

Selección Genómica

Luis Garreta

Electiva de Bioinformática
MAESTRÍA EN INFORMÁTICA BIOMÉDICA
Universidad del Bosque
Bogotá-Colombia

September 25, 2021

Selección Genómica

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

- La selección genómica se basa en los avances en el aprendizaje automático y la disponibilidad de datos de genotipificación para predecir rasgos fenotípicos agronómicamente relevantes.
- La selección altamente eficiente y precisa de genotipos conduce a un acortamiento dramático del ciclo de reproducción tanto en cultivos como en animales de interés agroalimenticio y agroindustrial.

Agrigenomics

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

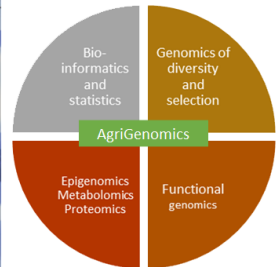
GS

Data

Training

Model

Quizás la mayor revolución en la agricultura en la última década es el surgimiento de la agrigenómica para mejorar los programas tradicionales de mejoramiento.



ADN, Cromosómas, y Ploidía

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

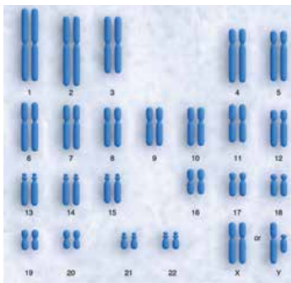
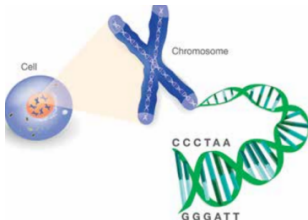
GS

Data

Training

Model

- Humanos son seres diploides: Tenemos 23 pares de cromosomas, 1 heredado del padre y el otro de la madre.
- Otros seres son poliploides: algunos animales y muchas plantas hereden más de dos cromosomas
 - La caña es hexaploide
 - Variedades de papa diploide y otras tetraploides
 - Algunos peces son poliploides.



Alelos

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

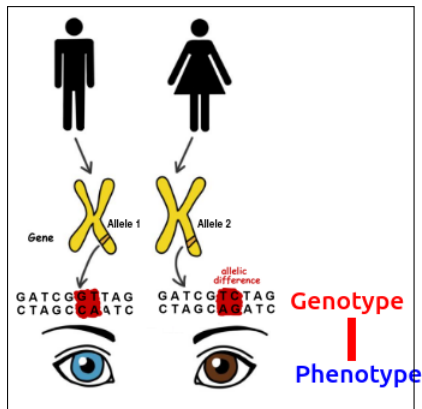
GS

Data

Training

Model

- Los alelos son formas alternativas del mismo gen que pueden diferir en 1 o más variaciones en las 4 bases A, G, C o T.
- Para cada gen, los individuos diploides tendrán cada uno dos alelos.
- Los alelos crean diversidad.
- La combinación de estos alelos es lo que define un genotipo para un individuo.
- El genotipo y otras variables externas determinan en gran medida el fenotipo de un individuo.



Genotipos y Fenotipos

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

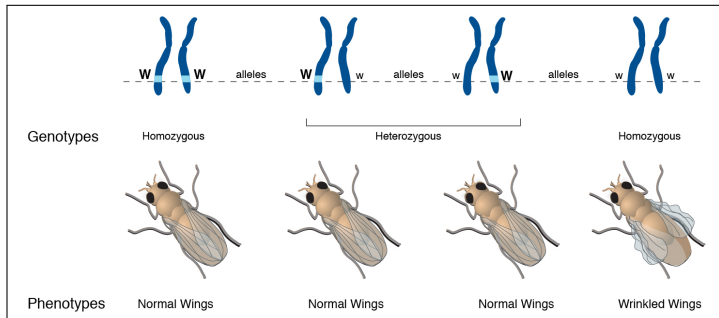
GS

Data

Training

Model

- El genotipo de un organismo es el conjunto de genes que porta.
- El fenotipo de un organismo son todas sus características observables



- El fenotipo es la suma de los efectos de su genotipo y del medio ambiente.
- Las diferencias en los genotipos pueden producir diferentes fenotipos.

Marcadores Genéticos: SNPs

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

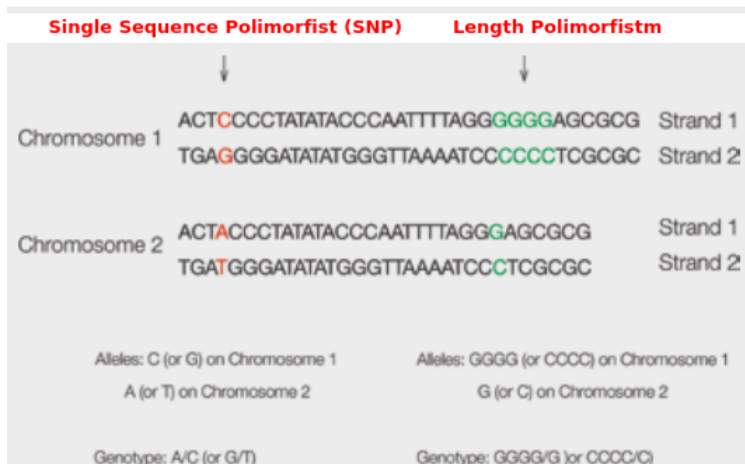
Mejoramiento

GS

Data

Training

Model



Los MG y los genes cercanos entre sí en un cromosoma generalmente se heredan juntos.

Mérito Genético

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

- **Breeding**: Mejoramiento, Crianza, Reproducción
- **Breeding Value (BV)**: Valor de Cría, Mérito Genético,
- **Estimated Breeding Value (EBV)**: Valor estimado de cría
- **Genomic Estimated Breeding Value (gEBV)**: Valor genético, valor genético estimado de cría

Advances in DNA sequencing

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

Los avances en las tecnologías de secuenciación de ADN ahora nos permiten aislar ADN (o ARN) de múltiples tipos de muestras, amplificar y secuenciar regiones del genoma y secuenciar genomas completos.



SNP chips y Estudios de Genoma Amplio

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

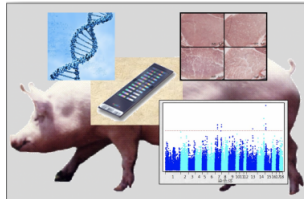
GS

Data

Training

Model

- Ya no es necesario secuenciar todo un genoma.
- Existe arreglos o chips con información de los SNPs de ese genoma:
 - Chips de calidad de la carne
 - Chips de algunas enfermedades



- Con estos chips, es mucho más fácil y económico realizar estudios:

GWAS

Asociar un único marcador o gene a una característica. Por ejemplo el color de los ojos (**características cualitativas**)

GS

O dar una medida (merito genético) de que un individuo vaya a desarrollar una característica en el futuro. Por ejemplo Peso. (**características cuantitativa**)

Quantitative Trait Loci (QTLs)

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

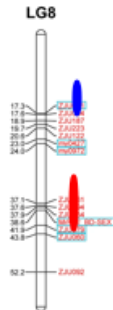
GS

Data

Training

Model

Cuando se encuentra que las ubicaciones en el genoma están correlacionadas con estos rasgos, los llamamos loci de rasgos cuantitativos (QTL)



Mejoramiento (Breeding)

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

- El mejoramiento denota el proceso de mejorar los rasgos fenotípicos deseables en las plantas o animales (ej.. rendimiento y calidad).
- Se basa en la selección artificial, que es el proceso de controlar los cambios genéticos para beneficiar las demandas humanas.
- Como resultado, los genotipos que exhiben rasgos fenotípicos mejorados son seleccionados, lo que conduce a la acumulación de alelos favorables.



Estimated Breeding Value (EBV) (Tradicional)

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

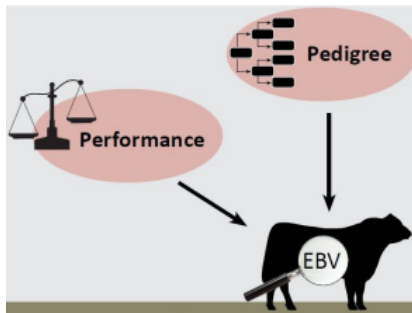
GS

Data

Training

Model

- Selection for breeding has historically been made using estimated breeding values (EBV), without identifying genes involved in phenotypes.
- EBVs were simply estimated from the study of pedigrees and phenotypic records with the knowledge of the heritability of each trait



Problemas con el mejoramiento tradicional

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

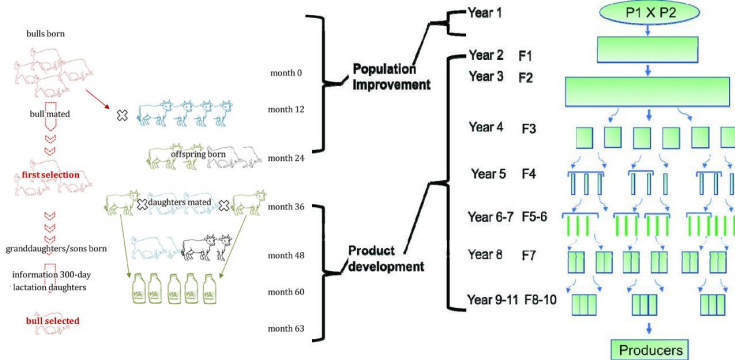
Mejoramiento

GS

Data

Training

Model



Sin embargo, la eficiencia de este método disminuye a medida que su uso se expande a rasgos que son difíciles de medir, tienen baja heredabilidad o pueden medirse solo después de varios años y / o generaciones.

Genomic Estimated breeding values (gEBV)

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

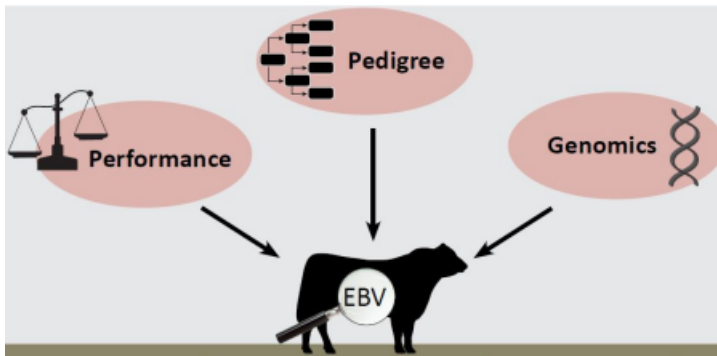
GS

Data

Training

Model

Por esta razón, la identificación y el conocimiento de los genes que subyacen a estos rasgos en animales y plantas es de gran valor en agrigenómica.



¿Qué es la selección genómica?

Selección Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

La selección genómica o predicción genómica es una nueva herramienta que puede predecir un mérito genético vegetal / animal basándose en la puntuación de marcadores de ADN, como los polimorfismos de un solo nucleótido (SNP).



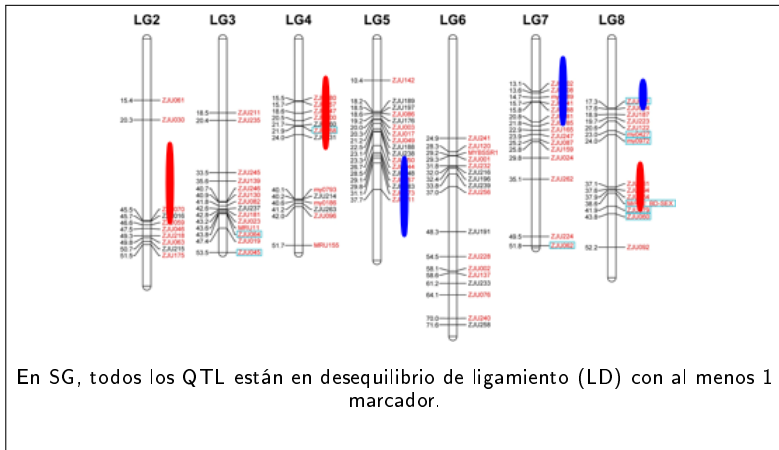
GS puede predecir cómo se comportará la progenie de un animal antes de que se midan los rasgos o incluso inmediatamente después del nacimiento sin necesidad de ninguna otra información (como el pedigrí).

Principio de Selección Genómica

Selección Genómica

Luis Garreta

SG se basa en el principio de que la información de un gran número de marcadores, sin saber dónde se encuentran los genes, se puede utilizar para estimar los valores genéticos.



Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

¿Cómo funciona la SG?

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

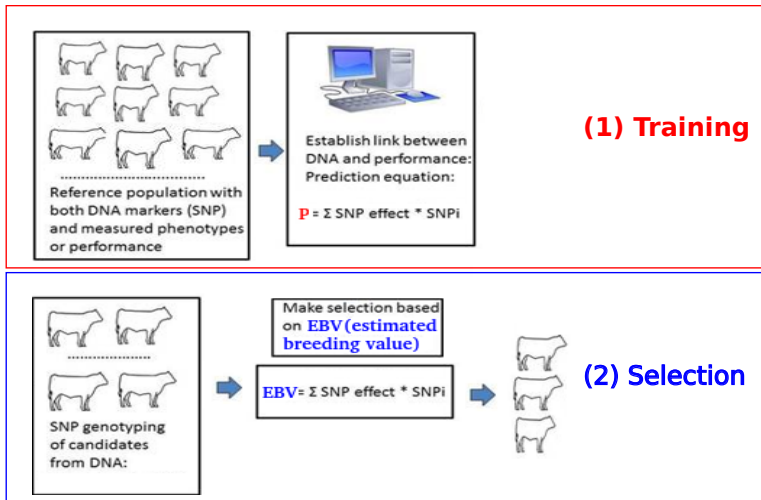
Mejoramiento

GS

Data

Training

Model



Flujo de trabajo de GS

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

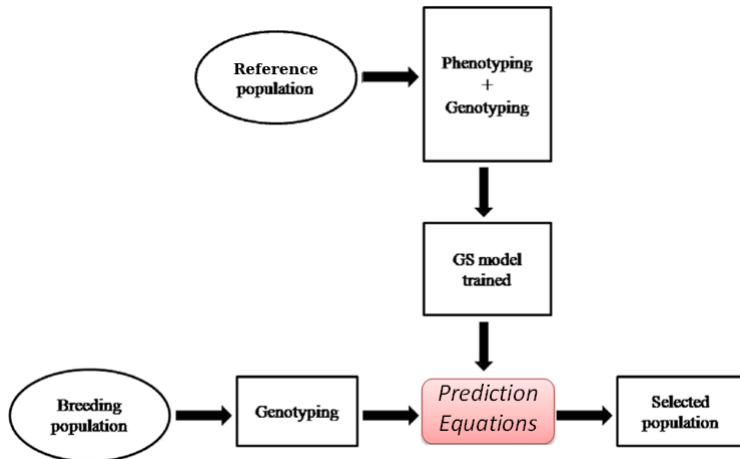
Mejoramiento

GS

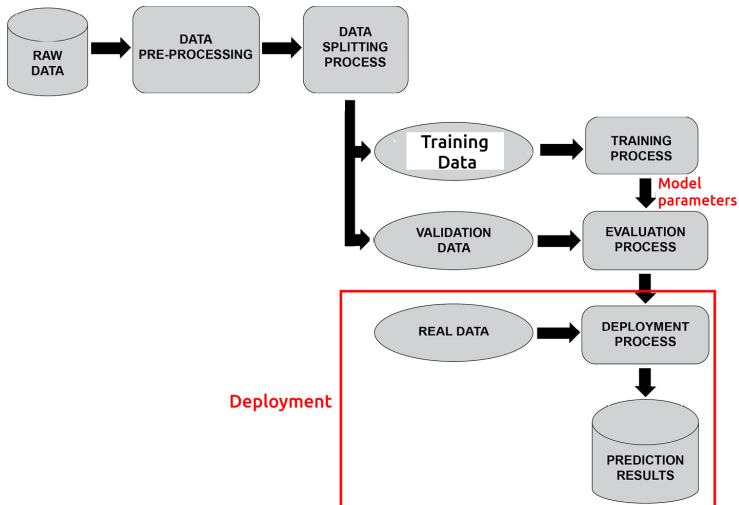
Data

Training

Model



Proceso de aprendizaje automático



Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

Data

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

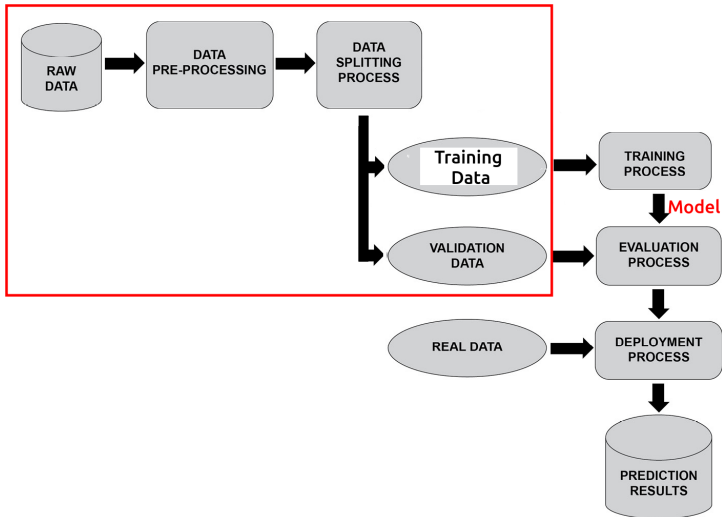
Mejoramiento

GS

Data

Training

Model



Genotypes and Phenotypes (traits)

- Documentation of **phenotypes**

Individuals	Value
ACBrador	3.59
AdirondackBlue	4.07
AlIBlue	4.73
AlpineRusset	4.85
Alturas	4.46
Andover	2.54
Atlantic	2.41
BannockRusset	4.86
BeaconChipper	2.54



- Documentation of **genotypes** for each marker of interest

Individuals	Markers (SNPs)						
	c2_41417	c2_24258	c2_21112	c2_21120	c2_21118	c2_21114	c2_4410
ACBrador	2	0	2	0	NA	1	NA
ACLP175195	0	0	2	0	NA	0	0
ADGP195204	0	0	2	0	NA	NA	1
AdirondackBlue	2	2	2	0	1	1	1
AdirondackRed	0	1	2	0	1	0	0
AlIBlue	0	2	1	0	1	0	0
Allegany	2	0	0	0	0	0	0
AlpineRusset	0	0	1	2	NA	1	1

Genotypic class

Individual sequences									
GAT	A	T	T	C	G	T	A	C	G
GAT	G	T	T	C	G	T	A	C	G
GAT	A	T	T	C	G	T	A	C	G
GAT	A	T	T	C	G	T	A	C	G
GAT	G	T	T	C	G	T	A	C	G
GAT	G	T	T	C	G	T	A	C	G
SNPs	A/G						G/T	A/T	

Genotype coding schemes

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

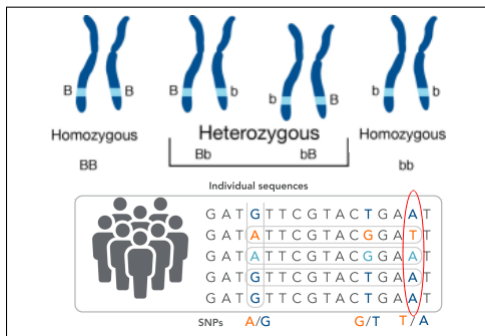
Mejoramiento

GS

Data

Training

Model



Minor Allele	BB	Bb	bB	bb
Binary	00	01	10	11
Numeric	0	1		2
ACGT	AA	AT	TA	TT

Data: Splitting into Training and Testing datasets

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

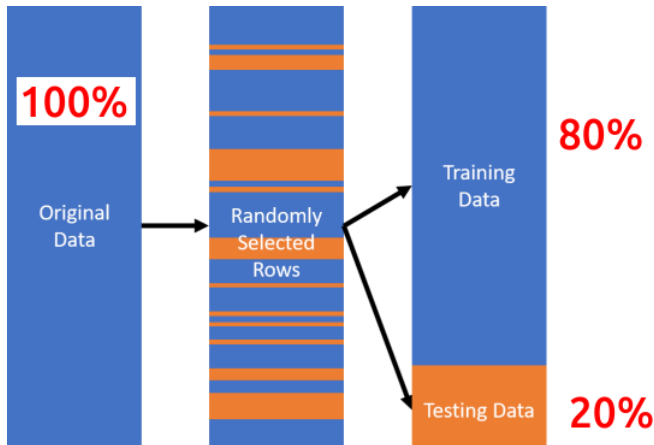
Mejoramiento

GS

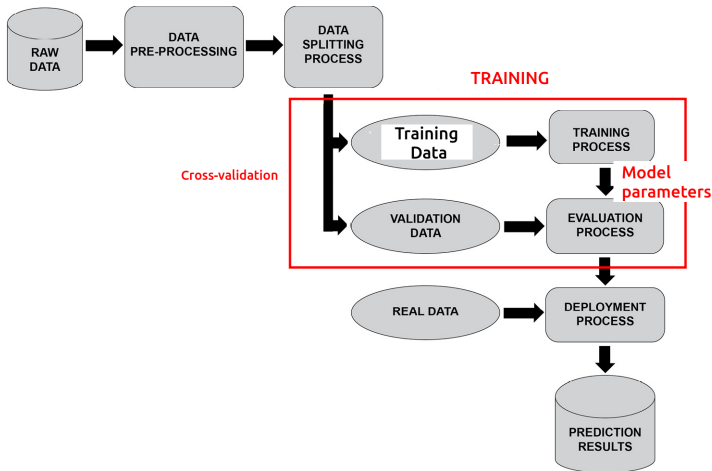
Data

Training

Model



Training



Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

Model

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

A simple but frequently used genetic model is that the phenotypic value of an individual (P) is expressed as the summation of the genetic value (G) and the residual environmental effect (E):

$$P = G + E$$

- P : Phenotype
- G : Genotype
- E : Environment

Linear model for a single marker

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

$$y_i = \beta_0 + x_{1i}\beta_1 + \varepsilon_i$$

- Y_i : Observed phenotype individual i
- β_1 : Marker fixed effect
- X_{1i} : Genotype of the marker 1 of the i th individual
- ε_i : Error term

Linear model for M marker

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

- General equation for M markers

$$\begin{aligned}y_i &= \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 \dots + x_{Mi}\beta_M + \varepsilon_i \\ &= \sum_{j=0}^M x_{ji}\beta_j + \varepsilon_i\end{aligned}$$

- Minimization term:

$$E = \sum_i (y_i - \beta_0 - x_{1i}\beta_1)^2$$

Large p , small n problem

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

- $p \gg n$
 - n : Number of individuals (observations)
 - p : Number of markers (variables)
- Regression model algorithms do not work with $p \gg n$
- These algorithms need to be modified to work with this kind of data

Genomic Selection Methods

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

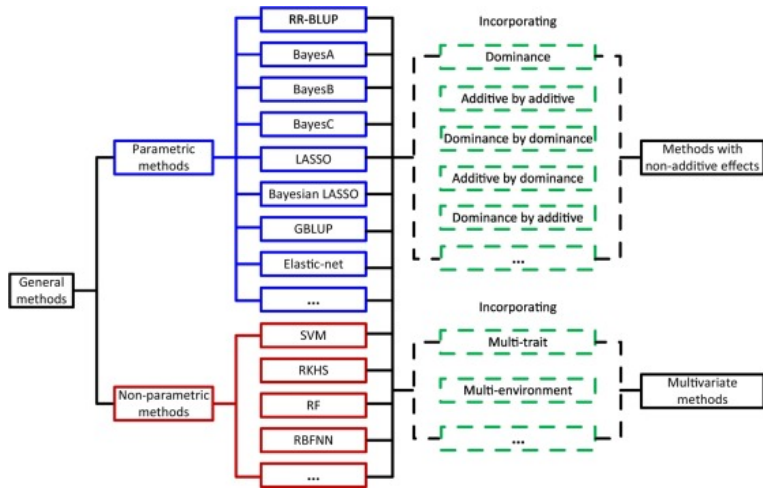
Mejoramiento

GS

Data

Training

Model



Source: Genomic selection methods for crop improvement: Current status and prospects

Evaluating a Model by Cross-Validation

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

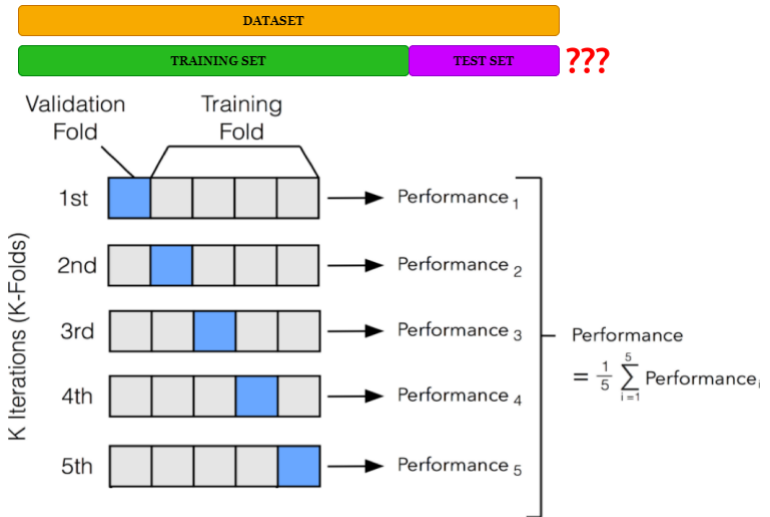
Mejoramiento

GS

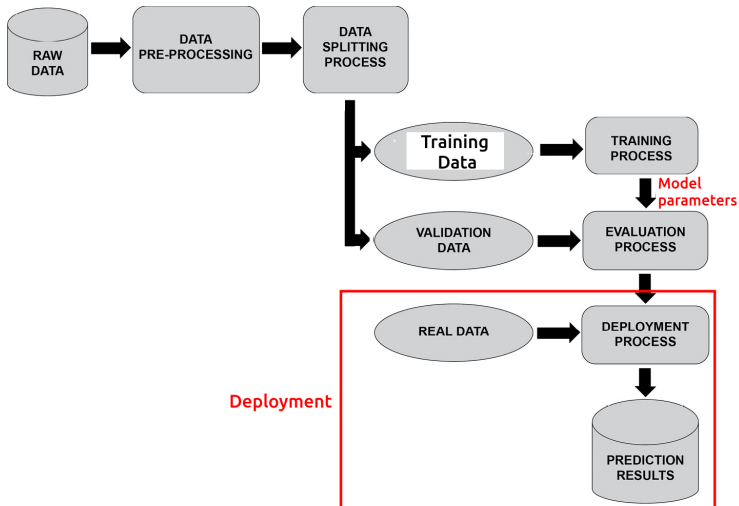
Data

Training

Model



Deployment



Deployment

Selection

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

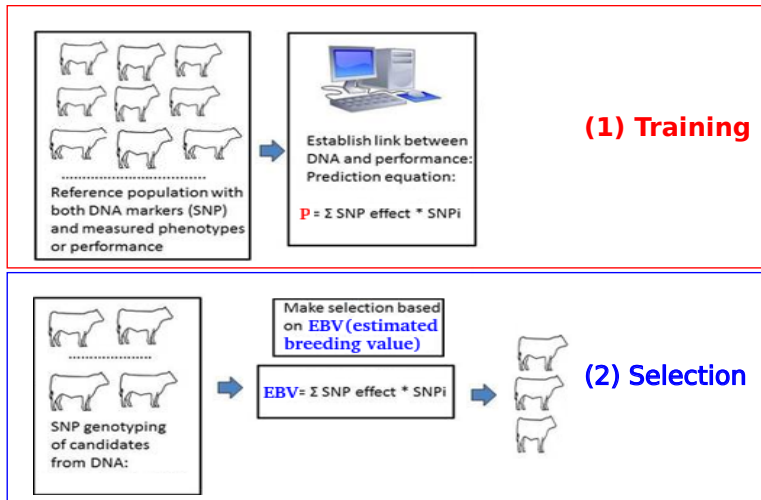
Mejoramiento

GS

Data

Training

Model



Tarea: Problema de $P \gg n$ más grande que n

Selección
Genómica

Luis Garreta

Intro

Conceptos

DNA sequencing

SNP chips

QTLs

Mejoramiento

GS

Data

Training

Model

Investigar sobre el problema de $p \gg n$ “P greater than N problem”. Específicamente realizar un ensayo corto describiendo en español o (inglés) plano los siguientes puntos:

- De qué trata este problema
- Qué implicaciones tiene el problema en los algoritmos de aprendizaje automático.
- Cómo solucionan o suavizan este problema.