https://doi.org/10.17221/91/2024-CJAS

Commonly used genomic estimators of individual inbreeding in livestock

Adrián Halvoník, Nina Moravčíková, Monika Chalupková, Radovan Kasarda*

Institute of Nutrition and Genomics, Faculty of Agrobiology and Food Resources, Slovak University of Agriculture in Nitra, Nitra, Slovakia

**Corresponding author: xhalvonik@uniag.sk*

Citation: Halvoník A., Moravčíková N., Chalupková M., Kasarda R. (2024): Commonly used genomic estimators of individual inbreeding in livestock. Czech J. Anim. Sci., 69: 269–279.

Abstract: Management of inbreeding is one of the crucial parts of breeding programs in livestock populations. Traditionally, the inbreeding coefficient is calculated using pedigree data; however, it can also be estimated from genomic data. Nowadays, various approaches to estimating genomic-based inbreeding coefficients are increasingly integrated into research and breeding practices. These genomic estimators can supplement or replace pedigree-based coefficients. Each genomic-based inbreeding coefficient has its own properties and different ranges of values, and some of them need specific settings for calculation. Moreover, depending on the methodological approach, genomic estimators are sensitive to the population structure, genotyping technology applied, and the quality control of obtained genomic data. It is important to consider all these factors when calculating and especially when interpreting the final genomic inbreeding values. For these reasons, using genomic-based inbreeding coefficients can be more challenging than using pedigree-based ones. In this review, we comprehensively evaluate the most commonly used genomic estimators of individual inbreeding in livestock, providing an in-depth analysis of their advantages and limitations while offering insights into the methodological considerations and best practices for their accurate calculation and interpretation.

Keywords: genomic relationship matrix; inbreeding coefficient; PLINK; runs of homozygosity; single nucleotide polymorphism (SNP)

INTRODUCTION

An increase in inbreeding in populations leads to the expression of recessive disorders, reduced genetic diversity and a decline in selection response. Furthermore, increased levels of inbreeding may cause a reduction in the average value of traits, known as inbreeding depression, which has been documented across all livestock species (Leroy 2014; Doekes et al. 2019; Gutierrez-Reinoso et al. 2022). For these reasons, sufficient attention must be paid to managing inbreeding in livestock populations. Inbreeding can be described as an increase in autozygosity due to the mating of related individuals. In this context, autozygosity refers to homozygosity in which two alleles are identical by descent (IBD), i.e. originating from a common ancestor. The level of inbreeding is measured

Supported by the Slovak Research and Development Agency (Grants Nos. APVV-17-0060, APVV-20-0161 and DS-FR-22-0016)

https://doi.org/10.17221/91/2024-CJAS

by the inbreeding coefficient (F). All measures of individual inbreeding seek to predict the proportion of the genome that is IBD (Kardos et al. 2015). Wright (1922) defined the F as a correlation between the parents' uniting gametes and Malecot (1948) as the probability that two homologous alleles in an individual are IBD. Villanueva et al. (2021) used another definition of F regarding loss or gain of variability relative to a reference base population. To properly understand and interpret each F calculated from genomic data, we need to know which of the abovementioned definitions is used. Genomic estimators of individual inbreeding are usually divided into two main groups: Those analysing each single nucleotide polymorphisms (SNPs) separately (F_{SNP}, F_{HOM}, F_{UNI}, F_{GRM}) or those testing uninterrupted segments of SNPs (F_{ROH}). Another possibility is dividing F estimators based on the extent to which they depend on allele frequency (Dadousis et al. 2022).

This paper reviews the most commonly used estimators of individual inbreeding in livestock derived from genomic data. We highlighted their advantages and limitations and provided a useful summary of how to calculate and interpret the genomic-based inbreeding coefficient in livestock populations.

GENOMIC-BASED INBREEDING COEFFICIENTS

The most commonly used methods to measure F in livestock using genomic data are simple heterozygosity or homozygosity measures (F_{SNP}; Bjelland et al. 2013), estimators of the excess of homozygosity (F_{HOM}; Wright 1948) and correlation between uniting gametes (F_{UNI}; Yang et al. 2010), estimators based on the diagonal elements of a genomic relationship matrix (F_{GRM}; VanRaden 2008) and the proportion of runs of homozygosity (F_{ROH}; McQuillan et al. 2008).

An overview of frequently used software for calculating different F estimators is presented in Table 1. All the mentioned software can run on commonly used operating systems, including Windows, Linux, and MacOS. However, the author of GCTA recommends using only the Linux version.

Despite several sophisticated approaches to determine F, it is still unclear which genomic estimator provides the most accurate information about genetic variability and inbreeding depression. Yengo et al. (2017) supported the use of SNP-by-SNP-based F estimates with more weight for rare alleles, but e.g. Nietlisbach et al. (2019) preferred ROH-based F estimates. Caballero et al. (2020) claimed that the reliability of obtained results strongly depends on the particular population considered.

F based on homozygosity measures (F_{SNP})

The inbreeding coefficient based on individual SNP (F_{SNP}) is defined as the proportion of homozygous SNP in each animal (Silio et al. 2013). It can be calculated for the whole genome as follows:

$$F_{SNP} = \frac{N_{AA} + N_{aa}}{N_{AA} + N_{Aa} + N_{aa}} \quad (1)$$

where:
N_{AA} – the number of SNP classified as AA;
N_{Aa} – the number of SNP classified as Aa;
N_{aa} – the number of SNP classified as aa.

Values of F_{SNP} can range from 0 to 1, where 0 means that each SNP is heterozygous and 1 means that each SNP is homozygous.

Table 1. The most commonly used software running on Windows, Linux and MacOS for estimating genomic-based inbreeding coefficients

Software	F estimator
PLINK	F_{HOM}, F_{UNI}, F_{GRM}, ROH screening (SWM)
GCTA	F_{HOM}, F_{UNI}, F_{GRM}
R package detectRUNS	ROH screening (SWM, CSM), F_{ROH} calculation
R package RZooRoH	ROH screening (HMM)

CSM = consecutive single nucleotide polymorphism method; F_{GRM} = estimators based on the diagonal elements of a genomic relationship matrix; F_{HOM} = estimators of the excess of homozygosity; F_{UNI} = correlation between uniting gametes; HMM = method based on the hidden Markov model; SWM = sliding-window method

https://doi.org/10.17221/91/2024-CJAS

This method is based on the assumption that individuals with higher homozygosity will be more inbred. Calculation of F_{SNP} was used in a few studies to measure F and quantify inbreeding depression effects (Bjelland et al. 2013; Martikainen et al. 2017; Reverter et al. 2017; Dadousis et al. 2022). Although it can provide some information about the level of inbreeding, it does not distinguish between irritable bowel syndrome (IBS) and inflammatory bowel disease (IBD) markers (Bjelland et al. 2013). This approach can serve as information about genetic erosion but it does not provide a detailed view of the genetic architecture of the analysed population (Szulkin et al. 2010).

F based on the difference between observed and expected homozygosity (F_{HOM})

F_{HOM} is equivalent to Wright's fixation index (F_{IS}), ranging from –1 to 1. The calculation of F_{HOM} is based on the difference between the observed and expected number of homozygous genotypes. Expected homozygosity is estimated from allele frequencies, and observed homozygosity is directly derived from individual genotypes. It depends on the amount of genetic variation in the population and the level of inbreeding (Ritland 1996). Values higher than 0 represent an excess of homozygosity (inbreeding) (Schiavo et al. 2020). On the other hand, values lower than 0 represent an excess of heterozygosity (outbreeding) (Zhang et al. 2015; Xu et al. 2021). Villanueva et al. (2021) claimed that even if the F_{HOM} is not a probability or a correlation, it can be useful for determining whether variability is lost or gained. F_{HOM} can be calculated by the formula proposed by Wright (1948):

$$F_{HOM} = \frac{[O(\#hom) - E(\#hom)]}{[1 - E(\#hom)]} \quad (2)$$

where:

O (#hom) – the observed numbers of homozygous genotypes in the sample;

E (#hom) – the expected numbers of homozygous genotypes in the sample.

The most commonly used software for calculating F_{HOM} is PLINK (command *--het* and *--ibc*; Chang et al. 2015), where F_{HOM} is labelled as Fhat2, and GCTA (command *--ibc*; Yang et al. 2011), where F_{HOM} is labelled as $\hat{F}_{II}$. Both methods are based on the same approach and show similar results, where *--het* calculates F_{HOM} as follows:

$$F_{HOM} = 1 - \frac{\sum_{k=1}^{S} x_k(2 - x_k)}{\sum_{k=1}^{S} 2p_k(1 - p_k)} \quad (3)$$

where:

S – the total number of markers;

x_k – the number of minor alleles of marker k (i.e. 0, 1 or 2 copies);

p_k – the current frequency of the minor allele in the population.

The only difference between those two approaches is that the sumation over markers is made differently in each case; *--het* is a ratio of sums, whereas *--ibc* is a sum of ratios.

F based on the correlation between uniting gametes (F_{UNI})

Inbreeding estimator F_{UNI} is based on the correlation between uniting gametes (Yang et al. 2010). It is directly related to a definition proposed by Wright (1922) and can be calculated as follows:

$$F_{UNI} = \frac{1}{S}\sum_{k=1}^{S} \frac{x_k^2 - (1 + 2p_k)xk_i + 2p_k^2}{2p_k(1 - p_k)} \quad (4)$$

where:

S – the total number of markers;

x_k – the genotype of the individual for SNP;

k – coded as 0, 1 or 2 for genotypes *AA*, *AB* and *BB*;

p_k – the frequency of the reference allele (allele *B*) of SNP k in the base (reference) population.

This estimator is implemented in PLINK (Chang et al. 2015) and GCTA (Yang et al. 2011) software (Fhat3 and $\hat{F}_{III}$ by command *--ibc*. Nietlisbach et al. (2019) claimed that the values F_{UNI} can range from –∞ to ∞. However, Villanueva et al. (2021) and Dadousis et al. (2022) reported that the values of F_{UNI} can range from –1 to ∞ and results indicate the loss (positive values) or gain (negative values) of variability in the genome compared to the reference population. In general, higher F_{UNI} values indicate a higher inbreeding rate in animals,

https://doi.org/10.17221/91/2024-CJAS

while lower (or negative) values suggest a lower rate of inbreeding.

This method is widely used in evaluating livestock populations, although it usually serves only as supplementary information and conclusions are based on other approaches (mainly F_{ROH}). Studies that have looked deeper into a pattern of F_{UNI} and examined its advantages and disadvantages have been conducted on simulated data or small populations (Yengo et al. 2017; Kardos et al. 2018; Nietlisbach et al. 2019; Caballero et al. 2020; Villanueva et al. 2021). F_{UNI} based on correlations between genetic effects gives more weight to homozygosity at rare alleles (Keller et al. 2011). Alemu et al. (2021) claimed that F_{UNI} could have a strong association with inbreeding depression. However, the situation might differ when recessive deleterious alleles reach higher frequencies like in populations with the low effective population size (N_e). Similar recommendations were provided by Nietlisbach et al. (2019) and Caballero et al. (2020). In populations with large N_e, where animals are less related and deleterious alleles are expected to be at lower frequencies, F_{UNI} provides almost unbiased average estimations of inbreeding depression. Dadousis et al. (2022) said that F_{UNI} is hard to interpret for a practical application in populations under intense directional selection. However, this estimator can provide some information about lost or gained variability (Villanueva et al. 2021).

F based on genomic relationship matrix (F_{GRM})

VanRaden (2008) presented multiple approaches to *F* calculation based on the diagonal element of the genomic relationship matrix. The first one, often referred to as VanRaden1 (VR1), can be calculated as follows:

$$F_{GRM} = \frac{\sum_{k=1}^{S}(x_k - 2p_k)}{2\sum_{k=1}^{S} p_k(1-p_k)} \quad (5)$$

where:

S – the total number of markers;

x_k – the genotype of the individual for SNP;

p_k – the frequency of the reference allele (allele *B*) of SNP *k* in the base (reference) population.

This approach weighted homozygous genotypes by the inverse of their allele frequency and, thus, rare homozygous genotypes contribute more to the inbreeding measure than common homozygous genotypes (Villanueva et al. 2021). The second approach, often referred to as VanRaden2 (VR2), can be calculated as follows:

$$F_{GRM} = \frac{1}{S}\sum_{k=1}^{S}\left(\frac{(x_k - 2p_k)^2}{2p_k(1-p_k)} - 1\right) \quad (6)$$

where:

S – the total number of markers;

x_k – the genotype of the individual for SNP;

p_k – the frequency of the reference allele (allele *B*) of SNP *k* in the base (reference) population.

This method is very similar to VR1, but the summation across markers is made differently, so the weight given to rare alleles is even greater. In this case, the contribution of each SNP is divided by its own variance, whereas in VR1, the contributions of all SNPs are divided by the same denominator. This method is included in PLINK (Chang et al. 2015) and GCTA (Yang et al. 2011) as Fhat1 and $\hat{F}_I$ under *--ibc* command. In general, when we compare results of inbreeding coefficient derived from VR1 and VR2, they behave similarly. Although values derived from VR2 are more extreme (Villanueva et al. 2021), F_{GRM} results highly depend on allele frequency assumptions. Optimally, allele frequencies should represent frequencies in the founder population. Because these frequencies are mostly unknown, they are replaced by the allele frequency of evaluated population or set to a fixed value. We need to point out that if the evaluated sample is small, the allele frequency can be very inaccurate, and the F_{GRM} results can be misleading. A simulation study by Forutan et al. (2018) showed that using a fixed allele frequency of 0.5 can be beneficial. Also, the authors who performed analyses on real populations argue that using fixed allele frequencies can be helpful (VanRaden et al. 2011; Doekes et al. 2019; Lozada-Soto et al. 2022). A more extensive overview of the usage of different GRM utilising fixed or real allele frequency was provided by Villanueva et al. (2021) and Dadousis et al. (2022).

VR2 without fixed allele frequency is often used in current studies, probably due to the implementation of this method in commonly used software. Villanueva et al. (2021) and Dadousis et al. (2022) claimed that values of this estimator can range from −1 to ∞, where positive values signal loss of variability and negative values signal gain of variability

https://doi.org/10.17221/91/2024-CJAS

relative to the base (reference) population. In general, the interpretation is same as for F_{UNI}. The higher F_{GRM} values indicate higher inbreeding in animals, while lower (or negative) values suggest lower inbreeding. Zhang et al. (2015) reported that if the population has a high level of heterozygosity and some rare alleles with low frequency, F_{GRM} will yield high positive inbreeding coefficients, which can be misleading. However, this is unusual in livestock populations under directional selection.

F based on runs of homozygosity (F_{ROH})

ROH are defined as continuous homozygous segments in the DNA sequence, and the general formula for calculating F_{ROH} is as follows:

$$F_{ROH} = \frac{L_{ROH}}{L_{autosome}} \quad (7)$$

where:

L_{ROH} – the sum of individual lengths of ROH;

$L_{autosome}$ – the total length of autosomes covered by SNPs.

Results of F_{ROH} can range from 0 to 1 and can be interpreted as a proportion of the genome that is IBD. Caballero et al. (2020) showed that for populations with low N_e, where individuals are more related to each other, F_{ROH} provides more accurate inbreeding depression estimates than other *F* estimators.

One of the main advantages of F_{ROH} is the ability to separately estimate recent and historical inbreeding by calculating F_{ROH} from ROH of different lengths. This can be very useful due to the different effects of recent and historical inbreeding on the population gene pool (Doekes et al. 2019). Short ROH indicates historical connectedness, and long ROH may reveal recent inbreeding (McQuillan et al. 2008). Ferencakovic et al. (2013a) placed ROH in five length classes: 1 to 2, 2 to 4, 4 to 8, 8 to 16 and >16 Mb, where ROH > 1, > 2, > 8 and > 16 Mb, representing up to ≈ 50, ≈ 25, ≈ 6 and ≈ 3 generations from common ancestor, respectively. However, this distribution of ROH length categories was addressed to cattle, and it is questionable whether its use in other species leads to sufficiently accurate results.

In general, we can say that the most problematic part of the F_{ROH} calculation is the detection of ROH in the genome. There are three methods to detect ROH: sliding-window method (SWM), consecutive SNP method (CSM) and method based on the hidden Markov model (HMM). The chosen method and its specific settings significantly impact the number and length of detected ROH. The first method, using a sliding window to scan the genome, is described in Figure 1.

The second method is window-free and it directly scans the genome of SNP by SNP. It was first proposed by Marras et al. (2015). In both approaches, multiple parameters need to be pre-set, which is a crucial part of ROH detection. The optimal value of run-related parameters depends on the species, number and length of chromosomes, genome length, and density of SNP chips. The parameters that need to be set before the analysis are summarised in Table 2. Several studies tested the effect of different settings that directly affect ROH screening (Hillestad et al. 2018; Meyermans et al. 2020; Macciotta et al. 2021) and F_{ROH} calculation (Yengo et al. 2017; Mulim et al. 2022) but there has been no standard yet. This fact makes it difficult to compare results from different studies, as mentioned by Peripolli et al. (2017).

In the case of cattle, Ferencakovic et al. (2013a) proposed the most commonly used settings, but some authors still use their own settings based on an empirical approach. In other species, no optimal settings have been established.

The minimal number of SNPs in ROH (*L*) can be determined by the formula proposed by Lencz et al. (2007) and adapted by Purfield et al. (2012):

$$L = \frac{\log_e \frac{\alpha}{n_s n_i}}{\log_e (1 - het)} \quad (8)$$

where:

α – the percentage of false positive ROH;

het – average heterozygosity across all SNPs;

n_i – the number of genotyped individuals;

n_s – the number of genotyped SNPs per individual.

Meyermans et al. (2020) proposed a formula for the calculation of the scanning window threshold (*t*) as follows:

$$t = \text{floor}\left(\frac{N_{out} + 1}{L}, 3\right) \quad (9)$$

where:

N_{out} – the desired number of final outer SNPs on either side of the homozygous segment that should not

https://doi.org/10.17221/91/2024-CJAS

be included in the final ROH;
L – the scanning window size.

In this formula, "+ 1" denotes the first SNP that will be tolerated of the final ROH and ",3" points at flooring with three decimals. For example, with L = 100 and N_{out} = 4, the threshold will be set at 0.05. This will lead to scanning the windows of 100 SNPs and in the obtained homozygous segment we discard the four outer SNPs on each side of the homozygous segment.

We reviewed 48 studies focusing on ROH detection based on medium-density SNP data (29 using SWM, 19 using CSM) to identify commonly used settings. The species included in these studies were cattle (29), pigs (8), goats (6), and sheep (5). Figure 2 shows that 1 Mb is the most commonly used minimum length for the ROH and the gap between two SNPs. A length of 100 kb is the most preferred SNP density. Regarding the minimum number of SNPs in ROH, it is commonly required that ROH contain at least 15 or 30 SNPs. The authors

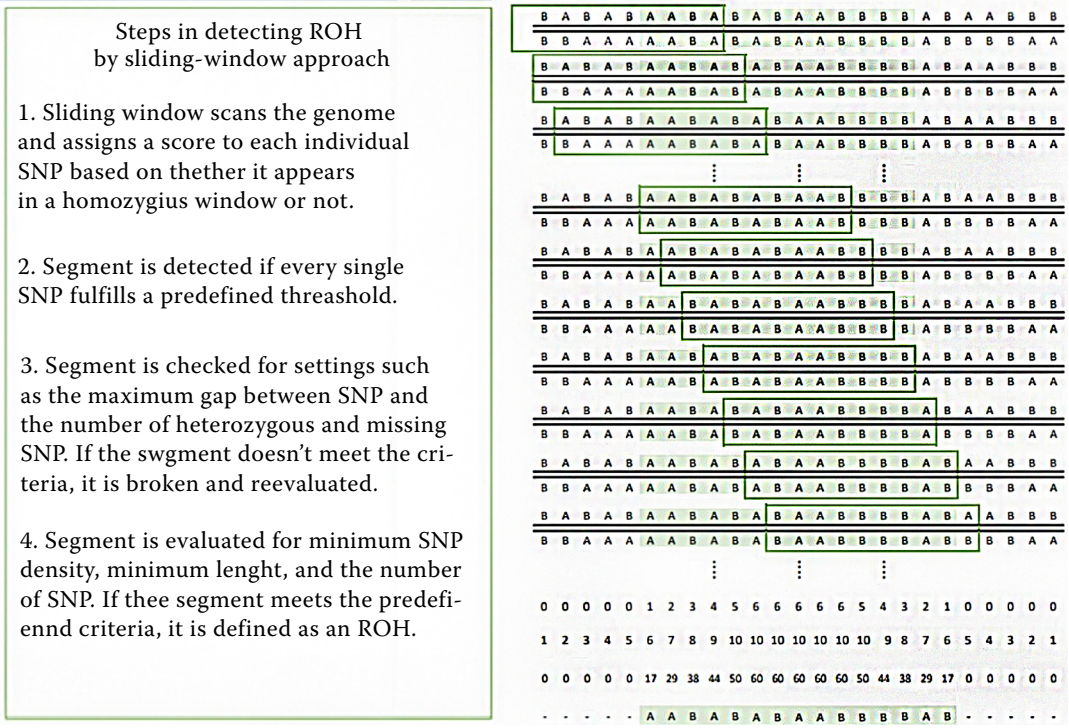

Figure 1. Description of the sliding-window approach
ROH = runs of homozygosity; SNP = single nucleotide polymorphism
Source: Bjelland et al. (2013); Winnberg (2020)

Table 2. Run-related parameters for ROH detection

Parameter	Sliding-window method	Consecutive SNP method
Minimum number of SNPs included in ROH	✓	✓
Minimum length of ROH in bp	✓	✓
The gap between two SNPs to be considered as a ROH	✓	✓
Number of heterozygous calls allowed in ROH	✓	✓
Number of missing calls allowed in ROH	✓	✓
Minimum number of 1 SNP per kb	✓	×
Number of heterozygous calls allowed in a window	✓	×
Number of missing calls allowed in a window	✓	×
Number of SNPs that the sliding window must have	✓	×
Window of threshold a sliding genomic window	✓	×

ROH = runs of homozygosity; SNP = single nucleotide polymorphism

https://doi.org/10.17221/91/2024-CJAS

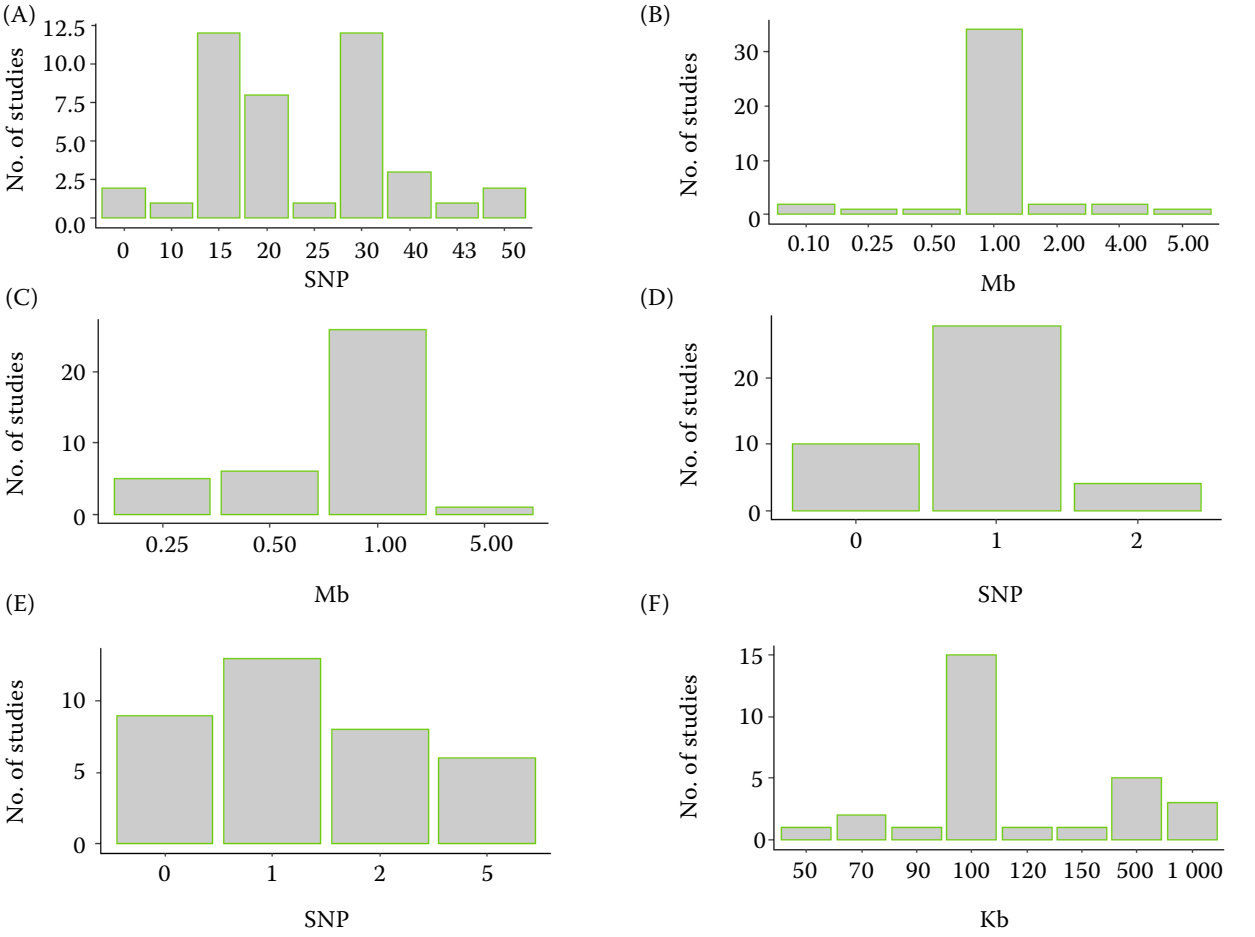

Figure 2. Review of settings for detecting ROH

(A) minimum number of SNPs included in ROH, (B) minimum length of ROH, (C) the gap between two SNPs to be considered as ROH, (D) number of heterozygous calls allowed in ROH, (E) number of missing calls allowed in ROH, and (F) minimum number of 1 SNP per kb

ROH = runs of homozygosity; SNP = single nucleotide polymorphism

most often allow 1 heterozygous and 1 missing SNP in ROH. It is necessary to mention that the number of heterozygous and missing SNPs allowed in ROH should depend on the ROH length. Although many authors used the same settings for all length categories, we highly recommend using the specific number for each length category of ROH like Ferencakovic et al. (2013b) or Nosrati et al. (2021). Because the authors do not often mention in the methodology all settings used for ROH detection, it was not possible to summarise the most commonly used settings for all run-related parameters.

ROH detection based on HMM is the newest method. This model describes the individual genome as mosaics of multiple homozygosity-by-descent (HBD) and non-HBD classes (k). The different HBD classes are defined by their specific rates R_k. The length of HBD segments from class k is exponentially distributed with the rate R_k and mean $1/R_k$. To compute the probability of such a sequence, the model requires the probability to stay in the current segment or to start a new segment between two consecutive markers and the probability to observe the particular genotype conditionally on the class specificities. The probability to continue a segment is, where R_k is the rate specific to class k. As a result, the length of HBD segments from class k is exponentially distributed with the rate R_k. The expected length is then equal to $1/R_k$ morgans (Bertrand et al. 2019).

One of the important steps before detecting ROH is quality control of the genomic data, primarily minor allele frequency (MAF) pruning. In general, when analysing SNP data, monomorphic SNPs are considered uninformative and should be removed during quality control. Usually, the value of MAF

https://doi.org/10.17221/91/2024-CJAS

pruning included in quality control ranged from 0.01 to 0.05. However, when detecting ROH by SWM or CSM, there is no consensus on whether we should include MAF pruning in quality control. Howrigan et al. (2011) recommended to include MAF pruning in data quality control based on their research performed on simulated data. However, a recent study by Meyermans et al. (2020) showed that MAF pruning can overlook large homozygous regions. Therefore, they do not recommend including MAF in data quality control before ROH analysis. There is a probability that if monomorphic markers are excluded due to MAF pruning, the segment should not be considered as ROH because it did not meet the criteria such as minimum of SNP per kb, maximum gap between two SNPs, or minimum number of SNPs included in the ROH. In the case of HMM, Druet et al. (2019) claimed that this approach is not sensitive to MAF pruning, and they found only negligible differences when applying different settings in MAF pruning. Due to the fact that the evidence and explanation of different settings for MAF filtering in ROH detection are inconsistent, further research is necessary to fully explain the optimal approach depending on the density of genomic data and species analysed.

Ferencakovic et al. (2013b) pointed out that the density of the SNP chips used to generate the data for ROH analysis and also the frequency of SNP genotyping errors can influence ROH identification. Many authors (Purfield et al. 2012; Aliloo et al. 2018; Ceballos et al. 2018) reported that the use of at least a 50k panel in detecting ROH with SWM and CSM is recommended. The use of low-density panels can underestimate the number of long ROH due to the low number of SNP markers and the high physical distance between them (Mulim et al. 2022). On the other hand, the 50k panel overestimated the number of small segments (1–4 Mb long) (Purfield et al. 2012; Ferencakovic et al. 2013b). Marras et al. (2015) pointed out that the use of medium-density chips may provide good estimates in populations with recent inbreeding and high linkage disequilibrium (LD), but the precise detection of autozygosity in populations with more "ancient" inbreeding and low LD will require high-density data. Therefore, the minimum ROH length that can be detected depends on the density of the SNP chips (Ferencakovic et al. 2013b).

Another factor that can influence the minimum ROH length that can be accurately detected is an applied statistical approach to ROH screening. HMM can be more accurate than SWM and CSM for analysing the genomic data that are sparser or less accurate, such as low coverage sequencing data, or low-density SNP data (Druet and Gautier 2017). This statement is also supported by the study conducted by Lavanchy and Goudet (2023) where they compared the accuracy of PLINK and RZooRoH in ROH detection. RZooRoH requires a significantly lower proportion of SNPs per Mb (2 SNPs/Mb compared to PLINK (22 SNPs/Mb) to precisely estimate inbreeding coefficients. Sole et al. (2017) claimed that despite using low-density SNP arrays (7K and 32K), HMM can efficiently capture recent autozygosity. The correlation between inbreeding coefficients obtained with low-density arrays and those from higher density (600K SNPs) was 0.934 for 7K and 0.975 for 32K. This can serve as an advantage of HMM. However, the computation time of this method is much longer than that of CSM or SWM, which can be a disadvantage of HMM (Ceballos et al. 2018).

The SNP genotyping error is another factor that needs to be considered. In SWM and CSM it is possible to consider the bias due to potential genotyping errors depending on the allowance of heterozygous genotypes in ROH. If we do not allow any heterozygous genotype in ROH, there is a possibility of splitting a very long ROH into two shorter ROH. In this case, these segments will still be considered as ROH, and the overall level of autozygosity will be unaffected. However, if this happens to short segments, they may not be able to reach the minimum size of a ROH, and the autozygosity of the examined animal will be underestimated. On the other hand, allowing too many heterozygous genotypes in ROH can lead to the detection of short segments that are most likely not autozygous (Ferencakovic et al. 2013b). When it comes to considering the genotyping error in ROH detection with HMM, it is necessary to specify it before detection. The authors mainly used the value of the genotyping error of 0.25% specified by Ferencakovic et al. (2013b).

CONCLUSION

Although each *F* estimator provides valuable insights into the genomic inbreeding levels within the studied population, they exhibit slightly

https://doi.org/10.17221/91/2024-CJAS

different features and usage. Based on recent studies in livestock, F_{ROH} provides the most reliable parameter for estimating individual inbreeding as well as tracking inbreeding trends, particularly with high-density genomic data and proper settings. On the other hand, F_{UNI} seems to be the most problematic of all the estimators described. Based on our research, we do not recommend using it in livestock populations without proper knowledge and experience. We hope our review helps better understand the features of commonly used genomic-based inbreeding coefficients and provides a short guide to their calculation and interpretation in livestock populations.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

Alemu SW, Kadri NK, Harland C, Faux P, Charlier C, Caballero A, Druet T. An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. Heredity (Edinb). 2021 Mar;126(3):410-23.

Aliloo H, Mrode R, Okeyo AM, Ni G, Goddard ME, Gibson JP. The feasibility of using low-density marker panels for genotype imputation and genomic prediction of crossbred dairy cattle of East Africa. J Dairy Sci. 2018 Oct; 101(10):9108-27.

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 Feb;10(6):860-6.

Bjelland DW, Weigel KA, Vukasinovic N, Nkrumah JD. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. J Dairy Sci. 2013 Jul;96(7):4697-706.

Caballero A, Villanueva B, Druet T. On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. Evol Appl. 2020 Oct 23; s14(2):416-28.

Ceballos FC, Hazelhurst S, Ramsay M. Assessing runs of homozygosity: A comparison of SNP array and whole genome sequence low coverage data. BMC Genomics. 2018 Jan 30;19(1):106.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience. 2015 Feb 25;4:7.

Dadousis C, Ablondi M, Cipolat-Gotet C, van Kaam JT, Marusi M, Cassandro M, Sabbioni A, Summer A. Genomic inbreeding coefficients using imputed genotypes: Assessing different estimators in Holstein-Friesian dairy cows. J Dairy Sci. 2022 Jul;105(7):5926-45.

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 Sep 27;51(1):54.

Druet T, Gautier M. A model-based approach to characterize individual inbreeding at both global and local genomic scales. Mol Ecol. 2017 Oct;26(20):5820-41.

Druet T, Bertrand A, Kadri N, Gautier M. The RZooRoH package. 2023. Available from https://mirror.rcg.sfu.ca/mirror/CRAN/web/packages/

Ferencakovic M, Hamzic E, Gredler B, Solberg TR, Klemetsdal G, Curik I, Solkner J. Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. J Anim Breed Genet. 2013a Aug;130(4):286-93.

Ferencakovic M, Solkner J, Curik I. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genet Sel Evol. 2013b Oct 29; 45(1):42.

Forutan M, Ansari Mahyari S, Baes C, Melzer N, Schenkel FS, Sargolzaei M. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. BMC Genomics. 2018 Jan 27;19(1):98.

Gutierrez-Reinoso MA, Aponte PM, Garcia-Herreros M. A review of inbreeding depression in dairy cattle: Current status, emerging control strategies, and future prospects. J Dairy Res. 2022 Feb;28:1-10.

Hillestad B, Woolliams JA, Boison SA, Grove H, Meuwissen T, Vage DI, Klemetsdal G. Detection of runs of homozygosity in Norwegian Red: Density, criteria and genotyping quality control. Acta Agr Scand A-An. 2018 Jul;67(3-4):107-16.

Howrigan DP, Simonson MA, Keller MC. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. BMC Genomics. 2011 Sep 23;12:460.

Kardos M, Luikart G, Allendorf FW. Measuring individual inbreeding in the age of genomics: Marker-based measures are better than pedigrees. Heredity (Edinb). 2015 Jul;115(1):63-72.

Kardos M, Nietlisbach P, Hedrick PW. How should we compare different genomic estimates of the strength of inbreeding depression? Proc Natl Acad Sci U S A. 2018 Mar 13;115(11):E2492-E3.

Keller MC, Visscher PM, Goddard ME. Quantification of inbreeding due to distant ancestors and its detection using

https://doi.org/10.17221/91/2024-CJAS

dense single nucleotide polymorphism data. Genetics. 2011 Sep;189(1):237-49.

Lavanchy E, Goudet J. Effect of reduced genomic representation on using runs of homozygosity for inbreeding characterization. Mol Ecol Resour. 2023 May;23(4):787-802.

Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, Kucherlapati R, Malhotra AK. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc Natl Acad Sci U S A. 2007 Dec 11;104(50): 19942-7.

Leroy G. Inbreeding depression in livestock species: Review and meta-analysis. Anim Genet. 2014 Oct;45(5):618-28.

Lozada-Soto EA, Tiezzi F, Jiang J, Cole JB, VanRaden PM, Maltecca C. Genomic characterization of autozygosity and recent inbreeding trends in all major breeds of US dairy cattle. J Dairy Sci. 2022 Nov;105(11):8956-71.

Macciotta NPP, Colli L, Cesarani A, Ajmone-Marsan P, Low WY, Tearle R, Williams JL. The distribution of runs of homozygosity in the genome of river and swamp buffaloes reveals a history of adaptation, migration and crossbred events. Genet Sel Evol. 2021 Feb 27;53(1):20.

Malecot G. Les mathematiques de l'heredite [The mathematics of heredity]. Paris: Masson; 1948. 63 p. French.

Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, Williams JL, Macciotta NP. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim Genet. 2015 Apr;46(2):110-21.

Martikainen K, Tyriseva AM, Matilainen K, Poso J, Uimari P. Estimation of inbreeding depression on female fertility in the Finnish Ayrshire population. J Anim Breed Genet. 2017 Oct;134(5):383-92.

McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, Macleod AK, Farrington SM, Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H, Wilson JF. Runs of homozygosity in European populations. Am J Hum Genet. 2008 Sep;83(3):359-72.

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 Jan 29;21(1):94.

Mulim HA, Brito LF, Pinto LFB, Ferraz JBS, Grigoletto L, Silva MR, Pedrosa VB. Characterization of runs of homozygosity, heterozygosity-enriched regions, and population structure in cattle populations selected for different breeding goals. BMC Genomics. 2022 Mar 16;23(1):209.

Nosrati M, Nanaei HA, Javanmard A, Esmailizadeh A. The pattern of runs of homozygosity and genomic inbreeding in world-wide sheep populations. Genomics. 2021 May;113(3):1407-15.

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 Oct 23;12(2):266-79.

Peripolli E, Munari DP, Silva MVGB, Lima ALF, Irgang R, Baldi F. Runs of homozygosity: Current knowledge and applications in livestock. Anim Genet. 2017 Jun;48(3):255-71.

Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. BMC Genet. 2012 Aug 14;13:70.

Reverter A, Porto-Neto LR, Fortes MRS, Kasarapu P, de Cara MAR, Burrow HM, Lehnert SA. Genomic inbreeding depression for climatic adaptation of tropical beef cattle. J Anim Sci. 2017 Sep;95(9):3809-21.

Ritland K. Estimators for pairwise relatedness and individual inbreeding coefficients. Genet Res. 1996 Apr;67(2):175-85.

Schiavo G, Bovo S, Bertolini F, Tinarelli S, DallOlio S, Nanni Costa L, Gallo M, Fontanesi L. Comparative evaluation of genomic inbreeding parameters in seven commercial and autochthonous pig breeds. Animal. 2020 May; 14(5):910-20.

Silio L, Rodriguez MC, Fernandez A, Barragan C, Benitez R, Ovilo C, Fernandez AI. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. J Anim Breed Genet. 2013 Oct;130(5):349-60.

Sole M, Gori AS, Faux P, Bertrand A, Farnir F, Gautier M, Druet T. Age-based partitioning of individual genomic inbreeding levels in Belgian Blue cattle. Genet Sel Evol. 2017 Dec 22;49(1):92.

Szulkin M, Bierne N, David P. Heterozygosity-fitness correlations: A time for reappraisal. Evolution. 2010 May; 64(5):1202-17.

VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008 Nov;91(11):4414-23.

VanRaden PM, Olson KM, Wiggans GR, Cole JB, Tooker ME. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. J Dairy Sci. 2011 Nov;94(11): 5673-82.

Villanueva B, Fernandez A, Saura M, Caballero A, Fernandez J, Morales-Gonzalez E, Toro MA, Pong-Wong R. The value of genomic relationship matrices to estimate levels of inbreeding. Genet Sel Evol. 2021 May 1;53(1):42.

Winnberg K. Trends in runs of homozygosity and inbreeding in Norwegian Red Cattle before and after implementation of genomic selection [master's thesis]. (As): Norwegian University of Life Sciences; 2020. p 8.

Wright S. . Genetics of populations. In: Yust W, editor. Encyclopaedia Britannica. Chicago: Encyclopædia Britannica; 1948. p. 111-2.

https://doi.org/10.17221/91/2024-CJAS

Wright S. Coefficients of inbreeding and relationship. Am Nat. 1922 Jul;56(645):330-8.

Xu Z, Mei S, Zhou J, Zhang Y, Qiao M, Sun H, Li Z, Li L, Dong B, Oyelami FO, Wu J, Peng X. Genome-wide assessment of runs of homozygosity and estimates of genomic inbreeding in a chinese composite pig breed. Front Genet. 2021 Sep 1;12:720081.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010 Jul;42(7):565-9.

Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. 2011 Jan 7;88(1):76-82.

Yengo L, Zhu Z, Wray NR, Weir BS, Yang J, Robinson MR, Visscher PM. Detection and quantification of inbreeding depression for complex traits from SNP data. Proc Natl Acad Sci U S A. 2017 Aug 8;114(32):8602-7.

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 Jul 22;16:88.

Received: June 6, 2024
Accepted: July 18, 2024
Published Online: Aug 1, 2024