Fundamentals of Genomic Selection

Jack Dekkers

Animal Breeding & Genetics
Department of Animal Science
Iowa State University
Past and Current Selection Strategies

Black box of Genes

Quantitative genetics

h² → Phenotype

Environment

Estimated Breeding Value

BLUP

Phenotype of relatives

selection
This approach has been very successful for many traits.
US Holsteins – Daughter Pregnancy Rate

Breeding value (%)

Phenotypic base = 21.53%

Holstein birth year


AIPL USDA
and has important limitations

E.g. Need to select Bulls by Progeny Test

E.g. Need to select Bulls by Progeny Test

X

Embryo Transfer

Which is best??

5 years and $$$$ later

Superior progeny tested bull

Limitations:
- Long generation intervals
- High cost of progeny test
- Difficult to improve low heritable traits (fertility, disease resistance)
'70 – '00: Promise of Molecular Genetics

Find major genes or markers linked to QTL and use these for Marker-Assisted Selection

<table>
<thead>
<tr>
<th>Mean weight (kg)</th>
<th>Effect 'G' allele = +5</th>
</tr>
</thead>
<tbody>
<tr>
<td>105</td>
<td>100</td>
</tr>
<tr>
<td>95</td>
<td>100</td>
</tr>
</tbody>
</table>

Effect ‘G’ allele = +5 = effect of # G alleles
Use of MAS to enhance Selection

- **Genes**
  - Major genes
  - Markers
  - QTL

- **Molecular genetics**

- **Phenotypic data**

- **Molecular data**

- **Marker-Assisted Selection**

- **BLUP**

- **Expressed in both sexes**
- **Expressed at early age**
- **Requires less phenotypic data**
Potential gains from MAS in livestock

Meuwissen & Goddard, 1996 (GSE)
QTL with 1/3 of genetic variance haplotype-marked

\[ h^2 = .27 \]

MAS is most beneficial for ‘difficult’ traits

<table>
<thead>
<tr>
<th>Generation</th>
<th>Carcass trait</th>
<th>Sex-limited trait</th>
<th>Phenotyped before selection</th>
<th>Phenotyped after selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>38</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>62</td>
<td>62</td>
<td>62</td>
</tr>
<tr>
<td>4</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>5</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>39</td>
</tr>
</tbody>
</table>

Extra response from MAS (%)
3 types of marker loci for MAS

**Direct markers**
- Functional mutations
  - known genes

**LD-markers**
- in pop.-wide Linkage Disequilibrium with QTL
  Marker-QTL linkage phase
  ~consistent across population

**LE-markers**
- used on a within-family basis
  - in pop.-wide Linkage Equil. with QTL
  Marker-QTL linkage phase NOT consistent across families
Many markers and QTL have been reported but few have been utilized.

Examples of gene tests in commercial breeding:

D = dairy cattle  
B = beef cattle  
C = poultry  
P = pigs  
S = sheep


<table>
<thead>
<tr>
<th>Trait</th>
<th>Direct markers</th>
<th>LD markers</th>
<th>LE markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congenital defects</td>
<td>BLAD (D)</td>
<td></td>
<td>RYR (P)</td>
</tr>
<tr>
<td></td>
<td>Citrulinaemia (D,B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DUMPS (D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CVM (D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maple syrup urine (D,B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mannosidosis (D,B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RYR (P)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appearance</td>
<td>CKIT (P)</td>
<td></td>
<td>Polled (B)</td>
</tr>
<tr>
<td></td>
<td>MC1R/MSHR (P,B,D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MGF (B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milk quality</td>
<td>κ-Casein (D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>β-lactoglobulin (D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FMO3 (D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meat quality</td>
<td>RYR (P)</td>
<td>RYR (P)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RN/PRKAG3 (P)</td>
<td>RN/PRKAG3 (P)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A-FABP/FABP4 (P)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H-FABP/FABP3 (P)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CAST (P, B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;15 PICmarq™ (P)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feed intake</td>
<td>MC4R (P)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease</td>
<td>Prp (S)</td>
<td>B blood group (C)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F18 (P)</td>
<td>K88 (P)</td>
<td></td>
</tr>
<tr>
<td>Reproduction</td>
<td>Booroola (S)</td>
<td>Booroola (S)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inverdale(S)</td>
<td>ESR (P)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hanna (S)</td>
<td>PRLR (P)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RBP4 (P)</td>
<td></td>
</tr>
<tr>
<td>Growth &amp; composition</td>
<td>MC4R (P)</td>
<td>CAST (P)</td>
<td>QTL (P)</td>
</tr>
<tr>
<td></td>
<td>IGF-2 (P)</td>
<td>IGF-2 (P)</td>
<td>QTL (B)</td>
</tr>
<tr>
<td></td>
<td>Myostatin (B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Callipyge (S)</td>
<td>Carwell (S)</td>
<td></td>
</tr>
<tr>
<td>Milk yield &amp; composition</td>
<td>DGAT (D)</td>
<td>PRL (D)</td>
<td>QTL (D)</td>
</tr>
<tr>
<td></td>
<td>GRH (D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>κ-Casein (D)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reasons for limited use of MAS in livestock

- # markers available is limited
- Markers only explain limited % of genetic variance
  - Only QTL with moderate – large effects detected
- Genotyping costs
- Marker/QTL effects are not consistent / not transferable to commercial breeding populations
  - ‘Beavis’ effect – effects of ‘significant’ markers tend to be overestimated
  - Marker effects were estimated within families or in experimental crosses
  - Interactions of marker/QTL effects with genetic background and / or environment
  - Inconsistent marker-QTL LD across populations
A Revolution in Molecular Technology

Since 2000: High-through-put SNP genotyping

NOW AVAILABLE: Illumina Bovine 50k Beadchip

50,000 DNA tests for <$250

Genomic Selection

International Swine Genome Sequencing Consortium
Nature 2004
2.8 million SNPs + discovery of many Single Nucleotide Polymorphisms

AAGCCTTGATAATT
maternal

AAGCCTTGCTAATT
paternal

Bovine Genome Project
How to use high-density SNP data?

**Statistical Analysis to detect QTL / estimate SNP effects**

- Genotype large # of Individuals
- for large numbers of SNPs
- + collect their phenotypes

Use only significant SNPs for MAS

Allows detecting more LD markers but still suffers from only using significant markers
- Small effects are missed
- Beavis effect
Solution: Genomic selection

Genetic Evaluation using high-density SNPs

- All SNPs are fitted simultaneously, i.e. 50,000 vs. 1 at a time
- SNP effects are fitted as random vs. fixed effects
  - enables all SNPs to be fitted simultaneously
  - shrinks SNP effect estimates to 0 depending on evidence from data

\[ y_i = \mu + \sum_{SNP_k} \beta_k g_{ik} + e_i \]

Estimates of SNP effects \( \hat{\beta}_k \)

Implemented using a variety of Bayesian methods (Bayes-A, -B, -C)
Or by using genomic vs. pedigree relationships in animal model BLUP (GBLUP)

Use to estimate breeding value of new animals based on genotypes alone
Genomic EBV = \( \sum \hat{\beta}_k g_{ik} \)
Example Genomic EBV based on 3 SNPs

with estimated effects ($\beta$ for # A alleles) of:

+10 for SNP 1
+5 for SNP 2
-10 for SNP 3

<table>
<thead>
<tr>
<th>Individual</th>
<th>SNP 1</th>
<th>SNP 2</th>
<th>SNP 3</th>
<th>Genomic Breeding Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Genotype</td>
<td>Value</td>
<td>Genotype</td>
<td>Value</td>
</tr>
<tr>
<td>1</td>
<td>AA</td>
<td>10</td>
<td>AA</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>AA</td>
<td>10</td>
<td>AA</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>AB</td>
<td>0</td>
<td>BB</td>
<td>-5</td>
</tr>
<tr>
<td>4</td>
<td>AB</td>
<td>0</td>
<td>BB</td>
<td>-5</td>
</tr>
<tr>
<td>5</td>
<td>BB</td>
<td>-10</td>
<td>AA</td>
<td>5</td>
</tr>
</tbody>
</table>
Genomic selection
Genetic Evaluation using high-density SNPs

Estimate marker effects

Predict BV from marker genotypes at early age

Select at young age

Training data

Phenotype

Genotype
for >50,000 SNPs

Genotype
for >50,000 SNPs

Genotype
for >50,000 SNPs

Meuwissen et al. 2001
Genomic EBV have greater reliability for young bulls and heifers than Parent Average EBV

E.g. for Young Holstein Bulls

(VanRaden and Tooker, 2009 USDA-AIPL)

ftp://aipl.arsusda.gov/pub/outgoing/GenomicReliability0608.doc

<table>
<thead>
<tr>
<th>Trait</th>
<th>Gain over parent average reliability (~39%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net merit</td>
<td>+ 23</td>
</tr>
<tr>
<td>Milk yield</td>
<td>+ 32</td>
</tr>
<tr>
<td>Fat yield</td>
<td>+ 36</td>
</tr>
<tr>
<td>Protein yield</td>
<td>+ 28</td>
</tr>
<tr>
<td>Productive life</td>
<td>+ 33</td>
</tr>
<tr>
<td>Dtr. Pregnancy rate</td>
<td>+ 20</td>
</tr>
</tbody>
</table>
How will Genomic Selection change Dairy Cattle Breeding?

- AI Studs will market young bulls / bull teams selected on Genomic EBV

- These young bulls will be from ET flushes of heifers contract-mated to young bulls selected on Genomic EBV

- Need for progeny-testing will decrease
Resulting shorter generation intervals can increase genetic gains

US Holsteins - Milk production

- Phenotypic base = 11,638 kg

Breeding value (kg)

Holstein birth year
Greater Reliabilities for Functional Traits can reverse declines in health, fertility

US Holsteins – Daughter Pregnancy Rate

Breeding value (%)

Phenotypic base = 21.53%

Holstein birth year
Costs to identify superior bulls could decline 5 yrs & $$$ later

Superior progeny-tested bull

Which is best??

Superior genome-tested young bull

DNA samples

< 6 mo & $$ later

Embryo Transfer

Illumina Bovine 50k Beadchip

Semen samples

5 yrs & $$$$$$$ later
Genomic Selection reduces the need to obtain phenotypes on selection candidates or their close relatives to obtain accurate EBV.

Instead, marker effects can be estimated on the target individuals in the target environment.

Current Pyramid Selection Programs  
(e.g. poultry and pigs)

Selection is primarily on phenotypes obtained on pure lines in high-biosecurity environments. But target is improvement of crossbreds in field environment.
Selection for Performance in Field

‘Traditional’ Breeding Solution:

Collect phenotypes on relatives in field

→ Combined Crossbred-Purebred Selection

Requirements / limitations:
- Costly logistics – Pedigree-based phenotyping in field
- Longer generation intervals
- Higher rates of inbreeding
  - use of family vs. own phenotype
Selection for Performance in Field
Molecular Breeding Solution:

Genomic selection based on SNP effects estimated in field

Advantages:
- Does not require pedigree-based phenotypes
- Opportunities to reduce generation intervals
- Opportunities to select for traits not observed in nucleus (disease, mortality)
Summary / Conclusions

Genomic Selection offers unique opportunities to enhance breeding programs by removing limitations on when, where, and on whom phenotypes are recorded.

- Not requiring phenotypes on individual/close relatives
- Selection on marker effects estimated on target individuals (crossbreds) and in target environment (field)

- With opportunities to
  - reduce generation intervals
  - reduce inbreeding
- Select for ‘difficult’ traits
- Address GxE
  - Genomic selection of pure-breds for field performance of crossbreds

Simulation results are very promising.
Initial empirical results are encouraging.
Acknowledgements

Rohan Fernando
Dorian Garrick
Iowa State University

Paul VanRadren and colleagues
USDA-AIPL
AB&G Summer Short Courses

Use of High-Density SNP Genotyping for Genetic Improvement of Livestock
June 1-10, 2009
Dekkers, Fernando, Garrick
Marker-assisted composite line development

QTL detection
Estimation of marker effects

\[ y = \text{marker genotype} + \text{BV} + e \]


10-20 cM marker distance sufficient
QTL detection, MA-Evaluation, and Selection

- Model:
  \[ y_{jk} = \text{generation}_k + \sum_{i=1}^{QTL} \beta_i \cdot MS_{ijk} + u_{jk} + e_{jk} \]

- Select on
  \[ \sum_{i=1}^{QTL} \hat{\beta}_i \cdot MS_{ijk} + \hat{u}_{jk} \]
Extra response (\%) over BLUP

Detected QTL explain 18\% of gen.var.
Marker-Assisted Evaluation and Selection

- Model:

\[ y_{jk} = \text{generation}_k + \sum_{i=1}^{90} \beta_i \ast MS_{ijk} + e_{jk} \]

- Select on:

\[ \sum_{i=1}^{90} \hat{\beta}_i \ast MS_{ijk} \]

# marker alleles from line 1 in interval \(i\)

\(\beta_i\) fixed or random

Genomic selection (Meuwissen et al. ‘01)
50% -ve QTL Effects, 20 cM

Cumulative Response (% over BLUP)

Generation

F3 F4 F5 F6 F7 F8 F9 F10 F11

Known QTL

Sign.QTL

R11

F11

BLUP

0

-50