Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Chromatography B, 833 (2006) 117–120
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Short communication

Assessment of population structure by single nucleotide polymorphisms (SNPs) in goat breeds☆

L. Pariset[a,*], I. Cappuccio[a], P. Ajmone Marsan[b], S. Dunner[c], G. Luikart[d], P.R. England[d], G. Obexer-Ruff[e], C. Peter[f], D. Marletta[g], F. Pilla[h], A. Valentini[a]

ECONOGENE Consortium[1]

[a] *Dipartimento di Produzioni Animali, Università della Tuscia, Viterbo, Italy*
[b] *Istituto di Zootecnica, Università Cattolica del Sacro Cuore, Piacenza, Italy*
[c] *Departimento de Producción Animal, Universidad Complutense, Madrid, Spain*
[d] *Laboratoire de Biologie des Populations d'Altitude, Université Joseph Fourier, Grenoble, France*
[e] *Institute of Animal Genetics, Nutrition and Housing, University of Berne, Berne, Switzerland*
[f] *Department of Animal Breeding and Genetics, University of Giessen, Giessen, Germany*
[g] *Dipartimento di Scienze Agronomiche, Agrochimiche e delle Produzioni Animali, Università di Catania, Catania, Italy*
[h] *Dipartimento di Scienze Animali Vegetali e dell'Ambiente, Università degli Studi del Molise, Campobasso, Italy*

Received 29 July 2005; accepted 15 January 2006
Available online 10 February 2006

Abstract

Single nucleotide polymorphisms (SNPs) may be used in biodiversity studies and commercial tasks like traceability, paternity testing and selection for suitable genotypes. Twenty-seven SNPs were characterized and genotyped on 250 individuals belonging to eight Italian goat breeds. Multilocus genotype data were used to infer population structure and assign individuals to populations. To estimate the number of groups (K) to test in population structure analysis we used likelihood values and variance of the bootstrap samples, deriving optimal K from a drop in the likelihood and a rise in the variance plots against K.

Keywords: SNPs; Goat; Population structure

1. Introduction

An ideal genetic marker for population and evolutionary studies should be abundant and distributed widely across the genome and the data must be comparable across laboratories using different genotype scoring methods or technologies [1]. Markers routinely used for molecular ecology and conservation biology include DNA fingerprinting, mitochondrial DNA sequencing, restriction fragment length polymorphism (RFLP) analysis, amplified fragment length polymorphisms (AFLPs) and microsatellites, approaches undergoing some technical and analytical matters [2–4]. Single nucleotide polymorphisms (SNPs) appear to have some of the characteristics of an ideal genetic marker, being abundant in the genome, easy to reproduce in different laboratories, simple to score. SNPs can be interesting for population genetic studies and to discover signature of selection [5–7].

A list of genes involved in key metabolic pathways influencing production, disease resistance and morphological traits have been selected for SNPs discovery in goat under the EU Econogene project. All the selected genes are candidate to have high adaptive value and to be under strong natural or artificial selection.

So far, there are not goat SNPs available in NCBI dbSNP database and only 26 are reported in the literature so far. Regardless of the difficulties in SNPs identification, the use of those markers could lead to a rapid, large scale and cost effective genotyping [8,9] for several purposes.

One of the uses of SNPs in population analysis is the possibility to assess the probable number of populations in a

☆ This paper was presented at the 2nd IPSo Congress on Proteomics and Genomics, Viterbo, Italy, 29 May to 1 June 2005.

* Corresponding author. Tel.: +39 076 135 7447; fax: +39 076 135 7434.
E-mail address: pariset@unitus.it (L. Pariset).

[1] http://lasig.epfl.ch/projets/econogene.

1570-0232/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2006.01.011

genotyped sample. This issue has been poorly investigated: Pritchard et al. [10] presented a method to compute the assignment of individuals to a predetermined number of groups using a Bayesian inference. However, the most probable number of groups in a population (K) cannot be directly assessed by their software. Here, we present a method that estimates K by plotting the likelihood and its variance in bootstrap replicates versus K.

2. Experimental

2.1. Material

A total of 264 animals, about one third males, belonging to eight Italian breeds were sampled. Analysed breeds are Argentata dell'Etna (Sicily), Sarda (Sardinia), Grigia molisana (Molise), Girgentana (Sicily), Bionda dell'Adamello (Lombardia e Trentino Alto Adige), Camosciata (Piemonte e Trentino Alto Adige), Valdostana (Valle d'Aosta e Piemonte), Orobica (Lombardia).

For each breed, blood samples of 33 individuals per farm were collected over the traditional rearing area of the breed, sampling no more than three animals per farm. DNA was extracted with standard techniques. The strategy followed for SNP discovery was using two SNP discovery panels, each including eight individuals belonging to different European breeds and representing the most of the variation across a wide geographic area. Some already identified polymorphisms were utilised [11–13]. From a total of 35 genes screened in goats, 27 SNPs have been discovered in exon and intron sequences of the following loci: activin receptor IIB, calpastatin, callipyge, αS1-casein, κ-casein, cathepsinK, desmin, MHC class II DQA gene and DRB gene, fatty acid binding protein 4, fibronectin 1, GDF9 fecundity gene, growth hormone receptor, Interleukin 2, Interleukin 4, integrin B1, b-lactoglobulin, lipase, melatonin, myostatin, prion-protein, toll-like receptor 4. Those SNPs were applied to the structure analysis of the selected Italian breeds.

2.2. Polymerase chain reaction (PCR) conditions

Pairs of PCR primers were designed on genomic sequences published in GeneBank belonging either to *Capra hircus* or to the most genetically related species available, more often human.

Amplification conditions were optimised to produce single amplicons of the predicted length. Amplicons were then used for sequence and DHPLC analyses. Each polymerase chain reaction was performed in volume of 30 μl containing 30 ng of genomic DNA, 1.6 pMol of each primer, 200 mM dNTPs, 1× PCR buffer and 0.2 unit Taq DNA polymerase (Amersham Pharmacia Biotech, Piscataway, NJ, USA). After initial denaturation at 94 °C for 5 min, 35 cycles of amplification with denaturation at 94 °C for 30 s, annealing for 30 s at the temperature optimised for each primer pair and extension at 72 °C for 1 min, followed by a final extension step at 72 °C for 10 min, were performed. Primer sequences and amplification conditions are in press [14].

Table 1

Locus	Analysis temperature	Slope (% of TEAA with 25% of acetonitrile per minute)
GHR	59.0	57.0
GHRHR	58.5	62.0
IGF1	58.5	54.0
MSTN	55.9	57.0
RYR1	61.0	62.0
SCD	60.3	57.0
TYRP1	63.5	57.0

2.3. Sequence analysis

Polymerase chain reaction products were purified through Sephadex G-50 (Amersham Pharmacia Biotech, Piscataway, NJ, USA) to remove residual primers and dNTPs and used as templates for two sequencing reactions. Sequencing was performed using a CEQ8800 sequencer using DTCS QuickStart Kit (both from Beckman Coulter, Fullereton, CA, USA) according to manufacturers instructions. The sequence data were analysed and aligned with Bioedit software [15] to ascertain polymorphisms and characterize SNPs.

2.4. Denaturing high pressure liquid chromatography (DHPLC) analysis

Some genes fragments have been subjected to DHPLC scanning on an automated HPLC instrument (Transgenomic WAVE^{TM} DNA Fragment Analysis System) for SNPs discovery. PCR products, loaded on a DNASep analytical column (Transgenomic, Omaha, NE, USA), were eluted from the column with a binary gradient of 0.1 M triethylammonium acetate (TEAA, pH 7.0) and 0.1 M TEAA with 25% of acetonitrile (pH 7.0) at a flow rate of 1.5 ml/min. Elution gradients and analysis temperatures are automatically predicted from the target sequences by WAVEMaker 4.1 software (Transgenomic, Omaha, NE, USA) (Table 1), and the eluted products detected by UV analysis at 260 nm. Polymorphic individuals have been sequenced to confirm and characterize each SNP.

2.5. SNPs analysis

Genotyping was performed in different laboratories participating to the Econogene project, choosing the appropriate technology on the basis of the specific competence and facilities of each lab, e.g. by direct sequencing, PCR-RFLP (polymorphism in the length of restriction fragments obtained by restriction digestion of a PCR fragment, [12]), single stranded conformational polymorphism (SSCP [11,13]), single nucleotide primer extension performed by incorporation of a dideoxy nucleotide using the SNaPshotTM Multiplex Kit (Applied Biosystems) according to the manufacturer's instructions or outsourced to *K* Biosciences, which uses both competitive allele specific PCR system (KASPar) and Taqman^{TM} chemistries (detection is achieved with proven 5′ nuclease chemistry by means of exonuclease cleavage of a 5′ allele-specific dye label, which

generates the assay signal; Applied Biosystems™, Foster City, CA, USA).

2.6. Data analysis

Individuals were clustered applying a parametric genetic mixture analysis implemented in the Structure 2.0 software [10,16], and a number of genetic clusters (K) ranging from 2 to 10 was tested using the no admixture model. Consistent results across runs were obtained using a burning period of 100,000 followed by 200,000 Markov chain Monte Carlo (MCMC) repeats and considering SNPs frequencies correlated among populations.

3. Results and discussion

A typical Structure analysis assumes a model in which there are K populations, each of which is characterized by a set of allele frequencies at each locus. Individuals are assigned (probabilistically) to populations, or jointly to two or more populations, if their genotypes indicate that they are admixed. It is assumed that within populations, the loci are at Hardy–Weinberg equilibrium, and in linkage equilibrium.

Individuals were clustered applying a parametric genetic mixture analysis implemented in the Structure 2.0 software [10], and a number of genetic clusters (K) ranging from 2 to 10 was tested using the no admixture model, so assuming that each individual comes purely from one of the K populations, and carrying out five runs for each K. Consistent results across runs were obtained using a burning period of 100,000 followed by 200,000 MCMC repeats and considering SNPs frequencies correlated among populations.

The likelihood values and the variance of the bootstrap samples so obtained were plotted against K (Fig. 1) for choosing the optimal K value leading to the most reliable results. The likelihood reaches a maximum around $K = 6$–7, and variance across runs shows a sharp rise for $K > 7$. We hypothesize that this kind of plot can be used to establish the most probable number of groups in which a population is subdivided. In fact, if the number of groups is higher than the true one, individuals are subdivided into groups they do not belong to. In each run individuals can be allocated to different groups and this determines a rise in the variance across bootstraps. In our case, we infer $K = 6$ by a parsimony principle since $K = 7$ has almost the same values of likelihood and variance.

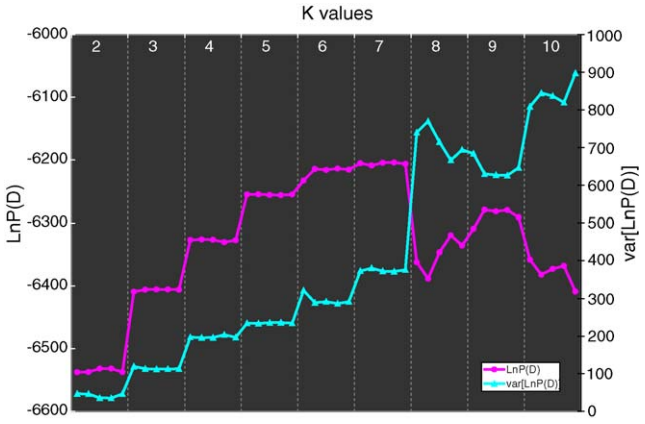

Fig. 1. ln of the probability of the data (ln $P(D)$ and its variance for K from 2 to 10.

Table 2
Proportion of membership of analysed goat breeds in each of the six clusters ($K = 6$)

Goat breed	Inferred clusters					
	1	2	3	4	5	6
Argentata dell'Etna	0.20	0.04	0.33	0.04	0.35	0.04
Sarda	0.25	0.01	0.35	0.05	0.23	0.12
Grigia molisana	0.05	0.08	0.18	0.09	0.41	0.19
Girgentana	0.48	0.23	0.11	0.03	0.10	0.05
Bionda dell'Adamello	0.07	0.48	0.16	0.12	0.04	0.14
Camosciata	0.16	0.50	0.22	0.03	0.03	0.06
Valdostana	0.06	0.39	0.05	0.09	0.02	0.38
Orobica	0.05	0.07	0.04	0.78	0.04	0.03

A graphic representation of the estimated membership coefficients to the six clusters for each individual, obtained running structure setting $K = 6$, is shown in Fig. 2. Each individual is represented by a single vertical line broken into K colored segments, whose lengths are proportional to each of the K inferred clusters. Table 2 demonstrates the proportion of membership of

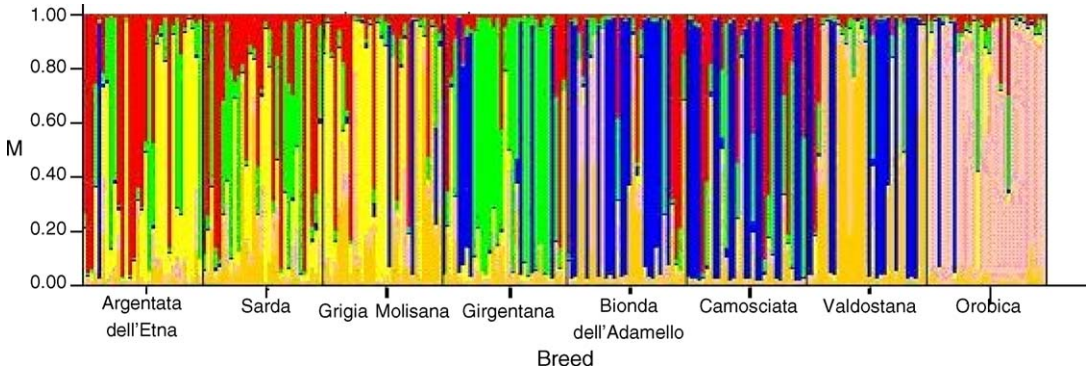

Fig. 2. Summary plot of estimates of Q (estimated membership coefficients for each individual, in each cluster) for $K = 6$ in the eight goat breeds. Each individual is represented by a single vertical line broken into K colored segments, with lengths proportional to each of the K inferred clusters. Each colour represents the proportion of membership (M) of each individual, represented by a vertical line, to the six clusters.

the eight goat breeds in each of the six clusters. Orobica breed is the most differentiated, with 0.78% of the individuals assigned to a single cluster, and most Girgentana individuals (0.48%) belong to a specific cluster. Of the other cluster inferred, one is predominant in Bionda dell'Adamello, Camosciata and Valdostana and another in Argentata dell'Etna and Grigia Molisana. Valdostana shows a high proportion of individuals assigned to a peculiar group, and Sarda breed shows the highest level of genetic admixture.

The results obtained show a genetic structure that seems related to geographic distribution, since northern breeds (Sarda, Grigia Molisana and Orobica) and southern ones (Argentata dell'Etna and Bionda dell'Adamello) belongs to distinct genetic cluster (Fig. 2).

4. Conclusions

This study produced consistent data relative to SNP discovery in *Capra hircus,* derived from analysis of individuals belonging to different breeds and representing a wide geographic area. SNPs have been successfully used for the structure analysis of eight Italian breeds, showing that the number of polymorphisms used for genotyping is enough to produce consistent data for individual assignment to populations and that there is a geographic variation between northern and southern breeds. Moreover, we produced a method for the identification of the right number of groups (K) to test in population analysis. In fact, assessing the probable number of populations in a genotyped sample is not always an obvious matter. Structure software elaborates population data giving the probability of assignment of individuals to groups, but the most probable number of groups in a population (K) cannot be directly assessed by the software. Here, we show how to estimate K by plotting the likelihood and its variance in bootstrap replicates versus K.

Acknowledgements

This work has been partially supported by the EU Econogene contract QLK5-CT-2001-02461. The content of the publication does not represent necessarily the views of the Commission or its services.

References

[1] P. Sunnucks, Trends Ecol. Evol. 15 (2000) 199.
[2] N. Aitken, S. Smith, C. Schwarz, P.A. Morin, Mol. Ecol. 13 (2004) 1423.
[3] C. Schlötterer, J. Pemberton, in: R. DeSalle, B. Schierwater (Eds.), Molecular Approaches to Ecology and Evolution, Birkhäuser–Verlag, Berlin, 1998, p. 71.
[4] G. Luikart, P.R. England, Trends Ecol. Evol. 14 (1999) 253.
[5] G. Luikart, P.R. England, D. Tallmon, S. Jordan, P. Taberlet, Nat. Rev. Genet. 4 (2003) 981.
[6] J.M. Akey, G. Zhang, K. Zhang, L. Jin, M.D. Shriver, Genome Res. 12 (2002) 1805.
[7] L. Pariset, I. Cappuccio, F. Pilla, D. Marletta, P. Ajmone Marsan, A. Valentini, ECONOGENE Consortium. Ital. J. Anim. Sci. 4 (2005) 131.
[8] A. Vignal, D. Milan, M. San Cristobal, A. Eggen, Genet. Sel. Evol. 34 (2002) 275.
[9] C. Schlötterer, Nat. Rev. Genet. 5 (2004) 63.
[10] J.K. Pritchard, M. Stephens, P. Donnelly, Genetics 155 (2000) 945.
[11] G. Lühken, A. Buschmann, M.H. Groschup, G. Erhardt, Arch. Virol. 149 (2004) 1571.
[12] L. Ramunno, G. Cosenza, M. Pappalardo, N. Pastore, D. Gallo, P. Di Gregorio, P. Masina, Anim. Genet. 31 (2000) 342.
[13] S. Chessa, E. Budelli, K. Gutscher, A. Caroli, G. Erhardt, J. Dairy Sci. 86 (2003) 3726.
[14] I. Cappuccio, L. Pariset, P. Ajmone-Marsan, S. Dunner, O. Cortes, G. Erhardt, G. Lühken, C. Peter, S. Joost, I.J. Nijman, J.A. Lenstra, G. Luikart, G. Obexer-Ruff, A. Valentini and the Econogene Consortium, Anim. Genet., 2006, submitted for publication.
[15] T.A. Hall, Nucleic Acids Symp. Ser. 41 (1999) 95.
[16] D. Falush, M. Stephens, J.K. Pritchard, Genetics 164 (2003) 1567.