



OPEN

A comparative analysis of *SLA-DRB1* genetic diversity in Colombian (creoles and commercial line) and worldwide swine populations

Carmen Teresa Celis-Giraldo^{1,2,9}, Michel David Bohórquez^{3,4,9}, Milena Camargo^{3,9}, Carlos Fernando Suárez³, Anny Camargo^{2,3}, Kewin Rodríguez-Obediente^{3,4}, Alejandra Martínez¹, Carlos Edmundo Lucero⁵, Byron Hernández⁵, Raúl Manzano-Román⁶ & Manuel Alfonso Patarroyo^{3,7,8}✉

Analysing pig class II major histocompatibility complex (MHC) molecules is mainly related to antigen presentation. Identifying frequently-occurring alleles in pig populations is an important aspect to be considered when developing peptide-based vaccines. Colombian creole pig populations have had to adapt to local conditions since entering Colombia; a recent census has shown low amounts of pigs which is why they are considered protected by the Colombian government. Commercial hybrids are more attractive regarding production. This research has been aimed at describing the allele distribution of Colombian pigs from diverse genetic backgrounds and comparing Colombian *SLA-DRB1* locus diversity to that of internationally reported populations. Twenty *SLA-DRB1* alleles were identified in the six populations analysed here using sequence-based typing. The amount of alleles ranged from six (Manta and Casco Mula) to nine (San Pedreño). Only one allele (01:02) having > 5% frequency was shared by all three commercial line populations. Allele 02:01:01 was shared by five populations (around > 5% frequency). Global F_{ST} indicated that pig populations were clearly structured, as 20.6% of total allele frequency variation was explained by differences between populations ($F_{ST} = 0.206$). This study's results confirmed that the greatest diversity occurred in wild boars, thereby contrasting with low diversity in domestic pig populations.

Pig species are of great interest as a source of protein for human beings and due to their close relationship to humans in terms of anatomy, physiology and genetics^{1–3}. This makes them a good model for biomedical research in areas such as immunology, therapeutics, production and xenotransplantation; characterising their immunological systems will lead to advances in such disciplines^{4–6}. MHC antigen presentation has been widely studied in various species regarding functions associated with the immune system⁷. Genetic diversity has been comparatively analysed in the *Suidae* family concerning *Sus scrofa* (wild boars) and their congener *Sus scrofa domesticus* (domestic pig), along with independent analysis^{8–12}.

The MHC is called the swine leukocyte antigen (SLA) in the *Sus scrofa* species; it is located on chromosome 7 and is formed by class I and class III (position 7p1.1) and class II complex regions (position 7q1.1)¹⁰. Class I and class II complex genes are related to immunity and inflammation and are characterised by being highly

¹Animal Science Faculty, Universidad de Ciencias Aplicadas y Ambientales (U.D.C.A), 111166 Bogotá, Colombia. ²PhD Programme in Biomedical and Biological Sciences, School of Medicine and Health Sciences, Universidad del Rosario, 112111 Bogotá, Colombia. ³Molecular Biology and Immunology Department, Fundación Instituto de Inmunología de Colombia (FIDIC), 111321 Bogotá, Colombia. ⁴Microbiology Postgraduate Programme, Universidad Nacional de Colombia, 111321 Bogotá, Colombia. ⁵Germplasm Bank, Agrosavia, 250047 Bogotá, Colombia. ⁶Proteomics Unit, Cancer Research Center (IBMCC/CSIC/USAL/IBSAL), 37007 Salamanca, Spain. ⁷Faculty of Medicine, Universidad Nacional de Colombia, 111321 Bogotá, Colombia. ⁸Health Sciences Division, Main Campus, Universidad Santo Tomás, 110231 Bogotá, Colombia. ⁹These authors contributed equally: Carmen Teresa Celis-Giraldo, Michel David Bohórquez and Milena Camargo. ✉email: mapatar.fidic@gmail.com

polymorphic⁴. Interestingly, the class II complex expresses proteins on the surface of antigen presenting cells (e.g. macrophages and dendritic cells); forming a protein complex, peptides anchored to the peptide-binding region (PBR) and the T-lymphocyte receptor (TCR) promote a humoral immune response^{13–17}. *SLA-DRB1* (99 alleles) and *SLA-DQB1* (53 alleles) have the greatest polymorphism (<https://www.ebi.ac.uk/ipd/mhc/>). These genes consist of six exons; exon 2 encodes the PBR and has the greatest polymorphism^{4,10,12}, thereby providing it with greater diversity for presenting peptides¹⁸. This study did not consider analysing the *SLA-DRA* gene because it has poor genetic variability, as happens in humans^{4,10,19}.

Studies have evaluated MHC-Class II haplotype relationship with such parameters as production²⁰, reproduction²¹ and/or genetic characterisation^{8,13,22}. Several studies in the field of biomedical research have concluded that *SLA-DRB1* is an xenoantigen⁵; some studies have even been aimed at determining which pig alleles have the greatest reactivity²³ whilst others have dealt with xenotransplantation^{5,24,25}. MHC gene variability could be affected by natural and/or artificial selection^{26,27}; regarding diversity, allele determination in a target population is thus relevant when designing epitope-based vaccines where it is recommended that the most prevalent pig population alleles (super-types) are identified so that such vaccines are designed to serve the majority^{28,29}.

Local pigs are used in various traditional production systems worldwide, including native pigs in Asia, Africa and the Mediterranean (Iberian pigs); they are called creole pigs in Latin American countries³⁰. The different breeds used in the modern pig industry have a greater impact on the rural economy because they are considered a low-cost protein source^{31,32}. The appearance of the domestic pig in America arose from the introduction of Iberian or hairless pigs during the time of the Conquest; they have become adapted to local conditions over a period lasting more than 500 years, having some special phenotypical characteristics^{31,33}. Three breeds are currently located in specific areas of Colombia. The nucleus of the Casco Mula breed is on Colombia's Eastern Plains (Llanos Orientales) and the foothills of the Eastern Plains region (Piedemonte Llanero), the San Pedreño breed is in the Antioquia and Viejo Caldas departments and the Zungo breed is found on the Atlantic coast (mainly in the Córdoba department). The latter region is considered the place where the first specimens arrived in Colombia^{34,35}.

Phenotypically, the San Pedreño breed is black, has a small head, a short snout and straight/upright ears; the Casco Mula breed's name arises from syndactyly, as the pigs are multi-coloured, have red and sometimes black hair, a medium-sized snout, concave head forward-sloping, large ears and strong, short legs. The Zungo breed is characterised by being black and hairless, large-sized and having floppy ears; it has been reported that pigs from the Caribbean region have been crossed with pigs from the Duroc, Poland China and Hampshire breeds³¹.

These breeds' living nuclei are kept in germplasm banks in the Colombian Agricultural Research Corporation's (AGROSAVIA) research centres because they have a nucleus consisting of less than 1000 individuals³⁵. According to previous reports these breeds are characterised by being able to produce meat and reproduce in difficult environmental situations³². Several studies have used genetic markers for better characterising breeds regarding their production potential^{32–34,36}; a germplasm bank-based study of the San Pedreño breed recently characterised its genetic structure³⁵.

The creole pig's adaptation to varying environmental conditions has promoted the development of their productive and reproductive abilities. Production-related genetic improvement programmes in Colombia have enabled the development of commercial hybrids whose production-related qualities abilities have favoured their widespread distribution in Colombia (www.solla.com)^{32,37}. This research was thus aimed at genotyping the *SLA-DRB1* locus in pig populations from Colombian creole breeds and commercial hybrids and comparing the diversity amongst Colombian populations and to that of some previously reported international ones.

Results

The study populations' genetic diversity indices and phylogenetic analysis. The study involved sampling 188 pigs from locations in Colombia; 91 pigs were creoles and 97 came from a commercial line (Fig. 1). Twenty *SLA-DRB1* alleles were found in the population, five of which were only found in creole pigs (i.e. *SLA-DRB1**02:05, *04:01, *09:01:01) and another seven were unique for the commercial pig group (i.e. *SLA-DRB1**01:01, *02:11, *10:01:01).

Regarding allele distribution according to population, the creole pig results showed that Casco Mula, San Pedreño and Zungo shared eight alleles; two alleles (*SLA-DRB1**01:02 and *02:01:01) being found in all three breeds. The *SLA-DRB1**02:01:01 allele was shared by five populations where it occurred with > 5% frequency; moreover, this allele had > 40% frequency in creole populations. The San Pedreño population had the greatest amount of alleles (Fig. 2a, Table S1). Eight alleles were shared amongst individuals from the commercial line (Remanso, Manta and Ubaque) and two alleles (*SLA-DRB1**01:01, *01:02) were shared amongst the three populations. Four unique alleles (*SLA-DRB1**01:05, *02:01:03, *04:02 and *10:04) were observed in the Remanso population (U.D.C.A) (Fig. 2b).

Commercial line individuals had greater homozygosity (64%), the highest being observed in the Manta population (74.2%), followed by Remanso (71.0%). The Casco Mula population had the greatest homozygosity in the creole pig group (73.3%) (Fig. 3).

Diversity index results per population showed that the San Pedreño creole pig breed had the fewest segregating sites (S: 39) and haplotypes (h: 9), whilst the Zungo breed had the greatest nucleotide (π : 0.09214) and haplotype (Hd: 0.753) diversity (Table 1). The pig production group had the greatest nucleotide diversity (π : 0.06942), theta from eta (Θ : 0.06404) and haplotype diversity (Hd: 0.797) in the Manta population (Table 1).

The amount of recombination events for creole pigs (3 events) was slightly higher than that for the production animals (2 events). No recombination events were found regarding Casco Mula; the recombinant *SLA-DRB1**09:02 allele was found in both populations (Fig. S2).

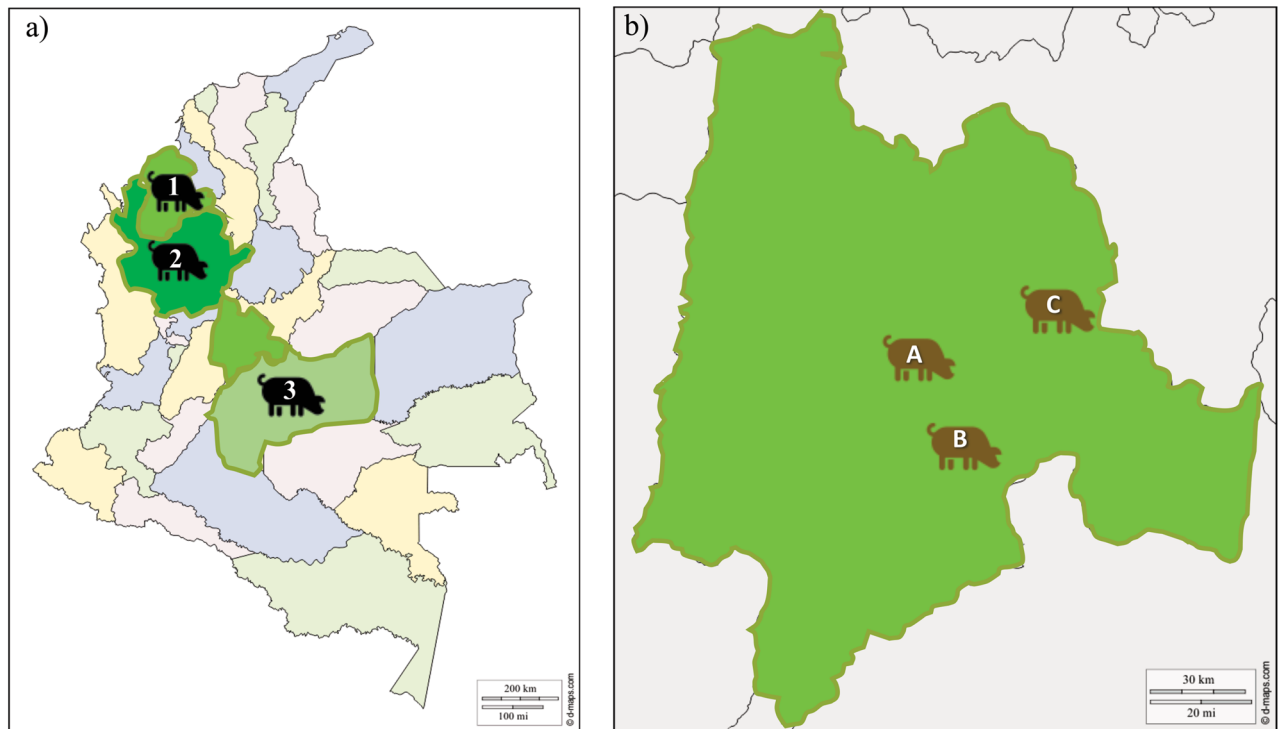


Figure 1. Sampling location sites. (a) Map of Colombia showing the location of AGROSAVIA's research centres (I.C) having creole pig populations (pictogram in black), 1 = Zungo (IC Turipaná, Córdoba, $n = 26$), 2 = San Pedroño (IC El Nus, Antioquia, $n = 35$) and 3 = Casco Mula (IC La Libertad, Meta, $n = 30$)⁸². (b) Enlargement of the Cundinamarca department showing the location of the commercial line population (pictogram in brown), A = El Remanso-U.D.C.A ($n = 31$), B = Ubaque ($n = 35$) and C = Manta ($n = 31$). These free maps were downloaded from d-maps (<https://d-maps.com/>); the images are freely available and modifiable in accordance with d-maps policies^{83,84}. Maps were modified using PowerPoint Software (Version 2002) Microsoft Office 365 ProPlus (U.D.C.A institutional license) by CTCG”.

SLA-DRB1 allele distribution compared to that in other populations. *DRB1**02:01:01, *10:01:01, *01:01, *09:01:01 and *01:02 alleles were found (often having very high frequencies) in several Colombian, European and Asian populations. Several alleles were unique to one population and often found to have low frequency, including the *DRB1**10:04 (Remanso) and *02:05 (San Pedroño) alleles. Some alleles were not found in Colombian populations, being restricted to just some European (*DRB1**06:02, *06:03, *06:07, *05:01 and *13:01) or both European and Asian populations (*DRB1**04:04, *08:01 and *14:01) (Table S3).

Hardy–Weinberg equilibrium. Observed and expected heterozygosity and F_{IS} were determined for these populations (Table 2) to examine *SLA-DRB1* locus variability and potential Hardy–Weinberg variation. It was found that h_o was lower in the commercial line than in creole pigs, the Zungo ($h_o:0.538$) breed having the greatest h_o amongst all six populations, whereas Casco Mula ($h_o:0.267$) had a h_o comparable to that of Remanso ($h_o:0.290$) and Manta ($h_o:0.258$). Casco Mula ($h_e:0.380$) had the lowest h_e value, which was twice as low as those for the San Pedroño ($h_e:0.728$) and Zungo ($h_e:0.743$) breeds. Remanso ($h_e:0.558$) and Manta ($h_e:0.563$) had the lowest h_e values amongst commercial line populations, due to the extremely biased allele frequency distribution in these populations. All populations had higher expected than observed heterozygosity and had significant heterozygote deficiency, being almost twice as high in the Manta (highest value) than the Casco Mula breed (lowest value). Both commercial line and creole pigs had the lowest observed heterozygosity (except for Duroc pigs) compared to that for other pig populations, although expected heterozygosity levels were more similar to those for other populations. Allele richness (the amount of alleles corrected for sampling size effects) had a clear sample size-related effect on the amount of alleles identified (Table 2). As expected, neither h_o nor h_e appeared to be affected by sample size ($h_o r = 0.13$ and $h_e r = 0.17$), indicating that these estimates should be good representatives for *SLA-DRB1* diversity in these pig populations. Colombian creole and commercial line pig populations had the highest Hardy–Weinberg variations (highest F_{IS} values).

The mean amount of nonsynonymous (N_d) and synonymous (S_d) polymorphisms was calculated to evaluate sequence diversity amongst populations. Considering the whole $\beta 1$ domain, S_d ranged from 2.32 (Remanso) to 8.65 (Micromini) and N_d from 7.80 (Casco Mula) to 19.02 (KNP). Colombian populations had the lowest values for both N_d and S_d , despite the amount of alleles or sample size being the lowest amongst all populations. This indicated a less divergent gene pool for the Colombian populations. When only considering PBR codons, S_d ranged from 0.70 (Remanso) to 2.79 (Micromini) and N_d from 7.0 (Duroc) to 11.11 (KNP); Colombian populations, Yorkshire_E, SNU and Duroc also had low levels of diversity. As expected, a large percentage (54%

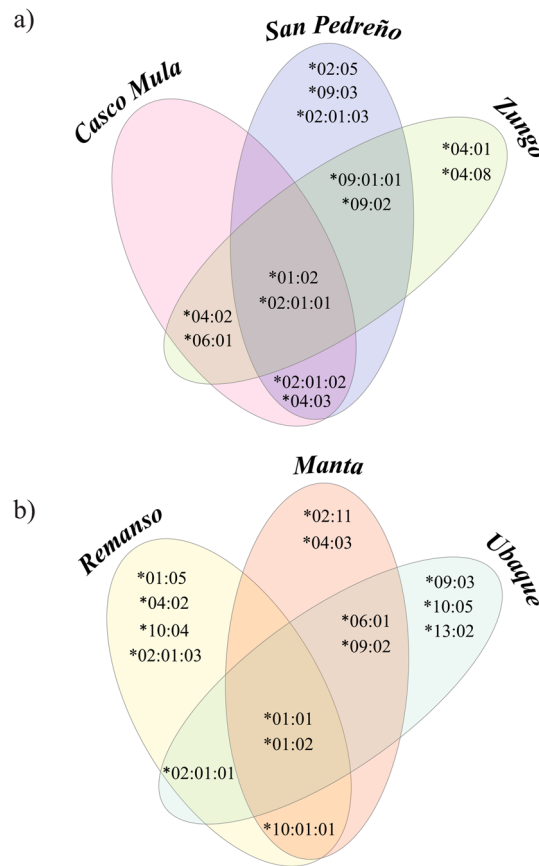


Figure 2. Venn diagram comparing the distribution of specific and shared *SLA-DRB1* alleles according to the target population. **(a)** Creole pigs. **(b)** commercial line.

to 82%) of N_d in the $\beta 1$ domain was due to PBR positions, confirming these positions being the most variable ones (Table 3).

Population structure and genetic differentiation. The overall F_{ST} indicated that the pig populations were clearly structured; 17.4% of total variation for allele frequency could be explained by differences between populations ($F_{ST}=0.174$; p -value < 0.0001) (Table S4). D_A genetic distances based on allele frequency were used to construct dendrograms using the NJ and BIONJ algorithms to further evaluate relationships between populations, the BIONJ algorithm (Fig. 4a) having the best adequacy ($BIONJ\ RQS=0.698$, $NJ\ RQS=0.667$). Multidimensional scaling (MDS), using the D_A genetic distance matrix, was explored to depict population relationships in two dimensions (Fig. 4b). MDS also gave good representation of real variability ($RQS=0.663$) and the results largely agreed with the BIONJ dendrogram.

Despite the populations' homogeneous MDS distribution, three well-supported clusters could be identified in the dendrogram (bootstrap support ≥ 85). The first was a compact group consisting of Duroc, Zungo, Casco Mula and San Pedroño. This group had a 0.26 mean D_A distance and the alleles shared by its members ($DRB1^*01:02$, $^*02:01:01$, $^*02:01:02$, $^*04:02$, $^*04:03$, $^*04:08$, $^*06:01$ and $^*09:01:01$) accounted for 76% of its mean absolute allele frequency (Table 2). The second cluster was made up of Ubaque and Manta; it had a 0.358 D_A distance and 75% mean absolute frequency (explained by alleles $DRB1^*01:01$, $^*01:02$, $^*06:01$ and $^*09:02$). The third cluster consisted of Landrace_C and Micromini, having a 0.432 D_A distance and its common alleles ($DRB1^*06:01$, $^*08:01$ and $^*10:01:01$) explained 63% of mean absolute allele frequency (Table 2). Yorkshire and Landrace clustered in both the MDS and the dendrogram and, although not very well supported by bootstrap, this cluster had a 0.259 D_A distance and 88% mean absolute allele frequency (alleles $DRB1^*01:01$, $^*01:02$, $^*02:01:01$, $^*04:02$, $^*04:04$, $^*06:03$, $^*09:01:01$ and $^*10:01:01$) (Fig. 4 and Table 2).

A clustering method was used for assigning individuals to populations according to their genotypes. Based on ΔK distribution, an inferred K value of 4 provided the best fit to the data and thus all individuals could be grouped in 4 populations (Fig. S5). Interestingly, KNP and SNU were well-defined populations having fewer admixed genotypes. Colombian populations could be separated into two groups. Creole pigs were admixed with SNU, agreeing with dendrogram and MDS ordination plots. Remanso and Manta genotypes were admixed with KNP, agreeing with first MSD coordinate (Fig. 5).

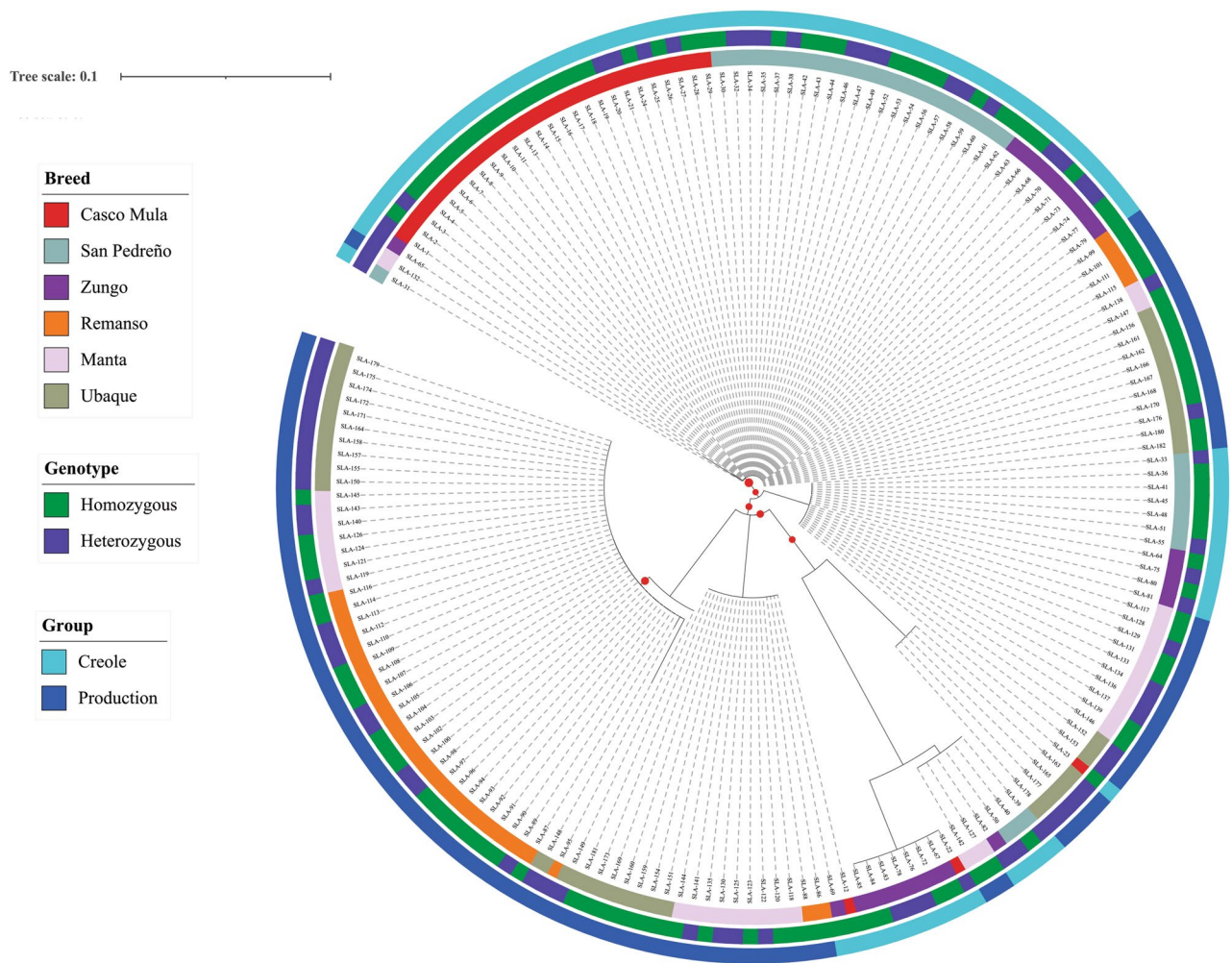


Figure 3. Homozygosity and heterozygosity distribution according to the populations analysed here.

| Parameter | Creole | | | Commercial line | | |
|-----------|------------|-------------|---------|-----------------|---------|---------|
| | Casco Mula | San Pedroño | Zungo | Remanso | Manta | Ubaque |
| n | 60 | 70 | 52 | 62 | 62 | 70 |
| π | 0.04018 | 0.04262 | 0.09214 | 0.03761 | 0.06942 | 0.05807 |
| Θ | 0.05847 | 0.04102 | 0.06767 | 0.05394 | 0.06404 | 0.04934 |
| S | 64 | 39 | 68 | 47 | 66 | 54 |
| h | 6 | 9 | 8 | 8 | 7 | 8 |
| Hd | 0.491 | 0.593 | 0.753 | 0.567 | 0.797 | 0.788 |
| SD | 0.078 | 0.042 | 0.038 | 0.070 | 0.030 | 0.031 |

Table 1. Genetic diversity indices for the breeds analysed in this study. n the amount of sequences, π nucleotide diversity, Θ theta (per site) from eta, S the amount of segregating sites, h the amount of haplotypes, Hd haplotype diversity, SD standard deviation.

Discussion

This study was aimed at comparing *SLA-DRB1* (Exon 2) genetic diversity for the first time in two Colombian pig populations. It was seen that the amount of alleles in each population; the Casco Mula breed had the least amount of alleles. Individuals from the San Pedroño breed had the most alleles and individuals from the Zungo breed the highest nucleotide variability (Table 1). Such data coincided with that reported by other groups^{26,38}, though differing from that reported by some authors^{8,26}, for example, twenty alleles have been identified in a population of 133 wild boars in Croatia⁸. Sequence diversity analysis showed that most polymorphisms in the $\beta 1$ domain were non-synonymous, Colombian pig populations having lower values in this region. However, synonymous and non-synonymous polymorphism distribution in the PBR was more uniform, despite

| Population | N | N _a | R _s | h _o | h _e | F _{IS} (S.E.)—p-value |
|-------------|-----|----------------|----------------|----------------|----------------|--------------------------------|
| Remanso | 31 | 8 | 5.816 | 0.290 | 0.558 | 0.4842 (0.0011)—0.0011 |
| Manta | 31 | 7 | 5.478 | 0.258 | 0.563 | 0.5463 (0.0000)—< 0.0001 |
| Ubaque | 35 | 8 | 5.478 | 0.514 | 0.731 | 0.2994 (0.0034)—0.0101 |
| Casco Mula | 30 | 6 | 4.542 | 0.267 | 0.380 | 0.3401 (0.0033)—0.0203 |
| San Pedroño | 35 | 9 | 7.128 | 0.457 | 0.728 | 0.3885 (0.0000)—< 0.0001 |
| Zungo | 26 | 8 | 6.566 | 0.538 | 0.743 | 0.3739 (0.0064)—0.0238 |
| Wild boar | 125 | 19 | 10.646 | 0.720 | 0.885 | 0.1879 (0.0000)—< 0.0001 |
| Micromini | 14 | 8 | 8.000 | 0.785 | 0.867 | 0.0978 (0.0150)—0.6968 |
| SNU | 114 | 4 | 3.993 | 0.657 | 0.715 | 0.0813 (0.0009)—0.0046 |
| Yorkshire | 31 | 12 | 9.792 | 0.741 | 0.878 | 0.1575 (0.0018)—0.0046 |
| Berkshire | 31 | 8 | 6.614 | 0.677 | 0.760 | 0.1108 (0.0120)—0.1202 |
| KNP | 105 | 6 | 4.939 | 0.771 | 0.744 | - 0.0368 (0.0145)—0.4926 |
| Duroc | 32 | 6 | 4.774 | 0.281 | 0.489 | 0.0986 (0.0152)—0.0018 |
| Landrace | 30 | 9 | 8.141 | 0.766 | 0.849 | 0.0986 (0.0138)—0.1454 |
| Landrace_C | 22 | 5 | 4.869 | 0.681 | 0.734 | 0.0735 (0.0056)—0.1069 |
| Yorkshire_C | 15 | 6 | 5.867 | 0.866 | 0.721 | - 0.2093 (0.0133)—0.3668 |
| Pietrain | 26 | 10 | 8.712 | 1.00 | 0.862 | - 0.1638 (0.0205)—0.3823 |

Table 2. Sample size (N), amount of alleles (N_a), allele richness (R_s), observed heterozygosity (h_o), expected heterozygosity (h_e) and F_{IS} values for analysed pig populations. *p* < 0.05.

| Population | β1 domain | | Peptide binding region | |
|-------------|---------------------|---------------------|------------------------|---------------------|
| | N _d (SE) | S _d (SE) | N _d (SE) | S _d (SE) |
| Remanso | 9.07 (1.94) | 2.32 (0.75) | 6.19 (1.38) | 0.70 (0.34) |
| Manta | 8.43 (1.60) | 2.55 (0.58) | 5.85 (1.13) | 0.89 (0.38) |
| Ubaque | 10.13 (2.29) | 2.72 (0.94) | 8.26 (1.77) | 1.30 (0.66) |
| Casco Mula | 7.80 (1.43) | 2.61 (0.65) | 5.17 (0.92) | 0.82 (0.37) |
| San Pedroño | 9.12 (2.06) | 3.47 (0.86) | 7.20 (1.55) | 1.31 (0.56) |
| Zungo | 15.95 (2.97) | 8.07 (1.64) | 10.52 (1.89) | 2.49 (0.86) |
| Wild boar | 17.34 (2.69) | 7.63 (1.51) | 10.44 (1.92) | 2.52 (0.75) |
| Micromini | 18.71 (3.00) | 8.65 (1.59) | 11.01 (1.69) | 2.79 (0.90) |
| SNU | 12.60 (2.58) | 6.74 (1.50) | 8.28 (1.81) | 2.05 (0.79) |
| Yorkshire | 17.98 (2.85) | 6.80 (1.22) | 11.09 (1.93) | 1.81 (0.56) |
| Berkshire | 13.91 (2.45) | 3.87 (1.00) | 9.48 (1.63) | 1.12 (0.51) |
| KNP | 19.02 (2.88) | 7.67 (1.57) | 11.11 (1.79) | 1.90 (0.70) |
| Duroc | 10.06 (1.73) | 4.92 (0.91) | 7.00 (1.12) | 1.56 (0.44) |
| Landrace | 17.54 (2.82) | 5.09 (1.15) | 10.56 (1.81) | 1.24 (0.48) |
| Landrace_C | 16.96 (2.73) | 5.70 (1.31) | 9.15 (1.70) | 1.88 (0.60) |
| Yorkshire_C | 12.10 (2.24) | 4.58 (1.08) | 7.59 (1.59) | 1.18 (0.50) |
| Pietrain | 18.24 (2.98) | 7.26 (1.48) | 10.61 (1.92) | 1.95 (0.73) |

Table 3. The amount of nonsynonymous (N_d) and synonymous (S_d) polymorphisms in the sequence encoding the whole β1 domain and peptide binding region (PBR) for the swine populations analysed here. SE standard error.

the Colombian populations having less diversity. An excess of non-synonymous polymorphisms was evident in the PBR; this was to be expected due to this region's role in antigen presentation^{11,12,26}.

This study's results confirmed greater diversity in wild boars, contrasting with low diversity in domestic pig populations. The clearest differences were observed regarding allele richness which was especially low in some populations (R_s < 5 in the Casco Mula, KNP, Duroc and Landrace_C populations). However, some populations having low R_s had high h_e (corrected for sample size), indicating a marked variation regarding allele frequency distribution.

Considering that genetic drift and the structure of animal crosses are factors affecting allele frequency distribution (having greater intensity in populations having smaller effective population size)^{39–43}, then variations regarding expected heterozygosity and allele richness could have been due to the combined effects of small effective population sizes and varying degrees of endogamy. The latter factor would have mainly been due to the

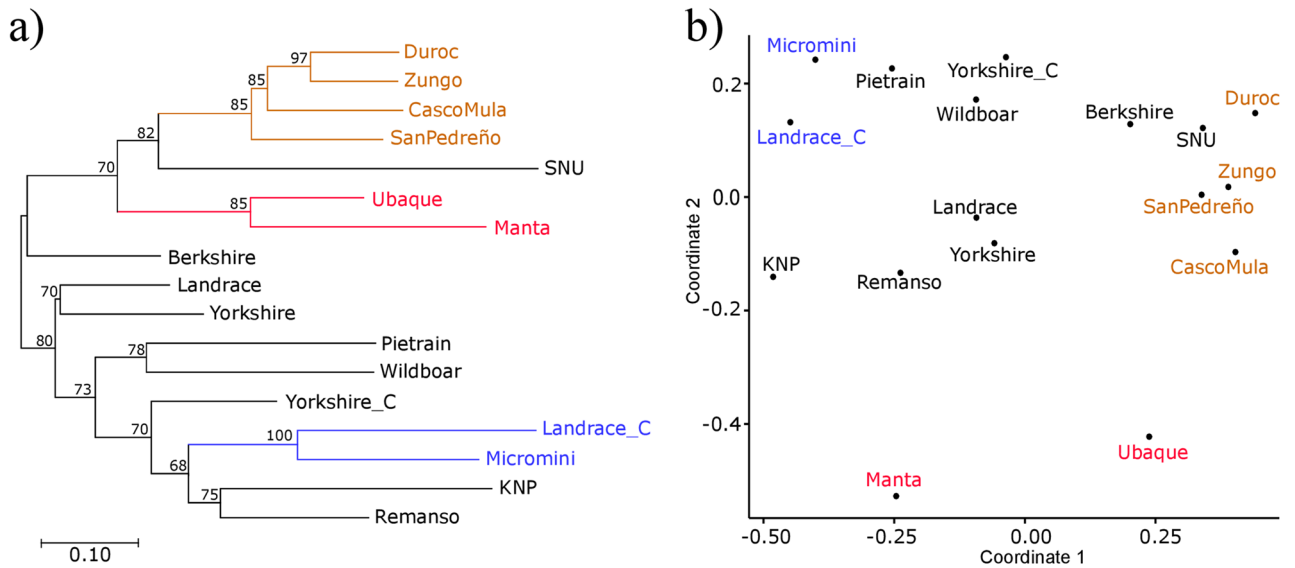


Figure 4. Relationships between pig populations. BIONJ dendrogram (a) and multidimensional scaling analysis (b) based on DA genetic distances calculated from *SLA-DRB1* allele frequency. Bootstrap values are shown for each node in the dendrogram and clusters identified by both methods are depicted in the same colour.

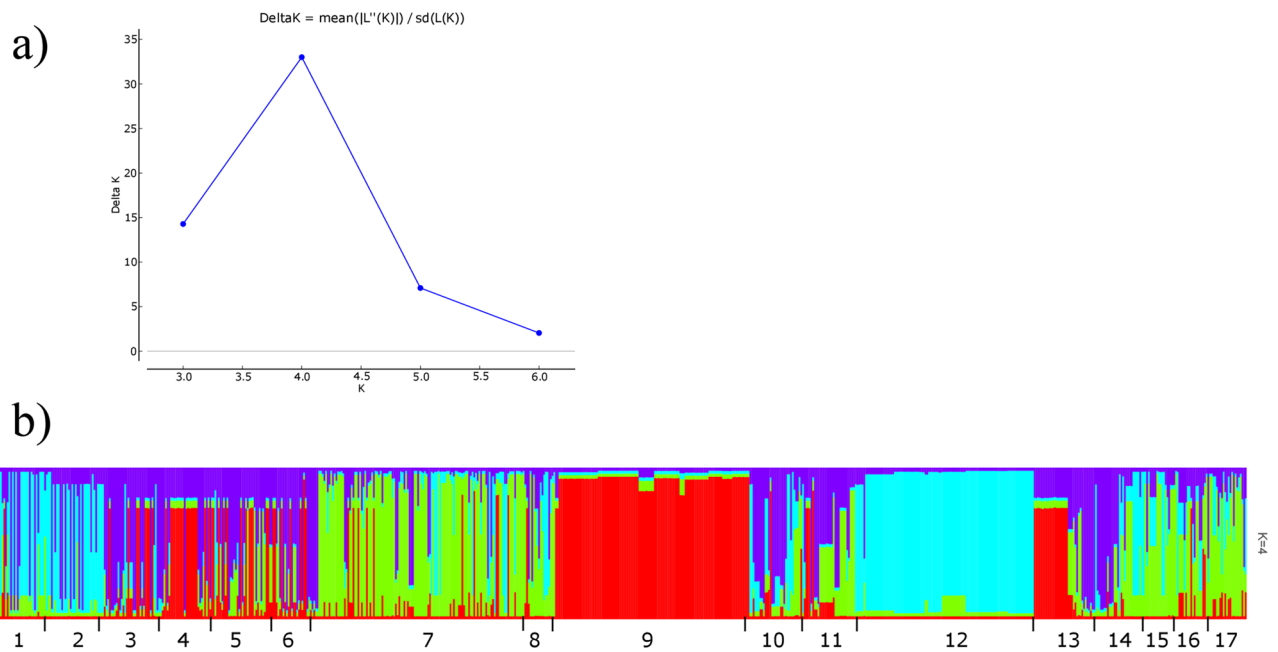


Figure 5. Pig population structure analysis based on Bayesian clustering. Delta K values as a function of K (a). Population clustering (b). Each vertical line represents an individual and each colour an inferred population. Numbers under (b) indicate populations: 1, Remanso; 2, Manta; 3, Ubaque; 4, Casco Mula; 5, San Pedroño; 6, Zungo; 7, Wild boar; 8, Micromini; 9, SNU; 10, Yorkshire; 11, Berkshire; 12, KNP; 13, Duroc; 14, Landrace; 15, Landrace_C; 16, Yorkshire_C; 17, Pietrain.

artificial selection of production features. For example, selection based on the Casco Mula breed’s characteristic feature (hoof) and its small effective population size³⁵ could lead to intensive in-breeding which would then produce differences regarding allele frequency and could eventually lead to a loss of some alleles. This idea could also be supported by the high F_{IS} values for some of these populations, indicating higher homozygote percentages than those to be expected in a population in equilibrium and could suggest higher endogamy levels.

Alleles *DRB1**01:02 and *10:01 were previously considered ancestral due to their presence in populations having different origins²⁶. It was seen in this study that alleles *DRB1**01:01, *01:02, *02:01:01, *09:01:01 and *10:01:01 were broadly distributed, often having high frequency in various populations. As the pigs’ domestication seems

to have been occurred during six independent events⁴⁴, these alleles' distribution could either indicate their presence in the founding populations in only some domestication events or their presence in the founding populations in all domestication events, accompanied by the subsequent loss of some alleles due to drift and/or natural selection. Allele *DRB1*02:01:01* may thus have been present in pig populations originating the creole and commercial line populations but which could have disappeared due to genetic drift (or its frequency could have become reduced and not been sampled) in the Manta population. In fact, several studies have established that mutationally robust populations might have less mutational load; however, it is not clear whether small populations have less susceptibility to drift-related fitness reduction⁴⁵.

The creole pig group's results particularly seem to have coincided with previously reported data regarding individuals from the San Pedreño breed³⁵, confirmed by analysing pedigree data from information collected from 2008 to 2017 for estimating the generation interval, the level of pedigree integrity, consanguinity and evolution. It was particularly interesting that this group had 6.73% consanguinity³⁵. There have been no reports concerning this for Casco Mula pigs to date.

Considering the group's experience⁴⁶, it was decided to use bioinformatics methods, involving Haplofinder software, for assigning alleles in heterozygous individuals; this tool has been used regarding bovine species for *DR*^{47,48} and *DQA*⁴⁹, thereby facilitating assigning a large amount of individuals, resorting to cloning in just 16% of the population. The data obtained from analysing variability between populations revealed a low level of heterozygotes, this being more evident in individuals from the Casco Mula breed and those from the Manta and Remanso groups.

The recombination events observed in this study coincided with that reported by other authors^{8,50} and have been considered a molecular mechanism associated with MHC genetic diversity; such phenomenon has been observed in ungulate⁵¹ and passerine species⁵².

Studies in Poland regarding commercial line pigs and native individuals have analysed runs of homozygosity (ROH), identifying genome regions having high ROH levels, for example, SSC4 (51.9–55.9 Mb) which is related to various functions, including immunological ones⁵³. Detecting high levels of homozygosity in various regions of the genome of inbred Babraham pigs led to developing specific MHC-related analysis, revealing a high level of homozygosity, thereby making it a candidate for research group use as a biomodel⁵⁴.

The pig populations had a clear genetic structure based on the *SLA-DRB1* locus. Interestingly, the low genetic differentiation in Colombian creole populations contrasted sharply with the great differentiation between commercial line populations. This could have arisen from crossing individuals for obtaining such lines. Commercial lines' impact on Landrace pigs' genetic diversity has been evaluated recently; the current breed provides evidence of the old lines' formation, but that such fusion has reduced Landrace genetic diversity⁵⁵. Various clusters were identified based on a small genetic distance and/or high bootstrap support values. This indicated that the populations forming these clusters have high genetic identity and detailed analysis has revealed that they share a large percentage of their mean absolute allele frequency and the alleles so involved. Such information could be used in infectious disease control programmes or when designing vaccines^{56–58}. The *SLA-DRB1*02:01:01* allele has 79% amino acid identity with *HLA-DRB1*01:01* in pigs and has been used for making a bioinformatics predictor based on pocket profile methodology²⁸, highlighting the need for characterising SLA diversity when designing vaccination and control programmes.

Materials and methods

Animals and DNA isolation. The study involved sampling 188 animals from six populations, thereby identifying a first group called the creole pig group consisting of individuals from the Zungo (n = 26), Casco de Mula (n = 30) and San Pedreño populations (n = 35). All blood samples from this group were provided by AGROSAVIA's research centres. A second group called the commercial line consisted of Super Mom 52 and PIC commercial hybrid populations; blood samples were taken on three pig farms in the Cundinamarca department, identified as follows: Remanso U.D.C.A (n = 31), Manta (n = 31) and Ubaque (n = 35) (Fig. 1). All the pigs were randomly sampled; variables such as age, gender, production stage or relationship were not considered.

Venous blood was obtained by jugular vein puncture and kept in tubes with anticoagulant (EDTA). Genomic material was extracted following Wizard Genomic DNA Purification Kit recommendations. All samples' DNA integrity was determined on 1.5% agarose gel; a μ Drop microplate reader was used for determining DNA concentration. The samples were then stored at -20°C until PCR. The sample collection procedure was approved by the Universidad de Ciencias Aplicadas y Ambientales' (U.D.C.A) bioethics committee (in minutes No.002/2019). All procedures followed the guidelines and regulations established in the Colombian code of bioethics (resolution 8430 of 1993) and Law 84/1989 regarding the protection of animals.

Genotypical data regarding the populations reported in the pertinent literature was included for carrying out the comparative analysis of diversity between Colombian and international populations, such as Wild boar⁸, Microminipig³⁹, SNU (Seoul National University), Yorkshire, Berkshire, Korean Native Pigs (KNP), Duroc and Landrace¹⁴, Landrace_C and Yorkshire_C²⁶ and Pietrain pigs⁶.

SLA-DRB1 genotyping. Polymerase chain reaction (PCR) was used for obtaining the *SLA-DRB1* exon 2 sequences, using a pair of primers reported in the literature: forward 5'-GTCCACGCAGCGCATTCTT-3'³⁸ and reverse 3'-ACACACTCTGCCCCCG-5'²⁶. Amplified fragment length was 328 bp; 2X high-fidelity Kodaq PCR MasterMix was used and reaction volume was 25 μL , distributed as follows: 7 μL nuclease-free water, 12.5 μL enzyme Master Mix, 1.25 μL for each primer [10 μM] and 3 μL genomic DNA (concentration ranged from 50–100 $\mu\text{g}/\mu\text{L}$). Amplification conditions were as follows: initial denaturing at 94°C for 3 min, 35 cycles consisting of denaturing at 94°C for 30 s., annealing at 68.2°C for 30 s., extension at 72°C for 1 min, followed

by a final extension step at 72° for 5 min. All PCR products were amplified in 2 independent reactions, revealed on 2% agarose gel and then purified and sent to Macrogen Inc. (South Korea) to be sequenced in both senses.

CLC Main Workbench v.3.6.5 software (QIAGEN, Aarhus, Denmark; <https://digitalinsights.qiagen.com>) was used for assembling and analysing the sequences. Heterozygotic individuals' sequences were manually edited, IUPAC/IUBMB single-letter code was used for ambiguous pairs⁶⁰ and Haplofinder Python script was used for assigning them^{47–49}. This involved creating a library with the exon 2 sequences from the alleles reported for *SLA-DRB1* in IPD-MHC (www.ebi.ac.uk/ipd/mhc). pGEM-T Easy vector (Promega) was used for ligating heterozygotic samples which could not be assigned by bioinformatics methods which were then transformed in competent *E. coli* JM109 cells to separate both alleles, following the protocol reported by FIDIC⁶¹. Selection involved blue/white colony screening; the amount of positive colonies was replicated and scaled for plasmid purification; 5–10 positive clones were obtained per sample, confirmed on 1.5% agarose gel and sent to Macrogen again to be sequenced in both senses. The complete genotyping results are shown in Table S6.

Genetic diversity indices, recombination events and phylogenetic analysis. The MUSCLE sequence alignment tool (<https://www.ebi.ac.uk/Tools/msa/muscle/>) was used for aligning the sequences; they were then evaluated for identifying indels (insertion and elimination events) which could have affected the length of any sequence to be analysed. Minor adjustments were then made to the alignments to eliminate gaps. DnaSP (v5) software (<http://www.ub.edu/dnasp/>) was used for evaluating each population's genetic diversity, nucleotide diversity (π), theta (per site) from eta (Θ), the amount of segregating sites (S), the amount of haplotypes (h) and haplotype diversity (H_d) along with their standard deviations (SD).

Maximum likelihood trees with double precision, considering Jukes-Cantor as a substitution model were built based on the alignments using FastTree Version 2.1.9 (<http://www.microbesonline.org/fasttree/>) for phylogenetic reconstruction; node robustness was evaluated by bootstrap method with 1000 replicates⁶². The phylogenetic trees were visualised in the web tool Interactive Tree Of Life V3 (<http://itol.embl.de>)⁶³; the Recombination Detection Program (version 4) (RDP4) was also used. In-depth statistical analysis included using RDP, GENECONV (for detecting gene conversion), BootScan (screening nucleotide sequence alignments for evidence of recombination), Maximum Chi-Squared test (MaxiChi) (for identifying potential recombination events between two sequences or between two sequences and a putative derived sequence), Chimaera (for identifying recombinant and parental sequences and estimate breakpoint positions), Sister Scanning (SiScan) (for assessing phylogenetic and compositional signals in various patterns of identity), 3Seq (for ascertaining whether any sequence in a data set is a recombinant or mosaic), VisRD (for visual recombination detection in a sequence alignment) and BURT (for testing significance in factor analysis/analysis of variance) methods⁶⁴.

Measures of genetic diversity and Hardy–Weinberg equilibrium. The maximum likelihood method was used for obtaining the amount of alleles (N_a) and allele frequencies by direct counting, along with their standard errors⁶⁵. FSTAT software was used for calculating allele richness⁶⁶. Observed heterozygosity (h_o) and unbiased expected heterozygosity (h_e) according to Nei⁶⁷ were estimated using Arlequin v.3.5 software for population genetic analysis⁶⁸. Potential Hardy–Weinberg variation percentages were estimated by F_{IS} statistic test⁶⁹ included in Genepop v.4.7.0, using the exact test of significance⁷⁰. Molecular Evolutionary Genetics Analysis (MEGA X)⁷¹ software was used for calculating the average amount of nonsynonymous (N_d) and synonymous (S_d) polymorphisms and their standard errors (by bootstrapping with 10,000 replicates) for all sequence pairs within populations. N_d and S_d were not normalised for sequence length to attain the percentage of polymorphisms in the PBR regarding the whole $\beta 1$ domain encoding sequence.

Population structure and genetic differentiation. Arlequin software v.3.5.2 analysis of molecular variance (AMOVA)^{68,72} was used for assessing variation amongst populations. POPTREE2 software⁷³ was used for estimating D_A genetic distances between populations from allele frequencies⁷⁴. R packages *ape*⁷⁵ and *poppr*⁷⁶ were used to estimate Nei standard genetic distances between populations from amino acid sequences⁷⁷. R packages *ape*, *poppr* and *stats* were used for inferring dendrograms with NJ and BIONJ algorithms. Confidence values for nodes in the dendrogram were obtained by bootstrapping alleles with 1,000 replicates and the trees were constructed using webserver Phylogeny.fr⁷⁸. Multidimensional Scaling (MDS) analysis based on D_A genetic distances involved using the R *cmdscale* function. R squared was used for evaluating the representation of the trees' original distances and MDS and correlation analysis was used for evaluating sample size effects on h_e and h_o . The Bayesian iterative algorithm in STRUCTURE multilocus genotype data for assigning individuals to a particular population⁷⁹ was used for further evaluation of population structure, inferring genetic clusters in which individuals could be assigned. Structure harvester⁸⁰ a programme for parsing Pritchard's STRUCTURE output and performing the Evanno method⁸¹ was used for determining the most likely K by ΔK distribution. Ten replicates were run for each K using the admixture model with correlated allele frequencies and Markov Chain Monte Carlo 50,000 burn-in steps, followed by 100,000 iterations.

Received: 1 September 2020; Accepted: 18 January 2021

Published online: 22 February 2021

References

1. Rothkötter, H. J. Anatomical particularities of the porcine immune system: A physician's view. *Dev. Comp. Immunol.* **33**, 5 (2009).

2. Dawson, H. A comparative assessment of the pig, mouse and human genomes. *In book: The Minipig in Biomedical Research*, 19. <https://doi.org/10.1201/b11356-28> (2011).
3. Dawson, H. D. *et al.* Structural and functional annotation of the porcine immunome. *BMC Genomics* **14**, 332. <https://doi.org/10.1186/1471-2164-14-332> (2013).
4. Hammer, S. E. *et al.* Importance of the Major Histocompatibility Complex (Swine Leukocyte Antigen) in Swine Health and Biomedical Research. *Annu. Rev. Anim. Biosci.* **8**, 171–198. <https://doi.org/10.1146/annurev-animal-020518-115014> (2020).
5. Ladowski, J. M. *et al.* Swine Leukocyte Antigen Class II Is a Xenoantigen. *Transplantation* **102**, 249–254. <https://doi.org/10.1097/TP.0000000000001924> (2018).
6. Essler, S. E. *et al.* Molecular characterization of swine leukocyte antigen gene diversity in purebred Pietrain pigs. *Anim. Genet.* **44**, 202–205. <https://doi.org/10.1111/j.1365-2052.2012.02375.x> (2013).
7. Ujvari, B. & Belov, K. Major Histocompatibility Complex (MHC) markers in conservation biology. *Int. J. Mol. Sci.* **12**, 5168–5186. <https://doi.org/10.3390/ijms12085168> (2011).
8. Arbanasic, H. *et al.* Evolution of MHC class II SLA-DRB1 locus in the Croatian wild boar (*Sus scrofa*) implies duplication and weak signals of positive selection. *Anim. Genet.* **50**, 33–41. <https://doi.org/10.1111/age.12734> (2019).
9. Ho, C. S. *et al.* Molecular characterization of swine leukocyte antigen class I genes in outbred pig populations. *Anim. Genet.* **40**, 468–478. <https://doi.org/10.1111/j.1365-2052.2009.01860.x> (2009).
10. Lunney, J. K., Ho, C. S., Wysocki, M. & Smith, D. M. Molecular genetics of the swine major histocompatibility complex, the SLA complex. *Dev. Comp. Immunol.* **33**, 362–374. <https://doi.org/10.1016/j.dci.2008.07.002> (2009).
11. Moutou, K. A. *et al.* Domestication does not narrow MHC diversity in *Sus scrofa*. *Immunogenetics* **65**, 195–209 (2013).
12. Barbisan, F., Savio, C., Bertorello, G., Patarnello, T. & Congiu, L. Duplication polymorphism at MHC class II DRB1 locus in the wild boar (*Sus scrofa*). *Immunogenetics* **61**, 145–151. <https://doi.org/10.1007/s00251-008-0339-6> (2009).
13. Gao, C. *et al.* Swine leukocyte antigen diversity in Canadian specific pathogen-free Yorkshire and Landrace Pigs. *Front. Immunol.* **8**, 282. <https://doi.org/10.3389/fimmu.2017.00282> (2017).
14. Thong, L. M. *et al.* Systematic analysis of swine leukocyte antigen-DRB1 nucleotide polymorphisms using genomic DNA-based high-resolution genotyping and identification of new alleles. *Tissue Antigens* **77**, 572–583. <https://doi.org/10.1111/j.1399-0039.2011.01662.x> (2011).
15. Garde, C. *et al.* Improved peptide-MHC class II interaction prediction through integration of eluted ligand and peptide affinity data. *Immunogenetics* **71**, 445–454. <https://doi.org/10.1007/s00251-019-01122-z> (2019).
16. Unanue, E. R., Turk, V. & Neefjes, J. Variations in MHC Class II antigen processing and presentation in health and disease. *Annu. Rev. Immunol.* **34**, 265–297. <https://doi.org/10.1146/annurev-immunol-041015-055420> (2016).
17. Schneidman-Duhovny, D. *et al.* Predicting CD4 T-cell epitopes based on antigen cleavage, MHCII presentation, and TCR recognition. *PLoS ONE* **13**, e0206654. <https://doi.org/10.1371/journal.pone.0206654> (2018).
18. Van Chanh Le, Q. *et al.* Analysis of peptide-SLA binding by establishing immortalized porcine alveolar macrophage cells with different SLA class II haplotypes. *Vet. Res.* **49**, 96. <https://doi.org/10.1186/s13567-018-0590-2> (2018).
19. Wang, Y. *et al.* Cloning and bioinformatics analysis of SLA-DR genes in Hunan Daweizi pigs. *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi* **25**, 770–773 (2009).
20. Imaeda, N. *et al.* Influence of swine leukocyte antigen haplotype on serum antibody titers against swine erysipelas vaccine and reproductive and meat production traits of SLA-defined selectively bred Duroc pigs. *J. Vet. Med. Sci.* **80**, 1662–1668. <https://doi.org/10.1292/jvms.18-0027> (2018).
21. Ando, A. *et al.* Genetic Association between Swine Leukocyte Antigen Class II Haplotypes and Reproduction Traits in Microminipigs. *Cells* **8**. <https://doi.org/10.3390/cells8080783> (2019).
22. Gao, C. *et al.* Molecular genetic characterization and haplotype diversity of swine leukocyte antigen in Chinese Rongshui miniature pigs. *Mol. Immunol.* **112**, 215–222. <https://doi.org/10.1016/j.molimm.2019.05.009> (2019).
23. Hundrieser, J. *et al.* Role of human and porcine MHC DRB1 alleles in determining the intensity of individual human anti-pig T-cell responses. *Xenotransplantation* **26**, e12523. <https://doi.org/10.1111/xen.12523> (2019).
24. Wolf, E., Kemter, E., Klymiuk, N. & Reichart, B. Genetically modified pigs as donors of cells, tissues, and organs for xenotransplantation. *Anim. Front.* **9**, 13–20. <https://doi.org/10.1093/af/vfz014> (2019).
25. Choi, N. R. *et al.* Analysis of Swine leukocyte antigen haplotypes in yucatan miniature pigs used as biomedical model animal. *Asian-Australas J. Anim. Sci.* **29**, 321–326. <https://doi.org/10.5713/ajas.15.0331> (2016).
26. Luetkemeier, E. S., Malhi, R. S., Beaver, J. E. & Schook, L. B. Diversification of porcine MHC class II genes: evidence for selective advantage. *Immunogenetics* **61**, 119–129. <https://doi.org/10.1007/s00251-008-0348-5> (2009).
27. Le, M. *et al.* Development of a simultaneous high resolution typing method for three SLA class II genes, SLA-DQA, SLA-DQB1, and SLA-DRB1 and the analysis of SLA class II haplotypes. *Gene* **564**, 228–232. <https://doi.org/10.1016/j.gene.2015.03.049> (2015).
28. Gutierrez, A. H. *et al.* Development and validation of an epitope prediction tool for swine (PigMatrix) based on the pocket profile method. *BMC Bioinformatics* **16**, 290. <https://doi.org/10.1186/s12859-015-0724-8> (2015).
29. Gutierrez, A. H. *et al.* In vivo validation of predicted and conserved T cell epitopes in a Swine Influenza model. *PLoS ONE* **11**, e0159237. <https://doi.org/10.1371/journal.pone.0159237> (2016).
30. Benitez-Ortiz, W. & Sánchez, M. D. Los cerdos locales en los sistemas tradicionales de producción. *E. Estudio FAO producción y sanidad animal* **148**, 191 (2001).
31. Espinosa, C. & Ly, J. Cerdos criollos colombianos y agricultura sostenible. *Revista Computarizada de Producción Porcina* **22**, 9 (2015).
32. Oslinger, A. *et al.* Characterization of Colombian creole pigs by RAMs. *Acta Agron.* **55**, 7 (2006).
33. Meléndez, I., Pardo, E. & Cavadia, T. Genetic characterization of the domestic pig (*Sus scrofa domestica*) in Cerete-Colombia, using microsatellite markers. *Rev. MVZ Córdoba* **19**, 7 (2014).
34. Barrera, G. P. *et al.* Cerdos criollos colombianos: caracterización racial, productiva y genética. *Repository Agrosavia* (2007).
35. Ocampo-Gallego, R. J., Tobón-Castaño, J. A., Martínez-Oquendo, P. Y., Ramírez-Toro, E. J. & Lucero-Casanova, C. E. Analysis of genetic diversity in san pedreño indigenous swine using pedigree data. *Ecosistema y Recursos Agropecuarios* **6**, 8. <https://doi.org/10.19136/era.a6n17.2049> (2019).
36. Pardo, C. A. *Evaluación de la calidad de la carne y caracterización de genes asociados a la calidad de tres razas de cerdos criollos colombianos* (Tesis Maestría, Facultad de Ciencias Agarias, 2016).
37. Diaz, C. A. *et al.* Characterization of pig farms in the main swine producing regions of Colombia. *Revista Colombiana de Ciencias Pecuarias* **24**, 13 (2011).
38. Ando, A. *et al.* SLA-DRB1 and -DQB1 genotyping by the PCR-SSOP-Luminex method. *Tissue Antigens* **78**, 49–55. <https://doi.org/10.1111/j.1399-0039.2011.01669.x> (2011).
39. Husemann, M., Zachos, F. E., Paxton, R. J. & Habel, J. C. Effective population size in ecology and evolution. *Heredity (Edinb)* **117**, 191–192. <https://doi.org/10.1038/hdy.2016.75> (2016).
40. Nei, M. & Tajima, F. Genetic drift and estimation of effective population size. *Genetics* **98**, 625–640 (1981).
41. Tajima, F. & Nei, M. Note on genetic drift and estimation of effective population size. *Genetics* **106**, 569–574 (1984).
42. Blackburn, H. D., Plante, Y., Rohrer, G., Welch, E. W. & Paiva, S. R. Impact of genetic drift on access and benefit sharing under the Nagoya Protocol: the case of the Meishan pig. *J. Anim. Sci.* **92**, 1405–1411. <https://doi.org/10.2527/jas.2013-7274> (2014).
43. Akashi, H., Osada, N. & Ohta, T. Weak selection and protein evolution. *Genetics* **192**, 15–31 (2012).

44. Larson, G. *et al.* Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**, 1618. <https://doi.org/10.1126/science.1106927> (2005).
45. LaBar, T. & Adami, C. Evolution of drift robustness in small populations. *Nat. Commun.* **8**, 1012. <https://doi.org/10.1038/s41467-017-01003-7> (2017).
46. Bohorquez, M. D. *et al.* Major histocompatibility complex class II (DRB3) genetic diversity in Spanish Morucha and Colombian Normande cattle compared to Taurine and Zebu populations. *Front. Genet.* **10**, 1293. <https://doi.org/10.3389/fgene.2019.01293> (2019).
47. Baxter, R., Hastings, N., Law, A. & Glass, E. J. A rapid and robust sequence-based genotyping method for BoLA-DRB3 alleles in large numbers of heterozygous cattle. *Anim. Genet.* **39**, 561–563 (2008).
48. Miltiadou, D., Law, A. S. & Russell, G. C. Establishment of a sequence-based typing system for BoLA-DRB3 exon 2. *Tissue Antigens* **62**, 55–65 (2003).
49. Takeshima, S. N., Miki, A., Kado, M. & Aida, Y. Establishment of sequence-based typing system for BoLA-DQA1 exon 2. *J. Comp. Gen.* **69**, 10 (2007).
50. Chardon, P., Renard, C., Gaillard, C. R. & Vaiman, M. The porcine major histocompatibility complex and related paralogous regions: A review. *Genet. Sel. Evol.* **32**, 109–128. <https://doi.org/10.1186/1297-9686-32-2-109> (2000).
51. Schaschl, H., Wandeler, P., Suchentrunk, F., Obexer-Ruff, G. & Goodman, S. J. Selection and recombination drive the evolution of MHC class II DRB diversity in ungulates. *Heredity (Edinb)* **97**, 427–437. <https://doi.org/10.1038/sj.hdy.6800892> (2006).
52. Balasubramaniam, S., Mulder, R. A., Sunnucks, P., Pavlova, A. & Melville, J. MHC class II beta exon 2 variation in pardalotes (Pardalotidae) is shaped by selection, recombination and gene conversion. *Immunogenetics* **69**, 101–111. <https://doi.org/10.1007/s00251-016-0953-7> (2017).
53. Szmatała, T. *et al.* Detection of runs of homozygosity in conserved and commercial pig breeds in Poland. *J. Anim. Breed. Genet.* <https://doi.org/10.1111/jbg.12482> (2020).
54. Schwartz, J. C. *et al.* The major histocompatibility complex homozygous inbred Babraham pig as a resource for veterinary and translational medicine. *HLA* <https://doi.org/10.1111/tan.13281> (2018).
55. Hulsegge, I. *et al.* Impact of merging commercial breeding lines on the genetic diversity of Landrace pigs. *Genet. Sel. Evol.* **51**, 60. <https://doi.org/10.1186/s12711-019-0502-6> (2019).
56. Kapetanovic, R., Fairbairn, L., Sester, D. P., Archibald, A. L. & Hume, D. A. Gene expression analysis of pig macrophages reveals similarities to humans. *Research Gate.net*, 1 (2012).
57. Baxter, R. *et al.* BoLA-DR peptide binding pockets are fundamental for foot-and-mouth disease virus vaccine design in cattle. *Vaccine* **28**, 28–37 (2009).
58. Chakraborty, S., Rahman, T. & Chakravorty, R. Characterization of the protective HIV-1 CTL epitopes and the corresponding HLA class I alleles: A step towards designing CTL based HIV-1 vaccine. *Adv. Virol.* **2014**, 321974. <https://doi.org/10.1155/2014/321974> (2014).
59. Gao, C. *et al.* Characterization of swine leukocyte antigen (SLA) polymorphism by sequence-based and PCR-SSP methods in Chinese Bama miniature pigs. *Dev. Comp. Immunol.* **45**, 87–96. <https://doi.org/10.1016/j.dci.2014.02.006> (2014).
60. Leonard, S. A. IUPAC/IUB Single-letter codes within nucleic acid and amino acid sequences. *Curr. Protocols Bioinf.* **00**(1), 1. <https://doi.org/10.1002/0471250953.bia01as00> (2002).
61. Lopez, C., Suarez, C. F., Cadavid, L. F., Patarroyo, M. E. & Patarroyo, M. A. Characterising a microsatellite for DRB typing in *Aotus vociferans* and *Aotus nancymae* (Platyrrhini). *PLoS ONE* **9**, e96973. <https://doi.org/10.1371/journal.pone.0096973> (2014).
62. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650. <https://doi.org/10.1093/molbev/msp077> (2009).
63. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–245. <https://doi.org/10.1093/nar/gkw290> (2016).
64. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003. <https://doi.org/10.1093/ve/vev003> (2015).
65. Li, C. *Fisrt course in population genetics.* Boxwood Press (1976).
66. Goudet, J. FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *J. Heredity* **86**, 2 (1995).
67. Nei, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 7 (1978).
68. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analysis in Linux and Windows. *Mol. Ecol. Resour.* **10**, 3 (2010).
69. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 12 (1984).
70. Rousset, F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 3 (2008).
71. Kumar, S., Stecher, G., Li, M., Nnyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 3 (2018).
72. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
73. Takezaki, N., Nei, M. & Tamura, K. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Mol. Biol. Evol.* **27**, 747–752. <https://doi.org/10.1093/molbev/msp312> (2010).
74. Nei, M., Tajima, F. & Tateno, Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol.* **19**, 13 (1983).
75. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 2 (2004).
76. Kavran, J. M. & Leahy, D. J. Coupling antibody to cyanogen bromide-activated sepharose. *Methods Enzymol.* **541**, 27–34. <https://doi.org/10.1016/B978-0-12-420119-4.00003-3> (2014).
77. Nei, M. Genetic distance between populations. *Am. Nat.* **106**, 10 (1972).
78. Dereeper, A. *et al.* Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469. <https://doi.org/10.1093/nar/gkn180> (2008).
79. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
80. Earl, D. A. & vonHoldt, B. M. Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361. <https://doi.org/10.1007/s12686-011-9548-7> (2012).
81. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* **14**, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x> (2005).
82. d-maps.com. Colombia-departments. (2007–2020).
83. d-maps.com. Cundinamarca (Colombia). (2007–2020).
84. d-maps.com. Terms and conditions of use. (2007–2020).

Acknowledgements

This work was supported by the Universidad de Ciencias Aplicadas y Ambientales (U.D.C.A). The authors wish to express their thanks to AGROSAVIA and the pig farm owners for allowing us access to their biological resources/data. We also would like to thank Jason Garry for translating the manuscript. The Universidad Nacional de Colombia partially covered open access publication costs.

Author contributions

Conceptualisation: M.A.P. and C.T.C.G. Data curation: C.T.C.G., M.D.B. and M.C. Formal analysis: C.T.G.C., M.D.B., M.C. and C.F.S. Funding acquisition: M.A.P. Investigation methodology: C.T.G.C., C.E.L., B.H., A.M., A.C., K.R. and C.F.S. Supervision: M.A.P., R.M.R. and C.F.S. Visualisation: C.T.C.G., M.D.B. and M.C. Writing-original draft: C.T.C.G., M.D.B. and M.C. Writing-review editing: all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-83637-8>.

Correspondence and requests for materials should be addressed to M.A.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021