Joint Genome-Wide Gametic and Zygotic Linkage Disequilibria Describe The Distinct Domestication Between Dairy and Beef Cattle Populations

Z. Wang[*], H. Li[*], R.-C. Yang[*,†] , X.-S. Hu[*,‡] , S. S. Moore[*], G. Plastow[*]

Introduction

The long time of domestication in *Bos taurus* brings about a substantial decline in effective population size and creates genome-wide linkage disequilibrium (LD) (Goddard and Hayes 2009, MacEachern et al. 2009). Such genome-wide LD gradually decays with time due to recombination. However, directional artificial selection modifies this process, facilitating heterogeneous genetic variation among chromosome regions. A transient positive correlation between genetic diversity at linked sites can be generated via the genetic hitchhiking between selective and adjacent neutral sites. A negative (or positive) correlation can be produced through the antagonistic (or synergistic) interactions between linked selective sites. Characterizing such genomic structure can be used to infer the population history, and/or to aid genomic selection. Here, we use pairwise gametic and zygotic LD between SNP sites along chromosomes to elucidate bovine genomic structure. Gametic LD is widely used in population genomic studies (Slatkin 2008), but zygotic LD is applied in a limited way. The differences between gametic and zygotic LD are complicated (Yang 2002). The linkage phase must be known or inferred in the gametic LD calculation but not necessarily in the zygotic LD analysis. Another crucial difference is that directional artificial selection has direct effects on the zygotic LD but only indirect effects on the gametic LD in cattle domestication. This may result in distinct footprints of gametic and zygotic LD in historical domestication. In this study, we select two populations contrasting in production purpose (meat vs. milk) to examine the inter-site correlations between heterozygosities and between allele frequencies. Patterns for the two types of correlations are compared to capture signals that contrast between the two populations.

Material and methods

The dairy population consists of 647 proven bulls born in North America born between 1985 and 2002 from Semex Alliance, Canada. Bulls were progenies of 71 sires and 492 dams with an average paternal family size of 9.1. The beef population consists of 1,023 hybrid steers maintained at the University of Alberta's Kinsella research station from 2003 to 2008. They were progenies of the crosses between Angus, Charolais, or University of Alberta hybrid bulls and the University of Alberta experimental hybrid dam line. In each population, genotyping was carried out using the Illumina BovineSNP50 BeadChip. Two measures are the square of standardized gametic LD (r_D^2) and the correlation between the heterozygositie (Yang 2002). Here we present the results for only two *Bos taurus* autosomes (BTAs 23 and 29) and details will be given elsewhere.

[*]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, T6G 2C8 Canada. [†]Alberta Agriculture and Rural Development, 307 J. G. O'Donoghue Building, 700-113 Street, Edmonton, Alberta, T6H 5T6 Canada. [‡]Corresponding author: xin-sheng.hu@ualberta.ca

Results

Both the dairy and beef populations have a similar range of heterozygosity (0.0~0.6) across all SNP sites on BTA23 and BTA29. Figure 1 shows a comparison in heterozygosity distribution between these two populations. The dairy population has more SNP sites fixed (~13.2% vs. 8.3%), and slightly fewer SNP sites with heterozygosities around the expected maximum ($H_e \approx 0.5$, ~ 21.2% vs. 23.0%). The beef population has larger genetic diversity on average across both chromosomes than the dairy population (Table1). Both populations have a certain proportion of sites with excess heterozygotes, 8~12%. This is probably due to artificial selection favouring heterozygotes in some chromosomal regions. It deserves further investigation for low heritability (narrow-sense) traits where over-dominance may be important.

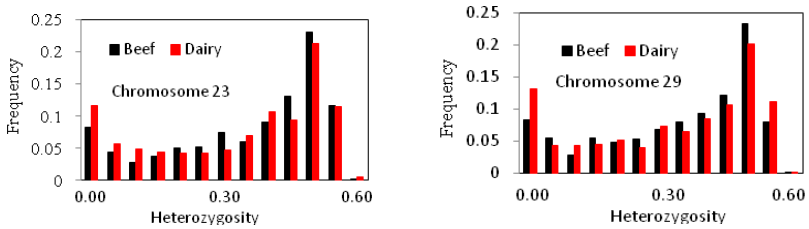

Figure 1. A comparison between the dairy and beef populations for the abundance of SNP sites at each heterozygosity interval on BTAs 23 and 29.

Among all pairwise correlations between heterozygosities, a slightly higher number of SNP pairs, 50.3~52.5%, have positive correlations in each population (Table 1). The average correlation coefficients change from -0.0409 ± 0.0310 to 0.0653 ± 0.0803. The dairy population has a greater variation than the beef population for either positive or negative correlations. Figure 2 shows the pattern for the change of the pairwise correlation between heterozygosities with distance. A few SNP pairs have very high correlations in each population. A usual negative exponential distribution exists for all positive correlations, with high correlations for neighbouring SNPs and low correlations for distant SNPs. However, this is not the case for all negative correlations that are essentially insensitive to the distance change.

Table 1. A comparison between the dairy and beef populations in H_e, r_H ,and r_D

	BTA23(dairy)	BTA23(beef)	BTA29(dairy)	BTA29(beef)
SNPs	1036	1023	1007	985
H_e	0.3082 ± 0.1862	0.3316 ± 0.1711	0.3029 ± 0.1842	0.3175 ± 0.1725
r_H (>0)%	52.5	50.8	51.4	50.3
r_H (>0)	0.0653 ± 0.0803	0.0406 ± 0.0458	0.0620 ± 0.0759	0.0422 ± 0.0483
r_H (<0)	-0.0402 ± 0.0304	-0.0332 ± 0.0258	-0.0409 ± 0.0310	-0.0343 ± 0.0264
r_H^2 (>0.3) pairs	802 (0.19%)	166 (0.04%)	542 (0.14%)	182 (0.04%)
r_D^2 (>0.3) pairs	3195 (0.76%)	411 (0.09%)	1869 (0.61%)	409 (0.10%)

To compare the pattern for the two types of correlations, the same set of markers was used to calculate pairwise r_D^2. As expected, a negative exponential distribution exists for the change of r_D^2 with distance (details not shown here), similar to the findings of previous studies in different species. Both r_H^2 and $r_D^2$are calculated, with a comparable number of SNP pairs using the same set of SNP markers on each chromosome. Compared with the beef population, the dairy population has a substantially larger number of SNP pairs with r_H^2 greater than 0.3. A similar result to the pattern of r_H^2 is observed for the number of SNP pairs with r_D^2greater than 0.3 (Table 1). The dairy population exhibits stronger gametic and zygotic LDs than the beef population.

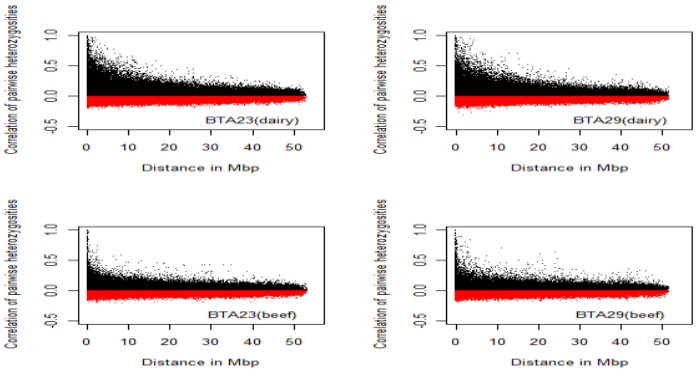

Figure 2. Patterns for the change of r_H with distance in Mbp show the presence of a negative exponential distribution for all positive correlations (black dots), but the absence of this trend for all negative correlations (red dots) in both populations.

Figure 3 (left 4 graphs) shows that the empirical cumulative distribution function (e.c.d.f.) goes up faster for r_H^2 than for r_D^2 in the vicinity to zero (fixation), indicating that a general weaker correlation (chromosome-based) exists between heterozygosities (e.g., the right graph in Fig. 3). Kolmogorov-Smirnov tests indicate that a significant difference between r_H^2 and r_D^2 ($p < 2.2 \times 10^{-16}$) exists between the patterns of two type correlation distributions.

Discussion

In this study, we employ two types of correlation to characterize the structure of genomic diversity in dairy and beef populations. Correlation between heterozygosities directly describes how the genetic diversity at one locus is associated with that at the other. One striking result is that the dairy population has more SNP pairs with strong zygotic LD (r_H^2 >0.3) than the beef population. This suggests that these regions are under stronger selection, signalling the occurrence of selection history during the dairy domestication. Selection for genetic improvement could result in strong positive correlation between heterozygosities at neighbouring sites. It is commonly held that the dairy population has experienced more intensive directional selection through artificial insemination in recent history than the beef population. This selection signature seems evident from the heterozygosity distribution (Fig.1) or from the lower average heterozygosity (Table 1). Unlike gametic LD eroding over generations due to recombination, zygotic LD can be maintained in the population as long as the selection direction and pressure are unchanged. Under this situation, the zygotic LD is

more informative on genetic improvement through genomic selection. In the dairy population, strong correlations between heterozygosities in some chromosome regions are signalling the presence of strong artificial selection.

In the beef population, a large heterozygosity, along with a very small percentage (0.04%) of SNP pairs with r_H^2 greater than 0.3, is consistent with less intensive selection for the beef population. This result is also in agreement with the breeding history of the composite beef population used here where multiple sires and dam lines might have experienced extensive recombination.

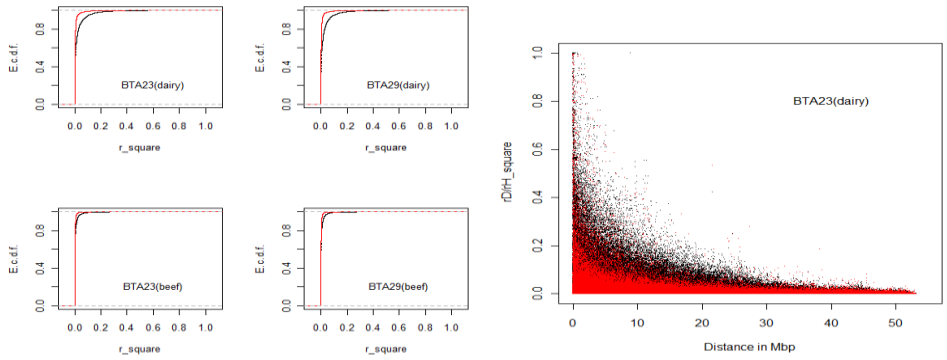

Figure 3. The left four figures show the patterns for the empirical cumulative distribution function (e.c.d.f.), indicating that more small values of r_H^2 (red line) exist compared with r_D^2 (black line). The right one figure shows that r_H^2 (red dots) collapses faster than r_D^2 (black dots) with distance on BTA23 in the dairy population.

Although a similarity exists between r_H^2 and r_D^2 in revealing the difference between the dairy and beef populations, they have different statistical properties (Table 1). Yang (2002) demonstrated that higher order LDs (three- and four-allele pairs) are involved in r_H, which can lead to a differentiation between r_H^2 and r_D^2. The present results demonstrate that the correlation between heterozygosities is generally weaker than the gametic LD and decays faster with distance (Fig.3). When linked to effects of evolutionary forces, use of r_H^2 for population parameter estimation, as implied from the r_D^2 theory, requires further exploration.

Acknowledgements

The work is financially supported by research grant #2008F175R from Alberta Livestock Industry Development Fund Ltd (ALIDF) and Alberta Agricultural Research Institute (AARI) to ZW.

References

Goddard, M., Hayes, B.J. (2009). *Nat. Rev. Genet.*10: 381-391.

MacEachern, S., Hayes, B., McEwan, J., et al. (2009). *BMC Genomics* 10:181.

Slatkin, M. (2008). *Nat. Rev. Genet*. 9:477-485.

Yang, R.C. (2002). *Genet*. 161: 435-445.