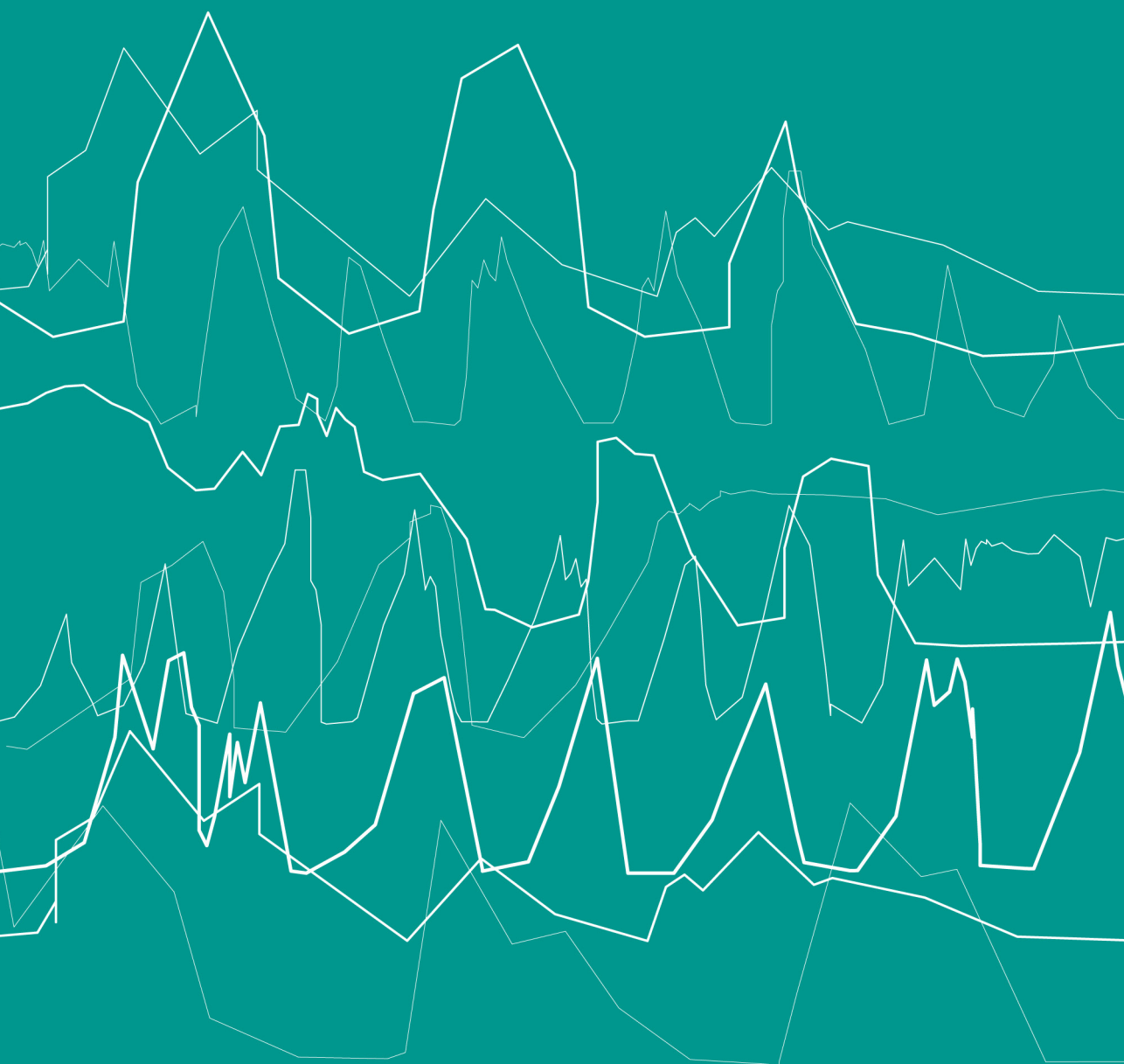


Milk progesterone measures to improve genomic selection for fertility in dairy cows

Amabel M. M. Tenghe





ISSN 1652-6880

ISBN (print version): 978-91-576-8847-7

ISBN (electronic version): 978-91-576-8848-4

**Milk progesterone measures to improve
genomic selection for fertility in dairy cows**

Amabel Manyu Mefru Tenghe

Thesis committee

Promotors

Prof. Dr R.F. Veerkamp
Professor of Numerical Genetics and Genomics
Wageningen University & Research

Prof. Dr B. Berglund
Professor in Animal Breeding and Genetics
Swedish University of Agricultural Sciences

Co-promotor

Prof. Dr D.J. de Koning
Professor in Animal Breeding and Genetics
Swedish University of Agricultural Sciences

Dr A.C. Bouwman
Researcher, Animal Breeding and Genomics Center
Wageningen University & Research

Other members

Prof. Dr B. Kemp, Wageningen University and Research
Dr S. van Beek, Cooperative Cattle Improvement Organization CRV BV
Dr B. Heringstad, Norwegian University of Life Sciences
Prof. U. Emanuelson, Swedish University of Agricultural Sciences

This research was conducted under the joint auspices of the Swedish University of Agricultural Sciences and the Graduate School Wageningen Institute of Animal Sciences of Wageningen University and is part of the Erasmus Mundus Joint Doctorate Program EGS-ABG.

Milk progesterone measures to improve genomic selection for fertility in dairy cows

Amabel Manyu Mefru Tenghe



ACTA UNIVERSITATIS AGRICULTURAE SUECIAE
DOCTORAL THESIS No. 2017:37

Thesis

submitted in fulfilment of the requirements for the joint degree of doctor between

Swedish University of Agricultural Sciences

by the authority of the board of the Faculty of Veterinary Medicine and Animal
Sciences

and

Wageningen University

by the authority of the Rector Magnificus, Prof. Dr A.P.J. Mol,
in the presence of the

Thesis Committee appointed by the Academic Board of Wageningen University and
the Board of the Faculty of Veterinary Medicine and Animal Sciences at
the Swedish University of Agricultural Sciences

to be defended in public

on Friday 28 April 2017

at 4 p.m. in the Aula of Wageningen University.

Amabel M.M. Tenghe.

Milk progesterone measures to improve genomic selection for fertility in dairy cows

179 pages.

Joint PhD thesis, Swedish University of Agricultural Sciences, Uppsala, Sweden and Wageningen University, the Netherlands (2017)

With references, with summary in English

ISSN: 1652-6880

ISBN: (print version): 978-91-576-8847-7

ISBN: (electronic version): 978-91-576-8848-4

ISBN: 978-94-6343-133-0

DOI: 10.18174/409680

Abstract

Tenghe, A.M.M. (2017). Milk progesterone measures to improve genomic selection for fertility in dairy cows. Joint PhD thesis, between Swedish University of Agricultural Sciences, Sweden and Wageningen University, the Netherlands

Improved reproductive performance has a substantial benefit for the overall profitability of dairy cattle farming by decreasing insemination and veterinary treatment costs, shortening calving intervals, and lowering the rate of involuntary culling. Unfortunately, the low heritability of classical fertility traits derived from calving and insemination data makes genetic improvement by traditional animal breeding slow. Therefore, there is an interest in finding novel measures of fertility that have a higher heritability or using genomic information to aid genetic selection for fertility. The overall objective of this thesis was to explore the use of milk progesterone (P4) records and genomic information to improve selection for fertility in dairy cows. In a first step, the use of in-line milk progesterone records to define endocrine fertility traits was investigated, and genetic parameters estimated. Several defined endocrine fertility traits were heritable, and showed a reasonable repeatability. Also, the genetic correlation of milk production traits with endocrine fertility traits were considerably lower than the correlations of milk production with classical fertility traits. In the next step 17 quantitative trait loci (QTL) associated with endocrine fertility traits, were identified on *Bos taurus* autosomes (BTA) 2, 3, 8, 12, 15, 17, 23, and 25 in a genome-wide association study with single nucleotide polymorphisms. Further, fine-mapping of target regions on BTA 2 and 3, identified several associated variants and potential candidate genes underlying endocrine fertility traits. Subsequently, the optimal use of endocrine fertility traits in genomic evaluations was investigated; using empirical and theoretical predictions for single-trait models, I showed that endocrine fertility traits have more predictive ability than classical fertility traits. The accuracy of genomic prediction was also substantially improved when endocrine and classical fertility traits were combined in multi-trait genomic prediction. Finally, using deterministic predictions, the potential accuracy of multi-trait genomic selection when combining a cow training population measured for the endocrine trait commencement of luteal activity (C-LA), with a training population of bulls with daughter observations for a classical fertility trait was investigated. Results showed that for prediction of fertility, there is no benefit of investing in a cow training population when the breeding goal is based on classical fertility traits. However, when considering a more biological breeding goal for fertility like C-LA, accuracy is substantially improved when endocrine traits are available from a limited number of farms.

To my family

Table of Contents

Chapter 1	General introduction	11
Chapter 2	Estimating genetic parameters for fertility in dairy cows from in-line milk progesterone profiles	25
Chapter 3	Genome-wide association study for endocrine fertility traits using single nucleotide polymorphism arrays and sequence variants in dairy cattle	49
Chapter 4	Opportunities for genomic prediction for fertility using endocrine and classical fertility traits in dairy cattle	81
Chapter 5	Improving accuracy of bulls' predicted genomic breeding values for fertility using daughters' milk progesterone profiles	107
Chapter 6	General discussion	135
	Summary	161
	Curriculum vitae	165
	Acknowledgements	173
	Colophon	177

1

General introduction

1.1 Importance of fertility

Fertility is one of the non-yield traits which is of great economic importance in dairy herds. Poor fertility increases costs due to fertility treatments, multiple inseminations, prolongs calving interval, and leads to a high replacement rate due to involuntary culling (Boichard, 1990; Dekkers, 1991; González-Recio et al., 2004). The economic consequences of poor fertility have been widely studied. Inchaisri et al., (2010) studied the economic consequences of non-optimal fertility of dairy cows under Dutch conditions, where non-optimal fertility was defined as “average” or “poor” reproductive performance, using different fertility traits. They reported mean net economic losses of €34 and €231 per cow per year for “average” and “poor” reproductive performance respectively, compared to a “good” fertility. The losses were mainly caused by decreased milk production and increased number of non-pregnant cows, especially in the situation of poor fertility. Also, a net cost of a one day increase in calving interval was €0.57 per cow per day for “average” fertility, and €0.70 per cow per day for “poor” fertility (Inchaisri et al., 2010). Other studies have reported costs of one day extra for of €0.06 to €1.10 for days open (de Vries and Conlin, 2003; Groenendaal et al., 2004; Meadows et al., 2005) and €2.07 to €2.95 for calving interval (Plaizier et al., 1997). This economic importance is one of the reasons why there is an interest in genetic improvement of fertility.

1.2 Fertility measures used in genetic evaluation

Overall fertility can be described and measured by several different traits. For example, it is useful to distinguish among traits which are affected by the cow (female fertility), and traits affected by the sire mated to the cow (male fertility), and traits affected by both. Traits like age at first puberty in heifers, estrous expression, and the time between calving and regular ovarian activity or first insemination measure female fertility, whereas traits like sperm count and semen quality measure male fertility. Traits that measure the results of insemination and conception rate are influenced by both male and female fertility, and a combination of them. They include non-return rate, number of inseminations per conception or service period, percentage of cows bred that are pregnant (conception rate), the period between calving and confirmed conception (days open), and pregnancy rate. This thesis focuses on female fertility. Female fertility is made up of different underlying traits (Figure 1.1). The traits used in genetic evaluation of female fertility can be categorized into two aspects of fertility. The first reflects the ability of the cow to return to cyclicity after calving, e.g., the interval from calving to first service. The

second reflects the ability of the cow to conceive following insemination, become pregnant, and maintain pregnancy, e.g., the time between the first and last insemination, and non-return rate after service. Therefore, because female fertility can be measured by different indicators, it is important to identify the optimal measures to facilitate genetic improvement.

The classical fertility traits derived from insemination and calving data which are used in genetic evaluation of fertility have the disadvantage that they have a low heritability (Jansen, 1985; Berry et al., 2003), which makes genetic improvement difficult. Heritability estimates for classical fertility traits generally range from 0.01 to 0.10 (Hou et al., 2009; Sun et al., 2009; Berry et al., 2012). The low heritability may be explained by the fact that classical fertility traits are highly influenced by farm management decisions and poor recording practices (Hayes et al., 1992; Campos et al., 1994; Marti and Funk, 1994). For example, a planned extended CI will delay CFS, not because a cow has a late start of cyclic activity, but because of the farmer's decision on when to inseminate. Such management practices are justified, but results in large residual variance and low heritability estimates. The low heritability sparks the interest for alternative approaches that might yield more accurate information for genetic evaluations of dairy cows' fertility. For instance, fertility measures that more directly reflect the cows' own reproductive physiology like endocrine fertility traits that are derived from progesterone concentration levels in milk, or the use of DNA information.

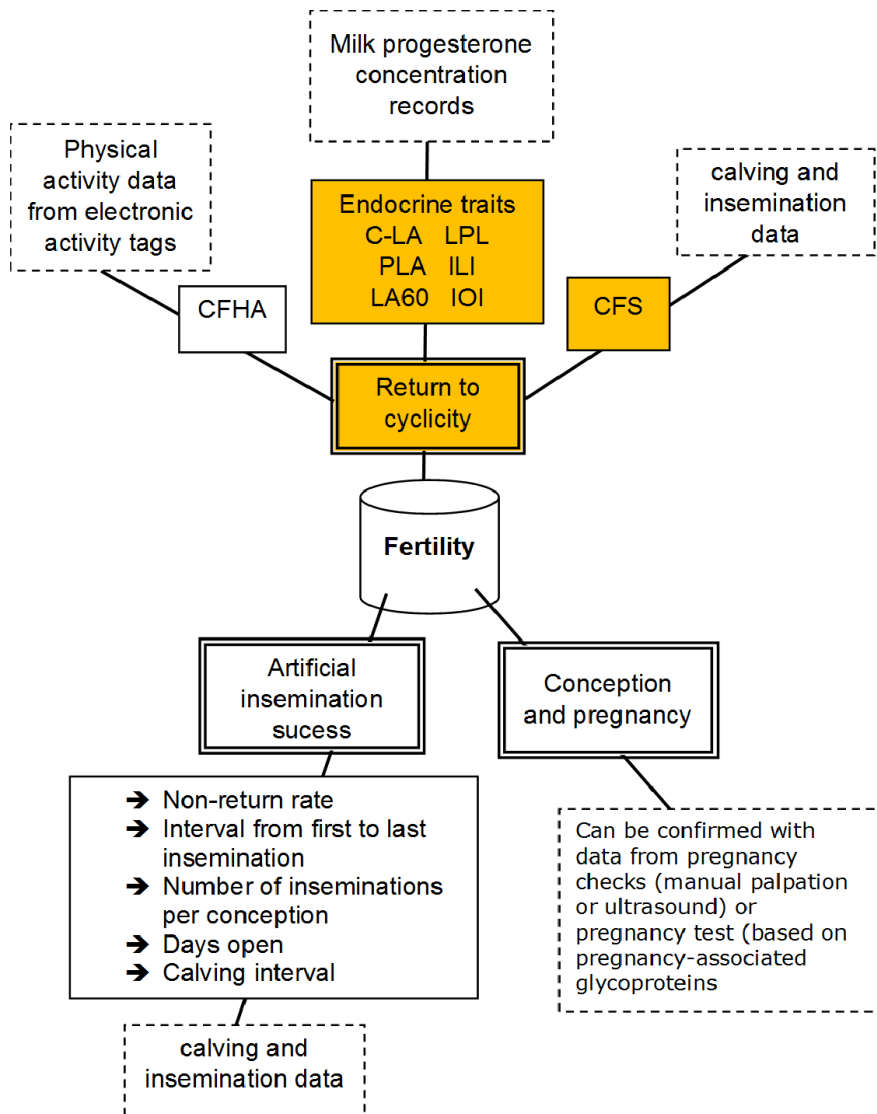


Figure 1.1 Schematic chart of different components of female fertility in dairy cattle (in double boxes), most widely used underlying phenotype measures associated with each component (in solid boxes), and source of data used to derive the phenotypes (in dashed boxes). This thesis investigated the traits in the orange boxes. C-LA = Interval from calving to commencement of luteal activity; PLA = Proportion of samples in luteal activity; LA60 = Occurrence of luteal activity during first 60 days in milk; LPL = Luteal phase length; ILI = Length of inter-luteal interval (ILI); IOI = Length of inter-ovulatory interval; CFHA = interval from calving to first high activity; CFS = interval from calving to first service

1.3 Progesterone defined fertility traits

Understanding the estrous cycle is important for managing and improving reproductive performance in dairy cattle. The estrous cycle is divided into two phases which are characterized by changes on the ovary (Figure 1.2). There is the follicular phase during which the pre-ovulatory follicle on the ovary, which contains the oocyte (or egg), produces estrogen. When estrogen concentrations are high enough, a surge of luteinizing hormone is released, initiating ovulation. The luteal phase begins after ovulation, when the follicle transforms into a corpus luteum which produces progesterone to maintain pregnancy (Senger, 2003). Early resumption of ovarian cyclicity postpartum facilitates a greater number of estrus cycles before insemination which, on average, increases the likelihood of subsequent conception (Darwash et al., 1997). Studies have shown that there is a high correlation between progesterone (P4) concentration in blood and milk (e.g., $r = 0.88$; Dobson and Fitzpatrick, 1975), hence, P4 analysis of milk samples can be used to study postpartum ovarian activity in dairy cows. A period of low P4 levels usually occurs after calving, when a cow exhibits anestrus (Lamming and Bulman, 1976). This period is followed by an increase in P4 levels, which is indicative of the first postpartum ovulation. The cavity of the ovulated follicle is gradually filled with progesterone-secreting luteal cells, which forms the corpus luteum. From about the fifth day after ovulation, the corpus luteum dominates the estrous cycle during the luteal phase with high P4 levels for about 14 days. After that, unless the cow becomes pregnant, the corpus luteum degenerates (luteolysis), and a new ovulation can occur. If the cow becomes pregnant, the corpus luteum is maintained during the pregnancy (Ball and Peters, 2004).

Progesterone and overall fertility of the dairy cow have been shown to be connected in several ways. Low probability of embryo survival was shown to be associated with both low and excessive P4 levels 5 to 7 days after insemination, which indicates that an optimum in P4 level is required for embryo survival after insemination (Stronge et al., 2005). In a study, repeat breeding heifers (i.e., heifers that fail to conceive from 3 or more regularly spaced services in the absence of detectable abnormalities) tended to have higher basal P4 concentrations at estrus, a late P4 rise in early luteal phase and low luteal concentrations of P4 compared to virgin heifers (Båge et al., 2002). In addition, increased basal P4 levels at insemination have been associated with increased probability of repeat-breeding for cows and higher return rate at insemination (Waldmann et al., 2001). The ovarian activity of a cow after calving also affects overall fertility. For example, early onset of estrous cyclicity after calving has been shown to increase probability of an early insemination after calving,

shorten the interval from calving to conception, increase conception rate and reduce the number of services per conception (Darwash et al., 1997).

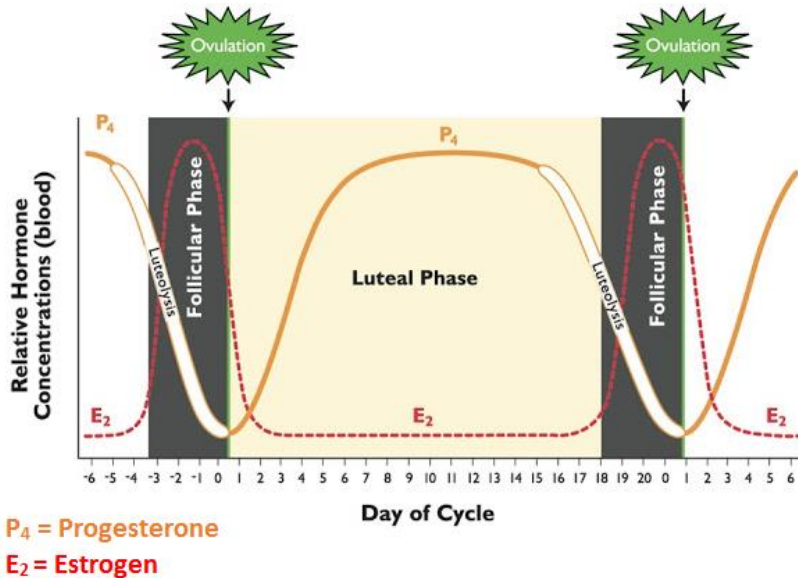


Figure 1.2 Stages within the follicular and luteal phases in association with respective changes in circulating concentrations of estrogen and progesterone during the bovine estrous cycle. Adapted from P.L. Senger Pathways to pregnancy and parturition Current Conceptions, Inc., Pullman, WA

1.4 Genomic regions associated to endocrine fertility traits

Several genome-wide association studies have attempted to locate genomic regions associated with reproductive performance in dairy cattle, see the cattle QTL (quantitative trait loci) database for an overview (Hu et al., 2016). To date, 81,653 QTL have been catalogued in cattle, for 521 traits of which 8,969 QTL are for fertility traits (<http://www.animalgenome.org>; 2016). For CFS, QTL have been reported on chromosomes 4, 7, 9, 13, 20, 23, and 25 (Druet et al., 2008; Sahana et al., 2010; Schulman et al., 2011; Höglund et al., 2014, 2015). Most QTL studies for fertility have been based on classical fertility traits, with few studies performed with endocrine fertility traits. In a genome-wide association study with 50,000 single nucleotide polymorphisms (SNP), Berry et al., (2012) reported QTL regions associated to C-LA on chromosomes 2 and 21 and these regions have been associated with reproductive performance in other studies (Huang et al., 2010; Sahana et al., 2010; Schulman et

al., 2011). Also, most of the QTL studies used low density (e.g., 50,000) SNP panels which do not allow for precision mapping of quantitative trait loci. The use of genome sequence data for association studies has been enabled by advances in next generation sequencing techniques which have led to sequencing numerous animals in cattle, e.g., projects like the 1000 Bull Genome Consortium (Daetwyler et al., 2014). It is expected that associations with endocrine fertility traits can be targeted to smaller chromosomal regions with sequence compared to smaller genetic marker panels like 50,000 SNP markers.

1.5 Genomic selection for fertility

Genomic selection is a breeding tool that uses SNP markers spread across the genome to predict genomic breeding values for individuals (Meuwissen et al., 2001; Hayes et al., 2009). Genomic selection has become the standard for dairy cattle breeding in most countries, because it increases the potential genetic gain by up to 80% due to reduced generation interval (Schaeffer, 2006). Two steps are required to perform genomic selection. The first step entails genotyping and phenotyping a set of animals to form the training population. The phenotypes and genotypes of the training population are then used to obtain prediction equations. In the second step, the prediction equations are used with genotypes of evaluation animals (e.g., selection candidates), to predict their genomic breeding values. The selection candidates usually are juveniles, and their genomic breeding values are more accurate than when estimated with traditional breeding tools that rely on parent average information (Meuwissen et al., 2001). The accuracy of genomic selection depends several factors including the number of animals in the training population, the heritability of the trait, and the genetic architecture of the trait, in particular the number of loci affecting the trait and distribution of their effects (Daetwyler et al., 2008; Meuwissen, 2009). An important feature of genomic selection which makes it interesting to apply to expensive or difficult to measure traits like endocrine fertility traits is that the traits do not have to be recorded on a routine scale and on a large number of daughters for each selection candidate. With a training population of limited size, genomic selection was shown as a promising tool for starting selection for scarcely recorded or difficult to measure traits (Calus et al., 2013).

1.6 This thesis

The main objective of this thesis was to explore the use of milk P4 records and genomic information to improve selection for fertility in dairy cows. This main objective was divided into two sub-objectives. The first sub-objective investigated the use of in-line milk P4 records to define endocrine fertility traits and identify genomic regions associated to the defined traits, using SNP markers and whole genome sequence variants. The second sub-objective aimed to determine the optimal use of endocrine fertility traits in genomic evaluations by investigating the impact of different phenotyping strategies on the accuracy of predicting genomic breeding values for fertility.

In **chapter 2** of this thesis, the use of in-line recorded milk progesterone levels to define endocrine fertility traits was investigated, genetic parameters were estimated and genetic correlations of the endocrine traits with classical fertility and milk production traits were estimated. In **chapter 3**, genomic regions associated to endocrine fertility traits were identified by performing a genome-wide association study with 85,485 SNP, and targeted QTL regions were fine-mapped using imputed sequence variants. **Chapter 4** investigated the added value in terms of accuracy, of using endocrine traits along with classical fertility traits in genomic prediction of fertility, by performing single-trait and multi-trait genomic predictions. The aims of **chapter 5** were: a) to investigate the potential accuracy of different scenarios when combining a cow training population measured for an endocrine trait with a training population of bulls with daughter observations for a classical trait for multi-trait genomic prediction of fertility, by using a deterministic prediction equation within and between populations, and b) to investigate recording strategies that optimally use the Herd Navigator for genomic prediction, in terms of, the number of farms, and recording period for endocrine fertility traits.

References

- Båge, R., H. Gustafsson, B. Larsson, M. Forsberg, and H. Rodriguez-Martinez. 2002. Repeat breeding in dairy heifers: follicular dynamics and estrous cycle characteristics in relation to sexual hormone patterns. *Theriogenology*. 57:2257–2269.
- Ball, P.J.H., and A.R. Peters. 2004. Reproduction in cattle. Wiley Online Library.
- Berry, D.P., J.W.M. Bastiaansen, R.F. Veerkamp, S. Wijga, E. Wall, B. Berglund, and M.P.L. Calus. 2012. Genome-wide associations for fertility traits in Holstein-Friesian dairy cows using data from experimental research herds in four

- European countries. *Animal*. 6:1206–15.
- Berry, D.P., F. Buckley, P. Dillon, R.D. Evans, M. Rath, and R.F. Veerkamp. 2003. Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows. *J. Dairy Sci.* 86:2193–2204.
- Boichard, D. 1990. Estimation of the economic value of conception rate in dairy cattle. *Livest. Prod. Sci.* 24:187–204.
- Calus, M.P.L., Y. De Haas, M. Pszczola, and R.F. Veerkamp. 2013. Predicted accuracy of and response to genomic selection for new traits in dairy cattle. *Animal*. 7:183–191.
- Campos, M.S., C.J. Wilcox, C.M. Becerril, and A. Diz. 1994. Genetic Parameters for Yield and Reproductive Traits of Holstein and Jersey Cattle in Florida. *J. Dairy Sci.* 77:867–873.
- Daetwyler, H.D., A. Capitan, H. Pausch, P. Stothard, R. Van Binsbergen, R.F. Brøndum, X. Liao, A. Djari, S.C. Rodriguez, C. Grohs, D. Esquerré, O. Bouchez, M.-N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P.J. Bowman, D. Coote, A.J. Chamberlain, C. Anderson, C.P. VanTassell, I. Hulsege, M.E. Goddard, B. Guldbrandsen, M.S. Lund, R.F. Veerkamp, D.A. Boichard, R. Fries, and B.J. Hayes. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* 46:858–865.
- Daetwyler, H.D., B. Villanueva, and J.A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*. 3:e3395.
- Darwash, A.O., G.E. Lamming, and J.A. Woolliams. 1997. The phenotypic association between the interval to post-partum ovulation and traditional measures of fertility in dairy cattle. *Anim. Sci.* 65:9–16.
- Dekkers, J.C.M. 1991. Estimation of economic values for dairy cattle breeding goals: bias due to sub-optimal management policies. *Livest. Prod. Sci.* 29:131–149.
- Dobson, H., and R.J. Fitzpatrick. 1975. Clinical application of the progesterone-in-milk test. *Br. Vet. J.* 132:538–542.
- Druet, T., S. Fritz, M. Boussaha, S. Ben-Jemaa, F. Guillaume, D. Derbala, D. Zelenika, D. Lechner, C. Charon, D. Boichard, I.G. Gut, A. Eggen, and M. Gautier. 2008. Fine Mapping of Quantitative Trait Loci Affecting Female Fertility in Dairy Cattle on BTA03 Using a Dense Single-Nucleotide Polymorphism Map. *Genetics*. 178:2227–2235.
- González-Recio, O., M.A. Pérez-Cabal, and R. Alenda. 2004. Economic value of female fertility and its relationship with profit in Spanish dairy cattle. *J. Dairy Sci.* 87:3053–3061.
- Groenendaal, H., D.T. Galligan, and H.A. Mulder. 2004. An Economic Spreadsheet Model to Determine Optimal Breeding and Replacement Decisions for Dairy Cattle. *J. Dairy Sci.* 87:2146–2157.
- Hayes, B.J., H.D. Daetwyler, P.J. Bowman, G. Moser, B. Tier, R. Crump, M. Khatkar, H.W. Raadsma, and M.E. Goddard. 2009. Accuracy of genomic selection: comparing theory and results. *In Proc Assoc Advmt Anim Breed Genet.* 34–37.

- Hayes, J.F., R.I. Cue, and H.G. Monardes. 1992. Estimates of Repeatability of Reproductive Measures in Canadian Holsteins. *J. Dairy Sci.* 75:1701–1706.
- Höglund, J.K., B. Guldbrandtsen, M.S. Lund, and G. Sahana. 2015. Identification of genomic regions associated with female fertility in Danish Jersey using whole genome sequence data. *BMC Genet.* 16:60.
- Höglund, J.K., G. Sahana, R. Brøndum, B. Guldbrandtsen, B. Buitenhuis, and M.S. Lund. 2014. Fine mapping QTL for female fertility on BTA04 and BTA13 in dairy cattle using HD SNP and sequence data. *BMC Genomics.* 15:790.
- Hou, Y., P. Madsen, R. Labouriau, Y. Zhang, M.S. Lund, and G. Su. 2009. Genetic analysis of days from calving to first insemination and days open in Danish Holsteins using different models and censoring scenarios. *J. Dairy Sci.* 92:1229–1239.
- Hu, Z.-L., C.A. Park, and J.M. Reecy. 2016. Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Res.* 44:D827–D833.
- Huang, W., B.W. Kirkpatrick, G.J.M. Rosa, and H. Khatib. 2010. A genome-wide association study using selective DNA pooling identifies candidate markers for fertility in Holstein cattle. *Anim. Genet.* 41:570–578.
- Inchaisri, C., R. Jorritsma, P. Vos, G.C. Van der Weijden, and H. Hogeveen. 2010. Economic consequences of reproductive performance in dairy cattle. *Theriogenology.* 74:835–846.
- Jansen, J. 1985. Genetic aspects of fertility in dairy cattle based on analysis of A.I. data — A review with emphasis on areas for further research. *Livest. Prod. Sci.* 12:1–12.
- Lamming, G.E., and D.C. Bulman. 1976. The use of milk progesterone radioimmunoassay in the diagnosis and treatment of subfertility in dairy cows. *Br. Vet. J.*
- Marti, C.F., and D.A. Funk. 1994. Relationship Between Production and Days Open at Different Levels of Herd Production. *J. Dairy Sci.* 77:1682–1690.
- Meadows, C., P.J. Rajala-Schultz, and G.S. Frazer. 2005. A Spreadsheet-Based Model Demonstrating the Nonuniform Economic Effects of Varying Reproductive Performance in Ohio Dairy Herds. *J. Dairy Sci.* 88:1244–1254.
- Meuwissen, T.H.E. 2009. Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* 41:1.
- Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 157:1819–1829.
- Plaizier, J.C.B., G.J. King, J.C.M. Dekkers, and K. Lissemore. 1997. Estimation of Economic Values of Indices for Reproductive Performance in Dairy Herds Using Computer Simulation. *J. Dairy Sci.* 80:2775–2783.
- Sahana, G., B. Guldbrandtsen, C. Bendixen, and M.S. Lund. 2010. Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. *Anim. Genet.* 41:579–588.
- Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123:218–223.

- Schulman, N.F., G. Sahana, T. Iso-Touru, S.D. McKay, R.D. Schnabel, M.S. Lund, J.F. Taylor, J. Virta, and J.H. Vilkki. 2011. Mapping of fertility traits in Finnish Ayrshire by genome-wide association analysis. *Anim. Genet.* 42:263–269.
- Senger, P.L. 2003. Placentation, the endocrinology of gestation and parturition. *Pathways to pregnancy parturition.* 2:304–325.
- Stronge, A.J.H., J.M. Sreenan, M.G. Diskin, J.F. Mee, D.A. Kenny, and D.G. Morris. 2005. Post-insemination milk progesterone concentration and embryo survival in dairy cows. *Theriogenology.* 64:1212–1224.
- Sun, C., P. Madsen, U.S. Nielsen, Y. Zhang, M.S. Lund, and G. Su. 2009. Comparison between a sire model and an animal model for genetic evaluation of fertility traits in Danish Holstein population. *J. Dairy Sci.* 92:4063–4071.
- de Vries, A., and B.J. Conlin. 2003. Economic Value of Timely Determination of Unexpected Decreases in Detection of Estrus Using Control Charts1. *J. Dairy Sci.* 86:3516–3526.
- Waldmann, A., O. Reksen, K. Landsverk, E. Kommisrud, E. Dahl, A.O. Refsdal, and E. Ropstad. 2001. Progesterone concentrations in milk fat at first insemination—effects on non-return and repeat-breeding. *Anim. Reprod. Sci.* 65:33–41.

2

Estimating genetic parameters for fertility in dairy cows from in-line milk progesterone profiles

A. M. M. Tenghe¹²³, A. C. Bouwman¹, B. Berglund³, E. Strandberg³, J. Y. Blom⁴, R. F. Veerkamp¹²

¹Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands

²Animal Breeding and Genomics Centre, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands

³Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, P. O. Box 7023, SE-750 07 Uppsala, Sweden

⁴Lattec I/S, Slangerupgade 69, DK 3400 Hillerød, Denmark

Abstract

The aim of this study was to define endocrine fertility traits from in-line milk progesterone (P4) records and to estimate genetic parameters for these traits. Correlations of classical fertility (calving interval and calving to first service) and milk production traits with endocrine fertility traits were also estimated. In-line milk P4 records ($n = 160,952$) collected from June 2009 through November 2013 for 2,273 lactations of 1,561 Holstein-Friesian cows in 12 commercial herds in the Netherlands were analyzed for (the log of) the number of days from calving till commencement of luteal activity (InC-LA), proportion of samples between 25 and 60 days in milk with luteal activity (PLA), presence or absence of luteal activity for a cow between 25 and 60 days in milk, interval from commencement of luteal activity to first service (CLAFS), first luteal phase length, length of first inter-luteal interval, and length of first inter-ovulatory interval. Milk P4 records were sampled on average every two days. Genetic parameters were estimated using a mixed linear animal model. Heritability estimates (\pm SE) of endocrine fertility traits were 0.12 ± 0.05 for InC-LA, 0.12 ± 0.05 for PLA, and 0.11 ± 0.06 for CLAFS, and their repeatability estimates were 0.29 ± 0.04 , 0.21 ± 0.04 and 0.15 ± 0.06 respectively. The genetic correlation of InC-LA with PLA was -0.91 ± 0.06 , and -0.56 ± 0.25 with CLAFS. The genetic correlations of InC-LA were 0.26 ± 0.33 with calving interval and 0.37 ± 0.21 with calving to first service. Genetic correlations of the milk production traits with InC-LA ranged from 0.04 to 0.18 and 0.07 to 0.65 with classical fertility traits. The phenotypic correlations of all endocrine fertility traits with milk production traits were close to zero (0.01 to 0.07). This study shows that in-line P4 records can be used to define and explore several heritable endocrine fertility traits in dairy cows, and might help in selection for improved fertility.

Key words: dairy cow, fertility, in-line progesterone, heritability

2.1 Introduction

Fertility is an important component of herd production efficiency because each additional estrus cycle that does not result in a planned pregnancy adds to the cost of dairy farming. The negative impact of fertility on production efficiency is often reflected in increased number of inseminations per conception, higher involuntary herd replacement, high cost of veterinary intervention and longer calving intervals. In addition to these negative effects, subfertility can affect the rate of genetic gain in other traits of economic importance.

Genetically improving fertility by selection with classical traits like interval from calving to first service (**CFS**), calving interval (**Clnt**), or days open is complicated by the low heritability (h^2) of these traits (Jansen, 1985; Berry et al., 2003). These low heritabilities ($h^2 < 0.1$) may be explained by the fact that classical fertility traits are highly influenced by farm management decisions and poor recording practices (Hayes et al., 1992; Campos et al., 1994; Marti and Funk, 1994). For instance, a planned extended Clnt will delay CFS not because a cow has a late start of cyclic activity, but because of the farmer's decision on when to inseminate, which results in large residual variance and low heritability estimates. Further indications that commercial farm management and recording might affect heritability estimates for classical fertility traits is the fact that larger heritability estimates ($h^2 = 0.13$, for days open and CFS) have been estimated from records collected on an experimental farm, under controlled management (Pryce et al., 1997). This low heritability of classical fertility traits makes it difficult to discriminate fertile genotypes at the cow level and consequently makes selection less effective.

Endocrine fertility traits have been suggested as alternative indicators for fertility in dairy cows because they more directly reflect a cow's physiology (Bulman and Lamming, 1978b; Lamming and Darwash, 1998; Darwash et al., 1999). For example, the interval from calving to first ovulation as determined by progesterone (**P4**) levels in milk could be used as a direct indicator of a cow's ability to return to luteal activity after calving, instead of an indirect indicator like CFS. Several studies have revealed that endocrine fertility traits yield higher heritability estimates than classical fertility traits. Notably, for interval from calving to commencement of luteal activity (**C-LA**), heritability estimates of 0.16–0.28 have been found, which is larger than for classical fertility traits (Darwash et al., 1997a; Veerkamp et al., 1998; Petersson et al., 2007). Furthermore, examination of phenotypic correlation of C-LA with classical fertility traits revealed that early re-establishment of cyclic activity in post-partum cows

increases the probability of an early insemination after calving, shortens the interval from calving to conception, increases conception rate, and reduces the number of services per conception (Darwash et al., 1997b). These results further suggest that early re-establishment of cyclic activity is an important prerequisite for high fertility. In addition to C-LA, Darwash et al. (1997a) equally confirmed that milk P4 profiles could provide several objective heritable endocrine fertility traits. For example, corpus luteum competence and inter-luteal interval (ILI) were highly correlated with conception rate. Similarly, (Petersson et al., 2006) showed that endocrine fertility traits have moderate repeatability (0.14 to 0.16), suggesting more influence by the cow itself compared with classical fertility traits.

Although milk P4 levels have been widely accepted as valid indicators of fertility in dairy cows (Bulman and Lamming, 1978b; Lamming and Darwash, 1998; Royal et al., 2002a), their application in routine genetic evaluation schemes has been constrained by the high cost associated with collecting these measures in sufficient number of samples per cow. Until recently, methods to measure milk P4 level were labor intensive. They often entailed manually taking several milk samples per cow, analyzing and recording the results. However, P4 level measuring technology has advanced to allow in-line systems to instantly measure milk P4 level (Friggens et al., 2008). In these systems, milk is automatically sampled; P4 level measured, and results recorded. Hence, more animals can be sampled at a lower cost, making it possible to have sufficient endocrine fertility traits for use in routine genetic evaluations. To use endocrine fertility traits from in-line milk P4 records in genetic improvement of fertility, the first step will be to examine whether these traits have sufficient genetic variation. Moreover, it will be important to know the correlation of in-line endocrine fertility traits with other traits included in the selection index before use in genetic improvement. To the best of our knowledge, no study has attempted to characterize heritable endocrine fertility traits in dairy cows from in-line milk P4 records.

The aim of this study was to define endocrine fertility traits from in-line milk progesterone records, estimate genetic parameters for these traits and their genetic correlations with classical fertility and milk production traits.

2.2 Materials and Methods

2.2.1 In-line milk progesterone records and filtering criteria

Milk P4 records ($n = 226,188$) collected from June 2009 through November 2013 were available for 2,514 Holstein-Friesian cows from 12 commercial dairy herds in the Netherlands. Milk sampling, measuring and recording of P4 level was performed by the Herd NavigatorTM. (HN, DeLaval Intl, Tumba, Sweden). The HN is a management tool for dairy herds, which samples and analyzes a number of milk constituents automatically during milking. One function of the HN is to monitor reproductive performance of cows by sampling and analyzing milk P4 level. This function is based on a bio-model that controls automatic in-line sampling, measuring and recording of milk P4 level at varying intervals during a cows' estrus cycle (Friggens et al., 2008). Progesterone record files from each herd consisted of four variables: 1) herd number, 2) cow identification, 3) sampling date and time, and 4) sample value (P4 level in ng/ml). These records were accompanied by corresponding data files with information on calving dates, parity and data files with insemination dates. Linking of P4 records to calving data resulted in 3,648 lactations of 2,340 cows with 213,877 P4 records. In total, 5% P4 records and 7% of cows were edited out after linking, because their corresponding calving information was not available. Most cows were sampled from 27 days in milk (DIM) to 176 DIM, with a mean sampling duration of 149 days. Milk P4 samples were taken on average every 2 days, with 25 percent of samples taken at an interval of less than 1 day, and 75 percent at an interval of less than 4 days.

Two sets of filtering criteria (FC) were applied to P4 records: herd level FC and lactation level FC. At the herd level, two criteria were applied. The management tool is recent and most herds installed the tool at a certain date when not all sampled cows were at the start of their lactation, therefore a lactation was retained if the interval from calving to start of herd P4 recording was ≤ 25 days. Similarly, because P4 recording ended at different times in each herd, a lactation was retained if the interval from calving to end of herd P4 recording was ≥ 60 days. At the lactation level, to reduce sampling variation due to difference in start of P4 recording per lactation, two criteria were applied: 1) P4 samples taken before 25 DIM in a lactation were excluded, and 2) lactations for which P4 recording started after 35 DIM were excluded. After data filtering, 160,952 (75%) P4 records of 2,353 (65%) lactations for 1,630 (70%) cows were retained. Data editing, filtering, and trait definition were carried out in R (R Core Team, 2013). The largest data loss was at herd level filtering.

The number of P4 records, lactations and cows retained after each filtering criterion are shown in Table 2.1.

Table 2.1 Number of progesterone (P4) records, lactations (Lact), and cows retained by filtering criteria (FC).

Criteria ¹	P4	% P4	Lact	% Lact	Cows	% Cows
Start Data	213,877	100	3,648	100	2,340	100
Herd FC	171,804	80	2,489	68	1,708	73
Lactation FC	160,952	75	2,353	65	1,630	70

¹Herd FC = exclude lactations with interval from calving to start of P4 recording > 25 days and lactations with interval from calving to end of P4 recording < 60 days; Lactation FC = exclude P4 records taken before 25 days in milk and lactations with start of P4 recording > 35 days in milk

2.2.2 Defining endocrine fertility traits

For each lactation, the following endocrine fertility traits were defined (Figure 2.1).

Commencement of luteal activity (C-LA). C-LA refers to the interval from calving to start of luteal activity. A threshold of 5 ng/ml was used as indication for luteal activity. Commencement of luteal activity (indicator I, Figure 2.1) was defined as the number of days between day of calving and first day on which milk P4 level was elevated (≥ 5 ng/ml).

Commencement of luteal activity to first service (CLAFS). CLAFS refers to the interval from the initiation of luteal activity to first service (indicator II, Figure 2.1), and was measured from the first day of elevated P4 level (≥ 5 ng/ml), fitting the luteal activity criteria, to day of first service. This trait is not strictly endocrine, but a hybrid trait as it is defined from both P4 and insemination records.

First luteal phase length (LPL). After ovulation, the period during which the corpus luteum secretes progesterone, measuring ≥ 5 ng/ml was referred to as the luteal phase (indicator III, Figure 2.1). The LPL was defined as the interval from the first day of elevated P4 level (≥ 5 ng/ml), to the last consecutive day of elevated P4 level (≥ 5 ng/ml).

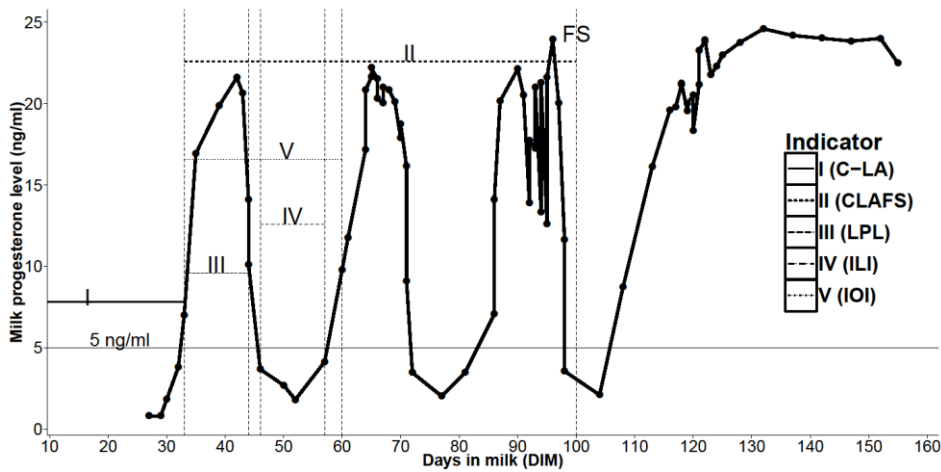


Figure 2.1 Milk progesterone level over days in milk of a hypothetical dairy cow, used to define endocrine fertility traits. C-LA = Interval from calving to commencement of luteal activity; CLAFS = Interval from commencement of luteal activity to first service; LPL = First luteal phase length; ILI = Length of first inter-luteal interval (ILI); IOI = Length of first inter-ovulatory interval; FS = First service; Line at 5 ng/ml represents threshold for luteal activity, records above 5 ng/ml indicate occurrence of luteal activity while records below 5 ng/ml indicate no luteal activity

Length of first inter-luteal interval (ILI). ILI refers to the period of anestrus between the demise of one corpus luteum and the rise of the next (indicator IV, Figure 2.1). Length of ILI was defined as the interval from the first day of decreased P4 level (< 5 ng/ml) following luteal phase, and the last consecutive day of decreased P4 level (< 5 ng/ml).

Length of first inter-ovulatory interval (IOI). IOI refers to the interval between P4 level rise from the corpus luteum of one estrus cycle and the P4 level rise from corpus luteum of the next estrus cycle (indicator V, Figure 2.1). Length of IOI was defined from the first day of elevated P4 level (≥ 5 ng/ml) and the first day of elevated P4 level (≥ 5 ng/ml) following ILI.

Luteal activity during first 60 days in milk (LA60). LA60 refers to the presence (LA60 = 1) or absence (LA60 = 0) of luteal activity between 25 and 60 days in milk.

Proportion of samples with luteal activity (PLA). PLA refers to the number of P4 records with luteal activity (P4 level ≥ 5 ng/ml), divided by total

number of P4 records in the period from 25 to 60 days in milk. This trait gives an indication of the reproductive activity of a cow within 25 – 60 DIM, which could be quantified by alternative measures, for example by taking the absolute number of P4 samples above a given maximum threshold of luteal activity. Although PLA could be affected by dynamic sampling, still taking the ratio of number of possible samples versus absolute numbers in luteal activity gives less bias in terms of bias that could be expected due to dynamic sampling.

In defining the endocrine fertility traits, a set of restrictions were applied to periods of non-sampling of P4 records that might occur during a lactation period. A period of no P4 sampling might be herd-specific and occurs when the HN has a breakdown, resulting in gaps of no sampling. If these gaps are not taken into account, they can bias the defined endocrine traits. Recording gaps were identified at herd and at cow-lactation level. When the cow's lactation had no P4 samples for more than 7 days, the following restrictions were applied: 1) if the gap occurred 1 or more days before C-LA (i.e. there was a low progesterone sample before CLA), and the gap duration was less than 15 days, all traits were retained, otherwise all traits were excluded; 2) if the gap occurred at least 25 days after C-LA, all traits were retained, otherwise all traits were excluded except C-LA. On the other hand, when the gap was herd-specific all lactations on the given herd could be affected, hence when a gap was herd-specific; the following restrictions were applied to all lactations on that herd. For each lactation, when a gap of 3 or more days occurred 1 or more days before C-LA all traits were retained. When a gap of 3 or more days occurred after C-LA, the following restrictions were applied: a) LPL was retained if the gap occurred at least 25 days after C-LA; b) ILI was retained if gap occurred 1 or more days before start of the inter-luteal interval; c) IOI was retained if gap occurred 1 or more days before start of the next luteal cycle. In total, 45 (2%) lactations were affected by the restrictions applied to cow-lactation-specific gaps, and 144 (6%) lactations for herd-specific gaps.

2.2.3 Defining classical fertility traits

For all cows with defined endocrine fertility traits, extra lactation records for these cows were obtained from the national database, to increase number of records. These records included the classical fertility traits calving interval (Clnt), and interval from calving to first service (CFS). The trait Clnt was restricted between 300 and 700 days, while CFS was restricted to 30 and 250 days. The following milk production traits were considered as well: accumulated milk yield (**MY**), protein yield (**PY**) and

fat yield (**FY**) over 305 days of lactation; of which, 55% of the lactations had their milk production traits from 305 DIM lactations whereas 39% were predicted milk production traits over 305 days, and 6% of lactations had no production traits. One herd did not participate in the national milk recording scheme, hence only records for fertility traits were used for that herd.

After trait definition, the data were linked to pedigree information. The pedigree ($n = 8,712$) was traced back in the national database for each animal ($n = 1,561$) with records. Pedigree data consisted of 514 paternal and 1,643 maternal half-sib groups. Maximum group size was 79 for paternal half-sibs and 4 for maternal half-sibs. Percentage Holstein genes of cows were obtained from pedigree data. The final data used in genetic analysis consisted of 5,792 lactations of 2,119 cows. There were 2,273 lactations (80 lactations excluded due to no pedigree) of 1,561 cows with endocrine fertility traits, and 1,954 extra lactations with classical fertility and production traits from the national database for these cows. In addition, 1,565 lactations of 558 cows in the 12 herds with classical fertility and production traits but no endocrine traits were also included from the national database. Of the final data, 1,963, 1,525, 1,011 and 1,293 lactations were in parity 1, 2, 3, and ≥ 4 respectively. The proportion of cows with repeated records per trait were: 33% for C-LA, 23% for CLAFL, 32% for PLA and LA60, 20% for LPL, 15% for ILI, 14% for IOI, 66% for CI, 62% for CFS and 74% for MY, PY, and FY.

2.2.4 Genetic analysis

Data were analyzed with mixed linear models that use the restricted maximum likelihood method in ASREML (Gilmour et al., 2009). To estimate the heritability of each trait, variance components were obtained with a univariate animal model. The traits C-LA, LPL, ILI and IOI were highly skewed, and hence were transformed (natural logarithm) to **lnC-LA**, **lnLPL**, **lnILI**, and **lnIOI** respectively. Genetic and phenotypic correlations were obtained with bivariate models. The model used was:

$$y_{ijklmn} = \mu + p_i + hys_j + b_1pch_k + b_2ca_l + b_3ca^2(p_i) + a_m + pe_m + e_{ijklmn}$$

where y_{ijklmn} was the analyzed trait, μ was the intercept, p_i was the fixed effect of parity i (cows in parity 4 or above were grouped to a common class); hys_j was the fixed effect of herd-year-season combination j , with calving season defined as winter: December-February, spring: March-May, summer: June-August, autumn: September-November; b_1pch_k was the fixed regression on percentage Holstein genes with coefficient b_1 ; b_2ca_l was the fixed regression on age at calving l (in months) with

coefficient b_2 ; $b_3ca^2(p_i)$ was the fixed regression on age at calving i , fitted as a quadratic covariate, nested within parity p_i , with coefficient b_3 ; a_m was the random genetic effect of cow m : $\sim N(0, A\sigma_a^2)$, A was the additive genetic relationship matrix and σ_a^2 was the genetic effects variance; pe_m was the random permanent environmental effect of cow m to account for repeated measures within cow: $\sim N(0, I\sigma_{pe}^2)$, I was an identity matrix including all animals (but animals without information on repeated records are automatically set at zero) and σ_{pe}^2 was the permanent environment variance; and e_{ijklmn} was the random error term, $\sim N(0, I\sigma_e^2)$, I was an identity matrix and σ_e^2 was the residual variance.

The same model was fitted for univariate and bivariate analysis. However, the number of repeated records was too low for the endocrine traits to estimate the correlation between traits for the permanent environmental effect, and to separate this covariance from the genetic covariance between traits. Therefore, also a reduced model was fitted without this correlation included, and a dataset was created without the repeated records per animal. The likelihood ratio test was used to test for the significant difference between the full bivariate model and the model without the permanent environmental covariance (the reduced model). From the likelihood ratio test, only 8 pairs of bivariate analysis (out of 33) gave a significant improvement for the full model. Still, some of the genetic correlations were affected, and it is difficult to ascribe the covariance to a genetic effect, if the data are not sufficient to disentangle genetic from permanent environmental effects. Therefore, the results of the full model will be presented in the results section, and compared with the alternative model's results in the discussion. Also, finally, the variance-covariance matrix was constrained to be positive definite. There was no large change in correlation estimates when the matrix was constrained. The absolute average change for correlations in the model without a constrained matrix and the model with a constrained matrix was 0.05 for correlations ≥ 0.75 and 0.06 for other correlations.

2.3 Results

Summary statistics for fertility and milk production traits are in Table 2.2. Overall mean C-LA was 38.1 d and 50.4 d for CLAFS. Minimum CLAFS was -46, which occurs when a cow is inseminated before start of recorded luteal activity; 9 (0.72%) records for CLAFS were negative. Days between calving and first service (CFS) was 92 d on average and mean Clnt 410 d, whereas average 305-day milk yield was 8,592 kg milk.

2 In-line progesterone and cow fertility

Mean LA60 was 0.88, implying that 88% of cows were in luteal activity during the first 60 DIM, whereas mean PLA was 58%. Large differences were observed at herd level, C-LA ranged from 32.4 d to 45.3 d, and milk yield from 7,070 to 10,140 kg milk, while CLAFS ranged from 37.4 d to 66.3 d across herds. Herd average C-LAs were also moderately correlated with longer calving intervals (0.20), but strongly correlated with longer CFS (0.55).

Table 2.2 Overall mean, standard deviation (S.D.), minimum (Min), maximum (Max), number of lactations (n), and skewness (Skew) for endocrine fertility traits, classical fertility traits, and milk production traits

Trait ¹	Mean	S.D.	Min	Max	n	Skew
C-LA (d)	38.1	17.8	25	187	2,165	2.42
CLAFS (d)	50.4	29.5	-46	181	1,245	0.91
LA60 (0-1)	0.88	0.33	0	1	2,192	-2.31
PLA (%)	58	30	0	1	2,192	-0.70
LPL (d)	13.0	11.5	2	106	1,055	3.77
ILI (d)	9.3	8.5	1	103	924	4.44
IOI (d)	26.2	13.3	5	115	932	2.54
CInt (d)	410	68	303	699	3,632	1.40
CFS (d)	92	36	30	250	4,413	1.36
MY (kg)	8,592	1,888	1,228	15,319	5,466	0.18
PY (kg)	301	63	43	532	5,466	0.13
FY (kg)	374	76	47	714	5,466	0.17

¹C-LA = Commencement of luteal activity, CLAFS = interval from commencement of luteal activity to first service, LA60 = luteal activity between 25 and 60 days in milk, PLA = proportion of samples in luteal activity between 25 and 60 days in milk, LPL = first luteal phase length, ILI = length of first inter-luteal interval, IOI = length of first inter-ovulatory interval, CInt = calving interval, CFS = interval from calving to first service, MY = milk yield over 305 days, PY = protein yield over 305 days, FY = fat yield over 305 days

Heritability estimates for the fertility traits reported in Table 2.3 ranged from 0.05 to 0.12. The endocrine fertility traits InC-LA and PLA had moderate heritability (0.12 ± 0.05), with genetic standard deviations on the original scale 5 d and 0.10 respectively, while heritability of CLAFS was 0.11 ± 0.06 with genetic standard deviation 9 d. For the classical traits, heritability was 0.05 ± 0.03 for CInt and 0.11 ± 0.03 for CFS. Repeatability estimates of fertility traits ranged from 0.09 for CInt to 0.29 for InC-LA, implying that a cow with an early C-LA has a low to moderate probability to have an early C-LA after the next calving. The traits InILI, InLPL and InIOI

had a close to or zero genetic variation (results not shown), and were therefore excluded from correlation analysis.

Table 2.3 Estimates of heritability (h^2), repeatability (t) and their standard errors (s.e.) for endocrine fertility traits, classical fertility traits and milk production traits, and number of lactations (n) per trait

Trait ¹	n	h^2	s.e.	t	s.e.
lnC-LA (ln(d))	2,165	0.12	0.05	0.29	0.04
CLAFS (d)	1,245	0.11	0.06	0.15	0.06
LA60 (0-1)	2,192	0.06	0.04	0.18	0.04
PLA (%)	2,192	0.12	0.05	0.21	0.04
CInt (d)	3,632	0.05	0.03	0.09	0.02
CFS (d)	4,413	0.11	0.03	0.12	0.02
MY (kg)	5,466	0.23	0.04	0.42	0.02
PY (kg)	5,466	0.22	0.04	0.42	0.02
FY (kg)	5,466	0.20	0.03	0.38	0.02

¹lnC-LA = natural logarithm of commencement of luteal activity, CLAFS = interval from commencement of luteal activity to first service, LA60 = luteal activity between 25 and 60 days in milk, PLA = proportion of samples in luteal activity between 25 and 60 days in milk, CInt = calving interval, CFS = interval from calving to first service, MY = milk yield over 305 days, PY = protein yield over 305 days, FY = fat yield over 305 days

²Heritability and repeatability estimates of lnLPL, lnILI, and lnIOI are not shown, as they were close to or zero; lnLPL = natural logarithm of first luteal phase length, lnILI = natural logarithm of length of first inter-luteal interval, lnIOI = natural logarithm of length of first inter-ovulatory interval

Correlations between fertility traits are in Table 2.4. Due to low to moderate heritability and low number of records for some of the endocrine traits, standard errors for those genetic correlations were large. Therefore, the phenotypic correlations are discussed as proxies for the genetic correlations for those cases. There was a strong genetic correlation of lnC-LA with PLA (-0.91 ± 0.06), LA60 (-0.86 ± 0.03), and a longer lnC-LA (late start of luteal activity) was associated with shorter CLAFS (-0.56 ± 0.17).

All correlations of endocrine with classical fertility traits were in the direction that is expected from the trait definition. There was a moderate genetic correlation of CFS with lnC-LA (0.37 ± 0.21), and PLA (-0.31 ± 0.20), suggesting that cows with an early start of luteal activity and high reproductive activity have shorter intervals from calving to first service.

2 In-line progesterone and cow fertility

Correlations of milk production traits with all fertility traits are in Table 2.5. The genetic correlations of milk production traits with lnC-LA ranged from 0.04 to 0.18 and from 0.03 to 0.34 with other endocrine traits. With the exception of CLAFS, the phenotypic correlations of all endocrine traits with milk production traits were low (0.01 to 0.07), and based on the standard errors (0.02 to 0.03), not significantly different from zero. For phenotypic correlations of CLAFS with milk production traits, estimates of up to 0.14 ± 0.03 were obtained.

Table 2.4 Estimates of genetic (below diagonal) and phenotypic (above diagonal) correlations, of endocrine and classical fertility traits, and standard errors below each estimate in parentheses

Trait ¹	lnC-LA	CLAFS	LA60	PLA	CInt	CFS
lnC-LA	—	-0.44 (-0.02)	-0.77 (-0.01)	-0.80 (0.01)	0.03 (0.04)	0.23 (0.03)
CLAFS	-0.56 (0.25)	—	0.34 (0.03)	0.41 (0.03)	0.28 (0.05)	0.86 (0.01)
LA60	-0.86 (0.03)	0.47 (0.27)	—	0.71 (0.01)	-0.03 (0.04)	-0.25 (0.03)
PLA	-0.91 (0.06)	0.65 (0.21)	0.82 (0.16)	—	-0.02 (0.04)	-0.14 (0.03)
CInt	0.26 (0.33)	0.29 (0.43)	-0.54 (0.33)	-0.28 (0.31)	—	0.46 (0.02)
CFS	0.37 (0.21)	0.46 (0.12)	-0.53 (0.21)	-0.31 (0.20)	0.75 (0.16)	—

¹lnC-LA = natural logarithm of commencement of luteal activity, CLAFS = interval from commencement of luteal activity to first service, LA60 = luteal activity between 25 and 60 days in

2.4 Discussion

The uniqueness of this study is that in-line milk P4 records were used to characterize endocrine fertility traits. As a consequence, P4 level sampling frequency was higher, and more P4 records were available per lactation than in previous studies. Moreover, previous studies on characterization of heritable endocrine fertility traits have been based on manually collected milk samples for P4 recording, which is not practical for

breeding programs, mainly because routine milk recording samples are taken only every 4 to 6 weeks. In this study, genetic variation of similar size as in previous studies was found for some of the endocrine fertility traits. Heritability estimates were greatest for InC-LA and PLA (0.12 ± 0.05), CLAFS (0.11 ± 0.06), and InC-LA was strongly correlated with PLA (-0.91 ± 0.06). Correlations of classical fertility traits (Clnt and CFS) and milk production traits (MY, PY, and FY) with endocrine fertility traits were also examined. All correlations of endocrine traits with classical fertility traits were as expected, and there was a near-zero phenotypic correlation of endocrine fertility traits with yield traits.

2.4.1 Characterization of endocrine fertility traits.

The overall mean for C-LA (38.1 d) can be used as an indication of the interval from calving to first ovulation, which occurs on average 5 days before C-LA (Bulman and Lamming, 1978a), hence on average ovulation would have occurred in this study on approximately day 33 post-partum. The mean C-LA in this study was 8.6 days longer than the estimate (29.5 d) reported by Veerkamp et al (2000) for Dutch cows. Petersson et al. (2006) reported a mean C-LA of 33.8 d for data collected in the period of 1987 to 2002 on Swedish Holstein-Friesian and Swedish Red and White cows, whereas using data collected from 2003 through 2005 on British cows, Pollott and Coffey (2008) observed mean C-LA of 32.4 d. Furthermore, Royal et al. (2000) found shorter C-LA, 27.9 d, for data collected between 1995 to 1998, in British cows. In addition, C-LA as low as 24.7 d for data collected between 1976 to 1977 has been reported (Bulman and Lamming, 1978b). One likely reason for the mean difference of C-LA between these studies is differences in data recording and editing. For example, in our study most herds did not start recording P4 before 25 DIM, hence this was used as the lower bound for the data, whereas in most of the previous studies, P4 recording started earlier after calving. However, the mean difference might also be the result of population differences or declining fertility in dairy cattle (Lucy, 2001; Pollott and Coffey, 2008).

Proportion of samples in luteal activity (PLA) might be an informative indicator for dairy cow fertility. In addition to C-LA, PLA not only measures start of luteal activity, but also gives an indication of a cow's cyclicity after first ovulation. Because PLA was quantified based on P4 records between 25 and 60 DIM, it gives an indication of the reproductive activity of a cow post-partum for this period of time, but the period can be easily extended, and could be useful for management purposes. Another advantage of PLA is that it can be quantified for more animals than C-LA at a given

time after calving. Animals without a C-LA in the first 60 days have a missing C-LA, but will have a score for PLA. In this study, more lactations with P4 records received a measure for PLA ($n = 2,192$), than for C-LA (2,165), but PLA may have the disadvantage that it is affected by the number of samples taken in a fixed period. Because P4 records were sampled on average every two days, mean PLA (58%) was higher than the values of 37% and 47% that have been reported elsewhere (Petersson et al., 2006; Petersson et al., 2007), where P4 records were sampled one to three times per week. Furthermore, a decrease in mean PLA was also observed when P4 sampling frequency was decreased from twice weekly to twice monthly (Petersson et al., 2007). Another disadvantage of PLA is that, very high values could be associated with poor reproductive performance, e.g. persistent corpus luteum or pyometra (Etherington et al., 1991; Petersson et al., 2007). The occurrence of luteal activity during the first 60 days in milk (LA60) is a new trait that has the same advantages as PLA, but LA60 has the disadvantage of being a binary trait.

Repeatability estimates for endocrine fertility traits (0.15 to 0.29) were generally moderate, while those for classical fertility traits were low (0.09 to 0.12), thus supporting the hypothesis that endocrine measures of fertility are more influenced by the cow itself, and better reflect the cow's own physiology than classical fertility measures, which are biased by farm management decisions and recording practices. The repeatability for InC-LA was similar to that of Darwash et al. (1997a). Moreover, a moderate repeatability for InC-LA (0.29) suggests that cows with shorter intervals from calving to first ovulation have the tendency to have such short intervals after every calving.

2.4.2 Genetic parameters

Heritability estimates for endocrine fertility traits varied from 0.06 to 0.12. The heritability of InC-LA (0.12 ± 0.05) was at the lower bound of previously reported estimates (0.13 to 0.21) (Darwash et al., 1997a; Veerkamp et al., 2000; Berry et al., 2012). However, the genetic standard deviation of InC-LA (0.12) in this study, which is equivalent to approximately 5 days, implies that there is considerable genetic variation in InC-LA. Most previous studies defined C-LA and other endocrine traits using 3 ng/ml as threshold for luteal activity, whereas this study used 5 ng/ml. To investigate the effect of the threshold on the heritability of InC-LA, the analysis with 3 ng/ml was also performed, but that did not affect the heritability of InC-LA. Phenotypic variation in CLAFS has been studied (Royal et al., 2000), but no other estimates are available for the genetic variation. The genetic standard deviation for

CLAFS was 9 d, so also considerable genetic variation was observed for this trait. Heritability of PLA (0.12 ± 0.05), was lower than the estimate (0.29 ± 0.06) of Petersson et al. (2007). However, the genetic standard deviation for PLA (0.10) implies that there is considerable genetic variation also for this trait. The heritability of Clnt in this study was in agreement with previous studies (Wall et al., 2003; Berry et al., 2012), but the heritability of CFS was slightly larger than generally observed from field data.

Besides possible reasons like sample size, population structure and other factors that might influence heritability estimates, one possible reason for the lower estimates of endocrine fertility traits in this study compared to previous reports, might be the data editing steps applied. By excluding all lactations for cows not sampled before 25 DIM or for at least 60 DIM, a sampling bias might have been introduced, and hence biased estimates for endocrine fertility traits. However, the considerable genetic variation and moderate repeatability estimates of the endocrine fertility traits found in this study led to the conclusion that no important bias existed, and substantial genetic improvement can be achieved if the accuracy of identifying genetically superior animals is high.

The correlations of InC-LA with PLA (-0.91 ± 0.06) and CLAFS (-0.56 ± 0.17) implies that direct selection for shorter InC-LA will lead to increased reproductive activity, but also to increased CLAFS. The strong correlation of PLA with InC-LA means that PLA could be used for selection in cases where C-LA cannot be quantified for all animals, though some checks for reproductive abnormalities should be considered, as very high values of PLA are sometimes associated with persistent corpus luteum or pyometra. Comparable correlations between InC-LA and PLA (-0.97 ± 0.05) have been reported (Petersson et al., 2007). Also, there was a moderate genetic correlation (-0.31 ± 0.20) between PLA and CFS, whereas Petersson et al. (2007) found a low estimate (-0.03 ± 0.22) but of same sign, and both with large standard errors. The genetic correlation of InC-LA with CFS (0.37 ± 0.21) and Clnt (0.26 ± 0.33), imply that decreasing InC-LA by 1 genetic standard deviation will decrease CFS by 4.1 days and Clnt by 3.7 days. Some studies have examined the correlations between endocrine and classical fertility traits. Nyman et al. (2014) found a correlation of 0.54 ± 0.27 between InC-LA and Clnt. Fitting a genetic regression of daughter InC-LA on sire PTA for Clnt, Royal et al. (2002b) inferred a genetic correlation of 0.36 between InC-LA and Clnt, while Berry et al. (2012) reported an estimate of 0.87 ± 0.33 from an animal model. Similarly, positive genetic correlations between InC-LA and CFS

(0.37 ± 0.34 and 0.35 ± 0.12) have been reported (Berry et al., 2012; Nyman et al., 2014).

The results discussed thus far were estimated including the correlation between permanent environmental effects across traits (full model). Since there were too few (repeated) records available to estimate both the genetic and permanent environmental correlation between traits very accurately, two alternative analyses were performed to investigate the effect on the estimated genetic correlations. First a reduced model was fitted that excluded the permanent environmental correlation between traits (but included the permanent effect for each trait), and secondly an analysis without repeated records was performed. Comparing the likelihood from the reduced and full model, the permanent environmental covariance between traits could have been excluded for 25 (out of 33) of the analysis. However, excluding the permanent environmental covariance or not had an impact on some of the genetic correlation estimates. Especially for CLAFS, its genetic correlations with milk production traits from the full model were: 0.00 ± 0.25 for MY, 0.04 ± 0.25 for PY, and 0.42 ± 0.15 for FY, but moved to 0.40 ± 0.17 for MY, 0.45 ± 0.17 for PY, and 0.53 ± 0.19 for FY with the reduced model. Hence, the two models seemed to be conflicting, either suggesting no genetic correlation (full model) between CLAFS and MY or a moderate genetic correlation (reduced model). This illustrates the confounding between the permanent and genetic correlation. The analysis where all repeated records were deleted and genetic correlations re-estimated gave low genetic correlations of CLAFS with MY (0.09 ± 0.31), but 0.24 ± 0.32 for PY and 0.70 ± 0.44 for FY. However, the genetic correlations of the different models were not statistically significantly different from each other, and therefore there is still an open question of whether there is a positive genetic correlation between CLAFS and MY, or not. However, it is clear that properly accounting for a permanent environmental effect for each cow, that is likely to be influenced by voluntary waiting period determined by the farmer when a cow is inseminated or not, and having enough (repeated) records to separate these effects is important when concluding on the association between yield and fertility traits.

The positive genetic correlation between InC-LA and milk production traits suggests that there is an unfavorable influence of genetic merit for milk yield on genetic merit for InC-LA, but this correlation was considerably lower than estimates reported in previous studies. Veerkamp et al. (2000) reported higher correlations of C-LA with MY (0.51), PY (0.48), and FY (0.65) for 305-day yield records, while Royal et al. (2002a) found an estimate (0.36) for InC-LA with predicted milk yield on day 56, and Nyman et al. (2014) found a genetic correlation of 0.45 ± 0.09 for InC-LA with energy-

corrected milk during the first 60 DIM. Furthermore, the negative genetic correlations between PLA and milk production traits suggest an unfavorable influence of genetic merit for yield on genetic merit for PLA. No previous study has reported correlations for PLA and milk production traits. Nonetheless, the correlations of endocrine fertility traits and milk production traits in this study are considerably lower than has been observed between classical fertility and milk production traits (Pryce et al., 1997; Pryce et al., 1998). The results from this study support and expand the work of (Darwash et al., 1997a; Royal et al., 2002b; Petersson et al., 2007) that endocrine fertility traits might serve as better indicators of dairy cow fertility. In addition, Berry et al. (2012) were able to identify two markers on the genome associated with C-LA that have not been identified with classical fertility traits, suggesting that endocrine fertility traits might be more useful for mapping genomic regions and studying the biology of fertility. Moreover, endocrine fertility traits can also help provide valuable insights to the consequences of selection for different aspects of fertility, or on the other hand be very useful in understanding how selection for other traits impacts on fertility in great detail.

2.4.3 Application of endocrine fertility traits in breeding schemes

The breeding goal trait for fertility would be cows that resume luteal activity early after calving, start showing heat so they can be inseminated at the optimal time point, are pregnant after first insemination and calve successfully. None of these traits are measured on a sufficient large scale to select on these traits directly. The use of endocrine fertility traits might provide a better trait definition, with the added advantage that they provide an elaborate definition of the breeding goal. However, the use of endocrine fertility traits has been limited by the costs associated with frequent sampling of P4 records per cow and on a large number of animals required for genetic evaluation. Alternatives like sampling of daughters of bulls only during normal milk recording (Van der Lende et al., 2004), or sampling at the cow level on a monthly basis (Petersson et al., 2007) have been suggested. In-line P4 records may provide a new opportunity to address these limitations.

This study has provided a number of endocrine fertility traits with genetic variation that might serve as breeding goal traits. For example in breeding programs where CFS is used as index trait as indirect measure of calving to first ovulation, InC-LA is likely to be a more direct measure of return to reproductive cyclicity, as there is only a delay of 4 to 5 days between C-LA and first ovulation. Hence, in accordance with

Petersson et al. (2007), we suggest using InC-LA as a trait in the breeding goal, rather than CFS. Also, CLAFS might provide a measure for heat and heat detection, although CLAFS is partly affected by insemination decision, and PLA might be an interesting trait to be considered indicating reproductive abnormalities, because most of the detailed traits like LPL, ILI, and IOI gave zero heritability in this study. Nonetheless, classical fertility traits are measured on a large scale and might still serve as predictor traits in the national genetic evaluations. Thus, endocrine fertility traits provide the option of a more precise definition of the breeding goal.

Also, to be considered is the fact that only few herds at present possess the in-line recording system, although the number of herds with this system is increasing, and in the near future it might be possible to use endocrine fertility traits in genetic evaluation schemes. However, because of the few herds, in-line endocrine fertility traits would be most useful in a genomic selection scheme, where cows from collaborator herds would be part of the reference population and for the national breeding schemes genomic prediction are used for endocrine fertility traits. Therefore, combining (genomic predictions for) endocrine and classical fertility traits may provide the most accurate information for the selection of a more precise definition of dairy cattle breeding goals based on endocrine fertility traits.

2.5 Conclusions

We showed that in-line progesterone records can be used to define and explore several endocrine fertility traits in dairy cows. Some of these traits have genetic variation, and might help to define better breeding goals in order to select for improved fertility. Nonetheless, because there are only few herds at present with the in-line recording system, endocrine fertility traits would only be more useful in a genomic selection scheme.

Acknowledgements

The authors acknowledge Lattec I/S (Denmark) for providing the in-line progesterone data and the Cooperative Cattle Improvement Organization CRV BV (Arnhem, the Netherlands) for providing the pedigree data for this study. A. M. M. Tenghe benefited from a joint grant from the European Commission and the European Union Seventh Framework Programme (FP7/ 2007-2013) under the grant agreement n°311776, within the framework of the Erasmus-Mundus joint doctorate "EGS-ABG". This study was also performed as part of the Prolific project. The Prolific project is

financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement n°311776.

References

- Berry, D. et al. 2003. Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows. *J. Dairy Sci.* 86: 2193-2204.
- Berry, D. P. et al. 2012. Genome-wide associations for fertility traits in Holstein-Friesian dairy cows using data from experimental research herds in four European countries. *Animal* 6: 1206-1215.
- Bulman, D. C., and G. Lamming. 1978a. Milk progesterone levels in relation to conception, repeat breeding and factors influencing acyclicity in dairy cows. *J. Reprod. Fertil.* 54: 447-458.
- Bulman, D. C., and G. E. Lamming. 1978b. Milk progesterone levels in relation to conception, repeat breeding and factors influencing acyclicity in dairy cows. *J. Reprod. Fertil.* 54: 447-458.
- Campos, M. S., C. J. Wilcox, C. M. Becerril, and A. Diz. 1994. Genetic Parameters for Yield and Reproductive Traits of Holstein and Jersey Cattle in Florida. *J. Dairy Sci.* 77: 867-873.
- Darwash, A. O., G. E. Lamming, and J. A. Woolliams. 1997a. Estimation of genetic variation in the interval from calving to postpartum ovulation of dairy cows. *J. Dairy Sci.* 80: 1227-1234.
- Darwash, A. O., G. E. Lamming, and J. A. Woolliams. 1997b. The phenotypic association between the interval to post-partum ovulation and traditional measures of fertility in dairy cattle. *Anim. Sci.* 65: 9-16.
- Darwash, A. O., G. E. Lamming, and J. A. Woolliams. 1999. The potential for identifying heritable endocrine parameters associated with fertility in post-partum dairy cows. *Anim. Sci.* 68: 334.
- Etherington, W. et al. 1991. Progesterone profiles in postpartum Holstein dairy cows as an aid in the study of retained fetal membranes, pyometra and anestrus. *Theriogenology* 35: 731-746.
- Friggens, N. C., M. Bjerring, C. Ridder, S. Højsgaard, and T. Larsen. 2008. Improved Detection of Reproductive Status in Dairy Cows Using Milk Progesterone Measurements. *Reprod. Domest. Anim.* 43: 113-121.
- Gilmour, A. R., B. R. Cullis, and R. Thompson. 2009. ASReml Update: What's new in Release 3.00 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

- Hayes, J. F., R. I. Cue, and H. G. Monardes. 1992. Estimates of Repeatability of Reproductive Measures in Canadian Holsteins. *J. Dairy Sci.* 75: 1701-1706.
- Jansen, J. 1985. Genetic aspects of fertility in dairy cattle based on analysis of A.I. data — A review with emphasis on areas for further research. *Livest. Prod. Sci.* 12: 1-12.
- Lamming, G. E., and A. O. Darwash. 1998. The use of milk progesterone profiles to characterise components of subfertility in milked dairy cows. *Anim. Reprod. Sci.* 52: 175-190.
- Lucy, M. 2001. Reproductive loss in high-producing dairy cattle: where will it end? *J. Dairy Sci.* 84: 1277-1293.
- Marti, C. F., and D. A. Funk. 1994. Relationship Between Production and Days Open at Different Levels of Herd Production. *J. Dairy Sci.* 77: 1682-1690.
- Nyman, S. et al. 2014. Genetic analysis of atypical progesterone profiles in Holstein-Friesian cows from experimental research herds. *J. Dairy Sci.* 97: 7230-7239.
- Petersson, K.-J. et al. 2007. Genetic analysis of postpartum measures of luteal activity in dairy cows. *J. Dairy Sci.* 90: 427-434.
- Petersson, K. J., E. Strandberg, H. Gustafsson, and B. Berglund. 2006. Environmental effects on progesterone profile measures of dairy cow fertility. *Anim. Reprod. Sci.* 91: 201-214.
- Pollott, G., and M. Coffey. 2008. The effect of genetic merit and production system on dairy cow fertility, measured using progesterone profiles and on-farm recording. *J. Dairy Sci.* 91: 3649-3660.
- Pryce, J. et al. 1998. Estimation of genetic parameters using health, fertility and production data from a management recording system for dairy cattle. *Animal Science* 66: 577-584.
- Pryce, J., R. Veerkamp, R. Thompson, W. 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.
- R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Royal, M. D. et al. 2000. Declining fertility in dairy cattle: changes in traditional and endocrine parameters of fertility. *Anim. Sci.* 70: 487-501.
- Royal, M. D., A. P. F. Flint, and J. A. Woolliams. 2002a. Genetic and Phenotypic Relationships Among Endocrine and Traditional Fertility Traits and Production Traits in Holstein-Friesian Dairy Cows. *J. Dairy Sci.* 85: 958-967.
- Royal, M. D., J. E. Pryce, J. A. Woolliams, and A. P. F. Flint. 2002b. The Genetic Relationship between Commencement of Luteal Activity and Calving Interval, Body

- Condition Score, Production, and Linear Type Traits in Holstein-Friesian Dairy Cattle. *J. Dairy Sci.* 85: 3071-3080.
- Van der Lende, T. et al. 2004. Infrequent milk progesterone measurements in daughters enable bull selection for cow fertility. *J. Dairy Sci.* 87: 3953-3957.
- Veerkamp, R., J. Oldenbroek, and T. Van der Lende. 1998. The use of milk progesterone measurements for genetic improvement of fertility traits in dairy cattle. *Proc. Int. Workshop on Genetic Improvement of Functional Traits in Cattle; Fertility and Reproduction*, Grub, Germany. *Interbull Bull.* 18, Interbull, Uppsala, Sweden pp. 62–67.
- Veerkamp, R. F., J. K. Oldenbroek, H. J. Van Der Gaast, and J. H. J. V. D. Werf. 2000. Genetic Correlation Between Days Until Start of Luteal Activity and Milk Yield, Energy Balance, and Live Weights. *J. Dairy Sci.* 83: 577-583.
- Wall, E., S. Brotherstone, J. Woolliams, G. Banos, and M. Coffey. 2003. Genetic evaluation of fertility using direct and correlated traits. *J. Dairy Sci.* 86: 4093-4102.

3

Genome-wide association study for endocrine fertility traits using single nucleotide polymorphism arrays and sequence variants in dairy cattle

A.M. M. Tenghe¹²³, A. C. Bouwman², B. Berglund³, E. Strandberg³, DJ. de Koning³, R. F. Veerkamp¹²

¹ Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands

²Animal Breeding and Genomics Centre, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands

³Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, P. O. Box 7023, SE-750 07 Uppsala, Sweden

Journal of Dairy Sciences (2016) 99:5470-5485

Abstract

Endocrine fertility traits which are defined from progesterone concentration levels in milk are interesting indicators of dairy cow fertility because they more directly reflect the cows own reproductive physiology than classical fertility traits which are more biased by farm management decisions. The aim of this study was to detect quantitative trait loci (QTL) for 7 endocrine fertility traits in dairy cows by performing a genome-wide association study with 85k SNP, and then fine-map targeted QTL regions, using imputed sequence variants. Two classical fertility traits were also analyzed for QTL with 85k SNP. The association between a SNP and a phenotype was assessed by single-locus regression for each SNP, using a linear mixed model that included a random polygenic effect. A total of 2,447 Holstein Friesian cows with 5,339 lactations with both phenotypes and genotypes were used for association analysis. Heritability estimates ranged from 0.09 to 0.15 for endocrine fertility traits and 0.03 to 0.10 for classical fertility traits. The genome-wide association study identified 17 QTL regions for endocrine fertility traits, on *Bos taurus* autosomes (BTA) 2, 3, 8, 12, 15, 17, 23, and 25. The highest number (5) of QTL regions from the genome-wide association study were identified for the endocrine trait “proportion of samples with luteal activity”. Overlapping QTL regions were found between endocrine traits on BTA 2, 3 and 17. For the classical trait calving to first service, 3 QTL regions were identified on BTA 3, 15, and 23, and an overlapping region on BTA23 with endocrine traits. Fine-mapping target regions for the endocrine traits on BTA 2, and 3 using imputed sequence variants confirmed the QTL from the genome-wide association study, and identified several associated variants that can contribute to an index of markers for genetic improvement of fertility. Several potential candidate genes underlying endocrine fertility traits were also identified in the target regions and are discussed. However, due to high linkage disequilibrium, it was not possible to specify genes or polymorphisms as causal factors for any of the regions.

Key words: quantitative trait loci, milk progesterone, dairy cattle, fertility

3.1 Introduction

Genome-wide association studies (**GWAS**) have identified thousands of single nucleotide polymorphisms (**SNP**) across the cattle genome associated with important economic traits (Hu et al., 2016). Previous studies that have attempted to locate QTL for reproductive performance in dairy cattle have used mostly classical fertility traits, see Khatkar et al., (2004) for a review, and the cattle QTL database (Hu et al., 2016) for more recent studies. Endocrine fertility traits defined from progesterone concentration levels in milk have been suggested as alternative indicators for dairy cow fertility, because they more directly reflect a cow's reproductive physiology and are less influenced by farm management decisions (Bulman and Lamming, 1978; Darwash et al., 1999), compared to classical fertility traits which are defined from calving data and insemination records. Hence endocrine fertility trait might be more useful to detect fertility QTL. In addition, endocrine fertility traits seem to be more heritable than classical traits (Veerkamp et al., 2000; Petersson et al., 2007; Tenghe et al., 2015), but few GWAS studies have been performed for endocrine fertility traits (Berry et al., 2012).

Most of the genetic variants detected in GWAS are not the causal mutations for the traits, but in linkage disequilibrium (**LD**) with these causal polymorphisms. This is so mainly because the marker panels used in most of the studies only represent a small fraction of the common variants segregating in the population. However, advances in next generation sequencing techniques have led to sequencing a large number of animals in cattle. Also, imputation techniques offer the possibility to reliably impute genotyped animals to sequence variants (Browning and Browning, 2009). The advantage of using sequence data over SNP arrays for GWAS arises from the expectation that there will be higher precision in detecting QTL as the data is expected to include the causal variant, and there is less dependence on population LD (Meuwissen and Goddard, 2010; Daetwyler et al., 2014; Druet et al., 2014). Furthermore, recent studies have shown that the precision of mapped QTL can be increased by the use of sequence data (Sahana et al., 2014; Höglund et al., 2014a, 2015).

Genomic prediction helps to select breeding animals for the next generation more accurately at an early age. In addition to revealing the genetic architecture that underlies the physiological and biological process of female reproduction, detected QTL could also be practically applied to genomic selection schemes to improve fertility. The introduction of high-density SNP arrays like the 777k did not increase

accuracy of genomic predictions in cattle substantially (Su et al., 2012; VanRaden et al., 2013), one of the reasons being the increase in the number of unknown parameters to be estimated with high density data. However, Meuwissen and Goddard (2010) demonstrated in simulations that genomic predictions based on sequence data were up to 40 % more accurate than predictions based on ~ 30k SNP, because the causal mutations were used in prediction. Furthermore, in another simulation study, Druet et al. (2014) showed that, if the minor allele frequency of QTL is very low, genomic predictions from imputed sequence data can have up to 20% advantage in accuracy of genomic predictions from SNP panels. If the causal mutations influencing female fertility are detected, this information could be included in genomic prediction models where additional weight is put on the regions that influence female fertility; this would especially improve predictions over generations.

The aim of this study was to detect QTL for 7 endocrine fertility traits in dairy cows by performing a GWAS with 85k SNP, and then fine-map targeted QTL regions using imputed sequence variants. Two classical fertility traits were also analyzed for QTL with 85k SNP.

3.2 Materials and Methods

3.2.1 Animal population and phenotypes

The data consisted of in-line progesterone (P4) records from 14 commercial herds in the Netherlands, and manually collected P4 records of 4 experimental herds from: Wageningen UR Livestock Research, the Netherlands; Teagasc, Moorepark, Ireland; Scottish Agricultural College, United Kingdom; and the Jälla herd of Swedish University of Agricultural Science. In total, phenotypic data were available for 2,447 Holstein cows with 5,339 lactations.

A detailed description of the experimental treatments imposed on animals in the different experimental herds, procedures for milk sampling and P4 level measuring have been given in Veerkamp et al. (2000), Horan et al. (2005), Petersson et al. (2006), and Pollott and Coffey (2008). In brief, milk sampling for P4 measurement was undertaken twice a week at the experimental herds in Sweden and the Netherlands, and thrice a week in Ireland and the United Kingdom. In the commercial herds, milk sampling, measuring and recording of P4 level was performed by the

3 Quantitative trait loci for endocrine fertility traits

Herd Navigator™ (HN, DeLaval Intl, Tumba, Sweden). Sampling frequency on the commercial herds is based on a biological model (Friggensen et al., 2008), but on average undertaken every 2 days.

For each lactation, endocrine fertility traits were defined using P4 records as described in Tenghe et al., (2015) as follows: 1) commencement of luteal activity (**C-LA**) as the number of days between calving and first day on which milk P4 level was elevated (≥ 5 ng/ml); 2) proportion of samples with luteal activity (**PLA**) as the number of P4 records with luteal activity (P4 level ≥ 5 ng/ml), divided by total number of P4 records in the period from 25 to 60 days in milk; 3) luteal activity during first 60 days in milk (**LA60**) as the presence (LA60 = 1) or absence (LA60 = 0) of luteal activity between 25 and 60 days in milk; 4) commencement of luteal activity to first service (**CLAFS**) as the interval from the first day of elevated P4 level (≥ 5 ng/ml), fitting the luteal activity criteria, to day of first service; 5) first luteal phase length (**LPL**) as the interval from the first day of elevated P4 level (≥ 5 ng/ml) to the last consecutive day of elevated P4 level (≥ 5 ng/ml); 6) length of inter-luteal interval (**ILI**) as the interval from the first day of decreased P4 level (< 5 ng/ml) following the luteal phase, and the last consecutive day of decreased P4 level (< 5 ng/ml); and 7) length of first inter-ovulatory interval (**IOI**) as the interval from the first day of elevated P4 level (≥ 5 ng/ml) of one oestrus cycle to the first day of elevated P4 level (≥ 5 ng/ml) of the following oestrus cycle. In defining the endocrine fertility traits, a set of restrictions was applied to periods of non-sampling of P4 records that might occur during a lactation period. When a gap was cow-specific, that is, when the cows lactation had no P4 samples for more than 7 days, the following restrictions were applied: 1) if the gap occurred 1 or more days before C-LA, and the gap duration was less than 15 days, all traits were retained, otherwise all traits were excluded; 2) if the gap occurred at least 25 days after C-LA, all traits were retained, otherwise all traits were excluded except C-LA. When a gap was herd-specific, the following restrictions were applied to all lactations on that herd; for each lactation, when a gap of 3 or more days occurred 1 or more days before C-LA and there were low P4 values after the gap, C-LA and all traits were retained, whereas when a gap of 3 or more days occurred after C-LA, the following restrictions were applied: a) LPL was retained if the gap occurred at least 25 days after C-LA; b) ILI was retained if gap occurred 1 or more days before start of the inter-luteal interval; c) IOI was retained if gap occurred 1 or more days before start of the next luteal cycle.

The classical fertility traits interval from calving to first service (**CFS**) and calving interval (**Clnt**) were also analyzed in the GWAS. The trait Clnt was restricted between 300 and 700 d, whereas CFS was restricted to 30 and 250 d.

3.2.2 Genotypes and imputation

Genotyping of cows from the experimental herds was performed with the Illumina BovineSNP50 v1 BeadChip (Illumina Inc., San Diego, CA, USA) on 1,946 cows and contained 54,001 SNP (50k), whereas 1,907 cows on commercial herds were genotyped with the GeneSeek Genomic Profiler Bovine HD (GeneSeek, Lincoln, NE, USA), containing 76,883 SNP (80k). An additional 6 cows from a commercial herd were genotyped with a custom Illumina 6k array. The 50k and 80k SNP arrays had 25,815 SNP in common, the rest were imputed from one dataset to the other and vice versa using BEAGLE 3.3.2. (Browning and Browning, 2009), resulting in a total of 102,062 SNPs. The quality criteria applied before imputation were minimum call rate of 95% per animal and 97% per SNP, and minimum GC-score of 0.3. A total of 120 animals failed the quality criteria and were excluded. After imputation, SNPs were retained for association analysis if they fulfilled all of the following filtering criteria: 1) call rate > 95%, minor allele frequency (MAF) > 0.01, and no extreme deviation from Hardy-Weinberg Equilibrium (p -value < 0.01). The final genotype dataset contained 85,485 SNP for 3,739 cows.

Table 3.1 Target regions from sequence data selected based on a genome-wide association study with 85,485 SNP in *Bos taurus*

BTA	Target region (Mb)	Position top SNP (bp)	No. of significant SNP	No. of SNP on 85k	No. of variants on sequence	Traits ¹ showing association
2	1.87 – 15.30	6,852,950	8	202	161,426	C-LA, CLAFL, LPL
2	31.87 – 41.96	36,891,323	3	320	97,548	LPL, IOI
3	85.68 – 95.66	90,669,666	6	348	89,632	C-LA, PLA, LA60

¹C-LA = commencement of luteal activity, PLA = proportion of samples in luteal activity between 25 and 60 days in milk, LA60 = luteal activity between 25 and 60 days in milk, CLAFL = interval from commencement of luteal activity to first service, LPL = first luteal phase length, IOI = length of first inter-ovulatory interval

3.2.3 Statistical analysis

A total of 2,447 Holstein Friesian cows with 5,339 lactations with both phenotypes and genotypes were used for all association analysis. For both fine mapping and the GWAS, the association between a SNP and a phenotype was assessed by a single-locus regression for each SNP separately, using a linear mixed model that included a

random polygenic effect. Because there were repeated lactations per cow, a random permanent environmental effect was also fitted. The model was as follows:

$$y_{ijklm} = \mu + \text{parity}_i + \text{hys}_j + b_1 \times \text{ca}_k + b_2 \times M_l + a_m + \text{pe}_m + e_{ijklm}$$

where y_{ijklm} is the phenotype; μ is the intercept, parity_i is the fixed effect of parity i (cows in parity 4 or above were grouped to a common class); hys_j is the fixed effect of herd-year-(calving) season combination j ($j=1-495$), with calving season defined as winter: December-February, spring: March-May, summer: June-August, autumn: September-November; $b_1 \times \text{ca}_k$ is the regression on calving age k (in months) with coefficient b_1 ; M_l is the genotype score (0, 1, 2 for the GWAS with SNP-chip data, but genotype dosage [any value between 0-2] for the regional association study with sequence variants) of SNP l for individual m ; b_2 is the additive effect of SNP l ; a_m is the random polygenic effect of animal m , a_m was assumed to be normally distributed with mean zero and variance $A\sigma_a^2$, where A is the additive genetic relationship matrix; pe_m is the random permanent environmental effect of animal m to account for repeated measures (i.e. more than one lactation) within animal, pe_m was assumed to be normally distributed with mean zero and variance σ_{pe}^2 , and e_{ijklm} is the random error term, with residuals assumed to be normally distributed, with mean zero and variance σ_e^2 . The variance components were re-estimated for each SNP analyzed. The F-test of the null hypothesis (no QTL, i.e., $\beta_{\text{SNP}} = 0$) against the alternative (there is a QTL, i.e., $\beta_{\text{SNP}} \neq 0$) was performed for each SNP.

Heritability and repeatability estimates were calculated from univariate analyses using the same model, but with the SNP effect excluded. Because LA60 is binary a trait, it was analyzed with a threshold model, which included the same fixed and random effects as described above. All statistical analysis were performed in ASReml 4.1 (Gilmour et al., 2014).

3.2.4 Imputation to sequence

The multi-breed sequenced population from the 1,000 Bulls Genomes Project Run 4 was used as reference population for imputation to sequence and contained 1,147 sequenced animals with on average 11-fold coverage. The reference population contained 311 Holstein bulls, but all individuals were used as reference, because earlier studies have shown that a multi-breed sequenced reference population can be beneficial for imputation accuracy, especially for variants with low minor allele frequency (Bouwman and Veerkamp, 2014; Daetwyler et al., 2014). Imputation was

done using MINIMAC2 (Fuchsberger et al., 2015) with pre-phased genotypes from BEAGLE. As described by Daetwyler et al. (2014), the genotype calls of 1,000 Bull Genomes reference population were improved with BEAGLE using genotype likelihoods from SAMtools (Li et al., 2009) and inferred haplotypes in the samples, the resulting phased genotypes were used directly as pre-phased reference genotypes in MINIMAC2. Also, the target population had pre-phased genotypes due to the imputation step in BEAGLE described above.

Defining QTL regions. QTL regions from the GWAS were defined based on the location of the significant SNPs. A chromosomal region was defined as a QTL when at least two SNPs within a 10 Mb region showed genome-wide significant association with a trait. Single, isolated, significant SNP were excluded from QTL definition, because they have a high risk of representing a false positive.

Significant associations. In association analyses with a dense marker map, results are obscured by issues related to multiple testing and high correlations between close markers due to LD. As suggested by Teyssèdre et al., (2012), three p-value thresholds were used to identify and describe regions of interest. The most stringent threshold 10^{-6} was chosen as an approximation of the tests corrected with Bonferroni. The threshold of 10^{-5} was used to provide moderate evidence of association (Burton et al., 2007). A less stringent threshold of 10^{-4} was used to detect associations from the GWAS scan, and to describe and compare QTL between traits.

Selecting target regions. Three chromosomal regions were targeted for fine mapping based on results of the GWAS with 85,485 SNPs for the endocrine fertility traits. The target regions were selected based on association signals across multiple traits. The full chromosomes containing these regions were then imputed to sequence level, and the regions for fine-mapping extracted. There were two target regions on BTA2, and one target region on BTA3. The number of imputed sequence variants in the target regions ranged from 89,632 to 161,426. The lengths of the target regions and number of imputed variants (SNP and short insertions and deletions) based on whole-genome sequence are listed in Table 3.1. Filtering out imputed variants, with $MAF < 0.01$, resulted in 109,051 and 57,622 variants for the target regions of BTA 2 (1.87 – 15.30 Mb and 31.87 – 41.96 Mb respectively), and 58,894 variants on BTA3 (85.68 – 95.66 Mb). Region-wise association analysis with the sequence variants was performed with the same model used in the GWAS.

Candidate genes. Based on the region-wise association results for each target region, putative candidate genes were searched for in the regions flanking the most significant variant. We kept the genes closest to the most significant variant. The candidate genes were identified based on the *Bos taurus* UMD3.1 genome assembly (Zimin et al., 2009), and visualized using locusZoom (Pruim et al., 2010). The gene information was extracted from the Ensembl data base <http://www.ensembl.org> (Cunningham et al., 2015): Ensembl *Bos taurus* 81.31, UMD3.1.

3.3 Results

For all the fertility traits analyzed, descriptive statistics, heritability and repeatability estimates are in Table 3.2. Heritability estimates ranged from 0.03 to 0.15, whereas repeatability estimates ranged from 0.14 to 0.34. The trait ILI had zero heritability, hence was excluded from further analysis.

3.3.1 Genome-wide association study

The average inflation factor of p -values from the GWAS was 1.08 ± 0.061 , with a minimum of 0.99 for IOI, and a maximum of 1.16, for PLA, indicating relatively good concordance between observed and assumed distributions of the test statistics. The quantile-quantile plots of all the GWAS are in Supplementary Figure 3. S1.

The Manhattan plots for all fertility traits analyzed with the 85,485 SNPs are in Figure 3.1, and Figure 3.2 shows the location of all SNP-trait associations on the 30 chromosomes. In total, 150 SNP-trait associations [$-\log_{10}(p\text{-value}) > 4$] were detected across all chromosomes except on BTA 4, 22, 26 and 28. For endocrine fertility traits, there were 123 SNP-trait associations on all chromosomes except on BTA 4, 22, 26 and 28. The greatest number of associated SNP were on BTA2, followed by BTA3 and BTA17. All associated SNP from the GWAS are in Supplementary Table 3. S1 (online at JDS).

A total of 20 QTL regions (of which 17 were for endocrine traits) were identified for all the fertility traits except Clnt, on 8 chromosomes (BTA 2, 3, 8, 12, 15, 17, 23, and 25), Table 3.3. The trait PLA had the highest number of QTL regions (5), followed by CLAFS (4). The traits LA60, LPL and CFS each had 3 QTL regions, whereas C-LA and IOI had a QTL region each. As shown in Table 3.4, the top associated SNP in a QTL for C-LA and PLA were observed on BTA3 at 90.67 Mb, and 95.24 Mb for LA60. The top QTL for CLAFS was on BTA2 (74.44 Mb), whereas that of LPL and IOI was on BTA2 at 36.89 Mb. The magnitude of estimated SNP effects (Table 3.4) was expressed in trait

phenotypic standard deviation units. The absolute SNP effects ranged from 0.12 for CFS to 0.93 for LA60.

Two QTL regions were shared by multiple endocrine fertility traits on BTA 2 and 3 (Table 3.3). The QTL region on BTA2 (36.88 – 37.05 Mb) was shared by LPL and IOI. The top associated SNP in this region was for LPL [$-\log_{10}(p\text{-value}) = 4.96$], with a total of 3 SNP associations for both LPL and IOI; also, the other 2 associated SNP had the same p -value as the top associated SNP, suggesting strong LD. The region on BTA3 (90.67 – 95.24 Mb) was shared by C-LA, PLA and LA60, with the topmost association being for PLA [$-\log_{10}(p\text{-value}) = 6.41$], followed by LA60 and C-LA (Table 3.4).

There were 29 SNP-trait associations [$-\log_{10}(p\text{-value}) > 4$] identified for classical fertility traits (CFS and Clnt) across 18 chromosomes (Figure 3.2). The most significant SNP association for Clnt [$-\log_{10}(p\text{-value}) = 5.78$] was on BTA11 (76.42 Mb). A total of 3 QTL regions were identified for CFS (on BTA 3, 15 and 23), and none for Clnt (Table 3.3). The top QTL region for CFS ranged from 27.78 – 30.35 Mb on BTA23 [$-\log_{10}(p\text{-value}) = 6.35$], which was also associated to CLAFS [$-\log_{10}(p\text{-value}) = 4.50$].

Table 3.2 Number of lactations (n) per trait, overall mean, standard deviation (SD), estimates of heritability (h^2), repeatability (t) and their SE for various endocrine fertility traits.

Trait ¹	n	Mean	SD	h^2	SE	t	SE
C-LA (d)	3,524	38.53	18.04	0.14	0.04	0.34	0.03
PLA (%)	3,597	0.56	0.30	0.15	0.04	0.32	0.03
LA60 (0-1)	3,597	0.87	0.33	0.15	0.06	0.19	0.03
CLAFS (d)	2,015	51.23	30.83	0.09	0.05	0.29	0.04
LPL (d)	1,785	12.54	11.09	0.15	0.05	0.15	0.05
IOI (d)	1,658	26.71	14.91	0.14	0.05	0.14	0.05
ILI (d)	1,637	10.95	11.04	0.00	0.00	0.00	0.06
CFS (d)	3,631	90.85	33.72	0.10	0.04	0.15	0.02
Clnt (d)	2,031	408.46	65.17	0.03	0.04	0.11	0.03

¹C-LA = Commencement of luteal activity; PLA = proportion of samples in luteal activity between 25 and 60 days in milk; LA60 = luteal activity between 25 and 60 days in milk, CLAFS = interval from commencement of luteal activity to first service; LPL = first luteal phase length; IOI = length of first inter-ovulatory interval; ILI = inter-luteal interval; CFS = interval from calving to first service; Clnt = calving interval

3 Quantitative trait loci for endocrine fertility traits

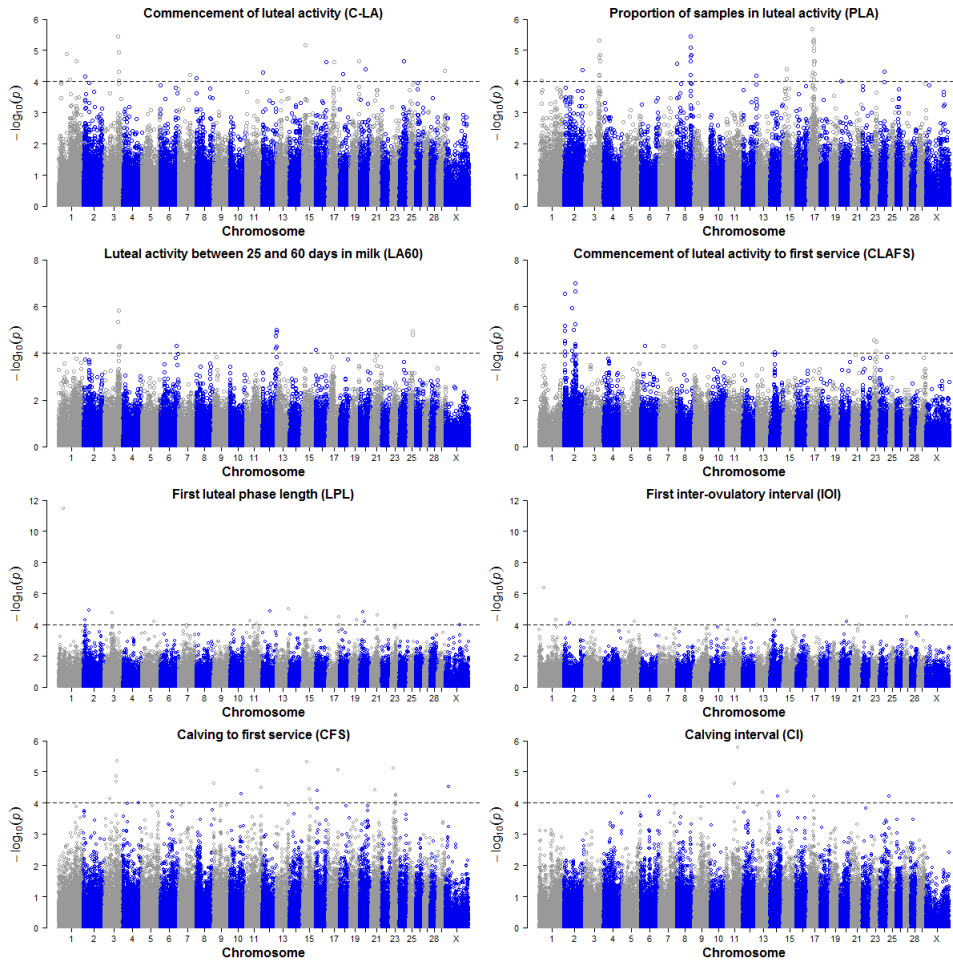


Figure 3.1 Genome-wide scan for 6 endocrine and 2 classical fertility traits using 85,485 SNPs of 2,447 cows with 5,339 lactations in dairy cattle. The y-axis shows the $-\log_{10}(p\text{-value})$ for each SNP. The x-axis shows the chromosomes and SNP position; chromosomes are shown in alternating colors for clarity. The dotted line represents the significance threshold [$-\log_{10}(p\text{-value}) = 4$] as considered in the present study

3 Quantitative trait loci for endocrine fertility traits

Table 3.3 The QTL³ regions on chromosomes of *Bos taurus* (BTA) associated to endocrine and classical fertility traits

BTA	Trait ¹	No. ²	Start QTL (bp)	End QTL (bp)
2	CLAFS	5	6,852,950	7,561,006
2	LPL	3	36,777,703	37,051,246
2	IOI	3	36,777,703	37,051,246
2	CLAFS	2	53,979,862	55,566,541
2	CLAFS	6	71,779,928	74,548,518
3	LPL	2	50,948,690	51,069,220
3	CFS	3	79,997,340	80,486,264
3	CLA	4	90,669,666	95,240,284
3	PLA	7	90,669,666	100,563,383
3	LA60	4	90,669,666	96,751,550
8	PLA	6	91,395,466	97,118,294
12	LA60	5	86,501,602	90,422,712
15	PLA	2	24,225,492	24,257,848
15	CFS	2	43,813,899	51,499,475
17	PLA	8	25,146,353	34,816,827
17	PLA	5	35,229,850	40,308,609
17	LPL	2	70,962,192	72,151,374
23	CFS	5	27,776,075	30,349,749
23	CLAFS	2	28,855,836	29,385,602
25	LA60	2	23,315,423	23,372,664

¹C-LA = Commencement of luteal activity; PLA = proportion of samples in luteal activity between 25 and 60 days in milk; LA60 = luteal activity between 25 and 60 days in milk, CLAFS = interval from commencement of luteal activity to first service; LPL = first luteal phase length; IOI = length of first inter-ovulatory interval; CFS = interval from calving to first service

²No. = Number of significant SNP within the QTL

³QTL region defined as a chromosomal region with multiple significant SNP within a window of 10 Mb

3 Quantitative trait loci for endocrine fertility traits

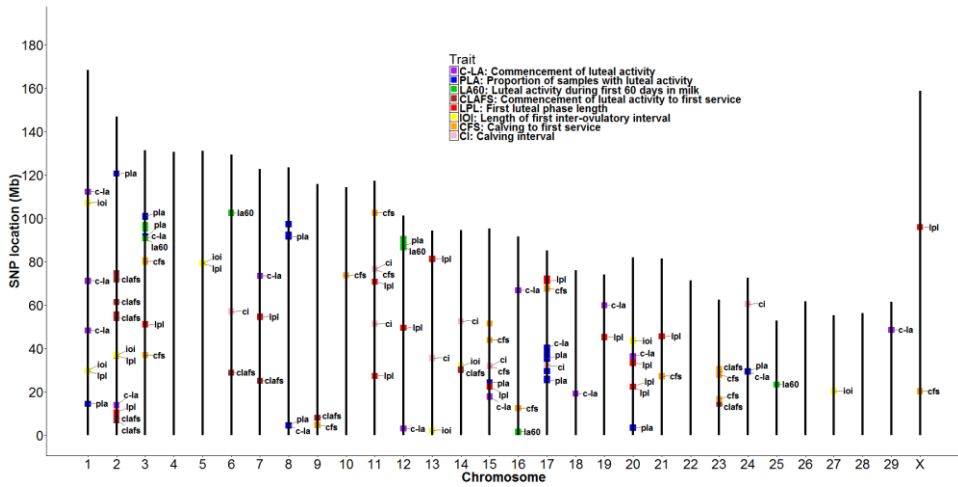


Figure 3.2 Distribution along the chromosomes of SNPs with p -value $< 10^{-4}$, identified for 6 endocrine and 2 classical fertility traits in dairy cattle, from a genome-wide scan using 85,485 SNP of 2,447 cows with 5,339 lactations

Table 3.4 The most significantly associated SNP in QTL4 regions identified on chromosomes of *Bos taurus* from a genome-wide association study of fertility traits with 85,485 SNP

BTA	Trait ¹	Position top SNP	MAF ²	$-\log(p\text{-value})$	Effect in SD ³	SE in SD ³
2	CLAFS	6,852,950	0.33	6.53	- 0.19	0.04
2	LPL	36,777,703	0.06	4.96	0.33	0.07
2	IOI	36,777,703	0.06	4.09	0.29	0.07
2	CLAFS	55,566,541	0.03	5.91	- 0.45	0.09
2	CLAFS	74,444,101	0.45	7.01	- 0.18	0.03
3	LPL	51,069,220	0.31	4.78	- 0.18	0.04
3	CFS	80,486,264	0.4	5.34	- 0.12	0.03
3	CLA	90,669,666	0.2	5.46	0.16	0.04
3	PLA	90,669,666	0.2	6.41	- 0.17	0.03
3	LA60	95,240,284	0.22	5.83	- 0.86	0.10
38	PLA	91,401,417	0.11	5.45	0.21	0.05
12	LA60	90,422,712	0.26	5.00	0.85	0.11
15	PLA	24,257,848	0.21	4.40	- 0.14	0.03
15	CFS	43,813,899	0.03	4.45	0.32	0.08

17	PLA	25,146,353	0.32	5.68	- 0.14	0.03
17	PLA	36,525,393	0.16	5.10	0.17	0.04
17	LPL	70,962,192	0.07	4.52	0.30	0.07
23	CLAFS	28,855,836	0.49	4.50	- 0.14	0.03
23	CFS	28,855,836	0.49	6.35	- 0.13	0.03
25	LA60	23,315,423	0.12	4.94	- 0.93	0.12

¹C-LA = Commencement of luteal activity; PLA = proportion of samples in luteal activity between 25 and 60 days in milk; LA60 = luteal activity between 25 and 60 days in milk, CLAFS = interval from commencement of luteal activity to first service; LPL = first luteal phase length; IOI = length of first inter-ovulatory interval; CFS = interval from calving to first service

²MAF = Minor allele frequency

³SD = Allele substitution effect expressed in phenotypic standard deviation of trait

⁴QTL region defined as a chromosomal region with multiple significant SNP within a window of 10 Mb

3.3.2 Fine-mapping with sequence variants

Three target regions on BTA 2 and 3 were fine-mapped for 6 endocrine fertility traits by region-wise association analysis using imputed sequence variants (Table 3.1). A total of 4,747 associations [$-\log_{10}(p\text{-value}) > 4$] between sequence variants and 6 endocrine fertility traits were identified on all target regions. All variants associated to the analysed endocrine fertility traits are in Supplementary Table 3. S2 (online at JDS). The most likely candidate genes for each target region, based on proximity to the most significant variant are in Table 3.6.

The target region on BTA2 (1.87 – 15.30 Mb) was fine-mapped for C-LA, CLAFS and LPL, and variant-trait associations were observed for all traits. In total, 45 associations were identified for C-LA, 716 for CLAFS and 564 for LPL. The most significant association was observed for CLAFS [$-\log_{10}(p\text{-value}) = 6.89$] at 7.18 Mb (Table 3.5), followed by LPL [$-\log_{10}(p\text{-value}) = 6.62$; at 8.45 Mb], and C-LA [$-\log_{10}(p\text{-value}) = 4.75$; at 14.02 Mb]. Based on the region-wise association results, candidate genes were searched for in the region spanning 12,499,856-15,303,929 bp for C-LA, and 3,499,960-10,999,836 bp for CLAFS and LPL (Figure 3.3). A total of 83 genes were identified for CLAFS and LPL, of which 61 are annotated and 22 uncharacterized. There were 24 genes for C-LA, with 16 annotated. The annotated genes closest to the most significant variant were *PDE1A* for C-LA, *SLC40A1* for CLAFS, and *GULP1* for LPL (Table 3.6).

3 Quantitative trait loci for endocrine fertility traits

Table 3.5 The most significantly associated variants from region-wise association analysis of 6 endocrine fertility traits in target regions of *Bos taurus* chromosomes using imputed sequence variants

BTA	Trait ¹	Top variant	Position	$-\log_{10}(p\text{-value})$	MAF ²	Effect in SD ³	SE in SD ³	rs-ID ⁴
2	C-LA	Chr2:14019875	14,019,875	4.75	0.01	3.68	0.85	rs472149103
2	CLAFS	Chr2:7175833	7,175,833	6.89	0.08	1.38	0.26	rs479466942
2	LPL	Chr2:8453071	8,453,071	6.62	0.04	- 7.68	1.47	rs43288488
2	LPL	Chr2:39502686	39,502,686	8.96	0.01	- 2.17	0.35	rs110097322
2	IOI	Chr2:40258911	40,258,911	5.31	0.01	- 1.91	0.42	rs378841057
3	C-LA	Chr3:89795380	89,795,380	6.10	0.18	- 0.22	0.04	rs451081085
3	PLA	Chr3:89360154	89,360,154	7.84	0.08	0.88	0.15	rs461884243
3	LA60	Chr3:89360154	89,360,154	7.11	0.08	4.45	0.45	rs461884243

¹C-LA = Commencement of luteal activity; PLA = proportion of samples in luteal activity between 25 and 60 days in milk; LA60 = luteal activity between 25 and 60 days in milk, CLAFS = interval from commencement of luteal activity to first service; LPL = first luteal phase length; IOI = length of first inter-ovulatory interval

²MAF = Minor allele frequency

³SD = Allele substitution effect expressed in phenotypic standard deviation of trait

⁴rs-ID = reference SNP identification

Two traits (LPL and IOI) were fine-mapped for the second target region on BTA2 (31.87 – 41.96 Mb). There were 477 variants associated with LPL and 70 with IOI. The most significant association for LPL [$-\log_{10}(p\text{-value}) = 9.09$] was at 39.45 Mb, and at 39.48 Mb for IOI [$-\log_{10}(p\text{-value}) = 6.03$]. The region between 33,500,042-40,999,921 bp was searched for candidate genes, and resulted in 62 genes, of which 45 were annotated and 17 were uncharacterized. The annotated gene closest to the most significant variant for both LPL and IOI was *NR4A2* (Figure 3.4), and another interesting candidate gene was *ACVR1*.

The target region on BTA3 (85.68 – 95.66 Mb) was fine-mapped for C-LA, PLA and LA60, and variant-trait associations were observed for all three traits. There were 594 variants associated with C-LA, 1,298 with PLA and 983 with LA60. The most significant association was observed for PLA [$-\log_{10}(p\text{-value}) = 7.84$] at 89.4 Mb (Table 3.5); this SNP was also the most significant for LA60 [$-\log_{10}(p\text{-value}) = 7.11$]. For C-LA, the most significant association [$-\log_{10}(p\text{-value}) = 6.10$] was at 89.8 Mb. Candidate genes were searched for in the region spanning 87,500,158-95,662,104 bp. There were 105 genes in this region, of which 76 have been annotated, and 26 were

described as uncharacterized. The annotated genes closest to the most significant variant were *DAB1* for C-LA and *C8B* for PLA and LA60 (Figure 3.5).

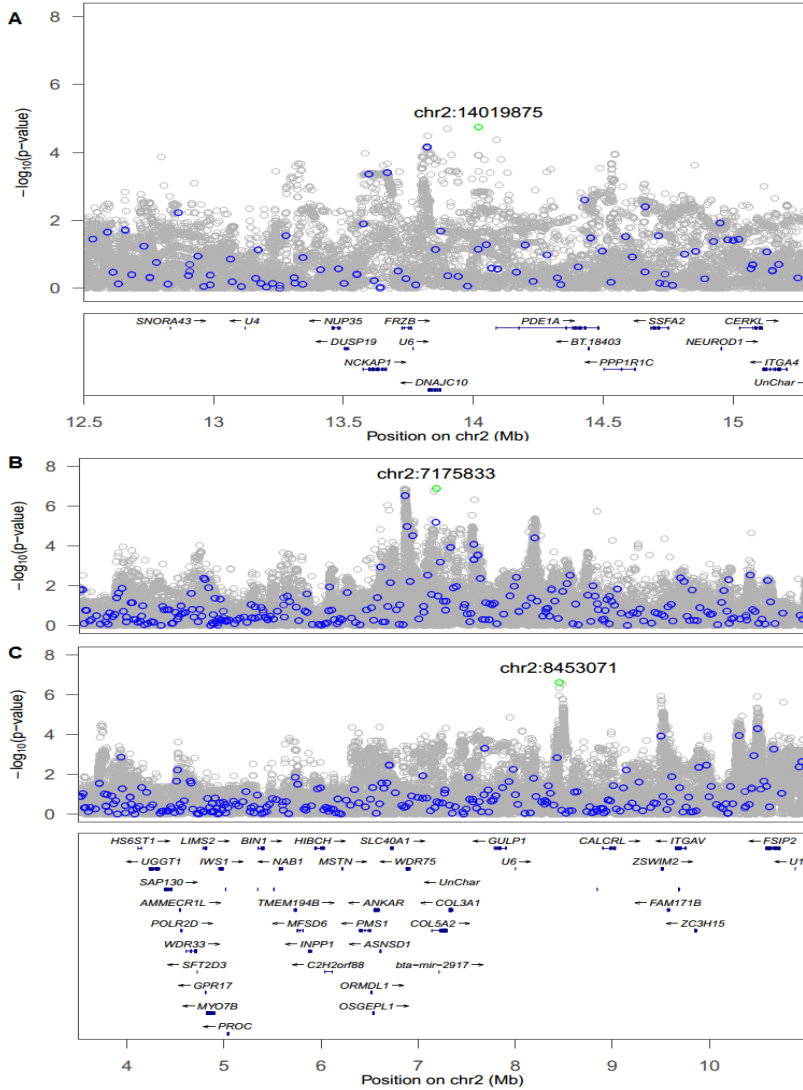


Figure 3.3 Region-wise association using sequence variants within the target region on BTA2 from 12,499,856-15,303,929 bp for A) C-LA, and 3,499,960-10,999,836 bp for B) CLAFS and C) LPL. Blue circles: genome-wide association with 85,485 SNPs of 2,447 cows; gray circles: region-wise association analysis with imputed sequence variants; labeled green circle is the most significant variant in the region. Candidate genes in the target regions are annotated in the boxes below each target region.

3 Quantitative trait loci for endocrine fertility traits

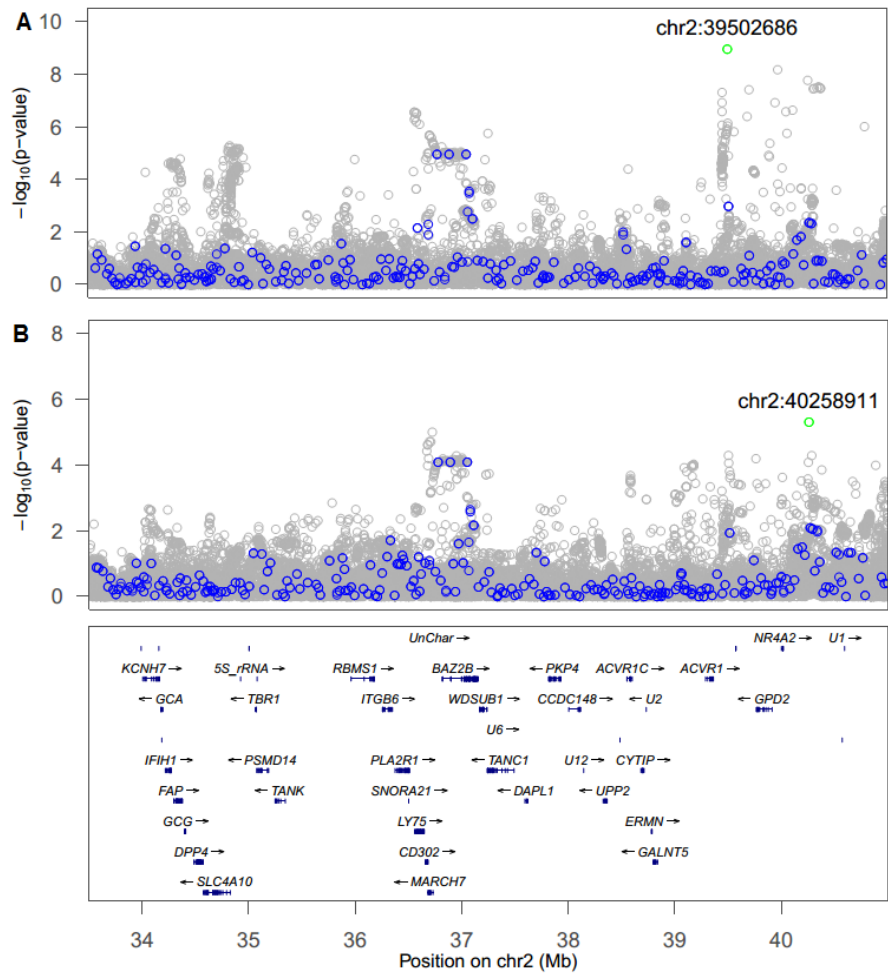


Figure 3.4 Region-wise association using sequence variants within the target region on BTA2 from 33,500,042-40,999,921 bp for A) LPL, and B) IOI. Blue circles: genome-wide association with 85,485 SNPs of 2,447 cows; gray circles: region-wise association analysis with imputed sequence variants; labeled green circle is the most significant variant in the region. Candidate genes are annotated below the target region.

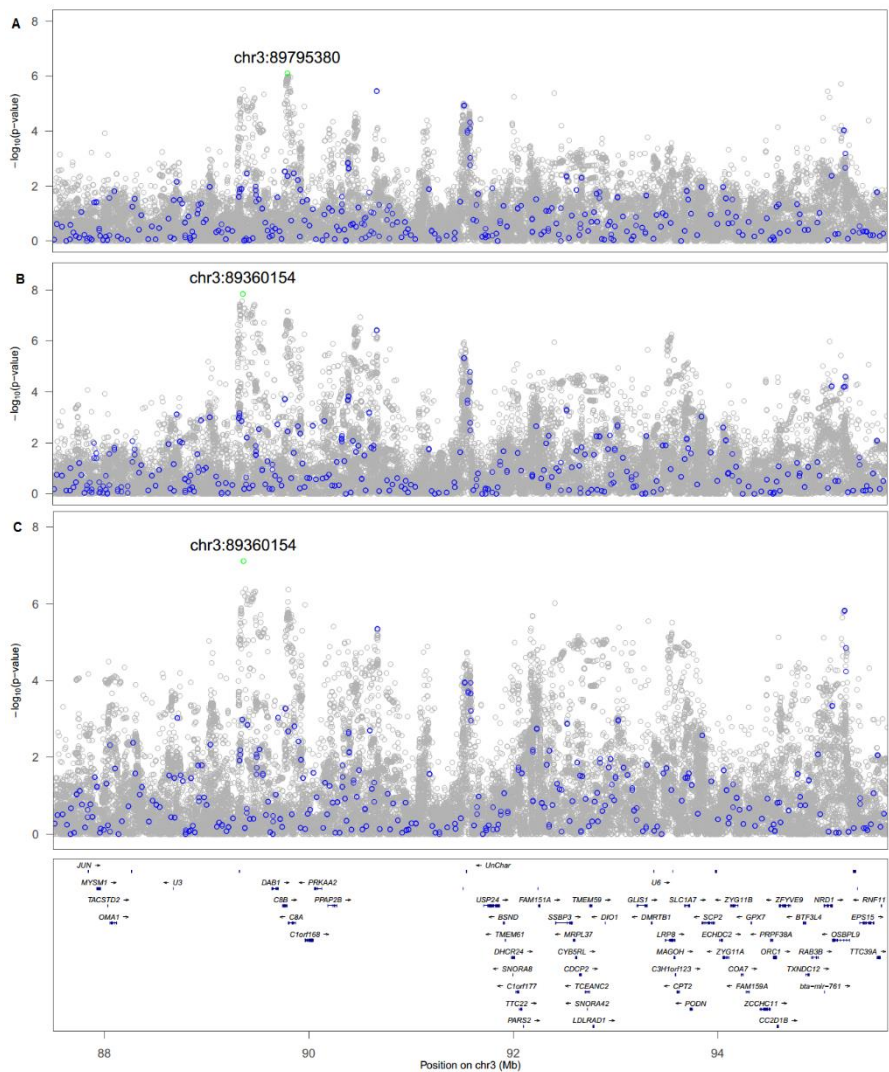


Figure 3.5 Region-wise association using sequence variants within the target region on BTA3 from 87,500,158-95,662,104 bp for A) C-LA, B) PLA, and C) LA60. Blue circles: genome-wide association with 85,485 SNPs of 2,447 cows; gray circles: region-wise association analysis with imputed sequence variants; labeled green circle is the most significant variant in the region. Candidate genes are annotated below the target region.

3 Quantitative trait loci for endocrine fertility traits

Table 3.6 The most likely candidate genes (Top candidate) for 6 endocrine fertility traits on target regions of *Bos taurus* chromosomes

BTA	Trait ¹	Top candidate	Start gene (Mb)	End gene (Mb)	Gene description
2	C-LA	PDE1A	14.09	14.48	Phosphodiesterase 1A, calmodulin-dependent
2	CLAFS	SLC40A1	6.72	6.74	Solute carrier family 40 (iron-regulated transporter)
2	LPL	GULP1	7.83	8.07	GULP, engulfment adaptor PTB domain containing 1
2	LPL, IOI	NR4A2	40.00	40.02	Nuclear receptor subfamily 4, group A, member 2
2	LPL, IOI	ACVR1	39.26	39.36	Activin A receptor, type I
3	C-LA	DAB1	88.37	89.72	Dab, reelin signal transducer, homolog 1 (Drosophila)
3	PLA, LA60	C8B	89.74	89.79	Complement component 8, beta polypeptide

¹C-LA = Commencement of luteal activity; PLA = proportion of samples in luteal activity between 25 and 60 days in milk; LA60 = luteal activity between 25 and 60 days in milk, CLAFS = interval from commencement of luteal activity to first service; LPL = first luteal phase length; IOI = length of first inter-ovulatory interval

3.4 Discussion

Previous studies have shown that endocrine fertility traits derived from milk progesterone profiles have higher heritability than classical fertility traits (Veerkamp et al., 2000; Petersson et al., 2007; Tenghe et al., 2015). Heritability estimates of the endocrine fertility traits in this study were similar to estimates reported by Veerkamp et al. (2000), and Nyman et al. (2014) who used a subset of the data used in this study, and estimates for classical fertility traits were similar to those of previous studies (Campos et al., 1994; Pryce et al., 1998; Berry et al., 2003; VanRaden et al., 2004). The trait ILI had zero heritability. Tenghe et al. (2015) also reported a heritability of zero for ILI, although (Nyman et al., 2014) reported a heritability of 0.08.

3.4.1 Genome-wide association study

Detecting QTL for fertility traits is of high importance due to the low heritability of these traits. In this study, a GWAS was performed for 6 endocrine and 2 classical fertility traits using 85k SNP. The GWAS identified 20 QTL regions for all the fertility traits except Clnt on BTA 2, 3, 8, 12, 15, 17, 23, and 25. Three selected target QTL regions on BTA 2 and 3, were then fine-mapped for endocrine fertility traits using imputed sequence variants. The endocrine trait PLA had the highest number (5) of QTL regions from the GWAS. The top QTL for PLA on BTA3 (90.67 Mb) was also associated with C-LA, and with LA60 (at 95.24 Mb), an observation consistent with the genetic correlation between these traits. Tenghe et al. (2015) reported strong

genetic correlations of -0.91 for C-LA with PLA and 0.82 for PLA with LA60. Up till now, no QTL have been reported for C-LA in this region. Using a subset (1,570 cows from the experimental herds; 50k SNP) of the data used in this study, Berry et al. (2012) identified QTL for C-LA at 130,141,723 bp on BTA2 and 9,375,095 bp on BTA21. No QTL was identified for C-LA on BTA21 in this study, however, a single SNP association was observed at 13,821,980 bp on BTA2, which is 116.32 Mb away from the QTL observed by Berry et al. (2012). Discrepancies between sample size, methods of analysis, and marker density could explain the differences between the results of Berry et al. (2012) and our study.

To the best of our knowledge, with the exception of C-LA, no QTL have been reported for the other endocrine fertility traits analyzed in this study. In this study, the top QTL for LPL on BTA2 was also associated to IOI. As expected from the trait definition, IOI combines information from LPL and ILI, hence it is likely that IOI and LPL are influenced by overlapping groups of genes.

The top QTL for CFS on BTA23 (28.86 Mb) was also associated to CLAFS. The shared QTL region by CFS and CLAFS could be expected, given the trait definition. The trait CLAFS is not entirely endocrine, but part endocrine and part classical, as it combines information from CFS and C-LA. In addition, there is a moderate genetic correlation (0.46) of C-LA and CFS, as reported by Tenghe et al. (2015). A number of fertility QTL in dairy cattle have been reported on BTA23 (Druet et al., 2008; Sahana et al., 2010; Höglund et al., 2014b). A QTL associated with veterinary treatment for reproductive disorders at 28.89 Mb was found in Danish and Swedish Holstein (Sahana et al., 2010). A QTL for CFS was found in Danish Holstein and validated in Danish Jersey at 28.09 Mb (Höglund et al., 2014b). They reported QTL 0.03 and 0.77 Mb away from the QTL identified in this study, suggesting it might be the same QTL. Furthermore, using sequence variants, (Höglund et al., 2015) identified a QTL region from 30 – 32 Mb for components of a fertility index (which included CFS as component trait) in Danish Jersey; the QTL for CFS was at 30.36 Mb, which may be identical to the one we observed at 30.35 Mb in this study.

3.4.2 Fine-mapping BTA2

The target region on BTA2 (1.87 – 15.30 Mb) was fine-mapped for C-LA, CLAFS and LPL. The most likely candidate genes in this region were *PDE1A* for C-LA, *SLC40A1* for CLAFS and *GULP1* for LPL. The *PDE1* belongs to the cyclic nucleotide phosphodiesterases (PDEs) that play a role in signal transduction by regulating intracellular cyclic nucleotide concentrations through hydrolysis of cyclic adenosine

monophosphate (cAMP). This gene complex is known to mediate oocyte maturation by regulating cAMP concentrations (Landim-Alvarenga and Maziero, 2014). A relative increase in the level of cAMP within the oocyte is essential for maintaining meiosis block, while a decrease in cAMP oocyte concentration allows the resumption of meiosis (Sela-Abramovich et al., 2006; Conti et al., 2012). The protein encoded by *SLC40A1* is a cell membrane protein that may be involved in iron export from duodenal epithelial cells, and defects in this gene result in hemochromatosis (Geer et al., 2010). In cattle, *SLC40A1* is suggested to be related to milk yield, but no fertility-related function is known (Fang et al., 2014). The gene *GULP1* encodes an adapter protein necessary for the engulfment of apoptotic cells by phagocytes, but no fertility-related function is known for this gene.

Another target region (31.87 – 41.96 Mb) on BTA2 was fine-mapped for LPL and IOI. The most significant variant in this region was located in close proximity of *NR4A2*, and *ACVR1* was another interesting candidate based on its function. The gene *NR4A2* is a member of the steroid receptor superfamily (Geer et al., 2010), and considered an important ovarian factor in regulation of female reproduction (Zhao et al., 2007). *ACVR1* (also known as *ALK2*) is a dimeric growth and differentiation factor which belongs to the transforming growth factor-beta (TGF-beta). The TGF-beta together with bone morphogenetic proteins (BMPs) form a signaling family that is necessary for follicle development and oocyte competence in mammals, and different studies have demonstrated the role of specific family members in theca cells, granulosa cells, cumulus cells and oocytes (Knight and Glister, 2006; Edson et al., 2009). In cattle, FSH and estradiol both play a role in regulating *ACVR1* expression during follicle development, and BMP signaling through *ACVR1* in granulosa cells is thought to play a role in follicle growth (Shimizu et al., 2006; Glister et al., 2010). Furthermore, *ACVR1* levels in granulosa cells were shown to increase with follicle size (Shimizu et al., 2006). Moreover, in equine, *ACVR1* is suggested to be involved with follicular fluid exosomes that regulate members of the TGFβ/BMP signaling pathway in granulosa cells, and possibly play a role in regulating follicle maturation (da Silveira et al., 2014).

3.4.3 Fine-mapping BTA3

The target region on BTA3 (85.68 – 95.66 Mb) was fine-mapped for C-LA, PLA and LA60. Similar to the GWAS, the most significant association (89,360,154 bp) was common to PLA and LA60, suggesting that indeed these traits are probably

influenced by the same genes. Moreover, detecting the same QTL for genetically correlated traits improves evidence for the QTL.

In the region from 87,500,158-95,662,104 bp on BTA3, *DAB1* and *C8B* were located in close proximity of the most significant variant in this region. The gene *DAB1* has been suggested to be involved in numerous potential physiological functions: cell adhesion processes, activation of neuron differentiation, stimulation of protein kinase activity, ventral spinal cord development, adult walking behavior, and cerebellum structural organization (Ashburner et al., 2000). In mice, the protein encoded by this gene is thought to be a signal transducer that interacts with protein kinase pathways to regulate neuronal positioning in the developing brain. The gene *C8B* is involved in encoding a membrane attack complex protein that mediates cell lysis. In humans, deficiency of this protein is associated with increased risk of meningococcal infection (Geer et al., 2010). Nevertheless, none of the candidate genes in this region had obvious fertility-related functions.

3.4.5 Limitations to mapping causal mutations

The LD in cattle is spread over a wide region due to low effective population size and strong selection. This LD breadth places limitations on mapping QTL to regions suitable for identifying causative mutations. However, using multiple-breed data may have advantage for mapping causal factors; for example, as shown in dogs (Karlsson et al., 2007). Also, studies on a more functional level may be required to be able to identify causative variants.

3.4.6 Effect size and application for breeding

Genomic predictions are now used routinely in selection of dairy cattle (Hayes et al., 2009; Goddard et al., 2010; Wiggans et al., 2011), however, an ongoing challenge is to improve the accuracy of these predictions, as the genetic gain that can be achieved is proportional to their accuracy. Genomic prediction from whole-genome sequence data is attractive, because the accuracy of genomic prediction is no longer bounded by the extent of linkage disequilibrium between DNA markers and causal mutations affecting the trait, given the causal mutations are in the data set. Although simulation studies show large gain in accuracy of genomic prediction due to use of sequence data, (Meuwissen and Goddard, 2010; Clark et al., 2011; Druet et al., 2014), this has not yet been realized in real data (Ober et al., 2012; Hayes et al., 2014; van Binsbergen et al., 2015). These simulation studies also show that the added value of using sequence data over SNP arrays to increase accuracy of genomic breeding is

dependent on the size, and distribution of QTL effects. The absolute effect sizes of QTL detected for endocrine fertility traits in this study were low to moderate, ranging from 0.12 to 4.45 (Table 3.4 and Table 3.5), suggesting that the effect of the causative mutations for endocrine fertility traits may be very small as observed for several quantitative traits (Kemper et al., 2011; Stahl et al., 2012). When QTL effects are small and MAF are low as observed for the fertility QTL in this study, large numbers of individuals are required to estimate these effects accurately, and even though the cost of genome resequencing has fallen dramatically, it is still too expensive to consider resequencing the tens of thousands of individuals that would be required to accurately estimate the likely small effects of mutations. An alternative strategy would be to sequence a relatively few individuals (key ancestors), and then impute genotypes for the variants discovered in the sequence into the whole population. Furthermore, simulation results (Druet et al., 2014) suggest that when the MAF of QTL is very low, genomic predictions from imputed sequence data can have up to 20% advantage in accuracy over use of SNP panels. In dairy cattle, such genetic architecture is most likely for a complex trait like fertility, which would benefit greatly from genomic prediction with whole-genome sequence. Another possibility to benefit from sequence data as suggested by Hayes et al., (2014) would be to preselect sequence variants based on annotation information, and include this biological information in genomic predictions. Yet another strategy could be to include the sequence variants from the causal regions on low density chips used in genomic prediction on which a large number of individuals can be genotyped. This would increase the power of QTL detection e.g., a mutation in the *SMC2* gene that causes embryonic loss was identified using such strategy in the 1000 bull genome project (Daetwyler et al., 2014). However, in terms of genomic prediction these low density chip do not provide more accuracy of prediction yet.

A sustainable breeding scheme also takes into account the correlated effects on other production and reproduction traits. In general, the genetic correlations among all the analyzed fertility traits are favorable, and therefore no serious negative effects on genetic progress due to selection are to be expected. However, the genetic correlations of milk yield with endocrine and classical fertility traits are generally unfavorable, hence special attention will be needed to optimize breeding schemes to account for these correlations. Although genetic correlation estimates of in-line endocrine fertility traits with milk yield seem lower (Tenghe et al., 2015) than genetic correlations of milk yield with classical fertility traits, more studies are needed to confirm these estimates. Hence a plausible solution to improve fertility using endocrine fertility traits would be optimized breeding schemes that jointly include

endocrine and classical fertility traits in a fertility index, with special attention given to the negative correlation of fertility with milk yield.

Nonetheless, it is not yet obvious on how to benefit directly from the endocrine traits, QTL, and sequence data, as there are still several challenges e.g., how to translate SNP effects to economic values and optimal integration in breeding decisions, where it is not clear how to deal with. Hence the findings in this study helped to learn more about the genetic architecture and underlying biology of fertility.

3.5 Conclusions

A total of 17 QTL regions for 6 endocrine fertility traits were identified on chromosomes 2, 3, 8, 12, 15, 17, 23, and 25 from a genome-wide association with 85,485 SNPs. Three QTL regions were identified for classical fertility traits. Overlapping QTL regions were found between endocrine traits on BTA 2, 3, and 17 and between endocrine and classical fertility traits on BTA23.

Fine-mapping of target regions for endocrine traits on BTA 2, and 3, using imputed sequence variants confirmed the QTL from the genome-wide association study, and identified several associated variants that can contribute to an index of markers for genetic improvement of fertility. Several potential candidate genes underlying endocrine fertility traits were also identified in the target regions. However, due to high LD, it was not possible to specify genes or polymorphisms as causal factors for any of the regions.

Acknowledgements

The authors acknowledge Lattec I/S (Denmark) for providing the in-line progesterone data, RobustMilk project "grant agreement KBBE-211708" for the manually recorded progesterone data, and the Cooperative Cattle Improvement Organization CRV BV (Arnhem, the Netherlands) for providing the pedigree data for this study. A. M. M. T benefited from a joint grant from the European Commission and the European Union Seventh Framework Programme (FP7/ 2007-2013) under the grant agreement n°311776, within the framework of the Erasmus-Mundus joint doctorate "EGS-ABG" (Paris, France). This study was performed as part of the Prolific project. The Prolific project is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement n°311776.

References

- Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, and J.T. Eppig. 2000. Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25:25–29.
- Berry, D.P., J.W.M. Bastiaansen, R.F. Veerkamp, S. Wijga, E. Wall, B. Berglund, and M.P.L. Calus. 2012. Genome-wide associations for fertility traits in Holstein-Friesian dairy cows using data from experimental research herds in four European countries. *Animal.* 6:1206–15.
- Berry, D.P., F. Buckley, P. Dillon, R.D. Evans, M. Rath, and R.F. Veerkamp. 2003. Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows. *J. Dairy Sci.* 86:2193–2204.
- van Binsbergen, R., M.P.L. Calus, M.C.A.M. Bink, F.A. van Eeuwijk, C. Schrooten, and R.F. Veerkamp. 2015. Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. *Genet. Sel. Evol.* 47:1–13.
- Bouwman, A.C., and R.F. Veerkamp. 2014. Consequences of splitting whole-genome sequencing effort over multiple breeds on imputation accuracy. *BMC Genet.* 15:105.
- Browning, B.L., and S.R. Browning. 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *Am. J. Hum. Genet.* 84:210–223.
- Bulman, D.C., and G.E. Lamming. 1978. Milk progesterone levels in relation to conception, repeat breeding and factors influencing acyclicity in dairy cows. *J. Reprod. Fertil.* 54:447–458.
- Burton, P.R., D.G. Clayton, L.R. Cardon, N. Craddock, P. Deloukas, A. Duncanson, D.P. Kwiatkowski, M.I. McCarthy, W.H. Ouwehand, and N.J. Samani. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 447:661–678.
- Campos, M.S., C.J. Wilcox, C.M. Becerril, and A. Diz. 1994. Genetic Parameters for Yield and Reproductive Traits of Holstein and Jersey Cattle in Florida. *J. Dairy Sci.* 77:867–873.
- Clark, S.A., J.M. Hickey, and J.H.J. Van der Werf. 2011. Different models of genetic variation and their effect on genomic evaluation. *Genet Sel Evol.* 43:10–1186.
- Conti, M., M. Hsieh, A.M. Zamah, and J.S. Oh. 2012. Novel signaling mechanisms in the ovary during oocyte maturation and ovulation. *Mol. Cell. Endocrinol.* 356:65–73.
- Cunningham, F., M.R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C.G. Girón, L. Gordon, T. Hourlier, S.E.

- Hunt, S.H. Janacek, N. Johnson, T. Juettemann, A.K. Kähäri, S. Keenan, F.J. Martin, T. Maurel, W. McLaren, D.N. Murphy, R. Nag, B. Overduin, A. Parker, M. Patricio, E. Perry, M. Pignatelli, H.S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S.P. Wilder, A. Zadissa, B.L. Aken, E. Birney, J. Harrow, R. Kinsella, M. Muffato, M. Ruffier, S.M.J. Searle, G. Spudich, S.J. Trevanion, A. Yates, D.R. Zerbino, and P. Flicek. 2015. Ensembl 2015. *Nucleic Acids Res.* 43:D662–D669.
- Daetwyler, H.D., A. Capitan, H. Pausch, P. Stothard, R. Van Binsbergen, R.F. Brøndum, X. Liao, A. Djari, S.C. Rodriguez, C. Grohs, D. Esquerré, O. Bouchez, M.-N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P.J. Bowman, D. Coote, A.J. Chamberlain, C. Anderson, C.P. VanTassell, I. Hulsege, M.E. Goddard, B. Guldbrandtsen, M.S. Lund, R.F. Veerkamp, D.A. Boichard, R. Fries, and B.J. Hayes. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* 46:858–865.
- Darwash, A.O., G.L. Ward, G.E. Lamming, and J.A. Woolliams. 1999. The effects of raising post-oestrus progesterone concentrations on luteal activity in post-partum dairy cows. *Anim. Sci.* 68:527–532.
- Druet, T., S. Fritz, M. Boussaha, S. Ben-Jemaa, F. Guillaume, D. Derbala, D. Zelenika, D. Lechner, C. Charon, D. Boichard, I.G. Gut, A. Eggen, and M. Gautier. 2008. Fine Mapping of Quantitative Trait Loci Affecting Female Fertility in Dairy Cattle on BTA03 Using a Dense Single-Nucleotide Polymorphism Map. *Genetics.* 178:2227–2235.
- Druet, T., I.M. Macleod, and B.J. Hayes. 2014. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity (Edinb).* 112:39–47.
- Edson, M.A., A.K. Nagaraja, and M.M. Matzuk. 2009. The mammalian ovary from genesis to revelation. *Endocr. Rev.* 30:624–712.
- Fang, M., W. Fu, D. Jiang, Q. Zhang, D. Sun, X. Ding, and J. Liu. 2014. A Multiple-SNP Approach for Genome-Wide Association Study of Milk Production Traits in Chinese Holstein Cattle.
- Friggens, N.C., M. Bjerring, C. Ridder, S. Højsgaard, and T. Larsen. 2008. Improved detection of reproductive status in dairy cows using milk progesterone measurements. *Reprod. Domest. Anim.* 43 Suppl 2:113–21.
- Fuchsberger, C., G.R. Abecasis, and D.A. Hinds. 2015. minimac2: faster genotype imputation. *Bioinformatics.* 31:782–784.
- Geer, L.Y., A. Marchler-Bauer, R.C. Geer, L. Han, J. He, S. He, C. Liu, W. Shi, and S.H. Bryant. 2010. The NCBI BioSystems database. *Nucleic Acids Res.* 38:D492–D496.

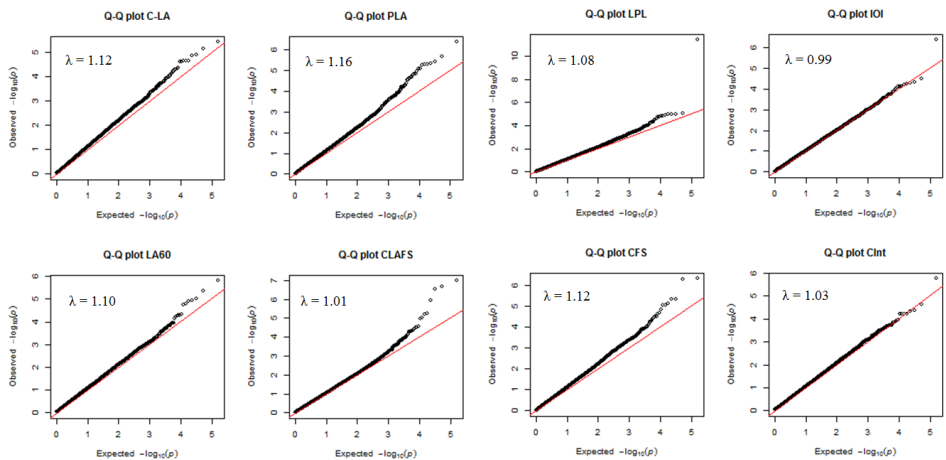
- Gilmour, A.R., B.J. Gogel, B.R. Cullis, S.J. Welham, and R. Thompson. 2014. ASReml User Guide Release 4.1 Structural Specification, VSN International Ltd, Hemel Hempstead, HP1 1ES, UK www.vsnl.co.uk.
- Glister, C., L. Satchell, and P.G. Knight. 2010. Changes in expression of bone morphogenetic proteins (BMPs), their receptors and inhibin co-receptor betaglycan during bovine antral follicle development: inhibin can antagonize the suppressive effect of BMPs on thecal androgen production. *Reproduction*. 140:699–712.
- Goddard, M.E., B.J. Hayes, and T.H.E. Meuwissen. 2010. Genomic selection in livestock populations. *Genet. Res. (Camb)*. 92:413–421.
- Hayes, B.J., P.J. Bowman, A.J. Chamberlain, and M.E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433–443.
- Hayes, B.J., A. MacLeod, H.D. Daetwyler, R.F. Veerkamp, C. Tassell, B. Gredler, T. Druet, A. Bagnato, J. Vilkkilä, and D.J. de Koning. 2014. Genomic prediction from whole genome sequence in Livestock: 1000 bull genomics project. In Proceedings of the 10th World congress of genetics applied to livestock production. 17–22 August 2014; Vancouver. 2014. https://asas.org/docs/default-source/wcgalp-proceedingsoral/183_paper_10441_manuscript_1644_0.pdf.
- Höglund, J.K., B. Guldbrandtsen, M.S. Lund, and G. Sahana. 2015. Identification of genomic regions associated with female fertility in Danish Jersey using whole genome sequence data. *BMC Genet.* 16:1–9.
- Höglund, J.K., G. Sahana, R. Brøndum, B. Guldbrandtsen, B. Buitenhuis, and M.S. Lund. 2014a. Fine mapping QTL for female fertility on BTA04 and BTA13 in dairy cattle using HD SNP and sequence data. *BMC Genomics*. 15:790.
- Höglund, J.K., G. Sahana, B. Guldbrandtsen, and M.S. Lund. 2014b. Validation of associations for female fertility traits in Nordic Holstein, Nordic Red and Jersey dairy cattle. *BMC Genet.* 15:8.
- Horan, B., J.F. Mee, P. O'Connor, M. Rath, and P. Dillon. 2005. The effect of strain of Holstein-Friesian cow and feeding system on postpartum ovarian function, animal production and conception rate to first service. *Theriogenology*. 63:950–971.
- Hu, Z.-L., C.A. Park, and J.M. Reecy. 2016. Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Res.* 44:D827–D833.
- Karlsson, E.K., I. Baranowska, C.M. Wade, N.H.C.S. Hillbertz, M.C. Zody, N. Anderson, T.M. Biagi, N. Patterson, G.R. Pielberg, and E.J. Kulbokas. 2007. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat. Genet.* 39:1321–1328.

- Kemper, K.E., D.L. Emery, S.C. Bishop, H. Oddy, B.J. Hayes, S. Dominik, J.M. Henshall, and M.E. Goddard. 2011. The distribution of SNP marker effects for faecal worm egg count in sheep, and the feasibility of using these markers to predict genetic merit for resistance to worm infections. *Genet. Res. (Camb)*. 93:203–219.
- Khatkar, M.S., P.C. Thomson, I. Tammen, and H.W. Raadsma. 2004. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol.* 36:163–90.
- Knight, P.G., and C. Glistler. 2006. TGF- β superfamily members and ovarian follicle development. *Reproduction*. 132:191–206.
- Landim-Alvarenga, F.C., and R.R.D. Maziero. 2014. Control of oocyte maturation. *Anim. Reprod.* 150–158.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25:2078–2079.
- Meuwissen, T., and M. Goddard. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics*. 185:623–631.
- Nyman, S., K. Johansson, D.J. de Koning, D.P. Berry, R.F. Veerkamp, E. Wall, and B. Berglund. 2014. Genetic analysis of atypical progesterone profiles in Holstein-Friesian cows from experimental research herds. *J. Dairy Sci.* 97:7230–7239.
- Ober, U., J.F. Ayroles, E.A. Stone, S. Richards, D. Zhu, R.A. Gibbs, C. Stricker, D. Gianola, M. Schlather, and T.F.C. Mackay. 2012. Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*. *PLoS Genet.* 8:e1002685.
- Petersson, K.-J., B. Berglund, E. Strandberg, H. Gustafsson, A.P.F. Flint, J.A. Woolliams, and M.D. Royal. 2007. Genetic analysis of postpartum measures of luteal activity in dairy cows. *J. Dairy Sci.* 90:427–434.
- Petersson, K.-J., E. Strandberg, H. Gustafsson, and B. Berglund. 2006. Environmental effects on progesterone profile measures of dairy cow fertility. *Anim. Reprod. Sci.* 91:201–14.
- Pollott, G.E., and M.P. Coffey. 2008. The effect of genetic merit and production system on dairy cow fertility, measured using progesterone profiles and on-farm recording. *J. Dairy Sci.* 91:3649–3660.
- Pruim, R.J., R.P. Welch, S. Sanna, T.M. Teslovich, P.S. Chines, T.P. Gliedt, M. Boehnke, G.R. Abecasis, and C.J. Willer. 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 26:2336–2337.
- Pryce, J.E., R.J. Esslemont, R. Thompson, R.F. Veerkamp, M.A. Kossaibati, and G. Simm. 1998. Estimation of genetic parameters using health, fertility and production data from a management recording system for dairy cattle. *Anim. Sci.* 66:577–584.

- Sahana, G., B. Guldbbrandtsen, C. Bendixen, and M.S. Lund. 2010. Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. *Anim. Genet.* 41:579–588.
- Sahana, G., B. Guldbbrandtsen, B. Thomsen, L.E. Holm, F. Panitz, R.F. Brøndum, C. Bendixen, and M.S. Lund. 2014. Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. *J. Dairy Sci.* 97:7258–7275.
- Sela-Abramovich, S., I. Edry, D. Galiani, N. Nevo, and N. Dekel. 2006. Disruption of gap junctional communication within the ovarian follicle induces oocyte maturation. *Endocrinology.* 147:2280–2286.
- Shimizu, T., B.C. Jayawardana, H. Nishimoto, E. Kaneko, M. Tetsuka, and A. Miyamoto. 2006. Involvement of the bone morphogenetic protein/receptor system during follicle development in the bovine ovary: Hormonal regulation of the expression of bone morphogenetic protein 7 (BMP-7) and its receptors (ActRII and ALK-2). *Mol. Cell. Endocrinol.* 249:78–83.
- da Silveira, J.C., E.M. Carnevale, Q.A. Winger, and G.J. Bouma. 2014. Regulation of ACVR1 and ID2 by cell-secreted exosomes during follicle maturation in the mare. *Reprod. Biol. Endocrinol.* 12:44.
- Stahl, E.A., D. Wegmann, G. Trynka, J. Gutierrez-Achury, R. Do, B.F. Voight, P. Kraft, R. Chen, H.J. Kallberg, and F.A.S. Kurreeman. 2012. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat. Genet.* 44:483–489.
- Su, G., R.F. Brøndum, P. Ma, B. Guldbbrandtsen, G.P. Aamand, and M.S. Lund. 2012. Comparison of genomic predictions using medium-density (~ 54,000) and high-density (~ 777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. *J. Dairy Sci.* 95:4657–4665.
- Tenghe, A.M.M., A.C. Bouwman, B. Berglund, E. Strandberg, J.Y. Blom, and R.F. Veerkamp. 2015. Estimating genetic parameters for fertility in dairy cows from in-line milk progesterone profiles. *J. Dairy Sci.* 98:5763–73.
- Teyssèdre, S., M.C. Dupuis, G. Guérin, L. Schibler, J.M. Denoix, J.M. Elsen, and A. Ricard. 2012. Genome-wide association studies for osteochondrosis in French Trotter horses. *J. Anim. Sci.* 90:45–53.
- VanRaden, P.M., D.J. Null, M. Sargolzaei, G.R. Wiggans, M.E. Tooker, J.B. Cole, T.S. Sonstegard, E.E. Connor, M. Winters, and J. van Kaam. 2013. Genomic imputation and evaluation using high-density Holstein genotypes. *J. Dairy Sci.* 96:668–678.
- VanRaden, P.M., A.H. Sanders, M.E. Tooker, R.H. Miller, H.D. Norman, M.T. Kuhn, and G.R. Wiggans. 2004. Development of a national genetic evaluation for cow fertility. *J. Dairy Sci.* 87:2285–2292.

- Veerkamp, R.F., J.K. Oldenbroek, H.J. Van Der Gaast, and J.H.J. Van Der Werf. 2000. Genetic Correlation Between Days Until Start of Luteal Activity and Milk Yield, Energy Balance, and Live Weights. *J. Dairy Sci.* 83:577–583.
- Wiggins, G.R., P.M. VanRaden, and T.A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, future. *J. Dairy Sci.* 94:3202–3211.
- Zhao, H., Z. Li, A.J. Cooney, and Z.-J. Lan. 2007. Orphan nuclear receptor function in the ovary. *Front. Biosci.* 12:3398–3405.
- Zimin, A. V, A.L. Delcher, L. Florea, D.R. Kelley, M.C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C.P. Van Tassell, and T.S. Sonstegard. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10:R42.

Supplementary material



Supplementary Figure 3. S1 Quantile-quantile plots (Q-Q plot) of $-\log_{10}$ of p -values resulting from a genome-wide scan for 6 endocrine and 2 classical fertility traits using 85,485 SNPs of 2,447 cows with 5,339 lactations in dairy cattle. Deviations from the slope line correspond to loci that deviate from the null hypothesis of no association; λ = genomic inflation factor

C-LA = Commencement of luteal activity; PLA = proportion of samples in luteal activity between 25 and 60 days in milk; LA60 = luteal activity between 25 and 60 days in milk, CLAFS = interval from commencement of luteal activity to first service; LPL = first luteal phase length; IOI = length of first inter-ovulatory interval; CFS = interval from calving to first service; CInt = calving interval.

4

Opportunities for genomic prediction for fertility using endocrine and classical fertility traits in dairy cattle

A. M. M. Tenghe,^{1,2,3} B. Berglund,³ E. Wall,⁴ R. F. Veerkamp,^{1,2} D.J. de Koning³

¹ Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands

² Animal Breeding and Genomics Centre, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands

³ Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, P. O. Box 7023, SE-750 07 Uppsala, Sweden

⁴ Animal and Veterinary Sciences, SRUC, Easter Bush Campus, EH25 9RG, Scotland, United Kingdom

Journal of Animal Science (2016) 94:3645-3654

Abstract

Endocrine fertility traits, defined from progesterone concentration levels in milk, have been suggested as alternative indicators for fertility in dairy cows because they are less biased by farm management decisions and more directly reflect a cow's reproductive physiology than classical traits derived from insemination and calving data. To determine the potential use of endocrine fertility traits in genomic evaluations, the improvement in accuracy from using endocrine fertility traits concurrent with classical traits in the genomic prediction of fertility was quantified. The impact of recording all traits on all training animals was also investigated. Endocrine and classical fertility records were available on 5,339 lactations from 2,447 Holstein cows in Ireland, the Netherlands, Sweden, and the United Kingdom. The endocrine traits were commencement of luteal activity (CLA) and proportion of samples in luteal activity (PLA); the classical trait was calving to first service (CFS). The interval from commencement of luteal activity to first service (CLAFS) which is a combination of an endocrine and classical trait was also investigated. The target (breeding goal) trait for fertility was CFS or CLAFS, whereas CLA and PLA served as predictor traits. Genomic estimated breeding values (GEBV) for fertility were derived using genomic BLUP in bivariate models with 85,485 single nucleotide polymorphisms. GEBV for the separate fertility traits were also computed, in univariate models. The accuracy of GEBV were evaluated by 5-fold cross-validation. The highest accuracy of GEBV was achieved using bivariate predictions, where both an endocrine and the classical fertility trait were used. Accuracy of GEBV for predicting adjusted phenotypes for CFS in the univariate model was 0.04, but when predicting CFS using a bivariate model with CLA, the accuracy increased to 0.14 when all training animals were phenotyped for CLA and (or not) for CFS. On phenotyping all training animals for both CLA and CFS, accuracy for CFS increased to 0.18; however, when validation animals were also phenotyped for CLA, there was no substantial increase in accuracy. When predicting CFS in bivariate analysis with PLA, accuracy ranged from 0.07 to 0.14. This first study on genomic predictions for fertility using endocrine traits suggests some improvement in the accuracy of prediction over using only the classical traits. Further studies with larger training populations may show greater improvements.

Key words: fertility, multi-trait genomic prediction, milk progesterone, dairy cattle, endocrine fertility traits, validation

4.1 Introduction

Endocrine fertility traits, based on progesterone (P4) levels in milk, have been suggested as indicators for fertility as they are less biased by farm management decisions and more directly reflect a cow's reproductive physiology than insemination and calving based indicators of fertility (Lamming and Darwash, 1998; Darwash et al., 1999). For example, the interval from calving to first ovulation as determined by P4 levels in milk could be used as indicator of a cow's ability to return to luteal activity after calving, instead of the indirect indicator calving to first service (CFS). However, the use of endocrine traits in routine genetic evaluation schemes has been constrained by the high cost associated with collecting P4 measures on sufficient number of animals. Nonetheless, in-line technology now exists to measure P4 during each milking (Friggens et al., 2008), enabling sampling of more animals at reasonable cost.

Because few herds currently have the in-line recording system, endocrine traits would be most useful in a genomic selection scheme, where cows from contract herds form part of the reference population. In comparison, classical fertility traits are routinely measured on a large scale and remain useful target traits in national genetic evaluations. Therefore, combining genomic predictions for endocrine and classical fertility traits may provide a more accurate prediction of dairy cow fertility. As a first step towards determining the optimum scenario for using endocrine fertility traits in genomic evaluations, the aim of this study was to quantify the improvement in accuracy achievable from using endocrine traits that describe the ability to return to normal cyclicity concurrent with the classical trait CFS in genomic prediction of fertility. This is only a first step, as CFS may not be the ideal breeding goal trait, and all traits related to fertility should be considered. Also evaluated were whether all traits should be measured on all animals in the training data and how well the genomic predictions translate across countries.

4.2 Materials and Methods

4.2.1 Phenotypes

The phenotype data consisted of in-line P4 records from 14 commercial herds in the Netherlands, and manually collected P4 records of 4 experimental herds from Wageningen UR Livestock Research, the Netherlands; Teagasc, Moorepark, Ireland; Dairy Research Centre, SRUC (Scotland's Rural College, previously SAC), Scotland;

and the Jälla herd of the Swedish University of Agricultural Science. In total, phenotypic data were available from 5,339 lactations on 2,447 Holstein-Friesian cows. A detailed description of the experimental treatments imposed on animals in the different experimental herds, procedures for milk sampling and P4 level measuring have been given in Veerkamp et al. (2000), Horan et al. (2005), Petersson et al. (2006), and Pollott and Coffey (2008). In brief, milk sampling for P4 measurement was undertaken twice a week at the experimental herds in Sweden and the Netherlands, and three times a week in Ireland and the United Kingdom. In the commercial herds, milk sampling, measuring and recording of P4 level was performed using the Herd Navigator™ (HN, DeLaval Intl, Tumba, Sweden). Sampling frequency on the commercial herds was based on a biological model (Friggens et al., 2008), but on average, was undertaken every 2 days.

For each lactation, the following endocrine fertility traits were defined using P4 records as described in Tenghe et al. (2015): 1) commencement of luteal activity (CLA) as the number of days from calving to start of luteal activity; 2) proportion of samples with luteal activity (PLA) as the number of P4 records with luteal activity, divided by total number of P4 records in the period from 25 to 60 days in milk; 3) commencement of luteal activity to first service (CLAFS) as the interval from start of luteal activity, to the day of first service. The trait CLAFS is a hybrid trait as it is composed of both an endocrine and a classical fertility trait. The classical fertility trait investigated was interval from calving to first service (CFS). The trait CFS was restricted between 30 to 250 days.

4.2.2 Genotype data and quality control

Genotyping of cows from the experimental herds was performed with the Illumina BovineSNP50 v1 BeadChip (Illumina Inc., San Diego, CA, USA) on 1,946 cows and contained 54,001 SNP (50k), whereas 1,907 cows from commercial herds were genotyped with the GeneSeek Genomic Profiler Bovine HD (GeneSeek, Lincoln, NE, USA), containing 76,883 SNP (80k). An additional 6 cows from a commercial herd were genotyped with a custom Illumina 6k array. The 50k and 80k SNP arrays had 25,815 SNP in common, the rest were imputed from one dataset to the other and vice versa using BEAGLE 3.3.2. (Browning and Browning, 2009), resulting in a total of 102,062 SNPs. The quality criteria applied before imputation were minimum call rate of 95% per animal and 97% per SNP, and minimum GC-score of 0.3. A total of 120 animals failed the quality criteria and were excluded. After imputation, SNPs were retained for analysis if they fulfilled all of the following quality criteria: 1) call rate >

4 Progesterone profiles in genomic prediction

95%, minor allele frequency (MAF) > 0.01, and no extreme deviation from Hardy-Weinberg Equilibrium ($p < 0.01$). Following all edits, 85,485 SNP for 3,739 cows remained.

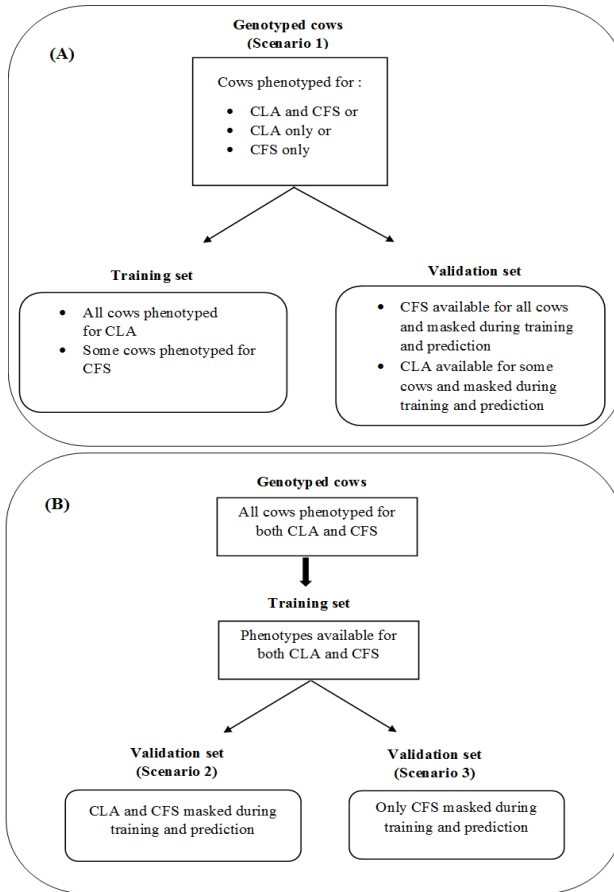


Figure 4.1 Flow chart showing the scenarios of genomic prediction for fertility investigated, using endocrine and classical fertility traits in dairy cows. The interval from calving to commencement of luteal activity (CLA) is used as an example for endocrine fertility traits, and the interval from calving to first service (CFS) as an example for classical fertility traits. The difference between scenarios is based on the fertility records in training and validation sets.

4.2.3 Statistical analysis

From the 3,739 cows, 2,447 cows with both phenotypes (5,339 lactations) and genotypes were used for analysis. Genomic estimated breeding values (GEBV) were predicted based on the genomic best linear unbiased prediction (GBLUP). The model was:

$$y_{ijkl} = \mu + hys_i + parity_j + b_1 \times ca_k(p_j) + b_2 \times ca_k^2(p_j) + g_l + pe_l + e_{ijkl}$$

where y_{ijkl} is the phenotype; μ is the intercept, hys_i is the fixed effect of herd-year- (calving) season combination i ($i=1-495$), with calving season defined as winter: December-February, spring: March-May, summer: June-August, autumn: September-November; $parity_j$ is the fixed effect of parity j (cows in parity 4 or above were grouped to a common class); $b_1 \times ca_k(p_j)$ is the linear regression on calving age k (in months) nested within parity p_j ; $b_2 \times ca_k^2(p_j)$ is the quadratic regression on age at calving, g_l is the genomic breeding value of the l^{th} individual, and pe_l is the permanent environmental effect of individual l to account for repeated measures (i.e., more than 1 lactation). The g were random, normally distributed effects with mean zero and variance $G\sigma_g^2$, where G is the genomic relationship matrix among all individuals, constructed according to VanRaden (2008), which was computed as $G = ZZ' / 2 \sum p_m q_m$, where Z is a matrix of centred genotypes, and p_m and q_m are the allelic frequencies of the m^{th} marker based on observed genotypes; pe were assumed to be normally distributed with mean zero and variance σ_{pe}^2 , and e_{ijkl} is the random error term, with residuals assumed to be normally distributed, with mean zero and variance σ_e^2 . ASReml 4.1 (Gilmour et al., 2014) was used for the GBLUP analysis.

A univariate model was used for single trait predictions, whereas bivariate models were used for multiple trait predictions, with the same fixed effects described above. Heritability estimates and pedigree-based estimated breeding values (EBV) were calculated from the univariate model, using the whole dataset.

4.2.4 Scenarios

Three scenarios for genomic predictions (Figure 4.1) were investigated using univariate and bivariate approaches. Scenario 1 aimed to evaluate the accuracy of predicting fertility for non-phenotyped candidate animals when all training animals were phenotyped for endocrine fertility traits, and (or not) for classical fertility traits. Scenario 2 aimed to evaluate whether the accuracy of prediction for non-

phenotyped candidate animals could be improved when all training animals were phenotyped for both endocrine and classical fertility traits. Scenario 3, investigated whether also phenotyping validation animals for endocrine fertility traits improved accuracy.

4.2.5 Accuracy of prediction

A 5-fold cross-validation was used to quantify the accuracy of prediction for the different scenarios. For each analysis (i.e., univariate or bivariate), individuals were randomly divided in five subsets and 5 training-validation groups were then formed from the subsets; each subset was used as validation set once. The accuracy was averaged across the 5 training-validation groups. To estimate accuracy of prediction, the target phenotype of the validation set in (i.e., univariate or bivariate) each analysis was excluded from the training dataset, and the GEBV were predicted from the training set. For the validation set, the GEBV were then correlated with a vector of phenotypes (Phen), corrected for the fixed effects previously described. In case a cow had multiple lactations, only the first available lactation was used in the validation set. However, the accuracy of GEBV may be best described as the correlation between GEBV and true breeding values. The accuracy of true breeding values (realized accuracy) was approximated as $r(\text{GEBV, Phen})/h$, where h was the square root of the estimated heritability of the phenotype from the univariate model. Prediction bias was then calculated by regressing the validation variables (Phen) on the predictor variables (GEBV). Prediction bias was also assessed by regressing the pedigree-based EBV on GEBV.

We also performed single trait predictions across country for scenario 1, with the univariate model. The data for this study originated from 4 countries (Ireland, the Netherlands, Sweden, and the United Kingdom). The number of cows and lactations available from each country are in Table 4.1. Because the Netherlands had the largest proportion of the data (88 %), it was not used as a validation set. The validation sets were the data from Ireland, Sweden or the United Kingdom, and for each validation set, the training set was composed of data from the Netherlands and the other two countries.

Table 4.1 Number of cows and lactation records for Holstein-Friesian cows from Ireland (IRE), the Netherlands (NL), Sweden (SWE), and the United Kingdom (UK), used in analysis

Trait ¹	No.	IRE	NL	SWE	UK	Total
CLA (d)	Records	63	2,946	473	42	3,524
	Cows	35	2,037	217	30	2,319
PLA (%)	Records	63	3,001	491	42	3,597
	Cows	35	2,074	220	30	2,359
CLAFS (d)	Records	61	1,542	380	32	2,015
	Cows	33	1,166	187	24	1,410
CFS (d)	Records	100	3,064	395	75	3,634
	Cows	35	1,407	189	30	1,661

¹CLA = Commencement of luteal activity, PLA = proportion of samples in luteal activity between 25 and 60 days in milk, CLAFS = interval from commencement of luteal activity to first service, CFS = interval from calving to first service

4.3 Results

The descriptive statistics for each trait in each country are in Table 4.2, together with the genomic heritability estimated from the whole dataset. The genomic heritability estimates of the fertility traits ranged from 0.04 to 0.12 (Table 4.2), whereas pedigree-based heritability estimates (\pm SE) were 0.14 (\pm 0.04) for CLA, 0.15 (\pm 0.04) for PLA, 0.09 (\pm 0.05) for CLAFS, and 0.10 (\pm 0.04) for CFS. The genomic and pedigree-based heritabilities were generally similar.

4.3.1 Accuracy

Accuracy of calculated from univariate and bivariate analysis are presented in Tables 4.3 and 4.4. The accuracy of prediction for endocrine fertility traits was in general greater than for the classical fertility trait in univariate predictions. In the univariate predictions, accuracy of prediction ranged from 0.13 to 0.15 for CLA across scenarios, and was 0.14 for PLA. The accuracy of prediction for CFS and CLAFS was 0.04 across scenarios.

There was an increase between 0.10 to 0.14 in accuracy of predicting CFS in bivariate analysis with endocrine fertility traits, compared to accuracy in the univariate predictions (Tables 4.3 and 4). The accuracy of predicting CFS in univariate predictions was 0.04, but when predicting CFS using the bivariate analysis with CLA as predictor trait, accuracy increased to 0.14 in scenario 1 (Table 4.3) where all training animals were phenotyped for CLA, and (or not) for CFS, and up to 0.18 in

4 Progesterone profiles in genomic prediction

scenario 2 (Table 4.4) where all training animals were phenotyped for both CLA and CFS. When all validation animals were also phenotyped for endocrine traits (scenario 3), there was no substantial increase in accuracy; that is accuracy of CFS was 0.15 (Table 4.4). For prediction of CFS using the bivariate analysis with PLA as predictor trait, the accuracy of predicting CFS was greatest in scenario 1 (0.14), decreased to 0.07 in scenario 2, and was 0.12 in scenario 3.

For the hybrid trait CLAFS, an increase in accuracy of prediction was observed only in scenario 3 where all candidate animals were phenotyped for endocrine fertility traits. That is, when predicting CLAFS in a bivariate analysis with CLA, the accuracy of predicting CLAFS increased to 0.18 and up to 0.21 in the bivariate analysis with PLA, but for the bivariate predictions of CLAFS in scenario 1 and 2, accuracy ranged from 0.02 to 0.07.

Accuracy of prediction across countries was low to moderate for the endocrine and classical fertility traits. The accuracy of prediction across countries ranged from 0.12 to 0.34 for CLA, 0.14 to 0.24 for PLA and 0.06 to 0.38 for CFS (Table 4.5). The accuracy of prediction for CLAFS was poor across all countries except for Sweden.

Table 4.2 Genomic heritability (h^2) and repeatability (t) estimates with their standard errors from the whole dataset and number of records, mean, and phenotypic standard deviation for each trait within each country

Trait ¹	h^2	SE	t	SE	Ireland		Netherlands		Sweden		United Kingdom	
					Mean	SD	Mean	SD	Mean	SD	Mean	SD
CLA	0.10	0.02	0.33	0.03	38.6	14.8	38.3	17.4	40.0	21.5	38.2	21.2
PLA	0.12	0.03	0.31	0.03	0.4	0.2	0.6	0.3	0.5	0.3	0.5	0.3
CLAFS	0.04	0.03	0.28	0.04	36.7	20.5	48.5	27.6	66.0	39.2	35.5	20.8
CFS	0.11	0.03	0.15	0.02	74.1	19.6	89.8	33.0	105.7	39.1	77.4	23.8

¹CLA = Commencement of luteal activity, PLA = proportion of samples in luteal activity between 25 and 60 days in milk, CLAFS = interval from commencement of luteal activity to first service, CFS = interval from calving to first service

Table 4.3 Accuracy of genomic estimated breeding values (GEBV) from GBLUP, using univariate and bivariate models (Scenario 1)

Trait ¹	Trait	Training ²		Validation ³		Realized ⁴ accuracy (SD)	Slope1	Slope2
		Average no. of lactations	Average no. of cows	Average no. of cows	Accuracy (SD)			
Training	Validation							
CLA	CLA	2,819	1,855	464	0.15 (0.05)	0.46 (0.16)	1.19	0.79
PLA	PLA	2,878	1,887	472	0.14 (0.02)	0.42 (0.07)	0.97	0.68
CLAFS	CLAFS	1,612	1,128	282	0.04 (0.07)	0.19 (0.37)	1.72	1.69
CFS	CFS	2,907	1,329	332	0.04 (0.07)	0.13 (0.21)	0.37	0.47
CLA, CFS	CFS	2,819	1,855	258	0.14 (0.07)	0.41 (0.21)	1.39	0.42
CLA, CLAFS	CLAFS	2,819	1,855	258	0.03 (0.09)	0.13 (0.44)	0.98	1.07
CLA, PLA	PLA	2,819	1,855	464	0.14 (0.05)	0.41 (0.14)	0.95	0.67
PLA, CFS	CFS	2,878	1,887	284	0.10 (0.03)	0.31 (0.09)	1.12	0.53
PLA, CLAFS	CLAFS	2,878	1,887	282	0.07 (0.03)	0.34 (0.16)	1.29	0.96

¹CLA = Commencement of luteal activity, PLA = proportion of samples in luteal activity between 25 and 60 days in milk, CLAFS = interval from commencement of luteal activity to first service, CFS = interval from calving to first service. Slope1 = regression coefficient of phenotypes on GEBV, Slope2 = regression coefficient of pedigree-based EBV on GEBV. Accuracy and average of training and validation sets were obtained by 5 random training-validation sets

²Standard deviation of average number of lactations in the training sets ranged from 5.89 to 26.24; Standard deviation of average number of cows in the training sets ranged from 0.00 to 0.45; ³Standard deviation of average number of cows in the validation sets ranged from 0.00 to 13.77, each cow in the validation set had one lactation

⁴Realized accuracy = Accuracy divided by square root of heritability of the validation trait

4.3.2 Bias

In the univariate prediction of the endocrine fertility traits, the slopes ranged from 0.97 to 1.20, indicating that the predictions were not severely biased (Tables 4.3 and 4); the confidence intervals for the slopes were wide, but the slopes were not different from 1. For univariate predictions of CFS and CLAFS, the variance of the GEBV were most biased in scenario 1, the slopes were 0.37 for CFS and 1.72 for CLAFS, but somewhat reduced in scenario 2 and 3, that is, the slope increased to 0.70 for CFS and reduced to 1.34 for CLAFS. In the cases with high prediction bias in bivariate predictions, we observed that the genomic heritability was very low. For example, in scenario 2 (Table 4.4), when CFS was predicted in bivariate analysis with CLA, the GEBV across the cross-validation sets were on a much smaller scale compared to pedigree-based EBV for CFS. That is, the pedigree-based EBV ranged from -15 to 20 whereas the GEBV ranged from -2 to 2 (Supplementary Figure 4.1), implying that the genomic heritability of CFS was under estimated in the cross-validation sets due to small sample sizes. Also, the GEBV from bivariate predictions

were in general on a much smaller scale than GEBV from univariate predictions (e.g., Supplementary Figure 4.2), implying that genomic heritability of the target trait was underestimated in the bivariate predictions.

There was generally a bias in estimated variance of GEBV across countries, with the most bias for CLAFS (Table 4.5). This bias could be due to the few number of animals in training and validation sets. For example, GEBV for CLAFS for the United Kingdom predicted with the whole dataset (scenario 1) ranged from -1 to 3, and those from across country predictions ranged from -1 to 4 (Supplementary Figure 4.3), implying that prediction of CLAFS was consistent in both cases, but biased due to very few animals ($n = 24$).

4.4 Discussion

The main aim of this study was to quantify the improvement in accuracy from using endocrine fertility traits concurrent with classical traits in genomic prediction of fertility. Univariate and bivariate predictions were used to obtain GEBV for two endocrine fertility traits (CLA and PLA), a classical trait (CFS), and a hybrid trait (CLAFS) in three scenarios of genomic prediction. With CFS as target (breeding goal) trait, for fertility, our results show that accuracy of GEBV for CFS was increased in bivariate analysis where endocrine and classical fertility traits were used, and there was a better predictive ability of CFS in bivariate analysis with CLA than with PLA. There was no substantial increase in accuracy of GEBV for CFS when both training and validation animals were phenotyped for endocrine fertility traits. Across country predictions were also evaluated in univariate predictions, and some predictive ability was observed for all traits across countries, except for CLAFS.

The mean phenotypic performance in the present study was similar to previous reports from Ireland (Berry et al., 2012), the Netherlands (Veerkamp et al., 2000), Sweden (Petersson et al., 2006), and the United Kingdom (Pollott and Coffey, 2008), all of which included some of the data used in this study. The genomic heritability estimates of the endocrine traits (CLA and PLA) in this study were slightly lower than pedigree-based estimates reported elsewhere (Veerkamp et al., 2000; Petersson et al., 2007; Tenghe et al., 2016), whereas heritability of the classical trait CFS was similar to those of previous studies (Pryce et al., 1998; Berry et al., 2003).

4.4.1 Accuracy

This is the first study to evaluate the usefulness of endocrine measures on genomic predictions for fertility. The study showed that the accuracy of predicting genomic breeding values for fertility can be increased in multi-trait predictions when endocrine and classical fertility traits are used. In the bivariate predictions, the target trait was CFS or CLA_{FS}, whereas the predictor trait was CLA or PLA. The gain in accuracy of GEBV of a target trait is expected to be higher when the correlation with the predictor trait is stronger. In a previous study, using a subset of the dataset used in this study, Tenghe et al. (2015) estimated genetic correlations of 0.37 between CFS and CLA, and 0.31 between CFS and PLA. Hence the gain in accuracy of GEBV for CFS was expected to be slightly higher when using CLA as the predictor trait than when using PLA, as confirmed by the results in this study. Similarly, the genetic correlation of CLA_{FS} was -0.56 with CLA and 0.65 with PLA (Tenghe et al., 2015), hence the gain in accuracy of GEBV for CLA_{FS} was expected to be higher when using PLA as predictor trait rather than CLA. With the exception of scenario 2, accuracies of predicting CLA_{FS} with CLA or PLA as predictor traits were as expected, that is the highest accuracy for GEBV of CLA_{FS} were obtained when PLA was used as the predictor trait. The superiority of multi-trait genomic analysis over single-trait has been shown in stochastic (Calus and Veerkamp, 2011; Jia and Jannink, 2012) and deterministic (Calus et al., 2013) simulations, as well as in real data (Aguilar et al., 2011; De Haas et al., 2012; Pszczola et al., 2013).

4 Progesterone profiles in genomic prediction

Table 4.4 Accuracy of genomic estimated breeding values (GEBV) from GBLUP, using univariate and bivariate models (Scenarios 2 and 3)

Trait ¹ Training	Trait Validation	Training ²		Validation ³		Realized ⁴ accuracy (SD)	Slope1	Slope2
		Average no. of lactations	Average no. of cows	Average no. of cows	Accuracy (SD)			
Scenario 2								
CLA	CLA	1,822	1,199	300	0.13 (0.08)	0.40 (0.26)	1.14	0.72
PLA	PLA	1,819	1,202	301	0.14 (0.04)	0.42 (0.12)	1.20	0.67
CLAFS	CLAFS	1,612	1,128	282	0.04 (0.04)	0.22 (0.21)	1.34	1.34
CFS	CFS	1,621	1,131	283	0.04 (0.05)	0.11 (0.15)	0.70	0.77
CLA, CFS	CFS	1,614	1,128	282	0.18 (0.04)	0.55 (0.13)	12.27	2.20
CLA, CLAFS	CLAFS	1,612	1,128	282	0.04 (0.03)	0.19 (0.17)	1.17	1.21
CLA, PLA	PLA	2,808	1,853	464	0.15 (0.04)	0.43 (0.13)	1.02	0.71
PLA, CFS	CFS	1,614	1,129	282	0.07 (0.06)	0.20 (0.17)	0.65	0.40
PLA, CLAFS	CLAFS	1,605	1,126	282	0.02 (0.02)	0.10 (0.09)	0.39	0.75
Scenario 3								
CLA, CFS	CFS	1,614	1,128	282	0.15 (0.05)	0.44 (0.16)	1.38	0.42
CLA, CLAFS	CLAFS	1,612	1,128	282	0.18 (0.07)	0.90 (0.37)	7.91	2.60
CLA, PLA	PLA	2,808	1,853	464	0.49 (0.03)	1.42 (0.07)	2.74	0.82
PLA, CFS	CFS	1,614	1,129	282	0.12 (0.07)	0.35 (0.20)	0.98	0.36
PLA, CLAFS	CLAFS	1,605	1,126	282	0.21 (0.09)	1.05 (0.47)	3.62	0.90

¹CLA = Commencement of luteal activity, PLA = proportion of samples in luteal activity between 25 and 60 days in milk, CLAFS = interval from commencement of luteal activity to first service, CFS = interval from calving to first service. Slope1 = regression coefficient of phenotypes on GEBV, Slope2 = regression coefficient of pedigree-based EBV on GEBV. Accuracy and average of training and validation sets were obtained by 5 random training-validation sets

²Standard deviation of average number of lactations in the training sets ranged from 5.89 to 26.24; Standard deviation of average number of cows in the training sets ranged from 0.00 to 0.45; ³Standard deviation of average number of cows in the validation sets ranged from 0.00 to 13.77, each cow in the validation set had one lactation; ⁴Realized accuracy = Accuracy divided by square root of heritability of the validation trait

The results from across country predictions were inconclusive when predicting GEBV for animals in Sweden with data from the Netherlands, Ireland and the United Kingdom. Predictive ability was not confirmed for Ireland or the United Kingdom, when using training data from the other three countries. The accuracy of GEBV for Sweden was highest for endocrine fertility traits and lowest for the classical trait. The accuracies in across country predictions were in general higher than accuracies observed in the univariate predictions that used the whole dataset. One reason might be the fact that the size of the training set with phenotypes is larger when combining data from multiple countries to predict another country.

Future studies with larger training populations within countries may show greater improvements in accuracy (Daetwyler et al., 2008; Calus et al., 2013). Also, it needs to be studied how this relates to using multiple traits across multiple countries (e.g., Wientjes et al., 2015)

4.4.2 Bias

The slopes of regression of phenotypes on GEBV are a measure of the bias in terms of the estimated variance of the GEBV. Slopes greater than one indicate underestimation of the variance of the GEBV whereas slopes smaller than one indicate overestimation. When CFS and CLA FS were predicted in univariate analyses, there was increased bias in variance of GEBV, but this bias was reduced in the bivariate analysis in scenario 1, where CLA or PLA were used as predictor traits in the training population, suggesting that inclusion of a predictor trait may decrease the bias. However, also phenotyping the validation animals (scenario 3) for CLA or PLA in the bivariate approach did not improve the bias, which is in contrast to the results of Pszczola et al., (2013), where in a similar study they found that recording predictor traits for both training and validation animals significantly increased accuracy of GEBV for the target trait and removed the bias observed when only training animals had phenotypes. The difference in results could be because the traits investigated by Pszczola et al. (2013) had higher heritability than the traits in this study. In Pszczola et al. (2013), heritability for the target trait ranged from 0.60 to 0.44 (when estimated using the pedigree or genomic relationship matrix), whereas heritability for the predictor traits ranged from 0.36 to 0.48.

4 Progesterone profiles in genomic prediction

Table 4.5 Accuracy of across country predictions for fertility from GBLUP, using endocrine and classical fertility records from Ireland (IRE), the Netherlands (NL), Sweden (SWE), and the United Kingdom (UK)

Trait ¹	Validation set	No. Training		No. Validation	Accuracy	Realized ²		
		Lactations	Cows	Cows		accuracy	Slope1	Slope2
CLA	IRE	3,461	2,284	35	0.12	0.37	1.02	0.29
CLA	SWE	3,051	2,102	217	0.16	0.52	2.68	0.42
CLA	UK	3,482	2,289	30	0.34	1.07	2.72	1.19
PLA	IRE	3,534	2,324	35	0.15	0.43	1.01	0.38
PLA	SWE	3,106	2,139	220	0.24	0.68	2.32	0.46
PLA	UK	3,555	2,329	30	0.14	0.41	0.99	0.64
CLAFS	IRE	1,954	1,377	33	0.00	0.01	0.05	0.43
CLAFS	SWE	1,635	1,223	187	0.12	0.61	5.84	0.65
CLAFS	UK	1,983	1,386	24	-0.11	-0.55	-1.14	-0.12
CFS	IRE	3,534	1,626	35	0.12	0.36	0.63	0.50
CFS	SWE	3,239	1,472	189	0.06	0.18	0.88	0.25
CFS	UK	3,559	1,631	30	0.38	1.15	1.80	0.20

¹CLA = Commencement of luteal activity, PLA = proportion of samples in luteal activity between 25 and 60 days in milk, CLAFS = interval from commencement of luteal activity to first service, CFS = interval from calving to first service. Slope1 = regression coefficient of phenotypes on GEBV, Slope2 = regression coefficient of pedigree-based EBV on GEBV.

²Realized accuracy = Accuracy divided by square root of heritability of the trait

4.4.3 Breeding goal for fertility and optimal selection strategy

In this study, we assumed the breeding goal trait for fertility to be the target trait (i.e., CFS or CLAFS), in the bivariate predictions. However, in practice, the breeding goal for fertility is usually a fertility index which is composed of several individual fertility traits. The current fertility index in most countries is composed mainly of classical fertility traits, with CFS often used as an indirect indicator of a cow's ability to return to luteal activity after calving. This study has shown that it is possible to increase the accuracy of predicting CFS when CLA is used as a predictor trait in multi-trait genomic prediction. For example, when predicting CFS in univariate predictions, the realized accuracy of GEBV was 0.13, which improved to between 0.41 to 0.55, when CLA was included in the training or validation sets of predictions for CFS (Tables 4.3 and 4.4). Also, the predictive ability for CLA in the univariate predictions was better than for CFS. Hence it might be advantageous to include CLA in the fertility index, as a direct indicator of a cow's ability to resume luteal activity after calving. Furthermore, there was no substantial increase in accuracy of GEBV for CFS when

both training and validation animals were phenotyped for endocrine fertility traits, suggesting that the optimal design of the training population for genomic prediction of fertility using endocrine and classical fertility traits would be scenario 1, in which all training animals were phenotyped for endocrine traits, and all available classical phenotypes for the training population were used.

Ideally, the breeding goal trait for fertility would be cows that resume luteal activity early after calving, start showing heat so they can be inseminated at the optimal time point, are pregnant after first insemination and calve successfully. However, not all of these components are captured by endocrine fertility traits at present, hence the feasible approach at the moment for using endocrine fertility traits in genetic evaluation schemes would be to combine current endocrine and classical fertility traits in a fertility index. Therefore, evaluating the predictive ability of the separate endocrine fertility traits is a first step towards incorporating endocrine fertility traits in genetic evaluations.

The greatest advantage of combining genomic data from multiple countries is that the size of the training set with phenotypes is increased. Although the achieved accuracy of GEBV in this study were lower than those currently reported for other dairy cattle traits (e.g., milk yield and type traits Lund et al., 2011), having GEBV from endocrine fertility traits is an important first step towards incorporating them in genomic selection for fertility. To further increase the accuracy of GEBV, one strategy can be to increase the training set for the endocrine fertility traits by expanding international collaboration. This will be especially useful for across country predictions, as international collaboration would open the way for the introduction of international multi-trait genomic predictions, and more accurate GEBVs. International collaboration will be important for traits like endocrine fertility traits that are difficult or expensive to measure on a large scale in individual countries, as this will make it possible to achieve a sufficiently large training population. Although not investigated in this study, another approach could be to use more endocrine (predictor) traits in a multi-trait analyses. In a similar study on investigating the accuracy of genomic prediction for a scarcely recorded trait using predictor traits, Pszczola et al. (2013) showed that three-trait analyses were more accurate than two-trait analyses, hence, adding more endocrine traits as predictor traits could further increase the accuracy. However, for practical application, the challenge of defining the breeding goal for fertility using endocrine and classical fertility traits remains and should be addressed.

Another important point to be addressed is how to combine the endocrine fertility traits with the large datasets of bull fertility EBV from national genomic evaluations. For national genomic evaluations, classical fertility traits are usually available for large number of daughters to predict GEBV for genotyped test bulls. One approach of incorporating endocrine traits in national genomic evaluations will be to combine the cow endocrine traits to the bull training population to improve accuracy of predicting test bulls (or cows). However, because the bull training populations with classical fertility traits for national genomic evaluations are usually large, a large number of cows with endocrine fertility traits will be required to obtain a substantial increase in accuracy of prediction. Nevertheless, with the increased use of automated systems coupled with new technologies like the Herd Navigator on farms, we can expect the number of cow records to increase. Furthermore, evaluating the required number of cows using simulations will be an important first step

4.5 Conclusions

This study showed that the accuracy of predicting genomic breeding values for fertility can be increased by using endocrine and classical fertility traits in multi-trait genomic prediction. The greatest accuracy of GEBV for CFS was achieved in the bivariate predictions, in which endocrine and classical fertility traits were used, and there was a better predictive ability of CFS in bivariate analysis with CLA than with PLA. There was no substantial increase in accuracy of GEBV for CFS when both training and validation animals were phenotyped for endocrine fertility traits. Across country predictions were also evaluated in univariate predictions, and some predictive ability was observed for all traits across countries, except for CLAFS. This first study on genomic predictions for fertility using endocrine traits suggests some improvement over using only the classical traits. Further studies with larger training populations may show bigger improvements.

Acknowledgements

The authors acknowledge Lattec I/S (Denmark) for providing the in-line progesterone data, RobustMilk project “grant agreement KBBE-211708” for the manually recorded progesterone data from Ireland, the Netherlands and Sweden, and the Cooperative Cattle Improvement Organization CRV BV (Arnhem, the Netherlands) for providing the pedigree data for this study. Data from Scotland was collected under a grant from the Scottish Executive Environment and Rural

Affairs Department. Donagh Berry; Animal and Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland is acknowledged for contributing data and comments to the manuscript. A. M. M. T benefited from a joint grant from the European Commission and the European Union under the grant agreement n°311776, within the framework of Erasmus-Mundus joint doctorate "EGS-ABG" (Paris, France). This study was performed as part of the Prolific project which is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement n°311776.

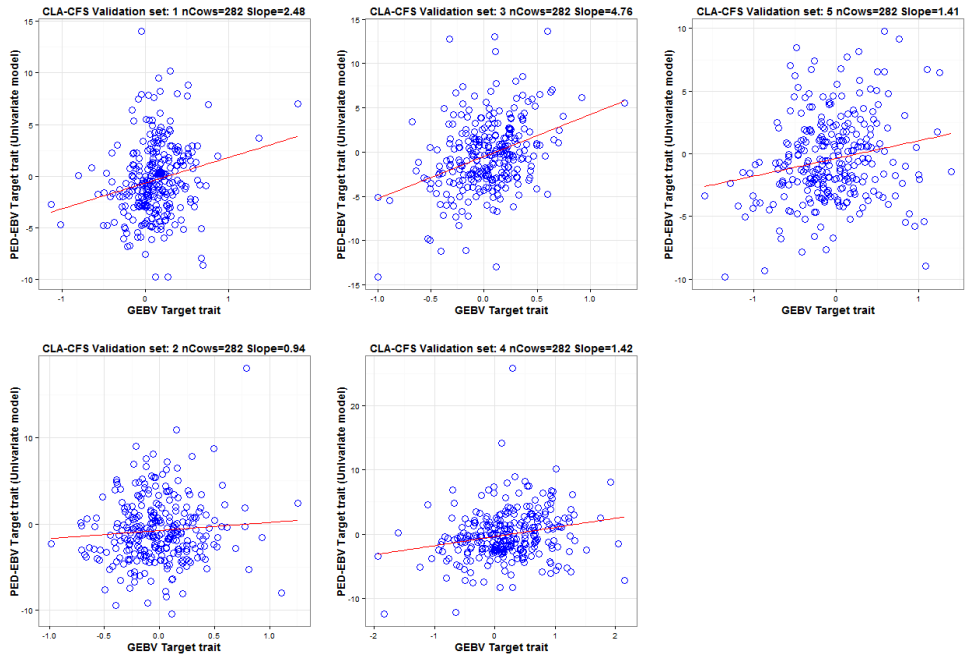
References

- Aguilar, I., I. Misztal, S. Tsuruta, G.R. Wiggans, and T.J. Lawlor. 2011. Multiple trait genomic evaluation of conception rate in Holsteins. *J. Dairy Sci.* 94:2621–2624.
- Berry, D.P., J.W.M. Bastiaansen, R.F. Veerkamp, S. Wijga, E. Wall, B. Berglund, and M.P.L. Calus. 2012. Genome-wide associations for fertility traits in Holstein-Friesian dairy cows using data from experimental research herds in four European countries. *Animal*. 6:1206–15.
- Berry, D.P., F. Buckley, P. Dillon, R.D. Evans, M. Rath, and R.F. Veerkamp. 2003. Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows. *J. Dairy Sci.* 86:2193–2204.
- Browning, B.L., and S.R. Browning. 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *Am. J. Hum. Genet.* 84:210–223.
- Calus, M.P.L., Y. De Haas, M. Pszczola, and R.F. Veerkamp. 2013. Predicted accuracy of and response to genomic selection for new traits in dairy cattle. *Animal*. 7:183–191.
- Calus, M.P.L., and R.F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. *Genet Sel Evol.* 43:1–14.
- Daetwyler, H.D., B. Villanueva, and J.A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*. 3:e3395.
- Darwash, A.O., G.L. Ward, G.E. Lamming, and J.A. Woolliams. 1999. The effects of raising post-oestrus progesterone concentrations on luteal activity in post-partum dairy cows. *Anim. Sci.* 68:527–532.
- Friggens, N.C., M. Bjerring, C. Ridder, S. Højsgaard, and T. Larsen. 2008. Improved detection of reproductive status in dairy cows using milk progesterone measurements. *Reprod. Domest. Anim.* 43 Suppl 2:113–21.
- Gilmour, A.R., B.J. Gogel, B.R. Cullis, S.J. Welham, and R. Thompson. 2014. ASReml

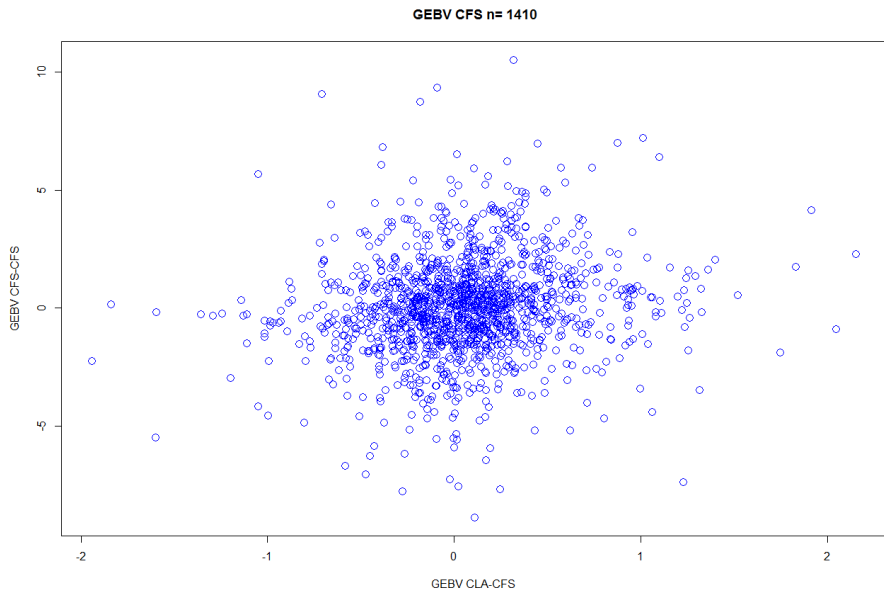
- User Guide Release 4.1 Structural Specification, VSN International Ltd, Hemel Hempstead, HP1 1ES, UK www.vsn.co.uk.
- Goddard, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*. 136:245–257.
- De Haas, Y., M.P.L. Calus, R.F. Veerkamp, E. Wall, M.P. Coffey, H.D. Daetwyler, B.J. Hayes, and J.E. Pryce. 2012. Improved accuracy of genomic prediction for dry matter intake of dairy cattle from combined European and Australian data sets. *J. Dairy Sci.* 95:6103–6112.
- Hayes, B.J., H.D. Daetwyler, P.J. Bowman, G. Moser, B. Tier, R. Crump, M. Khatkar, H.W. Raadsma, and M.E. Goddard. 2009. Accuracy of genomic selection: comparing theory and results. *In Proc Assoc Advmt Anim Breed Genet.* 34–37.
- Horan, B., J.F. Mee, P. O'Connor, M. Rath, and P. Dillon. 2005. The effect of strain of Holstein-Friesian cow and feeding system on postpartum ovarian function, animal production and conception rate to first service. *Theriogenology*. 63:950–971.
- Jia, Y., and J.-L. Jannink. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*. 192:1513–1522.
- Lamming, G.E., and a O. Darwash. 1998. The use of milk progesterone profiles to characterise components of subfertility in milked dairy cows. *Anim. Reprod. Sci.* 52:175–90.
- Lund, M.S., A.P.W. De Roos, A.G. De Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, and R. Reents. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet Sel Evol.* 43:43.
- Petersson, K.-J., B. Berglund, E. Strandberg, H. Gustafsson, A.P.F. Flint, J.A. Woolliams, and M.D. Royal. 2007. Genetic analysis of postpartum measures of luteal activity in dairy cows. *J. Dairy Sci.* 90:427–434.
- Petersson, K.-J., E. Strandberg, H. Gustafsson, and B. Berglund. 2006. Environmental effects on progesterone profile measures of dairy cow fertility. *Anim. Reprod. Sci.* 91:201–14.
- Pollott, G.E., and M.P. Coffey. 2008. The effect of genetic merit and production system on dairy cow fertility, measured using progesterone profiles and on-farm recording. *J. Dairy Sci.* 91:3649–3660.
- Pryce, J.E., R.J. Esslemont, R. Thompson, R.F. Veerkamp, M.A. Kossabati, and G. Simm. 1998. Estimation of genetic parameters using health, fertility and production data from a management recording system for dairy cattle. *Anim. Sci.* 66:577–584.

- Pszczola, M., R.F. Veerkamp, Y. De Haas, E. Wall, T. Strabel, and M.P.L. Calus. 2013. Effect of predictor traits on accuracy of genomic breeding values for feed intake based on a limited cow reference population. *animal*. 7:1759–1768.
- De Roos, A.P.W., B.J. Hayes, R.J. Spelman, and M.E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. *Genetics*. 179:1503–1512.
- Tenghe, A.M.M., A.C. Bouwman, B. Berglund, E. Strandberg, J.Y. Blom, and R.F. Veerkamp. 2015. Estimating genetic parameters for fertility in dairy cows from in-line milk progesterone profiles. *J. Dairy Sci*. 98:5763–73.
- Tenghe, A.M.M., A.C. Bouwman, B. Berglund, E. Strandberg, D.J. de Koning, and R.F. Veerkamp. 2016. Genome-wide association study for endocrine fertility traits using single nucleotide polymorphism arrays and sequence variants in dairy cattle. *J. Dairy Sci*. 99:5470–5485.
- Thompson, R., and K. Meyer. 1986. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livest. Prod. Sci*. 15:299–313.
- VanRaden, P.M. 2008. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci*. 91:4414–4423.
- Veerkamp, R.F., J.K. Oldenbroek, H.J. Van Der Gaast, and J.H.J. Van Der Werf. 2000. Genetic Correlation Between Days Until Start of Luteal Activity and Milk Yield, Energy Balance, and Live Weights. *J. Dairy Sci*. 83:577–583.
- Wientjes, Y.C.J., P. Bijma, R.F. Veerkamp, and M.P.L. Calus. 2015. An Equation to Predict the Accuracy of Genomic Values by Combining Data from Multiple Traits, Populations, or Environments. *Genetics*. doi:10.1534/genetics.115.183269.

Supplementary Material

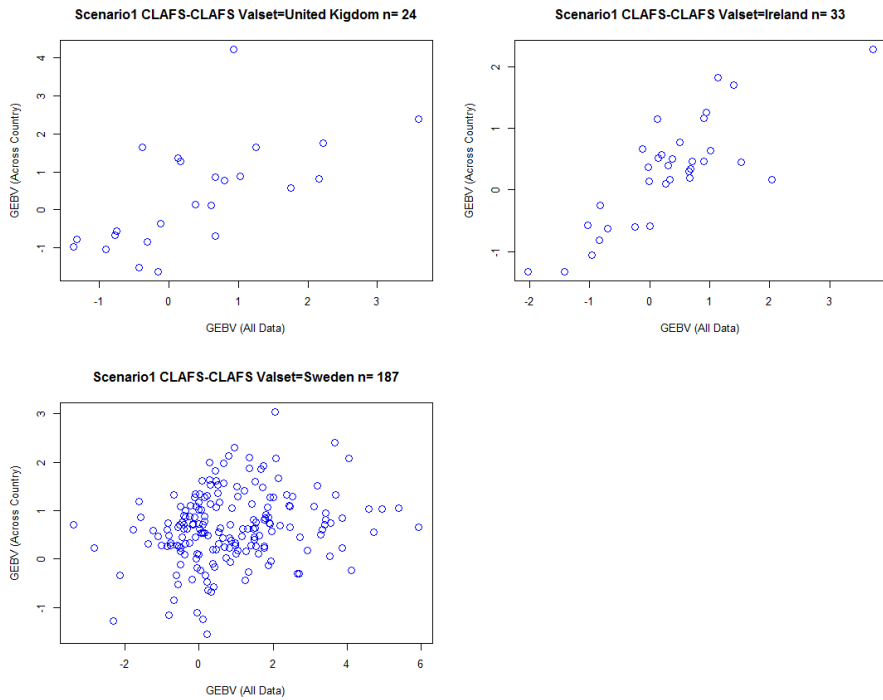


Supplementary Figure 4.1 Regression of pedigree-based estimated breeding values (PED-EBV) for calving to first service (CFS) obtained from univariate predictions using the whole dataset on genomic estimated breeding values (GEBV) for CFS obtained from prediction of CFS in bivariate analysis with commencement of luteal activity (CLA) as predictor trait (Scenario 2); CFS was the target trait



Supplementary Figure 4.2 Plot of genomic estimated breeding values (GEBV) for calving to first service (CFS) obtained from univariate prediction of CFS against GEBV for CFS obtained from prediction of CFS in bivariate analysis with commencement of luteal activity (CLA) as predictor trait (Scenario 2)

4 Progesterone profiles in genomic prediction



Supplementary Figure 4.3 Plot of genomic estimated breeding values (GEBV) for the interval from commencement of luteal activity to first service (CLAFS) obtained from across country predictions on GEBV for CLAFS obtained from univariate predictions of CLAFS using the whole dataset in scenario 1

5

Improving accuracy of bulls' predicted genomic breeding values for fertility using daughters' milk progesterone profiles

A. M. M. Tenghe,^{1,2,3} A.C. Bouwman,² B. Berglund,³ D.J. de Koning,³ R. F. Veerkamp^{1,2}

¹Animal Breeding and Genomics Centre, Wageningen Livestock Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands

²Animal Breeding and Genomics Centre, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands

³Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, P. O. Box 7023, SE-750 07 Uppsala, Sweden

Journal of Dairy Sciences (Submitted)

Abstract

The main objective of this study was to investigate the potential accuracy of multi-trait genomic selection when combining a cow training population measured for the endocrine fertility trait C-LA (commencement of luteal activity) with a training population of bulls with daughter observations for classical fertility traits. The potential accuracy of across country genomic prediction and optimal recording strategies of C-LA for genomic prediction was also investigated in terms of the number of farms and recording period for C-LA. Predicted accuracy was obtained by estimating population parameters for the traits in a dataset of 3,136 Holstein Friesian cows with 8,080 lactations, and using a deterministic prediction equation. The effect of different factors such as genetic correlations, heritability, and reliability of C-LA on the accuracy of genomic prediction were investigated. When considering a classical breeding goal trait like calving interval (Clnt), and when there is an existing large training population with bull EBV for Clnt, there is no benefit in adding cow C-LA records to the training population. However, when considering an endocrine breeding goal trait like C-LA, accuracy is substantially improved. For the endocrine breeding goal, and genetic correlation of 0.3 between C-LA and Clnt, when the existing training population was 10,000 bulls EBV, predicted accuracy increased from 0.22, to 0.57 when 15,000 cows with C-LA records were added to the bull training population; for genetic correlation of 0.7, accuracy increased from 0.51 to 0.63. In across country predictions, when the training population from Sweden was small, e.g., 200 cows, accuracy increased from 0.19 to 0.31 when 15,000 cows were added from the Netherlands, for a genetic correlation of 0.5 between countries, and from 0.19 to 0.48 for genetic correlation of 0.9. There was little to no increase in accuracy when the training population from Sweden was large. Optimizing the recording strategy for C-LA can maximize accuracy of genomic prediction. That is, for genomic prediction of fertility using C-LA, it is more important to have more animals with C-LA records than more C-LA records per animal. For example, 15,000 records obtained from 3 different recording strategies yield different accuracies due to different number of cows in the training population. That is, for 15,000 records obtained from 15 farms on 4,875 cows in 10 years, accuracy of prediction is 0.54, when the 15,000 records are obtained from 30 farms on 6,000 cows in 5 years, accuracy is 0.56, and accuracy is 0.57 for the 15,000 records obtained from 50 farms on 7,500 cows in 3 years. Therefore, for genomic prediction, it is more important to have records from more cows than more lactations per cow in the training population.

Key words: multi-trait genomic prediction, milk progesterone, dairy cattle, accuracy

5.1 Introduction

Although milk progesterone (P4) levels have been widely accepted as valid indicators of fertility in dairy cows, their application in routine genetic evaluation schemes has been constrained by the high cost associated with collecting sufficient samples per cow. Until recently, methods to measure milk P4 levels were labor intensive. They often entailed manually taking several milk samples per cow, analyzing, and recording the results. However, in-line technology like the Herd Navigator™ (HN, DeLaval Intl, Tumba, Sweden) now exists to measure P4 on individual cows during each milking, making it possible to sample a larger number of animals at a reasonable cost. But because few herds currently have the in-line recording system, endocrine fertility traits that are defined from milk P4 levels would be most useful in a genomic selection scheme, where cows from contract herds form the reference population. In comparison, classical fertility traits are routinely measured on a large scale and are still useful target traits in national genetic evaluations. Therefore, combining (genomic predictions for) endocrine and classical fertility traits may provide a more accurate prediction of fertility. In a previous study, using real data, the improvement in accuracy achievable from using endocrine fertility traits concurrent with a classical trait (interval from calving to first service) in genomic prediction of fertility was quantified (Tenghe et al., 2016a). However, the extent of exploring the impact of number of cows with endocrine traits in the training population in that study was limited by the data size. In this study, the limitation was overcome by using a deterministic prediction equation to evaluate accuracy.

The objectives of this study were to 1) investigate the potential accuracy of different scenarios when combining a cow training population measured for the endocrine fertility trait C-LA (commencement of luteal activity) with a training population of bulls with daughter observations for classical fertility traits for multi-trait genomic prediction of fertility; 2) investigate the potential accuracy of across country genomic prediction; 3) investigate recording strategies that optimally use the Herd Navigator for genomic prediction, in terms of, the number of farms, recording period for endocrine traits, and number of lactations per cow. For the above objectives, predicted accuracy was obtained by estimating population parameters for the traits in a dataset of Holstein Friesian cows, and using a deterministic prediction equation (Wientjes et al., 2015). The effect of different factors such as genetic correlation between populations, heritability and reliability of the endocrine fertility trait on the accuracy of genomic prediction were investigated.

5.2 Materials and Methods

5.2.1 Phenotypes

The phenotype data consisted of in-line progesterone (P4) records from 15 commercial herds in the Netherlands. In total, phenotypic data were available for 8,080 lactations on 3,136 Holstein Friesian cows. Milk sampling, measuring and recording of P4 level was performed with the Herd Navigator™ (HN, DeLaval Intl, Tumba, Sweden) on all farms. Sampling frequency for P4 measurement was based on a biological model (Friggens et al., 2008), but on average, was undertaken every 2 days. The endocrine fertility trait investigated was commencement of luteal activity (C-LA) which is the number of days from calving to start of luteal activity, as defined in Tenghe et al. (2015). The classical fertility traits were: interval from calving to first service (CFS) and calving interval (CInt).

5.2.2 Genotypes

The genotypes were used to calculate the effective number of chromosome segments (Me) across predicted and training populations; Me is one of the parameters used in the deterministic equation to predict accuracy. Genotypes were available from a previous study (Tenghe et al., 2016b), where 1,907 cows from the Netherlands were genotyped with the GeneSeek Genomic Profiler Bovine HD (GeneSeek, Lincoln, NE, USA), containing 76,883 SNP (80k), and an additional 6 cows were genotyped with a custom Illumina 6k array, whereas 1,946 cows from Ireland, Sweden and the United Kingdom were genotyped with the Illumina BovineSNP50 v1 BeadChip (Illumina Inc., San Diego, CA, USA) which contained 54,001 SNP (50k). The 50k and 80k SNP arrays had 25,815 SNP in common, the rest were imputed from one dataset to the other and vice versa using BEAGLE 3.3.2. (Browning and Browning, 2009), resulting in a total of 102,062 SNPs. The quality criteria applied before imputation were minimum call rate of 95% per animal and 97% per SNP, and minimum GC-score of 0.3. After imputation, SNPs were retained for analysis if they fulfilled all the following quality criteria: 1) minor allele frequency (MAF) > 0.01, and 2) no extreme deviation from Hardy-Weinberg Equilibrium ($p < 0.01$). Following all edits, the final data contained 85,485 SNP for 3,739 cows. Of the 3,739 cows with genotypes, 1,623 cows from the Netherlands and 223 from Sweden were used to calculate Me.

5.2.3 Estimating variance components

Variance components were estimated with mixed linear models that use the restricted maximum likelihood method. The model used was:

$$y_{ijkl} = \mu + p_j + hys_k + b_1 \times ca_l(p_j) + b_2 \times ca_l^2(p_j) + a_i + pe_i + e_{ijkl}$$

where y_{ijkl} was the phenotype of the analyzed trait for individual i , μ was the overall mean, p_j was the fixed effect of parity j (cows in parity 4 or above were grouped to a common class); hys_k was the fixed effect of herd-year-season combination k , with calving season defined as winter: December-February, spring: March-May, summer: June-August, autumn: September-November; $b_1 ca_l(p_j)$ was the fixed regression on age at calving l (in months), nested within parity p_j , with regression coefficient b_1 ; $b_2 ca_l^2(p_j)$ was the fixed regression on age at calving l , fitted as a quadratic covariate, nested within parity p_j , with regression coefficient b_2 ; a_i was the random genetic effect of cow i : $\sim N(0, A\sigma_a^2)$, A was the additive genetic relationship matrix and σ_a^2 was the additive genetic variance; pe_i was the random permanent environmental effect of cow i to account for repeated measures within cow: $\sim N(0, I\sigma_{pe}^2)$, I was an identity matrix including all animals (but animals without information on repeated records are automatically set at zero) and σ_{pe}^2 was the permanent environment variance; and e_{ijkl} was the random error term, $\sim N(0, I\sigma_e^2)$, I was an identity matrix and σ_e^2 was the residual variance. ASReml 4.1 (Gilmour et al., 2014) was used for estimating the variance components.

5.2.4 Deterministic prediction of accuracy

We used the deterministic equation of Wientjes et al. (2015), which uses population parameters as input, to predict the accuracy of genomic prediction when different populations are combined into one training population. The populations might be populations from different lines or environments, or populations measured for different traits. The prediction equation is:

$$r_{GEBV} = \sqrt{\left[rLD_{A,C} rG_{A,C} \sqrt{\frac{h_A^2}{Me_{A,C}}} rLD_{B,C} rG_{B,C} \sqrt{\frac{h_B^2}{Me_{B,C}}} \right] \left[\begin{array}{cc} \frac{h_A^2}{Me_{A,C}} + \frac{1}{N_A} & rG_{A,B} \frac{\sqrt{h_A^2 h_B^2}}{\sqrt{Me_{A,C} Me_{B,C}}} \\ rG_{A,B} \frac{\sqrt{h_A^2 h_B^2}}{\sqrt{Me_{A,C} Me_{B,C}}} & \frac{h_B^2}{Me_{B,C}} + \frac{1}{N_B} \end{array} \right]^{-1} \left[\begin{array}{c} rG_{A,C} \sqrt{\frac{h_A^2}{Me_{A,C}}} \\ rG_{B,C} \sqrt{\frac{h_B^2}{Me_{B,C}}} \end{array} \right]}$$

where r_{GEBV} is the accuracy of genomic estimated breeding values; A and B are the populations that are combined into one training population; C represents the

animals in the predicted population for which only genotypes are available; Me is the effective number of chromosome segments across predicted and training populations; h^2 is the reliability of the breeding value used, i.e. the trait heritability for a single record on a cow, a reliability for progeny tested bulls (a value of 0.8 was assumed), and the reliability when cows have multiple repeated records in each training population; N is the number of individuals in each training population; rG is the genetic correlation between the traits measured in the different populations; rLD is the genetic variance in the predicted population captured by SNP in the training populations.

5.2.5 Scenarios

Three main scenarios were examined for accuracy of genomic estimated breeding values (GEBV). The first two scenarios (Figure 5.1), represent multi-trait genomic prediction scenarios where prediction is for an endocrine or classical fertility breeding goal. A cow training population (population B) with own records for an endocrine fertility trait is combined with training population A, consisting of classical fertility records. Population A is either a bull training set with EBV based on daughter information for a classical trait, or a cow training set where cows have own phenotypes for a classical trait. The combined training set (population A and B) is then used to predict GEBVs for animals from population C. In scenario 1, the prediction in population C is for an endocrine breeding goal, and for a classical breeding goal in scenario 2. The animals from population C could be young genotyped bulls or cows with genotypes only and no phenotypes. Scenario 3 (Figure 5.2) represents multi-population prediction where countries collaborate to set-up a training population. For example, the training set of cows with endocrine fertility traits in Sweden (population A) is increased by adding a training set of cows with endocrine traits from the Netherlands (population B) to predict GEBVs for animals in Sweden (population C). The effect of heritability, genetic correlation between countries, number of cows in training population A and number of cows in training population B on predicted accuracy were investigated in scenario 3. For scenario 1 and 2, the parameters impacting predicted accuracy that were investigated included: heritability, genetic correlation between the endocrine and classical traits, and number of cows with endocrine traits added in training population B. The parameters used in all scenarios are in Table 5.1.

5 Progesterone profiles and bull fertility

Table 5.1 Input parameters of scenarios investigated for accuracy of genomic prediction when combining training populations and considering an endocrine or classical breeding goal for fertility in dairy cattle

	Scenario		
	1	2	3
	Multi-trait prediction for endocrine breeding goal	Multi-trait prediction for classical breeding goal	Across country prediction for endocrine breeding goal
Parameter ¹			
Trait training population (A)	Classical (Clnt or CFS) ²	Classical (Clnt or CFS)	Endocrine (C-LA) ²
Trait training population (B)	Endocrine (C-LA)	Endocrine (C-LA)	Endocrine (C-LA)
Trait predicted population (C)	Endocrine (C-LA)	Classical (Clnt or CFS)	Endocrine (C-LA)
N population A	10,000 bulls or cows	10,000 bulls or cows	200 -10,000 cows
N population B	0 - 15,000 cows	0 - 15,000 cows	0 - 15,000 cows
Range h^2A	0.8 (bulls) or 0.05 (cows)	0.8 (bulls) or 0.05 (cows)	0.13
Range h^2B	0.05 - 0.3	0.05 - 0.3	0.13
Range $rG_{A,B}$	0.1- 0.7	0.1 - 0.7	0.5 - 0.9
Range $rG_{A,C}$	0.1- 0.7	0.99	0.99
Range $rG_{B,C}$	0.99	0.1 - 0.7	0.5 - 0.9
$rLD_{A,C}$	0.8	0.8	0.8
$rLD_{B,C}$	0.8	0.8	0.8
$Me_{A,C}$	1,566	1,566	416
$Me_{B,C}$	1,566	1,566	2,683

¹A and B are the different populations that are combined into one training population; C is the predicted population;

N is the number of individuals in each training population; h^2 is the heritability in each training population;

rG is the genetic correlation between the populations; rLD is the genetic variance in the predicted population captured by SNP in the training population; Me is the effective number of chromosome segments across predicted and training populations ²C-LA = Commencement of luteal activity, Clnt = calving interval, CFS = interval from calving to first service

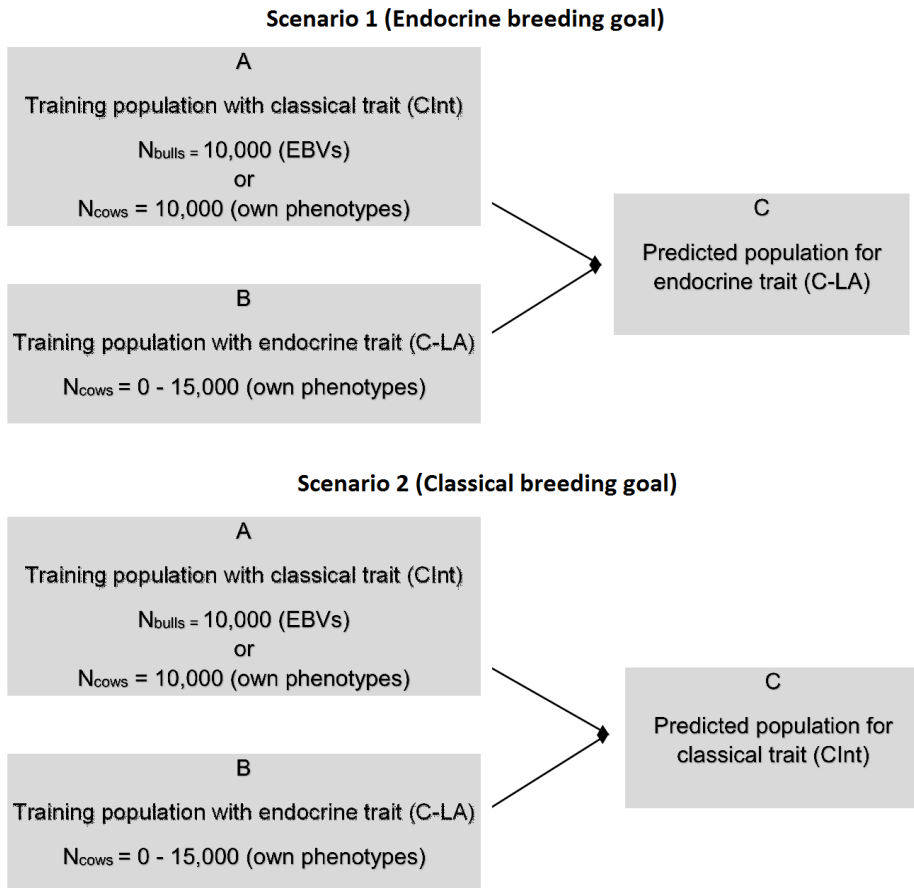


Figure 5.1 Flowchart of multi-trait scenarios investigated, where a cow training population (population B) with own records for an endocrine fertility trait is combined with training population A, consisting of classical fertility records. The endocrine trait is commencement of luteal activity (C-LA), and calving interval (CInt) is the classical trait. Population A is either a bull training set with estimated breeding values (EBV) based on daughter information from a classical trait, or a cow training set where cows have own records for a classical trait. The combined training set is then used to predict genomic breeding values for animals from population C. The prediction in population C is for an endocrine breeding goal in scenario 1, and for a classical breeding goal in scenario 2; N_{bulls} in population A is the number of bulls whose information comes from at least 100 daughters; N_{cows} represents the number of cows with their own records.

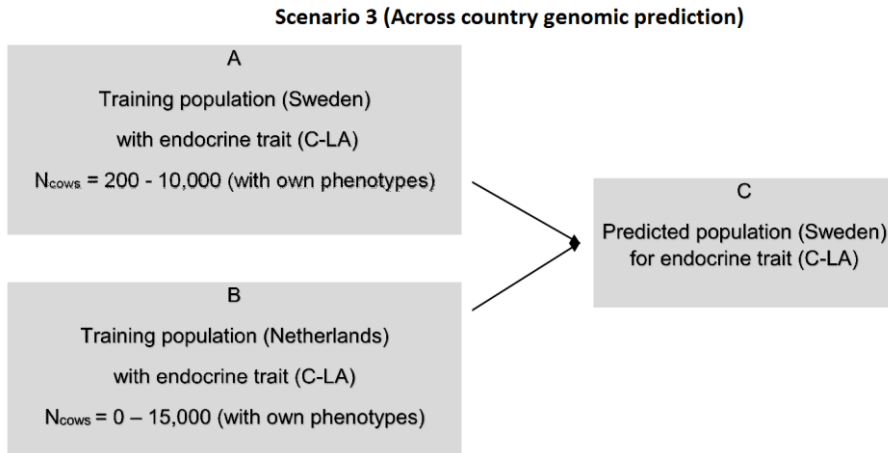


Figure 5.2 Flowchart of a multi-environment (or across country) genomic prediction scenario investigated in dairy cattle, where because the training set of cows with endocrine fertility traits in Sweden (population A) is limited, the training population is increased by adding a training set of cows with endocrine traits from the Netherlands (population B) to predict genomic estimated breeding values for animals in Sweden (population C). The interval from calving to commencement of luteal activity (C-LA) is used as an example for endocrine fertility traits; N_{cows} represents the number of cows with their own records

5.2.6 Recording strategies for endocrine fertility traits

To investigate different strategies that optimally use the investment in the Herd Navigator for genomic prediction, we evaluated the impact of the number of farms and number of cows on predicted accuracy for scenario 1. We also varied the recording period for endocrine traits because simply varying the number of records for genomic prediction is not sufficient since cows are recorded for several lactations on a single farm. We assumed a strategy with different number of farms (15, 30, and 50) equipped with the Herd Navigator and different lengths of recording periods, that is, 1, 2, 3, 5, 10, and 15 years of recording C-LA. Each farm was assumed to have 100 lactating cows at the start of recording, with a replacement and culling rate of 25% per year where the oldest animals were replaced, and each cow would have one calf per year. With multiple records per cow the heritability estimate for C-LA in the prediction equation defined above, was replaced with the reliability calculated using the formula; $\text{reliability} = \frac{mh^2}{(m-1)t+1}$, where m = number of records per cow, h^2 = heritability of the trait, and t = repeatability of the trait. The h^2 (0.13) and t (0.26) of C-LA obtained from the mixed model described above were used. The number of

cows with C-LA records for each recording period and corresponding reliability estimates were used in the prediction equation to predict accuracy of GEBVs. The details of the computed reliabilities, number of cows and records for each recording period are in Table 5.2.

Table 5.2 Input parameters used to investigate optimal recording strategies for endocrine fertility traits for use in genomic prediction [Text].

No. of farms	No. of recording years	No. of unique cows per farm	No. of records per farm	Total cows on all farms	Total records on all farms	No. of records per cow	Reliability C-LA ¹
15	1	100	100	1,500	1,500	1.0	0.13
	2	125	200	1,875	3,000	1.6	0.18
	3	150	300	2,250	4,500	2.0	0.21
	5	200	500	3,000	7,500	2.5	0.23
	10	325	1000	4,875	15,000	3.1	0.26
	15	450	1500	6,750	22,500	3.3	0.27
30	1	100	100	3,000	3,000	10	0.13
	2	125	200	3,750	6,000	1.6	0.18
	3	150	300	4,500	9,000	2.0	0.21
	5	200	500	6,000	15,000	2.5	0.23
	10	325	1000	9,750	30,000	3.1	0.26
	15	450	1500	13,500	45,000	3.3	0.27
50	1	100	100	5,000	5,000	1.0	0.13
	2	125	200	6,250	10,000	1.6	0.18
	3	150	300	7,500	15,000	2.0	0.21
	5	200	500	10,000	25,000	2.5	0.23
	10	325	1000	16,250	50,000	3.1	0.26
	15	450	1500	22,500	75,000	3.3	0.27

¹C-LA = Commencement of luteal activity; Reliability computed as $mh^2/(m-1)t+1$, m = number of records per cow, h² = heritability of the trait, and t = repeatability of the trait

5.3 Results

5.3.1 Genetic parameters

The estimated heritability and repeatability for each trait are in Table 5.3. Heritability estimates were 0.06 for Clnt, 0.11 for CFS and 0.13 for C-LA. Repeatability estimates were low to moderate, ranging from 0.09 to 0.26, with the highest value observed for C-LA. The estimated genetic correlations (\pm SE) were 0.58 (\pm 0.12) between C-LA and CFS, and 0.31 (\pm 0.21) between C-LA and Clnt.

Table 5.3 Overall mean, standard deviation (S.D), estimates of heritability (h^2), repeatability (t) and their standard errors (\pm SE) below estimates for endocrine and classical fertility traits from Holstein Frisian cows in the Netherlands

Trait ¹	Mean	SD	Number of Lactations	Number of cows	h^2	t
C-LA, d	38.96	18.98	5043	2,748	0.13 (0.03)	0.26 (0.02)
Clnt, d	409.76	66.54	4030	1,916	0.06 (0.03)	0.09 (0.02)
CFS, d	88.94	33.93	6655	2,898	0.11 (0.02)	0.11 (0.02)

¹C-LA = Commencement of luteal activity, Clnt = calving interval, CFS = interval from calving to first service

5.3.2 Multi-trait genomic prediction considering an endocrine breeding goal

First, we considered an endocrine breeding goal in scenario 1, where C-LA is the breeding goal trait, and training population A is either 10,000 bulls with EBV or 10,000 cows with own records for a classical trait (Clnt or CFS), whereas training population B is cows with own records for C-LA. The training populations are combined and used to predict C-LA for animals from population C. Predicted accuracy as a function of number of cows from population B added to training population A are shown in Figure 5.3, assuming a heritability of 0.05 to 0.3 for C-LA and genetic correlations of 0.1 to 0.7 between C-LA and the classical trait. When training population A consists of bull EBV (Figure 5.3a), and heritability of C-LA is 0.1, with a genetic correlation of 0.3, accuracy increases from 0.22 when only population A is used, to 0.57 when 15,000 cows are added from population B; with genetic correlation of 0.7, accuracy increases from 0.51 to 0.63. With a higher heritability for C-LA of 0.3, the accuracy increases more (from 0.22 to 0.68) when 15,000 cows are

added from population B for genetic correlation of 0.3, and from 0.51 to 0.70 with genetic correlation of 0.7. When training population A consists of cows with own records of Clnt or CFS instead of bull EBV (Figure 5.3b), and the heritability of C-LA is 0.1, with a genetic correlation of 0.3, the accuracy increases from 0.04 when only population A is used, to 0.55 when 15,000 cows are added from population B; with a genetic correlation of 0.7, the accuracy increases from 0.28 to 0.57. A similar trend of increased accuracy was observed with heritability of 0.3 for C-LA (Figure 5.3b). In general, the accuracy is higher when the genetic correlation between C-LA and the classical trait is higher, and the accuracy increases when an increasing number of cows with C-LA records is added to the training population, especially when the genetic correlation is small.

5.3.3 Multi-trait genomic prediction considering a classical breeding goal

The second scenario considered a classical breeding goal where Clnt or CFS is the breeding goal trait. Training population A consists of 10,000 bulls with EBV or 10,000 cows with own records for Clnt or CFS, and training population B is cows with own records for C-LA. The training populations are combined and used to predict CFS or Clnt for animals from population C. Predicted accuracy as a function of number of cows added to training population B is shown in Figure 4. When training population A consists of bull EBV, adding information on cows for C-LA did not increase accuracy, irrespective of the genetic correlation between traits, or heritability of C-LA (Figure 5.4a). However, when the 10,000 animals in training population A are cows with own records for the classical trait, there is some improvement in accuracy when adding cows with C-LA records from population B (Figure 5.4b). That is, when heritability of C-LA is 0.1, with a genetic correlation of 0.3, the accuracy increases from 0.39 when only population A is used, to 0.41 when 15,000 cows are added from population B, and from 0.39 to 0.50 with a correlation of 0.7. For heritability of 0.3, with a genetic correlation of 0.3, the accuracy increases from 0.39 when only population A is used, to 0.42 when adding population B, and from 0.39 to 0.55 with a genetic correlation of 0.7.

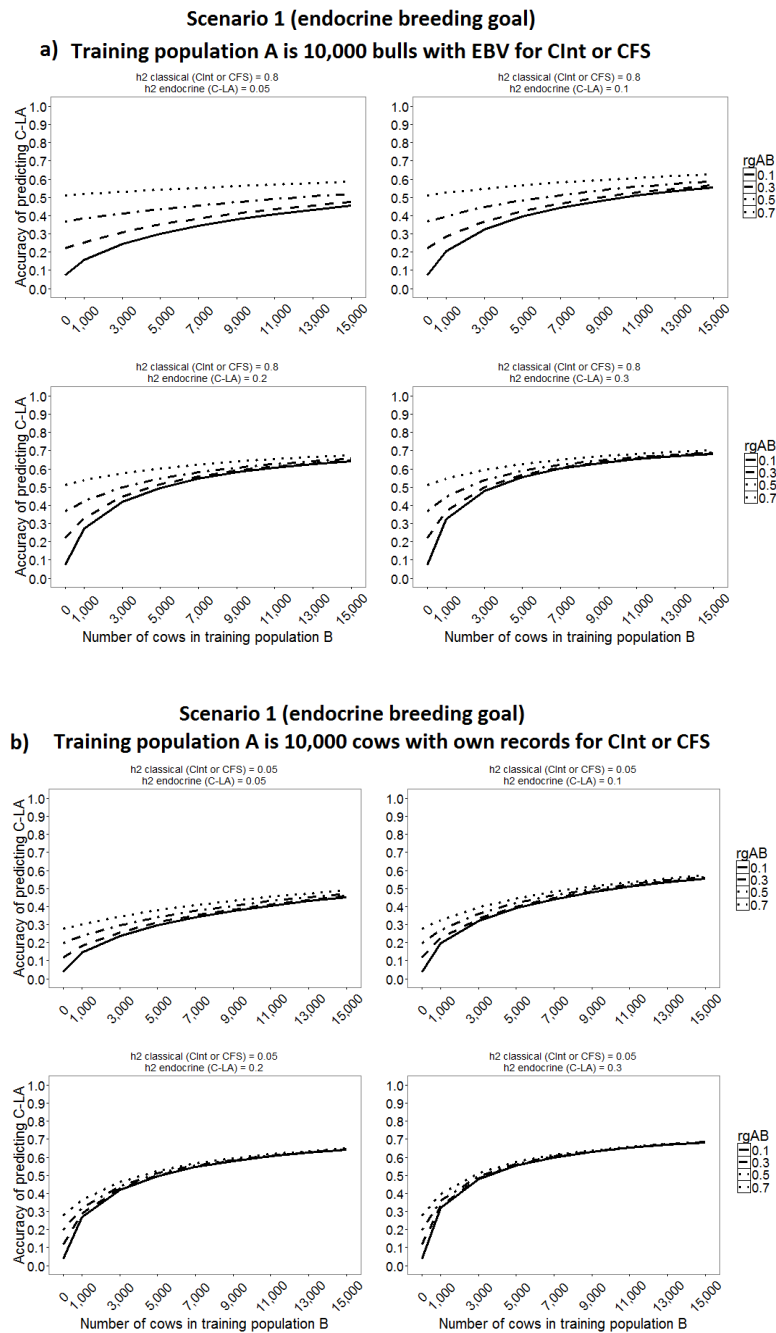


Figure 5.3 Predicted accuracies for fertility when considering the endocrine trait C-LA (commencement of luteal activity) as the breeding goal trait, and adding different number

of cows (population B) with own C-LA records to a training population (population A) consisting of 10,000 individuals (bulls or cows) with records for classical fertility traits. The classical traits considered were calving interval (Clnt) and calving to first service (CFS). When population A was a bull training population with estimated breeding values (EBV) based on daughter information, heritability or reliability of the classical trait was 0.8 (a); when population A was a cow training population with own records, heritability of the classical trait was 0.05 (b). The heritability for the endocrine trait in population B varied from 0.05 to 0.5, and genetic correlation between the endocrine and classical trait (r_{gAB}) varied from 0.1 to 0.7. The effective number of chromosomes (Me) was 1,566, and the proportion of variance captured by SNPs (r_{LD}) was 0.8. The accuracy was based on the formula of Wientjes et al. (2015)

5.3.4 Across country genomic prediction

The across country predictions show that when the training population from Sweden is small, a substantial increase in accuracy can be obtained by adding animals from the Netherlands. For example, the predicted accuracy with 200 cows from Sweden was 0.19, and increased to 0.31 when 15,000 cows were added from the Netherlands, for a genetic correlation of 0.5 between countries; and for a genetic correlation of 0.9, the accuracy increased from 0.19 to 0.48 (Figure 5.5a). Also, when the genetic correlation between countries is high, fewer animals from the Netherlands are needed to obtain substantial accuracy. For example, with 200 animals from Sweden and a correlation of 0.5, 15,000 animals from the Netherlands were needed to achieve an accuracy of 0.31, but when the correlation was 0.9, only 3,000 animals from the Netherlands were needed to achieve the same accuracy (Figure 5a). Similarly, with 2,000 animals from Sweden, 15,000 animals from the Netherlands were needed to achieve an accuracy of 0.52 when the correlation was 0.5, but only 3,000 animals were needed to attain the same accuracy when the correlation was 0.9 (Figure 5.5c).

However, when the training population from Sweden is large, for example 10,000 animals, there is little to no increase in accuracy when adding animals from the Netherlands. That is, for 10,000 animals from Sweden, the accuracy was 0.69 with and without adding animals from the Netherlands, for a genetic correlation of 0.5; for genetic correlation of 0.9, accuracy increased from 0.69 to 0.70 (Figure 5.5d).

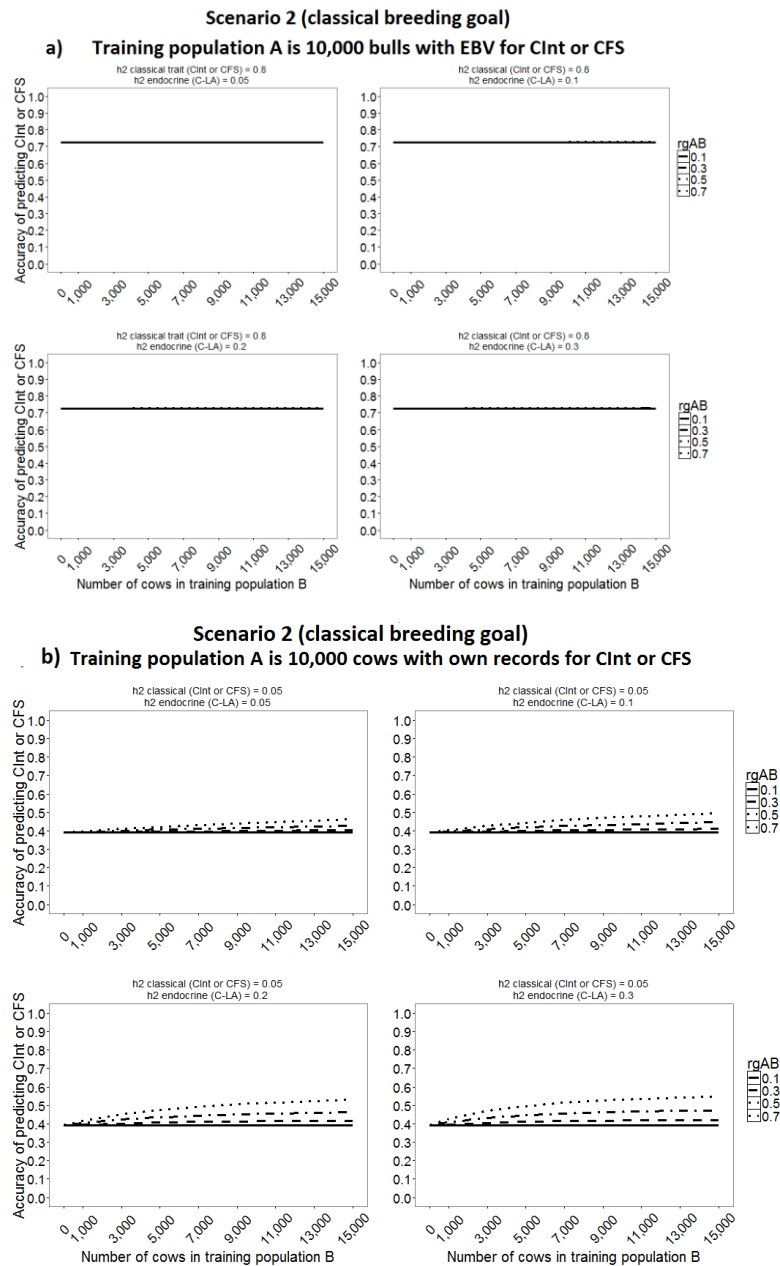


Figure 5.4 Predicted accuracies for fertility when considering the classical traits calving interval (Clnt) or calving to first service (CFS) as the breeding goal trait, and adding different number of cows (population B) with own measures for the endocrine trait C-LA

(commencement of luteal activity) to a training population (population A) consisting of 10,000 individuals (bulls or cows) with classical fertility traits. When population A was a bull training population with estimated breeding values (EBV) based on daughter information, heritability or reliability of the classical trait was 0.8 (a); when population A was a cow training population with own records, heritability of the classical trait was 0.05 (b). The heritability for the endocrine trait in population B varied from 0.05 to 0.5, and the genetic correlation between the endocrine and classical trait (rg_{AB}) varied from 0.1 to 0.7. The effective number of chromosomes (Me) was 1,566, and the proportion of variance captured by SNPs (r_{LD}) was 0.8. The accuracy was based on the formula of Wientjes et al. (2015)

Scenario 3 (across country prediction)

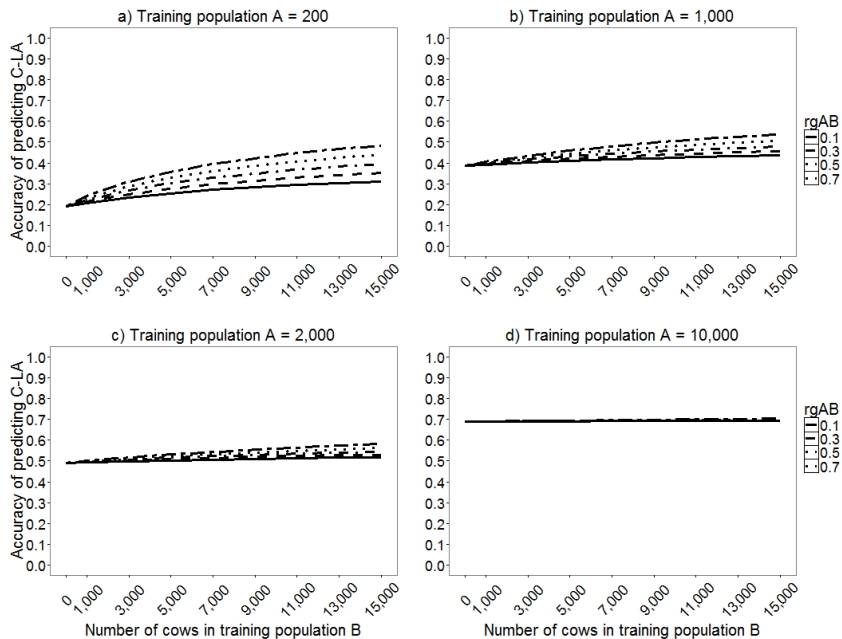


Figure 5.5 Predicted accuracies for fertility when adding different number of cows from the Netherlands (population B) with own measures for the endocrine trait C-LA (commencement of luteal activity) to a cow training population consisting of 200 to 10,000 cows from Sweden with own measures for C-LA, to predict animals in Sweden. The heritability of the C-LA was 0.13, and the genetic correlation between the populations (rg_{AB}) varied from 0.5 - 0.9. The effective number of chromosomes (Me) was 416 ($Me_{A,C}$) and 2,683 ($Me_{B,C}$), and proportion of variance captured by SNPs (r_{LD}) was 0.8. The accuracy was based on the formula of Wientjes et al. (2015)

5.3.5 Recording strategies for endocrine fertility traits

Figure 5.6 presents predicted accuracy for the endocrine breeding goal, as a function of recording strategies for C-LA in terms of different number of farms on which cows were recorded, and number of years during which C-LA was recorded. The recording strategies influence the number of cows recorded and the number of lactations for each cow added from training population B. The reliability of C-LA for the different recording strategies are in Table 5.3, they are 0.13 for 1 year of recording, 0.18 (for 2 years), 0.21 (for 3 years), 0.23 (for 5 years), 0.26 (for 10 years), and 0.27 (for 15 years). The number of lactations (records) per cow, number of cows, and total number of records for each recording strategy are in Table 5.3. The predicted accuracy increases substantially when multiple C-LA records are used in prediction compared to single records; the accuracy increases from 0.33 with single records, to 0.38 with multiple records (on average 1.6 records per cow) from 2 years of recording (Figure 5.6a). However, when C-LA is recorded for longer periods, there is no substantial benefit in accuracy from multiple records. For example, for 5 years of recording (on average 2.5 records per cow), the accuracy was 0.47, on doubling the recording period to 10 years (on average 3.1 records per cow), the accuracy increased by 0.07 units (Figure 5.6a), but when C-LA was recorded for 15 years (on average 3.5 records per cow) it increased only by 0.05 units. Therefore, the extra lactations for each cow after the second lactation did not add much benefit to the accuracy of prediction.

The results also indicate that in general, the number of cows with C-LA is more important than the number of records for genomic predictions. For example, 15,000 records obtained from 3 different recording strategies yield different accuracies due to different number of cows in the training population. That is, for 15,000 records obtained from 15 farms on 4,875 cows in 10 years, the accuracy of predicting C-LA is 0.54, when the 15,000 records are obtained from 30 farms on 6,000 cows in 5 years, accuracy is 0.56, and the accuracy is 0.57 for the 15,000 records obtained from 50 farms on 7,500 cows in 3 years (Figure 5.6a); a similar trend is observed for the other scenarios (Figure 5.6). Thus, it is more important to have more cows than more lactations per cow in the training population.

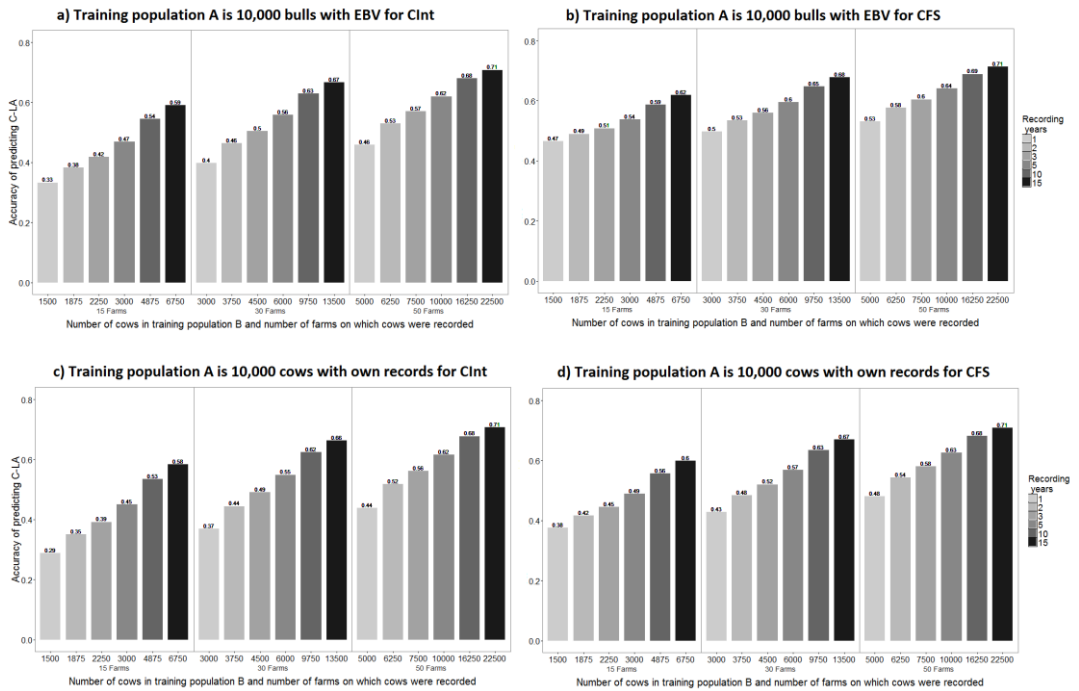


Figure 5.6 Predicted accuracies for commencement of luteal activity (C-LA), as a function of number of cows and recording period in years for C-LA, when considering an endocrine breeding goal. Training population A is combined with different number of cows (with different number of C-LA records per cow) in population B for genomic prediction. Heritability estimates were 0.13 for C-LA, 0.06 for calving interval (Clnt), and 0.11 for calving to first service (CFS). The genetic correlations were 0.31 between C-LA and Clint, and 0.58 between C-LA and CFS. The effective number of chromosomes (Me) was 1,566, and proportion of variance captured by SNPs (rLD) was 0.8. The accuracy was based on the formula of Wientjes et al. (2015)

5.4 Discussion

The main objective of this study was to investigate the potential accuracy multi-trait genomic selection when combining a cow training population (population B) measured for the endocrine fertility trait C-LA, with a training population (population A), measured for classical fertility traits (CFS or Clnt). Population A was either a bull training set with EBV based on daughter information for the classical trait, or a cow training set where cows had own phenotypes for the classical trait. Also investigated was the potential accuracy of across country genomic prediction where training populations from Sweden and the Netherlands were combined to predict animals in Sweden. To achieve these, predicted accuracies were obtained by using population parameters and a deterministic prediction equation, and the effect of genetic correlations, size of training populations, heritability and reliability of the endocrine trait on predicted accuracy were assessed. The results show that when considering a classical breeding goal trait like CFS or Clnt, and when there is an existing large training set with bull EBV for the classical trait, there is no benefit in adding cow C-LA records to the training set. However, when considering an endocrine breeding goal trait like C-LA, the accuracy is substantially improved. For across country prediction, when the training population from Sweden is small (e.g., 200 animals), substantial increase in accuracy can be obtained by adding animals from the Netherlands. But, when the training population from Sweden is large (e.g., 10,000 animals), there is little to no increase in accuracy when adding animals from the Netherlands. Another objective was to investigate recording strategies that optimally use the Herd Navigator for genomic prediction, in terms of, the number of farms, and recording period for C-LA. The results show that for genomic prediction of fertility using C-LA, it is more important to have more animals with C-LA records than more C-LA records per animal.

5.4.1 Accuracy

The general increase in accuracy from multi-trait genomic prediction is in accordance with previous studies that have shown that when the genetic correlation between traits is high, multi-trait prediction improves accuracy; this has been shown in stochastic (Calus and Veerkamp, 2011; Jia and Jannink, 2012) and deterministic (Calus et al., 2013) simulations as well as in real data (Calus et al., 2013; Cooper et al., 2016; Tenghe et al., 2016a). However, a substantial increase in accuracy was observed in scenario 1 where C-LA was the breeding goal trait, and cows with own C-LA records were added to a large training population of bulls with EBV for a classical trait for prediction. This may be explained by the fact that in scenario 1,

because the breeding goal trait (C-LA) was not measured in the bull training population, more information from the cows added to the training set was used to predict C-LA. But when CFS or Clnt were the breeding goal trait (scenario 2), adding 15,000 cows with C-LA records to the bull training set did not improve accuracy because the bull training population was measured for the breeding goal trait. This confirms that it is important to measure the breeding goal trait in the training population to benefit from the information in predictions. This was true for C-LA, but could equally apply to similar situations where a predictor trait is used to predict the breeding goal trait. For example, somatic cell counts can be used to predict breeding values for mastitis incidence (Philipsson et al., 1995), and birth weight can be used to predict calving ease or perinatal mortality (Johanson and Berger, 2003). However, it is beneficial to have actual measures of mastitis or perinatal mortality incidence on a set of cows. That is, it is beneficial to have actual measurements on the breeding goal trait if a correlated trait is used to predict it.

The accuracy of breeding values depends on the sources of information included in each phenotypic record. The accuracy of a single phenotypic measurement of an animal itself is equal to the square root of the heritability ($\sqrt{h^2}$), and the reliability = h^2 , however, this is not true when multiple measurements per animal are used. In practice, more than one measurement for C-LA will be available for cows with multiple lactations. Therefore, the predicted accuracy was also evaluated accounting for repeated records, by substituting the heritability of C-LA in the prediction equation with the reliability. When considering C-LA as breeding goal trait, the predicted accuracy increased substantially when more records were available either due to measuring on more farms (i.e. more animals), or due to measuring for a longer period on the farms (i.e. repeated records). However, measuring C-LA on more farms gave more increase in accuracy than measuring for longer periods on a few farms. Little to no increase in accuracy was observed when considering a classical breeding goal and accounting for repeat records, especially when training population A consisted of bulls with EBV for the classical trait.

The trend of increased accuracy from combining two populations from different countries in one training population was expected due to the increase in size of the training population. When the training population from Sweden was small, the accuracy was improved on adding animals from the Netherlands. Previous studies that investigated the accuracy of multi-population genomic prediction by combining Holstein populations from different countries found a similar trend (Lund et al., 2011; Haile-Mariam et al., 2015; de Haas et al., 2015). A similar trend has also been

observed in other studies involving different breeds, e.g., Jersey populations (Wiggans et al., 2015; Haile-Mariam et al., 2015), and Brown Swiss populations (Zumbach et al., 2010; Jorjani, 2012). Little to no improvement in accuracy was observed when the training population from Sweden was large, e.g., 10,000 because the accuracy obtained with 10,000 animals was already high.

5.4.2 Implications for practical dairy breeding

In genomic selection, the requirement of collecting many phenotypes on progeny is minimized, as opposed to conventional selection. Therefore, genomic selection provides a new opportunity to reconsider genetic improvement using new traits like endocrine fertility traits, and to start selecting for them. For genomic selection with new traits, the phenotypes need to be collected only for the animals in the training population, hence collection of the phenotypes may be conducted on contract farms where collecting expensive or difficult to measure data is feasible. The training population for new traits will be smaller, compared to the routinely recorded traits, and may result in lower accuracy in genomic prediction. The expected limited size of the training population for new traits stresses the need for methods that optimally record and use all available information. One objective of this study was therefore to investigate recording strategies that optimally use the Herd Navigator for genomic prediction. This was done by evaluating the effect of number of farms, and recording period for C-LA on accuracy when combining endocrine and classical fertility traits for prediction. This study showed that in terms of accuracy, it is more beneficial to record more cows for C-LA than more C-LA records per cow. That is, for example, it is more beneficial to collect 15,000 records for C-LA on 7,500 cows, than 15,000 records on 4,875 cows. This is important, as it raises an important question of whether it is beneficial to invest in more contract farms and record C-LA for a shorter period, or invest in fewer contract farms and record for longer period to attain a target training population size.

The investment option will depend partly on the breeding goal trait, and cost of measuring the trait. In our case, if we consider that each farm requires one Herd Navigator for collecting P4 records, and the breeding goal trait is C-LA, then it will be beneficial for the accuracy of estimating genomic breeding values, to invest in more contract farms. That is, for example, there is more benefit investing in 50 contract farms to collect 15,000 records on 7,500 cows in three years, than investing in 15 farms to collect the 15,000 records on 4,875 cows in 10 years. Such recording strategies could equally be applied to traits that are not routinely measured because

they are difficult or expensive to measure, and for which obtaining large number of phenotypes for the training population in genomic prediction is problematic. On the other hand, if the aim is for a classical breeding goal trait like Clnt or CFS, then collecting P4 data in general would be a waste of time and money, especially when there is a large existing population of bulls with EBV for the classical trait.

The substantial increase in accuracy due to increase in size of the training population by combining the Dutch and Swedish populations in one training population denotes that there is a benefit in collaboration between countries enabling an increased size of the training population for expensive or difficult to measure traits. This collaboration is especially beneficial when the genetic correlation between countries is high. In this study, when the genetic correlation was high, fewer number of animals were needed to attain the same level of accuracy as for when the correlation was lower. This also means that for mutual benefit or as incentive for collaboration between countries, it would be more interesting for both countries when setting up a training set of 5,000 C-LA records yields an accuracy of 0.52 provided the genetic correlation is 0.9, than collecting 17,000 C-LA records to obtain the same accuracy when the correlation is 0.5.

5.5 Conclusions

Genomic selection and new phenotyping technologies like the Herd Navigator for measuring P4 level in milk present an opportunity to fundamentally redesign dairy breeding programs and include selection for new phenotypes like endocrine fertility traits. Optimal use of these technologies may require a different balance in the investments that underpin breeding activities compared to current practice. This study showed that the use of C-LA in genomic selection can be maximized based on the breeding goal objective considered. When the breeding goal is a classical trait like CFS or Clnt, and when there is an existing large training set with bull EBV for the classical trait, there is no benefit in adding cow C-LA records to the training set. But when the breeding goal is an endocrine trait like C-LA, accuracy is substantially improved when cows with C-LA records are added to the bull training set, and the accuracy is further improved when multiple records of C-LA per cow are accounted for in predictions. The study also showed that optimizing the recording strategy for C-LA can maximize the accuracy of prediction, that is, for genomic prediction of fertility using C-LA, it is more important to have more animals with C-LA records than more C-LA record per animal.

Acknowledgements

The authors acknowledge Lattec I/S (Denmark) for providing the in-line progesterone data, and the Cooperative Cattle Improvement Organization CRV BV (Arnhem, the Netherlands) for providing the pedigree data for this study. A. M. M. T. benefited from a joint grant from the European Commission and the European Union under the grant agreement n°311776, within the framework of Erasmus-Mundus joint doctorate "EGS-ABG" (Paris, France). This study was performed as part of the Prolific project which is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement n°311776.

References

- Browning, B.L., and S.R. Browning. 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *Am. J. Hum. Genet.* 84:210–223.
- Calus, M.P.L., Y. De Haas, M. Pszczola, and R.F. Veerkamp. 2013. Predicted accuracy of and response to genomic selection for new traits in dairy cattle. *Animal*. 7:183–191.
- Calus, M.P.L., and R.F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. *Genet Sel Evol.* 43:1–14.
- Cooper, T.A., G.R. Wiggans, and P.M. VanRaden. 2016. Short communication: Analysis of genomic predictor population for Holstein dairy cattle in the United States-Effects of sex and age. *J. Dairy Sci.* 98:2785–2788.
- Daetwyler, H.D. 2009. Genome-wide evaluation of populations. Ph.D. Thesis, Animal Breeding and Genomics Centre, Wageningen, The Netherlands.
- Erbe, M., B. Gredler, F.R. Seefried, B. Bapst, and H. Simianer. 2013. A function accounting for training set size and marker density to model the average accuracy of genomic prediction. *PLoS One*. 8:e81046.
- Friggens, N.C., M. Bjerring, C. Ridder, S. Højsgaard, and T. Larsen. 2008. Improved detection of reproductive status in dairy cows using milk progesterone measurements. *Reprod. Domest. Anim.* 43 Suppl 2:113–21.
- Gilmour, A.R., B.J. Gogel, B.R. Cullis, S.J. Welham, and R. Thompson. 2014. ASReml User Guide Release 4.1 Structural Specification, VSN International Ltd, Hemel Hempstead, HP1 1ES, UK www.vsn.co.uk.

- Goddard, M.E., B.J. Hayes, and T.H.E. Meuwissen. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. Anim. Breed. Genet.* 128:409–421.
- de Haas, Y., J.E. Pryce, M.P.L. Calus, E. Wall, D.P. Berry, P. Løvendahl, N. Krattenmacher, F. Miglior, K. Weigel, D. Spurlock, K.A. Macdonald, B. Hulsege, and R.F. Veerkamp. 2015. Genomic prediction of dry matter intake in dairy cattle from an international data set consisting of research herds in Europe, North America, and Australasia. *J. Dairy Sci.* 98:6522–6534.
- Haile-Mariam, M., J.E. Pryce, C. Schrooten, and B.J. Hayes. 2015. Including overseas performance information in genomic evaluations of Australian dairy cattle. *J. Dairy Sci.* 98:3443–3459.
- Jia, Y., and J.-L. Jannink. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*. 192:1513–1522.
- Johanson, J.M., and P.J. Berger. 2003. Birth weight as a predictor of calving ease and perinatal mortality in Holstein cattle. *J. Dairy Sci.* 86:3745–3755.
- Jorjani, H. 2012. Genomic evaluation of BSW populations, InterGenomics: Results and Deliverables. *Interbull Bull.* 43.
- Lund, M.S., A.P.W. De Roos, A.G. De Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, and R. Reents. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet Sel Evol.* 43:43.
- Philipsson, J., G. Ral, and B. Berglund. 1995. Somatic cell count as a selection criterion for mastitis resistance in dairy cattle. *Livest. Prod. Sci.* 41:195–200.
- Tenghe, A.M.M., B. Berglund, E. Wall, R.F. Veerkamp, and D.J. de Koning. 2016a. Opportunities for genomic prediction for fertility using endocrine and classical fertility traits in dairy cattle. *J. Anim. Sci.* 94:3645–3654.
- Tenghe, A.M.M., A.C. Bouwman, B. Berglund, E. Strandberg, J.Y. Blom, and R.F. Veerkamp. 2015. Estimating genetic parameters for fertility in dairy cows from in-line milk progesterone profiles. *J. Dairy Sci.* 98:5763–73.
- Tenghe, A.M.M., A.C. Bouwman, B. Berglund, E. Strandberg, D. de Koning, and R.F. Veerkamp. 2016b. Genome-wide association study for endocrine fertility traits using single nucleotide polymorphism arrays and sequence variants in dairy cattle. *J. Dairy Sci.* 99:5470–5485.
- Wientjes, Y.C.J., P. Bijma, R.F. Veerkamp, and M.P.L. Calus. 2015. An Equation to Predict the Accuracy of Genomic Values by Combining Data from Multiple Traits, Populations, or Environments. *Genetics*.
- Wiggans, G.R., G. Su, T.A. Cooper, U.S. Nielsen, G.P. Aamand, B. Guldbrandtsen, M.S. Lund, and P.M. VanRaden. 2015. Short communication: Improving accuracy of

5 Progesterone profiles and bull fertility

Jersey genomic evaluations in the United States and Denmark by sharing reference population bulls. *J. Dairy Sci.* 98:3508–3513.

Zumbach, B., H. Jorjani, and J. Dürr. 2010. Brown Swiss genomic evaluation. *Interbull Bull.* 44.

6

General discussion

6.1 Introduction

The research in this thesis on exploring the use of milk progesterone (P4) measures to improve selection for fertility in dairy cows can be divided into two main areas. The first focused on the genetic aspects of fertility as measured using P4, which have been dealt with in chapters 2 and 3. Chapter 2 investigated the use of in-line measured milk P4 concentrations to define endocrine fertility traits and estimate genetic parameters. In chapter 3, a study on identifying genomic regions associated to endocrine fertility traits, using single nucleotide polymorphisms (SNP) panels as well as and imputed whole genome sequence variants was performed. The second part focused on the use of endocrine fertility traits in genomic prediction, to improve genomic selection for fertility. On the one hand, classical fertility traits derived from calving and insemination data are available for numerous cows for national genetic evaluations. On the other hand, the availability of endocrine fertility traits presents the opportunity to improve accuracy of predicting genomic breeding values of cows' fertility by combining endocrine and classical fertility traits in multi-trait genomic predictions (chapter 4). Also, an important question is how many cows with endocrine fertility records in the training population will be required to improve the accuracy of bulls' genomic breeding values (chapter 5).

In each of the chapters, the main results have already been discussed. The general discussion will focus on aspects related to breeding for fertility in dairy cows, optimizing accuracy of genomic prediction for endocrine fertility traits, practical integration of endocrine fertility traits (and novel traits in general) in routine genetic evaluations, and prospects for future research will be given.

6.2 Breeding for fertility

This thesis specifically looked at estrous related traits that reflect the ability of a cow to return to cyclicity after calving (see Figure 1, chapter 1), and how they can be used in genetic evaluations to improve selection for fertility. Different traits derived from different sources of information have been proposed as measures of the ability of a cow to resume cyclicity postpartum. In this thesis, we investigated endocrine fertility traits derived from milk P4 concentrations and classical fertility traits derived from insemination and calving data. Endocrine fertility traits are expected to be more objective measures that are closer to the physiology underlying fertility than the classical traits. For genetic evaluations, the most commonly used traits are classical fertility traits like calving to first service (CFS) and calving interval (CI). Although the

use of classical fertility traits for breeding has been facilitated by their ease of recording on a large scale and at feasible cost, the drawbacks are that classical traits are highly influenced by farm management decisions, are available late in life (e.g., CI), and have low heritability. Reported heritability estimates for CFS range from 0.05 to 0.10 (Roxström et al., 2001; Andersen-Ranberg et al., 2005; Sun et al., 2009; Berry et al., 2012), though estimates for CFS in this thesis were in the upper bound. The estimated heritability of CI in this thesis was low (0.03 to 0.05), and in agreement with previous studies (Wall et al., 2003; Berry et al., 2012). The low heritability of classical fertility traits makes selection for fertility less effective. Studies of physical activity derived traits e.g., the interval from calving to the first sign of high activity (CFHA) have reported higher heritability, ranging from 0.12 to 0.18 (Løvendahl and Chagunda, 2009; Ismael et al., 2015). Though CFHA has been suggested as a measure for resumption of ovarian activity, Ismael et al., (2016) showed that a drawback is that activity monitor derived fertility traits are seasonally sensitive.

The advantage of endocrine fertility traits is that they are less biased by farm management decisions and have higher heritability. In this thesis, several endocrine fertility traits were derived, but not all the traits were heritable. The traits with highest heritability were commencement of luteal activity (C-LA), and proportion of samples in luteal activity (PLA). The heritability of C-LA ranged from 0.12 to 0.14, whereas that of PLA ranged from 0.12 to 0.15. Reported heritability estimates for C-LA or transformed C-LA range from 0.16 to 0.30 (Veerkamp et al., 2000; Royal et al., 2002; Petersson et al., 2007), whereas Petersson et al., (2007) reported a heritability estimate of 0.30 for PLA. Thus, heritability estimates of C-LA and PLA in this thesis were in the lower bound of those reported in literature. Another advantage of endocrine fertility traits is that they more directly reflect a cow's reproductive physiology than classical traits. For example, in some breeding programs, CFS is used in a fertility index as an indirect indicator of calving to first ovulation, but C-LA is a more objective indicator of first ovulation because it more directly reflects a cow's physiology (Bulman and Lamming, 1978; Lamming and Darwash, 1998; Darwash et al., 1999). In addition, for animals with delayed insemination, CFS records will be available much later, whereas C-LA is independent of farm management decisions. Thus, C-LA is a better trait to use in genetic evaluations.

Whether a trait is suitable for use in breeding depends not only on that it has substantial genetic variation, but also its genetic correlation with other traits. As discussed in 6.2.2, results in chapter 2 demonstrated that the genetic correlation of endocrine fertility traits with milk production traits were considerably lower than

those observed between milk yield and classical fertility traits. The genetic correlation of C-LA was 0.18 with milk yield, and there was a favorable correlation of C-LA with fat yield (- 0.12) and with protein yield (- 0.04). This is another advantage of using endocrine traits compared with the classical traits. In this thesis, the genetic correlation between C-LA and CFS was positive (0.37 to 0.58), which indicates that selection for shorter C-LA will also reduce interval from CFS. Royal et al. (2003) reported a genetic correlation of 0.53 between CFS and C-LA. Furthermore, they found genetic correlations between a long C-LA and a long CI (0.39). In this thesis, the genetic correlation between C-LA and CI varied from 0.26 to 0.31. Ismael et al., (2015) also found a positive correlation between CFHA and CFS (0.96). This indicates that selection for a shorter interval from calving to resumption of ovarian activity would benefit overall fertility.

6.2.1 New tools and new breeding goals to improve selection for fertility in dairy cows

New breeding tools like genomic selection and in-line recording technologies like the Herd Navigator provide an opportunity to implement endocrine fertility traits in genetic evaluations to improve fertility in dairy cows. However, to benefit from the added value of endocrine traits in genetic evaluations, breeding goals for fertility will need to be redefined. For a long time, the use of endocrine fertility traits in genetic evaluations has been hindered by the fact that it is laborious and costly to measure sufficient milk P4 records per cow on a large scale. Also, all research on the use of endocrine fertility traits has focused on application in breeding schemes that generally target a classical breeding goal (e.g., cows with shorter calving intervals and shorter intervals from calving to first service). However, now that measuring sufficient P4 records on a large scale has been facilitated by in-line technologies, we now realize that endocrine traits will be useful for breeding only if they are important for the breeding goal. As shown in chapter 5, if the question is “can we improve selection for the biological breeding goal for fertility?”, that is, where the breeding goal trait is an endocrine trait like C-LA, then it is valuable to include endocrine fertility traits in genetic evaluations. However, if the question remains to improve the classical breeding goal for fertility (e.g., CI or CFS), there is little to no benefit of including endocrine traits in addition to the cheap classical traits in genetic evaluations. This means that if breeding programs want to make valuable use of endocrine fertility traits to improve selection for fertility, they will need to defining new breeding goals for fertility that include the endocrine traits.

A disadvantage of using the endocrine traits in progeny testing schemes is that it is unrealistic, as progeny testing schemes usually require a large number (100 - 150) of daughters with phenotypes of each selection candidate bull, and these daughters are often milked in many different herds. Another drawback of progeny testing is that the long generation interval between phenotyping and proving bulls hampers genetic gain. Genomic selection has become the standard in cattle breeding because it increases the potential genetic gain by up to 80% due to reduced generation interval (Schaeffer, 2006). Genomic selection has the added benefit that it also disconnects the phenotype recording in a training population from the selection of the candidates. This new feature of genomic selection provides a great opportunity for novel traits such as endocrine traits, because a trait can be selected as soon as several thousand of animals are recorded for the trait (Misztal, 2011). For endocrine fertility traits, deterministic predictions with the equation in chapter 5 show that for within population genomic prediction of C-LA with heritability of 0.13, and considering 1,566 for the effective number of chromosomes (Me), 15,000 animals in a cow training population with own C-LA records are needed to achieve 0.74 accuracy; to achieve 0.84 accuracy of prediction, 30,000 animals are required. Although these numbers are still high, it is easier to achieve than collecting records on progeny of every test bull. Also, across country collaboration can help. This is especially beneficial when one country has very limited number of animals in the training population. However, as shown in chapter 5, the accuracy of prediction when combining training populations across countries is largely influenced by the genetic correlation between the countries for the target trait. For example, in chapter 5, when the genetic correlation between Sweden and the Netherlands was high, fewer animals from the Netherlands were needed to obtain the same accuracy of genomic breeding values in Sweden than with lower genetic correlations. That is, with 200 animals from Sweden and a genetic correlation of 0.5 between countries, 15,000 animals from the Netherlands were needed to achieve 0.31 accuracy of predicting C-LA, but when the correlation was 0.9, only 3,000 animals from the Netherlands were needed to achieve the same accuracy.

Furthermore, recording strategies for endocrine fertility traits can be optimized to improve accuracy of genomic prediction. For example, in chapter 5, it was shown that it is more important to have more animals with phenotypes than more lactations per animal in the training population. Also, to maximize accuracy of genomic prediction, it is important to optimize the design of the training population. An optimal design of the training population should maximize the relationships between evaluated animals and animals in the training population, and minimize

relationships between training animals (Habier et al., 2007; Meuwissen, 2009; Pszczola et al., 2012). Because animals in the training population do not have to be progeny or relatives of evaluated bulls, this conveniently allows for optimizing recording strategies for endocrine fertility traits as shown in chapter 5, while at the same time maintaining an optimal design of the cow training population. Optimizing the training population can be done by choosing animals for phenotyping based on their genomic relationship with evaluated animals.

Therefore, genomic selection and in-line recording technologies provide a great opportunity to reconsider the use of endocrine fertility traits to further improve selection for fertility. However, investing in novel traits for a cow training population should only be done when the trait is in the breeding goal.

6.2.2 Association of fertility traits with milk yield and genetic progress

There is accumulating evidence of the antagonistic effect of increased milk yield on fertility in dairy cattle, for a review see Lucy, (2001), Veerkamp and Beerda, (2007), and Walsh et al., (2011). For several decades, breeding objectives in dairy cattle focused mainly on increasing milk yield and its components, with the exception of the Nordic countries that included fertility since the 1970s. The unfavorable genetic correlation between milk yield and fertility traits meant that fertility decreased with increasing milk yield, which led to a long-term decline in fertility. In recent years, selection indices worldwide have changed to a more balanced breeding approach that includes longevity, udder health, and fertility (Miglior et al., 2005, 2012). This has in turn reversed the undesirable genetic trend for fertility, while genetic progress for production has continued at about the same rate as before. For example, Figure 6.1 shows the genetic trend in Holstein Friesian cows from the Netherlands, from 1990 to 2014. In the Netherlands, fertility traits were included in the total merit index from 2006. As can be seen in Figure 1, the genetic trend has reversed after the inclusion of fertility in the breeding goal. This indicates that it is possible to improve fertility genetically, without ceasing selection for milk yield, as milk production traits still have a major impact on the revenue of each farm, and are still worth improving.

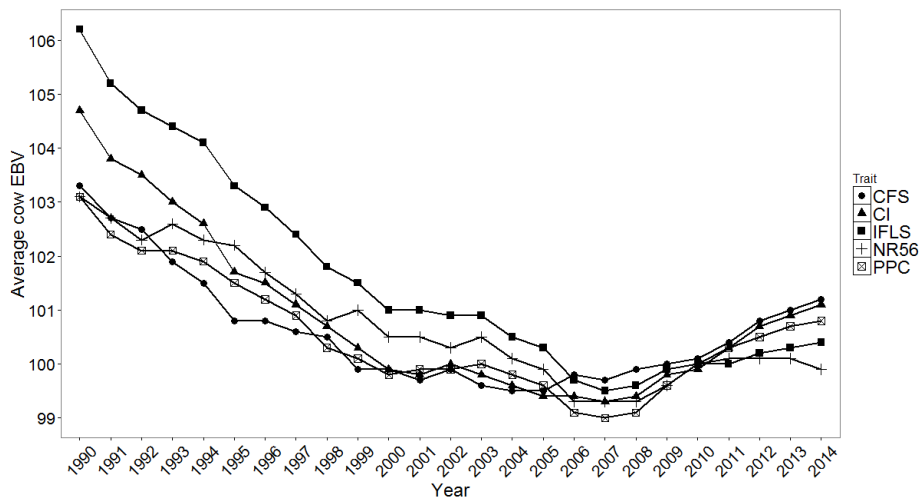


Figure 6.1 Genetic trend of fertility for Holstein Friesian cows from commercial farms in the Netherlands. CFS = calving to first service, CI = calving interval, IFLS = interval from first to last service, NR56 = non-return rate within 56 days, PPC = percentage of pregnant cows, EBV = estimated breeding value. Data from the Dutch herd book (https://crvnl-be6.kxcdn.com/wp-content/uploads/2016/09/gen_trend_koe_nl_20160816.pdf).

The traits currently used for fertility in the total merit index of the Netherlands are classical fertility traits derived from insemination and calving data. The current rate of genetic progress for fertility, as seen in Figure 6.1, is influenced by the unfavorable genetic correlation between milk production and classical fertility traits. In this thesis, the genetic correlation varied from 0.07 to 0.24 between milk production traits and CFS, and from 0.56 to 0.69 between milk production traits and CI, while phenotypic correlations of milk production traits with CFS and CI varied from 0.13 to 0.21. However, in chapter 2, the genetic correlations of endocrine fertility traits with milk production traits were considerably lower than those observed between milk yield and classical fertility traits. The genetic correlation of C-LA was 0.18 with milk yield, and there was a favorable correlation of C-LA with fat yield (- 0.12) and with protein yield (- 0.04). In addition, the phenotypic correlations of the endocrine fertility traits with milk production traits were close to zero (0.01 to 0.07). This indicates that novel fertility traits that are not (or less) unfavorably correlated with milk yield can improve genetic progress for fertility further. Higher genetic correlation estimates of C-LA with milk yield have been reported, but all unfavorable. Veerkamp et al., (2000) reported an unfavorable genetic correlation of 0.51 between C-LA and milk yield, and Nyman et al., 2014 found a correlation of 0.45. Although further research is needed to confirm the low genetic correlation between in-line

endocrine fertility traits and milk production traits, this thesis shows that the genetic progress of fertility can be improved by considering C-LA as a breeding goal trait in genetic evaluations, or index trait in a fertility index.

Another approach of improving the genetic progress for fertility is to put a higher economic weight (breeding goal weight) on fertility to get faster genetic improvement, compared with the other traits in the breeding. Miglior et al 2005, analyzed national selection indices for 15 major dairy countries and reported that average relative emphasis was 59.5% for production, 28% for durability, and 12.5% for health and reproduction across all countries. In another study, Miglior et al., (2012) reported interesting changes in relative emphasis on production traits in selection indices of four countries taken as examples. This suggest that perhaps we could put more emphasis on fertility traits to improve genetic progress. However, such a scenario would reduce genetic progress in milk yield due to its unfavorable correlation with fertility, but again this higher emphasis on fertility could be put on endocrine fertility traits like C-LA, as it is less unfavorably correlated with milk production traits.

6.3 Optimal use of genomic information in selection for fertility

The main interest in genome-wide association studies (GWAS) in dairy cattle breeding is to find markers to improve the accuracy of predicting breeding values and to increase the understanding of the genetic control of economically important traits. Identifying the genes that affect traits such as fertility will also provide insight into the likely effect of selection on these mutations on other traits in the breeding goal. With the development of methods that allow to perform genomic prediction based on a large number of markers (Meuwissen et al., 2001), and the availability of commercial SNP chips, genomic selection has become the standard tool for animal and plant breeders. The accuracy of genomic prediction is influenced by the number of phenotypes in the training population used to derive the prediction equation, the heritability of the trait, the effective population size, the size of the genome, the density of markers, and the genetic architecture of the trait, in particular the number of loci affecting the trait and distribution of their effects (Daetwyler et al., 2008; Meuwissen, 2009).

In this thesis, I investigated the heritability of fertility by defining different endocrine fertility traits in chapter 2. In chapter 3, the genetic architecture of fertility was

investigated by identifying quantitative trait loci (QTL) or genomic regions genes associated to fertility traits using SNP data. Using sequence variants, target QTL regions were also fine-mapped in chapter 3, and several candidate genes were identified. Chapter 4 evaluated empirical accuracy of predicting fertility using endocrine and classical fertility traits in multi-trait genomic prediction models, and in chapter 5 I evaluated the impact of the size of a cow training population with endocrine traits on accuracy of genomic prediction. In this section, I discuss the impact of the investigated factors on the accuracy of genomic prediction for fertility, and how to optimally use genomic information to improve accuracy of genomic prediction for endocrine traits.

The properties of QTL that control a trait like allele frequency spectra of QTL, and distribution of QTL effects are key factors that determine the accuracy of genomic predictions (Wientjes et al., 2015a). When the minor allele frequency (MAF) of QTL is on average lower than that of SNP marker, the accuracy of genomic prediction is reduced (Daetwyler et al., 2013; Wientjes et al., 2015a). This is because, a lower MAF of QTL than for SNPs, results in decreased strength of linkage disequilibrium (LD) between QTL and SNP markers (Khatkar et al., 2008; Yan et al., 2009; Wientjes et al., 2015b), hence reducing the proportion of the genetic variance captured by the SNP markers. The MAF of QTL underlying complex traits is expected to be lower than the MAF of SNP markers (Goddard, 2009; Yang et al., 2010; Kemper and Goddard, 2012), therefore, it is highly likely that not all the genetic variance can be captured by SNP marker panels in real data. This also indicates that, there is a probability of underestimating the heritability of complex traits using SNP markers.

In chapter 3, a genome-wide association study (GWAS) for endocrine and classical fertility traits was performed with 85,485 SNPs. The GWAS identified 17 QTL regions for 6 endocrine fertility traits and 3 regions for one classical trait. The average proportion of variation explained by SNPs in the QTL regions identified for each trait are in Figure 6.2. The proportion of variation was calculated as in Pryce et al., (2010). That is, the proportion of variation for any SNP can be calculated simply as, F/a , where F is the F statistic for that SNP based on number of animals (a) used in analysis. In the absence of any real effect, F statistics have an expected value of 1, so an unbiased estimate of the proportion of variation can be made by $(F-1)/a$. This estimate of the proportion of variation (R^2) was averaged over all N SNP that were statistically significant in each identified QTL region for each trait by calculating $R^2 = \left(\frac{\sum_{i=1}^N F}{N} - 1 \right) \times 100\% \times \frac{1}{a}$. The proportion of genetic variance explained by the QTL

was 0.70% for CLA, 0.72% for PLA, 0.72% for LA60, 0.71% for LPL, 0.60% for IOI, 0.77% for CLAFS, and 0.74% for CFS. In general, the variance explained by the QTL was very minimal, and there were no QTL with large effects. Also, the average MAF of the QTL detected was 0.27. This suggest that the remaining genetic variance of endocrine fertility traits is influenced by QTL with low MAF that may not have been detected in the GWAS due to incomplete LD between the SNP markers and underlying QTL. Consequently, genomic prediction for endocrine fertility traits will be more beneficial with genomic best linear unbiased (GBLUP) models that assume that all SNP effects are drawn from the same distribution and explain equal amount of genetic variance, than Bayesian variable selection models that accommodate for SNPs explaining a larger part of the genetic variance compared to other SNPs. For instance, in chapter 4, empirical accuracies of genomic prediction for endocrine fertility traits were estimated with GBLUP, and theoretical accuracies were estimated with the prediction equation of (Daetwyler et al., 2008, 2010) which was derived with the assumptions of a GBLUP model. The theoretical accuracies were in line with the empirical estimates. Therefore, based on the genetic architecture of endocrine fertility traits, I do not expect an improvement in accuracy of genomic prediction from approaches such as Bayesian variable selection.

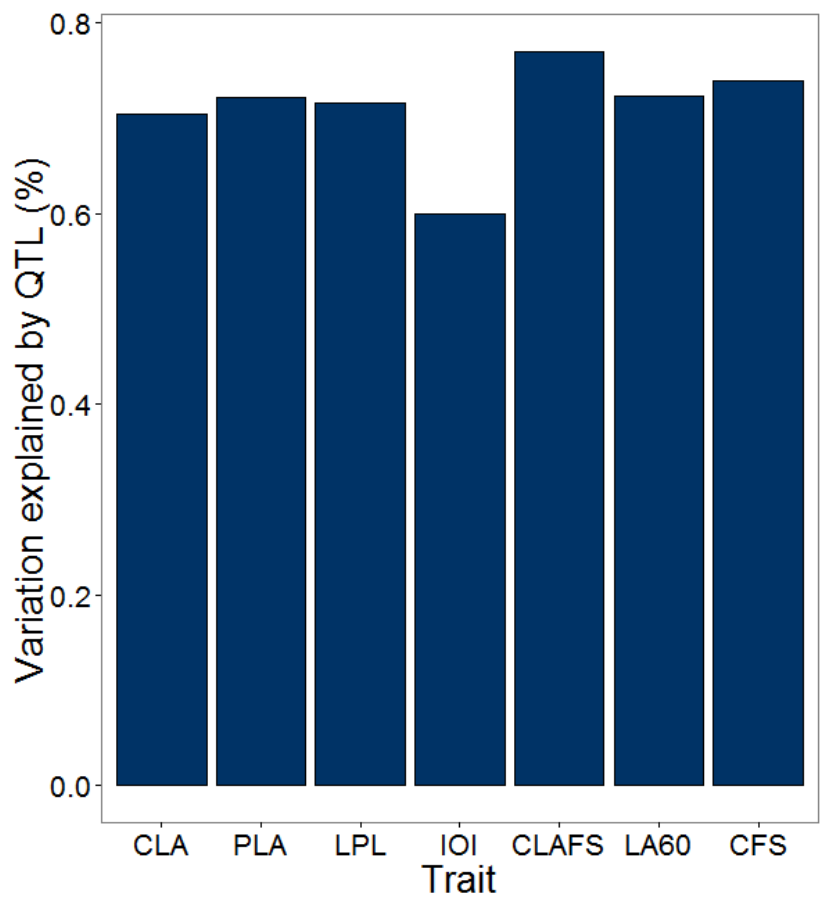


Figure 6.2 Average genetic variation explained by SNP markers in QTL regions identified in a genome-wide association study using 85,485 SNPs and 2,447 cows for endocrine and classical fertility traits in Holstein. CLA = interval from calving to commencement of luteal activity; PLA = proportion of samples in luteal activity during the first 60 days in milk; LPL = length of first luteal phase; IOI = length of first inter-ovulatory interval; CLAFS = interval from commencement of luteal activity to first service; LA60 = occurrence of luteal activity during the first 60 days in milk; CFS = interval from calving to first service

One option to increase the likelihood of identifying causal mutations is to increase marker density and (or) number of genotyped and phenotyped animals. Developments in genotyping technology have resulted in a reduction of costs, enabling the production of commercial high-density (HD) SNP chips (e.g., Illumina Bovine HD 770k SNP chip). With more animals genotyped, which increases the sample size, and with the genome more densely covered with markers, which leads

to a smaller distance between the SNP and the causative mutation, a more precise detection of QTL is expected. The ultimate level of genotypic information though is sequence data. Sequencing tries to determine the order of all nucleotides of the DNA of a given organism. Therefore, sequence data should contain the causative mutations of a trait. Hence a GWAS using sequence data is expected to find the causative mutation (Meuwissen and Goddard, 2010). There have been efforts to increase the number of sequenced animals (Daetwyler et al., 2014), but sequencing a large number of animals for GWAS is still expensive. Therefore, imputation from lower density genotypes to whole genome sequence using a sequenced training population offers a good alternative. The approach that has been taken in chapter 3 was to perform a GWAS with a low density (e.g., here 85k) SNP chip panel, and then focus on the identified peaks, performing a region-wise association study (RWAS) using imputed sequence data (e.g., Höglund et al., 2014; Sahana et al., 2014; Wu et al., 2015). In chapter 3, significant QTL regions from the GWAS with 85k SNP were fine-mapped for endocrine fertility traits using imputed sequence variants. The RWAS was able to refine the QTL regions from the GWAS, but it was not possible to identify the causative mutation, mainly because of long-range LD that exist in cattle due to low effective population size and strong selection. Similar observations have been reported in previous studies (Höglund et al., 2014; Wu et al., 2015). Another factor that might be hampering identification of the causative mutation is that imputation is not 100% accurate, especially for rare variants and small training populations. The RWAS was able to identify several candidate genes associated to endocrine fertility traits that can help to learn more about the genetic architecture and underlying biology of fertility. In addition, variants significantly associated to endocrine fertility traits detected in the RWAS can be integrated in genomic selection to improve accuracy of prediction. As these significant regions on the genome are continually being found and described, it is of interest to integrate the significant markers in genomic evaluations. This integration is relevant because while the causative mutations are not detected, these significant markers provide knowledge regarding the genetic architecture of the trait. Even though the effects found are not large, they might add to the prediction accuracy, and thus should be exploited. Although there is ongoing research to develop methods that efficiently make use of sequence variants in genomic predictions, at the moment, there is no consensus on how to benefit from the use of sequence information in genomic predictions. Wientjes et al., (2015a) demonstrated in a simulation study that adding causal QTL to SNP panels increased accuracy of genomic prediction, with a much larger increase achieved when the initial number of SNPs is lower. Because the significant variants from the RWAS are expected to be in LD with the underlying causal mutation, an

option would be to add the identified variants to a low-density SNP chip to improve accuracy of genomic prediction. Brondum et al., (2015) investigated this option in cattle wherein they performed a GWAS in Nordic cattle for three separate breeds using sequence variants for different traits. They then selected QTL and three to five variants to tag each QTL and combined 1,623 variants with a 54k SNP panel. Their results showed that the reliability of genomic prediction was improved when using the combined SNP panel compared to only the 54k panel, with the largest gains of up to 5 percentage points for production traits and lowest of 0.5 percentage point increase for classical fertility traits. Their results also showed that when using a Bayesian model accuracies were generally higher with only 54k data compared with the genomic BLUP. In chapter 4 of this thesis, I showed that there was more predictive ability for endocrine fertility traits than classical fertility traits. This suggests that combining the significant markers from the RWAS for endocrine fertility traits to a low SNP panel like 50k SNP might improve the accuracy of prediction further, compared to the 0.5 percentage point increase for classical fertility traits.

In general, the accuracy of genomic selection can be improved by spending time on trait definition to improve heritability, improving statistical modelling of the trait, or by increasing the size of the training population. Although all these factors are important to improve accuracy of prediction, for endocrine fertility traits and novel traits in general, the critical factor that needs to be improved is the size of the training population.

6.4 Practical integration of novel traits in routine genetic evaluations

Chapter 5 showed that for genomic prediction, it is more important to have phenotypes from more cows than more lactations with phenotypes per cow in the training population. That is, optimizing recording strategies for endocrine fertility traits will maximize the accuracy of genomic prediction for fertility. This is true for endocrine fertility traits, but will generally apply to novel traits that are expensive or difficult to measure, e.g., feed efficiency and methane emission. For endocrine fertility traits, some important questions that arise are “can we use the Herd Navigator for the national breeding goal?”, “who should pay for the Herd Navigator?”, and “should farmers be subsidized for collecting extra records?” There is no single answer to these questions, as this depends on several factors. In this

section, I will discuss the pros and cons of using the Herd Navigator in national breeding evaluations, phenotyping strategies for novel traits, and in general, challenges for integrating novel traits in genetic evaluation schemes.

6.4.1 Use of the Herd Navigator in genetic evaluations

In this thesis, we specifically looked at estrous related fertility traits derived from in-line milk P4 levels measured with the Herd Navigator, and how to include these traits in genetic evaluations. One downside of using the Herd Navigator for the national breeding goal, is that it is not realistic for progeny testing schemes. Also, chapter 5 showed that it is only worthwhile to invest in endocrine traits if the endocrine traits are themselves breeding goal traits. A key requirement for the recording of data is the motivation of the stakeholders involved. That is, recording needs to have benefits beyond genetic improvement, and the additional effort required for recording must result in added value, also for the farmers. The added advantage of the Herd Navigator is that farmers can also use the Herd Navigator for other management purposes like estrus detection, and monitoring of mastitis and ketosis. However, there is a lot of competition with other fertility systems e.g., fertility systems based on heat detection and movement like physical activity monitors; these other systems are generally cheaper than the Herd Navigator. Furthermore, fertility traits based on physical activity monitors like interval from calving to first high activity have been shown to have substantial genetic variation, and have the potential to improve genetic selection for fertility (Løvendahl and Chagunda, 2009; Ismael et al., 2015). However, comparisons of pedometer measures with progesterone-determined estrus has shown that activity monitors cannot detect silent behavioral estrus (Løvendahl and Chagunda, 2016). Studies show that body condition score has a favorable relationship with fertility (e.g., Pryce et al., 2001; Berry et al., 2003) and is growing in popularity as a novel predictor for fertility (Fogh et al., 2013), but body condition score cannot be used for estrus detection. Research on the use of mid-infrared predicted fatty acids as indicator traits for fertility (Bastin et al., 2012) and of mid-infrared for pregnancy status testing is also underway (Gengler, 2014). None the less, such novel traits are not designed to detect estrus, except for the activity meters, and P4 remains the ‘gold standard’ for estrus detection.

For the farms that have the Herd Navigator, P4 levels are usually recorded until pregnancy is established, and these P4 measure were available in the data used in this thesis. Therefore, these profiles could also be used to derive endocrine fertility traits that reflect the ability of a cow to conceive following insemination, and become

pregnant. This will be useful, as these endocrine traits might be more informative and objective than some of the classical fertility traits currently used to reflect conception and pregnancy in dairy cows e.g., non-return rate (see Figure 1, chapter 1). For example, perhaps describing the length of the last luteal phase before pregnancy might be a more informative trait that reflects the ability of a cow to get pregnant after insemination. The last luteal phase before pregnancy might be more informative because in general, after ovulation, the luteal phase is the period during which the endometrium (uterine lining) is prepared for implantation of an eventual fertilized egg (or embryo). That is, estrogen and progesterone together promote thickening of the endometrium. The length of the last luteal phase before pregnancy might be informative on whether a pregnancy is successful, as shorter luteal phase lengths might mean that there is not enough time for the endometrium to develop, which might contribute to embryo death and pregnancy failure. The last luteal phase length before pregnancy might be a more objective measure for the ability of a cow to get pregnant after insemination. Yet another option would be to describe lengths of all luteal phases before pregnancy, as animals with regular luteal phase lengths might have a higher probability of getting pregnant. Also, the interval from commencement of luteal activity to first service was investigated in chapter 2, but perhaps it will be more useful to look at the interval from C-LA to the last service before pregnancy, as this might be a more objective measure of the interval from calving to conception. For future studies, it would be interesting to consider defining and investigating this group of endocrine fertility traits that reflect the ability of a cow to conceive and get pregnant. Therefore, the Herd Navigator also makes it possible to exploit P4 measures further to improve fertility, because ideas that were considered unrealistic in the past may become feasible in the near future.

6.4.2 Phenotyping strategies for novel traits

A major benefit of genomic selection in dairy cattle breeding is that it is no longer required to measure a large number of phenotypes from progeny groups for each male selection candidate. This significantly decreases generation intervals and increases genetic gain per year for all breeding goal traits (Pryce et al., 2016; Lund et al., 2011). For traits that have been part of the breeding goal, current training populations consist of bulls with highly accurate phenotypes based on the average phenotype of hundreds of daughters (for reviews see: Hayes et al., 2009; Calus, 2010) and may contain up to as much as 16, 000 bulls (Lund et al., 2011). For novel traits, depending on the cost of a single phenotypic measurement, composing

training populations may only be feasible for a few thousand cows. Therefore, there is the need to optimize phenotyping strategies for novel phenotypes.

Genomic predictions with novel traits can be trained within a set of animals, representative for the whole population and then applied in the general population. In sheep or beef cattle breeding, selection for carcass and meat quality traits is implemented in a centralized approach, in which test animals are housed in information nucleus herds, and phenotypes are recorded on the breeding animals themselves (i.e., own phenotypes). One advantage of an information nucleus is that animals are well identified and similarly managed, and fixed effects are fully recorded. In its current form, dairy cattle breeding companies have contracts with farmers, where intensive recording of phenotypes is performed. However, this system results in lots of records but with lots of lactations per animal. In this thesis, I showed that for genomic prediction of novel traits, it is more important to have more cows with phenotypes than more lactations per cow in the training population (chapter 5). Perhaps the dairy industry could learn from phenotyping structures established for breeding in other species. That is, the dairy industry should think of optimizing phenotyping structures by establishing nucleus farms for phenotype recording, especially for novel phenotypes that need new technologies for optimal recording.

6.4.3 Phenotyping farms and labs

Advances in technology will have an impact on the future definition and availability of phenotypes. The challenge for dairy breeding companies is to start thinking of ways to optimize phenotyping strategies, which is something that has been lacking behind in the dairy industry compared to the meat industry. Some dairy companies (http://www.eaap.org/Annual_Meeting/2015_warsaw/S22_12.pdf), have currently set up partner herds called “DataPlus” farms for recording specific phenotypes. A next step for dairy breeding companies would be to optimize the phenotyping structure by establishing phenotyping labs where animals are phenotyped in batches for one lactation and dispatched. This will make space for the next batch of cows and leads to phenotypes recorded on as many animals as possible. Based on the results of chapter 5, I argue that this is an optimal way of collecting as many phenotypes on novel traits as possible to set up a training population for genomic prediction. In the case of endocrine traits, the Herd Navigator would be a great investment for phenotyping labs. Another advantage of having phenotyping labs is that animals to be measured can be selected in terms of their relatedness to important bulls (or offspring of the sires of the young bulls) so that there is always a close relationship

between the cow training population and the candidate bulls, as this will increase the accuracy of genomic prediction. Studies have shown that a close relationship between evaluated animals and animals in the training population is expected to give more reliable prediction (Habier et al., 2007; Meuwissen, 2009; Pszczola et al., 2012). These relationships are especially important for small training populations (Wientjes et al., 2013). That is, optimally, all evaluated animals should have at least some closely related animals in the training population. Therefore, it will be more interesting to phenotype heifers (i.e., first parity cows) for novel traits than later parity cows. It is also important to have an idea of genetic correlation estimates between endocrine fertility traits across parities, but that requires large datasets, which were not available in this thesis. However, genetic correlation estimates of classical fertility traits across lactations can give an idea of what to expect from endocrine traits. Reported genetic correlation of classical fertility traits between first and second lactation vary between 0.7 and 0.9 (Roxström et al., 2001; Haile-Mariam et al., 2003), and are 0.9 or higher between the second and third lactation. Based on these, I do not expect a large difference for the correlation of endocrine fertility traits across lactations. To enable the immediate use of endocrine fertility traits in genetic evaluations, these correlation estimates for classical traits could be used as pointers rather than waiting to first obtain correlation estimates of endocrine traits. Therefore, rather than the partner farms, maybe the dairy breeding companies should be thinking about phenotyping labs.

On the one hand, investing in phenotyping labs with the Herd Navigator will optimize the recording strategy for endocrine fertility traits in dairy cattle, but this investment will not be worthwhile if the only traits recorded are fertility traits. On the other hand, if investments for recording other novel traits e.g., feed intake and methane emission are included, this might become a realistic option for the future, where breeding companies establish high-tech farms with equipment for measuring different types of data, and where animals are kept for one lactation only. As an example, to optimize phenotyping strategies, a breeding company could invest in 10 phenotyping labs where all novel phenotypes of interest are collected. Heifers are bought, genotyped, phenotypes are recorded for one lactation, and then animals are sold. Another advantage of phenotyping labs is that all traits can be recorded on the same animals. The number of phenotyping labs needed to optimize phenotyping can be decided based on predicted accuracy and expected size of the training population, as shown in chapter 5. A critical question that arises concerns genotype-by-environment (G X E) interactions. With limited training populations in a given environment, selection will be more sensitive to G X E interactions than with

traditional progeny tests where daughters of candidate bulls are distributed over a large number of herds. Therefore, it is recommended to establish the phenotyping labs over a range of environments.

6.5 Concluding remarks

The research in this thesis investigated the use of in-line milk progesterone concentrations to define novel fertility traits that can be used in genetic evaluations to improve genomic selection for fertility in dairy cows. This was inspired by the availability of new in-line recording technologies like the Herd Navigator, and the opportunity due to genomic selection, that makes it feasible to implement novel traits in breeding programs for dairy cattle. In the first part of this thesis, I show how in-line milk progesterone records can be used to define several heritable endocrine fertility traits, and describe research that detected genetic markers significantly associated with endocrine fertility traits, fine-mapped target QTL regions and identified potential candidate genes. The less unfavorable genetic correlation of endocrine fertility traits like and their heritability makes them better alternatives for use in genetic improvement of fertility than classical fertility traits like CFS. I expect that GWAS will continue to be performed because they provide scientifically relevant results, especially with greater statistical power when more animals will be sequenced or genotyped using high density SNP chips. With more markers, the physical distance between markers and the causative mutation will be shortened, allowing QTL regions to be fine-mapped. However, finding the causal mutation will require more than just a GWAS using denser genotyping or sequence data. Linkage disequilibrium plays a major role in GWAS, and one may require additional evidence to distinguish associated variants. The results of GWAS with sequence variants can be used to augment low-density SNP chip panels like 50k, to increase the accuracy of genomic prediction.

In the second part of this thesis, I describe genomic prediction using endocrine fertility traits in dairy cattle. This is the first study on the use of endocrine traits in genomic prediction. I have shown that endocrine fertility traits have more predictive ability than classical fertility traits, and that combining endocrine fertility traits in multi-trait genomic prediction can substantially increase the accuracy of genomic prediction. Because endocrine traits are novel fertility traits, there is bound to be a limited training population size for these traits. I showed that for prediction, it is more important to phenotype more animals than more lactations per animal in the

training population. Therefore, breeding companies can optimize phenotyping strategies for endocrine fertility traits (and novel traits in general) by establishing phenotyping labs where animals are phenotyped in batches for one lactation and dispatched, as leads to phenotypes recorded on as many animals as possible. However, as shown in this thesis, investing in a cow training population for endocrine fertility traits (and novel traits in general), is relevant only when the novel trait is in the breeding goal.

In the final discussion, I show that new tools like, genomic selection and in-line recording technologies provide a great opportunity to reconsider the use of endocrine fertility traits to further improve selection for fertility, and should be explored further. These tools make it possible to optimize phenotype recording and increase the size of the training population, which is the main limiting factor of improving accuracy of prediction with novel traits. Also, the availability of more and more new electronic technologies for farm management and phenotype recording means that in the future, dairy farming might shift from the traditional system to a more electronic technology oriented farming system. Therefore, the dairy industry might be faced with challenges on how to handle/integrate new phenotypes from these new technologies in to genetic evaluations. The results from this thesis can server as pointers.

References

- Andersen-Ranberg, I.M., G. Klemetsdal, B. Heringstad, and T. Steine. 2005. Heritabilities, genetic correlations, and genetic change for female fertility and protein yield in Norwegian dairy cattle. *J. Dairy Sci.* 88:348–355.
- Bastin, C., D.P. Berry, H. Soyeurt, and N. Gengler. 2012. Genetic correlations of days open with production traits and contents in milk of major fatty acids predicted by mid-infrared spectrometry. *J. Dairy Sci.* 95:6113–6121.
- van den Berg, I., D. Boichard, and M.S. Lund. 2016. Sequence variants selected from a multi-breed GWAS can improve the reliability of genomic predictions in dairy cattle. *Genet. Sel. Evol.* 48:83.
- Berry, D.P., J.W.M. Bastiaansen, R.F. Veerkamp, S. Wijga, E. Wall, B. Berglund, and M.P.L. Calus. 2012. Genome-wide associations for fertility traits in Holstein-Friesian dairy cows using data from experimental research herds in four European countries. *Animal.* 6:1206–15.

- Berry, D.P., F. Buckley, P. Dillon, R.D. Evans, M. Rath, and R.F. Veerkamp. 2003. Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows. *J. Dairy Sci.* 86:2193–2204.
- Brøndum, R.F., G. Su, L. Janss, G. Sahana, B. Guldbrandtsen, and D. Boichard. 2015. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. *J. Dairy Sci.* 98. doi:10.3168/jds.2014-9005.
- Bulman, D.C., and G.E. Lamming. 1978. Milk progesterone levels in relation to conception, repeat breeding and factors influencing acyclicity in dairy cows. *J. Reprod. Fertil.* 54:447–458.
- Calus, M.P.L. 2010. Genomic breeding value prediction: methods and procedures. *Animal.* 4:157–164.
- Daetwyler, H.D., M.P.L. Calus, R. Pong-Wong, G. los Campos, and J.M. Hickey. 2013. Genomic prediction in animals and plants: Simulation of data, validation, reporting, and benchmarking. *Genetics.* 193.
- Daetwyler, H.D., A. Capitan, H. Pausch, P. Stothard, R. Van Binsbergen, R.F. Brøndum, X. Liao, A. Djari, S.C. Rodriguez, C. Grohs, D. Esquerré, O. Bouchez, M.-N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P.J. Bowman, D. Coote, A.J. Chamberlain, C. Anderson, C.P. VanTassell, I. Hulsege, M.E. Goddard, B. Guldbrandtsen, M.S. Lund, R.F. Veerkamp, D.A. Boichard, R. Fries, and B.J. Hayes. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* 46:858–865.
- Daetwyler, H.D., R. Pong-Wong, B. Villanueva, and J.A. Woolliams. 2010. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics.* 185:1021 LP-1031.
- Daetwyler, H.D., B. Villanueva, and J.A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One.* 3:e3395.
- Darwash, A.O., G.L. Ward, G.E. Lamming, and J.A. Woolliams. 1999. The effects of raising post-oestrus progesterone concentrations on luteal activity in post-partum dairy cows. *Anim. Sci.* 68:527–532.
- Duchemin, S.I., M. Glantz, D.-J. de Koning, M. Paulsson, and W.F. Fikse. 2016. Identification of QTL on Chromosome 18 Associated with Non-Coagulating Milk in Swedish Red Cows. *Front. Genet.* 7:57.
- Duchemin, S.I., M.H.P.W. Visker, J.A.M. Van Arendonk, and H. Bovenhuis. 2014. A quantitative trait locus on Bos taurus autosome 17 explains a large proportion of the genetic variation in de novo synthesized milk fatty acids. *J. Dairy Sci.* 97:7276–7285.

- Fogh, A., G.P. Aamand, L. Hjørtø, and U. Lauritsen. 2013. Recording of data and identification issues. *In* ICAR-Technical Workshop 2013.
- Gengler, N. 2014. Mitigation strategies versus adaptation strategies. Invited presentation at the ASGGN Workshop.
- Goddard, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*. 136:245–257.
- Habier, D., R.L. Fernando, and J.C.M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*. 177:2389–2397.
- Haile-Mariam, M., P.J. Bowman, and M.E. Goddard. 2003. Genetic and environmental relationship among calving interval, survival, persistency of milk yield and somatic cell count in dairy cattle. *Livest. Prod. Sci.* 80:189–200.
- Hayes, B.J., P.J. Bowman, A.C. Chamberlain, K. Verbyla, and M.E. Goddard. 2009. Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet. Sel. Evol.* 41:1.
- Höglund, J.K., G. Sahana, R. Brøndum, B. Guldbrandtsen, B. Buitenhuis, and M.S. Lund. 2014. Fine mapping QTL for female fertility on BTA04 and BTA13 in dairy cattle using HD SNP and sequence data. *BMC Genomics*. 15:790.
- Ismael, A., E. Strandberg, B. Berglund, A. Fogh, and P. Løvendahl. 2016. Seasonality of fertility measured by physical activity traits in Holstein cows. *J. Dairy Sci.* 99:2837–2848.
- Ismael, A., E. Strandberg, M. Kargo, A. Fogh, and P. Løvendahl. 2015. Estrus traits derived from activity measurements are heritable and closely related to the time from calving to first insemination. *J. Dairy Sci.* 98:3470–3477.
- Kemper, K.E., and M.E. Goddard. 2012. Understanding and predicting complex traits: Knowledge from cattle. *Hum Mol Genet.* 21.
- Khatkar, M.S., F.W. Nicholas, A.R. Collins, K.R. Zenger, J.A.L. Cavanagh, W. Barris, R.D. Schnabel, J.F. Taylor, and H.W. Raadsma. 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics*. 9:1.
- Lamming, G.E., and a O. Darwash. 1998. The use of milk progesterone profiles to characterise components of subfertility in milked dairy cows. *Anim. Reprod. Sci.* 52:175–90.
- Løvendahl, P., and M.G.G. Chagunda. 2009. Short communication: Genetic variation in estrus activity traits. *J. Dairy Sci.* 92:4683–4688.
- Løvendahl, P., and M.G.G. Chagunda. 2016. On the use of physical activity monitoring for estrus detection in dairy cows. *J. Dairy Sci.* 93:249–259.

- Lucy, M.C. 2001. Reproductive loss in high-producing dairy cattle: where will it end? *J. Dairy Sci.* 84:1277–1293.
- Lund, M.S., A.P.W. De Roos, A.G. De Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, and R. Reents. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet Sel Evol.* 43:43.
- Meuwissen, T., and M. Goddard. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics.* 185:623–631.
- Meuwissen, T.H.E. 2009. Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* 41:1.
- Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 157:1819–1829.
- Miglior, F., J. Chesnais, and B.J. Van Doormaal. 2012. Genetic improvement: a major component of increased dairy farm profitability. *38th Ann Meet ICAR.* 28.
- Miglior, F., B.L. Muir, and B.J. Van Doormaal. 2005. Selection indices in Holstein cattle of various countries. *J. Dairy Sci.* 88:1255–1263.
- Misztal, I. 2011. FAQ for genomic selection. *J. Anim. Breed. Genet.* 128:245–246.
- Petersson, K.-J., B. Berglund, E. Strandberg, H. Gustafsson, A.P.F. Flint, J.A. Woolliams, and M.D. Royal. 2007. Genetic analysis of postpartum measures of luteal activity in dairy cows. *J. Dairy Sci.* 90:427–434.
- Pryce, J.E., S. Bolormaa, A.J. Chamberlain, P.J. Bowman, K. Savin, M.E. Goddard, and B.J. Hayes. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *J. Dairy Sci.* 93:3331–3345.
- Pryce, J.E., M.P. Coffey, and G. Simm. 2001. The relationship between body condition score and reproductive performance. *J. Dairy Sci.* 84:1508–1515.
- Pryce, J.E., M.E. Goddard, H.W. Raadsma, and B.J. Hayes. 2016. Deterministic models of breeding scheme designs that incorporate genomic selection. *J. Dairy Sci.* 93:5455–5466.
- Pszczola, M., T. Strabel, H.A. Mulder, and M.P.L. Calus. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. *J. Dairy Sci.* 95:389–400.
- Roxström, A., E. Strandberg, B. Berglund, U. Emanuelson, and J. Philipsson. 2001. 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 — Anim. Sci.* 51:7–14.

- Royal, M.D., A.P.F. Flint, and J.A. Woolliams. 2002. Genetic and Phenotypic Relationships Among Endocrine and Traditional Fertility Traits and Production Traits in Holstein-Friesian Dairy Cows. *J. Dairy Sci.* 85:958–967.
- Royal, M.D., E. Wall, and A.P.F. Flint. 2003. The genetic relationship between interval to commencement of luteal activity postpartum and UK national fertility proofs for dairy cattle. *Proc. Br. Soc. Anim. Sci.(BSAS)*, York, UK. *Br. Soc. Anim. Sci., Edinburgh, UK.* 54.
- Sahana, G., B. Guldbrandtsen, B. Thomsen, L.E. Holm, F. Panitz, R.F. Brøndum, C. Bendixen, and M.S. Lund. 2014. Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle¹. *J. Dairy Sci.* 97:7258–7275.
- Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123:218–223.
- Sun, C., P. Madsen, U.S. Nielsen, Y. Zhang, M.S. Lund, and G. Su. 2009. Comparison between a sire model and an animal model for genetic evaluation of fertility traits in Danish Holstein population. *J. Dairy Sci.* 92:4063–4071.
- Veerkamp, R.F., and B. Beerda. 2007. Genetics and genomics to improve fertility in high producing dairy cows TT - Proceedings of the International Conference on Farm Animal Reproduction “From Egg to Embryo” International Conference on Farm Animal Reproduction. *Theriogenology.* 68, Supple: S266–S273.
- Veerkamp, R.F., J.K. Oldenbroek, H.J. Van Der Gaast, and J.H.J. Van Der Werf. 2000. Genetic Correlation Between Days Until Start of Luteal Activity and Milk Yield, Energy Balance, and Live Weights. *J. Dairy Sci.* 83:577–583.
- Wall, E., S. Brotherstone, J.A. Woolliams, G. Banos, and M.P. Coffey. 2003. Genetic evaluation of fertility using direct and correlated traits. *J. Dairy Sci.* 86:4093–4102.
- Walsh, S.W., E.J. Williams, and A.C.O. Evans. 2011. A review of the causes of poor fertility in high milk producing dairy cows. *Anim. Reprod. Sci.* 123:127–138.
- Wientjes, Y.C., M.P. Calus, M.E. Goddard, and B.J. Hayes. 2015a. Impact of QTL properties on the accuracy of multi-breed genomic prediction. *Genet Sel Evol.* 47.
- Wientjes, Y.C.J., R.F. Veerkamp, and M.P.L. Calus. 2015b. Using selection index theory to estimate consistency of multi-locus linkage disequilibrium across populations. *BMC Genet.* 16:1.
- Wu, X., M.S. Lund, G. Sahana, B. Guldbrandtsen, D. Sun, Q. Zhang, and G. Su. 2015. Association analysis for udder health based on SNP-panel and sequence data in Danish Holsteins. *Genet. Sel. Evol.* 47:1.
- Yan, J., T. Shah, M.L. Warburton, E.S. Buckler, M.D. McMullen, and J. Crouch. 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One.* 4: e8451.

Yang, J., B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin, and G.W. Montgomery. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–569.

Summary

Summary

Improved reproductive performance has a substantial benefit for the overall profitability of dairy cattle farming by decreasing insemination and veterinary treatment costs, shortening calving intervals, and lowering the rate of involuntary culling. Unfortunately, the low heritability of classical fertility traits derived from calving and insemination data makes genetic improvement by traditional animal breeding slow. Therefore, there is an interest in finding novel measures of fertility that have a higher heritability or using genomic information to aid genetic selection for fertility. Endocrine fertility traits based on progesterone concentration in milk have been widely accepted as valid indicators for fertility because they are not biased by farm management decisions and more directly reflect a cow's reproductive physiology than classical fertility traits. However, the use of endocrine traits in routine genetic evaluations has been constrained by the high cost associated with collecting progesterone records on sufficient number of animals. Nonetheless, in-line technology like the Herd Navigator now exists to automatically measure progesterone concentration during milking, enabling sampling of more animals at reasonable cost.

In this thesis, I explored the use of milk progesterone concentrations and genomic information to improve genetic selection for fertility in dairy cattle. In chapter 2, I investigated the use of in-line milk progesterone records to define endocrine fertility traits, and estimated genetic parameters. Several defined endocrine fertility traits were heritable, and showed a reasonable repeatability. The genetic correlation of milk production traits with endocrine fertility traits were considerably lower than the correlations of milk production with classical fertility traits. In chapter 3, genomic regions associated to endocrine fertility traits were identified on the cattle genome by genome-wide association and fine-mapping.

The genome-wide association study identified 17 quantitative trait loci (QTL) associated with endocrine fertility traits, on *Bos taurus* autosomes (BTA) 2, 3, 8, 12, 15, 17, 23, and 25. Overlapping QTL regions were found between endocrine traits on BTA 2, 3, and 17. For the classical trait calving to first service, three QTL regions were identified on BTA 3, 15, and 23, and an overlapping region on BTA23 with endocrine traits. Fine-mapping target regions for the endocrine traits on BTA 2 and 3 confirmed the QTL from the genome-wide association study, and identified several associated variants that can contribute to an index of markers for genetic improvement of

fertility. Several potential candidate genes underlying endocrine fertility traits were also identified in the target regions.

In the next two chapters, I investigated the optimal use of endocrine fertility traits in genomic evaluations. In chapter 4, using empirical and theoretical predictions for single-trait models, I showed that endocrine fertility traits have more predictive ability than classical fertility traits. The accuracy of genomic prediction was also substantially improved when endocrine and classical fertility traits were combined in multi-trait genomic prediction. Across country predictions were also evaluated in univariate predictions, and some predictive ability was observed. Because we were limited by sample size in chapter 4, in chapter 5, using a deterministic equation, I investigated the potential accuracy of multi-trait genomic selection when combining a cow training population measured for the endocrine fertility trait C-LA, with a training population of bulls with daughter observations for classical fertility traits. The results showed that for prediction of fertility, there is no benefit of investing in a cow training population when the breeding goal is based on classical fertility traits. However, when considering a more biological breeding goal for fertility like C-LA, accuracy is substantially improved when endocrine traits are available from a limited number of farms. Evaluation of the potential accuracy of across country genomic prediction showed that when the training population from one country (e.g., Sweden) is small, substantial increase in accuracy can be achieved by adding animals from another country (e.g., The Netherlands), however, the accuracy was highly dependent on the genetic correlation between countries. In chapter 5, recording strategies that optimally use the Herd Navigator for genomic prediction, in terms of, the number of farms, and recording period for endocrine traits were investigated. The results showed that for prediction of fertility using C-LA, it is more important to have more animals with C-LA records than more C-LA records per animal. That is, for prediction accuracy with novel traits in general, it is more beneficial to phenotype more animals than more lactations per animal in the training population.

Finally, in chapter 6, the relevance of the findings was discussed, i.e. how breeders can benefit from combining endocrine fertility traits with genomic information to further improve selection for fertility. Suggestions for future studies and how breeders can make use of the results generated from this study were given. I finalized by suggesting practical phenotyping strategies like phenotyping labs, that can be established by breeding companies to optimize phenotype recording for endocrine fertility traits (and novel traits in general).

Curriculum vitae

About the author

List of publications

Training and supervision plan

About the author

Amabel Manyu Mefru Tenghe was born on the 3rd of April 1984 in Pinyin, Southern Cameroons. She attended secondary school at Lycée Bilengue de Garoua, and graduated from high school in 2004 from Cameroon Protestant Colledge Bali. In 2007, she obtained a BSc in Zoology with a minor in medical laboratory technology from the University of Buea. Amabel's passion for applied agricultural sciences led her to pursue the European Masters in Animal Breeding and Genetics from 2008. During her Masters, Amabel did two major theses. One on estimating genetic parameters for pigmentation and shape in fish, at Wageningen University (Netherlands), and the other on detecting genomic regions associated to production traits in commercial pig populations at Agro ParisTech (France). Upon obtaining a double MSc degree in 2010, she worked as a research assistant at the Unit of Animal Genomics at Universite de Liège (Belgium). In January 2013, she started her PhD, which is part of the European Graduate School in Animal Breeding and Genetics. While most of her PhD was done at Wageningen University, she had the opportunity to spend one year at the Swedish University of Agricultural Sciences in Uppsala (Sweden). In her PhD, Amabel investigated the use of milk progesterone measures to improve genomic selection for fertility in dairy cows, the results of which are presented in this thesis.

List of publications

Peer reviewed publications

A.M.M. Tenghe, B. Berglund, R.F. Veerkamp, and D.J. de Koning. 2016. Opportunities for genomic prediction for fertility using endocrine and classical fertility traits in dairy cattle. *J. Anim. Sci.*

A.M.M. Tenghe, A.C. Bouwman, B. Berglund, E. Strandberg, D.J. de Koning, and R.F. Veerkamp. 2016. Genome-wide association study for endocrine fertility traits using single nucleotide polymorphism arrays and sequence variants in dairy cattle. *J. Dairy Sci.* 99:5470–5485

A.M.M. Tenghe, A.C. Bouwman, B. Berglund, E. Strandberg, J.Y. Blom, and R.F. Veerkamp. 2015. Estimating genetic parameters for fertility in dairy cows from in-line milk progesterone profiles. *J. Dairy Sci.* 98:5763–73

Sanchez M.-P., Tribout T., Iannuccelli N., Bouffaud M., Servin B., **Tenghe A.**, et al. (2014). A genome-wide association study of production traits in a commercial population of Large White pigs: evidence of haplotypes affecting meat quality. *Genet. Sel. Evol.* 46:12 10.1186/1297-9686-46-12

Submitted manuscript

A.M.M. Tenghe, A.C. Bouwman, B. Berglund, D.J. de Koning and R.F. Veerkamp. Improving accuracy of bulls 'predicted genomic breeding values for fertility using daughters' milk progesterone profiles. Submitted to *Journal of Dairy Science*

Conference proceedings

A.M.M. Tenghe, B. Berglund, R.F. Veerkamp, and D.J. de Koning. Accuracy of genomic breeding values from endocrine and traditional fertility traits in dairy cows. 67th Annual Meeting of the European Federation of Animal Science (EAAP), Belfast, UK.

Amabel Tenghe, Aniek Bouwman, Britt Berglund, Erling Strandberg, Dirk-Jan de Koning, Roel Veerkamp. Towards using endocrine fertility traits in genomic selection for dairy cow fertility. PROLIFIC 3rd annual consortium meeting Uppsala, Sweden, 2016

Tenghe, A.M.M., Bouwman, A.C, B Berglund, Strandberg, E, Veerkamp, R.F. Unravelling the genetic background of endocrine fertility traits. 66th Annual Meeting of the European Federation of Animal Science (EAAP), Warsaw, Poland in September 2015.

Tenghe, A.M.M, Bouwman, A.C, Berglund, B, Strandberg, E, Blom, J, Veerkamp, R.F. Unravelling the genetic background of endocrine fertility traits. PROLIFIC 2nd annual consortium meeting Versailles, France, 2015

A.M.M. Tenghe A.C. Bouwman, B. Berglund, E. Strandberg, J.Y. Blom, and R.F. Veerkamp Genetic Parameters for Endocrine Fertility Traits from In-line Milk Progesterone Records in Dairy Cows. In: Proceedings of 10th World Congress on Genetics Applied to Livestock Production (WCGALP), Vancouver, Canada in August 2014.

Tenghe, A.M.M, Bouwman, A.C, Berglund, B, Strandberg, E, Blom, J, Veerkamp, R.F. Genetic parameters for endocrine fertility traits in dairy cows. PROLIFIC 1st annual consortium meeting Wageningen, Netherlands, 2014

Veerkamp, R.F, **Tenghe, A.M.M,** Kaal, L.M.T.E, Bouwman, A.C. Genetics and genomics of fertility in dairy cows In: Proceedings of the Nottingham Cattle Fertility Conference 2014, p.25-30.



Training and supervision plan

The basic package (7 ECTS)

WIAS Introduction Course, Wageningen (Netherlands), 2014

Ethics and Philosophy in Life Sciences, Wageningen (Netherlands), 2014

EGS-ABG Summer School SLU, Uppsala (Sweden), 2016

EGS-ABG Summer School WUR, Wageningen (Netherlands), 2017

Scientific exposure (12 ECTS)

International conferences

World Congress on Genetics Applied to Livestock Production, Vancouver (Canada), 2014

European association of Animal Production (EAAP), Warsaw (Poland), 2015

European association of Animal Production (EAAP), Belfast (UK), 2016

Seminars and workshops

WIAS Science day, Wageningen (Netherlands), 2014

Prolific Annual Meeting, Wageningen (Netherlands), 2014

WIAS Science day, Wageningen (Netherlands), 2015

Prolific Annual Meeting, Paris (France), 2016

Prolific Annual Meeting, Uppsala (Sweden), 2016

Presentations

Oral, at Prolific Annual meeting, Wageningen (Netherlands), 2014

Oral, at World Congress on Genetics Applied to Livestock Production, Vancouver (Canada), 2014

Oral, at European association of Animal Production, Warsaw (Poland), 2015

Oral, at Prolific Annual Meeting, Paris (France), 2015

Oral, at Prolific Annual Meeting, Uppsala (Sweden), 2016

Oral, at European association of Animal Production, Belfast (UK), 2016

Disciplinary and interdisciplinary courses (14 ECTS)

Statistics for the life Sciences, Wageningen (Netherlands), 2014
Genetic Analysis using ASREML4.0, Wageningen (Netherlands), 2014
Genomic Selection Theory, Wageningen (Netherlands), 2014
Linear models in Animal Breeding and Genetics, Lofeton (Norway), 2015
From sequence data to genomic prediction, Wraclaw (Poland), 2015
Statistical Genetics, Faro (Portugal), 2016

Professional skills support courses (7 ECTS)

Techniques for Writing and Presenting a Scientific Paper, Wageningen (Netherlands), 2014
Information Literacy PhD including EndNote Introduction
Scientific Writing
Scientific Publishing
Data Management
Reviewing a scientific paper
Presentation Skills
Project and Time Management

Research skills training (10 ECTS)

Preparing own PhD research proposal, Wageningen (Netherlands), 2013
External training period, 1 year at SLU, Uppsala (Sweden), 2015-2016
WIAS science day organizing committee, Wageningen (Netherlands), 2014
Quantitative Genetics Discussion Group, Wageningen (Netherlands), 2013-2016

Education and training total: 50 ECTS

Acknowledgements

Acknowledgements

The work in this thesis truly has resulted from the contribution of many. The list of individuals to whom I owe the utmost gratitude is lengthy. Without their support, this work would simply not have been possible. I first give honor to God, who has provided this vision and purpose in my life, and endowed me with the resources and facilities to fulfill it. These past 4 years have been a worthy journey and adventure.

A special thank you to my supervisor team for their guidance and support throughout this journey. Roel, thank you for your critical and dynamic attitude, helping to push my thoughts in the right direction. I admire your dynamic skill, which I aspire to acquire. Aniek, thank you for the very detailed and useful remarks to my papers, even the minor details, they were very much appreciated. Britt, you always found a way to encourage me, and I learnt a lot about fertility from you, thank you. Erling, I am very grateful to have had you as a supervisor for the first half of my PhD, thank you for your always very insightful comments. DJ, although you joined the team half way, you immediately took interest in my work, and I very much enjoyed working with you on my last papers. Thank you for all the discussions, especially during the time spent in Uppsala.

Thank you, Sander de Roos, Donagh Berry, and Mike Coffey, for being in my thesis committee. Thank you for interesting and valuable discussions. Thanks to my opponents for reading the thesis, I look forward to the discussion during the public defense.

Mathijs, thank you for sharing the office with me at the start of my thesis, and for all the nice discussions we had. Your super-duper organized skills inspired me to want to improve on mine. Ernest, thank you for your friendship, and all the laughs :). Both of you, thanks for taking the responsibility of being my paranymphs. Marvis, thank for your hospitality, and for offering me a place to stay whenever I needed one. Thank you, Johan and Helena Valema, for opening your home to me, and always encouraging me. Special thanks to Coralia, Yvonne, Rianne, Sonia, Mathijs, Yvette, Mario, Ina, Eef, Marcin, and all at the Triton building for helping me fit in at the start of my PhD.

To my ICF family in Wageningen, thank you for receiving me always with warmth, for the great workshops and fellowships. Special thanks to the worship team members; Yuni, Peter, Gerson, Koenraad, Adnan, Sebastiaan, Ebenezer for all the worship

Acknowledgements

preparation time shared. Special thanks also go to the “Connect group Ede”, thank you for your companionship and elaborate discussion of the word.

This work was carried out at Wageningen University (WUR) and the Swedish University of Agricultural Sciences (SLU). To researchers and my colleagues at both Universities, thank you for your helpfulness, and for creating an atmosphere that I enjoyed a lot while working on my thesis. A special thank you to Ina, Zih-Hua, Mandy, Nancy, Henk, Jack, Jovanna, Juan, John, Lucia, Mahlet, Marcos, Marinus, Sandrine, Sandrine, Tessa, Hamed, Pascal, Hadi, Andre, Gabriel, Jeremie, Floor, Kasper, Sanne, Piter, Hans, Han, Bart Maria, Esther, Gareth, Biaty, Anouk, Anoop, Tom, Sabine, Shuwen, Robert, Asaah, Rita, from WUR. Thanks to Agnese, Merina, Chrissy, Beriuhu, Sofia, Emilie, Josh, Shizi, Thu, Cano, Suvi, Kim, Freddy, Iotta, Lisa, Valentina, from SLU, for all the Firka and nice dinners together. I would also like to acknowledge all administrative staff members for all the help I received from them. Thank you to Ada, Lisette, and Maya at WUR, Monika, and Helena at SLU, and Helena Fuchs and Isabelle Laissy at the EGS-ABG secretary.

I would like to acknowledge members of the EU Prolific project, of which this work was a subset of the larger project, thank you for lots of valuable interactions during the annual project meetings.

To Coralia, Claudia, and Merina, thank you for all the awesome trips we had together. These diversions made the journey more fun and exciting, and I am honored to have shared them with you girls, and to get to know you more. Thank you, Sonia, Zih-Hua, Maria, and Coralia for the ladies’ cocktail nights. I admit, you girls got the best cocktail skills ;).

Special thanks to my family for their patience and support given to me over the last years. Mom and Dad, thank you for your love, and for being the role models in my life. My dearest sisters; Makah, Aza, Bella, Glory, and brothers; Akum, big & small Muluh, Nda, thank you for cracking me up with jokes whenever I called or was home. You guys are the best, you’re too “nyanga”

Dear reader, thank you for taking the effort to read this thesis. I hope you enjoyed it and consider the time spent valuable.

Amabel

Colophon

Colophon

The research described in this thesis was supported by the European Commission (within the framework of the Erasmus-Mundus joint doctorate “EGS-ABG”), and Prolific (a project financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement n°311776)

The Dutch data used in this thesis were provided by Lattec I/S (Denmark), and the Cooperative Cattle Improvement Organization CRV BV (Arnhem, the Netherlands). Additional data from Sweden, Ireland, and the United Kingdom were contributed by RobustMilk (a project financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-211708).

The cover of this thesis was designed by Piotr Sell (www.pcell.it).

The thesis was printed by Digiforce.