# Final report

## Bull fertility update: historical data, new cohort and advanced genomics

Project code: L.GEN.1818

Prepared by: Laercio Porto-Neto, Sean McWilliam, Pamela Alexandre, Antonio Reverter
Commonwealth Scientific and Industrial Research Organisation
Michael McGowan, Marina Fortes, Ben Hayes
The University of Queensland
John Bertram
Agriculture Consultant, Livestock management and breeding

Date published: 30 June 2021

PUBLISHED BY
Meat and Livestock Australia Limited
PO Box 1961
NORTH SYDNEY NSW 2059

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

# Abstract

Considering that most beef breeding herds are bull mated worldwide, bull fertility is a main driver of on-farm productivity, including live-weight production, and it affects the whole beef supply chain. This project combined data from the CRC for Beef Genetic Technologies with new data, collected in partnership with producers, to create one of the most complete datasets on bull fertility traits with matched DNA profiles. The traits are those recorded in the bull breeding soundness examination (BBSE). The assembled dataset included nearly 7,000 records in six tropical breed types. Using this dataset, we: a) confirmed that BBSE traits were heritable and influenced by many genes, suggesting that the best method for trait improvement is via genomic selection approaches, b) were able to identify DNA markers and genomic regions that affect the fertility traits, and c) were able to, using this multibreed reference population, calculate GEBV for fertility traits with useful accuracy for the Beef Industry. Further work is needed to better characterize the genes affecting male fertility traits and to implement and consolidate the genomic selection approaches developed during this project.

# Executive summary

## Background

Worldwide most of the beef breeding herd is subjected to natural mating where bull fertility is essential to reproductive rates and live weight productivity, affecting the whole supply chain. The direct observation of fertility indicators, e.g. number of calves per sire, is rarely assessed due to high costs. Also, it would be performed later in the life of a sire when there is a reduced value of the information compared to a pre-purchase assessment. The standardized bull breeding soundness examination (BBSE) was developed to fill this gap; it is based on several criteria, from the animal's conformation to sperm morphology. Importantly the BBSE assesses traits that are known to be linked to the bull's calve getting ability and can be used to inform purchase and breeding decisions since the observed traits are heritable. To date, there hasn't been a comprehensive evaluation of the genetic background of BBSE traits, neither the exploration of their potential for multibreed genomic evaluation. The overarching aim of this project was to fill-in this knowledge gap by investigating the genomics of BBSE traits.

Bull breeders and bull buyers are the primary audiences for this research - both are interested in identifying more fertile animals. In Australia, only a small fraction of replacement herd bulls sold every year are part of any formal genetic improvement program aimed at productive attributes, including fertility. A better understanding of the genetic architecture of BBSE traits, combined with genomic selection approaches that were part of this project, can assist with adoption of the technology. The Beef Industry benefits from this research that builds capacity to use genomic information to inform purchasing decisions (by bull buyers that seek fertile animals) and to drive genetic improvement (by bull breeders). In short, this research paves the way to improved bull fertility, which will lift productivity for the Australian beef breeding herd.

## Objectives

There were two main objectives in this project:

- Update the Beed CRC legacy dataset to the latest technology. During the Beef CRC projects, especially Beef CRC III, an impressive dataset was collected on cattle fertility. The phenotypic records were paired with a SNP profile, mostly at around 50,000 SNP markers density. We proposed and successfully applied imputation approaches to fill in missing trait records and increase marker density from the original SNP chip up to the whole-genome sequence (around 22 million SNP).
- Expand the Beef CRC dataset on bull fertility and generate impactful outcomes. We achieved this objective by increasing the number of bulls evaluated from around 2,800 up to nearly 10,000 bulls with BBSE records. Around 7,000 of these bulls have a matching DNA profile. Using this newly combined dataset, we exploited the genetic architecture of BBSE traits to identify the most relevant genomic regions associated with the traits and we derived multibreed genomic breeding values with useful accuracies.

## Methodology

Several methodologies were applied to the Beef CRC and the newly combined datasets. The software SAS was used to a) impute missing phenotypes of the Beef CRC dataset using the multiple imputation procedure and b) adjust the phenotypes before the multibreed analyses. The genotype

imputation was performed in two steps, first from native SNP chip up to High-Density (HD) with around 700,000 SNP, then from HD up to whole-genome sequence with around 22 million SNP. Both steps included genotype phasing followed by imputation using two specialized software's (Eagle and Minimac). These genotypes were used to discover regions of the genome associated with BBSE traits in genome-wide association studies (GWAS) and to predict breeding values with GBLUP. The GWAS analyses used the GREML procedure implemented in GCTA software, and the parameter estimate used GBLUP implemented in Qxpak software.

## Results/key findings

The newly assembled dataset of BBSE records with matching SNP profiles includes nearly 7,000 bulls with SNP genotypes, representing six cattle breed types. This is possibly the most extensive dataset on bull fertility in Australia.

Heritability estimates varied from low (0.17 body condition score and 0.19 sperm progressive motility) to moderate (0.46 scrotal circumference and 0.52 sheath score). Although some traits showed a few genomic regions with highly significant associations, all traits seem to be controlled by several genes (meaning they have a polygenic architecture). Surprisingly, the genomic correlation of the same trait when observed in different breeds was very low for most of the traits, except for scrotal circumference and sheath score. The interpretation of this result is that in different breeds, different genes or mutations maybe driving the studied traits.

We applied a pleiotropy test to identify markers relevant to several traits at the same time. This test allowed the identification of two key chromosomes (5 and X) that harbor most of the top makers, with additional markers on chromosomes 6, 8, 10, 21, and 24. These genomic regions can be prioritized and further explored to better understand the genes controlling bull fertility traits.

Using a multibreed genomic selection approach, we derived estimated breeding values (GEBV) for fertility traits with useful accuracies. Additionally, we demonstrated the importance of a breed to being represented within the reference population when estimating breeding values in a multibreed scheme.

## Benefits to Industry

By utilizing the data contributed by five Queensland bull breeders, we collated a multibreed reference population on bull fertility traits of tropical cattle. All traits observed followed a polygenic pattern, meaning that multiple regions across the genome were associated with the bull fertility traits studied in this project. Therefore, the best approach for bull fertility improvement would be via genomic selection approaches for which this project sets the basis.

The reference population used was big and diverse enough to produce multibreed GEBV with useful accuracies for bull fertility traits for the Industry. Our analyses highlighted the importance for the specific breed of interest to be part of the reference population when estimating multibreed breeding values. To estimate the most accurate GEBVs for the particular breed in question, this breed needs to be included in the reference dataset. By creating a multibreed dataset, this project paved the way for the development of GEBVs for bull fertility in multiple breed types.

## Future research and recommendations

Based on the positive results of this project, we recommend:

- Working with Industry to grow the multibreed reference population used to investigate bull fertility traits. We suggest establishing a process to engage with Industry, to collect and compile BBSE records with matched DNA profiles to cement a multibreed population in our research arena. Importantly, many bull breeders already pay for BBSE and so it is of great value to leverage these existing records for genetic improvement.
- Undertaking further research on the genomic architecture of the bull fertility traits investigated in this project to a) better understand the genes controlling these traits, b) explore further avenues to assist in improving the accuracy of genomic prediction (GEBV), and c) investigate other potential uses of the defined functional mutations affecting these traits.
- Establishing a process for delivering multibreed genomic predictions for bull fertility traits to the Industry.

# Table of contents

# 1. Background

The Background, Objectives and Methodology sections are updated text extracted from the original proposal.

## 1.1 Background of research work and significance

We acknowledge the significant previous investments in the Beef CRC programs (I, II and III) made by MLA and other core participants. The Beef CRC legacy database represents investments of millions of dollars and 20 years of research. This database is still very relevant to beef industry R&D, and continue to be one of the most recorded populations of tropical cattle worldwide. In recent years a number of scientific publications were written using this resource. To date, most of these publications have been genome-wide association studies (GWAS) with just one or two identifications of causative mutations. Now, with current genomic sequencing resources it will be possible to advance beyond GWAS and identify many more causative mutations.

Recent publications include analyses on carcase traits (Bolormaa et al, 2011), the identification of markers that affect multiple traits (Bolormaa, et al. 2014), analyses on inbreeding depression (Reverter et al. 2017), identification of markers associated with traits related to tropical adaptation (Porto-Neto et al. 2014), male and female fertility (Hawken et al. 2012; Fortes et al. 2012), and developments towards multibreed parameter estimation (Porto-Neto et al. 2015). This important earlier work can now be upgraded, as none of the above-mentioned studies used whole-genome sequence data. Sequence data is fast becoming the gold standard for genomic selection and research, holding great potential for future commercial implementation.

Bull fertility-related phenotypes had not benefitted from recent advances in genetic technologies, even though they unquestionably contribute to the success of any beef breeding enterprise, until now. There are still several conditions (or pathologies) of the male reproductive tract with reasonably high prevalence in registered and unregistered cattle that directly impacts on-farm productivity. These are often detected only later in life (e.g. spiral deviation in Angus cattle), when the bull has already served for a few years, and potentially spread the condition in its herd. Therefore, a molecular test for those male conditions has the potential to guide early selection to remove those carriers from the breeding population. Removal of these pathologies that impact not only fertility but also the general health of bulls has animal welfare implications, in addition to improving herd fertility.

Bull breeding soundness evaluation (BBSE) is a very useful practice that can detect most of the conditions affecting a bull's calf-getting ability. Some of those conditions have early onset, e.g. early puberty (high percentage of normal sperm at early age and high scrotal circumference) and testicular hypoplasia. Other conditions have late onset, e.g. spiral penile deviation and persistence of penile frenulum. There were already several years-worth of records on BBSE in beef cattle available to this project, and many records also had a stored DNA sample that were genotyped in this project. The combination of CRC data, and additional historical records that were collected during the project allowed a comprehensive update of the research resources for bull fertility traits. Now, we have created a comprehensive dataset update to the latest genomic technologies.

CSIRO and UQ have contributed genome sequences to the 1000 Bulls Genome Project, which has granted access to the combined dataset. The 1000 Bulls Genome Project dataset comprises more than 4,000 cattle, and 42 million genetic markers. This is a marker density of at least 60x greater than all previous analyses undertaken during the Beef CRC. There is growing evidence about

accuracy improvements to be made when causative mutations are identified (Perez-Enciso et al. 2015), and their potential significant contribution for multibreed genetic evaluation (Kemper et al. 2015). In a nut-shell, using sequence data we have added value to the Beef CRC legacy, expanded greatly the resources for bull fertility RD&A and fostered technologies to overcome many of the challenges faced by the tropical Beef Industry when it comes to implementation of genomic selection.

A game-changer is the accurate imputation of genotypes from 50K SNP chips up to whole-genome sequence. In this project, we have invested in updating already available resources, along with the new male fertility dataset with the four-fold objective to:

- Leverage millions of dollars of previous R&D investment,
- Improve the power of previous analyses,
- Further our understanding on male fertility-related traits and identifying specific markers affecting those traits, and
- Assist future projects by supporting and guiding scientific endeavours in livestock genomics.

Additional gains were made with the application of advanced multi-variate analyses for the imputation of missing phenotypes aiding the validation of candidate causative mutations. The combination of genotype and phenotype imputation within the Beef CRC legacy database has strengthen our position as world leaders in tropical cattle genomics.

Re-organisation, re-purposing and "updating" of the Beef CRC Legacy Database with whole-genome sequence and phenotype imputation has also create the potential to identify new and unforeseen research opportunities. Within this project, we have updated and re-analyse the Legacy Database, followed by the analysis of the new data that was collected in partnership with Australian bull breeders, as part of the project. We have collaborated with producers that regularly perform Bull Breeding Soundness Evaluations (BBSE) and have therefore obtained high quality phenotypes and matching DNA genotypes.

## 1.2 Summary of benefits to a grassfed producer

The Beef CRC legacy dataset was collected over 20 years (three CRC rounds) and represents millions of dollars of previous R&D investment. The tropical cattle component of this dataset includes more than 10,000 cattle genotyped, as well as data on a large number of traits, ranging from health traits, to parasite resistance and male/female fertility. This dataset is still regarded as one of the most comprehensively annotated populations available for the Australian Beef Industry. However, with the end of the Beef CRC, the database has not benefitted from recent advances in genomic technologies. Indeed, recent years have shown exponential developments in genomics, with the smart use of whole-genome sequences as one of the next frontiers for livestock geneticists.

This project has leveraged previous investments in the Beef CRC by updating the existing legacy dataset with the most recent genomic technology: the use of whole-genome sequences and phenotype imputation. Additionally, this project expanded the Beef CRC dataset with the inclusion of new samples focused on male fertility traits, which is an important topic for grassfed producers that rely on bulls to breed their cow herd.

Male fertility traits are among those traits that would benefit the most with advanced genetic technologies. Common fertility traits of bulls are related to semen quality and scrotal size, but there are several conditions that are often neglected, specially, persistent penile frenulum, testicular

hypoplasia, sheath score, and penile spiral deviation. Importantly, all of these pathologies impact the calf-getting ability of a bull, and some have known genetic correlations to female fertility traits. Known correlations favour cross-sex selection (breeding approaches focused on both male and female fertility). It is worth noting that several pathologies of the male reproductive tract are only detectable later in life, after the bull has been mated for a few seasons and potentially spread the unfavourable genetic condition. Therefore, a molecular test able to detect those conditions would significantly impact the whole bull breeding system, allowing earlier detection and removal of disease carriers before joining. Research conducted in this project sets the scene for identifying more fertile bulls using genomic data. Further development of the genomic estimated breeding values (GEBV) as proposed in this project will benefit grassfed producers because it will lead to identifying bull calves with higher genetic merit for fertility traits, which can then be purchased as bulls that will lift herd reproductive rates.

The use of whole genome sequence provides the best chance of identifying causative mutations associated with Bull fertility traits. This project has advanced our understanding of bull fertility by linking biological knowledge, sequence resources and new datasets, for the identification of mutations that will underpin more accurate estimates of breeding values. The power of associated mutations was reported before (MLA project B.NBP.0786) and confirmed in this study.

# 2. Objectives

## Objective 1

Updated the Beed CRC legacy dataset to the latest technology. During the CRC projects, especially the Beef CRC III, an impressive dataset of traits (production, reproductive performance, and fertility data) was collected. This dataset was linked to DNA profiling done using state-of-the-art technology at that time. However, time has passed, and technology has evolved. The phenotypic data is still relevant, but the DNA profiling should be updated. We proposed using bioinformatic tools to bring the phenotypic and genotypic data up to the latest technology.

This objective was successfully met; genotype and phenotypic data were worked on, and new analyses undertaken.

## Objective 2

Expanded the Beef CRC dataset with the inclusion of new bull fertility data linked to DNA profiles. The additional recorded traits are from Bull Breeding Soundness Examinations (BBSE) including a full semen analysis.

The combined dataset (Beef CRC plus newly collected data) was used in genomic analyses that identified genomic regions and DNA markers associated to fertility traits. The combined data was used to develop GEBVs and evaluate the effectiveness and feasibility of multibreed genomic estimation of genetic merit for bull fertility traits.

This objective was successfully met. From the initial Beef CRC dataset of around 2,800 bulls with a bull breeding soundness examination, the dataset grew to nearly 10,000 bulls with phenotypic records, of which 7,000 have a matched DNA profile (SNP genotypes). This new dataset was then used to explore the genetic architecture of bull fertility traits and to derive multibreed genomic predictions for the same traits. We established that it is possible to predict genetic merit for BBSE traits using a multibreed approach and genomic data.

# 3. Methodology

## 3.1 Imputation of phenotypic data within the Beef CRC legacy dataset

### 3.1.1 Phenotype imputation using the multiple imputation procedure

A general linear model analysis (PROC GLM) fitted in SAS 9.4 (SAS Inst. Inc.) was used to adjust the observed phenotypes for its specific fixed effects (e.g. contemporary group and lab batch assay). In addition, the estimated indicine percent and the age at measurement were fitted as covariates. The adjusted phenotypes were then used as inputs for the phenotype imputation procedure.

Missing phenotypes were imputed using the Multiple Imputation (PROC MI) Procedure of SAS 9.4 with a Markov Chain Monte Carlo (MCMC) method that assumes multivariate normality and creates multiple imputed datasets for incomplete p-dimensional multivariate data with arbitrary missing patterns (https://support.sas.com/documentation/onlinedoc/stat/141/mi.pdf).

According to Schafer, 1997 ('Analysis of Incomplete Multivariate Data', New York: Chapman and Hall. pp. 147–148), inferences based on multiple imputations can be robust to departures from multivariate normality if the amount of missing information is not large because the imputation model is effectively applied not to the entire data set but only to its missing part. A recent example of imputing missing phenotypes in the context of livestock breeding and genetics can be found in Bolormaa et al. 2017 (Bolormaa S, Swan AA, Brown DJ, Hatcher S, Moghaddar N, van der Werf JH, Goddard ME and Daetwyler HD. Multiple-trait QTL mapping and genomic prediction for wool traits in sheep. Genet Sel Evol. 2017, 49(1):62). In that work, for animals without records for a particular trait, missing phenotypes were predicted using a single round of a multiple regression approach.

The advantage of the MI Procedure of SAS is that sensitivity analyses yielding a measure of phenotype imputation accuracy are possible because multiple imputed datasets are created with an internal cross-validation scheme.

### 3.1.2 Genomic predictions using the imputed dataset

Genotype information for all animals included in the analyses were used to compute a genomic relationship matrix (GRM) following method 1 of VanRaden (VanRaden, 2008). Variance components and heritability were estimated using GBLUP implemented in the Qxpak5 software (Pérez-Enciso and Misztal, 2011). Two rounds of univariate models were used for each trait. The first was undertaken using the reduced dataset that had missing values, while the second round used the fully complete dataset by combining the observed and the imputed phenotypes. After these initial runs, a five-way cross validation using the same GRM was undertaken, by setting to missing value 20% of the phenotype data. The GEBV from the analyses using all data was used as the calibration, and the subsequent cross validation were the validation.

Accuracy of predicted breeding values were calculated using the correlation-based approach and the Method LR (Legarra and Reverter, 2018) approaches. Additional, with the Method LR, bias and dispersion were also computed. For additional descriptive information please refer to section 3.4 of this report on “Estimating multibreed genomic breeding values for male-traits”.

## 3.2 Collection of new dataset with bull fertility indicator traits

### 3.2.1 Properties enrolled

Five bull breeders from Queensland contributed data to this new dataset. These were approached to be part of the project for four reasons primarily: : a) previous interest shown in collaborating with researchers, b) routine parent-verification of their bulls using DNA tests, implying that for most of their animals a DNA sample should be available in storage, c) routinely perform a BBSE (with the same technical officer), and d) representative of a diverse population of tropical cattle.

The collaborators' herd breeds were (Figure 1, North to South): Droughtmaster (Lisgar at Gumlu), Brahman (Jimarndy at May Downs), Belmont Tropical Composite (Tremere at Moura), Santa Gertrudis (Gyranda at Theodore), and Ultra-black (Nindooinbah at Beaudesert).

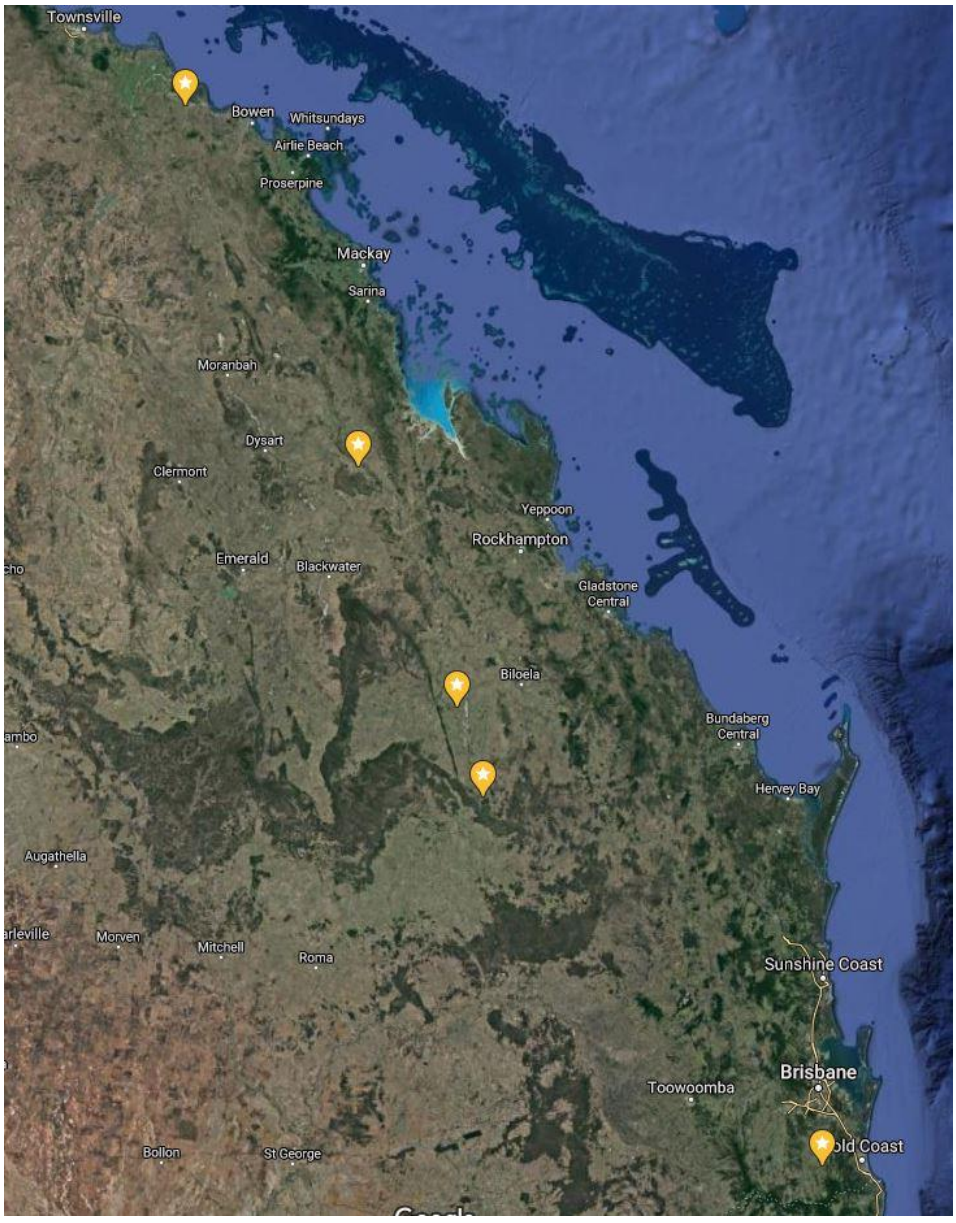


Figure 1. Location of the collaborating herds.

### 3.2.2 Bull breeding soundness examination (BBSE) data

A full bull breeding soundness examination (BBSE) makes observations on the physical aspects of the animal, and its semen quality, including a sperm morphology assessment and count of sperm defects (Table 1). The BBSE was developed to be a routine assessment of the serving capability of bulls, which would be ideally performed every year before each mating season. However, in most cases, BBSE is used as a one-off assessment before sale to indicate that the bull will perform as a sire. The combination of a physical evaluation with analyses on a semen sample gives the best proxy for routine evaluation and, importantly, can be also used to drive selection towards a preferred trait. The measurements in a BBSE test are good indicator traits for bull fertility (Fordyce et al., 2006).

**Table 1. Recorded information observed at the bull breeding soundness evaluation (BBSE).**

| Group | Observation | Scale |
|---|---|---|
| Animal description | Animal_id | - |
| | Farm/Breed | - |
| | Management group | - |
| | Sire | - |
| | Dam | - |
| | Date of birth | - |
| BBSE info | Date of BBSE | - |
| | Age at BBBSE | - |
| Physical evaluation | Weight at BBSE | Kg |
| | Body Condition Score (BCS) | 1-5 |
| | Leg Structure | 1-S/hck; 5-Post leg |
| | Leg Joints | N/A |
| | Feet | Sciss Claw 1-mild, 3-severe N/A |
| | Gait | N/A |
| | Head | N/A |
| | Sheath Score | 1-Tight; 5-V/Pend. |
| | Penis | N/A |
| | Preputial eversion | cm |
| | Scrotal size | cm |
| | Test tone | 1-Soft, 5 hard |
| Semen quality crunch side | Density | 1-5 |
| | Mass activity | 1-5 |
| | Progress motility | % |
| | Normal | % |
| Semen morphology | Normal sperm | % |
| | PDs (Proximal cytoplasmic droplets) | max 20% |
| | MP (Midpiece defects) | max 30% |
| | T&H (Tail & head defects) | max 30% |
| | Py (Pyriform heads) | max 20 % |
| | KA (Knobbed acrosomes) | max 30 % |
| | V&T (Vacuoles/ teratoids) | max 20 % |
| | SA (Swollen acrosomes) | max 30% |
| Obs | Comments. Specific conditions (e.g. spiral deviation, testicular hypoplasia) | |

### 3.2.3 Source of biological samples and DNA profiling – SNP genotypes available

We made use of DNA samples archived at the Neogen laboratory at Gatton (The University of Queensland campus). Most of these historical samples were submitted by the bull breeders for parent verification (or assignment). With their consent, we accessed the archived DNA samples and genotyped them using the Neogen Tropical cattle chip that contains around 54,000 markers (GGP TropBeef50K). In addition to these samples genotyped by the project, some bull breeds also made available DNA profiles on animals previously genotyped by them. These early DNA profiles often were done using a different SNP platform (e.g. Neogen 35K SNP). It is worth noting that during the SNP genotype imputation it is possible to accommodate and merge different SNP platforms, which we did.

## 3.3 Imputation of SNP genotypes from lower to higher density

The genotype imputation is a multi-step procedure. First we impute the native platform (20K, 35K, 50K or 80K SNP chip) up to Bovine high-density (HD, ~700,000 SNP), for this step we used Beef CRC data on animals that were actually genotyped with the Bovine HD chip directly, as the reference population. We then imputed the HD genotypes up to whole genome sequence, using data from the 1000 Bulls Genome Project (http://www.1000bullgenomes.com/, run8).

The step-by-step procedure was as follows:

- Samples with actual HD data (~700,000 SNP) had its genotypes phased using the software Eagle (Loh et al., 2016) to be used as "reference" in a later step;
- Genotypes of lower density chips (Bovine SNP50 v1 or v2 or Neogen Tropical Chip v1 and v2) were also phased using Eagle, but this time imputation of missing genotypes was not performed. These samples were the "target" in the following step;
- Genotype imputation of lower density ("target") up to high-density ("reference") was performed using the software Minimac 3 (Das et al., 2016) (autosomes) and Minimac 4 (X chromosome);
- SNP genotypes for 668 animals were extracted from the 1000 Bulls Genome Project (Table 5). The raw data was filtered such that only bi-allelic* DNA markers were kept, and these had to have at least four copies of the minor allele in this population. These whole-genome sequence samples were phased as per step (a) and used as a reference for the last step;
- Our samples recently imputed up to HD were then phased using the whole-genome data;
- Then, using the same procedure of (c) the samples were imputed from HD up to whole-genome sequence.

After the procedure, SNP with imputation quality score (rsq) >0.8 were kept for future analyses.

*Note: bi-allelic DNA markers are the informative markers because they vary in the population: two alleles are present. When only one allele is present, fixed, in the population the marker is not informative.

## 3.4 Estimating multibreed genomic breeding values for male-traits

**Data included in the analyses.** These multibreed analyses were performed using data from six populations: Beef CRC Brahman and Tropical Composite, Santa Gertrudis, Droughtmaster, Ultra-Black and Belmont Tropical Composite.

**Data adjustment and model implementation.** The phenotypes were adjusted using SAS 9.4 (www.sas.com) before the genomic analyses. The model for adjustment included the fixed effects of population, year of birth and cohort, and the covariates of age and the first two principal components based on imputed HD genotypes. We estimated genetic parameters and predicted genomic-estimated breeding values (GEBV) using the multi-variate genomic-relatedness-based restricted maximum likelihood (GREML) approach as implemented in Qxpak5 software (Pérez-Enciso and Misztal, 2011).

**Genomic correlations between traits and populations.** To estimate genetic correlation between pairs of traits, a total of 45 bi-variate GREML analyses were performed for all pair-wise phenotypes and across the six breeds (i.e., each using a GRM of dimension 6,063 to include all bulls). In addition, following the analytical approach described in Porto-Neto et al. (2015), the genomic correlation for a given phenotype in two breeds was estimated by treating each phenotype as a different trait in each breed pair. As a result, we performed a total of 150 bi-variates GREML analyses (15 pairs, and 10 BBSE phenotypes) each with a GRM of dimension equal to the number of animals in the breed pair.

**Assessment of accuracy of the estimated genomic breeding values.** A correlation-based approach (Bolormaa *et al.* 2013) and the Method LR (Legarra and Reverter 2018) were used to estimate accuracy of GEBVs. The Method LR was also used to calculate bias and dispersion of GEBVs.

The following four metrics were employed:

- Correlation-based Accuracy ($ACC_R$): In the context of cross-validation, the accuracy of a GEBV is traditionally computed from the Pearson correlation between a GEBV and the adjusted phenotype (*y**; phenotype *y* adjusted for fixed effects) for individuals in the validation population, and divided by the square root of heritability:

$$\mathrm{ACC_R} = \frac{r\left(\widehat{\boldsymbol{u}}_p, \boldsymbol{y}^*\right)}{\sqrt{h^2}}$$

- Method LR Accuracy ($ACC_{LR}$): For individuals in the validation population, the method LR accuracy was computed as follows:

$$\mathrm{ACC_{LR}} = \sqrt{\frac{cov\left(\widehat{\boldsymbol{u}}_w, \widehat{\boldsymbol{u}}_{\boldsymbol{p}}\right)}{\left(1 + \bar{F} - 2\bar{f}\right)\sigma^2_{g,\infty}}}$$

Where $\bar{F}$ is the average inbreeding coefficient, $2\bar{f}$ is the average relationship between individuals, and $\sigma^2_{g,\infty}$ is the genetic variance at equilibrium in a population under selection. Assuming the individuals in the validation population are not under selection, $\sigma^2_{g,\infty}$ can be estimated by the additive genetic variance estimated from the partial dataset.

Regardless of method, the interpretation of accuracies is the same: the higher the accuracy, the more reliable are the GEBVs and therefore the choice of bull for breeding.

- Method LR Bias ($\text{Bias}_{LR}$): Difference between the average GEBV of individuals in the validation population using the partial data minus that same parameter but using the whole data:

$$\text{Bias}_{\text{LR}} = \overline{\widehat{\boldsymbol{u}}_p} - \overline{\widehat{\boldsymbol{u}}_w}$$

In the absence of bias, the expected value of $\text{Bias}_{\text{LR}}$ is zero.

- Method LR Dispersion ($\text{Disp}_{\text{LR}}$): For individuals in the validation population, dispersion was measured from the slope of the regression of $\widehat{\boldsymbol{u}}_w$ on $\widehat{\boldsymbol{u}}_p$:

$$\text{Disp}_{\text{LR}} = 1 - \frac{cov(\widehat{\boldsymbol{u}}_w, \widehat{\boldsymbol{u}}_p)}{var(\widehat{\boldsymbol{u}}_p)}$$

In the absence of bias, the expected value of $\text{Disp}_{\text{LR}}$ is zero. Values less than zero indicate under-dispersion (or deflation) of $\widehat{\boldsymbol{u}}_p$ into $\widehat{\boldsymbol{u}}_w$ as phenotypes become available. Values greater than 1 indicate over-dispersion (or inflation) of $\widehat{\boldsymbol{u}}_p$ into $\widehat{\boldsymbol{u}}_w$.

## 3.5 Genome-wide association study and SNP selection

Genome-wide association studies (GWAS) were run within breed, at whole-genome sequence level using imputed genotypes described in 3.3, and the 10 adjusted phenotypes described in 3.4. In total, sixty GWAS were run. For these GWAS, we used GREML implemented in the GCTA software (Yang et al., 2010), using genomic relationship matrices (GRM) based on imputed HD genotypes (~700K SNP). We fit two GRMs for each chromosome analyses, the first GRM had all autosomes but one (the chromosome being tested) plus the pseudo autosomal region of the chromosome X, and the second GRM had the non-autosomal segment of the chromosome X. For example, when testing for association on chromosome 1, the first GRM included the chromosomes 2 to 29, plus the pseudo autosomal region of chromosome X, and the second GRM had the other segment of chromosome X. Significance was assessed by a t-test on the distribution of the SNP effects, with association p-values lower than $1x10^{-5}$ deemed significant.

Additionally, also based on the whole-genome sequence using two GRM implemented in GCTA (Yang et al., 2010), case-control association analyses were run for some specific conditions e.g. the spiral penile deviation, retention of frenulum, testicular hypoplasia, and swollen of the hocks.

To further characterize the pleiotropic potential of SNP, we estimated SNP effects for each trait using derivations from Stradén and Garrick (2009) and Wang et al. (2012) as follows:

$$\hat{\boldsymbol{s}} = \lambda \mathbf{M}^T \mathbf{G}^{-1} \widehat{\boldsymbol{u}}_w$$

in which $\hat{\boldsymbol{s}}$ is the vector of estimated SNP effects of dimension 680,758 for as many SNP included in the analyses, $\lambda$ is the ratio of SNP variance to genetic variance and assumed 0.85 throughout, $\mathbf{M}$ is the matrix of genotypes centred for current allele frequencies with dimension equal to the number of animals (6,063) by number of SNP (680,758), $\mathbf{G}$ is the GRM computing using Method 1 of VanRaden (2008) across all animals and SNP, and $\widehat{\boldsymbol{u}}_w$ is as defined earlier for the trait of interest. we

computed the multi-trait $\chi^2$ statistic for the *i*-th SNP following derivations from Bolormaa et al. (2014):

$$\chi_i^T = \hat{\boldsymbol{s}}_i^T \mathbf{V}^{-1} \hat{\boldsymbol{s}}_i$$

where $\hat{\boldsymbol{s}}_i$ is the 10 (number of traits) × 1 vector of z-score standardized effect of the *i*-th SNP and $\mathbf{V}^{-1}$ is the inverse of the 10 × 10 correlation matrix calculated over all estimated SNP effects. The $\chi^2$ value of each SNP was assessed for significance based on a $\chi^2$ distribution with 10 degrees of freedom to test against the null hypothesis that the SNP had no significant effect on any of the ten traits.

# 4. Results

## 4.1 Phenotype imputation of the Beef CRC dataset

### 4.1.1 Evaluation of the missing data before imputation and summary statistics of the imputed dataset

For the current work, data from 1,116 Brahman bulls (Table 2) and 1,453 Tropical Composite bulls (Table 3) were used across 38 phenotypes related to tropical adaptation, growth, fertility and semen characteristics. The Tropical Composite bulls had 429 records on EPG (worm eggs per gram of faeces), but this phenotype was not available in Brahman.

The fraction of missing data varied greatly from trait to trait, from 0% (e.g. Flight time (FT) and Sheath score (SHEATH), among others on both breeds) to 77% (Inhibin 4 (IN4)) in Brahman and ~90% of some semen observations in Tropical Composite (Figure 2). All traits were observed in 12.28% of the animals (or 137) while 30.47% of animals (or 340) had only one observation missing and 53.67% of animals (or 599) had 10 or more observations missing. These values compare with the Tropical Composite bulls dataset where 13.49% of records (or 196) had an observation for all traits while 18.10% of records (or 263) had only one observation missing and 50.58% of records (or 735) had 10 or more observations missing.

Although not being the main focus of this project, as part of the dealings with historical Beef CRC records, data from 995 Brahman cows and 1,094 Tropical Composite cows were used across 18 phenotypes related to tropical adaptation, growth, and fertility (Appendix 1). For the Brahman cows dataset, 36.68% of records (or 365) had observation for all traits while 36.88% of records (or 367) had only one observation missing. These values compare with the Tropical Composite cows dataset where 15.08% of records (or 165) had an observation for all traits while 47.17% of records (or 516) had only one observation missing.

For both bulls' and cows' datasets, phenotypes were corrected for fixed effects and regression covariates before imputation. The imputation was based on the MCMC method with a single chain to create 10 imputations each with 200 burn-in iterations and 500 iterations between imputations. After the completion of the 10 imputations the relative efficiency (measured in units of variance) was over 90% for all variables (Figure 2).

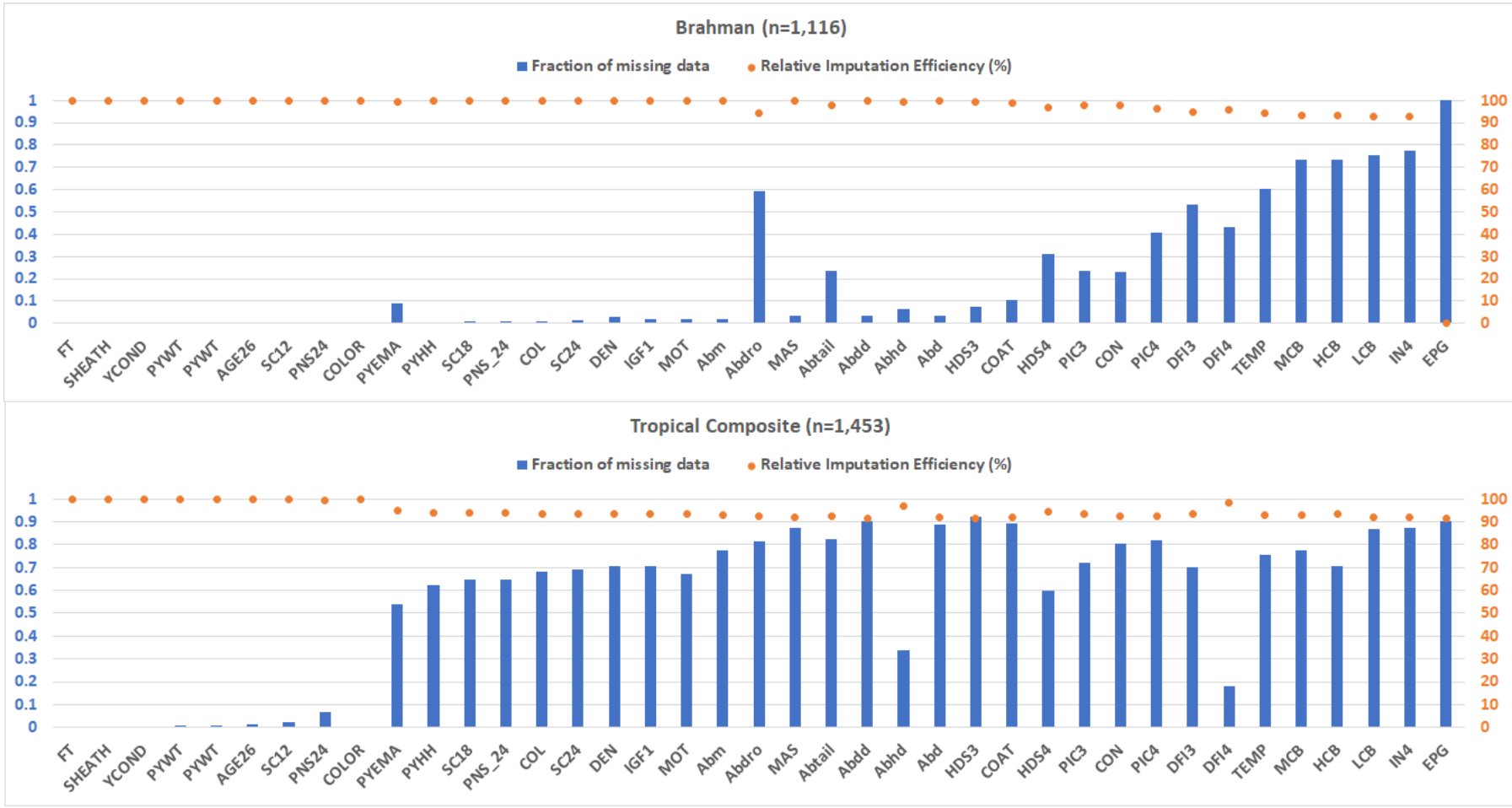


**Figure 2. Fraction of missing data and relative imputation efficiency on Beef CRC Brahman and Tropical Composite cattle (refer to table 2 and 3 for trait abbreviations).**

**Table 2. Summary statistics of phenotypes after imputation of missing data in Brahman (n=1,116).**

| Trait | Description | Mean | Std Dev | Min | Max |
|---|---|---|---|---|---|
| **FT** | Flight time | 164.47 | 33.66 | 70.67 | 281.39 |
| **TEMP** | Rectal temperature | 39.53 | 0.17 | 38.73 | 40.85 |
| **SHEATH** | Sheath score | 4.19 | 0.49 | 2.50 | 5.39 |
| **COLOR** | Coat color | 3.20 | 0.38 | 1.00 | 4.45 |
| **COAT** | Coat score | 5.17 | 1.26 | 1.73 | 8.31 |
| **YCOND** | Yearling condition score | 6.70 | 0.31 | 5.92 | 7.39 |
| **YWT** | Yearling weight | 308.72 | 38.67 | 180.00 | 430.00 |
| **PYHH** | Post-yearling hip height | 127.74 | 3.10 | 117.09 | 134.21 |
| **PYEMA** | Post-yearling eye muscle area | 46.16 | 5.72 | 27.00 | 57.34 |
| **PYWT** | Post-yearling weight | 308.72 | 30.25 | 204.01 | 376.02 |
| **IN4** | Inhibin 4 concentration | 6.79 | 1.26 | -0.65 | 10.95 |
| **IGF1** | IGF-I concentration | 541.52 | 297.41 | -91.13 | 1299.00 |
| **SC12** | Scrotal circumference at 12 mo | 21.24 | 1.66 | 16.50 | 25.94 |
| **SC18** | Scrotal circumference at 18 mo | 26.65 | 1.25 | 20.72 | 30.61 |
| **SC24** | Scrotal circumference at 24 mo | 29.66 | 1.64 | 20.95 | 33.06 |
| **AGE26** | Age at SC 26 cm | 556.34 | 54.60 | 440.96 | 902.34 |
| **DEN** | Semen density | 3.15 | 0.47 | 1.33 | 4.50 |
| **COL** | Semen color | 3.24 | 0.50 | 1.49 | 4.50 |
| **MOT** | Semen motility | 72.16 | 7.61 | 10.00 | 92.64 |
| **MAS** | Semen mass | 2.71 | 0.45 | 1.18 | 4.50 |
| **PNS24** | Percent normal sperm at 24 mo | 73.58 | 8.23 | 15.47 | 92.28 |
| **Abdd** | Sperm morphology defect 1 | 0.20 | 0.06 | 0.00 | 0.49 |
| **Abdrop** | Sperm morphology defect 2 | 0.37 | 0.07 | 0.21 | 0.65 |
| **Abhd** | Sperm morphology defect 3 | 0.41 | 0.04 | 0.22 | 0.86 |
| **Abmp** | Sperm morphology defect 4 | 0.42 | 0.05 | 0.26 | 0.72 |
| **Abd** | Sperm morphology defect 5 | 0.29 | 0.06 | 0.07 | 0.58 |
| **Abtail** | Sperm morphology defect 6 | 0.03 | 0.03 | -0.04 | 0.14 |
| **PIC3** | Sperm cytometer observation 1 | 6.46 | 0.13 | 5.89 | 6.63 |
| **DFI3** | Sperm cytometer observation 2 | 2.08 | 0.40 | 0.82 | 4.04 |
| **HDS3** | Sperm cytometer observation 3 | 2.82 | 0.96 | 1.29 | 7.17 |
| **PIC4** | Sperm cytometer observation 4 | 6.40 | 0.14 | 5.82 | 6.60 |
| **DFI4** | Sperm cytometer observation 5 | 2.59 | 0.37 | 1.23 | 4.53 |
| **HDS4** | Sperm cytometer observation 6 | 2.88 | 1.03 | 1.29 | 7.54 |
| **LCB** | Sperm cytometer observation 7 | 6.42 | 0.11 | 5.91 | 7.06 |
| **MCB** | Sperm cytometer observation 8 | 3.41 | 0.46 | 0.78 | 5.26 |
| **HCB** | Sperm cytometer observation 9 | 1.87 | 0.55 | -0.41 | 4.28 |

**Table 3. Summary statistics of phenotypes after imputation of missing data in Tropical Composite (n=1,453).**

| Trait | Description | Mean | Std Dev | Min | Max |
|---|---|---|---|---|---|
| **FT** | Flight time | 145.38 | 31.15 | 59.79 | 243.12 |
| **TEMP** | Rectal temperature | 39.57 | 0.31 | 37.46 | 41.42 |
| **EPG** | Worm eggs per gram | 7.35 | 1.35 | 1.55 | 15.22 |
| **SHEATH** | Sheath score | 7.00 | 0.79 | 1.34 | 9.15 |
| **COLOR** | Coat color | 3.84 | 0.30 | 2.00 | 4.60 |
| **COAT** | Coat score | 6.05 | 1.28 | 3.00 | 11.00 |
| **YCOND** | Yearling condition score | 6.76 | 0.61 | 5.14 | 9.50 |
| **YWT** | Yearling weight | 325.76 | 43.72 | 174.00 | 550.00 |
| **PYHH** | Post-yearling hip height | 124.75 | 3.79 | 110.11 | 135.00 |
| **PYEMA** | Post-yearling eye muscle area | 49.39 | 7.40 | 23.77 | 64.27 |
| **PYWT** | Post-yearling weight | 325.77 | 33.09 | 222.85 | 495.59 |
| **IN4** | Inhibin 4 concentration | 8.74 | 1.65 | 4.94 | 13.23 |
| **IGF1** | IGF-I concentration | 751.75 | 377.30 | 46.12 | 1690.00 |
| **SC12** | Scrotal circumference at 12 mo | 25.95 | 2.20 | 18.01 | 32.40 |
| **SC18** | Scrotal circumference at 18 mo | 29.76 | 1.53 | 23.89 | 33.75 |
| **SC24** | Scrotal circumference at 24 mo | 30.81 | 2.05 | 24.86 | 35.61 |
| **AGE26** | Age at SC 26 cm | 417.65 | 42.90 | 257.10 | 576.78 |
| **DEN** | Semen density | 3.31 | 0.27 | 2.46 | 4.00 |
| **COL** | Semen color | 3.50 | 0.36 | 2.44 | 4.43 |
| **MOT** | Semen motility | 71.19 | 6.42 | 29.41 | 95.00 |
| **MAS** | Semen mass | 2.77 | 0.32 | 1.66 | 4.00 |
| **PNS24** | Percent normal sperm at 24 mo | 0.73 | 0.06 | 0.43 | 0.90 |
| **Abdd** | Sperm morphology defect 1 | 0.19 | 0.06 | -0.04 | 0.49 |
| **Abdrop** | Sperm morphology defect 2 | 0.35 | 0.05 | 0.00 | 0.59 |
| **Abhd** | Sperm morphology defect 3 | 0.41 | 0.05 | 0.21 | 0.61 |
| **Abmp** | Sperm morphology defect 4 | 0.42 | 0.04 | 0.27 | 0.63 |
| **Abd** | Sperm morphology defect 5 | 0.27 | 0.05 | 0.00 | 0.51 |
| **Abtail** | Sperm morphology defect 6 | 0.01 | 0.01 | -0.01 | 0.11 |
| **PIC3** | Sperm cytometer observation 1 | 6.51 | 0.06 | 6.30 | 6.71 |
| **DFI3** | Sperm cytometer observation 2 | 2.07 | 0.45 | 0.55 | 3.94 |
| **HDS3** | Sperm cytometer observation 3 | 2.59 | 0.32 | 1.38 | 4.09 |
| **PIC4** | Sperm cytometer observation 4 | 6.47 | 0.13 | 5.92 | 6.90 |
| **DFI4** | Sperm cytometer observation 5 | 2.71 | 0.49 | 1.13 | 5.45 |
| **HDS4** | Sperm cytometer observation 6 | 2.60 | 0.31 | 1.45 | 4.11 |
| **LCB** | Sperm cytometer observation 7 | 6.46 | 0.07 | 6.21 | 6.62 |
| **MCB** | Sperm cytometer observation 8 | 3.35 | 0.38 | 2.23 | 4.53 |
| **HCB** | Sperm cytometer observation 9 | 2.31 | 0.26 | 1.07 | 4.05 |

### 4.1.2 Evaluation of the estimated genomic breeding values using the imputed phenotype dataset.

The amount of missing data was variable across the different breeds and traits. To characterise the potential effect of imputing the phenotypic data on the accuracy of GEBV, we selected ten traits representing the range of variation in terms of missing data in the Tropical Composite bulls (n=1,449) from the Beef CRC. We selected three traits with complete dataset or missing very few observations (YWT, COAR, SHEATH), four traits missing around 1/3 of observations (PNS24, IN4 and SC12; n~1,000), and three traits missing around 2/3 of the observations (PPIC3, EPG, HDS3; n~500).

Not surprisingly, the traits in which very few (if any) observation was missing, were not affected by the imputation of phenotypes (Table 4). On the other hand, when a significant amount of observations was missing, the imputation generated positive outcomes in terms of GEBV accuracies. However, heritability estimates were reduced when using the imputed dataset, this can be partly explained by the limited number of records for some of the traits and the fact that the imputation procedure does not add variation. Therefore, heritability estimates were affected.

Similarly, to the heritability estimates, the accuracies of GEBVs were not affected when only few records were missing (YWT, COAT, SHEATH). However, when a larger amount of data was missing, it was possible to estimat breeding values with higher accuracy when using imputed phenotypes, as compared to using only the observed dataset (Figure 3 and Table 4). The estimates were largely unbiased and without over- or under-dispersion. Possible exceptions to these positive outcomes were: IGF1 using the real observations the estimates appeared slightly biased, PIC3 both analyses were slightly over-dispersed, and HDS3 using the real observations the estimates appeared to be under-dispersed. Bias and dispersion values that are not distant from zero (Table 4) are desirable.

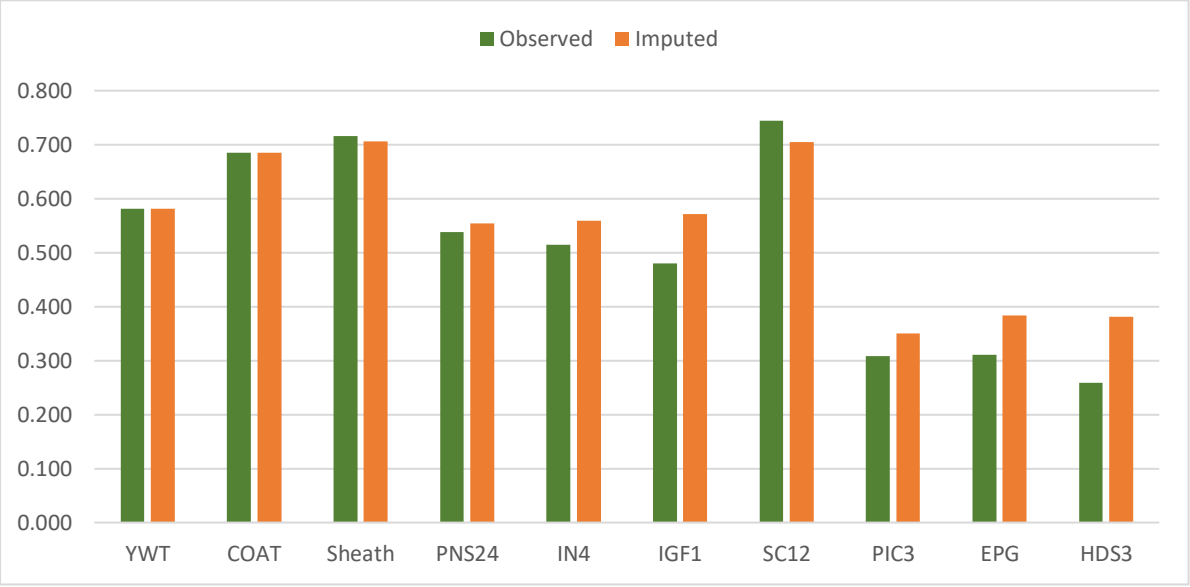


**Figure 3. Accuracy of genomic estimated breeding values (GEBV) via Method LR for ten traits that varied in terms of missing phenotypic data. YWT, COAT and Sheath had minimal, if any missing records; IN4, IGF1 and SC12 had around 1/3 of records missing, and PIC3, EPG and HDS3 had around 2/3 of the data missing.**

**Table 4. Mean heritability and GEBV accuracy (Traditional - correlation-based, and Method LD) for the five-way cross-validation using the observed dataset (observed) that had missing observations and the complete dataset (with the observed plus the imputed records).**

| Missingness | Trait | Heritability | Traditional | Method LR | | |
|---|---|---|---|---|---|---|
| | | | ACC | ACC | Bias | Dispersion |
| very few | YWT_observed | 0.507 | 0.516 | 0.582 | -0.092 | -0.025 |
| | YWT_imputed | 0.507 | 0.516 | 0.582 | -0.091 | -0.025 |
| | COAT_observed | 0.624 | 0.462 | 0.685 | 0.002 | 0.008 |
| | COAT_imputed | 0.624 | 0.462 | 0.685 | 0.002 | 0.008 |
| | SHEATH_observed | 0.745 | 0.601 | 0.716 | 0.005 | -0.108 |
| | SHEATH_imputed | 0.770 | 0.593 | 0.707 | 0.006 | -0.086 |
| about 1/3 | PNS24_observed | 0.421 | 0.400 | 0.538 | 0.000 | 0.020 |
| | PNS24_imputed | 0.339 | 0.480 | 0.555 | 0.000 | -0.037 |
| | IN4_observed | 0.514 | 0.467 | 0.515 | 0.001 | 0.001 |
| | IN4_imputed | 0.484 | 0.500 | 0.560 | 0.000 | 0.021 |
| | IGF1_observed | 0.492 | 0.385 | 0.481 | 0.526 | -0.019 |
| | IGF1_imputed | 0.402 | 0.470 | 0.572 | -0.038 | -0.003 |
| | SC12_observed | 0.565 | 0.524 | 0.744 | -0.009 | -0.005 |
| | SC12_imputed | 0.456 | 0.591 | 0.705 | -0.010 | 0.031 |
| about 2/3 | PIC3_observed | 0.426 | 0.117 | 0.309 | 0.000 | 0.145 |
| | PIC3_imputed | 0.209 | 0.212 | 0.350 | 0.000 | 0.125 |
| | EPG_observed | 0.520 | 0.189 | 0.311 | -0.002 | 0.080 |
| | EPG_imputed | 0.384 | 0.300 | 0.384 | -0.002 | 0.082 |
| | HDS3_observed | 0.064 | 0.400 | 0.259 | -0.001 | -0.286 |
| | HDS3_imputed | 0.159 | 0.271 | 0.381 | -0.002 | -0.045 |

The use of phenotype imputation for the calculation of GEBV is promising. The SAS procedure multiple imputation (MI) generated values with high (>90%) imputation efficiency that can safely replace missing values on a cattle dataset. It is worth noting that since the MI procedure uses correlation between vectors to inform the imputation, the ideal dataset has several observations on each animal, with occasional missing records. The amount of missing data is a key factor to be consider when using this phenotype imputation approach. It is also important to consider the relatedness between the animals that have imputed phenotypes. Since MI does not use any genetic information, the imputed phenotypes can be used for further genetic analyses (i.e. estimates of GEBVs). However, further research is needed to better evaluate the use of phenotype imputation in a variety of traits, e.g. with high and low environmental impact, and within and across breeds.

## 4.2 The bull fertility dataset: expanded and updated

### 4.2.1 Number of samples collected and their origins

The combined dataset with samples from the Beef CRC and the collaborating herds has nearly 10,000 BBSE records; of those, 6,202 animals have a paired genotype profile. The Beef CRC animals were born in the early 2000's while most of the cattle from collaborating herds were born after 2015 (Figure 4). This impressive genomic resource gives, for the first time, the opportunity to explore the genetic controls for bull fertility traits in tropical beef cattle in Australia and sets the basis of multibreed genomic predictions.


**Figure 4. Distribution of records on full bull breeding soundness examination (BBSE) per year of birth and breed type. Top panel – all animals with records, bottom panel – genotyped animals with records.**

### 4.2.2 Imputation up to HD and whole-genome sequence

We applied the same genotype imputation pipeline for the Beef CRC and the newly assembled dataset. This pipeline has two main stages, first the imputation from "native" genotypes, in our case around 50,000 SNP markers, up to around 700,000 SNP, and a second stage where the data generated on the first stage is imputed up to genome sequence variants ~25M SNP. Table 5 shows the number of animals per breed used in each stage for the imputation.

**Table 5. Number of animals in the newly collected dataset (assayed 50K), and those used as a reference panel for the two steps of genotype imputation to HD and to Sequence.**

| Breed | Beef CRC* assayed 50K SNP | BBSE** assayed 50K SNP | Reference HD 700K SNP | Reference Sequence 25M SNP |
|---|---|---|---|---|
| **Afrikander** | | | | 5 |
| **Angus** | | | 195 | 50 |
| **Angus Red** | | | | 30 |
| **Beefmaster** | | | | 16 |
| **Belmont Tropical Composite** | 764 | 664 | 130 | |
| **Bonsmara** | | | 32 | |
| **Boran** | | | 24 | 21 |
| **Brahman** | 5,040 | | 863 | 200 |
| **Brangus** | | | | 5 |
| **Charolais** | | | | 50 |
| **Composite** | 999 | | 12 | |
| **Droughtmaster** | 464 | 764 | 345 | 37 |
| **Gir** | | | | 7 |
| **Hereford** | | | | 50 |
| **Limousin** | | | | 50 |
| **Murray Grey** | | | | 2 |
| **Nelore** | | | | 12 |
| **Santa Gertrudis** | 1,566 | 952 | 467 | 28 |
| **Senepol** | | | | 12 |
| **Shaiwal** | | | | 2 |
| **Shorthorn** | | | | 33 |
| **Tropical Composite** | 3,744 | | 351 | 56 |
| **Tuli** | | | 33 | 2 |
| **Ultra-black** | | 860 | | |
| **total** | **12,577** | **3,240** | **2,452** | **668** |

* Beef CRC dataset includes bulls and cows.

**BBSE datasets of bulls assembled by this project.

The imputation $R^2$ is a metric used to evaluate the reliability of the imputed data, and often a threshold on this metric is applied before the imputed data is used on further analyses. Applying the conservative threshold of imputation $R^2 > 0.8$, resulted in 704,852 and 22M SNP for the first and second rounds of imputation using the CRC data, and 721,986 and approximately 16M SNP for the newly acquired dataset.

The Principal Components Analysis (PCA) was used to visualize the genetic divergence between individuals. In a PCA plot, similar animals cluster together. The PCA based on the imputed HD data showed clear clusters that agree with the breed designations (Figure 5), giving confidence that the genotype imputation worked as expected.

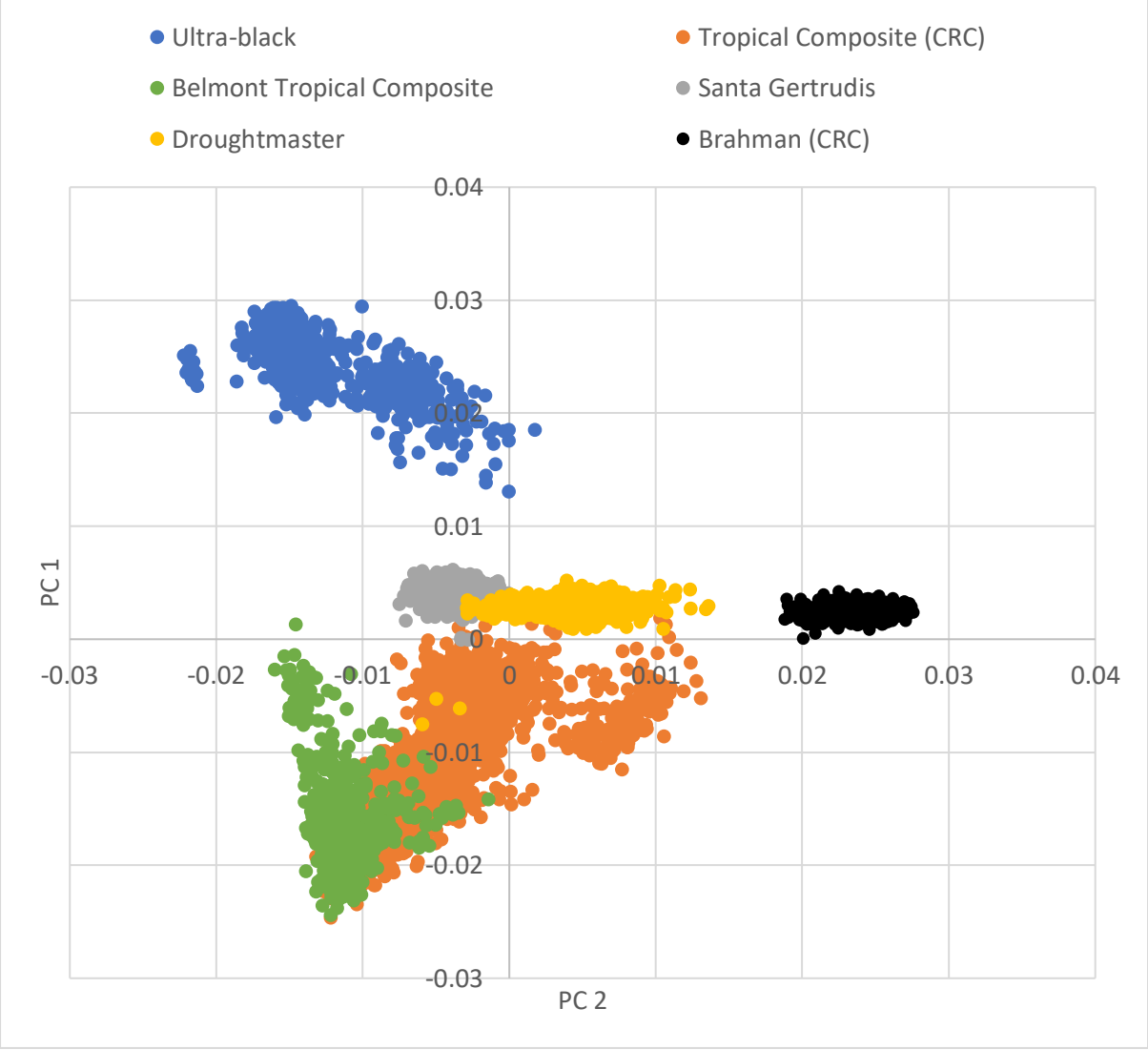


**Figure 5. Principal components analysis on imputed genotypes of tropical cattle from Australia** (680,758 SNP in 6,203 cattle used for analyses reported on 4.3 Multibreed genomic predictions for bull fertility traits).

## 4.3 Multibreed genomic predictions for bull fertility traits

Our aim was to assemble a large reference population of tropically adapted bulls with measurements for ten fertility traits and genotyped at high-density to generate multibreed GEBVs with useful accuracies. We also explored the factors affecting accuracy and bias of GEBVs. This multibreed reference population was comprised of 6,063 bulls across six breeds (Table 6) and ranging from 660 bulls (or 11% of the total, for BTC - Belmont Tropical Composite) to 1,819 (or 30%of the total, for TRC - Beef CRC Tropical Composite). The number of records per phenotype varied, ranging from 5,588 (92%) for MASS (mass movement of sperm in ejaculate) to 6,060 (100%) for sheath score.

**Table 6. Number of records by breed and trait.**

| Breed | Total | Trait | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | WT | COND | SC | SHEATH | DENS | MASS | MOT | PNS | PD | MP |
| BRM | 1,051 | 1,051 | 1,051 | 1,051 | 1,051 | 1,051 | 1,051 | 1,051 | 1,051 | 1,051 | 1,051 |
| TRC | 1,819 | 1,819 | 1,819 | 1,819 | 1,819 | 1,818 | 1,817 | 1,818 | 1,819 | 1,819 | 1,819 |
| SGT | 929 | 929 | 928 | 918 | 928 | 901 | 901 | 901 | 895 | 895 | 895 |
| DMT | 760 | 750 | 617 | 601 | 760 | 581 | 581 | 581 | 710 | 709 | 710 |
| UBK | 844 | 454 | 842 | 837 | 842 | 785 | 785 | 785 | 783 | 781 | 781 |
| BTC | 660 | 655 | 660 | 653 | 660 | 453 | 453 | 453 | 450 | 450 | 450 |
| Total | 6,063 | 5,658 | 5,917 | 5,879 | 6,060 | 5,589 | 5,588 | 5,589 | 5,708 | 5,705 | 5,706 |

***** WT – weight (Kg), COND – body condition score (1-5), SC – scrotal circumference (cm), SHEATH – Sheath score (1-5), DENS – Density of ejaculate (1-5), MASS – mass movement of sperm in ejaculate (1-5), MOT – sperm progressive motility (%), PNS – percentage of normal sperm (%), PD – proximal cytoplasmic droplets (%), MP – middle piece abnormalities (%).

** BRM – CRC Brahman, TRC – CRC Tropical Composite, SGT – Santa Gertrudis, DMT – Droughtmaster, UBK – Ultra-black, BTC – Belmont Tropical Composite.

One key observation was the large difference in age at measurement across breeds (Table 7). On average, oldest and youngest measurements were from BTC animals (Belmont Tropical Composite) measured at ~395 days, and SGT animals (Santa Gertrudis) measured at ~625 days. Importantly, within breed there was minimal variation of ages at measurement across the ten traits.

As expected, the across-breed variation in age at measurement is likely to impact on the observed measurements (Table 8). For instance, the younger BTC (Belmont Tropical Composite) weighing on average 284 kg, compared to the older SGT (Santa Gertrudis) weighing on average 507 Kg. Thus, the SGT being 223 kg heavier and 233 days older than BTC. This corresponds to a weight gain of 0.957 kg/d which is probably a ballpark figure for most bulls in the 12 to 20 months old period. Nevertheless, this also highlights the importance of fitting the covariate of age (in addition to contemporary group which included breed) in the analytical model for all the ten traits.

**Table 7. Average (SD) age at measurement in days by breed and trait.**

| Breed | WT | COND | SC | SHEATH | DENS | MASS | MOT | PNS | PD | MP |
|---|---|---|---|---|---|---|---|---|---|---|
| BRM | 580.2 (95.9) | 580.5 (95.6) | 580.2 (96.2) | 580.2 (96.2) | 582.9 (95.3) | 582.9 (95.3) | 582.9 (95.3) | 582.9 (95.3) | 582.9 (95.3) | 582.9 (95.3) |
| TRC | 489.5 (99.9) | 489.4 (99.6) | 488.5 (99.6) | 488.4 (99.6) | 491.5 (99.6) | 491.6 (99.6) | 491.7 (99.6) | 491.6 (99.6) | 491.6 (99.6) | 491.5 (99.6) |
| SGT | 626.1 (34.5) | 626.1 (34.5) | 625.9 (34.5) | 626.1 (34.5) | 625.8 (34.6) | 625.8 (34.6) | 625.8 (34.6) | 625.9 (34.5) | 625.9 (34.5) | 625.9 (34.5) |
| DMT | 586.8 (29.1) | 601.8 (25.5) | 601.7 (25.6) | 600.9 (24.9) | 601.8 (25.5) | 601.7 (25.5) | 601.8 (25.5) | 601.8 (25.6) | 601.0 (24.9) | 601.8 (25.5) |
| UBK | 452.2 (75.7) | 438.2 (75.2) | 438.4 (75.3) | 438.2 (75.2) | 443.1 (74.0) | 456.4 (74.2) | 443.1 (74.0) | 443.2 (74.3) | 443.3 (74.3) | 443.2 (74.2) |
| BTC | 393.5 (25.2) | 392.2 (30.1) | 392.1 (30.3) | 392.2 (30.1) | 397.1 (29.2) | 399.1 (20.9) | 397.1 (29.2) | 397.0 (29.3) | 397.0 (29.3) | 397.0 (29.3) |

***** WT – weight (Kg), COND – body condition score (1-5), SC – scrotal circumference (cm), SHEATH – Sheath score (1-5), DENS – Density of ejaculate (1-5), MASS – mass movement of sperm in ejaculate (1-5), MOT – sperm progressive motility (%), PNS – percentage of normal sperm (%), PD – proximal cytoplasmic droplets (%), MP – middle piece abnormalities (%).

** BRM – CRC Brahman, TRC – CRC Tropical Composite, SGT – Santa Gertrudis, DMT – Droughtmaster, UBK – Ultra-black, BTC – Belmont Tropical Composite.

**Table 8. Average (SD) measurement by breed and phenotype*.**

| Breed | WT | COND | SC | SHEATH | DENS | MASS | MOT | PNS | PD | MP |
|---|---|---|---|---|---|---|---|---|---|---|
| BRM | 359.05 (42.9) | 3.09 (0.37) | 27.92 (2.80) | 5.88 (1.11) | 2.64 (0.95) | 2.09 (1.07) | 60.22 (23.66) | 51.57 (30.24) | 24.50 (27.16) | 13.40 (11.25) |
| TRC | 328.50 (58.7) | 2.81 (0.44) | 28.89 (2.89) | 3.13 (1.77) | 2.80 (0.94) | 2.35 (1.07) | 64.71 (22.11) | 57.63 (27.23) | 14.51 (19.79) | 14.34 (12.47) |
| SGT | 507.28 (79.2) | 3.03 (0.28) | 34.46 (3.10) | 2.94 (0.78) | 2.67 (0.94) | 2.25 (0.93) | 62.83 (21.50) | 73.35 (21.43) | 6.88 (12.50) | 7.85 (8.28) |
| DMT | 459.53 (58.0) | 3.05 (0.29) | 33.65 (3.16) | 3.14 (0.68) | 2.39 (0.91) | 2.15 (0.93) | 69.59 (20.85) | 65.69 (25.16) | 9.88 (16.36) | 8.93 (9.67) |
| UBK | 439.16 (65.7) | 3.08 (0.23) | 33.80 (3.37) | 1.78 (0.80) | 2.58 (0.78) | 2.56 (0.88) | 73.58 (19.00) | 67.66 (26.57) | 10.20 (15.96) | 7.83 (8.24) |
| BTC | 283.68 (54.2) | 2.83 (0.25) | 27.27 (3.77) | 1.64 (0.58) | 2.14 (0.79) | 2.03 (0.90) | 66.04 (24.61) | 56.14 (28.07) | 13.80 (18.35) | 13.66 (11.96) |

Multibreed estimates of genomic heritabilities, genetic and residual correlations are given in Table 9. Using the genomic relationship matrix (GRM) across the 6,063 bulls, a total of 45 bi-variate analyses were performed for as many pair-wise trait combinations. Each trait was included in nine analyses (one with each of the remaining traits) and the heritabilities listed in Table 9 correspond to the average heritability estimated across the nine analyses. These reported estimates are well within those published in the literature for the same traits and, on occasions using a subset of this population (e.g. the Beef CRC cattle). For instance, using a population of Beef CRC Brahman and Tropical Composite bulls and cows, Porto-Neto et al (2014) reported a heritability estimate for sheath score of 0.51 and 0.57 for Brahman and Tropical Composite, respectively, and very similar to the 0.572 reported here. Similarly, and more recently reported, for the semen traits, Fortes et al. (2020) reported a heritability estimate for PNS (percent normal sperm) of 0.35 and 0.29 for Brahman and Tropical Composite bulls, respectively, slightly higher than the 0.244 found here. The positive correlations (both genetic and residual) between weight, condition score and scrotal circumference have been reported in the past. However, among the semen traits, the positive genetic correlations observed for density, mass, motility and percent normal sperm have, to the best of our knowledge (though perhaps not surprisingly), not been reported in the past. Again, the fact that these semen traits are genetically negatively correlated with the semen defect traits of PD (proximal cytoplasmic droplets) and MP (mid piece abnormalities) are both novel and encouraging. Importantly, PD and MP were uncorrelated with each other (genetic and residual correlation of 0.06 and 0.01, respectively) indicating that these two sperm defects are independent from each other. Taken together, these genetic parameter estimates offer hope for the possibility of successfully implementing genetic selection programs for these bull fertility traits.

**Table 9. Multibreed estimates of heritability (bold, diagonal), genetic (above diagonal) and residual correlations (below diagonal) from bi-variate analyses*.** A total of 45 bi-variate analyses were performed for as many pair-wise trait combinations and each using the "big" across-breed GRM. Heritabilities are the average across the 9 bi-variates in which a trait was included.

| | WT | COND | SC | SHEATH | DENS | MASS | MOT | PNS | PD | MP |
|---|---|---|---|---|---|---|---|---|---|---|
| WT | ***0.369*** | 0.311 | 0.431 | 0.071 | 0.081 | 0.084 | 0.034 | -0.083 | 0.000 | 0.021 |
| COND | 0.339 | ***0.174*** | 0.114 | -0.135 | 0.067 | 0.128 | -0.001 | 0.088 | -0.156 | 0.048 |
| SC | 0.439 | 0.147 | ***0.465*** | 0.238 | 0.168 | 0.143 | 0.058 | 0.047 | -0.043 | 0.048 |
| SHEATH | 0.034 | 0.002 | 0.039 | ***0.572*** | -0.058 | -0.072 | -0.151 | -0.259 | 0.273 | 0.121 |
| DENS | 0.062 | -0.002 | 0.173 | 0.011 | ***0.202*** | 0.725 | 0.340 | 0.293 | -0.178 | -0.180 |
| MASS | 0.077 | 0.009 | 0.151 | 0.025 | 0.683 | ***0.208*** | 0.681 | 0.525 | -0.324 | -0.304 |
| MOT | 0.049 | -0.014 | 0.113 | 0.059 | 0.340 | 0.635 | ***0.192*** | 0.527 | -0.216 | -0.398 |
| PNS | 0.111 | 0.067 | 0.201 | 0.046 | 0.263 | 0.355 | 0.353 | ***0.244*** | -0.697 | -0.512 |
| PD | -0.078 | -0.063 | -0.174 | -0.049 | -0.209 | -0.218 | -0.156 | -0.692 | ***0.224*** | 0.060 |
| MP | -0.040 | -0.002 | -0.079 | -0.022 | -0.078 | -0.163 | -0.220 | -0.442 | 0.010 | ***0.225*** |

* WT – weight (Kg), COND – body condition score (1-5), SC – scrotal circumference (cm), SHEATH – Sheath score (1-6), DENS – Density of ejaculate (1-5), MASS – mass movement of sperm in ejaculate (1-5), MOT – sperm progressive motility (%), PNS – percentage of normal sperm (%), PD – proximal cytoplasmic droplets (%), MP – middle piece abnormalities (%).

Table 10 lists the estimates of genomic correlations resulting from the breed pair-wise bi-variate analyses of ten traits when treating the same trait as a different phenotype in each breed. Hence, 150 bi-variate analyses from 15 breed-pairs times the ten traits are presented in Table 10. With the exception of SC (scrotal circumference) and sheath score for which the estimated genomic correlations were moderately positive (i.e. averaged across the 15 breed pairs was 0.363 for SC and 0.514 for sheath), all other estimated genomic correlations were near zero.

**Table 10. Estimates of genomic correlation within phenotype* across all pair wise breeds**.**

| Breed 1 | Breed 2 | WT | COND | SC | SHEATH | DENS | MASS | MOT | PNS | PD | MP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BRM | TRC | 0.729 | 0.048 | 0.489 | 0.673 | -0.012 | -0.062 | -0.042 | -0.006 | -0.001 | 0.031 |
| BRM | SGT | 0.006 | 0.011 | 0.620 | 0.940 | 0.027 | 0.005 | 0.006 | 0.014 | 0.003 | 0.009 |
| BRM | DMT | 0.020 | 0.005 | 0.433 | 0.135 | 0.057 | 0.012 | 0.016 | 0.025 | 0.026 | 0.021 |
| BRM | UBK | -0.001 | 0.018 | 0.370 | 0.360 | 0.018 | -0.060 | -0.082 | -0.008 | 0.007 | 0.004 |
| BRM | BTC | 0.044 | 0.039 | 0.030 | 0.141 | 0.000 | -0.002 | 0.003 | 0.003 | 0.003 | -0.001 |
| TRC | SGT | 0.025 | -0.030 | 0.326 | 0.599 | 0.100 | -0.012 | 0.025 | 0.011 | 0.026 | 0.057 |
| TRC | DMT | 0.012 | 0.009 | 0.538 | 0.019 | -0.003 | -0.079 | -0.005 | -0.009 | 0.008 | 0.010 |
| TRC | UBK | 0.017 | 0.135 | 0.486 | 0.804 | 0.079 | 0.073 | 0.013 | 0.028 | 0.029 | 0.018 |
| TRC | BTC | 0.683 | 0.066 | 0.109 | 0.630 | -0.041 | -0.044 | 0.045 | 0.006 | 0.003 | 0.014 |
| SGT | DMT | 0.011 | -0.163 | 0.750 | 0.800 | 0.004 | -0.005 | 0.004 | 0.003 | 0.020 | 0.012 |
| SGT | UBK | 0.006 | 0.003 | 0.625 | 0.568 | 0.000 | 0.018 | 0.005 | -0.002 | 0.016 | 0.001 |
| SGT | BTC | 0.008 | 0.074 | 0.055 | 0.938 | -0.002 | 0.004 | -0.005 | 0.021 | 0.006 | 0.002 |
| DMT | UBK | 0.023 | -0.004 | 0.283 | 0.040 | -0.003 | -0.011 | -0.009 | 0.007 | 0.005 | 0.006 |
| DMT | BTC | 0.019 | 0.004 | 0.056 | 0.676 | -0.002 | 0.004 | 0.004 | 0.009 | 0.007 | -0.004 |
| UBK | BTC | -0.005 | 0.040 | 0.274 | 0.387 | 0.000 | 0.000 | 0.006 | 0.014 | 0.003 | 0.015 |
| | Average | 0.106 | 0.017 | 0.363 | 0.514 | 0.015 | -0.011 | -0.001 | 0.008 | 0.011 | 0.013 |

***** WT – weight (Kg), COND – body condition score (1-5), SC – scrotal circumference (cm), SHEATH – Sheath score (1-5), DENS – Density of ejaculate (1-5), MASS – mass movement of sperm in ejaculate (1-5), MOT – sperm progressive motility (%), PNS – percentage of normal sperm (%), PD – proximal cytoplasmic droplets (%), MP – middle piece abnormalities (%).

** BRM – CRC Brahman, TRC – CRC Tropical Composite, SGT – Santa Gertrudis, DMT – Droughtmaster, UBK – Ultra-black, BTC – Belmont Tropical Composite.

We did not observe breed differences for the average GEBV within a trait as they were all non-significantly different from zero. However, there were some differences in the GEBV variation across breeds and these differences could reflect different accuracies with higher variations associated with higher accuracies (Table 11).

**Table 11. Standard deviation of GEBV for the ten bull fertility traits and by breed.**

| Breed | WT | COND | SC | SHEATH | DENS | MASS | MOT | PNS | PD | MP |
|---|---|---|---|---|---|---|---|---|---|---|
| BRM | 11.46 | 0.053 | 1.174 | 0.492 | 0.151 | 0.175 | 3.674 | 6.676 | 5.808 | 2.328 |
| TRC | 14.75 | 0.072 | 1.395 | 0.947 | 0.181 | 0.232 | 4.896 | 7.559 | 5.139 | 2.986 |
| SGT | 15.42 | 0.045 | 1.253 | 0.455 | 0.198 | 0.183 | 4.065 | 5.534 | 3.473 | 2.097 |
| DMT | 22.42 | 0.067 | 1.469 | 0.426 | 0.177 | 0.179 | 4.318 | 7.145 | 4.181 | 2.569 |
| UBK | 13.82 | 0.048 | 1.476 | 0.564 | 0.162 | 0.199 | 4.377 | 6.689 | 3.939 | 2.080 |
| BTC | 15.32 | 0.046 | 1.442 | 0.368 | 0.145 | 0.181 | 4.914 | 7.340 | 4.602 | 2.888 |

**Assessment of the accuracies of multibreed GEBV for ten bull fertility traits.**

To evaluate the accuracies of the GEBV, we used two validation schemes, 1) a whole population was removed from the reference population, and their GEBV calculated, and 2) a five-way cross validation where 20% of animals in each breed is removed from the reference population and their GEBV calculated. The estimates were largely unbiased, and not under- or over-dispersed (Table 12). For scheme 1, the average accuracies were around 20-25% for most traits, while for scheme 2 the accuracies ranged around 30-35%. Different from most traits, SC and SHEATH had higher accuracies in both schemes ~40-44% and 49-58% (Figure 6 and Table 12). The comparison between the validation schemes 1 and 2 demonstrated the importance of a breed being represented in the reference population when estimating their multibreed GEBV. For all traits and breeds more accurate GEBV were obtained when all breeds were represented in the reference population.

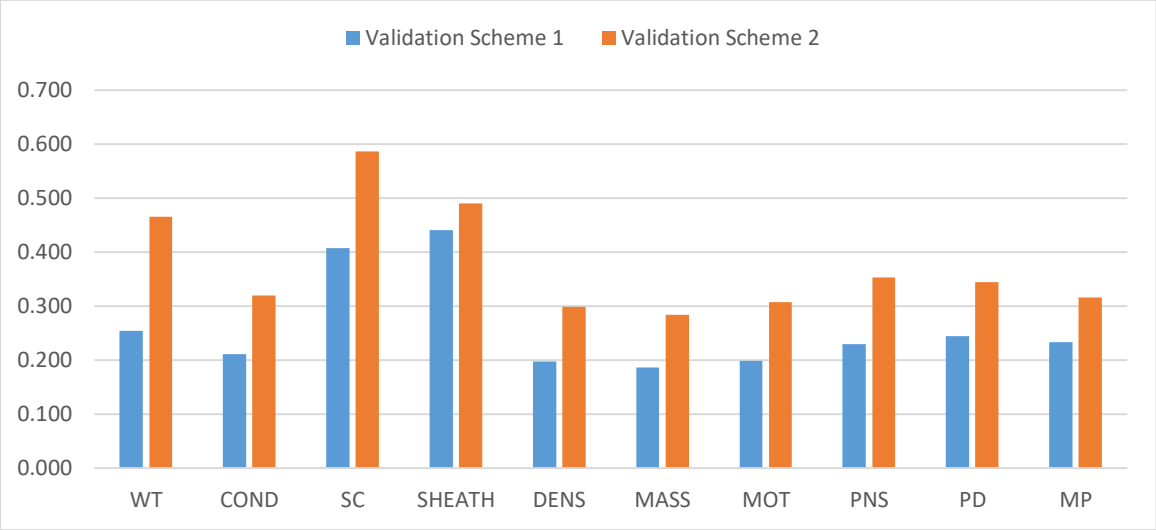


**Figure 6. Accuracy of GEBV (Method LR).** Validation Scheme 1: From a given validation breed, all measures set as missing in the reference population. Validation Scheme 2: From a given validation breed, a random 20% of measures missing in the reference (and then averaged across the five 80/20 cross-validation splits). The advantage of being represented in the reference population while estimating breeding values in a multibreed approach is clear.

**Table 12. Summary of accuracy estimates. Correlation-based accuracy ($ACC_R$) and method LR accuracy ($ACC_{LR}$), bias and dispersion (Disp.) of GEBV across the 10 phenotypes averaged across the six breeds and by two validation schemes.**
Validation Scheme #1: From a given validation breed, all measures set as missing in the reference population.
Validation Scheme #2: From a given validation breed, a random 20% of measures missing in the reference (and then averaged across the five 80/20 cross-validation splits).

| Phenotype* | Validation Scheme #1 | | | | Validation Scheme #2 | | | |
|---|---|---|---|---|---|---|---|---|
| | $ACC_R$ | $ACC_{LR}$ | Bias | Disp. | $ACC_R$ | $ACC_{LR}$ | Bias | Disp. |
| WT | 0.163 | 0.254 | -0.489 | 0.280 | 0.445 | 0.465 | 0.037 | 0.026 |
| COND | 0.120 | 0.211 | -0.001 | 0.162 | 0.189 | 0.320 | 0.000 | 0.135 |
| SC | 0.357 | 0.407 | -0.026 | 0.167 | 0.590 | 0.586 | 0.012 | 0.013 |
| SHEATH | 0.439 | 0.441 | 0.015 | 0.093 | 0.549 | 0.490 | 0.003 | 0.137 |
| DENS | 0.034 | 0.197 | 0.004 | 0.253 | 0.097 | 0.299 | 0.001 | 0.150 |
| MASS | -0.004 | 0.187 | 0.000 | 0.265 | 0.073 | 0.284 | 0.000 | 0.205 |
| MOT | 0.057 | 0.199 | 0.047 | 0.236 | 0.171 | 0.308 | 0.014 | 0.192 |
| PNS | 0.107 | 0.229 | 0.150 | 0.299 | 0.341 | 0.353 | 0.006 | 0.248 |
| PD | 0.203 | 0.244 | -0.012 | 0.152 | 0.323 | 0.344 | 0.015 | 0.230 |
| MP | 0.192 | 0.234 | 0.061 | 0.216 | 0.291 | 0.316 | 0.001 | 0.300 |

***** WT – weight (Kg), COND – body condition score (1-5), SC – scrotal circumference (cm), SHEATH – Sheath score (1-5), DENS – Density of ejaculate (1-5), MASS – mass movement of sperm in ejaculate (1-5), MOT – sperm progressive motility (%), PNS – percentage of normal sperm (%), PD – proximal cytoplasmic droplets (%), MP – middle piece abnormalities (%).

The increased accuracy observed in validation scheme 2 was not of the same magnitude across all breeds and traits (Figure 7). Across the board, scheme 2 showed 10% higher accuracy. For TRC and BTC the improvement was even better: ~15%. This information, combined with the comparison between validation schemes, (above Figure 6 and Table12) confirms the importance of a breed being represented in the reference population when estimating breeding values.

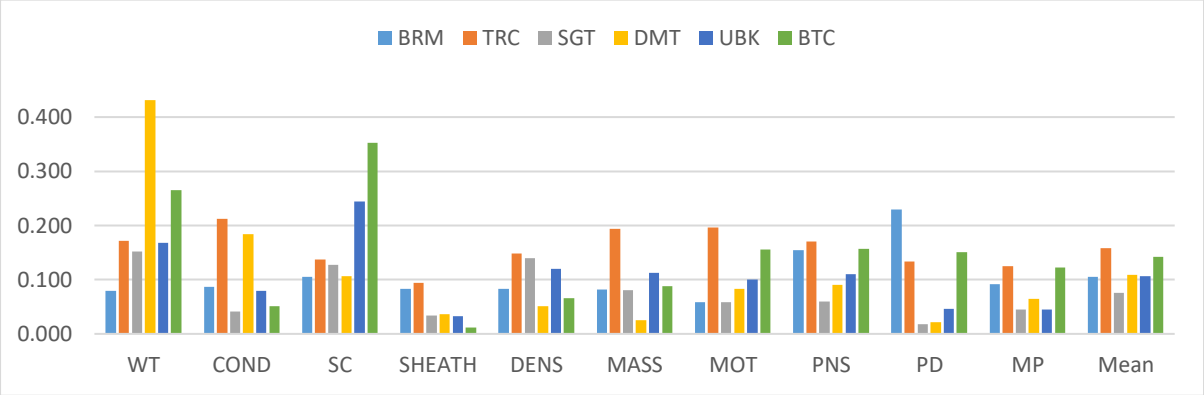


**Figure 7. Accuracy gain when the breed is part of the reference population for the estimation of GEBV (Validation Scheme 2 minus Validation Scheme 1)**.

## 4.4 Genome-wide association analyses

There are several different ways for undertaking genome-wide association analyses that aim at identifying genomic regions of interest. Here we used three approaches, a) case-control analyses for some specific conditions, e.g. spiral deviation of penis, b) GREML fitting two genomic relationship matrices, and c) the estimated SNP effect used in a pleotropic test.

All case-control analyses for specific conditions suggested that those were influenced by several genes (polygenic architecture). Using this approach, we assessed the spiral deviation of the penis, retention of the frenulum, testicular hypoplasia, and swolleness of the hocks. Unfortunately, none of these conditions had a single genomic region associated with the trait or a major region that explained a large proportion of the variance to be translated into a genomic test. Although these results supported the case for these conditions being polygenic, further analyses should be conducted trying to define a reduced number of markers that could be developed as diagnostic.

We ran within-breed genome-wide associations for ten BBSE traits (60 analyses: six breeds x ten traits). The analyses confirmed most of the observed traits' polygenic nature and highlighted some breed similarities and differences. Not surprisingly, WT, COND, DENS, MASS, MOT, PNS, PD, and MP were confirmed as polygenic in all breeds, although some more significant regions exist for some trait-breed combinations.

The significant associations for SHEATH are mostly confined to a broad region on chromosome 5 in all breeds. This genomic region had previously been identified (Porto-Neto et al., 2014), and a potential causative mutation pointed out in a later study (Aguiar et al., 2018). At this point, we cannot confirm whether the same causative mutation is segregating in our population; further analyses is needed to test this hypothesis.

The scrotal circumference (SC) is also a polygenic trait, however, it has some more relevant chromosomes. Specific regions of chromosomes 5 and X were significantly associated with SC in the majority of the six breeds (**Error! Reference source not found.**). These two chromosomes (5 and X) had significant associations with several traits, including PNS in TRC, and MP in BRM and DMT.

To prioritize genomic regions relevant to these ten traits, we applied a test designed to identify pleiotropic effects of molecular markers (Bolormaa et al., 2014). The input for this analysis was the back-solved SNP effects for each trait. Figure 9 shows the Manhattan plot of the significance of the pleiotropy test of the 680,758 SNP in the analyses. There were 788 highly significant pleiotropic SNP (p-value < $10^{-7}$), of which 646 were located in chromosome 5, 133 were in chromosome X, and 9 were spread in the rest of the genome.

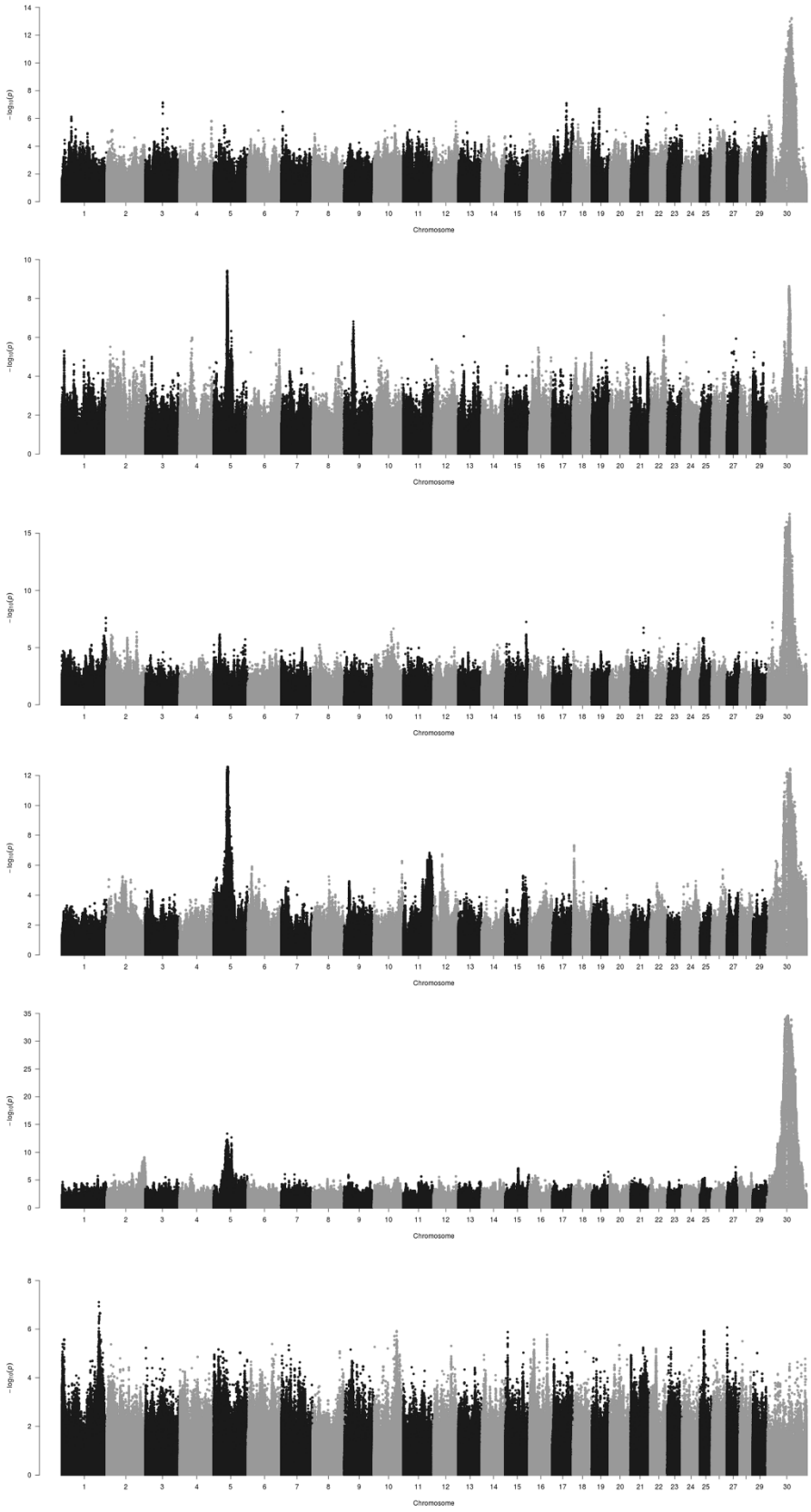


**Figure 8. Genome-wide association for SC in six different breeds.** Top to bottom Brahman (BRM), Santa Gertrudis (SGT), Droughtmaster (DMT), Ultra-black (UBK), Tropical Composite (TRC) and Belmont Tropical Composite (BTC).

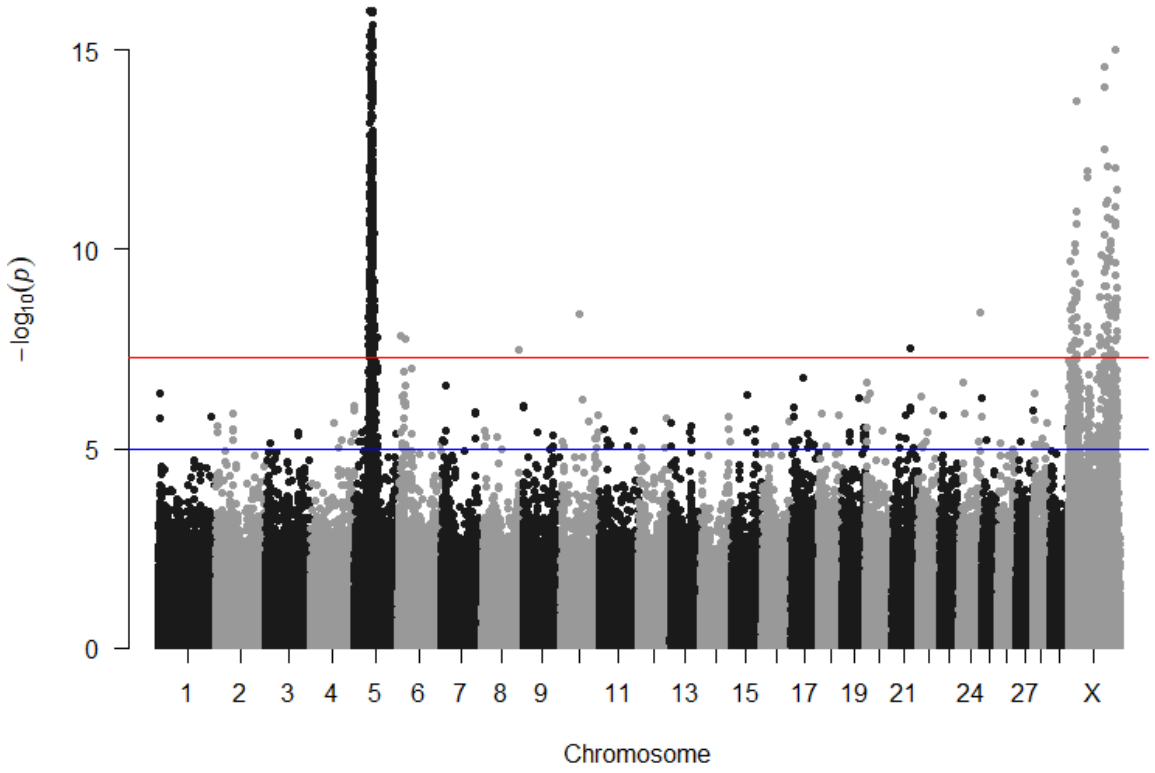


**Figure 9. Manhattan plot of the significance for the pleiotropy test.** There are 680,758 SNP included in the analyses. Blue and red line indicate significant thresholds at $P < 10^{-5}$ and $P < 10^{-7}$, respectively.

Specifically, for BTA5, Table 13 lists 34 genes harbouring SNP in their coding regions with a highly significant (p-value < $10^{-7}$) pleiotropic effect for bull fertility phenotypes. The most prominent region from the chi-square statistic is mapped from 47.2 Mb to 47.8 and contains five genes: *GRIP1, HELB, IRAK3, TMBIM4,* and *HMGA2*. A recent study by our group exploring selection signatures in tropical cattle revealed a missense mutation in the damage response gene *HELB* (Naval-Sánchez et al. 2020). Earlier work also by our group also pointed at *HELB* as a candidate gene for regulating the inhibin hormone, produced by Sertoli cells, which can be measured at four months of age and was suggested as an early biomarker for sexual development (Fortes et al. 2013).

Indeed, bovine chromosome 5 has long been the subject of great scrutiny. For instance, a literature search on PubMed.gov using the string "bovine chromosome 5" results in 138 publications. Of these, it is worth remarking the USDA work by McDaneld et al. (2014), who identified a deletion on chromosome 5 associated with reproductive efficiency in *Bos indicus*-influenced cattle. Importantly, and specifically for sheath score, Aguiar et al. (2018) reported an association of copy number variation at intron 3 of *HMGA2* with navel length in *Bos indicus* cattle.

**Table 13. Genes in Chromosome 5 harbouring SNP in their coding region with highly significant (P < $10^{-7}$) pleiotropic effect for bull fertility phenotypes.**

| Gene | SNP bp position | Chi-square | -Log(P) |
|---|---|---|---|
| CNTN1 | 39,774,560 | 55.577 | 7.608 |
| CPNE8 | 42,460,841 | 95.463 | 15.352 |
| PTPRR | 42,775,924 | 102.919 | 16.000 |
| PTPRB | 42,928,340 | 71.461 | 10.636 |
| MYRFL | 43,596,439 | 73.277 | 10.988 |
| RAB3IP | 43,697,199 | 58.797 | 8.213 |
| BEST3 | 43,767,053 | 74.885 | 11.300 |
| YEATS4 | 44,093,212 | 56.267 | 7.738 |
| LYZ1 | 44,397,878 | 68.456 | 10.056 |
| MDM2 | 44,997,514 | 60.333 | 8.504 |
| ENSBTAG00000051975 | 45,017,190 | 75.760 | 11.471 |
| DYRK2 | 46,125,383 | 124.361 | 16.000 |
| CAND1 | 46,506,894 | 129.648 | 16.000 |
| ENSBTAG00000053087 | 46,572,882 | 175.645 | 16.000 |
| GRIP1 | 47,206,794 | 234.019 | 16.000 |
| HELB | 47,495,826 | 269.899 | 16.000 |
| IRAK3 | 47,600,738 | 304.799 | 16.000 |
| TMBIM4 | 47,657,554 | 307.113 | 16.000 |
| HMGA2 | 47,834,873 | 253.515 | 16.000 |
| MSRB3 | 48,372,783 | 125.971 | 16.000 |
| LEMD3 | 48,608,772 | 119.131 | 16.000 |
| WIF1 | 48,755,876 | 115.579 | 16.000 |
| TBC1D30 | 49,010,441 | 130.908 | 16.000 |
| GNS | 49,074,733 | 92.030 | 14.676 |
| RASSF3 | 49,133,413 | 137.694 | 16.000 |
| TBK1 | 49,291,902 | 188.789 | 16.000 |
| XPOT | 49,417,084 | 220.690 | 16.000 |
| SRGAP1 | 49,724,856 | 134.623 | 16.000 |
| ENSBTAG00000051362 | 49,903,137 | 62.425 | 8.901 |
| RXYLT1 | 49,933,497 | 72.039 | 10.748 |
| PPM1H | 50,694,900 | 143.784 | 16.000 |
| FAM19A2 | 51,871,818 | 78.797 | 12.064 |
| R3HDM2 | 56,148,237 | 57.608 | 7.990 |
| ANKS1B | 62,983,219 | 56.608 | 7.802 |

Table 14 lists 23 genes mapped to chromosome X and harbouring SNP in their coding regions with highly significant (p-value < $10^{-7}$) pleiotropic effect for bull fertility phenotypes. Two of these genes, *AFF2* and *ZNF81*, are transcription factors, while two others, *PLP1* and *VEGFD* have differential expression in fertility-related bovine phenotypes. *PLP1* was differentially expressed between fresh and frozen-thawed sperm of Holstein bulls (Chen et al. 2015). With cryodamage being a major problem in semen cryopreservation, causing changes to sperm transcripts that may influence sperm function and motility, it is tempting to speculate the use of *PLP1* as a biomarker of sperm quality. More recently, Hayashi et al. (2019) demonstrated that the *VEGF* family is expressed and regulated in the bovine uterus during the peri-implantation period, which may be associated with uterine

functions, including vascular remodelling in maternal recognition of pregnancy and implantation. Its role for bull fertility is yet unknown.

**Table 14. Genes in Chromosome X harbouring SNP in their coding region with highly significant ($P < 10^{-7}$) pleiotropic effect for bull fertility phenotypes.**

| Gene | SNP bp position | Chi-square | -Log(P) |
|---|---|---|---|
| MCTS1 | 5,150,584 | 60.329 | 8.503 |
| ENOX2 | 14,779,188 | 62.764 | 8.965 |
| MOSPD1 | 18,596,694 | 68.795 | 10.121 |
| FGF13 | 22,261,944 | 53.628 | 7.245 |
| AFF2 | 31,323,415 | 55.864 | 7.662 |
| ENSBTAG00000050056 | 50,615,669 | 57.195 | 7.912 |
| DRP2 | 50,951,444 | 58.044 | 8.072 |
| PLP1 | 52,547,466 | 78.308 | 11.968 |
| ZNF81 | 86,318,735 | 61.990 | 8.818 |
| DGKK | 88,202,750 | 56.150 | 7.716 |
| KDM6A | 98,741,063 | 57.987 | 8.061 |
| CASK | 102,062,494 | 65.758 | 9.537 |
| RPGR | 104,975,526 | 56.209 | 7.727 |
| SRPX | 105,115,127 | 61.913 | 8.804 |
| SYTL5 | 105,316,656 | 60.674 | 8.568 |
| ENSBTAG00000047410 | 107,118,481 | 78.926 | 12.089 |
| ENSBTAG00000008248 | 111,970,902 | 68.309 | 10.027 |
| ENSBTAG00000050834 | 114,363,399 | 72.162 | 10.771 |
| NHS | 125,837,596 | 54.177 | 7.347 |
| PIR | 128,106,540 | 78.751 | 12.055 |
| VEGFD | 128,114,535 | 93.717 | 15.000 |
| PRPS2 | 130,811,730 | 60.133 | 8.466 |
| MID1 | 132,176,375 | 57.335 | 7.938 |

Finally,

Table **15** lists the nine SNPs in other chromosomes that were significant (p-value < $10^{-7}$) in the pleiotropic test for bull fertility phenotypes. Three of them are nearby or in the coding regions of genes previously mapped to selection signatures in cattle, including *LCORL, RORA,* and *WDR7* (Xu et al. 2015; Wang et al. 2019). Also, *LCORL* and *RORA* are transcription factors. It is tempting to guilt transcription factors when pleiotropy is discussed because their role is to drive the expression of multiple genes, and therefore we can hypothesise they are likely to influence multiple phenotypes. An example of a potentially pleiotropic transcription factor is *PLAG1* on chromosome 14, associated with many female traits (Fortes et al. 2013).

**Table 15. SNP in chromosomes other than 5 and X with highly significant ($P < 10^{-7}$) pleiotropic effect for bull fertility phenotypes.**

| SNP ID* | Chr | Bp | Distance to nearest Gene | Nearest Gene | Chi-square | -Log(P) |
|---|---|---|---|---|---|---|
| SNP_173334 | 6 | 6,553,411 | 0 | *SEC24D* | 56.724 | 7.823 |
| SNP_176525 | 6 | 18,130,874 | 0 | *DKK2* | 56.353 | 7.754 |
| SNP_181805 | 6 | 37,489,614 | 0 | *LCORL* | 52.422 | 7.021 |
| SNP_262199 | 8 | 102,041,877 | 0 | *SNX30* | 54.983 | 7.497 |
| SNP_262200 | 8 | 102,043,448 | 0 | *SNX30* | 54.946 | 7.491 |
| SNP_262202 | 8 | 102,045,298 | 0 | *SNX30* | 54.946 | 7.491 |
| SNP_305038 | 10 | 48,871,728 | 20,797 | *RORA* | 59.603 | 8.366 |
| SNP_543411 | 21 | 48,576,534 | 180,608 | *CLEC14A* | 55.121 | 7.523 |
| SNP_593856 | 24 | 56,309,078 | 0 | *WDR7* | 59.936 | 8.429 |

*Map sequence ID among the 680,758 SNP included in the analyses.

The within-breed genome-wide association analyses and the pleiotropic test applied to the SNP effects identified genomic regions associated with several bull fertility traits. These steps could be considered the first towards the fine mapping of functional genetic variants that affect the different traits, leading to their use in animal breeding. Additional research is needed to further explore the prioritized markers and genomic regions. In this context, the dataset created in this project is ideal for asking the following research questions: a) which are the genes and mutations driving these effects on the bull fertility traits? b) how does the biological function of these genes affect the traits? c) Can we use these prioritized markers to derive more accurate GEBVs? d) would GEBVs calculated using these prioritized markers be more stable across different breeds? e) What are the genomic correlations between these traits to other production traits, including female fertility traits?

These are some of the questions that should be explored in the future. The dataset assembled here is very well positioned to form the basis of future research on bull fertility.

# 5. Training and capability building in cattle genomics

This project assembled one of the most complete genomic datasets in bull fertility; it created many capability-building opportunities and will serve the industry in the years to come. During this short two yeas project, via the participants based at the University of Queensland, this project supported the development of four Honours Thesis (two completed, and two ongoing), one Master of Science (ongoing), and two PhD Thesis (both ongoing). Although the training of the next generation of livestock scientists was not part of the formal objectives of this project, we considered it a fundamental activity for the future of the industry. Therefore, we supported these students and their research activities and dedicated time to training them.

# 6. Conclusion

## 6.1 Key findings

The use of phenotype imputation was explored using the Beef CRC dataset. We found that imputing missing observations while calculating GEBV might return improved estimates. However, more research on this topic is needed to evaluate its effects and define a path for implementation.

The traits observed during the Bull Breeding Soundness Examination (BBSE) are heritable to different degrees and influenced by many genes (polygenic architecture). These observations were confirmed across breeds in this project and we have established a solid conceptual basis for using BBSE records in selective breeding.

The polygenic nature of the BBSE traits and the lack of genetic variants explaining a large proportion of the variance suggest that genomic selection approaches are the best approach for the genetic improvement of male fertility traits in beef herds.

We identified and prioritized DNA markers (and genomic regions) relevant to several bull fertility traits. Specific regions in chromosomes 5 and X are noteworthy for their association with fertility traits in the majority of the studied breeds. These findings can be used to fine-tune genomic selection approaches and as a starting point for further research to better understand the biological function of the genes that affect these traits.

Multibreed GEBV for bull fertility traits were obtained. A reference population of approximately 7,000 animals that included six breed types with a matched DNA profile was able to generate multibreed genomic predictions with useful accuracies for the Industry.

## 6.2 Benefits to industry

We have built great resources for genetics and genomics RD&A for the Beef Industry. We assembled a dataset of nearly 10,000 bulls with BBSE records, of which 7,000 have matched DNA profiles representing six tropical breed types. This resource is available for future research to better understand the biology and the genetics of bull fertility and further develop genomic selection approaches for BBSE traits.

Using the genetic resources assembled in this project, we confirm that bull fertility traits are heritable and controlled by several genes.

Additionally, we generated GEBV for BBSE traits with useful accuracies. Bull breeders could use these GEBV to rank and select bulls aiming at trait improvement.

This project set up a multibreed reference population that can be expanded and used to implement genomic selection approaches for bull fertility traits.

We demonstrated that a relatively small multibreed reference population (~7,000 records) could generate GEBV with useful accuracies. This is encouraging to the broader Beef Industry considering genomic selection approaches where only a limited number of records are available.

# 7. Future research and recommendations

Based on the positive results of this project, we recommend:

- Working with the Beef Industry to grow the multibreed reference population used to investigate bull fertility traits. We recommend establishing a process to engage with producers, to collect and compile BBSE records with matched DNA profiles to cement a multibreed population in our research arena.
- Undertaking further research on the genomic architecture of the bull fertility traits investigated in this project to a) better understand the genes controlling these traits, b) explore possible avenues to assist in improving the accuracy of genomic prediction (GEBV), and c) investigate other potential uses of the defined functional mutations affecting these traits; and
- Establishing a process for delivering multibreed genomic prediction for bull fertility traits to the industry.

# 8. Acknowledgments

We are thankful to the bull breeders that contributed data to this project and gave access to stored DNA samples for genotyping; without them, this project wouldn't have been possible. Specifically, we would like to acknowledge the contributions from Nikki and Peter Mahony (Gyranda), Swin Hudson (Tremere), Robert Rea Sr and Jr (Lisgar), Ian Murdock, Nick Cameron and Nat McGhee (Nindooinbah), Eric Simon (Jimarndy), and Rodger Jefferies (Elrose). It is worth noting that this project used historical BBSE records paid for by these producers that together represent an industry investment, by these producers, of more than $500,000 dollars.

A fraction of the BBSE data included in this project was collected during the project "Next gen beef breeding strategies for the Northern Australian Beef Industry" funded by the Queensland Government under the Smart Futures Research Partnership Program (2012-2015). We acknowledge the contribution of Prof Brian Burns, Dr Geoffry Fordyce, and their collaborators.

This project also utilized data from the CRC for Beef Genetic Technologies (Beef CRC). We acknowledge the previous investment and contributions from several pastoral companies and the Beef CRC participants, including AgForce, QLD DPI, UQ, AGBU, CSIRO, and MLA.

# 9. References

Aguiar TS, Torrecilha RBP, Milanesi M, Utsunomiya ATH, Trigo BB, Tijjani A, Musa HH, Lopes FL, Ajmone-Marsan P, Carvalheiro R, Neves HHR, do Carmo AS, Hanotte O, Sonstegard TS, Garcia JF, Utsunomiya YT (2018) **Association of Copy Number Variation at Intron 3 of HMGA2 With Navel Length in Bos indicus**. Front Genet. 2018 Dec 7;9:627. doi: 10.3389/fgene.2018.00627.

Bolormaa S., Porto Neto L.R., Zhang Y.D., Bunch R.J., Harrison B.E., Goddard M.E. & Barendse W. (2011) **A genome wide association study of meat and carcass traits in Australian cattle**. *J Anim Sci* 89, 2297-309.

Bolormaa S, Pryce JE, Kemper K, Savin K, Hayes BJ, et al. (2013) **Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in Bos taurus, Bos indicus, and composite beef cattle.** Journal of Animal Science 91: 3088–3104.

Bolormaa S., Pryce J.E., Reverter A., Zhang Y., Barendse W., Kemper K., Tier B., Savin K., Hayes B.J. & Goddard M.E. (2014) **A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle**. *PLoS Genet* 10, e1004198.

Chen X, Wang Y, Zhu H, Hao H, Zhao X, Qin T, Wang D. (2015) **Comparative transcript profiling of gene expression of fresh and frozen-thawed bull sperm.** Theriogenology. 1;83(4):504-11. doi: 10.1016/j.theriogenology.2014.10.015.

Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C. (2016) **Next-generation genotype imputation service and methods**. Nat Genet. 48(10):1284-1287. doi: 10.1038/ng.3656

Fordyce G, Entwistle K, Norman S, Perry V, Gardiner B, Fordyce P. (2006) **Standardising bull breeding soundness evaluations and reporting in Australia**. Theriogenology. 15;66(5):1140-8. doi: 10.1016/j.theriogenology.2006.03.009.

Fortes, MR, Kemper, K, Sasazaki, S, Reverter, A, Pryce, J, Barendse, W, Bunch, R, McCulloch, R, Harrison, BE, Bolormaa, S, Zhang, YD, Hawken, R, Goddard, M, Lehnert, SA (2013) **Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle**. Animal Genetics. 44(6):636-647. doi: 10.1111/age.12075.

Fortes M.R., Reverter A., Hawken R.J., Bolormaa S. & Lehnert S.A. (2012) **Candidate genes associated with testicular development, sperm quality, and hormone levels of inhibin, luteinizing hormone, and insulin-like growth factor 1 in Brahman bulls**. *Biol Reprod* 87, 58.

Fortes M.R., et al. 2020. **X chromosome variants are associated with male fertility traits in two bovine populations**. Genet Sel Evol 52:46.

Fortes MR, Reverter A, Kelly M, McCulloch R, Lehnert SA. (2013**) Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species**. Andrology. 2013 Jul;1(4):644-50. doi: 10.1111/j.2047-2927.2013.00101.x.

Hawken R.J., Zhang Y.D., Fortes M.R.S., Collis E., Barris W.C., Corbet N.J., Williams P.J., Fordyce G., Holroyd R.G., Walkley J.R.W., Barendse W., Johnston D.J., Prayaga K.C., Tier B., Reverter A. & Lehnert S.A. (2012) **Genome-wide association studies of female reproduction in tropically adapted beef cattle**. *J Anim Sci* 90, 1398-410.

Hayashi KG, Hosoe M, Fujii S, Kanahara H, Sakumoto R. (2019) **Temporal expression and localization of vascular endothelial growth factor family members in the bovine uterus during peri-implantation period**. Theriogenology. 2019 Jul 15;133:56-64. doi: 10.1016/j.theriogenology.2019.04.021.

Legarra A and Reverter A (2018) **Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method**. Genet Sel Evol. 6;50(1):53. doi: 10.1186/s12711-018-0426-6.

Loh PR, Palamara PF, Price AL. (2016) **Fast and accurate long-range phasing in a UK Biobank cohort.** Nat Genet. 48(7):811-6. doi: 10.1038/ng.3571.

Kemper K.E., Reich C.M., Bowman P.J., Vander Jagt C.J., Chamberlain A.J., Mason B.A., Hayes B.J. & Goddard M.E. (2015) **Improved precision of QTL mapping using a nonlinear Bayesian method in a multibreed population leads to greater accuracy of across-breed genomic predictions**. *Genet Sel Evol* 47:29.

McDaneld TG, Kuehn LA, Thomas MG, Pollak EJ, Keele JW. (2014) **Deletion on chromosome 5 associated with decreased reproductive efficiency in female cattle**. J Anim Sci. 2014 Apr;92(4):1378-84. doi: 10.2527/jas.2013-6821.

Perez-Enciso M, Misztal I (2011) **Qxpak.5: Old mixed model solutions for new genomics problems**. BMC Bioinformatics 12: 202.

Perez-Enciso M., Rincon J.C. & Legarra A. (2015) **Sequence- vs. chip-assisted genomic selection: accurate biological information is advised**. *Genet Sel Evol* 47:43.

Porto-Neto L.R., Reverter A., Prayaga K.C., Chan E.K.F., Johnston D.J., Hawken R.J., Fordyce G., Garcia J.F., Sonstegard T.S., Bolormaa S., Goddard M.E., Burrow H.M., Henshall J.M., Lehnert S.A. & Barendse W. (2014) **The Genetic Architecture of Climatic Adaptation of Tropical Cattle**. *PLoS ONE* 9, e113284.

Porto-Neto L.R., Barendse W., Henshall J.M., McWilliam S.M., Lehnert S.A. & Reverter A. (2015) **Genomic correlation: harnessing the benefit of combining two unrelated populations for genomic selection.** *Genet Sel Evol* 47:84.

Reverter A., Porto-Neto L.R., Fortes M.R.S., Kasarapu P., de Cara M.A.R., Burrow H.M. & Lehnert S.A. (2017) **Genomic inbreeding depression for climatic adaptation of tropical beef cattle**. *J Anim Sci* 95, 3809-21.

Strandén I, Garrick DJ. (2009) **Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit**. J Dairy Sci. 2009 Jun;92(6):2971-5. doi: 10.3168/jds.2008-1929.

VanRaden PM. (2008) **Efficient methods to compute genomic predictions**. J Dairy Sci. 2008 Nov;91(11):4414-23. doi: 10.3168/jds.2007-0980.

Xu L, Bickhart DM, Cole JB, Schroeder SG, Song J, Tassell CP, Sonstegard TS, Liu GE. (2015) **Genomic signatures reveal new evidences for selection of important traits in domestic cattle.** Mol Biol Evol. 2015 Mar;32(3):711-25. doi: 10.1093/molbev/msu333.

Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. (2012) **Genome-wide association mapping including phenotypes from relatives without genotypes.** Genet Res (Camb). 2012 Apr;94(2):73-83. doi: 10.1017/S0016672312000274

Wang Z, Ma H, Xu L, Zhu B, Liu Y, Bordbar F, Chen Y, Zhang L, Gao X, Gao H, Zhang S, Xu L, Li J. (2019) **Genome-Wide Scan Identifies Selection Signatures in Chinese Wagyu Cattle Using a High-Density SNP Array.** Animals (Basel). 2019 May 30;9(6):296. doi: 10.3390/ani9060296.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. (2010) **Common SNPs explain a large proportion of the heritability for human height.** Nat Genet. 42(7):565-9. doi: 10.1038/ng.608.

## Appendix 1 – Phenotype imputation of the Beef CRC data: summary statistics of Brahman and Tropical Composite cows.

| | | BRAHMAN (n=995) | | | | TROPICAL COMPOSITE (N=1094) | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **Description** | **Mean** | **Std Dev** | **Min** | **Max** | **Mean** | **Std Dev** | **Min** | **Max** |
| **FT** | Flight time | 139.59 | 41.28 | 65.00 | 253.79 | 135.10 | 31.13 | 52.00 | 200.10 |
| **TEMP** | Rectal temperature | 39.14 | 0.52 | 37.73 | 39.95 | 38.49 | 0.56 | 37.18 | 40.69 |
| **EPG** | Worm eggs per gram | 6.03 | 1.61 | -0.06 | 8.80 | 6.11 | 1.55 | 1.08 | 9.15 |
| **SHEATH** | Navel score | 5.38 | 0.46 | 3.84 | 7.00 | 7.90 | 0.57 | 6.03 | 9.34 |
| **COLOR** | Coat color | 3.38 | 0.26 | 2.40 | 4.26 | 3.75 | 0.37 | 2.00 | 4.37 |
| **FLY** | Fly lesions score | 1.06 | 0.31 | 0.19 | 3.20 | 0.74 | 0.30 | 0.13 | 1.50 |
| **TICK** | Tick score | 0.79 | 0.58 | -0.17 | 2.68 | 2.35 | 0.54 | 0.65 | 4.00 |
| **COAT** | Coat score | 5.02 | 0.73 | 3.26 | 7.55 | 7.38 | 1.27 | 1.89 | 11.00 |
| **YCOND** | Yearling condition score | 8.01 | 0.41 | 6.32 | 9.80 | 7.31 | 0.41 | 6.19 | 9.18 |
| **YWT** | Yearling weight | 209.73 | 24.65 | 135.18 | 275.31 | 216.69 | 18.29 | 164.26 | 264.00 |
| **ACL** | Age at first corpus luteum | 750.03 | 88.99 | 467.00 | 1056.00 | 650.62 | 63.17 | 476.66 | 786.64 |
| **MatHH** | Mature hip height | 140.91 | 1.63 | 132.35 | 146.73 | 137.30 | 1.65 | 129.43 | 141.24 |
| **MatWT** | Mature weight | 499.38 | 31.92 | 404.17 | 587.79 | 523.57 | 23.65 | 441.25 | 581.37 |
| **MatEMA** | Mature Eye muscle area | 61.70 | 4.11 | 51.16 | 75.34 | 60.24 | 4.02 | 49.03 | 77.00 |
| **PPAI** | Post-partum anoestus interval | 306.80 | 199.94 | 62.57 | 693.17 | 153.47 | 70.58 | 52.25 | 393.89 |
| **IGF** | IGF-I concentration | 191.42 | 75.09 | 60.75 | 352.67 | 228.43 | 51.23 | 41.73 | 642.00 |
| **DC1** | Days-to-calving first mating | 345.45 | 43.70 | 306.58 | 418.25 | 318.99 | 32.49 | 298.57 | 426.41 |
| **DC5** | Days-to-calving first five matings | 344.38 | 12.67 | 308.55 | 370.04 | 329.54 | 11.85 | 297.99 | 368.46 |