# Introduction to Genetics and Evolution

**Kenneth Kim**
Core Facility for Bioinformatics
Philippine Genome Center

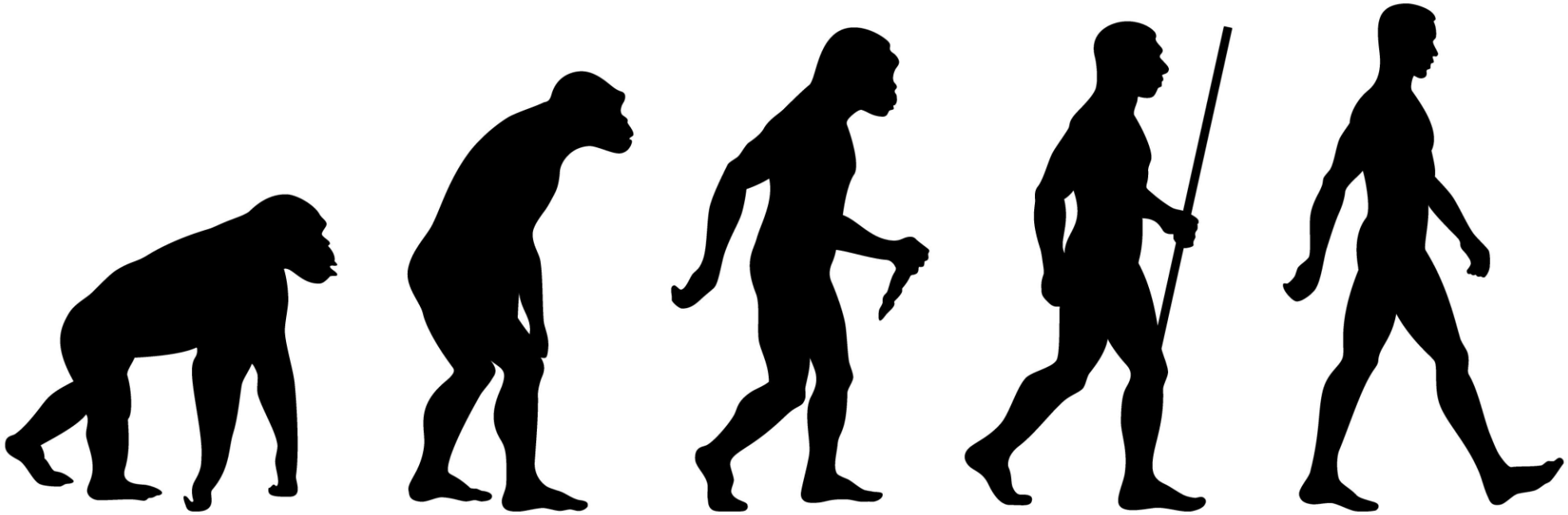
PGC
PHILIPPINE GENOME CENTER

Overview

Brief History – Charles Darwin on Natural Selection
Mendelian Inheritance and Cell Division
- Laws of Inheritance
- Codominance
- Sex Linked Genes
- Non-Mendelian Inheritance
- Continuous Variation
- Modern Synthesis

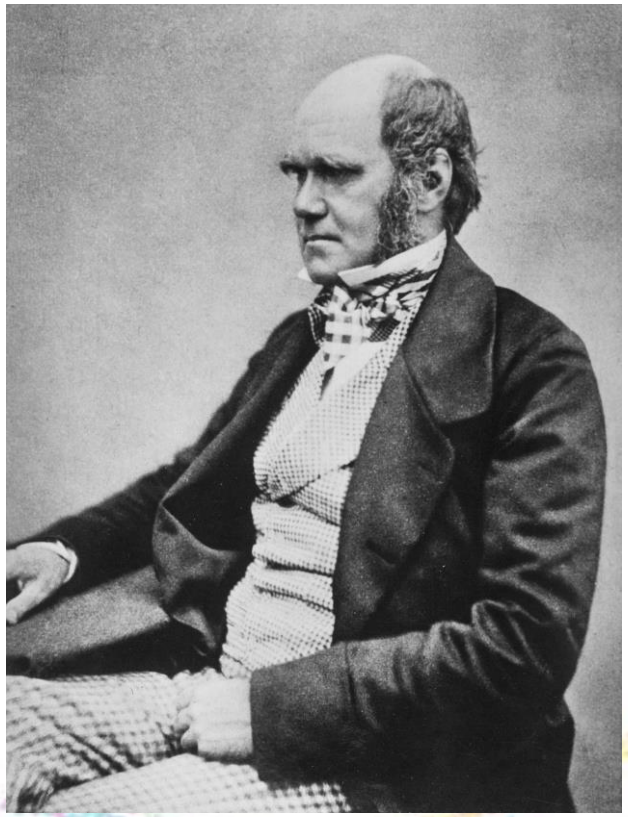
Forces of Evolution
Hardy-Weinberg Equilibrium
Detecting Molecular Variation

PGC
PHILIPPINE GENOME CENTER



PGC
PHILIPPINE GENOME CENTER

# Charles Darwin: Natural Selection

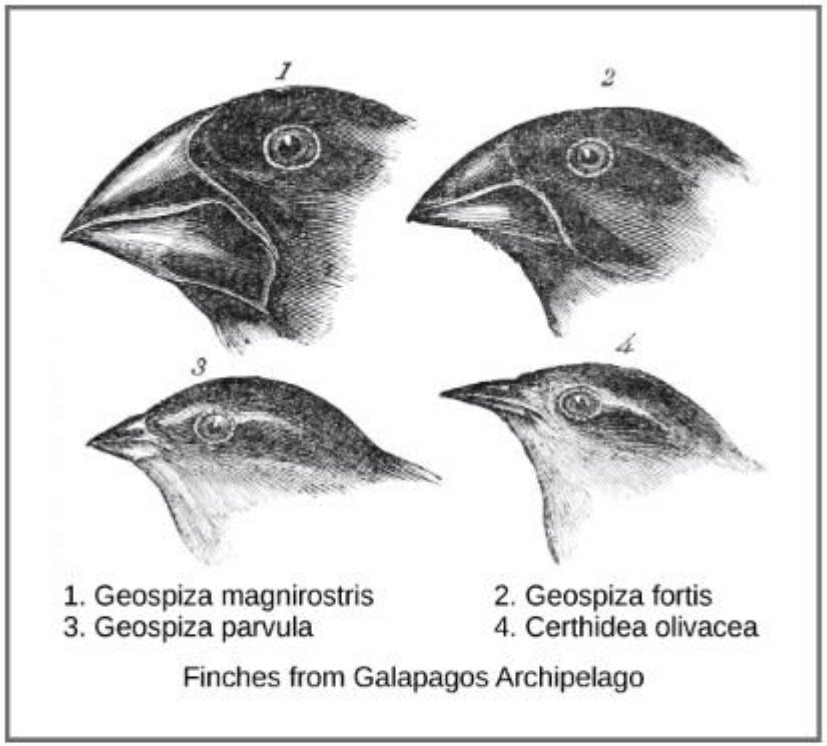


The theory of evolution **by natural selection** is the process by which organisms change over time as a result of changes in heritable trait. ("On the Origin of Species" in 1859)

Natural selection as the mechanism of change.

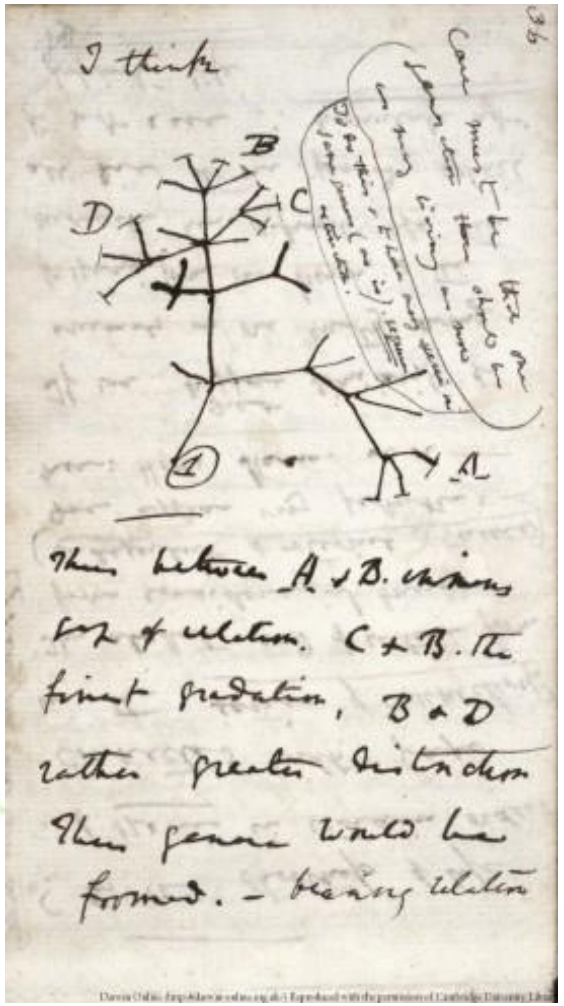
https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics

PGC
PHILIPPINE GENOME CENTER

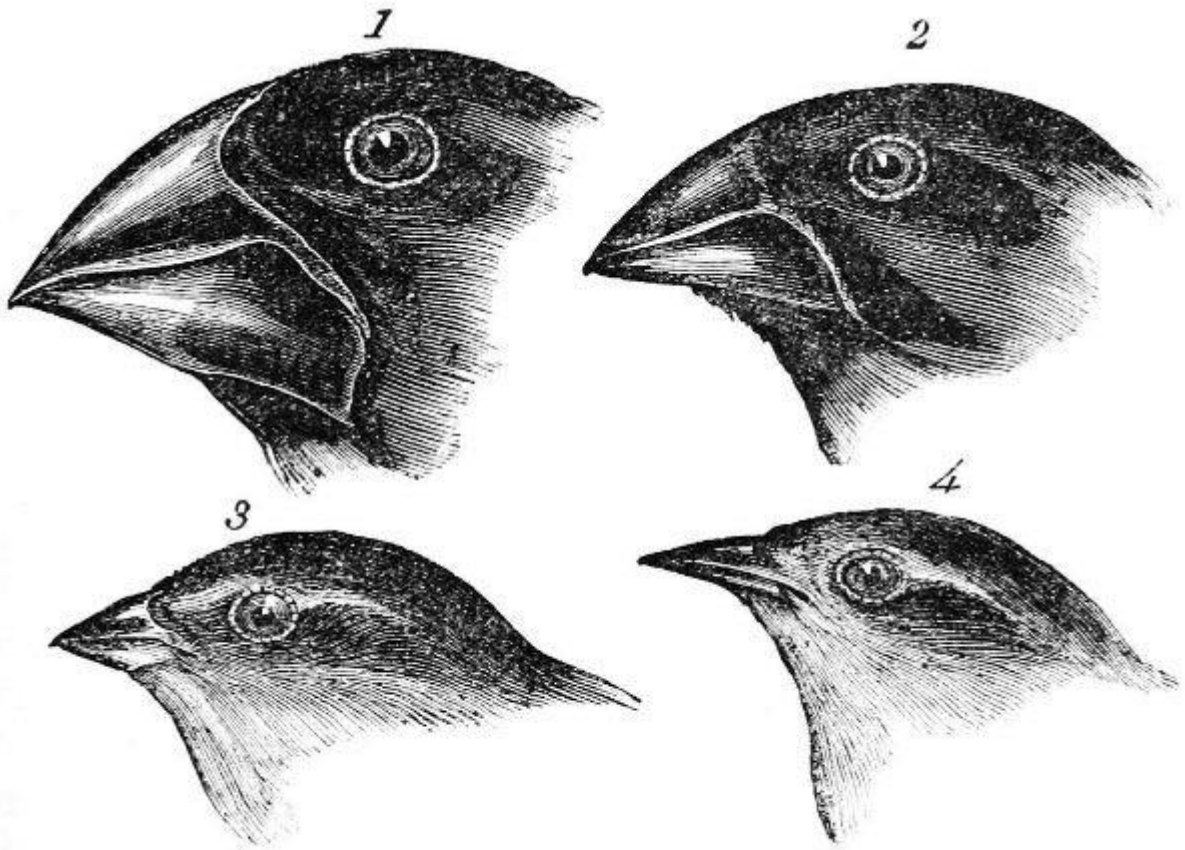



Finches from Galapagos Archipelago

**Beak Shape Among Finch Species**: Darwin observed that beak shape varies among finch species. He postulated that the beak of an ancestral species had adapted over time to equip the finches to acquire different food sources.




PGC
PHILIPPINE GENOME CENTER

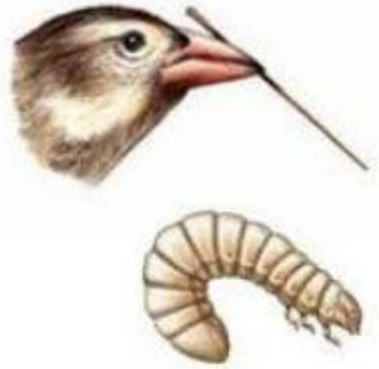



1. Geospiza magnirostris.
2. Geospiza fortis.
3. Geospiza parvula.
4. Certhidea olivacea.

Large ground finch (seeds)

Cactus finch
(cactus fruits and flowers)

Vegetarian finch (buds)



Woodpecker finch (insects)

PGC
PHILIPPINE GENOME CENTER

# Discrete Variation: Mendelian Inheritance

Principles of Mendelian inheritance (1866)

There exist hereditary factors, one of which is dominant. Each individual has two factors for each trait, one from each parent.

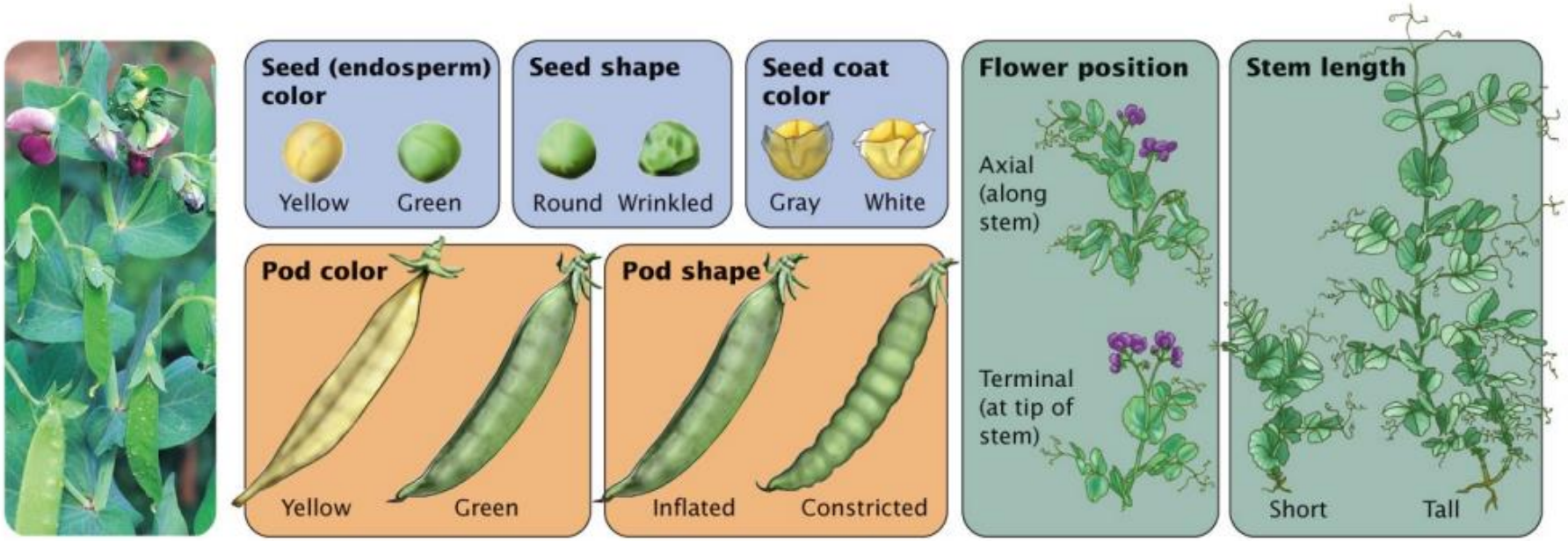
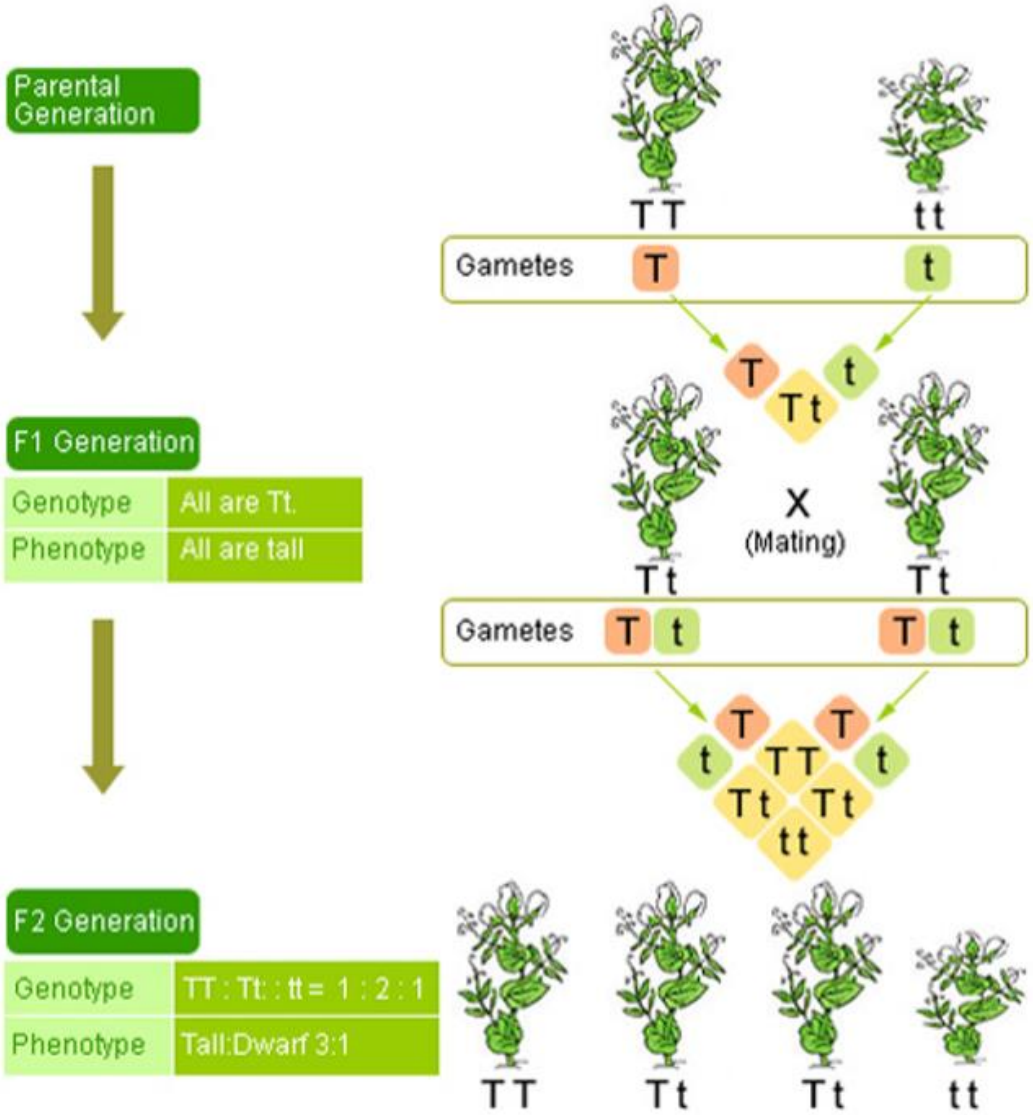



Fig_03-01 *Genetics, Second Edition* © 2005 W.H. Freeman and Company

https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics

PGC PHILIPPINE GENOME CENTER



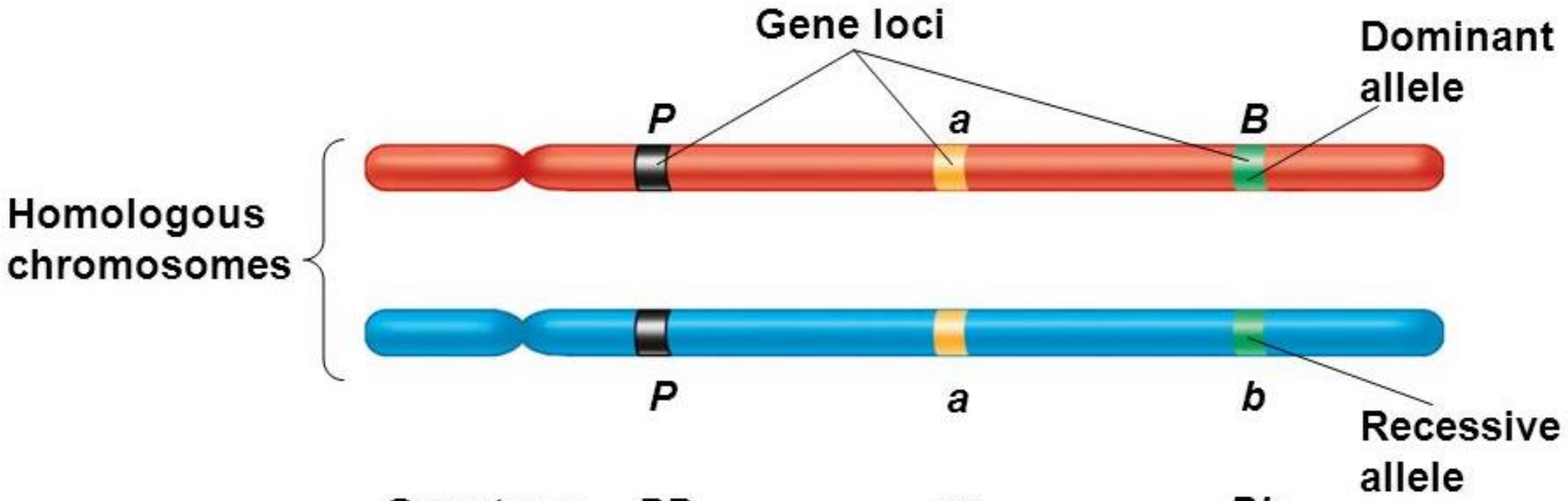

Law's of Inheritance

Law of Dominance: recessive alleles will always be masked by dominant alleles.

Law of Segregation: each gamete receives only one copy of a gene (allele)

Law of Independent Assortment: alleles of different genes sort independently during gamete formation

PGC
PHILIPPINE GENOME CENTER




**Genotype:** ***PP*** ***aa*** ***Bb***

**Homozygous for the dominant allele**

**Homozygous for the recessive allele**

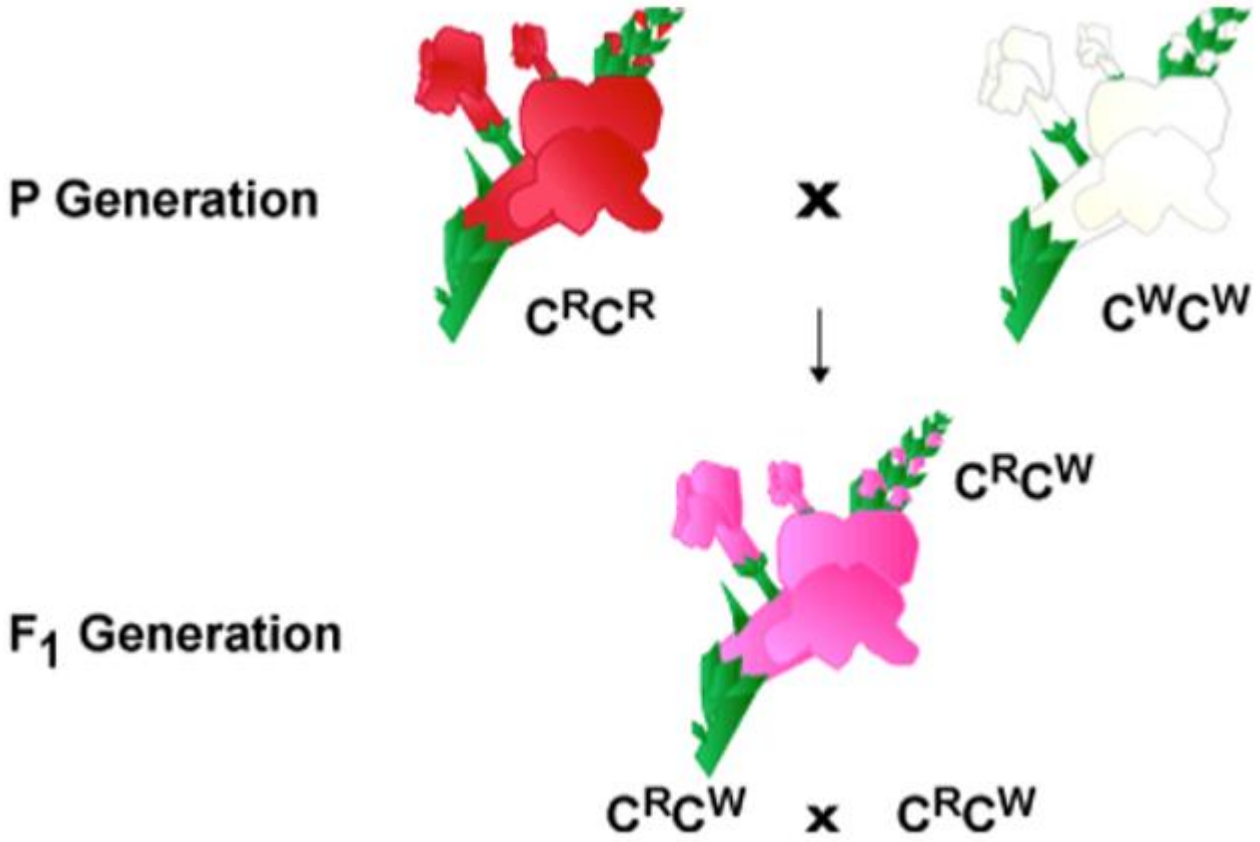
**Heterozygous, with one dominant and one recessive allele**



http://www.agilent.com/labs/features/2013_oliver.html

**Dominance may be incomplete or partial**

Codominance, Incomplete dominance
Example:  Snapdragon (Antirrhinum majus)
Incomplete dominance
Red x white = pink



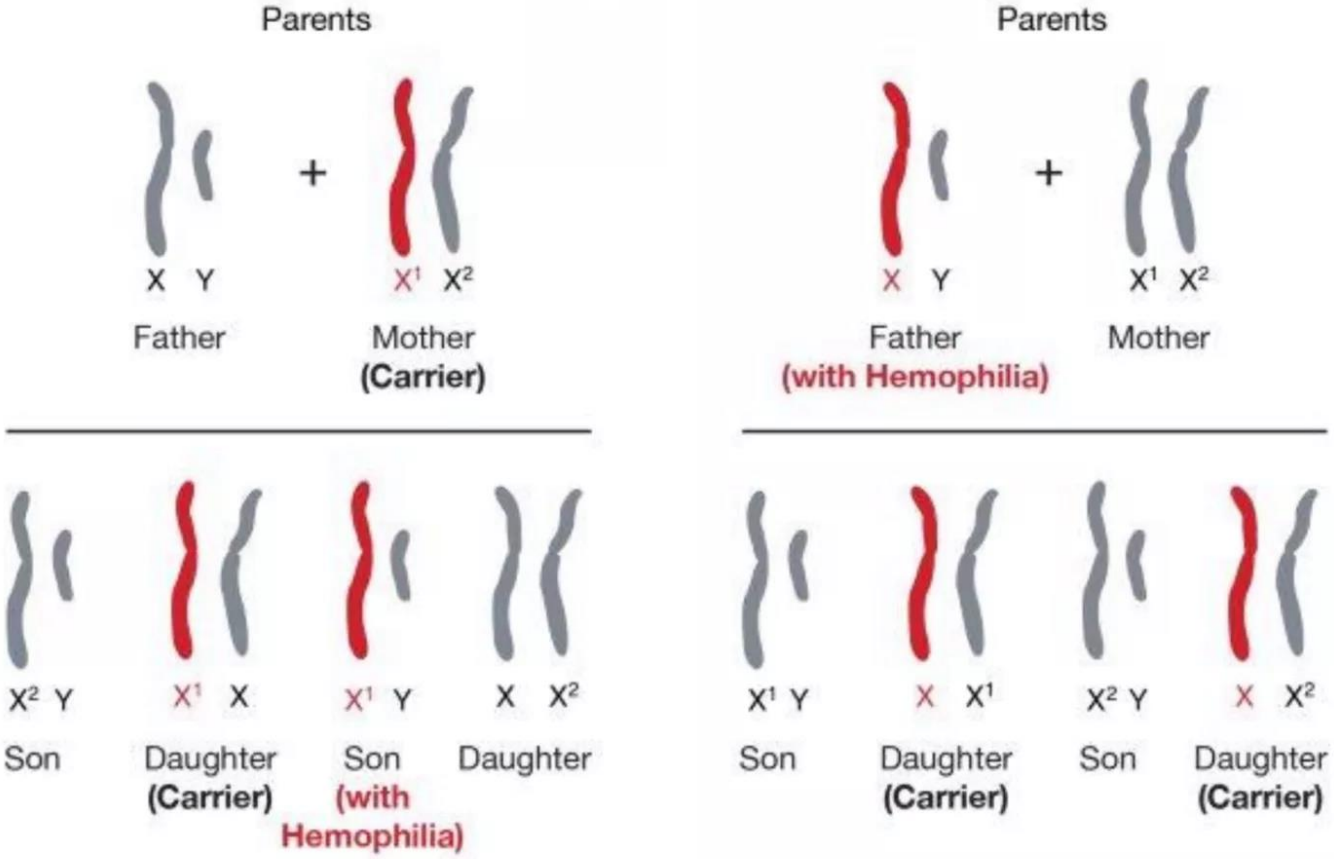

http://www.agilent.com/labs/features/2013_oliver.html

PGC
PHILIPPINE GENOME CENTER

**Genes that are found on sex chromosomes are called sex-linked genes**

If a gene is located on the Y chromosome, it is a **Y-linked gene**. These genes are only inherited by males because, in most instances, males have a genotype of **(XY)**. Females do not have the Y sex chromosome. Genes that are found on the X chromosome are called **X-linked genes**.

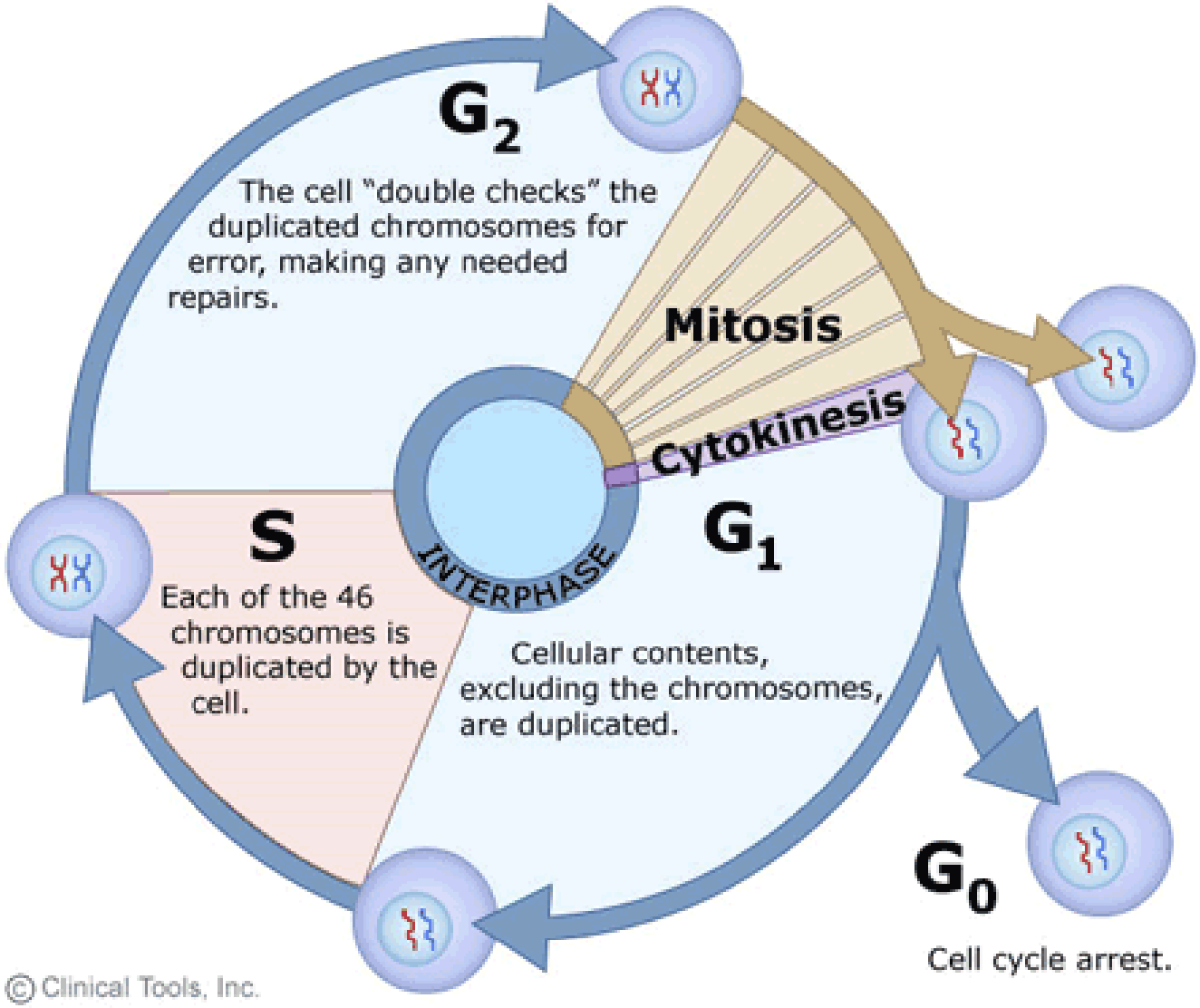
## Hemophilia




http://www.agilent.com/labs/features/2013_oliver.html

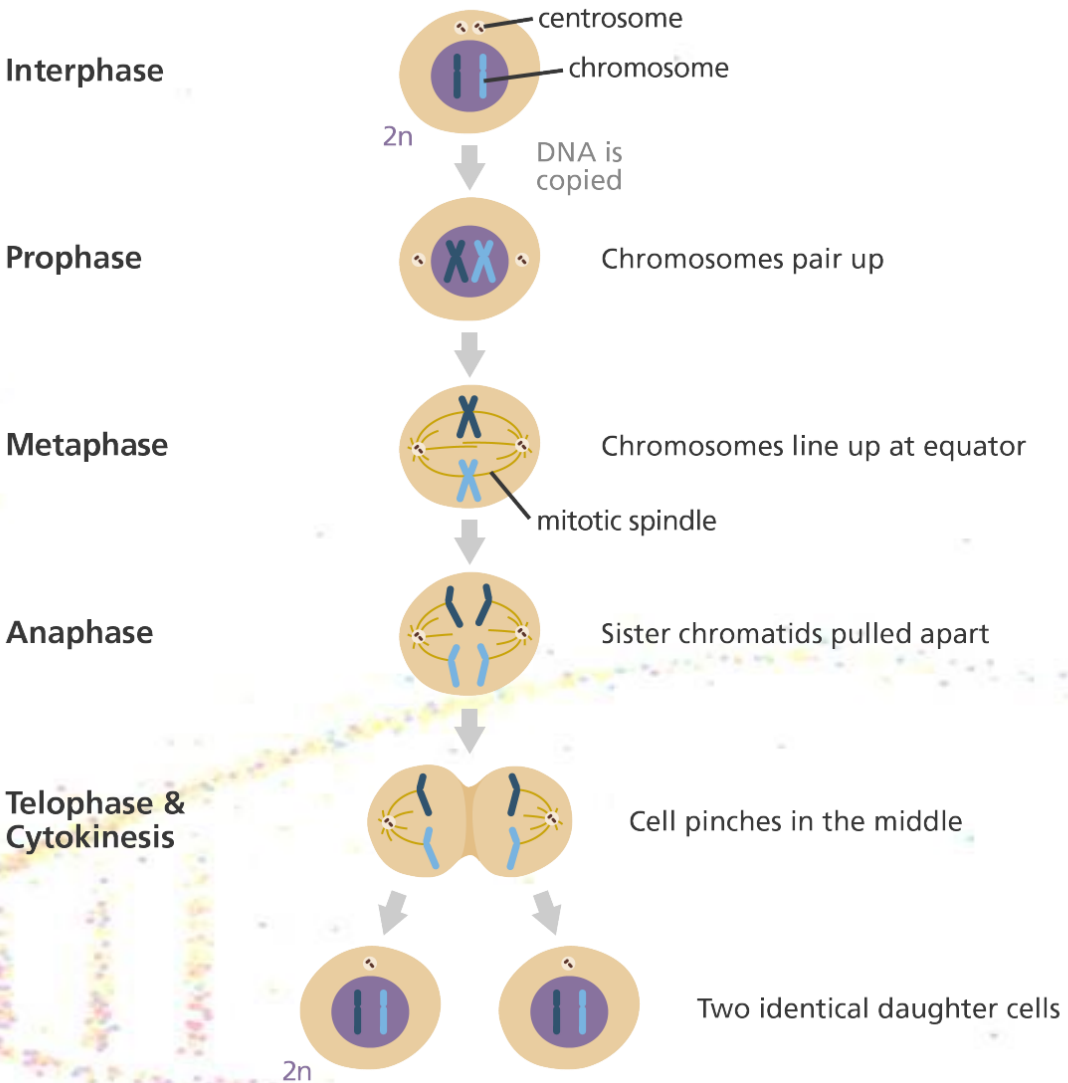
PGC
PHILIPPINE GENOME CENTER

# Cell Cycle



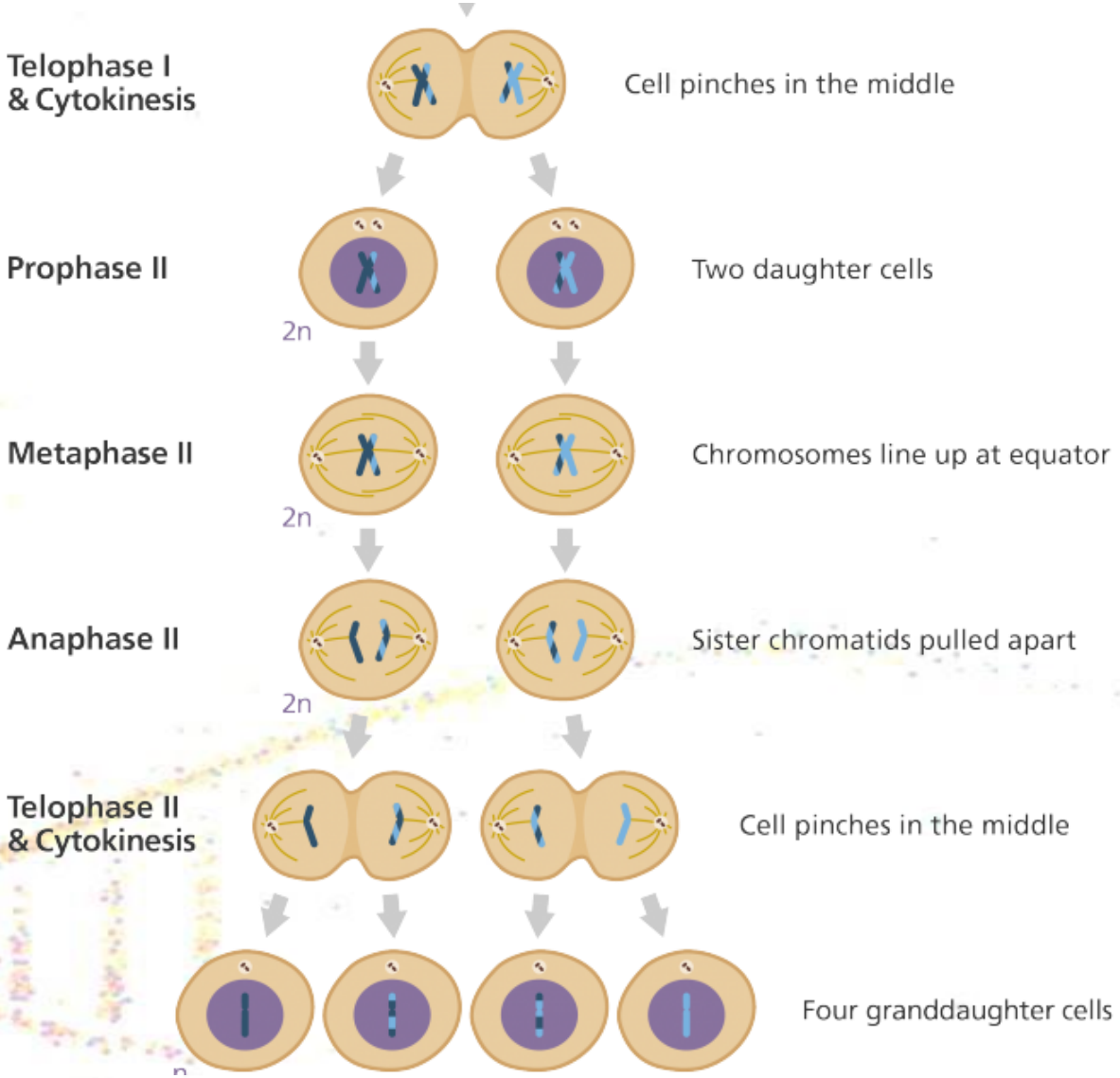

PGC
PHILIPPINE GENOME CENTER

# Cell Division: Mitosis



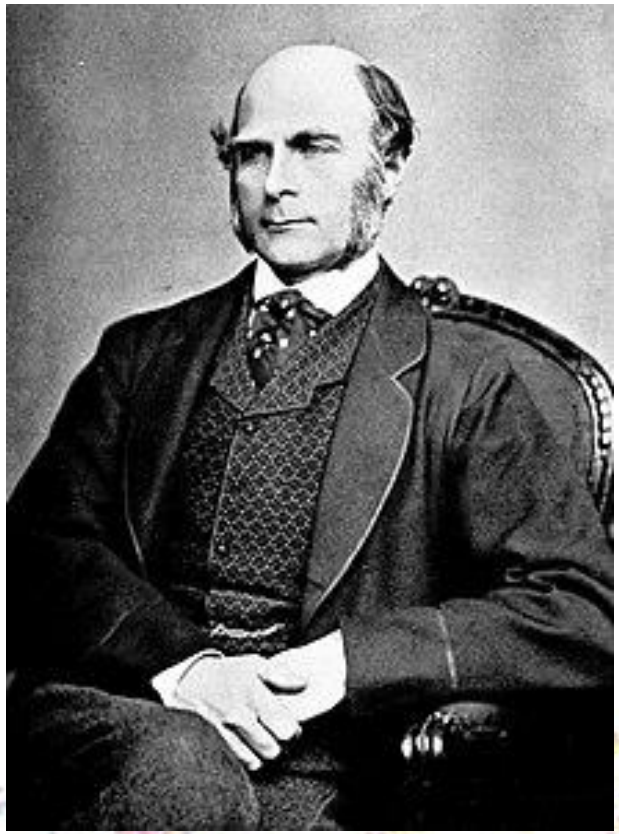

PGC
PHILIPPINE GENOME CENTER

# Cell Division: Meiosis




PGC
PHILIPPINE GENOME CENTER

# Continuous Variation: Biometrics



Continuous variation does not show a few discrete alternative states but is found as a continuum in a population

e.g. Height in Humans, Skin color, Milk yield in cattle, IQ (intelligence quotient), Time to run 100m, Weight in humans

Sir Francis Galton

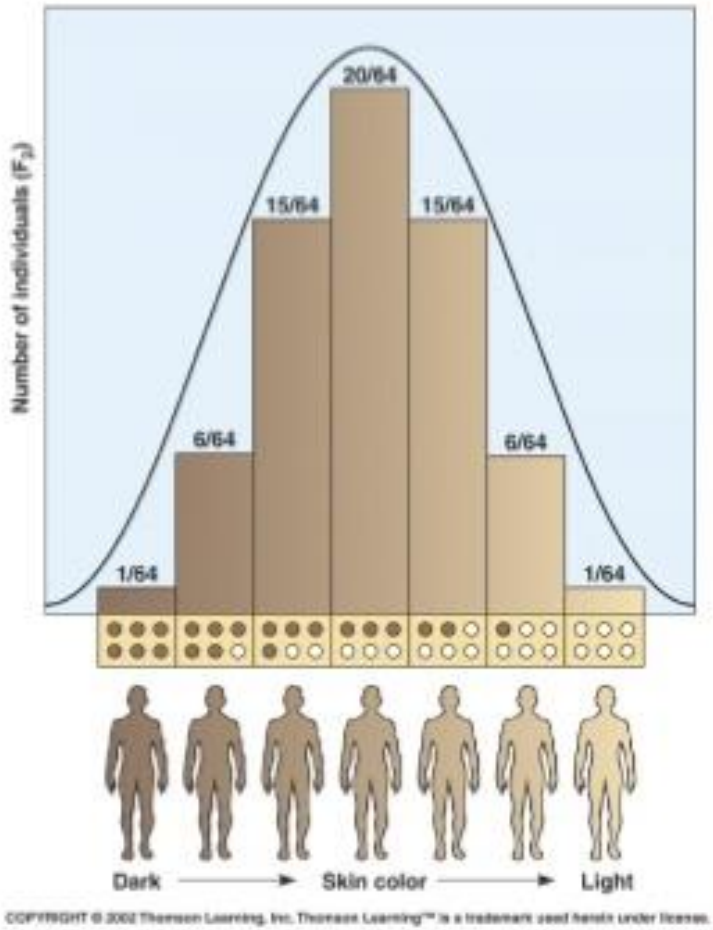
-introduced the Model for population stability and made a Statistical description of continuous traits. Continuous traits typically fit a normal distribution and they have a multifactorial inheritance meaning they are polygenic.

https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics

PGC
PHILIPPINE GENOME CENTER

# Polygenic Traits

- The control of a trait by more than one gene
  - Skin color is controlled by at least 6 genes
- Each gene product is **additive** to the others
- The hallmark of a polygenic trait's phenotype expression is:
  - A bell curve distribution
  - A continuous distribution





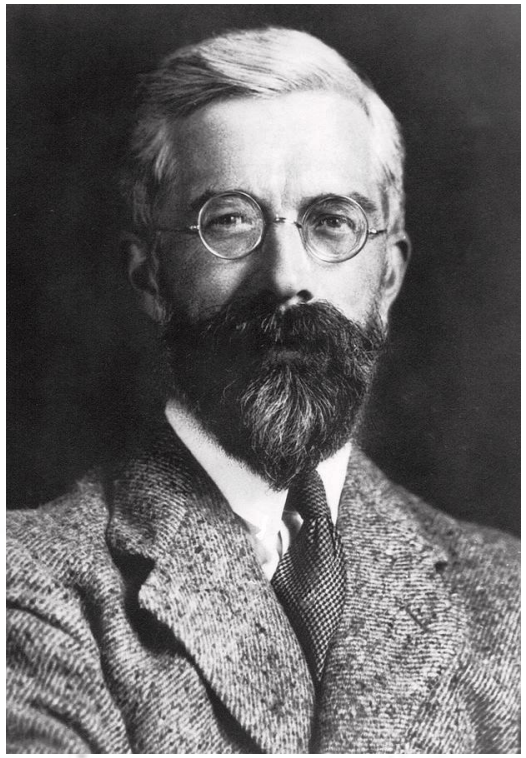
PGC
PHILIPPINE GENOME CENTER

# The inheritance pattern of discrete variation (mendelian) vs continuous variation.

Are they different?

PGC
PHILIPPINE GENOME CENTER

# Unifying Biometrics and Mendelian Inheritance



R.A. Fisher (1918): Biometrics meets Mendelian inheritance

- Inheritance of Continuous traits consistent with Mendelian genetics
- Additive effects of “many unblending particles of inheritance” could result in a normal distribution of the trait in a population
- Effects of genes are cumulative
- No one gene is dominant or recessive

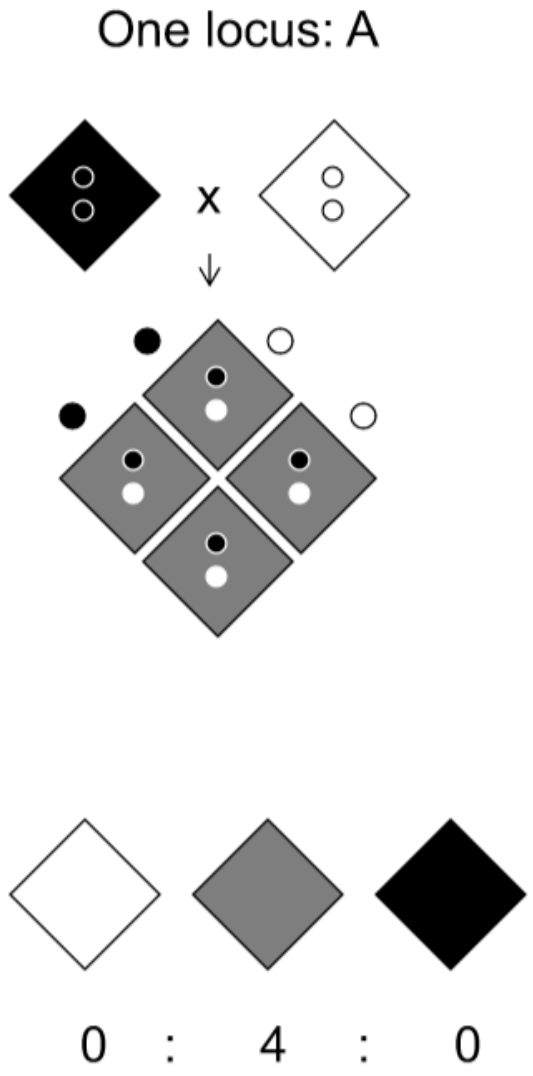
Fisher, R. A. (1918). The Correlation between Relatives on the Supposition of Mendelian Inheritance. Transactions of the Royal society of Edinburgh, 52(02), 399-433.

PGC
PHILIPPINE GENOME CENTER

Consider a hypothetical trait:
Phenotype: Hair color

Polygenic trait
Additive effects of 2 alleles at multiple loci

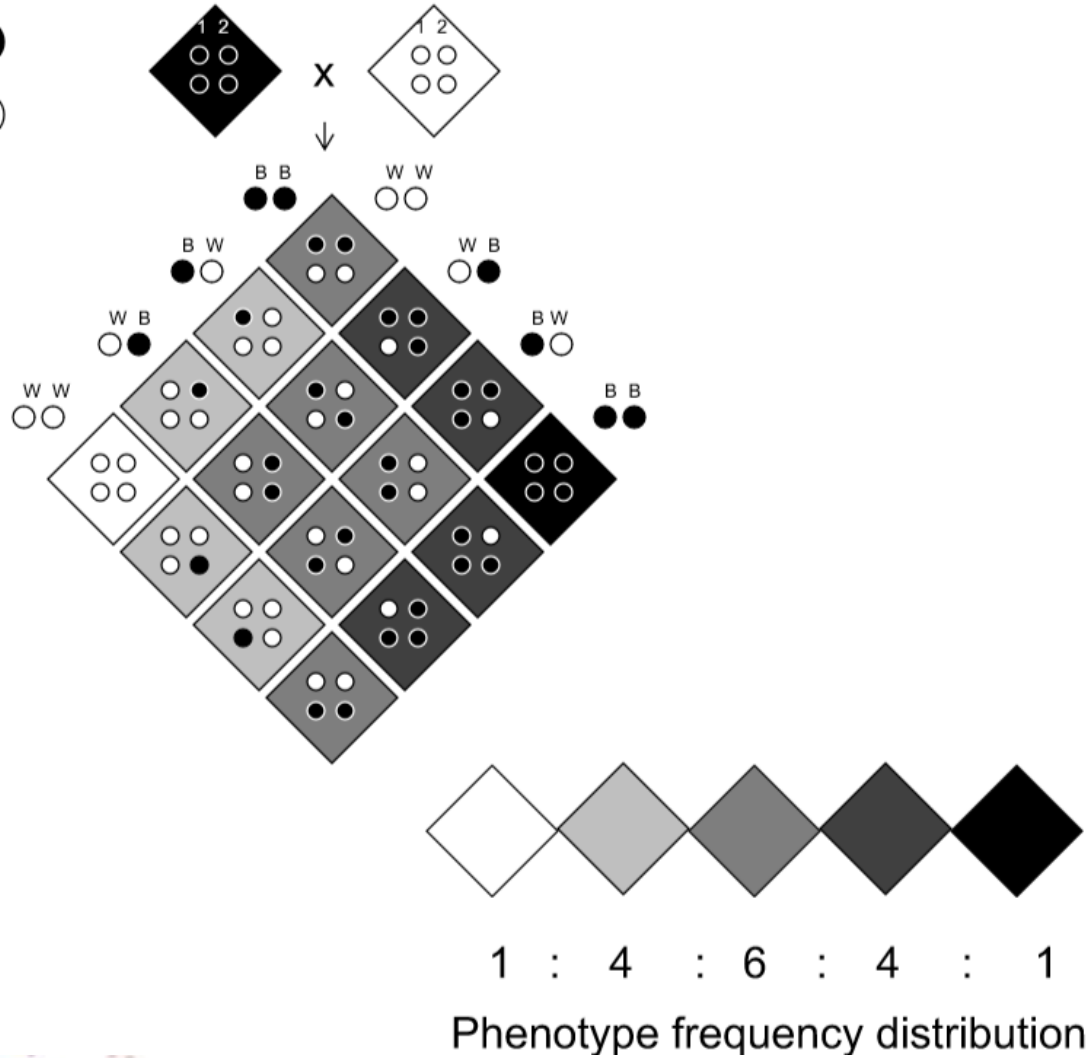
‘Black’ allele
‘White’ allele




PGC
PHILIPPINE GENOME CENTER

# Additive effects of 2 alleles at multiple loci

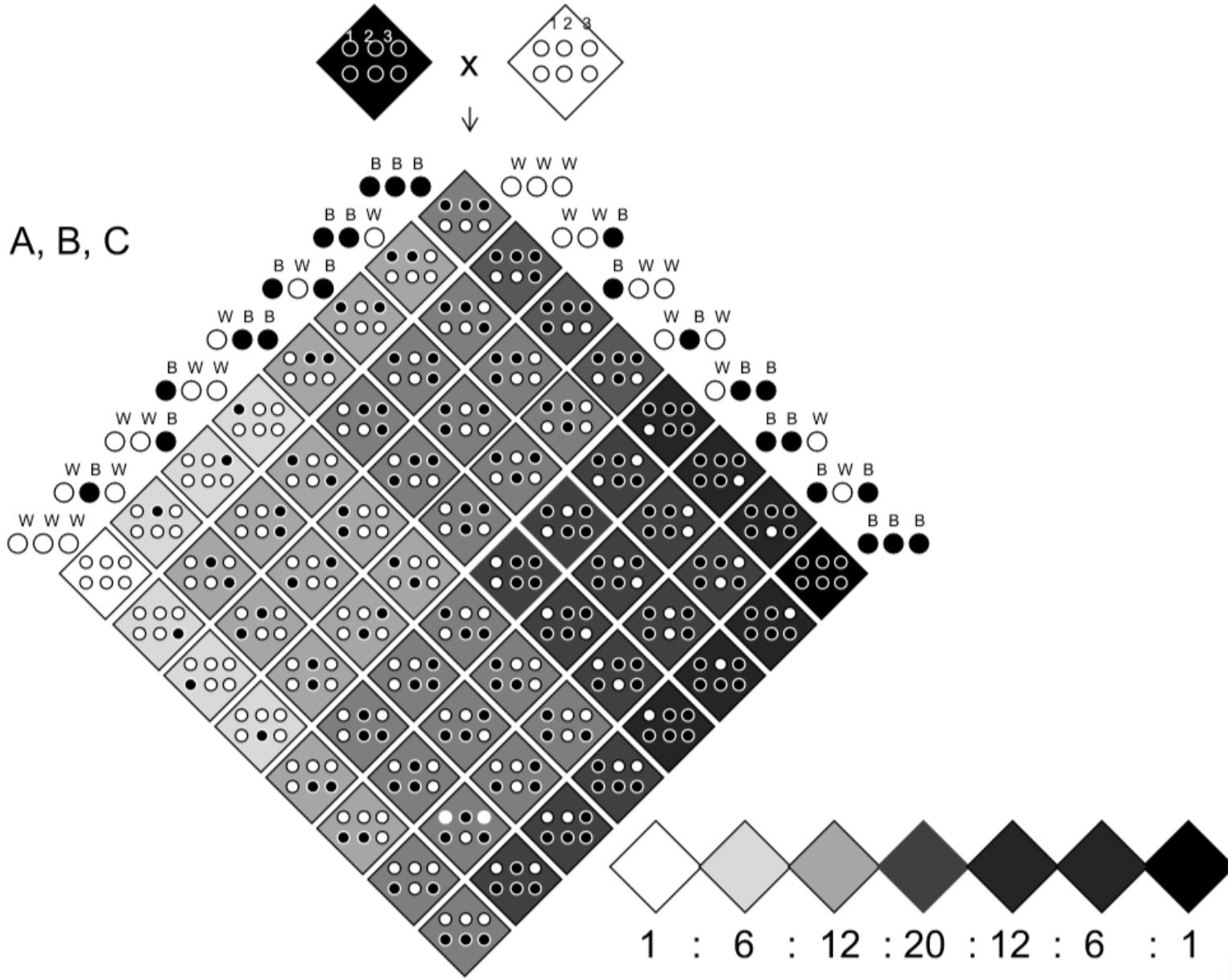
‘Black’ allele (B)
‘White’ allele (W)




http://www.agilent.com/labs/features/2013_oliver.html

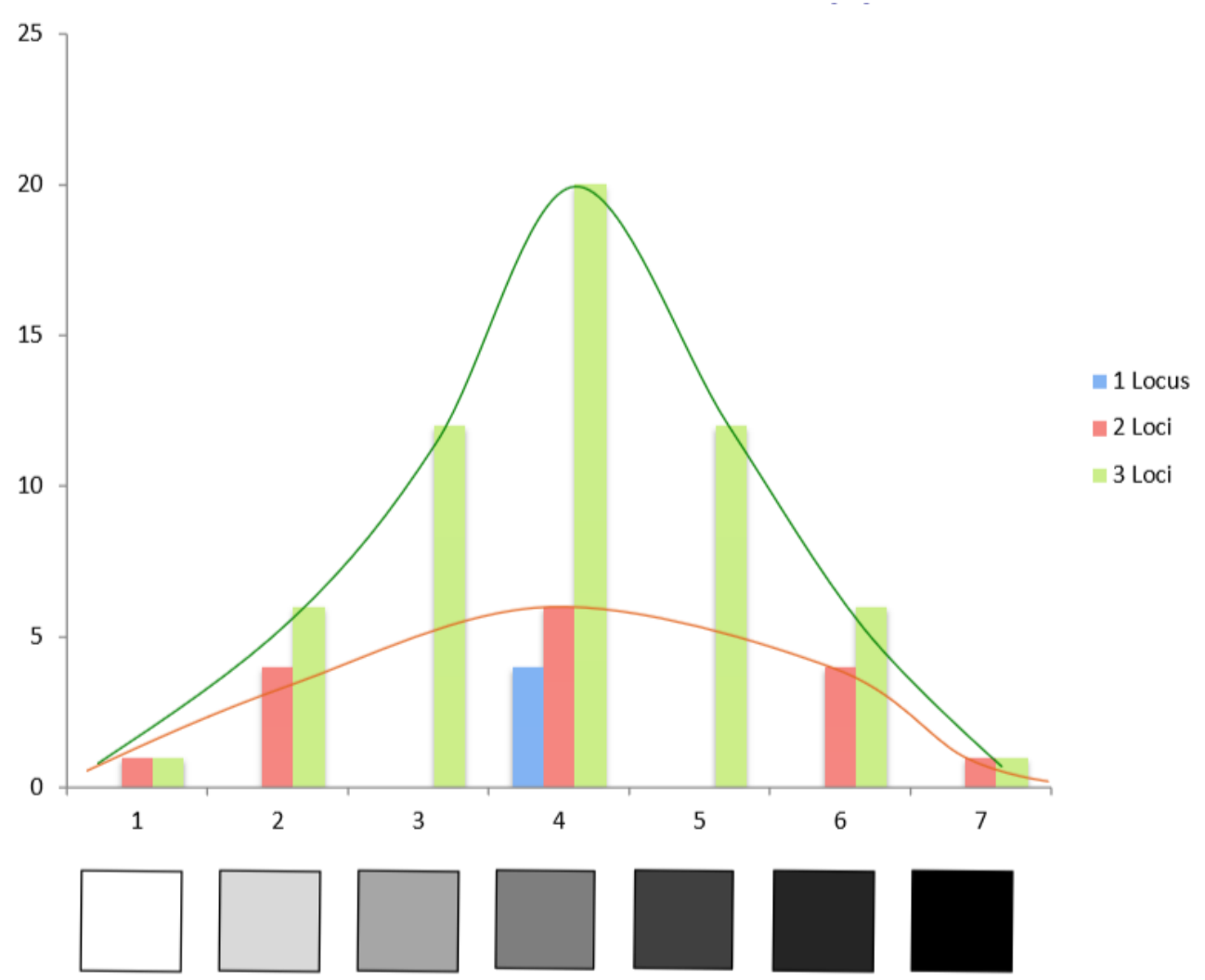
PGC
PHILIPPINE GENOME CENTER

Three loci: A, B, C




http://www.agilent.com/labs/features/2013_oliver.html

PGC
PHILIPPINE GENOME CENTER




http://www.agilent.com/labs/features/2013_oliver.html

PGC
PHILIPPINE GENOME CENTER

# Modern Synthesis (Huxley, 1943)

Natural Selection
(Charles Darwin)

The Correlation between Relatives on the Supposition of Mendelian Inheritance.
(RA Fisher 1918)

Discrete Variation:
Mendelian Inheritance
(Gregor Mendel)

Continuous Variation
(Francis Galton)

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4318712/pdf/nihms622574.pdf an interesting paper to read

# Population Genetics

PGC
PHILIPPINE GENOME CENTER

**Population genetics** is the study of genetic variation within populations, and involves the examination and modeling of changes in the frequencies of genes and alleles in populations over space and time.

PGC
PHILIPPINE GENOME CENTER

# Mutation

Mutation: Ultimate source of new variation

Point mutations, SNPs
Insertions, deletions = frameshift
Large-scale mutations in chromosomal structure
    Chromosomal inversions, translocations
    Transposable elements


PGC
PHILIPPINE GENOME CENTER

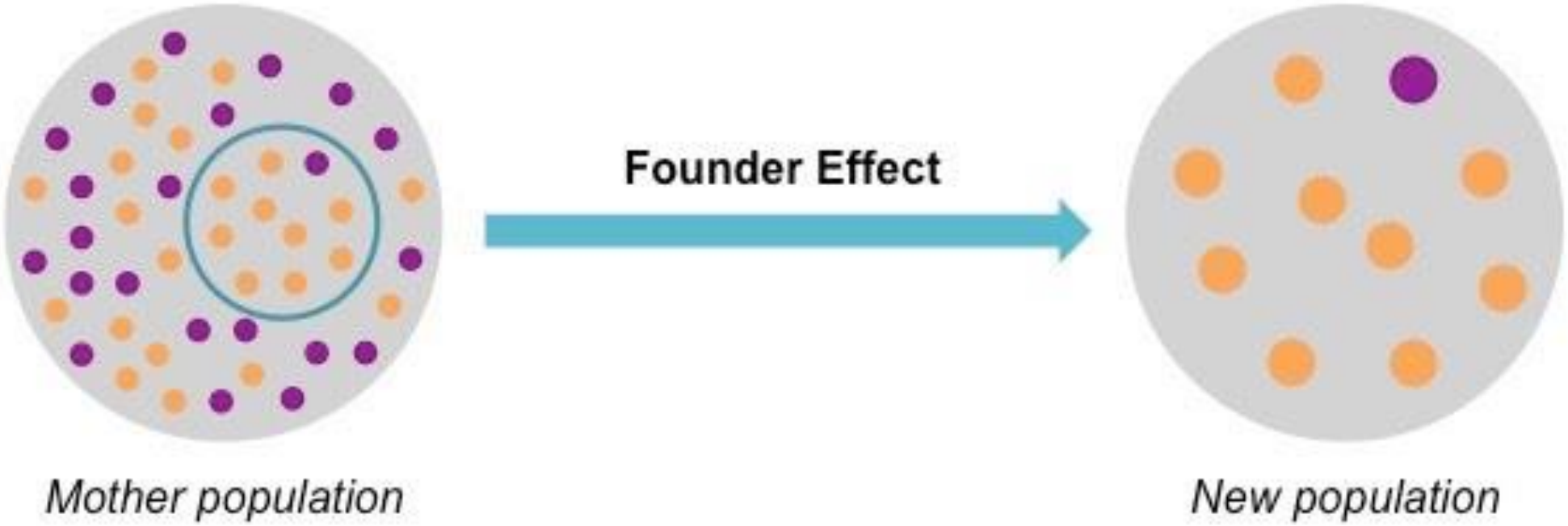
# Genetic Drift

Stochastic change in allele frequency due to random sampling in a finite population

Two types of Drift

- Founder Effect
- Bottleneck Effect

PHILIPPINE GENOME CENTER


Founder Effect
Mother population
New population

# Population bottleneck

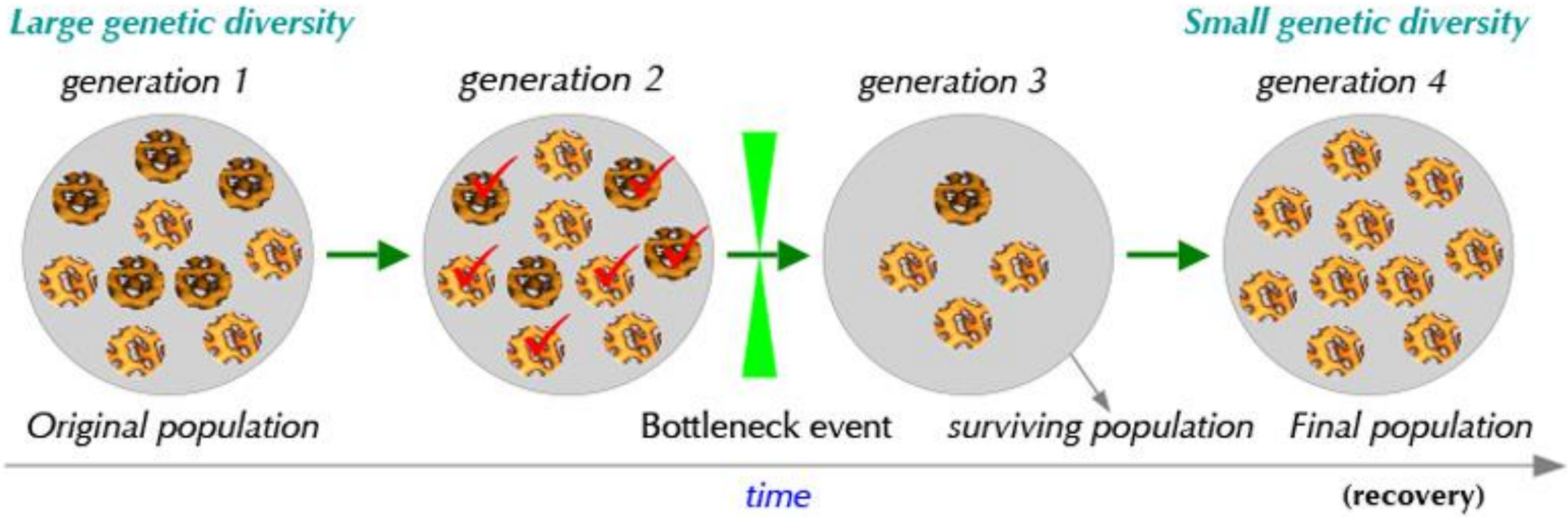
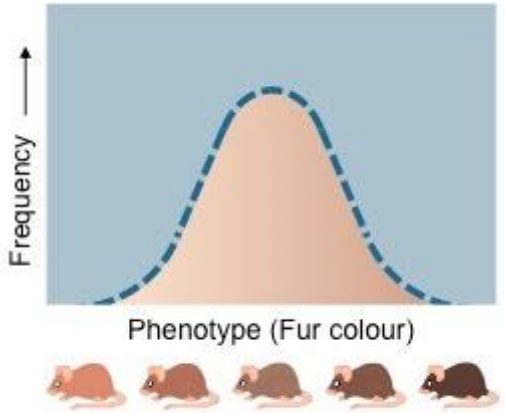



Image design: COSNET Lab

PGC
PHILIPPINE GENOME CENTER

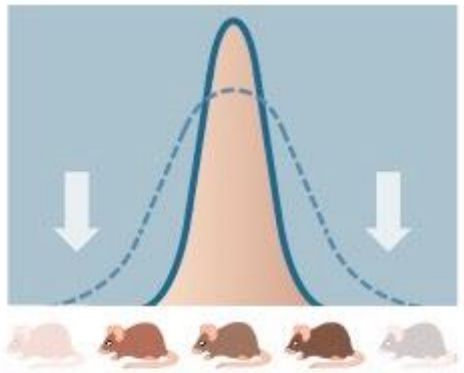
# Selection

Natural selection is the differential survival and reproduction of individuals due to differences in phenotype. It is a key mechanism of evolution, the change in the heritable traits characteristic of a population over generations.




**Normal Distribution**

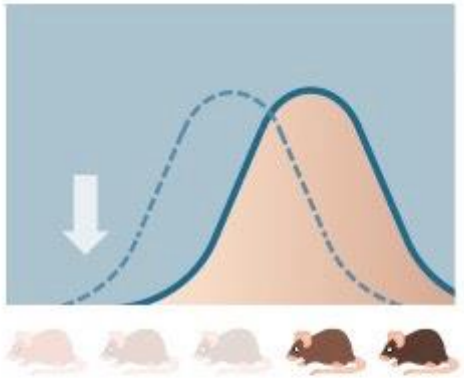
Gaussian (bell-shaped) trend



**Stabilising Selection**

Culls extreme variations

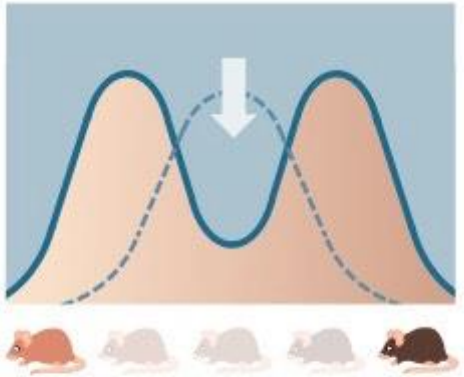
Narrows width of distribution



**Directional Selection**

Favours one extreme

Shifts distribution left / right



**Disruptive Selection**

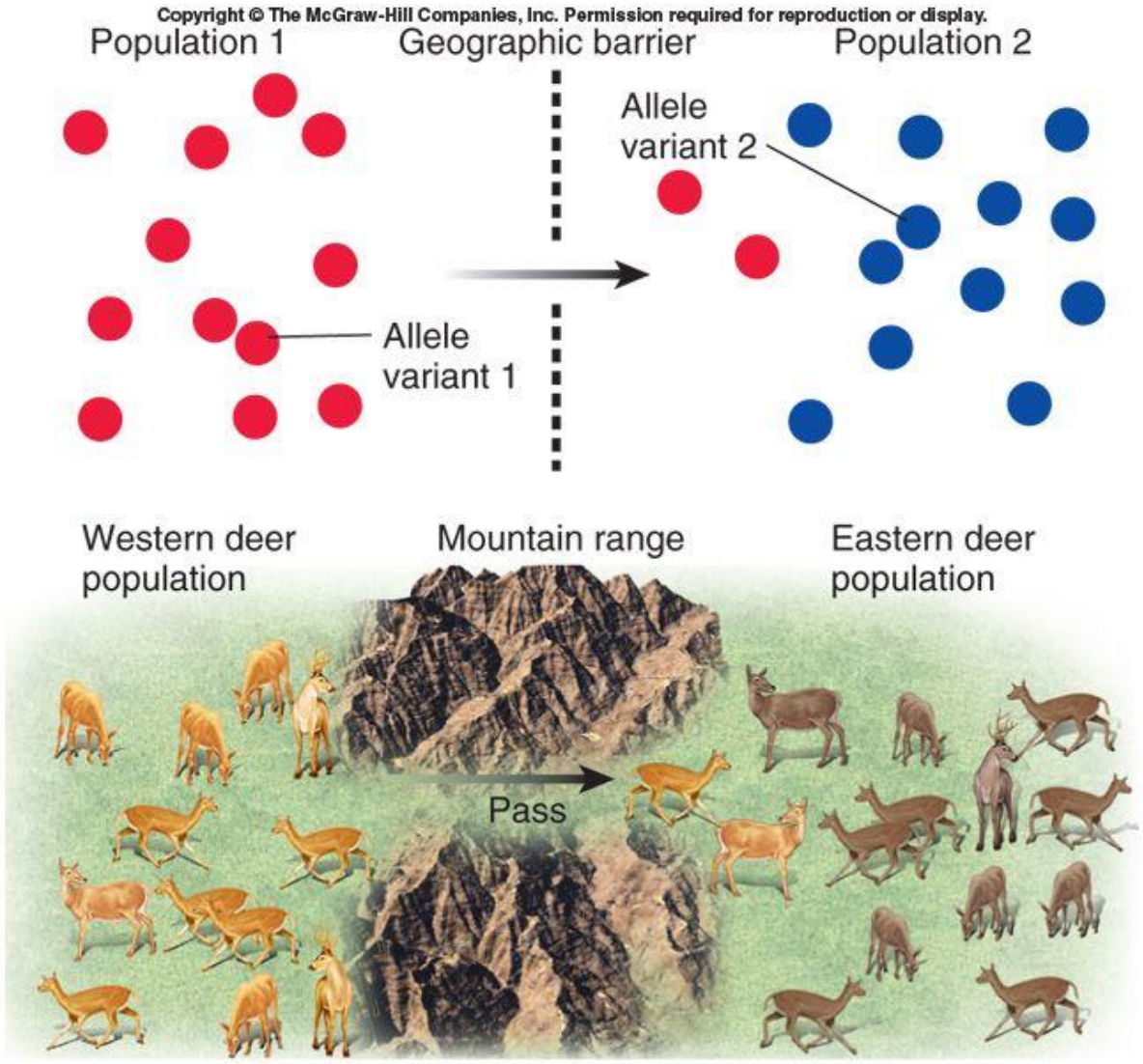
Favours both extremes

Creates bimodal distribution

PGC
PHILIPPINE GENOME CENTER

# Gene Flow

Gene flow is the movement of alleles from one genetic pool (population) to another.




PGC
PHILIPPINE GENOME CENTER

# Hardy-Weinberg Equilibrium (HWE)

Allelic frequencies will remain the same from one generation to the next with the following assumptions:

There is no mutation, no migration, no selection, population is infinitely large and randomly mating

$$(p + q)^2 = 1$$

$$p2 + 2pq + q2 = 1$$

For a diploid organism, consider a 2 allele locus (A, a), where A is dominant

Let:
*p* = frequency of allele A
*q* = frequency of allele a

PGC
PHILIPPINE GENOME CENTER

## Violations of HWE

- Small population size
- Deviations from random mating
- Assortative mating
- Disassortative mating
- Inbreeding
- Population structure
- Mutation
- Migration
- Selection

# Testing for deviations from HWE expectations:

Chi-square test: goodness of fit between observed & expected frequencies

| Genotype | Observed | Expected | |
|---|---|---|---|
| MM | 165 | | |
| MN | 562 | | |
| NN | 339 | | |
| Total | 1066 | | |

1. Calculate allele frequencies p and q (for M and N, respectively):

$$p = \frac{2 \times 165 + 562}{2 \times 1066} = \frac{892}{2132} = 0.4184$$

$$q = 1 - 0.4184 = 0.5816$$

PGC
PHILIPPINE GENOME CENTER

# Testing for deviations from HWE expectations:

Chi-square test: goodness of fit between observed & expected frequencies

| Genotype | Observed | Expected | |
|---|---|---|---|
| MM | 165 | 186.61 | |
| MN | 562 | 518.80 | |
| NN | 339 | 360.58 | |
| Total | 1066 | | |

1. Calculate allele frequencies p and q (for M and N, respectively):
2. Calculate expected number of individuals per genotype:

$MM = 0.4184^2 \times 1066 = 186.61$

$MN = 2(0.4184)(0.5816) \times 1066 = 518.80$

$NN = 0.5816^2 \times 1066 = 360.58$

PGC
PHILIPPINE GENOME CENTER

# Testing for deviations from HWE expectations:

Chi-square test: goodness of fit between observed & expected frequencies

| Genotype | Observed | Expected | |
|---|---|---|---|
| MM | 165 | 186.61 | |
| MN | 562 | 518.80 | |
| NN | 339 | 360.58 | |
| Total | 1066 | | |

1. Calculate allele frequencies p and q (for M and N, respectively):
2. Calculate expected number of individuals per genotype:
3. Calculate $X^2$

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

$$X^2 = \frac{(-21.6)^2}{181.61} + \frac{(43.2)^2}{518.80} + \frac{(-21.6)^2}{360.58} = 7.46$$

PGC
PHILIPPINE GENOME CENTER

# Testing for deviations from HWE expectations:

Chi-square test: goodness of fit between observed & expected frequencies

| Genotype | Observed | Expected | |
|---|---|---|---|
| MM | 165 | 186.61 | |
| MN | 562 | 518.80 | |
| NN | 339 | 360.58 | |
| Total | 1066 | | |

1. Calculate allele frequencies p and q (for M and N, respectively):
2. Calculate expected number of individuals per genotype:
3. Calculate $X^2$
4. Determine probability of $X^2$ from $X^2$ distribution table

Degrees of freedom = 3 -1 -1

Since P < 0.05, there is significant deviation from HWE.

PGC
PHILIPPINE GENOME CENTER

# How do we detect variation?

**MARKER : Object or feature used to identify, distinguish**

| Marker Class | Level of Analysis |
|---|---|
| Morphological | Phenotype |
| Biochemical | Gene Product (Protein) |
| Molecular | DNA sequence |

**MOLECULAR MARKER**

Specific locations in the genome where variation is observed

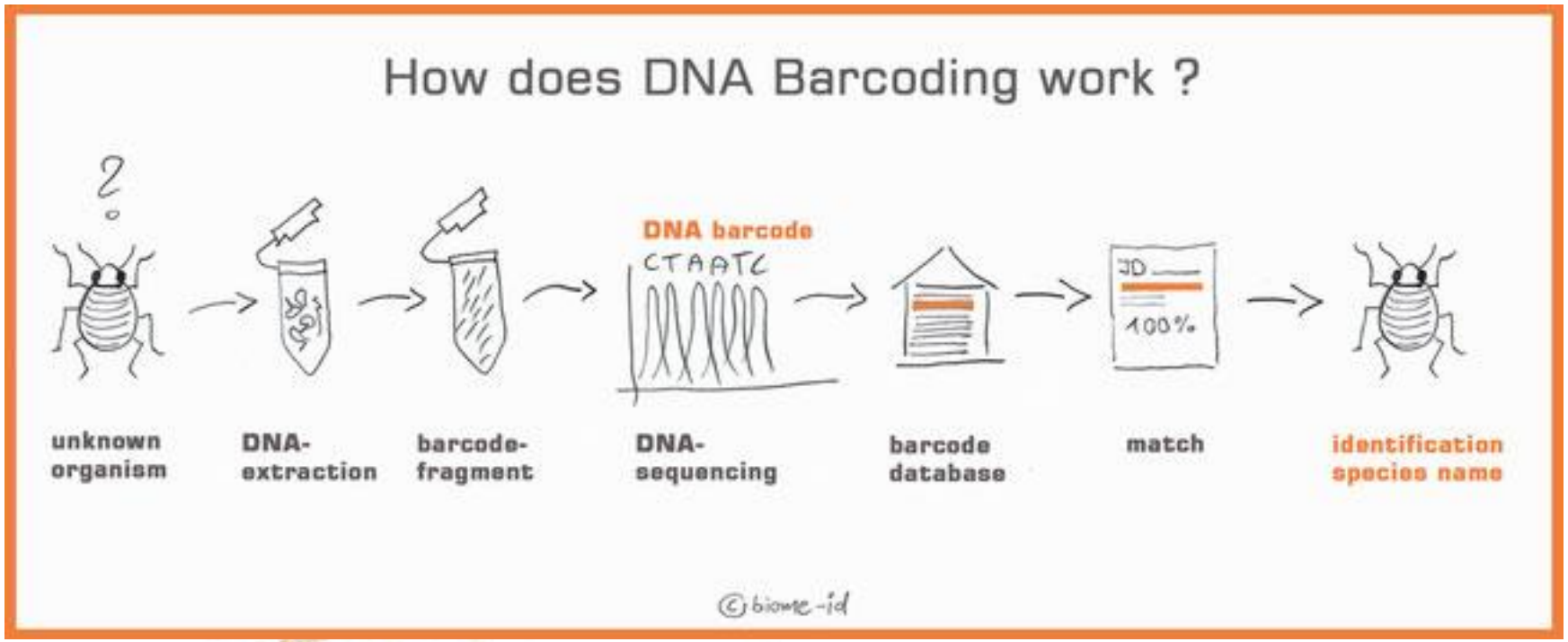
PGC
PHILIPPINE GENOME CENTER

# DNA Markers: Location + Types of variation

1. Location
    - Nuclear DNA
    - Cytoplasmic DNA –mitochondria, chloroplast

PGC
PHILIPPINE GENOME CENTER

# Mitochondrial DNA marker – COI (Cytochrome oxidase I)




PGC
PHILIPPINE GENOME CENTER

# DNA Markers: Location + Types of variation

1. Location
    - Nuclear DNA
    - Cytoplasmic DNA –mitochondria, chloroplast

2. Types of Mutation
    - Point mutations
    - Insertions/Deletions
    - Sequence repeats
    - Inversions
    - Translocations

PGC
PHILIPPINE GENOME CENTER

DNA Markers:

# Types of variation

1. Point mutations – sequence change

2. Insertions/Deletions – sequence AND length change
   - Tandem repeats



PGC
PHILIPPINE GENOME CENTER

DNA Markers:

# Tandem repeats

- 2 or more nucleotides are repeated, adjacent to each other
  - Microsatellites : 2 – 6 bp units
  - Minisatellites: 10 – 100 bp units
- Change in length (length variants = alleles)

```
Allele 1   AGTCTGCACACACACACA--------TGTCAAGCT   (CA)6
Allele 2   AGTCTGCACACACACACACACA----TGTCAAGCT (CA)8
Allele 3   AGTCTGCACACACACACACACACACACATGTCAAGCT (CA)10
```

PGC
PHILIPPINE GENOME CENTER

Table 1 | **Comparison of different molecular markers**

| Marker | Advantages | Disadvantages |
|---|---|---|
| SNPs | • Low mutation rate<br>• High abundance<br>• Easy to type<br>• New analytical approaches are being developed at present<br>• Cross-study comparisons are easy; data repositories already exist | • Substantial rate heterogeneity among sites<br>• Expensive to isolate<br>• Ascertainment bias<br>• Low information content of a single SNP |
| Microsatellites | • Highly informative (large number of alleles, high heterozygosity)<br>• Low ascertainment bias<br>• Easy to isolate | • High mutation rate<br>• Complex mutation behaviour<br>• Not abundant enough<br>• Difficult to automate<br>• Cross-study comparisons require special preparation |
| Allozymes | • Cheap<br>• Universal protocols | • Requirement for fresh or frozen material<br>• Some loci show protein instability<br>• Limited number of available markers<br>• Potentially direct target of selection |
| RAPDs and derivatives | • Cheap<br>• Produces a large number of bands, which can then be further characterized individually (for example, converted into single locus markers) | • Low reproducibility<br>• Mainly dominant<br>• Difficult to analyse<br>• Difficult to automate<br>• Cross-study comparisons are difficult |
| DNA sequencing | • Highest level of resolution possible<br>• Not biased<br>• Cross-study comparisons are easy; data repositories already exist | • Still significantly more expensive than the other techniques |

RAPD, randomly amplified polymorphc DNA; SNP, single nucleotide polymorphism.