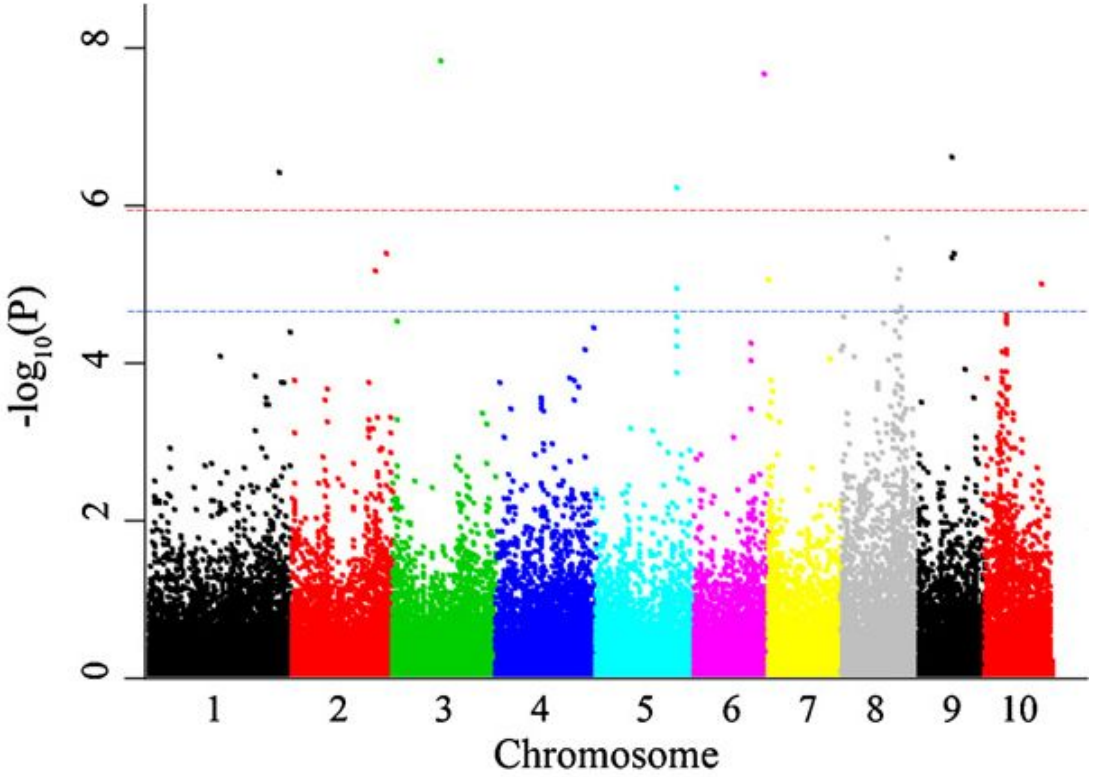
GWAS in plant breeding



Flavio Lozano Isla

Plant breeder

Ing. Agr. MSc.

flavjack@gmail.com

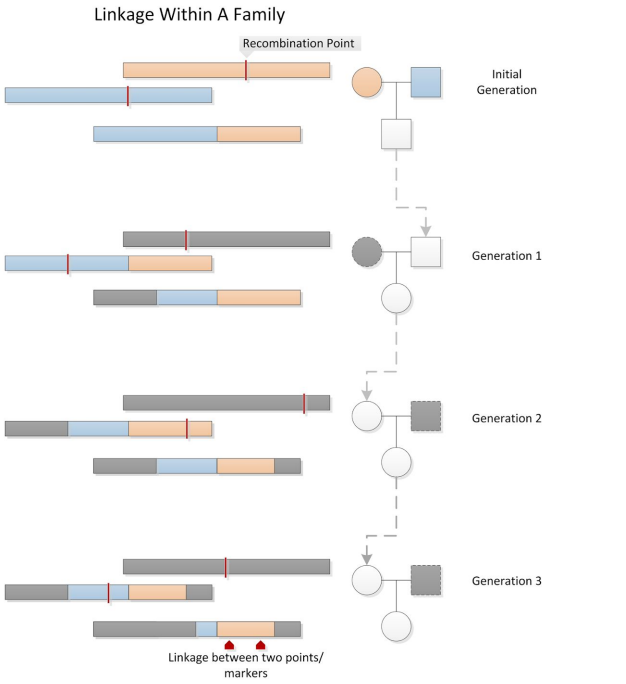
lectures

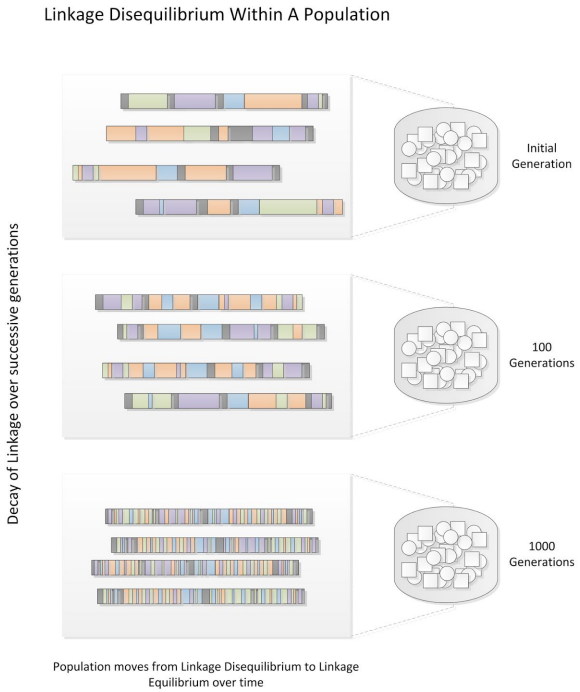
lozanoisla.com

Molecular marker

- Molecular markers are DNA sequences in the genome which can be detected using suitable techniques.
- A marker is a segment of DNA whose inheritance can be followed.
- DNA markers are used to detect polymorphism between individuals on the DNA level

Linkage and Linkage Disequilibrium

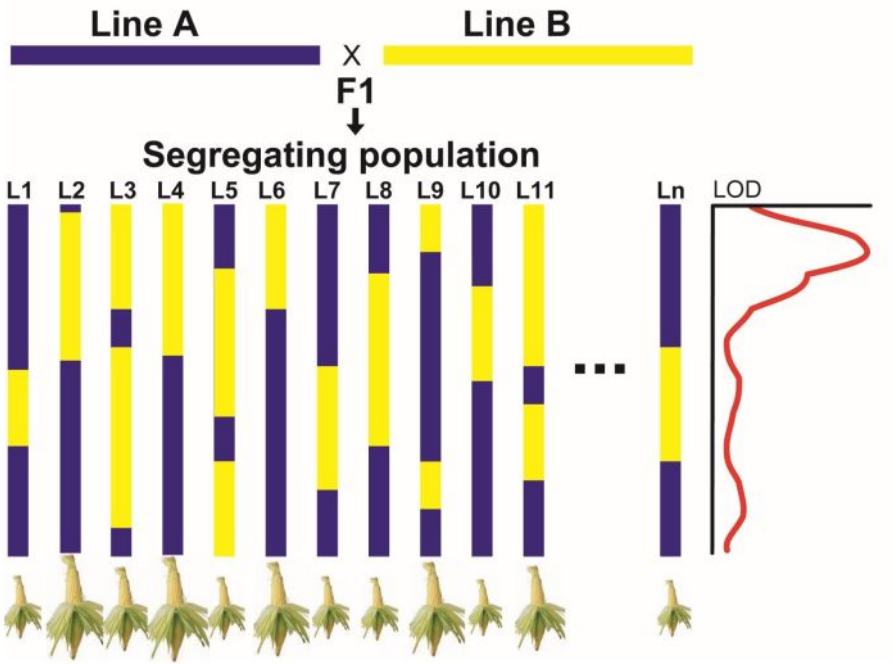




Bush WS, Moore JH (2012)

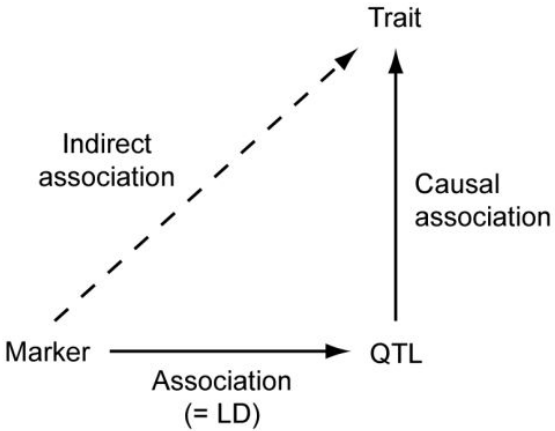
- W/in Family:
  Two genetics markers liked on a chromosome
- In a population:
  LD → LE (more recombination events)
- Linkage disequilibrium (LD) decays with time by recombination

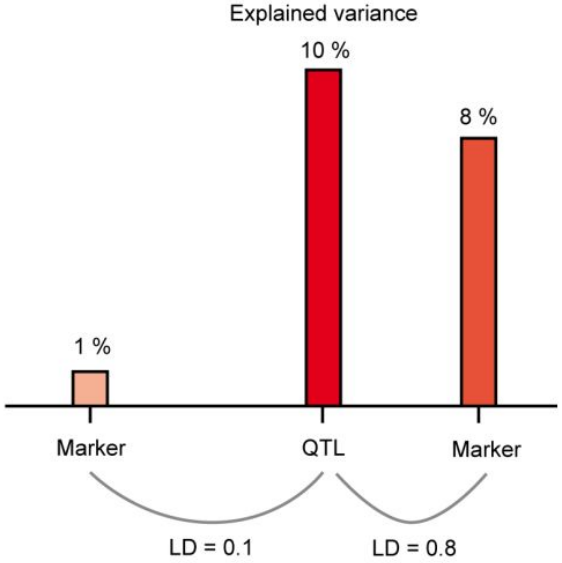
Quantitative Trait Loci (QTL)



- QTL are of interest for plant breeding for marker assisted selection (MAS)
- QTL detection power and mapping resolution depend on LD
- Limitations:
    - F2 or RIL population
    - Limit in mapping resolution

Marker vs Trait





- High LD required to detect medium or small QTL
- Indirect association between Trait & Marker
- LD is always population-specific

Single nucleotide polymorphisms (SNPs)

- SNPs and short insertions/deletions (InDels) are the most abundant source of DNA polymorphisms in a genome
- SNPs are biallelic in most cases

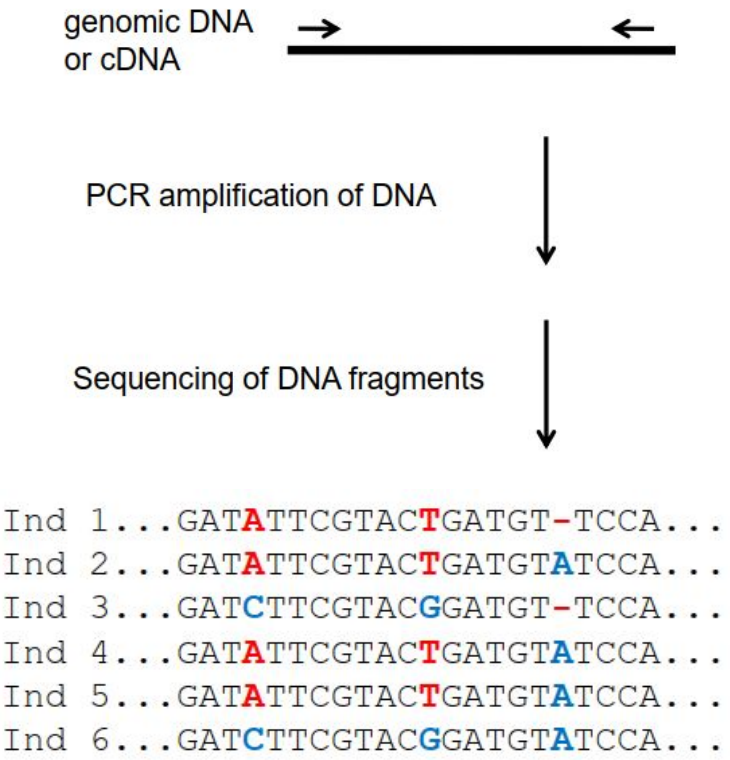
```
SNP C/A:
Allele 1: ACATGGATCCAGCCATGGCTTAG
          TGTACCTAGGTCGGTACCGAATC

Allele 2: ACATGGATCCAGACATGGCTTAG
          TGTACCTAGGTCTGTACCGAATC
```

```
InDel:
ACATGGATCCAGCCATGGCTTAG
TGTACCTAGGTCGGTACCGAATC

ACATGGATC------TGGCTTAG
TGTACCTAG------ACCGAATC
```

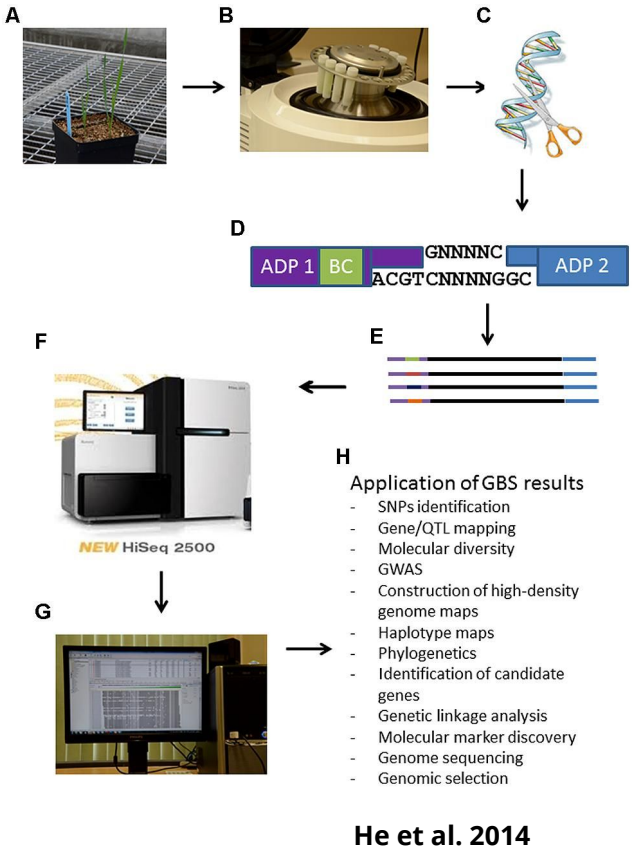
SNP discovery



- SNP is identified when a nucleotide from an accession read differs from the reference genome at the same nucleotide position
- In the absence of a reference genome → comparing reads from different genotypes using de novo assembly strategies

Genotyping by sequencing (GBS)

- GBS is a simple, highly multiplexed system for constructing libraries for next-generation sequencing
- GBS Methods:
  - GBS – Keygene and Cornell → very often used
  - DARTseq → many breeders use it
  - RAD-seq → often used in diversity studies
- GBS is a novel application of NGS protocols for discovering and genotyping SNPs in crop genomes and populations.



He et al. 2014

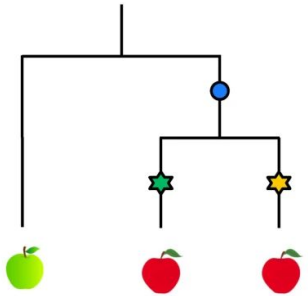
Genome-Wide Association Studies

GWAS

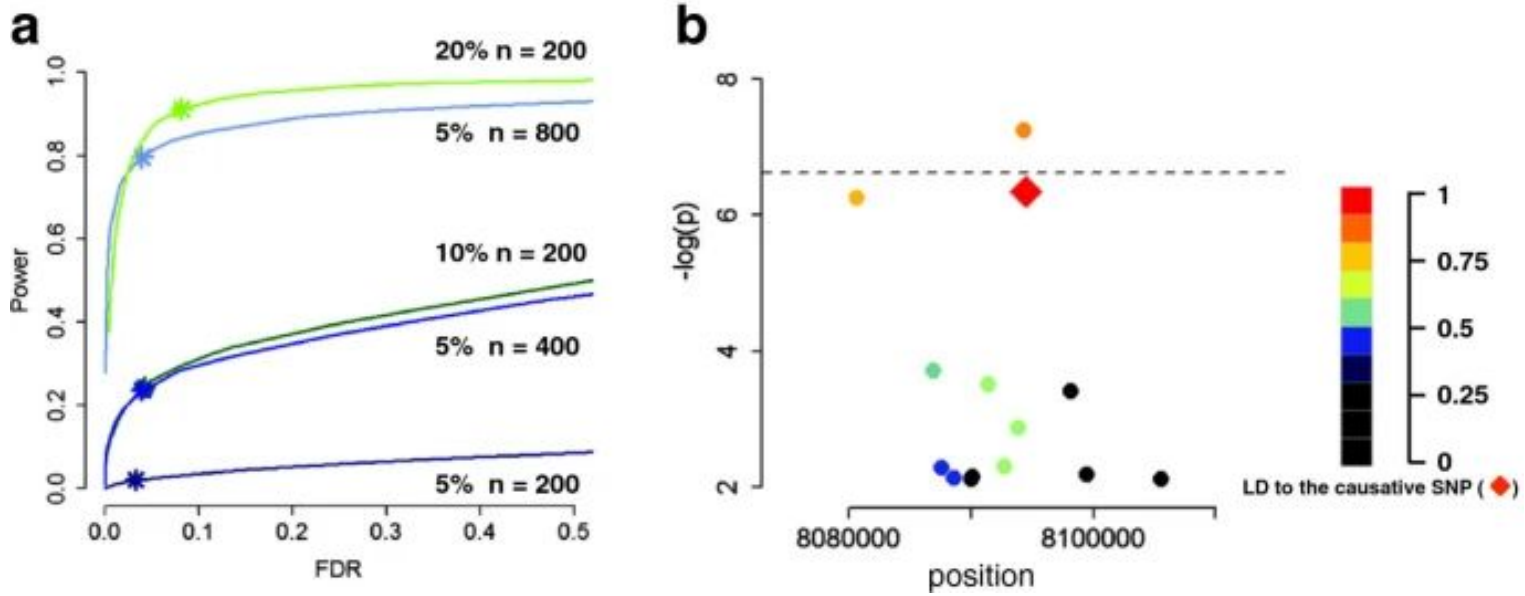
- GWAS enabled the mapping of genomic loci associated with economically important traits (yield, resistance to biotic and abiotic stresses, and quality)
- Used to perform marker-assisted selection (MAS) in breeding programs and discover genes underlying phenotypic variation

Considerations

- Sample size (> 100 accession)
- Mapping panel condition
- Statistical approach
- Genetic architecture

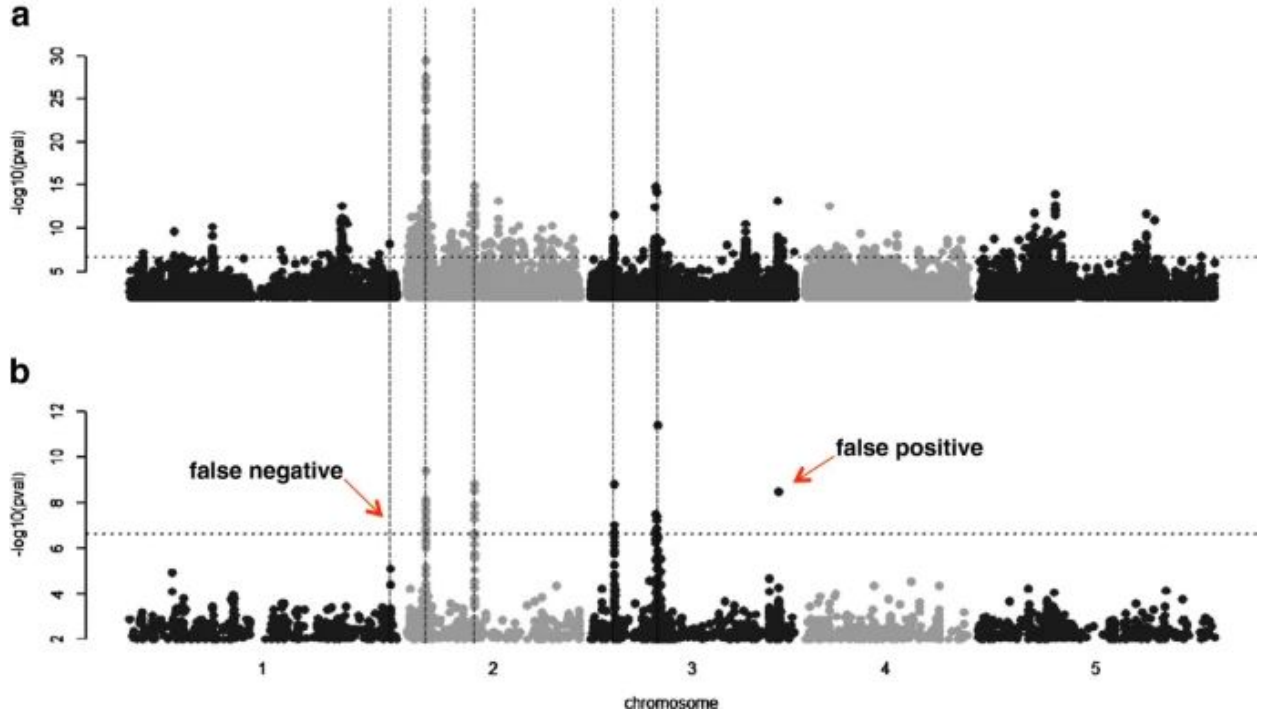


Sample size & Genetic architecture



Simulations in which a single random SNP explaining 5%, 10% or 20% of the phenotypic variance with heritability ~0.75

Statistical approach



a) Linear model

b) Mixed model

- Include population structure
- Mixed model dramatically reduces inflation of p -values

GWS

REVIEW ARTICLE

Front. Genet., 05 June 2020 | https://doi.org/10.3389/fgene.2020.00447

Recommendations for Choosing the Genotyping Method and Best Practices for Quality Control in Crop Genome-Wide Association Studies

 Stefano Pavan[1,2*], Chiara Delvento[1], Luigi Ricciardi[1], Concetta Lotti[3],  Elena Ciani[4] and  Nunzio D'Agostino[5*]

[1]Department of Soil, Plant and Food Science, Section of Genetics and Plant Breeding, University of Bari Aldo Moro, Bari, Italy

[2]Institute of Biomedical Technologies, National Research Council (CNR), Bari, Italy

[3]Department of Agricultural, Food and Environmental Sciences, University of Foggia, Foggia, Italy

[4]Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari Aldo Moro, Bari, Italy

[5]Department of Agricultural Sciences, University of Naples Federico II, Naples, Italy

GWS: genotyping methods

Three categories

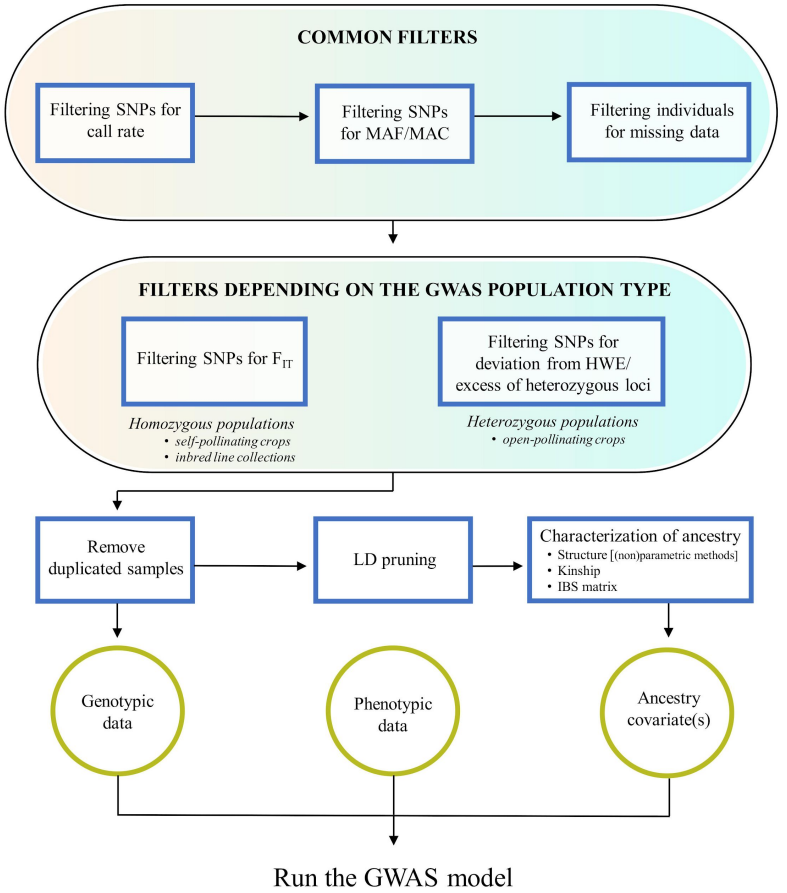
- Next-generation sequencing (NGS)
  - whole genome resequencing (WGR)
  - reduced representation sequencing (RRS)
- SNP arrays

Selection

- WGR is especially desirable for GWAS populations displaying rapid LD decay
- WGR: roughly costs $400 per sample for a genome of 1 Gb and 10× average sequencing depth
- RRS consists in sequencing only a small fraction of the genome → GBS, RADseq, ddRADseq
- RSS: $35 per sample independently from the genome size

Quality control procedures for crop GWAS

- Common filters:
  - Minor allele frequency (MAF)
  - Minor allele count (MAC)
- Filter by pop type
  - Homozygous
  - Heterozygous
- Sample duplication / Ancestral relation (STRUCTURE - ADMIXTURE)



Bioinformatic tools

Conversion of genotypic data

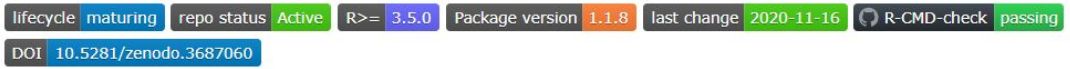
- Variant call format (VCF)
- haplotype map (hapmap)
- pedigree/map (ped/map)
- binary (bed/bim/fam)
- Affymetrix chip (chp)

> Tools:

- PGDSpider
- TASSEL
- PLINK

QC tools for filtering

- VCFtools



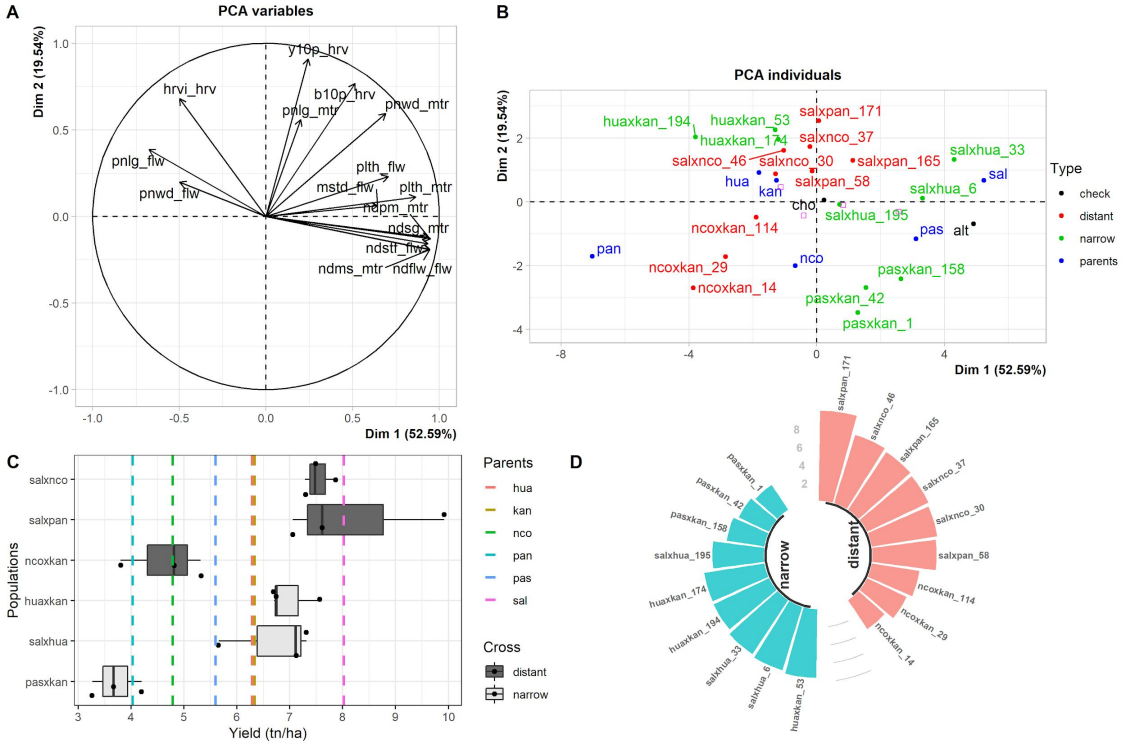
radiator: an R package for RADseq Data Exploration, Manipulation and Visualization

Most genomic analysis look for patterns and trends with various statistics. Bias, noise and outliers can have bounded influence on estimators and interfere with polymorphism discovery. Avoid bad data exploration and control the impact of filters on your downstream genetic analysis. Use radiator to: import, explore, manipulate, visualize, filter, impute and export your GBS/RADseq data.

radiator is designed and optimized for fast computations of diploid data using Genomic Data Structure GDS file format and data science packages in tidyverse. **radiator** handles VCF files with millions of SNPs and files of several GB.

Quinoa GWS

Phenotyping



- MET trials using LMM
- Calculation of H^2 of 19 traits
- Genotype selection
- Traits for direct and indirect selection
- First manuscript

Quinoa GWS

Genotyping

Description	Library	Crop	Genome (Gb)	Depth	Price (Gb)	Samples	Total (Gb)	Total (€)
Simple crosses parents	1	Quinoa	1.5	20	€10.00	6	180	€1,800.00
Simple crosses population 1	2	Quinoa	1.5	0.5	€10.00	90	67.5	€675.00
Simple crosses population 2	3	Quinoa	1.5	0.5	€10.00	90	67.5	€675.00
Simple crosses population 3	4	Quinoa	1.5	0.5	€10.00	90	67.5	€675.00
Simple crosses population 4	5	Quinoa	1.5	0.5	€10.00	90	67.5	€675.00
Simple crosses population 5	6	Quinoa	1.5	0.5	€10.00	90	67.5	€675.00
Simple crosses population 6	7	Quinoa	1.5	0.5	€10.00	90	67.5	€675.00
								€5,850.00

- Quinoa genome ~ 1.5 Gb
- Sequencing service in China using GBS
- Six population from Six parents
- Phenotypic data from MET experiments
- GWS analysis at 2021

Advances

OPEN ACCESS PEER-REVIEWED

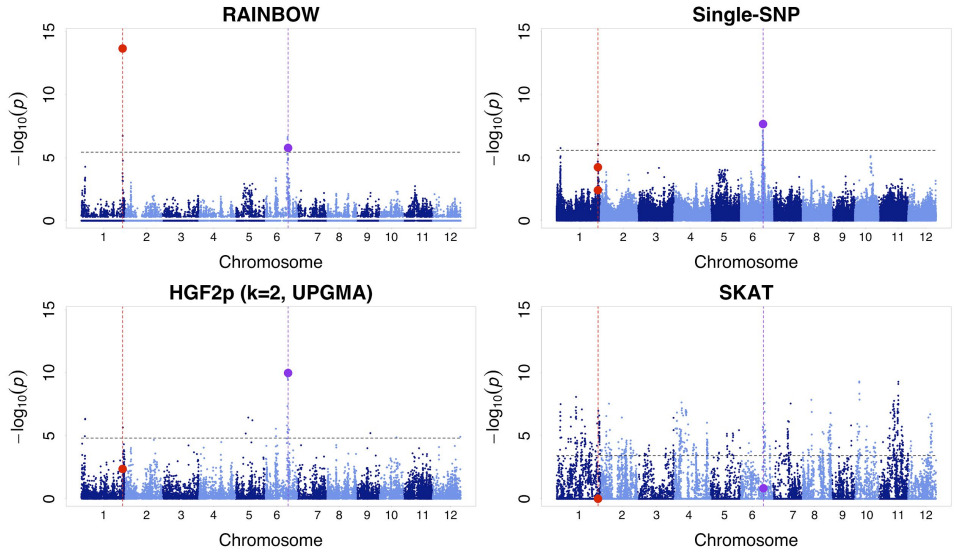
RESEARCH ARTICLE

RAINBOW: Haplotype-based genome-wide association study using a novel SNP-set method

Kosuke Hamazaki, Hiroyoshi Iwata

Version 2 Published: February 14, 2020 • https://doi.org/10.1371/journal.pcbi.1007663
• >> See the preprint

Iteration 48



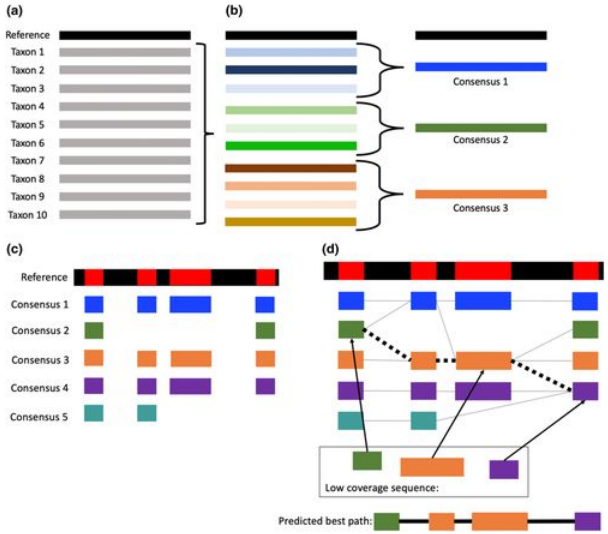
The Plant Genome OPEN ACCESS

ORIGINAL RESEARCH Open Access

A sorghum practical haplotype graph facilitates genome-wide imputation and cost-effective genomic prediction

Sarah E. Jensen, Jean Rigaud Charles, Kebede Muleta, Peter J. Bradbury, Terry Casstevens, Santosh P. Deshpande, Michael A. Gore, Rajeev Gupta, Daniel C. Ilut, Lynn Johnson ... See all authors

First published:25 March 2020 | https://doi.org/10.1002/tpg2.20009



Advances

About BGI Genomics

BGI Genomics is the world's leading provider of genomic sequencing services and proteomic services, now serving customers in more than 66 countries. We provide academic institutions, pharmaceutical companies, health care providers and other organizations with integrated genomic sequencing and proteomic services and solutions across a broad range of applications spanning:

- Basic research covering human, plant, animal and microbial species
- Clinical research in human health
- Genetic testing and screening
- Drug discovery and development
- Agriculture and Biodiversity preservation and sustainability

Project Workflow

We care for your samples from the start through to the result reporting. Highly experienced laboratory professionals follow strict quality procedures to ensure the integrity of your results.

Biotechnology / DNA testing

China's BGI says it can sequence a genome for just $100

Super-cheap DNA sequencing could boost cancer screening, prenatal tests, and research into population genetics.

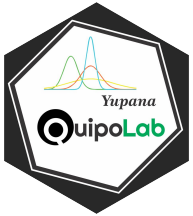
by **Antonio Regalado** February 26, 2020

Using technology originally acquired in the US, the Chinese gene giant BGI Group says it will make genome sequencing cheaper than ever, breaking the $100 barrier for the first time.

Inkaverse

Tarpuy

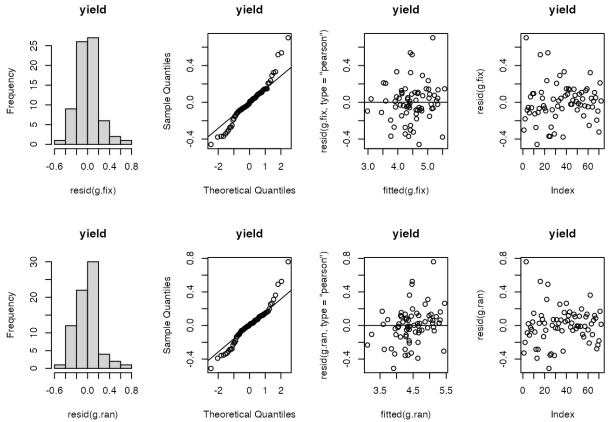
Yupana



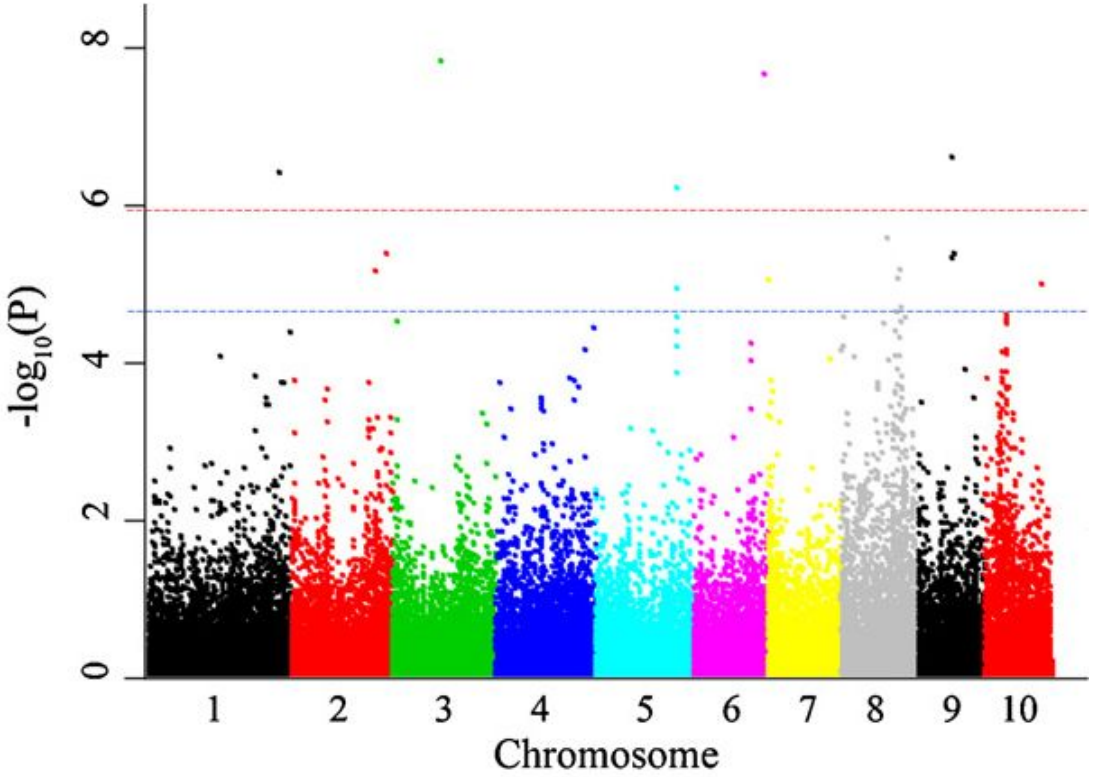
inti: Tools and Statistical Procedures in Plant Science

The 'inti' package is part of the 'inkaverse' project for developing different procedures and tools used in plant science and experimental designs. The mean aim of the package is to support researchers during the planning of experiments and data collection (tarpuy()), data analysis and graphics (yupana()) , and technical writing (rtciles()). Learn more about the 'inkaverse' project at <https://inkaverse.com/>.

Version:	0.1.2
Depends:	shiny, ggplot2, dplyr, tidyr, tibble
Imports:	lme4, agricolae, FactoMineR, purrr, stringr, DT, emmeans
Suggests:	agridat, googlesheets4, knitr, rmarkdown
Published:	2020-11-25
Author:	Flavio Lozano-Isla [aut, cre]
Maintainer:	Flavio Lozano-Isla <flavjack at gmail.com>
BugReports:	https://github.com/flavjack/inti/issues/



GWAS in plant breeding



Flavio Lozano Isla

Plant breeder

Ing. Agr. MSc.

flavjack@gmail.com

lectures

lozanoisla.com