Final report

Accelerating genetic gain for productivity and profitability in Northern beef cattle with genomic technologies

Project Code: P.PSH.0833

Prepared by: Ben Hayes, Shannon Speight, Elizabeth Ross, Elise Dodd, Geoffry Fordyce

University of Queensland

Date published: 27 January 2022

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

This is an MLA Donor Company funded project.

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

Abstract

Cow fertility is a key driver of profitability in northern Australian beef enterprises. While substantial genetic variation exists for this trait, it has been difficult to improve. This project has produced two products that will enable northern Australian beef enterprises to improve fertility of their cows: **Genomic breeding values** (GBV) from the project have useful accuracy for selecting teams of bulls, or for culling heifers that are unlikely to perform, across a wide range of breeds, crossbreds and composites. GBV traits included heifer puberty, pregnant 4 months after calving, weight, body condition score, temperament, tick resistance and buffalo fly lesion score. The GBV for a herd can also be summarised into a **herd profile**, that benchmarks the herd and can be used as a tool to identify which traits should be the focus in bull selection decisions. The GBV are based on data collected from more than 26,000 genotyped and trait recorded heifers in 54 collaborating herds across northern Australia, with a wide range of breeds, crossbreds and composites represented. The project has also identified a panel of 70 DNA markers (SNP) from whole genome sequence data that are highly associated with fertility, growth and adaptation traits. These SNP will be incorporated into future SNP arrays used for commercial genotyping, and will increase both the accuracy of GBV from this project, and also BREEDPLAN single step EBV for Northern breeds.

Executive summary

Background

Cow fertility is a key driver of productivity and profitability of beef production in northern Australia (Harburg et al. 2018). While substantial genetic variation exists for cow fertility, it is a difficult trait to select for, as with traditional EBV bulls must have many daughters recorded for the trait before these EBV are accurate. If accurate Genomic breeding values (GBV) for fertility were available, these could be used to select young bulls with high genetic merit, accelerating genetic gain. However, there are at least two challenges to developing accurate GBV to enable beef producers to accurately select for fertility. The herd in northern Australia consists of many breeds, crossbreds and composites, so the GBV must be multi-breed – that is be accurate across all these breeds, crossbreds and composites. Secondly, to generate accurate GBV for a trait such as fertility, very large numbers of heifers/cows must be recorded for fertility traits and genotyped for the reference population, which is challenging in the extensive conditions typical of northern Australia beef enterprises.

In this project, we collaborated with 54 herds across northern Australia to record fertility traits, growth traits, body condition score, temperament, tick scores and buffalo fly lesion scores on more than 26,000 heifers and cows, representing the breeds, crossbreds and composites used in the north. GBV prediction equations were derived that will enable commercial herds to select bulls and heifers for improved fertility and performance on an industry wide scale. The GBV can also be used to benchmark northern Australian beef enterprises for their genetic level for different traits.

Objectives

1. Develop methodology for calculating accurate genomic breeding value for key traits affecting productivity in Northern Beef cattle.
   This objective was achieved. Genomic breeding values (GBV) for two fertility traits (heifer puberty, pregnant four months after calving), weight, height, body condition score, temperament, tick resistance and buffalo fly lesion score were developed, based on trait records from 26,000 heifers.
2. Develop a low cost SNP array including i. SNP that improve accuracy of genomic breeding values for breeds and composites used in Northern beef production ii. SNP that predict polled/horned iii. SNP that can be used for pedigree reconstruction in these breeds/composites, iv. SNP that predict breed composition, and v. Any recessive defects.
   This objective was achieved. A panel of 512 SNP was derived using genome wide association studies in this data set with whole genome sequence that are highly associated with fertility traits, temperament, tick score and growth. Twenty of these SNP were incorporated into the 2019 update to Neogen's TropBeef Array, which is commercially available. The remainder of the SNP will be incorporated in the next TropBeef update. A breed composition predictor was developed, predicting breed content of animals from the genotype data for 14 breeds from the Tropbeef SNP array.
3. Develop novel ways of presenting genomic information that resonate and are meaningful to commercial producers in Northern Australia, to accelerate adoption of the use of genomic information in bull selection decisions.
   This objective was achieved. The owners of the 54 collaborating herds in the project participated in extensive discussions on the best way to present the GBV information. The

consensus was quintile plots, which can be used to quickly benchmark the genetic level for different traits, are very useful.

4. Establish a core group of up to 75 herds that represent the production systems, and breeds and composites present in Northern Australia for validating genomic breeding values and evaluating alternative ways of presenting genetic and genomic information.
   This objective was achieved. Fifty-four herds across northern Australia participated in the project, ranging from family farms to large corporates. Many breeds, cross bred and composites were represented in the data set. A proportion of these herds are now involved in the NB2 project.
5. Train the next generation of researchers and professionals that will have the capacity to deliver large research projects that impact on productivity of beef production in Northern Australia.
   The project trained two professionals in large scale beef enterprise data management and analysis, as well as the fundamentals of genomic breeding values. Two PhDs students are in training.

Methodology

Fifty four collaborating herds were selected on a range of criteria, including large cohorts of heifers in single management groups, tight bull management, representation of the environments and breeds, crossbreds and composites used in northern Australia, and willingness to record the required data over the life of the project (up to three heifer cohorts). Ultrasound scanning by trained vets was used to record heifer puberty (cycling or not at approximately 600 days). Weight, body condition score, height, pregnant or not 4 months after calving, tick score, temperament score, and buffalo fly lesion score were also required. All 26,000 heifers were genotyped with either 35k or 50k SNP markers with the TropBeef array, and these genotypes were imputed up to 709768 SNP for GBV derivation using a genomic best linear unbiased prediction (GBLUP) approach. The same data, with the genotypes imputed up to whole genome sequence (48.8 million variants) was used in genome wide association studies to identify genetic variants highly associated with the traits for inclusion on new SNP arrays that improve accuracy of GBV. Extensive dialogue with the fifty four collaborating herds was undertaken to determine the most useful way of reporting GBV results.

Results/key findings

Genomic heritabilities and estimates of heterosis from the data were consistent with previous studies, suggesting the data is of good quality. The GBV derived from the data were of useful accuracy (0.3-0.45) in a validation with independent herds, and the GBV also appropriately predicted related traits in the Beef CRC data (GBV for heifer puberty were correlated with age at first CL, GBV for P4M were correlated with post partum anoestrus interval).

Benefits to industry

The prediction equations from this project can be used to calculate GBV for young bulls and heifers for fertility, growth and adaptation traits across a wide range of breeds, crossbreds and composites, enabling selection for these traits on a industry wide scale. The GBV for a herd can also be used to benchmark the genetic level of the herd, identifying areas of focus to improve profitability through bull selection in the future.

Future research and recommendations

This technology should be commercialised and rolled out, together with case studies of beef enterprises actually using the GBV to demonstrate their value, to the Northern Australian beef industry.

Table of contents

1. Background

Cow fertility is a key driver of productivity and profitability for northern Australian beef enterprises (Taylor and Rudder 1986, Fordyce 2012, Johnstone et al. 2014, Harburg et al. 2018). Genomic estimated breeding values (GBV) for cow fertility would enable more rapid genetic gains for these traits, through accurate and early selection of bulls and heifers with high genetic merit. However, accurate genomic evaluations for low heritability traits such as fertility require large reference populations (e.g. Goddard and Hayes 2009), with thousands of cows measured for both the traits of interest and genotyped for genome wide markers. Assembling such large reference populations for each breed used in Northern Australia is likely to be challenging, as the cattle population consists of many breeds, crossbreds and composites. Cattle populations include high proportion Bos indicus breeds (e.g. Brahman), stabilised composites (e.g. Droughtmaster and Santa Gertrudis), adapted Bos taurus breeds, and many composites. An alternative to constructing reference populations within each breed is to use multi-breed genomic evaluations, where the reference set includes cows from across Northern Australia.

Recognising that large numbers of phenotypes and genotypes are required to develop accurate genomic breeding values and this is a major barrier particularly for fertility traits, we will adopt and enhance the data collection strategies used in the CashCow project to collect fertility data and other traits from up to 75 producers/seedstock herds and potentially pastoral company herds representing the four country types and major breeds and composites in Northern Australia. The CashCow project (MLA final report B.NBP.0382) demonstrated that useful fertility phenotypes can be collected on large scale in commercial herds. This workpackage will proceed in two stages, a sub project on Spyglass research station where will test the suitability of CashCow type data for calculating genomic breeding values, compared with the data being collected for the BINS – to make the comparison direct the two data sets will be collected on the same animals. If this is successful, we will then collect data from the 75 cooperating properties (30,000 cattle), over four years. One constraint for the use of this information will be collection of pedigrees – in this workpackage we will use genomic information rather than pedigrees for commercial cattle. Selection criteria for the collaborating herds will include integrity of existing records, willingness to muster at appropriate times for ovarian and fetal scanning/ageing, breed composition of the herd.

As recently demonstrated by Kemper et al. (2015) and Macleod et al. (2016), using 50K SNP genotypes and BLUP methodology does not enable multi-breed predictions. That is, the marker effects only work within breeds when 50K SNP are used. As demonstrated by both those authors, much higher density of markers, up to whole genome sequence, does begin to enable multi-breed predictions, where breeds not in the reference set, or with only limited numbers in the reference set, can have genomic evaluations with moderate accuracy. Of all the genetic variants in the sequence data, only a small proportion of these will be necessary to enable the multi-breed evaluations. The trick is to find these. The first step is to infer whole genome sequence data into all beef animals that have been genotyped, including the Beef CRC animals, Repronomics, Smart Futures, and the BIN animals, using the 1000 bull genome reference set of sequenced animals (Daetwyler et al. 2014). Where key breeds are underrepresented, we will add additional sequences of key bulls from those breeds to the 1000 bull genomes project. We will then use this sequence data to develop a new, low cost accuracy array highly enriched for predictive SNP that increases the accuracy of GBV for key traits in northern Australian beef cattle.

This project is a core and major project within the National Livestock Genetics Consortium, and is very closely aligned with the consortium's priorities and goals, as well as MISP especially pillar 4

(productivity and profitability in red meat and livestock enterprises). The overall goal for the National Livestock Genetics Consortium is to double the annual rate of genetic gain for profit by 2022, this project will make a major contribution to that goal by 1) developing methodology for multi-breed genomic estimated breeding values for Northern beef cattle that can be applied in BREEDPLAN, 2) increasing the accuracy of genomic breeding values for key traits affecting productivity in the North by significantly enlarging the number of animals genotyped and phenotyped, in up to 75 core commercial herds informed by and building on CashCow methodology, 3) reducing the cost of applying genomic selection, and 4) working closely with the same group of up to 75 commercial properties, develop new and innovative ways of presenting genomic breeding values that is meaningful and resonates with commercial producers, and increases uptake of the technology.

Specifically this project with contribute to the following key objectives of the National Livestock Genetics Consortium.

- Develop world leading genetic improvement technologies and resources. The multi-breed genomic evaluation approach that will be developed here, based on DNA sequence information will be world leading. The up to 75 herds that will participate in the project is a unique resource.
- Maximise the effectiveness and value of genetic improvement tools. The multi-breed approach will ensure genomic breeding values can be applied to any animal, regardless of breed composition.
- Maximise the adoption of genetic improvement tools. By working with up to 75 herds to identify the language that resonates with commercial producers for describing genomic breeding values for key traits, uptake of the technology will be accelerated.
- Fast track genetic gain. By developing accurate genomic breeding values that can be applied regardless of breed composition, delivered with a lost cost chip, genetic gain will be fast tracked.
- Develop easy to use, accessible products and services. By working with 75 herds to identify the language that resonates with commercial for describing genomic breeding values for key traits, ease of use of the technology will improve.
- The delivery of disruptive transformation change. Genomic breeding values with useful accuracy regardless of breed composition at low cost would represent a disruptive change for the industry, allowing significantly greater genetic gains.

2. Objectives

The objectives of this project were:

1. Develop methodology for calculating accurate genomic breeding value for key traits affecting productivity in Northern Beef cattle that can be implemented in BREEDPLAN. The methodology will exploit whole genome sequence data, and build on existing one step methodologies developed by Animal Genetics Breeding Unit (AGBU).

This objective was achieved. Genomic breeding values (GBV) for two fertility traits (heifer puberty, pregnant four months after calving), weight, height, body condition score, temperament, tick resistance and buffalo fly lesion score were developed, based on trait records from more than 26,000 heifers. Whole genome sequencing was performed on 200 foundation Brahman, Santa Gertrudis and

Droughtmaster bulls. This whole genome sequence data was exploited to improve the accuracy of the GBV. Incorporation of the data in BREEDPLAN, and appropriate models, have been developed and discussed with AGBU and ABRI staff.

2. Develop a low cost SNP array including i. SNP that improve accuracy of genomic breeding values for breeds and composites used in Northern beef production ii. SNP that predict polled/horned iii. SNP that can be used for pedigree reconstruction in these breeds/composites, iv. SNP that predict breed composition, and v. Any recessive defects.

This objective was achieved. A panel of 512 SNP was derived using genome wide association studies in this data set and others, that are highly associated with fertility traits, temperament, tick score and growth. 112 of these SNP were incorporated into the 2019 update to Neogen's TropBeef Array, which is commercially available and used in Breedplan one step analysis for some Northern breeds. The remainder of the SNP will be incorporated in the next TropBeef update. A breed composition predictor has been developed, predicting breed content of individual animals from the genotype data for 14 breeds.

3. Develop novel ways of presenting genomic information that resonate and are meaningful to commercial producers in Northern Australia, to accelerate adoption of the use of genomic information in bull selection decisions.

This objective has been achieved. The owners of the 54 collaborating herds in the project participated in extensive discussions on the best way to present the GBV information. The consensus was quintile plots, which can be used to quickly benchmark the genetic level for different traits, was very useful.

4. Establish a core group of up to 75 herds that represent the production systems, and breeds and composites present in Northern Australia for validating genomic breeding values and evaluating alternative ways of presenting genetic and genomic information.

This objective was achieved. Fifty-four herds across northern Australia participated in the project, ranging from family farms to large corporates. Many breeds, cross bred and composites were represented in the data set. A proportion of these herds are now involved in the NB2 project.

5. Train the next generation of researchers and professionals that will have the capacity to deliver large research projects that impact on productivity of beef production in Northern Australia

One of the professionals employed on the project, Shannon Speight, won a Zanda award in part due to her establishment of the large networks associated with the project. Shannon has now founded BlackBox Co, which uses large data sets to provide insights for beef producers. Two PhDs students are training on and actively using the project data.

3. Methodology

3.1 Develop methodology for calculating accurate genomic breeding value for key traits affecting productivity in Northern Beef cattle

Selection criteria for recruiting herds into the project was informed by previous work, including the Beef CRCs, the Smart Futures Project, and the Cash Cow project. Herds were selected to represent the environments in Northern Australia, and to capture the breeds used in the North (where the breed has reasonable numbers of cattle represented).

Essential criteria included:

- Large contemporary groups of heifers and cows (>50) run together since birth. This criterion is essential to enable genetic and genomic evaluations, as genetic differences between heifers and cows can only be clearly identified if there is a large group of animals that have been subject to the same nutrition, environment and management over their lifetime.
- All cattle must be individually identified by NLIS ear tags at the time of enrolment in the project. All NLIS tags must have either an existing business identifier linked with the NLIS tag, or a virtual number which will be needed for any sampling.
- Ensure all cattle in research groups are managed securely, ie, within defined boundaries and achieving exclusion of uncontrolled cattle movements into or out of research groups, wherever possible.
- Ensure all mated females are assessed for pregnancy status and foetal age at or after the end of annual mating each year (we specified that a veterinarian with PregCheck accreditation can be used, or the project team can provide this service).
- Ability to score cows for udders, teats, body condition score, fly lesions and tick burden (where appropriate) using simple 5-point scales developed by the project team.
- In order to calibrate the environment for each property, allow project team members to visit twice annually around the wet-dry and dry-wet interfaces to conduct rapid visual assessment of available pasture and conduct rapid assessments of fauna populations. Ground cover will be assessed using satellite imagery over digital mapping of paddocks.
- For the same purpose, include a group of at least six monitor steers or heifers within each management group of first lactation-pregnancy females as a reference point for live weight productivity. The reference animals will be selected and weighed annually using a consistent protocol agreed by the project team.

Highly desirable criteria include:

- Multiple breeds, crossbreds or composites run in the same contemporary group of cows and heifers. This will enable multi-breed and across breed analysis. Note that a small proportion of purebred herds have been included in the project, predominately stud herds that provide bulls to commercial herds in the project, in order to provide linkage within the project and linkage to other projects such as Repronomics and Smart Futures.

The project was profiled at MLA BeefUp days (Barcaldine, Nebo, Beaudesert, Karumba), by Department of Agriculture Staff at Various field days, and other events. Interested property owners/managers were contacted by project staff and an interview scheduled at their property. At the interview project staff assessed suitability of the property for the project, against the above selection criteria, and based on history of recording data (properties with a good history of recording) and using data preferred.

Fifty-four herds were recruited and became collaborators on the project. These herds are located across northern Australia, Fig. 1. Effort was made to include herds from Northern Territory and Western Australia, however only one herd in NT and WA that were considered met project criteria (controlled mating).

Figure 1. Location of herds collaborating in the Northern Genomics Project.

Three "phenotype training" days were held on properties in Northern, Central and Southern Queensland. At these well attended days, project collaborators were trained in trait recording required for the project, including for temperament, tick scoring and buffalo fly lesion scoring.

For each herd, data was collected on two or three heifer cohorts (eg heifers born in 2016, 2017 and 2018) for 11 traits. These are cycling or not cycling at 600 days, as determined by ovarian scanning (presence or absence of a corpus luteum when scanned about 600 days, trait designated as Heifer puberty), pregnancy diagnosis 1, pregnancy diagnosis 2 (eg rebreed) weight at scan, body condition score at scan, hip height at scan (and weights, hip height and body condition score at pregnancy diagnosis 1 and 2),tick score (ranging from 0 to 5) , fly lesion score (ranging from 0 to 5), temperament score, and udder and teat scores in lactating cows. The score traits (tick score, fly lesion score, temperament score, and udder and teat scores in lactating cows and body condition score) are scored as described by BREEDPLAN, and in the Northern Genomics Phenotyping manual. The number of records collected for each trait is given in Table 1.

Table 1. Traits recorded in the Northern Genomics Project and the number of animals recorded for each trait as of 30/06/2021.

Trait	Number records
Weight*	61894
Hip Height*	51011
Body Condition Score*	59103
Heifer puberty	26427
Pregnancy diagnosis 1	23530
Pregnancy diagnosis 2	11698
Tick score	2094
Buffalo Fly lesion score	15927
Teat Score	2370
Udder Score	2910
Temperament	3234
Total	**260198**

*Up to three records per animal

A trait indicating potential to rebreed after the first calving was created from pregnancy 1 and pregnancy 2 data. This trait was pregnant or not four months after calving (P4M). Cows that are not pregnant 4 months after calving will not have a calf every year.

All heifers were genotyped with the 35k or 50k tropBeef SNP array by Neogen, Australasia. Genotypes were imputed up to 709,768 SNP (Bovine HD array) using the findhap software (VanRaden et al. 2013), and a panel of 4650 cattle of relevant breeds genotyped for the Bovine HD array.

We first estimated breed proportions of each heifer for each of the 14 breeds known to be in the data set (using the 35K array data only). Previously, a separate large data set consisting of only purebred cattle was used to estimate SNP effects for breed composition. A GBLUP model was fitted, where the phenotype was 1 if the animal was of that breed and 0 if not (Dodds et al. 2014). The effects of each SNP for the proportion of each breed was then derived by back-solving for the SNP effects (Yang et al. 2011), and the resulting prediction equations for each breed were used to estimate breed proportions in the heifers.

Then the model fitted to the Heifer puberty, height, body condition score, weight and adaptation trait data was

$$\boldsymbol{y = \mu + cohort + year + het + breedprop + animal + error}$$

Where: **y** is a vector of trait records (Heifer puberty, weight, hip height or body condition score, temperament score, tick score, buffalo fly lesion score); μ is the population mean; **cohort** is the property+yeardrop+paddock that the heifers were in prior to mustering for trait recording; year is the

year of recording; **het** is the heterozygosity of each heifer as measured by the proportion of markers that were heterozygous (to capture heterosis effects) fitted as a liner effect; **breedprop** is a series of 15 covariates (14 breeds and *Bos indicus* content), measuring the proportion of each breed in the heifers as described above; **animal** is a vector of random effects ~ N(0, $\mathbf{G}\sigma_g^2$), with G the genomic relationship matrix among all heifers (Yang et al. 2011) and σ_g^2 the genetic variance captured by the SNP markers; **error** is a vector of random deviations ~ N(0, $\mathbf{I}\sigma_e^2$). There were at least 50 animals in each contemporary group. To report heterosis, which ranges in the data from marker heterozygosity ~ 0.25 (purebred) to 0.5 (F1), we have re-scaled the estimate to be on the scale of purebred to F1 (eg. ¼ of the actual estimate of marker heterozygosity which is scaled 0 to 1). Variance components were estimated in GCTA (Yang et al. 2011), and the heritability of the traits (actually the proportion of phenotypic variance captured by the SNP) was estimated as $h^2=\widehat{\sigma_g^2}/(\widehat{\sigma_g^2}+\widehat{\sigma_e^2})$.

The accuracy of GEBV was evaluated by dropping out 10 of the 2018 cohorts (the latest birth year of heifers for which data was captured) at random. The breed composition within the 10 herds (3790 heifers) ranged from purebred Brahman to crossbreds of Bos taurus breeds. There were 24,561 heifers in the reference population. GEBV were predicted for the heifers in the 10 excluded herds, then the GEBV were correlated with the actual phenotypes (adjusted for fixed effects) of the heifers within each herd. This correlation was divided by the square root of the heritability of the trait to get the accuracy of genomic prediction. We also assessed accuracy of genomic prediction in a completely independent data set, the Beef CRC cohort of 894 and 1088 Brahman and Tropical composite cattle (respectively). These cattle were phenotyped for a different fertility trait than we have used, age at first corpus luteum, assessed by ultrasound scanning at six-week intervals (Johnston et al. 2013). Genotypes for these cattle were imputed to the same 709,768 SNP used in the genomic predictions.

For multi-breed evaluations, head-to-head comparisons of breeds in the same herd/environment are necessary. We assessed how many head-to-head comparisons as $\sum_i^n X_i' X_i$ where for each of *n* cohorts in the data set, X_i is a matrix of breed proportions in cohort *i*, of dimensions number of heifers in the cohort x number of breeds.

3.2 Develop a low cost SNP array including i. SNP that improve accuracy of genomic breeding values for breeds and composites used in Northern beef production ii. SNP that predict polled/horned iii. SNP that can be used for pedigree reconstruction in these breeds/composites, iv. SNP that predict breed composition, and v. Any recessive defects

The first step in developing the new array was to generate a resource of whole genome sequence data representing the genomes of the Northern Australian beef herd. The most cost effective cattle to sequence are key ancestors, that is bulls that have left behind many progeny, grand progeny, and so on, and have therefore made a major contribution to the gene pool of the breeds used in Northern Australian beef production. The 1000 bull genomes project (Hayes and Daetwyler et al. 2019), of which UQ is key member, already includes a number of key sires from tropical beef breeds, particularly Brahman and Tropical Composites. To complement the already sequenced sires, we chose 200 sires for whole genome sequencing that were not in the 1000 bull genomes project, with 150 Brahman, 25 Droughtmaster and 25 Santa Gertrudis sires selected, given these are key breeds for beef production in Northern Australia.

The first step in choosing these sires was to identify from pedigree those that had many sons, and in particular many sons being used widely as bulls, in each breed. This was done in close collaboration with Tropical Beef Technical Services (TBTS) and DAF staff. Then within each breed a "Sequencing the Legends Project" was initiated, to encourage studs to donate semen straws from the key sires. The "Sequencing the Legends" project were run together with the breed societies or with key studs within each breed. Dr Brian Burns was a key driver of these projects. The final list sires for sequencing were chosen from the available samples using an algorithm that uses the pedigree within each breed to identify the bulls in each breed that have made the largest contribution, but avoids double counting genomes (that is if a sire is chosen his sons are unlikely to be chosen because sequencing his sons will just re-sequence half the sires genome). The algorithm we used is described in Daetwyler et al. (2014) and Druet et al. (2014).

Each sire was sequenced to 10 fold coverage (eg 10 times coverage of the genome) using Illumina paired end sequencing, and sent to the 1000 bulls project. In the 1000 bull genomes pipeline (Daetweyler et al. 2014), raw re-sequence reads were filtered based on chastity score and trimmed based on quality scores. Fastq files were then aligned to the reference bovine genome assembly ARS1.1 using the BWA program (Li and Durbin 2009). Variants (SNP and short insertions and deletions) were called using Samtools-0.1.18 mpileup (Li et al. 2009). BAM files were pooled for variant calling. Variants were then removed if they had: two or more alternative alleles, no observations of the alternative allele on either the forward or reverse reads, an overall quality score (QUAL) of <20, a mapping quality (MQ) score of <30, a read depth of <10 or more than median plus 3 SD read depth, >0.1 opposing homozygotes, or the same bp position (e.g. SNP overlapping with indel). In addition, we applied the following proximity filters: when an indel was within 10 bp of another indel, the indel with a lower QUAL score was removed; when any variant was within 3 bp of another variant, the lower QUAL variant was removed. The filters were implemented by extending the python vcf file parser PyVCF. The phred score genotype probabilities were converted to true probabilities and BEAGLE (Browning and Browning 2009). As a quality control step, concordance of re-sequence genotypes with Bovine HD chip genotypes was calculated as the proportion of identical genotypes pre and post the BEAGLE step. Opposing homozygotes were calculated as the proportion of non-matching homozygotes between parent-offspring pairs and collected per pair and per locus. Additional filtering was conducted to select a set of SNP that were polymorphic in tropical beef animals. This resulted in 48.8 million SNP and small insertions/deletions being identified.

The high density genotypes for the 26,000 project heifers were then imputed to the whole genome sequence data using the 1000 bull genomes reference panel (Run 8) (Hayes and Daetwyler 2020) which consists of 2703 cattle with genotypes for the 48.8 million called SNP and small insertion/deletions. The 2703 animals included several hundred Brahman, Droughtmaster and Santa Gertrudis founder animals sequenced as described above. Beagle 5.2 (Browning et al. 2021) was used for the imputation.

Then to identify which SNP to include on the new SNP panel, we performed a genome wide association study for each trait. Each SNP was tested in turn for association with the three traits. The model fitted for each trait, 48.8 million times (once for each SNP) was

$$\boldsymbol{y = \mu + SNP + cohort + heterozygosity + bos\ indicus\ proportion + animal + e}$$

Where terms are as described above, and SNP is a covariate 0=homozygous reference allele, 1=heterozygous, 2=homozygous alternate allele. The *mlma* package in the GCTA program (Yang et

al. 2011) was used to fit the above model. The P-value and SNP effect were recorded. A stringent P value of 5x10-8 was used to identify significant SNP, as this value accounts for multiple testing of so many SNP and the inheritance of the genome in independent chromosome chunks. SNP with P-values below the threshold were incorporated in the new design. The new SNP were added to the Neogen trop beef chip, which already includes SNP that track horned/polled, parentage, and Pompes disease.

3.3 Develop novel ways of presenting genomic information that resonate and are meaningful to commercial producers in Northern Australia, to accelerate adoption of the use of genomic information in bull selection decisions.

The GBV for each herd were sent back to each individual animal and were sent back to the herd owners, together with estimated breed composition and polled/horned status (as defined by Neogen, eg homozygous horned, HP and PP for horned/horned, heterozygous and homozygous polled respectively). After extensive consultation with owners at field days, and individually by phone and by email to discuss how best to present results, herd owners have also received a custom report detailing where each of their heifer cohorts benchmarks in relation to the overall population (proportion of animals in top 20%, proportion of animals in bottom 20%) and so on for each trait. An example report is given in Appendix 1. A webinar taking collaborators through GBV results and getting feedback was held in August 2021.

4. Results

4.1 Develop methodology for calculating accurate genomic breeding value for key traits affecting productivity in Northern Beef cattle

Heterosis as estimated from the marker data had a positive effect on all traits, increasing height, weight, body condition and a substantial effect on the chance of being pubertal at scanning, Table 2. These estimates are in line with those from other studies (eg. Frisch and Vercoe 1984).

Table 2. Heterosis effect (scaled purebred to F1 of *Bos indicus* x *Bos taurus*; trait range for reference in body condition score is 1 to 5, for heifer puberty 0 to 1).

Trait	Heterosis
Weight	40.5 kg
Hip height	43.8 mm
Body condition score	0.25 units
Heifer Puberty	0.27 units

The heritability of the traits estimated from the genomic data was moderate for most traits, and low for P4M and Buffalo fly lesion, Table 3.

Table 3. Trait genomic heritabilities and standard errors

Trait	Heritability	Standard error
Weight	0.29	0.01
Hip height	0.39	0.01
Body condition score	0.22	0.01
Heifer Puberty	0.22	0.01
Pregnant 4 months after calving	0.11	0.02
Temperament	0.37	0.03
Tick score	0.33	0.04
Buffalo fly lesion score	0.14	0.02

Heritabilities were consistent with previous estimates for these traits in tropical beef cattle data sets derived from pedigree (eg Johnston et al. 2013, Corbet et al. 2017), and SNP genotypes (including for temperament, tick score and buffalo fly lesion score (eg. Porto-Neto et al. 2014). The fact that heterosis estimates and heritability estimates from this project align closely with previous estimates suggest the data in this project is of high quality, and suitable for GBV prediction.

Across the Northern Genomics animals, there was a substantial range in GEBV for each trait, Table 4.

Table 4. Spread of quintiles for each GEBV in the Northern Genomics data set.

Likely to be		V low 1	Low 2	Moderate 3	High 4	V high 5
pubertal		-39% to -13%	-13% to -4%	-4% to +4%	+4% to +13%	+13% to +39%
P4M		-10% to -3%	-3% to -1%	-1% to +1%	+1% to +3%	+3% to +10%
Docile*	score	+0.12 to +0.35	+0.04 to +0.12	-0.04 to +0.04	-0.12 to -0.04	-0.35 to -0.12
heavier	Kg	-31 to -10	-10 to -3	-3 to +3	+3 to +10	+10 to +31
Taller	mm	-64 to -22	-22 to -7	-7 to +7	+7 to +22	+22 to +64
BCS	score	+0.24 to +0.08	+0.08 to +0.02	+0.02 to -0.02	-0.02 to -0.08	-0.08 to -0.24
tick res**	score	+0.09 to +0.03	+0.03 to +0.01	+0.01 to -0.01	-0.01 to -0.03	-0.03 to -0.09
fly res***	score	+0.06 to +0.02	+0.02 to +0.01	+0.01 to -0.01	-0.01 to -0.02	-0.02 to -0.06

* More negative values indicate more docile cattle

** More negative values indicate more tick resistant animals

*** More negative values indicate more fly lesion resistant animals

The accuracy of the GBV is cross validation was moderate, and consistent across the 10 validation herds (as indicated by the low standard error on the accuracy, Table 5. These accuracies are useful for selecting teams of bulls, or culling heifers which are unlikely to perform.

Table 5. Accuracy of genomic predictions (across 10 herds).

Trait	Accuracy of genomic prediction
Weight	0.31±0.04
Hip height	0.43±0.04
Body condition score	0.33±0.04
Heifer Puberty	0.42±0.05

The GBV also predicted performance in the Beef CRC data quite well, even though the traits were not exactly the same, Table 6. The closest trait in the beef CRC data to our Heifer puberty is age at first corpus luteum (AGECL), and these traits have been shown in the past to have a genetic correlation of -0.8 (heifers more likely to be pubertal have a lower age at first calving, hence the negative sign). The closest trait in the Beef CRC data to our P4M is post partum anoestrus interval (PPAI). For tick score, these was little to no variation for scores in the Brahman data, hence the low correlation.

Table 6. Prediction of beef CRC traits from Northern Genomics GBV

	Number	AGECL-CL600	PPAI-P4M	Tick score-Tick GBV
Brahman	894	-0.45	-0.17	0.02
Tropcomps	1088	-0.21	-0.52	0.19

The number of effective head-to-head comparisons possible from the data set, which enables estimates of breed effect, was reasonable for Angus versus Brahman, Brahman versus Droughtmaster and Brahman versus Santa Gertrudis, but was lower for other breed combinations, Table 7.

Table 7. Number of effective head-to-head breed comparisons in the data set, where each cell represents the number of genomes for a breed being compared to the number of genomes of the other breed. Empty cells indicate no comparisons for that breed combination. Only breeds with substantial numbers are shown (out of 14 breeds).

	Angus	Belmont Red	Brahman	Charolais	Drought master	Hereford	Limousin	Santa Gertrudis	Shorthorn	Wagyu
Angus										
Belmont Red	7									
Brahman	337	24								
Charolais	20	2	153							
Droughtmaster	123	15	924	44						
Hereford	47	4	76	12	58					
Limousin	21	3	97	13	33	13				
Santa Gertrudis	111	7	421	31	208	35	27			
Shorthorn	33	3	73	10	61	13	10	44		
Wagyu	3	0	14	2	16	2	3	8	2	
Boran	2	4	33	1	7	1	1	3	2	0

4.2 Develop a low cost SNP array

The accuracy of imputation to whole genome sequence genotypes was high, close to 0.90, for the majority of SNP, but lower at low minor allele frequencies, Figure 2. This high accuracy of imputation

indicates a GWAS can be conducted on the imputed sequence genotypes with confidence (eg Daetwyler et al. 2014).

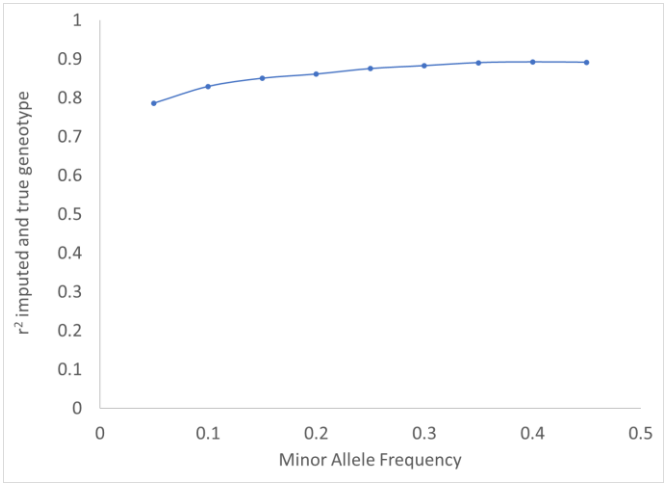
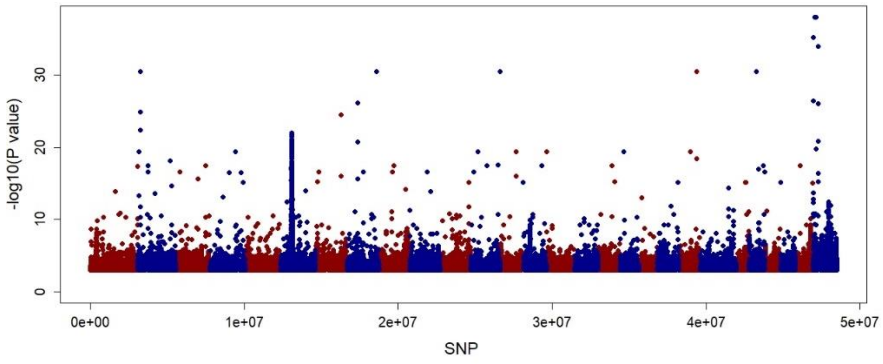


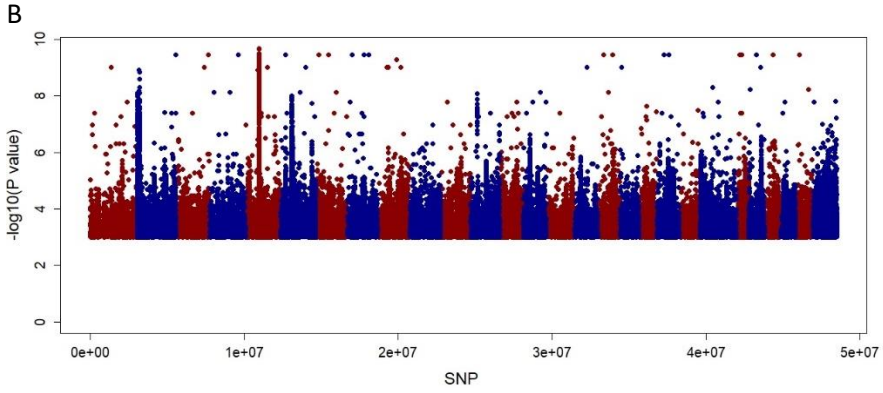
Figure 2. Accuracy of imputation (correlation squared of true and imputed genotypes) to whole genome sequence data for the northern genomics heifers using the 1000 bull genomes reference panel.

There were a large number of significant SNP for weight, height, body condition score, heifer puberty, and tick score (Pvalue<$5x10^{-8}$), Figure 3. There were no significant SNP for temperament, buffalo fly lesion score or P4M. The significant SNP clustered in sites across the genome, indicating a number of SNPs were tracking the same causative mutation, Figure 3.

A



B



C

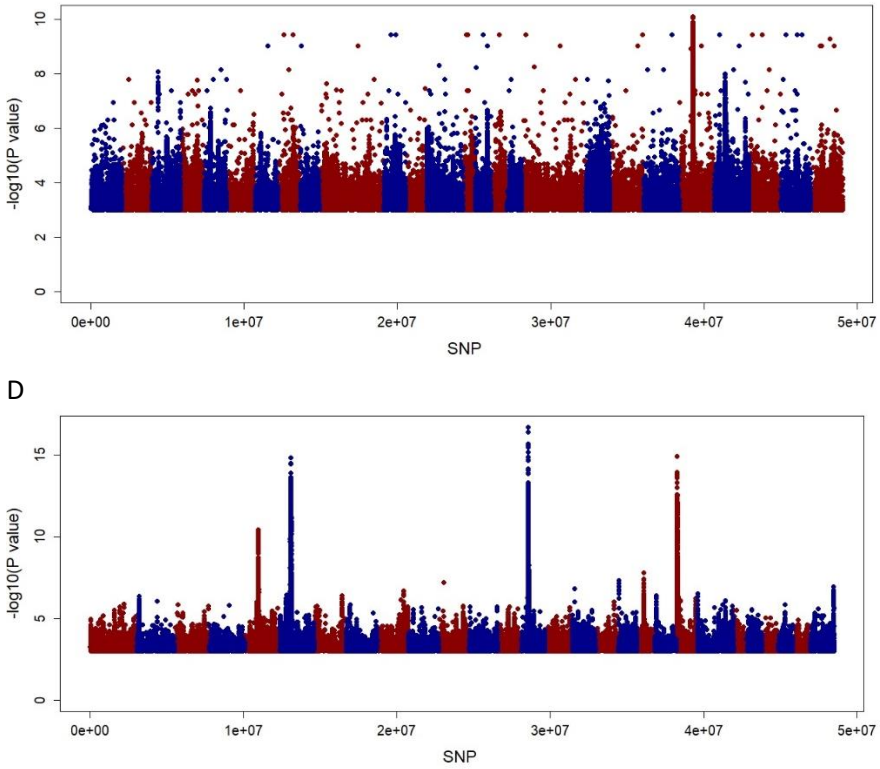


Figure 3. Genome wide association with 48.8 million SNP for weight (A), height (B), body condition score (C) and heifer puberty (D). Odd chromosomes are coloured blue, even chromosomes in red. The significance threshold on the y-axis is 8.3 (eg -log10(5x10^{-8}). Only SNPs with significance > 3 (P value<0.0001) are shown, otherwise there are too many SNP associations to plot.

Heifer puberty. There were four genome locations with significant SNP for Heifer puberty, Figure 3D and Table 8. When we investigated the genes closest to the most significant SNP, they included PLAG1, HMGA2, LCORL and SNRP, Table 8.

Table 8. Genome location, -log10(Pvalue) and closest gene for most significant SNP in each genome location for Heifer Puberty.

Chromosome	Location	-LOG$_{10}$(Pvalue)	Nearest Gene
5	47847003	11.32	HMGA2
6	37998696	14.82	LCORL
14	23300304	16.81	PLAG1
21	1970429	15.29	SNRPN

PLAG1 has been extensively reported to harbour mutations affecting both puberty traits and height, in cattle and many other mammals (Fortes et al. 2012, Bouwman et al 2018). In fact, the mutation we report here, 14: 23300304 G->A, is exactly the same as the one Bouwman et al. (2018) identified in a very large meta-analysis of height in cattle using imputed sequence data on more than 50k animals. Engle et al. (2022) demonstrated that this PLAG1 mutation affects heifer fertility traits, such as heifer days to calving and age at first calving, but has less effect on fertility later in life (days to calving for later lactations).

HMGA2 modulates the expression of PLAG1, which in turn affects the expression of IGF2, a key regulator of fetal growth. Mutations in HMGA2 and PLAG1 in humans and mice reduce early fetal growth, Habib et al. (2018). Mutations in HMGA2 were identified affecting cattle height (Bouwman et al. 2018).

LCORL has also previously been reported to harbour a mutation affecting height in cattle (Bouwman et al. 2018), and LCORL knockout mice have reduced fertility, supporting this gene as the gene affected by a mutation in this region. LCORL is also involved in fetal growth regulation.

There have been fewer reports on the SNRPN gene and it's effects on cattle traits, highlighting the power of this GWAS to reveal new chromosome locations affecting fertility.

Body condition score. There were a greater number of significant SNP for body condition score than for heifer puberty, but fewer than for weight, Figure 3C, Table 9. Some of the genome locations for significant body condition score SNP overlapped with those for heifer puberty, for example chromosome 5, and are likely to be driven by the same mutations (eg in HMGA2).

Table 9. Genome location, -log10(Pvalue) for most significant SNP in each genome location for body condition score.

Chromosome	Location	$-LOG_{10}$(Pvalue)
1	67247976	9.02
3	1.17E+08	9.46
4	99992127	9.46
5	47507791	10.05
6	16601349	9.46
7	9631765	9.46
7	43119397	9.46
8	16036836	9.46
8	59710232	9.46
9	27639619	9.02
9	58911044	9.29
9	75163207	9.02
16	39856168	9.02
17	17842491	9.46
17	45793858	9.46
18	3860734	9.02
20	24636534	9.46
20	41650474	9.46
22	42556083	8.32
26	24263988	9.46
26	39499084	9.02
27	23823737	9.46
29	11006590	9.46

Weight. Weight has the largest number of significant SNP, Figure 3A, Table 10. Some of the significant SNP clusters overlapped with Heifer puberty genome locations, and body condition score locations. The most significant SNP was on chromosome X. A gene very close to the most significant SNP, E3 ubiquitin ligase, was demonstrated to control body weight in knockout mice (Fujita et al 2015).

Table 10. Genome location, -log10(Pvalue) for most significant SNP in each genome location for weight.

Chromosome	Location	-LOG$_{10}$(Pvalue)
1	157712614	17.39
3	109004454	17.39
4	67202065	16.52
4	89340645	19.36
4	113711550	15.17
6	37266497	22.09
7	6182187	15.21
7	93366623	24.50
8	34531113	26.19
8	56276351	16.52
8	104400448	30.47
9	48731995	17.42
10	57545477	16.52
11	102366214	15.11
12	22989237	19.36
12	55566723	17.39
12	81850910	17.58
12	87045475	30.47
13	57556143	19.36
13	59061278	15.98
14	367491	15.11
14	69013922	17.39
15	621113	19.36
17	42353770	17.39
17	51396538	15.17
18	10966395	19.36
20	68647130	15.11
21	37871054	19.36
21	58740827	30.47
24	34047722	14.37
25	31004348	15.11
25	31065402	15.11
26	23660967	30.47
26	47542560	17.39
28	971221	15.11
29	13165610	17.39
30	2853494	35.16
30	8613303	38.02
30	15277634	38.02
30	30266970	33.94

Tick score. There was one significant region on chromosome 7 associated with tick score, Figure 4, Table 11.

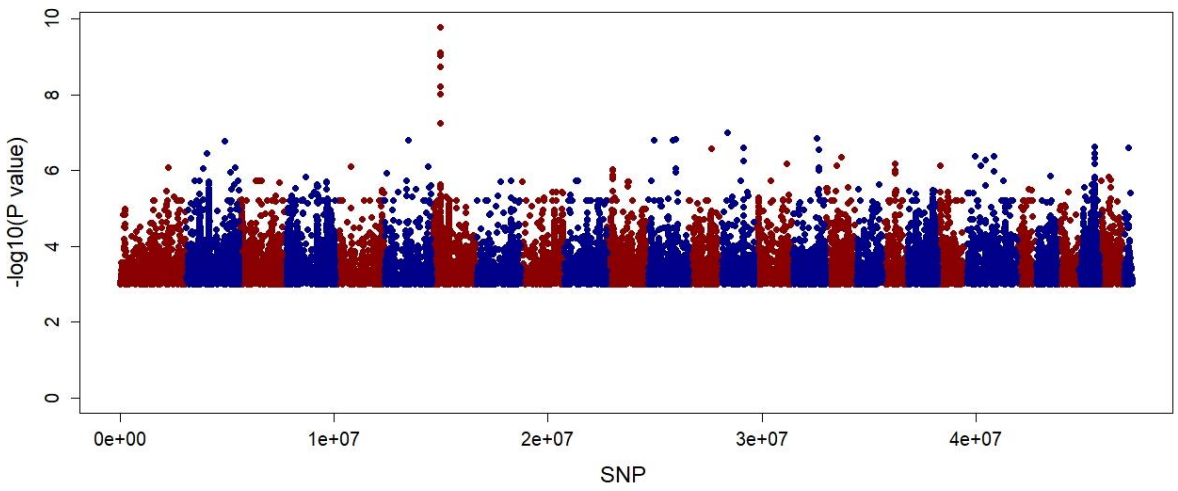


Figure 4. Genome wide association study for tick count. Odd chromosomes are coloured blue, even chromosomes in red. The significance threshold on the y-axis is 8.3 (eg -log10(5x10^{-8}). Only SNPs with significance > 3 (P value<0.0001) are shown, otherwise there are too many SNP associations to plot.

Table 11. Genome location, -log10(Pvalue) for most significant SNP in each genome location for tick score.

Chromosome	Location	-LOG$_{10}$(Pvalue)
7	15582076	9.10092
7	15606229	9.03386
7	15633281	9.78161
7	15676107	8.74595

Genes in this region include ANGPTL8, DOCK6, SPC24 and KANK2. KANK2 is a good candidate, as mutations in this gene have been associated with skin thickness in people (Ramot et al. 2014).

Pleiotropic effects of SNPs highly associated with heifer puberty. We investigated the effect of the most significant SNP in the four genome regions affecting heifer puberty on weight, body condition score, and height, Table 12. For the regions on chromosome 5,6 and 14, the allele that increased the probability of a heifer being pubertal at six hundred days decreased weight and height, and increased body condition score. These effects were highly significant. The SNP on chromosome 21 did not significantly affect any of the other three traits.

Table 12. Effects of SNPs highly associated with heifer puberty on weight, body condition score and height.

Chromosome	Location	Heifer Puberty	Weight	Body condition score	Height
5	47847003	**0.06**	**-2.70**	**0.023**	**-7.33**
6	37998696	**-0.04**	**3.91**	**-0.021**	**6.37**
14	23300304	**0.06**	**-3.49**	**0.017**	**-6.67**
21	1970429	**0.05**	0.25	0.001	-0.52

The pattern of effects for the SNP on chromosome 5,6 and 14 very likely reflects the effect of the genes on fetal and early growth. Even at this stage of development, it appears there is some competition for resources for growth and reproductive development. The identification of patterns of effects across traits gives some indication of selection responses once these SNP are included on commercially used arrays. Identification of new regions, such as that on chromosome 21, where

selection for this SNP could improve fertility without affecting weight, offer new opportunities for selection.

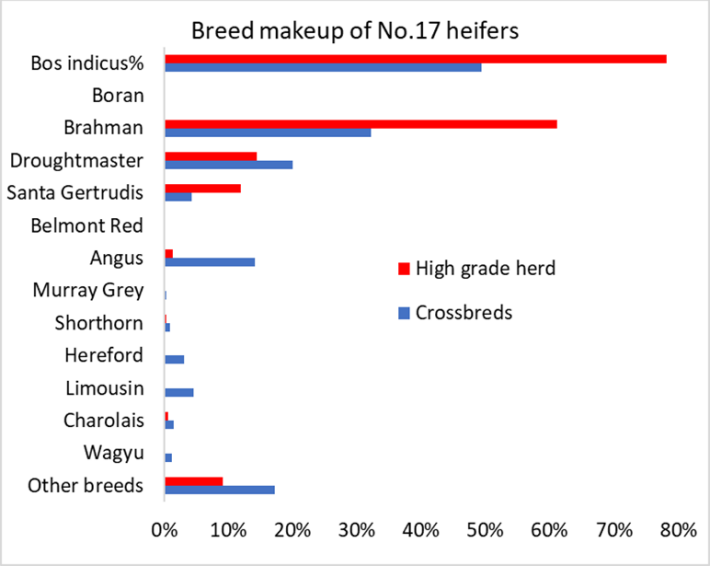
We have been in communication with SNP genotyping providers, particularly Neogen, regards to outcomes for the project, and the SNP identified in this milestone report will be included in the next round of array design.

4.3 Develop novel ways of presenting genomic information that resonate and are meaningful to commercial producers in Northern Australia, to accelerate adoption of the use of genomic information in bull selection decisions.

Following extensive dialogue with collaborating herds on the best way to present the GBV information to enable rapid decision making and benchmarking, a quintile approach was adopted, where each animal was identified as very good (top quintile for the whole population), good (in the best 20-40% of animals), moderate (in 40%-60% range), poor (bottom 60-80%), and very poor (bottom 80%).

All herd owners have received a spreadsheet of GBV, quintile classification for animal, and breed composition for each of their animals. In addition, herd owners have received a customised report benchmarking their herds' performance for all GEBV traits against the Northern Genomics population. In this report, they also received a summary of the breed composition of their herd. An example of the graphics in the report for one of the project collaborators is given in Figure 5.

A



B

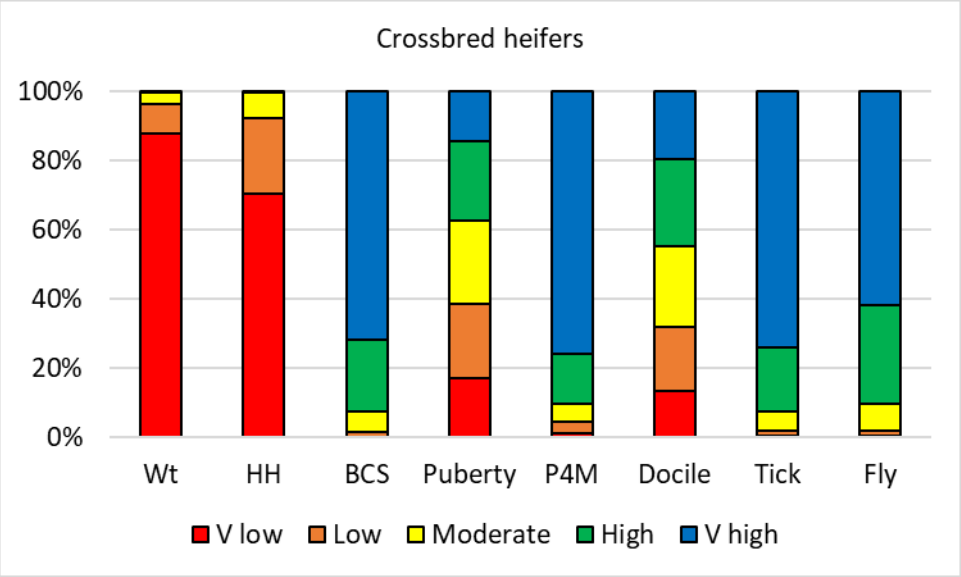


Figure 5. For a particular collaborators first cohort of heifers (n=257), A. breed proportions, and B. proportion of heifers in lowest quintile (v low) to highest quintile (v high), benchmarked against the whole Northern Genomics population of genotyped animals (y axis, where proportion in each quintile is 20%).

The full report outlines possible decisions that can be made with the information, Appendix 8.1. Project participants, along with all their GBV for all animals tested and an individual report, also received a copy of the "Hair of the cow" newsletter, which reported on project progress, profiled collaborators, and reminded project participants of key times for data collection.

5. Conclusion

This project has produced two products that will enable northern Australian beef enterprises to improve fertility and other key traits in their herds. The GBV from the project have useful accuracy for selecting teams of bulls, or for culling heifers that are unlikely to perform, across a wide range of breeds, crossbreds and composites. The GBV for a herd can also be summarised into a herd profile, that benchmarks the herd and can be used as a tool to identify which traits should be the focus in bull selection decisions.

The project has also identified a panel of 512 SNP from whole genome sequence data that are highly associated with fertility, growth and adaptation traits. These SNP will be incorporated into future SNP arrays used for commercial genotyping, and will increase both the accuracy of GBV from this project, and also BREEDPLAN single step EBV for Northern breeds.

Working with project collaborators (the 54 herds), new ways of presenting GBV information were developed that should enable rapid, accurate decision making. However, this process did highlight how challenging it is to convey such large volumes of information. This is an area for future development, perhaps with experts in conveying complex information visually (,graphic designers, artists for example?).

Finally, the project has demonstrated that moderate accuracies of multi-breed genomic prediction can be achieved from reference sets derived from large scale commercial data, provided the herd owners involved are well informed with data collection protocols. Genomic heritabilities from the project were consistent with previous estimates for these traits in tropical beef cattle data sets derived from pedigree, as were heterotic effects. The challenges of unknown breed composition and level of

heterosis in commercial data can be solved using estimates of these parameters obtained directly from the genotype data. Using crossbred data allows head-to-head comparison of chromosome segments derived from different breeds, disentangling breed effects from contemporary group effects, at least to some extent. The data set described here will contribute to multi-breed genomic evaluations for many, but not all breeds used in northern Australia.

5.1 Key findings

- Genomic heritabilities and estimates of heterosis from the data were consistent with previous studies, suggesting the data is of good quality. The methodology for estimating heterosis could be further validated in a population of structured crosses. The project has demonstrated that moderate accuracies of multi-breed genomic prediction can be achieved from reference sets derived from large scale commercial data, provided the herd owners involved are well informed with data collection protocols.
- The challenges of unknown breed composition and level of heterosis in commercial data can be solved using estimates of these parameters obtained directly from the genotype data. Using crossbred data allows head-to-head comparison of chromosome segments derived from different breeds, disentangling breed effects from contemporary group effects, at least to some extent.
- The GBV derived from the data were of useful accuracy for bull team selection and heifer culling (0.3-0.45) in a validation with independent herds with a range of breeds, crossbreds and composites, and the GBV also appropriately predicted related traits in the Beef CRC data (GBV for heifer puberty were correlated with age at first CL, GBV for P4M were correlated with post partum anoestrus interval).
- Working with project collaborators (the 54 herds), new ways of presenting GBV information were developed that should enable rapid, accurate decision making. However this process did highlight how challenging it is to convey such large volumes of information. This is an area for future development, perhaps with experts in conveying complex information visually (graphic designers, artists for example?).

5.2 Benefits to industry

The prediction equations from this project can be used to calculate GBV for young bulls and heifers for fertility, growth and adaptation traits across a wide range of breeds, crossbreds and composites, enabling selection for these traits on an industry wide scale. The GBV for a herd can also be used to benchmark the genetic level of the herd, identifying areas of focus to improve profitability through bull selection in the future.

6. Future research and recommendations

This technology should be commercialised and rolled out, together with case studies of beef enterprises actually using the GBV to demonstrate their value, to the Northern Australian beef industry. The challenge of presenting GBV in a useful way that enable rapid selection decisions was addressed to some extent in this project, and has been the focus in other MLA projects, but continuing effort in this area, particularly at the commercial level is warranted. Incorporation of the data into BREEDPLAN to improve accuracy of single step EBV has been discussed with ABRI and AGBU (Steve Skinner, Brad Crook, Steve Miller, David Johnston) and appropriate models identified, and transfer of the data is underway. Data for purebred herds has already been transferred.

7. References

Boerner V, Johnston DJ, Tier B. Accuracies of genomically estimated breeding values from pure-breed and across-breed predictions in Australian beef cattle. *Genet Sel Evol*. 2014 Oct 24;46:61.

Boichard D, Chung H, Dassonneville R, David X, Eggen A, Fritz S, Gietzen KJ, Hayes BJ, Lawley CT, Sonstegard TS, Van Tassell CP, VanRaden PM, Viaud-Martinez KA, Wiggans GR; Bovine LD Consortium. Design of a bovine low-density SNP array optimized for imputation*. PLoS One.* 2012;7(3):e34130.

Bolormaa, S., Pryce, J.E., Kemper, K., Savin, K., Hayes, B.J., Barendse, W., Zhang, Y., Reich, C.M., Mason, B.A., Bunch, R.J., Harrison, B.E., Reverter, A., Herd, R.M., Tier, B., Graser, H.-U., Goddard, M.E. 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 (2013) *Journal of Animal Science,* 91 (7), pp. 3088-3104*.*

Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, Sahana G, Govignon-Gion A, Boitard S, Dolezal M, Pausch H, Brøndum RF, Bowman PJ, Thomsen B, Guldbrandtsen B, Lund MS, Servin B, Garrick DJ, Reecy J, Vilkki J, Bagnato A, Wang M, Hoff JL, Schnabel RD, Taylor JF, Vinkhuyzen AAE, Panitz F, Bendixen C, Holm LE, Gredler B, Hozé C, Boussaha M, Sanchez MP, Rocha D, Capitan A, Tribout T, Barbat A, Croiseau P, Drögemüller C, Jagannathan V, Vander Jagt C, Crowley JJ, Bieber A, Purfield DC, Berry DP, Emmerling R, Götz KU, Frischknecht M, Russ I, Sölkner J, Van Tassell CP, Fries R, Stothard P, Veerkamp RF, Boichard D, Goddard ME, Hayes BJ. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet.* 2018 Mar;50(3):362-367.

Browning BL, X Tian, Y Zhou, and S R Browning (2021) Fast two-stage phasing of large-scale sequence data*. Am J Hum Genet* 108(10):1880-1890.

Burns, B. M., (2016). Queensland Government Smart Futures Research Partnerships Program (2012-2015). University of Queensland.

Corbet, N. J., J. M. Allen, A. R. Laing, G. Fordyce, M. R. McGowan, and B. M. Burns. (2017). Anim. Prod. Science. doi.org/10.1071/AN16616Dodds KG, Auvray B, Newman SA, McEwan JC. (2014) *BMC Genetics.* 16:15:92.Fordyce G (2012) Beef CRC exit Report.

Daetwyler HD, Capitan A, Pausch H, Stothard P, Van Binsbergen R, Brøndum RF, Liao X, Djari A, Rodriguez SC, Grohs C, Esquerré D, Bouchez O, Rossignol M-N, Klopp C, Rocha D, Fritz S, Eggen A, Bowman PJ, Coote D, Chamberlain AJ, Anderson C, Vantassell CP, Hulsegge I, Goddard ME, Guldbrandtsen B, Lund MS, Veerkamp RF, Boichard DA, Fries R, Hayes BJ. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics,* 46:858-865.

de Camargo GM, Porto-Neto LR, Kelly MJ, Bunch RJ, McWilliam SM, Tonhati H, Lehnert SA, Fortes MR, Moore SS. Non-synonymous mutations mapped to chromosome X associated with andrological and growth traits in beef cattle*. BMC Genomics.* 2015 16:384.

Engle, B, Hayes BJ (2022) Genetic variation in PLAG1 is associated with early fertility in Australian Brahman cattle. *Journal of Animal Science,* submitted.

Fortes MR, Kemper K, Sasazaki S, Reverter A, Pryce JE, Barendse W, Bunch R, McCulloch R, Harrison B, Bolormaa S, Zhang YD, Hawken RJ, Goddard ME, Lehnert SA. Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. *Anim Genet.* 2013 Dec;44(6):636-47. doi: 10.1111/age.12075.

Fujita H, Yagishita N, Aratani S, et al. The E3 ligase synoviolin controls body weight and mitochondrial biogenesis through negative regulation of PGC-1β*. EMBO J.* 2015;34(8):1042-1055. doi:10.15252/embj.201489897.

Habib Abi W, Brioude F, Edouard T, Bennett JT, Lienhardt-Roussie A, Tixier F, Salem J, Yuen T, Azzi S, Le Bouc Y, Harbison MD, Netchine I. Genetic disruption of the oncogenic HMGA2-PLAG1-IGF2 pathway causes fetal growth restriction. *Genet Med.* 2018 Feb;20(2):250-258.

Harbug, S and Lean, I (2018) ECONOMIC CONSEQUENCES OF SELECTION FOR COW FERTILITY IN NORTHERN AUSTRALIAN BEEF HERDS*. BushAgribusiness report.*

Johnston, D. J., Barwick. S.A. Fordyce G, Holroyd R,G, Williams P.J. Corbet N.J. Grant T. (2013). *Anim. Prod. Science* 54:1-15.

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. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions (2015) *Genetics Selection Evolution*, 47 (1), art. no. 29.

Moghaddar N, Khansefid M, van der Werf JHJ, Bolormaa S, Duijvesteijn N, Clark SA, Swan AA, Daetwyler HD, MacLeod IM. Genomic prediction based on selected variants from imputed whole-genome sequence data in Australian sheep populations*. Genet Sel Evol.* 2019 Dec 5;51(1):72. doi: 10.1186/s12711-019-0514-2.

Ramot Y, Molho-Pessach V, Meir T, Alper-Pinus R, Siam I, Tams S, Babay S, Zlotogorski A. Mutation in KANK2, encoding a sequestering protein for steroidreceptor coactivators, causes keratoderma and woolly hair. *J Med Genet.* 2014Jun;51(6):388-94. doi: 10.1136/jmedgenet-2014-102346.

Sargolzaei, M., Chesnais J.P. Schenkel F.S. (2014)*/ BMC Genomics.* 15:478.

Xiang R, Berg IVD, MacLeod IM, Hayes BJ, Prowse-Wilkins CP, Wang M, Bolormaa S, Liu Z, Rochfort SJ, Reich CM, Mason BA, Vander Jagt CJ, Daetwyler HD, Lund MS, Chamberlain AJ, Goddard ME. Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc Natl Acad Sci U S A.* 2019 Sep 24;116(39):19398-19408. doi: 10.1073/pnas.1904159116.

Yang J, Lee SH, Goddard ME, Visscher PM (2011) *Amer. Jour. Human. Genet.* 88:76-82.

8. Appendix

8.1 Example report delivered to each collaborating herd

Bovem Vallis Station

Genomic Summary Report from analysis in June 2021

Synopsis

Bovem Vallis has contributed phenotypes and DNA from 2 cohorts of heifers. All females have been genotyped, ie, ~50,000 SNPs have been read using the TropBeef chip. Genomic breeding values (GBVs) for P4M (pregnant or not 4 months after calving), puberty, temperament, weight, height, body condition and tick and fly lesion scores of all animals at approximately 18 months of age have been calculated.

The *Bos indicus* content of the Bovem Vallis cattle is around 50% and 80% for the crossbred heifers and high grade heifers, respectively. Of this genetic makeup, the predominate breed is Brahman with some Droughtmaster and Santa Gertrudis content.

Both the crossbreds and the high grade cattle have below par GBVs for height and growth, but are well adapted and have typical industry temperament. The high grade cattle are slower to reach puberty, but both herds have good ability to become pregnant during lactation.

Bovem Vallis has 1% of high grade heifers and 1% of crossbred heifers displaying elite genetics, ie, they are in the top 5% of all project cattle. These animals are well suited for breeding bulls to achieve genetic improvement.

The details in this report expand on the key points above. Initially there is a situation analysis of your herd to see where the genomics of your cattle sit within northern Australia. The next part is how to identify individual animals that may suit and not suit your breeding objectives.

Genomic Breeding Value (GBV) Targets

The table below presents the traits for which GBVs have been calculated (main traits shown in bold).

Trait Type	*Trait*	*Comment*	*Data #*	*Heritability*
Female fertility	**Heifer puberty***	Affects maiden pregnancy rate	27,707	22%
	Pregnant within 4 months of calving (P4M*)	Pregnant wet cows produce more live weight	8,477	11%
Growth	Weight at ~18 months	Similar to 600-day weight EBV*	24,528	29%
	Height at ~18 months	Tall cattle may have lower production	24,261	39%
Adaptation	Body condition (BCS*) at ~18 months	Adapted cattle have better condition	24,537	22%
	Temperament (Docility)	Measured after weaner training	3,234	37%

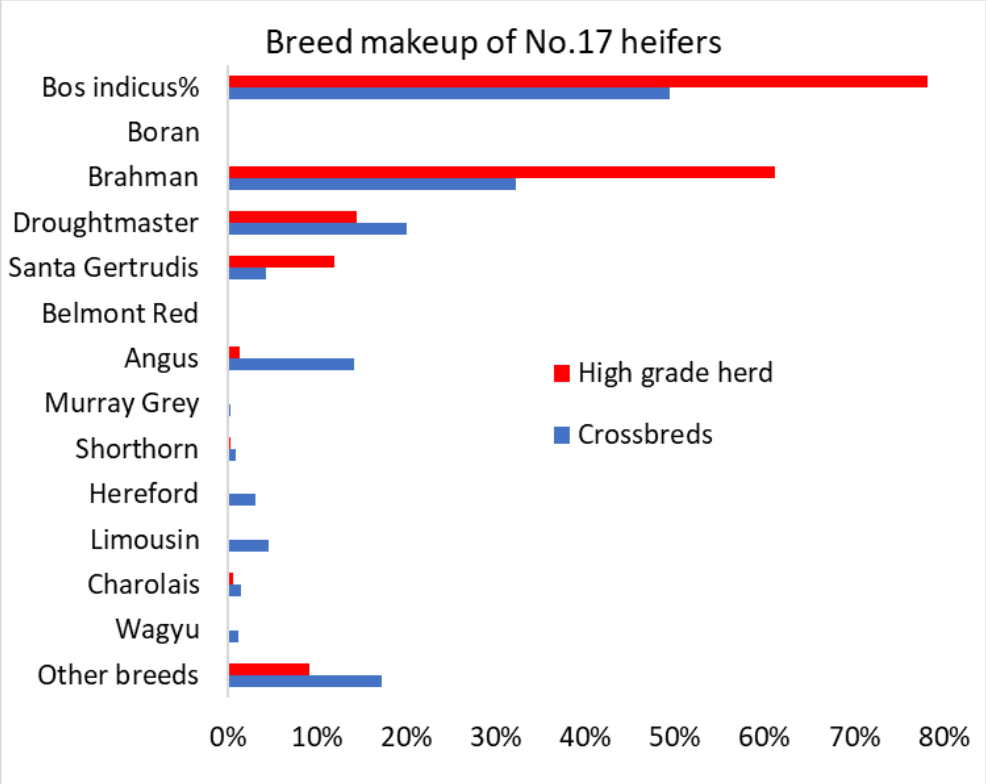
	Fly lesions	Affects market value	15,927	14%
	Tick score	Indicates tick resistance	2,094	33%

*See final page for definitions of terms. # In the recent analysis

How do the genomics of my herd look currently?

Breed makeup of the herd

The following graph shows the average breed makeup of the herd.
Note that breed prediction may have as much as 10% error for some breeds.



Where does your herd perform within northern Australia?

What's the range of GBVs in the northern Australian herd?

All 26,000 cattle included in the genomic analysis were divided into 5 equal-sized groups (called quintiles) on the basis of GBVs. The purpose of this is to give you a picture (situation analysis) for where your breeding is at present compared to the northern Australian cattle population (as represented by the 30,000 cattle in the project). These GBVs are accurate for a range of breeds and nutritional conditions. The spread of values in each quintile for each trait is described in the table below.

Spread of GBVs within each quintile for each trait

Likely to be		V low 1	Low 2	Moderate 3	High 4	V high 5
pubertal		-39% to -13%	-13% to -4%	-4% to +4%	+4% to +13%	+13% to +39%
P4M		-10% to -3%	-3% to -1%	-1% to +1%	+1% to +3%	+3% to +10%
Docile*	score	+0.12 to +0.35	+0.04 to +0.12	-0.04 to +0.04	-0.12 to -0.04	-0.35 to -0.12
heavier	kg	-31 to -10	-10 to -3	-3 to +3	+3 to +10	+10 to +31
taller	mm	-64 to -22	-22 to -7	-7 to +7	+7 to +22	+22 to +64
fatter	score	+0.24 to +0.08	+0.08 to +0.02	+0.02 to -0.02	-0.02 to -0.08	-0.08 to -0.24
tick res**	score	+0.09 to +0.03	+0.03 to +0.01	+0.01 to -0.01	-0.01 to -0.03	-0.03 to -0.09
fly res***	score	+0.06 to +0.02	+0.02 to +0.01	+0.01 to -0.01	-0.01 to -0.02	-0.02 to -0.06

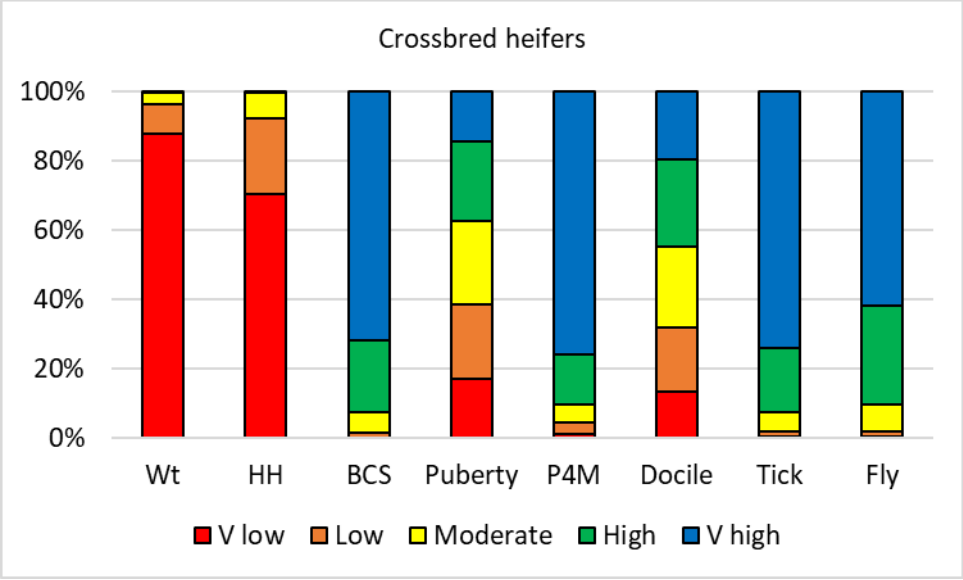
* More negative values indicate more docile cattle

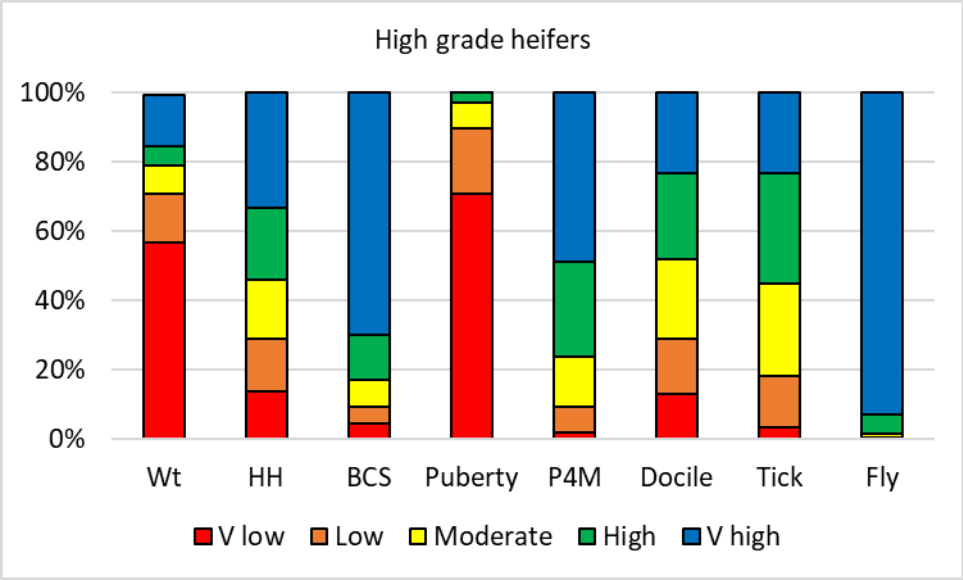
** More negative values indicate more tick resistant animals

*** More negative values indicate more fly lesion resistant animals

What is the range of GBVs in my herd?

The bars in the following graphs show the distribution of GBVs in your cattle compared to the average herd. If there is the same amount for each colour in a bar, then the proportion of animals in each quintile is equal. If the traits are showing a predominance of blue & green, you're already well advanced in genetic improvement. These animals have genetic merit above median within the north Australian herd. Yellow represents animals that are about average, while orange & red represent below median genetics within northern Australia. The genetic profile will influence bull selection in your herd. For example, if there is a problem with female fertility traits this analysis will identify this; using fertility GBVs to select bulls will be a way to tackle the problem.





What about individual GBVs?

How do I identify potential animals for bull breeding using the GBV data?

Using the filter function in excel can help you find specific animals, for example, animals with elite genetics. The following describes how to filter and find specific animals without the need to scroll the spreadsheet for days.
In your Microsoft Excel spreadsheet select the top row of headings (Row 1). Next, look to the top left and along from 'File' you will see a tab called 'Data'. Click 'Data' and around the top centre of screen you will notice a funnel-looking image called 'Filter'; click this. Now you will see little downward arrows in boxes on the right-hand side of each heading box in Row 1 of your spreadsheet. If you click the little arrow box it will show you all the data available in that column and you can search for and single out specific data.

Example 1: Finding Elites
If you find the column headed 'Elite' and filter by ticking the box to show all '1's' and click 'OK', this will filter and show all animals in your herd that have been tagged as elite animals.

Example 2: Finding Potential Culls
Search out quintile 1 (V Low) animals for puberty by finding the column headed 'Puberty' and filtering for number '1's. These are the animals that returned a very low likelihood to be pubertal.

Which animals in my herd are elite?

For this analysis we have identified elite animals as those that have genetics in the industry's top 50% for P4M and Temperament, top 70% for Puberty, Weight and Condition and top 75% for tick and fly lesion scores; ie, they are in the top 5% of animals. The following table indicates number of elite animals that can provide the opportunity for combined focus on fertility, growth, adaptation and temperament genetics if used in a bull breeding herd.

Number of elite animals identified by the project

	High grade heifers	Crossbred Heifers
Total	135	474
Elite	2 (1%)	4 (1%)

How do I interpret the percentile ranking in my data spreadsheet?

The number provided represents the ranking of the animal across the north Australian herd. For example, if an animal has a ranking of 63 then it ranks above 63% of cattle in northern Australia and there are 37% of cattle that rank higher.

GBVs within percentiles for each trait

Trait	Ranking of 1	Ranking of 50	Ranking of 100
Puberty	Lowest probability	Moderate probability	Highest probability
P4M	Lowest probability	Moderate probability	Highest probability
Weight	Lowest growth	Moderate growth	Highest growth
Height	Shortest cattle	Moderate height	Tallest cattle
Condition	Poorest adaptation	Moderately adapted	Well adapted to tropics
Temperament	Wildest	Moderate	Quietest
Fly lesions	Severe	Minor	Nil
Tick score	Highly resistant	Moderate resistance	Highly susceptible

How does this help me improve my herd?

How will the use of GBVs increase the amount of liveweight I sell?

The GBVs you attain from this project and from any future testing you do will allow you to profile the genetics in your herd and identify the genetic merit of individual animals. Animals with elite GBVs will allow you to slowly but steadily make genetic change within your herd leading to increased profitability. One example of this is female puberty for which scientific research has shown that earlier puberty can enable more liveweight to be produced per animal over their lifetime. This improvement in liveweight production and sales is key to improving your beef business's bottom line.

There is good opportunity for genetic improvement in the northern Australian beef herd and while on your road to genetic improvement, continue to select for improvements in all traits not just one or two. Research shows selection for fertility has little detrimental effect on other traits, which gives good opportunity to select for all traits.

When selecting on one trait it may indirectly cause a shift in another trait. This is called a genetic correlation. For example, in the table below, the genetic correlation between docility and P4M is 0.19. This means that if we select for P4M with no attention to temperament, docility will improve anyway, though not to the degree it would if we selected for both at once. A positive value is not always good; for example, the genetic correlations of 0.14 & 0.17 mean that animals that are less resistant to ticks and develop more fly lesions are likely to have a higher chance of becoming pregnant within 4 months of calving (P4M). Therefore, we select for both to avoid problems. Most traits are negatively correlated with hip height, indicating that tall cattle are less likely to be ideal for breeding.

Genetic correlations between traits

Note: For weight, height, BCS, puberty & P4M, higher values is better, but the reverse applies for temperament, tick and fly lesion scores

	Weight	Hip height	BCS	Puberty	P4M	Temperament	Tick score
Hip height	0.50						
BCS	0.29	-0.05					
Puberty	0.12	-0.36	0.15				
P4M	-0.02	-0.28	0.03	0.29			
Temperament	-0.02	-0.21	0.03	0.19	0.14		
Tick score	-0.02	-0.22	0.00	0.21	0.14	0.16	
Fly lesion score	0.09	-0.32	-0.02	0.27	0.17	-0.12	-0.21

How do I find animals that have the traits I want to select for?

In the spreadsheet we have provided a custom GBV selection calculator that you can use to specify the GBVs that have meaning to you and that you wish to select for in your herd. By entering the values that are your threshold for genetic selection for each trait into the coloured spaces you can find the animals you wish to select for breeding based on your breeding objectives.

Business impact

Extracts from one collaborator's comment on his results after the Sep 2020 analyses:

We usually get 90% heifers PTIC in 3 months but nearly 50% of the 17s were very late or didn't get in calf at all as heifers last year. We sold these thinking the result reflected the tough environment but we don't make (many) excuses and as always we were happy to apply the selection pressure. Now we can see they were almost all 1s and 2s and were genetic duds. Even when the environment was overwhelmingly bad, the genetics controlled the outcome. The vast majority of the PubQ 1s and 2s (lowest two quintiles for Puberty GBV) *have been culled for actual reproductive non-performance.*

I was disappointed with our results on the genomic analysis given the effort we've put in for over 4 decades. I can now see at least two of our purchased "star" sires directly or through their sons (representing at least 20% of our sire power) are reliably throwing PubQs at the bottom end of the spectrum indicating we are going forward with our female selection and then backward through our sires. One of them contributed 19 No 17 heifers and they are recorded as 1 PubQ 2 and 18 PubQ 1s. Only one of these 19 tested pregnant. The ability to get this information on our sires before using them is a game changer for us.

I believe Benomics is a huge advance along with the introduction of Brahman cattle, Buffel grass and Black pipe to the northern beef industry.

What's Next?

In the short term, collaborators have access to genomic analyses. If you send DNA, (eg, tail hairs or TSU from young bulls), to the lab for genotyping, the project will do genomics to produce GBVs. Please note that if you are doing any genotyping (= reading the DNA), our strong recommendation is to request the 50K Neogen TropBeef SNP chip for now. If you use this, it services our needs, it services BREEDPLAN needs, and you can have all sorts of secondary genomics done such as poll gene testing, testing for genetic diseases, and parentage verification. There are alternate (eg, Illumina 50K SNP chip) and cheaper options, but they are either restricted to one genomic analysis (eg, just parentage or just a specific disease test) or are less suited for genomics of north Australian cattle. We can run GBV with other chips, but a warning that the resulting GBV will be a bit less accurate (couple of %) than with the Tropbeef chip, particularly for high indicus content animals.

Definitions

Genotyping	Reading the DNA which is just a bunch of ACTGs and doesn't mean much to anyone until further analysis.
Genomics	Analysing the genotypes to produce meaningful & useful information aka GBVs which indicate the genetic value of an animal, that will be passed onto progeny.
Genotype	The genetic code for a particular trait.
Phenotype	A measurement or score of an animal that is the result of its genes and conditions during its life.
EBV	Estimated Breeding Value: An estimate of an animal's true breeding value.
GBV	Genomic Breeding Value: The genomic equivalent of an EBV where DNA is used to estimate the true genetic merit of potential breeding animals.
BCS	Body Condition Score: A score indicating the condition of an animal ranging from 1 (poor) to 5 (fat).
DNA	Deoxyribo Nucleic acid: The main constituent of chromosomes carrying genetic information (ie, the ACTGs).
Median	The midpoint where half the animals are higher, and half the animals are lower.
Puberty	A time when females can cycle, become pregnant and rear a calf. Typically, this occurs at 60-70% of mature liveweight.
P4M	Pregnant within 4 months of calving in wet cows. Therefore, you cannot have a P4M for a cow that loses its calf. P4M enables a cow to wean a calf in consecutive years.
Quintile	One division of 5 equal groups. The lowest 20% is the lowest quintile; the highest 20% is the top quintile.

Thank you

Thank you for your commitment and efforts as part of this project, we appreciate your support and hope you have gained valuable learnings from this project to further improve your business & herd fertility.
As always, best wishes from the team & if you have any questions, please get in contact with us.

Elsie Dodd	e.dodd@uq.edu.au	04 2799 3779
Geoffry Fordyce	g.fordyce@uq.edu.au	04 2810 9062
Ben Hayes	b.hayes@uq.edu.au	