Introduction to GWAS

Genotyping

Christian Werner

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

Filippo Biscarini

HerrFalloppio

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

Oscar González-Recio

OscarGenomics

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

Genotyping

- A very brief overview -

The first steps – Biomarkers

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

J. L. HUBBY AND R. C. LEWONTIN

Department of Zoology, University of Chicago, Chicago, Illinois

Received March 30, 1966

[1] The work reported here was supported in part by grants from the National Science Foundation (GB 3113) and the Public Health Service (GM-11226).

Genetics 54: 577–594 August 1966.

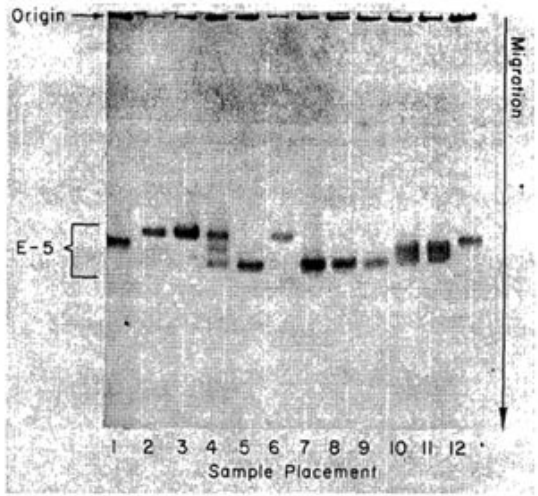


FIGURE 1.—Gel illustrating sample placement and typical results of strain analysis for Esterase-5. The first and the last samples were derived from the standard reference strain ($E\text{-}5^{1.00}$), while positions 2 through 6 were obtained from five individuals of one strain and positions 7 through 11 are from five individuals of a second strain. Positions 2, 3, and 6 contain Esterase-$5^{.95}$, position 5 contains Esterase-$5^{1.12}$, and position 4 contains Esterase-$5^{.95}$, Esterase-$5^{1.12}$, and a site of activity between them. Positions 7, 8, and 9 contain Esterase-$5^{1.12}$ and positions 10 and 11 contain Esterase-$5^{1.00}$ and Esterase-$5^{1.12}$. 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 usually located in non-coding regions of DNA
- are practically unlimited in number and are not affected by environmental factors and/or the developmental stage of the plant
- **RFLP, AFLP, RAPD, SSR (microsatellites), SNP**

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

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

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

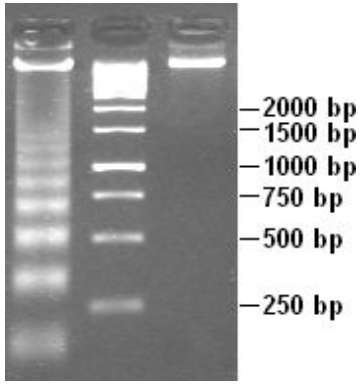
B.C.Y. Collard[1,4,*], M.Z.Z. Jahufer[2], J.B. Brouwer[3] & E.C.K. Pang[1]
[1]*Department of Biotechnology and Environmental Biology, RMIT University, P.O. Box 71, Bundoora, Victoria 3083, Australia;* [2]*AgResearch Ltd., Grasslands Research Centre, Tennent Drive, Private Bag 11008, Palmerston North, 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

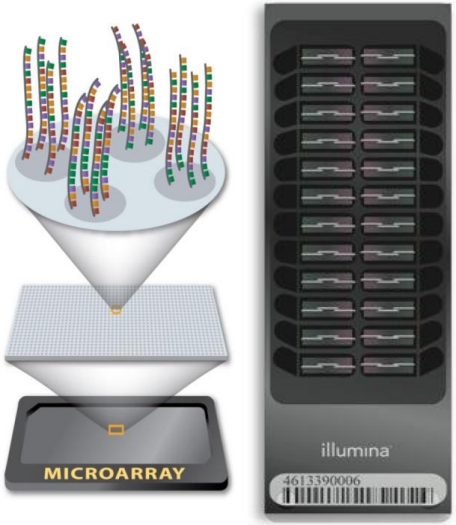
Marker gel

(a few markers)



SNP array (or GBS)

(100s -1,000,000s)

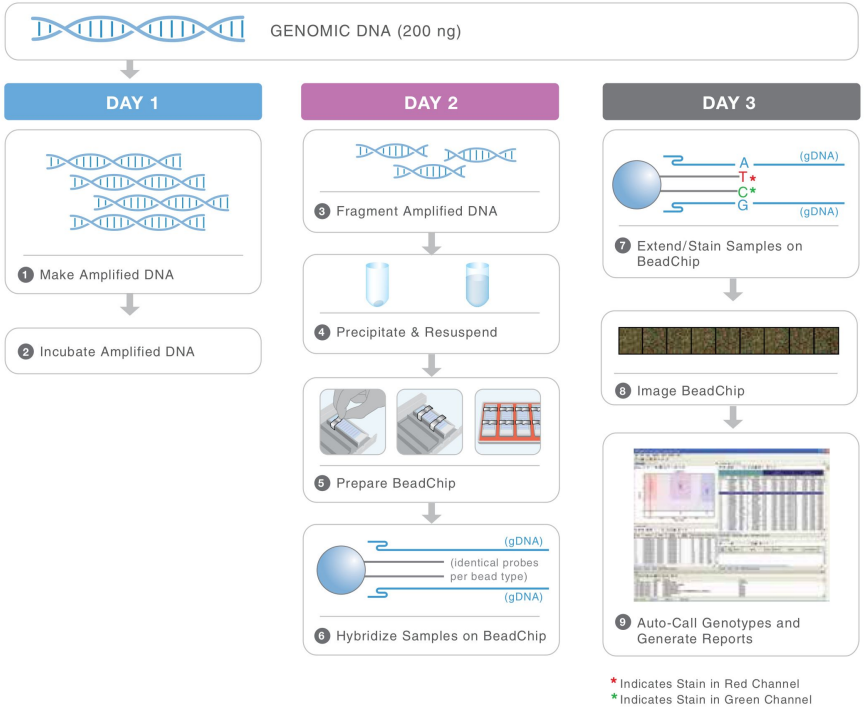


Genome sequencer

(1,000,000s +)

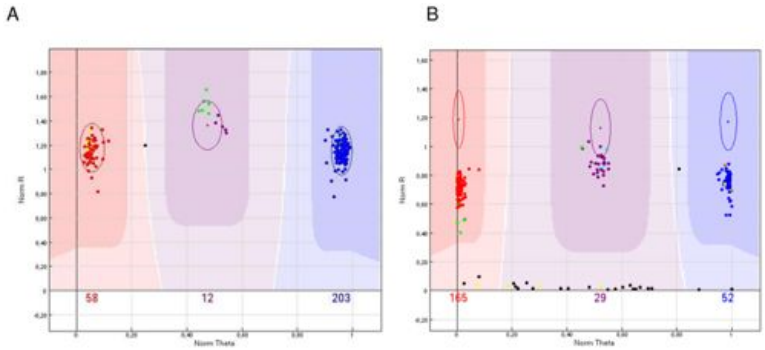
SNP array genotyping

Figure 1: Infinium Assay Protocol

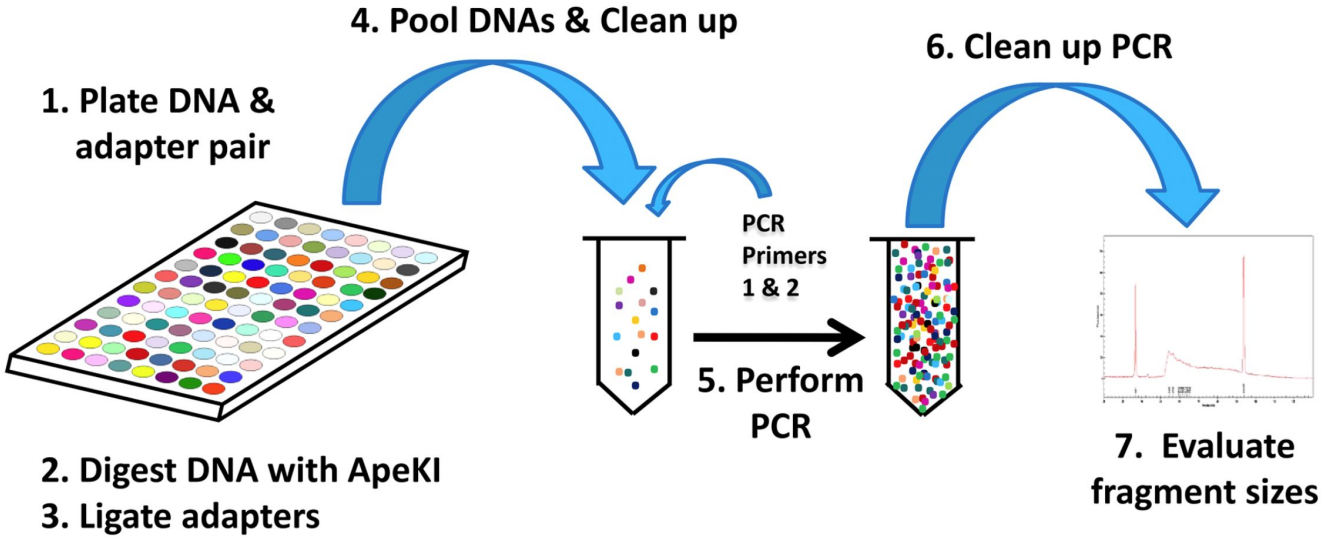


Genotype calling

3 genotypes: AA, AG, GG



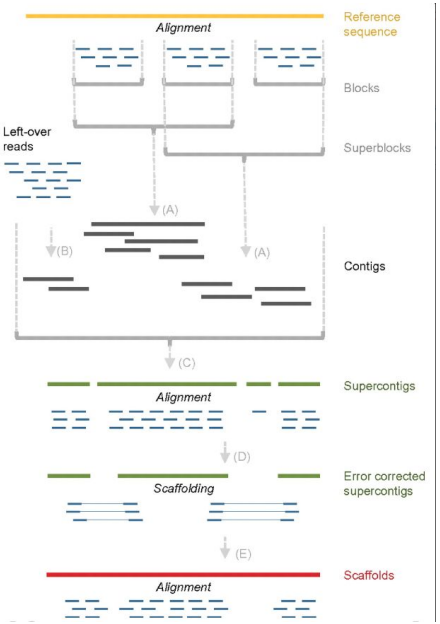
Reduced representation sequencing – Genotyping-by-Sequencing (GBS)



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

The Next Generation Sequencing Revolution

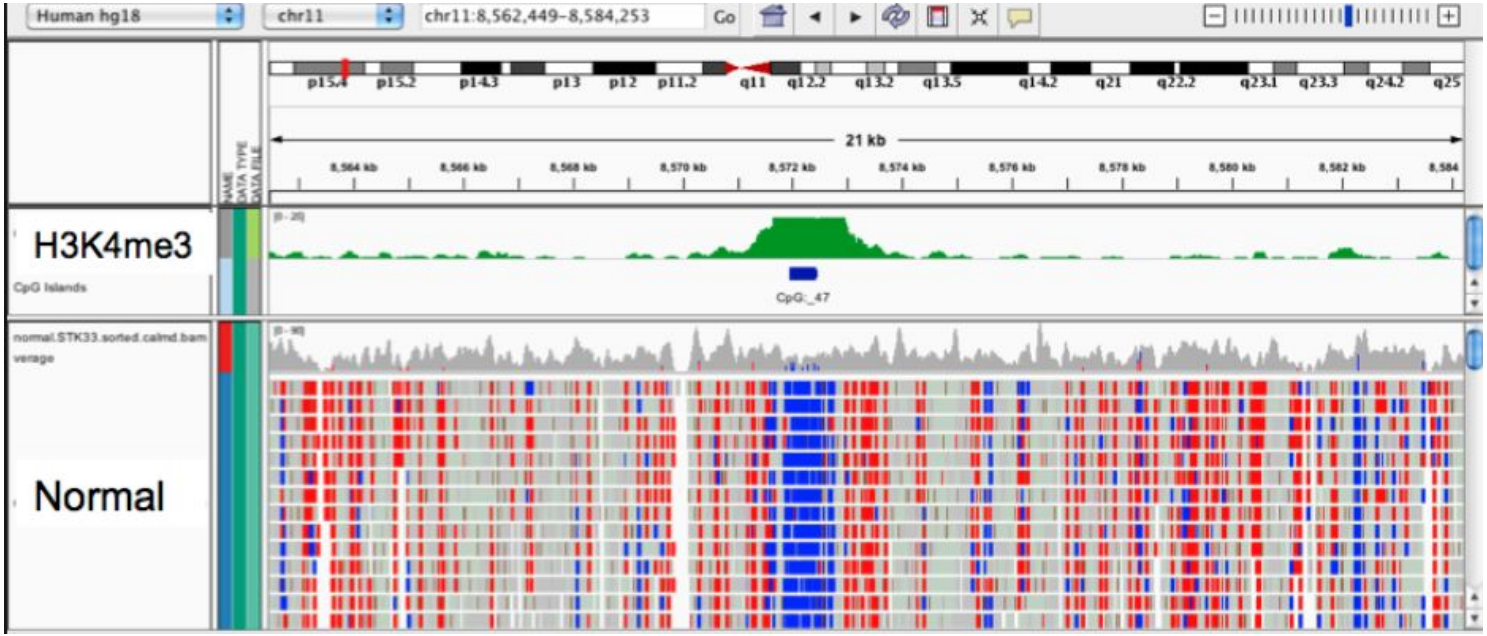
Hmmm...now the data is here….so what now?



Assembly!

The Next Generation Sequencing Revolution

Millions of polymorphisms in the genome sequences…

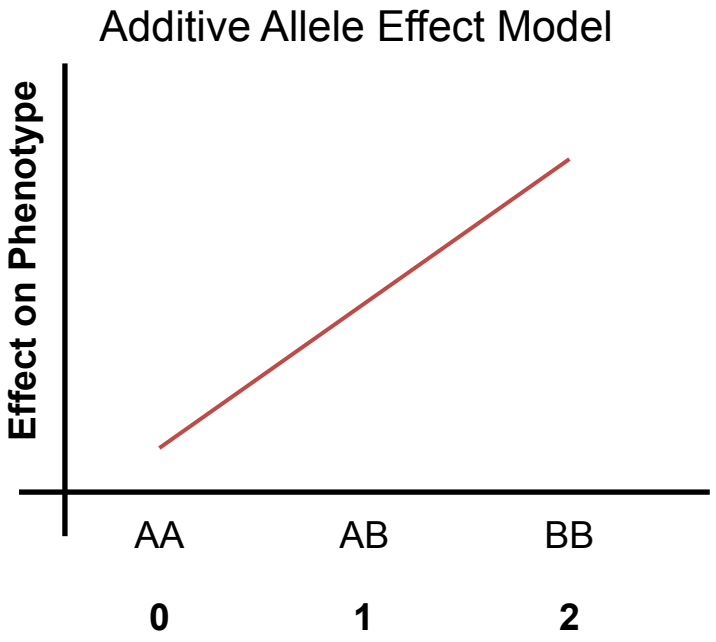


The diploid genotype matrix – “additive” effect modeling

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



Genotyping formats