12-2022

Variance Component Estimates for Growth Traits in Beef Cattle Using Selected Variants from Imputed Low-Pass Sequence Data

Chad A. Russell
University of Nebraska - Lincoln, crussell17@huskers.unl.edu

VARIANCE COMPONENT ESTIMATES FOR GROWTH TRAITS IN BEEF CATTLE USING SELECTED VARIANTS FROM IMPUTED LOW-PASS SEQUENCE DATA

by

Chad A. Russell

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Animal Science

Under the Supervision of Professor Matthew L. Spangler

Lincoln, Nebraska

December, 2022

VARIANCE COMPONENT ESTIMATES FOR GROWTH TRAITS IN BEEF CATTLE USING SELECTED VARIANTS FROM IMPUTED LOW-PASS SEQUENCE DATA

Chad A. Russell M.S.

University of Nebraska, 2022

Advisor: Matthew L. Spangler

A beef cattle population (n=2,343) was used to assess the impact of variants identified from imputed low-pass sequence (LPS) on the estimation of variance components and genetic parameters of birth weight (BWT) and post weaning gain (PWG). Variants were selected based on functional impact and were partitioned into four groups (Low, Modifier, Moderate, High) based on predicted functional consequences and re-partitioned based on consequence of mutation, such as missense and untranslated region variants, into six groups (G1-G6). Each subset was used to construct a genomic relationship matrix (GRM) for univariate animal models. Multiple analyses were conducted to compare the proportion of additive genetic variation explained by the different subsets individually and collectively, and these estimates were benchmarked against all LPS variants in a single GRM and array (e.g., GeneSeek Genomic Profiler 100K) genotypes. When all variants were included in a single GRM, heritability estimates for BWT and PWG were 0.43±0.05 and 0.38±0.05, respectively. Heritability estimates for BWT ranged from 0.10-0.42 dependent on which variant subsets were included. Similarly, estimates for PWG ranged from 0.05-0.38. Results showed that

variants in the subsets Modifier and G1 (untranslated region) yielded similar heritability estimates compared to the inclusion of all variants yielded the highest estimates, while estimates from GRM containing only variants in the categories High, G4 (non-coding transcript exon), and G6 (start and stop loss/gain) were the lowest. All variants combined provided similar heritability estimates to chip genotypes and provided minimal to no additional information when combined with chip data. This suggests that the chip data and the variants from LPS predicted to be less consequential are in relatively high linkage disequilibrium with the underlying causal variants and sufficiently spread throughout the genome to capture larger proportions of additive genetic variation.

Acknowledgements

We thank the U.S. Meat Animal Research Center staff for animal care and data recording and Holland Computer Center for computational resources. The USDA is an equal opportunity provider and employer. The mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.

Author's Acknowledgements

This thesis would not be possible without the assistance of my family, friends, and colleagues. I am especially thankful for my advisor Dr. Matthew Spangler, for his feedback, guidance, expertise and patience. I also appreciate my friends and family for believing in my ability to pursue my degree and complete my thesis. I would not have been able to do this without their motivation and guidance.

Table of Contents

CHAPTER I

LITERATURE REVIEW

Introduction

The inclusion of genomic data in genetic evaluations has allowed for increased accuracy of prediction for animal genetic merit. Genetic evaluations calculate estimated breeding values (EBV) of animals and associated accuracy for a variety of traits. Prior to the inclusion of genomic data, animals were evaluated using pedigree and available phenotypic data only. Pedigree data is used to estimate the relationship among individuals using the numerator relationship matrix, also known as the $\boldsymbol{A}$ matrix (Wright, 1922).

Pedigrees provide an estimate of the amount of genetic information shared among related individuals. While any individual inherits 50% of their genetic material from each of their parents, the expected values shared with each grandparent is 25% of their genetic material. This expectation however is not always the case, and assuming no inbreeding, individuals can share between 0 and 50% of their alleles with a single grandparent. Even though pedigree-based relationships are estimates, they still provide a framework for how phenotypic information can be "shared" among related individuals such that information on records for one animal can provide an EBV for related animals.

Methods for the Inclusion of Genomic Data in Evaluations

Fernando and Grossman (1989) theorized a generalized method for simultaneous evaluations with the inclusion of genetic markers and inverting relationship matrices as a

method to be used with best linear unbiased prediction (BLUP). As theorized and simulated by Meuwissen *et al*. (2001), if markers were distributed throughout the genome and then utilized in a genetic evaluation, there would be a benefit to the animal and plant industries by improving the rate of genetic change.

These theories combined with Van Raden's (2008) development of the genomic relationship matrix (GRM) also known as the $\boldsymbol{G}$ matrix, which accounts for similarity among animals based on SNP being Identical by State (IBS), provided a way for rapid development in the utilization of genomics. This development allowed for an efficient method for genetic predictions with the inclusion of genomic information. However, the methods described by Van Raden (2008) only worked assuming all animals are genotyped. The ability to combine genomic- and pedigree-based relationship matrices, as proposed by Christensen & Lund (2009) and Legarra *et al.* (2009), enabled the development of single-step evaluations and the eventual move away from the prior two-step approaches. Single-step methods estimate breeding values using all data at once, while two-step methods require the estimation of molecular-based EBV (MBV) in a separate step and then either the use of a multi-trait model that fits the EBV as a correlated trait or an indexing approach to combine the pedigree-based EBV and then, for genotyped animals, the MBV.

The use of genomic information via a GRM was quickly implemented with the use of BLUP, to make genomic best linear unbiased prediction (GBLUP). In the simplest case of GBLUP, a single trait animal model can be described as

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e}$$

where $\boldsymbol{y}$ represents a vector of phenotypic records, $\boldsymbol{b}$ represents a vector of fixed effects, $\boldsymbol{u}$ represents a vector of random additive genetic effects, $\boldsymbol{e}$ represents a vector of random residual components which is part of the phenotype which is not explained by effects in $\boldsymbol{b}$ or $\boldsymbol{u}$. The $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices relate observations in $\boldsymbol{y}$ to fixed effects in $\boldsymbol{b}$ and random additive genetic effects in $\boldsymbol{u}$, respectively. It is also assumed that

$$\boldsymbol{u} \sim N(0, \boldsymbol{G}\sigma_u^2),\ \boldsymbol{e} \sim N(0, \boldsymbol{I}\sigma_e^2),\ \text{Cov}[\boldsymbol{u}, \boldsymbol{e}] = 0,\ \boldsymbol{y} \sim \text{N}(\boldsymbol{Xb}, \boldsymbol{ZG}\sigma_u^2\boldsymbol{Z}' + \boldsymbol{I}\sigma_e^2)$$

where $\boldsymbol{G}\sigma_u^2$ and $\boldsymbol{I}\sigma_e^2$ are variance and covariance matrices for genetic and residual effects, respectively. These specifications are similar for a pedigree BLUP model, where $\boldsymbol{A}$ replaces $\boldsymbol{G}$.

Van Raden (2008) described three methods to form $\boldsymbol{G}$. The first, and most popular, method scales the $\boldsymbol{G}$ matrix to be analogous to the $\boldsymbol{A}$ matrix. $\boldsymbol{G}$ can be defined as

$$\boldsymbol{G} = \frac{\boldsymbol{MM'}}{2\sum p_i(1-p_i)},$$

where $\boldsymbol{M}$ denotes a centered SNP matrix with rows corresponding to each animal and columns representing each locus, and p_i is the minor allele frequency at the ith locus.

Historically, the process of incorporating genomic data into EBV was done through what is known as multi-step, where the calculation of genomically-enhanced estimated breeding values (GEBV) would require a traditional evaluation using the $\boldsymbol{A}$ matrix, then the extraction of peusdo-observations such as deregressed EBV, and then the estimation of SNP effects of genotyped animals, and then the combining of genomic predictions with traditional EBV. This process was then simplified into a single-step evaluation with one of the first being by Aguilar *et al*. (2010). The authors found that single-step nearly matched the accuracy and bias from that of the multi-step. Misztal *et*

al. (2009) derived the ***H*** matrix, a matrix the includes both pedigree-based relationships and genomic-based relationships. Afterwards, Aguilar *et al*. (2010) and Christensen & Lund (2009) both independently demonstrated how to easily invert ***H***. This matrix is best utilized when there is information on both genotyped and non-genotyped animals.

GBLUP is not the only modeling approach, some of the other methods include Bayesian methods, such as BayesR and BayesRC, among others. Kemper *et al*. (2015) used BayesR, described as a non-linear model by the authors, and high-density SNP genotypes and reported estimates of GEBV with similar to improved accuracies compared to GBLUP for within- and across-breed analyses. Both BayesR and GBLUP assume that each variant is equally likely to affect the trait, although with BayesRC this equality is not assumed. BayesRC does use similar prior information but classifies variants into classes, where this change has been shown to improve accuracy over BayesR in dairy cattle (MacLeod *et al*., 2016).

It was also shown by Erbe *et al*. (2012), that BayesR could perform better in terms of correlation of GEBV and daughter deviations compared to GBLUP. Fernando *et al*. (2014) used single step Bayesian regression (SSBR) in which they demonstrated combining genotyped and non-genotyped animals in a single evaluation using BayesCπ. Zhou *et al*. (2014b) observed that when combining genotyped and non-genotyped Nordic Red cattle, genetic predictions of animals for a whole genome analysis with BayesR yielded higher accuracies than GBLUP. However, Warburton *et al*. (2020) reported that GBLUP was still able to provide a higher prediction accuracy than the BayesR method when using a small number of highly relevant variants.

Generally, GBLUP methods focus solely on predicting additive genetic effects using the ***G*** matrix, although other matrices can be used in addition. The dominance relationship matrix, ***D***, follows the design of ***G*** but takes on values that are determined from the dominance level of a locus and the heterozygosity of the individual at that locus. The ***D*** matrix follows as

$$\boldsymbol{D} = \frac{\boldsymbol{TT'}}{\sum 2p_i q_i (1 - 2p_i q_i)}$$

where ***T*** is a matrix of heterozygosity coefficients and p_i and q_i are minor and major allele frequencies at the i^{th} locus (Su *et al.*, 2012). The findings of Raiden *et al.* (2018) suggest the use of a dominance matrix along with epistasis and heterozygosity may only provide a slight benefit when estimating genetic parameters and that for genetic prediction an additive genetic effects model was adequate. The authors saw an increase in accuracy from 0.28 to 0.33 when including nonadditive genetic effects for one group of cattle and an increase of 0.18 to 0.23 in another group. The magnitude of these increases is reasonable, given dominance and epistatic effects have been shown to account for a 4-6 % of phenotypic variation for several traits across breeds in dairy cattle (Marete *et al*., 2018). Computation can become even more complex when considering dominance and epistatic effects.

Differences created by multi-breed or multi-population evaluations can impact resulting accuracy of predictions and can require changes to the kinship matrix that is used. Differences can be due to different linkage disequilibrium (LD) patterns between breeds, where fewer shared alleles between breeds provide less information for other breeds in the matrix (Zhou *et al*., 2014a). LD is the non-random association of alleles

from two loci, which can reflect recombination history of alleles. As a population becomes older, more recombination events can occur that can lead to differences in LD phases between populations. It is likely that LD will differ across different populations and across diverged breeds (de Roos *et al.*, 2008). The authors also found that these breeds can differ so much that in order to obtain consistent marker effects across Angus, Jersey, and Holstein-Friesian it would take approximately 300,000 markers. Similarly, as shown by Kachman *et al*. (2013), genetic correlations between molecular breeding values and growth traits (weaning weight and post weaning gain) were low when trained in one breed and used to predict in a different, albeit it closely related, breed such as the case with Angus and Red Angus. However, the authors found that across breed and within breed molecular breeding values accuracies were similar when across-breed training sets were representative of the target breed(s).

The weighting of $\boldsymbol{G}$ was demonstrated by Van Raden (2008) where a weighting can be applied to matrices to add more or less emphasis to certain markers in order to better meet a certain objective. An example of this can be weighting based on QTL effect in order to account for differences in how SNPs may affect a trait differently. The use of weighted relationship matrices can also be used to account for the covariance of SNP effects across two breeds with different LD phases. The matrices are weighted for individual SNP effects when accounting for heterogenous variances and covariances. The genomic relationship coefficient for the weighted $\boldsymbol{G}$ matrix can be described as:

$$g_{ij} = \frac{\sum_{k=1}^{m} \boldsymbol{M}_{1(i,k)} \boldsymbol{M}_{2(j,k)} \omega_k}{\sqrt{\sum 2p_{1,k}\left(1 - p_{1,k}\right) \sum 2p_{2,k}\left(1 - p_{2,k}\right)}},$$

where ***M*** is a genotype matrix for individual *i* and individual *j* at marker *k*, ω_k is the weight of marker *k*, and p_k is the allele frequency for each of the two breeds, 1 and 2 (Zhou *et al*., 2014a). The weights used by the authors were measured as pairwise correlations between two breeds as measurements of LD between markers.

Genomic Data for Genetic Evaluations

Genomic sequencing technology has progressed substantially in recent years. Sanger (1975) developed a method to sequence DNA through dideoxy chain termination that was utilized for the next 40 years through improved processes. Some of these improvements came through the development of fluorescent detection systems that detected different colored dyed fluorescents for each nucleic base. The development of polymerase chain reaction (PCR) in the 1980s led to bacterial artificial chromosomes (BAC) that could be cloned to be used in sequencing and combining chains to compile a whole genome. Next-Gen sequencing (NGS) surfaced in 2008 as several companies released the high throughput technologies that utilized PCR and fluorescent detection in such a way that the whole genome could be sequenced in a matter of a few hours.

The use of NGS allows for a simpler way to identify segments of the genome which can be selected for. In order to select for features of the genome, a large variety of panels have been developed to test a variety of SNPs in order to provide more information. The SNPs for these panels are chosen to include significant variants or variants somewhat uniformly distributed throughout the genome and at intermediate

allelic frequency across a large number of breeds. These panels can be broken into the categories below as described by the Beef Cattle Research Center (Crowley, 2016):

Small panels: include between 5 and 2,000 SNPs. These are generally used for parentage, genetic abnormality testing, coat color, horns, etc.

Low density panels: include between 5,000 and 30,000 SNPs. These are used for genomic prediction/selection purposes when there are a lot of higher density genotypes available to use for imputation.

Medium density panels: include between 50,000 and 150,000 SNPs. This density is common in beef genomic selection programs. Usually, a population will begin by genotyping a reference set (influential animals in the breed) using a medium density panel realizing the definition of "medium" evolves overtime. A reference set is chosen to represent a population based on how well they represent the diversity or current population.

High density panels: include between 500,000 to 1 million SNPs. These panels are less popular in beef as it comes at a high cost.

Whole Genome Sequencing: occurs when an animal's entire genome is sequenced. This type of test is presently used primarily for research and is too expensive to be employed on farm.

Accuracy of Different Variant Densities and Causal Variants

There have been mixed results on increases in accuracy from the inclusion of additional genomic data. As found by Veerkamp *et al*. (2016), there was a marginal increase of less than 0.04 in heritability when comparing estimates from a low-density panel and a high-density panel. Additionally, as found by Chang *et al*. (2019) via simulation, there was no significant increase in accuracy from high density array panels or whole genome sequence data when compared to low and medium density panels. The authors attribute this to the limitations of current methodology for evaluation implementation that may be caused from the large number of unknown parameters which is limited by small marker effects and high false positives. On the other hand, the 1000 Bull Genomes project found that the inclusion of imputed sequence data provided a 2% improvement in prediction accuracy compared to 800k array data (Hayes *et al.*, 2014). The authors also demonstrate that BayesRC was able to identify some causal mutations in the imputed data (e.g., PLAG1). It was also found by Meuwissenn *et al*. (2021) that using WGS led to a 3% accuracy improvement over a 600k SNP chip for Australian Red cows.

Some traits may be monogenic, where one gene controls a trait, while other traits are polygenic, where more than one gene controls a trait. The polygenic traits are much more complex and can be difficult to estimate or predict. Causative variants, variants found to be the true cause of variation in traits, have been found recently that can account for a relatively large amount of variation for polygenic traits. The search for causal variants in complex traits can be a difficult process. As reported in a summary of findings, Casas *et al*. (2016) outlined the genomic regions for several traits where some

causative variants have been identified. As reported by Utsunomiya *et al*. (2013), a GWAS was able to associate five variants with birth weight in Nellore cattle. The most significant single variant was able to account for 4.62% of variation in birth weight, while a non-significant variant may only account for 0.4% or less of the variation.

Although this summary is not comprehensive, these findings can then be used to make better selection decisions. Although causal variants and marker-assisted selection may be useful, it is important to make sure that the information used from evaluations have an emphasis on phenotypic record collection (Dekkers, 2004).

Imputation

The process of imputation takes genotypes of variants from lower density arrays and uses animal relationships and/or haplotype groups from animals that have been genotyped or sequenced at higher densities to determine the likely genotypes of the animals with lower density genotypes. Marchini *et al*. (2007) reported that animals that have been genotyped via lower density assays had evaluations that were not as accurate as those of animals genotyped with high-density arrays, until the inclusion of imputation.

Imputation can be done by comparing the genomic relationships of animals based on pedigree, haplotype groups, or both. Imputation has been utilized in several studies, for example in Ventura *et al.* (2016), the authors were able to improve imputation accuracy in sheep by testing different panels, software, imputation processes, and which animals to genotype with the higher density. The authors found that by conducting a two-step imputation process from 5k to 50k and then to high density outperformed going

directly from 5K to high density. They also found that a large reference set of animals provided increases across many animals both purebred and crossbred, especially for larger breeds. However, large reference sets resulted in a slight loss in accuracy for some animals when compared to small within-breeds reference sets, but the increase across most animals justified the larger sets (Ventura *et al*., 2016).

In order to improve accuracy of imputation for rare variants, the use of a two-step imputation process can be used. As found by Kreiner-Møller *et al*. (2015), going from 550k to WGS yielded accuracies ranging from 0.806-0.942 for markers involving rare homozygotes. By using a two-step approach going from the 550k to ~4.3 million markers and then up to WGS they were able to improve accuracies for the same markers to range from 0.830-0.955. There was not any improvement for more common homozygote markers. However, given that the heterozygote and rare homozygote marker calls were improved, the overall accuracy of imputation as also improved.

As imputation is optimized, more avenues for genomic testing are opened. Accuracy of imputation can be improved in several ways such as when the reference set of animals is larger in size especially for low frequency alleles (Browning and Browning, 2009). The relationship of the animals included in the reference set compared to those in the target dataset may also be important for increasing accuracy as shown by Moghaddar *et al*. (2015) when comparing purebred or crossbred Merino sheep.

Given more work is needed to better understand the impact of including potentially causative variants in genetic evaluations, and more data is needed to identify causative variants, a new process to identify these markers for a reasonable cost could be

beneficial. With the recent development of low pass sequencing (LPS), a process of using a low coverage sequencing at around 0.5x, and then imputing markers across the genome, similar genotype identification accuracy might be achieved as compared to SNP array sequenced data. As found in Snelling *et al*. (2020), imputed LPS was able to produce similar genotypes calls to that of SNP arrays by looking at correlations between the two sets. Along with this, LPS provides a reasonable cost-effective alternative to genotyping with high accuracy.

Summary

Genetic evaluations have progressed greatly with the development of new methods. With the use of genomics, accuracy of prediction has been increased and will likely continue to be improved. One of these improvements could come in the form of LPS and imputation providing a new way of obtaining genomic information that is more dynamic with the potential to increase accuracy.

Literature Cited

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752. doi:10.3168/jds.2009-2730.

Browning, B. L., and S. R. Browning. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84:210-223. doi:10.1016/j.ajhg.2009.01.005.

Casas, E., and M. E. Kehrli. 2016. A review of selected genes with known effects on performance and health of cattle. Front. Vet. Sci. 15:1-11. doi: 10.3389/fvets.2016.00113.

Chang, L. Y., S. Toghiani, S. E. Aggrey, and R. Rekaya. 2019. Increasing accuracy of genomic selection in presence of high density marker panels through the prioritization of relevant polymorphisms. BMC Genet. 20:21. doi:10.1186/s12863-019-0720-5.

Christensen, O. F., and M. S. Lund. 2009. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2. doi:10.1186/1297-9686-42-2.

Crowley, J. 2016. BCRC and CBBC Fact Sheet Genomics and Genotyping in Beef Production. https://www.beefresearch.ca/files/pdf/BCRC_and_CBBC_Fact_Sheet_Genomics_and_Genotyping_in_Beef_Production.pdf.

de Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. Genetics. 179:1503–1512. doi:10.1534/genetics.107.084301.

Dekkers, J. C. M. 2004. Commercial application of marker and gene assisted selection in livestock- Strategies and lessons. J. Anim. Sci. 82:E313–E328. doi:10.2527/2004.8213_supplE313x.

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P.J. Bowman, C.M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95:4114-4129. doi:10.3168/jds.2011-5019.

Fernando, R. L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21:467–477. doi:10.1186/1297-9686-21-4-467.

Fernando R. L., J. C. M. Dekkers, and D. J. Garrick. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genet. Sel. Evol. 46:59. doi:10.1186/1297-9686-46-50.

Hayes, B.J., I. M. Macleod, H. D. Daetwyler, P. J. Bowman, A. J. Chamberlian, C. J. V. Jagt, A. Capitan, H. Pausch, P. Stothard, X. Liao, C. Schrooten, E. Mullaart, R. Fries, B. Guldbrandtsen, M. S. Lund, D. Biochard, R. F. Veerkamp, C. P. Vantassell, B. Gredler, T. Druet, A. Bagnato, J. Vikki, D. J. Dekoning, E. Santus, and M. E. Goddard. 2014. Genomic prediction from whole genome sequence in livestock: the

1000 Bull Genomes Project. In: Proceedings of the 10th World Congress of Genetics Applied to Livestock Production.

Kachman, S. D., M. L. Spangler, G. L. Bennett, K. J. Hanford, L. A. Kuehn, E. J. Pollak, W. M. Snelling, R. M. Thallman, M. Saatchi, D. J. Garrick, R. D. Schnabel, J. F. Taylor, and E. J. Pollak. 2013. Comparison of within and across breed trained molecular breeding values in seven breeds of beef cattle. Genet. Sel. Evol. 45:30. doi:10.1186/1297-9686-45-30.

Kemper, K. E., C. M. Reich, P. J. Bowman, C. J. V. Jagt, A. J. Chamberlain, B. A. Mason, B. J. Hayes, and M. E. Goddard. 2015. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. Genet. Sel. Evol. 47:29. doi:10.1186/s12711-014-0074-4.

Kreiner-Møller, E., C. Medina-Gomez, A. Uitterlinden, F. Rivadeneira and K. Estrada. 2015. Improving accuracy of rare variant imputation with a two-step imputation approach. Eur. J. Hum. Genet. 23:395–400. doi:10.1038/ejhg.2014.91.

MacLeod, I. M., P. J. Bowman, C. J. Vander Jagt, M. Haile-Mariam, K. E. Kemper, A. J. Chamberlain, C. Schrooten, B. J. Hayes, and M. E. Goddard. 2016. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genomics. 17:144. doi:10.1186/s12864-016-2443-6.

Marchini, J., B. Howie, S. Myers, G. McVean, and P. Donnelly. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. 39:906–913. doi:10.1038/ng2088.

Marete, A., L. Janns, B. Guldbrandtsen, C. Hoze, G. Sahana, D. Boichard, and M. S. Lund. 2018. Exploring non-additive variance using genome-wide dense SNP in four French and Nordic dairy cattle breeds. In: Proceedings of the 11th World Congress of Genetics Applied to Livestock Production.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157(4):1819-1829. doi:10.1093/genetics/157.4.1819.

Meuwissen, T. H. E., I. van den Berg, and M. Goddard. 2021. On the use of whole-genome sequence data for across-breed genomic prediction and fine-scale mapping of QTL. Genet. Sel. Evol. 53:19. doi:10.1186/s12711-021-00607-4.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92(9):4648-4655. doi:10.3168/jds.2009-2064.

Moghaddar, N., K. P. Gore, H. D. Daetwyler, B. J. Hayes, and J. H. J. van der Werf. 2015. Accuracy of genotype imputation based on random and selected reference sets in purebred and crossbred sheep populations and its effect on accuracy of genomic prediction. Genet. Sel. Evol. 47:97. doi:10.1186/s12711-015-0175-8.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92(9):4656-4663. doi:10.3168/jds.2009-2061.

Raidan, F. S. S., L. R. Porto-Neto, Y. Li, S. A. Lehnert, Z. G. Vitezica, and A. Reverter. 2018. Evaluation of nonadditive effects in yearling weight of tropical beef cattle. J. Anim. Sci. 96(10):4028-4034. doi:10.1093/jas/sky275.

Sanger, F., and A. R. Coulson. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. 94(3):441-448. doi:10.1016/0022-2836(75)90213-2.

Snelling, W. M., J. L. Hoff, J. H. Li, L. A. Kuehn, B. N. Keel, A. K. Lindholm-Perry, and J. K. Pickrell. 2020. Assessment of imputation from low-pass sequencing to predict merit of beef steers. Genes. 11(11):1312. doi:10.3390/genes11111312.

Su, G., O. F. Christensen, T. Ostersen, M. Henryon, and M. S. Lund. 2012. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PLoS One. 7(9):e45293. doi:10.1371/journal.pone.0045293.

Utsunomiya, Y. T., A. S. do Carmo, R. Carvalheiro, H. H. R. Neves, M. C. Matos, L. B. Zavarez, A. M. P. O'Brien, J. Sölkner, J. C. McEwan, J. B. Cole, C. P. Van Tassell, F. S. Schenkel, M. V. G. B. da Silva, L. R. P. Neto, T. S. Sonstegard, and J. F. Garcia. 2013. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. BMC Genet. 14:52. doi:10.1186/1471-2156-14-52.

Van Raden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91(11):4414–4423. doi:10.3168/jds.2007-0980.

Veerkamp, R. F., A. C. Bouwman, C. Schrooten, and M. P. L. Calus. 2016. Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein–Friesian cattle. Genet. Sel. Evol. 48:95. doi:10.1186/s12711-016-0274-1.

Ventura, R. V., S. P. Miller, K. G. Dodds, B. Auvray, M. Lee, M. Bixley, S. M. Clarke, and J. C. McEwan. 2016. Assessing accuracy of imputation using different SNP panel densities in a multi-breed sheep population. Genet. Sel. Evol. 48:71. doi:10.1186/s12711-016-0244-7.

Warburton, C. L., B. N. Engle, E. M. Ross, R. Costilla, S. S. Moore, N. J. Corbet, J. M. Allen, A. R. Laing, G. Fordyce, R. E. Lyons, M. R. McGowan, B. M. Burns, and B. J. Hayes. 2020. Use of whole-genome sequence data and novel genomic selection strategies to improve selection for age at puberty in tropically-adapted beef heifers. Genet. Sel. Evol. 52:28. doi:10.1186/s12711-020-00547-5.

Wright, S. 1922. Coefficients of Inbreeding and Relationship. The American Naturalist. 56(645):330–338. doi:10.1086/279872.

Zhou, L., M. S. Lund, Y. Wang, and G. Su. 2014a. Genomic predictions across Nordic Holstein and Nordic Red using the genomic best linear unbiased prediction model with different genomic relationship matrices. J. Anim. Breed. Genet. 131:249-257. doi:10.1111/jbg.12089.

Zhou, L., B. Heringstad, G. Su, B. Guldbrandtsen, T. H. E. Meuwissen, M. Svendsen, H. Grove, U. S. Nielsen, and M. S. Lund. 2014b. Genomic predictions based on a joint reference population for the Nordic Red cattle breeds. J. Dairy Sci. 97:4485-4496. doi:10.3168/jds.2013-7580.

CHAPTER II

VARIANCE COMPONENT ESTIMATES FOR GROWTH TRAITS IN BEEF CATTLE USING SELECTED VARIANTS FROM IMPUTED LOW-PASS SEQUENCE DATA

Abstract

A beef cattle population (n=2,343) was used to assess the impact of variants identified from imputed low-pass sequence (LPS) on the estimation of variance components and genetic parameters of birth weight (BWT) and post weaning gain (PWG). Variants were selected based on functional impact and were partitioned into four groups (Low, Modifier, Moderate, High) based on predicted functional consequences and re-partitioned based on consequence of mutation, such as missense and untranslated region variants, into six groups (G1-G6). Each subset was used to construct a genomic relationship matrix (GRM) for univariate animal models. Multiple analyses were conducted to compare the proportion of additive genetic variation explained by the different subsets individually and collectively, and these estimates were benchmarked against all LPS variants in a single GRM and array (e.g., GeneSeek Genomic Profiler 100K) genotypes. When all variants were included in a single GRM, heritability estimates for BWT and PWG were 0.43±0.05 and 0.38±0.05, respectively. Heritability estimates for BWT ranged from 0.10-0.42 dependent on which variant subsets were included. Similarly, estimates for PWG ranged from 0.05-0.38. Results showed that variants in the subsets Modifier and G1 (untranslated region) yielded similar heritability estimates compared to the inclusion of all variants yielded the highest estimates, while

estimates from GRM containing only variants in the categories High, G4 (non-coding transcript exon), and G6 (start and stop loss/gain) were the lowest. All variants combined provided similar heritability estimates to chip genotypes and provided minimal to no additional information when combined with chip data. This suggests that the chip data and the variants from LPS predicted to be less consequential are in relatively high linkage disequilibrium with the underlying causal variants and sufficiently spread throughout the genome to capture larger proportions of additive genetic variation.

Introduction

Although advances have been made in recent years relative to the incorporation of genomic data into routine genetic evaluations of beef cattle and corresponding increases in predictive accuracy, hope remains that the inclusion of causal variants into prediction models will lead to further gains in accuracy. Such efforts require the use of whole-genome sequencing (WGS), and despite decreases in cost, wide-spread deep sequencing of seedstock cattle is cost prohibitive. The use of low-pass sequencing at much shallower depths (e.g., 0.5x) is a much more attractive alternative, when coupled with imputation, to garner genotypes of potentially causal variants on large numbers of individuals for a lesser cost. Such an approach also represents a means of making marker subsets actually used in genetic evaluations more flexible and dynamic.

Large effect variants for complex traits such as birth weight in beef cattle are known to exist and are predictive in populations external to where they were identified (Snelling *et al.*, 2017). However, the general benefits of including putative causal variants

in genetic predictions or the use of WGS for prediction more generally has had mixed results in the literature (e.g., Heidaritabar *et al.*, 2016; Veerkamp *et al.*, 2016; Warburton *et al.*, 2020). Several challenges exist relative to the use of WGS for genetic prediction, including the inability to accurately estimate the effect of all WGS Single Nucleotide Polymorphisms (SNP) using currently available phenotypic datasets to allow for accurate pre-selection of variants. Additionally, currently available high-density panels may sufficiently capture enough genetic variation through LD with causal variants to limit gains from using WGS (e.g., Frischknecht *et al.*, 2018). However, evidence in the literature supports the concept of pre-selected WGS variants to improve prediction accuracy in multi-breed populations (e.g., Raymond *et al.*, 2018). Consequently, the objective of the current study was to investigate the benefit of fitting variants based on predicted functional impact and variant consequence in prediction models for birth weight (BWT) and post weaning gain (PWG) in a multi-breed beef cattle population using a genomic best linear unbiased prediction framework.

Materials and Methods

Animal Care

All methods and animal care described in this study followed the Guide for the Care and Use of Agricultural Animals in Agricultural Research and Teaching (Ag Guide, 2020) and were approved by the U. S. Meat Animal Research Center (USMARC) Animal Care and Use Committee.

Data

A crossbred beef cattle population (n=2,343) with sequence variant genotypes imputed from low-pass (~0.5x) WGS and recorded BWT and PWG was used. Animals were part of the continuously sampled USMARC Germplasm Evaluation (GPE) project (Retallick *et al.*, 2017). The low-coverage sequence was submitted to the Gencove pipeline for imputation with loimpute (Loimpute-Public, 2020) to a haplotype reference panel constructed from WGS of 946 cattle (598 available from NCBI Sequence Read Archive; 348 GPE sires) (Snelling *et al.*, 2020). As described by Snelling *et al*. (2022), functional impact and consequences of the imputed variants was assessed with snpEff using the Ensembl annotation of the ARS-UCD1.2 assembly of the bovine genome (Aken *et al*., 2016; Cingolani *et al*., 2012; Rosen *et al*., 2020). Genotypes for interesting variants were extracted from the imputed calls of each individual with low-pass sequence. Interesting variants included variants in exons of protein-coding genes, which may affect gene function, in untranslated regions (UTR) and non-coding transcript exon variants, which may impact gene regulation. Low-pass calls were required to have a genotype probability greater than 0.95, then a 0.95 call rate filter by animal and variant was applied to the set of interesting variants.

In total, a set of 1,145,892 variants was identified for further analysis. The variants were partitioned into two different categories, predicted functional impact and consequence of mutation. For the functional impact classifications, variants were divided into four groups: High, Moderate, Modifier, and Low. For the consequence of mutation classifications, variants were divided into six groups: G1 (untranslated region; UTR), G2

(synonymous), G3 (missense), G4 (non-coding transcript exon variants), G5 (splice region), and G6 (start and stop loss/gain). Animals also had LPS genotypes called that corresponded to the GGP Bovine 100K SNP Array with a total of 72,997 variants (Chip).

In order to determine the effects of linkage disequilibrium (LD), pruning at differing levels of r^2 values were tested (r^2 = 0.9, 0.8 and 0.7). Variants were prioritized for pruning based on the level of predicted function from highest potential effect to the lowest. For example, for the functional impact subsets this meant that if a variant was in LD with a variant contained in the High subset, the lower-level variant was removed. For functional impact variants, this was ordered as High, Moderate, Modifier, and then Low. For consequence variants, this was ordered as G6, G5, G3, G4, G1, and then G2. If a variant was in LD with another variant from the same subset, both variants were kept. Results herein include only those from the most stringent LD cutoff.

After removing those that were fixed (minor allele frequency (MAF) of 0) or were in LD with another marker at a level of r^2=0.7 data were reduced to 461,806 variants for the functional impact classifications and 435,538 for consequence classifications. These values differ due to the way pruning was conducted. Variants in the same subset that were in high LD were retained while those in high LD across subsets were removed. Consequently, when classifications changed, subsets changed and different variants were removed based on LD. Total variants in the functional impact subsets were as follows: Low (98,557 variants across 17,885 genes), Modifier (183,823 variants across 16,970 genes), Moderate (171,690 variants across 19,599 genes) and High (7,736 variants across 5,360 genes). The other category based on consequence of mutation were partitioned as

follows: G1 (138,364 variants across 14,199 genes), G2 (79,787 variants across 16,784 genes), G3 (147,908 variants across 19,185 genes), G4 (32,949 variants across 2,607 genes), G5 (31,440 variants across 12,009 genes), and G6 (5,090 variants across 3,812 genes).

Statistical analysis

All analyses were conducted using the ASREML 4.1 software package (Gilmour *et al*., 2015). Genomic Relationship Matrices (GRM) were constructed to contain variants from each of the partitioned subsets, and one GRM that included all variants. Genetic parameters were obtained from univariate animal models in a GBLUP framework by fitting fixed effects of sex, contemporary group (concatenation of year, season of birth, and age of dam (BW only)) and linear covariates of expected heterozygosity and breed fractions. Random effects included the additive genetic effect associated with each GRM and a residual. In matrix notation, the model for one GRM is:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e}$$

where $\boldsymbol{y}$ represents a vector of observations for the traits BW or PWG, $\boldsymbol{b}$ represents a vector of fixed effects, $\boldsymbol{u}$ represents a vector of random additive genetic effects, $\boldsymbol{e}$ represents a vector of random residual components which is part of the phenotype which is not explained by effects in $\boldsymbol{b}$ or $\boldsymbol{u}$. The $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices relate observations in $\boldsymbol{y}$ to fixed effects in $\boldsymbol{b}$ and random additive genetic effects in $\boldsymbol{u}$, respectively. It is also assumed that

$$\boldsymbol{u} \sim N(0, \boldsymbol{G}\sigma_{\boldsymbol{u}}^2),\ \boldsymbol{e} \sim N(0, \boldsymbol{I}\sigma_{\boldsymbol{e}}^2),\ \mathrm{Cov}[\boldsymbol{u}, \boldsymbol{e}] = 0,\ \boldsymbol{y} \sim \mathrm{N}(\boldsymbol{Xb}, \boldsymbol{ZG}\sigma_{\boldsymbol{u}}^2\boldsymbol{Z}' + \boldsymbol{I}\sigma_{\boldsymbol{e}}^2)$$

where $\boldsymbol{G\sigma_u^2}$ and $\boldsymbol{I\sigma_e^2}$ are variance and covariance matrices for genetic and residual effects, respectively. The addition of more than one GRM required fitting additional corresponding random effects (***u)*** and incidence matrices.

As detailed by Yang *et al*. (2011), models can contain multiple GRM to be used as random effects in a joint analysis. In the current study this included fitting a single random effect associated with the GRM comprised of a subset of variants (e.g., High, G3, etc.), models that included multiple random effects each associated with a different GRM (e.g., Low & Modifier; G1,2,4), models that included all variants in a single GRM (Full) and models that included all variants by fitting multiple GRM (All). Variants from the Chip genotypes were also included in their own GRM. To measure any additional information from Full when compared to Chip, a single GRM was built to include all variants in both the Full and Chip sets (Full & Chip). There were 2,820 SNP that were shared between these two sets, whether this was from Full for functional impact or consequence of mutation partitions, so the variants were removed from the Chip set to avoid double counting of these variants. To test if the number of markers impacted estimates, ten random sets of variants from Low were taken with 7,736 variants, the same number of variants found in High. Each random set was then included in single GRM models for comparison.

Riemannian distance between GRMs was used as an additional method to determine differences among GRMs by considering intra-class variance and distance. The Riemannian distance between two semi-positive definite covariance matrices, A and B is given by:

$$\delta_R(A,B) = \sqrt{\sum_{i=1}^{n} \{\ln[\lambda_i(A^{-1}B)]\}^2}$$

where the λ_i are the real and strictly positive eigenvalues of the matrices A and B (Moakher, 2005).

Results

Estimates of variance components and heritability for single GRM models of BWT and PWG for functional impact subsets are reported in Tables 1 and 2, respectively, while Tables 3 and 4 report findings for consequence separated subsets. These single GRMs include Chip, Full, Full & Chip and the subsets (e.g., Low, Moderate, G1, G2, etc.). Differences in heritability estimates between models were negligible except for models that omitted the Modifier or G1 variants. The Full and Chip models resulted in similar estimates and when they were combined into one GRM, they provided minimal or no increase in heritability estimates.

Estimates of variance components, heritability, and percentage of additive genetic variance explained by each combination of GRMs for models with more than one GRM for BWT and PWG for functional impact subsets are reported in Tables 5 and 6, respectively, while Tables 7 through 12 report findings for consequence separated subsets. Tables 7, 8, and 9 are models for BWT and Tables 10, 11, and 12 for PWG. The tables are separated based on number of GRM in each model where Tables 7 and 10 contain models with two GRMs, 8 and 11 for three GRMs, and 9 and 12 for four or more GRMs. These contain every combination of GRMs where the models included multiple GRM as different random effects (i.e. Low & Modifier, Low, Modifier & Moderate, etc.).

Estimates from models that included GRM based on High impact or G4 and G6 variants resulted in heritability estimates similar to what was obtained when they were not included. In general, the highest estimates of heritability were from models containing Modifier or G1 and the lower estimates from models containing High, G4, or G6.

The 10 random sets of Low variants had an average heritability estimate of 0.16 with a standard deviation of 0.03, while the heritability estimate from High was 0.10. The Riemannian distances did not differ dramatically from one another for either partition except when the High, G4, or G6 subsets was involved. The distances for functional impact combinations that did not contain High ranged from 26.4-29.8, while the combinations that contained High had a range of 93.0-94.5. For the consequence of mutation combinations, the distances ranged from 27.1-38.5 for combinations not containing G4 or G6. Combinations that contained G4 had distances of 51.7-55.5, and G6 containing combinations ranged from 122.4-124.6, with the G4:G6 combination falling into the latter range. This indicates that the High impact variants, the non-coding transcript exon (G4), and start/stop loss/gain (G6) variants are different from the rest of the subsets.

Discussion

Estimates of heritability from models fitting the Full GRMs from either categorization were 0.43±0.05 and 0.38±0.05 for BWT and PWG, respectively, both in the range of estimates previously reported for these traits in beef cattle (Koots *et al.*, 1994). The Chip GRM yielded higher or similar estimates of heritability to that of the

Full GRMs despite having only 2,820 SNP in common between the two, suggesting that any differences in heritability estimates were due to the content not in common.

The models that contained all subsets in separate GRMs provided a very similar estimate of heritability compared to Full. Warburton *et al.* (2020) reported results from a beef cattle population comparing the use of a single GRM with pre-selected WGS variants added to panel SNP and a multiple GRM model whereby panel SNP comprised one GRM and WGS variants comprised a second GRM. The multiple GRM model produced greater accuracy than the single GRM model in their study. Interestingly, only fitting the GRM comprised of Modifier and G1 variants resulted in a negligible decrease in the additive genetic variance compared to the Full GRM.

Perhaps more interesting are results from fitting only the GRM comprised of High impact variants. These loss-of-function variants are predicted to substantially alter protein-coding genes and yet only accounted for 22.5% of the additive genetic variation in BWT and 25.7% in PWG in this population. Approximately 62% of annotated genes in the current study's data include the high impact variants, and the low amount of phenotypic variation explained by high impact variants could indicate that genes with similar function may compensate for genes affected by high impact alleles (El-Brolosy and Stainier, 2017). Similarly, Veerkamp *et al*. (2016) reported a marginal gain in the proportion of genetic variation explained when fitting a GRM based on full sequence data as compared to a GRM based on a common SNP array for traits in Holstein cattle despite identifying 42 variants that explained ~23% of the genetic variance when they were fitted alone.

The classification of mutation category results also follows previous literature. Zhang *et al.* (2020) reported that UTR, synonymous, and missense variants explained a large amount of genetic variance for growth traits such as average daily gain (ADG) or residual feed intake (RFI) in beef cattle. In the current study, similar variants are accounted for in G1, G2, and G3. These variants account for a total of 62.7% of additive variance in PWG and 83.2% of the additive variance for BWT when they were fit in All. While many of these variants are not believed to be causal variants, they are still able to account for large proportions of genetic variation, which may imply further that the inclusion of causal variants might not substantially improve prediction accuracy. This might be particularly true within-population given that linkage disequilibrium between observed variants and underlying causal variants is sufficiently high. In across-breed or admixed populations, where LD structures differ, the use of causal variants might prove more helpful in increasing prediction accuracy or more specifically in providing more robust prediction accuracies across breeds.

The Full GRMs provided the same heritability estimates across both traits and partitions. Although the four subsets in the functional impact partition are not quite equivalent to the six subsets in the consequence of mutation partition, but in general the synonymous variants, G2, were in Low while High contained most of the variants found in G4 and G6. This can be further seen in the heritability estimates with similar estimates between G2 and Low and the lowest estimates coming from High and G6.

In general, as the number of variants included in a GRM increased the estimate of heritability increased. However, when taking 10 random samples from Low, the resulting

heritability estimate was consistently higher than the estimate from the High subset. This could be due to variants in Low covering more of the genome and having higher MAF on average.

Models that included GRM from Chip and Full variants resulted in similar heritability estimates, a result that is supported by Frischknecht *et al*. (2018). The authors found no difference in genomic prediction accuracy when comparing a 50k SNP chip compared to 50k imputed WGS variants in Brown Swiss cattle. They also found slight deviations in accuracy between traits, which was also observed by Lopez *et al*. (2021) when using pre-selected variants where marbling score accuracy increased slightly while carcass weight accuracy had a slight decrease. Additionally, Lopez *et al*. (2021) found that WGS was more accurate than using a specific genomic region, such as intronic or synonymous regions.

Conclusion

. Results suggest that if reduced subsets based on predicted function or consequence, as done in the current study, were to be included for genetic prediction, additional information (i.e., pedigree or more global SNP representation) would be needed to fully capture additive genetic variation. In the current study similar results were found using SNP from a standard platform compared to pre-selected variants from imputed low-pass sequence. However, different categorizations or prioritizing different regions of the genome could yield improvements in results. Additionally, the current study relied on current annotation to initially choose variants. Improvement in annotation

could also substantially improve initial variant selection. The use of low-pass sequencing may allow for more robust prediction in multi-breed populations and allow for more frequent changes to genomic content without increasing genotyping cost. Future research including different classifications for partitioning variants into GRM, investigating different trait complexes and quantifying predictive accuracy across-breeds is needed.

Literature Cited

Ag Guide. 2020. Guide for care and use of agricultural animals in research and teaching. 4th ed. Champaign (IL): American Dairy Science Association, American Society of Animal Science, and Poultry Science Association.

Aken, B. L., S. Ayling, D. Barrell, L. Clarke, V. Curwen, S. Fairley, J. F. Banet, K. Billis, C. G. Girón, T. Hourlier, K. Howe, A. Kähäri, F. Kokocinski, F. J. Martin, D. N. Murphy, R. Nag, M. Ruffier, M. Schuster, Y. A. Tang, J. Vogel, S. White, A. Zadissa, P. Flicek, and S .M. J. Searle. 2016. The Ensembl gene annotation system. Database 2016 baw093. doi:10.1093/database/baw093.

Cingolani P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6:80–92. doi:10.4161/fly.19695.

El-Brolosy, M. A., and D. Y. R. Stainier. 2017. Genetic compensation: A phenomenon in search of mechanisms. PloS Genet. 13(7):e1006780. doi:10.1371/journal.pgen.1006780.

Frischknecht, M., T. H. E. Meuwissen, B. Bapst, F. R. Seefried, C. Flury, D. Garrick, H. Signer-Hasler, C. Stricker, Intergenomics Consortium, A. Bieber, R. Fries, I. Russ, J. Sölkner, A. Bagnato, and B. Gredler-Grandl. 2018. Short communication: Genomic prediction using imputed whole-genome sequence variants in Brown Swiss Cattle. J. Dairy Sci. 101:1292–1296. doi:10.3168/jds.2017-12890.

Gilmour A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2015. VSN International Ltd. Hemel Hempstead. HP1 1ES, UK. www.vsni.co.uk.

Heidaritabar M., M. P. L. Calus, H-J. Megens, A. Vereijken, M. A. M. Groenen, and J. W. M. Bastiaansen. 2016. Accuracy of genomic prediction using imputed whole-genome sequence data in white layers. J. Anim. Breed. Genet. 133:167–79. doi:10.1111/jbg.12199.

Koots, K. R., J. P. Gibson, C. Smith, and J. W. Wilton. 1994. Analyses of published genetic parameter estimates for beef production traits. 1. Heritability. Anim. Breed. Abstr. 62:309-337.

Loimpute-Public. 2020. Available at: https://gitlab.com/gencove/loimpute-public.

Lopez, B. I. M., N. An, K. Srikanth, S. Lee, J.-D. Oh, D.-H. Shin, W. Park, H.-H. Chai, J.-E. Park, and D. Lim. 2021. Genomic Prediction Based on SNP Functional Annotation Using Imputed Whole-Genome Sequence Data in Korean Hanwoo Cattle. Front. Genet. 11:603822. doi:10.3389/fgene.2020.603822.

Moakher, M. 2005. A differential geometric approach to the geometric mean of symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl. 26(3):735-747. doi:10.1137/S0895479803436937.

Raymond, B., A. C. Bouwman, C. Schrooten, J. Houwing-Duistermaat, and R. F. Veerkamp. 2018. Utility of whole-genome sequence data for across-breed genomic prediction. Genet. Sel. Evol. 50:27. doi:10.1186/s12711-018-0396-8.

Retallick, K. J., J. M. Bormann, R. L. Weaber, M. D. MacNeil, H. L. Bradford, H. C. Freetly, K. E. Hales, D. W. Moser, W. M. Snelling, R. M. Thallman, and L. A. Kuehn. 2017. Genetic variance and covariance and breed differences for feed intake

and average daily gain to improve feed efficiency in growing cattle. J. Anim. Sci. 95(4):1444–1450. doi:10.2527/jas.2016.1260.

Rosen, B. D., D. M. Bickhart, R. D. Schnabel, S. Koren, C. Elsik, E. Tseng, T. N. Rowan, W. Y. Low, A. Zimin, C. Couldrey, R. Hall, W. Li, A. Rhie, J. Ghurye, S. D. McKay, F. Rhibaud-Nissen, J. Hoffman, B. M. Murdoch, W. M. Snelling, T. G. McDaneld, J. A. Hammond, J. C. Schwartz, W. Nandolo, D. E. Hagen, C. Dreischer, S. J. Schultheiss, S. G. Schroeder, A. M. Phillippy, J. B. Cole, C. P. Van Tassell, G. Liu, T. P. L. Smith, and J. F. Medrano. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience 9. doi:10.1093/gigascience/giaa021.

Snelling, W. M., L. A. Kuehn, B. N. Keel, R. M. Thallman, and G. L. Bennett. 2017. Linkage disequilibrium among commonly genotyped SNP variants detected from bull sequence. Anim. Genet. 48:516–522. doi:10.1111/age.12579.

Snelling, W. M., J. L. Hoff, J. H. Li, L. A. Kuehn, B. N. Keel, A. K. Lindholm-Perry, and J. K. Pickrell. 2020. Assessment of imputation from low-pass sequencing to predict merit of beef steers. Genes. 11(11):1312. doi:10.3390/genes11111312.

Snelling, W. M., R. M. Thallman, M. L. Spangler, and L. A. Kuehn. 2022. Breeding Sustainable Beef Cows: Reducing Weight and Increasing Productivity. Animals. 12:1745. doi:10.3390/ani12141745.

Veerkamp, R. F., A. C. Bouwman, C. Schrooten, and M. P. L. Calus. 2016. Genomic prediction using preselected DNA variants from a GWAS with whole-genome

sequence data in Holstein–Friesian cattle. Genet. Sel. Evol. 48:95. doi:10.1186/s12711-016-0274-1.

Warburton, C. L., B. N. Engle, E. M. Ross, R. Costilla, S. S. Moore, N. J. Corbet, J. M. Allen, A. R. Laing, G. Fordyce, R. E. Lyons, M. R. McGowan, B. M. Burns, and B. J. Hayes. 2020. Use of whole-genome sequence data and novel genomic selection strategies to improve selection for age at puberty in tropically-adapted beef heifers. Genet. Sel. Evol. 52:28. doi:10.1186/s12711-020-00547-5.

Yang, J., T. Manolio, L. Pasquale, E. Boerwinkle, N. Caporaso, J. M. Cunningham, M. de Andrade, B. Feenstra, E. Feingold, M. G. Hayes, W. G. Hill, M. T. Landi, A. Alonso, G. Lettre, P. Lin, H. Ling, W. Lowe, R. A. Mathias, M. Melbye, E. Pugh, M. C. Cornelis, B. S. Weir, M. E. Goddard and P. M. Visscher. 2011. Genome partitioning of genetic variation for complex traits using common SNPs. Nat. Genet. 43:519–525. doi:10.1038/ng.823.

Zhang, F., Y. Wang, R. Mukiibi, L. Chen, M. Vinsky, G. Plastow, J. Basarab, P. Stothard, and C. Li. 2020. Genetic architecture of quantitative traits in beef cattle revealed by genome wide association studies of imputed whole genome sequence variants: I: feed efficiency and component traits. BMC Genomics. 21:36. doi:10.1186/s12864-019-6362-1.

Table 1. Estimates of variance components and heritability (±SE) for birth weight using different genomic relationship matrices based on predicted functional impact.

GRM[1]	Additive variance (kg^2)	Residual variance (kg^2)	Heritability (h^2)	# of variants
Chip	23.25±3.08	29.64±2.43	0.44±0.05	72,997
Full	23.03±3.13	29.95±2.49	0.43±0.05	461,806
Low	17.37±2.74	34.97±2.34	0.33±0.05	98,557
Modifier	20.49±2.83	32.30±2.26	0.39±0.05	183,823
Moderate	17.80±2.80	34.69±2.35	0.34±0.05	171,690
High	5.19±1.38	46.25±1.76	0.10±0.03	7,736
Chip & Full	25.42±3.29	27.78±2.57	0.48±0.05	531,983

[1]GRM: Genomic relationship matrices; Chip: GRM containing variants from the 100k array; Full: GRM containing a set of variants with varying predicted impact on gene function; Low: GRM containing subset of variants of low impact; Modifier: GRM containing subset of variants of low or varying impact; Moderate: GRM containing subset of variants of moderate impact; High: GRM containing subset of variants of high impact; Chip & Full: GRM containing a set of variants from the Chip and Full GRMs.

Table 2. Estimates of variance components and heritability (±SE) for post weaning gain using different genomic relationship matrices based on predicted functional impact.

GRM[1]	Additive variance (kg^2)	Residual variance (kg^2)	Heritability (h^2)	# of variants
Chip	383.66±52.70	495.97±42.11	0.44±0.05	72,997
Full	334.29±51.67	541.27±42.96	0.38±0.05	461,806
Low	245.19±45.83	621.82±41.02	0.28±0.05	98,557
Modifier	303.27±48.37	573.80±40.35	0.35±0.05	183,823
Moderate	277.40±45.78	591.38±39.61	0.32±0.05	171,690
High	85.95±24.14	768.73±30.43	0.10±0.03	7,736
Chip & Full	382.95±54.71	496.73±44.27	0.44±0.05	531,983

[1]GRM: Genomic relationship matrices; Chip: GRM containing variants from the 100k array; Full: GRM containing a set of variants with varying predicted impact on gene function; Low: GRM containing subset of variants of low impact; Modifier: GRM containing subset of variants of low or varying impact; Moderate: GRM containing subset of variants of moderate impact; High: GRM containing subset of variants of high impact; Chip & Full: GRM containing a set of variants from the Chip and Full GRMs.

Table 3. Estimates of variance components and heritability (±SE) for birth weight using different genomic relationship matrices based on classification of mutation.

GRM[1]	Additive variance (kg^2)	Residual variance (kg^2)	Heritability (h^2)	# of variants
Chip	23.25±3.08	29.64±2.43	0.44±0.05	72,997
Full	22.73±3.14	30.19±2.51	0.43±0.05	435,538
G1	20.22±2.77	32.58±2.21	0.38±0.05	138,364
G2	17.67±2.75	34.72±2.33	0.34±0.05	79,787
G3	16.67±2.79	35.71±2.39	0.32±0.05	147,908
G4	8.85±1.87	42.95±1.91	0.17±0.04	32,949
G5	14.26±2.35	37.97±2.08	0.27±0.04	31,440
G6	3.29±1.13	48.05±1.70	0.06±0.02	5,090
Chip & Full	25.30±3.30	27.88±2.58	0.48±0.05	505,715

[1]GRM: Genomic relationship matrices; Chip: GRM containing variants from the 100k array; Full: GRM containing a set of variants with varying predicted impact on gene function; G1: GRM containing subset of UTR variants; G2: GRM containing subset of synonymous variants; G3: GRM containing subset of missense variants; G4: GRM containing subset of non-coding transcript exon variants; G5: GRM containing subset of splice region variants; G6: GRM containing subset of start or stop variants; Chip & Full: GRM containing a set of variants from the Chip and Full GRMs.

Table 4. Estimates of variance components and heritability (±SE) for post-weaning gain using different genomic relationship matrices based on classification of mutation.

GRM[1]	Additive variance (kg^2)	Residual variance (kg^2)	Heritability (h^2)	# of variants
Chip	383.66±52.70	495.97±42.11	0.44±0.05	72,997
Full	331.06±52.05	544.40±43.52	0.38±0.05	435,538
G1	282.85±47.14	593.44±39.88	0.32±0.05	138,364
G2	241.79±45.71	625.40±40.98	0.28±0.05	79,787
G3	262.45±45.80	604.97±40.30	0.30±0.05	147,908
G4	150.70±32.98	711.42±33.23	0.17±0.04	32,949
G5	214.26±39.03	651.18±35.90	0.25±0.04	31,440
G6	42.71±19.07	809.43±29.67	0.05±0.02	5,090
Chip & Full	382.76±54.99	497.15±44.59	0.44±0.06	505,715

[1]GRM: Genomic relationship matrices; Chip: GRM containing variants from the 100k array; Full: GRM containing a set of variants with varying predicted impact on gene function; G1: GRM containing subset of UTR variants; G2: GRM containing subset of synonymous variants; G3: GRM containing subset of missense variants; G4: GRM containing subset of non-coding transcript exon variants; G5: GRM containing subset of splice region variants; G6: GRM containing subset of start or stop variants; Chip & Full: GRM containing a set of variants from the Chip and Full GRMs.

Table 5. Estimates of variance components and heritability (±SE) for birth weight using different combinations of genomic relationship matrices based on predicted functional impact and percent of additive variance explained by each category.

Model[1]	Additive Variance (kg^2)	Residual Variance (kg^2)	Heritability (h^2)	% V_a[2] Low	% V_a Modifier	% V_a Moderate	% V_a High	Total # of variants in model
Low & Modifier	21.36±3.00	31.48±2.42	0.40±0.05	15.1	84.9	-	-	282,380
Low & Moderate	19.98±2.98	32.64±2.46	0.38±0.05	51.6	-	48.4	-	270,247
Low & High	17.47±2.76	34.87±2.35	0.33±0.05	97.1	-	-	2.9	106,293
Modifier & Moderate	22.27±3.05	30.68±2.43	0.42±0.05	-	77.6	22.4	-	355,513
Modifier & High	20.62±2.86	32.17±2.29	0.39±0.05	-	98.3	-	1.7	191,559
Moderate & High	18.02±2.82	34.50±2.36	0.34±0.05	-	-	95.3	4.7	179,426
Low, Modifier & Moderate	22.40±3.10	30.55±2.48	0.42±0.05	4.5	75.1	20.4	-	454,070
Low, Modifier & High	21.37±3.01	31.46±2.43	0.40±0.05	14.8	84.8	-	0.4	290,116
Low, Moderate & High	19.99±2.99	32.63±2.46	0.38±0.05	51.4	-	48.3	0.3	277,983
Modifier, Moderate & High	22.27±3.29	30.68±2.43	0.42±0.05	-	77.6	22.4	~0	363,249
All	22.40±3.10	30.55±2.48	0.42±0.05	4.5	75.1	20.4	~0	461,806

[1]Models contain multiple GRM each fit as random effects; Low: GRM containing subset of 98,557 variants of low impact; Modifier: GRM containing subset of 183,823 variants of low or varying impact; Moderate: GRM containing subset of 171,690 variants of moderate impact; High: GRM containing subset of 7,736 variants of high impact; All: Model including four GRM corresponding to the four variant subsets.

[2]V_a: Additive Variance.

Table 6. Estimates of variance components and heritability (±SE) for post-weaning gain using different combinations of genomic relationship matrices based on predicted functional impact and percent of additive variance explained by each category.

Model[1]	Additive Variance (kg^2)	Residual Variance (kg^2)	Heritability (h^2)	% V_a [2] Low	% V_a Modifier	% V_a Moderate	% V_a High	Total # of variants in model
Low & Modifier	315.97±49.13	560.95±40.77	0.36±0.05	47.4	52.6	-	-	282,380
Low & Moderate	295.88±49.04	574.41±42.05	0.34±0.05	24.3	-	75.7	-	270,247
Low & High	251.04±46.24	615.98±41.20	0.29±0.05	88.2	-	-	11.8	106,293
Modifier & Moderate	330.46±50.63	545.15±42.00	0.38±0.05	-	54.7	45.3	-	355,513
Modifier & High	308.76±48.79	568.08±40.69	0.35±0.05	-	92.6	-	7.4	191,559
Moderate & High	280.32±46.15	588.62±39.85	0.32±0.05	-	-	95.2	4.8	179,426
Low, Modifier & Moderate	330.46±50.63	545.16±42.00	0.38±0.05	~0	54.7	45.3	-	454,070
Low, Modifier & High	316.99±50.87	559.76±42.70	0.36±0.05	14.2	79.7	-	6.1	290,116
Low, Moderate & High	297.35±49.20	573.02±42.16	0.34±0.05	23.1	-	73.5	3.4	277,983
Modifier, Moderate & High	331.73±50.77	543.93±42.10	0.38±0.05	-	54.0	43.5	2.5	363,249
All	331.73±50.77	543.93±42.10	0.38±0.05	~0	54.0	43.5	2.5	461,806

[1]Models contain multiple GRM each fit as random effects; Low: GRM containing subset of 98,557 variants of low impact; Modifier: GRM containing subset of 183,823 variants of low or varying impact; Moderate: GRM containing subset of 171,690 variants of moderate impact; High: GRM containing subset of 7,736 variants of high impact; All: Model including four GRM corresponding to the four variant subsets.

[2]V_a: Additive Variance.

Table 7. Estimates of variance components and heritability (±SE) for birth weight using combinations of two genomic relationship matrices based on classification of mutation and percent of additive variance explained by each category.

Model	Additive Variance (kg^2)	Residual Variance (kg^2)	Heritability (h^2)	% V_a G1	% V_a G2	% V_a G3	% V_a G4	% V_a G5	% V_a G6	Total # of variants in model
G1,2	21.30±2.97	31.54±2.40	0.40±0.05	82.4	17.6	-	-	-	-	218,151
G1,3	21.15±2.99	31.72±2.42	0.40±0.05	89.1	-	10.9	-	-	-	286,272
G1,4	20.59±2.82	32.21±2.26	0.39±0.05	93.9	-	-	6.1	-	-	171,313
G1,5	21.38±2.90	31.53±2.30	0.40±0.05	82.1	-	-	-	17.9	-	169,804
G1,6	20.22±2.73	32.58±2.21	0.38±0.05	~1	-	-	-	-	~0	143,454
G2,3	19.29±2.96	33.27±2.47	0.37±0.05	-	71.3	28.7	-	-	-	227,695
G2,4	18.43±2.82	34.02±2.37	0.35±0.05	-	85.9	-	14.1	-	-	112,736
G2,5	18.90±2.85	33.64±2.37	0.36±0.05	-	64.2	-	-	35.8	-	111,227
G2,6	17.67±2.75	34.72±2.33	0.34±0.05	-	~1	-	-	-	~0	84,877
G3,4	17.17±2.82	35.20±2.41	0.33±0.05	-	-	81.2	18.8	-	-	180,857
G3,5	18.11±2.86	34.42±2.41	0.34±0.05	-	-	47.9	-	52.1	-	179,348
G3,6	16.67±2.79	35.71±2.39	0.32±0.05	-	-	~1	-	-	~0	152,998
G4,5	15.23±2.47	37.01±2.15	0.29±0.04	-	-	-	19.8	80.2	-	64,389
G4,6	9.42±1.97	42.37±1.97	0.18±0.04	-	-	-	86.9	-	13.1	38,039
G5,6	14.40±2.39	37.83±2.12	0.28±0.04	-	-	-	-	97.5	2.5	36,530

[1]Models contain multiple GRM each fit as random effects; Full: GRM containing a set of variants with varying predicted impact on gene function; G1: GRM containing subset of UTR variants; G2: GRM containing subset of synonymous variants; G3: GRM containing subset of missense variants; G4: GRM containing subset of non-coding transcript exon variants; G5: GRM containing subset of splice region variants; G6: GRM containing subset of start or stop variants.
[2]V_a: Additive Variance.

Table 8. Estimates of variance components and heritability (±SE) for birth weight using combinations of three genomic relationship matrices based on classification of mutation and percent of additive variance explained by each category.

Model[1]	Additive Variance (kg^2)	Residual Variance (kg^2)	Heritability (h^2)	% V_a[2] G1	% V_a G2	% V_a G3	% V_a G4	% V_a G5	% V_a G6	Total # of variants in model
G1,2,3	21.56±3.06	31.30±2.47	0.41±0.05	80.4	15.0	4.6	-	-	-	366,059
G1,2,4	21.52±3.00	31.33±2.42	0.41±0.05	79.4	15.9	-	4.7	-	-	251,100
G1,2,5	21.79±3.01	31.13±2.41	0.41±0.05	76.0	9.5	-	-	14.5	-	249,591
G1,2,6	21.79±3.01	31.13±2.41	0.41±0.05	76.0	9.5	-	-	-	14.5	223,241
G1,3,4	21.28±3.01	31.58±2.43	0.40±0.05	86.5	-	8.6	4.9	-	-	319,221
G1,3,5	21.63±3.03	31.30±2.43	0.41±0.05	79.9	-	3.6	-	16.5	-	317,712
G1,3,6	21.15±2.99	31.72±2.42	0.40±0.05	89.1	-	10.9	-	-	~0	291,362
G1,4,5	21.54±2.93	31.37±2.32	0.41±0.05	79.9	-	-	3.8	16.3	-	202,753
G1,4,6	20.59±2.82	32.21±2.26	0.39±0.05	93.9	-	-	6.1	-	~0	176,403
G1,5,6	21.38±2.90	31.53±2.30	0.40±0.05	82.1	-	-	-	17.9	~0	174,894
G2,3,4	19.59±2.98	32.98±2.48	0.37±0.05	-	66.5	22.6	10.9	-	-	260,644
G2,3,5	19.73±2.99	32.89±2.48	0.38±0.05	-	53.1	16.2	-	30.7	-	259,135
G2,3,6	19.29±2.96	33.27±2.47	0.37±0.05	-	71.3	28.7	-	-	~0	232,785
G2,4,5	19.28±2.88	33.28±2.39	0.37±0.05	-	59.4	-	9.2	31.4	-	144,176
G2,4,6	18.43±2.82	34.02±2.37	0.35±0.05	-	85.9	-	14.1	-	~0	117,826
G2,5,6	18.90±2.85	33.64±2.37	0.36±0.05	-	64.2	-	-	35.8	~0	116,317
G3,4,5	18.32±2.88	34.18±2.42	0.35±0.05	-	-	41.8	10.8	47.4	-	212,297
G3,4,6	17.17±2.82	35.20±2.41	0.33±0.05	-	-	81.2	18.8	-	~0	185,947
G3,5,6	18.11±2.86	34.42±2.41	0.34±0.05	-	-	47.9	-	52.1	~0	184,438
G4,5,6	15.28±2.49	36.97±2.18	0.29±0.04	-	-	-	19.6	79.6	0.8	69,479

[1]Models contain multiple GRM each fit as random effects; Full: GRM containing a set of variants with varying predicted impact on gene function; G1: GRM containing subset of UTR variants; G2: GRM containing subset of synonymous variants; G3: GRM containing subset of missense variants; G4: GRM containing subset of non-coding transcript exon variants; G5: GRM containing subset of splice region variants; G6: GRM containing subset of start or stop variants.

[2]V_a: Additive Variance.

Table 9. Estimates of variance components and heritability (±SE) for birth weight using combinations of four or more genomic relationship matrices based on classification of mutation and percent of additive variance explained by each category.

Model[1]	Additive Variance (kg^2)	Residual Variance (kg^2)	Heritability (h^2)	% V_a[2] G1	% V_a G2	% V_a G3	% V_a G4	% V_a G5	% V_a G6	Total # of variants in model
G1,2,3,4	21.67±3.07	31.19±2.49	0.41±0.05	78.4	14.3	2.8	4.5	-	-	399,008
G1,2,3,5	21.83±3.07	31.10±2.48	0.41±0.05	75.7	9.1	0.8	-	14.4	-	397,499
G1,2,3,6	21.56±3.06	31.30±2.47	0.41±0.05	80.4	15.0	4.6	-	-	~0	371,149
G1,2,4,5	21.92±3.03	31.01±2.43	0.41±0.05	74.4	8.8	-	3.4	13.4	-	282,540
G1,2,4,6	21.52±3.00	31.33±2.42	0.41±0.05	79.4	15.8	-	4.7	-	~0	256,190
G1,2,5,6	21.79±3.01	31.13±2.41	0.41±0.05	76.0	9.5	-	-	14.5	~0	254,681
G1,3,4,5	21.70±3.04	31.22±2.44	0.41±0.05	78.5	-	2.5	3.5	15.5	-	350,661
G1,3,4,6	21.28±3.01	31.58±2.43	0.40±0.05	86.5	-	8.6	4.9	-	~0	324,311
G1,3,5,6	21.63±3.03	31.30±2.43	0.41±0.05	79.9	-	3.6	-	16.5	~0	322,802
G1,4,5,6	21.54±2.93	31.37±2.32	0.41±0.05	79.9	-	-	3.8	16.3	~0	207,843
G2,3,4,5	19.91±3.00	32.70±2.49	0.38±0.05	-	51.2	13.0	8.0	27.8	-	292,084
G2,3,4,6	19.59±2.98	32.98±2.48	0.37±0.05	-	66.5	22.6	10.9	-	~0	265,734
G2,3,5,6	19.73±2.99	32.89±2.48	0.38±0.05	-	53.1	16.2	-	30.7	~0	264,225
G2,4,5,6	19.28±2.88	33.28±2.39	0.37±0.05	-	59.4	-	9.2	31.4	~0	149,266
G3,4,5,6	18.32±2.88	34.18±2.42	0.35±0.05	-	-	41.8	10.8	47.4	~0	217,387
G1,2,3,4,5	21.91±3.03	31.01±2.43	0.41±0.05	74.4	8.8	~0	3.4	13.4	-	430,448
G1,2,3,4,6	21.67±3.07	31.19±2.49	0.41±0.05	78.4	14.3	2.8	4.5	-	~0	404,098
G1,2,3,5,6	21.83±3.07	31.10±2.48	0.41±0.05	75.7	9.1	0.8	-	14.4	~0	402,589
G1,2,4,5,6	21.91±3.03	31.01±2.43	0.41±0.05	74.4	8.8	-	3.4	13.4	~0	287,630
G1,3,4,5,6	21.70±3.04	31.23±2.44	0.41±0.05	78.5	-	2.5	3.5	15.5	~0	355,751
G2,3,4,5,6	19.91±3.00	32.71±2.49	0.38±0.05	-	51.2	13.0	8.0	27.8	~0	297,174
All	21.91±3.03	31.01±2.43	0.41±0.05	74.4	8.8	~0	3.4	13.4	~0	435,538

[1]Models contain multiple GRM each fit as random effects; Full: GRM containing a set of variants with varying predicted impact on gene function; G1: GRM containing subset of UTR variants; G2: GRM containing subset of synonymous variants; G3: GRM containing subset of missense variants; G4: GRM containing subset of non-coding transcript exon variants; G5: GRM containing subset of splice region variants; G6: GRM containing subset of start or stop variants; All: Model including six GRM corresponding to the six variant subsets.

[2]V_a: Additive Variance.

Table 10. Estimates of variance components and heritability (±SE) for post-weaning gain using combinations of two genomic relationship matrices based on classification of mutation and percent of additive variance explained by each category.

Model[1]	Additive Variance (kg^2)	Residual Variance (kg^2)	Heritability (h^2)	% V_a[2] G1	% V_a G2	% V_a G3	% V_a G4	% V_a G5	% V_a G6	Total # of variants in model
G1,2	303.41±50.53	572.93±42.76	0.35±0.05	72.5	27.5	-	-	-	-	218,151
G1,3	314.95±50.39	560.09±42.33	0.36±0.05	56.3	-	43.7	-	-	-	286,272
G1,4	301.48±48.52	575.29±40.60	0.34±0.05	78.4	-	-	21.6	-	-	171,313
G1,5	295.51±48.21	579.35±40.68	0.34±0.05	64.4	-	-	-	35.6	-	169,804
G1,6	282.84±47.18	593.42±39.91	0.32±0.05	~1	-	-	-	-	~0	143,454
G2,3	290.89±49.45	579.36±42.63	0.33±0.05	-	35.2	64.8	-	-	-	227,695
G2,4	258.72±46.72	609.51±41.32	0.30±0.05	-	71.2	-	28.8	-	-	112,736
G2,5	259.05±46.84	609.48±41.38	0.30±0.05	-	43.8	-	-	56.2	-	111,227
G2,6	241.79±45.69	625.40±40.98	0.28±0.05	-	~1	-	-	-	~0	84,877
G3,4	280.91±46.98	587.96±40.75	0.32±0.05	-	-	76.6	23.4	-	-	180,857
G3,5	279.77±46.79	589.04±40.65	0.32±0.05	-	-	58.2	-	41.8	-	179,348
G3,6	262.44±45.79	604.97±40.30	0.30±0.05	-	-	~1	-	-	~0	152,998
G4,5	255.10±42.58	613.93±37.39	0.29±0.04	-	-	-	32.3	67.7	-	64,389
G4,6	154.64±34.48	707.35±34.49	0.18±0.04	-	-	-	94.6	-	5.4	38,039
G5,6	215.15±39.86	650.33±36.66	0.25±0.04	-	-	-	-	99.1	0.9	36,530

[1]Models contain multiple GRM each fit as random effects; Full: GRM containing a set of variants with varying predicted impact on gene function; G1: GRM containing subset of UTR variants; G2: GRM containing subset of synonymous variants; G3: GRM containing subset of missense variants; G4: GRM containing subset of non-coding transcript exon variants; G5: GRM containing subset of splice region variants; G6: GRM containing subset of start or stop variants.

[2]V_a: Additive Variance.

Table 11. Estimates of variance components and heritability (±SE) for post-weaning gain using combinations of three genomic relationship matrices based on classification of mutation and percent of additive variance explained by each category.

Model[1]	Additive Variance (kg^2)	Residual Variance (kg^2)	Heritability (h^2)	% V_a[2] G1	% V_a G2	% V_a G3	% V_a G4	% V_a G5	% V_a G6	Total # of variants in model
G1,2,3	320.68±51.79	554.74±43.54	0.37±0.05	50.5	10.4	39.1	-	-	-	366,059
G1,2,4	312.26±50.91	564.39±42.89	0.36±0.05	64.6	17.0	-	18.4	-	-	251,100
G1,2,5	303.08±50.50	572.06±42.85	0.35±0.05	57.8	11.4	-	-	30.8	-	249,591
G1,2,6	303.41±50.53	572.93±42.76	0.35±0.05	72.5	27.5	-	-	-	~0	223,241
G1,3,4	324.68±50.94	550.90±42.57	0.37±0.05	48.3	-	36.1	15.6	-	-	319,221
G1,3,5	316.50±50.48	557.95±42.43	0.36±0.05	44.0	-	33.2	-	22.8	-	317,712
G1,3,6	314.95±50.39	560.09±42.33	0.36±0.05	56.3	-	43.7	-	-	~0	291,362
G1,4,5	311.17±49.21	564.49±41.14	0.36±0.05	49.8	-	-	19.1	31.1	-	202,753
G1,4,6	301.48±48.52	575.29±40.60	0.34±0.05	78.4	-	-	21.6	-	~0	176,403
G1,5,6	295.51±48.21	579.36±40.69	0.34±0.05	64.4	-	-	-	35.6	~0	174,894
G2,3,4	299.54±49.87	571.18±42.75	0.34±0.05	-	24.9	56.5	18.6	-	-	260,644
G2,3,5	291.84±49.45	578.24±42.67	0.34±0.05	-	16.5	49.5	-	34.0	-	259,135
G2,3,6	290.89±49.45	579.36±42.63	0.33±0.05	-	35.2	64.8	-	-	~0	232,785
G2,4,5	274.63±47.57	595.46±41.52	0.32±0.05	-	22.7	-	26.0	51.3	-	144,176
G2,4,6	258.72±46.72	609.51±41.32	0.30±0.05	-	71.2	-	28.8	-	~0	117,826
G2,5,6	259.05±46.84	609.48±41.38	0.30±0.05	-	43.8	-	-	56.2	~0	116,317
G3,4,5	296.37±47.81	574.18±41.01	0.34±0.05	-	-	43.3	20.2	36.5	-	212,297
G3,4,6	280.91±46.98	587.96±40.75	0.32±0.05	-	-	76.6	23.4	-	~0	185,947
G3,5,6	279.77±46.79	589.04±40.65	0.32±0.05	-	-	58.2	-	41.8	~0	184,438
G4,5,6	255.10±42.58	613.93±37.39	0.29±0.04	-	-	-	32.3	67.7	~0	69,479

[1]Models contain multiple GRM each fit as random effects; Full: GRM containing a set of variants with varying predicted impact on gene function; G1: GRM containing subset of UTR variants; G2: GRM containing subset of synonymous variants; G3: GRM containing subset of missense variants; G4: GRM containing subset of non-coding transcript exon variants; G5: GRM containing subset of splice region variants; G6: GRM containing subset of start or stop variants.

[2]V_a: Additive Variance.

Table 12. Estimates of variance components and heritability (±SE) for post-weaning gain using combinations of four or more genomic relationship matrices based on classification of mutation and percent of additive variance explained by each category.

Model[1]	Additive Variance (kg^2)	Residual Variance (kg^2)	Heritability (h^2)	% V_a[2] G1	% V_a G2	% V_a G3	% V_a G4	% V_a G5	% V_a G6	Total # of variants in model
G1,2,3,4	326.72±51.98	549.10±43.58	0.37±0.05	46.3	4.0	34.6	15.1	-	-	399,008
G1,2,3,5	317.44±51.61	557.09±43.56	0.36±0.05	43.3	1.7	32.7	-	22.3	-	397,499
G1,2,3,6	320.68±51.79	554.74±43.54	0.37±0.05	50.5	10.4	39.1	-	-	~0	371,149
G1,2,4,5	311.58±50.80	564.02±42.86	0.36±0.05	49.5	0.7	-	19.0	30.8	-	282,540
G1,2,4,6	312.26±50.91	564.39±42.89	0.36±0.05	64.6	17.0	-	18.4	-	~0	256,190
G1,2,5,6	303.09±50.50	572.07±42.85	0.35±0.05	57.8	11.4	-	-	30.8	~0	254,681
G1,3,4,5	326.01±50.98	549.15±42.60	0.37±0.05	36.5	-	26.2	15.4	21.9	-	350,661
G1,3,4,6	324.68±50.94	550.90±42.57	0.37±0.05	48.3	-	36.1	15.6	-	~0	324,311
G1,3,5,6	316.50±50.48	557.96±42.43	0.36±0.05	44.0	-	33.2	-	22.8	~0	322,802
G1,4,5,6	311.18±49.21	564.50±41.11	0.36±0.05	49.8	-	-	19.1	31.1	~0	207,843
G2,3,4,5	300.41±49.78	570.47±42.73	0.35±0.05	-	6.0	40.9	19.2	33.9	-	292,084
G2,3,4,6	299.54±49.87	571.18±42.75	0.34±0.05	-	24.9	56.5	18.6	-	~0	265,734
G2,3,5,6	291.84±49.45	578.24±42.67	0.34±0.05	-	16.5	49.5	-	34.0	~0	264,225
G2,4,5,6	274.63±47.57	595.46±41.52	0.32±0.05	-	22.7	-	26.0	51.3	~0	149,266
G3,4,5,6	296.37±47.81	574.18±41.01	0.34±0.05	-	-	43.3	20.2	36.5	~0	217,387
G1,2,3,4,5	326.01±50.97	549.16±42.60	0.37±0.05	36.5	~0	26.2	15.4	21.9	-	430,448
G1,2,3,4,6	326.68±51.99	549.01±43.57	0.37±0.05	46.3	4.0	34.6	15.1	-	~0	404,098
G1,2,3,5,6	317.44±51.61	557.10±43.56	0.36±0.05	43.3	1.7	32.7	-	22.3	~0	402,589
G1,2,4,5,6	311.59±50.79	564.09±42.86	0.36±0.05	49.5	0.7	-	19.0	30.8	~0	287,630
G1,3,4,5,6	326.01±50.97	549.16±42.60	0.37±0.05	36.5	-	26.2	15.4	21.9	~0	355,751
G2,3,4,5,6	300.41±49.78	570.50±42.70	0.34±0.05	-	6.0	40.9	19.2	33.9	~0	297,174
All	326.01±50.97	549.16±42.60	0.37±0.05	36.5	~0	26.2	15.4	21.9	~0	435,538

[1]Models contain multiple GRM each fit as random effects; Full: GRM containing a set of variants with varying predicted impact on gene function; G1: GRM containing subset of UTR variants; G2: GRM containing subset of synonymous variants; G3: GRM containing subset of missense variants; G4: GRM containing subset of non-coding transcript exon variants; G5: GRM containing subset of splice region variants; G6: GRM containing subset of start or stop variants; All: Model including six GRM corresponding to the six variant subsets.
[2]V_a: Additive Variance.