


Article

Identification of Ancestry Informative Marker (AIM) Panels to Assess Hybridisation between Feral and Domestic Sheep

Elisa Somenzi, Paolo Ajmone-Marsan and Mario Barbato * 

Department of Animal Science, Food and Nutrition – DIANA, and Nutrigenomics and Proteomics Research Center – PRONUTRIGEN, Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy; elisa.somenzi@unicatt.it (E.S.); paolo.ajmone@unicatt.it (P.A.-M.)

* Correspondence: mario.barbato@unicatt.it

Received: 1 March 2020; Accepted: 25 March 2020; Published: 30 March 2020



Simple Summary: Once present in the entirety of Europe, mouflon (wild sheep) became extinct due to intense hunting, but remnant populations survived and became feral on the Mediterranean islands of Corsica and Sardinia. Although now protected by regional laws, Sardinian mouflon is threatened by crossbreeding with domestic sheep causing genetic hybridisation. The spread of domestic genes can be detrimental for wild populations as it dilutes the genetic features that characterise them. This work aimed to identify diagnostic tools that could be applied to monitor the level of hybridisation between mouflon and domestic sheep. Tens of thousands of genetic markers known as single nucleotide polymorphisms (SNPs) were screened and we identified the smallest number of SNPs necessary to discriminate between pure mouflon and sheep. We produced four SNP panels of different sizes which were able to assess the hybridisation level of a mouflon and we verified that the SNP panels efficacy is independent of the domestic sheep breed involved in the hybrid. The implementation of these results into actual diagnostic tools will help the conservation of this unique and irreplaceable mouflon population, and the methodology applied can easily be transferred to other case studies of interest.

Abstract: Hybridisation of wild populations with their domestic counterparts can lead to the loss of wildtype genetic integrity, outbreeding depression, and loss of adaptive features. The Mediterranean island of Sardinia hosts one of the last extant autochthonous European mouflon (*Ovis aries musimon*) populations. Although conservation policies, including reintroduction plans, have been enforced to preserve Sardinian mouflon, crossbreeding with domestic sheep has been documented. We identified panels of single nucleotide polymorphisms (SNPs) that could act as ancestry informative markers able to assess admixture in feral \times domestic sheep hybrids. The medium-density SNP array genotyping data of Sardinian mouflon and domestic sheep (*O. aries aries*) showing pure ancestry were used as references. We applied a two-step selection algorithm to this data consisting of preselection via Principal Component Analysis followed by a supervised machine learning classification method based on random forest to develop SNP panels of various sizes. We generated ancestry informative marker (AIM) panels and tested their ability to assess admixture in mouflon \times domestic sheep hybrids both in simulated and real populations of known ancestry proportions. All the AIM panels recorded high correlations with the ancestry proportion computed using the full medium-density SNP array. The AIM panels proposed here may be used by conservation practitioners as diagnostic tools to exclude hybrids from reintroduction plans and improve conservation strategies for mouflon populations.

Keywords: hybridisation; ancestry informative marker; mouflon; sheep; random forest

1. Introduction

Genetic hybridisation among related species is increasingly studied due to its role in the adaptation, evolution, and diversification of species [1–5]. Hybridisation due to crossbreeding between interfertile species can occur due to range overlap of the populations [6], or can be promoted by anthropogenic activities [7–9]. Habitat degradation and livestock translocation have recently increased the rate of hybridisation events worldwide, contributing to the erosion of genetic diversity, and, in some cases, to the extinction of locally adapted wild animals [10,11]. In Europe, the genetic diversity represented by many locally adapted wild species is threatened due to hybridisation with their domestic counterpart [6]. Among these species is the European mouflon (*Ovis aries musimon*) present on the Mediterranean islands of Corsica and Sardinia [12,13]. The European mouflon is considered a remnant of the first wave of sheep domestication that occurred in the Fertile Crescent ~11,000 years ago (YA), whereas the current domestic sheep (*Ovis aries*) have been associated to a second wave of domestication which occurred ~5,000 years later [14]. The European mouflon was introduced in Corsica and Sardinia ~6–7,000 YA following human migrations and established feral populations (i.e., domesticates that have returned to the wild state) which survived until today in the harshest mountainous area of the islands [14]. Sardinia hosts the largest extant autochthonous European mouflon population [12,13,15]. Throughout the last century the Sardinian mouflon was reduced to the brink of extinction due to intense hunting and habitat erosion, until it was declared endangered and protected under local Governmental laws (Legge Regionale n23 del 1998). Since then, conservation efforts have been pursued to increase the mouflon presence in the island through translocation and genetic rescue. Since the arrival of the second wave of domestication in Europe, the sheep population brought to Sardinia has lived in sympatry with the already established mouflon population residing in the islands. Crosses between mouflon and domesticated sheep are documented since Roman times [12] and evidence of introgression from domestic sheep to mouflon has been occasionally reported by microsatellites and single nucleotide polymorphism (SNP) genotyping [12,16]. Importantly, strong signals of recent introgression were detected in an enclosed mouflon population used for restocking across the island [12]. Such an occurrence is not surprising, as the identification of hybrids based on morphological features is unreliable, especially when backcrosses occur and individuals no longer show a distinguishable intermediate phenotype between parental taxa [10]. Hence, establishing new populations from crossbred founders represents a threat for the Sardinian autochthonous genetic diversity. The use of genome-wide analysis using DNA arrays with tens or hundreds of thousands of SNPs allows the accurate detection of hybrid individuals despite confounding morphological evidence [17]. However, the large-scale use of DNA arrays can be challenging for the average financial availability of conservation projects [18].

Small sets, or panels, of ancestry informative markers (AIMs) have been developed and used to infer population genetic parameters in several species. AIM panels have been used to estimate biogeographical ancestry and structure in human populations [19–22], to discriminate among breeds and geographical origin of Italian sheep breeds [23], to identify cattle breeds [24–26], to trace the origin of animal products [27,28], and for breed assignment and analysis of individual ancestry in cattle [29–32] and pig populations [33]. Moreover, AIM panels have been developed to identify hybrids in wildlife conservation projects on wolf [34], wild cat [9], and mule deer [35].

In this work, we applied supervised machine learning approaches on mid-density DNA array data and identified AIMs able to correctly discriminate mouflon \times domestic sheep crosses when tested both on real and simulated data. Our results provide a fundamental conservation tool for the affordable identification of hybrid mouflon \times domestic sheep and for the quantification of their admixture level.

2. Materials and Methods

We collected publicly available genotype data (~50k SNPs) from 23 non-admixed feral Sardinian mouflon (MSar) and 23 Sarda sheep (SAR) (Figure S1, Supplementary Materials), and from 28 Sardinian mouflon hybrids (MxS) showing extensive levels of admixture according to previous analyses [12,36].

The Sarda sheep is an autochthonous Sardinian sheep breed counting almost four million heads in Sardinia. It is reared for its high milk production and accounts for almost all of the sheep presence in the island. Additionally, we collected genotypic data from 26 Lacaune (LAC), 28 Australian Poll Merino (MER), and 23 New Zealand Texel (TEX) made available by the Sheep HapMap project [37]. Lacaune is a milk-producing breed which has been recently imported into Sardinia, whereas Merino and Texel are two cosmopolitan breeds reared for wool and meat production, respectively [37].

Non-autosomal variants, markers with missing data (> 0), minor allele frequency (MAF) ≤ 0.1 , and markers significantly out of Hardy–Weinberg equilibrium (HWE ≤ 0.001) were excluded from the subsequent analyses. Pruning was performed using PLINK v1.9 [38].

2.1. Aims Identification and Panel Development

To identify the markers from the medium-density SNP array genotyping data for MSar and SAR able to detect admixture between mouflon and domestic sheep, we applied a two-fold approach which included (1) a preselection step aimed at removing the least significant markers, and (2) supervised machine learning classification approaches to identify the most informative markers among those left after preselection.

2.2. Preselection

The dataset was first reduced based on Principal Component Analysis (PCA) results. PCA is a widely used dimensionality reduction technique that generates new uncorrelated variables (principal components; PCs) sorted according to the variance they explain. The contribution of each SNP to every PC is expressed as a loading score. A PCA analysis of the reference breeds was performed using R v3.4.3 [39]. The first PC discriminated mouflon against domestic sheep, whereas the second PC described within-mouflon structure (see Results). Similarly, the following PCs identified within-population substructure. Consequently, we only considered the first PC. The loading scores of the first PC were squared (see [21,24]), and those SNPs associated with loading values exceeding 1.5x the interquartile range of the loading distribution were retained for further analyses (hereon: GW1; Figure S2, Supplementary Materials).

2.3. Supervised Classification

SNPs which passed preselection were submitted to supervised machine learning classification, as implemented in Boruta v6.0.0 [40], in order to select the most discriminant markers. This algorithm is based on a random forest learning process that allows, after a training step, to derive a set of rules for data classification. The Boruta algorithm measures the importance of an attribute (here, a SNP) through the loss of classification accuracy due to random permutation of the attribute values between objects [40]. Eventually, Boruta assigns the binary classification “confirmed important” and “confirmed unimportant” to each of the features tested. A single run of the Boruta algorithm was performed genome-wide to identify a first subset of significant SNPs (GW1) and those SNPs listed as “confirmed important” were assembled as a GW2 panel. To obtain an even smaller AIM panel, we performed 20 independent iterations of the classification algorithm on the GW1 panel; those SNPs listed as “confirmed important” in at least 19 scans out of 20 were pooled in a third panel (GW3) (Figure S3, Supplementary Materials).

With the aim to identify AIM panels which forcibly intercepted all autosomes, we applied the iteration algorithm to each autosome separately (Chromosome Wide, CW). Feature selection was performed for each autosome and SNPs listed as important were assembled in a panel named CH1. Lastly, to further reduce the number of significant SNPs, a last panel (CH2) was generated by selecting from CH1 the three most discriminant SNPs per each autosome.

2.4. AIM Panel Validation

The accuracy of each of the five AIM panels (three genome-wide and two chromosome-wide) to assess ancestry levels, or levels of admixture, was tested using genotype data from both simulated and real hybrid populations (MxS). To test the AIM panels with simulated hybrids, we developed Hybridiser v0.1, an R script able to simulate F1 hybrid individuals (available at <https://github.com/barbatom/Hybridiser>). Given two parental populations, Hybridiser computes the allele frequency at each locus, and then generates hybrid genotypes at each locus by selecting an allele from each of the parental populations with probability equal to the parental allele frequency. A dataset of 270 simulated hybrids (HYBs) was generated pooling 90 MSar \times SAR first-generation crosses (F1) and the reciprocal backcrosses F1 \times MSar (BC1M), and F1 \times SAR (BC1S), each comprising 90 individuals.

Admixture v1.3.0 [41] was used to perform supervised clustering tests to evaluate the ancestry proportions of both simulated and real hybrids using the pure mouflon and domestic populations as ancestry sources. To measure how well the AIM panels estimated the admixture level compared to that determined by the full set of markers, we compared the admixture results from each of the AIM panels (selected markers) with those from the full medium-density genotype array (full set of markers) using the coefficient of determination (r^2).

To test if the AIM panels performed better than an equally sized set of SNPs chosen at random, we generated 5,000 random AIM sets, and for each random set we performed supervised admixture analysis. Finally, we computed coefficients of determination values between the ancestry assignment of the full set and the reduced random panel. Being computationally challenging, this test was performed on the smallest AIM panel set we generated (GW3). The coefficient of determination values obtained using the 5,000 random SNP sets were standardised by z-scores. An empirical p -value for the correlation obtained using the GW3 panel was calculated according to Davison and Hinkley [42] as $p = (1 + x)/(1 + n)$, where x is the number of random sets that produced an r^2 score greater than or equal to that calculated using GW3 and n is the total number of random set tested (Figure S4, Supplementary Materials). Significance was set for p -values lower than 0.01.

Finally, we evaluated the ability of the five AIM panels to detect mouflon hybrids with domestic sheep breeds other than Sarda. For each mouflon \times domestic sheep population, we generated a dataset of simulated hybrids composed of: i) 90 F1 offspring between MSar and the test domestic breed, ii) 90 backcrosses between F1 and the test domestic breed and iii) 90 individuals obtained as backcross between F1 and MSar. The AIM panels were tested on simulated hybrid populations of MSar with MER, LAC, and TEX, respectively. Coefficients of determination were computed between the supervised admixture results obtained using AIM panels (selected markers) and the full medium-density genotype array (full set of markers).

3. Results

Pruning for missing data, MAF, and HWE left 33,481 non-rare, neutral SNPs. PCA was performed on the two reference populations (MSar and SAR) to evaluate how many PCs contributed in discriminating mouflon and domestic sheep. As expected, PC1 (18.7% of the total variance), split mouflon and domestic sheep as two distinct clusters (Figure 1, Figure S1, Supplementary Materials). PC2 (7.07% of the total variance) identified subpopulation structure in mouflon exclusively (Figure 1). Consequently, PC1 was considered the only relevant component for the preselection step.

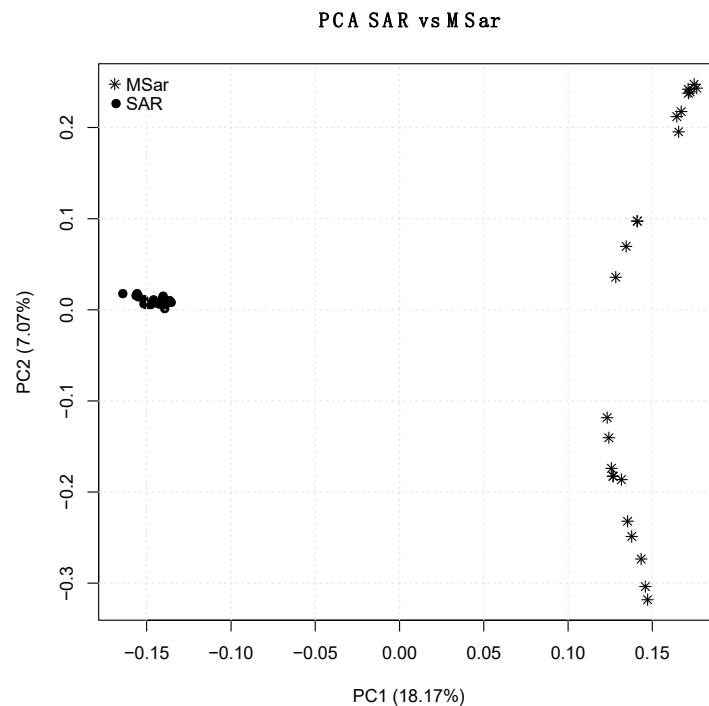


Figure 1. Principal Components Analysis (PC1 vs. PC2) of the two reference populations (MSar and SAR) analysed using the full single nucleotide polymorphism (SNP) set. In brackets are the percentage of variance explained by each component.

3.1. AIM Identification and Panel Development

The PCA-based preselection step identified 1,279 SNPs that contributed the most to discriminate between mouflon and domestic sheep reference populations (GW1; Table 1, Figure S2, Supplementary Materials). This first panel was submitted to both a single run and 20 iterations of random forest selection that identified 131 (GW2), and 51 SNPs (GW3), respectively (Table 1). A single run of random forest was applied chromosome-wide and identified 933 SNPs (CH1), with a range of 10 to 73 SNPs per chromosome (Table S1; Figure S5, Supplementary Materials). The selection of the three most significant SNPs per chromosome from CH1 selected 78 SNPs (CH2) (Table 1).

Table 1. Characteristics of the ancestry informative marker panels. The data processes used for marker selection were: preselection (Pre), Random Forest (RF), iterated Random Forest (iRF), and top-markers choice (tc). N is the number of SNPs in each panel. The SNP distribution per chromosome can be found in Supplementary Table S1.

Panel Name	Scope	Method	N
GW1	Genome-wide	Pre	1279
GW2	Genome-wide	Pre + RF	131
GW3	Genome-wide	Pre + RF + iRF	51
CH1	Chromosome-wide	Pre + RF	933
CH2	Chromosome-wide	Pre + RF + tc	78

3.2. Panels Validation

PCA analyses were used to visually compare the performance of the full set of SNPs and the five AIM panels applied to simulated (HYB) and real hybrid (MxS) populations.

The PCA of HYB using the full set of SNPs discriminated the parental populations (SAR and MSar) at opposite sides of the graph and positioned the hybrid populations according to their ancestry proportions, with F1 at the centre of the plot and the two backcrosses BC1S and BC1M closer to

SAR and MSar, respectively (Figure 2; Figure S6, Supplementary Materials). PCA of MxS using the full set of markers identified four individuals overlapping with the pure ancestry mouflon cluster, while the others were distributed along a gradient between MSar and SAR (Figure 3; Figure S7, Supplementary Materials).

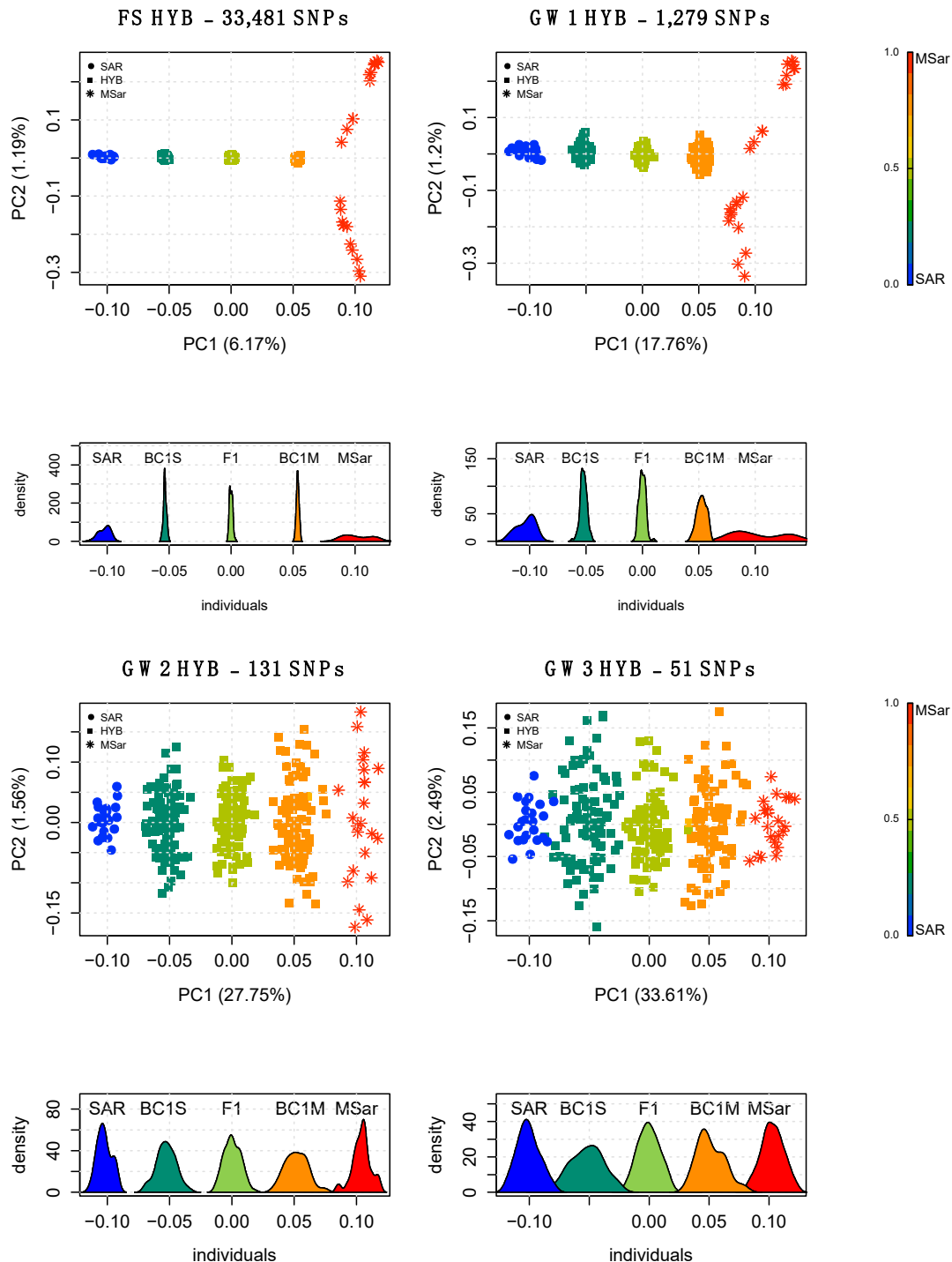


Figure 2. PCA and density distribution of the PC1 obtained using the full SNP set (top-left panel) and three AIMs on reference populations and simulated hybrids (HYB) using the genome-wide discovery approach. BC1S and BC1M are the simulated F1 backcrossed with SAR and MSar, respectively. The gradient legend on the right side of the plot shows the transition gradient from sheep to mouflon genetic components.

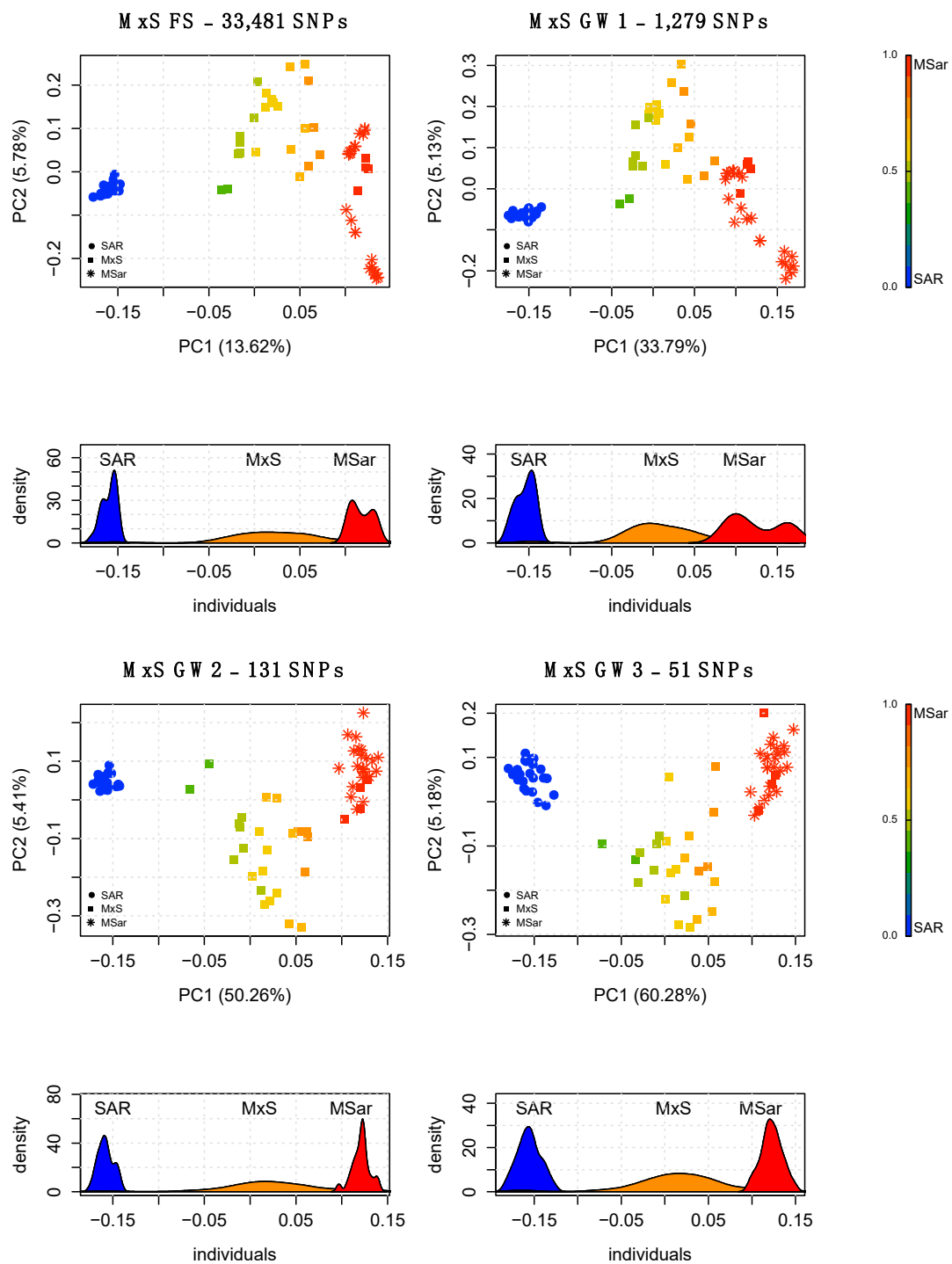


Figure 3. PCA and density distribution of the PC1 obtained using the full SNP set (top-left panel) and three AIMs on reference populations and real mouflon x domestic hybrids (MxS) using the genome-wide discovery approach. The gradient legend on the right side of the plot shows the transition gradient from sheep to mouflon genetic components. The analysis was performed using 33,481 SNPs and the three GW panels.

PCA performed on HYB using the AIMs showed clustering comparable with the PCA results obtained using the full set of SNPs (Figure 2; Figure S6). Each AIM panel was able to discriminate the simulated ancestry proportions and showed distinct clusters for each simulated hybrid population, with only minimal cluster overlap for three individuals when using the smallest SNP set (GW3;

Figure 2). When applied to the real population (MxS), GW1, GW2, GW3 and CH1 discriminated the pure and admixed individuals (Figure 3; Figure S7), whereas CH2 showed some overlap occurring between the pure mouflon and the admixed individuals having the smallest amount of domestic ancestry (Figure S7).

A quantitative assessment of the ancestry proportions in HYB was obtained through supervised Admixture using MSar and SAR as reference populations. HYB showed ancestry proportions coherent with the expected values of each offspring group (Figure S8, Supplementary materials); mean and standard deviation were 0.497 ± 0.0059 for F1, 0.234 ± 0.0059 for BC1S, and 0.76 ± 0.005 for BC1M. Supervised Admixture on MxS using the full set highlighted the heterogeneity of ancestry proportions in this population, and the presence of four individuals showing pure ancestry (Figure 4).

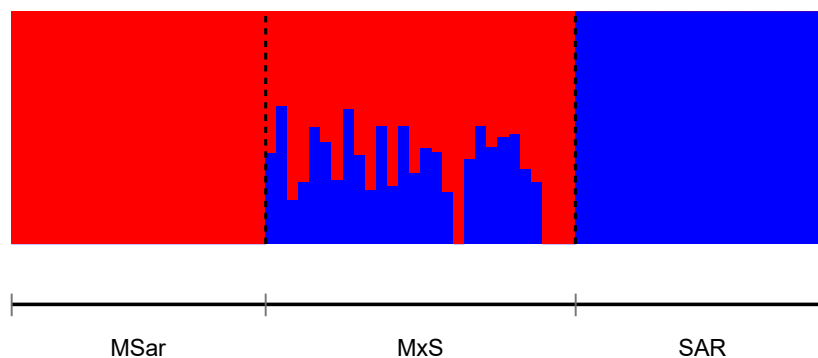


Figure 4. Supervised Admixture plot of MxS dataset obtained using the full set of SNPs. MSar and SAR were used as prior populations.

Coefficients of determination were then calculated between the ancestry proportions obtained using the full set of SNPs and AIM panels. The coefficients of determination were high overall across all panels ($r^2 \geq 0.960$), with GW1 and CH2 scoring the highest and lowest, respectively (Table 2). As expected, values were proportional to the number of AIMS in the panels. In simulated individuals, GW1 scored $r^2 = 0.990$ and CH1 $r^2 = 0.989$, while MxS had $r^2 = 0.997$ for both the panels. Coefficients of determination were slightly higher for AIMS identified using a genome-wide approach instead of chromosome-wide. The genome-wide GW3 panel performed better than the chromosome-wide CH2 panel despite the lower number of SNPs. AIM panels tested on HYB recorded higher correlation values compared to MxS.

Table 2. Coefficient of determination values (r^2) calculated between the ancestry percentages using the full set of SNPs and the AIM panels in the simulated (HYB) and case study (MxS) populations. N is the number of SNPs in each panel.

AIMs	N	HYB	MxS
GW1	1279	0.997	0.99
GW2	131	0.985	0.971
GW3	51	0.966	0.966
CH1	933	0.997	0.989
CH2	78	0.961	0.946

We further tested the ancestry assignments obtained using GW3 (the smallest panel among the AIMS generated) against the null hypothesis that the same results could be obtained using any equal sized set of SNPs chosen at random. The null hypothesis was rejected with high significance (p -value < 0.001 ; Figure S4, Supplementary Materials).

AIM panels were further tested to detect admixture between mouflon (MSar) and three commercial sheep breeds (TEX, APM, and LAC). Coefficients of determination calculated between the ancestry percentage obtained using the AIM panels and full set of SNP ancestry resulted in $r^2 > 0.92$ (Table 3).

Table 3. Coefficient of determination values calculated between the ancestry percentages obtained using the full SNP set and the AIMs in three commercial sheep breeds.

Breed	Acronym	GW1	GW2	GW3	CH1	CH2
New Zealand Texel	TEX	0.995	0.971	0.945	0.993	0.920
Australian Poll Merino	APM	0.995	0.968	0.938	0.993	0.923
Lacaune	LAC	0.994	0.962	0.928	0.992	0.923

4. Discussion

Hybridisation between wild species and their domestic counterpart may lead to genetic homogenization and loss of local adaptation [10]. Sardinian mouflon is the largest extant autochthonous European mouflon population and represents a unique genetic heritage threatened by hybridisation with domestic sheep [12,15,16].

In this work we implemented a supervised machine-learning-based classification approach to identify highly discriminant markers which can be included in rapid and low-cost diagnostic assays with the aim to support mouflon conservation [43]. The accuracy of AIM panels to detect introgression depends on the quality and sample size of the reference populations, with increasing probability of capturing most of the existing within-population variability when purer/non-admixed and larger reference datasets are used [44]. To the best of our knowledge, previous work on AIM panel discovery performed on humans, domestic and wild animals always accounted >100 reference samples [19,22–24,26,27,34,35]. Due to the large genetic distance occurring between sheep and mouflon, we achieved reliable feature selection using a sample size of 23 individuals for each reference population. Our results showed that the AIMs we identified can accurately discriminate Sardinian mouflon ancestry from other cosmopolitan sheep breeds as well. As expected, panels counting a higher number of SNPs, such as GW1 and CH1, resulted in better performances in hybrid identification than other panels. However, previous research showed that reliable results in individual assignment can be achieved also with a number of SNPs lower than 100 [21,27,29,32,43,45,46]. Accordingly, we obtained positive results also with less than 100 AIMs, as shown by tests performed with GW3 and CH2 panels. We tested AIM identification using both chromosome-wide and genome-wide approaches as done in previous research (see [45]). Our results confirmed the genome-wide approach to perform better than chromosome-wide. Chromosome-wide selection may introduce feature redundancy by selecting several SNPs on different chromosomes but carry identical discriminant power. Conversely, the genome-wide approach might be better at selecting highly discriminant SNPs which also capture a wider range of the focal population diversity. Noticeably, the distribution of AIMs in the genome-wide approach appears to be not homogeneous along the genome. Indeed, two phylogenetically close references, as in this case, are likely to share most of the genome, with few genomic regions responsible for much of the genomic distance, further supporting the genome- over the chromosome-wide approach as best suited to intercept such regions.

We assessed the performance of the panels in identifying crosses between domestic sheep and feral mouflon using both simulated and real data. As expected, using the AIMs on simulated data performed moderately better than on real admixed mouflon samples. Real admixed populations present a more complex genetic make-up, influenced by demography, selection, inbreeding and introgression events occurred during centuries. Conversely, the simulated individuals were generated from the same reference populations used to select the best AIMs. In addition, the mating system applied in simulations generates simplified admixture patterns with respect to those occurring in real populations. However, the use of simulated hybrid genotypes may aid in AIM panels testing to overcome the lack of real admixed samples. When available, real hybrids samples can guarantee the accuracy of the AIMs panels.

5. Conclusions

The detection of hybrids is a fundamental task in biodiversity conservation, and genetics provides an unmatched tool to identify even cryptic levels of hybridization. Applying a machine learning-based approach, we identified reduced SNP panels which proved effective in identifying domestic sheep introgression in Sardinian mouflon. The AIM panels we propose can be applied for large-scale assessment of the Sardinian mouflon population, and aid in the conservation of this unique genetic resource. Lastly, the method we present can be easily applied to other wildlife and domestic species conservation to develop accurate and affordable tools for hybrid identification.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2076-2615/10/4/582/s1>, Figure S1: Unsupervised Admixture analysis for K=2 evaluated on the full set of reference populations (SAR and MSar). Figure S2: Plot of PC1 loadings, squared and ordered. Red dots represent SNPs selected. Figure S3: For each one of the 20 subsequent iterations is shown the consensus number of SNPs “confirmed important” in all the performed iterations. Figure S4: plot of the Pearson correlation values density distribution of random panels test. The green line represents the correlation value obtained for the GW3 panel. Figure S5: distribution of SNPs per chromosome in panel CH1. Figure S6: Principal Components Analysis (PC1 vs PC2) analysis and density distribution of mouflon (MSar), hybrid (HYB), and domestic sheep (SAR) populations. The gradient legend on the right side of the plot shows the transition gradient from sheep to mouflon genetic components. The analysis was performed using panel CH1 and CH2. Figure S7: Principal Components Analysis (PC1 vs PC2) analysis and density distribution of mouflon (MSar), mouflon x domestic hybrid (MxS), and domestic sheep (SAR) populations. The gradient legend on the right side of the plot shows the transition gradient from sheep to mouflon genetic components. The analysis was with CH1 and CH2. Figure S8: supervised admixture K=2 plot evaluated with full set on the HYB dataset. Table S1: distribution of SNPs per chromosome: for each AIMs panel is reported the exact number of SNPs per chromosome and the percentage related to FS.

Author Contributions: Conceptualization, M.B.; methodology, M.B.; software, M.B., E.S.; validation, M.B., E.S.; formal analysis, M.B., E.S.; investigation, M.B., E.S.; data curation, M.B.; writing—original draft preparation, E.S.; writing—review and editing, M.B., P.A.-M.; visualization, E.S.; supervision, P.A.-M.; funding acquisition, P.A.-M. All authors have read and agreed to the published version of the manuscript.

Funding: Elisa Somenzi was partially supported by the Doctoral School on the Agro-Food System (Agrisystem) of the Università Cattolica del Sacro Cuore (Italy), and the ‘Fondazione Romeo ed Enrica Invernizzi’, Milan, Italy.

Acknowledgments: We thank the two reviewers for useful suggestions to improve this paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Barbato, M.; Hailer, F.; Upadhyay, M.; Del Corvo, M.; Colli, L.; Negrini, R.; Kim, E.-S.; Crooijmans, R.P.M.A.; Sonstegard, T.; Marsan, P.A. Adaptive introgression from indicine cattle into white cattle breeds from Central Italy. *Sci. Rep.* **2020**, *10*, 1–11. [\[CrossRef\]](#)
2. Bruford, M.W.; Ginja, C.; Hoffmann, I.; Joost, S.; Orozco-Terwengel, P.; Alberto, F.J.; Amaral, A.; Barbato, M.; Biscarini, F.; Colli, L.; et al. Prospects and challenges for the conservation of farm animal genomic resources, 2015–2025. *Front. Genet.* **2015**, *6*, 337. [\[CrossRef\]](#)
3. Dowling, T.E.; Secor, A.C.L. The role of hybridization and introgression in the diversification of animals. *Annu. Rev. Ecol. Syst.* **1997**, *28*, 593–619. [\[CrossRef\]](#)
4. Hedrick, P.W. Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* **2013**, *22*, 4606–4618. [\[CrossRef\]](#)
5. Hu, X.-J.; Yang, J.; Xie, X.-L.; Lv, F.-H.; Cao, Y.-H.; Li, W.-R.; Liu, M.-J.; Wang, Y.-T.; Li, J.-Q.; Liu, Y.-G.; et al. The Genome Landscape of Tibetan Sheep Reveals Adaptive Introgression from Argali and the History of Early Human Settlements on the Qinghai–Tibetan Plateau. *Mol. Biol. Evol.* **2018**, *36*, 283–303. [\[CrossRef\]](#)
6. Iacolina, L.; Corlatti, L.; Buzan, E.; Safner, T.; Šprem, N. Hybridisation in European ungulates: An overview of the current status, causes, and consequences. *Mammal Rev.* **2018**, *49*, 45–59. [\[CrossRef\]](#)
7. Miller, J.M.; Poissant, J.; Hogg, J.T.; Coltman, D.W. Genomic consequences of genetic rescue in an insular population of bighorn sheep (*Ovis canadensis*). *Mol. Ecol.* **2012**, *21*, 1583–1596. [\[CrossRef\]](#)

8. Olden, J.D.; Poff, N.L.; Douglas, M.R.; Douglas, M.; Fausch, K.D. Ecological and evolutionary consequences of biotic homogenization. *Trends Ecol. Evol.* **2004**, *19*, 18–24. [[CrossRef](#)]
9. Oliveira, R.; Randi, E.; Mattucci, F.; Kurushima, J.D.; Lyons, L.A.; Alves, P.C. Toward a genome-wide approach for detecting hybrids: Informative SNPs to detect introgression between domestic cats and European wildcats (*Felis silvestris*). *Hered.* **2015**, *115*, 195–205. [[CrossRef](#)]
10. Allendorf, F.W.; Leary, R.F.; Spruell, P.; Wenburg, J.K. The problems with hybrids: Setting conservation guidelines. *Trends Ecol. Evol.* **2001**, *16*, 613–622. [[CrossRef](#)]
11. Mallet, J. Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **2005**, *20*, 229–237. [[CrossRef](#)]
12. Barbato, M.; Hailer, F.; Orozco-Terwengel, P.; Kijas, J.; Mereu, P.; Cabras, P.; Mazza, R.; Pirastru, M.; Bruford, M.W. Genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Sci. Rep.* **2017**, *7*, 7623. [[CrossRef](#)]
13. Sanna, D.; Barbato, M.; Hadjisterkotis, E.; Cossu, P.; DeCandia, L.; Trova, S.; Pirastru, M.; Leoni, G.G.; Naitana, S.; Francalacci, P.; et al. The First Mitogenome of the Cyprus Mouflon (*Ovis gmelini ophion*): New Insights into the Phylogeny of the Genus *Ovis*. *PLoS ONE* **2015**, *10*, 0144257. [[CrossRef](#)]
14. Chessa, B.; Pereira, F.; Arnaud, F.; Amorim, A.; Goyache, F.; Mainland, I.; Kao, R.R.; Pemberton, J.M.; Beraldi, D.; Stear, M.; et al. Revealing the History of Sheep Domestication Using Retrovirus Integrations. *Science* **2009**, *324*, 532–536. [[CrossRef](#)]
15. Mereu, P.; Pirastru, M.; Barbato, M.; Satta, V.; Hadjisterkotis, E.; Manca, L.; Naitana, S.; Leoni, G.G. Identification of an ancestral haplotype in the mitochondrial phylogeny of the ovine haplogroup B. *PeerJ* **2019**, *7*, e7895. [[CrossRef](#)]
16. Lorenzini, R.; Cabras, P.; Fanelli, R.; Carboni, G.L. Wildlife molecular forensics: Identification of the Sardinian mouflon using STR profiling and the Bayesian assignment test. *Forensic Sci. Int. Genet.* **2011**, *5*, 345–349. [[CrossRef](#)]
17. Rhymer, J.M.; Simberloff, D. Extinction by hybridization and introgression. *Annu. Rev. Ecol. Syst.* **1996**, *27*, 83–109. [[CrossRef](#)]
18. Ruane, A.; Sonnino, J. Background document to the e-mail conference on the role of biotechnology for the characterization and conservation of crop, forest, animal and fishery genetic resources in developing countries. In *The Role of Biotechnology in Exploring and Protecting Agricultural Genetic Resources*; FAO: Rome, Italy, 2006; pp. 151–172.
19. Halder, I.; Shriver, M.D.; Thomas, M.; Fernández, J.R.; Frudakis, T. A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: Utility and applications. *Hum. Mutat.* **2008**, *29*, 648–658. [[CrossRef](#)]
20. Kosoy, R.; Nassir, R.; Tian, C.; White, P.A.; Butler, L.M.; Silva, G.; Kittles, R.; Riquelme, M.E.A.; Gregersen, P.K.; Belmont, J.W.; et al. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Huma. Mutat.* **2009**, *30*, 69–78. [[CrossRef](#)]
21. Paschou, P.; Ziv, E.; Burchard, E.G.; Choudhry, S.; Rodriguez-Cintron, W.; Mahoney, M.W.; Drineas, P. PCA-Correlated SNPs for Structure Identification in Worldwide Human Populations. *PLoS Genet.* **2007**, *3*. [[CrossRef](#)]
22. Paschou, P.; Drineas, P.; Lewis, J.; Nievergelt, C.M.; Nickerson, D.A.; Smith, J.D.; Ridker, P.M.; Chasman, D.I.; Krauss, R.M.; Ziv, E. Tracing Sub-Structure in the European American Population with PCA-Informative Markers. *PLoS Genet.* **2008**, *4*, e1000114. [[CrossRef](#)]
23. DiMauro, C.; Nicoloso, L.; Cellesi, M.; Macciotta, N.P.P.; Ciani, E.; Moioli, B.; Pilla, F.; Crepaldi, P. Selection of discriminant SNP markers for breed and geographic assignment of Italian sheep. *Small Rumin. Res.* **2015**, *128*, 27–33. [[CrossRef](#)]
24. Bertolini, F.; Galimberti, G.; Calò, D.; Schiavo, G.; Matassino, D.; Fontanesi, L. Combined use of principal component analysis and random forests identify population-informative single nucleotide polymorphisms: Application in cattle breeds. *J. Anim. Breed. Genet.* **2015**, *132*, 346–356. [[CrossRef](#)]
25. Bertolini, F.; Galimberti, G.; Schiavo, G.; Mastrangelo, S.; Di Gerlando, R.; Strillacci, M.G.; Bagnato, A.; Portolano, B.; Fontanesi, L. Preselection statistics and Random Forest classification identify population informative single nucleotide polymorphisms in cosmopolitan and autochthonous cattle breeds. *Animal* **2017**, *12*, 12–19. [[CrossRef](#)]

26. Wilkinson, S.; Wiener, P.; Archibald, A.; Law, A.; Schnabel, R.D.; McKay, S.; Taylor, K.; Ogden, R. Evaluation of approaches for identifying population informative markers from high density SNP Chips. *BMC Genet.* **2011**, *12*, 45. [CrossRef]
27. DiMauro, C.; Cellesi, M.; Steri, R.; Gaspa, G.; Sorbolini, S.; Stella, A.; Macciotta, N.P.P. Use of the canonical discriminant analysis to select SNP markers for bovine breed assignment and traceability purposes. *Anim. Genet.* **2013**, *44*, 377–382. [CrossRef]
28. Orrù, L.; Catillo, G.; Napolitano, F.; De Matteis, G.; Scatà, M.C.; Signorelli, F.; Moioli, B. Characterization of a SNPs panel for meat traceability in six cattle breeds. *Food Control.* **2009**, *20*, 856–860. [CrossRef]
29. Frkonja, A.; Gredler, B.; Schnyder, U.; Curik, I.; Sölkner, J. Prediction of breed composition in an admixed cattle population. *Anim. Genet.* **2012**, *43*, 696–703. [CrossRef]
30. Kumar, H.; Panigrahi, M.; Chhotaray, S.; Pal, D.; V., B.; Ka, S.; Shandilya, R.; Parida, S.; Bhushan, B. Identification of breed-specific SNP panel in nine different cattle genomes. *Biomed. Res.* **2019**, *30*, 78–81. [CrossRef]
31. Gorbach, D.; Makgahlela, M.; Reecy, J.; Kemp, S.; Baltenweck, I.; Ouma, R.; Mwai, O.; Marshall, K.; Murdoch, B.M.; Moore, S.; et al. Use of SNP genotyping to determine pedigree and breed composition of dairy cattle in Kenya. *J. Anim. Breed. Genet.* **2010**, *127*, 348–351. [CrossRef]
32. Lewis, J.; Abas, Z.; Dadousis, C.; Lykidis, D.; Paschou, P.; Drineas, P. Tracing Cattle Breeds with Principal Components Analysis Ancestry Informative SNPs. *PLoS ONE* **2011**, *6*, e18007. [CrossRef]
33. Liang, Z.; Bu, L.; Qin, Y.; Peng, Y.; Yang, R.; Zhao, Y. Selection of Optimal Ancestry Informative Markers for Classification and Ancestry Proportion Estimation in Pigs. *Front. Genet.* **2019**, *10*, 183. [CrossRef]
34. Randi, E.; Hulva, P.; Fabbri, E.; Galaverni, M.; Galov, A.; Kusak, J.; Bigi, D.; Bolfíková, B.Č.; Smetanová, M.; Caniglia, R. Multilocus Detection of Wolf x Dog Hybridization in Italy, and Guidelines for Marker Selection. *PLoS ONE* **2014**, *9*, e86409. [CrossRef]
35. Russell, T.; Cullingham, C.I.; Kommadath, A.; Stothard, P.; Herbst, A.; Coltman, D.W. Development of a Novel Mule Deer Genomic Assembly and Species-Diagnostic SNP Panel for Assessing Introgression in Mule Deer, White-Tailed Deer, and Their Interspecific Hybrids. *G3 Genes Genomes Genet.* **2019**, *9*, 911–919. [CrossRef]
36. Ciani, E.; Crepaldi, P.; Nicoloso, L.; Lasagna, E.; Sarti, F.M.; Moioli, B.; Napolitano, F.; Carta, A.; Usai, G.; D'Andrea, M.; et al. Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. *Anim. Genet.* **2013**, *45*, 256–266. [CrossRef]
37. Kijas, J.W.; Lenstra, J.A.; Hayes, B.; Boitard, S.; Porto-Neto, L.R.; Cristobal, M.S.; Servin, B.; McCulloch, R.; Whan, V.; Gietzen, K.; et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* **2012**, *10*, e1001258. [CrossRef]
38. Chang, C.C.; Chow, C.; Tellier, L.C.A.M.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **2015**, *4*, 7. [CrossRef]
39. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing **2012**. Available online: <https://www.R-project.org/> (accessed on 30 March 2020).
40. Kursa, M.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*, 1–13. [CrossRef]
41. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **2009**, *19*, 1655–1664. [CrossRef]
42. Davison, A.C.; Hinkley, D.V. *Bootstrap Methods and Their Application*; Cambridge University Press: Cambridge, UK, 1997.
43. Vonholdt, B.M.; Pollinger, J.P.; Earl, D.A.; Parker, H.G.; Ostrander, E.A.; Wayne, R.K. Identification of recent hybridization between gray wolves and domesticated dogs by SNP genotyping. *Mamm. Genome* **2013**, *24*, 80–88. [CrossRef]
44. Hulsege, I.; Calus, M.P.L.; Windig, J.; Hoving-Bolink, A.H.; Eijndhoven, M.H.T.M.-V.; Hiemstra, S.J. Selection of SNP from 50K and 777K arrays to predict breed of origin in cattle1. *J. Anim. Sci.* **2013**, *91*, 5128–5134. [CrossRef]

45. Bertolini, F.; The ADAPTMAP Consortium; Cardoso, T.F.; Marras, G.; Nicolazzi, E.; Rothschild, M.F.; Amills, M. AdaptMap Consortium Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genet. Sel. Evol.* **2018**, *50*, 59. [[CrossRef](#)]
46. Kersbergen, P.; Van Duijn, K.; Kloosterman, A.D.; Dunnen, J.T.D.; Kayser, M.; De Knijff, P. Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans. *BMC Genet.* **2009**, *10*, 69. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).