

# **GENOMIC APPROXIMATIONS TO IMPROVE PATHOGEN RESISTANCE IN TETRAPLOID POTATO**

**Felix Enciso-Rodríguez, Gustavo de los Campos, Marco Lopez-Cruz  
Joseph Coombs, David Douches.**

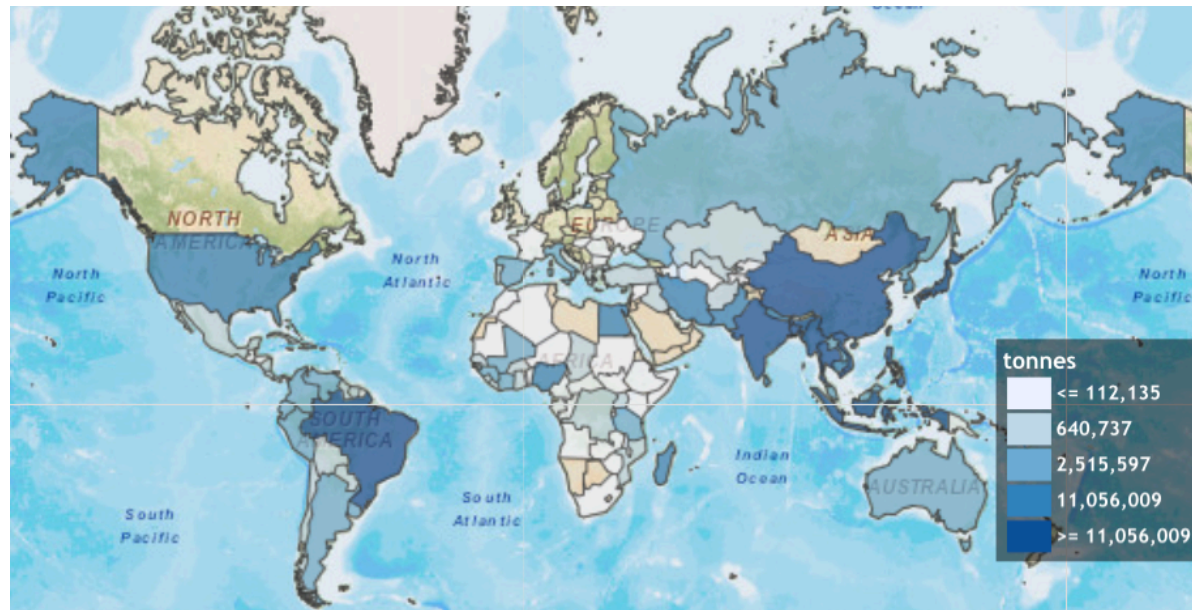
MSU Potato Breeding and Genetics Program  
Department of Plant, Soil, and Microbial Sciences  
Department of Epidemiology and Biostatistics  
Michigan State University

# Outline

- Introduction
- Objectives
- Background
- Methodology
- Preliminary results
- Further steps

## Introduction

### Potato (*Solanum tuberosum* L)



- 5<sup>TH</sup> agricultural commodity, worldwide.
- Worldwide production (2013): 372, 452,524 tons
- United States: 4<sup>th</sup> producer with 20,459,369.33 tons

Source: FAO STATS, 2015

## Introduction

### Late blight (*Phytophthora infestans* )



MSU potato breeding program.

- Irish potato famine in the 18th century
- Losses until 16% of the total global production
- Fungicides: Can be ineffective or indirectly select resistant strains

## Introduction

### **Scab (*Streptomyces scabies*)**



MSU potato breeding program.

- Soil-borne organism: worldwide distribution
- Persistent pathogen
- Reduce commercial and market price

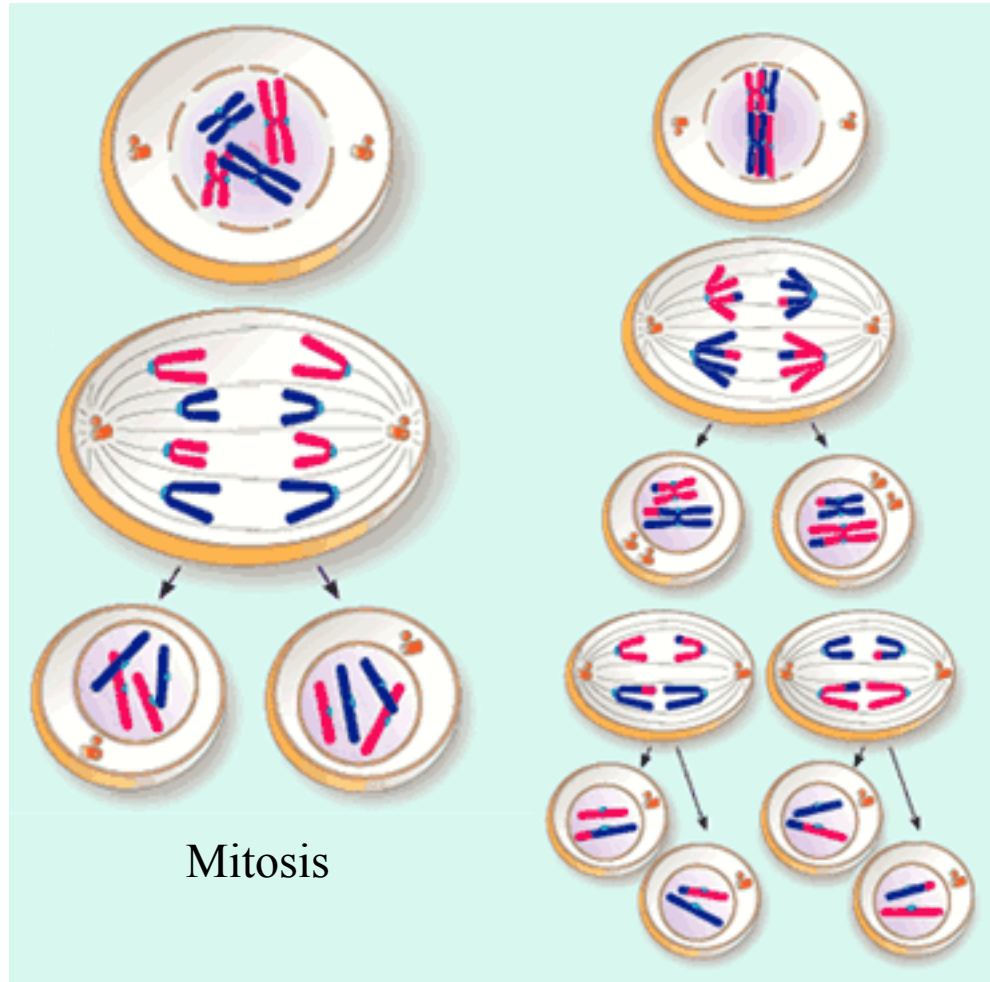
## Objectives

1. Modification of the Coiled-Coil domain from the RB gene in transgenic plants, using genome editing (CRISPR/Cas9)
2. Estimation of Markers effects through Genome Wide Association Analysis and Genome Selection for Late Blight and Scab resistance in a tetraploid potato population

### **Activities:**

- 2.1 To find genomic regions associated with late blight and scab
- 2.2 To evaluate models of genomic selection for late blight and scab resistance

## Why Potato is a challenge?



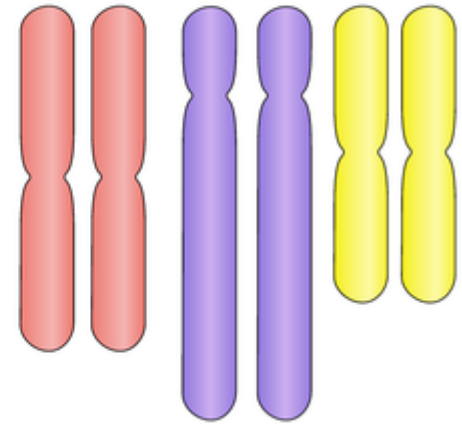
Mitosis

Meiosis

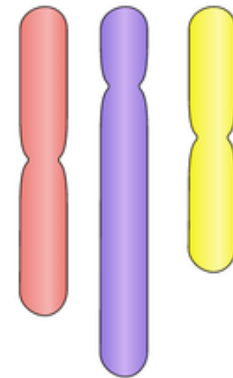
## Why Potato is a challenge?

***Ploidy:*** set of chromosomes in one somatic cells

***Haploid:*** set of chromosomes in gametic cells



Diploid somatic cell:  $2n=2x$



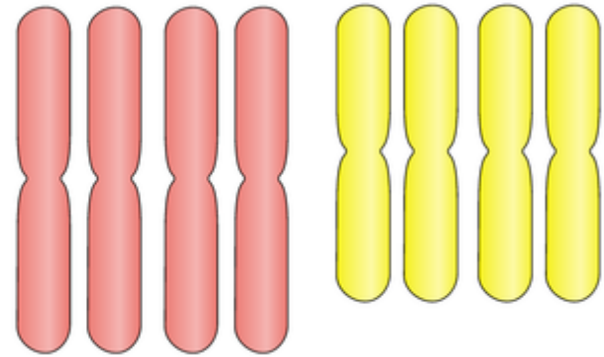
Diploid gametic cell:  $n=x$



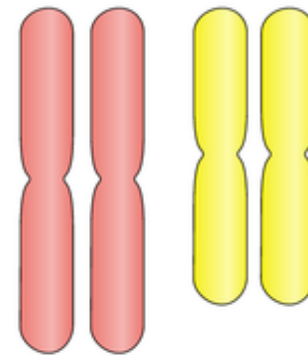
## Why Potato is a challenge?

***Polyploidy***: more than two complete set of chromosomes:

- *Allopolyploid*: inter-specific
- *Autopolyploid*: duplication



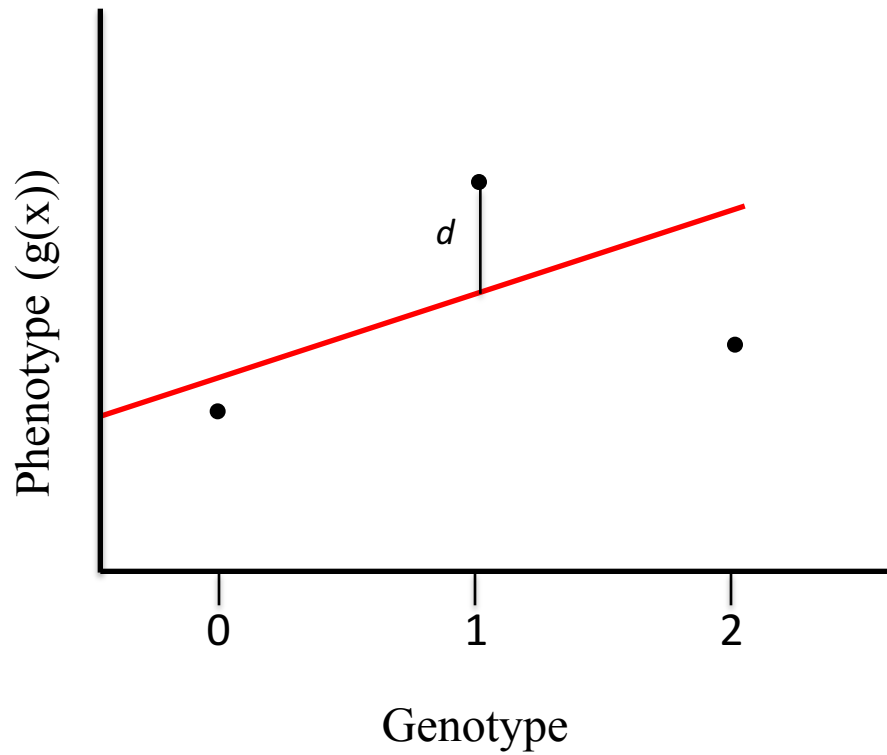
Tetraploid somatic cell:  $2n=4x$



Tetraploid gametic cell:  $n=2x$

## Background

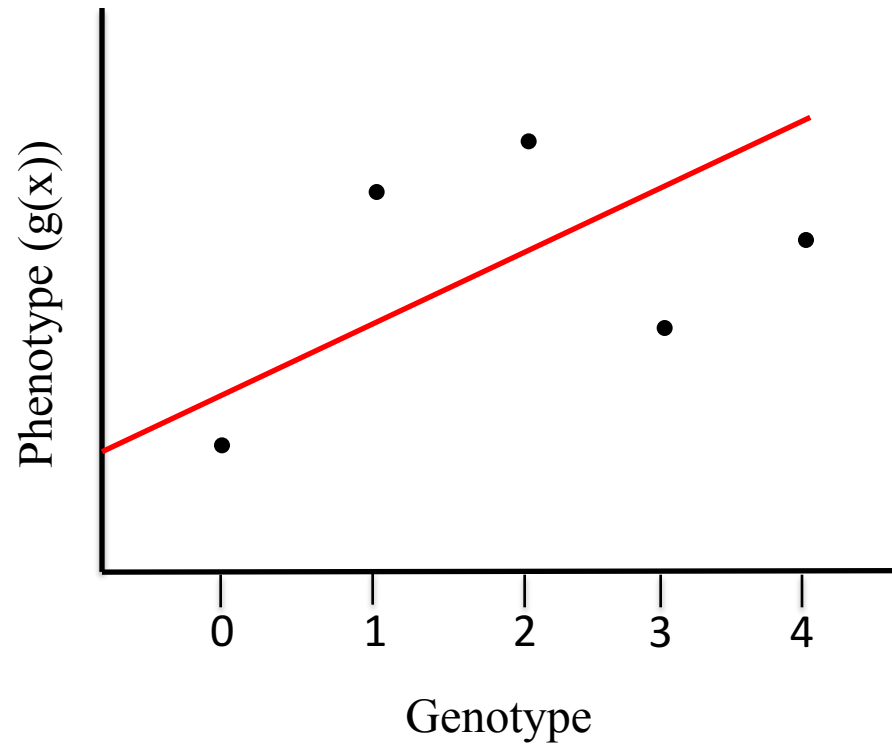
### Diploid model



Gen.	Adi.	Dom.
AA	0	0
Aa	1	1
aa	2	0

## Background

### Tetraploid model

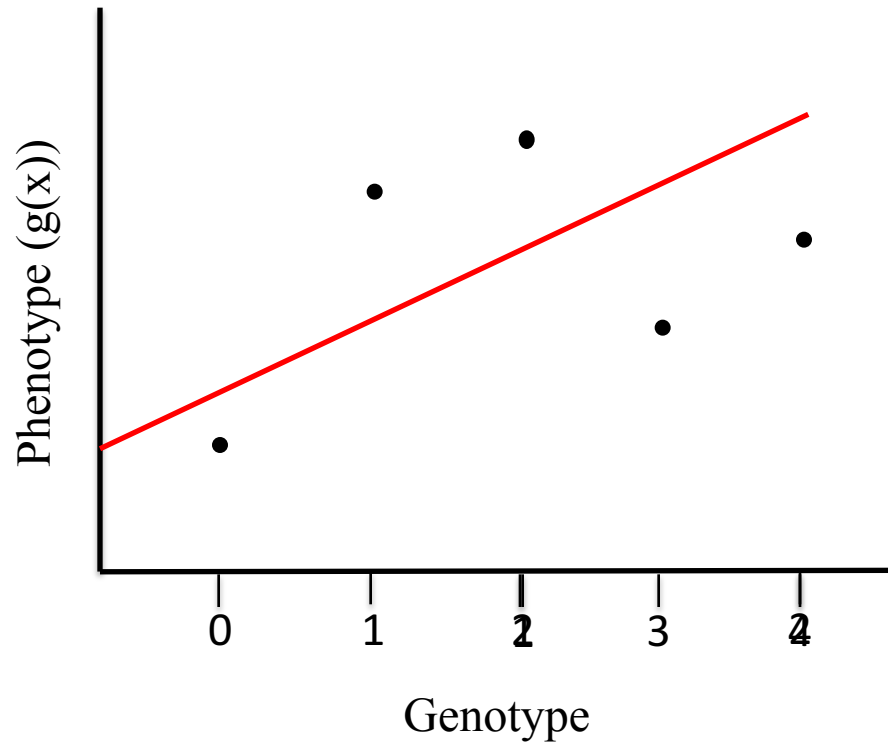


Gen.	Adi.
AAAA	0
AAAB	1
AABB	2
ABBB	3
BBBB	4

## Background

### Diplodised model

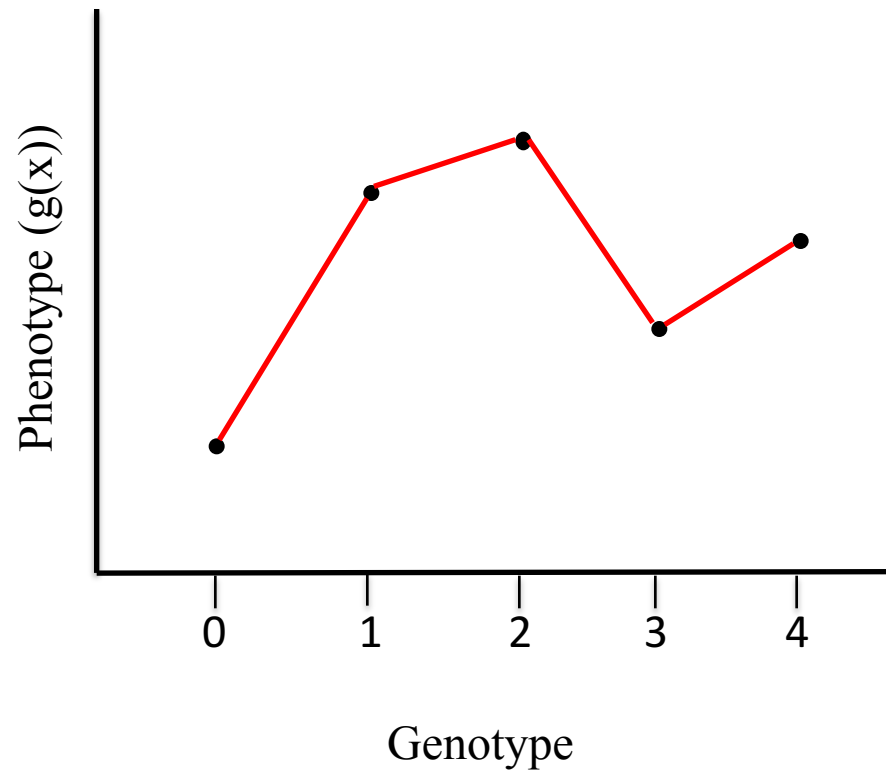
- Collapse heterozygous from tetraploid model



Gen.	Adi.
AAAA	0
AAAB AABB ABBB	1
BBBB	2

Background

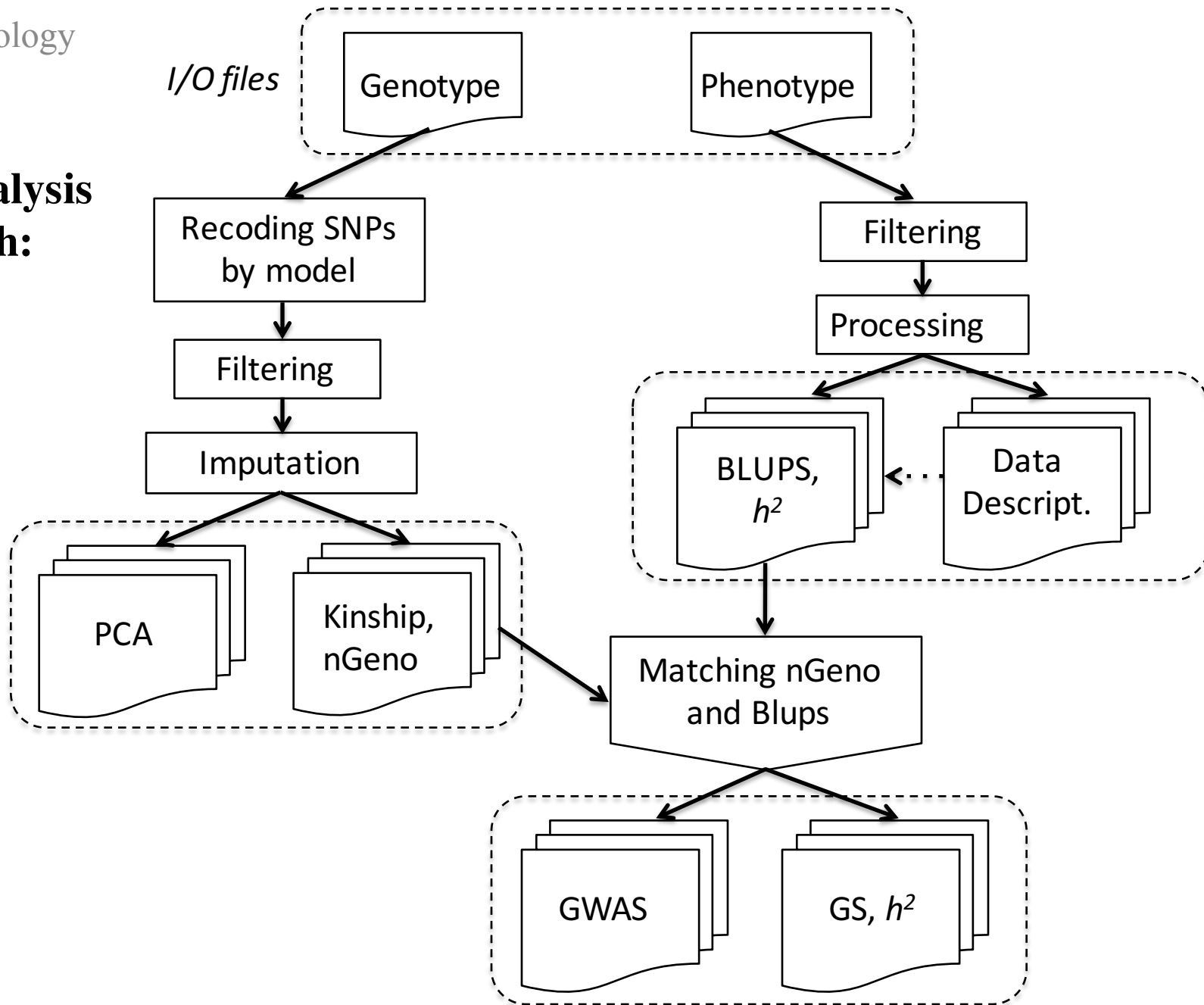
## Step model



	$Z_1$	$Z_2$	$Z_3$	$Z_4$
<b>0</b>	0	0	0	0
<b>1</b>	1	0	0	0
<b>2</b>	1	1	0	0
<b>3</b>	1	1	1	0
<b>4</b>	1	1	1	1

## Methodology

### Data analysis approach: *pipeline*



## Preliminary results

### **Potato data description:**

#### *Genotype:*

- Filtered file with 4859 SNPs coded as a tetraploid
- 185 lines

#### *Phenotype:*

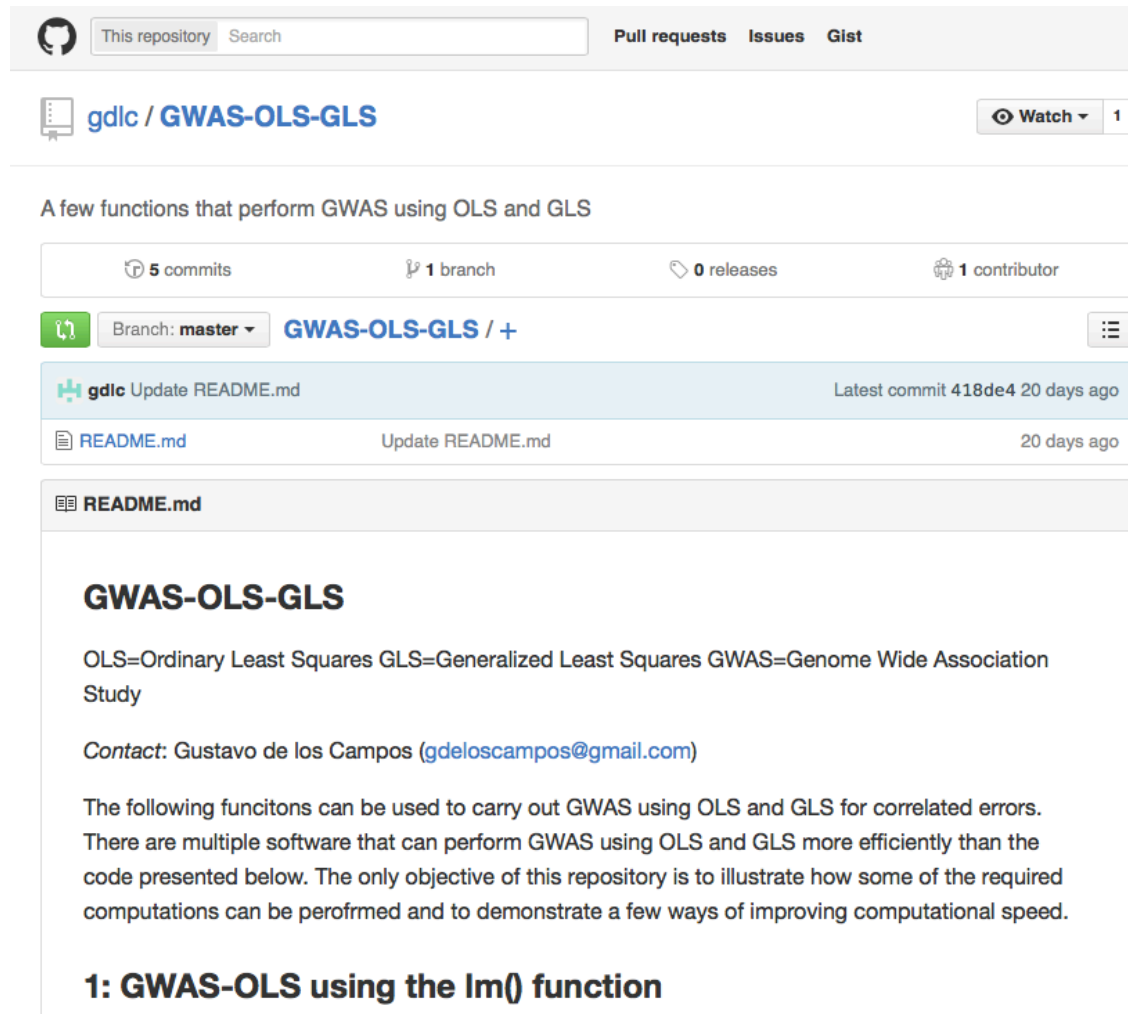
- ~ Five years records for late blight and Scab

#### *Observations :*

- Late blight: ordinal variables (0-9), transformed to RAUDPC
- Scab: ordinal variables (0-5)

Preliminary results

## Data analysis approach: *web-based Git repository*



The screenshot shows a GitHub repository page for 'gdlc / GWAS-OLS-GLS'. At the top, there's a search bar and navigation links for 'Pull requests', 'Issues', and 'Gist'. Below the repository name, there's a 'Watch' button with a dropdown arrow and the number '1'. The description states: 'A few functions that perform GWAS using OLS and GLS'. Below this, there's a summary bar showing '5 commits', '1 branch', '0 releases', and '1 contributor'. A green 'i' icon is next to the 'Branch: master' dropdown. The repository name 'GWAS-OLS-GLS' is followed by a '+' icon. Below the summary bar, there's a commit history table with columns for the commit message, the commit hash, and the time ago. The first commit is 'gdlc Update README.md' with hash '418de4' and '20 days ago'. Below the commit history, there's a section for the 'README.md' file. The README content includes the title 'GWAS-OLS-GLS', a definition of OLS, GLS, and GWAS, contact information for Gustavo de los Campos, and a description of the repository's purpose. The first section of the README is titled '1: GWAS-OLS using the lm() function'.

gdlc / GWAS-OLS-GLS

Watch 1

A few functions that perform GWAS using OLS and GLS

5 commits 1 branch 0 releases 1 contributor

Branch: master GWAS-OLS-GLS / +

Commit Message	Commit Hash	Time Ago
gdlc Update README.md	418de4	20 days ago

README.md

### GWAS-OLS-GLS

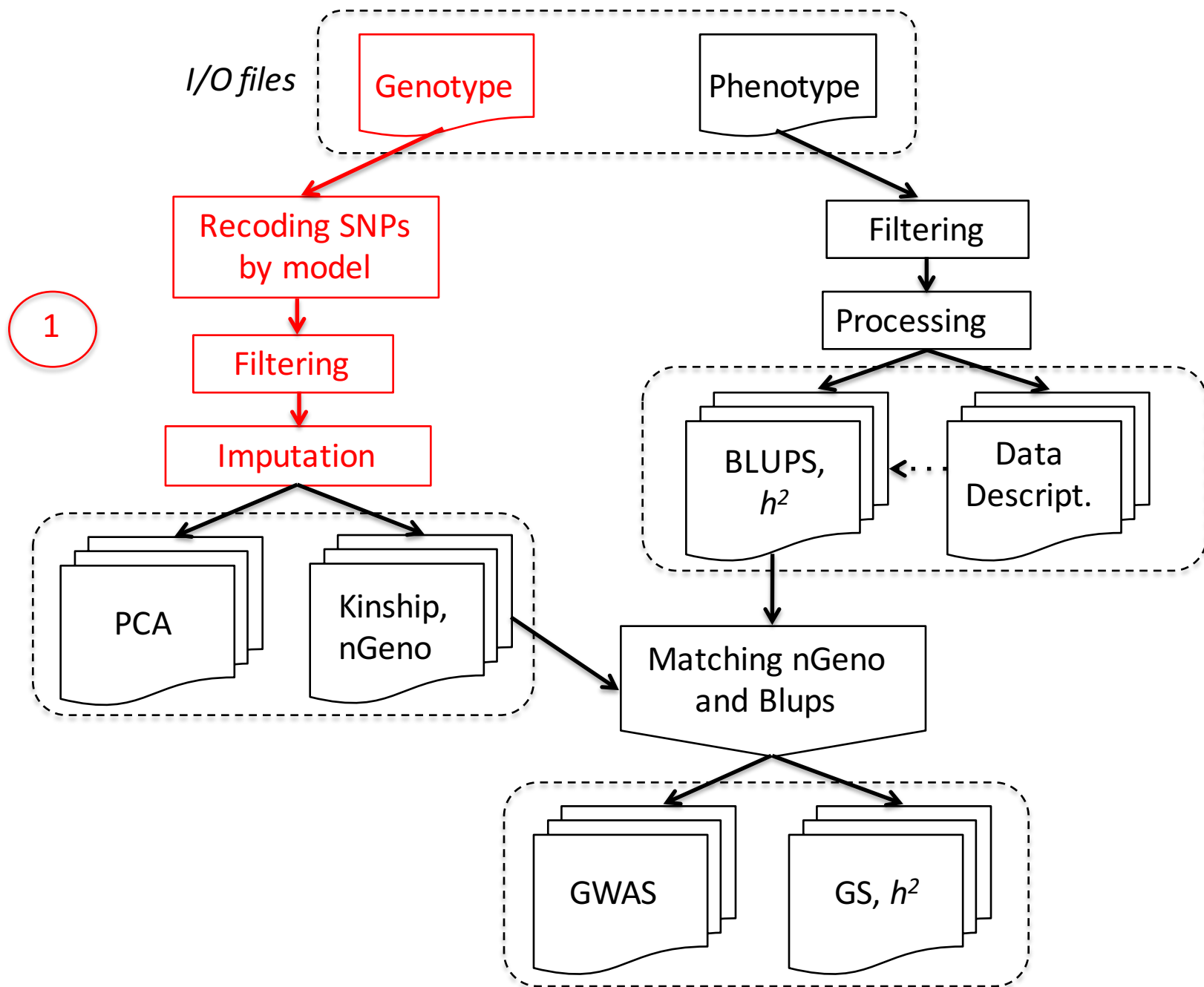
OLS=Ordinary Least Squares GLS=Generalized Least Squares GWAS=Genome Wide Association Study

Contact: Gustavo de los Campos ([gdeloscampos@gmail.com](mailto:gdeloscampos@gmail.com))

The following functions can be used to carry out GWAS using OLS and GLS for correlated errors. There are multiple software that can perform GWAS using OLS and GLS more efficiently than the code presented below. The only objective of this repository is to illustrate how some of the required computations can be performed and to demonstrate a few ways of improving computational speed.

#### 1: GWAS-OLS using the `lm()` function





Preliminary results

1

# Genotype filtering

## Preliminary results

### Genotype recoding

	stub-71	stub-640	stub-10	stub-18	stub-33	stub-35	stub-37	stub-38	stub-41	stub-42
[1,]	"ABBB"	"AABB"	"AABB"	"ABBB"	"AAAB"	"AABB"	"ABBB"	"AABB"	"AAAB"	"AABB"
[2,]	"AABB"	"BBBB"	"AAAA"	"AAAB"	"AAAB"	"AAAB"	"AAAB"	"AAAB"	"AABB"	"BBBB"
[3,]	"AAAB"	"AABB"	"AABB"	"AAAA"	"ABBB"	"AAAB"	"."	"AAAA"	"AABB"	"AAAB"
[4,]	"ABBB"	"ABBB"	"AABB"	"BBBB"	"AABB"	"ABBB"	"AAAB"	"BBBB"	"AABB"	"BBBB"
[5,]	"AABB"	"ABBB"	"ABBB"	"AAAA"	"AABB"	"AAAB"	"BBBB"	"."	"AABB"	"AABB"
[6,]	"AABB"	"AAAB"	"AAAB"	"BBBB"	"AABB"	"ABBB"	"AAAA"	"."	"AABB"	"AABB"
[7,]	"BBBB"	"ABBB"	"BBBB"	"BBBB"	"BBBB"	"BBBB"	"BBBB"	"BBBB"	"BBBB"	"BBBB"
[8,]	"AAAB"	"AAAB"	"AAAA"	"AABB"	"AAAA"	"AAAA"	"AAAB"	"AAAA"	"ABBB"	"AAAB"
[9,]	"ABBB"	"ABBB"	"ABBB"	"ABBB"	"BBBB"	"ABBB"	"AABB"	"BBBB"	"BBBB"	"BBBB"
[10,]	"ABBB"	"BBBB"	"BBBB"	"ABBB"	"ABBB"	"BBBB"	"AABB"	"ABBB"	"ABBB"	"BBBB"

	stub-71	stub-640	stub-10	stub-18	stub-33	stub-35	stub-37	stub-38	stub-41	stub-42
[1,]	"3"	"2"	"2"	"3"	"1"	"2"	"3"	"2"	"1"	"2"
[2,]	"2"	"4"	"0"	"1"	"1"	"1"	"1"	"1"	"2"	"4"
[3,]	"1"	"2"	"2"	"0"	"3"	"1"	"NA"	"0"	"2"	"1"
[4,]	"3"	"3"	"2"	"4"	"2"	"3"	"1"	"4"	"2"	"4"
[5,]	"2"	"3"	"3"	"0"	"2"	"1"	"4"	"NA"	"2"	"2"
[6,]	"2"	"1"	"1"	"4"	"2"	"3"	"0"	"NA"	"2"	"2"
[7,]	"4"	"3"	"4"	"4"	"4"	"4"	"4"	"4"	"4"	"4"
[8,]	"1"	"1"	"0"	"2"	"0"	"0"	"1"	"0"	"3"	"1"
[9,]	"3"	"3"	"3"	"3"	"4"	"3"	"2"	"4"	"4"	"4"
[10,]	"3"	"4"	"4"	"3"	"3"	"4"	"2"	"3"	"3"	"4"

*I/O files*

Genotype

Phenotype

Recoding SNPs  
by model

Filtering

Filtering

Processing

Imputation

BLUPS,  
 $h^2$

Data  
Descript.

2

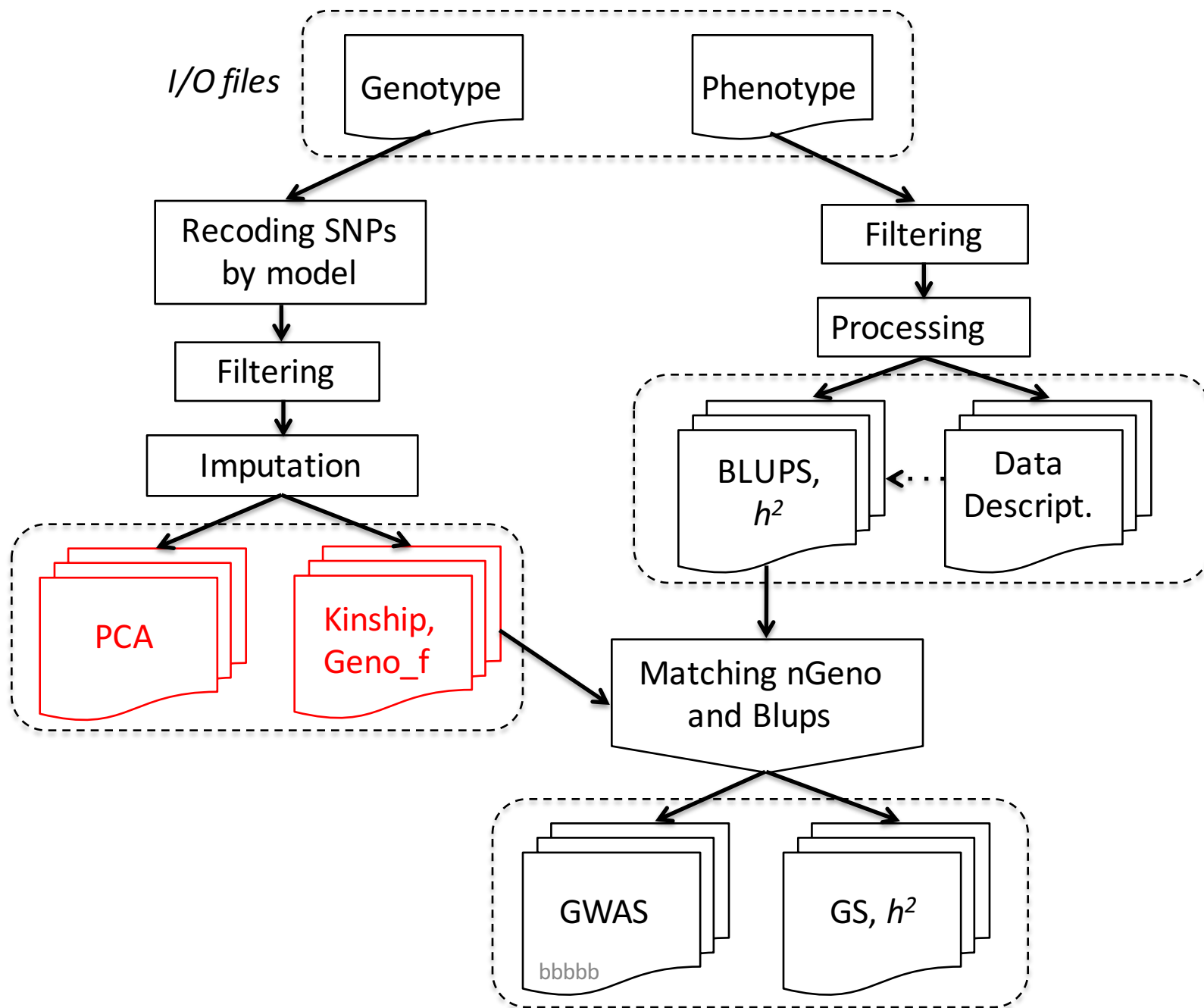
PCA

Kinship,  
Geno\_f

Matching nGeno  
and Blups

GWAS  
bbbbbb

GS,  $h^2$



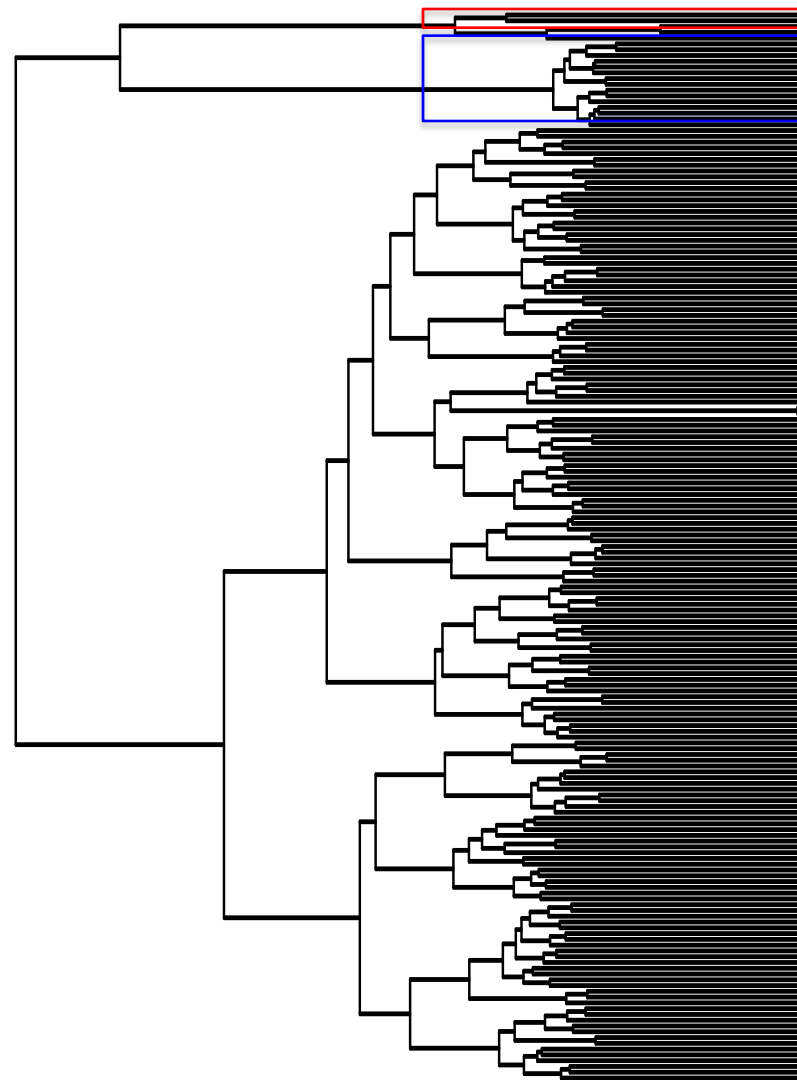
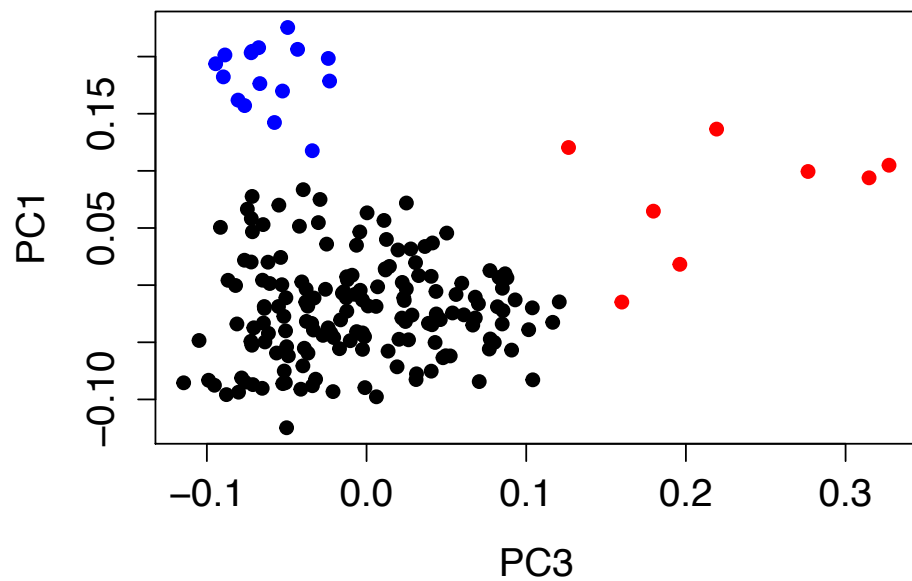
Preliminary results

2

## **Genotype Analysis**

Preliminary results

## Population structure analysis

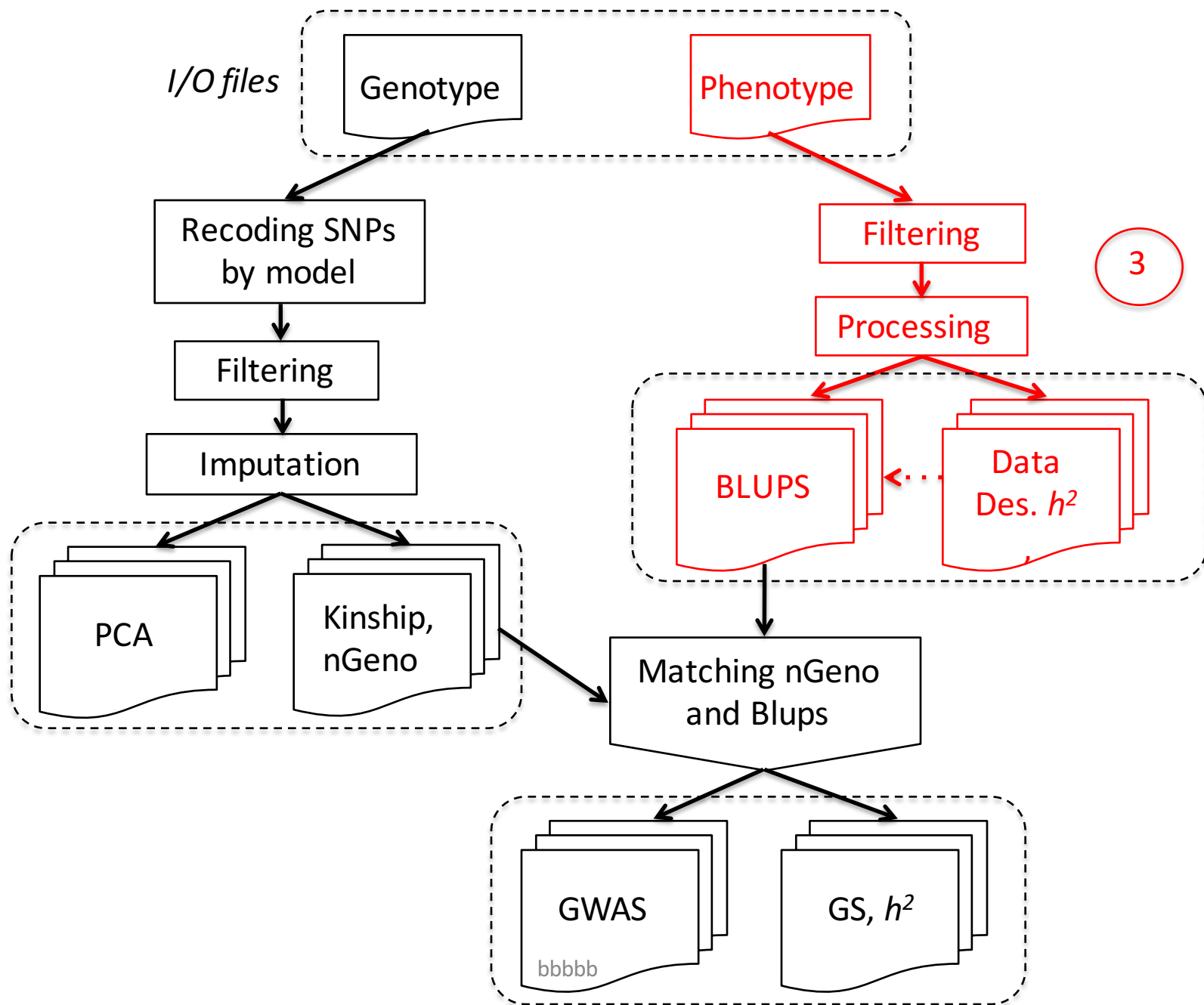


bbbbbb

Euclidian distance

## Variance covariance matrix: G matrix

	stub-71	stub-640	stub-10	stub-18	stub-33	stub-35	stub-37
stub-71	0.89787165	0.086173055	-0.05062016	0.032772548	0.015506595	-0.075634906	0.018645819
stub-640	0.08617305	1.050017555	0.03971775	0.154745944	-0.058620959	0.007746332	-0.006709244
stub-10	-0.05062016	0.039717750	0.98994481	0.101756337	0.276440337	0.086377640	-0.036960244
stub-18	0.03277255	0.154745944	0.10175634	0.896993434	0.005895964	0.031152034	-0.057628145
stub-33	0.01550659	-0.058620959	0.27644034	0.005895964	0.948981053	0.079531064	0.044278701
stub-35	-0.07563491	0.007746332	0.08637764	0.031152034	0.079531064	0.973861052	-0.026634186
stub-37	0.01864582	-0.006709244	-0.03696024	-0.057628145	0.044278701	-0.026634186	0.793946725





3

# Phenotype processing

## Preliminary results

### Phenotype processing: *Data description*

Late blight



Description	
Measure	RAUDPC
Years	5
Lines	416
Total indiv.	1758

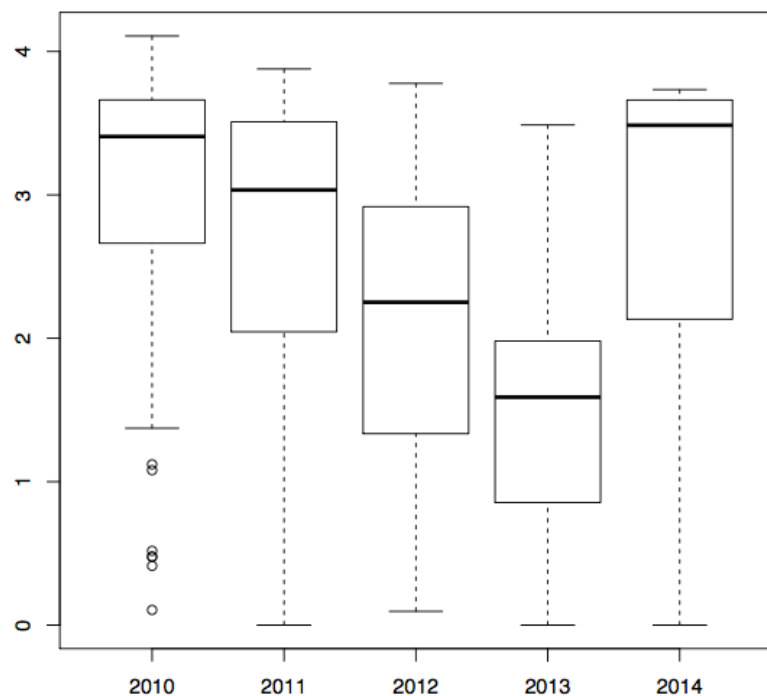
Scab



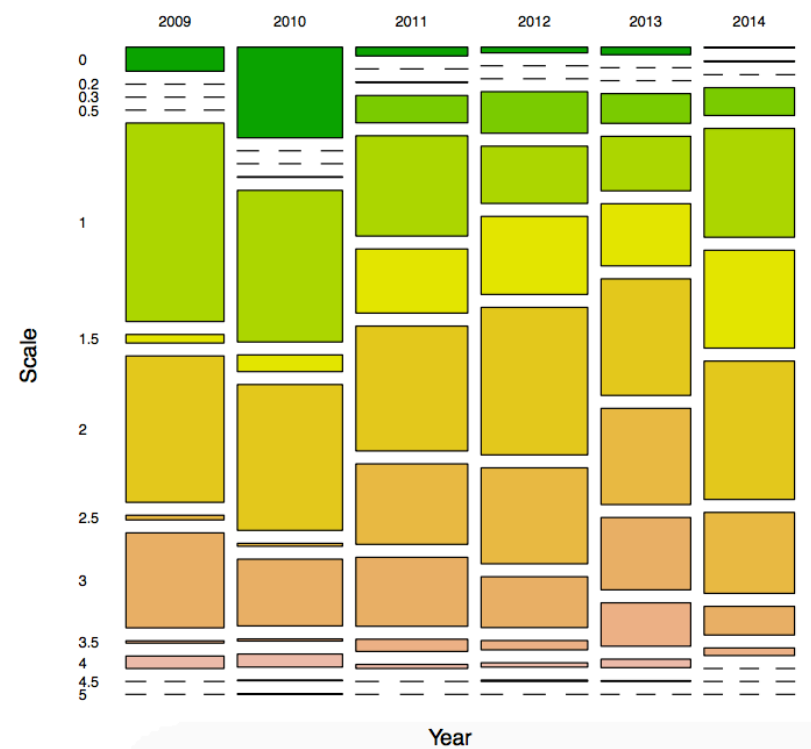
Description	
Measure	Scale (0-4)
Years	6
Lines	548
Total indiv.	3909

## Preliminary results

### Phenotype processing: *Data description*



*Late blight*



*Scab*

## Preliminary results

### Phenotype processing: *BLUPs estimation*

- Fixed model

$$Y = X\beta + Z_1u_1 + Z_2u_2 + \varepsilon$$

Assumptions:

$$u_1 \sim N(\mathbf{0}, \mathbf{I}\sigma_{u1}^2) \quad u_2 \sim N(\mathbf{0}, \mathbf{I}\sigma_{u2}^2) \quad \text{and} \quad \varepsilon \sim N(0, \sigma^2)$$

- *Pre-corrections:*

$$Y = \underline{Trait + Year} + \underline{Line + (Line \times Year)} + error$$

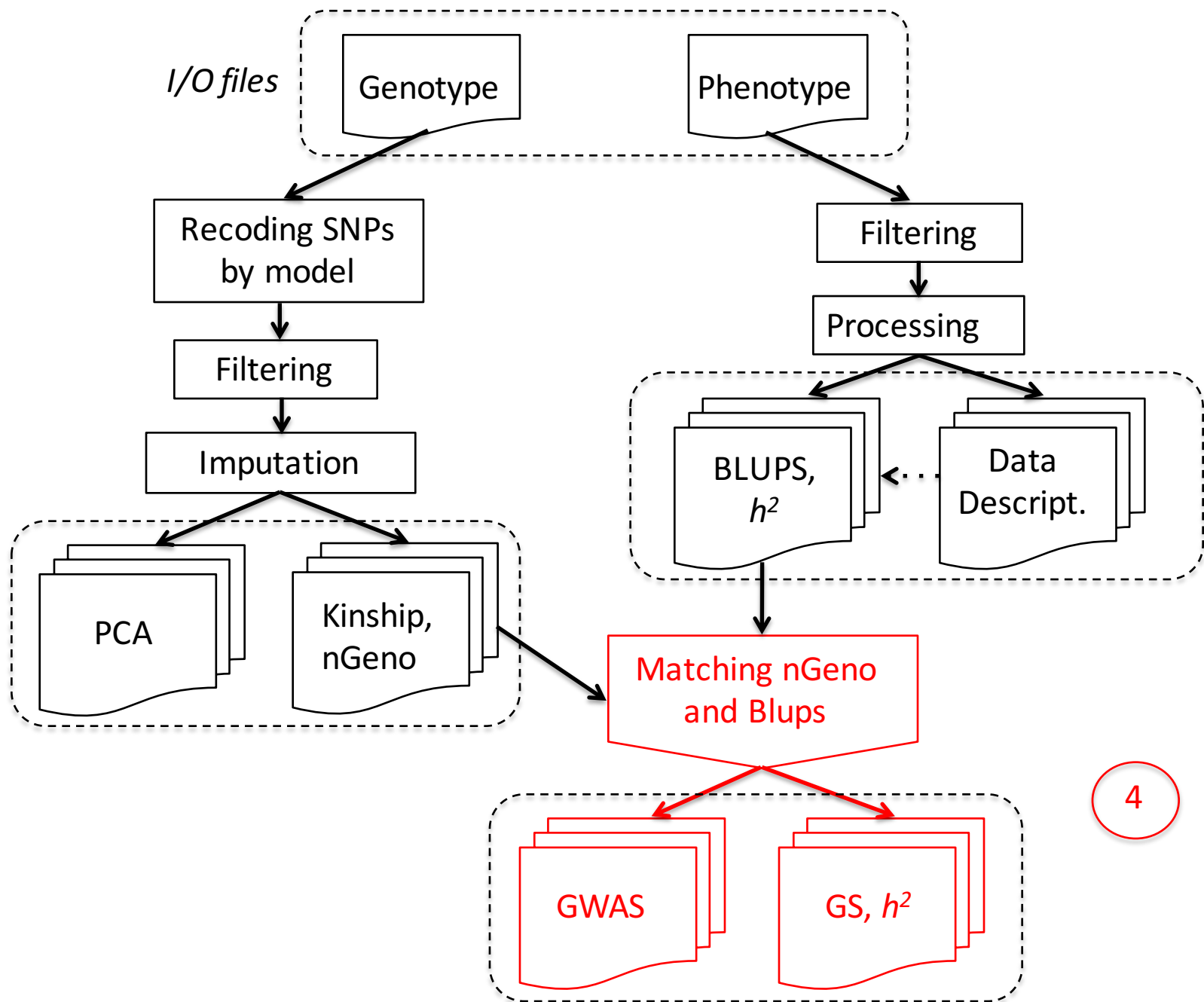
Fixed effects

Random effects

## Preliminary results

### Phenotype processing: *Heritability estimates*

<i>Grups</i>	<i>Variance</i>	
	<i>LB</i>	<i>Scab</i>
Genotype: $\sigma^2(G)$	0.36	0.43
Residual $\sigma^2(E)$	0.13	0.71
Interaction: $\sigma^2(G \times E)$	0.25	0.03
$h^2$	0.59	0.93



Preliminary results

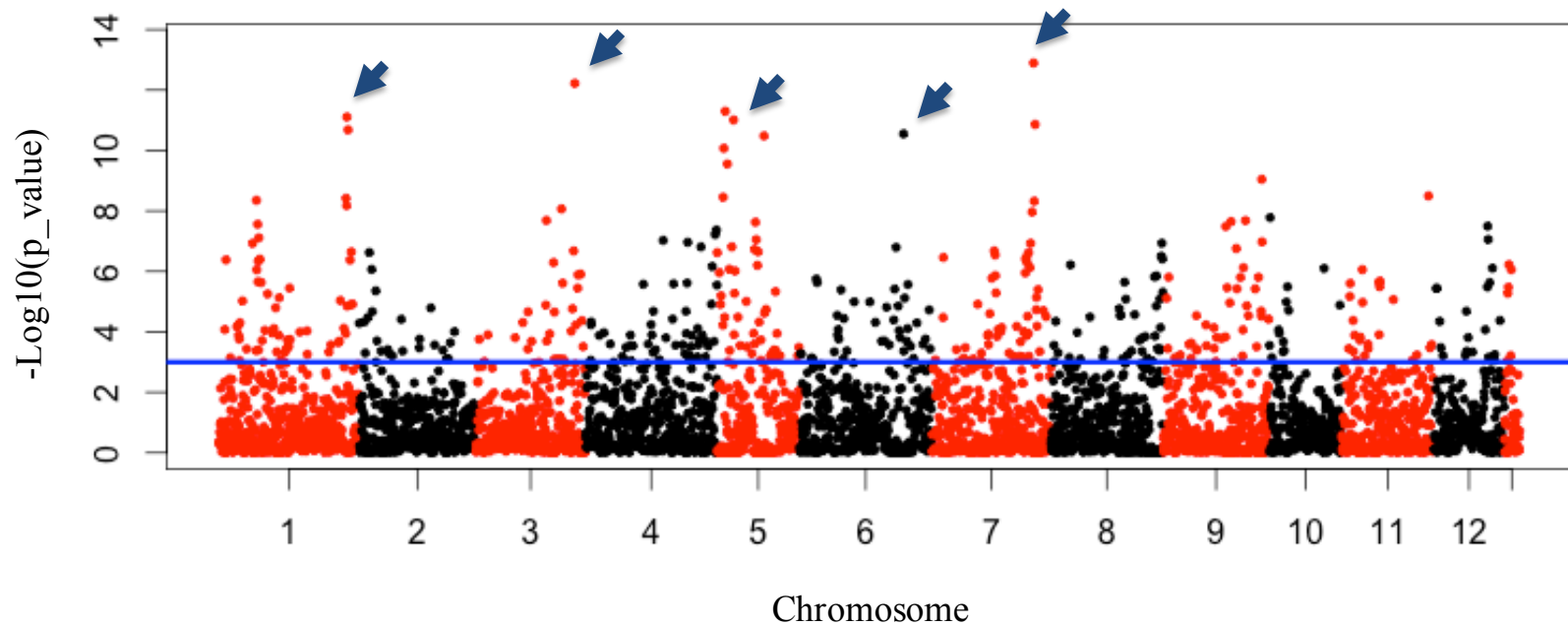
4

## **GWAS and GS**

Preliminary results

## GWAS: OLS

$$\hat{\beta}_j = (X'X)^{-1}X'Y$$



Threshold: P value = 0.001



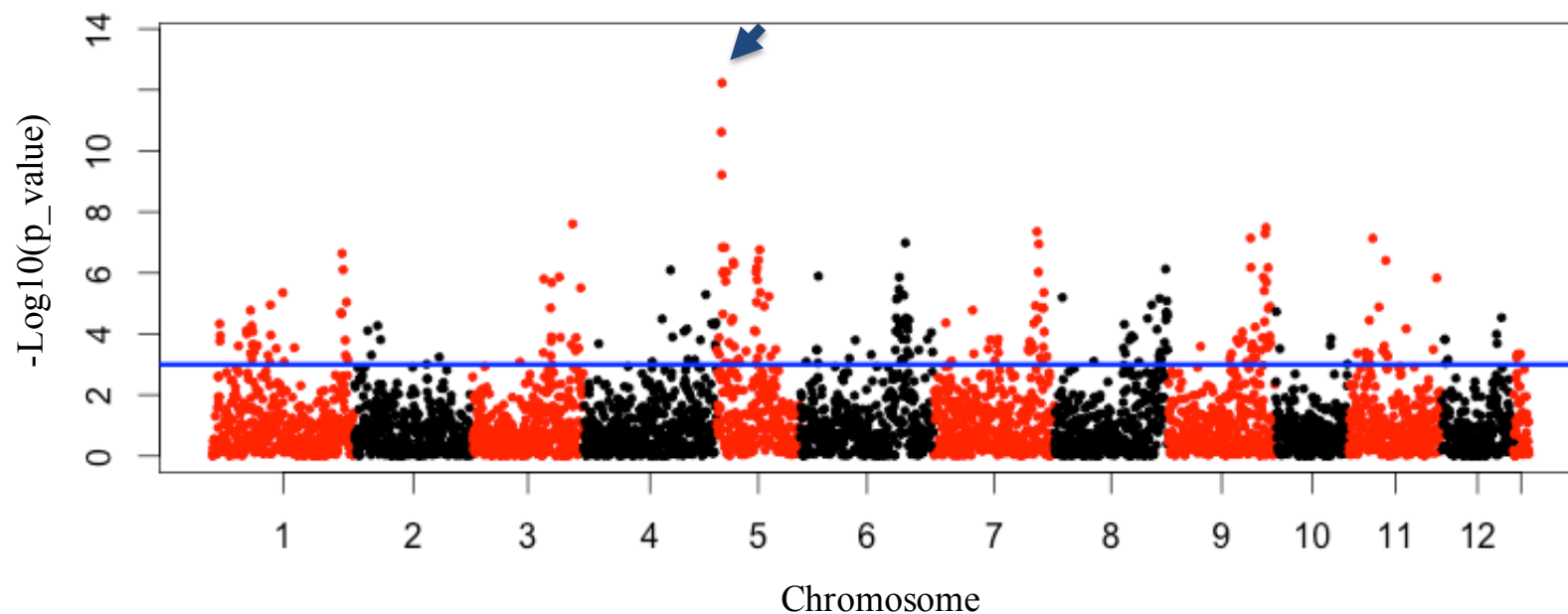
Preliminary results

## GWAS: GLS

$$\hat{\beta}_j = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

where:

$$V = Va(U + \varepsilon) = G\sigma_U^2 + I\sigma_\varepsilon^2$$



Threshold: P value = 0.001

## Preliminary results

### Genome Selection: Prediction accuracy

- *BGLR package: GBLUP*

#### Cross-validation:

Subsampling: 20 individuals randomly selected x 100 times

	<i>Late blight</i>	<i>Scab</i>
Mean Pred. accu.	0.42	0.34

## Next steps

- Increase the number of lines for GWAS/GS:
  - \* ~ 150 new lines from *in vitro* culture bank and potato programs across US
  - \* 156 F1 population (MSL603)
- Continue testing different model for GWAS/GS
- Compare our results with the ones obtained by:
  - \* gwa.R(<https://github.com/steibelj/gwaR>)
  - \* SKAT2 (<https://github.com/lian0090/SKAT2/wiki>)