Conference on Applied Statistics in Agriculture 2016 - 28th Annual Conference Proceedings

A BAYESIAN GWAS METHOD UTILIZING HAPLOTYPE CLUSTERS FOR A COMPOSITE BREED POPULATION

Danielle F. Wilson-Wells
University of Nebraska-Lincoln, twinonedil@centurytel.net

Stephen D. Kachman
University of Nebraska-Lincoln

Recommended Citation

Wilson-Wells, Danielle F. and Kachman, Stephen D. (2016). "A BAYESIAN GWAS METHOD UTILIZING HAPLOTYPE CLUSTERS FOR A COMPOSITE BREED POPULATION," *Conference on Applied Statistics in Agriculture*. https://doi.org/10.4148/2475-7772.1469

A BAYESIAN GWAS METHOD UTILIZING HAPLOTYPE CLUSTERS FOR A COMPOSITE BREED POPULATION

Danielle F. Wilson-Wells and Stephen D. Kachman

Department of Statistics, University of Nebraska-Lincoln

Abstract

Commercial beef cattle are often composites of multiple breeds. Current methods used to produce genomic predictors are based on the underlying assumption of animals being sampled from a homogeneous population. As a result, the predictors can perform poorly when used to predict the relative genetic merit of animals whose breed composition are different. In part, this is due to the changes in linkage disequilibrium between the markers and the quantitative trait loci as we move from one breed to the next. An alternative model based on breed specific haplotype clusters was developed to allow for differences in linkage disequilibrium across multiple breeds. The haplotype clusters were modeled as hidden states in a hidden Markov model where the genomic effects are associated with loci located on the unobserved clusters. Similar to the Bayes C model, we can model the genomic effects at the loci using a prior, which consists of a mixture of a multivariate normal and a point mass at zero distribution. The model will be used to construct genomic predictors using records on 6,552 cattle genotyped for 99,827 mapped SNPs representing various fractions of three different breeds.

1 Introduction

Current genome-wide association study (GWAS) methods focus on improving predictability within purebred population by utilizing the assumption that animals are being sampled from an homogeneous population in order to produce genomic predictors. This assumption makes these methods difficult to implement in an commercial setting since commercial beef cattle are often composites of multiple breeds [Rolf et al., 2015]. When the current GWAS methods are used it results in predictors which may perform poorly when used to predict the relative genetic merit of animals whose breed composition differs from the training population. One explanation for this poor performance is the changes in linkage disequilibrium between the markers and the quantitative trait loci depending on which breed the chromosomal segment originated in.

There have been few studies which analyzed the effect on prediction accuracy when a composite breed population was used. Bolormaa et al. [2013] showed that, when you have a composite breed population, prediction accuracy is improved when you include animals composite and pure bred individuals together in the training set versus only including breed specific individuals. Studies using pure breed populations can also give us some insight. Hayes et al. [2009]and Kachman et al. [2013] used pure breed populations to demonstrate that, in most cases, prediction accuracy is improved when the breed with which you wish to make evaluations on is included in the training set. Muijibi et al. [2011] showed that prediction accuracy was improved when only one breed was used for both training and evaluation rather than using all the breeds together for training and evaluation.

Current Bayesian genome wide association studies (GWAS) methods utilize single nucleotide polymorphisms (SNPs) to find potential quantitative trait loci (QTL). A QTL is a location on a chromosome where variation in the genotype correlates with the variation observed within a phenotype. We use linkage disequilibrium between SNPs, which are single base changes, and a QTL in order to identify QTLs. Since a true QTL may be located at a position which we have not genotyped, identification of a QTL is dependent on the location of the SNPs and their association with the QTL. GWAS methods identify QTLs using the pattern of SNPs around the putative QTL alleles.

Kachman [2015] proposed a method which uses the information of SNPs close to the putative QTL. This method is called Bayes IM. Bayes IM uses the genotypes from the SNP data and adds putative QTL along the chromosome which can be placed at a SNP location or between SNP locations. Using a set of haplotype clusters, Bayes IM estimates the haplotype effect at each potential QTL. However, like other Bayesian GWAS methods, Bayes IM fails to account

for the breed composition of the individual. The proposed solution is to adapt Bayes IM to include breed specific haplotype clusters rather than using a common set of haplotype clusters. This adapted model will be called Bayes IM Comp.

To estimate the haplotype effects we need to consider which haplotypes make up an individuals genotype. Each individuals genotype is made up of two haplotypes, one received from the sire and one from the dam. These haplotypes themselves come from a population of haplotypes where some are more common than others and some may be very rare. We could treat each possible haplotype as unique, but unless each segment is short there will be many unique haplotypes. Also, some possible haplotypes may not be observed within our sample. We are going to cluster the haplotypes together based on similarity since this requires fewer covariates and the novel haplotypes will be placed in clusters based on similarity. To incorporate haplotypes into a GWAS, partition the genome into segments, identify the haplotypes within these segments, form haplotype clusters, and use the identified haplotype clusters as covariates.

Next, we shift the segment down by one SNP locus and, due to crossing, we expect to be able to identify recombination locations as we move down the chromosome. Finally, we can extend the clusters across each chromosome. The paternal and maternal haplotypes will be broken up into segments with each segment belonging to a cluster. For a particular segment, the haplotype being sampled is based on the frequency of the A allele at a locus. The probability that a segment transitions to a different cluster between two loci is based on a function of the distance between those two loci.

2 Materials and Methods

Data Description

Population

Genotypes on 6,552 Simmental and Simmental composite cattle from the American Simmental Association were used. The genotypes consisted of a total of 99,827 mapped autosomal SNPs from two different genotyping platforms, in which 27,562 were common between the two platforms. The average breed composition of the genotyped individual were 63% Simmental, 30% Angus, and 2% Hereford. With the remaining 29 breed together accounting for the other 5%. Thus we considered the percentage of each individual which came from a Simmental, Angus, Hereford, and combined breed background. Expected progeny differences (EPDs) for the following traits were deregressed to account for variable accuracies [Garrick et al., 2009]. Traits include body weights for birth weight (BWT), maternal weaning weight (MILK), mature weight (MWWT), weaning weight (WWT), and yearling weight (YWT). In addition carcass traits were collected including back fat (BFAT), carcass weight (CWT), docility (DOC), marbling score (MARB), ribeye muscle area (REA), and yield grade (YG). Calving ease (CE) and maternal calving ease (MCE) were two other traits collected.

All individuals whom had a SNP genotype and breed composition were included in analysis. Table 1 shows the number of individuals remaining after removing animals missing EPDs for a particular trait. Training and evaluation sets were created at random with 2/3 of individuals being used in training and 1/3 of individuals used for evaluation. For each trait 3 models were considered, which will be described in more detail below:

1. Bayes IM 8: This is the Bayes IM method with a total of 8 haplotype clusters.
2. Bayes IM 16: This is the Bayes IM method with a total of 16 haplotype clusters.
3. Bayes IM Comp: This is the Bayes IM Comp method with 8 haplotype clusters assigned to the Simmental breed, 4 assigned to Angus, 2 to Hereford, and 2 to the combined breed group with a total of 16 haplotype clusters all together.

Models

Bayes IM

Bayes IM begins by constructing a hidden Markov Model (HMM) and then uses the EM-algorithm to estimate the parameter values. Bayes IM uses a fixed number of haplotype clusters, N, which we will denote by $S = \{S_1, S_2, ..., S_N\}$. We use the genotypic data on n markers and P individuals, which for individual i, is denoted by $X_i = \{x_{i1}, x_{i2}, ..., x_{in}\}$,

Table 1: Number of individuals used in the training and evaluation sets per trait

Trait	Training	Evaluation	Total
BFAT	1652	846	2498
BWT	2516	1291	3807
CE	2481	1264	3745
CWT	2487	1235	3722
DOC	1436	727	2163
MARB	1633	805	2438
MCE	2436	1151	3587
MILK	2314	1145	3459
MWWT	2409	1208	3617
REA	1588	790	2378
WWT	2488	1206	3694
YG	1712	829	2541
YWT	2493	1212	3705

where x_{ik} is the genotype for individual i at position k. The genotype covariate is equal to the number of copies of the A allele x_{ik} can have one of four possible values, $O = \{0,1,2,.\}$, where $\{.\}$ represents a missing value. Since an individuals genotype at a particular position is made up of two haplotypes, we will let π_{ik} denote the unordered pair of haplotype cluster that generated genotype x_{ik}. Then the haplotype path for individual i can be denoted as $\pi_i = \{\pi_{i1}, \pi_{i2}, ..., \pi_{in}\}$. The initial state probabilities, transition probabilities, and emission probabilities will now be defined. We do this using a model and notation similar to that developed by Scheet and Stephens [2006] for use in fastPHASE.

Using the assumption that each haplotype cluster is equally likely to occur, we can define probability that we are in haplotype cluster S_l at position k as:

$$\alpha^k_{S_l} = \frac{1}{N}$$

given there are N total haplotype clusters. We can then define the probability that a chromosome transitions from cluster S_l at position $k-1$ to cluster S_m at position k as:

$$a^k_{S_l S_m} = \begin{cases} e^{-\frac{d_k}{\lambda}} + \left(1 - e^{-\frac{d_k}{\lambda}}\right)\left(\frac{1}{N}\right) & S_l = S_m \\ \left(1 - e^{-\frac{d_k}{\lambda}}\right)\left(\frac{1}{N}\right) & S_l \neq S_m \end{cases},$$

[Scheet and Stephens, 2006] where d_k is the physical distance between markers $k-1$ and k and λ is a parameter which needs to be estimated. The emission probability, $e_{\pi_{ik}}(x_{ik})$, is the probability that animal i at position k has observed genotype x_{ik} given we are in cluster pair $\pi_{ik} = \{S_l, S_m\}$. The emission probability is written:

$$e_{\pi_{ik}}(x_{ik}) = \begin{cases} \left(1 - \theta_{S_l k}\right)\left(1 - \theta_{S_m k}\right) & x_{ik} = 0 \\ \theta_{S_l k}\left(1 - \theta_{S_m k}\right) + \theta_{S_m k}\left(1 - \theta_{S_l k}\right) & x_{ik} = 1 \\ \theta_{S_l k}\theta_{S_m k} & x_{ik} = 2 \end{cases},$$

[Scheet and Stephens, 2006] where $\theta_{S_l k}$ is the frequency of allele A in cluster S_l at position k.

Once complete, Bayes IM uses the parameter values obtained from the HMM to:

1. Sample the haplotypes for each individual
2. Sample for an effect at each QTL
3. Sample the individual haplotype cluster effects at each QTL
4. Sample the variance components.

Similar to the Bayes C model [Meuwissen and Goddard, 2004], a QTL at locus k has an effect of 0 with probability π and an effect that is drawn from a normal distribution with a mean of 0 and a variance of σ_α^2 with probability $1-\pi$. The residuals, $\mathbf{e}$, have a multivariate normal prior with a mean of $\mathbf{0}$ and a covariance matrix of $\mathbf{R}\sigma_e^2$, where $\mathbf{R}$ is a diagonal matrix. Each variance component, σ_α^2 and σ_e^2, have a scaled-inverse chi-square prior. Unlike Bayes C model, the haplotypes themselves are unknown and must be sampled for each individual and, in addition, the haplotype effects for each cluster are sampled from a normal with a mean of 0 and a variance of σ_b^2 Kachman [2015].

Bayes IM Comp

Now, we want to consider a population of individuals made up of B breeds. We assume that each breed, b, is make up of N_b haplotype clusters. Then the overall number of clusters is equal to $\sum_b N_b = N$. Let $C_i = \{C_{i1}, C_{i2}, ..., C_{iB}\}$ be the breed composition of individual i, where C_{ib} is the proportion of individual i which is from breed b. The probability that, at position k, we are in haplotype cluster S_{l_b} given we are in a haplotype cluster from breed b is:

$$\alpha^k_{S_{l_b}} = \frac{1}{N_b}.$$

This assumes that within a particular breed each haplotype is equally likely.

Next we can further weight these probabilities by an individuals breed compositions. Thus the probability that, at position k, individual i is in haplotype cluster S_{l_b} is

$$C_{ib} \cdot \alpha^k_{S_{l_b}} = C_{ib} \cdot \frac{1}{N_b}.$$

The probability that individual i transitions on a chromosome from cluster S_{l_b} in breed b at position $k-1$ to cluster S_{m_f} in breed f at position k is defined by

$$a^k_{S_{l_b}S_{m_f}} = \begin{cases} e^{-\frac{d_k}{\lambda}} + \left(1 - e^{-\frac{d_k}{\lambda}}\right)\left(C_{if} \cdot \frac{1}{N_f}\right) & S_{l_b} = S_{m_f} \\ \left(1 - e^{-\frac{d_k}{\lambda}}\right)\left(C_{if} \cdot \frac{1}{N_f}\right) & S_{l_b} \neq S_{m_f} \end{cases}$$

where, as above, d_k is the physical distance between markers $k-1$ and k and λ is a parameter which needs to be estimated. All other pieces match with the Bayes IM model described above.

3 Results and Discussions

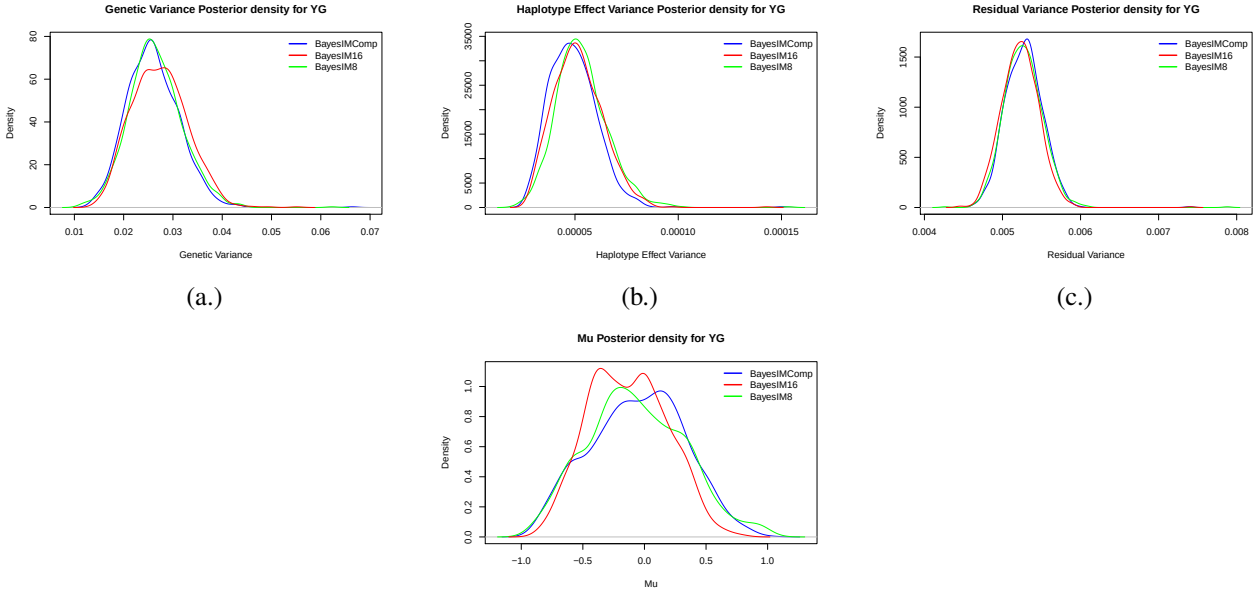
The performance of the three models will be compared by first training each of the models using the training data and then evaluating their performance in the evaluation data set. We will:

1. Examine the posterior distribution plots for the mean and the variance components
2. Examine the estimated QTL effects using Metropolis plots
3. Examine the plots for the haplotype effects
4. Compare the estimated genetic correlations for each model in order to access accuracy.

Since many of the same trends were observed among the traits, we have chosen to only look at a few of the traits in detail for the first three points list above.

For all 13 traits, the genetic variance, haplotype effect variance, residual variance, and mean plots, shown in Figure 1, showed very little differences between the three models. To confirm that the models has converged the trace plots

Figure 1: Comparison of Posterior Distributions of Variances and Mean for Yield Grade



(a.) (b.) (c.)

(d.)

were examined for all the parameters, which can be found in Appendix A. The trace plots for the genetic variance, haplotype effect variance, residual variance, and mean are consistent with chains that were mixing well and had converged to their stabilizing distribution. Table 2 gives a numerical summary of the mean and standard deviation for the posterior distributions of the parameters. For yield grade, the mean and standard deviation for all parameters appear to match fairly well.

Table 2: Comparison of the Mean and Standard Deviation of the Posterior Distribution for Yield Grade

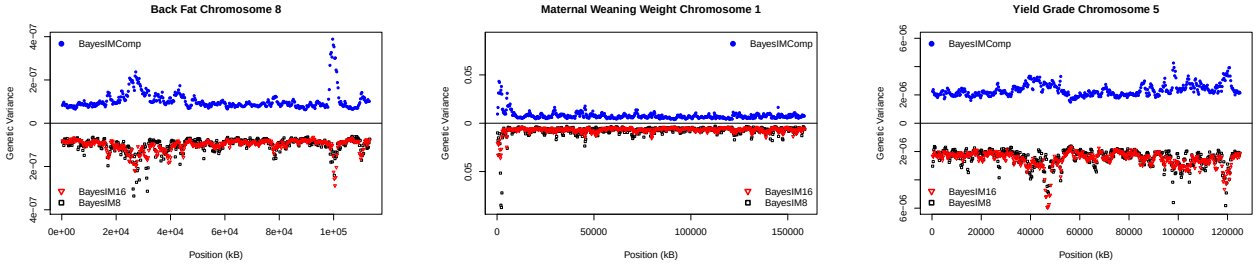
Model	Genetic Variance	Haplotype Effect Variance	Residual Variance	Mean
BayesIMComp	0.026 (0.0053)	4.91e-05 (1.14e-05)	0.0053 (0.00024)	-0.214 (0.010)
BayesIM8	0.026 (0.0054)	5.33e-05 (1.26e-05)	0.0053 (0.00025)	-0.214 (0.010)
BayesIM16	0.027 (0.0056)	5.17e-05 (1.20e-05)	0.0052 (0.00024)	-0.213 (0.010)

Note: Mean(SD)

Next, we examined the metropolis plots for genetic variation in order to examine the differences in the putative QTL effects size and location between the three models. Figure 2 shows the plots for three traits; Back fat on chromosome 8, maternal weaning weight on chromosome 1, and yield grade on chromosome 5. Figure 2(a.) shows that around 100,000 kilobases on chromosome 8 Bayes IM Comp is showing the QTL for back fat that has a larger effect than either Bayes IM 8 or Bayes IM 16. In addition, Bayes IM 16 has a larger peak than Bayes IM 8. Figure 2(b.) shows that around 2,000 kilobases on chromosome 1 Bayes IM 8 has a QTL for maternal weaning weight with a sharp peak that is not seen in either Bayes IM Comp or Bayes IM 16. Here Bayes IM Comp has an effect which is slightly larger than the effect seen in Bayes IM 16. In figure 2(c.) between 40,000 and 50,000 kilobases on chromosome 5, we see a QTL for yield grade with sharp peak from Bayes IM 16, but lower peaks with more variance in the Bayes IM Comp and Bayes IM 8 models. However, the peak observed within Bayes IM 8 is higher than the peak for Bayes IM Comp.

We can now examine the haplotype effect estimates for each haplotype in each model at the locations described above. Figure 3 examines the haplotype estimates for back fat between 96,000 and 104,000 kilobases on chromosome

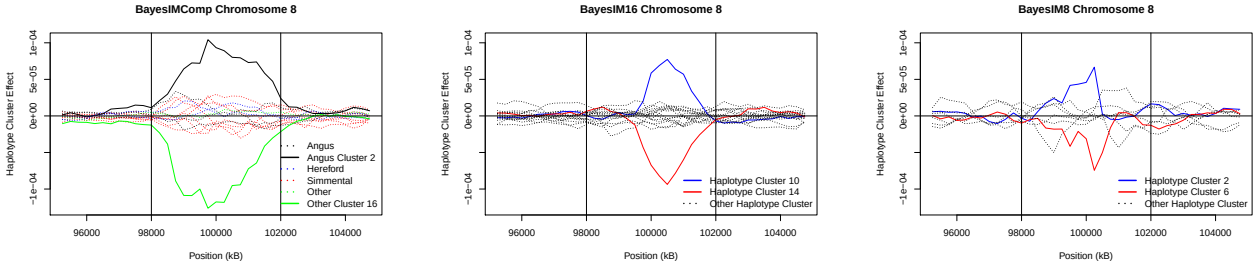
Figure 2: Comparison of Genetic Variance for 3 Traits



(a.) (b.) (c.)

a. Genetic variances for Bayes IM 8 and Bayes IM 16 are plotted as deviations below zero.

Figure 3: Haplotype Estimates for Back Fat on Chromosome 8 for each model



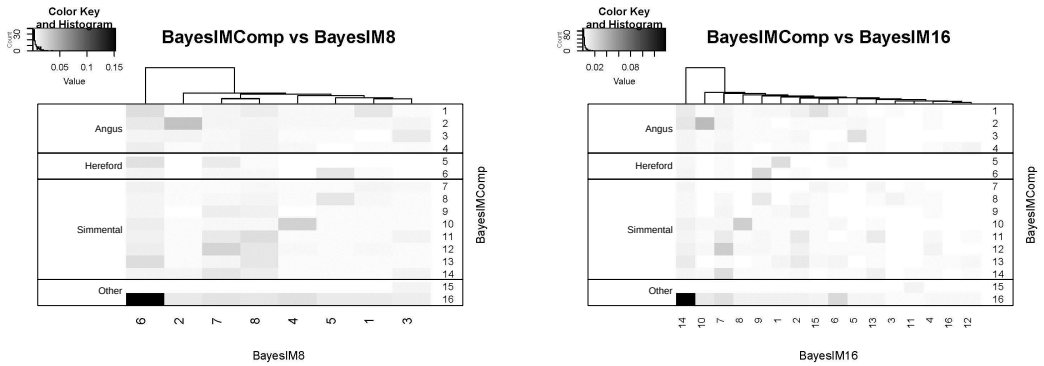
(a.) (b.) (c.)

8. Bayes IM Comp has a large positive effect on Angus haplotype cluster 2 and a large negative effect on other haplotype cluster 16. The remaining haplotypes had very little effect on back fat. The large positive Angus cluster and large negative other breed group cluster have effects that are much larger than any of the clusters within Bayes IM 16 or Bayes IM 8. The large haplotype cluster effects in Figure 3(a.) accounts for the large peak in the genetic variance, seen in Figure 2(a.), within Bayes IM Comp that was not present in the other two models. This is because the genetic variance at a locus is a function of the haplotype effects and the frequencies of the haplotype genotypes. When we see large haplotype effects at a particular locus, we will also observe a large genetic variance at that same locus. Bayes IM 16 has one haplotype cluster with a large positive effect and one with a large negative effect; however the effects of those haplotypes are smaller than what was seen within the Bayes IM Comp model. Bayes IM 8 is more dispersed. It is showing one haplotype cluster with a positive effect and one haplotype cluster with a negative effect around the 100,000 kilobases but the effects are much smaller and are not smooth.

To investigate the relationship between the haplotype clusters from the different models, we select a location, around the genetic variance peak on chromosome 8 between 98,000 and 102,000 kilobases, and calculate the probability of being assigned jointly to a particular haplotype cluster in one model and a particular haplotype cluster in another model. Figure 4(a) is a heat map of the probability of being assigned to a particular haplotype cluster within Bayes IM 8 and a particular haplotype cluster within Bayes IM Comp and Figure 4(b) is a heat map of the probability of being assigned to a particular haplotype cluster within Bayes IM 16 and a particular haplotype cluster within Bayes IM Comp. It is important to note that for Bayes IM Comp haplotype clusters 1 through 4 belong to the Angus breed group, 5 and 6 to Hereford, 7 through 14 to Simmental, and 15 and 16 to the combined breed group.

Figure 4(a) shows that combined breed cluster 16 in the Bayes IM Comp model is mapping mainly to haplotype cluster 6 within the Bayes IM 8 model. Haplotype cluster 6 in the Bayes IM 8 model has a muted effect due to Angus cluster 1, Hereford cluster 5, and Simmental cluster 13, which all have a small effect size, also mapping the haplotype cluster 6 in Bayes IM 8. In addition, Angus cluster 2 within the Bayes IM Comp model maps primarily to haplotype

Figure 4: Probability of being jointly assigned to a particular haplotype cluster in Bayes IM Comp and a particular haplotype cluster in Bayes IM 8 or Bayes IM 16 on chromosome 8 between 96,000 and 104,000 kilobases



(a.) (b.)

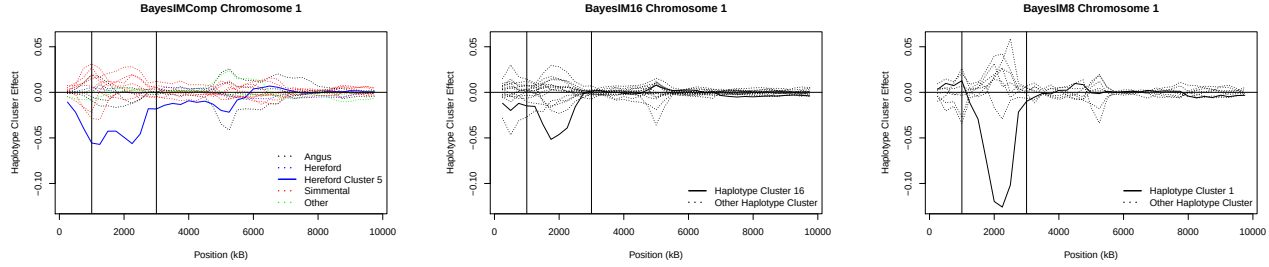
Note: The darker the color the higher the probability of being jointly assigned to a particulate haplotype cluster in Bayes IM Comp and a particular haplotype cluster in Bayes IM 8 or Bayes IM 16
b. Histogram shows the distribution of the probabilities within the heatmap
c. Dendogram was formed using hierarchical clustering based on complete linkage and euclidean distance

cluster 2 within the Bayes IM 8 model. However, haplotype clusters from the Angus, Simmental and other breed groups within Bayes IM Comp are also mapping to haplotype cluster 2 within Bayes IM 8. Thus we have haplotype clusters from Bayes IM Comp with both large and small effects that are being mapped to a single cluster within the Bayes IM 8 model. This is consistent with the muted effects seen in the haplotype effects of the Bayes IM 8 model.

The effect is much more obvious for the Bayes IM 16 model than it was for the Bayes IM 8 model. Figure 4(b) shows that combined breed cluster 16 maps to haplotype cluster 16 within the Bayes IM 16 model. The overall effect, however, is being muted by the presence of haplotype clusters, with small effect sizes or effect sizes in the opposite direction, from Angus, Hereford, and Simmental haplotype clusters within Bayes IM Comp. Additionally, Angus cluster 2 within Bayes IM Comp is mapping to haplotype cluster 10 within the Bayes IM 16 model. The larger haplotype cluster effects for Bayes IM 16 compared to Bayes IM 8 can be explained by the lower probabilities of the small effect clusters within Bayes IM Comp mapping to the Bayes IM 16 haplotype clusters. However, the effects are still smaller than that observed within Bayes IM Comp due to the clusters with small effect sizes within the Bayes IM Comp model occasionally being mapped to haplotype clusters within the Bayes IM 16 model.

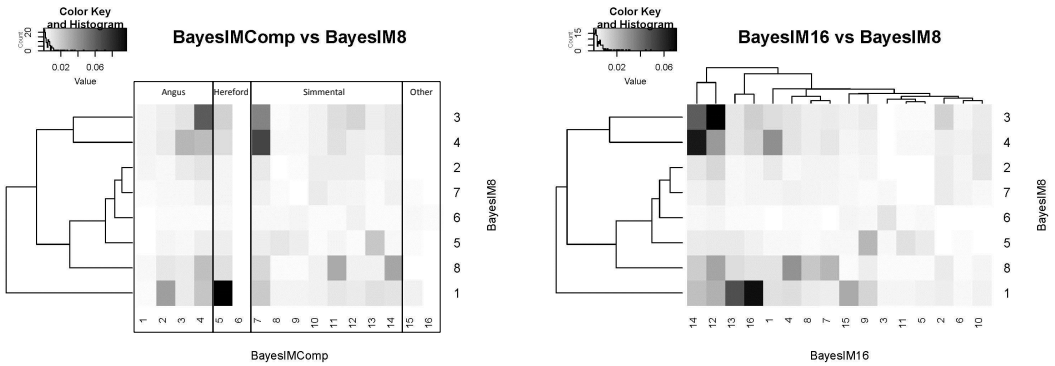
Figure 5 shows the haplotype estimates for maternal weaning weight on chromosome 1 between 0 and 10,000 kilo bases. Bayes IM 8 is showing a large negative effect from one of the haplotypes and smaller positive effects from most of the other haplotypes. Bayes IM 16 and Bayes IM Comp show a much smaller negative effect. We should also note

Figure 5: Haplotype Estimates for Maternal Weaning Weight on Chromosome 1 for each model



(a.) (b.) (c.)

Figure 6: Probability of being jointly assigned to a particular haplotype cluster in Bayes IM 8 and a particular haplotype cluster in Bayes IM Comp or Bayes IM 16 on chromosome 1 between 1,000 and 3,000 kilobases



(a.) (b.)

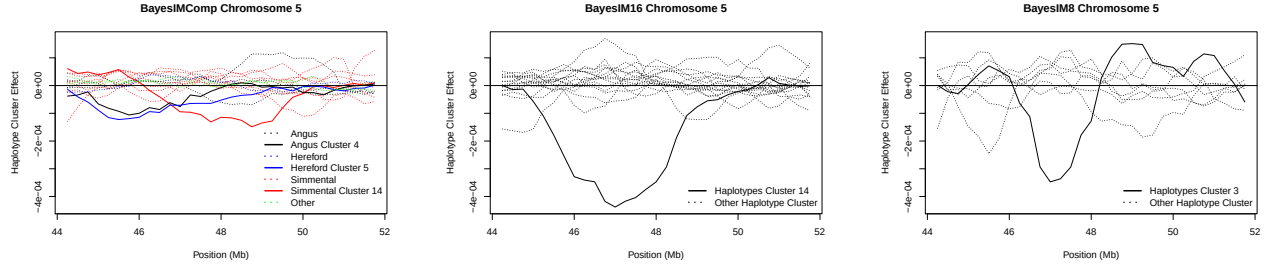
a. The darker the color the higher the probability of being jointly assigned to a particulate haplotype cluster in Bayes IM 8 and a particulate haplotype cluster in Bayes IM Comp or Bayes IM 16
b. Histogram shows the distribution of the probabilities within the heatmap
c. Dendogram was formed using hierarchical clustering based on complete linkage and euclidean distance

that the negative haplotype in Bayes IM Comp has been attributed to the Hereford breed which is only represented by 2% of the data. The haplotype effects seen here explain the large peak, seen in Figure 2(b), within Bayes IM 8 that was not present in the other two models. Since Bayes IM 8 is using a few number of non-breed specific haplotypes it is possible this effect is amplified when haplotypes from other breeds are combined into a common haplotype cluster.

Figure 6(a) and (b) are heat maps of the probability of being in a particular haplotype cluster within the Bayes IM Comp and Bayes IM 16 models, respectively, and a particular haplotype cluster within the Bayes IM 8 model. Figure 6(a) shows that haplotype cluster 1 within Bayes IM 8 has the strongest map to Hereford cluster 5 in the Bayes IM 8 model, but haplotype cluster 1 is also mapping to Angus clusters 1 and 4 and Simmental cluster 7. This is spreading the large effect of cluster 1 from Bayes IM 8 among several clusters within Bayes IM Comp causing the lower effect size seen in the Bayes IM Comp model. Figure 6(b) shows that haplotype cluster 1 in the Bayes IM 8 model has the strongest map to haplotype cluster 16 in the Bayes IM 16 model, but is also mapping to Bayes IM 16 haplotype clusters 12, 13, 14, and 15. Since the effect of cluster 1 from Bayes IM 8 is being spread among several haplotype clusters within Bayes IM 16 the overall effect size is much smaller.

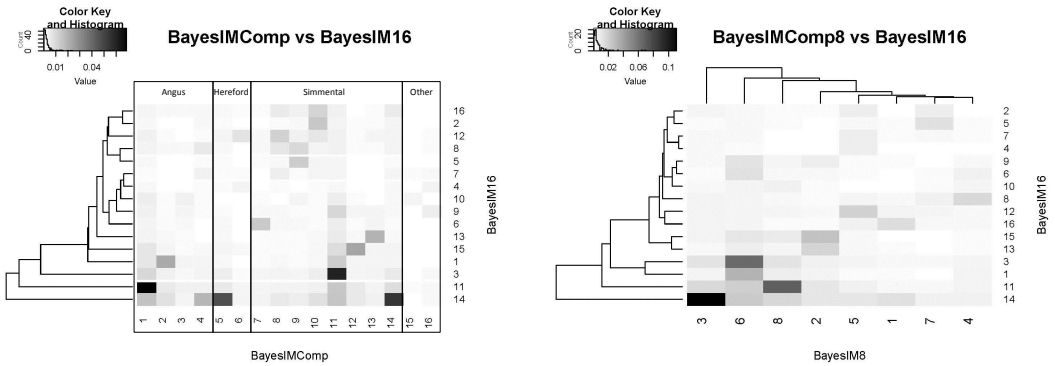
Figure 7 shows the haplotype estimates for yield grade on chromosome 5 between 46,000 and 48,000 kilo bases. Bayes IM 16 is showing a large negative effect from one of the haplotypes and Bayes IM 8 is showing a smaller negative effect from 1 haplotypes in the same region. Bayes IM Comp, however, shows no significant positive or

Figure 7: Haplotype Estimates for Yearling Growth on Chromosome 5 for each model



(a.) (b.) (c.)

Figure 8: Probability of being jointly assigned to a particular haplotype cluster in Bayes IM 16 and a particular haplotype cluster in Bayes IM Comp or Bayes IM 8 on chromosome 5 between 46,000 and 48,000 kilobases



(a.) (b.)

a. The darker the color the higher the probability of being jointly assigned to a particulate haplotype cluster in Bayes IM 16 and a particulate haplotype cluster in Bayes IM Comp or Bayes IM 8
b. Histogram shows the distribution of the probabilities within the heatmap
c. Dendogram was formed using hierarchical clustering based on complete linkage and euclidean distance

negative haplotype effects. Once again this explains the large peak, seen in Figure 2(c), within Bayes IM 16 that was not as obvious in the Bayes IM Comp.

Figure 8 shows the heat maps for (a) the probability of being in a particular haplotype cluster within Bayes IM Comp and a particular haplotype cluster within Bayes IM 16 and (b) the probability of being in a particular haplotype cluster within Bayes IM 8 and a particular haplotype cluster within Bayes IM 16 for chromosome 5 between 46,000 and 48,000 kilobases. Figure 8(a) shows that haplotype cluster 14 within the Bayes IM 16 model is being mapped to Angus cluster 4, Hereford cluster 5, and Simmental cluster 14. This divides the effect cluster 14 from Bayes IM 16 among several clusters within Bayes IM Comp and causes there to be a muted effect within the Bayes IM Comp model. Figure 8(b) shows that haplotype cluster 14 within the Bayes IM 16 model maps to cluster 3 within the Bayes IM 8 model with a small mapping to cluster 6 within Bayes IM 8. This is consistent with the large effect seen within the Bayes IM 8 model which is slightly smaller than the one seen within the Bayes IM 16 model since a small percentage of cluster 14 from Bayes IM 16 is being taken away from cluster 3 within Bayes IM 8 and given to cluster 6 within Bayes IM 8.

Estimated genetic correlation and standard errors between the estimated breeding value and the actual breeding value were calculated using ASReml [Gilmour et al., 2014]. Table 3 displays these results. Within the table the model with the highest genetic correlation is in bold. By examining the values in Table 3, we can conclude there was no best model. Bayes IM Comp was higher about 50% of the time and one of the Bayes IM models were higher the other 50% of the time. No model was ever more or less than one standard error away from any other model. Thus, estimation wise, we did just as well with 8 haplotypes as we did with 16 and we also did just as well accounting for breed composition as we did not accounting for breed composition.

4 Conclusions and Discussions

There are obvious differences between Bayes IM 8, Bayes IM 16, and Bayes IM Comp in the size of the putative QTLs. There were also differences in the individual haplotype cluster effect estimates. It is apparent that at times a haplotype cluster which is common among all breed groups is necessary as was seen with yield grade. It is also apparent that having a common haplotype cluster among all breeds hindered our ability to identify QTL as was the case with back fat. A solution may be to add a common haplotype cluster among all the breed groups and see if Bayes IM Comp is able to distinguish the difference between theses two situations.

However, when it comes to the overall estimation of an animals breeding value all models appear to preform

Table 3: Estimated genetic correlations and standard errors

Trait	BayesIMComp	BayesIM8	BayesIM16
BFAT	0.367 (0.051)	0.363 (0.052)	**0.368 (0.051)**
BWT	0.610 (0.032)	0.610 (0.032)	**0.613 (0.032)**
CE	**0.786 (0.029)**	0.776 (0.030)	0.783 (0.029)
CWT	0.569 (0.042)	**0.588 (0.042)**	0.576 (0.042)
DOC	0.430 (0.059)	0.414 (0.060)	**0.432 (0.059)**
MARB	**0.622 (0.060)**	0.613 (0.061)	0.612 (0.061)
MCE	**0.607 (0.050)**	0.579 (0.051)	0.595 (0.050)
MILK	**0.410 (0.056)**	0.406 (0.056)	0.393 (0.056)
MWWT	0.510 (0.048)	0.522 (0.048)	**0.527 (0.048)**
REA	**0.442 (0.073)**	0.440 (0.073)	0.438 (0.073)
WWT	0.491 (0.042)	**0.499 (0.042)**	0.479 (0.042)
YG	**0.481 (0.063)**	0.479 (0.064)	0.478 (0.064)
YWT	**0.534 (0.040)**	0.533 (0.041)	0.518 (0.041)

Note: Bold values represent the model with the highest genetic correlation for each trait

equally well. A major advantage to accounting for breed is that we were able to see which breeds had a positive effect QTL and which had a negative effect, as was the case with back fat. We also have an advantage in being able to identify locations were combining breed haplotypes results in the false identification of a QTL as we saw with yield grade.

There are several additional changes which may improve the performance of Bayes IM Comp. Bayes IM Comp uses a single λ parameter. Thus we are equally likely to jump to a haplotype cluster in another breed as we are to jump to a haplotype cluster within the same breed. As transitions between clusters within the same breed versus transition between clusters across different breed may represent recombination events on different time scales, having λ vary between and across breeds may be reasonable. In addition, Kachman [2015] has shown that in pure breed population Bayes IM performs as well as, if not slightly better than, the current Bayesian GWAS methods such as Bayes B and Bayes C. We still need to compare the performance of Bayes IM Comp and Bayes IM versus the Bayesian GWAS methods on a population of composite breed individuals.

References

S. Bolormaa, 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 Goddard M.E. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in Bos taurus, Bos indicus, and composite beef cattle. *Journal of Animal Science*, 91(7):3088–3104, 2013.

Dorian J Garrick, Jeremy F Taylor, and Rohan L Fernando. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution*, 41(1):1, 2009.

A.R. Gilmour, B.J. Gogel, B.R. Cullis, S.J. Welham, R. Thompson, D. Butler, M. Cherry, D. Collins, G. Dutkowski,

S.A. Harding, et al. Asreml user guide. release 4.1 structural specification. *VSN International Ltd*, 2014.

B.J. Hayes, P.J. Bowman, A.C. Chamberlain, K. Verbyla, and M.E. Goddard. Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution*, 41(1):51, 2009.

S.D. Kachman. Genomic prediction model based on haplotype clusters. In *Joint Statistical Meetings*, Seattle, WA, August 2015.

S.D. Kachman, 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, and E.J. Pollak. Comparison of molecular breeding values based on within- and across-breed training in beef cattle. *Genetics Selection Evolution*, 45:30, 2013.

T.H. Meuwissen and M.E. Goddard. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genetics Selection Evolution*, 36(3):261–279, 2004.

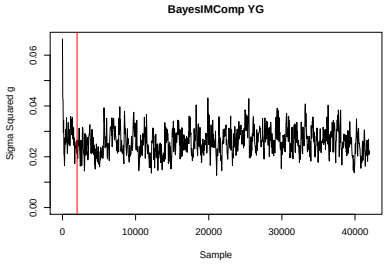
F.D.N. Mujibi, J.D. Nkrumah, O.N. Durunna, P. Stothard, J. Mah, Z. Wang, J. Basarab, G. Plastow, D.H. Crews, and S.S. Moore. Accuracy of genomic breeding values for residual feed intake in crossbred beef cattle. *Journal of Animal Science*, 89(11):3353–3361, 2011.

M.M. Rolf, D.J. Garrick, T. Fountain, H.R. Ramey, R.L. Weaber, J.E. Decker, E.J. Pollak, R.D. Schnabel, and J.F. Taylor. Comparison of bayesian models to estimate direct genomic values in multi-breed commercial beef cattle. *Genetics Selection Evolution*, 47(1):1, 2015.

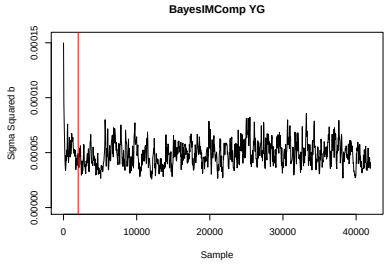
P. Scheet and M. Stephens. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *The American Journal of Human Genetics*, 78(4):629–644, 2006.

A Additional Plots

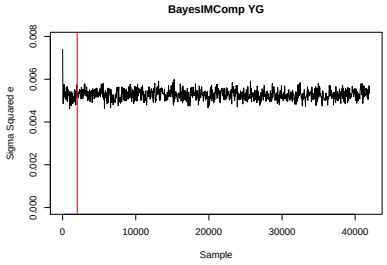
Figure 9: YG Trace Plots for Bayes IM Comp



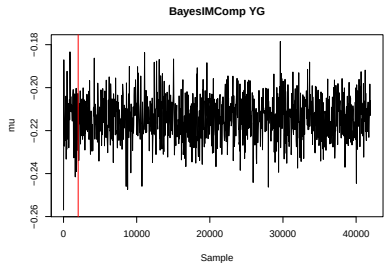
(a.)



(b.)

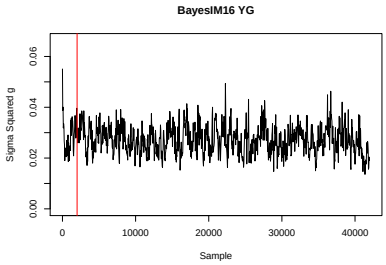


(c.)

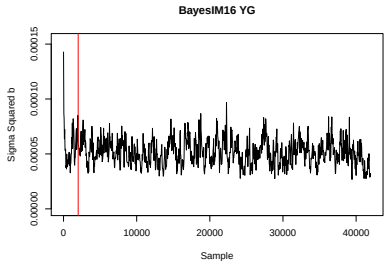


(d.)

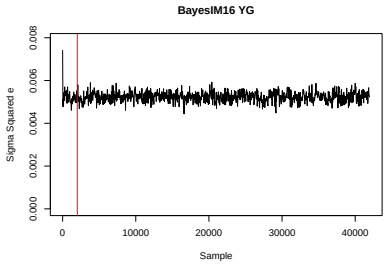
Figure 10: YG Trace Plots for Bayes IM 16



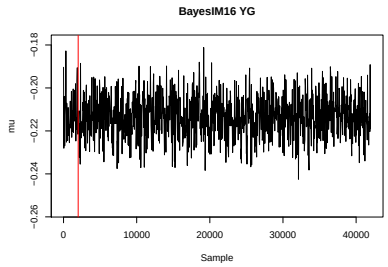
(a.)



(b.)

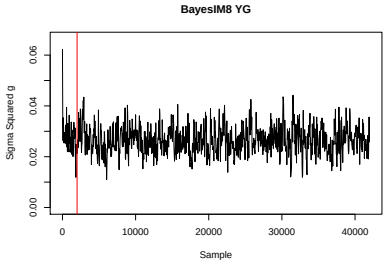


(c.)

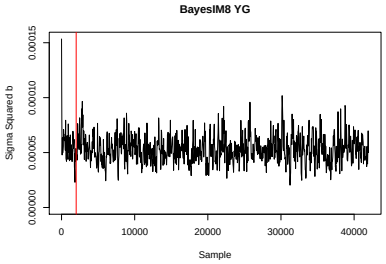


(d.)

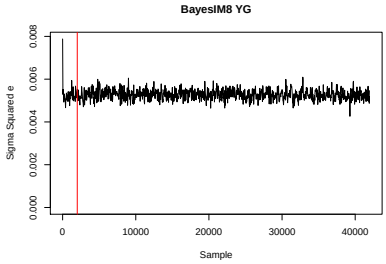
Figure 11: YG Trace Plots for Bayes IM 8



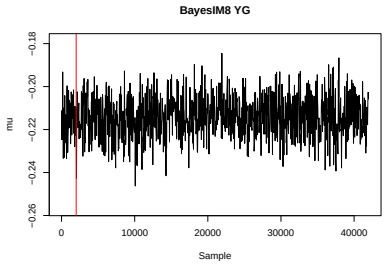
(a.)



(b.)



(c.)



(d.)