Gene Mapping of Reproduction Traits in Dairy Cattle

Johanna Karolina Höglund
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics, Uppsala

Faculty of Science and Technology
Department of Molecular Biology and Genetics,
Center for Quantitative Genetics and Genomics, Aarhus

Doctoral Thesis
Swedish University of Agricultural Sciences
Aarhus University
2013

Acta Universitatis agriculturae Sueciae
2013:50

Biggan med kalven Andy från Bergs Mjölkgård AB, Strandskogen, Färjestaden (photo: VikingGenetics)

ISSN 1652-6880
ISBN 978-91-576-7838-6

Print: SLU Service/Repro, Uppsala 2013

Gene Mapping of Reproduction Traits in Dairy Cattle.

Abstract

In this thesis my aims were to map genes affecting reproduction in cattle and to explore the correlated effect of these genes on yield traits. This knowledge is expected to provide novel insights into the biological processes underlying reproduction traits and to identify individual causal polymorphisms or genetic markers for practical application in marker-assisted and genomic selection strategies.

A wide range of female fertility traits has been recorded for cattle indigenous to the Nordic countries. Moreover, the trait definitions have been standardized across these countries and the cattle industry has made its data available for research. The research findings, themselves, highlight the benefit of such a comprehensive data source and increasingly detailed genomic information ranging from relatively few microsatellites to full sequence profiles.

In manuscript I, a genome scan was performed using linkage analysis with microsatellite markers in order to identify genomic areas of interest. Twenty-six quantitative trait loci (QTL) affecting female fertility were identified.

In manuscript II, the trait fertility treatments were decomposed into sub-traits, four QTL for retained placenta were revealed. These QTLs and additional 24 QTL regions discovered in manuscript I, were analyzed for effects on yield traits. Sixteen of the genomic regions containing QTL for female fertility also harbored QTL for milk production or milk composition traits. Twelve QTL regions with effects on nine different fertility traits did not harbor any QTL for milk production or milk composition traits. When.

An association study based on 50k single nucleotide polymorphism (SNP) data indicated significant associations for 4,474 SNPs with eight different female fertility traits in Nordic Holstein. Of these SNPs, 152 were validated in both the Danish Jersey and the Nordic Red breeds, the most significant results were found on Chromosomes1, 4, 9, 11 and 13. Small significant regions on chromosomes 4 and 13 were analyzed by sequencing to further focus the genomic region in which candidate genes and possible causative mutations may be localized.

Subsequently, new calving data was added and a validation study was performed, which confirmed 321 of 424 SNPs significantly associated with 14 calving traits. It was determined, however, that the analysis of the new data alone had low power, and an analysis of the full data set was more powerful.

Keywords: mapping, dairy cattle, reproduction, quantitative trait loci, female fertility, calving, validation

Author's address: Johanna Höglund, Aarhus University, Department of Molecular Biology and Genetics, Blichers allé 20, 8830 Tjele, Denmark
E-mail: Johanna.hoglund@agrsci.dk

Dedication

Till min familj

Contents

List of Publications

This thesis is based on the work contained in the following publications*, referred to by Roman numerals in the text:

I. Höglund J K, Guldbrandtsen B, Su G, Thomsen Bo, Lund M S. (2009) Genome scan detects quantitative trait loci affecting female fertility traits in Danish and Swedish Holstein Cattle. *Journal of Dairy Science* Vol. 92 (Issue 5), Pages 2136-2143.

II. Höglund J K, Buitenhuis A J, Guldbrandtsen B, Su G, Thomsen Bo, Lund M S. (2009) Overlapping chromosomal regions for fertility traits and production traits in the Danish Holstein population. *Journal of Dairy Science* Vol. 92 (Issue 11), Pages 5712-5719.

III. Höglund J K, Guldbrandtsen B, Sahana G, Lund M S. Validation of Associations for Female Fertility Traits in Nordic Holstein, Nordic Red and Jersey Dairy Cattle. Submitted to *PLoS One*.

IV. Höglund J K, Guldbrandtsen B, Sahana G, Lund M S. (2013) Fine mapping of specific genomic regions across breeds for female fertility traits. Manuscript in preparation.

V. Höglund J K, Guldbrandtsen B, Lund M S, Sahana G. (2012) Analysis of genome-wide association follow-up study for calving traits in dairy cattle. *BMC Genetics* doi:10.1186/1471-2156-13-71

*All data is reproduced with the permission of the corresponding publisher.

Abbreviations

AI	artificial insemination
AIS	number of inseminations
BI	birth index
BLUP	best linear unbiased prediction
CE	calving ease
CI	calving index
CS	calf size
D	direct effect
EBV	estimated breeding value
F	first pregnancy
FRT	fertility treatments
FSH	follicle-stimulating hormone
FTI	fertility index
GDD	granddaughter design
GnRH	gonadotropin-releasing hormone
GWAS	genome-wide association study
HST	heat strength
ICF	calving-to-first insemination
IFL	first-to-last insemination
L	later pregnancy
LA	linkage analysis
LD	linkage disequilibrium
LH	luteinizing hormone
MAF	minor allele frequency
M	maternal effect
MAS	marker assisted selection
NAV	Nordisk Avlsværdivurdering
NRR	non-return rate

QTL	quantitative trait loci
QTN	quantitative trait nucleotide
RACE	rapid amplification of cDNA ends
SB	stillbirth
SNP	single nucleotide polymorphism
STBV	single trait breeding value
WGS	whole genome sequence

1 Introduction

The cow's ability to reproduce is essential for milk production and is the key biological feature on which the dairy industry is based. Impaired reproduction capacity results in additional inseminations, higher replacement rate and increased culling rate. In fact, reproduction problems are cited as the most common reason for industry culling (Ahlman *et al.,* 2011).

It should be recognized that application of breeding strategies for improved reproduction should be accompanied by efforts to reduce reproduction diseases. To this end, risk profiles of particular diseases in particular periods have been described; specifically, the dry period represents a higher risk of all production diseases and the period after calving represents a higher risk of infectious diseases, such as that from retained placenta. In fact, the risk for all production diseases is highest during the first 100 days after calving. Reducing the incidence of reproduction-related diseases will provide benefits to the animals' welfare and industry's efficiency, namely reducing veterinary treatments, shortening calving intervals, and reducing the number of inseminations.

Increasing the dairy industry's efficiency has broader effects on human health and wellbeing. As the global human population continues to grow, so does its demand for essential nutrients derived from animal products. Thus, as the competition for land and water resources intensifies, more efficient livestock production is required. Female fertility is a trait which has a large impact on the efficiency of dairy cattle production: fertile females have more offspring and therefore dilute their own feed requirements over this increased number of offspring and number of lactations (Hayes *et al.,* 2013). Efficient reproduction is therefore also expected to be associated with lower emission of methane and nutrients per unit of product.

Detection of quantitative trait loci (QTL) is an important first step in identifying genes affecting a trait, such as female fertility in cattle. QTL are

stretches of DNA containing a gene or groups of genes that explain the variation in quantitative manifestation of a trait among a population. Even when the causal gene or genes are not identified, it is still possible to incorporate (and put a higher weighted value on) a specific QTL region(s) in a genomic selection scheme (Boichard *et al*., 2012). During the course of my PhD thesis, genome sequencing technologies and bioinformatic analysis approaches advanced tremendously. Widespread application of these tools has led to remarkable increases in the numbers of trait markers available and thus enhanced precision of QTL mapping. Indeed, the in-depth sequence data generated by these new technologies (see Figure 2) overcomes the dependence on linkage disequilibrium (LD) that limited the traditional genetic approaches, so that causal mutations can be directly detected in the sequence data.

The breeding industry has worked in conjunction with and exploited these more powerful analytical strategies to improve its efficiency. For example, the best linear unbiased prediction (BLUP) statistical model which incorporates information from the pedigree and phenotypes only is now used in conjunction with Genomic selection (GS) which makes breeding decisions based on genomic information. Thus, breeders now predict a breeding value from a large number of genetic markers by adding the predicted effects of markers from individuals.

The genome-wide association study (GWAS) of fertility traits has led to the identification of specific genes or chromosomal regions related to individual traits affecting overall female fertility. The availability of full genome sequence data can also help to identify causal mutations underlying variation in female fertility. In addition to revealing the genetic architecture that underlies the physiological and biological processes of female reproduction, this information could be practically applied to genomic selection schemes. By assigning higher weights to certain genomic regions that influence female fertility, more effective combinations of chromosomal regions can be selected to increase the number of calvings without increasing the incidence of reproductive diseases.

1.1 Trends in female fertility

Female fertility has declined within the last decades especially in the Holstein population. In recent years, however, the Nordic Holstein population has experienced a slight improvement in female fertility, (Årsstatistik Avl 2011/12). The overall decline in fertility rates has been attributed, at least partially, to genetic factors (Lucy, 2001; Royal *et al*., 2008). A study by Shook (2006) estimated that approximately one-fourth of the decline in pregnancy

rate was due to genetic variables. One of the reasons for the decline in female fertility is believed to be a result of an unfavorable genetic correlation between female fertility and milk yield (Roxström *et al.,* 2010). When considered in the context of the intense genetic selection for improved yield that has been carried out in recent decades, the unfavorable genetic correlation may have contributed to the decline in female fertility, at least in part, due to an insufficient weight having been placed on female fertility in the breeding goal. Moreover, the extremely intense use of a selected pool of bull sires has led to a dramatic reduction of effective population sizes, promoting accumulation of deleterious or even lethal recessive alleles. Some of these act in early embryonic development and therefore result in as reduced fertility (Van Raden *et al.,* 2011)

Despite the recent slight increase in female fertility of the Nordic Holstein breed, there is still a lot of room for improvement. The recent genetic improvement is in part due to increased emphasis on female fertility in the breeding goal. Moreover, availability of more information on female fertility characteristics for foreign bulls along with proactive removal of bulls producing daughters with particularly low indices for female fertility from the breeding program (Lars Nielsen, VikingGenetics, personal communication). To this end, investigations will first need to determine whether unfavorable genetic correlations exist between the different components of female fertility, such as reproductive diseases and other traits currently included in the selection index. If such unfavorable genetic correlations do exist than that fact should be taken into account when including the particular trait in the selection index for fertility in order to avoid adverse effects on other traits. However, determining the genetic correlations between the traits affecting female fertility is challenging due to complexity in female fertility traits.

1.2 Inheritance

Fertility traits are modeled as quantitative traits. This means we consider the influence of multiple genes, alleles and the environment.

The heritability of fertility traits is characterized as low, typically in the range of 0.015-0.08 for the Nordic Holstein population (Sun *et al*., 2009). Moreover, low heritability may explain why the task of identifying genes that affect female fertility traits has proven particularly difficult, as each causal gene only explains a small fraction of the total phenotypic variance. However, even though the heritability is low, the additive genetic variation is considerable in the Nordic Holstein population (Sun *et al*., 2009), which represents a

potentially useful opportunity to breed for improved female fertility despite the low heritability.
Genetic progress through selective breeding requires accurate breeding values. A trait with low heritability requires large progeny groups in order to attain an acceptable level of accuracy for the breeding value. To this end, international cooperation will be beneficial. Dairy farmers' organizations in Denmark, Sweden, and Finland established a joint breeding value estimation system (Nordisk Avlsværdivurdering, NAV) in 2002 (Aamand, 2005). NAV predicts breeding values using a large number of registrations, thereby increasing the accuracy of estimated breeding values (EBVs). The closer the predicted breeding values are to the true genetic values, the higher the power to identify QTL.

1.3 Reproduction traits

The fertility trait of a cow is ultimately composed of a number of characteristics, including the cow's abilities to re-start its estrous cycle after parturition, exhibit signs of heat, get pregnant when inseminated, and sustain a pregnancy. Likewise, variation in fertility is influenced by variations in a broad range of physiological processes and their related factors.
A number of recorded traits inform us about the characteristics making up female fertility. For instance, the traits of first-to-last insemination (IFL) and number of inseminations (AIS) are measures a Cow´s/heifer's ability to show signs of heat and ability to conceive (defined as the pregnancy rate). These two traits represent all of the events of the heat state, and also represent the factors involved in conception. The trait of calving to first insemination (ICF) is a measure of both the ability to show heat and the ability to return to cycling after calving. The trait of 56-day non-return rate (NRR) also represents the cow´s/heifer´s ability to conceive after insemination pregnancy rate.
A highly simplified illustration of a heifer's estrous cycle is provided below. However, this description reveals the remarkable complexity of physiological processes and factors involved in the estrous portion alone for the traits of IFL, AIS, and NRR.
The estrous cycle is regulated by the hypothalamic-pituitary-gonadal axis, which produces hormones that determine reproductive events. Three organs—hypothalamus, pituitary and ovary—are involved in this reproductive signaling axis. As previously described by Rick Rasby and Rosemary Vinton at the University of Lincoln-Nebraska, (2012) *"The sequence of hormonal release essentially begins with the synthesis and release of gonadotropin-releasing hormone (GnRH) from the hypothalamus. This polypeptide hormone*

is transported to the anterior pituitary through a highly specialized capillary network called the hypothalamo-hypophyseal portal system. GnRH functions to stimulate the anterior pituitary to produce and release follicle-stimulating hormone (FSH) and luteinizing hormone (LH). FSH and LH are transported through systemic blood circulation to the ovaries, where they initiate a series of morphological changes that lead to ovulation and pregnancy if fertilization occurs. Morphological changes also occur on the ovary throughout the cow and heifer's oestrous cycle. Once ovulation has occurred and the egg is released, the cells on the ovary that made up the ovulatory follicle differentiate to form luteal cells" (this text was reproduced with the permission of the author). In order to investigate the underlying biology behind the trait fertility treatments (FRT), FRT was divided into its underlying sub traits in manuscript II. Each of these sub traits could affect reproductive efficiency. Presented below are the traits analyzed in this thesis.

Infective Reproductive Disorders. Uterine disease occurring in the period after calving has serious negative influence on the heifer's/cow´s subsequent reproductive performance (Gilbert *et al*., 2005). The uterine infection itself and antibiotics delivered to the uterus and retained placenta is traits recorded in the breeding scheme. The underlying disease state is defined as endometritis or metritis, with metritis being the more severe form. Retained placenta can lead to the development of endometritis and metritis, thereby impairing the subsequent reproductive performance of the affected heifer/cow (LeBlanc, 2008). Due to the relation of retained placenta to endometritis and metritis, the two are considered as a single phenotype in this study.

Spontaneous abortion. The overall frequency of spontaneous abortions among recorded pregnancies is 1.5% (in the Nordic Holstein population). This trait, in particular, is believed to be affected by a significant underestimation of its true frequency, as it is difficult for a farmer to detect an abortion occurring early in pregnancy, the cow is than subsequently considered as not pregnant.

Calving traits. The Calving trait also contributes to the overall reproduction. Calving traits are reflecting different aspects of calving. Calving ease (CE) reflects different degrees of difficulties in the birth of the calf. In addition, the size of the calf is registered by farmers, as large calves are considered at high risk of dystocia. Stillbirth (SB) is defined as calf mortality before and during the 24 hours after parturition. The SB rate is likely multifactorial in nature where birth weight, difficult calving, recessive lethals, pathogens and incompatibility between calf size and dam size have been suggested reasons (Berglund *et al*., 2003).

1.4 Gene mapping

The current strategies of QTL analysis include a number of statistical methods that link two types of information: phenotypic data (trait measurements) and genotypic data (usually molecular markers). In this study, linkage analysis and association studies were selected as the methods to carry out the gene mapping.

1.5 Linkage analysis

Linkage analysis (LA) is based on the fact that genetic markers proximal to a QTL tend to be inherited together because only few recombinations occur during the meiosis. This is termed linkage. The number of recombinations in a single meiosis is relatively low. When DNA is transmitted to offspring it is therefore passed on in large chromosome blocks. Inheritance of markers therefore indicates the inheritance of a large chromosomal region. If the inheritance of markers statistically associates with the pattern of (dis)similarity in trait phenotypes, there is reason to believe that the chromosome region marked by the markers harbors a gene or genes affecting the trait in question.
In a QTL genome scan (reported in manuscripts I and II) using linkage analysis with a granddaughter design (GDD) see figure 1 (Weller *et al*., 1990), the phenotypes were set as breeding values of bulls based on the information from many daughters of the sire. The method studies the co-inheritance of markers and traits from grandsires to their progeny tested sons. The simultaneous inheritance of multiple markers is used to scan the genome. This method is advantageous over LA, as the latter has low precision for the QTL position since it only utilizes recombination events that occur in the grandsire. Moreover, the low number of recombination events prevents localization of the QTL to segments smaller than 10~20 cM.

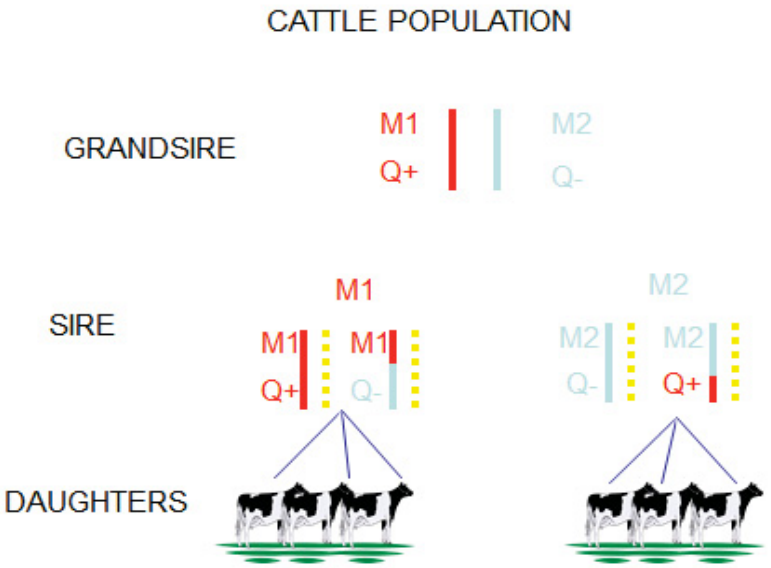

Figure 1. A simplified description of Linkage analysis illustrating recombination with a granddaughter design (GDD). In a GDD design genotypes are obtained on grandsires and their sons and the phenotypes of the sons are based on the performance of their daughters. The two marker alleles, M1 and M2, are linked to the QTL alleles Q+ and Q-. The two marker alleles can be passed on from the grandsire to the sire in two different chromosome blocks, depending on whether or not recombination has occurred. The yellow dotted lines represent the maternal contribution. Picture by Mogens Sandø Lund.

1.6 Association mapping

With today's dense marker maps, such as those represented on SNP chips and by whole genome sequence data, it has become possible to perform GWAS). This method narrows down the region on the genome where causative polymorphisms are located.

The basis of association studies is linkage disequilibrium. In contrast to linkage analyses which only exploit information from current recombinations, GWAS takes advantage of historical recombinations as well. GWAS performed on quantitative traits is usually performed by marker regression, where each marker by regressing the dependent variables onto the number of copies of one of the alleles. This strategy has been used in manuscripts III-V in this thesis. Figure 2 illustrates a GWAS performed with the Bovine SNP50 BeadChip (Illumina, Inc.). While an advantage of GWAS is the simple nature of the

analysis of each marker, the efficiency is complicated by the impact of several inheritance- and study-related factors, such as allele effect size, density of

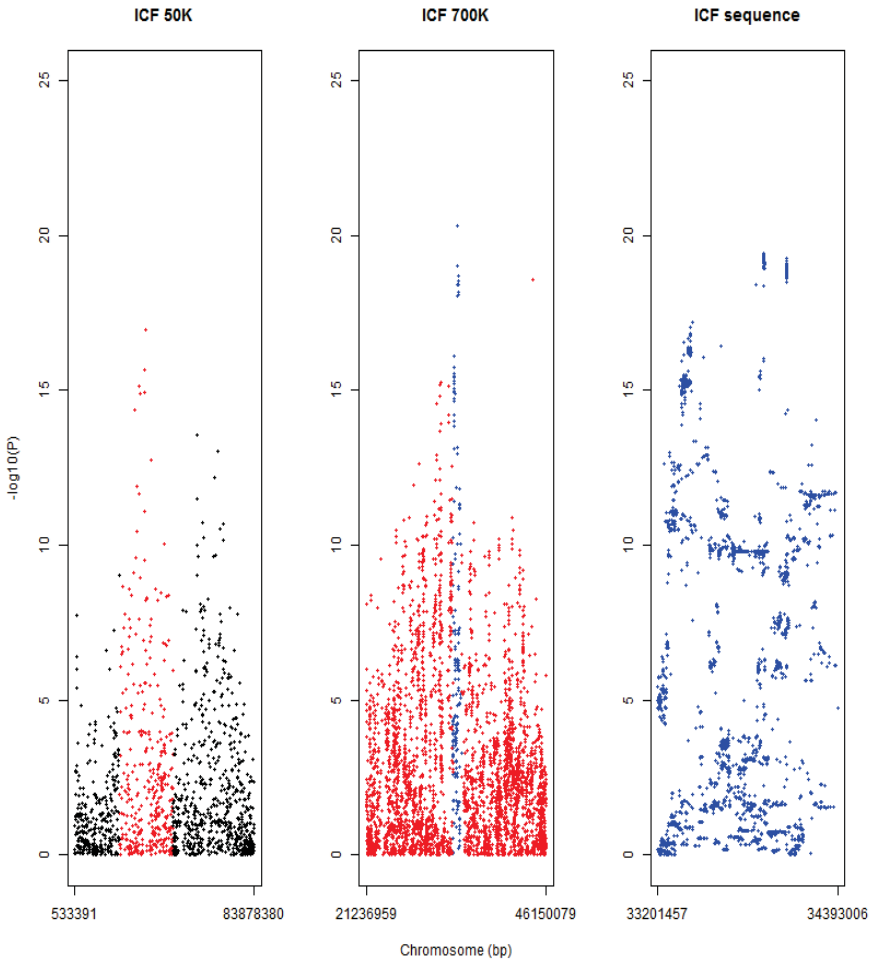

Figure 2. By increasing the density of markers, a genomic region of interest may be narrowed down for focused analysis. (Far left panel) The trait of ICF was analyzed using a BovineSNP50 bead chip (50K) on BTA13. (Middle panel) The selected (red) region from the first panel was analyzed using a High-Density BeadChip (700K; HD BeadChip by Illumina, Inc.). (Far right panel) The blue area from the middle panel was analyzed with sequence data. The dotted line represents genome-wide significance for the individual chromosome positions.

Informative markers, numbers of animals applied with genotype and phenotype data, as well as the degree of LD. Many large species of farm animals, including dairy cattle, have in recent generations had small effective population

sizes, causing extensive LD. Therefore, within-breed LD is a significant factor in large domestic animals. While this fact makes it possible to reduce the number of markers used in analyses, it also limits the accuracy of mapping that is achievable. Figure 3 illustrates the potential to narrow down a genomic region by adding more markers; the last panel shows an actual analysis performed with sequence data. Two key advantages of using sequence data are that the causal mutation is likely to be present in the data set and that bioinformatic information is available to qualify competing variants.

1.7 Validation

There are often pronounced disagreements in the results obtained by LA and GWAS conducted in different populations. However, it is necessary to perform studies in different populations in order to detect condition-specific (such as breed) polymorphisms. Such studies have also revealed that effects of the same polymorphism may differ in different genetic backgrounds. In these studies, failure of validation of QTL data may indicate that the original detection was a false positive, but may also occur when detection power is low. The former finding highlights the need to confirm detected QTL before extensive resources are invested into gene identification within the region.

In manuscript IV of this thesis, our efforts to validate associations across breeds are reported. The probability of observing spurious associations between a particular trait and a SNP in multiple populations by chance are small, especially if two or more validation populations of different breeds are used (Karlsson *et al*., 2007). There are different strategies to analyze and validate GWAS results. In time more and more data accumulate giving more animals with breeding values and higher accuracy to the predicted breeding values. In manuscript V of this thesis we compared two strategies to validate previously detected QTL when new data have accumulated from the same population. The comparison is made by analyzing all the data collected on all published sires presently available. We compare this to analyzing only the new accumulated data and compare associations to those found on older sires previously analyzed and published. The principle is illustrated in figure 4.

2 Objective of the thesis

The overall goals of this PhD thesis research were to map genes affecting reproductive traits in cattle and to contribute to the overall genetic improvement of cattle reproduction. It is important to validate genomic areas of interest before substantial resources are invested in efforts to identify causal polymorphisms. Ultimately, these data will help to expand our understanding of the genetic basis of female reproductive biology and to improve genomic predictions within and across breeds.

The specific aims of the individual subprojects are as follows:

I. The objective of this study was to detect QTL for multiple female fertility traits in Danish and Swedish populations of Holstein cattle.

II. The objectives of this study were to refine fertility phenotypes and explore whether QTL that segregate for female fertility also segregate for yield traits.

III. The objectives of this study were to detect significant SNP associations for female fertility traits in the Nordic Holstein population and validate these associations in the Nordic Red and Jersey populations.

IV. The objective of this study was to use sequence data to narrow down our focus on certain regions in the genome where causative mutations for female fertility traits are likely to be located.

V. The objectives of this study were to detect associations and evaluate methods to validate associations for calving traits in the Holstein population.

3 Summary of Investigations

3.1 Phenotypes

For the studies reported in manuscripts I-III and V, the single trait breeding values (STBVs) were used. The STBVs had been generated specifically for QTL mapping by the Nordic Cattle Genetic Evaluation Company (http://www.nordicebv.info). STBVs were predicted for each animal by using the BLUP procedure with a sire model, wherein sires were treated as unrelated. Pedigree information other than the link between sire and daughter was not included in this prediction model of STBV. Thus, the STBVs of a sire were predicted from its daughters' information only. The STBVs were adjusted for the same systematic environmental effects as in the official routine evaluations. In manuscript IV, de-regressed proofs were used as phenotypes.

3.2 Recordings of the phenotypes

Many characteristics related to fertility are routinely recorded in the Nordic countries. All inseminations are performed by artificial insemination (AI) technicians or licensed farmers and are recorded at the time of insemination. The corresponding data are subsequently entered into each country's national database for the particular sire and date of insemination. These records form the basis for predicting breeding values. For four female fertility traits (AIS, 56-day NNR, IFL, and HST), the recordings were split into two groups: heifers (H) and lactating cows (C). Separate breeding values were calculated for the heifer and lactating cow phenotypes.

3.3 Fertility traits

3.3.1 Number of inseminations (AIS)

This trait represents the number of inseminations a cow or heifer needs to get pregnant. It describes the cow's or heifer's ability to go into heat and achieve a pregnancy. Signs of heat are used to indicate the optimal time for insemination. As such, AIS also reflects HS.

3.3.2 56-day non return rate (NRR)

NRR is a measure of whether a cow or heifer has a second insemination within 56 days of the first insemination. All cows and heifers not inseminated within 56 days after the first insemination are considered pregnant. NRR describes the pregnancy rate. The recording unit is days.

3.3.3 First to last insemination (IFL)

IFL measures the time from the first insemination to the last insemination, and the recording unit is days. IFL is used as a measure of pregnancy rate and HST.

3.3.4 Heat strength (HST)

HST measures a cow's or heifer's ability to show estrous. The trait is measured subjectively by the individual farmer by using a predefined scale, with scores ranging from 1 to 5. HST is only measured in Sweden.

3.3.5 Calving to first insemination (ICF)

This trait is only described for cows and reflects both HST and the ability to return to cycling after calving. The recording unit is days. For a cow to be inseminated, it must first return to cycling after calving.

3.3.6 Fertility treatments 1^{st} 2^{nd} 3^{rd} lactation (FRT)

In Denmark, fertility treatments are recorded by veterinarians and AIS technicians. Fertility treatments are divided into three groups. Group 1 represents hormonal reproductive disorders, and consists of ovarian cysts treatments. Group 2 represents infective reproductive disorders, and consists of recordings of endometritis, metritis, and vaginitis treatments. Group 3 consists of treatments for spontaneous abortion, uterine prolapse, uterine torsion, and other reproductive disorders. A disorder score of 1 is assigned if the female has the corresponding disease; otherwise, the female is assigned a score of 0.
The three lactations are considered as different traits. In Sweden and Finland, similar fertility treatments are recorded by the same strategy; however, veterinary strategies and regulations vary across countries.

3.3.7 Fertility index (FTI)

Female fertility represents a joint Nordic index, which is based upon insemination data collected from Denmark (since 1985), Sweden (since 1982), and Finland (since 1994). The traits that were included in the national FTI calculation were: AIS in cows and heifers; ICF in cows; IFL in cows and heifers; NRR in cows and heifers; and HST in cows and heifers. FTI is considered to reflect the ease with which a cow or heifer is able to conceive.

3.4 Calving traits

For each calving trait described below except the combined indices, four single trait breeding values are calculated: one for each combination of direct effect (D) (the evaluated sire is the father of the calf), maternal effect (M) (the evaluated sire is the maternal grandfather of the calf), and first (F) and later (L) pregnancy.

3.4.1 Calving ease (CE)

In Denmark and Finland, calvings are divided into four groups according to the degree of calving difficulty. Group 1 represents easy calving, without help. Group 2 represents easy calving with help. Group 3 represents difficult calving, without veterinarian assistance. Group 4 represents difficult calving with veterinarian assistance and includes Caesarean sections. In Sweden, two categories are recorded at the time of calving, which are later transformed into the four groups described above.

3.4.2 Calf size (CS)

CS is measured only in Denmark and is divided into four groups. Group 1 represents small calves. Group 2 represents calves below the average size. Group 3 represents calves above the average size. Group 4 represents calves considered large in size.

3.4.3 Stillbirth (SB)

A calf is considered stillborn if it dies before birth or within 24 hours after parturition.

3.4.4 Birth index (BI)

A compound index describing a sire´s total direct additive genetic effect on calving ease by combining DSBF, DSBL, DCEF, DCEL, DCSF, and DCSL.

3.4.5 Calving index (CI)

A compound index describing the maternal additive genetic effect on calving ease by combining MSBF, MSBL, MCEF, MCEL, MCSF, and MCSL.

3.5 Materials

In manuscript I, a total of 36 Danish and Swedish Holstein grandsires were used for a linkage-based QTL analyses with a granddaughter design (Weller et al., 1990). The number of progeny sons tested per grandsire family ranged from 16 to 160 (average: 61). In total, 2182 sons were genotyped.

In manuscript II, only grandsire families from Danish Holstein were used. This study design was used according to the differences between the data recording systems and different treatment strategies across the Nordic countries. The study population consisted of 34 grandsire families. The number of progeny sons tested per grandsire family ranged from 16 to 105 (average: 55). In total, 1888 sires were genotyped. The 34 families analyzed in this manuscript constituted a subset of the same Danish Holstein grandsire families analyzed in manuscript I.

In manuscript III, a total of 3475 sires from Danish, Swedish, and Finnish Holstein with breeding values for female fertility traits were genotyped with a 50k SNP array covering the entire genome. The Holstein data were used for discovery of association between SNPs and female fertility traits. The identified SNP associations were then validated in two other cattle populations, namely Nordic red sires (n = 4998) and Danish Jersey sires (n = 1225).

In manuscript IV, 3918 animals with recorded phenotypes were used. Genotypes of all individuals were imputed to the whole genome sequence WGS level, and sequence variants were applied to GWAS.

In manuscript V, a total of 4258 sires with genotypes and phenotypes for calving traits were used.

3.6 Marker data

In manuscripts I and II, QTL mapping using within-family LA was performed with microsatellite markers. A total of 416 microsatellite markers were genotyped, covering the 29 autosomes. The total length of the linkage map was 3179 cM, with an average marker spacing of 7.64 cM. In manuscript III, the GWAS mapping was performed with SNP markers. All sires were genotyped with the Bovine SNP50 BeadChip, which assayed 54,001 SNP markers (Matukumalli *et al.,* 2009). After filtering for poor quality data, a total of 38545 SNPs on 29 bovine autosomes remained.

In manuscript IV, First HD SNP Beadchip was used. The number of SNPs after imputation to Bovine HD chip was 648,219 and as a second step imputed sequence data for selected regions of the genome were used. Data included, on average, 3000-5000 bi-allelic markers per 1 Mb region (Sahana *et al*. 2012; EAAP presentation).

3.7 Methods QTL Analysis

Manuscript I and II study design

The same statistical analysis, linkage mapping, was used in the first two manuscripts for detection of QTL for female fertility traits.

The traits were analyzed with the linear regression mapping procedure adapted from Haley and Knott (1992). Each trait and chromosome was analyzed separately and tested for the presence of a single QTL affecting one single trait for both the across and within family analysis. The linkage phases of the markers in the grandsires were determined based on the marker types of the sons. Marker allele frequencies were estimated using an expectation–maximization algorithm (Dempster et al., 1977). Segregation probabilities for each position were calculated using all markers on the chromosome simultaneously, together with allele frequencies where segregation was ambiguous. Phenotypes were then regressed onto the segregation probabilities. The following regression model was applied in analyses both across and within families.

$$Y_{ij} = \mu_i + b_i^{(p)} X_{ij}^{(p)} + e_{ij}^{(p)}$$

where Y_{ij} is the single-trait EBV of son j from grandsire i; μ_i is the overall mean of grandsire i; $b_i^{(p)}$ is the regression coefficient within grandsire i at position p; $X_{ij}^{(p)}$ is the probability of QTL allele 1 being transmitted from grandsire i, given all of the informative markers of son j; and $e_{ij}^{(p)}$ is the residual effect, given QTL position p.

Manuscripts I and II: Significance level

Two types of test statistics were calculated: an *F*-statistic for each grandsire family, chromosome, and trait; and a joint *F*-statistic for each trait and chromosome across grandsire families. A QTL was considered significant if it exceeded the chromosome threshold by 5%, which was determined by a

permutation test with 1000 permutations (Churchill and Doerge, 1994). Churchill and Doerge (1994) introduced the use permutation testing in QTL mapping. By many times analyzing datasets where phenotypes have been randomly re-assigned to genotypes the distribution of the maximal test statistic under the null hypothesis are obtained.

Manuscript III: Study design

A SNP-by-SNP analysis was carried out, with each SNP being tested sequentially for association with phenotypes. The following linear mixed model was used to estimate SNP effects in the Nordic Holstein population:

$$y_{ij} = \mu + bx_{ij} + s_i + e_{ij}$$

where y_{ij} is the single-trait EBV of individual j, belonging to the half-sib (sire) family i;, μ is the general mean; b is the allelic substitution effect; x_{ij} is the number of copies of an allele, with an arbitrary labeling of the SNP count in individual j (corresponding to 0, 1 or 2 copies); s_i is the random effect of the i -th half-sib family, which is assumed to have covariances according to the relationship among sires (such that $s = \{si\}$ is normally distributed $N(0, \sigma_g^2 A_s)$, where σ_g^2 is the polygenic genetic variance and $A_s$is the additive relationship matrix among the sires derived from the pedigree), and e_{ij} is a random residual of individual i , which is assumed to follow a normal distribution with mean zero and unknown variance. Testing was done by *t*-test against a null hypothesis of H_0: b = 0.

Manuscript III, Significance level

The significance threshold was determined using a Bonferroni correction (within trait using 38,545 SNP markers). The genome-wide significance threshold was 1.3×10^{-6}, and was calculated by dividing the nominal significance threshold of 0.05 by the total numbers of SNPs included in the analyses. For the LD studies in manuscript III-IV permutation testing were replaced by the use of the Bonferroni correction (Dunn, 1961). This testing assumes independence of tests. However, this is usually not the case with GWAS as there is LD between adjacent markers. Correlation between tests means that the effective number of independent tests is less than the number of markers. Thus use of the Bonferroni correction is conservative. The nominal test level is 5% throughout manuscript III-V under this model assumption,

which reduces the risk of making a type I error to no more than 5% experiment wise.

Manuscript IV: Study design

Association analysis was carried out for each trait using a linear mixed model, which fitted a fixed effect for each SNP along with a random polygenic effect. The random polygenic effect was included to account for residual genetic variance that was not explained otherwise by the SNP in the model or the population structure because of the presence of half-sib families (Yu *et al.*, 2006).

$$y = \mu + S\alpha + Zu + e$$

where y is a vector of de-regressed breeding values; μ is the general mean; Z is a matrix relating additive polygenic effects to individuals; u is a vector of additive polygenic effects; α is a vector of SNP effects; S is an incidence matrix relating α to the individuals; and e is a vector of random residual effects. The random variables u and e are assumed to be normally distributed. Specifically, u is normally distributed with $(0, \sigma_g^2$ A), where σ_g^2 is the polygenic genetic variance, and A is the additive relationship matrix derived from the pedigree.

Manuscript V: Study design

The linear model used was:

$$y_i = \mu + bx_i + s_i + e_i$$

where y_i is the single-trait EBV of individual i; μ is the general mean; x_i is a count in individual i for one of the two alleles (with an arbitrary labeling);, b is the allele substitution effect;, s_i is the fixed effect of the sire of individual i; and e_i is a random residual of individual i, which was assumed to follow a normal distribution with mean zero and unknown variance. Testing was done with a *t*-test against a null hypothesis of H_0: b = 0.

Manuscript V: Significance level

The same significance level as used in manuscript III was applied to the analysis of the comprehensive combined dataset. A chromosome-wise

significance threshold was used for the reference population. A threshold of >0.05 was used for the validation population.

3.8 Results

In manuscript I, a total of 26 QTL were identified on 17 chromosomes. Among the QTL reported here, eight had effects on FRT, two on HST, five on NRR, two on AIS, and nine on ICF.

In manuscript II, we identified 12 QTL regions with effects on nine different fertility traits. None of these QTL harbored QTL for milk production or milk composition traits. However, an additional 16 selected genomic regions containing QTL for fertility also harbored QTL for milk production or milk composition traits. The FRT trait was decomposed into its underlying components: infective reproductive disorders, hormonal reproductive disorders, and other reproductive disorders. No QTL were detected for the subtrait of spontaneous abortion. Four different QTL were identified for the trait of retained placenta.

In manuscript III, 4474 significant SNPs were associated with eight different female fertility traits. Of these, 836 SNPs were validated in the Nordic Red breed, 686 SNPs were validated in the Nordic Jersey population, and 152 SNPs were validated in both breeds. There was evidence for QTL where multiple traits were segregating, and many SNPs were validated on chromosomes 1, 4, 7, 9, 11, and 13 (Fig. 3).

In manuscript V, 424 significant SNPs were found by a genome-wide scan for 14 calving traits. Of these, 321 SNPs were confirmed in the validation dataset (Fig. 4).

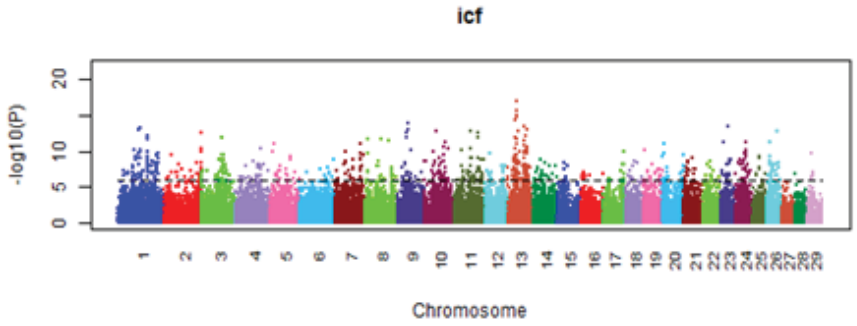

Figure 3. Illustration of GWAS for the trait calving to first insemination (ICF) The dotted line represents genome wide significance. The data was analysed with the BovineSNP50 Beadchip, which assayed 54,001 SNP markers (Matukumalli *et al.,* 2009).

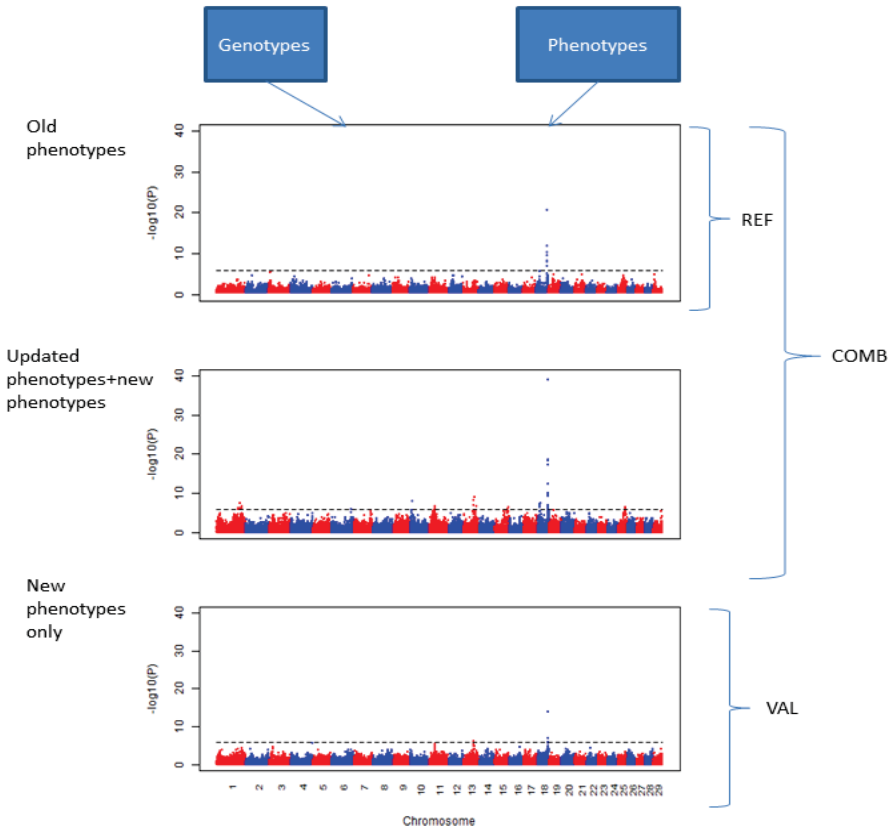

Figure 4. Illustration of a validation study (from manuscript V), with association results plotted for the trait of birth index. The left side indicates the data used. The right side shows the names of the different datasets, as referred to in manuscript V. The figure illustrates the number of markers reaching significance in the different datasets (*i.e.*, where the dotted line exceeds the significance threshold depicted by the broken line). The REF dataset included 2,219 animals and the VAL dataset included 2,039 animals.

4 General discussion

The identification of several QTL through this thesis work, along with the parallel progression of genome-based technologies and analytical approaches, have provided the opportunity to fine-map genomic regions influencing female fertility in cattle. These data contribute to a deeper understanding of the genetic basis of female reproductive biology and improve genomic predictions within and across breeds.

4.1 Studies

In manuscript I, several QTL were identified by LA. Strong evidence of segregating QTL for several of the female fertility traits were found on BTA1, BTA7, and BTA10. In the association study presented in manuscript III one can assume that the significant associations on BTA1 represent the same QTL found in the LA study in manuscript I. Other studies have also detected QTLs and associations in the same areas on BTA1 (Sahana et al., 2010, Ben-Jema *et al.,* 2008, Schulman *et al.,* 2011, Boichard *et al.,* 2003). This constitutes strong evidence for QTL for fertility segregating on BTA1. On BTA13 we found QTL segregating in manuscript I. Our subsequent association study found strong association with fertility in close vicinity of the same region. Again other studies also detected QTL effects in the same region on BTA13 (Sahana *et al.,* 2010, Schulman *et al.,* 2011 and Huang *et al.,* 2010). This makes BTA13 interesting for further analysis. This QTL affects the trait ICF which reflects the ability to show heat and the ability to return to cycling after calving. In the fine-mapped region, augmenting the association with full sequence in the region identified intragenic variants as the most likely causative variants. This could suggest that the causative variant control some genes involved in the physiology of heat or returning to cycling. The top SNP (most significant SNP) on BTA13 explained 3% of the variance of the de-regressed proof (used as a

response variable in the sequence analysis in manuscript IV). The minor allele frequency (MAF) for this SNP was 38.5% i.e. in high frequency in the population. The frequency and effect size indicate that this SNP can be exploited in breeding.

Linkage analysis as used in manuscript I, has low precision for determining the QTL position because it relies solely on recombination events that occur between two generations (i.e., the lowest possible number). The low number of recombination events and the use of relatively few markers prevent localization of QTL to segments smaller than about 10 to 20 cM. Besides, the localization of the QTL in a family will depend on the informative markers for that family. Therefore, it is generally difficult to draw any firm conclusion concerning overlap in QTL positions from manuscript I and the later association studies for fertility traits in this thesis. The power of LA studies depends on the number of segregating families. Typically, when working with functional traits, most QTL effects will be small compared to the amount of residual error. Low power means that estimates of effects of the detected QTL will suffer from "the winner's curse". Therefore, the QTL that are detected will represent those with overestimated effect sizes. Furthermore, false positives may account for a significant fraction of the QTLs detected. Thus, it is important to find ways to validate the detected QTLs and/or identified associations.

An unfavorable genetic correlation is known to exist between the traits of female fertility and those of milk production. Therefore, it is expected that there are overlapping QTL regions for these two traits across the genome. This could be either due to pleiotropic effect of a QTL affecting either traits or separate QTL affecting each trait but linked together. If we could identify the QTL that affect both traits (pleiotropic) or that only affect fertility (linked), then it may be possible to optimize simultaneous breeding for the two traits. In manuscript II, 27 genomic regions harboring QTL for fertility traits were screened for different milk production traits, and 16 of those regions were found also to harbor QTL for milk production traits. This finding is in line with our expectations, given that the genetic correlation between fertility and production is ranging from about -0.2 to -0.5 (Roxström *et al.,* 2001 and Pryce *et al.,*1997). Using a combined breeding goal in the selection scheme a positive genetic gain in fertility is most likely to come from the QTL that only affect fertility.

It remains unknown whether it is possible to break parts of the negative genetic correlation to increase the benefit for both trait groups. To do so, it is first necessary to know if overlapping QTL are due to pleiotropic genes affecting both traits or linked genes each affecting separate traits; however, the

confidence intervals for QTL achieved in LA studies are not suitable for this type of analysis.

It would be highly beneficial to define the traits more accurately and to decompose the traits into sub-traits, to gain a better understanding of the physiology behind them. In manuscript II, the trait "fertility treatments" was dissected. The FRT trait was analyzed to identify the underlying sub-traits. Reduction of fertility related disease is highly relevant due to its immediate contribution to better animal welfare and reduction of veterinary costs. The sub-trait of retained placenta was represented by four different QTL across the genome. It would have been highly useful to combine data across countries, so that the breeding values could be more reliably predicted. Unfortunately, the Nordic countries employ different veterinary strategies for this feature, which precluded combination of the data. No QTL were detected for the sub-trait of spontaneous abortion. However, the frequency of reported spontaneous abortion is believed to be low, due to the inherent underestimation of this feature, which is very difficult to detect in early pregnancy.

In manuscript III, fine-mapping of fertility traits were conducted by using the Bovine SNP50 BeadChip, and the results were validated in two different breeds. The strongest evidence of associations was found on chromosomes BTA1, BTA4, BTA7, BTA9, BTA11, and BTA13.

Several QTL and associations have been detected through the studies comprising this thesis work. It is challenging to conclude the number or nature of the loci involved in a particular trait. This type of information is even more challenging to obtain for female fertility for which trait definitions were not generated with the aim of determining the physiology behind the trait but with the aim of improving a breeding scheme. Therefore, many physiological sub-traits are combined in the definition of a single trait/phenotype. On many positions on the genome, several female fertility traits are significantly associated with the QTL. That is to be expected as the female fertility traits are genetically correlated. It is however, not known how common non additive gene action (dominance vs. epistasis) in determining a fertility trait value.

The genetic correlation between cow and heifer fertility traits in the Nordic countries is around 0.4 (Årsstatistik Avl, 2011/12). This indicates that cows and heifers have partially separate genetic bases. This is reflected in the results manuscript I and III, in where only limited overlap between cow and heifer fertility QTL was found.

In manuscript IV, the fine-mapping objective was taken two steps further. First, with the analysis of a high-density SNP array (including >777,000 evenly spaced SNPs across the entire bovine genome). Secondly, genotypes were imputed to the full sequence level. Sequence variants were then associated to

phenotypes, to narrow down the genomic regions on BTA4 and BTA13 in the search for possible causative mutations (Figure 3). In total, nine genes were annotated in the focused region on BTA4 (see Table 1 in manuscript IV). A search of the Ensemble database indicated semaphorin-3C (SEMA3C) as the most promising candidate gene, according to the description of its biological function. SEMA3C binds to plexin family members to exert its effects as a regulator of developmental processes. SEMA3C is essentially involved in cardiovascular development during embryogenesis and plays an important role in axon growth and guidance through its functions as an attractant for growing axons (http://www.uniprot.org/uniprot/A7MB70). However, the SNP markers within this gene were not found among the genome-wide significant markers.

When the test statistics of the SNP markers (i.e., $-\log_{10}$(P-value)) was considered, CD36 was identified as the most promising candidate gene. CD36 is involved in cell adhesion, a key process of many physiological processes. However, the mechanism by which CD36 influences fertility in cattle remains unknown. In general, only some of the SNP markers of this study were located within a gene and the SNP markers with the highest association to AISC and IFLC (on BTA4) were not annotated, which highlight the need for a better annotation of the bovine genome. Currently, the sites of gene transcription, initiation, termination, and differential splicing remain to be fully defined, even though the tools and resources for annotation and gene discovery are available for the bovine genome (Brent, 2005; Childers *et al.,* 2011). Rapid amplification of cDNA ends (RACE, a PCR-based method) is a well-established tool for empirically annotating the transcription start / end sites for a single gene. RACE has been successfully supplied for large-scale structural transcript annotation (Salehi-Ashtiani *et al.,* 2009).

Although we used the 'full' sequence-level variants in our analysis, we could not with certainty, pinpoint any particular causal factor underlying the fertility QTL. Several reasons may explain this issue. First, half of the total genetic variants identified in the WGS had been filtered out, due to low quality scores at the sequencing level or low imputation accuracy. Second, all of the variants that were not bi-allelic had been removed, due to the inherent limitation of imputing such data to all of the genotyped animals. Therefore, there was a high risk that the causal factor had not been included in the analysis. Third, there were many SNPs with very high $-\log_{10}$(P-values), as a result of the high LD among these SNPs. Therefore, our ability to make any firm conclusion on which SNPs to pick for further analysis was limited.

A number of further studies may help in the identification of causal variants. Re-sequencing animals in the region of interest and among multiple breeds may also prove helpful in the detection of SNPs and causative mutations

(Boyko, 2011). This could help to validate the set of detected SNPs across breeds, which could be a useful approach for identifying causative mutations. For example, the selected SNPs from the Nordic Holstein cattle population can be validated by comparison to sequence data from Nordic Red or Danish Jersey cattle. Combining the association results across these three breeds may help in lowering the number of candidate Quantitative Trait Nucleotide (QTN). Another potentially useful strategy is to genotype a large number of cows with the most promising candidate QTN. This approach would avoid imputation errors and provide a statistically independent data set, while providing high power for distinguishing between candidates QTN. As stated above, there might be other transcripts in the analyzed area that remain unrecognized because not all of the genes are annotated. Indeed, the genomic structures of candidate genes may differ according to species-specific and tissue-specific expression patterns, or even variations in expression during different developmental stages. Expression analysis (transcriptome sequencing data) may prove useful to investigate the expression of genes in relevant organs at different developmental stages. Evolutionary analysis of nucleotide sequences may also be useful, in particular for predicting the deleteriousness or potential functions of noncoding variants (Cooper and Shendure, 2011). For example, on BTA13, only transcripts located in the noncoding regions were found. The causal variants for these genes may exist in the regulatory elements of noncoding regions (Keane *et al.,* 2011). The most common regulatory elements are enhancers, but other regulatory sequences, such as promoters, insulators and silencers, may also be involved (Maston *et al.,* 2006).

In manuscript V, the strategies used for validation studies were evaluated. This type of study is highly relevant for dairy cattle populations, in particular, because new data are continuously accumulating and made available in the national databases. Our findings from a GWAS follow-up study indicated that the choice of method for validation study depends on the objective of the study. If the aim of the study is to detect QTL, it is recommended that all of the presently available data be analyzed. If the aim is to identify causal mutations, then confirmation of SNP association should be performed with the new data.

4.2 Markers

This thesis spans years where marker technologies have developed very quickly. At the start of my PhD project we only had access to microsatellite markers with 416 markers distributed across the whole genome. (Manuscripts I and II). The average spacing of microsatellites was around 20 cM in linkage studies. Next technology progressed to use of DNA chips with first (2008)

54,001 marker across the genome and more recently more than 777,000. Now, with whole genome sequence data used to fine map certain areas of the genome we now have 8,000-9,000 bi-allelic markers per Mb. This represents a significant difference in density of markers available for mapping compared for the last manuscript in this thesis. With the lower density microsatellite maps, QTL mapping was conducted using linkage analysis, which makes use of co-segregation between marker and QTL within family. Now with the increase in number of markers it has been possible to perform association studies which make use of population based LD.

4.3 Choice of model

The underlying statistical method used for our linkage and association studies was regression analysis. The choice of model depends on factors such as properties of the "phenotypes" (predicted breeding values, de-regressed proofs, models used for prediction etc.) used as the response variable, the distribution of accuracies for the phenotypes, the family/population structure of the individuals and the marker density. In QTL mapping in dairy cattle response variables are EBVs or de-regressed proofs of sires, both of which are estimated from the performance of a sire's progeny with or without information from other relatives. The reliability of breeding values from a sire depends on the number of that sire's progeny with phenotype records. This is considered when using de-regressed proofs. STBVs have been used throughout this thesis with the exception of manuscript IV were de-regressed proofs were used as response variable. STBV was used to avoid contamination with information from correlated traits. Without this step, it would not have been possible to interpret the presence of QTL for a particular trait, as the presence of one of the QTL might have been inferred based on a correlated trait. Because the breeding values represent the additive genetic effect, the QTL mapping model can only detect the additive genetic effect of the QTL. To detect dominance effects the actual phenotypes would have to be analyzed.

4.4 Gene Mapping methods

In our LA studies, we attempted to identify markers that co-segregate with trait values within families. In contrast, in our association studies, we sought to identify a direct correlation between a specific genetic variation and a trait variation among a group of animals, with the aim of implicating a causal role for the variant. The fact that LA is based on the relationship between markers and phenotypes allows for the identification of a trait locus that is nearby the

marker locus. On the other hand, association between a marker of genomic variation and the phenotype of interest at the population level is observable under two circumstances. In the first, the functional variant is measured directly; and in the second, the marker variant is in LD with the actual functional variant. Direct measurement of the causal factor increases the power of the association study. However, depending on how dense and detailed the marker map is, it may be more likely that neighboring polymorphisms will be identified in LD. By using whole sequence data, as we did in manuscript IV, it is feasible to map the functional variant directly. However, only bi-allelic markers mapped with the sequence (3,000-5,000 bi-allelic markers per 1 Mb region (Sahana *et al.,* 2012; EAAP presentation)) because imputation methods only work for bi-allelic variants. Therefore, information on other markers remains to be investigated.

False-positive associations may result from familial or population stratification, which frequently occurs in many domestic animal populations. When populations are divided into subgroups, each of which may differ in marker allele frequency and disease frequency, and the combined population is analyzed without accounting for the structure, spurious associations between markers and traits may occur. These associations do not reflect the proximity of the marker and QTL. In a mapping sense, they reflect false-positive associations. This issue can be resolved by applying pedigree information to the model used for analyzing the data (Yu *et al.,* 2006).

False positives can also arise from incorrect adjustment for multiple testing. For more information regarding corrections for multiple testing see section 3.6

4.5 Applied use in breeding

The animal breeding industry in the Nordic countries has applied genomic selection, based on the individual's marker information, over approximately the last 5 years. Genomic selection is based on the individual´s marker information. In this strategy, the markers' effects are predicted in a (preferably large) reference population of animals with marker information and reliably predicted breeding values. A breeding value for a selection candidate can then be predicted from its genetic markers by adding their predicted effects. Genomic prediction can help to select the animals for the next generation with the most desirable traits. Information about identified QTL can be applied to this genome selection scheme to further enhance the prediction reliability (Boichard *et al.,* 2012).

However, accurately estimating the prediction equations becomes complicated when the number of SNPs is much larger than the number of phenotypes. Thus,

applying WGS data to a genomic prediction scheme has proven a challenge. Indeed, the sequence data are useful for identifying the regions on the genome with influence on female fertility (by association studies), and the availability of WGS data could identify the causal mutations underlying female fertility not only markers in LD with the causal mutation. This information could be used in a genomic prediction scheme with higher weights being put on certain genomic regions or causal variants that influence female fertility. There is benefit in such genomic predictions accounting for identified causative mutations, because they are expected to have better predictive ability across breeds, particularly in highly structured populations, or when candidate progenitors are distantly related to the reference.

5 Conclusions

- Several QTL have been identified in this thesis by LA study: BTA1, BTA7, and BTA10 showed the strongest evidence of segregating QTL for several female fertility traits.

- Twelve QTL regions with effects on nine different cattle fertility traits did not harbor coinciding QTL for milk production traits. Sixteen genomic regions harboring female fertility traits also contained QTL for milk production traits.

- Dissection of the fertility treatments trait identified no QTL for the sub-trait of spontaneous abortion. However, four QTL were identified across the genome for the sub-trait of retained placenta.

- BTA1, BTA4, BTA7, BTA9, BTA11, and BTA13 showed the strongest evidence of association with cattle fertility traits across breeds in a validation study.

- Focused (narrowed-down) regions of interest on BTA4 and BTA13 were identified after applying a denser marker map and conducting sequence data analysis to identify annotated genes.

- Finally, selecting an optimal method for processing new accumulated data from the ever-expanding national databases depends on the objective of the study. If the aim is to detect QTL, it is recommended to analyze all of the data that are presently available. If the aim is to identify causal mutations, a validation study design is recommended.

6 Future outlook

A key goal of dairy cattle breeding is to increase the reproductive performance and decrease the amount of disease in the next generation. Both of these features benefit the welfare of the animals, the output of dairy product, and the efficiency of the dairy industry's future breeding efforts.
Additional fine mapping using high-density sequencing array technology and WGS data is expected to facilitate even finer mapping of genomic areas affecting female fertility in cattle. For the results of this thesis work to facilitate such future efforts, a number of steps should be taken:

- Directly measured phenotypes and genotypes of cows are necessary to provide a more direct link between phenotype and genotype and to identify the temporal-specific aspects of gene expression.

- Identification and analysis of phenotypic measures that reflect more directly the physiologic background of the reproduction traits could be helpful for determining the precise physiological aspect represented by a specific QTL. For example, progesterone level has been suggested as a measure of HST.

- Application of genomic data from other breeds and species of cattle may be useful for increasing the overall annotations and knowledge of functions for specific genes with potential association to a fertility trait.

- Gaining more information on the animals in the research population i.e. add additional bulls with many daughters in order to gain more power to the analysis; this is also shown in the last manuscript (V).

- Genomic prediction can help to select the animals for the next generation more efficiently; however there are challenges in estimating the prediction equations when the number of SNPs is much larger than the number of phenotypes. So far the whole genome sequence data is therefore difficult to use in a genomic prediction scheme. The sequence data can however be used to identify the regions on the genome influencing female fertility. This information could be used in a genomic selection scheme where weights can be put on certain genomic regions which influence female fertility.

7 Sammanfattning

Mål

Under flera decennier har honlig fruktsamhet hos kor av Holsteinras försämrats samtidigt som mjölkavkastningen har stigit kraftigt, även om man kan se ett trendbrott de senaste åren. Det finns flera orsaker till denna minskning: arvbarheten är låg, i tjurindexet har mjölkavkastning en högre ekonomisk vikt än fruktsamhet och före år 2000 saknade många utländska Holsteintjurar avelsvärden för honlig fruktsamhet. Dessutom är den genetiska korrelationen mellan fruktsamhetsegenskaper och produktionsegenskaper ogynnsamma. Genetisk korrelation mäter hur samma eller kopplade genetiska faktorer påverkar två egenskaper på samma gång.

Målsättningen med denna studie var att i första hand med hjälp av genetiska markörer kunna förstå vilka gener som styr reproduktion. En genetisk markör är en bit DNA där det finns variation mellan individer som man har utvecklat en metod för att observera. I andra hand var målet att se om dessa markörer även interagerar med produktionsegenskaper. Den tredje målsättningen var att utvärdera en metod för att bekräfta de markörer man har funnit associerade med reproduktionsegenskaper.

Resultat

I den första studien sökte vi efter genetiska markörer för honlig fruktsamhet. Vi identifierade totalt 26 markörer på 17 kromosomer som hade effekt på honlig fertilitet. På kromosom 1, 7 10 och 26 fann vi hos olika familjer flera sådana markörer inom begränsade områden. Våra resultat överensstämmer med resultat från andra undersökningar på Holstein.

I den andra studien undersökte vi om det i områden med markörer för fruktsamhetsegenskaper även fanns markörer för produktionsegenskaper. Vi identifierade 16 kromosomregioner med markörer för båda typerna av egenskaper och 12 områdena var unika för fruktsamhetsegenskaperna. Detta

betyder att större delen av de områden vi fann i genomet påverkar båda typerna av egenskaper.
I den tredje studien hade vi tillgång till många fler markörer. 4,474 markörer var associerade med honlig fruktsamhet och här hade vi möjlighet att mer precist identifiera var på genomet dessa markörer befann sig. Vi försökte sedan bekräfta associationerna i de röda nordiska mjölkoraserna och i Dansk Jersey. 152 markörer kunde bekräftas i alla tre raserna. De mest intressanta kromosomer var kromosom nummer 1, 4,7, 9, 11 och 13.
I den fjärde studien analyserade vi mindre regioner på kromosom 4 och 13 med sekvensdata där hela genomet sekvenserats. På båda kromosomerna fanns det gener i området men fler analyser behövs för att kunna peka ut någon uppenbar kandidatgen.
I den femte studien bekräftade vi markörassociationer för kalvningsegenskaper, Vi kunde bekräfta 68 % av markörerna i ett valideringsdata-set hos Holstein.

Material och metoder

I Norden har vi unika förutsättningar att studera honlig fruktsamhet hos nötkreatur. Egenskaper registrerade i Kokontrollen och avelsvärden skattade av Nordisk Avelsvärdering, NAV, utgör underlag för studien. Vi har undersökt följande egenskaper: antal semineringar per serie, 56 dagars non-return %, brunststyrka (endast svenska tjurar), intervall mellan kalvning och första seminering samt fruktsamhetsbehandlingar.
I den första undersökningen ingick 2182 tjurar från Sverige och Danmark. Data analyserades både inom och mellan familjer med ca 416 markörer.
I den andra studien utgick vi från de områden med genetiska markörer som vi funnit i den första studien och undersökte om dessa områden även påverkade produktionsegenskaper. Det ingick 1888 danska tjurar och samma genetiska markörer och analysmetod användes som i den första studien.
I en tredje studie försökte vi finkartlägga de områden på genomet som påverkar honlig fruktsamhet genom en associations studie. Här hade vi möjlighet att använda oss av ca 3500 tjurar och ca 38500 markörer.
I den fjärde studien valde vi ut regioner med de mest övertygande resultat från den tredje studien, för att om möjligt finna de underliggande generna i dessa regioner. I den femte studien använde vi samma metod som i tredje manuskriptet, associationsstudie, för att validera associationer för kalvningsegenskaper.

Slutsatser

Användning av genomisk information i avelsarbetet kan möjliggöra en säker och tidig selektion av de djur som har den största potentialen som avelsdjur.

Kunskap om olika markörer i kromosomsegment (QTL) och deras inverkan på olika egenskaper kan användas för att välja ut vilka regioner som skall ingå i och läggas vikt på vid avelsvärdering. Resultatet från vår studie visar att det är delvis samma och delvis olika regioner på kromosomerna som har QTL för fruktsamhets- och produktionsegenskaper. I QTL-regioner kan sekvensdata ge möjlighet att identifiera specifika kromosomsegment för respektive egenskap och på så vis öka förståelsen för vad det är som orsakar variationer i egenskaperna.

Markörinformation används idag vid genomisk selektion. Därför går det att identifiera tjurkalvar med övervägande positiva anlag för både produktions- och reproduktionsegenskaper.

Våra studier syftar också till att ge en bättre förståelse av de biologiska processer som via specifika kromosomsegment påverkar fertilitet.

8 References

Aamand GP: Nordisk Avlsværdivurdering (NAV) – Joint Nordic Genetic Evaluation. 2005, URL: http://www.nordicebv.info/NR/rdonlyres/33A7018F-0BDD-4A1F-A3B8-627170AE6F33/0/0016GAPSkara24102005.pdf

Ahlman, T., Berglund, B., Rydhmer, R. and Strandberg, E: Culling reasons in organic and conventional dairy herds and genotype by environment interaction for longevity. J. Dairy Sci. 2011, 94:1568–1575.

Årsstatistik Avl 2011/12: https://www.landbrugsinfo.dk/Kvaeg/Avl/Avlsanalyser/Sider/aarsstat_2012.pdf

Ben-Jemaa S, Fritz S, Guillaume F, Druet T, Denis C, Eggen A, Gautier M: Detection of quantitative trait loci affecting non-return rate in French dairy cattle. J. Anim. Breed. Genet. 2008, 125:280-288.

Boichard, D., Guillaume, F., Baur, A., Croiseau, P., Rossignol, M.N., Boscher, M.Y., Druet, T., Genestout, L., Colleau, J.J., Journaux, L., Ducrocq, V. and Fritz, S: Genomic selection in French dairy cattle. Anim. Prod. Sci. 2012, 52:115-120.

Boichard D, Grohs C, Bourgeois F, Cerqueira fF, Faugeras R, et al: Detection of genes influencing economic traits in three French dairy cattle breeds. Genet. Sel. Evol. 2003, 35:77-101

Boyko AR: The domestic dog: man's best friend in the genomic era. Genome Biol. 2011, 12:216

Brent MR: Genome annotation past, present, and future: how to define an ORF at each locus. Genome Res. 2005, 15:1777-1786.

Berglund B, Steinbock L, Elvander M: Causes of stillbirth and time of death in Swedish Holstein calves examined post mortem. Acta Vet Scand. 2003; 44:111-20.

Carroll SB: Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008, 134:25-36.

Childers CP, Reese JT, Sundaram JP, Vile DC, Dickens CM, Childs KL, Salih H,Bennett AK, Hagen DE, Adelson DL, Elsik CG: Bovine Genome

Database: integrated tools for genome annotation and discovery. Nucleic Acids Res. 2011, 39(Database issue):D830-D834.
Churchill GA, Doerge RW: Empirical threshold values for quantitative trait mapping. Genetics. 1994, 138:963-971.
Cooper GM, Shendure J: Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat Rev Genet. 2011, 12:628-640.
Dempster AP, Laird NM, Rubin, DB: Maximum likelihood from incomplete data via the EM algorithm. J. Royal Stat. Soc. Ser. B (Methodological) 1977, 39:1-38.
Dunn OJ: Multiple comparisons among means. J. Am. Stat. Assoc. 1961, 56:52-64.
Gilbert RO, Shin ST, Guard CL, Erb HN, Frajblat M: Prevalence of endometritis and its effects on reproductive performance of dairy cows. Theriogenology. 2005, 64:1879-1888.
Haley CS, Knott SA: A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 1992, 69:315-324.
Hayes BJ, Lewin HA, Goddard ME: The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. Trends Genet. 2013, 29:206-214.
Hoekstra HE, Coyne JA: The locus of evolution: evo devo and the genetics of adaptation. Evolution. 2007, 61:995-1016.
Huang W, Kirkpatrick BW, Rosa GJ, Khatib H: A genome-wide association study using selective DNA pooling identifies candidate markers for fertility in Holstein cattle. Anim. Genet. 2010, 41:570-578.
Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, Biagi TM, Patterson N, Pielberg GR, Kulbokas EJ 3rd, Comstock KE, Keller ET, Mesirov JP, von Euler H, Kämpe O, Hedhammar A, Lander ES, Andersson G, Andersson L, Lindblad-Toh K: Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet. 2007, 39:1321-1328
Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A,Slater G, Goodson M, Furlotte NA, Eskin E, Nellåker C, Whitley H, Cleak J,Janowitz D, Hernandez-Pliego P, Edwards A, Belgard TG, Oliver PL, McIntyre RE,Bhomra A, Nicod J, Gan X, Yuan W, van der Weyden L, Steward CA, Bala S, Stalker J, Mott R, Durbin R, Jackson IJ, Czechanski A, Guerra-Assunção JA, Donahue LR, Reinholdt LG, Payseur BA, Ponting CP, Birney E, Flint J, Adams DJ. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011, 477:289-294.
LeBlanc SJ: Postpartum uterine disease and dairy herd reproductive performance: a review. Vet J. 2008, 176:102-114.
Lucy MC: Reproductive loss in high-producing dairy cattle: where will it end? J Dairy Sci. 2001, 84:1277-1293.

Maston GA, Evans SK, Green MR: Transcriptional regulatory elements in the Human genome. Annu. Rev. Genomics Hum. Genet. 2006, 7:29-59.
Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS, Van Tassell CP: Development and characterization of a high density SNP genotyping assay for cattle. PLoS One, 2009, 4:e5350.
Pryce, J. E. R. F. Veerkamp, R Thompson, W.G Hill, and G. Simm. 1997. Genetic aspects of common health disorders and measures of fertility in Holstein Friesian dairy cattle. Anim. Sci. 65:353-360.
Roxström A, Strandberg E, Berglund B, Emanuelson U, Philipsson J: Genetic and Environmental Correlations Among Female Fertility Traits and Milk Production in Different Parities of Swedish Red and White Dairy Cattle. Acta Agric. Scand., Sect. A. 2001, 51:7-14.
Royal MD, Smith RF, Friggens NC: Fertility in dairy cows: bridging the gaps. Animal, 2008, 2:1101-1103.
Salehi-Ashtiani K, Lin C, Hao T, Shen Y, Szeto D, Yang X, Ghamsari L, Lee H,Fan C, Murray RR, Milstein S, Svrzikapa N, Cusick ME, Roth FP, Hill DE, Vidal M: Large-scale RACE approach for proactive experimental definition of C. elegansORFeome. Genome Res. 2009, 19:2334-2342.
Sahana G, Guldbrandtsen B, Bendixen C, Lund MS: Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. Anim. Genet. 2010, 41:579-588.
Sahana G, Guldbrandtsen B, Lund MS. Refining QTL with high-density SNP genotyping and whole genome sequence in three cattle breeds. Book of Abstracts. Vol. 18 2012. udg. The Netherlands : Wageningen Academic Publishers, 2012. s. 354.
Schulman NF, Sahana G, Iso-Touru T, McKay SD, Schnabel RD, Lund MS, Taylor JF, Virta J, Vilkki JH: Mapping of fertility traits in Finnish Ayrshire by genome-wide association analysis. Anim Genet. 2011, 42:263-269.
Shook GE: Major advances in determining appropriate selection goals. J Dairy Sci. 2006, 89:1349-1361.
Sun C, Madsen P, Nielsen US, Zhang Y, Lund MS, Su G: Comparison between a sire model and an animal model for genetic evaluation of fertility traits in Danish Holstein population. J Dairy Sci. 2009, 92:4063-4071.
VanRaden PM, Olson KM, Null DJ, Hutchison JL: Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J Dairy Sci. 2011, 94: 6153-6161.
Weller JI, Kashi Y, Soller M: Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J Dairy Sci. 1990, 73: 2525-2537.
Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES: A

unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006, 38: 203-208.

9 Acknowledgements

This study was conducted at the Department of Molecular Biology and Genetics, Center for Quantitative Genetics and Genomics, Aarhus University, Denmark and at the Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics in Sweden as a double degree. VikingGenetics is greatly acknowledged for my employment.

Till min farfar
Jag lovade att du skulle bli den första att bli bjuden till min disputation, jag hoppas att du ser från ovan och ja, jag tror man kan säga att jag har slutat skolan nu.

Till min handledargrupp
Min svenska huvudhandledare Lena för ditt engagemang och förmåga att konstruktivt få saker på plats.
Min danske hovedvejleder Mogens for din måde at formidle din faglige viden og sætte tingene i rette sammenhæng.
Bernt for dit engagement og din entusiasme for nye ideer.
Hans för ditt stöd och intresse för projektet.

गौतम सहना,
अवश्य होने पर उपलब्द होने केलिए ,
हर प्रश्न की तुरंत उत्तर केलिए , तांत्रिक मदत केलिए ,
मैं आपकी बहुत बहुत आभारी हूँ.

Jeg vil gerne takke alle mine kolleger og venner ved Instituttet for et interessant og fantastisk arbejdsmiljø.

Specielt; Hanne Skovsgaard Pedersen, Louise Dybdahl, Jørn Rind og Elise Norberg

Min mamma och pappa för oändlig uppmuntran och stöd.

Wim en Cori voor de eindeloze steun en omdat jullie altijd klaar staan om te helpen. Jullie zorgen voor mijn rust en vredigheid.

Tina Skau Nielsen og Jehan Ettema fordi de altid er der.

Tack till hela besättningen där hemma, Cavour för avslappnande turer och för inga frågor eller krav ställda.

Mijn gezin: Bart, Jakob, Allis en Erik, van jullie komt alles! Jullie laten me inzien wat het belangrijkste is in het leven en dat een proefschrift misschien toch niet zo heel veel betekent.