Combining Genotypic, Phenotypic and Pedigree Information to Analyze Functional Traits in Dairy Cattle

by
Aline Rocha Guarini

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Doctor of Philosophy
in
Animal Biosciences

Guelph, Ontario, Canada

© Aline Rocha Guarini, December, 2017
ABSTRACT

COMBINING GENOTYPIC, PHENOTYPIC AND PEDIGREE INFORMATION TO ANALYZE FUNCTIONAL TRAITS IN DAIRY CATTLE

Aline Rocha Guarini
University of Guelph, 2017

Advisor:
Professor Flavio S. Schenkel

Reliabilities of genomic estimated breeding values (GEBV) for functional traits, e.g. health, fertility and reproduction, remain low compared to those for production in dairy cattle. This is likely because large training populations are required for evaluation of lowly heritable traits. Different strategies have been proposed to overcome this limitation, such as the use of genotyped cows, inclusion of external bulls (multi-trait across-country evaluation, MACE), and adoption of different methodologies, such as the simultaneous use of genotyped and non-genotyped animals, known as the single-step genomic BLUP (ssGBLUP). Thus, the main objectives of this thesis were to evaluate strategies for combining genotypic, phenotypic and pedigree information to analyze functional traits in Holstein cattle and investigate the effects of two deleterious recessive haplotypes (AH1 and AH2) on reproduction performance of Canadian Ayrshire cattle.

Data for various functional traits recorded in Canada and MACE estimated breeding values were obtained from the Canadian Dairy Network (CDN, Guelph, Canada). Additionally, information of carriers and non-carriers bulls for AH1 and AH2 were used to investigate their effects on reproductive performance. Genomic predictions were obtained using multi-step and single-step GBLUP under different strategies, such as adding genotyped cows in the evaluation, integrating MACE information, blending traditional and genomic evaluations, and using different
proportions of polygenic effect. A genome-wide association study and functional analyses were also performed for three fertility disorders, namely retained placenta, metritis and cystic ovaries.

Genomic predictions for functional traits benefited greatly from simultaneous use of phenotypes, pedigree, and genotypes. Integration of MACE data using ssGBLUP yielded the highest reliabilities compared to other methods and also helped reduce bias of genomic predictions. Effects of AH1 and AH2 on reproductive performance of Canadian Ayrshire cattle were validated. A negative effect of AH1 on stillbirth rates was observed, whereas AH2 had a negative impact on 56-day non-return rate.

These findings provide valuable information on strategies to more accurately predict GEBV for functional traits in dairy cattle by adopting ssGBLUP and different sources of domestic and foreign information. Biological understanding of reproductive disorders and validated lethal haplotypes affecting fertility will help enhance accuracy of selection and mating plans in the Canadian dairy cattle breeding programs.
Acknowledgements

To start my list of acknowledgements, I would like to extend my sincere gratitude to my advisor, Dr. Flavio Schenkel, who gave me the opportunity to move to Canada and start a whole new chapter of my life. Being around you have always inspired and motivated me in so many ways. Maybe even more than you can think of. You are not only an example of an excellent professor, but also a great father, grandfather, husband and friend. In the end, I think I wanted to make you proud of me more than I wanted to finish my PhD. I hope I have accomplished that and I hope that this work is what you hoped it would be.

I am also grateful to all the members of my advisory committee: Dr. Christine Baes, Dr. Daniela Lourenço, Dr. Filippo Miglior, and Dr. Mehdi Sargolzaei. I have no words to describe how grateful I am for having the opportunity to work with you and, most of all, learn from you. Thank you for putting up with my non-stopping questions about everything. And, most of all, thank you for always having you door open when I needed (Or Skype, Email, and WhatsApp). You have taught me so much and I will be forever grateful. This whole journey would not have been possible without your support, your great knowledge, and, most of all, your willingness to help when I needed the most (And always with a smile).

I also want to express my sincere gratitude to CAPES (Brazilian Federal Agency for Support and Evaluation of Graduate Education, Brazil) and the Brazilian government for financial support during my PhD.

Dr. Luiz Brito, thank you for the huge help in putting this thesis together, for all the times when you said you were not busy to help (even though you had a million things to do), and most of all, thank you for being such an amazing friend! I would need, at least, 10 pages to thank everyone for being part of my life in these past years, but since I don’t have that, here are some of
the people that I couldn't go on without acknowledging: Hinayah Rojas, Emily Miller, Kristen Alves, Fabieli Feitosa, and Guilherme Batista do Nascimento. You all made my life in Canada so much better and I hope to have you as friends for the rest of my days. Thank you to all my friends that I was not able to list here but have place in my heart nonetheless. You all made a difference in my life somehow.

To my parents, Leonice and Ronaldo, my sister, Naiara, and my brother-in-law, Kleber. You are the ones that truly made me be who I am today. You taught me the importance of always treating others with respect and kindness. You taught me the importance of being an honest person and being thankful for what I have in life. You taught me how to take care of myself and never let others put me down. You taught me so many things, and I will be forever thankful for that, but the one thing I am grateful the most is for your unconditional love and support always.

And my final acknowledgement goes to my partner, Ala. Thank you! You have been with me during the best and worst moments of this graduate life. And for sure you made it more interesting and special. After all this time and after helping me practice for so many presentations, you probably know more about the single-step GBLUP than I do. I cannot wait to see what life has reserved for us. I love you with all my heart!
**Table of Contents**

List of Tables ........................................................................................................................................... ix  
List of Figures ........................................................................................................................................... xii  
List of Abbreviations ................................................................................................................................. xiii  

**CHAPTER 1: General introduction** ........................................................................................................ 1  
1.1 *INTRODUCTION* ................................................................................................................................. 1  
1.2 *THESIS OUTLINE* ............................................................................................................................... 6  
1.3 *THESIS OBJECTIVES* ............................................................................................................................ 7  
1.4 *REFERENCES* ....................................................................................................................................... 9  

**CHAPTER 2: Comparison of genomic predictions for lowly heritable traits using multi-step**  
and single-step genomic BLUP in Holstein cattle ....................................................................................... 14  
2.1 *ABSTRACT* .......................................................................................................................................... 14  
2.2 *INTRODUCTION* .................................................................................................................................. 16  
2.3 *MATERIALS AND METHODS* .............................................................................................................. 18  
   2.3.1 Data ............................................................................................................................................... 18  
   2.3.2 Genomic predictions .......................................................................................................................... 19  
   2.3.3 Statistical models .............................................................................................................................. 20  
      2.3.3.1 Multi-step GBLUP Model ........................................................................................................... 20  
      2.3.3.2 Single-Step GBLUP Model ....................................................................................................... 21  
   2.3.4 Validation ....................................................................................................................................... 22  
2.4 *RESULTS AND DISCUSSION* ............................................................................................................ 23  
2.5 *CONCLUSIONS* .................................................................................................................................. 28  
2.6 *ACKNOWLEDGEMENTS* .................................................................................................................... 29  
2.7 *TABLES* ............................................................................................................................................ 30  
2.8 *REFERENCES* ..................................................................................................................................... 34  

**CHAPTER 3: Use of a single-step approach for integrating foreign information into national**  
genomic evaluation in Holstein cattle ........................................................................................................... 39  
3.1 *ABSTRACT* .......................................................................................................................................... 39  
3.2 *INTRODUCTION* .................................................................................................................................. 40  
3.3 *MATERIALS AND METHODS* ............................................................................................................. 42
3.3.1 Data ..............................................................................................................................................42
  3.3.1.1 Genotypes ..................................................................................................................................42
  3.3.1.2 Phenotypes ..................................................................................................................................42
  3.3.1.3 Domestic and MACE EBVs ......................................................................................................43
3.3.2 Methods of Evaluation ...................................................................................................................43
  3.3.2.1 Single-Step GBLUP Model ....................................................................................................43
  3.3.2.2 Multi-step GBLUP Models ....................................................................................................45
3.3.3 Implementation ..............................................................................................................................46
3.3.4 Validation ........................................................................................................................................46
3.4 RESULTS AND DISCUSSION .............................................................................................................48
3.5 CONCLUSIONS ....................................................................................................................................53
3.6 ACKNOWLEDGEMENTS ..................................................................................................................53
3.7 TABLES ................................................................................................................................................54
3.8 FIGURES ...............................................................................................................................................57
3.9 REFERENCES .......................................................................................................................................58

CHAPTER 4: Genetics and genomics of reproductive disorders in Holstein cattle ........ 61

4.1 ABSTRACT .........................................................................................................................................61
4.2 INTRODUCTION ...............................................................................................................................62
4.3 MATERIALS AND METHODS ..........................................................................................................65
  4.3.1 Data ..............................................................................................................................................65
    4.3.1.1 Database ...................................................................................................................................65
    4.3.1.2 Data Validation and Editing .................................................................................................66
    4.3.1.3 Trait Definitions ..................................................................................................................66
  4.3.2 Variance Component Estimation ..................................................................................................67
  4.3.3 Genomic Prediction of Breeding Values ....................................................................................68
  4.3.4 Genome-Wide Association Study ..............................................................................................69
  4.3.5 Functional Analyses ..................................................................................................................71
4.4 RESULTS AND DISCUSSION .............................................................................................................72
4.4.1 Heritabilities and Genetic and Phenotypic Correlations ..........................................................72
4.4.2 Prediction of Breeding Values ........................................................................................................73
4.4.3 Genome-wide Association Study and Functional Analyses ..........................................................75
4.5 CONCLUSIONS ....................................................................................................................................79
4.6 ACKNOWLEDGEMENTS .......................................................................................................................80
4.7 TABLES ..................................................................................................................................................81
4.8 FIGURES ...............................................................................................................................................88
4.9 REFERENCES .........................................................................................................................................91

CHAPTER 5: Estimating the impact of the deleterious recessive haplotypes on reproduction performance of Ayrshire cattle .........................................................................................................................101
5.1 ABSTRACT ...........................................................................................................................................101
5.2 INTRODUCTION ..................................................................................................................................102
5.3 MATERIALS AND METHODS ...............................................................................................................104
5.4 RESULTS AND DISCUSSION ...............................................................................................................107
5.5 CONCLUSIONS ....................................................................................................................................113
5.6 ACKNOWLEDGEMENTS .......................................................................................................................113
5.7 TABLES ..................................................................................................................................................114
5.8. FIGURES .............................................................................................................................................122
5.9 REFERENCES .........................................................................................................................................123

CHAPTER 6: General conclusions ..................................................................................................................129
6.1 FINAL REMARKS .................................................................................................................................129
6.2 REFERENCES .........................................................................................................................................133
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Descriptive statistics and heritability ($h^2$) of reproductive and workability traits in first and later parities</td>
<td>30</td>
</tr>
<tr>
<td>2.2</td>
<td>Structure of training and validation datasets for msGBLUP and ssGBLUP methods</td>
<td>31</td>
</tr>
<tr>
<td>2.3</td>
<td>Observed reliabilities of genomic predictions obtained using different methods</td>
<td>32</td>
</tr>
<tr>
<td>2.4</td>
<td>Regression coefficient of the linear regression of de-regressed EBV (DDc) on genomic predictions (GEBV) or on standard BLUP evaluations (PA) obtained using different methods</td>
<td>33</td>
</tr>
<tr>
<td>3.1</td>
<td>Descriptive statistics for Multi-trait across-country evaluation (MACE) EBVs used as external information in single-step GBLUP for evaluation of milking temperament and milking speed</td>
<td>54</td>
</tr>
<tr>
<td>3.2</td>
<td>Descriptive statistics of EBVs used for genomic evaluations using multi-step GBLUP</td>
<td>55</td>
</tr>
<tr>
<td>3.3</td>
<td>Observed reliabilities ($R^2$) of genomic predictions and regression coefficient ($b_1$) of de-regressed EBVs on genomic predictions for 554 genotyped bulls, with at least 50 daughters, for milking temperament (MT) and milking speed (MS) for different genomic evaluations</td>
<td>56</td>
</tr>
<tr>
<td>4.1</td>
<td>Summary statistics for health disorders in first-lactation Canadian Holstein cows</td>
<td>81</td>
</tr>
<tr>
<td>4.2</td>
<td>Heritability estimates (in bold on the diagonal, with posterior standard deviation, SD, in parentheses), genetic correlations (above diagonal, with posterior SD in parentheses), and phenotypic correlations (below diagonal) from multiple-trait analysis of first-lactation Holstein cows</td>
<td>82</td>
</tr>
</tbody>
</table>
4.3 Mean reliability of sire EBV (number of sires) computed with pedigree and genomic information for first-lactation records using single-trait analyses …… 83

4.4 Mean reliability of sire EBV computed with pedigree and genomic information for first-lactation records using multiple-trait analysis .......................... 84

4.5 Percentage of affected daughters of all sires, top 10% sires, and bottom 90% sires according to their GEBV for cystic ovaries, metritis, and retained placenta from single-trait analyses of first-lactation Holstein cows .......................... 85

4.6 Twenty-SNP windows that explained more than 1% of genetic variance (Var) for cystic ovaries (CYST), metritis (METR), and retained placenta (RETP) in Canadian Holstein dairy cows and the corresponding list of annotated genes within each window ........................................................................................................ 86

4.7 Pathways identified, associated enzymes, and associated genes for cystic ovaries (CYST), metritis (METR), and retained placenta (RETP) in the Kyoto Encyclopedia of Genes and Genomes (KEGG). Biological pathways known to be related to CYST, METR, and/or RETP are shown in bold ......................... 87

5.1 Descriptive statistics for stillbirth and 56-day non-return rate in Canadian Ayrshire ........................................................................................................... 114

5.2 Distribution of livability scores for stillbirth by parity and sex of calf in the edited data-set used for the genetic evaluation .................................................. 115

5.3 Observed distribution of stillbirth (SB) and 56-day non-return rate (NRR) by carrier status of service sire within carrier status of maternal grandsire for AH1 haplotype in Canadian Ayrshire .......................................................... 116

5.4 Observed distribution of stillbirth (SB) and 56-day non-return rate (NRR) by carrier status of service sire within carrier status of maternal grandsire for AH2 haplotype in Canadian Ayrshire .......................................................... 117
5.5 Numbers of observations (N) and effects (± SE) of carrier status of maternal grandsire on 56-day non-return rate within carrier status of service sire for AH1 haplotype for heifers and cows in Canadian Ayrshire ………………………………………118

5.6 Numbers of observations and effects (± SE) of carrier status of maternal grandsire on stillbirth within carrier status of sire of calf for AH1 haplotype for heifers and cows in Canadian Ayrshire ………………………………………119

5.7 Numbers of observations and effects (± SE) of carrier status of maternal grandsire on 56-day non-return rate within carrier status of service sire for AH2 haplotype for heifers and cows in Canadian Ayrshire ………………………………………120

5.8 Numbers of observations and effects (± SE) of carrier status of maternal grandsire on stillbirth within carrier status of sire of calf for AH2 haplotype for heifers and cows in Canadian Ayrshire ………………………………………121
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Information flow for calculation of national and international genetic evaluations. EBV = estimated breeding values; GEBV = genomic estimated breeding values; MACE = multiple-trait across-country evaluations; GMACE = genomic enhanced MACE.</td>
<td>57</td>
</tr>
<tr>
<td>4.1</td>
<td>Increase in reliability of predictions when adding genomic information plotted against number of daughters for each sire in single-step analysis for cystic ovaries.</td>
<td>88</td>
</tr>
<tr>
<td>4.2</td>
<td>Manhattan plot for the proportion of genetic variance explained by the 20-SNP moving windows for cystic ovaries.</td>
<td>89</td>
</tr>
<tr>
<td>4.3</td>
<td>Manhattan plot for the proportion of genetic variance explained by the 20-SNP moving windows for metritis.</td>
<td>89</td>
</tr>
<tr>
<td>4.4</td>
<td>Manhattan plot for the proportion of genetic variance explained by the 20-SNP moving windows for retained placenta.</td>
<td>90</td>
</tr>
<tr>
<td>5.1</td>
<td>Phenotypic trend for 56-day non-return rate (A) and stillbirth (B) in heifers (dotted line), and cows (broken line).</td>
<td>122</td>
</tr>
</tbody>
</table>
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Pedigree relationship matrix</td>
</tr>
<tr>
<td>AcMcX</td>
<td>Age at current calving by month of current calving by sex of calf by parity</td>
</tr>
<tr>
<td>AFS</td>
<td>Age at first insemination</td>
</tr>
<tr>
<td>AH1</td>
<td>Ayrshire Haplotype 1</td>
</tr>
<tr>
<td>AH2</td>
<td>Ayrshire Haplotype 2</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial insemination</td>
</tr>
<tr>
<td>An</td>
<td>Animal additive genetic effect</td>
</tr>
<tr>
<td>ApMf</td>
<td>Age at previous calving by month of first insemination by parity</td>
</tr>
<tr>
<td>BLUP</td>
<td>Best Linear Unbiased Prediction</td>
</tr>
<tr>
<td>BTA</td>
<td><em>Bos taurus</em> autosome</td>
</tr>
<tr>
<td>CDCB</td>
<td>Council on Dairy Cattle Breeding</td>
</tr>
<tr>
<td>CDN</td>
<td>Canadian Dairy Network</td>
</tr>
<tr>
<td>CE</td>
<td>Calving ease</td>
</tr>
<tr>
<td>CTFS</td>
<td>Days from calving to first insemination</td>
</tr>
<tr>
<td>CYST</td>
<td>Cystic ovaries</td>
</tr>
<tr>
<td>CZ</td>
<td>Calf size</td>
</tr>
<tr>
<td>DD</td>
<td>De-regressed breeding values</td>
</tr>
<tr>
<td>DE&lt;sub&gt;PA&lt;/sub&gt;</td>
<td>Daughter equivalent from parent average</td>
</tr>
<tr>
<td>DE&lt;sub&gt;Prog&lt;/sub&gt;</td>
<td>Daughter equivalent from progeny information</td>
</tr>
<tr>
<td>DGV</td>
<td>Direct genomic value</td>
</tr>
<tr>
<td>E</td>
<td>Error term</td>
</tr>
<tr>
<td>EBV</td>
<td>Estimated breeding value</td>
</tr>
<tr>
<td>EBV&lt;sub&gt;M&lt;/sub&gt;</td>
<td>Estimated breeding value for multi-trait Across-country evaluation bulls</td>
</tr>
<tr>
<td>EBV&lt;sub&gt;rel&lt;/sub&gt;</td>
<td>Reliability of estimated breeding value</td>
</tr>
<tr>
<td>FSTC</td>
<td>Days from first service to conception</td>
</tr>
<tr>
<td>G</td>
<td>Genomic relationship matrix</td>
</tr>
<tr>
<td>GBLUP</td>
<td>Genomic Best Linear Unbiased Prediction</td>
</tr>
<tr>
<td>Abbr</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>GEBV</td>
<td>Genomic Estimated Breeding Value</td>
</tr>
<tr>
<td>GL</td>
<td>Gestation length</td>
</tr>
<tr>
<td>GS</td>
<td>Genomic Selection</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association studies</td>
</tr>
<tr>
<td>H</td>
<td>Matrix that combines pedigree and genomic relations</td>
</tr>
<tr>
<td>HY</td>
<td>Herd by year of birth</td>
</tr>
<tr>
<td>INT</td>
<td>Interaction term</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto Encyclopedia of Genes and Genomes</td>
</tr>
<tr>
<td>LD</td>
<td>Linkage disequilibrium</td>
</tr>
<tr>
<td>MACE</td>
<td>Multi-trait across-country evaluation</td>
</tr>
<tr>
<td>MAF</td>
<td>Minor allele frequency</td>
</tr>
<tr>
<td>Mb</td>
<td>Mega-base pairs</td>
</tr>
<tr>
<td>METR</td>
<td>Metritis</td>
</tr>
<tr>
<td>Mf</td>
<td>Month of first insemination</td>
</tr>
<tr>
<td>MGS</td>
<td>Maternal grandsire</td>
</tr>
<tr>
<td>MS</td>
<td>Milking speed</td>
</tr>
<tr>
<td>msGBLUP</td>
<td>Multi-step Genomic Best Linear Unbiased Predictor</td>
</tr>
<tr>
<td>MT</td>
<td>Milking temperament</td>
</tr>
<tr>
<td>N</td>
<td>Number of observations</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>NEB</td>
<td>Negative energy balance</td>
</tr>
<tr>
<td>NEFA</td>
<td>Non-esterified fatty acids</td>
</tr>
<tr>
<td>NR</td>
<td>Non-redundant</td>
</tr>
<tr>
<td>NRR</td>
<td>56-day non-return rate</td>
</tr>
<tr>
<td>NS</td>
<td>Number of services</td>
</tr>
<tr>
<td>PA</td>
<td>Parent average</td>
</tr>
<tr>
<td>PA_{rel}</td>
<td>Reliability of parent average</td>
</tr>
<tr>
<td>PCOS</td>
<td>Polycystic ovary syndrome</td>
</tr>
<tr>
<td>PEV</td>
<td>Prediction error variance</td>
</tr>
</tbody>
</table>
QTL  Quantitative trait loci
R²   Reliability
RETP Retained placenta
RYM  Region by year of birth by month of birth effect
SB   Stillbirth
SCR  Sire conception rate
SE   Standard error
SNP  Single Nucleotide Polymorphism
ssGBLUP Single-step Genomic Best Linear Unbiased Predictor
T    Technician
TMI  Total Merit Index
US   United States
Var  Variance
WssGWAS Weighted Single-step genome-wide association studies
δ    Polygenic weight
CHAPTER 1

General Introduction

1.1 Introduction

Since domestication, livestock species have undergone artificial selection, which was initially performed based on morphological and phenotypic records. Over the past few decades, selection for economically important traits have been based on more advanced statistical methods, such as Best Linear Unbiased Predictor (BLUP; Henderson, 1984), which consists of combining phenotypic and pedigree information of individuals and its relatives to generate estimated breeding values (EBV). In dairy cattle, genetic improvement was achieved with the use of progeny testing schemes in conjunction with artificial insemination (AI). However, this strategy has some drawbacks, such as the time required and costs associated with progeny tests (Schaeffer, 2006). Meuwissen et al. (2001) proposed the use of genomic selection (GS) as an alternative to traditional breeding program schemes and showed that breeding values for young animals could be predicted with accuracies as high as 0.85, which represented a gain of 110% compared to traditional methods. The proposed method relies on the assumption that, by using a dense marker map, markers will be very close to all quantitative trait loci (QTL) affecting a trait and in linkage disequilibrium (LD) with them.

In GS, a prediction equation estimated from a population with genotyped and phenotyped individuals (i.e., training population) is used to calculate genomic estimated breeding values (GEBV) for individuals without phenotypic information (Meuwissen et al. 2001). GEBV are calculated as the sum of effects of dense markers weighted by the genotype content. In this way, because markers are assumed to be in linkage disequilibrium with the QTL, most genetic variation of a given trait would be captured by those markers (Meuwissen et al., 2001). Schaeffer (2006)
suggested that conventional progeny-testing schemes should be modified (or replaced) by the use of GS. The author showed that if GEBV for both sexes could be calculated with high levels of accuracy at an early age, thus, due to a substantial reduction in generation intervals, annual genetic gain could be doubled and costs of proving a bull would be reduced by 92%. Multiple factors made dairy cattle a promising candidate for GS: large number of phenotypic and pedigree records collected over the years along with a well-designed evaluation infrastructure; longer generation interval compared to other livestock species; and breeding organization’s capability to invest in novel technologies (Wiggans et al., 2017).

In Canada, genomic evaluation for various traits has been successfully implemented in Holstein cattle and other dairy breeds since 2009 through a multi-step genomic BLUP (msGBLUP) approach. This procedure consists of running a traditional genetic evaluation based on only pedigree and phenotypic information. Then, traditional evaluations are de-regressed to make the data resemble individual records (Garrick et al., 2009). With that, direct genomic values (DGV) are obtained for genotyped animals, which are then blended with EBV or parent average (PA) to obtain the GEBV. However, the incorporation of genomic information through a multi-step procedure may be suboptimal. For instance, msGBLUP requires the use of de-regressed evaluations which can introduce biases and errors into the genomic evaluation due to losses of information (Legarra et al., 2014). Moreover, given that genotyped animals represent only a small portion of the total population being evaluated, bias can also be introduced by the fact that not all information used for selection decisions is included in the evaluation and thus not accounted for.

With the development of technologies such as frozen semen and embryos, the use of sires from international populations in dairy cattle breeding programs became routine and bulls started having progeny in several countries (Přibyl et al., 2014; Vandenplas and Gengler, 2015). However,
with genetic evaluations being performed based on phenotypic and pedigree information recorded in an internal level, i.e., within each country, the comparison of breeding values of individual bulls across countries cannot be performed. Several approaches have been developed to make comparison among domestic and foreign evaluations possible, such as computation of EBV for each country through the multiple across-country evaluation (MACE; Schaeffer, 1994). This EBV is also de-regressed for use in national evaluations through the msGBLUP method. Besides all possible problems associated with the use of de-regressed EBV (DD), i.e., loss of information, reduction in accuracies and biases (Legarra et al., 2014), double-counting of information and overestimation of reliabilities can occur if the model used for a combined evaluation of domestic and foreign data does not properly account for differences in DD reliabilities and daughters contributions to DD (Calus et al., 2016; Vandenplas et al., 2017).

Genomic selection has had a profound effect on genetic improvement of dairy cattle. Increased rates of genetic gain were expected as a result of higher intensity of selection, higher accuracy of genetic evaluations, and reduced generation intervals (Meuwissen et al., 2001; Schaeffer, 2006). The improvements achieved are especially large for traits with low heritability (Ponsart et al., 2013; García-Ruiz et al., 2016). In Canada and US Holstein populations, genetic gain has increased more than 50% for yield traits and 3-fold to 4-fold for lowly heritable traits, including female fertility, herd life and somatic cell score (García-Ruiz et al., 2016, Van Doormaal, 2017). Increased rates of genetic gain obtained, especially for lowly heritability traits, are a result of the increased accuracies of genetic evaluations due to inclusion of genomic information. In addition, A.I. companies are now able to genotype thousands of young candidate bulls before any selection decision is made, thus greatly increasing selection intensity and reducing generation interval. For instance, as a result of genomic selection, generation intervals were reduced from
about 7 to 2.5 years and from about 4 to 2.5 years for sire of bulls and dam of bulls pathways, respectively, in the US Holstein population, positively impacting the rates of genetic gain (García-Ruiz et al., 2016). The same trend was observed in Canada (Van Doormaal, 2014).

Besides being able to accurately select animals based on genomic breeding values, another promising aspect of GS is tracking and controlling of inbreeding rates. Prior to the introduction of GS this was done based on pedigree relationships. However, pedigree information is based on the expected proportion of the genome that is identical by descent (IBD) and does not capture variation due to Mendelian sampling and linkage during gamete formation. With genomics, it is now possible to estimate the actual proportion of the genome that is shared among 2 individuals (Hill and Weir, 2011). This should help breeders to more effectively select mating pairs and find a balance between the genetic value of the progeny and possible undesirable effects arising due to higher rates of inbreeding (Howard et al., 2017). Nevertheless, even though more young bulls are being screened by A.I. companies, average inbreeding levels among those bulls is the highest seen in the past years (Miglior and Beavers, 2014). Proper control of inbreeding levels is of extreme importance in dairy cattle breeding programs especially due to its detrimental effects on fitness traits, such as fertility and reproduction (Pryce et al., 2014). For instance, as a result of higher rates of inbreeding, increased frequencies of homozygous deleterious recessive alleles are observed in the population (Pryce et al., 2014). Previous studies have already identified genomic regions affecting female fertility in dairy cattle as a consequence of higher rates of inbreeding (e.g., VanRaden et al., 2011; Fritz et al 2013; Pryce et al., 2014; Adams et al., 2016). Genomic information enables the identification of such regions and, therefore, the implementation of mating programs, which take into account the probabilities of an animal being a carrier of recessive alleles, preventing known carrier males to be mated to known carrier females.
Along with higher inbreeding rates, many other factors have contributed to a general decline in functional traits, e.g. health, fertility and reproduction, such as intense selection for increased milk production and milk quality traits, increasing herd sizes, and greater use of confinement housing (Lucy, 2001; Chesnais et al., 2016). More recently, however, especially due to recent declines observed in fitness traits and animal health, there has been an increasing interest in including new functional traits in selection indexes, as they play an important role for the overall profitability of the production system and are directly related to animal welfare (Chesnais et al., 2016). In spite of substantial advancement in genomics, reliabilities of genomic predictions for functional traits are still low compared to those obtained for production traits (Egger-Danner et al., 2015). These reliabilities are highly dependent on the heritability of the trait being evaluated, statistical method used to estimate SNP effects in the training population, and, most importantly, the size of the training population (Goddard, 2009; Hayes et al., 2009; Hozé et al., 2014). For instance, traits with heritability close to 30%, require a training population size of more than 10,000 animals to achieve an accuracy of genomic prediction of 0.8, whereas around 4,000 animals are required for highly heritable traits (h² > 0.5; Hayes et al., 2009). In Canada, building a large training population for functional traits, which are often lowly heritable traits is challenging given that only progeny-tested bulls are included in the training population.

To overcome problems associated with the use of DD and small training populations, a new methodology that simultaneously uses genomic, pedigree, and phenotypic information of all individuals for genetic evaluation into one model was proposed (Misztal et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). This method is called single-step genomic BLUP (ssGBLUP) and consists of augmenting the pedigree relationship matrix with contributions from genomic relationships into a matrix of realized relationships. In this way, genotyped and non-
genotyped animals can be included simultaneously in the evaluation. As several steps are avoided, ssGBLUP is, in many instances, simpler to use and has the potential to deliver more accurate and less biased genomic evaluations. According to Legarra et al. (2014), the main advantage of ssGBLUP over multi-step methodologies, especially for dairy cattle, is its ability to account for genomic pre-selection. Moreover, it easily accommodates more complicated models, such as multiple trait and threshold models. Accuracy of ssGBLUP method is usually as high as, if not greater than other methods of evaluation (Aguilar et al., 2010; Christensen et al., 2012; Baloche et al., 2014; Zhang et al., 2016). In addition, benefits of ssGBLUP are greater for lowly heritable traits and those where information of animals with both phenotypes and genotypes is scarce, such as functional traits (Chen et al., 2011).

1.2 Thesis Outline

In order to better understand the impact of different methodologies on genomic evaluation, Chapters 2 and 3 focused on comparing genomic predictions obtained with ssGBLUP and msGBLUP methods in terms reliability and bias of genomic predictions for various functional traits. Multiple scenarios were tested for each methodology, such as adding cows’ genomic information into the training population, use of a blending approach for calculation of GEBV and different polygenic weights when blending the genomic and pedigree relationship matrices. We also applied a modification of the current ssGBLUP for genomic evaluation with integration of domestic and foreign information and compared the results obtained with those from msGBLUP.

Chapter 4 focused on implementation of genomic evaluation for three reproductive disorders using ssGBLUP. A genome-wide association study was also performed to identify
significant genomic regions associated with those traits and better describe the genetic mechanisms underlying the biology of reproductive disorders.

Chapter 5 evaluated the impact of two known recessive haplotypes in reproductive performance of Ayrshire cattle. Functional analyses were also performed aiming to better understand the biological mechanisms of the studied haplotypes.

1.3 Thesis Objectives

The overall goal of this thesis was to evaluate different strategies for combining genotypic, phenotypic, and pedigree information to analyze functional traits in dairy cattle. The main specific objectives of this thesis were:

1) To evaluate the impact of adopting a single-step approach for genomic evaluation of various workability and reproductive traits in Canadian Holstein cattle and compare the accuracy and bias of genomic predictions obtained with this method to those from multi-step method under the following strategies: 1) adding genomic information of cows in the analysis 2) testing different adjustments of the genomic relationship matrix, and 3) using a blending approach to obtain GEBV from the multi-step method.

2) To apply a single-step genomic BLUP approach that simultaneously evaluates domestic and foreign (MACE) information for genomic evaluation of workability traits in Canadian Holstein cattle; and compare reliability and bias of genomic predictions obtained by this methodology to those obtained by multi-step GBLUP.

3) To perform pedigree- and genomic-based analyses of three producer-recorded reproductive disorders, retained placenta (RETP), metritis (METR), and cystic ovaries (CYST), using traditional BLUP and ssGBLUP to assess the impact of including
genomic information into the genetic evaluation of these traits; to perform GWAS to prospect significant genomic regions and biological pathways associated with the three reproductive disorders; and explore the genetic mechanisms underlying these traits.

4) To assess if the recessive lethal haplotypes Ayrshire haplotype 1 (AH1) and Ayrshire haplotype 2 (AH2) are segregating in the Canadian Ayrshire cattle population; and investigate the possible deleterious effects of these haplotypes on stillbirth and 56-day non-return rate.
1.4 References


CHAPTER 2

Comparison of genomic predictions for lowly heritable traits using multi-step and single-step genomic BLUP in Holstein cattle


*Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada
†Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, USA
‡The Semex Alliance, Guelph, Ontario, Canada
§Canadian Dairy Network, Guelph, Ontario, Canada

2.1 Abstract

The success and sustainability of a breeding program incorporating genomic information is largely dependent on the accuracy of genomic predictions. For low heritability traits, large training populations are required in order to achieve high accuracies of genomic estimated breeding values (GEBV). By including genotyped and non-genotyped animals simultaneously in the evaluation, the single-step genomic BLUP (ssGBLUP) approach has the potential to deliver more accurate and less biased genomic evaluations. The aim of this study was to compare the accuracy and bias of genomic predictions for various traits in Canadian Holstein cattle using ssGBLUP and multi-step genomic BLUP (msGBLUP) under different strategies, such as 1) adding genomic information of cows in the analysis, 2) testing different adjustments of the genomic
relationship matrix, and 3) using a blending approach to obtain GEBV from msGBLUP. De-
regressed EBV for reproductive and workability traits were used as response variables. The
following genomic predictions were evaluated regarding accuracy and bias: 1) GEBV estimated
by ssGBLUP; 2) direct genomic value (DGV) estimated by msGBLUP (GBLUP_D) with polygenic
effects of 5% and 20%; and 3) GEBV calculated by a blending approach of DGV with EBV
(GBLUP_G) using polygenic effects of 5% and 20%. The impact of adding genomic information of
cows in the evaluation was also assessed for each approach. When genomic information was
included in the analyses, there was an average improvement in observed reliability of predictions
of 7 and 13 percentage points for reproductive and workability traits, respectively, compared to
traditional BLUP. Absolute deviation from 1 of the regression coefficient of the linear regression
of de-regressed EBV on genomic predictions went from 0.19 when using traditional BLUP to 0.22
when using the msGBLUP method, and to 0.14 when using the ssGBLUP method. The use of
polygenic weight of 20% in the msGBLUP slightly improved the reliability of predictions, while
reducing the bias. A similar trend was observed when a blending approach was used. Adding
genomic information of cows increased reliabilities, while decreasing bias of genomic predictions
when using the ssGBLUP method. Differences between using a training population with cows and
bulls or with only bulls for the msGBLUP method were small, likely due to the small number of
cows included in the analysis. Predictions for lowly heritable traits benefit greatly from genomic
information, especially when all phenotypes, pedigree, and genotypes are used in a single-step
approach.
2.2 Introduction

In Canada, genomic evaluation for a series of traits has been successfully implemented in Holstein cattle and other dairy breeds since 2009 through a multi-step genomic BLUP method (msGBLUP, Van Doormaal et al., 2009). This procedure consists of (1) running a traditional genetic evaluation based solely on pedigree and phenotypic information, (2) calculation of pseudo-phenotypes such as daughter deviations or de-regressed breeding values (DD) using results obtained from the previous step, (3) estimation of SNP effects to obtain direct genomic values (DGV) for genotyped animals, and (4) blending of genomic predictions with EBV or parent averages (PA). Loss of information during the de-regression step can introduce bias and errors into the genomic evaluation (Legarra et al., 2014), thus reducing the contribution of genomic information. Moreover, in multi-step evaluation, only information on genotyped animals is directly used. Given that only a small portion of the animals are genotyped, bias can also be introduced by the fact that not all information used for selection decisions is accounted for.

To overcome potential problems, Misztal et al. (2009), Aguilar et al. (2010), and Christensen and Lund (2010) developed the single-step genomic BLUP (ssGBLUP), a methodology that combines genotypic, pedigree, and phenotypic information into a single evaluation. It consists of augmenting the pedigree relationship matrix with contributions from genomic relationships into a matrix of realized relationships, the inverse of which is used in the BLUP mixed model equations. In this way, genotyped and non-genotyped animals can be included simultaneously in the evaluation. As several steps are avoided, ssGBLUP is, in many instances, simpler to use and has the potential to deliver more accurate and less biased genomic evaluations. According to Legarra et al. (2014), the main advantage of ssGBLUP over multi-step methodologies, especially for dairy cattle, is its ability to account for genomic pre-selection.
Accuracy of ssGBLUP method is usually as high as, if not greater than other methods of genetic merit evaluation (Aguilar et al., 2010; Christensen et al., 2012; Baloche et al., 2014; Lourenco et al., 2014a).

Accuracy of genomic predictions is critical for the expected genetic gains resulting from genomic selection and is dependent on many factors, such as heritability of the trait being evaluated, the statistical method used to estimate single nucleotide polymorphism (SNP) effects in the training population, and, most importantly, the size of the training population (Goddard, 2009; Hayes et al., 2009; Hozé et al., 2014). For the latter, the number of genotyped progeny-tested bulls could be a limitation, especially for lowly heritable traits or those traits that are difficult or expensive to measure. One way of overcoming the training population size problem is by incorporating genomic information of cows into the evaluation (Calus et al., 2013; Tsuruta et al., 2013; Uemoto et al., 2016). It is important, however, that cow and bull traditional evaluations are comparable to avoid a decrease in reliability. Wiggans et al. (2011) reported a decrease in reliability of genomic predictions with the inclusion of cows in the training population. The authors believe that many cows were subject to preferential treatment for having higher genetic merit. A pre-adjustment was then developed to reduce the mean and variance of cows’ traditional EBV so they would be comparable to those of bulls (Wiggans et al., 2012). Tsuruta et al. (2013) suggested that by using ssGBLUP, the inclusion of cows in the training population would be possible without any pre-adjustments. Although some countries have already adopted the inclusion of cows in their training populations, in Canada, genomic evaluations are currently based solely on bulls’ genomic information.

The objective of this study was to compare the accuracy and bias of genomic predictions for various workability and reproductive traits in Canadian Holstein cattle using multi-step and
single-step GBLUP methods under the following strategies: 1) adding genomic information of cows in the analysis, 2) testing different adjustments of the genomic relationship matrix, and 3) using a blending approach to obtain GEBV from the multi-step method.

2.3 Materials and Methods

2.3.1 Data

Genotypic data consisted of 10,590 bulls born between 1960 and 2012 genotyped with the Illumina Bovine SNP50 BeadChip (50K, Illumina Inc., San Diego, CA) and 6,842 cows born between 1997 and 2015 genotyped either with a 50K SNP panel or a low-density panel (6K), and then imputed to the 50K panel by using the FImpute software (Sargolzaei et al., 2014). Genotypes were coded as 0, 1 or 2 for calculation of the genomic relationship matrix (G). Genotype quality control excluded monomorphic SNPs, SNPs and individuals with call lower than 90%, SNPs that were out of Hardy-Weinberg equilibrium with very low probability (P-value < 10⁻⁶) or with minor allele frequency (MAF) less than 0.05, and individuals with parent-progeny Mendelian conflicts. After quality control, the number of genotyped animals retained was 17,430 and the final genotype dataset included 40,635 informative SNPs. PREGSF90 software was used for SNP and sample quality control (Misztal et al., 2002).

Data on Holstein reproductive and workability traits were extracted from the April 2017 genetic evaluation carried out by the Canadian Dairy Network (CDN, Guelph, Ontario, Canada, www.cdn.ca). The traits selected for this study were: milking speed (MS), milking temperament (MT), age at first insemination (AFS), days from calving to first insemination (CTFS), number of services (NS), 56-day non-return rate (NRR), days from first service to conception (FSTC), calving ease (CE), stillbirth (SB), gestation length (GL), and calf size (CZ). Traits that were
recorded during or before first calving were considered as heifer traits and were coded as parity 0. Traits measured in first-lactation cows were coded as parity 1 and the same approach was applied to subsequent lactations (parity > 1). The animal models used for genetic evaluation can be found in Jamrozik and Kistemaker (2016). Descriptive statistics for each trait are shown in Table 1.

De-regressed EBV were used as response variables for genomic prediction in the msGBLUP approach. De-regression was computed following Wiggans et al. (2011):

\[
\text{DE}_{\text{Prog}} = \left[ \frac{\text{EBV}_{\text{rel}}}{(1-\text{EBV}_{\text{rel}})} \right] - \text{DE}_{\text{PA}},
\]

\[
\text{DE}_{\text{PA}} = \frac{\text{PA}_{\text{rel}}}{(1-\text{PA}_{\text{rel}})},
\]

\[
\text{DD} = \text{PA} + \left[ \frac{\text{EBV}-\text{PA}}{\text{DE}_{\text{Prog}}/(\text{DE}_{\text{Prog}}+\text{DE}_{\text{PA}}+1)} \right],
\]

where PA is parent average, DE\text{PA} is daughter equivalent from PA, DE\text{Prog} is daughter equivalent from progeny information, PA\text{rel} is reliability of parent average, and EBV\text{rel} is reliability of EBV.

The pedigree file used for the analysis was generated by tracing the pedigree of animals with records for each trait up to four generations back (Lourenco et al., 2014b) and its size ranged from 2,343,158 animals for CTFS to 5,070,448 for MS.

2.3.2 Genomic Predictions

The following genomic predictions were evaluated with regarding accuracy and bias: GEBV estimated by ssGBLUP (\text{ssGEBV}); DGV estimated by msGBLUP (hereafter, called the GBLUPb) with either 5\% weight for the expected additive relationship matrix of genotyped individuals (A22) when combined with the genomic relationship matrix (DGV - 5\%A) or 20\% weight (DGV - 20\%A); and GEBV calculated by blending traditional EBV/PA with DGV with
either 5% (GEBV - 5%A) or 20% (GEBV - 20%A) weight for $A_22$ (hereafter, called the GBLUP$_B$ method). In addition, all analyses were carried out including or not cows’ genomic information. Genomic predictions were computed using the gebv software (Sargolzaei et al., 2009) for the msGBLUP method and BLUP90IOD2 program (Tsuruta et al., 2001) for the ssGBLUP method.

For this study, different datasets were prepared: 1) a full dataset containing all records of cows from which EBVs were calculated and de-regressed to be used as benchmarks; and, 2) a reduced dataset that included records of cows after deleting the last four years of observations. Furthermore, animals from the reduced dataset were then split into training and validation populations according to year of birth. Genotyped bulls with no daughters in the reduced dataset but with at least 50 daughters in the full dataset were used as validation bulls. Table 2 summarizes the number of bulls and cows in each dataset as well as the average reliability of DD for the msGBLUP method for each trait.

2.3.3 Statistical Models

The statistical models used for the analysis in different scenarios are described below.

2.3.3.1 Multi-step GBLUP Model. This approach consisted of four steps:

1. Estimation of breeding values using traditional BLUP;
2. De-regression of EBV;
3. Prediction of DGV;
4. Blending of DGV with PA/EBV.

Consider the following model:

$$y = 1\mu + Zg + e,$$
where \( \mathbf{y} \) is a vector of de-regressed evaluations of genotyped animals, \( \mathbf{1} \) is a vector of ones, \( \mu \) is the overall mean, \( \mathbf{Z} \) is a design matrix that allocates records to breeding values, \( \mathbf{g} \) is a vector of DGV to be estimated, and \( \mathbf{e} \) is a vector of random residual effects. It was assumed that \( \mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}\sigma_g^2) \), where \( \mathbf{G} \) is the genomic relationship matrix based on SNP markers, and \( \sigma_g^2 \) is the additive genetic variance; \( \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}\sigma_e^2) \), where \( \mathbf{I} \) is an identity matrix and \( \sigma_e^2 \) is the residual variance.

Here, DGV were estimated with either 5% or 20% weight (\( \delta \)) for the expected additive relationship matrix (\( \mathbf{A} \)) when combined with the genomic relationship matrix. Thus, the model becomes:

\[
\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g}_\delta + \mathbf{e},
\]

where, \( \mathbf{g}_\delta \sim \mathcal{N}(\mathbf{0}, \mathbf{G}_\delta\sigma_g^2) \), where \( \mathbf{G}_\delta \) is a combined relationship matrix, \( \mathbf{G}_\delta = \delta \mathbf{A} + (1 - \delta)\mathbf{G} \).

2.3.3.1 Blending. The blending approach used to generate GEBV was as follows:

\[
\text{GEBV} = w_1 \text{DGV} + w_2 \text{EST},
\]

\[
w_i = \frac{\text{Rel}_i}{\text{Rel}_i + \text{Rel}_j},
\]

where \( i = 1 \) and \( j = 2 \) for \( w_1 \), and \( i = 2 \) and \( j = 1 \) for \( w_2 \), \( w_1 \) is the reliability of DGV estimated from the multi-step GBLUP method, and \( w_2 \) is the reliability of EST, i.e. either EBV or PA from traditional BLUP.

2.3.3.2 Single-Step GBLUP Model. This method uses information from both genotyped and non-genotyped animals simultaneously. The mixed model equations then become:

\[
\begin{bmatrix}
\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\
\mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{H}^{-1}\alpha
\end{bmatrix}
\begin{bmatrix}
\mathbf{b} \\
\mathbf{u}
\end{bmatrix}
= \begin{bmatrix}
\mathbf{X}'\mathbf{y} \\
\mathbf{Z}'\mathbf{y}
\end{bmatrix},
\]

where \( \alpha \) is the variance ratio and \( \mathbf{H}^{-1} \) is derived as in Aguilar et al., (2010) and Christensen and Lund (2010):
\[ H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & (G^{-1} - \omega A_{22}^{-1}) \end{bmatrix}, \]

where \( A \) is the pedigree relationship matrix, \( G \) is the genomic relationship matrix, as \( 0.95G^* + 0.05A_{22} \) (to avoid singularity problems), and \( G^* \) is constructed as in VanRaden (2008); \( A_{22} \) is the pedigree relationship matrix among genotyped animals. Inbreeding was considered in the construction of all the relationship matrices. Differences between \( A_{22} \) and \( G \) may occur for several reasons that should be taken into account. For instance, incompleteness and shortness of pedigree, pedigree mistakes, incorrect assignment of genotypes, and poor quality of genotypes can cause such differences resulting in upward bias of genomic EBV for young animals and poor convergence when iteratively solving the mixed model equations for ssGBLUP (Misztal et al., 2013). To avoid convergence problems and bias, a weight (\( \omega \)), ranging from 1.0 to 0.6 for different traits, was given to \( A_{22}^{-1} \). All models described above were also used for analyses including (or not including) cows’ genomic information.

### 2.3.4 Validation

The reliability of genomic predictions for the studied traits were measured as squared correlation between the predicted breeding values (obtained with the reduced dataset) and de-regressed EBV (from the full dataset) in the validation datasets. In order to assess the bias (spread) of genomic predictions for each method, the following regression model was used:

\[ DD_C = b_0 + b_1 X_P + e, \]

where \( DD_C \) are de-regressed EBV, obtained from the complete dataset, of genotyped bulls with no daughters in the reduced dataset, but with at least 50 daughters in the complete dataset; \( b_0 \) is the
intercept; $b_1$ is the linear regression coefficient indicating bias of the predictions; $X_P$ is the bull’s PA or GEBV obtained with the reduced dataset; and $e$ is the residual.

### 2.4 Results and Discussion

Advantages of incorporating genomics into a breeding program are greatest for lowly heritable traits, especially due to the increase in reliabilities of predictions achieved for these traits by using genomic information (García-Ruiz et al., 2016; Wiggans et al., 2017). García-Ruiz et al. (2016) reported substantial improvements in genetic trends and consequently rapid genetic improvement for traits with low heritability. In the present study, when genomic information was included in the analyses, there was an average improvement in observed reliability of predictions of 7 and 13 percentage points for reproductive traits and workability traits, respectively, compared to traditional BLUP (Table 3). Moreover, averaged over the 18 traits, the absolute deviation from 1 of the regression coefficient of the linear regression of de-regressed EBV on genomic predictions went from 0.19 when using traditional BLUP to 0.22 when using msGBLUP, and to 0.14 when using ssGBLUP (Table 4). Regression coefficients ranged from 0.51 to 1.35 for msGBLUP, while for ssGBLUP, the coefficients ranged from 0.66 to 1.04 (Table 4).

Regression coefficient values close to 1 suggest that 1-unit difference in the genomic evaluation results in a 1-unit change in the trait and, therefore, they are considered to be on the same scale (Sullivan, 2009; Wiggans et al., 2011). A regression coefficient less than 1 indicates inflation of the variance of genomic predictions, while deflation occurs when the coefficient is larger than 1. Traditional BLUP models consider all information available in the evaluation, i.e., phenotypes and pedigree, and, therefore, less bias derived from selection is observed (Sorensen and Kennedy, 1984). Bias on multi-step genomic predictions may arise from the fact that genomic
prediction models do not simultaneously fit all data used for selection decisions and, therefore, must assume no selection in the population being evaluated. Yet, in dairy cattle, a large number of genotyped populations usually consist of progeny-tested bulls, which are highly selected animals (Hayes et al., 2009; Vitezica et al., 2011).

By including all information available in the evaluation simultaneously, i.e., phenotypes, genotypes and pedigree, ssGBLUP is able to partially account for pre-selection, thus minimizing bias of predictions (Legarra et al., 2014). Gao et al. (2012) applied three multi-step GBLUP and two single-step blending methods for genomic prediction in Nordic Holsteins. In their study, the two single-step methods led to less bias than the other GBLUP methods. Similarly, Ma et al. (2015) showed that regression coefficients went from 0.69 when using a multi-step GBLUP to 0.78 for the single-step GBLUP method. In the present study, averaged over all traits, ssGBLUP resulted in regression coefficients with the smallest absolute deviation from 1. For reproductive and workability traits, the deviation of the regression coefficient from 1 for ssGBLUP was equal to 0.11 and 0.19, respectively, while for msGBLUP, this deviation was 0.21 and 0.25, respectively. The single-step approach also resulted in a slight increase in reliability of genomic predictions, compared to those of msGBLUP.

Historically, ssGBLUP could not converge or had a large bias in some cases, and one measure to improve on both counts was a weight parameter ($\omega$) given to $A_{22}^{-1}$. Such a parameter was shown to account for missing inbreeding, especially when some pedigrees were missing (Misztal et al., 2017). When inbreeding was accounted for in $A$ including in phantom parent groups, the $b_1$ parameter improved from 0.75 to 0.90. Accounting for reduced heritability due to genetic selection improved the parameter $b_1$ to 0.96 and slightly increased accuracy. These
measures indicated that special modeling is required to account for selection practices under genomic selection (Misztal et al., 2017).

Obtaining unbiased predictions is of fundamental importance for an accurate ranking and fair comparison of animals across generations (Patry and Ducrocq, 2009; Aguilar et al., 2010). Gao et al. (2012) showed that absolute deviations of the regression coefficients from 1 tended to decrease with increasing polygenic weights in a GBLUP model, which is consistent with the literature (Aguilar et al., 2010; Liu et al., 2011). This could be explained by the fact that genetic markers are not able to explain the total genetic variance of a given trait. Thus, the polygenic effect would account for the residual genetic variance, which is not accounted for by using only genetic markers. In the present study, the polygenic weight of 20% in the msGBLUP slightly improved the observed reliability of predictions while reducing the bias. Averaged over all traits, the deviation of the regression coefficient from 1 for GBLUP_D went from 0.28 when using a 5% polygenic weight to 0.24 with a 20% polygenic weight.

Alternatively, one way to account for the residual genetic variance and ensure that all polygenes are considered is by adopting a blending approach (Sullivan, 2009). In Canada, when using this blending approach, GEBV are calculated by combining EBV obtained from traditional pedigree-based (BLUP) evaluations with DGV from GBLUP evaluations through a selection index, weighted by their reliabilities (Van Doormaal et al., 2009). Predictions obtained with GBLUP_B were, on average, less biased and slightly more reliable than those obtained with GBLUP_D. When a polygenic weight of 5% was used, reliabilities ranged from 0.037 to 0.454 (average: 0.206) and from 0.042 to 0.464 (average: 0.217), for GBLUP_D and GBLUP_B, respectively, while the range went from 0.040 to 0.464 (average: 0.21) and from 0.042 to 0.463 (average: 0.218), for GBLUP_D and GBLUP_B, respectively, when using 20% polygenic weight.
(Table 3). These results are in agreement with Lourenco et al. (2014). The authors reported higher reliabilities for GEBV compared to DGV for milk yield, fat and protein percentages. Conversely, Su et al. (2012) reported higher reliabilities for DGV compared to GEBV for various traits of Nordic Red Cattle. However, according to the authors the reliability of DGVs could be overestimated, given the fact that different scaling factors were used as weights for the genomic relationship matrices in each method.

In the present study, additionally to the increase in reliability caused by blending, the GEBV bias was reduced. For a polygenic weight of 5%, the deviation of the regression coefficient from 1 was 0.28 and 0.19 for GBLUP$_D$ and GBLUP$_B$, respectively, whereas for predictions obtained with a 20% polygenic weight, this deviation was 0.24 for GBLUP$_D$ and 0.18 for GBLUP$_B$ (Table 4). It is important to point out that traditional evaluations might be biased by pre-selection, which increased after a wider use of genomic selection. Commonly, only animals selected based on their GEBV will have their progeny’s phenotypes recorded. However, such information is not included in the evaluation and, therefore, pre-selection is not properly accounted for; this will lead to biased pedigree-based predictions, i.e., EBV (Patry and Ducrocq, 2011) and, consequently, affect the blending approach. The use of all information available simultaneously to predict GEBV could potentially help avoid such bias.

In dairy cattle genomic evaluation, training populations are usually composed of progeny-tested bulls, as those have more reliable information. However, for low heritable traits, such as health and fertility, a very large training population is required in order to achieve high accuracies of predictions (Hayes et al., 2009), which could be a challenge, given the fact that progeny-tested bulls are limited for those traits. Moreover, progeny-tested bulls are intensively selected, which could potentially lead to loss of genetic variability. The inclusion of cows’ genomic data in training
populations is, somehow, appealing, as they can contribute to increase the size of a training population and are also of major importance during the selection decision process (Buch et al., 2011; Wiggans et al., 2011; Tsuruta et al., 2013; Ma et al., 2015).

The benefits of adding cows’ genomic information into the training populations may depend on the strategy adopted (Tsuruta et al., 2013; Gao et al., 2015). When using a multi-step approach, accuracies of genomic predictions are contingent on the accuracies of de-regressed EBV used and for the cows’ EBV lower accuracies are expected compared to those of bulls. In the present study, differences between using a training population with cows and bulls or with bulls only for the msGBLUP methods were small (Tables 3 and 4). This is likely due to the small number of cows included in the analyses, and their low reliabilities (Table 2). However, for ssGBLUP the inclusion of cows’ information had a positive effect on genomic predictions. Across all traits, the single-step approach including bulls and cows’ genomic information led to the highest reliability of genomic predictions, followed by ssGBLUP with bulls only in the training population, and msGBLUP methods. On average, reliabilities of ssGBLUP including bulls and cows’ genomic information were 1.21% higher than those of msGBLUP methods with a bulls-only training population, which is the method currently used for Canadian genomic evaluations. Tsuruta et al. (2013) also reported an increase in reliability of 1.9% to 2.7% when cow genotypes were added in the evaluation using ssGBLUP.

The inclusion of cows’ genomic information can also help reduce the bias of genomic predictions under a single-step approach. In the present study, the average deviation of the regression coefficient from 1 was 0.18 and 0.19 for GBLUPB for bulls and cows, and bulls only in the training population, respectively. However, when using the single-step approach this deviation decreased to 0.12 when bulls and cows’ genomic information was used for the analysis and when
only bulls’ genomic information was used, the average deviation across all traits was 0.16. The use of a single-step approach also resulted in a slightly increased mean reliability (0.221) compared to GBLUPB (0.217).

The average deviation from 1 for msGBLUP was 0.22 and 0.25 for reproductive and workability traits, respectively, when only bulls were included in the training, while for ssGBLUP the deviation from 1 was 0.13 and 0.21 for reproductive and workability traits, respectively. When both cows and bulls were included in the training population the average deviation for msGBLUP was 0.20 and 0.25 for reproductive and workability traits, respectively, and was 0.09 and 0.16 for ssGBLUP, respectively. The results obtained for regression coefficients are in agreement with those of previous findings (Tsuruta et al., 2013; Ma et al., 2015; Uemoto et al., 2016).

### 2.5 Conclusions

Single-step genomic BLUP can reduce bias while slightly increasing observed reliability of genomic predictions. Predictions obtained with the single-step approach by adding cows’ genomic information showed a small gain in reliability and a reduction in bias compared to those obtained by only including bulls’ genomic information. For the multi-step GBLUP method, no effects on reliability or bias of predictions were observed when genomic information of cows was considered. This may be explained by the fact that only a small number of cows were included in the analyses. The inclusion of a polygenic effect in the multi-step analysis as well as the use of a blending approach had a small effect on reliabilities of predictions, but an important effect on bias. However, the improvements obtained with different strategies adopted in the multi-step method were smaller compared to the use of the single-step approach.
2.6 Acknowledgements

This research was supported in main part by Agriculture and Agri-Food Canada and by additional contributions from Dairy Farmers of Canada, the Canadian Dairy Network and the Canadian Dairy Commission under the Agri-Science Clusters Initiative. The first author is also thankful to CAPES (Brazilian Federal Agency for Support and Evaluation of Graduate Education, Brazil) for financial support.
## 2.7 Tables

**Table 2.1.** Descriptive statistics and heritability ($h^2$) of reproductive and workability traits in first and later parities

<table>
<thead>
<tr>
<th>Parity</th>
<th>Trait$^1$</th>
<th>$h^2$</th>
<th>N. records</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reproductive traits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First</td>
<td>AFS</td>
<td>0.09</td>
<td>1,217,701</td>
<td>478.97</td>
<td>55.81</td>
<td>274</td>
<td>639</td>
</tr>
<tr>
<td></td>
<td>NRR</td>
<td>0.03</td>
<td>1,404,619</td>
<td>0.72</td>
<td>0.45</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>0.02</td>
<td>1,420,060</td>
<td>1.62</td>
<td>1.04</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>FSTC</td>
<td>0.03</td>
<td>1,086,405</td>
<td>18.25</td>
<td>33.64</td>
<td>0</td>
<td>206</td>
</tr>
<tr>
<td></td>
<td>GL</td>
<td>0.19</td>
<td>1,000,420</td>
<td>278.78</td>
<td>6.41</td>
<td>240</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>CE</td>
<td>0.12</td>
<td>1,075,704</td>
<td>1.57</td>
<td>0.68</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>SB</td>
<td>0.11</td>
<td>1,068,520</td>
<td>0.89</td>
<td>0.31</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CZ</td>
<td>0.11</td>
<td>1,040,971</td>
<td>2.01</td>
<td>0.57</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>NRR</td>
<td>0.04</td>
<td>3,771,204</td>
<td>0.55</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>0.06</td>
<td>3,815,035</td>
<td>2.31</td>
<td>1.62</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Later</td>
<td>CTFS</td>
<td>0.07</td>
<td>3,803,907</td>
<td>86.95</td>
<td>31.54</td>
<td>14</td>
<td>244</td>
</tr>
<tr>
<td></td>
<td>FSTC</td>
<td>0.08</td>
<td>2,545,203</td>
<td>37.34</td>
<td>47.98</td>
<td>0</td>
<td>206</td>
</tr>
<tr>
<td></td>
<td>GL</td>
<td>0.16</td>
<td>2,359,727</td>
<td>279.88</td>
<td>6.54</td>
<td>240</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>CE</td>
<td>0.08</td>
<td>2,747,813</td>
<td>1.40</td>
<td>0.59</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>SB</td>
<td>0.01</td>
<td>2,733,684</td>
<td>0.95</td>
<td>0.22</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CZ</td>
<td>0.09</td>
<td>2,704,295</td>
<td>2.19</td>
<td>0.58</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Workability traits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First$^2$</td>
<td>MT</td>
<td>0.13</td>
<td>2,921,775</td>
<td>3.29</td>
<td>0.75</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>0.14</td>
<td>3,573,346</td>
<td>3.12</td>
<td>0.70</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

1Age at first service (AFS); calf size (CZ); calving ease (CE); days from calving to first service (CTFS); days from first service to conception (FSTC); gestation length (GL); milking speed (MS); milking temperament (MT); 56-day non-return rate (NRR); number of services (NS); stillbirth (SB).

2For both MS and MT, scoring is made only once during the first 6 months of first lactation.
### Table 2.2. Structure of training and validation datasets for multi-step genomic BLUP (msGBLUP) and single-step genomic BLUP (ssGBLUP) methods

<table>
<thead>
<tr>
<th>Parity</th>
<th>Trait(^1)</th>
<th>Validation population</th>
<th>Training population</th>
<th>ssGBLUP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N(^2)</td>
<td>Birth year</td>
<td>msGBLUP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N(^3)</td>
<td>Rel(_{DD}) (%)(^4)</td>
<td>N</td>
</tr>
<tr>
<td>Reproductive traits</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First</td>
<td>AFS</td>
<td>508</td>
<td>2009 - 2011</td>
<td>5,365</td>
</tr>
<tr>
<td></td>
<td>NRR</td>
<td>317</td>
<td>2009 - 2011</td>
<td>4,785</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>241</td>
<td>2009 - 2011</td>
<td>5,146</td>
</tr>
<tr>
<td></td>
<td>FSTC</td>
<td>209</td>
<td>2008 - 2010</td>
<td>4,126</td>
</tr>
<tr>
<td></td>
<td>GL</td>
<td>418</td>
<td>2009 - 2011</td>
<td>5,111</td>
</tr>
<tr>
<td></td>
<td>CE</td>
<td>438</td>
<td>2009 - 2011</td>
<td>5,359</td>
</tr>
<tr>
<td></td>
<td>SB</td>
<td>438</td>
<td>2009 - 2011</td>
<td>5,330</td>
</tr>
<tr>
<td></td>
<td>CZ</td>
<td>420</td>
<td>2009 - 2011</td>
<td>5,316</td>
</tr>
<tr>
<td></td>
<td>NRR</td>
<td>489</td>
<td>2009 - 2011</td>
<td>5,917</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>570</td>
<td>2008 - 2010</td>
<td>5,568</td>
</tr>
<tr>
<td></td>
<td>CTFS</td>
<td>568</td>
<td>2008 - 2010</td>
<td>5,568</td>
</tr>
<tr>
<td></td>
<td>FSTC</td>
<td>406</td>
<td>2007 - 2009</td>
<td>5,060</td>
</tr>
<tr>
<td></td>
<td>GL</td>
<td>342</td>
<td>2008 - 2010</td>
<td>5,416</td>
</tr>
<tr>
<td></td>
<td>CE</td>
<td>363</td>
<td>2008 - 2010</td>
<td>6,089</td>
</tr>
<tr>
<td></td>
<td>SB</td>
<td>285</td>
<td>2008 - 2010</td>
<td>6,040</td>
</tr>
<tr>
<td></td>
<td>CZ</td>
<td>350</td>
<td>2008 - 2010</td>
<td>6,060</td>
</tr>
<tr>
<td>Later</td>
<td>NRR</td>
<td>489</td>
<td>2009 - 2011</td>
<td>5,917</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>570</td>
<td>2008 - 2010</td>
<td>5,568</td>
</tr>
<tr>
<td></td>
<td>CTFS</td>
<td>568</td>
<td>2008 - 2010</td>
<td>5,568</td>
</tr>
<tr>
<td></td>
<td>FSTC</td>
<td>406</td>
<td>2007 - 2009</td>
<td>5,060</td>
</tr>
<tr>
<td></td>
<td>GL</td>
<td>342</td>
<td>2008 - 2010</td>
<td>5,416</td>
</tr>
<tr>
<td></td>
<td>CE</td>
<td>363</td>
<td>2008 - 2010</td>
<td>6,089</td>
</tr>
<tr>
<td></td>
<td>SB</td>
<td>285</td>
<td>2008 - 2010</td>
<td>6,040</td>
</tr>
<tr>
<td></td>
<td>CZ</td>
<td>350</td>
<td>2008 - 2010</td>
<td>6,060</td>
</tr>
<tr>
<td>Workability traits</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First(^6)</td>
<td>MT</td>
<td>554</td>
<td>2009 - 2011</td>
<td>7,047</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>554</td>
<td>2009 - 2011</td>
<td>7,369</td>
</tr>
</tbody>
</table>

\(^1\)Age at first service (AFS); calf size (CZ); calving ease (CE); days from calving to first service (CTFS); days from first service to conception (FSTC); gestation length (GL); milking speed (MS); milking temperament (MT); 56-day non-return rate (NRR); number of services (NS); stillbirth (SB).

\(^2\)Only bulls with at least 50 daughters were considered in the validation population.

\(^3\)Only bulls with at least 10 daughters were considered in the training population.

\(^4\)Average reliability of de-regressed EBV in the training population.

\(^5\)Number of animals with phenotypes and number of animals with both phenotypes and genotypes.

\(^6\)For both MS and MT, scoring is made only once during the first 6 months of first lactation.
<table>
<thead>
<tr>
<th>Parity</th>
<th>Trait (^2)</th>
<th>Average Reliability of DDe(^3)</th>
<th>PA</th>
<th>Observed reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DGV - 5%A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bulls</td>
<td>Bulls and Cows</td>
</tr>
<tr>
<td>Reproductive traits</td>
<td></td>
<td></td>
<td>Bulls</td>
<td>Bulls and Cows</td>
</tr>
<tr>
<td>First</td>
<td>AFS</td>
<td>0.712</td>
<td>0.129</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td>NRR</td>
<td>0.389</td>
<td>0.045</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>0.516</td>
<td>0.103</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>FSTC</td>
<td>0.331</td>
<td>0.038</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>GL</td>
<td>0.797</td>
<td>0.280</td>
<td>0.355</td>
</tr>
<tr>
<td></td>
<td>CE</td>
<td>0.758</td>
<td>0.155</td>
<td>0.237</td>
</tr>
<tr>
<td></td>
<td>SB</td>
<td>0.575</td>
<td>0.131</td>
<td>0.302</td>
</tr>
<tr>
<td></td>
<td>CZ</td>
<td>0.718</td>
<td>0.149</td>
<td>0.243</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>0.599</td>
<td>0.129</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>NRR</td>
<td>0.326</td>
<td>0.021</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>0.681</td>
<td>0.161</td>
<td>0.237</td>
</tr>
<tr>
<td></td>
<td>CTFS</td>
<td>0.731</td>
<td>0.138</td>
<td>0.187</td>
</tr>
<tr>
<td></td>
<td>FSTC</td>
<td>0.474</td>
<td>0.093</td>
<td>0.155</td>
</tr>
<tr>
<td></td>
<td>GL</td>
<td>0.808</td>
<td>0.315</td>
<td>0.454</td>
</tr>
<tr>
<td></td>
<td>CE</td>
<td>0.695</td>
<td>0.092</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>SB</td>
<td>0.328</td>
<td>0.067</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>CZ</td>
<td>0.722</td>
<td>0.180</td>
<td>0.325</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>0.596</td>
<td>0.133</td>
<td>0.200</td>
</tr>
<tr>
<td>Workability traits</td>
<td></td>
<td></td>
<td>Bulls</td>
<td>Bulls and Cows</td>
</tr>
<tr>
<td>First</td>
<td>MT</td>
<td>0.797</td>
<td>0.106</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>0.815</td>
<td>0.219</td>
<td>0.385</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>0.806</td>
<td>0.163</td>
<td>0.285</td>
</tr>
</tbody>
</table>

\(^1\)GEBV using a blending approach with a polygenic effect with a weight of 0.05 (GEBV - 5\%A); GBLUP using a blending approach with a polygenic effect with a weight of 0.2 (GEBV - 20\%A); GBLUP with a polygenic effect with a weight of 0.05 (DGV - 5\%A); GBLUP with a polygenic effect with a weight of 0.2 (DGV - 20\%A); Single-Step GBLUP (ssGEBV).

\(^2\)Age at first service (AFS); calf size (CZ); calving ease (CE); days from calving to first service (CTFS); days from first service to conception (FSTC); gestation length (GL); milking speed (MS); milking temperament (MT); 56-day non-return rate (NRR); number of services (NS); stillbirth (SB).

\(^3\)Average reliability of de-regressed EBV obtained from complete dataset for validation population.

\(^4\)For both MS and MT, scoring is made only once during the first 6 months of first lactation.
Table 2.4. Regression coefficient of the linear regression of de-regressed EBV (DDc) on genomic predictions (GEBV) or on standard BLUP evaluations (PA) obtained using different methods

<table>
<thead>
<tr>
<th>Parity</th>
<th>Trait</th>
<th>Average Reliability of DDc</th>
<th>DGV - 5%A Bulls and Cows</th>
<th>GEBV - 5%A Bulls and Cows</th>
<th>DGV - 20%A Bulls and Cows</th>
<th>GEBV - 20%A Bulls and Cows</th>
<th>SsGEBV Bulls and Cows</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reproductive traits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First</td>
<td>AFS</td>
<td>0.71</td>
<td>0.79</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>NRR</td>
<td>0.39</td>
<td>0.85</td>
<td>0.64</td>
<td>0.63</td>
<td>0.69</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>0.61</td>
<td>0.76</td>
<td>0.53</td>
<td>0.51</td>
<td>0.57</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>FSTC</td>
<td>0.33</td>
<td>0.76</td>
<td>1.25</td>
<td>1.21</td>
<td>1.29</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td>GL</td>
<td>0.80</td>
<td>0.95</td>
<td>0.85</td>
<td>0.84</td>
<td>0.91</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>CE</td>
<td>0.76</td>
<td>0.84</td>
<td>0.71</td>
<td>0.72</td>
<td>0.76</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>SB</td>
<td>0.66</td>
<td>0.96</td>
<td>1.09</td>
<td>1.09</td>
<td>1.14</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>CZ</td>
<td>0.72</td>
<td>0.96</td>
<td>0.91</td>
<td>0.90</td>
<td>0.97</td>
<td>1.15</td>
</tr>
<tr>
<td>Mean deviation from 1</td>
<td>0.62</td>
<td>0.14</td>
<td>0.26</td>
<td>0.26</td>
<td>0.24</td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td>Later</td>
<td>NRR</td>
<td>0.33</td>
<td>0.97</td>
<td>1.05</td>
<td>1.04</td>
<td>1.13</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>0.68</td>
<td>0.83</td>
<td>0.78</td>
<td>0.78</td>
<td>0.97</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>CTFS</td>
<td>0.77</td>
<td>0.61</td>
<td>0.56</td>
<td>0.56</td>
<td>0.68</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>FSTC</td>
<td>0.47</td>
<td>1.00</td>
<td>0.92</td>
<td>0.92</td>
<td>1.12</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>GL</td>
<td>0.83</td>
<td>0.88</td>
<td>0.89</td>
<td>0.89</td>
<td>1.00</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>CE</td>
<td>0.69</td>
<td>0.74</td>
<td>0.57</td>
<td>0.56</td>
<td>0.76</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>SB</td>
<td>0.50</td>
<td>0.79</td>
<td>0.69</td>
<td>0.68</td>
<td>0.81</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>CZ</td>
<td>0.72</td>
<td>0.78</td>
<td>0.73</td>
<td>0.73</td>
<td>0.86</td>
<td>0.90</td>
</tr>
<tr>
<td>Mean deviation from 1</td>
<td>0.62</td>
<td>0.18</td>
<td>0.24</td>
<td>0.24</td>
<td>0.17</td>
<td>0.20</td>
<td>0.16</td>
</tr>
<tr>
<td>Workability traits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First</td>
<td>MT</td>
<td>0.80</td>
<td>0.54</td>
<td>0.54</td>
<td>0.54</td>
<td>0.63</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>0.81</td>
<td>0.97</td>
<td>0.79</td>
<td>0.79</td>
<td>0.99</td>
<td>0.84</td>
</tr>
<tr>
<td>Mean deviation from 1</td>
<td>0.81</td>
<td>0.25</td>
<td>0.34</td>
<td>0.34</td>
<td>0.19</td>
<td>0.19</td>
<td>0.29</td>
</tr>
</tbody>
</table>

1. GBLUP using a blending approach with a polygenic effect with a weight of 0.05 (GEBV - 5% A); GBLUP using a blending approach with a polygenic effect with a weight of 0.2 (GEBV - 20% A); GBLUP with a polygenic effect with a weight of 0.05 (DGV - 5% A); GBLUP with a polygenic effect with a weight of 0.2 (DGV - 20% A); Single-Step GBLUP (ssGEBV).
2. Age at first service (AFS); calf size (CZ); calving ease (CE); days from calving to first service (CTFS); days from first service to conception (FSTC); gestation length (GL); milking speed (MS); milking temperament (MT); 56-day non-return rate (NRR); number of services (NS); stillbirth (SB).
3. Average reliability of de-regressed EBV in the validation population.
4. Mean of absolute deviation of the regression coefficient from 1.
5. For both MS and MT, scoring is made only once during the first 6 months of first lactation.
2.8 References


CHAPTER 3

Use of a single-step approach for integrating foreign information into national genomic evaluation in Holstein cattle

A.R. Guarini*, 1, D.A.L. Lourenco†, L.F. Brito*, M. Sargolzaei*, ‡, C.F. Baes*, F. Miglior, *, §, S. Tsuruta†, I. Misztal†, and F.S. Schenkel*

*Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada
†Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, USA
‡The Semex Alliance, Guelph, Ontario, Canada
§Canadian Dairy Network, Guelph, Ontario, Canada

3.1 Abstract

The use of multi-trait across-country evaluation (MACE) and the exchange of genomic information among countries allows national breeding programs to combine foreign and national data to increase the size of the training populations and potentially increase accuracy of genomic prediction of breeding values. By including genotyped and non-genotyped animals simultaneously in the evaluation, the single-step GBLUP (ssGBLUP) approach has the potential to deliver more accurate and less biased genomic evaluations. A single-step genomic BLUP approach, which enables integration of data from MACE evaluations can be used to obtain genomic predictions while avoiding double-counting of information. The objectives of this study were to apply a single-step approach that simultaneously includes domestic and external (MACE) information for
genomic evaluation of workability traits in Canadian Holstein cattle, and compare the results obtained with this methodology to those obtained using a multi-step approach (msGBLUP). By including MACE bulls in the training population, msGBLUP led to an increase in reliability of genomic predictions of 4.8% and 15.4% for milking temperament (MT) and milking speed (MS), respectively, compared to a traditional evaluation using only pedigree and phenotypic information. Integration of MACE data through a single-step approach (ssGBLUPIM) yielded the highest reliabilities compared to other methods considered. Integration of MACE data also helped reduce bias of genomic predictions. When using ssGBLUPIM, the bias of genomic predictions decreased by half compared to msGBLUP using domestic and MACE information. Therefore, the reliability and bias of genomic predictions for both traits improved substantially when a single-step approach was used for evaluation compared to a multi-step approach.

3.2 Introduction

In dairy cattle breeding programs, the use of sires from international populations has become common since the development of technologies such as frozen semen and embryo transfers (Vandenplas and Gengler, 2015). When breeding programs rely on the use of foreign bulls, predictions will likely be biased if foreign pedigree, performance records, genetic merit, and different genetic levels between base populations in each country are not taken into account for domestic evaluations (Bonaiti and Boichard, 1995). With the implementations of multi-trait across-country evaluation (MACE; Schaeffer, 1994) performed by Interbull (Uppsala, Sweden) and the exchange of sire genotypes, national breeding programs are now able to properly aggregate foreign information into domestic evaluations in order to increase their training populations and consequently obtain more accurate genomic predictions. In Canada, genomic evaluations have been successfully implemented in Holstein cattle and other dairy breeds through a multi-step
GBLUP method since 2009 (msGBLUP, Van Doormaal et al., 2009). When MACE information is available, foreign data are added to the evaluations in an extra step after domestic (national) evaluations are performed (Figure 1). Although national breeding programs perform such evaluations routinely, several critical assumptions required by multi-step procedures are not fully met, hence, there are issues associated with them. Since genomic selection (GS) was implemented in 2009, selection intensity increased; by using models that do not properly account for the existing pre-selection of imported bulls, predictions become less accurate and more biased (Vitezica et al., 2011).

As phenotypic data of foreign bulls cannot be used, multi-step approaches rely on pseudo-phenotypes, i.e., deregressed estimated breeding values (DD). However, the use of such pseudo-phenotypes could lead to loss of information, reduction in accuracies and increase in biases (Legarra et al., 2014). Moreover, one assumption for the use of DD is independence. For instance, highly reliable estimated breeding values (EBV) have more contributions from daughters’ performance than from relatives, thus, de-regression is expected to have less impact on estimates. On the other hand, de-regression is expected to have a significant impact on EBVs with substantially lower reliabilities, i.e., those with more contributions from relatives (Calus et al., 2016). If the model does not properly account for differences in reliabilities and contributions, double-counting of information and overestimation of reliabilities can occur (Calus et al., 2016; Vandenplas et al., 2017).

Simultaneous use of pedigree, phenotypic, and genomic information in a single-step approach should theoretically overcome problems arising with the use of DD and consequent reduction in accuracies and biases. Legarra et al. (2009), Aguilar et al. (2010), and Christensen and Lund (2010) developed the single-step genomic BLUP (ssGBLUP), a methodology that
combines genotypes, pedigree, and phenotypes into a single evaluation. However, because international phenotypic data are not available at the domestic level, the standard ssGBLUP is not able to integrate MACE and domestic data into a joint evaluation. Legarra et al. (2007) proposed a modification to the traditional BLUP that simultaneously integrates data from foreign evaluations and considers those as priors for calculations of domestic EBVs. This modification can be further extended to ssGBLUP to integrate MACE data into a domestic single-step genomic evaluation.

The objectives of this study were to (1) apply a modification of the current single-step genomic BLUP that simultaneously evaluates domestic and external (MACE) information for genomic evaluation of workability traits in Canadian Holstein cattle, and (2) compare reliability and bias of genomic predictions obtained by this methodology to those obtained by multi-step GBLUP approach.

3.3 Materials and Methods

3.3.1 Data

3.3.1.1 Genotypes. Genotype data consisted of 33,201 Holstein bulls born between 1960 and 2011 generated using the Illumina Bovine SNP50 BeadChip (50K, Illumina Inc., San Diego, CA) or a greater density panel and then imputed to 50K (non-overlapping SNPs). Genotypes were coded as 0, 1 or 2 for calculation of the genomic relationship matrix (G). Genotype quality control excluded monomorphic SNPs, SNPs and individuals with call rate lower than 90%, SNPs that were out of Hardy-Weinberg equilibrium with very low probability (P-value $< 10^{-6}$) or with minor allele frequency (MAF) less than 0.05, and individuals with parent-progeny Mendelian conflicts. After quality control, the number of genotyped animals was 33,196 and the final genotype dataset included 40,894 SNPs.
3.3.1.2 Phenotypes. A total of 3,173,169 records for milking speed (MS) and 2,519,932 for milking temperament (MT) in first-lactation Holstein cows were extracted from the phenotypes used in the April 2017 genetic evaluation carried by Canadian Dairy Network (CDN, Guelph, Ontario, Canada, www.cdn.ca).

3.3.1.3 Domestic and MACE EBVs. Official relative breeding values (RBV), i.e., domestic traditional (hereafter identified by subscript D) and MACE EBVs (subscript M) for bulls registered in the CDN database were obtained from the April 2013 genetic evaluation for workability traits of Holstein cows. In Canada, EBV are standardized to RBV with a mean of 100 and standard deviation of 5. In order to use MACE EBVs in ssGBLUP evaluations, RBV were transformed back to EBV as:

\[
EBV = \frac{SD \times (RBV - 100)}{5} + \bar{x}
\]

where SD and \(\bar{x}\) are the actual standard deviation and mean of the trait, respectively.

3.3.2 Methods of Evaluation

This section describes the fundamental concepts behind (1) the standard ssGBLUP, (2) the modification of the standard ssGBLUP for simultaneously analyzing national and international data, (3) the required corrections to avoid double-counting, and (4) the msGBLUP.

3.3.2.1 Single-Step GBLUP model. This method uses information from both genotyped and non-genotyped animals simultaneously. The mixed model equations then become:

\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & Z'Z + H^{-1}\alpha
\end{bmatrix}
\begin{bmatrix}
\hat{b} \\
\hat{u}
\end{bmatrix}
= \begin{bmatrix}
X'y \\
Z'y
\end{bmatrix}
\]

where \(\alpha\) is the variance ratio and \(H^{-1}\) is derived as in Aguilar et al., (2010) and Christensen and Lund (2010):

\[
H^{-1} = A^{-1} + \begin{bmatrix}
0 & 0 \\
0 & (G^{-1} - A^{-1})
\end{bmatrix}
\]

\[\text{[3]}\]
where $A^{-1}$ is the inverse of the pedigree relationship matrix, $A_{22}^{-1}$ is the inverse of the pedigree relationship matrix among genotyped animals, $G^{-1}$ is the inverse of the genomic relationship matrix. The G matrix was constructed as in VanRaden (2008) and then blended with 5% of $A_{22}$ to avoid singularity problems.

### 3.3.2.1.1 Integration of external MACE evaluation.

When external information is used, $H^{-1}$ is adapted from Legarra et al. (2007):

$$H^{*-1} = \begin{bmatrix} H^{MM} + T^{-1} - H^{-1}_{MM} & H^{MD} \\ H^{DM} & H^{DD} \end{bmatrix}$$

where $H^{MM}$, $H^{MD}$, $H^{DM}$, and $H^{DD}$ are submatrices of the relationship matrix for external, MACE (M), and domestic (D) animals, $H^{-1}_{MM}$ is the inverse of the relationship matrix among external animals, and $T$ is a diagonal matrix with prediction error variance for external animals.

BLUP90MBE software, a modification of BLUP90IOD2 (Tsuruta et al. 2001), was used to calculate single-step genomic estimated breeding values (ssGEBV) using iteration on data by the method of the preconjugated gradients. Following Tsuruta et al. (2011), a convergence criterion of $10^{-14}$ was used.

### 3.3.2.1.2 Corrections to avoid double-counting.

Consider a national and multi-national evaluation for the same set of individuals with the same pedigree and genotype data, but independent phenotypic data. The aim is to integrate predictions obtained with multi-national evaluation into a national evaluation. One concern is that double-counting of information due to relationships among MACE animals is not taken into account. The EBV of an animal is calculated considering information from its own records as well as records from its parents and offspring. Hence, EBV for an individual that contributed to MACE, when used in national evaluation, leads to information being double counted (Vandenplas et al., 2014).
In the present study, bulls whose domestic traditional proof (i.e., EBV and associated reliability) contributed to MACE EBVs were identified based on the non-zero national effective daughter contributions reported by Interbull. A total of 1,729 and 1,788 bulls had MACE EBVs for MT and MS, respectively. From those, 100 and 109 had non-zero national effective daughter contributions, i.e., bulls whose national information contributed to MACE. Thus, domestic EBVs and associated reliabilities for those animals were used and corrections to avoid double-counting were applied (Ignacy Misztal, personal communication):

\[ EBV^* = \frac{((DE_M - DE_D) + k) \times EBV_M - (DE_D + k) \times EBV_D}{(DE_M - DE_D) + k} \]  

\[ Rel_{EBV^*} = \frac{DE_M - DE_D}{(DE_M - DE_D) + k} \]

where EBV* are MACE EBVs obtained from Interbull (EBV_M) after domestic information (EBV_D), used to generate foreign EBV, is removed to avoid double-counting, DE_M and DE_D are daughter equivalent from foreign and national data, respectively, and k is the variance ratio calculated as \((4 - 2h^2)/h^2\). Descriptive statistics for MACE EBVs used in the analysis are shown in Table 1.

3.3.2.2 Multi-step GBLUP Models. Consider the following model:

\[ y = 1 \mu + Zg + e \]  

where \( y \) is a vector of de-regressed evaluations of genotyped animals, \( 1 \) is a vector of ones, \( \mu \) is the overall mean, \( Z \) is a design matrix that assigns records to direct genomic breeding values (DGV), \( g \) is a vector of DGV to be estimated, and \( e \) is a vector of random residual effects. It was assumed that \( g \sim N(0, G \sigma_g^2) \), where \( G \) is the genomic relationship matrix based on SNP marker genotypes, and \( \sigma_g^2 \) is the additive genetic variance; \( e \sim N(0, I \sigma_e^2) \), where \( I \) is an identity matrix and \( \sigma_e^2 \) is the residual variance.
Here, DGV were estimated with 20% weight ($\delta$) for the expected additive relationship matrix ($A$) when combined with the genomic relationship matrix. Thus, the model becomes:

$$ y = 1\mu + Zg_\delta + e $$  \[8\]

where, $g_\delta \sim N(0, G_\delta \sigma^2_g)$, and where $G_\delta$ is a combined relationship matrix, $G_\delta = \delta A + (1 - \delta)G$.

**3.3.2.2.1 Blending.** The blending approach used to generate GEBV was as follows:

$$ \text{GEBV} = w_1 \text{DGV} + w_2 \text{EST} $$  \[9\]

$$ w_i = \frac{\text{Rel}_i}{\text{Rel}_i + \text{Rel}_j} $$  \[10\]

where $i = 1$ and $j = 2$ for $w_1$, and $i = 2$ and $j = 1$ for $w_2$, $w_1$ is the reliability of DGV estimated from the multi-step GBLUP method, and $w_2$ is the reliability of EST, i.e., either EBV or PA from traditional BLUP for training and validation populations, respectively.

**3.3.3 Implementation**

Six different evaluations were performed in order to evaluate the impact of including MACE information in the domestic evaluation:

1. Domestic traditional BLUP evaluation with domestic phenotypic and pedigree information (BLUP$_D$)


4. Joint evaluation where domestic information of foreign bulls was analyzed with domestic phenotypic, genomic, and pedigree information in a single-step approach (ssGBLUP$_J$).
(5) Domestic evaluation with domestic phenotypic, genomic, and pedigree information with integration of MACE information \((\text{EBV}_M)\) using a single-step method \((\text{ssGBLUP}_{\text{DM}})\).

(6) Multi-national evaluation with domestic and MACE de-regressed EBV and genotypic information using a multi-step method \((\text{msGBLUP}_{\text{MD}})\).

For analyses using the single-step approach, different data-sets were prepared: 1) a full data-set containing all phenotypic records of cows from which EBV were calculated and de-regressed to be used as benchmarks in the evaluation process; 2) a reduced data-set that included records of cows after deleting the last four years of observations; and, for analysis with integration of MACE information: 3) a data-set containing EBV\(_M\) (obtained after corrections to avoid double-counting were applied) and associated reliabilities for foreign bulls. For the multi-step approach, de-regressed EBVs of domestic and foreign genotyped bulls obtained from the 2013 evaluation carried out by CDN were used for prediction, and EBVs obtained from the 2017 evaluation were used for validation. Genomic predictions (GEBV) obtained with the alternate prediction methods were compared based on their observed validation reliabilities \((R^2)\) and bias of predictions.

### 3.3.4 Validation

The reliability of genomic predictions for the validation bulls were measured as squared correlations between the predicted breeding values (obtained with the reduced dataset) and \(\text{DD}\) (from the full dataset) in the validation dataset. In order to assess bias of genomic predictions for each method, the following regression model was used:

\[
\text{DD}_C = b_0 + b_1 X_P + e
\]

where \(\text{DD}_C\) are de-regressed EBVs obtained from complete dataset of genotyped bulls with no daughters in the reduced dataset, but with at least 50 daughters in the complete dataset; \(b_0\) is the
intercept; $b_1$ is the linear regression coefficient indicating bias of predictions; $X_P$ is the bull’s prediction obtained with the reduced dataset; and $e$ is the residual.

3.4 Results and Discussion

The calculation of an animal’s breeding value aggregates information from the animal’s own records (i.e., contributions due to records) and from records of its relatives (i.e., contributions due to relationships) (Misztal and Wiggans, 1988). When a national evaluation is based on integration of EBVs from external sources, i.e., MACE EBVs, avoiding double-counting of information becomes of utmost importance inasmuch as a failure to do so could lead to biased predictions and inflated reliabilities (Vandenplas et al., 2017). In the present study, a total of 1,729 and 1,788 bulls with Interbull evaluations were available for the analysis of MT and MS, respectively. Because double-counting of contributions due to relationships can lead to inflation of predictions, corrections were applied to avoid possible double-counting. From the total number of bulls with MACE EBVs available, 100 and 109 bulls for MT and MS, respectively, also had national (domestic) evaluations that were sent to Interbull and therefore contributed to MACE EBVs calculation (Table 1). Double-counting of information was avoided by applying a discount factor on EBV$_M$ based on daughter equivalents from national and MACE EBVs using equations [5-6]. Although the correction for double-counting used in our study potentially adds an extra-step to ssGBLUP, other studies have successfully implemented such corrections simultaneously in the analysis through a modification of the system of equations (Vandenplas et al., 2014; 2017).

The number of genotyped bulls in the training population has a great impact on reliability of genomic predictions (VanRaden et al., 2009). Thus, the use of foreign (MACE) bulls in genetic evaluations becomes an appealing tool, as it has the potential to increase reliability compared to
the use of domestic bulls only. Their inclusion will increase the size of the training population, which will likely lead to an increase in predictive ability of GBLUP. For msGBLUP, 6,189 and 6,321 domestic bulls were available for MT and MS, respectively. With the inclusion of MACE bulls in the evaluation, there was an increase of about 28% in the size of training population for both traits.

The use of genomic information for prediction increased $R^2$ by 4.0% and 13.0% points for MT and MS, respectively, for msGBLUP$_D$ compared to BLUP$_D$ (Table 3). When MACE bulls were included in the training population, msGBLUP$_{MD}$ led to an increase in $R^2$ of 4.8% and 15.4% points for MT and MS, respectively, compared to BLUP$_D$. Although reliabilities of MACE EBVs were, on average, lower than those for domestic EBVs (Table 2), adding this extra information in the training population still helped increase reliabilities, which underlines the importance of training population size and amount of information available.

The availability of MACE EBVs enables the exchange of information from several countries, however, there are still several limitations associated with their use in the context of a national evaluation through a multi-step approach. For instance, in genetic evaluations, traditional EBVs are regressed toward parent averages based on their associated reliabilities. Thus, the use of such EBVs in genomic evaluations could lead to double-counting of information, shrinkage of the GEBV and inaccuracies of predictions, which is especially true for EBVs with low reliability (Garrick et al., 2009). The use of de-regressed information is then viewed as an alternative response variable that could overcome such problems. However, although de-regression of domestic EBVs is quite straightforward, steps for obtaining de-regressed MACE EBVs can be complicated and potentially cause inaccuracies during evaluation. This might be so because the parent average used for de-regression of an animal’s EBV in the national evaluation may not match to that used by
Interbull for calculations of MACE EBVs (Sargolzaei and Chesnais, 2014). Further complications arise when multi-trait evaluations are used (Schaeffer 2001), which is especially true when each country uses a different number of traits (Vandenplas and Gengler, 2015). Moreover, when MACE EBV of a bull is considered official, but it greatly deviates from its national EBV, inflation may arise due to the fact that EBVs of that bull’s close relatives, i.e., progeny, cows, bulls without international EBV, are no longer in agreement with the bull’s official breeding value (Vandenplas and Gengler, 2015). All these might help explain the very small to non-existent improvements observed on regression coefficients (bias) of genomic predictions obtained with msGBLUP$_{MD}$ compared to msGBLUP$_{D}$ (Table 3).

As expected, the use of a single-step approach led to higher reliabilities of genomic predictions. An increase of 5.2% and 6.3% for MT and MS, respectively, was observed when ssGBLUP$_{D}$ was used compared to msGBLUP$_{D}$ (Table 3). The highest reliabilities were obtained with ssGBLUP$_{IM}$, which used integration of MACE information into the domestic evaluation, while correcting for double-counting. Integration also improved bias of genomic predictions (regression coefficient values were closer to 1.00). Although overestimation of predictions was already minimum for pedigree-based evaluation for milking speed ($b_1 = 0.975$), by using a single-step approach, regression coefficients approached 1 (Table 3). However, there was an underestimation of genomic predictions for the same trait when a multi-step approach was adopted. For MT, less inflated predictions were observed with integration of MACE EBVs into domestic evaluation. Absolute deviations of the regression coefficients from 1 for MT went from 0.46 with BLUP$_{D}$ to 0.42 msGBLUP$_{MD}$. When integration of MACE information was implemented through a single-step approach (ssGBLUP$_{IM}$), the absolute deviation from 1 decreased by half compared to msGBLUP$_{MD}$ (0.21).
In the present study, ssGBLUP\(_J\) can be considered as a reference. This is a hypothetical scenario where all available information, i.e., phenotypes, pedigree, and genotypes in the national database, which also includes domestic information of foreign bulls, is used simultaneously in a joint evaluation. Given that ssGBLUP\(_{IM}\) uses all domestic information available while incorporating external information of foreign bulls (i.e., MACE EBVs), the slightly increased reliabilities and reduced inflation of genomic predictions obtained with this method, compared to ssGBLUP\(_J\), showed that integration of MACE information through a single-step approach was successful. These findings were in agreement with previous studies that also attempted to implement an integration method using a unified approach (Vandenplas et al., 2014; 2017).

Besides the obvious importance of the size of the training data set, genetic relationships between its individuals and young bulls whose breeding values are to be predicted, are also a critical factor impacting the accuracy of genomic breeding values (Habier et al., 2010; Clark et al 2012). In a simulation study, Clark et al. (2012) compared the accuracy of an animal’s breeding value when: 1) animals from validation and training data sets were closely related, 2) animals from validation and training data sets were distantly related, and 3) animals from both data sets were unrelated. Relationships were calculated based on pedigree. When animals in the two data sets were unrelated, accuracies of GEBV were 0.34. However, there was an increase in accuracy of 7 and 23 percentage points by using training data sets with distant and close relationships, respectively, showing that the relationship between animals in training and validation data sets has an great impact on the accuracy of genomic predictions. The gains obtained with the inclusion of MACE bulls in the present study could be partially explained by their genetic relationship with those from the validation data set. When MACE bulls are not included in the analysis, the estimation of breeding values is largely dependent on distant relatives, whence, lower reliabilities
are obtained. However, when MACE bulls are included, given the close relationship with animals from validation data set, more weight is placed on their information, instead of distantly related animals, and hence an increase in reliability is expected. Thus, given the importance of those animals for breeding programs, this study emphasizes the need for a methodology that properly incorporates MACE information into a domestic evaluation.

To our knowledge, no studies have been previously published on the use of MACE information in domestic evaluations using a single-step approach for dairy cattle in Canada and United States using real data. The proposed method greatly improved reliabilities and bias of genomic predictions for the studied traits compared to a multi-step approach, which was expected given the fact that ssGBLUP uses all available information simultaneously. Nonetheless, improvements obtained with ssGBLUP\textsubscript{IM} compared to ssGBLUP\textsubscript{J} shows that, even though the amount of MACE information was small, it had a favorable impact on genomic predictions and, therefore, integration should be considered. Our study highlights the importance of the method used for prediction of breeding values, and also the importance of the size and structure of the training population used for genomic evaluations. The proposed method also enabled integration of external information while avoiding double-counting of information. For a bull with high reliability the amount of “own” information (i.e., contribution from progeny) used to calculate its EBV is usually larger compared to contribution from relatives (Calus et al., 2016). Thus, avoiding double-counting of information is especially important for evaluations based on bulls with lower reliabilities (e.g., bulls with a small number of progeny).

We expect that the proposed methodology for integration of MACE information through a single-step approach will be more beneficial for countries where the number of proven domestic bulls for genomic evaluation is limited. For instance, an improvement of 4 percentage points in
reliability was observed with the inclusion of MACE information for genomic evaluation of udder depth in US Holsteins using a population of 105,116 domestic genotyped animals (Shogo Tsuruta, unpublished data). However, when external information was integrated using a similar methodology in a simulated population with genomic information of only 1,000 domestic animals, reliabilities doubled when compared to a scenario where genomic evaluation was performed considering only domestic information (Andonov et al., 2017).

3.5 Conclusions

The use of a single-step approach with integration of MACE information provides an alternative to the current method used in Canadian genomic evaluations. The proposed method leads to better results compared to the multi-step approach, with less bias and more reliable predictions. Moreover, because phenotypes, genotypes, pedigree, and MACE EBVs are evaluated simultaneously, the single-step approach overcomes the limitations that arise from de-regressing MACE EBVs for use in a multi-step approach. Further developments will allow corrections to avoid double-counting of information simultaneously in the single-step analysis.

3.6 Acknowledgements

This research was supported in main part by Agriculture and Agri-Food Canada and by additional contributions from Dairy Farmers of Canada, the Canadian Dairy Network and the Canadian Dairy Commission under the Agri-Science Clusters Initiative. The first author is also thankful to CAPES (Brazilian Federal Agency for Support and Evaluation of Graduate Education, Brazil) for financial support.
### 3.7 Tables

**Table 3.1.** Descriptive statistics for Multi-trait across-country evaluation (MACE) EBVs used as external information in single-step GBLUP for evaluation of milking temperament (MT) and milking speed (MS)

<table>
<thead>
<tr>
<th>Trait</th>
<th>EDC¹ ≠ 0²</th>
<th>EDC = 0³</th>
<th>All⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N. Bulls</td>
<td>RelEBV (%)</td>
<td>N. Bulls</td>
</tr>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>100</td>
<td>62.1</td>
<td>47.4</td>
</tr>
<tr>
<td>MS</td>
<td>109</td>
<td>74.6</td>
<td>66.5</td>
</tr>
</tbody>
</table>

¹Effective daughter contributions (EDC).
²Bulls whose domestic information contributed to MACE.
³Bulls whose MACE EBVs had no contribution from domestic data.
⁴Total number of MACE bulls included in the analysis.
⁵Reliability of EBV before and after corrections to avoid double-counting were applied.
Table 3.2. Descriptive statistics of EBVs used for genomic evaluations using multi-step genomic BLUP for milking temperament (MT) and milking speed (MS)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Validation</th>
<th>Training</th>
<th>MACE&lt;sup&gt;2&lt;/sup&gt;</th>
<th>Domestic&lt;sup&gt;3&lt;/sup&gt;</th>
<th>All&lt;sup&gt;4&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N. Bulls</td>
<td>Rel&lt;sub&gt;PA&lt;/sub&gt; (%)</td>
<td>N. Bulls</td>
<td>Rel&lt;sub&gt;DD&lt;/sub&gt; (%)&lt;sup&gt;5&lt;/sup&gt;</td>
<td>N. Bulls</td>
</tr>
<tr>
<td>MT</td>
<td>554</td>
<td>25.8</td>
<td>1,729</td>
<td>36.1</td>
<td>6,189</td>
</tr>
<tr>
<td>MS</td>
<td>554</td>
<td>26.3</td>
<td>1,788</td>
<td>54.5</td>
<td>6,321</td>
</tr>
</tbody>
</table>

<sup>1</sup>Only domestic bulls with at least 50 daughters in the 2017 evaluation were considered in validation population.
<sup>2</sup>Bulls with multi-trait across-country evaluation (MACE) EBVs included in the training population.
<sup>3</sup>Bulls with domestic EBVs included in the training population.
<sup>4</sup>All bulls included in the training population.
<sup>5</sup>Average reliability of EBVs used in the training population of multi-step genomic BLUP.
Table 3.3. Observed reliabilities ($R^2$) of genomic predictions and regression coefficient ($b_1$) of de-regressed EBVs on genomic predictions for 554 genotyped bulls, with at least 50 daughters, for milking temperament (MT) and milking speed (MS) for different genomic evaluations.

<table>
<thead>
<tr>
<th>Trait</th>
<th>BLUP$_D^1$</th>
<th>ssGBLUP$_D^2$</th>
<th>msGBLUP$_D^3$</th>
<th>ssGBLUP$_J^4$</th>
<th>ssGBLUP$_{IM}^5$</th>
<th>msGBLUP$_{MD}^6$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_1$</td>
<td>$R^2$</td>
<td>$b_1$</td>
<td>$R^2$</td>
<td>$b_1$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>MT</td>
<td>0.541</td>
<td>0.105</td>
<td>0.662</td>
<td>0.201</td>
<td>0.570</td>
<td>0.149</td>
</tr>
<tr>
<td>MS</td>
<td>0.975</td>
<td>0.219</td>
<td>0.917</td>
<td>0.412</td>
<td>1.086</td>
<td>0.349</td>
</tr>
</tbody>
</table>

1Domestic traditional BLUP evaluation with domestic phenotypic and pedigree information.
2Domestic evaluation with domestic phenotypic, genomic, and pedigree information using a single-step approach.
3Domestic evaluation with domestic de-regressed EBV and genotypic information using a multi-step approach.
4Joint evaluation where all available information (Domestic and multi-national phenotypic, genomic, and pedigree) were used in a single-step approach.
5Domestic evaluation with domestic phenotypic, genomic, and pedigree information with integration of multi-trait across-country evaluation (MACE) information (EBV$_{IM}$) using a single-step method.
3.8 Figures

Figure 3.1. Information flow for calculation of national and international genetic evaluations. EBV = estimated breeding values; GEBV = genomic estimated breeding values; MACE = multiple-trait across-country evaluations; GMACE = genomic enhanced MACE.
3.9 References


CHAPTER 4
Genetics and genomics of reproductive disorders in Holstein cattle


*Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada
†Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, USA
‡The Semex Alliance, Guelph, Ontario, Canada
§Canadian Dairy Network, Guelph, Ontario, Canada

4.1 Abstract
In Canada, reproductive disorders known to affect the profitability of dairy cattle herds are being recorded by producers on a voluntary basis since 2007. Previous studies have shown the feasibility of using producer-recorded health data for genetic evaluations. Despite low heritability estimates and limited availability of phenotypic information, sufficient genetic variation has been observed for those traits, indicating that genetic progress, although slow, can be achieved. Pedigree- and genomic-based analyses were performed on producer-recorded health data of reproductive disorders including retained placenta (RETP), metritis (METR), and cystic ovaries (CYST) using traditional BLUP and single-step genomic BLUP. Genome-wide association studies and functional analyses were carried out to unravel significant genomic regions, biological
pathways, and to better understand the genetic mechanisms underlying RETP, METR, and CYST. Heritability estimates were 0.02 (0.003), 0.01 (0.004), and 0.02 (0.003) for CYST, METR and RETP, respectively. A moderate to strong genetic correlation of 0.69 (0.102) was found between METR and RETP. Averaged over all traits, sire proof reliabilities increased approximately 11 percentage points with the incorporation of genomic data using a multiple-trait linear model. Biological pathways and associated genes underlying the studied traits were identified and will contribute to a better understanding of the biology of these three health disorders in dairy cattle.

4.2 Introduction

For many years, selection for dairy cattle in Canada was focused on the improvement of traits related to milk production and cow conformation (Chesnais et al., 2016). Some of those traits have antagonistic genetic correlations with functional traits, and a decline in reproductive performance of dairy cows over the last decades has already been documented (Pryce and Veerkamp, 2001; Mackey et al., 2007; Berry et al., 2016). However, in the past few years there has been a shift in this scenario, and several countries have started to include health traits into their selection objectives (Chesnais et al., 2016).

A decline in health, reproduction, and longevity will lead to a decrease in the overall profitability of herds due to increased culling rates, veterinary expenses, replacement costs, and decreased milk sales (Pritchard et al., 2013; Parker Gaddis et al., 2014). Reproductive disorders, such as cystic ovaries (CYST), retained placenta (RETP), and metritis (METR), are well known to affect the profitability of dairy farms (Kelton et al., 1998). In Canada, disease prevalence were found to be about 8.2%, 10.8%, and 4.6% for CYST, METR, and RETP, respectively (Koeck et al., 2012). The average cost per case of RETP was estimated to be $315.00, which included costs
related to extra labor for treatment, veterinary fees, drugs, milk losses, and culling (Guard 2008). Kelton et al. (1998) estimated a cost of approximately $39.00 per case of CYST in each lactation. The average cost of METR per case has been estimated at approximately $162.30 (Mahnani et al., 2015).

In Nordic countries, data recording, genetic evaluations, and selection for reproduction and health traits started in the 1960s. The adoption of an index (Total Merit Index, TMI) that comprises not only production, but also reproduction and health traits, enabled those countries to improve functional efficiency of dairy cows while maintaining a steep increase in production (Philipsson and Lindhé, 2003). In those countries, recording of health traits is mandatory. However, in Canada health data are recorded by producers using their on-farm herd management software or calendar log books (Van Doormaal, 2009). Koeck et al. (2012) demonstrated the feasibility of using producer-recorded health data for genetic evaluations in Canada, where more than 40% of all dairy producers under milk recording also participate in the health-recording system. Despite the low heritability estimates reported for such traits (Koeck et al., 2012; Jamrozik et al., 2016) and the limited availability of phenotypic information, sufficient genetic variation was observed, indicating that genetic progress, although slow, can be achieved (Koeck et al., 2012) and will be cumulative over generations.

The use of dense genomic marker data has been widely adopted in many species (Meuwissen et al., 2016) and has been proven to overcome the limitations of traditional breeding programs and progeny test schemes, such as high cost and time associated with proving a bull, and long generation intervals (Schaeffer, 2006; Dekkers, 2007; Hayes et al., 2009). In Canada and the United States, the implementation of a genetic-based selection program for novel traits, such as reproductive disorders, is under implementation and actively investigated. For instance, official
genetic evaluations for mastitis resistance and metabolic disease resistance became available in Canada in 2014 and 2016, respectively. In the United States, the Council on Dairy Cattle Breeding (CDCB, www uscdcb com) plans to release official genetic evaluations for resistance to Hypocalcemia, Displaced Abomasum, Ketosis, Mastitis, Metritis and Retained Placenta in April 2018. Two recent studies (Parker Gaddis et al., 2014; Vukasinovic et al., 2016) applied a single-step Genomic Best Linear Unbiased Prediction (ssGBLUP) procedure for evaluation of health traits in dairy cattle based on on-farm recorded information. The authors concluded that genetic selection for those traits using data recorded by producers is feasible. Moreover, with the inclusion of genomic information, there was an increase in theoretical reliability of 9 to 29 percentage points.

More recently, the ssGBLUP approach that combines genotypes, pedigree, and phenotypes into a single evaluation (Misztal et al., 2009; Aguilar et al., 2010; and Christensen and Lund, 2010) has become a routine procedure in various breeding schemes. By replacing the pedigree-based relationship matrix (A) with the augmented H matrix to include genomic information, both genotyped and non-genotyped animals can be included simultaneously in the evaluation. Previous studies suggest that ssGBLUP yields accuracy of genomic estimated breeding values (GEBV) that are equal to or greater than other methods, but with reduced inflation (Tsuruta et al., 2011; VanRaden, 2012).

Other than performing genomic prediction of breeding values, the use of dense markers for genome-wide association studies (GWAS) becomes a powerful tool to uncover relevant genomic regions associated with phenotypic variation, which will also contribute to faster genetic progress for the traits of interest. The identification of such regions is crucial to better understand the genetic architecture and biological mechanisms of a trait, and possibly improve breeding schemes by putting more selection weight on those regions to increase the number of favorable alleles in the
population through selection. Wang et al. (2012) proposed the use of ssGBLUP for GWAS (ssGWAS), which combines genotypic, pedigree, and phenotypic information into a single evaluation, thus eliminating the need for pseudo-observations (e.g., de-regressed estimated breeding values, DD). In this way, single nucleotide polymorphisms (SNPs) are considered simultaneously in the analysis along with information of non-genotyped animals, which can lead to more accurate estimates compared to other methods, such as BayesB or classical GWAS using DD (Wang et al., 2012). As the number of individuals with producer-recorded data for reproductive disorder traits and genomic information is scarce, using all sources of information (i.e., genotypes and pedigree and phenotypes for both genotyped and non-genotyped animals) simultaneously may be of great value when performing genome associations.

The objectives of the current study were to 1) perform pedigree- and genomic-based analyses of three producer-recorded reproductive disorders, RETP, METR, and CYST, using traditional BLUP and ssGBLUP, 2) perform GWAS analysis to prospect significant genomic regions and biological pathways associated with the three reproductive disorders, and 3) explore the genetic mechanisms underlying RETP, METR, and CYST.

4.3 Materials and Methods

4.3.1 Data

4.3.1.1 Database. Reproductive disorder records measured between April 1, 2007 and March 31, 2017 were obtained from the Canadian Dairy Network (CDN, Guelph, Ontario, Canada, www.cdn.ca) and included events for CYST, METR, and RETP from first lactation Holstein cows. Table 1 shows a descriptive analysis and summary of the data structure by trait.
Genomic data from 33,575 Holstein bulls, born between 1960 and 2015 and genotyped with the Illumina Bovine SNP50 BeadChip (50K, Illumina Inc., San Diego, CA) or a greater density panel and then imputed to 50K (non-overlapping SNPs), and 6,842 Holstein cows, born between 1988 and 2012 and genotyped with either the 50K panel or a low-density panel (6K), and then accurately imputed (Larmer et al., 2014) to the 50K panel using the FImpute software (Sargolzaei et al., 2014) were available for analysis. Genotypes were coded as 0, 1 or 2 for calculation of the genomic relationship matrix (G). Monomorphic SNPs, animal and SNPs with call rate less than 90%, SNPs that were out of Hardy-Weinberg equilibrium with very low probability ($P$-value < $10^{-6}$) or with minor allele frequency less than 0.05, and individuals with parent-progeny Mendelian conflicts were excluded from the analysis. After quality control, the number of genotyped individuals was 40,412 and the final genomic data included 40,852 SNPs. PREGSF90 software was used for SNP and sample quality control (Misztal et al., 2002).

4.3.1.2 Data Validation and Editing. To ensure that all cows were from herds with reliable recording, several editing criteria were applied separately for each disease. Only herds having at least 2 records of a specific disease were considered. The first and last record had to be at least 180 days apart to remove herds that had done recording for only a short time period. A minimum disease frequency of 1% for all traits was applied to ensure continuous data recording within individual herds. Only records from cows with age at first service between 19 and 43 months were considered.

4.3.1.3 Trait Definitions. Reproductive disorders were defined as binary traits (0 = no case, 1 = at least one case) based on whether or not the cow had at least 1 disease case recorded within the first 305 days after calving for CYST, within 150 days after calving for METR, and within 14 days after calving for RETP.
4.3.2 Variance Component Estimation

Both single- and multi-trait linear models were used for analyses of CYST, METR, and RETP and the results were compared. Despite the binary nature of the studied traits, previous studies have shown that the use of threshold models did not improve the goodness of fit compared with linear models (e.g., Neuenschwander et al., 2012). Thus, following the current genetic evaluation models used in Canada, a linear model was applied. For the multiple-trait analysis, a cow was allowed to have only one missing record. Variance components were estimated using pedigree only for all traits with the models described below using the GIBBS2F90 software from the BLUPF90 family (Misztal et al., 2002). A total of 200,000 iterations were completed with the first 50,000 discarded as burn-in, saving one every 25 samples. Post Gibbs analyses were performed using the POSTGIBBSF90 software (Misztal et al., 2002). Trace plots were inspected visually to ensure that convergence had been reached. Posterior standard deviations were calculated for each estimate. The model used for each trait was:

\[ y = X\beta + Z_h h + Z_a a + e, \]

where \( y \) is a vector of observations for the disease trait; \( \beta \) is a vector of systematic effects, including fixed effects of age at calving and year-season of calving; \( h \) is a vector of random herd-year of calving effects, where \( h \sim N(0, I\sigma_h^2) \); \( a \) is a vector of random additive genetic effects, where \( a \sim N(0, H\sigma_a^2) \); \( e \) is a vector of random errors, where \( e \sim N(0, I\sigma_e^2) \); and \( X, Z_h, \) and \( Z_a \) are the corresponding incidence matrices. Eight classes of age at calving (\( \leq 22, 23, 24, 25, 26, 27, 28-29, \) and \( \geq 30 \) months) and 4 seasons of calving (January to March, April to June, July to September, and October to December) were defined. For the multiple-trait analysis, the covariance structure among traits was modelled as:
\[
\text{Var} \begin{bmatrix} \mathbf{h} \\ \mathbf{a} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \otimes \mathbf{H}_0 & 0 & 0 \\ \mathbf{A} \otimes \mathbf{G}_0 & 0 & 0 \\ \text{symm.} & \mathbf{I} \otimes \mathbf{R}_0 \end{bmatrix},
\]

where \( \mathbf{H}_0 = \begin{bmatrix} \sigma_{hl}^2 & \sigma_{h2} & \sigma_{hl3} \\ \sigma_{h2} & \sigma_{h2} & \sigma_{h3} \\ \text{symm.} & \sigma_{h3} & \sigma_{h3} \end{bmatrix} \) is the (co)variance matrix between traits due to random herd-year of calving effects; \( \mathbf{I} \) is an identity matrix; \( \mathbf{G}_0 = \begin{bmatrix} \sigma_{a1}^2 & \sigma_{a1a2} & \sigma_{a1a3} \\ \sigma_{a2} & \sigma_{a2a3} & \sigma_{a3} \\ \text{symm.} & \sigma_{a3} & \sigma_{a3} \end{bmatrix} \) is the (co)variance matrix between traits due to animal additive genetic effects; \( \mathbf{A} \) is the additive genetic relationship matrix; and \( \mathbf{R}_0 = \begin{bmatrix} \sigma_{e1}^2 & \sigma_{e1e2} & \sigma_{e1e3} \\ \sigma_{e2} & \sigma_{e2e3} & \sigma_{e3} \\ \text{symm.} & \sigma_{e3} & \sigma_{e3} \end{bmatrix} \) is the (co)variance matrix between traits due to random error effects.

### 4.3.3 Genomic Prediction of Breeding Values

By using the same models described above, genomic information was incorporated using the ssGBLUP approach and results were compared. The ssGBLUP method replaces the traditional pedigree relationship matrix (\( \mathbf{A} \)) by the blended \( \mathbf{H} \) matrix: a modified genetic relationship matrix that combines pedigree- (\( \mathbf{A} \)) and genomic-based (\( \mathbf{G} \)) relationship matrices (Legarra et al., 2009; Aguilar et al., 2010). Its inverse is represented by:

\[
\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & (\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}) \end{bmatrix}
\]
where \( \mathbf{A} \) is the pedigree relationship matrix, \( \mathbf{G} \) is the genomic relationship matrix, created as 
\[ 0.95\mathbf{G}^* + 0.05\mathbf{A}_{22} \] 
to avoid singularity problems, where \( \mathbf{A}_{22} \) is the pedigree relationship matrix among genotyped animals, and \( \mathbf{G}^* \) was constructed as in VanRaden (2008):

\[
\mathbf{G}^* = \frac{\mathbf{ZDZ}'}{2\sum p_i(1-p_i)}
\]

where \( \mathbf{Z} \) is a matrix of centered SNP content; \( \mathbf{D} \) is a diagonal matrix of SNP variance; \( p_i \) is the minor allele frequency of SNP \( i \).

The GIBBS2F90 software (Misztal et al., 2002) was used for the analyses. The GEBVs were obtained by solving the mixed model equations using GIBBS2F90, using 5,000 iterations with the first 1,000 discarded as burn-in, saving one every 25 samples, with fixed variance components. Approximate reliabilities of the EBVs and GEBVs were calculated as 
\[ 1 - \frac{\text{PEV}_j}{1 + \mathbf{f}_j\sigma^2_g} \] 
where \( 1 + \mathbf{f}_j \) is the diagonal element of \( \mathbf{A} \), \( \sigma^2_g \) is the genetic variance for a given trait, PEV\(_j\) is the prediction error variance for animal \( j \), obtained based on the standard error of the samples.

**4.3.4 Genome-Wide Association Study**

Genome-wide association analyses for CYST, METR, and RETP were performed using a weighted ssGWAS (\textbf{WssGWAS}) approach considering a multiple-trait linear model, as described above. Following concepts presented by Nagamine et al. (2012) for detection of QTL, Wang et al. (2012) showed that the concept of GBLUP being equivalent to SNP-BLUP could be extended to the ssGBLUP approach. VanRaden (2008) showed that from the selection index equation for GBLUP:

\[
\mathbf{a} = \mathbf{G} \left[ \mathbf{G} + \mathbf{R} \left( \frac{\sigma^2_a}{\sigma^2_e} \right) \right]^{-1} (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}),
\]
where $\mathbf{a}$ is GEBV and $\mathbf{R}$ is a diagonal matrix accounting for heterogeneous residual variance. If $\mathbf{a} | \mathbf{u} = \mathbf{Z} \mathbf{u}$, replacing the leftmost $\mathbf{G}$ by $\mathbf{Z}'$, and weighting by the ratio of SNP to additive direct variances would allow the calculation of SNP effects ($\mathbf{u}$):

$$\mathbf{u} = \mathbf{Z}' \lambda \left[ \mathbf{G} + \mathbf{R} \left( \frac{\sigma_u^2}{\sigma_e^2} \right) \right]^{-1} (\mathbf{y} - \mathbf{X} \mathbf{b}),$$

where $\lambda$ is $\sigma_u^2 / \sigma_a^2$, and $\sigma_u^2$ is the SNP variance. It can be assumed that $\sigma_u^2 = \sigma_a^2 / 2 \sum p_i (1 - p_i)$. Therefore, $\lambda$ can be reduced to $1 / 2 \sum p_i (1 - p_i)$.

Assuming that $\mathbf{w} = \left[ \mathbf{G} + \mathbf{R} \left( \frac{\sigma_u^2}{\sigma_e^2} \right) \right]^{-1} (\mathbf{y} - \mathbf{X} \mathbf{b})$ and $\mathbf{u} = \lambda \mathbf{Z}' \mathbf{w}$, then $\mathbf{a} = \mathbf{G} \mathbf{w}$. In this way, $\mathbf{w} = \mathbf{G}^{-1} \mathbf{a}$, and finally, the SNP effects can be calculated as:

$$\mathbf{u} = \lambda \mathbf{Z}' \mathbf{G}^{-1} \mathbf{a},$$

as $\text{Var}(\mathbf{u}) = \mathbf{D}$, the conditional mean of SNP effects given the GEBV is:

$$\mathbf{u} | \mathbf{a} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}^{-1} \mathbf{a}.$$

Thus, given that GEBV from ssGBLUP were available, SNP effects were calculated as (Wang et al., 2012):

$$\mathbf{u} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}^{-1} \mathbf{a},$$

Estimates of SNP effects were then used to calculate individual variance of each SNP effect (Zhang et al., 2010):

$$\sigma_{u,i}^2 = \{d_{ii}\} = \hat{u}_{i}^2 2p_i (1 - p_i).$$

Weighted ssGWAS is an iterative process with several steps (Wang et al., 2012), considering $t$ as the iteration number, the steps are:

1. $t = 0$; $\mathbf{D}_t = \mathbf{I}$; $\mathbf{G}_{(t)} = \mathbf{Z} \mathbf{D}_t \mathbf{Z}' \lambda$.
2. GEBV for all animals in data set are estimated using ssGBLUP.
3. SNP effects are obtained as: $\hat{\mathbf{u}}_{(t)} = \lambda \mathbf{D}_{(t)} \mathbf{Z}' \mathbf{G}_{(t)}^{-1} \mathbf{a}_g$.
4. SNP variances are calculated as: $d_{i(t+1)}^* = \hat{u}_{i(t)}^2 2p_i (1 - p_i)$. 

70
5. SNP values are normalized so additive genetic variance remains constant:

\[ D_{(t+1)} = \frac{\text{tr}(D_{(t)})}{\text{tr}(D_{(t+1)})} D'_{(t+1)} \]

6. Computed: \( G^*_{(t+1)} = ZD_{(t+1)}Z'\lambda \).

7. \( t = t + 1 \).

8. Loop to step 2.

In the present study, the effects of markers were obtained after iterating steps 2 to 6 3 times.

The variance explained by 20-SNP moving windows was successively calculated across the whole genome. Thus, the percentage of genetic variance explained by the \( i^{th} \) window was:

\[ \frac{\text{var}(a_i)}{\sigma_a^2} = \frac{\text{var}(\sum_{j=1}^{20} z_j \hat{u}_j)}{\sigma_a^2}, \]

where \( a_i \) is the genetic value of the \( i^{th} \) window, \( \sigma_a^2 \) is the total genetic variance, \( z_j \) is a vector of SNP content of the \( j^{th} \) SNP for all individuals, and \( \hat{u}_j \) is the marker effect within the \( i^{th} \) region.

### 4.3.5 Functional Analyses

Genomic windows that explained more than 1% of the total genetic additive variance were further investigated. Genes located within those windows were mapped using the bovine genome assembly UMD3.1 (release 90) available at [http://www.ensembl.org/Bos_taurus/Info/Index](http://www.ensembl.org/Bos_taurus/Info/Index) and considered for functional analysis. *Bos taurus* nucleotide and protein sequences of the genes located within windows that explained more than 1% of the total genetic variance were downloaded from the ENSEMBL Biomart Martview application ([http://www.ensembl.org/biomart/martview](http://www.ensembl.org/biomart/martview)). All unique sequences were aligned against the sequences in the National Center for Biotechnology Information (NCBI, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) non-redundant (NR) database using the BLASTx and BLASTp algorithms from the Blast2GO software (Götz et al., 2008). Sequences with a BLAST match were mapped and annotated.
Metabolic pathways associated with significant sequences were identified using the Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa and Goto, 2000; Kanehisa et al., 2016, 2017).

4.4 Results and discussion

4.4.1 Heritabilities and Genetic and Phenotypic Correlations

Incidences for each trait during first lactation were 3.6%, 4.4%, and 2.8% for CYST, METR, and RETP, respectively (Table 1). All incidences found in the present study were within the range reported in literature (e.g., Kelton et al., 1998; Parker Gaddis et al., 2012; Dhakal et al., 2015). Heritability estimates and genetic correlations (posterior means and standard error, SE) obtained with a multiple-trait linear model are shown in Table 2. Heritability estimates obtained from multiple-trait analyses were similar to those obtained with single-trait analyses (results not shown) and in accordance with previous studies that also used producer-recorded data from Holstein cows in Canada (Koeck et al., 2012; Jamrozik et al., 2016). In general, most studies reported heritability estimates for health traits lower than 10% and, although environmental factors play a large role in such diseases, there is still a genetic component associated with them, which implies that improvement can be achieved through selection. Moreover, all studied traits displayed moderate-to-strong, positive (favorable) genetic correlations. Therefore, selection to improve one trait will potentially lead to a favorable response in others (Table 2). The strongest genetic correlation was found between RETP and METR (0.69 ± 0.102, Table 2), which is in accordance with previous studies (e.g., Koeck et al., 2012; Parker Gaddis et al., 2014; Jamrozik et al., 2016). Dhakal et al. (2015) reported a significant positive recursive effect from liability of RETP on liability of METR and, although no direct effects were found from the studied health disorders on
milk yield, a strong causal relationship was observed between health disorders and culling, which causes economic losses.

4.4.2 Prediction of Breeding Values

Tables 3 and 4 show the reliabilities of sire EBVs computed with pedigree and genomic information for single-trait and multiple-trait analyses, respectively. Multiple-trait analysis utilizes more information to evaluate the animals and, therefore, leads to higher reliabilities of estimated breeding values for the traits being evaluated (Schaeffer, 1984). Moreover, increased reliabilities are obtained with a reduction of PEV, which is dependent on the absolute difference between error and genetic correlations, i.e., greater absolute difference leads to greater accuracies for both traits (Schaeffer, 1984). In the present study, by using a multiple-trait analysis, reliabilities of sire EBV were, averaged over all three traits, 9 percentage points higher compared to single-trait analyses (Tables 3 and 4).

An improvement on reliabilities of sire EBV for all three traits was observed when genomic information was included in the analysis, as shown in Tables 3 and 4, for single- and multiple-trait analyses, respectively. The increase in mean reliability, calculated as the difference in overall mean reliability between genomic and pedigree based models, was 11, 9, and 12 percentage points for CYST, METR, and RETP, respectively, using a multiple-trait analysis (Table 4). Results from single-trait analyses were very similar (Table 3). Increases in reliability by adding genomic information varied among sires. Figure 1 shows a trend in increase based on number of daughters for each sire for CYST. Most often, sires with larger number of daughters already have enough information to attain higher reliabilities of estimated breeding values, whereas greater benefits arising from the inclusion of genomic information are observed for sires with only a few daughters.
This trend was similar for the other traits. The average number of daughters per sire was approximately 23, 19, and 24 for CYST, METR, and RETP, respectively.

The percentage of affected daughters of all sires, the best (top 10%) and worst (bottom 90%) sires, according to their GEBV for CYST, METR, and RETP was calculated (Table 5). A difference among the best and worst sire categories is expected and indicates that, despite the low heritability for such traits, genetic variation exists and could be useful in breeding programs aiming to genetically improve health traits. For all three traits, differences between the percentages of daughters affected for each sire group were observed. For instance, 1 out of 18 daughters of the worst bulls developed RETP, while for the best sires, the incidence was of only 1 out of 42 daughters (Table 5).

Reliabilities of genomic predictions are dependent on the level of linkage disequilibrium (LD) between the markers and the quantitative trait loci (QTL), the distribution of QTL effects, heritability of the trait being evaluated, and the size of the training population from which SNP effects are obtained (Hayes et al., 2009). For lowly heritable traits, a large number of animals with phenotypes and genotypes is required to achieve reliabilities of GEBV comparable to those of traits with higher heritabilities, such as production. For instance, more phenotypic records available means more observations per SNP allele, which will lead to higher accuracies of genomic selection (Hayes et al., 2009). However, obtaining a large number of animals with both phenotypes and genotypes can be a challenge for those traits. The advantages of augmenting the genotypic data with information from phenotypes from the complete population through a ssGBLUP approach are therefore greater for lowly heritable traits (Chen et al., 2011). In addition, as shown in the present and other studies (e.g., Sun et al., 2010; Chen et al., 2011), adoption of a multiple-
trait model will likely result in increased accuracies compared to those of single-trait analysis. A multiple-trait model can be easily implemented through ssGBLUP.

4.4.3 Genome-wide Association Study and Functional Analyses

Various GWAS methods have already been described in the literature (e.g., Hayes, 2013; Feng, 2014). In one of these methods, known as classical GWAS, SNPs are fitted, one at a time, as a fixed effect in the model. A SNP is considered significantly associated with the trait of interest if its $P$-value is below a defined significance threshold (Hayes, 2013). However, by fitting one SNP at a time, many thousands of tests are run, which can cause a multiple testing problem, although multiple testing correction methods can partially solve this issue. Moreover, finding causative mutations is a difficult task given the fact that, due to the extensive LD in dairy cattle, even SNPs positioned far away from the QTL may have significant effects (Pryce et al., 2010). Another method, which follows a Bayesian approach, fits all SNPs simultaneously, as random effects, with different prior assumptions on the distribution of possible SNP effects (Hayes, 2013). Yet, a potential issue associated with this method is that when markers are closely linked, existing effects may be split among all the other markers (Xu, 2003), thus an approach that takes into account the variance explained by specific genomic regions, instead of single SNPs, is needed. Those methods mentioned above include solely individuals with both phenotypes and genotypes and therefore, to enable incorporating phenotypes from non-genotyped individuals, they require the use of pseudo-phenotypes, such as DD. A multi-step approach can lead to increased biases and losses of accuracy (Legarra et al., 2014), thus reducing the contribution of genomic information to the detection of important genomic regions associated with the trait of interest.
In the present study, WssGWAS was implemented as an alternative GWAS method to uncover genomic regions associated with fertility disorders. The WssGWAS allows for unequal variances of markers, rather than assuming they all follow the same Gaussian distribution, which may not be a realistic assumption (Zhang et al., 2016). The use of a weighted approach for association analyses, as implemented in WssGWAS, can be advantageous for small genotyped populations, leading to greater power of estimating QTL effects and positions when compared to other methods. However, the benefits are offset when more animals are genotyped. Moreover, as the number of QTL controlling a trait increases (polygenic traits), accuracies obtained with WssGWAS become similar to those obtained with ssGWAS (Lourenco et al., 2017). In addition, one drawback of ssGWAS approaches is the inability to determine SNP significance levels (Wang et al., 2012). Despite of that, various important genomic regions have been unravelled by using this method in a variety of traits in cattle (Silva et al., 2017; Melo et al., 2017; Teixeira et al., 2017), aquaculture (Vallejo et al., 2017), and pigs (Wu et al., 2017).

Analyses were performed considering 20-SNP moving windows, which are expected to reduce the noise arising from over- or under-estimation of window variances, given that a fixed number of SNPs might not match haplotype blocks patterns (Wang et al., 2014). The Manhattan plots of additive genetic variance explained by 20-SNP moving windows for CYST, METR, and RETP are shown in Figures 2, 3 and 4, respectively. As expected, the results found in this study support the polygenic nature of the studied traits, with many regions across the genome having small contributions to the total genetic variation. However, some chromosomal regions appeared to have a larger contribution to the total genetic variation. Thus, windows that explained more than 1% of the variance for each trait were identified and genes located within those regions were selected for further investigation (Table 6). When combined, those regions explained 8.7%, 12.6%,...
and 10.7% of the total genetic variance for CYST, METR, and RETP, respectively. Biological pathways known to be related to CYST, METR, and/or RETP are shown in Table 7.

Two important biological pathways identified, “cysteine and methionine metabolism” and “nicotinate and nicotinamide metabolism”, were found to be related to oxidative stress in mammals (Sordillo and Aitken, 2009; Martínez et al., 2017). Oxidative stress can cause damage to macromolecules and cells of an organism as a result of an unbalanced production of reactive oxygen species and antioxidant defenses (Gonsette, 2008; Martínez et al., 2017). This phenomenon plays an important role in the susceptibility of dairy cows to health disorders, such as mastitis, mammary edema, METR, and RETP, especially during periods of physiological, metabolic and endocrine changes, such as those in periparturient and early lactation (Sordillo and Aitken, 2009).

Cows intensively selected for higher milk production are more likely to have increased incidence of health disorders (Walsh et al., 2011). Tsuruta et al. (2017) identified a major effect around the location of the *DGAT1* gene on *Bos Taurus* autosome (BTA) 14 for milk yield and mortality for US Holsteins, implying that regions affecting production have a negative impact on cow survival and longevity. During early and peak lactation, daily energy requirements of a cow exceed daily energy availability, a phenomenon known as negative energy balance (NEB). This is especially true for high milk producing cows (Tetens et al., 2012). During periods of NEB, there is an increased production of non-esterified fatty acids (NEFA), which are reported to be associated with increased risk for displaced abomasum, clinical ketosis, RETP and METR during the first 30 days in milk (Ospina et al., 2010). Corroborating these findings, Tetens et al. (2012) reported a significant association between NEB and the “fructose and mannose metabolism” pathway, which was also significantly associated with METR in the present study. This pathway was associated with the *PKFB2* gene, a mediator in glycolysis located on BTA16. The expression
of *PFKFB2* has been reported to be up-regulated in the endometrium of heifers with embryos exhibiting retarded development and hence might be linked to subfertility in heifers (Beltman et al., 2010).

Dysfunctions in the “glyoxylate and dicarboxylate metabolism” and the “citrate cycle (TCA cycle)”, which were significantly associated with CYST and METR, have been reported as highly associated with early pregnancy loss in ewes (Zhao et al., 2015). The authors also identified malfunctioning proteins that are essential for the establishment of pregnancy in the endometrium of ewes that failed to conceive. One of these proteins was the malate dehydrogenase (*MDH2*), which was present in most pathways associated with CYST in the present study. In an earlier study, polycystic ovary syndrome (*PCOS*) in women was associated with marked alterations of the TCA cycle (Zhao et al., 2012).

Another important association identified in the present study is related to “Th1 and Th2 cell differentiation” pathway. T-helper (*Th*) cells are, among others, responsible for modulating immune responses. Based on the cytokine production, Th cells are classified into Th1 and Th2. Both Th1 and Th2 cells are regulated by progesterone, a hormone secreted by the female reproductive system associated with establishment and maintenance of pregnancy. Th1-related cells appear to be associated with non-pregnancy and embryonic loss in cows (Hansen et al. 2004; Maeda et al., 2013). In addition, evidence suggests that Th1 and Th2 cells are differentially regulated by stress hormones, which might induce pro-inflammatory activities in certain tissues and may influence the onset and/or course of infectious, autoimmune/inflammatory, allergic and neoplastic diseases (Elenkov and Chrousos, 1999). This pathway was associated with the cyclin-dependent kinase inhibitor 3 gene (*CDKN3*), which may play a role in negative regulation of cell cycle in placenta in women (Mikheev et al., 2008).
In a similar manner, the “cyanoamino acid metabolism” pathway, which was significant for both METR and RETP, mediates inflammatory responses, cellular proliferation, cell movement, the cell cycle and apoptosis in the bovine endometrium (Hailemariam et al., 2014). This pathway was associated with the ASRGL1 gene located on BTA29, which explained the highest amount of variation for RETP (2.4%), followed by METR (1.32%, Table 6). The “T cell receptor signaling pathway” is also related to immune response, inflammatory diseases and is associated with “Th1 and Th2 cell differentiation” pathway in KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2016; Kanehisa et al., 2017).

4.5 Conclusions

Despite the low heritability estimates observed for reproductive disorders (i.e., metritis, retained placenta and cystic ovaries), substantial increase in the theoretical reliability of estimated breeding values can be achieved by incorporating genomic information into evaluations through a single-step approach, especially when using multiple-trait analysis. As genetic gains are cumulative over generations, genetic progress is expected in the long-term if continued genetic selection is implemented. High and positive (favorable) genetic correlations were found between retained placenta and metritis, implying that selection to improve one trait will potentially lead to a positive response in the other. Genome-wide association studies using single-step GBLUP and functional analyses revealed biological pathways and genes associated with retained placenta, metritis, and cystic ovaries. These findings will contribute to increase the accuracy of genomic predictions for these traits, and therefore contribute to enlarge reproductive efficiency of dairy cattle herds.
4.6 Acknowledgements

This research was supported in main part by Agriculture and Agri-Food Canada and by additional contributions from Dairy Farmers of Canada, the Canadian Dairy Network and the Canadian Dairy Commission under the Agri-Science Clusters Initiative. The authors thank Dr. Angela Cánovas (Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada) for providing advice and support with the functional analyses. The first author is also thankful to CAPES (Brazilian Federal Agency for Support and Evaluation of Graduate Education, Brazil) for financial support.
4.7 Tables

Table 4.1. Summary statistics for health disorders in first-lactation Canadian Holstein cows

<table>
<thead>
<tr>
<th>Trait</th>
<th>Incidence (%)</th>
<th>Days from calving</th>
<th>Number of records</th>
<th>Number of year-seasons</th>
<th>Number of herd-years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cystic ovaries</td>
<td>3.2</td>
<td>0 to 305</td>
<td>185,258</td>
<td>42</td>
<td>5,685</td>
</tr>
<tr>
<td>Metritis</td>
<td>4.4</td>
<td>0 to 150</td>
<td>143,936</td>
<td>42</td>
<td>3,238</td>
</tr>
<tr>
<td>Retained placenta</td>
<td>2.8</td>
<td>0 to 14</td>
<td>205,079</td>
<td>42</td>
<td>5,511</td>
</tr>
</tbody>
</table>
Table 4.2. Heritability estimates (in bold on the diagonal, with posterior standard deviation, SD, in parentheses), genetic correlations (above diagonal, with posterior SD in parentheses), and phenotypic correlations (below diagonal) from multiple-trait analysis of first-lactation Holstein cows.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Cystic ovaries</th>
<th>Metritis</th>
<th>Retained placenta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cystic ovaries</td>
<td>0.02 (0.003)</td>
<td>0.38 (0.174)</td>
<td>0.32 (0.119)</td>
</tr>
<tr>
<td>Metritis</td>
<td>0.05</td>
<td>0.01 (0.004)</td>
<td>0.69 (0.102)</td>
</tr>
<tr>
<td>Retained placenta</td>
<td>0.03</td>
<td>0.13</td>
<td>0.02 (0.003)</td>
</tr>
</tbody>
</table>
Table 4.3. Mean reliability of sire EBV (number of sires) computed with pedigree and genomic information for first-lactation records using single-trait analyses

<table>
<thead>
<tr>
<th>Trait</th>
<th>Pedigree information</th>
<th>Genomic information</th>
<th>Overall gain$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall mean</td>
<td>Unproven sires$^1$</td>
<td>Proven sires$^2$</td>
</tr>
<tr>
<td>Cystic ovaries</td>
<td>0.09 (7,835)</td>
<td>0.14 (4,715)</td>
<td>0.25 (3,120)</td>
</tr>
<tr>
<td>Metritis</td>
<td>0.31 (7,501)</td>
<td>0.37 (5,037)</td>
<td>0.45 (2,464)</td>
</tr>
<tr>
<td>Retained Placenta</td>
<td>0.10 (8,515)</td>
<td>0.16 (5,058)</td>
<td>0.29 (3,457)</td>
</tr>
</tbody>
</table>

$^1$ Unproven sires considered as sires with less than 10 daughters.

$^2$ Proven sires considered as sires with at least 10 daughters.

$^3$ Increase in mean reliability calculated as the difference in overall mean reliability between genomic and pedigree based models.
Table 4.4. Mean reliability of sire EBV computed with pedigree and genomic information for first-lactation records using multiple-trait analysis

<table>
<thead>
<tr>
<th>Trait</th>
<th>Pedigree information</th>
<th>Genomic information</th>
<th>Overall gain$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall mean</td>
<td>Unproven sires$^1$</td>
<td>Proven sires$^2$</td>
</tr>
<tr>
<td>Cystic ovaries</td>
<td>0.22</td>
<td>0.30</td>
<td>0.37</td>
</tr>
<tr>
<td>Metritis</td>
<td>0.40</td>
<td>0.46</td>
<td>0.52</td>
</tr>
<tr>
<td>Retained Placenta</td>
<td>0.16</td>
<td>0.24</td>
<td>0.35</td>
</tr>
</tbody>
</table>

$^1$Unproven sires considered as sires with less than 10 daughters (N = 4,516).
$^2$Proven sires considered as sires with at least 10 daughters (N = 1,693).
$^3$Increase in mean reliability calculated as the difference in overall mean reliability between genomic and pedigree based models.
Table 4.5. Percentage of affected daughters of all sires, top 10% sires, and bottom 90% sires according to their GEBV for cystic ovaries, metritis, and retained placenta from single-trait analyses of first-lactation Holstein cows

<table>
<thead>
<tr>
<th>Trait</th>
<th>Number of sires(^1)</th>
<th>Percentage of affected daughters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All</td>
</tr>
<tr>
<td>Cystic ovaries</td>
<td>3,120</td>
<td>7.4</td>
</tr>
<tr>
<td>Metritis</td>
<td>2,464</td>
<td>7.7</td>
</tr>
<tr>
<td>Retained placenta</td>
<td>3,457</td>
<td>5.2</td>
</tr>
</tbody>
</table>

\(^1\)Only sires with at least 10 daughters were considered.
Table 4.6. Twenty-SNP windows that explained more than 1% of genetic variance (Var) for cystic ovaries (CYST), metritis (METR), and retained placenta (RETP) in Canadian Holstein dairy cows and the corresponding list of annotated genes within each window

<table>
<thead>
<tr>
<th>Trait</th>
<th>Var, %</th>
<th>Chr</th>
<th>Start, bp</th>
<th>End, bp</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYST</td>
<td>1.868</td>
<td>29</td>
<td>48580965</td>
<td>49478288</td>
<td><em>NADSYN1, OSBPL5, CARS, NAP1L4, PHLD2, CDKN1C, U6</em></td>
</tr>
<tr>
<td></td>
<td>1.589</td>
<td>25</td>
<td>17245976</td>
<td>18590810</td>
<td><em>CCP110, C16orf6, IQCK, ACSM5, ACSM1, ACSM4</em></td>
</tr>
<tr>
<td></td>
<td>1.531</td>
<td>25</td>
<td>34663848</td>
<td>35532374</td>
<td><em>POR, MDH2, SRRM3, HSPB1, DTX2, UPK3B,</em></td>
</tr>
<tr>
<td></td>
<td>1.305</td>
<td>16</td>
<td>45150817</td>
<td>46887759</td>
<td><em>bta-mir-34a, U1, GPR157, U6, RERE, ERRF11, TNFRSF9</em></td>
</tr>
<tr>
<td></td>
<td>1.199</td>
<td>21</td>
<td>58503314</td>
<td>6059264</td>
<td><em>CCNK, HHIP1L1, CYP46A1, EML1, U1, EVL, bta-mir-342, YY1, bta-mir-345</em></td>
</tr>
<tr>
<td>METR</td>
<td>2.106</td>
<td>10</td>
<td>66735975</td>
<td>68089016</td>
<td><em>CDK3, CGRRF1, bta-mir-2292, SAMD4A, SOCS4, MAPK1P1L, LGALS3, FBXO34</em></td>
</tr>
<tr>
<td></td>
<td>1.966</td>
<td>16</td>
<td>4567154</td>
<td>6385951</td>
<td><em>FCMR, PFKFB2, C4BPB, C4BPA, CR2, KCNT2</em></td>
</tr>
<tr>
<td></td>
<td>1.611</td>
<td>2</td>
<td>35126</td>
<td>1164256</td>
<td><em>LGSN, NIPA1, NIPA2</em></td>
</tr>
<tr>
<td></td>
<td>1.323</td>
<td>29</td>
<td>41166786</td>
<td>41939696</td>
<td><em>BEST1, ASRGL1, U6atac, SCGB1A1, ROM1, METTL12, SNORA57, UQCC3, GNG3, POLR2G, TAF6L, TMEM179B, SLC3A2</em></td>
</tr>
<tr>
<td></td>
<td>1.163</td>
<td>20</td>
<td>24179723</td>
<td>25463818</td>
<td><em>SNORA17, COX8A, ARL15</em></td>
</tr>
<tr>
<td></td>
<td>1.160</td>
<td>28</td>
<td>1135437</td>
<td>2751851</td>
<td><em>GALNT2, 5S_rRNA, OR5A1, OR5D14</em></td>
</tr>
<tr>
<td></td>
<td>1.130</td>
<td>14</td>
<td>16902500</td>
<td>18000539</td>
<td><em>MTSS1, TATDN1, TMEM65, ANXA13, KLHL38, FBXO32</em></td>
</tr>
<tr>
<td></td>
<td>1.084</td>
<td>8</td>
<td>679513</td>
<td>1604017</td>
<td><em>PALLD, 5S_rRNA, bta-mir-2466, CLCN3</em></td>
</tr>
<tr>
<td></td>
<td>1.052</td>
<td>20</td>
<td>6157936</td>
<td>7050201</td>
<td><em>MSX2, FAM169A, GFM2, ENC1</em></td>
</tr>
<tr>
<td>RETP</td>
<td>2.415</td>
<td>29</td>
<td>41264801</td>
<td>4198397</td>
<td><em>ASRGL1, U6atac, SCGB1A1, ROM1, METTL12, SNORA57, UQCC3, GNG3, POLR2G, TAF6L, TMEM179B, SLC3A2</em></td>
</tr>
<tr>
<td></td>
<td>1.733</td>
<td>17</td>
<td>6483488</td>
<td>7472614</td>
<td><em>SH3D19, LRB1</em></td>
</tr>
<tr>
<td></td>
<td>1.436</td>
<td>16</td>
<td>44176019</td>
<td>46093561</td>
<td><em>U6, CTNNBIP1, CLSTN1, U6, 5S_rRNA, bta-mir-34a, U1, GPR157, U6, RERE</em></td>
</tr>
<tr>
<td></td>
<td>1.410</td>
<td>14</td>
<td>81864817</td>
<td>83027085</td>
<td><em>-</em></td>
</tr>
<tr>
<td></td>
<td>1.272</td>
<td>6</td>
<td>2343794</td>
<td>3505409</td>
<td><em>NPY1R, NAF1, U6, BBS7, CCNA2</em></td>
</tr>
<tr>
<td></td>
<td>1.243</td>
<td>21</td>
<td>31278884</td>
<td>33475697</td>
<td><em>IREB2, HYKK, PSMA4, CHRNA5, UBE2Q2, FBXO22, SNORA3, TMEM266, ISL2, RCN2, PSTPIP1, HMG20A</em></td>
</tr>
<tr>
<td></td>
<td>1.227</td>
<td>10</td>
<td>44733798</td>
<td>45488421</td>
<td><em>GN2, C14orf166, PTGDR, PIFI, RBPMS2, OAZ2</em></td>
</tr>
</tbody>
</table>

1Single-step genomic-BLUP was used to obtain marker effects.
2Important candidate genes for each trait are shown in bold face. Any gene with start and stop positions fully or partially overlapping the window were also considered.
<table>
<thead>
<tr>
<th>Pathway</th>
<th>Enzyme</th>
<th>Gene</th>
<th>Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Glutaminase I</td>
<td>NADSYN1</td>
<td>CYST</td>
</tr>
<tr>
<td>Alanine, aspartate and glutamate metabolism</td>
<td>Synthetase, Asparaginase II</td>
<td>ASRGL1, LGSN</td>
<td>METR, RETP</td>
</tr>
<tr>
<td>Aminoaoyl-tRNA biosynthesis</td>
<td>Ligase</td>
<td>CARS</td>
<td>CYST</td>
</tr>
<tr>
<td>Aminobenzoate degradation</td>
<td>Nitrophenyl phosphatase</td>
<td>CdkN3, PFKFB2</td>
<td>METR</td>
</tr>
<tr>
<td>Arginine biosynthesis</td>
<td>Synthetase</td>
<td>LGSN</td>
<td>METR</td>
</tr>
<tr>
<td>Arginine biosynthesis</td>
<td>Glutaminase I</td>
<td>NADSYN1</td>
<td>CYST</td>
</tr>
<tr>
<td>Biosynthesis of antibiotics</td>
<td>Dehydrogenase</td>
<td>MDH2</td>
<td>CYST</td>
</tr>
<tr>
<td>Butanoate metabolism</td>
<td>Ligase</td>
<td>ACSM1</td>
<td>CYST</td>
</tr>
<tr>
<td>Carbon fixation in photosynthetic organisms</td>
<td>Dehydrogenase</td>
<td>MDH2</td>
<td>CYST</td>
</tr>
<tr>
<td>Carbon fixation pathways in prokaryotes</td>
<td>Dehydrogenase</td>
<td>MDH2</td>
<td>CYST</td>
</tr>
<tr>
<td>Citrate cycle (TCA cycle)</td>
<td>Dehydrogenase</td>
<td>MDH2</td>
<td>CYST</td>
</tr>
<tr>
<td>Cyanooamino acid metabolism</td>
<td>Asparaginase II</td>
<td>ASRGL1</td>
<td>METR, RETP</td>
</tr>
<tr>
<td>Cysteine and methionine metabolism</td>
<td>Dehydrogenase</td>
<td>MDH2</td>
<td>CYST</td>
</tr>
<tr>
<td>D-Glutamine and D-glutamate metabolism</td>
<td>Glutaminase I</td>
<td>NADSYN1</td>
<td>CYST</td>
</tr>
<tr>
<td>Fructose and mannose metabolism</td>
<td>2-phosphatase, Phosphofructokinase 2</td>
<td>PFKFB2</td>
<td>METR</td>
</tr>
<tr>
<td>Glyoxylate and dicarboxylate metabolism</td>
<td>Synthetase, Dehydrogenase</td>
<td>LGSN, MDH2</td>
<td>METR, CYST</td>
</tr>
<tr>
<td>Lysine degradation</td>
<td>Kinase</td>
<td>HYKK</td>
<td>RETP</td>
</tr>
<tr>
<td>Methane metabolism</td>
<td>Dehydrogenase</td>
<td>MDH2</td>
<td>CYST</td>
</tr>
<tr>
<td>Mucin type O-glycan biosynthesis</td>
<td>N-acetylgalactosaminyltransferase</td>
<td>GALNT2</td>
<td>METR</td>
</tr>
<tr>
<td>Nicotinate and nicotinamide metabolism</td>
<td>Synthase (glutamine-hydrolysing)</td>
<td>NADSYN1</td>
<td>CYST</td>
</tr>
<tr>
<td>Nitrogen metabolism</td>
<td>Synthetase</td>
<td>LGSN</td>
<td>METR</td>
</tr>
<tr>
<td>Purine metabolism</td>
<td>Phosphatase, RNA polymerase,</td>
<td>GNG2, GNG3, POLR2G,</td>
<td>METR, RETP</td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td>RNA polymerase</td>
<td>POLR2G</td>
<td>METR, RETP</td>
</tr>
<tr>
<td>Pyruvate metabolism</td>
<td>Dehydrogenase</td>
<td>MDH2</td>
<td>CYST</td>
</tr>
<tr>
<td>T cell receptor signaling pathway</td>
<td>Phosphatase</td>
<td>CDKN3</td>
<td>METR</td>
</tr>
<tr>
<td>Th1 and Th2 cell differentiation</td>
<td>Phosphatase</td>
<td>CDKN3</td>
<td>METR</td>
</tr>
<tr>
<td>Thiamine metabolism</td>
<td>Phosphatase</td>
<td>GNG2, GNG3, PIF1</td>
<td>METR, RETP</td>
</tr>
</tbody>
</table>
Figure 4.1. Increase in reliability of predictions when adding genomic information plotted against number of daughters for each sire in single-step analysis for cystic ovaries.
Figure 4.2. Manhattan plot for the proportion of genetic variance explained by the 20-SNP moving windows for cystic ovaries.

Figure 4.3. Manhattan plot for the proportion of genetic variance explained by the 20-SNP moving windows for metritis.
Figure 4.4. Manhattan plot for the proportion of genetic variance explained by the 20-SNP moving windows for retained placenta.
4.9 References


CHAPTER 5

Estimating the impact of the deleterious recessive haplotypes AH1 and AH2 on reproduction performance of Ayrshire cattle


*Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada
†Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, USA
‡The Semex Alliance, Guelph, Ontario, Canada
§Canadian Dairy Network, Guelph, Ontario, Canada

#Animal Genomics and Improvement Laboratory, ARS, USDA, Beltsville, MD.

5.1 Abstract

The effects of two deleterious recessive haplotypes on reproduction performance of Ayrshire cattle, Ayrshire Haplotype 1 (AH1) and Ayrshire Haplotype 2 (AH2), were investigated in Ayrshire cattle evaluated in Canada. Their phenotypic effects on stillbirth (SB) rate and 56-day non-return rate (NRR) were calculated by estimating the interaction of service sire carrier-status with maternal grandsire (MGS) carrier-status using the official Canadian evaluation models for those two traits. The interaction term included 9 subclasses for the 3 possible conditions of each bull: haplotype carrier, non-carrier, or not genotyped. A total of 394 carriers and 1,433 non-carriers were available for AH1, whereas numbers of carriers and non-carriers for the AH2 haplotype were
313 and 1,543 respectively. The number of matings considered for SB was 34,312 for heifers (first parity) and 115,935 for cows (later parities). For NRR, 49,479 matings for heifers and 160,528 for cows were used to estimate haplotype effects. A negative effect of AH1 on SB rates was observed, which were 2.0% higher for matings of AH1-carrier sires to dams that had an AH1-carrier sire; this effect was observed for both heifers and cows. However, no significant effect was observed for the impact of AH1 on NRR. The AH2 haplotype had a negative impact on NRR, with 5.1% more heifers and 4.0% more cows returning to service, while no significant effect was found on SB rates. The harmful effects of AH1 and AH2 on reproduction traits were validated in the Canadian Ayrshire population. This information is of great interest for the dairy industry, allowing producers to make mating decisions that would reduce reproductive losses.

5.2 Introduction

The rapid and advanced development of genomic technologies has revolutionized modern livestock breeding programs, especially in dairy cattle, enabling early selection of breeding animals based on genomic EBV (GEBV) and genomic tests, such as for lethal haplotypes or single genes.

Genomic selection has brought significant improvements in productivity and profitability in dairy cattle breeding programs due to intensive genetic selection for production and conformation traits (e.g., milk yield). However, genetic improvement of traits related to health and reproduction/fertility is limited especially due to lower heritability estimates, lower accuracy of measurements, and less phenotypic data compared to production and conformation traits. Moreover, the extensive use of elite sires through artificial insemination (AI) in the dairy industry has led to increased inbreeding rates, which has a negative impact on fitness traits (Pryce et al.,
Most dairy cattle breeds in Canada are genetically small populations that have originated from a limited number of founders. Melka et al. (2013) reported effective population sizes for five Canadian breeds, ranging from 40 for Canadienne to 66 for Milking Shorthorn. The authors also investigated the effective number of ancestors, which was observed to be the lowest in the Ayrshire breed, indicating an intensive use of few sires of superior genetic merit, and higher selection intensity for this breed. This finding could explain the increased inbreeding levels and the loss of genetic diversity observed in this breed over the past five decades (Sewalem et al., 2006; Melka et al., 2013). More recently, the average inbreeding levels for heifers born in 2016 were reported to be 7.34%, 6.96%, 6.43%, and 6.36% for Holstein, Brown Swiss, Ayrshire, and Jersey breeds, respectively (CDN, 2017).

A reduction in genetic variance is expected with higher levels of inbreeding, thus shrinking the potential for genetic gains (Falconer and Mackay, 1996). Furthermore, inbreeding can negatively affect fitness traits, including fertility and reproduction, by increasing the frequency of homozygous deleterious recessive alleles in the population, which can have an unfavorable impact in the profitability of the production system (Pryce et al., 2014). Previous studies have already identified genomic regions affecting female fertility in dairy cattle as a consequence of increased rates of inbreeding (e.g., VanRaden et al., 2011; Fritz et al 2013; Pryce et al., 2014; Adams et al., 2016). A recently discovered haplotype on BTA17 in the Ayrshire breed, named Ayrshire Haplotype 1 (AH1) was found to be associated with juvenile mortality, developmental disorders, and reduced fertility (Cooper et al., 2014; Venhoranta et al., 2014). Cooper et al. (2014) identified the oldest Ayrshire ancestor carrying the AH1 haplotype, named Selwood Betty’s Commander, which was also reported to have the highest expected inbreeding for the Ayrshire breed (11.1%; VanRaden and Smith, 1999). The intense use of Selwood Betty’s Commander through AI and the
popularity of his descendants might explain the high and steady frequencies for the AH1 haplotype found in the Ayrshire population (Cooper et al., 2014); the current estimate of the frequency of AH1 in the United States Ayrshire population is 22.2% (Null et al., 2017). A second recessive haplotype located on BTA3, known as Ayrshire haplotype 2 (AH2), was also recently identified and a decreased sire conception rate (SCR) was reported for its carriers (Null et al., 2017). Although lower rates for SCR were observed for AH2 carriers, the effect was not statistically significant, likely due to the small number (n = 247) of matings of carrier sires with carrier maternal grandsires (MGS) available for the analysis (Null et al., 2017). The RNA polymerase 2 associated protein (RPAP2) gene was identified as the leading candidate mutation in the haplotype and associated with embryonic development (Null et al., 2017). The carrier frequency of AH2 is of concern as it has grown from 6% in 1990 to 21.7% in 2017 (Null et al., 2017).

Due to high carrier frequencies reported for AH1 and AH2 in the United States and the limited availability of data on previous studies, it is crucial to investigate and potentially validate the effects of such haplotypes on reproduction traits using a larger dataset. Therefore, the present study aimed to 1) test if the recessive lethal haplotypes reported by Cooper et al. (2014) and Null et al. (2017) are segregating in the Canadian Ayrshire cattle population, and 2) investigate their possible deleterious effects on stillbirth (SB) and 56-day non-return rate (NRR).

5.3 Materials and Methods

Data on Ayrshire reproduction traits (SB and NRR) were extracted from the June 2017 genetic evaluation carried out by Canadian Dairy Network (CDN, Guelph, Ontario, Canada; www.cdn.ca). The SB trait was defined as 0 if the calf was born dead and as 1 if the calf was born alive. NRR was coded as 1 when there was no subsequent insemination between 15 and 56-day
following the first service, and 0 otherwise. All traits recorded during or before the first calving were coded as parity 1 and considered heifer traits, and those measured after first calving (later parities) were considered as cow traits. Only records up to the sixth calving were kept in the data.

After data filtering, a total of 150,247 and 210,008 records for SB and NRR were used for further analyses. Information on animals’ haplotype statuses was obtained from the Animal Genomics and Improvement Laboratory (ARS, USDA, Beltsville, MD).

The animal model proposed by Jamrozik and Kistemaker (2016) was modified to evaluate the traits in heifers and cows by including an interaction between haplotype carrier status of the sire of the embryo or newborn calf (service sire) and carrier status of the cow’s sire (MGS) (VanRaden et al., 2011). For heifers (first parity) the single trait models were:

\[
\text{NRR} = \text{INT} + \text{RYM} + \text{HY} + \text{Mf} + \text{SS} + \text{T} + \text{A} + \text{E}; \quad \text{and},
\]

\[
\text{SB} = \text{INT} + \text{RYM} + \text{HY} + \text{AcMcX} + \text{SC} + \text{A} + \text{E},
\]

Similarly, for cows (later parities) the single trait models were:

\[
\text{NRR} = \text{INT} + \text{RYM} + \text{HY} + \text{ApMf} + \text{SS} + \text{T} + \text{A} + \text{PE} + \text{E}; \quad \text{and},
\]

\[
\text{SB} = \text{INT} + \text{RYM} + \text{HY} + \text{AcMcX} + \text{SC} + \text{A} + \text{PE} + \text{E},
\]

where INT is the interaction effect between haplotype carrier statuses; RYM is region by year of birth by month of birth effect (12 classes); HY is herd by year of birth effect; Mf is month of first insemination effect; AcMcX is age at current calving by month of current calving by sex of calf by parity effect; ApMf is age at previous calving by month of first insemination by parity effect; A is the animal additive genetic effect; SS is the service sire by year of insemination effect; SC is sire of calf effect; PE is the permanent environment effect; T is AI technician effect, and E is the random error term. INT, RYM, Mf, AcMcX and ApMf were considered as fixed effects, while all the other effects were treated as random. The INT effect included 9 subclasses for the 3
possible conditions of each sire: carrier, non-carrier, or not genotyped. For AH1 haplotype, a total of 394 carriers and 1,433 non-carriers were available in the data, whereas for AH2 haplotype, the numbers of carriers and non-carriers were 313 and 1,543 respectively.

At the time of this study, status of AH2 haplotypes was unknown for many genotyped Canadian Ayrshire animals because the test was not part of official US genomic evaluation. However, AH2 status was known for 6,344 genotyped animals (1,306 carriers and 5,308 non-carriers) from a test carried out by the Animal Genomics and Improvement Laboratory, USDA. In order to increase the number of Canadian bulls with known status and with phenotype for SB and NRR, the affected haplotype was located by searching the bulls’ genomes with sliding windows of different sizes using snp1101 software (Sargolzaei, 2014). Phased imputed 50k genotypes were used for the haplotype search. Only one haplotype was consistent with the already known carrier/non-carrier status for AH2. Moreover, the identified haplotype was not present as homozygous in the population and the animal’s status was compatible with that from its parents. The affected haplotype was located between 50.8 to 52.3 Mb of BTA3, which was slightly shorter (by 0.2 Mb) than the one later published by Null et al., (2017). For the detected location, haplotypes of all genotyped animals with unknown status were checked against the affected haplotype and carrier/non-carrier statuses were determined. Single trait analyses of SB and NRR were performed using BLUPF90 programs (Misztal et al., 2002), using pedigrees of 125,087 and 108,298 animals for NRR and SB, respectively. A modified version of airemlf90 software (Misztal et al., 2002) was used to perform t-test for specific contrasts between interactions.

Functional analyses were also performed aiming to better understand the biological mechanisms of AH1 and AH2 on SB and NRR. The reported locations for the studied haplotypes were BTA17 between 65.9 and 66.2 Mbp for AH1 (Cooper et al., 2014) and BTA3 between 50.7
and 52.4 Mbp for AH2 (Null et al., 2017 and results from the present study). *Bos taurus* protein sequences of the genes located within those regions or within a distance of 500 kb on each side of the regions were further investigated using the ENSEMBL Biomart Martview application (www.ensembl.org/biomart/martview). All unique sequences were aligned against the reference sequences in the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov) Non-Redundant (NR) database using the *BLASTp* algorithm from Blast2GO software (Götz et al., 2008). Sequences with a BLAST match were mapped and annotated. Metabolic pathways associated with the genes were identified using the Kyoto Encyclopedia of Genes and Genomes (KEGG, Kanehisa and Goto, 2000; Kanehisa et al., 2016, 2017).

### 5.4 Results and Discussion

The reproduction performance of a population is a key indicator of its sustainability in the medium and long term. Therefore, identifying genetic factors associated with reduced reproduction performance is of crucial interest of the dairy cattle industry worldwide. In this study we investigated if the recessive lethal haplotypes reported by Cooper et al. (2014) and Null et al. (2017) were segregating in the Canadian Ayrshire cattle population and estimated their effects on SB and NRR, which had not been previously explored using a large dataset. Descriptive statistics for each trait are shown in Table 1. The total numbers of observations for SB was 34,312 for heifers, with 91.51% and 8.48% live births and stillbirths, respectively; and 115,935 records for cows, with 93.81% live births and 6.19% stillbirths. For NRR, there were 49,479 observations for heifers and 160,529 for cows, with 67.23% and 56.74% non-returns for heifers and cows, respectively (Table 1). A total of 59% and 58% of SB were male calves for heifers and cows, respectively (Table 2). This is in agreement with previous studies that reported male calves are
more likely to be stillborn than female calves, which could be associated with calf size (Meyer et al., 2001; Bicalho et al., 2007; Cole et al., 2007; Sewalem et al., 2008). The proportion of male calves born dead were also greater than of females when comparing heifers against cows. However, the total (males and females) proportion of stillbirths was higher for heifers (10.0% and 6.9% for males and females, respectively) compared to cows (7.1% and 5.3% for males and females, respectively).

The phenotypic trends for NRR and SB are shown in Figure 1. Averaged over the years (1996 to 2014), SB for heifers and cows were 7.7% and 6.3%, respectively. Up to 2009, a steady increase was observed in the number of SB for cows and heifers, with a higher increase for the latter. The incidence of SB for heifers went from 3.5% in 1996 to 10.9% in 2009, followed by a decrease to 7.5% in 2014. For cows, the same trend was observed, although SB occurred in a lesser degree, ranging from 3.6% in 1996 to 7.1% in 2014, with a higher incidence in the years from 2005 (7.6%) to 2006 (7.9%). Heifers showed more year-to-year variations than cows (Figure 1). No major differences were observed along the years for NRR, with an average of 67.4% and 56.5% of non-returns for heifers and cows, respectively. Despite the low heritabilities estimated for SB and NRR, ranging from 1% for SB to 2% for NRR, additive genetic variation could be enough to allow effective selection for these traits. In addition, the current Canadian Daughter Fertility index for all breeds is based on age-at-first-service (11%), NRR in heifers (16%), first-service-to-conception in heifers (8%), calving-to-first-service (15%), NRR in cows (34%), and first-service-to-conception in cows (16%) (Filippo Miglior, personal communication). A higher weight is placed on NRR, which emphasizes the importance of this trait for reproductive performance in dairy cattle breeds.
Tables 3 and 4 show the observed distribution of each trait occurrence by carrier status of MGS within carrier status of service sire for AH1 and AH2 haplotypes, respectively. For NRR, slightly higher (favorable) occurrence was observed for carrier MGS and carrier service sire (67.7% and 57.3% for heifers and cows, respectively) compared to non-carrier MGS and non-carrier service sire (66.6% and 55.7% for heifers and cows, respectively) for the AH1 haplotype (Table 3). However the differences in the occurrence were not significant between carrier service sire x carrier MGS matings compared to non-carrier service sire x non-carrier MGS (Table 5). This implies that no evidence of detrimental effects of AH1 on NRR was found. NRR were considerably higher for non-tested MGS and service sires for both haplotypes (Tables 3 and 4), which could be explained by the small number of observations within those classes, as seen in Tables 5 through 8.

To our knowledge, there are no other reports in the literature of association between AH1 or AH2 haplotypes and NRR. Thus, these findings greatly contribute to understanding the association between AH1 or AH2 haplotypes and important reproduction traits.

For cows, an increase of 1.4% in stillbirths was observed for carrier service sire x carrier MGS matings compared to non-carrier service sire x non-carrier MGS for the AH1 haplotype, whereas for heifers, a slightly higher increase (2.1%) was observed (Table 3). Significant negative effects (P-value < 0.01) were found for all matings (more stillbirths) where service sire was a carrier of the AH1 haplotype compared to non-carrier service sire x non-carrier MGS (Table 6). This could be an indication of detrimental effects on fertility, especially during mid- to late-gestation (VanRaden et al., 2011). Venhoranta et al. (2014) suggested that PIRM syndrome, a disorder affecting Ayrshire calves, could be linked to the AH1 haplotype. The authors identified a common region on BTA17 with extended homozygosity present in all animals affected with the disease while no homozygosity was observed on the unaffected animals, which suggests a
recessive pattern of inheritance. The most typical symptoms of animals with PIRM syndrome is ptosis, intellectual disability, retarded growth and mortality of affected animals.

Stillbirth is an economically important trait for dairy cattle, as its occurrence is not only linked with the costs associated with the loss of a calf, but also costs associated with subsequent impaired productive and reproductive performance of cows (Mahnani et al., 2017). Thus, better understanding of the biology underlying this trait and potentially reducing its incidence is of utmost importance, as it can affect the profitability of a dairy farm. Three important genes were located in a region in common with AH1 haplotype and were further investigated: Slingshot protein phosphatase 1 (SSH1), hepatocyte nuclear factor 1 alpha (HNF-1α), and transient Receptor Potential Vanilloid 4 (TRPV4).

The functional analyses performed here indicated that SSH1 gene is involved in the “Th1 and Th2 cell differentiation” pathway, in which both Th1 and Th2 cells are regulated by progesterone, a hormone secreted by the female reproductive system associated with establishment and maintenance of pregnancy. Th1-related cells appear to be associated with non-pregnancy and embryonic loss in cows (Hansen et al. 2004; Maeda et al., 2013). HNF-1α gene has been previously reported to affect fertility in mice (Lee et al., 1998). HNF-1α gene has an essential role in postnatal growth and development. Moreover, mice became infertile when the region encoding the first 108 amino acids of HNF-1α were deleted (Lee et al., 1998). According to the authors, the deletion results in the functional inactivation of the HNF-1α gene in addition to blocking HNF-1α mRNA maturation, which was similar to results found by Baker et al. (1996). The authors associated infertility of male mutants with a failure of androgenization resulting in absence of mating behavior, due to reduced levels of serum testosterone. TRPV4 gene has been shown to be linked with sperm thermotaxis (i.e., a form of sperm guidance according to a temperature gradient) in
mice and humans (Castellano et al., 2003; Hamano et al., 2016). In a more recent study, Mondal et al. (2017) examined sperm migration and expression of thermotaxis in the fertilization mechanism of bulls. The authors confirmed thermotaxis of bulls’ sperm and suggested that this phenomenon is involved with calcium channels and intracellular stored calcium, which are known to be involved with \textit{TRPV4} gene.

As shown in Table 7, AH2 had a very strong effect (\(P\)-value \(< 0.001\)) on NRR associated with carrier x carrier matings, with a decrease of more than 5\% in 56-day non-return rate compared to non-carrier x non-carrier matings. Estimated subclass effects for the 9 interactions of carrier status of service sire with carrier status of MGS for AH2 haplotype on SB were small and not significant (\(P\)-value \(\geq 0.05\)) for both heifers and cows (Table 8). Although non-significant negative effects on SB were observed in the present study, Null et al. (2017) reported a reduction of 6.1\% on SCR for matings of carrier sires to cows with carrier MGS, which could be explained by the fact that AH2 has an early-acting effect leading to embryonic losses. Embryonic loss is of high concern especially for high producing cows and accounts for the majority of pregnancy loss events in dairy cattle (Diskin et al., 2006; Wiltbank et al., 2016). Moreover, embryonic lethality is often difficult to observe, thereby leading to increased frequencies of the haplotype in the population. Thus, the early-acting effect observed for AH2 might have played an important role in the increase of carrier frequency in the population because a carrier cow with affected embryo can be inseminated again and carry a pregnancy with more than 50\% chance of giving birth to a carrier calf.

Another important gene identified within the reported location for AH2 haplotype is the transforming growth factor beta receptor 3 \((TGF\beta R3)\) gene. The \(TGF\beta\) superfamily of growth factors regulates many aspects of reproductive biology and is known to be expressed in adult mice.
testis and ovary (Sarraj et al., 2007). Markers located in the region of the \(TGF\beta R3\) gene were associated with testicular dysgenesis syndrome, a common disease related to male infertility and impaired development of the testis in men (Dalgaard et al., 2012). It also plays an important role in early embryonic development in cattle (Huang et al., 2010).

Due to intensive selection and reduced effective population size, most dairy cattle breeds display increased inbreeding rates and an unfavorable impact on production efficiency and fertility is expected (Wiggans et al., 2017). Moreover, higher inbreeding rates can lead to expression of recessive defects, which has become an increasing issue. Other recessive haplotypes have already been identified in other breeds, such as BH1 and BH2 in Brown Swiss (VanRaden et al., 2011; Schwarzenbacher et al., 2016), HCD, HH1 and HHM in Holstein (Duchesne et al., 2006; Adams et al., 2012; Menzi et al., 2016), and JH1 and JH2 in Jersey (Sonstegard et al., 2013; VanRaden et al., 2014). The advent of GS, along with large enough phenotypic and pedigree databases, have enabled the identification of such recessive lethal haplotypes. Strategic actions should be taken in order to properly address the issue of increasing incidences of deleterious recessive haplotypes in a population. However, complete elimination of carriers could have a downside if not efficiently performed with a well-designed breeding plan. For instance, effective population size could be substantially reduced as well as, consequently, the genetic variability. This could generate inbreeding depression in fitness traits and, as a result, cause a reduction of genetic gain. Implementing a mating program where the probabilities of an animal being a carrier is taken into account to prevent that known carrier males are mated to known carrier females could be an efficient strategy.
5.5 Conclusions

The effects of two deleterious haplotypes, AH1 and AH2, were evaluated in the Canadian Ayrshire population. Increased stillbirth rates are observed for matings of carrier service sire x carrier MGS of AH1 haplotype, whereas AH2 haplotype leads to a reduction on 56-day non-return rate. It is believed that AH2 causes early embryonic mortality while the effect of AH1 occurs much later during the pregnancy. Action should be taken to prevent an increase in the frequency of AH1 and AH2 carriers in the Ayrshire population. It can be argued that controlling frequency of AH2 is more urgent due to its early-acting effect. Validation of the effects of AH1 and AH2 haplotypes on other reproductive traits will be investigated next.

5.6 Acknowledgements

This research was supported in main part by Agriculture and Agri-Food Canada and by additional contributions from Dairy Farmers of Canada, the Canadian Dairy Network and the Canadian Dairy Commission under the Agri-Science Clusters Initiative. The first author is also thankful to CAPES (Brazilian Federal Agency for Support and Evaluation of Graduate Education, Brazil) for financial support.
### Table 5.1. Descriptive statistics for stillbirth and 56-day non-return rate in Canadian Ayrshire

<table>
<thead>
<tr>
<th></th>
<th>Stillbirth</th>
<th></th>
<th>56-day non-return rate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N. Records</td>
<td>Incidence(^1), %</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Heifers</td>
<td>34,312</td>
<td>8.48</td>
<td>0.92</td>
<td>0.28</td>
</tr>
<tr>
<td>Cows</td>
<td>115,935</td>
<td>6.19</td>
<td>0.94</td>
<td>0.24</td>
</tr>
</tbody>
</table>

\(^1\)Calf was born dead or died within 48 h of birth.  
\(^2\)No subsequent insemination between 15 and 56 days following the first service
Table 5.2. Distribution of livability scores for stillbirth by parity of dam and sex of calf in the edited data-set used for the genetic evaluation

<table>
<thead>
<tr>
<th>Scores</th>
<th>Heifers</th>
<th></th>
<th></th>
<th>Cows</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male²</td>
<td>Female</td>
<td></td>
<td>Male</td>
<td>Female</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>0</td>
<td>1,716</td>
<td>10.0</td>
<td>1,192</td>
<td>6.9</td>
<td>4,191</td>
<td>7.1</td>
</tr>
<tr>
<td>1</td>
<td>15,422</td>
<td>90.0</td>
<td>15,982</td>
<td>93.1</td>
<td>55,102</td>
<td>92.9</td>
</tr>
<tr>
<td>Total</td>
<td>17,138</td>
<td></td>
<td>17,174</td>
<td></td>
<td>59,293</td>
<td></td>
</tr>
</tbody>
</table>

¹Stillbirth scores: 1 = calf born alive, 0 = calf born dead.

²Percentages sum to 100% within parity-sex groups
Table 5.3. Observed distribution of stillbirth (SB) and 56-day non-return rate (NRR) by carrier status of service sire within carrier status of maternal grandsire for AH1 haplotype in Canadian Ayrshire

<table>
<thead>
<tr>
<th>Carrier status of service sire</th>
<th>Carrier status of maternal grandsire</th>
<th>Carriers</th>
<th>Non-carriers</th>
<th>Not tested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SB, %</td>
<td>NRR, %</td>
<td>SB, %</td>
<td>NRR, %</td>
</tr>
<tr>
<td>Heifers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carrier</td>
<td>10.7</td>
<td>67.7</td>
<td>8.7</td>
<td>66.8</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>7.3</td>
<td>66.2</td>
<td>8.6</td>
<td>66.6</td>
</tr>
<tr>
<td>Not tested</td>
<td>6.7</td>
<td>72.2</td>
<td>8.0</td>
<td>75.3</td>
</tr>
<tr>
<td>Cows</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carrier</td>
<td>7.5</td>
<td>57.3</td>
<td>6.5</td>
<td>56.5</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>5.9</td>
<td>56.6</td>
<td>6.1</td>
<td>55.7</td>
</tr>
<tr>
<td>Not tested</td>
<td>6.3</td>
<td>58.8</td>
<td>6.4</td>
<td>61.9</td>
</tr>
</tbody>
</table>

1Calf was born dead or died within 48 h of birth.
2No subsequent insemination between 15 and 56 days following the first service.
<table>
<thead>
<tr>
<th>Carrier status of service sire</th>
<th>Carrier status of maternal grandsire</th>
<th>Heifers</th>
<th>Cows</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Carrier</td>
<td>Non-carrier</td>
<td>Not tested</td>
</tr>
<tr>
<td></td>
<td>SB(^1), %</td>
<td>NRR(^2), %</td>
<td>SB, %</td>
</tr>
<tr>
<td>Carrier</td>
<td>7.4</td>
<td>61.9</td>
<td>8.7</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>8.9</td>
<td>67.2</td>
<td>8.5</td>
</tr>
<tr>
<td>Not tested</td>
<td>6.3</td>
<td>73.1</td>
<td>7.3</td>
</tr>
<tr>
<td>Carrier</td>
<td>6.7</td>
<td>51.6</td>
<td>6.7</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>7.1</td>
<td>55.9</td>
<td>5.9</td>
</tr>
<tr>
<td>Not tested</td>
<td>5.9</td>
<td>62.6</td>
<td>6.2</td>
</tr>
</tbody>
</table>

\(^1\) Calf was born dead or died within 48 h of birth.

\(^2\) No subsequent insemination between 15 and 56 days following the first service.
<table>
<thead>
<tr>
<th>Carrier status of service sire</th>
<th>Carrier status of maternal grandsire</th>
<th>Heifers</th>
<th>Cows</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Effect</td>
<td>N</td>
</tr>
<tr>
<td>Carrier</td>
<td>4,182</td>
<td>0.008 ± 0.009</td>
<td>9,165</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>10,376</td>
<td>0.000</td>
<td>21,152</td>
</tr>
<tr>
<td>Not tested</td>
<td>453</td>
<td>0.050 ± 0.022</td>
<td>1,093</td>
</tr>
<tr>
<td></td>
<td>13,492</td>
<td>0.000 ± 0.006</td>
<td>29,540</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>34,447</td>
<td>0.000</td>
<td>63,617</td>
</tr>
<tr>
<td>Not tested</td>
<td>1,385</td>
<td>0.002 ± 0.014</td>
<td>2,904</td>
</tr>
</tbody>
</table>

1. Coded as either 1 if no subsequent insemination between 15 and 56 days following the first service or 0 otherwise.
<table>
<thead>
<tr>
<th>Carrier status of service sire</th>
<th>Carrier status of maternal grandsire</th>
<th>Heifers</th>
<th>Cows</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Carrier</td>
<td>N</td>
<td>Effect</td>
</tr>
<tr>
<td>Carrier</td>
<td>2,820</td>
<td>-0.031 ± 0.007</td>
<td>6,345</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>7,099</td>
<td>0.000</td>
<td>14,648</td>
</tr>
<tr>
<td>Not tested</td>
<td>341</td>
<td>0.004 ± 0.016</td>
<td>956</td>
</tr>
<tr>
<td>Carrier</td>
<td>10,585</td>
<td>-0.020 ± 0.003</td>
<td>21,523</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>24,533</td>
<td>-0.001 ± 0.002</td>
<td>43,857</td>
</tr>
<tr>
<td>Not tested</td>
<td>1,361</td>
<td>-0.011 ± 0.007</td>
<td>2,758</td>
</tr>
</tbody>
</table>

1Coded as either 0 if calf was born dead or died within 48 h of birth or 1 otherwise.
Table 5.7. Numbers of observations and effects (± SE) of carrier status of maternal grandsire on 56-day non-return rate\(^1\) within carrier status of service sire for AH2 haplotype for heifers and cows in Canadian Ayrshire

<table>
<thead>
<tr>
<th>Carrier status of service sire</th>
<th>Carrier status of maternal grandsire</th>
<th>Heifers</th>
<th></th>
<th>Cows</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Carrier</td>
<td>Non-carrier</td>
<td>Not tested</td>
<td>Carrier</td>
<td>Non-carrier</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Effect</td>
<td>N</td>
<td>Effect</td>
<td>N</td>
</tr>
<tr>
<td>Carrier</td>
<td>2,477</td>
<td>-0.043 ± 0.011</td>
<td>9,222</td>
<td>0.00</td>
<td>541</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>8,058</td>
<td>0.011 ± 0.009</td>
<td>25,579</td>
<td>0.008 ± 0.006</td>
<td>1,939</td>
</tr>
<tr>
<td>Not tested</td>
<td>264</td>
<td>0.054 ± 0.029</td>
<td>1,045</td>
<td>0.078 ± 0.016</td>
<td>353</td>
</tr>
<tr>
<td>Carrier</td>
<td>8,185</td>
<td>-0.040 ± 0.008</td>
<td>33,949</td>
<td>-0.001 ± 0.005</td>
<td>3,886</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>23,450</td>
<td>-0.002 ± 0.006</td>
<td>84,356</td>
<td>0.00</td>
<td>11,212</td>
</tr>
<tr>
<td>Not tested</td>
<td>716</td>
<td>0.038 ± 0.020</td>
<td>3,005</td>
<td>0.026 ± 0.011</td>
<td>1,519</td>
</tr>
</tbody>
</table>

\(^1\)Coded as either 1 if no subsequent insemination between 15 and 56 days following the first service or 0 otherwise.
Table 5.8. Numbers of observations and effects (± SE) of carrier status of maternal grandsire on stillbirth\(^1\) within carrier status of sire of calf for AH2 haplotype for heifers and cows in Canadian Ayrshire.

<table>
<thead>
<tr>
<th>Carrier status of service sire</th>
<th>Carrier status of maternal grandsire</th>
<th>Carrier</th>
<th>Non-carrier</th>
<th>Not tested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Effect</td>
<td>N</td>
<td>Effect</td>
</tr>
<tr>
<td>Heifers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carrier</td>
<td>1,630</td>
<td>0.015 ± 0.008</td>
<td>6,313</td>
<td>0.00</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>5,975</td>
<td>0.000 ± 0.006</td>
<td>17,296</td>
<td>-0.003 ± 0.005</td>
</tr>
<tr>
<td>Not tested</td>
<td>268</td>
<td>0.017 ± 0.018</td>
<td>846</td>
<td>0.004 ± 0.011</td>
</tr>
<tr>
<td>Cows</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carrier</td>
<td>4,787</td>
<td>0.002 ± 0.004</td>
<td>21,821</td>
<td>0.00</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>15,789</td>
<td>-0.003 ± 0.004</td>
<td>58,810</td>
<td>0.004 ± 0.003</td>
</tr>
<tr>
<td>Not tested</td>
<td>647</td>
<td>0.005 ± 0.010</td>
<td>2,950</td>
<td>-0.005 ± 0.005</td>
</tr>
</tbody>
</table>

\(^1\)Coded as either 0 if calf was born dead or died within 48 h of birth or 1 otherwise.
5.8. Figures

**Figure 5.1.** Phenotypic trend for 56-day non-return rate (A) and stillbirth (B) in heifers (dotted line), and cows (broken line).
5.9 References


CHAPTER 6

General Conclusions

6.1 Final Remarks

Breeding objectives in dairy cattle selection programs have changed over the years in order to breed animals that are more profitable, adapted to current environmental conditions, able to produce high-quality food, are healthy and reduce industry footprints (e.g., less methane emissions, manure). Functional traits are key when aiming to meet these goals and, encouragingly, there is significant genetic variation, as documented in various studies (e.g., Egger-Danner et al., 2015; Jamrozik et al., 2016; Pryce et al., 2016). Despite the low heritability of functional traits, genetic progress, although slow, can be achieved and will be cumulative over generations. Genomic selection has played a major role in improving the accuracy of selection for traits with low heritability, which are frequently difficult or expensive to measure or even measured late in life. One drawback of genomic evaluation of lowly heritable traits, such as functional traits, is the size of the training population required to accurately estimate SNP effects to be used for prediction. In addition, a better understanding of the genetic architecture of these traits and the identification of causal mutations or lethal haplotypes is of great value to improve accuracy of selection for these traits. Thus, this thesis investigated alternative strategies to predict breeding values of functional traits more accurately, by using alternative statistical approaches and different sources of domestic and foreign information. In addition, biological understanding of the traits investigated and lethal haplotypes affecting fertility were investigated in order to maximize selection accuracy and design mating plans in Canadian dairy cattle in a more efficient way.

In Chapter 2, the feasibility of applying the single-step genomic BLUP method to predict genomic breeding values for various reproductive and workability traits in Holstein cattle was
investigated and results were compared to those obtained with multi-step genomic BLUP, which is the method currently used in Canadian dairy cattle genomic evaluations. The comparisons were made based on reliabilities and bias of genomic predictions obtained with both methods under different scenarios, such as testing different adjustments of the genomic relationship matrix and using a blending approach to obtain GEBV from the multi-step genomic BLUP method. The single-step genomic BLUP reduced bias while slightly increasing the reliability of genomic predictions for the studied traits. Given that training populations in Canada are composed of genotyped bulls only, the effect of adding cows’ genomic information was also investigated. Predictions obtained with the single-step approach by adding cows’ genomic information showed a small gain in reliability and a reduction in prediction bias compared to those obtained by only including bulls’ genomic information. However, for the multi-step GBLUP method, no effect on reliability or bias of predictions were observed when genomic information of cows was considered. This may be explained by the fact that only a small number of cows were included in the analyses for this method. The inclusion of a polygenic effect in the multi-step analysis, as well as the use of a blending approach had a small effect on reliabilities of predictions, but an important effect on bias. However, the improvements obtained with different strategies adopted in the multi-step method were still smaller compared to the use of the single-step approach.

The results presented in Chapter 2 were obtained considering a genomic evaluation with inclusion of domestic information only, i.e., without considering foreign (MACE) data. However, in Canada and other countries, genomic evaluations for some traits are performed using both sources of information (i.e., foreign and domestic). Thus, in Chapter 3 a new approach for integration of MACE information based on a modification of the current single-step genomic BLUP was proposed and compared to the results obtained from a multi-step genomic evaluation.
The proposed method led to better results compared to the multi-step approach, with more reliable and less biased predictions. Moreover, because phenotypes, genotypes, pedigree, and MACE EBVs are evaluated simultaneously, the single-step approach overcomes the limitations that arise from de-regressing MACE EBVs for use in a multi-step approach. This new methodology will be especially beneficial for countries, traits and breeds with small training populations composed mainly of domestic proven bulls.

In Chapter 4, the feasibility of using fertility disorders data recorded by producers for genomic evaluation in Canada was demonstrated. Despite the low heritability estimates observed for these traits, substantial increase in the theoretical reliability of estimated breeding values were achieved by incorporating genomic information into evaluations through a single-step approach. As expected, the benefits were greater when using multiple-trait analysis. These findings are extremely important given their substantial impact on a dairy farm profitability. Moreover, as genetic gains are cumulative over generations, genetic progress is expected in the long-term if continued genetic selection is performed. Another important aspect to be considered, is the favorable genetic correlation found between retained placenta and metritis, implying that selection to improve one trait will potentially lead to a positive response in the other. Genome-wide association studies using single-step GBLUP and functional analyses revealed important biological pathways and genes associated with retained placenta, metritis, and cystic ovaries, and helped broaden the understanding of the genetic architecture and biological mechanisms underlying the reproductive disorders, which are of great relevance to the dairy cattle industry. These findings will add in the future to increase the accuracy of genomic predictions for these traits and, therefore, contribute to improve reproductive efficiency of dairy cattle herds.
In Chapter 5, the effects of two deleterious recessive haplotypes, AH1 and AH2, on reproduction traits were investigated in the Canadian Ayrshire breed. Phenotypic effects were confirmed by estimating the interaction of service sire carrier status with maternal grand sire carrier status in the Canadian national evaluation for 56-day non-return rate and stillbirth. Stillbirth rate for heifers and cows was 2% higher for carrier service sire x carrier MGS matings compared to non-carrier service sire x non-carrier MGS for the AH1 haplotype. However, no significant effect was observed on 56-day non-return rate. The opposite effect was observed for AH2, with no significant effect found on stillbirth rates, whereas 56-day non-return rate was reduced by 5.1% in heifers and 4.0% in cows. Based on these findings and the results obtained in the functional analyses, it is possible that AH2 causes early embryonic mortality whereas the effect of AH1 occurs much later during the pregnancy. It is important that strategic actions are taken to properly address the issue of increasing incidences of these two deleterious recessive haplotypes in the Canadian Ayrshire breed. As a result, the Canadian Dairy Network has recently implemented a genomic test for carrier identification of the AH1 and AH2 haplotypes. This information becomes available for producers and can be used for mating decisions.
6.2 References

