# Genetic Difference of Five Beef Cattle Breeds Characterized by Genome-wide SNPs and Haplotypes

## A.S. Leaflet R3064

Xiaochen Sun, Postdoctoral Research Associate;
Hailin Su, Postdoctoral Research Associate;
Dorian Garrick, Professor, Department of Animal Science,
Iowa State University

**Summary and Implications**

The objective of this study was to characterize breed differences using SNP available on commonly-used marker panels compared to using genome-wide SNP haplotypes derived from the same markers. The percentage of breed-specific segregating haplotypes was much higher than that of SNPs. Principal components of haplotypes characterized breed difference better than SNP genotypes. Results indicate that haplotypes characterize breed differences more adequately than SNP genotypes and hence offer promise to improve genomic prediction and fine mapping in multi breed populations.

**Introduction**

Linkage Disequilibrium (LD) reflects the extent to which a particular allele at one locus is inherited together with a particular allele at another locus. Genomic prediction using SNP panels will be more effective when a marker locus on the panel is in high LD with a causal mutation at a quantitative trait locus (QTL) having a major effect on a trait. In beef cattle populations, LD between QTL and SNPs can be high in one breed but low and inconsistent across breeds, because QTL can be fixed within certain breeds or have different minor allele frequencies (MAF) among breeds. The SNPs used for genotyping panels are selected to be informative by having moderate to high MAF, and these cannot well characterize breed differences at QTL with low frequency alleles. Thus breed differences characterized by SNP genotypes are inadequate in representing differences at QTL, which leads to low accuracy of across-breed genomic prediction.

Alleles at nearby SNP loci tend to be inherited together in what is referred to as a haplotype. The inadequate LD between SNPs and QTL can be improved using haplotypes within short genomic windows that are constructed from phased SNP genotypes. With sufficient SNP density, haplotypes 1) can be in complete LD with QTL irrespective of the MAF of QTL, and 2) can capture genetic variance at multi-allelic loci. The objective of this study was to investigate the advantage of haplotypes over SNP genotypes in characterizing breed difference of multiple beef cattle breeds.

**Materials and Methods**

The beef cattle datasets include 1,905 Angus (AAN), 1,449 Charolais (CHA), 1,214 Gelbvieh (GVH), 1,500 Hereford (HER) and 1,500 Simmental (SIM) animals that were genotyped with a Bovine 50K SNP chip. Within each breed, the 50K genotypes were imputed to a much higher density HD panel using a sample of HD-genotyped animals from each of the same breeds as reference populations. The imputed HD genotypes were phased for haplotype analyses. A total of 607,039 HD SNPs on chromosome 1 ~ 29 were used in analysis after removing SNPs with MAF $\leq 0.01$.

Each chromosome was divided into consecutive non-overlapping windows with length 0.1 or 1Mbp corresponding to about 1,000 or about 100 windows per chromosome. Haplotypes at each window with allele frequency $\leq 0.01$ were defined as rare haplotypes and otherwise as common haplotypes. At each window, the LD between each SNP and all unique haplotypes was evaluated using a $\chi^2$ statistic, which equaled 1 under complete LD and 0 under no LD, and the LD between pairwise SNPs was evaluated as the correlation coefficient between their genotypes. Principal component analysis was performed for animals in five breeds using SNP genotypes or haplotypes.

**Results and Discussion**

When the window length increased from 0.1 to 1 Mbp, the average number of SNPs within each window increased more than 10 fold, however, the average number of unique haplotypes only increased one fold, and the number of common haplotypes remained constant (Table 1). The average LD between pairwise SNPs was < 0.2 at low MAF and increased to 0.4 at high MAF. In contrast, the average LD between haplotypes and SNPs was close to 1.0 irrespective of MAF of SNPs (Figure 1). Results suggest that haplotypes are in almost complete LD with SNPs used in their construction - much higher than the LD between pairwise SNPs.

The percentage of SNPs segregating in all five breeds was 88%, with only 3.5% unique to CHA and less than 1% unique to each of the other four breeds. In contrast, less than 10% of common 0.1Mbp haplotypes were segregating in all five breeds, with a substantial percentage being unique to each of the five breeds (Figure 2). Separation of breeds by principal components was much clearer using 0.1Mbp haplotypes compared to SNP genotypes (Figure 3). Results suggest that haplotypes capture breed differences better than SNP genotypes due to the high proportion of breed-specific haplotypes.

In conclusion, haplotypes characterize breed differences much better than SNP genotypes. Fitting haplotypes can potentially improve accuracy of genomic prediction and

assist in fine mapping of QTL in structured populations including multiple breeds.

**Acknowledgments**

This work was supported by the Zoetis Cattle Call 2014 Research Grant.

**Table 1. Numbers of SNP loci, haplotype alleles and common haplotype alleles (allele frequency > 0.01) derived from 50K SNP genotypes imputed to HD panel densities in five beef cattle breeds**

| Window length | No. SNPs[1] | No. unique[2] | No. common[3] |
|---|---|---|---|
| 0.1Mbp | 18.8 | 227.3 | 18.7 |
| 1Mbp | 240.7 | 521.4 | 19.4 |

[1]Average number of HD imputed SNPs within each window in five breeds
[2]Average number of all unique haplotypes within each window in five breeds
[3]Average number of common haplotypes within each window in five breeds

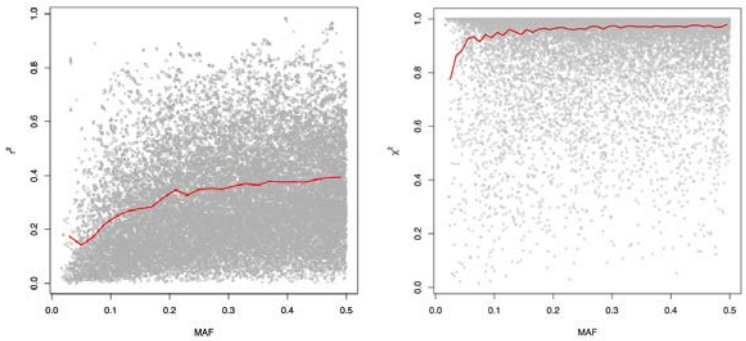


**Figure 1. Linkage Disequilibrium (LD) in relation to minor allele frequency (MAF) between pairwise SNPs (left) and between 0.1 Mb window haplotypes and SNPs (right).** Red lines show average LD at each MAF.

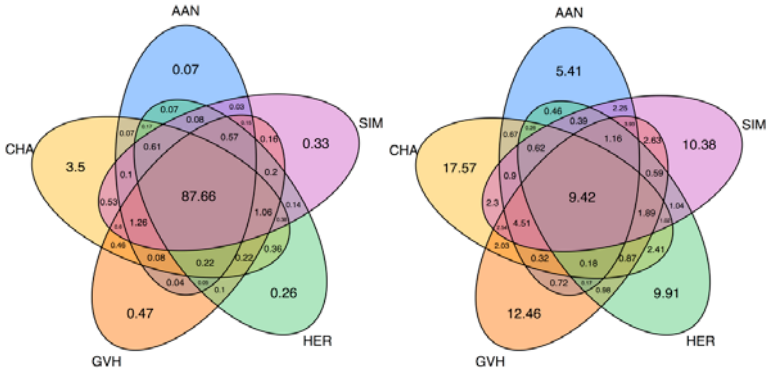


**Figure 2. Percentage of shared SNPs (left) and shared 0.1Mbp haplotypes (right) in five breeds.**

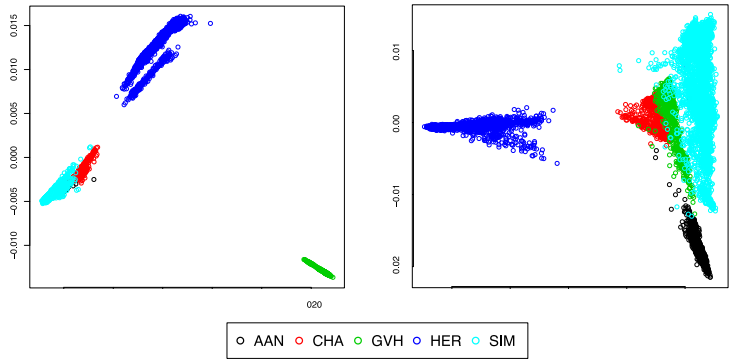


**Figure 3. First 2 principal components of five breeds using SNP genotypes (left) and 0.1Mbp haplotypes (right).**