Description of the French genomic Marker Assisted by Selection program

P. Croiseau, C. Hoze, S. Fritz, F. Guillaume, C. Colombani, A. Legarra, A. Baur, C. Robert-Granié, D. Boichard and V. Ducrocq
Use of the MA-BLUP in France

- A strong background with MA-BLUP/QTL-BLUP for national genomic evaluations
- This strategy was tested in Holstein and provided better results than others genomic selection approaches

<table>
<thead>
<tr>
<th>Correlation between DYD_{obs} and GEBV</th>
<th>Milk</th>
<th>Protein</th>
<th>Fat</th>
<th>Protein %</th>
<th>Fat %</th>
<th>Conception rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>pedigree-based BLUP</td>
<td>0.38</td>
<td>0.44</td>
<td>0.40</td>
<td>0.47</td>
<td>0.44</td>
<td>0.29</td>
</tr>
<tr>
<td>GBLUP</td>
<td>0.56</td>
<td>0.55</td>
<td>0.59</td>
<td>0.73</td>
<td>0.72</td>
<td>0.35</td>
</tr>
<tr>
<td>PLS</td>
<td>0.53</td>
<td>0.55</td>
<td>0.58</td>
<td>0.71</td>
<td>0.70</td>
<td>0.33</td>
</tr>
<tr>
<td>Elastic-Net</td>
<td>0.57</td>
<td>0.57</td>
<td>0.63</td>
<td>0.75</td>
<td>0.80</td>
<td>0.34</td>
</tr>
<tr>
<td>French BLUP-QTL</td>
<td>0.60</td>
<td>0.57</td>
<td>0.7</td>
<td>0.73</td>
<td>0.81</td>
<td>0.39</td>
</tr>
</tbody>
</table>

*WCGALP, Liepzig 2010
QTL-BLUP
(Fernando and Grossman, 1989)

\[ y_i = \sum_{j=1}^{QTL} (h_{ij}^{sire} + h_{ij}^{dam}) + u_i + e_i \]

- \( Y_i \rightarrow \) performances for individual i
- \( h_{ij} \rightarrow \) gametic effect from sire and dam for the QTL j
- \( u_i \rightarrow \) polygenic effect for individual i
- \( e_i \rightarrow \) residual effect for individual i
Particularities of the QTL-BLUP

• Haplotypes are taken into account
  ➢ LD between QTL and markers is increased using haplotypes instead of SNP
  ➢ Since the causal mutations are probably rarely genotyped, the use of haplotypes should improve prediction equations

• Since the same list of QTL is used for successive national genomic evaluations
  ➢ GEBV stability over time

• Since a restricted number of QTL in included in the model, such an evaluation is:
  ➢ less computationally demanding
  ➢ adapted to higher density chips once QTL are chosen
Which list of haplotypes for the MA-BLUP?

1. LRT peaks from a QTL detection (LDLA, Meuwissen and Goddard)

2 criteria to define a LRT peak:

- Have the highest LRT value in a window of SNP (1 or 2 cM)
- Have a LRT value higher than a threshold (3 or 5)

(Croiseau et al., Leibzig 2010)
Which list of haplotypes for the QTL-BLUP?

1. LRT peaks from a QTL detection (LDLA, Meuwissen and Goddard)

2. Use of a genomic selection approach based on a variable selection method
   - Elastic-Net (EN); sparse Partial Least Squares (sPLS)
     - EN is a linear combination of Ridge Regression and LASSO
     - sPLS* is the variable selection version of the PLS (citation)
   - From the set of SNP obtained using EN or sPLS, the SNP which are in the same cM were grouped in haplotypes

(* Colombani et al., WCGALP 2011)
• This strategy worked well in Holstein
• In this study we wanted to test it in a smaller reference population (Montbéliarde)
  ▪ 1392 animals in Montbéliarde breed
    ➢ Training : 1170 individuals
    ➢ Validation: 222 individuals
  ▪ 5 traits
    ➢ Rear udder width
    ➢ height at sacrum
    ➢ Somatic Cell Counts
    ➢ milk yield
    ➢ protein yield
  ▪ Performances were DYD (Daughter Yield Deviation)
Analysis

- Weighted correlation between observed DYD and DGV are calculated
  - Weight = Equivalent Daughter Contribution (EDC)
- QTL-BLUP were performed using QTL lists defined by
  - LDLA
  - Elastic-Net
  - sPLS
- Comparison with pedigree-based BLUP, GBLUP, EN and sPLS
- For LDLA, results for the best definition of LRT peaks are shown
- For EN and sPLS, results of the best combination of parameters are shown
### Correlation between $DYD_{obs}$ and DGV

<table>
<thead>
<tr>
<th>Trait</th>
<th>pedigree-based BLUP</th>
<th>GBLUP</th>
<th>EN</th>
<th>sPLS</th>
<th>QTL-BLUP</th>
<th>LDLA</th>
<th>EN</th>
<th>sPLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>somatic cell count</td>
<td>0.50</td>
<td>0.59</td>
<td>0.57</td>
<td>0.47</td>
<td>0.57</td>
<td>0.55</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>rear udder width</td>
<td>0.39</td>
<td>0.55</td>
<td>0.54</td>
<td>0.48</td>
<td>0.50</td>
<td>0.50</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>milk</td>
<td>0.27</td>
<td>0.41</td>
<td>0.44</td>
<td>0.41</td>
<td>0.45</td>
<td>0.46</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>protein yield</td>
<td>0.28</td>
<td>0.43</td>
<td>0.47</td>
<td>0.42</td>
<td>0.47</td>
<td>0.49</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>height at sacrum</td>
<td>0.41</td>
<td>0.54</td>
<td>0.54</td>
<td>0.49</td>
<td>0.53</td>
<td>0.55</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>mean over the 5 traits</td>
<td>0.37</td>
<td>0.50</td>
<td>0.51</td>
<td>0.45</td>
<td>0.50</td>
<td>0.51</td>
<td>0.44</td>
<td></td>
</tr>
</tbody>
</table>
French QTL-BLUP

• In the QTL-BLUP based on LDLA, QTL variances were estimated.
• In the QTL-BLUP based on EN list, all the QTL have the same variance.
  ✓ Total QTL variance was 60% in all cases (40% for the polygenic component).
  ✓ Proportion selected to improve the slope of regression.
• In the French QTL-BLUP.
  ✓ QTL come from the combine LDLA + EN list.
## Correlation between $DYD_{obs}$ and GEBV

<table>
<thead>
<tr>
<th>Trait</th>
<th>pedigree-based BLUP</th>
<th>GBLUP</th>
<th>EN</th>
<th>sPLS</th>
<th>BLUP-QTL LDLA</th>
<th>BLUP-QTL EN</th>
<th>BLUP-QTL sPLS</th>
<th>French QTL-BLUP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somatic cell count</td>
<td>0.50</td>
<td><strong>0.59</strong></td>
<td>0.57</td>
<td>0.47</td>
<td>0.57</td>
<td>0.55</td>
<td>0.49</td>
<td>0.56</td>
</tr>
<tr>
<td>Rear udder width</td>
<td>0.39</td>
<td><strong>0.55</strong></td>
<td>0.54</td>
<td>0.48</td>
<td>0.50</td>
<td>0.50</td>
<td>0.46</td>
<td>0.53</td>
</tr>
<tr>
<td>Milk</td>
<td>0.27</td>
<td>0.41</td>
<td>0.44</td>
<td>0.41</td>
<td>0.45</td>
<td><strong>0.46</strong></td>
<td>0.43</td>
<td>0.42</td>
</tr>
<tr>
<td>Protein yield</td>
<td>0.28</td>
<td>0.43</td>
<td>0.47</td>
<td>0.42</td>
<td>0.47</td>
<td><strong>0.49</strong></td>
<td>0.39</td>
<td>0.45</td>
</tr>
<tr>
<td>Height at sacrum</td>
<td>0.41</td>
<td>0.54</td>
<td>0.54</td>
<td>0.49</td>
<td>0.53</td>
<td><strong>0.55</strong></td>
<td>0.45</td>
<td>0.53</td>
</tr>
<tr>
<td>Mean over the 5 traits</td>
<td><strong>0.37</strong></td>
<td><strong>0.50</strong></td>
<td><strong>0.51</strong></td>
<td><strong>0.45</strong></td>
<td><strong>0.50</strong></td>
<td><strong>0.51</strong></td>
<td>0.44</td>
<td>0.50</td>
</tr>
</tbody>
</table>
## Number of SNP/QTL used in the prediction equation

<table>
<thead>
<tr>
<th>Pedigree-based BLUP</th>
<th>GBLUP</th>
<th>EN SNP number</th>
<th>BLUP-QTL LDLA</th>
<th>BLUP-QTL EN</th>
<th>French BLUP-QTL SNP</th>
<th>French BLUP-QTL QTL</th>
</tr>
</thead>
<tbody>
<tr>
<td>somatic cell count</td>
<td>-</td>
<td>38490</td>
<td>13687</td>
<td>2350 470 940 312</td>
<td>1568 392</td>
<td></td>
</tr>
<tr>
<td>rear udder width</td>
<td>-</td>
<td>38490</td>
<td>19957</td>
<td>2750 550 652 400</td>
<td>1920 480</td>
<td></td>
</tr>
<tr>
<td>milk</td>
<td>-</td>
<td>38490</td>
<td>25713</td>
<td>3025 605 1007 628</td>
<td>2832 708</td>
<td></td>
</tr>
<tr>
<td>protein yield</td>
<td>-</td>
<td>38490</td>
<td>14879</td>
<td>3145 629 1211 479</td>
<td>2236 559</td>
<td></td>
</tr>
<tr>
<td>heigth at sacrum</td>
<td>-</td>
<td>38490</td>
<td>22703</td>
<td>2065 413 784 344</td>
<td>1696 424</td>
<td></td>
</tr>
<tr>
<td>mean over the 5 traits</td>
<td>-</td>
<td>38490</td>
<td>19388</td>
<td>2667 533 919 433</td>
<td>2050 582</td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

- Optimal correlations were obtained using
  - EN
  - QTL-BLUP with a list of QTL defined using EN

- French QTL-BLUP brings more benefits for Holstein than for Montbéliarde
  - Maybe due to the reference population size which is smaller
    - haplotype effects estimation is more difficult for rare variants
    - We need to work on haplotype clustering to avoid this problem
  - Results should be improved with HD chip

- French QTL-BLUP maintained the stability of GEBV
  - When the same list of QTL is used over successive genomic evaluations
LABOGENA for genotyping

ANR and APISGEN for funding AMASGEN program