Characterization of Genetic Variation in Icelandic Cattle

L.-E. Holm,[#] A. Das,[#] J. Momeni,[#] F. Panitz,[#] C. Bendixen,[#] E. Eythorsdottir[¤]
[#]:Dept. of Molecular Biology & Genetics, Aarhus University, Denmark;
[¤]:Faculty of Land and Animal Resources, Agricultural University of Iceland, Iceland

Abstract: Identification of genetic variation in cattle breeds using next-generation sequencing technology has focused on the modern production cattle breeds. We focused on one of the oldest indigenous breeds, the Icelandic cattle breed. Sequencing of two individuals enabled identification of more than 8 million SNPs and more than one million short indels. Annotation of the genetic variants identified a substantial number of functional SNPs and variants. The number of genetic variants identified in the Icelandic cattle breed is on the same level as previously seen in other studies on Holstein cattle.

Keywords: Icelandic cattle**;** whole-genome resequencing; genetic variation; annotation.

INTRODUCTION

The recent technological advances and reduction of costs in next generation sequencing has enabled whole genome resequencing in individuals and not only for *denovo* sequencing in important species. A number of different breeds of cattle has been resequenced, and most notably has the 1000 bull genomes initiative with the aim of gathering a large number of individual sequences and making these available for the scientific community been initiated (Hayes et al. (2012)). These increasing numbers of studies describe the genetic variation present in numerous breeds by describing large numbers of SNPs and small indels. However, these studies concentrate on the description of the modern and commercially important breeds.

The modern breeds like Holstein are subject to intense breeding programmes resulting in a high level of selection on the breeds. This reduces the amount of genetic variation in the breeds and imposes a risk of loss of important genetic variants. It is therefore of interest to include old or indigenous breeds to the list of resequenced animals enabling description of additional genetic variants which could be interesting to reintroduce into the modern breeds.

Icelandic cattle is a cattle breed that has been isolated from other breeds for more than 1,000 years (Kantanen et al. (2000a)). The Icelandic breed is a phenotypic diverse breed which has not been standardized to any specific appearance (Kantanen et al. (2000b)). Icelandic cattle are probably closely related to some of the indigenous breeds from western Norway, in particular the Blacksided Troender and Nordland breed (Kantanen et al. (2000a)). There might have been small introductions from other breeds from Denmark in the 19th century and possibly some Jersey animals at one stage (Brænd et al. (1962)) but characterization of genetic variation in these breeds using microsatellites and blood types did not indicate a significant contribution from these breeds to Icelandic cattle (Kantanen et al. (2000a)). The Icelandic settlers coming from Norway probably stayed in the northern parts of the British Isles before continuing to Iceland and it is not possible to exclude a genetic influence from some of the indigenous British breeds in the original Icelandic cattle (Brænd et al. (1962)). However, these potential genetic contributions seem not to have had a marked genetic impact on the Icelandic cattle (Kantanen et al. (2000a); Brænd et al. (1962)).

In this study, we used whole-genome resequencing to characterize the genetic variation present in the Icelandic cattle breed.

METHODS

Animals. Animals were sampled from the Icelandic cattle breed in such a way to ensure they were unrelated (at least back to grandparents) based on herdbook information.

Sequencing. DNA was extracted from whole blood from two unrelated cows using a salt precipitation method (Miller et al. (1988)). Two sequencing libraries with insert sizes of 300 bp and 800 bp, respectively, were constructed for each individual according to the manufacturers instructions (Illumina, San Diego, CA, USA). Paired-end sequencing was performed on a HiSeq2000 (Illumina) using 101 cycle reads.

Table 1. Annotation of identified SNPs within coding regions in Icelandic cattle based on the genome annotation from ENSEMBL and UCSC.

Functional class	ENSEMBL	UCSC
Non-synonymous	15,229	22,443
Synonymous	21,323	27,145
Splice sites	1,346	979
Stop gain	114	269
Stop loss	17	35
Initiation codon	17	52
Unknown	22,660	8,818

Table 2. Annotation of identified indels within coding regions in Icelandic cattle based on the genome annotation from ENSEMBL and UCSC.

Functional class	ENSEMBL	UCSC
Within frame insertion	493	546
Within frame deletion	571	488
Frameshift insertion	307	255
Frameshift deletion	203	238
Splice site	204	93
Stop gain	15	12
Initiation codon	1	7
Unknown	640	647

Sequence mapping and annotation. The Burrows-Wheeler Aligner (BWA) (Alkan et al., 2009) was used for mapping towards the *Bos taurus* genome assembly UMD 3.1 (Zimin et al. (2009)). SNP and indel calling were performed using the UnifiedGenotyper procedure from the Genome Analysis Toolkit v.2.4.7 (GATK) (McKenna et al. (2010)) with options "--min_base_quality_score 20 -dcov 200" and keeping other parameters as default. SNPs and indels from dbSNP build 133 (Sayers et al. (2011)) were used as known sites. Annotation of functional SNPs and indels were performed using the Variant classification procedure in SVS vers. 8 (Golden Helix, Bozeman, MT, USA) using both the ENSEMBL and the UCSC annotation based on bosTau7.

RESULTS

The two individuals were both sequenced to more than 20x coverage. The GATK procedure identified 6,591,169 and 6,630,223 SNPs and 761,029 and 768,029 indels in the two individuals, respectively. The data shows a total of 8,790,026 SNPs and 1,038,881 indels in the two individuals combined. Approx. 73% of the SNPs were intergenic and 24% intronic, which is comparable to the results obtained by Stothard et al. (2011). Approx. 48% of the SNPs are annotated in dbSNP. The distribution of the annotated SNPs and indels on different functional classes are presented in Table 1 and Table 2. There is a marked difference between the two annotation databases with lower numbers of non-synonymous and synonymous variants identified using the ENSEMBL annotation while the number of variants with unknown effect is higher in the ENSEMBL annotation.

DISCUSSION

Here, we present a description of the amount of genetic variation present in the Icelandic cattle breed based on whole genome resequencing of two individuals. Previous studies on blood types and microsatellites revealed an apparent reduced level of variation in Icelandic cattle compared to Nordic cattle breeds showing lower heterozygosities and allelic diversity (Brænd et al. (1962); Kantanen et al. (1999, 2000a)). This could be caused by a founder effect with relatively few animals founding the population and/or random drift especially in periods of low effective population sizes resulting in loss of alleles. However, the identified number of SNPs and indels in this study is not markedly different from other studies of modern breeds (Stothard et al. (2011); Köks et al. (2013)) although direct comparisons are difficult due to different sequencers being used.

These data provide a solid backbone for analysis of genetic variation in chromosomal regions that has been reduced in variation in the modern breeds as a result of the intense selection in the breeding programmes. At the same time, these data are valuable for the description of the genetic variation present in an old breed that has been separate from the modern breeds for a very long period and therefore it is possible to use this breed as a basis for genetic comparisons towards the modern breeds.

CONCLUSION

Whole genome resequencing of two individuals from the Icelandic cattle breed revealed a substantial amount of genetic variation and provides a backbone for future studies into the genetic differences between this isolated cattle breed and the modern production breeds.

ACKNOWLEDGMENTS

We want to thank technician Mette Jeppesen for excellent work in the laboratory. This work was supported by the Danish Ministry of Food and Agriculture.

REFERENCES

Alkan, C., Kidd, J., Marques-Bonet, T. et al. (2009). Nat. Genet. 41:1061- 067.

Brænd, M., Rendel, J., Gahne, B. et al. (1962). Hereditas 48:264-283.

Kantanen J., Olsaker I., Adalsteinsson S., et al. (1999). Anim. Genet. 30:16-27.

Kantanen J., Olsaker I., Holm L.-E. et al. (2000a). J. Hered. 91:446-457.

Kantanen J, Olsaker I., Brusgaard K., et al. (2000b). Genet. Sel. Evol. 32:561-576.

Hayes B., Fries R., Lund M.S. et al. (2012). 1000 bull genomes consortium project. PAG XX, San Diego, US, Abstract W139.

Köks S., Lilleoja R., Reimann E. et al. (2013). Mamm. Genome 24:309-321.

McKenna, A., Hanna, M., Banks, E. et al. (2010). Genome Res.20(9):1297-1303.

Miller S.A., Dykes D.D. & Polesky H.F. (1988). Nucleic Acids Res. 16:1215.

Sayers, E., Barrett, T., Benson, D. et al. (2011). Nucleic Acids Res.39:D38-D51.

Stothard P., Choi J.-W., Basu U. et al. (2011). BMC Genomics 12:559.

Zimin, A., Delcher, A., Florea, L. et al. (2009). Genome Biol. 10(4):R42.