8-2016

Genome-wide association study for feed efficiency traits using SNP and haplotype models

Kashly Rene Schweer
University of Nebraska-Lincoln, kash@wbsnet.org

GENOME-WIDE ASSOCIATION STUDY FOR FEED EFFICIENCY TRAITS USING SNP AND HAPLOTYPE MODELS

By

Kashly Rene Schweer

A THESIS

Presented to the Faculty of

The Graduate College of 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

August, 2016

GENOME-WIDE ASSOCIATION STUDY FOR FEED EFFICIENCY TRAITS USING SNP AND HAPLOTYPE MODELS

Kashly Rene Schweer, M.S.

University of Nebraska, 2016

Advisor: Matthew L. Spangler

Feed efficiency has long been recognized for its economic importance in beef cattle production systems. Feed intake is not a routinely collected trait on-farm due to the expense of measuring individual feed intake. Identification of genomic regions associated with feed efficiency and its component traits would be advantageous for selection programs. The objective of this study was to uncover genomic regions associated with average daily gain (ADG) and average daily feed intake (ADFI) through genome-wide association studies (GWAS) with SNP and haplotype Bayesian models. Phenotypes of ADG and ADFI along with genotypes from the BovineSNP50v2 array from crossbred steers and heifers (n=748) representing seven sire breeds were used in the analyses. Traits were analyzed independently through SNP (BayesC) and haplotype models (BayesIM) and together in a bivariate haplotype model. For the BayesIM model, haplotypes were mapped to haplotype clusters based on local haplotype similarity using a hidden Markov model (HMM) and the haplotype clusters were fitted as covariates in the Bayesian mixture model. The number of haplotype clusters at each location was assumed to be either 8 (BayesIM8) or 16 (BayesIM16), resulting in a total of three univariate analyses for each trait and two bivariate analyses. Posterior genomic heritability estimates (PSD) for ADG were 0.28 (0.08), 0.46 (0.12), 0.46 (0.12), 0.44 (0.10) and 0.44 (0.12) for BayesC, BayesIM8, BayesIM16, BayesIM8 bivariate and BayesIM16 bivariate,

respectively. Average daily feed intake posterior genomic heritability estimates (PSD) were 0.30 (0.07), 0.46 (0.10), 0.45 (0.09), 0.45 (0.09) and 0.46 (0.11) for the same analyses. Top genomic regions across all univariate analyses for ADG exist on BTA 3, 4, 13, 20 and 22 while ADFI was associated with regions on BTA 1, 13 and 21. Potential pleiotropic genomic regions associated with both traits were identified through bivariate analyses on BTA 1, 13, 20 and 25. The use of multi-trait models for GWAS is currently unexplored in beef cattle. A multi-trait model can use information from one trait to inform the other when the traits are genetically correlated.

Acknowledgements

There is truly no environment parallel to that within the Animal Breeding and Genetics Department at the University of Nebraska-Lincoln. The vast amount of knowledge and experience within its halls is unmatched. It was a privilege to be a part of this group for two years while completing my M.S. I would like to thank my advisor, Dr. Matt Spangler, for bringing me into this atmosphere for my graduate studies and opening my eyes to what is possible within the realm of genomics. Sincere thanks are extended to my committee members, Dr. Steve Kachman and Dr. Larry Kuehn, for their time and dedication to my education and thesis project. I am also very grateful for Sherri Pitchie's work with travel arrangements and meeting registrations. Aside from the immeasurable expertise of the faculty, the department is filled with numerous bright students. Their friendship, advice and support are truly appreciated.

This opportunity would not have been attainable without the love and support of my family. Mom and Dad, you have pushed me to achieve things in life that I didn't think were possible. I cannot list all the things you have done for me over the years. I am grateful for every sacrifice you have made in order for me to pursue my dreams. Thanks also, to my sisters, who have served as role models since I was a child and to continue to do so today. And to Jarek, thank you for the constant encouragement and for sharing all of life's moments with me.

Table of Contents

Introduction

Feed costs comprise the majority of variable expenses in beef cattle systems making feed efficiency an important economic consideration within the beef industry (Koch et al., 1963; Dickerson et al., 1974). Aside from the direct economic impact of this trait complex at the individual producer level, the projections of global population growth provide extra pressure for efficient beef cattle production as producers try to combat the growing food demand with limited resources (Eggen, 2012). Improved feed efficiency also has an environmental impact through a decreased carbon footprint as more efficient cattle have fewer days to finish, emitting less methane throughout their lifetime (Freetly and Brown-Brandl, 2013).

There are multiple measures of feed efficiency. The most common used in the fed cattle sector is feed conversion ratio (FCR), the ratio of feed to gain (F:G), or gain to feed (G:F). This ratio is simply the raw pounds of feed required for raw pounds of weight gained, or the reciprocal. It makes no adjustments for age and weight differences of the cattle or energy content differences of the diet being fed. For these reasons, unadjusted FCR should be limited to use within contemporary groups. Due to the positive genetic correlation between feed intake and gain, selection to improve FCR has the potential to lead to larger, more maintenance intensive animals in the breeding herd (Archer et al., 1999).

One proposed alternative to FCR is residual feed intake (RFI). The concept of RFI was introduced by Koch et al. (1963) by suggesting that feed intake should be adjusted for body weight and weight gain, making RFI the difference between actual feed intake and the predicted feed intake of an animal based its requirements for maintenance and

gain. More desirable or efficient animals will have a negative RFI value with an average individual having an RFI of zero (Koch et al., 1963; Archer et al., 1999). Predicted feed intake is calculated by regressing actual feed intake on ADG and metabolic weight within a contemporary group, meaning the sum of RFI values across the contemporary group in which it was calculated should equal zero and thus contemporary group definition becomes vital. It is sometimes considered the preferred definition of feed efficiency because RFI is phenotypically independent of production traits (e.g. growth and body weight) used in the prediction equation (Kennedy et al., 1993). Ultimately, selection on RFI is equivalent to using a restricted selection index containing the component traits. Since genetic variation in RFI exists, genetic progress towards more efficient cattle through selection on this trait is possible.

The use of RFI as a measure of feed efficiency is occasionally contested for a variety of reasons including difficult interpretation, differences in the frequency of recording for the component traits, and the fact that it is a restricted selection index and not an economically relevant trait. Additionally, if any genetic or residual correlations exist between feed intake and maintenance traits, the resulting heritability estimates can be flawed (Lu et al., 2015; Kennedy et al., 1993). In the dairy industry, Lu et al. (2015) proposed a multi-trait model as an alternative approach to feed efficiency. This may represent a more robust measure of feed efficiency and comprehensive investigation into the genetic relationship between intake and gain.

Feed intake, and consequently feed efficiency traits, are difficult to obtain and expensive to measure. Therefore a genomics approach seems warranted. Although it is an

expensive initiative, the detection of genetic markers for feed efficiency has the potential for great returns in the beef industry.

Literature Review

Genetic parameters for feed efficiency traits

Moderate heritability estimates for average daily gain (ADG), dry matter intake (DMI), metabolic mid-test body weight (MMBW; $kg^{0.75}$) and RFI suggest genetic variation exists and genetic progress can be garnered. Average daily gain is defined as the difference between the start and end test weights divided by the total number of days on feed. Arthur et al. (2001a) used data from 1,180 young Angus bulls and heifers on performance tests to estimate genetic and phenotypic parameters. Direct heritability of ADG was estimated as 0.28 (Arthur et al., 2001a). Arthur et al. (2001b) estimated the heritability of ADG in Charolais bulls between 15 and 19 months of age to be 0.34 and 0.41, respectively. These estimates are similar to previous reports from Robinson and Oddy (2004) and Schenkel et al. (2003) of 0.23 and 0.35, respectively.

Dry matter intake is the cumulative on-test feed intake on a dry matter basis. Nkrumah et al. (2007a) estimated the heritability of daily DMI as 0.54 using crossbred beef steers, which is higher than a previous estimate of 0.44 by Schenkel et al. (2003). Feed intake can also be measured on an as-fed basis. Heritability estimates for feed intake as total feed consumed (as-fed) are also moderate with reports of 0.27, 0.48 and 0.39 from Robinson and Oddy (2004), Arthur et al. (2001b) and Arthur et al. (2001a), respectively.

Mid-test body weight (MBW) can be calculated by the average of the initial and end weights or through regression techniques. Metabolic mid-test body weight (MMBW)

is $MBW^{0.75}$ (Schenkel et al., 2003). Arthur et al. (2001a) reported the direct heritability estimate of 0.40 for MMBW. This result agrees with the estimates of 0.35 and 0.41 from Schenkel et al. (2003), and Robinson and Oddy (2004), respectively.

Direct heritability estimates are moderate for RFI. Arthur et al. (2001a) estimated a heritability of 0.39. Schenkel et al. (2003) used two definitions of RFI. The first was the classical definition of the trait, the difference between actual feed intake and expected feed intake required for body weight and growth (RFI_P), and the second included an adjustment for end of test backfat thickness (RFI_b). Heritability estimates for both versions of RFI were very similar at 0.38 and 0.39 for RFI_P and RFI_b, respectively. Robinson and Oddy (2004) reported heritability estimates much lower for RFI (0.18) when cattle from varying breed types (temperate and tropical) at near market-ready weights were fed an ad libitum feedlot diet. Estimates for other feed efficiency related traits including FCR, feeding time and number of eating sessions per day are also moderate (Robinson and Oddy, 2004; Herd and Bishop, 2000; Arthur et al., 2001a; Arthur et al., 2001b).

Two of the main causes for genetic correlations between traits are the existence of pleiotropy and linkage (Bolormaa et al., 2014). Genetic and phenotypic correlations exist among feed efficiency traits and between feed efficiency and production traits. Phenotypic and genetic correlations between MMBW and ADG were 0.24 and 0.53, respectively (Arthur et al., 2001a). Moderate-to-strong positive genetic correlations exist between ADG and feed intake (as-fed or dry matter basis) with estimates of 0.54 (Arthur et al., 2001a), 0.87 (Nkrumah et al., 2007a) and 0.50 (Schenkel et al., 2004). Several authors have reported moderate phenotypic correlations between gain and feed intake

ranging from 0.41 to 0.60 (Arthur et al., 2001a; Nkrumah et al., 2007a; Schenkel et al., 2004). Additionally, MMBW has been reported to be positively correlated with feed intake both phenotypically (r_p=0.77), and genetically (r_g=0.71) by Schenkel et al. (2004). By definition, RFI should be phenotypically independent of its component traits (Koch et al., 1963). Estimates from Arthur et al. (2001a) illustrate this independence with reported phenotypic correlations between RFI and MMBW (r_p=0.02) and between RFI and ADG (r_p=-0.06). Nkrumah et al. (2007) reported RFI was also genetically independent of its component traits, ADG and MMBW, with estimates close to zero (r_g=-0.04, r_g=-0.06). Nkrumah et al. (2007a) reported that the genetic correlation between RFI and feed intake was 0.72, while feed intake was genetically correlated with F:G to a lesser degree (r_g=0.31). Schenkel et al. (2003) also found RFI to be more strongly genetically correlated with feed intake than F:G, thus suggesting selecting for low RFI could decrease feed intake more substantially than selecting for FCR. This result is positive if the level of gain is not decreased as well.

Feed intake tends to be positively genetically correlated with postweaning growth traits including 200-d weight direct and 400-d weight direct with estimates of r_g =0.28 and 0.56, respectively (Arthur et al., 2001a). Additionally, RFI was negatively correlated with 200-d weight direct (r_g=-0.45) and 400-d weight direct (r_g=-0.26; Arthur et al., 2001a). Both FCR and RFI are negatively correlated with longissimus muscle area (LMA). This suggests more efficient cattle have larger LMA (Schnekel et al., 2003). More efficient cattle may also produce a leaner product, as RFI is genetically correlated with intra-muscular fat percentage (r_g=0.22) and rump fat (r_g=0.72) (Robinson and Oddy, 2004). Robinson and Oddy (2004) further investigated the association between RFI and

fat by holding age and carcass weight constant. Regardless of adjustment, the magnitude and sign of the relationships were similar.

Initial use of genomic information

Traditionally, selection of economically relevant traits in livestock has been based on phenotypic records and pedigree information. Estimated breeding value (EBV) is an estimate of the additive genetic merit for a given trait. It is expected that an individual will express one-half of its EBV in its offspring's performance, which is referred to as expected progeny difference (EPD). The calculation of the EBV combines pedigree information, the individual's own performance records and the performance records of one's offspring or relatives. Selection based on EBV has resulted in genetic change in National Cattle Evaluations (NCE). For animals to have EBV with high accuracy, many offspring with performance records are typically needed. If animals with high accuracy EBV are desired, the generation interval is the limiting factor on the rate of genetic progress. The concept of identifying genes to improve certain traits and selecting candidates for breeding based on the presence of favored alleles is advantageous (Goddard and Hayes, 2009). Molecular markers are used to identify the chromosomal regions that affect both single-gene traits and quantitative traits. Single-gene traits include genetic defects, coat color, horned/polled status, etc. and are often binary in nature. Genetic tests to identify animals with favorable markers have been successfully implemented into the beef industry. Due to quantitative traits being influenced by many loci, it is more difficult to identify the quantitative trait loci (QTL).

Molecular analysis of quantitative traits leads to the identification of two types of genetic loci. Direct markers code for the functional mutation and are the most preferred

type of loci for marker-assisted selection (MAS). However, these causative mutations are difficult to find and hard to validate. Presumed non-functional or indirect markers are the second type of genetic marker that can lead to the identification of QTL and are candidates for MAS. Indirect markers are abundant across the genome and can be identified by candidate-gene, fine-mapping or association approaches as linkage between the marker and QTL can be established (Dekkers, 2004). Linkage disequilibrium (LD) is the nonrandom relationship between alleles present at two or more loci; it reflects the recombination history of a specific haplotype within a population (Conner and Hartl, 2004). The extent of LD surrounding the causative mutation is dependent on differences in recombination, the effective population size, the age of the mutation and the population structure (Botstein and Risch, 2003; Cardon and Bell, 2001).

Linkage analysis was historically performed using microsatellite makers within families. However, the results of linkage analysis often involved large intervals for QTL positions which made the identification of the direct causative mutations difficult (Georges et al., 1995). Furthermore, linkage analysis was not efficient due to the degree of LD varying between families. This meant the linkage phase had to be determined within each family before the marker could be used for selection of a QTL (Goddard et al., 2010).

Two additional approaches have been used in the past to discover genes and polymorphisms that are responsible for genetic variation in livestock species. With the first approach, candidate genes were targeted based on their biological functions in relation to the trait. The second approach mapped genes of interest to a chromosomal location using genetic markers (Falconer and McKay, 1996; Anderson and Georges,

2004; Dekkers and Hospital, 2002; Van Laere et al., 2003). Much like linkage analysis, both of these approaches resulted in slow progress and lacked the ability to identify the specific causal genes or mutations (Goddard and Hayes, 2009).

The initial discovery of polymorphisms within genes affecting quantitative traits had little effect on the overall accuracy of breeding values. Phenotypic or pedigree selection is very effective when the heritability of a trait is high and the addition of genomic information gives way to little additional accuracy in the EBV. The proportion of genetic variance explained by a single gene is minimal, making the gain in response from selection based on individual genes small. Moreover, with many genes affecting a single trait, the effects of each gene are difficult to estimate accurately. Few genes are known that are directly responsible for large amounts of variation in economically relevant traits (Goddard, 2008). When indirect markers are used as the basis of selection, selection decisions are ultimately being made on the marker itself instead of the QTL directly. As LD decays between the marker and the QTL due to recombination, the response to selection decreases.

To counteract the difficulties of accounting for molecular information in selection decisions, Meuwissen et al. (2001) proposed the idea of genomic selection. Genomic selection refers to the use of genome-wide dense marker maps to estimate the total genetic value of a candidate for selection. Marker maps are made possible by the use of single nucleotide polymorphisms (SNP) that are easy to evaluate and interpret due to their bi-allelic nature (Fan et al., 2010). Sparse marker maps (15-50 cM spacing between markers) were introduced by Georges et al. (1995) and resulted in the detection of QTL. However, this method was found to be limited in outcrossing species due to the lack of

linkage between the marker and the QTL. This lends to a key advantage of the dense marker map. With more total SNP spaced across the genome, some markers are very close to QTL and possibly in LD with the QTL itself (Hastbacka et al., 1992). Close markers can be combined to form haplotype blocks that would carry the same QTL and can potentially be conserved over generations (Meuwissen et al., 2001).

Whole-genome sequencing of livestock species, HapMap projects and reduced representation library (RRL) studies uncovered many genetic variants, most of which were SNP (Fan et al., 2010). Approximately 2.2 million SNP were detected across the cattle genome (The Bovine HapMap Consortium, 2009). Detection of a massive quantity of SNP enabled the commercial production of genotyping platforms. Illumina's BeadArray and Affymetrix's GeneChip are two of the largest and most competitive producers of genotyping platforms. Genotyping platforms are developed in various sizes from low to high density (e.g., 3k, 50k, 770k).

The creation of high density genotyping platforms has enabled the implementation of genome wide association studies (GWAS; Fan et al., 2010). For GWAS, a sample from a population of animals with phenotypic information is chosen for genotyping. Statistical associations between the phenotypic data and genotypes yield estimations of genetic variance components and marker effects. Results from the analysis of complex traits through GWAS typically include several mutations each with a small effect. Large numbers of individuals are required for GWAS in order to have sufficient power to detect the genes responsible for the majority of the genetic variance for a complex trait (Goddard and Hayes, 2009).

Methods for genomic prediction

One of the primary hurdles which must be overcome with the current genomic advances is continuing to develop methodology to accurately estimate marker effects in a computationally efficient manner. The evolution of methodology is almost as vast as the changes in technology, or possibly parallel to some degree. Historically, BLUP has given animal breeders a powerful tool for the prediction of breeding value based on performance records. Henderson (1984) realized the advantages of prediction with BLUP over least squares, regressed least squares or selection index due to the reduction of error and the greater correlations between the predictors and the predictions. BLUP methodologies were augmented to include maker effects into breeding value predictions through mixed models for the introduction of genomic BLUP (GBLUP; Fernando and Grossman, 1989). With BLUP, pedigree information is used to derive the relationship matrix making full sibs have the same EBV when no individual or progeny data is included (parent average value). If instead, genomic information (i.e. SNP genotypes) is used to form the relationship matrix, the Mendelian Sampling term is partially predicted and allows for individual deviation from the parent average. This increases the accuracy of the EBV and consequently, the response to selection (Hayes et al., 2009). As with most linear predictions, GBLUP can assume that all markers contribute to the overall genetic variance therefore meaning each of the SNP have small effects (VanRaden, 2008).

Regression techniques were also used as a method for MAS where least square analysis was used as a way to estimate marker effects. Lande and Thompson (1990) found MAS to be feasible through the multiple regression of phenotype on genotype at a

given marker loci to determine markers associated with a given trait due to LD with the QTL. If a sufficient amount of markers are linked with the QTL and a large enough sample size of individuals exist, the regression should be able to account for most of the additive genetic variance within the trait due to a particular QTL (Lande and Thompson, 1990). In reality, the massive amount of marker effects cannot be estimated to be included into the standard regression model. With the advent of high-density genotyping platforms, the number of makers exceeds the number of genotyped individuals within the population. In order to overcome the challenge of dimensionality, only a subset of the marker effects are estimated to be included in the regression model and results in larger errors and poor estimates of the genetic value of an individual (Zhang and Smith, 1992; Whittaker et al., 2000).

Ridge regression is a method that has better predictive ability than when only a subset of markers can be used (Breiman, 1995) as more markers are able to be included with the estimates of marker effects shrunk towards zero by a constant factor (λ) known as the smoothing factor (Whittaker et al., 2000). However, Xu (2002) demonstrated that this penalty approach is not a valid method for QTL mapping when genome-dense SNP are used as it fails to capture the true essence of the data. Through a simulation study, Whittaker et al. (2000) showed ridge regression outperformed regression with a subset of marker effects estimated and traditional phenotypic selection.

Perhaps the most appealing method for genomic prediction lies within the "Bayesian alphabet." Bayesian analysis has captured the attention of animal breeders for a number of reasons. First, Bayesian procedures have the ability to handle situations where the number of markers exceeds the number of observations (Gianola et al., 2009).

Meuwissen et al. (2001) demonstrated how to make the transition from traditional BLUP estimation to analysis of marker effects with each SNP having a specific variance through Bayesian techniques. Bayesian analysis takes into account the degree of uncertainty revolving around each of the unknowns within the model (Gianola et al., 2009). Nonlinear equations, such as those within Bayesian methods, assume a prior distribution of SNP effects. This assumption may be a more realistic approach as the effects from each marker may not all contribute small effects as assumed with linear predictions. In fact, major genes may exist on some chromosomes therefore having corresponding markers that explain a greater amount of genetic variance (VanRaden, 2008).

In Bayesian models, all unknown parameters are treated as random variables each with its own distribution. Variables are further classified as observables or unobservables. The observable variables include the phenotype (y_i for i=1,…,n where n number of individuals) and the marker information. The QTL effect (b_j) and the variance of each marker effect (σ^2_j for j=1,…,p where p is the total number of markers) are considered unobservable. The distribution of the unobservable variables is referred to as the prior distribution, $f(\Theta)$. The distribution of the observable variables is a function of the unobservables; the likelihood function, $f(y|\Theta)$. The likelihood function represents the contribution of the phenotypic information to knowledge of the prior (Θ). The posterior distribution is the conditional distribution of the parameters given the observable variables or simply the combination of the likelihood function and the prior distribution, $f(\Theta|y)$. The Markov Chain Monte Carlo (MCMC) sampling technique draws samples from the posterior distribution to estimate the posterior means and variances (Xu, 2002; Gianola and Fernando, 1986).

Meuwissen et al. (2001) proposed two Bayesian methods at the advent of genomic selection; BayesA and BayesB. BayesA shares similarities with the previous regression models as it assumes that all markers have an effect and the prior distribution of the marker effects is normal with a marker-specific variance from a scaled inverse chi-square distribution. The normal distribution of SNP effects allows for some SNP to have larger effects than others, but with BayesA every SNP is treated as though it has a non-zero effect. However, if the number of QTL is substantially less than the number of markers and, given the multitude of SNP on current genotyping arrays, it seems logical to assume some markers will have no effect (Meuwissen et al., 2001; Goddard et al., 2010). For this reason, BayesB was introduced. BayesB allows for a proportion of the makers to have an effect (1-π) following a normal distribution and a proportion of the markers to have no effect (π). The proportion of markers that have no effect (π) is assumed *a priori*. The variance of the marker effects is sampled from a scaled inverse chi square distribution similar to the BayesA approach. One of the criticisms of BayesA and BayesB is the magnitude of influence the prior has on the shrinkage of the marker effects (Gianola et al., 2009). BayesC (Habier et al.,2011) assumes that every marker will not have an effect parallel to BayesB, but BayesC uses an equal variance for all SNP. BayesC can be extended by assuming that pi is not known and instead is estimated from the data (BayesCπ). Estimating π requires additional samples (Habier et al., 2011).

Methods of genomic prediction primarily exploit LD between SNP markers and QTL. Although SNP-based models offer promise to discover genomic regions associated with traits of interest, models utilizing haplotypes consisting of multiple SNP markers may provide greater power for association experiments. This is primarily justified as

haplotypes may be in greater LD with the QTL than the individual SNP marker. As the number of SNP markers within a chromosomal segment increase, the likelihood that identical haplotypes carried by different animals are identical by descent increases as well. Given haplotypes are identical by descent, QTL alleles would be conserved within the haplotype. Still, it is difficult to identify the number of haplotype states at a given loci and their corresponding frequency (Hayes, 2013).

Augmentation of molecular information into beef cattle evaluations

The augmentation of genomic information into NCE is critical for the progression of breeding programs. The advent of SNP panels to genotype large numbers of animals at a reasonable cost has made genomic prediction feasible. The QTL can be detected through LD with the markers, even though in practice the position of the QTL and the effects are not known. Summation of the product of the marker effects and SNP genotypes across all loci can estimate the breeding value of an individual based on markers effects only, or the molecular breeding value (MBV). The objective of this estimation focuses on the total genetic value of the animal instead of the precise discovery of QTL (Goddard et al., 2010).

Molecular breeding values have been augmented into NCE for the majority of popular beef cattle breeds. Genomically-enhanced expected progeny differences (GE-EPD) are calculated similarly to traditional EPDs with the addition of genomic test results. The way the genomic information is augmented into EPDs differs and can be divided into multi-step and single-step approaches. The multi-step approach requires the estimation of marker substitution effects, the prediction of the MBV and the combination of MBV with EPD. Single-step approaches include all phenotypic, pedigree and genomic

information by modifying the relationship matrix of the mixed model equations (Fernando and Garrick, 2013).

In order to augment genomic information into NCE for one succinct value for selection decisions, a method referred to as “blending” was proposed. Blending is an index-like approach that utilizes traditional matrix calculations to establish the weighting factors b from Pb=g. These weightings will differ for each trait according to the accuracy of the MBV and for each animal according to the EBV reliability (Garrick and Saatchi, 2013). For most breeds, blending is done post-evaluation and thus the MBV only influences the genotyped individual (Spangler, 2013).

Kachman (2008) introduced methodology to incorporate marker scores into NCE by integrating MBV as a correlated indicator trait in a multi-trait model. This approach was very adaptable for breed associations. Contrary to the blended approach, treating the MBV as a correlated trait had the ability for the genomic information to influence other animals in the pedigree that did not have genotypic data (Spangler, 2013). This approach was later adopted by MacNeil et al. (2010) for the American Angus Association to incorporate ultrasound data and MBV as indicator traits for predicting carcass EBV. Exploiting the knowledge of genetic correlations among traits and between traits and MBV allow for multiple sources of information to be used to predict hard to measure traits, such as carcass traits that can only be obtained after an animal is slaughtered.

Recently, a single-step approach to GBLUP has been adopted (Misztal, 2009). The single-step approach combines phenotypic, pedigree and SNP data in a single analysis. It creates the G relationship matrix using animals with genotypic data as with GBLUP and a sub-matrix using pedigree data for individuals without genotypic

information. It combines those matrices into a relationship matrix, H. Evidence shows that single-step GBLUP (ssGBLUP) produces improved accuracies over other genomic prediction methods due to the better estimation of genetic relationships (Misztal et al., 2013). It allows combined phenotypes from nongenotyped animal into the analysis. The limitations of ssGBLUP are the massive computing power needed. For large datasets, the inversion of the H matrix can be computationally expensive. Additionally, the G and A matrices need to be scaled appropriately (Misztal et al., 2013). Several breed associations are moving toward this single-step approach for their genetic analyses.

As with any method of genomic augmentation, animals with preexisting high accuracy EBV do not notice additional gains in accuracy by incorporating molecular information for a given trait. However, lowly accurate animals (i.e. those without progeny) do see gains in accuracy. The increase in accuracy through the incorporation of genomic information is directly related to the correlation between the phenotype and the MBV as the amount of genetic variance explained is equal to the square of the correlation. This is best illustrated by an example adopted from Spangler (2011) using results reported by MacNeil et al. (2010). Assuming the correlation (r) for marbling score was 0.37, 13.7% (r^2x100) of the additive genetic variance of marbling score was explained by the genomic test. Moreover, if the heritability of marbling is known to be moderate (h^2=0.3), the gain in accuracy for an animal with no ultrasound record or progeny information is now equivalent to the accuracy of having 5 progeny with carcass records in its pedigree or ultrasound information on the individual itself through the incorporation of genomic data (Spangler, 2011).

The primary justification for incorporating molecular information into traditional selection methods is the faster rate of genetic gain than could be achieved by phenotypic data alone (Meuwissen et al., 2001). Meuwissen and Goddard (1996) predicted 8-38% extra genetic gain through the incorporation of marker information into BLUP breeding values. Additional advantages of genomic selection include improved accuracy of young, unproven animals as selection candidates (Kachman et al., 2013) such as yearling bulls who have not produced offspring. In the dairy industry, it is estimated that the use of genomic selection with reduce the costs of bull testing by upwards of 90% (Eggen, 2012). The ability to make more accurate selection decisions at a younger age will in turn reduce the generation interval, speeding the rate of genetic progress as Meuwissen et al. (2001) anticipated. Within the beef industry, genomic predictors allow for selection of economically relevant traits that have phenotypes that are only expressed late in life, phenotypes that are expensive or difficult to measure, traits that are limited by sex, lowly heritable traits or phenotypes that can only be collected once the animal has been harvested (Dekkers, 2004; Bolormaa et al., 2013).

In regards to feed efficiency, GWAS is conducted as the association between genotypic data and measures of feed efficiency such as FCR, RFI or component traits including DMI or ADG. Multi-trait models have been proposed in the dairy (Lu et al., 2015) and swine (Strathe et al., 2014) industries as a more comprehensive investigation of feed efficiency. Lu et al. (2015) modeled DMI with energy sink traits including milk and MMBW. The classical calculation of RFI assumes relationships at the genetic and nongenetic levels are constant. The proposed multi-trait model allows these relationships to differ. Aside from the gains in genetic prediction accuracies with the multi-trait model

over RFI, the multi-trait model allows the inclusion of all animals, even those with missing records (Lu et al., 2015).

Given multi-trait models can be deployed for genetic merit prediction, it seems possible to use the same approach with GWAS. A multi-trait model for GWAS including intake and gain is currently an unexplored area in the beef cattle industry. The frequency of intake and gain phenotypes differ considerably with gain measured more routinely on-farm. Since a strong genetic correlation exists between the two traits, a bivariate model would exploit the knowledge of the highly recorded trait to inform the limited phenotype.

QTL associated with feed efficiency traits

Association studies help lead to the discovery of QTL. Nkrumah et al. (2007b) identified several putative QTL for feedlot growth, feed intake and feed efficiency using 400 steers across 20 half-sib families that were genotyped with 455 markers. Evidence for QTL associated with ADG was detected on 14 different chromosomes with the most significant QTL from the across-family analysis detected on BTA 6 and 19. Both these QTL were within the same chromosomal regions as growth QTL reported by Kneeland et al. (2004). The most significant QTL associated with ADG from the within-family analyses was located on BTA 5 at 67 cM, in the same region reported by Li et al. (2002). Across-family QTL affecting DMI were identified on 14 different chromosomes while the most significant QTL were detected on BTA 1, 18, 20 and 26 with the largest effect on BTA 18 at 49 cM. The within-family analyses of the same trait detected the most significant effect on BTA 14 at 101 cM. Two measures of residual feed intake, RFI_g and RFI_p, as described by Kennedy et al. (1993), were evaluated. One QTL on BTA 5 was significant for both RFI_g and RFI_p for the across-family analysis while the within-family

analyses detected QTL for RFI_p on chromosomes 5, 20 and 21 and QTL for RFI_g on chromosomes 2 and 28 (Nkrumah et al., 2007b).

Following the preliminary study by Nkrumah et al. (2007b), Sherman et al. (2009) aimed to refine QTL locations by using additional markers. A total of 2,194 markers were used for the identification and fine mapping of QTL for RFI, DMI and FCR across 24 cattle chromosomes, excluding BTA 2, 5, 10, 20 and 29 which were the focus of previous work (Moore et al., 2006; Sherman et al., 2008b). From the across-family analysis, the most significant QTL for RFI was on BTA 3 at 82 cM. Two QTL reached the genome-wise threshold for FCR on BTA 13 ($P < 0.05$) and 24 ($P < 0.001$) with greater significance detected on BTA 24 at 59 cM. The most significant QTL for DMI was detected on BTA 7 at 54 cM (Sherman et al., 2009). The within-family analysis confirmed the prior conclusion drawn by Nkrumah et al. (2007b) that the QTL location can be influenced by a small number of families that have highly significant QTL effects or the opposite can occur when many families can contribute moderate QTL effects. For example, the significant QTL on BTA 3 for RFI has two sire families with very significant effects at the location of the QTL. Alternatively, the FCR QTL on BTA 24 is influenced by 6 of the 9 families having a significant effect at the QTL location. By comparing both the across-family and within-family analyses, the QTL interval can be narrowed. Fifteen QTL were confirmed from the original genome scan (Nkrumah et al., 2007b) with the QTL intervals being narrowed on average by 21.9 cM. The most consistent QTL for RFI are located on BTA 1, 7, 18 and 19 (Nkrumah et al., 2007b; Barendse et al., 2007; Sherman et al., 2009).

Following the Nkrumah et al. (2005, 2007b) primary genome scan for QTL associated with RFI, five cattle chromosomes were selected for fine mapping to evaluate markers associated with QTL first by Moore et al. (2006) then by Sherman et al. (2008b). The latter study analyzed an average of 15.8 SNPs per chromosome (BTA2, 5, 10, 20 and 29). The largest effect on RFI was rs29019569:C>T on BTA2 with an estimated effect of -0.25 kg/day. The same allele substitution was responsible for the largest effect on FCR (-0.26 kg DMI/kg gain). The largest effect on DMI was identified at rs29021916:A>T on BTA10 with an estimated effect of 0.31 kg/day. Both rs29021916:A>T and rs29021102:C>T were associated with all traits, RFI, FCR and DMI. The combined effects of the five most significant SNP for RFI explained 6.9% of the phenotypic variation in the trait (Sherman et al., 2008b).

Fine mapping of QTL for feed intake and feed efficiency (Sherman et al., 2009) compared location of SNP from the whole-genome analysis (Barendse et al., 2007) to QTL location. The closest SNP, rs29020547:A>G, was identified approximately 200kb from the RFI QTL on BTA 25. Additionally, rs29022381:C>T was within 1 Mb of QTL on BTA 18. Six additional SNP were identified within 5 Mb of RFI QTL on BTA 1, 7, 11, 19, 22 and 23 (Sherman et al., 2009).

Two methods were used by Sherman et al. (2010) to create a panel of 150 informative SNP for RFI; a multivariate model where unique SNP effects were combined and a sequential MBV that added estimated SNP effects one at a time keeping only the SNP that improved the proportion of variance explained. One SNP from the sequential MBV panel matched previous work by Bardense et al. (2007), SNP rs29022779:C>T on BTA 1. Three SNP were located in coding regions of genes while only one SNP,

rs29016422>C:T, produced a nonsynonymous change in the protein-coding contactin 1 (CNTN1) gene. In total, 9.5% of the SNP from both panels are within 5 cM of QTL previously reported to be associated with RFI (Sherman et al., 2008a, 2009).

Nkrumah et al. (2007b) discovered a QTL on BTA 6 at 42 cM associated with ADG. QTL associated with birth weight and carcass quality traits were also detected on BTA 6 between 30 and 42 cM (Gutierrez-Gill et al., 2009; Takasuga et al., 2007). An additional region of BTA 6 (37.96-38.53 cM) was found to be correlated to feed intake and gain in a crossbred population of cattle (Snelling et al., 2011). Seven genes exist on the Btau 4.0 genome assembly in the region of interest. Of the seven genes on BTA 6 between 30 and 42 cM, leucine aminopeptidase (LAP3), non-SMC condensing I complex, subunit G (NCAPG) and LOC540095 (bovine ortholog of ligand depend nuclear receptor corepressor-like, LCORL) were used as potential candidate genes for feed intake and growth by Lindholm-Perry et al. (2011).

A recent GWAS study of feed efficiency and its component traits performed by Saatchi et al. (2014a) discovered ten significant 1-Mb SNP windows located on eight autosomes for RFI with the largest effect 1-Mb SNP window detected on BTA 15 in a Simmental x Angus population which explained 2.40% of total additive genetic variance. Large-effect QTL associated with RFI were also detected on chromosomes 6, 10, 14, 18, 19, 20 and 25. In the same study, six large-effect QTL associated with feedlot DMI were identified. Three pleiotropic or closely linked QTL on BTA 21 at 13 Mb and BTA 7 at 0 and 23 Mb were associated with both DMI and MBW. BTA 7 at 23 Mb in an Angus population explained 10.39% of additive genetic variance of DMI and also explained 14.25% of additive genetic variance of MBW. Two pleiotropic or closely linked QTL

with ADG and MBW were identified on BTA 6 at 38 Mb and BTA 7 at 93 Mb. BTA 6 at 38 Mb has previously been reported with feed efficiency (Snelling et al., 2011; Lindholm-Perry et al., 2013; Lindholm-Perry et al., 2011). BTA 7 at 93 Mb explained 9.18% of additive genetic variance for ADG in a Hereford population while the same location explained 2.55% additive genetic variance for MBW also in Hereford (Saatchi et al., 2014a).

Pleiotropic studies for QTL identification

Pleiotropy exists when polymorphisms affect multiple traits and is the leading cause for genetic correlations between traits while LD between QTL for different traits can also lead to genetic correlations. Identification of QTL expressing different patterns of pleiotropy can help explain the underlying physiological aspects (Bolormaa et al., 2014). Cross phenotype associations occur when two traits are found to be associated with SNP within the same chromosomal region. It is difficult to decipher whether the cross phenotype association is a result of one pleiotropic QTL or two linked QTL (Solocieff et al., 2013). Multi-trait GWAS may lead to an increase in power to detect QTL (Bolormaa et al., 2014). Saatchi et al. (2014b) analyzed the pleiotropic effects of 12 traits across 10 breeds of cattle. Four pleiotropic QTL were detected in two or more breeds with the largest impact detected on BTA 6 at 37-42 Mb. This large-effect QTL was associated with body weights, calving ease direct and weaning weight maternal for all breeds in the study except Angus, while the largest effect of the QTL was detected in Hereford where it explained 32% of the additive genetic variance for calving ease (Saatchi et al., 2014b). The location of this pleiotropic QTL is in the same region as previously reported feed efficiency QTL (Lindholm-Perry et al., 2013; Lindholm-Perry et

al., 2011; Snelling et al., 2011). The pleiotropic QTL on BTA 7 at 93 Mb explained more than 16% of the additive genetic variance for ribeye muscle area in Gelbvieh. It was also associated with ribeye muscle area in Angus, Hereford, Red Angus and Simmental and growth traits in Angus, Hereford and Simmental. A third pleiotropic QTL on BTA 14 at 24-26 Mb was associated with five traits for growth in Brangus, Gelbvieh and Simmental. The final across-breed pleiotropic QTL was detected on BTA 20 at 4 Mb. It was found to be associated with 9 of the 16 traits including growth and calving ease in Angus, Shorthorn, Hereford, Red Angus and Simmental (Saatchi et al., 2014b). The advantage of pleiotropic QTL lies in the greater understanding of the biological mechanisms of beef cattle in relation to a variety of traits.

Future implications of genomic predictors for feed efficiency

Genomics has proven to be an exciting time within the beef industry; however, it is not a cure-all type of solution. With expensive to measure phenotypes, such as feed intake, it is practical to assume that only superior animals will be chosen for feeding trials. This non-random selection creates a bias in the genomic predictions (Spangler, 2013). Genomic prediction requires a large number of animals with phenotypic and genotypic data for training. For traits that are routinely recorded and have existing EPD, the transition to GE-EPD has been made. However, novel traits require greater effort to build resource populations of thousands of animals representing multiple breeds to establish genomic predictions that are robust across beef cattle populations. This has been the focus of a multi-institutional research effort in beef cattle (Saatchi et al., 2014). Genomic information could also be improved by having a greater understanding of the underlying biological mechanisms of distinct phenotypes (Eggen, 2012). Molecular

breeding values work well when used within the same cattle breed as training, but lose efficacy when applied across breeds (Kachman et al., 2013).

At its current state, genomics serves as a tool to compliment selection techniques in order to gain higher accuracies. Although genomics was unable to serve as the magic bullet for animal breeding, it does bring forth advantages. Aside from the expected genetic gains through greater accuracies and decreased generation intervals, genomics has the ability to aid in parent identification and traceability. As industry and social demands continue to increase, it is vital for livestock producers to implement all possible selection techniques to produce the most efficient animals. The world population is expected to increase 40 to 50% by 2020 to 2030. To accommodate the growing demand for protein with the decreasing land resources, cattle must be more efficient in converting feed to consumable product. Increased environmental awareness also drives the demand for greater feed efficiency with concerns of the carbon footprint resulting from livestock production (Green, 2008). Cattle producers will face these contests and many others in years to come, but the opportunities through beef cattle genomics are considerable.

Literature Cited

Andersson, L., and M. Georges. 2004. Domestic animal genomics: deciphering the genetics of complex traits. Nature Rev. Genet. 5:202-212.

Archer, J.A., E.C. Richardson, R.M. Herd, and P.E. Arthur. 1999. Potential for selection to improve efficiency of feed use in beef cattle: a review. Aust. J. Agric. Res. 50:147-161.

Arthur, P.F., G. Renand, and D. Krauss. 2001b. Genetic and phenotypic relationships among different measures of growth and feed efficiency in young Charolais bulls. Livest. Prod. Sci. 68:131-139.

Arthur, P.F., J.A. Archer, D.J. Johnston, R.M. Herd, E.C. Richardson, and P.F. Parnell. 2001a. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. J. Anim. Sci. 79:2805-2811.

Barendse, W., A. Reverter, R.J. Bunch, B.E. Harrison, W. Barris, and M.B. Thomas. 2007. A validated whole-genome association study of efficient food conversion in cattle. Genetics 176:1893-1905.

Bolormaa, S., J.E. Pryce, A. Reverter, Y. Zhang, W. Barendse, K. Kemper, B. Tier, K. Savin, B.J. Hayes, and M.E. Goddard. 2014. A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. PLOS Genetics 10(3):e1004198.

Bolormaa, S., J.E. Pryce, K. Kemper, K. Savin, B.J. Hayes, W. Barendse, Y. Zhang, C.M. Reich, B.A. Mason, R.J. Bunch, B.E. Harrison, A.Reverter, R.M. Herd, B.Tier, H.-U. Graser, and M.E. Goddard. 2013. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in Bos Taurus, Bos indicus, and composite beef cattle. J. Anim. Sci. 91:3088-3104.

Botstein, D., and N. Risch. 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 33:Suppl 228-237.

Breiman, L. 1995. Better subset selection using the nonnegative garrote. Technometrics 37:373-384.

Cardon, L.R., and J.I. Bell. 2001. Association study designs for complex diseases. Nat. Rev. Genet. 2:91-99.

Conner, J.K., and D.L. Hartl. 2004. A primer of ecological genetics: Sunderland, Massachusetts. Sinauer Associates Inc.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. J. Anim. Sci. 82:ESuppl E313-328.

Dekkers, J.C.M., and F. Hospital. 2002. Multifactorial genetics: the use of molecular genetics in the improvement of agricultural populations. Nature Rev. Genet. 3:22-32.

Dickerson, G.E., N. Kunzi, L.V. Cundiff, R.M. Koch, V.H. Arthaud, and G.E. Gregory. 1974. Selection criteria for efficient beef production. J. Anim. Sci. 39:659-673.

Eggen, A. 2012. The development and application of genomic selection as a new Breeding paradigm. Anim. Front. 2:10-15.

Falconer, D.S., and T.F.X. McKay. 1996. Introduction to Quantitative Genetics , Ed. 4, Longman, New York.

Fan, B., D. Zhi-Qiang, D.M. Gorbach, and M.F. Rothchild. 2010. Development and application of high-density DNP arrays in genomic studies of domestic animals. Asian-Aust. J. Anim. Sci. 23:833-847.

Fernando, R.L., and D.J. Garrick. 2013. Bayesian regression as an alternative implementation of genomic-enhanced genetic evaluation. Proc. 10th Genetic Prediction Workshop. Kansas City, MO, p. 38-43.

Fernando, R.L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21:467-477.

Freetly, H.C., and T.M. Brown-Brandl. 2013. Enteric methane production from beef cattle that vary in feed efficiency. J. Anim. Sci. 91:4826-4831.

Garrick, D.J., and M. Saatchi. 2013. Practical experiences in developing breed-specific predictions for genome-enhanced EPDs. Proc. 10th Genetic Prediction Workshop. Kansas City, MO, p. 24-34.

Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, A.T. Pasquino, L.S. Sargeant, A. Sorensen, M.R. Steele, and X. Zhao. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907-920.

Gianola D., and R.L. Fernando. 1986. Bayesian methods in animal breeding theory. J. Anim. Sci. 63:217-244.

Gianola D., G. de los Campos, W.G. Hill, E. Manfredi, and R. Fernando. 2009. Additive genetic variability and the Bayesian alphabet. Genetics 183:347-363.

Goddard, M. 2008. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245-257.

Goddard, M.E., and B.J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature 10:381-391.

Goddard, M.E., B.J. Hayes, and T.H.E. Meuwissen. 2010. Genomic selection in livestock populations. Genet. Res. 92:413-421.

Green, R.D. 2008. ASAS Centennial Paper: Future needs in animal breeding and genetics. J. Anim. Sci. 87:793-800.

Gutierrez-Gill, B., J.L. Williams, D. Homer, D. Burton, C.S. Haley, and P. Wiener. 2009. Search for quantitative trait loci affecting growth and carcass traits in a cross population of beef and dairy cattle. J. Anim. Sci. 87:24-36.

Habier, D., R. L. Fernando, K. Kizilkaya, and D.J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics 12:186.

Hastbacka, J.A., A. de la Chapelle, I. Kaitila, P. Sistonen, A. Weaver, and E. Lander. 1992. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat. Genet. 2:204-211.

Hayes, B.J. 2013. Overview of statistical methods for genome-wide association studies (GWAS). In: Cedric Gondro et al., editors, Genome-Wide Association Studies and Genomic Prediction, Methods in Molecular Biology. Spring Science+Business Media, LLC, p. 149-169.

Hayes, B.J., P.M. Visscher, and M.E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. 91:47-60.

Henderson, C.R. 1984. Best linear unbiased prediction of performance and breeding value. Proc. 33rd National Breeders' Roundtable, St. Louis, Mo, p. 172.

Herd, R.M., and S.C. Bishop. 2000. Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. Livest. Prod. Sci. 63:111-119.

Kachman, S. 2008. Incorporation of marker scores into national cattle evaluations. Proc. 9th Genetic Prediction Workshop, Kansas City, MO, p. 88-91.

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

Kennedy, B.W., J.H.J. van der Werf, and T.H.E. Meuwissen. 1993. Genetic and statistical properties of residual feed intake. J. Anim. Sci. 71:3239-3250.

Kneeland, J.C., C. Li, J. Basarab, W.M. Snelling, B. Benkel, B. Murdoch, C. Hansen, and S.S. Moore. 2004. Identification and fine mapping of quantitative trait loci for growth traits on bovine chromosomes 2, 6, 14, 19, 21, and 23 within one commercial line of Bos taurus. J. Anim. Sci. 82:3405-3414.

Koch, R.M., L.A. Swinger, D. Chambers and K.E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22:486-494.

Lande, R., and R. Thompson. 1990. Efficiency of marker assisted selection in the improvement of quantitative traits. Genetics 124:743-756.

Li, C., J. Basarab, W.M. Snelling, B. Benkel, B. Murdoch, and S.S. Moore. 2002. The identification of common haplotypes on bovine chromosome 5 with commercial lines of Bos taurus and their associations with growth traits. J. Anim. Sci. 80:1187-1194.

Lindholm-Perry, A.K., L.A. Kuehn, W.T. Oliver, A.K. Sexton, J.R. Miles, L.A. Rempel, R.A. Cushman and H.C. Freetly. 2013. Adipose and muscle tissue gene expression of two genes (NCAPG and LCORL) located in a chromosomal region associated with cattle feed intake and gain. PLoS One 8:e80882.

Lindholm-Perry, A.K., A.K. Sexton, L.A. Kuehn, T.P. Smith, D.A. King, S.D. Shackelford, T.L. Wheeler, C.L. Ferrell, T.G. Jenkins, W.M. Snelling, and H.C. Freetly. 2011. Association, effects and validation of polymorphisms with the NCAPG – LCORL locus located on BTA 6 with feed intake, gain, meat and carcass traits in beef cattle. BMC Genet. 12:103.

Lu, Y., M.J. Vandehaar, D.M. Spurlock, K.A. Weigel, L.E. Armentano, C.R. Staples, E.E. Connor, Z. Wang, N.M. Bello, and R.J. Templeman. 2015. An alternative approach to modeling genetic merit of feed efficiency in dairy cattle. J. Dairy Sci. 98:6535-6551.

MacNeil, M.D., J.D. Nkrumah, B.W. Woodward, and S.L. Northcutt. 2010. Genetic evaluation of Angus cattle for carcass marbling using ultrasound and genomic indicators. J. Anim. Sci. 88:517-522.

Meuwissen, T.H.E., and M.E. Goddard. 1996. The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. 28:161-176.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. J. Anim. Sci. 88:544-551.

Misztal, I., S.E. Aggrey, and W.M. Muir. 2013. Experiences with single-step genome evaluation. Poul. Sci. 92:2530-2534.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic valuation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655.

Moore, S., D.H. Crews, and D. Nkrumah. 2006. Multiple and candidate gene approaches to genetic evaluation of feed efficiency in beef cattle. Proc. 8th World Congr. Genet. Appl. Livest. Prod., Belo Horizonte, Brazil. pp. 3-6.

Nkrumah, J.D., C. Li, Z. Wang, R. Bartusiak, B. Murdoch, J. Basarab, D. Crews, and S. Moore. 2005. Full genome scan of quantitative trait loci (QTL) for net feed efficiency in beef cattle. J. Anim. Sci. 83 Suppl 1:13.

Nkrumah, J.D., E.L. Sherman, C. Li, E. Marques, D.H. Crews Jr., R. Bartusiak, B. Murdoch, Z. Wang, J.A. Basarab, and S.S. Moore. 2007b. Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency in beef cattle. J. Anim. Sci. 85:3170-3181.

Nkrumah, J.D., J.A. Basarab, Z. Wang, C. Li, M.A. Price, E.K. Okine, D.H. Crews, Jr., and S.S. Moore. 2007a. Genetic and phenotypic relationships of feed intake and measures of efficiency with growth and carcass merit of beef cattle. J. Anim. Sci. 85:2711-2720.

Robinson, D.L., and V.H. Oddy. 2004. Genetic parameters for feed efficiency, fatness, Muscle area and feeding behaviour of feedlot finished beef cattle. Lvstk. Prod. Sci. 90:255-270.

Saatchi, M., J.E. Beever, J.E. Decker, D.B. Faulkner, H.C. Freetly, S.L. Hansen, H. Yampara-Iquise, K.A. Johnston, S.D. Kachman, M.S. Kerley, J. Kim, D.D. Loy, E. Marques, H.L. Neibergs, E.J. Pollak, R.D. Schnabel, C.M. Seabury, D.W. Shike, W.M. Snelling, M.L. Spangler, R.L. Weaber, D.J. Garrick, and J.F. Taylor. 2014a. QTLs associated with dry matter intake, metabolic mid-test weight, growth and feed efficiency have little overlap across 4 beef cattle studies. BMC Genomics 15:1004.

Saatchi, M., R.D. Schnabel, J.F. Taylor, and D.J. Garrick. 2014b. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. BMC Genomics 15:442.

Schenkel, F.S., S.P. Miller, and J.W. Wilton. 2004. Genetic parameters and breed differences for feed efficiency, growth, and body composition traits of young beef bulls. Can. J. Anim. Sci. 84:177-185.

Sherman, E.L., J.D. Nkrumah, C. Li, R. Bartusiak, B. Murdoch, and S.S. Moore. 2009. Fine mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. J. Anim. Sci. 87:37-45.

Sherman, E.L., J.D. Nkrumah, B.M. Murdoch, C. Li, Z. Wang, A. Fu, and S.S. Moore. 2008a. Polymorphisms and haplotypes in the bovine neuropeptide T, growth hormone receptor, insulin-like growth factor 2, and uncoupling proteins 2 and 3 genes and their associations with measures of growth, performance, feed efficiency, and carcass merit in beef cattle. J. Anim. Sci. 86:1-16.

Sherman, E.L., J.D. Nkrumah, B.M. Murdoch, S.S. Moore. 2008b. Identification of polymorphisms influencing feed intake and efficiency in beef cattle. Amin. Genet. 39:225-231.

Sherman, E.L., J.D. Nkrumah, and S.S. Moore. 2010. Whole genome single nucleotide polymorphism associations with feed intake and feed efficiency in beef cattle. J. Anim. Sci. 88:16-22.

Snelling, W.M., M.F. Allan, J.W. Keele, L.A. Kuehn, R.M. Thallman, G.L. Bennett, C.L. Ferrell, T.G. Jenkins, H.C. Freetly, M.K. Nielsen, and K.M. Rolfe. 2011. Partial-genome evaluation of postweaning feed intake and efficiency of crossbred beef cattle. J. Anim. Sci. 89:1731-1741.

Solovieff, N., C. Cotsapas, P.H. Lee, S.M Purcell, and J.W. Smoller. 2013. Pleiotropy in complex traits: challenges and strategies. Nat. Rev. Genet. 14:483-495.

Spangler, M.L. 2011. Integration of genomic information into genetic evaluation. NCBA Cattlemen's College.

Spangler, M.L. 2013. Strengths and weaknesses of methods of incorporating genomics into genetic evaluations. Proc. 10th Genetic Prediction Workshop. Kansas City, MO, p. 1-4.

Strathe, A.B., T. Mark, B. Nielsen, D.N. Doo, H.N. Kadarmideen, and J. Jensen. 2014. Deriving genomic breeding values for residual feed intake from covariance functions of random regression models. Proc 10th World Congr. Genet. Appk. Livest. Prod., Vancouver, Canada.

Takasuga, A., T. Watanabe, Y. Mizoguchi, T. Hirano, N. Ihara, A. Takano, K. Yokouchi, A. Fujikawa, K. Chiba, N. Kobayashi, K. Tatsuda, T. Oe, M. Furukawa-Kuroiwa, A. Nishimura-Abe, T. Fruita, K. Inoue, K. Mizoshita, A. Ogino, and Y. Sugimoto. 2007. Identification of bovine QTL for growth and carcass traits in Japanese Black cattle by replication and identical-be-descent mapping. Mamm Genome 18:125-136.

The Bovine HapMap Consortium. 2009. Genome-wide survery of SNP variation uncovers the genetic structure of cattle breeds. Science 324:528-532.

Van Laere, A.S., M. Nguyen, M. Braunschweig, C. Nezer, C. Collette, L. Moreau, A.L. Archibald, C.S. Haley, N. Buys, M. Tally, G. Andersson, M. Georges, and L. Andersson. 2003. A regulatory mutation in IFG2 causes a major QTL effect on muscle growth in the pig. Nature 425:832-836.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 92:16-24.

Whittaker, J.C., R. Thompson, and M.C. Denham. 2000. Marker-assisted selection using ridge regression. Genet. Res., Camb. 75:249-252.

Xu, S. 2002. Estimating polygenic effects using markers of the entire genome. Genetics 163:789-801.

Zhang, W. and C. Smith. 1992. Computer simulation of marker-assisted selection utilizing linkage disequilibrium. Theor. and Appl. Genet. 83:813-820.

Abstract

Feed costs comprise the majority of variable expenses in beef cattle systems making feed efficiency an important economic consideration within the beef industry. Due to the expense of recording feed intake phenotypes, identification of genomic regions associated with feed efficiency traits is advantageous for facilitating selection programs. Genome-wide association studies (GWAS) were performed using 748 crossbred steers and heifers representing seven sire breeds with phenotypes for average daily gain (ADG) and average daily feed intake (ADFI). Animals were genotyped with the BovineSNP50v2 BeadChip. Both traits were analyzed using univariate SNP-based (BayesC) and haplotype-based (BayesIM) models and jointly using BayesIM to perform a bivariate GWAS. A hidden Markov model (HMM) of variable length haplotype segments was built where haplotypes were mapped to haplotype clusters based on local haplotype similarity. The estimated HMM was then used to assign haplotype cluster genotypes, instead of SNP genotypes, as latent covariates in a Bayesian mixture model. Haplotype cluster effects at loci with non-zero effects were modeled as normal random variables. In the bivariate model, loci where both traits had non-zero effects, haplotype cluster effects were modeled as bivariate normal random variables. The number of haplotype clusters at each location was assumed to be either 8 (BayesIM8) or 16 (BayesIM16). A total of three univariate analyses for each trait and two bivariate analyses were performed. In all cases, the effect of fitting sire breed (SB) as a fixed effect or not fitting sire breed as a fixed effect (NSB) was quantified. From the SB analysis, posterior mean genomic heritability estimates (PSD) for ADG were 0.28 (0.08), 0.46 (0.12), 0.46 (0.12), 0.44 (0.10) and 0.44 (0.12) for BayesC, BayesIM8, BayesIM16, BayesIM8

bivariate and BayesIM16 bivariate, respectively. Estimates were similar for NSB. Average daily feed intake posterior mean genomic heritability estimates (PSD) were 0.30 (0.07), 0.46 (0.10), 0.45 (0.09), 0.45 (0.09) and 0.46 (0.11) for the same analyses for SB. Genomic heritability estimates for ADFI increased for NSB. The top 1% of 1-Mb windows in common between univariate SNP and haplotype models ranged from 0.24 to 0.52 and from 0.08 to 0.28 for ADG and ADFI, respectively, across both SB and NSB.

Introduction

The majority of variable expenses in beef cattle systems are due to feed costs making feed efficiency an important economic consideration (Koch et al., 1963; Dickerson et al., 1974). It is estimated that a 10% increase in gain would lead to an 18% advantage in profit, but if instead efficiency is improved by 10%, a 43% increase in profit would occur (Fox et al., 2001). Aside from the economic considerations, improved feed efficiency also has an environmental impact as more efficient cattle have fewer days to finish and produce less methane throughout their lifetime (Freetly and Brown-Brandl, 2013).

In the fed cattle sector, feed conversion ratio (FCR) is the most widely-used measure of efficiency. Due to the positive genetic correlation between feed intake and gain, selection to improve FCR often results in larger, more maintenance-intensive animals in the breeding herd (Archer et al., 1999). One proposed alternative to FCR is residual feed intake (RFI). The concept of RFI was introduced by Koch et al. (1963) by suggesting that feed intake should be adjusted for body weight and weight gain, making RFI the difference between actual feed intake and the predicted feed intake of an animal based on its requirements for maintenance and gain (Koch et al., 1963; Archer et al., 1999). It is sometimes considered the preferred definition of feed efficiency because RFI is phenotypically independent of the production traits (e.g. growth and metabolic body weight) used in the prediction equation (Kennedy et al., 1993). An alternative approach to feed efficiency is a multi-trait model that contains the component traits of efficiency measures. An index of these traits may represent a more robust measure of feed

efficiency and comprehensive investigation into the genetic relationship between intake and gain (Lu et al., 2015).

Moderate-to-high heritability estimates for feed efficiency traits (Arthur et al., 2001a; Arthur et al., 2001b; Nkrumah et al., 2007) suggest feed efficiency would respond favorably to selection. Still, feed intake, and consequently feed efficiency traits, are difficult to obtain and expensive to measure. Therefore a genomics approach seems warranted. Genome-wide association studies (GWAS) have identified several QTL associated with feed efficiency traits (Saatchi et al., 2014; Snelling et al., 2011). The objective of this study was to identify genomic regions associated with average daily gain (ADG) and average daily feed intake (ADFI) in an admixed population of beef cattle using univariate SNP and univariate and bivariate haplotype models.

Materials and Methods

Description of population

Feedlot ADG and ADFI (on a dry matter basis) were recorded from crossbred steers and heifers (n=777) at the U.S. Meat Animal Research Center (USMARC) in Clay Center, Nebraska and the University of Missouri (MU) in Columbia, Missouri. Commercial dams were mated to seven purebred sire breeds including Angus, Red Angus, Charolais, Simmental, Hereford, Gelbvieh and Limousin and one commercial sire group comprised of ½ Angus, ¼ Simmental and ¼ South Devon. The number of offspring by breed of sire is presented in Table 1.

Animals used in the current study were the product of three matings across two years and two locations. The first calf crop (n=213) was born in May 2012 at the Rex Ranch near Ashby, Nebraska. Calves were weaned in August of the same year and placed

in a dry lot for backgrounding before entering individual feed intake facilities. The steer calves from the Rex Ranch were placed in GrowSafe facilities at MU with the feeding period beginning on March 20, 2013 and ending on May 30, 2013 for a 70 d feeding period. Weights were recorded for two consecutive days at the start and end of the feeding period. Initial and final weights were determined as the mean of the two consecutive weights. While in the individual feeding facilities, the ration consisted of 8.9% corn silage, 52.1% whole corn, 26.4% DDGS and 12.6% premix on a dry matter basis.

The second calf crop (n=309) was born in August of 2012 at USMARC and fence line weaned in January of 2013. They entered Calan Gate feeding facilities at USMARC. The feeding period began July 9, 2013. Initial and final weights were determined as the mean of the two consecutive weights at the start and end of the feeding period. The on-test finishing ration consisted of 8% ground alfalfa, 67.75% rolled corn, 20% WDGS, and 4.25% supplement containing rumensin at 700g/ton on a dry matter basis. Steers were implanted with Revalor XS and heifers were implanted with Revalor IH. Cattle were removed from the facilities on October 1, 2013 for an 83 d feeding period. Forty steers were used in a metabolism study prior to the feeding period, and therefore were treated as a separate contemporary group.

The final group of calves (n=255) were born across April and May of 2013 at USMARC and weaned into the feedlot in September. They entered Calan Gate feeding facilities with the feeding period beginning on February 11, 2014 and were removed from the facilities on May 6, 2014 for an 84 d feeding period. Initial and final weights were determined by the mean of the two consecutive weights at the start and end of the feeding

period. The finishing ration and implant regimen were consistent with the 2012-born USMARC cattle.

Animals were genotyped with Illumina BovineSNP50v2 Beadchip (Illumina, San Diego, CA).

Data editing

Animals with unidentified sires or sire breeds (n=6), those with missing birth dates (n=19), missing genotypic data (n=1) or late castrated steers (n=3) were removed from the analysis. A total of 748 animals remained after data editing. Phenotypic means (SD) for ADG and ADFI after correcting for breed of sire, contemporary group (concatenation of location, year and sex) and initial weight when entering the feeding facilities are presented in Table 2.

Quality scores (GenCall) for each genotype were assigned through Illumina data analysis software. Missing genotypes or genotypes with GenCall scores less than 0.20 were replaced with the mean genotype score at that marker calculated within subgroups based on location and birth year. No pre-analysis filtering was performed based on minor allele frequency. Unmapped and sex chromosome SNP were removed leaving 52,890 SNP for analysis.

Statistical analyses

Both traits were analyzed independently through SNP (BayesC; Habier et al., 2011) and haplotype (BayesIM; Kachman, 2016) models and together using BayesIM to perform a bivariate GWAS. Contemporary group (concatenation of year, location and sex) was fitted as a classification effects with initial weight, calculated as the average of the two consecutive weights at the start of the feeding period, fitted as a fixed covariate.

Additionally, the impact of breed of sire was quantified by including (SB) or excluding (NSB) sire breed as a classification effect with the composite sire group treated as its own breed. BayesC was implemented via GenSel (Version 0.9.2.045; Fernando and Garrick, 2009). The proportion of SNP assumed to have a null effect on the trait, π, was assumed to be 0.99 which corresponded to fitting approximately 500 markers in each Markov chain Monte Carlo (MCMC) iteration. A chain length of 41,000 iterations was run with the first 1,000 discarded as burn-in. A prior heritability estimate was selected by starting with high and low *a priori* heritability estimates until the posterior heritability estimates were trending up and down, respectively and a value in the middle was chosen as the *a priori* heritability estimate. The genome was also separated into 1-Mb non-overlapping windows (n=2,536) with the additive genetic variance calculated within each window.

Haplotype association analyses were performed using a Bayesian mixture model fitting haplotype effects as covariates (BayesIM). BayesIM begins with a hidden Markov model (HMM) that models haplotypes using haplotype clusters and transitions between the clusters. Haplotypes are mapped to haplotype clusters based on local haplotype similarity. The estimated HMM is then used to sample haplotype cluster genotypes. The haplotype cluster genotypes, instead of SNP genotypes, are used as latent covariates in a Bayesian mixture model. Haplotype cluster effects at loci with non-zero effects were modeled as independent normal random variables. In the bivariate model, loci where both traits had non-zero effects, haplotype cluster effects were modeled as bivariate normal random variables. The number of haplotype clusters at each location was assumed to be either 8 (BayesIM8) or 16 (BayesIM16). A pooled within sire breed genetic variance was calculated. To keep approximately equal number of markers in the SNP and haplotype

model, π was assumed to be 0.98 for both the univariate and bivariate haplotype analyses. The average haplotype length was set to 500 kb with QTL evenly spaced every 100 kb. A total MCMC chain length of 100,000 iterations was used with the first 10,000 iterations discarded as burn-in. A prior heritability estimate was again selected by starting with high and low *a priori* heritability estimates until the posterior heritability estimates were trending up and down, respectively. A middle value was then chosen as the *a priori* heritability estimate. Overlapping 1-Mb QTL regions (n=25,200) were built in a stair-step fashion by offsetting the region starting position by 100 kb. From these QTL regions, 1-Mb non-overlapping windows (n=2,536) were extracted for a direct comparison to the SNP-based model.

Calculation of genetic and residual correlations

Bivariate haplotype analyses estimate the genetic and residual (co)variances for both traits in the model at each iteration. Given the additive genetic merit of the jth animal from j=1,…,748 is:

$$A_j = \sum_{i=1}^{25,200} H_i G_{ij}$$

where H_i is the effect of the ith haplotype loci from i=1,…,25,200 and G_{ij} is the unobserved haplotype genotype of the ith haplotype loci for the jth animal, the additive genetic covariance between ADG and ADFI is derived as:

$$\sigma_{A_{ADG,AFI}} = \frac{\sum_{j=1}^{748}(A_{ADG_j} - \overline{A_{ADG}})(A_{ADFI_j} - \overline{A_{ADFI}})}{748}$$

where A_{ADG_j} is the additive genetic merit of each animal from j=1,…,748 for ADG, $\overline{A_{ADG}}$ is the mean additive genetic merit for ADG, $A_{ADFI_j}$is the additive genetic merit of each animal from j=1,…,748 for ADFI and $\overline{A_{ADFI}}$ is the mean additive genetic merit for ADFI.

The genetic correlation was calculated as:

$$r_{A_{ADG,ADFI}} = \frac{\sigma_{A_{ADG,ADFI}}}{\sigma_{A_{ADG}} x \sigma_{A_{ADFI}}}.$$

where $\sigma_{A_{ADG,ADFI}}$ is the additive genetic covariance between ADG and ADFI, $\sigma_{A_{ADG}}$ is the additive genetic standard deviation of ADG and $\sigma_{A_{ADFI}}$ is the additive genetic standard deviation of ADFI.

The residual correlation was calculated as:

$$r_{E_{ADG,ADFI}} = \frac{\sigma_{E_{ADG,ADFI}}}{\sigma_{E_{ADG}} x \sigma_{E_{ADFI}}}$$

where $\sigma_{E_{ADG,ADFI}}$ is the residual covariance between ADG and ADFI, $\sigma_{E_{ADG}}$ is the residual standard deviation of ADG and σ_{ADFI} is the residual standard deviation of ADFI.

Calculation of rank correlations

A molecular breeding value (MBV) for each animal was estimated as the total genetic value of that individual based on the summation of the product of the marker effect and animal's genotype across all loci. Pearson rank correlations were calculated based on MBV within trait for ADG and ADFI between all possible univariate analyses.

Gene Ontology

Top windows of interest were extended by 0.5 Mb in each direction to determine candidate genes associated with feed efficiency traits using the *Bos taurus* build UMD_3.1 assembly (Zimin et al., 2009). The BioMart data mining tool available through Ensembl (Ensembl Genes 84) was used to determine gene ontology terms of candidate genes.

Results and Discussion

Posterior mean genomic heritability estimates

The posterior means (PSD) of genomic heritability, additive genetic and residual variances for ADG and ADFI from SB are presented in Table 3. When fitting breed of sire in the models, posterior mean genomic heritability estimates for ADG ranged from 0.28 to 0.46 for the three univariate and two bivariate analyses with the SNP model producing the lowest estimate and the haplotype analyses being similar despite the number of haplotype clusters assumed. A similar trend was observed with ADFI as posterior mean genomic heritability estimates ranged from 0.30 to 0.46 with the SNP model again producing the lowest estimate.

Posterior mean (PSD) genomic heritability and variance component estimates from NSB are presented in Table 4. A similar trend was observed with the SNP-based models producing the lowest posterior mean genomic heritability estimates. Estimates for ADG were invariable regardless of the inclusion of breed of sire as a classification effect with estimates ranging from 0.28 to 0.48. Average daily feed intake posterior mean genomic heritability estimates tended to increase with NSB as estimates ranged from 0.42 to 0.62. When breed of sire was removed from the model, the variance due to breed of sire differences was no longer being accounted for. The increased genomic heritability estimates for ADFI suggest the variance due to breed of sire differences was partitioned to the additive genetic variance.

Saatchi et al. (2014) reported genomic heritability estimates of 0.30 for ADG and 0.35 for ADFI from a BayesB analysis with an admixed population while Abo-Ismail et al. (2014) reported slightly higher estimates of 0.35 and 0.42 for ADG and ADFI,

respectively. Pedigree-based estimates of heritability range from 0.23 to 0.41 for ADG and 0.27 to 0.54 for ADFI (Arthur et al., 2001a; Arthur et al., 2001b; Nkrumah et al., 2007; Robinson and Oddy, 2004; Schnekel et al., 2003). These results are in agreement with the findings of the current study. Differences between the haplotype- and SNP-based models may be due to the breed admixture of the population as BayesIM has the potential to be more sensitive to breed admixture with the HMM possibly building haplotype clusters that are breed specific.

Genetic and residual correlations from bivariate analyses

The posterior mean genetic correlations (PSD) between ADG and ADFI from SB were 0.59 (0.09) and 0.58 (0.13) for BayesIM8 bivariate and BayesIM16 bivariate, respectively. When NSB was considered, estimates were similar with genetic correlations of 0.61 (0.07) and 0.60 (0.07) for the same analyses. Previous estimates of genetic correlations between ADG and ADFI range from 0.50 to 0.87 (Arthur et al., 2001a; Nkrumah et al., 2007; Schenkel et al., 2004).

The posterior mean residual correlations (PSD) between ADG and ADFI were 0.55 (0.07) and 0.55 (0.08) for SB from BayesIM8 bivariate and BayesIM16 bivariate, respectively. NSB yielded similar results with estimates of 0.56 (0.07) and 0.57 (0.08). Robinson and Oddy (2004) reported a higher residual correlation of 0.68 between feed intake and weight gain.

Rank correlations of Molecular Breeding Values

The rank correlations of MBV estimates across univariate analyses for ADG and ADFI are presented in Tables 5 and 6, respectively. Animals ranked similarly across SNP- and haplotype-based models for ADG with correlations > 0.95 for both SB and

NSB. Rank correlations were similar for ADFI with correlation > 0.96 for SB and > 0.95 for NSB. High rank correlation estimates between SNP and haplotype models indicated that both models would lead to similar animals being selected based on MBV.

Comparison of genomic regions across univariate SNP and haplotype models

The chromosomes and positions of top 1% of 1-Mb windows (n=25) based on the percentage of genetic variance explained from univariate analyses for SB and NSB are detailed in Tables 7 and 8, respectively. The top 1% of 1-Mb windows (n=25) from the bivariate analyses were determined by the top joint model frequency when both traits have a non-zero effect. The results for SB and NSB are presented in Tables 9 and 10, respectively. Commonality between top genomic regions across univariate SNP and haplotype models was described as the proportion of top 1% 1-Mb windows shared. Results are presented in Tables 11 and 12 for SB and NSB, respectively.

Within the top 1% of 1-Mb windows from SB, BayesC and BayesIM8 shared 44% of 1-Mb windows for ADG and 12% for ADFI. As the number of clusters increased, the commonality between models decreased as the genetic variance is partitioned across more haplotype effects with 32% of 1-Mb windows shared between BayesC and BayesIM16 for ADG and 8% for ADFI.

When NSB was considered, BayesC and BayesIM8 shared 40% of 1-Mb windows for ADG. The commonality between SNP and haplotype models was higher with NSB compared to SB for ADFI with 28% in common between BayesC and BayesIM8. BayesC and BayesIM16 shared 24% of 1-Mb windows for ADG and 12% for ADFI.

Genomic regions associated with ADG

Metropolis plots of the model frequency of each SNP or haplotype loci across the genome for BayesC, BayesIM8 and BayesIM16 for ADG from SB are in Figures 1, 2 and 3, respectively. For the same analyses from NSB, metropolis plots are presented in Figures 4, 5, and 6, respectively. The model frequency of the SNP model begins at zero due to monomorphic SNP being excluded from the analysis. However, with the haplotype model, all haplotype loci are included in the analysis. The magnitude of the model frequencies also differ between models. As the number of covariates fitted at a given loci increases, the magnitude of the model frequency decreases resulting in the SNP analysis having higher model frequencies than BayesIM8 and BayesIM16. BayesIM8 also has higher model frequencies than BayesIM16 again due to half as many covariates being fitted at a given loci. Additionally, the SNP model gives the location of the marker in LD with the QTL. The haplotype model builds a peak around the QTL with overlapping windows. This could provide more precise QTL locations by finding the interval around the QTL. The top 1% of 1-Mb windows was similar across SB and NSB. Specific QTL peak maximums and ranges presented are from SB.

For ADG, a peak on BTA 22 is seen across all univariate analyses based on model and window frequency (Figures 7 and 8). The top SNP on BTA 22 was *BTA-54550-no-rs* at 44.94 Mb identified through the BayesC analysis. A second SNP with high model frequency, *ARS-BFGL-NGS-81286*, was nearby at 45.12 Mb. The maximum point of the QTL peak was at 45.0 Mb for both BayesIM8 and BayesIM16. The width of the QTL peak was defined as the position when the model frequency was greater than the mean model frequency of the chromosome and ends when the model frequency returns to the

mean. The peak surrounding the top QTL on BTA 22 ranged from 42.69 to 47.70 Mb and 42.81 to 47.38 Mb for BayesIM8 and BayesIM16, respectively.

Gene ontology results for the extended 1-Mb window on BTA 22 include positive regulation of lipid formation (*ABHD6*), skeletal muscle tissue development (*E1BKX7*), muscle contraction (*SLMAP*), metabolic processes (*LOC100847355*) and regulation of glucose (*APPL1*). Snelling et al. (2011) discovered one significant SNP associated with ADG and two significant SNP associated with mid-test metabolic weight (MMBW) in the adjacent region from 45 to 46 Mb on BTA 22. Within that same 1-Mb region, Bolormaa et al. (2011) found two additional SNP associated with MMBW.

The QTL peak within the chromosomal region on BTA 13 from 83 to 84 Mb was common among the top 1% of 1-Mb windows across all univariate ADG analyses (Figures 9 and 10). While the top SNP based on model frequency on BTA 13 was at 75.27 Mb, two SNP within the region of interest, *ARS-BFGL-NGS-114977* and *ARS-BFGL-NGS-87042*, were located at 83.80 and 83.76 Mb, respectively. The maximum of the QTL peak was at 83.7 Mb from BayesIM8 while it shifted slightly to 83.8 Mb for BayesIM16. The QTL region spanned from BTA 13 at 80.76 to the end of the chromosome and BTA 13 at 79.24 Mb to the end of BTA 13 for BayesIM8 and BayesIM16, respectively.

Within this region are genes associated with the perception of smell (*LOC532472*) and nervous system development (*DOK5*). Lu et al. (2013) found *ARS-BFGL-NGS-87042* to have a significant allelic substitution effect on dry matter intake (DMI). An additional SNP in the same region, *ARS-BFGL-NGS-89423*, had a significant effect on birth weight (Lu et al., 2013).

Also common across univariate analyses for ADG was BTA 3 from 119 to 120 Mb (Figures 11 and 12). The top SNP based on model frequency, *ARS-BFGL-NGS-57851* was located at 119.55 Mb. The maximum point of the QTL peak was observed at 119.7 Mb from BayesIM8 and spanned from 118.74 to 120.08 Mb. When 16 clusters were assumed, the maximum point remained at 119.7 Mb and the width of the QTL peak increased, 118.83 to 121.49 Mb.

This region houses genes that are involved in cell proliferation (*HDAC4*), skeletal system development (*HDAC4*), metabolic processes (*LOC782114*), regulation of insulin and glucose (*CAPN10*) and positive regulation of skeletal muscle differentiation (*GCP1*). Serão et al. (2013) identified one SNP associated with residual ADG at 120.79 Mb.

Genomic regions associated with ADFI

Metropolis plots for ADFI for BayesC, BayesIM8 and BayesIM16 from SB are presented in Figures 13, 14 and15, respectively. Metropolis plots for the same analyses from NSB are presented in Figures 16, 17 and 18, respectively. The region from 151 to 152 Mb on BTA 1 was common across univariate analyses for ADFI based on model and window frequency (Figures 19 and 20). Three SNP with top model frequencies were within this window, *ARS-BFGL-NGS-14751*, *ARS-BFGL-NGS-57499* and *ARS-BFGL-NGS-70523*, located at 151.00, 151.15 and 151.13 Mb, respectively. The haplotype analyses agreed with the maximum of the QTL peak located at 151.1 Mb from BayesIM8 and BayesIM16. Within this region there were two peaks; a sharp, narrow QTL region followed by a broader area that began approximately 1 Mb following the tail of the first peak. For BayesIM8, the first peak ranged from 150.45 to 151.91 Mb with the subsequent interval starting at 152.63 Mb and extending to 155.99 Mb. When the number of clusters

assumed increased, the range of the first peak became narrower (150.15 to 151.61 Mb), while the second QTL interval spanned nearly 6 Mb from 152.63 to 158.34 Mb.

At the frontend of this region, Lu et al. (2013) found a SNP located at 149.55 Mb to have a significant allelic substitution effect on ADG. At the end of the QTL peak at 157.50 Mb, the same authors found *BTB-01633159* to be associated with RFI. Gene ontology results for the extended window on BTA 1 include nervous system development (SIM2), cell proliferation (HLCS, RIPPLY3) and cell differentiation (ERG).

On BTA 21, two adjacent windows from 26 to 28 Mb were associated with ADFI (Figures 21 and 22). Three SNP, *Hapmap53212-rs29015272*, *BTB-01168615* and *Hapmap49382-BTA-9378*, were located at 25.70, 26.12 and 27.89 Mb, respectively. These top SNP were potentially in LD with the same QTL as the QTL window ranged from 24.95 to 28.59 Mb from BayesIM8 and 25.14 to 29.15 Mb from BayesIM16. Within this region, gene ontology results include positive regulation of cell proliferation (CTSH, RASGRF1) and metabolic processes (FAH, ABHD17C). Abo-Ismail et al. (2014) found two SNP with significant associations with RFI and one SNP significantly associated with DMI on BTA 21 near 29 Mb.

Potential pleiotropic genomic regions associated with ADG and ADFI

In the bivariate analyses, top regions identified have the potential to influence both traits. Metropolis plots of BayesIM8 bivariate and BayesIM16 bivariate from SB are in Figures 23 and 24, respectively. Metropolis plots for the same analyses from NSB are presented in Figures 25 and 26, respectively.

It was expected that top regions in common across traits from the univariate analyses would also be apparent in the bivariate associations. The proportion of the top

1% 1-Mb windows shared between ADG, ADFI and bivariate haplotype association studies are presented in Tables 13 and 14 for BayesIM8 and BayesIM16, respectively.

Previously discussed regions on BTA 1 and 13 were also identified in the top 1% of 1-Mb windows from haplotype bivariate associations. Comparison of QTL peaks from BayesIM8 and BayesIM16 for ADG, ADFI and the bivariate association for BTA 1 are in Figures 27 and 28, respectively. The QTL intervals for ADFI and the bivariate analyses were similar while ADG lacked a pronounced peak. The BayesIM8 bivariate analysis displayed a sharp QTL peak from 1521.43 to 152.57 Mb. When 16 clusters were assumed, the QTL interval widened from 152.65 Mb to 158.49 Mb. On BTA 13, both traits and the bivariate analyses performed similar with all QTL intervals beginning at approximately 80 Mb and extending to the end of BTA 13 (Figures 29 and 30).

Additionally, BTA 25 from 34 to 35 Mb was common across all bivariate analyses. When compared to univariate haplotype analyses, no distinct peak was observed for ADG. However, QTL locations were similar for ADFI and bivariate haplotype analyses. Comparisons of QTL peaks on BTA 25 for ADG, ADFI and bivariate analyses are in Figures 31 and 32 for BayesIM8 and BayesIM16, respectively. For ADFI, the QTL intervals shifted with the differing number of clusters assumed. The peak ranged from 35.64 to 37.57 Mb from BayesIM8 and 33.01 to 35.79 Mb from BayesIM16. The bivariate associations had similar QTL intervals of 32.97 to 34.73 Mb and 33.81 to 35.25 Mb for BayesIM8 bivariate and BayesIM16 bivariate, respectively.

Gene ontology results of bovine candidate genes in this region include growth factor activity (RABEP2), metabolic processes (ATP2A1, SULT1A1, GDPD3, ALDOA), cell growth (NUPR1), skeletal muscle differentiation (NUPR1) and positive regulation of

protein modification (NUPR1). Two SNP located on BTA 25 at 36.28 Mb were found to have a significant effect on MMBW in a population of crossbred cattle (Abo-Ismail et al., 2014).

The region on BTA 20 from 6 to 7 Mb was observed in the majority of bivariate analyses. The QTL peak spanned approximately 2 Mb for both bivariate analyses with the intervals positioned from 5.78 to 7.68 Mb from BayesIM8 bivariate and 5.77 to 7.95 Mb when 16 clusters were assumed. This peak was observed in the univariate haplotype analysis for ADG. The interval location was similar to the bivariate analyses ranging from 5.76 to 7.57 Mb and 5.48 to 8.41 Mb for BayesIM8 and BayesIM16, respectively. Comparison of genomic regions across univariate and bivariate haplotype analyses on BTA 20 are illustrated in Figures 33 and 34.

This region is flanked by two large-effect pleiotropic QTL discovered by Saatchi et al. (2014). A QTL associated with RFI and MMBW in a Hereford population was identified in the 1-Mb window from 4 to 5 Mb on BTA 20. The same study identified a significant window on BTA 20 from 8 to 9 Mb associated with ADG in an admixed population of beef cattle and MMBW in a Simmental x Angus population. Genes within this region are involved in metabolic processes (ENSBTAG00000015512, ENSBTAG00000034138) and fat cell differentiation (MSX2).

Although bivariate analyses detect genomic regions that are potentially associated with two traits, it is unknown if these regions are pleiotropic. Caution must be taken when interpreting genetic correlations between complex traits when conducting associations with molecular markers as linkage between markers can create phantom correlations between traits (Gianola et al., 2015).

The number of haplotype clusters was assumed *a priori* to be either 8 or 16. Parameter estimates were similar regardless of the number of haplotype clusters assumed. Additionally, the models ranked animals similarly based on rank correlations of MBV. A slight difference was observed between BayesIM8 and BayesIM16 when determining the QTL intervals as the QTL intervals tended to be wider when more clusters were assumed.

Implications

Moving towards the use of haplotype models for genomic association studies has the ability to define the QTL locations more precisely. The traditional method of fitting SNP genotypes as covariates relies on LD between the marker and QTL. The corresponding location is that of the marker instead of the actual QTL of interest. Haplotype models allow for a QTL interval to be established which may help identify the casual variant when studies are advanced for fine mapping.

Multi-trait GWAS is currently unexplored in the beef cattle industry. The deployment of such association studies would not only allow for the identification of potential pleiotropic regions, but offer a more comprehensive biological investigation into genetic variants affecting feed efficiency. Although not the case in this study, animals with missing phenotypic records can be included in the analysis. In beef cattle systems, gain is recorded frequently while feed intake is an expensive phenotype to collect. Since a moderate-to-strong genetic correlation exists between ADG and ADFI, it is logical to exploit the knowledge of one trait to inform the other. Due to the cost of individual feeding systems, usually superior animals are chosen for phenotyping. This introduces selection bias. By including animals with missing phenotypes in the analysis, selection bias can be reduced.

Literature Cited

Abo-Ismail, M.K., G.V. Voort, J.J. Squires, K.C. Swanson, I.B. Mandell, X. Liao, P. Stothard, S. Moore, G. Plastow, and S.P. Miller. 2014. Single nucleotide polymorphisms for feed efficiency and performance in crossbred beef cattle. BMC Genetics 15:14.

Archer, J.A., E.C. Richardson, R.M. Herd, and P.E. Arthur. 1999. Potential for selection to improve efficiency of feed use in beef cattle: a review. Aust. J. Agric. Res. 50:147-161.

Arthur, P.F., G. Renand, and D. Krauss. 2001b. Genetic and phenotypic relationships among different measures of growth and feed efficiency in young Charolais bulls. Livest. Prod. Sci. 68:131-139.

Arthur, P.F., J.A. Archer, D.J. Johnston, R.M. Herd, E.C. Richardson, and P.F. Parnell. 2001a. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. J. Anim. Sci. 79:2805-2811.

Bolormaa, S., B.J. Hayes, K. Savin, R. Hawken, W. Barendse, P.F. Arthur, R.M. Herd, and M.E. Goddard. 2011. Genome-wide association studies for feedlot and growth traits in cattle. J. Anim. Sci. 89:1684-1697.

Dickerson, G.E., N. Kunzi, L.V. Cundiff, R.M. Koch, V.H. Arthaud, and G.E. Gregory. 1974. Selection criteria for efficient beef production. J. Anim. Sci. 39:659-673.

Fernando, R.L. and D.J. Garrick. 2009. GenSel User manual for a portfolio of genomic selection related analysis. Animal Breeding and Genetics. Ames: Iowa State University.

Fox, D.G., L.O. Tedeschi, and P.J. Guiroy. 2001. Determining feed intake and feed efficiency of individual cattle fed in groups. Proc. Beef Improvement Federation Conference, San Antonio, TX, p. 80-98.

Freetly, H.C., and T.M. Brown-Brandl. 2013. Enteric methane production from beef cattle that vary in feed efficiency. J. Anim. Sci. 91:4826-4831.

Gianola, D., G. de los Campos, M.A. Toro, H. Naya, C.C. Schon and D. Sorensen. 2015. Do molecular markers inform about pleiotropy? Genetics 201:23-29.

Habier, D., R. L. Fernando, K. Kizilkaya, and D.J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics 12:186.

Kachman, S.D. 2016. Genomic prediction using a model based on haplotype clusters. J. Anim. Sci. 94: Suppl. 2.

Kennedy, B.W., J.H.J. van der Werf, and T.H.E. Meuwissen. 1993. Genetic and statistical properties of residual feed intake. J. Anim. Sci. 71:3239-3250.

Koch, R.M., L.A. Swinger, D. Chambers and K.E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22:486-494.

Lu, D., S. Miller, M. Sargolzaei, M. Kelly, G. Vander Voort, T. Caldwell, Z. Wang, G. Plastow, and S. Moore. 2013. Genome-wide association analyses for growth and feed efficiency traits in beef cattle. J. Anim. Sci. 91:3612-3633.

Lu, Y., M.J. Vandehaar, D.M. Spurlock, K.A. Weigel, L.E. Armentano, C.R. Staples, E.E. Connor, Z. Wang, N.M. Bello, and R.J. Templeman. 2015. An alternative approach to modeling genetic merit of feed efficiency in dairy cattle. J. Dairy Sci. 98:6535-6551.

Nkrumah, J.D., J.A. Basarab, Z. Wang, C. Li, M.A. Price, E.K. Okine, D.H. Crews, Jr., and S.S. Moore. 2007. Genetic and phenotypic relationships of feed intake and measures of efficiency with growth and carcass merit of beef cattle. J. Anim. Sci. 85:2711-2720.

Robinson, D.L., and V.H. Oddy. 2004. Genetic parameters for feed efficiency, fatness, Muscle area and feeding behaviour of feedlot finished beef cattle. Lvstk. Prod. Sci. 90:255-270.

Saatchi, M., J.E. Beever, J.E. Decker, D.B. Faulkner, H.C. Freetly, S.L. Hansen, H. Yampara-Iquise, K.A. Johnston, S.D. Kachman, M.S. Kerley, J. Kim, D.D. Loy, E. Marques, H.L. Neibergs, E.J. Pollak, R.D. Schnabel, C.M. Seabury, D.W. Shike, W.M. Snelling, M.L. Spangler, R.L. Weaber, D.J. Garrick, and J.F. Taylor. 2014. QTLs associated with dry matter intake, metabolic mid-test weight, growth and feed efficiency have little overlap across 4 beef cattle studies. BMC Genomics 15:1004.

Schenkel, F.S., S.P. Miller, and J.W. Wilton. 2004. Genetic parameters and breed differences for feed efficiency, growth, and body composition traits of young beef bulls. Can. J. Anim. Sci. 84:177-185.

Serão, N.V.L., D. González-Peña, J.E. Beever, D.B Faulkner, B.R. Southey and S.L. Rodiguez-Zias. 2013. Single nucleotide polymorphisms and haplotypes associated with feed efficiency in beef cattle. BMC Genetics 14:94.

Snelling, W.M., M.F. Allan, J.W. Keele, L.A. Kuehn, R.M. Thallman, G.L. Bennett, C.L. Ferrell, T.G. Jenkins, H.C. Freetly, M.K. Nielsen, and K.M. Rolfe. 2011. Partial-genome evaluation of postweaning feed intake and efficiency of crossbred beef cattle. J. Anim. Sci. 89:1731-1741.

Zimin, A.V., A.L Delcher, L. Florea, D.R. Kelley, M.C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C.P. van Tassell, T.S. Sonstegard, G. Marcais, M. Roberts, P. Subramanian, J.A. Yorke, and S.L. Salzberg. 2009. A whole-genome assembly of the domestic cow, bos Taurus. Genome Biol. 10(4):R42.

Table 1. Number of calves by sire breed

[1]Sire Breed	No. Progeny
AN	204
AR	67
CH	64
COM	54
GV	148
HH	23
LM	73
SM	115

[1]AN=Angus, AR=Red Angus, CH=Charolais, COM=Commercial, GV=Gelbvieh, HH=Hereford, LM=Limousin, SM=Simmental

Table 2. Number of observations (N) and mean (SD) for ADG and ADFI[1]

Trait	N	Mean
ADG, kg/d	748	1.81 (0.22)
ADFI, kg/d	748	10.00 (1.13)

[1]ADG=average daily gain and ADFI=average daily feed intake adjusted for contemporary group (concatenation of year, location and sex), breed of sire and initial weight at the start of the feeding period

Table 3. Genomic heritability (h^2), additive genetic variance (V_A) and residual variance (V_E) for ADG and ADFI from SB[1,2]

Trait	BayesC			[3]BayesIM8			[4]BayesIM16			[3]BayesIM8 bivariate			[4]BayesIM16 bivariate		
	h^2	V_A	V_E	h^2	V_A	V_E	h^2	V_A	V_E	h^2	V_A	V_E	h^2	V_A	V_E
ADG,	0.28	0.02	0.05	0.46	0.03	0.03	0.46	0.03	0.03	0.44	0.03	0.04	0.44	0.03	0.04
kg/d	(0.08)	(0.01)	(0.01)	(0.12)	(0.01)	(0.01)	(0.12)	(0.01)	(0.01)	(0.10)	(0.01)	(0.01)	(0.12)	(0.01)	(0.01)
DMI,	0.30	0.28	0.67	0.46	0.43	0.50	0.45	0.42	0.51	0.45	0.43	0.53	0.46	0.44	0.51
kg/d	(0.07)	(0.07)	(0.07)	(0.10)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.11)	(0.09)	(0.11)	(0.13)	(0.10)

[1]ADG=average daily gain and ADFI=average daily feed intake

[2]SB= sire breed was included as a fixed effect in the model

[3]BayesIM8=BayesIM 8 clusters

[4]BayesIM16=BayesIM 16 clusters

Table 4. Genomic heritability (h^2), additive genetic variance (V_A) and residual variance (V_E) for ADG and ADFI from NSB[1,2]

Trait	BayesC			[3]BayesIM8			[4]BayesIM16			[3]BayesIM8 bivariate			[4]BayesIM16 bivariate		
	h^2	V_A	V_E	h^2	V_A	V_E	h^2	V_A	V_E	h^2	V_A	V_E	h^2	V_A	V_E
ADG, kg/d	0.28 (0.07)	0.02 (0.01)	0.05 (0.01)	0.47 (0.11)	0.03 (0.01)	0.03 (0.01)	0.48 (0.10)	0.03 (0.01)	0.03 (0.01)	0.41 (0.10)	0.03 (0.01)	0.04 (0.01)	0.42 (0.11)	0.03 (0.01)	0.04 (0.01)
DMI, kg/d	0.42 (0.06)	0.44 (0.07)	0.60 (0.06)	0.61 (0.08)	0.57 (0.09)	0.37 (0.07)	0.60 (0.08)	0.57 (0.09)	0.37 (0.07)	0.58 (0.07)	0.63 (0.10)	0.44 (0.07)	0.62 (0.07)	0.66 (0.10)	0.40 (0.07)

[1]ADG=average daily gain and ADFI=average daily feed intake
[2]NSB=sire breed was not included as a fixed effect in the model
[3]BayesIM8=BayesIM 8 clusters
[4]BayesIM16=BayesIM 16 clusters

Table 5. Rank correlations for ADG and ADFI across univariate analyses from SB [1,2]

	ADG		ADFI	
	[3]BayesIM8	[4]BayesIM16	[3]BayesIM8	[4]BayesIM16
BayesC	0.95	0.95	0.96	0.96
[3]BayesIM8	-	1.00	-	1.00

[1]ADG=average daily gain and ADFI=average daily feed intake
[2]SB=sire breed was included as a fixed effect in the model
[3]BayesIM8=BayesIM 8 clusters
[4]BayesIM16=BayesIM 16 clusters

Table 6. Rank correlations for ADG and ADFI across univariate analyses from NSB [1,2]

	ADG		ADFI	
	[3]BayesIM8	[4]BayesIM16	[3]BayesIM8	[4]BayesIM16
BayesC	0.95	0.95	0.95	0.95
[3]BayesIM8	-	1.00	-	1.00

[1]ADG=average daily gain and ADFI=average daily feed intake
[2]NSB=sire breed was not included as a fixed effect in the model
[3]BayesIM8=BayesIM 8 clusters
[4]BayesIM16=BayesIM 16 clusters

Table 7. Chromosome and position of the top 1% 1-Mb windows for ADG and ADFI from each univariate analysis from SB[1,2,3]

Chromosome	ADG			ADFI		
	BayesC	[4]BayesIM8	[5]BayesIM16	BayesC	[4]BayesIM8	[5]BayesIM16
1		156-157	156-157	4-5, 150-151, 151-152	57-58, 151-152, 155-156, 157-158	151-152, 153-154, 155-156, 157-158
2		103-104		14-15, 68-69		
3	119-120	20-21, 119-120	119-120	98-99		
4	5-6	5-6, 35-36	5-6			
6	38-39			14-15		16-17
7			59-60, 60-61	102-103		9-10
8	1-2, 72-73	1-2, 72-73		82-83		109-110
9		21-22, 53-54	21-22	101-102	6-7	19-20
10	27-28, 45-46, 68-69, 101-102	27-28		20-21	53-54, 55-56	53-54
11	7-8	4-5	4-5, 5-6			106-107
13	75-76, 81-82, 83-84	81-82, 83-84	81-82, 83-84	79-80, 82-83	83-84	82-83, 83-84
14	13-14, 24-25	24-25	24-25, 26-27	9-10	29-30	
15	81-82	68-69	67-68	20-21, 44-45, 82-83	47-48, 56-57, 71-72	45-46, 46-47, 62-63, 64-65, 67-68, 68-69
16		24-25				
17			63-64		14-15	4-5
18	63-64		10-11		40-41	54-55
19	25-26	21-22		17-18, 30-31		
20	6-7	6-7, 7-8	6-7, 7-8			16-17
21				25-26, 27-28	21-22, 26-27, 30-31, 31-32	27-28, 31-32
22	9-10, 44-45, 45-46, 55-56	32-33, 44-45, 45-46, 46-47	17-18, 43-44, 44-45, 45-46, 46-47		9-10	
24				7-8		
25				6-7	6-7, 34-35	34-35
26	19-20		19-20			
27	15-16	29-30	32-33	31-32	10-11, 31-32	32-33
29				34-35		

[1]Position refers to the location in megabases (Mb) for a particular chromosome derived from the *Bos Taurus* build UMD_3.1 assembly (Zimin et al., 2009)
[2]ADG=average daily gain and ADFI=average daily feed intake
[3]SB=sire breed was included as a fixed effect in the model
[4]BayesIM8=BayesIM 8 clusters
[5]BayesIM16=BayesIM 16 clusters

Table 8. Chromosome and position of the top 1% 1-Mb windows for ADG and ADFI from each univariate analysis from NSB[1,2,3]

Chromosome	ADG			ADFI		
	BayesC	[4]BayesIM8	[5]BayesIM16	BayesC	[4]BayesIM8	[5]BayesIM16
1		156-157	156-157	4-5, 150-151, 151-152, 156-157	37-38, 57-58, 151-152	151-152, 154-155, 157-158
2				68-69	68-69	
3	96-97, 119-120	119-120	96-97, 119-120	59-60, 69-70, 90-91, 98-99	90-91	79-80
4	5-6, 7-8	5-6				
5				73-74		
6	38-39	94-95		14-15, 24-25	8-9	7-8, 8-9, 16-17, 17-18
7	64-65	59-60, 60-61	59-60, 60-61		14-15	14-15
8	1-2, 72-73	1-2, 72-73		82-83, 99-100	107-108	108-109, 109-110
9		21-22	20-21		19-20, 65-66	
10	27-28, 101-102,	27-28			53-54	53-54
11	21-22	4-5, 5-6, 72-73	4-5, 5-6			
13	75-76, 83-84	81-82, 83-84	83-84	38-39, 79-80, 82-83	41-42, 83-84	82-83, 83-84
14	13-14, 24-25, 26-27					
15	68-69	68-69		15-16, 20-21	16-17, 20-21, 71-72, 80-81	25-26, 45-46, 46-47, 56-57, 62-63, 67-68
16			24-25			
17		67-68		31-32	9-10, 14-15	4-5, 14-15, 31-32
18	63-64					
19	25-26	61-62		17-18, 60-61		
20	6-7	4-5, 6-7	3-4, 6-7, 7-8			
21				26-27	15-16, 26-27, 27-28	27-28
22	44-45, 45-46, 55-56	44-45, 45-46	16-17, 17-18, 34-35,43-44, 44-45, 45-46, 56-57			
23		2-3			43-44	
25				6-7	6-7, 34-35	
26	19-20	31-32	32-33	37-38		
27		29-30	32-33			12-13
28			36-37			
29	44-45					

[1]Position refers to the location in megabases (Mb) for a particular chromosome derived from the *Bos Taurus* build UMD_3.1 assembly (Zimin et al., 2009)
[2]ADG=average daily gain and ADFI=average daily feed intake
[3]NSB=sire breed was not included as a fixed effect in the model
[4]BayesIM8=BayesIM 8 clusters
[5]BayesIM16=BayesIM 16 clusters

Table 9. Chromosome and position of the top 1% 1-Mb windows for ADG and ADFI from the bivariate haplotype model from SB[1,2,3]

Chromosome	[4]BayesIM8 Bivariate	[5]BayesIM16 bivariate
1	151-152, 155-156	155-156
2	136-137	
3	119-120	96-97
6	14-15	
7		14-15
8	107-108	
9	21-22	
10	41-42, 53-54	41-42
11	4-5	4-5
13	83-84	82-83, 83-84
14		24-25, 26-27, 30-31
15		67-68
17	63-64, 67-68	
18		40-41
19	17-18	17-18
20	6-7	3-4, 6-7, 7-8, 16-17
21	15-16, 49-50	21-22, 27-28
22	29-30, 34-35, 44-45	29-30, 37-38, 43-44, 44-45, 46-47
23	2-3	
25	6-7, 34-35	34-35
27	4-5, 10-11	

[1]Position refers to the location in megabases (Mb) for a particular chromosome derived from the *Bos Taurus* build UMD_3.1 assembly (Zimin et al., 2009)
[2]ADG=average daily gain and ADFI=average daily feed intake
[3]SB=sire breed was included as a fixed effect in the model
[4]BayesIM8=BayesIM 8 clusters
[5]BayesIM16=BayesIM 16 clusters

Table 10. Chromosome and position of the top 1% 1-Mb windows for ADG and ADFI from the bivariate haplotype association from NSB[1,2,3]

Chromosome	[4]BayesIM8 Bivariate	[5]BayesIM16 bivariate
1	57-58, 155-156	155-156, 157-158
2	68-69, 125-126	
3	90-91, 119-120	79-80, 90-91, 92-93
7		13-14, 14-15
8	86-87	
9	21-22	
10		41-42, 50-51, 55-56
11		4-5
13	83-84	82-83, 83-84
15		47-48, 56-57, 64-65, 67-68, 71-72
17	31-32, 67-68	31-32
19	17-18	
20	6-7	
21	15-16, 49-50	21-22, 25-26
22	9-10, 34-35, 44-45	
24	46-47	
25	6-7, 34-35	34-35
27	4-5, 10-11, 29-30	27-28, 31-32
29	44-45	44-45

[1]Position refers to the location in megabases (Mb) for a particular chromosome derived from the *Bos Taurus* build UMD_3.1 assembly (Zimin et al., 2009)
[2]ADG=average daily gain and ADFI=average daily feed intake
[3]NSB=sire breed was not included as a fixed effect in the model
[4]BayesIM8=BayesIM 8 clusters
[5]BayesIM16=BayesIM 16 clusters

Table 11. Proportion of the top 1% 1-Mb windows shared between univariate analyses for ADG and ADFI from SB[1,2]

	ADG		ADFI	
	[3]BayesIM8	[4]BayesIM16	[3]BayesIM8	[4]BayesIM16
BayesC	0.44	0.32	0.12	0.08
[3]BayesIM8	-	0.52	-	0.20

[1]ADG=average daily gain and ADFI=average daily feed intake
[2]SB=sire breed was included as a fixed effect in the model
[3]BayesIM8=BayesIM 8 clusters
[4]BayesIM16=BayesIM 16 clusters

Table 12. Proportion of the top 1% 1-Mb windows shared between univariate analyses for ADG and ADFI from NSB[1,2]

	ADG		ADFI	
	[3]BayesIM8	[4]BayesIM16	[3]BayesIM8	[4]BayesIM16
BayesC	0.40	0.24	0.28	0.12
[3]BayesIM8	-	0.44	-	0.24

[1]ADG=average daily gain and ADFI=average daily feed intake
[2]NSB=sire breed was not included as a fixed effect in the model
[3]BayesIM8=BayesIM 8 clusters
[4]BayesIM16=BayesIM 16 clusters

Table 13. Proportion of the top 1% 1-Mb windows shared between BayesIM8 univariate analyses for ADG and ADFI and BayesIM8 bivariate analysis from SB and NSB[1,2,3,4]

	SB		NSB	
	ADFI	Bivariate	ADFI	Bivariate
ADG	0.04	0.24	0	0.28
ADFI	-	0.24	-	0.20

[1] BayesIM8=BayesIM 8 clusters
[2] ADG=average daily gain and ADFI=average daily feed intake
[3]SB=sire breed is included as a fixed effect in the model
[4]NSB=sire breed is not included as a fixed effect in the model

Table 14. Proportion of the top 1% 1-Mb windows shared between BayesIM16 univariate analyses for ADG and ADFI and BayesIM16 bivariate analysis from SB and NSB[1,2,3,4]

	SB		NSB	
	ADFI	Bivariate	ADFI	Bivariate
ADG	0.12	0.40	0.04	0.08
ADFI	-	0.28	-	0.20

[1]BayesIM16=BayesIM 16 clusters
[2]ADG=average daily gain and ADFI=average daily feed intake
[3]SB=sire breed is included as a fixed effect in the model
[4]NSB=sire breed is not included as a fixed effect in the model

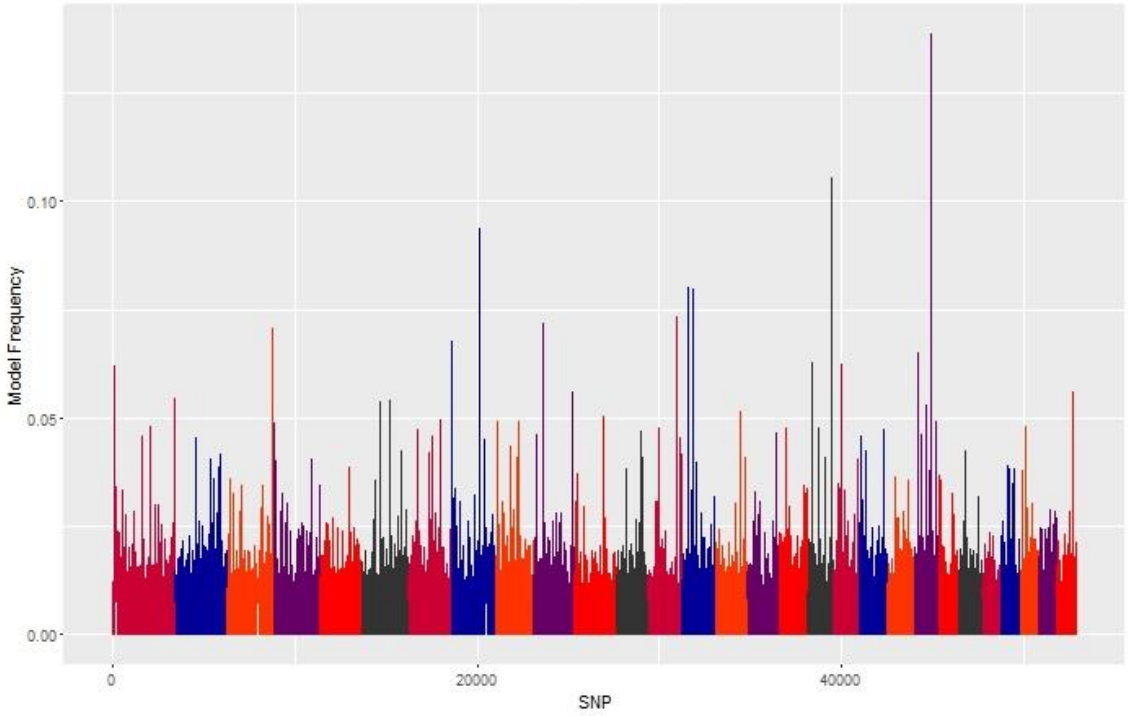

Figure 1. Genome-wide association analysis between SNP genotypes and average daily gain (ADG) from BayesC when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each marker. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

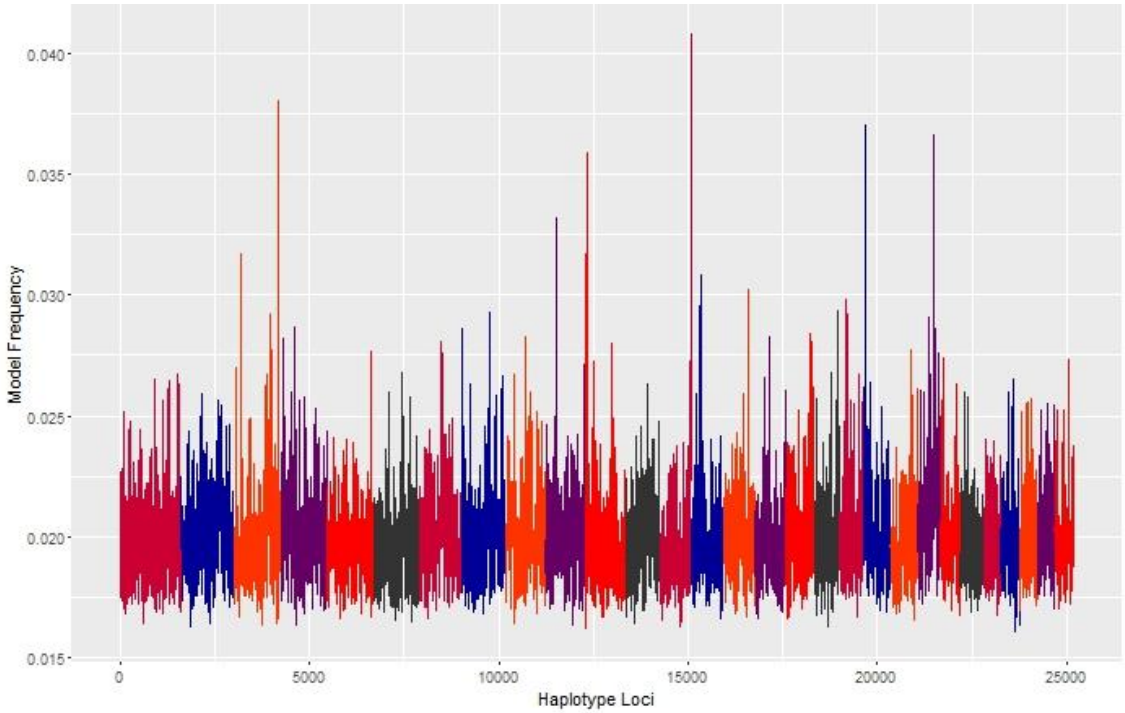

Figure 2. Genome-wide association analysis between haplotype genotypes and average daily gain (ADG) from BayesIM 8 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each haplotype loci. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

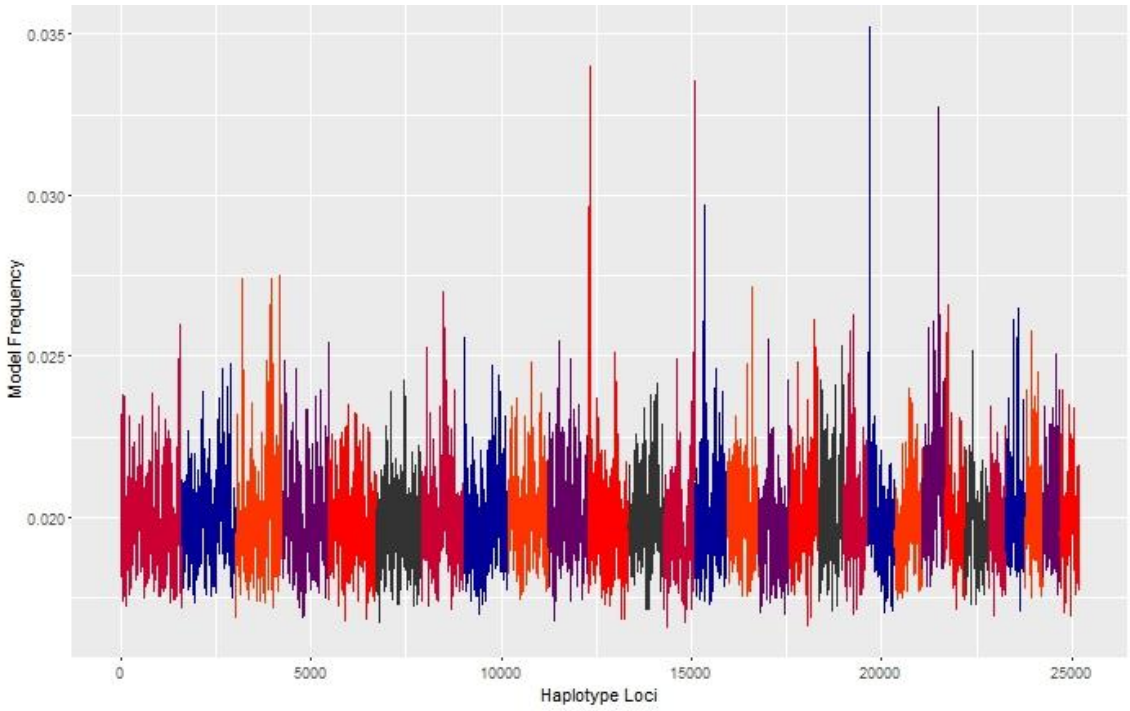

Figure 3. Genome-wide association analysis between haplotype genotypes and average daily gain (ADG) from BayesIM 16 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each haplotype loci. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

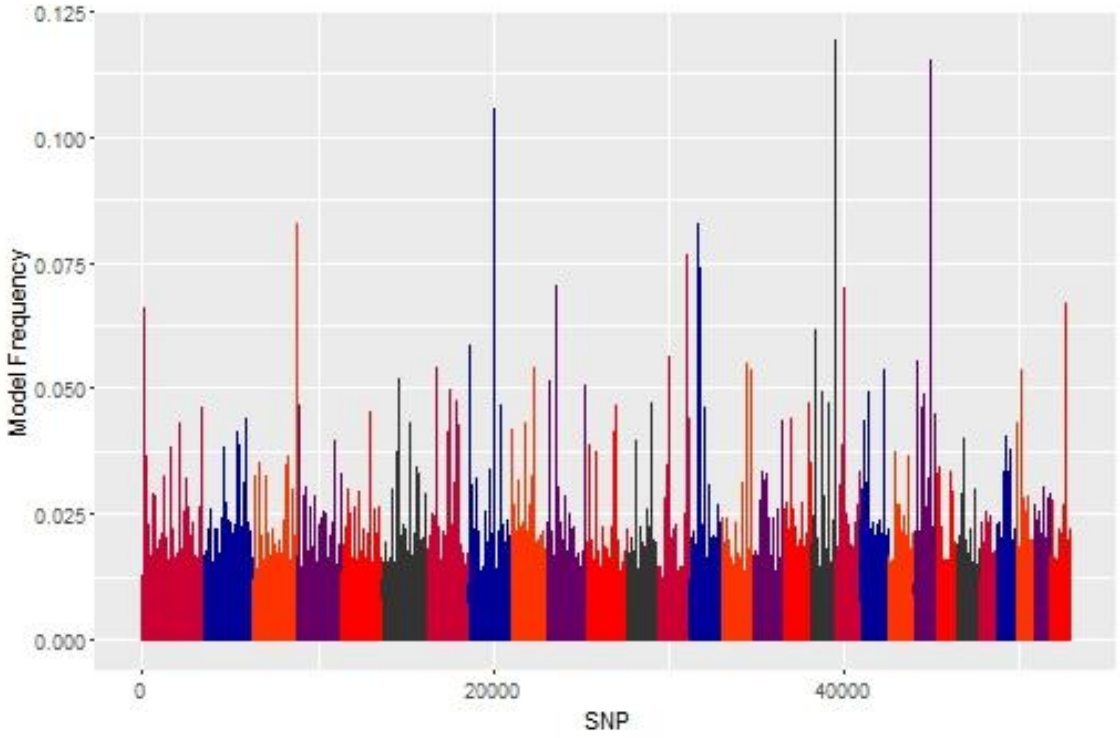

Figure 4. Genome-wide association analysis between SNP genotypes and average daily gain (ADG) from BayesC when sire breed was not included as a fixed effect in the model (NSB). The Y-axis represents the model frequency of each marker. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

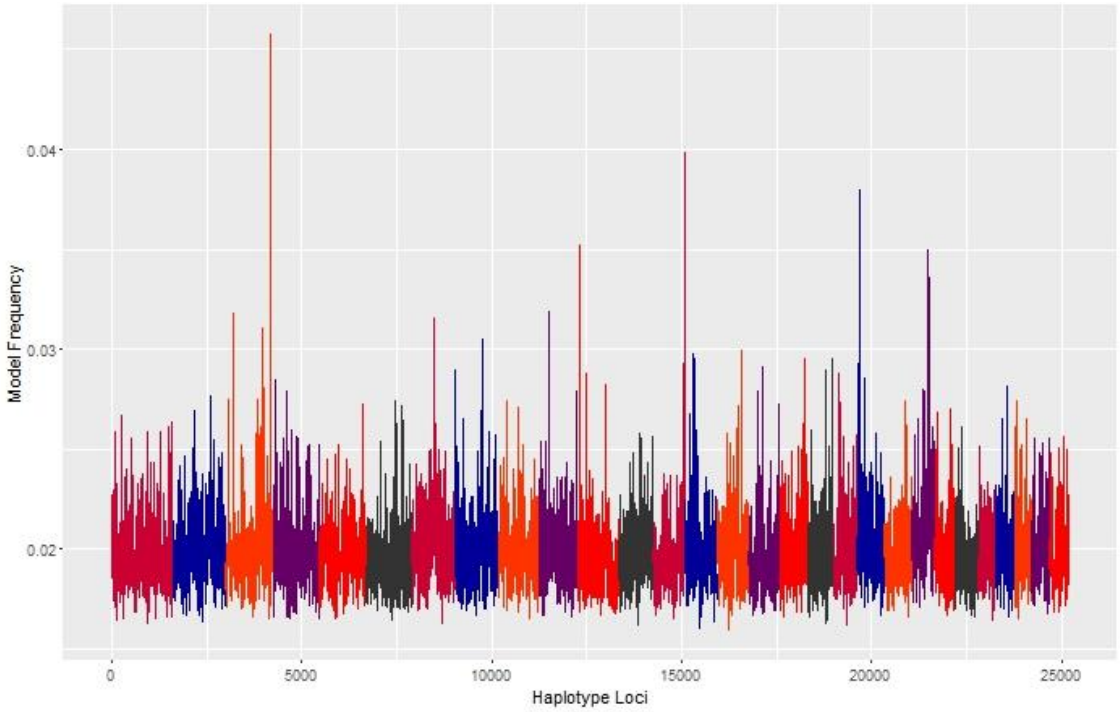

Figure 5. Genome-wide association analysis between haplotype genotypes and average daily gain (ADG) from BayesIM 8 clusters when sire breed was not included as a fixed effect in the model (NSB). The Y-axis represents the model frequency of each haplotype loci. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

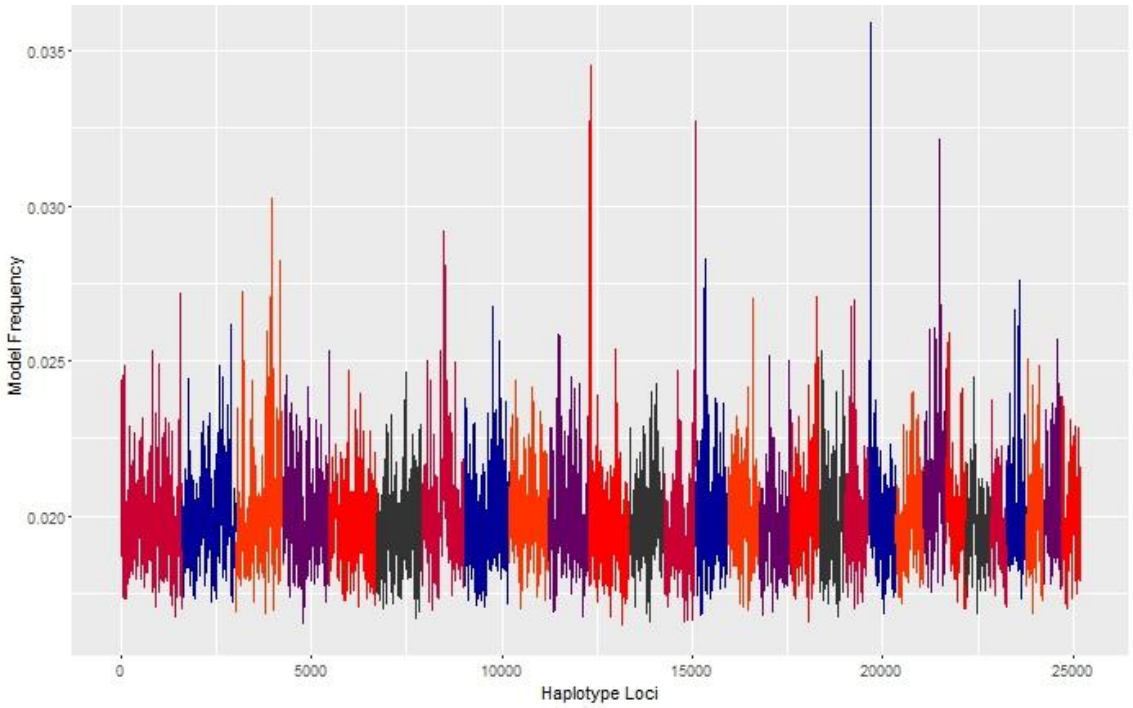

Figure 6. Genome-wide association analysis between haplotype genotypes and average daily gain (ADG) from BayesIM 16 clusters when sire breed was not included as a fixed effect in the model (NSB). The Y-axis represents the model frequency of each haplotype loci. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

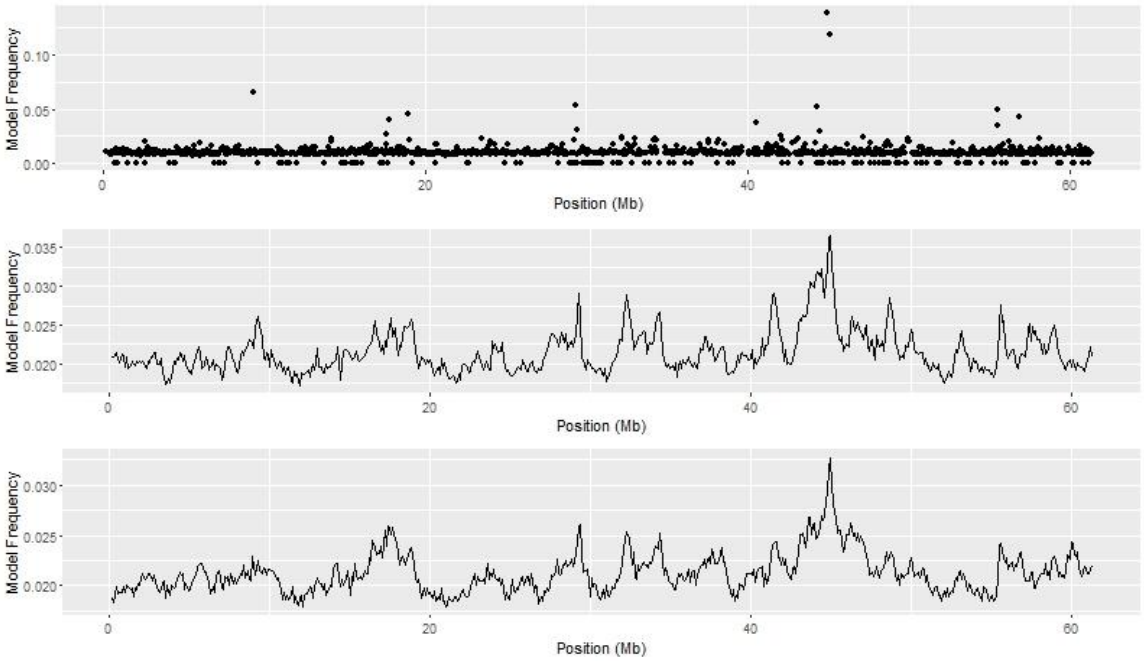

Figure 7. BTA 22 from genome-wide association analyses for average daily gain (ADG) from BayesC (top), BayesIM 8 clusters (middle) and BayesIM 16 clusters (bottom) when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each marker or haplotype loci. The X-axis is position in megabases (Mb) on BTA 22.

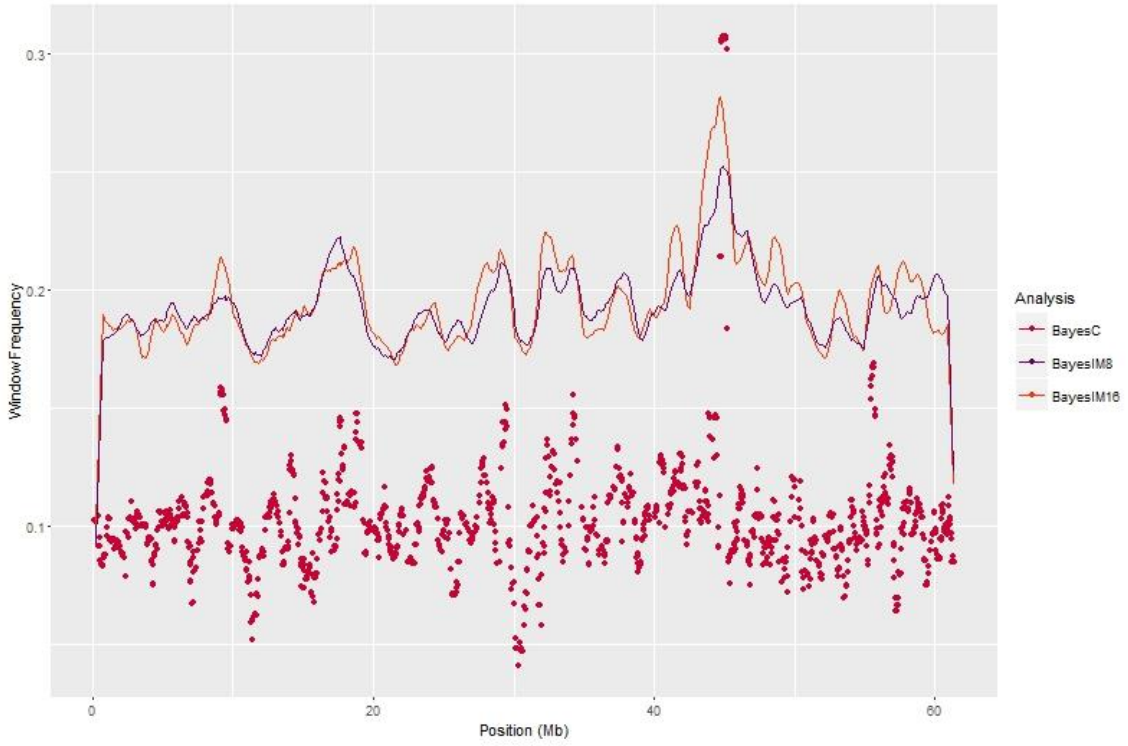

Figure 8. BTA 22 from genome-wide association analyses for average daily gain (ADG) from BayesC, BayesIM 8 clusters and BayesIM 16 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the window frequency of each marker or haplotype loci. The X-axis is position in megabases (Mb) on BTA 22.

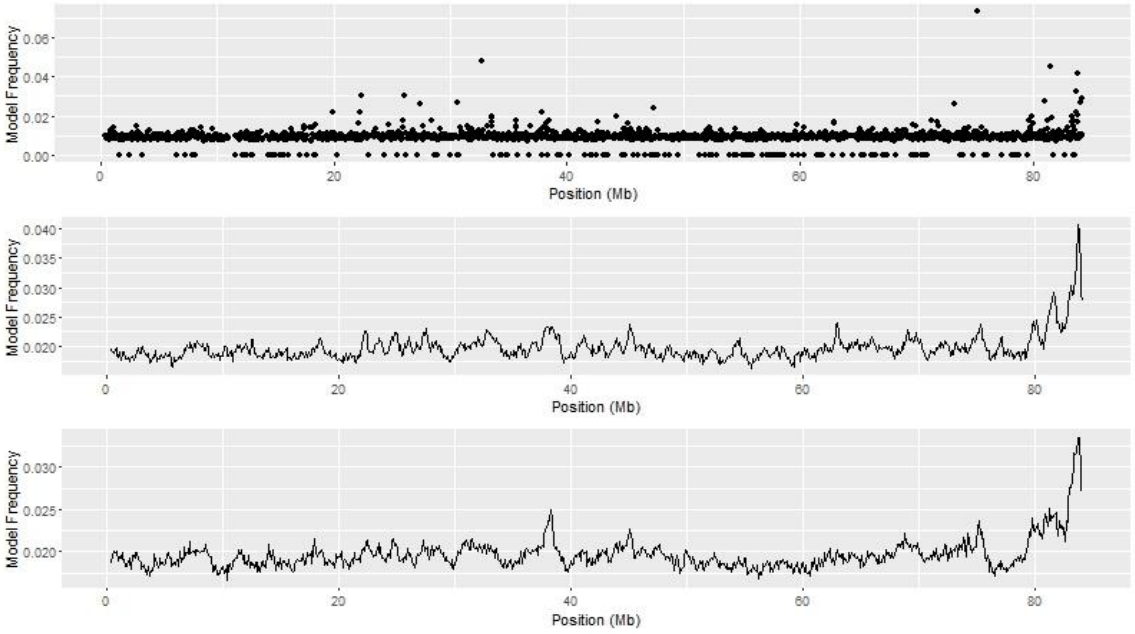

Figure 9. BTA 13 from genome-wide association analyses for average daily gain (ADG) from BayesC (top), BayesIM 8 clusters (middle) and BayesIM 16 clusters (bottom) when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each marker or haplotype loci. The X-axis is position in megabases (Mb) on BTA 13.

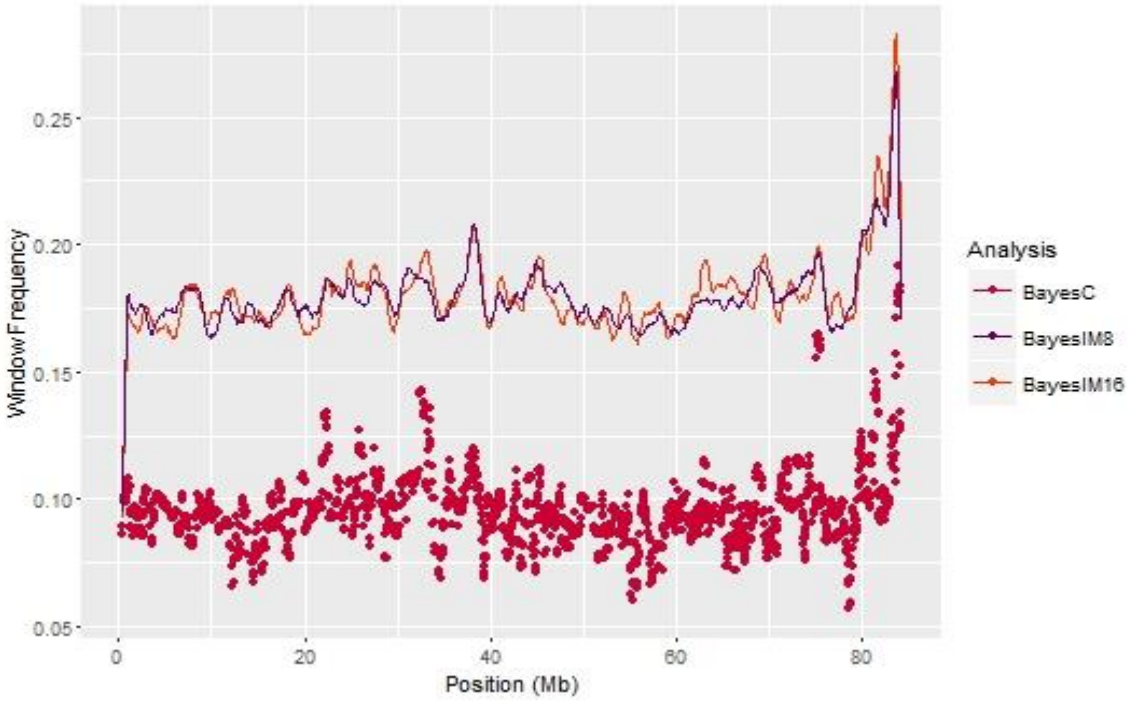

Figure 10. BTA 13 from genome-wide association analyses for average daily gain (ADG) from BayesC, BayesIM 8 clusters and BayesIM 16 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the window frequency of each marker or haplotype loci. The X-axis is position in megabases (Mb) on BTA 13.

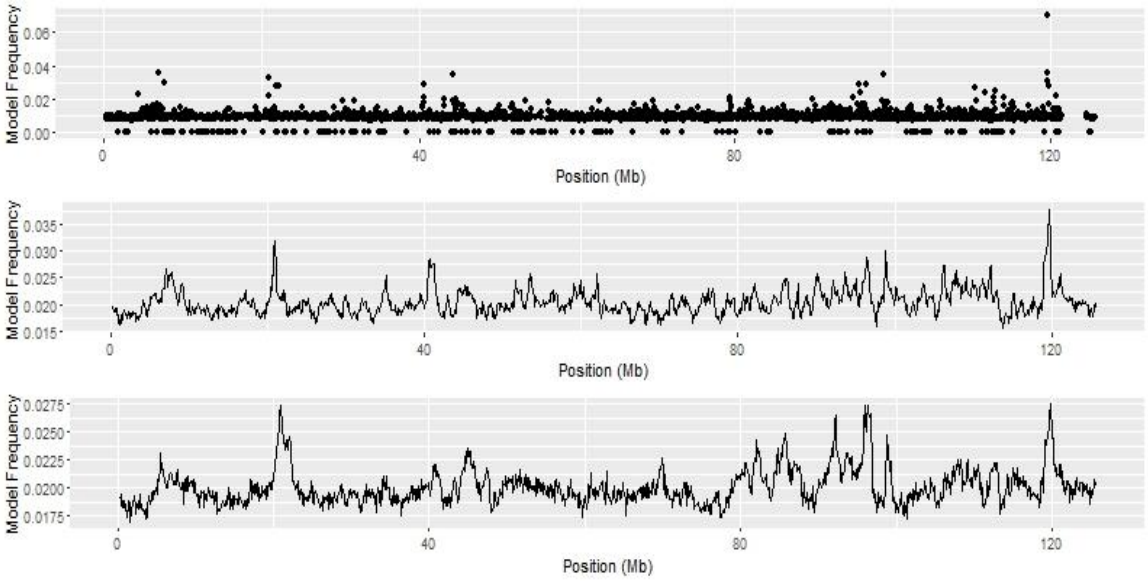

Figure 11. BTA 3 from genome-wide association analyses for average daily gain (ADG) from BayesC (top), BayesIM 8 clusters (middle) and BayesIM 16 clusters (bottom) when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each marker or haplotype loci. The X-axis is position in megabases (Mb) on BTA 3.

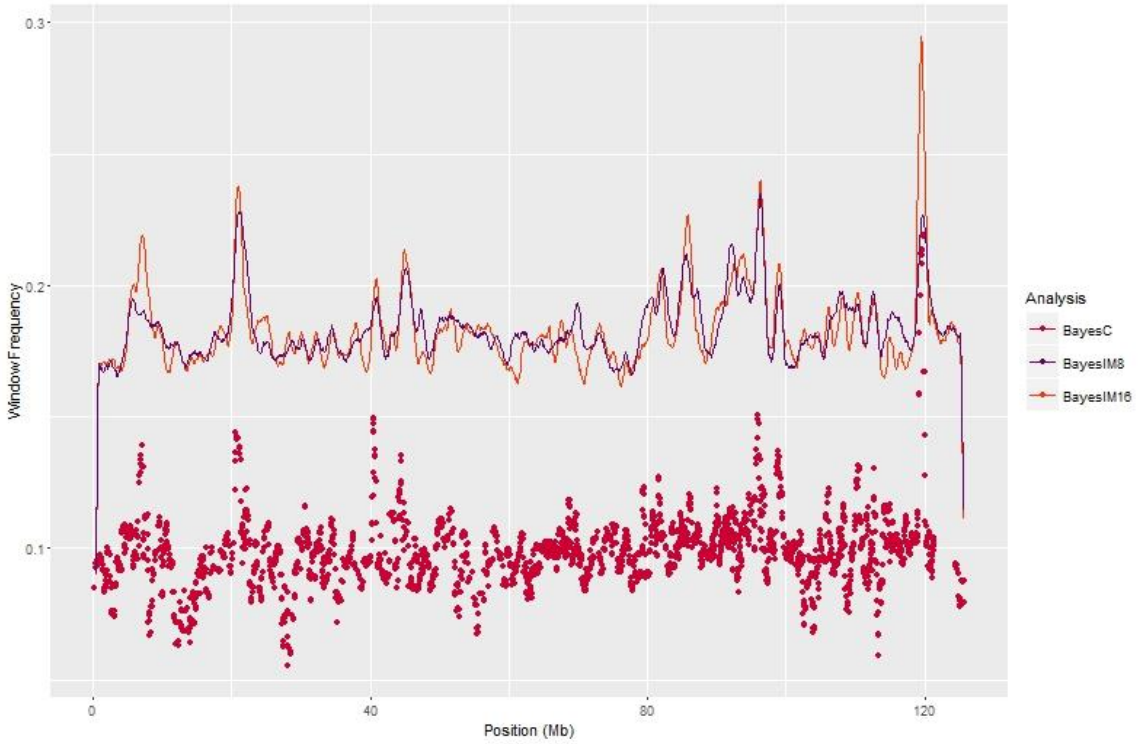

Figure 12. BTA 3 from genome-wide association analyses for average daily gain (ADG) from BayesC, BayesIM 8 clusters and BayesIM 16 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the window frequency of each marker or haplotype loci. The X-axis is position in megabases (Mb) on BTA 3.

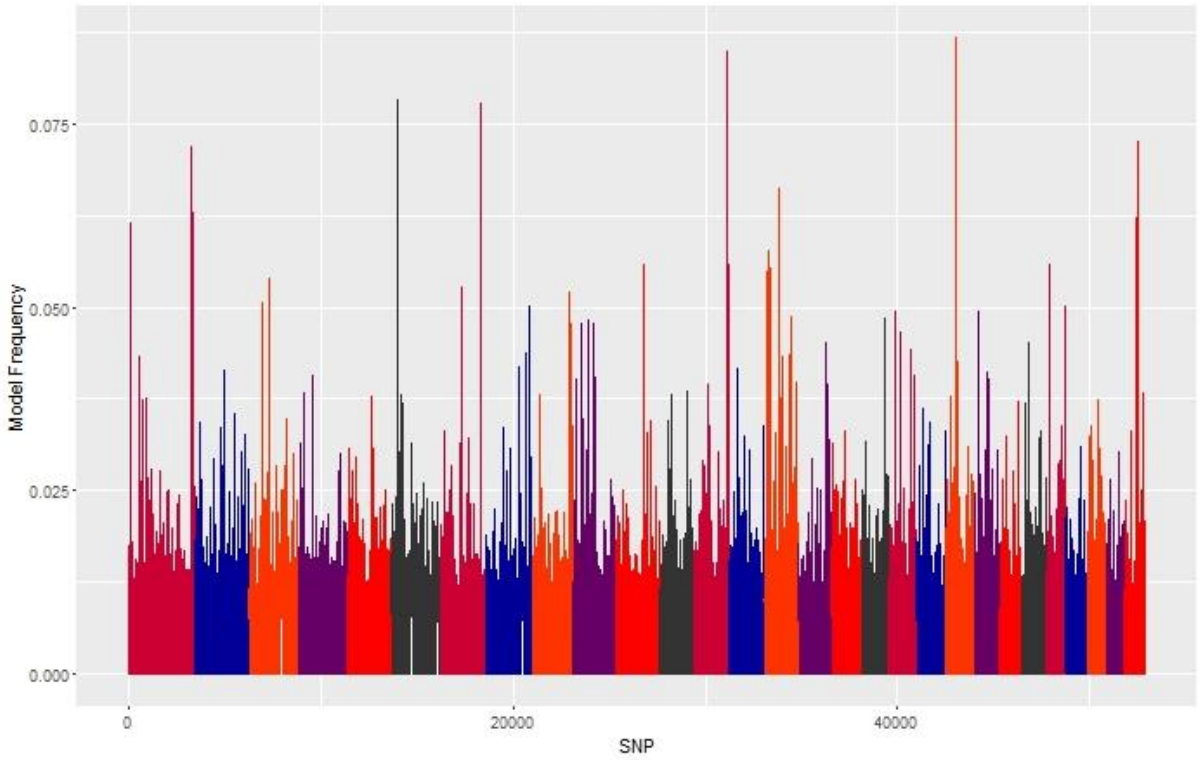

Figure 13. Genome-wide association analysis between SNP genotypes and average daily feed intake (ADFI) from BayesC when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each marker. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

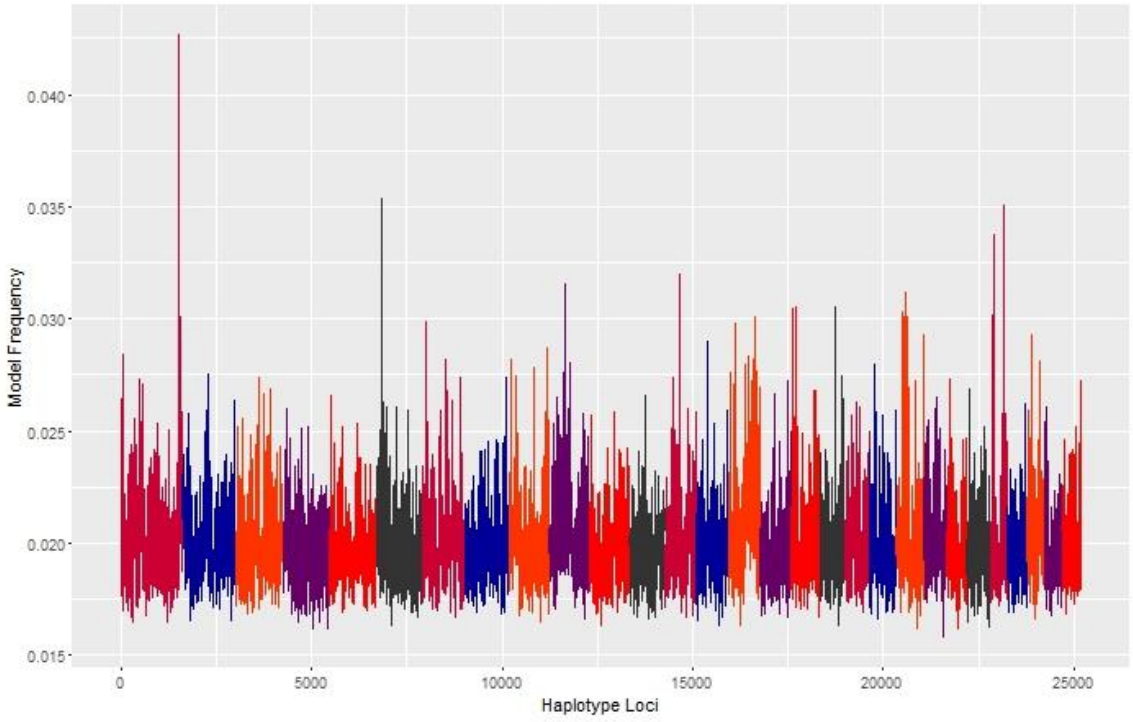

Figure 14. Genome-wide association analysis between haplotype genotypes and average daily feed intake (ADFI) from BayesIM 8 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each haplotype loci. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

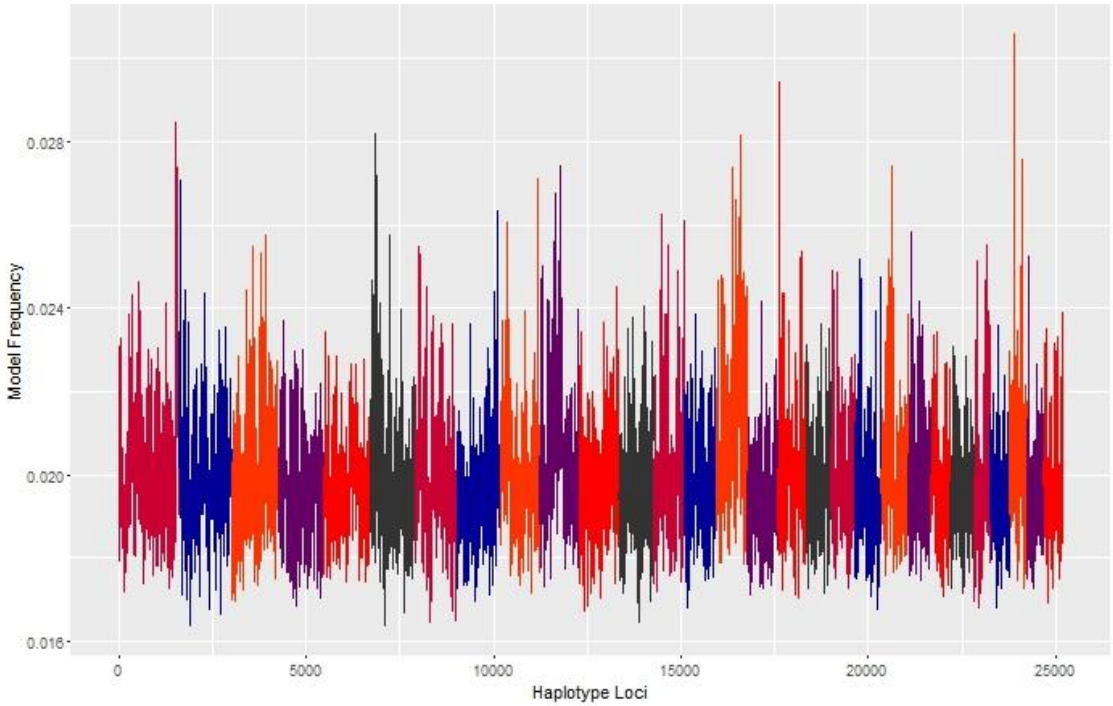

Figure 15. Genome-wide association analysis between haplotype genotypes and average daily feed intake (ADFI) from BayesIM 16 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each haplotype loci. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

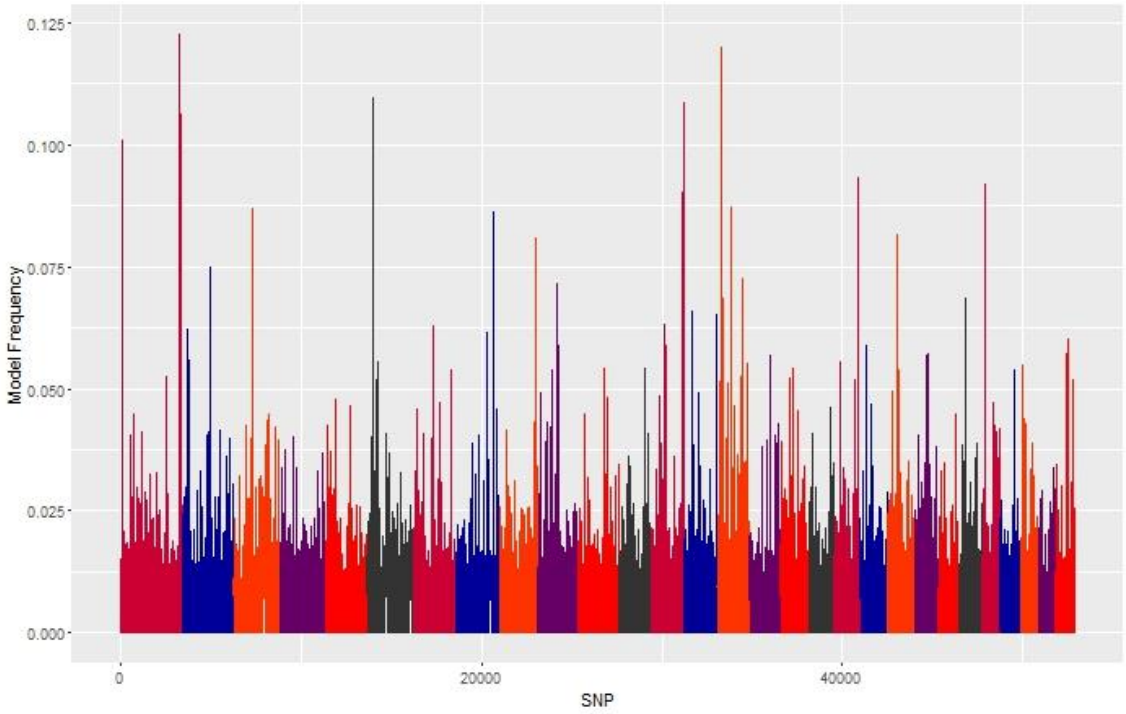

Figure 16. Genome-wide association analysis between SNP genotypes and average feed daily intake (ADFI) from BayesC when sire breed was not included as a fixed effect in the model (NSB). The Y-axis represents the model frequency of each marker. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

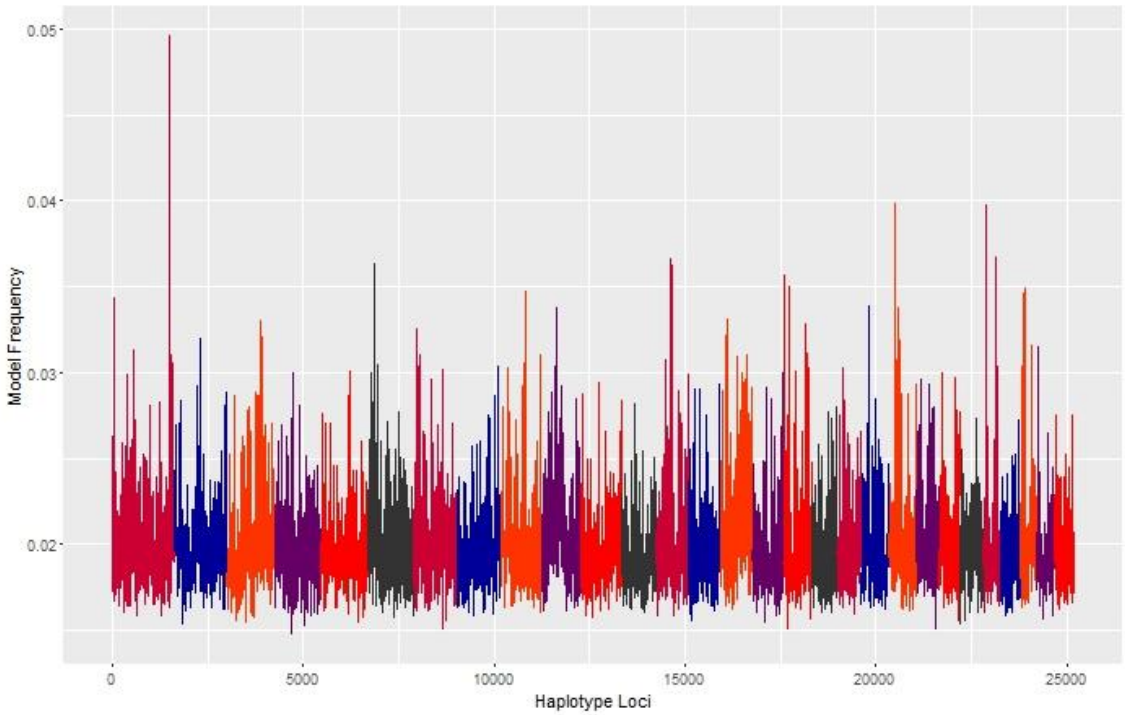

Figure 17. Genome-wide association analysis between haplotype genotypes and average daily feed intake (ADFI) from BayesIM 8 clusters when sire breed was not included as a fixed effect in the model (NSB). The Y-axis represents the model frequency of each haplotype loci. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

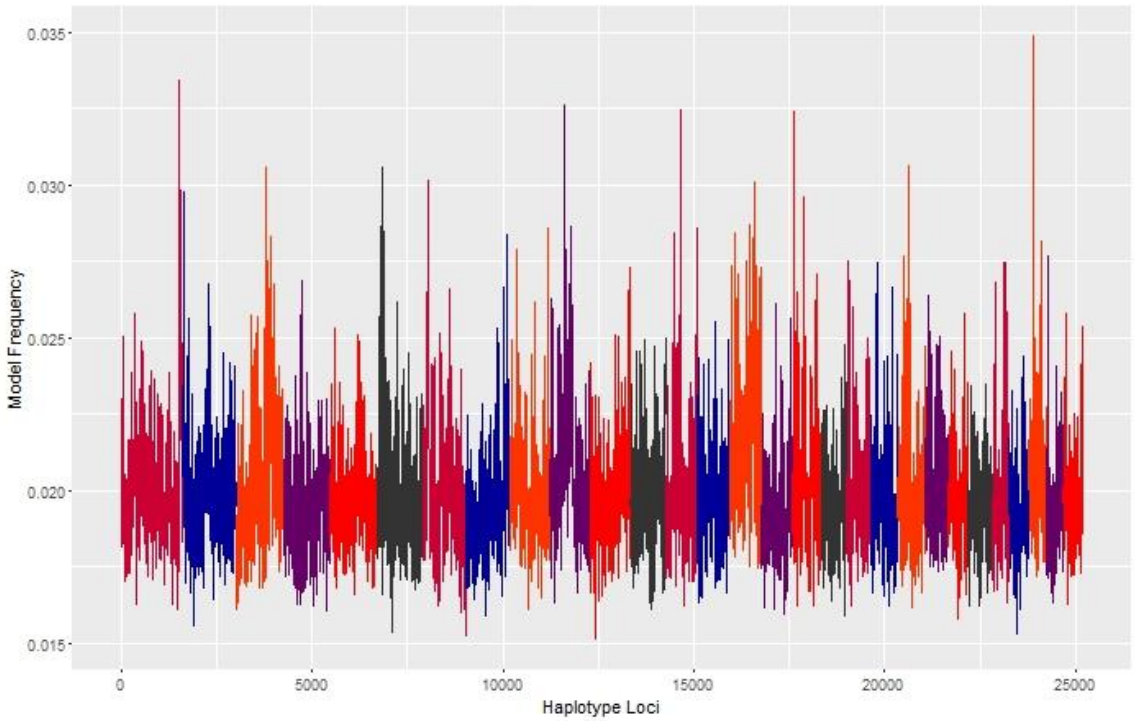

Figure 18. Genome-wide association analysis between haplotype genotypes and average daily feed intake (ADFI) from BayesIM 16 clusters when sire breed was not included as a fixed effect in the model (NSB). The Y-axis represents the model frequency of each haplotype loci. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

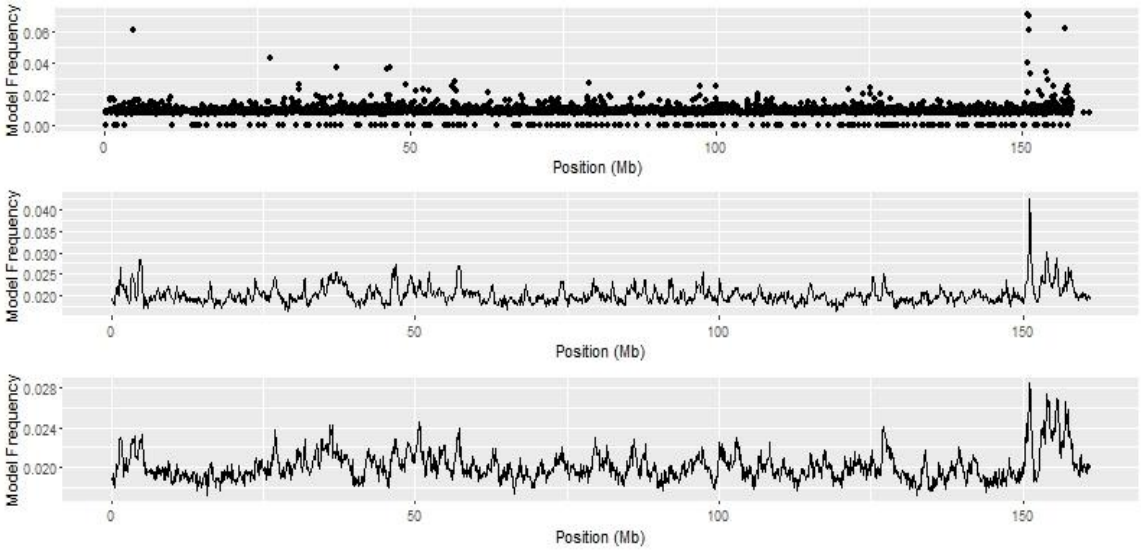

Figure 19. BTA 1 from genome-wide association analyses for average daily feed intake (ADFI) from BayesC (top), BayesIM 8 clusters (middle) and BayesIM 16 clusters (bottom) when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each marker or haplotype loci. The X-axis is position in megabases (Mb) on BTA 1.

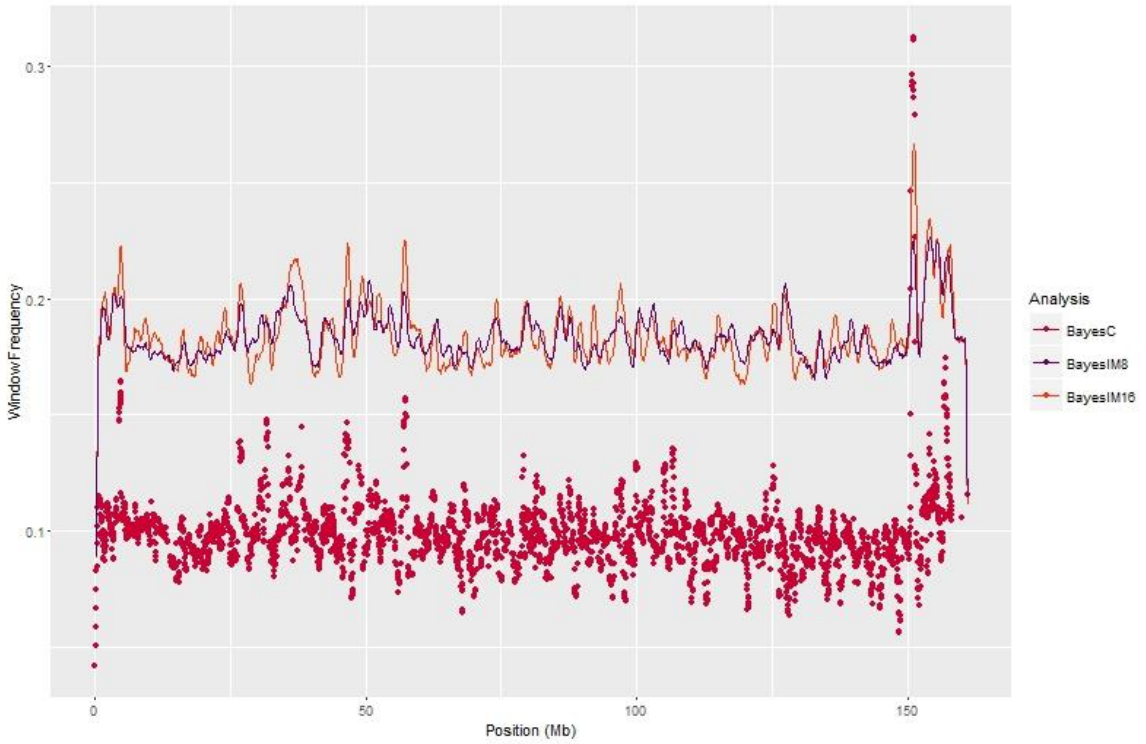

Figure 20. BTA 1 from genome-wide association analyses for average daily feed intake (ADFI) from BayesC, BayesIM 8 clusters and BayesIM 16 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the window frequency of each marker or haplotype loci. The X-axis is position in megabases (Mb) on BTA 1.

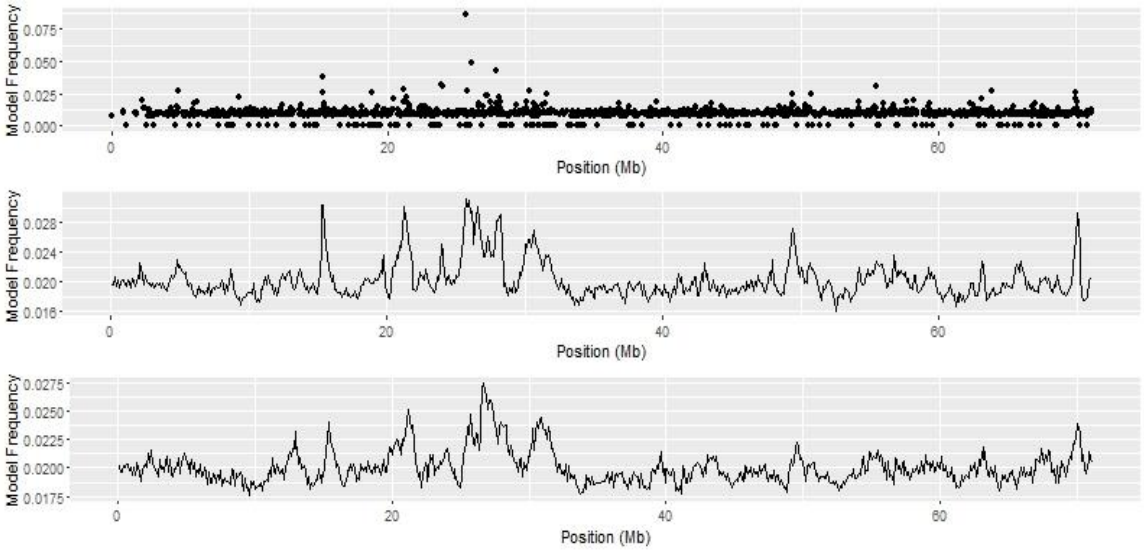

Figure 21. BTA 21 from genome-wide association analyses for average daily feed intake (ADFI) from BayesC (top), BayesIM 8 clusters (middle) and BayesIM 16 clusters (bottom) when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each marker or haplotype loci. The X-axis is position in megabases (Mb) on BTA 21.

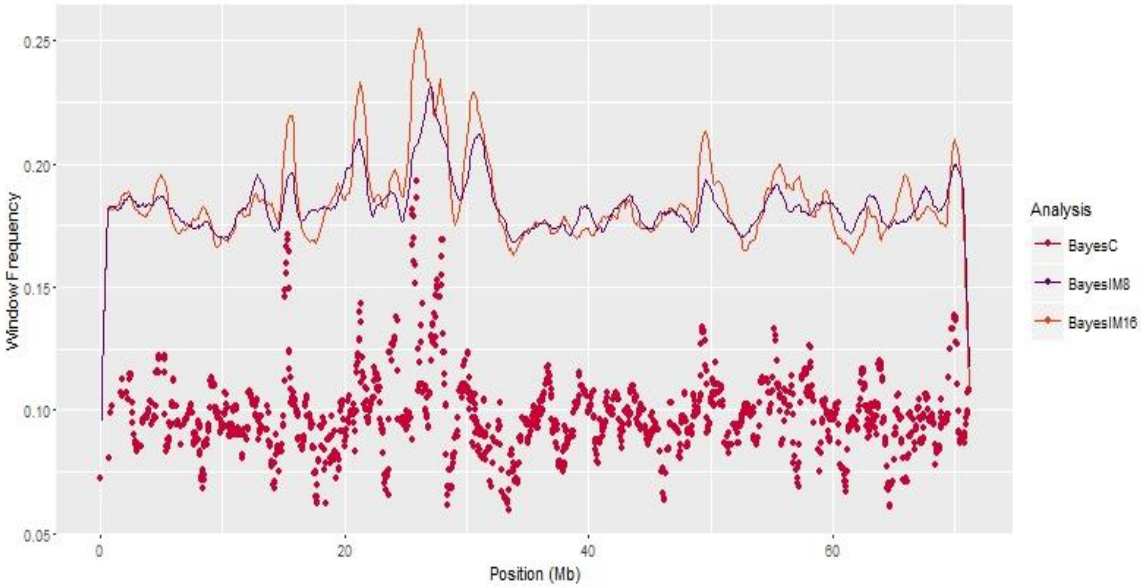

Figure 22. BTA 21 from genome-wide association analyses for average daily feed intake (ADFI) from BayesC, BayesIM 8 clusters and BayesIM 16 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the window frequency of each marker or haplotype loci. The X-axis is position in megabases (Mb) on BTA 21.

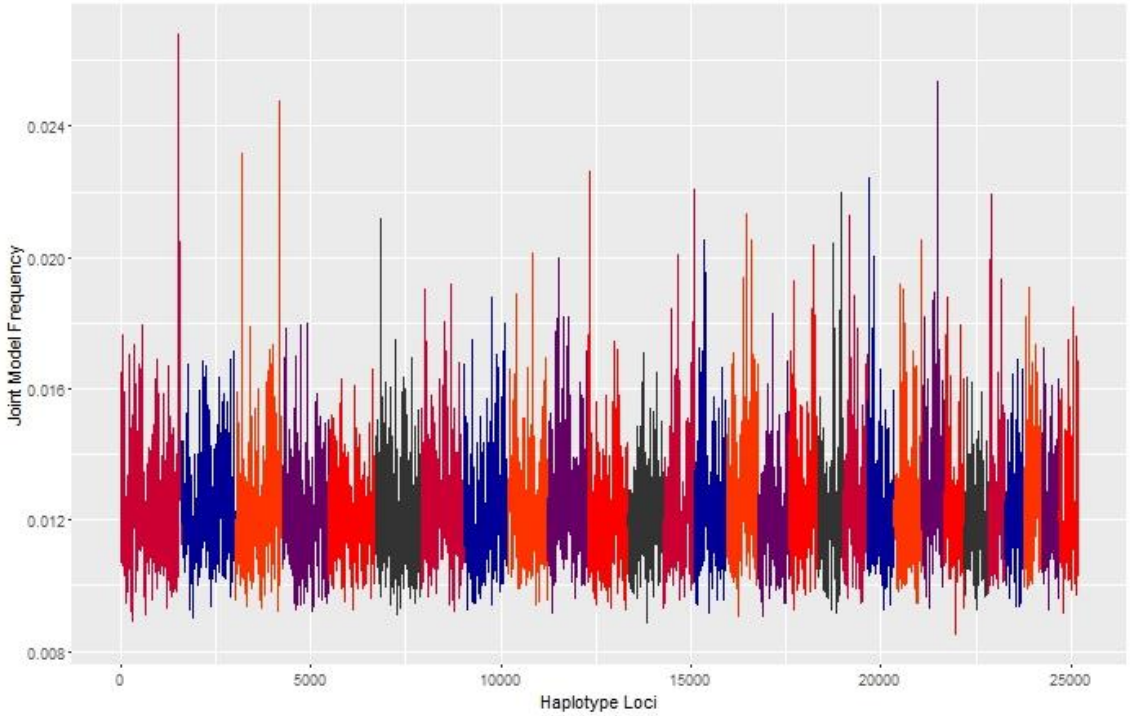

Figure 23. Genome-wide association analysis between haplotype genotypes and average daily gain (ADG) and average daily feed intake (ADFI) from BayesIM bivariate 8 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the joint model frequency of each haplotype loci when both traits are included in the model. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

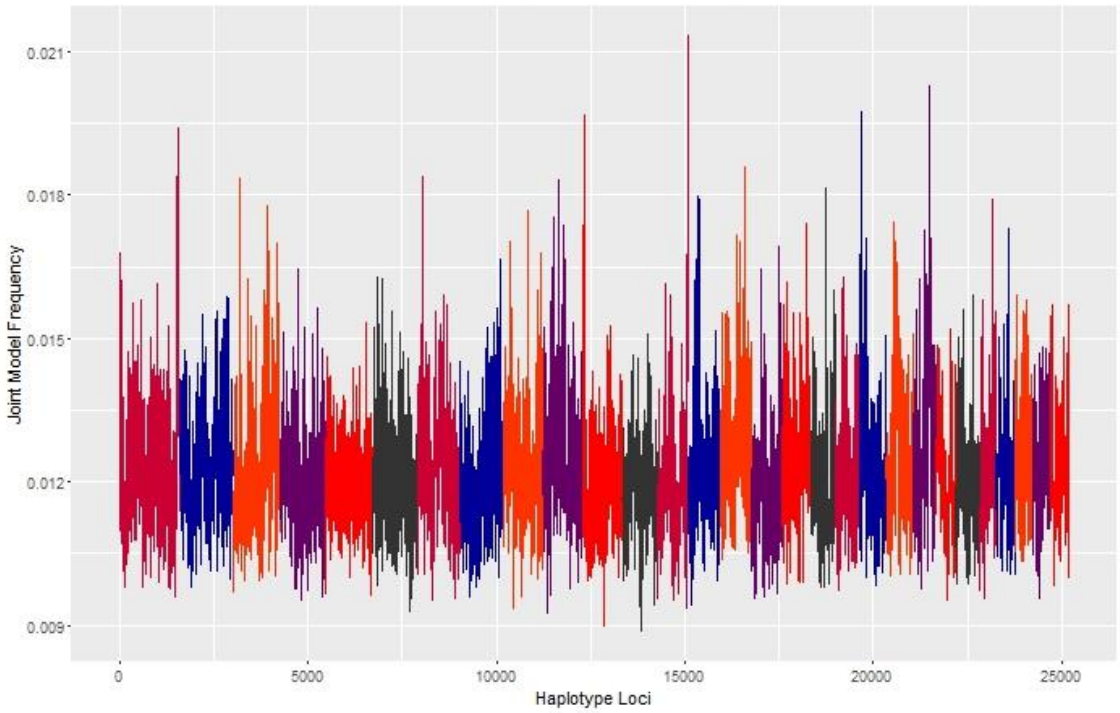

Figure 24. Genome-wide association analysis between haplotype genotypes and average daily gain (ADG) and average daily feed intake (ADFI) from BayesIM bivariate 16 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the joint model frequency of each haplotype loci when both traits are included in the model. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

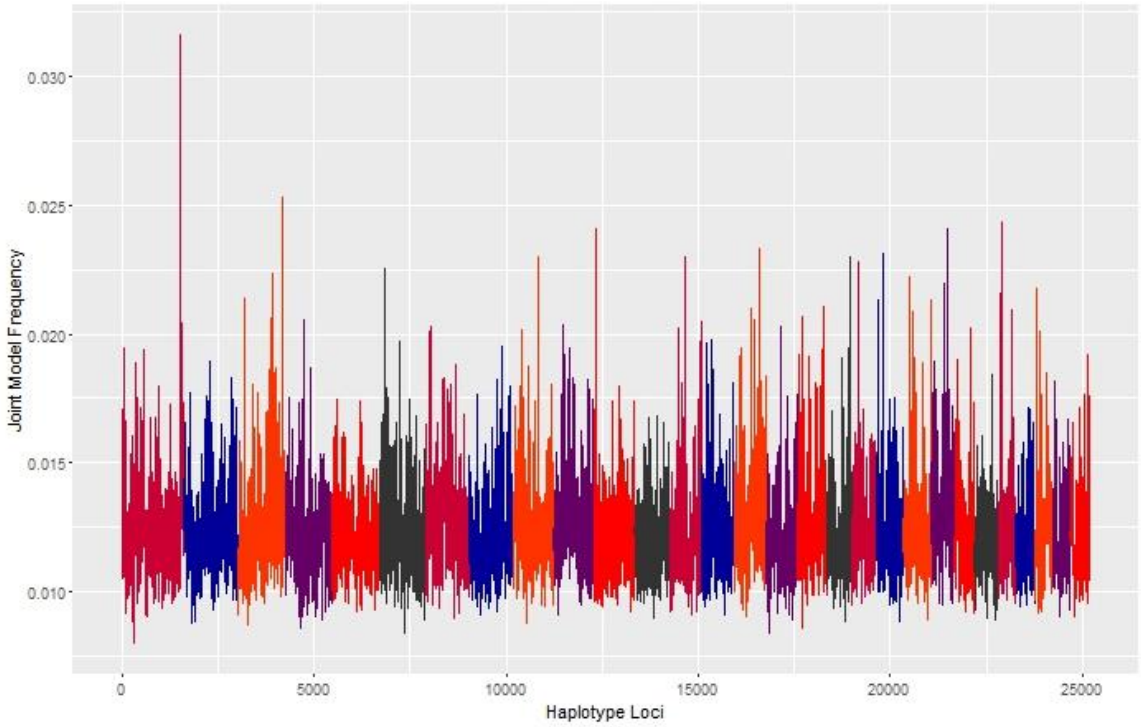

Figure 25. Genome-wide association analysis between haplotype genotypes and average daily gain (ADG) and average daily feed intake (ADFI) from BayesIM bivariate 8 clusters when sire breed was not included as a fixed effect in the model (NSB). The Y-axis represents the joint model frequency of each haplotype loci when both traits are included in the model. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

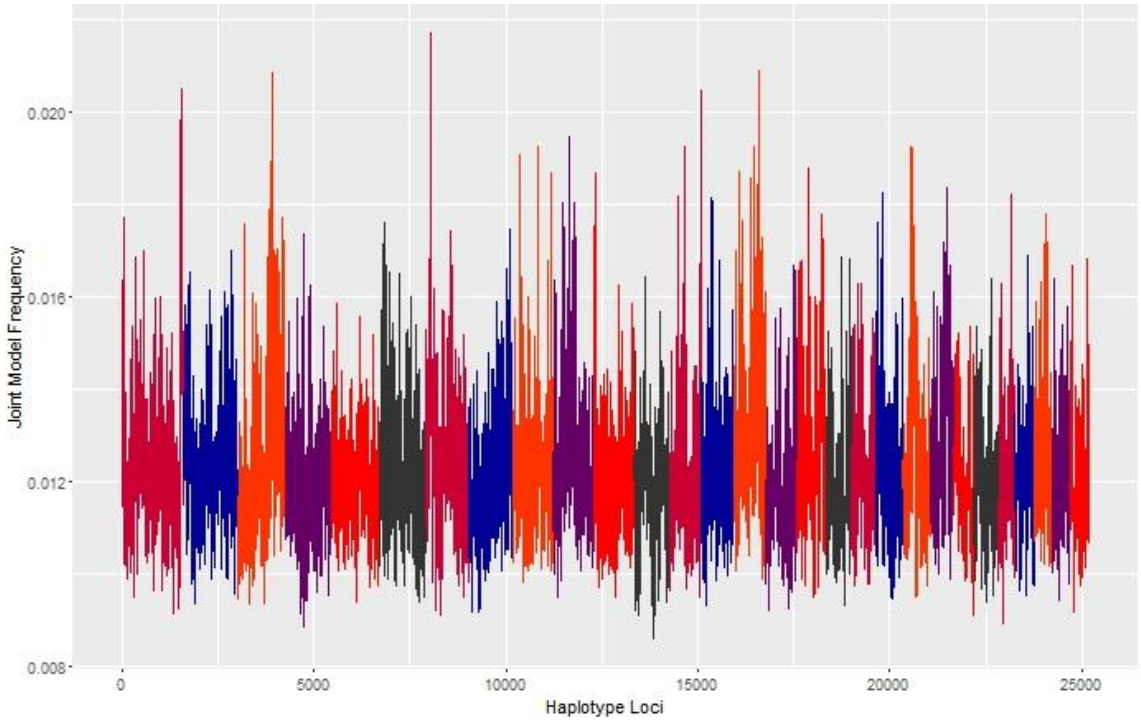

Figure 26. Genome-wide association analysis between haplotype genotypes and average daily gain (ADG) and average daily feed intake (ADFI) from BayesIM bivariate 16 clusters when sire breed was not included as a fixed effect in the model (NSB). The Y-axis represents the joint model frequency of each haplotype loci when both traits are included in the model. On the X-axis, alternate colors represent different chromosomes from BTA 1 to BTA 29.

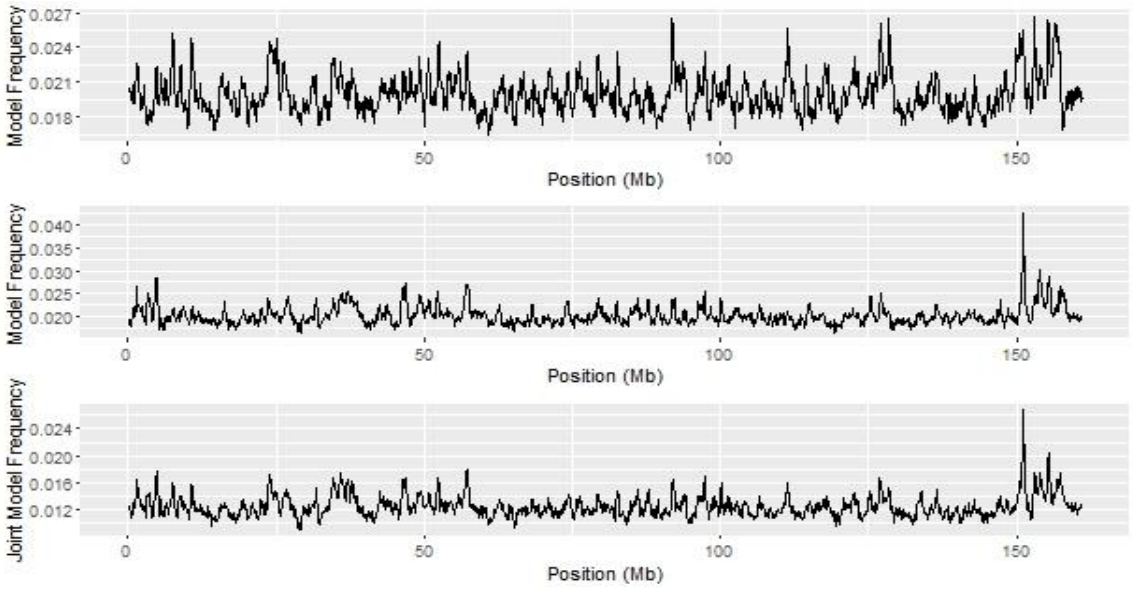

Figure 27. Comparison of univariate, average daily gain (ADG; top) and average daily feed intake (ADFI; middle), and bivariate (bottom) haplotype associations for BTA 1 from BayesIM 8 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each haplotype loci or the joint model frequency when both traits are in the model. The X-axis is position in megabases (Mb) on BTA 1.

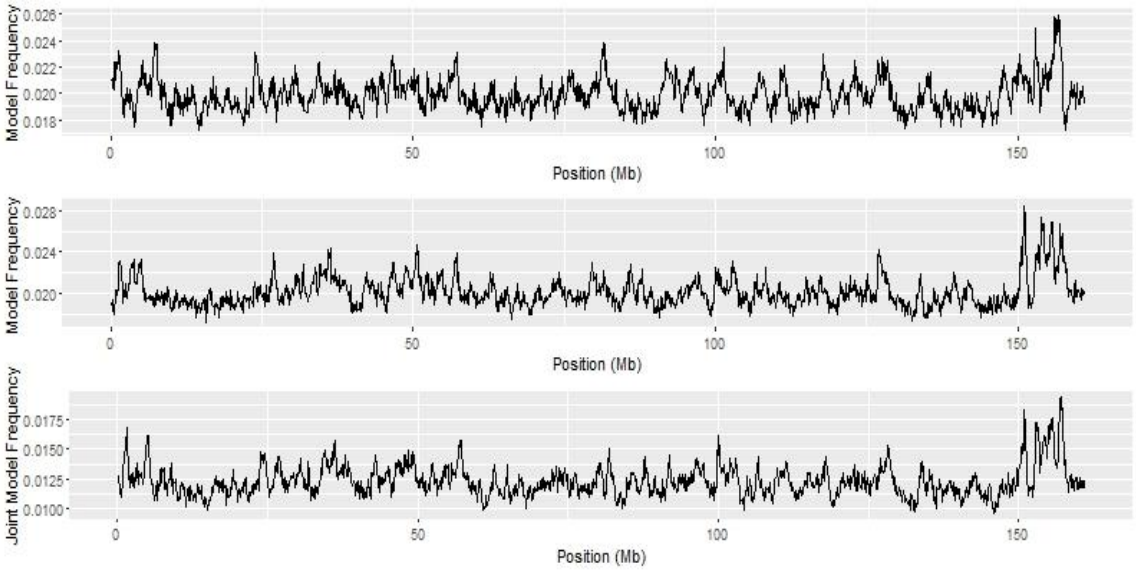

Figure 28. Comparison of univariate, average daily gain (ADG; top) and average daily feed intake (ADFI; middle), and bivariate (bottom) haplotype associations for BTA 1 from BayesIM 16 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each haplotype loci or the joint model frequency when both traits are in the model. The X-axis is position in megabases (Mb) on BTA 1.

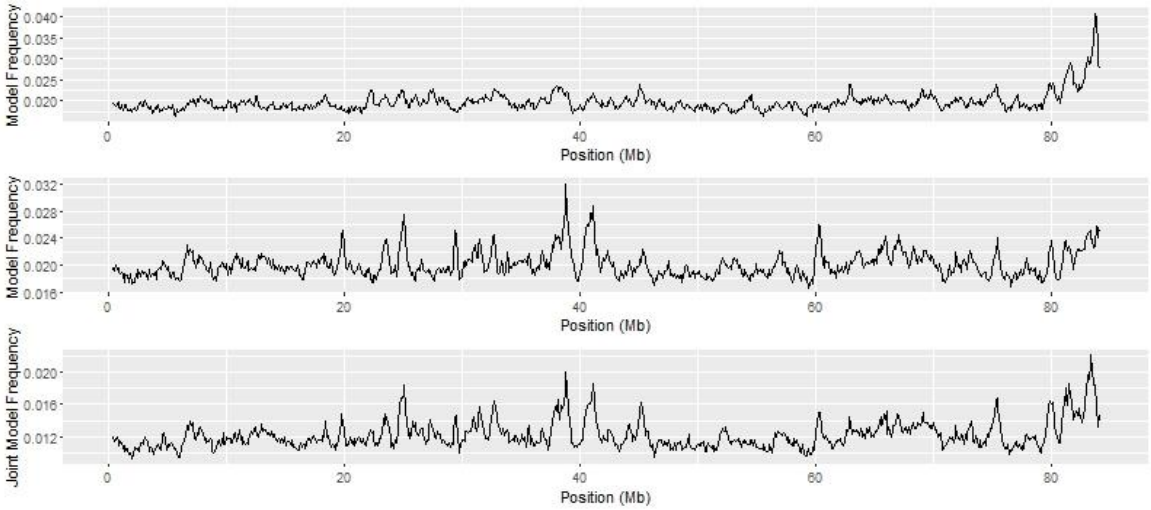

Figure 29. Comparison of univariate, average daily gain (ADG; top) and average daily feed intake (ADFI; middle), and bivariate (bottom) haplotype associations for BTA 13 from BayesIM 8 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each haplotype loci or the joint model frequency when both traits are in the model. The X-axis is position in megabases (Mb) on BTA 13.

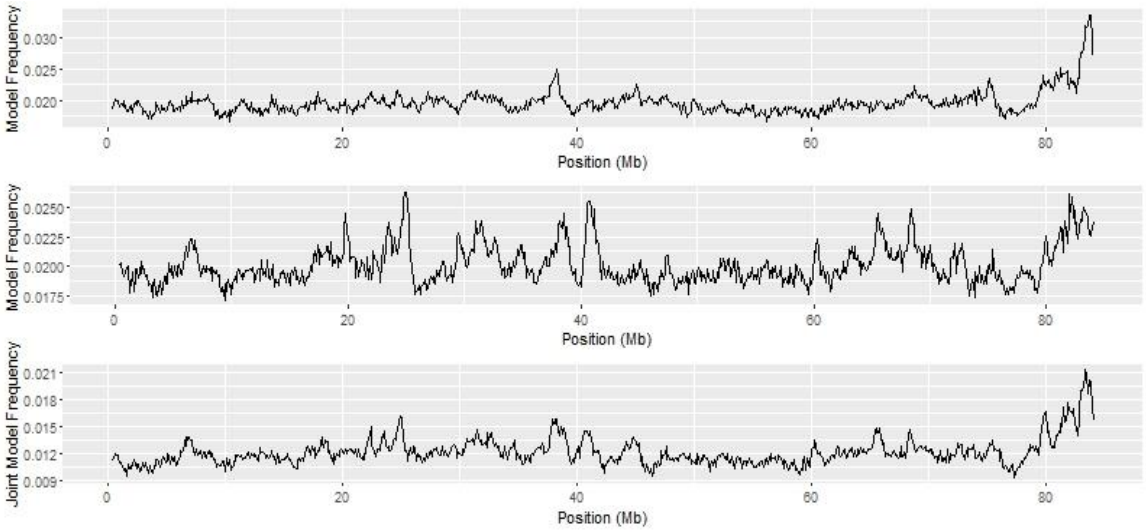

Figure 30. Comparison of univariate, average daily gain (ADG; top) and average daily feed intake (ADFI; middle), and bivariate (bottom) haplotype associations for BTA 13 from BayesIM 16 clusters with when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each haplotype loci or the joint model frequency when both traits are in the model. The X-axis is position in megabases (Mb) on BTA 13.

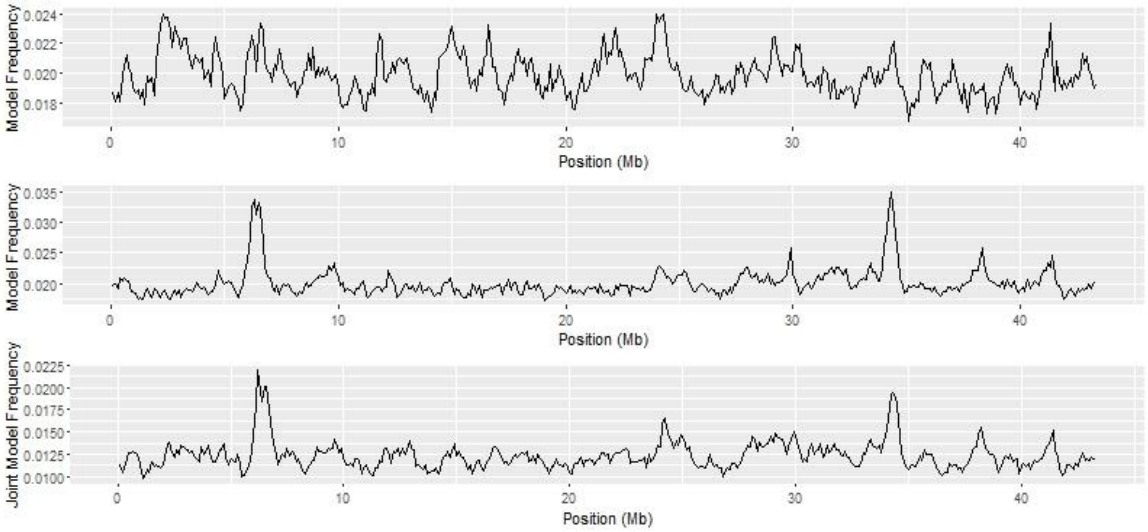

Figure 31. Comparison of univariate, average daily gain (ADG; top) and average daily feed intake (ADFI; middle), and bivariate (bottom) haplotype associations for BTA 25 from BayesIM 8 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each haplotype loci or the joint model frequency when both traits are in the model. The X-axis is position in megabases (Mb) on BTA 25.

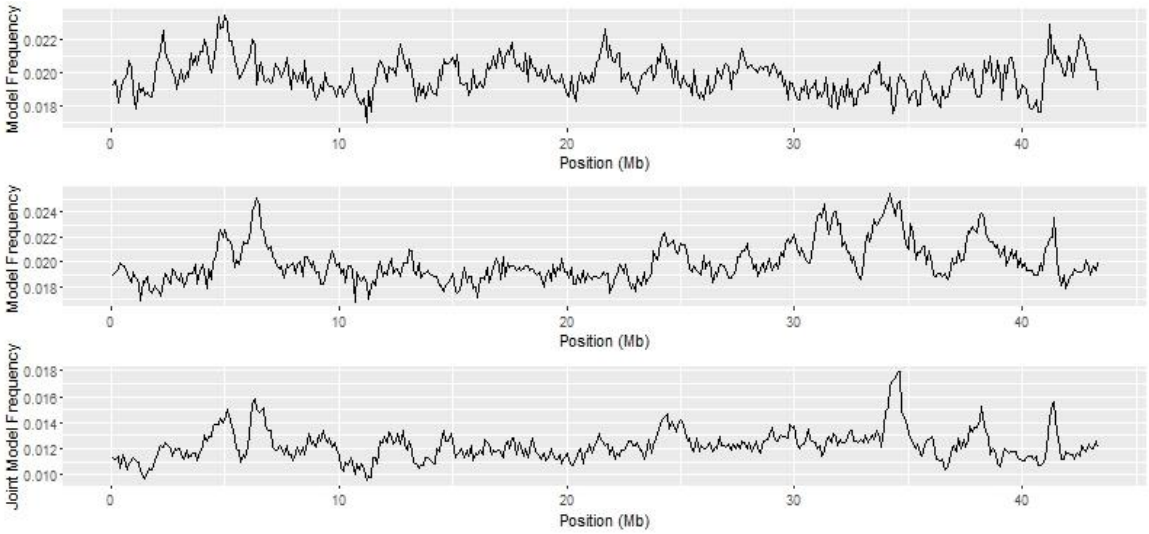

Figure 32. Comparison of univariate, average daily gain (ADG; top) and average daily feed intake (ADFI; middle), and bivariate (bottom) haplotype associations for BTA 25 from BayesIM 16 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each haplotype loci or the joint model frequency when both traits are in the model. The X-axis is position in megabases (Mb) on BTA 25.

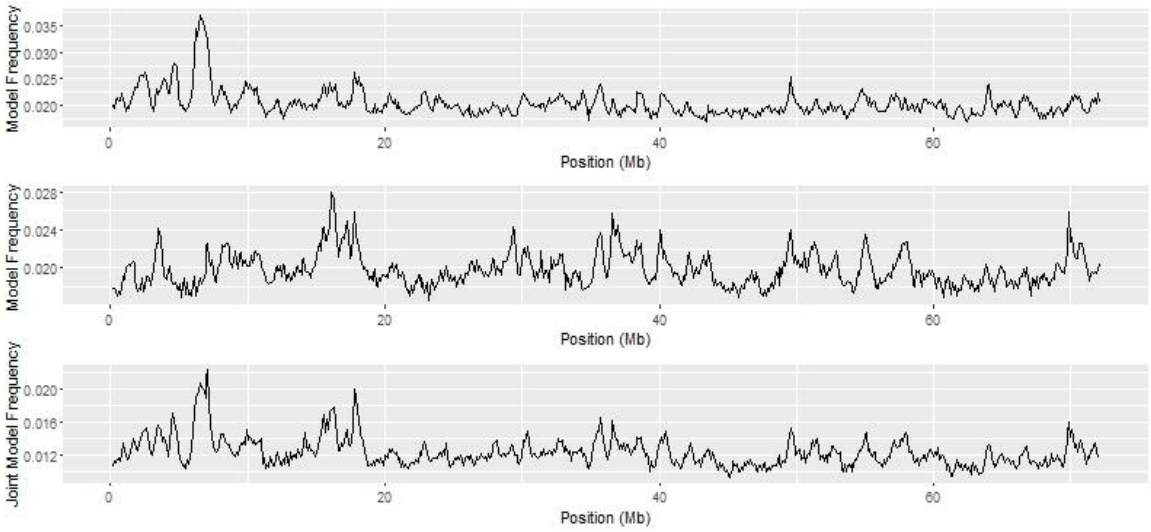

Figure 33. Comparison of univariate, average daily gain (ADG; top) and average daily feed intake (ADFI; middle), and bivariate (bottom) haplotype associations for BTA 20 from BayesIM 8 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each haplotype loci or the joint model frequency when both traits are in the model. The X-axis is position in megabases (Mb) on BTA 20.

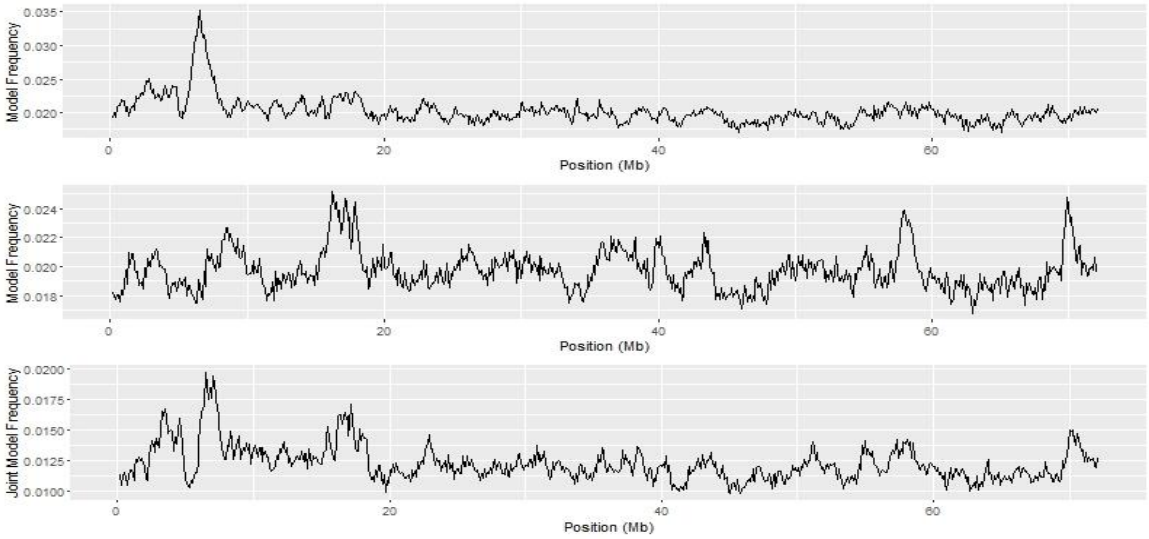

Figure 34. Comparison of univariate, average daily gain (ADG; top) and average daily feed intake (ADFI; middle), and bivariate (bottom) haplotype associations for BTA 20 from BayesIM 16 clusters when sire breed was included as a fixed effect in the model (SB). The Y-axis represents the model frequency of each haplotype loci or the joint model frequency when both traits are in the model. The X-axis is position in megabases (Mb) on BTA 20.