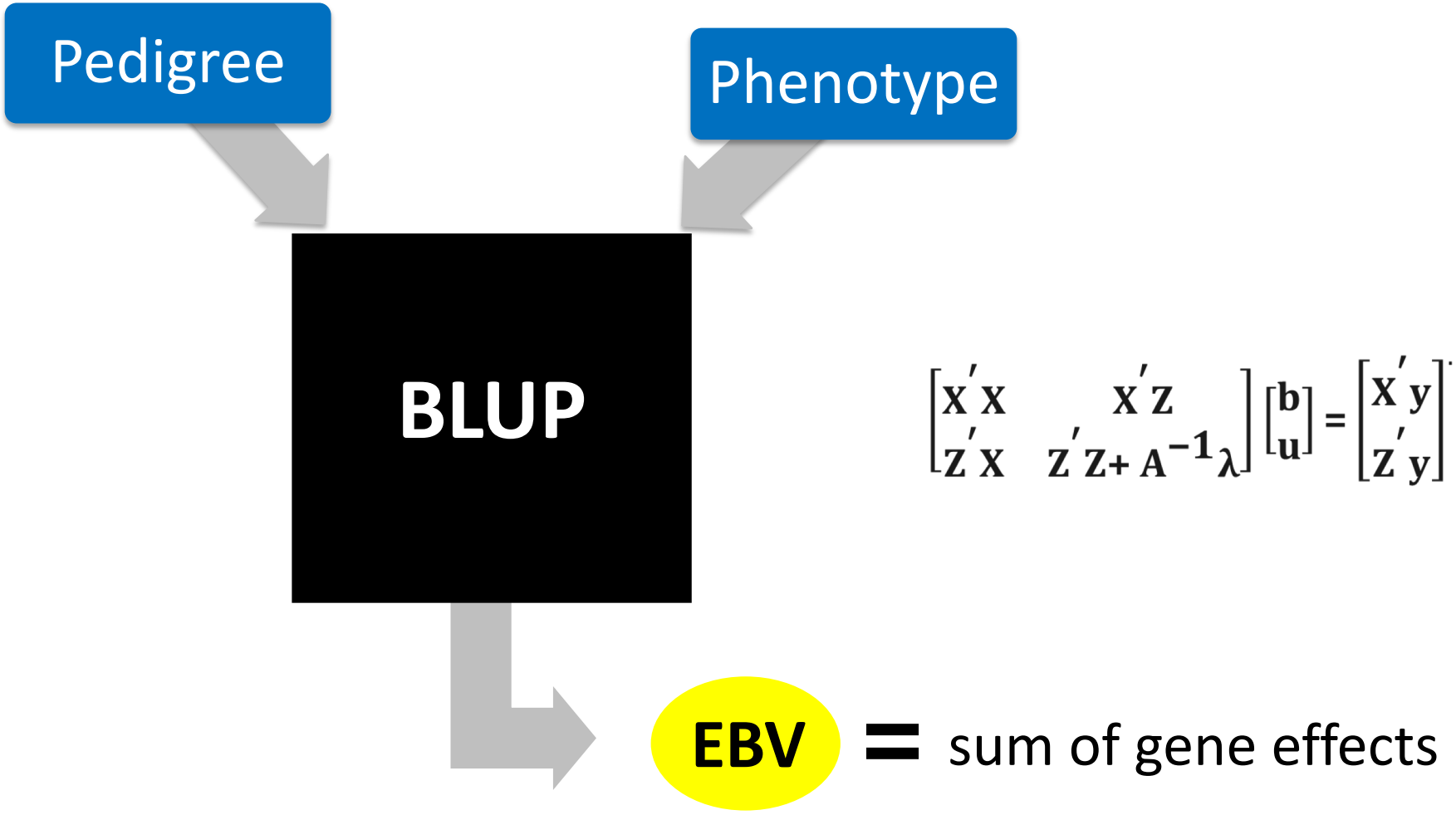
UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

The promise of genomics for breeding and genetics

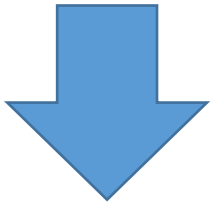
Daniela Lourenco

Athens, GA - 05/21/2018

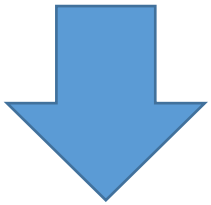
Traditional evaluation



What if we could know the genes/DNA variants that affect the trait?



Would we have more accurate EBV?



Genomics in livestock breeding

Genomic information

Theor Appl Genet (1983) 67:25–33

Genetic polymorphism in varietal identification and genetic improvement *

M. Soller [1] and J. S. Beckmann [2]

[1] Department of Genetics, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel
[2] Institute of Field and Garden Crops, Agricultural Research Organization, The Volcani Center 50250 Bet Dagan, Israel

Received July 14, 1982; Accepted July 3, 1983
Communicated by A. Robertson

Summary. New sources of genetic polymorphisms promise significant additions to the number of useful genetic markers in agricultural plants and animals, and prompt this review of potential applications of polymorphic genetic markers in plant and animal breeding. Two major areas of application can be distinguished. The first is based on the utilization of genetic markers to determine genetic relationships. These applications include varietal identification, protection of breeder's rights, and parentage determination. The second area of application is based on the use of genetic markers to identify and map loci affecting quantitative traits, and to monitor these loci during introgression or selection programs. A variety of breeding applications based on

Use of DNA polymorphisms as genetic markers

- Construct genetic relationships
- Parentage determination
- Identification of QTL

RFLP

Genomic information

articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

* A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

The rediscovery of Mendel's laws of heredity in the opening weeks of the 20th century[1–5] sparked a scientific quest to understand the nature and content of genetic information that has propelled biology for the last hundred years. The scientific progress made falls naturally into four main phases, corresponding roughly to the four quarters of the century. The first established the cellular basis of heredity: the chromosomes. The second defined the molecular basis of heredity: the DNA double helix. The third unlocked the informational basis of heredity, with the discovery of the biological mechanism by which cells read the information contained in genes and with the invention of the recombinant DNA technologies of cloning and sequencing by which scientists can do the same.

The last quarter of a century has been marked by a relentless drive to decipher first genes and then entire genomes, spawning the field of genomics. The fruits of this work already include the genome sequences of 599 viruses and viroids, 205 naturally occurring plasmids, 185 organelles, 31 eubacteria, seven archaea, one fungus, two animals and one plant.

Here we report the results of a collaboration involving 20 groups from the United States, the United Kingdom, Japan, France, Germany and China to produce a draft sequence of the human genome. The draft genome sequence was generated from a physical map covering more than 96% of the euchromatic part of the human genome and, together with additional sequence in public databases, it covers about 94% of the human genome. The sequence was produced over a relatively short period, with coverage rising from about 10% to more than 90% over roughly fifteen months. The sequence data have been made available without restriction and updated daily throughout the project. The task ahead is to produce a finished sequence, by closing all gaps and resolving all ambiguities. Already about one billion bases are in final form and the task of bringing the vast majority of the sequence to this standard is now straightforward and should proceed rapidly.

The sequence of the human genome is of interest in several respects. It is the largest genome to be extensively sequenced so far, being 25 times as large as any previously sequenced genome and eight times as large as the sum of all such genomes. It is the first vertebrate genome to be extensively sequenced. And, uniquely, it is the genome of our own species.

Much work remains to be done to produce a complete finished sequence, but the vast trove of information that has become available through this collaborative effort allows a global perspective on the human genome. Although the details will change as the sequence is finished, many points are already clear.

- The genomic landscape shows marked variation in the distribution of a number of features, including genes, transposable elements, GC content, CpG islands and recombination rate. This gives us important clues about function. For example, the developmentally important HOX gene clusters are the most repeat-poor regions of the human genome, probably reflecting the very complex coordinate regulation of the genes in the clusters.
- There appear to be about 30,000–40,000 protein-coding genes in the human genome—only about twice as many as in worm or fly. However, the genes are more complex, with more alternative splicing generating a larger number of protein products.
- The full set of proteins (the 'proteome') encoded by the human genome is more complex than those of invertebrates. This is due in part to the presence of vertebrate-specific protein domains and motifs (an estimated 7% of the total), but more to the fact that vertebrates appear to have arranged pre-existing components into a richer collection of domain architectures.
- Hundreds of human genes appear likely to have resulted from horizontal transfer from bacteria at some point in the vertebrate lineage. Dozens of genes appear to have been derived from transposable elements.
- Although about half of the human genome derives from transposable elements, there has been a marked decline in the overall activity of such elements in the hominid lineage. DNA transposons appear to have become completely inactive and long-terminal repeat (LTR) retroposons may also have done so.
- The pericentromeric and subtelomeric regions of chromosomes are filled with large recent segmental duplications of sequence from elsewhere in
- Analysis o
standing my
suggests that
retention of
elements ma
- The muta
meiosis, sho
- Cytogenet
tions that la
G-bands' in
- Recombin
(around 20
chromosome
occurrence
meiosis.
- More than
in the huma
allow the
mapping of

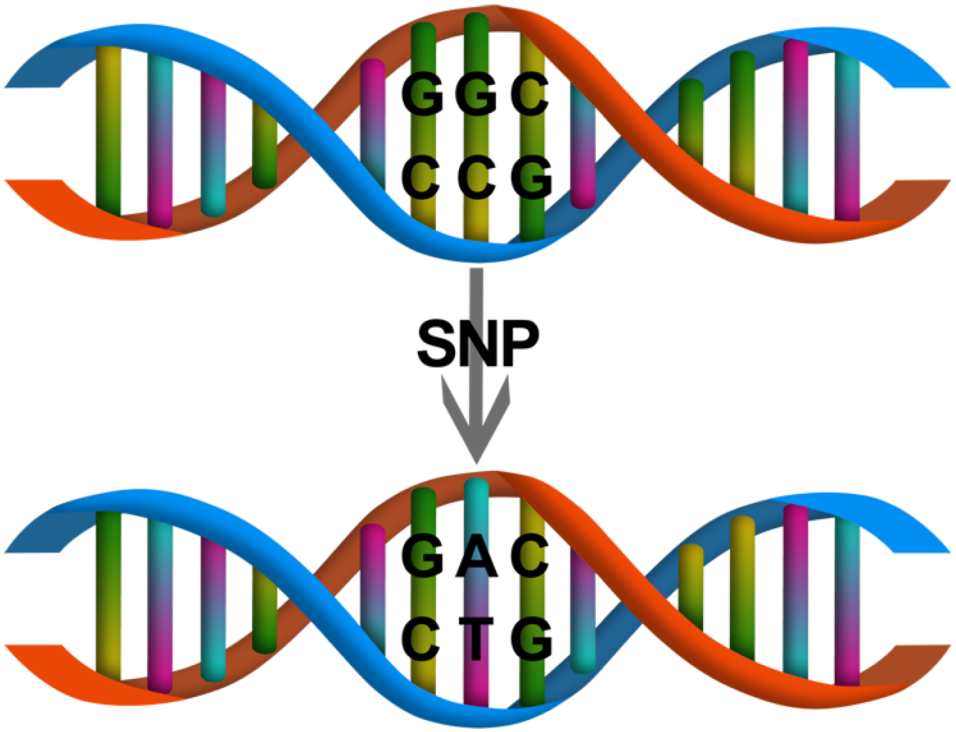
In this pa
the project a
of the draft g
the sequenc
elements an
biological p
proteins and

"The majority of the genome sequence variation can be attributed to single nucleotide polymorphisms (SNP)"

"SNPs have become the bread-and-butter of DNA sequence variation"
(Stonecking, 2001)

Single Nucleotide Polymorphisms

Individual 1



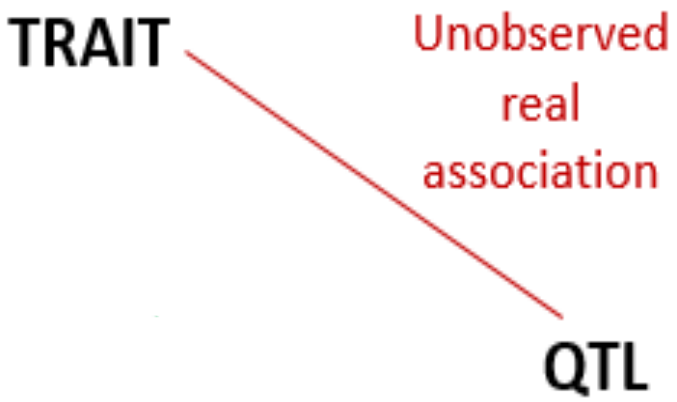
Individual 2

http://www.thinnergene.com/about-thinnergene/genetics-101/

- **Errors in the DNA**
- Most are repaired
- Some are transmitted
- Some influence performance
- Some are beneficial
- Some are harmful

- **Why SNP?**
- Abundant
- Found everywhere in the genome
- Introns, Exons, Promoters
- Enhancers, Intergenic regions
- ~ 1 every 100 nucleotides

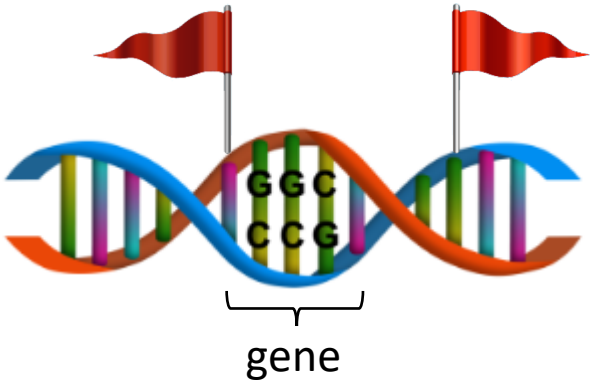
SNP tracing genes or QTL



Marker Assisted Selection - MAS

MAS

Select parents with a desired marker profile



- Few SNPs
- Meat quality
- Feed efficiency
- Disease
- Expensive!!!

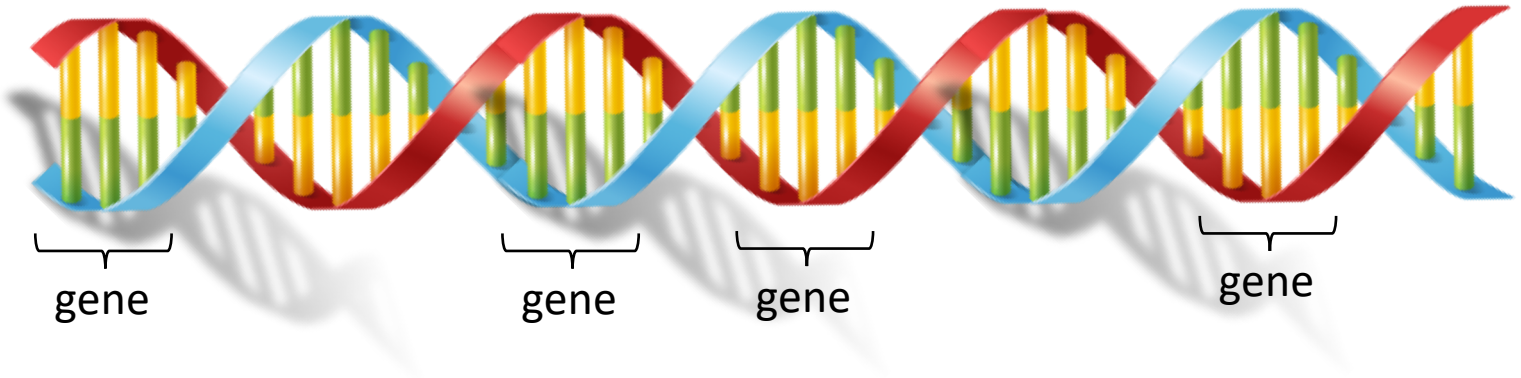
Methods to apply MAS in AB&G

Nejati-Javaremi et al. BLUP with Total allelic relationships

Fernando & Grossman BLUP to MAS

Why MAS did not quite work?

- Traits of interest are polygenic



Fisher (1918): phenotypic variation is backed up by a large number of Mendelian factors with additive effects - Infinitesimal Model

Thousands of genes → Thousands of SNP

What if we could use thousands of SNPs?

Meuwissen,
Hayes
&
Goddard

The promises…

Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

T. H. E. Meuwissen,* B. J. Hayes† and M. E. Goddard†,‡

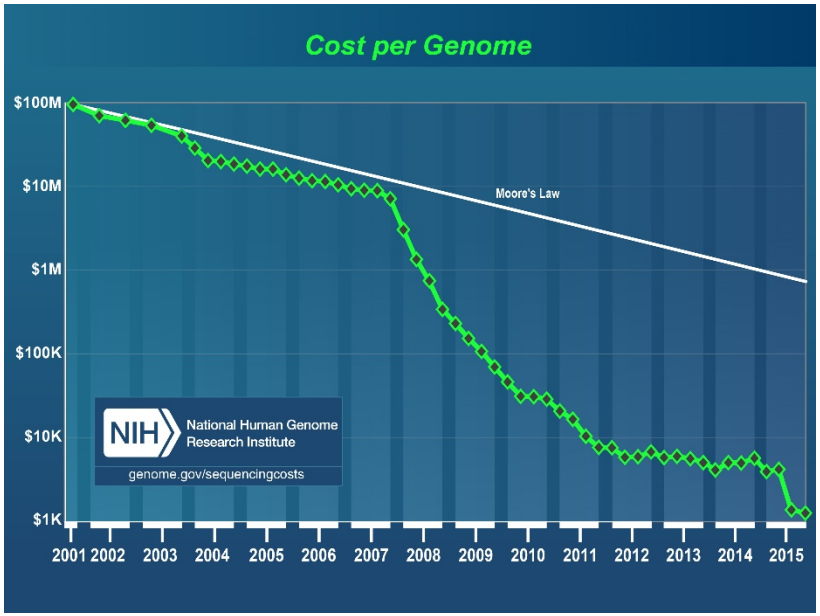
*Research Institute of Animal Science and Health, 8200 AB Lelystad, The Netherlands, †Victorian Institute of Animal Science, Attwood 3049, Victoria, Australia and ‡Institute of Land and Food Resources, University of Melbourne, Parkville 3052, Victoria, Australia

Manuscript received August 17, 2000
Accepted for publication January 17, 2001

- We can use thousands of SNPs
- Genotyping thousands of SNPs will become cheap
- We can calculate EBV based on SNPs (e.g., DGV, MBV)
  - Without own performance or progeny records
- **Accuracy of predicting EBV more than double (0.40 vs. 0.85)**
- Increase in accuracy for traits with low h^2 and hard to measure
- We can select animals earlier (reducing generation interval)

Cost of genotyping

What is 100,000 cheaper NOW than in 2001?



https://www.genome.gov/images/content/costpergenome2015_4.jpg

Peak of excitement

Human genome project = $3Bi
Bovine genome project = $53Mi

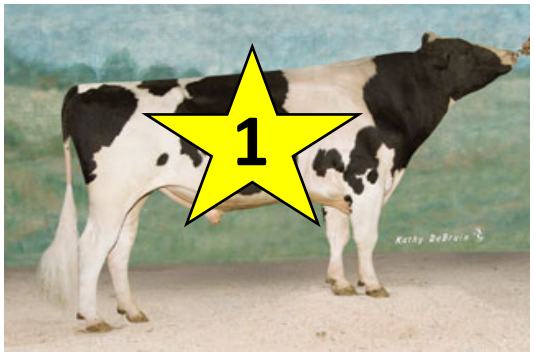
Cheaper genotyping 2009

Illumina 50k SNP chip 2009

Who would go first?

The Dairy Cattle Industry

First genomic evaluation in 2009



http://genex.crinet.com

50K SNP + parent information
No daughters with records

Net merit = $792

7 bulls > $700

Evaluation in 2012

Parent information +
100s of daughters with records

The promises…

Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

T. H. E. Meuwissen,* B. J. Hayes† and M. E. Goddard†,‡

*Research Institute of Animal Science and Health, 8200 AB Lelystad, The Netherlands, †Victorian Institute of Animal Science, Attwood 3049, Victoria, Australia and ‡Institute of Land and Food Resources, University of Melbourne, Parkville 3052, Victoria, Australia

Manuscript received August 17, 2000
Accepted for publication January 17, 2001

- We can use thousands of SNPs ✓
- Genotyping thousands of SNPs will become cheap ✓
- We can calculate EBV based on SNPs (e.g., DGV, MBV) ✓
  - Without own performance or progeny records ✓

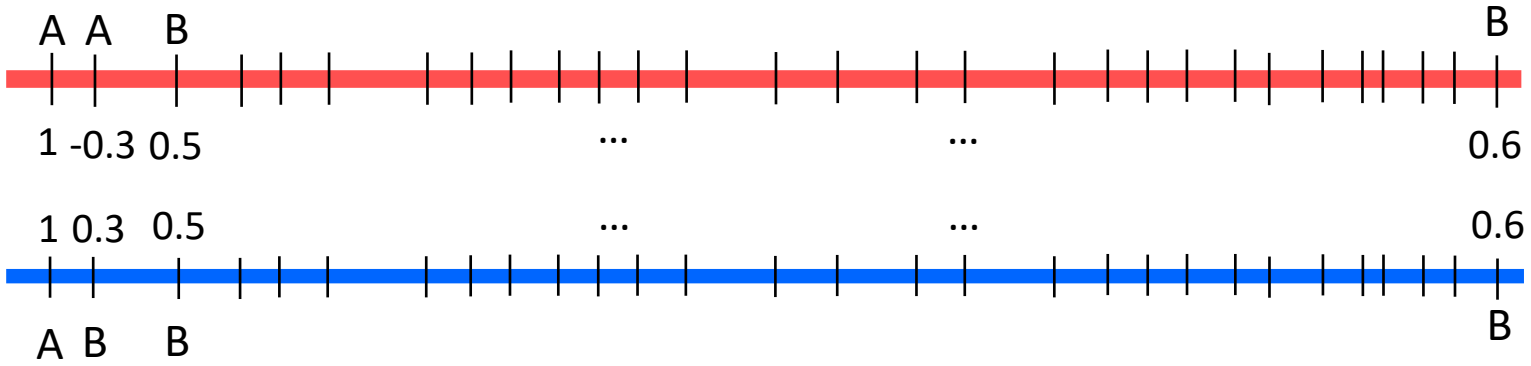
- Accuracy of predicting EBV more than double (**0.40** vs. **0.85**)
- Increase in accuracy for traits with low h^2 and hard to measure
- We can select animals earlier (reducing generation interval)

The Beef Cattle Industry

- 2009-2010: Angus
- 2012: Simmental, Hereford, Red Angus, Limousin
- 2013-2016: Charolais, Santa Gertrudis, Shorthorn, Brangus, Guelbvieh

How is genomic incorporated?

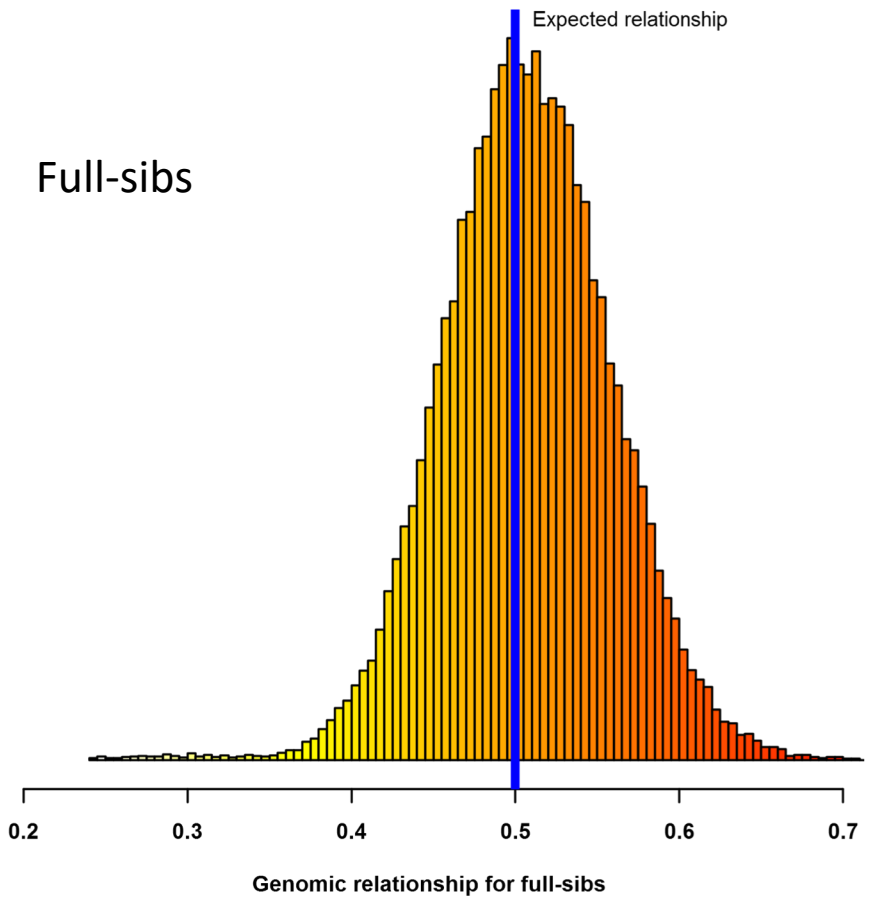
1. SNP effects: compute the effect each SNP has on the trait



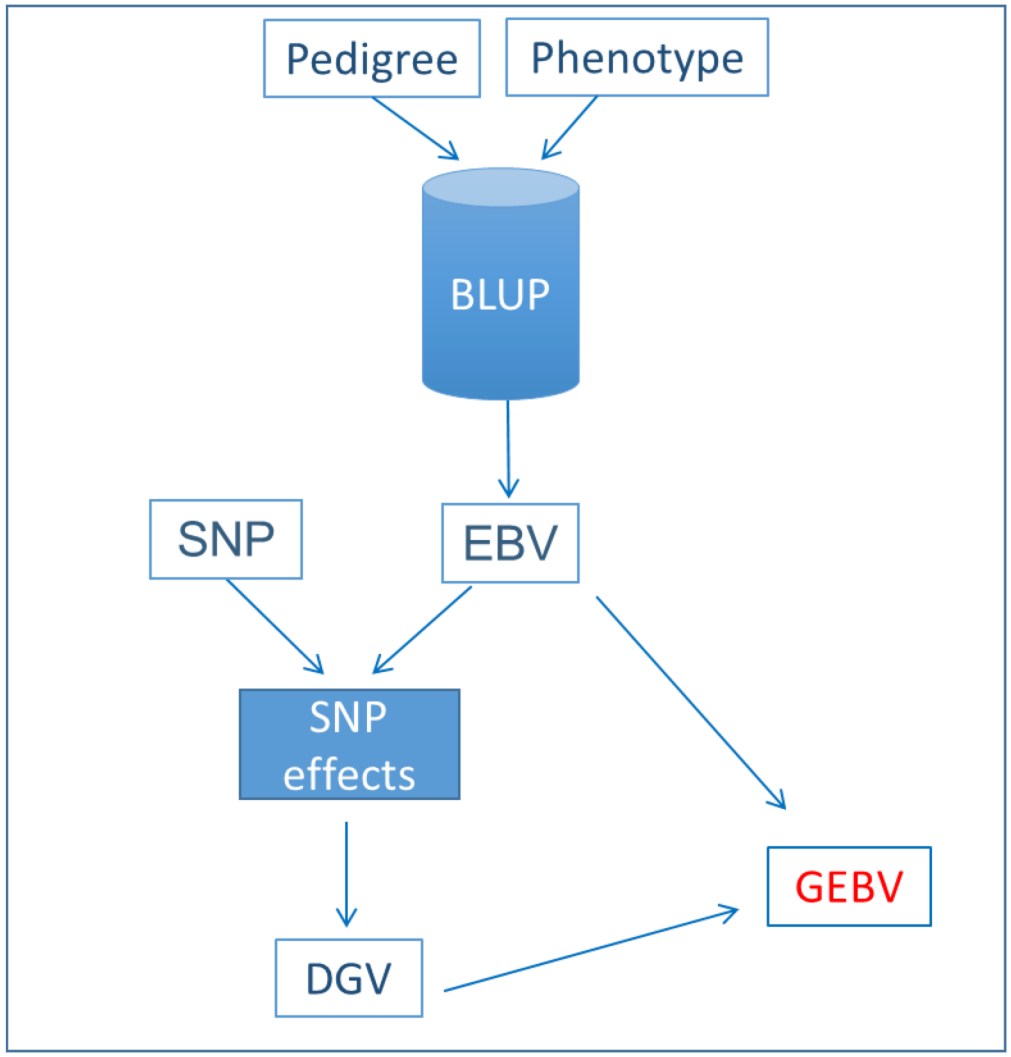
DGV = 1+1-0.3+0.3+0.5+0.5+…+0.6+0.6 = 5.7

How is genomic incorporated?

2. Better relationships: proportion of alleles shared



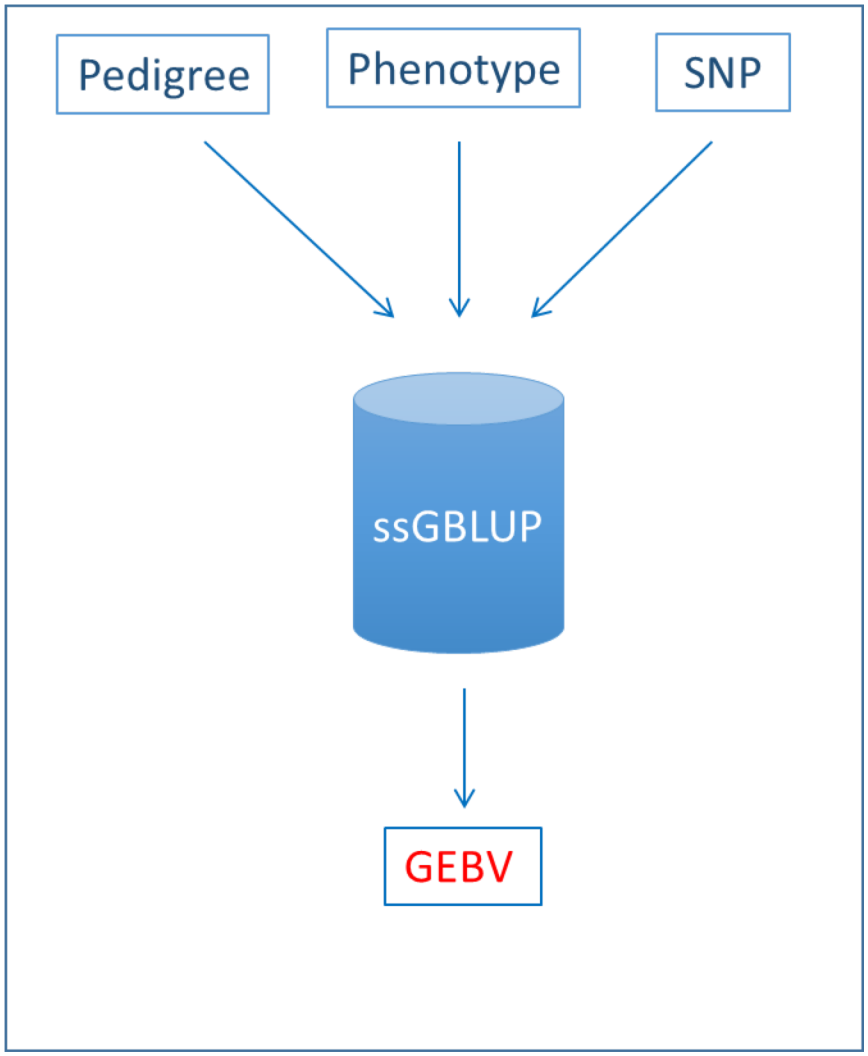
Which methods?



Multistep

first method developed and implemented for genomic selection in livestock

Which methods?

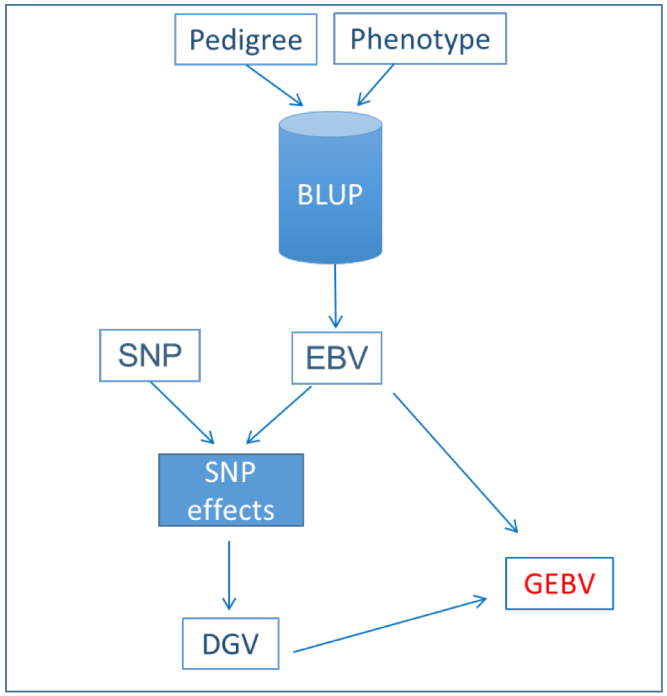


Single-step

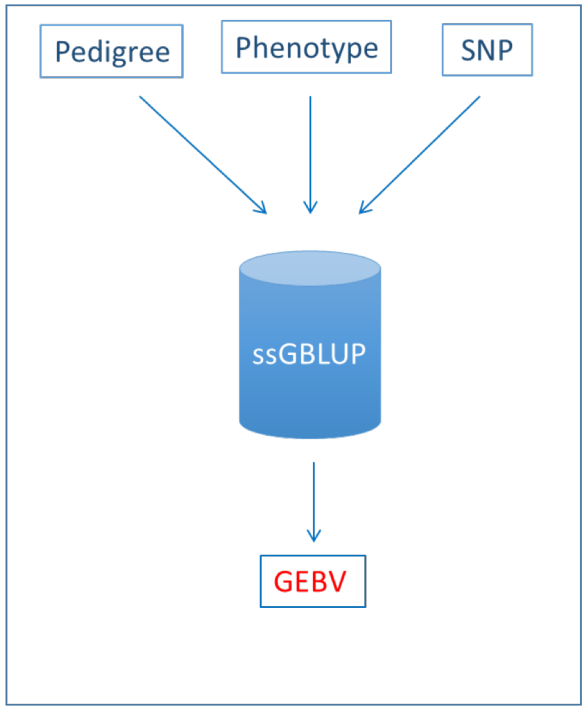
Initially developed by UGA team in 2009

Trending now

Multistep



Single-step



Simplicity

The promises…

Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

T. H. E. Meuwissen,* B. J. Hayes† and M. E. Goddard†,‡

*Research Institute of Animal Science and Health, 8200 AB Lelystad, The Netherlands, †Victorian Institute of Animal Science, Attwood 3049, Victoria, Australia and ‡Institute of Land and Food Resources, University of Melbourne, Parkville 3052, Victoria, Australia

Manuscript received August 17, 2000
Accepted for publication January 17, 2001

- We can use thousands of SNPs ✓
- Genotyping thousands of SNPs will become cheap ✓
- We can calculate EBV based on SNPs (e.g., DGV, MBV) ✓
  - Without own performance or progeny records ✓

- Accuracy of predicting EBV more than double (**0.40** vs. **0.85**) ?
- Increase in accuracy for traits with low h^2 and hard to measure
- We can select animals earlier (reducing generation interval)

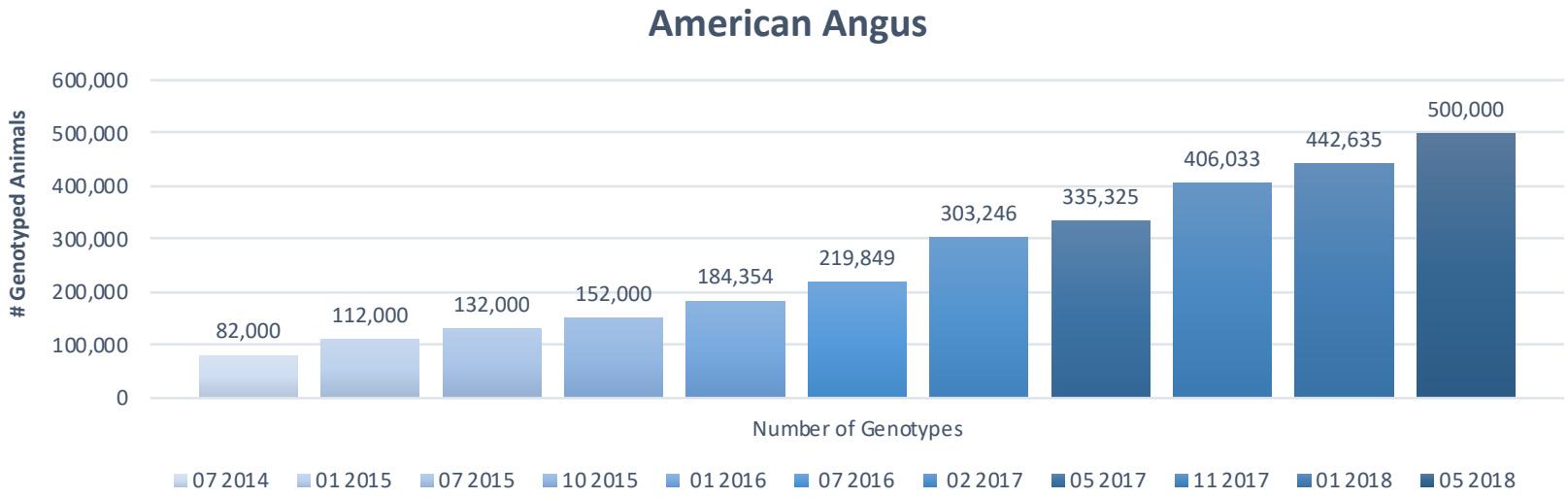
Accuracy gains

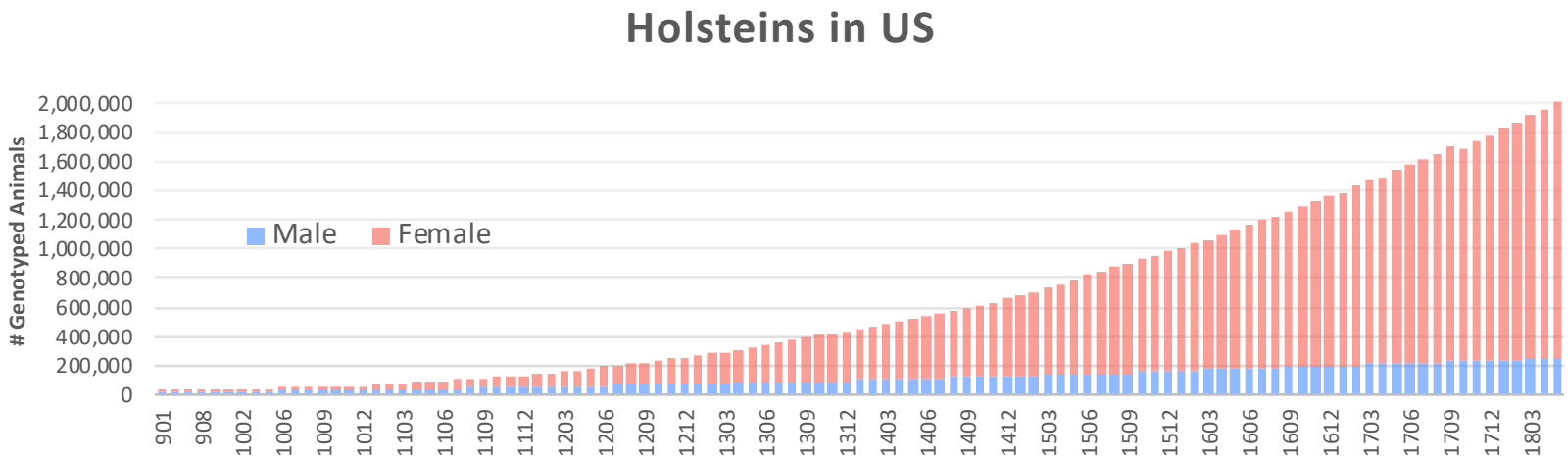
Trait	Breed	number of genotyped animals	EBV accuracy	GEBV or DGV accuracy	Gain %	Author
Simulated	-	2,000	0.40	0.84	**112**	Meuwissen et al., 2001

Small gain due to small number of genotyped animals
~ 2,000

“You should genotype more animals”

You should genotype more animals





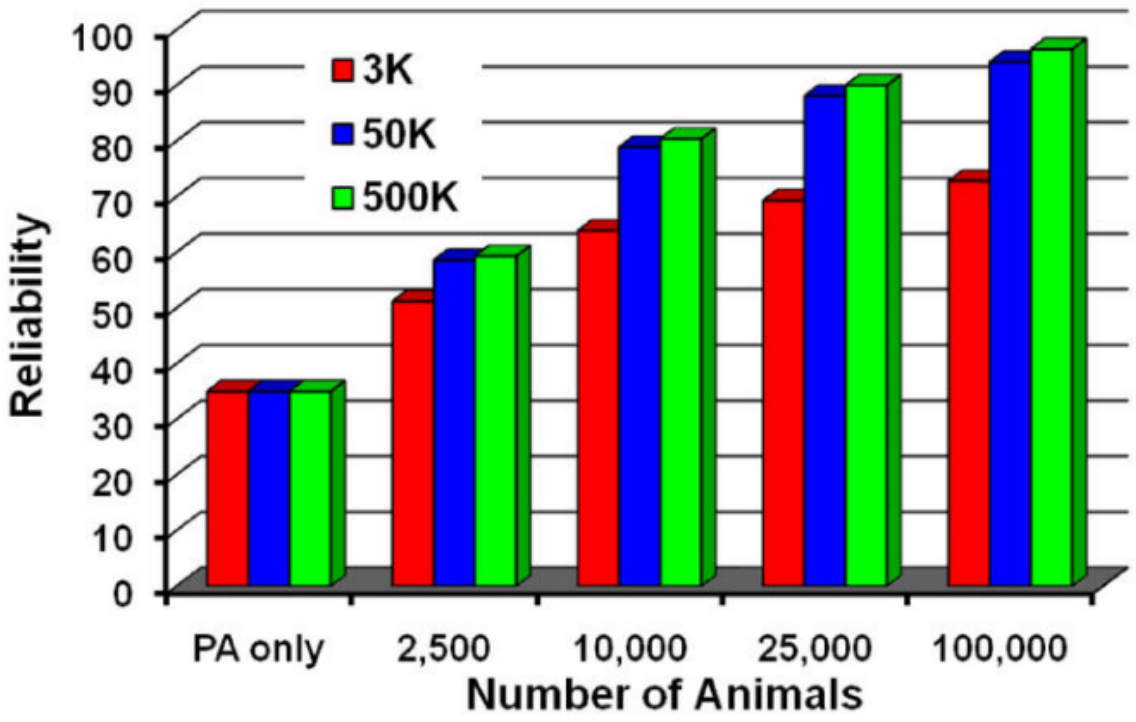
Accuracy gains

Trait	Breed	number of genotyped animals	EPD accuracy	GE-EPD or MBV accuracy	Gain %	Author
Simulated	-	2,000	0.40	0.84	**110**	Meuwissen et al., 2001
Growth	Angus	2,000	0.29	0.32	10	Lourenco et al., 2015
Growth	Angus	33,000	0.29	0.35	21	Lourenco et al., 2015

"You should genotype more animals"

"You are using only 50k SNP… not enough… you should use over 300,000"

You should use more SNP

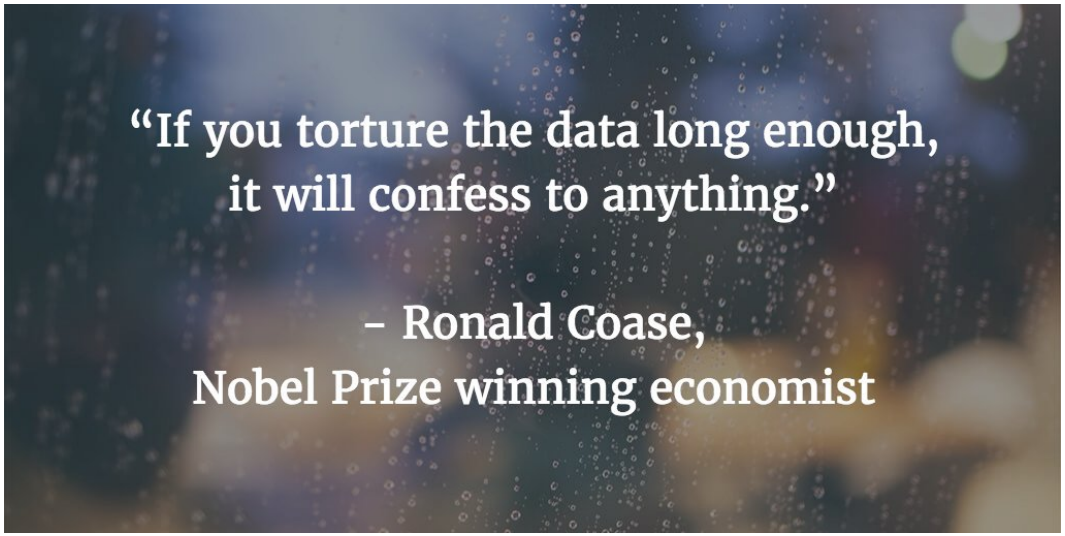


Gain = 0.02

VanRaden et al., 2011

"You are using only 50k SNP… not enough…
you should u[illegible]ver 300,000"

Sequence the whole genome



Sequence vs. genotyping

> 30M

50k

Sequence information for predictions

Overall gains in REL

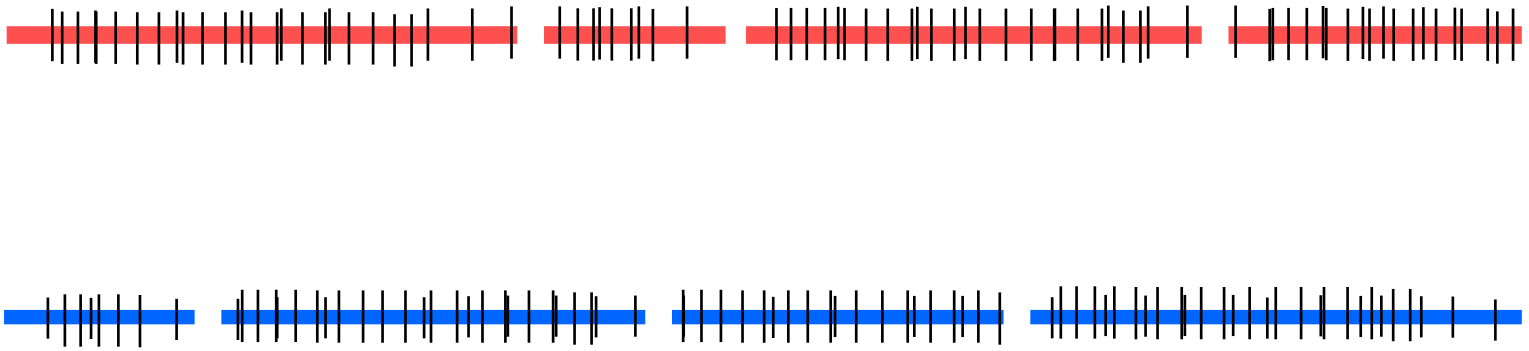
Trait group	60K + 17K
Production	1.5
Health	2.5
Calving	3.3
Type	3.2
All traits	2.7

VanRaden et al., 2017

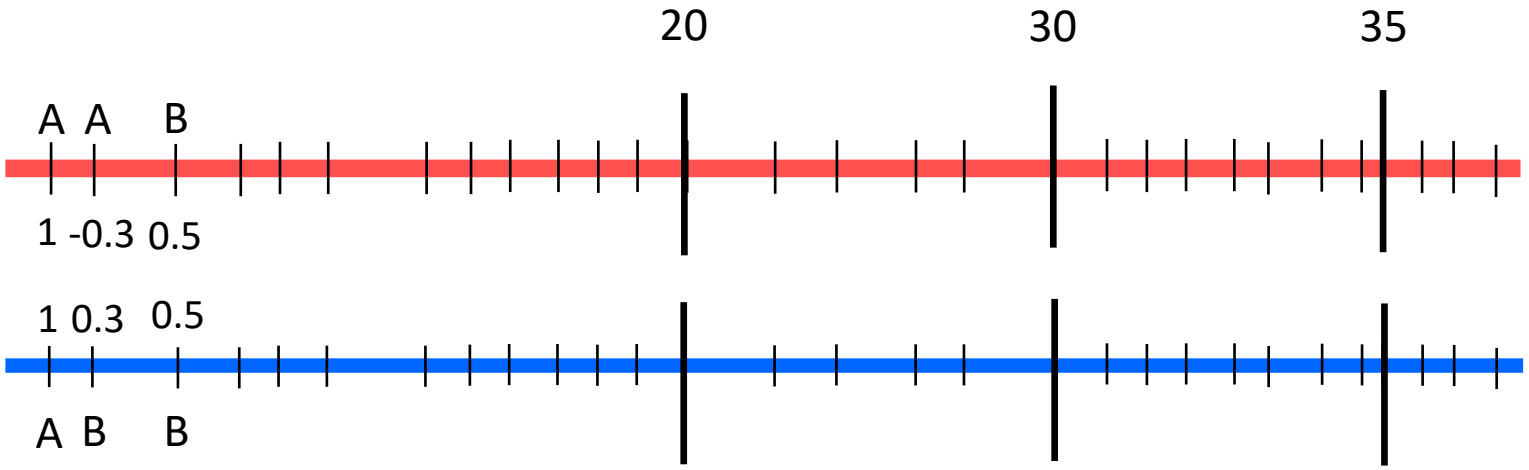
Small gain with more SNP

1. Better relationships: already accurate with 50k

2. SNP effects: only more SNP to estimate effects without increasing phenotypes



Why Meuwissen et al. (2001) got it but we did not?



- Assumed few SNP with large effect
- Large SNP explained large proportion of genetic variance

Traits of interest are polygenic: several genes with small effect

The promises…

Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

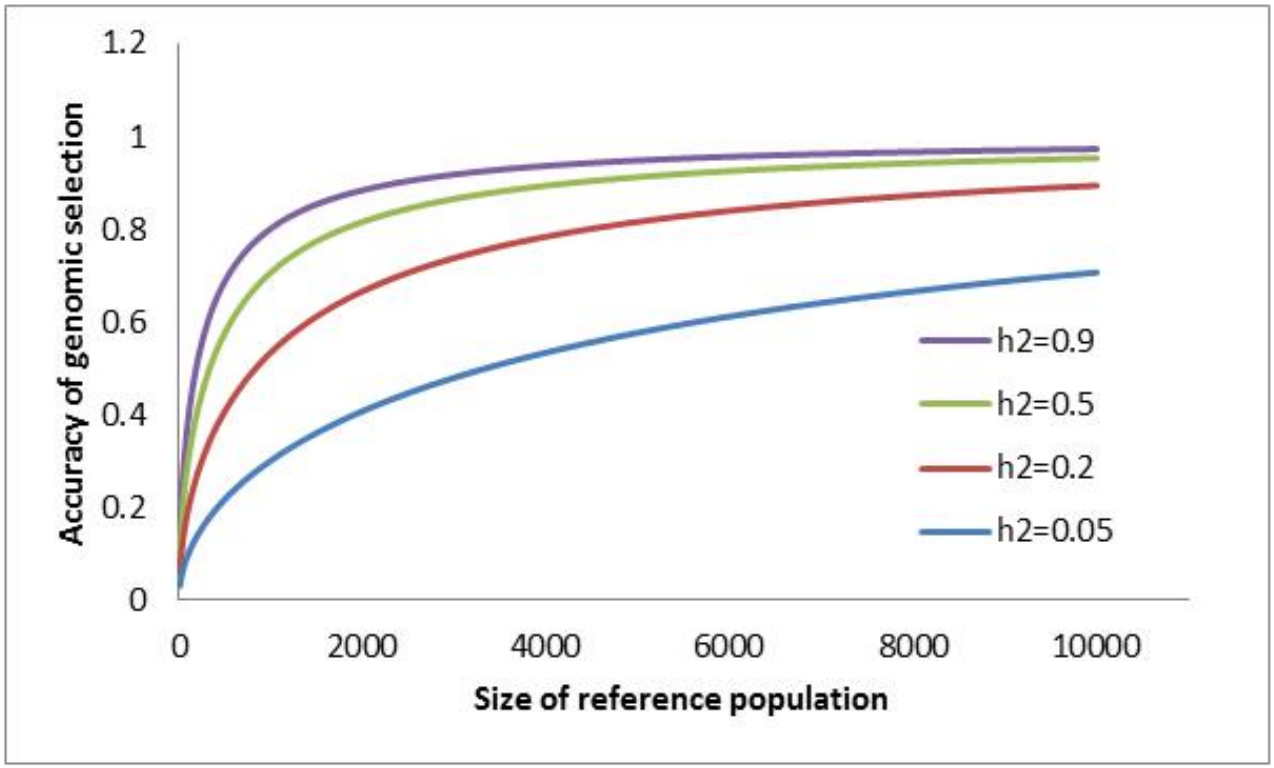
T. H. E. Meuwissen,* B. J. Hayes† and M. E. Goddard†,‡

*Research Institute of Animal Science and Health, 8200 AB Lelystad, The Netherlands, †Victorian Institute of Animal Science, Attwood 3049, Victoria, Australia and ‡Institute of Land and Food Resources, University of Melbourne, Parkville 3052, Victoria, Australia

Manuscript received August 17, 2000
Accepted for publication January 17, 2001

- We can use thousands of SNPs ✓
- Genotyping thousands of SNPs will become cheap ✓
- We can calculate EPD based on SNPs (e.g., DGV, MBV) ✓
  - Without own performance or progeny records ✓
- Accuracy of predicting EBV Increases ✓
- Increase in accuracy for traits with low h^2 and hard to measure ✓
- We can select animals earlier (reducing generation interval)

Increase in accuracy for traits with low h^2 and hard to measure



Kor Oldenbroek and Liesbeth van der Waaij, 2015

- Increase depends on the number of genotypes and phenotypes

The promises…

Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

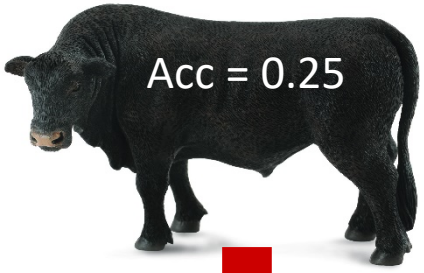
T. H. E. Meuwissen,* B. J. Hayes† and M. E. Goddard†,‡

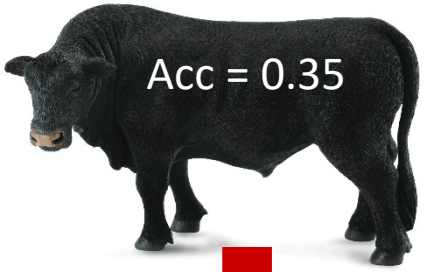
*Research Institute of Animal Science and Health, 8200 AB Lelystad, The Netherlands, †Victorian Institute of Animal Science, Attwood 3049, Victoria, Australia and ‡Institute of Land and Food Resources, University of Melbourne, Parkville 3052, Victoria, Australia

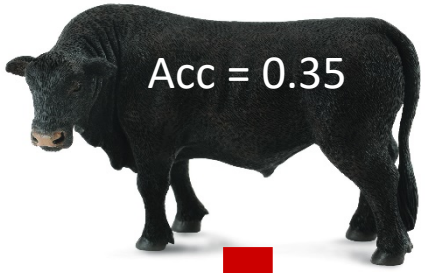
Manuscript received August 17, 2000
Accepted for publication January 17, 2001

- We can use thousands of SNPs
- Genotyping thousands of SNPs will become cheap
- We can calculate EPD based on SNPs (e.g., MBV)
  - Without own performance or progeny records
- Accuracy of predicting EBV Increases
- Increase in accuracy for traits with low h^2 and hard to measure
- We can select animals earlier (reducing generation interval)

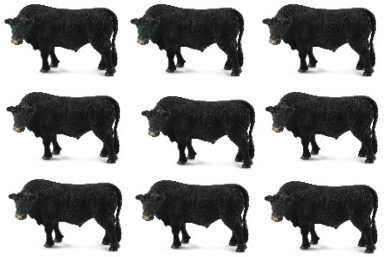
We can select animals earlier







Parent Average

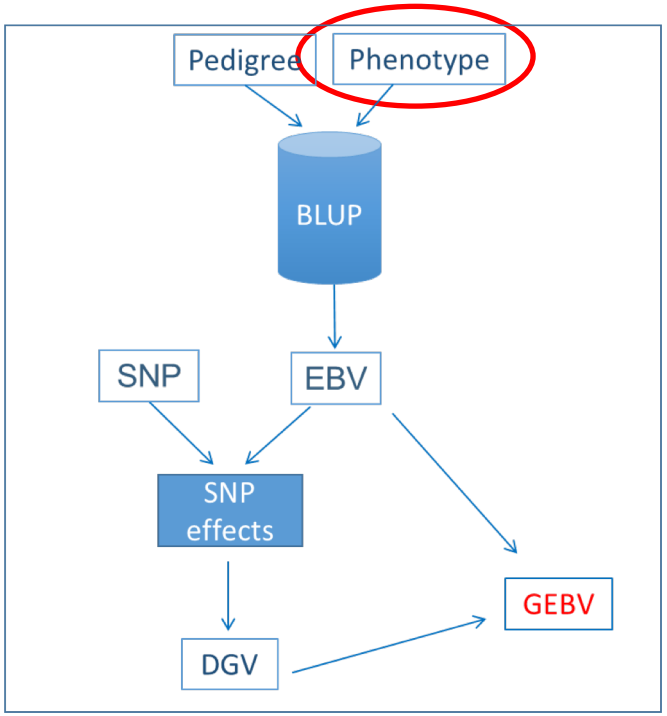


$$\Delta G = \frac{i\, r\, \sigma_a}{L}$$

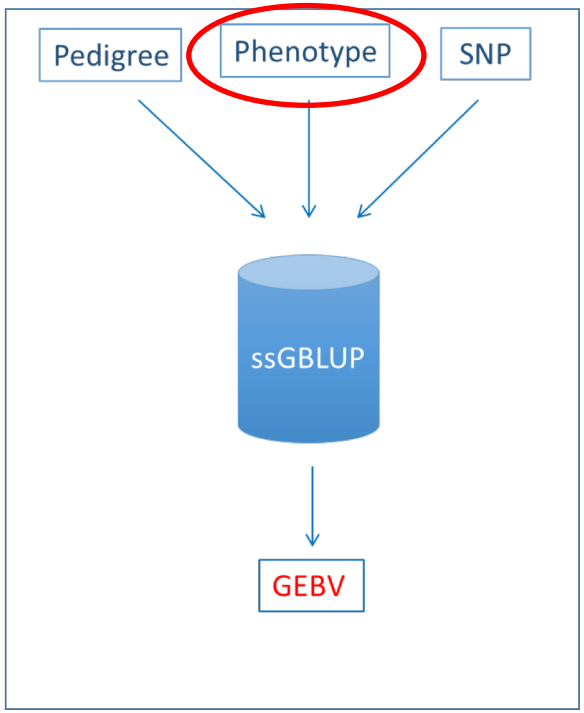
Does it mean we do not need to collect phenotypes?

There is no magic here

Multistep



Single-step



Genotype

Phenotype

Millions of genotyped animals
How is it possible?

More information = higher accuracy

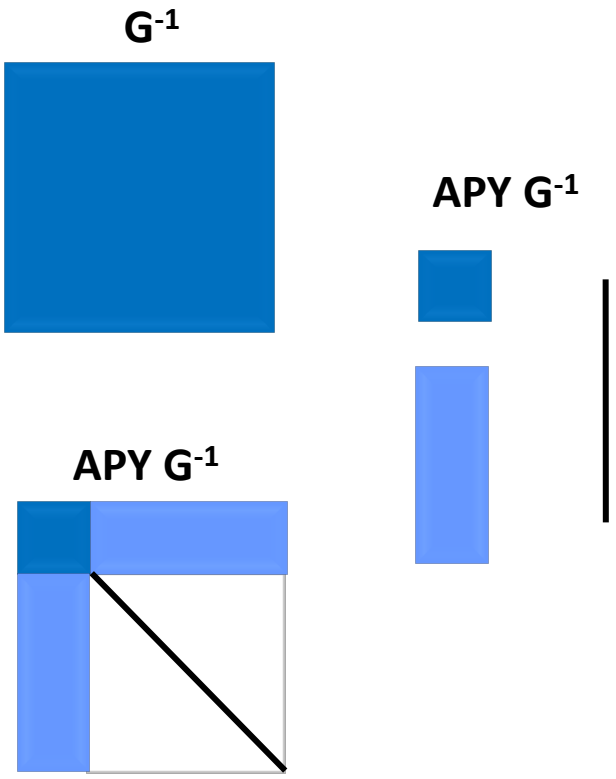
More genotypes, phenotypes, pedigree

Challenge

Millions of genotyped animals

- Is it possible to use genotypes for millions of animals?

APY – Algorithm for Proven and Young



Misztal et al., 2014

- US Holstein type trait data
  - 18 trait-model
  - 13.6M animals in pedigree
  - 10.2M phenotypes
  - 2.3M genotyped animals
- APY ssGBLUP with 15k core
  - 1 day to build $\mathbf{G}_{\mathrm{APY}}^{-1}$ and $\mathbf{A}_{22}^{-1}$
  - ~2.5 days to converge
  - 1000 PCG rounds

Tsuruta et al. (2019)

Keep in mind

- Idea of using genomics in Breeding & Genetics is not new
- Initial studies were driven by Meuwissen et al. 2011
- Lower genotyping cost was essential for the adoption
  - Dairy, Beef, others
- Promises were higher than the realized
- But still a great improvement in accuracy
- Reduced generation interval
  - ~20% to 30% genetic gain
- Genomic information set new standards in Breeding & Genetics