UNIVERSITY OF HELSINKI
DOCTORAL SCHOOL IN ENVIRONMENTAL, FOOD AND BIOLOGICAL SCIENCES
Doctoral Programme in Sustainable Use of Renewable Natural Resources
Department of Agricultural Sciences
(Animal Science)

PUBLICATION No. 20/2020

GENOMIC ANALYSIS OF INBREEDING DEPRESSION ON FERTILITY TRAITS IN FINNISH AYRSHIRE CATTLE

DOCTORAL THESIS

KATJA MARTIKAINEN

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Agriculture and Forestry of the University of Helsinki, for public examination in Lecture Hall 108 (LS B3), Latokartanonkaari 7, Helsinki, on November 25th, 2020, at 1 p.m.

HELSINKI, FINLAND
2020
Custos:  Professor Pekka Uimari  
Department of Agricultural Sciences  
University of Helsinki, Finland

Supervisor:  Professor Pekka Uimari  
Department of Agricultural Sciences  
University of Helsinki, Finland

Reviewers:  Associate Professor Christine Baes  
Department of Animal Biosciences  
University of Guelph, Canada

Dr. Luis Gomez-Raya  
Department of Animal Breeding  
The National Institute for Agricultural and Food Research and Technology, Spain

Opponent:  Professor Erling Strandberg  
Department of Animal Breeding and Genetics  
Swedish University of Agricultural Sciences, Sweden

Cover photo: Terhi Laurila

Dissertationes Schola Doctoralis Scientiae Circumiectalis, Alimentariae, Biologicae  
ISBN 978-951-51-6195-6 (PDF)  
ISSN 2342-5423 (Print)  
ISSN 2342-5431 (Online)  
Electronic publication at http://ethesis.helsinki.fi

The Faculty of Agriculture and Forestry of the University of Helsinki uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

Unigrafia  
Helsinki 2020
ABSTRACT

Inbreeding increases homozygosity, which in turn increases the frequency of harmful recessive alleles and decreases advantageous heterozygosity, resulting in inbreeding depression. Inbreeding depression on fertility decreases the lifetime milk production of cows and increases the need for veterinary treatments, inseminations and involuntary cullings, which has a negative impact on the profitability, sustainability and animal welfare of the dairy sector.

The main goal of this thesis was to estimate the effect of inbreeding on female fertility in Finnish Ayrshire cattle. More specifically, this study aimed to estimate inbreeding depression using different metrics based on pedigree and genomic information, and to dissect inbreeding depression to chromosomes, chromosomal segments and specific homozygous genotypes. Moreover, it identified specific ROH genotypes that had detrimental effect on female fertility and milk production traits.

The data used included over 12 000 genotyped cows with corresponding phenotypic records of fertility and milk production traits. The fertility traits studied included the non-return rate at 56 days after first insemination (NRR), number of inseminations per conception (AIS), interval from calving to first insemination (ICF) and interval from first to last insemination (IFL). Milk production traits included milk, protein and fat yields. Inbreeding coefficients were estimated based on pedigree ($F_{\text{PED}}$), percentage of homozygous SNPs ($F_{\text{PH}}$) and runs of homozygosity ($F_{\text{ROH}}$).

$F_{\text{PED}}$ was not found to be associated with inbreeding depression in any of the fertility traits. However, statistically significant inbreeding depression was observed when using genomic measures of inbreeding ($F_{\text{PH}}$ and $F_{\text{ROH}}$) in the model. For example, a 10% increase in $F_{\text{ROH}}$ was associated with 4 and 6 days longer IFL in heifers and first-parity cows, respectively.

When $F_{\text{ROH}}$ was used to estimate inbreeding depression separately for each autosomal chromosome, a 10% increase in $F_{\text{ROH}}$ on chromosomes 2, 18 and 22 was observed to increase IFL in heifers by 1.6, 0.9 and 0.7 days, respectively. Similarly, a 10% increase in $F_{\text{ROH}}$ on chromosome 15 was associated with the lengthening of IFL in second-parity cows by 2.3 days. Haplotype analysis for the detected regions revealed haplotypes that, when occurring as homozygous, were associated with the lengthening of IFL of approximately 4 and 8 days in heifers and second-parity cows, respectively.

Finally, an analysis of unique ROH genotypes revealed several ROHs with unfavourable effects on fertility and milk production traits. The estimated effects of ROHs with the highest statistical significance varied between parities from 13 to 38 days longer IFL, from 9 to 17 days longer ICF and from 0.3 to 1.0 more AIS. For milk production traits, the ROHs with the highest statistical significance resulted in reductions of 208 kg for milk yield, 7 kg for protein yield and 16 kg for fat yield. In
addition, this study found regions in which inbreeding is particularly harmful due to ROHs displaying unfavourable effects across multiple traits.

The detrimental effects of inbreeding observed in this study highlight the importance of managing the levels of inbreeding in the Finnish Ayrshire breeding programme. The findings of this study can be utilized for more efficient control of inbreeding depression as well as investigations of the mechanisms of inbreeding depression by examining the identified regions showing inbreeding depression.
ACKNOWLEDGEMENTS

The path towards this dissertation, carried out at the Department of Agricultural Sciences, University of Helsinki, has been long and full of challenges but at the same time very exciting and fulfilling. My gratitude is owed to all contributors who have made this journey possible.

I am grateful for the funding received from Naudanjalostussäätiö sr, the Ministry of Agriculture and Forestry, the Oiva Kuusisto Foundation, the Doctoral Program in Sustainable Use of Renewable Natural Resources and the University of Helsinki.

I would like to express the deepest appreciation to my supervisor, Professor Pekka Uimari. I am thankful for your great expertise, guidance and support throughout the years of my master and PhD studies. You have made my journey meaningful and inspiring by always knowing when to challenge me to be independent and solve problems by myself, and when to provide advice, help and encouragement. Without your supervision, I would not have grown to be the researcher that I am today.

I warmly acknowledge my coauthors Minna Koivula and Jukka Pösö. The collaboration with you and your help with my data was essential for this thesis. I sincerely thank my other coauthors, Maria Leino, Kaarina Matilainen and Anu Sironen for sharing their expertise as well as for providing valuable and heartening feedback in the preparation of the manuscripts.

I am grateful to the reviewers, Associate Professor Christine Baes and Dr. Luis Gomez-Raya, for their constructive comments and suggestions to improve this thesis. My sincere gratitude is owed to Professor Erling Strandberg for honouring me by being my opponent and Professor Pekka Uimari for being my custos. I would like to gratefully acknowledge Professors Asko Mäki-Tanila, Päivi Onkamo and Johanna Vilkki, the members in the follow-up group of my doctoral studies, and Professor Helena Korpelainen, the faculty representative in the evaluation of my dissertation.

I thank all my colleagues, past and present, at the Department of Agricultural Sciences: it has been a real pleasure to work with you. I particularly thank Johanna Häggman, Puuliina Kokko and Hanni Kärkkäinen for scientific and not-so-scientific help to survive PhD.

My loving gratitude goes to my family and friends. You were always there for me during the times of joy and celebration, and times of frustration and despair. And finally, to my dearest fiancé Tuomas, who has been my inspiration, motivation and support. My rock!


Rakkain Tuomas, ilman tukeasi ja kannustustasi en olisi rohjennut edes aloittaa jatko-opintoja, enkä varmasti olisi selvinnyt maaliin. Kiitos kun jaksoit kaikki suuret tunteeni, juhlit kanssani ilon hetkellä ja käräjän jälkeen kannattelit yli epätoivon karikoiden. Oot mun lepardi!
ABSTRACT ................................................................................................................ 3
ACKNOWLEDGEMENTS............................................................................................ 5
LIST OF ORIGINAL PUBLICATIONS........................................................................... 8
ABBREVIATIONS ..................................................................................................... 9
1 INTRODUCTION .................................................................................................. 10
  1.1 Effects of low fertility ...................................................................................... 10
  1.2 Inbreeding and inbreeding depression ............................................................. 11
  1.3 Measures of inbreeding .................................................................................... 12
    1.3.1 Pedigree inbreeding coefficient ................................................................. 13
    1.3.2 Genomic inbreeding coefficients ............................................................... 13
  1.4 Estimation of inbreeding depression ............................................................... 15
    1.4.1 Whole genome estimation of inbreeding depression .................................. 15
    1.4.2 Genomic regions contributing to inbreeding depression ......................... 15
2 OBJECTIVES ......................................................................................................... 17
3 MATERIALS AND METHODS ........................................................................... 18
  3.1 Materials .......................................................................................................... 18
    3.1.1 Pedigree data .............................................................................................. 18
    3.1.2 Genotypic data .......................................................................................... 18
    3.1.3 Quality control and pruning of SNP ............................................................ 19
    3.1.4 Phenotypes of female fertility traits ............................................................ 19
    3.1.5 Deregressed proofs of milk production traits ............................................ 20
    3.1.6 Merging datasets ....................................................................................... 21
  3.2 Methods ............................................................................................................ 21
    3.2.1 Estimation of inbreeding coefficients ......................................................... 21
3.2.2 Estimation of inbreeding depression on fertility ........................................ 23
3.2.3 Identification of haplotypes associated with reduced fertility ............... 23
3.2.4 Identification of ROH genotypes associated with reduced fertility and milk production traits ................................................................. 24

4 RESULTS AND DISCUSSION ............................................................................. 26
4.1 Runs of homozygosity .............................................................................. 26
4.2 Inbreeding coefficients ............................................................................. 29
4.3 Correlations between different measures of inbreeding ......................... 30
4.4 Whole genome estimates of inbreeding depression on fertility ............... 32
4.5 Intrachromosomal estimates of inbreeding depression on fertility .......... 34
4.6 Haplotypes associated with inbreeding depression ............................... 37
4.7 ROH genotypes associated with reduced fertility and milk production .... 39
4.8 Implications and future research ............................................................... 43

5 CONCLUSIONS ..................................................................................................... 46
REFERENCES ........................................................................................................... 47
LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following articles, which are referrer in the text by their Roman numerals:


Contribution of the author to articles I to III:

The author participated in planning the studies, data editing, statistical analyses, interpretation of results, and dissemination of research outcomes to the journals as the main author.
# ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Artificial insemination</td>
</tr>
<tr>
<td>AIS</td>
<td>Number of inseminations</td>
</tr>
<tr>
<td>DRP</td>
<td>De-regressed proof</td>
</tr>
<tr>
<td>EBV</td>
<td>Estimated breeding value</td>
</tr>
<tr>
<td>F</td>
<td>Inbreeding coefficient</td>
</tr>
<tr>
<td>FAT</td>
<td>Deregressed proof of fat yield</td>
</tr>
<tr>
<td>FAY</td>
<td>Finnish Ayrshire</td>
</tr>
<tr>
<td>GRM</td>
<td>Genomic relationship matrix</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome wide association study</td>
</tr>
<tr>
<td>HWE</td>
<td>Hardy-Weinberg equilibrium</td>
</tr>
<tr>
<td>IBD</td>
<td>Identical by descent</td>
</tr>
<tr>
<td>IBS</td>
<td>Identical by state</td>
</tr>
<tr>
<td>ICF</td>
<td>Interval from calving to first insemination</td>
</tr>
<tr>
<td>IFL</td>
<td>Interval from first to last insemination</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase</td>
</tr>
<tr>
<td>LD</td>
<td>Linkage disequilibrium</td>
</tr>
<tr>
<td>Mb</td>
<td>Megabase</td>
</tr>
<tr>
<td>MAF</td>
<td>Minor allele frequency</td>
</tr>
<tr>
<td>MILK</td>
<td>Deregressed proof of milk yield</td>
</tr>
<tr>
<td>NAV</td>
<td>Nordic Cattle Genetic Evaluation</td>
</tr>
<tr>
<td>NRR</td>
<td>Non-return rate</td>
</tr>
<tr>
<td>NTM</td>
<td>Nordic Total Merit</td>
</tr>
<tr>
<td>PROT</td>
<td>Deregressed proof of protein yield</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative trait locus</td>
</tr>
<tr>
<td>RDC</td>
<td>Nordic Red dairy cattle</td>
</tr>
<tr>
<td>ROH</td>
<td>Runs of homozygosity</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

Not only is good fertility of dairy cows necessary for milk production, it is also economically important for farmers and impacts the sustainability of the dairy sector. What is considered to be good female fertility is a combination of clear expression of heat, successful ovulation, ability to start and resume ovarian cyclicity and ability to conceive and maintain pregnancy (Royal et al., 2000). Hence, fertility is a complex breeding goal affected by several environmental and genetic factors that complicate the achievement of genetic progress through breeding (Berry et al., 2014).

In Nordic countries (Finland, Sweden and Denmark) the breeding values of Red Dairy Cattle (RDC, including Finnish Ayrshire) cows in milk recording are calculated jointly by Nordic Cattle Genetic Evaluation (NAV). Selection for female fertility is based on fertility index, which is part of the Nordic Total Merit (NTM) breeding goal (NAV Nordic Cattle Genetic Evaluation, 2019). Fertility index is composed of the following traits: interval from calving to first insemination for cows (ICF), interval from first to last insemination for heifers and cows (IFL) and number of inseminations for heifers and cows (AIS) (NAV Nordic Cattle Genetic Evaluation, 2019). Fertility index reflects the genetic potential of a heifer or a cow for good reproductive ability (NAV Nordic Cattle Genetic Evaluation, 2019).

A commonly known difficulty of breeding for fertility is the low heritability of the desired traits, which is why the genetic progress of fertility is slow. The estimates for heritability of fertility traits are mostly below 0.05 (e.g. Mäntysaari and Van Vleck, 1989; Berry et al., 2013; El-Bayoumi et al., 2015; Muuttoranta et al., 2019). However, low heritability may not reflect a lack of genetic variance, provided the phenotypic variance of fertility in the population is large. Muuttoranta et al. (2019) obtained coefficients of genetic variation estimates for fertility traits (ICF, IFL and conception rate) in the Nordic RDC and Holstein population similar or even higher than coefficients of genetic variation estimates found for milk production traits. Hence, exploitable genetic variance exists for female fertility traits in the RDC population (Muuttoranta et al., 2019).

1.1 Effects of low fertility

Although female fertility is an important breeding goal in dairy cattle breeding in Finland, fertility problems exist and they reduce the profitability and sustainability of dairy production in many ways. For example, infertility contributes to decreased milk production per cow per year by increasing the calving interval. Over the past 10 years (from 2008 to 2018), the calving interval of all cows in the Finnish milk production recording system has varied between 407 days and 418 days (ProAgria, 2019). For the Finnish Ayrshire (FAY) breed, the calving interval results based on milk recording from 2018 was 405 days (S. Nokka, ProAgria, personal communication, November
The calving interval in Finnish dairy cattle is far from optimal, since the economically most advantageous calving interval would generally be approximately one year (e.g. Strandberg and Oltenacu, 1989; Heikkilä, 1999). Problems with bringing cows to calve in the desirable season will further decrease profitability due to seasonal differences (in feed costs and milk components, for example) (Heikkilä, 1999). In addition, poor fertility leads to a reduction in the number of calves born per cow, which impacts the profitability of milk production. In Finland, 85% of beef is a by-product of milk production (Luke Natural Resources Institute Finland, 2016) and calves that are not used for replacement are sold to beef production. Furthermore, the number of inseminations per calving in FAY in the Finnish milk recording was 1.92 in 2018 (S. Nokka, ProAgria, personal communication, November 15, 2019). These repeated inseminations increase costs through, for example, increased working hours and semen dose purchases.

An additional economical effect of infertility is increased veterinary costs. According to results from the Finnish health management system in 2018, fertility treatments represented 34% of all veterinary treatments in FAY, and 21.6% of all FAY cows were treated for fertility problems (Faba, 2019). Hence, the most common reason for veterinary treatments in FAY is fertility problems (Faba, 2019).

Above all, fertility problems lead to involuntary cullings, which increase costs. In 2018, the most common reason for culling of first-calf FAY cows included in milk recording was poor fertility, representing 23.2% of all first-calf FAY cullings (S. Nokka, ProAgria, personal communication, November 15, 2019). For older FAY cows, poor fertility was the second most common reason for culling after mastitis in 2018, representing 17.1% of all the cullings of FAY cows included in milk recording with two or more calves (S. Nokka, ProAgria, personal communication, November 15, 2019).

In addition to negatively impacting profitability, poor fertility deteriorates the sustainability of dairy production. First, veterinary treatments and premature cullings due to reproductive problems affect animal welfare (Royal et al., 2000). Second, premature cullings increase the number of replacement heifers required, which has negative environmental effects through increased methane and ammonia emissions (Garnsworthy, 2004). Maintaining the profitability of dairy production at reasonable levels, while responding to the growing public interest towards animal welfare and the effect of dairy production on climate change, is a challenge for the dairy sector. Therefore, improving fertility will play an important role in facing this challenge.

1.2 Inbreeding and inbreeding depression

The success of modern breeding programmes on dairy production is undeniable. For example, in Finland the average milk yield per cow per year has increased by approximately 1000 kg from 2008 to 2018 (ProAgria, 2019). Though the increase is partly explained by improved management, intensive selection with the help of
reproductive techniques (e.g. artificial insemination (AI) or embryo transfer) together with accurate breeding value estimation has been advantageous for genetic progress. However, animals that receive the highest breeding values with the commonly used best linear unbiased prediction (BLUP) method tend to be related to each other (Vahlsten et al., 2004). The breeding practices, including high intensity of selection with relative-favouring genetic evaluations, are expected to result in increased inbreeding (Miglior and Burnside, 1995). This hypothesis was supported by Vahlsten et al. (2004), who observed an increase of 0.20 in the rate of inbreeding (ΔF) per generation between the years 1976 and 1999 in the Finnish Ayrshire population.

The general harmful effects of increased inbreeding are increased homozygosity and reduced genetic variance (Falconer & Mackay, 1996). These events lead to inbreeding depression, which is defined as impairment in phenotypic performance within a population (Falconer & Mackay, 1996). According to the partial dominance hypothesis, inbreeding depression is caused by the increased homozygosity of both rare lethal (or near lethal), almost completely recessive alleles and detrimental, partially recessive alleles (Charlesworth and Charlesworth, 1999). Second hypothesis for the genetic background of inbreeding depression is overdominance, where the alleles with heterozygous advantage are decreased due to increased homozygosity (Hedrick and Garcia-Dorado, 2016). A third possible mechanism behind inbreeding depression may be epistasis, where the increased homozygosity decreases the frequency of heterozygous genotypes that have favourable interactions (Jain and Allard, 1966).

Inbreeding is particularly harmful for dairy production, since inbreeding depression impairs more fitness traits (e.g. fertility and health) than other types of traits (e.g. conformation traits) (Kristensen and Sorensen, 2005). Furthermore, inbreeding is expected to reduce heritability through reduced genetic variance (Kristensen et al., 2005), which will hamper genetic progress especially for low heritability traits such as fertility.

1.3 Measures of inbreeding

The level of inbreeding of an individual is expressed as an inbreeding coefficient (F), which was originally defined as a correlation coefficient of the uniting gametes (Wright, 1921). A more typical definition of F is the probability of two alleles at any locus being identical by descent (IBD) (Malecot, 1948). Locus is IBD when it carries two allele copies both originated from the same common ancestor. A number of methods to estimate F based on pedigree or genomic information have been developed. Described here are some of the most commonly used approaches.
1.3.1 Pedigree inbreeding coefficient

The classical method used to estimate inbreeding is to use path analysis (Wright, 1922) of a pedigree ($F_{\text{PED}}$). The calculation of $F_{\text{PED}}$ usually includes only the effect of known common ancestors of parents; the pedigree founders are assumed to be unrelated and non-inbred (Wang, 2014). This assumption may result in an underestimation of $F_{\text{PED}}$, especially in shallow or incomplete pedigrees, because the inbreeding caused by ancestors missing from the pedigree is not accounted for (Kardos et al., 2015). Although VanRaden (1992) proposed a method to include the effect of unknown ancestors in the estimation of $F_{\text{PED}}$, the estimates can still be inaccurate due to the variation in IBD among individuals with the same pedigree as a result of recombination events (Keller et al., 2011). Moreover, errors in pedigree due to misidentification are common and will further impair the accuracy of $F_{\text{PED}}$ (Israel and Weller, 2000).

1.3.2 Genomic inbreeding coefficients

The availability of genetic data has made it possible to improve the accuracy of IBD estimation compared with pedigree information. As reported by Keller et al. (2011), the use of genomic data in the estimation of $F$ has many advantages over pedigree-based estimation: first, the reliability of genomic estimates of $F$ do not suffer from unknown ancestors or pedigree errors. Second, genomic methods can estimate the actual proportion of the genome in IBD more accurately than pedigrees. Third, $F$ can be estimated on certain regions on the genome, for example by estimating separate $F$ for different chromosomes. Lastly, genomic $F$ can be estimated for individuals for which pedigree information is not available. However, genomic $F$ can be obtained using several different methods, each of which may produce different results. Next, some of the commonly used genomic estimates of $F$ are presented.

Percent of homozygous genotypes

Single nucleotide polymorphisms (SNP) are the most commonly used genetic variants in estimation of genomic based inbreeding. The simplest method for estimating the levels of inbreeding from SNP data is to determine the percentage of homozygosity ($F_{\text{PH}}$) of all SNPs. Although $F_{\text{PH}}$ can provide some information about the levels of inbreeding, it does not distinguish between SNPs that are identical by state (IBS) from those that are IBD (Bjelland et al., 2013). Since $F_{\text{PH}}$ reflects the levels of overall homozygosity, the estimates are on a different scale than those estimates that take into account the expected number of homozygous IBS genotypes. The advantage of $F_{\text{PH}}$ is that it results in uniform estimates throughout the individuals, because it is not affected by the estimation of the allele frequencies in the base population.
**Excess of homozygosity**
The excess of homozygosity ($F_{HOM}$) estimate by Wright (1948) is based on the observed versus expected number of homozygous genotypes. $F_{HOM}$ is sensitive to the estimation of allele frequencies in the reference population (Zhang et al., 2015a). $F_{HOM}$ can take negative values if an individual has less homozygosity than what is expected based on the reference population (Wang, 2014). Furthermore, $F_{HOM}$ estimates may produce different results depending on the SNP panel used, because different panels contain different SNPs with different allele frequencies (Wang, 2014). Kardos et al. (2015) reported that $F_{HOM}$ underestimates the levels of inbreeding because allele frequencies are derived from the current population. Unbiased estimates of $F_{HOM}$ would require a reference population in the Hardy-Weinberg equilibrium (Kardos et al., 2015).

**Genomic relationship matrix**
Genomic data can be used to construct a genomic relationship matrix (GRM) (e.g. VanRaden, 2008). Values on the diagonal of GRM represent the relationship of the individual to itself (VanRaden, 2008). To obtain GRM-based inbreeding coefficients ($F_{GRM}$), one is subtracted from the diagonal values of GRM. As well as $F_{HOM}$, $F_{GRM}$ is influenced by the assumptions regarding allele frequencies in the reference population (VanRaden, 2008). $F_{GRM}$ may rise to incorrectly high levels for individuals whose allele frequencies differ from the population unless GRM is calculated with allele frequencies fixed at 0.5 (Pryce et al., 2014). In addition, likewise other SNP-by-SNP estimates of inbreeding (here $F_{PH}$ and $F_{HOM}$), also $F_{GRM}$ suffers from an inability to truly distinguish between IBD and IBS alleles.

**Runs of homozygosity**
Broman and Weber (1999) proposed identifying homozygous IBD segments from stretches of uninterrupted homozygous SNP genotypes called runs of homozygosity (ROH). Since ROH are assumed to result only from inbreeding, estimation of $F$ based on ROH ($F_{ROH}$) is currently considered to be the most effective procedure by which to quantify genomic inbreeding (Curik et al., 2017). Moreover, estimation of $F_{ROH}$ does not require estimation of allele frequencies (McQuillan et al., 2008).

$F_{ROH}$ can be obtained, for example, as the total length of all ROH divided by the length of the genome covered by SNPs in the genotyping panel (McQuillan et al., 2008), or as the number of SNPs in ROH divided by the total number of SNPs in the genotyping panel (e.g. Pryce et al., 2014). The length of ROH segments can reflect the distance (in generations) to a common ancestor, since recombination events may interrupt long ROHs. Thus, short ROHs are expected to reflect ancient inbreeding, whereas long ROHs imply that inbreeding happened only a few generations ago (Howrigan et al., 2011). The limitation of ROH analysis is the lack of standardized criteria for defining the ROHs, which complicates the comparison of results from different studies using their own criteria for determining ROHs (Ku et al., 2011).
1.4 Estimation of inbreeding depression

1.4.1 Whole genome estimation of inbreeding depression

Inbreeding depression is usually estimated by regressing the phenotypic value of an individual on its pedigree-based inbreeding coefficient. By this method it has been shown that inbreeding affects many economically important traits of dairy cattle. For example, McParland et al. (2007) reported that 12.5% of inbred Irish Holstein-Friesians cows had calving interval increased by 8.8 days and milk, fat and protein yields reduced by 61.8, 5.3 and 1.2 kg, respectively, compared to non-inbred cows. Pryce et al. (2014) found an increase of 0.18 days in calving interval, a decrease of 21.1 litres for milk yield and a decrease of 0.73 and 0.63 kg for fat and protein yields, respectively, per 1% increase in FPED in Holstein cattle.

As explained in Chapter 1.3.1, FPED is an error-prone, genome-wide estimate of inbreeding that does not account for the variation in the levels of inbreeding between individuals with the same pedigree. This results in limited ability to estimate inbreeding depression, because the low variation of FPED impairs the ability to detect inbreeding depression by regression analysis (Keller et al., 2011). Thus, genomic alternatives to FPED are expected to serve as more precise estimates of inbreeding depression.

Keller et al. (2011) showed that FROH outperforms both FPED and genomic SNP-by-SNP estimates of inbreeding (e.g. FHOM) in the detection of inbreeding depression. Furthermore, Szpiech et al. (2013) found the frequency of detrimental homozygous variants in ROH to be considerably higher than the corresponding frequency of neutral homozygotes. The effect of increased FROH on inbreeding depression in dairy cattle is well documented. Bjelland et al. (2013) observed an increase of 1.72 days in days open and a decrease of 20kg of 205-day milk yield per 1% increase in FROH in Holstein cattle. Doekes et al. (2019) reported an increase of 0.48 days in calving interval, and a decrease of 36.25, 1.34 and 1.20 kg in 305-day milk, fat and protein yields, respectively, per 1% increase in FROH in Dutch Holstein-Friesian cattle.

1.4.2 Genomic regions contributing to inbreeding depression

The effect of inbreeding can vary between individuals with the same whole genome F due to homozygosity being more detrimental at some regions of the genome than others (Howard et al., 2017a). Furthermore, lethal recessive mutations with large contribution to inbreeding depression can only be identified by the absence of homozygosity in certain regions (VanRaden et al., 2011). Therefore, estimating inbreeding depression separately for different chromosomal segments can provide more insight into the effects of inbreeding compared to whole genome estimates.

Recently, the dissection of inbreeding depression has been performed using a variety of methods. One of the approaches used is fragmentation of the genome into
segments and regression of the phenotypic values on inbreeding coefficients of the segments. Using this method, Saura et al. (2015) located a region between 27 and 55 Mb from chromosome 13 of Iberian pigs associated with a reduction in the number of piglets born alive and the total number of piglets born. Another approach utilizes genome wide association analysis (GWAS) added to the ROH status of an SNP (value 1 if SNP is part of ROH, otherwise 0). Results from the study conducted by Pryce et al. (2014) showed a detrimental effect of ROH on calving interval on chromosomes 2, 5, 8, 9, 15 and 24 for the Holstein breed. This approach was upgraded by Howard et al. (2015), who presented a machine-learning tree based regression algorithm and detected inbreeding depression on milk production traits from chromosomes 13, 23 and 25 of US Jersey cows and from chromosomes on 3, 7 and 17 of Australian Jersey cows.

As pointed out by Howard et al. (2017a), the effects of inbreeding depression are expected to be population specific due to differences in genetic properties (e.g. allele frequencies) between populations. Because the information pertaining to unfavourable effects of inbreeding found in other populations cannot be directly exploited in Finnish Ayrshire cattle, estimation of the levels of end effects of inbreeding in this population may provide new information for the breeding programme to more efficiently balance between genetic progress and the negative consequences of inbreeding.
2 OBJECTIVES

The main goal of this thesis was to estimate the effect of inbreeding on female fertility in the Finnish Ayrshire population. The goal included the investigation of various estimates of the inbreeding coefficient, determining the associations of the estimates and inbreeding depression and dissecting the associations down to genomic regions and specific genotypes.

The main objective was divided into three more specific goals (with the article number given in parenthesis):

1) To estimate inbreeding coefficients from pedigree and genomic data, to compare the estimates and to use them to estimate inbreeding depression on female fertility traits in Finnish Ayrshire cattle (I).

2) To detect genomic regions responsible for inbreeding depression on female fertility traits using ROH-based measurement of inbreeding coefficient and to examine identified regions in more detail using varying lengths of haplotypes (II).

3) To identify specific ROH genotypes with detrimental effects on female fertility and milk production traits (III).
3 MATERIALS AND METHODS

3.1 Materials

The Nordic Red dairy cattle (RDC) population consists of the Finnish Ayrshire breed and the Scandinavian red breeds. In this thesis, all analyses are based on the Finnish Ayrshire (FAY) dairy cattle population. The original data included records for 38,265 RDC cows. FAY was defined by subsetting cows with the NAV country code 12 (Finland) from the full RDC data, which resulted in records for 19,075 FAY cows. All cows in the subsetted data were born between 2002 and 2014.

Data used in this thesis contained four different datasets: pedigree data (I - III), genotypic data (I - III), phenotypes of female fertility traits (I - III) and deregressed proofs of milk production traits for first lactation (III). Datasets and all supporting data (e.g. solutions for systematic effects related to fertility traits) were obtained from NAV (Aarhus, Denmark) and from Faba, The Finnish Animal Breeding Association (Hollola, Finland).

3.1.1 Pedigree data

Original pedigree data consisted of 2,482,163 records. Average pedigree depth of FAY was 10 generations, varying from 3 to 12 generations. Prior to statistical analyses of inbreeding depression (I, II), pedigree data was pruned using the Relax2 program (Strandén and Vuori, 2006) with the varcomp option, which prunes out animals that do not contribute to the variance component estimation by animal model. After pruning, pedigree data included information from 85,285 individuals. In Study III, generation limit 4 (number of ancestral generations for animals with records) was used in the pruning of pedigree using the varcomp procedure in the Relax2 program (Strandén and Vuori, 2006), resulting in 44,547 pedigree records.

3.1.2 Genotypic data

Genotypic data included 46,914 SNP and was available for 19,075 FAY cows. It was originally generated using Illumina BovineLDv.2 BeadChip low-density panel (Illumina Inc., 2015) that contains 7,931 SNP. Genotypes were then imputed to 50K density using Fimpute software (Sargolzaei et al., 2014) with default settings and a reference population of Nordic Red AI-bulls with 50K genotypes before genotypic data was obtained for this study. The SNP positions were based on Illumina’s assembly (I, II). Illumina’s assembly refers to SNP positions obtained from the panel producer, but as reported by Boichard et al. (2012), UMD3 assembly was used when selecting SNPs to the Illumina BovineLD BeadChip. For Study III the SNP positions
were updated to the UMD 3.1 assembly. Updating changed the position of 1103 SNP and removed 225 SNP that did not have positions in UMD3.1 assembly.

3.1.3 Quality control and pruning of SNP

As the genotypic data used in this thesis consisted of imputed genotypes, both SNP and animal call rates were greater than 0.99 and no pruning based on call rate was done. SNP with minor allele frequency (MAF) less than 0.05 or p-value of the chi-square test for Hardy-Weinberg equilibrium (HWE) of less than 0.0001 were pruned out prior to the analysis (I). After quality control, the number of SNP remaining to be used in the analyses was 39,144 (I).

Next, SNP with MAF less than 0.05 were removed from the genotypic data (II). No pruning for HWE was performed, as inbreeding is one possible cause for the deviation from HWE. After quality control, the number of SNP remaining to be used in the analyses was 40,554 (II).

Lastly, no pruning for HWE was performed and the limit for MAF was lowered to 0.01 (III). Lowering the limit for MAF was based on a suggestion by Hillestad et al. (2018), who reported that strict criteria for including SNP in ROH analyses may result in an inability to detect all ROH. However, SNP were pruned for linkage disequilibrium (LD), as the Unfavorable Haplotype Finder software (Howard et al., 2017b) used in ROH analyses did not allow for the setting of minimum length for segments to be accepted as ROH. As local high LD may result in accepting non-autozygous segments as ROH (Howrigan et al., 2011), pruning of SNP for LD was accomplished using PLINK version 1.9 (Chang et al., 2015; Purcell and Chang, 2015). In LD pruning, SNP with a variance inflation factor (VIF) greater than 10 (corresponding to r^2 > 0.9) within a 50 SNP window were excluded from the data. After quality control, the number of SNP remaining to be used in the analyses was 29,227 (III).

3.1.4 Phenotypes of female fertility traits

Raw phenotypes of female fertility were available for 1,805,454 cows. Data included observations for four female fertility traits: non-return rate at 56 days after first insemination (NRR), number of inseminations per conception (AIS), interval (in days) from calving to first insemination (ICF) and interval from first to last insemination (IFL). Fertility traits were considered separately for heifers (parity 0), and for first-, second- and third-parity cows (parities 1 to 3). The raw phenotypic values were pre-corrected for the main systematic effects prior to statistical analyses using the solutions from the full Nordic fertility evaluation model. These solutions were used because several systematic effect classes contained only one or a few observations in this study. The systematic effects included herd-birth year (for heifers) or herd-year of first calving (for cows), insemination year-month (for NRR, AIS and IFL), calving year-
month (for ICF) and heifer’s age at first insemination. The dataset of fertility phenotypes used throughout this thesis was comprised of these pre-corrected phenotypes. Figure 1 shows the variation of the pre-corrected fertility phenotypes.

![Box plots of pre-corrected phenotypic values](image)

**Figure 1.** Box plots of pre-corrected phenotypic values of non-return rate, number of inseminations, interval from calving to first insemination, interval from first to last insemination and deregressed proofs of milk, protein and fat yields

### 3.1.5 Deregressed proofs of milk production traits

Deregressed proofs (DRP) of the first lactation of milk production traits included milk yield (MILK), protein yield (PROT) and fat yield (FAT) expressed in kilograms (III). DRPs were calculated based on the estimated breeding values (EBV) of milk, protein and fat yield. EBVs were obtained from the routine NAV milk production test day data. DRPs with reliability $r^2_{\text{DRP}} < 40\%$ were pruned from the data. The variation of the milk production trait DRPs are shown in Figure 1.
3.1.6 Merging datasets

For the estimation of inbreeding depression, genotypic and phenotypic datasets were combined and only cows with both genotypic and phenotypic records were kept. The total number of animals with genotypes and records of fertility traits was 13,712. The total number of animals with genotypes and DRPs for milk production traits was 12,233. The number of observations per trait is presented in Table 1.

Table 1. Number of observations per trait for FAY cows with phenotypic and genotypic records

<table>
<thead>
<tr>
<th>Trait</th>
<th>Parity 0 (heifers)</th>
<th>Parity 1</th>
<th>Parity 2</th>
<th>Parity 3</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRR</td>
<td>13 368</td>
<td>9 474</td>
<td>5 043</td>
<td>1 540</td>
<td>I, II</td>
</tr>
<tr>
<td>AIS</td>
<td>13 261</td>
<td>9 323</td>
<td>4 918</td>
<td>1 453</td>
<td>I, III</td>
</tr>
<tr>
<td>ICF</td>
<td>9 453</td>
<td>5 067</td>
<td>1 509</td>
<td>I, II, III</td>
<td></td>
</tr>
<tr>
<td>IFL</td>
<td>12 878</td>
<td>9 546</td>
<td>5 131</td>
<td>1 567</td>
<td>I, II, III</td>
</tr>
<tr>
<td>MILK</td>
<td></td>
<td>12 233</td>
<td></td>
<td></td>
<td>III</td>
</tr>
<tr>
<td>PROT</td>
<td></td>
<td>12 233</td>
<td></td>
<td></td>
<td>III</td>
</tr>
<tr>
<td>FAT</td>
<td></td>
<td>12 233</td>
<td></td>
<td></td>
<td>III</td>
</tr>
</tbody>
</table>

3.2 Methods

3.2.1 Estimation of inbreeding coefficients

Five different methods were used for the estimation of inbreeding coefficients (I): pedigree-based inbreeding coefficient (FPED), percent of homozygous SNP (FPH) and three estimates based on runs of homozygosity (FROH_1, FROH_2 and FROH_3). FPED was estimated only for animals with a pedigree completeness value (MacCluer et al., 1983) of at least 0.80 based on five generations.

The first genomic measure of inbreeding, FPH, represented the proportion of homozygous genotypes. It was derived using the following formula:

\[
F_{PH} = \frac{n_{AA} + n_{BB}}{n_{AA} + n_{AB} + n_{BB}}
\]  

(1),

where \(n_{AA}\), \(n_{AB}\) and \(n_{BB}\) represent the number of SNPs with the genotypes AA, AB and BB, respectively.

The other genomic measures of inbreeding were based on ROH detected from autosomal chromosomes with three different criteria in PLINK v.1.07 (Purcell et al.,
The criteria for ROH_1, ROH_2 and ROH_3 are presented in Table 2. Following ROH detection, inbreeding coefficients based on ROH_1, ROH_2 and ROH_3 were derived for each animal using the following formula:

\[ F_{ROH,1/2/3} = \frac{\sum ROH_{SNP,1/2/3}}{\sum SNP} \]  

(2),

where \( ROH_{SNP,1/2/3} \) refers to SNPs in ROHs detected using either ROH_1, ROH_2 or ROH_3 criteria, and \( SNP \) refers to all SNPs.

### Table 2. Criteria for ROH detection

<table>
<thead>
<tr>
<th>Criteria</th>
<th>ROH_1</th>
<th>ROH_2</th>
<th>ROH_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>-- homozyg-density(^1)</td>
<td>1 SNP / 120 kb</td>
<td>1 SNP / 1000 kb</td>
<td>1 SNP / 1000 kb</td>
</tr>
<tr>
<td>-- homozyg-kb(^2)</td>
<td>500 kb</td>
<td>10 kb</td>
<td>10 kb</td>
</tr>
<tr>
<td>-- homozyg-snp(^3)</td>
<td>0</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>-- homozyg-gap(^4)</td>
<td>1000 kb</td>
<td>1000 kb</td>
<td>1000 kb</td>
</tr>
<tr>
<td>-- homozyg-window-snp(^5)</td>
<td>50 SNP</td>
<td>20 SNP</td>
<td>20 SNP</td>
</tr>
<tr>
<td>-- homozyg-window-het(^6)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>-- homozyg-window-threshold(^7)</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

\(^1\)Minimum density of SNP in ROH  
\(^2\)Minimum length of ROH  
\(^3\)Minimum number of SNP in ROH  
\(^4\)Maximum gap between two SNP in ROH  
\(^5\)Size of sliding window  
\(^6\)Number of heterozygous SNP allowed in a window  
\(^7\)SNP is accepted as part of ROH using this threshold for the proportion of overlapping homozygous windows containing that SNP

Estimation of inbreeding coefficients was performed for each autosomal chromosome as well as for different regions within the autosomal chromosomes (II). These inbreeding coefficients were based on ROHs detected with ROH_1 criteria (Table 2). Chromosomal inbreeding coefficients for each autosomal chromosome \( K \) were derived using the following formula:

\[ F_{ROH,K} = \frac{\sum ROH_{SNP,K}}{\sum SNP_K} \]  

(3),

where \( ROH_{SNP,K} \) refers to the SNPs in ROHs in chromosome \( K \), and \( SNP_K \) refers to all SNPs in chromosome \( K \). For regions within autosomal chromosomes, a sliding
window method was used. Inbreeding coefficient was estimated for window W using the following formula:

\[ F_{ROH,W} = \frac{\sum ROH\_SNP_W}{\sum SNP_W} \]  

(4),

where \( ROH\_SNP_W \) refers to the SNPs in ROHs in window W and \( SNP_W \) refers to all SNPs in window W.

3.2.2 Estimation of inbreeding depression on fertility

Inbreeding depression was estimated by regressing the phenotypic values of fertility traits on the estimated inbreeding coefficients (I: \( F_{PED}, F_{PH} \) or \( F_{ROH,1} \), II: \( F_{ROH,K} \) or \( F_{ROH,W} \)). Estimation was performed separately for each of the studied traits (I: NRR, AIS, ICF and IFL, II: NRR, ICF and IFL) with the multi-trait model, where heifer and cow traits are estimated jointly:

\[ y_{ij} = \mu + bF_{X,i} + a_i + e_i \]  

(5),

where \( y_{ij} \) is the pre-corrected fertility phenotype of animal i, \( \mu \) is the general mean, \( b \) is the estimate of inbreeding depression, \( F_{X} \) is the inbreeding coefficient (\( F_{PED}, F_{PH} \) or \( F_{ROH,1} \) in Study I and \( F_{ROH,K} \) or \( F_{ROH,W} \) in Study II) of animal i, \( a_i \) is the additive genetic effect of animal i and \( e_i \) is the residual effect of animal i. Furthermore, it was assumed that additive genetic effects were normally distributed with \( N(0, A \otimes G) \) and residual effects were normally distributed with \( N(0, R) \), where \( A \) is the pedigree-based additive relationship matrix, \( G \) is the additive genetic variance-covariance matrix between heifer and cow traits and \( R \) is the residual variance-covariance matrix between heifer and cow traits. The variance components used in the analyses were derived from the Nordic fertility evaluation (Muuttoranta et al., 2019). Genetic groups were used for animals with unknown parents and treated as random effects in the analyses. Statistical analyses of inbreeding depression were conducted using the DMU program package (Madsen and Jensen, 2013).

3.2.3 Identification of haplotypes associated with reduced fertility

Chromosomal regions showing inbreeding depression on female fertility when \( F_{ROH,W} \) was used as a covariate in the model were selected for haplotype analysis (II). First, all haplotypes within the regions with lengths of 1 to 10 SNP were screened. The association of the haplotypes in a homozygous state with reduced fertility was tested using a simple linear model without the additive genetic effect as:

\[ y_{ij} = \mu + g_{ij} + e_i \]  

(6),
where $y_{ij}$ is the pre-corrected fertility phenotype of animal i, $\mu$ is the general mean, $g_{ij}$ is either 1 or 0 depending whether the animal i is homozygous for the haplotype j or not and $e_i$ is the residual effect of animal i. Next, based on the initial screening of haplotypes, from each region only the haplotype with the smallest p-value and homozygous frequency greater than 0.05 was selected to be validated using the full model:

$$y_{ij} = \mu + g_{ij} + a_i + e_i$$

(7),

where $a_i$ is the additive genetic effect of animal i and $y_{ij}$, $\mu$ and $e_i$ are the same as in formula 6. Observation of inbreeding depression requires that the trait is affected by dominance (Howard et al., 2017a). Therefore, in the validation step, both recessive ($g_{ij}$ in formula 7 equals to 1 if animal i is homozygous for haplotype j, otherwise 0) and additive ($g_{ij}$ in formula 7 equals to 0, 1 or 2 depending on the number of copies of the haplotype j on animal i) modes of inheritance were tested.

3.2.4 Identification of ROH genotypes associated with reduced fertility and milk production traits

The detrimental effect of ROH was estimated separately for each different ROH segment (III). In addition to female fertility traits, milk production traits (MILK, PROT and FAT) were considered. Identification of the ROHs associated with impaired fertility or milk production traits was performed using the Unfavorable Haplotype Finder software (Howard et al., 2017b). A detailed description of the algorithm of the software is presented in Howard et al. (2017b).

Briefly, the first step is to scan the genome using a sliding window method to detect ROH genotypes that are associated with an unfavourable effect on the phenotype. A genotype was accepted as ROH and placed into the ROH category if it did not contain any heterozygous SNPs and had a frequency at least 0.01. Genotypes that did not meet these criteria were placed into the non-ROH category. Genotypes in the ROH category were considered unfavourable if the mean phenotypic value of the ROH genotype was above (fertility traits) or below (milk production traits) a certain cutoff value. The cutoff values of each trait were determined by generating a t-statistic distribution based on 1000 permutations for random regions in the genome. The mean phenotype of the samples with a significance ranking from 0.05 to 0.10 was selected as the cutoff value.

Since ROHs from the same origin may slightly differ from each other due to recombination events, the algorithm compromises multiple aggregation steps to detect the core genotypes of each ROH. These core ROH genotypes then serve as tags for the full ROHs. The core genotypes were detected using a window of decreasing size from 50 SNPs to 15 SNPs at 5-SNP intervals.

In step 2, the significance of the ROH genotypes that were considered unfavourable in step 1 is tested using the full model that includes pedigree information. In step 2,
one-sided t-tests are performed between animals with unique ROH genotypes and animals with no ROH in that window. Because each region may have several unique ROH genotypes, each set of cows with unique ROH genotypes was tested independently against a set of cows with no ROH in that window. The following linear mixed model was used to test the significance of ROH genotypes in step 2:

\[ y_{ij} = \mu + ROH_{ij} + a_i + e_i \]  \hspace{1cm} (8),

where \( y_{ij} \) is the pre-corrected phenotype (fertility traits) or DRP (milk production traits) of animal \( i \), \( \mu \) is the general mean, \( ROH_{ij} \) is a vector including the effect of ROH\(_j\) genotype of animal \( i \), \( a_i \) is the additive genetic effect of animal \( i \) and \( e_i \) is the residual effect of animal \( i \).

Within the fertility traits, heifer and cow traits were analysed separately in contrast to formula 5 presented in Chapter 3.2.3. Both the additive genetic and residual effects were assumed to be normally distributed with \( N(0, A\sigma^2_a) \) and \( N(0, I\sigma^2_e) \), where \( A \) is the pedigree-based additive relationship matrix, \( \sigma^2_a \) the additive genetic variance, \( I \) is an identity matrix and \( \sigma^2_e \) is the residual variance.

The variance components of the production traits were estimated using the Average Information Restricted Maximum Likelihood (AI-REML) method with the DMU software (Madsen and Jensen, 2013).
4 RESULTS AND DISCUSSION

This thesis consisted of three studies that focused on the effects of inbreeding on female fertility traits in the Finnish Ayrshire (FAY) population using genomic information. The studies included estimation of inbreeding coefficient using pedigree-based and genomic approaches (I), comparison of the different estimates of inbreeding (I), estimation of whole genome (I) and regional (II) inbreeding depression on female fertility traits, identification of homozygous haplotypes associated with inbreeding depression on fertility traits (II) and identification of specific ROH genotypes associated with inbreeding depression on fertility and, additionally, on milk production traits (III).

4.1 Runs of homozygosity

Detection of ROHs was performed for all 19 075 genotyped FAY cows using PLINK software (Purcell et al., 2007) and three different criteria for defining ROHs (ROH_1, ROH2 and ROH_3) (I). The total number of ROHs detected using ROH_1, ROH_2 and ROH_3 criteria were 411 541, 838 383 and 165 843 ROHs, respectively (Table 3). The average number of ROHs per cow was 22 (varying from 0 to 50), 44 (varying from 6 to 82) and 9 (varying from 0 to 33) with ROH_1, ROH_2 and ROH_3, respectively.

Table 3 presents the frequency distribution of all ROHs obtained with parameter setting ROH_1, ROH_2 and ROH_3. ROH lengths varied between 0.5 - 125.1 Mb with ROH_1, 0.9 - 138 Mb with ROH_2 and 3.3 - 138 Mb with ROH_3. The number of SNPs in ROH varied from 6 to 1897 SNP with ROH_1, from 30 to 2123 SNP with ROH_2 and from 100 to 2123 SNP with ROH_3.

<table>
<thead>
<tr>
<th>ROH length</th>
<th>Number of ROH_1</th>
<th>Number of ROH_2</th>
<th>Number of ROH_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 5 Mb</td>
<td>198 595 (48.3%)</td>
<td>602486 (71.9%)</td>
<td>1 345 (0.8%)</td>
</tr>
<tr>
<td>5 – 10 Mb</td>
<td>140 677 (34.2%)</td>
<td>158924 (19.0%)</td>
<td>87 719 (52.9%)</td>
</tr>
<tr>
<td>10 – 20 Mb</td>
<td>55 202 (13.4%)</td>
<td>58887 (7.0%)</td>
<td>58 693 (35.4%)</td>
</tr>
<tr>
<td>20 – 30 Mb</td>
<td>11 412 (2.7%)</td>
<td>12131 (1.4%)</td>
<td>12 131 (7.3%)</td>
</tr>
<tr>
<td>30 – 40 Mb</td>
<td>3 637 (0.9%)</td>
<td>3 832 (0.4%)</td>
<td>3 832 (2.3%)</td>
</tr>
<tr>
<td>40 – 50 Mb</td>
<td>1 235 (0.3%)</td>
<td>1 285 (0.2%)</td>
<td>1 285 (0.8%)</td>
</tr>
<tr>
<td>&gt; 50 Mb</td>
<td>783 (0.2%)</td>
<td>838 (0.1%)</td>
<td>838 (0.5%)</td>
</tr>
<tr>
<td>Total number of ROH</td>
<td>411 541</td>
<td>838 383</td>
<td>165 843</td>
</tr>
</tbody>
</table>
The criteria used to accept ROHs has an effect on the number and length of ROHs found, which was demonstrated in this study. Changing the minimum required length of ROH can be compared to changing the depth of pedigree in pedigree-based analysis because long ROH are present due to recent inbreeding, while short ROH display older inbreeding (Forutan et al., 2018). As reported by Fisher (1954), the expected length of ROH follows an exponential distribution with mean 100/2g centiMorgan, where g is the number of generations to the common ancestor. Using the average conversion by Arias et al. (2009), where 1 Mb equals to 1.25 centiMorgan, ROH_1 with minimum ROH length of 0.5 Mb could capture inbreeding arising from 80 generations ago. Similarly, for ROH_2 and ROH_3 with the minimum (observed) ROH lengths of 0.9 Mb and 3.3 Mb the captured inbreeding could have occurred 44 and 12 generations ago, respectively. However, it is not likely that the ROHs of different length detected in this study merely represent the number of generations from which ROHs are included in the analysis. As pointed out by Howrigan et al. (2011), incorrect setting of the parameters to derive ROH may result in accepting IBS segments as ROHs. Moreover, the rate of recombination was not accounted for in this study. Therefore, an ROH of a certain length on a low recombination rate region of the genome reflects more ancient inbreeding than an ROH of the same length on a high recombination rate region (Gomez-Raya et al., 2019).

The observed number of ROHs was greatest for ROH_2, with twice the number of ROHs than for ROH_1. This can be explained by the difference in the number of short ROHs between ROH_1 and ROH_2. The numbers of ROHs with less than 50 SNP (around 3 Mb) were 24 427 (5.9%) and 400 015 (47.7%), for ROH_1 and ROH_2, respectively. Because the minimum ROH length was longer for ROH_2 than for ROH_1, the difference in the number of ROHs must be explained by parameters other than the minimum required ROH length.

ROH_1 and ROH_2 differed by the required minimum density (1 SNP per 120 kb and 1 SNP per 1 Mb, respectively) and by the sliding window size (50 SNP and 20 SNP, respectively). The low required density may falsify ROH length (Purfield et al., 2012). In addition, the small window size may result in overestimation of short ROHs. The effect of the small window size was reported by Forutan et al. (2018), who observed an overestimation of FROH with a window size of 5 SNPs possibly due to a large number of short IBS segments accepted as ROHs. However, a window size of 10 to 25 SNP was not associated with the overestimation of FROH (Liu et al., 2014, Forutan et al., 2018). Moreover, the --homozyg-window-threshold parameters may have contributed to the overestimation of ROHs in this study, especially when ROH_2 criteria were used. In a 20 SNP window, each SNP has 20 possibilities to belong to a homozygous (one heterozygous SNP allowed) window. The 5% threshold for the proportion of homozygous windows with a 20 SNP window size results in a calling of ROH when an SNP belongs only to one homozygous window (provided other criteria for ROH are met). When one heterozygous SNP is allowed per window, the 5% threshold for a 20 SNP window size may result in the first or last SNP or ROH being heterozygous. For a 50 SNP window, an SNP must belong to 2.5 windows (rounded
to 3 windows) in order to be accepted into ROH, and therefore the first or last SNP of ROH cannot be heterozygous. It can be concluded that a large proportion of ROHs detected using ROH_2 is most likely due to a combined effect of accepting ROHs containing approximately 5% (1/20) heterozygous SNPs from sparse regions with low homozygous window threshold, thus including short ROHs that are not IBD.

The criteria in ROH_3 analysis was otherwise the same as in ROH_2 analysis, but the minimum number of SNPs required in ROH was set as 100, which resulted in a five times smaller number of ROHs than what was observed with ROH_2 criteria. Albeit the 100 SNP minimum ROH length in ROH_3 prevented the formation of short segments, the low density criteria may have resulted in a creation of incorrectly long ROHs. Furthermore, the 100 SNP minimum length of ROH likely resulted in underestimation of short ROHs.

In the ROH_1 analysis a minimum length of ROH was set at 500 kb and no restriction was placed for the number of SNP in ROH, which may have led to acceptance of some non-IBD segments as ROHs. An approach by Lencz et al. (2007) applied by Purfield et al. (2012) determines the minimum number of SNP in ROH according to SNP density of the genotypic data. Following this method, the minimum number of SNP required for calling an ROH would have been 50 in this study. As mentioned above, only 5.9% of SNPs detected using ROH_1 criteria were shorter than 50 SNP, suggesting that ROH_1 succeeded in capturing short ROHs while excluding non-IBD ROHs better than ROH_2 and ROH_3 did.

The effect of quality control of SNPs also impacts the detection of ROHs, as the algorithm of PLINK does not account for MAF, HWE or LD (Howrigan et al., 2011). However, because the quality control was performed prior to the ROH detection, its effect is expected to be similar across all three ROH analyses. Removing low MAF SNPs (< 0.05) was done based on the suggestion by Howrigan et al. (2011), who noted that many low MAF SNPs in a row can result in the acceptance of non-IBD segments as ROHs. Common practice in genome-wide association studies is to remove SNPs that deviate extensively from HWE because the deviation may indicate genotyping errors (Anderson et al., 2010). Pruning of SNPs deviating from HWE (< 0.0001) was performed in Study I, but not Study II because the deviation can be due inbreeding (Cox and Kraft, 2006). When otherwise identical ROH_1 criteria were used, 15 212 more ROHs were detected for data that was not pruned for HWE (II). No SNPs were pruned based on LD, but the short non-IBD ROHs that occur due to high LD within dense SNP regions were excluded by setting the minimum ROH length to 500 kb (ROH_1) or above, as suggested by Purfield et al. (2012). Ferenčaković et al. (2013a), however, pointed out that due to long range LD in cattle, a minimum length of 1 Mb should be used in order to exclude ROHs that occur due to LD.

In addition to quality control of SNP, the program used also influences ROH detection. ROHs were detected using PLINK software (Purcell et al., 2007), a commonly used program that has been shown to perform well in ROH analysis (Howrigan et al., 2011; Ceballos et al., 2018). Ferenčaković et al. (2013b) noted that the sliding window approach used in PLINK may produce artificial ROHs that are not
IBD. A simulation study by Forutan et al. (2018) reported underestimation of \( F_{\text{ROH}} \) using PLINK. The authors expected this to be a result of the fixed window approach used by PLINK, and suggested that overlapping windows of decreasing size could more efficiently identify true IBD ROHs.

Furthermore, SNP sampling has an effect on ROH analysis. Ferenčaković et al. (2013b) reported that the medium density panels overestimate the number of ROHs shorter than 4 Mb. In the present study, 31.8%, 61.5% and 0.03% of all ROHs observed were shorter than 4 Mb in ROH_1, ROH_2 and ROH_3 analyses, respectively. The results indicate that both ROH_1 and ROH_2 have resulted in overestimation of short ROHs in the present study. In this study, the genotypes were originally generated with low density panels and imputed to 50K density. However, imputation accuracy over 97% using Illumina BovineLD genotyping panel for BovineSNP50 genotypes was reported by Boichard et al. (2012). A simulation study by Zhang et al. (2018) showed that ROHs detected in PLINK using imputed data with an imputation accuracy of 96.8% had a correlation over 0.9 with ROHs detected using real data. Therefore, the overall effect of imputation on ROH detection in the present study was presumably small. In addition, the allowance of one heterozygous SNP per window was expected to reduce the number of undetected ROHs due to real homozygous genotypes imputed as heterozygous. On the other hand, this may have created some short non-IBD ROHs due to real heterozygous genotypes imputed as homozygous.

ROH detection in Study III was somewhat different to ROH detection in Studies I and II. First, the Unfavorable Haplotype Finder (Howard et al., 2017b) program used in Study III did not allow the setting of a minimum length for ROH, which is why pruning for LD was performed in order to prevent spurious ROH calls due to high LD. Second, the MAF threshold was lowered to 0.01 based on the suggestion of Hillestad et al. (2018), who found that including low MAF SNP is important for avoiding the loss of short ROHs or the splitting of longer ROHs. Third, no heterozygous genotype was allowed within ROH, which may have reduced the calling of short non-IBD ROHs originating from imputation errors compared to Studies I and II (Hillestad et al., 2018). However, as the program does not output full ROH segments but only the core genotypes of full ROHs, no comparison of ROH results between the two programs used for ROH analyses in this thesis was made.

4.2 Inbreeding coefficients

Level of inbreeding in the Finnish Ayrshire population was estimated using information from two different sources: pedigree and imputed 50K SNP genotype data (I). There were five estimates of inbreeding coefficients \( F_{\text{PED}}, F_{\text{ROH}_1}, F_{\text{ROH}_2}, F_{\text{ROH}_3} \) and \( F_{\text{PHI}} \) for each animal. The results of the 13 712 FAY with both genotypic and phenotypic records are presented in Table 4: the average inbreeding coefficients of
FAY were 0.02, 0.06, 0.09, 0.04 and 0.63 using \( \text{FPED}, \text{FROH}_1, \text{FROH}_2, \text{FROH}_3 \) and \( \text{FPH} \), respectively.

**Table 4.** Estimated mean, standard deviation (SD) and range of values with different measures of inbreeding for 13 712 FAY with genotypic and phenotypic data (I).

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{FPED} )</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00 – 0.29</td>
</tr>
<tr>
<td>( \text{FROH}_1 )</td>
<td>0.06</td>
<td>0.03</td>
<td>0.00 – 0.28</td>
</tr>
<tr>
<td>( \text{FROH}_2 )</td>
<td>0.09</td>
<td>0.03</td>
<td>0.008 – 0.30</td>
</tr>
<tr>
<td>( \text{FROH}_3 )</td>
<td>0.04</td>
<td>0.02</td>
<td>0.00 – 0.27</td>
</tr>
<tr>
<td>( \text{FPH} )</td>
<td>0.63</td>
<td>0.01</td>
<td>0.60 – 0.71</td>
</tr>
</tbody>
</table>

\( \text{FPED} \): pedigree-based inbreeding coefficient, \( \text{FROH} \): ROH-based inbreeding coefficients (using criteria ROH_1, ROH_2 or ROH_3), \( \text{FPH} \): percent of homozygous SNP –based inbreeding coefficient

\( \text{FPED} \) resulted in the lowest average inbreeding coefficient, but the range of \( \text{FPED} \) estimates was similar to the range of \( \text{FROH} \) estimates (Table 4). Low variation was observed within \( \text{FPED} \) values (Table 4). A majority (95%) of the \( \text{FPED} \) values fell at the lower end of the range of values (\( \text{FPED}<0.05 \)) (I). Similar low variation was also observed for \( \text{FPH} \) (Table 4). A majority (97%) of the \( \text{FPH} \) values were lower than 0.65 (data not shown).

Average \( \text{FROH} \) values followed the number of observed ROHs, as the lowest average estimate was observed for \( \text{FROH}_3 \) and the highest for \( \text{FROH}_2 \) (Table 4). Variation within \( \text{FROH} \) estimates was larger than what was observed for \( \text{FPED} \) and \( \text{FPH} \) (Table 4).

### 4.3 Correlations between different measures of inbreeding

Correlations between the ROH-based measures of inbreeding were high, with correlations between \( \text{FROH}_1 \) and \( \text{FROH}_2 \) of 0.98, \( \text{FROH}_1 \) and \( \text{FROH}_3 \) of 0.96 and \( \text{FROH}_2 \) and \( \text{FROH}_3 \) of 0.92 (Table 5). Another genomic measure of inbreeding, \( \text{FPH} \), also had large correlations with ROH-based estimates: 0.94, 0.95 and 0.90 between \( \text{FPH} \) and \( \text{FROH}_1 \), \( \text{FROH}_2 \) or \( \text{FROH}_3 \), respectively (Table 5). Pedigree-based measure of inbreeding had more modest correlations to ROH-based estimates (around 0.70) and to \( \text{FPH} \) (0.66) (Table 5).
Table 5. Correlations coefficient between different estimates of inbreeding for 13,712 FAY with genotypic and phenotypic data (I).

<table>
<thead>
<tr>
<th></th>
<th>$F_{PED}$</th>
<th>$F_{PED}$</th>
<th>$F_{ROH_1}$</th>
<th>$F_{ROH_2}$</th>
<th>$F_{ROH_3}$</th>
<th>$F_{PH}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{PED}$</td>
<td><strong>1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F_{ROH_1}$</td>
<td>0.71</td>
<td><strong>1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F_{ROH_2}$</td>
<td>0.70</td>
<td>0.98</td>
<td><strong>1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F_{ROH_3}$</td>
<td>0.69</td>
<td>0.96</td>
<td>0.92</td>
<td><strong>1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F_{PH}$</td>
<td>0.66</td>
<td>0.94</td>
<td>0.95</td>
<td>0.90</td>
<td><strong>1</strong></td>
<td></td>
</tr>
</tbody>
</table>

$F_{PED}$: pedigree-based inbreeding coefficient, $F_{ROH}$: ROH-based inbreeding coefficients using three different criteria for ROH ($F_{ROH_1}$, $F_{ROH_2}$ and $F_{ROH_3}$), $F_{PH}$: percent of homozygous SNP-based inbreeding coefficient.

Correlations between the different measures of inbreeding observed in this study are in the range of values observed by previous studies. Moderate correlations varying on average from 0.50 to 0.75 have been observed between $F_{PED}$ and $F_{ROH}$ (Purfield et al., 2012; Ferenčaković et al., 2013a; Pryce et al., 2014; Rodríguez-Ramilo et al., 2015; Saura et al., 2015; Zhang et al., 2015a). Moderate correlations are expected to arise from the fact that pedigree can only capture quite recent inbreeding that occurred due to known ancestors in the pedigree, while ROH can capture much more ancient inbreeding. The minimum required ROH length has an effect on the correlation between $F_{PED}$ and $F_{ROH}$. Thus, correlation is higher when only longer ROHs that reflect recent inbreeding are included in $F_{ROH}$ estimation (Zhang et al., 2015a). Zhang et al. (2015a) reported an increased correlation between $F_{PED}$ and $F_{ROH}$ when ROHs shorter than 1 or 3 Mb were excluded. However, greater correlation was not observed in this study for $F_{ROH_3}$ including only ROHs longer than 3.3Mb than correlation for $F_{ROH_1}$ and $F_{ROH_2}$. This can be due to the formation of incorrectly long ROHs and underestimation of short ROHs in ROH_3 analysis, as discussed in Chapter 4.1.

Correlation between $F_{PH}$ and other measures of inbreeding observed in this study followed the same trend as what has been observed previously: strong correlation between genomic measures of inbreeding and notably lower correlations between $F_{PH}$ and $F_{PED}$. Pryce et al. (2014) found correlations of 0.90 and 0.91 between $F_{PH}$ and $F_{ROH}$ in Holstein and Jersey populations, respectively, while Bjelland et al. (2013) observed a correlation of 0.81 in the Holstein population. Pryce et al. (2014) reported lower correlation between $F_{PH}$ and $F_{PED}$ (approximately 0.40) than what was observed in this study. However, the criteria for pedigree depth and completeness used in this study was more stringent than that used in the study by Pryce et al. (2014). This is expected to reduce the number of animals that receive $F_{PED}$ values of zero due to shallow or incomplete pedigree, and therefore should increase the correlation between $F_{PED}$ and genomic measures of inbreeding.
Strong correlations were observed between all ROH-based estimates of inbreeding despite varying criteria used for the calling of ROHs. As discussed in Chapter 4.1, the criteria used to define ROH has a major effect on the number and length of detected ROHs. Although the absolute inbreeding coefficient values varied between $F_{ROH\_1}$, $F_{ROH\_2}$ and $F_{ROH\_3}$, the strong correlation between the estimates shows that the relative values between individuals are very similar. Therefore, only $F_{ROH\_1}$ was selected for the analysis of inbreeding depression in addition to $F_{PED}$ and $F_{PH}$, as it had the highest correlation with $F_{PED}$ and also was expected to represent the true proportion of the genome covered by ROH as discussed in Chapter 4.1.

### 4.4 Whole genome estimates of inbreeding depression on fertility

Inbreeding depression on female fertility traits was estimated by regressing fertility traits (NRR, AIS, ICF and IFL) on pedigree-based ($F_{PED}$) or genomic-based ($F_{ROH\_1}$, $F_{PH}$) inbreeding coefficients ($I$). Table 6 presents the traits that showed statistically significant inbreeding depression. The estimated effect of 1% increase in $F_{ROH\_1}$ was associated with 0.01 more AIS0 and AIS1, 0.43 days longer IFL0 and 0.55 days longer IFL1. A suggestive association (p-value 0.08) was observed between 1% increase in $F_{ROH\_1}$ and 0.28 days longer ICF2. The estimated effect of 1% increase in $F_{PH}$ was associated with 0.02 more AIS0, 0.03 more AIS1, 1.07 percent point decrease in NRR1 and 0.89 days longer IFL0. A suggestive association (p-value 0.08) between 1% increase in $F_{PH}$ and 1.10 days longer IFL1 was also observed.

<table>
<thead>
<tr>
<th>Trait</th>
<th>$F_{PED}$</th>
<th>$F_{ROH_1}$</th>
<th>$F_{PH}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRR1</td>
<td>ns</td>
<td>ns</td>
<td>-1.07 (0.045)</td>
</tr>
<tr>
<td>AIS0</td>
<td>ns</td>
<td>0.01 (0.001)</td>
<td>0.02 (0.02)</td>
</tr>
<tr>
<td>AIS1</td>
<td>ns</td>
<td>0.01 (0.03)</td>
<td>0.03 (0.02)</td>
</tr>
<tr>
<td>ICF2</td>
<td>ns</td>
<td>0.28 (0.08)</td>
<td>ns</td>
</tr>
<tr>
<td>IFL0</td>
<td>ns</td>
<td>0.43 (0.0009)</td>
<td>0.89 (0.004)</td>
</tr>
<tr>
<td>IFL1</td>
<td>ns</td>
<td>0.55 (0.04)</td>
<td>1.10 (0.08)</td>
</tr>
</tbody>
</table>

1NRR: non-return rate (%), AIS: number of inseminations, ICF: interval from calving to first insemination in days, IFL: interval from first to last insemination in days (0: heifer, 1: first-parity cow, 2: second-parity cow)

2$F_{PED}$: pedigree-based inbreeding coefficient, $F_{ROH\_1}$: ROH-based inbreeding coefficients using ROH_1 criteria, $F_{PH}$: percent of homozygous SNP -based inbreeding coefficient
Statistically significant inbreeding depression was observed in this study mainly for the AIS and IFL of heifers and first-parity cows (Table 6). The results may be affected by the fact that the data consisted mostly of records from cows with low levels of inbreeding and that the number of highly inbred cows was small (1). This may reduce the power of linear regression analysis and thus result in a number of non-significant results (Cassell et al., 2003). Additionally, using a linear model may not have been optimal, especially for NRR, which is a discrete trait with a bimodal distribution. For example, the threshold model or the generalized linear mixed model may have produced more reliable estimates of inbreeding depression on NRR. The use of a more sophisticated model may also have benefited the estimation of inbreeding depression on ICF and AIS with substantial skewedness to the right and IFL with zero-inflated distribution. Moreover, the power of regression analysis depends on the number of observations, and fewer observations were available for later parities, which may explain the non-significance of the results.

No inbreeding depression was observed for FPED (Table 6). Low variance in FPED values may have resulted in low power to detect inbreeding depression by regression analysis (Curik et al., 2017). However, increased FPH was associated with decreased fertility even though variance in FPH values was similarly low as in FPED. As noted by Curik et al. (2017), FPED values are mostly low with some individuals having very high values, which makes it difficult to estimate the association between increased FPED and phenotypic values. The range of FPED values (0 - 0.29) observed in this study was wider than the range of FPH values (0.6 – 0.71), which may be one factor affecting the power of inbreeding depression analysis. Moreover, a simulation study by Kardos et al. (2015) reported low precision and downward biased estimates in detection of the true levels of inbreeding with FPED. A potential source for the errors in FPED is incomplete and inaccurate pedigree information (Cassell et al. 2003). As a pedigree completeness value of 0.80 based on five generations was used in this study, a larger error is probably the result of assigning the same FPED values to all individuals with the same pedigree and ignoring the variation in actual IBD (Curik et al., 2017).

The results of this study correspond with the results from previous studies that have obtained small but statistically significant associations between increased genomic inbreeding and fertility traits. For example, Bjelland et al. (2013) found that a 1% increase in FROH and FPH was associated with 1.72 and 1.76 days increase in days open (trait that consists of ICF and IFL), respectively. Doekes et al. (2019) reported an increase of 0.27 days in IFL per 1% increase in FROH. Results for ICF by Doekes et al. (2019) were not significant for any of the estimates of inbreeding (FPED, FROH, FGRM), which was also observed in this study.

The different methods used to estimate inbreeding result in different scales of inbreeding coefficient and thus different estimates of inbreeding depression, which complicates the comparison of the results between and within studies. In this study, the difference in scale between FPH and FROH,1 resulted in higher estimates of inbreeding depression for FPH than for FROH,1 per 1% increase in inbreeding coefficient (Table 6) because a 1% increase in FPH was associated with an average 2.4% increase
in \( F_{ROH,1} \) (data not shown). This means that the 1% increase is much higher for \( F_{PH} \) than for \( F_{ROH,1} \) and the effect for the magnitude of inbreeding depression is likewise larger.

For a more accurate comparison between \( F_{ROH,1} \) and \( F_{PH} \), the difference in phenotype was estimated for cows with low (5th percentile) and high (95th percentile) \( F_{ROH,1} \) and \( F_{PH} \), as suggested by Doekes et al. (2019). High \( F_{ROH,1} \) cows (\( F_{ROH,1} \): 0.11) were estimated to have 0.09 more AIS0, 0.1 more AIS1, 3.7 days longer IFL0 and 4.6 days longer IFL1 than low \( F_{ROH,1} \) (\( F_{ROH,1} \): 0.02) cows. Corresponding results for high \( F_{PH} \) (\( F_{PH} \): 0.65) cows were 0.06 more AIS0, 0.1 more AIS1, 3.0 longer IFL0 and 3.8 days longer IFL1 than low \( F_{PH} \) cows (\( F_{PH} \): 0.61). The results are similar to those observed by Doekes et al. (2019), who reported different estimates on IFL when using \( F_{ROH} \) and \( F_{GRM} \) (0.27 and 0.42 days longer IFL per 1% increase in \( F_{ROH} \) and \( F_{GRM} \), respectively). However, when they reported inbreeding depression expressed as the difference between the 5th percentile and 95th percentile cows, the results were more similar between \( F_{ROH} \) and \( F_{GRM} \) (2.2 and 2.7 days longer IFL for \( F_{ROH} \) and \( F_{GRM} \), respectively).

\( F_{ROH,1} \) estimated the levels of inbreeding by accounting for very distant ancestors. Szpiech et al. (2013) reported that long ROHs are more enriched by detrimental recessive alleles than short ROHs in the human genome, because detrimental recessive alleles are removed by selection and purging over many generations. In this study, individuals with the same \( F_{ROH} \) value may have had different proportion of short and long ROHs. Accounting for only long ROHs in the estimation of \( F_{ROH} \) may have increased the power of the inbreeding depression analysis. However, a large part of inbreeding depression is assumed to be the result of many rare detrimental recessive alleles with small effects (Charlesworth and Willis, 2009). Because purging is inefficient against the rare recessive alleles, including the effect of the distant ancestors is important in order to obtain full insight into inbreeding depression (Kardos et al., 2015).

### 4.5 Intrachromosomal estimates of inbreeding depression on fertility

A large variation in the patterns of inbreeding has been observed in cattle (e.g. Kleinman-Ruiz et al., 2016; Goszczynski et al., 2018; Peripolli et al., 2018). Weak correlations between whole genome estimates and region-specific estimates of inbreeding imply that the contribution of the whole genome estimates to inbreeding depression is low (Curik et al., 2017). This was observed in this study when the ROH_1 criteria were applied to analyse correlations between whole genome \( F_{ROH} \) and chromosomal \( F_{ROH,K} \) of all 29 autosomes. Correlations varied from low (0.19 between \( F_{ROH} \) and \( F_{ROH,27} \)) to moderate (0.40 between \( F_{ROH} \) and \( F_{ROH,1} \)). To account for the varying patterns of inbreeding within the genome, the phenotypic values of fertility traits were regressed first on the chromosomal \( F_{ROH,K} \) of all autosomal chromosomes. Next, the chromosomes showing inbreeding depression were further partitioned into
smaller segments to locate genomic regions associated with inbreeding depression on fertility traits.

Table 7 summarises the results of the chromosomal and intrachromosomal analyses (I). Chromosomes 2, 15, 18 and 22 showed inbreeding depression when the threshold for p-values was set to <0.01 to account for multiple testing (I). The estimated effect of the 10% increase in FROH,K in chromosomes 2, 15, 18 and 22 was associated with 1.6 days longer IFL0, 2.3 days longer IFL2, 0.9 days longer IFL0 and 0.7 days longer IFL0, respectively.

Detected chromosomal inbreeding was dissected down to regional inbreeding (FROH,W) using a sliding window method. The first screening was performed using a window size of half of the chromosome size (measured in the number of SNPs) and a window overlapping of half of the window size. Further screenings were performed in the same way except the window size was decreased to half of the window size used in the previous screening.

The first screening revealed a 64.3 Mb region from the beginning of chromosome 2 that was associated with 1.6 days longer IFL0 per 10% increase in FROH,W (Table 7). When the window size was reduced, two overlapping windows with the highest statistical significance at position 0 – 47.1 Mb were associated with the lengthening of IFL0 by 0.9 and 1.0 days per 10% increase in FROH,W. The first screening for chromosome 15 found a 44.4 Mb region at the end of the chromosome where a 10% increase in FROH,W was associated with 2.2 days longer IFL2. The next screening revealed a region between 63.4 – 85.1 Mb where a 10% increase in FROH,W was associated with 1.7 days longer IFL2. The first screening for chromosome 18 did not produce statistically significant results, but the next screening showed a region between 47.7 – 65.7 Mb, where a 10% increase in FROH,W was associated with 0.5 days longer IFL0. The first screening for chromosome 22 resulted in two overlapping windows between 0 – 47.5 Mb associated with 0.6 days longer IFL0 per 10% increase in FROH,W. The region was more accurately located between 0 – 21.9 Mb in the next screening, which resulted in two overlapping windows associated with 0.5 days longer IFL0.
Table 7. Chromosomes and regions within chromosomes associated with inbreeding depression on interval from first to last insemination on heifers (IFL0) and second-parity cows (IFL2) with p-value < 0.01 (II)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Chromosomal</th>
<th>Regional, first round</th>
<th>Regional, second round</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CHR</td>
<td>b in days</td>
<td>Region (Mb)</td>
</tr>
<tr>
<td>IFL0</td>
<td>2</td>
<td>1.64***</td>
<td>0 – 64.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFL2</td>
<td>15</td>
<td>2.32*</td>
<td>40.8 – 85.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFL0</td>
<td>18</td>
<td>0.85*</td>
<td>32.8 – 65.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFL0</td>
<td>22</td>
<td>0.69*</td>
<td>14.3 – 47.5</td>
</tr>
</tbody>
</table>

1Whole chromosome sized window
2Half of the chromosome sized (in SNPs) window
3Quarter of the chromosome sized (in SNPs) window
4Estimate of inbreeding depression for a 10% increase in inbreeding coefficient.

*p-value < 1e-02, **p-value < 1e-03, ***p-value < 1e-04, ****p-value < 1e-05

Chromosomal inbreeding depression was detected only for IFL with the applied p-value threshold of 0.01. As discussed in Chapter 4.4, the power of the linear model used in this study may not have been sufficient to reveal inbreeding depression with sufficient statistical significance for NRR and ICF. It should be noted that AIS was excluded from this analysis and only IFL was used to reflect inbreeding depression of insemination period.

The summed chromosomal effect for IFL0 obtained in this study was a lengthening of 3.18 days per 10% increase in \( F_{ROH} \) (Table 7), while the whole genome estimate was 4.3 days (Table 6). It is likely that other regions contribute to inbreeding depression as well, but these possibly rare and small effects were not discoverable by the intrachromosomal approach applied in this study. Previous studies have mainly applied the ROH effect as a discrete variable (i.e. presence/absence of ROH) for identification of genomic regions associated with inbreeding depression (e.g. Keller et al., 2012; Simón-Sánchez et al., 2012; Pryce et al., 2014). Keller et al. (2012) and Simón-Sánchez et al. (2012) were able to detect specific regions in the human genome with significant associations to disease status of schizophrenia and early-onset Parkinson's disease, respectively. Pryce et al. (2014) investigated cattle genome and identified regions where the presence of ROH was associated with the lengthening of the calving interval. As pointed out by Saura et al. (2015), the advantage of using the
continuous $F_{ROH}$ variable in this study was that it produced more power to determine the associations between inbreeding and phenotype within specific genomic regions.

However, the use of sliding windows of fixed size in the current study may have resulted in certain limitations for the estimation of inbreeding depression. First, the ROH coverage of the region increased when the window size decreased, thereby decreasing the variation in the $F_{ROH,W}$. This led to decreased efficiency of the analysis with decreasing window size. Second, the same $F_{ROH,W}$ value in a given window may consist of varying proportions of short and long ROHs. As discussed in Chapter 4.4, the association between ROH and inbreeding depression is expected to depend on ROH length, but the results from the literature are inconsistent in terms of whether long or short ROHs harbour more detrimental recessive alleles in cattle populations. Pryce et al. (2014) found that only ROHs longer than 3.5 Mb were associated with inbreeding depression on milk yield. They suggested that natural selection will purge detrimental variants and that therefore only long ROHs representing recent inbreeding are associated with inbreeding depression. On the contrary, Zhang et al. (2015b) reported that the enrichment of detrimental variants was greater in short (<0.1Mb) and medium (0.1 – 3 Mb) ROHs than in long (> 3 Mb) ROHs. The authors pointed out that the strong artificial selection applied in cattle populations has resulted in long-time selection of favourable variants but has also led to the hitchhiking of detrimental variants in short ROHs. Doekes et al. (2019) reported both short and long ROHs contributing to inbreeding depression in cattle, although they found that the most detrimental ROH length varied between traits.

Within a large window the neutral long ROHs can mask the effects of the short detrimental ROHs. Because only those large windows that contributed to inbreeding depression were dissected down to shorter segments, some inbreeding depression due to short ROHs may have remained undetected. In turn, within the short windows the neutral short ROHs may have masked the effect of the long detrimental ROHs and lowered the accuracy of locating regions showing inbreeding depression. Moreover, the bias caused by short non-IBD segment further impairs the efficiency of detecting inbreeding depression with short windows. However, investigating the effects of short ROHs in larger windows would have required more dense SNP data, which was not available for this study. Despite these limitations, the sliding window approach effectively located chromosomal regions contributing to impaired fertility. Moreover, the sliding window analysis was complemented by the haplotype analysis to overcome the inaccuracy of the short windows to pinpoint the regions associated with inbreeding depression.

4.6 Haplotypes associated with inbreeding depression

A haplotype analysis was carried out on the regions showing inbreeding depression in the first round of intrachromosomal analysis (Table 7). Different haplotype lengths of 1 to 10 SNP, 20 SNP and 30 SNP were tested to find the haplotype lengths with the
Table 8. Effects of haplotypes on interval from first to last insemination of heifers (IFL0) or second-parity cows (IFL2) estimated using the recessive model (II)

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>CHR</th>
<th>Trait</th>
<th>Position (Mb)</th>
<th>Frequency of homozygotes</th>
<th>b1 in days (SE)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>IFL0</td>
<td>28.6 – 28.8</td>
<td>0.10</td>
<td>4.38 (1.01)</td>
<td>1.55e-05</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>IFL2</td>
<td>70.4 – 70.9</td>
<td>0.18</td>
<td>7.57 (2.16)</td>
<td>4.73e-04</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>IFL0</td>
<td>49.7 – 49.9</td>
<td>0.15</td>
<td>3.23 (0.83)</td>
<td>1.01e-04</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>IFL0</td>
<td>42.0 – 42.1</td>
<td>0.05</td>
<td>4.13 (1.29)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

1Difference between individuals homozygous/not homozygous for the haplotype

Both recessive and additive modes of inheritance were tested for the haplotypes, and the results supported the recessive mode of inheritance for IFL (II). The result is congruent with the partial dominance hypothesis which assumes that inbreeding depression is caused by detrimental recessive variants in homozygous state (Charlesworth and Willis, 2009). In addition, recessivity of the haplotypes indicate that the variants that contribute to impaired IFL have been selected against in the FAY breeding programmes.

The magnitude of the haplotype effects obtained in this study were in line with the results obtained by Pryce et al. (2014), who reported an increase in calving interval of between 5.1 and 7.6 days for chromosomes 2, 5, 8, 9, 15 and 24 for Holsteins using ROH status in GWAS analysis. It should be pointed out that the comparison of results is directive, as calving interval comprises ICF, IFL and gestation length (Pereira et al. [2016] showed that gestation length is not affected by inbreeding), and the haplotype analysis applied in this study did not include ROH status of SNPs. ROH was present for 26.5%, 51.4%, 18.5% and 31.9% of the haplotypes 1, 2, 3 and 4, respectively, which indicates that not all haplotype homozygosity was due to inbreeding and that some haplotypes may have been IBS.

Despite the origin of the homozygosity, accumulation of homozygosity for the regions detected in this study is expected to result in impairment of the phenotypic performance of FAY. The aim of the haplotype analysis was to pinpoint the regions...
in which homozygosity contributes to inbreeding depression on fertility, but a more detailed analysis is required in order to obtain full insight into the biological consequences of inbreeding. Possible candidate genes found from the detected regions are discussed in II.

4.7 ROH genotypes associated with reduced fertility and milk production

Intrachromosomal analysis of the regions showing inbreeding depression was performed by regressing the phenotypic value on region-specific $\text{F}_{\text{ROH}}$, which assumes that all ROHs within a region have a detrimental effect on the trait of interest. Howard et al. (2017b), however, pointed out that the detrimental effect is likely due to specific ROH genotype, while different ROH genotypes result in no effect. Furthermore, selection may have also created ROHs with favourable effect through a process known as purging of inbreeding depression (Boakes and Wang, 2005). Since inbreeding increases the frequency of detrimental homozygous (partially) recessive genotypes, the efficiency of selection against these unfavourable alleles also increases (Hedrick and Garcia-Dorado, 2016). Purging of inbreeding depression can be the result of natural selection (Doekes et al., 2019) or of artificial selection, through which inbred animals with a good phenotypic performance have been selected, while the inbred animals with a poor performance have been culled (McParland et al., 2009). If a genomic region contains ROHs with detrimental effect together with ROHs of neutral of favourable effect and a region specific $\text{F}_{\text{ROH}}$ is used as a covariate in the model, detrimental ROHs may not be found, as the effect is diluted by the neutral and favourable ROHs.

On the other hand, inbreeding depression can exist in seemingly outbred individuals (Gibson et al., 2006). As the authors pointed out, some long ROHs can be detected in outbred individuals due to unusual mutation, LD or low recombination rates in certain genomic regions. However, the overall genomic inbreeding coefficients of these individuals would lack variation (all estimates come close to zero), resulting in an inability to detect inbreeding depression by regression analysis (Keller et al., 2011).

Identification of specific ROH genotypes that result in unfavourable phenotypes was performed in this study to account for the varying levels of detrimental recessives within ROHs (III). In addition to female fertility traits, milk production traits were targeted in order to obtain a broader perspective on the effects of inbreeding. The analyses were performed using the Unfavorable Haplotype Finder program (Howard et al., 2017b), which identifies the core genotypes of ROHs associated with detrimental effects on the traits of interest. The outcome of the program is a list of core ROH genotypes that serve as tags for (usually much larger) ROHs. However, because the program does not run all the tag ROH genotypes simultaneously, the output list can contain several ROH genotypes that in reality tag the same full ROH. Therefore,
the statistically most significant ROH genotypes for ICF, IFL, AIS, milk yield (MILK), protein yield (PROT) and fat yield (FAT) are presented in Table 9. The most significant ROH genotypes were estimated to result in 9 to 17 days longer ICF, 13 to 38 days longer IFL, 0.3 to 1.0 more inseminations, 208 kg less milk, 7 kg less protein and 16 kg less fat compared to individuals with no ROH present on the identified regions (Table 9).

Table 9. The most significant unfavourable genotypes within ROH for the interval from calving to first insemination (ICF), interval from first to last insemination (IFL), number of inseminations (AIS), milk yield (MILK), protein yield (PROT) and fat yield (FAT) (III).

<table>
<thead>
<tr>
<th>Trait</th>
<th>CHR</th>
<th>Position (Mb)</th>
<th>Freq.</th>
<th>Full ROH (Mb)</th>
<th>b (SE)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICF1</td>
<td>12</td>
<td>65.13 – 66.37</td>
<td>0.010</td>
<td>6.76</td>
<td>9.2 (2.5)</td>
<td>2.22e-04</td>
</tr>
<tr>
<td>ICF2</td>
<td>10</td>
<td>62.34 – 64.20</td>
<td>0.017</td>
<td>7.48</td>
<td>9.5 (2.5)</td>
<td>1.92e-04</td>
</tr>
<tr>
<td>ICF3</td>
<td>12</td>
<td>61.67 – 62.73</td>
<td>0.018</td>
<td>7.29</td>
<td>17.1 (4.6)</td>
<td>2.20e-04</td>
</tr>
<tr>
<td>IFL0</td>
<td>17</td>
<td>9.21 – 10.31</td>
<td>0.015</td>
<td>7.98</td>
<td>12.8 (2.4)</td>
<td>4.84e-08</td>
</tr>
<tr>
<td>IFL1</td>
<td>5</td>
<td>27.54 – 31.16</td>
<td>0.032</td>
<td>10.71</td>
<td>13.2 (3.2)</td>
<td>2.85e-05</td>
</tr>
<tr>
<td>IFL2</td>
<td>8</td>
<td>65.01 – 69.12</td>
<td>0.015</td>
<td>11.90</td>
<td>29.6 (6.5)</td>
<td>4.96e-06</td>
</tr>
<tr>
<td>IFL3</td>
<td>18</td>
<td>8.01 – 8.85</td>
<td>0.021</td>
<td>4.83</td>
<td>37.7 (9.8)</td>
<td>1.16e-04</td>
</tr>
<tr>
<td>AIS0</td>
<td>17</td>
<td>9.47 – 10.59</td>
<td>0.018</td>
<td>3.15</td>
<td>0.26 (0.05)</td>
<td>8.59e-07</td>
</tr>
<tr>
<td>AIS1</td>
<td>14</td>
<td>81.76 – 83.07</td>
<td>0.010</td>
<td>5.39</td>
<td>0.48 (0.10)</td>
<td>2.26e-06</td>
</tr>
<tr>
<td>AIS2</td>
<td>15</td>
<td>7.31 – 9.15</td>
<td>0.010</td>
<td>3.33</td>
<td>0.58 (0.14)</td>
<td>1.83e-05</td>
</tr>
<tr>
<td>AIS3</td>
<td>18</td>
<td>4.90 – 6.93</td>
<td>0.018</td>
<td>4.55</td>
<td>0.95 (0.22)</td>
<td>2.34e-05</td>
</tr>
<tr>
<td>MILK</td>
<td>20</td>
<td>31.55 – 33.15</td>
<td>0.049</td>
<td>7.36</td>
<td>-207.843 (38.9)</td>
<td>9.30e-08</td>
</tr>
<tr>
<td>PROT</td>
<td>13</td>
<td>53.66 – 55.64</td>
<td>0.056</td>
<td>12.75</td>
<td>-7.04 (1.12)</td>
<td>3.60e-10</td>
</tr>
<tr>
<td>FAT</td>
<td>5</td>
<td>87.02 – 88.11</td>
<td>0.012</td>
<td>13.28</td>
<td>-15.68 (1.69)</td>
<td>8.60e-09</td>
</tr>
</tbody>
</table>

1Fertility traits considered separately for heifers (0) and for parities 1, 2 and 3
2Median length of the full ROH segment
3Estimate b represents the difference between animals with the ROH genotype and animals without ROH (in days for ICF and IFL, in number of inseminations for AIS and in kg for MILK, PROT and FAT)

Although the effects of the most significant ROHs on the traits were quite large, they corresponded well to the estimates obtained from previous studies using the same program in cattle populations (Marras et al., 2018; Baes et al., 2019). Baes et al. (2019) reported that cows with the most extreme ROH were likely to have approximately 9
days longer ICF compared to cows without ROH (as least squares means difference). Corresponding results by Marras et al. (2018) indicated approximately 11 days longer ICF. Here, an effect of 9 days was obtained for ICF1 and ICF2 and 17 days for ICF3.

Similarly for AIS, Baes et al. (2019) reported an effect of 0.5 and 0.6 more AIS for heifers and cows, respectively, while Marras et al. (2018) obtained effects of 0.6 more AIS for both heifers and cows. An effect of 0.3 more AIS for heifers was found in the current study and a varying effect from 0.5 to 1 more AIS was found for cows.

Furthermore, the effects of the most extreme ROH for days open (ICF+IFL) were 27 days obtained by Baes et al. (2019) and 26 days obtained by Marras et al. (2018), while the combined effects of ICF and IFL obtained in this study varied from 22 days to 55 days between first- and third-parity cows. Results for IFL can be compared to the results for interval from first insemination to calving reduced by the results for the gestation length reported by the authors. Using this comparison, the effect for IFL was approximately 14 days for heifers and 16 days for cows by Baes et al. (2019) and 11 days for heifers and 19 days for cows by Marras et al. (2018), while in this study an effect of 13 days was obtained for heifers and an effect varying from 13 days to 38 days was obtained for cows. Previous studies did not separate the fertility phenotypes by parities, which may explain the larger effects for the later parities obtained in this study compared to results from previous studies.

For milk production traits, the effects obtained in the current study were approximately 208 kg for MILK, 7 kg for PROT and 16 kg for FAT, whereas Baes et al. (2019) found effects of 918 kg, 29 kg and 32 kg for 305-day milk, protein and fat yields, respectively, and Marras et al. (2018) reported effects of 684 kg, 22 kg and 29 kg for 305-day milk, protein and fat yields, respectively.

Howard et al. (2017b) and Marras et al. (2018) pointed out that ROH genotypes impacting multiple traits could pinpoint regions in the genome where inbreeding is particularly detrimental. Figure 2 illustrates the locations of unfavourable ROH genotypes with the p-value threshold $< 1.0e-05$ set to minimize false associations due to multiple testing.
Figure 2. Genomic regions containing ROH genotypes with unfavourable effects for interval from first to last insemination (IFL), number of inseminations (AIS) and milk, protein and fat yields (MILK, PROT, FAT) with p-value < 1e-05.

Multiple regions contained ROHs with negative effects across fertility traits or across milk production traits, but no overlapping regions were observed between fertility and milk production traits (Figure 2). However, as the ROH genotypes tag much larger ROHs, the associated QTLs may locate away from the tag genotypes. Therefore, the regions containing unfavourable ROH genotypes should be considered as starting points for more detailed investigation of the impact of inbreeding at specific genomic regions.

Interestingly none of the regions containing unfavourable ROH genotypes with p-value < 1e-05 overlapped with the regions detected using the intrachromosomal analysis (II). This result supports the hypothesis that the unfavourable effect of a region is likely due to a single or a few ROH genotypes, with the other ROH genotypes having neutral or even favourable effect on the trait of interest (Howard et al., 2017b). Here, the frequency of the unfavourable ROH genotypes were mostly between 0.01
and 0.02, which indicates that these regions were undetectable in the intrachromosomal study due to a larger frequency of non-harmful ROH genotypes diluting the overall effect towards zero. On the contrary, the regions on chromosomes 2, 15, 18 and 22 identified in the intrachromosomal analysis contained several ROH genotypes detected in this analysis, but the statistical significance of these ROH genotypes was lower than the applied p-value threshold. Detrimental variants may have accumulated on these regions, which is why the individual ROH genotypes did not stand out in the unfavourable ROH analysis, although the combined effect of all ROHs showed inbreeding depression in the intrachromosomal analysis. However, the difference in identified regions between these analyses may be due to other factors (overdominance or epistasis, for example), which is why further research is needed in order to understand the genetic basis of inbreeding depression.

A problem in all ROH analyses throughout this study, which may have been particularly significant in the ROH genotype analysis where no threshold was set for the minimum ROH length, is the effect of hemizygosity. This may have resulted in there being some improperly detected ROHs that are in reality deletions. For example, a 660 kb deletion in chromosome 12 with detrimental effects for all the traits reflecting pregnancy success (AIS, IFL and NRR) has been found to occur in 32% of the Finnish Red Cattle animals (Kadri et al., 2014). However, McQuillan et al. (2008) analysed the effect of deletions for the ROH detection and found that the ROHs identified were in reality homozygous segments rather than deletions or other chromosomal abnormalities.

4.8 Implications and future research

The main goal of this thesis was to investigate the suitability of different measures of inbreeding to detect inbreeding depression and to locate genomic regions associated with inbreeding depression. The information produced in this thesis can be exploited for more efficient control of inbreeding depression or to investigate the mechanism of inbreeding depression by examining the identified regions showing inbreeding depression.

Managing the levels and rates of inbreeding in dairy cattle populations is crucial to maintaining their fitness and genetic variation. However, the current practice of controlling inbreeding by minimizing the kinship of the parents based on pedigree may not be optimal. Pryce et al. (2012) showed that mating decisions based on genomic information outperform the use of pedigree data in controlling the level of inbreeding of the next generation. Additionally, the results of this thesis demonstrate that the pedigree-based inbreeding coefficients correlated weakly with inbreeding depression.

Although genomic measures in the current study produced more precise views of the effects of inbreeding, further research is needed to effectively incorporate genomic information into the Finnish Ayrshire breeding programme. ROH can provide the most effective measure of the true levels of inbreeding that are influenced by selection
(Curik et al., 2017), but the need to standardize ROH definition criteria to obtain comparable inbreeding coefficients was demonstrated in this study. Moreover, future work should examine how the quality control of SNPs, panel density, recombination rate and presence of deletions each affect ROH detection.

In addition to increasing the accuracy of ROH detection, the optimal method by which to incorporate ROH information in the breeding programme should be investigated. The region-specific metrics of ROH-based inbreeding examined in this study could be an interesting approach. Overall, genomic selection is expected to reduce the whole genome levels of inbreeding per generation by accounting for the Mendelian sampling variation between individuals with the same pedigree (Hayes et al., 2009). However, genomic selection could create stronger selection at specific chromosomal regions than traditional selection does, which can be seen as an increased number of ROH segments (Forutan et al., 2018). Moreover, the implementation of genomic selection has changed the distribution of short and long ROHs such that short ROHs have become more frequent (Forutan et al., 2018). Pryce et al. (2012) suggested a mating strategy based on ROH, where mating of parents that carry common ROHs is avoided. Even though this strategy could decrease the overall levels of inbreeding in the next generation, it may not be effective in decreasing the effects of inbreeding depression because common ROHs could have been selected due to their favourable effects or because ROHs with detrimental effects segregate in the population with low frequency, as was observed in this study. A more effective method may be to identify unfavourable ROH genotypes, as was done in this study, and incorporate those into a mating decision algorithm. For example, Howard et al. (2017b) presented a method by which to construct an inbreeding load matrix based on the unfavourable ROHs, where the off-diagonals represent the predicted outcome in the phenotypic performance of the offspring when mating parents that carry one or two copies of the haplotypes identified as detrimental when homozygous. However, controlling inbreeding depression cannot rely solely on ROH; the effect of lethal variants should also be considered. This is because lethal variants cannot be traced by examining homozygosity in a population, but rather by analysing the lack of homozygosity at specific regions (VanRaden et al., 2011).

From a biological point of view, understanding the mechanism of inbreeding depression would require a large-scale analysis using high-density data to resolve the inheritance patterns and the genes involved in the regions contributing to inbreeding depression. Due to the continuously decreasing costs of whole genome sequencing, sequence data will be available for a large number of both bulls and cows in the near future (Curik et al., 2017). This will open up opportunities to research the biological consequences of inbreeding in greater detail. Two different approaches designed to identify genomic regions showing inbreeding depression were used in this study: overall homozygosity at specific regions and homozygosity of specific genotypes. Targeting the identified regions may provide more insight into the mechanisms of inbreeding depression, as the increased overall homozygosity may reflect the presence of common recessive detrimental variants with small effects or decreased beneficial
heterozygosity, whereas specific harmful ROH genotypes can represent rare variants with large effects. In addition, the methods presented in this study could also be used to detect the effects of selection and purging by identifying homozygosity with favourable associations with the traits of interest.
5 CONCLUSIONS

This study investigated the levels and effects of inbreeding in Finnish Ayrshire cattle using pedigree information and genomic approaches. Based on the results, different metrics of inbreeding coefficient produced estimates that differed in magnitude. The estimates derived from genotypic data were strongly correlated with each other, whereas pedigree-based estimates were only moderately correlated to genomic estimates. This study found that the criteria used to define ROH has major impacts on the number and length of observed ROH. The results of this study highlight the importance of careful consideration of criteria used in ROH definition in order to avoid spurious ROH calls.

In the analysis of inbreeding depression, increased inbreeding based on genomic estimates was associated with decreased fertility. However, no inbreeding depression was observed with pedigree-based estimates of inbreeding, albeit animals with shallow or incomplete pedigrees were pruned out. Inbreeding based on ROH and overall homozygosity resulted in inbreeding depression estimates of varying magnitude. Therefore, results from inbreeding depression analyses must be interpreted by taking into account the methods used to derive inbreeding coefficients.

This study showed that patterns of inbreeding depression vary across the genome in Finnish Ayrshire cattle. Dissecting ROH-based inbreeding to chromosomes and chromosomal segments revealed regions where increased inbreeding was associated with reduced fertility. However, the effect of different ROH genotypes on the traits of interest can vary, which is why a further analysis was carried out to estimate the effects of different ROH genotypes separately. Based on the results, several regions contained ROH genotypes with detrimental effects on female fertility and milk production traits. These regions showing detrimental effects on fertility did not overlap with the previously identified regions. The results of this study indicate that the unfavourable effect of a region could be due to either several ROHs with small effects or specific ROHs with larger effects. However, further research is needed in order to understand the genetic background of inbreeding depression.

The results of this study can be used to improve the estimation of the effects of inbreeding in the breeding programme for more efficient control of inbreeding depression, or to investigate the mechanism of inbreeding depression, such as by examining the genes in the identified regions. Overall, the results of this study emphasize the importance of balancing between genetic gain and levels of inbreeding in the Finnish Ayrshire breeding programme, as detrimental effects of inbreeding on reproductive ability (and also on milk production performance) were clearly observed.
REFERENCES


