

# Introduction to Whole Genome Sequencing from DNA to Phenotypes



Rodrigo Pelicioni Savegnago  
Department of Animal Science



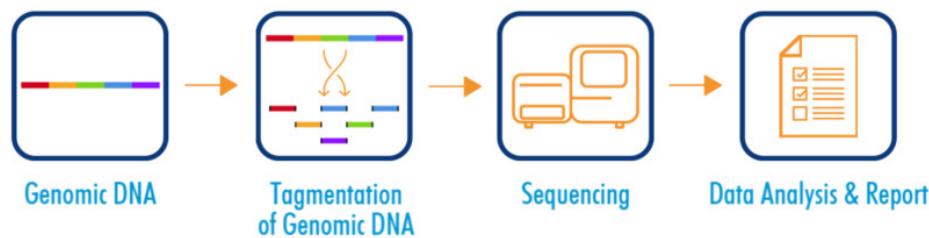
- Overview of genetic variation



- File formats

TCTTAGCCCAAGACTTCCC GTGT CCTTT  
TGATGCTGTGGTCTTCATCTGCAGGT  
CAGGGTGCAAGCTGAGCACTGGAGTG

- WGS workflow



- IVDP – Integrated Variant Discovery Pipeline



# Variation: from DNA to phenotypes

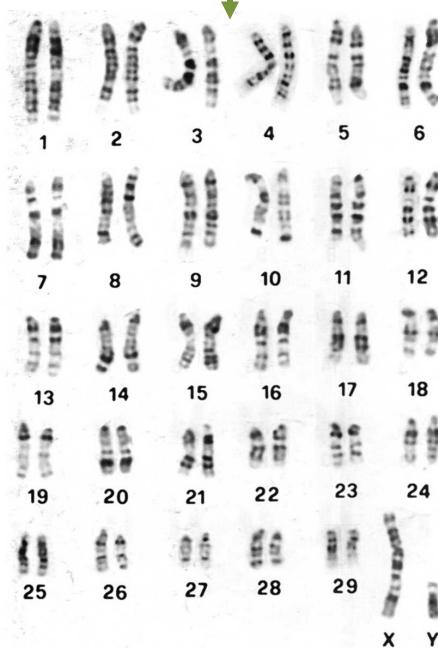
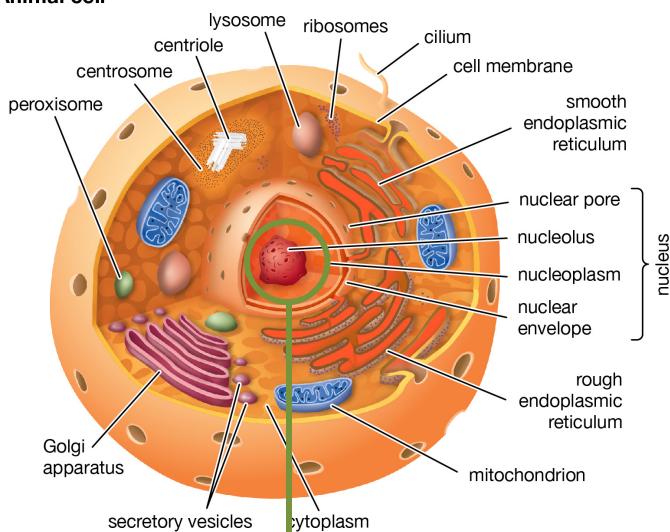


# Variation: from DNA to phenotypes

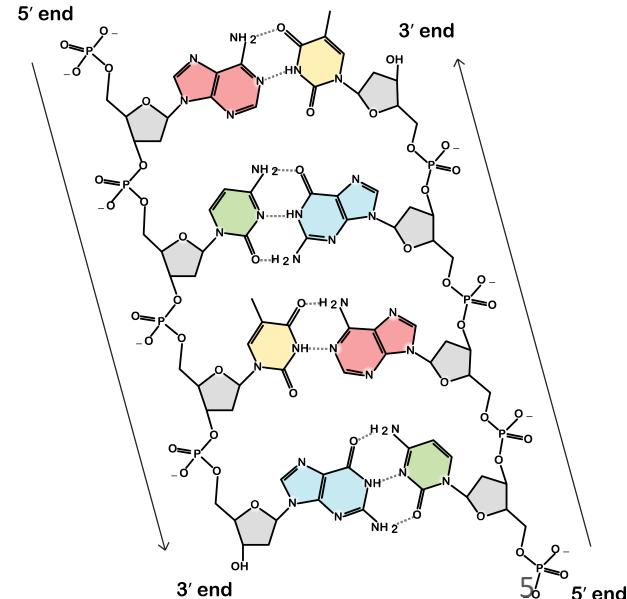
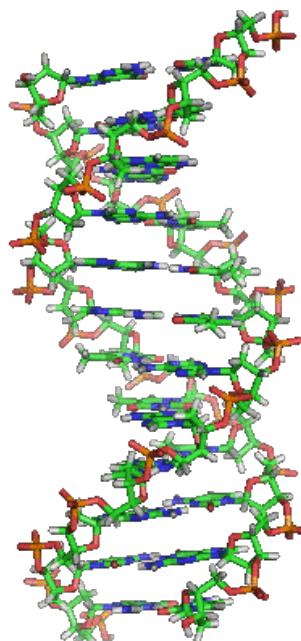


# Variation: from DNA to phenotypes

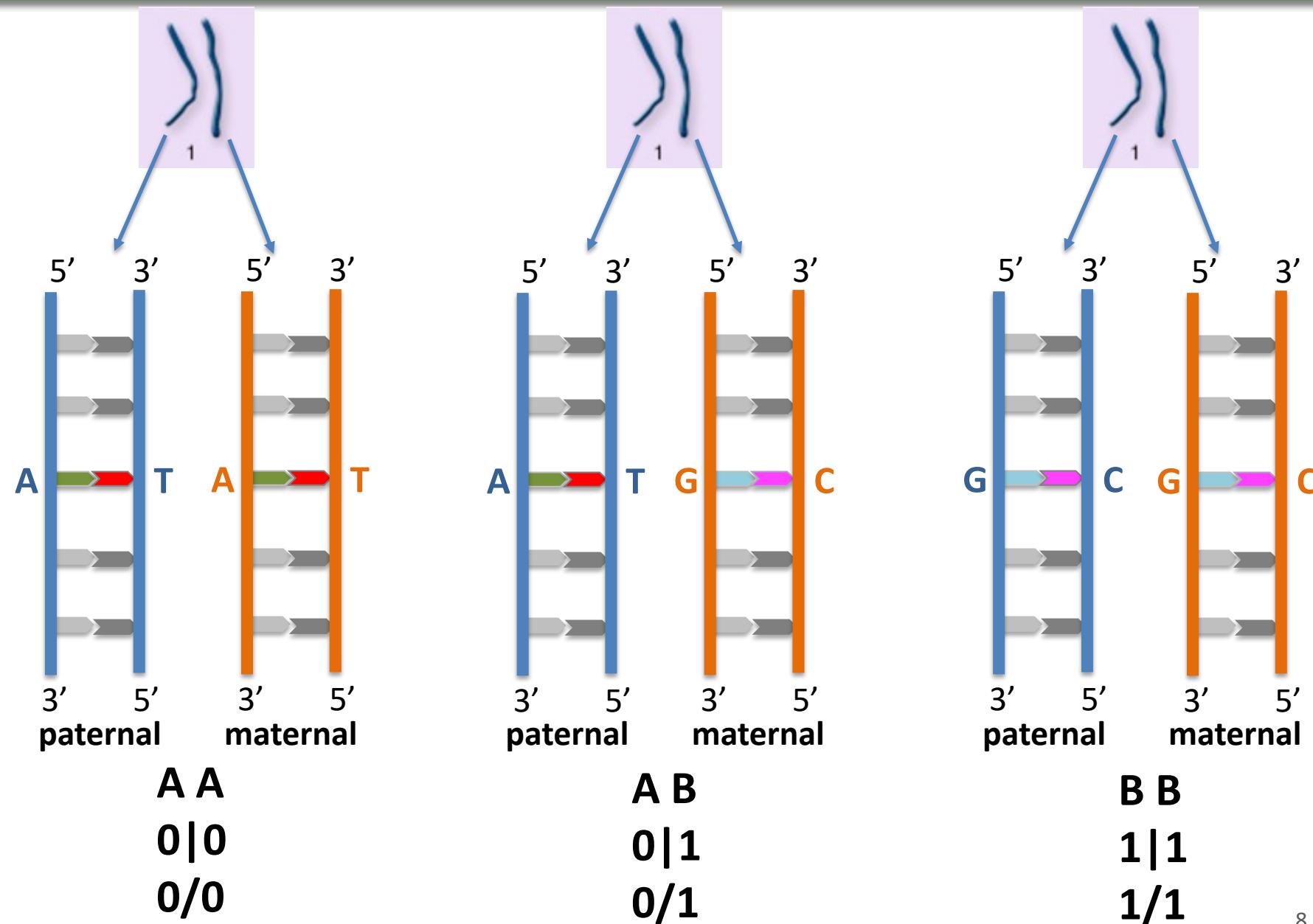
Animal cell



© Encyclopædia Britannica, Inc.

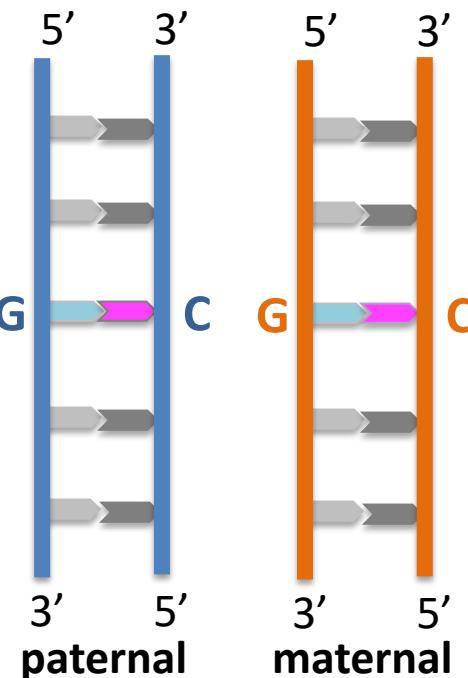
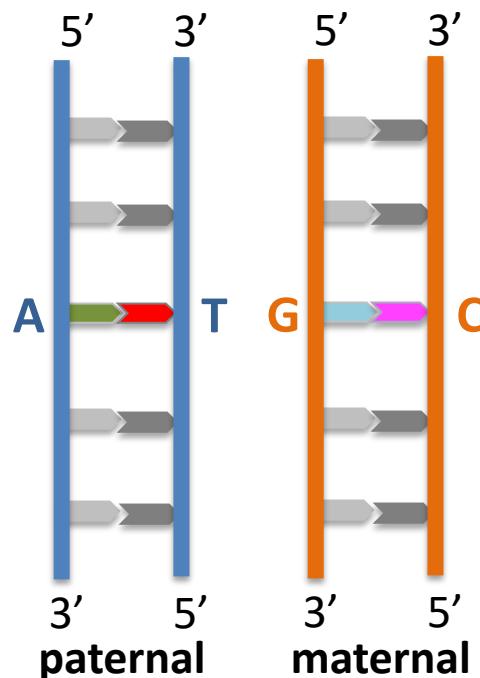
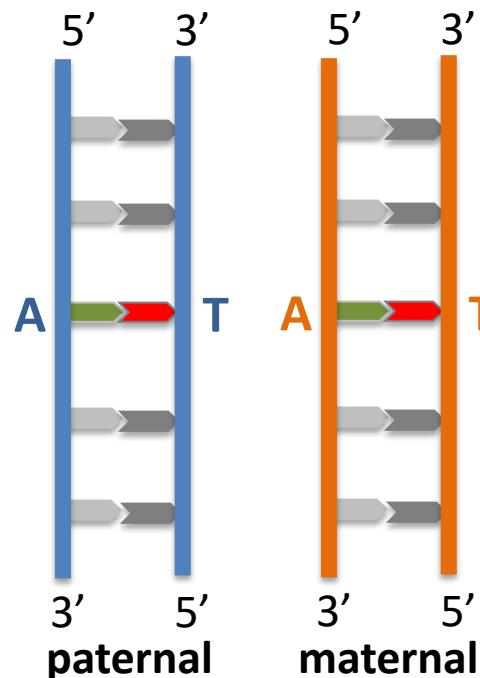
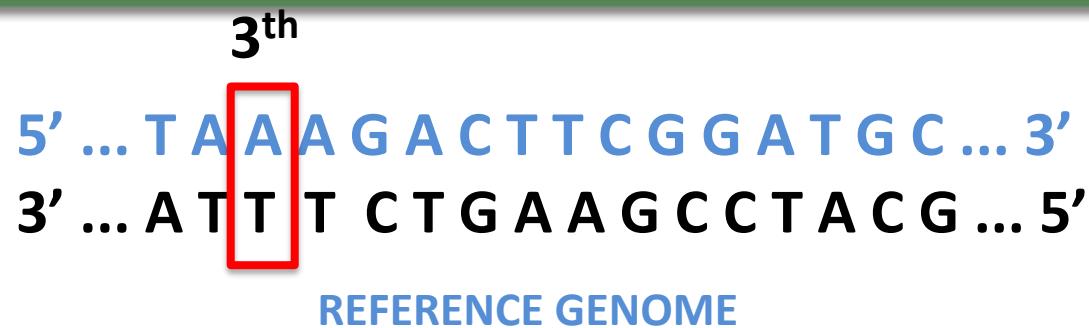


# Variation: from DNA to phenotypes





# Reference Genome



AA  
0|0  
0/0

AB  
0|1  
0/1

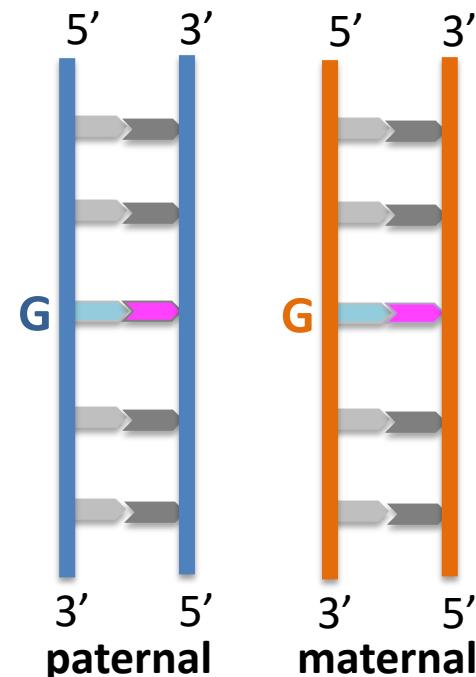
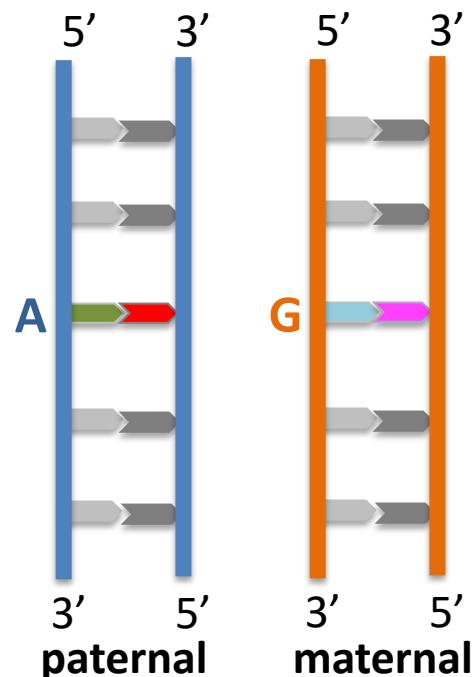
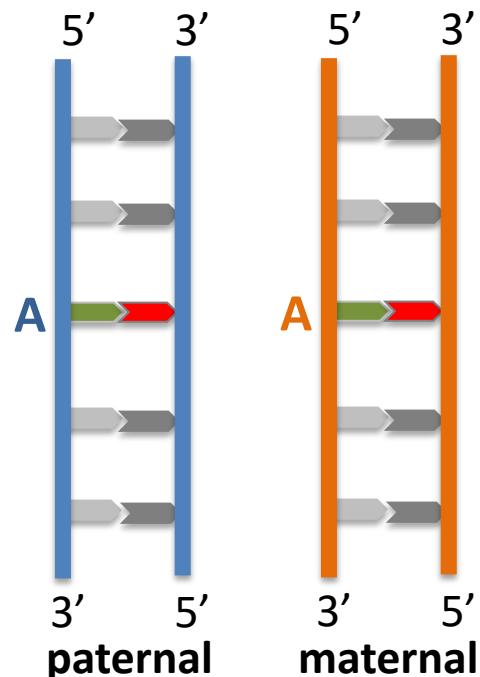
BB  
1|1  
1/1

# Reference Genome

3<sup>th</sup>

5' ... TAAAGACTTCGGATGC ... 3'

REFERENCE GENOME



A A

0|0

0/0

A B

0|1

0/1

B B

1|1

1/1

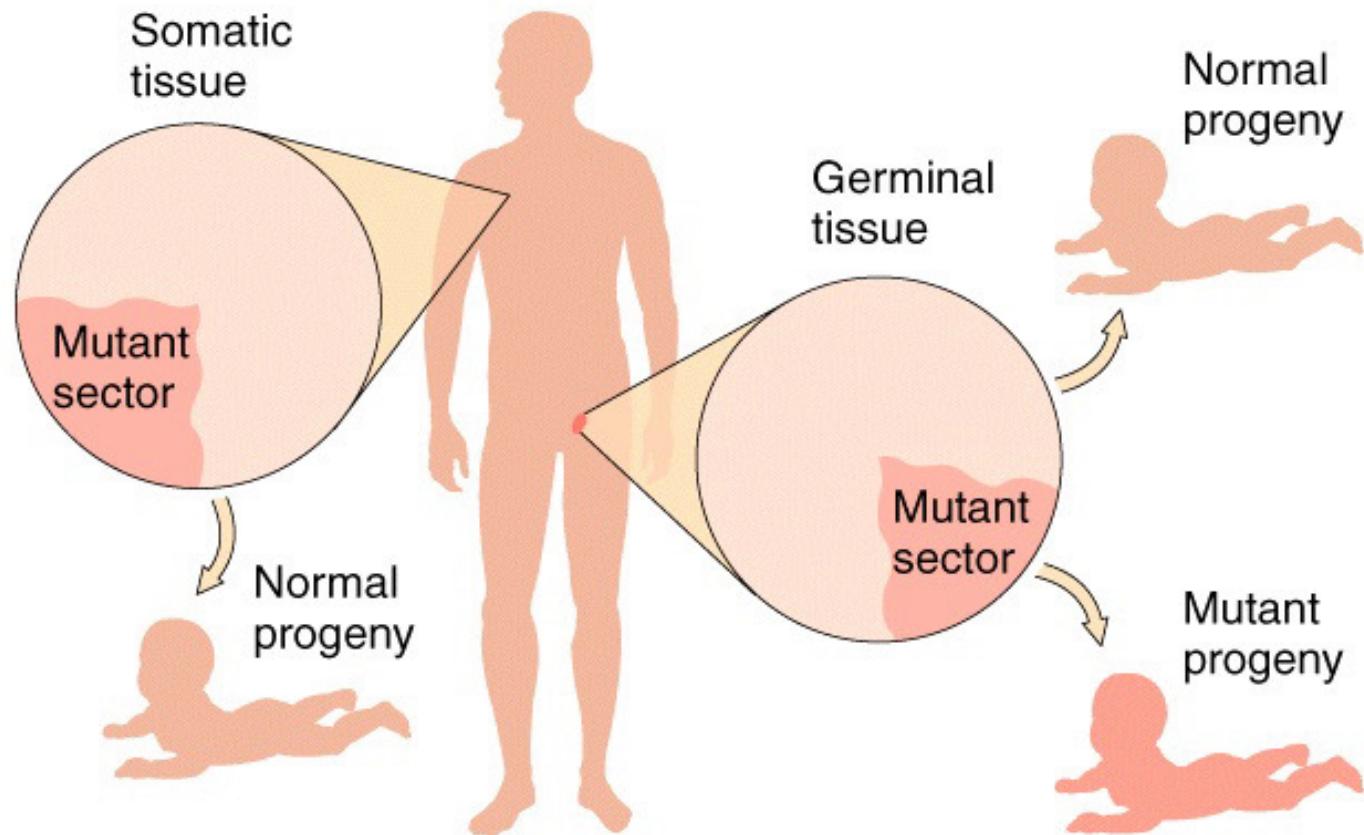
- **Endogenous factors**

Errors in cellular replication and oxidative damage.

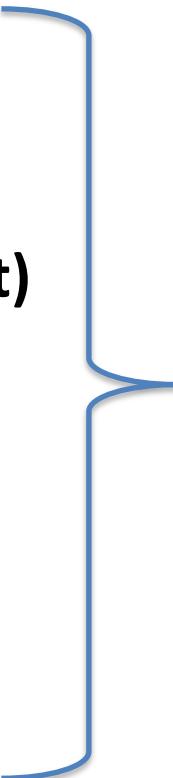
- **Exogenous factors**

Exposure to harmful substances, which damage the DNA of somatic and/or germ cells. This damage can then either be repaired perfectly, and no mutations will be present, or repaired imperfectly, resulting in a variety of mutations

# Mutation: location



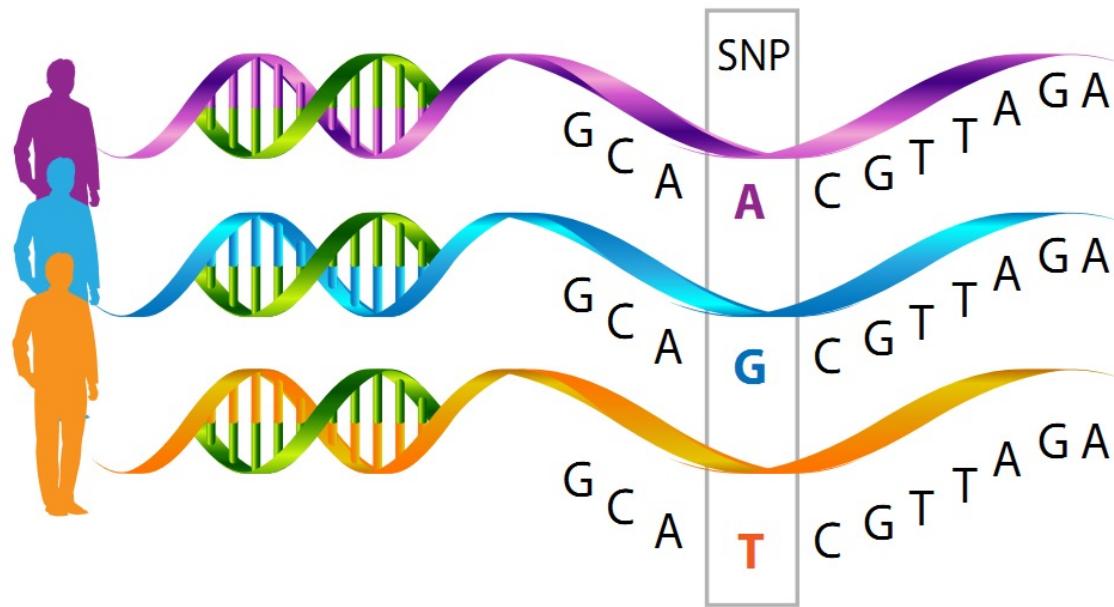
The unique source to **create** variation is mutation!

- Migration
  - Genetic drift (bottleneck effect)
  - Non-random mating
  - Selection
- 
- They do not create variation!
  - They just shuffle and change the allele frequencies.
  - They play with variation that already exist.

# Single Nucleotide Polymorphism

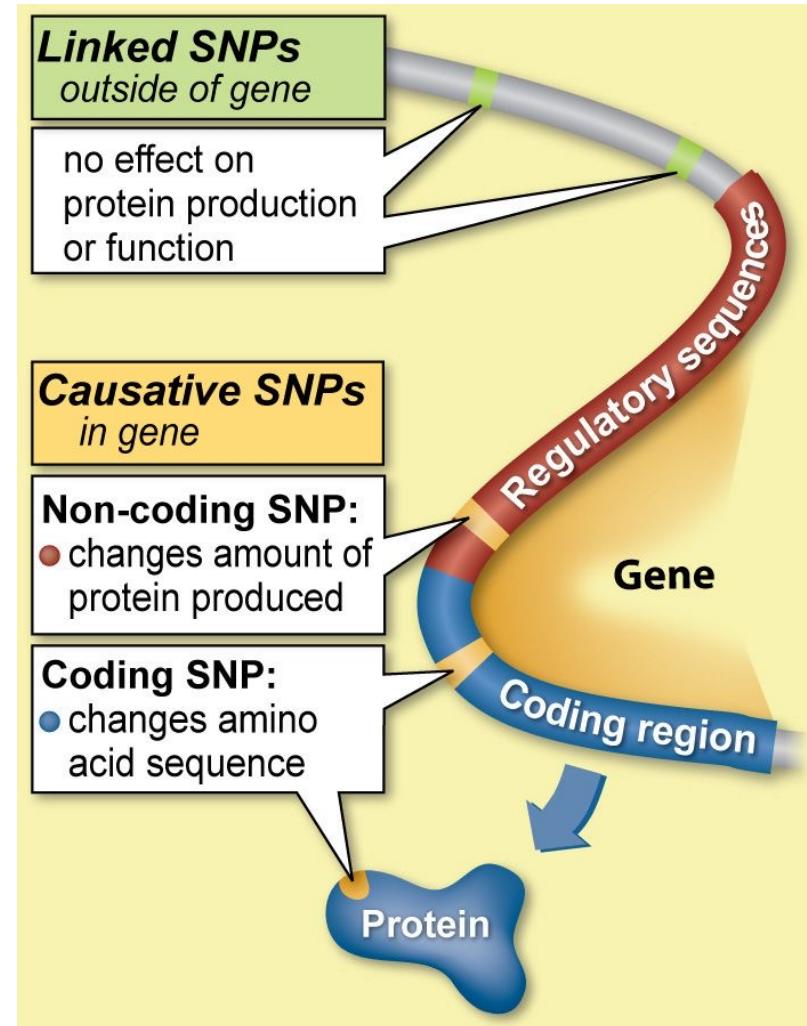
## Definition

- A single-nucleotide polymorphism (SNP) is a substitution of a single nucleotide at a specific position in the genome, that is present in a sufficiently large fraction of the population (e.g. 1% or more).

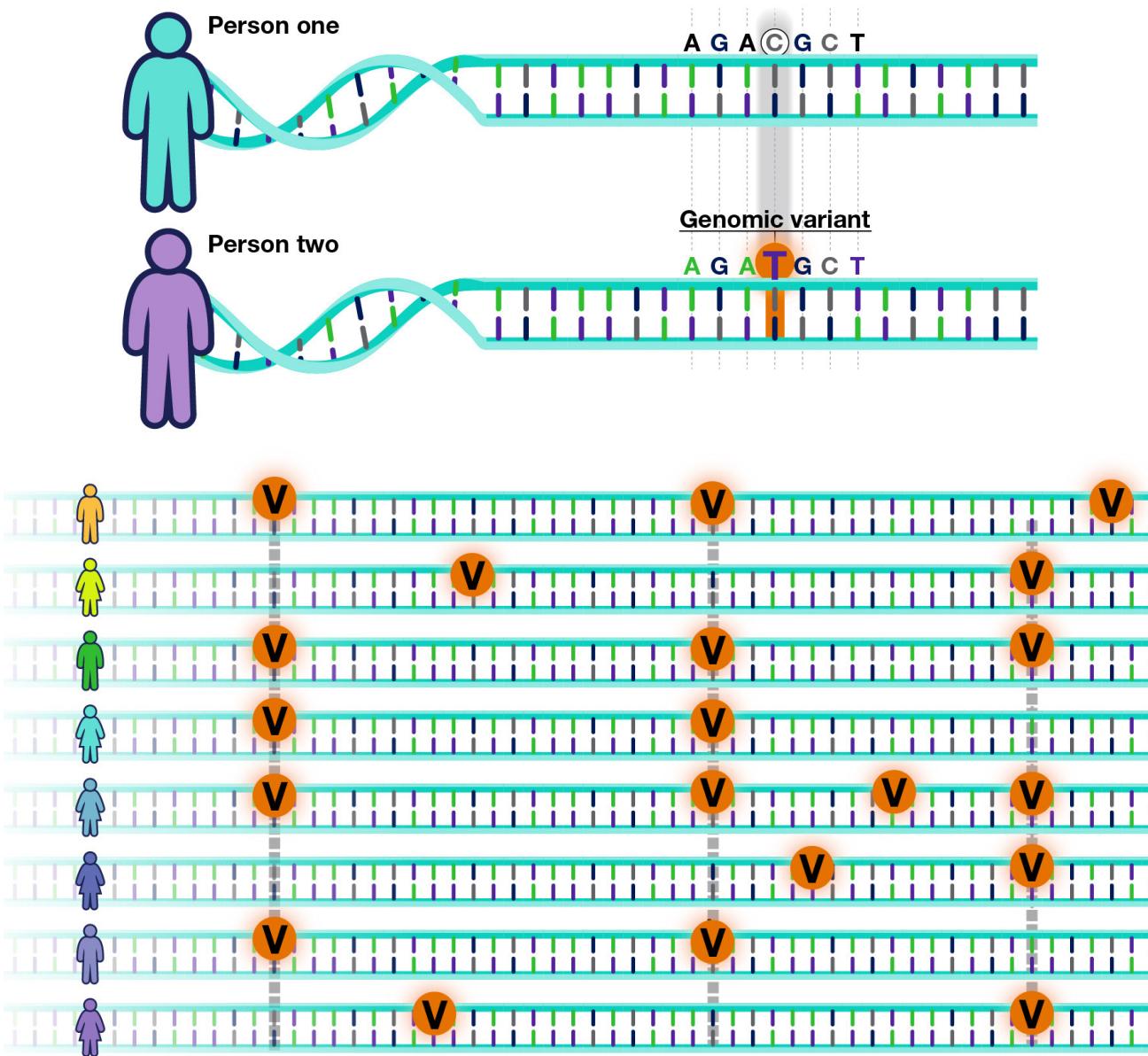


# Single Nucleotide Polymorphism

**SNPs that are not in protein-coding regions** may still affect gene splicing, transcription factor binding, messenger RNA degradation, or the sequence of noncoding RNA. Gene expression affected by this type of SNP is referred to as an **eSNP (expression SNP)** and may be upstream or downstream from the gene.



# SNPs and quantitative traits

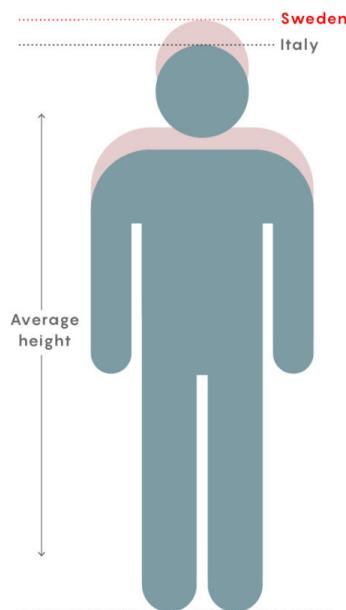


# SNPs and quantitative traits

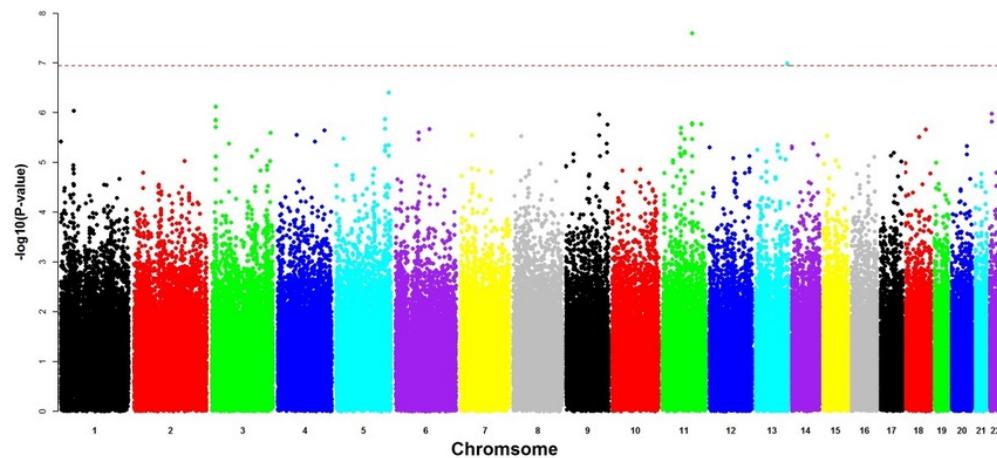
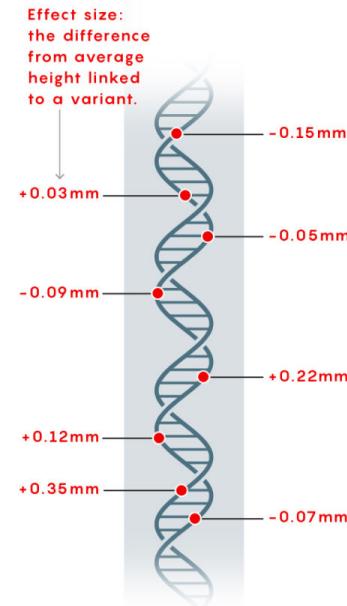
## The Problem With Polygenic Scores

Genome-wide association studies (GWAS) have helped identify thousands of genetic markers that are statistically linked to certain traits. More recently, some researchers have tried using those results to make quantitative predictions about individuals and the evolutionary histories of populations — and problems have emerged.

For example, researchers have wanted to know whether the summed effects of many genes could explain differences in height between European populations.



GWAS analyses identified thousands of relevant genetic variants. Polygenic scores represent a way of adding up all the statistical effects of the variants.



But if the estimated effect sizes are slightly biased, as recent studies suggest is often the case, then predictions of individuals' heights from their polygenic scores will be flawed.

# SNPs and quantitative traits: SNP Matrix

	Homologous chromosomes		Homologous chromosomes		Homologous chromosomes		Homologous chromosomes		Homologous chromosomes	
	ID 1		ID 2		ID 3		ID 4		ID 5	
Pos1	A	A	A	A	A	A	A	A	A	A
Pos2	G	G	G	G	G	G	G	G	G	G
Pos3	A	A	A	A	A	A	A	A	A	A
Pos4	T	T	T	T	T	T	T	T	T	T
Pos5	A	C	A	A	A	C	A	A	C	C
Pos6	G	G	G	G	G	G	G	G	G	G
Pos7	T	T	T	T	T	T	T	T	T	T
Pos8	A	A	A	A	A	A	A	A	A	A
Pos9	C	A	A	A	C	C	C	C	C	C
Pos10	T	T	T	T	T	T	T	T	T	T

Position	Major Allele Frequency	Minor Allele Frequency (MAF)
5	0.6 ( A )	0.4 ( C )
9	0.7 ( C )	0.3 ( A )

	Homologous chromosomes		Homologous chromosomes		Homologous chromosomes		Homologous chromosomes		Homologous chromosomes	
	ID 1		ID 2		ID 3		ID 4		ID 5	
Pos5	0	1	0	0	0	1	0	0	1	1
Pos9	0	1	1	1	0	0	0	0	0	0

# SNPs and quantitative traits: SNP Matrix

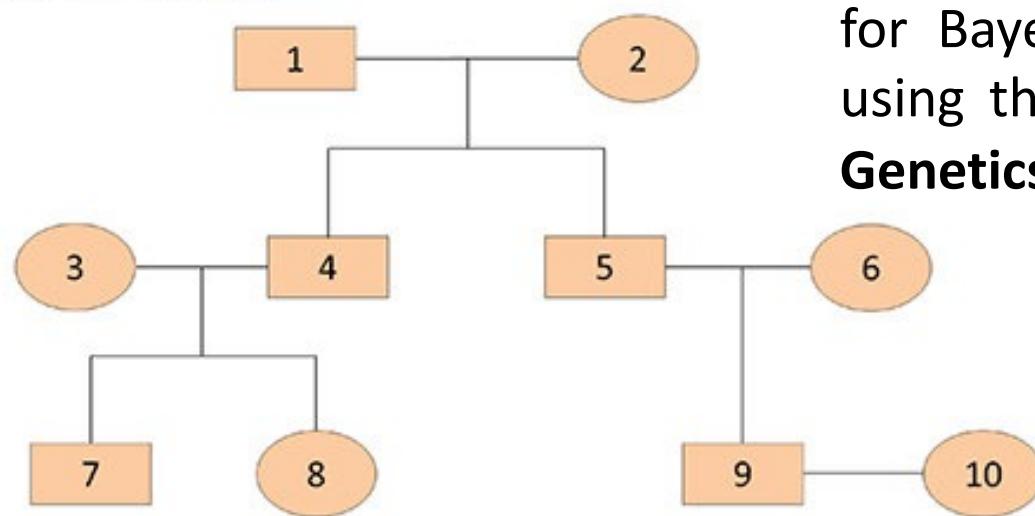
		Homologous chromosomes				
		ID 1	ID 2	ID 3	ID 4	ID 5
Pos5	0	1	0	0	0	0
	0	1	1	0	0	1
Pos9	0	1	1	0	0	0
	0	1	1	0	0	0

Animal	Pos5	Pos9
ID1	1	1
ID2	0	2
ID3	1	0
ID4	0	0
ID5	2	0

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2\sum p_i (1 - p_i)}$$

# SNPs and quantitative traits: SNP Matrix

A Example Pedigree



Bae et al (2014). An efficient technique for Bayesian modeling of family data using the BUGS software. **Frontiers in Genetics**, 5, 390

B Corresponding Additive Genetic Relationship Matrix

	1	2	3	4	5	6	7	8	9	10
1	1	0	0	0.5	0.5	0	0.25	0.25	0.25	0
2	0	1	0	0.5	0.5	0	0.25	0.25	0.25	0
3	0	0	1	0	0	0	0.5	0.05	0	0
4	0.5	0.5	0	1	0.5	0	0.5	0.5	0.25	0
5	0.5	0.5	0	0.5	1	0	0.25	0.25	0.5	0
6	0	0	0	0	0	1	0	0	0.5	0
7	0.25	0.25	0.5	0.5	0.25	0	1	0.5	0.125	0
8	0.25	0.25	0.5	0.5	0.25	0	0.5	1	0.125	0
9	0.25	0.25	0	0.25	0.5	0.5	0.125	0.125	1	0
10	0	0	0	0	0	0	0	0	0	1

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{K}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

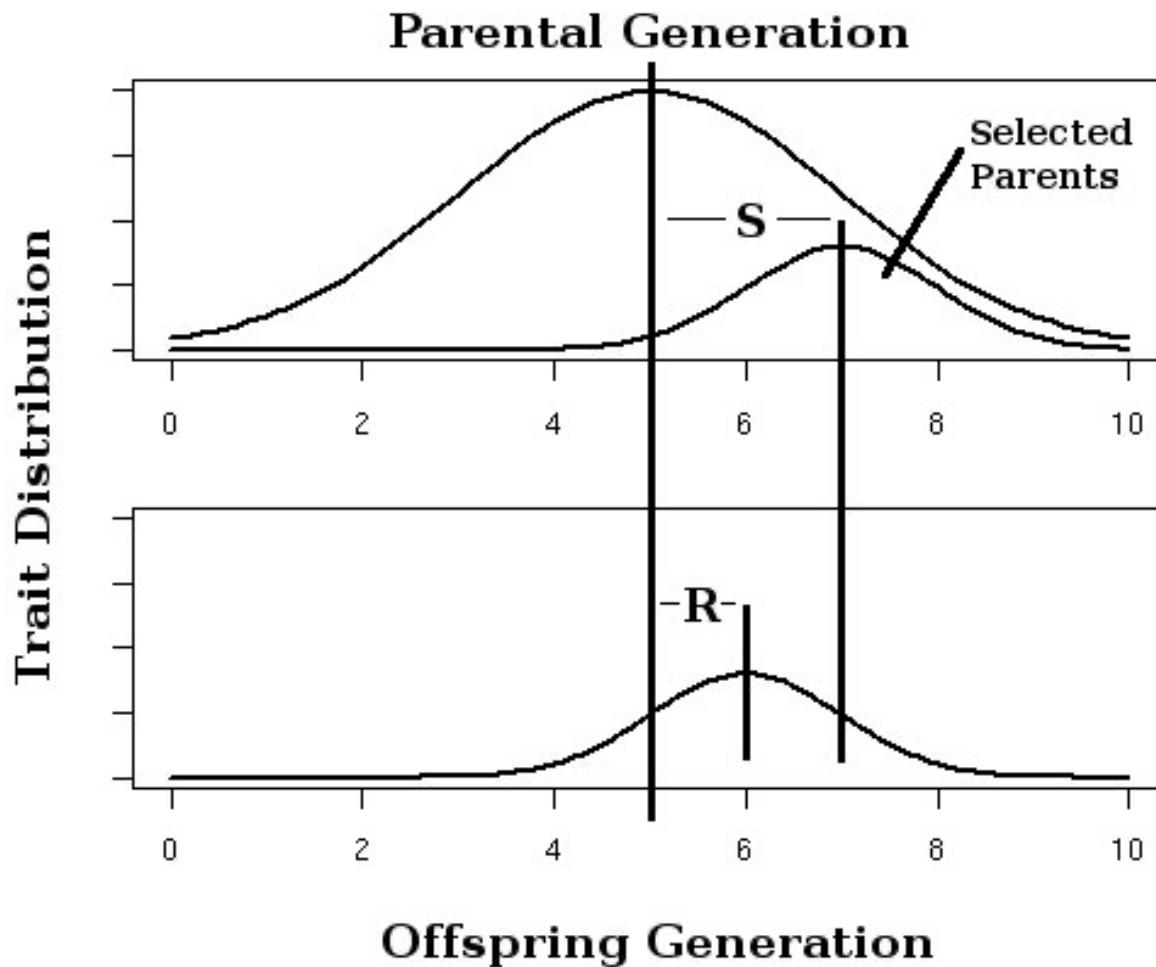
$\mathbf{K} = \mathbf{A}$  → pedigree-based matrix

$\mathbf{K} = \mathbf{G}$  → Genomic relationship matrix (Van Raden, 2008)

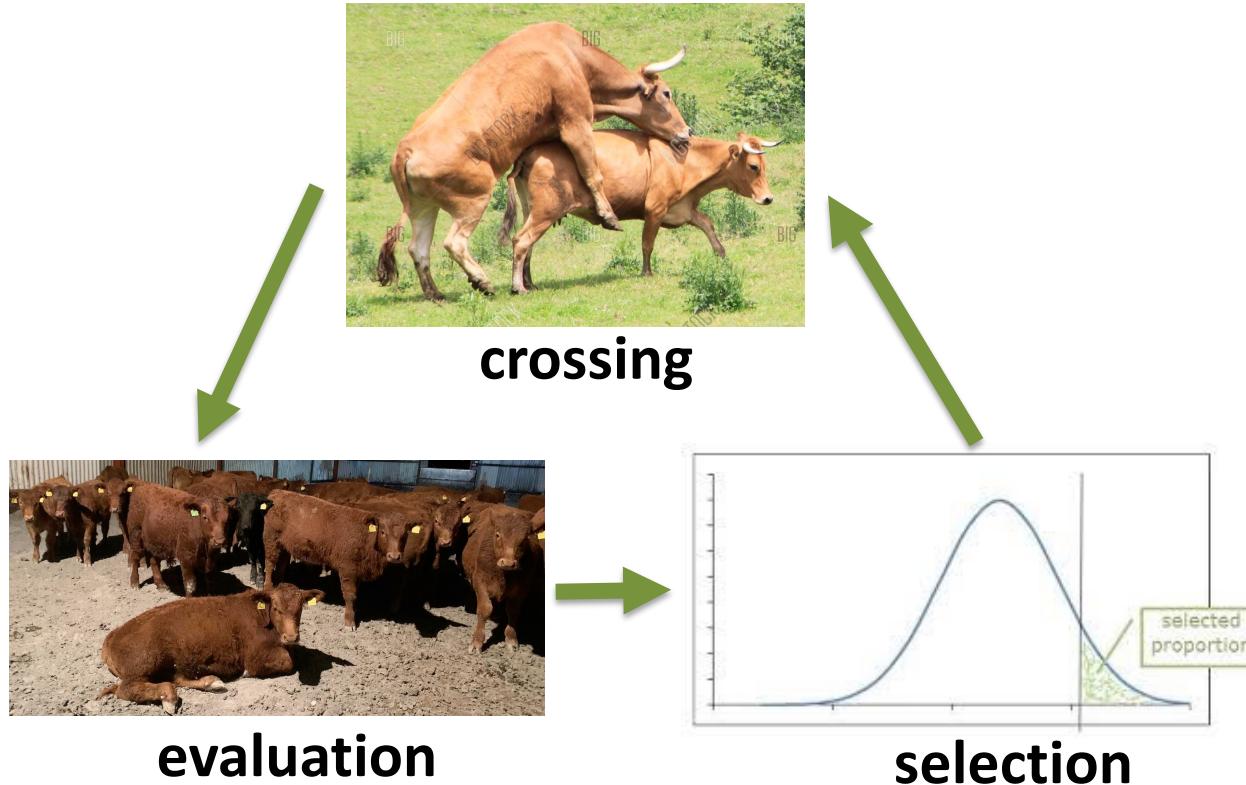
$\mathbf{K} = \mathbf{H}$  → Pedigree-based relationships augmented by genomic relationship matrix (Misztal et al. 2009).

$$\lambda = \frac{\sigma_e^2}{\sigma_a^2} \text{ or } \frac{\sigma_e^2}{\sigma_g^2} \text{ or } \frac{\sigma_e^2}{\sigma_u^2}$$

# SNPs and quantitative traits: Genetic gains



# SNPs and quantitative traits: Genetic gains



## Breeder's equation

$$\Delta G = \frac{i \cdot r \cdot \sigma_a}{L}$$

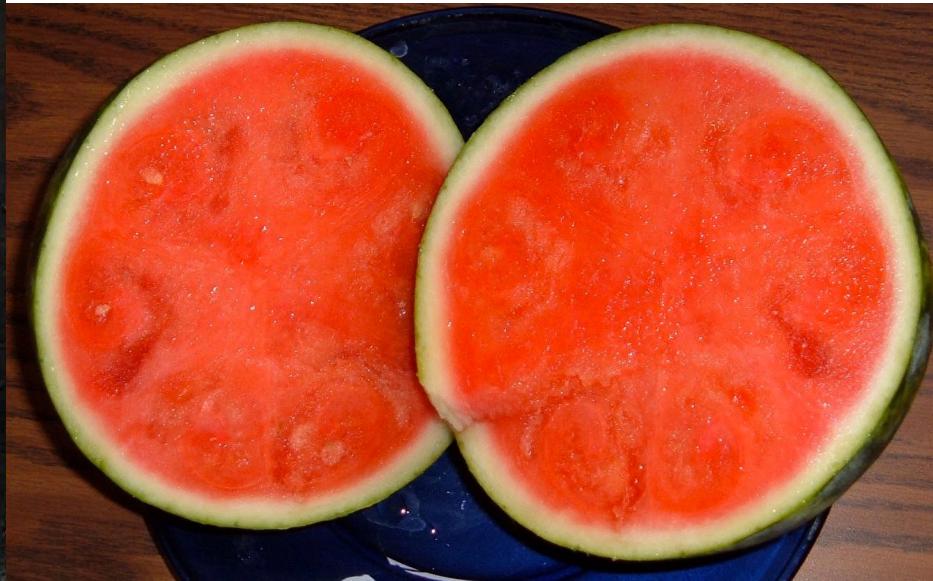
$\Delta G$  = genetic gain  
 $i$  = selection intensity  
 $r$  = selection accuracy  
 $\sigma_a$  = additive genetic variance  
 $L$  = generation interval

# SNPs and quantitative traits: Genetic gains



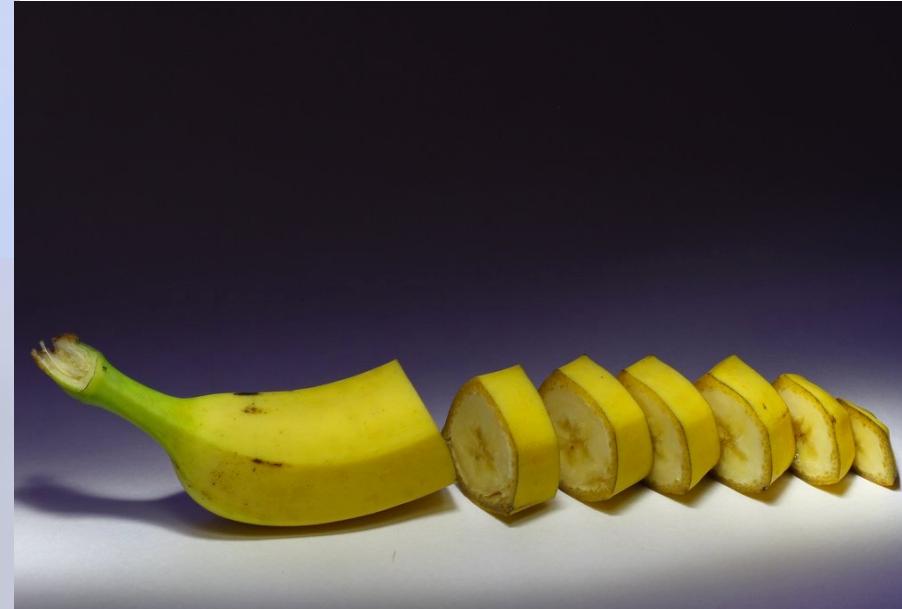
<https://commons.wikimedia.org/wiki/File:Maize-teosinte.jpg>

# SNPs and quantitative traits: Genetic gains



# SNPs and quantitative traits: Genetic gains

MICHIGAN STATE  
UNIVERSITY



# SNPs and quantitative traits: Genetic gains



# SNPs and quantitative traits: Genetic gains



Artificial selection from 1957 to 2001

ACRBC Males - 2001 Feed

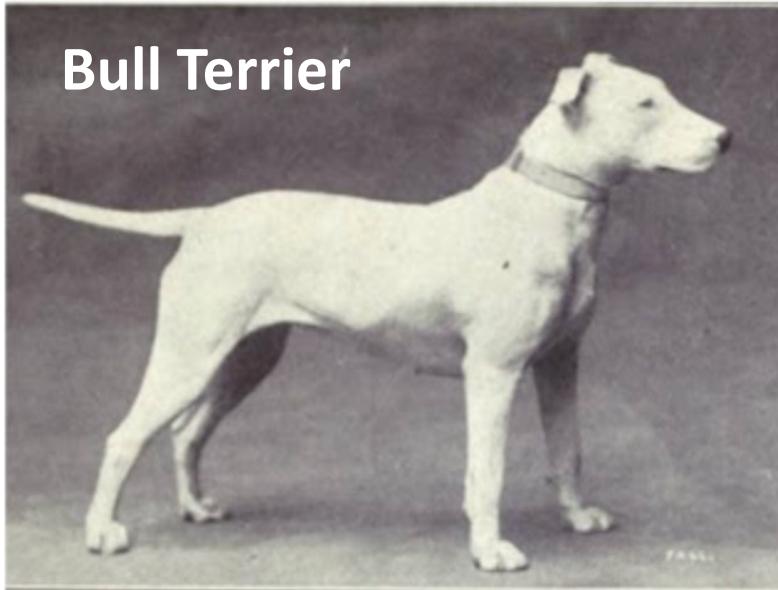


Ross Males - 2001 Feed



# SNPs and quantitative traits: Genetic gains

Bull Terrier



Boxer

