

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 -

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 34: 517-524 August 1968.



582

J. L. HUBBY AND R. C. LEWONTIN

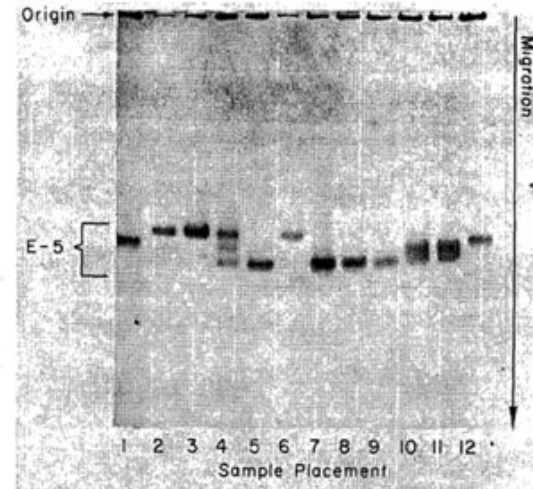


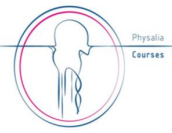
FIGURE 1.—Gel illustrating sample placement and typical results of strain analysis for Esterase-5. The first and the last samples were derived from the standard reference strain (E-5^{1.00}), while positions 2 through 6 were obtained from five individuals of one strain and positions 7 through 11 are from five individuals of a second strain. Positions 2, 3, and 6 contain Esterase-5⁹⁵, position 5 contains Esterase-5^{1.12}, and position 4 contains Esterase-5⁹⁵, 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.

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

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 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), SNP, CNV, InDel ...**

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

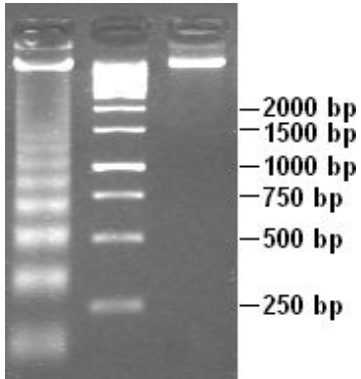
¹*Department of Biotechnology and Environmental Biology, RMIT University, P.O. Box 71, Bundoora, Victoria 3083, Australia;* ²*AgResearch Ltd., Grasslands Research Centre, Tennent Drive, Private Bag 11008, Palmerston North, New Zealand;* ³*P.O. Box 910, Horsham, Victoria, Australia 3402;* ⁴*Present address: 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

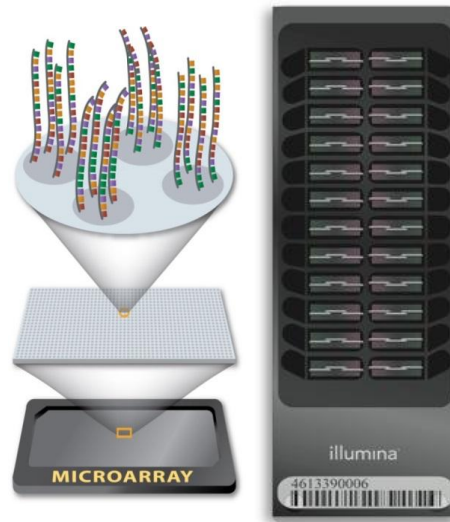
Marker gel

(a few markers)



SNP array

(100s -1,000,000s)



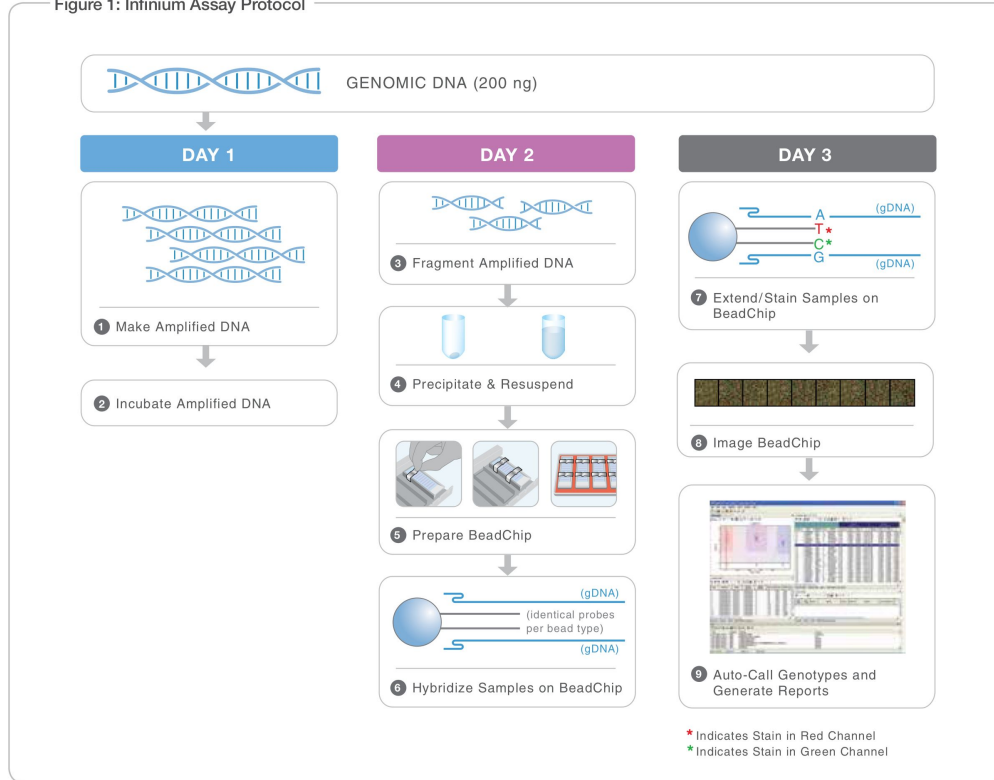
→(GBS) → Genome sequencer

(1,000,000s +)

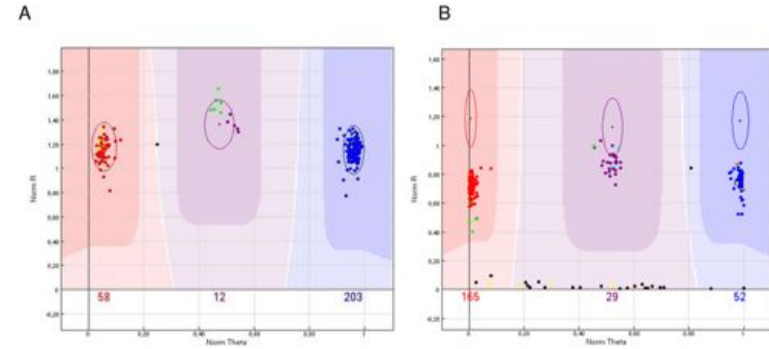


SNP array genotyping

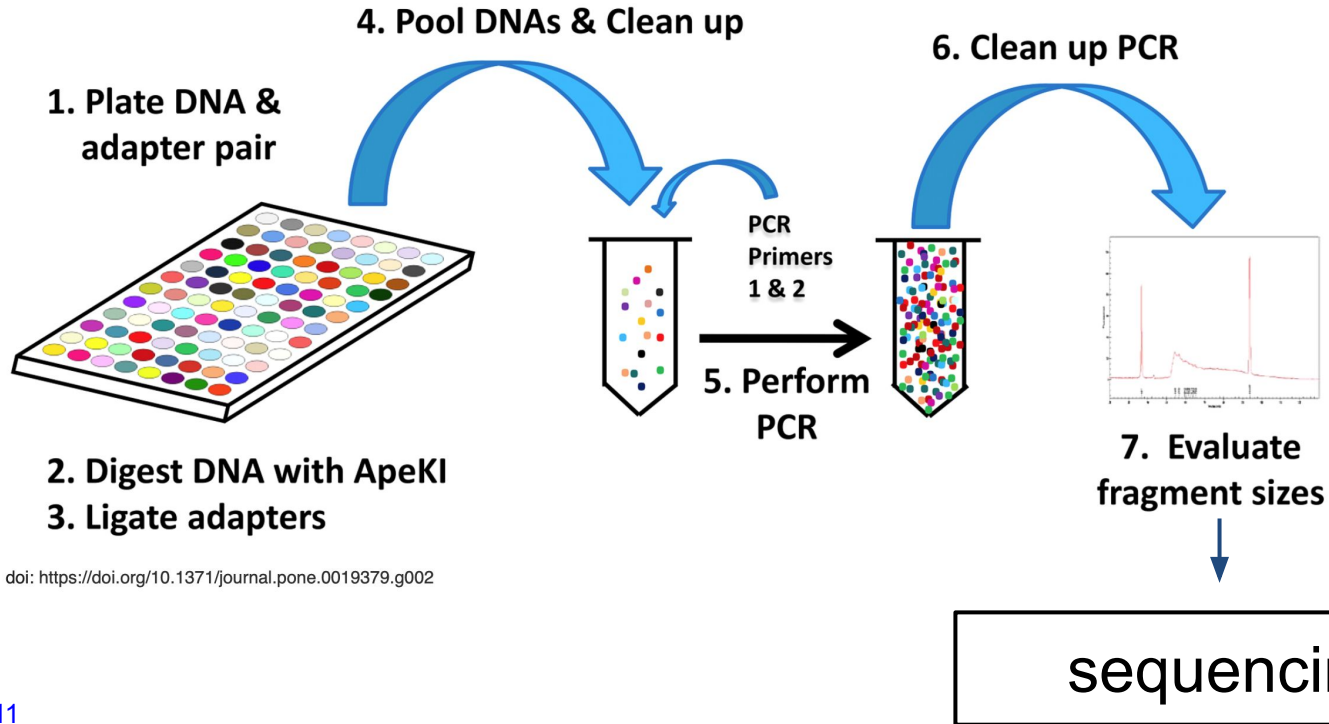
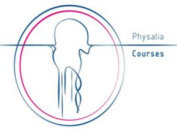
Figure 1: Infinium Assay Protocol



Genotype calling
3 genotypes: AA, AG, GG



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



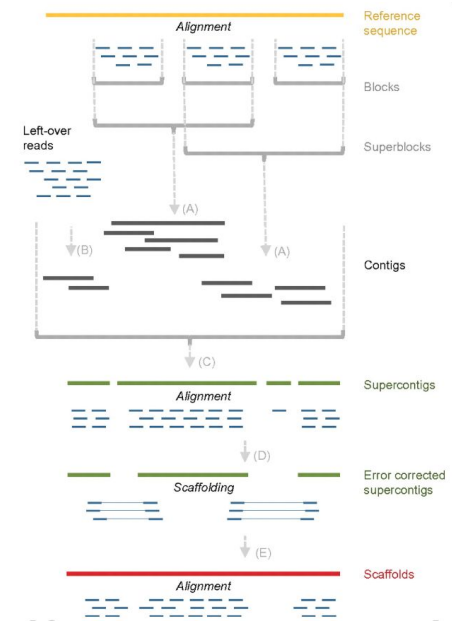
doi: <https://doi.org/10.1371/journal.pone.0019379.g002>

[Elshire et al. 2011](#)

The Next Generation Sequencing Revolution



Hmmm...now the data is there....so what now?

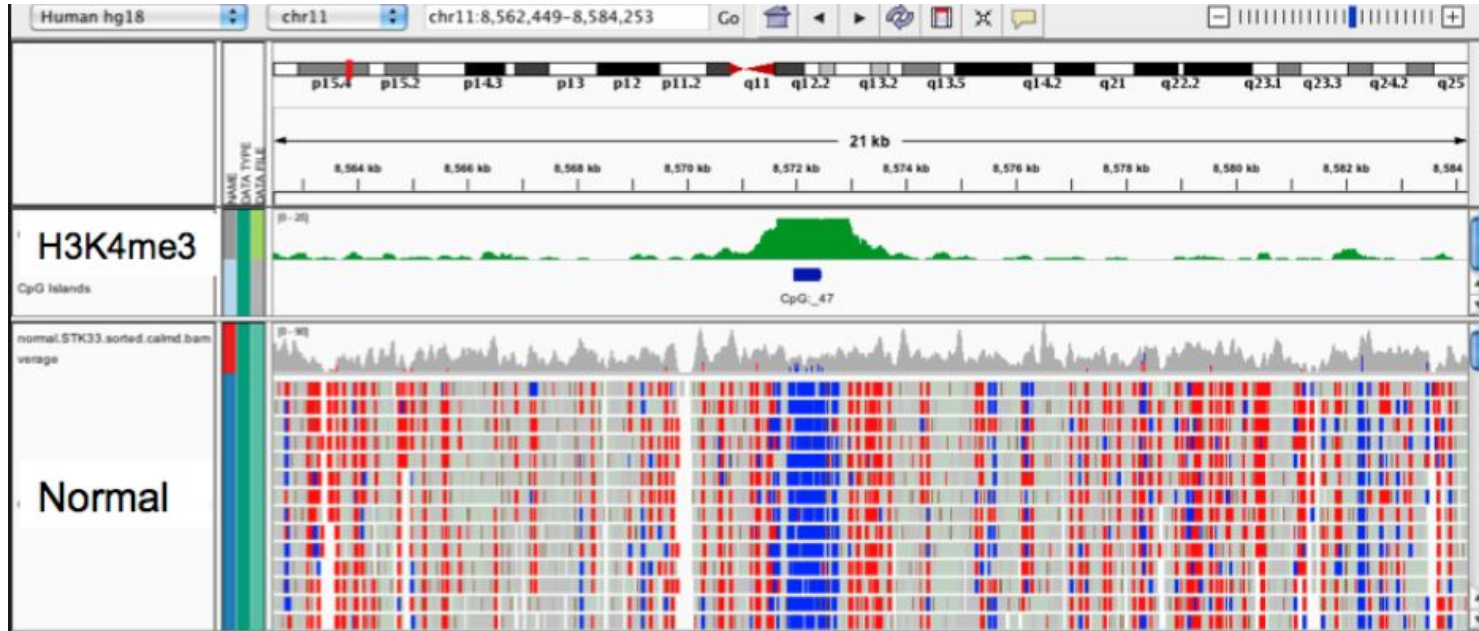


Assembly!



The Next Generation Sequencing Revolution

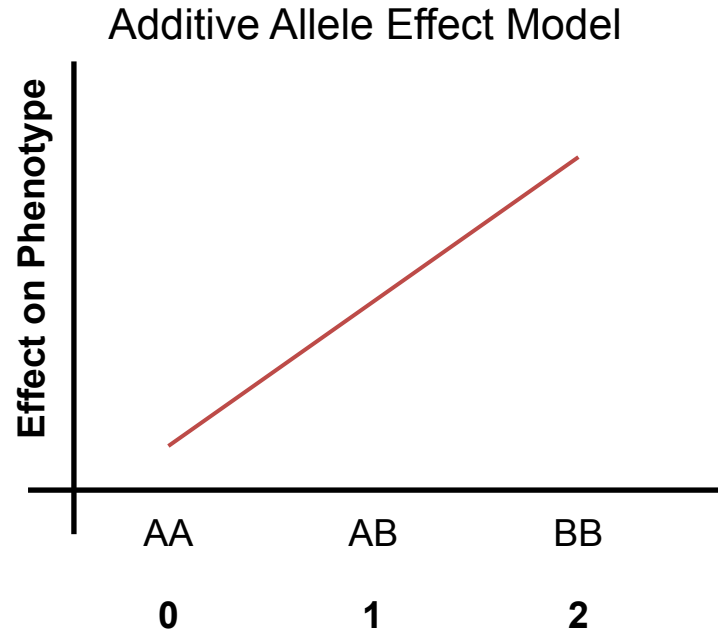
Millions of polymorphisms in the genome sequences...



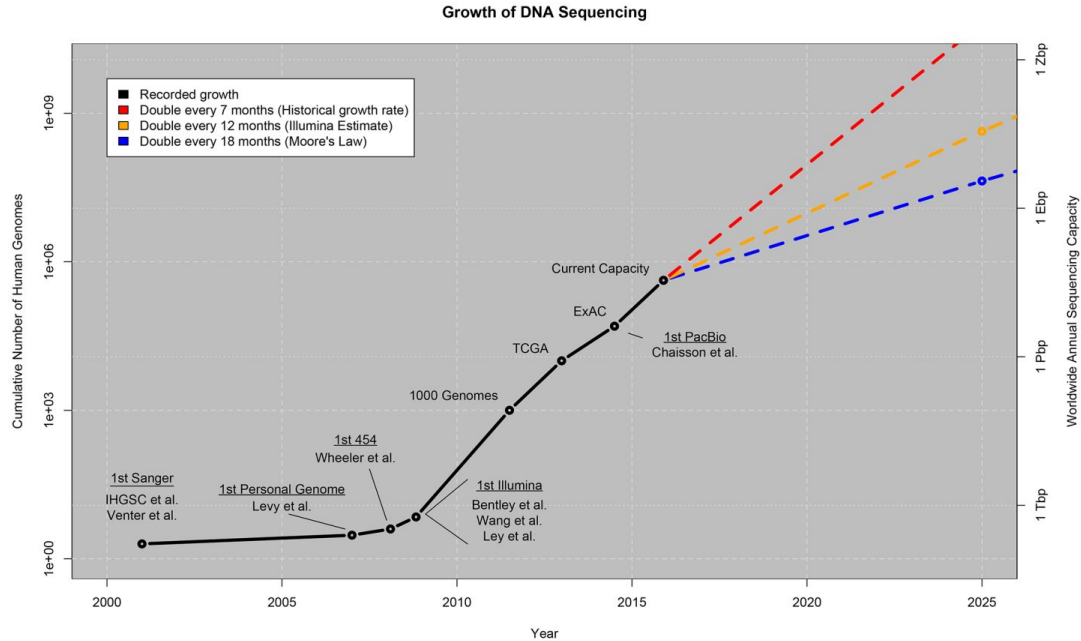
https://link.springer.com/chapter/10.1007/10_2017_46

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



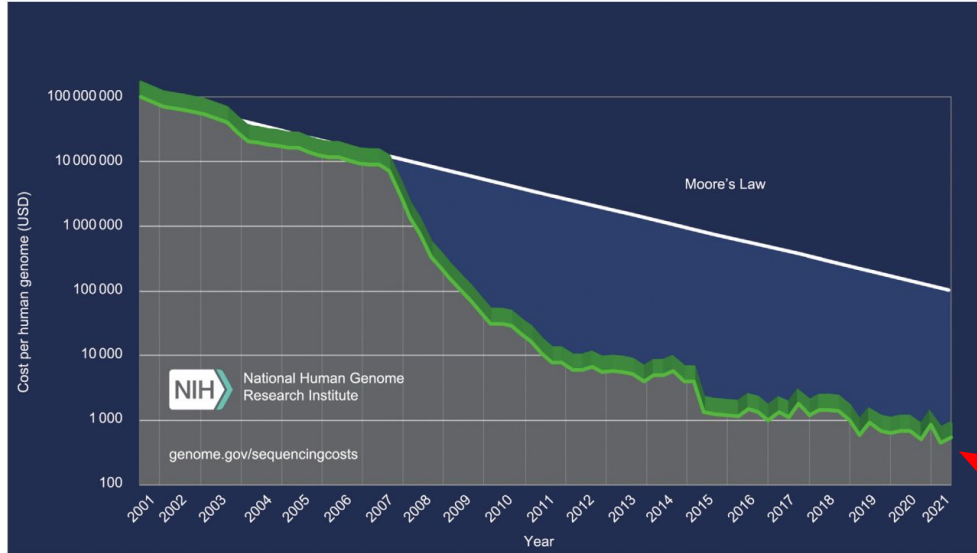
Big data: astronomical or genomical?



- new technologies e.g. Nanopore long-reads sequencing
- single-cell sequencing
- tons of sequencing/genotyping projects (plants, animals, humans)
- metagenomics
- “four-headed beast”:
 - i) data acquisition: 1 ZB/year (10^{21})
 - ii) data storage: 2-40 EB/year (10^{18})
 - iii) data distribution: few massive, many small data exchanges
 - iv) data analysis: highly heterogeneous and intensive

Stephens et al. 2015. Big data: astronomical or genomical?. PLoS biology, 13(7), p.e1002195.

\$1 genome?



- emerging/new technologies
- new companies
- streamlined (more efficient) processes (e.g. innovative chemistry)
- \$100 genome: not yet there, but promised by several companies
- \$1 genome? we are headed towards that!

on the brink of the next revolution in sequencing?

- Pennisi, E. "\$100 genome? New DNA sequencers could be a 'game changer' for biology, medicine." Washington, DC: Science; 2022. doi: 10.1126/science.add5060
- Pollie, Robert. "Genomic sequencing costs set to head down again." Engineering 23 (2023): 3-6.