Livestock Science 188 (2016) 72–80

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland

Tomasz Szmatoła*, Artur Gurgul, Katarzyna Ropka-Molik, Igor Jasielczuk, Tomasz Ząbek, Monika Bugno-Poniewierska

National Research Institute of Animal Production, Department of Genomics and Animal Molecular Biology, Krakowska 1, 32-083 Balice, Poland

ARTICLE INFO

Article history:
Received 6 November 2015
Received in revised form
31 March 2016
Accepted 8 April 2016

Keywords:
Runs of homozygosity
Autozygosity
Selection signatures

ABSTRACT

Runs of homozygosity (ROH) are defined as contiguous homozygous regions of the genome where the two haplotypes inherited from the parents are identical. It has been shown that the length and frequency of ROH may describe the history of the population in which an individual occurs; they may also reveal the level of inbreeding within populations, recent population bottlenecks or signatures of positive selection. In this study, BovineSNP50 whole-genome genotyping assay was used to analyse the lengths and distributions of the ROH found in the genomes of four cattle breeds maintained in Poland (Holstein, Polish Red, Limousin and Simmental) to assess both the level of autozygosity of each breed and to identify the genomic regions most commonly associated with ROH that may reflect directional selection pressure. Visible differences in the length and distribution of homozygous regions across the genome between selected breeds were observed. The breeds also varied in the level of autozygosity (inbreeding) estimated by F_{ROH}, which was lower for unselected cattle. Moreover, within the regions of the genome most commonly associated with ROH that may reveal signatures of recent selection a number of genes potentially connected with different production features characteristic for individual breeds were detected.

1. Introduction

Runs of homozygosity (ROH) are contiguous homozygous segments of the genome where the two haplotypes inherited from the parents are identical. These haplotypes are most likely identical because the parents inherited them from a common ancestor (Purfield et al., 2012; Curik et al., 2014; Gurgul et al., 2014). Broman and Weber (1999) originally described long homozygous segments of the genome in human populations using microsatellite markers. These analyses have shown that ROH are not only common, but that they may also have an effect on gene mapping and health. Currently, homozygous segments of the genome can be easily identified using information obtained from high-density SNP (single nucleotide polymorphism) arrays encompassing genome-wide marker panels. In humans, a deep research about runs of homozygosity has been conducted and now ROH are considered as a valuable method of identifying susceptibility of an individual to recessive diseases (Gibson et al., 2006; Lencz et al., 2007; McQuillan et al., 2008; Hildebrandt et al., 2009; Kirin et al., 2010). In animal genetics, runs of homozygosity may be utilized as a measure of inbreeding by estimating the level of autozygosity of the genome (Ferenčaković et al., 2013b; Curik et al., 2014) and to identify regions having an unfavourable effect on a phenotype when being homozygous (Pryce et al., 2014; Saura et al., 2015).

It has been shown that the length and frequency of ROH may help to describe the history of the population in which an individual occurs and the history of that individual's ancestors (Kirin et al., 2010; Purfield et al., 2012; Curik et al., 2014). It has also been shown, that long ROH can be used to identify consanguinity (Kirin et al., 2010). In particular, recent inbreeding resulting from the mating of closely-related ancestors leads to a high occurrence of long ROH. On the other hand, very long ROH sometimes occur in outbred populations (Gibson et al., 2006). Shorter ROH appear because chromosomal segments are broken up by repeated meiosis and have older origins (Kirin et al., 2010). The fact that ROH are mainly autozygous, makes them a useful tool to estimate inbreeding, by the estimation of genome portion covered by ROH (F_{ROH}) (Broman and Weber, 1999; McQuillan et al., 2008).

The homozygous segments of the genome, with reduced variability within a population, may also arise as a result of strong artificial selection which leads to the fixation of favourable alleles in the population in a sort of "hitchhiking" process (Ron et al., 1996; McQuillan et al., 2008; Hildebrandt et al., 2009). The

* Corresponding author.
E-mail address: tomasz.szmatola@izoo.krakow.pl (T. Szmatoła).

http://dx.doi.org/10.1016/j.livsci.2016.04.006
1871-1413/

identification of such regions which display reduction or elimination of polymorphism (selective sweeps) across population may indicate traces of recent selection and help to understand biological processes behind the selected traits (Smith et al., 1974; Carothers et al., 2006). Analysis of allele frequency spectra as well as extended homozygosity across the genome were shown to be an effective method for identification of selective sweeps (Smith et al., 1974; Sabeti et al., 2002). In this context, the size and position of runs of homozygosity are expected to correlate with direction of selection and suggest that ROH should be non-randomly distributed across the genome (Stella et al., 2010). The availability of modern genome scan technologies such as high density SNP arrays directed to the bovine genome has provided an opportunity for investigating ROH regions in various breeds. This leads to possibilities of comparing the extent and patterns of homozygosity between different populations.

The aim of this study was to present the characteristics of ROH in four different cattle breeds (Holstein, Polish Red, Limousin and Simmental) with different breeding schemes and production specializations. Moreover, it attempts to identify the regions of the genome with high ROH frequencies, namely ROH islands, that may have occurred due to selection directed on QTLs for functionally important traits. We also compared F_{ROH} statistics for this breeds, to evaluate the differences in the level of inbreeding between the studied populations.

2. Methods

2.1. Samples and genotyping

The study was performed on 1387 (462 females and 925 males) randomly selected animals belonging to four breeds: 853 Holstein, 255 Polish Red, 201 Limousin and 78 Simmental (hereafter referred as HO, RP, LM and SM, respectively).

Genomic DNA was extracted from the semen or whole blood of the animals using Sherlock AX (A&A Biotechnology) or QuickGene DNA Whole Blood S (Kurabo) kits. After quality control, the DNA was normalised to the required concentration and genotyped using the Illumina BovineSNP50 v2 BeadChip assay (Illumina Inc., San Diego, CA) in accordance with the standard Infinium Ultra protocol. After genotyping, the results were assessed for quality and only samples with call rates above 0.98 were used in further analyses.

To asses genetic diversity within cattle breeds, expected and observed heterozygosity, as well as F_{IS} statistic were calculated for all breeds with the use of Plink software (Purcell et al., 2007).

2.2. Data filtering and ROH identification

Data filtration and ROH identification was performed separately for each breed. Before quality pruning, the dataset comprised 54,609 SNPs. For further analysis only SNPs with a genotyping rate higher than 90%, GenCall ≥ 0.7 and GenTrain ≥ 0.4 were used. Finally, SNPs on chromosome X, Y or without determined position were also removed.

ROH were defined for each animal individually using cgaTOH software (Zhang et al., 2013). The following parameters were used for ROH identification: minimum number of 15 consecutive homozygous SNP in ROH with a maximum gap of 1 Mb. The detected ROH were assigned to five length categories: 1–2 Mb, 2–4 Mb, 4–8 Mb, 8–16 Mb and above 16 Mb. To calculate how many heterozygotes should be allowed for each ROH length category we initially identified ROH without allowing any heterozygotes. This enabled us to calculate the number of SNPs falling into each of ROH length categories so that we could estimate the number of heterozygotes permitted in each category. Assuming a genotype error of 0.2% (Howrigan et al., 2011) 0 heterozygotes were permitted for first four ROH categories (1–2 Mb, 2–4 Mb, 4–8 Mb, 8–16 Mb) and 1 heterozygote for the last category of ROH with size over the 16 Mb. The detailed number of SNPs falling into each selected ROH length category are presented in Supplementary File 1: Table S1. The number of missing genotypes allowed in the ROH was calculated according to Ferenčaković et al. (2013a) and the following thresholds were applied: for the 1–2 Mb and 2–4 Mb categories no missing SNPs, for the 4–8 Mb category one missing SNP, 8–16 Mb category two missing SNPs, >16 Mb category four missing SNPs. Mean sums of the ROH lengths within selected length categories were calculated by summing all ROH per each animal in every category and averaged within breed.

Autozygosity (inbreeding coefficient, F_{ROH}) was calculated according to McQuillan et al. (2008) by dividing a total length of all individual's ROH in a selected ROH length category by the length of the autosomal chromosomes covered by SNPs. The ROH length categories were as follows: above 1 Mb, above 2 Mb, above 4 Mb, above 8 Mb and above 16 Mb.

2.3. Identification of the genomic regions most commonly associated with ROH within the breeds

To identify the genomic regions that were most commonly associated with ROH within each breed, we analysed files generated using cgaTOH software (Zhang et al., 2013) that specified how many times each SNP appeared in ROH (for all ROHs above 1 Mb). These values were divided by the number of animals in each breed to calculate the proportion of SNP occurrence in ROH for each breed. The genomic regions most commonly associated with ROH were identified by selecting the top 1% of the SNPs most commonly observed in ROH in each breed. The adjacent SNPs with proportion of ROH occurrences over the adopted threshold formed long genomic regions (called ROH islands). To maintain continuity of the regions we allowed two inconsecutive SNP with a value just below the adopted threshold within region. Subsequently, the regions were screened for overlapping RefSeq genes using the UCSC Genome Browser (Rosenbloom et al., 2015) and the UMD3.1 cattle genome assembly. The genes were further analysed with the Panther Classification System (Mi et al., 2013) to identify their molecular functions and associated biological processes. Moreover, genomic locations of known strong QTLs for production traits in HO cattle like *DGAT1, FABP5, GHR* and *FASN* were screened to evaluate ROH patterns at these specific locations (control genes) (Thaller et al., 2003; Ogorevc et al., 2009).

Assuming that some of the detected ROH islands arose as a result of strong artificial selection, we additionally screened three ROH islands observed in at least three of the analysed breeds, to evaluate whether they carry a shared and common ROH variant (haplotype) segregating across populations and potentially associated with a beneficial allele under selection. This analysis was based on the regions located on chromosome: 7, 14 (HO, LM, RP) and 16 (LM, RP, SM). To account for complicated ROH structure at a specific ROH island, we narrowed down the screened areas to "core regions" formed by ROHs overlapping in all animals with ROH at a specific locus (Supplementary File 2). Non-overlapping (across population) ROH ends were truncated. At a core region, the frequency of each ROH variant was calculated and only ROHs with frequency higher than 5% were taken into consideration. Frequent ROHs with single heterozygotes (allowed in the ROH detection process) were treated as a separate allelic variants, due to the uncertainty associated with allocation of this variant to different similar haplotypes present in population. Analogous analysis was performed at an evident ROH island detected in the LM breed near position of the *MSTN* gene on BTA2, which is a known strong QTL