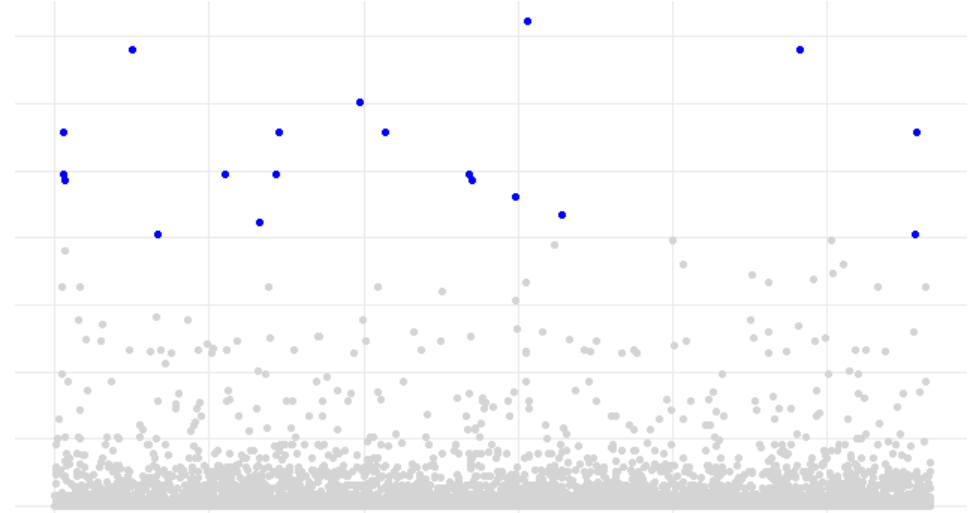
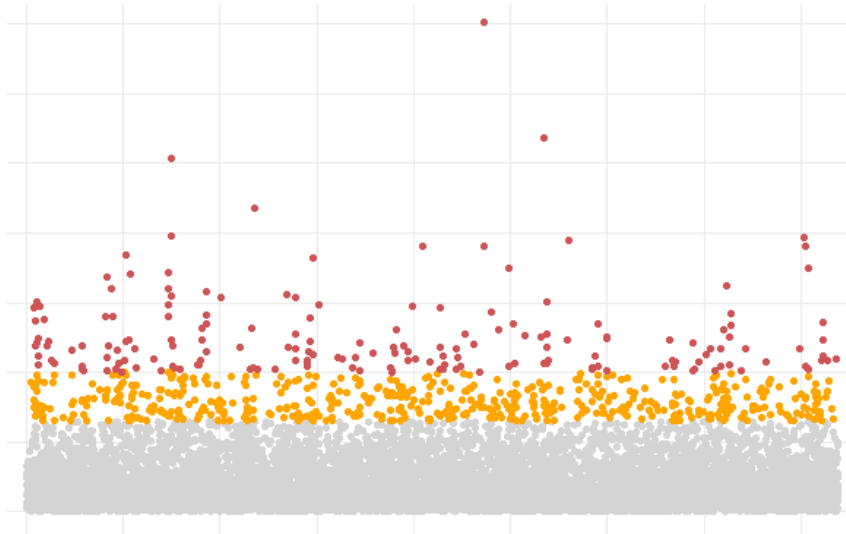


# DETECCIÓN DE REGIONES GENÓMICAS RESPONSABLES DE LA ADAPTACIÓN LOCAL

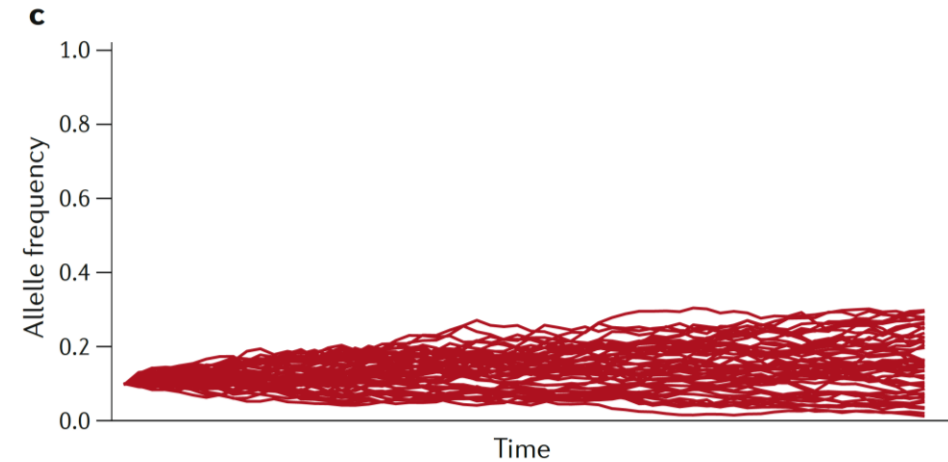
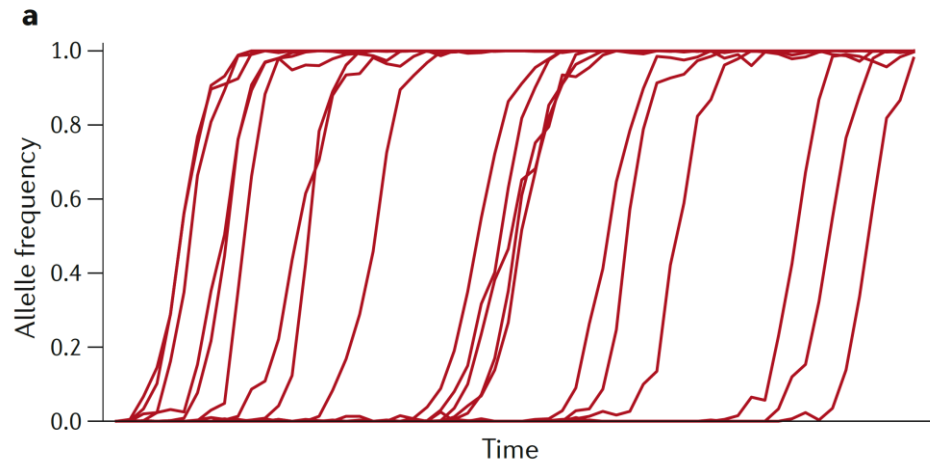
---



JOSÉ LUIS BLANCO PASTOR

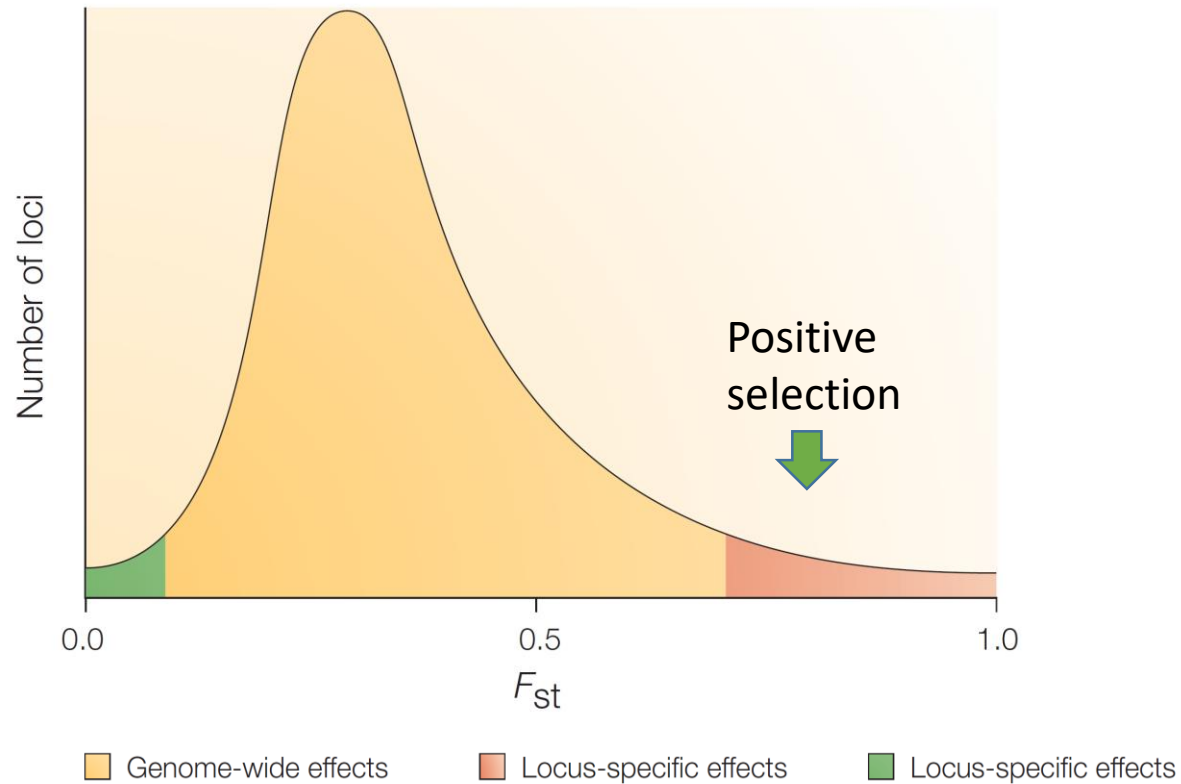
# Las marcos teóricos de la genética evolutiva poblacional para el análisis de la adaptación local:

- **Genética poblacional molecular** (análisis de diferenciación):
  - Cambios brusco en las frecuencias alélicas de uno en uno. Hard sweep model.
    - Versiones modernas incluyen “soft sweeps” (cambios lentos en las frecuencias alélicas, alelos seleccionados no fijados).
- **Genética cuantitativa** (asociación genotipo-ambiente GEAs).
  - Cambios sutiles de las frecuencias alélicas en múltiples genes a la vez. Infinitesimal model, “soft shifts”.
    - Versiones modernas incluyen “large effect loci” (cambios rápidos en las frecuencias alélicas, alelos seleccionados próximos a la fijación).

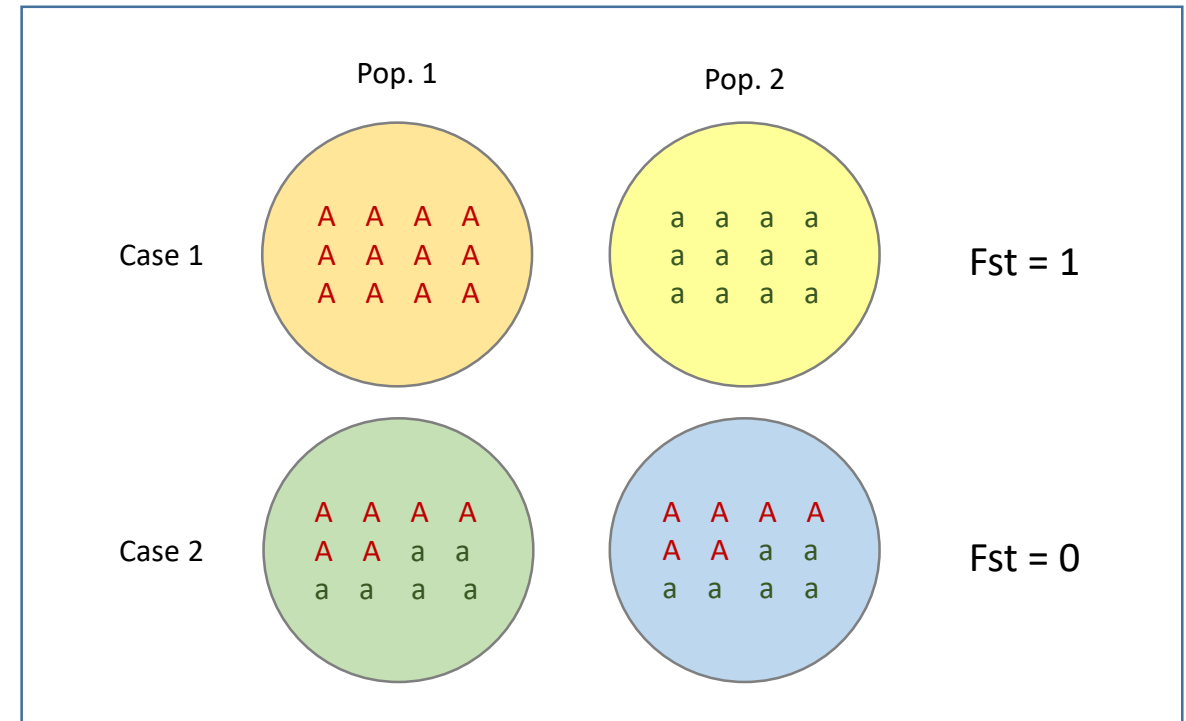


# Methods:

## 1. Genética poblacional → “Population differentiation methods”



$$F_{st} = \frac{\text{Variation between pops.} - \text{Variation within pops.}}{\text{Variation between pops.}}$$



## Methods:

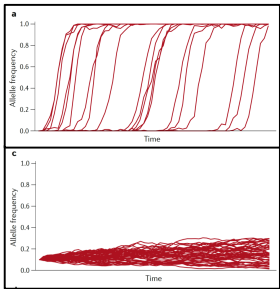
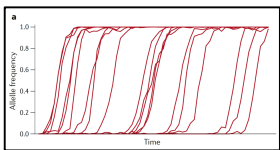
1. Genética poblacional → “Population differentiation methods” → Fst-based tests are blind tests

- Caso facil:
  - *Flowering locus T*
- Caso difícil:
  - PF01535//PF12854//PF13041// PF14432 – PPR repeat (PPR) // PPR repeat (PPR\_1) // PPR repeat family (PPR\_2) // DYW family of nucleic acid deaminases (DYW\_deaminase)



# Methods:

- 1. Genética poblacional → Population differentiation methods
- 2. Genética cuantitativa → Métodos de asociación



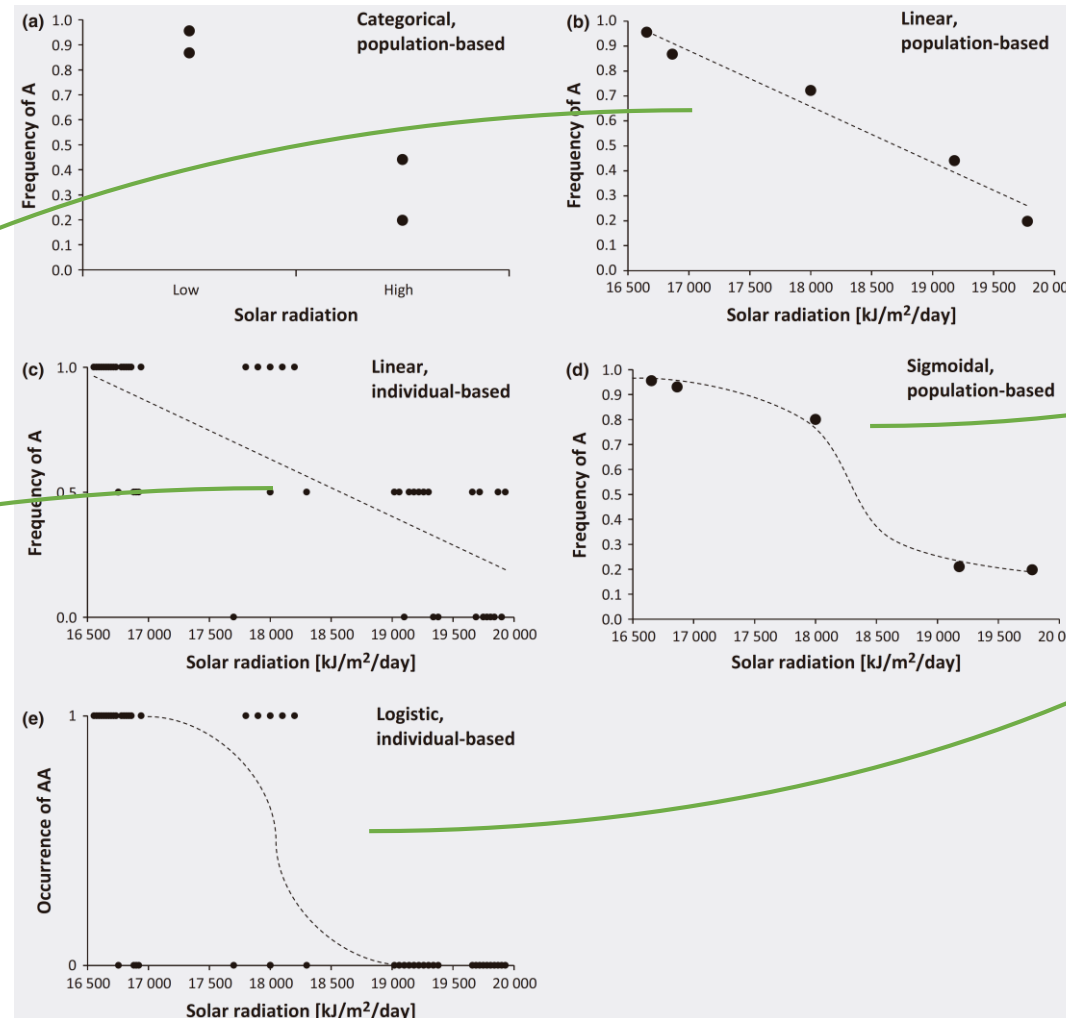
Test	Original references
1. Population differentiation methods	
Bayescan2/ Bayescan3	Beaumont and Balding (2004); Foll and Gaggiotti (2008)
PCAdapt	Duforet-Frebourg et al. (2014) Luu et al (2017)
FLK	Bonhomme et al. (2010)
OutFLANK	Whitlock and Lotterhos (2015)
SelEstim	Vitalis et al. (2013)
2. Association methods (logistic, linear)	
SAM/SAMβADA	Joost et al. (2007); Stucki (2014)
Bayenv/Bayenv2	Coop et al. (2010); Günther and Coop (2013)
BayPass	Gautier (2015)
RDA	Rao (1964); Forester et al. (2016)
LFMM	Frichot et al. (2013); Frichot and François (2015); Gain and François (2021)

Poder estadístico superior y tasa de falsos positivos menor en comparación a otros métodos



# Methods:

## 2. Genética cuantitativa → Métodos de asociación ambiental – “Genotype Environment Association analyses (GEAs)”



Logistic regression

Linear regression

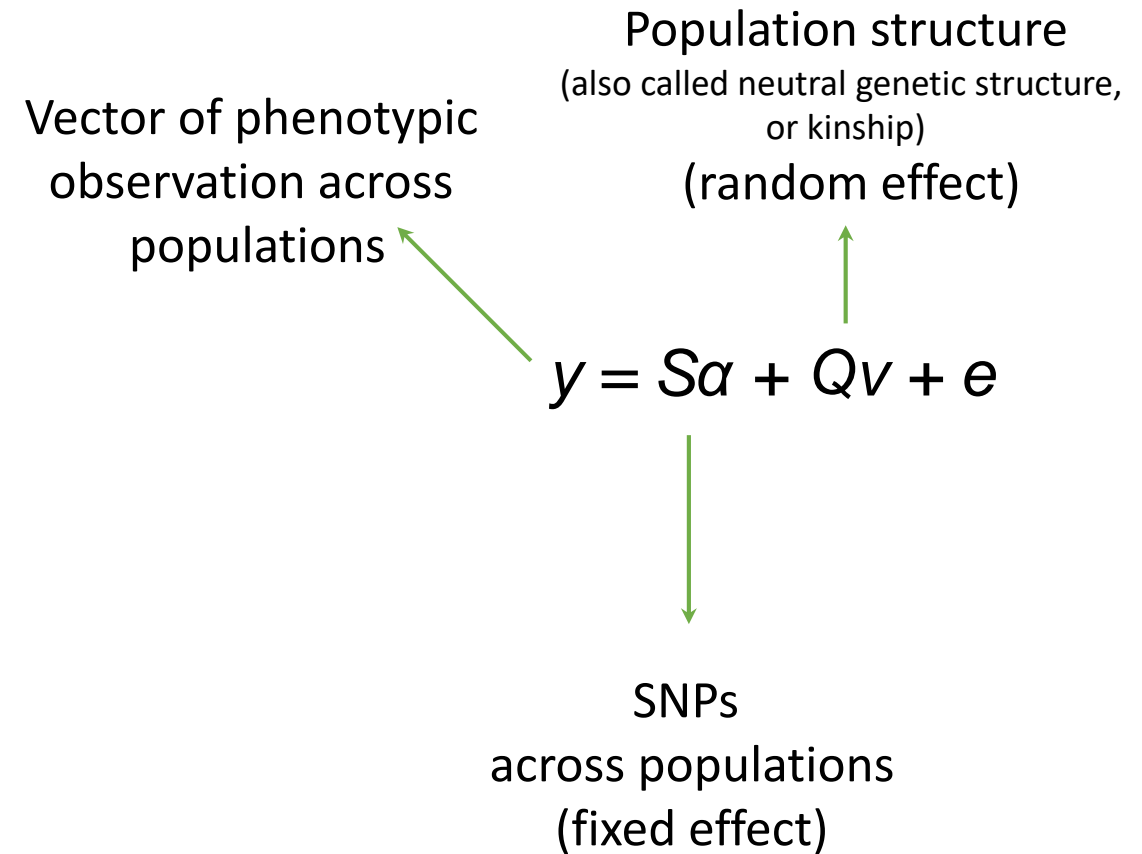
*Modelar frecuencias alélicas  
utilizando descriptors  
bioclimáticos*

## Methods:

### 2. Genética cuantitativa → Métodos de asociación

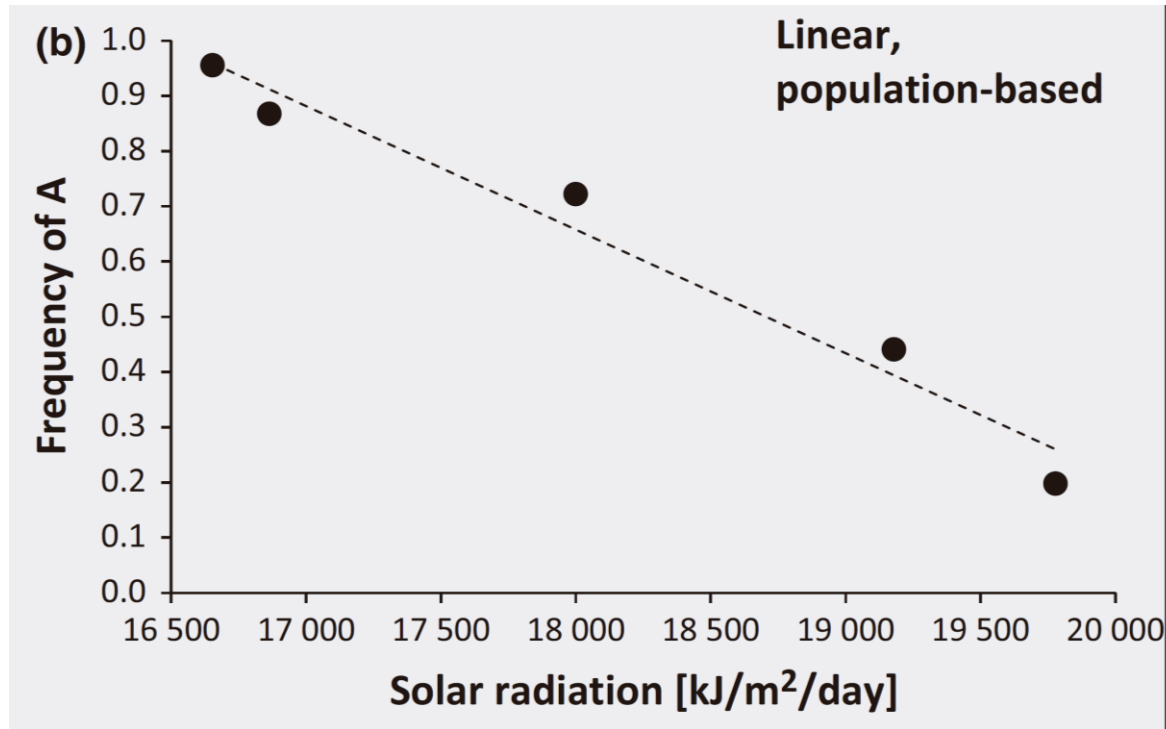
*Las diferencias fenotípicas dentro de una especie van a venir determinadas por presiones selectivas y por el reparto diferencial de linajes genéticos causado por la pérdida de flujo genético entre las poblaciones.*

#### Traditional Mixed Model (GWAS)

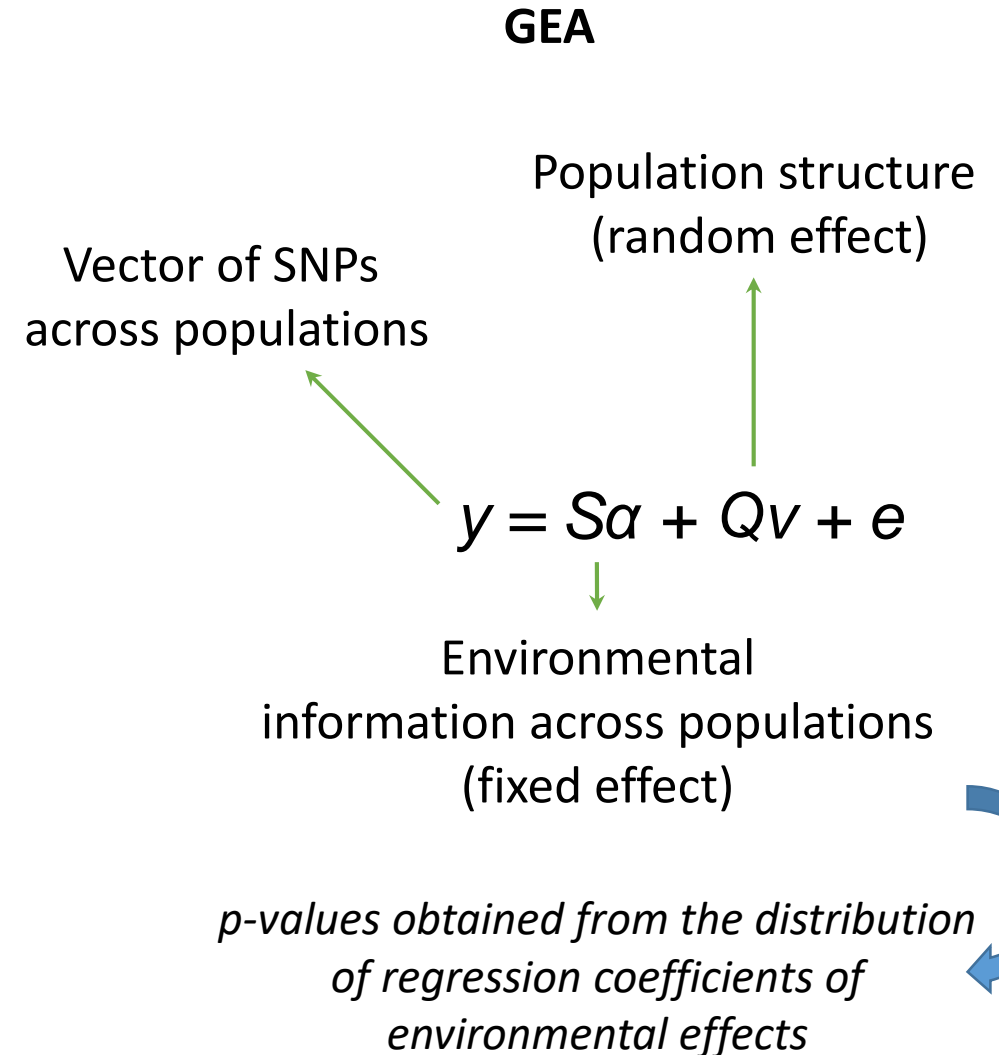


## Methods:

### 2. Genética cuantitativa → Métodos de asociación



Reilstab et al. 2015. Molecular Ecology



Notation following Yu et al. 2006. Nature Genetics



# Methods:

## 2. Genética cuantitativa → Métodos de asociación

### Desafíos de los métodos de asociación

#### 1. *Necesidad de corrección de pruebas múltiples*

- Bonferroni correction

$$\alpha_{\text{corregido}} = \frac{\alpha}{\text{número de tests independientes}} ;$$

Ejemplo:

$\alpha = 0.05$  en GEAs con 10.000 SNPs equivale a

$$\alpha_{\text{corregido}} = 0.000005$$

La corrección Bonferroni es demasiado conservadora para estudios genómicos

# Methods:

## 2. Genética cuantitativa → Métodos de asociación

### Desafíos de los métodos de asociación

#### 1. *Necesidad de corrección de pruebas múltiples*

- **False Discovery rate (FDR)**, método Benjamini & Hochberg (1995)  
Proporción esperada de falsos positivos
  - Ordenar los  $n$  test de menor a mayor  $p$ -value ( $p_1, p_2, \dots p_n$ )
  - Se define  $k$  como la última posición para la que se cumple que  $p \leq \alpha \frac{i}{n}$  (where  $\alpha = 0.05$ ,  $i$  is the test index and  $n$  is the number of tests = 10)
  - Se consideran significativos todos los  $p$ -value hasta la posición  $k$  ( $p_1, p_2, \dots p_k$ )

Ejemplo:

Significativos para  $p$ -value < 0.05 con corrección Bonferroni  
 $\alpha_{\text{corregido}} = 0.005$

Significativos para un  $p$ -value < 0.05

test	p_values	Index (i)	$\alpha*i/n$	Significance
test_4	0.0001	1	0.005	TRUE
test_10	0.0002	2	0.010	TRUE
test_7	0.0100	3	0.015	TRUE
test_3	0.0130	4	0.020	TRUE
test_9	0.0300	5	0.025	FALSE
test_6	0.0400	6	0.030	FALSE
test_2	0.0700	7	0.035	FALSE
test_8	0.1500	8	0.040	FALSE
test_5	0.2600	9	0.045	FALSE
test_1	0.5200	10	0.050	FALSE

Significativos para un FDR = 0.05

# Methods:

## 2. Genética cuantitativa → Métodos de asociación

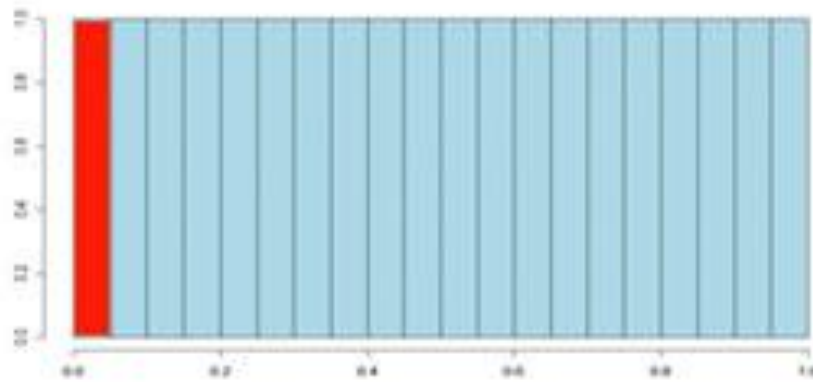
### Desafíos de los métodos de asociación

#### 1. *Necesidad de corrección de pruebas múltiples*

- **False Discovery rate (FDR)**, método q-value, Storey & Tibshirani (2003)

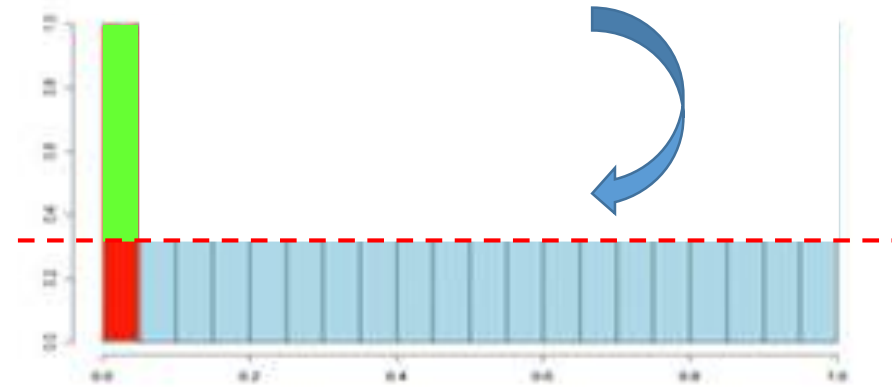
$$\text{FDR} = \frac{\text{no. falsos positivos}}{\text{no. total positivos}}$$

Distribución uniforme de p-valores en un experimento sin verdaderos positivos (todos los p-valores son nulos). **Misma probabilidad de p-valor de 0 a 1.**



*p-value*

*p-valores calibrados (distribución uniforme de p-valores nulos)*



# Methods:

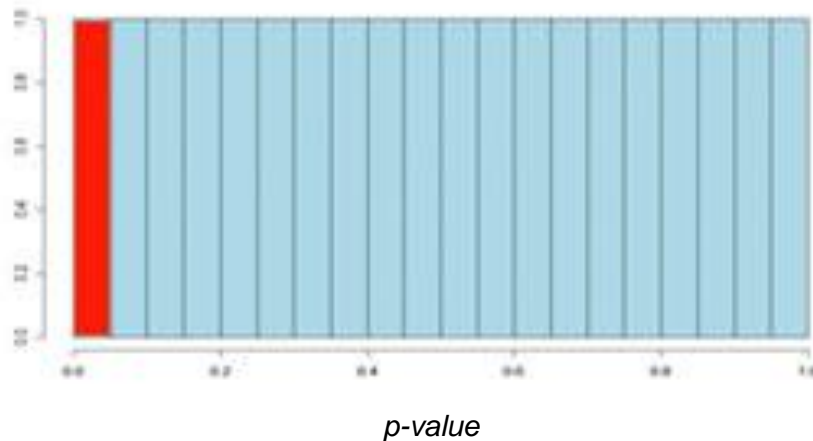
## 2. Genética cuantitativa → Métodos de asociación

### Desafíos de los métodos de asociación

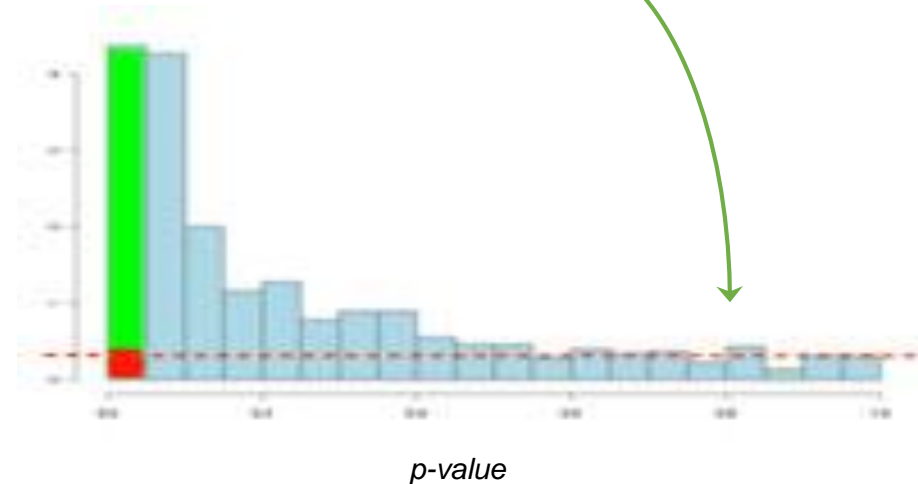
#### 1. *Necesidad de corrección de pruebas múltiples*

- **False Discovery rate (FDR)**, método q-value, Storey & Tibshirani (2003)

$$\text{FDR} = \frac{\text{no. falsos positivos}}{\text{no. total positivos}}$$



En experimentos genómicos los p-valores nulos no tienen una distribución uniforme. Los p-valores están “inflados” a causa de la estratificación (estructura genética neutral, estructura poblacional, kinship, etc.) no capturada en el modelo.



# Methods:

## 2. Genética cuantitativa → Métodos de asociación

### Desafíos de los métodos de asociación

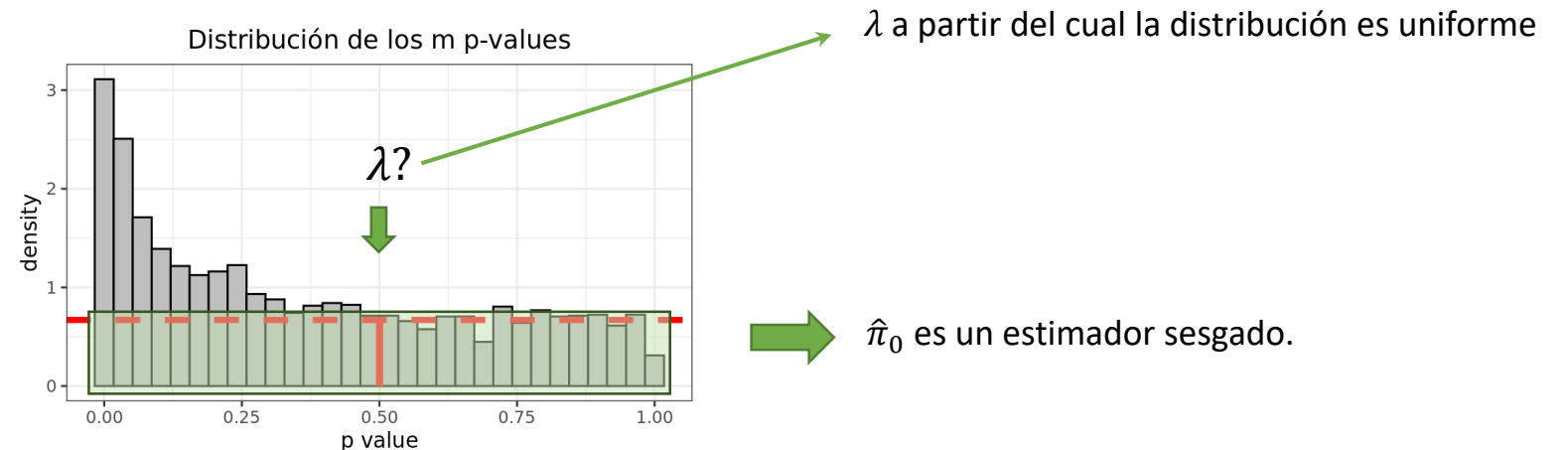
#### 1. *Necesidad de corrección de pruebas múltiples*

- **False Discovery rate (FDR)**, método q-value, Storey & Tibshirani (2003)

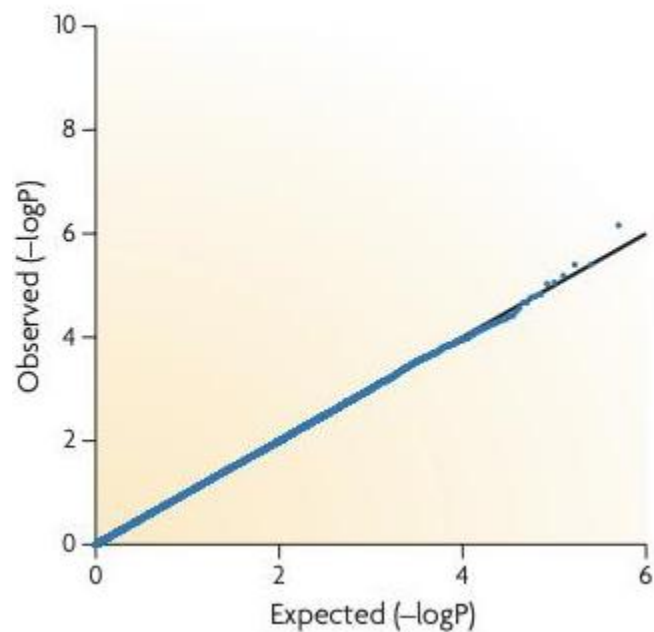
$$\text{FDR} = \frac{\text{no. falsos positivos}}{\text{no. total positivos}}$$

$$\hat{q}(p_i) = \widehat{\text{FDR}} = \frac{\hat{\pi}_0 * \text{all } p\text{values} * \alpha}{\text{no. } p\text{values} \leq \alpha}$$

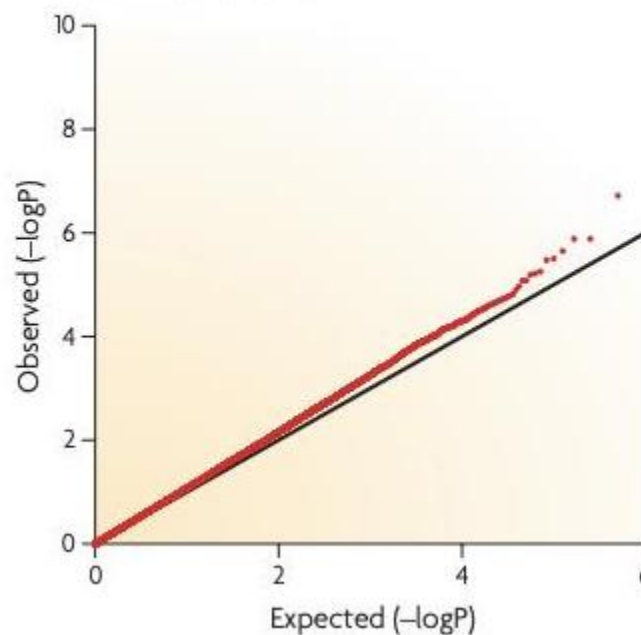
$$\hat{\pi}_0 = \frac{\sum p\text{values} > \lambda}{\text{all } p\text{values} * (1 - \lambda)}$$



**a No stratification**



**b Stratification**

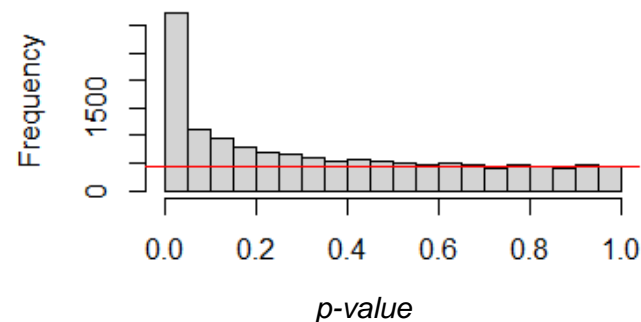
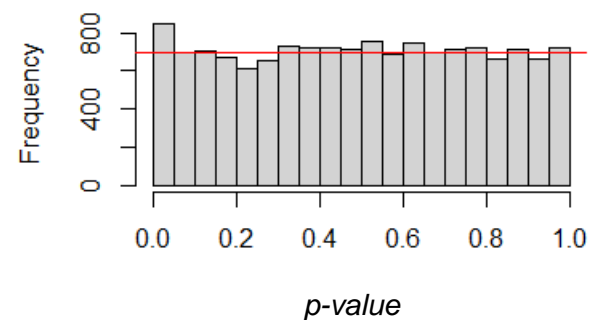


Price et al. 2010. Nature Reviews Genetics

$$y = S\alpha + Qv + e$$



*Not all stratification  
is captured here*



*Divide by the Genomic Inflation Factor (GIF)*



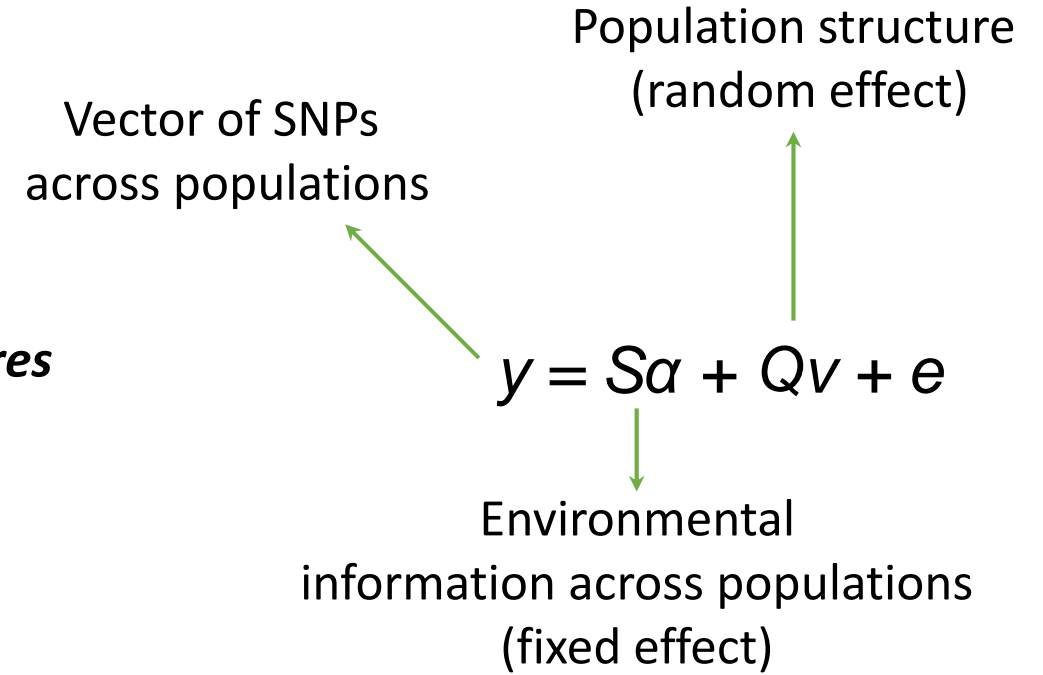
## Methods:

### 2. Genética cuantitativa → Métodos de asociación

#### Desafíos de los métodos de asociación

##### 2. *Cuando los valores ambientales (y/o valores de los caracteres fenotípicos) co-varían con la estructura poblacional*

- Bajo poder estadístico (corrección excesiva).
- Exceso de falsos positivos (corrección insuficiente).



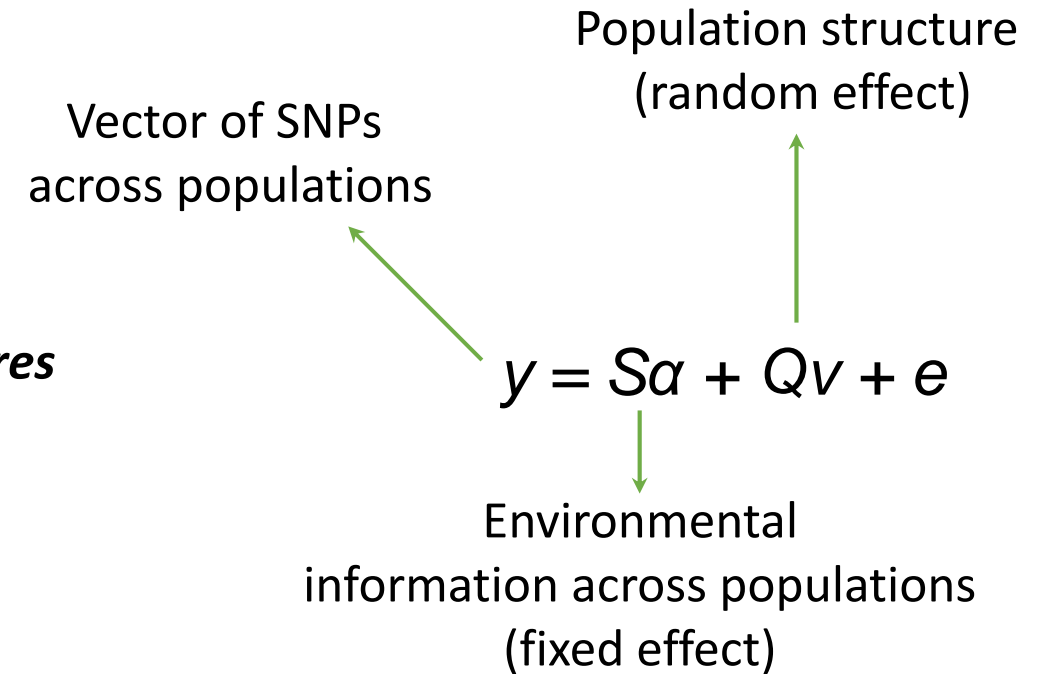
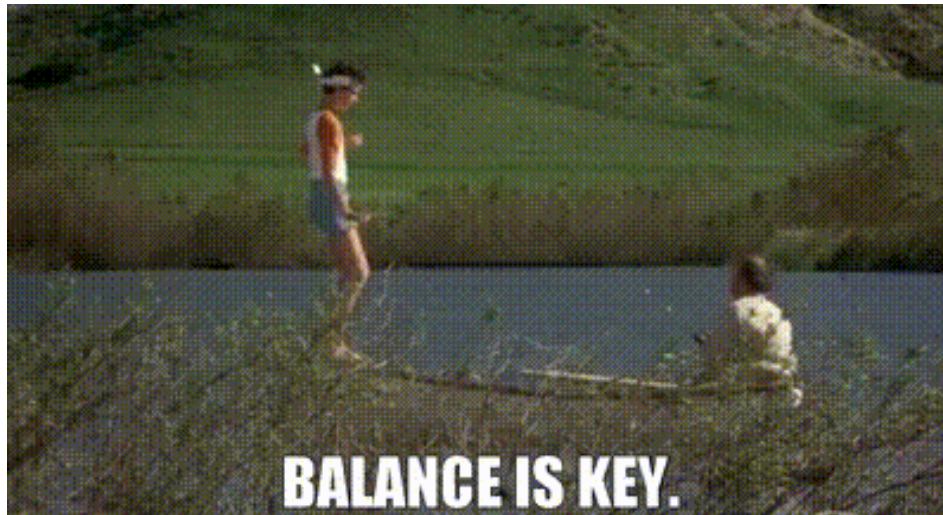
## Methods:

### 2. Genética cuantitativa → Métodos de asociación

#### Desafíos de los métodos de asociación

#### 2. *Cuando los valores ambientales (y/o valores de los caracteres fenotípicos) co-varían con la estructura poblacional*

- Bajo poder estadístico (corrección excesiva).
- Exceso de falsos positivos (corrección insuficiente).





# Methods:

## 2. Genética cuantitativa → Métodos de asociación

### Desafíos de los métodos de asociación

1. *Necesidad de corrección de pruebas múltiples*
2. *Los valores ambientales (y/o valores de los caracteres fenotípicos) co-varían con la estructura poblacional*

Aplicar corrección por FDR  
(no Bonferroni que es demasiado conservadora)  
o bien, ningún tipo de  
corrección de pruebas  
múltiples

Evidencia a partir de  
diferentes aproximaciones  
analíticas en la que una de  
ellas (o todas) no aplique(n)  
corrección por estructura  
poblacional

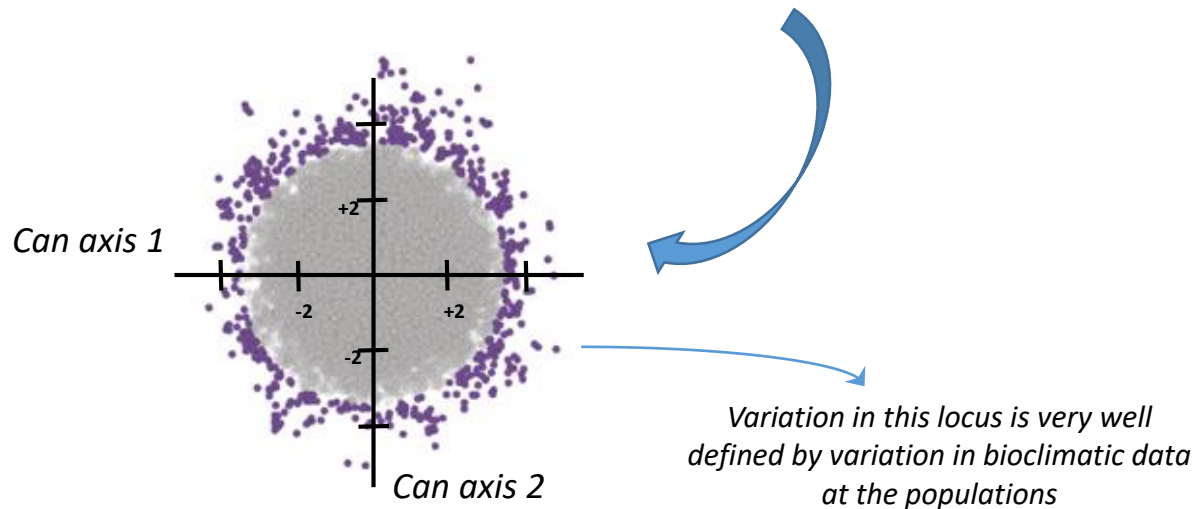
Evidencia a partir de fuentes  
de datos diferentes (e.g.  
variables ambientales y  
fenotípicas)

# RDA

1. **multiple linear regression** of loci on bioclimatic data
2. **PCA** de los valores ajustados ( $\hat{Y}$ ).

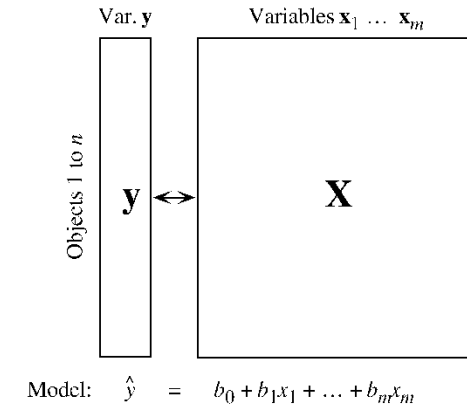
PCA of  $\hat{Y}$  is a set axes defined as a linear combination of loci constrained by bioclimatic data

Ordenación donde loci outliers identificados por una excesiva dispersion con respecto al resto (distancia al centro de los ejes RDA)

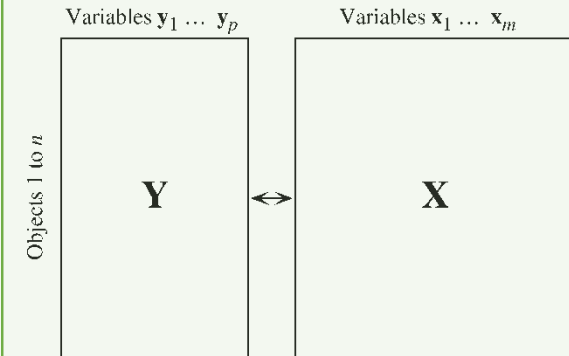


## RDA – Rao (1964); Forester *et al.* (2015)

(b) Ordination of  $y$  (single axis) under constraint of  $X$ : multiple regression



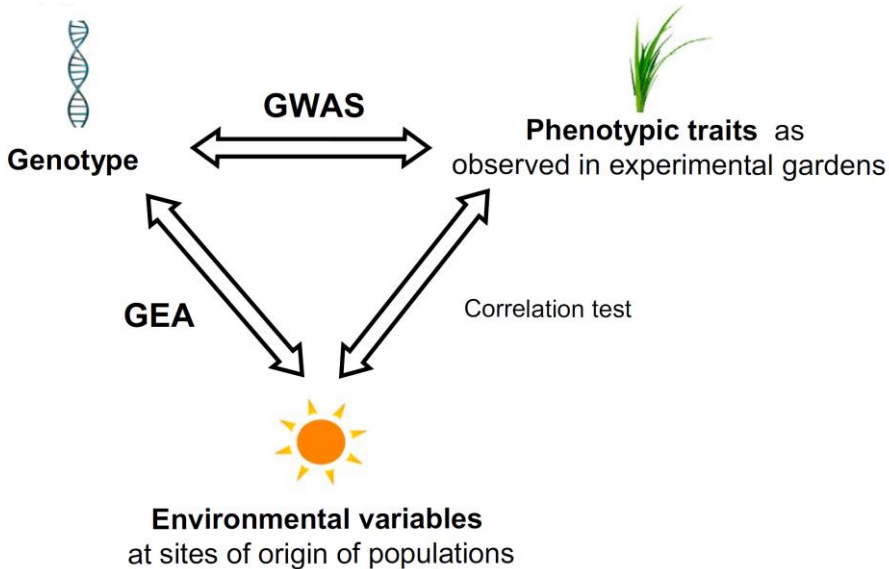
(c) Ordination of  $Y$  under constraint of  $X$ :  
redundancy analysis (RDA)  
canonical correspondence analysis (CCA)



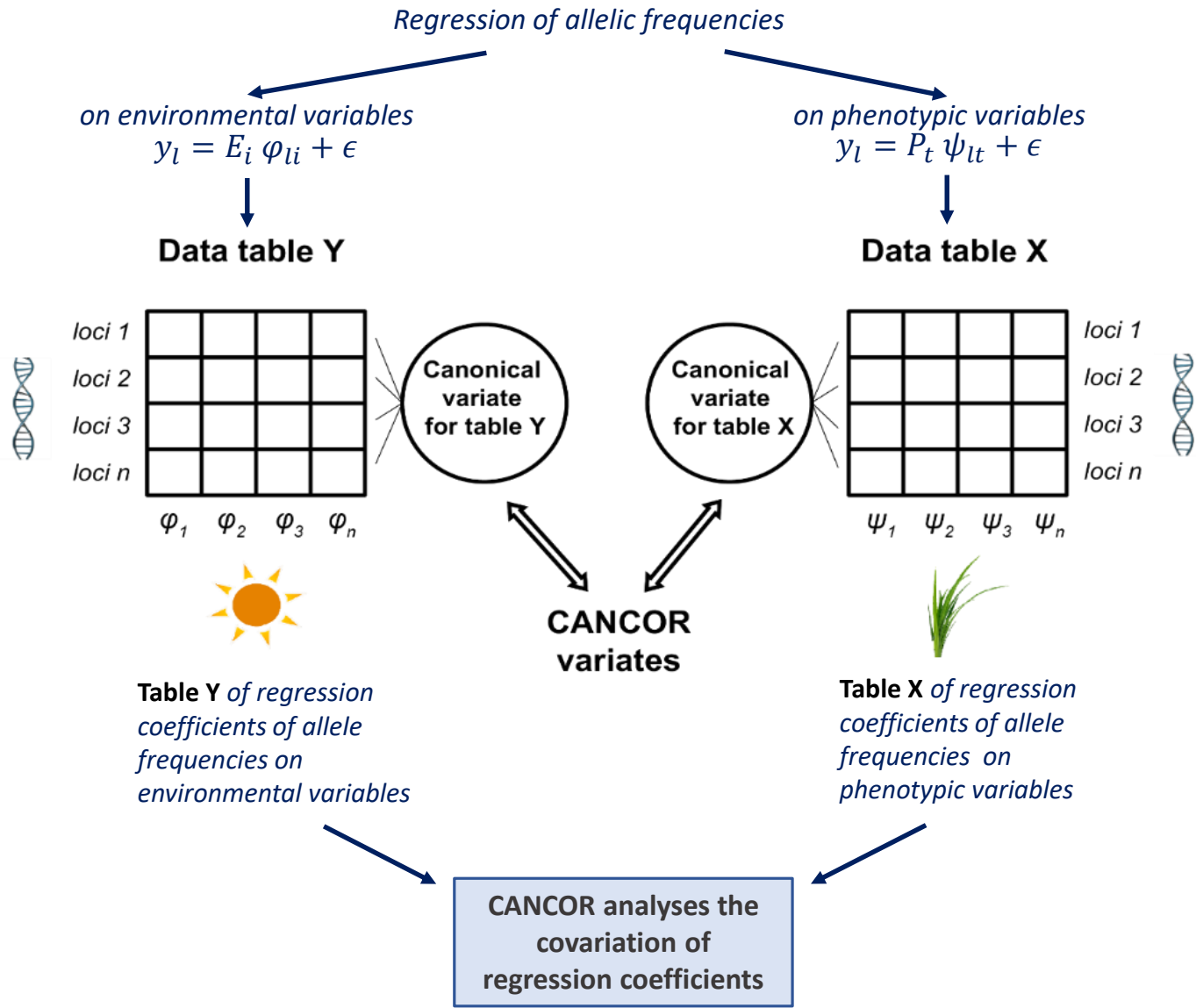
**HANDS ON!**





# GEA + GWAS



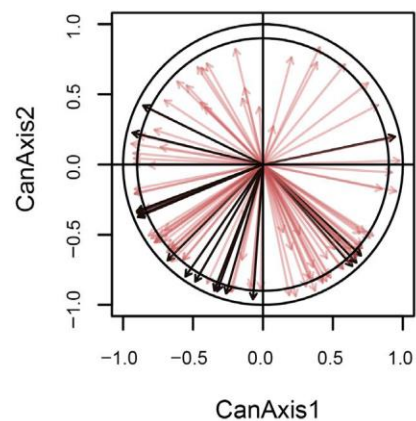
# CANCOR test



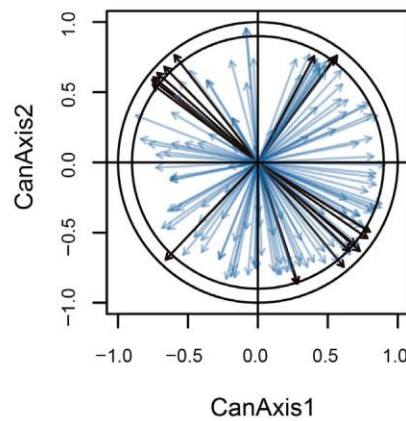
## Canonical correlations reveal adaptive loci and phenotypic responses to climate in perennial ryegrass

José Luis Blanco-Pastor<sup>1</sup>  | Philippe Barre<sup>1</sup> | Thomas Keep<sup>1</sup> | Thomas Ledauphin<sup>1</sup> | Abraham Escobar-Gutiérrez<sup>1</sup> | Anna Maria Roschanski<sup>2</sup> | Evelyn Willner<sup>2</sup> | Klaus J. Dehmer<sup>2</sup> | Matthew Hegarty<sup>3</sup> | Hilde Muylle<sup>4</sup> | Elisabeth Veeckman<sup>4,5,6</sup> | Klaas Vandepoele<sup>4,5,7</sup> | Tom Ruttink<sup>4</sup> | Isabel Roldán-Ruiz<sup>4,6</sup> | Stéphanie Manel<sup>8</sup>  | Jean-Paul Sampaou<sup>1</sup>

(b) CCorA environment plot  
First data table (Y)



(c) CCorA phenotype plot  
Second data table (X)



## CANCOR test

