Characterizing Haplotype Diversity in Ten US Beef Cattle Breeds

A.S. Leaflet R2846

Hailin Su, Postdoctoral Research Associate; James Koltes, Postdoctoral Research Associate; Mahdi Saatchi, Postdoctoral Research Associate; Jungjae Lee, Postdoctoral Research Associate; Rohan Fernando, Professor; Dorian Garrick, Professor, Department of Animal Science

Summary and Implications

The distributions of haplotype diversity across the whole genome among 10 US beef cattle breeds were constructed. In most chromosomes for all the breeds, consistent low haplotype diversity were observed in some specific regions, 55% of which was found to match the positions of reported gene duplications. Further work is required to determine whether the low haplotype diversity is real, or a result of problems in sequencing which have limited our ability to identify informative markers in those regions. Haplotype diversity will be the subject of ongoing work to identify haplotypes that are under-represented as homozygotes, to fine-map regions with major gene effects, and to fit haplotype rather than SNP models for genomic prediction.

Introduction

Single-nucleotide polymorphism (SNP) markers are extremely common throughout the genome. Individual genes or quantitative trait loci (QTL) carrying causal mutations often include multiple SNPs in its coding areas and surrounding regions. The alleles at these mutations 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 may be separated by recombination hot spots. The size and number of haplotype blocks are fundamental factors influencing the results of association studies and the accuracy of genomic prediction.

Haplotypes are of direct scientific interest as fitting them in genome-wide association studies can provide greater statistical power to capture effects of causal genes because they determine the observed linkage disequilibrium (LD) between SNP markers (Schork et al. 2000). Distinct populations such as cattle breeds may carry some haplotypes in common. In this study, we reconstructed haplotypes within each 1-Mb SNP window for 10 US beef cattle breeds that had been genotyped with the Illumina BovineSNP50K.

Materials and Methods

A dataset comprising over 32,000 animals genotyped for 50K SNPs and representing 5,692 Hereford, 1,794 Red Angus, 5,242 Simmental, 1,418 Brangus, 11,360 Angus, 3,275 Limousin, 1,467 Gelbvieh, 996 Charolais, 956 Maine-Anjou and 450 Shorthorn, pure-bred and cross-bred individuals was used. The linkage phase of haplotypes were inferred and missing genotypes were imputed using Beagle separately for data from each breed association, and then all the SNP alleles were grouped into 1-Mb windows according to UMD3.1 map positions, forming haplotypes consisting of on average 20 SNPs in width. We investigated the number of haplotype alleles and their frequencies within each window, and the diversity of common haplotypes, which we defined as those observed at a frequency of at least 1 in 100 in each breed. We related the regions where the common haplotype diversity was low to the reported QTL database.

Results and Discussion

The average number of common haplotypes across the entire genome was 18 and ranged from 15~25 in individual breeds. Some specific windows showed consistent increased or decreased haplotype diversity in all breeds. Decreased haplotype diversity was observed in some windows of most chromosomes for all the breeds. Some 55% of these regions matched positions of reported gene duplications (Table 1). This information provides direction for future studies to characterize haplotype diversity in relation to annotated gene-rich regions, copy number variation, published QTL, selection signals and loss-of-function mutations.

Table 1. Positions of low-diversity haplotype regions overlap or flank gene duplications in the bovine genome.

Chromosome	Position*	Avg # Alleles	Duplicated Gene Family
5	59	5.5	Olfactory receptors
7	7, 9, 11	6.0, 4.0, 4.9	Olfactory receptors
7	51	4.7	Procadherin beta family
10	23, 24	3.9, 4.2	Olfactory receptors
12	73, 74, 75	1.8, 1.7, 2.4	ABCC4 family
14	24	10.0	PLAG1 region
18	59	5.0	Kallikrein-related peptidase family; Zinc Finger family
23	26	5.6	BOLA families
27	5	3.7	Beta-defensin family
29	39	4.4	Pregnancy associated glycoprotein family
All	All	20.4	-

* N^{th} 1-Mb window of the same chromosome started from zero

Acknowledgments

Hailin Su is funded by Endowment of the Lush Chair.