MS-thesis November 2008

Genetic variation within the Icelandic cattle breed

Assessment using microsatellites and analysis of single nucleotide polymorphisms in the *Leptin* and *DGAT1* genes

Margrét Guðrún Ásbjarnardóttir

Faculty of Land and Animal Resources

MS-thesis November 2008

Genetic variation within the Icelandic cattle breed

Assessment using microsatellites and analysis of single nucleotide polymorphisms in the *Leptin* and *DGAT1* genes

Margrét Guðrún Ásbjarnardóttir

Academic advisor: Jón Hallsteinn Hallsson

Agricultural University of Iceland
Faculty of Land and Animal Resources

Clarification of contribution

I hereby declare that the writing of the following thesis and the two accompanying manuscripts is my work, done under the supervision and with the assistance of my advisor Dr. Jón Hallsteinn Hallsson and the other two members of my masters committee, Emma Eyþórsdóttir and Dr. Magnús B. Jónsson.

Manuscript 1: In this manuscript we describe the analysis of genetic variation within the Icelandic cattle breed using microsatellite analysis. My contribution was to isolate genomic DNA from blood and organize samples for analysis. I was responsible for the processing of the data (analysed microsatellites). This included choosing diversity indices to be calculated, selecting the appropriate software programs, carrying out the calculation, as well as interpreting the results. Calculation of relationship indices was done in cooperation with Þorvaldur Kristjánsson. I wrote the accompanying manuscript in collaboration with Dr. Jón Hallsteinn Hallsson.

Manuscript 2: In this manuscript we describe the analysis of parts of two genes shown previously to influence economically important traits in cattle. My contribution was to prepare the samples, perform PCR amplification and the clean-up steps, as well as preparing samples for sequencing. I organized the samples, processed the sequencing results and interpreted the data. I wrote the accompanying manuscript in collaboration with Dr. Jón Hallsteinn Hallsson.

Unpublished results: A part of the processing of sequencing data was to perform an association study in the Icelandic cattle breed, exploring the connection between the C/T SNP located in exon 2 of the *Leptin* gene and two traits. Record data for the traits was provided by the Farmers Association of Iceland. Statistical analysis was done by me.

_____

Margrét Guðrún Ásbjarnardóttir

Abstract

Sufficient genetic variation within domestic breeds is a crucial factor in all breeding work. It is a prerequisite for future breeding progress and is used in research and development of breeding methods at the molecular level. The present study estimates genetic variation within the Icelandic cattle breed using two approaches. The data consisted of genomic DNA samples from 100 heifers in first pregnancy collected at 45 dairy farms located in the four major dairy regions in Iceland. First, the samples were genotyped with a set of eleven microsatellite markers and various diversity indices calculated. Second, parts of the *Leptin* and *DGAT1* genes were sequenced in order to search for single nucleotide polymorphisms (SNPs), (several in the *Leptin* gene and a well known dinucleotide substitution (K232A) in the *DGAT1* gene). Polymorphisms in these regions have been identified and associated with commercial traits in different dairy breeds and the purpose was to explore their frequency in Icelandic cattle, as well as to search for breed specific polymorphisms. The Icelandic cattle breed is the only cattle breed in Iceland and has been, more or less, isolated for over 1000 years; therefore it is considered a closed population.

Results from the microsatellite analysis revealed a mean number of observed alleles per locus of 6.182, ranging from 4 (ETH3) to 9 (TGLA53) for individual markers. Mean observed and expected heterozygosity were calculated as 0.626 and 0.685, respectively. Polymorphism information content was high (≥0.5) for all the markers indicating that they can be considered suitable for further research of the Icelandic cattle breed and could be adopted for breeding purposes and parentage testing. Average within population inbreeding coefficient of Icelandic cattle ranged between 8.8 and 9.7% for the three approaches applied and the effective number of individuals was estimated to be 111 individuals (lower and upper 95% confidence limits set as 99.76 and 127.39, respectively). No subdivision of the sample was observed and the breed is not likely to have experienced recent bottlenecks. Three out of the five previously known SNPs were identified in the *Leptin* gene of Icelandic cattle and a new polymorphism, not previously described in other cattle breeds, was found in intron 2. All heifers analysed were homozygous (carrying the GC/GC dinucleotide) for the K232A polymorphisms in the *DGAT1* gene. Together, these results indicate a substantial amount of genetic variation within the Icelandic cattle breed despite its long isolation and the progressive breeding strategies used for the population in later decades.

Ágrip

Allt kynbótastarf í búfjárstofnum byggist á erfðafjölbreytileika. Hann er grunnforsenda ræktunarframfara og nýtist ennfremur við rannsóknir og þróun kynbótaaðferða sem byggja á sameindaerfðafræði. Í rannsókn þessari var erfðafjölbreytileiki innan íslenska kúastofnsins metinn með tveimur aðferðum sem beitt var á safn erfðaefnis úr 100 íslenskum fyrsta kálfs kvígum. Sýnum var safnað á 45 kúabúum af fjórum megin nautgriparæktarsvæðum Íslands. Í fyrsta lagi voru sýnin greind með 11 örtunglum og greiningin síðan notuð til að reikna algenga breytileikastuðla fyrir stofninn. Í öðru lagi voru hlutar *Leptin* og *DGAT1* genanna raðgreindir í leit að einbasabreytileikum (SNP) (nokkrum slíkum í *Leptin* geninu og einum vel þekktum tvíbasabreytileika (K232A) í *DGAT1* geninu). Þessir breytileikar hafa verið tengdir hagnýtum eiginleikum í ýmsum öðrum mjólkurkúakynjum og tilgangur greininganna var að skoða tíðni þeirra í íslenska kúastofninum, ef til staðar, ásamt því að leita að öðrum breytileikastöðum sem einkennt gætu íslensku kúna. Íslenski kúastofninn er eina kúakynið á Íslandi og hefur verið einangrað kyn í meira en 1000 ár. Stofninn er þess vegna lokaður erfðahópur. Niðurstöður örtunglagreiningar sýndu að meðalfjöldi samsæta í hverju sæti (MNA) var 6,182 og lágu gildi fyrir einstök örtungl á bilinu 4 (ETH3) til 9 (TGLA53). Meðal fundin (H_O) og væntanleg (H_E) arfblendni var 0,626 og 0,685. Breytileikagildi örtunglanna, PIC-gildi, var hátt (≥0.5) fyrir öll örtunglin sem bendir til þess að þau henti til frekari rannsókna á íslenska kúakyninu og gætu nýst við ræktun og ætternisgreiningar. Meðal skyldleikaræktarstuðull innan stofnsins var reiknaður með þremur aðferðum á bilinu 8,8 – 9,7%. Virk stofnstærð var metin sem 111 einstaklingar (95% öryggismörk voru 99,76 – 127,39). Hvorki fundust merki um skiptingu sýnanna í sérstaka undirhópa né þess að stofninn hafi gengið í gegnum nýlega erfðafræðilega flöskuhálsa.
Í *Leptin* geninu fundust þrír af fimm áður þekktum einbasabreytileikastöðum. Einn nýr erfðabreytileiki, sem ekki hefur verið lýst í öðrum kúakynjum, fannst í innröð 2. Öll raðgreind sýni voru arfhrein (GC/GC) fyrir K232A breytileikann í *DGAT1* geninu. Samantekið benda þessar niðurstöður til þess að umtalsverður erfðabreytileiki sé til staðar í íslenska kúakyninu þrátt fyrir langa einangrun og þær framsæknu aðferðir sem beitt hefur verið í kynbótastarfinu í seinni tíð.

Acknowledgements

I wish to express my gratitude to my advisor Dr. Jón Hallsteinn Hallsson for great assistance, encouragement, and valuable guidance throughout this whole project.

I would like to thank my co-advisors, Emma Eyþórsdóttir and Dr. Magnús B. Jónsson for all their helpful comments on the manuscript and assistance during the work of this project. Magnús also receives gratitude for providing the blood samples, as do the veterinarians and farmers involved in the collection of the samples.

Thanks go to Gunnfríður E. Hreiðarsdóttir for providing useful information, assistance, and data.

Thanks go to Dr. Áslaug Helgadóttir for valuable comments on the manuscript.

Thanks go to Þorvaldur Kristjánsson for assisting with relationship calculation and pedigree related work of this project.
Being my fiancé as well, Þorvaldur also receives gratitude for his patience, support, and good humour when most needed.

I would also like to thank my son, Þorvaldur Örn, for showing a great deal of patience for work he does not yet understand the point of.

Thanks go to my family for support and encouragement.

Financial support for this study was provided by the following organizations, which hereby receive my gratitude: Icelandic Board of Genetic Resources in Agriculture, The Agricultural Productivity Fund, RANNIS, and Blikastaðasjóður (e. Blikastaðir Fund).

Table of content

Original papers included in the thesis

The following thesis is based on two original paper manuscripts, which are referred to by roman numerals.

I. Asbjarnardottir, M.G., Kristjansson, T., Jonsson, M.B. & Hallsson, J.H. (2008). Analysis of genetic variation within the Icelandic cattle population using molecular markers. To be submitted.

II. Asbjarnardottir, M.G. & Hallsson, J.H. (2008). Genetic analysis of the Icelandic cattle breed with respect to single nucleotide polymorphisms in the *Leptin* and *DGAT1* genes. To be submitted.

List of tables

List of figures

List of abbreviations

ADG: Average daily gain

AI: Artificial insemination

BLUP: Best linear unbiased prediction

BTA: Bos Taurus autosome

Chr: Chromosome number

$\hat{D}$: Gene diversity

DGAT1: AcylCoA:diacylglycerol acyltransferase 1

EM: Expectation-maximization

F: Inbreeding coefficient

FAO: Food and Agricultural Organization of the United Nations

F_{IS}: Within population inbreeding coefficient

FY: Fat yield

F%: Fat percentage

H_O: Observed heterozygosity

H_E: Expected heterozygosity

HWE: Hardy-Weinberg equilibrium

IAM: Infinite allele model

ISAG: International Society for Animal Genetics

MLE: Maximum likelihood estimation

MNA: Mean number of alleles

MY: Milk yield

N_e: Effective population size

N_O: Number of observed alleles

PCR: Polymerase chain reaction

PIC: Polymorphism information content

PY : Protein yield

P% : Protein percentage

QTL: Quantitative trait locus

SSC : Soluble solids content

SMM: Stepwise mutation model

SNP: Single nucleotide polymorphism

TPM: Two-phased model of mutation

1 Introduction

1.1 The importance of genetic variation

Genetic variation has been defined as the variation between and within breeds of a given species, influenced by interaction of different forces such as selection, genetic drift, mutation, and migration (Talle *et al*., 2005). Within domestic breeds it is the fuel for all breeding work, giving breeders the opportunity to improve the traits present or to develop new characteristics in response to changes in the environment, new diseases or market demands (Hill, 2000; Eyþórsdóttir, Tómasson & Helgadóttir, 2001; Maudet, Luikart & Taberlet, 2002; Talle *et al*., 2005).

Despite of the importance of genetic variation, global domestic animal diversity is declining substantially (Rischkowsky & Pilling, 2007). About 300 of the 6000 breeds of farm animals identified by FAO globally have become extinct over the past 15 years and genetic resources in cattle, sheep, and goats are considered to be highly at risk, in particular in developed countries were efficient selection methods successfully increase the production at the cost of genetic variability (Taberlet *et al*., 2007). This trend has been seen as a result of market extensions and economic globalization which calls for an increase in the use of fewer, high-output breeds, specialization in production along with decoupling of agriculture from the natural environment (Tisdell, 2003). Due to this, many native cattle breeds adapted to local environment have become endangered (Kantanen *et al*., 1999; Taberlet *et al*., 2007). In Europe, it is estimated that 171 cattle breeds have already become extinct and 122 are currently categorized as endangered or in a critical state (Scherf, 2000). As an example, several local Scandinavian cattle breeds, known for their phenotypic diversity, have been displaced by commercial breeds; either red-and-white Ayrshire or black-and-white Holstein-Friesian based stock (Kantanen *et al*., 2000b). Here it can be mentioned that some breeds have special qualities making them desirable for a special group of buyers which can lead to a niche market. Such markets can help to preserve original livestock breeds and add to the importance of genetic diversity among breeds (Tisdell, 2003; Eyþórsdóttir *et al*., 2001).

When trying to estimate the future breeding potential of a given livestock breed it is necessary to consider the level of genetic variation within the breed. Using genetic material for this purpose commonly involves genetic markers and the analysis is based on using the marker's degree of polymorphism as an indicator of the genetic variation level. Moreover,

the use of various genetic markers to accelerate genetic progress within a particular trait, the trait being quantitative or not, is referred to as marker-assisted selection and is of interest to breeders. For dairy breeding programmes, it has been estimated that markers linked to quantitative trait loci (QTL) might increase annual response by up to 30% (Simm, 2000). A more direct use of a marker involves exploiting certain polymorphic sites in protein coding genes that have been associated with altered protein function.

Considering all the factors above, the maintenance of genetic variation within domestic animals should be a major concern to breeders worldwide.
The study presented here deals with genetic variation within the Icelandic cattle breed using methods of molecular genetics. To introduce the matter, findings of related research are discussed and the Icelandic cattle breed described including its development and current status as a dairy breed.

1.2 Genetic variation within a population

An alternative to using genetic markers for evaluation of genetic variation within a population is to apply methods of quantitative genetics based on pedigree information. Both of these methodologies have been used for Icelandic cattle and are reviewed in the following two chapters.

1.2.1 Genetic variation estimated by marker analysis

Commonly used genetic markers include microsatellites (also termed simple/short tandem repeats) and single nucleotide polymorphisms (SNPs), both of which are used in genetic diversity studies. Searching for polymorphisms like SNPs in likely candidate genes is also a standard procedure when trying to identify a causative mutation (i.e. the quantitative trait nucleotide) that affects a physiological trait (Ron & Weller, 2007). Microsatellite analysis has been widely used in studies of domestic animals like cattle, both to evaluate genetic relationships between different breeds and also to estimate genetic diversity within a population (e.g. Kantanen *et al*., 2000b; Maudet *et al*., 2002; Cymbron, Freeman, Malheiro, Vigne & Bradley, 2005; Radko, Żyga, Ząbek & Słota, 2005; Li *et al*., 2007). In addition, genotypic arrays score multiple microsatellite loci in samples of individuals and are useful for individual identification, parentage and relatedness tracking (Sunnucks, 2000). Microsatellites have also been used for archaeological purposes e.g. when studying ancient cattle remains (Edwards *et al*., 2003).

Kantanen *et al*. (2000b) studied genetic diversity in 20 North European cattle breeds including the Icelandic cattle breed, using microsatellite data as well as red cell antigen and protein data. Their results showed a somewhat low genetic variation for the Icelandic cattle breed. Kantanen *et al*. (1999) also detected a decrease in the average heterozygosity and number of alleles in the Icelandic cattle breed when estimated over a 34 years period using red cell antigen and plasma protein loci as markers.

The relatively low genetic variation found in the Icelandic cattle breed in these two studies is inconsistent with the findings of Kantanen *et al*. (2000a) that show a relatively high phenotypic and genetic variation within the Icelandic cattle breed when estimated at six coat colour loci. The breed was classified as multicoloured along with Western Fjord cattle and Dølafe, and colour variation within multicoloured breeds appeared to depend on a high frequency of the E^+ allele at the extension locus (Klungland, Olsen, Hassanane, Mahrous & Våge, 2000). It should be noted here that the studies of Klungland *et al*. (2000) and Kantanen *et al*. (1999, 2000a, 2000b) are all based on the same sample of animals; only the markers differ between studies.

Using three microsatellite markers specially chosen to suit ancient DNA, Edwards *et al*. (2003) found genetic diversity values for the Icelandic cattle to be similar to that of three other Scandinavian breeds and one continental. These five breeds all had higher diversity values than six existing breeds from Britain and Ireland as well as a group of ancient cattle studied. Table 1 gives an overview of results from microsatellite analysis carried out on the Icelandic cattle breed.

Table 1. Results from various studies showing observed (H_O) and expected (H_E) heterozygosity within the Icelandic cattle breed (N is the number of analyzed individuals).

	Markers				
N	**Microsatellites (number used)**	**Other**	**H_O**	**H_E**	**Ref.**
15	3	-	0.580	0.680	Edwards *et al*. (2003)
48	-	Red cell antigen and plasma protein	-	0.264	Kantanen *et al*. (1999)
44	10	-	0.560	0.550	Kantanen *et al*. (2000b)
44	20	-	0.608	0.607	Tapio *et al*. (2006)

According to Kantanen *et al*. (2000b) the gene pools of North European cattle breeds have developed through a breed-specific evolution resulting in significant differentiation. Same

author found expected heterozygosity (H_E) values for 20 North European cattle breeds to vary between 0.45-0.69 and observed heterozygosity (H_O) for the same groups of animals to range between 0.49-0.68. In a similar study carried out by Tapio *et al*. in 2006 using 35 North European cattle breeds H_E was calculated between 0.56-0.72 and H_O between 0.53-0.72. Genetic material for 19 of the 35 breeds examined by Tapio *et al*. (2006) was the same as in Kantanen *et al*. (2000b).

Based on the results from the 20 breeds studied, Kantanen *et al*. (2000b) suggested the existence of a geographical cline with microsatellite-based heterozygosity and number of alleles from the southeast towards the northwest of Northern Europe. These results are similar to the results from a study on 103 different cattle breeds/populations from Europe, Africa, the Near East, and Asia (no Scandinavian breed was included) where the highest values for H_E were observed in cattle in the Near East and the H_E values for European cattle decreased gradually with greater distance from the Near East (Freeman, Bradley, Nagda, Gibson & Hanotte, 2005). A similar geographical trend can be observed when looking at the results of Cymbron *et al*. (2005). They found H_E values for three cattle breeds from the British Islands, ranging between 0.52 and 0.63. These were somewhat higher values than the ones found by MacHugh, Shriver, Loftus, Cunningham and Bradley (1997) for four breeds from the British Islands. However, both studies estimated the H_E to be overall higher for cattle breeds from continental Europe than the British breeds.

1.2.2 Genetic variation estimated from pedigree information

Estimation of the level of inbreeding within a population is a way to estimate its genetic variation. Inbreeding is defined as the mating of individuals more closely related than the average for a given population (Bourdon, 2000). For an individual, it is measured by the inbreeding coefficient (*F*) which estimates the probability that two genes at any locus in a given individual are identical by descent (Falconer & Mackey, 1996). Inbreeding leads to an increase in the frequency of homozygous genotypes and, thus, a decrease in the frequency of heterozygous genotypes.

It should be noted that both a theoretical model and empirical data have suggested a very weak correlation between *F* and heterozygosity measured by small number of molecular markers (10-20) unless the population under study exhibits an unusually high variance in *F*. Therefore, heterozygosity measured by molecular markers has been suggested to be a poor indicator of *F* (Slate *et al*., 2004; Balloux, Amos & Coulson, 2004).

For the years 1995-2000 the rate of inbreeding for the Icelandic cattle breed was calculated to be 0.42% over a generation and a related measurement, the effective population size (N_e), was estimated as 118 individuals compared to 0.34%, and 147 individuals respectively, ten years earlier (Kristjánsson, Jónmundsson & Benjamínsson, 2006). These numbers are considered to be at an agreeable level and N_e is substantially higher for Icelandic cattle than for many of the other commercial cattle breeds with larger census sizes (e.g. Sørensen, Sørensen & Berg, 2005; Taberlet *et al*., 2007 and ref. within). Modern breeding methods, using a best linear unbiased prediction (BLUP) animal model, can calculate the predicted transmitting ability of animals with great accuracy. Such models can, however, lead to an increase in the co-selection of related animals. Along with advanced reproductive technology like artificial insemination (AI) and embryo transfer there is also the threat of using relatively few outstanding animals, especially sires. Combined, this can increase inbreeding (Weigel & Lin, 2002). The breeding scheme for the Icelandic cattle breed has from the beginning considered the development of inbreeding by using more sires and bull dams at the cost of immediate progress and limited the use of each sire (around 7000 doses of semen are collected per individual) (Jónsson, Jónmundsson & Kristjánsson, 2007).

A genetic bottleneck is defined as a reduction in N_e and can therefore increase inbreeding rates and cause a loss of genetic variation (Cornuet & Luikart 1996). A recently bottlenecked population is likely to have lost rare alleles but may nevertheless still contain substantial heterozygosity. A recently bottlenecked population is defined by Luikart, Allendorf, Cornuet and Sherwin (1998) as a population bottlenecked within the past few dozen generations, based on the time during which a distortion of allele frequency distributions is likely to be detectable. They further claim that a bottleneck is likely to be detectable for only 40 to 80 generations, assuming that the maximum bottleneck size to be detected is around $N_e = 20$ and that bottlenecks can be detected for approximately 2-4 times N_e generations. After that, genetic drift and new mutations begin to re-establish mutation drift equilibrium. Using a method based on heterozygosity excess, Cornuet and Luikart (1996) suggested that a bottleneck of $N_e = 50$ was likely to be detectable for 25-250 generations after the initiation of a population reduction.
A population with a large census size can experience a genetic bottleneck in the absence of a demographic bottleneck, e.g. if only a few males mate with all the females (Luikart *et al*.,

1998). It can be estimated whether a population has experienced a recent bottleneck by using methods based on allele frequency data like microsatellite genotypes.

1.3 Two genes affecting milk production

The majority of economically important traits in livestock is quantitative and continuously distributed phenotypes result from the combined action of multiple genes, each with marginal, additive effects (Szyda & Komisarek, 2007). Such quantitative traits are therefore often called multigenic or complex traits and genes that contribute to them are known as polygenes (Glazier, Nadeau & Aitman, 2002). Phenotypic variation of complex traits can be caused by one or more single-nucleotide polymorphisms located in a single gene or closely linked genes. Kühn *et al*. (2004) hypothesized that QTLs associated with complex traits could often be determined by a complex interaction between haplotypes consisting of a number of mutations within one or many genes.

Knowledge of the genetic background of a complex trait, including SNPs in genes with great impact, is considered important in order to better understand the mechanism underlying the physiology of the trait (Kühn *et al*., 2004). Two genes, the *Leptin* and *DGAT1* genes, have over the past few years received an increasing interest as both have been shown to influence important economic traits in cattle. Therefore, identifying causative mutations in those genes has been of interest to researchers hoping to gain better understanding of their breeding material. The following two chapters give a description of the *Leptin* and *DGAT1* genes and related research in cattle.

1.4 The *Leptin* gene

The *Leptin* gene encodes a 167 amino-acid long protein that is mainly synthesized in white adipose tissue (Zhang *et al*., 1994) although its production has been located in other tissues in different mammalian species (for a review see van der Lende, te Pas, Veerkamp & Liefers, 2005). The *Leptin* gene, previously known as the *Obese* (*ob*) gene, was first characterized in mice through positional cloning and shown to be conserved in various vertebrate species including cattle (Zhang *et al*., 1994). The amino acid sequence of the bovine leptin protein was found to show 91% and 97% homology with porcine and ovine leptin sequences, respectively (Ji *et al*., 1998). The mature leptin protein is only 146 amino-acids long due to cleavage of a 21 amino-acid long signal peptide before the protein is excreted from the cell (Buchanan *et al*., 2002).

Leptin has multiple roles in mammals, most of which are related to energy balance control and general feeding behaviour. However, it has been implied that the protein plays a role in other biological processes such as regulation of reproduction and immune responses (Liefers *et al*., 2003b; Houseknecht, Baile, Matteri & Spurlock, 1998). Leptin lacks an internal membrane-spanning domain, is secreted into the bloodstream (Zhang *et al*., 1994) and is believed to exhibit its negative effect on food intake in the hypothalamus where it has been associated with mediators contributing to both starvation and obesity responses (e.g. Schwartz, Seeley, Campfield, Burn & Baskin, 1996).
Due to the diverse biological role of leptin, it has gained interest in domestic animals, especially ruminants and pigs. By identifying important features of the *Leptin* gene, scientists hope to gain better understanding of economically important processes in livestock. Such processes include direct and indirect targets of the hormone, like eating behaviour, milk production and carcass composition. In cattle, *Leptin* is located on chromosome 4 (BTA 4) and consists of three exons, of which exon 1 is not translated.

Several polymorphisms have been identified in the *Leptin* gene. These polymorphisms are found in exons and introns as well as in the promoter region of the gene. In a study performed on 22 animals of 13 diverse breeds, 20 SNPs were detected by sequencing 1788 base pairs (bp) of the *Leptin* gene giving a frequency of 1 SNP per 89 bp (Konfortov, Licence & Miller, 1999). In cattle, polymorphisms in the *Leptin* gene have been associated with milk production traits, carcass content, fertility, and feed consumption, as well as the expression level of the gene itself (e.g. Adamowicz, Flisikowski, Starzynski, Zwirezchowski & Switonski, 2006; Almeida, Almeida, Moraes & Weimer, 2003; Buchanan *et al*., 2002; Di Stasio, Brugiapaglia, Galloni, Destefanis & Lisa, 2007; Liefers, te Pas, Veerkamp & van der Lende, 2002; Liefers *et al*., 2005). Table 2 gives an overview of five different bi-allelic SNPs identified in the bovine *Leptin* gene, within exon 2 and parts of the flanking introns.

Table 2. Identified polymorphisms in exon 2 (Ex2) and parts of intron 1 (Int1) of the *Leptin* gene in various cattle breeds, their allele frequency, and association with milk production or other traits. Polymorphisms located in exon 2 are distinguished in bold and changes in amino acid with an asterisk. (N: number of individuals, *SSC: soluble solids content, **ADG: Average daily gain).

Allele	Population (N)	Allele frequency	Association	Reference
C/T [Int1 -102]	22 animals of 13 different breeds	C: 0.59 – T: 0.41	-	Konfortov *et al.* (1999)
C/T [Int1 -102]	Charolais x Holstein bull calves (168)	C: 0.65 – T: 0.35	-	Lagonigro, Wiener, Pilla, Woolliams & Williams (2003)
C/G [Int1 -79]	22 animals of 13 different breeds	C: 0.41 – G: 0.59	-	Konfortov *et al.* (1999)
C/G [Int1 -79]	Charolais x Holstein bull calves (168)	C: 0.35 – G: 0.65	-	Lagonigro *et al.* (2003)
C/T [Int1 -62]	22 animals of 13 different breeds	C: 0.98 – T: 0.02	-	Konfortov *et al.* (1999)
A/T [Ex2 +48] Y→F*	Charolais x Holstein bull calves (166)	A: 0.86 – T: 0.14	AT: greater feed intake compared to AA	Lagonigro *et al.* (2003)
C/T [Ex2 +101] R→C*	22 animals of 13 different breeds	C: 0.59 – T: 0.41	-	Konfortov *et al.* (1999)
C/T [Ex2 +101] R→C*	Charolais x Holstein bull calves (168)	C: 0.65 – T: 0.35	-	Lagonigro *et al.* (2003)
C/T [Ex2 +101] R→C*	Holstein cows (416)	C: 0.54 – T: 0.46	TT: more milk compared to CC	Buchanan, Van Kessel, Waldner & Christensen, (2003)
	Ayrshire (17)	C: 0.38 – T: 0.62	TT: more protein compared to CC	
	Brown Swiss (21)	C: 0.55 – T: 0.45	TT: increase in SSC* linear score	
	Canadienne (9)	C: 0.89 – T: 0.11		
	Guernsey (16)	C: 0.94 – T: 0.06		
	Jersey (20)	C: 0.47 – T: 0.53		
C/T [Ex2 +101] R→C*	Angus (60)	C: 0.42 – T: 0.58	T: fatter carcasses C: leaner carcasses	Buchanan *et al.* (2002)
	Charolais (55)	C: 0.66 – T: 0.34		
	Hereford (22)	C: 0.45 – T: 0.55	TT: higher leptin mRNA expression	
	Simmental (17)	C: 0.68 – T: 0.32		
C/T [Ex2 +101] R→C*	Blonde d´Aquitaine bulls (59)	C: 0.56 – T: 0.44	C: Higher ADG**, lower dressing %, higher marbling	Di Stasio *et al.* (2007)

The polymorphisms located in intron 1 (C/T [Int1 -102], C/G [Int1 -79] and C/T [Int1 -62]) were described by Konfortov *et al*. (1999) but have so far not been connected to any trait or condition in cattle. The A/T [Ex2 +48] polymorphisms in exon 2 was first described by Lagonigro *et al*. (2003).

Konfortov *et al*. (1999) also described the C/T [Ex2 +101] SNP in exon 2. It is a first position, non-conservative substitution, changing an arginine (R) into a cysteine (C). This polymorphism has been linked to various traits in cattle but results are inconsistent between studies. For example, it has been associated with production traits and higher serum leptin concentrations at different pregnancy intervals (Buchanan *et al*., 2003; Liefers *et al*., 2003a) although in another study no connection with production traits could be established (Madeja, Adamowitz, Chmurzynska, Jankowski & Melonek, 2004). The amino acid change has been suggested to affect the tertiary structure of the protein and thereby influence its binding to a receptor (Liefers *et al*., 2003a). However, this substitution is located at position four in the first of four helices of the leptin protein and outside the four most conserved regions (Zhang *et al*., 1997). Moreover, the amino acid at this position varies substantially between animals, being glutamine in primates and mice, arginine in cattle and dogs, tryptophan in pigs and histidine in rats (Zhang *et al*., 1997).

1.5 The *DGAT1* gene

Milk lipids consist primarily of triglycerides (Reece, 1997). Triglycerides are made of a glycerol molecule and three variable fatty acids which can differ in length (number of carbon atoms; C). The *DGAT1* gene encodes an enzyme; acylCoA:diacylglycerol acyltransferase 1 which is an integral membrane enzyme of 489 amino acids with 6-12 possible transmembrane domains (Cases *et al*., 1998). It catalyzes the final step in triglyceride synthesis; the linkage of a sn-1,2-diacylglycerol with a fatty acyl CoA to form a triacylglyceride (Chen & Farese, 2005).

In an experiment carried out with mice, Smith *et al*. (2000) found that knock-out mice, that lacked both copies of *DGAT1* (*DGAT1*-/-), were viable, healthy and fertile. The *DGAT1*-/- mice were also able to synthesize triglycerides which points to the intervention of other triglyceride-synthesizing enzymes. However, *DGAT1*-/- females were completely incapable of producing milk for their young indicating the importance of *DGAT1* for lactation.

Beside its role in lactation, the enzyme is also important for various other physiological processes relating to triacylglycerol metabolism, such as intestinal fat absorption, lipoprotein assembly, and adipose tissue formation (Cases *et al*., 1998 and ref. within). In cattle, *DGAT1* is located in the centromeric region of chromosome 14 (BTA14) and consists of seventeen exons.

A QTL associated with milk yield and composition has been described in the centromeric region of bovine chromosome 14 (Coppieters *et al*., 1998). Due to the role of the *DGAT1* enzyme in milk synthesis and the location of the gene, it became both a positional and a functional candidate for the previously mentioned QTL. Studies led to the discovery of a non-conservative dinucleotide substitution in exon 8 of the *DGAT1* gene. The substitution is a change from AA to GC which, at the protein level, results in a lysine (Lys, K) to alanine (Ala, A) substitution at amino acid number 232, therefore this polymorphism is commonly named K232A (Grisart *et al*., 2002). This substitution has been regarded as the causative mutation or the quantitative trait nucleotide for the QTLs effect (Winter *et al*., 2002; Grisart *et al*., 2002).

The lysine encoding variant is considered the ancestral state of *DGAT1* and the mutation is believed to have taken place early in the history of domestication of cattle or even before domestication (Winter *et al*., 2002; Grisart *et al*., 2002).

Since its identification, the K232A substitution has been associated with milk yield and composition in various dairy cattle breeds (e.g. Grisart *et al*., 2002; Spelman, Ford, McElhinney, Gregory & Snell, 2002; Thaller *et al.,* 2003; Pareek*,* Czarnik, Zabolewicz, Pareek & Walawski, 2005). The lysine encoding-allele has been shown to increase milk fat synthesis and to some extent, protein content (Winter *et al.,* 2002; Grisart *et al.,* 2002; Thaller *et al*., 2003; Gautier *et al*., 2007; Schennink *et al*., 2007). Indeed, the relative effect of the lysine allele on the amount of triglycerides synthesized was measured as 1.5 times more in comparison to the alanine allele when studied in cell culture (Grisart *et al*., 2004). The lysine allele has also been shown to cause a decrease in protein and milk yield (Grisart *et al*., 2002; Spelman *et al*., 2002; Thaller *et al*., 2003; Gautier *et al*., 2007; Schennink *et al*., 2007). The difference found in the fat effect of the allele has been suggested to be due to interactions with background genes of different populations (Spelman *et al*., 2002). No significant effects of the K232A polymorphism were found on 18 nonproduction traits (management, size, longevity and conformation traits) (Spelman *et al*., 2002). Table 3 shows results from studies on the K232A polymorphism.

Table 3. The K232A polymorphism in the bovine *DGAT1* gene of various cattle breeds, frequency of the K allele (lysine), and its association with milk production traits. (N: number of individuals, FY: Fat yield, MY: Milk yield, PY: Protein yield, F%: Fat percentage, P%: Protein percentage).

Population (N)	Allele frequency	Association	Reference
New Zealand (NZ) Holstein-Friesian bulls (1,527)	K: 0.60	K: Increases FY, decreases MY and PY	Spelman *et al.* (2002)
NZ Jersey bulls (1,053)	K: 0.88		
NZ Ayrshire bulls (113)	K: 0.78		
Fleckvieh bulls (833)	K: 0.07	K: Increases FY, F% and P%; decreases MY and PY	Thaller *et al.* (2003)
German Holstein bulls (858)	K: 0.55		
Swedish Red Polled bulls (65)	K: 0.07	-	Umeland (2006)
Montbéliarde bulls	K: 0.04	K: Increases FY, F% and P%; decreases MY and PY	Gautier *et al.* (2007)
Normande bulls	K: 0.13		
French Holstein bulls	K: 0.37		
Dutch Holstein Friesian cows (1,762)	K: 0.40	K: Increases FY, F% and P%; decreases MY and PY	Schennink *et al.* (2007)

As can be seen in Table 3 frequency of the wild-type lysine (K) variant varies a great deal between different cattle populations. This might be due to different breeding objectives regarding milk composition in different countries and the genetic background of the cattle breeds (Gautier *et al.*, 2007). Another explanation, as Kühn *et al.* (2004) pointed out, might involve other sources of genetic variation that existed in the genomic region of *DGAT1* and contributed to the variation of milk fat content. An interesting aspect of the *DGAT1* K232A polymorphisms is its influence on milk-fat composition. Milk-fat is relatively high in saturated fatty acids but low in polyunsaturated fatty acids, the latter commonly being regarded as a healthier dietary option. The lysine variant (K) has been found to increase the ratio of saturated to unsaturated fatty acids (Schennink *et al.*, 2007).

1.6 The Icelandic cattle breed

The Icelandic cattle breed is the only cattle population found in Iceland. It has been postulated that Icelandic cattle has descended from old Norwegian landraces brought to Iceland by settlers that populated the island in the years 874-930 (Adalsteinsson, 1981). Findings by Kantanen *et al.* (2000b) support this and they suggest that the animals brought

to Iceland may have originated from breeding areas of the present-day Blacksided Troender and Nordland cattle in Norway. The separation interval between the Icelandic cattle breed and the Blacksided Troender and Nordland cattle was estimated as 1100-1300 years but longer for other Norwegian breeds studied, such as Western Fjord cattle, Doela cattle and Telemark cattle. The Icelandic cattle breed is believed to have remained almost completely isolated since the settlement and is therefore regarded as a closed population, i.e. no major attempts have been made to import foreign genetic material.
Since the settlement, the Icelandic cattle breed is believed to have suffered several population bottlenecks, for example due to harsh weather conditions and catastrophic events such as volcanic eruptions. It has been estimated that during the 13th and 14th centuries, the breed counted around 100,000 animals (Sigurðsson, 1937). At the beginning of the 18th century, the number of cattle in Iceland was estimated to be approximately 36,000 but the population was reduced to less then 10,000 animals following the Skaftáreldar volcanic eruption later that century (Sigurðsson, 1937; Torfason & Jónmundsson, 2001). During the 19th century the population size stayed at an average of 20,000 animals but then started to increase around 1930. Today, the population counts around 69,000 individuals of which approximately 31,000 are recorded as milking cows (Bændasamtök Íslands, 2007, 2008b).
Regarding the import of foreign cattle, records show that during the 19th century a limited number of cattle was brought to Iceland from Denmark, the last ones being imported around 1870 (Sigurðsson, 1937). However, this import is not considered to have permanently affected the Icelandic cattle breed (Jónsson *et al*., 2007).

The current breeding programme for Icelandic cattle is based on the proposals of Jónsson and Jónmundsson published in 1974, inspired by contemporary Scandinavian breeding programmes (Jónmundsson & Jónsson, 1974). Their suggestions led to the founding of one extensive breeding databank serving the whole breed. In 1993, methods for the breeding assessment were changed, incorporating a BLUP animal model instead of a contemporary comparison (Sigurðsson, 1993).
The breeding objectives are defined by a breeding committee (i: Fagráð í nautgriparækt) spearheaded by the Farmers Association of Iceland (i: Bændasamtök Íslands) and consist of a number of traits, protein content and milk yield being the current major factors. In the years 1950-70 there was also emphasis on increasing the milk fat content and therefore a gain in fat content, but since 1970 fat content has not played an important role in the

breeding work (Jónmundsson & Jónsson, 1991). One bull centre serves the whole country and AI is the main breeding method for the Icelandic cattle breed including little less than 80% of producing cows (Jónmundsson, 2000). The use of home bulls in Iceland is therefore still considerable, compared to other Scandinavian countries, which is a concern and might reduce breeding progress (Jónmundsson & Jónsson, 1991; G. Hreiðarsdóttir, personal communication, 2008).
Traits not linked to performance, such as coat colour, have never been a part of the breeding objectives for the Icelandic cattle breed and it therefore contains great coat colour diversity, with six basic colours and over a 100 colour schemes (Klungland *et al*., 2000). However, selection is against horned animals so the majority of the population is polled.

There has been a steady increase in average production per cow estimated as milk yield, as well as protein and fat content, for several years now, following a change to fewer and larger dairy farms, each having larger herds than before (G. Hreiðarsdóttir, personal communication, 2008). Today, the average milk yield is close to 5.500 litres per cow per lactation period (Bændasamtök Íslands, 2008b).
The use of molecular genetic techniques has so far only been used to a limited extent for research of Icelandic cattle. Studies have been carried out on protein polymorphisms (Ólafsson, Eyþórsdóttir & Hafberg, 2003) but this has not been integrated into the breeding work. Direct studies in order to identify desirable alleles or other possible causative genes or chromosomal regions have not been carried out for Icelandic cattle, partly due to the small size of the breed. Identification of known polymorphisms should, however, be straight-forward and could benefit the breeding work (Eyþórsdóttir & Jónmundsson, 2004). Furthermore, the detection of DNA polymorphisms, whether in coding or non-coding regions of chromosomes, has enabled the description of the genetic uniqueness for a particular breed (Talle *et al*., 2005). When an estimation of the genetic variation level has been established for the breed in question it is possible to consider its future potentials. It can be speculated that an island breed like the Icelandic cattle is likely to suffer from a low level of genetic variation considering its relatively small founder population and a history of population fluctuations combined with progressive breeding strategies in recent decades. This, however, needs a further verification.

2 Aims of study

This study had three main objectives:

Firstly, to increase knowledge of the genetic variation within the Icelandic cattle breed. For this purpose a sample of heifers was genotyped using a panel of eleven microsatellite markers. Microsatellite analysis is a common method when genetic variation in cattle is studied (e.g. Kantanen *et al*., 2000b; Tapio *et al*., 2006). The potential use of the markers in parentage testing was also evaluated.

Secondly, to establish an unbiased DNA collection for Icelandic cattle. Such a collection was a prerequisite for the study and will be of importance for future research.

Thirdly, to see if certain polymorphisms that have been identified and associated with commercial traits in various dairy breeds exist in Icelandic cattle and explore their frequency in the population as well as search for other polymorphisms distinctive for the breed. The identification of reported SNPs and/or finding new ones may benefit the breeding work and yield valuable information on the genetic variation level for the breed. For this purpose the sample was genotyped in specified regions of two genes known to affect milk production (see Table 2 and 3 for references).

3 Material and methods

3.1 Sample collection and DNA isolation

Blood samples of 440 Icelandic heifers were collected from 58 dairy farms covering all major dairy farming regions of Iceland. Genomic DNA was extracted from buffy coat using the MasterPure™ DNA Purification Kit (EPICENRTE® Biotechnologies) according to manufacturer's recommendations.

The samples were collected from September 2006-February 2007. The regions were divided into the Northern region (including Skagafjörður (61 samples) and Eyjafjörður along with a small area of Suður-Þingeyjarsýsla (122 samples)), the Western region (78 samples) and the South-Western region (179 samples). Milking cows within these regions (Figure 1, see Manuscript I) are estimated as 84.1% of all recorded milking cows. Samples were obtained in cooperation with another independent study and sampling limited to heifers in first pregnancy (see Jónsson, 2008).

The 100 samples used in this study for genetic analysis were chosen randomly from the pool of 440 and included samples from 45 farms. The 13 farms not included in the sample of 100 were divided between the three parts of the country such that five belonged to the Northern part, one to the Western part and the remaining seven were located in the South-Western part. A complete list of the 100 heifers used in the analysis is given in Appendix 1.

Efforts were made to evaluate the relationship of the heifers in the pool of 440 and the sample of 100 in order to see if the samples were typical for the breed and showing the same average interrelationship as the population in general. For this purpose, pedigree information obtained from the Farmers Association of Iceland was used to count the number of sires and to compute the coefficient of relationship (R). This was done by applying the programs of Boichard (2002). R was calculated for the pool of 440, the sample of 100 and all the individuals of the Icelandic cattle breed born in the years of 2003, 2004 and 2005, a total of 14,505 individuals. These three years were chosen as the majority of the heifers studied were born in that period.

3.2 Microsatellite analysis

The following eleven bovine microsatellite markers were used to genotype the heifers: BM1824, BM2113, ETH10, ETH225, ETH3, INRA23, SPS115, TGLA122, TGLA126, TGLA227, and TGLA53. The markers are distributed over the bovine genome, covering

11 of the 29 autosomes. All of them are included on a list comprising 30 microsatellite markers and jointly recommended by the FAO and the International Society for Animal Genetics, (ISAG) to be used for analysis of genetic diversity in cattle (Hoffmann *et al*., 2004). In addition, nine of the eleven microsatellites used in this study are recommended by ISAG for parentage testing (Roslin Institute, 2002).
The heifers were genotyped for the microsatellites at Eurofins Medigenomix GmbH (http://www.medigenomix.de/en/index.html) using the ABI StockMarks Cattle® Bovine Genotyping Kit including the recommended ISAG marker sets.

3.3 Genetic diversity analysis

Several different estimators can be used to analyse marker data. The basic diversity indices, (i.e. H_O, H_E, mean number of alleles (MNA), and polymorphism information content (PIC)) were calculated for all the 11 microsatellites using the POWERMARKER package (Liu & Muse, 2005). Total number of alleles (observed alleles, N_O) and allele frequency were calculated over individual loci. Mean number of alleles is the total number of all distinct alleles at all loci divided by the number of loci. As the observed number of alleles in a sample depends to a large extent on the size of the sample (Goudet, 2001) allelic richness was also calculated as a measure of the number of alleles irrespective of sample size.
Observed heterozygosity (H_O) is a simple measure of genetic variation in a population and can be reported for a single locus or as an average over a number of loci (Weir, 1996). The proportion of heterozygous individuals in a population at a single locus was calculated as:

$$\hat{H}_l = 1 - \sum_{u=1}^{k} \tilde{P}_{luu}$$

where l is the the lth locus and P_{luu} is the population (sample) frequency for the genotype A_uA_u (the symbol A means any genetic locus with a series of alleles A_u) (Liu & Muse, 2005).
Gene diversity (D) hereafter referred to as expected heterozygosity (H_E) is defined as the probability that two randomly chosen alleles from a population are different. An unbiased estimator of gene diversity calculated at the lth locus is:

$$\hat{D}_l = (1 - \sum_{u=1}^{k} \tilde{p}_{lu}^2) / (1 - \frac{1+f}{n})$$

where p_{lu} is the population frequency of an allele A_u at the lth locus, n is the number of individuals and f is the inbreeding coefficient (Liu & Muse, 2005).

Polymorphism information content was developed by Botstein, White, Skolnick and Davis (1980) by using two hypothetical loci, one containing a rare dominant allele called the "index locus" and the other being the "marker locus". The informative value of the marker locus was then defined as the probability that a given offspring which had inherited the rare allele at the index locus would allow deduction of the parental genotype at the marker locus. The higher the PIC value, the more informative the marker and a locus with PIC value > 0.5 is regarded as highly informative whereas a locus with a PIC value < 0.25 is regarded as slightly informative (Botstein *et al*., 1980). The PIC value can be calculated as:

$$PIC_l = 1 - \sum_{u=1}^{k} \tilde{p}_{lu}^2 - \sum_{u=1}^{k-1} \sum_{v=u+1}^{k} 2\tilde{p}_{lu}^2 \tilde{p}_{lv}^2$$

where allele A_v has the population frequency p_{lv} at the lth locus (Liu & Muse 2005). Guo and Elston (1999) generalized the definition of the PIC value of Botstein *et al*. (1980) to a general measure of how informative a marker is, regardless of the mode of inheritance of the trait being linked. Being a measure of a marker's polymorphism a high PIC value reflects an informative marker and depends on the number of alleles and the frequency of each allele at the marker locus (Guo & Elston, 1999). Reasonably high PIC values for microsatellite markers have been regarded as indicative of the marker's usefulness for biodiversity evaluation (Sodhi, Mukesh, Prakash, Ahlawat & Sobti, 2006).

The POWERMARKER package (Liu & Muse, 2005) was used to estimate inbreeding within the Icelandic cattle breed by calculating the within-population inbreeding coefficient f of Cockerham. This corresponds to Wright's within population inbreeding coefficient F_{IS} and is defined as the correlation of alleles within individuals within one population (Weir, 1996). Two different methods were applied: first an EM (expectation-maximization) algorithm to find the MLE (maximum likelihood estimation) of F_{IS} and, secondly, the method of moments (using statistics that are unbiased for F_{IS}) (Weir, 1996; Liu & Muse, 2005). In addition, the within population inbreeding estimates were

calculated for each locus and overall loci using the FSTAT computer program version 2.9.3.2. (Goudet, 2001).
ONeSAMP 1.0 (http://genomics.jun.alaska.edu/; Tallmon *et al*., 2008) was applied to estimate the N_e of the Icelandic cattle breed using the microsatellite data. FSTAT and STRUCTURE 2.0 (Pritchard, Stephens, & Donnelly, 2000) were used to examine population subdivision. For FSTAT, the sample of 100 heifers was divided into two main groups according to their geographical location; the Northern region and the combined Western and South-Western regions (North/South). The null hypothesis was that there were no differences between the groups estimated as allelic richness, observed and expected heterozygosity, F_{IS} and F_{ST} (F_{ST} defined as the correlation of alleles of different individuals in the same population, see Weir, 1996). The sample was not divided when using STRUCTURE.

The Hardy-Weinberg equilibrium (HWE) implies that allele frequencies are constant from one generation to the next. If a population deviates from HWE it indicates that some evolutionary force, e.g. selection, mutation, or migration, is changing the allele frequencies between generations, i.e. allele frequencies is not constant for the population in question (Hartl & Clark, 1989). Exact tests for deviations from HWE per locus and for the population were performed using the GENEPOP program version 4 (Rousset, 2007) applying a Markov chain to compute unbiased estimates of the exact probabilities (*P*-values) of being wrong in rejecting HWE.
The chi-square goodness-of-fit was also applied to calculate deviations from HWE per locus using POWERMARKER. The two groups, (North/South) were also used to estimate HWE over the whole population by permuting (5000 times) alleles among groups. The statistic used to compare the randomised data sets to the observed was the overall inbreeding coefficient *F,* i.e. Wright's F_{IT}, which is defined as the correlation of alleles within individuals over all populations (Weir, 1996; Goudet, 2001).
Exact test for genotypic linkage disequilibrium for pairs of markers was performed using FSTAT.

Two methods were employed in order to estimate whether the Icelandic cattle breed has experienced a recent reduction in the N_e or a genetic bottleneck. The first method was based on heterozygosity excess as described by Cornuet and Luikart (1996). It is based on the fact that for neutral loci, allele number and frequency distribution results from the

equilibrium between mutation and genetic drift, this being governed by the mutation rate and N_e (Cornuet & Luikart, 1996). If the difference between the heterozygosity observed for a sample of genes is significantly larger than the heterozygosity expected from the number of alleles observed if the population were at mutation drift equilibrium then the population exhibits a heterozygosity excess and is considered to have experienced a recent genetic bottleneck. Two different statistical tests were applied, a sign test, and a Wilcoxon test. These tests were used following three different models of microsatellite evolution: infinite allele model (IAM), stepwise mutation model (SMM) and two-phased model of mutation (TPM). Briefly, the IAM is based on the equilibrium between the loss of variation caused by drift and the introduction of new variation by mutation, each mutation producing a new allele different from all existing ones (Weir, 1996), while the SMM better accounts for the exact changes of an allele caused by mutation before attaining a steady state (Hartl & Clark, 1989) and TPM is an intermediate stage, incorporating the mutational process of the SMM while allowing for mutations of a larger magnitude to occur (Murray, 1996).
The second approach involved the graphical method of Luikart *et al*. (1998) stating that in a nonbottlenecked population alleles at low frequency (0.0 – 0.1) are always more abundant that alleles at more intermediate frequency. A mode-shift distribution of allele frequency (i.e. fewer alleles in low frequency classes compared with intermediate frequency classes) is expected in recently bottlenecked populations.
Both of the above approaches were carried out using the computer software program BOTTLENECK (http://www.montpellier.inra.fr/URLB/bottleneck/bottleneck.html) performing 5000 replicates.

The Excel Microsatellite Toolkit version 3.1 (Park, 2001) was used to transform data to a format acceptable by the GENEPOP and FSTAT program.

3.4 Amplification and sequencing of parts of the *Leptin* and *DGAT1* genes

For designing PCR (polymerase chain reaction) and sequencing primers/oligos reported, sequences of the *Leptin* (Genbank Accession number AY138588 (Lagonigro *et al*., 2003), refGene NM_173928) and *DGAT1* genes (Genbank accession number AJ318490 (Winter *et al*., 2002)) were retrieved from the NCBI and UCSC databases (http://www.ncbi.nlm.nih.gov and http://genome.cse.ucsc.edu). PCR primer pairs were designed using the program Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). The primers were synthesized at Eurofins Medigenomix

GmbH (http://www.medigenomix.de/). The *Leptin* gene amplicon spanned exon 2, and parts of introns 1 and 2, a total of 598 base pair fragment using primers BtLepEx2_F (5´-ACA CCT CCT GTG GTT TTC TTG ATT CCG-3') and BtLepEx2_R (5´-GGC ACT AGG ATT CCG GTC TGG-3') (Figure 1). The *DGAT1* gene amplicon included exons 7, 8, and 9, and introns 7 and 8 (443 base pairs) and was amplified using following primers: BtDGAT1_F (5´-TGC TGG CCC TGA TGG TCT ACA CCA TC-3') and BtDGAT1_R (5´-GTC GCC GCA GCA GGA AGC GCT TTC G-3') (Figure 2). The fragment covered the site in exon 8 where a nonconservative lysine to alanine substitution (K232A) has been described by Grisart *et al*. (2002).

The PCR was carried out in a 25 µL volume containing approximately 15 ng of genomic DNA, 10 pmol of each primer and 12.5 µL of Taq 2x Master Mix (as supplied by New England BioLabs®) which included 0.4 mM dNTPs, 50 U/ml Taq polymerase, 3.0 mM MgCl2, Standard Taq Reaction Buffer and stabilizers.

The PCR program used for amplification was as follows: initial denaturation at 94°C (4 min) for one cycle; denaturation at 94°C (30 sec), annealing at 52°C (45 sec), extension at 72°C (2 min) for 35 cycles, and final extension step at 72°C (4 min) for one cycle. A Px2 Thermal Cycler (Thermo Electron Corporation) was used to perform the amplifications. After amplification, PCR products were run on 1% agarose gel stained with ethidum bromide and visualized by UV light before purification. PCR products were purified either directly or after gel electrophoresis using a NucleoSpin® Extract II PCR clean-up/Gel extraction kit (Macherey-Nagel GmbH & Co. KG) according to the manufacturer's manual with DNA eluted in 40 µl of elution buffer. The purified PCR product was checked by running it on a 1% agarose gel. For sequencing, the appropriate primers were added to the samples and they were then commercially sequenced (http://www.eurofinsdna.com/).

Sequences were analyzed using the Vector NTI Advance® Software from Invitrogen. Haplotypes were examined using POWERMARKER.

An association study was carried out for the SNP located in exon 2 of the *Leptin* gene (Ex2 +101) using data on milk yield and protein content provided by the Farmers Association of Iceland. Milk yield was estimated as the milk (kg) produced in the first lactation period which is used for breeding value estimation. Milk yield value could not be provided for all the heifers (a total of 26 heifers); some due to a missing record value and some because they had not completed their first lactation period at the time of the assessment or had been excluded from it. The protein content was estimated as the average value from parturition

until the end of the year 2007 (values were missing for 11 heifers). A one-way analysis of variance (ANOVA) was constructed using the MINITAB® Release 14.20 Statistical Software in order to estimate the association separately for the two traits. Due to missing records for the majority of homozygous TT heifers (four heifers for milk yield and three for protein) they were excluded from the analysis.

4 Results

4.1 Sampling

For analysis of the relationship between sampled individual pedigree information could not be obtained for 10 and 40 heifers occupying the sample of 100 and group of 440 heifers, respectively. Relationship calculation was therefore based on the remaining 90 and 400 heifers.

The coefficient of relationship, R was calculated between and within all groups. When calculating R between groups of 400 and 14,505 three animals born outside the 2003-2005 period were removed. Therefore, R between these two groups is calculated using 397 and 14,505 animals. Overall, R was approximately 0.03 (0.025-0.030) within and between all groups included (Table 4).

Table 4. Coefficient of relationship (R) calculated for sample and groups of Icelandic cattle. (*Three animals excluded when calculating R between this group and the 14,505 animals).

	Group of 90	**Group of 400***	**Group of 14,505**
Sample of 90	0.028	0.030	0.026
Group of 400*		0.030	0.027
Group of 14,505			0.025

4.1.1 Sires

The sample of 90 heifers and the group of 14,505 animals had a total of 45 and 740 known sires, respectively. The heifers in the groups were found to share the same most common male ancestors (Table 5). The strong influences of Þráður 86013 and Bassi 86021 are clear. Together they contribute, either as fathers or grandfathers of sires, to a total of 2,211 heifers in the group of 14,505 and 21 heifers in the group of 90, or 18% and 26%, respectively.

Table 5. The proportion of the ten most common sires for the sample of 90 heifers and for the Icelandic cattle population (all animals born in 2003, 2004, and 2005). Ancestors are shown in parenthesis. Names and individual number of each bull is in accordance with the breeding databank recording (Bændasamtök Íslands 2008a). F: father; Grf: grandfather.

Name and record number of sire	% of daughters in group of 14,505 (% of daughters in group of 90)	Name and record number of sire	% of daughters in group of 90 (% of daughters in group of 14,505)
Soldán 95010 (F: Bassi 86021)	7.19 (0)	Hófur 96027 (F: Þráður 86013)	8.54 (3.22)
Punktur 94032 (F: Þráður 86013)	3.36 (0)	Ölvir 02366 (Grf: Bassi 86021)	4.88 (0.06)
Hófur 96027 (F: Þráður 86013)	3.22 (8.54)	Flói 02029 (Grf: Þráður 86013)	4.88 (0.35)
Túni 95024 (F: Daði 87003)	3.04 (0)	Sendill 02013 (Grf: Þráður 86013)	3.66 (0.57)
Fróði 96028 (F: Óli 88002)	2.84 (3.66)	Glæðir 02001 (Grf: Þráður 86013)	3.66 (1.04)
Hvítingur 96032 (F: Óli 88002)	2.78 (3.66)	Villingur 01036 (Grf: Bassi 86021)	3.66 (0.75)
Frískur 94026 (F: Bassi 86021)	2.53 (1.22)	Harrason 01912 (Grf: Daði 87003)	3.66 (0.22)
Sproti 95036 (F: Daði 87003)	2.13 (1.22)	Hvítingur 96032 (F: Óli 88002)	3.66 (2.78)
Pinkill 94013 (F: Bassi 86021)	1.84 (0)	Fróði 96028 (F: Óli 88002)	3.66 (2.85)
Prakkari 96007 (F: Holti 88017)	1.46 (2.43)	Dúri 96023 (F: Holti 88017)	3.66 (1.05)
Sire unknown%	*19*	*Sire unknown%*	*9.76*

4.2 Microsatellites and genetic diversity analysis

4.2.1 Diversity indices

A total of 91 samples were successfully analyzed with the exception of one marker for one sample. Altogether, 68 alleles were detected across the 11 loci giving a mean number of 6.182 alleles per locus (MNA). The frequency of the most common allele never exceeded 0.95 so all the loci were polymorphic according to Hartl and Clark (1989). Number of alleles ranged between 4 (ETH3) and 9 (TGLA53). Genetic diversity measures showed mean observed heterozygosity (H_O) of 0.626 and mean expected heterozygosity (H_E) of 0.685. PIC values exceeded 0.5 for all loci except SPS115 (Table 6). For a graphical distribution of alleles see Figure 2 in Manuscript I. The inbreeding coefficient for the Icelandic cattle population was estimated over all loci as 8.8%, 9.7%, and 9.7% according to the EM algorithm, the method of moments, and Weir and Cockerham's F_{IS}, respectively (see Table 6 for values for each locus, FSTAT output). The N_e of the Icelandic cattle breed

was estimated to be 111 individuals (lower and upper 95% confidence limits set as 99.76 and 127.39, respectively).

Table 6. Diversity indices calculated for 91 Icelandic heifers. Chromosome number (Chr), number of observed alleles (N_O), size range in base pairs (reported values for other cattle breeds also included), frequency of alleles, heterozygosity (H_O observed, H_E expected), polymorphism information content (PIC), and within population inbreeding estimates (F_{IS}). The allele at highest frequency and its respective frequency is given in italic for each individual marker. The highest and lowest values for H_E and H_O and underlined. Reported values from Roslin (2002). †: Aberdeen-Angus (AA), Ayrshire (A), Friesian (F) and Holstein (H) cattle breeds; ††: AA, A, F, H and Limousin (L); ‡: H, ‡‡: H and L; the rest of unlabelled range values are for AA and H.

			Size range of alleles						
Marker	**Chr**	**N_O**	**Icelandic cattle breed**	**Reported values**	**Allele frequency**	**H_O**	**H_E**	**PIC**	**F_{IS}**
BM1824	1	5	178; *180*; 182; 186; *188*	179-191†	0.137; *0.423*; 0.005; 0.011; *0.423*	0.571	0.619	0.545	0.088
BM2113	2	8	123; *125*; 131; 133; 135; 137; 141; 143	126-142†	0.016; *0.368*; 0.044; 0.143; 0.115; 0.170; 0.137; 0.005	<u>0.769</u>	<u>0.776</u>	0.753	0.020
ETH10	5	6	213; 215; 217; *219*; 223; 225	207-223‡	0.055; 0.093; 0.066; *0.615*; 0.143; 0.027	0.528	0.581	0.555	0.102
ETH225	9	5	*140*; 144; 146; 148; 150	137-153††	*0.390*; 0.071; 0.093; 0.379; 0.066	<u>0.462</u>	0.681	0.631	0.332
ETH3	19	4	117; 119; *125*; 127	109-131‡	0.203; 0.247; *0.368*; 0.181	0.736	0.725	0.681	-0.004
INRA23	3	6	192; 202; 208; 212; *214*; 216	199-219‡‡	0.011; 0.005; 0.231; 0.209; *0.418*; 0.126	0.648	0.708	0.665	0.096
SPS115	15	5	*248*; 252; 254; 256; 260	234-256	*0.689*; 0.006; 0.094; 0.128; 0.083	0.511	<u>0.491</u>	0.462	-0.031
TGLA122	21	7	141; *143*; 147; 149; 151; 171; 173	136-184	0.022; *0.368*; 0.242; 0.225; 0.077; 0.005; 0.060	0.604	0.740	0.705	0.194
TGLA126	20	6	113; *115*; 117; 121; 123; 125	114-127	0.022; *0.341*; 0.280; 0.137; 0.049; 0.170	0.758	0.751	0.715	0.001
TGLA227	18	7	81; 89; 91; 93; 95; *97*; 101	79-105	0.049; 0.291; 0.055; 0.170; 0.011; *0.418*; 0.005	0.714	0.702	0.658	-0.006
TGLA53	16	9	154; 160; 162; 164; 168; 170; 172; *176*; 178	151-183	0.198; 0.093; 0.209; 0.027; 0.005; 0.027; 0.044; *0.379*; 0.016	0.582	0.756	0.728	0.240
Mean		**6.182**				**0.626**	**0.685**	**0.645**	**0.097**

FSTAT yielded no statistically significant difference between the two groups comparing values for allelic richness, estimated as 4.305 and 4.414 for the Northern and South-Western groups, respectively, observed and expected heterozygosity and average relatedness ($P > 0.05$). No further subdivision of the breed was observed applying STRUCTURE (data not shown).

Three loci (ETH225, TGLA122, and TGLA53) showed significant ($P < 0.05$) deviations from HWE according to the exact test. Significant deviations were found for the same three loci as well as the TGLA126 locus when using chi-square test. When examined over all loci the results showed significant deviations from HWE in the population. Exact test for genotypic linkage disequilibrium revealed insignificant *P* values (> 0.05), and therefore independent assortment, for all but three pairs of loci (BM2113 and TGLA126, ETH225 and SPS115, INRA23 and TGLA53).

No recent bottleneck was revealed in the Icelandic cattle breed. Both the Sign test and the Wilcoxon test (a non-parametric test) yielded significant probability values for IAM ($P < 0.05$) therefore rejecting the null hypothesis of mutation drift equilibrium. The probability values for SMM and TPM were in the same two cases insignificant, however (Table II in Manuscript I). Taken together, the results from these two tests indicate the absence of a recent genetic bottleneck (null hypothesis accepted). The mode-shift test showed a normal "L" shaped distribution (Figure 3 in Manuscript I) characteristic for a non-bottlenecked population.

4.3 Analysis of single nucleotide polymorphisms

The amplified fragment of the *Leptin* gene was observed to be of the desirable length (data not shown). Around 60 heifers were successfully sequenced for the *Leptin* gene fragment and four polymorphisms, all bi-allelic single nucleotide substitutions were detected (Figure 1 and Table 7). Three SNPs were located in the introns flanking exon 2 (C/T at Int1 -102, C/G at Int1 -79 and G/A at Int2 +21) but one was located in exon 2 (C/T at Ex2 +101). The last one causes a non-conservative amino acid substitution, from arginine (Cgc) to cysteine (Tgc). Three of the four SNPs found in this study have already been identified before (Konfortov *et al*., 1999) but the G/A substitution located in intron 2 has not been reported previously.

An A/T substitution in exon 2 (Ex2 +48) described by Lagonigro *et al*. (2003) was not found in the Icelandic cattle, all 58 heifers sequenced were homozygous AA (tyrosine).

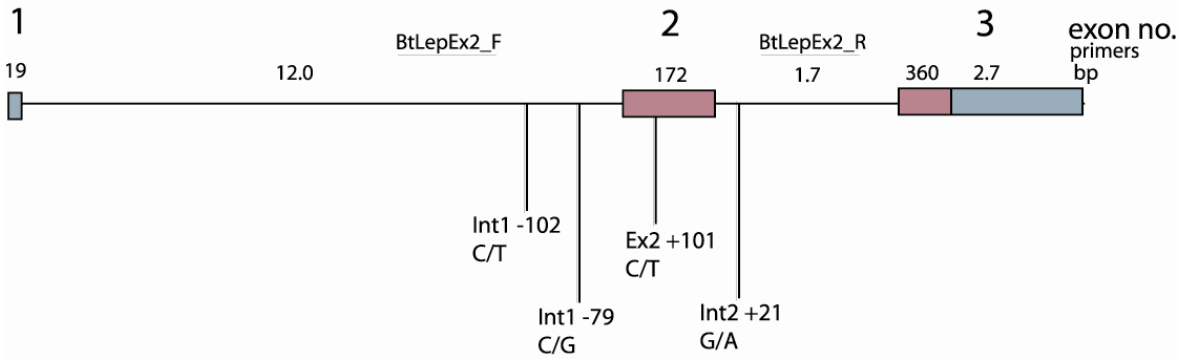
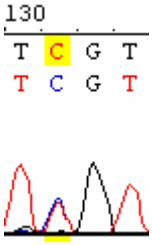
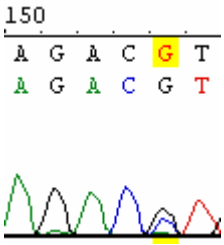
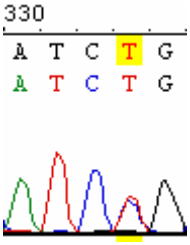
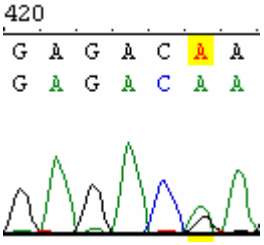


Figure 1. A schematic view of the bovine *Leptin* gene showing its exon-intron structure as well as location of primers and the four SNPs detected. Pink areas are translated, gray ones not. Gene organization based on Genbank Accession no. AY138588 (exon 2), AJ132764 (exon 3), RefGene NM_173928 and Lagonigro *et al.* (2003).

Table 7. A summary of sequencing results for a *Leptin* gene fragment in Icelandic heifers. (*Numbers in parenthesis are as in Genbank accession no. AY138588). Chromatograms show heterozygosity at polymorphic location (highlights) and no. according to contig are shown.

Location	Int1 -102 (103)*	Int1 -79 (126)*	Ex2 +101 (305)*	Int2 +21 (397)*
SNP	C/T	C/G	C/T	G/A
Chromatograms				
Genotype	CC: 10 CT: 28 TT: 25	CC: 27 CG: 25 GG: 9	CC: 24 CT: 26 TT: 7	GG: 53 AG: 6 AA: 0
Allele frequency (%)	C: 38.1 T: 61.9	C: 64.8 G: 35.2	C: 64.9 T: 35.1	G: 94.9 A: 5.1
Number of heifers	63	61	57	59
Amino acid change	-	-	*C*gc (R) *T*gc (C)	-

When haplotypes were examined for the two identified polymorphic sites in exon 2 and intron 2, the A variant in intron 2 was often inherited along with the C allele in exon 2 (frequency of the C-A haplotype was 0.05). Of the six heifers heterozygous for A/G, four were homozygous C/C and the remaining two heterozygous C/T.

Heifers heterozygous C/T produced on average more milk with higher protein content than heifers homozygous C/C (5,124 L and 3.35% compared with 4,750 L and 3.26%, respectively). However, the difference was not significant in either case ($P > 0.05$).

The amplified fragment of the *DGAT1* gene was observed to be of the desirable length (data not shown). All the 97 heifers successfully sequenced were homozygous for the alanine residues, (GC/GC) at the K232A site (Figure 2 and Table 8).

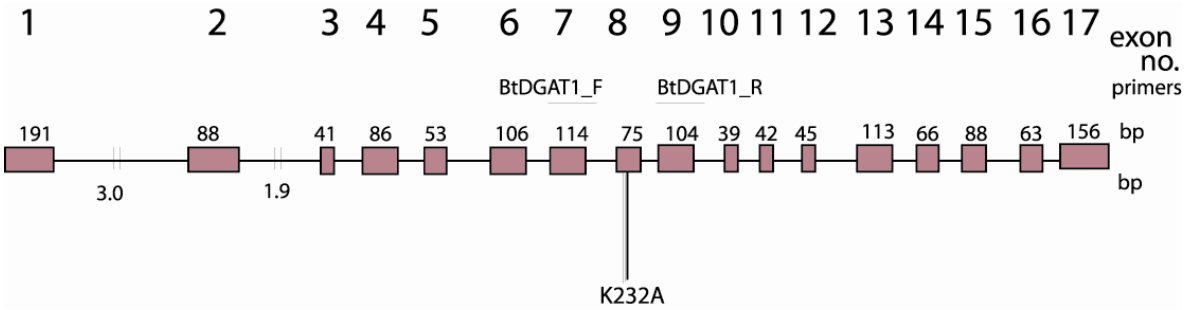
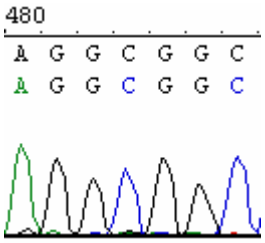


Figure 2. A schematic overview of the bovine *DGAT1* gene showing its relative structure, location of primers, and the K232A substitution. Genbank accession number AJ318490. Gene organization modified from Winter *et al*. (2002).

Table 8. A summary of sequencing results for a fragment of the *DGAT1* gene in Icelandic heifers. (*Numbers in parenthesis are as in Genbank accession no. AJ318490). Chromatogram shows homozygosity at the K232A (no. 482 and 483 according to contig).

Location	Ex8 15 and Ex8 16 (10,433 and 10,434)*
SNP	A/G and A/C
Chromatogram	
Genotype	GC/GC: 97 GC/AA: 0 AA/AA: 0
Allele frequency (%)	AA (K): 0,0 GC (A): 1,0
Number of heifers	97
Amino acid change	*AA*g (K) *GC*g (A)

5 Discussion

5.1 The sample

The coefficient of relationship was very similar when calculated between and within the three groups of animals under study. Moreover, the same five ancestors (fathers or grandfathers of individual sires) are present in the Icelandic cattle breed (group of animals born in 2003, 2004, and 2005) as in the sample of heifers studied. Together, these results support the assumption of a normally related sample representative for the population. Currently, the relationship of cattle in and between different regions of the country is more or less the same. This is considered a result of the collective breeding work and centralized AI beginning in 1970 (Jónmundsson, Kristjánsson & Benjamínsson, 2007). The fact that no obvious subdivision of the breed was observed in this study further suggests a representative sample. Therefore, although some areas of Iceland were not covered in this study, this is not believed to have caused a bias of the results. The existence of a different cline within the breed not found here cannot, though, be ruled out without using a larger sample covering all parts of the country.

5.2 Microsatellites and genetic diversity analysis

The results show that there exists a considerable level of genetic variation within the Icelandic cattle breed. The sizes of observed alleles were comparable to sizes in a number of prevalent commercial breeds (Table 6). The values for mean observed and expected heterozygosity (H_O and H_E) are considerably larger than the ones found by Kantanen *et al.* (2000b) for the Icelandic cattle breed but more similar to those found by Tapio *et al.* (2006). Six of the 20 loci used by Tapio *et al.* (2006) were used in this study whereas none were identical to the loci used by Kantanen *et al.* (2000b). The low values obtained by Kantanen *et al.* (2000b) were partly explained by the fact that island breeds tend to exhibit a lower level of heterozygosity than continental breeds because of the finite founder population as well as geographical limitations to gene flow. Indeed, the Icelandic cattle breed has been almost a completely closed population for over a thousand years. It is an island breed and considering the difficulties of transporting animals at the time of settlement, it is likely that the founding population for the breed was relatively small. However, it has been suggested that import of cattle took place over the whole settlement period and therefore a constant flow of genetic material took place during that period (Jónsson *et al.*, 2007).

Earlier results by Kantanen *et al.* (1999) suggested a decrease in the average heterozygosity and number of alleles for the Icelandic cattle when estimated over a 34 years period (1962-1996) using red cell antigen and protein markers. Here it can be mentioned that protein-based markers may be influenced by selection, developmental or various environmental factors and have become widely replaced by DNA-based methods (Talle *et al.*, 2005).

It has been suggested that the genetic divergence of Nordic cattle breeds and their within-population diversity can be explained by the combined effects of breed origin, admixture during foundation and development, and random genetic drift due to limitations in N_e either when the breed was founded or more recently (Kantanen *et al.*, 2000b).
Allelic richness calculated for the two groups (North/South) of heifers was in both cases found to be a little higher than the value, 3.948, calculated for the Icelandic cattle breed by Tapio *et al.* in 2006. Nevertheless, it was still lower than the values from same study obtained for the three main Nordic cattle breeds (the Norwegian Red, the modern Swedish Red-and-White, and the Swedish Friesian), which all exceeded 4.750. Collectively, H_E and H_O values for the three Nordic breeds were found to range between 0.67-0.72 (Tapio *et al.*, 2006). These values are similar to the ones obtained here for the Icelandic cattle breed.

Findings suggest that microsatellite-based diversity decreases across a geographical cline from east to west, i.e. from the Middle East to Western Europe (Kantanen *et al.*, 2000b; Freeman *et al.*, 2005). Being an island breed with a North-European origin, it can be expected that the genetic diversity of the Icelandic cattle breed is somewhat lower than for most continental breeds. However, the value for H_E in the Icelandic cattle breed is close to the same value pooled for eleven Northern continental breeds by Cymbron *et al.* (2005) and higher than the average value for three continental breeds in the study of MacHugh *et al.* (1997). The number of common loci in these studies were 6 of 19 (Cymbron *et al.*, 2005) and 2 of 20 (MacHugh *et al.*, 1997). However, the H_E value of the Icelandic breed is considerable lower than a value pooled for eight countries of the Near East (Cymbron *et al.*, 2005), supporting the previously mentioned geographical cline.
All the markers used in the present study were polymorphic and regarded as highly informative according to Botstein´s definition (> 0.5) (Botstein *et al.*, 1980), with the exception of the marker SPS115. Since all the PIC values were reasonably high and considering that they are all recommended for genetic diversity studies in cattle (Hoffmann

et al., 2004), nine also being specially recommended for parentage testing, this suggests that these eleven markers can be further applied for biodiversity evaluation and breeding purpose in the Icelandic cattle breed. Microsatellite analysis is a widely applied method in parentage testing and currently used for a proportion of the Icelandic horse population but it has not been incorporated in the breeding work for Icelandic cattle. If incorporated, it would increase the credibility of pedigree recording.

In addition, genomic marker data has been used to assess conservation priorities for different animal breeds. This has, however, received some criticism, based on the need to improve current methods used for assessment and to implement other important criteria such as historical and cultural values (Talle *et al*., 2005; European Cattle Diversity Consortium, 2006).

In general, comparison of different microsatellite data sources is difficult and should be considered as indirect (Sunnucks, 2000). This can be explained by the inconsistency in the microsatellites used, caused partly by the great availability of loci to choose from. Even when the same loci are used for a number of studies, difference in the microsatellite genotyping methods applied can cause various allele sizes (Freeman *et al*., 2005).

The within inbreeding coefficient for the Icelandic cattle population estimated from the microsatellite data (ranging from 8.8-9.7%) was in three cases substantially higher than when estimated for the population using pedigree information (3.5%) (Kristjánsson *et al*., 2006). It has been established that *F* for the Icelandic cattle breed increases when more restrictions are placed on pedigree data completeness (Sigurdsson & Jonmundsson, 1995; Kristjánsson *et al*., 2006). From this, it can be inferred that more complete pedigree data might yield an *F* somewhat closer to the value found in this study. Another explanation for the large F_{IS} values obtained when compared to values from pedigree records might be the fact that inbreeding coefficient and heterozygosity assessed using molecular markers does not measure the same quantity and the correlation between these two estimators has been observed as weak (Slate *et al*., 2004; Balloux *et al*., 2004). Furthermore, although conditions have been found where heterozygosity and inbreeding are closely linked, an accurate estimation of inbreeding coefficient from genetic data is regarded to require a large number of markers and strong population subdivision, small population sizes or highly skewed mating systems (Balloux *et al*., 2004). Given this, the rather large F_{IS} values obtained herein should be interpreted with caution. The N_e value found in this study

was smaller than the one found by Kristjánssson *et al*. (2006). Confidence limits were, however, wide and covered the value previously found (118).
Measures are taken in order to maintain a desirable N_e and minimizing inbreeding for the Icelandic cattle (Jónsson *et al*., 2007). A large proportion of the population is nevertheless inbred (Sigurdsson & Jónmundsson, 1995). A compromise between rapid selection response and a high N_e is considered necessary in order to minimize the loss of genetic variation (Dempfle, 1990).

The fact that the Icelandic cattle breed was not in Hardy-Weinberg equilibrium was expected given that it is a relatively small population under artificial selection (Simm, 2000).
The methods based on heterozygous excess revealed the absence of a recent bottleneck in the breed (all loci fitted the mutation drift equilibrium). A normal graphical distribution of allele frequency reinforced that result.

5.3 Sequencing of the *Leptin* gene

Neither of the SNPs found in intron 1 have been associated with altered function of the leptin protein at the molecular level or with observed phenotypic difference. Whether these polymorphisms affect the splicing mechanism of the *Leptin* gene or contribute to phenotypic difference was beyond the scope of this study. Their existence is nevertheless an indicator of genetic variation.
Regarding the SNP found in exon 2, however, the T allele seemed to have positive although insignificant (95% confidence level), effects on milk yield and protein content which is in accordance with previous results (Buchanan *et al*., 2003). This result is however based on limited records on few individuals and should be explored further.

The novel SNP found in intron 2 [+21] has not been reported previously despite being located within the area sequenced by both Konfortov *et al*. (1999) and Lagonigro *et al*. (2003). It seems to be a rare nucleotide substitution changing a guanine to adenosine. Because of its location, it does not alter the amino acid sequence of the protein directly but its effects on splicing processes can not be excluded. Nevertheless, its identification is important considering that the detection of DNA polymorphisms, whether in coding or non-coding regions of chromosomes is regarded as an important part when evaluating the genetic uniqueness of a breed (Talle *et al*., 2005). The haplotype study revealed some

connection between the A allele in intron 2 and the C allele in exon 2. A connection between traits and different haplotypes (CT-AG / CC-AG) could not be examined due to few individuals and missing record information.

5.4 Sequencing of the *DGAT1* gene

No polymorphism was found in the K232A site in the *DGAT1* gene and all heifers were homozygous for the alanine variant (GC/GC, hereafter referred to as AA). In the French Montbéliarde dairy breed the A allele is nearly fixed, a result comparable to the one presented here for the Icelandic cattle breed. The few K alleles still present in the Montbéliarde breed are thought to have been recently introduced through a limited crossbreeding event in the 1970s (Gautier *et al*., 2007). In the Swedish Red Polled breed the majority of animals were homozygous for the alanine residue and frequency of the lysine variant was very low (0.07) (Umeland, 2006). A low frequency of the K allele was also observed in Fleckvieh bulls (Thaller *et al*., 2003). The K residue positively affects content traits, mainly fat, at the cost of milk yield. Different breeding goals with respect to milk composition for different breeds in different countries can therefore explain the variations in the frequency of the K allele (Gautier *et al*., 2007) (see Table 3). The low frequency of the K allele obtained for Fleckvieh bulls was explained by genetic drift or special emphasis on milk yield in selection (Thaller *et al*., 2003). In the USA, selection has mainly been for milk yield and recently protein. This is reflected in an increased frequency of the alanine residue in the NZ Friesian population as a consequence of gene flow from the USA Holstein cattle (Spelman *et al*., 2002). The breeding goal for the Icelandic cattle breed placed the main emphasis on milk yield until in the 1990s when the emphasis was shifted to protein yield. However, there was a gain in fat content during the period 1950-1970. Since the K allele negatively affects milk yield it can be postulated that the A allele has been inadvertently selected for since 1970 by using relatively more AA sires, thereby causing a constant decrease in the frequency of the K allele and fixation of the A allele. The fact that strict selection through a number of years is generally regarded as a way to establish desirable alleles in a population (Grisart *et al*., 2002, Taberlet *et al*., 2007) supports this. It is also possible that the frequency of the K allele has been very low (even zero) already in the founder population and if present, dwindled with time perhaps following bottlenecks in the population size. Yet another interesting speculation regarding high frequency of the A allele is that dairy farmers might inadvertently prefer AA bulls when selecting sires for their cows due to the desirable effect of the A allele over the K

allele or to some other unknown favourable effects on traits like fertility (Thaller *et al*., 2003).

The findings of Schennink *et al*. (2007) suggest that the A allele is associated with a more suitable milk-fat composition than the K allele. Therefore, a high frequency of the A allele is desirable when considering milk consumption in relation to public health. From this point of view, the frequency of the A allele should be regarded as an advantage for Icelandic cattle. The fatty acid composition however needs to be further verified for the breed.

6 Conclusions

The findings presented here show that the Icelandic cattle breed still possesses a considerable amount of genetic variation, despite its long isolation and the postulated low number of founding members. The diversity indices obtained from microsatellite analysis were similar for the Icelandic cattle breed and other North European cattle breeds studied, suggesting that the genetic potential of Icelandic cattle could be similar to other North European cattle breeds.

Two genes were analyzed with respect to single nucleotide polymorphisms. Three previously known SNPs were identified in the Icelandic cattle breed and a new polymorphism was found in intron 2 of the *Leptin* gene, not previously described in other cattle breeds. This finding underlines the uniqueness of the Icelandic cattle breed.

The fact that all heifers analysed were homozygous AA for the K232A polymorphisms in the *DGAT1* gene might be traced to a limited founder population harbouring a low frequency of the K allele. Fluctuations in population size may have contributed to the loss of the K allele as might selection (unintentional) against the allele since it decreases milk yield. This however is difficult to verify.

The microsatellite analysis presented here shows that the eleven microsatellite markers tested are suitable for further research of the Icelandic cattle breed and could be implemented for breeding purposes and parentage testing. Parentage testing would add credibility to the pedigree information and thereby support progress in the breeding work.

Future studies aimed at evaluating the genetic variation of the Icelandic cattle breed are necessary incorporating animals from all areas of the country and a larger number of markers. Moreover, studies should also focus on the connection between single or few polymorphisms, like the ones reported here, and economically important traits. Such connection might benefit the breeding work directly. Finally, a thorough comparison at the genetic variation level between the Icelandic cattle breed and other cattle breeds is necessary before importing foreign genetic material for future breeding work.

In summary, this study adds to the genetic characterization of the Icelandic cattle breed.

References

Adalsteinsson, S. (1981). Origin and conservation of farm animal populations in Iceland. *Z. Tierzüchtg. Züchtgsbiol*, *98*, 258 -264.

Adamowicz, T., Flisikowski, K., Starzynski, R., Zwierzchowski, L. & Switonski, M. (2006). Mutation in the Sp1 motif of the bovine leptin gene affects its expression. *Mammalian Genome*, *17*, 77-82.

Almeida, S.E.M., Almeida E.A., Moraes, J.C.F. & Weimer T.A. (2003). Molecular markers in the *LEP* gene and reproductive performance of beef cattle. *Journal of Animal Breeding and Genetics*, *120*, 106-113.

Balloux, F., Amos, W. & Coulson, T. (2004). Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology*, *13*, 3021-3031.

Boichard, D. (2002). Pedig: A fortran package for pedigree analysis suited for large populations. Communication No. 28-13 on CD-ROM. Proc.7th World Cong. *Genetics Applied to Liveststock Production*, Montpellier, France.

Botstein, D., White, R.L., Skolnick, M. & Davis R.W. (1980). Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *American Journal of Human Genetics*, *32*, 314-331.

Bourdon, R.M. (2000). *Understanding Animal Breeding* (2nd ed.). New Jersey: Prentice Hall Inc.

Buchanan, F.C, Fitzsimmons, C.J., Van Kessel, A.G., Thue, T.D., Winkelman-Sim D.C. & Schmutz S.M. (2002). Association of a missense mutation in the bovine leptin gene with carcass fat content and leptin mRNA levels. *Genetics Selection Evolution*, *34*, 105-116.

Buchanan, F.C, Van Kessel, A.G., Waldner, C. & Christensen, D.A. (2003). *Hot Topic*: An Association Between a Leptin Single Nucleotide Polymorphism and Milk and Protein Yield. *Journal of Dairy Science*, *86*, 3164-3166.

Bændasamtök Íslands (2007). *Hagtölur landbúnaðarins 2007*. Reykjavík: Bændasamtök Íslands.

Bændasamtök Íslands (2008a) *Nautaskrá Nautastöðvar BÍ árin 2007-2008*. Reykjavík: Bændasamtök Íslands.

Bændasamtök Íslands (2008b). *Vefsvæði Bændasamtaka Íslands*. Retrieved July 9, 2008, from http://bondi.is/lisalib/getfile.aspx?itemid=327/

Cases, S., Smith, S.J., Zheng, Y.W., Myers, H.M., Lear, S.R., Sande, E., *et al*. (1998). Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 13018-13023.

Chen, H.C. & Farese, R.V. Jr. (2005). Inhibition of Triglyceride Synthesis as a Treatment Strategy for Obesity. *Arteriosclerosis Thrombosis and Vascular Biology, 25*, 482-486.

Coppieters, W., Riquet, J., Arranz, J.-J., Berzi, P., Cambisano, N., Grisart, B., *et al*. (1998). *Mammalian Genome*, *9*, 540-544.

Cornuet, J.M. & Luikart, G. (1996). Description and Power Analysis of Two Tests for Detecting Recent Population Bottlenecks From Allele Frequency Data. *Genetics*, *144*, 2001-2014.

Cymbron, T., Freeman, A.R., Malheiro, M.I., Vigne, J-D. & Bradley, D.G. (2005). Microsatellite diversity suggests different histories for Mediterranean and Northern European cattle populations. *Proceedings of the Royal Society Biological Sciences Series B*, *272*, 1837-1843.

Dempfle, L. (1990). Conservation, Creation, and Utilization of Genetic Variation. *Journal of Dairy Science*, *73*, 2593-2600.

Di Stasio, L., Brugiapaglia, A., Galloni, M., Destefanis, G. & Lisa, C. (2007). Effects of the *leptin* c.73T>C mutationon carcass traits in beef cattle. *Animal Genetics*, *38*, 315-317.

Edwards, C.J., Connellan, J., Wallace, P.F., Park, S.D.E., McCormick, F.M., Olsaker, I., *et al*. (2003). Feasibility and utility of microsatellite markers in archaeological cattle remains from a Viking Age settlement in Dublin. *Animal Genetics*, *34*, 410-416.

European Cattle Genetic Diversity Consortium (2006). *Marker-assisted conservation of European cattle breeds: an evaluation*. *Animal Genetics*, *37*, 465-481.

Eyþórsdóttir, E. & Jónmundsson J.V. (2004). Nýting erfðatækni í búfjárkynbótum. *Fræðaþing landbúnaðarins 2004*. Reykjavík: BÍ, LBH, L.r., RALA and S.r.

Eyþórsdóttir E., Tómasson Þ., & Helgadóttir Á. (2001). Erfðalindir í landbúnaði. *Ráðunautafundur 2001*. Reykjavík: BÍ, LBH, RALA.

Falconer, D.S. & Mackey, T.F.C. (1996). *Introduction to quantiative genetics* (4th ed.). Essex: Longman.

Freeman, A.R., Bradley, D.G., Nagda, S., Gibson, J.B. & Hanotte, O. (2005). Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle. *Animal Genetics*, *37*, 1-9.

Gautier, M., Capitan, A., Fritz, S., Eggen, A., Boichard, D. & Druet, T. (2007). Characterization of the *DGAT1* K232A and Variable Number of Tandem Repeat Polymorphisms in French Dairy Cattle. *Journal of Dairy Science*, *90*, 2980-2988.

Glazier, A.M., Nadeau, J.H. & Aitman, T.J. (2002). Finding Genes That Underlie Complex Traits. *Science*, *298*, 2345-2349.

Goudet, J. (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html/. Updated from Goudet J. (1995) FSTAT (vers.1.2): a computer program to calculate F-statistics. *J. Hered.*, *86*, 485-486.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., *et al*. (2002). Positional Candidate Cloning of a QTL in Dairy Cattle: Identification of a Missense Mutation in the

Bovine DGAT1 Gene with Major Effects on Milk Yield and Composition. *Genome Research*, *12*, 222-231.

Grisart, B., Farnir, F., Karim, L., Cambisano, N., Kim, J.-J., Kvasz, A., *et al*. (2004). Genetic and functional confirmation of the causality of the *DGAT1 K232A* quantitative trait nucleotide in affecting milk yield and composition. *Proceedings of the National Academy of Sciences*, *101*, 2398-2403.

Guo, X. & Elston, R.C. (1999). Linkage Information Content of Polymorphic Genetic Markers. *Human Heredity*, *49*, 112-118.

Hartl, D.L. & Clark, A.G. (1989). *Principles of Population Genetics* (2nd ed.). Sunderland, Massachusetts: Sinauer Associates, Inc.

Hill, W.G. (2000). Maintenance of quantitative genetic variation in animal breeding programmes. *Livestock Production Science*, *63*, 99-109.

Hoffmann, I., Marsan, P.A., Barker, J.S.F., Cothran, E.G., Hanotte, O., Lenstra, J.A., Milan, D., Weigend, S. & Simianer, H. (2004). *New MoDAD marker sets to be used in diversity studies for the major farm animal species: recommendations of a joint ISAG/FAO working group*. Retrieved July 8, 2008, from: http://lprdad.fao.org/cgi-bin/EfabisWeb.cgi?sid=a23f29106a1360e9ffd36b2930e37b51,refcat_50000048

Houseknecht, K.L., Baile, C.A., Matteri, R.L. & Spurlock, M.E. (1998). The Biology of Leptin: A Review. *Journal of Animal Science*, *76*, 1405-1420.

Ji, S., Willis, G.M., Scott, R.R. & Spurlock, M.E. (1998). Partial cloning and expression of the bovine Leptin gene. *Animal Biotechnology*, *9*, 1-14.

Jónmundsson, J.V. & Jónsson, M.B. (1991). Nautgripakynbætur. *Ráðunautafundur 1991*. Reykjavík: Búnaðarfélag Íslands, Rannsóknastofnun landbúnaðarins.

Jónmundsson, J.V. (2000). Þróun í nautgripasæðingum. *Freyr*, *96*(11-12), 16-18.

Jónmundsson, J.V., Kristjánsson, Þ., & Benjamínsson, B.H. (2007). Erfðaframlag þekktra kynbótagripa í íslenska kúastofninum á síðari hluta 20.aldar. In Dýrmundsson, Ó.R. (Ed.), *Íslensk búfjárrækt*. *Málstofa til heiðurs Hjalta Gestssyni nýræðum Hótel Sögu*, *Reykjavík 17. nóvember 2006* (p.21). Reykjavík: Landbúnaðarháskóli Íslands.

Jónsson, M.B. & Jónmundsson, J.V. (1974). Kynbótaskipulag fyrir íslenska kúastofninn. *Íslenzkar landbúnaðarrannsóknir*, *6*(1), 49-63.

Jónsson, M.B., Jónmundsson, J.V., & Kristjánsson Þ. (2007). Ræktunarstarf í litlum erfðahópum. In Dýrmundsson, Ó.R. (Ed.), *Íslensk búfjárrækt*. *Málstofa til heiðurs Hjalta Gestssyni nýræðum Hótel Sögu*, *Reykjavík 17. nóvember 2006* (p.33). Reykjavík: Landbúnaðarháskóli Íslands.

Jónsson, M.B. (2008). Orsakir kálfadauða hjá fyrsta kálfs kvígum. *Fræðaþing landbúnaðarins 2008*. Reykjavík: Bændasamtök Íslands, Landbúnaðarháskóli Íslands, Landgræðsla Ríkisins, Skógrækt Ríkisins, Hólaskóli, Veiðimálastofnun, Matvælastofnun, Hagþjónusta Landbúnaðarins og Matís ohf.

Kantanen, J., Olsaker, I., Adalsteinsson, S., Sandberg, K., Eythorsdottir, E., Pirhonen, K. & Holm, L.-E. (1999). Temporal changes in genetic variation of North European cattle breeds. *Animal Genetics*, *30*, 16-27.

Kantanen, J., Olsaker, I., Brusgaard, K., Eythorsdottir, E., Holm, L.-E., Lien, S., *et al*. (2000a). Frequencies of genes for coat colour and horns in Nordic cattle breeds. *Genetics Selection Evolution*, *32*, 561-576.

Kantanen, J., Olsaker, I., Holm, L.-E., Lien, S., Vilkki, J., Brusgaard, K., *et al*. (2000b). Genetic Diversity and Population Structure of 20 North European Cattle Breeds. *The Journal of Heredity*, *91*(6), 446-457.

Klungland, H., Olsen, H.G., Hassanane, M.S., Mahrous, K. & Våge, D.I. (2000). Coat colous genes in diversity studies. *Journal of Animal Breeding and Genetics*, *117*, 217-224.

Konfortov, B.A., Licence, V.E., & Miller, J.R. (1999). Re-sequencing of DNA from a diverse panel of cattle reveals a high level of polymorphism in both intron and exon. *Mammalian Genome*, *10*, 1142-1145.

Kristjánsson, Þ., Jónmundsson, J.V. & Benjamínsson, B.H. (2006). Þróun skyldleikaræktar í íslenska kúastofninum. *Fræðaþing landbúnaðarins 2006*. Reykjavík: LbhÍ, L.r., S.r.

Kühn, C., Thaller, G., Winter, A., Bininda-Emonds, O.R.P., Kaupe, B., Erhardt, G., *et al*. (2004). Evidence for Multiple Alleles at the *DGAT1* Locus Better Explains a Quantitative Trait Locus With Major Effect on Milk Fat Content in Cattle. *Genetics*, *167*, 1873-1881.

Lagonigro, R., Wiener, P., Pilla, F., Woolliams, J.A. & Williams, J.L. (2003). A new mutation in the coding region of the bovine leptin gene associated with feed intake. *Animal Genetics*, *34*, 371-374.

Li, M.H., Tapio, I., Vilkki, J., Ivanova, Z., Kiselyova, T., Marzanov, N., *et al*. (2007). The genetic structure of cattle populations (*Bos taurus*) in northern Eurasia and the neighbouring Near Eastern regions: implications for breeding strategies and conservation. *Molecular Ecology*, *16*, 3839-3853.

Liefers, S.C., te Pas, M.F.W., Veerkamp, R.F. & van der Lende, T. (2002). Association between Leptin Gene Polymorphisms and Production, Live Weight, Energy Balance, Feed Intake, and Fertility in Holstein Heifers. *Journal of Dairy Science*, *85*, 1633-1638.

Liefers, S.C., te Pas, M.F.W., Veerkamp, R.F., Chilliard Y., Delavaud C., Gerritsen R. & van der Lende, T. (2003a). Association of leptin gene polymorphisms with serum leptin concentration in dairy cows. *Mammalian Genome*, *14*, 657-663.

Liefers, S.C., Veerkamp, R.F, te Pas, M.F.W., Delavaud, C., Chilliard Y. & van der Lende, T. (2003b). Leptin Concentrations in Relations to Energy Balance, Milk Yield, Intake, Live Weight, and Estrus in Dairy Cows. *Journal of Dairy Science*, *86*, 799-807.

Liefers, S.C., Veerkamp, R.F., te Pas, M.F.W., Delavaud, C., Chilliard, Y., Platje, M. & van der Lende, T. (2005). Leptin promoter mutations affect leptin levels and performance traits in dairy cows. *Animal Genetics*, *36*, 111-118.

Liu, K. & Muse, S.V. (2005). POWERMARKER: Integrated analysis environment for genetic marker data. *Bioinformatics*, *21*, 2128-2129.

Luikart, G., Allendorf, F.W, Cornuet, J.-M. & Sherwin, W.B. (1998). Distortion of Allele Frequency Distributions Provides a Test for Recent Population Bottlenecks. *The Journal of Heredity*, *89*, 238-247.

MacHugh, D.E., Shriver, M.D., Loftus, R.T., Cunningham, P. & Bradley, D.G. (1997). Microsatellite DNA Variation and the Evolution, Domestication and Phylogeography of Taurine and Zebu Cattle (*Bos taurus* and *Bos indicus*). *Genetics*, *146*, 1071-1086.

Madeja, Z., Adamowicz, T., Chmurzynska, A., Jankowski, T. & Melonek, J. (2004). Short Communication: Effect of Leptin Gene Polymorphisms on Breeding Value for Milk Production Traits. *Journal of Dairy Science*, *87*, 3925-3927.

Maudet, C., Luikart, G. & Taberlet, P. (2002). Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis. *Journal of Animal Science*, *80*, 942-950.

Murray, B.W. (1996). The Estimation of Genetic Distance and Population Substructure from Microsatellite allele frequency data. Retrieved July 23, 2008, from http://helix.mcmaster.ca/brent/node6.html

Ólafsson, B.L., Eyþórsdóttir, E. & Hafberg, H.B. (2003). Erfðabreytileiki próteina í mjólk íslenskra kúa. *Freyr*, *99*(4), 48-53.

Pareek, C.S., Czarnik, U., Zabolewicz, T., Pareek, R.S. & Walawski, K. (2005). *DGAT1 K232A* quantitative trait nucleotide polymorphism in Polish Black-and-White cattle. *Journal of Applied Genetics*, *46*, 85-87.

Park, S. D. E. (2001). Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection (Ph.D. thesis). University of Dublin, Dublin.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945-959.

Radko, A., Żyga, A., Ząbek, T. & Słota, E. (2005). Genetic variability among Polish Red, Hereford and Holstein-Friesian cattle raised in Poland based on analysis of microsatellite DNA sequences. *Journal of Applied Genetics*, *46*(1), 89-91.

Reece, W.O. (1997). *Physiology of Domestic Animals* (2nd ed.). Baltimore: Williams & Wilkins.

Rischkowsky, B., & Pilling, D. (Eds.) (2007). *The State of the World's Animal Genetic Resources for Food and Agriculture*. Rome: FAO.

Ron, M. & Weller, J.I. (2007). From QTL to QTN identification in livestock – winning by point rather than knock-out: a review. *Animal Genetics*, *38*, 429-439.

Roslin Institute (2002). *CaDBase Genetic Diversity in Cattle*. Retrieved July 9, 2008, from http://www.projects.roslin.ac.uk/cdiv/markers.html

Rousset, F. (2007). GENEPOP'007: a complete reimplementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Notes* in press.

Schennink, A., Stoop, W.M., Visker, M.H.P.W., Heck, J.M.L., Bovenhuis, H., van der Poel, J.J., van Valenberg, H.J.F. & van Arendonk, J.A.M. (2007). *DGAT1* underlies large genetic variation in milk-fat composition of dairy cows. *Animal Genetics*, *38*, 467-473.

Scherf, B.D. (2000). *World Watch List for Domestic Animal Diversity* (3rd ed.). Rome: Food and Agricultural Organization of the United Nations.

Schwartz, M.W., Seeley, R.J., Campfield, L.A., Burn, P. & Baskin, D.G. (1996). Identification of Targets of Leptin Action in Rat Hypothalamus. *Journal of Clinical Investigation*, *98*, 1101-1106.

Sigurdsson, A. & Jonmundsson, J.V. (1995). Inbreeding and Its Impact in the Closed Population of Icelandic Dairy Cattle. *Acta Agriculturae Scandinavica Section A Animal Science*, *45***,** 11-16.

Sigurðsson, Á. (1993). *Nýja kynbótamatskerfið*. Retrieved July 10, 2008, from Bændasamtök Íslands Web site: http://www.bondi.is/

Sigurðsson, S. (1937). *Búnaðarhagir*. Reykjavík: Ríkisprentsmiðjan Gutenberg.

Simm, G. (2000). *Genetic Improvement of Cattle and Sheep*. Tonbrigde: Farming Press.

Slate, J., David, P., Dodds, K.G., Veenvliet, B.A., Glass, B.C., Broad, T.E. & McEwan, J.C. (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity*, *93*, 255-265.

Smith, J.S., Cases, S., Jensen, D.R., Chen, H.C., Sande, E., Tow, B., et al. (2000). Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat. *Nature Genetics*, *25*, 87-90.

Sodhi, M., Mukesh, M., Prakash, B., Ahlawat, S.P.S. & Sobti, R.C. (2006). Microsatellite DNA typing for assessment of genetic variability in Tharparkar breed of Indian zebu (*Bos indicus*) cattle, a major breed of Rajasthan. *Journal of Genetics*, *85*(3), 165-170.

Spelman, R.J., Ford, C.A., McElhinney, P., Gregory, G.C. & Snell, R.G. (2002). Characterization of the DGAT1 Gene in the New Zealand Dairy Population. *Journal of Dairy Science*, *85*, 3514-3517.

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology & Evolution*, *15*, 199-203.

Szyda, J. & Komisarek, J. (2007). Statistical Modeling of Candidate Gene Effects on Milk Production Traits in Dairy Cattle. *Journal of Dairy Science*, *90*, 2971-2979.

Sørensen, A.C., Sørensen, M.K. & Berg P. (2005). Inbreeding in Danish dairy cattle breeds. *Journal of Dairy Science*, *88*(5), 1865-1872.

Taberlet, P., Valentini, A., Rezaei, T., Naderi, S., Pompanon, F., Negrini, R. & Ajmone-Marsan P. (2007). Are cattle, sheep, and goats endangered species? *Molecular Ecology*, *17*, 275-284.

Talle, S.B., Chenyabuga, W.S., Fimland, E., Syrstad, O., Meuwissen, T. & Klungland H. (2005). Use of DNA technologies for the conservation of animal genetic resources: A review. *Acta Agricultural Scand Section A*, *55*, 1-8.

Tallmon, D.A., Koyuk, A., Luikart, G. & Beaumont, A. (2008). Computer programs. OnESAMP: a program to estimate effective population size using approximate Bayesian computation. *Molecular Ecology Resources*, *8*, 299-301.

Tapio, I., Värv, S., Bennewitz, J., Maleviciute, J., Fimland, E., Grislis, Z., *et al*. (2006). Prioritization for Conservation of Northern European Cattle Breeds Based on Analysis of Microsatellite Data. *Conservation Biology*, *20*(6), 1768-1779.

Thaller, G., Kramer, W., Winter, A., Kaupe, B., Erhardt, G. & Fries, R. (2003). Effects of DGAT1 variants on milk production traits in German cattle breeds. *Journal of Animal Science*, *81*, 1911-1918.

Tisdell, C. (2003). Socioeconomic causes of loss of animal genetic diversity: analysis and assessment. *Ecological Economics*, *45*, 365-376.

Torfason, J. & Jónmundsson, J.V. (2001). *Íslenska mjólkurkýrin*. Hjaltadalur: Bókaútgáfan Hofi.

Umeland, P. (2006). *Genetisk variation av betydelse för mjölkkvalitet i Rödkullerasen*. (Examensarbete 283, Institutionen för husdjursgenetik, Sveriges Lantbruksuniversitet, 2006). Retrieved from http://www.slu.se/

van der Lende, T., te Pas, M.F.W., Veerkamp, R.F. & Liefers, S.C. (2005). Leptin gene polymorphisms and their phenotypic associations. *Vitamins and Hormones*, *71*, 373-40.

Weigel, K.A. & Lin, S.W. (2002). Controlling Inbreeding by Constraining the Average Relationship Between Parents of Young Bulls Entering AI Progeny Test Programs. *Journal of Dairy Science*, *85*, 2376-2383.

Weir, B.S. (1996). *Genetic data analysis II*, Sunderland, MA: Sinauer Associates, Inc.

Winter, A., Krämer, W., Werner, F.A.O., Kollers, S., Kata, S., Durstewitz, G., *et al*. (2002). Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglyerol acyltransferase (*DGAT1*) with variation at a quantitative trait locus for milk fat content. *Proceedings of the National Academy of Sciences*, *99*, 9300-9305.

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L. & Friedman, J.M. (1994). Positional cloning of the mouse *obese* gene and its human homologue. *Nature*, *372*, 425-432.

Zhang, F., Basinski, M.B., Beals, J.M., Briggs, S.L., Churgay, L.M., Clawson, D.K., *et al*. (1997). Crystal structure of the *obese* protein leptin-E100. *Nature*, *387*, 206-209.

Appendix 1

Sample of 100 heifers, their origin and individual sample number.

Name of farm	Number of heifer	Year of birth	Number of sample
Melur	370	2005	1
Súlunes	368	2004	2
Helgavatn	330	2004	3
Stóra-Hildisey	447	2005	4
Helgavatn	320	2004	5
Skeiðháholt	666	2004	6
Akurey	557	2004	7
Ölkelda	416	2004	8
Dagverðareyri	852	2003	9
Ytra-Laugaland	1058	2004	10
Hundastapi	232	2004	11
Guðnastaðir	370	2004	12
Steinsholt	347	2004	13
Guðnastaðir	385	2004	14
Reykir	405	2004	15
Reykir	402	2004	16
Svalbarð	1109	2004	17
Hrosshagi	817	2004	18
Akurey	538	2004	19
Stóru-Reykir	329	2004	20
Hóll	415	2004	21
Hrafnkelsstaðir	190	2004	22
Stóra-Mörk	325	2004	23
Helluvað	523	2003	24
Dagverðareyri	871	2004	25
Svalbarð	1144	2004	26
Auðbrekka	403	2004	27
Keta	457	2005	28
Litli-Dunhagi	138	2003	29
Litli-Dunhagi	148	2004	30
Sakka	552	2004	31
Stóra-Hildisey	420	2004	32
Stakkhamar	170	2004	33
Geirakot	421	2005	34
Akurey	555	2004	35
Akurey	551	2004	36
Hlíðarendi	8204	2003	37
Helgavatn	1009	2004	38
S.-Völlur	237	2004	39
Eystra-Seljaland	412	2004	40
Núpur	331	2003	41
Dagverðareyri	868	2004	42
Höfði	243	2003	43
Ytra-Laugaland	9374	2004	44
Stóru-Akrar	283	2004	45
Súlunes	386	2004	46
Birtingaholt	5380	2004	47
Bakki	242	2004	48
Guðnastaðir	384	2004	49
Stóra-Mörk	331	2004	50

Name of farm	Number of heifer	Year of birth	Number of sample
Stóra-Hildisey	923	2003	51
Stóra-Hildisey	422	2004	52
Miðhjáleiga	1064	2005	53
Skeiðháholt	690	2004	54
Stóru-Reykir	335	2004	55
Stóru-Reykir	333	2004	56
Höfði	244	2003	57
Höfði	255	2004	58
Hundastapi	230	2004	59
Stóru-Reykir	334	2004	60
Daufá	313	2004	61
Sakka	576	2004	62
Stærri Bær	440	2004	63
Stóru-Reykir	337	2004	64
Hrafnkelsstaðir	199	2004	65
Hrosshagi	812	2004	66
Skeiðháholt	669	2004	67
Nes	429	2004	68
Skeiðháholt	668	2004	69
Hóll	401	2004	70
Vestri Reyn	269	2004	71
Daufá	317	2004	72
Helluvað	543	2004	73
Hundastapi	9133	2004	74
Hlíðarendi	8277	2004	75
Vestri Reyn	270	2004	76
Stærri Bær	452	2004	77
Höfði	254	2004	78
Höfði	239	2003	79
Bakki	256	2004	80
Búrfell	313	2004	81
Sakka	559	2004	82
Sakka	596	2004	83
Stóra-Mörk	317	2004	84
Vík	337	2004	85
Hvanneyri	1083	2004	86
Melur	348	2004	87
Bakki	254	2004	88
Svalbarð	1136	2004	89
Dagverðareyri	857	2004	90
Stóra-Hildisey	925	2004	91
Stóru-Akrar	9018	2004	92
Sakka	556	2004	93
Trésstaðir	361	2005	94
Steinsholt	520	2004	95
Flugumýrarhvammur	519	2004	96
Hagi	311	2004	97
Núpur	330	2004	98
Búrfell	263	2004	99
Helluvað	550	2004	100

Manuscript in preparation for Genet. Sel. Evol.

Analysis of genetic variation within the Icelandic cattle population using molecular markers

Margret Gudrun Asbjarnardottir [a], Thorvaldur Kristjansson [a], Magnus B. Jonsson [a], and Jon Hallsteinn Hallsson [a,*].

[a] Department of Land and Animal Resources, Agricultural University of Iceland, Keldnaholt, 112 Reykjavik, Iceland

*Corresponding author: jonhal@lbhi.is

Running head: Genetic variation within Icelandic cattle

Abstract - The Icelandic cattle breed (*Bos taurus*) is believed to have originally been brought to Iceland from Norway over 1000 years ago or around the time of settlement. Since then the breed is believed to have been almost completely isolated and is known to have gone through large fluctuations in population size. To assess the current genetic variation within the Icelandic cattle breed microsatellite markers were used on a sample of 100 heifers shown to be representative for the population. Measures of genetic variability, such as mean number of alleles, mean observed and expected heterozygosity, and mean polymorphism information content showed substantial genetic variation within the breed. Average inbreeding coefficient (F_{IS}) and mean effective population size (N_e) was also estimated. Our findings do not support the occurrence of recent genetic bottleneck in the Icelandic cattle breed and our analysis suggest that the population is uniform in its genetic variation, suggesting that the current breeding scheme has broken the isolation of any previous subpopulations caused by geographical barriers to gene flow.

Icelandic cattle / microsatellites / genetic variation / effective population size

1. INTRODUCTION

Genetic variation within domestic breeds is a very important asset and is the fuel in all breeding work. It gives breeders the opportunity to improve the traits present or to develop new characteristics in response to changes in the environment, new diseases or market demands [6, 16, 26]. Despite of this, domestic animal diversity is declining substantially in global terms [20]. This has been seen as the result of market extensions and economic globalization which calls for an increase in the use of fewer, high-output breeds, specialization in production along with decoupling of agriculture from natural environment [29].

Microsatellites are a widely used marker system for genetic characterization due to their polymorphic nature combined with their abundance in the genome. Due to their informative value, they have been applied in various studies aimed at domestic animals, both to evaluate genetic relationships between different breeds and also to estimate genetic diversity within single population [3, 10, 13, 16, 19, 23, 28].

The Icelandic cattle breed is the only dairy breed in Iceland and has been so for over 1000 years or since the settlement. No major attempts have been made to import dairy cattle and the Icelandic cattle breed is therefore regarded as a genetically closed population. Moreover, it is one of a few indigenous cattle breeds remaining in North Europe that still comprise a large population size and a well defined purpose. The breed has a

63 reasonably well documented history and pedigree records, but information
64 from the field of molecular genetics is deficient. Today, the population
65 consists of approximately 69.000 individuals of which less than half is
66 recorded as milking cows. The main objective of this study was to gain
67 additional information regarding the genetic makeup of the Icelandic cattle
68 breed. Only a few studies have been conducted for Icelandic cattle using
69 molecular methods and in all cases the analysis was aimed at several
70 different breeds rather than focusing on the breed alone [9-11, 28].
71 Moreover, these studies used the same sample collection and differ only in
72 the markers used. A prerequisite for this study was to establish a new,
73 reasonably large and unbiased collection of DNA samples representative for
74 the breed.
75 Information gained from this study will increase our understanding of the
76 breed's population structure with a potential usefulness regarding breeding
77 work as well as contribute to the global database of cattle genetic resources.
78 In addition, these results will be an input into the ongoing discussion
79 whether the Icelandic cattle breed has the necessary genetic variation to be a
80 sustainable dairy production breed or whether it should be substituted with
81 another higher-yielding commercial breed. Such substitution of native cattle
82 breeds for more commercial ones is a development observed in other
83 Scandinavian countries [10] and can be seen as a trend towards lesser
84 genetic diversity.

2. MATERIALS AND METHODS

2.1 Sampling and DNA extraction

Blood samples from 440 Icelandic heifers were collected from 58 dairy farms (see Fig. 1 for distribution of farms) covering regions of Iceland where numbers of milking cows is estimated as 84.1% of all milking cows and approximately 80% of the whole population. One hundred samples were chosen randomly from the sample collection and used for the molecular analysis. The majority of the heifers (98.4%) included in the study were born in the years 2003-2005. The relationship within and between the heifers analyzed and the whole population was examined as well the main sires in common with these groups in order to see if the individuals analyzed were representative for the breed as a whole.

Genomic DNA was extracted from buffy coat using a standard MasterPure™ DNA Purification Kit commercially available from EPICENTRE ® Biotechnologies.

2.2 Microsatellite analysis

The following eleven bovine microsatellite markers were used to genotype the heifers: *BM1824*, *BM2113*, *ETH10*, *ETH225*, *ETH3*, *INRA23*, *SPS115*, *TGLA122*, *TGLA126*, *TGLA227*, and *TGLA53*. The markers are distributed throughout the bovine genome, covering 11 of the 29 autosomes. All are jointly recommended by the Food and Agricultural Organization of the

106 United Nations (FAO) and the Internal Society for Animal Genetics (ISAG)
107 to be used in analysis of genetic diversity in cattle [7]. In addition, nine of
108 the eleven microsatellites used in this study are recommended by
109 International Society for Animal Genetics for parentage testing. The heifers
110 were genotyped for the microsatellites in a single multiplex at Eurofins
111 Medigenomix GmbH (http://www.medigenomix.de/en/index.html) applying
112 the ABI StockMarks Cattle® Bovine Genotyping Kit.

113 2.3 Analysis of molecular data

114 Basic diversity indices, revealing variability at the DNA level, included
115 mean number of alleles (MNA), allele frequencies, observed and expected
116 heterozygosity (H_O and H_E, respectively) as well as the polymorphism
117 information content (PIC) of Botstein *et al*. [1]. The higher the PIC value,
118 the more informative the marker and a locus with a PIC value higher than
119 0.5 is regarded as highly informative. A locus was defined as polymorphic if
120 the most common allele had a frequency of less than 0.95 [5]. All basic
121 diversity indices were calculated using the POWERMARKER package [14].
122 The within population inbreeding estimates (F_{IS}) were calculated for each
123 locus and overall loci using the FSTAT program version 2.9.3.2 [4]. The
124 significance level was set as: $p<0.05$. ONeSAMP 1.0 [27] was applied to
125 estimate the effective population size (N_e) of the Icelandic cattle breed using
126 the microsatellite data.

FSTAT and Structure 2.0 [18 2000] were used in order to examine population subdivision. For the FSTAT analysis the sample of 100 heifers was divided into two groups according to their geographical location; on one hand individuals from the Northern side and on the other hand individuals from the South-Western part of the country.

Allelic richness was also estimated separately for the two groups.

Exact test for deviations from Hardy-Weinberg equilibrium (HWE) per locus and per population was performed using the GENEPOP program version 4 [21] applying a Markov chain to compute unbiased estimates of the exact probabilities (p-values) of being wrong in rejecting HWE.

The Excel Microsatellite Toolkit version 3.1 [17] was used to transform data to a format acceptable by the GENEPOP and FSTAT programs.

Two methods were employed in order to estimate if the Icelandic cattle breed has experienced a recent reduction in effective population size (N_e) or a genetic bottleneck. A recently bottlenecked population is here defined according to Luikart *et al*. [15] as a population bottlenecked within the past few dozen generations. The first method was based on heterozygosity excess as described by Cornuet and Luikart [2]. Two different statistical tests were applied, a sign test, and a Wilcoxon test. These tests were applied under three following models of microsatellite evolution, the infinite allele model (IAM), the stepwise mutation model (SMM), and the two-phased model of mutation (TPM).

149 The second approach involved the graphical method of Luikart *et al*. [15]
150 stating that in a non-bottlenecked population alleles at low frequency (0.0-
151 0.1) are always more abundant that alleles at more intermediate frequency.
152 A mode-shift distribution of allele frequency (i.e. fewer alleles in low
153 frequency classes compared with intermediate frequency classes) is
154 expected in recently bottlenecked populations.
155 Both of the above approaches were carried out using the computer software
156 program BOTTLENECK
157 (http://www.montpellier.inra.fr/URLB/bottleneck/bottleneck.html)
158 performing 5000 replicates.
159

160 ## 3. RESULTS

161 ### 3.1 Analysis of sample collection

162 Shared male ancestors (fathers or grandfathers of individual sires) and
163 similar relationship values (approximately 0.03) between the sample
164 collection and the population supported the prerequisite of a sample
165 representative for the population regarding gene pool (data not shown).

166 ### 3.2 Genetic diversity indices

167 Of the 100 heifers, 91 were successfully analyzed with the exception of one
168 marker for one sample. The number of observed alleles, size range of PCR
169 products, observed (H_O) and expected (H_E) heterozygosity, and PIC values

are presented in Table I. Allele frequency distribution for each microsatellite marker is shown in Figure 2.

A total of 68 alleles were detected across the 11 loci giving a mean number of 6.182 alleles per locus. All the loci were polymorphic (data not shown). The number of observed alleles ranged from 4 (*ETH3*) to 9 (*TGLA53*). Mean observed heterozygosity was 0.626 and mean expected heterozygosity was 0.685.

PIC values exceeded 0.5 for all loci except *SPS115*, which had a PIC value of 0.462. No statistically significant difference was obtained between the two groups comparing values for allelic richness, estimated as 4.305 and 4.414 for the Northern and South-Western groups, respectively, observed and expected heterozygosity and average relatedness ($p>0.05$).

The inbreeding coefficient for the Icelandic cattle population estimated over all loci was 9.7% according to Weir and Cockerham's F_{IS} (Tab. II). The N_e of the Icelandic cattle breed was estimated to be 111 individuals (lower and upper 95 % credible limits ranging from 99.76 to 127.39).

Three loci (*ETH225*, *TGLA122*, and *TGLA53*) gave significant ($p<0.05$) deviations from Hardy-Weinberg equilibrium when the exact test was applied. Furthermore, the Icelandic cattle breed deviated significantly from the Hardy-Weinberg equilibrium when examined over all loci. Both the method based on heterozygous excess and the graphical distribution of allele frequency revealed the absence of a bottleneck in the Icelandic cattle

192 breed (Tab. II and Fig. 3). Therefore the null hypothesis of mutation drift
193 equilibrium was not rejected in the breed.

194 4. DISCUSSION

195 The results of this study suggest a considerable level of genetic variation
196 within the Icelandic cattle breed despite its long isolation. Moreover, these
197 findings are of value for the current breeding work; the most obvious one
198 being the potential use of the eleven microsatellites markers for parentage
199 testing (all PIC values were high). Such analysis is presently used for a
200 proportion of the Icelandic horse population but has not yet been
201 incorporated in the breeding work for the Icelandic cattle. It is likely to add
202 credibility to pedigree information thereby supporting an ongoing progress
203 in the breeding work.
204 Both the values obtained for mean observed (H_O=0.626) and expected
205 (H_E=0.685) heterozygosity were quite larger than the ones for the Icelandic
206 cattle breed found by Kantanen *et al*. [11] but more similar to the ones
207 found by Tapio *et al*. [28]. Further comparison using values from Kantanen
208 *et al*. [11] and Tapio *et al*. [28] reveals that H_O and H_E values for the major
209 Scandinavian cattle breeds (Norwegian Red, modern Swedish Red-and-
210 White, modern Swedish Friesian, Finnish Ayrshire and Finnish Holstein
211 Friesian) have been found to range between 0.63-0.72 and 0.62-0.70,
212 respectively (numbers not available for the three most prevalent Danish
213 breeds). The values for allelic richness observed were little higher than the

214 value calculated for the Icelandic cattle breed by Tapio *et al*. [28] (3.948)
215 but a little lower than the values from same study obtained for the five
216 previously mentioned Nordic cattle breeds. Collectively, these results
217 indicate a similar level of genetic variation for the Icelandic cattle breed and
218 major commercial breeds in Scandinavia. These results can therefore be
219 interpreted as a valuable supplement to the discussion on the conservation
220 role of the breed and whether it should be substituted with another higher-
221 yielding commercial breed. Due to the demand for more economical breeds,
222 the development in Scandinavia has for the last decades led to extensive
223 displacement of the local cattle breeds by commercial breeds; either red-
224 and-white Ayrshire or black-and-white Holstein-Friesian based [11]. The
225 rather high genetic variation found within the Icelandic cattle breed,
226 estimated with standard diversity indices as well as effective population
227 size, indicate a possibility for further development and progress in breeding
228 work. The absence of recent bottlenecks in the breed supports this view.
229 The fact that no obvious subdivision of the sample was observed might be
230 explained by the finding of Kristjánsson *et al*. [12], that the current
231 relationship of cattle within and between different geographical regions of
232 Iceland has been observed to be more or less the same. This is considered to
233 be a result of the collective breeding work and artificial insemination
234 beginning around 1970 and now including little less then 80% of producing
235 cows [8]. Therefore, it can be concluded that the Icelandic cattle breed is

236 genetically uniform, despite its relatively large geographical distribution;
237 this pattern is similar to what has recently been found true for the Icelandic
238 horse population [Hreidarsdottir, personal communication].
239 The inbreeding status for the Icelandic cattle breed found in this study is
240 higher than has previously been estimated for the population using pedigree
241 information (a value of 3.5%) [12]. This discrepancy might be explained, at
242 least partly by the level of restriction placed on the completeness of the
243 pedigree data. It has been established that the inbreeding coefficient for the
244 Icelandic cattle breed based on pedigree data increases when more
245 restrictions are placed on pedigree data completeness [12, 22].
246 Results revealing a N_e of 111 individuals for the Icelandic cattle are
247 concordant with what has formerly been found for the breed using pedigree
248 data; estimating N_e as 118 individuals for the interval of 1995-2000 and as
249 146 individuals 10 years earlier [12]. As a guideline, it is estimated that the
250 threshold of N_e is between 50 and 100, below which the fitness of the
251 population decreases steadily [20]. Effective population size of the Icelandic
252 cattle breed is therefore considered to be at an agreeable level and
253 substantially higher than values found for many other commercial cattle
254 breeds with larger census sizes; e.g. the French Holstein breed (N_e = 46), the
255 USA Holstein breed (N_e =39), Danish Holstein (N_e = 49), Danish Jersey (N_e
256 = 53), and Danish Red (N_e = 47) [24, 25].

To conclude, the result presented here indicate a considerable amount of genetic variation within the Icelandic cattle breed despite of its long geographical isolation and progressive breeding methods applied for the breed in later decades.

ACKNOWLEDGEMENT

Financial support for this study was provided by the Icelandic Board of Genetic Resources in Agriculture and the Icelandic Agricultural Productivity Fund. Gunnfríður E. Hreiðarsdóttir receives gratitude for useful conversation regarding the Icelandic cattle breed.

267 REFERENCES

268 [1] Botstein D., W.R.L., Skolnick M., and Davis R.W., *Construction of*
269 *a Genetic Linkage Map in Man Using Restriction Fragment Length*
270 *Polymorphisms.* Am. J. Hum. Genet., 1980. **32**: p. 314-331.
271 [2] Cornuet, J.M.a.L., G., *Description and Power Analysis of Two Tests*
272 *for Detecting Recent Population Bottlenecks From Allele Frequency Data.*
273 Genetics, 1996. **144**: p. 2001-2014.
274 [3] Cymbron T., F.A.R., Malheiro M.I., Vigne J-D., and Bradley D.G.,
275 *Microsatellite diversity suggests different histories for Mediterranean and*
276 *Northern European cattle populations.* Proc. R. Soc. Biol. Sci. Ser. B.,
277 2005. **272**: p. 1837-1843.
278 [4] Goudet, J., *Fstat vers. 1.2: a computer program to calculate F-*
279 *statistics.* J. Hered., 1995. **86**: p. 485-486.
280 [5] Hartl, D.L.a.C., A.G., *Principles of Population Genetics (2nd ed.).*
281 Sunderland, Massachusetts: Sinauer Associates, Inc., 1989.
282 [6] Hill, W.G., *Maintenance of quantitative genetic variation in animal*
283 *breeding programmes.* Livest. Prod. Sci., 2000. **63**: p. 99-109.
284 [7] Hoffmann, I., Marsan, P.A., Barker, J.S.F., Cothran, E.G., Hanotte,
285 O., Lenstra, J.A., Milan, D., Weigend, S., and Simianer, H., *New MoDAD*
286 *marker sets to be used in diversity studies for the major farm animal*
287 *species: recommendations of a joint ISAG/FAO working group.* 2004.

288 [8] Jónmundsson, J.V., *Þróun í nautgripasæðingum.* Freyr, 2000. **96(11-**
289 **12)**: p. 16-18.
290 [9] Kantanen, J., Olsaker, I., Adalsteinsson, S., Sandberg, K.,
291 Eythorsdottir, Pirhonen, K., Holm, L.-E., *Temporal changes in genetic*
292 *variation of North European cattle breeds.* Anim. Genet., 1999. **30**: p. 16-
293 27.
294 [10] Kantanen, J., Olsaker, I., Brusgaard, K., Eythorsdottir, E., Holm, L.-
295 E., Lien, S., Danell, B., Adalsteinsson, S., *Frequencies of genes for coat*
296 *colour and horns in Nordic cattle breeds.* Genet. Sel. Evol., 2000. **32**: p.
297 561-576.
298 [11] Kantanen, J., Olsaker, I., Holm, L.-E., Lien, S., Vilkki, J., Brusgaard,
299 K., Eythorsdottir, E., Danell, B., Adalsteinsson, S., *Genectic Diversity and*
300 *Population Structure of 20 North European Cattle Breeds.* J. Hered., 2000B.
301 **91**: p. 446-457.
302 [12] Kristjánsson Þ., J.J.V.a.B.B.H., *Þróun skyldleikaræktar í íslenska*
303 *kúastofninum [Icelandic].* Fræðaþing landbúnaðarins 2006, 2006: p. 133-
304 139.
305 [13] Li M.H., T.I., Vilkki J., Ivanova Z., Kiselyova T., Marzanov N.,
306 Cinkulov M., Stojanovic S., Ammosov I., Popov R., and Kantanen J., *The*
307 *genetic structure of cattle populations (Bos taurus) in northern Eurasia and*
308 *the neighbouring Near Eastern regions: implications for breeding strategies*
309 *and conservation.* Mol. Ecol., 2007. **16**: p. 3839-3853.

[14] Liu, K.M., S.V., *PowerMarker: an integrated analysis environment for genetic marker analysis.* Bioinformatics (Oxf.), 2005. **21**(9): p. 2128-2129.

[15] Luikart, G., Allendorf, F.W., Cornuet, J.-M. & Sherwin, W.B., *Distortion of Allele Frequency Distributions Provides a Test for Recent Population Bottlenecks.* J. Hered., 1998. **89**: p. 238-247.

[16] Maudet C., L.G., and Taberlet P., *Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis.* J. Anim. Sci., 2002. **80**: p. 942-950.

[17] Park, S.D.E., *Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection (Ph.D. thesis).* University of Dublin, Dublin., 2001.

[18] Pritchard, J.K., Stephens, M., and Donnelly, P., *Inference of population structure using multilocus genotype data.* Genetics, 2000(155): p. 945-959.

[19] Radko A., Z.A., Zabek T. and Slota E., *Genetic variability among Polish Red, Hereford and Holstein-Friesian cattle raised in Poland based on analysis of microsatellite DNA sequences.* J. Appl. Genet., 2005. **46**(1): p. 89-91.

[20] Rischkowsky, B., and Pilling D., edited by. Rome., *FAO: The State of the World's Animal Genetic Resources for Food and Agriculture,.* 2007.

[21] Rousset, F., *GENEPOP'007: a complete reimplementation of the GENEPOP software for Windows and Linux.* Molecular Ecology Notes, 2007. **In press**.

[22] Sigurdsson, A.a.J., J.V., *Inbreeding and Its Impact in the Closed Population of Icelandic Dairy Cattle.* Acta Agric. Scand. Sect. A Anim. Sci., 1995(45): p. 11-16.

[23] Sodhi M., M.M., Prakash B., Ahlawat S.P.S., and Sobti R.C., *Microsatellite DNA typing for assessment of genetic variability in Tharparkar breed of Indian zebu (Bos indicus) cattle, a major breed of Rajasthan.* J.Genet.., 2006. **85**(3): p. 165-170.

[24] Sørensen A.C., S.M.K., and Berg P., *Inbreeding in Danish dairy cattle breeds.* J. Dairy Sci., 2005. **88**(5): p. 1865-1872.

[25] Taberlet P., V.A., Rezaei H.R., Naderi S., Pompanon F., Negrini R. & Ajmone-Marsan P., *Are cattle, sheep and goats endangered species?* Mol. Ecol. (2007), 2007: p. 1-10.

[26] Talle S.B., C.W.S., Fimland E., Syrstad O., Meuwissen T., and Klungland H., *Use of DNA technologies for the conservation of animal genetic resources: A review.* Acta Agric. Scand. Sect. A Anim. Sci., 2005. **55**: p. 1-8.

[27] Tallmon, D.A., Koyuk A., Luikart, G. & Beaumont M.A., *ONeSAMP: a program to estimate effective population size using*

352 *approximate Bayesian computation.* Mol. Ecol. Resources, 2008. **8**(299-
353 301).
354 [28] Tapio I., V.S., Bennewitz J., Maleviciute J., Fimland E., Grislis Z.,
355 Meuwissen T.H.E., Miceikiene I., Olsaker I., Viinalass H., Vilkki J. and
356 Kantanen J., *Prioritization for Conservation of Northern European Cattle*
357 *Breeds Based on Analysis of Microsatellite Data.* Conserv. Biol., 2006.
358 **20**(6): p. 1768-1779.
359 [29] Tisdell, C., *Socioeconomic causes of loss of animal genetic*
360 *diversity: analysis and assessment.* Ecol. Econ., 2003. **45**: p. 365-376.

361 FIGURE CAPTIONS

362 Figure 1: Map of Iceland showing the distribution of farms where samples
363 were collected, farms included in the sample of 100 are coloured green. In
364 parenthesis are the numbers of milking cows in each region shown as
365 percentage of the total number of milking cows in Iceland.

366

367 Figure 2: Frequency distribution of alleles for the 11 microsatellite markers.

368

369 Figure 3: Distribution of allele frequency (mode-shift analysis) of 11
370 microsatellites in the Icelandic cattle breed.

371

372 Table I: Diversity indices calculated for 91 Icelandic heifers. Number of
373 observed alleles (N_O), size range in base pairs (bp), heterozygosity (H_O
374 observed, H_E expected), polymorphism information content (PIC) and
375 within population inbreeding estimates (F_{IS}). The highest and lowest values
376 for H_E and H_O and underlined. * $p<0.05$ for F_{IS} within samples.

Marker	**Chr**	**N_O**	**Size range (bp)**	**H_O**	**H_E**	**PIC**	**F_{IS}**
BM1824	1	5	178-188	0.571	0.619	0.545	0.088
BM2113	2	8	123-143	<u>0.769</u>	<u>0.776</u>	0.753	0.020
ETH10	5	6	213-225	0.528	0.581	0.555	0.102
ETH225	9	5	140-150	<u>0.462</u>	0.681	0.631	0.332*
ETH3	19	4	117-127	0.736	0.725	0.681	-0.004
INRA23	3	6	192-216	0.648	0.708	0.665	0.096
SPS115	15	5	248-260	0.511	<u>0.491</u>	0.462	-0.031
TGLA122	21	7	141-173	0.604	0.740	0.705	0.194*
TGLA126	20	6	113-125	0.758	0.751	0.715	0.001
TGLA227	18	7	81-101	0.714	0.702	0.658	-0.006
TGLA53	16	9	154-178	0.582	0.756	0.728	0.240*
Mean		**6.182**		**0.626**	**0.685**	**0.645**	**0.097**

Table II: Estimation of loci with heterozygosity excess and probabilities obtained from three microsatellite evolution models for bottleneck test in the Icelandic cattle breed (*deviation from the mutation drift equilibrium $p<0.05$).

Test/Model	**IAM**	**SMM**	**TPM**
Sign test: number of loci with heterozygosity excess (probability)	*Expected* = 6,45 (0,02559)* *Observed* = 10	6,55 (0,106455) 4	6,51 (0,10824) 9
Wilcoxon rank test (probability of heterozyogsity excess)	0,00098*	0,27832	0,06738

Authors comment: Ideal position of figures and tables

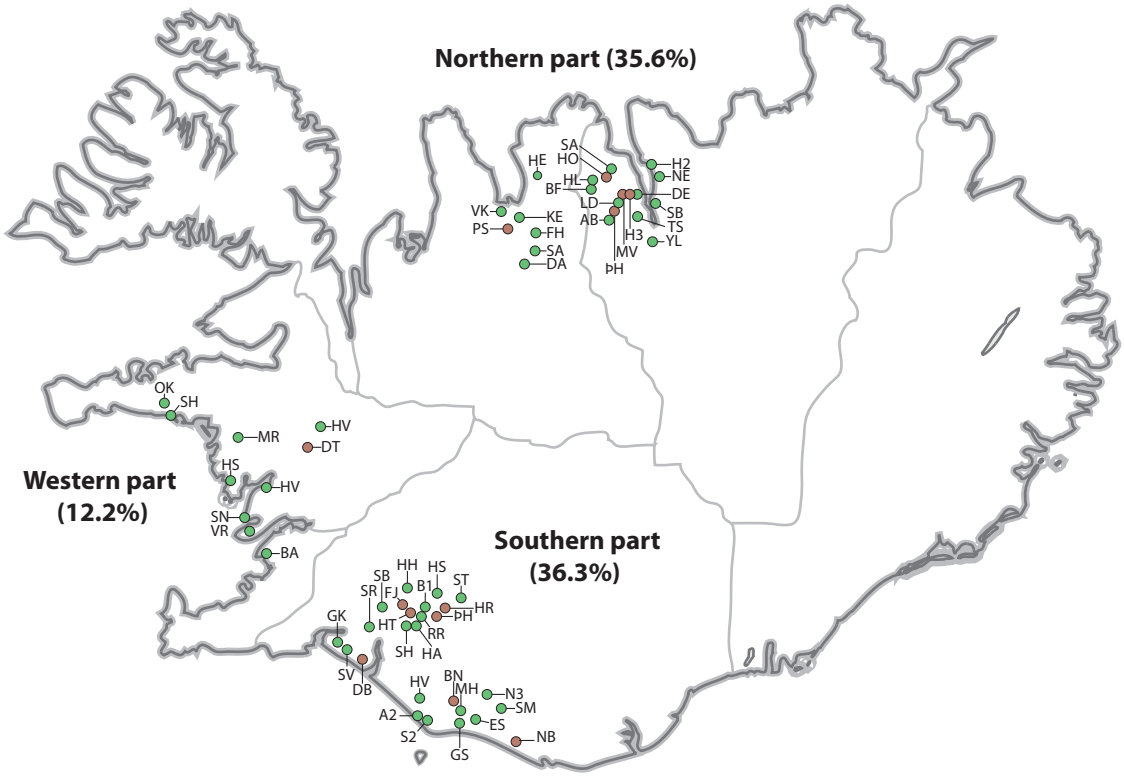
Figure 1: In the beginning of section 2.1

Figure 2: After the first paragraph in section 3.2

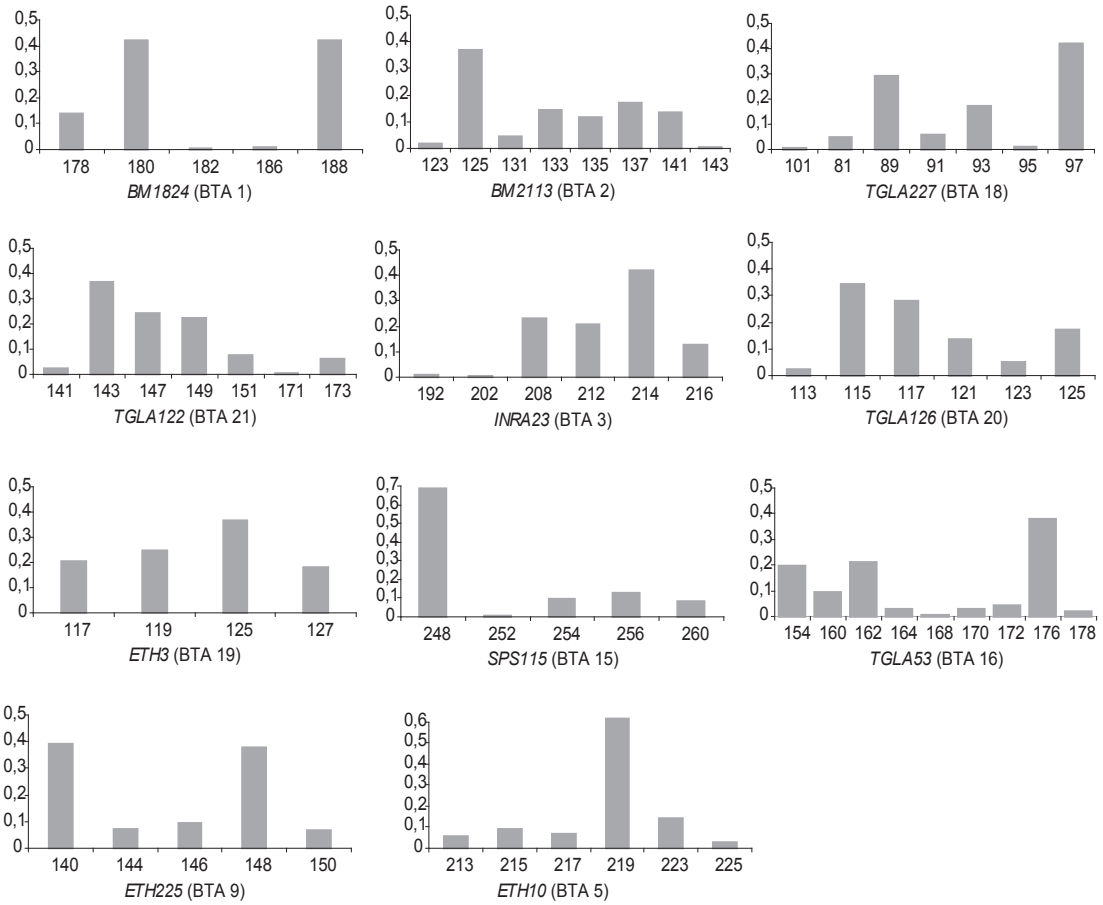
Figure 3: In the end of section 3.2

Table I: After the first paragraph in section 3.2

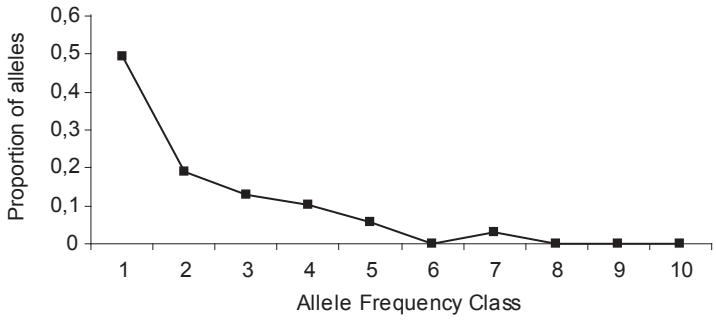
Table II: In the end of section 3.2



Asbjarnardottir et al., Fig. 1



Asbjarnardottir et al., Fig. 2



Asbjarnardottir et al., Fig. 3

Manuscript in preparation for Genet. Sel. Evol.

Genetic analysis of the Icelandic cattle breed with respect to single nucleotide polymorphisms in the *Leptin* and *DGAT1* genes

Margret Gudrun Asbjarnardottir[a] and Jon Hallsteinn Hallsson[a,*]

[a] Department of Land and Animal Resources, Agricultural University of Iceland, Keldnaholt, 112 Reykjavik, Iceland

*Corresponding author: jonhal@lbhi.is

Running head: Polymorphisms in the Icelandic cattle breed

Abstract – The existence of several single nucleotide polymorphisms in regions of the *Leptin* and *DGAT1* genes was explored in the Icelandic cattle breed by DNA sequencing. The breed has been geographically and genetically isolated for well over 1000 years. Here we report the first assessment of SNP frequency conducted for the breed for these two genes. Analysis of the *DGAT1* gene showed all the heifers to be homozygous for the alanine variant, suggesting fixation of the Ala232 variant in exon 8 of the *DGAT1* gene or at least very low levels of the Lys232 variant in the breed. Sequencing results further revealed the presence of three out of five previously identified SNPs in the *Leptin* gene in the Icelandic cattle breed. Also, a new polymorphism, previously unknown in other cattle breeds, was identified in intron 2 of the *Leptin* gene. This new SNP is a guanine to adenine substitution at location +21 (*G/A* +21). Together; these results indicate some level of genetic variation present within the breed despite its long period of isolation and fluctuation in population size.

single nucleotide polymorphism / Icelandic cattle / leptin / DGAT1

INTRODUCTION

The dairy industry in Iceland is based on one indigenous breed; the Icelandic cattle breed. Due to the geographical location and strict limits on import of foreign genetic material, the breed has remained almost completely genetically isolated since the immigration of the island around 900 AD. Information on general characteristics of the Icelandic cattle breed is provided in Adalsteinsson [1]. At present, no molecular genetic data is systematically gathered and used for breeding or research purpose in the Icelandic cattle despite progressive breeding methods based on knowledge at the genetic level and the general deficiency of such information for the breed.

Genetic polymorphisms in genes with various metabolic function have been associated, either directly or indirectly, with milk production and composition in ruminants [2, 4, 14]. In fact, the identification of genes underlying genetic variability of milk production traits with the potentials to be implemented in breeding programs is considered a major objective of today dairy cattle genomics [14]. Examples of such candidate genes in cattle are the *Leptin* and *DGAT1* genes, widely studied due to their association with economically important traits like energy balance (*Leptin*) and milk composition (*DGAT1*). A number of polymorphisms have been identified in the *Leptin* gene [8] and some associated with traits such as feed intake and milk yield [3, 9].

The *DGAT1* gene was found to be both a functional [12] and positional candidate for a QTL effect found for milk production on bovine autosomal chromosome 14 (BTA14) [5]. A non-conservative dinucleotide substitution in exon 8, leading to an amino acid change Lysine 232 to Alanine (often referred to as *K232A*) has been suggested as the causative mutation for the effect [7, 18]. Since its identification, the *K232A* polymorphism has been associated with milk yield and composition in various cattle breeds [7, 11, 13, 16].

The main objectives of this study was to genotype a sample of the Icelandic cattle breed for a number of several single nucleotide polymorphisms in regions of the *Leptin* and *DGAT1* genes, including the site of the *K232A* polymorphism in *DGAT1*. So far, little research has been carried out in the field of molecular genetics for this particular breed. Our work was done in order to see if polymorphisms identified and associated with commercial traits in other dairy breed exist in the Icelandic cattle breed and explore their frequency in the population as well as search for other polymorphisms that could be distinctive for the breed. These findings therefore add to the genetic characterization of the breed and might also be beneficial to the breeding work.

MATERIALS AND METHODS

1.1 Animals

Blood samples of 440 Icelandic heifers at first pregnancy were collected from 58 dairy farms covering regions of Iceland where dairy farming is substantial. 100 samples were chosen randomly from the pool of 440 and used for analysis. The majority of the heifers included in the study were born in 2003, 2004, and 2005 (98.4%). Efforts were made to evaluate the relationship of the heifers in order to see if the sample were typical for the breed as a whole. Calculation of coefficient of relationship (R) and examination of sires in common yielded prove for a representative sample (data not shown).

1.2 DNA extraction and primers

Genomic DNA was extracted from buffy-coat using a standard MasterPure™ DNA Purification Kit commercially available from EPICENRE ® Biotechnologies. Reported sequences of the *Leptin* gene (Genbank Accession number AY138588 (Lagonigro *et al*. 2003), refGene NM_173928) and the *DGAT1* gene (Genbank accession number AJ318490 (Winter *et al*. 2002)) were available at the NCBI and UCSC databases (http://www.ncbi.nlm.nih.gov and http://genome.cse.ucsc.edu). Primers were designed using Primer 3 (http://frodo.wi.mit.edu/cgi-

bin/primer3/primer3_www.cgi). The primers were synthesized at MWG Biotech (www.mwg-biotech).

The *Leptin* gene amplicon was 598 bp and spanned exon 2, and parts of introns 1 and 2 using primers BtLepEx2_F (5´-ACA CCT CCT GTG GTT TTC TTG ATT CCG-3') and BtLepEx2_R (5´-GGC ACT AGG ATT CCG GTC TGG-3'). Five SNPs have been reported within this region [8, 9], (see Figure 1 for location of primers and polymorphic sites found in the *Leptin* gene). The *DGAT1* gene amplicon was 443 base pairs, including exons 7, 8, and 9, and introns 7 and 8, covering the site of *K232A* and was amplified using following primers: BtDGAT1_F (5´-TGC TGG CCC TGA TGG TCT ACA CCA TC-3') and BtDGAT1_R (5´-GTC GCC GCA GCA GGA AGC GCT TTC G-3').

1.3 PCR reaction and sequencing preparation

PCR was carried out in a 25 µl volume containing approximately 15 ng of genomic DNA, 10 pmol of each primer and 12.5 µl of Taq 2x Master Mix (as supplied by New England BioLabs).

The PCR program used to amplify both *Leptin* and *DGAT1* genes was as follows: initial denaturation for 4 min at 94°C for one cycle; denaturation for 30 sec at 94°C, annealing for 45 sec at 52°C, extension for 2 min at 72°C for 35 cycles; and final extension step for 4 min at 72°C for one cycle. A Px2 Thermal Cycler (Thermo Electron Corporation) was used to perform the amplifications.

PCR products were run on 1% agarose gel stained with ethidum bromide and visualized by UV light before purification, either directly or after gel electrophoresis using a NucleoSpin® Extract II PCR clean-up/Gel extraction kit according to the manufacturer's manual, with DNA eluted in 40 µl of elution buffer. The purified PCR product was checked by running it on a 1% agarose gel. The correct primers were added to the samples before being sent to MWG Biotech for sequencing.

Sequences were analyzed using the Vector NTI Advance® Software from InvitrogenTM. Haplotypes were examined using POWERMARKER computer package [10].

RESULTS

A total number of 97 heifers were successfully sequenced for the *DGAT1* fragment and all turned out to be homozygous for the Ala232 variant at the *K232A* site and so the *A* allele seems fixed in the breed. Around 60 heifers were successfully sequenced for the *Leptin* gene fragment and four polymorphisms, all bi-allelic single nucleotide substitution (SNPs) were detected. Three SNPs were located in the introns flanking exon 2 (*C/T* at Int1 -102, *C/G* at Int1 -79 and *G/A* at Int2 +21) but one was located in exon 2 (*C/T* at Ex2 +101). The last one causes a non-conservative amino acid substitution, from arginine (Cgc) to cysteine (Tgc). Three of the SNPs found in this study have already been identified before [8] but the *G/A* substitution

located in intron 2 has not been reported previously despite being located within the area sequenced by both Konfortov *et al*. [8] and Lagonigro *et al*. [9]. No animals were identified homozygous *A/A* for the substitution. An *A/T* substitution in exon 2 (Ex2 +48) described by Lagonigro *et al*. (2003) was not found in the Icelandic cattle, all the heifers sequenced were homozygous *A/A* (tyrosine).

Figure 1 gives an overview of the *Leptin* gene with identified polymorphic sites and the frequency of homo- and heterozygotes for the sites located in this study.

Examination of haplotypes for polymorphic sites in exon 2 and intron 2 revealed that the *A* variant in intron 2 was often inherited along with the *C* allele in exon 2 (frequency of the *C-A* haplotype was 0.05). Of the six heifers heterozygote *A/G*, four were homozygote *C/C* and the remaining two heterozygote *C/T*.

DISCUSSION

Here we report for the first time a new SNP in the *Leptin* gene of the Icelandic cattle breed which appears to be unique for the breed. Being located in a non-protein coding region, this mutation does not alter the amino acid sequence of the protein directly but due to its location close to the splice donor of intron 2 it might effect splicing of the *Leptin* gene.

Neither one of the SNPs found in intron 1 of the *Leptin* gene have been

associated with altered function of the protein at the molecular level or with observed phenotypic difference concerning traits. The existence of these four SNPs nevertheless indicates a level of genetic variation within the breed considering that the detection of DNA polymorphisms whether in coding or non-coding regions of chromosomes is an important part when evaluating the genetic uniqueness of a breed [15].

Since all heifers were homozygous for the alanine variant at the *K232A* site, it can be inferred that the *A* allele is fixed within the Icelandic cattle breed. However, due the limited size of the sample this needs further investigation. The *A* allele has been observed to be nearly fixed in the French Montbéliarde dairy breed and present at very high frequency (>0.90) in Swedish Red Polled and Fleckvieh bulls [6, 16, 17]. The *K* residue positively affects content traits, mainly fat, at the cost of milk yield. Different breeding goals regarding milk composition for different breeds in different countries can therefore explain the variations in the frequency of the *K* allele [6]. The breeding goal for the Icelandic cattle breed was mainly to increase milk volume until in the 1990s when emphasis was set on protein as well. However, there was a gain in fat content over the years of 1950-1970 due to a current breeding emphasis. Since the *K* allele negatively affects milk yield it can be suggested that the *A* allele has been selected for since 1970 by using relatively more *AA* sires, thereby causing a constant decrease in the frequency of the *K* allele and fixation of the *A* allele. It can

also be suggested that the frequency of the *K* allele might have been very low already in the founder population and since then dwindled with time, perhaps following fluctuations in population size during harsh whether condition like the ones following the infamous volcano eruption in the 18th century.

In order to deduct on the exact frequency or fixation of alleles in the breed a further investigation needs to be carried out implementing a larger sample. The association of single or few polymorphisms, like the ones reported herein, and economically important traits is also of interest in the Icelandic cattle as such connection might be of direct use in breeding work.

ACKNOWLEDGEMENTS

Financial support for this study was provided by the Icelandic Board of Genetic Resources in Agriculture and the Icelandic Agricultural Productivity Fund.

REFERENCES

[1] Aðalsteinsson, S., *Origin and conservation of farm animal populations in Iceland.* J. Anim. Breed. Genet., 1981. **98**: p. 258-264.

[2] Badaoui, B., Serrdilla, J.M., Tomas, A., Urrutia, B., Ares, J.L., Carrizosa, J., Sanchez, A., Jordana, J. & Amills, M., *Short Communication: Identification of Two Polymorphisms in the Goat Lipoprotein Lipase Gene*

and Their Association with Milk Production Traits. J. Dairy Sci., 2007. **90**: p. 3012-3017.

[3] Buchanan, F.C., Van Kessel, A.G., Waldner, C., Christensen, D.A., *Hot topic: An Association Between a Leptin Single Nucleotide Polymorphism and Milk and Protein Yield.* J. Dairy Sci., 2003. **86**: p. 3164-3166.

[4] Calvo JH, M.-R.A., Beattie AE, Dodds KG, Marcos-Carcavilla A, Serrano M., *Fine mapping of genes on sheep chromosome 1 and their association with milk traits.* Anim. Genet., 2006. **37**(3): p. 205-10.

[5] Coppieters W., R.J., Arranz J.-J., Berzi P., Cambisano N., Grisart B., Karim L., Marcq F., Moreau L., Nezer C., Simon P., Vanmanshoven P., Wagenaar D., and Georges M., *A QTL with major effect on milk yield and composition maps to bovine Chromosome 14.* Mamm. Genome, 1998. **9**: p. 540-544.

[6] Gautier M., C.A., Fritz S., Eggen A., Boichard D. & Druett T., *Characterization of the DGAT1 K232A and Variable Number of Tandem Repeat Polymorphisms in French Dairy Cattle.* J. Dairy Sci., 2007. **90**(6): p. 2980-2988.

[7] Grisart B., C.W., Farnir F., Karim L., Ford C., Berzi P., Cambisano N., Mni M., Reid S., Simon P., Spelman R., Georges M. & Snell R., *Positional candidate cloning of a QTL in dairy cattle: identification of a*

missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res., 2002. **12**: p. 222-231.

[8] Konfortov B.A., L.V.E.M.J.R., *Re-sequencing of DNA from a diverse panel of cattle reveals a high level of polymorphism in both intron and exon.* Mamm. Genome, 1999. **10**: p. 1142-1145.

[9] Lagonigro R., W.P., Pilla F., Woolliams J.A. & Williams J.L., *A new mutation in the coding region of the bovine leptin gene associated with feed intake.* Anim. Genet., 2003. **34**: p. 371-374.

[10] Liu, K.M., S.V., *PowerMarker: an integrated analysis environment for genetic marker analysis.* Bioinformatics (Oxf.), 2005. **21**(9): p. 2128-2129.

[11] Pareek C.S., C.U., Zabolewicz T., Pareek R.S. & Walawski K., *DGAT1 K232A quantitative trait nucleotide polymorphism in Polish Black-and-White cattle.* J. Appl. Genet. 46, 2005. **1**: p. 85-87.

[12] Smith J.S., C.S., Jensen D.R., Chen H.C., Sande E., Tow B., Sanan D.A., Raber J., Eckel R.H. and Farese R.V. Jr., *Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat.* Nat. Genet., 2000. **25**: p. 87-90.

[13] Spelman R.J., C.A.F., McElhinney P., Gregory G.C., and Snell R.G., *Characterization of the DGAT1 Gene in the New Zealand Dairy Population.* J. Dairy Sci., 2002. **85**: p. 3514-3517.

[14] Szyda J., a.K., J., *Statistical Modeling of Candidate Gene Effects on Milk Production Traits in Dairy Cattle.* J. Dairy Sci., 2007. **90**: p. 2971-2979.

[15] Talle S.B., C.W.S., Fimland E., Syrstad O., Meuwissen T., and Klungland H., *Use of DNA technologies for the conservation of animal genetic resources: A review.* Acta Agric. Scand. Sect. A Anim. Sci., 2005. **55**: p. 1-8.

[16] Thaller G., K.W., Winter A., Kaupe B., Erhardt G. & Fries R., *Effects of DGAT1 variants on milk production traits in German cattle breeds.* J. Anim. Sci., 2003. **81**: p. 1911-1918.

[17] Umeland, P., *Genetisk variation av betydelse för mjölkkvalitet i Rödkullerasen [Swedish].* Examensarbete 283. Institutionen för husdjursgenetik. Sveriges Lantbruksuniversitet., 2006.

[18] Winter A., K.W., Werner F.A.O., Kollers S., Kata S., Durstewitz G., Buitkamp J., Womack J.E., Thaller G. & Fries R., *Association of a lysine-232/alanine polymorpghism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content.* PNAS, 2002. **99**(44): p. 9300-9305.

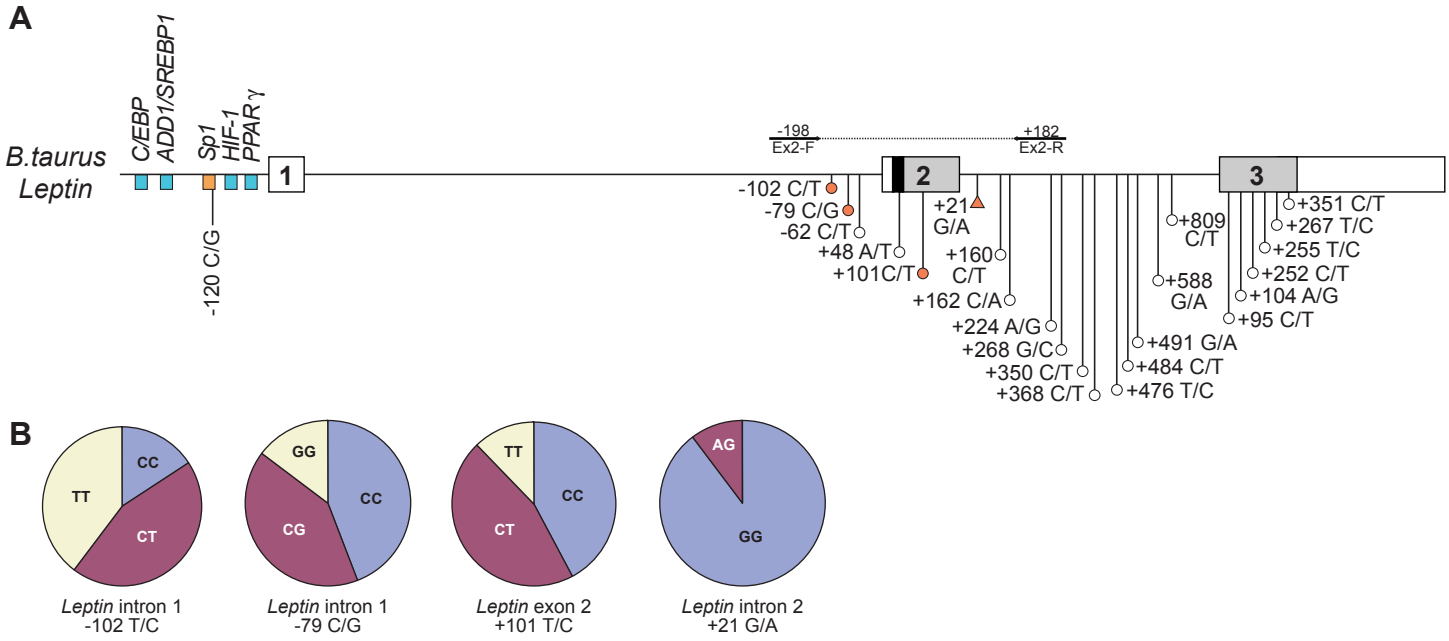
FIGURE CAPTIONS

Figure 1: (A) A schematic view of the bovine *Leptin* gene showing its exon-intron structure as well as location of primers and all known SNP polymorphisms. Location of SNP polymorphisms is given in relation to the nearest exon-intron boundary. Sites found to be polymorphic in the Icelandic cattle breed are marked with a filled circle. The novel SNP found in intron 2 is marked with a filled triangle.

(B) The frequency of homo- and heterozygotes for each polymorphic site.

Authors comment: Ideal position of figures and tables

Figure 1: At the end of section 3 (results).



Leptin intron 1
-102 T/C

Leptin intron 1
-79 C/G

Leptin exon 2
+101 T/C

Leptin intron 2
+21 G/A

Asbjarnardottir et al., Fig. 1