Interest and limits of the bovine 50k chip to study *Bos taurus x Bos indicus* crossbred animals in an Indian bull stud

Dennis Cruz Hidalgo,
M. Swaminathan, P. Deshpande,
M. Boussaha, R. Saintilan, D. Laloe,
Vincent Ducrocq*

BAIF Development Research Foundation, Central Research Station Uruli Kanchan, Maharashtra, India

Bovine Genetics and Genomics, GABI, INRA Jouy-en-Josas
BAIF: Bharatiya Agro Industries Foundation
Largest Indian NGO in Agriculture, founded in 1967
... committed to sustainable development in rural India

Dr Manibhai Desai (1920-1993)
BAIF: Bharatiya Agro Industries Foundation
Largest Indian NGO in Agriculture, founded in 1967
... committed to sustainable development in rural India

Dr Manibhai Desai (1920-1993)

http://www.baif.org.in
BAIF main activity: Livestock Development

Genetic improvement of local cattle through AI in tiny farms (< 2 cows)
BAIF main activity: Livestock Development

Genetic improvement of local cattle through AI in tiny farms (< 2 cows)
BAIF main activity: Livestock Development

Genetic improvement of local cattle through AI in tiny farms (< 2 cows)

⇒ 5.3 million families spread over 90,000 villages in 12 states
Semen production

• In the late 70’s, a **nucleus herd** of ~200 pure Holstein and Jersey heifers from Denmark and Canada + some imported semen

⇒ a **bull stud** producing now > **8 millions doses a year** ...
Semen production

• In the late 70’s, a nucleus herd of ~200 pure Holstein and Jersey heifers from Denmark and Canada + some imported semen
→ a bull stud producing now > 8 millions doses a year ...
= pure « exotic » (HF/J) and crossbred bulls
In the late 70’s, a **nucleus herd** of ~200 pure Holstein and Jersey heifers from Denmark and Canada + some imported semen ➔ a **bull stud** producing now > 8 millions doses a year ... = pure « exotic » (HF/J) and crossbred bulls but also pure **indigenous** (Bos Indicus)
Semen production

• In the late 70’s, a nucleus herd of ~200 pure Holstein and Jersey heifers from Denmark and Canada + some imported semen ➔ a bull stud producing now > 8 millions doses a year ...

= pure « exotic » (HF/J) and crossbred bulls

but also pure indigenous (Bos Indicus) as well as buffaloes
Joint BAIF-INRA genomic project

• 288 animals genotyped on Illumina Bovine SNP50 Beadchip®

<table>
<thead>
<tr>
<th>Breed</th>
<th>Bulls</th>
<th>Cows</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indigenous</td>
<td>58</td>
<td>(8 breeds, mainly Sahiwal and Gir)</td>
</tr>
<tr>
<td>Holstein</td>
<td>59</td>
<td>22</td>
</tr>
<tr>
<td>Holstein crossbreds</td>
<td>78</td>
<td>7</td>
</tr>
<tr>
<td>Jersey</td>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>Jersey crossbreds</td>
<td>28</td>
<td>(8 J 75%, 16 J 50%, 4 J 62.5%)</td>
</tr>
</tbody>
</table>

• Question: **how to use this genomic information to help the management of the bull stud?**
Principal Component Analysis (PCA)

1 animal = 1 point = (up to) 43,167 SNP

ID PCA axes 1 and 2

15.3%

6.4%
Principal Component Analysis (PCA)

1 animal = 1 point = (up to) 43,167 SNP

- Jersey and Jersey crosses
- Holstein and Holstein crosses
- All indigenous breeds

6.4% 15.3%
PCA for Indigenous breeds

ID PCA axes 1 and 2

- Khillar
- Amritmahal
- Haryana
- Tharparkar
- Gir
- Sahiwal

5.3%
4.2%
PCA for Indigenous breeds

Draft breeds

Dairy Breeds
(higher heterogeneity in Sahiwal)

ID PCA axes 1 and 2

khilar
Amritmahal
Haryana
Tharparkar
Gir
Sahiwal

5.3%
4.2%
3D PCA for exotic purebreds

Holstein

Jersey

12.9%

4.3%

3.2%
3D PCA for exotic purebreds

Much larger heterogeneity among Holstein animals!
Cluster analysis within pure Holsteins
Cluster analysis within pure Holsteins

Clusters:
sons of ...

1. « recent » CD HF
2. « old » Danish
3. « old » Canadian
4. « recent » HF
Why?

- **Genomic evaluation** on ~40 traits on the *French scale*
  (obviously not necessarily an adequate scale, given the huge Gx E)
Why?

- **Genomic evaluation** on ~40 traits on the **French scale**
  (obviously not necessarily an adequate scale, given the huge Gx E)

![Milk yield chart](chart.png)
Why?

• Genomic evaluation on ~40 traits on the French scale (obviously not necessarily an adequate scale, given the huge GxE)

![Milk yield and Fat % graphs](image-url)
Why?

• Genomic evaluation on ~40 traits on the French scale (obviously not necessarily an adequate scale, given the huge Gx E)

► Clearly reveals huge differences in genetic merit, consistent with the history of semen procurement and use at BAIFF
Breed composition: Admixture

Jersey crosses
Breed composition: Admixture

Jersey crosses

Holstein crosses
Breed composition: Admixture

Jersey crosses

Holstein crosses
## Breed composition and breed assignation

- **Purebreds**: **Holstein vs Jersey vs Gir vs Sahiwal**: 100%
- **Jersey crosses**: % Jersey: good  
  % Sahiwal/Gir: not always good

<table>
<thead>
<tr>
<th>N°</th>
<th>From pedigree information</th>
<th>Jersey</th>
<th>Sahiwal</th>
<th>Gir</th>
</tr>
</thead>
<tbody>
<tr>
<td>170</td>
<td>Jersey 50% Sahiwal 50%</td>
<td>0.51</td>
<td>0.49</td>
<td>-</td>
</tr>
<tr>
<td>203</td>
<td>Jersey 62.5% Gir 37.5%</td>
<td>0.69</td>
<td>-</td>
<td>0.31</td>
</tr>
<tr>
<td>155</td>
<td>Jersey 75% Sahiwal 25%</td>
<td>0.76</td>
<td>0.24</td>
<td>-</td>
</tr>
<tr>
<td>152</td>
<td>Jersey 75% Sahiwal 12.5% Gir 12.5%</td>
<td>0.77</td>
<td>0.22</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Breed composition and breed assignation

- Purebreds: **Holstein vs Jersey vs Gir vs Sahiwal**: 100%
- Jersey crosses: % Jersey: good  % Sahiwal/Gir: not always good

<table>
<thead>
<tr>
<th>N°</th>
<th>From pedigree information</th>
<th>Jersey</th>
<th>Sahiwal</th>
<th>Gir</th>
</tr>
</thead>
<tbody>
<tr>
<td>170</td>
<td>Jersey 50% Sahiwal 50%</td>
<td>0.51</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>203</td>
<td>Jersey 62.5% Gir 37.5%</td>
<td>0.69</td>
<td>-</td>
<td>0.31</td>
</tr>
<tr>
<td>155</td>
<td>Jersey 75% Sahiwal 25%</td>
<td>0.76</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>152</td>
<td>Jersey 75% Sahiwal 12.5% Gir 12.5%</td>
<td>0.77</td>
<td>0.22</td>
<td>0.01</td>
</tr>
</tbody>
</table>

- Holstein crosses: % **Holstein**: often overestimated  % Sahiwal/Gir: poor

<table>
<thead>
<tr>
<th>N°</th>
<th>From pedigree information</th>
<th>Holstein</th>
<th>Sahiwal</th>
<th>Gir</th>
</tr>
</thead>
<tbody>
<tr>
<td>106</td>
<td>Holstein 50% Sahiwal 50%</td>
<td>0.52</td>
<td>0.10</td>
<td>0.38</td>
</tr>
<tr>
<td>108</td>
<td>Holstein 50% Gir 50%</td>
<td>0.70</td>
<td>-</td>
<td>0.30</td>
</tr>
<tr>
<td>92</td>
<td>Holstein 75% Gir 25%</td>
<td>0.80</td>
<td>-</td>
<td>0.20</td>
</tr>
<tr>
<td>63</td>
<td>Holstein 75% Sahiwal 3.1% Gir 21.9%</td>
<td>0.80</td>
<td>0.16</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Why ?: the 50K Chip

- Only 40 animals of two *bos indicus* breeds used to construct and validate the chip ...

Source:
BovineSNP50 ® BeadChip  Illumina Datasheet
Why ?: the 50K Chip

- Much less polymorphic
  (only ~50% SNP are polymorphic in Gir, in the BAIF project:
  \[>18000\] SNP are completely fixed)

### Table 3: BovineSNP50 BeadChip Content Validation

<table>
<thead>
<tr>
<th>Breed</th>
<th>Samples</th>
<th>Polymorphic Loci*</th>
<th>Mean MAF</th>
<th>Median MAF**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angus</td>
<td>24</td>
<td>41,073</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>Beefmaster</td>
<td>23</td>
<td>43,114</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>Bos indicus Gir</td>
<td>21</td>
<td>23,587</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>Bos indicus Nelore</td>
<td>19</td>
<td>25,492</td>
<td>0.11</td>
<td>0.03</td>
</tr>
<tr>
<td>Brahman</td>
<td>22</td>
<td>29,444</td>
<td>0.13</td>
<td>0.07</td>
</tr>
<tr>
<td>Brown Swiss</td>
<td>21</td>
<td>33,971</td>
<td>0.10</td>
<td>0.17</td>
</tr>
<tr>
<td>Cherkala</td>
<td>19</td>
<td>43,723</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Cuernsey</td>
<td>21</td>
<td>33,743</td>
<td>0.19</td>
<td>0.17</td>
</tr>
<tr>
<td>Hereford</td>
<td>24</td>
<td>42,132</td>
<td>0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>Holstein</td>
<td>49</td>
<td>42,049</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>Jersey</td>
<td>23</td>
<td>33,346</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>Limousin</td>
<td>39</td>
<td>42,517</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>N'Dama</td>
<td>23</td>
<td>28,369</td>
<td>0.14</td>
<td>0.07</td>
</tr>
<tr>
<td>Norwegian Red</td>
<td>17</td>
<td>42,055</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Piedmontese</td>
<td>21</td>
<td>41,912</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Red Angus</td>
<td>10</td>
<td>42,388</td>
<td>0.21</td>
<td>0.20</td>
</tr>
<tr>
<td>Romagnola</td>
<td>21</td>
<td>38,524</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>Santa Gertrudis</td>
<td>21</td>
<td>41,783</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Sheko</td>
<td>15</td>
<td>35,084</td>
<td>0.17</td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td>434</td>
<td>47,168</td>
<td>0.24</td>
<td>0.25</td>
</tr>
</tbody>
</table>

* MAF > 0.05
** Across all 54,099 loci
Why ?: the 50K Chip

- **Much less polymorphic**
  (only ~50% SNP are polymorphic in Gir, in the BAIF project: >18000 SNP are completely fixed)
- **Much less informative**
  (50% of SNP have a MAF<=0.02 in Gir)
Why ?: the 50K Chip

• Much less polymorphic (only ~50% SNP are polymorphic in Gir, in the BAIF project: >18000 SNP are completely fixed)
• Much less informative (50% of SNP have a MAF<=0.02 in Gir)

⇒ An Indian (BAIF + NDDB) initiative:
• create a new 50k chip adapted to Indian indigenous breeds and crossbreds: the « IndusChip » (on going project)
Use of genomic tools in India is promising

- Bring original and useful information at bull stud level
- Can partly compensate the absence of large scale pedigree (and performance) recording
- Can contribute to a better characterization and understanding of the current genetic diversity between and within breeds
- Can improve the way crossbreeding is done
- Can be a first step for genomic selection of better bulls in crossbred and even indigenous populations
Interest and limits of the bovine 50k chip to study *Bos taurus x Bos indicus* crossbred animals in an Indian bull stud

Dennis Cruz Hidalgo, M. Swaminathan, P. Deshpande, M. Boussaha, R. Saintilan, D. Laloe, Vincent Ducrocq*
Why? Quality control at SNP level:

- Gen Train (cluster reliability) > 0.7
- Call Freq > 0.97
- SNP from chromosome X, Y or unknown
- SNP monomorphic or with very rare alleles \( \Rightarrow 4749 \) SNP removed
  \( \Rightarrow \) overall \( \textbf{43,167} \) SNP used ... but ...

For indigenous breeds only: \( \textbf{11,957} \) monomorphic SNP
For Sahiwal + Gir only: \( \textbf{18,191} \) monomorphic SNP

(For Holstein + Jersey only: \( \textbf{6,134} \) monomorphic SNP)

- Call Rate > 0.95 \( \Rightarrow 7 \) animals removed
Why?: SNP “contributions” to axes 1 and 2

• Much more information to distinguish between exotic and indigenous breeds that between indigenous breeds
PCA for exotic purebreds

Heterogeneity within Holstein (much larger!)

Jersey vs Holstein

12.9%

4.3%