Norwegian University
of Life Sciences

Master's Thesis 2020 60 ECTS
Faculty of Life Sciences

Trends in runs of homozygosity and inbreeding in Norwegian Red Cattle before and after implementation of genomic selection

Kirsti Winnberg
Animal breeding and genetics

Takk til

Nesten et år har gått siden jeg begynte å lese min aller første artikkel om genomisk innavl. Det har vært en lang og tøff prosess. 2020 har vært et svært spesielt år, og jeg har kjent ekstra på takknemligheten over dem jeg har rundt meg i denne tida. Uten deres støtte hadde jeg aldri fått til dette. De fortjener derfor en stor takk.

Først og fremst vil jeg takke Peer som på ulastelig vis har veiledet meg gjennom dette arbeidet. Han skapte et sårt tiltrengt rammeverk jeg kunne lene meg på. Jeg følte at han bestandig var tilgjengelig og jeg fikk alltid raskt svar på det jeg lurte på. Jeg tror virkelig ikke jeg kunne fått en bedre veileder å samarbeide med. Takk også til medveileder Arne i Geno som kunne hjelpe meg med alt av data-fiksfakseri og stilte opp midt i ferien for å ordne opp da jeg fikk servertrøbbel. Jeg vil også takke selvutnevnt mentor, Cathrine som jeg kunne prate med om absolutt alle utfordringer, (også de teite).

Gjennom 5 år har jeg tilhørt et enestående fagmiljø på IHA. Det er jeg svært takknemlig for. Jeg har ikke tall på hvor mange timer jeg har fått bruke på kontorene til professorer på huset. Der har jeg kunnet stille dumme spørsmål, smarte spørsmål eller bare snakke om løst og fast. Takk for at dere holder kontordørene deres åpen for oss studentene!

Til sist må jeg takke de næreste. Takk, Idun for at du dro meg gjennom korona. Takk, Anne for utallige pep-talks. Takk, Marius for at du stakk innom bare for å sjekke om jeg hadde det bra, og takk, Kristin for kameratskapet i skriveprosessen. Lunsjklubben 2018/2019, hysydyr-veterangruppa og alle lunsjere i vrimlerommet som jeg (bokstavelig talt) kunne kaste ball med fortjener også en stor takk!

Men mest av alt, takk til mamma som aldri er mer enn en telefonsamtale unna og alltid er lutter øre når jeg vil dele noe. Takk for at du har gitt meg selvtillit, ambisjoner og en uanstendig mengde stahet.

Ås, 27. juli 2019

Abstract

Background: Despite commercial breeding being part of the genomic era, routine use of genotype data to govern inbreeding is still scarce. Recent studies have found acceleration of inbreeding rates after implementation of GS. Development of robust and reliable genomic inbreeding metrics should therefore be a priority. Aim of this study was to optimize detection of runs of homozygosity (ROH) and use these along with inbreeding coefficients based on pedigree and genomic relationship matrix to examine trends in genome wide and region-specific inbreeding after implementation of GS in Norwegian Red Cattle (NR).

Methods: Pedigree data from whole population and genotype data from 80.999 animals (on 777K chip) was used to estimate inbreeding coefficients and rates of inbreeding. F_{PED}, F_{GRM} and F_{ROH} was used to assess inbreeding trends in NR before and after implementation of GS. ROH was also used to examine inbreeding on individual chromosomes. Detection of ROH in PLINK was optimized using genome coverage validation method.

Results: Parameter settings and density of data set was found to strongly influence ROH detection. No significant increase in rates of inbreeding was found after implementation of GS in NR, neither in genome wide nor chromosomal estimates. We detected an abundance of short ROH in the genome of NR, indicating little recent inbreeding. Rates of inbreeding were well within recommended 0.5-1% limits. High correlations between F_{ROH} and F_{GRM} indicate that these metrics can be used for routine inbreeding estimation in NR.

Conclusion: We lay the foundation for a framework that can be used to develop methodology for genomic inbreeding evaluation in NR. Calculations in this thesis only had 3,5 years of GS to base upon, and paucity of data strongly limits estimates. Estimates should be repeated when more time from implementation of GS has elapsed. Development of methods using genomic information to manage inbreeding in NR is advisable.

Background

Inbreeding is an increase in autozygous (identical by descent (IBD)) alleles due to mating of related individuals. This may give expression of detrimental recessive alleles, reduce the genetic variation in the population and give decline in selection response. Governing of inbreeding is important, and to achieve this, reliable and robust estimators are required. The inbreeding coefficient (F) is an extensively used statistic for this purpose. Traditionally, F has been based on expected proportions of autozygous alleles between relatives given by pedigree (F_{PED}). F_{PED} is however prone to serious flaws (Howard et al., 2017). Firstly, their reliability highly depends on depth and quality of pedigree records and secondly, they are unable to capture added genetic variation due to random process of mendelian sampling and recombination during meiosis (Keller et al., 2011). Hence, development of alternative methods to compute F is alluring. In the genomic era, F can be estimated using genotypes rather than pedigrees. Genomic inbreeding metrics are deemed more accurate than F_{PED} (e.g. Bjelland et al., 2013; Ferdosi et al., 2016) because they are based on realized rather than expected autozygosity, i.e. more adept at capturing true inbreeding. F_{PED} often underestimate F (Keller et al., 2011), and studies conclude that genomic estimators can enhance inbreeding management (e.g. Ferenčaković et al., 2013; Solé et al., 2017). Use of SNP data is prevalent. As with F_{PED}, increases in genomic F give decrease in production and reproductive ability in dairy cattle (Bjelland et al., 2013).

Genomic F can be estimated using segment-based methods looking at regions of consecutive SNPs or marker-by-marker methods considering single SNPs. Latter method includes genomic relationship matrix (GRM). This gives F as $(1 + F_{GRM})$ on its diagonal. F_{GRM} depends on allele frequency assumptions. These are unknown and challenging to assign. They can be estimated from sample or set to fixed value. Results show that using frequency of 0.5 can be beneficial. Simulation study by Forutan et al. (2018) found higher correlations between F_{GRM} and F_{TRUE} when using 0.5 compared to using known base allele frequencies. Also, VanRaden et al. (2011) and Bjelland et al. (2013) used frequencies of 0.5 and got higher correlations between F_{PED} and F_{GRM} than with base allele frequencies. A GRM constructed using 0.5 frequencies is basically an estimator of homozygosity that is adjusted to fit the distribution of the pedigree-based relationship matrix (A) (Bjelland et al., 2013). For segment-based methods, a common method is identifying homozygous regions called runs of homozygosity (ROH), and calculate F based on these (F_{ROH}). ROH have putatively arisen due to parental relatedness. They are presumed to be autozygous rather than allozygous because long homozygous stretches are unlikely to occur by chance. The term ROH was coined by Lencz et al. (2007), whose work validated antecedent presumptions of ROH reflecting autozygosity (Curik et al., 2014). Examination of ROH length, number of ROH and ROH distribution is useful as it gives indication of demographic history (Purfield et al., 2012), can help us detect selection sweeps (Hillestad, 2017) and is correlated with recombination rate and GC content

(Bosse et al., 2012). ROH also allows us to examine region-specific inbreeding, a useful feature as inbreeding in some regions of the genome is more detrimental than others (Howrigan et al., 2011).

Application of genomic information in estimating breeding values (EBVs) has gained precedence. Routine use of same data to manage inbreeding in actual breeding schemes is however still scarce (Howard et al., 2017). Most metrics are underdeveloped. This applies especially for ROH for which there exists little consensus regarding definition. Neither identification criteria nor characterization is uniform even within species (Peripolli et al., 2016). This makes comparison of results challenging, and accuracy of estimates might be weakened (Hillestad et al., 2017). The most common program for detecting ROH is probably PLINK (Chang et al., 2015). By specifying parameter settings in PLINK we set defining criteria for ROH. Few studies has however looked at influence of these on detection, and only recently a method was developed to validate choice of parameters (Meyermans et al., 2020).

Selection scheme influence trends in F. Since early 2000's genomic selection (GS) (Meuwissen et al., 2001) has predominated cattle breeding. GS gave reduction of generation intervals which could lead to higher annual inbreeding rates (ΔF). But because GS allows us to capture mendelian sampling, it was predicted that GS would give less co-selection of siblings and reduced ΔF (Daetwyler et al., 2007; Sonesson et al., 2012). Simulation studies showing reduced ΔF due to GS supported these expectations (e.g. Lillehammer et al., 2011; VanRaden et al., 2011). Despite this, recent studies have found the exact opposite effect of GS. Substantial increase in ΔF after GS implementation has been found in several Holstein populations (Doekes et al., 2018; Doublet et al., 2019; Forutan et al., 2018; Makanjuola et al., 2020) and in Jersey (Makanjuola et al., 2020).

In view of this, we aim to examine GS' influence on ΔF in Norwegian Red Cattle (NR). NR is a composite dual-purpose cattle breed based on imported and national genetics. Effective population size is 197 and census size ~210.000 (Geno, 2019). According to Geno (2018), a high priority in NR is maintenance of a broad breeding objective combining fertility and health with high productivity. Objective is ⅓ production, ⅓ functionality and ⅓ health and fertility traits (Nordbø et al., 2019). Because inbreeding effects fitness the most (Howard et al., 2017), improvement of management methods for ΔF might be especially important in NR. GS was introduced relatively late for NR, replacing traditional pedigree testing selection (PTS) in 2016 after a combination period (PTS/GS) using GS for preselection of young bulls. Comparison of these three selection schemes using stochastic simulation found that both GS and PTS/GS could increase genetic gain and reduce ΔF, and that GS could give higher genetic gain at same inbreeding rate as PTS/GS when selecting 20 sires (Lillehammer et al., 2011).

The aim of this thesis was to 1) determine optimal detection of ROH using dense marker data from NR population 2) assess effect of GS on inbreeding trends in NR using F_{PED}, F_{GRM} and F_{ROH}, and 3) use ROH to assess genetic diversity and inbreeding on a region-specific level.

Materials and method

Animals

All data used was provided by Geno. Data set with genotypes consisted of animals born from 1960-2019, and pedigree data with animals from 1990-2019. Due to paucity of some data, only animals from 1994-2019 were used for calculations in this thesis. Pedigree data constituted ~6.5 million individuals and genomic data consisted of 80.999 genotyped individuals. Figure 1 and 2 shows number of animals per year in pedigree and genomic data sets respectively.

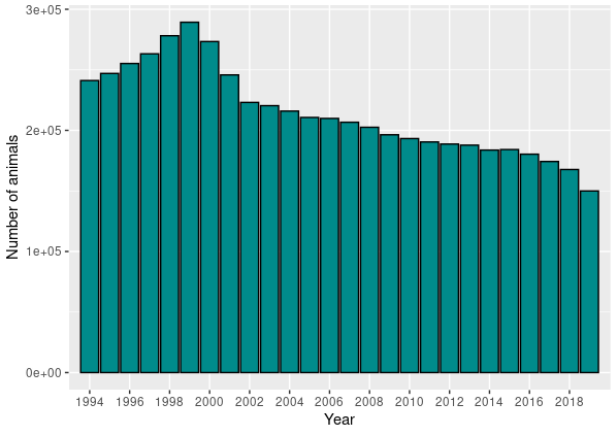


Figure 1: Number of animals by year of birth 1994-2019 in pedigree data set.

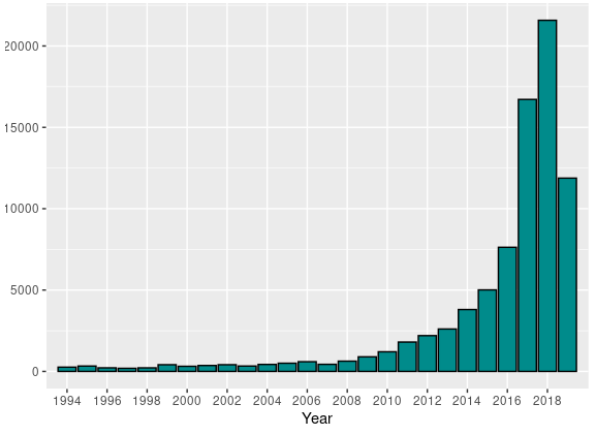


Figure 2: Number of genotyped animals 1994-2019 in genomic data set

Genotyping had been done using different platforms: an Affymetrix 25K chip, a custom made Affymetrix 54K chip (Affymetrix, Santa Clara), an Illumina 54K v.1, Illumina 54K v.2 and a high density (HD) Illumina 777K chip (Illumina, San Diego). All data had been imputed to a subset of the HD Illumina 777K chip using Fimpute (Sargolzaei et al., 2014). Prior to imputation. SNPs with minor allele frequency of less than 0.01, more than 10% missing, and/or deviations from Hardy-Weinberg Equilibrium with p-value < $1e^{-7}$ was discarded. Pruning for mendelian errors was conducted, and samples missing for more than 20% were removed. Details on imputation procedure are also given in Nordbø et al. (2019)

Inbreeding coefficients and inbreeding rates

Inbreeding coefficients were estimated using three different methods; pedigree-based method (F_{PED}), genomic method using GRM (F_{GRM}) and method based on ROH (F_{ROH}). Coefficients based on pedigree and genomic relationship matrix were computed by Geno. Pedigree-based estimates were calculated using relax2 (Strandén, 2014) based on all animals in population born from 1990-2019. F_{PED} estimation was done using VanRaden method (VanRaden, 1992). The genomic relationship matrix, GRM was computed using SNP data with allele frequencies set to 0.5 for all SNPs. Matrix was then scaled by multiplying a parameter to all matrix elements. This was done in order to make average diagonal elements equal to 1 (Nordbø et al., 2019). Inbreeding coefficients, F_{GRM} were derived from diagonal elements in matrix given by (1 + F_{GRM}). ROH was detected using PLINK and F_{ROH} was calculated used

method from Meyermans et al. (2020) (details given below). Strength of association between F_{PED}, F_{GRM} and F_{ROH} was assessed using Pearson correlations. In order to look at how changes in selection scheme from PTS through PTS/GS and finally, pure GS effected inbreeding, rates of inbreeding ΔF per year were calculated for all three metrics ΔF_{PED}, ΔF_{GRM} and ΔF_{ROH} Two methods of calculating ΔF was performed; one based on average F-values using formula given in formula *i.* and one based on multiple model regression given by formula *ii*. Inbreeding rate per year was given by:

i. $$\Delta \boldsymbol{F}_{\boldsymbol{year}} = \frac{\boldsymbol{F_t - F_{t-1}}}{\boldsymbol{1 - F_{t-1}}}$$

Where $\boldsymbol{F_t}$ is the average inbreeding coefficient for year $\boldsymbol{t}$. Regression was performed by fitting following multiple regression model to average inbreeding coefficient for each year:

ii. $$\boldsymbol{y_{ij} = \beta_i + \beta_1 \times x_j + \beta_{2i} \times x_j + \varepsilon_{ij}}$$

Where $\boldsymbol{y_{ij}}$ is the average inbreeding coefficient, $\boldsymbol{F}$, for year $\boldsymbol{j}$ and breeding scheme $\boldsymbol{i}$, $\boldsymbol{x_j}$ is the year, $\boldsymbol{\beta_i}$ is the intersect estimator for each breeding scheme, $\boldsymbol{\beta_1}$ is the general regression on year and $\boldsymbol{\beta_{2i}}$ the regression on year for breeding scheme $\boldsymbol{i}$ respectively. Inbreeding rates per generation was attained by multiplying generation interval, $\boldsymbol{L}$, with annual rate of inbreeding.

Runs of homozygosity detection

Detection of ROH was done using PLINK 1.9 (Chang et al., 2015). ROH detection in PLINK is greatly influenced by a set of pre-defined parameters (Howrigan et al., 2011; Meyermans et al., 2020). PLINK provides default parameter settings, but customization of these to own data set provides more robust and reliable analysis (e.g. Hillestad et al., 2017; Meyermans et al., 2020). Settings used in this thesis, as well as their function and PLINK command are listed in table 1. Figure 2 illustrates the detection process. Choice of settings are for the most part based on recommendations from Meyermans et al. (2020). These authors' validation method, genome coverage, was also used.

Table 1: Parameter settings chosen for ROH detection in PLINK 1.9 (Chang et al. 2015). Recommendations from Meyermans et al. (2020) has been followed in choice of parameters. Genome coverage validation method was used for SNP density and gap length setting.

DETECTION	PARAMETER	PLINK 1.9 COMMAND	VALUES
Defining sliding window	Sliding window size (number of SNPs)	-homozyg-window-snp	64
	Number of heterozygotes allowed within window	-homozyg-window-het	0
	Number of missing SNPs allowed within window	-homozyg-window-missing	1
Identifying ROH	Proportion completely homozygous windows	-homozyg-window-threshold	0.07
Check point for putative ROH	Largest interval between consecutive SNPs	-homozyg-gap	200
	Number of heterozygotes allowed in final segment	-homozyg-het	0
Minimum SNP density and ROH length	Minimum SNP density to call ROH	-homozyg-density	60
	Minimum length in kb. to call ROH	-homozyg-kb	500 kb
	Minimum number of SNPs to call ROH	-homozyg-snp	64

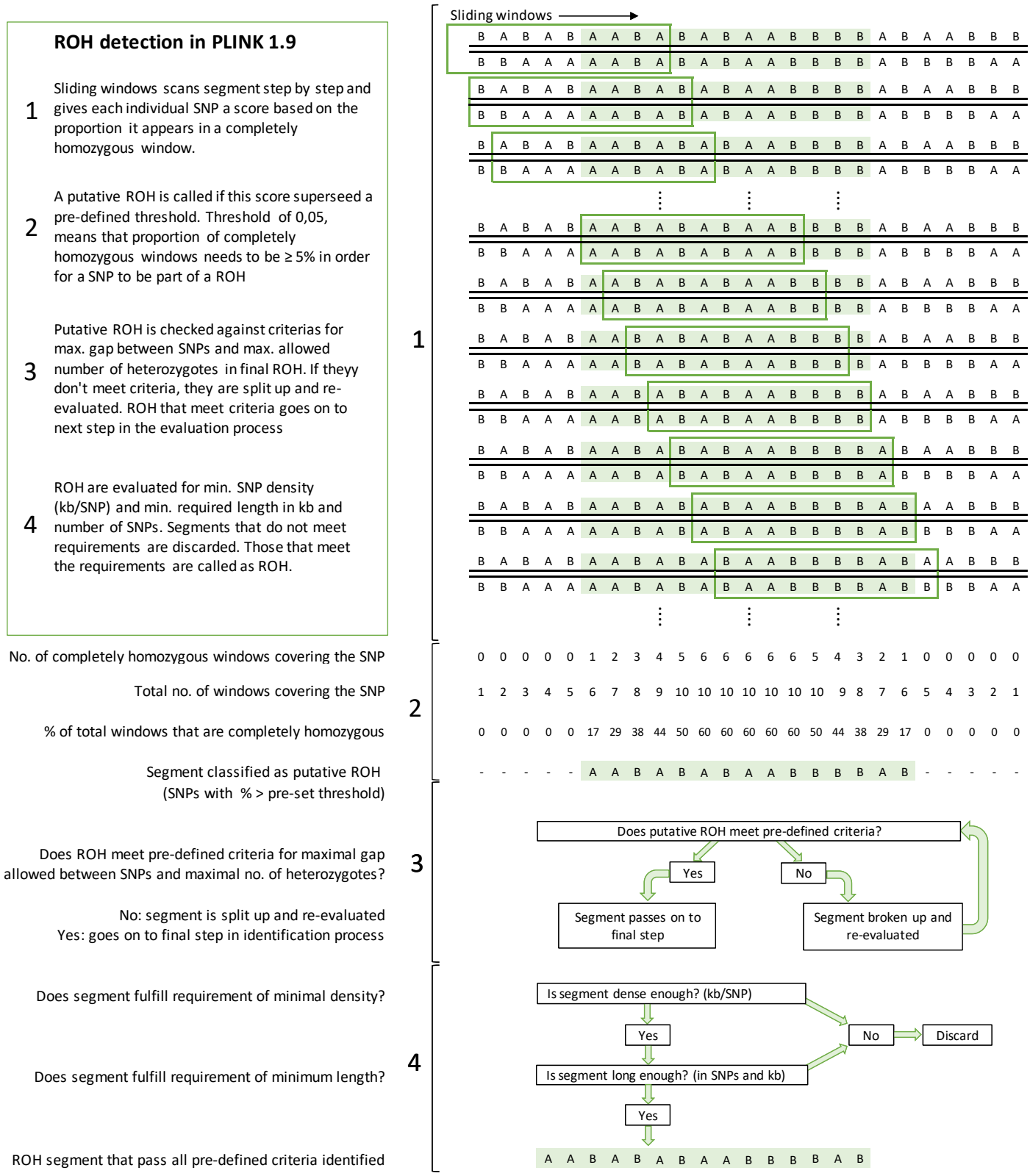


Figure 3: Step by step process for detection and classification of ROH. Figure inspired by Bjelland et al. (2013). PLINK 1.9 (Chang et al. 2015) is program used for detection.

Settings for length of sliding window and length of final ROH segment was calculated using formula from Lencz et al. (2007) and adapted by Purfield et al. (2012):

ii. $$\boldsymbol{L = \frac{log_e \frac{\alpha}{n_s n_i}}{log_e(1-het)}}$$

Where $\boldsymbol{L}$ is the length of sliding window/final ROH segment, $\boldsymbol{n_s}$ number of SNPs per individual (622.179), $\boldsymbol{n_i}$ the number of animals (80.999), $\boldsymbol{het}$ the mean heterozygosity across all SNPs (0.35) and $\boldsymbol{\alpha}$ the chosen significance level for type I errors. $\boldsymbol{L}$ was calculated to 64 SNPs. Plotting $\boldsymbol{L}$ against α showed that varying $\boldsymbol{\alpha}$ from 1%-5% only gave a change in L from 68-64 SNPs (data not shown), i.e. $\boldsymbol{L}$ was not deemed to be very sensitive of significance level, and a 5% level was chosen. With an average heterozygosity of 35% in population there's a 65% chance of a SNP to be homozygous. When we have 622.179 SNPs for 80.999 individuals, a minimum ROH length of 64 SNPs would be needed to produce <5% false positives across all subjects if we assume independence of all SNPs. PLINK provides a setting for defining length of final ROH segments in kb rather than number of SNPs, but Howard et al. (2017) found basing parameters on SNP outperformed detection done by using kb length.

PLINK evaluates whether every single SNP is part of a ROH. Number of windows that cover the SNP and are completely homozygous is evaluated, and if this number supersede a pre-defined threshold, SNP is called as part of a ROH. Threshold was calculated using formula in Meyermans et al. (2020):

iii. $$\boldsymbol{t = floor(\frac{N_{out}+1}{L}, 3)}$$

Where threshold is **t**, $\mathbf{N_{out}}$ is desired number of SNPs on outer sides of ROH segment that should not be included in ROH and L is same as in formula iii. $\mathbf{N_{out}}$ was set to 4 and t was calculated to 0.07.

In ROH detection it is possible to allow heterozygotes within ROH in order to account for genotyping errors. However results show that allowance of heterozygotes gives false positives (Hillestad et al., 2017), and poorer detection results (Howrigan et al., 2011). Parameter was therefore set to 0 both in sliding window and final segment. Allowance of missing SNPs is another genotyping error setting. Hillestad et al. (2017) found that allowing 3 vs. 1 missing only had a minor effect on detection. Detection in this thesis uses 1, which is quite common in the literature (e.g. Bjelland et al., 2013; Meyermans et al., 2020; Scraggs et al., 2014).

Genome coverage method (Meyermans et al., 2020) was used to validate settings for gap length and SNP density. Two completely homozygous individuals were simulated. One on 777K SNP array and one on a thinned down version of 54K SNPs. When the whole genome is completely homozygous, all the chromosomes will in practice be one long ROH, and the ROH detected using specific parameter settings in PLINK will be the maximum detectable ROH length.

Estimation of inbreeding coefficients based on ROH

Inbreeding coefficients based on ROH measure the proportion of the total genome covered by ROH. Most common estimation method is from McQuillan et al. (2008). But Meyermans et al. (2020) has adjusted this to account for variation of SNP coverage across the genome. Rather than using total length of autosome (L_{auto}) calculated as sum of distance between first and last SNP on all chromosomes, the size of autosome where ROH is detectable ($\boldsymbol{L_{auto\ cov}}$) is used in formula. $\boldsymbol{L_{auto\ cov}}$ was calculated by doing ROH detection with specified parameters on completely homozygote individual. Total ROH length found for this is the maximum detectable ROH length in any individual. Formula from McQuillan et al. (2008) and adapted by McQuillan et al. (2008) is given by:

iii. $$\boldsymbol{F_{ROH\ cov}} = \frac{\sum \boldsymbol{L_{ROH}}}{\boldsymbol{L_{auto\ cov}}}$$

Where $\boldsymbol{F_{ROH\ cov}}$ is the inbreeding coefficient, $\sum \boldsymbol{L_{ROH}}$ is the sum of all ROH detected and $\boldsymbol{L_{auto\ cov}}$ is length of autosome covered by SNPs. For 777K data set, $\boldsymbol{L_{auto\ cov}}$ equaled $2.49e^9$ while L_{auto}, was $2.51e^9$. For simplicity, $\boldsymbol{F_{ROH\ cov}}$ is referred to as F_{ROH} in this thesis.

ROH distribution and chromosomal inbreeding and genetic diversity

ROH was used to look at distribution of inbreeding across the genome. Changes in distribution and frequency in different ROH length classes before and after GS was examined. For comparisons sake, length classes was derived from previous studies (e.g. Ferenčaković et al., 2013; Forutan et al., 2018). They were <2 Mb, 2-4 Mb, 4-8 Mb, 8-16 and >16 Mb. On a chromosomal level, average length and average number of ROH as well as average F_{ROH} before and after GS was assessed. F per chromosome were estimated using the same method as was used for genome wide estimates. Formula was:

iv. $$\boldsymbol{F_{CHR\ k}} = \frac{\sum \boldsymbol{L_{ROH\ k}}}{\boldsymbol{L_{k\ cov}}}$$

Where $\boldsymbol{F_{CHR\ k}}$ is inbreeding coefficient for chromosome $\boldsymbol{k}$, $\sum \boldsymbol{L_{ROH\ k}}$ is the sum of all ROH on chromosome $\boldsymbol{k}$ and $\boldsymbol{L_{k\ cov}}$ is the length of chromosome $\boldsymbol{k}$ covered by SNPs (calculated same way as $L_{auto\ cov}$ from formula iii). Inbreeding rates per chromosome $\Delta F_{CHR\ k}$, were estimated with linear regression for each chromosome by fitting following multiple regression model to data:

v. $$\boldsymbol{y_{ij}} = \boldsymbol{\beta_{ik}} + \boldsymbol{\beta_{1k}} \times \boldsymbol{x_j} + \boldsymbol{\beta_{2ik}} \times \boldsymbol{x_j} + \boldsymbol{\varepsilon_{ijk}}$$

Where $\boldsymbol{y_{ij}}$ is the $\Delta F_{CHR\ k}$, for chromosome $\boldsymbol{k}$, year $\boldsymbol{j}$ and breeding scheme $\boldsymbol{i}$, $\boldsymbol{x_j}$ is the year, $\boldsymbol{\beta_{ik}}$ is the intersect estimator for each breeding scheme for chromosome $\boldsymbol{k}$, $\boldsymbol{\beta_{1k}}$ is the general regression on year for chromosome $\boldsymbol{k}$, and $\boldsymbol{\beta_{2ik}}$ the regression on year for breeding scheme $\boldsymbol{i}$ for chromosome $\boldsymbol{k}$.

Results

Runs of homozygosity parameter settings

Results obtained from varying PLINK parameter settings in ROH detection for 777K and 54K data sets are given in figure 4. For 777K data set, an increase in SNP density setting from 4 to 4.9 kb/SNP gave coverage increase from 55% to peak value of 99.7%. After this, no change in coverage was observed for increased densities, i.e. setting did not influence detection. For 54K data set on the other hand, the SNP density had a large effect on ROH detection. Genome coverage obtained was kept at 0% until 46 kb/SNP where a steep increase occurred, and a 61% coverage was obtained. Maximum coverage reached was 98.2% at density of 65 kb/SNP. This coverage was not exceeded even when testing with very large values (1000 kb/SNP, data not shown).

When it came to maximum gap length setting (figure 4 b), patterns are similar to those shown for SNP density. Both data sets display a steep increase in genome coverage. 777K data being the most pronounced with a coverage that rises from 5 to 85% when maximal gap length allowed is increased from 20 to 30 kb. Increase in coverage for 54K data is steepest around 200 kb, moving from 4.5 to 52% when increasing gap length from 195 to 200 kb, and declines in rate of increase after this, not reaching 77% until using a gap length of 250 kb, and peak coverage of 100% at 1200 kb (data not shown). A genome coverage of >99% was obtained at maximum gap set to 200 kb in 777K data set, while a length of 900 kb was required to exceed 99% in 54K data set. It is favorable to keep gap length as small as possible. Maximum gap length of 200 kb gave >99% and was chosen for detection in this thesis.

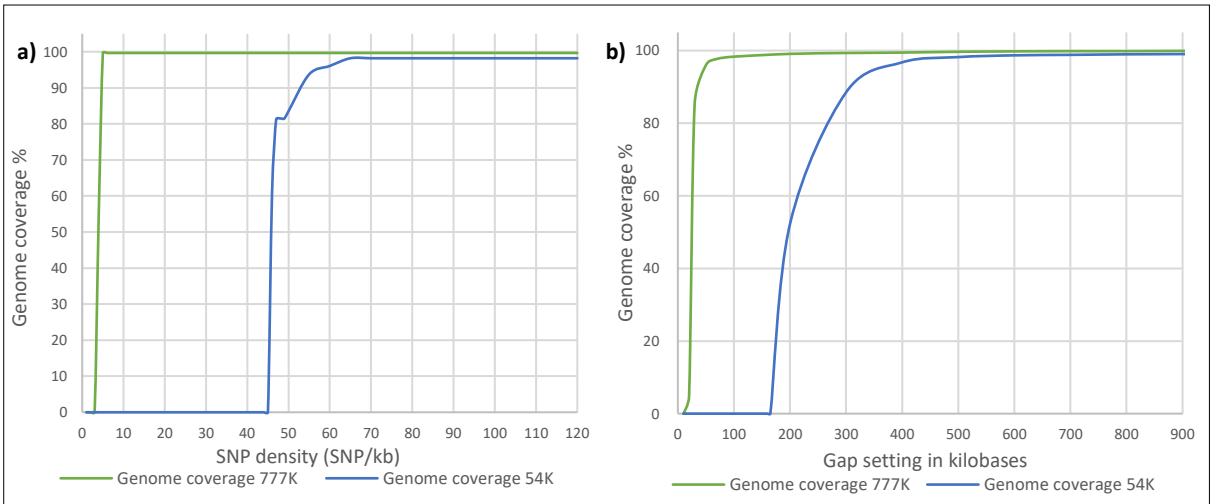


Figure 4: Genome coverage obtained by varying a) SNP density and b) maximum gap length in ROH detection using PLINK 1.9 (Chang et al. 2015) for two different density data sets (777K and 54K).

Inbreeding coefficients and rate of inbreeding

Inbreeding coefficients, F, were estimated using pedigree (F_{PED}), genomic relationship matrix (F_{GRM}) and runs of homozygosity (F_{ROH}). To examine trends in inbreeding before and after implementation of GS, rates of inbreeding per year and per generation were compared for the three different selection schemes (PTS, PTS/GS and GS). Rates per generation was calculated by multiplying annual rate with generation intervals which were 4.55, 4.58 and 3.83 years for PTS, PTS/GS and GS respectively.

Average F_{PED}, F_{GRM} and F_{ROH} per year from 1994-2019 are shown in figure 6. GRM-based estimates are centered around 0 because of scaling. All three metrics display an increase across the years, and the slopes of the linear trend lines (estimated increase for whole period) are very similar with 0.0004 for both F_{ROH} and F_{PED} and 0.0003 for F_{GRM}. Estimates for F_{ROH} display the largest fluctuations with a R^2 value of 0.592 compared to 0.878 and 0.968 for F_{GRM} and F_{PED} respectively.

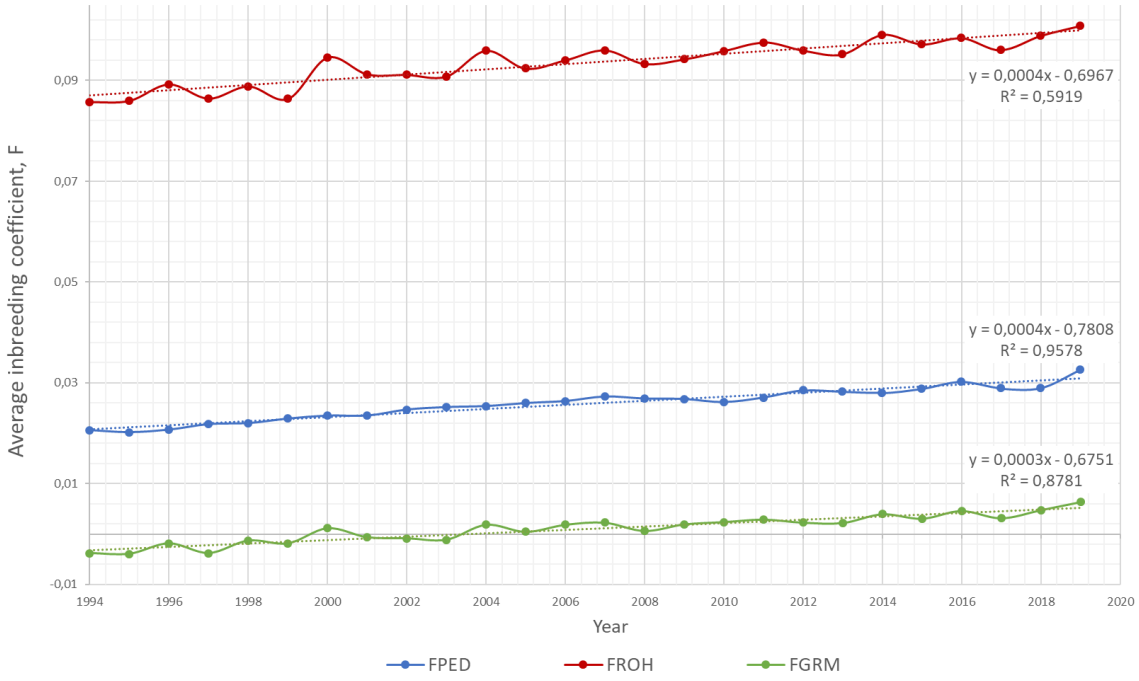


Figure 6: Average estimated inbreeding coefficients per year 1994-2019. F based on pedigree (F_{PED}), genomic relationship matrix (F_{GRM}) and runs of homozygosity (F_{ROH}).

Average inbreeding coefficients (F_{PED}, F_{ROH}, and F_{GRM}) and rate of inbreeding, ΔF, per year and generation for the three different selection schemes are presented in table 2. All three inbreeding metrics showed a continual increase in average F from progeny testing scheme (PTS) through combination PTS/GS and to genomic selection scheme (GS). Increase is moderate in F_{PED} and F_{ROH} and more prominent for F_{GRM}. When it came to inbreeding rates, ΔF_{PED} was the only metric that showed continual increase from PTS through PTS/GS and onto GS with rates of 0.038, 0.044 and 0.098% for the three schemes respectively. Increase in rate is small from PTS to PTS/GS period, but more than doubled during GS period when it rises from 0.044 to 0.098% per year.

Both ΔF_{ROH} and ΔF_{GRM} displayed noticeable increases when looking at PTS versus GS period isolated. The former increased from 0.060 to 0.102% and the latter from 0.035 to 0.083% per year. However, both estimates also exhibit a considerable drop during the combination PTS/GS scheme. ΔF_{ROH} from 0.060 to -0.010% and ΔF_{GRM} from 0.035 to 0.005% per year. Inbreeding rates per generation displays the same general trend as rates per year, but increases are lower, with e.g. F_{PED} not doubling from PTS/GS to GS period. Average generation interval across the three periods change, with a slight increase from PTS to PTS/GS period from 4.55 to 4.58 and then decreasing during GS to 3.83.

Large standard errors are observed for both inbreeding rates per year and per generation. Standard errors for ΔF_{ROH} are the highest and for both ΔF_{ROH} and ΔF_{GRM} they surpass estimated rates for all periods. ΔF_{PED} standard errors are smallest, but still large. Generally, the standard errors increased throughout the three periods being lower for PTS than for PTS/GS and GS periods.

Table 2: Estimated inbreeding coefficients and rates of inbreeding. PTS = progeny testing scheme, GS = genomic selection, PTS/GS = combination selection scheme, F = inbreeding coefficients, PED = pedigree, ROH = runs of homozygosity, GRM = genomic relationship matrix, % ΔF = percentage inbreeding rates, std.error = standard error.

		F		% ΔF per year		% ΔF per generation	
		Average	**Std.error**	**Average**	**Std.error**	**Average**	**Std.error**
F_{PED}	PTS	0.024	0.001	0.038	0.013	0.174	0.058
	PTS/GS	0.028	0.000	0.044	0.042	0.199	0.194
	GS	0.030	0.001	0.098	0.107	0.348	0.388
F_{ROH}	PTS	0.092	0.001	0.060	0.081	0.283	0.373
	PTS/GS	0.097	0.001	-0.010	0.148	-0.040	0.677
	GS	0.098	0.001	0.102	0.128	0.335	0.481
F_{GRM}	PTS	0.000	0.001	0.035	0.037	0.164	0.171
	PTS/GS	0.003	0.000	0.005	0.059	0.024	0.270
	GS	0.005	0.001	0.083	0.076	0.299	0.299

Looking at standard errors, we see that none of the inbreeding rates can be said to be significantly different from 0. Estimation of significance in changes of rates can therefore not be done accurately. Because of this, a multiple regression model was used to try and examine changes more closely.

Percentage estimated inbreeding rates per year and per generation from regression are presented in table 3. As seen, ΔF_{ROH} is the only statistic exhibiting continual increase from PTS through PTS/GS and onto GS period, with rates of 0.035, 0.040 and 0.068% per year and 0.157, 0.183 and 0.259% per generation. ΔF_{ROH} estimates are also the highest for all periods except for during PTS for which ΔF_{PED} gives the largest estimate. Standard errors for ΔF_{ROH} are the smallest of all metrics and for all periods. ΔF_{PED} shows a drop from PTS to PTS/GS period with 0.042 to -0.035% per year and from 0.193 to -0.162% per generation respectively. During GS period, ΔF_{PED} increases to 0.031% per year and 0.117% per generation, but these rates are still lower than during PTS period, and the lowest for GS period

compared to other two metrics. For ΔF_{GRM}, the same pattern can be discerned in table 3 as that in table 2. The rate of increase is higher during GS than PTS period, but (as for ΔF_{PED}) rate displays a considerable drop during middle (PTS/GS scheme). Rates are 0.029, 0.013 and 0.041% per year and 0.13, 0.06 and 0.16% per generation for PTS, PTS/GS and GS respectively. ΔF_{GRM} has the largest standard errors of the three metrics, but standard errors for PTS/GS and GS periods are also considerable for both ΔF_{PED} and ΔF_{GRM} estimates. PTS period exhibits the smallest standard errors for all metrics.

Table 3: Estimated rate of inbreeding from regression. PTS = progeny testing scheme, GS = genomic selection, PTS/GS = combination selection scheme, % ΔF = percentage estimated rate of inbreeding, PED = pedigree, ROH = runs of homozygosity, GRM = genomic relationship matrix.

		% ΔF per year		% ΔF per generation	
		Estimate	**Std.error**	**Estimate**	**Std.error**
ΔF_{PED}	PTS	0.042	0.003	0.193	0.012
	PTS/GS	-0.035	0.034	-0.162	0.157
	GS	0.031	0.034	0.117	0.132
ΔF_{ROH}	PTS	0.035	0.000	0.157	0.002
	PTS/GS	0.040	0.012	0.183	0.053
	GS	0.068	0.012	0.259	0.044
ΔF_{GRM}	PTS	0.028	0.003	0.129	0.012
	PTS/GS	0.013	0.060	0.059	0.273
	GS	0.041	0.060	0.156	0.228

Using ANOVA to compare ΔF before and after implementation of GS gave p-values of 0.408, 0.794 and 0.774 for ΔF_{PED}, ΔF_{ROH} and ΔF_{GRM} respectively. This means that rates of inbreeding were not significantly different for the three selection schemes. High correlations were found between F_{ROH} and F_{GRM} while correlations between both genomic estimation methods and pedigree method were moderate. Correlations were 0.966 for F_{ROH} and F_{GRM}, 0.612 for F_{ROH} and F_{PED} and 0.683 for F_{PED} and F_{GRM}. Correlations were observed to be similar both across and within the three selection periods.

ROH distribution and chromosomal inbreeding

ROH was used to look at inbreeding differences across the genome and for individual chromosomes. A small increase in average number of ROH per animal was seen during GS with 137.3 compared to 131.1 for PTS and 127.6 for PTS/GS (table 4). Figure 7 clearly shows that the majority of ROH found were small (>2 Mb), constituting 73.8-74.8% of all ROH. Change in frequencies of length classes is negligible across the three periods. Relative frequencies of ROH in the different length classes ranged around 74% (<2 Mb), 14% (2-4 Mb), 7.5 (4-8 Mb), 2.4% (8-16 Mb) and 1.2% (>16 Mb).

Table 4: Percentage distribution of ROH in different length classes <2, 2-4, 4-8, 8-16 and >16 Mb. nROH = average number of ROH per animal, n = number of ROH, % = percentage distribution

	PTS		PTS/GS		GS	
	%	n	%	n	%	n
nROH	131.1		127.6		137.3	
<2	74.8	98.0	73.8	94.2	74.3	102.0
2-4	14.2	18.6	14.1	18.0	14.8	20.3
4-8	7.3	9.6	7.7	9.9	7.4	10.1
8-16	2.6	3.4	2.2	3.9	2.5	3.4
>16	1.2	1.5	1.3	1.7	1.1	1.5

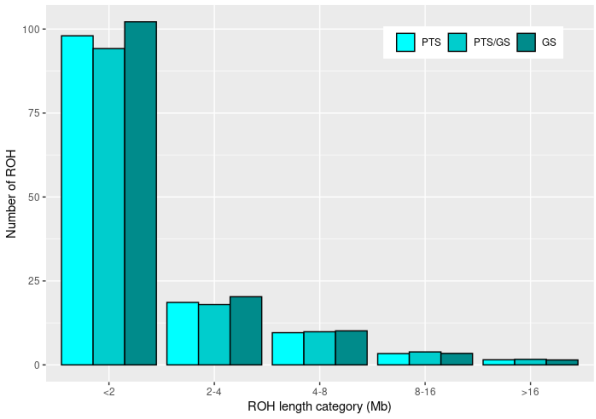


Figure 7: Distribution of ROH in different length classes.

Except for BTA 12, 13 and 14, all chromosomes show a higher number of ROH after GS than during PTS (figure 8). However, only BTA 11, 17, 22 and 27 display continual increase from PTS through PTS/GS and onto GS. All other chromosomes display a drop in number during PTS/GS scheme. Average number of ROH across all chromosomes was 4.63, 4.49 and 4.84 for PTS, PTS/GS and GS respectively.

Figure 9 shows average total length of all ROH on each chromosome for the three periods. All chromosomes except BTA 6, 12, 13, 18 and 23 have longer average ROH for GS than PTS. Only 12 of 29 chromosomes show continual increase. I.e. drop in average length during PTS/GS applies for over half of the chromosomes. Both average number of ROH and average length displays a pattern that coincide with total length of chromosomes; chromosome 1 displaying the longest average length and highest average number, but both measures generally decreasing for chromosomes of higher number (that are also shorter). Notable exceptions are number of ROH on BTA 5 and 14 in figure 8 and average length of ROH at BTA 4, 5, 14 and 20, representing peaks in figure 9. Total ROH length across genome was 8308.5 for PTS, 8607.4 for PTS/GS and 8705.6 for GS scheme.

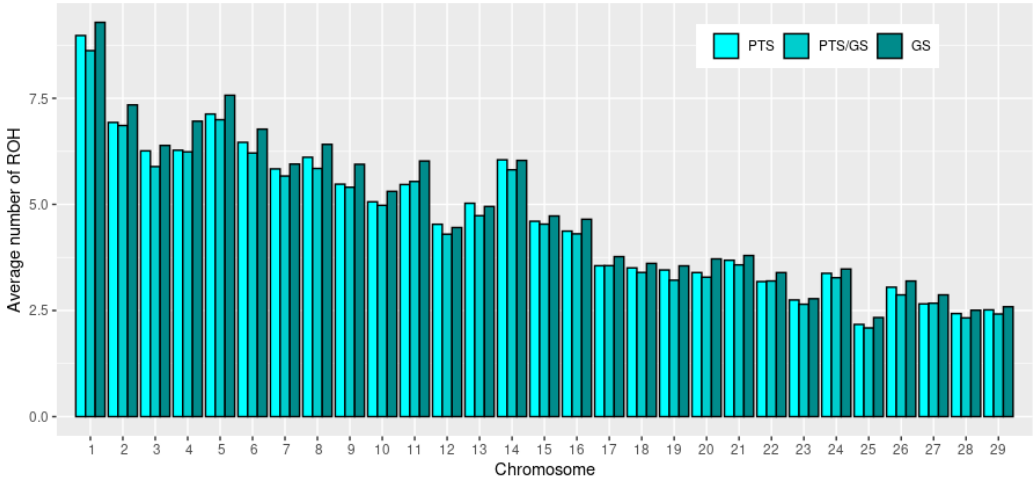


Figure 8: Average number of ROH per chromosome for three selection periods

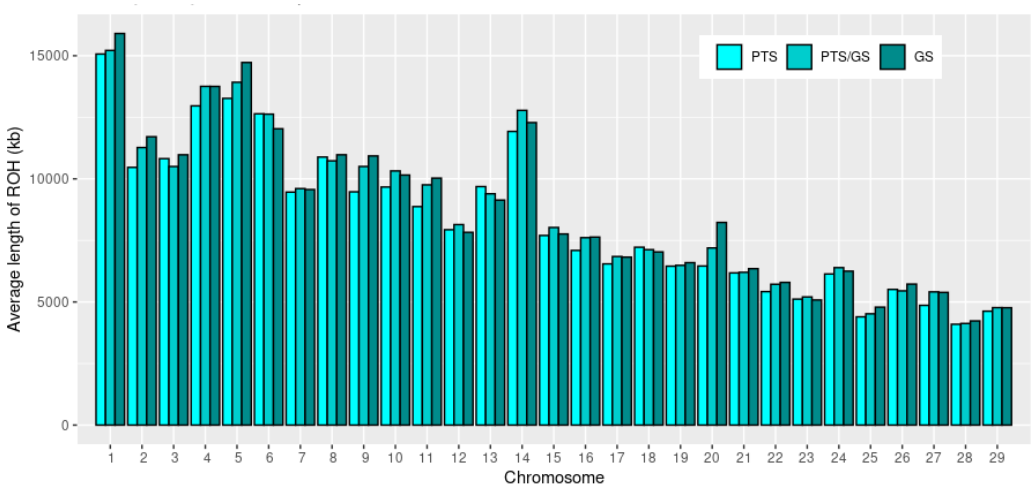


Figure 9: Average length of ROH per chromosome for the three selection periods

Variation in F_{ROH} for different chromosomes is given in figure 10. Relatively large variations in F_{ROH} can be seen, ranging from 0.077 on BTA 2 during PTS to 0.155 on BTA 14 for PTS/GS. BTA 14 stands out, clearly having the largest F_{ROH} for all three periods (0.145, 0.155 and 0.149 for PTS, PTS/GS and GS respectively). BTA 4, 5, 20 and 27 also distinguishes themselves as peaks in the figure, all these have a $F_{ROH} \geq 0.114$. In contrast to figure 8 and 9, F_{ROH} does not follow pattern of general chromosome length. Lowest F_{ROH} is observed for BTA 2 with 0.077, 0.082 and 0.086 for PTS, PT/GS and GS periods respectively. BTA 7 and 21 also display low F_{ROH}, (<0.09 for GS period).

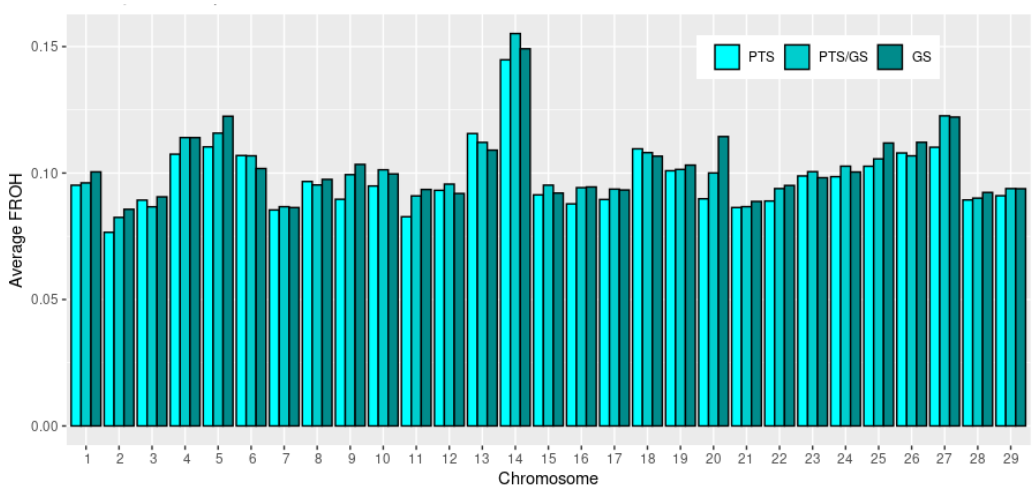


Figure 10: Average F_{ROH} per chromosome for the three selection periods

Estimated percentage ΔF per year for all chromosomes during the three selection schemes are given in table 5. Higher ΔF can be seen when comparing GS to PTS for 18 of the chromosomes. However, using ANOVA to compare ΔF showed that increase was not significant for any chromosome. Only 7 chromosomes display continual increase in ΔF from PTS, through PTS/GS and onto GS. Both the highest and lowest ΔF was observed during PTS/GS period (ΔF = 0.396% for BTA 27, and -0.546% for BTA 1). PTS/GS generally displayed the largest ΔF. BTA 13, 17, 27, 28 and 29 had ΔF ranging from 0.295 to 0.396%. During GS period the highest observed F was found on BTA 1, 5, 9, 10, 20 and 26 (from 0.219 to 0.338%). GS and PTS period display fewer extreme values than PTS/GS period. This is reflected in

average standard errors which are 0.297 for PTS and 0.296 for GS, but noticeably higher for PTS/GS (0.418). Standard errors are however large for all periods and all chromosomes, superseding actual estimates for all but four cases (these being negligibly larger). BTA 14 distinguishes itself during PTS period by having a ΔF of 0.254%, being by far the largest for this period, but also in comparison to estimates for PTS/GS. ΔF for BTA14 decreases and reaches a negative value during GS period (-0.079%).

Table 5: Estimated rate of inbreeding per year on individual chromosomes (BTA) for three selection schemes. ΔF_{ROH} = rate of inbreeding based on runs of homozygosity. PTS = progeny testing scheme. PTS/GS = combination scheme. GS = genomic selection scheme, std.error = standard error of estimates. avg. = averages

	PTS		PTS/GS		GS	
BTA	**% ΔF_{ROH}**	**std.error**	**% ΔF_{ROH}**	**std.error**	**% ΔF_{ROH}**	**std.error**
1	-0.245	0.305	-0.546	0.430	0.308	0.304
2	-0.027	0.153	0.135	0.215	0.087	0.152
3	-0.020	0.243	0.108	0.342	0.067	0.242
4	0.018	0.196	0.143	0.276	0.099	0.195
5	-0.148	0.299	-0.024	0.421	0.221	0.297
6	0.122	0.269	0.172	0.378	-0.037	0.267
7	0.004	0.303	-0.213	0.427	0.080	0.302
8	-0.049	0.335	-0.254	0.472	0.104	0.334
9	-0.207	0.282	-0.043	0.396	0.253	0.280
10	-0.149	0.275	-0.008	0.386	0.246	0.273
11	-0.157	0.147	-0.058	0.207	0.189	0.146
12	0.051	0.255	-0.091	0.359	-0.006	0.254
13	0.068	0.291	0.295	0.410	-0.088	0.290
14	0.254	0.452	0.156	0.636	-0.079	0.450
15	0.057	0.237	0.033	0.334	0.006	0.236
16	-0.091	0.260	-0.330	0.365	0.160	0.258
17	0.122	0.237	0.342	0.334	-0.103	0.236
18	0.123	0.367	-0.278	0.516	-0.033	0.365
19	-0.081	0.344	-0.231	0.484	0.142	0.342
20	-0.150	0.322	-0.102	0.453	0.338	0.320
21	-0.023	0.232	0.038	0.326	0.022	0.231
22	0.128	0.265	0.084	0.373	-0.098	0.263
23	-0.267	0.280	-0.154	0.394	0.198	0.279
24	0.026	0.365	-0.139	0.513	0.129	0.363
25	-0.133	0.327	0.167	0.460	0.196	0.325
26	-0.232	0.382	-0.361	0.538	0.219	0.380
27	0.128	0.426	0.396	0.599	-0.078	0.424
28	-0.025	0.404	0.310	0.569	-0.030	0.402
29	0.152	0.365	0.274	0.514	-0.184	0.364

Discussion

Using genomic and pedigree inbreeding measures. trends in inbreeding before and after genomic selection in Norwegian Red Cattle was examined. First aim was to determine optimal method of detecting runs of homozygosity in PLINK. Secondly, F_{PED}. F_{GRM} and F_{ROH} was used to examine inbreeding trends and thirdly, patterns in ROH was used to assess inbreeding rates and genetic diversity on a chromosomal level before and after GS implementation.

Optimizing runs of homozygosity detection

ROH-based inbreeding metrics are increasingly popular. Lack of consensus regarding definition is however an issue that makes it challenging to compare results across studies (Hillestad et al., 2017). Also. few efforts have been put towards development of robust detection methods. resulting in questionable reliability of estimates. By using genome coverage method proposed by Meyermans et al. (2020) we were able to evaluate effect of minimum SNP density and maximum gap length settings on ROH detection using PLINK (Chang et al., 2015) for high density (HD) and medium density data.

SNP density parameter (kb/SNP) lets us exclude autosomal regions with low SNP coverage and avoid false positives (type I error). Our results show that SNP density is rendered redundant in HD data. This is no surprise. Because, as argued by Hillestad et al. (2017), SNPs in a 777K array will on average be positioned >5 kb apart. Thus, density criterion does not take effect unless using <5 kb/SNP. For our data, using 4 kb/SNP gave genome coverage of 55.04% while 4.9 kb/SNP resulted in maximum coverage (99.7%). For 54K data, density criterion strongly influenced ROH detection. Maximum genome coverage was obtained at 65 kb/SNP. This coincide with results in Meyermans et al. (2020) who got maximum genome coverage at 60-70 kb/SNP. Default setting in PLINK is 50 kb/SNP. Using 50 kb/SNP in our data, genome coverage dropped to 83.5%, and Meyermans et al. (2020) showed that using 50 kb/SNP in Australian Polled Merino Sheep data, coverage of only 0.6% was obtained. Their results, and ours, suggest that default PLINK setting might weaken reliability of ROH detection in medium density data. Meyermans et al. (2020) also found parameter to be population dependent. Hence, it is advisable that optimal setting is determined for individual populations and array density.

As with SNP density, gap length reflects our expectation regarding true homozygosity status of nucleotides positioned between SNP markers. Results in this thesis show that gap length setting should be adjusted to density of data set. While maximum gap length of 200 kb gave genome coverage >99% in HD data. 54K data needed 900 kb for coverage to exceed 99%. PLINK default setting of 500 kb resulted in coverage of 98% and using a setting of 200 kb (as in HD data) led coverage to drop to 52% in 54K data. To the best of our knowledge, no previous study has examined effect of gap length on ROH detection in PLINK. Meyermans et al. (2020) was unable due to lack of HD data. In the literature.

values ranging from 100-1000 is used. and few justify their choice (Howrigan et al., 2011; Meyermans et al., 2020). Hillestad et al. (2017) argued that gap length setting was redundant in ROH detection in HD, but our results show that using lengths <200 kb would give slight coverage reduction (e.g. 98% for 100 kb). As with density setting. optimization of gap length setting to individual data set is preferable.

Not surprisingly, we found that density of array strongly influences robustness and reliability of ROH detection. When using HD array, assumption regarding true homozygosity status is more likely to hold true, avoiding type I errors. Our results are in accordance with Hillestad et al. (2017) who also analyzed ROH in NR. Their study concluded that higher density data set contributed to increased accuracies in ROH detection by discarding false positives, detecting shorter ROH and resulting in redistribution of long ROH to shorter ROH. Data set used in our study was not pruned for LD. Although Howrigan et al. (2011) recommended this based on simulation results. Meyermans et al. (2020) found that no pruning was preferable. Our data was however pruned for MAF (0.01) prior to imputation. Meyermans et al. (2020) advice against this in medium density data as it can lead to ignoring long homozygous regions and Hillestad et al. (2017) found the same to be true for HD data as it led to detection of fewer ROH (especially short ones) and that low MAF ROH can signalize selection signatures and trace selection. Optimally, analysis should be repeated using unpruned data.

Our results suggest that using HD data is preferred for ROH detection. As argued by Bosse et al. (2012), reduced number of markers makes it challenging to discover variation in inbreeding across the genome. These authors found that use of 60K sufficed for detection of ROH > 5 Mb, but that short ROH were challenging to find and medium density arrays underestimated cumulative ROH size. This is supported by Purfield et al. (2012) who found that 50K was enough for detection of almost all ROH > 5 Mb, but that detection accuracy was strongly reduced for 0.5-1 Mb ROH. Challenges was especially pronounced for populations with many short ROH. Although Zhang et al. (2015a) found that ROH detection using 50K data gave similar results as when using sequence data, it is important to consider differences in populations and distribution of ROH. As Marras et al. (2015) points out, use of medium density arrays may provide good estimates in populations with recent inbreeding and high linkage disequilibrium (LD), but precise detection of autozygosity in populations with more ancient inbreeding and low LD will require higher density data. Considering that NR is a population with many short ROH (figure 7) and more ancient inbreeding, use of HD arrays might be advisable when detecting ROH.

Genome wide inbreeding trends before and after genomic selection

Genomic selection has had a major impact on animal breeding. Despite this. investigation into GS' effect on inbreeding trends in real populations are still scarce (Doekes et al., 2018; Doublet et al., 2019). In a simulation, Forutan et al. (2018) found that GS gave decrease of ΔF, but using real data

from North American Holstein, the opposite was seen; ΔF increased at a faster pace after implementation of GS.

This study assesses inbreeding trends after implementation of GS in Norwegian Red Cattle. Estimates for F_{PED}. F_{GRM} and F_{ROH} were used to compare ΔF before and after GS. Results show that increase in ΔF were not significant. This does not coincide with other studies. Increase in ΔF after GS has been found in several Holstein breeds (Doekes et al., 2018; Forutan et al., 2018; Makanjuola et al., 2020) and in Jersey (Makanjuola et al., 2020). Our results are however in accordance with Doublet et al. (2019) who found significant increases in French Holstein, but not in national breeds (Montbéliarde and Normande). Their results, and ours, might suggest that Holstein, which is an international breed with large census size, but relatively small effective population size, might be more prone to GS mechanisms that accelerate ΔF than smaller national breeds. As discussed by Doublet et al. (2019), factors here might be use of imported bulls or number of inseminations per bull. Among other things, they point to massive use of few elite bulls in Holstein as probable cause for ΔF increase. Also, Miglior and Beavers (2014) found that although GS led to a higher number of bulls screened, a corresponding increase in diversity of selected bulls was not seen in US Holstein. On the contrary, number of bulls siring 50% of the young bulls entering artificial insemination was kept rather constant. Hence, even though it is possible, opportunity of less co-selection of relatives may not have been exploited in Holstein.

Results in our study is not in accordance with expectations regarding inbreeding and GS in NR. A simulation study done by Lillehammer et al. (2011) showed stagnation of decrease of ΔF for both PTS/GS and GS schemes. Decrease in ΔF was expected if ≥20 elite bulls were selected (Lillehammer et al., 2011), but although as many as 50, 34, 46 and 46 elite sires was selected for the four GS years (2016-2019 respectively), no decrease in ΔF is observed in our results. A decrease in PTS/GS was found. but standard errors suggest that these might be inaccurate estimates. Most likely due to paucity of data. This is supported by observation that different statistical analysis gave large variations in PTS/GS period (data not shown). Also, the number of elite bulls chosen for these years (2012-2015) was uncharacteristically low for NR with 10, 11, 9 and 6. This should lead to an increase in ΔF, not a decrease. Paucity of data is also a problem for GS estimates. PTS/GS estimates are based on 4 years and GS only 3.5 years. This is a short period, most likely giving insufficient data to base estimates upon. This is reflected in large standard errors given for PTS/GS and GS period in regression results.

Correlations between the different inbreeding metrics were in range with previous studies. We found correlations of 0.683 between F_{ROH} and F_{PED}. Ferenčaković et al. (2013) who also looked at NR and found correlations of 0.53-0.62 between F_{PED} and F_{ROH}. Studies in other breeds show corresponding results; Pryce et al. (2012) got 0.65. Makanjuola et al. (2020) 0.52-0.76 and Doublet et al. (2019) 0.50-0.59. Similarly. our observed correlations between F_{ROH} and F_{GRM} of 0.96 coincide with other findings

ranging from 0.81 (Bjelland et al., 2013), 0.90 (Makanjuola et al., 2020) to 0.94 (Forutan et al., 2018). Moderate correlations between genomic metrics and F_{PED} but high correlations between F_{ROH} and F_{GRM}, support assumption that genomic metrics are more adept at capturing biological sources of variation between relatives such as mendelian sampling and recombination. When we compare ΔF_{PED} estimates obtained in regression to ΔF_{ROH}, our results also point toward the possibility of underestimating ΔF when this is based on pedigree rather than genomic data. Hillestad (2017) found that F_{PED} underestimated ΔF compared to F_{ROH} in NR. Because of this, in addition to large consensus regarding increased accuracy of genomic rather than pedigree based inbreeding metrics (e.g. Baes et al., 2019; Bjelland et al., 2013; Ferdosi et al., 2016; Howard et al., 2017), it is advisable to improve NR breeding scheme by implementing routine evaluation of genomic inbreeding. Our results suggest that both F_{ROH} and F_{GRM} are proficient at capturing inbreeding levels in NR population. An advantage of using GRM is that Geno already constructs this as part of ssGBLUP procedure and development of methodology framework should require little effort. The downside is that GRM is sensitive to allele frequencies, and studies has shown that F_{GRM} overestimates inbreeding because it is less proficient than F_{ROH} at distinguishing IBS from IBD alleles (Baes et al., 2019). Forutan et al. (2018) found that F_{ROH} was a more appropriate metric because it was not sensitive to allele frequencies. Using F_{ROH} for routine inbreeding management will most likely require some more effort on part of NRs' breeding scheme, but our study provides a starting foundation and framework for doing this that can be built upon.

Region-specific inbreeding and trends in ROH

ROH is a useful feature of the genome that allows us to investigate inbreeding on a region-specific level. This study has looked at genome wide distribution of ROH and their frequency in different length classes. Average length and number of ROH as well as average inbreeding coefficient and annual inbreeding rates per chromosome was examined to try and discern trends in inbreeding before and after implementation of GS.

Comparing ΔF_{ROH} for the three selection schemes allowed us to examine inbreeding trends per chromosome before and after GS. Results show that increase in ΔF after GS was not significant for any chromosomes. Using observed homozygosity, Hillestad (2017) also investigated chromosomal ΔF per in NR, and found the highest rates for BTA 5, 6, 14, 20 and 24. In our study, three of these were among the chromosomes with highest ΔF. Those were BTA 5, 14 and 20. We showed that BTA 14 also distinguished itself by having high average F_{ROH} compared to other chromosomes (figure 10). This chromosome contains many gene variants influencing economically important traits in cattle (Hillestad, 2017), and it is the chromosome at which the well-known DGAT1 gene with major effect on milk characteristics is positioned (Grisart et al., 2002). Our results show high average F_{ROH} for BTA 14 for all three periods, but noticeably decline in ΔF from 0.254 to -0.079% for PTS and GS periods

respectively. This might correspond with Hillestad (2017) who found fixed haplotypes on this chromosome and signs of a historical selective sweep, but no ongoing sweep nor total fixation.

We found that short (<2 Mb) and medium (2-4 Mb) ROH are highly predominant in NR. This coincides with other studies (Forutan et al., 2018; Marras et al., 2015; McQuillan et al., 2008). Frequency distribution is however inconsistent with other results. Forutan et al. (2018) found relative frequencies in North American Holstein to be 43.5% (<2 Mb), 23.9% (2-4 Mb), 17.7% (4-8 Mb), 10.5% (8-15 Mb) and 4.7% (>16 Mb). Similar frequencies is shown in Italian Holstein (Marras et al., 2015). We found relative frequencies of 74% (<2 Mb), 14% (2-4 Mb), 7.5% (4-8 Mb), 2.4% (8-16 Mb) and 1.2% (>16 Mb). (averaged for three periods). These results point towards ROH in NR being much more accumulated as short and medium regions than ROH in Holstein. Looking at 5 different cattle breeds, Marras et al. (2015) found that dual-purpose and beef cattle had fewer long ROH compared to dairy cattle. Highest frequency of short ROH was 66.6% and found in dual-purpose Italian Simmental. This coincides with Kim et al. (2013) who found higher frequency of long ROH in populations with small effective population size and intense selection. Also, Zhang et al. (2015b) observed that New Danish Red Cattle, which is a composite breed, displayed more small size ROH than other Danish cattle breeds. As NR is both a composite and a dual-purpose cattle breed with relatively large effective population size, finding high frequencies of short and medium ROH in this population is thus no surprise.

However, selection scheme cannot explain all the differences in our results compared to other studies. Ferenčaković et al. (2013) looked at ROH distribution in NR and found mean number of ROH to be 80.8 per animal. This is similar to 82.3 and 81.7 found in Holstein breeds by Forutan et al. (2018) and Marras et al. (2015), but considerably lower than number found in our study (which ranged 127.6-137.3 for three periods). This difference in results are probably due to differences in ROH detection. We have most likely detected many short ROH not found by Ferenčaković et al. (2013). For one, we used HD rather than medium density data. As Hillestad et al. (2017) showed, use of HD data detected more ROH and gave redistribution from long to short ROH. Also, unlike Ferenčaković et al. (2013), we were able to validate our choice of detection parameters. This way, we could set appropriate criteria and detect ROH that would have been omitted otherwise. Also, minimum length for ROH detection used in this study was 500 kb, while Ferenčaković et al. (2013) and Marras et al. (2015) used 1000. 0.5-1 Mb ROH is thus not detected. On the one hand, exclusion of short ROH such as these is understandable because chances are higher that they originate due to chance. Also, they rarely contribute considerably to total autozygosity in the genome (Forutan et al., 2018) On the other hand however, short ROH (> 3 Mb) has been found to be enriched with deleterious variants (Zhang et al., 2015b), and detection of these contribute to our understanding of inbreeding in the genome.

Our results with many short ROH indicate little recent inbreeding in NR. Other studies has found significant changes in ROH after GS, Forutan et al. (2018) detecting a higher number of ROH and increased frequency of short and medium ROH, while Doublet et al. (2019) observed a significant increase in mean ROH length for French Holstein, (but not Montbéliarde and Normande) after GS. Although small, some increase in average length and average number of ROH can be discerned in our results. Considering that our estimates are based on only 3.5 years of data, development should be monitored closely in the future. It should be noted however that both for genome wide and chromosomal estimates, ΔF are well within recommended limits of 0.5-1% (FAO, 1998). However, our estimates found, that estimated genome wide ΔF can hide considerably higher rates on individual chromosomes. For instance, genome wide ΔF_{ROH} was 0.068% for GS period, but ΔF for same period was estimated at 0.338% on BTA 20. Using region-specific metrics will in other words give a more thorough control of inbreeding rates in the population. Although not done in this study, consideration of intra-chromosomal regions might also help us to detect regions that need specific management (Kleinman-Ruiz et al., 2016). These methods also hold great potential in selecting for genetic gain as, pointed out by Howrigan et al. (2011), it allows us to breed animals with similar genome wide inbreeding levels if these are not in same regions, or found to be in regions less detrimental than others.

Large standard errors apply for both genome wide and chromosomal estimates obtained in this study. The reason is most likely lack of data. This is supported by observation that PTS/GS period displays the largest standard error. These estimates are based on only 4 years of data and consists of much fewer genotyped animals than for GS period (figure 2). Number of genotyped animals probably also influenced fluctuations in F_{ROH} and F_{GRM} estimates, as these were noticeably large for initial period (figure 6) when number of genotypes is low. Another factor influencing standard errors is that regression was based on yearly averages rather than individual data points. This reduces effective number of degrees of freedom and power of statistical analysis. Regression on individual data was preformed to compare results from different statistical treatments. The largest differences were however found for PTS/GS period which has contributed to a lot of statistical noise in all analysis in this study. Ideally, source of this noise should be identified, and estimations repeated with statistical treatments that can buffer against it.

Conclusion

An important aim of this study was to try and investigate trends in inbreeding after implementation of GS in NR. Results from F_{ROH}, F_{GRM} and F_{PED} revealed no significant increase in rates of inbreeding. Detection of many short ROH indicate little recent inbreeding in population and both genome wide and chromosomal inbreeding rates were well within recommended limit of 0.5-1%. Paucity of data due to few years having elapsed since GS implementation makes it necessary to repeat evaluations in some years' time. Our study lay preliminary foundations for development of methods to manage genomic inbreeding in NR population. Comparison of inbreeding metrics indicate that F_{PED} might underestimate inbreeding rate, but high correlations between F_{ROH} and F_{GRM} indicate aptitude of both metrics when it comes to estimating inbreeding in NR.

Optimization of ROH detection in HD data from NR enabled us to investigate trends in inbreeding and genetic diversity both genome wide and region specific. An abundance of short ROH was detected, and our detection was able to locate larger number of short ROH than previous studies. Results show that using HD data and validating detection parameters highly influence ROH analysis. We found that ROH are a genomic feature that can provide much information regarding both inbreeding and genetic diversity. Looking at distribution, length and number of ROH in the genome gives us indications of demographic history as well as the intrinsic structures around genomic regions of interest. Region-specific inbreeding holds great promise for more meticulous management of inbreeding in commercial breeding schemes.

References

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.

Bjelland, D., Weigel, K., Vukasinovic, N. & Nkrumah, J. (2013). Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *Journal of Dairy Science*, 96 (7): 4697-4706.

Bosse, M., Megens, H.-J., Madsen, O., Paudel, Y., Frantz, L. A., Schook, L. B., Crooijmans, R. P. & Groenen, M. A. (2012). Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS genetics*, 8 (11).

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M. & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4 (1): 7.

Curik, I., Ferenčaković, M. & Sölkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livestock Science*, 166: 26-34.

Daetwyler, H. D., Villanueva, B., Bijma, P. & Woolliams, J. A. (2007). Inbreeding in genome-wide selection. *Journal of Animal Breeding and Genetics*, 124 (6): 369-376.

Doekes, H. P., Veerkamp, R. F., Bijma, P., Hiemstra, S. J. & Windig, J. J. (2018). Trends in genome-wide and region-specific genetic diversity in the Dutch-Flemish Holstein–Friesian breeding program from 1986 to 2015. *Genetics Selection Evolution*, 50 (1): 15.

Doublet, A.-C., Croiseau, P., Fritz, S., Michenet, A., Hozé, C., Danchin-Burge, C., Laloë, D. & Restoux, G. (2019). The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. *Genetics Selection Evolution*, 51 (1): 52.

FAO. (1998). *Initiative for Domestic Animal Diversity Secondary guidelines for development of national farm animal genetic resources management plans: Management of small populations at risk*: FAO.

Ferdosi, M. H., Henshall, J. & Tier, B. (2016). Study of the optimum haplotype length to build genomic relationship matrices. *Genetics Selection Evolution*, 48 (1): 75.

Ferenčaković, M., Hamzić, E., Gredler, B., Solberg, T., Klemetsdal, G., Curik, I. & Sölkner, J. (2013). Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding and Genetics*, 130 (4): 286-293.

Forutan, M., Mahyari, S. A., Baes, C., Melzer, N., Schenkel, F. S. & Sargolzaei, M. (2018). Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC genomics*, 19 (1): 98.

Geno. (2018). Norwegian Red characteristics. Available at: https://www.norwegianred.com/Start/Norwegian-Red/about-norwegian-red/Norwegian-Red-characteristics/?utm_source=forside&utm_medium=banner&utm_campaign=NR-characteristics (accessed: 28.08.19).

Geno. (2019). Årsberetning og regnskap for Geno 2019. Available at: https://www.geno.no/globalassets/geno-sa/02_dokumenter/11_nytt_for_tillitsvalgte/2020/arsmote-2020/signert-arsberetning-og-regnskap-geno-sa-2019.pdf.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S. & Simon, P. (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome research*, 12 (2): 222-231.

Hillestad, B. (2017). Inbreeding determined by the amount of homozygous regions in the genome.

Hillestad, B., Woolliams, J. A., Boison, S. A., Grove, H., Meuwissen, T., Våge, D. I. & Klemetsdal, G. (2017). Detection of runs of homozygosity in Norwegian Red: Density, criteria and genotyping quality control. *Acta Agriculturae Scandinavica, Section A—Animal Science*, 67 (3-4): 107-116.

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.

Howrigan, D. P., Simonson, M. A. & Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC genomics*, 12 (1): 460.

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.

Kim, E.-S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassell, C. P., Crooker, B. A., Liu, G., Da, Y. & Sonstegard, T. S. (2013). Effect of artificial selection on runs of homozygosity in US Holstein cattle. *PloS one*, 8 (11): e80813.

Kleinman-Ruiz, D., Villanueva, B., Fernández, J., Toro, M., García-Cortés, L. & Rodríguez-Ramilo, S. (2016). Intra-chromosomal estimates of inbreeding and coancestry in the Spanish Holstein cattle population. *Livestock Science*, 185: 34-42.

Lencz, T., Lambert, C., DeRosse, P., Burdick, K. E., Morgan, T. V., Kane, J. M., Kucherlapati, R. & Malhotra, A. K. (2007). Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proceedings of the National Academy of Sciences*, 104 (50): 19942-19947.

Lillehammer, M., Meuwissen, T. & Sonesson, A. (2011). A comparison of dairy cattle breeding designs that use genomic selection. *Journal of Dairy Science*, 94 (1): 493-500.

Makanjuola, B. O., Miglior, F., Abdalla, E. A., Maltecca, C., Schenkel, F. S. & Baes, C. F. (2020). Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. *Journal of Dairy Science*.

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., Williams, J. L. & Macciotta, N. P. (2015). Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Animal genetics*, 46 (2): 110-121.

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. (2008). Runs of homozygosity in European populations. *The American Journal of Human Genetics*, 83 (3): 359-372.

Meuwissen, T., Hayes, B. & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157 (4): 1819-1829.

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): 1-14.

Miglior, F. & Beavers, L. (2014). Genetic diversity and inbreeding: before and after genomics. Available at: https://www.progressivedairy.com/topics/a-i-breeding/genetic-diversity-and-inbreeding-before-and-after-genomics (accessed: 15.08.19).

Nordbø, Ø., Gjuvsland, A. B., Eikje, L. S. & Meuwissen, T. (2019). Level-biases in estimated breeding values due to the use of different SNP panels over time in ssGBLUP. *Genetics Selection Evolution*, 51 (1): 76.

Peripolli, E., Munari, D., Silva, M., Lima, A., Irgang, R. & Baldi, F. (2016). Runs of homozygosity: current knowledge and applications in livestock. *Animal genetics*, 48 (3): 255-271.

Pryce, J., Hayes, B. & Goddard, M. (2012). Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. *Journal of dairy science*, 95 (1): 377-388.

Purfield, D. C., Berry, D. P., McParland, S. & Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC genetics*, 13 (1): 70.

Sargolzaei, M., Chesnais, J. P. & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC genomics*, 15 (1): 478.

Scraggs, E., Zanella, R., Wojtowicz, A., Taylor, J., Gaskins, C., Reeves, J., de Avila, J. & Neibergs, H. (2014). Estimation of inbreeding and effective population size of full-blood wagyu cattle registered with the American Wagyu Cattle Association. *Journal of Animal Breeding and Genetics*, 131 (1): 3-10.

Solé, M., Gori, A.-S., Faux, P., Bertrand, A., Farnir, F., Gautier, M. & Druet, T. (2017). Age-based partitioning of individual genomic inbreeding levels in Belgian Blue cattle. *Genetics Selection Evolution*, 49 (1): 92.

Sonesson, A. K., Woolliams, J. A. & Meuwissen, T. H. (2012). Genomic selection requires genomic control of inbreeding. *Genetics Selection Evolution*, 44 (1): 27.

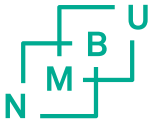
Strandén, I. (2014). *RelaX2 program for pedigree analysis, User's guide for version 1.65*.

VanRaden, P. (1992). Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *Journal of Dairy Science*, 75 (11): 3136-3144.

VanRaden, P., Olson, K., Wiggans, G., Cole, J. & Tooker, M. (2011). Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *Journal of Dairy Science*, 94 (11): 5673-5682.

Zhang, Q., Calus, M. P., Guldbrandtsen, B., Lund, M. S. & Sahana, G. (2015a). Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC genetics*, 16 (1): 88.

Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M. S. & Sahana, G. (2015b). Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC genomics*, 16 (1): 542.



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

Postboks 5003
NO-1432 Ås
Norway