Identifying family clusters within the US Holstein population to manage genetic diversity

Yvette Steyn, T. Lawlor, Y. Masuda, S. Tsuruta, D.A.L. Lourenco, I. Misztal
AI reach

Round Oak Rag Apple  Elevation  Pawnee Farm Arlinda  Chief
Introduction

Objectives

- Identify potential family clusters in the US Holstein population and their key founders

- Evaluate the expected inbreeding that would result from within- or across-cluster mating
Data

- Sires of animals born after 1985
- Genotyped sires with most progeny
- 1,145 Sires
- Nr progeny: 312 to 49,146
- 7 Countries
Relationships

\[ G = \frac{ZZ'}{2 \sum p_i (1 - p_i)} \]

\( G \): Genomic relationship matrix

\( Z \): SNP marker matrix (0, 1, 2) minus 2 times the allele frequency

\( p_i \): allele frequency of SNP i

VanRaden et al. 2008
Principal Component Analysis

- Eigenvalue decomposition
- $1^{st}$ PC:
  - Captures most variance
- $2^{nd}$ PC:
  - Uncorrelated to PC1
  - Captures $2^{nd}$ most variance
PC1 to PC3
K-means clustering

- Hard clustering
- Reduces within-cluster variation
- Increases between-cluster variation
- User choice of clusters
- Simple
- Fast
- Large data
5 Clusters
50 replicates
Clusters are reflecting different populations

Blackstar (2) contains almost all U.S. bulls.

Starbuck (4) is the most “international” group
The smallest group includes Valiant (Cluster 1), older U.S. genotyped bulls. Bulls in this group have a strong relationship to a single progenitor, **PAWNEE FARM ARLINDA CHIEF**.

The largest group is from Tradition (Cluster 5). Multiple U.S. sire families are represented. Tradition is the son of **ROUND OAK RAG APPLE ELEVATION**.

<table>
<thead>
<tr>
<th>Sire Family</th>
<th>Cluster</th>
<th>Progenitors</th>
<th>Key Sires</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster</td>
<td>Progenitors</td>
<td>Key Sires</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Chief</td>
<td>Mark and Valiant</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Ivanhoe &amp; Fond Matt</td>
<td>Chairman &amp; Blackstar</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Ivanhoe Star</td>
<td>Bell</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Canadian breeding</td>
<td>Starbuck, Outside, etc.</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>US breeding Multiple families</td>
<td>Tradition, Leadman, etc.</td>
<td></td>
</tr>
</tbody>
</table>
Cluster 1: Chief

1986
Cluster 2: Blackstar

1985 & 2001
Cluster 3: Bell

1998
Canadian

Cluster 4: Starbuck

2002
Oldest clusters

Chief (1)

Tradition (5)

1986

1995
## Traits means within cluster

### Table: Ranking and Traits

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Stature</th>
<th>FUA</th>
<th>RUH</th>
<th>RUW</th>
<th>FL</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>Starbuck</td>
<td>Starbuck</td>
<td>Starbuck</td>
<td>Starbuck</td>
<td>Starbuck</td>
</tr>
<tr>
<td>Second</td>
<td>Bell</td>
<td>Bell</td>
<td>Bell</td>
<td>Bell</td>
<td>Blackstar</td>
</tr>
<tr>
<td>Third</td>
<td>Blackstar</td>
<td>Blackstar</td>
<td>Blackstar</td>
<td>Blackstar</td>
<td>Bell</td>
</tr>
<tr>
<td>Fourth</td>
<td>Tradition</td>
<td>Tradition</td>
<td>Tradition</td>
<td>Tradition</td>
<td>Tradition</td>
</tr>
<tr>
<td>Fifth</td>
<td>Chief</td>
<td>Chief</td>
<td>Chief</td>
<td>Chief</td>
<td>Chief</td>
</tr>
</tbody>
</table>

### Chart: Indirect Prediction

The chart above illustrates the indirect prediction for various traits within different clusters, including Stature, FUA, RUH, RUW, and FL. The clusters are represented by different colors, with a legend indicating:
- **Chief**
- **Blackstar**
- **Bell**
- **Starbuck**
- **Tradition**

The ranking of characters based on their traits is as follows:

1. **Starbuck**
2. **Bell**
3. **Blackstar**
4. **Tradition**
5. **Chief**
Practical application

• Mating decisions
• Across-cluster mating
• Selection candidates
Data

• Sires of animals born after 2010 & at least 25 progeny
• Females measured after 2012 (only data to 2014)
• 20,099 selection candidates with genotypes
• Years: 1987 and up
• 14 countries
• 5 Clusters
Larger data: Selection candidates

Principal Component Plots - 3 Dimensions
Expected inbreeding when using Cluster 3 sires

C1 Dams

C2 Dams

C3 Dams

C4 Dams

C5 Dams
Expected inbreeding when using Cluster 4 sires

C1 Dams

C2 Dams

C3 Dams

C4 Dams

C5 Dams
Expected inbreeding: 8 Clusters

<table>
<thead>
<tr>
<th>Sires</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0.26</td>
<td>0.20</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.13</td>
</tr>
<tr>
<td>C2</td>
<td>0.20</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>0.18</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.13</td>
</tr>
<tr>
<td>C4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>0.12</td>
<td>0.12</td>
<td></td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td>0.13</td>
</tr>
<tr>
<td>C6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>0.13</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.22</td>
<td>0.13</td>
</tr>
<tr>
<td>C8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.13</td>
<td></td>
<td>0.12</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Expected inbreeding of all candidate sires and dams: 0.121
\[ \frac{\text{between cluster } SS}{\text{total } SS} \times 100 \]
Conclusion

• PCA show important bulls
• Corresponds to k-means clustering
• Can reduce expected inbreeding
• Does not take genetic merit into account
• Selection intensity?
• Genetic redundancy
Thank you

yvette.steyn@uga.edu