*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

# PROPERTIES OF RUNS OF HOMOZYGOSITY AS A MEASURE OF IDENTITY BY DESCENT

Oda B. Wæge[1], Tom Druet[2], Peer Berg[1], Theo Meuwissen[1]

[1]*Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway*
[2]*Unit of Animal Genomics, GIGA- R & Faculty of Veterinary Medicine, University of Liège, Liège, Belgium*
*Correspondence: Oda B. Wæge (oda.braar.wage@nmbu.no)*

**Abstract**

Runs of Homozygosity (ROH) are commonly used to quantify autozygosity/identity- by- descent (IBD) in an individual or population. However, the method's accuracy at the segment level in livestock populations has only been evaluated in a few studies. Thus, the aim of this study was to determine to what extent ROH are truly IBD and estimate the proportion of IBD segments that go undetected in a simulated livestock population. We simulated a population of randomly mating animals for 100 generations. The genome consisted of a single chromosome with a SNP density of either 46 or 92 SNPs per mega base (Mb). In addition, a set of founder markers tracing IBD was recorded. ROH were detected using four different parameter combinations. Using the two sets of markers, we calculated the true positive rate, power, and overall correlation between true ($F_{IBD}$) and estimated inbreeding ($F_{ROH}$). Additionally, a new measure for within- ROH inbreeding (F|ROH) was introduced and calculated the level of homozygosity within a ROH compared to the general expectation in the genome. The results indicate that ROH longer than 2 Mb are a reliable indicator of IBD, with the F|ROH being over 0.9 for all ROH lengths and parameter combinations. True positive rates only exceeded 0.9 consistently for ROH over 9 Mb, indicating that many of the identified ROH may be associated with common ancestors more ancient than the base population. The power was mainly controlled by the parameter stringency, that is, allowing for shorter ROH increased the power. The ROH- based individual measure of inbreeding $F_{ROH}$ was highly correlated to $F_{IBD}$ while also having regression coefficients close to 1 (i.e., a 1% variation in $F_{ROH}$ corresponded to a 1% variation in $F_{IBD}$). Using stringent ROH parameters resulted in underestimation of the rate of inbreeding in the population. Increasing marker density improved predictions, including a higher true positive rate, power, higher correlations, and less underestimation of inbreeding rates.

## 1 Introduction

In her review, Thompson (2013) defines two DNA segments in different current gametes as identical- by- descent (IBD) if they are a copy of a single segment of DNA in an ancestral individual. Further, she states that IBD is a measure that changes in relation to a specific ancestral population referred to as the base, reference, or founder population where all segments are assumed non- IBD. For example, in an experimental or livestock population this could correspond to a founding stock of animals. IBD is used in different applications such as mapping genes associated with disease (Houwen et al. 1994; Te Meerman et al. 1995), QTL mapping (Meuwissen et al. 2001; Riquet et al. 1999), estimating relatedness between individuals (Malécot 1948), past demographic history (Harris and Nielsen 2013; Palamara et al. 2012),

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

optimal genetic diversity management of populations (Meuwissen 1997), and the study of fundamental processes such as mutation and recombination rates (Zhou et al. 2020).

Within individuals, inheritance of the same segment of DNA from the same ancestor on both homologues (i.e., IBD at one locus) results from inbreeding (the mating between related individuals). Such IBD segments appear as long stretches of homozygous markers called runs- of- homozygosity (ROH). Therefore, a practical method used for measuring IBD within individuals is the identification of ROH. It was first mentioned by Broman and Weber (1999) who hypothesised that long stretches of homozygosity represented previous inbreeding and could be associated with heritable diseases in humans. This was later validated by Lencz et al. (2007) who first coined the term "ROH" and defined it as a long track of homozygosity.

Gibson et al. (2006) studied the distribution of ROH in multiple outbred human populations. They observed ROH in the same regions of the genome regardless of population, thus showing the potential use for tracking ancestral haplotypes. McQuillan et al. (2008) examined the potential of estimating individual autozygosity using ROH. For this, they defined the inbreeding coefficient $F_{ROH}$. This coefficient represents the proportion of the autosomal genome that is found within a ROH. This new measure described inbreeding in terms of haplotype homozygosity and, assuming long homozygous haplotypes do not occur by chance, it measures IBD.

If a ROH is IBD, we can infer that every non- genotyped locus within this region is also homozygous (except in case of a mutation), that is, the entire region is inherited from a common ancestor. Conversely, if the ROH is IBS, it only reflects homozygosity at the genotyped SNPs, without providing information about non-genotyped loci within the segment. This distinction enhances our ability to identify markers associated with disease (Kenny et al. 2009), as understanding inheritance patterns can reveal demographic structures underlying complex genetic disorders (Bosse et al. 2019; Luan et al. 2012; Sticca et al. 2021).

Early applications in livestock populations used $F_{ROH}$ to measure inbreeding in cattle (Ferencakovic et al. 2011; Hamzic 2012; Sölkner et al. 2010). Genomic prediction with the Genomic Relationship Matrix (GRM) estimated from ROH has also shown to be more accurate than prediction using GRM based on linkage analysis or IBS (Identical by State) information in both real and simulated data (Luan et al. 2014). With all these studies using ROH, the question arises: how accurate are ROH in assessing IBD?

Despite the widespread use of ROH, its accuracy has not been rigorously evaluated, particularly in livestock species. Indeed, several studies assessed the accuracy of ROH in human data (Howrigan et al. 2011; Lavanchy and Goudet 2022; Narasimhan et al. 2016), but the population structure, including past effective population size, Linkage Disequilibrium (LD) structure, levels of inbreeding, and relatedness differs significantly from livestock populations. In livestock species, the few evaluations conducted were mostly empirical, thus not knowing the true IBD status (Ferenčaković et al. 2013). Similarly, Meyermans et al. (2020) proposed rules to optimise ROH parameters, but without knowing the true IBD. In addition, for many studies, the focus is on the accuracy of the inbreeding coefficient or the estimation of inbreeding depression, such as the simulation study by Lavanchy and Goudet (2022). Therefore, a direct evaluation of the accuracy of ROH in assessing IBD segments in livestock is lacking.

The aim of this paper is to estimate to what extent ROH are truly IBD, and what proportion of true IBD segments are missed when detecting ROH. For this, a computer simulation that mimics livestock populations is performed, where true IBD relative to a founder population is known by simulating additional loci that carry unique founder alleles. Different parameters, marker densities, and distance from the base

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

population are used to evaluate the rate of true positives and false negatives. Since the definition of IBD is highly dependent on the choice of the base population, a new marker- based measure of inbreeding within ROH is developed using additional loci that are not involved in the estimation of the ROH.

# 2 Material and Method

## 2.1 Simulation

The data was simulated using ADAM (Version 272) (Pedersen et al. 2009), a forward in time stochastic simulation program (Figure 1). A single 100 cM chromosome was simulated for 1000 generations with an effective population size of 200 and a mutation probability of $1 \times 10^{-6}$ to generate genomic variation and linkage disequilibrium. The recombination rate was 1 per cM and that followed a Poisson distribution. From this historical population, 100 males and 100 females (or 50 males and 50 females) were selected for the base population. In each of the next 100 generations, 100 males and 100 females (or 50 males and 50 females) were selected using random selection, resulting in 4 offspring from each mating. This resulted in an average of 2 (with a variation of 0.98) offspring that contributed to the next generation. We ran 10 replicates of each effective population size scenario.

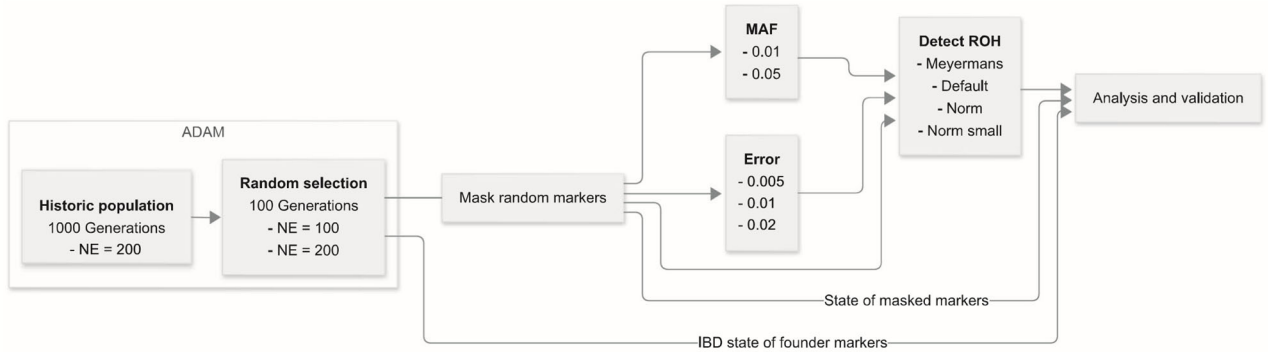


**Figure 1** Flowchart for the simulation and scenario generation. The genotypes are simulated using ADAM. A 100 cM chromosome was generated by simulating a 1000- generation deep historic population. From here, 100 males and 100 females (or 50 males and 50 females) were randomly selected and bred for 100 generations. The genotypes were then split into high- and low- SNP density datasets before a set of markers (MAF > 0.05) were masked from the dataset for validation.

The genome had 10,200 Single Nucleotide Polymorphism (SNP) markers with positional data, selected randomly from the segregating loci when sampling the base population. The SNP's positional data was given in cM and converted to base pair position using the assumption of 1 million base pairs per cM. These SNPs were split into a full set, with all the markers, and a 50% random subset to mimic a lower marker density. 30% of the SNPs with a MAF (minor allele frequency) over 5% in the final generation were masked (hidden for the ROH detection but available for other inferences) from the dataset used for detecting ROH. These loci were masked to examine the effect of a ROH on non- genotyped loci, such as the probability of inbreeding within a ROH (F|ROH). MAF filtering was done using PLINK 2.0. After masking these markers, the low- density scenario had an average of 46.09 SNP/Mb and the high density had 92.65 SNP/Mb.

ADAM also recorded a second set of 10,001 loci that carried unique founder alleles numbered from 1 to 200 for Ne = 100 and 1 to 400 for the Ne = 200 simulation. These markers were evenly spaced along the

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

genome. These unique founder alleles determine unequivocally whether two sampled alleles are IBD (i.e., copies of the same founder allele) or non- IBD (i.e., derived from two different founder alleles).

## 2.2 ROH

To detect the ROH we used the–homozyg function in PLINK 1.9 (Purcell et al. 2007), combined with four different combinations of parameters (Table 1). The four different settings were PLINKs default (Default), parameters commonly seen in livestock studies (Norm) (Álvarez et al. 2020; Biscarini et al. 2020; Meyermans et al. 2020; Monau et al. 2022; Pacheco et al. 2023), more lenient parameters with a lower minimum number of SNPs in a run and window (Norm small), and finally parameters based on the recommendations by Meyermans et al. (2020). The Meyerman parameters "scanning window size"(*L*) and "window threshold" (*t*) were calculated for each replication and marker density using a simplified version of the script shared by Gorssen et al. (2021). Since simulated data do not contain sequencing errors, nor missing genotypes, no quality control was performed to remove individuals or markers in any scenario. In accordance with the recommendations by Meyermans et al. (2020), no MAF filtering was performed in the analysis not exploring the effect of MAF.

## 2.3 Recording True IBD Segments

True IBD segments were detected by searching for continuous strings of founder markers homozygote for the founder alleles. Since the founder markers were evenly distributed, the marker distance was not considered. Results were categorised into seven bins based on segment and ROH length in Mb. The seven bins were [0,1], [1,2], [2,3], [3,4], [4,5], [5,9], and [9,100].

**Table 1** Parameters used to detect ROH using the–homozyg function in PLINK 1.9. The range of the *L* and *t* parameter is presented in parentheses.

| | Homozyg-window | | | Homozyg | | | | |
|---|---|---|---|---|---|---|---|---|
| **Settings** | **snp** | **Threshold** | **Missing** | **gap** | **snp** | **- kb** | **- Density** | **- Hetero** |
| Default | 50 | 0.05 | 1 | 1000 | 100 | 1000 | 50 | 0 |
| Norm | 50 | 0.05 | 1 | 1000 | 50 | 500 | 100 | 0 |
| Norm small | 20 | 0.05 | 1 | 1000 | 20 | 500 | 100 | 0 |
| Meyermans | *L* (188–228) | *t* (0.015–0.017) | 1 | 500 | *L* (188–228) | 1000 | 60 | 0 |

Abbreviations: L = scanning window size, *t* = window threshold.

## 2.4 True Positive Rate and the Power of ROH

The rate of true positive and power was evaluated by comparing ROH against true IBD segments. To mitigate the issue with ROH and IBD segments having different marker maps, the analysis asked two questions: (1) are the markers in a ROH also within an IBD segment, and (2) are the founder markers in an IBD segment also within a ROH?

The first step in this assessment involved converting the true IBD segments from the founder map to the SNP map. It is assumed that all areas outside the start and stop position of the true IBD segment are unknown in state. An IBD segment without any SNPs between the start and stop position of the segment

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

was excluded. Now with the same positional data, the true positive rate (IBD|ROH) and the power (ROH|IBD) was calculated. True positive was calculated as the fraction of IBD markers in a ROH, and power was the fraction of markers in a IBD segment detected by PLINK.

## 2.5 Measure of IBD Increase in a ROH: F|ROH

We define the measure F|ROH for the increase in inbreeding within a ROH inspired by Wright (1950) *F*-statistics, in particular $F_{IT}$, which compares the observed to the expected heterozygosity in a (sub)population. We use the $F_{IS}$ statistic here to measure the inbreeding within a ROH relative to the population mean inbreeding, which we will call F|ROH, (Wright 1950):

$$\mathrm{F|ROH}_i = 1 - \frac{O(\mathrm{Het})_i}{E(\mathrm{Het})_i}$$

where F|ROH is masked marker *i*'s individual inbreeding within a ROH relative to the average inbreeding within the population, the expected heterozygosity of the individuals is $(\mathrm{Het})_i = 2q_{ti}(1 - q_{ti})$, where $q_{ti}$ is the allele frequency in generation *t* at masked marker *i*, and the observed heterozygosity ($O(\mathrm{Het})_i$) is the heterozygosity within a ROH for masked marker *i*. Markers used to detect the ROH are by definition not heterozygous, since they are in a ROH. Hence, we used masked markers to estimate *O*(Het), that is, markers that were not used to detect the ROH. Because F|ROH estimation is based on heterozygosity, MAF > 0.05 filtering is used for the masked loci to ensure substantial frequencies of heterozygosity. Thus, F|ROH measures the increase in IBD (loss of heterozygosity) for masked markers in a ROH compared to the general expectations of these masked markers' heterozygosity in the genome. The masked markers may be seen as ungenotyped causal variants, whose IBD status we want to capture (e.g., for genomic predictions and IBD mapping).

## 2.6 Correlation, Regression Coefficients and Rate of Inbreeding

The individual $F_{IBD}$ was calculated as the total fraction of founder markers that were IBD. Similarly, $F_{ROH}$ was calculated as the fraction of the genome in a ROH. We computed the correlation between these two inbreeding measures as well as the regression of $F_{IBD}$ on $F_{ROH}$.

Since $F_t = 1 - (1-\Delta F)^t$, a constant $\Delta F$ will cause $\log(1 - F_t)$ to change linearly over time (Hinrichs et al. 2007; Meuwissen et al. 2020). $\Delta F$ was estimated as the slope of a regression of $-\log(1 - F_t)$ on generation t.

## 2.7 Effect of MAF and Genotyping Errors

To assess the impact of MAF filtering the data was filtered at 1% and 5% MAF. The resulting ROH were evaluated by examining the regression coefficient of $F_{IBD}$ on $F_{ROH}$ across different MAF levels. The effect of MAF was also assessed by calculating F|ROH for the different scenarios.

Genotyping errors were introduced by randomly altering heterozygotes to homozygotes and vice versa at a rate of 0.5%, 1%, and 2%. The effect of these genotyping errors was then assessed by regressing $F_{IBD}$ on $F_{ROH}$ calculated at each error level.

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

# 3 Results

## 3.1 Distribution of IBD Segments and ROH

The average total KB of IBD segments per individual increased from 10,400 KB of IBD in generation 10–21,255 KB in generation 100. When going from Ne = 200 to Ne = 100 there was a 72% increase in total KB per individual. The average length of the IBD segments were 8075 and 1343 KB in generation 10 and 100, respectively.

Figure 2 shows that the stringency of parameters used for ROH detection causes substantial changes in ROH length distribution. More stringent ROH detection results in fewer detected ROH, and a shift towards the detection of longer runs. Hence, a substantial proportion of total ROH length goes undetected if the minimum SNP threshold is set too high. We see that over time there is minimal change in ROH distribution. This is to be expected as the populations simulated are at equilibrium. The heights of the histograms also show that having a lower marker density causes a smaller total length of ROH to be detected with Norm, and default settings and Meyermans, while for Norm Small the opposite happens.

## 3.2 True Positive Rate in ROH

For each ROH, the true positive rate was calculated and averaged over each variable combination. The true positive rate was observed to be affected by the ROH detection settings, marker density, effective population size, distance to base population, and the length of the ROH. Looking at the overall picture, none of the methods reach an overall true positive rate over 0.9 (Table 2). For the within- bin results, it is only ROH over 9Mb that consistently achieves a rate over 0.9 (Table S1). This is, however, only true for generation 50 and 100 and all settings besides default. For the default setting, true positive rates over 0.9 are only achieved for generation 100 in the high- density scenario. It is notable that ROH 4–5 Mb or smaller will mostly be non- IBD in generation 10 and 20, even with a high minimum threshold for the number of SNPs. Reduced effective population size increases the true positive rate, with the values approximately being doubled in the Ne = 100 population compared to the Ne = 200 populations.

Since the stricter methods find fewer short ROH, and long ROH have on average higher true positive rates, the Meyermans setting achieves the highest overall true positive rate. For the shorter ROH, Meyermans achieves slightly higher true positive rates compared to Norm. But within each bin above 3–4 Mb there are no strong differences between the methods.

## 3.3 Power

The power was calculated by finding the fraction of markers in an IBD segment that were detected in a ROH. The results were then averaged over the different combinations of variables. Table 3 shows that generations more distant to the base population are associated with less power. This may be a result of having a higher proportion of shorter IBD segments (compared to earlier generations) that are usually more difficult to detect. That a higher proportion of short IBD segments is affecting the overall power is further supported when we calculated the results per length bin (Table S2). Here we see that across generations a bin performs similarly. The per bin power results do however show a clearer difference between the

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

settings for the smaller segments (< 2Mb), with the most stringent method missing more than 95% of the markers in the 0–1 Mb bin. Having a higher marker density has a positive effect on the power with all but Norm small having higher power. The power is thus mainly controlled by the lower minimum threshold for ROH set by the user.

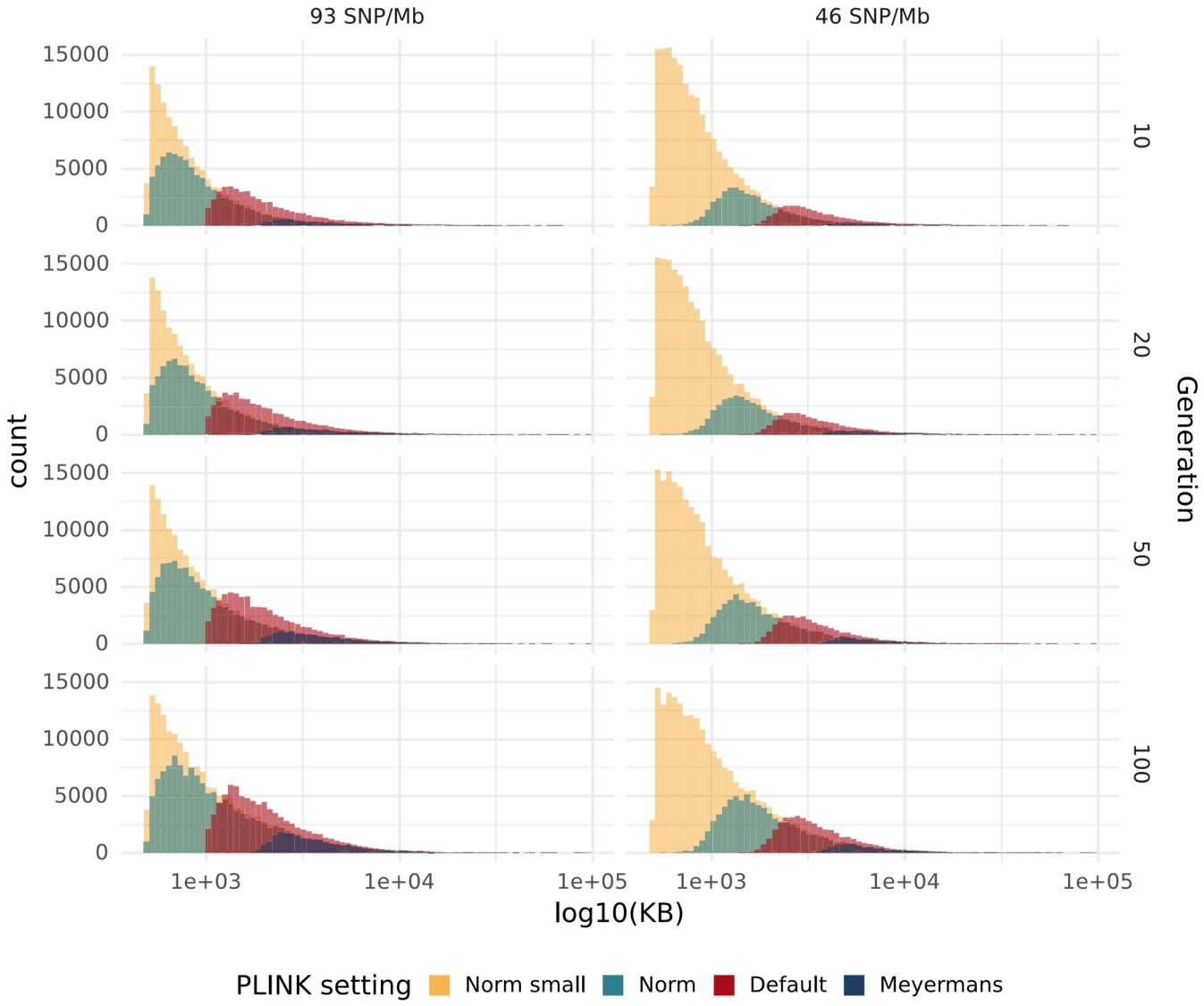


**Figure 2** Histogram of ROH length (KB) split by marker density and number of generations after the base population. The colours of the histograms represent the parameter settings used to detect ROH in the populations. [Colour figure can be viewed at wileyonlinelibrary.com]

## 3.4 F in ROH (F|ROH)

We calculated F|ROH for each masked marker in the datasets; the results were then aggregated based on relevant variable combinations. Table 4 shows that more stringent methods such as Meyermans resulted in overall higher inbreeding levels in the ROH. This effect of stringency is mainly explained by the lower portion of smaller ROH detected (Table S3). In the high- density scenario, most ROH in or over the 2–3 Mb bin have a F|ROH over 0.9, which indicates that these ROH generally reflect true inbreeding. For the low-density scenario, this increases to 3–4 Mb. When comparing F|ROH for the same bin for different detection

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

settings, we see a minimal difference between settings. Being further away from the base population reduces the F|ROH. This may be a result of the general population becoming more inbred and thus increasing the expected homozygosity, making it more unlikely to achieve high F|ROH estimates.

## 3.5 Regression and Correlation Between $F_{IBD}$ and $F_{ROH}$

Table 5 shows the regression coefficient and correlation of $F_{ROH}$ on $F_{IBD}$ calculated from the fraction of kb found within a ROH or IBD segments, respectively. For all settings beside Norm small and Meyermans higher density caused both the correlation and regression coefficients to approach one. This was true both for overall estimates and when inbreeding was calculated with only segments and ROH in a specific bin (Table S4). Having a lower effective population size increases the correlation and moves the regression coefficient towards one for all settings beside Meyermans. Meyermans does however reach a correlation above 0.9 in most scenarios beside generation 10. The other settings only achieve a correlation over 0.9 in generation 80 in the population with low effective population size.

**Table 2** The mean true positive rate (IBD given ROH) per ROH. The results are split by parameter settings used to detect ROH, effective population size (NE), marker density and number of generations after the base population. All standard error < 0.015.

| Setting | NE | Density | Generation 10 | Generation 20 | Generation 50 | Generation 100 |
|---|---|---|---|---|---|---|
| Norm small | NE100 | 93 SNP/Mb | 0.018 | 0.058 | 0.209 | 0.414 |
| | | 46 SNP/Mb | 0.011 | 0.035 | 0.137 | 0.305 |
| | NE200 | 93 SNP/Mb | 0.009 | 0.030 | 0.111 | 0.250 |
| | | 46 SNP/Mb | 0.006 | 0.019 | 0.070 | 0.170 |
| Norm | NE100 | 93 SNP/Mb | 0.025 | 0.081 | 0.265 | 0.492 |
| | | 46 SNP/Mb | 0.044 | 0.127 | 0.346 | 0.559 |
| | NE200 | 93 SNP/Mb | 0.013 | 0.044 | 0.149 | 0.312 |
| | | 46 SNP/Mb | 0.023 | 0.073 | 0.212 | 0.391 |
| Default | NE100 | 93 SNP/Mb | 0.041 | 0.114 | 0.317 | 0.521 |
| | | 46 SNP/Mb | 0.076 | 0.189 | 0.409 | 0.572 |
| | NE200 | 93 SNP/Mb | 0.021 | 0.064 | 0.186 | 0.347 |
| | | 46 SNP/Mb | 0.038 | 0.108 | 0.262 | 0.419 |
| Meyermans | NE100 | 93 SNP/Mb | 0.258 | 0.498 | 0.733 | 0.843 |
| | | 46 SNP/Mb | 0.409 | 0.659 | 0.828 | 0.867 |
| | NE200 | 93 SNP/Mb | 0.132 | 0.335 | 0.632 | 0.787 |
| | | 46 SNP/Mb | 0.248 | 0.488 | 0.766 | 0.850 |

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

**Table 3** The mean power rate (ROH given IBD) per IBD segment. The results are split by parameter settings used to detect ROH, effective population size (NE), marker density and number of generations after base population. All standard errors < 0.018.

| Setting | NE | Dens | Generation 10 | Generation 20 | Generation 50 | Generation 100 |
|---|---|---|---|---|---|---|
| Norm small | NE100 | 93 SNP/Mb | 0.951 | 0.936 | 0.881 | 0.842 |
| | | 46 SNP/Mb | 0.958 | 0.948 | 0.915 | 0.887 |
| | NE200 | 93 SNP/Mb | 0.97 | 0.936 | 0.869 | 0.806 |
| | | 46 SNP/Mb | 0.974 | 0.952 | 0.899 | 0.856 |
| Norm | NE100 | 93 SNP/Mb | 0.939 | 0.917 | 0.839 | 0.79 |
| | | 46 SNP/Mb | 0.9 | 0.845 | 0.716 | 0.648 |
| | NE200 | 93 SNP/Mb | 0.949 | 0.912 | 0.823 | 0.734 |
| | | 46 SNP/Mb | 0.888 | 0.83 | 0.677 | 0.561 |
| Default | NE100 | 93 SNP/Mb | 0.903 | 0.869 | 0.761 | 0.71 |
| | | 46 SNP/Mb | 0.851 | 0.766 | 0.594 | 0.548 |
| | NE200 | 93 SNP/Mb | 0.911 | 0.861 | 0.728 | 0.622 |
| | | 46 SNP/Mb | 0.815 | 0.736 | 0.535 | 0.427 |
| Meyermans | NE100 | 93 SNP/Mb | 0.787 | 0.653 | 0.404 | 0.305 |
| | | 46 SNP/Mb | 0.619 | 0.445 | 0.204 | 0.139 |
| | NE200 | 93 SNP/Mb | 0.751 | 0.617 | 0.367 | 0.218 |
| | | 46 SNP/Mb | 0.609 | 0.394 | 0.17 | 0.084 |

**Table 4** Mean increased level of inbreeding within a ROH (F|ROH). The results are split by parameter settings used to detect ROH, effective population size (NE), marker density and number of generations after base population. All standard errors < 0.036.

| Setting | NE | Density | Generation 10 | Generation 20 | Generation 50 | Generation 100 |
|---|---|---|---|---|---|---|
| Norm small | NE100 | 93 SNP/Mb | 0.818 | 0.821 | 0.802 | 0.787 |
| | | 46 SNP/Mb | 0.613 | 0.613 | 0.605 | 0.592 |
| | NE200 | 93 SNP/Mb | 0.809 | 0.805 | 0.784 | 0.765 |
| | | 46 SNP/Mb | 0.606 | 0.600 | 0.577 | 0.553 |
| Norm | NE100 | 93 SNP/Mb | 0.877 | 0.877 | 0.860 | 0.848 |
| | | 46 SNP/Mb | 0.854 | 0.843 | 0.828 | 0.799 |
| | NE200 | 93 SNP/Mb | 0.871 | 0.867 | 0.847 | 0.831 |
| | | 46 SNP/Mb | 0.849 | 0.840 | 0.813 | 0.787 |
| Default | NE100 | 93 SNP/Mb | 0.852 | 0.839 | 0.802 | 0.766 |
| | | 46 SNP/Mb | 0.833 | 0.807 | 0.770 | 0.704 |
| | NE200 | 93 SNP/Mb | 0.850 | 0.837 | 0.799 | 0.762 |
| | | 46 SNP/Mb | 0.822 | 0.802 | 0.754 | 0.707 |
| Meyermans | NE100 | 93 SNP/Mb | 0.978 | 0.977 | 0.962 | 0.946 |
| | | 46 SNP/Mb | 0.978 | 0.968 | 0.955 | 0.930 |
| | NE200 | 93 SNP/Mb | 0.977 | 0.976 | 0.964 | 0.956 |
| | | 46 SNP/Mb | 0.975 | 0.968 | 0.956 | 0.945 |

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

**TABLE 5** Correlation, regression coefficient ($\beta$) and deviation between true inbreeding ($F_{IBD}$) and estimated inbreeding ($F_{ROH}$). The results are split by parameter settings used to detect ROH, effective population size (NE), marker density and number of generations after base population.

| | | | Gen. 10 | | | Gen. 50 | | | Gen. 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Method** | **NE** | **Density** | $\boldsymbol{\beta}$ | **Corr.** | $\overline{\boldsymbol{F_{ROH} - F_{IBD}}}$ | $\boldsymbol{\beta}$ | **Corr.** | $\overline{\boldsymbol{F_{ROH} - F_{IBD}}}$ | $\boldsymbol{\beta}$ | **Corr.** | $\overline{\boldsymbol{F_{ROH} - F_{IBD}}}$ |
| Norm small | NE100 | 93 SNP/Mb | 0.700 | 0.744 | 0.190 | 0.960 | 0.891 | 0.180 | 1.020 | 0.897 | 0.170 |
| | | 46 SNP/Mb | 0.710 | 0.679 | 0.320 | 1.000 | 0.823 | 0.310 | 1.070 | 0.827 | 0.280 |
| | NE200 | 93 SNP/Mb | 0.520 | 0.646 | 0.200 | 0.840 | 0.826 | 0.200 | 0.930 | 0.862 | 0.190 |
| | | 46 SNP/Mb | 0.500 | 0.580 | 0.330 | 0.830 | 0.747 | 0.330 | 0.950 | 0.774 | 0.320 |
| Norm | NE100 | 93 SNP/Mb | 0.710 | 0.769 | 0.150 | 0.940 | 0.905 | 0.140 | 0.990 | 0.917 | 0.120 |
| | | 46 SNP/Mb | 0.670 | 0.751 | 0.140 | 0.880 | 0.894 | 0.120 | 0.850 | 0.877 | 0.100 |
| | NE200 | 93 SNP/Mb | 0.540 | 0.674 | 0.150 | 0.850 | 0.855 | 0.150 | 0.930 | 0.892 | 0.140 |
| | | 46 SNP/Mb | 0.510 | 0.671 | 0.140 | 0.770 | 0.837 | 0.120 | 0.830 | 0.870 | 0.100 |
| Default | NE100 | 93 SNP/Mb | 0.600 | 0.694 | 0.170 | 0.860 | 0.866 | 0.160 | 0.880 | 0.866 | 0.140 |
| | | 46 SNP/Mb | 0.570 | 0.698 | 0.130 | 0.760 | 0.860 | 0.110 | 0.710 | 0.834 | 0.090 |
| | NE200 | 93 SNP/Mb | 0.440 | 0.607 | 0.180 | 0.720 | 0.789 | 0.160 | 0.780 | 0.827 | 0.150 |
| | | 46 SNP/Mb | 0.410 | 0.614 | 0.140 | 0.650 | 0.790 | 0.110 | 0.680 | 0.823 | 0.080 |
| Meyermans | NE100 | 93 SNP/Mb | 0.790 | 0.873 | 0.030 | 0.890 | 0.943 | 0.000 | 0.810 | 0.919 | −0.070 |
| | | 46 SNP/Mb | 0.800 | 0.888 | 0.020 | 0.990 | 0.920 | −0.040 | 0.910 | 0.818 | −0.140 |
| | NE200 | 93 SNP/Mb | 0.670 | 0.811 | 0.040 | 0.910 | 0.942 | 0.010 | 0.910 | 0.929 | −0.040 |
| | | 46 SNP/Mb | 0.680 | 0.828 | 0.020 | 1.050 | 0.950 | −0.010 | 1.230 | 0.900 | −0.050 |

*Note:* Corr. = Correlations, $F_{ROH} - F_{IBD}$ = The group mean of difference between estimated and true inbreeding.

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

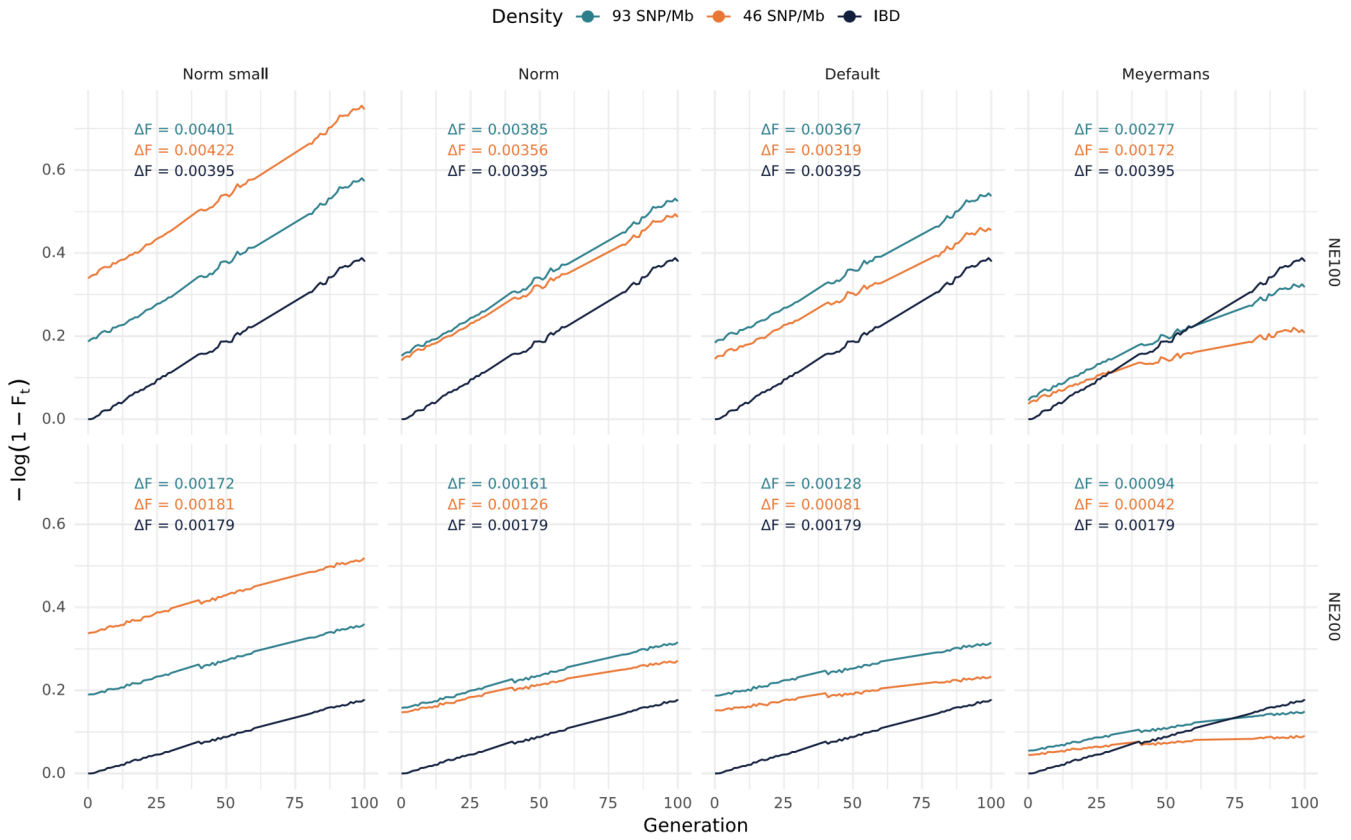


**Figure 3** The regression of −log(1- $F_t$) on generation. The lines are coloured by marker density. The black line is the regression of true inbreeding for NE = 100 and NE = 200, resulting in true Δ*F* of 0.004 and 0.0018, respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

The differences between settings become less apparent when we compare $F_{ROH}$ and $F_{IBD}$ calculated from ROH and IBD segments in a specific bin. For all generations the 9 + Mb bin achieves the highest regression coefficients, and 5–9 Mb only achieves coefficient values around 0.7 after 50 generations. For all bins and detections settings a lower SNP density negatively affects the quality of the $F_{ROH}$ estimates.

## 3.6 Rate of Inbreeding

Rates of inbreeding were estimated for both $F_{IBD}$ and $F_{ROH}$ as the regression of −log(1- $F_t$) on generation *t*. The regression coefficient for the rate of true inbreeding was 0.004 for Ne = 100, and 0.0018 for Ne = 200. These rates of inbreeding estimates are 25% and 39% smaller than expected for populations of these sizes. This decreased rate of inbreeding is possibly a result of the reduced variation in family size that tends to increase the estimated effective population size (Santiago and Caballero 1995). Figure 3 shows that stringent ROH parameters cause reduced estimates of inbreeding rates, with Meyermans having the lowest estimates for all scenarios. Marker density has a positive effect, with the regression coefficients for the high- density scenarios achieving values closer to the true inbreeding rate.

## 3.7 Effect of Genotyping Error and MAF Filtering

When regressing $F_{IBD}$ on $F_{ROH}$, increasing rates of genotyping errors cause the regression coefficient to diverge from one, regardless of the ROH detection parameters (Table 6). The more stringent parameters,

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

such as Meyermans, are more affected by errors than the more lenient parameter combinations. For most ROH parameters, except Default, higher error rates are more detrimental in the high- density scenario compared to the low- density scenario. MAF filtering does; however, not appear to be a stable mitigation strategy, as it has varying results across detection parameters, error rates, and generation (Table S5).

In the scenarios without genotyping errors, MAF filtering has varying effects on the regression coefficient (Table 6, Table S5). In generation 100, MAF filtering can be favourable for some parameters, with 5% MAF filtering for Norm resulting in the regression coefficient that is closest to 1. For generation 20, the best regression coefficients were achieved by Meyermans with 1% MAF and a coefficient of 0.83 in the high-density scenario and 5% MAF and 0.85 in the low- density scenario. For generation 50, 1% MAF filtering reaches a coefficient of 0.95 with the Norm small detection setting, and in low density, Norm small got a regression coefficient of 0.97 with 5% MAF filtering. However, in the low- density scenario after 5% MAF filtering, the Default and Meyermans parameters struggle to detect ROH.

**Table 6** Average regression coefficient of true inbreeding on estimated inbreeding for different levels of genotyping errors and MAF filtering. Values are from generation 100, and calculated for high and low marker density.

| | | | Gen. 100 | | | |
|---|---|---|---|---|---|---|
| **Density** | **Error** | **MAF** | **Default** | **Meyermans** | **Norm** | **Norm small** |
| 93 SNP/Mb | 0 | 0.00 | 0.91 | 0.97 | 1.02 | 1.06 |
| | | 0.01 | 0.86 | 1.03 | 1.02 | 1.09 |
| | | 0.05 | 0.71 | 1.02 | 1 | 1.07 |
| | 0.005 | 0.00 | 0.91 | 1.24 | 1.08 | 1.09 |
| | | 0.01 | 0.9 | 1.09 | 1.09 | 1.1 |
| | | 0.05 | 0.73 | 1.08 | 1.07 | 1.09 |
| | 0.01 | 0.00 | 0.93 | 1.6 | 1.17 | 1.16 |
| | | 0.01 | 0.98 | 1.23 | 1.21 | 1.09 |
| | | 0.05 | 0.78 | 1.17 | 1.15 | 1.11 |
| | 0.02 | 0.00 | 1.02 | 2.25 | 1.46 | 1.36 |
| | | 0.01 | 1.1 | 1.94 | 1.56 | 1.32 |
| | | 0.05 | 0.9 | 1.36 | 1.37 | 1.17 |
| 46 SNP/Mb | 0 | 0.00 | 0.79 | 0.92 | 0.91 | 1.15 |
| | | 0.01 | −0.18 | 0.13 | 0.88 | 0.8 |
| | | 0.05 | NA | −0.43 | 0.77 | 0.56 |
| | 0.005 | 0.00 | 0.8 | 1.12 | 0.97 | 1.15 |
| | | 0.01 | −0.06 | 0.51 | 0.97 | 0.91 |
| | | 0.05 | −1.81 | −0.18 | 0.86 | 0.64 |
| | 0.01 | 0.00 | 0.83 | 1.33 | 1.05 | 1.18 |
| | | 0.01 | 0.69 | 1.22 | 1.15 | 1.01 |
| | | 0.05 | −2.42 | 0.12 | 0.92 | 0.67 |
| | 0.02 | 0.00 | 0.93 | 1.78 | 1.27 | 1.24 |
| | | 0.01 | 1.01 | 1.62 | 1.37 | 1.19 |
| | | 0.05 | −0.28 | 0.77 | 1.14 | 0.8 |

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

For each generation, different levels of MAF filtering moved the distribution of ROH closer to the distribution of IBD segments (Figure S1). For generations 10–50, 5% MAF filtering resulted in the most similar distribution, while for generation 100, avoiding MAF filtering maintains the ability to detect the shorter IBD segments.

MAF filtering was also evaluated using F|ROH (Figure 4). Here MAF filtering causes an overall higher F|ROH, and less variation in the results for all detection parameters and generations. A notable effect of MAF filtering is that since ROH no longer can appear in low frequency areas, F|ROH increases over time, while without MAF filtering, F|ROH decreases over time.

# 4 Discussion

In the present study, we explored to what degree ROH represent IBD segments in a simulated livestock population, as validation studies in such populations remain rare despite the widespread use of these approaches. We simulated populations with different effective population sizes and marker densities and further examined how different parameters in rule- based ROH detection methods affect the quality of the results. While model- based approaches have been developed for the detection of IBD segments, where the genome is described as a mosaic of IBD and non- IBD segments (Druet and Gautier 2017; Leutenegger et al. 2003; Narasimhan et al. 2016; Vieira et al. 2016), the rule- based methods remain by far the most widely used approach. Model- and rule- based approaches have already been compared using simulated data (Druet and Gautier 2017; Lavanchy and Goudet 2022) or in empirical studies (Alemu et al. 2021; Duntsch et al. 2021; Solé et al. 2017). The advantage of model- based approaches is higher at low marker densities (Druet et al. 2020) or at variable marker spacing (Forneris et al. 2025), whereas at higher marker densities (e.g., 50K array in cattle), the genome- wide estimates obtained with the model- based approach and the rule- based ROH are highly correlated (Alemu et al. 2021; Solé et al. 2017). Given their popularity and the simulated marker densities, we have focused on a rule- based approach here. This also gives us a possibility to better understand how specific physical aspects of ROH, such as marker density and length, affect the ease of correct detection.

Our results indicate that parameter settings for rule- based detection approaches need to be adjusted to suit the properties of the dataset and the application using ROH. This dependency is exemplified by the contrasted results obtained when focusing only on longer ROH, which proved efficient for estimating the inbreeding coefficient $F_{ROH}$ but less so for estimating inbreeding rates. Indeed, when we separate the results based on the length of ROH, the difference between methods is reduced and shows that most of the overall estimates are controlled by the distribution of ROH length for the different detection parameters. However, it is notable that either focusing on long ROH or maximal coverage does not improve the quality of estimates across all applications. We find that the true positive rate of long ROH (9+ Mb) is consistently higher, and practical applications such as the estimation of $F_{ROH}$ produce estimates closer to the simulated truth when longer ROH represent a larger portion of the detected ROH. This is comparable with what has been seen in model- based approaches (Lavanchy and Goudet 2022; Tang et al. 2022; Zhou et al. 2020). However, the drawback of focusing on longer ROH is the loss of overall detection power, especially of shorter ROH. This loss of detection power becomes more detrimental further into the simulation, where the distribution of ROH leans more towards shorter ROH. This results in stricter methods underestimating

*Published in : Journal of Animal Breeding and Genetics (2025), pp. 1-13*
*DOI: 10.1111/jbg.70013*
*Status : Postprint (Author's version)*

the rate of inbreeding in the population. This reduced estimate may indicate that the shorter ROH are needed to correctly capture the accumulative nature of inbreeding.

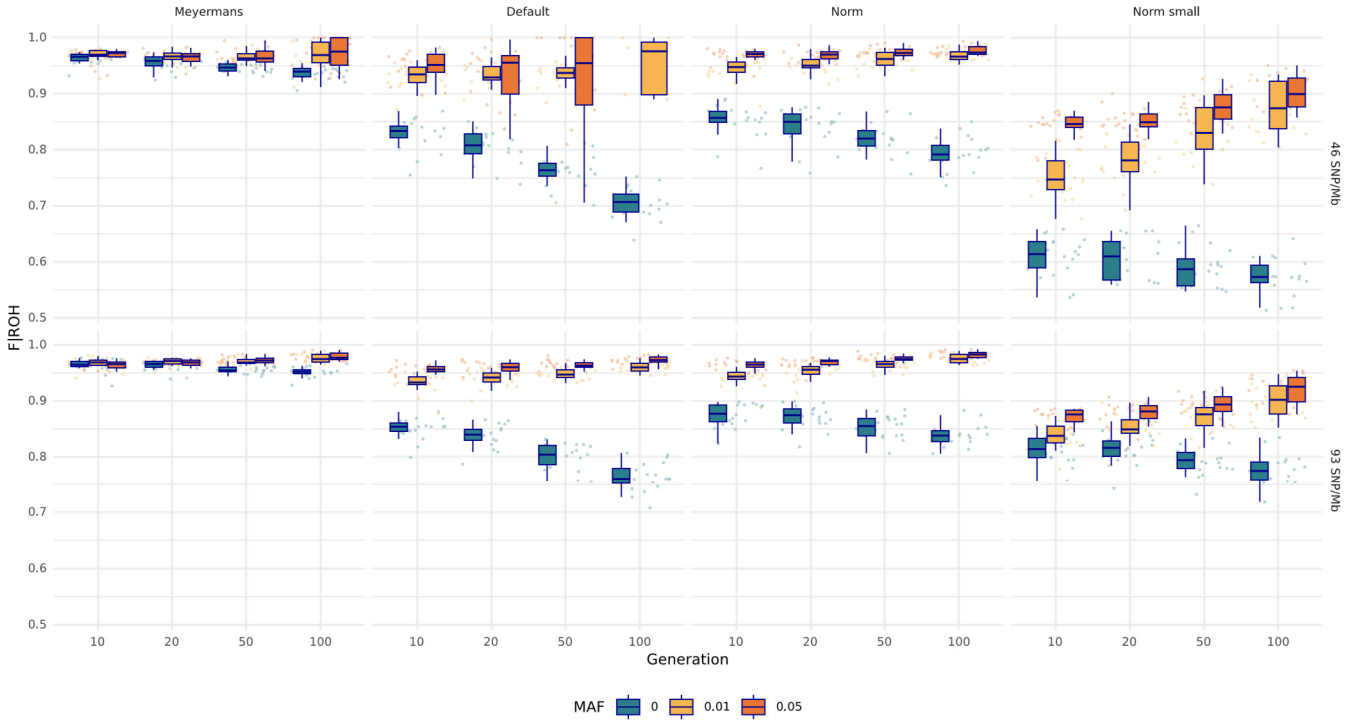


**Figure 4** The effect of MAF filtering on F|ROH estimates. The lower and upper hinges correspond to the first and third quartiles and the whisker extends from the hinge to the most extreme value no further than 1.5 × inter- quartile range from the hinge. [Colour figure can be viewed at wileyonlinelibrary.com]

For all analysis, the worse results were obtained with the lower marker density scenario (46 SNP/Mb). This is in line with the findings of previous studies, which argue that higher marker density allows for both more accurate and deeper historical information (Ceballos et al. 2018). A departure from the previous observations is that when a dataset includes genotyping errors, $F_{ROH}$ estimates become dependent on including shorter ROH to achieve similar values to $F_{IBD}$. This is due to genotyping errors breaking long ROH into shorter ones, a behaviour previously observed by Ceballos et al. (2018), who argued that due to the higher level of genotyping errors in whole genome sequencing data, more errors should be allowed in ROH detection to achieve a similar ROH structure to that found with lower density SNP data. Although inclusion of shorter ROH results in $F_{ROH}$ estimates more alike the $F_{IBD}$ values in our simulation, this might not be an advantageous strategy in all situations as our results show that these ROH less often represent true IBD. Accepting some heterozygous genotypes within ROH, which is already a common practice, or using a model- based approach may thus be more risk averse approaches to dealing with genotyping errors.

Previous studies by Hillestad et al. (2017), Ferenčaković et al. (2013) and Meyermans et al. (2020) that examined the effect of MAF filtering on ROH detection used coverage as their benchmark (coverage referring to the proportion of the genome where ROH can be detected based on marker density and parameter settings). These studies recommend not performing MAF filtering as they lost ROH coverage. Here, MAF filtering did not improve $F_{ROH}$ estimation but resulted in increased F|ROH values. Homozygosity at rare alleles is strong evidence for IBD, so the use of markers with low MAF may be beneficial. However, the vast majority of homozygous genotypes at these markers are homozygous for the common allele and

provide little evidence for IBD and should probably not be considered for calling ROH. Note that this dilemma is resolved when working with methods using allele frequencies.

In our simulations, the distribution of ROH and their length was relatively uniform across the genome. This would not be the case in the presence of selection or when recombination rates are heterogeneous. The impact of heterogeneous recombination rates on the identification of ROH has, for example, been highlighted in *Ficedula Flycatcher* by Kardos et al. (2017). Due to the presence of micro- and macro-c hromosomes, there is a much larger variability in recombination rates than in most mammals. Druet and Gautier (2017) showed that ignoring variation in recombination rate has minor consequences in simulations with moderate variation in recombination rates, and that this aspect can be taken into account when using genetic maps in model- based approaches. Higher recombination rates would reduce the length of the ROH, an effect similar to having more generations to the common ancestors, while the effect on the number of markers per ROH would be the same as reducing marker density. Therefore, even if we did not directly simulate such variations in recombination rate, we could infer the effects of higher or lower recombination rates on the relationships between ROH and IBD. Selection reduces the time to the most common ancestor and can increase the level of IBD, but only around favourable loci. The effect of selection can be seen as reduced Ne around these loci. Therefore, the effect of selection on the detection of ROH and their correlation with IBD can also be understood by comparing results in simulations with different Ne. IBD can result from both selection and drift, and ROH do not distinguish between these two processes, although it would be interesting to be able to disentangle these two forces.

Our study shows us there is no single parameter setting that works for all data sets (population structure, marker density, etc.) and is best for all applications. Therefore, parameters should ideally be validated for each data set. For example, Ferenčaković et al. (2013) and Silva et al. (2024) proposed measuring the robustness of ROH detection to different parameter settings. Forneris et al. (2025) suggested relying on simulations that mimic the target data. Interestingly, the F|ROH statistic we introduced in the present study could be used for this purpose when some SNPs (e.g., 1 in 10 SNPs) are masked for the ROH detection and subsequently used for the estimation of F|ROH. For example, for a selection of different parameter settings, the F|ROH measure would be calculated, and the parameter settings that reach a threshold of 0.9 could be deemed acceptable. In addition, F|ROH avoids the problem of having to define a base population, causing the results across generations to be more stable. An alternative would be to use model- based approaches that have the advantage of requiring less parameter definition and adjusting more automatically to the data. For example, in Forneris et al. (2025), parameters had to be adjusted when rule-based approaches were applied to different genotyping panels, whereas the model- based approach of ZooRoH used the same model across all panels. The model- based approaches are also useful for positions where the classification is ambiguous, as they provide HBD probabilities, unlike rule- based approaches that make a binary classification (ROH vs. non- ROH). Such positions with intermediate HBD probabilities would correspond to shorter segments (more ancient ancestors), but possibly also to segment boundaries.

# 5 Conclusions

In conclusion, for a 150 and 300 k SNP chip, ROH is a reliable indicator for IBD if the ROH are longer than 2 Mb. The accurate detection of long segments yields high correlations between estimated and true inbreeding. The high correlations imply that individual differences in inbreeding are accurately predicted

by $F_{ROH}$. Furthermore, the regression coefficients being close to one signify that a 1% variation in $F_{ROH}$ corresponds to a 1% variation in $F_{IBD}$. In contrast to $F_{ROH}$, when estimating the rate of inbreeding in a population, including shorter ROH is essential. When applying rule- based approaches using parameter settings that provide an overall F|ROH over 0.9, it may be a strategy for optimising ROH detection.

**Acknowledgements**

We are very grateful to two anonymous reviewers for their helpful comments. The authors acknowledge the Orion High Performance Computing Centre (OHPCC) at the Norwegian University of Life Sciences (NMBU) for providing computational resources that have contributed to the research results reported within this paper. Tom Druet is Research Director from the Fonds de la Recherche Scientifique—FNRS (F.R.S-FNRS).

**Conflicts of interest**

The authors declare no conflicts of interest.

**Data availability statement**

The data utilised in this study was generated through simulations and does not include any proprietary or confidential information. The scripts used for these simulations and data analysis are publicly available and can be accessed in the GitHub repository (https:// github. com/ odawa ge/ ROH-vs- IBD). For further inquiries or Supporting Information, please refer to the corresponding author.

## References

Alemu, S. W., N. K. Kadri, C. Harland, et al. 2021. "An Evaluation of Inbreeding Measures Using a Whole- Genome Sequenced Cattle Pedigree." *Heredity* 126: 410–423. https://doi.org/10.1038/s41437-02000383-9.

Álvarez, I., I. Fernández, A. Traoré, L. Pérez- Pardal, N. A. Menéndez- Arias, and F. Goyache. 2020. "Ancient Homozygosity Segments in West African Djallonké Sheep Inform on the Genomic Impact of Livestock Adaptation to the Environment." *Animals* 10: 1178. https://doi.org/10.3390/ani10071178.

Biscarini, F., S. Mastrangelo, G. Catillo, G. Senczuk, and R. Ciampolini. 2020. "Insights Into Genetic Diversity, Runs of Homozygosity and Heterozygosity- Rich Regions in Maremmana Semi- Feral Cattle Using Pedigree and Genomic Data." *Animals* 10: 2285. https://doi.org/10.3390/ani10122285.

Bosse, M., H.- J. Megens, M. F. L. Derks, Á. M. R. de Cara, and M. A. M. Groenen. 2019. "Deleterious Alleles in the Context of Domestication, Inbreeding, and Selection." *Evolutionary Applications* 12: 6–17. https://doi.org/10.1111/eva.12691.

Broman, K. W., and J. L. Weber. 1999. "Long Homozygous Chromosomal Segments in Reference Families From the Centre D'étude du Polymorphisme Humain." *American Journal of Human Genetics* 65: 1493–1500. https://doi.org/10.1086/302661.

Ceballos, F. C., S. Hazelhurst, and M. Ramsay. 2018. "Assessing Runs of Homozygosity: A Comparison of SNP Array and Whole Genome Sequence Low Coverage Data." *BMC Genomics* 19: 106. https://doi.org/10.1186/s12864-018-4489-0.

Druet, T., and M. Gautier. 2017. "A Model- Based Approach to Characterize Individual Inbreeding at Both Global and Local Genomic Scales." *Molecular Ecology* 26: 5820–5841. https://doi.org/10.1111/mec.14324.

Druet, T., K. Oleński, L. Flori, et al. 2020. "Genomic Footprints of Recovery in the European Bison." *Journal of Heredity* 111: 194–203. https://doi.org/10.1093/jhered/esaa002.

Duntsch, L., A. Whibley, P. Brekke, J. G. Ewen, and A. W. Santure. 2021. "Genomic Data of Different Resolutions Reveal Consistent Inbreeding Estimates but Contrasting Homozygosity Landscapes for the Threatened Aotearoa New Zealand Hihi." *Molecular Ecology* 30: 6006–6020. https://doi.org/10.1111/mec.16068.

Ferencakovic, M., E. Hamzic, B. Gredler, I. Curik, and J. Sölkner. 2011. "Runs of Homozygosity Reveal Genome- Wide Autozygosity in the Austrian Fleckvieh Cattle." *Agriculturae Conspectus Scientificus* 76: 325–329.

Ferenčaković, M., J. Sölkner, and I. Curik. 2013. "Estimating Autozygosity From High- Throughput Information: Effects of SNP Density and Genotyping Errors." *Genetics Selection Evolution* 45: 42. https://doi.org/10.1186/1297-9686-45-42.

Forneris, N. S., M. Bosse, M. Gautier, and T. Druet. 2025. "Genomic Prediction of Individual Inbreeding Levels for the Management of Genetic Diversity in Populations With Small Effective Size." *Molecular Ecology Resources* 25: e14068. https://doi.org/10.1111/1755-0998.14068.

Gibson, J., N. E. Morton, and A. Collins. 2006. "Extended Tracts of Homozygosity in Outbred Human Populations." *Human Molecular Genetics* 15: 789–795. https://doi.org/10.1093/hmg/ddi493.

Gorssen, W., R. Meyermans, S. Janssens, and N. Buys. 2021. "A Publicly Available Repository of ROH Islands Reveals Signatures of Selection in Different Livestock and Pet Species." *Genetics, Selection, Evolution* 53: 2. https://doi.org/10.1186/s12711-020-00599-7.

Hamzic, E. 2012. "Levels of Inbreeding Dervied From Runs of Homozygosity: A Comparison of Austrian and Norwegian Cattle Breeds (Master Thesis)." 17. Norwegian University of Life Sciences, Ås.

Harris, K., and R. Nielsen. 2013. "Inferring Demographic History From a Spectrum of Shared Haplotype Lengths." *PLoS Genetics* 9: e1003521. https://doi.org/10.1371/journal.pgen.1003521.

Hillestad, B., J. A. Woolliams, S. A. Boison, et al. 2017. "Detection of runs of homozygosity in Norwegian Red: Density, criteria and genotyping quality control." *Acta Agriculturae Scandinavica, Section A — Animal Science* 67, no. 3- 4: 107–116. https://doi.org/10.1080/09064702.2018.1501088.

Hinrichs, D., T. H. E. Meuwissen, J. Ødegard, M. Holt, O. Vangen, and J. A. Woolliams. 2007. "Analysis of Inbreeding Depression in the First Litter Size of Mice in a Long- Term Selection Experiment With Respect to the Age of the Inbreeding." *Heredity* 99: 81–88. https://doi.org/10.1038/sj.hdy.6800968.

Houwen, R. H. J., S. Baharloo, K. Blankenship, et al. 1994. "Genome Screening by Searching for Shared Segments: Mapping a Gene for Benign Recurrent Intrahepatic Cholestasis." *Nature Genetics* 8: 380– 386. https://doi.org/10.1038/ng1294-380.

Howrigan, D. P., M. A. Simonson, and M. C. Keller. 2011. "Detecting Autozygosity Through Runs of Homozygosity: A Comparison of Three Autozygosity Detection Algorithms." *BMC Genomics* 12: 460. https://doi.org/10.1186/1471-2164-12-460.

Kardos, M., A. Qvarnström, and H. Ellegren. 2017. "Inferring Individual Inbreeding and Demographic History From Segments of Identity by Descent in Ficedula Flycatcher Genome Sequences." *Genetics* 205: 1319–1334. https://doi.org/10.1534/genetics.116.198861.

Kenny, E. E., A. Gusev, K. Riegel, et al. 2009. "Systematic Haplotype Analysis Resolves a Complex Plasma Plant Sterol Locus on the Micronesian Island of Kosrae." *Proceedings of the National Academy of Sciences of the United States of America* 106: 13886–13891. https://doi.org/10.1073/pnas.0907336106.

Lavanchy, E., and J. Goudet. 2022. "Effect of Reduced Genomic Representation on Using Runs of Homozygosity for Inbreeding Characterization." *Molecular Ecology Resources* 23: 787–802. https://doi.org/10.1111/1755-0998.13755.

Lencz, T., C. Lambert, P. DeRosse, et al. 2007. "Runs of Homozygosity Reveal Highly Penetrant Recessive Loci in Schizophrenia." *Proceedings of the National Academy of Sciences* 104: 19942–19947. https://doi.org/10.1073/pnas.0710021104.

Leutenegger, A.- L., B. Prum, E. Génin, et al. 2003. "Estimation of the Inbreeding Coefficient Through Use of Genomic Data." *American Journal of Human Genetics* 73: 516–523. https://doi.org/10.1086/378207.

Luan, T., J. A. Woolliams, J. Ødegård, et al. 2012. "The Importance of Identity- by- State Information for the Accuracy of Genomic Selection." *Genetics Selection Evolution* 44: 28. https://doi.org/10.1186/1297-9686-44-28.

Luan, T., X. Yu, M. Dolezal, A. Bagnato, and T. H. Meuwissen. 2014. "Genomic Prediction Based on Runs of Homozygosity." *Genetics Selection Evolution* 46: 64. https://doi.org/10.1186/s12711-014-0064-6.

Malécot, G. 1948. "The Mathematics of Heredity." Math. Hered.

McQuillan, R., A.- L. Leutenegger, R. Abdel- Rahman, et al. 2008. "Runs of Homozygosity in European Populations." *American Journal of Human Genetics* 83: 359–372. https://doi.org/10.1016/j.ajhg.2008.08.007.

Meuwissen, T. H. 1997. "Maximizing the Response of Selection With a Predefined Rate of Inbreeding." *Journal of Animal Science* 75: 934–940. https://doi.org/10.2527/1997.754934x.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. "Prediction of Total Genetic Value Using Genome- Wide Dense Marker Maps."

*Genetics* 157: 1819–1829. https://doi.org/10.1093/genetics/157.4.1819.

Meuwissen, T. H. E., A. K. Sonesson, G. Gebregiwergis, and J. A. Woolliams. 2020. "Management of Genetic Diversity in the Era of Genomics." *Frontiers in Genetics* 11: 880. https://doi.org/10.3389/fgene.2020.00880.

Meyermans, R., W. Gorssen, N. Buys, and S. Janssens. 2020. "How to Study Runs of Homozygosity Using PLINK? A Guide for Analyzing Medium Density SNP Data in Livestock and Pet Species." *BMC Genomics* 21: 94. https://doi.org/10.1186/s12864-020-6463-x.

Monau, P. i., K. Raphaka, and S. J. Nsoso. 2022. "Genome Wide Characterisation of Runs of Homozygosity on the Main Indigenous Goat Breed of Botswana." In *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*, 956–959. Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-940-4_225.

Narasimhan, V., P. Danecek, A. Scally, Y. Xue, C. Tyler- Smith, and R. Durbin. 2016. "BCFtools/RoH: A Hidden Markov Model Approach for Detecting Autozygosity From Next- Generation Sequencing Data." *Bioinformatics* 32: 1749–1751. https://doi.org/10.1093/bioinformatics/btw044.

Pacheco, H. A., A. Rossoni, A. Cecchinato, and F. Peñagaricano. 2023. "Identification of Runs of Homozygosity Associated With Male Fertility in Italian Brown Swiss Cattle." *Frontiers in Genetics* 14: 1227310. https://doi.org/10.3389/fgene.2023.1227310.

Palamara, P. F., T. Lencz, A. Darvasi, and I. Pe'er. 2012. "Length Distributions of Identity by Descent Reveal Fine- Scale Demographic History." *American Journal of Human Genetics* 91: 809–822. https://doi.org/10.1016/j.ajhg.2012.08.030.

Pedersen, L. D., A. C. Sørensen, M. Henryon, S. Ansari- Mahyari, and P. Berg. 2009. "ADAM: A Computer Program to Simulate Selective Breeding Schemes for Animals." *Livestock Science* 121: 343–344. https://doi.org/10.1016/j.livsci.2008.06.028.

Purcell, S., B. Neale, K. Todd- Brown, et al. 2007. "PLINK: A Tool Set for Whole- Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81: 559–575. https://doi.org/10.1086/519795.

Riquet, J., W. Coppieters, N. Cambisano, et al. 1999. "Fine- Mapping of Quantitative Trait Loci by Identity by Descent in Outbred Populations: Application to Milk Production in Dairy Cattle." *Proceedings of the National Academy of Sciences* 96: 9252–9257. https://doi.org/10.1073/pnas.96.16.9252.

Santiago, E., and A. Caballero. 1995. "Effective Size of Populations Under Selection." *Genetics* 139: 1013–1030. https://doi.org/10.1093/genetics/139.2.1013.

Silva, G. A. A., A. M. Harder, K. B. Kirksey, S. Mathur, and J. R. Willoughby. 2024. "Detectability of Runs of Homozygosity Is Influenced by Analysis Parameters and Population- Specific Demographic History." *PLoS Computational Biology* 20: e1012566. https://doi.org/10.1371/journal.pcbi.1012566.

Solé, M., A.- S. Gori, P. Faux, et al. 2017. "Age- Based Partitioning of Individual Genomic Inbreeding Levels in Belgian Blue Cattle." *Genetics, Selection, Evolution* 49: 92. https://doi.org/10.1186/s12711-017-0370-x.

Sölkner, J., M. Feren, H. Schwarzenbacher, B. Gredler, and C. Fürst. 2010. "Genomic Metrics of Individual Autozygosity, Applied to a Cattle Population, Proceedings of the 61st Annual Meeting of the European Association for Animal Production." 2010. Presented at the European Association for Animal Production. 2010.

Sticca, E. L., G. M. Belbin, and C. R. Gignoux. 2021. "Current Developments in Detection of Identity- By- Descent Methods and Applications." *Frontiers in Genetics* 12: 722602. https://doi.org/10.3389/fgene.2021.722602.

Tang, K., A. Naseri, Y. Wei, S. Zhang, and D. Zhi. 2022. "Open- Source Benchmarking of IBD Segment Detection Methods for Biobank- Scale Cohorts." *GigaScience* 11: giac111. https://doi.org/10.1093/gigascience/giac111.

Te Meerman, G. J., M. A. Van Der Meulen, and L. A. Sandkuijl. 1995. "Perspectives of Identity by Descent (IBD) Mapping in Founder Populations." *Clinical and Experimental Allergy* 25: 97–102. https://doi.org/10.1111/j.1365-2222.1995.tb00433.x.

Thompson, E. A. 2013. "Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations." *Genetics* 194: 301–326. https://doi.org/10.1534/genetics.112.148825.

Vieira, F. G., A. Albrechtsen, and R. Nielsen. 2016. "Estimating IBD Tracts From Low Coverage NGS Data." *Bioinformatics* 32: 2096–2102. https://doi.org/10.1093/bioinformatics/btw212.

Wright, S. 1950. "Genetical Structure of Populations." *Nature* 166: 247– 249. https://doi.org/10.1038/166247a0.

Zhou, Y., B. L. Browning, and S. R. Browning. 2020. "Population- Specific Recombination Maps From Segments of Identity by Descent." *American Journal of Human Genetics* 107: 137–148. https://doi.org/10.1016/j.ajhg.2020.05.016.