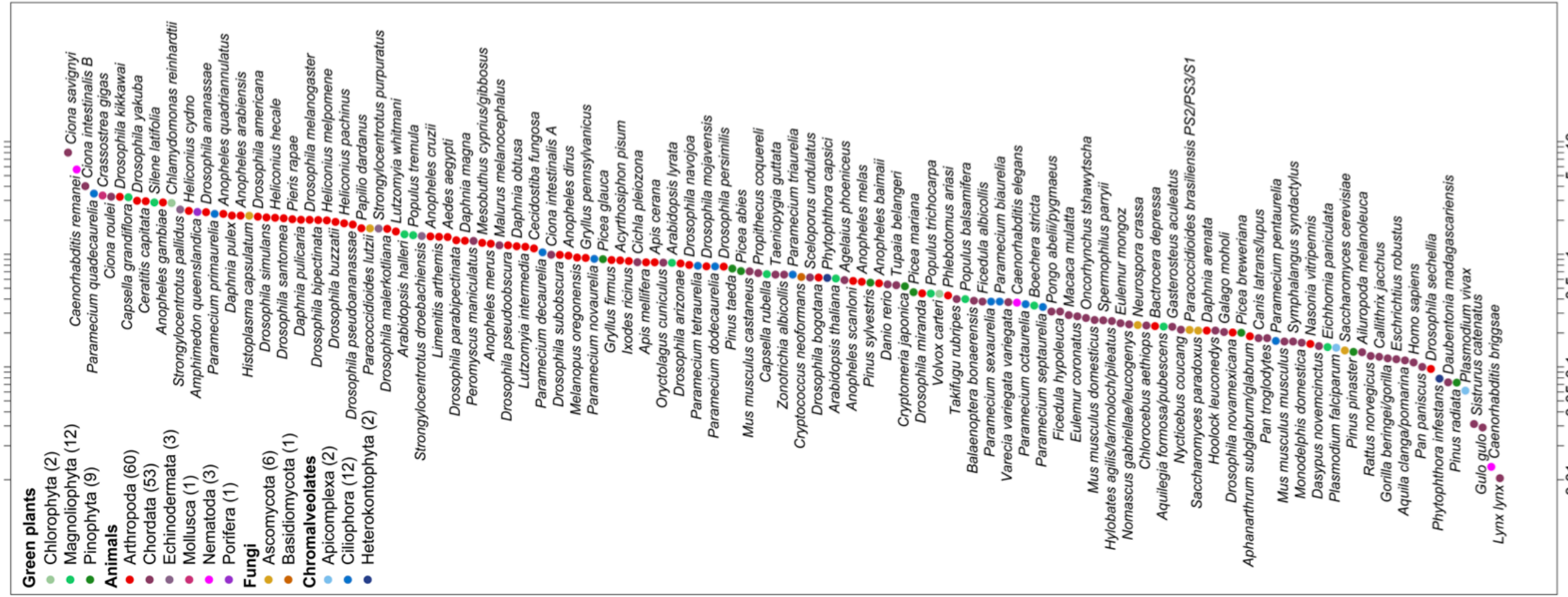
2.1.1 Measures of Genetic Variability

- With measurements across species, population geneticists have been surprised at the relatively small range of genetic variability given the large variation in census size.

- Leffler *et al.* (2012) compiled genetic diversity (π) data for 167 species across 14 phyla

- 800-fold difference in diversity from least (lynx) to most (sea squirt) diverse species; census-size variation is much more dramatic

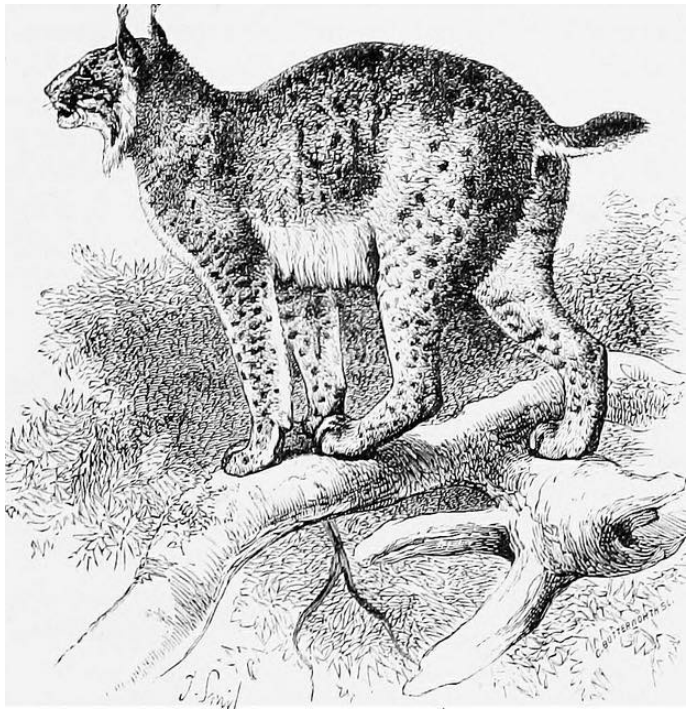
2.1.1 Measures of Genetic Variability



2.1.1 Measures of Genetic Variability

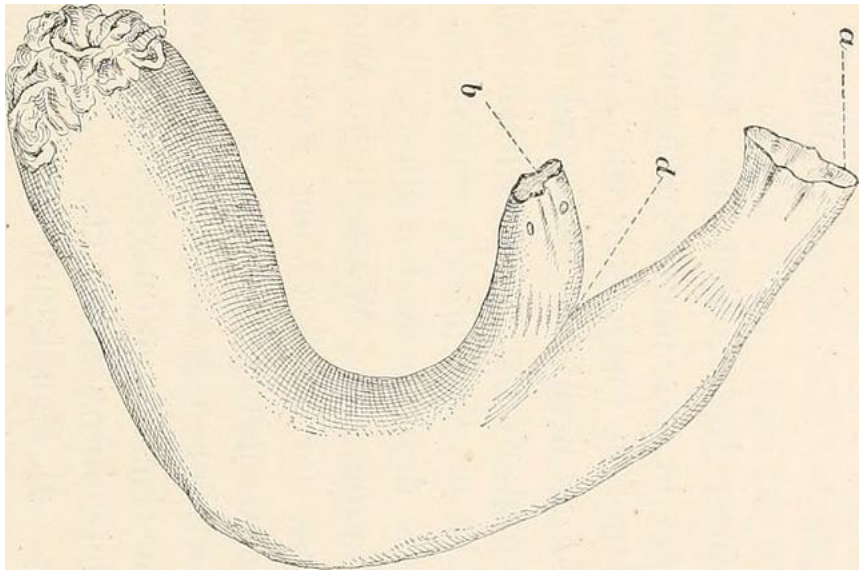
Eurasian Lynx (*Lynx lynx*)

π = 0.01% (1 in 10,000 bases differ)



Sea Squirt (*Ciona savignyi*)

π = 8.3% (1 in 12 bases differ)

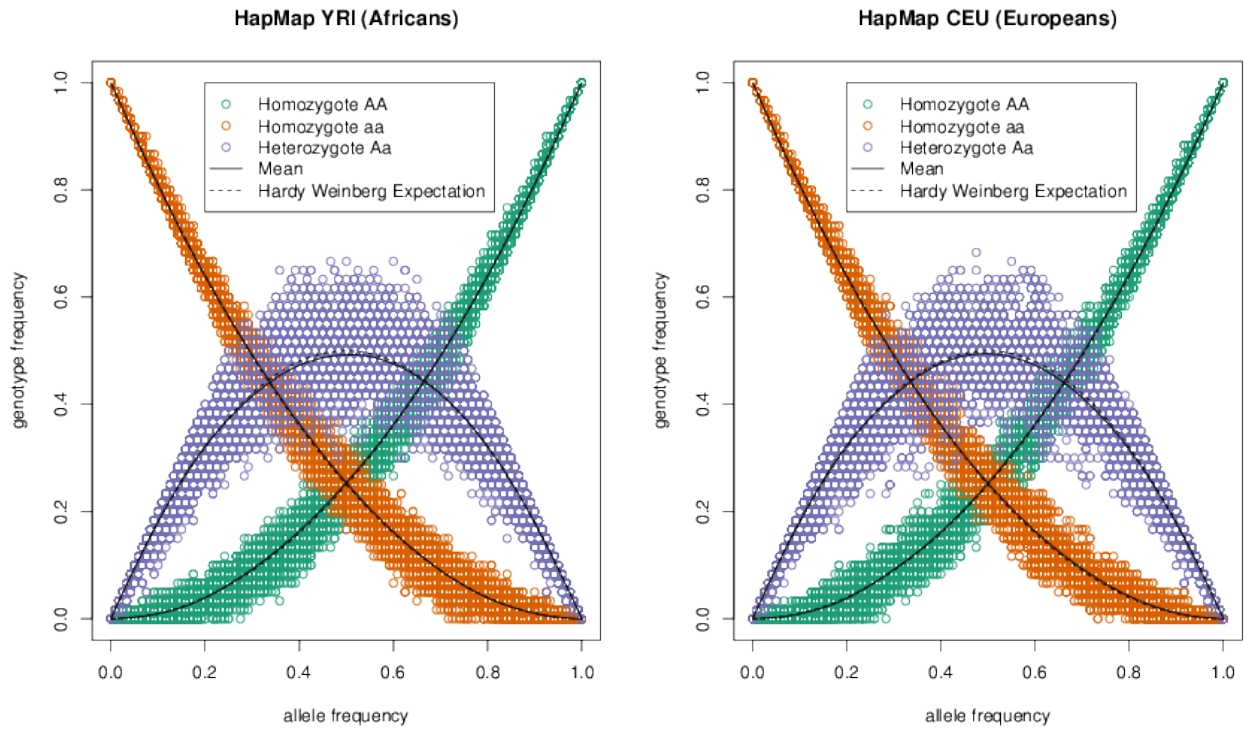


Coop, Chapter 2: 2.1.2-2.1.3

Allele and Genotype Frequencies

Hardy-Weinberg Proportions

Assortative Mating



2.1.2: Hardy-Weinberg Proportions

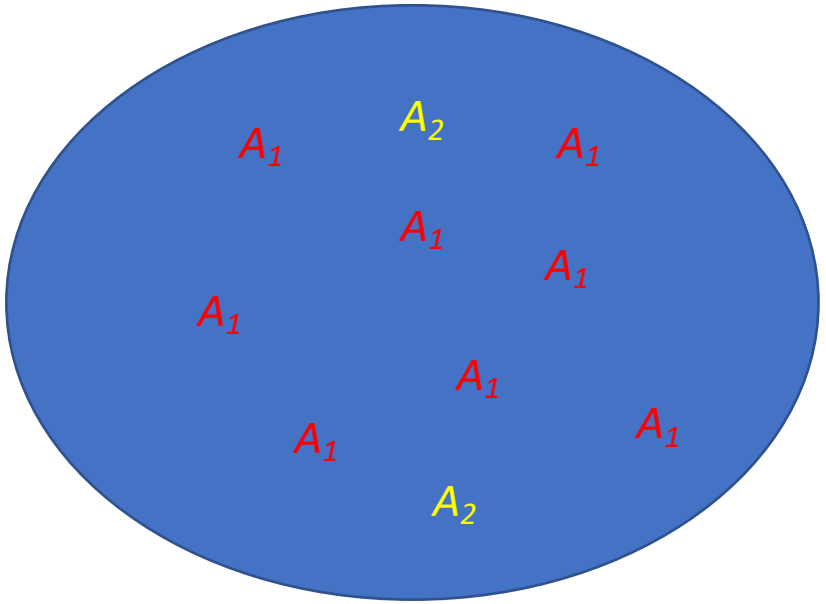
For Hardy-Weinberg proportions we assume:

- Random Mating with respect to genotypes
- No inbreeding
- No assortative mating
- No population structure
- No sex differences in allele frequencies

2.1.2: Hardy-Weinberg Proportions

For example, let's say the frequency of the A_1 allele at the time of reproduction is p:

What is p in this example?

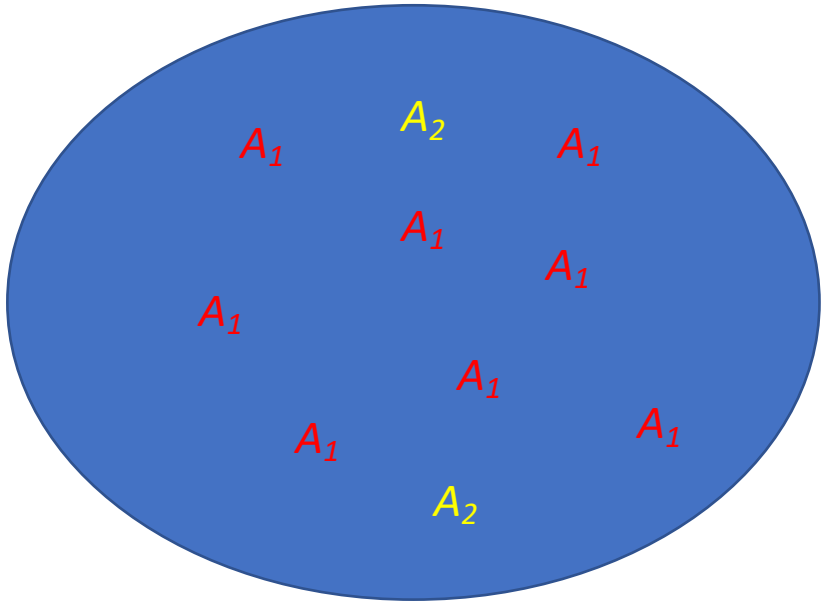


2.1.2: Hardy-Weinberg Proportions

For example, let's say the frequency of the A_1 allele at the time of reproduction is p:

What is p in this example?

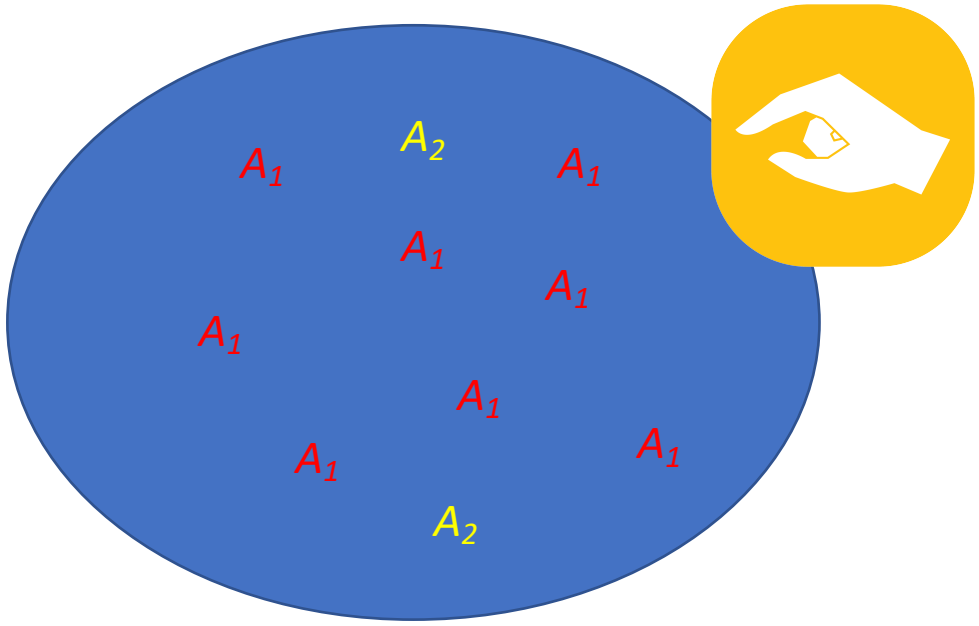
$p = 8/10 = 0.8$



2.1.2: Hardy-Weinberg Proportions

An A_1A_1 diploid genotype is made by making two random A_1 draws from the haploid gamete pool of the population:

What is the probability/expected frequency of an A_1A_1 diploid genotype?

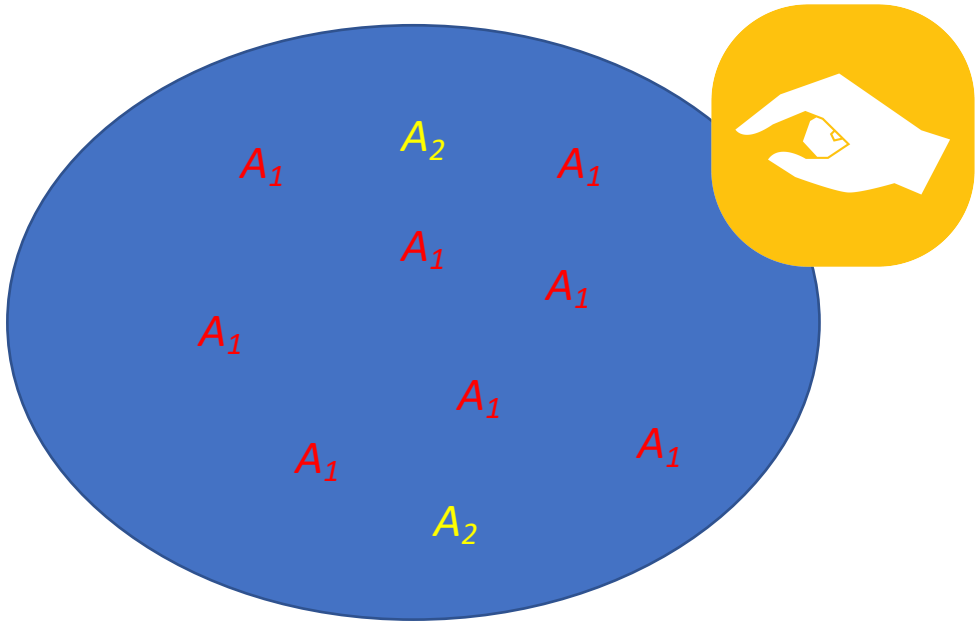


2.1.2: Hardy-Weinberg Proportions

An A_1A_1 diploid genotype is made by making two random A_1 draws from the haploid gamete pool of the population:

What is the probability/expected frequency of an A_1A_1 diploid genotype?

$f_{11} = p^2 = (0.8)^2 = 0.64$



2.1.2: Hardy-Weinberg Proportions

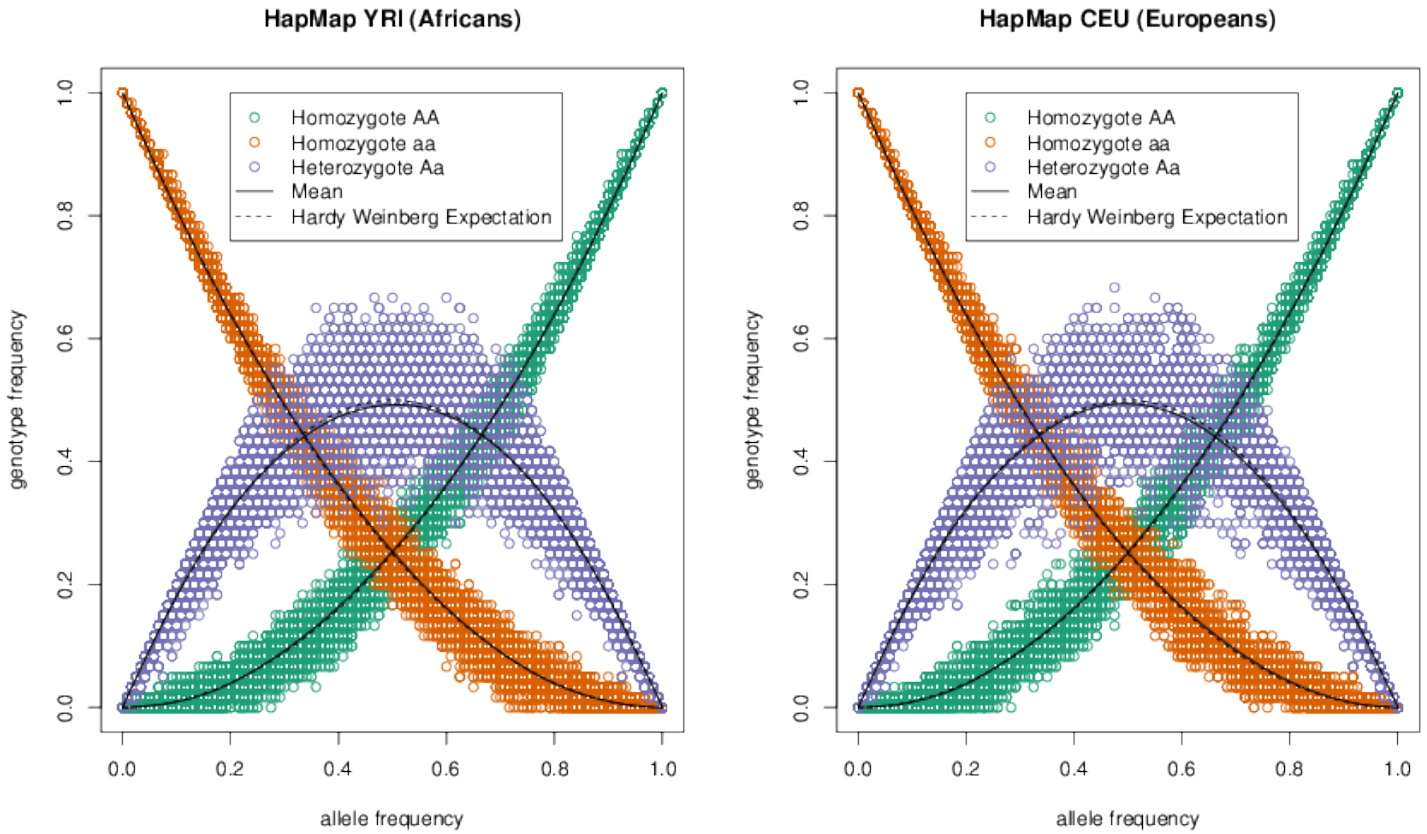
The expected frequency of the three possible diploid genotypes is thus:

f_{11}	f_{12}	f_{22}
p^2	$2pq$	q^2

*Note that selection can change frequencies within a generation, but our expectations hold as long as *p* is the frequency of the A_1 allele when gametes fuse to form a zygote

2.1.2: Hardy-Weinberg Proportions

An example of Hardy-Weinberg proportions in human populations based on 10,000 HapMap SNPs:



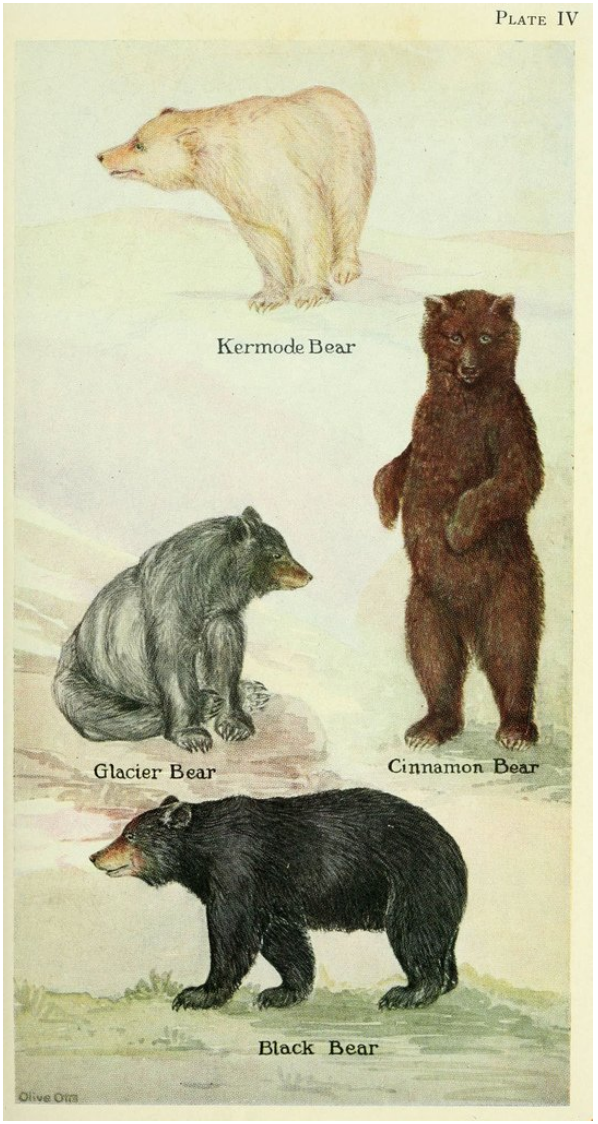
2.1.2: Hardy-Weinberg Proportions

On the coastal islands of British Columbia there is a subspecies of black bear (*Ursus americanus kermodei*, Kermode's bear). Some, called spirit bears, are white. They are homozygotes for a recessive allele at the MC1R gene. Individuals who are GG at this SNP are white while AA and AG individuals are black.

Here are the empirical genotype counts:

AA	AG	GG
42	24	21

What are the expected Hardy-Weinberg frequencies of these genotypes?



2.1.2: Hardy-Weinberg Proportions

What are the expected frequencies of the three genotypes under HWE?

AA	AG	GG
42	24	21

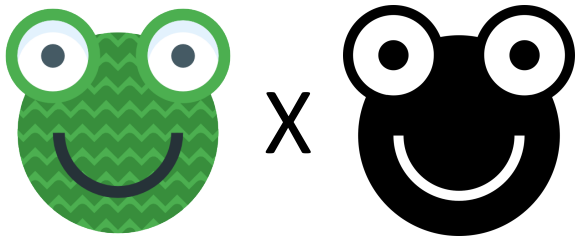
- First, we calculate the total number of individuals: 42+24+21 = 87
- Then, we calculate the allele frequency of the A allele (p): p = 42/87 + .5 * 24/87 = 0.62
- Then the allele frequency of the G allele (q): $q = 1 - p = 0.38$
- Finally, we apply Hardy-Weinberg to calculate expected allele frequencies:

AA	AG	GG
p^2	$2pq$	q^2
0.38	0.47	0.14

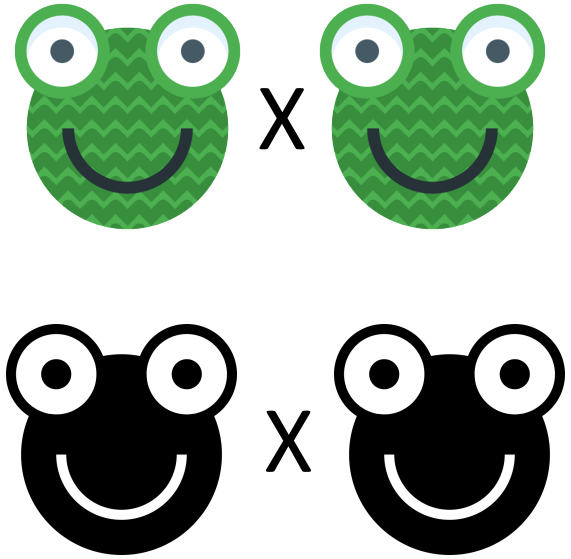
2.1.3: Assortative Mating

Hardy-Weinberg assumptions can be violated in a number of ways. For example, if the probability of mating depends on a phenotype encoded by a particular genotype at a locus:

Negative Assortative, Disassortative Mating



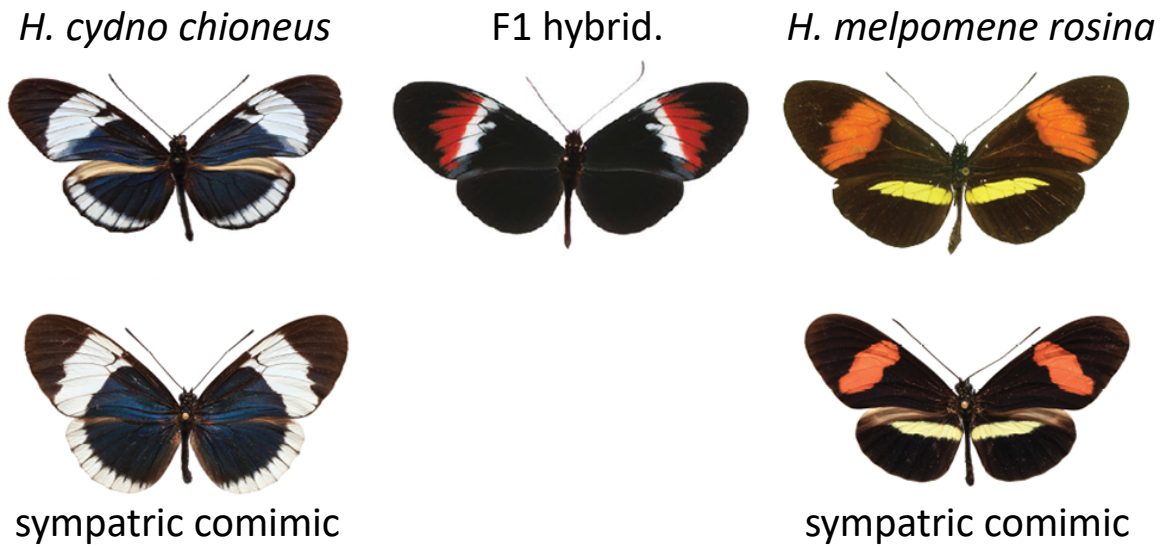
Positive Assortative Mating



2.1.3: Assortative Mating

An example of positive assortative mating

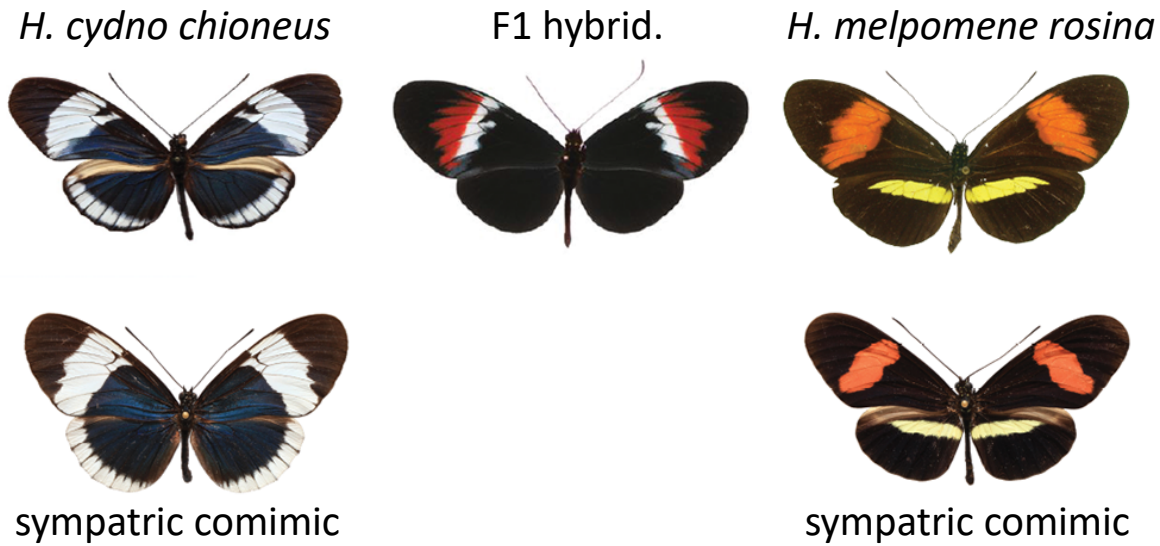
- Wing coloration of *Heliconius* butterflies signals “I’m poisonous, don’t eat me!” to predators
- Different coloration of hybrids relative to other wing coloration morphs results in them being highly predated
- These species show strong positive assortative mating (like with like), presumably due to the selective consequences of mating with a different morph



2.1.3: Assortative Mating

An example of positive assortative mating

- Under positive assortative mating, an excess of homozygotes is observed



2.1.3: Assortative Mating

An example of disassortative mating

- White-throated sparrows have a white-striped and a tan-striped morph (these are not gender differences)

- Mating pairs show strong enrichment for consisting of different morphs (1099 of 1116 pairs)

2.1.3: Assortative Mating

An example of disassortative mating

- Under disassortative mating, an excess of heterozygotes is observed

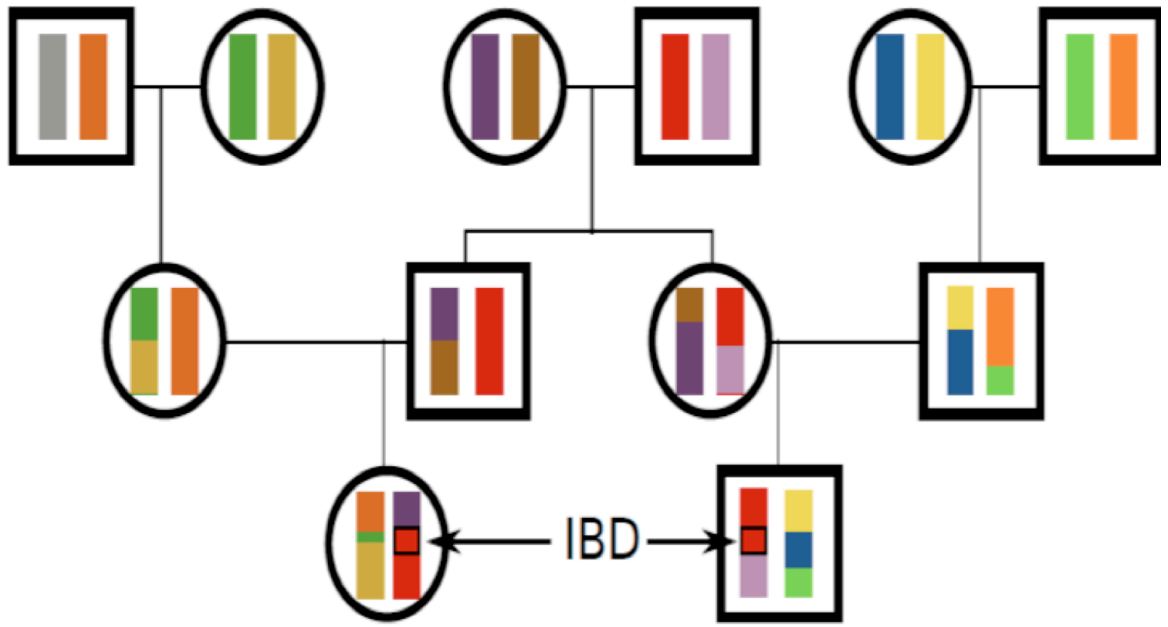
- Alleles in this empirical example show expected heterozygote enrichment and lack of homozygotes:

Tan	White	(Super)White
2/2	2/2m	2m/2m
978	1011	3

Coop, Chapter 2: 2.2

Allele and Genotype Frequencies

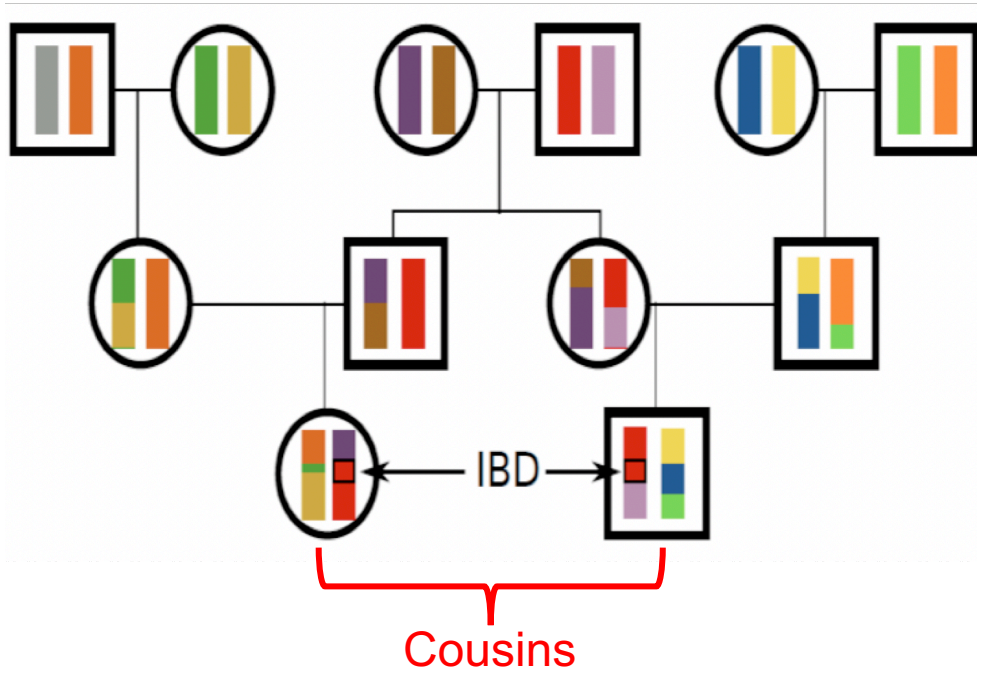
Allele sharing among related individuals and Identity by Descent



2.2: Allele Sharing, Identity by Descent (IBD)

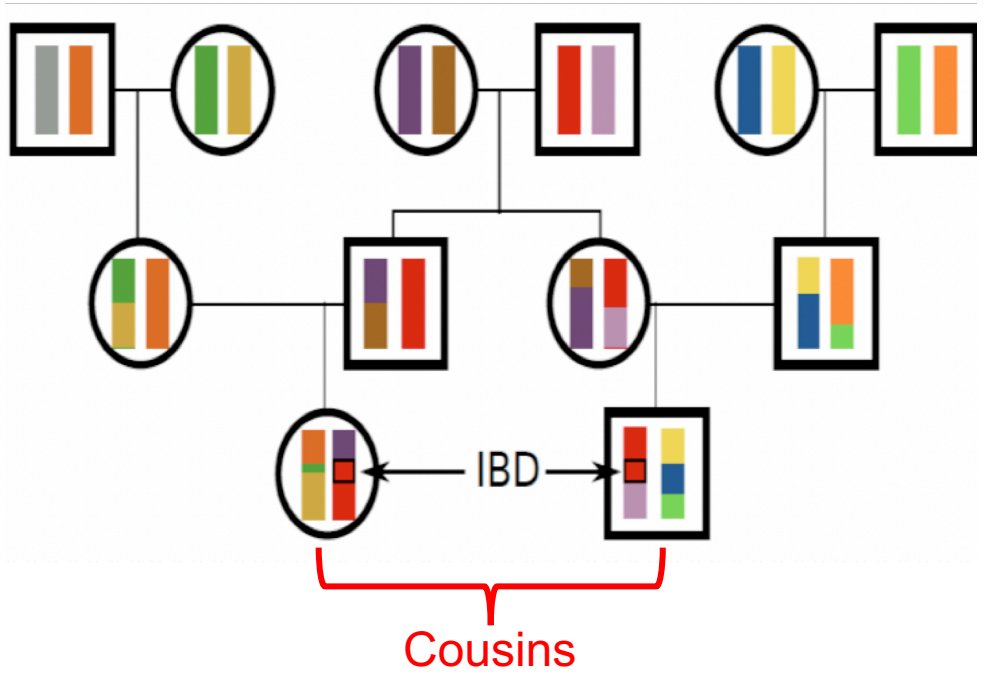
- All individuals in a population are related through a giant pedigree (*i.e.*, a family tree)

- While relatedness varies across pairs of individuals, they all show some level of *kinship*



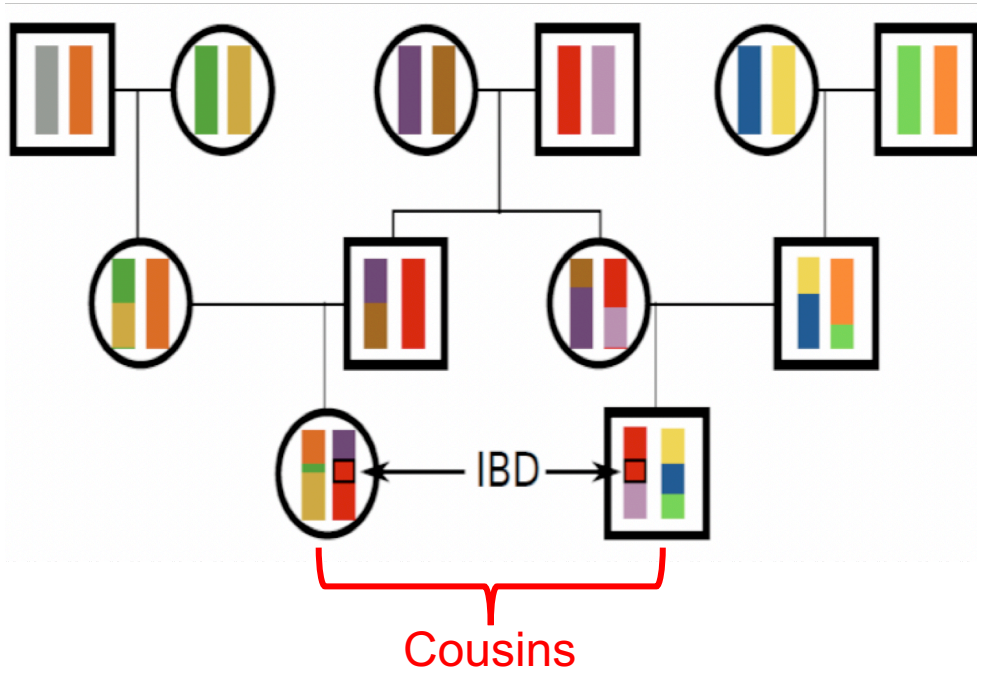
2.2: Allele Sharing, Identity by Descent (IBD)

- Related individuals share alleles from their common ancestor, with close relatives sharing more alleles due to separation by fewer meioses
- Many pop/quant genetics theories rely on how closely related individuals are--we need to know kinship



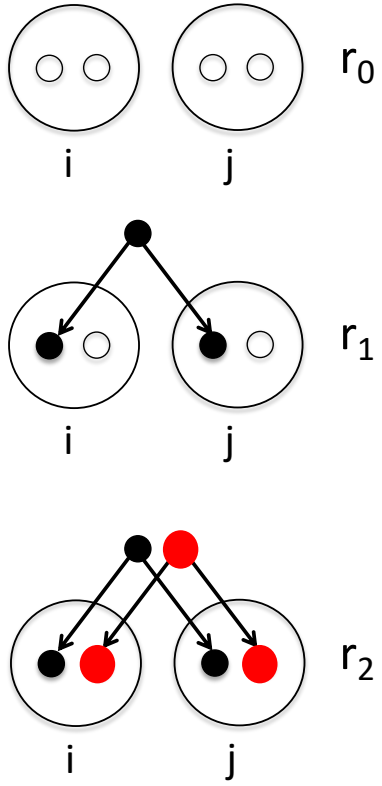
2.2: Allele Sharing, Identity by Descent (IBD)

- Definition: Alleles are Identical by Descent (IBD) if they are identical due to transmission from the same ancestor within the last few generations
- For example, parent and child share exactly 1 allele IBD



2.2: Allele Sharing, Identity by Descent (IBD)

- We can summarize how related two individual are by calculating the probability that they have IBD for 0, 1, or 2 alleles: ***r_0, r_1, r_2***
- These values can also be summarized as genome-wide averages (*e.g.*, a quarter of all loci in full siblings have zero alleles that are IBD; $r_0 = 0.25$)

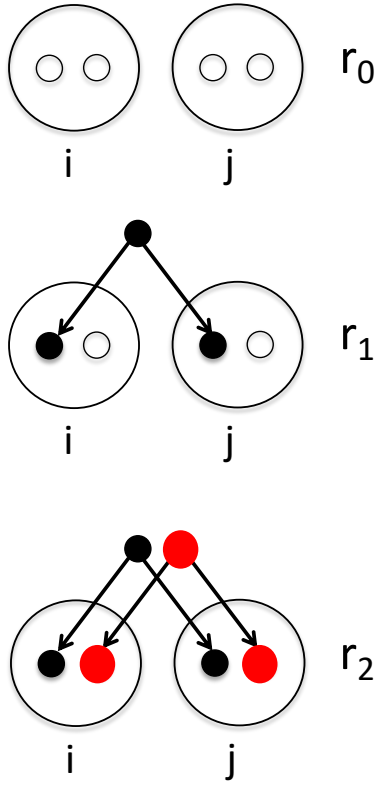


2.2: Allele Sharing, Identity by Descent (IBD)

- One important summary is the *coefficient of kinship* (F_{ij}) which is the probability that two alleles drawn at random (*I* & *J*) from two individuals (*i* & *j*) are IBD

- Mathematically, this can be expressed as:

$$
\begin{aligned}
F_{ij} =& P(\text{I\&J IBD}) && (2.3)\\
=& P(\text{I\&J IBD}|\ \text{i\&j 0 IBD})P(\text{i\&j 0 IBD})\\
&+ P(\text{I\&J IBD}|\ \text{i\&j 1 IBD})P(\text{i\&j 1 IBD})\\
&+ P(\text{I\&J IBD}|\ \text{i\&j 2 IBD})P(\text{i\&j 2 IBD}) && (2.4)\\
=& 0 \times r_0 + \frac{1}{4}r_1 + \frac{1}{2}r_2. && (2.5)
\end{aligned}
$$



2.2: Allele Sharing, Identity by Descent (IBD)

- In 2.4 we sum the conditional probabilities that alleles *I* and *J* are IBD over whether parents *i* and *j* have 0, 1, or 2 alleles that are IBD
- In 2.5 we've taken advantage of the fact that we can calculate these conditional probabilities based on rules of Mendelian transmission

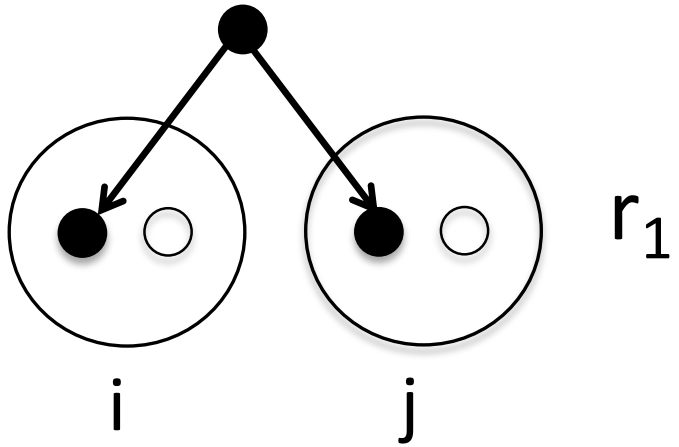
$$
\begin{aligned}
F_{ij} =& P(\text{I\&J IBD}) && (2.3)\\
=& P(\text{I\&J IBD}|\ \text{i\&j 0 IBD})P(\text{i\&j 0 IBD})\\
&+ P(\text{I\&J IBD}|\ \text{i\&j 1 IBD})P(\text{i\&j 1 IBD})\\
&+ P(\text{I\&J IBD}|\ \text{i\&j 2 IBD})P(\text{i\&j 2 IBD}) && (2.4)\\
=& 0 \times r_0 + \frac{1}{4} r_1 + \frac{1}{2} r_2. && (2.5)
\end{aligned}
$$

2.2: Allele Sharing, Identity by Descent (IBD)

$$P(\text{I\&J IBD}|\ \text{i\&j 1 IBD})$$

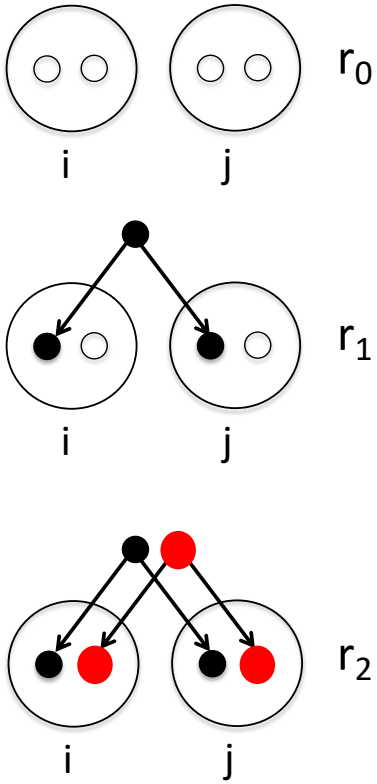
- The pair of alleles (*I* & *J*) drawn from individuals *i* and *j* are IBD, given these individuals share 1 allele that is IBD

- This probability is ¼ because we need to draw the IBD allele twice (once from each individual *i* and *j*): ½ x ½ = ¼



½ x ½ = ¼

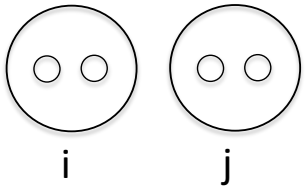
2.2: Allele Sharing, Identity by Descent (IBD)



When we know that 0 alleles are IBD, there is no way we can draw IBD alleles from two individuals

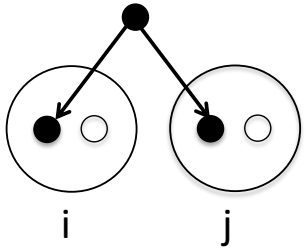
$$=0 \times r_0 + \frac{1}{4} r_1 + \frac{1}{2} r_2. \qquad (2.5)$$

2.2: Allele Sharing, Identity by Descent (IBD)



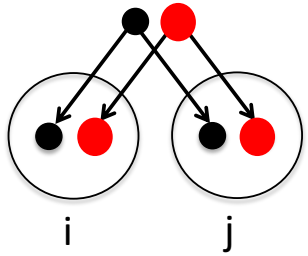
r_0

When we know that 0 alleles are IBD, there is no way we can draw IBD alleles from two individuals



r_1

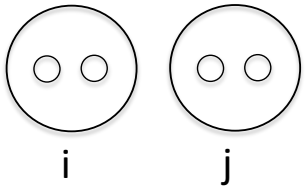
When we know that 1 allele is IBD, we have to draw it twice (from both individuals), each time with probability of ½: ½ × ½ = ¼



r_2

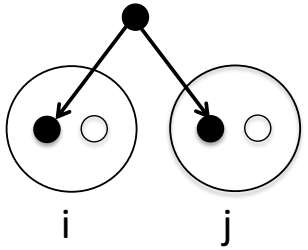
$$=0 \times r_0 + \frac{1}{4} r_1 + \frac{1}{2} r_2. \qquad (2.5)$$

2.2: Allele Sharing, Identity by Descent (IBD)



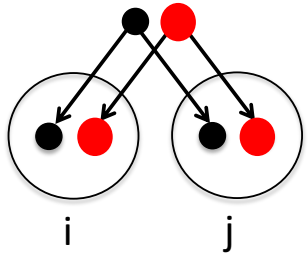
r_0

When we know that 0 alleles are IBD, there is no way we can draw IBD alleles from two individuals



r_1

When we know that 1 allele is IBD, we have to draw it twice (from both individuals), each time with probability of ½: ½ × ½ = ¼



r_2

When we know that 2 distinct alleles (black and red here) are IBD, we can draw an IBD allele in two ways, black with probability ¼ or red with probability of ¼: ¼ + ¼ = ½

$$=0 \times r_0 + \frac{1}{4} r_1 + \frac{1}{2} r_2. \qquad (2.5)$$

2.2: Allele Sharing, Identity by Descent (IBD)

- Based on equation 2.5 below which incorporates Mendelian transmission, convince yourself that the coefficient of kinship for first cousins is 1/16:

Relationship (i,j)*	P(i&j 0 IBD)	P(i&j 1 IBD)	P(i&j 2 IBD)	P(I&J IBD)
Relationship (i,j)*	r_0	r_1	r_2	F_{ij}
parent–child	0	1	0	$1/4$
full siblings	$1/4$	$1/2$	$1/4$	$1/4$
Monozygotic twins	0	0	1	$1/2$
1^{st} cousins	$3/4$	$1/4$	0	$1/16$

$$
\begin{aligned}
F_{ij} &= P(\text{I\&J IBD}) && (2.3)\\
&= P(\text{I\&J IBD}|\ \text{i\&j 0 IBD})P(\text{i\&j 0 IBD})\\
&\quad + P(\text{I\&J IBD}|\ \text{i\&j 1 IBD})P(\text{i\&j 1 IBD})\\
&\quad + P(\text{I\&J IBD}|\ \text{i\&j 2 IBD})P(\text{i\&j 2 IBD}) && (2.4)\\
&= 0 \times r_0 + \frac{1}{4} r_1 + \frac{1}{2} r_2. && (2.5)
\end{aligned}
$$

2.2: Allele Sharing, Identity by Descent (IBD)

- Our *r* coefficients have more uses. For example, we can calculate the probability of a homozygous genotype A_1A_1 when we know both the allele frequency and the coefficient of inbreeding

$$\begin{aligned} P(A_1A_1) =& P(A_1A_1|0 \text{ alleles IBD})P(0 \text{ alleles IBD}) \\ &+ P(A_1A_1|1 \text{ allele IBD})P(1 \text{ allele IBD}) \\ &+ P(A_1A_1|2 \text{ alleles IBD})P(2 \text{ alleles IBD}) \end{aligned} \tag{2.6}$$

Or, in our r_0, r_1, r_2 notation:

$$\begin{aligned} P(A_1A_1) =& P(A_1A_1|0 \text{ alleles IBD})r_0 \\ &+ P(A_1A_1|1 \text{ alleles IBD})r_1 \\ &+ P(A_1A_1|2 \text{ alleles IBD})r_2 \end{aligned} \tag{2.7}$$

2.2: Allele Sharing, Identity by Descent (IBD)

- If our individuals share 0 alleles IBD, then probability that they are A_1A_1 is:

$$P(A_1A_1|0 \text{ alleles IBD}) = p^2 \times p^2$$

- If our individuals share 1 allele IBD, then shared allele is of type A_1 with probability p, and other, non-IBD allele is A_1 with probability p^2

$$P(A_1A_1|1 \text{ alleles IBD}) = p \times p^2$$

- If our individuals share 2 allele IBD, then shared allele is of type A_1 with probability p^2, because if one individual is homozygous A_1, both are

$$P(A_1A_1|2 \text{ alleles IBD}) = p^2$$

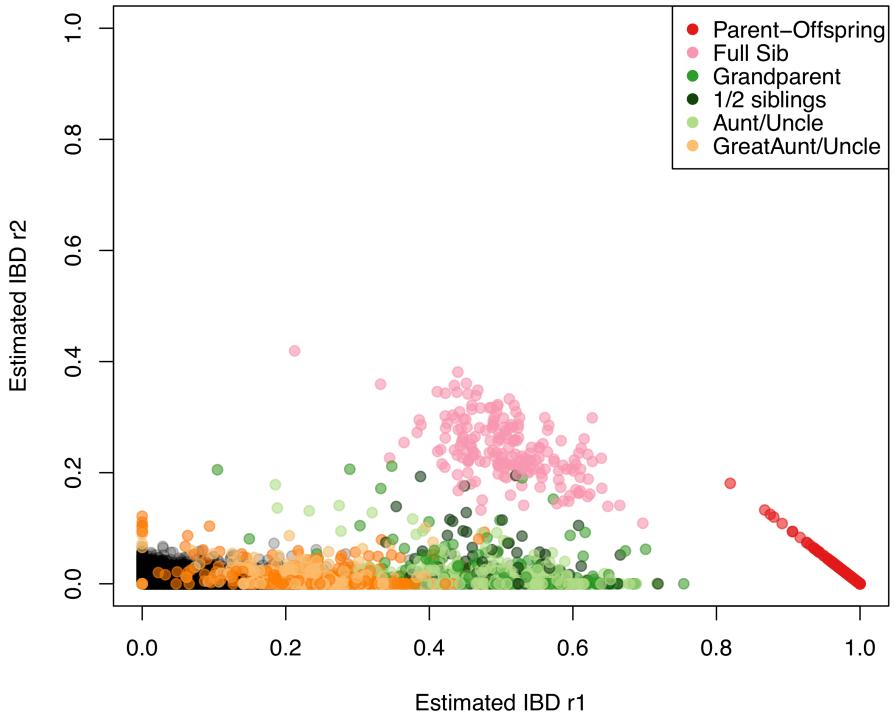
2.2: Allele Sharing, Identity by Descent (IBD)

$$\begin{aligned} P(A_1A_1) =& P(A_1A_1|0 \text{ alleles IBD})r_0 \\ &+ P(A_1A_1|1 \text{ alleles IBD})r_1 \\ &+ P(A_1A_1|2 \text{ alleles IBD})r_2 \end{aligned} \tag{2.7}$$

simplifies to:

$$P(A_1A_1) = p^4r_0 + p^3r_1 + p^2r_2 \tag{2.8}$$

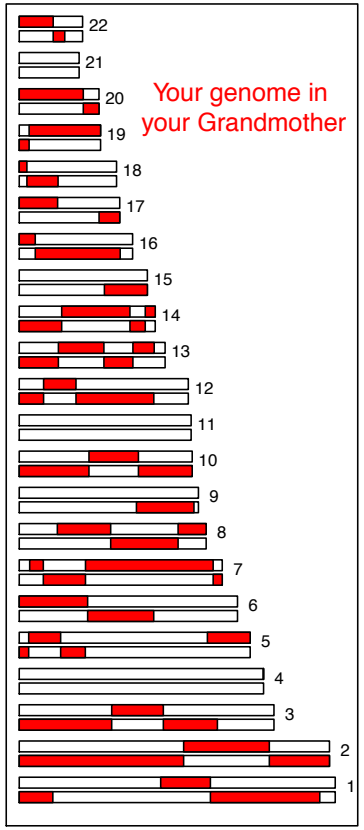
2.2: Allele Sharing, Identity by Descent (IBD)

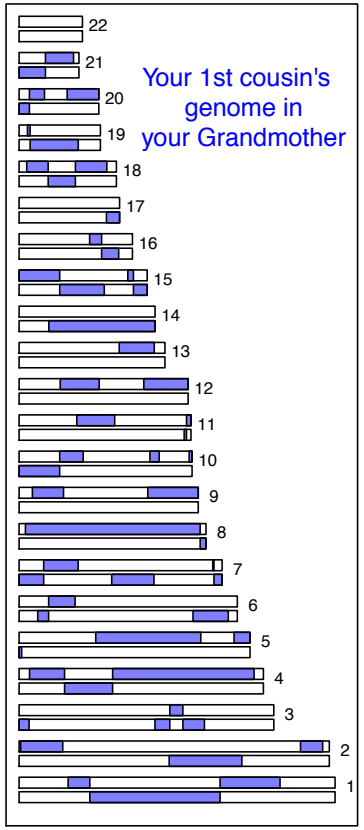


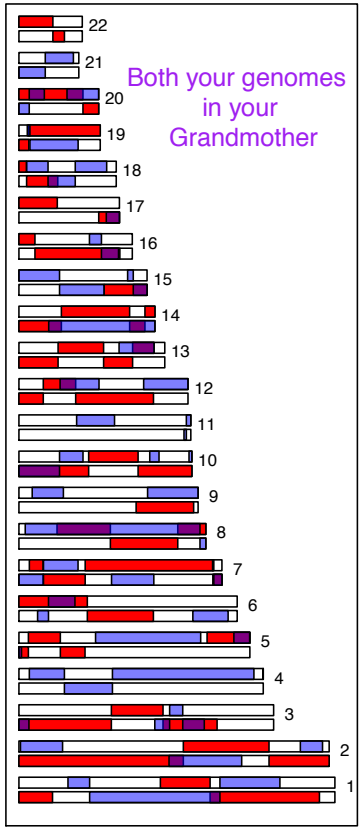
- In a scrub jay population, pedigree information has been kept for decades.
- Estimates of r_1 and r_2 match theoretical predictions and pedigree well

Relationship (i,j)*	P(i&j 0 IBD)	P(i&j 1 IBD)	P(i&j 2 IBD)	P(I&J IBD)
Relationship (i,j)*	r_0	r_1	r_2	F_{ij}
parent–child	0	1	0	$1/4$
full siblings	$1/4$	$1/2$	$1/4$	$1/4$
Monozygotic twins	0	0	1	$1/2$
1^{st} cousins	$3/4$	$1/4$	0	$1/16$

2.2: Allele Sharing, Identity by Descent (IBD)

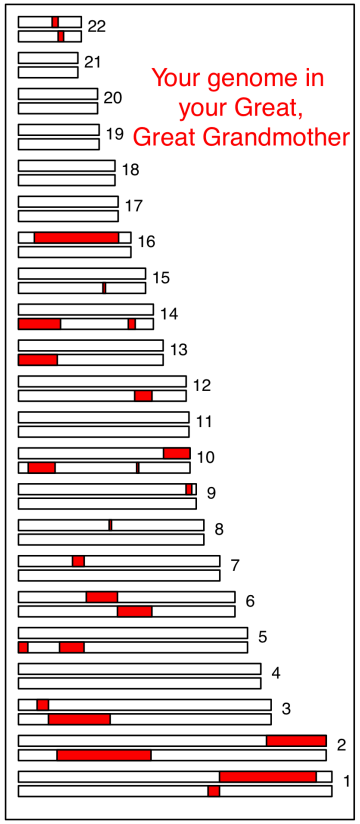


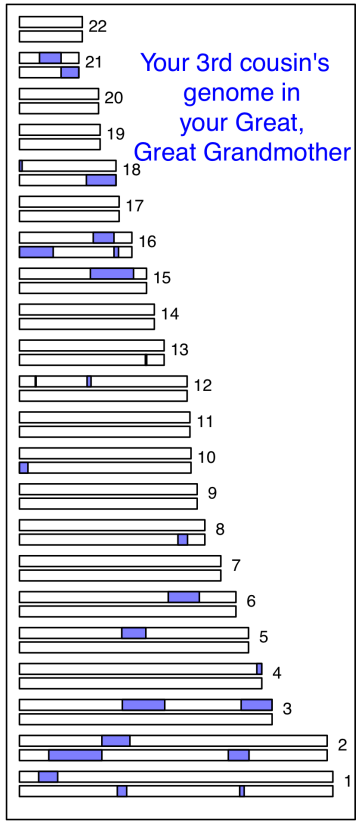


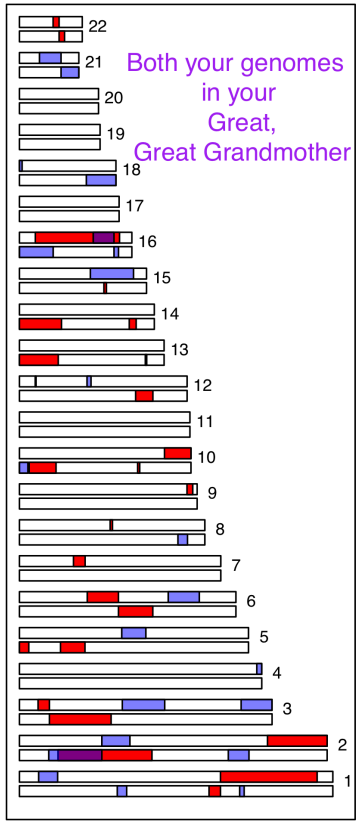


- Plot of expected chromosome sharing of two cousins with their shared grandmother
- Purple regions are IBD
- Companies like 23&me look for such blocks of IBD and use their cumulative extent and length to predict relatedness of individuals

2.2: Allele Sharing, Identity by Descent (IBD)





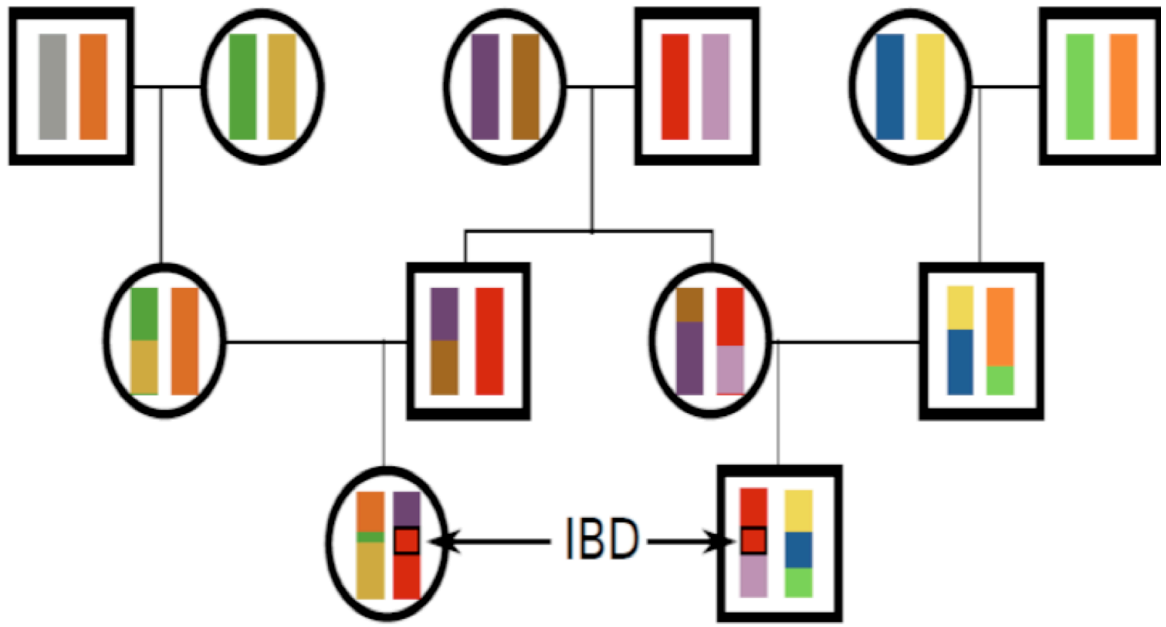


- With 3rd cousins, there is less IBD from the shared great, great grandmother
- IBD blocks are also shorter due to the increased number of meioses
- Beyond this level of relatedness, IBD can be difficult to detect due to short block lengths

Coop, Chapter 2: 2.2.1

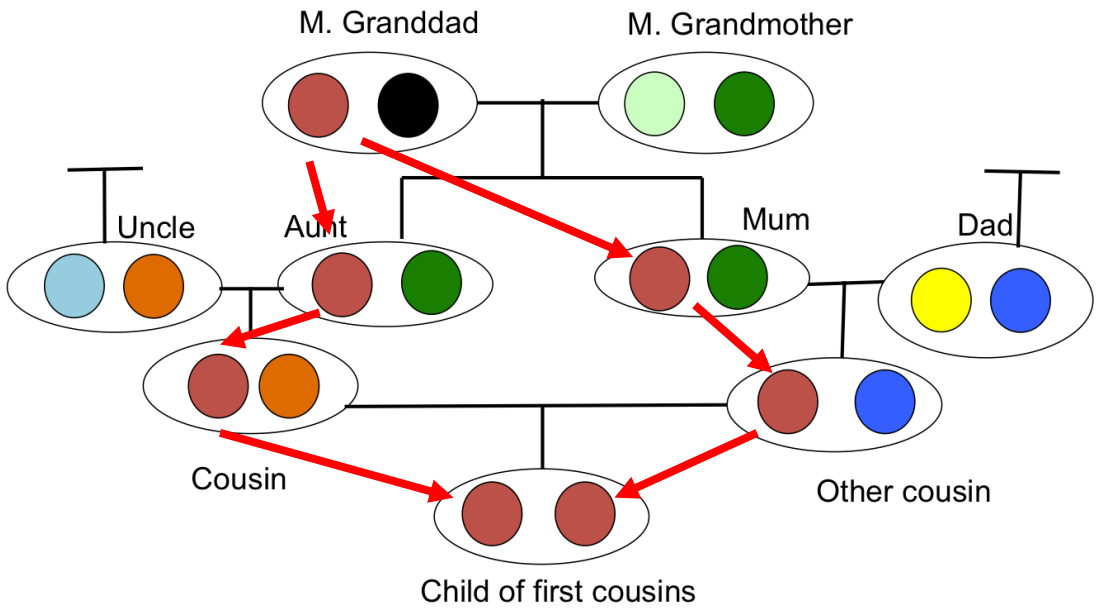
Allele and Genotype Frequencies

Inbreeding



2.2.1: Inbreeding

- Inbred: an individual whose parents are more closely related to each other than two random individuals drawn from a population
- Inbred individuals can have 2 alleles that are IBD, homozygous by descent through 2 paths in the pedigree
- Increased likelihood of being homozygous due to non-random mating → inbreeding



2.2.1: Inbreeding

- Remember, the probability of two alleles being IBD is the coefficient of kinship

$$F_{ij} = 0 \times r_0 + \frac{1}{4} r_1 + \frac{1}{2} r_2.$$

- The only way an individual can be heterozygous is to not be IBD (1 - *F*)

$$\frac{f_{12}}{(1-F)2pq}$$

2.2.1: Inbreeding

An individual could be an A_1A_1 homozygote in one of two ways:

- They have two A_1 alleles that are not IBD but happen to both be A_1

$$(1-F)p^2$$

- The two alleles are IBD

$$Fp.$$

An individual could be an A_2A_2 homozygote in one of two ways:

- They have two A_2 alleles that are not IBD but happen to both be A_2

$$(1-F)q^2$$

- The two alleles are IBD

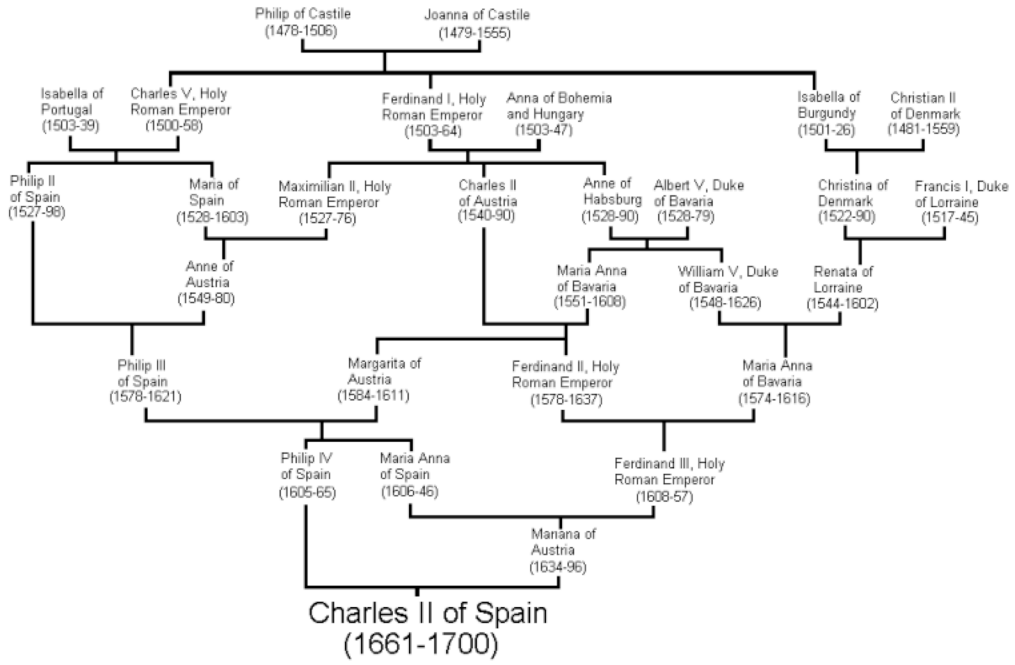
$$Fq$$

2.2.1: Inbreeding

- This leads to the generalized Hardy-Weinberg Form with Inbreeding:

f_{11}	f_{12}	f_{22}
$(1-F)p^2 + Fp$	$(1-F)2pq$	$(1-F)q^2 + Fq$

2.2.1: Inbreeding

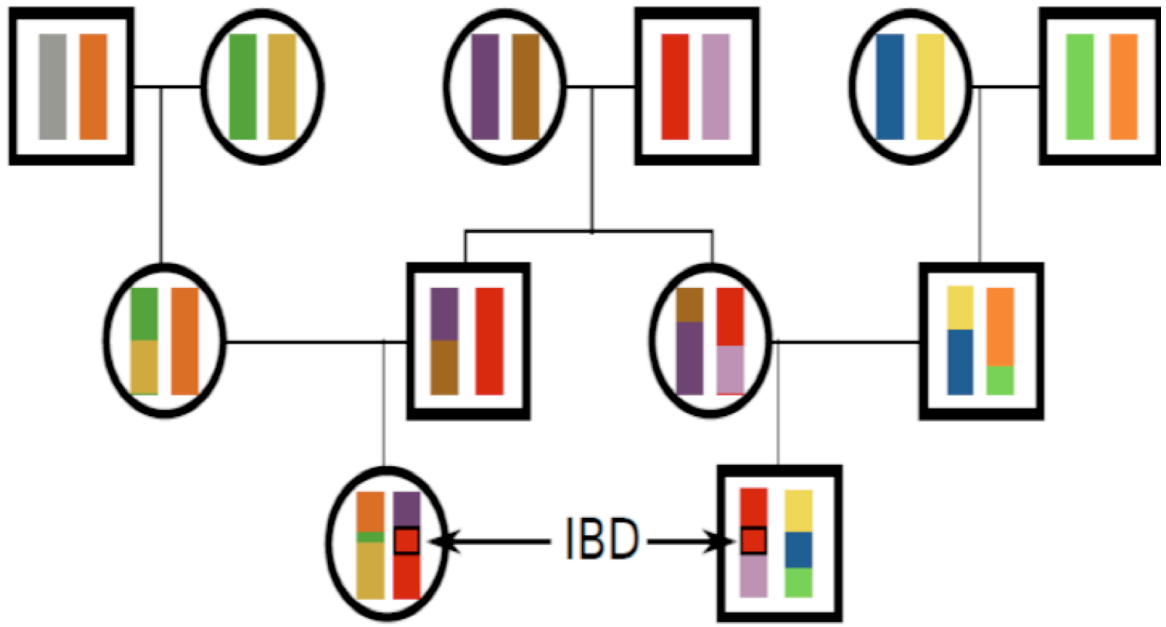


- Up until now, we've talked about single loops of inbreeding within a pedigree, but there can often be many more
- More loops increases probability of IBD and inbreeding
- Charles II of Spain is a classic example: his inbreeding coefficient was calculated at 0.254, the equivalent of full sib mating
- Expression of recessive disease alleles might have explained Charles II's poor health

Coop, Chapter 2: 2.2.2

Allele and Genotype Frequencies

Calculating inbreeding coefficients from genetic data



2.2.2 Calculating inbreeding coefficients from genetic data

Assuming Hardy-Weinberg…

$$\hat{F} = 1 - \frac{f_{12}}{2pq} = \frac{2pq - f_{12}}{2pq} \qquad (2.12)$$

2.2.2 Calculating inbreeding coefficients from genetic data

Assuming Hardy-Weinberg…

$$\hat{F} = 1 - \frac{f_{12}}{2pq} = \frac{2pq - f_{12}}{2pq} \qquad (2.12)$$

Where…

$\hat{F}$ = estimated inbreeding coefficient

2.2.2 Calculating inbreeding coefficients from genetic data

Assuming Hardy-Weinberg…

$$\hat{F} = 1 - \frac{f_{12}}{2pq} = \frac{2pq - f_{12}}{2pq} \qquad (2.12)$$

Where…

$\hat{F}$ = estimated inbreeding coefficient

f_{12} = observed frequency of heterozygotes (also referred to as H_O)

2.2.2 Calculating inbreeding coefficients from genetic data

Assuming Hardy-Weinberg…

$$\hat{F} = 1 - \frac{f_{12}}{2pq} = \frac{2pq - f_{12}}{2pq} \qquad (2.12)$$

Where…

$\hat{F}$ = estimated inbreeding coefficient

f_{12} = observed frequency of heterozygotes (also referred to as H_O)

2pq = expected frequency of heterozygotes under HWE (H_E)

2.2.2 Calculating inbreeding coefficients from genetic data

Which can be reduced to

$$\hat{F} = \frac{H_E - H_O}{H_E} = 1 - \frac{H_O}{H_E} \qquad (2.13)$$

Where…

H_E= expected frequency of heterozygotes under HWE

H_O= observed frequency of heterozygotes under HWE

2.2.2 Calculating inbreeding coefficients from genetic data

Reminder: What deviation from HWE do we expect when a population is inbred? What do you expect $\hat{F}$ to be when a population is inbred?

$$\hat{F} = \frac{H_E - H_O}{H_E} = 1 - \frac{H_O}{H_E} \qquad (2.13)$$

Important: This equation measures deviation of heterozygote frequency from the expected frequency under random mating (HWE)

2.2.2 Calculating inbreeding coefficients from genetic data

- Genetic markers are commonly used in conservation genetics to gauge inbreeding in species of concern like Mexican wolves
- The deficit of heterozygotes in Mexican wolves suggested substantial inbreeding:

H_E=0.18, H_O=0.12

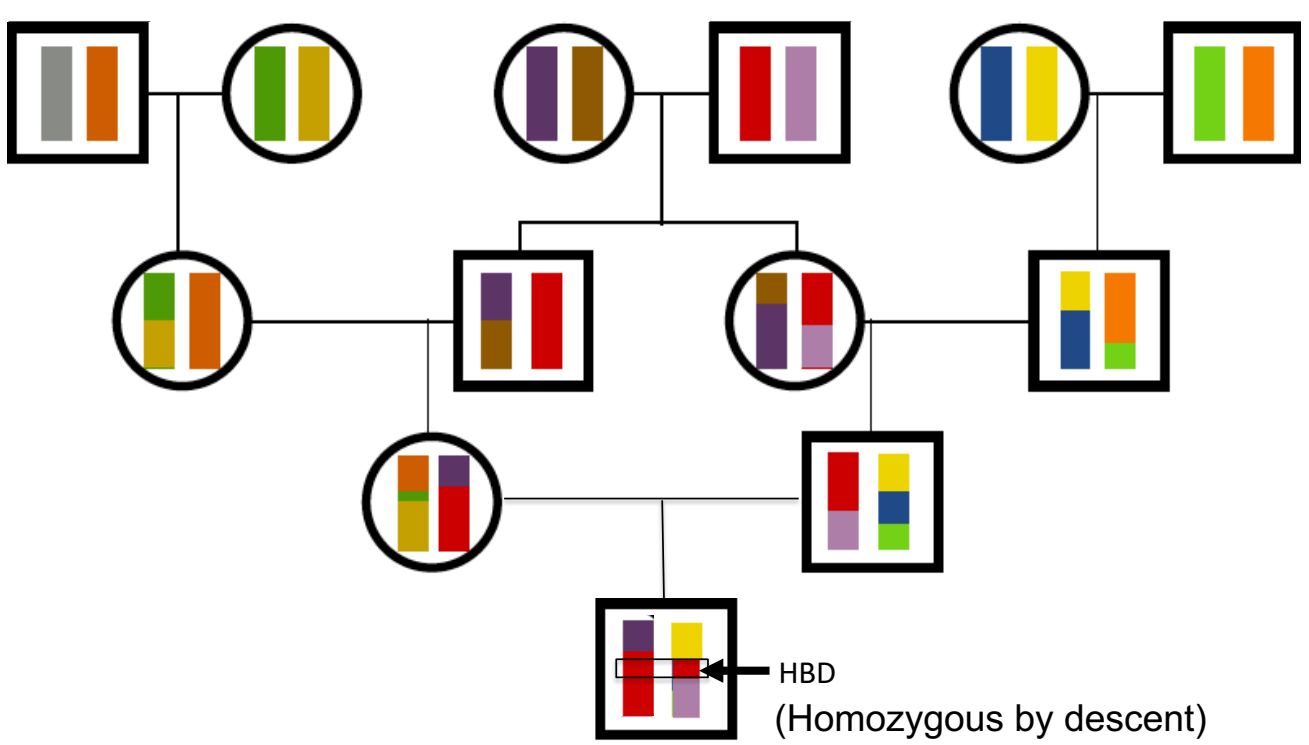
$$\hat{F} = 1 - \frac{H_O}{H_E}$$

$$\hat{F} = 0.33$$

- 33% of Mexican wolves' genome was homozygous due to inbreeding

2.2.2 Calculating inbreeding coefficients from genetic data

Genomic blocks of homozygosity: a product of inbreeding

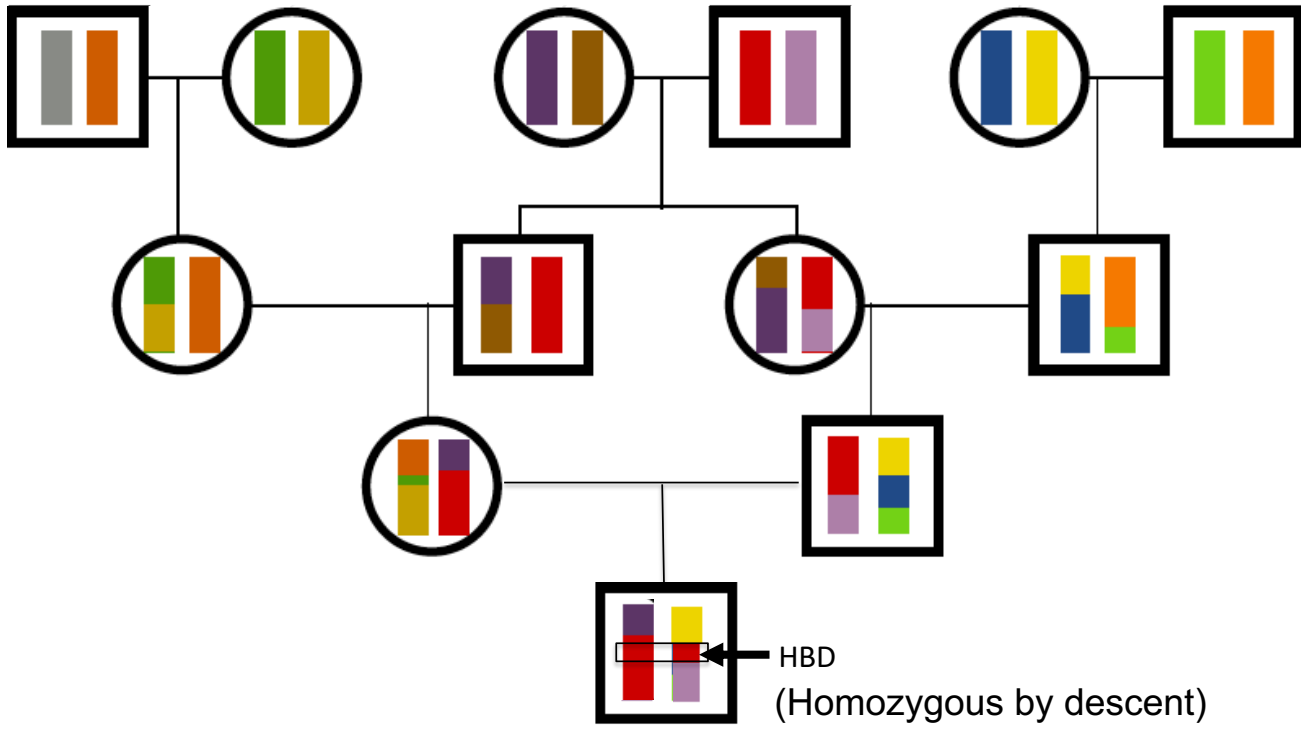


For each progressive generation, there is an additional crossing-over event

Runs of homozygosity (ROH) decrease in size

2.2.2 Calculating inbreeding coefficients from genetic data

Genomic blocks of homozygosity: a product of inbreeding

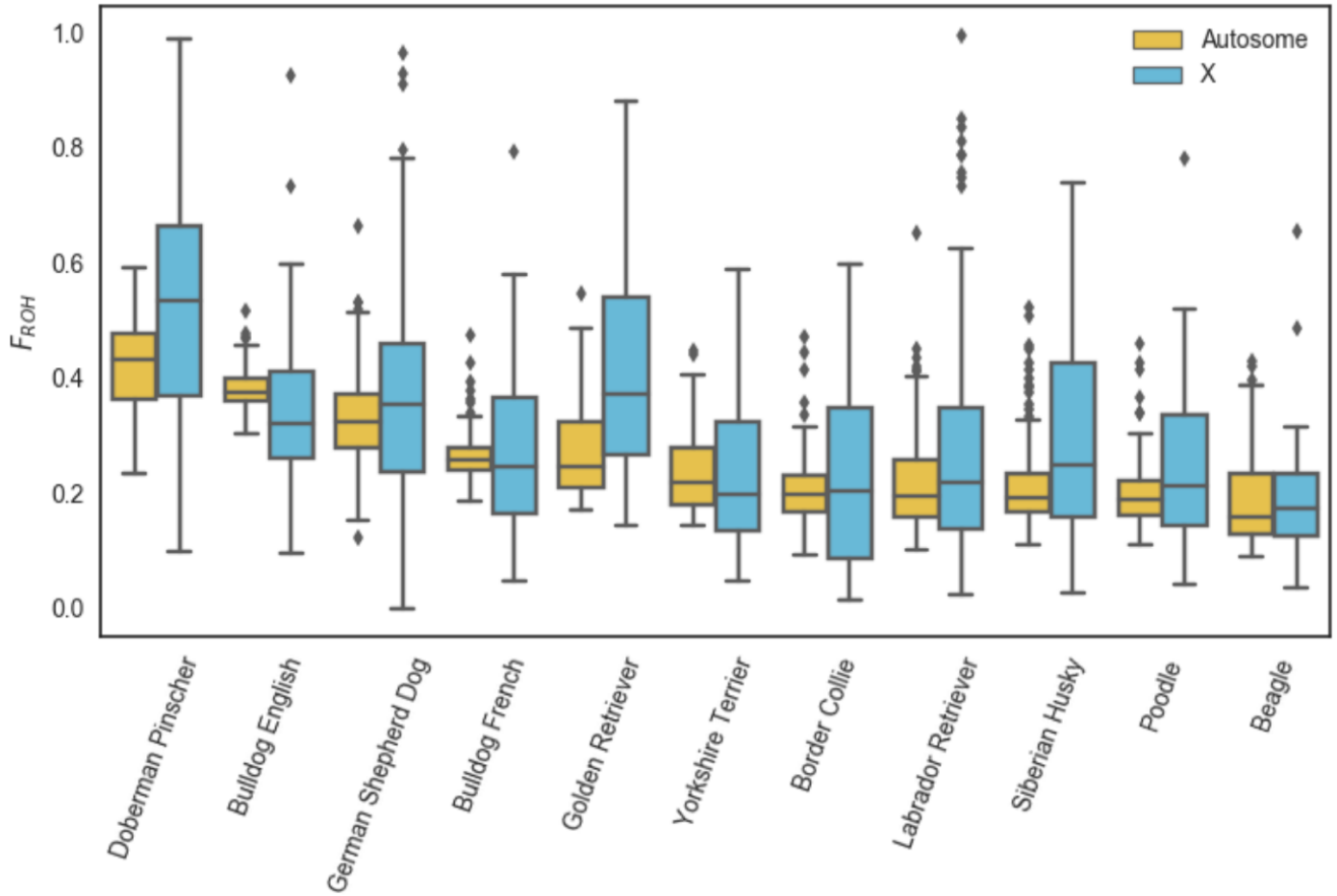


ROH: Runs of homozygosity

Inbreeding increases the frequency of ROH by increasing the likelihood of homozygosity by descent (HBD)

2.2.2 Calculating inbreeding coefficients from genetic data

Genomic blocks of homozygosity: a product of inbreeding



ROH: Runs of homozygosity
HBD: Homozygosity by descent

Measuring the frequency of ROH (F_{ROH}) can tell us about inbreeding

Fig. 2.21

2.2.2 Calculating inbreeding coefficients from genetic data

Genomic blocks of homozygosity: a product of inbreeding

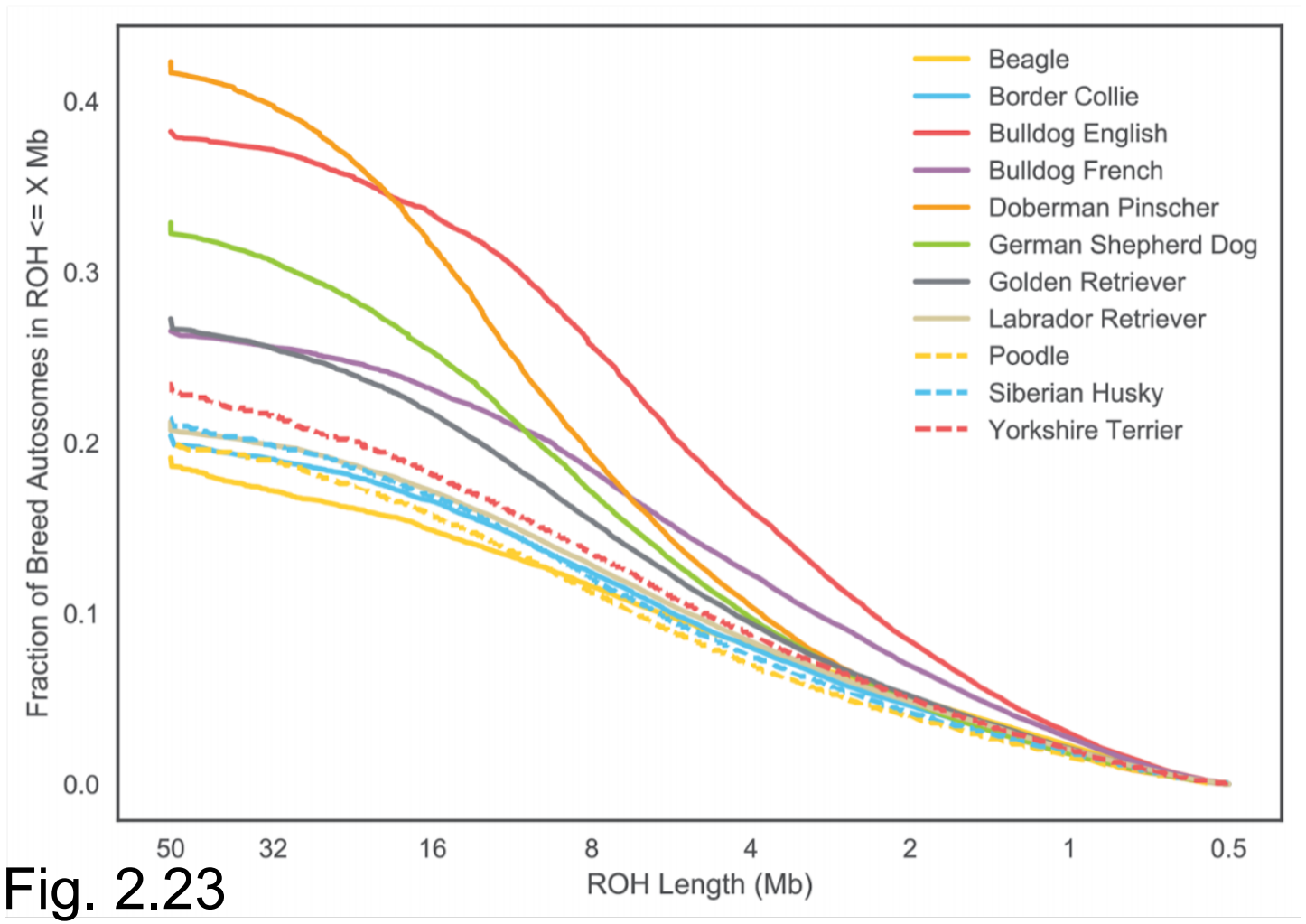


Fig. 2.23

ROH: Runs of homozygosity
HBD: Homozygosity by descent
F_{ROH}: Frequency of ROH

Notice the different shapes of the curves:

- Doberman Pinscher has more large ROH > recent inbreeding
- English Bulldog has more short ROH > older inbreeding

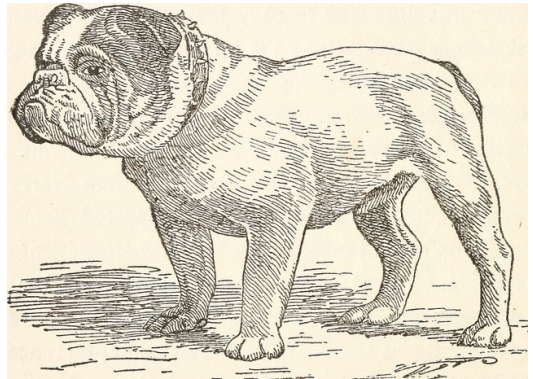


Fig. 2.22