

Introduction to **GWAS**

Genotyping

Christian Werner

(Quantitative geneticist and biostatistician) **EiB, CIMMYT**, Texcoco (Mexico)

Filippo Biscarini

(Biostatistician, bioinformatician and quantitative geneticist) **CNR-IBBA**, Milan (Italy)



HerrFaloppio

Oscar González-Recio

(Computational biologist and quantitative geneticist) **INIA-UPM**, Madrid (Spain)



OscarGenomics



Genotyping

- A very brief overview -

What is a marker?

“A measurable biological feature used to identify or track genetic variation, biological state, or trait association.”

Visible mutants: eye colour, body colour, wing shape in *Drosophila*, ...

Blood group markers: ABO and other inherited blood types, ...

Biochemical markers: haemoglobin variants, enzyme variants,

Plant markers: flower colour, seed colour, waxy starch, ...



The first steps – Biomarkers to describe genetic variation

A MOLECULAR APPROACH TO THE STUDY OF GENIC
 HETEROZYGOSITY IN NATURAL POPULATIONS.
 I. THE NUMBER OF ALLELES AT DIFFERENT
 LOCI IN *DROSOPHILA PSEUDOBSCURA*¹

J. L. HUBBY AND R. C. LEWONTIN

Department of Zoology, University of Chicago, Chicago, Illinois

Received March 30, 1966

¹The work reported here was supported in part by grants from the National Science Foundation (GB 3013) and the Public Health Service (GM-13126).
 Genetics 84: 577-594 August 1966.



<https://academic.oup.com/genetics/article/54/2/577/5988209>

582

J. L. HUBBY AND R. C. LEWONTIN

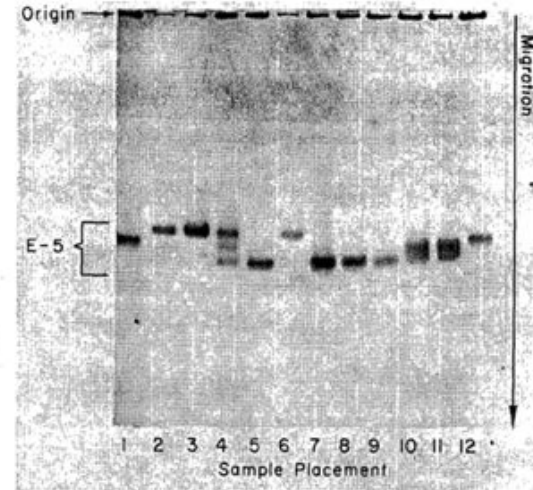
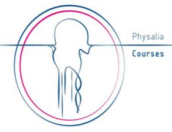


FIGURE 1.—Gel illustrating sample placement and typical results of strain analysis for Esterase-5. The first and the last samples were derived from the standard reference strain (E-5¹⁻⁰⁰), while positions 2 through 6 were obtained from five individuals of one strain and positions 7 through 11 are from five individuals of a second strain. Positions 2, 3, and 6 contain Esterase-5⁹⁵, position 5 contains Esterase-5¹⁻¹², and position 4 contains Esterase-5³⁰, Esterase-5¹⁻¹², and a site of activity between them. Positions 7, 8, and 9 contain Esterase-5¹⁻¹² and positions 10 and 11 contain Esterase-5¹⁻⁰⁰ and Esterase-5¹⁻¹². A site of activity midway between the latter two is barely discernible. In all the figures the direction of migration of the protein is down toward the anode.

From few to many markers – Molecular markers (DNA markers)

- arise from **different classes of DNA mutations** such as substitution mutations (point mutations), rearrangements (insertions or deletions) or errors in replication of tandemly repeated DNA
- are often located in **non-coding regions** of DNA
- are virtually **unlimited in number** and are **not affected by environmental factors** and/or the developmental stage of the plant/animal/human
- RFLP, AFLP, RAPD, SSR (microsatellites), CNV, InDel, **SNP**, ...

From few to many markers – Molecular markers (DNA markers)



Euphytica (2005) 142: 169–196
DOI: 10.1007/s10681-005-1681-5

© Springer 2005

An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts

B.C.Y. Collard^{1,4,*}, M.Z.Z. Jahufer², J.B. Brouwer³ & E.C.K. Pang¹

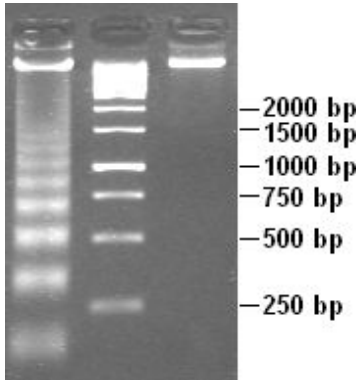
¹*Department of Biotechnology and Environmental Biology, RMIT University, P.O. Box 71, Bundoora, Victoria 3083, Australia;* ²*AgResearch Ltd., Grasslands Research Centre, Tennent Drive, Private Bag 11008, Palmerston North, New Zealand;* ³*P.O. Box 910, Horsham, Victoria, Australia 3402;* ⁴*Present address: Plant Breeding, Genetics and Biotechnology Division, International Rice Research Institute (IRRI), DAPO Box 7777, Metro Manila, Philippines;* (*author for correspondence: e-mail: bcycollard@hotmail.com)

Received 11 July 2004; accepted 2 February 2005

Genotyping Systems

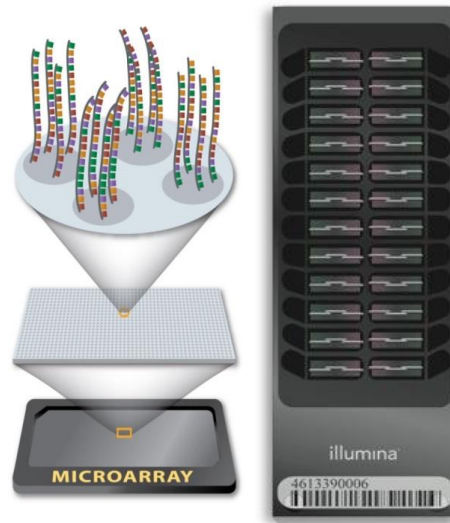
Marker gel

(a few markers)



SNP array

(100s -1,000,000s)

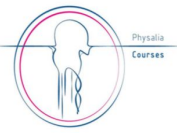


Genome sequencing

(1,000,000s +)



Genotyping for GWAS – Common Technologies

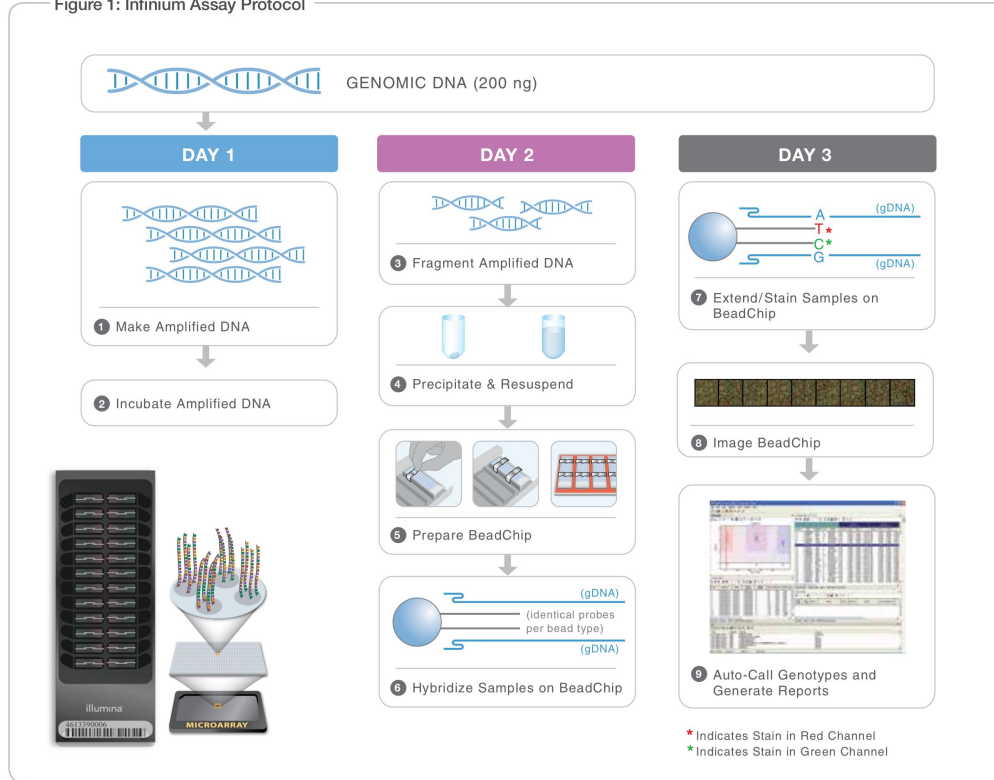


| Technology | Description |
|--|---|
| Marker arrays | Fixed set of known SNPs |
| Reduced Representation Sequencing (GBS, RADseq) | Sequence a repeatable subset of the genome (same restriction enzyme cuts - “semi reproducible”) |
| Low-coverage WGS (skim sequencing) | Sequence the whole genome shallowly, usually with imputation |
| Whole Genome Sequencing (WGS) | Sequence the genome more completely (for dense variant discovery) |

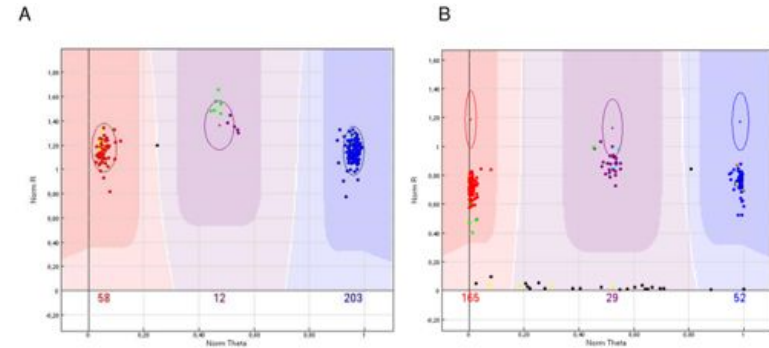


SNP array genotyping

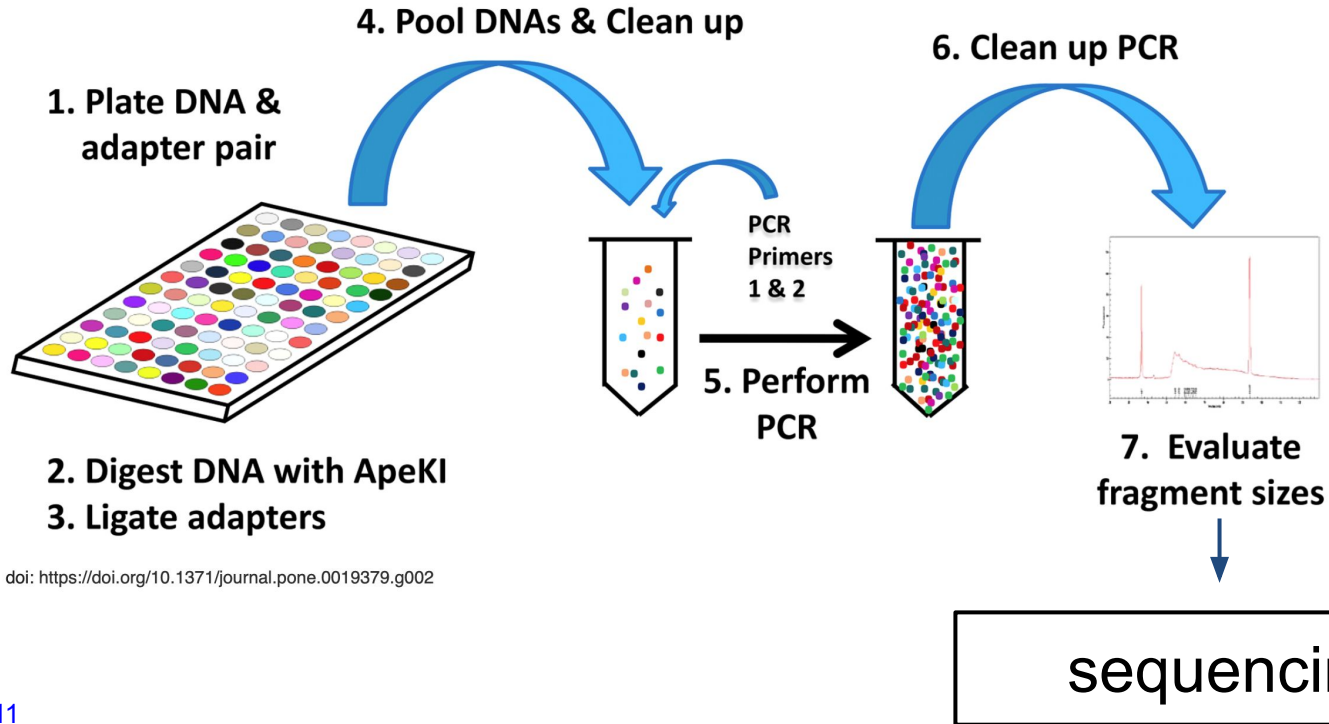
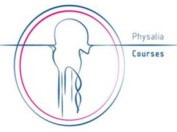
Figure 1: Infinium Assay Protocol



Genotype calling
3 genotypes: AA, AG, GG



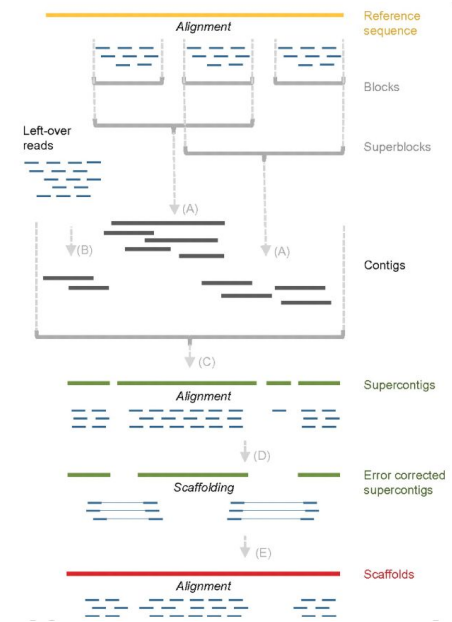
Reduced representation sequencing – Genotyping-by-Sequencing (GBS)



doi: <https://doi.org/10.1371/journal.pone.0019379.g002>

[Elshire et al. 2011](#)

The Next Generation Sequencing Revolution



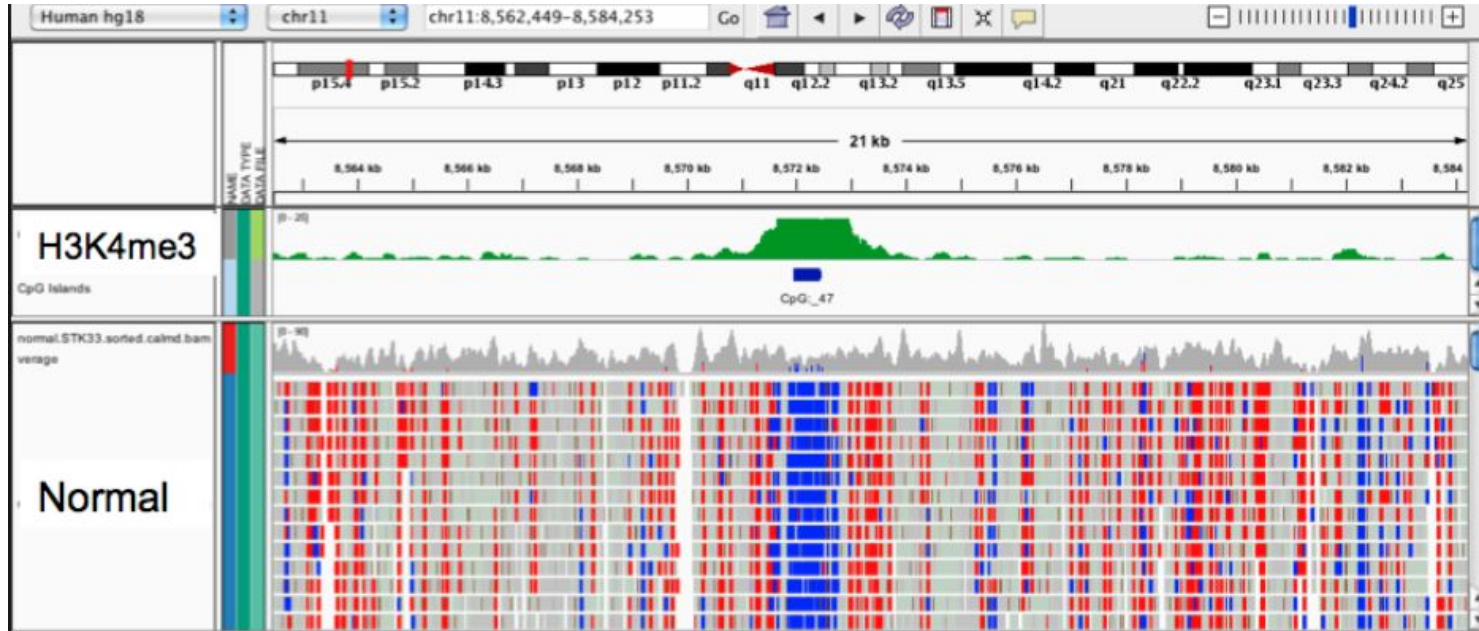
Assembly!

| | |
|---|------------------------|
| The Human Genome Project (1990-2003) | USD 2.7 billion |
| Today | < USD 1,000 per genome |



The Next Generation Sequencing Revolution

Millions of polymorphisms in the genome sequences...



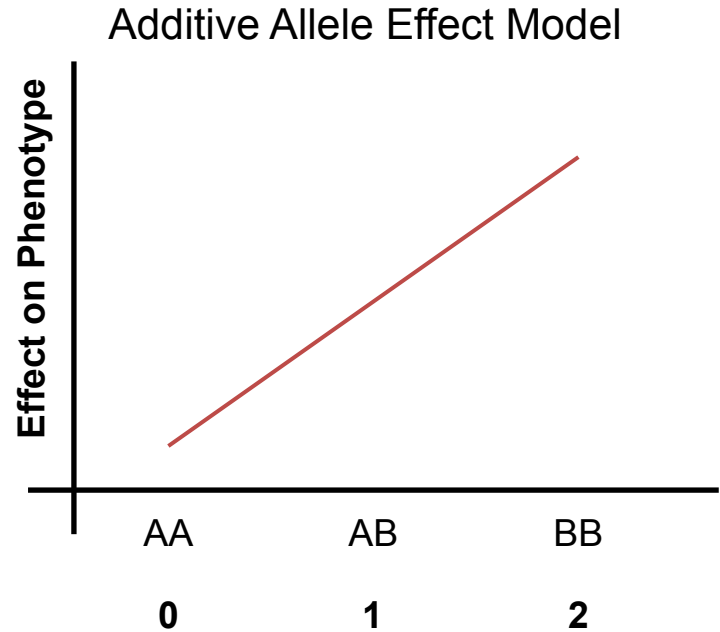
https://link.springer.com/chapter/10.1007/10_2017_46

Genotyping

- Why SNP? -

The diploid genotype matrix – “additive” effect modeling

| | Individuals | | | | | | |
|--------|-------------|----|---|---|---|----|---|
| Marker | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| | 2 | 2 | 2 | 2 | 2 | -1 | 2 |
| | 2 | -1 | 2 | 2 | 2 | 2 | 1 |
| | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| | 0 | 1 | 0 | 0 | 0 | -1 | 0 |
| | 0 | 0 | 0 | 0 | 0 | -1 | 0 |
| | 2 | 1 | 1 | 0 | 0 | -1 | 0 |
| | 0 | 2 | 1 | 1 | 2 | 2 | 1 |
| | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| | 1 | 2 | 1 | 1 | 2 | 2 | 1 |



The diploid genotype matrix – “dominance” effect modeling

| | Individuals | | | | | | |
|--------|-------------|----|---|---|---|----|---|
| Marker | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| | 2 | 2 | 2 | 2 | 2 | -1 | 2 |
| | 2 | -1 | 2 | 2 | 2 | 2 | 1 |
| | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| | 0 | 1 | 0 | 0 | 0 | -1 | 0 |
| | 0 | 0 | 0 | 0 | 0 | -1 | 0 |
| | 2 | 1 | 1 | 0 | 0 | -1 | 0 |
| | 0 | 2 | 1 | 1 | 2 | 2 | 1 |
| | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| | 1 | 2 | 1 | 1 | 2 | 2 | 1 |

Dominance Genotype Effect Model

