181. Impact of breed ancestry on the performance of crossbred dairy cattle

M.A. Jaafar[1*], B.J. Heins[2], C. Dechow[3], C.P. Van Tassell[4] and H.J. Huson[1]

[1]Cornell University, Animal Science Department, 149 Morrison Hall, Ithaca NY 14850, USA; [2]University of Minnesota, West Central ROC, 46352 State Hwy 329, Morris, MN 56267, USA; [3]Penn State University, Department of Animal Science, University Park, State College, PA 16802, USA; [4]United States Department of Agriculture, Animal Genomics and Improvement Lab, 10300 Baltimore Ave, Beltsville, MD 20705, USA; maj223@cornell.edu

Abstract

Crossbred dairy cattle populations have different proportions of parental breeds. However, individuals who display the same breed composition could exhibit remarkable differences in the distribution of local ancestry across the genome. Thus, this study examined the distribution of local ancestry blocks using Illumina BovineSNP50 genotypes and correlated them to key production traits. Two rotational crossbred populations, referenced as ProCROSS and Grazecross were assessed. The results showed higher performance ProCROSS animals in milk fat production more commonly have Holstein ancestry on BTA 2, 6, 13, 22 and 23, where the genes Prion, Lactotransferrin, and Cytochrome P450 reside. We found evidence of an excess proportion of Viking Red ancestry and decreased Jersey ancestry on BTA 1, 10, 12, 14 and 25 at known health QTL in the Grazecross population. In conclusion, we believe that localized breed ancestry plays an important role; thus, selecting and maintaining particular ancestry blocks is crucial to ensure optimal trait performance in crossbred cattle.

Introduction

Generally, crossbreeding is the breeding among individuals from at least two distinct populations. The admixed individuals will inherit haplotype blocks that contain genetic variants of one or more populations. The melding of these haplotype blocks increases the heterozygosity level that contributes to heterosis or hybrid vigour in the admixed offspring. However, this heterozygosity may not always exist, as recombination events occur randomly in gamete formation, creating unique combinations of chromosome segments from each population. Due to this, admixed individuals may display the same breed composition but exhibit differences in the distribution of local ancestry blocks that directly affect complex production trait performance. Mapping local ancestry composition instead of global ancestry could aid in identifying causal variants in complex traits, especially ancestry-specific traits in admixed individuals (Ding *et al.*, 2011). Therefore, our primary goals were to describe the distribution of local ancestry blocks across the genome of the ProCROSS and Grazecross dairy cattle populations. Furthermore, we aimed to explore the correlation of the local ancestry blocks with various production traits.

Materials & methods

Study population. The University of Minnesota (UMN) granted access to ProCROSS and Grazecross datasets. The datasets included 378 ProCROSS and 229 Grazecross cattle genotypes and four traits (milk yield (MY), fat yield (FY), protein yield (PY) and somatic cell score (SCS)). To compare ancestry variation for trait performance, the top 20% high-performance animals were compared to the lowest 20% low-performance animals for each trait. To investigate admixture of the ProCROSS and Grazecross, genotypes representing the purebred and admixed breeds used in the two rotational systems were needed for comparison. ProCROSS are a product of rotational crossbreeding of Viking Red (VKR), Holstein (HOL), and Montbeliarde (MON). In contrast, Grazecross consists of VKR, Normande (NOR), and Jersey (JER). VKR breeds were represented by 3 genotype breeds Danish Red (DNR, n=4), Swedish Red

(SWD, n=32) and Finnish Ayrshire (FAY, n=27) (Upadhyay *et al.*, 2017; Iso-Touru *et al.*, 2016, Viking Genetics, Denmark). HOL genotypes (n=92) were provided by UMN obtained from animals directly used in the development of these two admixed populations. Jersey (JER, n=73), Montbeliarde (MON, n=34), and Normande (NOR, n=35) genotypes were obtained from the Gautier *et al.* (2010) publication and represented purebred animals from the USA (JER) and Europe (MON & NOR), respectively. SNP data of all animals were obtained through bovine genotyping array kits BovineHD DNA Analysis Kit (HD150K), BovineSNP50 DNA Analysis BeadChip (50K) and CLARIFIDE® 50k (ZL5) (Illumina- Neogen, Lansing, MI, USA; Clarifide-Zoetis, San Diego, CA, USA).

Filtering and quality control of genomic data. Quality control (QC) analyses for both sets of autosomal SNPs were calculated using SNP and Variation Suite (SVS) v8.x (Golden Helix, Inc., Bozeman, MT, USA). SNPs were excluded if they were unmapped to the UMD 3.1 bovine genome assembly (Zimin *et al.*, 2009) or mapped to the sex chromosomes, had a call rate less than 0.95, or a minor allele frequency less than 0.05. SNPs were further pruned for linkage disequilibrium (LD) using a threshold of $r^2>0.75$, leaving 9,792 SNPs and 10,795 SNPs for ProCROSS and Grazecross analysis respectively. Overall, SNPs used provided relatively clear separation between ancestral and admixed populations.

Admixture analysis. The global breed composition was estimated from genomic data using a maximum likelihood model implemented in ADMIXTURE 1.23 software. The datasets were analyzed using unsupervised clustering analysis with K=3, where K represented the expected number of ancestral populations in the 3-breed rotational crossbreeding systems. Average global ancestry was determined for each extreme trait performance group with significance in ancestry variation between performance groups assessed using a T-test.

Local ancestry. The pipeline established by Martin *et al.* (2017) was used in developing the local ancestry inference. Crossbred individuals were first phased based on their respective reference population using SHAPEIT2 (v.2) before all haplotypes were merged using custom scripts. These combined phased haplotypes were used as input into the PopPhased version of RFMix v.1.5.471 with the following flags: -w 0.2, -e 1, -n 5, – use-reference-panels-in-EM, and – forward-backward EM. Node size was set to five to reduce bias in random forests resulting from unbalanced reference panel sizes (HOL panel n=92, MON panel n=34, and VKR panel n=63). Several methods were assessed to determine the best representation of local ancestry for each extreme performance group. The method shown here calculated the number of occurrences of the particular breed ancestry for each SNP within each performance group. The ancestry having the highest number of occurrences was considered the ancestry origin for a SNP within a specific performance group.

Results & discussion

Global ancestry composition. Admixture analysis was run to assess the global ancestry composition of ProCROSS and Grazecross. The admixture plots (Figure 1) for both populations revealed proportions of ancestral breeds between 23 and 34%, expected with the 3-breed crossbreeding system. Comparing the average of global ancestry composition between high- and low-performance animals showed that HOL composition plays a significant role in MY and FY in ProCROSS populations, while in the Grazecross population, VKR and NOR composition play a significant role in SCS performance (*P*-value<0.05). This finding provided the basis for understanding the impact of breed-specific ancestry on the performance of crossbred cattle.

Local ancestry. Local ancestry karyograms representing each extreme group were plotted for significant production traits in each admixed population (Figure 2) so that a comparison can be made to identify several potential regions correlated to the performance. A set of breed-specific QTL was extracted from
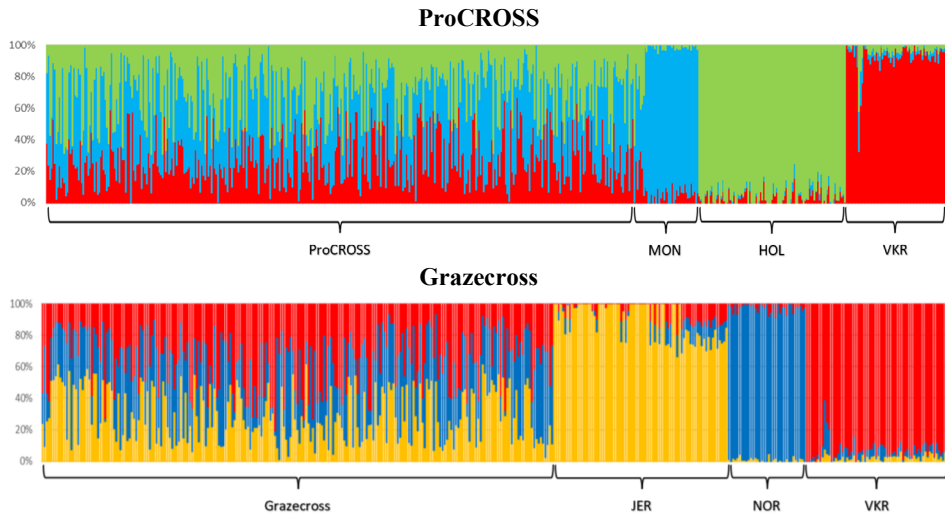
ProCROSS

Grazecross

Figure 1. Admixture analysis reflecting global ancestry of breeds for both ProCROSS and Grazecross admixed individuals and all parental populations for K=3.
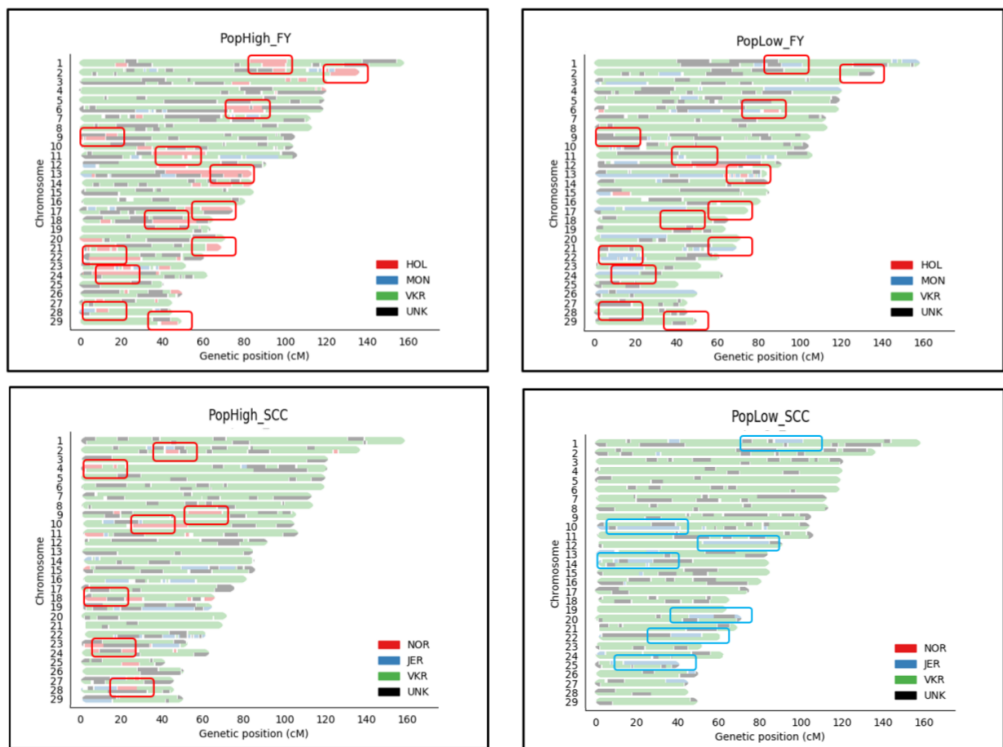

Figure 2. Local ancestry analysis compares high-performance groups (Left side diagrams) and low-performance groups (Right side diagrams) for both ProCROSS (Top) and Grazecross (Bottom) populations.

the Animal QTL database (https://www.animalgenome.org) and several recent publications. Higher ProCROSS FY production animals more commonly have visually larger HOL ancestry on BTA 2, 6, 13, 22 and 23, where genes for Prion, Lactotransferrin, and Cytochrome P450 reside and are denoted by the red box in Figure 2 (top). On the other hand, we found evidence of an excess proportion of VKR ancestry (red box) together with a decrease of Jersey ancestry (blue box) on BTA 1, 10, 12, 14 and 25 at known health QTL for high- and low-performance Grazecross animals respectively (Figure 2 bottom). These findings agreed with the global ancestry results but warrant further investigation to determine if the local ancestry variation visually seen is significantly different between performance groups. In addition, approaches for defining local ancestry for a performance group is still being assessed so that haplotype blocks define group ancestry as opposed to individual SNPs.

References

Ding L., Wiener H., Abebe T., Altaye M., *et al.* (2011) BMC Genom., 12, 622. https://doi.org/10.1186/1471-2164-12-622

Gautier M., Laloë D., and Moazami-Goudarzi K. (2010). PloS One, 5(9), e13038. https://doi.org/10.1371/journal.pone.0013038

Iso-Touru T., Tapio, M., Vilkki, J., Kiseleva T., *et al.* (2016). Animal Genetics, 47(6), 647-657. https://doi.org/10.1111/age.12473

Martin A. R., Gignoux C. R., Walters R. K., Wojcik G. L., *et al.* (2017). Am. J. Hum. Gen., 100(4), 635-649. https://doi.org/10.1016/j.ajhg.2017.03.004

Upadhyay M. R., Chen W., Lenstra J. A., Goderie C. R. J., MacHugh D. E., *et al.* (2017). Heredity, 118(2), 169-176. https://doi.org/10.1038/hdy.2016.79