THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Exploiting most informative markers to predict group membership of North Aegean sheep

Citation for published version:
Kominakis, A, Tarsani, E, Hager, A & Hadjigeorgiou, I 2020, 'Exploiting most informative markers to predict group membership of North Aegean sheep', EAAP 2020: 71st Annual Meeting of the European Federation of Animal Science, 1/12/20 - 4/12/20. https://doi.org/10.13140/RG.2.2.22685.13289

Digital Object Identifier (DOI):
10.13140/RG.2.2.22685.13289

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 27. Oct. 2022

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/346730480

Exploiting most informative markers to predict group membership of North Aegean sheep

Poster · December 2020

DOI: 10.13140/RG.2.2.22685.13289

CITATIONS
0

READS
51

4 authors, including:

Eirini Tarsani
The University of Edinburgh
9 PUBLICATIONS **27** CITATIONS
SEE PROFILE

Ariadne Loukia Hager-Theodorides
Agricultural University of Athens
46 PUBLICATIONS **1,366** CITATIONS
SEE PROFILE

Ioannis Hadjigeorgiou
Agricultural University of Athens
78 PUBLICATIONS **914** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Sheep and goats organic farming systems View project

Project Technical innovation in Cretan dairy sheep farming View project

All content following this page was uploaded by Ioannis Hadjigeorgiou on 08 December 2020.

The user has requested enhancement of the downloaded file.

Exploiting most informative markers to predict group membership of North Aegean sheep

A. Kominakis, E. Tarsani, A. Hager, I. Hadjigeorgiou

Department of Animal Science, Agricultural University of Athens, Iera Odos, 11855, Athens, Greece, email: acom@aua.gr

Abstract

Aim of the present study was to discover a small panel of SNPs with the highest discriminatory power to classify sheep to: (i) three neighboring North Aegean islands (Lemnos, Lesvos and Agios Efstratios) and (ii) two agro-ecological zones (rough and flat landscape) within Lemnos. A total number of 256 ewes belonging to n=15 herds dispersed on the three islands and the two agro-ecological zones were genotyped with the 50K SNP array. Application of quality criteria at the marker level resulted in about 37K SNPs retained. Of these markers, we searched for a limited number (n=50 or 100) of strongly differentiated SNPs between groups (islands/zones) defined as those with highest values for the F_{ST} fixation index and combined results of principal components (PCA) and Nonparametric Discriminant Analysis (NPDA) to derive the classification criterion (DC) across cases. Application of the DC derived on the first two principal components (PCs) constructed on n=100 strongly differentiated SNPs resulted in a lowest average misclassification error rates (0.019) across the islands, whereas the respective DC derived on the first PC of n=50 strongly differentiated SNPs could accurately assign individuals to the two agro-ecological zones (average misclassification error rate: 0.012). Current results suggest that the proposed approach could be implemented to unequivocally determine the origin/habitat of local sheep and possibly their associated specific products.

Introduction

Application of Discriminant Analysis (DA) on genomic data can summarize genetic differentiation between groups and allow for discrimination of individuals into pre-defined groups (supervised classification). Instead of using raw genomic data, discrimination between groups could alternatively be achieved by using limited number of discriminatory variables such as Principal Components (PCs) of the original genetic data that are uncorrelated and comprise most of the variation of the original data. Following this rationale, in the present study, first, we detected a limited number of highly differentiated SNPs, then we constructed PCs on the highly differentiated markers and finally we conducted nonparametric DA on the constructed PCs in attempts to classify animals into three neighboring islands (Lemnos, Lesvos and Agios Efstratios) and two agro-ecological zones (rough and flat landscape) within Lemnos island. Present results are expected to allow for the development of a method that could explicitly determine the origin of sheep within and across islands.

Material and methods

Data and quality control

A total number of 268 ewes were genotyped with the medium density SNP array (Illumina Ovine 50K SNP array). Quality control (QC) was performed at a sample and at a marker level.

After QC criteria:

- **256 samples** (n=197 in Lemnos, n=38 in Lesvos and n=21 in Agios Efstratios) were remained due to: call rate>0.90.
- **37,201 autosomal SNPs** were retained due to: call rate>0.95, minor allele frequency>0.05, Hardy–Weinberg equilibrium (HWE) using a Fisher exact test p-value>10^{-6} and SNP pruning due to linkage disequilibrium (LD) r^2 levels (r^2<0.50, window size: 50 SNPs, increment: 5 SNPs)

QC was performed in SNP & Variation Suite software.

Discriminant analyses

We performed DA between: i) islands (n=256 samples) and ii) agro-ecological zones (n=60 samples in rough and n=137 samples in flat landscape) of Lemnos, respectively. The same analyses were conducted as described below:

i. We searched for a limited number (n=50 or 100) of strongly differentiated SNPs between groups (islands/agro-ecological zones) defined as those with highest fixation index (F_{ST}) values.
ii. We then performed Principal Component Analysis (PCA) on the various sets of differentiated markers to obtain smaller sets i.e. PCs that contained most of the variance in the original genetic data.
iii. Finally, we employed Nonparametric Discriminant Analysis (NPDA) using the *k*-th nearest-neighbor method on PCs constructed at step (ii) to derive the discriminant (classification) criterion (DC). In the *k*-nearest-neighbor method, the DC is based on non-parametric density estimates per each group and proximity between groups is based on the estimated Mahalanobis distances estimated on the pooled covariance matrix.

Analyses of F_{ST} and PCs were performed in SNP & Variation Suite software while NPDA was carried out in SAS (2013). Results of NPDA are presented as crossvalidation misclassification error rates per group (island and/or landscape) and on average.

Acknowledgements

This work was supported by the 'Terra Lemnia' project funded by the 'MAVA Fondation pour la Nature'

Results

Discriminatory power of 50 or 100 markers to assign sheep to islands

Table 1. Misclassification error rates of the discriminant criterion (DC) into islands based on *3* nearest-neighbors NPDA. DC was derived on first two PCs constructed on 50 or 100 strongly differentiated markers (SD: standard deviation)

Number of markers	Average F_{ST} (SD)	Misclassification error rate			
		Lemnos	Lesvos	Agios Efstratios	Average
50	0.211 (0.030)	0.167	0.237	0.048	0.150
100	0.187 (0.032)	0.010	0.001	0.048	0.019

Figure 1. *Geographical map of the studied sheep populations*

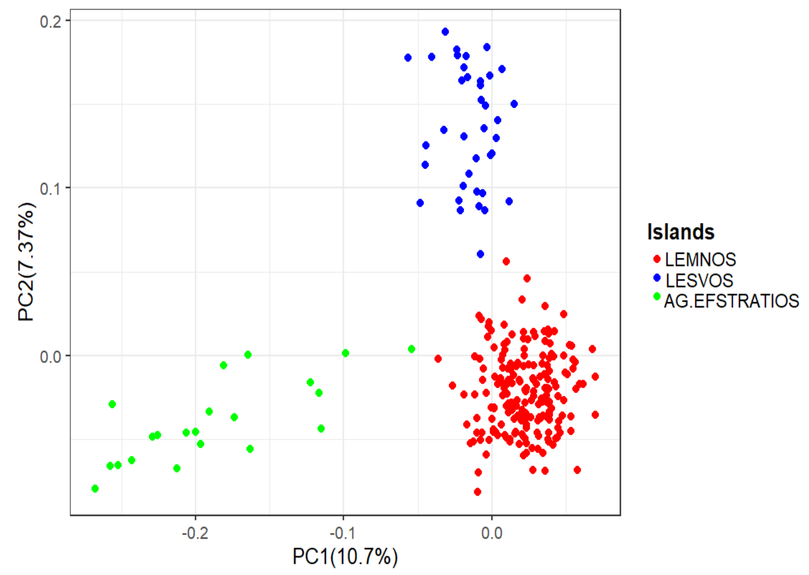

Figure 2. *Two-dimensional plot of the first two PCs using 100 most informative SNPs of n=256 ewes' genotypes. Note the high overall success rate of correct discrimination of animals to the three islands (see Table 1)*

Discriminatory power of 50 markers to assign sheep to agro-ecological zones

Table 2. Misclassification error rates of the discriminant criterion (DC) into two agro-ecological zones based on *2* nearest-neighbors NPDA. DC was derived on first PC constructed on 50 strongly differentiated markers (SD: standard deviation)

Number of markers	Average F_{ST} (SD)	Misclassification error rate		
		rough landscape	flat landscape	Average
50	0.249 (0.031)	0.007	0.017	0.012

Figure 3. *Two-dimensional plot of the first PC using 50 highly differentiated SNPs of n=197 Lemnos ewes' genotypes. Note the high overall success rate of correct discrimination of animals to different landscapes (see Table 2)*

Conclusions

- A number of 100 highly differentiated SNPs was sufficient to determine the origin (island) of the studied sheep while 50 discriminatory SNPs could effectively classify Lemnos sheep to rough/flat landscape habitats.
- The current approach could be exploited for the development of a specialized panel of markers that could be used for traceability purposes of animals across and within islands and possibly the associated specific products of local sheep.
- Current findings warrant a larger sample size to be verified.