Norwegian University of Life Sciences

Master's Thesis 2023 30 ECTS
Faculty of Biosciences

Impact of method choice on inbreeding coefficient estimator: A study of inbreeding in five beef cattle breeds in Norway

Påvirkning av metodevalg for beregning av innavlskoeffisient: En studie om innavl i fem kjøttferaser i Norge

Malin Olsen Wedaa
Animal science

Abstract

In 2020 was genotyping implemented into the populations of Norwegian beef cattle breeds of Hereford, Charolais, Aberdeen Angus, Limousin and Simmental. This development provides an opportunity to estimate genomic inbreeding within the population of these Norwegian beef cattle breeds.

Various methods were used for estimating inbreeding coefficients, including $F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ROH} and F_{ped}. When inbreeding is defined as identical-by-decent, estimators must range between 0 and 1 since they estimate probability (F_{ROH} and F_{ped}). In contrast, when inbreeding is defined as the correlation between uniting gametes, estimators can take both positive and negative values ranging from -1 to 1 ($F_{VanRaden1}$, F_{hom}, F_{uni}). The software tools PLINK.v1.9, CFC and R-studio has been used for conducting the estimates. Two approaches of analyzes for $F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ROH} in PLINK.v1.9 has been used. The first involved separate analysis by breeds, followed by a second approach where dataset contained all breeds, but each breed was coded as a family, and post-analysis was the breeds separate into individual datasets. The two approaches gave different estimates from $F_{VanRaden1}$, F_{hom}, F_{uni}, suggesting that the methods are sensitive to the reference allele frequencies. Correlations within methods conducted in the two approaches reveals that $F_{VanRaden1}$ showed to have the lowest correlation with itself. F_{ROH} had a perfect correlation and are only influenced by the fit of parameter settings.

Mean inbreeding coefficient showed to be lower than estimates from populations from other countries. The results from F_{ped} analysis in this study align closely with the finding of Kleiven (2007) for the same Norwegian beef cattle breeds in 2006. This could be explained with a substantial import of genetic material from outside Norway.

F_{ROH} divided the breeds into two groups where Hereford, Charolais and Aberdeen Angus all had an inbreeding of ~12% and Limousin and Simmental had an inbreeding of 4% and 5% respectively.

Research has consistently shown the complexity of selecting an appropriate inbreeding coefficient, and this complexity is also evident in this study. However, the study highlights the advantages of using F_{ROH}, which is not sensitive to reference allele frequencies, allows for breed comparisons, and can differentiate between new and old inbreeding.

Sammendrag

I 2020 ble genotyping implementert i populasjonene av norske kjøttferaser. Dette inkluderer rasene; Hereford, Charolais, Aberdeen Angus, Limousin og Simmental. Genotyping åpner for å utvikle rutiner for å estimere genomisk innavl i rasenes populasjoner.

Flere metoder ble brukt for å estimere innavlskoeffisienter, inkludert $F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ROH} og F_{ped}. Når innavl er definert som «identical-by-decent» må estimatorer resultere i verdier mellom 0 og 1, siden det blir estimert en sannsynlighet (F_{ROH} and F_{ped}). I kontrast, når innavl er definert som korrelasjon mellom gameter, kan verdiene ta både negative og positive verdier rangert fra -1 til 1 ($F_{VanRaden1}$, F_{hom}, F_{uni}). Dataprogrammene PLINK.v1.9, CFC og R-studio ble brukt til å utføre beregningene. To tilnærminger for å analysere datasettene ble brukt for $F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ROH} i PLINK.v1.9. Første tilnærming involverte å separat kjøre et datasett per rase gjennom PLINK.v1.9. Den andre tilnærmingen kjørte et felles datasett for alle raser der hver rase ble kodet som en familie. Etter analyse i PLINK.v1.9 ble rasene separert i individuelle datasett. De to tilnærmingene ga ulike estimater for $F_{VanRaden1}$, F_{hom}, F_{uni}, noe som antyder at metodene er følsomme for referanse allel frekvenser. Resultatene for korrelasjoner innad i metodene viser at $F_{VanRaden1}$ hadde den lavest korrelasjon med seg selv. F_{ROH} hadde en perfekt korrelasjon og påvirkes kun av parameterinnstillingene.

Resultatene fra gjennomsnittlig innavlskoeffisient har vist seg å være lavere enn estimater fra populasjoner i andre land. F_{ped} i denne studien samsvarer med funnene til Kleiven (2007) for de samme populasjonene av norske kjøttferaser i 2006. En grunn til at det ikke er funnet en økning i innavl, er den betydelige importen av genetisk materiale fra utlandet.

Resultater fra F_{ROH} delte rasene inn i to grupper, der Hereford, Charolais og Aberdeen Angus hadde en innavlsgrad på ~12 %, mens Limousin og Simmental hadde en innavlsgrad på henholdsvis ~4 % og ~5 %.

Forskning har konsekvent vist kompleksiteten ved å velge en passende innavlskoeffisient og denne kompleksiteten kommer tydelig frem i denne studien. F_{ROH} har vist seg å ha en rekke fordeler, slik som at metoden ikke er følsom for referanse allel frekvenser, tillater sammenligning av raser og kan skille mellom ny og gammel innavl.

Acknowledge

I would like to express my deepest appreciation to my three supervisors Peer Berg, Hadi Esfandyari and Siri Furre!

Peer, thank you for providing me with clear explanations of theoretical concepts and assisting me with everything from finding relevant literature to making estimations in R and PLINK. I appreciate your patience and for answering my endless stream of questions.

Hadi, your coding skills and computer knowledge have been invaluable to me. Thank you for helping me feel comfortable with my datasets and guiding me through the data analysis process.

Siri, your support has been invaluable, not just in relation to my master thesis but also in my personal and professional life, including job applications and motivation. I have felt your genuine wish for me to succeed. I am also grateful to you and your team at TYR for introducing me to beef cattle breeding, which has been a valuable learning experience.

I am also thankful to the TYR team for support and guidance throughout my internship in the autumn of 2022 and my thesis in the spring of 2023. Thank you for generously sharing your datasets and answering all of my questions. It has been a pleasure working with all of you! I sincerely hope that the digitization of inbreeding measurements I have conducted has been beneficial to your organization and has inspired you to continue exploring this innovative approach to estimating inbreeding.

I would be remiss in not mentioning my family and friends who have been an immense source of support during my years of study, and especially during my masters` writing. Despite not always understanding the field of genetics, you have always been there to listen and challenge me.

Finally, I want to express my appreciation for NMBU. It has provided a safe and stimulating environment for learning, situated right at the heart of scientific research. Additionally, the opportunity to meet and befriend students from different parts of Norway and the world has been a great privilege for which I am truly thankful for.

Best regards,

Malin Olsen Wedaa

Table of content

1.0 Introduction

1.1 Inbreeding

Since the start of the 20th century, inbreeding has been a problem in the field of genetics (Fedota O.M et al., 2016). Inbreeding is the probability of two alleles in an individual being identical by decent (IBD) due to mating of related individuals (Ghoreishifar, S., M. et al. 2020). IBD is a fundamental concept in genomics, which detects shared segments inherited from a common ancestor (Sticca et al., 2021). Inbreeding can also be described as a measure of similarity or dissimilarity of genes (VanRaden, 1992). In populations of finite size mating of related individuals is unavoidable, because the number of ancestors increases exponentially per generation (Howard et al., 2017). Efficiently characterizing and managing inbreeding levels is crucial in breeding programs to ensure that future generations meet breeding goals, maintain genetic diversity, and avoid harmful effects associated with accumulated inbreeding levels (Howard et al., 2017).

Inbreeding leads to higher levels of homozygosity in the genome, which usually results in a decreased overall fitness in the population. Homozygosity unmasks recessive deleterious alleles, because of the advantages in heterozygote dominance, is reduced (Charlesworth and Willis, 2009). When inbreeding occurs, the number of alleles which the genetic drift randomly chooses between becomes limited, which means the likelihood of drifting towards homozygous alleles is higher than without inbreeding. This genetic drift, the random fluctuation in allele frequencies, changes the allele frequencies in the population (Pekkala et al., 2014). Which may lead to reduced fitness by causing an accumulation and fixation of deleterious alleles within the population. The only way of getting rid of these deleterious alleles is introducing new individuals into the population (Lande, 1994; Lynch et al. 1995; Lynch, 1991).

1.2 Consequences of inbreeding

Inbreeding leads to several negative outcomes such as a higher probability of expressing deleterious recessive alleles, reduced fitness, reduced genetic variation within a population and a decline in response to selection (Klug et al, 2016). Mating of relatives increases the proportion of homozygous loci, which in turn increases the likelihood of offspring being affected by recessive deleterious traits (Fedota O.M et al., 2016). Due to the genetic variation in the segregating genomes from parents to offspring being less diverse than if mating would have been between non-relatives had mated. In the bigger picture this makes the gene pool, that the offsprings in the population can draw from, gets smaller and smaller with each generation. This also occurs naturally though genetic drift in finite populations.

The reduced genetic variation within a population can limit the potential for adaptation and evolutionary responses to changing environmental conditions (Leung et al., 2023). The rate of change in a population also gets reduced, as the particular traits do not have enough diversity to evolve from (Falconer, 1981).

Inbreeding, in particular, has a detrimental impact on performance and/or fitness related traits, resulting from a reduction in heterozygosity as inbreeding accumulates (Carolino and Gama, 2008). An example is the reduced performance of productive and reproductive traits in dairy cattle, where there has been an increase in the frequency of deleterious or non-deleterious recessive homozygous genotypes, leading to a decrease in genetic dominance and other non-additive-effects (Ghoreishifar, S., M. et al. 2020).

The severity of the negative effects inbreeding may have on fitness can depend on the timing of inbreeding within a population. Older inbreeding (from mating of relatives further back in the pedigree) has more time for natural selection to act and eliminate deleterious alleles that may have accumulated as a result of inbreeding (Doekes et al., 2019). In contrast, more recent instances of inbreeding may not have had sufficient time for natural selection to act and may therefore have a greater impact on fitness (Lozada-Solo et al., 2021). In their 2021 study Lozada-solo et al. (2021) found a significant effect on inbreeding accumulation on economically important growth traits particularly with recent inbreeding compared to previous inbreeding.

Doekes et al. (2019) found that recent inbreeding is more detrimental to milk and milk components yields, heifer and cow reproduction and health traits than ancient inbreeding in Dutch Holstein-Friesians. Recent inbreeding has also in a study on a population of red deer, found to be associated with reduced survival and reproduction rates, while ancient inbreeding had little to no effect (Walling, C. A. et al., 2011).

1.3 Inbreeding depression

Inbreeding depression was first observed through experimentations in plants (Darwin, 1879), but was later extrapolated to animal species (Miglior et al., 1995). Inbreeding depression causes an unfavorable increase or decrease in mean phenotype value of a particular trait in individuals (Lozada-Solo et al., 2021).

The cause of inbreeding depression has been explained by two main hypotheses. The first one is an overdominance hypothesis assuming that fitness is higher in heterozygotes than in any of the homozygotes. If this is the case, selection will favor heterozygous individuals and thus recessive alleles would be maintained (Carolino and Gama, 2008). The second one is a dominance hypothesis assuming that recessive deleterious alleles may affect fitness and that heterozygotes have a fitness closer to the wild type. Recessive alleles will in this case be purged through selection, but if mutations occur continuously the genetic load of deleterious recessive alleles will be maintained. The latter hypothesis also allows for selection under a slow increase in inbreeding, which means that the inbreeding depression will be lower than if inbreeding increased at a faster rate (Carolino and Gama, 2008). Both proposed hypotheses result in distinct consequences for the allele frequencies. Overdominant alleles are maintained at intermediate frequencies due to the balancing selection, while deleterious alleles at low frequencies are mostly young and segregate at low frequency or gets removed from the population (Alemu et al., 2020).

Inbreeding depression is expected to differ among breeds and populations, due to the function of allele frequencies at the loci affecting the traits of interest (Carolino and Gema, 2008). Furthermore, according to Carolino and Gama (2008) populations living in harsh environmental conditions often have limited access to unrelated mates, resulting in increased levels of inbreeding. This can lead to a higher expression of deleterious recessive alleles and reduced genetic variation within the population, ultimately resulting in inbreeding depression.

Additionally, harsh environmental conditions may worsen the negative effects of inbreeding on fitness, as individuals with reduced genetic variation may be less able to adapt to change in environmental conditions.

Inbreeding depression can lead to decrease in performance traits in beef cattle, such as growth rate and meat quality, in beef cattle. This is due to an increased frequency of some deleterious genotype frequencies at loci where heterozygotes deviate from the average value of homozygotes (Lozada-Solo et al., 2021).

Parallel to the previously mentioned studies, multiple analyses on various livestock species have been conducted. Table 1 provides an overview of the potential impact that inbreeding depression can have on various traits. These authors report unfavorable outcomes in growth and reproduction traits, both direct and maternal traits, production traits, as well as functional traits affecting animal welfare. The research is giving clear indications on the importance of focusing on inbreeding in populations.

Table 1: An overview of various studies that highlight the negative consequences of inbreeding depression. The first column indicates the overall effect, while the second column lists the traits that are affected. The third column represents the extent of the impact, either in terms of units of measurements, or as a decrease/increase. The fourth column contains the sources which explain the observed effects.

Inbreeding depression effect	**Traits**	**Change in traits**	**Sources**
Growth and reproduction traits	Growth weaning weight Growth on post-weaning growth First calving interval	Decrease Decrease Increase	Lozada-Solo et al., 2021 and Gutiérrez-Reinoso, M. A., 2022
Direct and maternal traits.	Calf birth weight Weight at 3, 7 and 12 months of age	Decrease Decrease	Carolino and Gama, 2008
	Fat concentration in milk Protein concentrations in milk Somatic cell scores	0.05% decrease 0.01% decrease 0.03-0.86 units increase	Gutiérrez-Reinoso, M. A., 2022
Production traits	Dystocia Stillbirths Male calves Calving interval Age at first parturition	2% increase 1% increase 0.7% increase 8.8 days increase 2.5 days increase	Gutiérrez-Reinoso, M. A., 2022
Animal welfare	Survival	Decrease	Baes et al., 2019 and Cassell et al., 2003

1.4 Methods for estimation of inbreeding coefficients

Professionals have consistently been focusing on the selective breeding of beef cattle over the last few centuries, which has led to the achievement of high-performance phenotypes. This has been achieved through the process of continuous mating of genetically related linages, resulting in an increased inbreeding levels in finite populations (Villanueva et al., 2021). To combat increased inbreeding, and ensure achievement of phenotypes in the population, it is necessary to properly estimate inbreeding coefficient (F) (Villanueva et al., 2021).

Traditionally F has been obtained from pedigree data. However, the developments of high-density single nucleotide polymorphism (SNP) panels give opportunities to implement genomic methods to estimate F (Caballero et al., 2022). There are different definitions for F, with Wright`s (1922) definition of F as a correlation between the parents` uniting gametes being the most commonly cited. In 1948 Malécot offered an alternative definition based on the probability that two homologous alleles in an individual IBD.

Measures of inbreeding can be estimated in different manners and will fit either definition better or worse, giving estimators of F different properties. When inbreeding is defined as IBD, estimators must range between 0 and 1 since they estimate probability (Alemu et al., 2020). In contrast, when inbreeding is defined as the correlation between uniting gametes, estimators can take both positive and negative values ranging from -1 to 1. The reason is that alleles are weighted alternatively, and the methods rely on correlations or covariances between genetic effects (VanRaden, 2008; Keller et al., 2011).

The method of F can provide information on relatedness among parents, population structure, recent demographic events, and mating systems (Alemu et al., 2020). There are several ways the estimation of F can be helpful for breeding programs:

(i) Conservation of genetic diversity in breeding programs, especially for endangered species or populations with limited genetic diversity to begin with.

(ii) Information on relationships between individuals before mating occurs, giving the advantage of reducing the likelihood of offspring from highly related individuals, thus maintaining genetic diversity.

(iii) Look at gene flow within population and over several years.

(iv) Calculate the rate of inbreeding.

1.5 Pedigree-based estimators

Pedigree information has been used for a long time to estimate the degree and consequences of inbreeding (Darwin, 1879; Weigel et al., 2000). However, these estimations have been shown to be fluctuating and imprecise, as well as presenting several challenges (Ghoreishifar, S., M. et al. 2020).

For instance, the need to include a, presumably unrelated, reference population present a challenge (Alemu et al., 2020). Another challenge is that pedigree-based estimation takes the expected relationship between individuals, and not the true relationship. That could give an over- or underestimating of F. An example on this is the imprecise expected relationship between full siblings, which is traditionally assumed to be 50%. However, recent study has revealed that this assumption does not accurately reflect the true relationship for all pairs of full siblings (Kenny et al., 2023).

To improve the accuracy of inbreeding coefficients based on pedigree measurements, additional sources of information are beneficial, such as estimates of genome autozygosity based on molecular markers (Gurgul, A. et al., 2016).

1.6 Genomic estimators

With the introduction of genomic markers, an increase in the prediction accuracy of an animal's genetic value is seen. Combining genomic markers with the breeding program has resulted in improved genetic progress, mainly because of its improved prediction of breeding values and the shortened generations intervals (Lozada-Solo et al., 2021). Genomic estimators have also improved the estimations of inbreeding in individuals and populations. Geno, the breed organization for Norwegian Red (dairy) are the only cattle breeding organization that routinely use genomic selection in their breeding program in Norway, while TYR, the breed organization for beef cattle plans to implement genomic selection within the next year.

Genomic estimators are robust to pedigree errors such as incomplete or missing pedigree records (Alemu et al., 2020). Genomic F captures the variation due to mendelian sampling and can therefore differentiate among individuals with the same pedigree, which can make the estimations more accurate than pedigree-based measures (Villanueva et al., 2021).

This means that even though it is expected to see full sibling sharing 50% of their segregating genome, genomic markers can find the actual proportion of genomic information they share (Kenny et al., 2023). Furthermore, the expected pedigree relationship and the absence of a member, will not change the fact of how much of the genome that is comparable between the individuals.

The availability of genomic data allows for estimations of F based on molecular markers (Caballero et al., 2022). For instance, inbreeding measures can be estimated from the diagonal elements of a genomic relationship matrix (GRM) (VanRaden, 2008), from simple heterozygosity or homozygosity measures (Szulkin et al., 2010; Bjelland et al., 2015; Ritland, 1996; Purcell et al., 2007) or from the proportion of the genome within runs of homozygosity (ROH) (McQuillan et al., 2008; Ferenčaković et al., 2013).

F estimation methods have been a topic of discussion in the scientific community (Goudet and Weir, 2018). Alemu et al. (2020) explored various methods for estimating F and assessed their efficacy in capturing allele frequencies and homozygosity. The study emphasizes the complexity of determining the best F estimator, as the optimal estimator may depend on several factors, such as allele frequency, age of inbreeding, and population demographic history. Some F estimators, such as F_{UNI} and $F_{VanRaden1}$, performed better for rare alleles, while other, such as F_{HOM} and F_{ROH}, were better suited for homozygosity at frequent alleles. Furthermore, the article stated that the optimal F estimator depends on the intended application. For instance, if the inbreeding coefficient is used to measure heterozygosity at all alleles, regardless of their frequencies or age, methods related to the proportion of autozygous genotypes (e.g., F_{hom}, F_{ROH} and F_{ped}) are recommended.

1.7 Beef cattle breeding in Norway

TYR is the national beef cattle breeding organization in Norway, responsible for improving the genetic quality of the commercial beef populations. The organization`s primary focus is to ensure genetic progress by providing breeding programs, genetic evaluations, and research to breeders and farmers (TYR, 2023). The breeding program use performance data from a national test-station for beef bulls in addition to field data from farms and slaughterhouses to perform genetic evaluation, and subsequently provide bulls for artificial insemination to promote genetically superior beef cattle. The main breeding goal is to increase the profitability and sustainability of the Norwegian beef industry through the selection of high-quality breeding stock, which includes bulls and heifers with superior growth, improved fertility and health, and good temperament. TYR works closely with farmers, veterinarians, and other stakeholders in the beef industry to ensure that the breeding program is adapted to the needs and requirements of the industry, and that it remains relevant.

The organization mainly focuses on the five largest beef cattle breeds, Hereford, Charolais, Aberdeen Angus, Limousin, and Simmental. The breeding programs for these breeds have been in place since the 1970s and have resulted in significant improvements in the quality of beef cattle in Norway. TYR currently conducts manual monitoring and control of inbreeding in these breeds based on pedigree information. The organization is currently planning to adapt genomic data as a tool for evaluating genomic inbreeding levels within the population.

The aim of this study is to assess the inbreeding status of the five major beef cattle breeds in the Norwegian breeding population of Hereford, Charolais, Aberdeen Angus, Limousin, and Simmental. This is done through comparing various methods of genomic F estimations and investigate how well pedigree-based F-estimates reflect the inbreeding estimates from genomic data.

2.0 Material and methods

2.1 Datasets

All data for this study was provided by TYR. The research material was genomic DNA obtained from routine genotyping of cows and bulls from same and different herds belonging to the beef cattle breeds Hereford, Charolais, Aberdeen Angus, Limousin and Simmental kept in Norway. Samples were collected from individuals mostly born since 2015, but there are also some individuals born before this. In Table 2 gives a detailed characterization of the genomic research material regarding the number of individuals of a given breed.

The pedigree file included 36905 individuals collectively from all breeds and in the imputed genotype file there was 5878 genotyped animals. Table 3 shows number of individuals per breed and their missing pedigree rate in percent. There are fewer records of genomic data as this is only a recent implementation in beef cattle in Norway. Genotyped data has been imputed from a 50K chip to a 100K chip with the use of two chips. In total there are 86,138 SNPs and 29 autosome chromosomes in the datasets. Imputation was already done in FImpute software (version 3.0) as a part of preparation of the datasets.

Table 2: Overview of the distribution of males, females, and total individuals for all five breeds.

Breed	**Males**	**Females**	**Total**
Hereford	347	447	794
Charolais	741	1170	1911
Aberdeen Angus	543	635	1178
Limousin	483	627	1155
Simmental	332	508	840
In total	**2446**	**3387**	**5833**

Table 3: Distribution of animals in pedigree.

Breed	No. individuals	% missing pedigree
Hereford	5612	3.56
Charolais	12207	4.92
Aberdeen Angus	5285	4.21
Limousin	8169	3.19
Simmental	5682	6.30
In total	**36905**	**22.12**

2.2 Quality control and pruning

In the datasets, 8 animals (2 Charolais, 1 Aberdeen Angus, and 5 Simmental) were excluded because of incorrect registration. The datasets has not undergone pruning for minor allele frequencies (MAF) and linkage disequilibrium (LD) as per the recommendations of Meyermans et al. (2020) since this may significantly affect medium density genotype analyses. Quality control measures for Hardy-Weinberg (p-value), observed heterozygosity per SNP, and MAF did not reveal any concerning outcomes, and hence, no SNPs were eliminated due to these factors.

2.3 Estimating inbreeding coefficients in PLINK.v.1.9

Software PLINK.v1.9, integrated with Genome-wide Complex Trait Analysis (GCTA), can estimate the inbreeding coefficient (F) from SNP data, which reflects the relationship between haplotypes within an individual (Yang et al., 2011). PLINK.v1.9 offers various estimators to calculate F including one based on the variance of additive genetic values ($F_{VanRaden1}$), and the other based on SNP homozygosity (F_{hom}). In addition, PLINK.v1.9 can also calculate F_{uni}, another F-measure based on the correlation between uniting gametes (Yang et al., 2011). Moreover, F_{ROH}, which uses information from neighboring SNPs by identifying homozygous marker sequences, can estimate the number of generations to the common ancestor by determining the length of these stretches (Alemu et al., 2020).

There are two approaches to executing PLINK.v1.9 commands on a genomic dataset. The first approach involves running all breeds jointly in a single dataset, treating each breed as a family to account for their distinct populations. The second entails running each breed separately one at a time. The choice of approach can impact the determination of the reference allele frequency, potentially leading to different outcomes for methods that are sensitive to it. Running each breed separately tends to result in a more accurate fit of the reference allele frequency to the data, whereas running each breed together, and separating them post-analysis may result in the larger breeds in determining the reference allele frequency.

2.4 Determining inbreeding coefficients

2.4.1 Software

Software PLINK v.1.9 with ported GCTA has been used for estimating genomic-based F measures. Two input files were provided to PLINK.v1.9. The first file was a .ped file consisting of family ID, individual ID, paternal ID, maternal ID, sex and a missing phenotype column. The second file was a .map file that contained information about each marker, including chromosome number (limited to autosomal chromosomes 1-29), SNP identifier, and base-pair position in bp units. Estimations for mean inbreeding coefficient, standard error, maximum and minimum values as well as the creation of figures, were performed using R-studio version 2022.7.1.554. Pedigree-based inbreeding measures (F_{ped}) were calculated using the CFC 1.0 software, which is a tool used to monitor genetic diversity from pedigree.

2.4.2 Methods for estimating inbreeding coefficient

To estimate inbreeding coefficient for the five breeds, one pedigree-based measure and four genomic measures have been utilized. The genomic measures include $F_{VanRaden1}$, F_{uni}, F_{hom} and F_{ROH}. $F_{VanRaden1}$ and F_{uni} are based on correlations between allele frequencies. F_{hom} utilized the observed and expected number of homozygous genotypes, which can be described as the excess homozygosity in the genome. F_{ROH} is based on the proportion of homozygous segments relative to the total length of the autosomal genome. The pedigree-based measure, F_{ped}, expresses the probability that the two alleles present at a given locus are IBD, meaning they have the same replicate of an ancestor's allele.

2.4.3 Equations

Following equations and theoretical background is from Yang et al., (2011) who has developed GCTA.

$F_{VanRaden1}$

$$F_{VanRaden1} = \sum[\frac{(x_i - E(x_i))^2}{h_i} - 1] = \sum[\frac{(x_i - 2p_i)^2}{h_i} - 1] \text{ and}$$

$$Var(\hat{F}_i^{VanRaden1}|F) = \sum[\frac{1-h_i}{h_i} + 7\frac{(1-2h_i)F}{h_i} - F^2],$$

where $h_i = 2p_i(1-p_i)$ is the expected heterozygotes (i.e., the number of copies of the reference allele). $(1 - p_i)^2 + p_i(1-p_i)F$, $2p_i(1 - p_i)(1 - F)$, and $p_i^2 + p_i(1 - p_i)F$ are the frequencies of the three genotypes of a SNP *i*. X_i is the number of copies of the reference allele for the i^{th} SNP. $E(x_i)$is the expected number of copies of the reference allele for the i^{th} SNP.

F_{hom}

$$F_{hom} = \sum[O(\#hom) - E(\#\text{hom})/[1 - E(\#\text{hom})] = \sum[1 - \frac{x_i(2-x_i)}{h_i}]$$

and $$Var(\hat{F}_i^{hom}|F) = \sum[\frac{(1-h_i)}{h_i} - \frac{(1-2h_i)F}{h_i} - F^2],$$

where O(# hom) and E(# hom) are the observed and expected number of homozygous genotypes in the sample, respectively. X_i is the number of copies of the reference allele for the i^{th} SNP. $H_i = 2p_i(1-p_i)$ is the expected heterozygotes (i.e., the number of copies of the reference allele).

F_{uni}

$$F_{uni} = \sum[\frac{(x_i^2 - (1+2p_i)x_i + 2p_i^2)}{h_i}] \text{ and } Var(\hat{F}_i^{uni}|F) = \sum[1 + 2\frac{(1-2h_i)F}{h_i} - F^2],$$

where the parameters are the same as in $F_{vanraden1}$ and F_{hom}.

F_{ROH}

Genomic-based inbreeding measures for runs of homozygosity (F_{ROH}) have been estimated according to Meyermans et al. (2020) recommendations with the use of PLINK.v1.9s integrated F_{ROH} estimator as follow:

$F_{ROH} = L_{ROH}/L_{aut}$, where L_{ROH} is the total length of all ROHs in the individual`s genome and L_{aut} is the length of the autosomal genome.

To define a runs-of homozygosity (ROH), PLINK.v1.9 uses a sliding window approach searching SNP data to detect homozygous stretches covering a SNP. ROH detection is performed by applying the parameters outlined in Table 4, which have been calibrated based on recommendations from Winnberg (2020) and Meyermans et al. (2020). To determine the best fitting parameter values, a simulated animal with a fully homozygous genotype was introduced and analyzed, following the methodology described by the before mentioned authors.

The PLINK.v1.9 algorithm for ROH detection operates in the following manner:

1. The algorithm loads SNP data and continuously scans the genome using a sliding window approach. For each individual SNP, a score is assigned based on the proportion of times it appears in a completely homozygous window.
2. A predetermined threshold is used to determine if a putative ROH is valid for further analysis. For instance, if the threshold is set to 0.05, a SNP must appear in at least 5% of a completely homozygous window to be considered part of a ROH.
3. The algorithm checks for certain criteria, such as the maximum allowed gap between SNPs and the maximum allowed number of heterozygotes, for the final ROH. If the criteria are met, the putative ROH proceeds to the next step. Otherwise, the segments are split up and re-evaluated.
4. Finally, the putative ROH are evaluated for minimum SNP density (kb/SNP) and minimum required length in kb and number of SNPs. If these criteria are not met, the segment is discarded.

Table 4: PLINK.v1.9 commando line commands and parameter settings with this study's parameter settings.

Detection	Parameter	PLINK.v.1.9 command	Values
Defining sliding window.	The size of the sliding window measured in the number of SNPs.	-homozyg-window-snp	57*
	The maximum allowed number of heterozygotes within a window.	-homozyg-window-het	0**
	The maximum number of missing SNPs within a window.	-homozyg-window-missing	1**
Identifying ROH.	The proportion of completely homozygous windows.	-homozyg-window-threshold	0.026*
Check point for putative ROH.	The largest allowable distance between consecutive SNPs.	-homozyg-gap	300*
	The maximum allowed number of heterozygoutes in the final segment.	-homozyg-het	0**
Minimum SNP density and ROH length.	The minimum SNP density required to call a ROH.	-homozyg-density	40*
	The minimum length in kb required to call a ROH.	-homozyg-kb	400**
	The minimum number of SNPs required to call a ROH.	-homozyg-snp	57*

*Estimated values shown in the attachment ** Recommended values

Fped

CFC utilizes the efficiency of the underlying algorithm by utilizing the inverse of the numerator relationship matrix (A^{-1}) (Sargolzaei, M. et al, 2006). This approach has been proven to be fast and minimizes memory requirements when calculating inbreeding and coancestries (Colleau, 2002; Sargolzaei et al., 2006).

Henderson introduced the technique for computing A^{-1} in 1979 and theory for the next paragraph are from his originally article from that year. The animals involved in the calculation of A-1 are labeled as 1, 2, …, n and must be arranged in a specific order where parents come before their offspring. Additionally, animals labeled 1, …, b are considered a base population, which is assumed to be unrelated and non-inbred. To begin the upper left submatrix of the base population are set to 1. Then, the submatrix is successively expanded in chronological order until the full matrix is obtained.

2.5 Pearson correlation matrix

In R-studio, the "ggpubr" package is used to estimate Pearson correlation. This measure indicates a linear dependence between two variables and is dependent on the distribution of the data. The equation for Pearson correlation (r) is as follows:

$$r = \frac{\sum(x - m_x)(y - m_y)}{\sqrt{\sum(x - m_x)^2 \sum(y - m_y)^2}}$$

3.0 Results

3.1 F estimation

The estimates provided by $F_{VanRaden1}$, F_{hom} and F_{uni} are all on a scale from -1 to 1 where 0 represents the expectation of random mating. In contrast, F_{ped} and F_{ROH} provide estimates on a scale from 0 to 1, where 1 represents an animal that is 100% inbred.

The F_{ROH} calculations gives results that divides the breeds into two groups with different mean inbreeding levels. The group of Hereford, Charolais and Aberdeen Angus have the highest level of inbreeding, with same mean inbreeding values of 12%. Limousin and Simmental are less inbred with the similar mean inbreeding values of 4% and 5%.

When the breeds are analyzed separately the results from $F_{VanRaden1}$, F_{hom} and F_{uni} indicate less inbreeding than expected under random mating. However, when the datasets are analyzed jointly and then separated by breed in post-analysis, the results indicate more inbreeding than expected under random mating. In contrast, F_{ROH} yields consistent results regardless of the analytical approach, which also applies as expected to F_{ped}.

Table 5 and 6 highlights that some animals exhibit remarkably high levels of mean inbreeding, as noted in the maximum column. No method generates an F-estimate of zero. A method to measure the uncertainty in the estimate of the average F is the standard error (std.error) (Altman and Bland, 2005). Estimates in Table 6 reveals that $F_{VanRaden1}$ has the highest std. error compared to F_{hom} and F_{uni} and indicates that the average F for $F_{VanRaden1}$ is the furthest away from the true average F. F_{ped} and F_{ROH} have a small std.error.

Table 5: Results from analysis of F-coefficient for the five methods. Datasets are run separately through PLINK.v.1.9. Standard error (std.error), Maximum value (max) and minimum value (min) are for the estimated mean inbreeding.

Method	Breed	Mean	Std.error	Max	Min
$F_{VanRaden1}$	Hereford	-0.0334	0.0018	0.2993	-0.1621
	Charolais	-0.0164	0.0011	0.2418	-0.1682
	Aberdeen Angus	-0.0285	0.0020	0.5618	-0.1762
	Limousin	-0.0223	0.0011	0.2227	-0.0974
	Simmental	-0.0028	0.0017	0.4159	-0.2109
F_{hom}	Hereford	-0.0039	0.0016	0.2869	-0.2864
	Charolais	-0.0015	0.0008	0.3023	-0.1922
	Aberdeen Angus	-0.0010	0.0016	0.3199	-0.5454
	Limousin	-0.0016	0.0009	0.2830	-0.2824
	Simmental	-0.0053	0.0013	0.2506	-0.3604
F_{uni}	Hereford	-0.0039	0.0009	0.3271	-0.0527
	Charolais	-0.0015	0.0005	0.2796	-0.0469
	Aberdeen Angus	-0.0010	0.0009	0.2748	-0.0612
	Limousin	-0.0016	0.0005	0.2583	-0.0354
	Simmental	-0.0052	0.0008	0.2568	-0.0635
F_{ped}	Hereford	0.0124	0.0018	0.3131	2.86102e-06
	Charolais	0.0078	0.0004	0.2537	9.53674e-07
	Aberdeen Angus	0.0159	0.0007	0.2684	4.577764e.05

	Limousin	0.0074	0.0006	0.2518	7.15256e-07
	Simmental	0.0112	0.0007	0.2535	1.43051e-06
F_{ROH}	Hereford	0.1169	0.0009	0.3707	0.0633
	Charolais	0.1168	0.0005	0.3703	0.0632
	Aberdeen Angus	0.1175	0.0008	0.3872	0.0348
	Limousin	0.0389	0.0005	0.3036	0.0014
	Simmental	0.0519	0.0007	0.2973	0.0177

Table 6: Results from analysis of F-coefficient for the five methods. Datasets are run jointly for all breeds through PLINK.v.1.9 and divided into separate dataset for each breed afterward.

Method	Breed	Mean	Std.error	Max	Min
$F_{VanRaden1}$	Hereford	0.0553	0.0009	0.2972	0.0056
	Charolais	0.0017	0.0008	0.2583	-0.0639
	Aberdeen Angus	0.1241	0.0010	0.5487	0.0636
	Limousin	0.0194	0.0006	0.2586	-0.0228
	Simmental	0.0553	0.0009	0.2972	0.0056
F_{hom}	Hereford	0.0771	0.0009	0.3211	0.0259
	Charolais	0.0471	0.0007	0.0341	-0.1581
	Aberdeen Angus	0.0849	0.0010	0.3730	-0.24742
	Limousin	0.0761	0.0006	0.3294	-0.0628
	Simmental	0.0771	0.0009	0.3212	-0.1159
F_{uni}	Hereford	0.0721	0.0007	0.3149	0.0331
	Charolais	0.0302	0.0005	0.3048	-0.0047
	Aberdeen Angus	0.1103	0.0009	0.3841	0.0344
	Limousin	0.0536	0.0005	0.3049	-0.0142
	Simmental	0.0721	0.0007	0.3150	0.0331
F_{ped}	Hereford	0.0124	0.0018	0.3707	2.86102e-06
	Charolais	0.0078	0.0004	0.3703	9.53674e-07
	Aberdeen Angus	0.0159	0.0007	0.3872	4.577764e.05

	Limousin	0.0074	0.0006	0.3036	7.15256e-07
	Simmental	0.0112	0.0007	0.2973	1.43051e-06
F_{ROH}	Hereford	0.1169	0.0009	0.3707	0.0633
	Charolais	0.1168	0.0005	0.3703	0.0632
	Aberdeen Angus	0.1175	0.0008	0.3872	0.0348
	Limousin	0.0389	0.0005	0.3036	0.0014
	Simmental	0.0519	0.0007	0.2973	0.0177

3.2 Mean inbreeding trends

To give a clearer representation of how trends in the breeds have been evolving annually, trendlines have been restricted to the years between 2015 and 2021 due to limited records outside of this time period ($n < 4$ as max records per breed/year). When reviewing the following plots, it should be noted that the methods $F_{VanRaden1}$, F_{hom} and F_{uni} will produce values ranging from -1 to 1, while F_{ROH} and F_{ped} will produce values ranging from 0 to 1.

The methods based on correlation, $F_{VanRaden1}$, F_{hom} and F_{uni}, display different trends. It appears that $F_{VanRaden1}$ deviates the most from the trendlines indicated by the two other correlation methods. This is especially clear in Hereford figure 1 and Limousin figure 4.

Estimates from methods, F_{ped} and F_{ROH}, exhibit similar trends over time. However, F_{ped} consistently yields lower F-values than F_{ROH}.

The trendlines for all methods indicate a stable inbreeding trend across all breeds, except for $F_{VanRaden1}$ which shows a greater decresing slope of inbreeding in Hereford, Charolais, and Limousin.

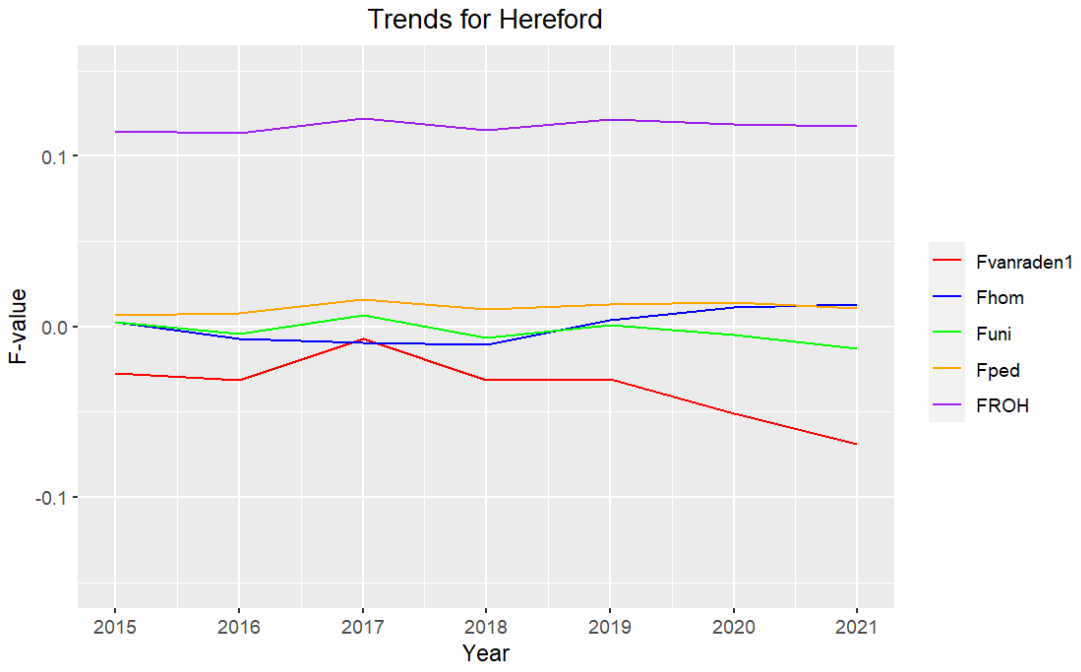

Figure 1: Mean inbreeding coefficient in the Hereford population per year from 2015 to 2021. Trendlines are shown for five different methods ($F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ped} and F_{ROH}).

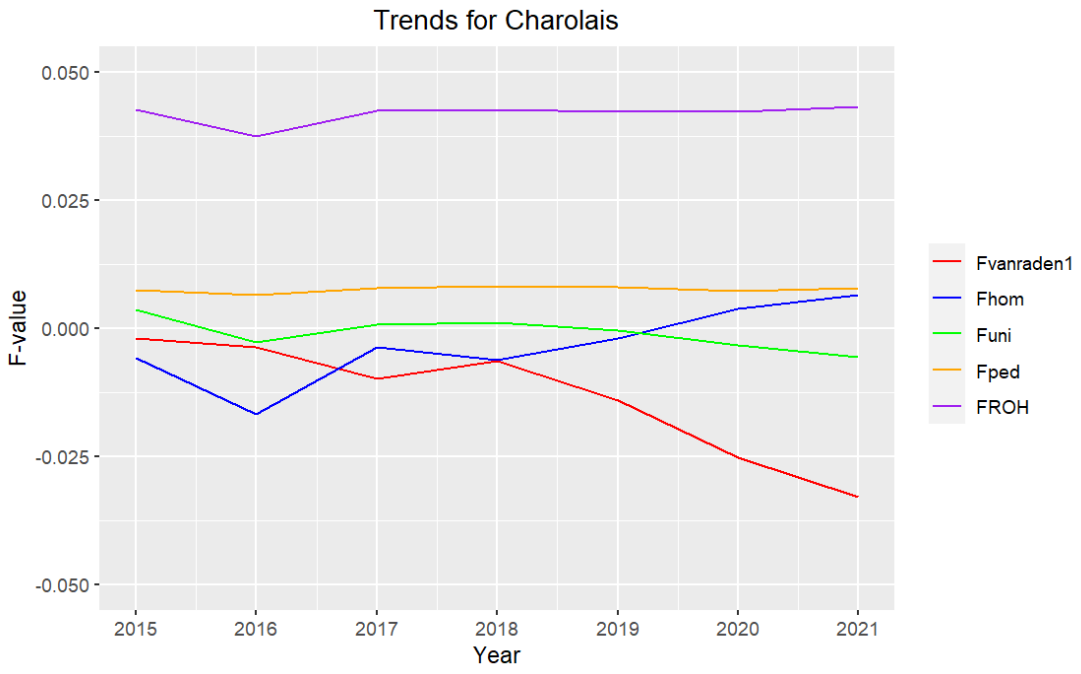

Figure 2: Mean inbreeding coefficient in the Charolais population per year from 2015 to 2021. Trendlines are shown for five different methods ($F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ped} and F_{ROH}).

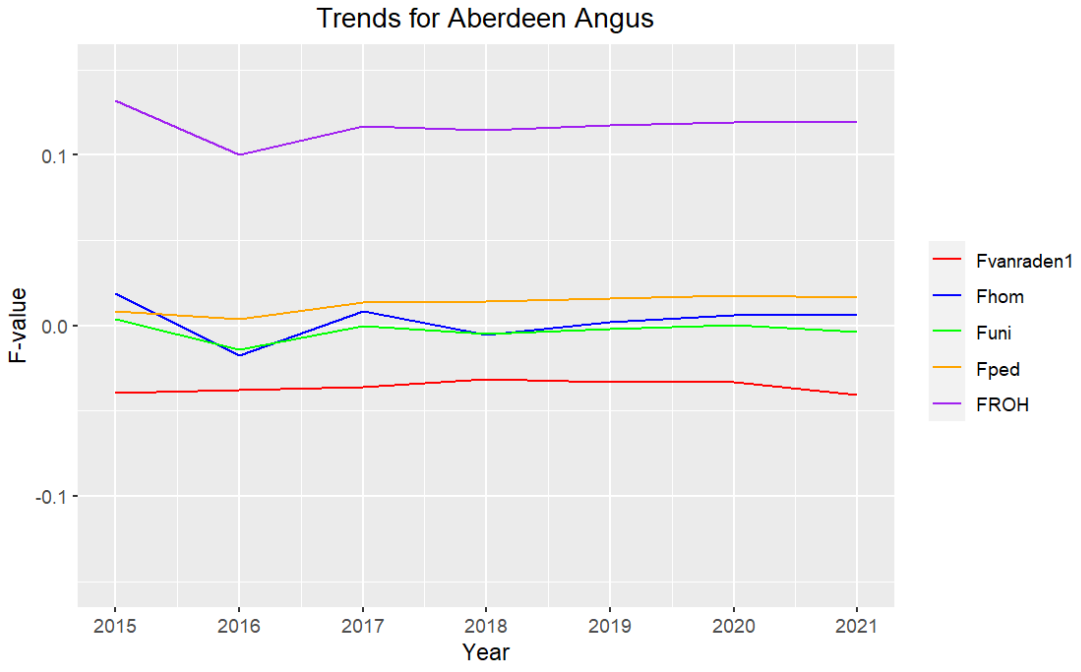

Figure 3: Mean inbreeding coefficient in the Aberdeen Angus population per year from 2015 to 2021. Trendlines are shown for five different methods ($F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ped} and F_{ROH}).

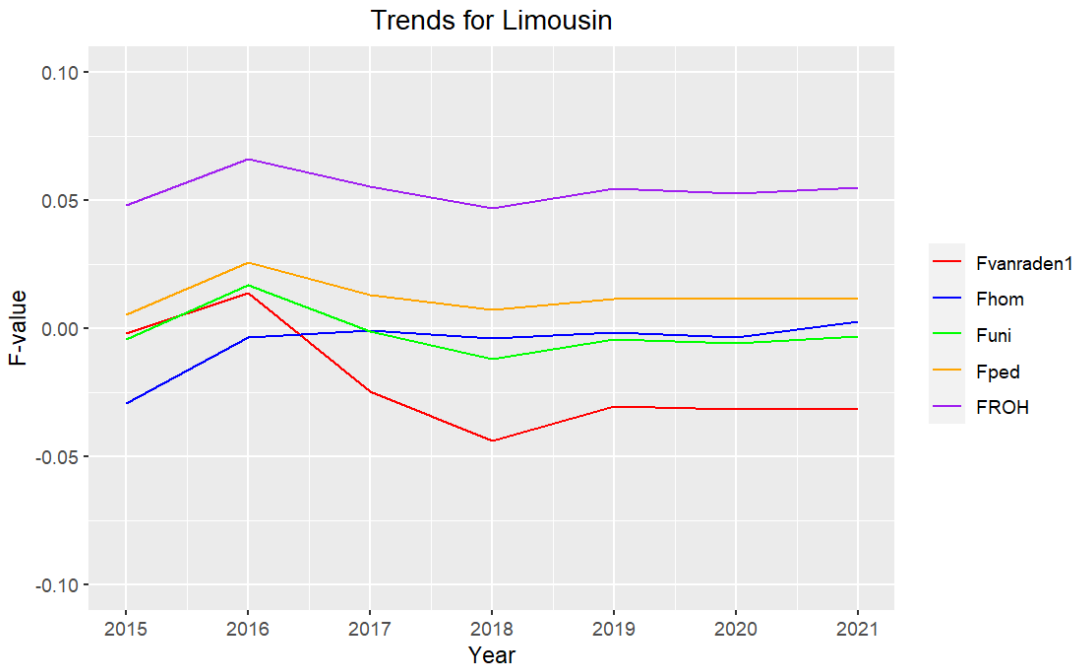

Figure 4: Mean inbreeding coefficient in the Limousin population per year from 2015 to 2021. Trendlines are shown for five different methods ($F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ped} and F_{ROH}).

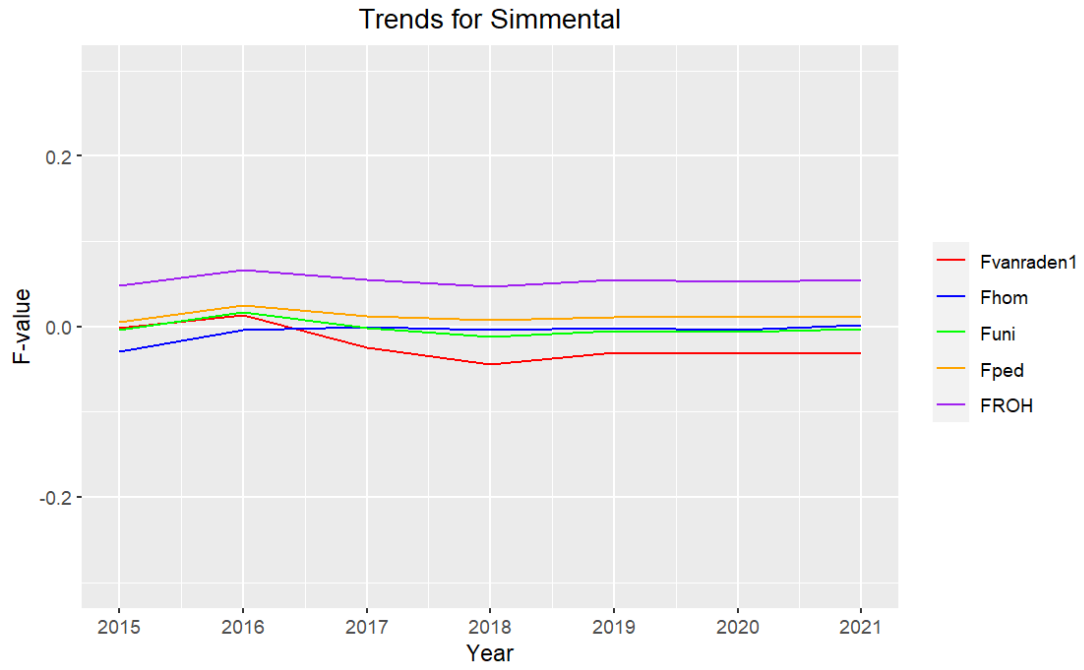

Figure 5: Mean inbreeding coefficient in the Simmental population per year from 2015 to 2021. Trendlines are shown for five different methods ($F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ped} and F_{ROH}).

3.2 Pearson correlation matrix

3.2.1 Correlation between methods of F-estimates

Table 7 to 11 shows Pearson correlation matrices for the different breeds when data were ran separately by breed through PLINK.v.1.9. $F_{VanRaden1}$ exhibits the highest correlation with F_{uni}, while F_{hom} and F_{uni} shows the highest correlation with F_{ROH}. This result is consistent for all breeds. Furthermore, F_{ped} displays the highest correlation with F_{ROH} in four out of five breeds, except for Hereford where F_{uni} has the highest correlation with F_{ROH}. Finally, F_{ROH} is highest correlated with F_{uni} in all breeds.

Results show that $F_{VanRaden1}$ yields the most divergent F-estimates compared to the other methods, and therefore has the lowest correlations with the other methods. On the other hand, F_{ROH} exhibits the highest degree of concordance with the methods.

Table 7: Correlation matrix between five different methods ($F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ped} and F_{ROH}) for the Hereford dataset. Highlighted in orange are the strongest correlation between each of the methods.

Method	$F_{VanRaden1}$	F_{hom}	F_{uni}	F_{ROH}	F_{ped}
$F_{VanRaden1}$	1.000				
F_{hom}	-0.284	1.000			
F_{uni}	0.660	0.532	1.000		
F_{ROH}	0.244	0.807	0.848	1.000	
F_{ped}	0.313	0.434	0.617	0.555	1.000

Table 8: Correlation matrix between five different methods ($F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ped} and F_{ROH}) for the Charolais dataset. Highlighted in orange are the strongest correlations between each of the methods.

Method	$F_{VanRaden1}$	F_{hom}	F_{uni}	F_{ROH}	F_{ped}
$F_{VanRaden1}$	1.000				
F_{hom}	-0.430	1.000			
F_{uni}	0.687	0.361	1.000		
F_{ROH}	0.318	0.649	0.851	1.000	
F_{ped}	0.280	0.418	0.625	0.695	1.000

Table 9: Correlation matrix between five different methods ($F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ped} and F_{ROH}) for the Aberdeen Angus dataset. Highlighted in orange are the strongest correlations between each of the methods.

Method	$F_{VanRaden1}$	F_{hom}	F_{uni}	F_{ROH}	F_{ped}
$F_{VanRaden1}$	1.000				
F_{hom}	-0.301	1.000			
F_{uni}	0.650	0.529	1.000		
F_{ROH}	0.333	0.751	0.895	1.000	
F_{ped}	0.262	0.441	0.584	0.605	1.000

Table 10: Correlation matrix between five different methods ($F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ped} and F_{ROH}) for the Limousin dataset. Highlighted in orange are the strongest correlations between each of the methods.

Method	$F_{VanRaden1}$	F_{hom}	F_{uni}	F_{ROH}	F_{ped}
$F_{VanRaden1}$	1.000				
F_{hom}	-0.384	1.000			
F_{uni}	0.573	0.537	1.000		
F_{ROH}	0.344	0.667	0.907	1.000	
F_{ped}	0.290	0.414	0.632	0.678	1.000

Table 11: Correlation matrix between five different methods ($F_{VanRaden1}$, F_{hom}, F_{uni}, F_{ped} and F_{ROH}) for the Simmental dataset. Highlighted in orange are the strongest correlations between each of the methods.

Method	$F_{VanRaden1}$	F_{hom}	F_{uni}	F_{ROH}	F_{ped}
$F_{VanRaden1}$	1.000				
F_{hom}	-0.406	1.000			
F_{uni}	0.716	0.347	1.000		
F_{ROH}	0.386	0.620	0.870	1.000	
F_{ped}	0.263	0.353	0.540	0.600	1.000

3.2.2 Correlations within methods of estimating F

The correlation analysis (Table 12) within F methods when datasets are run separately vs. jointly, indicates that $F_{VanRaden1}$ is dependent on data structure as it has a low and inconsistent correlation coefficient between the two approaches, in all breeds. F_{hom} has the weakest correlation after $F_{VanRaden1}$, followed by F_{uni}, which exhibits a slightly stronger correlation. On the other hand, F_{ROH} shows a perfect correlation of 1 for all breeds. F_{ped} is not included as the pedigree of animals in one breed is not affected by the pedigree of animals in other breeds.

Results suggest that $F_{Vanraden1}$, F_{hom} and F_{uni} is influenced by the reference allele frequency. Therefore, will the composition of the animals included for estimation be crucial for the results.

Table 12: Correlation between F methods when datasets were ran separately or jointly in all breeds.

Method/breed	Hereford	Charolais	Aberdeen Angus	Limousin	Simmental
$F_{VanRaden1}$	0.369	0.798	0.500	0.493	0.493
F_{hom}	0.859	0.873	0.792	0.790	0.790
F_{uni}	0.881	0.937	0.934	0.950	0.950
F_{ROH}	1.000	1.000	1.000	1.000	1.000

4.0 Discussion

4.1 Results for F compared to the latest F-estimates in the Norwegian population

The latest inbreeding estimates in the Norwegian population of beef cattle dates back to 2007 (Kleiven, 2007), when a population structure analysis of the five big beef cattle breeds in Norway were conducted. Kleiven (2007) used F_{ped} to calculate her F-measurements, and Table 13 compares F-estimates from 2006 and from this study. Table 13 shows no increase in F-estimates which is unexpected in a finite population.

Table 13: Comparison of F_{ped} mean inbreeding coefficients from this study and the latest inbreeding estimates done in the Norwegian population of beef cattle (Kleiven, 2007).

Breed/year	2006 (Kleiven, 2007)	2022 (this study)
Hereford	0.012	0.015
Charolais	0.008	0.008
Aberdeen Angus	0.013	0.012
Limousin	0.007	0.007
Simmental	0.013	0.011

The substantial use of foreign breeding sires, facilitated by artificial insemination and embryo import, has an impact on the estimated mean inbreeding values of the Norwegian population. The imported genetic material will not change the individual inbreeding coefficients for the current population when using F_{ped} as a method of estimation. Thereof, it could change the mean inbreeding as that is the average over all individuals in the population. Imported animals are in most cases unrelated to the population, which leads to the migration of new allele frequencies to the population. As a result, it is considered that the foreign sires have an impact on the genetic diversity and evolution of the population. One might argue that in a population without prior imports, animals that were already inbred could appear even more inbred when new individuals are introduced. This is because the imported animals are not related to the existing population, and therefor contribute to genetic diversity in a different manner than the existing population, leading to a portion of the population looking even more similar to each other.

4.2 F-estimates from two approaches

The result for mean inbreeding, when the datasets are run jointly or separately, only differs between $F_{VanRaden1}$, F_{hom} and F_{uni}, but not for F_{ped} and F_{ROH}. These three former methods are based on reference allele frequencies and results indicate that the methods are sensitive for changes in the frequencies. It was expected that F_{ped} did not change between the two approaches of running datasets, as the method is based on relatives that won`t change either way the dataset is runed.

The jointly analysis of $F_{VanRaden1}$, F_{hom} and F_{uni}, suggest that the breeds appear to be more inbred than at random mating, compared to the separately analysis found that they are less inbred at random mating. This could be due to the fit and sensitivity of the reference allele frequencies. When the reference allele frequencies are averaged across all five breeds, individuals may appear to have more rare alleles than expected with random mating. This was particularly evident in Aberdeen Angus, which resulted in a higher inbreeding estimate compared to the other breeds for $F_{VanRaden1}$, F_{hom} and F_{uni}. This difference was not so apparent when the datasets was analyzed separately.

An advantage of running the datasets jointly, is the possibility to compare inbreeding across breeds. On the other hand, when the datasets are analyzed separately, the breeds do not have the same reference allele frequencies, and as a result they cannot easily be compared to each other. This could also be implied as an argument for comparisons of the same breed across different research material. A solution could be to set the same reference allele frequencies manually for the methods, but then it would not give the best fit for the populations.

Hereford and Simmental had the same estimates of mean inbreeding when ran jointly for all methods dependent on the reference allele frequency ($F_{VanRaden1}$, F_{hom} and F_{uni}), but exhibited difference in maximum and minimum values. These two breeds had the smallest number (Table 2) of individuals in the dataset, which suggest that their impact on the reference allele frequency may have been limited, resulting in a less-than-optimal fit. Then, when the datasets were ran separately, thereof showing a reference allele frequency better fitted to each breed, F_{hom} and F_{uni} gave the identical mean inbreeding in all breeds. There is on the other hand a different in maximum and minimum values, indicating that different animals are detected with high or low inbreeding depending on the approach.

F_{ROH} showed no change in results in the two approaches for analyzing the datasets. This could be attributed to the fact that F_{ROH} is conditional on the settings used for its calculation, and if the settings remain the same, the results for the same population will not change. Before estimating F_{ROH} is it crucial to ensure that the parameters are properly fitted to the model before estimating F_{ROH}, rather than relying on default settings in a software such as PLINKv.1.9 (Meyermans et al., 2020). In order to make proper comparisons of the results, it is important to be aware of the parameter settings and calculations used. The results from runs with different parameter setting for F_{ROH} is included in the attachment to facilitate comparisons with other studies.

When datasets are run jointly there is a clear difference between Aberdeen Angus F-estimates for $F_{VanRaden1}$, F_{hom} and F_{uni} and the other four breeds. Indicating that Aberdeen Angus has a higher inbreeding than in the other four breeds.

4.3 Comparisons of F estimation in other literature

In the study conducted by Decker et al. (2014), an examination of ancestry, divergence, and admixture in domesticated cattle, revealed interesting patterns. The results indicated a closer genetic drift between Hereford and Aberdeen Angus, while Simmental and Limousin showed a closer affinity to each other. Charolais got positioned between the groups, but more towards the group of Simmental and limousine. Based on these results, one would have expected to see the same pattern in the study, but that is not the case.

The F_{ROH} results grouped Hereford, Charolais and Aberdeen Angus together, while Simmental and Limousin were in a separate group. This contradicts the findings of Decker et al. (2014), which found that Charolais would be grouped with Simmental and Limousin. Charolais and Limousin both have origins in France, while Simmental is from neighboring Switzerland. Meanwhile, Hereford and Aberdeen Angus are both from the United Kingdom (UK), making it more logical for Charolais to be grouped with the French-origin breeds rather than the UK group.

Considering the population structure of the five breeds, the categorization made by F_{ROH} appears logical because some of the breeds are made up from different lines. Limousin and Simmental are divided into two distinct sublines within their respective breeds, resulting in greater genetic diversity compared to Hereford and Aberdeen Angus, which has closed pedigree containing only one line. Charolais is derived from two smaller sublines within the breed, further contributing to its genetic variability, thereof more alike Limousin and Simmental in population structure.

Charolais is the most numerous breed in the study, it is also the breed with the overall lowest mean inbreeding for the three methods of $F_{VanRaden1}$, F_{hom} and F_{uni}. The results of this study might be biased due to the animals included in the analyzes. Results from Lozada-Soto et al. (2021) shows a higher mean genomic and pedigree inbreeding of the American Angus than what the results of this study have been estimating. On the other hand, another study Kasarda et al. (2020) found Hereford and Aberdeen Angus to have a higher level of genomic inbreeding compared to among others, Charolais and Limousin. Which is corroborated in this study.

4.4 Comparisons and correlation of methods to estimate F

$F_{VanRaden1}$ estimates are distinct to the other methods, as they show major differences in the correlation values. The correlations show $F_{VanRaden1}$ generally had a very low correlation coefficient with the other methods, except for F_{uni} where it had a medium correlation coefficient. Both F_{uni} and $F_{VanRaden1}$ are two methods relying on correlations and give more weight to homozygosity at rare alleles (VanRaden, 2008; Keller et al., 2011). This has been discussed in Alemu et al. (2020) which found that F_{uni} and $F_{VanRaden1}$ are good at capturing rare alleles with an allele frequency under 0.10. However, when the frequency is increasing F_{hom}, F_{ROH} and F_{ped} are more efficient. This might explain the differences in correlations found in the present study.

F_{uni} gives greater weight to SNP`s with rare alleles compared to F_{hom} which has equal weight to all SNPs (Alemu et al., 2020). Zhang et al. (2021) also pointed out that there is a bias of F_{uni}, namely the role of individual average kinship. Even though the methods exhibit the same mean inbreeding for the five breeds, they have a low to medium correlation between them (r = 0.3-0.5). The statement of Zhang et al. (2021) and the low correlation can indicate that the methods measure F of individual animals differently, but are in the total amount of inbreeding estimating the same degree of inbreeding to the population.

Zhang et al. (2021) argued that when using allele-sharing inbreeding estimators (in this study: $F_{VanRaden1}$, F_{uni} and F_{hom}) for inbreeding, it's more of a "within-population" measure rather than a "within-breed" measure. This is because the reference for pairwise coancestry is determined by the population of individuals whose allele frequency is used to measure inbreeding. Therefore, the choice of a reference population for inbreeding measure can introduce bias in the results. This has implications for this study, as the limited amount of available research material may be biased towards the animals which are genotyped, and the results for F-values may only represent the "within-genotyped population". There are approximately 111000 beef cattle in Norway (Animalia, 2023) while the number of genotyped individuals in this study is 5878. The low number of genotyped in this study compared to the total number of beef cattle in Norway indicates that the F-measurements might be biased and represent the genotyped-population.

Schiavo et al. (2020) found that their estimates of F_{hom} correlated strongly with F_{ROH} measures in their studies of pig breeds, which is consistent with the results of the thesis study and with those of Zhai et al. (2015). The authors concluded that F_{ROH} better captures inbreeding information in the breeds analyzed and could complemented pedigree-based inbreeding coefficients for the management of these genetic resources.

Lozada-Solo et al. (2021) investigated the correlations among various methods including F_{ROH}, $F_{VanRaden1}$, and F_{ped} to estimate inbreeding in American Angus cattle. In general, their results showed slightly higher positive correlations between F_{ROH}, $F_{VanRaden1}$, and F_{ped} than those found in the thesis study. In this study F_{ROH} shows the highest correlation with all the other four methods, in particular with F_{ped} and F_{uni}. However, Lozada-Solo et al. (2021) operated with a much larger sample size and a longer time span than this study.

Cortellari et al. (2022) concluded that pedigree depth can influence the correlations between F_{ped} with F_{ROH} and $F_{VanRaden1}$. There has been evidence in this study that there is also a low correlation between these methods. Lozada-Solo et al. (2021) highlights that $F_{VanRaden1}$ values are particularly susceptible to variation as they are strongly influenced by the frequency of rare allele variants. Villanueva et al. (2021) concluded that obtaining genomic inbreeding coefficients from diagonal elements of genomic matrices can result in inconsistent outcomes.

The results of this study found that $F_{VanRaden1}$ exhibits the lowest correlation among the tested methods, which may indicate inconsistencies in the methods ability to capture genetic variation. A reason might be that the genetic variation detected by $F_{VanRaden1}$ method has been influenced by factors such as rare allele variants. This may also explain the variation observed when running datasets jointly or separately in PLINK.v.1.9.

Ghoreishifar, S., M. et al. (2020) found that F_{uni}, F_{hom} and $F_{VanRaden1}$ are highly influenced by allelic frequencies, which is consistent with the results presented in Tables 5 and 6 of this study. This also leads to the observed variations in correlations between methods both within and between breeds. The authors also reported a strong correlation ($r > 0.9$) between F_{ROH} and F_{hom}/F_{uni}, which is consistent with the presents study`s finding of a strong correlation between F_{ROH} and F_{uni} ($r > 0.8$), while F_{hom} and F_{ROH} showed slightly lower but still strong correlation ($r > 0.6$).

4.5 Trendlines

Trendlines obtain from the results did not show any steep lines indicating an increase of the mean inbreeding over the last seven years, but these results could have been different if there were more animals with genotype records, including older animals, had been available for inclusion. It would have been better if it was possible to know the rate of inbreeding over several generations, rather than mean inbreeding per year. According to food and Agriculture Organization of the United Nations (FAO) guidelines, it is recommended to limit the rate of inbreeding to a maximum of 1% per generation. This recommendation should be considered when sufficient data is available to estimate the rate of inbreeding per generation in the Norwegian beef cattle breeds.

4.6 Consequences of inbreeding

Breeding programs have the object of maintaining genetic diversity and limit the increase in inbreeding to maximize response to selection. This can be achieved by increasing the effective population size and controlling the rate of inbreeding (Ghoreishifar, S., M. et al. (2020). There is a difference between breeding programs, while the commercial breeding programs aim for an increase in some particular traits (e.g., meat quality and/or protein in milk) others aim for conservation of the breed. There are several studies (Table 1) concluding that inbreeding has a negative effect on various traits, that could impact the profitability for the farmer and the breeding program. This is traits as survival (Baes et al., 2019 and Cassell et al., 2003), calf birth weight (Carolino and Gama, 2008) and calving interval (Gutiérrez-Reinoso, M. A., 2022). For the commercial breeding programs the aim should be to continue taking care of the genetic diversity in the population, as without this diversity there would not be possible to improve the population in the long term. If the genetic diversity is lost, it could only be increased through crossbreeding. Maintaining as much genetic diversity as possible could be done by having broad breeding goal and use an approach to limit the mating between relatives.

Since these populations have a finite size, inbreeding is expected to increase over time (Howard et al., 2017) and the goal should be to limit its rate. With incorporation of genotyping into beef cattle breeding programs in Norway, it is also possible to facilitate the use of genomic estimators to monitoring and estimate inbreeding in the populations.

Incorporating such estimators into routine evaluations and mating decisions would be a valuable tool to limit the rate of inbreeding over time.

There is no doubt that high levels of inbreeding can be harmful and should be monitored in populations. As mentioned in the Table 1 high levels of inbreeding can negatively impact not only animal welfare, but also traits that impact profitability of farmers due to reduced performance in various traits. Gutiérrez-Reinoso et al. (2022) recommended that genotypes for health traits (e.g., immune system and general health status of the individual) should be incorporated into the aspects of dairy cattle evaluations, selection, and breeding systems. This approach could also be considered for beef cattle breeding programs now that genotyping is developing. It is presumably that it is better to limit the damage inbreeding can cause to the population, as this only can be fixed through crossbreeding.

5.0 Conclusion

Low correlation within the methods $F_{vanraden1}$, F_{uni} and F_{hom} suggests that the animals with high inbreeding in the two approaches (jointly or separately) are not the same. It also implies that the methods are sensitive to the reference allele frequencies, and the makeup of population in the analysis can lead to biased results. Correlations between the two approaches of analyzing datasets were particularly evident for $F_{VanRaden1}$. The three methods are sensitive to reference allele frequencies, which is undoubtably important to consider when studying the levels of inbreeding and correlations between methods.

$F_{VanRaden1}$ is sensitive to rare alleles, and as demonstrated by Villanueva et al. (2021), it may not accurately reflect genetic variation. Results indicate that the method for $F_{VanRaden1}$ is the most sensitive to the composition of the population's allele frequencies and thereover the composition of the population analyzed. This may suggest that $F_{VanRaden1}$ is more biased than the other methods.

F_{ped} does not accurately represent the realized relationship among families and is therefore not the best parameter for determining inbreeding levels in a population, especially when used alone.

F_{ROH} is highly dependent on the chosen settings and should be tailored to fit the data, with full disclosure of the chosen parameters in any research using this method. F_{ROH} has shown the highest correlation among the methods used in this study and is shown to give additional advantages in comparisons across breeds or populations. Also, it also has the property of being able to differentiate new and old inbreeding. Therefore, it would be recommended for reflecting inbreeding levels in a population. It is unique in that it can accurately reflect true inbreeding rather than just expected values, as is the case with F_{ped}.

Depending on what the main goal with the estimations of F, it could be advantageous to use one or the other approach when analyzing datasets for different breeds. If the goal is to compare across the beef cattle breeds, F_{ROH} has shown to be consistent for its estimates without getting influenced by number or the assembly of the individuals in the dataset. It also does not depend on allele frequencies, but parameter settings which is the same for all five breeds. If the goal is to monitor the rate of inbreeding within the breed of the same population, the correlations methods ($F_{VanRaden1}$, F_{hom} and F_{uni}) could be used. The results do tell that these are quite sensitive models, especially $F_{VanRaden1}$. It`s recommended to use F_{uni} if one of the options are between the three correlation methods. The F_{ped} is better than not considering it as it shows historical events and are easy to understand for producers.

The mean inbreeding in the Norwegian populations are currently lower than other countries`. Comparisons of mean inbreeding from 2006 and today shows no change over time using F_{ped}. The trendlines of mean inbreeding per year from 2015-2021 showed a low increase in annual inbreeding. Because the use of imported genetic material and embryos, this is not unexpected.

Limited available genomic data made it impossible to properly estimate inbreeding rates over generations. However, with the recent implementation of routine genotyping (since 2020), it will be interesting to observe how the rate of inbreeding develops over time in future generations. It is recommended to monitor this closely to manage inbreeding levels and maintain genetic diversity, which is crucial for the sustainability of a healthy and productive beef cattle population.

6.0 References

Alemu, S.W., Kadri, N.K., Harland, C., Faux, P., Charlier, C., Caballero, A. & Druet, T. (2020). An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. *Heredity*, 126(3): 410–423. doi: 10.1038/s41437-020-00383-9.

Altman, D.G. & Bland, J.M. (2005). Standard deviations and standard errors. *BMJ*, 331(7521): 903. doi: 10.1136/bmj.331.7521.903

Animalia, 2023. *Statistikk fra storfekjøttkontrollen*. Available at: https://app.powerbi.com/view?r=eyJrIjoiNzE5MjY5ZDYtNjM5Yy00ZjdlLTkzZGUtYjMwODk3Yjk0YzRhIiwidCI6ImE1MjU5NDllLTEzNjItNGE3My1hYzFkLWU2MjM3NWNmZWViZiIsImMiOjh9&pageName=ReportSection (accessed: 05.05.2023).

Baes, C.F., Makanjuola, B.O., Miglior, F., Marras, G., Howard, J.T., Fleming, A. & Maltecca, C. (2019). Symposium review: The genomic architecture of inbreeding: How homozygosity affects health and performance. *Journal of Dairy Science*, 102(3): 2807–2817. doi: 10.3168/jds.2018-15520.

Bjelland, D.W., Weigel, K.A., Coburn, A.D. & Wilson, R.J. (2015). Using a family-based structure to detect the effects of genomic inbreeding on embryo viability in Holstein cattle. *Journal of Dairy Science*, 98(7), 4934-4933. doi: 10.3168/jds.2014-9014.

Caballero, A., Almudena Fernández, Villanueva, B. & Toro, M.A. (2022). A comparison of marker-based estimators of inbreeding and inbreeding depression. *Genetics Selection Evolution*, 54(1). doi: 10.1186/s12711-022-00772-0.

Carolino, N. & Gama, L.T. (2008). Inbreeding depression on beef cattle traits: estimates, linearity of effects and heterogeneity among sire-families. *Genetics Selection Evolution*, 40(5): 511-27. doi: 10.1186/1297-9686-40-5-511.

Cassell, B.G., Adamec, V. & Pearson, R.E. (2003). Effect of Incomplete Pedigrees on Estimates of Inbreeding and Inbreeding Depression for Days to First Service and Summit Milk Yield in Holsteins and Jerseys. *Journal of Dairy Science*, 86(9): 2967–2976. doi: 10.3168/jds.s0022-0302(03)73894-6.

Charlesworth, D. & Willis, J.H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, 10(11): 783–796. doi:10.1038/nrg2664.

Colleau, J. (2002). An indirect approach to the extensive calculation of relationship coefficients. *Genetics Selection Evolution*, 34(4): 409–21. doi: 10.1186/1297-9686-34-4-409.

Cortellari, M., Negro, A., Bionda, A., Grande, S., Cesarani, A., Carta, A., Macciotta, N., Stefano Biffani & Crepaldi, P. (2022). Using Pedigree and Genomic Data toward Better Management of Inbreeding in Italian Dairy Sheep and Goat Breeds. *Animals*, 12(20): 2828–2828. doi: 10.3390/ani12202828.

Darwin, C. (1876). The effects of cross and self-fertilization in the vegetable kingdom. London: John Murray.

Decker, J.E., McKay, S.D., Rolf, M.M., Kim, J., Molina Alcalá, A., Sonstegard, T.S., Hanotte, O., Götherström, A., Seabury, C.M., Praharani, L., Babar, M.E., Correia de Almeida Regitano, Yildiz, L., Heaton, M.A., Liu, M.P., Lei, W.-S., Reecy, C.-Z., Saif-Ur-Rehman, J.M., Schnabel, M.R.D., & Taylor, J.F. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLoS Genetics*, 10(3):1004254. doi: 10.1371/journal.pgen.1004254.

Doekes, H.P., Veerkamp, R.F., Bijma, P., de Jong, G., Hiemstra, S.J. & Windig, J.J. (2019). Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. *Genetics Selection Evolution*, 51(1). doi: 10.1186/s12711-019-0497-z.

Falconer DS (1981) Introduction to quantitative genetics, 2nd edn. Longman House, London

FAO, I. (2004). Guidelines for development of national farm animal genetics resources management plans. Measurement of domestic animal diversity (MoDAD): Recommended microsatellite markers. New microsatellite marker sets. Recommendations of Joint ISAG/FAO Standing Committee.

Fedota, O.M, Ruban, S. Yu, Lysenko N.G, Goraichuk, I.V, Kolisnuk, A.I & Tyzhnenko, T.V. (2016). Estimation of inbreeding depression by CAPN1, CAST, GH, GHR, CYP3A genes for quantitative and reproduction traits in seven Aberdeen-Angus lines. *Journal for veterinary medicin, biotechnology and biosafety*, 2(4): 2411-3174. doi: 574.113.1:636.223.1

Ferenčaković, M., Sölkner, J., & Curik, I. (2013). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution*, 45(1). doi: 10.1186/1297-9686-45-42.

Ghoreishifar, S.M., Moradi-Shahrbabak, H., Fallahi, M.H., Jalil Sarghale, A., Moradi-Shahrbabak, M., Abdollahi-Arpanahi, R. and Khansefid, M. (2020). Genomic measures of inbreeding coefficients and genome-wide scan for runs of homozygosity islands in Iranian river buffalo, Bubalus bubalis. *BMC Genetics*, 21(1). doi: 10.1186/s12863-020-0824-y.

Goudet, J., Kay, T. & Weir, B.S. (2018). How to estimate kinship. *Molecular Ecology*, 27(20): 4121–4135. doi: 10.1111/mec.14833.

Gurgul, A., Szmatoła, T., Topolski, P., Jasielczuk, I., Żukowski, K. & Bugno-Poniewierska, M. (2016). The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. *Journal of Applied Genetics*, 57(4): 527–530. doi: 10.1007/s13353-016-0337-6.

Gutiérrez-Reinoso, M. A., Aponte, P. M. & García-Herreros, M. (2022). A review of inbreeding depression in dairy cattle: current status, emerging control strategies, and future prospects. *Journal of Dairy Research*, 89(1): 3–12. doi: 10.1017/S0022029922000188.

Henderson, C.R. (1976). A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics*, 32(1): 69. doi: 10.2307/2529339.

Howard, J.T., Pryce, J.E., Baes, C. & Maltecca, C. (2017). Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. *Journal of Dairy Science,* 100(8): 6009–6024. doi:10.3168/jds.2017-12787.

Kasarda, R., Moravčíková, N., Vostrý, L., Krupová., Krupa, E., Lehocká, K., Olšanská, B., Trakovická, A., Nádaský, R., Polák, P., & Židek, R., Belej, Ľ., & Golian, J. (2020). Fine-scale analysis of six beef cattle breeds revealed patterns of their genomic diversity. *Italian Journal of Animal Science* (19): 1542-1557. doi: 10.1080/1828051X.2020.1852894.

Keller, M.C., Visscher, P.M. & Goddard, M.E. (2011). Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. *Genetics*, 189(1): 237–249. doi:1534/genetics.111.130922.

Kenny, D., Berry, D.P., Thierry Pabiou & Rafter, P. (2023). Variation in the proportion of the segregating genome shared between full-sibling cattle and sheep. *Genetics Selection Evolution,* 55(1): 27. doi: 10.1186/s12711-023-00802-5.

Kleiven, K. (2007). Populasjonsstudie av de fem største kjøttferasene i Norge. Master thesis. Ås: Norges miljø- og biovitenskapelige universitet. Available at: arkivet til NMBU (accessed: 01.03.2023)

Klug, W. S., Cummings, R. M., Spencer, A. C., & Palladino, M. A. (2016). *Essentials of Genetics. 9.edt.* Harlow: Pearson.

Lande, R. (1994) Risk of population extinction from fixation of new deleterious mutations. *Evolution,* 48(5): 1460-1469. doi: 10.1111/j.1558-5646.1994.tb02188.x

Leung, C., Guscelli, E., Chabot, D., Bourret, A., Calosi, P. & Parent, G.J. (2023). The lack of genetic variation underlying thermal transcriptomic plasticity suggests limited adaptability of the Northern shrimp, Pandalus borealis. *Frontiers in Ecology and Evolution*, 11. doi: 10.3389/fevo.2023.1125134.

Lozada-Soto, E.A., Maltecca, C., Lu, D., Miller, S., Cole, J.B. & Tiezzi, F. (2021). Trends in genetic diversity and the effect of inbreeding in American Angus cattle under genomic selection. *Genetics Selection Evolution*, 53(1). doi: 10.1186/s12711-021-00644-z.

Lynch, M. (1991).The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution.* 45(3):622–629.

Lynch, M., Conery, J., & Burger, R (1995). Mutation accumulation and the extinction of small population. *The American Naturalist.* 146(4):489–518. doi: 10.1086/285812

Malécot, G. (1948). Les mathématiques de l'hérédité. Paris: Masson et Cie;

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A.K., Farrington, S.M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S.H., Dunlop, M.G., Wright, A.F. &

Campbell, H. (2008). Runs of Homozygosity in European Populations. *The American Journal of Human Genetics*, 83(3): 359–372. doi: 10.1016/j.ajhg.2008.08.007.

Meyermans, R., Gorssen, W., Buys, N. & Janssens, S. (2020). How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*, 21(1). doi: 10.1186/s12864-020-6463-x.

Miglior, F., Burnside, E.B. & Dekkers, J.C.M. (1995). Nonadditive Genetic Effects and Inbreeding Depression for Somatic Cell Counts of Holstein Cattle. *Journal of Dairy Science*, 78(5): 1168–1173. doi: 10.3168/jds.s0022-0302(95)76734-0.

Pekkala, N., Knott, K.E., Kotiaho, J.S., Nissinen, K. and Puurtinen, M. (2014). The effect of inbreeding rate on fitness, inbreeding depression and heterosis over a range of inbreeding coefficients. *Evolutionary Applications*, 7(9): 1107–1119. doi: 10.1111/eva.12145.

Purcell, S., Neale, B.M., Todd-Brown, K., Thomas, L.L., Ferreira, M., Bender, D., Maller, J., Sklar, P., I.W, P., Daly, M.J. & Sham, P.C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, 81(3): 559–575. doi: 10.1086/519795.

R Core Team (2014). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Url: http://www.R-project.org/

Ritland, K. (1996). Estimators for pairwise relatedness and individual inbreeding coefficients. Genetical Research Cambridge, 67:175–185. doi: 10.1017/S0016672300033620

RStudio Team (2022). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Sargolzaei, M., H. Iwaisaki, & Colleau. J. J. (2006). "CFC: A tool for monitoring genetic diversity." Proceedings of the 8th world congress on genetics applied to livestock production. Belo Horizonte, Minas Gerais.

Schiavo, G., Luca Fontanesi, Bertolini, F., Tinarelli, S., Dall'Olio, S., Costa, L. & Gallo, M. (2020). Comparative evaluation of genomic inbreeding parameters in seven commercial and autochthonous pig breeds. *Animal*, 14(5): 910–920. doi: 10.1017/s175173111900332x.

Sticca, E.L., Belbin, G.M. & Gignoux, C.R. (2021). Current Developments in Detection of Identity-by-Descent Methods and Applications. *Frontiers in Genetics* (12). doi: 10.3389/fgene.2021.722602.

Szulkin, M., Bierne, N. & David, P. Heterozygosity-fitness correlations: a time for reappraisal. (2010). *Evolution*, 64(5):1202-17. doi: 10.1111/j.1558-5646.2010.00966.x.

TYR. (2023). *TYR - Fra avl til biff*. Available at: https://www.tyr.no/ (accessed: 01.05.23)

VanRaden, P.M. (1992). Accounting for Inbreeding and Crossbreeding in Genetic Evaluation of Large Populations. *Journal of Dairy Science*, 75(11): 3136–3144. doi: 10.3168/jds.s0022-0302(92)78077-1.

VanRaden, P.M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, 91(11): 4414–4423. doi: 10.3168/jds.2007-0980.

Villanueva, B., Fernández, A., Saura, M., Caballero, A., Fernández, J., Morales-González, E., Toro, M.A. & Pong-Wong, R. (2021). The value of genomic relationship matrices to estimate levels of inbreeding. *Genetics Selection Evolution,* 53(1). doi: 1186/s12711-021-00635-0.

Walling, C.A., Nussey, D.H., Morris, A., Clutton-Brock, T.H., Kruuk, L.E. & Pemberton, J.M. (2011). Inbreeding depression in red deer calves. *BMC Evolutionary Biology*, 11(1): 318. doi: 10.1186/1471-2148-11-318

Weigel, K.A. & Lin, S.W. (2000). Use of Computerized Mate Selection Programs to Control Inbreeding of Holstein and Jersey Cattle in the Next Generation. *Journal of Dairy Science*, 83(4): 822–828. doi: 10.3168/jds.s0022-0302(00)74945-9.

Winnberg, E. (2020). Inbreeding depression in the wild. *Journal of Animal Ecology,* 89(6): 1407-1420. doi: 10.1111/1365-2656.13261

Wright, S. (1922). Coefficients of inbreeding and relationship. *The American Naturalist* (56): 330–338. doi: 10.1086/279872

Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics*, 88(1): 76–82. doi: 10.1016/j.ajhg.2010.11.011.

Zhai, J.M., Guldbrandtsen, B., Lund, S. M., & Sahana, G. (2015). Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genetics*, 16(1). doi: 10.1186/s12863-015-0227-7.

Zhang, Q.S., Goudet, J. & Weir, B.S. (2021). Rank-invariant estimation of inbreeding coefficients. *Heredity*, 128(1):1–10. doi:10.1038/s41437-021-00471-4.

Attachment 1

Calculation for F_{ROH} parameters

Following parameters have been estimated to use in PLINKs for estimating F_{ROH}.

$$L = \frac{Log_e \frac{\alpha}{n_s n_i}}{Log_e(1 - het)} = \frac{Log_e \frac{0.05}{86138 \cdot 795}}{Log_e(1 - 0.31)} \approx 57$$

, ns = number of genotyped SNPs per individual, ni = number of genotyped individuals, α = significance level (percentage false positives) and het = mean heterozygosity across all SNPs. Equation is used for size of sliding window measures in SNPs and minimum number of SNPs required to call a ROH. This is an example from Hereford data.

$$T = floor\left(\frac{N_{out} + 1}{L}, 3\right) = floor\left(\frac{4 + 1}{57}, 3\right) = 0.026$$

, Nout = desired number of final outer SNPs on either side of the homozygous segments that should not be included in the final ROH and L = scanning window size.

The length of the autosomal genome was measured to be 2,485,159 kilo base pairs.

Different settings were experimented with to determine the optimal values for maximal gap and density. The objective was to set them as low as possible without deviating to far from 100%. Figures 6 and 7 illustrates this.

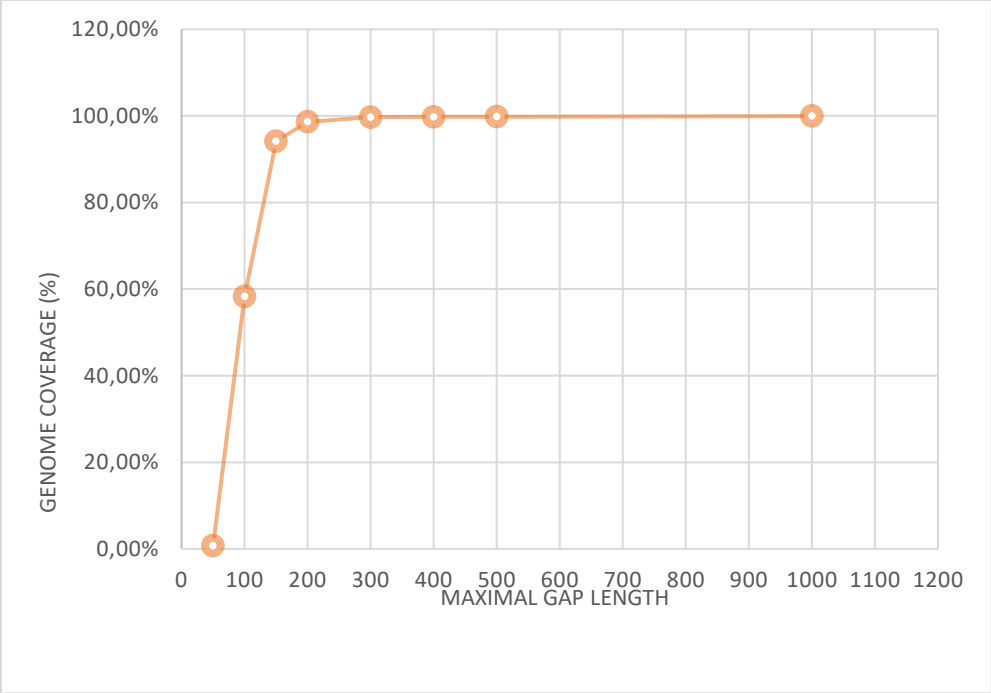

Figure 6: Output for experimenting with different parameter settings for maximal gap length.

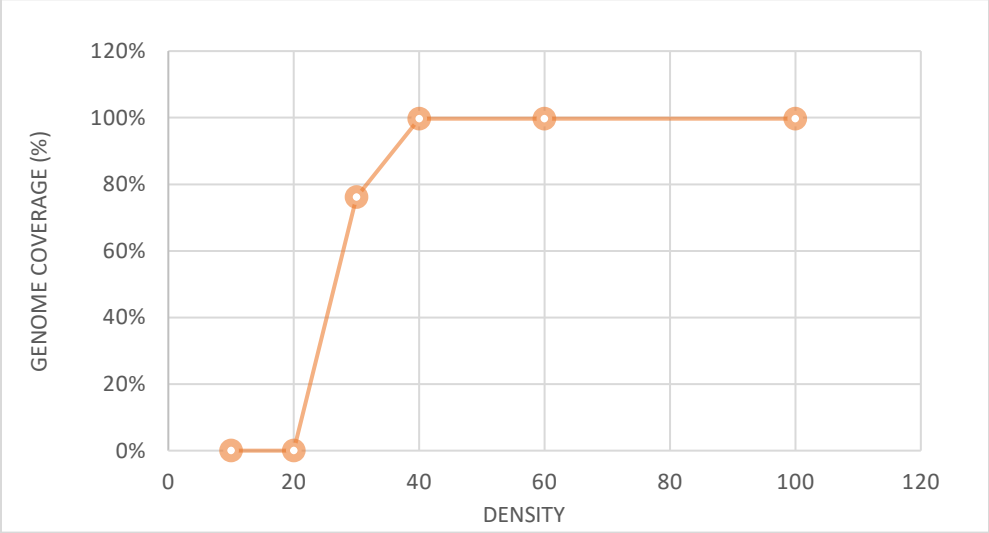

Figure 7: Output for experimenting with different parameter settings for density.

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway