Haplotype Diversity of Ten US Beef Cattle Breeds Captured by Different Definitions of Haplotypes Based on BovineSNP50K Chip

***H. Su*[1], R. L Fernando[1] and D. J Garrick[1]**
[1]Department of Animal Science, Iowa State University, Ames, IA 50011, USA.

ABSTRACT: Different ways of defining haplotypes from marker alleles fundamentally affects the size of haplotype blocks and the corresponding number of alleles. Here we studied a dataset comprising over 32,000 animals genotyped with BovineSNP50K and representing 10 beef cattle breeds, seeking to reveal the extent of genetic diversity captured by haplotypes constructed from adjacent marker alleles (1) of arbitrary number of SNPs and (2) within predefined Mb-based windows according to UMD3.1 map positions. The average numbers of common haplotype alleles, which we defined as those that were observed at a frequency of at least 0.01 in each breed, were investigated in relation to the width of haplotypes. The results suggested that 20 SNP haplotypes and/or 1-Mb haplotypes were preferred and these are typically characterized by 20 alleles.
Keywords: beef cattle, haplotype definition, BovineSNP50K

Introduction

Single-nucleotide polymorphism (SNP) markers are extremely common throughout the genome. Individual genes or quantitative trait loci (QTL) carrying causal mutations likely have multiple SNPs in their coding areas and surrounding regions. The alleles at these sites tend to be inherited together in discrete blocks known as haplotypes that are preserved unless there is a recombination event within the block to create a new haplotype. Haplotypes are of direct scientific interest as they may be in perfect linkage disequilibrium (LD) with QTL alleles, and fitting haplotypes in genome-wide association studies can provide greater statistical power to capture effects of causal genes, because haplotypes determine the observed LD between SNP markers (Fallin & Schork, 2000). Thus the width and therefore number of haplotype blocks, and their diversity in terms of number of alleles are fundamental factors influencing the computational effort, results of association studies, and the accuracy of genomic prediction. Simulation studies have tended to construct haplotypes with constant numbers of adjacent SNP alleles (Calus et al, 2007; Hickey et al, 2012), while in field data 1Mb window haplotypes have been commonly used (Garrick and Fernando, 2013).

The objective of this research was to determine the number of markers to use to define haplotypes for research in beef cattle based on BovineSNP50K markers. This was done by revealing the extent of haplotype diversity captured by haplotypes constructed from adjacent marker alleles (1) of arbitrary number of SNPs and (2) within predefined Mb-based windows according to UMD3.1 map positions.

Materials and Methods

Data. A dataset comprising over 32,000 animals genotyped for 50K SNPs and representing 5,692 Hereford (HER), 1,794 Red Angus (RAN), 5,242 Simmental (SIM), 1,418 Brangus (BRG), 11,360 Angus (AAN), 3,275 Limousin (LIM), 1,467 Gelbvieh (GVH), 996 Charolais (CHA), 956 Maine-Anjou (RDP) and 450 Shorthorn (BSH) was used.

Phasing and imputation. The genotype data were processed to Beagle-input files using GenSel4r (Fernando and Garrick, 2010). The linkage phase of haplotypes were inferred and missing genotypes were imputed using Beagle v3.3.2 (Browning and Browning, 2009) separately for data from each breed.

Haplotype construction. Two methods were adopted in building haplotypes from SNP genotype data. According to UMD3.1 map positions, method 1 treated every constant number (2~100) of adjacent SNP markers as a haplotype block, and method 2 grouped the SNP alleles into Mb-based (0.1~8.0) haplotype windows.

Haplotype diversity analysis. Within each haplotype block, we counted the number of haplotype alleles, calculated their frequencies, and defined common haplotype alleles as those observed at a frequency of at least 1 in 100 in each breed. The average number of common haplotype alleles over all haplotype blocks was used to measure the diversity captured by different methods of defining haplotypes.

Results and Discussion

The results showed that, in the haplotype diversity analysis with both haplotype definitions, the general trend with increasing haplotype width was explosive growth in the average number of observed common haplotype alleles, followed by a smooth long-tail drop down after reaching the maximum peak value (Figures 1 and 2). The increase in haplotype diversity suggested that more haplotype alleles were formed when taking more SNP markers into account, and in this process common haplotype alleles would become less and less common. Some would become rare if they were in complete LD with low-frequency SNP alleles. When much more rare haplotype alleles than common ones were being generated, the increasing trend of diversity of common haplotypes slowed down, and the number of common alleles started to decline. The genome is most simply characterized when the width of haplotype is chosen to maximize the proportion of the genome represented by haplotype alleles that were common by our definition.

For the ten beef cattle breeds, both figures 1 and 2 showed similar trends of genomic diversity: CHA, GVH and LIM showed faster increases and decreases in diversity than AAN, BRG and HER. This suggested that the latter breeds have a relatively more constant and longer haplotype blocks, which is evidence of their smaller effective population sizes. This inference is quite similar to some of the results of reported studies on effective population size in CHA 958[*] (Leroy et al, 2013), LIM 740[*] (Leroy et al, 2013), AAN 654 (Saatchi et al, 2011), BRG 95.4[†] (Ron Garrido et al, 2008), and HER 85 (Cleveland et al, 2005).

Conclusion

Theoretically, the combination of n SNP markers produces 2^n different haplotype alleles. However, when n is greater than 2, the number of observed haplotype alleles in finite populations is much smaller than 2^n due to LD between adjacent SNP alleles, and an even smaller proportion of those haplotype alleles are presented as common ones. To capture haplotype diversity as much as possible, it is suggested from the results of 50K marker analysis that 20 SNP haplotype blocks and/or 1-Mb haplotype windows were the appropriate widths where the captured diversity is maximized and these blocks are typically characterized by ~20 common haplotype alleles per locus. Further work using higher-density SNP panels is required to determine the extent that haplotype alleles that appear identical using 50K markers are in fact identical.

Literature Cited

Browning, B. L., and Browning, S. R. (2009). Am. J. Hum. Genet., 84:210-223.

Calus, M. P. L., Meuwissen, T. H. E., de Roos A. P. W. et al. (2008). Genet., 178-1:553-561.

Cleveland, M. A., Blackburn, H. D., Enns, R. M. et al. (2005). J. Anim. Sci., 83:992-1001.

Fallin, D., and Schork, N. J. (2000). Am. J. Hum. Genet., 67:947-959.

Fernando, R. L., and Garrick, D. J. (2010). Iowa State University.

Garrick, D. J., and Fernando, R. L. (2013). Chapter 11, Methods in Mol. Biol. Vol 1019. Springer.

Hickey, J. M., Kinghorn, B.P., Tier, B. et al. (2013). J. Anim. Breed. Genet., 130:259-269.

Leroy, G., Mary-Huard, T., Verrier, E. et al. (2013). G. S. E., 45:1.

Ron Garrido, L., Birchmeier, A. N., Munilla, S. et al. (2008). Livest. Sci., 117:43-51.

Saatchi, M., McClure, M. C., McKay, S. D. et al. (2011). G. S. E., 43:40.

Figure 1: Average common haplotype alleles observed per haplotype block with different haplotype widths by number of SNP markers

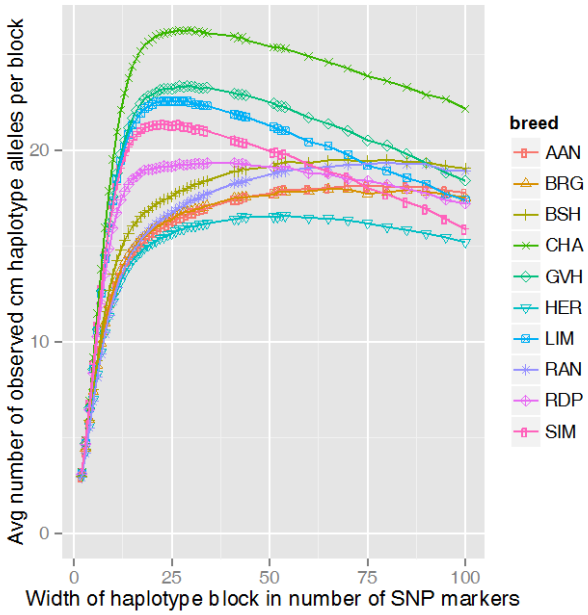
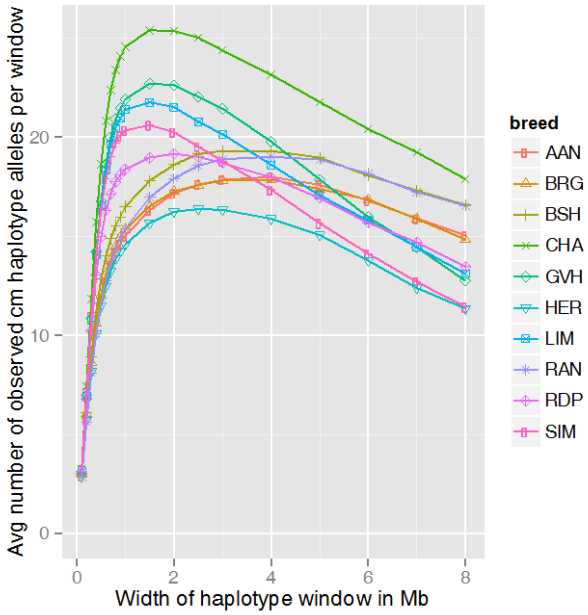


Figure 2: Average common haplotype alleles observed per haplotype window with different haplotype widths by Mb



[*] Based on individual co-ancestry rate model.
[†] Under random mating.