

Today's schedule

Monday 13 February: Genetic diversity

Genetic diversity in great apes

- Variation in quantitative characters (phenotypes)
- Variation at the chromosomal level
- Variation at single genes
- Variation in DNA sequences

13¹⁵-14⁰⁰ (L) Genetic diversity: From phenotype over chromosomes and genes to DNA (HRS)

14¹⁵-16⁰⁰ (E) [Exercise in estimating nucleotide diversity](#)  (Patricia Pečnerová , HRS)

[Here you can get an introduction and comments to the exercise.](#) 

Preparation

Read N&S Chapter 3

Recommended background reading:

See [Prado-Martinez et al. \(2013\)](#) for a comprehensive great ape paper

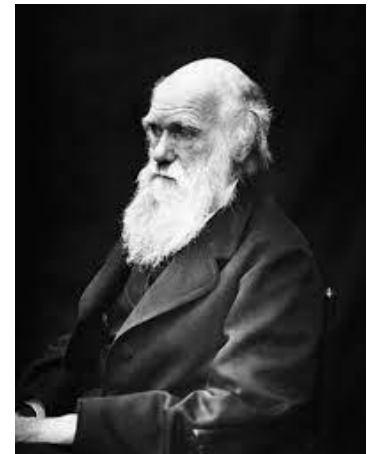
Why is genetic variation important?

Darwin's theory for natural selection

Tendency for geometric growth

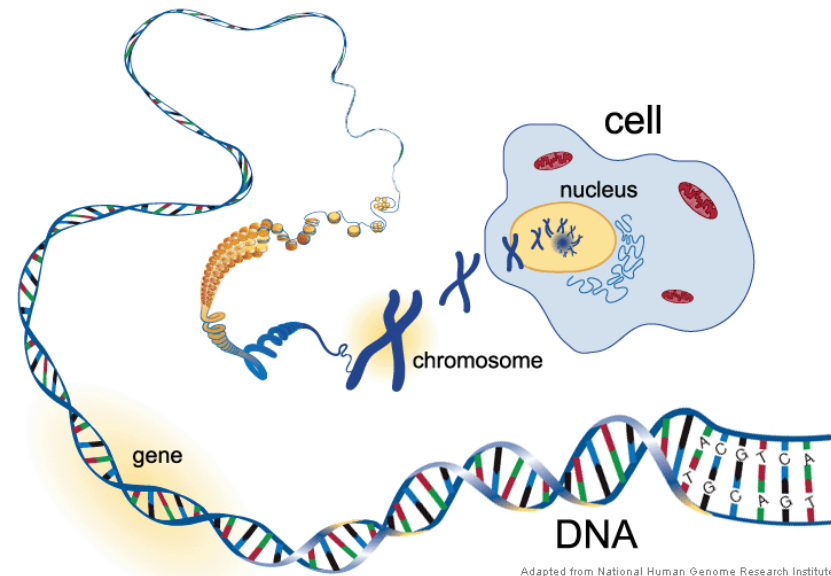
Variation
Inheritance

There is **no** evolution without genetic variation



Genetic diversity: from phenotype over chromosomes, genes to DNA

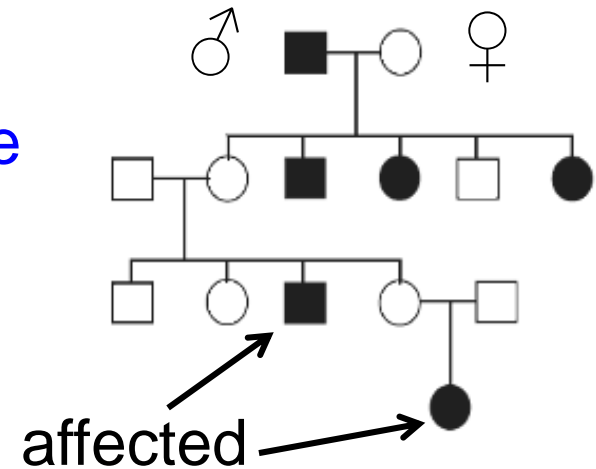
- Variation in quantitative characters (phenotypes)
- Variation at the chromosomal level
- Variation at single genes
 - Polymorphism
 - Gene diversity
 - Enzyme electrophoresis
- Variation in DNA sequences
 - Exercise:
Quantify nucleotide diversity on chromosome 22 in humans and chimpanzees



Characters

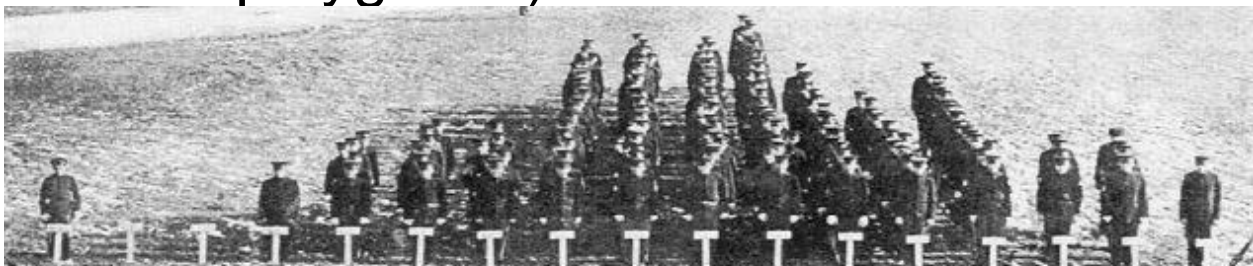
Qualitative characters

genes have **large** effect on **phenotype**
one or few genes affect the trait
(major genes)



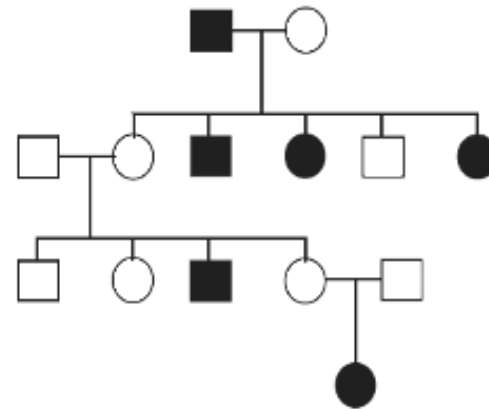
Quantitative characters

genes have **small** effect on **phenotype**
many genes affect the trait
(polygenic characters)
(minor genes or polygenes)



How is the trait inherited?

- A. Autosomal dominant
- B. Autosomal recessive
- C. Sex-linked dominant
- D. Sex-linked recessive



Quantitative characters

Nilsson-Ehle (1910)

Kernel color in wheat

P

$R_1R_1R_2R_2 \times r_1r_1r_2r_2$

F₁

$R_1r_1R_2r_2$

Simple model:

Two independent genes that interact additively



	R_1R_1	R_1r_1	r_1r_1
R_2R_2	4	3	2
R_2r_2	3	2	1
r_2r_2	2	1	0

Five phenotypes

Quantitative characters

Nilsson-Ehle (1910)

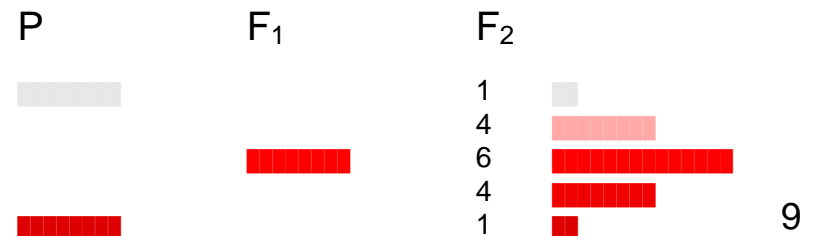
Kernel color in wheat

Two unlinked genes

Qualitative and quantitative characters do **not** differ genetically



P	$R_1R_1R_2R_2 \times r_1r_1r_2r_2$	
F ₁	$R_1r_1R_2r_2$	
F ₂	1	$r_1r_1r_2r_2$
	4	$R_1r_1r_2r_2$ $r_1r_1R_2r_2$
	6	$R_1R_1r_2r_2$ $R_1r_1R_2r_2$ $r_1r_1R_2R_2$
	4	$R_1R_1R_2r_2$ $R_1r_1R_2R_2$
	1	$R_1R_1R_2R_2$



Quantitative characters

Examples

Mice

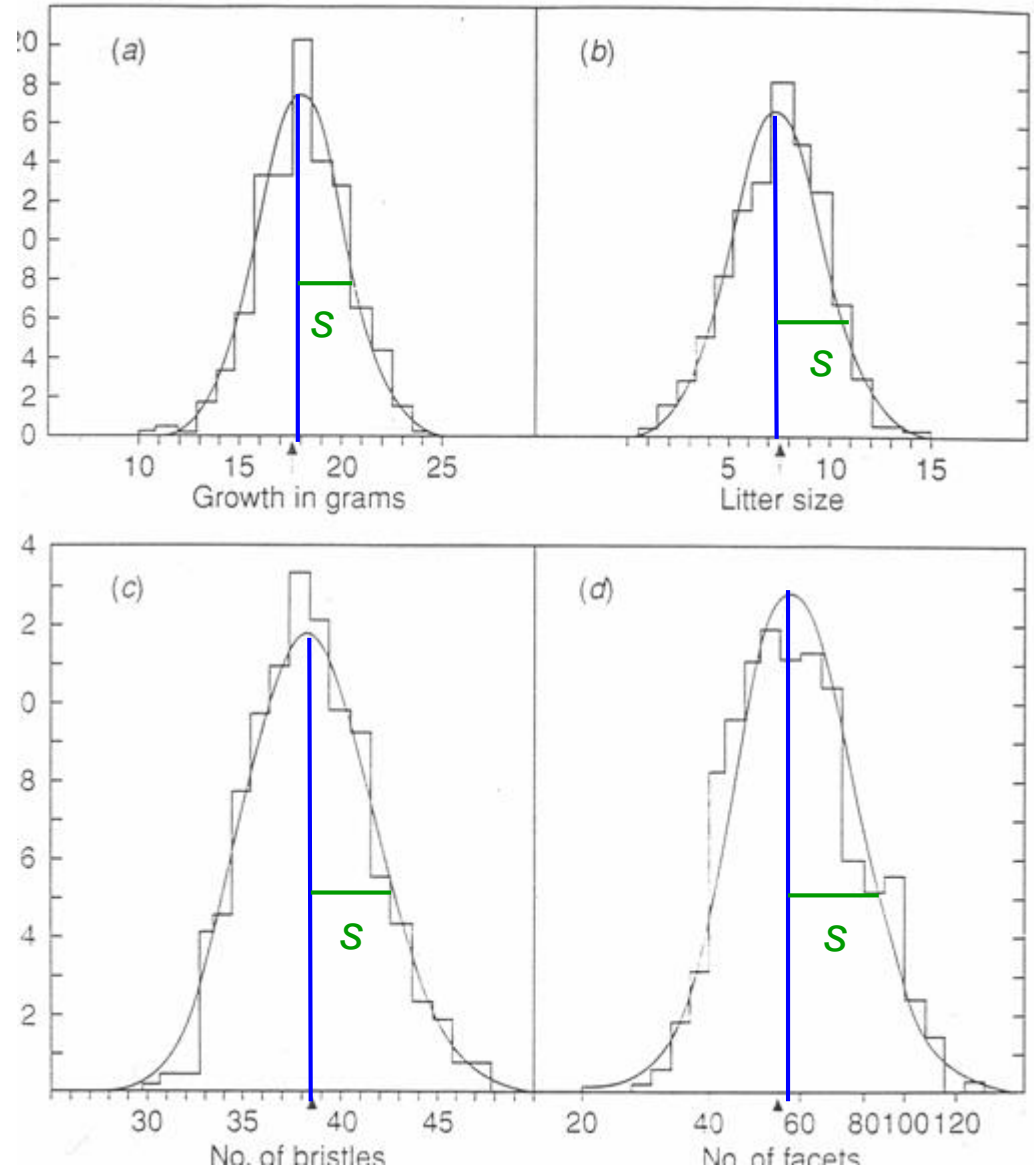
Drosophila

Important parameters

Average \bar{x}

Variance $V_x = s^2$

(Phenotypic)



Quantitative characters

How much do environment and genes influence a trait?

Components of the phenotypic variance

V_P : phenotypic variance (which we can estimate directly)

V_G : genetic variance

V_E : environmental variance

$$V_P = V_G + V_E$$

Quantitative characters

“Broad sense” heritability

$$h_B^2 = \frac{V_G}{V_P}$$

“Narrow sense” heritability

$$h_N^2 = \frac{V_A}{V_P}$$

(Will be described later)

Additive contributions
of alleles within loci

$$0 \leq h^2 \leq 1$$

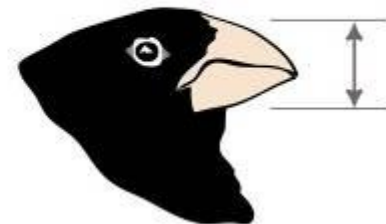
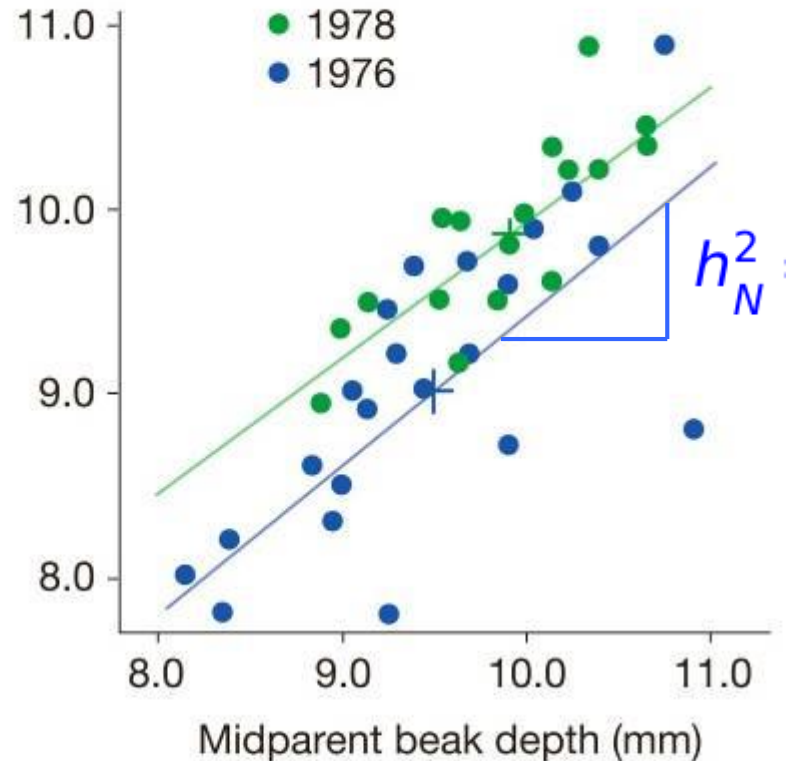
Quantitative characters

Estimating the heritability h^2_N

Darwin's finches

“Narrow sense” heritability

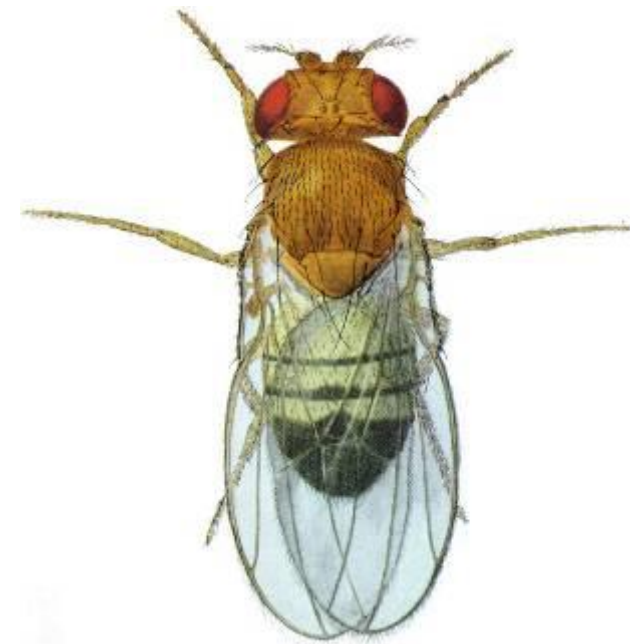
$$h^2_N = \frac{V_A}{V_P} = 0.9$$



Quantitative characters

Heritability for different traits in *Drosophila*

Category	Heritability
Morphology	0.46
Behavior	0.30
Physiology	0.33
Life history	0.26



The genetic variation for quantitative characters is very high

Chromosomal variation

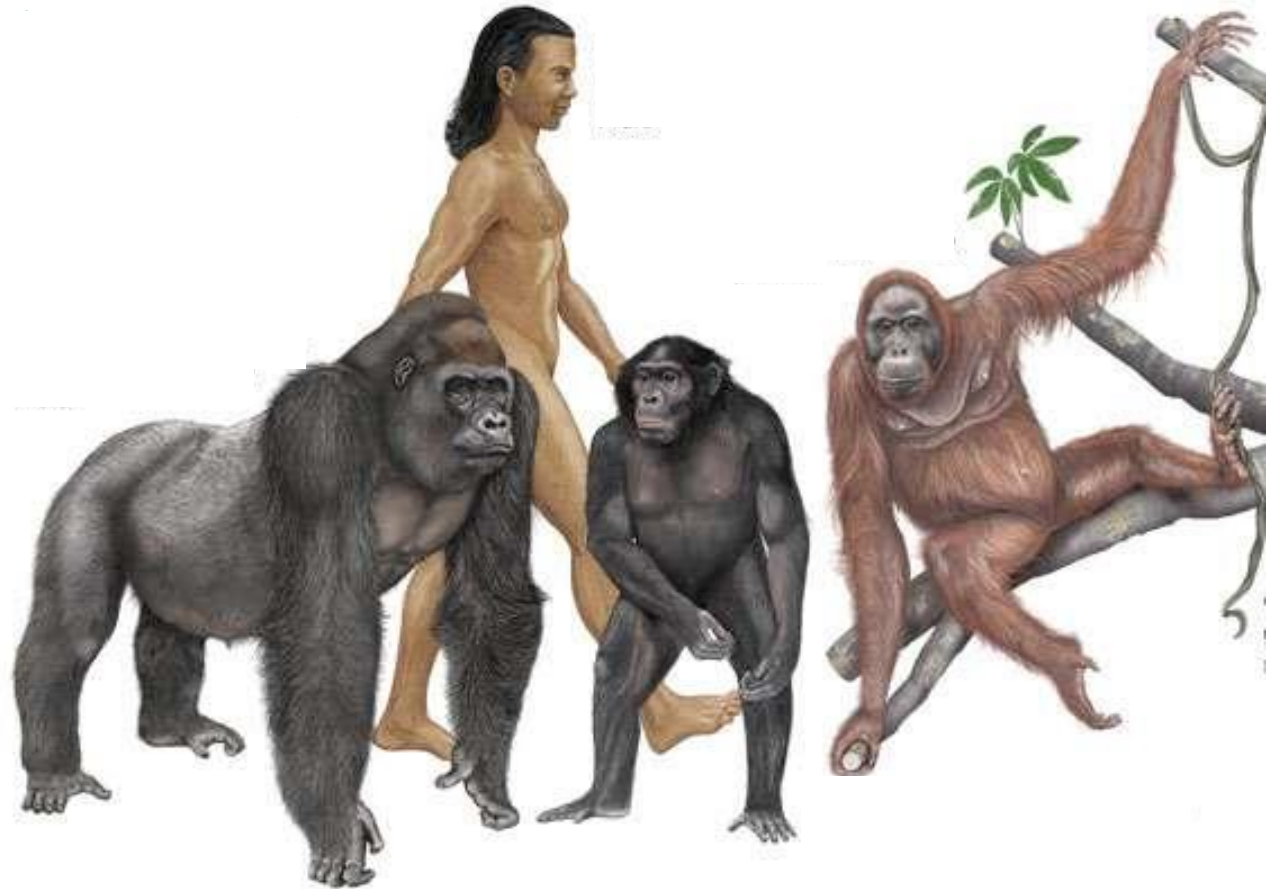
Primates

Human

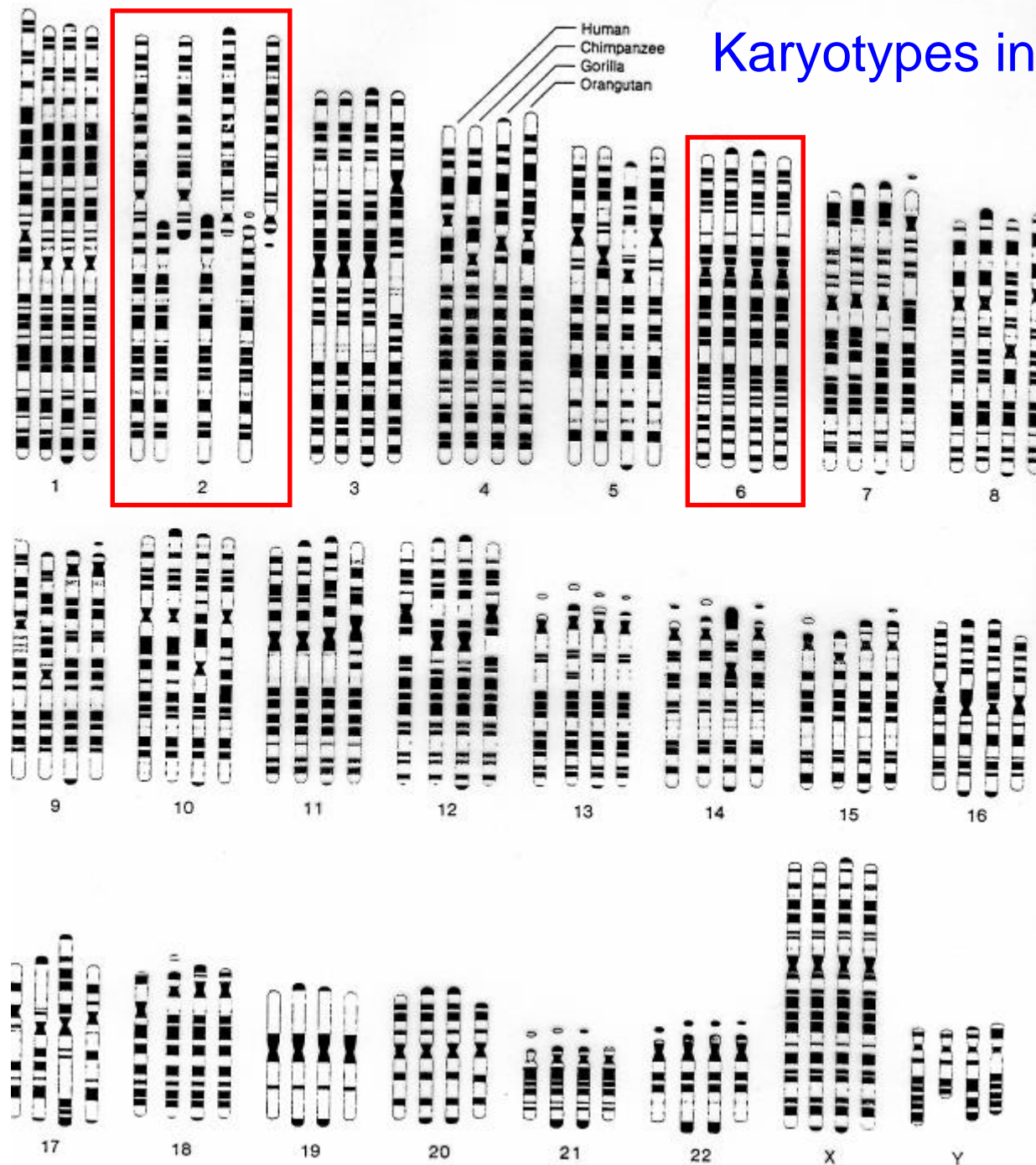
Chimpanzee

Gorilla

Orangutang



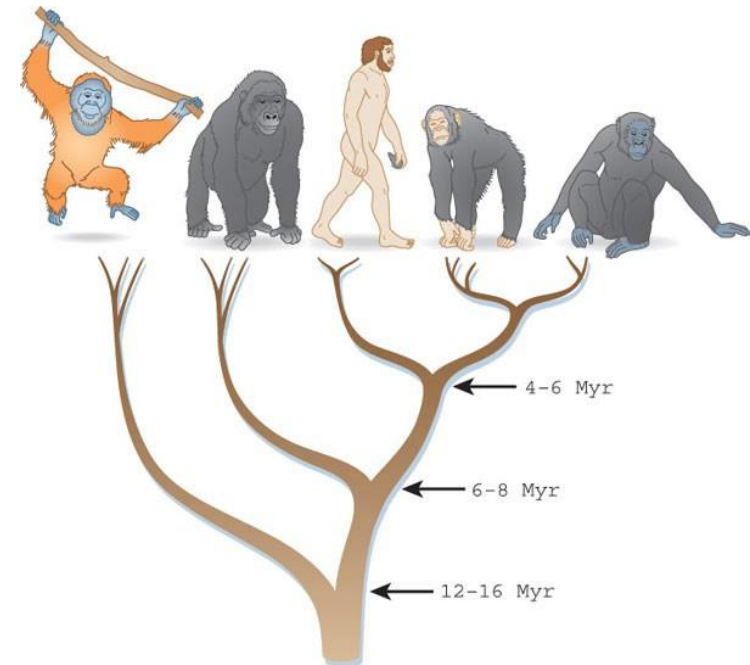
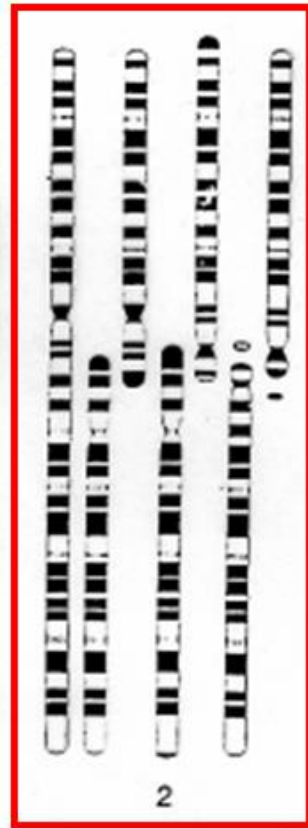
Karyotypes in primates



Which chromosomal mutation has occurred?

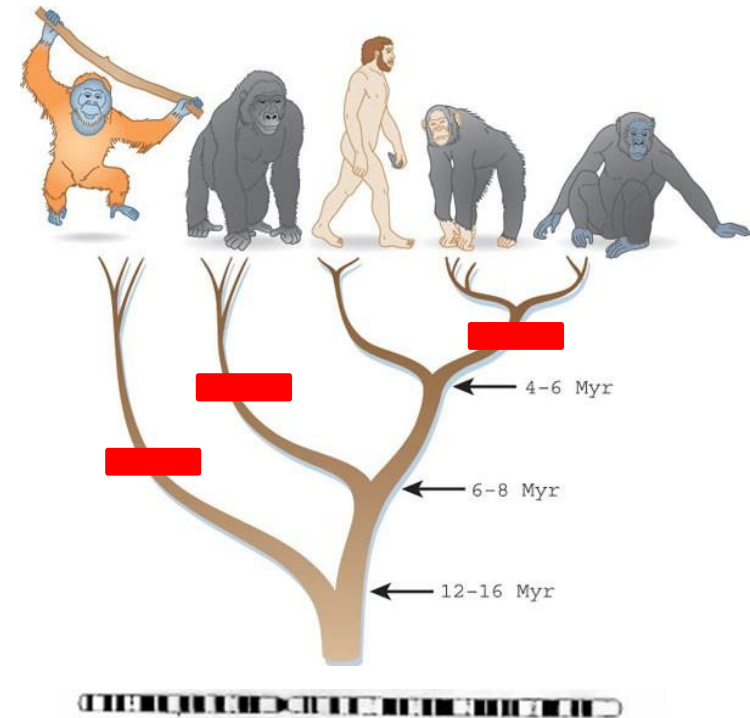
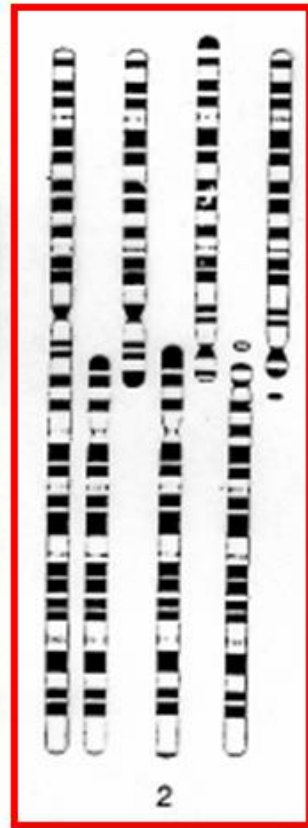
Fusion

Fission



Which chromosomal mutation has occurred?

Fission

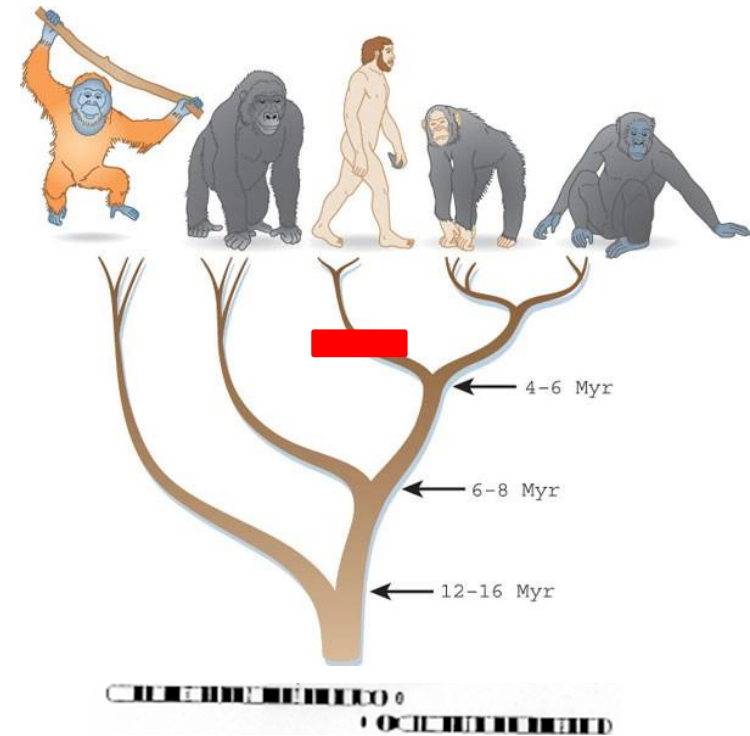
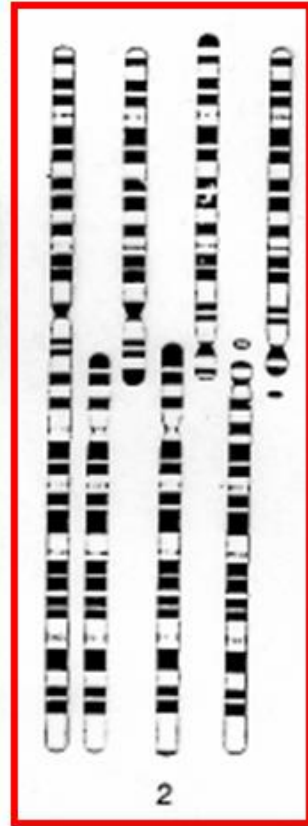


Common ancestor had one set of chromosomes

Requires 3 mutations

Which chromosomal mutation has occurred?

Fusion

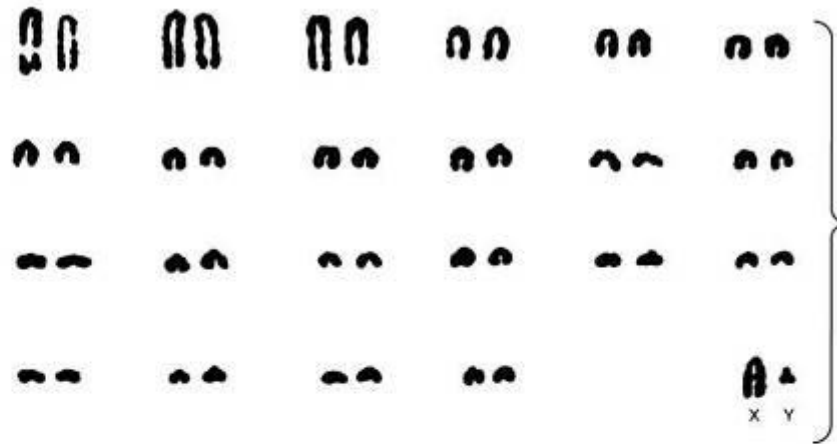


Common ancestor had two set of chromosomes

Requires 1 mutation
Most parsimonious

Karyotype of Chinese and Indian muntjac

Muntiacus reevesii ($2N = 46$)



Muntiacus muntiacus ($2N = 8$)



Chromosomal variation

Extensive variation

Difficult to study and quantify

Very variable among taxonomic groups

Important for reproductive isolation among species.

Variation in single genes

- Polymorphism
 - a) More than two alleles present in a population
 - b) $p(\text{rare allele}) > 0.01$
- Gene diversity
 - $p(\text{individual is heterozygous})$
 - (measured as observed or expected according to HW)
- Which is most affected by the sample size?

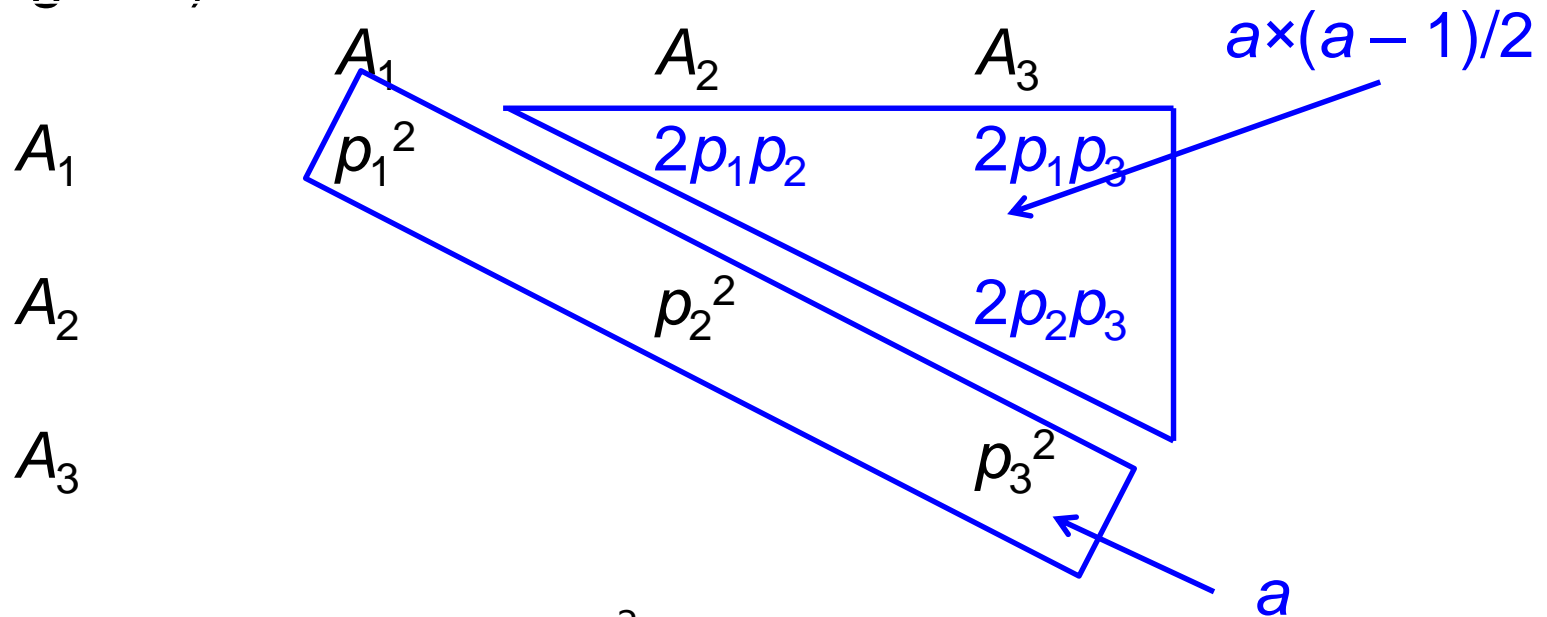
Polymorphism and gene diversity

Locus	A_1A_1	A_1A_2	A_2A_2	Number
1	100	0	0	100
2	81	18	1	100

Locus	Polymorphism	Observed gene diversity
1	no	0
2	yes	0.18

Polymorphism and gene diversity

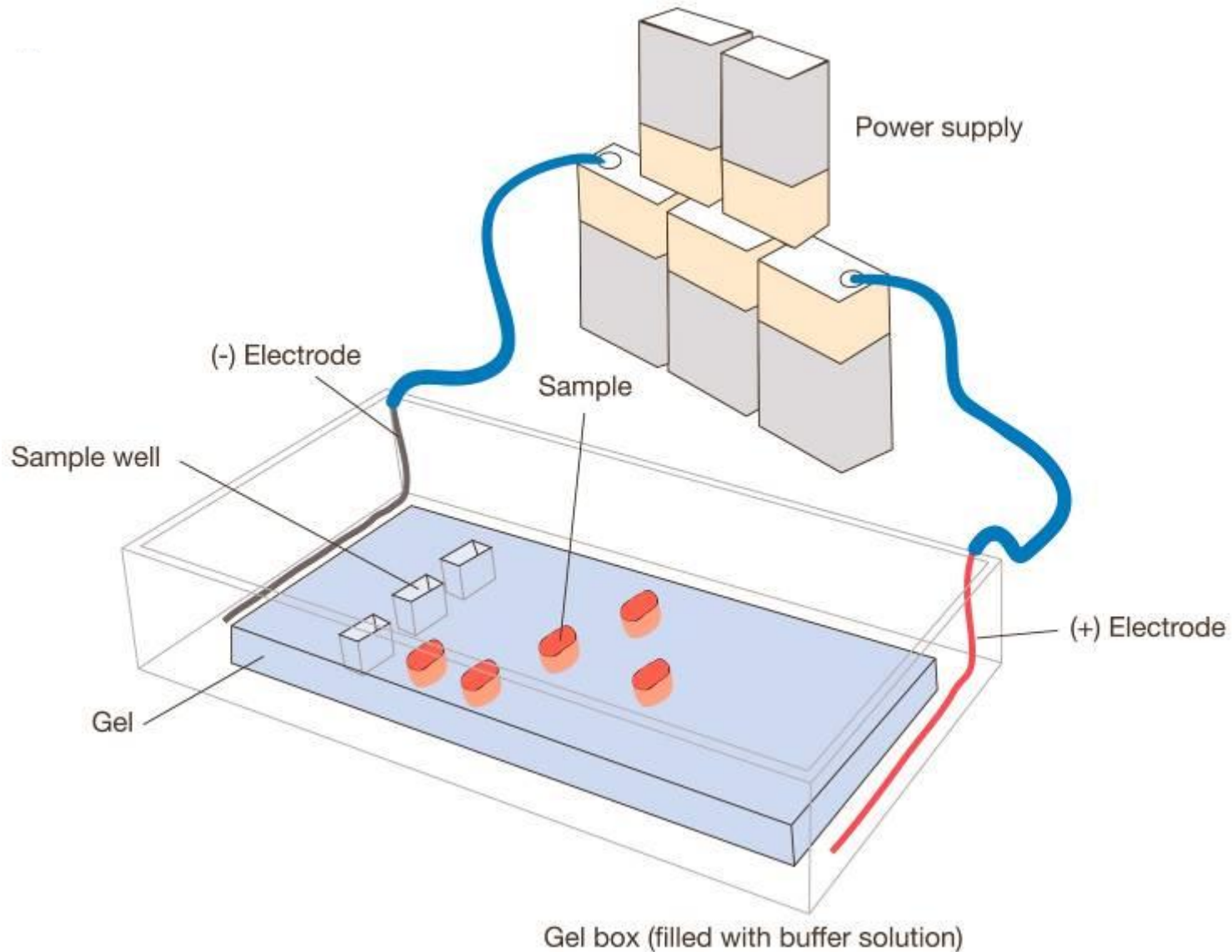
Expected gene diversity: Estimation for multiallelic (a) locus:
(assuming HW)



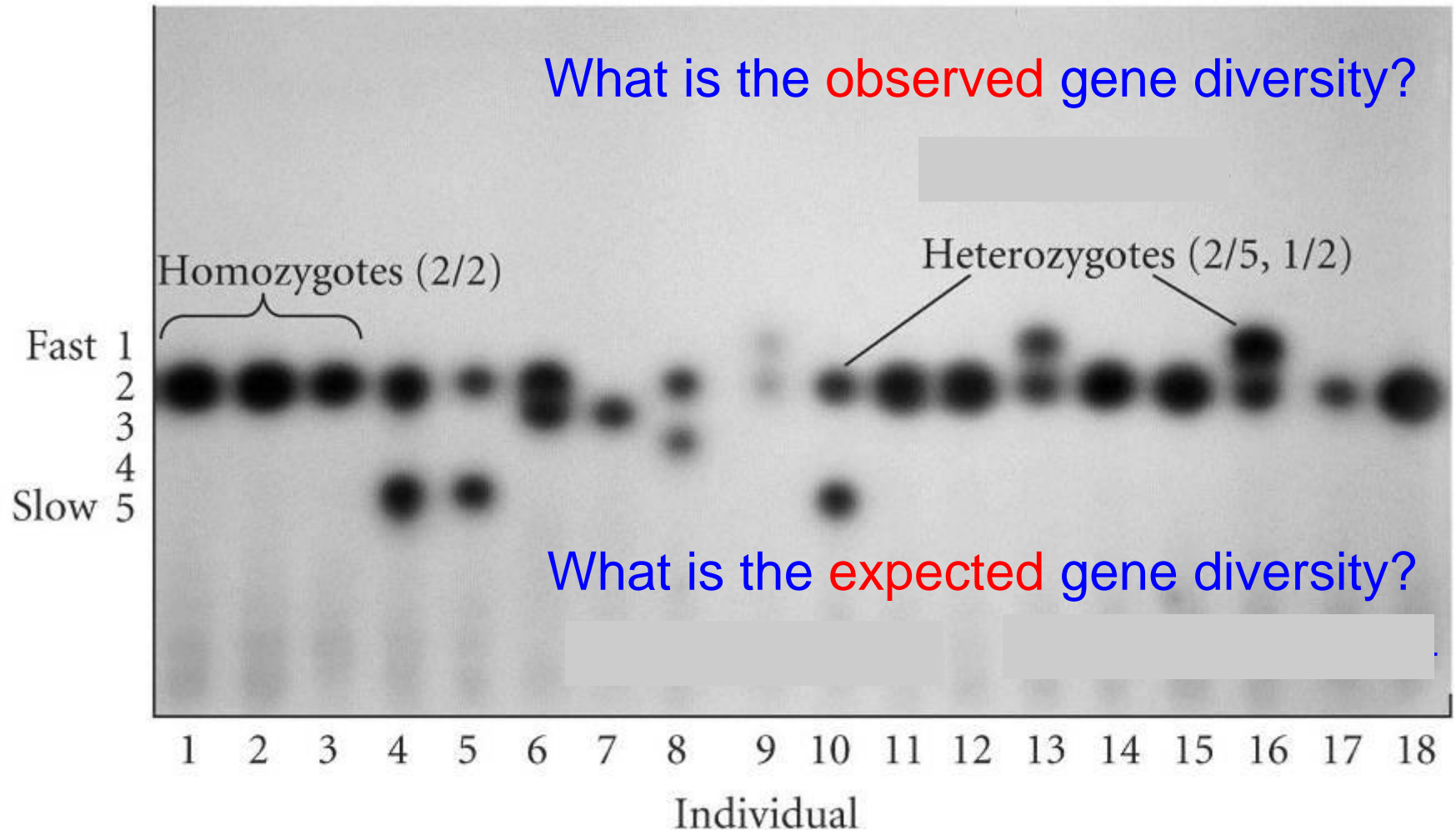
Gene diversity:

$$H = 1 - \sum_{i=1}^3 p_i^2$$

Enzyme electrophoresis



Enzyme electrophoresis

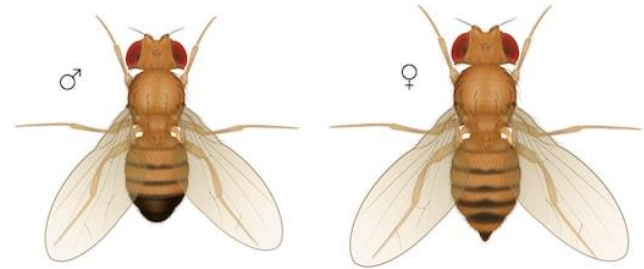


Enzyme electrophoresis

First estimates in natural populations (allozymes)

Lewontin & Hubby (1966)

Drosophila $H = 0.12$

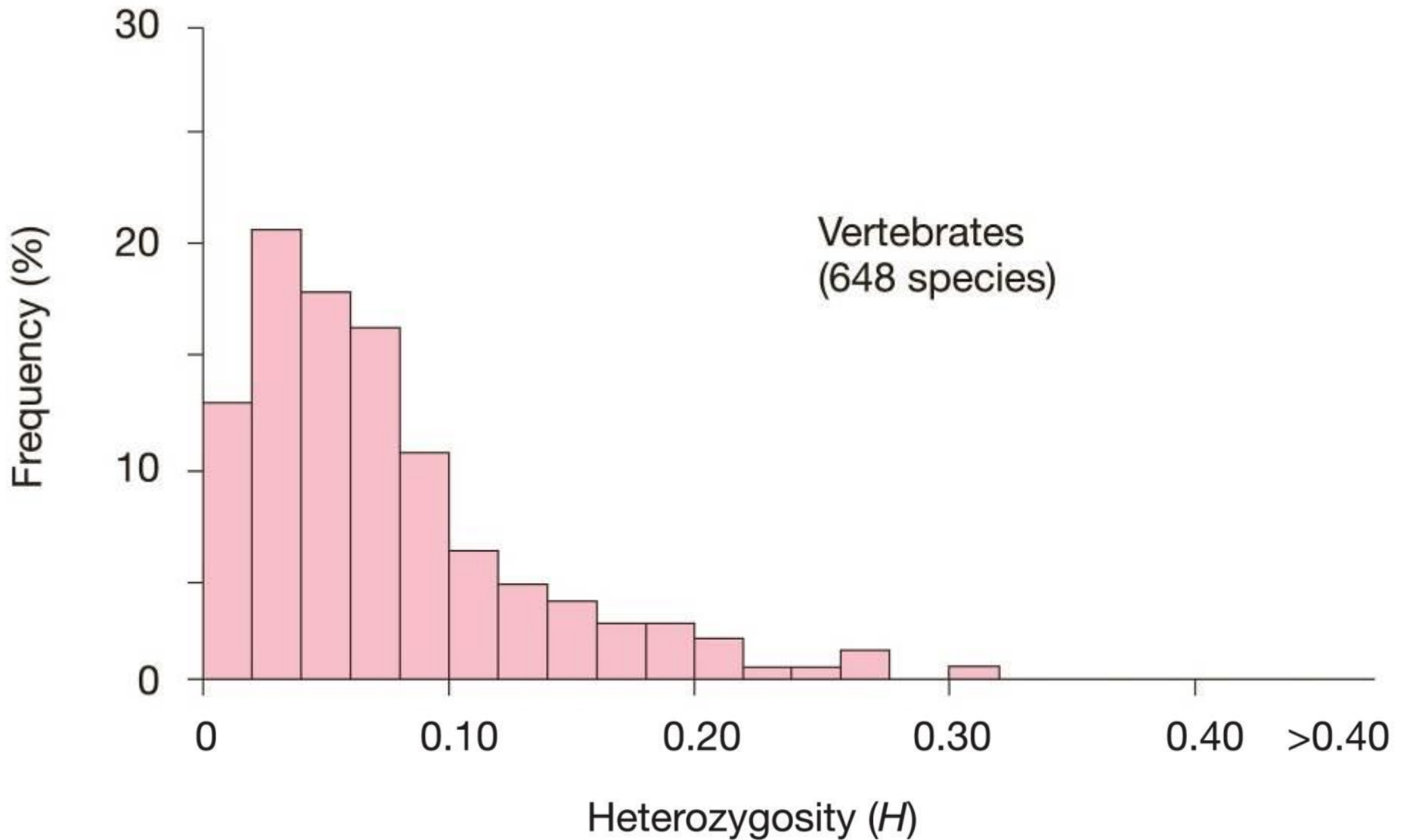


Harris (1966)

Humans $H = 0.10$



Enzyme electrophoresis

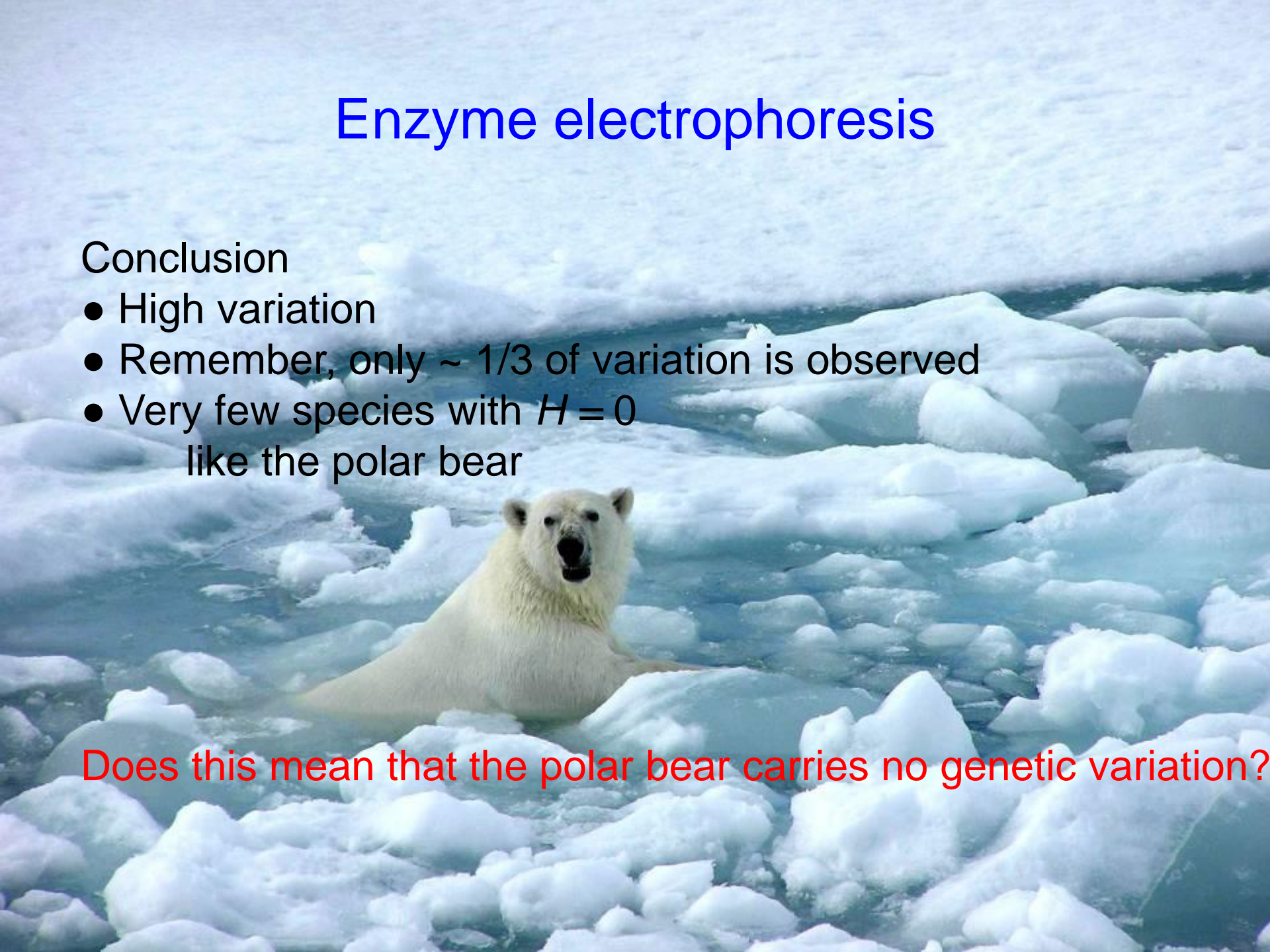


Enzyme electrophoresis

Conclusion

- High variation
- Remember, only $\sim 1/3$ of variation is observed
- Very few species with $H = 0$
like the polar bear

Does this mean that the polar bear carries no genetic variation?



Variation in DNA sequences

Single nucleotide polymorphisms (SNP)

1 every few hundred bp, mutation rate* $\approx 10^{-9}$

TGCATT**G**CGTAGGC
TGCATT**C**CGTAGGC

Short indels (=insertion/deletion)

1 every few kb, mutation rate v. variable

TGCATT---TAGGC
TGCATT**CCG**TAGGC

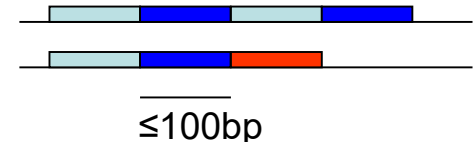
Microsatellite (STR[§]) repeat number

1 every few kb, mutation rate $\leq 10^{-3}$

TGCT**TCATCATCATCA**GC
TGCT**TCATCA**-----GC

Minisatellites

1 every few kb, mutation rate $\leq 10^{-1}$



*Mutation rates are measured per unit per generation or per year

§STR: Simple tandem repeats

Variation in DNA sequences

Microsatellites

[Short Tandem Repeats (STR)]

2, 3, 4...9 base pairs copies

Variation in copy number

Few to 100s of copies

Neutral?

Very abundant in most eukaryotic species

TGCTCATCATCATCAGC
TGCTCATCA-----GC

Advantages

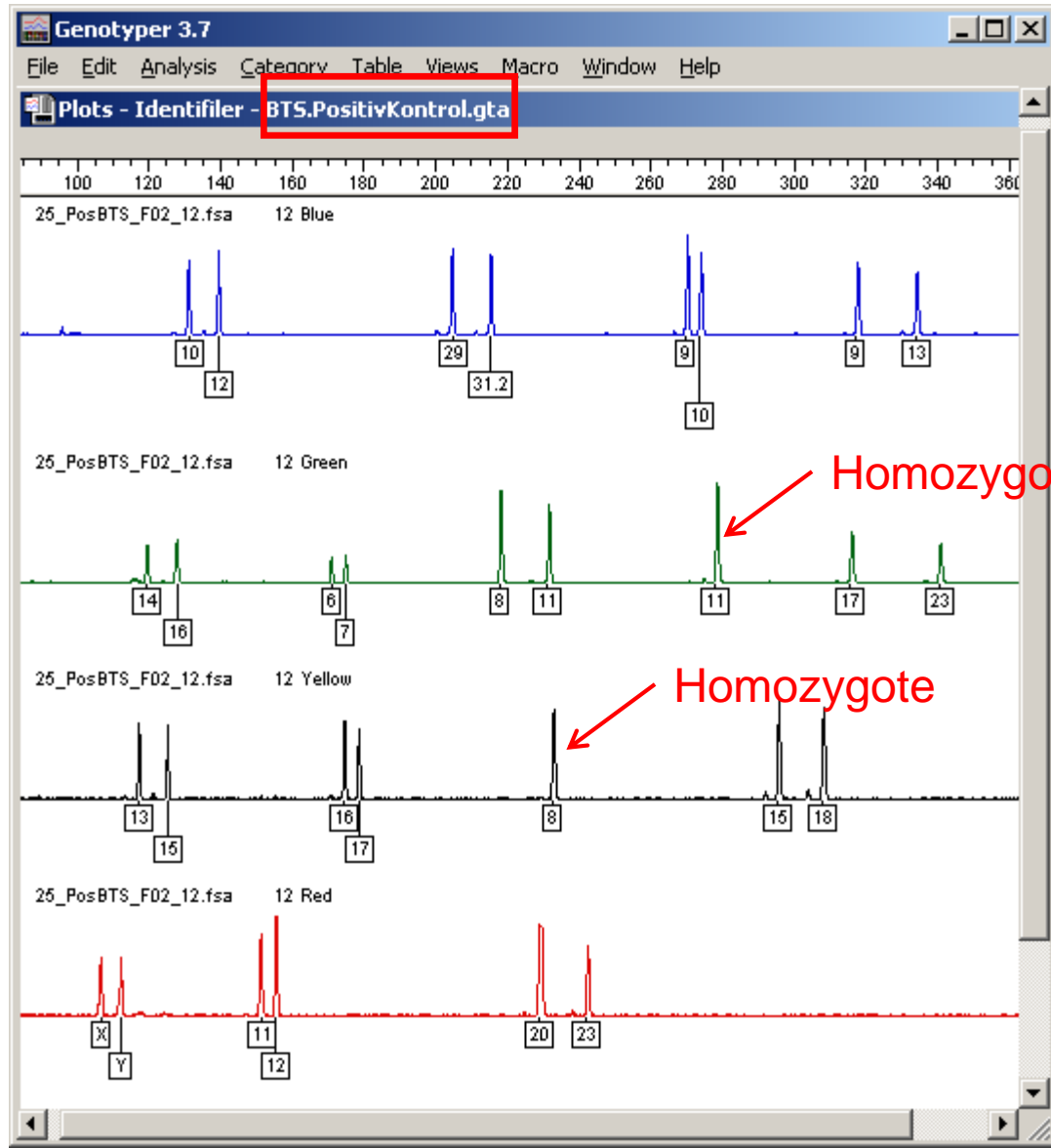
- Codominant

- High mutation rate

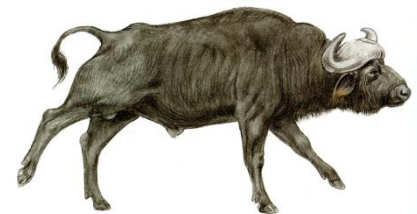
- Highly variable

Variation in DNA sequences

Microsatellites



Bo T. Simonsen
Head of
Forensic Genetics

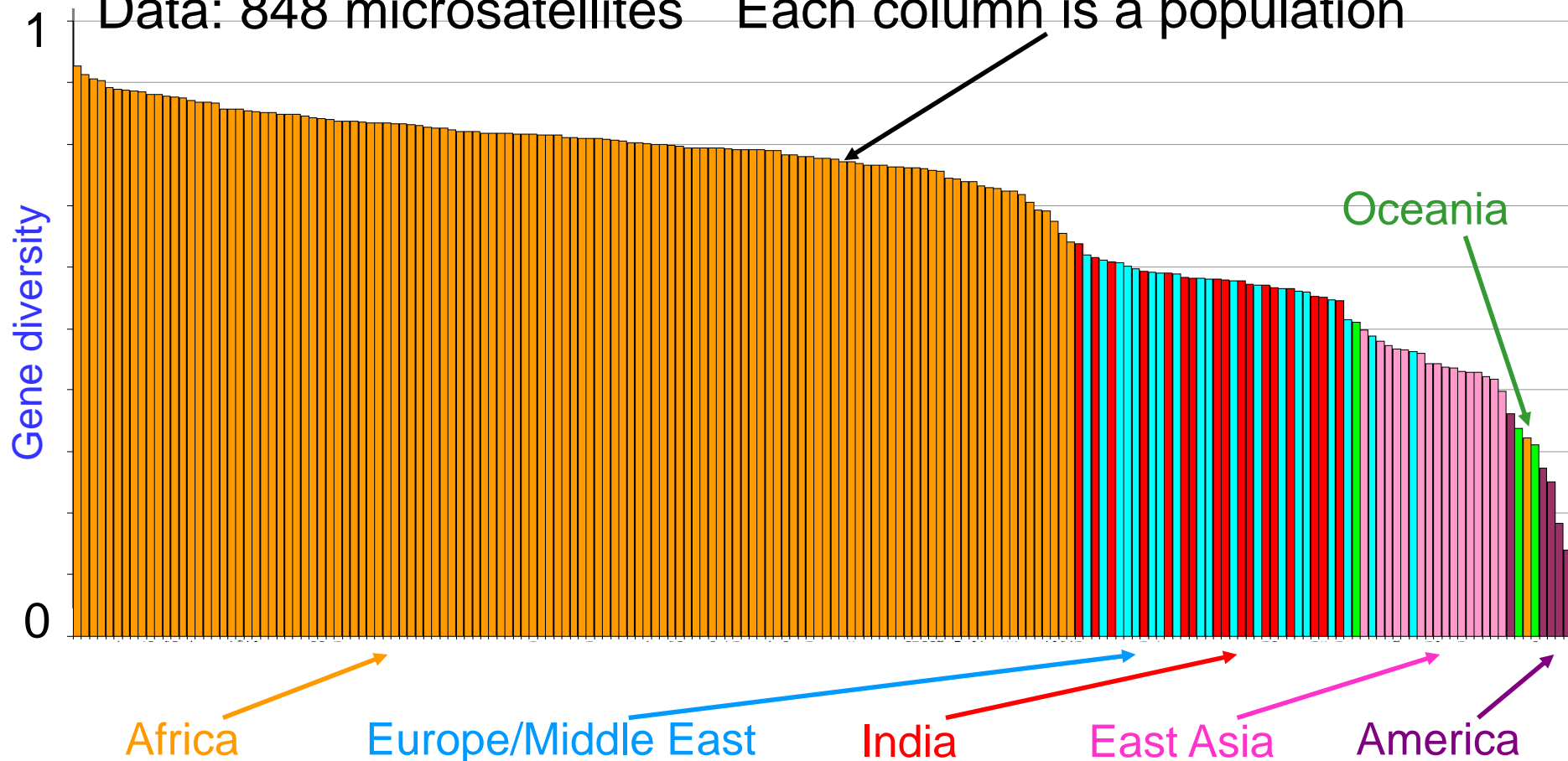


Variation in DNA sequences

Microsatellites

Gene diversity in 181 human populations ($H = 1 - \sum p_i^2$)

Data: 848 microsatellites Each column is a population



Diversity decreases with distance from Africa

Variation in DNA sequences

Sequence 1	AAAGGTCTAG
Sequence 2	AAAGGTCTAT
Sequence 3	AAAAGGCTAT
Sequence 4	AAAAGGCTAT

How much variation do we expect?

How do we quantify the variation?

- Haplotype diversity

- Pairwise differences

- Segregating sites

- Expected nucleotide diversity

Variation in DNA sequences

Sequence 1 AAAGGTCTAG
Sequence 2 AAAGGTCTAT
Sequence 3 AAAAGGCTAT
Sequence 4 AAAAGGCTAT

Haplotype diversity
($H = 1 - \sum p_i^2$)

Count of haplotypes: 3 (seq 3 = seq 4)

Frequencies:

Haplotype 1: $p_1 = 1/4$

Haplotype 2: $p_2 = 1/4$

Haplotype 3: $p_3 = 1/2$

Haplotype diversity = $1 - (1/4)^2 - (1/4)^2 - (1/2)^2 = 0.625$

Is haplotype diversity a useful measure?

Haplotype: A sequence that differs from other homologous sequences in at least one position

Variation in DNA sequences

Expectations from the **infinite-site model**

Every mutation hits a new base position:

Constant population size (N)

Neutral variation

Expected nucleotide diversity:

$$\Theta = 4N\mu,$$

$$2N\mu$$

$$2N\mu$$

$$2N$$

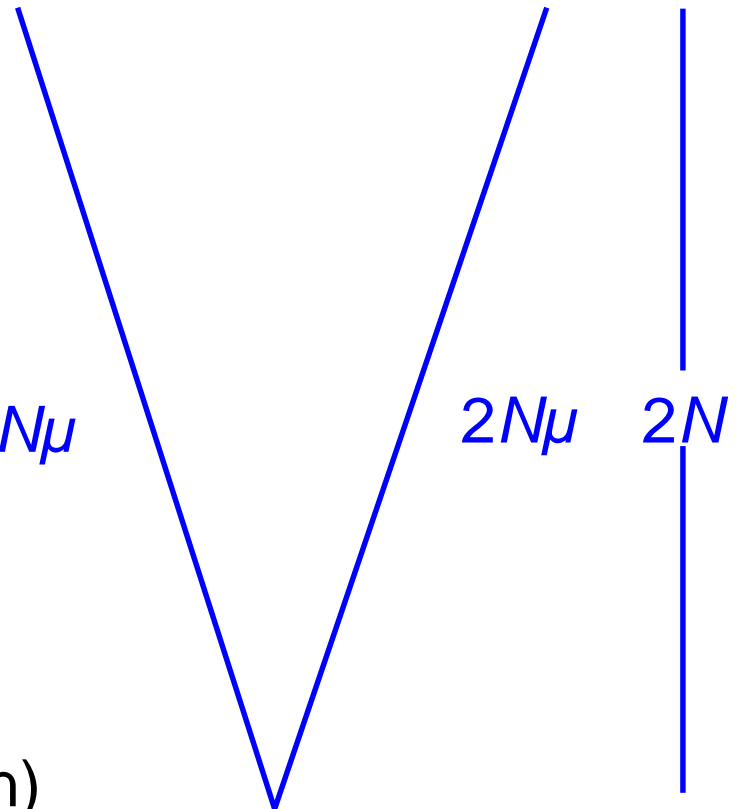
where

N = population size

μ = mutation rate

(per base per generation)

Most Recent Common Ancestor $t_{MRCA}(2) = 2N$



Variation in DNA sequences

Sequence 1	AAAGGTCTAG
Sequence 2	AAAGGTCTAT
Sequence 3	AAAAGGCTAT
Sequence 4	AAAAGGCTAT

Watterson's
estimator of θ

$$\hat{\theta}_w = \frac{S}{\sum_{k=1}^{n-1} \frac{1}{k}}$$

Segregating Sites (S)	0001010001	3
------------------------------	------------	---

n number of sequences	4
-------------------------	---

$$\theta_w = 3 / (1/1 + 1/2 + 1/3) / 10 = 0.16$$

10: length of sequence

Variation in DNA sequences

Sequence 1 AAAGGTCTAG
 Sequence 2 AAAGGTCTAT
 Sequence 3 AAAAGGCTAT
 Sequence 4 AAAAGGCTAT

Tajima's
estimator of θ

$$E[\pi] = \frac{\sum_{i < j} d_{ij}}{n(n-1)/2}$$

Pairwise nucleotide differences

	1	2	3	4
2	1			
3	3	2		
4	3	2	0	

n number of sequences 4

$$\theta_T = (1+3+3+2+2+0)/(4(4-1)/2 \times 10) = 0.18$$

Nucleotide diversity: (Total number of differences) /
 ((Number of pairwise comparisons) x (Length of sequence))

Variation in DNA sequences

Sequence 1 AAAGGTCTAG

Sequence 2 AAAGGTCTAT

Sequence 3 AAAAGGCTAT

Sequence 4 AAAAGGCTAT

Expected nucleotide
diversity

$p(\text{base 1})$ 1 1 1 $\frac{1}{2}$ 1 $\frac{1}{2}$ 1 1 1 $\frac{1}{4}$

$q(\text{base 2})$ 0 0 0 $\frac{1}{2}$ 0 $\frac{1}{2}$ 0 0 0 $\frac{3}{4}$

$2pq$ 0 0 0 $\frac{1}{2}$ 0 $\frac{1}{2}$ 0 0 0 $\frac{3}{8}$

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

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

Expected nucleotide diversity

$$\pi = (0 + 0 + 0 + \frac{1}{2} + 0 + \frac{1}{2} + 0 + 0 + 0 + \frac{3}{8})/10$$

$$= 0.14$$

Variation in DNA sequences

Do we expect the same diversity among and along chromosomes?

$$\begin{aligned}\mu &= \text{neutral mutation rate} \\ &= f_n \times \mu_T\end{aligned}$$

where

$$\begin{aligned}f_n &= \text{fraction of all mutations that are neutral} \\ \mu_T &= \text{Total mutation rate}\end{aligned}$$

Variation in DNA sequences

Do we expect the same diversity among and along chromosomes?

Selective constraints vary

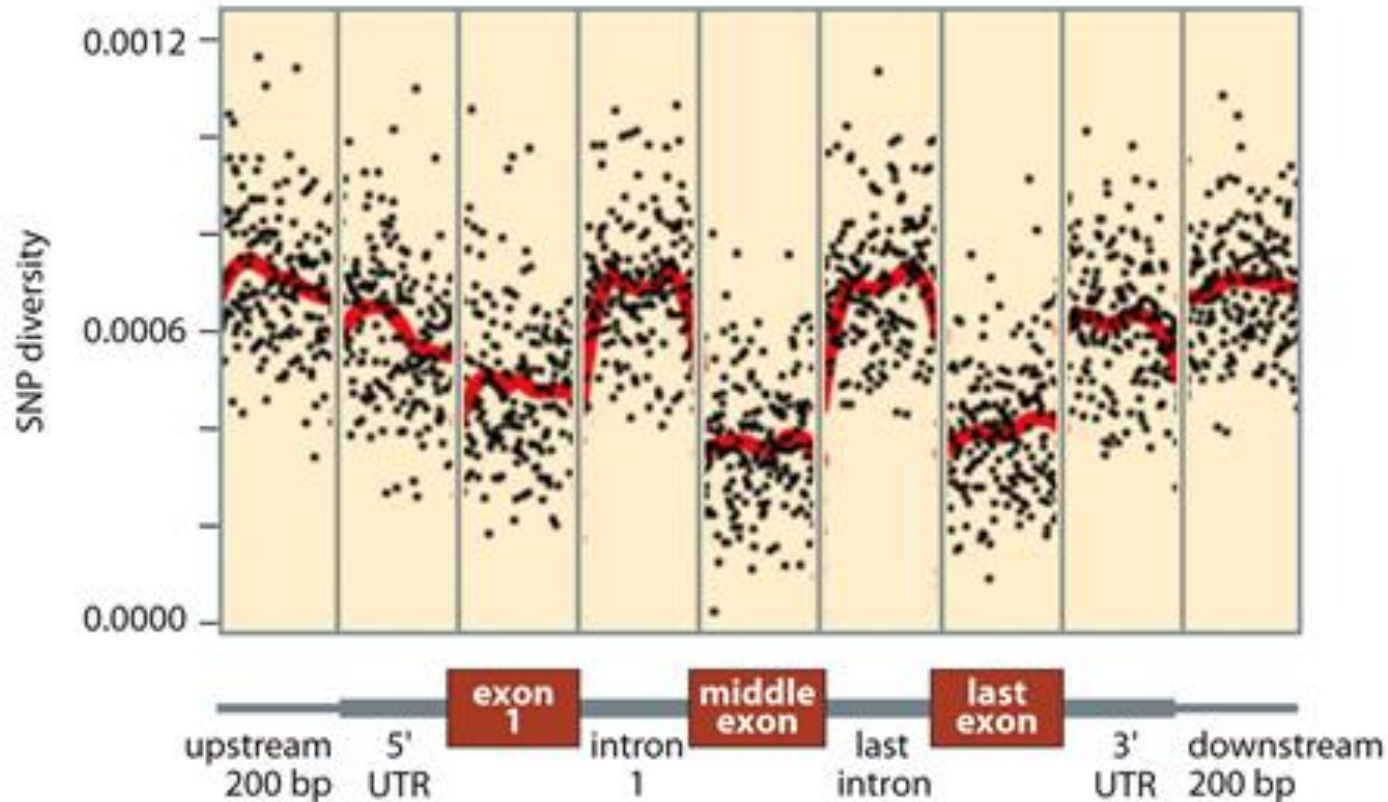
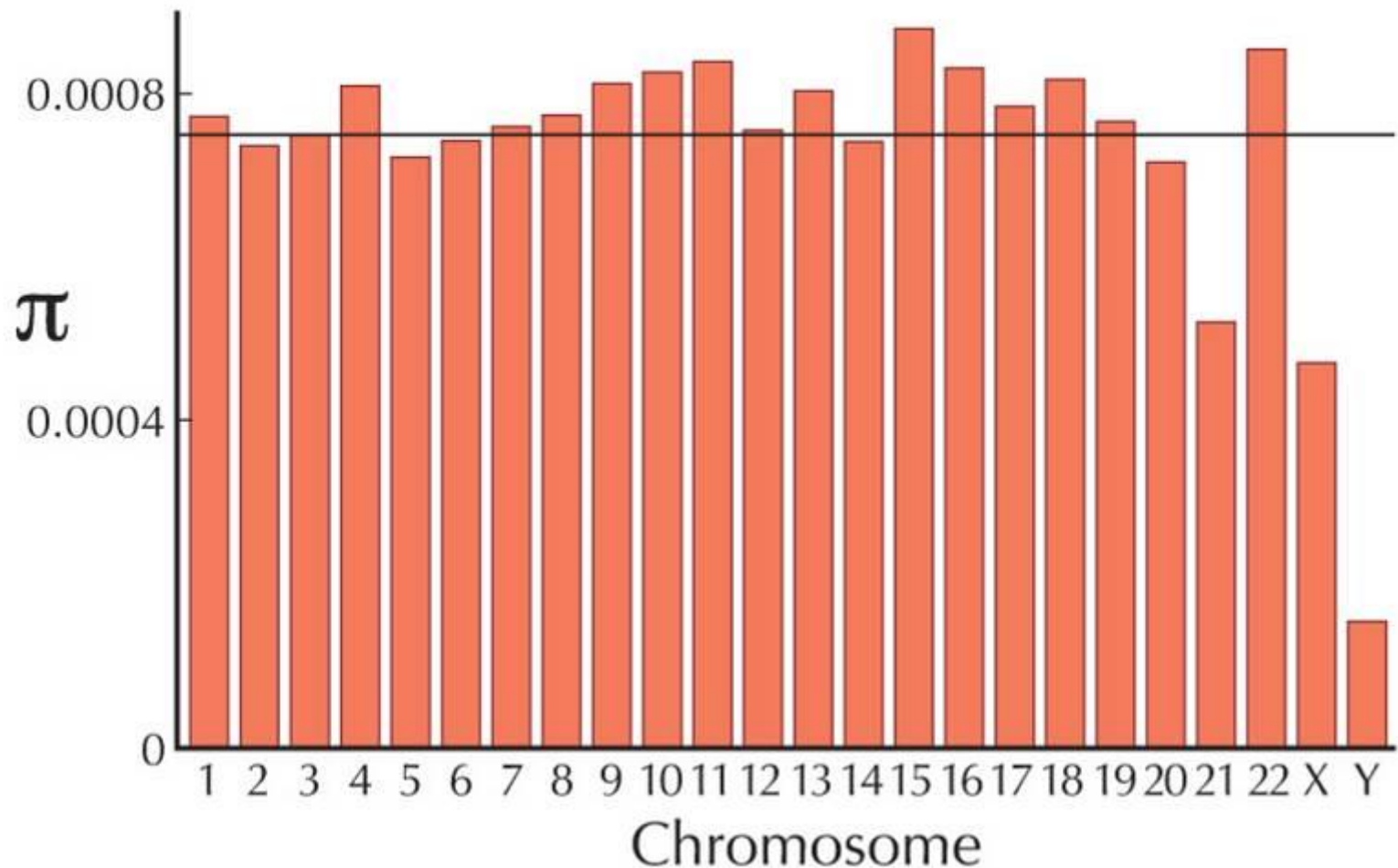


Figure 3.11 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

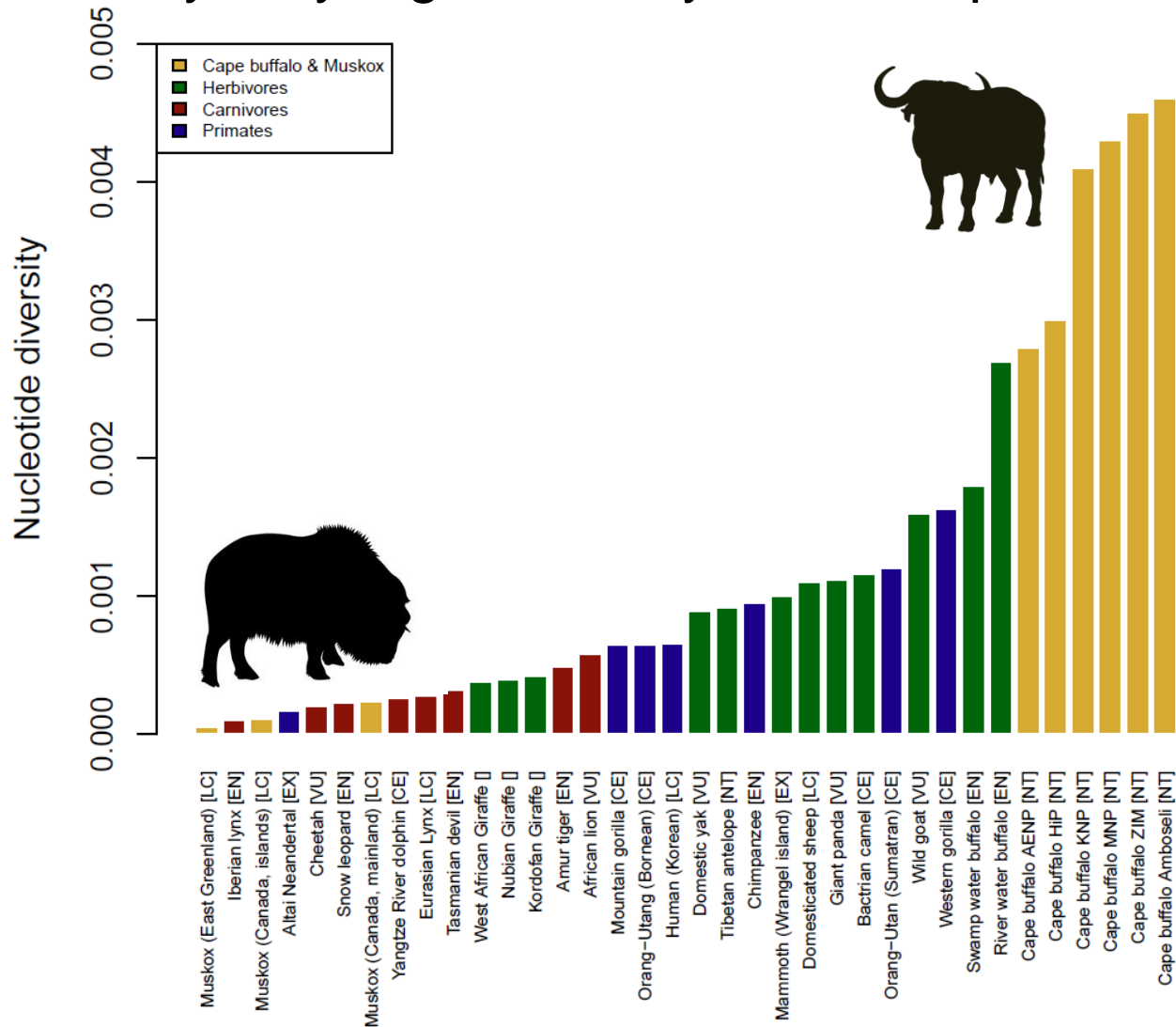
Variation in DNA sequences

Nucleotide diversity in humans ≈ 0.001



Variation in DNA sequences

Generally very high diversity in most species.



Variation in natural populations

Conclusions

Quantitative characters

Genetic variance (heritability)

Single genes

Gene diversity

Microsatellites

Gene diversity

DNA sequences

Nucleotide diversity

The genetic variation is very high.

BREAK before
Exercise on estimating nucleotide diversity in chimpanzees

