188. Patterns of inbreeding and selection using runs of homozygosity in North American dairy cattle

E.A. Lozada-Soto[1*], F. Tiezzi[2], J.B. Cole[3], P.M. VanRaden[4] and C. Maltecca[1]

[1]Department of Animal Science, North Carolina State University, Raleigh, NC 27607, USA; [2]Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, 50144 Florence, Italy; [3]URUS Group LP, Madison, WI 53718, USA; [4]Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Service, USDA, Beltsville, MD 20705, USA; ealozada@ncsu.edu

Abstract

The main objective of this study was to leverage genomic information to ascertain patterns of inbreeding and selection in five North American dairy cattle populations. We obtained genotypes for over 4 million individuals of the Ayrshire, Brown Swiss, Guernsey, Holstein, and Jersey breeds. Inbreeding based on runs of homozygosity was calculated in each population. The average inbreeding ranged from 0.11 for Ayrshire to 0.17 for Jersey. We calculated a coefficient of homozygosity for each marker. Highly homozygous markers were joined into larger genomic segments of interest that ranged from 0.08 to 7.83 Mb in length and spanned 14 chromosomes across breeds. Annotation of genes and QTLs in the highly homozygous regions revealed selection for economically important traits, notably for udder and cow health, productive life, and reproductive traits. We found differences across breeds on inbreeding load, genomic regions of high inbreeding, and selection signatures.

Introduction

Runs of homozygosity (ROH) are continuous stretches of markers in a homozygous state in the genome of an individual. The characterization of the length, number, and genomic distribution of ROH allow us to ascertain the level of inbreeding in individuals and decipher the inbreeding history of populations (Gibson *et al.*, 2006). The proportion of the autosomal genome covered by ROH segments has been routinely used as a measure of inbreeding in cattle populations (Howard *et al.*, 2015; Makanjuola *et al.*, 2020) due to its high correlation with other inbreeding coefficients and autozygosity (Peripolli *et al.*, 2017). We can also use the level of homozygosity at the marker level to identify regions of selection within a population and uncover differences in the selection history of multiple populations. Therefore, our aim was to characterize ROH segments, calculate the level of inbreeding, and identify regions of high homozygosity as selection signatures in the five major breeds of North American dairy cattle.

Materials & methods

Animals and genotypes. Imputed genotypes were obtained for 76,389 autosomal markers of 4,173,679 animals of the Ayrshire (AY), Brown Swiss (BS), Guernsey (GU), Holstein (HO), and Jersey (JE) breeds. Quality control of the genomic data was first performed within each breed, including the removal of animals with a marker call rate lower than 95%, markers with a minor allele frequency lower than 0.01%, and markers with Mendelian error rates larger than 10%. Subsequently, we removed markers with an across-breed call rate lower than 95%. After genomic quality control, a total of 4,010,718 animals (8,413 AY, 39,040 BS, 3,834 GU, 3,525,992 HO, and 433,439 JE) and 62,546 markers remained for subsequent analyses.

Detection of ROH and calculation of segment-based inbreeding. Runs of homozygosity (ROH) were detected using a sliding window approach with the PLINK v.1.9 software (Chang *et al.*, 2015). Parameters to define windows and detect ROH were the following: a minimum window size of 20 markers, at most 1 heterozygote call in a window, at most 2 missing calls in a window, a minimum physical length of 1 Mb for ROH, a maximum gap of 0.5 Mb between consecutive markers, a minimum marker density of 1 marker

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_188, © E.A. Lozada-Soto *et al.* 2022

per 0.1 Mb, a minimum number of 60 markers to declare a ROH. Parameters were chosen according to previous results (Purfield *et al.*, 2012; Forutan *et al.*, 2018; Meyermans *et al.*, 2020). Descriptive statistics of the detected ROH were calculated for each breed. A segment-based inbreeding coefficient (F_{ROH}) was calculated for each animal as the proportion of the autosomal genome covered by ROH.

Region-specific homozygosity and ROH islands. A coefficient of homozygosity ($F_{ROH.i}$) was calculated for each i^{th} marker as the proportion of animals (within breed) for which the marker is within a ROH segment. Afterwards, we identified markers that fell at or above the 99.5th percentile for $F_{ROH.i}$ and joined significant adjacent markers into larger segments, henceforth identified as ROH islands.

Annotation of genes and QTLs located within ROH islands. Annotation of genes located within the identified ROH islands was done using the 'btaurus_gene_ensembl' (ARS-UCD1.2 genome assembly) database with the 'biomaRt' R package (v.2.48.3) (Durinck *et al.*, 2009). We utilized the GALLO R package (Fonseca *et al.*, 2020) to query the Animal QTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/index) for quantitative trait loci (QTL) that have been previously identified in the regions of interest. Furthermore, a trait enrichment analysis was performed on the annotated QTLs to identify traits whose presence is not due to chance (FDR adjusted *P*-value<0.05).

Results & discussion

Descriptive statistics of detected ROH per individual are presented in Table 1. The average number of ROH ranged from 39.20 for AY to 65.87 for JE, the average length ranged from 6.48 Mb in GU to 7.31 Mb in BS, and the average combined length ranged from 268.39 Mb in AY to 431.77 Mb in JE. The inbreeding followed a similar trend with AY having the lowest average inbreeding (0.11) and JE having highest average inbreeding (0.17).

Region-specific homozygosity for every marker across the genome for the five breeds is presented in Figure 1. We found significant markers in chromosomes 4, 5,6, and 8 for AY; 5, 6, and 16 for BS; 6, 11, 13, 15, and 19 for GU; 10 and 20 for HO; and 2, 3, 7, and 20 for JE. ROH islands ranged from 0.08 to 7.83 Mb in length. We found two overlaps between ROH islands in different breeds, a 3.44 Mb overlap in chromosome 6 (BTA6: 86.76-90.1) between AY and BS, and a 2.57 Mb overlap in chromosome 20 (BTA20: 25.61-28.17) between HO and JE. We found a total of 26 genes shared between AY and BS, most notably *ADAM metallopeptidase with thrombospondin type 1 motif 3* (ADAMTS3). The ADAMTS3 gene has been linked to longevity (Mészáros *et al.*, 2014), embryonic development (Janssen *et al.*, 2016), and milk and milk component yield (Costa *et al.*, 2019) in cattle. The region that overlapped between HO and JE contained 5 genes, most notably *ISL LIM homeobox 1* (ISL1) and *pelota mRNA surveillance and ribosome rescue factor* (PELO). Both of these genes have been associated with spermatogenesis (Fernández *et al.*, 2014). The most

Table 1. Descriptive statistics of detected ROH per individual and average F_{ROH} in 5 breeds.[1]

Breed[2]	Avg. n.	Avg. length (Mb)	Avg. combined length (Mb)[3]	F_{ROH}
AY	39.20 (8.47)	6.78 (1.36)	268.39 (85.11)	0.11 (0.03)
BS	49.58 (8.72)	7.31 (1.08)	363.55 (87.12)	0.15 (0.04)
GU	57.39 (7.71)	6.48 (1.03)	371.99 (79.22)	0.15 (0.03)
HO	42.03 (8.66)	7.24 (1.17)	305.88 (87.68)	0.12 (0.04)
JE	65.87 (8.52)	6.55 (0.90)	431.77 (83.49)	0.17 (0.03)

[1] Standard deviation shown within parenthesis.

[2] AY= Ayrshire; BS= Brown Swiss; GU=Guernsey; HO= Holstein; JE= Jersey.

[3] Average combined length of genome in ROH.

gene-dense island was found for Jersey cattle in chromosome 7 (BTA7: 39.76-45.56), where we identified a total of 167 genes. A notable gene found in this region was *NLR family pyrin domain containing 3* (NLRP3), this gene has been previously associated with resistance to *Mycobacterium avium* ssp. paratuberculosis (MAP) infection (Mallikarjunappa *et al.*, 2018).

The top 10 (based on *P*-value) enriched traits from the annotated QTLs are presented for each breed in Figure 2. Milk production traits were found to be enriched in all five breeds, including milk or milk protein yield in AY, BS, HO, and JE. Other traits that were found enriched in two or more breeds included productive life, immunoglobin G level, clinical mastitis, and ketosis. This highlights the high selective pressure that has been put on traits that increase the health, reproductive ability, and length of productive life of U.S. dairy cows. Interestingly, 7 out of the top 10 enriched traits in the Guernsey breed related to meat and carcass traits, mainly content of saturated and monounsaturated fatty acids.

In summary, the differences observed in the genomic patterns of homozygosity suggest substantial heterogeneity in the management and selection strategies implemented in the different populations of U.S. dairy cattle.

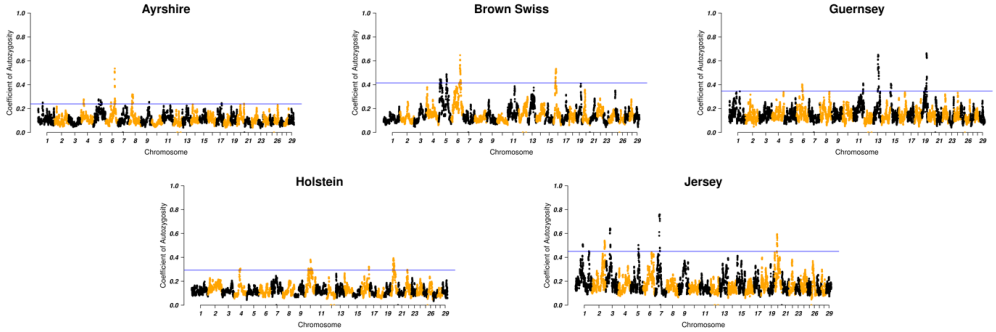

Figure 1. Manhattan plot of region-specific homozygosity ($F_{ROH,i}$) for Ayrshire (A), Brown Swiss (B), Guernsey (C), Holstein (D), and Jersey (E) cattle.

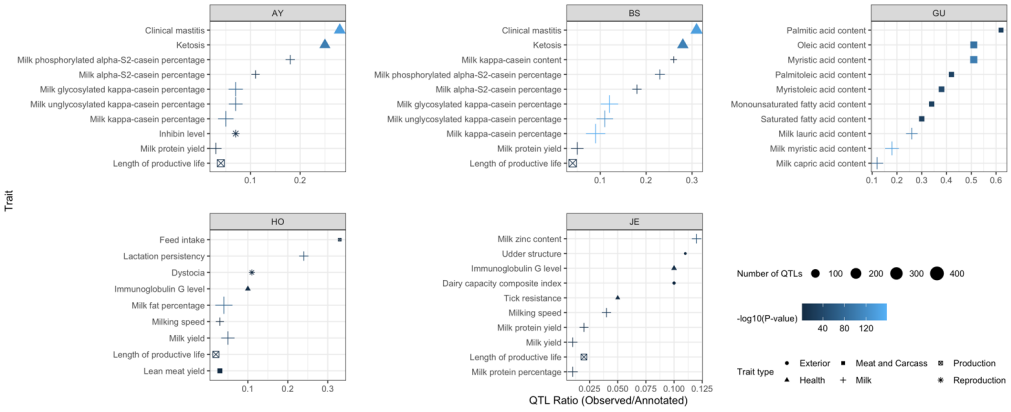

Figure 2. Top 10 enriched traits for annotated QTL's mapped to ROH islands in 5 breeds.

References

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., *et al.* (2015) Gigascience 4:7. https://doi.org/10.1186/S13742-015-0047-8.

Costa A., Schwarzenbacher H., Mészáros G., Fuerst-Waltl B., Fuerst C., *et al.* (2019) J. Dairy Sci. 102:10088-10099. https://doi.org/10.3168/JDS.2019-16663.

Durick S., Spellman P.T., Birney E., and Huber W. (2009) Nat. Protoc. 4:1184-1191. https://doi.org/10.1038/nprot.2009.97.

Fernández M.E., Lirón J.P., Prando A., Rogberg-Muñoz A., Peral-García P., *et al.* (2014) Livest. Sci. 167:9–18. https://doi.org/10.1016/J.LIVSCI.2014.05.009.

Fonseca P.A.S., Suárez-Vega A., Marras G., and Cánovas Á. (2020) Gigascience 9. https://doi.org/10.1093/GIGASCIENCE/GIAA149.

Forutan M., Ansari Mahyari S., Baes C., Melzer N., Schenkel F.S., *et al.* (2018) BMC Genomics 19:98. https://doi.org/10.1186/s12864-018-4453-z.

Gibson J., Morton N.E., and Collins A. (2006) Hum. Mol. Genet. 15:789–795. https://doi.org/10.1093/HMG/DDI493.

Howard J.T., Maltecca C., Haile-Mariam M., Hayes B.J, and Pryce J.E. (2015) BMC Genomics 16:1–13. https://doi.org/10.1186/s12864-015-1352-4.

Janssen L., Dupont L., Bekhouche M., Noel A., Leduc C., *et al.* (2016) Angiogenesis 19:53-65. https://doi.org/10.1007/S10456-015-9488-Z.

Makanjuola B.O., Miglior F., Abdalla E.A., Maltecca C., Schenkel F.S., *et al.* (2020) J. Dairy Sci. 103:5183–5199. https://doi.org/10.3168/jds.2019-18013.

Mallikarjunappa S., Sargolzaei M., Brito L.F., Meade K.G., Karrow N.A., *et al.* (2018) J. Dairy Sci. 101:7280–7286. https://doi.org/10.3168/JDS.2018-14388.

Mészáros G., Eaglen S., Waldmann P., and Sölkner J. (2014) Open J. Genet. 2014:46–55. https://doi.org/10.4236/OJGEN.2014.41007.

Meyermans R., Gorssen W., Buys N., and Janssens S. (2020) BMC Genomics 21:1-14. https://doi.org/10.1186/S12864-020-6463-X.

Peripolli E., Munari D.P., Silva M.V.G.B, Lima A.L.F., Irgang R., *et al.* (2017) Anim. Genet. 48:255–271. https://doi.org/10.1111/AGE.12526.

Purfield D.C., Berry D.P., McParlan S., and Bradley D.G. (2012) BMC Genet. 13:1-11 https://doi.org/10.1186/1471-2156-13-70.