Genetic Variation:
What it is and how to summarize it

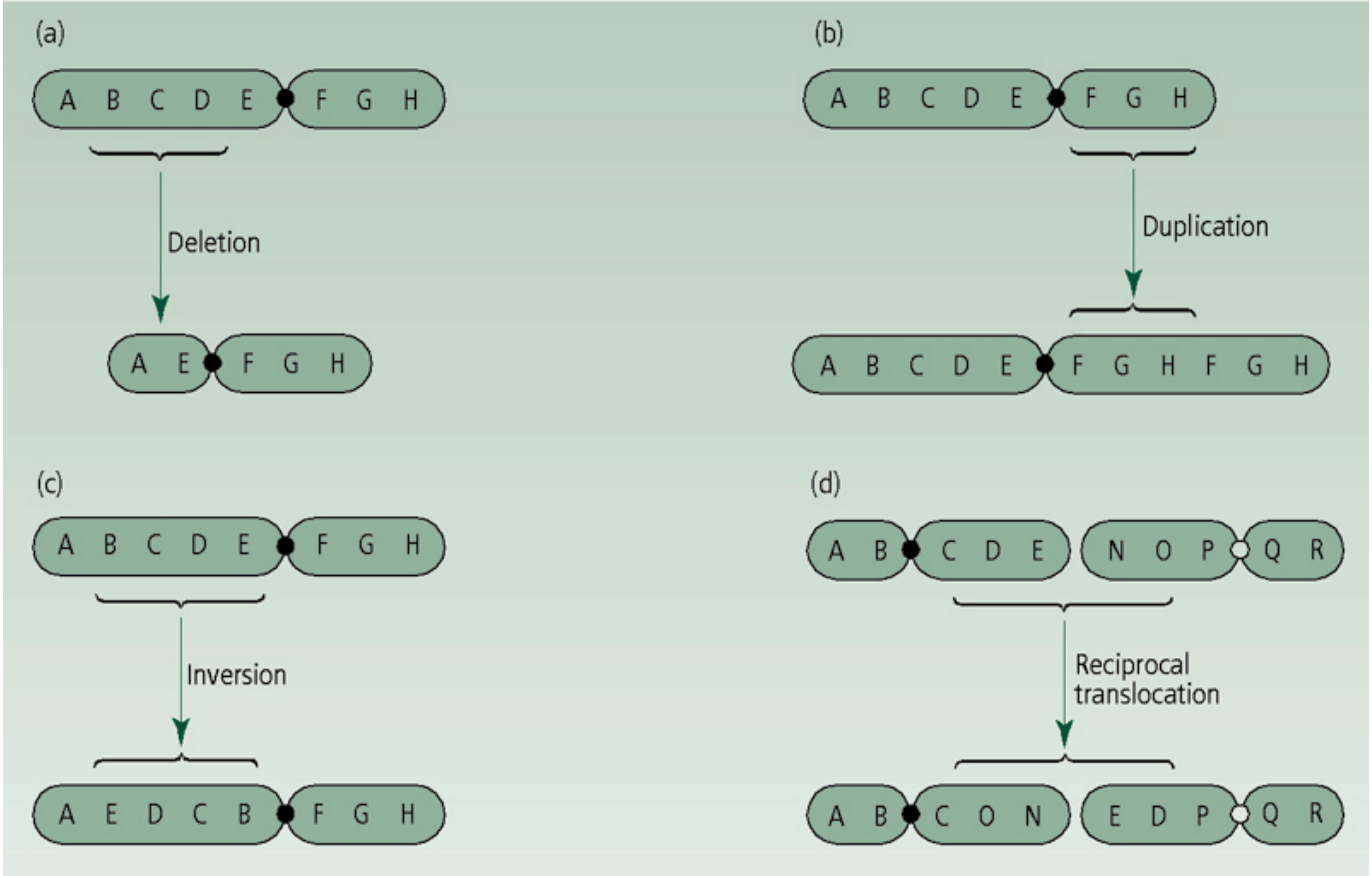
Genetic Variation

- Chromosomal variation
- Protein variation
- DNA variation

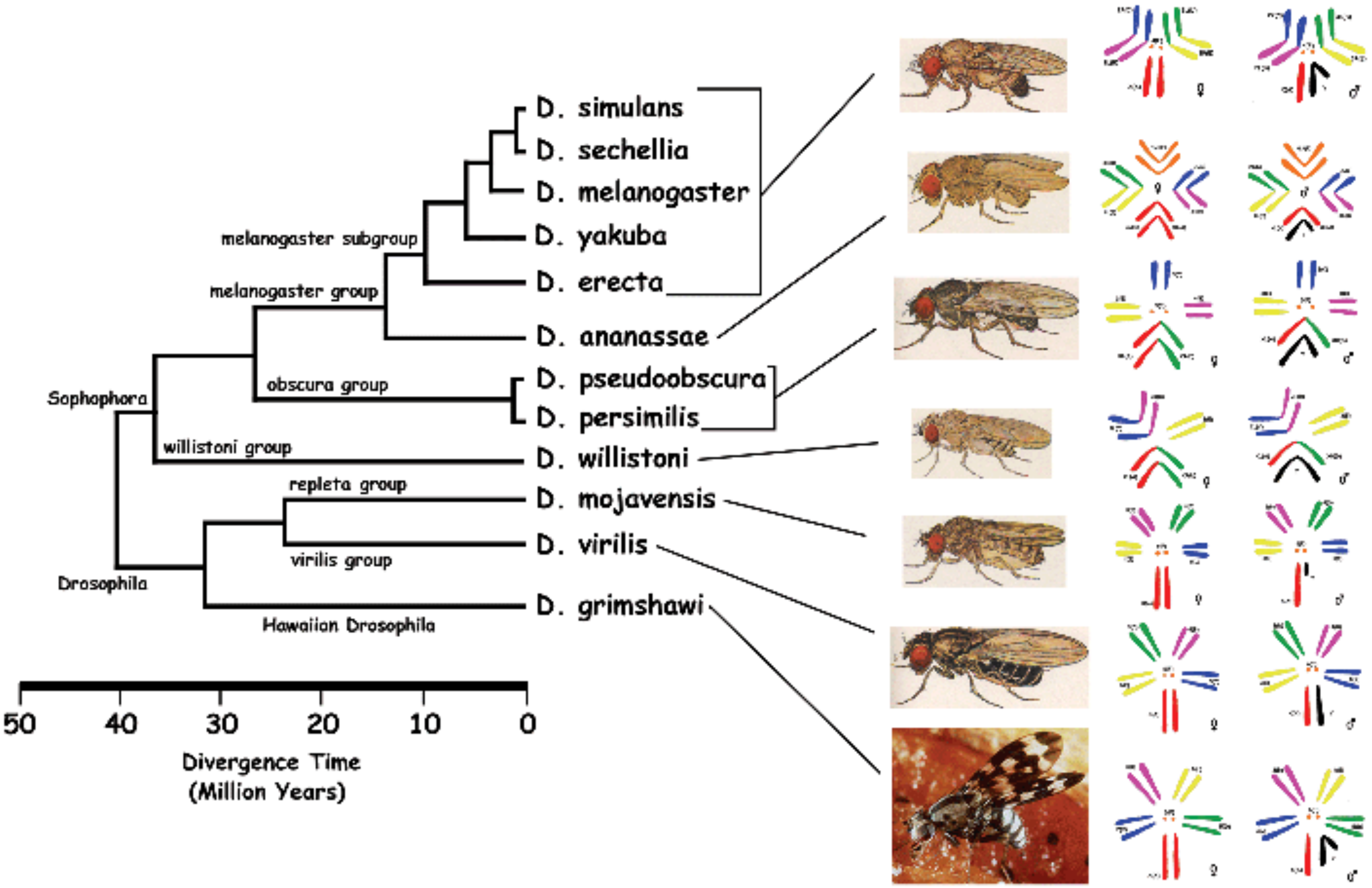
Chromosomal Variation

- Variation in chromosome number, gene number, gene order etc.

Chromosomal rearrangements

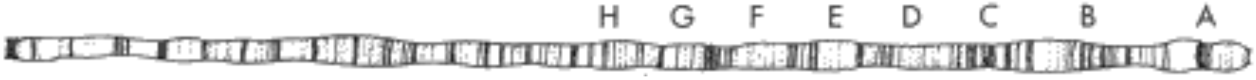


Chromosomal Variation

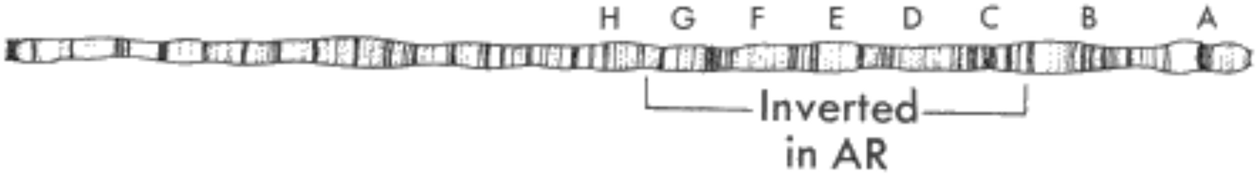


Chromosomal Variation

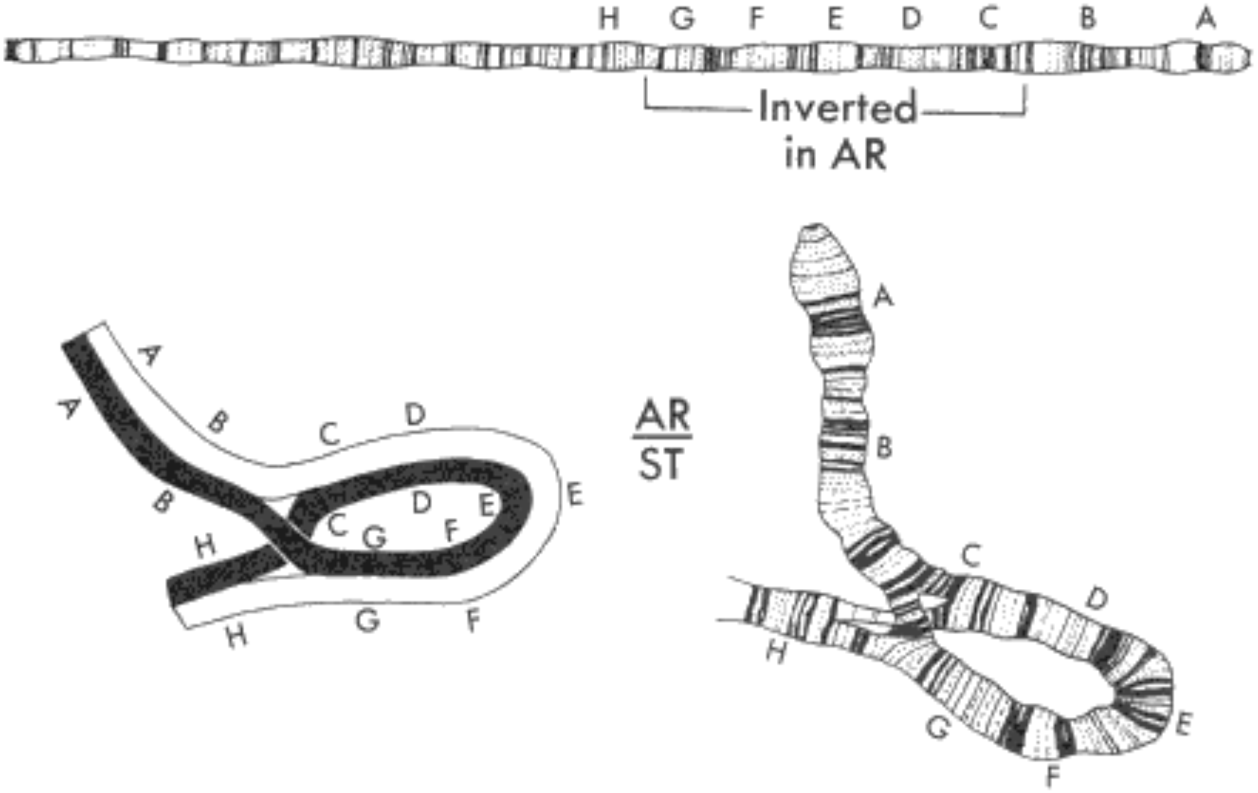
Chromosomal Variation



Chromosomal Variation

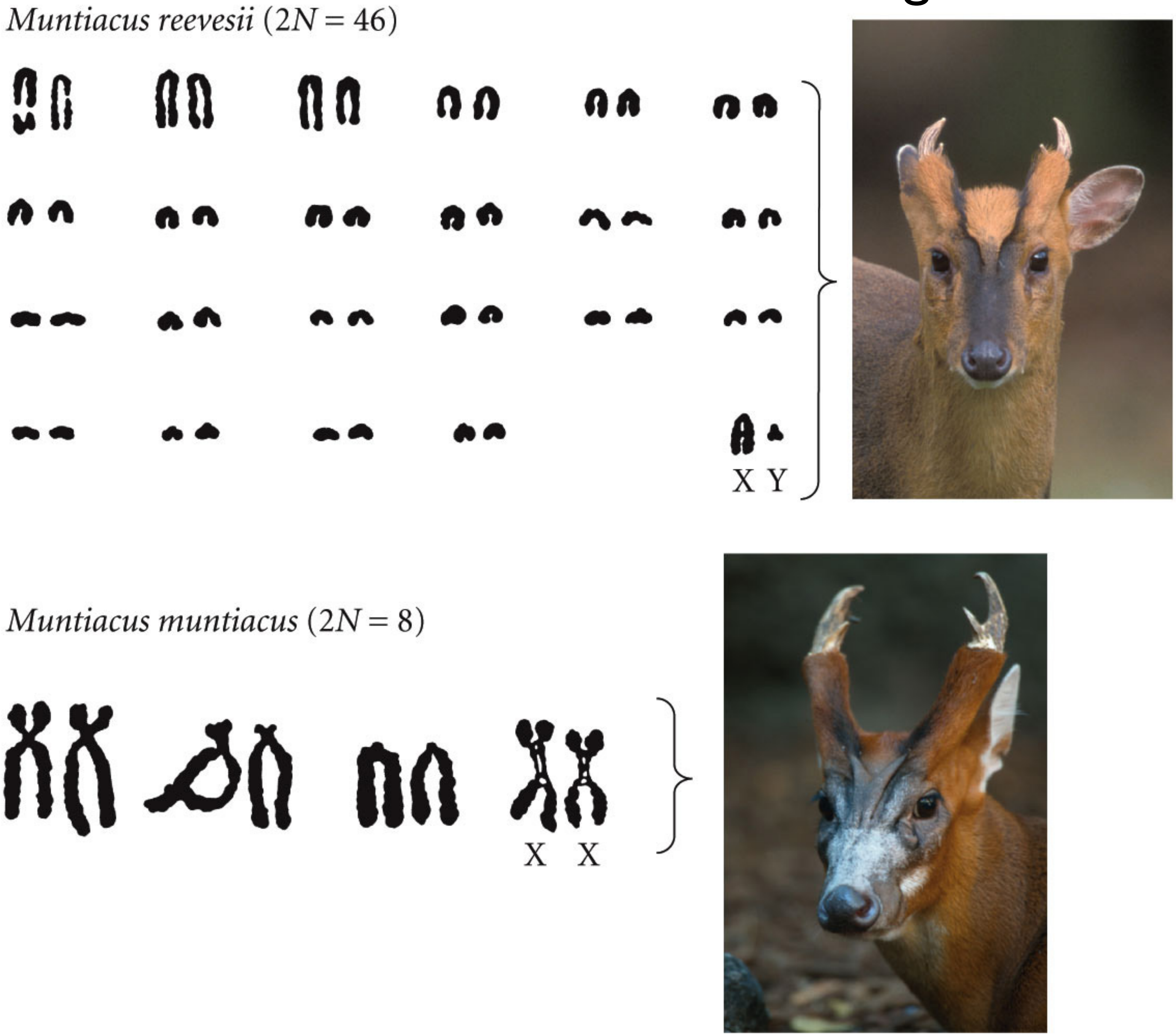


Chromosomal Variation



From Dobzhansky and Sturtevant

Chromosomal number changes

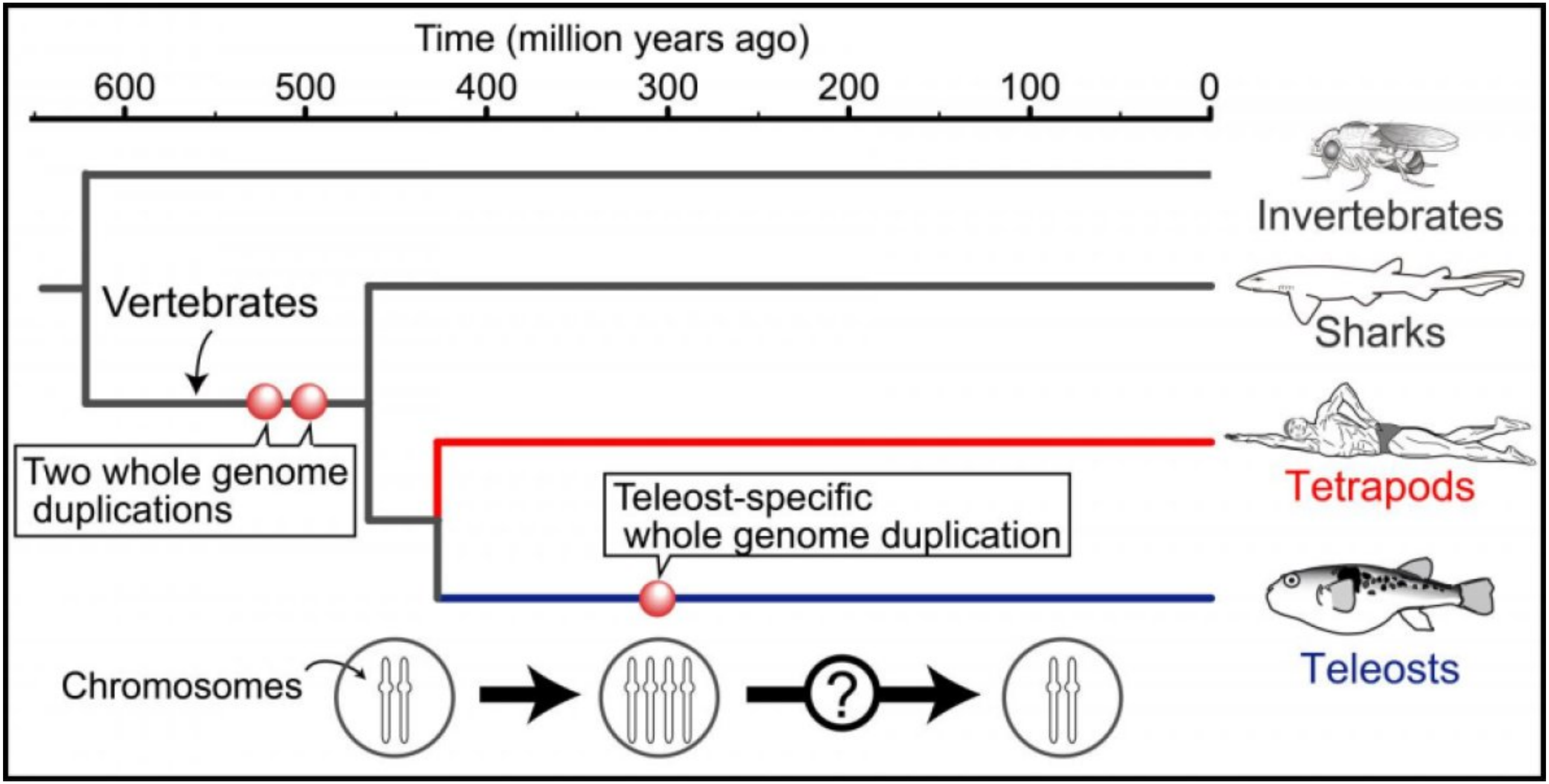


Muntiacus reevesii (2N = 46)

Muntiacus muntiacus (2N = 8)

***EVOLUTION 2e*, Figure 8.26**

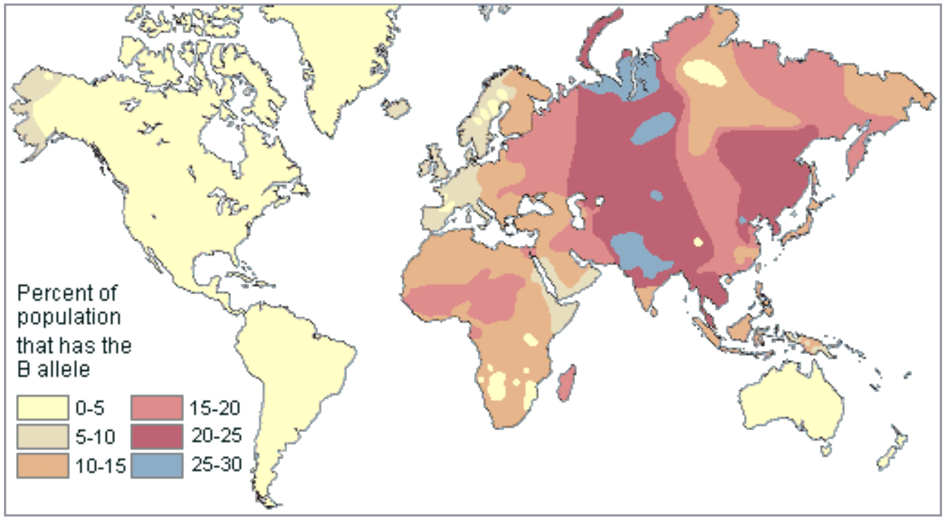
Genome Duplication

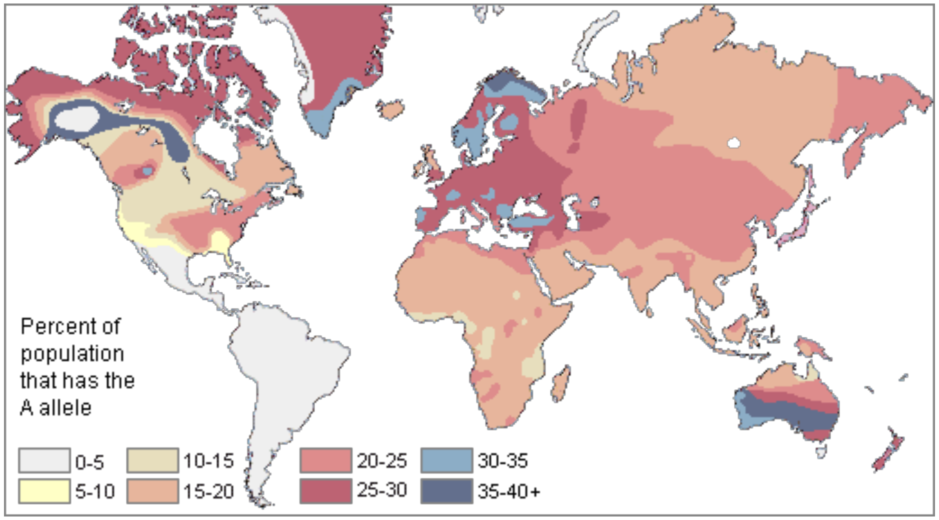


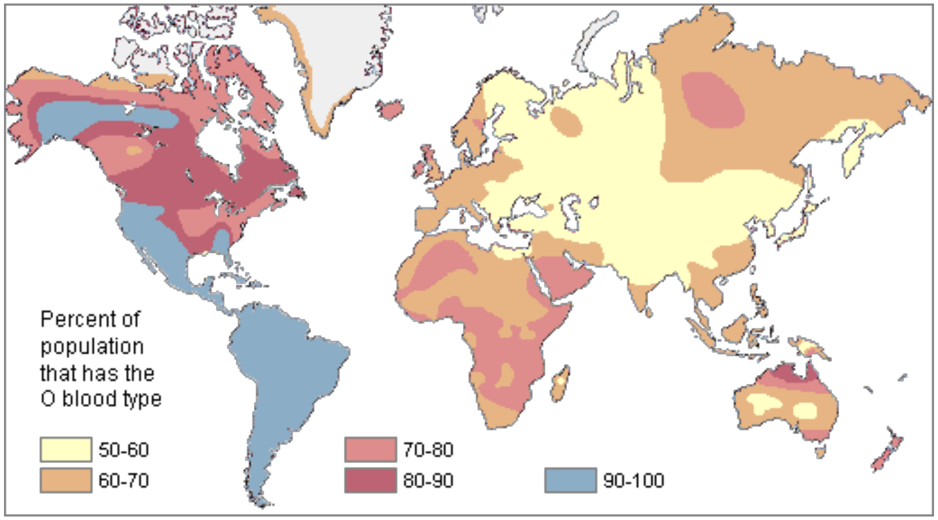
OIST

Protein variation

- ABO blood groups







Protein variation

- Allozymes: variant forms of an enzyme encoded by different alleles at the same locus
- Variation revealed using electrophoresis

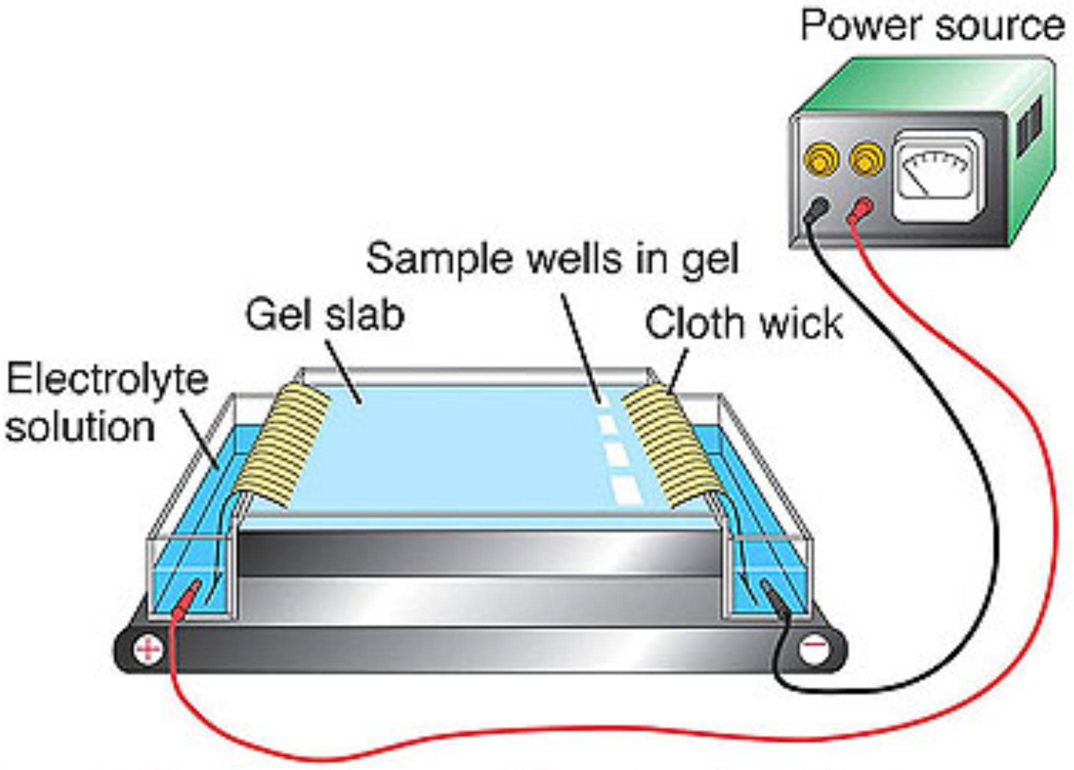
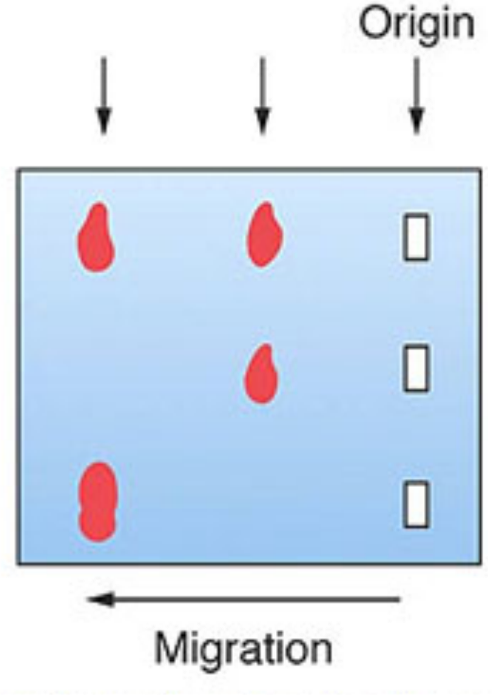


Figure 4: Schematic of devices used in protein electrophoresis

Protein variation

- Allozymes: variant forms of an enzyme encoded by different alleles at the same locus
- Variation revealed using electrophoresis

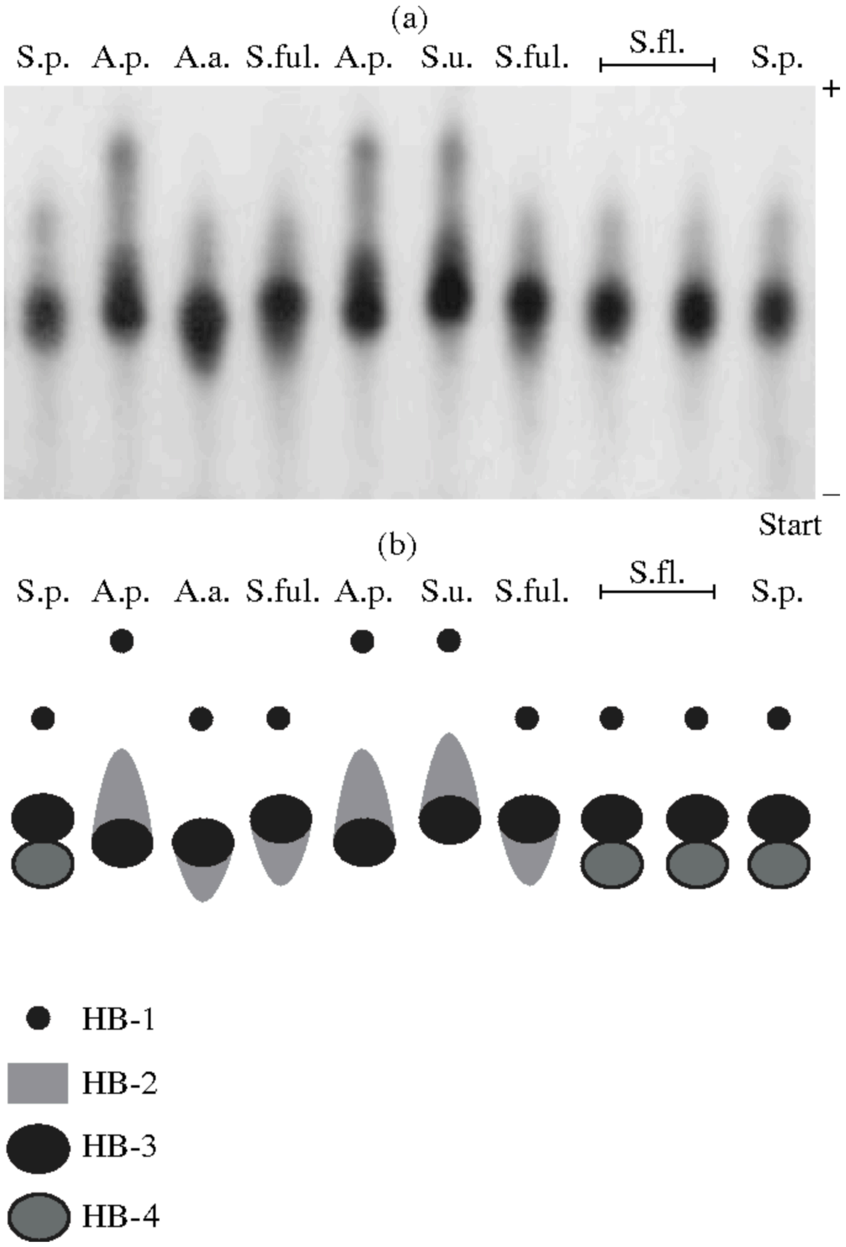


How does it work?

- Nonsynonymous mutations can change enzyme’s overall ionic charge
- Leads to differences in electrophoretic mobility

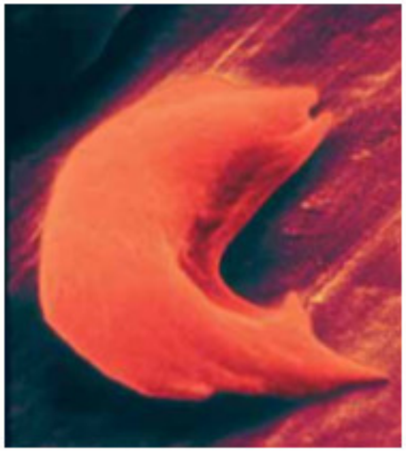
Amino Acid	3-Letter	1-Letter	Side chain polarity	Side chain charge (pH 7.4)
Alanine	Ala	A	nonpolar	neutral
Arginine	Arg	R	polar	positive
Asparagine	Asn	N	polar	neutral
Aspartic acid	Asp	D	polar	negative
Cysteine	Cys	C	nonpolar	neutral
Glutamic acid	Glu	E	polar	negative
Glutamine	Gln	Q	polar	neutral
Glycine	Gly	G	nonpolar	neutral
Histidine	His	H	polar	positive(10%), neutral(90%)
Isoleucine	Ile	I	nonpolar	neutral
Leucine	Leu	L	nonpolar	neutral
Lysine	Lys	K	polar	positive
Methionine	Met	M	nonpolar	neutral
Phenylalanine	Phe	F	nonpolar	neutral
Proline	Pro	P	nonpolar	neutral
Serine	Ser	S	polar	neutral
Threonine	Thr	T	polar	neutral
Tryptophan	Trp	W	nonpolar	neutral
Tyrosine	Tyr	Y	polar	neutral
Valine	Val	V	nonpolar	neutral

Protein variation

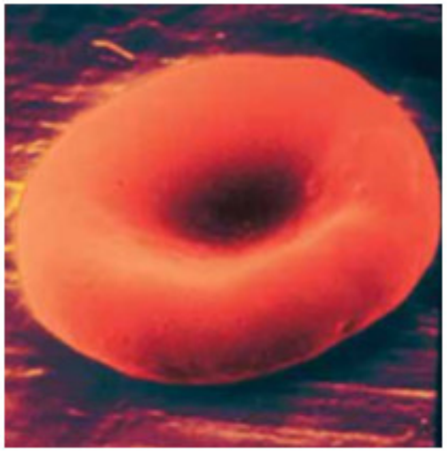


Protein variation

- Amino acid variation
  - Alternative forms of proteins arising from variation in the amino acid sequence
    - ❖ Sickle-cell disease (HbS): ONE amino acid change in beta-globin chain of hemoglobin



Sickle-cell phenotype



Normal phenotype

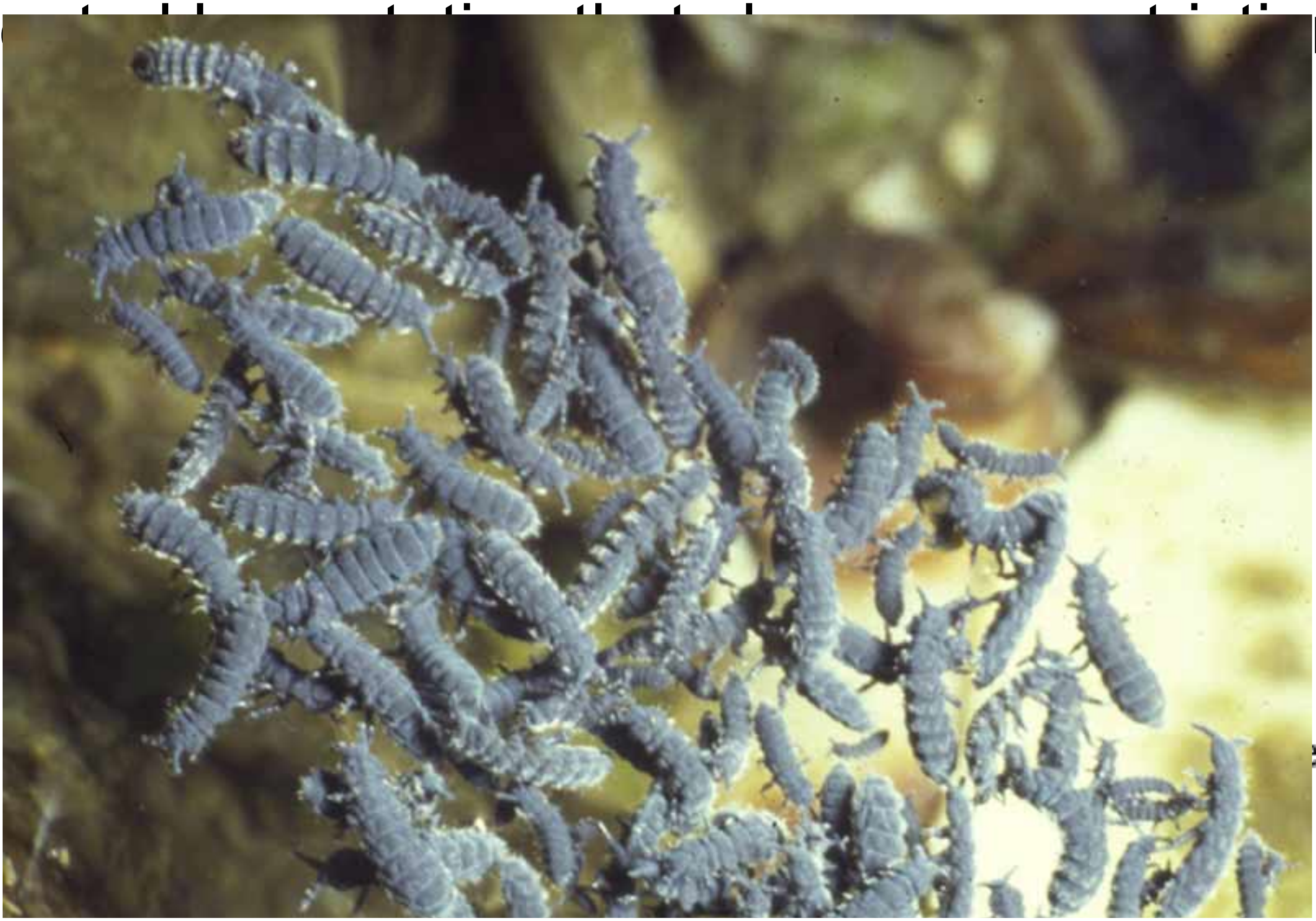
DNA variation

- RFLP: Restriction fragment length polymorphism
  - Created by mutation that changes a restriction site

```
GCCGCATTCTA          GCCGAATTCTA
CGGCGTAAGAT          CGGCTTAAGAT
```

DNA variation

- RFLP: Restriction fragment length polymorphism
  - Cr n site



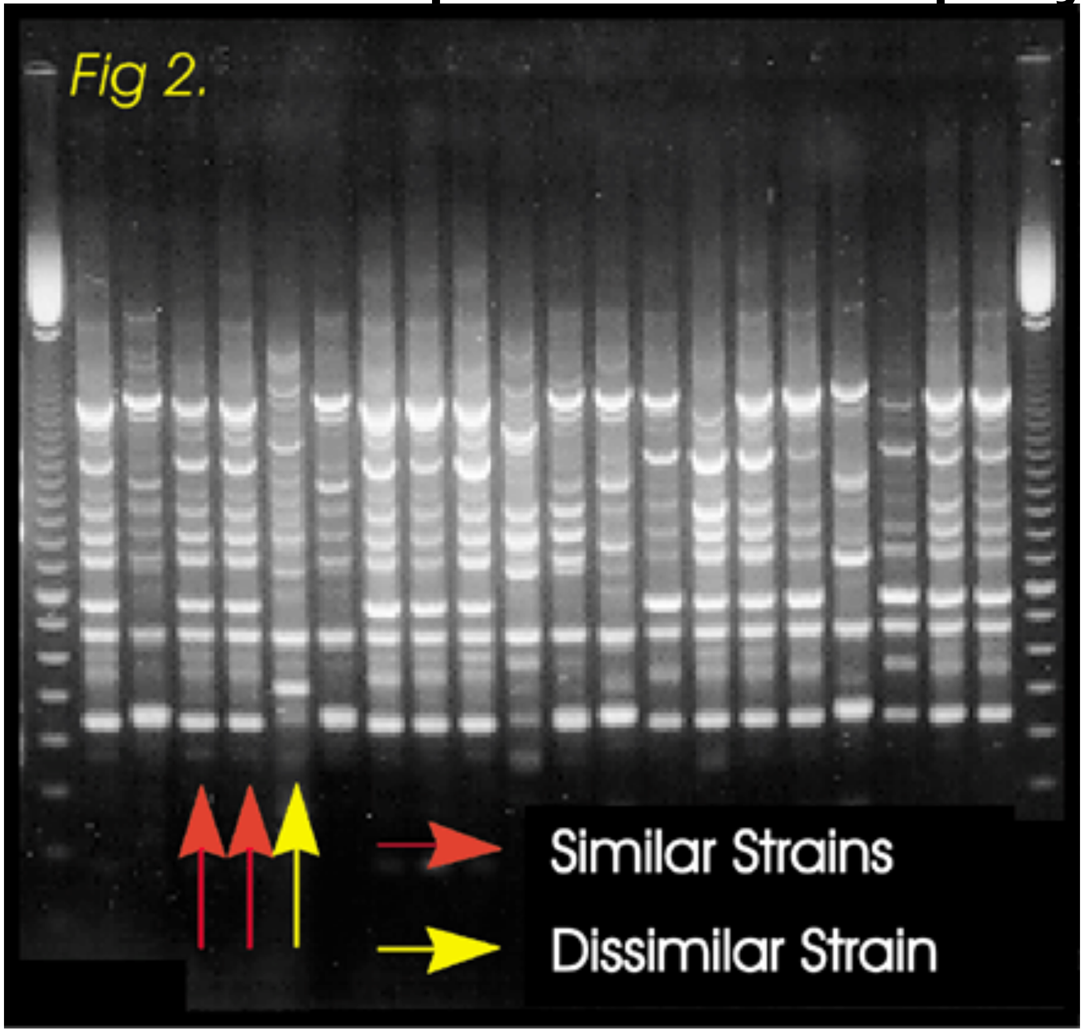
DNA variation

- RFLP
- RAPD: Random amplification of polymorphic DNA
  - Like PCR, but segments are amplified randomly
  - Employs several arbitrary, short primers
  - Need no knowledge of underlying sequence
  - Variation in RAPD profile comes from variation in primer binding sites across individuals

DNA variation

- RFLP
- RAPD: Random amplification of polymorphic DNA

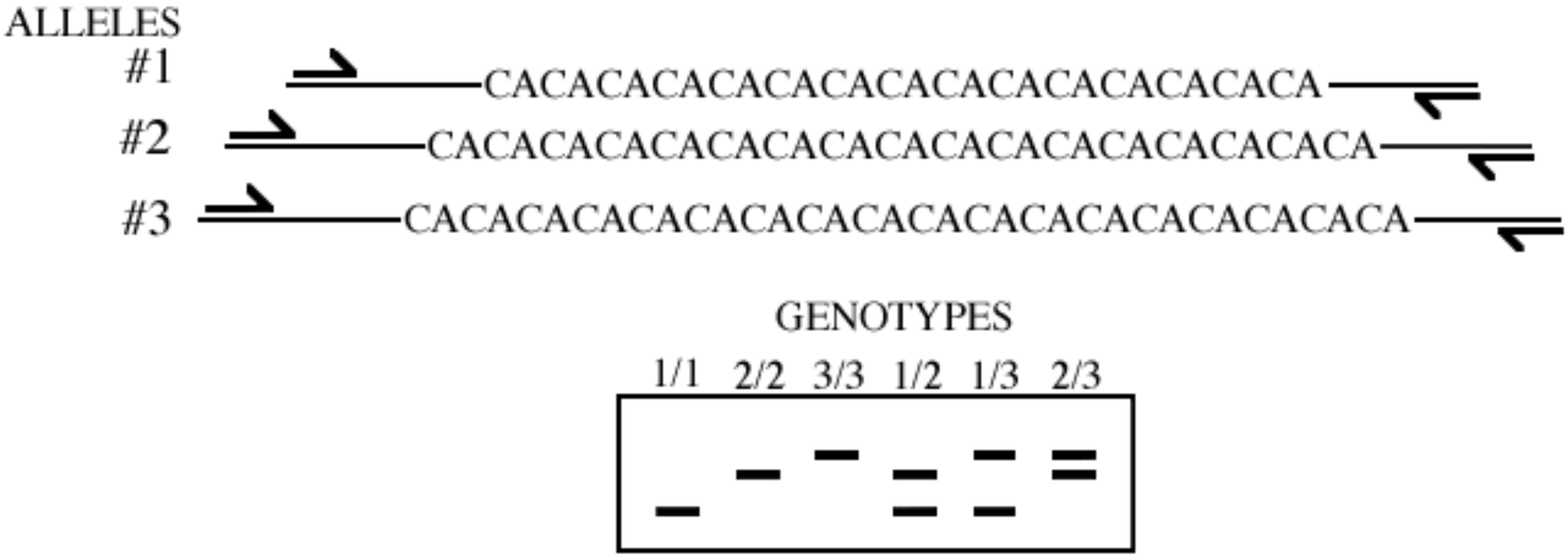
Strains of Lactobacillus from 18 types of Cheddar Cheese



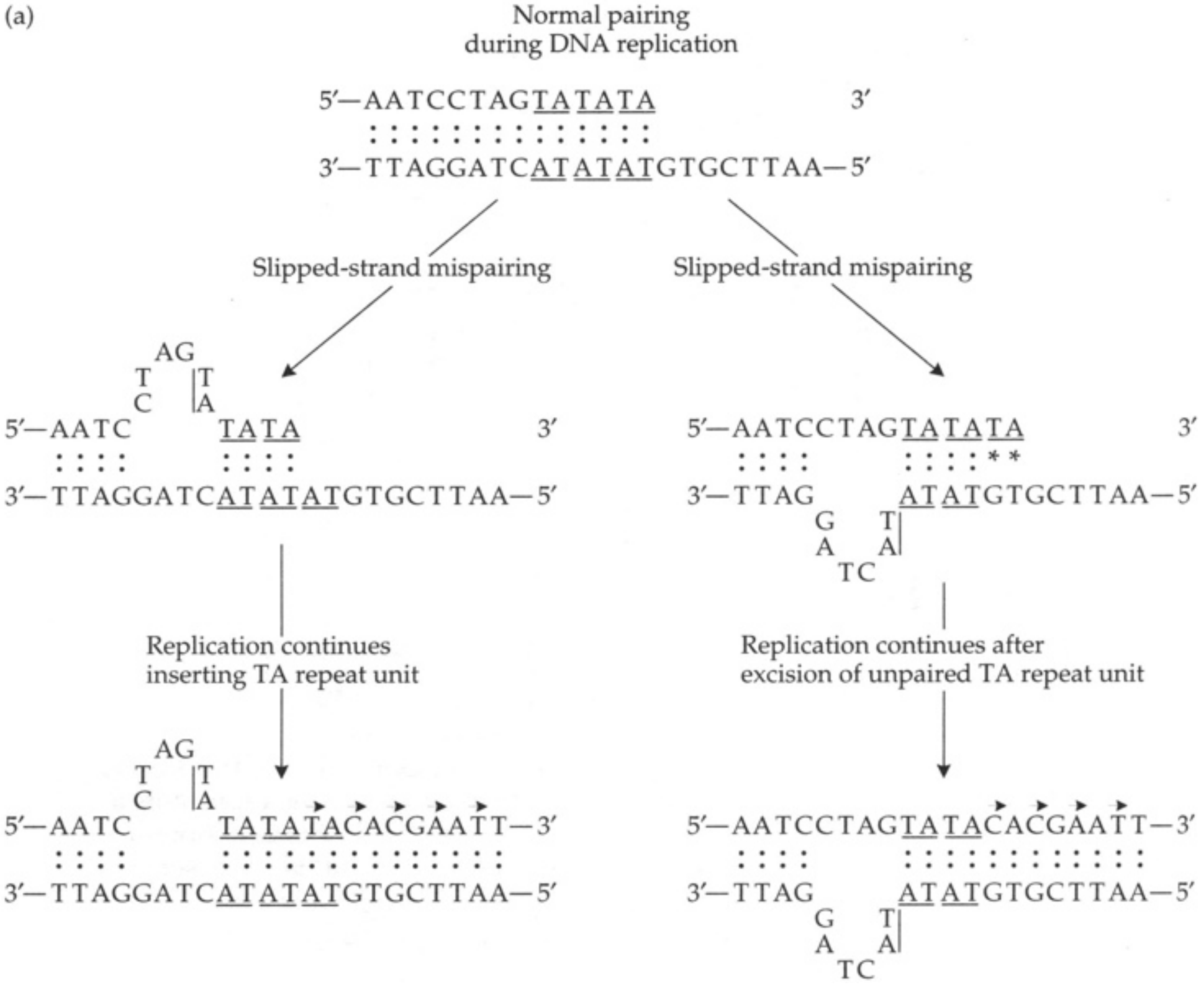
Teagasc 1998

DNA variation

- RFLP
- RAPD
- Microsatellites/minisatellites/VNTRs/SSRs: Tandemly repeated short sequences

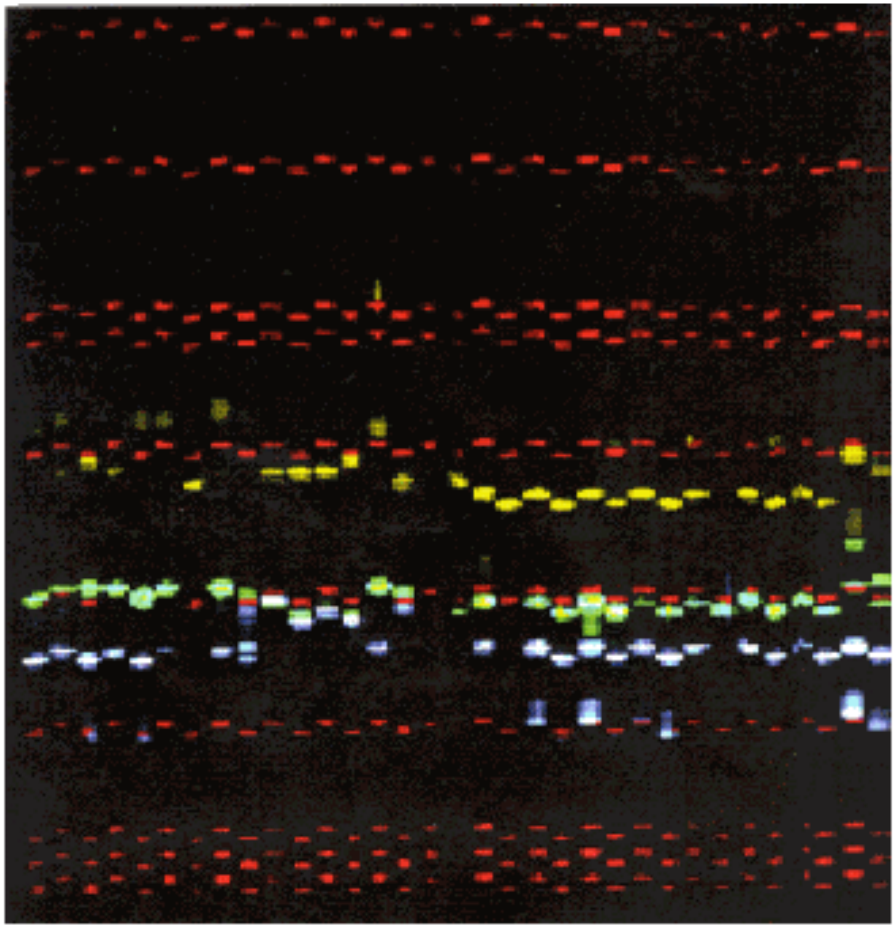


Errors in mispairing during replication and DNA repair



DNA variation

- RFLP
- RAPD
- Microsatellites/minisatellites/VNTRs/SSRs:



Leishmania
(Viannia) isolates

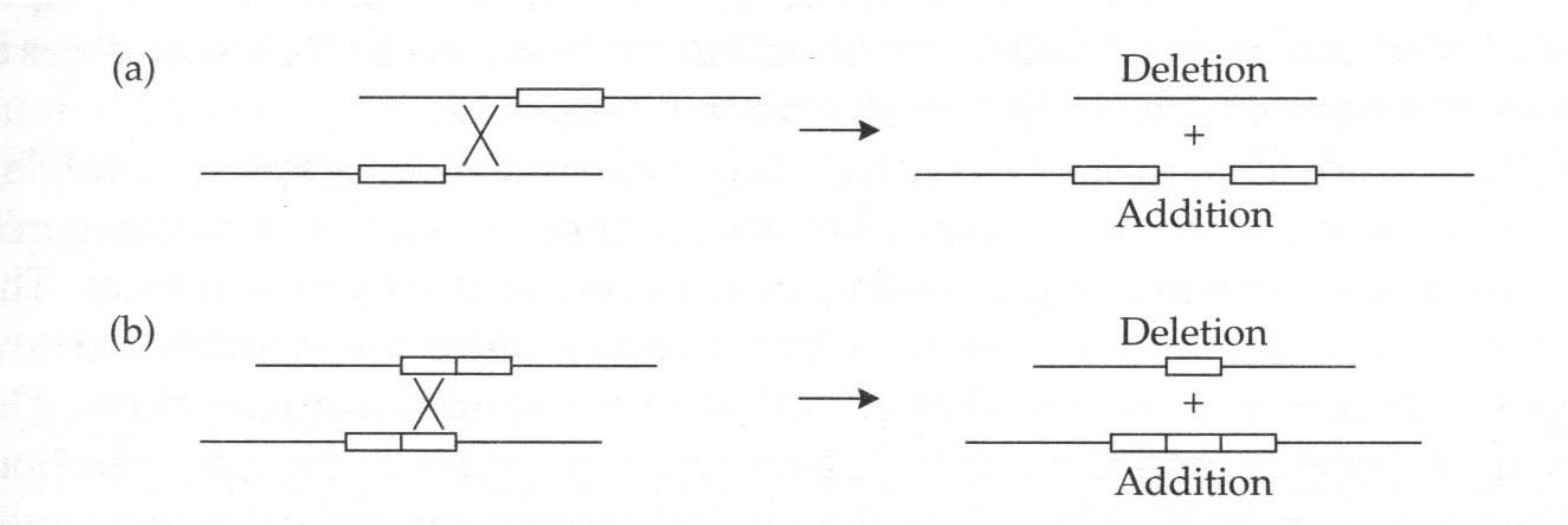
From London School of
Hygiene and Tropical
Medicine

DNA variation

- RFLP
- RAPD
- Microsatellites/minisatellites/VNTRs/SSRs
- Insertion/Deletion: Gain or loss of DNA segment

From Wei et al. 2011

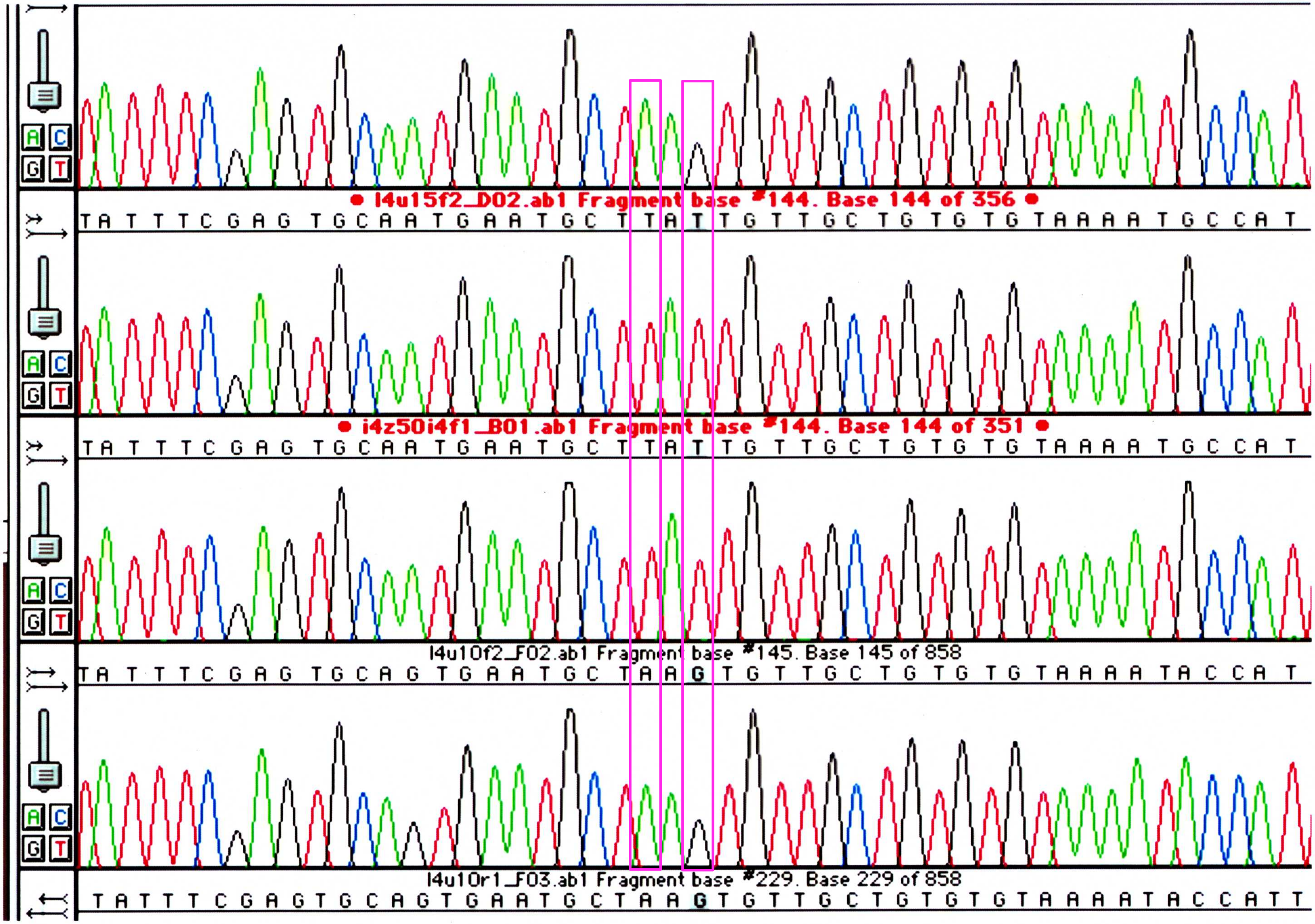
Small Scale Insertion & Deletions (InDels)



DNA variation

- RFLP
- RAPD
- Microsatellites/minisatellites/VNTRs/SSRs
- Insertion/Deletion
- Single Nucleotide Polymorphism: Differences at a single nucleotide

Sequence reads from 4 individuals



```
● l4u15f2_D02.ab1 Fragment base #144. Base 144 of 356 ●
TA T T T C G A G  T G C A A T G A A T G C  T  T A T  T G T T G C  T G  T G  T G T A A A A  T G C C A  T
● i4z50i4f1_B01.ab1 Fragment base #144. Base 144 of 351 ●
TA T T T C G A G  T G C A A  T G A A T G C  T  T A T  T G  T T G C  T G T G  T G  T A A A A  T G C C A  T
l4u10f2_F02.ab1 Fragment base #145. Base 145 of 858
TA T T T C G A G  T G C A G  T G A A T G C  T A A G  T G  T T G C  T G  T G  T G  T A A A A  T A C C A  T
l4u10r1_F03.ab1 Fragment base #229. Base 229 of 858
T A T T T C G A G T G C A G T G A A T G C T A A G T G T T G C T G T G T G T A A A A T A C C A T T
```

Discrete Genetic Variation

- Chromosomal variation
  - Inversions, chromosomes fusions/fissions
- Protein variation
  - Immunological, allozymes, amino acid variation
- DNA variation
  - RFLP, RAPD, VNTR, Indel, SNP

Where does genetic variation come from?

Mutation is the substrate of evolution

- All (genetic) polymorphisms originate with mutation
- Point mutation (one base for another)

Mutation is the substrate of evolution

- All polymorphisms originate with mutation
- Point mutation (one base for another)
- Insertion (addition of DNA)

Mutation is the substrate of evolution

- All polymorphisms originate with mutation
- Point mutation (one base for another)
- Insertion (addition of DNA)
- Deletion (loss of DNA)

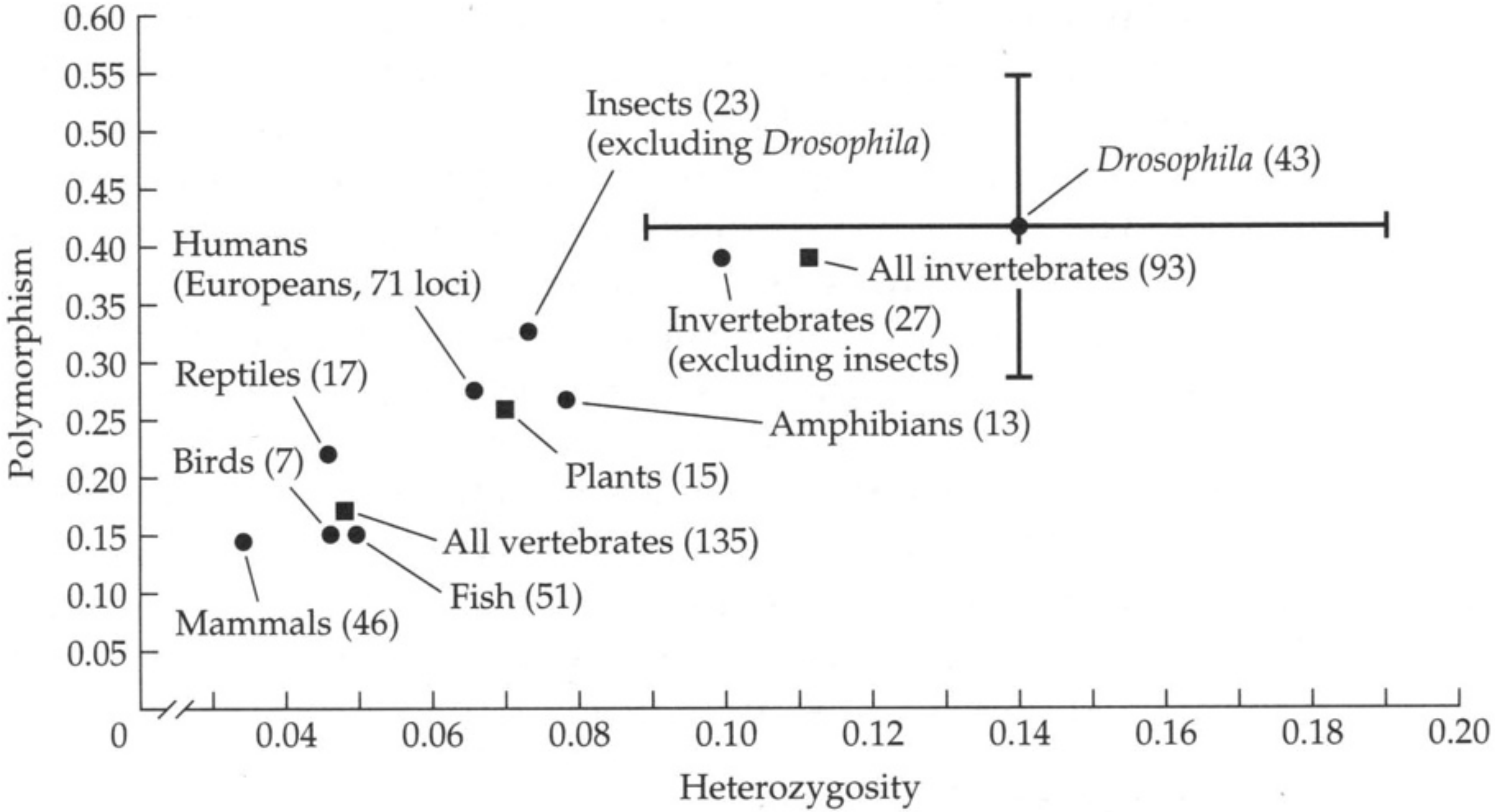
Mutation is the substrate of evolution

- All polymorphisms originate with mutation
- Point mutation (one base for another)
- Insertion (addition of DNA)
- Deletion (loss of DNA)
- Chromosomal mutations

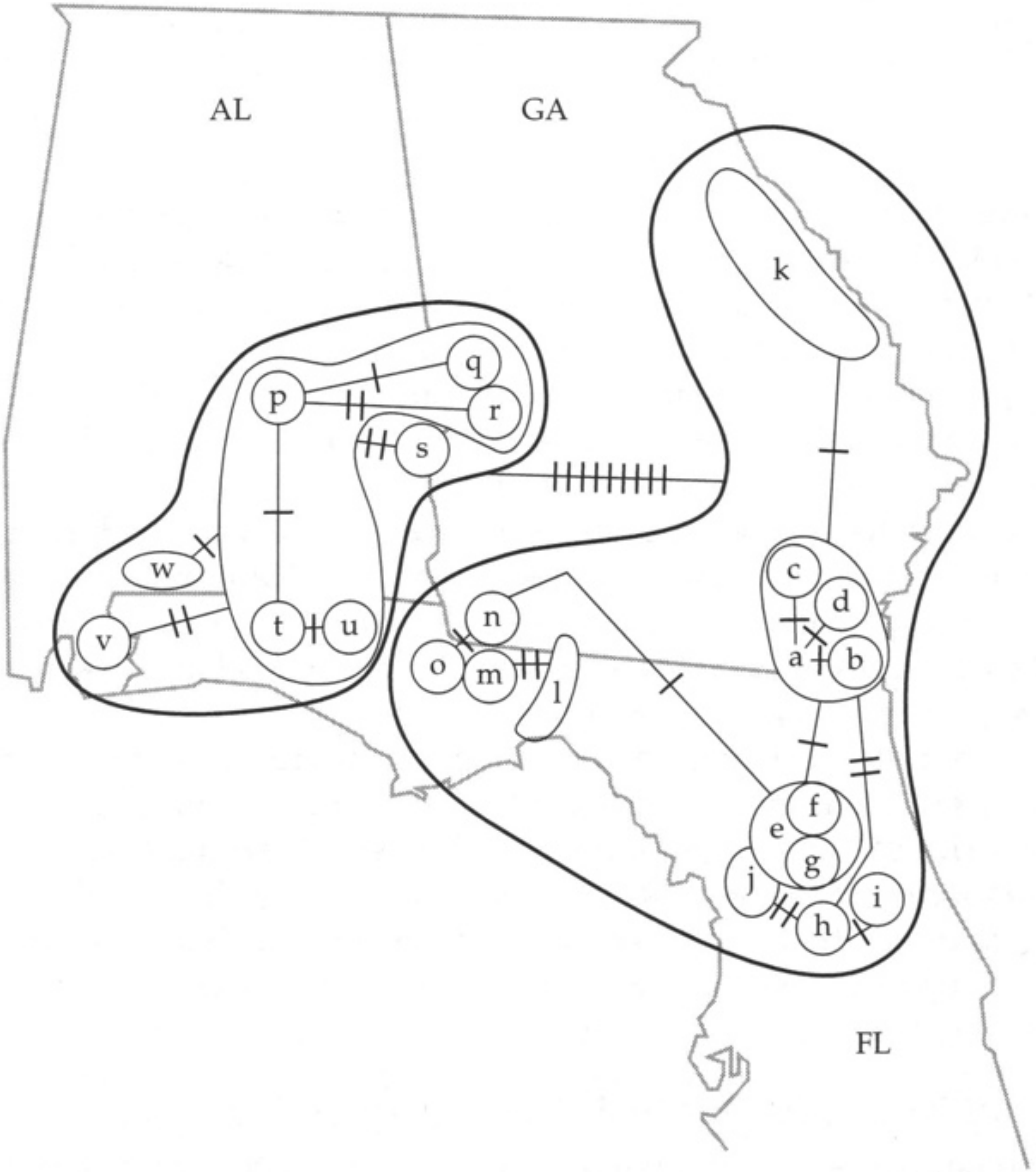
Genetic Variation

- All genetic variation originates with mutation
  - Mutation is the substrate of evolution
- All levels of organization from single base pairs to entire genomes
- Understanding genetic variation has deep implications

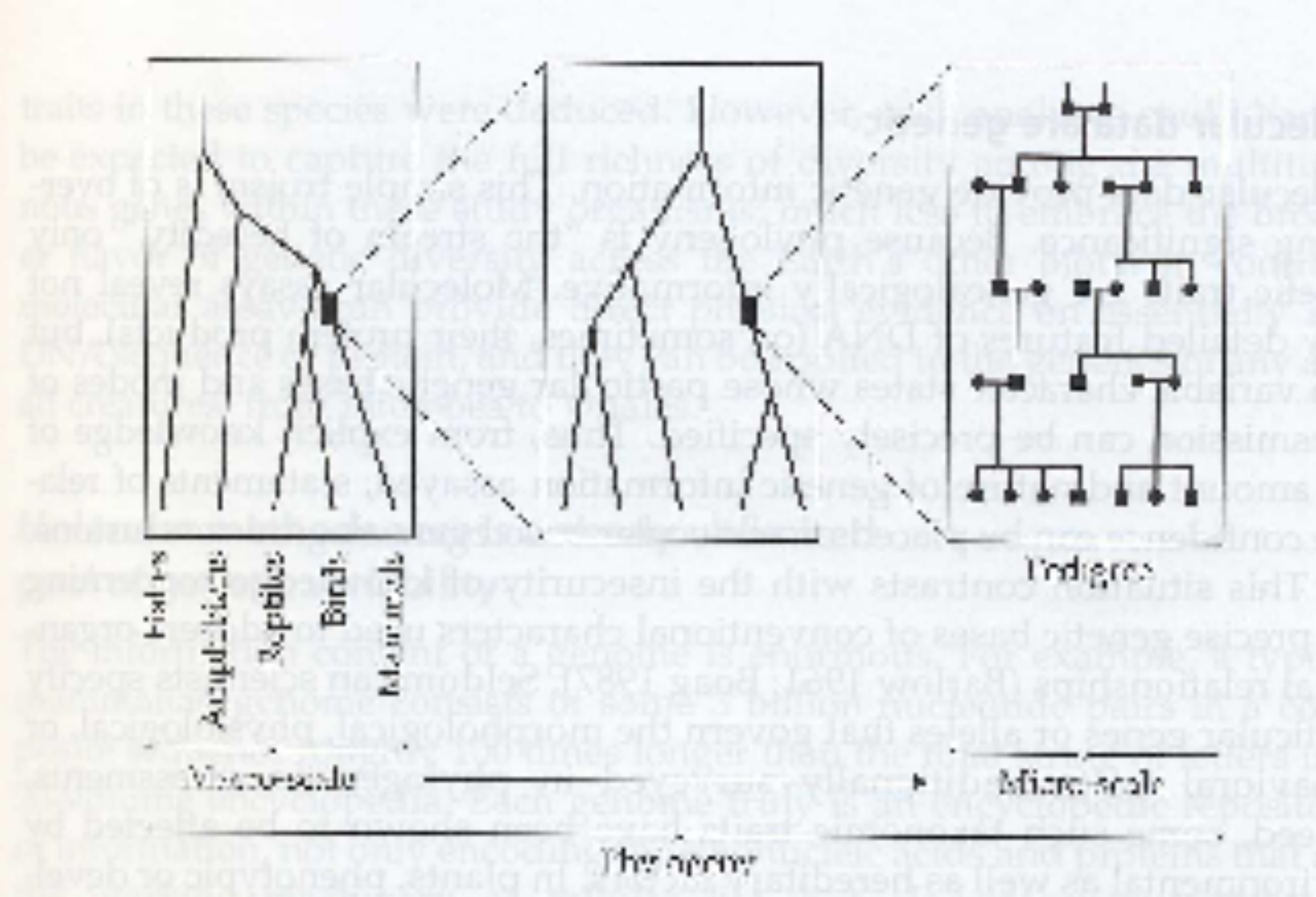
Genetic variation varies greatly among organisms

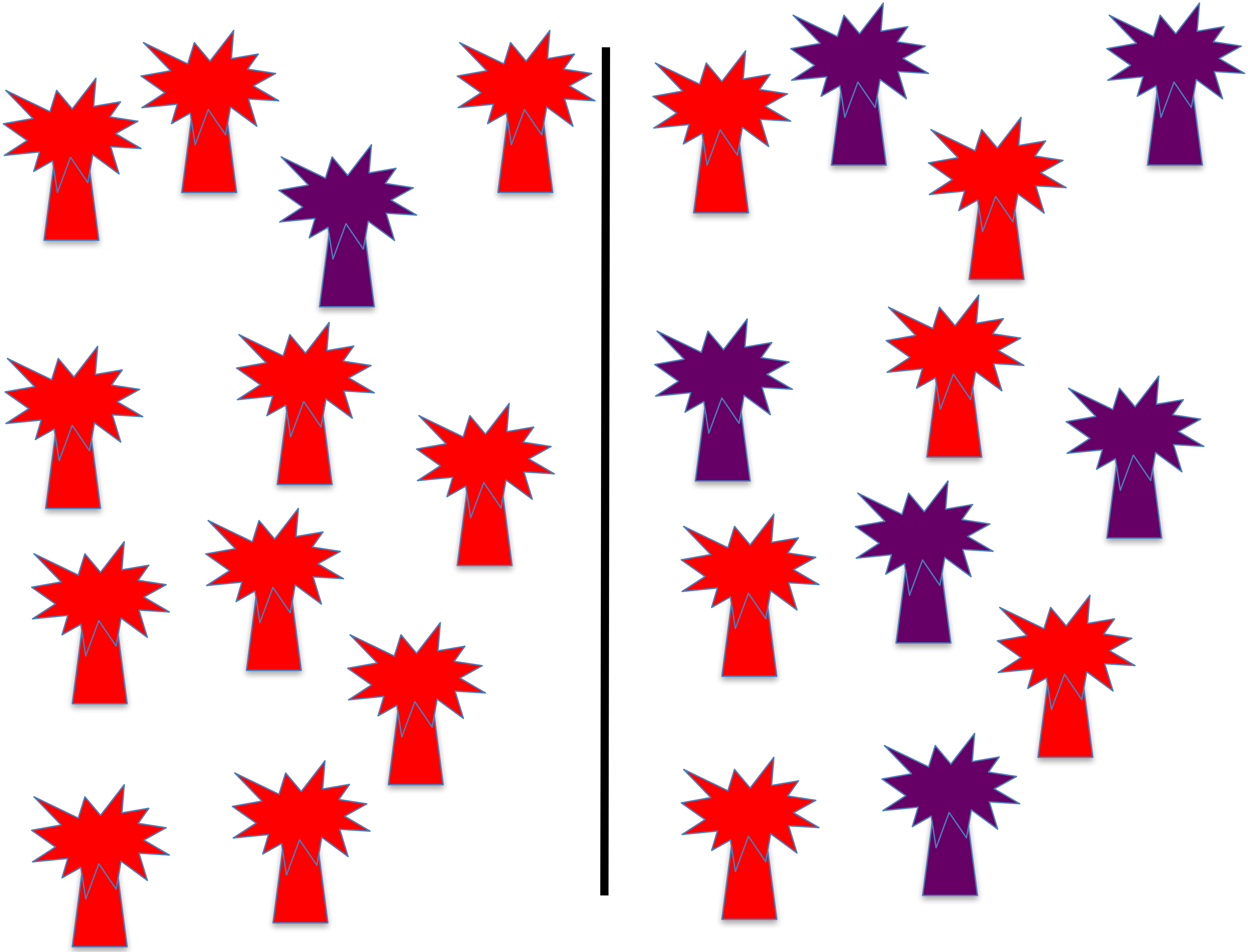


Genetic variation is distributed geographically



Genetic variation is nested phylogenetically





Summarizing molecular data

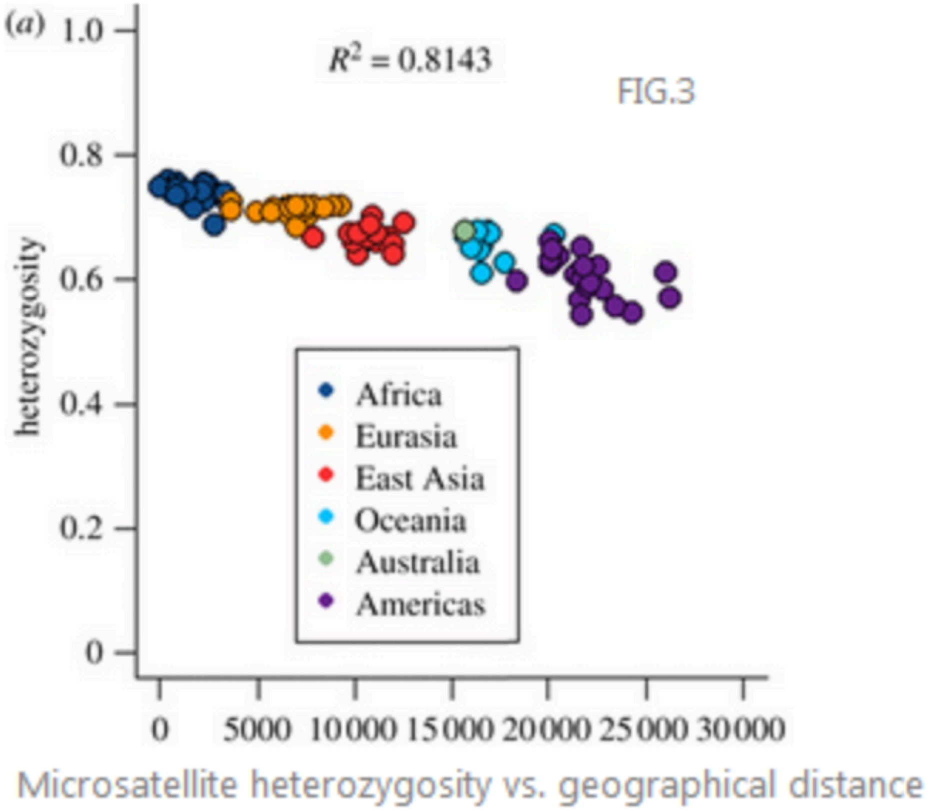
```
#1   CTGCCCACCTTTTGTTTGGGTCTTAGTCCGCAGTGCACTTGTGCCGCCGAGGGGAATGTGGTGCGTTTCCATTGTCCGGATG
#2   ....................C................................T....................C.......
#3   ..................................................................................
#4   ..A.................C................................T....................C.......
#5   ........G...................................A.....................................
#6   ..................................................................................
#7   ....................C.....................................................C.......
#8   ..............................T.............A.....................................
#9   ..............A...................................................................
#10  ..............A...................................................................
#11  ....................C...............C.....................................C.......
#12  ....................C...............C.....................................C.......
```

Key terms

- Gene
- Locus
- Allele
- Genotype
- Haplotype
- Homozygote
- Heterozygote
- Polymorphism

Heterozygosity

- Expected probability that an individual will be heterozygous at a locus
- Expected proportion of heterozygotes in a population at a locus



Hunley et al. 2012

Allelic diversity

- Number of alleles per locus

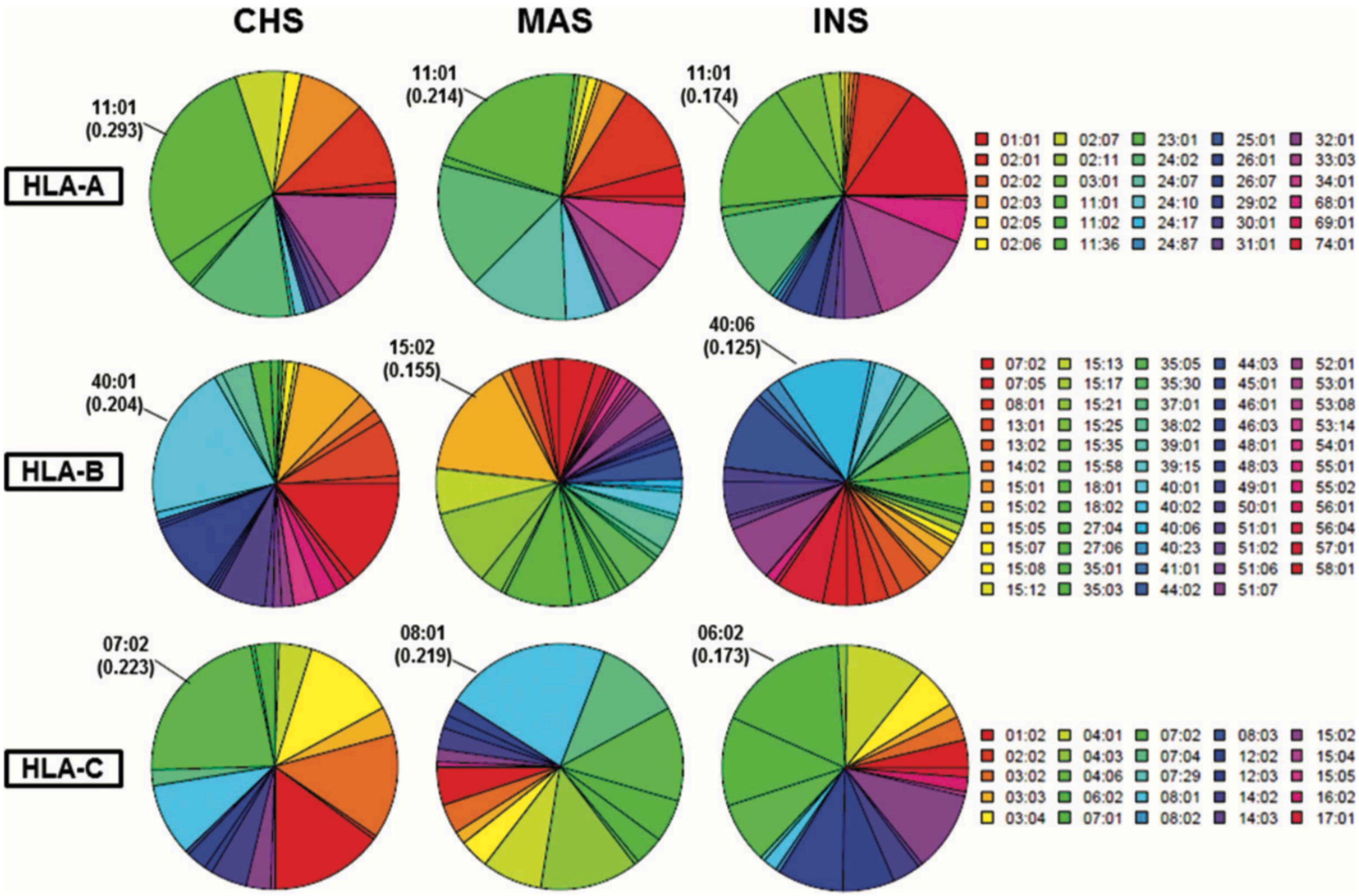
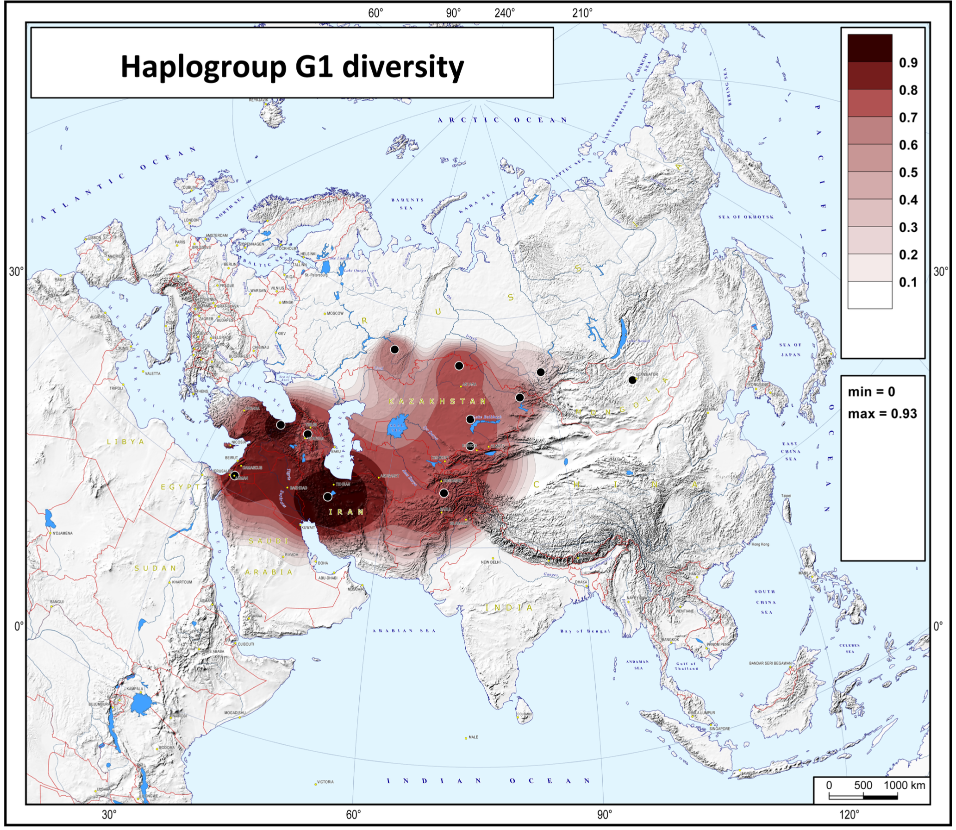


Figure 1. Allelic diversity and distribution in HLA Class I genes. Pie charts illustrating the allelic diversity of the three genes in HLA Class I in the three populations,

Pillai et al. 2014

Haplotype indices

- Haplotype number
- Unique (‘private’) haploytpes
- Haplotype diversity: probability that two randomly chosen haplotypes are different



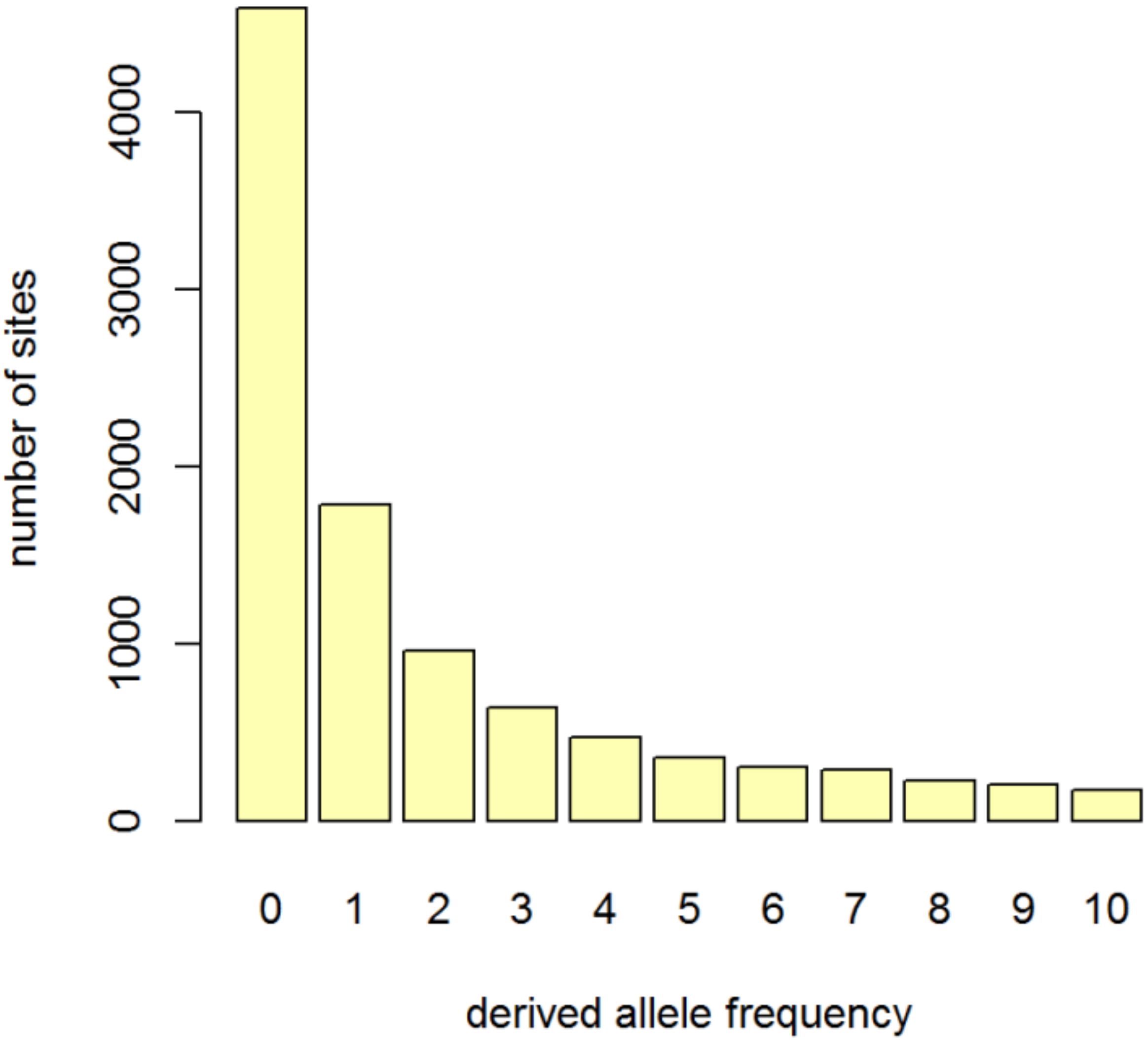
Balanovksy et al. 2015

Nucleotide-level Indices

- Θ_w
  - Watterson's estimator
  - $\frac{S}{\sum_{i=1}^{n-1}\frac{1}{i}}$
- π
  - Average pairwise difference between alleles

Site Frequency Spectrum

5 diploid individuals
10,000



Genetic Variation: Software

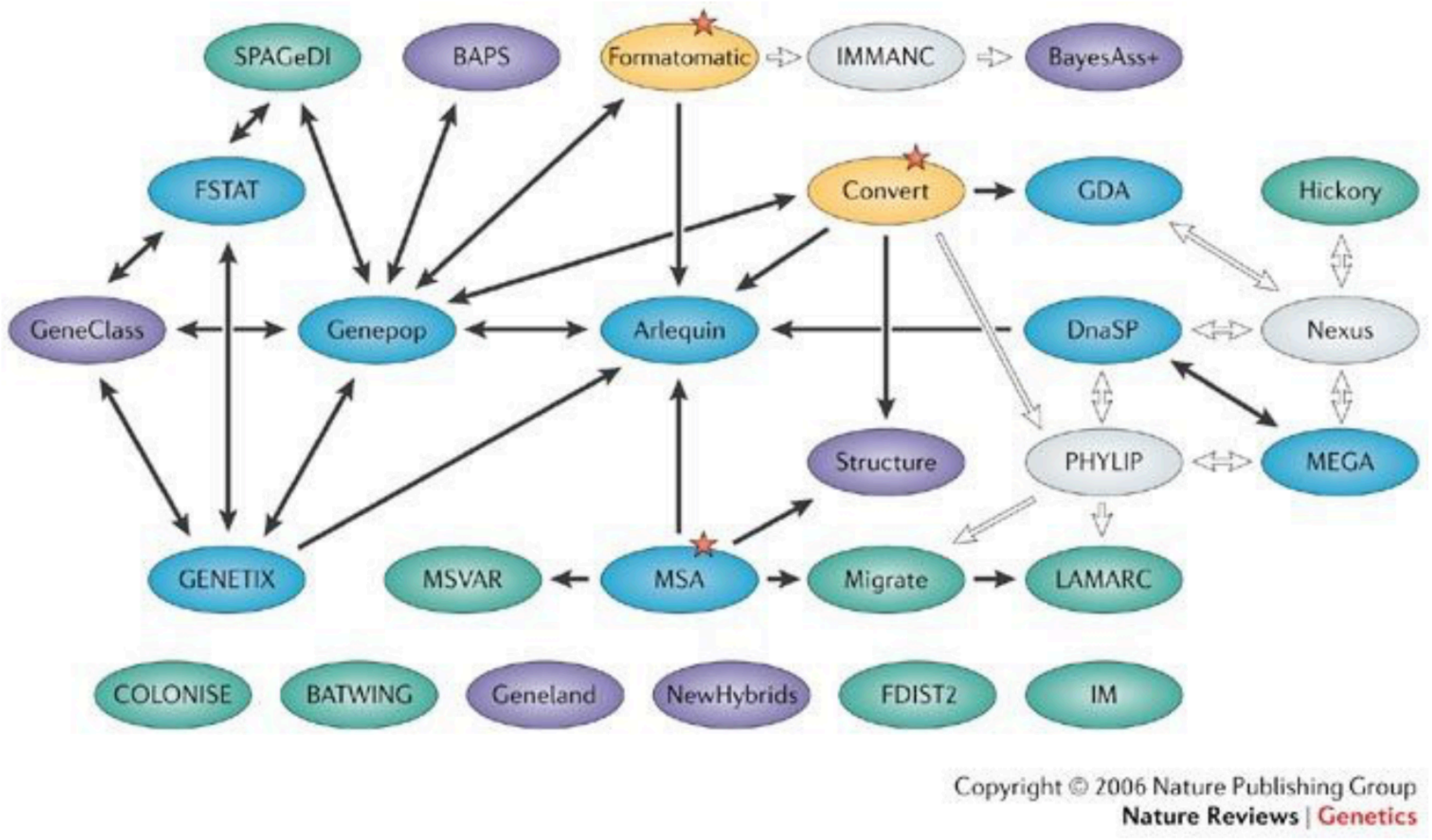
- DnaSP (DNA sequence polymorphism)
- DNA sequence
- Basic population genetic analyses
- Estimate genetic diversity
- Within/Between populations
- LD, recombination, gene flow, gene conversion
- Neutrality tests

Genetic Variation: Software

- Arlequin
- MEGA
- PHYLIP
- …
- Excoffier and Heckel 2006: Computer programs for population genetic data analysis: a survival guide

Genetic Variation: Software

From: Computer programs for population genetics data analysis: a survival guide



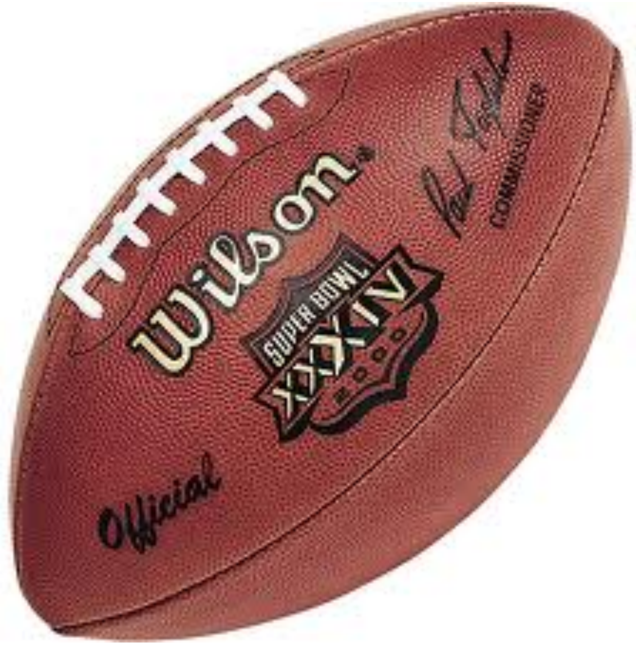
Nature Reviews | Genetics

Why do we care about genetic variation?

- Genetic variation underlies phenotypic differences among individuals
  - Including disease risk and responses to drugs and environmental factors

Why do we care about genetic variation?

- Genetic variation underlies phenotypic differences among individuals
  - Including disease risk and responses to drugs and environmental factors
- Individual identification



Why do we care about genetic variation?

- Genetic variation underlies phenotypic differences among individuals
  - Including disease risk and responses to drugs and environmental factors
- Individual identification
- Manage resources

Why do we care about genetic variation?

- Genetic variation underlies phenotypic differences among individuals
  - Including disease risk and responses to drugs and environmental factors
- Individual identification
- Manage resources
- Public Health

Why do we care about genetic variation?

- Genetic variation underlies phenotypic differences among individuals
  - Including disease risk and responses to drugs and environmental factors
- Individual identification
- Manage resources
- Public Health
- Improve plant and animal food products
- Understand genetic basis of disease and other complex phenotypes including behaviours
- Insights into evolutionary history, ancestry

Conservation genetics

- Use of genetic theories and techniques to study the loss and restoration of genetic diversity
- Derived from ecological, evolutionary, and quantitative genetics
- Focuses on small populations

Why is conservation important?

- Sixth mass extinction
- Aesthetic value
- Bioresources
- Ecosystem services
- Ethical, moral considerations

How can genetics help?

- Minimize inbreeding and loss of genetic variation
- Identify populations of concern
- Resolve population structure
- Resolve taxonomic uncertainty
- Define management units
- Detect hybridization
- Detect and define invasive species
- Estimate population size and sex ratio
- Establish parentage
- Understand population connectivity
- Aid in management
- Enhance reproductive capacity of organisms
- How do genetic factors affect extinction risk?
- How can we best genetically rescue populations?
- What can we use genetics to identify hotspots?
- Applied genetic detective work
- …

Amato et al. 2009

Future of conservation genetics

- Open questions:
  - Genetic variation and fitness
  - Mechanism connecting genetic variation and fitness
  - Genotype by environment interaction
  - Intermediate phenotypes
- New technologies
  - DNA sequencing