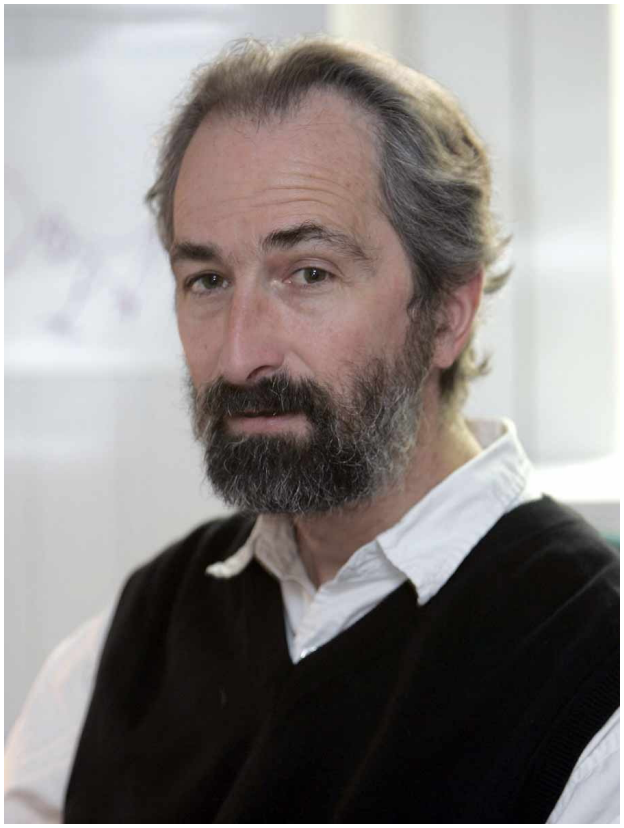
Genetic Variation:
What is it and why is it important?

“Nothing in biology makes sense except in light of evolution”

Theodosius Dobzhansky 1973

“Nothing in evolution makes sense except in light of population genetics”

Michael Lynch 2007



Population genetics

- The study of distribution and change in allele frequencies and genotype frequencies over time

Population genetics

- The study of distribution and change in allele frequencies and genotype frequencies over time
  - Evolution defined as change in allele frequency in gene pool
  - Influenced by:
    - ➢Natural Selection
    - ➢Random Genetic Drift
    - ➢Mutation
    - ➢Gene Flow
    - ➢Recombination
    - ➢Population structure

Population genetics

- The study of distribution and change in allele frequencies and genotype frequencies over time
  - Evolution defined as change in allele frequency in gene pool
  - Influenced by:
    - ➢Natural Selection
    - ➢Random Genetic Drift
    - ➢Mutation
    - ➢Gene Flow
    - ➢Recombination
    - ➢Population structure

Population genetics

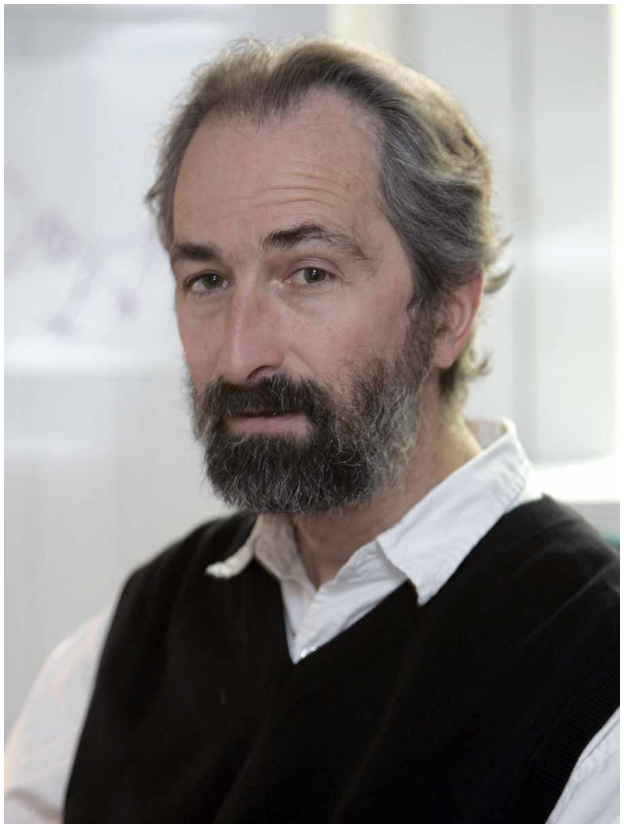
- The study of distribution and change in allele frequencies and genotype frequencies over time
  - Evolution defined as change in allele frequency in gene pool
  - Influenced by:
    - ➢Natural Selection
    - ➢Random Genetic Drift
    - ➢Mutation
    - ➢Gene Flow
    - ➢Recombination
    - ➢Population structure
- Goal: To determine the genetic basis of evolution

“Nothing in biology makes sense except in light of evolution”

Theodosius Dobzhansky 1973

“Nothing in evolution makes sense except in light of population genetics”

Michael Lynch 2007



Types of Variation

- Continuous variation: Complete range of measurements from one extreme to the other
  - Color

Types of Variation

- Continuous variation: Complete range of measurements from one extreme to the other
  - Color

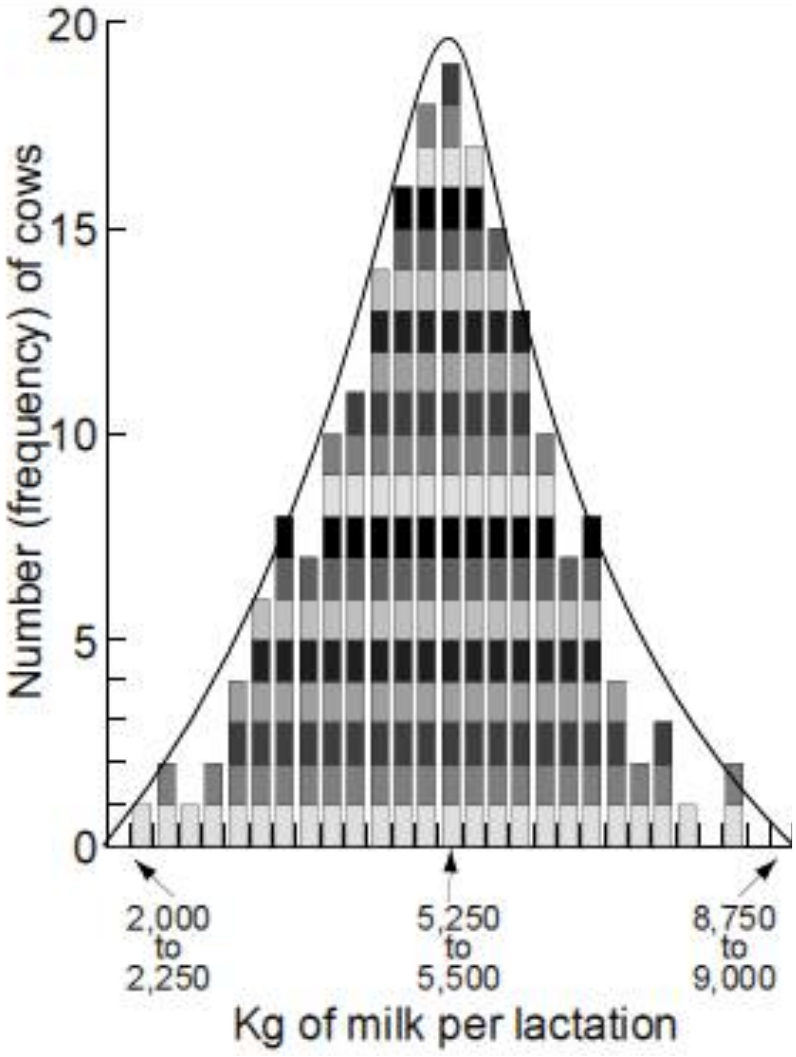
Donax variabilis

Types of Variation

- Continuous variation: Complete range of measurements from one extreme to the other
  - Color
  - Height

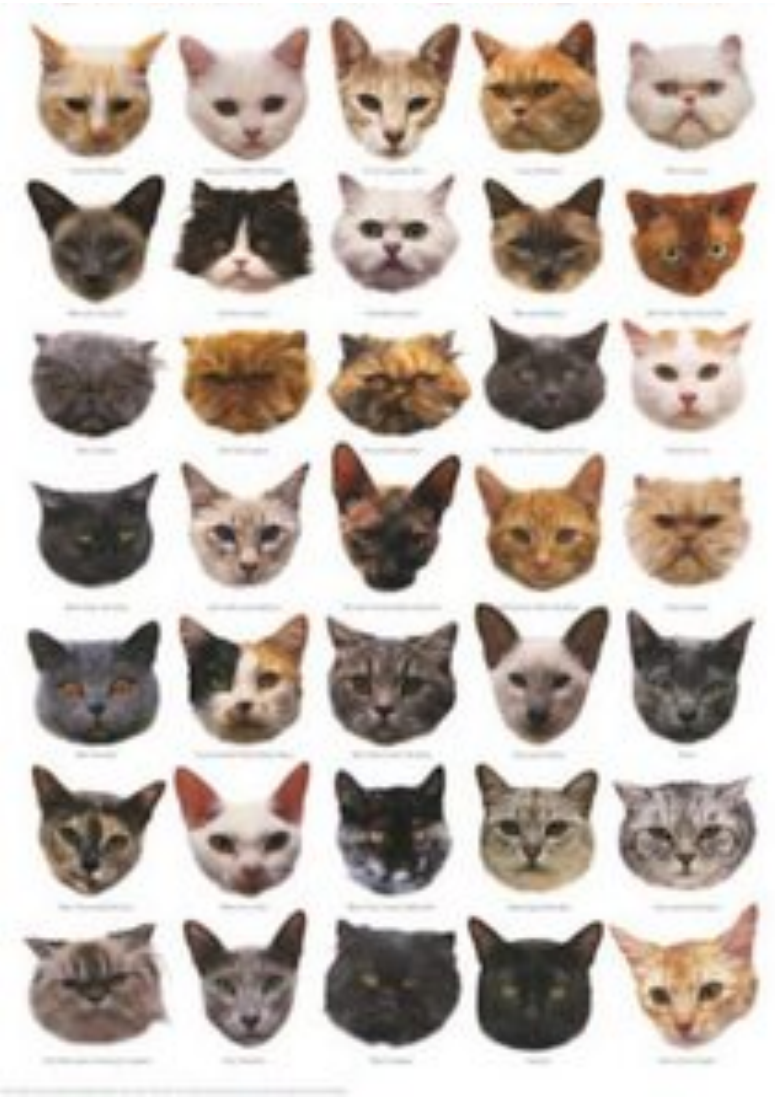
Types of Variation

- Continuous variation: Complete range of measurements from one extreme to the other
  - Color
  - Height
  - Milk production in cows



Types of Variation

- Continuous variation: Complete range of measurements from one extreme to the other
  - Color
  - Height
  - Milk production in cows
  - …



Types of Variation

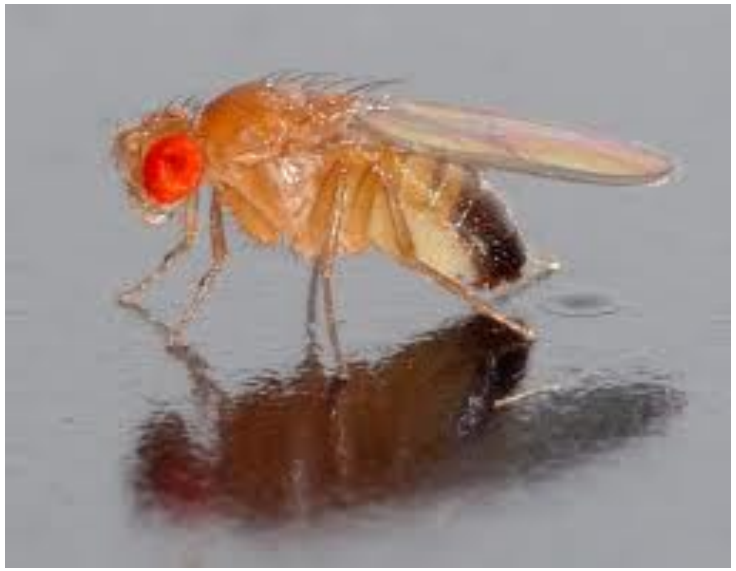
- Continuous variation: Complete range of measurements from one extreme to the other
- Discrete variation: Individuals fall into a number of distinct classes or categories

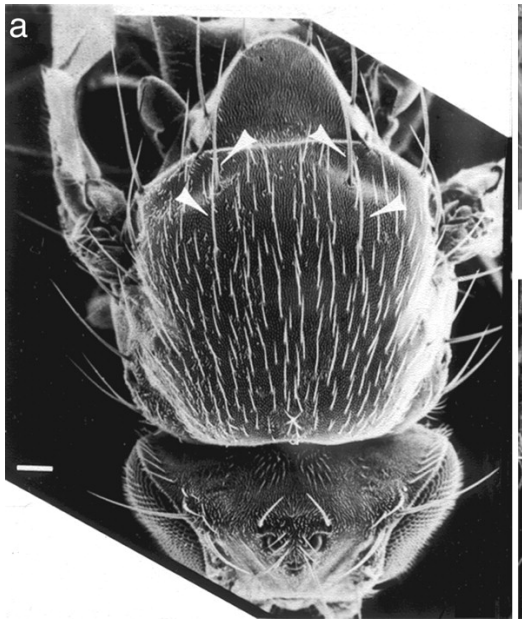
Types of Variation

- Continuous variation: Complete range of measurements from one extreme to the other
- Discrete variation: Individuals fall into a number of distinct classes or categories
  - Bristle number in *D. melanogaster*

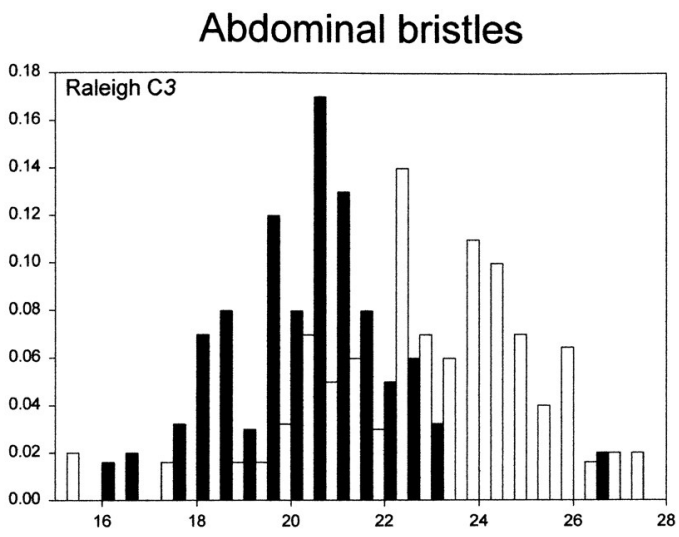
Types of Variation

- Continuous variation: Complete range of measurements from one extreme to the other
- Discrete variation: Individuals fall into a number of distinct classes or categories
  - Bristle number in *D. melanogaster*





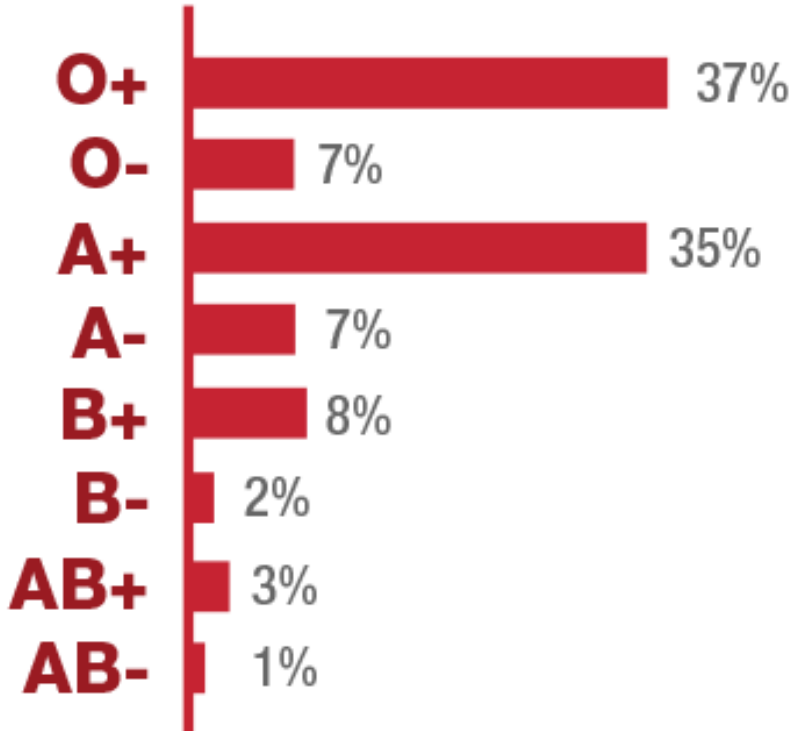
Tilney L G & DeRosier D J 2005



Lyman & Mackay 1998

Types of Variation

- Continuous variation: Complete range of measurements from one extreme to the other
- Discrete variation: Individuals fall into a number of distinct classes or categories
  - Bristle number in *D. melanogaster*
  - Human blood groups



Types of Variation

- Continuous variation: Complete range of measurements from one extreme to the other
- Discrete variation: Individuals fall into a number of distinct classes or categories

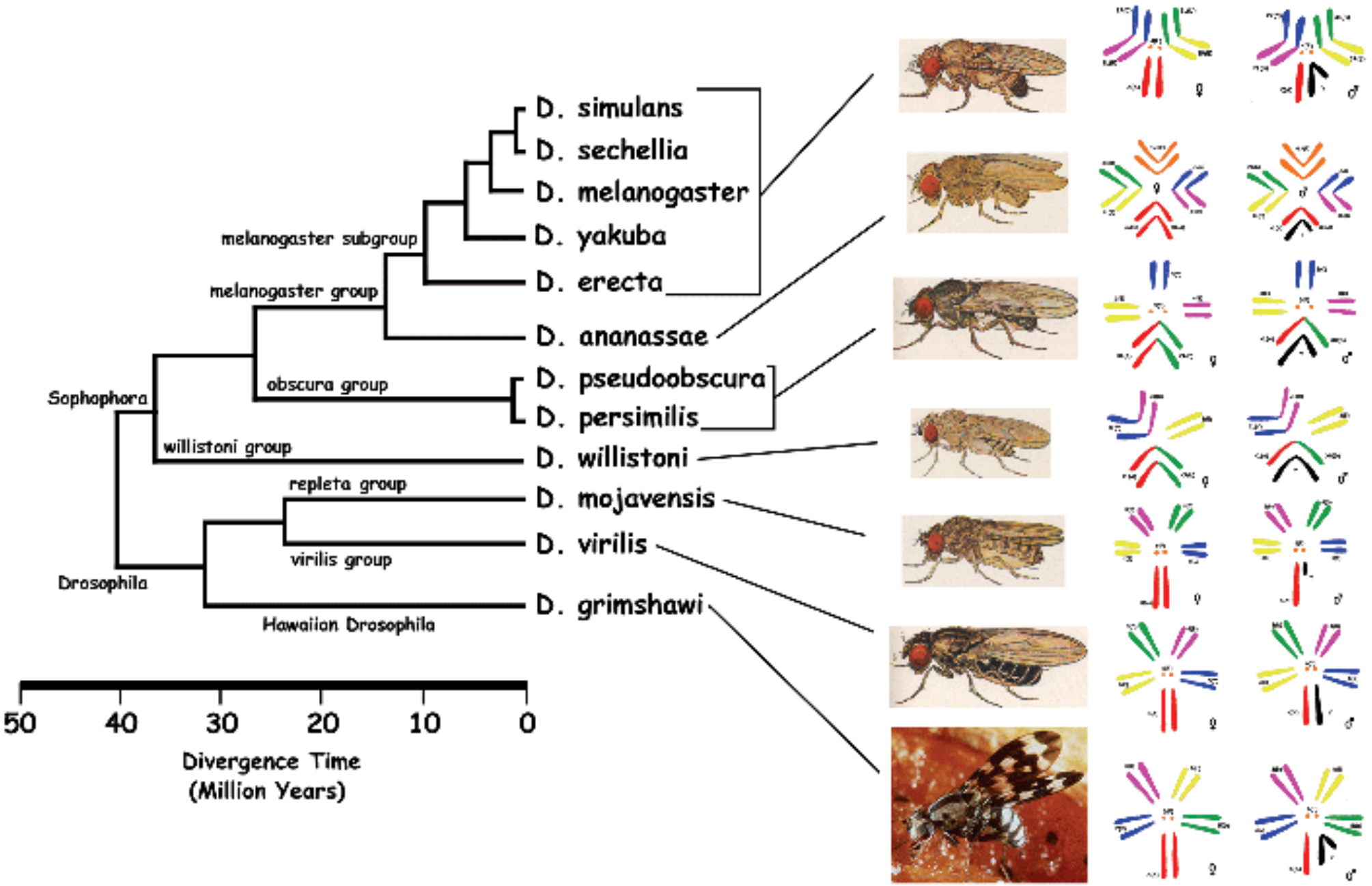
Discrete Genetic Variation

- Chromosomal variation
- Protein variation
- DNA variation

Chromosomal Variation

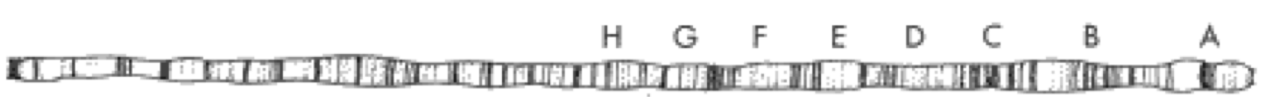
- Variation in chromosome number, gene number, gene order etc.

Chromosomal Variation

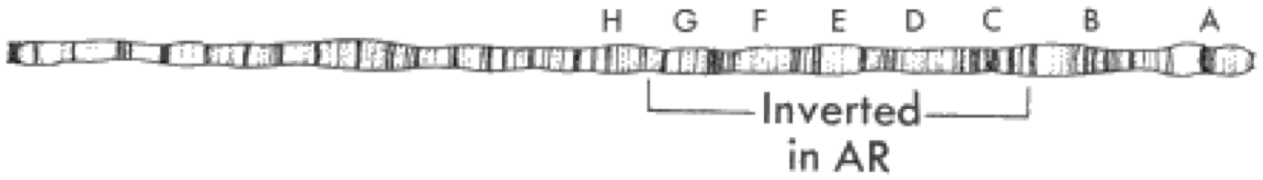


Chromosomal Variation

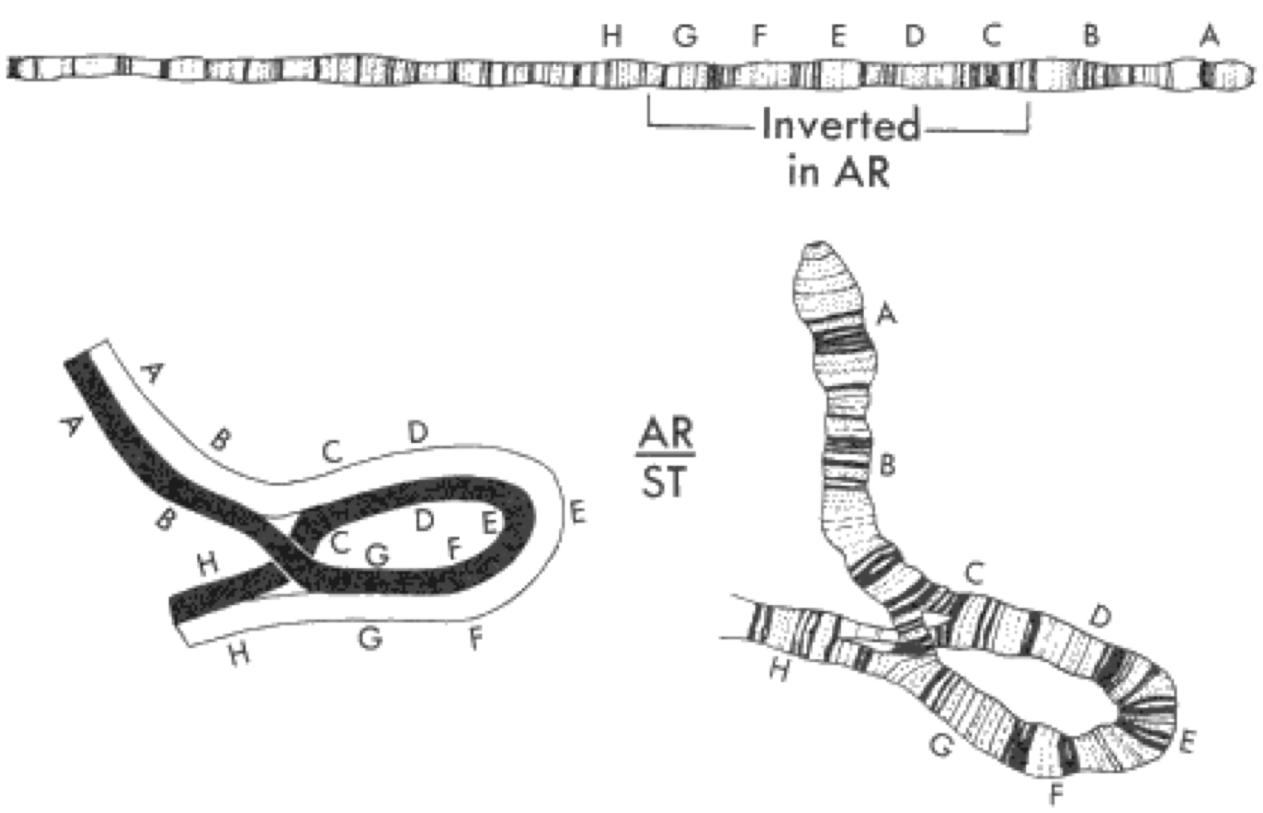
Chromosomal Variation



Chromosomal Variation



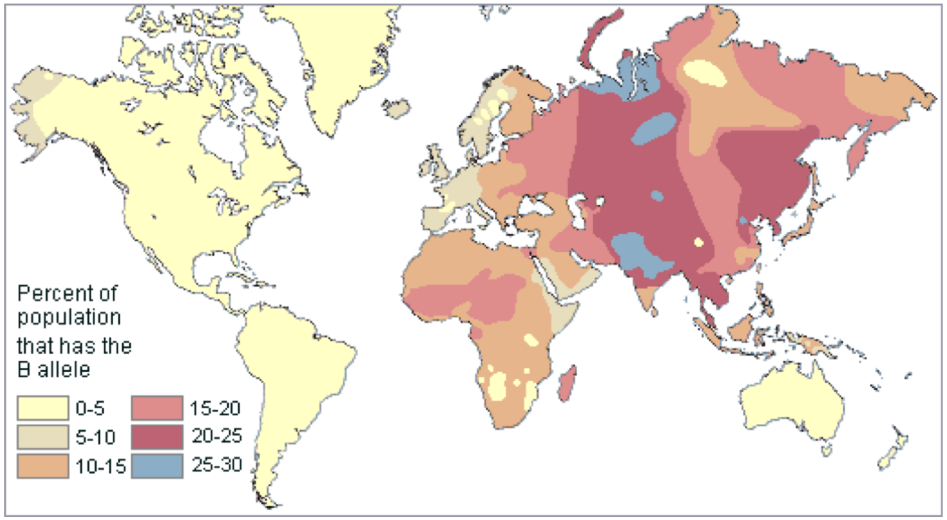
Chromosomal Variation

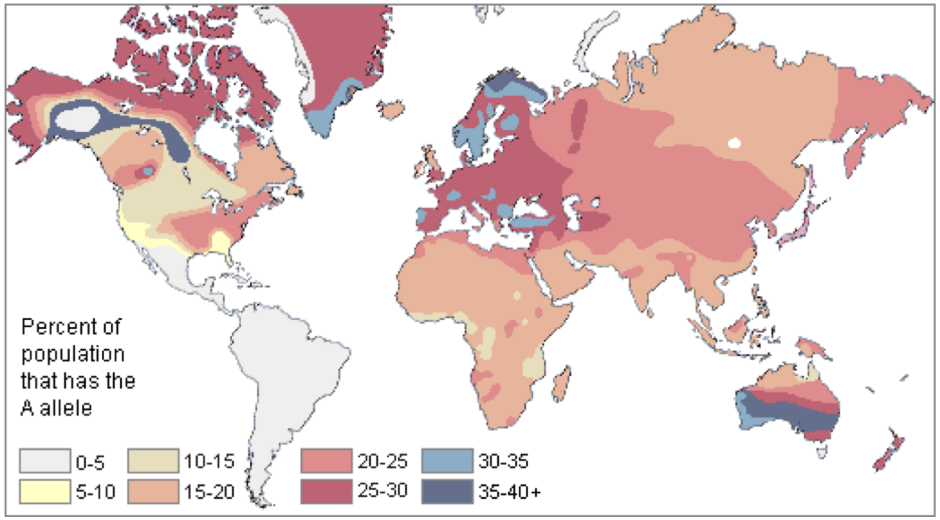


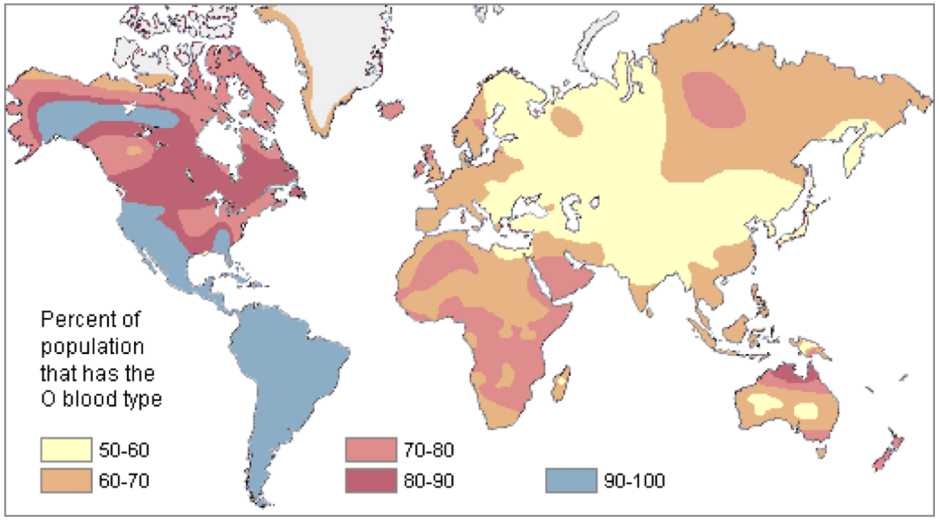
From Dobzhansky and Sturtevant 1938

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

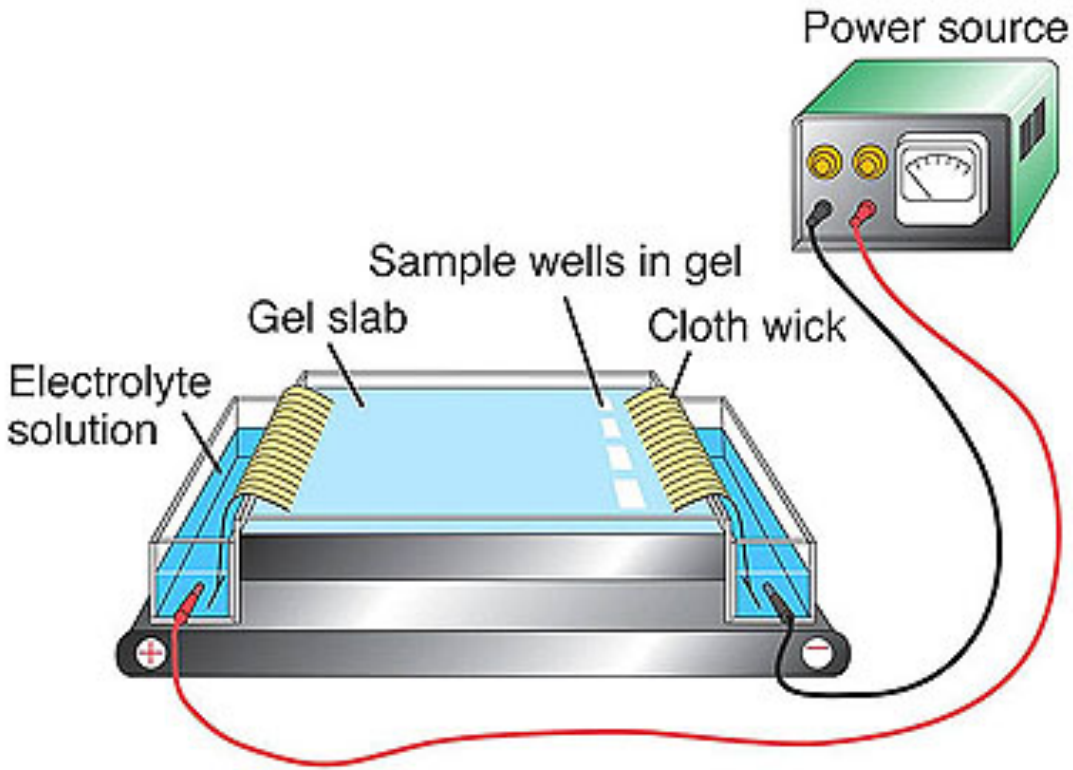
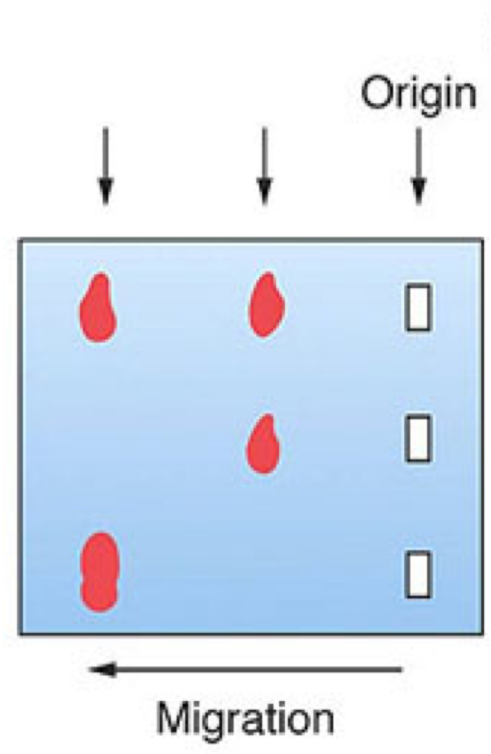


Figure 4: Schematic of devices used in protein electrophoresis

Protein variation

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



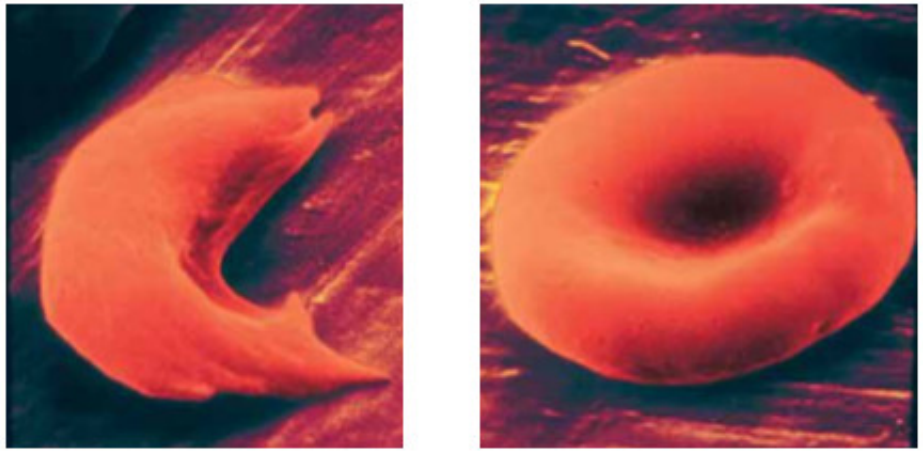
How does it work?

- 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

- 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

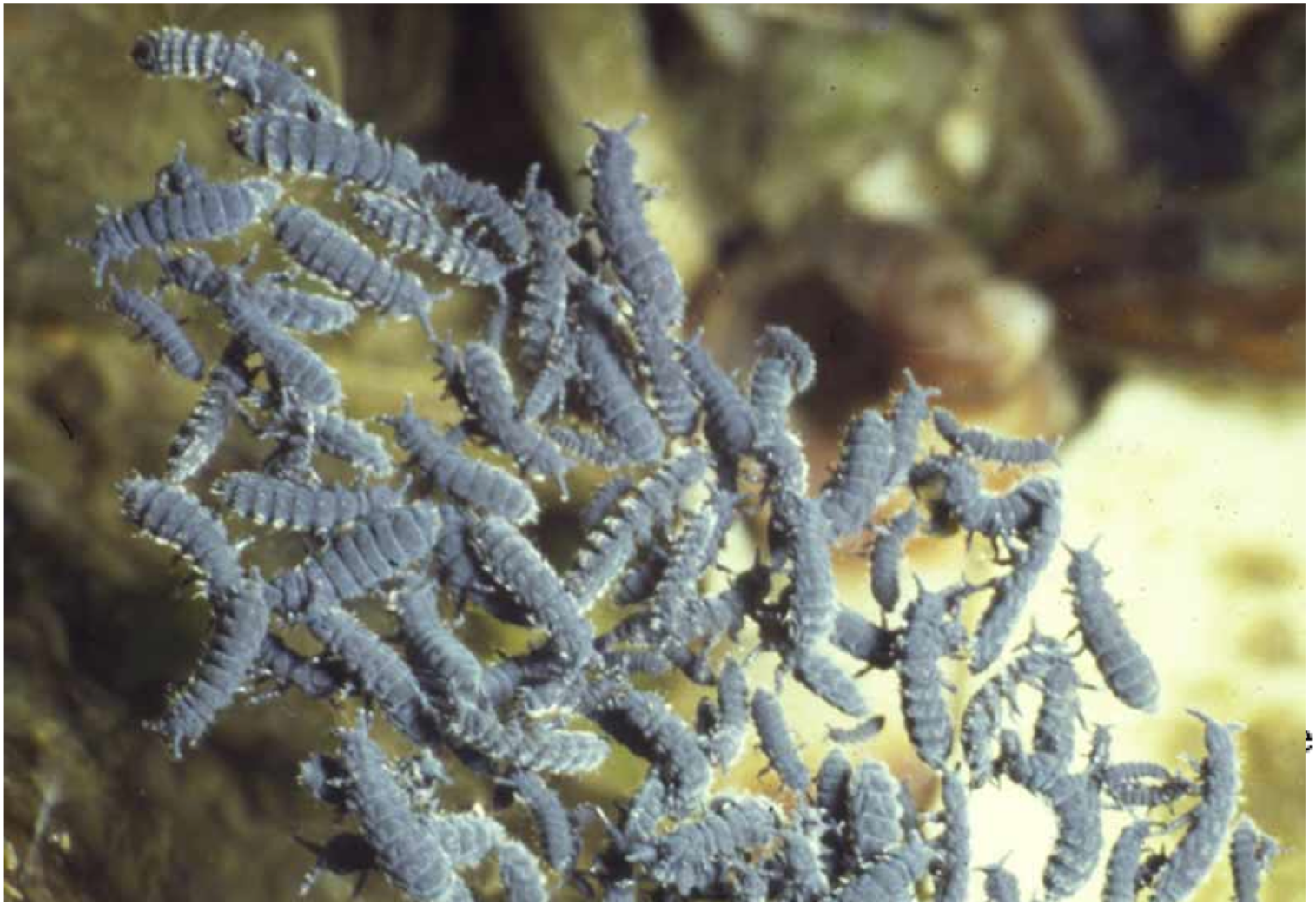
- RFLP: Restriction fragment length polymorphism
  - Created by mutation that changes a restriction site

```
GCCGCATTCTA
CGGCGTAAGAT
```

```
GCCGAATTCTA
CGGCTTAAGAT
```

DNA variation

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



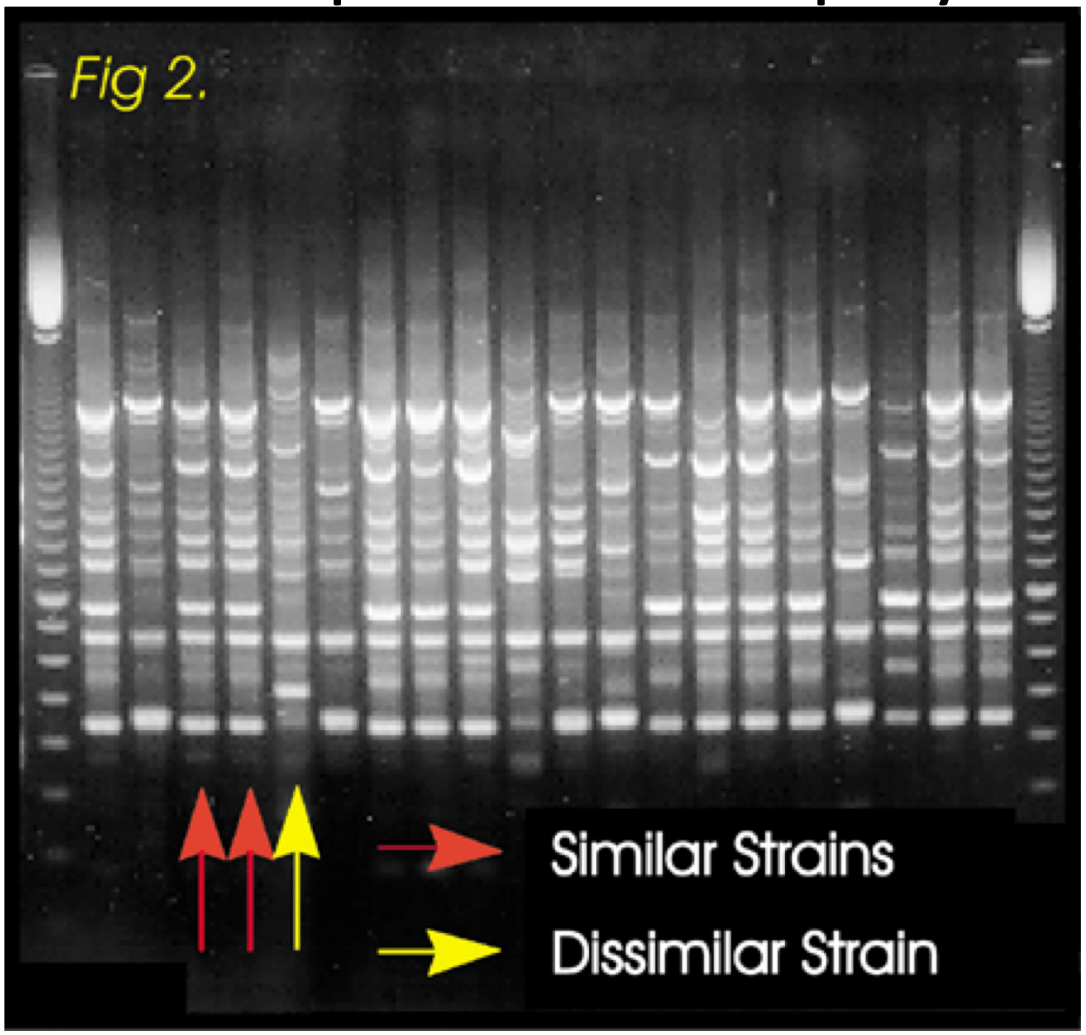
DNA variation

- 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

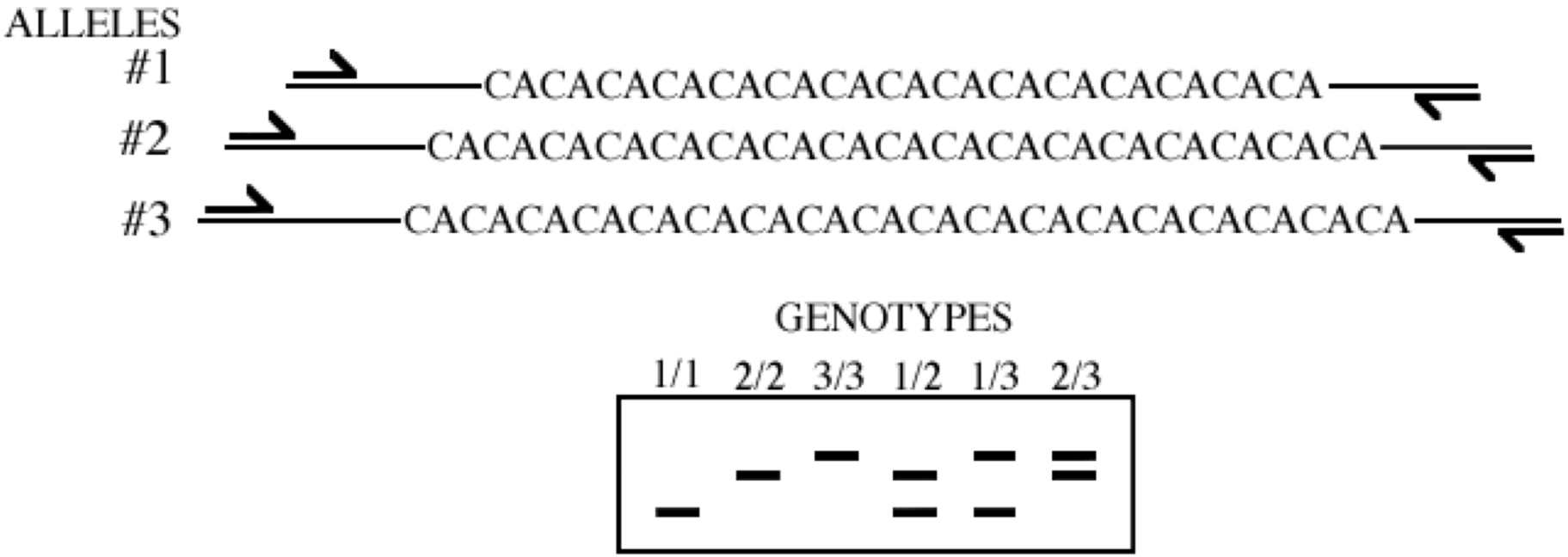
Strains of Lactobacillus from 18 types of Cheddar Cheese



Teagasc 1998

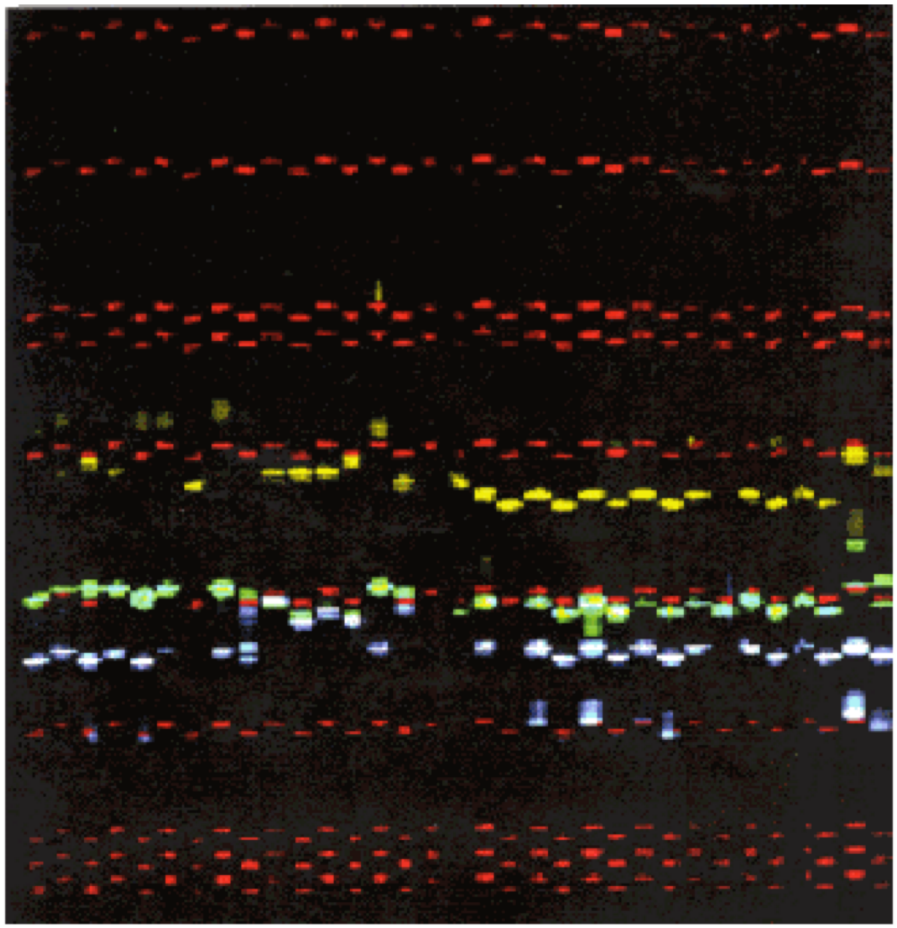
DNA variation

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



DNA variation

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

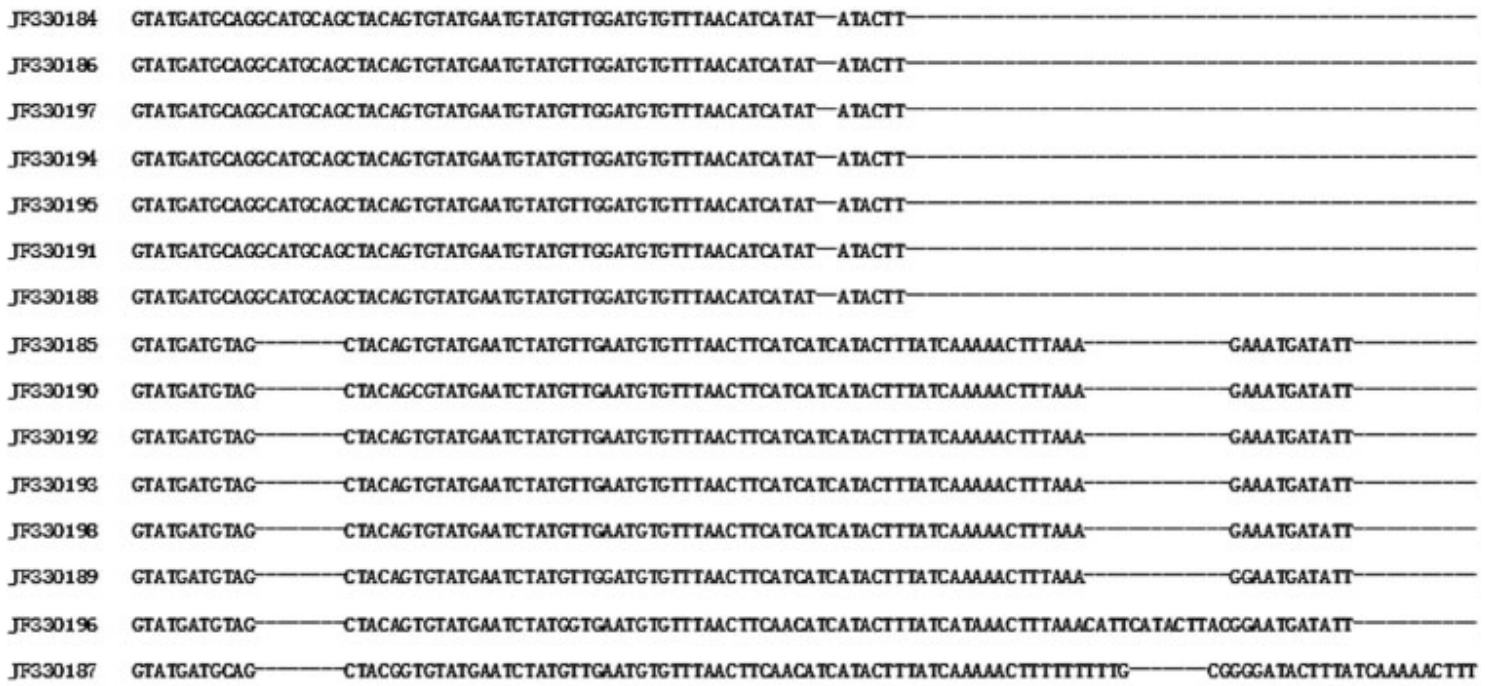


Leishmania (Viannia) isolates

From London School of Hygiene and Tropical Medicine

DNA variation

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

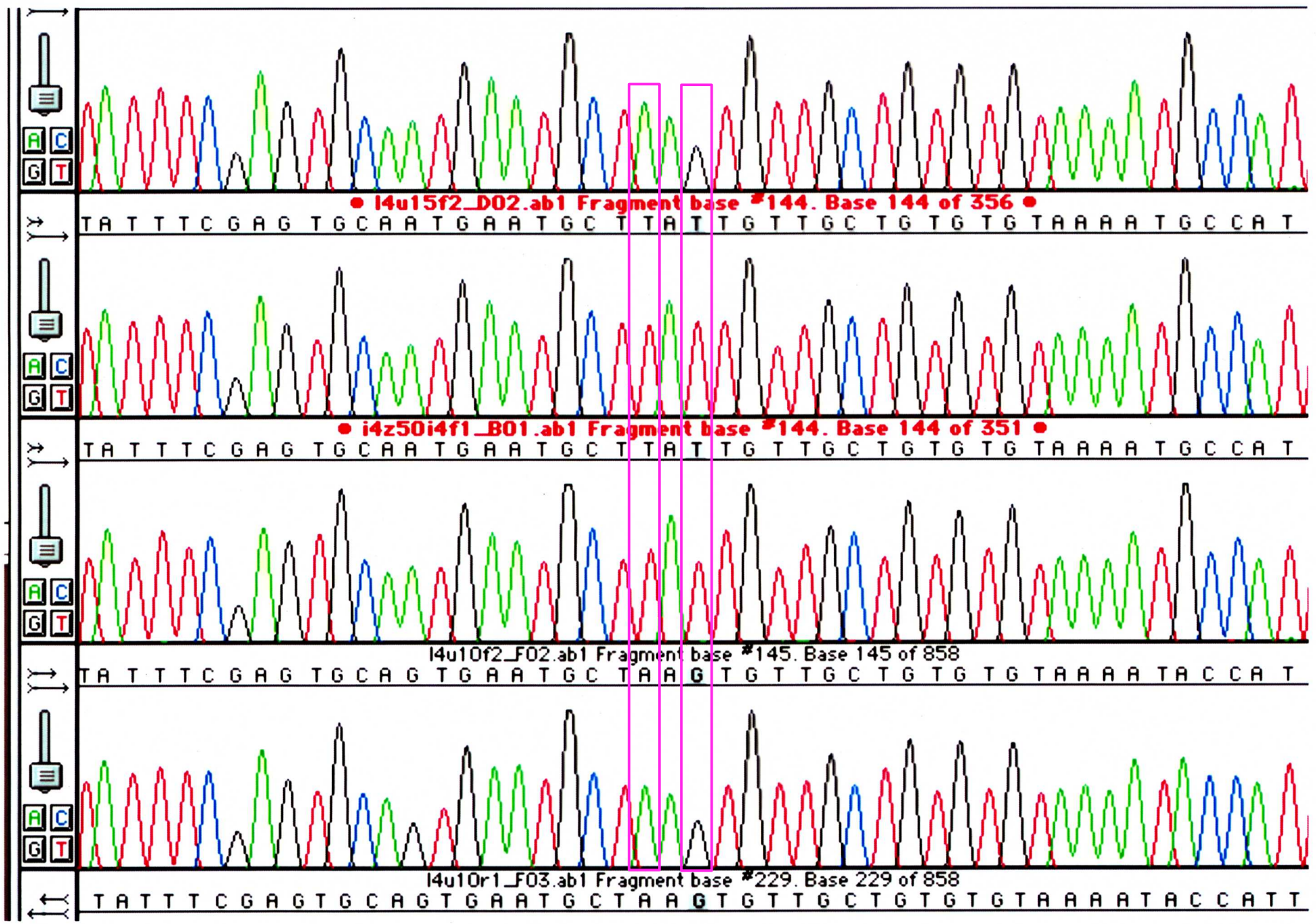


From Wei *et al*. 2011

DNA variation

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

Sequence reads from 4 individuals



Discrete Genetic Variation

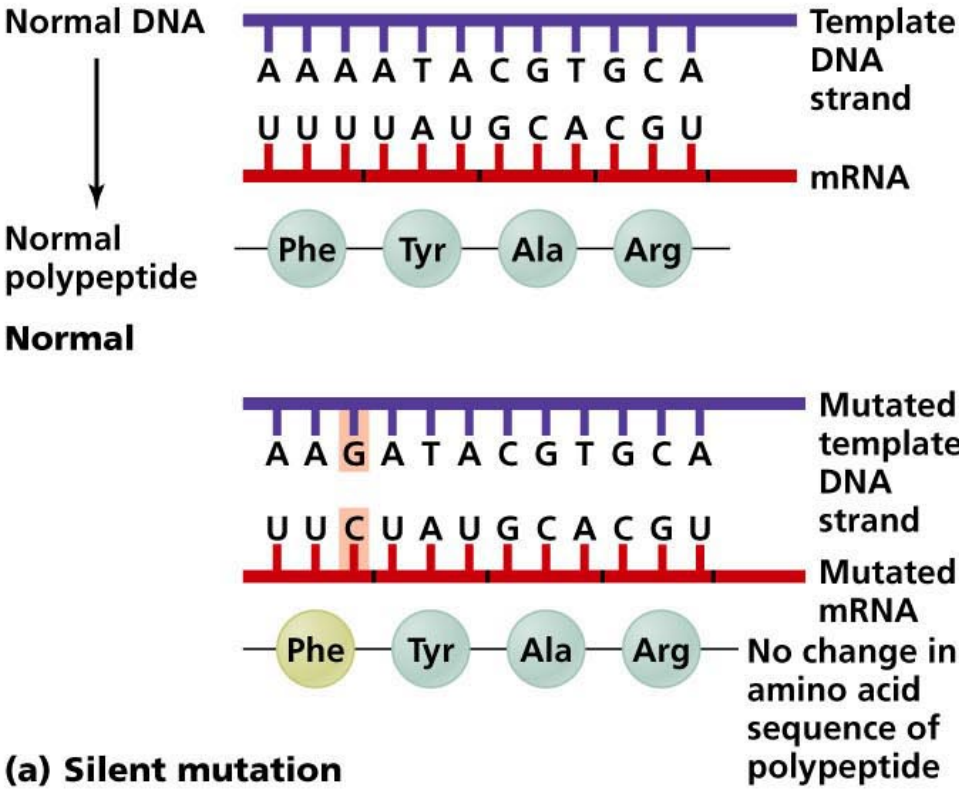
- Chromosomal variation
  - Inversions, chromosomes fusions/fissions
- Protein variation
  - Immunological, allozymes, amino acid variation
- DNA variation
  - RFLP, RAPD, VNTR, Indel, SNP

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



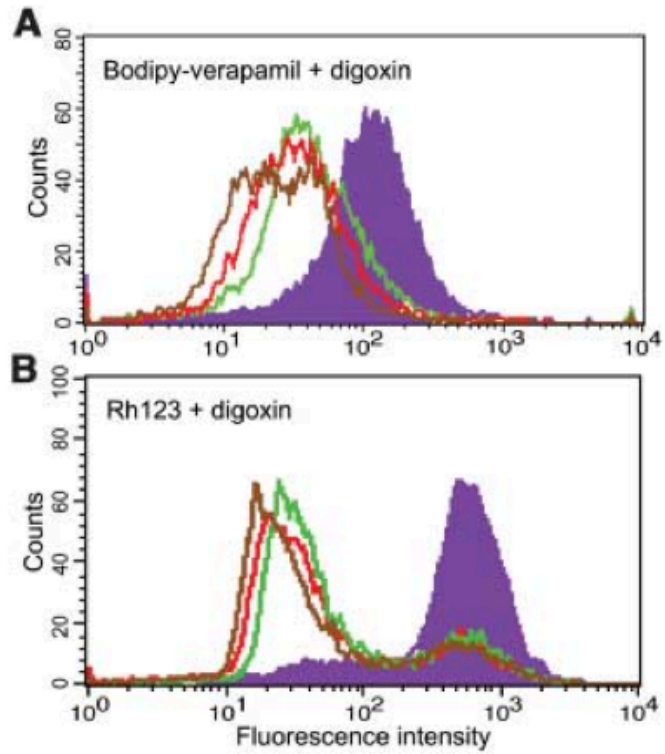
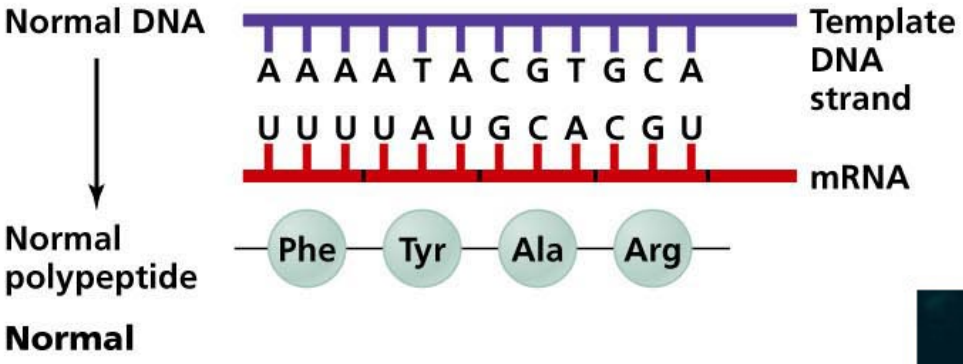
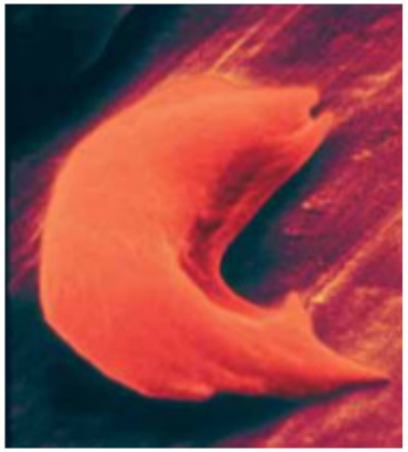


Fig. 4. Drug transport function of wild-type and two MDR1 haplotypes. The drug efflux of vaccinia infected/transfected HeLa cells was determined by FACS analysis (*14*). Cells were transfected with pTM1 (control; purple), MDR1, (wild-type P-gp; green), C1236T-G2677T-C3435T (red), and C1236T-G2677T-C3435A (brown). (**A**) 0.5 μM bodipy-FL-verapamil in the presence of 500 μM digoxin; (**B**) 0.5 μM Rh123 in the presence of 150 μM digoxin.

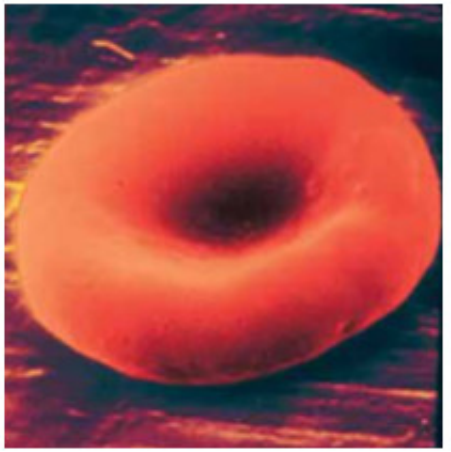
Kimchi-Sarfaty et al. 2007

Mutation is the substrate of evolution

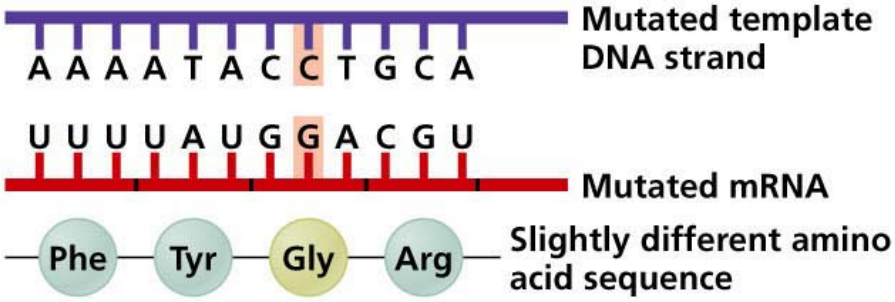




Sickle-cell phenotype

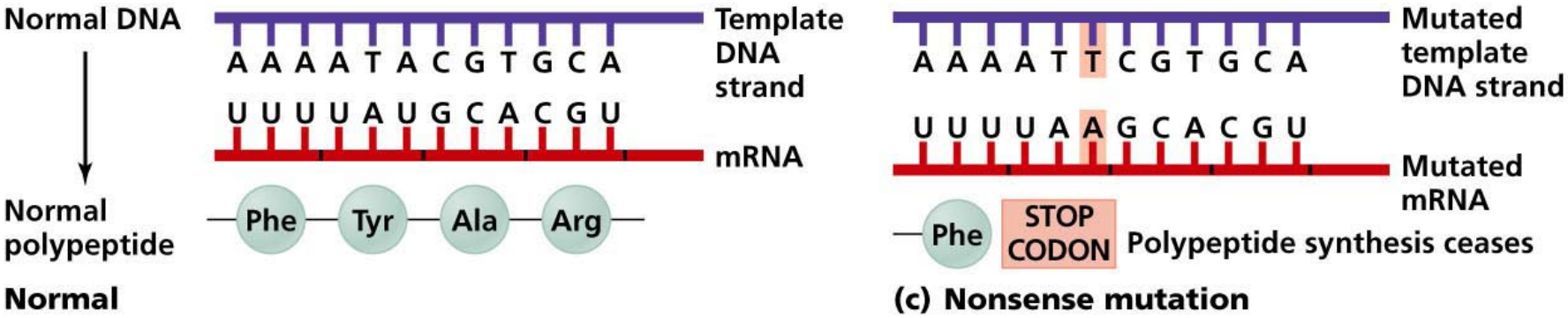


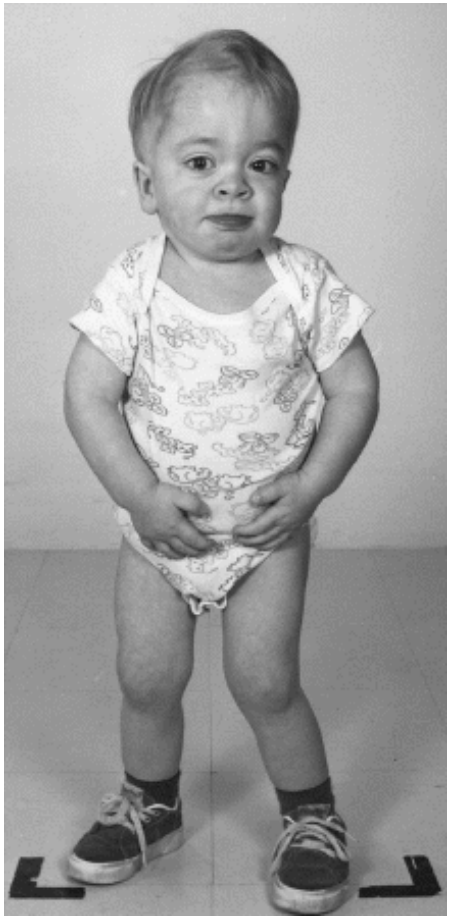
Normal phenotype



(b) Missense mutation

Mutation is the substrate of evolution

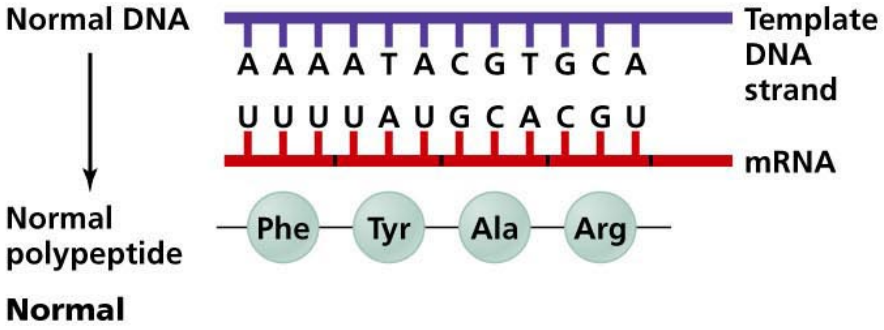


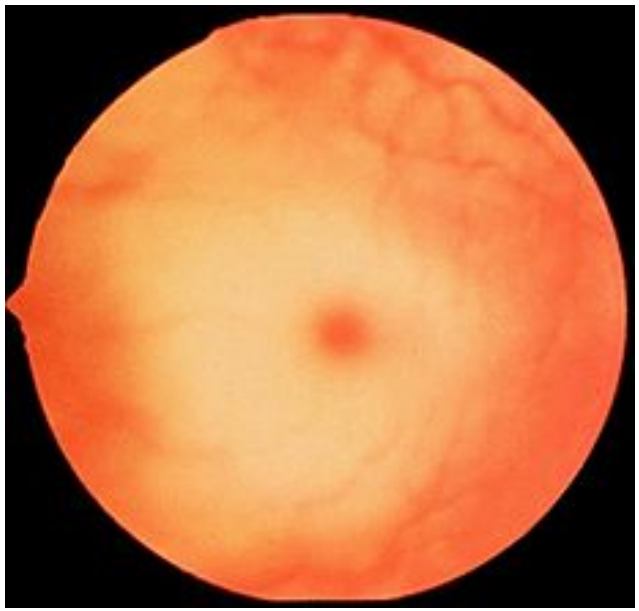


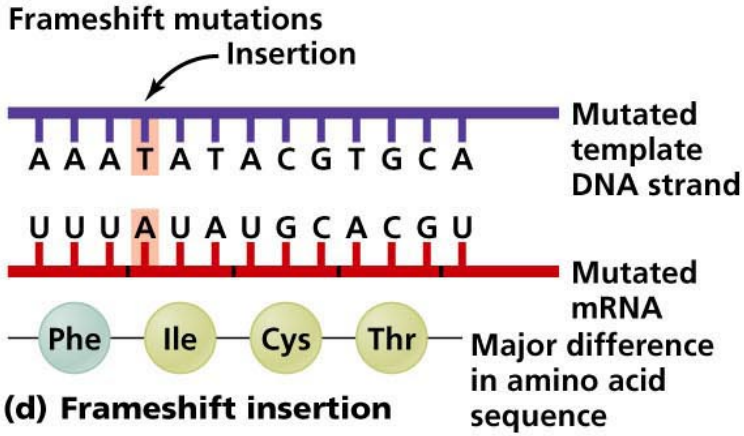
Mutation is the substrate of evolution

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

Mutation is the substrate of evolution



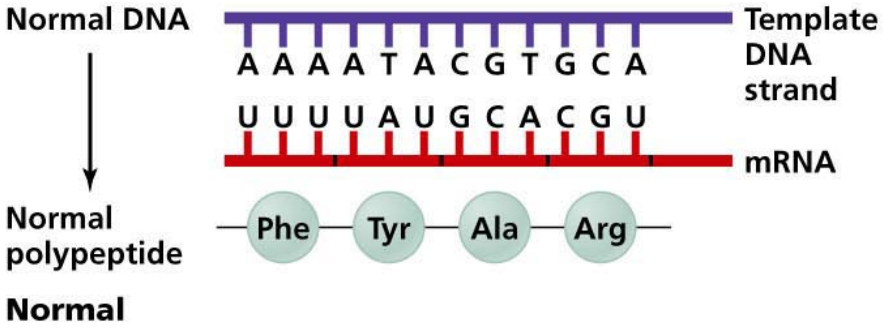


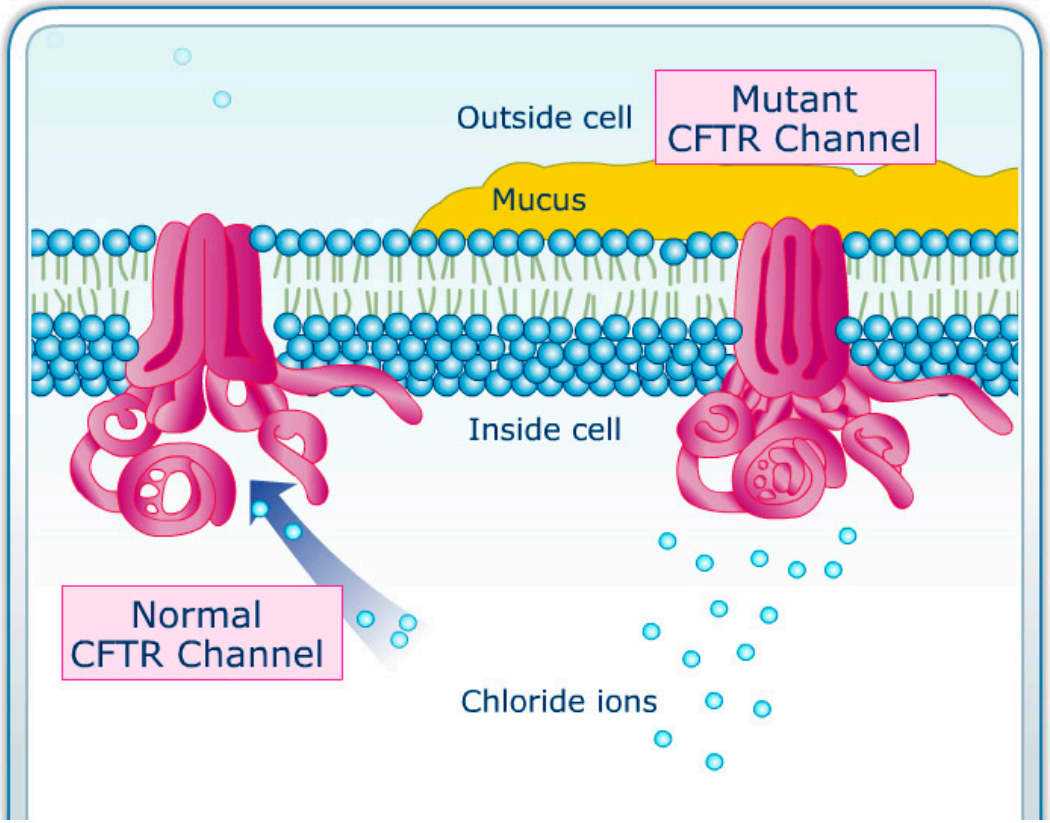


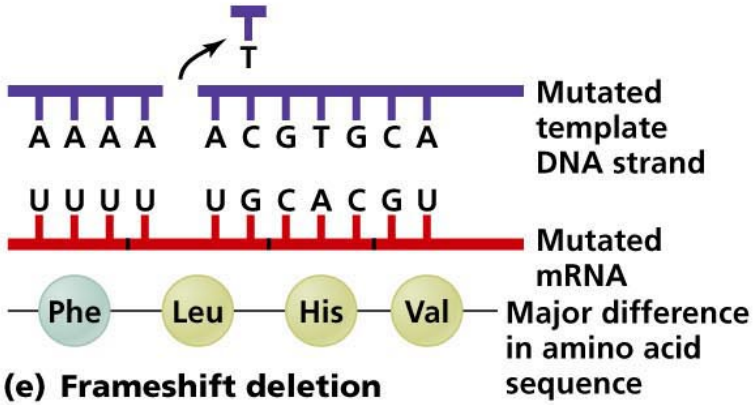
Mutation is the substrate of evolution

- 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



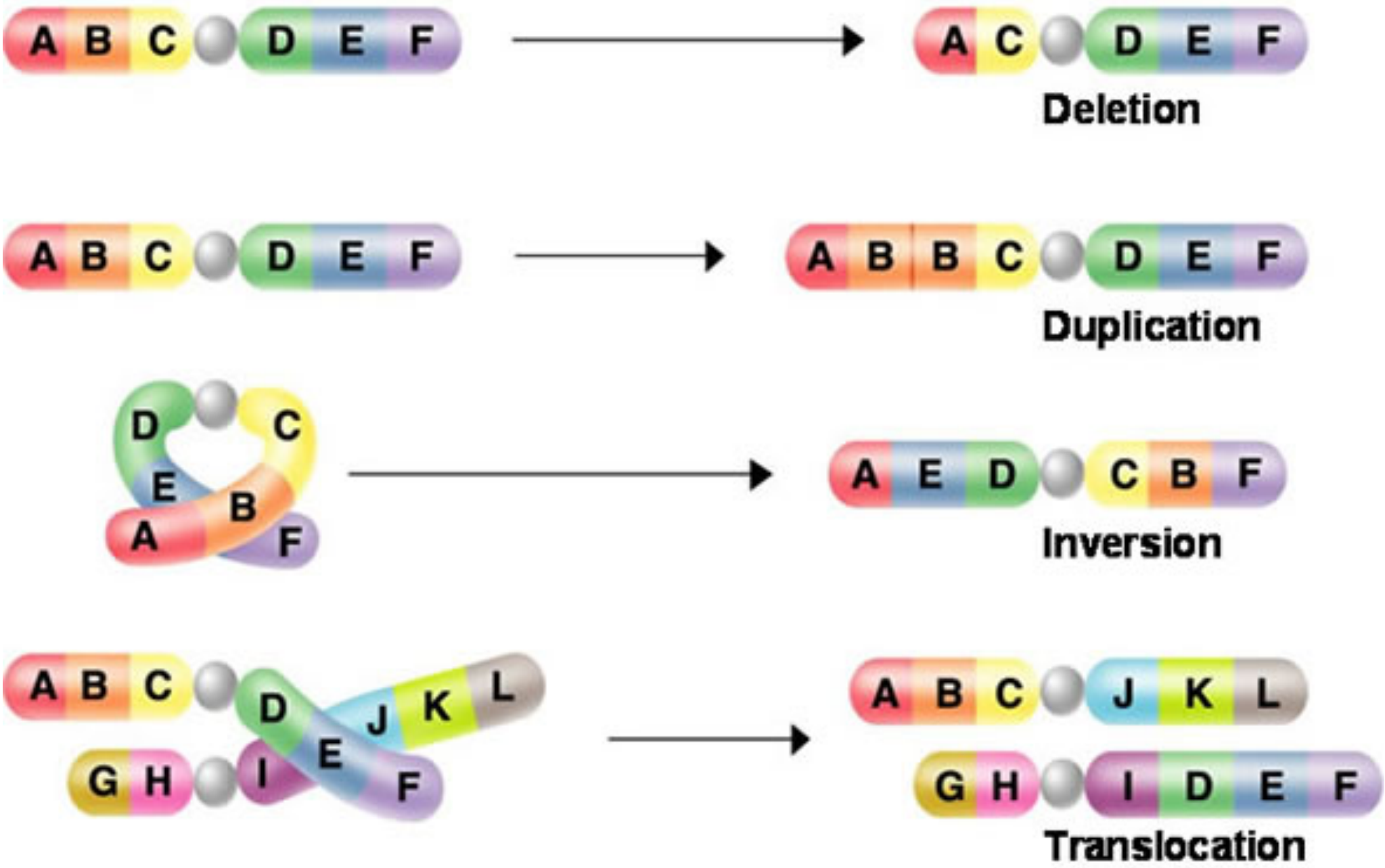




Mutation is the substrate of evolution

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

Mutation is the substrate of evolution



Genetic Variation

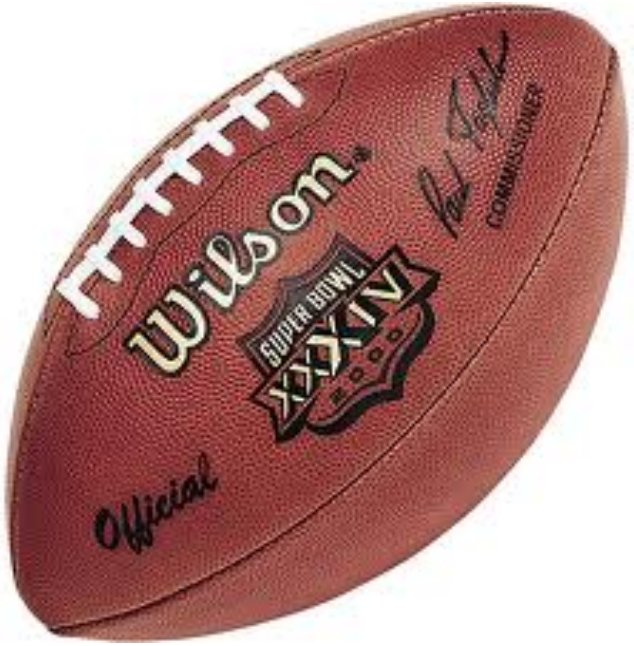
- All genetic variation originates with mutation
  - Mutation is the substrate of evolution
- All levels of organization from single base pairs to entire genomes
- Understanding genetic variation has deep implications
- Population genetics aimed at understanding genetic variation within populations

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

Key Terms

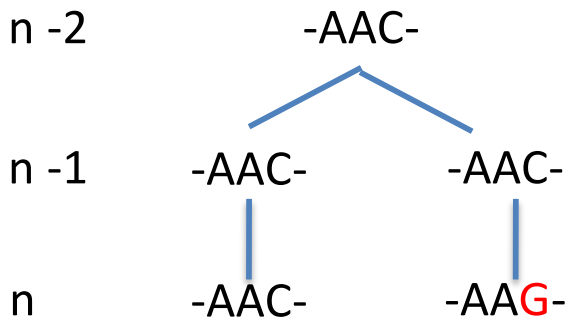
- Population: A group of individuals of a single species that, if sexual, is capable of exchanging genes
- Sample: A finite number of individuals collected from a population
- Polymorphism: Character that is variable within a species
- Allele: Variant at a specific locus in the genome
- Genotype: The allelic makeup of an individual
  - Homozygote (identical alleles at a locus in an individual)
  - Heterozygote (different alleles at a locus in an individual)

Key Terms

- Identical by state
  - Nucleotide level: Alleles have identical nucleotide sequences
  - Amino acid level: Alleles have identical amino acid sequences

Key Terms

- Identical by state
  - Nucleotide level: Alleles have identical nucleotide sequences
  - Amino acid level: Alleles have identical amino acid sequences
- Identical by descent: Alleles share a common ancestor allele (in the short term)
  - Need not be identical by state



Allele Frequency

Individual	Allele																			
1	a	C	A	T	A	G	A	A	C	C	T	G	G	G	C	A	C	T	T	C	A
2	b	.	.	.	A	.	T	.	.	.	.	G	.	.	.	C	.	.	.	T	.
3	c	.	.	.	G	.	A	.	.	.	.	G	.	.	.	A	.	.	.	C	.
4	d	.	.	.	G	.	A	.	.	.	.	T	.	.	.	A	.	.	.	C	.
5	b	.	.	.	A	.	T	.	.	.	.	G	.	.	.	C	.	.	.	T	.

- Allele frequency: number of instances of allele/total number of alleles sampled
  - $p_a = 1/5 = 0.2$
  - $p_b = 2/5 = 0.4$
  - $p_c = 1/5 = 0.2$
  - $p_d = 1/5 = 0.2$

Allele Frequency

Individual	Allele																			
1	a	C	A	T	A	G	A	A	C	C	T	G	G	G	C	A	C	T	T	C	A
2	b	.	.	.	A	.	T	.	.	.	.	G	.	.	.	C	.	.	.	T	.
3	c	.	.	.	G	.	A	.	.	.	.	G	.	.	.	A	.	.	.	C	.
4	d	.	.	.	G	.	A	.	.	.	.	T	.	.	.	A	.	.	.	C	.
5	b	.	.	.	A	.	T	.	.	.	.	G	.	.	.	C	.	.	.	T	.

- Allele frequency: number of instances of allele/total number of alleles sampled

Allele Frequency

Individual	Allele																			
1	a	C	A	T	A	G	A	A	C	C	T	G	G	G	C	A	C	T	T	C	A
2	b	.	.	.	A	.	T	.	.	.	.	G	.	.	.	C	.	.	.	T	.
3	c	.	.	.	G	.	A	.	.	.	.	G	.	.	.	A	.	.	.	C	.
4	d	.	.	.	G	.	A	.	.	.	.	T	.	.	.	A	.	.	.	C	.
5	b	.	.	.	A	.	T	.	.	.	.	G	.	.	.	C	.	.	.	T	.

- Allele frequency: number of instances of allele/total number of alleles sampled
  - $p_A = 3/5 = 0.6$
  - $p_G = 2/5 = 0.4$

Allele Frequency

Individual	Allele																			
1	a	C	A	T	A	G	A	A	C	C	T	G	G	G	C	A	C	T	T	C	A
2	b	.	.	.	A	.	T	.	.	.	.	G	.	.	.	C	.	.	.	T	.
3	c	.	.	.	G	.	A	.	.	.	.	G	.	.	.	A	.	.	.	C	.
4	d	.	.	.	G	.	A	.	.	.	.	T	.	.	.	A	.	.	.	C	.
5	b	.	.	.	A	.	T	.	.	.	.	G	.	.	.	C	.	.	.	T	.

- Allele frequency: number of instances of allele/total number of alleles
  - Whether estimated per locus or per site, allele frequencies must sum to 1

Genotype Frequency

Genotype	AA	Aa	aa

Genotype Frequency

Genotype Frequency

Genotype	AA	Aa	aa
Counts	17	15	14
Frequency	0.37	0.33	0.30

$p_A = (2*17 + 15)/(2*46) = 0.53$

$p_a = (2*14 + 15)/(2*46) = 0.47$