COBISS: 1.08
Agris category code: L10

MAPPING OF HETEROZYGOSITY RICH REGIONS IN AUSTRIAN PINZGAUER CATTLE

Maja FERENČAKOVIĆ [1, 2], Maja BANADINOVIĆ [3], Mario MERCVAJLER [4], Negar KHAYATZADEH [5], Gábor MÉSZÁROS [6], Vlatka CUBRIC-CURIK [7], Ino CURIK [8], Johann SÖLKNER [9]

ABSTRACT

Heterozygosity, the state of possessing different alleles at a given locus of an individual, is functionally related to inbreeding, heterosis and biodiversity. We questioned the appearance of regions with extraordinary high rates of heterozygosity, here "Heterozygosity Rich Regions" (HRR) in the genomes of a cattle population. We used 120 Pinzgauer bulls genotyped with 611102 SNPs and detected 14702 HRR unequally dispersed in the genome. Mean coverage of SNP chip data with HRR was 0.99 %. In total we found 11 regions with high frequency of SNPs being in HRR on nine chromosomes yielding 21 genes of which 17 have described functions. We further identified genes located in HRR and discussed their importance and function. The results of this study point to the analysis of HRR providing additional understanding of the genomes of livestock.

Key words: cattle, Pinzgauer, heterozygosity rich regions, SNP data

1 INTRODUCTION

Heterozygosity, the state of possessing different alleles at a given locus of an individual, is functionally related to inbreeding, heterosis and biodiversity. Balancing selection is a common term for three types of selection (heterozygote advantage, negative frequency dependent selection, or fluctuating selection) that maintain higher than expected levels of heterozygosity and allelic diversity within populations. Heterotic balancing selection is caused by selective advantage of heterozygous genotypes showing overdominance. The existence of the overdominance has been proved empirically, but its occurrence is generally considered as a rare phenomenon. However, such genes have been identified for traits that are "multiplicatively" determined (Gemmell and Slate, 2006; Krieger *et al.*, 2010) and might be present more commonly. The other explanation for the maintenance of polymorphism is negative frequency dependent selection, also type of balancing selection, with the mechanism explained for MHC inheritance (Hedrick and Thompson, 1983; Hedrick, 1994. In fluctuating selection, the last type of balancing selection, polymorphism is maintained in a population by fluctuation of the selective pressure in a relatively short time (Bell, 2010). A good example of fluctuating selection in *Cepaea* is shown in Cain *et al.* (1990).

Furthermore, the relationship between genetic diversity and fitness is an important issue in different areas

1 Department of Animal Science, Faculty of Agriculture, University of Zagreb, Svetosimunska 25, 10000 Zagreb, Croatia
2 Corresponding author, e-mail: mferencakovic@agr.hr
3 Same address as 1, e-mail: majabanadinovic@hotmail.com
4 Same address as 1, e-mail: mario-215@hotmail.com
5 University of Natural Resources and Life Sciences Vienna, Department of Sustainable Agricultural Systems, Division of Livestock Sciences, Gregor Mendel Str. 33, A-1180 Vienna, Austria, e-mail: negar.khayatzadeh@students.boku.ac.at
6 Same address as 5, e-mail: gabor.meszaros@boku.ac.at
7 Same address as 1, e-mail: vcubric@agr.hr
8 Same address as 1, e-mail: icurik@agr.hr
9 Same address as 5, e-mail: johann.soelkner@boku.ac.at

of evolutionary biology (Charlesworth and Charlesworth, 1987; Ellegren and Sheldon, 2008). The importance of understanding the relationship between genetic diversity and fitness is in assessing the evolutionary potential of the population and to simultaneously predict the reduction of their genetic variability. In addition, heterozygosity-fitness correlations (HFC) have been used to study the relationship between genetic diversity and fitness at the individual level in a variety of organisms (Coltman and Slate, 2003). Most HFC studies in animal populations report a linear, positive relationship between measures of individual heterozygosity and fitness-related traits (Olano-Marin *et al.*, 2011).

In livestock population, besides the overall genome heterozygosity (Curik *et al.*, 2010, 2014), not a lot has been done in analyzing genomic aspects of heterozygosity such as existence of regions rich in heterozygosity. Williams *et al.* (2016) analyzed the heterozygosity in the Chillingham cattle, using the genotypes obtained through a set of single nucleotide polymorphisms (SNP chip), and confirmed the lack of variability in this extremely homozygous breed. Although Runs of Homozygosity (ROH) segments covered 95 % of the genome in this breed, they also found some regions that were strictly heterozygous and called them "Runs of SNP Heterozygosity". The authors consider that such regions, unlike ROH regions, could contain loci that contribute to the survival rate, fertility and other fitness traits (McParland *et al.*, 2009), and can be segments of the genome where diversity could be very beneficial. However, the term "Runs of Heterozygosity" is somewhat misleading. While for runs of homozygosity all base pairs between genotyped SNPs are considered to be homozygous, the non-genotyped base pairs between genotyped heterozygous calls are surely not all heterozygous.

The main goal of this study was to analyze genomic aspects of heterozygosity and frequency of HRR in Austrian Pinzgauer cattle as well as to identify genes, together with their functions, that are in HRR.

2 MATERIALS AND METHODS

2.1. QUALITY CONTROL OF GENOTYPES

A total of 121 Austrian Pinzgauer bulls were genotyped with Illumina BovineHD Genotyping BeadChip containing 777972 SNPs. DNA was isolated from the sperm obtained during the regular procedure of taking ejaculate in artificial insemination stations. Using SAS 9.4. software, we first excluded all SNPs that were not assigned to any chromosome and those assigned to sex chromosomes and mitochondrial DNA. In the next step, we removed SNPs with "GenTrain Score" smaller or equal to 0.4 and SNPs with "GenCall Score" smaller or equal to 0.7. Further we used PLINK v1.07. (Purcell *et al.*, 2007) to exclude all SNPs that were missing in more than 10 % of individuals and individuals that lacked more than 5 % of the SNPs.

2.2. ESTIMATES OF GENETIC PARAMETERS RELATED TO DIVERSITY

The following genetic parameters were estimated: (i) the number of polymorphic SNPs, (ii) the number of monomorphic SNPs, (iii) observed heterozygosity (HET_{OBS}) which was calculated as a proportion of homozygous individuals for each SNP and averaged over the individual chromosome or an entire genome, (iv) expected heterozygosity (HET_{EXP}) based on allele frequencies ($2pq$). We also estimated the inbreeding coefficient F_{IS} which is defined as 1-(HET_{OBS}/HET_{EXP}). This inbreeding coefficient is equivalent to Wright's (Wright, 1949) within-subpopulation fixation index with values in the range of −1 to +1. For the evaluation we used PLINK v1.07 and SNP &Variation Suite (v8.4.0 Win64; Golden Helix, Bozeman, MT, USA www.goldenhelix.com).

2.3. DETECTION OF HETEROZYGOSITY RICH REGIONS

HRR were detected using SNP &Variation Suite. For this purpose, we prepared a data set in which we replaced the status of each SNP and converted homozygous SNPs into heterozygous and vice versa in order to trick the algorithm for detection of ROH segments. HRR were defined as a sequence of at least 50 heterozygous SNPs in a row, where the minimum length of the HRR segment had to be at least 1kb. The density of SNPs had to be at least one SNP per every 50 kb. To account for genotyping errors, we allowed a maximum of two missing SNPs, four homozygous genotypes within HRR (Williams *et al.*, 2016) but these genotypes were not allowed to be in a row (Ferencakovic *et al.*, 2013). To detect the parts of genome in which the SNPs are often found in HRR we calculated the proportion of each SNP in a HRR in the total sample of animals. Then we chose 0.1 % of SNPs with the highest frequency and analyzed the regions in which they were located and checked whether we could find genes in them. The functions of the genes were taken from the online databases http://www.uniprot.org and http://www.genecards.org (last access 31.05.2016). We used genetic map UMD 3.1.1. (http://bovinegenome.org/?q=node/61) for the genome mapping.

3 RESULTS AND DISCUSSION

3.1. QUALITY CONTROL OF GENOTYPES AND ESTIMATES OF GENETIC PARAMETERS

After quality control, we were left with genotypes of 120 animals, having 611102 SNPs on autosomal chromosomes covering 2507812473 bp of the genome. The estimated total genetic parameters for this population were: (i) the number of polymorphic SNPs; 603076 (98.7 %) while the number of monomorphic SNPs was complementary (8026; 1.3 %), (ii) the average observed heterozygosity (HET_{OBS}); 0.346 (range: 0.320 – 0.363), (iii) expected heterozygosity (HET_{EXP}); 0.341 (range: 0.329–0.342). Estimated inbreeding coefficient F_{IS} was −0.0133 (range: −0.0638 to 0.0644). We also calculated the number of monomorphic and polymorphic SNPs and estimate of HET_{OBS} for each chromosome. The lowest HET_{OBS} was on chromosome 2 (0.316) while the highest HET_{OBS} was on chromosome 27 (0.365). Genetic parameters evaluated in this study show a high polymorphism rate of SNP markers (98.7 %) obtained by Illumina BovineHD Genotyping BeadChip. Chromosomal or total HET_{EXP} (0.34) values in Pinzgauer cattle were not deviating from those obtained in other breeds (0.25–0.34) (e.g. Williams *et al.*, 2014). Estimated inbreeding coefficient F_{IS} was found to be negative (mean: −0.0133, 95% confidence interval: −0.0177; −0.0087), suggesting the avoidance of close pedigree inbreeding. Negative values could also indicate potential problems in genotyping, but we think this was not the case as we have applied severe quality control. Furthermore, Ferenčaković *et al.* (2013) showed that, in comparison to other breeds, inbreeding level, F_{ROH} (1.4 to 6.2 %, depending on minimum ROH size) and F_{PED} (1.9 %) is low in this breed.

3.2. HETEROZYGOSITY RICH REGIONS IN THE GENOME OF THE PINZGAUER CATTLE

In the genome of Pinzgauer cattle, the total number of detected HRR was 14702. The largest region, regarding the length in base pairs, (1,386964 Mb) was located on chromosome 21, and the shortest (0,058072 Mb) was observed on chromosome 10. Regarding the number of SNPs, the largest segment contained 210 heterozygous SNPs in a row, and the shortest 50, which, was our default minimum. In comparison to the length of ROH segments found in this breed by Ferenčaković *et al.* (2013.), HRR were much smaller and rarer. The lowest coverage of the genome represented by SNP chip data with HRR was on chromosome 13 (0.28 %) while the highest coverage was observed on chromosome 5 (1.40 %). In total, the average coverage of the with HRR was 0.99 % (0.57–1.13 %). These results could not be compared with any other research since, for now, there has been only one published study with a similar research objective (Williams *et al.*, 2016).

3.3. PARTS OF THE GENOME WITH A HIGH PROPORTION OF SNP_S IN HETEROZYGOSITY RICH REGIONS

After we determined the threshold of 0.1 % SNPs with the highest frequency in HRR, there were 611 SNPs passing it. Those SNPs formed 11 regions on nine chromosomes (Table 1). Chromosomes 1, 3, 9, 11, 16, 18 and 19 had only one region, while chromosomes 2 and 6 had two regions. In these 11 regions, we found a total of 21 genes. The second region on chromosome 6 (Table 1) did not have recorded genes, nor did have regions on chromosomes 9 and 19. Of the 21 recorded genes, 17 had

Table 1: *Parts of the genome with the highest proportion of SNPs in HRR*

Chromosome	Beginning of the region (bp)	End of the region (bp)	Number of heterozygous SNP in the region
1	131553025	131702250	18
2	65395949	65574548	77
	90526660	90590242	48
3	54166354	54261102	23
6	7770842	7857073	16
	80607709	8072311	47
9	43960964	44119040	68
11	61932165	62057011	13
16	42625201	42840188	76
18	25753024	25855179	20
19	47269324	47452230	51

known and described function, while the others were annotated as coding genes, but without function.

Based on gene functions obtained on http://www.uniprot.org and http://www. genecards.org (last access 31.05.2016), we concluded that genes found in HRR are important in biological processes. Our premises are finding good example in 5 from 17 genes with known and well described function. Interesting was *ALS2CR11* gene located on chromosome 2, which, in humans, has a function related to juvenile amyotrophic lateral sclerosis. The disease has different symptoms of the better known amyotrophic lateral sclerosis (ALS) and its inheritance is recessive (http://www.malacards.org/card/amyotrophic_lateral_sclerosis_2_ juvenile, last access 31.05.2016.). There is also a group of genes on chromosome 3 (*F1N4W2*, *GBP6* and *GBP5*) whose role is associated with with binding and metabolism of guanosine triphosphate (GTP) in innate immune and inflammatory response. On chromosome 11 we found the gene *MDH1* encoding the enzyme malate dehydrogenase. Malate dehydrogenase catalyzes the oxidation of malate to oxaloacetate in the Krebs cycle.

Here we have presented only five of 17 genes with known and well described functions. Those were chosen as examples because their functions in important biological processes are familiar to broader audience. Their presence in HRR could indicate presence of balancing selection (VanRaden *et al.*, 2011), but such premise must be further investigated and confirmed.

4 CONCLUSIONS

This research represents a pilot study in which we identified HRR in the cattle genome as well as detected genes that are located in HRR. We speculate that appearance of HRR is a consequence or trace of the balancing selection. However, readers should be aware that further analyses of HRR pattern are needed as experimental evidence is scarce while theoretical explanation is missing. On the other side, results from this pilot study question the reasons for the HRR presence and indicate potential importance of HRR as a tool that will provide additional understanding of livestock genomics.

5 REFERENCES

Bell, G. (2010). Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 365*(1537), 87–97.

Cain, A. J., Cook, L. M., Currey, J. D. (1990). Population size and morph frequency in a long-term study of Cepaea nemoralis. *Proceedings of the Royal Society of London B: Biological Sciences, 240*(1298), 231–250.

Charlesworth, D., Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual review of ecology and systematics*, 237–268.

Coltman, D., Slate, J. (2003). Microsatellite measures of inbreeding: A meta-analysis. *Evolution, 57*(5), 971–983.

Curik, I., Ferenčaković, M., Gredler, B., and Sölkner, J. (2010). Genome-wide heterozygosity and pedigree inbreeding coefficients in Simmental cattle population. In *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*. Leipzig, Germany August 1–6

Curik, I., Ferenčaković, M., Sölkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livestock Science, 166*, 26–34.

Ellegren, H., Sheldon, B. C. (2008). Genetic basis of fitness differences in natural populations. *Nature, 452*(7184), 169–175.

Ferenčaković, M., Sölkner, J., Curik, I. (2013). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution, 45*(1), 42.

Gemmell, N. J., Slate, J. (2006). Heterozygote advantage for fecundity. *PLoS One*, *1*(1), e125. Retrieved from http://dx.doi.org/10.1371/journal.pone.0000125.

Hedrick, P. W., Thomson, G. (1983). Evidence for balancing selection at HLA. *Genetics, 104*(3), 449–456.

Hedrick, P. W. (1994). Evolutionary genetics of the major histocompatibility complex. *American Naturalist*, *143*(6), 945–964.

Krieger, U., Lippman, Z. B., Zamir, D. (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nature genetics, 42*(5), 459–463.

McParland, S., Kearney, F., Berry, D. P. (2009). Purging of inbreeding depression within the Irish Holstein-Friesian population. *Genetics Selection Evolution, 41*, 16.

Olano-Marin, J., Mueller, J. C., Kempenaers, B. (2011). Correlations between heterozygosity and reproductive success in the blue tit (Cyanistes caeruleus): an analysis of inbreeding and single locus effects. *Evolution, 65*(11), 3175–3194.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559–575.

VanRaden, P., Olson, K., Wiggans, G., Cole, J., Tooker, M. (2011). Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *Journal of Dairy Science, 94*(11), 5673–5682.

Williams, J., Hall, S., Del Corvo, M., Ballingall, K., Colli, L., Ajmone Marsan, P., Biscarini, F. (2016). Inbreeding and purging at the genomic Level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. *Animal Genetics, 47*(1), 19–27.

Wright, S. (1949). The genetical structure of populations. *Annals of eugenics, 15*(1), 323–354.