Closing the gap on genomic technologies applied to tropical beef cattle

L.R. Porto-Neto¹, S.S. McWilliam¹, M. Naval-Sanchez¹, S.S. Moore², A. Reverter¹, B.J. Hayes²

¹ Commonwealth Scientific and Industrial Research Organization (CSIRO), Agriculture and Food, St Lucia, QLD 4067, Australia
laercio.portoneto@csiro.au (corresponding author)
² Queensland Alliance for Agriculture and Food Innovation (QAAFI), Centre for Animal Science, The University of Queensland, St Lucia, QLD 4067, Australia

Summary

The adoption of modern genomic technologies has not been equally distributed across different animal production systems and targeted industries. Although the application of genomic selection has been shown to increase productivity in several animal species, its use in tropical beef cattle is still slow in comparison to dairy cattle. Of 2,703 sequenced animals in the latest run of the 1000 Bull Genome Project (run6), around 16% represented tropical breeds, demonstrating a growing move towards more diversity and collaborative efforts. Here it is described a genomic resource of about twenty-two thousand animals with phenotypes on several different traits and imputed genetic variants up to whole-genome sequenced level. It is presented the procedure used for the imputation along with an example of the utilization of this resource in a chromosome level association analysis.

Keywords: whole-genome sequencing, tropical cattle, imputation, GWAS

Introduction

The adoption of modern genomic technologies has not been equally distributed across different production systems, targeted industries or animal species. Although the application of genomic selection has been shown to improve productivity in several animal production schemes (Weller et al., 2017), its use in tropical beef cattle is still lagging behind other schemes such as dairy cattle. There are examples of tropical beef production being supported by the use of genomic technologies, but in general, these initiatives have been made by individual producers or companies or by a group of producers, still with little potential impact worldwide or even nationally.

The international 1000 Bull Genome Project (Daetwyler et al., 2014) focused initially on dairy cattle, but soon adopted other taurine breeds, including beef producing breed, and latter also indicine breeds. Of 2,703 sequenced animals in the latest run (run6), around 16% represented tropical breeds, demonstrating a growing move towards diversity and collaborative efforts. SNP genotyping companies are also observing the growing market and are investing in technologies that have potentially less ascertainment bias on SNP selection, and allow accurate imputation to favour tropical beef cattle.

The number of phenotyped and genotyped animals is a key factor for the success of a
genomic selection program. With this in mind, a population of around twenty-two thousand tropical cattle from different experiments were consolidated into a single dataset. These cattle were genotyped using several SNP arrays and imputed up to whole-genome sequence. Additional to the genotypes, these cattle were phenotyped for many production-related traits. Here the procedure used for the imputation is presented along with an example of the utilization of this resource in a chromosome level association analysis.

**Material and methods**

### SNP genotyped and whole-genome sequenced cattle

Cattle genotypes from several experiments were combined to generate this large dataset (Table 1). The main source of data were from the Beef CRC I, II, III, and an Accelerate Partnership project co-funded by the Queensland Government with the CSIRO, the North Australian Pastoral Company and the University of Queensland. Whole-genome sequences were retrieved from the 1000 Bull Genome Project (run6, taurus and indicus). The selected breeds chosen as reference for imputation were the Brahman, Composites, Afrikander, Boran, Nelore, Gir, Charolais and Angus.

**Table 1. Cattle samples with SNP array genotypes that were imputed up to sequence.**

<table>
<thead>
<tr>
<th>Cattle breed</th>
<th>&lt;30K(^1)</th>
<th>Bovine SNP50v1 or v2</th>
<th>GGP Indicus ~80K</th>
<th>GGPHD ~150K</th>
<th>BovineHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belmont Red</td>
<td>289</td>
<td>378</td>
<td>97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brahman</td>
<td>396</td>
<td>4,124</td>
<td>519</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tropical Composite</td>
<td>2,044</td>
<td>10,832</td>
<td>30</td>
<td>391</td>
<td></td>
</tr>
<tr>
<td>Crossbred</td>
<td></td>
<td>999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Droughtmaster</td>
<td></td>
<td>419</td>
<td></td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Santa Gertrudis</td>
<td>358</td>
<td>1,040</td>
<td>168</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3,087</td>
<td>16,793</td>
<td>999</td>
<td>30</td>
<td>1,220</td>
</tr>
</tbody>
</table>

\(^1\)Illumina low density (~7K) or Zoetis low density (~25K).

### Imputation of SNP genotyped using arrays up to whole-genome sequence

All genetic variants derived from the sequence of selected animals were extracted and filtered to keep only bi-allelic variants with at least four copies of the minor allele within the selected animals. The genotypes of the filtered variants were then phased using Eagle (Loh *et al.*, 2016). The genotypes derived from the diverse range of SNP arrays were impute up to BovineHD (Illumina Inc. San Diego, ~800K SNP) using either Beagle v4.1 (Browning and Browning, 2016; Browning and Browning, 2007) or FImpute 2.2 (Sargolzaei *et al.*, 2014). A second filter was applied on imputed SNP keeping only variants from the BovineHD that were segregating on the sequenced animals (~650K SNP). Moreover, allele calls had to be equal between technologies (some SNP had to have its strain flipped). The final imputation of SNP derived from array up to sequence was performed using FImpute 2.2 with phased sequenced animals as reference.

### Genome-wide association for sheath score
A subsample of 964 tropical composite cattle had been recoded for sheath score (Porto-Neto et al., 2014; Prayaga et al., 2009). Adjusted phenotypes were regressed on SNP genotypes. The BLUE solutions for the fixed effects were used to adjust the phenotype, they were calculated from fitting a mixed model into Qxpak v5 (Perez-Enciso and Misztal, 2011) that included two fixed effects (management group and year of birth), and animal as a random effect via a genomic relationship matrix (VanRaden, 2008) calculated using BovineHD SNP. Subsequently, after capturing the most associated SNP, a second GWAS was performed fitting this SNP as an additional fixed effect.

Results and discussion

A dataset of more than twenty-two thousand animals from seven breed types, with imputed genotypes for more than 40 million markers was generated. The number of markers per chromosome varied from 674,062 (BTA25) to 2,415,205 (BTA1) (Figure 1). These animals were part of several different experiments with a number of traits recorded. Many of them are part of a commercial breeding herd, which has been followed by more than ten generations. The number of records per trait is variable, and traits included male and female fertility indicators (age at puberty, scrotal circumference), conformation (coat score, colour, condition score, height), adaptation (tick counts, rectal temperature, flight time), growth (body weight and weight gains), carcase traits, and others.

Figure 1. Distribution of genomic variants imputed per chromosome.

The cattle chromosome 5 (BTA5) is known to harbor several QTL for production and adaptation-related traits (Porto-Neto et al., 2014), it also has been associated to QTL of pleiotropic effects (Saatchi et al., 2014). Using the newly imputed dataset, a GWAS for sheath score on tropical beef cattle was conducted on BTA5 (Figure 2, black line) with over 1.8M SNP. The previously described broad association peak at ~ 48Mb was detected also in this study. Of the fifteen most associated SNP, twelve of them had genotyped imputed, which gives confidence on the imputation process (more information on this imputation pipeline and additional analysis using this dataset can be found at Hayes et. all 2018, this proceedings).

When the most associated SNP was fitted as a fixed effect, most of initial association signals vanished (Figure 2, red line), suggesting that if this SNP is not the functional mutation, it is highly linked to. The variance component analysis showed that the inclusion of the top significant SNP as a fixed effect had only impacted on the genetic component, leaving the residual variance unaffected by the additional fixed effect, keeping it within a standard error (0.31 and 0.29 respectively). Among the top ten highly associated SNP at the first GWAS run, five of them fall within predicted regulatory regions in the cow genome (Nguyen et al., 2017). Therefore, suggesting a potential impact in the gene expression levels of neighbouring genes.
Figure 2. Association between makers imputed on chromosome 5 and sheath score of tropical composite beef cattle. Black line is the first round of association, and the red line is the second round, where the most associated SNP from round one was fitted as a fixed effect.

A genomic resource of around twenty-two thousand cattle with more than forty million imputed genetic variants was created, opening several scientific opportunities not yet available for tropical beef cattle. The resolution for QTL mapping and marker selection to form tailored SNP panels for genomic selection increased several folds. It is envisaged that this resource will be further explored and expanded for genetic analyses in the near future.

Acknowledgements

The authors are thankful to all participants of the Beef CRC I, II and III that contributed to the generation of this resource. The authors are also grateful to all contributors of the 1000 Bulls Genome Project that made genetic variants available. The authors would like to acknowledge the contributions from the North Australian Pastoral Company and the Queensland Government that partly funded this project.

List of references


