Brief introduction to GWAS

Overview of its advantages and caveats

-

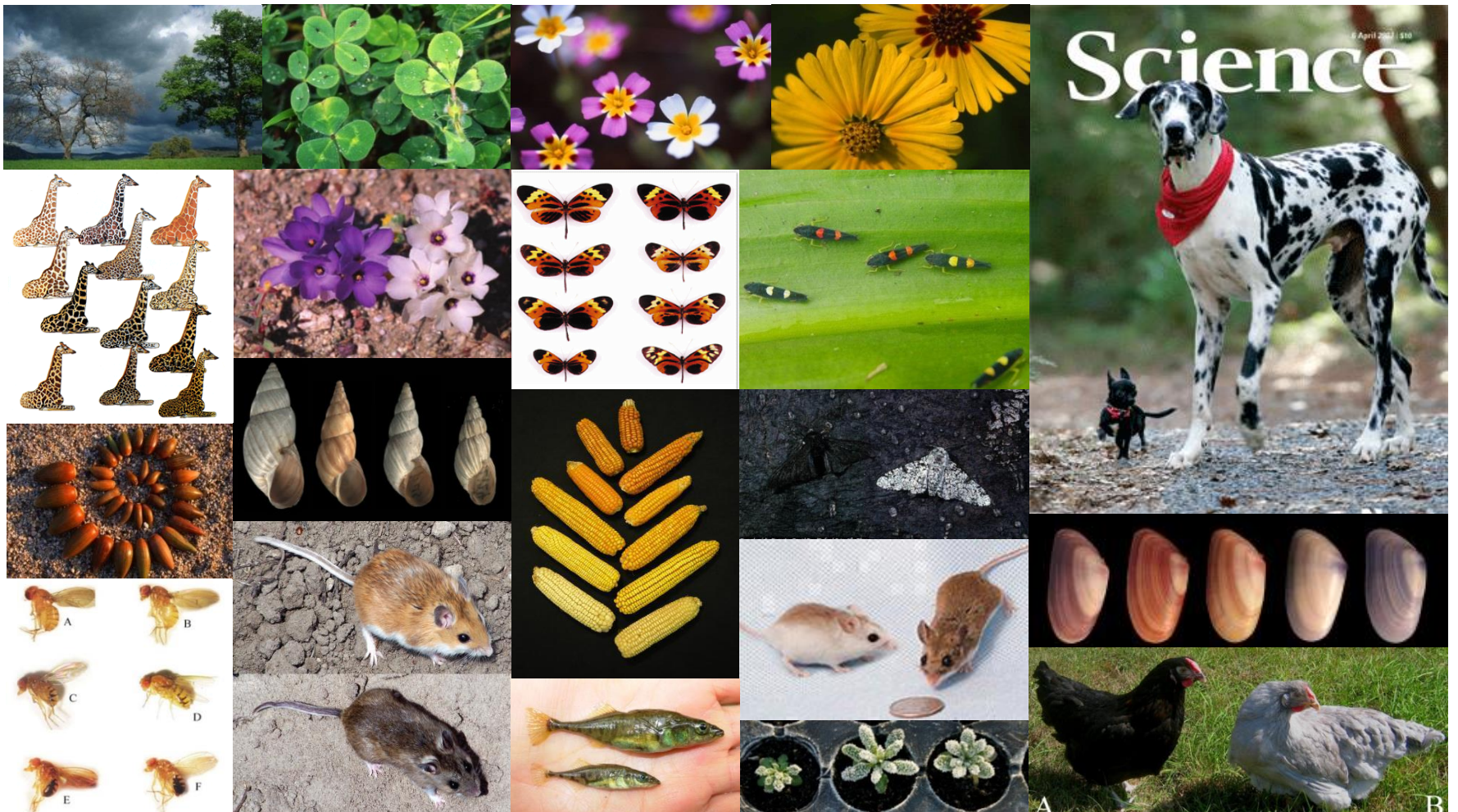
Population Genomics Workshops 2017 @ Sheffield

-

Romain Villoutreix

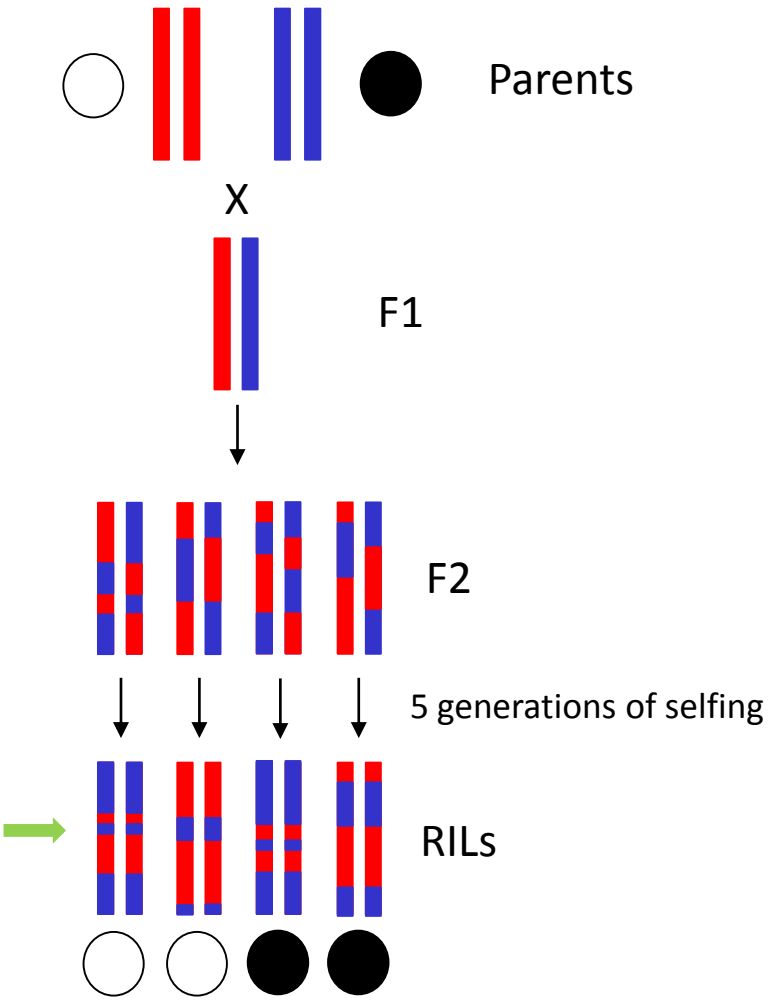
r.villoutreix@sheffield.ac.uk

Research Associate (Nosil lab)



Genetic architecture ? (Which genes, number of genes, gene function …)

QTL mapping: example of design



Advantage:

'low number' of marker needed

Problems:

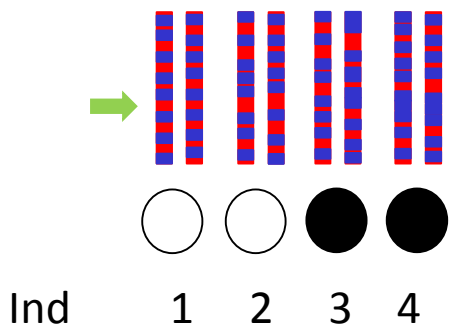
Take time to make crosses (ex: 2-3 years in *A. thaliana* with short generation time)

Discovery of allele not 'evolutionary meaningful' = rare variants

Lack of recombination = QTL are usually huge genomic regions ~ 200 genes

→ GWAS (Genome-wide association study)

GWAS: making use of recombination occurring in natural populations



Advantages:

Smaller genomic regions identified
No need for crosses

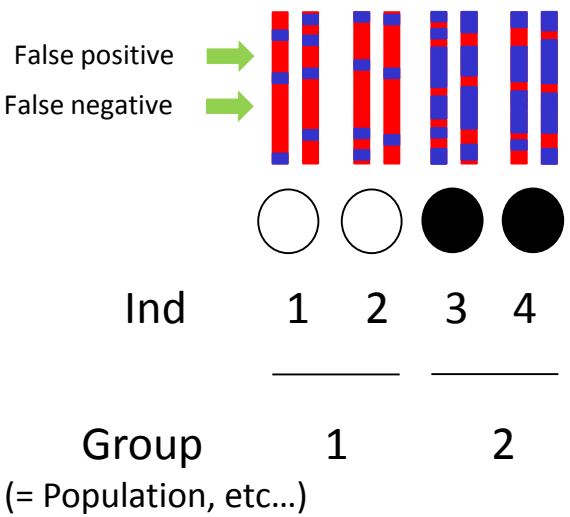
Problems:

Need for a high number of markers, well distributed over the genome and in linkage disequilibrium with causal variants

Sensitive to rare allele (MAF cut-off)

Population structure = demographic history

One of GWAS's major caveat: Population structure



False positives:

When neutral marker correlated with variation because of population structure

Solution: control with measure of genetic relatedness (kinship matrix, …), but:

False negatives :

When 'causal' markers are 'masked' because of control for genetic relatedness

Solution: Do GWAS with 'homogeneous' sample (if possible)

Combine with crosses (will invalidate false positive)

Korte & Farlow, 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods ***9****:29 (****DOI:*** *10.1186/1746-4811-9-29)*

GWAS work !

Li et al. 2010. Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana.* 107, 49: 21199–21204 (doi: 10.1073/pnas.1007431107)

Successful identification of genes already known to be implicated in variation of flowering time using GWAS on natural lines.

Brachi et al. 2010. Linkage and Association Mapping of *Arabidopsis thaliana* Flowering Time in Nature. PLoS Genetics. DOI: 10.1371/journal.pgen.1000940

Overlap of genomic regions identified both with QTL mapping and GWAS on natural lines

Comeault et al. 2015. Selection on a Genetic Polymorphism Counteracts Ecological Speciation in a Stick Insect. Current Biology. 25, 15: 1975–1981

Identification of genetic basis of colour crypsis in a non model organism using GBS

Many other examples …

GWAS: which software? non-exhaustive list

To cite a few: EMMAX, **GenABEL**, GAPIT, GWAspi, **gemma**, piMASS ... and many others !

google search: http://omictools.com/association-mapping-category

2 main types:

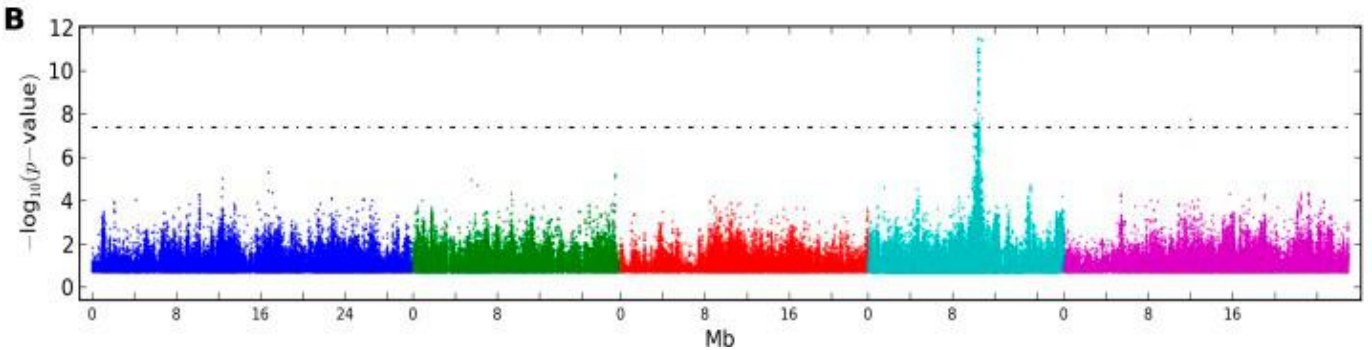
Single marker association: ***GenABEL*** **<- Micheal's session**

Multi-marker association: ***gemma*** **<- my session (Victor's original session)**

GWA: Single marker methods

To cite a few: EMMAX, **GenABEL**

Association computed for every marker independently along the genome



Advantage: Very fast (30 seconds to 5 minutes)

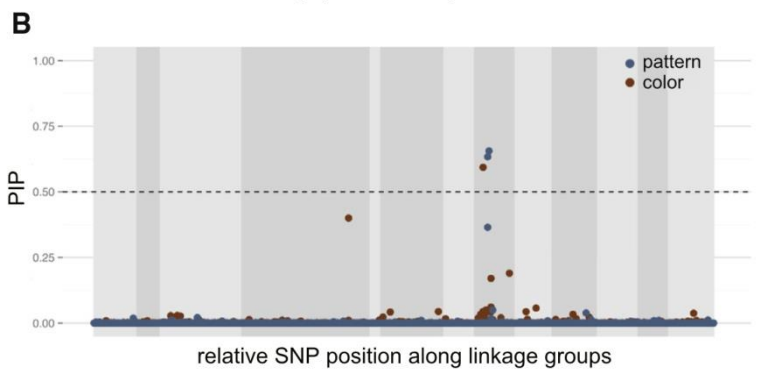
Caveats: Results with LD for downstream analysis (enrichment in GO term...).

Need to redo analysis with SNPs as covariate to detect epistasis (good exemple in Dubin et al. 2015. DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. eLife. doi: 10.7554/eLife.05255.)

GWA: Multi-marker methods

To cite a few: **gemma**, piMASS

Association computed with combination of markers



Advantages: Results without LD for downstream analysis (enrichment in genetic features, GO etc...).

~Detection of epistasis ?

Caveats: Bayesian framework and MCMC = takes time(1.5 days per run)

Marker with the highest association score != causal variant

The marker with the highest association score can sometimes be the causal variant but:

It is only the most associated marker in your dataset and the causal variant might not be in it.

Most GWAS only use bi-allelic SNPs (indel, inversion…)

Any questions ?