Livestock Science 185 (2016) 34–42

Contents lists available at ScienceDirect

Livestock Science

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

Intra-chromosomal estimates of inbreeding and coancestry in the Spanish Holstein cattle population

D. Kleinman-Ruiz [a], B. Villanueva [a], J. Fernández [a], M.A. Toro [b], L.A. García-Cortés [a], S.T. Rodríguez-Ramilo [a,*]

[a] Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28040 Madrid, Spain
[b] Departamento de Producción Animal, Escuela Técnica Superior de Ingenieros Agrónomos, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 18 September 2015
Received in revised form
17 December 2015
Accepted 5 January 2016

Keywords:
Genome-wide information
Holstein
Coancestry
Inbreeding
Genomic regions

ABSTRACT

In recent years, inbreeding and coancestry are being estimated from genome-wide molecular information using a large number of SNPs. Molecular inbreeding and coancestry can be calculated for the whole genome or for particular regions of the genome. In this study, genome-based inbreeding and coancestry were estimated per chromosome and at intra-chromosomal level in a group of Holstein animals genotyped with the Illumina BovineSNP50 BeadChip. After applying filtering criteria, the genomic dataset included 36,693 autosomal SNPs and 10,569 animals. Genome-based inbreeding and coancestry at intra-chromosomal level were calculated using sliding windows of approximately 5 Mb. The results showed differential patterns of inbreeding and coancestry on specific chromosome regions. These patterns provide a more detailed picture of genetic diversity that could be used, for example, for the detection of regions with low levels of genetic diversity that require a specific genetic management in conservation programmes.

1. Introduction

Over the last decades, inbreeding and coancestry in livestock populations have been alternatively measured using pedigree information or microsatellite data when pedigree information was not reliable or available. Currently, the availability of high-density SNP chips allows a more detailed evaluation of inbreeding and coancestry based on genomic estimates. The reasons why these measures are expected to be more accurate than the pedigree-based ones are basically as follows. First, genomic estimates reflect the percentage of homozygous positions given the genomic information, or the realised pairwise coancestry, while pedigree based estimates are just the expectations of these measures. Second, genomic estimates can detect relationships due to very distant common ancestors, which pedigree-based estimates do not take into account (Keller et al., 2011). An additional advantage of genome-based estimates is that they can be calculated for particular regions of the genome.

Published studies have shown that the variation in genetic diversity across regions could be substantially large (e.g. The International SNP Map Working Group, 2001; Engelsma et al., 2012; Esteve-Codina et al., 2013). The causes for these patterns are diverse. For example, it is well known that changes in genotype frequencies caused by selection affect the frequencies of neutral variants and other genetically linked sites in the genome, according to the theory of 'hitch-hiking' (Maynard-Smith and Haigh, 1974). As the favourable (unfavourable) allele in the selected locus increases (decreases) its frequency in a population, in the surrounding loci a parallel increase in frequency of one allele can also be observed due to the linkage between the selected locus and the rest. Consequently, there will be a greater loss of genetic diversity in regions harbouring selective loci and higher levels of inbreeding and coancestry.

In recent times, due to the rapid development of molecular resources, large numbers of SNPs have been used for the implementation of genome-wide evaluations (Meuwissen et al., 2001) in many commercial breeding programmes, including Holstein populations (VanRaden and Cooper, 2011). Despite the higher accuracy of genome-wide evaluations (which could lead to more intense selection and a faster increase in inbreeding), Daetwyler et al. (2007) showed that the global levels of diversity maintained per generation are higher when performing genome-wide selection than when using conventional BLUP based on pedigree relationships. However, Heidaritabar et al. (2014) showed that genome-wide selection applies selection pressure much more locally than BLUP, resulting in larger allele frequency changes in the selected loci and the linked ones. Therefore, the loss of genetic diversity on specific genomic regions associated to selection processes may be larger when genome-wide selection is implemented.

* Corresponding author.
E-mail address: rodriguez.silvia@inia.es (S.T. Rodríguez-Ramilo).

http://dx.doi.org/10.1016/j.livsci.2016.01.002
1871-1413/

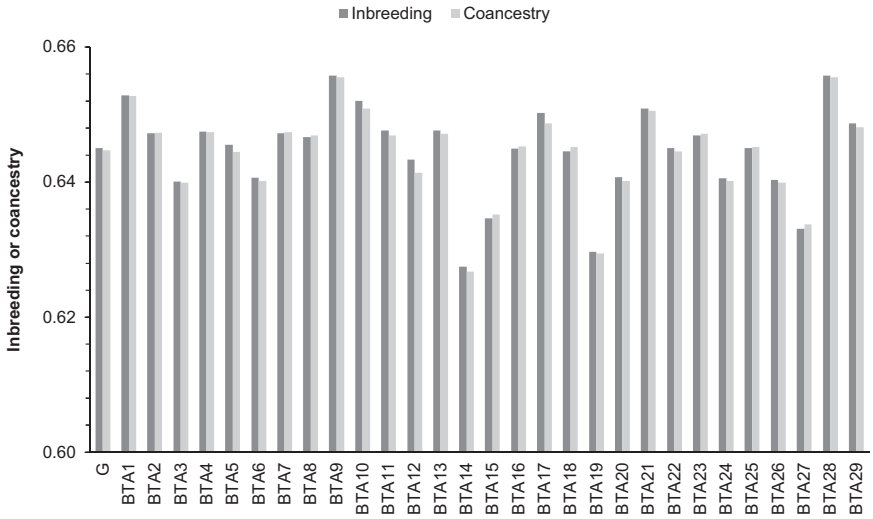

Fig. 1. Averaged inbreeding and coancestry across the genome (G) and for each autosome (BTA1-BTA29).

On the other hand, genomic regions harbouring loci involved in general disease resistance (such the MHC, where a high level of genetic diversity ensures that the population can deal with potential new disease challenges) may show low levels of inbreeding and coancestry (Birch et al., 2006). Both types of genomic regions could be essential for the potential adaptation and survival of any population.

Moreover, even the action of genetic drift may lead to some genomic regions having less genetic variation than others. Therefore, it may be arguable that regions with lower levels of diversity should be detected and specific management implemented to control diversity on such regions (Gómez-Romano et al., 2014). Genome-wide inbreeding and coancestry could be useful to detect such variation across the genome.

The objective of this study was to evaluate the variation of inbreeding and coancestry based on SNP marker information over the whole genome, at the chromosomal level and within chromosomes. For this purpose we used data from individuals belonging to the Spanish Holstein population and genotyped for the Illumina BovineSNP50 BeadChip.

2. Materials and methods

2.1. Genomic data

Genomic data used in this study were the same as in Rodríguez-Ramilo et al. (2015). Genomic information from 11,135 animals belonging to the Spanish Holstein population was analysed. These individuals were genotyped with the Illumina BovineSNP50 BeadChip (versions v1 or v2). Only SNPs common to both chip versions were selected for the analysis (52,340 SNPs). SNP positions within the genome were obtained from the UMD 3.0 bovine genome assembly. Unmapped SNPs (523) and those mapped on chromosomes X or Y (1056) were excluded. In addition, 14,068 SNPs with missing genotypes for more than 5% of the individuals were discarded. After that, 566 animals with more than 5% missing genotypes for the remaining 36,693 SNPs were also removed. The final dataset included 36,693 autosomal SNPs and 10,569 animals (9990 bulls and 579 cows).

2.2. Estimates of coancestry and inbreeding coefficients

The molecular coancestry coefficient between individuals i and j (f_{ij}) at a given locus can be defined as the probability that two alleles taken at random from each individual are alike in state. When dealing with several loci, it is the average across loci. In this study, f_{ij} was calculated following Nejati-Javaremi et al. (1997) as

$$f_{ij} = (1/n_s) \sum_{s=1}^{n_s} \left[\left(\sum_{k=1}^{2} \sum_{m=1}^{2} I_{S_{k_i m_j}} \right) /4 \right],$$

where n_s is the number of SNPs used and $I_{S_{k_i m_j}}$ is the identity of the kth allele from individual i with the mth allele from individual j at SNP s, and takes a value of 1 if both alleles are identical and 0 otherwise. For a single locus, the molecular inbreeding coefficient of individual i (F_i) is 1 if the individual is homozygous for this locus and 0 if it is heterozygous. For a group of loci, F_i is the proportion of homozygous loci. The inbreeding coefficient of individual i was calculated in this study as $F_i = 2f_{ii} - 1$.

Chromosomal estimates of inbreeding and coancestry were obtained using all the SNPs genotyped on each chromosome after applying filtering criteria. The number of genotyped SNPs was different between chromosomes, however the density of SNPs was very similar across chromosomes (data not shown).

Within chromosomes, estimates of inbreeding and coancestry were also obtained for different regions. Instead of calculating coefficients for individual SNPs, inbreeding and coancestry were estimated over sliding windows with a window size of approximately 5 Mb. Following this approach, the noise of single-locus estimates can be reduced by combining data from several adjacent markers. This procedure was based on the method used by Weir et al. (2005) and Engelsma et al. (2012). For each chromosome, the first window was identified by taking the SNPs at the first 5 Mb of the chromosome. Subsequently, the window moves along the genetic map in single-SNP increments, until the end of the chromosome is reached. In each window, the same number of SNPs for that specific chromosome has been maintained. The average number of SNPs in each window across chromosomes was 70.76 ± 3.00. It must be pointed out that, in that way, window size will not always be exactly 5 Mb. For each window, inbreeding and coancestry were estimated by averaging the values for all SNPs lying in that window. Afterwards, values were averaged over animals (or pairs of animals). A total of 10,569 animals were used for inbreeding and $(10,569 \times 10,568)/2$ pairs for coancestry estimates.

To detect regions showing significantly different levels of inbreeding and coancestry, their frequency distribution was plotted for each chromosome. As observed by Weir et al. (2005) for F_{ST} values across the genome, if it is assumed that differences in coancestry