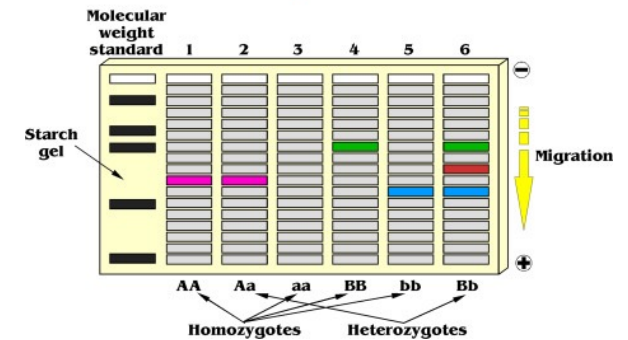
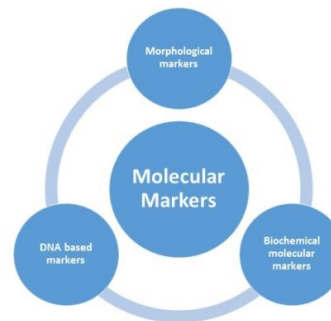
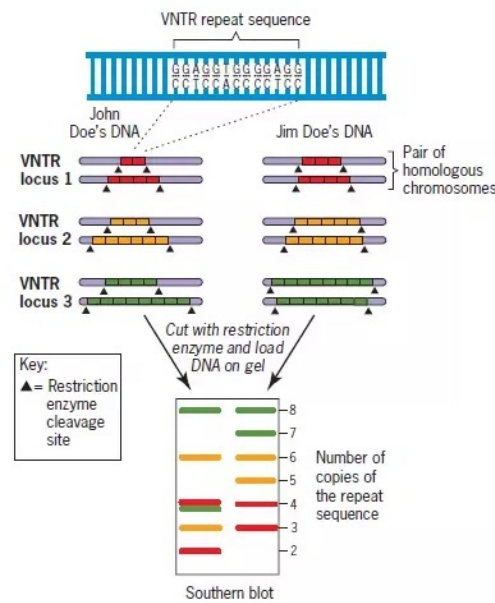


GENÉTICA E MELHORAMENTO DE PLANTAS

Licenciaturas em Eng^a Agronómica

Marcadores moleculares - Prática

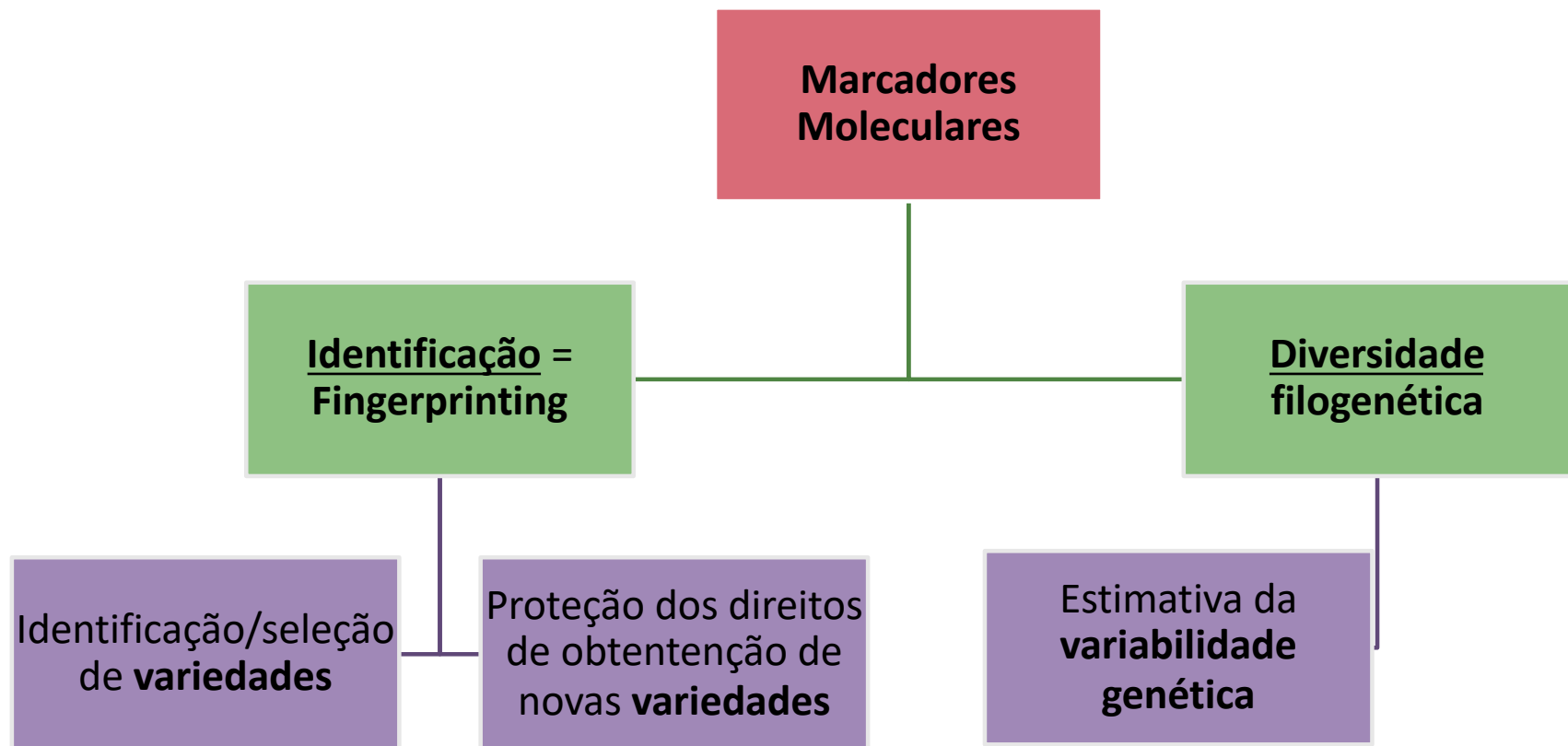


Bibliography:

An Introduction to Genetic Analysis 11th Ed
Genetics: From Genes to Genomes 6th Ed

Manuela Gomes da Silva
manuelasilva@isa.ulisboa.pt

Aplicações dos Marcadores Moleculares



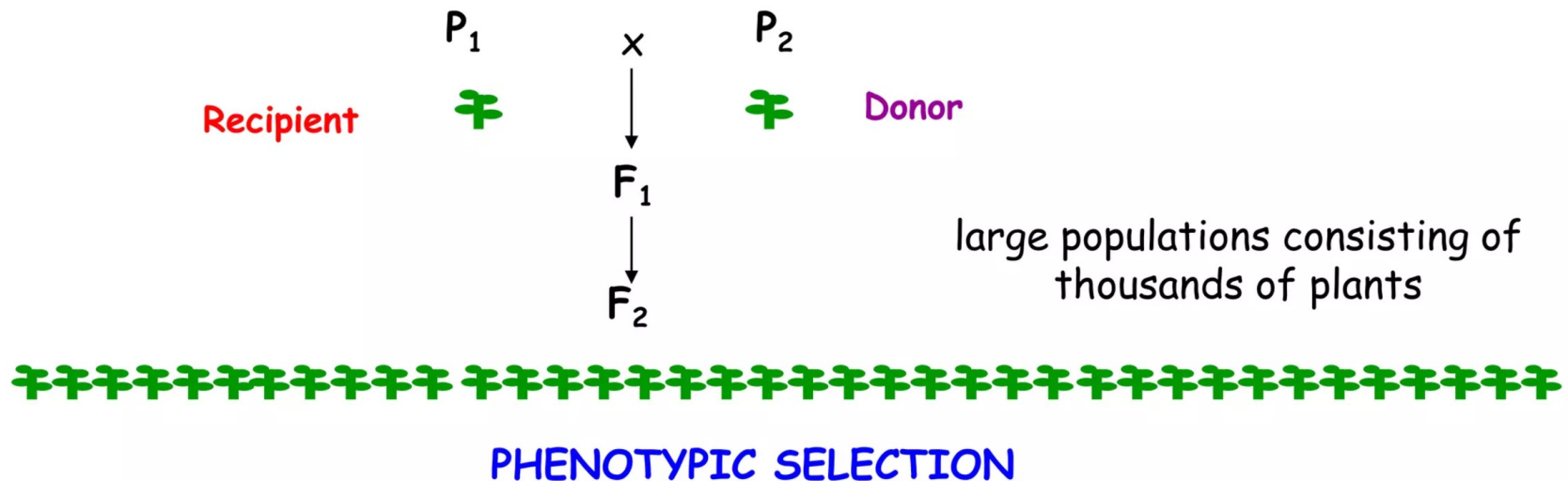
Plant Breeding

Conventional plant breeding strategies

Selection of individual plants in the progeny of a cross based on their **phenotypical characteristics**:

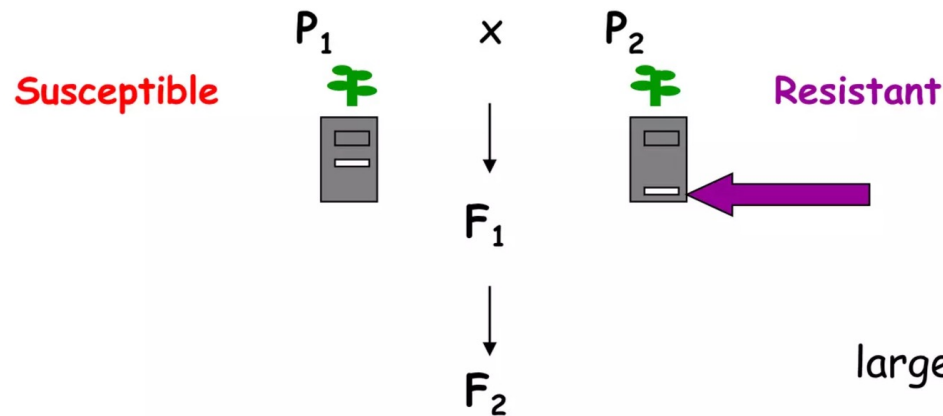
- extremely time consuming
- depends on the developmental

CONVENTIONAL PLANT BREEDING



Marker-Assisted Selection (MAS) - Plant Breeding

MARKER-ASSISTED BREEDING

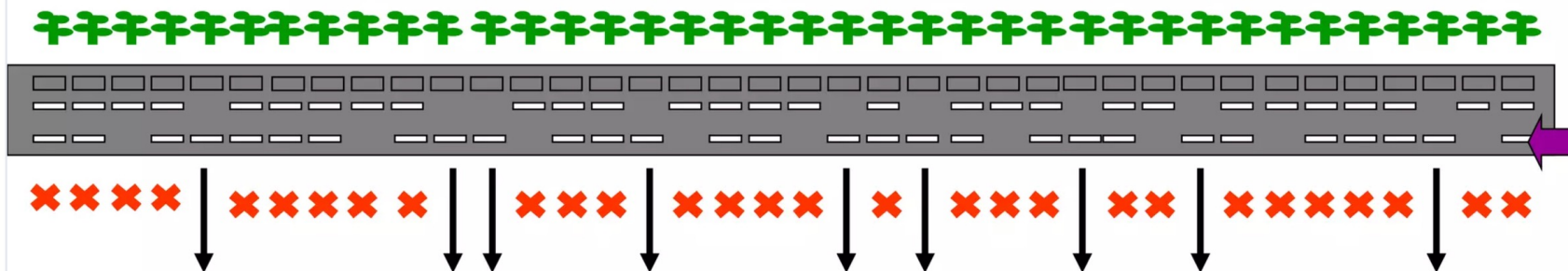


MAS Plant Breeding

Selection of individual plants in the progeny of a cross based on **molecular markers**:

- more rapid
- allows the screening of a higher number of plants

large populations consisting of thousands of plants



MARKER-ASSISTED SELECTION (MAS)

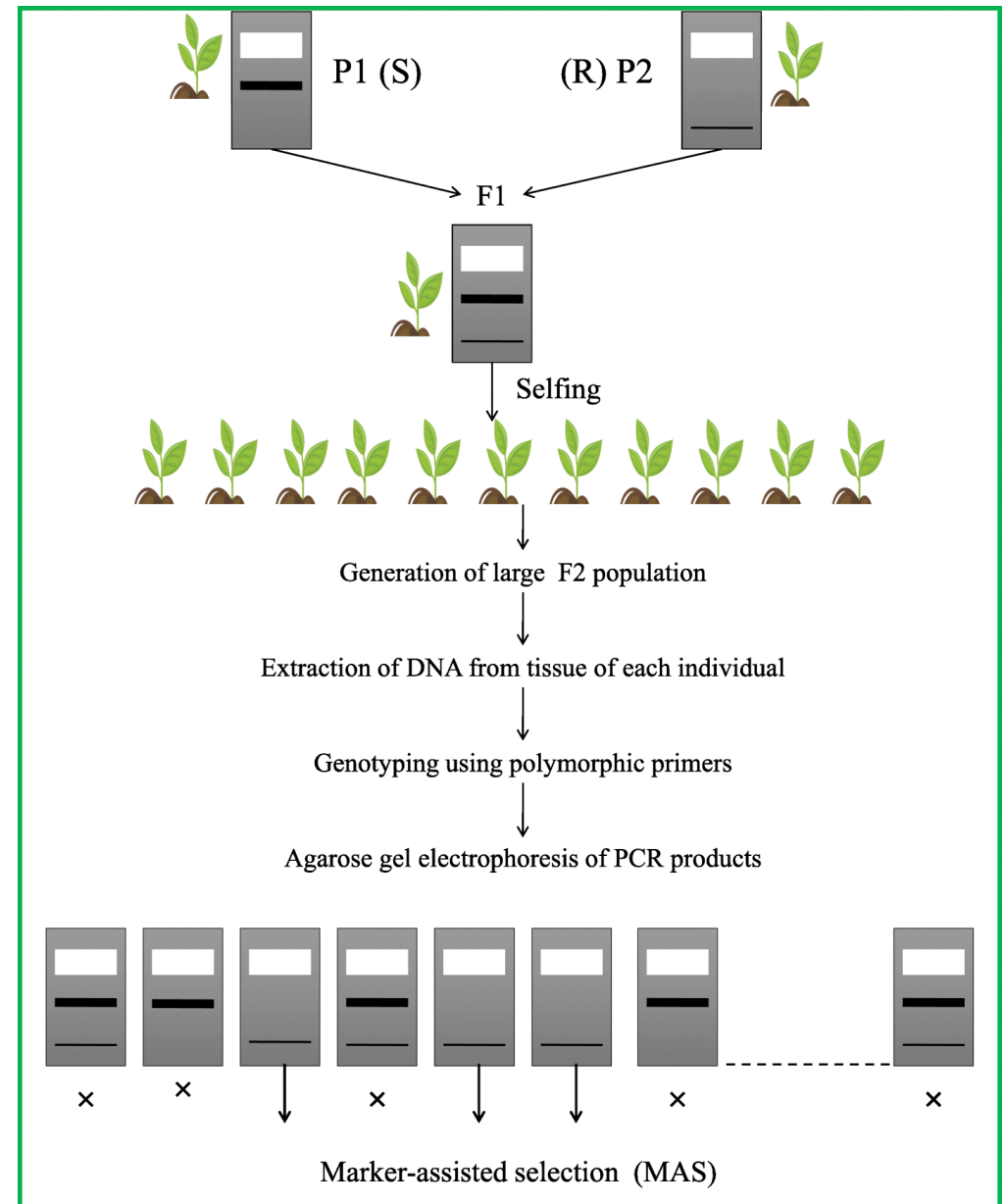
Marker-Assisted Selection Plant Breeding

Basic procedure of molecular
marker-assisted selection

Ex:

Use of two polymorphic markers flanking
the genomic regions of interest

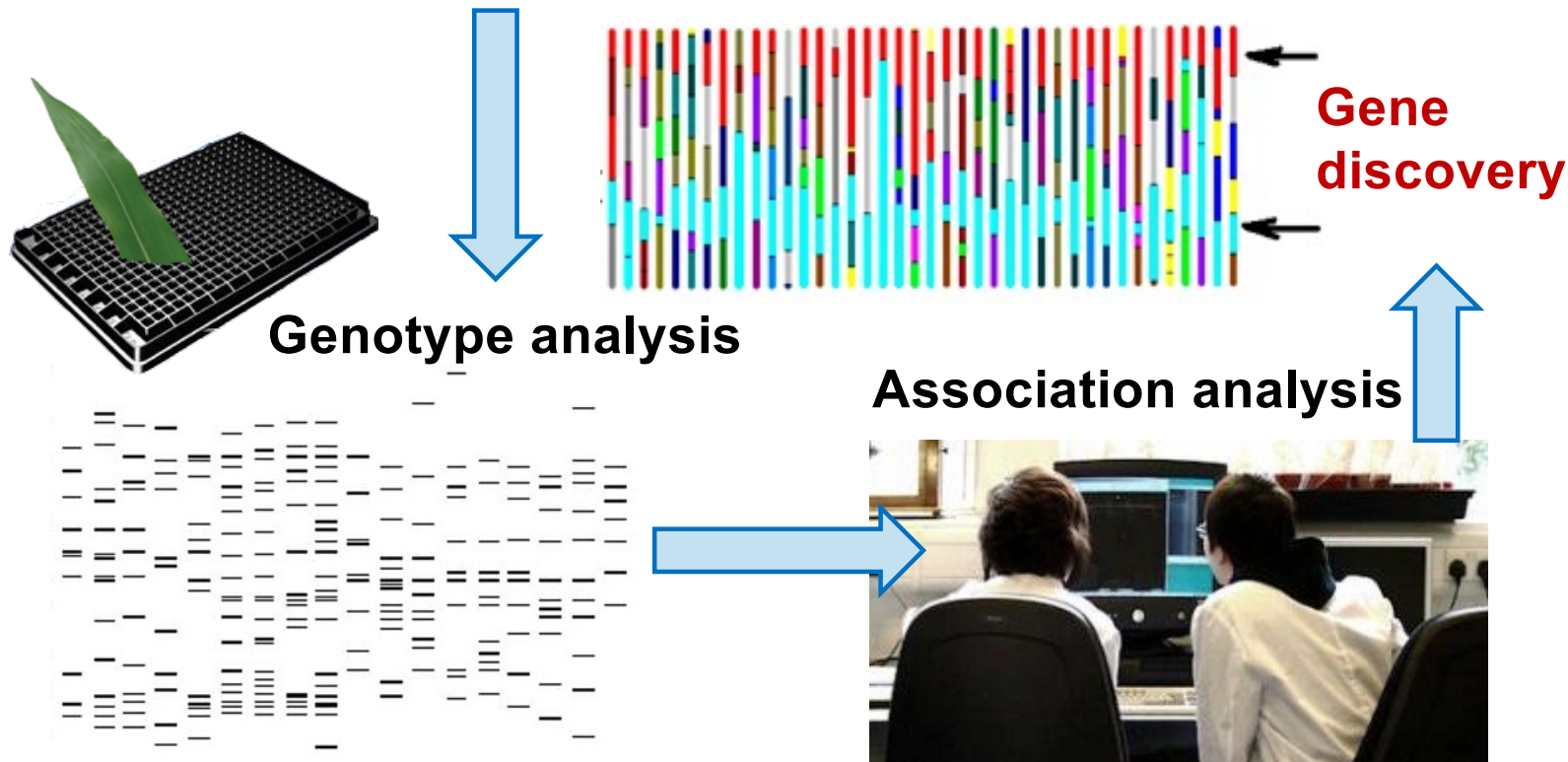
MAS is useful to characteristics controlled
by few genes in a chromosome region





Genome Wide Association Studies

→ identify **genes** associated with complex **traits**
(Ex: yield, water use efficiency)



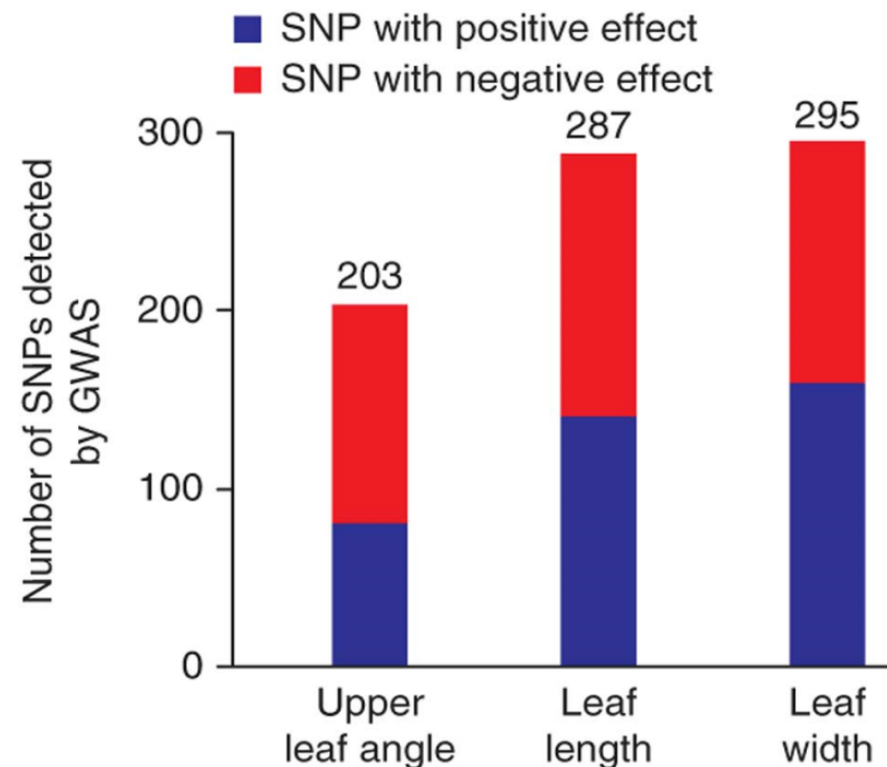
GWAS – Genome Wide Association Studies

This approach allows hundreds of genes with small effects to be identified

In **maize**, **grain yields** are correlated with leaf size and angle

GWAS revealed hundreds of single-nucleotide polymorphisms (**SNPs**) associated with these traits

→ valuable information for breeders

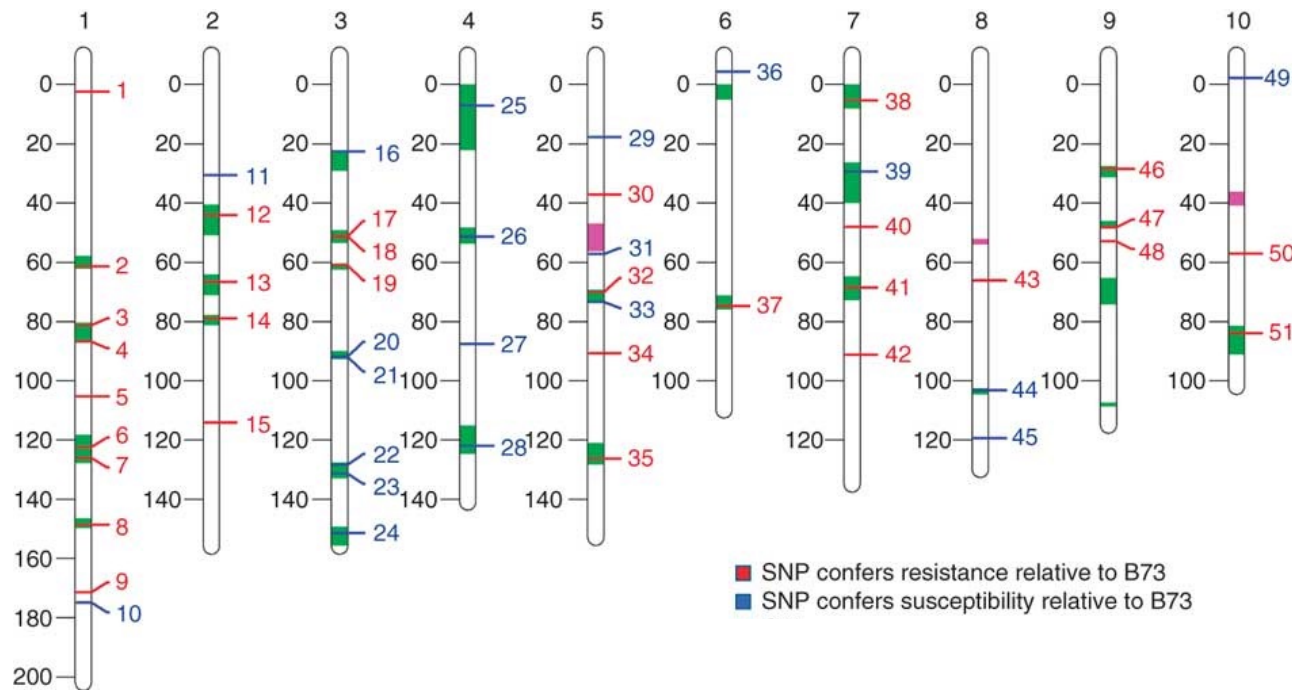


Reprinted by permission from Macmillan Publishers Ltd. Tian, F., Bradbury, P.J., Brown, P.J., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T.R., McMullen, M.D., Holland, J.B., and Buckler, E.S. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43: [159-162](#).

© 2013 American Society of Plant Biologists

GWAS – Genome Wide Association Studies

GWAS reveals **SNPs** that contribute to biotic and abiotic tolerance



GWAS studies allowed the **identification of genes** contributing to other **agronomic traits** including **drought tolerance** (in relation to variety B73)

Reprinted by permission from Macmillan Publishers Ltd Kump, K.L., Bradbury, P.J., Wissner, R.J., Buckler, E.S., Belcher, A.R., Oropeza-Rosas, M.A., Zwonitzer, J.C., Kresovich, S., McMullen, M.D., Ware, D., Balint-Kurti, P.J., and Holland, J.B. (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43: [163-168](#).

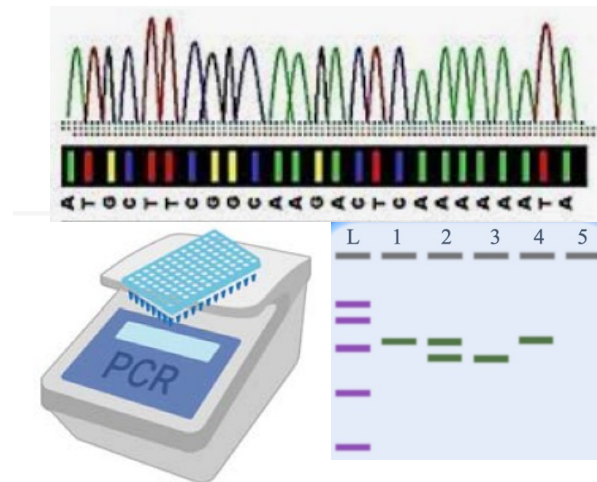
SNP - Single Nucleotide Polymorphism

Single nucleotide polymorphism (SNP)

- variation at a single position in a DNA sequence – **1 bp**
 - SNP within a gene → gene with more than one allele
→ potential variations in the amino acid sequence
 - SNPs in noncoding regions of DNA

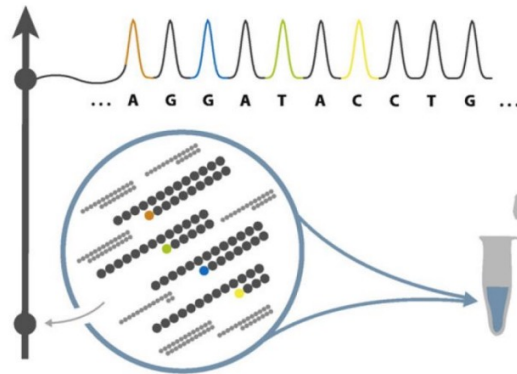
Detected by:

- Sequencing
- PCR

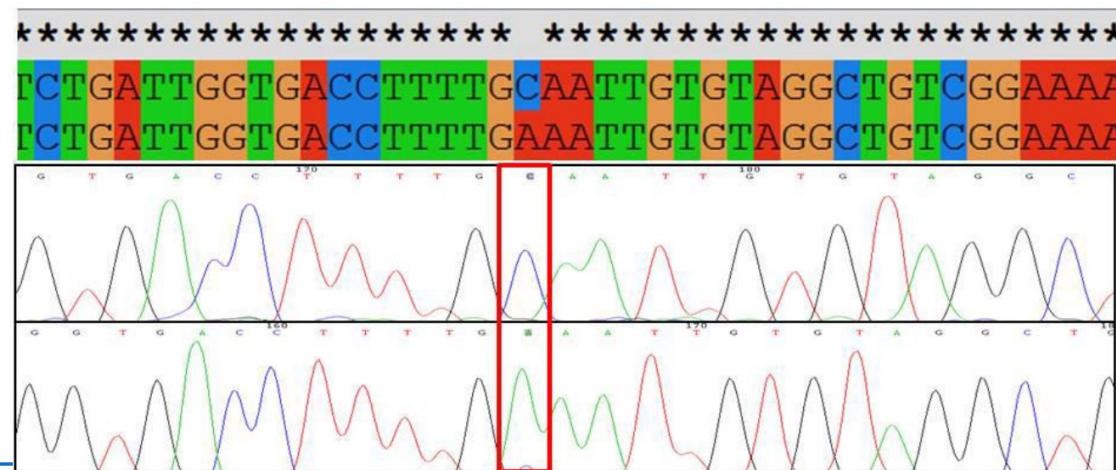


SNP - Single-nucleotide polymorphism

Sequenciação



Alinhamento de
sequências



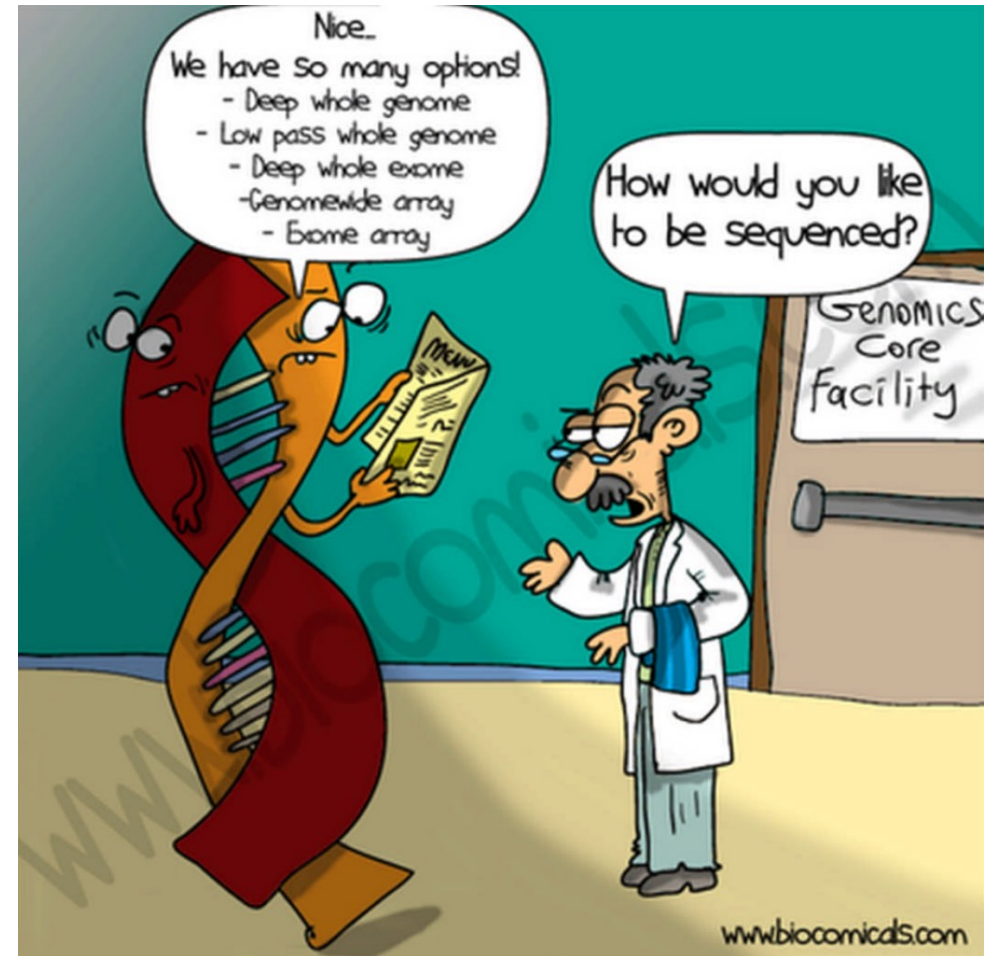
DNA sequencing

Sequencing

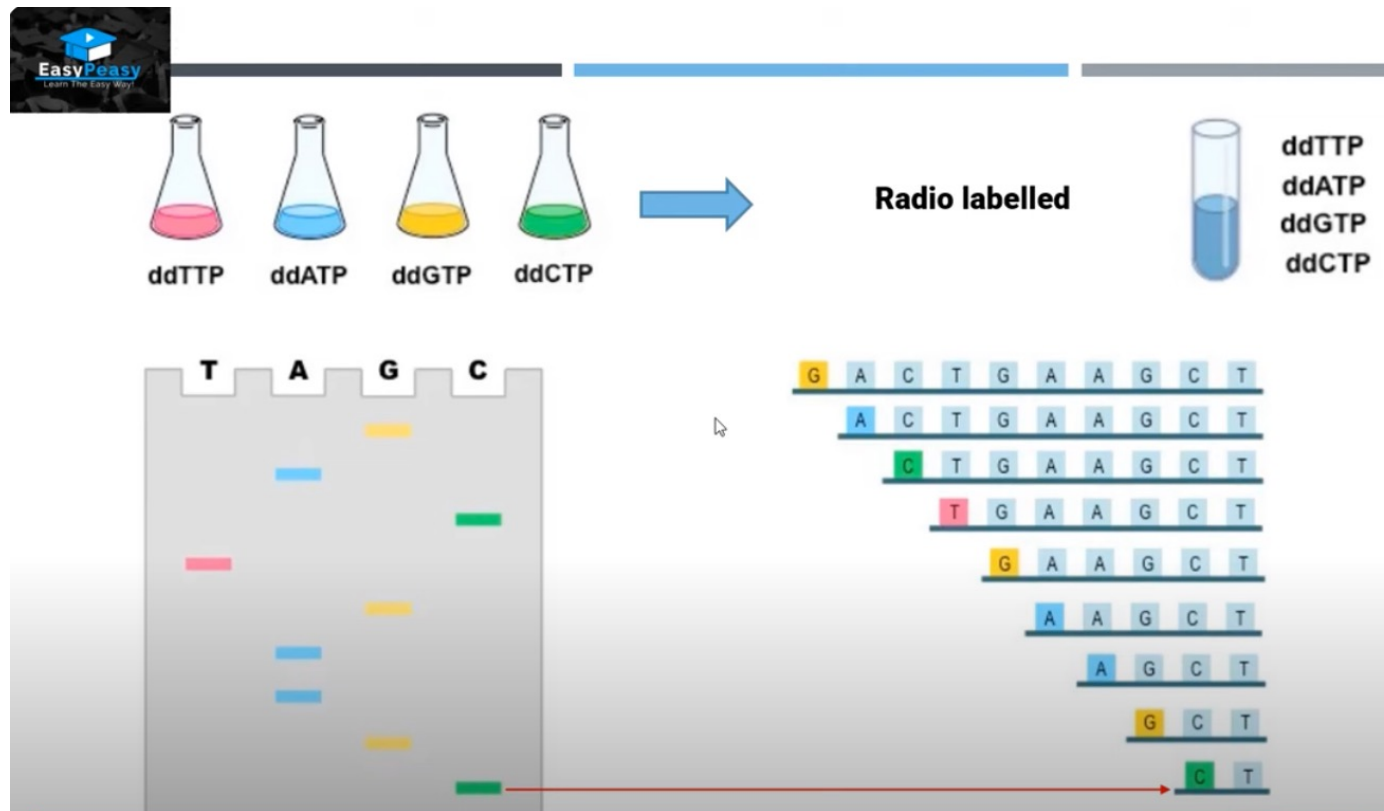
→ **order** of nucleotides in the DNA molecules

Nucleotides order determines the aminoacids order and therefore the protein structure and function

Changes in the DNA sequence can lead to changes in the protein

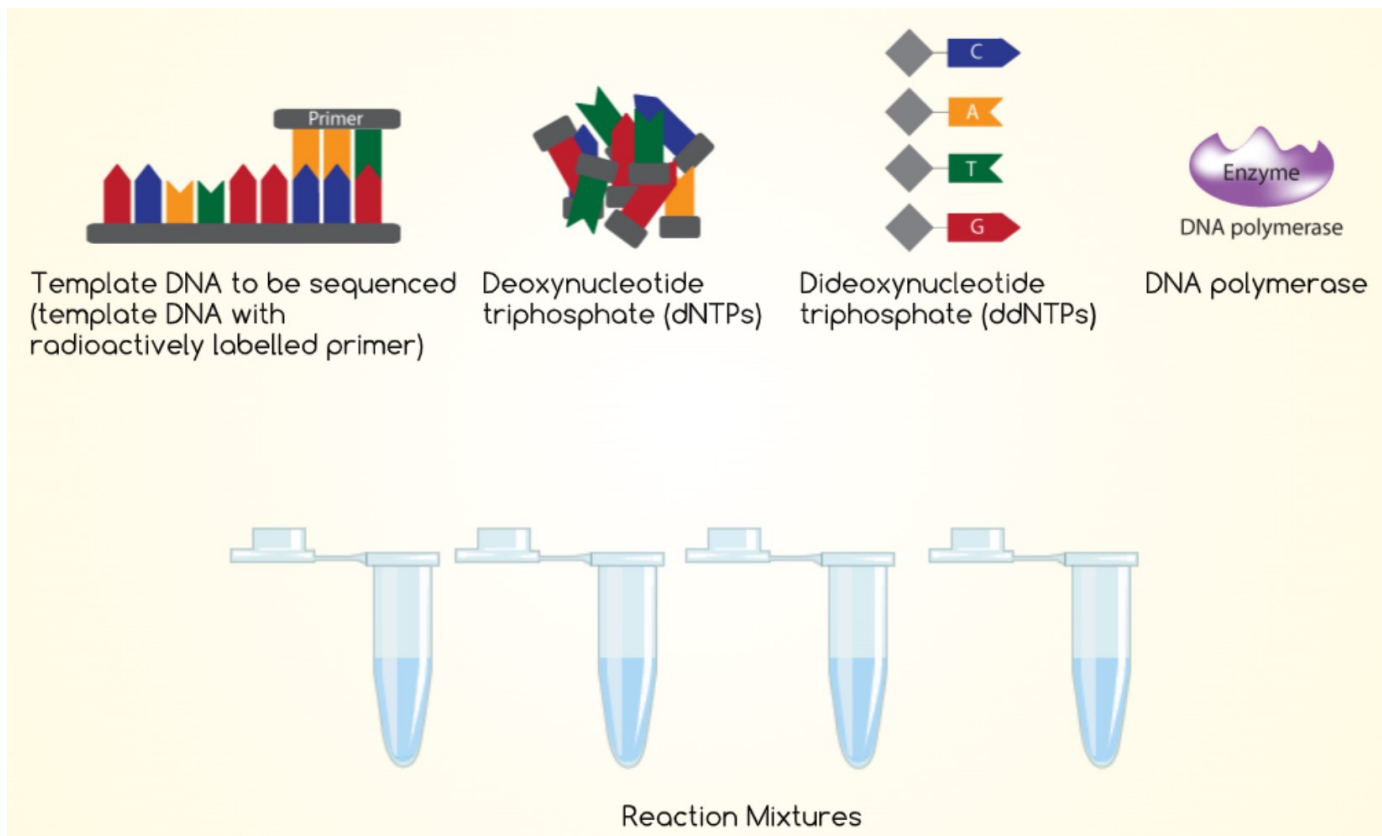


DNA sequencing



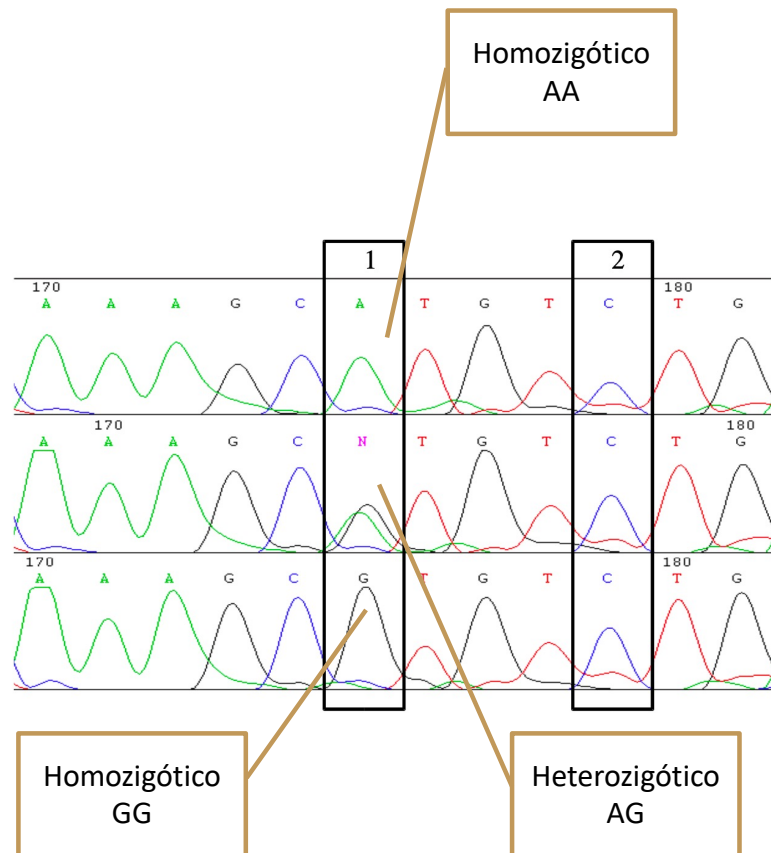
<https://www.youtube.com/watch?v=l0JVVPt4vNw>
5'40''

DNA sequencing activity



<https://tinyurl.com/DNASequencingGMP>

SNP - Single Nucleotide Polymorphism



Podem surgir de

- erro raro na replicação
- devido a mutação química

Podem ocorrer em

- genes → diferente **fenótipo**
- zonas não codificantes → **silenciosos**

Comparação entre espécies

Human 1	TTGACG	TATAAATGATCTTTATAT	T	TTTCAGAAGTC	
Human 2	TTGACG	TATAAATGATCTTTATAT	C	TTTCAGAAGTC	
Chimp	TTGAC	A	TATAAATGATCTTTATAT	C	TTTCAGAAGTC

C é o alelo original pois é partilhado entre humanos e chimpanzés

SNP - Single-nucleotide polymorphism

Gene A from person 1	Codon results in particular Amino Acid (AA) sequence	GCA	AGA	GAT	AAT	TGT
		Ala	Arg	Asp	Asn	Cys
Gene A from person 2	Codon change makes no difference in AA (redundant code)	GCG	AGA	GAT	AAT	TGT
		Ala	Arg	Asp	Asn	Cys
Gene A from person 3	Codon change results in different AA sequence	GCA	AAA	GAT	AAT	TGT
		Ala	Lys	Asp	Asn	Cys

Ex. Sickle Cell Anemia, Cystic Fibrosis

SNPs and Haplotypes

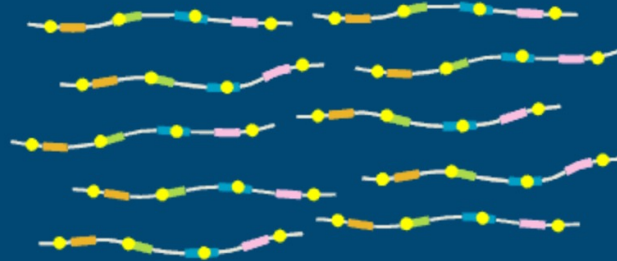
Making SNPs Make Sense

What Is a SNP?

What Is a Haplotype?

Applying SNP Profiles to Drug Choices

Single Nucleotide Polymorphisms (SNPs)



Scientists have found that many of these differences are single-nucleotide substitutions in the DNA sequence. These are referred to as Single Nucleotide Polymorphisms, or SNPs. What do these SNPs look like at the DNA level?



<https://tinyurl.com/SNPsHaplotype>

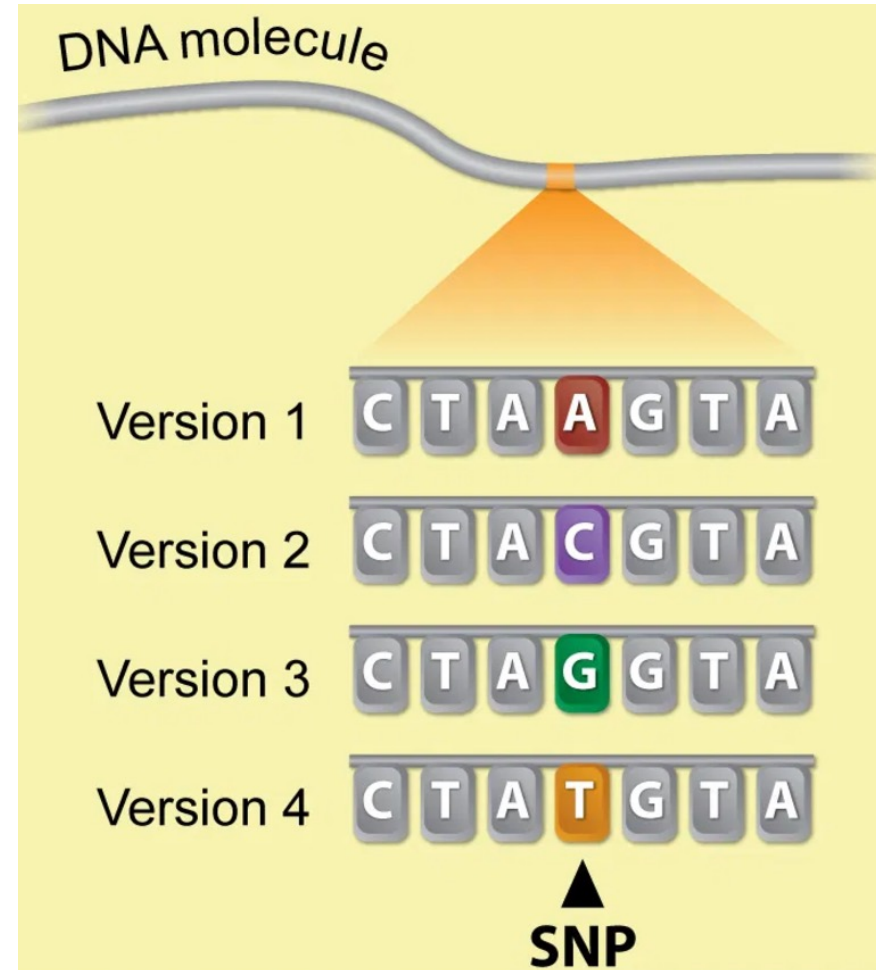
SNP quick Reference

SNP - single-nucleotide substitutions

Each SNP have **4 possible versions**:
1 for each nucleotide, A, C, G, and T

At least **2 versions**
each present in at least **1% of the population**

(human genome: 1 SNP every ~300 bp)



Human Genetics

SNPs and disease-causing mutations: Not the same!

SNPs and **disease-causing mutations**: both **point mutations**

- SNP - at least **1%** of the population
- disease-causing mutations are not this common

SNPs are not necessarily located within genes

- **Linked/associated SNPs** (indicative SNPs)
associated with a trait but not within genes controlling it

- **Causative SNPs**

affect the way a protein functions:

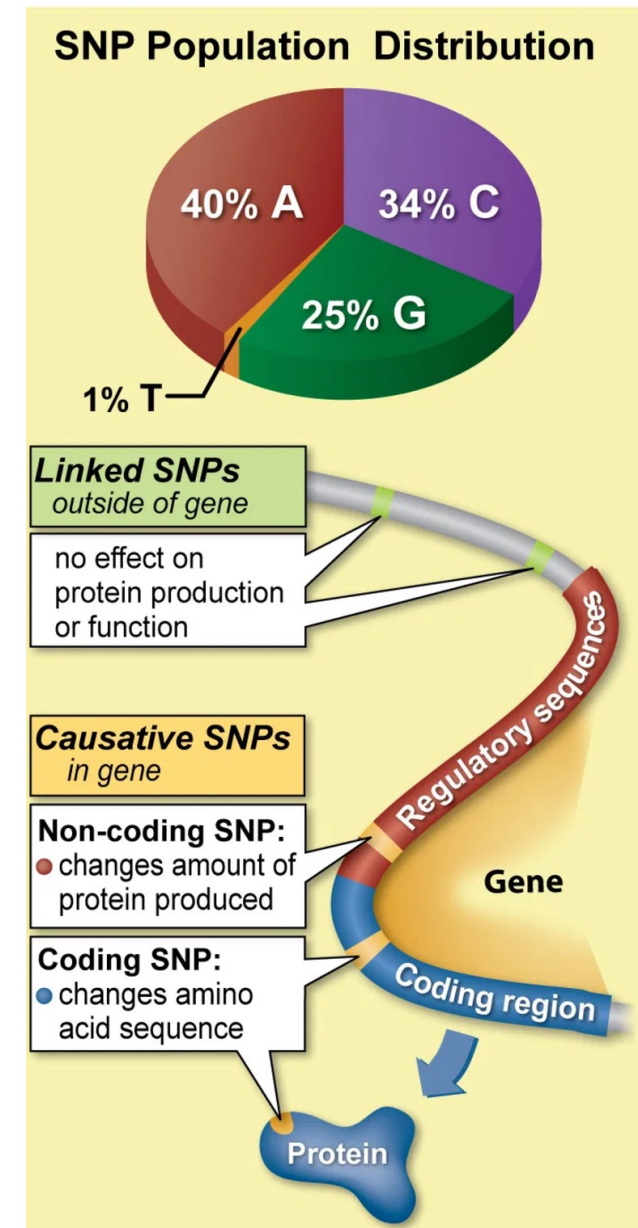
Coding SNPs - coding region of a gene

→ change the amino acid sequence of a protein

Non-coding SNPs - regulatory sequences

→ change timing, location, or level of gene expression

<https://learn.genetics.utah.edu/content/precision/snips/>



SNP association with a certain trait

SNP completely associated with a trait if:

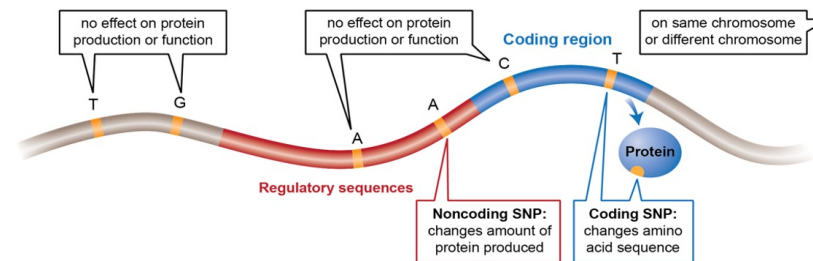
- all plants with the same trait share the **same alleles** at a certain position, and
- all plants with contrasting trait (ex: seed color) share **different alleles** at that position

SNP completely unassociated with a trait if:

- its alleles occur with **equal frequency** in plants with contrasting traits
- unlikely to be located within or near the gene responsible for that trait

SNP association with a certain trait

Genome Wide Association Studies (**GWAS**) based on **genetic linkage**



- Two DNA sequences / *loci* on the same chromosome are said to be “**linked**” if they tend to be **inherited together** during Meiosis
- If meiotic **crossing-over** occurs between the two sequences, they will **not be inherited together**
- The **chance of crossing-over** occurring between the two sequences depends on the **physical distance** between them

< distance → < **crossing over** probability → > **linked**

GWAS – Genome Wide Association Studies

O que é um **GWAS**?

→ Abordagem que envolve o screening de marcadores em genomas de **muitos indivíduos** para encontrar **variações genéticas (SNPs)**

→ Estes estudos podem ser particularmente úteis para encontrar variações genéticas que contribuam para **características importantes** das culturas

Identificar genes associados a um fenótipo

Pesquisar os genomas de **muitos indivíduos** para encontrar **variações no DNA** associadas a **características / fenótipos** particulares → **SNP associados a um fenótipo**

- Sequenciação do genoma de muitos indivíduos

→ identificação de **milhões de variações** comuns entre esses genomas

→ **Algumas** destas variações podem ser associadas a um fenótipo que varia entre os indivíduos

GWAS in Plants

GWAS design is usually **case vs control**

case - variety with the desired trait

control - variety without traits

Comparison of DNA from varieties/plants with contrasting phenotypes for a particular trait

Identification of SNPs by GWAS Analysis

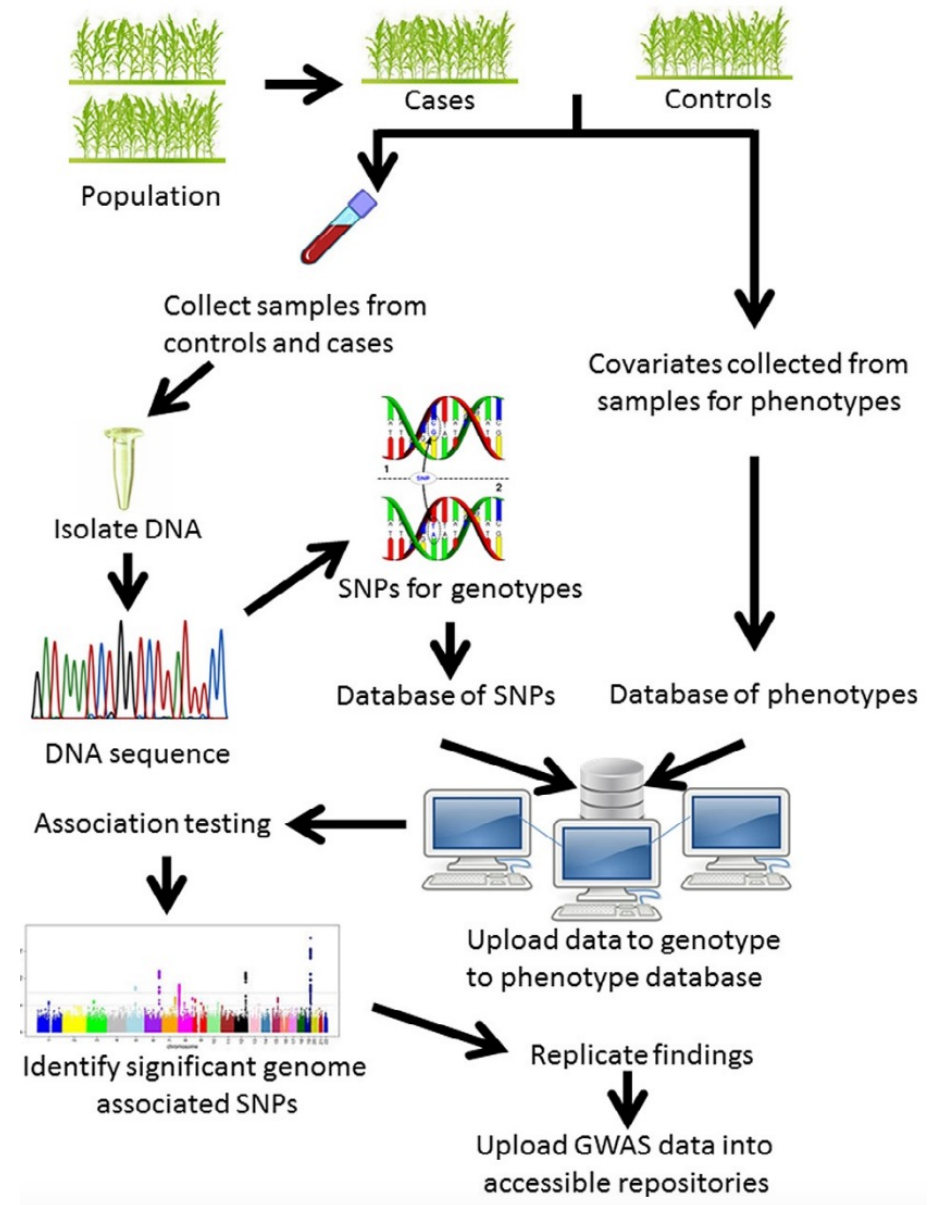
More than million

SNPs can be captured using genotyping

Correlation between a trait and SNP

in a population

→ Identification of **significantly associated SNPs**



Domesticação e raças de cães



Canine domestication began some
15 000 years ago from wolves

Actual breeds have no **more than**
200 years and were **selected from**
a unique population

GWAS in dog

Reference dog genome

- data from just one boxer dog (Tasha)

Other dogs from **other breeds** were sequenced and looking for **differences** compared to Tasha's genome

→ **Public database SNPs**

https://www.ncbi.nlm.nih.gov/assembly/GCF_000002285.5

Dog10K_Boxer_Tasha

Organism name: [Canis lupus familiaris \(dog\)](#)

Intraspecific name: Breed: boxer

Isolate: Tasha

Sex: female

BioSample: [SAMN02953603](#)

BioProject: [PRJNA13179](#)

Submitter: Dog Genome Sequencing Consortium

Date: 2020/10/06

Synonyms: canFam6

Assembly level: Chromosome

Genome representation: full

RefSeq category: representative genome

GenBank assembly accession: GCA_000002285.4 (latest)

RefSeq assembly accession: GCF_000002285.5 (latest)

RefSeq assembly and GenBank assembly identical: no ([hide details](#))

- Different: chromosome MT.
- Data displayed for RefSeq version

WGS Project: [AAEX04](#)

Assembly method: Canu v. 1.8

Expected final version: yes

Genome coverage: 100.0x

Sequencing technology: PacBio Sequel



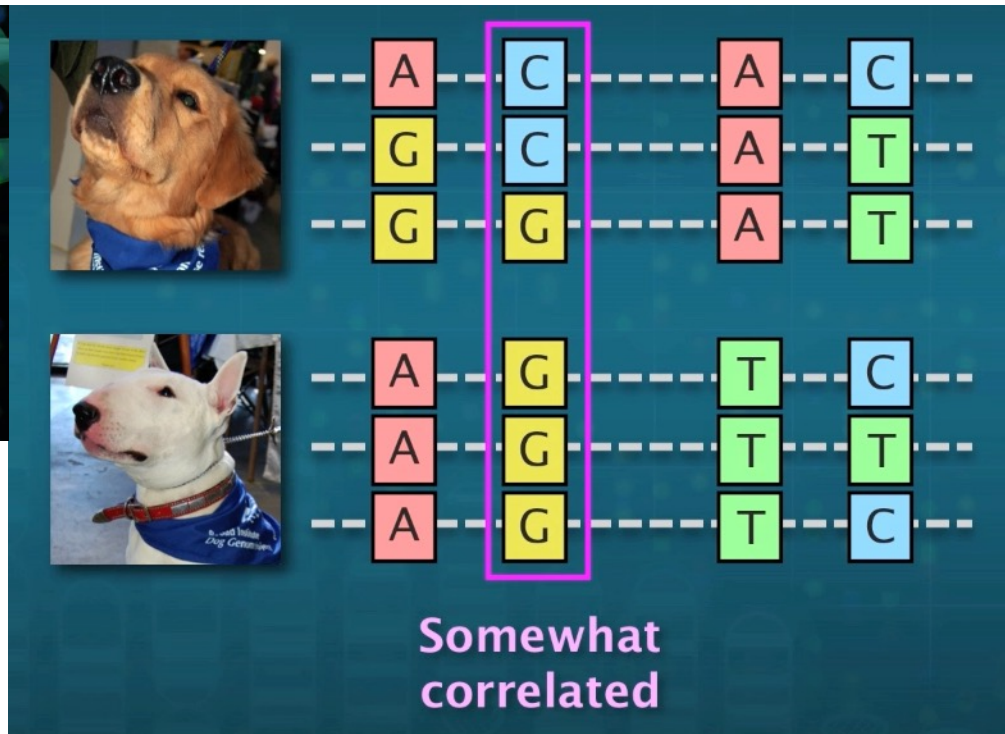
National Human Genome
Research Institute

Dog Genomics and Dogs as Model Organisms



Elinor Karlsson

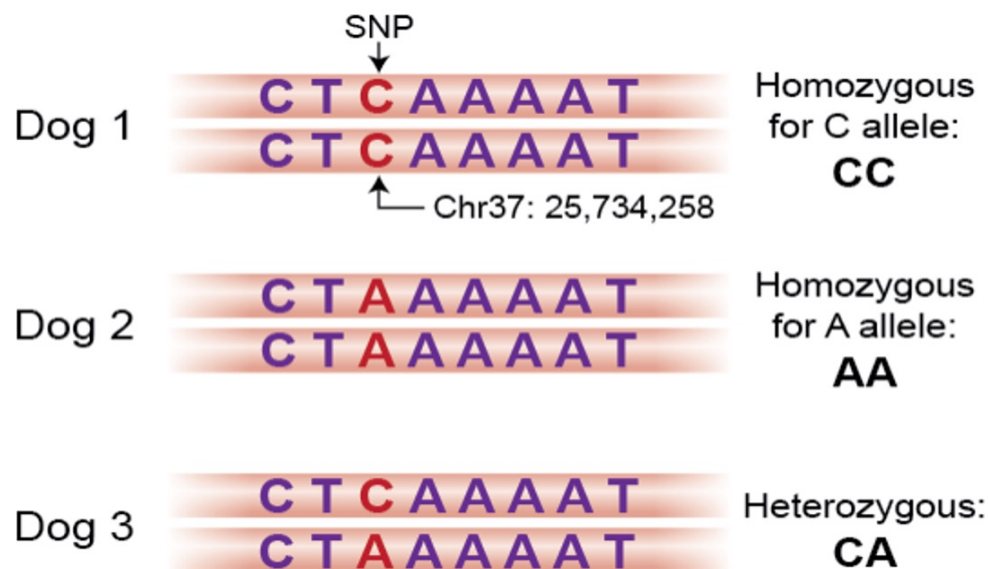
Director of the Vertebrate
Genomics Group at the Broad
Institute of MIT and Harvard



<https://media.hhmi.org/hl/13Discussion1.html>

5'27" - 12'30"

Como identificar SNPs?



GWAS usa associações entre um **SNP** específico e uma característica de interesse

Sequências de DNA do cromossoma 37 em três **cães**

Duas sequências para cada cão
- uma de **cada** cromossoma **parental**
(as cadeias complementares não são apresentadas)

As letras **vermelhas** e a seta indicam um **SNP** comum na posição nucleotídica 25 734 258 do cromossoma 37 determinado pela contagem de uma extremidade à outra do cromossoma

A maioria dos cães tem um **C** ou um **A** nesta posição

As duas versões do SNP são chamadas de **alelo C** e **alelo A**

GWAS aplicado à cor do pêlo dos cães

Alelos em **17 loci** diferentes em cães com pêlo **preto** e cães com pêlo **branco**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	CC	AT	CC	GG	AA	TC	TT	CC	GG	AA	TT	GT	AG	AA	CC	GG	AT
	CC	AT	AC	GG	GG	TT	TT	CC	GG	AG	TT	GG	AG	AG	CC	GG	AT
	CC	AA	AC	CG	GG	TT	TT	CT	GG	AA	TT	GT	AG	AG	CC	TT	AT
	CC	AA	AC	GG	AG	TT	TT	CT	GG	GG	TT	GG	AG	AG	CC	GT	AT
	CC	AT	CC	CG	AA	TT	AA	CT	GG	AA	TT	GT	AA	AA	CC	GT	AT
	CC	AT	CC	GG	AA	TT	AA	TC	GG	AA	TT	GG	AG	AA	CC	GG	TT
	CC	AT	CC	GG	AA	TT	AA	CC	GG	AA	TT	GG	AA	AA	CC	GT	AT
	CC	AA	CC	GG	AG	TT	AA	TC	GG	GG	TT	GG	AA	AA	CC	TT	AT

Se um **SNP** é encontrado com muito maior frequência em cães com pelo branco do que em cães com pelo preto, o SNP está **associado à cor branca** do pelo

SNP/cor do pêlo dos cães - Análise dos dados

1. Qual o SNP que está **completamente associado** à cor do pelo?
2. Quais os SNPs que estão **completamente não associados** com a cor da pele?
3. Qual o SNP que tem a **próxima associação mais forte** com a cor da pele, depois do SNP completamente associado identificado antes?

Actividade em grupos

Cada grupo faz **download** do Fénix do ficheiro '**SNP Exercises.xlsx**' para realizar os exercícios

Análise dos dados

Como avaliar quais os **SNPs** têm a associação mais forte com uma característica?

1. Contar o número de vezes que cada alelo aparece no SNP

Se por exemplo um cão tem os alelos TT para esse SNP, contar o alelo T duas vezes
se tiver os alelos TC, contar o alelo T uma vez e o alelo C uma vez

2. Registrar os números de cada alelo em todos os cães com cada fenótipo na tabela do SNP

3. Calcular a diferença nos números de cada alelo entre os dois fenótipos

4. Somar as diferenças para todos os alelos

Diferença total maior → **associação mais forte entre esse SNP** e a **característica**

SNP associated with coat length

Alelos em **7 loci** diferentes em cães com pêlo curto e cães com pêlo longo

Dog	Coat Length	chr32 7420804	chr32 7472206	chr32 7473337	chr32 7479580	chr32 7482867	chr32 7490570	chr32 7492364
	Short Coat	TC	AA	GT	TT	AG	TT	CG
	Short Coat	TC	AA	GG	TT	GG	CC	GG
	Short Coat	CC	GA	GT	CT	AG	CT	GG
	Short Coat	TC	GA	GG	CT	AA	TT	CG
	Short Coat	CC	GA	GT	CT	AG	CT	GG
	Short Coat	TC	AA	GG	TT	GG	CC	CG

Dog	Coat Length	chr32 7420804	chr32 7472206	chr32 7473337	chr32 7479580	chr32 7482867	chr32 7490570	chr32 7492364
	Long Coat	CC	AA	TT	TT	GG	TT	GG
	Long Coat	TC	AA	TT	TT	GG	TT	GG
	Long Coat	TC	AA	TT	TT	GG	TT	GG
	Long Coat	CC	AA	TT	TT	GG	CT	CG
	Long Coat	TT	AA	TT	TT	GG	TT	GG
	Long Coat	CC	AA	TT	TT	GG	TT	GG

Cada grupo faz **download** do Fénix do ficheiro '**SNP Exercises.xlsx**' para realizar os exercícios

Análise dos dados – Tamanho da pelagem

Uma **diferença total maior** indica uma **associação mais forte** entre esse SNP e a característica

1. Qual destes SNP tem a **associação mais forte** com o comprimento do pêlo?
2. Qual destes SNP tem a **associação mais fraca** com o comprimento do pêlo?
3. Com base nestes sete SNPs, onde procuraria se quisesse encontrar um **gene** envolvido no comprimento da pelagem do cão?

Análise dos dados – Tamanho da pelagem

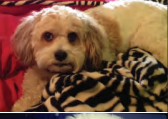
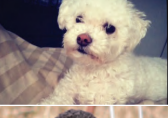
Uma **diferença total maior** indica uma **associação mais forte** entre esse SNP e a característica

1. Qual destes SNP tem a **associação mais forte** com o comprimento do pêlo?
2. Qual destes SNP tem a **associação mais fraca** com o comprimento do pêlo?
3. Com base nestes sete SNPs, onde procuraria se quisesse encontrar um **gene** envolvido no comprimento da pelagem do cão?

chr32 7420804			
Allele	Short Coat	Long Coat	Difference
T			
C			
		Total N. of differences	
chr32 7472206			
Allele	Short Coat	Long Coat	Difference
A			
G			
		Total N. of differences	
chr32 7473337			
Allele	Short Coat	Long Coat	Difference
G			
T			
		Total N. of differences	
chr32 7479580			
Allele	Short Coat	Long Coat	Difference
T			
C			
		Total N. of differences	
chr32 7482867			
Allele	Short Coat	Long Coat	Difference
A			
G			
		Total N. of differences	
chr32 7482870			
Allele	Short Coat	Long Coat	Difference
T			
C			
		Total N. of differences	
chr32 7492364			
Allele	Short Coat	Long Coat	Difference
C			
G			
		Total N. of differences	

SNP associated with coat texture

Alelos em 6 *loci* diferentes em cães com pêlo encaracolado e cães com pêlo liso

Dog	Hair Type	chr27 5525002	chr27 5541113	chr27 5542806	chr27 5545082	chr27 5557298	chr27 5568947	Dog	Hair Type	chr27 5525002	chr27 5541113	chr27 5542806	chr27 5545082	chr27 5557298	chr27 5568947
	Curly	CC	CT	CC	AA	CT	CC		Straight	TC	CT	CC	GA	TT	TC
	Curly	TC	CT	TC	GA	CT	TC		Straight	TT	TT	CC	AA	TT	TC
	Curly	TT	CT	TC	GA	CT	TT		Straight	CC	CC	CC	GA	CT	CC
	Curly	TT	CC	TT	GG	CT	TC		Straight	TC	CT	CC	AA	CT	TT
	Curly	TC	CC	TT	GG	TT	TC		Straight	TC	TT	CC	AA	TT	CC

Cada grupo faz **download** do Fénix do ficheiro '[SNP Exercises.xlsx](#)' para realizar os exercícios

Análise dos dados – Tamanho da pelagem

1. Qual destes SNP tem a **associação mais forte** com a textura da pelagem?
2. Com base nestes seis SNP, onde procuraria se quisesse encontrar um **gene** envolvido na textura da pelagem do cão?

chr27 5525002			
Allele	Curly Coat	Straight Coat	Difference
C			
T			
Total number of differences			
chr27 5541113			
Allele	curly Coat	Straight Coat	Difference
C			
T			
Total number of differences			
chr27 5542806			
Allele	curly Coat	Straight Coat	Difference
C			
T			
Total number of differences			
chr27 5545082			
Allele	curly Coat	Straight Coat	Difference
A			
G			
Total number of differences			
chr27 5557298			
Allele	curly Coat	Straight Coat	Difference
C			
T			
Total number of differences			
chr27 5568947			
Allele	curly Coat	Straight Coat	Difference
C			
T			
Total number of differences			