Small Ruminant Research 155 (2017) 72–80

Contents lists available at ScienceDirect

Small Ruminant Research

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

# Genetic diversity within economically important loci in European, Middle Eastern, and African sheep breeds: An insight into their development

Stefania Chessa[a], Isabella Jasmin Giambra[b], Horst Brandt[b], Anna Maria Caroli[c], Elisha Gootwine[d], Georg Erhardt[b,*]

[a] Institute of Agricultural Biology and Biotechnology, National Research Council, Lodi, Italy
[b] Department of Animal Breeding and Genetics, Justus-Liebig-University, Giessen, Germany
[c] Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy
[d] Department of Ruminant Science, Institute of Animal Science, Agricultural Research Organization, Bet Dagan, Israel

ARTICLE INFO

*Keywords:*
Ovine species
Genetic variability
Single nucleotide polymorphism
Casein
Haplotype
Economic trait loci

ABSTRACT

The aims of this study were to analyze sequence variability within genes related to important economic traits in dairy and non-dairy sheep breeds, and to evaluate the genetic variability as an insight into their development. In total, 31 single nucleotide polymorphisms (SNPs), one indel, and one microsatellite, all located in ovine milk protein genes (*CSN1S1*, *CSN1S2*, *CSN2*, *CSN3*, *LGB*), and 13 SNPs located in genes associated with production and reproduction traits of economic importance (*DGAT1*, *DGAT2*, *GHR*, *FASN*, *GHRHR*, *CTSB*, *MCR1*, *KRT1*, *IL2*, *IGF1*, *HR*, *GDF8*, *GDF9*) were genotyped. The analysis included 752 samples derived from 23 different European, Middle-Eastern (Turkey and Israel) and African (Nigeria and Cameroon) sheep breeds, and the European Mouflon.

We could not observe the often described decline of heterozygosity with increasing distance to the domestication centre maybe because of high gene flow and gene introgression between breeds following domestication. Tendencies of clustering according to the local origin of the animals were visible, whereas a clear breed grouping was not identified. No significant difference was found by principal component analysis of casein haplotypes between dairy and non-dairy breeds. However, using Canonical Discriminant Analysis most of the non-dairy individuals were correctly assigned to the non-dairy group, with an increasing precision when considering both casein haplotypes and the remaining SNP. The distribution of variation in the production-related SNP analyzed reflects both the long history of migrations, adaption, isolation, and the most recent effect of gene flow or isolation due to selection in the different breeds.

## 1. Introduction

Sheep were one of the first domesticated animals, about 9000 years before present in the Fertile Crescent in the Near East (Bruford et al., 2003; Zeder, 2008). Domestication and a long history of migrations, selection and adaption have created an enormous variety of breeds, whereas additionally mutations, selective breeding, adaption, isolation and genetic drift have created an enormous diversity of local populations (Groeneveld et al., 2010). As a consequence, sheep is the livestock species with the highest number of recorded breeds: 1294 worldwide distributed and 160 extinct breeds were described in 2010 (FAO, 2011). The loss of diversity of livestock breeds is affected by genetic drift, inbreeding, introgression and artificial and natural selection (Bruford, 2004). Selection has proceeded on traits such as coat colour, environmental tolerance, wool characteristics, and meat and milk production (Meadows et al., 2005), and was responsible for the development of highly productive and well-adapted sheep populations (Pereira et al., 2006). When less productive, locally adapted, native breeds are substituted with highly productive cosmopolitan breeds, undocumented diversity, which can contribute to future traits of interest for the reaction to future breeding objectives, is endangered (Bruford et al., 2003; Groeneveld et al., 2010; Pariset et al., 2012). An overview about the remaining genetic variability and evolutionary and diversity studies are necessary for an effective management of farm animal genetic resources (Groeneveld et al., 2010).

Markers used for the characterisation of livestock diversity are rarely causative mutations for phenotypic variation, as those markers are lying within genes under selection for economically important traits (Bruford et al., 2003) and may be non-optimal for calculating population parameters. However, those markers may be useful in assessing

* Corresponding author.
*E-mail address:* Georg.Erhardt@agrar.uni-giessen.de (G. Erhardt).

http://dx.doi.org/10.1016/j.smallrumres.2017.08.021
Received 1 February 2017; Received in revised form 3 August 2017; Accepted 31 August 2017
Available online 17 September 2017
0921-4488/ 

local speciation, adaption, and the effects of human (Pariset et al., 2012), as they allow tracing direct and indirect selection events leading to the spread of economically important alleles (Bruford et al., 2003).

Amongst genetic markers, SNPs occur frequently in the mammalian genome and are useful for rapid, large-scale, and cost-effective genotyping (Cappuccio et al., 2006; Pariset et al., 2006a). SNPs within genes involved in key metabolic pathways influencing production, disease and morphological traits (e.g. casein genes or haplotypes) were already used for evolutionary and diversity studies in cattle (Jann et al., 2004; Kamiński et al., 2005; Kaupe et al., 2004), and goat (Cappuccio et al., 2006; Finocchiaro et al., 2008; Pariset et al., 2006b). The distribution of the casein haplotypes in cattle, for example, shows a clear dependence on the geographic origin of the breeds and is due to selection of the breeds for different purposes (Jann et al., 2004).

To our knowledge, markers within genes coding for economic traits were not included in diversity studies in sheep in a broader way. The aim of this study was to analyse the genetic variability in such genes in local sheep breeds from different countries and in highly productive breeds from various parts of the world. We focused on autosomal genes with known influence on milk composition and performance, especially genes coding for milk proteins (*CSN1S1*, *CSN1S2*, *CSN2*, *CSN3*, *LGB*), and genes influencing synthesis of milk fat (*DGAT1*, *DGAT2*, *IGF1*, *GHR*, *FASN*). Furthermore, we studied genes with influence on other economic important traits like growth traits (*GDF8*, *GHRHR*), fertility (*CTSB*, *GDF9*), wool characteristics (*MC1R*, *HR*, *KRT1*), and immune response (*IL2*).

## 2. Material and methods

### 2.1. Ethics statements

No animal experiments were performed specifically for this manuscript. According to the German Animal Welfare Law (released on 05/18/2006) no notification or approval by the Animal Protection Unit of the Regional Council of Gießen, Germany, was necessary for this study (article 8 (7) 2a). No endangered or protected species were involved in this study.

### 2.2. Samples

The study includes 752 DNA-samples of 23 European, Middle-Eastern, and African sheep breeds, as well as 32 samples of *O. musimon,* the European Mouflon (Table 1), considered the descendant of feral sheep (FAO, 2008) and probable representative of a remnant of the first domestic sheep that entered Europe (Rezaei et al., 2010).

### 2.3. Markers and typing methods

In total, 46 markers were selected as follows: 31 SNPs, one indel, and one microsatellite were located in milk protein genes (*CSN1S1*, *CSN1S2*, *CSN2*, *CSN3*, *LGB*), and 13 SNPs in genes associated with production and reproduction traits of economic importance (*DGAT1*, *DGAT2*, *GHR*, *FASN*, *GHRHR*, *CTSB*, *MCR1*, *KRT1*, *IL2*, *IGF1, HR*, *GDF8*, *GDF9*). A list of all analysed markers and the relevant information about them are presented in Table S1.

Most of the SNPs were typed by matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) by Eurofins Medigenomixs GmbH (Ebersberg, Germany) using the Sequenom Massarray iPLEX Gold System (Sequenom, San Diego, USA). Genotyping of SNPs 6, 15, 16 and 17 was done by polymerase-chain-reaction followed by restriction-fragment-length-polymorphism (PCR-RFLP) analysis. In particular, SNP 6 was typed according to Giambra et al. (2010) and SNPs 15 and 16 according to Giambra and Erhardt (2012). PCR-RFLP-test for SNP 17 was based on an amplification created restriction site (ACRS; Haliassos et al., 1989).

The ABI Genetic Analyzer 3130 was used for typing the microsatellite within intron 3 of *CSN3*, following the protocol described by Brownstein et al. (1996). Comparisons of the *CSN3* microsatellite sequences were made using ChromasPro 1.32 (Technelysium Pty Ltd. Queensland, Australia).

### 2.4. Statistical analyses

#### 2.4.1. Markers anlysis

Allele frequencies of all loci and possible deviations from Hardy-Weinberg equilibrium (HWE), and heterozygosity were calculated using PopGene V 1.31 (Yeh et al., 1997).

#### 2.4.2. Genetic distances

All parameters of the F-statistics according to Weir and Cockerham (1984) and Nei (1973) were computed using FSTAT v2.9.3.2 (Goudet, 1995).

Pairwise Reynolds' genetic distances $D_R$ between the 24 sheep breeds were calculated with PHYLIP package 3.69 (Felsenstein, 2005). These distance data were used to construct a phylogenetic unweighted pair group method with arithmetic mean (UPGMA) tree with SplitsTree4 V4.12.6 (Huson and Bryant, 2006).

#### 2.4.3. Principal component analysis

A principal component analysis (PCA) was made to show pairwise differentiation between breeds absorbing the breed specific allele frequencies with the software package SAS® 9.3 (SAS Institute Inc., Cary, NC, USA).

#### 2.4.4. Bayesian clustering methods

STRUCTURE 2.3.4. program (Pritchard et al., 2000) was used to estimate the most likely number of subpopulations K independently of breed affiliation, performing the analyses for K from 2 to 6. All runs used a burn-in period of 20,000 iterations and a period of data collection of 10,000 iterations. The posterior probability for K was then calculated for each value of K using the mean estimated log-likelihood of K, L (K), to choose the optimal K. Following Evanno et al. (2005), we also calculated Delta K (ΔK), an *ad hoc* statistic based on the second order rate of change of the likelihood function, L"(K), with respect to K. Graphic representation of these statistics were obtained with the web-based STRUCTURE HARVESTER software v0.6.93 (Earl and VonHoldt, 2012). Populations were then assigned to a subpopulation based on a membership percentage (q) ≥50%. DISTRUCT (Rosenberg, 2004) was used for graphical illustration of the results of STRUCTURE.

#### 2.4.5. Casein-haplotype analysis

Intragenic haplotypes were calculated for each casein gene using the expectation and maximization (EM) algorithm (Excoffier and Slatkin, 1995; Hawley and Kidd, 1995; Long et al., 1995) of the HAPLOTYPE procedures of SAS® 9.3 (SAS Institute Inc., Cary, NC, USA) to obtain iteratively the maximum likelihood estimates of haplotype frequencies. The HTSNP procedure of SAS® 9.3 (SAS Institute Inc., Cary, NC, USA) was then adopted to select a reduced number of SNPs among the 31 available casein SNPs and reconstruct the entire casein cluster haplotypes (including in the order *CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*). The HTSNP procedure captures the most relevant SNP, termed haplotype tagging SNP (htSNP), characterizing the haplotype structure in a block as described by Johnson et al. (2001).

Frequencies of casein haplotypes between dairy and non-dairy breeds, classified as reported in Table 1, were compared using a principal component analysis (PCA) with the PRINCOMP procedure of SAS® 9.3 (SAS Institute Inc., Cary, NC, USA). A canonical discriminant analysis (CDA) was also applied using SNPs and haplotype data to assign individual sheep to the dairy and non-dairy groups. MASS 7.3–35 (Venables and Ripley, 2002) and Candisc 0.6–5 (Friendly and Fox, 2013) R packages were used to perform the analysis. Individuals were included in the analysis only if their haplotype assignment probability