Genomic selection: use of DNA data in animal breeding

Theo Meuwissen

Norwegian University of Life Sciences, Ås, Norway.

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, †Vict... stitute of Animal Science, Attwood 3049, Victoria, Australia and ‡Institute of Land and F... University of Melbourne, Parkville 3052, Victoria

Manuscript received August 17...
Accepted for publication Jan...

AB...

Recent advances in molecular genetic tech... ...nse marker maps available and genotyping many individuals for these markers feasi... ...pted to estimate the effects of ~50,000 marker haplotypes simultaneously from a... ...phenotypic records. A genome of 1000 cM was simulated with a marker spacin... ...kers surrounding every 1-cM region were combined into marker haplotypes. Due to... ...e (N_e = 100), the marker haplotypes were in linkage disequilibrium with the QTL ... markers. Using least squares, all haplotype effects could not be estimated simul... ...y the biggest effects were included, they were overestimated and the accuracy of pre... ...values of the offspring of the recorded animals was only 0.32. Best linear unbiased predict... ...plotype effects assumed equal variances associated to each 1-cM chromosomal segment, which yie... ...d an accuracy of 0.73, although this assumption was far from true. Bayesian methods that assumed a prior distribution of the variance associated with each chromosome segment increased this accuracy to 0.85, even when the prior was not correct. It was concluded that selection on genetic values predicted from markers could substantially increase the rate of genetic gain in animals and plants, especially if combined with reproductive techniques to shorten the generation interval.

The beginning...

GS: a paradigm shift in animal breeding

- Previous century selection technology:
  - –for phenotype
  - –combine with info from relatives (BLUP)
- 2000-2007, few breeding companies:
  - –Marker Assisted Selection (MAS)
  - –Selected for some genes (QTL)
- 2007+ :
  - –Genomic selection (first in dairy)

AIM

- Review the use of DNA information in animal breeding
  - –How?
  - –Why?
  - –Level : general (non-breeding) animal scientists

- Peek into the future where animal breeding is going

Why was MAS not the solution?

- Recipe for MAS:
  - Find genes (QTL) underlying traits
  - Select for positive QTL effects
- Problems:
  - Traits of interest are complex
    - Many small genes + environment
  - Detected QTL explained only small fraction of Vg
    - limits value of MAS
    - (traditional) selection for majority of genes stays important
    - In human genetics : missing heritability

Low uptake of MAS

Three breakthroughs

1. Detection of large numbers of SNP markers
   – Byproduct of sequencing efforts
2. SNP-chip genotyping technology
   – Affordable to genotype 50,000 SNPs
3. Development of GS / GP technology

Genomic Selection in a nutshell

Reference Population

- Phenotyped
- Genotyped (~100,000 SNPs)

$$Weight = b_1 x_1 + \cdots + b_{100000} x_{100000}$$

Prediction Equation

Selection Candidates

- Only genotyped

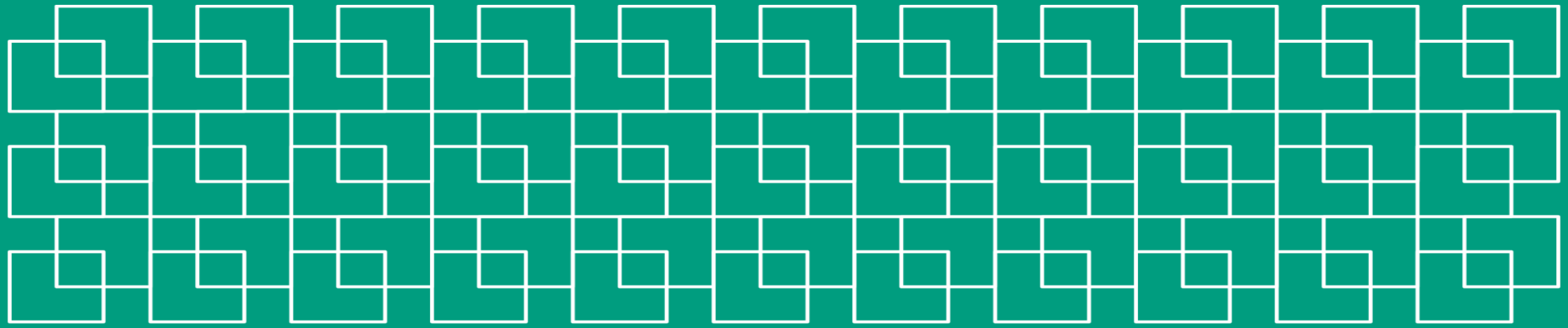
$$Expected(Weight) = b_1 x_1 + \cdots + b_{100000} x_{100000}$$

Predicted Weight based on genomic data alone

Some key features

- SNP effects are not tested for statistical significance
  - –(Many tests implies very stringent testing)
  - –All SNPs effects are used => all genetic variance is addressed
- Accurate GEBV for animals without records
  - –Decouples elite breeding from recording
    - Invasive recordings (slaughter traits; disease challenges; poor environm.)
    - Late in life records (slow down the breeding scheme)
    - Sex limited traits (milkproduction; fertility)
- No pedigree recording needed

Statistical Methods for GS



SNP-BLUP

- Statistical model to estimate SNP effects:

$$y_i = \mu + X_{1i}*b_1 + X_{2i}*b_2 + \ldots + X_{50000i}*b_{50000} + e_i$$

  - Bayesian statistics: use prior information i.e. $b_i \sim N(0,\sigma^2)$
    - σ^2 is same for all SNPs
- Prediction of genomic estimates of breeding value:

$$GEBV_j = X_{1j}*\hat{b}_1 + X_{2j}*\hat{b}_2 + \ldots + X_{50000j}*\hat{b}_{50000}$$

GBLUP

- is traditional BLUP but family relationships come from SNPs instead of from pedigree
  - Relationship matrix A is replaced by G
  - Instead of traditional EBV we get GEBV
- GBLUP is equivalent to SNP-BLUP
  - With carefull parameterization the GEBV are the same
  - GBLUP is computationally easier (no of equations = no of genotyped animals)

Non-linear methods for GS

- Called BayesA, BayesB, BayesC etc.
  - More sophisticated prior information
- Model for BayesB:

$$y_i = \mu + l_1 * X_{1i} * b_1 + l_2 * X_{2i} * b_2 + \ldots + l_{50000} * X_{50000i} * b_{50000} + e_i$$

- l_j=0/1 is indicator whether SNP j has effect or not
- Prior information is that fraction π of SNPs have effect & (1-π) have no effect
- Try to give extra weight to some SNPs and less to others

GBLUP vs. BayesA/B/C

- Simulation studies:

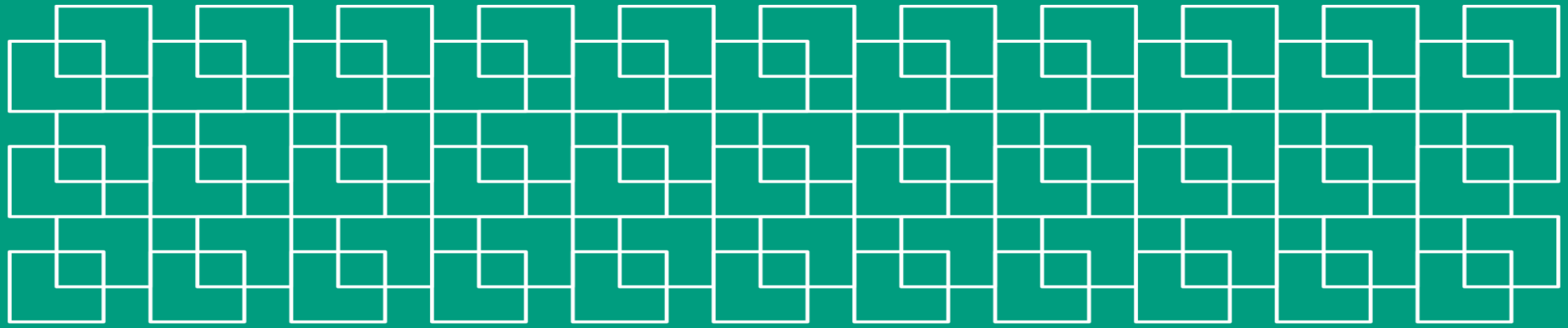
  GBLUP < BayesA < BayesB/C

- Real data studies:
  - Differences small
  - Often GBLUP is as good as BayesA/B
- Theoretically:
  - If few genes & dense SNPs: BayesB best
  - If many genes / not enough density: GBLUP best

Whole genome sequence data (WGSD)

- Affordable? ~1000 € /genome
  - Costs are coming down
- QTL mutations are amongst the SNPs
  - But: hidden amongst ~12 million SNPs
  - Need e.g. BayesB
    - Expect large improvements using BayesA,B,C,etc
- However if population size is small
  - Confounding between the SNPs make it impossible to identify the QTL
  - Research is needed to avoid this problem of WGSD

Applications of GS



General idea for any ’difficult’ trait

- Perform experiment: phenotype and genotype
- Estimate SNP effects
- Select for SNP effects forever after
- Turned out too good to be true:
  - Reference population needs to be huge
  - Reference population needs to be updated with young animals

Dairy cattle breeding

- Selection of young bulls /no progeny test (Schaeffer 2006):
  - Saves costs of progeny test: ~40,000 $/bull
  - Reduce generation interval by factor 2
  - Also use GS on bull dams
  - double ΔG
  - Saves $23 million / yr in Canada

- Breeding companies got very interested
  - Double gain and less costs

Dairy cattle: what happened in practice?

- Preselection of test-bulls
  - –Bulls that entered progeny test were preselected on GEBV
  - –Always better than random selection
  - –No costs reduction / only extra genotyping costs
- Abandon formal progeny test
  - –Costs reductions are realised
  - –Genetic gains were markedly increased
    - Although not doubled

US Holstein population

- 2003 data predicting progeny-proofs of 2008 bulls:

	Reliability:	
	Traditional	GenomicSel
MilkYield	28%	49%
FatYield	15%	44%
ProteinY	27%	47%

VanRaden, 2008

- Large increases in accuracy were realised

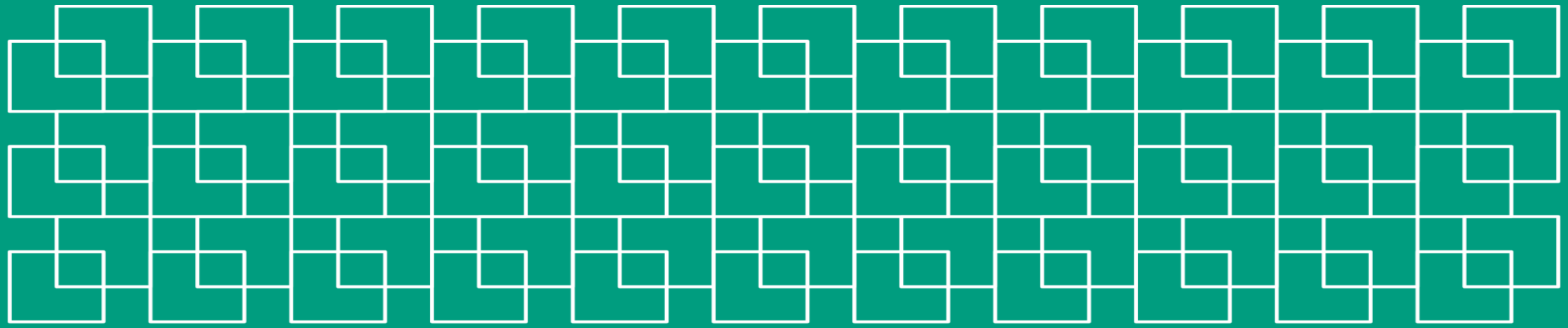
Need large reference populations

- North American Holsteins: >15,000 phen+genot bulls
  - Bulls have ‘daughter average’ as phenotype
  - Collaboration US and Canada
- Europe: >19,000 phen+genot Holstein bulls
  - France, Germany, Holland, Scandinavia, Ireland
- Accuracies: ~90%
- Recently: 10,000s of cows
- Small breeds: across breed GS

GS in pig breeding

- Main traits recorded in boar test before selection
  - No generation interval effect
- Main interests:
  - Select boars for sow-traits
  - Select purebred nucleus animals for crossbred performance
    - Estimate SNP effects in crossbreds under practical circumstances

Towards the future : Precision breeding



Precision breeding (Flint & Woolliams, 2007)

1. Accurate GEBV
2. Avoid deleterious side effects from breeding
   - Select for broad breeding goal
3. Manage genetic variation
   - Realize precision breeding also into the future

Ad 1: Accurate GEBV

- Accuracy > 90%
- Large scale genotyping and phenotyping
  - In the age of GS phenotype is king (Mike Coffey)
  - Make GS work accross large genetic distances

Ad 2: Broad breeding goal

- Phenomics to predict many traits
  - –Use of novel recording technologies
    - On a large scale on practical data
    - E.g: IR, CT, gene-expression data
  - – Combine with genotypes to get GEBV
    - For broad spectrum of traits for all animals

Ad 3: manage genetic variation

- Optimum Contribution Selection:
  - –Maximises genetic progress
  - –Manages the inbreeding
    - At the level of the DNA (Sonesson et al, 2012)
    - Using high density SNP data

Conclusions

- Genomic selection caused a paradigm shift in animal breed.
- Traits of interest are very complex
  - –Many thousands of genes + environment
- Can design very novel breeding structures
  - – Decouples accurate recording from elite breeding animals
- Large scale intensive recording remains key
- In the future : realise precision breeding
  1. Accuracy > 90%
  2. Broad breeding goal
  3. Manage genetic variation

Thanks to:

- Mike Goddard, Ben Hayes, John Woolliams, Roel Veerkamp, Anna Sonesson, Jørgen Ødegård, Marie Lillehammer, Matt Baranski, Christian Sørensen, Odd Vangen, Gunner Klemetsdal, Tormod Ådnøy, Hans-Magnus Gjøen, Torunn Åsmundstad, Borghild Hillestad, Eli Grindflek, Dan Olsen, Ina Ranberg, Eli Gjerlaug, Øyvind Nordby, Trygve Solberg, Bjørg Heringstad, Morten Svendsen, Xijiang Yu, Tu Luan, Kristine Martinsen, Sini Wallen, Kahsay Nirea, Binyam Dagnachew, Oscar Iheshuilor, Gebreyohans Tesfaye, and many more…
- Sponsors: Norwegian Research Council, EU and many industry partners.