


RESEARCH

Open Access



Fine-tuning genomic and pedigree inbreeding rates in equine population with a deep and reliable stud book: the case of the Pura Raza Española horse

Davinia Isabel Perdomo-González^{1*} , Nora Laseca², Sebastián Demyda-Peyrás^{3,4}, Mercedes Valera¹, Isabel Cervantes⁵ and Antonio Molina²

Abstract

Background: Estimating inbreeding, which is omnipresent and inevitable in livestock populations, is a primary goal for management and animal breeding especially for those interested in mitigating the negative consequences of inbreeding. Inbreeding coefficients have been historically estimated by using pedigree information; however, over the last decade, genome-base inbreeding coefficients have come to the forefront in this field. The Pura Raza Española (PRE) horse is an autochthonous Spanish horse breed which has been recognised since 1912. The total PRE population (344,718 horses) was used to estimate Classical (F), Ballou's ancestral, Kalinowski's ancestral, Kalinowski's new and the ancestral history coefficient values. In addition, genotypic data from a selected population of 805 PRE individuals was used to determine the individual inbreeding coefficient using SNP-by-SNP-based techniques (methods of moments $-F_{\text{HOM}}$, the diagonal elements of the genomic $-F_{\text{G}}$, and hybrid matrixes $-F_{\text{H}}$) and ROH measures (F_{RZ}). The analyse of both pedigree and genomic based inbreeding coefficients in a large and robust population such as the PRE horse, with proven parenteral information for the last 40 years and a high degree of completeness (over 90% for the last 70 years) will allow us to understand PRE genetic variability better and the correlations between the estimations will give the data greater reliability.

Results: The mean values of the pedigree-based inbreeding coefficients ranged from 0.01 (F for the last 3 generations $-F_3$) to 0.44 (ancestral history coefficient) and the mean values of genomic-based inbreeding coefficients varied from 0.05 (F_{RZ} for three generations, F_{H} and F_{HOM}) to 0.11 (F_{RZ} for nine generations). Significant correlations were also found between pedigree and genomic inbreeding values, which ranged between 0.58 (F_3 with F_{HOM}) and 0.79 (F with F_{RZ}). In addition, the correlations between F_{RZ} estimated for the last 20 generations and the pedigree-based inbreeding highlight the fact that fewer generations of genomic data are required when comparing total inbreeding values, and the opposite when ancient values are calculated.

Conclusions: Ultimately, our results show that it is still useful to work with a deep and reliable pedigree in pedigree-based genetic studies with very large effective population sizes. Obtaining a satisfactory parameter will always be

*Correspondence: dperdomo@us.es

¹ Departamento Agronomía, Escuela Técnica Superior de Ingeniería Agromónica, Universidad de Sevilla, Ctra Utrera Km 1, 41013 Sevilla, Spain
Full list of author information is available at the end of the article



© The Author(s) 2022. **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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

desirable, but the approximation obtained with a robust pedigree will allow us to work more efficiently and economically than with massive genotyping.

Keywords: Genomic inbreeding, Horses, Pedigree inbreeding

Background

In populations of a genetically finite size, the mating of closely-related individuals is inevitable even in large populations, resulting in inbred offspring. Inbreeding, with the accompanying increase in the genome homozygosity, results in a reduction of genetic diversity and is often related with a phenomenon known as inbreeding depression, which is the reduction in performance per unit increase in inbreeding coefficient [1–3]. The magnitude of inbreeding depression often poses a considerable threat to the survival of inbred populations [4]. Although conservation geneticists and those responsible for genetic improvement plans are constantly looking for efficient strategies to overcome the negative consequences of inbreeding [5], as optimal contribution selection strategies [6], such strategies can not be implemented systematically in all populations. For this reason, it is important to obtain accurate estimates of inbreeding in order to permit a better management of the animal populations under selection.

Historically, in animal populations, inbreeding has been calculated from pedigree information and extended literature exists to estimate different pedigree-based inbreeding coefficients such as classical inbreeding [7, 8], ancestral inbreeding [9, 10], partial inbreeding coefficients [11, 12] and the ancestral history coefficient [13]. Nevertheless, as the cost of genotyping falls, genomic inbreeding coefficients can also be easily obtained. Genomic inbreeding coefficients are expected to be more accurate than pedigree-based coefficients because they do not depend on the quality and completeness of the pedigree. In addition, genomic inbreeding coefficients measure real homozygosity while pedigree-based inbreeding coefficients make average estimations.

While pedigree-based inbreeding coefficients use deterministic or stochastic methods to distinguish recent from ancestral inbreeding, genomic inbreeding attempts to detect the proportion of the genome covered by homozygous regions (runs of homozygosity, ROH) of a certain length, following the theory proposed by Fisher [14]. Knowledge about the proportion of recent and ancestral inbreeding is especially relevant in order to identify if an individual's ancestor has been also subjected to inbreeding. Animals with inbred ancestors are less susceptible to inbreeding depression than individuals with non-inbred ancestors, due to the purging effects associated with ancestral inbreeding. Therefore, looking

for signs of potential purging provides a better mirroring of the genetic load than simply measuring classical inbreeding [15].

In the same way as pedigree-based inbreeding coefficients, genomic inbreeding values can also be estimated with different metrics. The simplest SNP-based methods include the methods of moments technique [16] and inbreeding derived from the diagonal of the genomic-relationship matrix [17, 18]. More recently, the methodologies based on the estimation of ROH have become the state-of-art procedures. Among them, observational approaches, in which ROH are determined by a moving window of fixed size which scans each chromosome to determine the presence of a certain number of consecutive homozygous markers (implemented in PLINK [16]) are the most commonly used in livestock studies [19]. However, their reliability is lower in comparison with those based on the use of hidden Markov models [20], which estimates the probabilities of identity by descent at each marker of an individual [21]. Nowadays, it is still open to debate which methodology provides the most reliable and accurate inbreeding estimations, even more so when the number of studies that compare them with the pedigree records is still extremely low [22] and when different thresholds and constraint could be applied within each genomic estimating methodology [19].

The Pura Raza Española (PRE) is one of the oldest European horse breeds and the best known in the Iberian Peninsula, with over 250,000 active individuals. In addition, PRE horses are currently distributed in over 60 countries but managed by a single association, the Real Asociación Nacional de Criadores de Caballos de Pura Raza Española (ANCCE) which carries out breeding as a meta-population, a term which refers to a group of populations with some possible gene flow among them [23]. The PRE studbook was created in 1912, and from that moment, animals were only registered if at least two complete generations were known. Next, the studbook was completely closed and registration of new PRE horses was restricted to animals with parents already registered. Later, in the early 1980's, paternity controls were carried out in PRE horses using different molecular tools, such as blood grouping, serum biochemical polymorphism and DNA microsatellites [24–26]. As a result, the PRE population has over 40 years of proven parental information, as well as having a large and deep pedigree with a high degree of completeness [27], which has remained over 90% for

the last 70 years (the last 7 generational intervals). Since its studbook creation, both ANCCE and PRE breeders have been working to maintain the genetic variability of breeding. In fact, inbreeding increased exponentially up to 8.5% in the 1980s, and then decreased to the current figure of 7.5% due to measures to control inbreeding carried out in recent years.

In horses, populational genomic studies are still scarce in comparison with other livestock species. Nowadays, the availability of array-based SNP genotyping (from 65 K to 670 K markers per individual) in the species is increasing, allowing for the development of genomic tools and studies aimed at characterizing populations [28–30] and determining the genetic basis of traits [31], among others. However, despite the fact that the number has recently grown [32–34], ROH-based studies in large cohorts and populations of horses are still very scarce [35].

The aim of this paper was to analyse the evolution of 8 pedigree and 4 different genomic-based inbreeding estimations in a large cohort of nearly 300,000 horses with a large and robust pedigree. In addition, we determined the correlations among all the estimations (pedigree and genomic based) using partial correlations in order to establish the reliability of different inbreeding estimations in horses.

Methods

Dataset

In this study, the genealogical information of all individuals registered in the PRE horse studbook was analysed, in which 97.97% (337,712 horses) have at least 3 complete generations. The total PRE population is composed of 344,718 horses (168,301 males and 176,417 females), born from the end of the nineteenth century to 2020. The mean of equivalent generations was 9.46 and the average numbers of full and maximal generations traced were 5.66 and 17.17, respectively. Animals with no known parents in the pedigree data were considered as founders and assumed to be unrelated.

Pedigree-based inbreeding coefficients

Pedigree-based inbreeding coefficients were computed using Endog [36] and GRain [13] programmes. The classical inbreeding coefficient (F) according to Wright [7] is defined as the probability that the two alleles at any locus in an individual are identical by descent (IBD). In addition, depending on the number of generations taken into account, inbreeding coefficients at the 3rd, 6th and 9th generations can be computed (F_3 , F_6 and F_9 , respectively). Otherwise, ancestral inbreeding coefficients were calculated following the approaches by Balou [9] (F_a _Bal) and Kalinowski et al. [10] (F_a _Kal). While F_a _Bal is defined as the probability that any allele in an

individual has been autozygous (IBD) in previous generations at least once, F_a _Kal is defined as the probability that any allele in an individual is currently IBD and has been IBD in previous generations at least once. At the same time, the Kalinowski approach allows us to split F into two parts: alleles which have undergone inbreeding in the past (F_a _Kal) and alleles IBD which have done so for the first time (F_{new_Kal}). On the other hand, partial inbreeding coefficients according to Lacy et al. [11, 12] (F_{ij}) examine whether alleles contributing to inbreeding have been distributed uniformly across founder genomes or from specific founders. Finally, the ancestral history coefficient, developed by Baumung et al. [13], has been defined as the number that tells us how many times, during the pedigree segregation, a randomly-taken allele has had IBD status. The idea behind distinguish recent and ancestral inbreeding coefficients is that alleles which have experienced inbreeding more often in the past are less likely to be deleterious than alleles which have undergone IBD less often.

SNP genotyping

Genotypic data from 805 PRE individuals was analysed, whose selection was based on a low average relatedness from 365 studs. To do this, blood samples of each individual were obtained by jugular venepuncture using sterile tubes with EDTA. Next, DNA was obtained from 200 μ L of whole blood using commercial kits following the manufacturer's instructions.

The genotypes of 670,776 SNPs markers were determined for each individual using the HD Axiom™ Equine SNP Genotyping Array (ThermoFisher, Madrid, Spain). The raw data (.CEL files) were first analysed following the “best genotyping practices” workflow in the Axiom Analysis Suite 5.0 software [37]. All the samples passed the genotyping quality threshold (dish quality check ≥ 0.82 and plate call rate ≥ 0.97). However, only 540,294 SNPs markers (located in 31 chromosomes) showing a high-quality genotyping rate (SNP call rate $> 95\%$ and Fisher's Linear Discriminant parameter > 3.6) were kept for inbreeding analysis. No minor allele frequency or linkage disequilibrium filtering was performed, following the latest ROH estimation guidelines [19].

Genomic-based inbreeding coefficients

Four different approaches were employed to determine the individual inbreeding coefficient using genomic data. First, we carried out a multistep methodology based on a hidden Markov model (HMM) framework developed by Druet and Gautier [21], which take into account the allele frequency of each SNP, its genetic position, and the genotyping error rate. All the ROH per individual were determined using a 7-class model ($K=6, 12, 18,$

36, 72, 144, and 144), with a *mixing coefficient* = 0.01, and a *genotyping error probability* = 0.001, implemented in the RZooROH R package [38]. Thereafter, the individual inbreeding value, F_{RZ} , was estimated as the relationship between the length of the genome covered by ROH (defined as continuous homozygous stretches with a minimum length of one megabase in the genome) and the total genome length of the genome fragments, as proposed by McQuillan et al. [39] using the *summaryRuns* function of *DetectRUNS* R package [40]. In this estimation, only ROH longer than 1 Mb were taking into account to avoid detecting ROH IBD segments [19].

Inbreeding was also calculated as the individual autozygosity (diagonal) of the genomic (F_G) and hybrid (F_H) matrixes proposed by VanRaden [17] and Martini et al. [18], respectively. The matrixes were constructed using only the genomic data (G) and a combination of genomic data and pedigree records available (H) using the AGHMatrix R package [41]. Finally, the method-of-moments methodology was applied to estimate F_{HOM} , the inbreeding coefficient as the difference between the observed and the expected homozygotes counts per individual. The calculations were made using the *--het* option in PLINK v1.90 [16].

In addition, genomic inbreeding value explained by matings occurred during the last 20 generations was determined following the theoretical approach proposed by Fisher [14] to analyse the theoretical increase of individual inbreeding across the last 20 generations. To do this, all the ROH fragments detected per individual using the RZooROH package were analysed, and 20 iterative runs were performed in which only ROH with a minimum length of $1/2g$ Morgans (where g is the number of generations theoretically elapsed since the inbreeding event) were retained in the inbreeding calculation. This produced 19 additional inbreeding values (F_{RZG2} to F_{RZG20}) per individual.

Statistical analysis

Mean values for the different pedigree-based inbreeding coefficient estimations for the total PRE population and the genotyped individuals were calculated. The total PRE population was divided according to each generational interval (each generational interval comprising 10 years, [42]) to analyse the evolution of founders and common ancestors and the partial inbreeding coefficients (F_{ij}) transmitted by them. While the effective number of founders (f_e) defines the number of equally contributing founders that would be expected to produce the same genetic diversity as in the total population, the effective number of ancestors (f_a) refers to the minimum number

of ancestors, not necessarily founders, which account for the complete genetic diversity, as in the total population.

In addition, for every trio of inbreeding coefficients, x , y and z , the three first-order partial correlation coefficients were computed to estimate the correlation between each pair of parameters with the third estimated variables fixed (partial correlation). The partial correlation coefficient between x and y given z indicates the strength of the linear relationship between x and y , that is, independent of, and uncorrelated with, z . The comparison with the ordinary (or unconditional or zero-order) correlation coefficient, allows us to determine whether the association between the two inbreeding coefficients has been sharply reduced after eliminating the effect of the third inbreeding coefficient. For every trio of inbreeding coefficients, in order to obtain the tolerance level (ϵ) to be used as the local threshold for determining significant associations, the mean ratio of partial to direct correlation was calculated according to Reverter and Chan [43]:

$$\mathcal{E} = \frac{1}{3} \left(\frac{r_{xy,z}}{r_{xy}} + \frac{r_{xz,y}}{r_{xz}} + \frac{r_{yz,x}}{r_{yz}} \right)$$

with $r_{xy,z}$, $r_{xz,y}$ and $r_{yz,x}$ the three partial correlations, and r_{xy} , r_{xz} and r_{yz} the ordinary correlations.

A correlation between inbreeding coefficients x and y is discarded if:

$$|r_{xy}| \leq |\mathcal{E} r_{xz}| \quad \text{and} \quad |r_{xy}| \leq |\mathcal{E} r_{yz}|$$

Otherwise, the association is defined as significant.

This procedure was extended to all the coefficient parameters different from x and y to determine which correlations exceeded the estimated thresholds and could be considered non-spurious relationships [44]. The first-order partial correlation coefficients together with the thresholds for determining significant associations were calculated with the software *PCIT* package in R [45].

Data handling

Management of the pedigree record and molecular dataset was performed entirely in the R environment using the following packages: *dplyr* [46], *tibble* and *tidyr* from *tidyverse* [47], and *data.table* [48]. Data visualization was conducted in *ggplot2* [49].

Results

Population structure

The number of individuals born in each generational interval (Table 1) has increased since the year 1950. The average F increased to 8.4% in 1980–1989, then started to decrease, a trend which continues to date (7.28%). The number of founders decreased until the generational interval of

Table 1 Inbreeding, founders, ancestors, and common ancestors of Pura Raza Española horse population by generational intervals

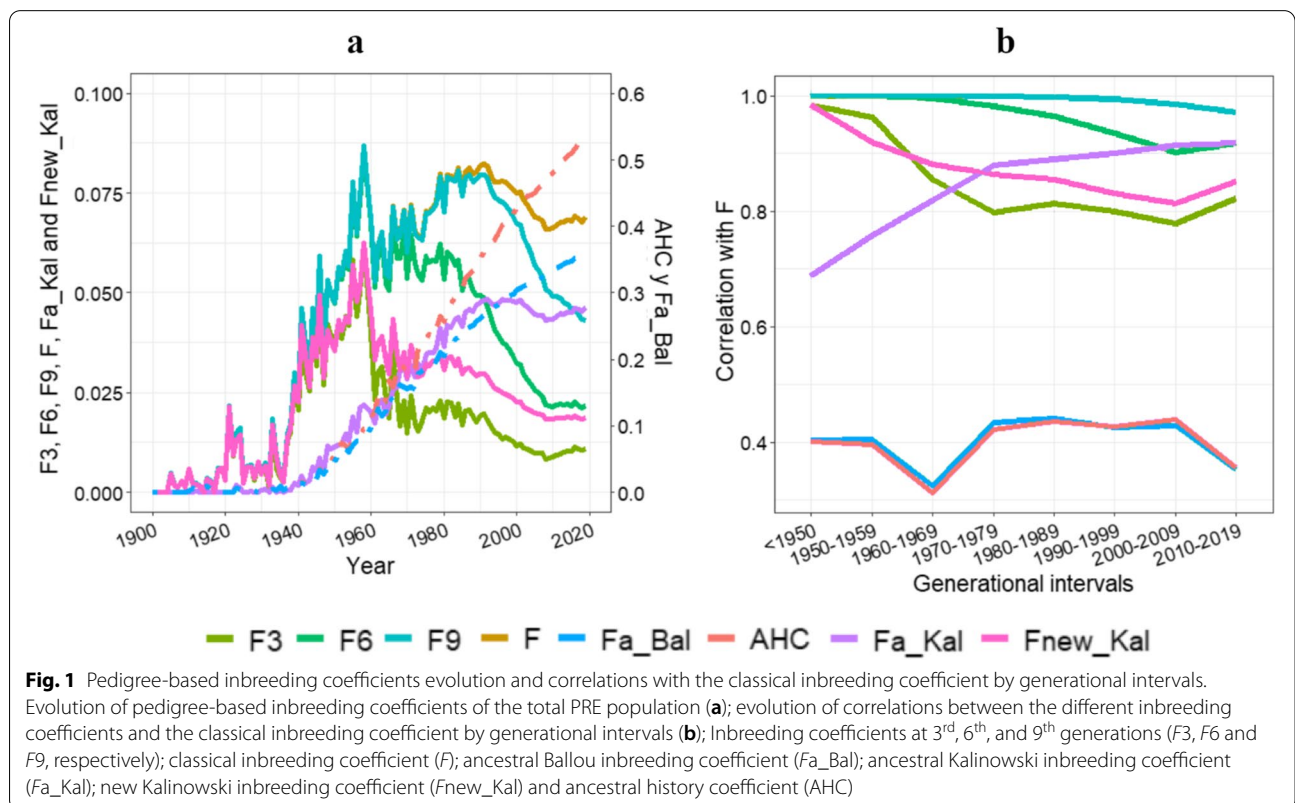
GI	N	F, %	Funders	NEF	Ancestors	fe	fa	fe/fa	MCA [Min, Max]	F_{ij} %
1950–1959	1818	6.87	410	330	350	58	33	1.75	18.29 [1, 63]	0.73
1960–1969	2357	6.73	407	323	348	44	26	1.69	28.45 [1, 132]	0.26
1970–1979	6046	7.48	387	309	344	34	20	1.70	49 [1, 173]	0.17
1980–1989	17,091	8.40	394	317	351	30	18	1.66	81.83 [1, 248]	0.12
1990–1999	53,102	8.32	424	337	394	29	17	1.70	124.76 [1, 336]	0.08
2000–2009	150,129	7.45	411	324	383	32	19	1.68	183.36 [26, 521]	0.05
2010–2019	109,062	7.28	339	315	379	34	19	1.78	242.40 [34, 771]	0.03

GI Generational interval, N number of individuals born, F mean classical inbreeding, NEF number of equivalent founders, fe effective number of founders, fa effective number of ancestors, fe/fa ratio between fe and fa, MCA mean common ancestors, Min minimum number of common ancestors, Max maximum number of common ancestors, F_{ij} mean partial inbreeding coefficient

1990–1999, when there was a slight increase, followed by a decrease in the next generations. Similar results can be seen for ancestors and equivalent founders. Nevertheless, while the number of founders, equivalent founders and ancestors have remained relatively constant over the history, the number of effective founders and effective ancestors have decreased to almost half. The numbers of effective founders and ancestors have oscillated from 58 to 34 and from 33 to 19, respectively, over the last 70 years, while the fe/fa ratio for each generational interval has varied between 1.66 at the 1980–1989 generational interval to 1.78 at the 2010–2019 generational interval. Finally, the

average common ancestors (those in both, maternal and paternal lineage) of individuals in each generational interval has increased from 18.29 in the 1950–1959 interval (transmitting an average F_{ij} of 0.73%) to 242.40 in the present day (transmitting an average F_{ij} of 0.03%).

The evolution of the different pedigree-based inbreeding coefficients over the years can be seen in Fig. 1a. All the pedigree-based inbreeding coefficients started to increase around 1940. While AHC and Fa_Bal increased exponentially to 0.541 and 0.367 by 2019, the other coefficients increased more rapidly around the 1960s but then has started to decrease in recent years, with some values



appearing constant. In 2019, F_3 , F_6 and F_{new_Kal} showed similar values, 0.011, 0.023 and 0.022, respectively, while F_9 and Fa_Kal also showed similar values between them, 0.047 and 0.051, respectively, and finally, F was 0.074.

The correlations between F and the other pedigree-based inbreeding coefficients (Fig. 1b) reveal differences over the generational intervals. The correlation values for Fa_Bal and AHC seem to be the lowest, at around 0.4, while those for the other coefficients were over 0.8 for most of the intervals. F_9 and F_6 had the most constant coefficients always close to 1 while the Fa_Kal correlation values increased from 0.69 to 0.91. However, F_{new_Kal} and F_3 correlation values have fallen over the last 50 years, at nearly 0.8.

Pedigree-based inbreeding estimations

The different pedigree-based inbreeding coefficients mean values and Pearson's correlations between them for the total PRE population and genotyped individuals can be seen at Table 2. The mean values ranged between 0.01 (F_3) and 0.08 (F) for most pedigree-based inbreeding coefficients, except for Fa_Bal (0.31) and AHC (0.44). Most of the Pearson's correlations between pedigree-based inbreeding coefficients were significant, positive and with moderate to high value. Similar correlation values can be seen between inbreeding coefficients for the total population and for genotyped individuals. The highest positive significant values were found between F and F_9 (0.98 and 0.99 for the total population and genotyped individuals, respectively) and between Fa_Bal and AHC (0.98 and 0.99 for total population and genotyped individuals, respectively). The lowest positive significant values were found between Fa_Bal and F (0.45) and between Fa_bal and F_9 (0.36), both for genotyped animals. There were no negative and significant correlations between pedigree-based inbreeding coefficients.

Genomic-based inbreeding estimations

The average results of the 4 different genomic inbreeding coefficients, Pearson's correlations within the genomic-based inbreeding coefficients and the Pearson's correlations between the genomic and pedigree inbreeding coefficients are shown at Table 3. The average values show clear differences based on the methodology employed. Within the ROH-based methodology (F_{RZ}), the average values were 0.05, 0.09 and 0.11 for F_{RZ3} , F_{RZ6} , F_{RZ9} , respectively, while for the total generations (F_{RZ}), it was 0.17. In comparison, the matrixial and method-of-moments based methods showed low and very similar values (0.05, 0.06 and 0.05 in F_H , F_G and F_{HOM} , respectively).

The Pearson's correlations within the genomic-based inbreeding coefficients are mostly, positive and of a high magnitude. F_{RZ} , F_{RZ3} , F_{RZ6} and F_{RZ9} showed a significant, high correlation between each other and with F_{HOM} , ranging between 0.84 ($F_{RZ3}-F_{HOM}$) and 0.99 ($F_{RZ6}-F_{RZ9}$). On the other hand, while F_G showed high and significant correlation values with F_{RZ3} (0.67) and F_{RZ6} (0.56), the other correlations between F_G and F_H and the other of the genomic estimations were lower and not significant. Interestingly, the correlation between both matrix-based methodologies (F_H and F_G) was high and significant (0.83).

Correlations between pedigree and genomic based inbreeding estimations

The Pearson's correlations between the genomic and pedigree-based inbreeding coefficients (right part of Table 3) were mostly significant, positive and of a high magnitude, ranging between 0.58 (F_3-F_{HOM}) and 0.79 ($F-F_{RZ}$). The classical pedigree-based estimation (F) showed the best fit with the non-matrixial genomic estimations among all the pedigree-based F values, including 0.79 for $F-F_{RZ}$ and 0.78 for $F-F_{HOM}$, $F-F_{RZ6G}$ and $F-F_{RZ9G}$. On the contrary, F_G and F_H showed

Table 2 Pearson's correlations between pedigree-based inbreeding coefficients for genotyped individuals and total population

	Mean	F_3	F_6	F_9	F	Fa_Bal	AHC	Fa_Kal	F_{new_Kal}
F_3	0.01		0.93 ^a	0.81 ^a	0.79 ^a	-0.07	-0.10	0.45	0.96 ^a
F_6	0.03	0.89 ^a		0.93 ^a	0.90 ^a	0.05	0.01	0.62	0.95 ^a
F_9	0.06	0.81 ^a	0.95 ^a		0.99 ^a	0.36 ^a	0.33	0.85 ^a	0.84 ^a
F	0.07	0.79 ^a	0.91 ^a	0.98 ^a		0.45 ^a	0.42	0.89 ^a	0.81 ^a
Fa_Bal	0.31	-0.01	0.02	0.22	0.36		0.99 ^a	0.76 ^a	-0.10
AHC	0.44	-0.02	0.01	0.22	0.36	0.98 ^a		0.74 ^a	-0.13
Fa_Kal	0.04	0.52	0.68	0.85 ^a	0.90 ^a	0.60 ^a	0.63 ^a		0.46
F_{new_Kal}	0.02	0.92 ^a	0.94 ^a	0.87 ^a	0.84 ^a	-0.07	-0.08	0.53	

Mean values are for different pedigree-based inbreeding coefficient estimations of the total PRE population. Over the diagonal: Pearson's correlations between pedigree-based inbreeding coefficient for genotyped individuals; under the diagonal: Pearson's correlations between pedigree-based inbreeding coefficient for total PRE population.^aSignificant according to Watson-Haigh et al. [45]. Standard mean errors were lower than 0.0005. Inbreeding coefficients at 3rd, 6th, and 9th generations (F_3 , F_6 and F_9 , respectively), classical inbreeding coefficient (F), ancestral Ballou inbreeding coefficient (Fa_Bal), ancestral history coefficient (AHC), ancestral Kalinowski inbreeding coefficient (Fa_Kal) and new Kalinowski inbreeding coefficient (F_{new_Kal})

Table 3 Pearson's correlations within genomic and between genomic and pedigree-based inbreeding coefficient estimations for genotyped animals

	Genomic inbreeding						Genealogical inbreeding								
	Mean	F_{RZ3}	F_{RZ6}	F_{RZ9}	F_G	F_H	F_{HOM}	$F3$	$F6$	$F9$	F	Fa_Bal	AHC	Fa_Kal	$Fnew_Kal$
F_{RZ}	0.17	0.85 ^a	0.94 ^a	0.97 ^a	0.41	0.26	0.98 ^a	0.59 ^a	0.68 ^a	0.78 ^a	0.79 ^a	0.4	0.38	0.72 ^a	0.61 ^a
F_{RZ3}	0.05		0.94 ^a	0.91 ^a	0.67 ^a	0.52	0.84 ^a	0.75 ^a	0.77 ^a	0.75 ^a	0.75 ^a	0.15	0.13	0.55	0.75 ^a
F_{RZ6}	0.09			0.99 ^a	0.56 ^a	0.39	0.93 ^a	0.68 ^a	0.74 ^a	0.78 ^a	0.78 ^a	0.28	0.25	0.65 ^a	0.70 ^a
F_{RZ9}	0.11				0.51	0.34	0.95 ^a	0.64 ^a	0.72 ^a	0.78 ^a	0.78 ^a	0.32	0.29	0.68 ^a	0.66 ^a
F_G	0.06					0.83 ^a	0.42	0.74 ^a	0.68 ^a	0.45	0.46	-0.31	-0.32	0.13	0.73 ^a
F_H	0.05						0.26	0.79 ^a	0.71 ^a	0.49	0.42	-0.43	-0.46	0.03	0.77 ^a
F_{HOM}	0.05							0.58 ^a	0.67 ^a	0.76 ^a	0.78 ^a	0.39	0.37	0.71 ^a	0.60 ^a

Mean values are for different genomic-based inbreeding coefficient estimations of the PRE genotyped animals; Pearson correlations within genomic inbreeding coefficients (left part) and between genomic and pedigree-based inbreeding coefficient estimations (right part).^aSignificant according to Watson-Haigh et al. [45]. Standard means errors were lower than 0.0005. ZooROH approach genomic inbreeding coefficient for all generations and 3, 6, and 9 generations (F_{RZ} , F_{RZ3} , F_{RZ6} , F_{RZ9} , respectively), genomic inbreeding based on the autozygosity (the matrix diagonal) using genomic data (F_G , matrix G according to VanRaden [17]), and genomic and pedigree data (F_H , matrix H according to Martini et al. [18]), inbreeding coefficient using Plink (F_{HOM}), inbreeding coefficients at 3rd, 6th, and 9th generations ($F3$, $F6$ and $F9$, respectively), classical inbreeding coefficient (F), ancestral Ballou inbreeding coefficient (Fa_Bal), ancestral history coefficient (AHC), ancestral Kalinowski inbreeding coefficient (Fa_Kal) and new Kalinowski inbreeding coefficient ($Fnew_Kal$)

significant correlations with pedigree data only when the information of 3 or 6 generations was employed in the estimations. The pedigree-based AHC and Fa_Bal showed low, non-significant correlations with all the pedigree and genomic estimations, showing negative values with F_H and F_G . Finally, both Kalinowski estimations showed high and significant correlations with the rest of the genomic and pedigree based values. However, the ancestral Kalinowski estimation showed low correlations with both matrixial estimators (F_G and F_H).

The correlations between the genomic inbreeding values estimated for the last 20 generations (from generation two, F_{RZ-2G} , to generation 20, F_{RZ-20G}) and the pedigree-based inbreeding estimations are shown in Fig. 2. Partial correlation values with $F3$, $F6$, $F9$ and F showed a similar pattern, in which correlations increased up to a certain point and then started to decrease with the increase of generations considered. Interestingly, the best correlation for $F3$ was observed using genomic data from only 3 generations (F_{RZ-3G}). Similarly, the best correlation for $F6$ was observed using 4 generations of genomic data (F_{RZ-4G}) and the best correlation for $F9$ was observed using 6 generations of genomic data (F_{RZ-6G}). In contrast, Fa_Kal showed an increasing correlation value across the generations, with the best correlation at F_{RZ-20G} , whereas Fa_Bal and AHC showed the lowest correlation values with F_{RZ} of all the analyses performed.

Discussion

Inbreeding estimations based on pedigree information can help us to understand the population structure and trends in inbreeding in each breed, and has been

historically used to manage populations and control inbreeding depression within populations. Nevertheless, it is not enough to know the population inbreeding level, as it does not exactly reflect the real degree of genome homozygosity, because inbreeding is not directly related to inbreeding depression, since natural and artificial selection over time may have meant that deleterious recessive alleles are purged or advantageous recessive alleles are fixed in a population [50–53]. Here, the distinction between ancient and recent inbreeding is also a highly relevant factor, since inbreeding arising from a distant common ancestor should have less effect on fitness than inbreeding from recent common ancestors [54]. For this reason, different methods of pedigree-based inbreeding estimations have been developed. However, there have been few studies in which they have been compared [44, 55–58] and none have analyzed all the different methods together in a large population with a deep, well-established, complete pedigree. In these sense, the PRE is an exceptional case in which to analyse the different methodologies of pedigree-based inbreeding estimations and compare them with genomic-based parameters where the 98.7% of the total PRE population is inbred (with an F value higher than 0) [27].

The PRE population census has undergone uneven progress during the generational interval analysed (Table 1). The number of horses born in latest generational interval is over one hundred thousand times that in the 1950s. The fe/fa ratio (which describe the unbalanced representation of the founder contributions) for each generational interval found in this study has undergone major changes in the genetic management of the population and demonstrated that different genetic

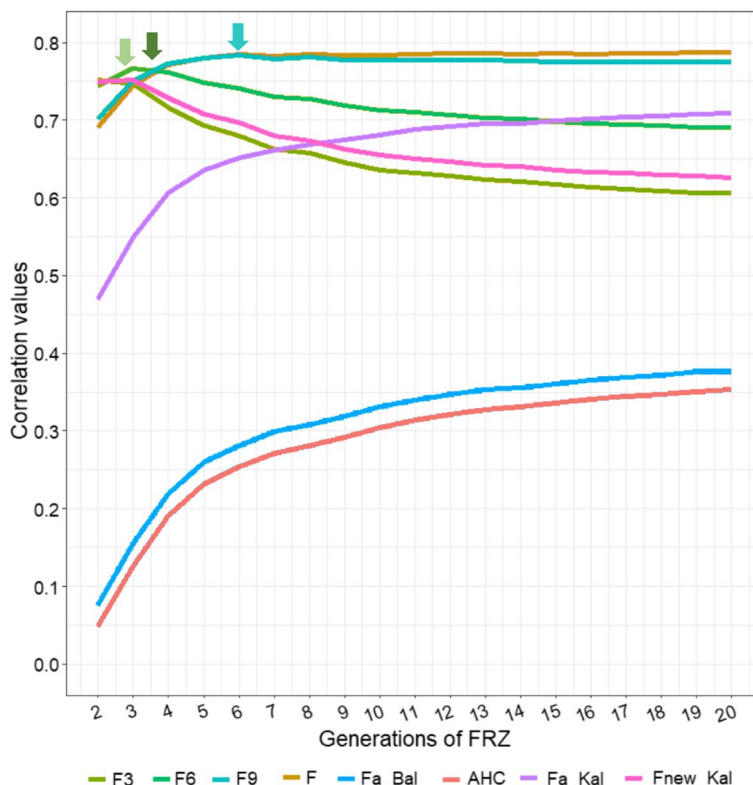


Fig. 2 Pearson's correlations between pedigree-based and F_{RZ} inbreeding for the last 20 generations. The ZooROH approach genomic inbreeding coefficient (F_{RZ}) was determined, which had elapsed since the hypothetical inbred mating in the PRE population; Inbreeding coefficients at 3rd, 6th, and 9th generations (F_3 , F_6 and F_9 , respectively); classical inbreeding coefficient (F); ancestral Ballou inbreeding coefficient (Fa_{Bal}); ancestral Kalinowski inbreeding coefficient (Fa_{Kal}); new Kalinowski inbreeding coefficient ($Fnew_{Kal}$) and ancestral history coefficient (AHC). The narrows depict the best correlations obtained with F_3 , F_6 and F_9

bottlenecks have taken place along the history of the PRE horse. The bottleneck phenomenon has also been seen in other horse breeds, such as the Old Kladruber horse [59], the Spanish Arab Horse [60] and the Campolina horse population [61] with an fe/fa ratio of 5.40, 2.03, and 1.51, respectively, although all of these have a much smaller clearly census. On the other hand, the increase in the average number of common ancestors and the decrease in the mean partial inbreeding coefficient that those common ancestors transmit to their descendants agrees with the latest studies of the genetic structure of the PRE population [27] and reflects a problem which is increasingly worrying of breeders, the selection of unrelated animals for breeding, instead of leading to an increase in genetic diversity in the PRE population, results in a loss.

Table 2 shows the mean values of the pedigree-based inbreeding coefficient of the total PRE population and, as expected, when different pedigree-based inbreeding coefficients are compared, the average values increase from recent to ancient inbreeding estimates. The total

population average Fa_{Bal} value (0.31) was seven times higher than the total population average Fa_{Kal} value (0.04). By definition, Fa_{Bal} is a value which tells us which individuals or populations possess fewer detrimental genes. It therefore follows that, the higher the value of Fa_{Bal} , the lower the probability of having detrimental genes [57]. As can be seen in Fig. 1, average Fa_{Bal} values have been increasing from the 1940s and we can safely say that on average, the PRE population is prone to limiting incidents of inbreeding depression as the generations progress. Studies analysing the differences between Fa_{Bal} and Fa_{Kal} have also shown higher values for Fa_{Bal} as in the case of the German Anger and the Red-and-White cattle breeds [57], and a crossbred rabbits population [44]. Nevertheless, Suwanlee et al. [62] found that Ballou's formula overestimate the real proportion of alleles within a genome which has undergone inbreeding by stochastic simulations of different settings for population size and initial allele frequencies, even though the overestimation seems to be more pronounced in small populations. Similar behaviour can be observed

in Fig. 1 for AHC for each generational interval, the average value has increased to the current value of 0.51. Our results show that, according to the total population average value (Table 2), an allele at random gene of a random individual have been IBD 0.44 times. In this sense, it has been shown in a population of Australian Thoroughbred horses [63], with a smaller pedigree (257,249 horses), but with an F value (0.139) and a mean equivalent generation (24.6) higher than the PRE population, and an AHC value of 1.973, that this coefficient best captures the effectiveness of selective breeding practices in increasing the frequency of favourable alleles and the purging of highly and mildly deleterious alleles. In the populations of German Anger and the Red-and-White cattle breeds studied [57], AHC and Fa_Bal values were also the highest inbreeding coefficient values and had very similar values between them, with an AHC value of 3.94 and a Fa_Bal value of 3.69 for German Anger and an AHC value of 1.49 and a Fa_Bal value of 1.39 for the Red-and-White breeds, while in the crossbred rabbit population [44], the AHC value was three times higher than Fa_Bal , 2.72 and 0.85, respectively.

Pearson's correlations between pedigree-based inbreeding coefficients (Table 2) were similar for genotyped individuals to that of the total population, which reveals the representativeness of the sample selected for genotyping. While positive and significant correlations were detected between classical inbreeding coefficients and classical inbreeding coefficients in different generations ($F3$, $F6$ and $F9$), no significant correlations were found between them with Fa_Bal and AHC. On the other hand, high positive significant correlations were found here between Fa_Kal with Fa_Bal (0.60 for total population and 0.76 genotyped individuals) and AHC (0.63 and 0.74), being lower than those obtained in Thoroughbred horses [63], 0.90 Fa_Kal - Fa_Bal , and 0.85 Fa_Kal -AHC, but similar to those in a combined pedigree of the German Anger and the Red-and-White cattle breeds populations [57], 0.63 Fa_Kal - Fa_Bal and 0.65 Fa_Kal -AHC. Nevertheless, both Mc Parland et al. [64] in an Irish Holstein-Friesian cattle population and Schäler et al. [55] in the Angler saddleback pig population reported weak correlation values between them, while Rodríguez-Ramilo et al. [44] reported no significant correlations in crossbred rabbits. At the same time, high, positive significant correlations were found here between Fa_Kal with $F9$ (0.85) and F (0.90), which is in line with both Rodríguez-Ramilo et al. [44] and Schäler et al. [55]. The correlation between Fa_Bal and AHC (0.98) was also positive, strong and significant in Rodríguez-Ramilo et al. [44], Addo et al. [57] and Todd et al. [63]. Finally, $Fnew_Kal$ had high, positive correlations with $F3$ (0.92), $F6$ (0.94), $F9$ (0.87), and F (0.84) but no significant correlations with ancestral

coefficients (Fa_Bal , AHC, and Fa_Kal). Ultimately, the high inbreeding coefficients on PRE population are more related to ancestral inbreeding than to recent inbreeding, which seem to be under control, and thus implies being less prone to inbreeding depression due to purging.

Although the classical inbreeding coefficient has often been considered as the best measure of population inbreeding, it may be an unrealistic measure of individual IBD. The variability between real levels of autozygosity and probability-based estimations can be due to recombination (with a stochastic nature) and a change on allele frequencies due to selection, in addition to the fact that pedigree-based inbreeding coefficients depend on the completeness of studbook and the reliability of the documentation [65]. Many of the problems evidenced in pedigree-based inbreeding can be overcome by using different methods based on analysis of the genomic data, which has been stated as a more precise estimate of IBD [65, 66]. Despite that studies analysing the correlation between pedigree and genomic-based inbreeding coefficients have shown strong correlations between genomic and pedigree-based inbreeding coefficients in human [39], cattle [67–69], rabbit [44] or pig populations [55], the data available in horse is still scarce.

An unresolved question is still how to determine the most reliable methodology to estimate genomic-based inbreeding values among the current alternatives available since comparison of their results can be contradictory due to diverse factors [22]. For instance, recent studies have reported that ROH-based F measures (F_{RZ} and F_{ROH}) are more powerful in detecting inbreeding depression than SNP-by-SNP-based F measures (F_G , F_H and F_{HOM}) [65, 70], but even more when using information from medium-density arrays since the accuracy of the later strongly relies on the number of markers per individual [71]. Moreover, Caballero et al. [72] indicated that the use of F_G (named F^I in that study), and F_{HOM} (to a lesser extent), was not advised to estimate inbreeding depression; as well as Villanueva et al. [73] reported that matrixes have proven to be very effective in increasing the accuracy of genomic predictions, however, they do not always provide a useful measure of inbreeding. Similarly, Ceballos et al. [20] found that methodologies based on statistical models that estimate probabilities of IBD at each marker such as hidden Markov models were more reliable and robust since they take into account more factors to estimate inbreeding, such as marker allele frequencies, genetic distances, genotyping error rates and the sequences of observed genotypes. In addition, hidden Markov model-based methodologies allow to provide a better fit for individual genetic data and to refine the genomic partitioning of inbreeding into stretches of IBD segments from possibly different ancestral origins

[21]. In this sense, a recent comprehensive meta-analysis [22] which compared the performance of genomic methodologies employed in this study (F_{RZ} , F_{HOM} , F_G and F_H) suggested that hidden Markov model-based F estimations (such as F_{RZ}) seems to be the more reliable. Same results were obtained in this study, when they were compared with the classical inbreeding coefficient F , but also with several different pedigree-based inbreeding estimators. However, the same authors mentioned that the size and genetic structure of the population evaluated must be taken into account to consider one methodology better than the other, suggesting that the best fitting genomic estimates for a population of a relatively small size (<1000) are those based on ROH (F_{RZ}), while those based on SNP by SNP analysis (F_{HOM} , F_G and F_H) are better for large populations (10,000) [74]. This hypothesis agrees with our results.

The correlations among different genomics F measures showed that matrixial methods (F_H , and to a lesser extent F_G) are less related and showed lower average values in comparison with the rest of the genomic measures. On the contrary, F_{HOM} and F_{RZ} showed the highest correlation (0.98), in agreement with the reported by Solé et al. [75] in which the correlations between hidden Markov model-based estimates and measures based on homozygosity and ROH were extremely high (F_{HOM} , $r=0.95$ and F_{ROH} , $r=0.95$, respectively). Moreover, they found that the correlations between the model based on the matrix proposed by VanRaden [17] and F_{HOM} and F_{RZ} were lower, agreeing with our results. Similar results were provided by Caballero et al. [72], which reported the lowest correlation between the matrix-based method (F^I) and hidden Markov model based methodologies. In addition, we obtained a correlation of 0.42 between F_G and F_{HOM} higher than obtained by Villanueva et al. [73], whose values were 0.25 and 0.28 for two independent datasets. Our results support the previous data suggesting that the estimations of inbreeding values obtained from matrixial methods should be used with caution, since both methods tends to underestimate the existence of ancestral inbreeding, and therefore, the F value. In a recent study, Meyermans et al. [19] used the RZooRoH (F_{RZ}) model based approach to validate their study of runs of homozygosity using PLINK. They obtained a high Pearson correlation of individual F_{ROH} between PLINK and RZooRoH (F_{RZ}) (0.89–0.99) which was similar to our previous results (0.95) [76]. Therefore, in agreement with above results, we believe it is a robust methodology for estimating genomic inbreeding.

At the present day, genomic estimates of F are considered more reliable and accurate than the pedigree-based for two main reasons. First, pedigree based estimations depend on the quality of the pedigree data available,

which sometimes is scarce and/or not entirely reliable [71]. But also, pedigree-based estimations assume an equal distribution of the “inbreeding” across the entire genome as well as a proportional passage of IBD alleles from the common ancestor, which has been proven as unrealistic in several species [33, 69, 71, 77]. However, the validation of this fact in a livestock population with a deep and robust pedigree is still scarce. In our study, the genealogical estimations were made by using the Pura Raza Española studbook pedigree, which deepness and reliability is largely proven.

Our results showed high correlations between classical pedigree-based estimation (F) and genomic estimations F_{RZ} (0.79) and F_{HOM} (0.78) in comparison with recent studies in other breeds such as the Mangalarga Marchador horse (0.02) [33]. However, those differences are more likely to be produced by the fact that Mangalarga pedigree included only 4 equivalent generations, in comparison with the 9.46 (and 17.17 maximal generations in average) of the Pura Raza Española horses. This effect produced by the incompleteness of pedigree information has been previously pointed as a serious constrain for the estimation of the real value of inbreeding [78]. In this sense, Polak et al. [79] reported an increase in the correlation coefficient between pedigree-based F and F_{ROH} along with the increase of the number of generations registered in pedigree data, which was also observed by [33] which used a pedigree database with a median depth of 15 generations. Our results are in agreement with both reports, highlighting the importance of the pedigree robustness in the estimation of inbreeding. Similarly, two recent studies compared genomic-based with pedigree-based inbreeding coefficients in the Norwegian–Swedish Coldblooded trotter and the Polish Cold-Blooded horses [79, 80]. In the second one, authors reported a moderate correlation between pedigree (F_{PED}) and genomic (F_{ROH}) inbreeding estimations, being 0.56 the highest value. In the first [79], authors found similar correlations between genealogical F and the genomic coefficient estimated based on ROH (0.443), but even lower when genomic inbreeding was estimated using the diagonal of the genomic relationship matrix. This agrees with our results, as the highest correlations were between pedigree-based F coefficients and ROH-based F measures (F_{RZ}) and the weakest between inbreeding coefficients based on the diagonal of the genomic (F_G) and hybrid (F_H) matrixes. However, it also need to be taking into account the differences in the number of individuals genotyped (805 in our case, 566 and 192 in their studies), as well the use of individuals from different generations, which is also required for an accurate estimation of F using pedigree values [22].

It was noteworthy that the inclusion of pedigree information in the G matrix (by estimating the hybrid matrix

H) improved the correlations of this methodology with pedigree-based methods but decreased the same correlations with genomic-based estimations. Despite the fact this is highly expectable, Martini et al. [18] demonstrated that the best fitting is highly dependant on the the correct estimation of τ and ω coefficients, and therefore, it should be adjusted very carefully.

Although most studies only correlated genomic inbreeding measures with the classical inbreeding coefficient (F), we have also made correlations between genomic estimates with ancestral inbreeding coefficients (F_{a_Bal} and F_{a_Kal}). F_{a_Bal} was moderately correlated with F_{RZ} (0.4) and F_{HOM} (0.39). This was in accordance with Schäler et al. [55], whose found a correlation of 0.39 between F_{a_Bal} and F_{HOM} , and 0.49 between F_{a_Bal} and F_{ROH} (ROH-based measure as F_{RZ}). However, their results do not agree with our correlations between F_{a_Kal} and F_{HOM} and F_{RZ} . While our correlations were high and significant between F_{HOM} and F_{RZ} (0.71 and 0.72, respectively), their correlations with F_{a_Kal} were low and non-significant (0.19 and 0, respectively). This can be explained not only by the lower genomic data analysed (from 76 individuals) but also by the population structure since Schäler et al. [55] analysed a population pedigree of 1273 individuals with an average F of 0.03, an average F_{a_Bal} of 0.024 and an average F_{a_Kal} of 0.09, which implies that recent inbreeding is more worrying than in the PRE population, whose recent inbreeding has been controlled. In addition, the results suggest that matrixial methods are less accurate in capturing the ancestral inbreeding than those based on more complex algorithms.

Nowadays, it is a common practice to determine recent and ancient inbreeding using the pedigree information of 3, 6, 9, or ROH fragments longer than 16.6, 8.3 and 5.5 Mb in pedigree and genomic-based estimations respectively. In our case, it was first employed a more refined approach by estimating the hypothetically F_{RZ} value during the last 20 generations, according to the minimum size of IBD fragments described by Fisher [14] (Fig. 2). Best correlations for F_3 , F_6 and F_{new_Kal} were observed using genomic data from 3, 4, and 20 generations (according the length of the ROH), demonstrating good fit. However, correlations decays in the three cases when shorter ROH fragments were included, which is highly expectable since all of them better estimate the recent inbreeding. On the contrary, F_9 and F increased the correlations rapidly with genomic values estimated using the first 6 generations, after which they become asymptotic in a value close to 0.78 despite the inclusion of more genomic information into the analysis. But also, this fits with the average pedigree information available on the PRE database (close to 9 generations on average), suggesting that the inclusion of additional genomic

information associated to this value (ROH shorter than 5 Mb approximately) is futile. However, it is also noteworthy that the differences observed among correlations before the lines start to decay (between 2nd and 5th generation in F_3 , F_6 and F_{new_Kal}) or become asymptotic (from 9th to 20th generation in F and F_9) are extremely low (less than 2% in average), and therefore, we can state that the correlation between genomic and pedigree-based estimations follows accurately the model proposed by Fisher [14] 70 years ago.

On the contrary, both ancestral estimators (AHC and F_{a_Bal}) as well the original F_{a_Kal} showed an increased correlation with the ROH-based genomic estimations, which was more marked when the genomic information included was lower than 7 generations (ROH < 7.1 Mb, according to Fisher [14]), after which the inclusion of additional genomic information did not provided a great improvement. However, it is noteworthy that AHC and F_{a_Bal} showed extremely lower correlations, in agreement with their aim in capturing the ancestral F , whereas F_{a_Kal} showed much higher values which were even better (in terms of correlations) with the obtained F_3 and F_{new_Kal} after the 9th generation and after the 15th in F_6 . Interestingly, our results fits adequately with both, the original theory of Fisher and the later validation of Howrigan [81], but in this case, based on the analysis of a real horse population worldwide bred as the Pura Raza Española horse.

Finally, Wang [82] suggested that pedigrees cannot be replaced completely by genomic data, because the pedigree estimations allow for the calculation of more complicated IBD coefficients for which the genomic estimations may have reduced capacity or limited power. Additionally, Todd et al. [63] suggested that the use of pedigree data allows inferences to be made for individuals from many generations ago, for whom biological samples might not be available for genotyping, particularly in horse.

Conclusions

In conclusion, our results show that if the pedigree has sufficient depth and reliability (especially if it includes individuals from different generations), the estimates obtained with the classical parameters present an acceptable correlation and therefore continue to be the most useful for the reproductive management of populations, even in this genomic era. In addition, the high correlation between the classical F and Kalinowski's F allows us to ensure that, in the current population, the possible founder effect is sufficiently diluted. In fact, the comparison of correlation between the F_{RZ} and the two previous pedigree-based estimates shows a slightly higher correlation of F_{RZ} with the classical F .

Obviously, it will also be preferable to obtain a real parameter instead of an estimated average, and the substantial incorporation of genomic information in livestock breeding programs gives us the opportunity to develop and implement new routines to manage populations at the genomic level. Nevertheless, the approximation obtained with a robust pedigree will allow us to work efficiently and more cheaply than with massive genotyping of the population, if the economic cost is a limitation. Obviously, in those populations in which there is a poor pedigree, the use of genomic information is the only valid way to obtain parameters for the genetic management of said population.

Abbreviations

AHC: Ancestral history coefficient; F : Wright classical inbreeding value; fa : Number of ancestors; Fa_Bal : Ballou's ancestral inbreeding; Fa_Kal : Kalinowski's ancestral inbreeding; fe : Number of founders; F_G : Genomic inbreeding value by the diagonal elements of the genomic matrix; F_{ij} : Genomic inbreeding value by the diagonal elements of the hybrid matrix; F_{HOM} : Genomic inbreeding value applying the method-of-moments methodology; F_{ij} : Partial inbreeding coefficients; F_{new_Kal} : Kalinowski's new inbreeding; F_{RZ} : Genomic inbreeding value applying the RZooROH R package; F_3 : Classical inbreeding considering 3 generations; F_6 : Classical inbreeding considering 6 generations; F_9 : Classical inbreeding considering 9 generations; IBD: Identity by descent; ROH: Runs of homozygosity.

Acknowledgements

We wish to thank the Pura Raza Española Horse Breeders Association for the data.

Authors' contributions

AM and MV conceived and designed the analyses. DPG and MV performed the pedigree data analyses. NL and SDP performed the genomic data analysis. DPG and NL wrote the first draft of the manuscript. All authors contributed to the interpretations of the results, the discussion and prepared the final manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the AGL-2017-84217-P Research project from Ministerio de Economía, Industria y Competitividad, Gobierno de España.

Availability of data and materials

The dataset supporting the results of this study was supplied by the National Association of Pura Raza Española Horse Breeders (ANCCE). The datasets generated and/or analysed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Departamento Agronomía, Escuela Técnica Superior de Ingeniería Agromónica, Universidad de Sevilla, Ctra Utrera Km 1, 41013 Sevilla, Spain.

²Departamento de Genética, Universidad de Córdoba, Córdoba, Spain.

³Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, La Plata,

Argentina. ⁴Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), La Plata, Argentina. ⁵Departamento de Producción Animal, Facultad de Veterinaria, Universidad Complutense de Madrid, Madrid, Spain.

Received: 8 May 2022 Accepted: 13 September 2022

Published online: 07 November 2022

References

- Wright S. Evolution and the genetics of populations, volume 3: Experimental results and evolutionary deductions. Chicago: University of Chicago Press; 1977.
- Hedrick PW, Kalinowski ST. Inbreeding depression in conservation biology. *Annu Rev Ecol Evol Syst.* 2000;31:139–62.
- Leroy G. Inbreeding depression in livestock species: review and meta-analysis. *Anim Genet.* 2014;45:618–28.
- Frankham R, Gilligan DM, Morris D, Briscoe DA. Inbreeding and extinction: effects of purging. *Conserv Genet.* 2001;2:279–85.
- De Cara MAR, Villanueva B, Toro MÁ, Fernández J. Purging deleterious mutations in conservation programmes: combining optimal contributions with inbred matings. *Heredity (Edinb).* 2013;110:530–7.
- Meuwissen THE. Maximizing the response of selection with a predefined rate of inbreeding. *J Anim Sci.* 1997;75:934–40.
- Wright S. Coefficients of inbreeding and relationship. *Am Nat.* 1922;56:330–8.
- Meuwissen THE, Luo Z. Computing inbreeding coefficients in large populations. *Genet Sel Evol.* 1992;24:305–13.
- Ballou JD. Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *J Hered.* 1997;88:169–78.
- Kalinowski ST, Hedrick PW, Miller PS. Inbreeding depression in the speke's gazelle captive breeding program. *Conserv Biol.* 2000;14:1375–84.
- Lacy RC, Alaks G, Walsh A. Hierarchical analysis of inbreeding depression in *Peromyscus polionotus*. *Evolution (N Y).* 1996;50:2187.
- Lacy RC, Alaks G, Walsh A, Galen C. Errata: hierarchical analysis of inbreeding depression in *Peromyscus polionotus*. *Evolution (N Y).* 1997;51(3):1025.
- Baumung R, Farkas J, Boichard D, Mészáros G, Sölkner J, Curik I. Grain: a computer program to calculate ancestral and partial inbreeding coefficients using a gene dropping approach. *J Anim Breed Genet.* 2015;132:100–8.
- Fisher RA. A fuller theory of 'junctions' in inbreeding. *Heredity (Edinb).* 1954;8:187–97.
- Woolliams JA, Berg P, Dagnachew BS, Meuwissen THE. Genetic contributions and their optimization. *J Anim Breed Genet.* 2015;132:89–99.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
- VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;91:4414–23.
- Martini JWR, Schrauf MF, Garcia-Baccino CA, Pimentel ECG, Munilla S, Rogberg-Muñoz A, et al. The effect of the h-1 scaling factors τ and ω on the structure of h in the single-step procedure. *Genet Sel Evol.* 2018;50:16. <https://doi.org/10.1186/s12711-018-0386-x>.
- Meyermans R, Gorssen W, Buys N, Janssens S. How to study runs of homozygosity using plink? A guide for analyzing medium density snp data in livestock and pet species. *BMC Genomics.* 2020;21:94.
- Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet.* 2018;19:220–34.
- Druet T, Gautier M. A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Mol Ecol.* 2017;26:5820–41.
- Doekes HP, Bijma P, Windig JJ. How depressing is inbreeding? A meta-analysis of 30 years of research on the effects of inbreeding in livestock. *Genes (Basel).* 2021;12:926.
- Hanski I, Gilpin ME, Ray C, Hoopes M, Hanski I, Gilpin ME. Metapopulation biology: ecology, genetics, and evolution. *Ecology.* 1997;78:2270.
- de Andres CD, Kaminsky M. Identification of 3 rare electromorphs among andalusian horses from Spain and genetic diversity of this population. *CR Acad Sci Paris.* 1985;299:189–94.

25. Kaminski M, de Andres Cara DF. Electrophoretic markers of andalusian horses: comparison of spanish and lusitanian lineages. *Comp Biochem Physiol.* 1986;83:575–88.
26. Negro S, Solé M, Pelayo R, Gómez MD, Azor PJ, Valera M. Molecular diversity between two cohorts of six spanish riding-horse breeds: impact of selection in crossbred vs purebred populations. *Livest Sci.* 2016;193:88–91.
27. Perdomo-González DI, Sánchez-Guerrero MJ, Molina A, Valera M. Genetic structure analysis of the pura raza español horse population through partial inbreeding coefficient estimation. *Animals.* 2020;10:1360.
28. Al Abri MA, Brooks SA, Al-Saqri N, Alkharousi K, Johnson EH, Alqaisi O, et al. Investigating the population structure and genetic diversity of arabian horses in oman using snp markers. *Anim Genet.* 2021;52:304–10.
29. Ablondi M, Dadousis C, Vasini M, Eriksson S, Mikko S, Sabbioni A. Genetic diversity and signatures of selection in a native italian horse breed based on snp data. *Animals.* 2020;10:1005.
30. Cosgrove EJ, Sadeghi R, Schlamp F, Holl HM, Moradi-Shahrbabak M, Miraei-Ashtiani R, et al. Genome diversity and the origin of the arabian horse. *Sci Rep.* 2020;10:9702.
31. Al Abri MA, Posbergh C, Palermo K, Sutter NB, Eberth J, Hoffman GE, et al. Genome-wide scans reveal a quantitative trait locus for withers height in horses near the ankr1 gene. *J Equine Vet Sci.* 2018;60:67–73.
32. Mancin E, Ablondi M, Mantovani R, Pigozzi G, Sabbioni A, Sartori C. Genetic variability in the italian heavy draught horse from pedigree data and genomic information. *Animals.* 2020;10:1310.
33. Bizarria dos Santos W, Pimenta Schettini G, Fonseca MG, Pereira GL, Loyola Chardulo LA, Rodrigues Machado Neto O, et al. Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the mangalarga marchador horse breed. *J Anim Breed Genet.* 2020;138:161–73.
34. Poyato-Bonilla J, Laseca N, Demyda-Peyrás S, Molina A, Valera M. 500 years of breeding in the carthusian strain of pura raza español horse: an evolutionary analysis using genealogical and genomic data. *J Anim Breed Genet.* 2021;139:84–99.
35. Laseca N, Molina A, Ramón M, Valera M, Azcona F, Encina A, et al. Fine-scale analysis of runs of homozygosity islands affecting fertility in mares. *Front Vet Sci.* 2022;9:754028.
36. Gutierrez JP, Goyache F. A note on endog: a computer program for analysing pedigree information. *J Anim Breed Genet.* 2005;122:172–6.
37. Affymetrix. Axiom analysis suite 5.0 software. Santa Clara: Affymetrix Inc.; 2016.
38. Bertrand AR, Kadri NK, Flori L, Gautier M, Druet T. RZooRoH: an R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. *Methods Ecol Evol.* 2019;10:860–6.
39. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in european populations. *Am J Hum Genet.* 2008;83:359–72.
40. Biscarini F, Cozzi P, Gaspa G, Marras G. detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes. CRAN (The Comprehensive R Archive Network); 2018. <https://orca.cardiff.ac.uk/id/eprint/108906/>.
41. Amadeu RR, Cellon C, Olmstead JW, Garcia AAF, Resende MFR, Muñoz PR, et al. AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. *Plant Genome.* 2016;9(3):2016.01.0009. <https://access.onlinelibrary.wiley.com/doi/10.3835/plantgenome2016.01.0009>.
42. Poyato-Bonilla J, Sánchez-Guerrero MJ, Santos RD, Valera M. Population study of the pura raza español horse regarding its coat colour. *Ann Anim Sci.* 2018;18:723–39.
43. Reverter A, Chan EKF. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics.* 2008;24:2491–7.
44. Rodríguez-Ramilo ST, Reverter A, Sánchez JP, Fernández J, Velasco-Galilea M, González O, et al. Networks of inbreeding coefficients in a selected population of rabbits. *J Anim Breed Genet.* 2020;137:599–608.
45. Watson-Haigh NS, Kadarmideen HN, Reverter A. PCIT: an R package for weighted gene co-expression networks based on partial correlation and information theory approaches. *Bioinformatics.* 2009;26:411–3.
46. Wickham H, François R, Henry L, Müller K. A grammar of data manipulation [R package dplyr version 1.0.0]. Media; 2020.
47. Wickham H, Averick M, Bryan J, Chang W, D'Agostino McGowan L, François R, et al. Welcome to the tidyverse. *J Open Source Softw.* 2019;4:1686.
48. Dowe M, Srinivasan A, Gorecki J, Chirico M, Stetsenko P, Short T, et al. Package 'data.table': extension of 'data.frame'. file LICENSE. 2019.
49. Wickham H. Ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016.
50. Varona L, Altarriba J, Moreno C, Martínez-Castillero M, Casellas J. A multivariate analysis with direct additive and inbreeding depression load effects. *Genet Sel Evol.* 2019;51:78.
51. Poyato-Bonilla J, Perdomo-González DI, Sánchez-Guerrero MJ, Varona L, Molina A, Casellas J, et al. Genetic inbreeding depression load for morphological traits and defects in the pura raza española horse. *Genet Sel Evol.* 2020;52:62.
52. Perdomo-González DI, Molina A, Sánchez-Guerrero MJ, Bartolomé E, Varona L, Valera M. Genetic inbreeding depression load for fertility traits in pura raza española mares. *J Anim Sci.* 2021;99:12.
53. Martínez-Castillero M, Varona L, Pegolo S, Rossoni A, Cecchinato A. Bayesian inference of the inbreeding load variance for fertility traits in brown swiss cattle. *J Dairy Sci.* 2021;104:10040–8.
54. Charlesworth D, Charlesworth B. Inbreeding depression and its evolutionary consequences. *Annu Rev Ecol Syst.* 1987;18:237–68.
55. Schärer J, Krüger B, Thaller G, Hinrichs D. Comparison of ancestral, partial, and genomic inbreeding in a local pig breed to achieve genetic diversity. *Conserv Genet Resour.* 2020;12:77–86. <https://doi.org/10.1007/s12686-018-1057-5>.
56. Doekes HP, Curik I, Hagy I, Farkas J, Köver G, Widing JJ, et al. Revised calculation of kalinowski's ancestral and new inbreeding coefficients. *Diversity.* 2020;12:155.
57. Addo S, Schärer J, Hinrichs D, Thaller G. Genetic diversity and ancestral history of the german angler and the red-and-white dual-purpose cattle breeds assessed through pedigree analysis. *Agric Sci.* 2017;08:1033–47.
58. Antonios S, Rodríguez-Ramilo S, Aguilar I, Astruc J, Legarra A, Vitezica Z. Genomic and pedigree estimation of inbreeding depression for semen traits in the basco-béarnaise dairy sheep breed. *J Dairy Sci.* 2021;104:3221–30.
59. Vostrá-Vydrová H, Vostrý L, Hofmanová B, Krupa E, Zavadilová L. Pedigree analysis of the endangered old kladruber horse population. *Livest Sci.* 2016;185:17–23.
60. Cervantes I, Molina A, Goyache F, Gutiérrez JP, Valera M. Population history and genetic variability in the spanish Arab horse assessed via pedigree analysis. *Livest Sci.* 2008;113:24–33.
61. Bussiman FO, Perez BC, Ventura RV, Peixoto MGCD, Curi RA, Balieiro JCC. Pedigree analysis and inbreeding effects over morphological traits in campolina horse population. *Animal.* 2018;12:2246–55.
62. Suwanlee S, Baumung R, Sölkner J, Curik I. Evaluation of ancestral inbreeding coefficients: ballou's formula versus gene dropping. *Conserv Genet.* 2007;8:489–95.
63. Todd ET, Ho SYW, Thomson PC, Ang RA, Velie BD, Hamilton NA. Founder-specific inbreeding depression affects racing performance in thoroughbred horses. *Sci Rep.* 2018;8:6167.
64. McParland S, Kearney F, Berry DP, McParland S, Kearney F, Berry DP. Purging of inbreeding depression within the irish Holstein-friesian population. *Genet Sel Evol.* 2009;41:16.
65. Keller MC, Visscher PM, Goddard ME. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics.* 2011;189:237–49.
66. Hoffman JL, Simpson F, David P, Rijks JM, Kuiken T, Thorne MAS, et al. High-throughput sequencing reveals inbreeding depression in a natural population. *Proc Natl Acad Sci U S A.* 2014;111:3775–80.
67. Doekes HP, Veerkamp RF, Bijma P, de Jong G, Hiemstra SJ, Windig JJ. Inbreeding depression due to recent and ancient inbreeding in dutch Holstein-friesian dairy cattle. *Genet Sel Evol.* 2019. <https://doi.org/10.1186/s12711-019-0497-z>.
68. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. *BMC Genet.* 2012;13:70.
69. Goszczynski D, Molina A, Terán E, Morales-Durand H, Ross P, Cheng H, et al. Runs of homozygosity in a selected cattle population with extremely inbred bulls: descriptive and functional analyses revealed highly variable patterns. *PLoS One.* 2018;13:7.

70. Nietlisbach P, Muff S, Reid JM, Whitlock MC, Keller LF. Nonequivalent lethal equivalents: models and inbreeding metrics for unbiased estimation of inbreeding load. *Evol Appl.* 2019;12:266.
71. Kardos M, Luikart G, Allendorf FW. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity (Edinb).* 2015;115:63–72.
72. Caballero A, Villanueva B, Druet T. On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. *Evol Appl.* 2021;14:416–28.
73. Villanueva B, Fernández A, Saura M, Caballero A, Fernández J, Morales-González E, et al. The value of genomic relationship matrices to estimate levels of inbreeding. *Genet Sel Evol.* 2021;53:42.
74. Yengo L, Zhu Z, Wray NR, Weir BS, Yang J, Robinson MR, et al. Detection and quantification of inbreeding depression for complex traits from snp data. *Proc Natl Acad Sci U S A.* 2017;114:8602–7.
75. Solé M, Gori A-S, Faux P, Bertrand A, Farnir F, Gautier M, et al. Age-based partitioning of individual genomic inbreeding levels in belgian blue cattle. *Genet Sel Evol.* 2017;49:92.
76. Laseca N, Molina A, Valera M, Antonini A, Demyda-Peyrás S. Comparison of different roh-based genomic inbreeding coefficients in a large cohort of horses. In: 73rd annual meeting of EAAP; 2022.
77. Howard JT, Pryce JE, Baes C, Maltecca C. Invited review: inbreeding in the genomics era: inbreeding, inbreeding depression, and management of genomic variability. *J Dairy Sci.* 2017;100:6009–24.
78. Cassell BG, Adamec V, Pearson RE. Effect of incomplete pedigrees on estimates of inbreeding and inbreeding depression for days to first service and summit milk yield in holsteins and jersey. *J Dairy Sci.* 2003;86:2967–76.
79. Polak G, Gurgul A, Jasielczuk I, Szmatola T, Krupiński J, Bugno-Poniewierska M. Suitability of pedigree information and genomic methods for analyzing inbreeding of polish cold-blooded horses covered by conservation programs. *Genes (Basel).* 2021;12:429.
80. Velie BD, Solé M, Fegraeus KJ, Rosengren MK, Røed KH, Ihler C-F, et al. Genomic measures of inbreeding in the norwegian-swedish cold-blooded trotter and their associations with known qtl for reproduction and health traits. *Genet Sel Evol.* 2019;51:22.
81. Howrigan DP, Simonson MA, Keller MC. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics.* 2011;12:460.
82. Wang J. Pedigrees or markers: which are better in estimating relatedness and inbreeding coefficient? *Theor Popul Biol.* 2016;107:4–13.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

