

500 years of breeding in the *Carthusian Strain* of Pura Raza Español horse: An evolutionary analysis using genealogical and genomic data

Julia Poyato-Bonilla¹  | Nora Laseca²  | Sebastián Demyda-Peyrás^{3,4}  |
Antonio Molina²  | Mercedes Valera¹ 

¹Dpto. Agronomía. Escuela Técnica Superior de Ingeniería Agronómica, Universidad de Sevilla, Sevilla, Spain

²Dpto. Genética. Facultad de Veterinaria, Universidad de Córdoba, Córdoba, Spain

³Departamento de Producción Animal, Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, Buenos Aires, Argentina

⁴CONICET, CCT La Plata, La Plata, Argentina

Correspondence

Julia Poyato-Bonilla, Dpto. Agronomía, Escuela Técnica Superior de Ingeniería Agronómica, Universidad de Sevilla, Sevilla, Spain.

Emails: julpoybon@alum.us.es; jpoybon@gmail.com

Funding information

This study was financed by the AGL-2017-84217-P Research project from the Ministerio de Economía, Industria y Competitividad of the Spanish Government. Nora Laseca is a doctoral fellow (PRE 2018-083492)

Abstract

The Carthusian horse is a Pura Raza Español (PRE) strain (CS), bred as a closed population since its creation more than 500 years ago. The aim of this study was to analyse for the first time its population structure and situation of variability combining both genealogical (GEL) and genomic (GEN) data. The GEL data comprised 348,429 pedigree records (56,105 CS horses), while the GEN analysis included the high-density genotypes (670,804 SNPs) of 287 horses. Pedigree completeness demonstrated its accuracy, showing a good correlation of GEL (F) and GEN (F_{ROH}) inbreeding coefficient in the case of PRE subpopulations partially related and non-related to Carthusian strain (0.68) but a lower value in the 100% Carthusian horses (0.42), due to the high weight of founders not detected by GEL analysis. GEN (PCA, AMOVA, and Admixture) and GEL analysis showed a good differentiation of subpopulations, but also a high level of introgression of the CS in the breed during past decades. A recent change in this trend was noteworthy, with a considerable reduction in CS variability and a genetic bottleneck (effective population sizes of 31.57 and 30.20 in GEL and GEN analysis, respectively, in last generation). The PRE has maintained its variability, and a considerable difference in estimated N_e by GEL (60.77) and GEN (188.0) data was observed. Using two sources of complementary information, it was found the existence of an ancient PRE strain with a unique genetic landmark, practically free from the influence of other equine populations.

KEYWORDS

bottleneck, effective number, equine, genetic diversity, population structure, SNP

1 | INTRODUCTION

With over 250,000 active individuals registered in its stud-book, the Pura Raza Española (PRE) breed, also known as the Andalusian breed, is the most important and widespread horse breed in Spain, although 23.30% of its census is distributed over 62 different countries around the world (Solé et al., 2018). It is mainly used as a leisure horse and

in equestrian sports, usually dressage. In 2018, the Spanish Ministry of Agriculture, Fisheries and Food (MAPA) officially recognized the Carthusian strain (CS) within the PRE breed (BOE-19, 2018).

The origins of CS go back to the 15th century, when Carthusian monks from Jerez de la Frontera, Andalusia, established a stud using autochthonous mares and stallions (Valera et al., 1998). The selection and inbred breeding

performed by the monks over more than three centuries was followed by a number of different breeders from 1,810 to the present day (Altamirano, 2017). As a result, Carthusian horses exhibit some morphological singularities with respect to PRE horses; for instance, Carthusians horses tend to be smaller in size (Valera & Molina, 1995) and also have different coat colours: grey is very predominant in CS, while PRE horses show more balanced frequencies between grey and bay colours (Solé et al., 2019).

Otherwise, the influence of the CS on the PRE has been remarkable. A previous study by Valera et al. (2005) using genealogical analysis showed that there was no genetic differentiation between Carthusians and most of the PRE population, due to the high number of descendants from Carthusian founders. Nonetheless, this also implied that the concentration of common genes is very high in both populations and that the contribution of certain founders may be greater than others, leading to an imbalance (Cunningham et al., 2001; Dell et al., 2020).

The facts that PRE studbook is closed and that an unequal use of founders and ancestors may have led to a bottleneck effect are both conditions that favour an increase in inbreeding. It has been demonstrated that high levels of inbreeding negatively affect morphological traits in this breed (Gómez et al., 2009, 2020) leading to morphological defects such as knock knee, cresty and ewe neck (Poyato-Bonilla et al., 2020; Ripollés et al., 2020; Sánchez et al., 2016) disorders such as melanoma and vitiligo (Sánchez-Guerrero et al., 2019) and reduced reproductive parameters (Perdomo-González et al., 2020). An increase in the inbreeding coefficient is also related to the loss of genetic diversity and presents negative effects on the effective population size (Bussiman et al., 2018).

For these reasons, the monitoring of the population structure and genetic variability is key to avoid the risk of reducing genetic diversity. In the CS, where more inbred matings inevitably occur, the importance is even higher. Traditionally, a population has been characterized using genealogical (GEL) information. However, when there is the lack of knowledge about the relationship among founder animals of the pedigree, the use of molecular information can amend these data (Dell et al., 2020).

In the past decades, populational analysis in animals started to rely in the use molecular data to estimate structure and genetic variability (Vignal et al., 2002). In horses, first molecular analysis included STR data (Hasler et al., 2011; Janssens et al., 2010). But nowadays, two array-based SNP chips are available in the species allowing to obtain large and reliable genotypes (from 65K to 670K markers per individual) at an affordable cost (Schaefer & McCue, 2020). This new technology allowed to perform high-throughput genomic analysis which are more reliable and accurate including, among others, populational analysis (Brumfield et al., 2003) individual inbreeding values (Kardos et al., 2015) or ancestry levels (Alexander et al., 2009). In horses, high-density chips were recently employed to analyse Persian Arabian (Sadeghi et al., 2019) and Arabian horse

populations (Cosgrove et al., 2020) as well several European equine breeds (Solé, Ablondi, et al., 2019). In all the cases, the use of genomic information increased the reliability of the estimations, providing a more meaningful view of populational parameters since their dependence of pedigree completeness and reliability is much lower than using GEL data.

This study aimed to clarify the population structure of the PRE breed and the situation of genetic variability and demographic evolution of the Carthusian PRE horse strain throughout its 500 years of life. This progress was assessed using a quantitative approach and genealogical data and it is the first time that a combination of pedigree records and high-density genomic information has been employed for this purpose in the breed.

2 | MATERIALS AND METHODS

2.1 | Genealogical analysis

The Studbook information was provided by the National Pura Raza Español Breeders' Association (ANCCE) and comprised pedigree records of a total of 348,429 horses (170,129 males and 178,300 females) from its foundation in 1912 to 2019. As animals were initially registered with known generations, horses born before 1900 can also be found in the pedigree. Of these, 56,105 (27,326 males and 28,729 females) belong to the CS of the PRE. The percentage of Carthusian blood was computed for the entire PRE population based on the number of Carthusian ancestors of each horse. From this, three different subpopulations were established:

- PRE_Carthusian (PRE_C). Composed of PRE horses with 100% Carthusian blood.
- PRE_Without Carthusian Influence (PRE_NC). PRE horses with no relationship with CS.
- PRE_With Carthusian Influence (PRE_CI). The rest of PRE population.

The pedigree completeness of the populations was studied across the equivalent complete generations (EqG). This parameter was calculated as the addition over all known ancestors of the terms computed as the sum of $(\frac{1}{2})^n$, where n is the number of generations separating the individual from each known ancestor (Maignel et al., 1996). This parameter allows us to estimate pedigree depth and, therefore, its quality.

2.1.1 | Genetic variability by probability of gene origin

To assess the genetic variability of PRE and Carthusian horses, the effective number of founders (f_e), effective

number of ancestors (f_a), founder genome equivalents (f_g), f_a/f_e and f_g/f_e ratios and the number of ancestors that explain 50% of the gene diversity (N_{50}) were computed.

f_e is the reciprocal of the probability that two alleles selected at random in the studied population originated from the same founder, while f_a is the minimum number of ancestors, not necessarily founders, explaining the complete genetic diversity of the populations. For its part, f_g is defined as the number of founders that would be expected to produce the same genetic diversity as in the population under study if the founders were equally represented and no loss of alleles occurred. This parameter was estimated as the inverse of twice the average coancestry of the individuals included in a predefined reference population (Caballero & Toro, 2000).

The Genetic Conservation Index (GCI) assesses the genetic contribution of founders that each animal retains from the initial genetic information of the breed or population (Alderson, 1992).

Inbreeding (F) and average relatedness (AR) coefficients were also computed. F is defined as the probability that an individual possesses two identical alleles by descent at a randomly chosen locus (Málecot, 1948). In terms of population level, this parameter is of high importance as it represents the percentage of heterozygous loci in the original population that have become homozygous in the current one due to related matings. It was computed by the algorithm for computing inbreeding coefficients in large populations formulated by Meuwissen and Luo (1992). Recent inbreeding at the 3rd generation (F_3) and ancient inbreeding over 3 generations were also estimated. Meanwhile, AR is the likelihood that an allele randomly chosen from the population belongs to an individual. It can also be interpreted as the representation of an animal in the whole pedigree, regardless of the knowledge of its own pedigree.

The effective size (N_e) is defined as the size of an idealized population with the same amount of inbreeding as the population under study and was computed via the individual increase in inbreeding following the approach of Gutierrez et al., (2008) in the way proposed by Gutiérrez et al. (2009). The choice of this method was done according to POPREP (Groeneveld et al., 2009) decision cascade side conditions.

2.1.2 | Genetic differentiation

The proportional contribution of PRE and CS to the total diversity of the breed was characterized by the loss or gain of genetic diversity in the whole population after the removal of one of these groups of animals from the whole population (Caballero & Toro, 2002). Remaining gene diversity (GD) is the proportional contribution of each population to the global coancestry, calculated as the average coancestry of the sub-population minus its average distance with all the others.

ENDOG v4.8 (Gutiérrez & Goyache, 2005) and POPREP (Groeneveld et al., 2009) softwares were used to carry out the analyses of the above genealogical parameters.

2.2 | Genomic analysis

2.2.1 | Horses and genomic collection

GEN data were obtained from 732 horses selected from 322 studs showing the present diversity of the population, following the Food and Agriculture Organization of the United Nations recommendations for the genomic characterization of animal genetic resources to select samples (FAO, 2021). Animals with a strong relationship (parent-foal, full siblings, half-siblings and first-degree cousins) were eliminated. A total of 287 PRE horses from 194 studs (53 stallions, 234 mares) were selected and classified according to the percentage of Carthusian blood in the same way as in the GEL analyses as follows: 38 with 100% Carthusian blood (PRE_C), 64 with no relationship with Carthusian blood (PRE_NC) and 185 with different degrees of Carthusian blood (PRE_CI).

2.2.2 | Genotyping and quality control

Genomic DNA was isolated from blood or hair samples using DNeasy Blood & Tissue Kit extraction kits (Qiagen). Thereafter, all the individuals were genotyped using the Affymetrix Axiom™ Equine 670K SNP Genotyping Array (Thermo Fisher), including 670,804 markers uniformly distributed across the entire genome (Schaefer et al., 2017). Genotype calls were performed following the “best practices workflow” procedure in the AXIOM ANALYSIS SUITE package v5.0 (Thermo Fisher Scientific, 2019) with default parameters (DQC \geq 0.82, individual call rate (QC) \geq 0.95 and SNP call rate \geq 97). Thereafter, SNPs with a minor allele frequency (MAF) $<$ 0.01 were excluded using PLINK software v1.9 (Chang et al., 2015), retaining 449,393 SNPs. An additional data set including 159,543 SNPs was created for AMOVA and admixture analysis by pruning by linkage disequilibrium in PLINK (*--indep-pairwise 50 5 0.5* option).

2.3 | Statistical analysis

2.3.1 | Population Structure and Genomic differentiation

Principal component analysis (PCA) was performed to assess population stratification, using the *pcair* function of the GENESIS package (Gogarten et al., 2019) in the R statistical environment V4.1.0 (R-Core-Team, 2021). This procedure

included the molecular kinship among each pair of individuals to correct the negative effect of relatedness in the analysis.

In addition, the genomic differences between subpopulations (PRE_C and PRE_NC) were also estimated by an analysis of molecular variance (AMOVA) described by Excoffier et al., (1992). To this, we first converted the genotypes to a “*genind*” object using the *RADIATOR* package (Gosselin, 2020) and then performed the AMOVA test using the *POPPR* package (Kamvar et al., 2015) in R. The statistical model included comparisons among subpopulations, studs within subpopulations and individuals within a studs.

Finally, the probability of allocation of each individual to a given group based on molecular data was estimated using the *unsupervised* and *supervised* options in *ADMIXTURE* software V1.3 (Alexander et al., 2009). The optimal K value in *unsupervised* mode was selected based on the lowest cross-validation (CV) error.

2.3.2 | Genomic variability and effective population size (N_e) estimates

Genomic variability was characterized using the heterozygosity, molecular inbreeding and kinship parameters.

Observed (H_o) and expected (H_e) heterozygosity was calculated in the whole population and within each subpopulation (PRE_C and PRE_NC) separately using *PLINK* (*--hardy* option).

The molecular inbreeding (F_{ROH}) was estimated using the methodology originally described by McQuillan et al. (2012). To this, we first estimated runs of homozygosity using the “*sliding windows*” method implemented in *DETECTRUNS* R package (Biscarini et al., 2018) with the following parameters: windows size = 50, minimum number of SNPs per ROH = 100, minimum ROH length = 1,000,000 bp, and a maximum gap between contiguous SNPs = 100,000 bp, maximum number of missing and heterozygous SNP per window = 1, and a threshold of overlapping windows of the same state = 0.05. Then, the molecular inbreeding was estimated per individual at chromosome level as the sum of all the ROH fragments divided by the total length (in bp). Finally, F_{ROH} values, F_{ROH} produced during the last 3 and 9 generations, were estimated based on the expected length of homozygous tracts produced by a inbred mate occurred N generations ago (ROH > 16.6Mb for 3 generations and ROH > 5.5Mb for 9 generations) as was proposed by Fisher (1954). Next, we determined the regions in which ROH abundance was overrepresented (ROH islands, ROHi). To this, we first identified the SNPs statistically enriched within ROH by performing a permutation test according to the methodology described by Goszczynski et al., (2018). To this, we compared one million times the number of ROH at each locus to a random value taken from the population. A p-value per loci was determined

as the fraction of times that randomized value was greater than the observed. All the regions of the genome showing a 100-fold increase in the Carthusian group were selected as ROHi.

The estimation of molecular kinship (KIN) was estimated following the methodology described by Manichaikul et al. (2010) in R using the *SNPRELATE* package (Zheng et al., 2012). Genotypes were converted to *GDS* objects applying the *snpgdsBED2GDS* function and analysed *snpgdsIBDKING* (options *type*= “robust” and *remove.monosnp* =T). Results were linearized in pairs for further comparisons with pedigree-based kinship analysis using self-made R scripts.

The variation in the effective population size (N_e) during the last 100 generations was estimated using a linkage disequilibrium (LD)-based approach (Hill, 1981). The analysis was implemented in *SNEP* software (Barbato et al., 2015) with the following parameters: bins per analysis = 250, minimum items per bin = 200 and maximum distance between markers included in the estimations = 2Mb.

Data manipulation and plotting was performed in R using the following packages: *data.table* (Dowle & Srinivasan, 2019), *ggplot* (Wickham, 2016) and *Tidverse* (Wickham et al., 2019).

3 | RESULTS

3.1 | Genealogical analyses

In order to compare the results of the genealogical parameters that characterize subpopulations and the most recent generation, they were also calculated for animals born in the last 10 years, considering it a generation interval (Table 1). The number of PRE_CI and PRE_CN (PRE_CI_NC) horses born since 2010 represented 36.46% of the total, while only 8.72% of individuals of the PRE_C were born in this decade. Meanwhile, the percentage of males and females was balanced in all populations, with the number of females being higher in almost all cases and the number of stallions being lower than the number of broodmares.

Pedigree completeness was higher in all cases than 8 equivalent generations, ranging between 10.05 (PRE_C) and 10.44 (PRE_CI_NC last generation). The probability of gene origin studied across the f_c, f_a, f_g and f_a/f_c and f_g/f_c ratios was very similar between both the subpopulations studied, although the latest generation PRE_C horses showed less population diversity. In both PRE_CI_NC and PRE_C, parameter f_c , which ranged from 20 in PRE_C last generation to 36 in PRE_CI_NC, is higher than f_a (ranging from 11 to 21, respectively) and, in turn, f_g values are lower than them (from 5.23 in PRE_C last generation to 8.53 in PRE_CI_NC last generation). The number of ancestors that explain the 50% of

TABLE 1 Genealogical information

Parameters	Latest generation		Latest generation	
	PRE_CI_NC	PRE_CI_NC ^a	PRE_C	PRE_C ^a
<i>N</i>	292,324	106,589	56,105	4,891
<i>N</i> males (% <i>N</i>)	142,753 (48.83)	52,980 (49.70)	27,376 (48.79)	2,446 (50.01)
<i>N</i> females (% <i>N</i>)	149,571 (51.17)	53,609 (50.30)	28,729 (51.21)	2,445 (49.99)
<i>N</i> stallions	23,222	12,587	5,231	1,082
<i>N</i> brood mares	74,305	42,244	13,903	2,715
EqG	9.60	10.44	8.42	10.05
f_e	36	36	33	20
f_a	21	21	19	11
f_a/f_e ratio	0.58	0.58	0.58	0.55
f_g	8.58	8.73	6.47	5.23
f_g/f_e	0.24	0.24	0.21	0.26
N_{50}	8	8	7	5
GCI	12.70	13.23	10.00	9.69
GD	0.9461609	0.9446687	0.9469435	0.9447329
N_e (\pm S.D.)	60.77 \pm 5.83	67.10 \pm 5.24	31.57 \pm 3.85	29.72 \pm 2.63
<i>F</i>	6.73%	6.80%	10.88%	14.02%
F_3	1.01%	0.96%	2.17%	2.24%
$F_{>3}$	5.79%	5.91%	8.92%	12.06%
AR	10.80%	10.85%	12.78%	13.63%
$N_{50\%}$ variability	-	-	7	-
$N_{100\%}$ variability	968	-	224	-

Note: Parameters describing pedigree quality, probability gene origin and genetic variability of the PRE_With Carthusian Influence and PRE_Without Carthusian Influence considered as a whole subpopulation (PRE_CI_NC) and PRE_Carthusian (PRE_C) of the Pura Raza Español (PRE) breed.

^aGeneration interval: 10 years. *N*: number of horses; EqG: equivalent complete generations; f_e : effective number of founders; f_a : effective number of ancestors; f_g : founder genome equivalents; N_{50} : number of ancestors that explain 50% of the gene diversity; GCI: genetic conservation index; GD: remaining gene diversity; N_e : effective population size; *F*: inbreeding coefficient; F_3 : recent inbreeding at 3rd generation; $F_{>3}$: ancient inbreeding over 3 generations; AR: average relatedness; $N_{50\%}$ variability: number of animals that explain 50% of population variability; and $N_{100\%}$ variability: number of animals that explain 100% of population variability.

the gene diversity varied from 5 (latest generation PRE_C) to 8 (PRE_CI_NC and latest generation PRE_CI_NC). The GCI ranged from 9.69 (latest generation PRE_C) to 13.23 (latest generation PRE_CI_NC). The GD of all subpopulations in the study was approximately 0.95.

The genetic variability of PRE_C, with higher levels of *F* (10.88% and 14.02% in the whole population and the last generation, respectively) and AR (12.78% and 13.63%), was lower than in the PRE_CI_NC, with *F* values of 6.73% and 6.80% and AR values of 10.80% and 10.85%.

The estimated value of N_e estimated via individual increase in inbreeding varied from 29.72 in the last generation in CS to 67.10 in the latest PRE_CI_NC generation. Interestingly, N_e of the latest generation of PRE_CI_NC is higher than the effective size of the whole population, while the opposite occurs in the PRE_C.

The number of horses explaining the 50% of genetic variability was 7 (Table 1), all of which were 100% Carthusian, and 100% of the variability can be attributed to 1,192 animals, with an average percentage of Carthusian blood of 30.28.

In Figure 1a, the percentage of PRE_NC, PRE_CI and PRE_C horses compared with the total number of animals in their group shows remarkable differences. The percentage of non-related to CS (PRE_NC) horses does not exceed the 1% of the total in any of the decades, which shows that most of PRE_NC animals are concentrated in previous years and that from the 1970s onwards, the majority of PRE horses were related to CS. On the other hand, the last years studied show a high, increasing concentration of PRE_CI and PRE_C horses compared with the whole history of the studbook. The average percentage of Carthusian blood and the number of both PRE_CI and PRE_C populations are also shown (Figure 1a). The percentage of Carthusian blood, which had previously remained stable at values of approximately 96%, has seen a decrease since the 90s.

Matings among PRE_CI_NC horses, with a maximum value of 4.49% in 2000–2009, exhibited the lowest average values of coancestry of the three analysed cases (Figure 1b). This parameter remained stable throughout time among these animals, while coancestries within PRE_C almost doubled

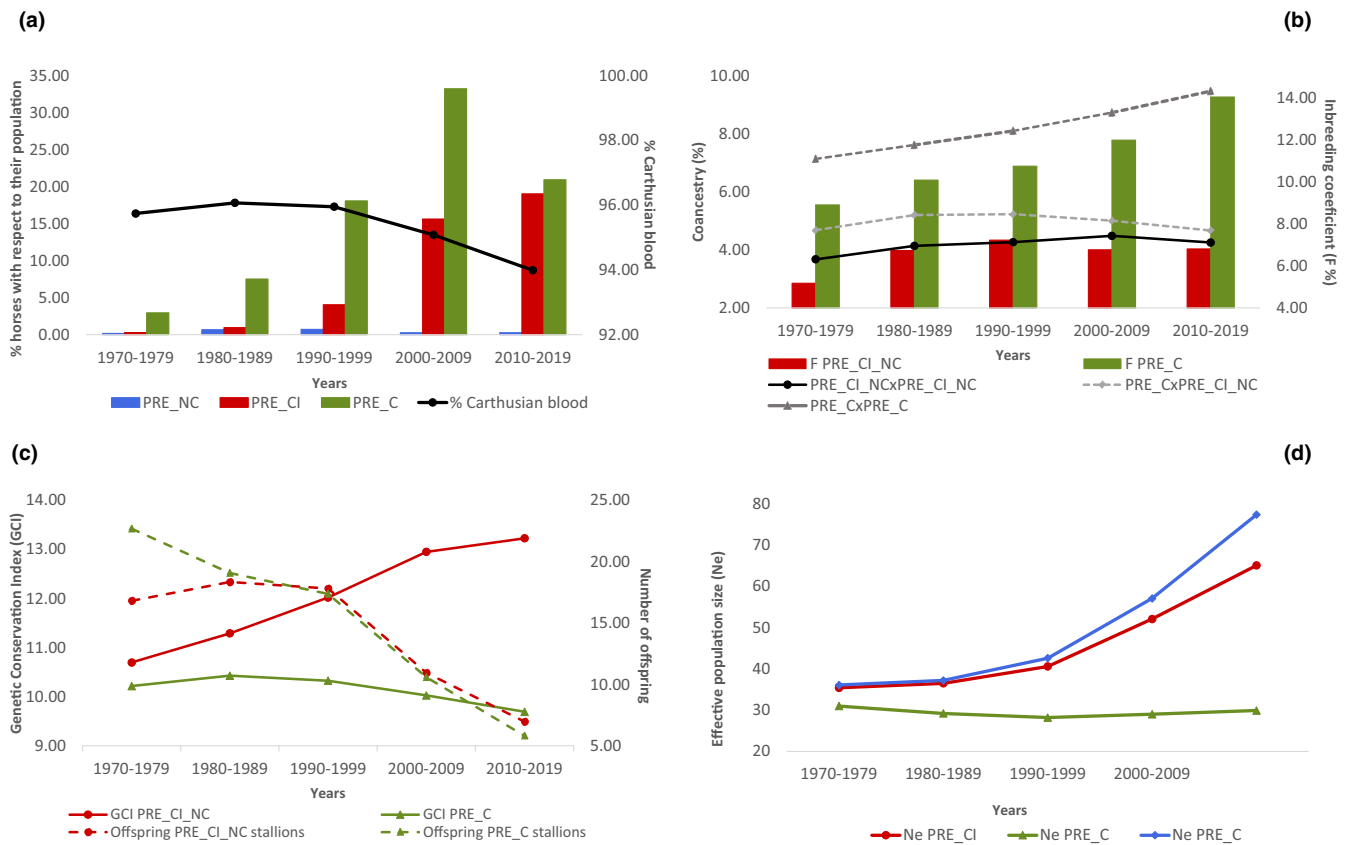


FIGURE 1 Genetic parameters throughout the last 5 decades of Pura Raza Español (PRE) subpopulations analysed. (a) Percentage of PRE_Without Carthusian Influence (PRE_NC) horses, PRE_With Carthusian Influence (PRE_CI) horses and PRE_Carthusian (PRE_C) horses compared with the total number of animals of each group (primary axis), percentage of Carthusian blood (secondary axis); (b) coancestry of matings between PRE_C; between PRE_CI and PRE_NC considered as a whole subpopulation (PRE_CI_NC) and PRE_C; and between PRE_CI_NC horses (primary axis), and inbreeding coefficient (F , secondary axis) of PRE_CI_N and PRE_C horses; (c) genetic conservation index (GCI, primary axis) and number of offspring (secondary axis) of PRE_CI_NC and PRE_C horses; and (d) effective population size of PRE_NC, PRE_CI and PRE_C horses

them (8.75% in 2000–2009 and 9.49% in 2010–2019) and followed a growing trend. The changes in F , also displayed (Figure 1b), indicate that inbreeding in PRE_C horses has always been the highest (reaching 14.02% in the last decade) and, unlike in the PRE_CI_NC (with values oscillating between 5.16% and 7.22%), it has not been possible to reduce nor maintain its values.

Regarding GCI in the PRE_CI_NC (Figure 1c), it has followed an upward trend, rising from 10.70 in the 1970s to 13.23 in 2010–2019. Meanwhile, this parameter has remained more stable and lower in PRE_C, oscillating between 10.43 (1980–1989) and 9.69 (2010–2019). The number of offspring, higher in PRE_C horses in the first decade studied (22.69), has undergone similar changes in both populations, decreasing from approximately 17 sons or daughters per breeding animal in the 1990s to only 6.97 (PRE_CI_NC) and 5.82 (PRE_C) in the last decade.

The effective population size, computed via the individual increase in inbreeding (Figure 1d), has remained at values close to 30 in the PRE_C. Meanwhile, in PRE_CI and

PRE_NC, N_e has seen a similar rise over the same period: from 35.40 (PRE_CI) and 36.10 (PRE_NC) to 65.10 and 77.40, respectively. Moreover, the highest values of N_e correspond in all decades to PRE_NC.

3.2 | Genomic analyses

3.2.1 | Population structure and genomic differentiation

PCA demonstrated a differentiation between PRE_NC and PRE_C horses (7.88% of the variance explained by PC1 and PC2, Figure 2a). This is also depicted by the intermediate position observed when PRE_CI horses were also included (Figure 2b).

Similar results were observed in the supervised (Figure 3a) and unsupervised (Figure 3b) admixture analysis. The first analysis ($K = 2$) showed that 49 horses of the unassigned PRE_CI subpopulation are closer to the PRE_C, 6 of which

would be genomically indistinguishable from PRE_C, while 136 PRE_CI animals could be assigned to the PRE_NC subpopulation, 38 of which with an assignment proportion higher than 0.95 ($K = 2$). In the non-supervised analysis, the best fitting number of clusters was K3 (cross-validation error of 0.47), which indicates a genomic differentiation between the PRE_C and PRE_NC subpopulations (Table S2). However, while 94.73% of the PRE_C horses were assigned to subpopulation 2, the PRE_NC subpopulation was not homogeneous, being differentiated in two subpopulations at the same time (subpop. 1 and subpop. 3).

The molecular variability (% variance) among individuals within stud was very low (Table 2), whereas the existence of two subpopulations of PRE breed explained the 5.23% of variance. In the same way, the genetic variation among each stud within subpopulations was lower (1.21%). The rest of the genetic variability was explained by intra-individual variability (93.30%).

The AMOVA for each subpopulation separately showed genetic diversity among each stud was greater in the PRE_NC than in the PRE_C subpopulation (Table 2).

3.2.2 | Genomic variability

The observed (H_o) and expected (H_e) overall mean heterozygosity was 0.26 and 0.27, respectively, for the PRE breed (Table 3). The PRE_C subpopulation showed lower values of 0.238 and 0.233 for H_o and H_e , respectively, than the PRE_NC subpopulation (0.27 and 0.279, respectively).

The evolution of H_o in the whole population (PRE) showed a slight decrease in the last decades, while H_e has practically remained constant. In the PRE_C subpopulation, H_o showed a slight increase in the last 10 years, as did the PRE_NC population, with higher values of 0.278. Finally, H_e decreased in both subpopulations in the last decade, with the

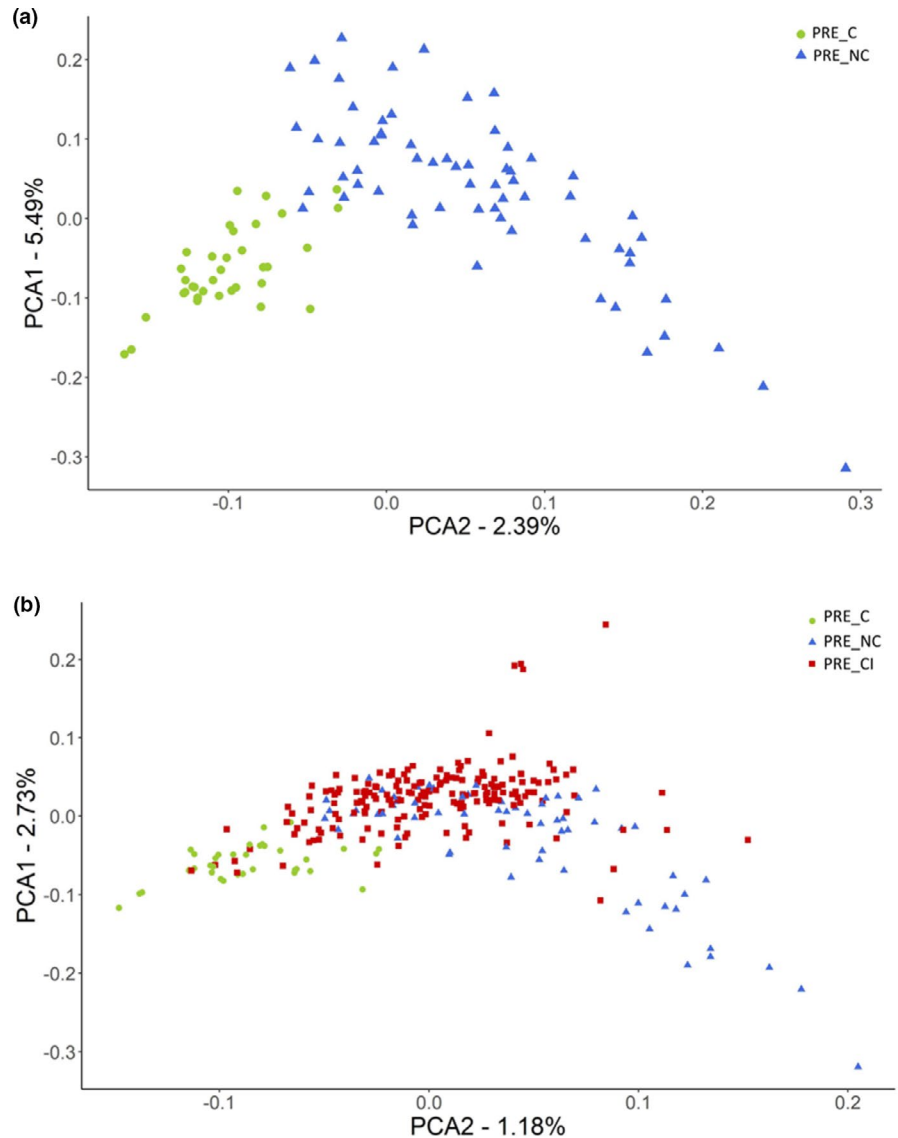


FIGURE 2 Principal component analysis for analysed subpopulations of Pura Raza Español (PRE) horses: (a) first component (PCA1) and second component (PCA2) for PRE_Carthusian (PRE_C) and PRE_Without Carthusian Influence (PRE_NC) subpopulations. (b) PCA1 and PCA2 for all subpopulations. Legend: PRE_C: PRE_Carthusian; PRE_NC: PRE_Without Carthusian Influence; and PRE_CI: PRE_With Carthusian Influence

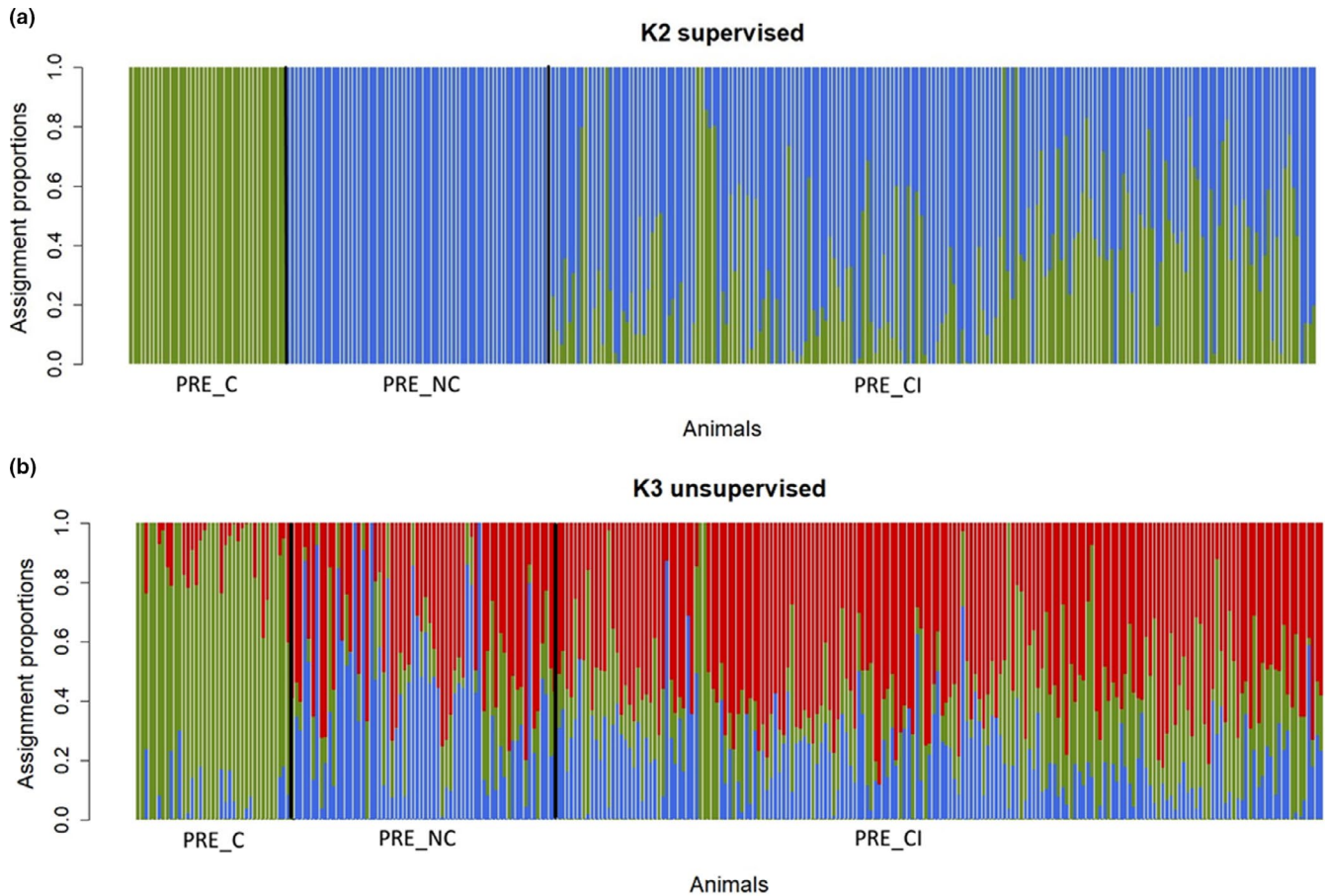


FIGURE 3 Admixture analysis in Pura Raza Español (PRE) subpopulations: (a) $K = 2$ for supervised analysis and (b) $K = 3$ for non-supervised analysis. Legend: PRE_C: PRE_Carthusian; PRE_NC: PRE_Without Carthusian Influence; and PRE_CI: PRE_With Carthusian Influence

TABLE 2 Analysis of molecular variance among subpopulations of Pura Raza Español (PRE) breed, among the stud within a subpopulation, among individuals within a stud and among individuals

Population	N	N Stud	Variance components (%)			
			Subpop.	Stud within subpop.	Individual within stud	Within samples
PRE_CI	102	22	5.219	1.210	0.428	93.302
PRE_C	38	9		0.438	0.429	99.135
PRE_NC	64	13		4.350	0.469	95.184

Note: PRE_CI: PRE_With Carthusian Influence; PRE_C: PRE_Carthusian; PRE_NC: PRE_Without Carthusian Influence; N : number; and Subpop.: subpopulation.

PRE_C subpopulation showing a greater decrease from 0.279 to 0.264 (Table 3).

Inbreeding, measured as the proportion of the genome covered by ROH, resulted in an average F_{ROH} of 0.15 in the PRE population. The PRE_C subpopulation was characterized by higher levels of homozygosity ($F_{ROH} = 0.195$) than in the PRE_NC subpopulation ($F_{ROH} = 0.133$) (Table 3). In the last ten years, inbreeding levels in the PRE breed have increased to 0.17, whereas in the PRE_C and PRE_NC

subpopulations, they have decreased to values of 0.186 and 0.106, respectively.

To differentiate ancient and recent inbreeding, we calculated F_{ROH} considering different ROH lengths (Table 3). The evolution of inbreeding over generations 3 ($F_{ROH > 16Mb}$) to 9 ($F_{ROH < 6Mb}$) has followed the same trend in all populations. In the PRE_C subpopulation, F_{ROH} has decreased from 0.066 ($F_{ROH \text{ Ancient}}$) to 0.049 ($F_{ROH \text{ Recent}}$) and in the PRE_NC subpopulation from 0.048 to 0.039. The $F_{ROH \text{ Ancient}}$ ($F_{ROH < 6Mb}$)

TABLE 3 Genomic variability of Pura Raza Español (PRE) breed and subpopulations analysed (mean \pm SD)

Parameters	PRE breed		PRE_C		PRE_NC	
	Whole	Latest generation ^a	Whole	Latest generation ^a	Whole	Latest generation ^a
<i>n</i>	287	41	38	8	64	8
Ho (\pm SD)	0.260 \pm 0.164	0.253 \pm 0.170	0.238 \pm 0.190	0.243 \pm 0.227	0.270 \pm 0.166	0.278 \pm 0.221
He (\pm SD)	0.270 \pm 0.168	0.268 \pm 0.171	0.233 \pm 0.181	0.226 \pm 0.187	0.279 \pm 0.166	0.264 \pm 0.180
F _{ROH} (\pm SD)	0.150 \pm 0.071	0.170 \pm 0.091	0.195 \pm 0.070	0.186 \pm 0.067	0.133 \pm 0.07	0.106 \pm 0.042
F _{ROH Ancient (<6Mb)} (\pm SD)	0.054 \pm 0.017	0.053 \pm 0.015	0.066 \pm 0.023	0.059 \pm 0.012	0.048 \pm 0.011	0.043 \pm 0.012
F _{ROH Recent (>16Mb)} (\pm SD)	0.040 \pm 0.044	0.057 \pm 0.059	0.049 \pm 0.039	0.049 \pm 0.039	0.039 \pm 0.05	0.029 \pm 0.030
Kinship (\pm SD)	0.003 \pm 0.017	0.005 \pm 0.024	0.035 \pm 0.031	0.030 \pm 0.031	0.008 \pm 0.024	0.009 \pm 0.019
N _e	188	190	29	30.2	51	50.2

^aGeneration interval: 10 years. PRE_C: PRE_Carthusian; PRE_NC: PRE_Without Carthusian Influence; *N*: number of animals; He: expected heterozygosity; Ho: observed heterozygosity; F_{ROH}: genomic inbreeding; F_{ROH Ancient}: ancient genomic inbreeding; F_{ROH Recent}: recent genomic inbreeding; and N_e: effective population size.

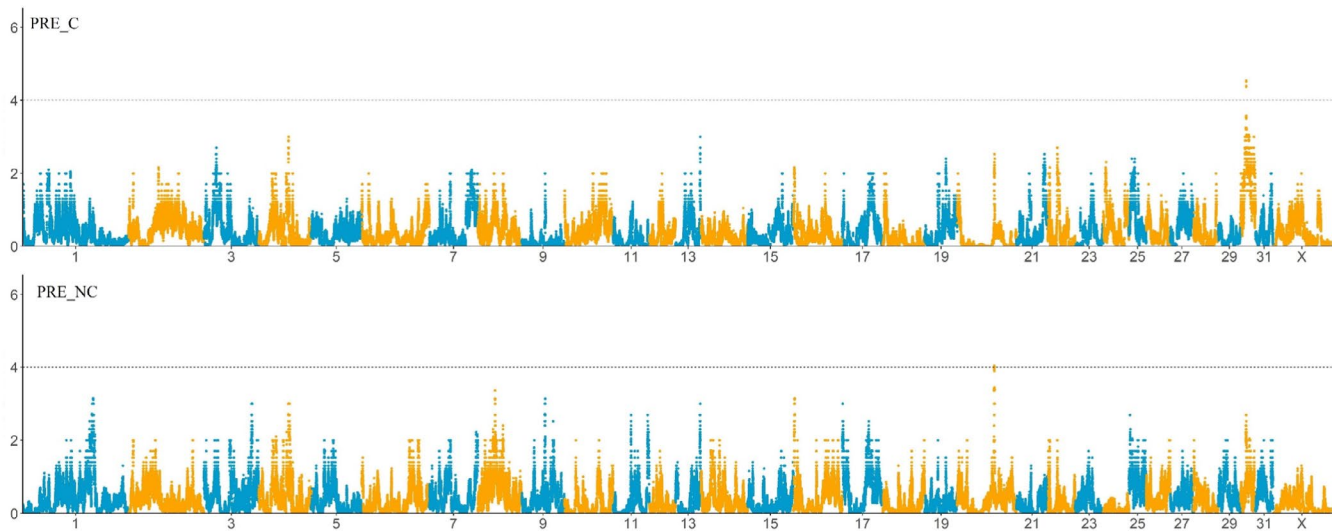


FIGURE 4 Whole-genome homozygosity mapping. The enrichment of runs of homozygosity (ROH) was evaluated using the permutation test. The Manhattan plot shows the significance of each ROH region (ordinate axis) across the entire autosome genome (abscissa axis) in Pura Raza Español Carthusian (PRE_C) and Pura Raza Español Without Carthusian Influence (PRE_NC) subpopulations. Legend: PRE_C: PRE_Carthusian and PRE_NC: PRE_Without Carthusian Influence

was 0.054 in the PRE population and decreased to 0.04 F_{ROH Recent} (F_{ROH > 16Mb}).

Three ROH islands showed an increase in homozygosity of two folds (100 \times) in the PRE_C group (data not shown), all of which were located in ECA30, with a clear peak only present in the PRE_C (Figure 4).

3.2.3 | Effective population size

The estimate of N_e in the whole PRE breed was 188, while in the PRE_C and PRE_NC subpopulation, N_e was estimated to be approximately 29 and 51, respectively (Table 3). The evolution of effective population size (N_e) across the last four decades is shown in Figure S1. In the

last two decades, N_e has seen a considerable rise in the PRE_C and PRE_NC subpopulations, from 5 (PRE_C) and 12.2 (PRE_NC) to 30.2 and 50.2, respectively, while the rise was a decade earlier in the case of the PRE_With Carthusian influence (CI).

4 | DISCUSSION

The history of PRE is commonly associated with endogamous matings, aimed to obtain specific morphotypes and set of skills. This practice has produced some degree of differentiation in certain genetic lines. Among these, the CS (the only officially recognized subpopulation within the breed) has acquired great prestige for its historical and economic

importance throughout the 5 centuries since its creation in the 15th century in the Jerez monastery of La Cartuja.

Previous studies were based on the PRE studbook and revealed that the current Carthusian population derives from a few horses which were protected from the introduction of Central European heavy horses after the 16th century (Valera et al., 2005). Subsequently, the breed was recovered by the studs of religious orders between the 18th and 19th centuries with virtually no influence from foreign breeds. The morphological differentiation between PRE_C and PRE_NC individuals remains to this day.

The prestige of the breed has been growing throughout the last centuries, making breeders seek to mate their breeding horses with the most outstanding individuals of this lineage. However, a change can be seen in many breeders' view of this population in the last 3 decades (Valera et al., 2005). The breed has increased in terms of census since 1970 (Figure 1a), but the number of births of PRE_CI individuals is highly concentrated over the last generation. In contrast, the percentage of PRE_NC horses has seen a decrease in the last 20 years and currently, the CS (PRE_C) represents only 8.72% of living animals and the last decade has seen a fall in PRE_C births. This, together with the decline of the percentage of Carthusian blood in the whole breed since the 2000s, could be attributed to different factors. Firstly, to the recognition of the breeders' association (ANCCE) by the Ministry of Agriculture in 2007 as the leading representative entity to manage the PRE studbook, which promoted both the breed and the development of the improvement programme. The PRE breeding programme aims at maintaining the genetic variability, which involves avoiding inbred matings, which are hardly found among PRE_C and PRE_NC horses. Secondly, the recent interest of PRE breeders towards different coat colours from grey (Poyato-Bonilla et al., 2018), which is the prevalent coat colour in the Carthusian lineage (around 80% according to PRE studbook data), could have contributed to an increase of matings between PRE_CI and PRE_NC animals and PRE_C individuals compared with those within the strain. Finally, the current preference of breeders for larger horses more suitable for sports disciplines such as dressage (Valera & Molina, 1995) has contributed to this differential growth of the census of both subpopulations.

The small size of the PRE_C subpopulation has forced breeders to mate in purity (Valera et al., 2005), triggering a considerably higher mean inbreeding coefficient for the PRE_C than the F of the rest of PRE population (Table 1). The PRE breeding programme, which avoids inbred matings as much as possible to reduce the upward trend of inbreeding and coancestry, has been successful in PRE_CI_NC animals, but not in PRE_C (Figure 1b). The scarce number of Carthusian individuals without common ancestors available for reproduction makes it especially difficult to plan non-inbred matings within pure Carthusians. In this situation,

Valera et al. (2005) found inbreeding coefficients of 13.01% in the PRE_C and 8.06% in the PRE (PRE_NC).

Recent and ancient F values tally with the differences between PRE_C and PRE_CI_NC subpopulations, with the biggest differences between recent and ancient F percentages of PRE_C. The inbreeding observed in the PRE population may well be linked to ancient Carthusian individuals (Valera et al., 2005). The average relatedness of PRE_CI_NC was lower than that of the PRE_C, and this parameter has become higher in the strain in the last generation, while in PRE_CI_NC, it has hardly increased. This maintenance of AR values in PRE_CI_NC together with the decrease in recent F in the last generation indicates a successful policy in the PRE breed. Valera et al. (2005) found an average AR of 12.09% in PRE_CI_NC horses and 14.00% in PRE_C. Previous studies in other equine breeds related to PRE estimated F average values of 4.06% in the Lusitano horse (da Silva Faria et al., 2018) and 10.8% in the Lipizzan horse, with a population size quite similar to the PRE_C subpopulation (Zechner et al., 2002). AR values in these breeds were lower: 3.73% and 5.41%, for Lipizzan and Lusitano horse, respectively (Pjontek et al., 2012; da Silva Faria et al., 2018).

The degree of relationship within a breed with such a high census may be due, firstly, to a limited number of founders, as well as to the interest of breeders to increase the genetic resemblance to certain outstanding individuals. The founder with the highest representation in the whole PRE breed (mare Destinada, Table S1) accounted for an AR of 11.65%, while the two major ancestors explained 13.84% and 11.81%, respectively, of the variability of the breed. Furthermore, only 7 ancestors were needed to explain 50% of the genetic variability. All the major founders and ancestors were 100% Carthusian animals, which shows the influence of the CS on the origins of the breed.

Therefore, the current PRE_CI_NC subpopulation show a high census, with high variability, and a sustained increase in F (despite the endogamous matings of many of its breeders), due to measures to prevent its increase. In contrast, in the case of the PRE_C subpopulation, with a small current population size, the possible existence of population bottlenecks and the type of inbred matings that Carthusian horse breeders follow, monitoring of the increase in F that is taking place is needed.

The reliability of these estimates is often conditioned by the quality of the pedigree. Here, the pedigree completeness, studied across equivalent complete generations, with a mean value of 8.42 in PRE_C (9.01 on average in the whole population), far exceeds the values of other important Spanish horse populations such as Spanish Sport Horse (5.31), Anglo Arab Horse (5.50) or Spanish Trotter Horse (3.87) (Gómez et al., 2020), and also of Lusitano (5.70) (da Silva Faria et al., 2018). The Lipizzan horse, a breed derived from ancient PRE horses, presented a value of the same magnitude, with 10.25 EqG (Pjontek et al., 2012). A considerable increase in

known generations can be observed in the last 10 years (up to 1.63 EqG in the case of the PRE_C). These results demonstrate the consistency and accuracy of the pedigree.

However, numerous studies from the last decade (de Cara et al., 2011; Engelsma et al., 2011) have shown that the real situation of a population cannot be defined exactly by GEL analyses. Although the degree of knowledge and reliability of the pedigree of a breed is very high (as is the case of the PRE horse, in which affiliations with genetic markers have been confirmed since the 1980s), analyses with GEN data can detect differences in most parameters related to genetic variability and population stratification and differentiation, the degree of kinship and even the actual effective size of the populations, which are not detectable with a GEL analysis (de Cara et al., 2011; Engelsma et al., 2011; Wiener et al., 2017).

Nowadays, it is accepted that genomic estimations of inbreeding (genomic homozygosity) are more reliable than using pedigree records since most of the variability parameters estimated with the genealogical information assume a probable average rather than an actual realized value (Howard et al., 2017), despite our genealogical estimations being highly reliable (which are based on a large data set including more than 8 ECG on average).

In general, GEN inbreeding coefficients (F_{ROH}) were higher than those estimated from the GEL data, but showed a correlation of 0.68. This value is higher than that reported for the Noriker and Austrian Haflinger horse breeds (0.19 and 0.38, respectively) (Druml et al., 2017) and similar to Italian Heavy Draught horses (0.65) (Mancin et al., 2020). Those differences were even more remarkable in PRE_C horses, but showed a lower correlation. This could be explained by the high degree of homozygosity detected in the first founders of this strain.

Comparison with recent studies reported in other horse populations with small population sizes reveals that the PRE_C has a particularly high level of inbreeding. For example, breed values were found in the Croatian Posavje horse of 8.6% (Grilz-Seger et al., 2018), Belgian draft horse 10.1% (Schurink et al., 2019), Bosnian mountain horse 13% (Druml et al., 2017), Austrian Noriker between 8% and 13% (Grilz-Seger et al., 2019) and Persian Arabian (14%) (Sadeghi et al., 2019). Similar levels of F_{ROH} for the PRE_C were estimated in other populations, such as Bardigiano horses (17%) (Ablondi et al., 2020), and in Purebred Arabian horses (17.7%), despite having a large population size (Druml et al., 2017). In contrast, a higher level of inbreeding was found in Friesian horses (22.3%) (Schurink et al., 2019). However, this comparison should be made with some caution since the estimation of the F_{ROH} depends to a great extent on the calculation setup and the SNP number used (Meyermans et al., 2020).

Since the problems of a high degree of genomic homozygosity (inbreeding depression) are in principle independent

of whether this homozygosity is in state or by descent, these results indicate that the use of genomic data to estimate the degree of homozygosity of a population and to detect inbreeding effects can complement the results estimated from the pedigree data (Howrigan et al., 2011). F_{ROH} has increased during the last 10 years in the PRE breed, while it decreased in PRE_NC and remained stable in PRE_C (Table 3).

The PRE_C subpopulation also showed the highest values of ancient inbreeding and the lowest values of heterozygosity (H_o), which fits in with the hypothesis that Carthusians originated from a very closed group of founders. It is noteworthy that the H_o values observed in PRE_C (0.238) were even lower than those reported in small, autochthonous breeds such as the Bosnian mountain horse (0.3), the Haflinger Italy horse (0.294) or closed breeds with a low level of admixture such as in Noriker (0.256) (Druml et al., 2017).

Linked to this increase in inbreeding between generations, the effective population size estimated by GEL data in whole PRE_CI_NC was double that of the PRE_C (Table 1), clearly reflecting closed inbreed matings within the CS. However, these differences have been increasing over the last few decades (Figure 1d). In PRE_C, average generation intervals (12.4 years) are higher than those for the rest of PRE (10.11 years) (Valera et al., 2005). These results are also consistent with a remarkable preference for famous studs and specific stallions, which decreases the variety of breeding animals within the strain (Valera et al., 2005). In the PRE_CI_NC, the N_e values obtained agree with but are slightly higher than those obtained in the study by Cervantes et al. (2008), which may be due to the effort in recent years to maintain the increased control of F and the broadening of the preference for one type of animal by breeders. In addition, the PRE_C N_e resembles more the values of an endangered population such as the Hungarian Hucul horse (49.54) (Somogyvári et al., 2018) and reflects the fact that PRE_C are used for longer periods, hence they exhibit longer generation intervals. In PRE_NC and PRE_CI animals, this parameter has followed a marked increase coinciding with the years where the largest number of PRE_CI animals was grouped together, and especially since the implementation of the breeding programme, which demonstrates its effectiveness in controlling genetic parameters by balancing the genetic representation of the founders among the breed. N_e has continued with almost the same values since 1970 in PRE_C. The effective size reflects not only the degree of kinship between the breeding animals, but also the management of the matings (in relation to a panmictic mating). In the case of the PRE_C, in recent years a great effort has been made to search for the least inbred matings, which has determined a lower increase in F in the next generation compared to what would be obtained with random matings. This determines an increase in N_e . Frankham et al. (2002) indicated the need for a value of at least 50 to avoid inbreeding depression. According to this, the

CS may be affected by this phenomenon and must be monitored in order to avoid inbreeding depression on fitness and reproductive traits in the long term (Bussiman et al., 2018).

Although N_e estimated with GEN data was larger than that measured with the GEL information in PRE_CI and PRE_NC, those differences were not detected in PRE_C, where they were very similar (Tables 1 and 3). To our knowledge, our N_e estimation for PRE_C is possibly the lowest population size reported to date, compared with other breeds such as Berdigiano (39) (Ablondi et al., 2020), Italian Heavy Draught horse (100) (Mancin et al., 2020), Persian Arabian Horses (113) (Sadeghi et al., 2019), Noriker-Carinthia (104.7), Croatian Coldblood (87.2) and Altmaerkisch Draught Horse (64) (Druml et al., 2007). In fact, many of these breeds are catalogued as endangered. However, it is also noteworthy that N_e has remained constant in the PRE_C subpopulation during the last 20 years, in agreement with the results obtained using pedigree data. In contrast, PRE_CI and PRE_NC showed a significant increase, caused by the breeding programme applied in this population during the last 25 years, which takes into account several conservation criteria to maintain the genetic variability.

The effective number of ancestors and the effective number of founders are complementary parameters which account for bottlenecks (the reduction of the number of breeding animals of a population) produced by the unbalanced use of reproductive individuals. Although population size can increase from these animals, their genetic variance cannot be higher than that of the lowest population size before it grows (Somogyvári et al., 2018). The ratio between the effective number of ancestors and the effective number of founders (f_a/f_e) has shown the existence of bottleneck effect in PRE breed and also in its CS (Table 1). Moreover, f_g values are lower than f_e and f_a , which indicates a reduction in genetic variability. The ratio between a founder's genome equivalents and the effective number of founders (f_g/f_e) is close to 0.24 in both populations, indicating that the population has been affected by genetic drift. Genetic drift involves and supports the reduction in genetic variability. In a large, closed population such as the PRE, this phenomenon can be probably attributed to over-use of the same breeding animals. The number of ancestors explaining 50 per cent of the gene diversity barely differs between the total PRE_CI_NC and the PRE_C populations. However, while 8 ancestors have accounted for the PRE_CI_NC diversity throughout all its history, the number of ancestors that explain the diversity of PRE_C has been reduced from 7 to 5 in the last generation. This decrease could be associated with the preferential use of certain Carthusian stallions and broodmares and the longer generation intervals within the strain.

Genetic conservation index (GCI) has seen a similar evolution in both PRE_CI_NC and PRE_C subpopulations: over the last decades, a great change has been detected. At the

individual level, GCI is a particularly important parameter in genetic conservation programmes, as horses with higher indexes exhibit a greater balance in the number of founders and should be used preferentially in the choice of breeding animals to maintain the genes transmitted by founders (da Silva Faria et al., 2018). The values of the GCI (Table 1) are in accordance with the value of 11.14 obtained in a previous work in the PRE breed including CS (Poyato-Bonilla et al., 2018). In recent literature, a slightly lower (9.60) index has been reported in a related breed, such as Lusitano (da Silva Faria et al., 2018). The evolution of GCI in the PRE_CI_NC and PRE_C has been the opposite since 1980 (Figure 1c): the number of founders contributing to PRE_CI_NC horses has increased, while this parameter has decreased in the PRE_C. This can be associated with the loss of interest on the part of many breeders for continuing exploiting this strain and the disappearance of the genetic lines present in a certain Carthusian stud.

Despite the fact that Carthusian horses were largely employed as breeders during the last decades of the last century within the PRE population, both populations remain differentiated from a genomic point of view. This can be observed in the PCA (Figure 2) and admixture analysis (Figure 3), in which PRE_CI and PRE_C horses are well clustered. However, there is a high number of individuals not assigned to any of these subpopulations, with a highly variable level of Carthusian gene introgression. This fact can be also observed as 3 "inbreeding peaks" (ROHi) (Figure 4), which are located in different positions of the genome in each group (particularly clear in ECA30), where the PRE_C population was highly homozygous, in contrast to the increased heterozygosity observed in PRE_NC.

Other finding of this study was the degree of intrapopulation differentiation at the stud and individuals within stud level. In PRE_CI, most of the variation was detected between individuals, followed by the differences explained by the studs (Table 2). In this case, the differences between animals of the same stud were very low (0.43%). In contrast, the molecular variation between studs within the PRE_C subpopulation was practically absent, with the genetic differences between and within studs being very similar. This situation contrasts with that of PRE_NC, where the differences between studs are much higher than those existing within each stud, which can probably be explained by breeders having different ideas about the type of animal to breed, and a larger or greater number of genetic lines with different degrees of differentiation.

Despite this, our results analysing the genetic diversity at the population level (studied across GD, a parameter that shows the global diversity when each of the subpopulations studied is removed (Bartolomé et al., 2010) demonstrated that the PRE_C group contributed with almost the same gene diversity to the complete population as PRE_CI_NC animals. Because the number of horses of the PRE_C subpopulation is

much lower than in PRE_CL_NC, GD demonstrates the current importance of the strain within the PRE breed.

5 | CONCLUSIONS

A moderate to high equivalence between genomic and genealogical information has been found comparing to the results obtained in other equine breeds and species, demonstrating the reliability and great depth of the pedigree of the PRE breed. The influence of the CS in the current PRE population is still very high, even if the number of individuals of this strain has decreased in the last years. Despite demonstrating some control of breeders, inbreeding parameters must be monitored regularly and the degree of relatedness of matings should be studied before being applied, especially in the CS.

ACKNOWLEDGEMENT

The authors wish to thank Yeguada de La Cartuja Hierro del Bocado and the National Association of Pura Raza Española Horse Breeders (ANCCE) for their collaboration in this study and for providing biological samples and the genealogical data used in this study.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

Data Availability Statement


The data that support the findings of this study are available from the Asociación Nacional de Criadores de Caballos de Pura Raza Española (ANCCE). Restrictions apply to the availability of these data, which were used under licence for this study. The data are available from the authors with the permission of ANCCE.

ORCID

Julia Poyato-Bonilla  <https://orcid.org/0000-0002-7287-5012>

<https://orcid.org/0000-0002-7287-5012>

Nora Laseca  <https://orcid.org/0000-0003-3753-6725>

Sebastián Demyda-Peyrás  <https://orcid.org/0000-0003-3286-2441>

<https://orcid.org/0000-0003-3286-2441>

Antonio Molina  <https://orcid.org/0000-0002-9566-6600>

Mercedes Valera  <https://orcid.org/0000-0003-1742-550X>

<https://orcid.org/0000-0003-1742-550X>

REFERENCES

Ablondi, M., Dadousis, C., Vasini, M., Eriksson, S., Mikko, S., & Sabbioni, A. (2020). Genetic diversity and signatures of selection in a native Italian horse breed based on SNP data. *Animals*, *10*(6), 1005. <https://doi.org/10.3390/ani10061005>

Alderson, G. L. H. (1992). A system to maximize the maintenance of genetic variability in small populations. In *Genetic conservation of domestic livestock* (Vol. 2, pp. 18–29). CAB International.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. <https://doi.org/10.1101/gr.094052.109>

Altamirano, J. C. (2017). *Historia y origen del caballo español*. Ediciones Ecuestres.

Barbato, M., Orozco-terWengel, P., Tapio, M., & Bruford, M. W. (2015). SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics*, *6*(109). <https://doi.org/10.3389/fgene.2015.00109>

Bartolomé, E., Goyache, F., Molina, A., Cervantes, I., Valera, M., & Gutiérrez, J. P. (2010). Pedigree estimation of the (sub) population contribution to the total gene diversity: The horse coat colour case. *Animal*, *4*(6), 867–875. <https://doi.org/10.1017/s1751731110000182>

Biscarini, F., Cozzi, P., Gaspa, G., & Marras, G. (2018). DetectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes in R. Retrieved from <https://github.com/bioinformatics-ptp/detectRUNS/tree/master/detectRUNS>

BOE-19. (2018). Orden APM/26/2018, de 11 de Enero, Por La Que Se Modifica El Anexo I Del Real Decreto 2129/2008, de 26 de Diciembre, Por El Que Se Establece El Programa Nacional de Conservación, Mejora y Fomento de Las Razas Ganaderas. Boletín Oficial del Estado Español. Retrieved from <https://www.boe.es/boe/dias/2018/01/22/pdfs/BOE-A-2018-749.pdf>

Brumfield, R. T., Beerli, P., Nickerson, D. A., & Edwards, S. V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution*, *18*(5), 249–256. [https://doi.org/10.1016/S0169-5347\(03\)00018-1](https://doi.org/10.1016/S0169-5347(03)00018-1)

Bussiman, F., Perez, B., Ventura, R., Peixoto, M., Curi, R., & Balieiro, J. (2018). Pedigree analysis and inbreeding effects over morphological traits in Campolina horse population. *Animal*, *12*(11), 2246–2255. <https://doi.org/10.1017/s175173111800023x>

Caballero, A., & Toro, M. (2000). Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetical Research*, *75*, 331–343. <https://doi.org/10.1017/S0016672399004449>

Caballero, A., & Toro, M. A. (2002). Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics*, *3*(3), 289–299. <https://doi.org/10.1023/A:1019956205473>

Cervantes, I., Goyache, F., Molina, A., Valera, M., & Gutiérrez, J. (2008). Application of individual increase in inbreeding to estimate effective sizes from real pedigrees. *Journal of Animal Breeding and Genetics = Zeitschrift Für Tierzüchtung Und Züchtungsbiologie*, *125*, 301–310. <https://doi.org/10.1111/j.1439-0388.2008.00755.x>

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. <https://doi.org/10.1186/s13742-015-0047-8>

Cosgrove, E. J., Sadeghi, R., Schlamp, F., Holl, H. M., Moradi-Shahrbabak, M., Miraei-Ashtiani, S. R., Abdalla, S., Shykind, B., Troedsson, M., Stefaniuk-Szmukier, M., Prabhu, A., Bucca, S., Bugno-Poniewierska, M., Wallner, B., Malek, J., Miller, D. C., Clark, A. G., Antczak, D. F., & Brooks, S. A. (2020). Genome diversity and the origin of the Arabian horse. *Scientific Reports*, *10*(1), 9702. <https://doi.org/10.1038/s41598-020-66232-1>

Cunningham, E. P., Dooley, J. J., Splan, R. K., & Bradley, D. G. (2001). Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Animal Genetics*, *32*(6), 360–364. <https://doi.org/10.1046/j.1365-2052.2001.00785.x>

- da Silva Faria, R. A., Vicente, A. P. A., dos Santos, R. I. D. G., Maiorano, A. M., Curi, R. A., Chardulo, L. A. L., Vasconcelos Silva, J. A., & Josineudson Augusto, I. I. (2018). Genetic diversity of Lusitano horse in Brazil using pedigree information. *Journal of Equine Veterinary Science*, *69*, 149–158. <https://doi.org/10.1016/j.jevs.2018.07.009>
- de Cara, M. A. R., Fernández, J., Toro, M. A., & Villanueva, B. (2011). Using genome-wide information to minimize the loss of diversity in conservation programmes. *Journal of Animal Breeding and Genetics*, *128*(6), 456–464. <https://doi.org/10.1111/j.1439-0388.2011.00971.x>
- Dell, A., Curry, M., Yarnell, K., Starbuck, G., & Wilson, P. B. (2020). Genetic analysis of the endangered Cleveland Bay horse: A century of breeding characterised by pedigree and microsatellite data. *PLoS One*, *15*(10), e0240410. <https://doi.org/10.1371/journal.pone.0240410>
- Dowle, M., & Srinivasan, A. (2019). data.table: Extension of data.frame. Retrieved from <https://CRAN.R-project.org/package=data.table>
- Druml, T., Curik, I., Baumung, R., Aberle, K., Distl, O., & Sölkner, J. (2007). Individual-based assessment of population structure and admixture in Austrian. *Croatian and German Draught Horses. Heredity*, *98*(2), 114–122. <https://doi.org/10.1038/sj.hdy.6800910>
- Druml, T., Neuditschko, M., Grilz-Seger, G., Horna, M., Ricard, A., Mesarič, M., Cotman, M., Pausch, H., & Brem, G. (2017). Population networks associated with runs of homozygosity reveal new insights into the breeding history of the Haflinger horse. *Journal of Heredity*, *109*(4), 384–392. <https://doi.org/10.1093/jhered/esx114>
- Engelsma, K. A., Veerkamp, R. F., Calus, M. P. L., & Windig, J. J. (2011). Consequences for diversity when prioritizing animals for conservation with pedigree or genomic information. *Journal of Animal Breeding and Genetics*, *128*(6), 473–481. <https://doi.org/10.1111/j.1439-0388.2011.00936.x>
- Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479–491. <https://doi.org/10.1093/genetics/131.2.479>
- FAO (2021). 11th Session of the ITWG on Animal Genetic Resources for Food and Agriculture. Retrieved from <http://www.fao.org/animal-genetics/events/events-detail/en/c/1369166/>
- Fisher, R. A. (1954). A fuller theory of junctions in inbreeding. *Heredity*, *8*(2), 187–197. <https://doi.org/10.1038/hdy.1954.17>
- Frankham, R., Ballou, J. D., & Briscoe, D. A. (2002). *Introduction to conservation genetics*. Cambridge University Press.
- Gogarten, S. M., Sofer, T., Chen, H., Yu, C., Brody, J. A., Thornton, T. A., Rice, K. M., & Conomos, M. P. (2019). Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics*, *35*(24), 5346–5348. <https://doi.org/10.1093/bioinformatics/btz567>
- Gómez, M. D., Sánchez, M. J., Bartolomé, E., Cervantes, I., Poyato-Bonilla, J., Demyda-Peyrás, S., & Valera, M. (2020). Phenotypic and genetic analysis of reproductive traits in horse populations with different breeding purposes. *Animal*, *14*(7), 1351–1361. <https://doi.org/10.1017/S1751731120000087>
- Gómez, M. D., Valera, M., Molina, A., Gutiérrez, J. P., & Goyache, F. (2009). Assessment of inbreeding depression for body measurements in Spanish Purebred (Andalusian) horses. *Livestock Science*, *122*(2–3), 149–155. <https://doi.org/10.1016/j.livsci.2008.08.007>
- Gosselin, T. (2020). radiator: RADseq Data Exploration, Manipulation and Visualization using R. v.1.1.8. Retrieved from <https://thierrygosselin.github.io/radiator/>
- Goszczynski, D., Molina, A., Terán, E., Morales-Durand, H., Ross, P., Cheng, H., Giovambattista, G., & Demyda-Peyrás, S. (2018). Runs of homozygosity in a selected cattle population with extremely inbred bulls: Descriptive and functional analyses revealed highly variable patterns. *PLoS One*, *13*(7), e0200069. <https://doi.org/10.1371/journal.pone.0200069>
- Grilz-Seger, G., Druml, T., Neuditschko, M., Mesarič, M., Cotman, M., & Brem, G. (2019). Analysis of ROH patterns in the Noriker horse breed reveals signatures of selection for coat color and body size. *Animal Genetics*, *50*(4), 334–346. <https://doi.org/10.1111/age.12797>
- Grilz-Seger, G., Mesarič, M., Cotman, M., Neuditschko, M., Druml, T., & Brem, G. (2018). Runs of homozygosity and population history of three horse breeds with small population size. *Journal of Equine Veterinary Science*, *71*, 27–34. <https://doi.org/10.1016/j.jevs.2018.09.004>
- Groeneveld, E., Westhuizen, B., Maiwashe, N., Voordewind, F., & Ferraz, J. (2009). POPREP: A generic report for population management. *Genetics and Molecular Research: GMR*, *8*, 1158–1178. <https://doi.org/10.4238/vol8-3gmr648>
- Gutiérrez, J. P., Cervantes, I., & Goyache, F. (2009). Improving the estimation of realized effective population sizes in farm animals. *Journal of Animal Breeding and Genetics*, *126*(4), 327–332. <https://doi.org/10.1111/j.1439-0388.2009.00810.x>
- Gutierrez, J. P., Cervantes, I., Molina, A., Valera, M., & Goyache, F. (2008). Individual increase in inbreeding allows estimating effective sizes from pedigrees. *Genetics Selection Evolution*, *40*(4), 359–378. <https://doi.org/10.1051/gse:2008008>
- Gutiérrez, J. P., & Goyache, F. (2005). A note on ENDOG: A computer program for analysing pedigree information. *Journal of Animal Breeding and Genetics = Zeitschrift Für Tierzüchtung Und Züchtungsbiologie*, *122*, 172–176. <https://doi.org/10.1111/j.1439-0388.2005.00512.x>
- Hasler, H., Flury, C., Menet, S., Haase, B., Leeb, T., Simianer, H., Poncet, P., & Rieder, S. (2011). Genetic diversity in an indigenous horse breed - Implications for mating strategies and the control of future inbreeding. *Journal of Animal Breeding and Genetics*, *128*(5), 394–406. <https://doi.org/10.1111/j.1439-0388.2011.00932.x>
- Hill, W. G. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genetical Research*, *38*(3), 209–216. <https://doi.org/10.1017/S0016672300020553>
- Howard, J. T., Pryce, J. E., Baes, C., & Maltecca, C. (2017). Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. *Journal of Dairy Science*, *100*(8), 6009–6024. <https://doi.org/10.3168/jds.2017-12787>
- Howrigan, D. P., Simonson, M. A., & Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC Genomics*, *12*(1), 460. <https://doi.org/10.1186/1471-2164-12-460>
- Janssens, S., Stinckens, A., Schroyen, M., Peeters, L., De Keyser, K., De Wael, R., Lamberigts, C., Luyten, T., Ons, E., & Buys, N. (2010). Genetic diversity in the Belgian Draught Horse breed as revealed by pedigree analysis and molecular marker data. *Animal Genetics*, *41*, 205–206. <https://doi.org/10.1111/j.1365-2052.2010.02114.x>
- Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, *6*, 208. <https://doi.org/10.3389/fgene.2015.00208>

- Kardos, M., Luikart, G., & Allendorf, F. W. (2015). Measuring individual inbreeding in the age of genomics: Marker-based measures are better than pedigrees. *Heredity*, *115*(1), 63–72. <https://doi.org/10.1038/hdy.2015.17>
- Maignel, L., Boichard, D., & Verrier, E. (1996). Genetic variability of French dairy breeds estimated from pedigree information. *Interbull Bull*, *14*, 49–54.
- Malécot, G. (1948). *Les mathématiques de l'Heredité*. Masson et Cie.
- Mancin, E., Ablondi, M., Mantovani, R., Pigozzi, G., Sabbioni, A., & Sartori, C. (2020). Genetic variability in the Italian heavy draught horse from pedigree data and genomic information. *Animals*, *10*(8), 1–19. <https://doi.org/10.3390/ani10081310>
- Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867–2873. <https://doi.org/10.1093/bioinformatics/btq559>
- McQuillan, R., Eklund, N., Pirastu, N., Kuningas, M., McEvoy, B. P., Esko, T., Corre, T., Davies, G., Kaakinen, M., Lyytikäinen, L.-P., Kristiansson, K., Havulinna, A. S., Gögele, M., Vitart, V., Tenesa, A., Aulchenko, Y., Hayward, C., Johansson, Å., Boban, M., ... Wilson, J. F. (2012). Evidence of inbreeding depression on human height. *PLoS Genetics*, *8*(7), e1002655. <https://doi.org/10.1371/journal.pgen.1002655>
- Meuwissen, T. H. E., & Luo, Z. (1992). Computing inbreeding coefficients in large populations. *Genetics Selection Evolution*, *24*(4), 305–313. <https://doi.org/10.1051/gse:19920402>
- Meyermans, R., Gorssen, W., Buys, N., & Janssens, S. (2020). How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*, *21*(1), 94. <https://doi.org/10.1186/s12864-020-6463-x>
- Perdomo-González, D. I., Sánchez-Guerrero, M. J., Molina, A., & Valera, M. (2020). Model comparison for estimation of genetic parameters in fertility of Pura Raza Española mares. Paper presented at the 71th Annual meeting of the European Federation of Animal Science, Porto, Portugal.
- Pjontek, J., Kadlečík, O., Kasarda, R., & Horný, M. (2012). Pedigree analysis in four Slovak endangered horse breeds. *Czech Journal of Animal Science*, *57*(2), 54–64. <https://doi.org/10.17221/5132-CJAS>
- Poyato-Bonilla, J., Perdomo-González, D. I., Sánchez-Guerrero, M. J., Varona, L., Molina, A., Casellas, J., & Valera, M. (2020). Genetic inbreeding depression load for morphological traits and defects in the Pura Raza Española horse. *Genetics Selection Evolution*, *52*(1). <https://doi.org/10.1186/s12711-020-00582-2>
- Poyato-Bonilla, J., Sánchez-Guerrero, M. J., Santos, R. D., & Valera, M. (2018). Population study of the Pura Raza Español Horse regarding its coat colour. *Annals of Animal Science*, *18*, 723–739. <https://doi.org/10.2478/aoas-2018-0016>
- R-Core-Team. (2021). R: A language and environment for statistical computing V4.1.0 "Camp Pontanezen". Retrieved from <https://www.R-project.org/>
- Ripollés, M., Sánchez-Guerrero, M. J., Perdomo-González, D. I., Azor, P., & Valera, M. (2020). Survey of risk factors and genetic characterization of ewe neck in a world population of Pura Raza Español Horses. *Animals*, *10*(10), 1789. <https://doi.org/10.3390/ani10101789>
- Sadeghi, R., Moradi-Shahrbabak, M., Miraei Ashtiani, S. R., Schlamp, F., Cosgrove, E. J., & Antczak, D. F. (2019). Genetic diversity of Persian Arabian horses and their relationship to other native Iranian horse breeds. *Journal of Heredity*, *110*(2), 173–182. <https://doi.org/10.1093/jhered/esy061>
- Sánchez, M. J., Azor, P. J., Molina, A., Parkin, T., Rivero, J. L. L., & Valera, M. (2016). Prevalence, risk factors and genetic parameters of cresty neck in Pura Raza Español horses. *Equine Veterinary Journal*, *49*(2), 196–200. <https://doi.org/10.1111/evj.12569>
- Sánchez-Guerrero, M. J., Solé, M., Azor, P. J., Sölkner, J., & Valera, M. (2019). Genetic and environmental risk factors for vitiligo and melanoma in Pura Raza Español horses. *Equine Veterinary Journal*, *51*(5), 606–611. <https://doi.org/10.1111/evj.13067>
- Schaefer, R. J., & McCue, M. E. (2020). Equine genotyping arrays. *Veterinary Clinics of North America - Equine Practice*, *36*(2), 183–193. <https://doi.org/10.1016/j.cveq.2020.03.001>
- Schaefer, R. J., Schubert, M., Bailey, E., Bannasch, D. L., Barrey, E., Bar-Gal, G. K., Brem, G., Brooks, S. A., Distl, O., Fries, R., Finno, C. J., Gerber, V., Haase, B., Jagannathan, V., Kalbfleisch, T., Leeb, T., Lindgren, G., Lopes, M. S., Mach, N., ... McCue, M. E. (2017). Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. *BMC Genomics*, *18*(1), <https://doi.org/10.1186/s12864-017-3943-8>
- Schurink, A., Shrestha, M., Eriksson, S., Bosse, M., Bovenhuis, H., Back, W., Johansson, A.M., & Ducro, B. J. (2019). The genomic makeup of nine horse populations sampled in the Netherlands. *Genes*, *10*(6), 480. <https://doi.org/10.3390/genes10060480>
- Solé, M., Ablondi, M., Binzer-Panchal, A., Velie, B. D., Hollfelder, N., Buys, N., Ducro, B. J., François, L., Janssens, S., Schurink, A., Viklund, S., Eriksson, S., Isaksson, A., Kultima, H., Mikko, S., & Lindgren, G. (2019). Inter- and intra-breed genome-wide copy number diversity in a large cohort of European equine breeds. *BMC Genomics*, *20*(1), 759. <https://doi.org/10.1186/s12864-019-6141-z>
- Solé, M., Valera, M., & Fernández, J. (2018). Genetic structure and connectivity analysis in a large domestic livestock meta-population: The case of the Pura Raza Español horses. *Journal of Animal Breeding and Genetics*, *135*(6), 460–471. <https://doi.org/10.1111/jbg.12352>
- Solé, M., Valera, M., Sánchez, M. J., Azor, P. J., & Fernández, J. (2019). Drawbacks and consequences of selective strategies in the design of semen banks: Case study of the Pura Raza Español horse breed. *Livestock Science*, *226*, 93–98. <https://doi.org/10.1016/j.livsci.2019.06.010>
- Somogyvári, E., Posta, J., & Mihók, S. (2018). Genetic analysis of the Hungarian population of endangered Hucul horses. *Czech Journal of Animal Science*, *63*, 237–246. <https://doi.org/10.17221/54/2017-CJAS>
- Thermo Fisher Scientific (2019). Axiom CNV Summary Tool user manual. Retrieved from https://tools.thermofisher.com/content/sfs/manuals/axiom_cnv_summary_tool_usermanual.pdf
- Valera, M. & Molina, A. (1995). Análisis Zoométrico de los Caballos de Pura Raza Española de Estirpe Cartujana. Paper presented at the 30th Jornadas luso-españolas de genética, Lisboa, Portugal.
- Valera, M., Molina, A., Gutiérrez, J., Gómez, J., & Goyache, F. (2005). Pedigree analysis in the Andalusian horse: Population structure, genetic variability and influence of the Carthusian strain. *Livestock Production Science*, *95*, 57–66. <https://doi.org/10.1016/j.livprodsci.2004.12.004>
- Valera, M., Molina, A., & Rodero, A. (1998). Índice de Conservación Genética En La Estirpe Cartujana Del Caballo de Pura Raza Española. *Archivos De Zootecnia*, *47*(178), 175–180.
- Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, *34*(3), 275–305. <https://doi.org/10.1051/gse:2002009>

- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer International Publishing.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wiener, P., Sánchez-Molano, E., Clements, D. N., Woolliams, J. A., Haskell, M. J., & Blott, S. C. (2017). Genomic data illuminates demography, genetic structure and selection of a popular dog breed. *BMC Genomics*, 18(1), 609. <https://doi.org/10.1186/s12864-017-3933-x>
- Zechner, P., Sölkner, J., Bodo, I., Druml, T., Baumung, R., Achmann, R., Marti, E., Habe, F., & Brem, G. (2002). Analysis of diversity and population structure in the Lipizzan horse breed based on pedigree information. *Livestock Production Science*, 77(2–3), 137–146. [https://doi.org/10.1016/s0301-6226\(02\)00079-9](https://doi.org/10.1016/s0301-6226(02)00079-9)
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24), 3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Poyato-Bonilla, J., Laseca, N., Demyda-Peyrás, S., Molina, A., & Valera, M. (2021). 500 years of breeding in the *Carthusian Strain* of Pura Raza Español horse: An evolutionary analysis using genealogical and genomic data. *Journal of Animal Breeding and Genetics*, 00, 1–16. <https://doi.org/10.1111/jbg.12641>