Practical integration of genomic selection in dairy cattle breeding schemes

A. BOUQUET & J. JUGA
Introduction

Genomic selection: a revolution for animal breeders
- Big data: information at 1000s of markers for 1000s of animals
- Bigger and bigger data:
  - Huge increase in the number of genotyped animals
  - Denser SNP chips or even sequence data
- Change in the breeding program structures:
  - Reduction in generation intervals
  - More frequent (on-demand) genomic evaluations

In this context, how to make effective selection decisions?
In practice, which strategies were implemented?
Outline

1. The novel idea underlying genomic selection
2. Integrating big genomic data in genetic evaluations
3. Impact on the structure of breeding schemes and expected benefits
4. Practical breeding strategies for implementing genomic selection across countries
Related animals resemble more than unrelated ones because they share common genetic material (genes, QTL)

$$\text{Var}(G) = \text{Additive relationship matrix} \times \text{genetic variance}$$

Traditionally, additive relationships estimated from pedigrees

⇒ Full sibs without performance have the same estimated breeding values (EBV)

⇒ Progeny-testing (PT) of bulls necessary before selection to:
   ⇒ Get high EBV accuracy
   ⇒ Be able to differentiate between sibs
   ⇒ But long (5 yrs) and expensive (40 000€ / bull)
Genetic markers (SNPs) used as proxies to detect the QTL mix inherited by an animal from its parents.

SNP technology: profiling many animals for thousands of markers in a single analysis.

Dense marker maps covering the whole genome:
- Follow the transmission of chromosomal regions across generations.
- Strong linkage disequilibrium (LD) between SNPs and QTL.
  \[ \Rightarrow \text{within breed, 1 marker allele strongly associated with 1 QTL allele} \]

Genomic EBV = $£$ (effects estimated at marker loci).
Challenges to be faced

- To handle & manipulate the data (storage, quality checks)
- To compute GEBVs (construction & inversion of genomic matrices, genotype imputations...)
- To integrate this flow of information in breeding schemes
Reference Population with genotypes AND phenotypes used to build prediction equations of genetic merit

⇒ Usually bulls with daughters

Genomic equation applied to young candidates having genotypes but no phenotype to estimate their GEBV

Different methods to estimate GEBV, mainly:

- Genomic-BLUP approach
- Bayesian models

>>> In practice, increase in computing time compared to conventional BLUP
2-step approach (Van Raden, 2008):
⇒ Construct the additive relationship matrix based on genotypes - the so-called G-matrix (Van Raden, 2008; Forni et al., 2011)
⇒ Estimate a genomic value at markers for genotyped animals (DGV)
⇒ Combine it with pedigree-based EBV to get GEBV

1-step approach (Misztal et al., 2009; Christensen & Lund, 2010):
⇒ directly include genotyped and ungenotyped individual in a single relationship matrix H

\[ H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{\text{genotyped}}^{-1} \end{bmatrix} \]
Genomic relationship matrix = dense matrix without known structure

⇒ Multi-trait analysis of type traits in the USA, time needed for computing GEBV was doubled when including ~7000 genotyped animals (Tsuruta et al., 2011)

New strategies to reduce computations:
⇒ Directly build an approximate inverse of the G-matrix (Faux et al., 2012)
⇒ Strategies to directly update the inverse of the G matrix with only new genotypes (Meyer et al., 2013)
Other efficient approaches were developed to boost computations based on:

- Dimensionality reduction through principal component analysis (Solberg et al., 2009; Macciotta et al., 2009; Long et al., 2011)

- SNP selection (Legarra et al., 2011; Croiseau et al., 2011; Colombani et al., 2011)

Whatever the genomic evaluation approach, high interest in developing high-throughput computing (Wu et al., 2011, 2012; Cole et al., 2012)

- Parallel programming
- Use new hardware: graphical processing units (GPU)
Benefits of genomic selection

Increase in EBV reliability for young candidates without performance (CD ~ 0.25 ’ 0.65 for milk production traits in Holsteins)

⇒ GBLUP and Bayesian approaches render similar GEBV reliabilities for real data

Possibility to differentiate young candidates from both sexes on GEBV:
• without own performance or offspring
• both within- and between-families
Strategies to use genomic information

Genomic pre-selection (PS) scheme

- Male calves (0y)
  - Selection for PT-step on GEBV (1y)
  - Birth of progeny sample (2y)
  - Calving of daughters (4y)
  - Selection of AI bulls on (G)EBV (5y)

Genomic juvenile scheme (TURBO)

- Male calves (0y)
  - Selection of AI bulls on GEBV (1y)
  - Birth of progeny sample (2y)
  - Calving of daughters (4y)
  - Lactation records
  - Selection of AI bulls (5y)

Genotyping at birth
Impact of implementing genomic selection in dairy cattle

Compared to the PT scheme, selection on a trait with $h^2=0.3$

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Generation interval</th>
<th>Annual genetic gain</th>
<th>Inbreeding rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS</td>
<td>~constant</td>
<td>+9 - 16%</td>
<td>Reduced</td>
</tr>
<tr>
<td>TURBO</td>
<td>halved</td>
<td>+29 - 100%</td>
<td>Constant to reduced</td>
</tr>
</tbody>
</table>

>> Advantage of genomic selection schemes larger for low heritability traits

(* Review from Buch et al., 2012; Colleau et al., 2009; Lillehammer et al., 2011; Pryce et al., 2010)
Increase in $G$ on the sire-to-offspring genetic pathway:

- reduction in the generation interval
- although reduction in selection accuracy of AI sires

Increase in $G$ on the dam-to-offspring pathway:

- Reduction in the generation interval
- Higher selection accuracy, especially for low heritability trait
- Less bias in GEBV than EBV: lesser impact of preferential treatments

Large gains in $G$ when genotyping - even small proportion of - females (Sørensen and Sørensen, 2009)

- Strategies needed to optimally allocate genotypes to males / females
Increasing the reference population size

The bigger the reference population, the higher the accuracy of GEBV

⇒ Creation of across-country consortia to exchange genotypes
  ▪ 3 main for Holsteins: Eurogenomics (Lund et al., 2010), Consortium USA-Canada-UK-Italy, Consortium Australia-Ireland-NZ (LIC)
  ▪ 1 worldwide consortium for the Brown Swiss (Santus, EAAP 2013)

⇒ Integration of cows in reference populations
  ⇒ Simulation of the Irish Holstein population (McHugh et al., 2011):

<table>
<thead>
<tr>
<th></th>
<th>TURBO</th>
<th>TURBO</th>
<th>TURBO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref. Population size</td>
<td>+500 B / yr</td>
<td>+ 500 B / yr</td>
<td>+ 500 B / yr</td>
</tr>
<tr>
<td></td>
<td>+ 500 @ / yr</td>
<td>+3 500 @ / yr</td>
<td></td>
</tr>
<tr>
<td>Annual • G</td>
<td>0.34 Ŕ (-)</td>
<td>0.37 Ŕ (+9%)</td>
<td>0.49 Ŕ (+44%)</td>
</tr>
<tr>
<td>Annual • F</td>
<td>+0.7%</td>
<td>+0.4%</td>
<td>+0.4%</td>
</tr>
</tbody>
</table>
Optimal design of the reference population when:

- Low relationship among reference sires
- High relationship between reference population and candidates (Pszczola et al., 2012)

Constitution of multi-breed reference populations:

- No increase in reliability with 54K if distant breeds and low heritability (Karoui et al., 2012)
- Would require denser maps to get strong LD between SNP and QTL across breeds (Schrooten et al., EAAP 2013)
Managing the risk of TURBO schemes

At the herd level, use teams of bulls to mitigate the risk of using bulls with poor merit (Schefers and Weigel, 2012)

To curb annual inbreeding rates

⇒ Enlarge the number of young genomic bulls and ensure balanced use of them

⇒ Avoid further use of genomic bulls once they have milking daughters, except if strict rules for managing genetic diversity (Colleau et al., 2009)

Promising results when integrating genomic data in methods for preserving genetic diversity

⇒ Selection based on optimal contributions (Sonesson et al., 2010, 2012)

⇒ Mate allocation with constraints on progeny inbreeding (Pryce et al., 2012)
Current-state implementation of genomic selection: Holstein breed

<table>
<thead>
<tr>
<th></th>
<th>USA</th>
<th>UK</th>
<th>FRA</th>
<th>GER</th>
<th>NLD</th>
<th>AUS</th>
<th>IRL</th>
<th>NZL (LIC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref. Pop. (# of bulls)</td>
<td>~23,000</td>
<td>~23,000</td>
<td>~26,000</td>
<td>~26,000</td>
<td>~26,000</td>
<td>~4,800</td>
<td>~4,800</td>
<td>~4,800</td>
</tr>
<tr>
<td>Including cows in Ref. Pop</td>
<td>Yes, 53,396 @</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes, 8716 @</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**>> Large increase in reference population sizes since Apr. 2011 (Pryce & Daetwyler, 2012)**

a Calus et al., 2008
### Current-state implementation of genomic selection: Holstein breed

<table>
<thead>
<tr>
<th></th>
<th>USA</th>
<th>UK</th>
<th>FRA</th>
<th>GER</th>
<th>NLD</th>
<th>AUS</th>
<th>IRL</th>
<th>NZL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nb of genotyped calves / yr</strong></td>
<td>25 000B 135 000@</td>
<td>2500</td>
<td>50 000</td>
<td>12 000 B</td>
<td>2500 B</td>
<td>580 B</td>
<td>2000 @</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Young sires selected /yr</strong></td>
<td>1 000</td>
<td>NA</td>
<td>~ 400</td>
<td>~ 500</td>
<td>20</td>
<td>NA</td>
<td>40</td>
<td>10-15</td>
</tr>
<tr>
<td><strong>Organized PT of bulls</strong></td>
<td>Yes</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>Yes, 120</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Market-share of genomic bulls</strong></td>
<td>48%</td>
<td>NA</td>
<td>57%</td>
<td>55%</td>
<td>35%</td>
<td>NA</td>
<td>54%</td>
<td>30%</td>
</tr>
</tbody>
</table>

Compared to a 2011 survey (Pryce & Daetwyler, 2012):

- Large increase in the number of genotyped calved in all countries
- Ceasing of organized PT in several countries
- Increase in the market-share of genomic bulls
Breeding schemes are being changed to integrate genomic selection:

- **Key drivers**: Improvement of the technical & economical efficiency
- **New opportunities**:
  - improve genetic gain on new traits / traits difficult to select
  - better control of inbreeding rates

Increasing amounts of genotypic information collected through time

Need to adapt on several aspects:

- Computing infrastructure (data storage, hardware architecture)
- Computing strategies & programming to gain in efficiency
Thank you!