

RESEARCH ARTICLE

Open Access



Approaching autozygosity in a small pedigree of Gochu Asturcelta pigs

Katherine D. Arias¹ , Juan Pablo Gutiérrez² , Iván Fernández¹, Isabel Álvarez¹ and Félix Goyache^{1*}

Abstract

Background In spite of the availability of single nucleotide polymorphism (SNP) array data, differentiation between observed homozygosity and that caused by mating between relatives (autozygosity) introduces major difficulties. Homozygosity estimators show large variation due to different causes, namely, Mendelian sampling, population structure, and differences among chromosomes. Therefore, the ascertainment of how inbreeding is reflected in the genome is still an issue. The aim of this research was to study the usefulness of genomic information for the assessment of genetic diversity in the highly endangered Gochu Asturcelta pig breed. Pedigree depth varied from 0 (founders) to 4 equivalent discrete generations (t). Four homozygosity parameters (runs of homozygosity, F_{ROH} ; heterozygosity-rich regions, F_{HRR} ; Li and Horvitz's, F_{LH} ; and Yang and colleague's F_{YAN}) were computed for each individual, adjusted for the variability in the base population (BP; six individuals) and further jackknifed over autosomes. Individual increases in homozygosity (depending on t) and increases in pairwise homozygosity (i.e., increase in the parents' mean) were computed for each individual in the pedigree, and effective population size (N_e) was computed for five subpopulations (cohorts). Genealogical parameters (individual inbreeding, individual increase in inbreeding, and N_e) were used for comparisons.

Results The mean F was 0.120 ± 0.074 and the mean BP-adjusted homozygosity ranged from 0.099 ± 0.081 (F_{LH}) to 0.152 ± 0.075 (F_{YAN}). After jackknifing, the mean values were slightly lower. The increase in pairwise homozygosity tended to be twofold higher than the corresponding individual increase in homozygosity values. When compared with genealogical estimates, estimates of N_e obtained using F_{YAN} tended to have low root-mean-squared errors. However, N_e estimates based on increases in pairwise homozygosity using both F_{ROH} and F_{HRR} estimates of genomic inbreeding had lower root-mean-squared errors.

Conclusions Parameters characterizing homozygosity may not accurately depict losses of variability in small populations in which breeding policy prohibits matings between close relatives. After BP adjustment, the performance of F_{ROH} and F_{HRR} was highly consistent. Assuming that an increase in homozygosity depends only on pedigree depth can lead to underestimating it in populations with shallow pedigrees. An increase in pairwise homozygosity computed from either F_{ROH} or F_{HRR} is a promising approach for characterizing autozygosity.

Background

In small populations in which preservation of genetic variability is the goal, possible losses of genetic diversity have been traditionally assessed using genealogical information by computing coefficients of inbreeding (F) [1], defined as the probability that two alleles sampled at random at a locus are identical-by-descendent (IBD), or coefficients of coancestry (f) [2], defined as

*Correspondence:

Félix Goyache

fgoyache@serida.org

¹ Área de Genética y Reproducción Animal, SERIDA-Deva, Camino de Rioseco 1225, 33394 Gijón, Spain

² Departamento de Producción Animal, Universidad Complutense de Madrid, Avda. Puerta de Hierro S/N, 28040 Madrid, Spain



the probability that two alleles sampled at random in two different individuals are IBD. However, how F and f coefficients are reflected in the genome remains unclear. Although genealogical analyses assume that each individual in the base population has unique alleles at each locus, genomic data are subject to finite sampling and, thus, homozygosity at a locus can occur by chance. In other words, two alleles at a locus can be identical-by-state (IBS) rather than IBD. Therefore, estimating homozygosity due to IBD in an individual (autozygosity sensu, Cockerham [3]) has major difficulties because estimates widely vary from expectations as a consequence of Mendelian sampling [4] and the variance in IBD sharing per chromosome [5]. In this scenario, genomic estimates of homozygosity tend to measure the ‘realized inbreeding’ of an individual rather than the effect of autozygosity [6].

Ascertainment of the relationships between genealogical and genomic estimates of F has usually been approached using either simulated [7–13] or real [14–17] data with deep pedigrees. However, inconsistencies between these two sources of information may be greater in the case of small populations with shallow pedigrees and high proportions of full and half-siblings. Such scenarios could lead to weak correlations between pedigree-based and genomic estimates of inbreeding [8]. Moreover, mating policies can also affect the informativeness of genomic estimators. When matings between close relatives are avoided, allele frequencies tend to remain balanced across years and generations [18, 19]. In such scenarios, parameters that characterize homozygosity may not show clear trends [15].

There is consensus that the use of genomic estimators of inbreeding (homozygosity) may be beneficial to conservation programmes by allowing them to maintain the largest possible amount of genetic diversity [20, 21]. However, their efficiency could be affected by the genotyping platform [10, 19, 22], as well as by the characteristics of the population [13, 23–25]. Furthermore, genomic coefficients of inbreeding are highly dependent on the assumptions underlying their definition and do not always provide useful estimates of inbreeding due to inconsistent results in terms of gain and loss of genetic variability [15].

The rate of inbreeding computed using genealogical information is a standard criterion considered in conservation programmes, as it translates to the effective population size (N_e) of a population [22, 26]. It can be efficiently estimated using the individual increase in inbreeding coefficients (ΔF_i) [27, 28]. The parameter ΔF_i is not affected by changes in mating policy, allows a flexible definition of the reference population in a pedigree, and gives useful estimates of N_e even with shallow

pedigrees [28, 29]. However, similar approaches have not been attempted using genomic data.

The Gochu Asturcelta breed is an extremely endangered Celtic–Iberian pig population [30, 31] for which a conservation programme was started with six founders, only four of which provided viable offspring [32]. The start of the recovery programme has been well documented [32, 33]: its first stages were affected by incorrect management practices, including full-sib matings [34], causing a sudden increase in inbreeding. Immediately thereafter, a strict mating policy prohibiting matings between close relatives and prolonging the reproductive career of the founders and their direct descendants as much as possible was implemented to keep founder contributions balanced across generations and to preserve their genetic background in the present population [32]. The mating policies generated a complex pedigree useful for testing different hypotheses on the relationship between pedigree and genomic information and the preservation of genetic diversity [32, 33, 35].

The main objective of this research was to obtain empirical evidence on the usefulness of genomic information for the assessment of genetic diversity in small populations with shallow pedigrees using a sample of Gochu Asturcelta pigs. The performance of four estimators of homozygosity and the influence of genomic variability in the base population will be assessed, as well as the partitioning of genomic diversity into autosomes. An increase in homozygosity will be estimated at the individual level following two approaches: (a) computation of individual increases in homozygosity as an extension of the genealogical approach proposed by Gutiérrez et al. [27] and (b) computation of the increase in homozygosity of an individual over the mean homozygosity of its parents. Based on these estimates of increase in homozygosity, genomic estimates of N_e will be computed and further compared with genealogical estimators (ΔF_i and realized N_e). Insights for programmes aimed at the preservation of genetic variability in small animal populations will be discussed.

Methods

SERIDA is associated with the Ethical Committee in Research of the University of Oviedo (Spain), which ensures that all research with biological agents follows Good Laboratory Practices and European and Spanish regulations on biosecurity (Regulation of February 13th, 2014; BOPA no.47, February 26th, 2014). Tissue and hair root samples used in this project were collected by veterinary practitioners working for the Gochu Asturcelta Breeders’ Association (ACGA), with the permission and in the presence of the owners of the animals. For this reason, permission from the Ethical Committee in Research

of the University of Oviedo was not required. In all instances, ACGA veterinarians followed standard procedures and relevant national guidelines to ensure appropriate animal care.

Samples and genotyping

The analysed pedigree included 534 genotyped individuals, forming 526 parents–offspring trios, corresponding to 76 families (descendants of the same boar–sow pair) and 95 litters registered between 2000 and 2010. Data include the genotypes of 24 boars, 42 sows, and their offspring. Family size (offspring per family) varied from 1 to 34, with 48 families having 5 offspring or more. Based on pedigree, up to 105 genotyped individuals were non-inbred ($F=0$). Two genotyped founders and four full-sib individuals (born from 2003 to 2005 and direct descendants from two non-genotyped founders with offspring in the dataset [32]) were considered as the base population (BP); their genotypes were used to calculate genomic estimators of inbreeding relative to the BP.

Individuals were genotyped with Axiom-PorcineHDv1 of Affymetrix (658,692 single nucleotide polymorphisms, SNPs). The software Axiom Analysis Suite v4.0.3 (Thermo Fisher Scientific, Waltham, MA, USA) was used to create standard.ped and .map files. The genotyped SNPs were mapped using the *Sscrofa* genome build 11.1 [36]. Only SNPs on autosomal chromosomes with known positions were considered. To avoid the presence of null and false alleles, genotypes were only filtered on the basis of Mendelian errors [37]. In total, 503,043 SNPs (151,226 of them monomorphic) with a minimum call rate of 0.97 were retained for further analysis.

Pedigree analyses

The pedigree data were analysed using the program ENDOG v4.8 [38]. The pedigree was described by computing the number of fully traced generations (G) and the equivalent discrete generations (t) for each individual. G is defined as the number of generations separating the offspring from the farthest generation where the two ancestors of the individual are known, while t is the sum of $(1/2)^n$, where n is the number of generations separating the individual from each known ancestor [39].

Inbreeding coefficients (F_i) were computed for each individual i in the pedigree following Meuwissen and Luo [40]. The increase in inbreeding (ΔF_i) was computed as proposed by Gutiérrez et al. [28] as $\Delta F_i = 1 - \sqrt[t_i]{1 - F_i}$. The realized effective population size was computed for each cohort by averaging the individual increase in inbreeding coefficients of the n individuals included in the cohort as $Ne_{F_i} = 1/2\Delta F_i$ [27, 29]. As a complementary approach, N_e was also computed from the increase in pairwise

coancestry [41] as $Ne_{C_{ij}} = 1/(2\Delta C)$, with ΔC being the mean value of the increase in pairwise coancestry $\Delta C_{ij} = 1 - \sqrt[(t_i+t_j)/2]{1 - C_{ij}}$, where C_{ij} is the coancestry coefficient between individuals i and j and t_i and t_j are their respective equivalent discrete generations.

Reference subpopulations

For descriptive purposes, most analyses were presented for both the whole genotyped population and the subpopulation of individuals with two or more equivalent discrete generations (t) in their pedigree (t -subset; 480 individuals). Furthermore, different cohorts were defined according to (a) the year of birth (C2007, 112 offspring; C2008, 208 offspring; and C2009, 125 offspring, including 20 individuals born in 2010) and (b) the number of complete generations in their pedigree (two complete generations—CG2, 325 offspring; and three complete generations—CG3, 149 individuals).

Estimates of homozygosity

Four parameters that characterize homozygosity (F_{ROH} , F_{HRR} , F_{LH} , and F_{YAN}) were computed:

- F_{ROH} is the proportion of the genome covered by stretches of homozygous sites, usually referred to as runs of homozygosity (ROH) [42]. F_{ROH} for individual i was computed as $F_{ROH_i} = \frac{\sum L_{ROH_i}}{L_{AUTO}}$, where $\sum L_{ROH_i}$ is the length of the genome covered by ROH in individual i and L_{AUTO} is the length of the autosomal genome [42]. Identification of ROH for each individual was carried out using the consecutive runs approach implemented in the package detectRuns in R [43], using the following parameters: a maximum number of five heterozygous SNPs and a maximum number of five missing SNPs in a run; a minimum number of 20 SNPs in a run; a minimum length of 1 kb; and a maximum distance of 1 Mb allowed between consecutive SNPs. These parameters are expected to allow almost complete detection of ROH [44] and to avoid bias caused by recombination events [45]. For descriptive purposes only, the ROH segments identified in each individual were summarized into ROH regions using the *intersectBed* function of the BedTools version 2.28.0 software [46], and the results were plotted as an ideogram using the package RIdiogram of R [47].
- F_{HRR} is the proportion of the genome covered by heterozygosity-rich regions (HRR; formerly known as runs of heterozygosity) [48]. F_{HRR} for individual i was computed as $F_{HRR_i} = 1 - \frac{\sum L_{HRR_i}}{L_{AUTO}}$, where $\sum L_{HRR_i}$ is the length of the genome covered by

HRR in individual i . Identification of HRR for each individual was performed using the consecutive runs approach implemented in the package `detectRuns` in R [43]. Given that HRR segments are expected to be less frequent and shorter (<1 Mb) than ROH and that the main parameter leading to their identification is the number of homozygous SNPs allowed in a segment [17, 44, 49], the following parameters were used to detect HRR: a maximum number of five homozygous SNPs in a run; a maximum number of five missing SNPs in a run; a minimum number of 10 SNPs in a run; a minimum length of 1 kb; and a maximum distance of 1 Mb allowed between consecutive SNPs. For descriptive purposes, the HRR segments identified in each individual were summarized into HRR regions using the `intersectBed` function of the `BedTools` version 2.28.0 software [46], and the results were plotted as an ideogram using the package `RIdeogram` of R [47].

- (c) F_{LH} is the Li and Horvitz [50] estimator of the deviation of the observed frequency of homozygotes from that expected in the BP under Hardy–Weinberg proportions. The diagonals of the Li and Horvitz [50] relationship matrix were computed using custom R code as $F_{LH} = \frac{SF_{NEJ} - \sum_{k=1}^S [1 - 2p_{k(0)}(1 - p_{k(0)})]}{S - \sum_{k=1}^S [1 - 2p_{k(0)}(1 - p_{k(0)})]}$, where S is the number of loci, F_{NEJ} is the proportion of loci homozygous in individual i [51] and $p_{k(0)}$ is the frequency of the reference allele of SNP k in the BP. F_{LH} is equivalent to F_{IS} [52, 53] after correction for the homozygosity in the BP and, therefore, gives information in terms of IBD.
- (d) F_{YAN} is equal to the diagonals of the genomic relationship matrix constructed using the Yang et al. [54] approach based on the correlation between uniting gametes. Using custom R code, F_{YAN} was computed as $F_{YAN} = \frac{1}{S} \sum_{k=1}^S \frac{x_k^2 - (1 + 2p_{k(0)})x_{k_i} + 2p_{k(0)}^2}{2p_{k(0)}(1 - p_{k(0)})}$, where x_k is the genotype of the individual for SNP k . Yang et al.'s [54] estimator gives more weight to homozygotes for the minor allele than to homozygotes for the major allele.

The four parameters computed are expected to characterize differences: (a) in the length of homozygous genomic stretches (F_{ROH}); (b) in the length of non-random heterozygous genomic segments (F_{HRR}); (c) in expected homozygosity (F_{LH}); and (d) on the diagonal of the genomic relationship matrix (F_{YAN}). Unlike F_{ROH} and F_{HRR} , both F_{LH} and F_{YAN} account for allele frequencies in a previously defined BP.

For descriptive purposes, possible trends in homozygosity estimates by pedigree depth (t) were assessed by computing regression coefficients using the package `STAT` of R [55].

Adjustment of estimates of homozygosity

To improve robustness in the estimates of F_{ROH} , F_{HRR} , F_{LH} and F_{YAN} for each individual in the dataset, two consecutive adjustment strategies were applied:

- Following Powell et al. [56], individual estimates of homozygosity were adjusted for the mean homozygosity in the BP as $F_{ia} = \frac{F_i - F_{BP}}{1 - F_{BP}}$, where F_{ia} is the individual homozygosity estimate adjusted for the BP, F_i is the homozygosity estimate (F_{ROH} , F_{HRR} , F_{LH} or F_{YAN}) for individual i and F_{BP} is the mean of the BP for each corresponding homozygosity estimate. Note that this BP adjustment is independent of the use of the allele frequencies in the BP for the computation of F_{LH} and F_{YAN} .
- Considering the partitioning of genomic diversity into chromosomes, BP-adjusted estimates were also computed for each autosome and further averaged at the individual level by jackknifing over autosomes [57].

Hereafter, references to unadjusted estimates will be referred to as ‘raw’ estimates, estimates adjusted for the mean homozygosity in the BP will be referred to as ‘BP-adjusted’ estimates, and those computed with additional jackknifing over autosomes as ‘jackknifed’ estimates.

Increases in genomic inbreeding and N_e

For the four parameters used, the increase in homozygosity was computed at the individual level in two ways:

- Individual increase in homozygosity, computed as $\Delta t F_i = 1 - \sqrt[t]{1 - F_i}$, where F_i is either the BP-adjusted or the jackknifed estimate of homozygosity and t is the equivalent discrete generations in the pedigree of the individual.
- Increase in pairwise homozygosity, computed as $\Delta p F_{ijk} = \frac{F_i - [0.5(F_j + F_k)]}{1 - [0.5(F_j + F_k)]}$, where F_i , F_j and F_k are either the BP-adjusted or the jackknifed estimates of homozygosity of offspring i , parent j , and parent k , respectively.

Note that both these approaches are applied for each individual in the dataset as the deviations from either an “ideal” population with $t = 0$ (individual increase in

Table 1 Mean ($\pm$ standard deviation) of genealogical parameters (inbreeding, F_i ; individual increase in inbreeding, ΔF_i ; and equivalent discrete generations, t) and estimates of homozygosity (F_{ROH} , F_{HRR} , F_{LH} and F_{YAN}) and their increase computed from genomic information

Genealogical information	Whole population		<i>t</i> -subset		
	Raw values	BP-adjusted values	Jackknifed values	Raw values	BP-adjusted values
F_i	0.120 $\pm$ 0.074 [0.00;0.250]			0.134 $\pm$ 0.066 [0.000;0.250]	
ΔF_i	0.066 $\pm$ 0.048 [0.00;0.250]			0.074 $\pm$ 0.045 [0.000;0.250]	
t	2.8 $\pm$ 0.7 [0.0;4.0]			2.9 $\pm$ 0.5 [2.0;4.0]	
Homozygosity estimates					
F_{ROH}	0.522 $\pm$ 0.052 [0.390;0.661]	0.123 $\pm$ 0.095 [-0.118;0.379]	0.116 $\pm$ 0.096 [-0.136;0.407]	0.530 $\pm$ 0.047 [0.399;0.661]	0.139 $\pm$ 0.085 [-0.102;0.379]
F_{HRR}	0.887 $\pm$ 0.013 [0.854;0.922]	0.111 $\pm$ 0.104 [-0.155;0.382]	0.104 $\pm$ 0.104 [-0.18;0.428]	0.889 $\pm$ 0.012 [0.855;0.922]	0.127 $\pm$ 0.095 [-0.149;0.382]
F_{LH}	-0.041 $\pm$ 0.093 [-0.283;0.226]	0.099 $\pm$ 0.081 [-0.110;0.330]	0.092 $\pm$ 0.081 [-0.127;0.384]	-0.026 $\pm$ 0.083 [-0.264;0.226]	0.112 $\pm$ 0.072 [-0.094;0.330]
F_{YAN}	0.030 $\pm$ 0.086 [-0.230;0.296]	0.152 $\pm$ 0.075 [-0.076;0.384]	0.148 $\pm$ 0.072 [-0.059;0.359]	0.043 $\pm$ 0.078 [-0.145;0.296]	0.163 $\pm$ 0.069 [-0.001;0.384]
Individual increase in homozygosity					
ΔF_{ROH}	0.243 $\pm$ 0.054 [0.153;0.521]	0.046 $\pm$ 0.041 [-0.112;0.196]	0.043 $\pm$ 0.042 [-0.136;0.23]	0.234 $\pm$ 0.043 [0.153;0.406]	0.052 $\pm$ 0.036 [-0.044;0.196]
ΔF_{HRR}	0.556 $\pm$ 0.087 [0.416;0.883]	0.041 $\pm$ 0.044 [-0.097;0.198]	0.038 $\pm$ 0.045 [-0.149;0.244]	0.537 $\pm$ 0.063 [0.416;0.715]	0.047 $\pm$ 0.039 [-0.064;0.198]
ΔF_{LH}	-0.018 $\pm$ 0.045 [-0.269;0.120]	0.037 $\pm$ 0.035 [-0.098;0.181]	0.034 $\pm$ 0.036 [-0.122;0.215]	-0.009 $\pm$ 0.032 [-0.110;0.120]	0.042 $\pm$ 0.031 [-0.041;0.181]
ΔF_{YAN}	0.010 $\pm$ 0.037 [-0.182;0.131]	0.060 $\pm$ 0.032 [-0.034;0.183]	0.058 $\pm$ 0.031 [-0.038;0.178]	0.016 $\pm$ 0.030 [-0.052;0.131]	0.061 $\pm$ 0.030 [0.000;0.182]
Increase in pairwise homozygosity					
$\Delta p_{F_{ROH}}$	0.087 $\pm$ 0.106 [-0.204;0.368]	0.087 $\pm$ 0.106 [-0.204;0.368]	0.090 $\pm$ 0.102 [-0.201;0.387]	0.098 $\pm$ 0.101 [-0.204;0.368]	0.098 $\pm$ 0.101 [-0.204;0.368]
$\Delta p_{F_{HRR}}$	0.091 $\pm$ 0.112 [-0.222;0.385]	0.090 $\pm$ 0.112 [-0.222;0.385]	0.092 $\pm$ 0.109 [-0.205;0.421]	0.102 $\pm$ 0.107 [-0.222;0.385]	0.102 $\pm$ 0.107 [-0.222;0.385]
$\Delta p_{F_{LH}}$	0.065 $\pm$ 0.090 [-0.211;0.291]	0.065 $\pm$ 0.090 [-0.211;0.291]	0.067 $\pm$ 0.088 [-0.219;0.322]	0.075 $\pm$ 0.083 [-0.158;0.291]	0.075 $\pm$ 0.083 [-0.158;0.291]
$\Delta p_{F_{YAN}}$	0.071 $\pm$ 0.088 [-0.160;0.352]	0.071 $\pm$ 0.088 [-0.160;0.352]	0.071 $\pm$ 0.083 [-0.158;0.352]	0.077 $\pm$ 0.084 [-0.156;0.352]	0.077 $\pm$ 0.084 [-0.156;0.352]

Minimum and maximum values are given in brackets. Genomic parameters are presented as raw values as well as values adjusted for the mean homozygosity in the base population (BP) and values obtained via jackknifing over autosomes. Data are given for the whole population as well as for the individuals with two or more equivalent discrete generations in their pedigree (*t*-subset). ΔF values correspond to the individual increase in inbreeding computed for each of the four homozygosity parameters tested, while ΔpF values indicate the corresponding increase in pairwise homozygosity values

homozygosity) or the mean genomic inbreeding of the parents of the individual, assumed to be a “reference population” from which each offspring is differentiated (increase in pairwise homozygosity).

Effective population sizes were subsequently computed by averaging ΔtF_i and ΔpF_{ijk} among all individuals in a cohort and computed as $\bar{N}_{eF_i} = \frac{1}{2\Delta tF_i}$ and $\bar{N}_{eF_{ijk}} = \frac{1}{2\Delta pF_{ijk}}$,

respectively. Uncertainty of the $\bar{N}_{eF_i}$ and $\bar{N}_{eF_{ijk}}$ estimates obtained for the two yearly and genealogical cohorts defined were measured via root-mean-squared error (RMSE) as $RMSE = \sqrt{\frac{\sum (N_e - \bar{N}_{eF})^2}{n}}$, where N_e is either the estimate of N_{eF_i} or of $N_{eC_{ij}}$ obtained from genealogical information, $\bar{N}_{eF}$ is the corresponding estimate of effective population size computed from either ΔtF_i or ΔpF_{ijk} for F_{ROH} , F_{HRR} , F_{LH} or F_{YAN} , and n is the sample size (5, in this case).

Results

A full description of the pedigree of the Gochu Asturcelta pig used is given in association with the estimates of genomic parameters computed for each individual as Additional files (see Additional file 1: Tables S1, S2, S3, S4). Table 1 gives mean values for the genealogical parameters describing the pedigree. The mean pedigree-based inbreeding level of the population was 0.120 ± 0.074 , with a mean pedigree depth of 2.8 ± 0.7 equivalent discrete generations (t). These values were only slightly higher (mean $F_i = 0.134 \pm 0.066$; $t = 2.9 \pm 0.5$) for the *t*-subset, which included 90% of the individuals in the dataset.

Raw homozygosity estimates

In total, 1,411,686 and 2,176,412 ROH and HRR segments, respectively, were identified across the 534 genotyped individuals. These segments were summarized into 3017 ROH areas covering 1.04 Gb and 4646 HRR areas covering 290.7 Mb over the 18 autosomes (Fig. 1).

Figure 2a illustrates the dispersion of the raw estimates of homozygosity by autosome in the individuals belonging to the BP. The raw estimates showed wide dispersion and evident differences in scale among homozygosity parameters. Raw estimates of the four parameters used (F_{ROH} , F_{HRR} , F_{LH} and F_{YAN}) computed at the individual level are given in Additional file 1: Table S1 and summarised in Table 1. As expected, before BP adjustment of the estimates for each individual, F_{ROH} and F_{HRR} always took positive and high mean values (0.522 and above, [see Additional file 1: Table S3]). However, both F_{LH} and F_{YAN} took significantly lower ($F_{YAN} = 0.030 \pm 0.086$) and even negative ($F_{LH} = -0.041 \pm 0.093$) mean values. Raw estimates for both F_{LH} and F_{YAN} took negative values in a substantial number of individuals (355 individuals, 66%, and 203 individuals, 38%, respectively), including those belonging to the BP.

BP-adjusted homozygosity estimates

Figure 2b illustrates the effect of adjustment for the mean variability in the BP on the estimates of homozygosity, by autosome, in the BP individuals. BP adjustment gave less dispersed estimates across chromosomes. However, some individuals still had estimates of homozygosity departing from expectations on some chromosomes (namely, *Sus scrofa* (SSC) chromosome SSC2, SSC12 and SSC13).

Genomic parameters computed after adjustment for the mean homozygosity in the BP are given for each individual in the pedigree in Additional file 1: Table S2. Mean values for genomic parameters after adjustment for variability in BP are in Table 1 for both the whole population and the t -subset. After adjustment, both F_{ROH} and F_{HRR} took negative values for some individuals belonging to the first generations of the pedigree (57 individuals, 11%, and 96 individuals, 18%, respectively), whereas the frequency of individuals with negative values for F_{LH} and F_{YAN} was lower (73 individuals, 14%, and 11 individuals, 2%, respectively). However, the means for F_{ROH} (0.123 ± 0.095), F_{HRR} (0.111 ± 0.104), F_{LH} (0.099 ± 0.081), and F_{YAN} (0.152 ± 0.075) were always positive. The corresponding means for the t -subset were slightly higher, ranging from 0.112 ± 0.072 (F_{ROH}) to 0.163 ± 0.069 (F_{YAN}).

Figure 3 (left column) illustrates the variation in the BP-adjusted individual homozygosity values per t . The

four homozygosity parameters showed a large amount of variation. When families with five or more offspring were considered, the within-family coefficient of variation ranged from 11 to 519% for F_{ROH} , from 17 to 334% for F_{HRR} , from 10 to 1,366% for F_{LH} , and from 55 to 1,688% for F_{YAN} . Although the homozygosity values did not show clear trends, regression of individual homozygosity estimates on t indicated that F_{ROH} , F_{HRR} , F_{LH} and F_{YAN} tended to increase significantly ($p < 0.001$) with pedigree depth. However, this scenario was not confirmed in the t -subset [see Additional file 1: Table S3]: regression of both F_{LH} and F_{YAN} estimates on t became statistically non-significant ($p > 0.10$), while those corresponding to F_{ROH} and F_{HRR} continued to be significant ($p < 0.05$).

Jackknifed estimates of homozygosity

Genomic parameters computed via jackknifing over autosomes are given for each individual in the pedigree

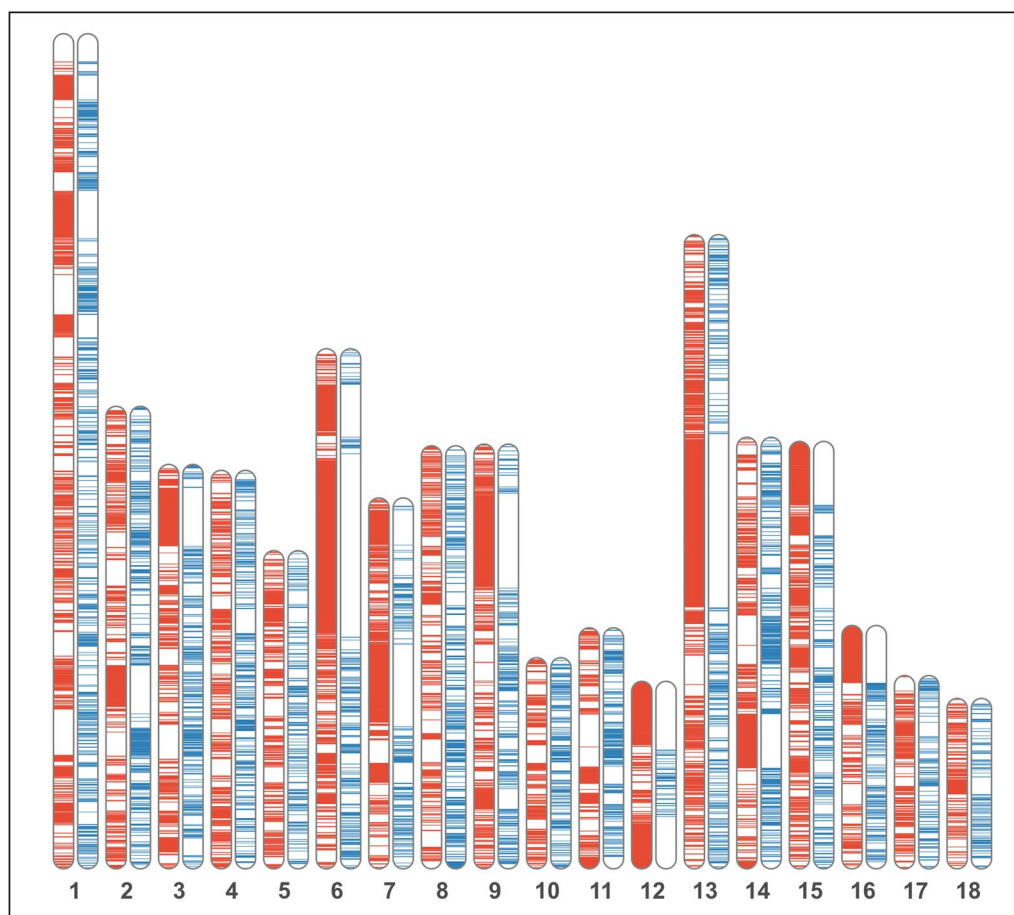


Fig. 1 Ideogram illustrating, per autosome, the distribution of the 3017 and 4646 genomic areas in which ROH (in red) and HRR (in blue) segments were identified across the 534 genotyped Gochu Asturcelta individuals

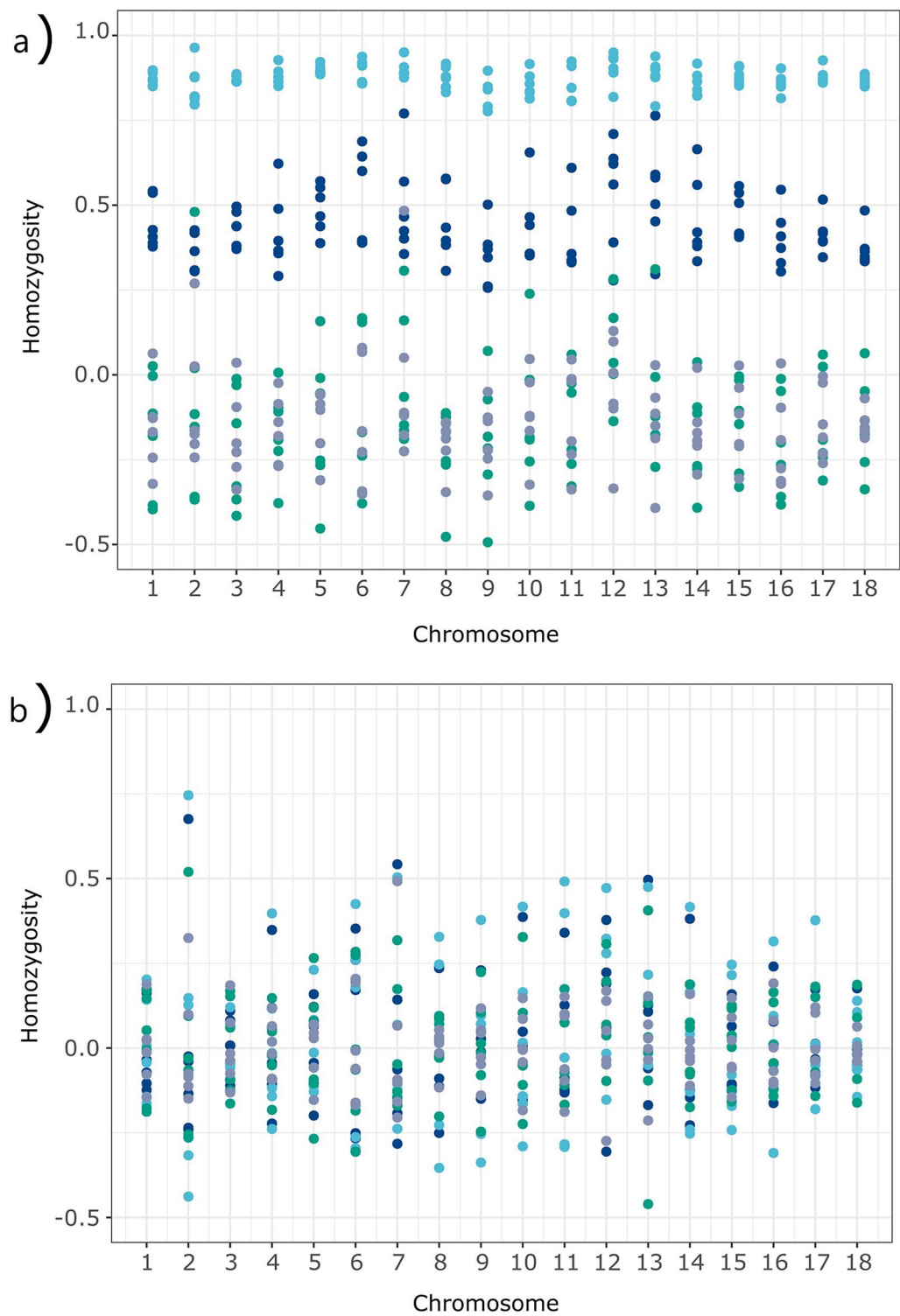


Fig. 2 Dispersion, over the 18 autosomes, of the four estimates of homozygosity tested (F_{ROH} , in dark blue; F_{HRR} , in light blue; F_{LH} , in grey; and F_{YAN} in green) in the six individuals of the Gochu Asturcelta pedigree used as the base population. Plot (a) illustrates the raw mean values, whereas plot (b) illustrates the mean values after adjustment for the mean for each autosome in the base population

in Additional file 1: Table S4. The corresponding mean values for the whole population and the t -subset are also in Table 1. After jackknifing, mean values were slightly lower than those obtained adjusting for BP only. A higher mean was obtained for F_{YAN} (0.148 ± 0.072), and a lower mean was obtained for F_{LH} (0.092 ± 0.081), whereas the means for F_{ROH} (0.116 ± 0.096) and F_{HRR} (0.104 ± 0.104) had intermediate values. The ranges (maximum and minimum values) of the jackknifed values computed for F_{ROH} , F_{HRR} , and F_{LH} were higher than those computed after BP adjustment only. However, the range for F_{YAN} was narrower than that observed after correcting for variability in the BP only.

Figure 3 also illustrates variation in the individual homozygosity values per t after jackknifing over autosomes. Compared with the BP-adjusted estimates, no clear differences in variation could be assessed. However, both regression coefficients and goodness-of-fit (R^2) of the regressions of homozygosity estimates on t were slightly higher than those obtained for the BP-adjusted values (Fig. 3; Additional file 1: Table S3). The within-family coefficients of variation (for families with five or more offspring) ranged from 11 to 726% for F_{ROH} , from 17 to 723% for F_{HRR} , and from 7 to 1865% for F_{LH} . However, the range of within-family coefficients of variation for F_{YAN} was significantly narrower (ranging from 7 to 178%).

Increases in inbreeding and homozygosity

Table 1 also gives mean values for the genealogical individual increase in inbreeding (ΔF_i), as well as for the individual increase in homozygosity (ΔtF_i) and the increase in pairwise homozygosity (ΔpF_{ijk}) values computed from genomic data. Although they were computed via jackknifing over autosomes or adjusting for the variability in the BP only, mean ΔpF_{ijk} values were roughly twofold higher than the corresponding ΔtF_i values, except for F_{YAN} , for which the mean values were more similar ($\Delta pF_{YAN} = 0.071 \pm 0.088$ and $\Delta tF_{YAN} = 0.060 \pm 0.032$, and $\Delta pF_{YAN} = 0.071 \pm 0.083$ and $\Delta tF_{YAN} = 0.058 \pm 0.031$, respectively). In spite of the similar computation methods applied, mean ΔtF_i values computed for F_{ROH} , F_{HRR} , and F_{LH} were lower than the genealogical increase in inbreeding (0.066 ± 0.048), which, in turn, was closer to the BP-adjusted and jackknifed mean values of

ΔtF_{YAN} . These trends were the same for the t -subset. Figure 4 shows the variation in the genealogical increase in inbreeding and both the individual increase in homozygosity and the increase in pairwise homozygosity by pedigree depth. As expected, for $t \geq 2$, the ΔtF_i values tended to fit the genealogical reference (ΔF_i) better than the ΔpF_{ijk} values. Furthermore, while ΔtF_i values tended to be lower than the mean genealogical increase in inbreeding, particularly for ΔtF_{LH} , the ΔpF_{ijk} values could take higher or lower values than the genealogical reference at different levels of t .

Consistent with the lower mean homozygosity values computed via jackknifing over autosomes, mean jackknifed ΔtF_i values were slightly lower than those computed using BP-adjusted values (Table 1). However, the increase in pairwise homozygosity values showed the opposite pattern, with mean jackknifed ΔpF_{ijk} values higher than the corresponding BP-adjusted values, except for F_{YAN} . In practical terms, the mean values for both ΔtF_{YAN} and ΔpF_{YAN} were the same.

Effective population size

Table 2 gives the effective population size computed from genealogical and genomic information for the five yearly and genealogical cohorts defined. Regarding genealogical information, N_e computed from individual increases in inbreeding was slightly higher than that computed from increases in pairwise coancestry, except for C2007 (which was approximately twofold higher) and CG3, for which $N_e C_{ij}$ (7.8) was higher than $N_e F_i$ (7.0).

The N_e values estimated using jackknifed individual increases in homozygosity were always higher than those obtained after adjustment for the BP only. However, the opposite pattern was observed for the N_e estimates based on the increase in pairwise homozygosity. Estimates of N_e computed using ΔpF_{ijk} tended to be roughly half of those computed using ΔtF_i . In any case, independently of the method for the calculation of increases in homozygosity, estimates of $\bar{N}_e F_{YAN_{ijk}}$ tended to be nearer to their $\bar{N}_e F_{YAN_i}$ counterparts.

N_e estimates obtained using F_{YAN} had stable root-mean-squared errors, independent of the approach used for computation of the increase in homozygosity (Table 3). However, when an increase in pairwise

(See figure on next page.)

Fig. 3 Individual variation in the four different homozygosity estimates used (F_{ROH} , F_{HRR} , F_{LH} , and F_{YAN}). Values computed for each individual in the dataset (in circles) are plotted by pedigree depth (equivalent discrete generations) of the animals. Regression of the individual homozygosity values over pedigree depth (t) is also displayed. The corresponding regression coefficients (b) and their statistical significance (p) are also given. The left column illustrates the variation in the base population-adjusted values, whereas the right column illustrates the variation in the jackknifed values

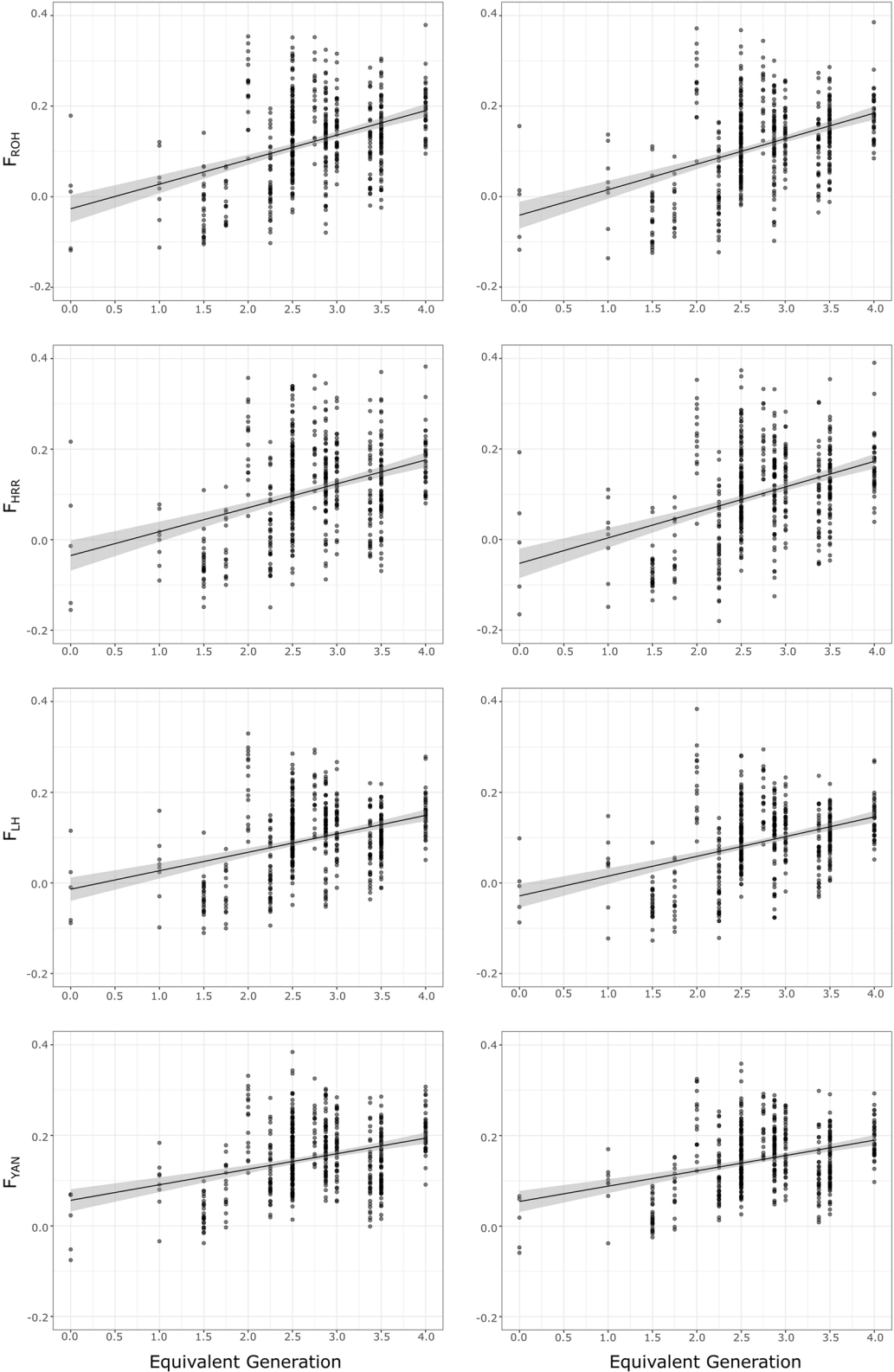


Fig. 3 (See legend on previous page.)

homozygosity was considered, $\bar{N}_e F_{ROH_{ijk}}$ and particularly $\bar{N}_e F_{HRR_{ijk}}$ resulted in lower RMSE values.

Discussion

Three groups of inbreeding coefficients were assessed in this study: (a) genealogical F_i , as a standard indicator of the genetic variability in a pedigree; (b) F_{ROH} and F_{HRR} , estimating genomic identity from homozygous and heterozygous genomic segments, respectively; and (c) F_{LH} and F_{YAN} , estimating homozygosity while accounting for the allele frequencies in the BP. The pedigree-based inbreeding coefficients reflect IBD probabilities at an infinite number of unlinked loci over descendants that share a pedigree. However, this assumption does not fit with genomic inbreeding estimates. It is unlikely that individuals of a founder population, for instance, in populations under conservation programmes [58], do not show different degrees of genomic relatedness and, possibly, some degree of structuring. This is why the definition of a BP for adjustment of genomic inbreeding estimates is challenging [13].

Usefulness of estimates of homozygosity

Comparison among homozygosity parameters is not straightforward, not only because of differences in computation but also because they are dependent on the variability in the BP [13, 15]. The probabilistic process of Mendelian segregation results in large variance across loci, among individuals, and at the population level. Consequently, homozygosity parameters have higher variation than expected from genealogies, even in full sibs [19, 48]. The four homozygosity parameters used here were adjusted for the mean homozygosity values in the BP, making the different estimates more comparable (Fig. 2).

The first group of homozygosity parameters tested, F_{ROH} and F_{HRR} , are not computed using allele frequencies [59]. Their main advantage is that, before BP adjustment, their definition always yields positive values (ranging from 0 to 1), leading to a straightforward interpretation. However, technical issues, such as marker density and the computational methods used (and the parameters assumed) for identification of either ROH or HRR segments, affect the estimates due to a possible less-efficient identification of shorter segments. This may be of

importance for both F_{ROH} and F_{HRR} . Since we used SNP array data with sufficient density to capture long homozygous or heterozygous segments, we applied the consecutive runs approach to identify both ROH and HRR stretches. This approach has been reported to have better performance for the identification of short (homozygous or heterozygous) genomic stretches than the well-known sliding window approach [17, 60].

Different studies have suggested that F_{ROH} can be an accurate estimator of IBD [13, 16, 61–63]. However, in small pedigrees, F_{ROH} may depend to a large extent on long ROH segments [6]. ROH patterns in a population are frequently conditioned by either recent demographic events or selection in the populations under study, leading to the occurrence of long ROH segments [63, 64]. Homozygosity estimates are greater when using long ROH segments than when using shorter ROH segments [13]. Therefore, fitting well with the Gochu Asturcelta scenario analysed, F_{ROH} better characterizes recent inbreeding than old inbreeding.

This is the first study that uses F_{HRR} to estimate genomic inbreeding. Unlike ROH, the nature of HRR is not well understood. Although recombination and other processes can shape the frequency and distribution of ROH over the genome, ROH are significantly more common in regions with high linkage disequilibrium and low recombination, tending to be identically inherited from the parents and therefore provide a good IBD estimator (see Gibson et al. [65] and many others later). Although the occurrence of HRR has been reported to be statistically non-random and to cluster in specific chromosomal regions [48], it seems to be independent from linkage disequilibrium in the regions [66]. Furthermore, HRR have been recently shown to be independent of the expected heterozygosity in the population [67] and informative about population structure and demographic history [49].

Altogether, the published evidence described above suggests that the evolutionary forces that shape ROH and HRR segments can be different. Furthermore, our results (Fig. 1) confirm that the chromosomal region with ROH and HRR segments do not overlap. Although the usefulness of HRR for monitoring genomic diversity in livestock needs further confirmation, we consider

(See figure on next page.)

Fig. 4 Comparison between the variation of different estimates of individual increase in homozygosity and the genealogical individual increase in inbreeding (dotted line) by pedigree depth (equivalent discrete generations; t). In the left column, the individual increase in homozygosity (red line) and the increase in pairwise homozygosity (green line) computed from the base population-adjusted values of individual homozygosity are given for each of the four homozygosity parameters (F_{ROH} , F_{HRR} , F_{LH} and F_{YAN}) tested. In the right column, the individual increase in homozygosity (blue line) and the increase in pairwise homozygosity (purple line) computed from the values of individual homozygosity obtained via jackknifing over autosomes are given for the same parameters

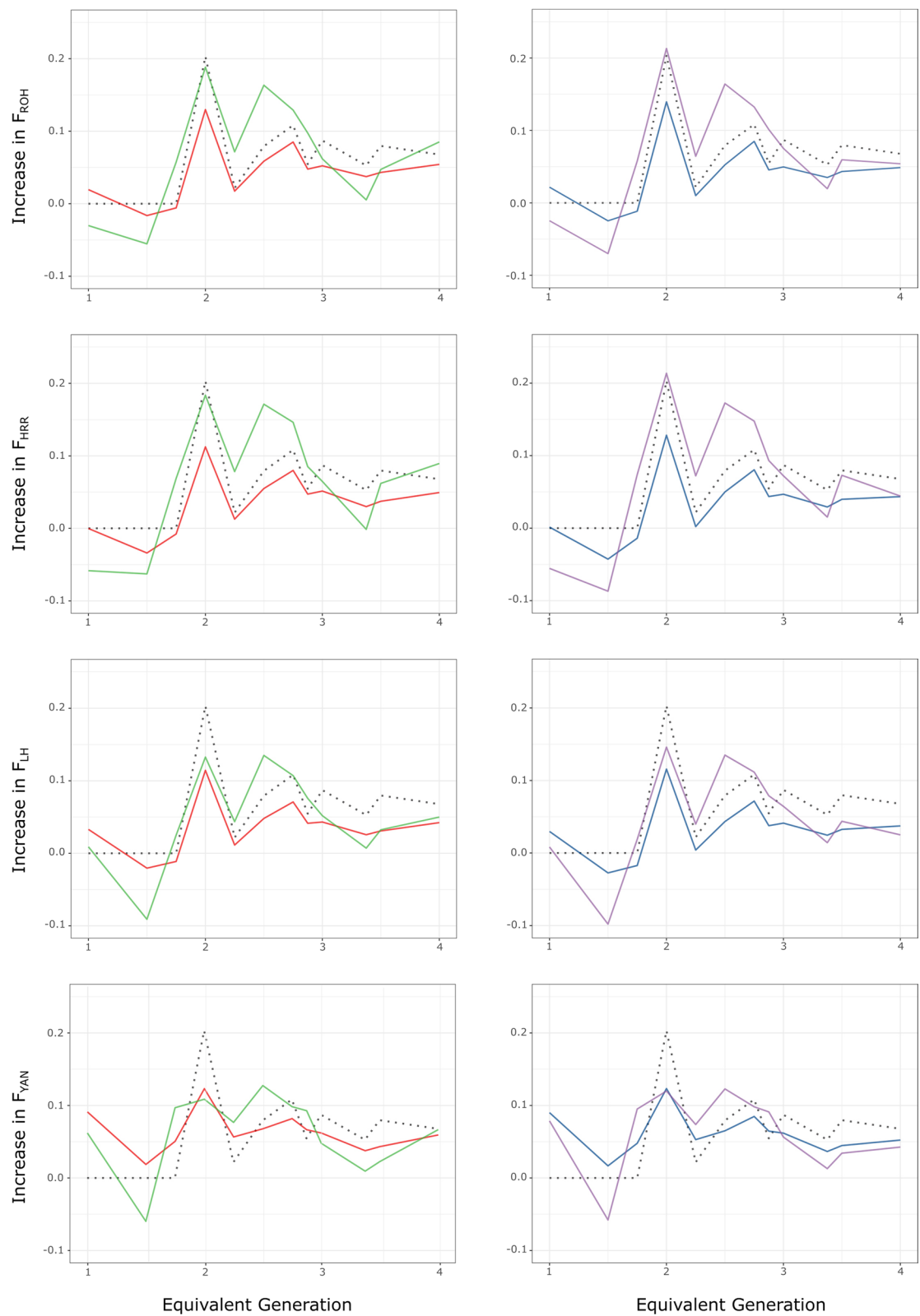


Fig. 4 (See legend on previous page.)

Table 2 Estimates of effective population size computed using genealogical information (individual increase in inbreeding, Ne_{F_i} and coancestry, $Ne_{C_{ij}}$) and genomic information (individual increase in homozygosity and increase in pairwise homozygosity)

	Yearly cohorts						Genealogical cohorts			
	C2007	C2008	C2009	C2009	C2009	C2009	CG2	CG3	CG3	CG3
Cohort size	112	208	125	125	125	125	325	149	149	149
Genealogical information										
Ne_{F_i}	8.3	7.7	7.5	7.5	7.5	7.5	6.5	7.0	7.0	7.0
$Ne_{C_{ij}}$	4.5	6.3	6.5	6.5	6.5	6.5	4.4	7.8	7.8	7.8
t	2.5 ± 0.3	3.1 ± 0.5	3.1 ± 0.5	3.1 ± 0.5	3.1 ± 0.5	3.1 ± 0.5	2.6 ± 0.3	3.6 ± 0.2	3.6 ± 0.2	3.6 ± 0.2
Individual increase in homozygosity										
	BP-adjusted values	Jackknifed values	BP-adjusted values	Jackknifed values	BP-adjusted values	Jackknifed values	BP-adjusted values	Jackknifed values	BP-adjusted values	Jackknifed values
$\bar{N}_e F_{ROH_i}$	9.9	10.24	12.0	12.94	10.0	11.30	9.1	9.73	11.4	11.86
$\bar{N}_e F_{HRR_i}$	11.5	11.72	13.3	14.36	10.4	12.18	9.7	10.49	13.2	13.23
$\bar{N}_e F_{LH_i}$	12.7	13.06	16.1	17.19	11.7	13.47	10.9	11.89	15.7	15.97
$\bar{N}_e F_{YAN_i}$	9.2	9.24	9.9	10.03	7.3	7.79	7.3	7.46	11.1	11.36
Increase in pairwise homozygosity										
	BP-adjusted values	Jackknifed values	BP-adjusted values	Jackknifed values	BP-adjusted values	Jackknifed values	BP-adjusted values	Jackknifed values	BP-adjusted values	Jackknifed values
$\bar{N}_e F_{ROH_{ijk}}$	4.2	3.97	7.3	6.87	5.7	5.96	4.1	3.97	11.5	10.91
$\bar{N}_e F_{HRR_{ijk}}$	4.0	3.83	6.8	6.55	5.7	5.98	4.0	3.87	9.9	10.15
$\bar{N}_e F_{LH_{ijk}}$	5.7	5.30	9.6	8.84	7.1	7.51	5.2	5.00	18.3	17.15
$\bar{N}_e F_{YAN_{ijk}}$	5.4	5.32	9.9	9.60	6.1	6.61	5.0	5.05	18.9	18.15

Increases in genomic homozygosity were computed for four different genomic parameters: F_{ROH} , F_{HRR} , F_{LH} and F_{YAN} . Genomic-based estimates of N_e were computed from raw homozygosity values as well as from unbiased estimates obtained via jackknifing over autosomes. Estimates were computed for the five yearly and genealogical cohorts defined. Mean equivalent discrete generations (t ; standard deviation in brackets) are also given

that the information provided by HRR through the different evolutionary forces acting on different chromosomal regions cannot be disregarded and is likely to be, at least, complementary to that provided by ROH. In any case, the increasing availability of sequence data will further clarify the evolutionary and demographic drivers of HRR and the role of these regions in livestock adaptation.

The literature suggests that HRR are short or very short and that they could be subject to balancing or counter-vailing selection [17, 44, 48, 49]. Our results confirm that HRR segments are short (always shorter than 1 Mb, and 89% of them shorter than 100 kb). Therefore, theoretically, the ability of ROH and HRR to characterize overall homozygosity (gain of homozygous stretches or loss of heterozygous stretches) in an individual may differ. However, our results suggest that after BP adjustment, both F_{ROH} and F_{HRR} values are highly consistent (Table 1; Fig. 2).

In contrast with F_{ROH} and F_{HRR} , which are putatively useful to characterize the overall homozygosity in an individual or population, the second group of homozygosity parameters, F_{LH} and F_{YAN} , have been reported as useful estimators of homozygosity due to IBS [9, 16, 45, 53, 59]. The results of F_{LH} and F_{YAN} depend on the allele frequencies of a defined reference population [45]. Furthermore, F_{YAN} is heavily weighted by the frequency of rare alleles [13, 15, 53, 59]. Values obtained for either F_{LH} or F_{YAN} can range from -1 to 1, and in our data, individual estimates of F_{LH} and F_{YAN} took negative values with some frequency. Caballero et al. [13] showed that both F_{LH} and F_{YAN} can take lower values than expected due to the excess of heterozygotes expected for markers in finite populations. Therefore,

the behaviour of F_{LH} and F_{YAN} cannot be straightforwardly compared to that of F_{ROH} and F_{HRR} [15, 53] unless BP adjustment is applied. Although F_{ROH} , F_{HRR} , and F_{LH} weight each genotyped allele equally, Yang et al.'s [54] method tends to yield higher homozygosity estimates in individuals that are homozygous for rare alleles [13, 15]. Other methods, particularly F_{ROH} , were less efficient in capturing homozygosity of rare alleles [52].

Adjustment of homozygosity estimates

The reasons summarized above make different homozygosity parameters difficult to compare. To overcome this issue, we adjusted the four homozygosity measures according to the mean values in the BP, allowing more comparable behaviours among them. However, estimates of homozygosity are strongly affected by Mendelian sampling and the genomic relatedness between individuals in the BP [8]. The adjustment for variability in the BP does not allow complete correction of the estimates (Fig. 2). In this scenario, outliers can still occur, therefore suggesting that additional strategies (here, 'jackknifing over autosomes') should be applied to improve robustness of the estimates.

After correction for the mean homozygosity in the BP, the main differences in behaviour were found for F_{YAN} , confirming that this parameter is highly sensitive to this adjustment. There is no consensus on how to best define the BP. Some authors reported that using the actual allele frequencies in the BP can give inconsistent F_{YAN} values across generations with unrealistic patterns of increases or decreases in variability [15]. Other authors reported that fixing the

Table 3 Root-mean-squared error (RMSE) for the genomic-based N_e estimated for the five yearly and genealogical cohorts defined using the corresponding Ne_{F_i} and $Ne_{C_{ij}}$ values as references

	Ne_{F_i}		$Ne_{C_{ij}}$	
	BP-adjusted values	Jackknifed values	BP-adjusted values	Jackknifed values
Individual increase in homozygosity				
$\overline{Ne}_{F_{ROH_i}}$	3.2	3.3	3.0	3.1
$\overline{Ne}_{F_{HRR_i}}$	3.4	3.5	3.2	3.3
$\overline{Ne}_{F_{LH_i}}$	3.7	3.8	3.4	3.5
$\overline{Ne}_{F_{YAN_i}}$	3.0	3.0	2.8	2.8
Increase in pairwise homozygosity				
$\overline{Ne}_{F_{ROH_{ijk}}}$	2.6	2.5	2.5	2.5
$\overline{Ne}_{F_{HRR_{ijk}}}$	2.5	2.5	2.4	2.4
$\overline{Ne}_{F_{LH_{ijk}}}$	3.0	3.0	2.9	2.9
$\overline{Ne}_{F_{YAN_{ijk}}}$	3.0	3.0	2.9	2.9

allele frequencies in the base population to 0.5 [8, 68] allows stronger correlations of genealogical inbreeding estimates with genomic estimates computed from the genomic relationship matrix to be obtained. However, defining the current population as the BP yields variable correlations, even negative, with pedigree-based estimates [13].

Furthermore, lack of information about the real founders in a population can bias estimates of inbreeding parameters. This is particularly important when the aim is to compare estimates of homozygosity and inbreeding because the number of loci genotyped is always finite and the base (or founder) population has a different proportion of homozygous loci assumed to be IBS, whereas when using genealogies, founders are assumed to be unrelated and F is always dependent on the pedigree depth [69, 70]. Although small in size, our BP included two out of four of the founders of the population of Gochu Asturcelta and four direct descendants of the non-genotyped founders [32, 33]. Estimates of inbreeding that account for allele frequencies (F_{LH} and F_{YAN}) are expected to capture relationships due to the existence of common ancestors in a population, tracing the genetic variability of the present population back to the BP [16]. However, the literature suggests that founder allele frequencies in small populations have little meaning: in the absence of mutation, all current copies of an allele may be IBD, and its current frequency simply represents the reproductive success of the founder and its descendants [45] and could be related to the census size of breeders [13]. In our data, homozygosity tended to increase

significantly with pedigree depth (t), and BP adjustment resulted in acceptably consistent behaviour among homozygosity parameters (Fig. 3; Table 1). However, this approach may not be sufficient and jackknifing over autosomes suggested that additional adjustments would be necessary to obtain robust estimates; thus, BP-adjusted estimates of homozygosity may be biased upwards.

In general, our results suggest that BP adjustment does not remove differences in performance among homozygosity parameters, therefore suggesting that the genetic variability of a population should be assessed with caution. Of course, allele frequencies may be strongly affected by both editing of genotype data and the presence of allelic dropouts and null alleles in arrays [37], which can bias homozygosity estimates. Removal of rare alleles can cause an apparent increase in inbreeding due to IBS rather than IBD [9, 62]. To avoid this undesirable effect, we edited the SNP array data, filtering only loci that have Mendelian errors only, which have been previously shown to be mainly caused by technical issues and cause spurious diversity [37].

Demographic causes of variation in genomic estimates of homozygosity

In our population, homozygosity did not show clear trends of variation in relation to pedigree depth (Fig. 3). In fact, the accumulation of homozygosity with pedigree depth in inbred populations may be lower than expected when minimum coancestry matings are implemented, as in our case [32]. The Gochu Asturcelta conservation programme applied a strict mating policy to avoid matings between close relatives and to prolong the reproductive career of the founders and their direct descendants as much as possible to keep founder contributions balanced across generations and to preserve their genetic background in the population [32]. Theoretical and empirical evidence suggests that mating policies involving minimum coancestry in undivided populations balance allele frequencies at neutral genetic markers across generations [18, 23], therefore minimizing Mendelian segregation variance [22].

Although genomic estimates of inbreeding reflect the percentage of the genome that is homozygous, pedigree-based estimates are only expectations [20]. This suggests that the ability of homozygosity parameters to characterize the genetic variability of a population should be tested in each particular case, making it difficult to provide general guidelines. The results obtained in a particular population may be shaped by its particular breeding scenario. The literature suggests that between-individual variation in inbreeding estimators is strongly increased by presence of population structure [45] as well as by relatedness within the population, the mating policy applied, and the presence of large families increase [13, 58].

Increases in homozygosity and effective population size

Here, we evaluated two approaches to compute the increase in homozygosity. All estimates had large standard deviations (Table 1) as a consequence of the relatively small sample size and the strong effect of Mendelian sampling on the estimates of genomic inbreeding. The individual increase in homozygosity is conceptually similar to the individual increase in inbreeding (ΔF_i) proposed by Gutiérrez et al. [27, 28], which has been used as a standard for comparisons. Although ΔF_i allows a flexible definition of the reference population and is independent of mating policies [29, 41], it is based on the assumption that losses in diversity depend on pedigree depth only [71]. However, this assumption cannot be straightforwardly applied to individual increases in homozygosity: although founder individuals and their direct descendants in a pedigree are not inbred, all individuals in a pedigree have different degrees of homozygosity that can partially be explained by IBS. Our data suggest that the accumulation of the pedigree tends to increase homozygosity (Fig. 3). However, matings between relatives may be more important. In this regard, the increase in pairwise homozygosity could be intuitively considered a more accurate estimate of the increase in homozygosity due to IBD. In any case, increases in homozygosity are not constant across the genome [72]. IBD estimates in a limited genomic area or at the genome-wide level exhibit more variance than those computed for chromosomes [5]. To avoid this, jackknifing over autosomes was applied to the BP-adjusted genomic inbreeding estimates.

Although the differences were small, the results obtained after jackknifing suggest that BP-adjusted estimates of individual increases in homozygosity can be overestimates, while BP-adjusted values for increases in pairwise homozygosity can be underestimates (Table 1; Fig. 4). Values obtained for ΔtF_i displayed consistent variation with ΔF_i , which can be explained by their similar computation methods. The two approaches tested for increases in homozygosity showed consistent behaviour for F_{ROH} , F_{HRR} , and F_{LH} . However, the increases in homozygosity computed using F_{YAN} departed from this general behaviour. F_{YAN} is expected to have a smaller sampling variance than other estimates of homozygosity [54]. Our results confirmed this, as ΔtF_{YAN} had lower variation (54.1%) than ΔtF_{HRR} (105.9%).

The usefulness of the two approaches tested for increases in homozygosity to compute N_e has been compared with that of the genealogical “realized” N_e computed from ΔF_i but also with that derived from the increase in pairwise coancestry proposed by Cervantes et al. [73]. The last approach is expected to give more stable estimates of N_e in cases of shallow pedigrees due to the larger amount of information used in the computations.

Estimates of N_e obtained from genealogical and genomic information are expected to have large differences [16, 22, 68, 74, 75]. Indeed, criteria based on pedigree information refer to an infinite number of loci [71], while criteria based on observed genomic polymorphism mirror phenomena related to temporal changes in allele frequency at a limited number of loci [22, 76]. Furthermore, genealogical estimates of N_e , frequently used as a reference in the literature, may not reflect the actual N_e . Even with suitable genealogical data, genealogical N_e does not consider natural selection against homozygotes. Therefore, the genetic variability at the genomic level could be higher than expected. Although different genomic methods have been proposed to obtain reliable N_e estimates using half- and full-sib allele frequencies [77] and heterozygote excess [78], computation of reliable estimates of N_e is still a challenge. Mendelian sampling causes large differences in allele frequencies among close relatives (e.g., full-sibs) and at the population level. This may be particularly important in small populations with shallow pedigrees and frequent matings between relatives.

Genomic approaches are expected to give higher estimates of N_e than approaches based on genealogical data [14, 33], probably varying with the actual number of breeders in the sample [13]. This expectation fits well with the estimates of N_e obtained using individual increases in homozygosity. However, this is not so clear for increases in pairwise homozygosity. The assumption that an increase in homozygosity only depends on pedigree depth causes ΔtF_i values to underestimate the IBD scenario of the subpopulations analysed when compared with ΔpF_i . Our results suggest that pedigree depth itself may influence the accumulation of homozygosity (see Additional file 1: Table S3). However, this effect may be weaker than that of matings between relatives. In any case, jackknifing over autosomes can help avoid under- and overestimation of N_e . For individual increases in homozygosity, estimates of $\overline{N_e F_{YAN_i}}$ tended to be closer to both $N_e F_i$ and $N_e C_{ij}$ estimates. However, this was not the case for the increase in pairwise homozygosity, which was more similar to $\overline{N_e F_{ROH_{ijk}}}$ and, particularly, to $\overline{N_e F_{HRR_{ijk}}}$. The lower variance of the F_{YAN} values may lead to better adjustment of the influence of pedigree depth on increases in homozygosity.

Estimates of N_e based on ΔpF_{ROH} and, particularly, on ΔpF_{HRR} had lower RMSE values in the dataset. If ΔpF_{ijk} values are useful for characterizing increases in homozygosity due to IBD, our results suggest that increases in homozygosity would be caused by losses of small HRR genomic segments, which could be more sensitive to mating between relatives than to the increase in the length of long stretches of ROH.

In any case, as expected, estimates of genomic N_e were affected by cohort size: the larger the subpopulation assessed is, the higher the estimates of genomic N_e . In this regard, the values for the cohorts C2008 and CG3 had the highest estimates for a given method. The yearly cohort C2008 included several non-inbred individuals ($F=0$), which could lead to higher N_e estimates. However, this did not happen in CG3, therefore confirming some dependency on genomic N_e and sampling size. In the Gochu Asturcelta population, cohort sizes have been reported to be crucial for obtaining reliable estimates of N_e [33]. Our results, obtained using completely different information and methods, confirm this point.

Conclusions

Here, we analysed the ability of different genomic estimates of homozygosity to characterize the genetic background of a livestock population under mating policies aiming at conserving genetic variability. Our study showed that the mating policy applied to the Gochu Asturcelta pig population was successful in maintaining balanced allele frequencies. However, in such scenarios, parameters that characterize homozygosity are limited in their ability to depict trends in losses of variability. In spite of the presence of closely-related individuals from different families that vary in size, the four homozygosity parameters used (F_{ROH} , F_{HRR} , F_{LH} , and F_{YAN}) characterized the evolution of the pedigree to some extent. However, different adjustments were necessary to improve robustness of the estimates: although the adjustment of the homozygosity values by the mean of the BP appears indispensable, additional adjustment via jackknifing over autosomes helps to account for within-individual genome variation. Unlike genealogical approaches, the assumption that increases in homozygosity depend on pedigree depth only may not apply to small, shallow pedigrees, and an increase in pairwise homozygosity may better characterize autozygosity. Although increases in homozygosity computed using F_{YAN} seem to have good general behaviour, increases in pairwise homozygosity computed using F_{ROH} and F_{HRR} may be particularly useful if efficient identification of short ROH and HRR segments can be ensured.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12711-023-00846-7>.

Additional file 1: Table S1. Pedigree used and raw individual estimates of homozygosity. Description of the pedigree of the Gochu Asturcelta pigs used and raw estimates of homozygosity parameters computed at the individual level. **Table S2:** BP-adjusted homozygosity estimates. Description of the BP-adjusted homozygosity values and the corresponding increase in homozygosity parameters computed at the individual level.

Table S3: Regression of homozygosity against equivalent discrete generations. Regression of four homozygosity estimates and their corresponding individual increase in homozygosity (Δt) and paired increase in homozygosity (Δp) against equivalent discrete generations (t). **Table S4:** Jackknifed homozygosity values and their corresponding individual increases in homozygosity (Δt) and paired increases in homozygosity (Δp) values. Description of the jackknifed homozygosity values and the corresponding increase in homozygosity parameters computed at the individual level.

Acknowledgements

The authors kindly thank the members of the Gochu Asturcelta Breeders Association (ACGA; <https://www.gochuasturcelta.org/>) for their support and kind collaboration. The genotyping service was carried out at CEGEN-PRB3-ISCIII; it was supported by grant PT17/0019 of PE I+D+i 2013-2016, funded by ISCIII and ERDF. The authors are indebted to Dr. Gonçalo Carvalho (Thermo Fisher Scientific) for his willingness and kind contribution to the execution of this project.

Author contributions

FG and IA conceived and planned the project; KDA, FG, and JPG wrote the paper; KDA, FG, and IF performed the data analyses; KDA, FG, and JPG interpreted the models; IA and IF performed sampling and discussed and interpreted genetic data in light of the statistical and breeding evidence; and IA performed the laboratory work. All the authors read and approved the final manuscript.

Funding

This research was partially funded by AEI-FEDER, grant no. PID2019-103951RB/AEI/10.13039/501100011033. K.D.A. was funded by AEI-ESF, grant no. PRE2020-092905. For the purpose of open access, the author has applied a CC BY public copyright licence to any author-accepted manuscript version arising from this submission.

Availability of data and materials

All data necessary to replicate the results presented are provided as Additional Tables (.xlsx). SNP array data are currently under analysis within the "AutoGenome" project framework. However, the dataset used is 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.

Received: 27 January 2023 Accepted: 3 October 2023

Published online: 25 October 2023

References

- Wright S. Evolution in Mendelian populations. *Genetics*. 1931;16:97–159.
- Malécot G. *Les mathématiques de l'hérédité*. Paris: Masson et Cie; 1948.
- Cotterman CW. A calculus for statistico-genetics. PhD thesis, Ohio State University; 1940.
- Hill WG, Weir BS. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genet Res (Camb)*. 2011;93:47–64.
- Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet*. 2006;2: e41.
- Gómez-Raya L, Rodríguez C, Barragán C, Silió L. Genomic inbreeding coefficients based on the distribution of the length of runs of homozygosity in a closed line of Iberian pigs. *Genet Sel Evol*. 2015;47:81.

7. Morales-González E, Fernández J, Pong-Wong R, Toro MÁ, Villanueva B. Changes in allele frequencies when different genomic coancestry matrices are used for maintaining genetic diversity. *Genes* (Basel). 2021;12:673.
8. Wang J. Pedigrees or markers: which are better in estimating relatedness and inbreeding coefficient? *Theor Popul Biol*. 2016;107:4–13.
9. Wang J. Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. *J Evol Biol*. 2014;27:518–30.
10. 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.
11. Maltecca C, Tiezzi F, Cole JB, Baes C. Symposium review: Exploiting homozygosity in the era of genomics: selection, inbreeding, and mating programs. *J Dairy Sci*. 2020;103:5302–13.
12. 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.
13. Caballero A, Fernández A, Villanueva B, Toro MA. A comparison of marker-based estimators of inbreeding and inbreeding depression. *Genet Sel Evol*. 2022;54:82.
14. Silió L, Barragán C, Fernández AI, García-Casco J, Rodríguez MC. Assessing effective population size, coancestry and inbreeding effects on litter size using the pedigree and SNP data in closed lines of the Iberian pig breed. *J Anim Breed Genet*. 2016;133:145–54.
15. 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.
16. Rodríguez-Ramilo ST, Elsen JM, Legarra A. Inbreeding and effective population size in French dairy sheep: comparison between genomic and pedigree estimates. *J Dairy Sci*. 2019;102:4227–37.
17. Mulim HA, Brito LF, Pinto LFB, Ferraz JBS, Grigoletto L, Silva MR, et al. Characterization of runs of homozygosity, heterozygosity-enriched regions, and population structure in cattle populations selected for different breeding goals. *BMC Genomics*. 2022;23:209.
18. Royo LJ, Álvarez I, Gutiérrez JP, Fernández I, Goyache F. Genetic variability in the endangered Asturcón pony assessed using genealogical and molecular information. *Livest Sci*. 2007;107:162–9.
19. Saura M, Fernández A, Rodríguez MC, Toro MA, Barragán C, Fernández AI, et al. Genome-wide estimates of coancestry and inbreeding in a closed herd of ancient Iberian pigs. *PLoS One*. 2013;8:e78314.
20. Toro MA, Villanueva B, Fernández J. Genomics applied to management strategies in conservation programmes. *Livest Sci*. 2014;166:48–53.
21. Fernández J, Toro MA, Gómez-Romano F, Villanueva B. The use of genomic information can enhance the efficiency of conservation programs. *Anim Front*. 2016;6:59–64.
22. Wang J, Santiago E, Caballero A. Prediction and estimation of effective population size. *Heredity* (Edinb). 2016;117:193–206.
23. Fernández J, Villanueva B, Pong-Wong R, Toro MÁ. Efficiency of the use of pedigree and molecular marker information in conservation programs. *Genetics*. 2005;170:1313–21.
24. de Cara MAR, Fernández J, Toro MA, Villanueva B. Using genome-wide information to minimize the loss of diversity in conservation programmes: using genome-wide information to minimize the loss of diversity. *J Anim Breed Genet*. 2011;128:456–64.
25. Gómez-Romano F, Villanueva B, Rodríguez de Cara MÁ, Fernández J. Maintaining genetic diversity using molecular coancestry: the effect of marker density and effective population size. *Genet Sel Evol*. 2013;45:38.
26. Meuwissen THE, Sonesson AK, Gebregiwergis G, Woolliams JA. Management of genetic diversity in the era of genomics. *Front Genet*. 2020;11:880.
27. Gutiérrez JP, Cervantes I, Molina A, Valera M, Goyache F. Individual increase in inbreeding allows estimating effective sizes from pedigrees. *Genet Sel Evol*. 2008;40:359–78.
28. Gutiérrez JP, Cervantes I, Goyache F. Improving the estimation of realized effective population sizes in farm animals. *J Anim Breed Genet*. 2009;126:327–32.
29. Cervantes I, Goyache F, Molina A, Valera M, Gutiérrez JP. Application of individual increase in inbreeding to estimate realized effective sizes from real pedigrees. *J Anim Breed Genet*. 2008;125:301–10.
30. Arias KD, Lee H, Bozzi R, Álvarez I, Gutiérrez JP, Fernández I, et al. Ascertaining the genetic background of the Celtic-Iberian pig strain: a signatures of selection approach. *J Anim Breed Genet*. 2023; 17829268. <https://doi.org/10.1111/jbg.12829>.
31. Menéndez J, Goyache F, Beja-Pereira A, Fernández I, Menéndez-Arias NA, Godinho R, et al. Genetic characterisation of the endangered Gochu Asturcelta pig breed using microsatellite and mitochondrial markers: insights for the composition of the Iberian native pig stock. *Livest Sci*. 2016;187:162–7.
32. Menéndez J, Álvarez I, Fernández I, Goyache F. Genealogical analysis of the Gochu Asturcelta pig breed: insights for conservation. *Czech J Anim Sci*. 2016;61:140–9.
33. Menéndez J, Álvarez I, Fernández I, Menéndez-Arias NA, Goyache F. Assessing performance of single-sample molecular genetic methods to estimate effective population size: empirical evidence from the endangered Gochu Asturcelta pig breed. *Ecol Evol*. 2016;6:4971–80.
34. Menéndez J, Álvarez I, Fernández I, de la Roza B, Goyache F. Multiple paternity in domestic pigs under equally probable natural matings—a case study in the endangered Gochu Asturcelta pig breed. *Arch Anim Breed*. 2015;58:217–20.
35. Arias KD, Gutiérrez JP, Fernández I, Menéndez-Arias NA, Álvarez I, Goyache F. Segregation patterns and inheritance rate of copy number variations estimated from pedigree assessed in a Gochu Asturcelta pig pedigree. *Gene*. 2023;854: 147111.
36. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*. 2012;491:393–8.
37. Arias KD, Álvarez I, Gutiérrez JP, Fernández I, Menéndez J, Menéndez-Arias NA, et al. Understanding Mendelian errors in SNP arrays data using a Gochu Asturcelta pig pedigree: genomic alterations, family size and calling errors. *Sci Rep*. 2022;12:19686.
38. Gutiérrez JP, Goyache F. A note on ENDOG: a computer program for analysing pedigree information. *J Anim Breed Genet*. 2005;122:172–6.
39. Maignel L, Boichard D, Verrier E. Genetic variability of French dairy breeds estimated from pedigree information. *Interbull Bull*. 1996;14:49–53.
40. Meuwissen T, Luo Z. Computing inbreeding coefficients in large populations. *Genet Sel Evol*. 1992;24:305–13.
41. Cervantes I, Pastor JM, Gutiérrez JP, Goyache F, Molina A. Computing effective population size from molecular data: the case of three rare Spanish ruminant populations. *Livest Sci*. 2011;138:202–6.
42. McQuillan R, Leutenegger A-L, 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.
43. Biscarini F, Cozzi P, Gaspa G, Marras G. Package 'detectRUNS'. <https://cran.r-project.org/web/packages/detectRUNS/detectRUNS.pdf/>. Accessed 8 Sept 2023.
44. Biscarini F, Mastrangelo S, Catillo G, Senczuk G, Ciampolini R. Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in Maremmana semi-feral cattle using pedigree and genomic data. *Animals* (Basel). 2020;10:2285.
45. Thompson EA. Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics*. 2013;194:301–26.
46. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
47. Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, et al. Rldiogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci*. 2020;6: e251.
48. Williams JL, Hall SJG, Del Corvo M, Ballingall KT, Colli L, Ajmone Marsan P, et al. Inbreeding and purging at the genomic level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. *Anim Genet*. 2016;47:19–27.
49. Bizarria Dos Santos W, PimentaSchettini G, Gonçalves Fonseca M, Pereira GL, Loyola Chardulo LA, 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*. 2021;138:161–73.
50. Li CC, Horvitz DG. Some methods of estimating the inbreeding coefficient. *Am J Hum Genet*. 1953;5:107–17.
51. Nejat-Javaremi A, Smith C, Gibson JP. Effect of total allelic relationship on accuracy of evaluation and response to selection. *J Anim Sci*. 1997;75:1738–45.
52. Wright S. The genetical structure of populations. *Ann Eugen*. 1949;15:323–54.

53. Alemu SW, Kadri NK, Harland C, Faux P, Charlier C, Caballero A, et al. An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. *Heredity* (Edinb). 2021;126:410–23.
54. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42:565–9.
55. Bolar K. Package 'STAT'. <https://cran.r-project.org/web/packages/STAT/STAT.pdf>. Accessed 8 Sept 2023.
56. Powell JE, Visscher PM, Goddard ME. Reconciling the analysis of IBD and IBS in complex trait studies. *Nat Rev Genet*. 2010;11:800–5.
57. Efron B. The Jackknife, the bootstrap and other resampling plans. 6th ed. Philadelphia: Society for Industrial and applied mathematics; 1994.
58. Goudet J, Kay T, Weir BS. How to estimate kinship. *Mol Ecol*. 2018;27:4121–35.
59. Zhang Q, Calus MP, Guldbrandtsen B, Lund MS, Sahana G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genet*. 2015;16:88.
60. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4:7.
61. Peripolli E, Munari DP, Silva MVGB, Lima ALF, Irgang R, Baldi F. Runs of homozygosity: current knowledge and applications in livestock. *Anim Genet*. 2017;48:255–71.
62. Ferenčaković M, Sölkner J, Curik I. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet Sel Evol*. 2013;45:42.
63. Curik I, Ferenčaković M, Sölkner J. Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest Sci*. 2014;166:26–34.
64. Purfield DC, McParland S, Wall E, Berry DP. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS One*. 2017;12:e0176780.
65. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet*. 2006;15:789–95.
66. Sellì A, Ventura RV, Fonseca PAS, Buzanskas ME, Andrietta LT, Balieiro JCC, et al. Detection and visualization of heterozygosity-rich regions and runs of homozygosity in worldwide sheep populations. *Animals* (Basel). 2021;11:2696.
67. Kenny D, Carthy TR, Murphy CP, Sleator RD, Evans RD, Berry DP. The association between genomic heterozygosity and carcass merit in cattle. *Front Genet*. 2022;13: 789270.
68. Lozada-Soto EA, Maltecca C, Lu D, Miller S, Cole JB, Tiezzi F. Trends in genetic diversity and the effect of inbreeding in American Angus cattle under genomic selection. *Genet Sel Evol*. 2021;53:50.
69. Álvarez I, Royo LJ, Gutiérrez JP, Fernández I, Arranz JJ, Goyache F. Relationship between genealogical and microsatellite information characterizing losses of genetic variability: empirical evidence from the rare Xalda sheep breed. *Livest Sci*. 2008;115:80–8.
70. Fernández J, Meuwissen THE, Toro MA, Mäki-Tanila A. Management of genetic diversity in small farm animal populations. *Animal*. 2011;5:1684–98.
71. Leroy G, Mary-Huard T, Verrier E, Danvy S, Charvolin E, Danchin-Burge C. Methods to estimate effective population size using pedigree data: examples in dog, sheep, cattle and horse. *Genet Sel Evol*. 2013;45:1.
72. Gossmann TI, Woolfit M, Eyre-Walker A. Quantifying the variation in the effective population size within a genome. *Genetics*. 2011;189:1389–402.
73. Cervantes I, Gutiérrez JP, Molina A, Goyache F, Valera M. Genealogical analyses in open populations: the case of three Arab-derived Spanish horse breeds. *J Anim Breed Genet*. 2009;126:335–47.
74. Granado-Tajada I, Rodríguez-Ramilo ST, Legarra A, Ugarte E. Inbreeding, effective population size, and coancestry in the Latxa dairy sheep breed. *J Dairy Sci*. 2020;103:5215–26.
75. Charlesworth B. Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet*. 2009;10:195–205.
76. Toro MA, Meuwissen THE, Fernández J, Shaat I, Mäki-Tanila A. Assessing the genetic diversity in small farm animal populations. *Animal*. 2011;5:1669–83.
77. Wang J. A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Mol Ecol*. 2009;18:2148–64.
78. Pudovkin AI, Zaykin DV, Hedgecock D. On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics*. 1996;144:383–7.

Publisher's Note

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

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

