Norwegian University of Life Sciences
Faculty of Biosciences
Department of Animal and Aquacultural Sciences

Philosophiae Doctor (PhD)
Thesis 2019:25

Management of genomic inbreeding in breeding schemes

Handtering av genomisk innavl i seleksjonsplanar

Gebreyohans Tesfaye Gebregiwergis

Management of genomic inbreeding in breeding schemes

Handtering av genomisk innavl i seleksjonsplanar

Philosophiae Doctor (PhD) Thesis

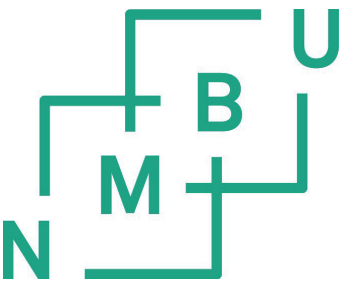
Gebreyohans Tesfaye Gebregiwergis

Norwegian University of Life Sciences

Faculty of Biosciences

Department of Animal and Aquacultural Sciences

Ås (2019)



Thesis number 2019:25

ISSN 1894-6402

ISBN 978-82-575-1587-4

PhD supervisors

Prof. Theo H.E. Meuwissen

Department of Animal and Aquaculture Sciences

Norwegian University of Life Sciences

P.O. Box 5003

1432 Ås, Norway

Dr. Anders Christian Sørensen

Center for Quantitative Genetics & Genomics (QGG)

Department of molecular Biology and Genetics

Aarhus University

Blichers Alle 20,Postboks 50

Dk-8830 Tjele

Dr. Mark Antione Henryon

The University of Western Australia, School of Animal Biology,

35 Stirling Highway, CRAWLEY WA 6009, Australia

SEGES, Danish Pig Research Centre, Axeltorv 3, 1609 Copenhagen V, Denmark

PhD Evaluation committee

Prof. Beatrice Villanueva

Animal Breeding department

National Institute for the Agriculture and Food Research (INIA)

Ctra. de A Coruña Km. 7,5

28040 Madrid

Spain

Dr. Jack Windig

Animal Breeding and Genomic center

Wageningen University & Research

PO Box 338, 6700AH Wageningen

The Netherlands

Prof. Bjørg Heringstad

Department of Animal and Aquacultural Sciences

Norwegian University of Life Sciences

P.O.Box 5003, 1432 Ås

Norway

ACKNOWLEDGEMENTS

This PhD-project was a collaborative effort from the Department of Animal and Aquacultural Sciences at the Norwegian University of Life Sciences and the Center for Quantitative Genetics & Genomics (QGG), at Aarhus University, Denmark. I am grateful to the Center for Genomic selection in Animals and Plants (GenSAP) for funding of the work.

I would like to express my sincerest thanks to my supervisors, main supervisor prof. Theo H.E. Meuwissen and co-supervisors Dr. Anders Christian Sørensen and Dr. Mark Henryon for their support and guidance through the period of my PhD study. Theo, I greatly appreciate your time, invaluable comments, suggestions and all your support. Anders Christian, I highly appreciate your invaluable inputs, suggestions and very clear explanations of complicated topics. Thanks a lot for all your efforts to make things clear especially about the ADAM program and interpretation of the results. I am grateful for adding the BayesP program into the ADAM software and welcoming me to Center for Quantitative Genetics & Genomics (QGG), Aarhus University. Mark: thanks a lot for your fast reply to my e-mails when I was struggling with ADAM, your interest to help me and for the materials you sent me. I am also grateful for all the skype discussions we had about the results and writing the manuscript.

I would also like to express my gratitude to Dr. Solomon Antwi Boison for your support, encouragement and co-authorship. I am very grateful to prof. Peer Berg for your input and interest to discuss the results. Thank you Assoc. prof. Tormod Ådnøy for translating the summary of the thesis into Norwegian and your comments. I also record my sincere thanks to all the co-authors, colleagues and friends at Quantitative Genetics group at the Department of Animal and Aquaculture Sciences (IHA), Norway and at Center for Quantitative Genetics & Genomics (QGG),

Aarhus University. I would like to extend my special thanks to Dr. Rajesh Joshi, Haile Welearegay Gebreslase, Dr. Tesfaye Kebede, Dr. Kahsay Gebremariam Nirea, Dr. Binyam Sime Dagnachew and Katarzyna Piechowiak for your nice friendship and encouragement. I am also grateful to all the administrative staff at IHA for their help whenever needed. I would like to thank Dr. Elise Norberg for facilitating the administrative link with Aarhus University and her encouragement.

During the course of my PhD, I visited North Carolina State University as a visiting scholar and I am grateful to Norwegian University of life Sciences research department for covering my expenses. I would like to thank Dr. Christian Maltecca for arranging the facilities, letting me be part of your research group and your contribution to paper 3. I would also like to thank all research group members for their help during my stay at North Carolina State University. I extend my special thanks to Dr. Francesco Tiezzi for making my stay as enjoyable as possible and for your support whenever needed. I want to thank Smithfield Premium Genetics (SPG) for providing the data used for paper 3 in this thesis.

I would like to thank Yonatan Ayalew, Dr. Bethlehem Legesse, Dr. Daniel Muluwork, Hailay Kidane, Belay Meshesha, Dr. Addis Asmelash, Dr. Fisseha Abenet, Dr. Mekonnen Haileselassie, Migbaru Wubia and Dr. Mengesteab Hailu for being my dear friends who always care about me and send encouragement in many ways.

Finally, a special appreciation goes to all my family members and thank you a lot for your motivation, support and understanding.

Ås, June, 2019

Gebreyohans Tesfaye Gebregiwergis

TABLE OF CONTENTS

SUMMARY

The overall aim of this research work was to explore strategies that manage genetic variation at the genomic level in the era of powerful genomic selection methods.

In paper 1, the role of alternative genomic relationship matrices on the rate of genetic gain at the same rate of true inbreeding in genomic optimum contribution selection breeding schemes were assessed using a stochastic simulation study. For prediction three alternative genomic relationship matrices were calculated based on the genomic information used, i.e. QTL only, markers only, or both markers and QTL. For control of inbreeding, markers only, or both markers and QTL were used. With 7,702 QTL, all genomic relationship matrix combinations used for prediction and control of inbreeding gave similar rates of genetic gain at the same rate of true inbreeding. However, with 1,000 QTL, prediction with QTL and control of inbreeding using markers realized a 29.7% higher rate of genetic gain at a 1% rate of true inbreeding than when using markers for both prediction and control of inbreeding. Hence, the effect of alternative genomic relationship matrices on rates of genetic gain at equal rates of true inbreeding depended on the number of QTL controlling the trait. With a large number of QTL, it is not critical which genomic relationship matrices to use for both prediction and control of inbreeding. However, it is critical with small numbers of QTL, or if few genetic markers can be pinpointed that track (most of) the genetic variance.

In paper 2, we evaluated the effect of different prediction methods on rates of genetic gain at equal rates of true inbreeding in genomic optimum contribution breeding schemes, and on the distribution of the genetic drift across the genome. Use of Bayesian variable selection genomic prediction (BayesP) outperformed GBLUP and realized 5.7%, 1.7% and 2.7 % more short-term

genetic gain with 180, 1000 and 7702 QTL, respectively, at a short-term rate of inbreeding of 1%. BayesP also had higher accuracies of selection than GBLUP. The difference in accuracies between BayesP and GBLUP were higher with few QTL. Moreover, BayesP resulted in smaller selective sweeps around the QTL region than GBLUP. In conclusion, it is advantageous to use BayesP compared to GBLUP in genomic optimum contribution selection breeding schemes.

In paper 3, we compared alternative pedigree and genomic estimators of inbreeding to estimate inbreeding depression in semen quality traits in a Large White pig population. Inbreeding coefficients were measured based on pedigree, average homozygosity (Fhomo), excess of homozygosity (FEx_homo), probability of uniting gametes (Fu), and from the diagonal of genomic relationships matrix (FGRM). Based on Runs of Homozygosity (ROH), three different genomic estimators of inbreeding were obtained: FROH< 5Mb (ROH between 1Mb and 5Mb), FROH>= 5Mb (ROH greater or equal to 5Mb), and FROH> 2Mb (ROH above 2Mb). Significant inbreeding depression effects on semen quality traits were found using FROH> 2Mb, FROH>= 5Mb, FEx_homo, and Fhomo. Moreover, FROH> 2Mb and FROH>= 5Mb showed the strongest association with inbreeding depression in semen quality traits. Hence, we recommend using the ROH based measures of genomic inbreeding to quantify inbreeding depression. Moreover, pig industries should consider implementing control of inbreeding in their breeding plan to maintain high-quality semen in sufficient quantities.

In conclusion: The use of genomic prediction methods and genomic relationship matrices for EBV prediction that focus on chromosomal regions with causative effects benefit rates of genetic gain at a controlled rate of inbreeding. Genomic relationship matrices for the control of inbreeding should represent the inbreeding across the entire genome as well as possible. ROH based inbreeding coefficients were best at picking up the inbreeding depression aspect of inbreeding

amongst the investigated measures of molecular inbreeding. Additional research is warranted on whether this also holds for other negative aspects of inbreeding, such as genetic drift of detrimental alleles to high frequency, and loss of genetic variation.

SAMANDRAG

Det overordna målet for denne avhandlinga var å utforska strategiar som forvaltar genetisk variasjon på genomnivå i vår tid når effektive genomiske seleksjonsmetodar er tilgjengelege.

I Paper 1 blir effekten av bruk av ulike genomiske slektskapsmatriser på genetisk framgang ved same sanne innavlsrate samanlikna. Avlsprogram med genomisk optimal seleksjon blei simulerte stokastisk. Tre typar genomiske slektskapsmatriser blei laga for genomisk prediksjon: enten med bare QTL (kvantitativ-eigenskaps-loci), med bare genetiske markørar, eller med både markørar og QTL. For å kontrollera innavl brukte ein enten bare genetiske markørar, eller både markørar og QTL . Med 7702 QTL fekk vi om lag same avlsframgang gitt same sanne innavlsgrad for alle dei tre genomiske slektskapsmatrisene vi prøvde. Men med 1000 QTL ga prediksjon med QTL-slektskap og innavlskontroll med markørar 27,9% større genetisk framgang når det var 1% sann innavlsauke, samanlikna med bruk av markørar for både prediksjon og innavlskontroll. Så avhengig av kor mange QTL som eigenskapen er styrt av kan det spelar ei rolle kva type genomiske slektskapsmatriser ein bruker til genomisk avlsverdiprediksjon av eigenskapen. Når det er mange QTL som styrer eigenskapen spelar det liten rolle kva slag genomisk slektskap ein bruker for prediksjon og innavlskontroll. Men når få QTL styrer eigenskapen, eller når (mesteparten av) den genetiske variansen til eigenskapen kan forklarast av få markørar, blir det viktig kva slag genomiske matriser ein bruker for prediksjon og innavlskontroll.

I Paper 2 samanlikna vi ulike avlsverdiprediksjonsmetodar når det gjaldt genetisk framgang og genetisk drift i ulike delar av genomet. Samanlikningane blei gjorde ved lik sann innavlsgrad og med genomisk optimal seleksjon. Bruk av Bayes' variabelseleksjon genomisk prediksjon (BayesP) var betre enn GBLUP og ga 5,7%, 1,7% og 2,7% høgare kortvarig genetisk framgang med 180,

1000 og 7702 QTL. Den kortvarige innavlsgraden var 1%. Og BayesP ga sikrare seleksjon enn GBLUP, spesielt når det var få QTL. Dessutan ga BayesP kortare seleksjonssveip i QTL-områda enn GBLUP. Konklusjonen er at det svarer seg å bruka BayesP framfor GBLUP i genomisk optimale seleksjonsprogram.

I Paper 3 samanlikna vi genomiske og anetavle-estimatorar til å detektera innavlsdepresjon for sædkvalitet hos rånar i ein yorkshire-populasjon. Innavlskoeffisientane blei rekna ut enten med anetavle, gjennomsnittleg homosygoti (Fhomo), overskotshomosygoti (FEx_homo), korelasjonen til foreldregametar (F_U), eller frå diagonalane i genomiske slektskapsmatriser (FGRM). Avhengig av homosygositetslengder (ROH) valde ein tre genomiske innavlsestimatorar: FROH< 5Mb (ROH frå 500kb til 5Mb), FROH>= 5Mb (ROH større eller lik 5Mb), og FROH> 2Mb (ROH over 2Mb). Signifikante innavlsdepresjonseffektar på sædkvalitet blei funne med FROH> 2Mb, FROH>= 5Mb, FEx_homo og Fhomo. FROH> 2Mb og FROH>= 5Mb ga høgare assosiasjon med genomisk innavlsdepresjon for sædkvalitet enn dei andre måla. Difor anbefaler vi bruk av ROH-baserte mål for genomisk innavl for kvantifisering av innavlsdepresjon. Elles bør svineproduksjonsselskap vurdera å nytta innavlskontroll i avlsplanane sine for å halda fram med å ha nok høgkvalitets-sæd i framtida.

Til konklusjon: Avlsverdiprediksjon med bruk av genomiske prediksjonsmetodar og genomiske slektskapsmatriser som fokuserer på kromosomområde som har effekt på eigenskapen det blir avla for gir høgare avlsframgang ved kontrollert innavslgrad. Genomiske slektskapsmatriser til kontroll av innavl bør visa innavl over heile genomet så godt som muleg. Blant måla for innavl var ROH-baserte best til å måla innavlsdepresjon av dei som blei testa. Det krevst meir forsking for å få veta om dette også gjeld for andre negative effektar av innavl, slik som genetisk drift for øydeleggjande allel til høge genfrekvensar, og tap av genetisk variasjon.

ABBREVIATIONS

ADAM	Program to simulate selective-breeding schemes for animals
cM	Centimorgan
F	Individual Inbreeding Coefficient
F_{ped}	Pedigree-based inbreeding coefficient
FROH>= 5Mb	Genomic inbreeding coefficient based on long ROH (5 Mb minimum ROH length)
FROH< 5Mb	Genomic inbreeding coefficient due to short ROH (between 1Mb and 5 Mb)
FROH> 2Mb	ROH based genomic inbreeding coefficient (using above ROH length 2Mb)
$F_{U/}$ F_{hat3}	Genomic inbreeding coefficient based on correlation between uniting gametes
FEx_homo	Genomic inbreeding coefficient based on excess of homozygosity
Fhomo	Inbreeding coefficient using the proportion of observed number of homozygous loci
FGRM	Genomic inbreeding coefficients from the genomic relationship matrix
G	Genomic relationship matrix
G_A	Genomic relationship matrix constructed using both markers and QTL
G_M	Genomic relationship matrix constructed using markers
G_Q	Genomic relationship matrix constructed using QTL
GBLUP	Genomic Best Linear Unbiased Prediction
GEBV	Genomic Estimated Breeding Value
GOCS	Genomic optimum contribution selection
IBD	Identical By Descent
IBS	Identical By State
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency

Mb	Megabase pair
Ne	Effective population size
ROH	Runs of Homozygosity
OCS	Optimum contribution selection
SNP	Single Nucleotide Polymorphism
QTL	Quantitative trait loci
WGS	Whole genome sequence
ΔIBD	Rate of true inbreeding
ΔG	Rate of genetic gain

LIST OF PAPERS

The following papers are included in the thesis

Paper 1:

Controlling inbreeding in optimum-contribution selection with alternative genomic relationship matrices.

G.T. Gebregiwergis, Anders C Sørensen, Mark Henryon, Theo HE Meuwissen

(Submitted to Frontiers in Genetics)

Paper 2:

Use of Bayesian genomic prediction methods in genomic optimum contribution selection.

G.T. Gebregiwergis, A Christian Sørensen, Theo HE Meuwissen

(Submitted to Genetic Selection Evolution)

Paper 3:

Estimation of inbreeding depression of semen quality traits using different measures of genomic inbreeding in a swine breeding population.

G.T.Gebregiwergis, Christian Maltecca, Francesco Tiezzi, Kent A.Gray, Vance D.Brown, Yijian Huang, Jeremy T.Howard, Solomon Antwi Boison and Theo HE Meuwissen

(Submitted to Animal Breeding and Genetics)

GENERAL INTRODUCTION

Estimation of inbreeding

The inbreeding coefficient (F) is a quantitative measure of an individual's inbreeding level and represents the proportion of the genome that is identical by descent (IBD) with respect to a base population (Wright, 1922, Malécot, 1948). Conventionally, it was estimated using pedigree information giving the pedigree based inbreeding coefficient (F_{ped}). This gives a probability of being IBD for neutral loci that are unlinked to loci under selection (Malécot, 1948). However, the assumptions used to estimate inbreeding coefficients using pedigree information may not hold in the genomic era, e.g., the presence of unlinked neutral loci is questionable in the genomics era (Sonesson et al., 2012) . In addition, the pedigree based inbreeding coefficient assumes there is no preferential selection between the two alleles at the same locus on the two homologous chromosomes (Wright, 1949). However, in reality this assumption is not true for all loci across the genome (Fernández et al., 2000). At some loci, the two alleles at the same loci may have different effects on a naturally or artificially selected trait, or may be in linkage disequilibrium (LD) with non-neutral loci (Curik et al., 2002, Fernández et al., 2000). As a result of these different effects on a trait, the two alleles at the same loci will have unequal probabilities of transmission to offspring violating the assumption in pedigree inbreeding calculations (Curik et al., 2002, Hill and Weir, 2011). This difference in probabilities of transmission between the two alleles at the same loci could be more noticeable under genomic selection since genomic selection acts on marker alleles in the genome. Hence, the pedigree based estimate of inbreeding may underestimate the true inbreeding at the genomic level since some genomic regions are more often transmitted to offspring than others (Sonesson et al., 2012).

Single Nucleotide Polymorphism - based measures of genomic inbreeding

Currently, genome wide single nucleotide polymorphism (SNP) markers are available at high density that cover the entire genome (Zhang et al., 2010). Dense panels of SNP markers can be used to trace which allele an individual inherited from his parents at each locus (Jones et al., 2010, Hill and Weir, 2011). As the result of this, an allele's inheritance within full sib families due the Mendelian segregation of alleles can be traced using SNP markers, which is called linkage analysis (Hill and Weir, 2011). And, the proportion of an individual genome that is IBD can be calculated more accurately using SNP markers than using pedigree information (Keller et al., 2011, Hayes et al., 2009). Since, variation in inbreeding level within a genome of an individual due to selection or recombination can be detected using SNP markers, inbreeding levels of specific genomic regions can be estimated more accurately using SNP markers than by the expected inbreeding level across all loci based on pedigree information (Hill and Weir, 2011). Hence, SNP markers create new opportunities to measure and manage inbreeding at the genome-wide level and at specific genomic regions.

There are many methods for estimating genomic inbreeding from SNP marker data (Gomez-Raya et al., 2015, Howrigan et al., 2011, Keller et al., 2011, McQuillan et al., 2008, VanRaden, 2008, Kardos et al., 2015). A direct way of measuring genomic inbreeding based on SNP genotypes is the proportion of homozygous genotypes per individual (Saura et al., 2015, Keller et al., 2011). This differs from the aforementioned linkage analysis method in that no pedigree is needed. However, the proportion of homozygous genotypes does not differentiate between alleles being IBD or identical by state (IBS) (Bérénos et al., 2016). Instead of using observed proportions of homozygous genotypes of an individual as a measure of the genomic inbreeding coefficient of the individual, a method that accounts for the population allele frequencies and the individual's

homozygosity has been developed, namely excess of homozygosity (Purcell et al., 2007). This measures the number of homozygous genotypes within an individual relative to the Hardy-Weinberg expected mean number of homozygous genotypes (Purcell et al., 2007, Kardos et al., 2015). In random mating populations and with allele frequencies estimated from the current sample population, this measure of genomic inbreeding will be centered near zero (Kardos et al., 2015). Genomic inbreeding coefficients using this method can also have a negative value indicating that the individual has less homozygote genotypes relative to the frequencies expected by assuming Hardy-Weinberg equilibrium (Kardos et al., 2015, Wang, 2014). Another method of measuring the genomic inbreeding coefficient is F_{hat3} (Yang et al., 2011) which gives more weight to rare homozygote genotypes by scaling with the expected heterozygosity at each marker locus. It is claimed to be the most accurate SNP-by-SNP based genomic inbreeding estimator (Yang et al., 2011, Keller et al., 2011, Bérénos et al., 2016). The genomic inbreeding coefficients estimated using this method can also be negative since its estimates represent the correlation coefficient between uniting gametes (Bérénos et al., 2016, Wright, 1922). Unlike the F_{hat3} measure of genomic inbreeding, by scaling each locus by the average variance (heterozygosity) of all loci, the genomic inbreeding coefficient can also be obtained from genomic relationship matrices. This scaling gives relatively more weight to high MAF (minor allele frequency) SNPs since they contribute more (co)variance than low MAF SNPs (VanRaden, 2008). This measure of genomic inbreeding and the excess of homozygosity based measures of genomic inbreeding have a higher sampling variance than the F_{hat3} measure of genomic inbreeding (Yang et al., 2011). Hence, all these SNP-by-SNP based genomic inbreeding estimators are based on the IBS concepts and except the observed homozygosity of an individual, they are affected by allele frequencies (Curik et al., 2017). The range of the genomic inbreeding coefficients extends beyond the 0-1 range (Wright,

1922, VanRaden, 2008, Wang, 2014). Hence, they are a proxy rather than a direct measure of true inbreeding (the proportion of genome which is IBD) since they do not really separate IBD from IBS (Kardos et al., 2015).

Runs of homozygosity based measures of genomic inbreeding

Alternative measures of realized autozygosity using genomic information were proposed by McQuillan et al. (2008) based on runs of homozygosity (ROH). This method is becoming increasingly popular for estimating individual inbreeding coefficients and has several advantages above SNP-by-SNP based measures of genomic inbreeding (Keller et al., 2011). First, the inbreeding coefficient using ROH provides an estimate of the proportion of the genome being IBD similar to the pedigree based measure of inbreeding coefficient (Bérénos et al., 2016). Second, the ROH length gives insight into the age of inbreeding and this enables the partitioning of recent and old inbreeding (Howrigan et al., 2011, Curik et al., 2017, Curik et al., 2014). Third, it enables estimation of inbreeding at chromosomal level or for specific chromosomal segments (Curik et al., 2014). Fourth, ROH inbreeding coefficients improve the detection of overall burden of rare recessive mutations (Keller et al., 2011). So ROH based inbreeding coefficients have advantages, but different studies differ in their definition of ROH. There are also differences in the detection of autozygosity among the different ROH detection softwares (Howrigan et al., 2011). As a result, there is no consensus on which ROH detection program is optimal at detecting autozygosity (Howrigan et al., 2011). Since ROH based measures of inbreeding implicitly differentiate IBD (long ROH) from IBS (incidental marker homozygosity), these measures get a lot of attention in livestock research (Howrigan et al., 2011, Ferenčaković et al., 2013, Gomez-Raya et al., 2015, Curik et al., 2014, Forutan et al., 2018, Curik et al., 2017, Zhang et al., 2015).

Optimum contribution selection

Optimum contribution selection (OCS) was developed to maximize rate of genetic gain for a given rate of inbreeding by computing the optimum contribution of all selection candidates to the next generation (Meuwissen, 1997, Wray and Goddard, 1994). This selection method was initially developed based on pedigree based relationships among selection candidates and pedigree based measures of inbreeding. However, measures of inbreeding or relationship among selection candidates using pedigree information has limitations as described above since it does not fully account for the genomic relationships among selection candidates. As a result, it is not an accurate measure of genomic relationship among the selection candidates and genetic markers may be more accurate in the genomic era (Goddard, 2009). In the genomic era, the genomic selection methodology was proposed (Meuwissen et al. (2001)) to obtain breeding values of selection candidates using their phenotypes and their genetic marker data. The development and implementation of genomic selection in animal breeding schemes and relationship among individual calculated based on markers being more accurate lead to the extension of OCS to genomic optimum contribution selection (Sonesson et al., 2012). The extension of the method is by replacing the pedigree based relationship matrix by a realized genomic relationship matrix for both prediction and management of inbreeding. As a result, the genomic optimum contribution selection takes into consideration variation in relationships within full-sibs/half-sibs families and variation between genomic regions (Gómez-Romano et al., 2016, Goddard, 2009, Sonesson et al., 2012). Moreover, it enables the use of different genomic relationship matrices for genomic prediction and for inbreeding management employing different genomic information sources. As outlined above, genomic relationship matrices can be constructed using all SNP markers, or using fewer markers which have associations with traits of interest, or based on regions of the genome

that harbor QTL if we know these genome regions in the future (Nejati-Javaremi et al., 1997, Fragomeni et al., 2017, Zhang et al., 2010). Furthermore, genomic relationship matrices can be constructed using several methods to improve accuracies of genomic prediction and increase genetic gain (VanRaden, 2008, Yang et al., 2010, Jannink, 2010, Liu et al., 2015). In addition to SNP-by-SNP based genomic relationship matrices, it is possible to use genomic relationship matrices constructed based on haplotype segments or runs of homozygosity for both prediction as well as management of genomic inbreeding in genomic optimum contribution selection breeding schemes. In conclusion, genomic optimum contribution selection provides a wide range of opportunities to use different prediction methods and alternative genomic relationship matrices to manage genomic inbreeding. The effect of the use of these alternative genomic relationship matrices and prediction methods on rates of genetic gain, rates of inbreeding and distribution of genetic drift over the genome will be investigated in this thesis.

AIM AND OUTLINE OF THE THESIS

The main objective of this thesis is to generate sustainable livestock breeding schemes that manage genetic variation at the genomic level in the era of powerful genomic selection methods.

The thesis has four sub-goals:

- ✓ To investigate rate of genetic gain and rate of true genomic inbreeding in breeding schemes that apply genomic optimum contribution selection applying different genomic relationship matrices
- ✓ To investigate whether variable selection methods improve genetic progress when applied in genomic optimum contribution selection
- ✓ To investigate the effect of variable selection methods and GBLUP on the distribution of genetic drift over the genome
- ✓ To compare alternative pedigree and genomic estimators of inbreeding to estimate inbreeding depression in semen quality traits in a large White pig population

This thesis is divided into three main parts: Paper 1 assesses the role of alternative genomic relationship matrices on rate of genetic gain at the same rate of true inbreeding in genomic optimum contribution selection. Paper 2 evaluates the impact of different prediction methods (variable and non-variable prediction methods) on rate of genetic gain at the same rate of true inbreeding in genomic optimum contribution selection. The impact of the prediction methods on the distribution of IBD profiles across the genome was also investigated. Paper 3 assesses the impact of an increase of inbreeding on semen quality traits in a large white pig population using

different measures of inbreeding coefficients. Paper 3 also assesses which measures of genomic inbreeding are best at detecting inbreeding depression.

GENERAL DISCUSSION

Goals of breeding schemes at the genomic level

In livestock genetic improvement breeding programs, we want the allele-frequencies of positive QTL-alleles to increase towards homozygosity, and thus desire high IBD profiles at such QTL regions (Sonesson et al., 2012). However, we want also to avoid the loss of favorable rare alleles in order to increase the long-term genetic gain (Jannink, 2010, Goddard, 2009, Liu et al., 2014a). At the same time, we want to minimize the loss of variation at non-QTL positions. We especially want to minimize the impact of selection at non-QTL positions, and thus minimize selection signatures. I.e., we want to have very narrow IBD peaks at the QTL' positions (Liu et al., 2014b, Smith and Haigh, 1974). The use of appropriate genomic prediction and inbreeding management methods are essential to achieve the desired changes at genomic level in most animal-breeding schemes. Papers 1 and 2 investigate the effects of different combinations of genomic prediction and inbreeding management methods.

Inbreeding management is a key component in a breeding program for sustainable livestock genetic improvement. The goal of inbreeding management in genetic improvement programs is to avoid / minimize the following consequences of inbreeding in breeding schemes (Meuwissen et al., 2018):

(1) Occurrence of inbreeding depression for breeding goal and non-breeding goal traits. The latter may be responsible for the general functioning of the animals (e.g. reproduction, health and survival).

(2) Loss of genetic variation for breeding goal and non-breeding goal traits. With respect to breeding goal traits only the loss of positive alleles at QTL is problematic. Genetic variation at

non-breeding goal traits may be important for future breeding directions or for specific genomic regions where variability is desirable (e.g. the major histocompatibility complex to fight with diseases).

(3) Random drift and possible increases of recessive disease-alleles to substantial frequencies such that these alleles start to occur in homozygous form, which results in diseased animals. Many recessive disease-alleles might be segregating at low frequencies with hardly any consequences for the health of the animals since heterozygous animals are not affected by the disease. However, an increase in frequency due to random drift will result in the occurrence of homozygous, diseased animals.

Genomic inbreeding management

Based on the pedigree we can predict the probability of IBD at neutral loci unlinked to any loci under selection. In current genomic selection breeding schemes such unlinked loci do not exist, but could be imagined as the inbreeding at a 'neutral chromosome' that is not involved in the selection in any way (no QTL and no SNP markers used for selection). The genetic drift and inbreeding at the chromosomes that are involved in the selection will be higher than this. Thus, with genomic selection, inbreeding consists of two components: (1) a component due to (the loops within) the pedigree, which is depicted by the 'neutral chromosome' inbreeding; and (2) an additional component which is due to the preferential selection of specific chromosomal regions or loci linked to such regions. Genomic management of inbreeding addresses both these components simultaneously.

With the availability of genomic information, we have an array of tools to address the aforementioned consequences of inbreeding. For example, we can maintain variability at specific targeted genomic regions or at fitness related loci using region-specific genomic coancestry

matrices together with genome-wide matrices, in optimum contribution selection (Woolliams et al., 2015, Gómez-Romano et al., 2016). With the use of variable selection prediction methods, loss of variation around the QTL regions can be minimized without the use of additional constraints on the genomic inbreeding around the QTL regions in genomic optimum contribution selection (Paper 2). Disease alleles and regions that contribute to inbreeding depression can be identified and mapped using genomic information (Curik et al., 2017, Kardos et al., 2016, VanRaden et al., 2011, Charlier et al., 2008). Once the responsible genomic regions or disease alleles are identified and mapped, we can select for the healthy allele or haplotype (Cole, 2015, Charlier et al., 2008). However, in practical breeding schemes, simultaneous selection against many new disease alleles together with the use of region-specific matrices to maintain variability at specific genomic regions could be more difficult to implement, or even impossible, as their number increases. Moreover, this strategy can distract substantial selection pressures away from breeding goal traits that improve production efficiency.

All the above consequences of inbreeding occur because of loss of heterozygosity at fitness related loci and/or at genomic regions where variability is desired (currently or in the future (Leroy, 2014)). The additive genetic variance of any trait is proportional to the heterozygosity at the loci controlling the trait (Falconer and Mackay, 1996). Apart from genomic regions that are of special interest for genetic variability, the overall goal of genomic inbreeding management in breeding schemes is to maintain genomic heterozygosity at all loci that are not affecting the breeding goal traits. This approach addresses all the above consequences of inbreeding using a single molecular genetic parameter namely the average heterozygosity at non-QTL loci (Meuwissen et al., 2018).

True inbreeding

The rate of true inbreeding is the rate at which heterozygosity is lost throughout the genome. Since the number QTL is small relative to the total number of loci in the genome [e.g. ~84,000,000 in cattle (1000 bull genomes project)(Hayes and Daetwyler, 2018)], excluding QTL (since allele frequency changes at QTL are desirable) or not excluding QTL hardly affects the rate of true inbreeding. True inbreeding measures both components of inbreeding, i.e. inbreeding due to pedigree and due to co-selection of chromosomal regions. We compared the role of alternative genomic relationship matrices and prediction methods on the rate of genetic gain at the same rate of true inbreeding (Paper 1 and 2), measured with IBD markers. In practice, measures of true genomic inbreeding (IBD markers) are not available to the breeder. Additionally, the relationship between rate of true inbreeding based on the IBD markers and SNP-chip based inbreeding, i.e. the marker panel used for genomic selection, is not obvious. However, our results in Paper 1 suggest that if the number DNA polymorphisms is large, alternative panels of DNA polymorphisms yield very similar estimates of genomic relationships (e.g. a large QTL panel of QTL polymorphisms resulted in the same relationships and thus accuracies of selection as a large SNP marker panel). The lack of observable measures of true inbreeding and the unclear relationship between the rate of true inbreeding based on IBD markers and genomic inbreeding could hinder the efficiency of inbreeding control in practical genomic optimum contribution breeding schemes. Thus, it is an area of urgent research priority to develop an observable estimator of true genomic inbreeding.

In this regard, with the availability of whole genome sequence (WGS) data, Meuwissen et al. (2018) recommended the use of average heterozygosity/homozygosity at neutral linked loci as true measures of inbreeding. This is because WGS data consists of all the fitness, disease and other loci of future interest, and can be directly used to manage inbreeding. Moreover, this information will

probably become available in the near future for practical use. For future research it is thus important to verify our findings in Papers 1 and 2, with respect to the role of alternative genomic relationship matrices and their impact on selection sweeps including the effect of Bayesian variable selection methods, when using whole genome sequence data and average heterozygosity at neutral linked loci as true measures of inbreeding.

The genomic relationship matrix

Currently the availability of large numbers of genetic markers as in whole genome sequencing data (WGS) is increasing in livestock species. Our results with 7702 QTL (Paper 1), namely no differences in rates of genetic gain using alternative genomic relationship matrices at the same rate of true inbreeding could have practical implications on how to use the big genomic data in genomic optimum contribution breeding schemes. The increased availability of such big genomic data enables us to detach the set of markers used for prediction from the set of markers used to control genomic inbreeding in genomic optimum contribution selection as suggested by (Meuwissen et al., 2018). If both sets of markers are sufficiently large, their sampling errors on relationship estimates are negligible. Therefore, both sets of markers give very similar genomic relationship estimates among individuals, which finally results in our finding of no differences in rates of genetic gain using one set of markers for prediction and another set of markers for genomic inbreeding management.

Alternatively, in the presence of abundant loci like in WGS data, a G-matrix for the control of inbreeding can be constructed weighing the SNPs along the genome differently, in order to vary the level of inbreeding control across the genome. We can construct the G-matrix by applying more weight at all the fitness, disease and potential future interest loci, but we do not give weight at the QTL regions. Use of this weighted G-matrix to control inbreeding could help to relax

inbreeding control in the QTL regions, but increase the inbreeding control at the loci where we apply more weight to reduce genetic drift. By doing this, we may manage better inbreeding depression and loss of genetic variation around the QTL regions. This relaxed inbreeding control in the QTL regions could also allow the selection to increase the frequencies of favorable QTL-alleles and may result in higher rates of genetic gain. Hence, the role of varying the inbreeding control across the genome using weighted G-matrices on the management of consequence of inbreeding and the interaction with the prediction methods warrants a promising line of future research.

Runs of homozygosity

Genomic inbreeding can be measured in many ways. All measures of genomic inbreeding have limitations and there is no full agreement on which measures of genomic inbreeding to use. In paper 3, we assessed the impact of inbreeding depression on semen quality traits in a pig population using different measures of genomic inbreeding. Our results showed that semen quality traits had significant association with inbreeding estimated by ROH, excess of homozygosity, and homozygosity based measures of genomic inbreeding. However, FROH> 2Mb and FROH>= 5Mb were the most closely associated with inbreeding depression since they resulted in more significant p-values for inbreeding depression than the other measures of genomic inbreeding coefficients. Other studies also reported significant effects of inbreeding on production and fertility traits using the ROH based measures of inbreeding (Martikainen et al., 2017, Bjelland et al., 2013, Ferenčaković et al., 2017, Howard et al., 2015). In the literature, the ROH based measures of inbreeding especially using the long ROHs are currently considered the most precise estimates of genomic inbreeding (Curik et al., 2014, Keller et al., 2011, Martikainen et al., 2018). Since, the long ROHs unlikely occur by chance they are likely rather to represent autozygosity which occurs

as the result of inbreeding (Curik et al., 2014). In particular the long runs of homozygosity occur due to inbreeding loops with a recent common ancestor, i.e. they are due to recent inbreeding. Inbreeding depression is expected to be more severe for recent inbreeding than for old inbreeding, since in the case of old inbreeding purging of deleterious alleles is more effective due the long time span. Hence, it is expected that ROH based inbreeding associates well with inbreeding depression and we recommended these measures of inbreeding to study inbreeding depression. More research is needed into the interrelationships between the length of the ROHs, the age of the inbreeding and how well it associates to inbreeding depression, where the latter may well depend on the selection history of the population.

Genomic relationship matrices based on ROHs have been used to assess the effects of genomic regions on economically important traits (Howard et al., 2015), to predict genomic breeding values (Luan et al., 2014), and to maintain diversity and fitness in genetic conservation programmes (de Cara et al., 2013). ROHs naturally measure inbreeding but may also be used to measure genomic relationships. The latter requires the phasing of the animals' genotypes (Sargolzaei et al., 2014, Browning et al., 2018) such that the genotypes of possible, putative offspring of the animals can be predicted including their expected ROHs. The inbreeding of the putative offspring equal half the relationship of the parents, and ROH-based genomic relationships of the parents (the animals we are interested in) is easily estimated. Given the success of ROH based inbreeding in predicting inbreeding depression (Paper 3), and given our interests here in the management of inbreeding, we consider below whether ROH based estimates of genomic relationships are useful for genomic inbreeding management.

ROHs based management of genomic inbreeding has not been well studied in genomic optimum contribution selection breeding schemes. Preliminary simulation results (results not presented in thesis) showed that management of genomic inbreeding using ROH based relationship matrices achieved larger genetic gains compare to SNP-by-SNP based genomic relationship matrix in genomic optimum contribution selection (Christian Maltecca et al., 2018). However, the rate of genetic gains were compared at the same rate of pedigree inbreeding instead of at the same rate of true genomic inbreeding. With this limitation in mind, the results showed that the ROH based genomic relationship matrix could be considered for the management of genomic inbreeding in genomic optimum contribution selection.

Despite of its advantages, the ROH based measures of inbreeding and relationship have limitations. (1) Inbreeding or relationship among individuals calculated using ROHs could under-estimate the inbreeding and the relationship among individuals, since the shorter IBD regions were not considered in the estimation of inbreeding and relationship (and some minimum length limitation is needed for defining ROHs). (2) Unlike the pedigree inbreeding coefficients, ROH based inbreeding coefficients do not accumulate over generations since the ROHs break up over generations due to chromosomal recombination (Meuwissen et al., 2018). Moreover, the pedigree inbreeding coefficient is expressed relative to a well-defined base populations but ROH based inbreeding coefficients have a less well defined base population (Curik et al., 2014, Malécot, 1948). Thus, inbreeding coefficients based on ROH cannot be directly compared to the pedigree based inbreeding coefficient. (3) When long stretches of homozygous genomic regions are advantageous, ROH based genomic relationship matrices to manage genomic inbreeding in genomic optimum contribution selection can limit the frequencies of the favorable long stretch of homozygous genomic region in the population. (4) Measuring recent inbreeding, as performed by

ROHs, was advantageous for the prediction of inbreeding depression (as explained above), but the management of inbreeding in breeding schemes may also require the estimation of relationships due to more ancient inbreeding. Thus, the use of ROH based measures of inbreeding/relationship for the management of genomic inbreeding in genomic optimum contribution selection needs further investigation in future studies.

CONCLUSIONS

- The role of alternative genomic relationship matrices in genomic optimum contribution selection depends on the number of QTL controlling the traits. Alternative genomic relationship matrices can be used to control inbreeding versus for genomic prediction.
- If the traits are controlled by a large number of QTL it is not critical whether we use the QTL based or marker based genomic relationship matrices to predict breeding values.
- With low numbers of QTL, genomic relationship matrices, which give higher accuracy and allow to vary the rates of inbreeding across the genome, yield higher genetic gains at the same overall rate of genomic inbreeding in genomic optimum contribution selection.
- The variable selection method realized higher accuracy, which resulted in higher rates of genetic gain, than non-variable selection methods.
- The variable selection method is in line with the objective of genomic optimum contribution selection breeding schemes in that it gives higher rates of genetic gain as well as smaller selective sweeps in the QTL regions than GBLUP.
- Among the genomic measures of inbreeding, ROH based measures of inbreeding showed stronger association with inbreeding depression. We recommended to use long ROHs to measure inbreeding and to study inbreeding depression.
- An increase of inbreeding affects semen quality traits and control of inbreeding strategies should be considered by the pig industry in order to get high-quality semen in sufficient quantity.

REFERENCES

BÉRÉNOS, C., ELLIS, P. A., PILKINGTON, J. G. & PEMBERTON, J. M. 2016. Genomic analysis reveals depression due to both individual and maternal inbreeding in a free-living mammal population. *Molecular ecology,* 25**,** 3152-3168.

BJELLAND, D., WEIGEL, K., VUKASINOVIC, N. & NKRUMAH, J. 2013. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *Journal of Dairy Science,* 96**,** 4697-4706.

BROWNING, B. L., ZHOU, Y. & BROWNING, S. R. 2018. A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics,* 103**,** 338-348.

CHARLIER, C., COPPIETERS, W., ROLLIN, F., DESMECHT, D., AGERHOLM, J. S., CAMBISANO, N., CARTA, E., DARDANO, S., DIVE, M. & FASQUELLE, C. 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature genetics,* 40**,** 449.

CHRISTIAN MALTECCA, GEBREYOHANS GEBREGIWERGIS, JEREMY T. HOWARD, CHRISTINE F. BAES & TIEZZI, F. 2018. Managing population diversity through genomic optimal contribution selection. Paper presented to American Dairy Science Association annual meeting, Knoxville, Tennessee.

COLE, J. B. 2015. A simple strategy for managing many recessive disorders in a dairy cattle breeding program. *Genetics Selection Evolution,* 47**,** 94.

CURIK, I., FERENČAKOVIĆ, M. & SÖLKNER, J. 2014. Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livestock Science,* 166**,** 26-34.

CURIK, I., FERENČAKOVIĆ, M. & SÖLKNER, J. 2017. Genomic dissection of inbreeding depression: a gate to new opportunities. *Revista Brasileira de Zootecnia,* 46**,** 773-782.

CURIK, I., SÖLKNER, J. & STIPIC, N. 2002. Effects of models with finite loci, selection, dominance, epistasis and linkage on inbreeding coefficients based on pedigree and genotypic information. *Journal of Animal Breeding and Genetics,* 119**,** 101-115.

DE CARA, M. Á. R., VILLANUEVA, B., TORO, M. Á. & FERNÁNDEZ, J. 2013. Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular ecology,* 22**,** 6091-6099.

FALCONER, D. S. & MACKAY, T. F. 1996. Introduction to quantitative genetics. Longman, Essex United Kingdom.

FERENČAKOVIĆ, M., SÖLKNER, J. & CURIK, I. 2013. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution,* 45**,** 42.

FERENČAKOVIĆ, M., SÖLKNER, J., KAPŠ, M. & CURIK, I. 2017. Genome-wide mapping and estimation of inbreeding depression of semen quality traits in a cattle population. *Journal of dairy science,* 100**,** 4721-4730.

FERNÁNDEZ, B., SANTIAGO, E., TORO, M. A. & CABALLERO, A. 2000. Effect of linkage on the control of inbreeding in selection programmes. *Genetics Selection Evolution,* 32**,** 249.

FORUTAN, M., MAHYARI, S. A., BAES, C., MELZER, N., SCHENKEL, F. S. & SARGOLZAEI, M. 2018. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC genomics,* 19**,** 98.

FRAGOMENI, B. O., LOURENCO, D. A., MASUDA, Y., LEGARRA, A. & MISZTAL, I. 2017. Incorporation of causative quantitative trait nucleotides in single-step GBLUP. *Genetics Selection Evolution,* 49**,** 59.

GODDARD, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica,* 136.

GOMEZ-RAYA, L., RODRÍGUEZ, C., BARRAGÁN, C. & SILIÓ, L. 2015. Genomic inbreeding coefficients based on the distribution of the length of runs of homozygosity in a closed line of Iberian pigs. *Genetics Selection Evolution,* 47**,** 81.

GÓMEZ-ROMANO, F., VILLANUEVA, B., FERNÁNDEZ, J., WOOLLIAMS, J. A. & PONG-WONG, R. 2016. The use of genomic coancestry matrices in the optimisation of contributions to maintain genetic diversity at specific regions of the genome. *Genetics Selection Evolution,* 48**,** 2.

HAYES, B. J. & DAETWYLER, H. D. 2018. 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. *Annual review of animal biosciences*.

HAYES, B. J., VISSCHER, P. M. & GODDARD, M. E. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics research,* 91**,** 47-60.

HILL, W. & WEIR, B. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research,* 93**,** 47-64.

HOWARD, J. T., HAILE-MARIAM, M., PRYCE, J. E. & MALTECCA, C. 2015. Investigation of regions impacting inbreeding depression and their association with the additive genetic effect for United States and Australia Jersey dairy cattle. *BMC genomics,* 16**,** 813.

HOWRIGAN, D. P., SIMONSON, M. A. & KELLER, M. C. 2011. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC genomics,* 12**,** 460.

JANNINK, J.-L. 2010. Dynamics of long-term genomic selection. *Genetics Selection Evolution,* 42**,** 1.

JONES, A. G., SMALL, C. M., PACZOLT, K. A. & RATTERMAN, N. L. 2010. A practical guide to methods of parentage analysis. *Molecular ecology resources,* 10**,** 6-30.

KARDOS, M., LUIKART, G. & ALLENDORF, F. 2015. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity,* 115**,** 63.

KARDOS, M., TAYLOR, H. R., ELLEGREN, H., LUIKART, G. & ALLENDORF, F. W. 2016. Genomics advances the study of inbreeding depression in the wild. *Evolutionary applications,* 9**,** 1205-1218.

KELLER, M. C., VISSCHER, P. M. & GODDARD, M. E. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics,* 189**,** 237-249.

LEROY, G. 2014. Inbreeding depression in livestock species: review and meta-analysis. *Animal genetics,* 45**,** 618-628.

LIU, H., MEUWISSEN, T. H., SØRENSEN, A. C. & BERG, P. 2015. Upweighting rare favourable alleles increases long-term genetic gain in genomic selection programs. *Genetics Selection Evolution,* 47**,** 19.

LIU, H., SORENSEN, A. & BERG, P. Optimum contribution selection combined with weighting rare favourable alleles increases long-term genetic gain. Proceedings of the 10th world congress on genetics applied to livestock production, 2014a. 17-22.

LIU, H., SØRENSEN, A. C., MEUWISSEN, T. H. & BERG, P. 2014b. Allele frequency changes due to hitch-hiking in genomic selection programs. *Genetics Selection Evolution,* 46**,** 8.

LUAN, T., YU, X., DOLEZAL, M., BAGNATO, A. & MEUWISSEN, T. H. 2014. Genomic prediction based on runs of homozygosity. *Genetics Selection Evolution,* 46**,** 64.

MALÉCOT, G. 1948. Les Mathématiques de l'Hérédité, Masson, Paris (translated as The Mathematics of Heredity.

MARTIKAINEN, K., SIRONEN, A. & UIMARI, P. 2018. Estimation of intrachromosomal inbreeding depression on female fertility using runs of homozygosity in Finnish Ayrshire cattle. *Journal of dairy science,* 101**,** 11097-11107.

MARTIKAINEN, K., TYRISEVÄ, A.-M., MATILAINEN, K., PÖSÖ, J. & UIMARI, P. 2017. Estimation of inbreeding depression on female fertility in the Finnish Ayrshire population. *Journal of Animal Breeding and Genetics,* 134**,** 383-392.

MCQUILLAN, R., LEUTENEGGER, A.-L., ABDEL-RAHMAN, R., FRANKLIN, C. S., PERICIC, M., BARAC-LAUC, L., SMOLEJ-NARANCIC, N., JANICIJEVIC, B., POLASEK, O. & TENESA, A. 2008. Runs of homozygosity in European populations. *The American Journal of Human Genetics,* 83**,** 359-372.

MEUWISSEN, T. 1997. Maximizing the response of selection with a predefined rate of inbreeding. *Journal of animal science,* 75**,** 934-940.
MEUWISSEN, T. H., A.K. SONESSON & WOOLLIAMS, J. A. 2018. Genomic management of inbreeding in breeding schemes. *Joint sessions ICAR-WCGALP.* paper presented to joint sessions ICAR-WCGALP, Auckland.
MEUWISSEN, T. H., HAYES, B. J. & GODDARD, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics,* 157**,** 1819-1829.
NEJATI-JAVAREMI, A., SMITH, C. & GIBSON, J. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of animal science,* 75**,** 1738-1745.
PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. & DALY, M. J. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics,* 81**,** 559-575.
SARGOLZAEI, M., CHESNAIS, J. P. & SCHENKEL, F. S. 2014. A new approach for efficient genotype imputation using information from relatives. *BMC genomics,* 15**,** 478.
SAURA, M., FERNÁNDEZ, A., VARONA, L., FERNÁNDEZ, A. I., DE CARA, M. Á. R., BARRAGÁN, C. & VILLANUEVA, B. 2015. Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. *Genetics Selection Evolution,* 47**,** 1.
SMITH, J. M. & HAIGH, J. 1974. The hitch-hiking effect of a favourable gene. *Genetics Research,* 23**,** 23-35.
SONESSON, A. K., WOOLLIAMS, J. A. & MEUWISSEN, T. H. 2012. Genomic selection requires genomic control of inbreeding. *Genetics Selection Evolution,* 44**,** 1.
VANRADEN, P. 2008. Efficient methods to compute genomic predictions. *Journal of dairy science,* 91**,** 4414-4423.
VANRADEN, P., OLSON, K., NULL, D. & HUTCHISON, J. 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *Journal of dairy science,* 94**,** 6153-6161.
WANG, J. 2014. Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. *Journal of Evolutionary Biology,* 27**,** 518-530.
WOOLLIAMS, J., BERG, P., DAGNACHEW, B. & MEUWISSEN, T. 2015. Genetic contributions and their optimization. *Journal of Animal Breeding and Genetics,* 132**,** 89-99.
WRAY, N. & GODDARD, M. 1994. Increasing long-term response to selection. *Genetics Selection Evolution,* 26**,** 431.
WRIGHT, S. 1922. Coefficients of inbreeding and relationship. *The American Naturalist,* 56**,** 330-338.
WRIGHT, S. 1949. The genetical structure of populations. *Annals of eugenics,* 15**,** 323-354.
YANG, J., BENYAMIN, B., MCEVOY, B. P., GORDON, S., HENDERS, A. K., NYHOLT, D. R., MADDEN, P. A., HEATH, A. C., MARTIN, N. G. & MONTGOMERY, G. W. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics,* 42**,** 565-569.
YANG, J., LEE, S. H., GODDARD, M. E. & VISSCHER, P. M. 2011. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics,* 88**,** 76-82.
ZHANG, Q., CALUS, M. P., GULDBRANDTSEN, B., LUND, M. S. & SAHANA, G. 2015. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC genetics,* 16**,** 88.
ZHANG, Z., LIU, J., DING, X., BIJMA, P., DE KONING, D.-J. & ZHANG, Q. 2010. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PloS one,* 5**,** e12648.

Paper 1

Controlling inbreeding in optimum-contribution selection with alternative genomic relationship matrices.

G.T. Gebregiwergis, Anders C Sørensen, Mark Henryon, Theo HE Meuwissen

Submitted to Journal: Frontiers in Genetics

Controlling inbreeding in optimum-contribution selection with alternative genomic relationship matrices

Authors: G.T. Gebregiwergis[1*], Anders C Sørensen[2], Mark Henryon[3,4], Theo HE Meuwissen [1]

[1] Department of Animal and Aquaculture Sciences, Norwegian University of life Sciences, Norway, [2] Department of Molecular Biology and Genetics, Aarhus University, Tjele DK-8830, Denmark, [3] Seges, Axeltorv 3, DK-1609 København V, Denmark, [4]The University of Western Australia, School of Animal Biology, 35 Stirling Highway, CRAWLEY WA 6009, Australia

***Correspondence:**

Gebreyohans tesfaye Gebregiwergis
Gebreyohans.tesfaye.gebregiwergis@nmbu.no

Keywords: true inbreeding, genetic gain, genomic optimum contribution selection, genomic relationship matrices, prediction

Abstract

We tested the consequences of using alternative genomic relationship matrices to predict genomic breeding values (GEBVs) and control inbreeding in optimum contribution selection, where the relationship matrix used to calculate GEBVs was not necessarily the same as that used to control inbreeding.

A stochastic simulation study was carried out to investigate genetic gain and true genomic inbreeding in breeding schemes that applied genomic optimum contribution selection (GOCS) with different genomic relation matrices. Three genomic-relationship matrices were used to predict the genomic breeding values based on three information sources: markers (G_M), QTL (G_Q), and markers and QTL (G_A). Two genomic-relationship matrices were used to control inbreeding: G_M

and G_A. Three genetic architectures were simulated: with 7702, 1000 and 500 QTL together with 54218 markers. Selection was for a single trait with heritability 0.2. All selection candidates were phenotyped and genotyped before selection. With 7702 QTL, there were no significant differences in rates of genetic gain at the same rate of true inbreeding using different genomic relationship matrices in GOCS. However, as the number of QTL was reduced to1000, prediction of GEBVs using a genomic relationship matrix constructed based on G_Q and control of inbreeding using G_M realized 29.7% higher genetic gain than using G_M for both prediction and control of inbreeding. These findings indicate that with large numbers of QTL, it is not critical what information, i.e. markers or QTL, is used to construct genomic-relationship matrices. However, it becomes critical with small numbers of QTL. This highlights the importance of using genomic-relationship matrices that focus on QTL regions when the number of QTL is small in GOCS, where the relationships used to predict GEBVs may differ from that used to control the inbreeding.

Background

Optimum contribution selection (OCS) is a selection method that maximizes genetic gain while controlling inbreeding (Meuwissen, 1997). It does this by optimizing the genetic contribution of selection candidates to the next generation using estimated breeding values and genetic relationships between candidates. A pedigree-based relationship matrix (**A**) was initially used to control inbreeding (Meuwissen, 1997). However, pedigree relationships have limitations. The **A**-matrix measures relationships and inbreeding at neutral, unlinked, and independent loci. But, genomic regions flanking Quantitative Trait Loci (QTL) under selection lose more variation than neutral regions of the genome (Roughsedge et al., 2008). It also does not consider variation due to Mendelian sampling during gamete formation, assuming the same relationship between all full-sibs (Nejati-Javaremi et al., 1997; Avendaño et al., 2005). Dense panels of Single Nucleotide Polymorphism (SNP) markers may be used to trace Mendelian segregation at marker loci (Hayes et al., 2009). Therefore, genomic markers might help to overcome some of the limitations imposed by pedigree.

There are several methods available to calculate genomic-relationships matrices (Nejati-Javaremi et al., 1997; Eding and Meuwissen, 2001; VanRaden, 2008; Yang et al., 2010; VanRaden et al., 2011). They have been used in different settings to realise high accuracies of genomic-prediction of breeding values and increase genetic gain (Jannink, 2010; Gómez-Romano et al., 2016). Moreover, genomic relationships that incorporate QTL information realise higher accuracies of genomic breeding values than genomic relationships constructed based on markers only (Nejati-Javaremi et al., 1997, Zhang et al., 2010). Accuracies of genomic breeding values close to one have been achieved using genomic relationships constructed from QTL (Fragomeni et al. 2017). Although, adding QTL information in the construction of genomic relationship matrices improved

accuracies of prediction, there is no full understanding on the interaction of the use of alternative genomic relationship matrices in genomic optimum contribution selection schemes.

Optimum contribution selection can be extended into genomic optimum contribution selection by using genomic information for both prediction of breeding values and estimation of relationship among selection candidates to manage inbreeding (Sonesson et al., 2012; Woolliams et al., 2015). Although, genomic relationship matrices that are based on dense SNP markers can reflect true relationship between individuals with a high degree of precision (Goddard, 2009), the covariances between additive genetic values of individuals for a specific trait are more accurately estimated using the relationships based on causal loci than SNP markers (Zhang et al., 2010; Habier et al., 2013; Yang et al., 2015). Hence, our hypothesis is that the use of genomic relationship matrices based on QTL for the prediction of genomic breeding value (GEBV) and marker based genomic relationships for OCS to control inbreeding could have a synergistic action on the rate of genetic improvement.

In this study, we investigated the use of alternative genomic relationship matrices for the prediction of GEBV and for the management of inbreeding, where these relationship matrices are not necessarily identical. We investigated these combinations of relationship matrices by simulating three genetic architectures with 7702, 1000 and 500 QTL. Alternative genomic relationship matrices were calculated using different genomic information sources such as SNP markers, QTL, and both SNP markers and QTL.

Methods

We used stochastic simulations of breeding schemes to estimate rates of genetic gain realized by OCS at the same rate of true inbreeding with three matrices for prediction and two matrices to control inbreeding using three genetic architectures. The three prediction matrices were constructed using genetic markers (G_M), QTL (G_Q), and markers and QTL (G_A). The two matrices to control inbreeding were G_M and G_A. G_Q was not used to control inbreeding because allele frequency changes at the QTL were assumed desirable (increasing positive allele frequencies) in order to realise genetic gains. The six alternative genomic relationship matrix combinations for prediction and control of inbreeding were A_A (both prediction and inbreeding-control use G_A), M_A (prediction of GEBV using G_M and control of inbreeding using G_A), Q_A (prediction of GEBV using G_Q and control of inbreeding using G_A), A_M (prediction of GEBV using G_A and control of inbreeding using G_M), M_M (both prediction and control of inbreeding using G_M), Q_M (prediction of GEBV using G_Q and control of inbreeding using G_M). The three investigated genetic architectures were: very many QTL (7702), a large number of QTL (1000), and relatively few QTL (500). QTL were randomly positioned on the genome of 18 chromosomes of equal length (167 cM), which resembles the pig genome.

Simulation of Genome and Population

Founder population

A schematic representation of the simulated breeding scheme is presented in Figure 1. Using 25 males and 25 females, a founder population was initiated and simulated for 1000 discrete generation. And, the effective-population size (Ne) of 50 was kept constant in each generation. The founder population genomes consisted of 3006 cM contained 30000,000 equidistant monomorphic loci (both markers and QTL; 1 x 10^4 loci per cM). The ratio of QTL loci to marker

loci was 1:7. As a result 1/8 of the monomorphic loci (3.75×10^6) were QTL loci and the remaining loci were SNP markers. Mutation rate was assumed to be 4×10^{-6} per locus in order to generate bi-allelic polymorphism at mutated loci.

Recombination-drift-mutation-selection equilibrium of the founder population was reached after 1000 generations. Moreover, linkage disequilibrium between the QTL and markers alleles was established during the simulation of the founder population with a Fisher-Wright inheritance model (Fisher 1930, Wright 1931).

In each generation, male and female parents of the next generation were randomly sampled with replacement from the 25 males and 25 females of the current generation. The additive-genetic effect of the original allele at each QTL locus was set to zero. The additive-genetic effect of the mutant allele at each QTL were sampled from an exponential distribution and it was assumed to be positive with a probability of 0.1. Selection was introduced by culling and resampling approximately 5% of animals with the lowest true breeding value (TBV) in each generation.

The TBV of an individual in the founder population was calculated as:

$TBV_i = \sum_{j=1}^{n} (x_{ij}\, g_j)$, where n is the number of QTL across the genome in the *i-th* founder animal, x_{ij} is the number of copies of the mutant allele that animal i inherited at the jth QTL ($x_{ij} = 0, 1, 2$), and g_j is the additive effect of the mutant allele at the jth QTL.

The average decay of linkage disequilibrium with distance between the segregating marker loci in generation 1000 of the founder population was similar to the average decay seen in the three commercial breeds of Danish pigs (Wang et al., 2013).

The founder population had 61920 (7702 QTL and 54218 marker) segregating loci at generation 0. All 54218 segregating marker loci were used in our breeding schemes. The number of segregating QTL used in the breeding schemes were all segregating QTL (7702 QTL), 1000 or

500 QTL. The number of segregating QTL across the genome were reduced to a desired number, i.e., 1000 QTL or 500 QTL by random sampling from the 7702 QTL.

The additive-genetic effects of the segregating mutant QTL alleles (7702, 1000 or 500) were standardised in order to get a total additive–genetic variance of 1 for the trait under selection in the founder population at generation 0.

Chromosomes of the 50 founder animals were pooled in to 18 pools of 100 chromosomes (i.e., 50 founder animals x 2 chromosome per chromosome pair) in generation 0.

Base Population

In each replicate of the simulation, a unique base population with a size of 110 animals (10 males and 100 females) was sampled from the pools of chromosomes of the founder population to initiate the breeding schemes. The genotype of each base animal was sampled from the 18 pools of chromosomes in the founder population. For chromosome j ($j = 1 \ldots 18$), two chromosomes were randomly sampled without replacement from the jth pool of 100 chromosomes. The sampled chromosomes were replaced before the next base animal was sampled. Base animals were assumed unrelated and non-inbred based on pedigree and IBD alleles. They were genotyped, but not phenotyped.

Identical by Descent markers

A total of 12024 (Indentical By Descent) IBD loci were used to measure the true rate of inbreeding (ΔIBD) and were placed evenly across the genomes in the base population at 4 IBD loci per cM. Base animals were assigned unique alleles at each IBD locus (i.e., $2n$ distinct alleles at each IBD locus among the n animals in the base population), such that identical alleles in any later generation indicates that the loci are IBD with a single unique base population allele. The IBD loci were not involved in any way in the selection.

Simulation of phenotypic values

Phenotypic value of animal *i*, P_i, were simulated as: $P_i=TBV_i+e_i$, where e_i is an error term for individual *i* sampled from $e_i \sim N(0, \sigma_e^2 = 4)$ resulting in a trait heritability of 0.2; and TBV_i is the true breeding value of animals which was obtained as described above.

Genomic estimated breeding values

The G-BLUP model (Meuwissen et al., 2001) was used to predict GEBVs:

$$\mathbf{y}=\mathbf{1}\mu+\mathbf{Zg}+\mathbf{e}, \quad (1)$$

where **y** is a vector of phenotypes, μ is the overall mean, **1** is a vector of ones, **Z** is a design matrix allocating records to breeding values, **g** is a vector of breeding values for all animals with Var (**g**)= $\mathbf{G}\boldsymbol{\sigma}_g^2$, **G** is the genomic relationship matrix, and $\boldsymbol{\sigma}_g^2$ is the additive genetic variance. The term **e** is a vector of normal independent and identical distributed residuals with variance σ_e^2.

Genomic-relationship matrices

Genomic relationship matrices (**G**) were computed using VanRaden method 2:

$$G = \frac{ZD^{-1}Z\prime}{L},$$

where $\boldsymbol{Z}$ is a matrix of centered marker genotypes by subtracting the mean of the marker or QTL genotypes; L is the number of loci; D is a diagonal matrix with entries $2p_i(1-p_i)$; and p_i is frequency of the minor allele at locus *i* in the base population. All animals in the base population were used to calculate p_i to center and scale genotypes at locus *i*. After scaling, each locus obtained equal weight. The prediction and control of inbreeding matrices, G_M, G_Q, and G_A, were constructed using marker, QTL, and marker and QTL loci.

Truncation selection

The base population animals were randomly mated to produce 500 offspring with equal sex ratio in generation 1. A truncation selection breeding program was conducted in generation 2-5 in order

to mimic a population that had undergone selection. Ten sires and 100 dams were truncation selected based on genomic-estimated breeding values. Each selected sire was randomly mated with 10 dam and each mating produced five offspring. As the result of these matings, 500 offspring with an equal sex ratio were obtained.

Optimum contribution selection

EVA (Evolutionary Algorithms) was used to optimize individual genetic contributions by maximizing the function U_t with respect to **c**:

$$U_t(\mathbf{c}) = \boldsymbol{c}'\hat{\mathbf{g}} - \omega\mathbf{c}'\mathbf{G}\mathbf{c}, \quad \textbf{(2)}$$

where **c** is a *n* vector of genetic contributions of the current generation to the next which is proportional to the number of offspring each animal obtains; $\hat{\mathbf{g}}$ is a *n* vector of genomic estimated breeding values, ω is a penalty applied on the average relationship of the selected parents for the next generation, and **G** is a *n* x *n* genomic relationship matrix among all animals in the population calculated as $\mathbf{G}_M$ or $\mathbf{G}_A$. In the above function, $\boldsymbol{c}'\hat{\mathbf{g}}$ and $\mathbf{c}'\mathbf{G}\mathbf{c}$ represent the average genetic value and average relationship of the new generation. For a detailed description of the EVA method see Henryon et al. (2015).

Optimum contribution selection was carried out in generations 6-11. A total of *25* matings were allocated to *500* selection candidates (approx. 250 males and 250 females*)* by OCS in each generation. Each male was allocated 0, 1, 2 … or *25* matings in correspondence to their optimum contributions, **c**. Each of 25 selected female was allocated a single mating. The 25 sire and dam were mated randomly. Each mating produced 20 offspring, resulting in 25 full-sib families and 500 offspring. Offspring were assigned as males / females with a probability of 0.5.

Data analyses

We plotted the rate of genetic gain against the rate of true inbreeding at different penalties (ω) for

the schemes with 7702 and 500 QTL. For the scheme with 1000 QTL, we presented rates of genetic gain at 1% and 0.5% rates of true inbreeding. The 1% and 0.5% rates of true inbreeding were realized by calibrating the penalty, ω, in equation (2).We also compared the accuracies of males and females selection candidates for 1000 QTL at 1% rate of true inbreeding. Rates of genetic gain were calculated as the slope of the linear regression of G_t on t where G_t is the average true genetic value of animals born in generation t (t = 6 … 11). Rates of true inbreeding (using the IBD markers) and rates of inbreeding based on pedigree were calculated as 1-exp(β), where β is a linear regression of ln(1-F_t) on t and F_t is the average coefficient of true inbreeding or pedigree inbreeding for animals born at generation t (t = 6 … 11). F_t for true inbreeding was calculated using the d = 12024 IBD markers as $F_t = \frac{1}{n_t d}\sum_{i=1}^{n_t}\sum_{j=1}^{d}\delta_{ij}$, where n_t is the number of animals born in generation t-and δ_{ij} is the IBD status at IBD-marker locus j (j = 1 … d) for animal i (i = 1 … n_t). δ_{ij} was equal to 1 if IBD locus j for animal i was IBD for a unique (base) allele, and 0 otherwise.

Software

The simulations were run using the program, ADAM (Pedersen et al., 2009). BLUP-breeding values were estimated using DMU6 (Madsen et al., 2006). OCS was carried out by EVA (Berg et al., 2006).

Results

500 QTL

QTL based prediction of breeding values with marker control of inbreeding, Q_M, realized more genetic gain at the same rate of true inbreeding than the other 5 prediction-control G-matrix combinations (see Figure 2).

1000 QTL

Also here QTL based prediction with marker based control of inbreeding, Q_M, realized more genetic gain at the same rate of true inbreeding than the other 4 prediction-control inbreeding combinations. Q_M realized between 21.5-29.7% more genetic gain than A_M, A_A, M_M and M_A at 1% rate of true inbreeding (Table 1). At 0.5 % rate of true inbreeding, it realized between 29.9 -53 % more genetic gain than A_M, A_A, M_M and M_A. Q_M realized almost the same rate of genetic gain as Q_A with both 1% and 0.5% rate of true inbreeding.Use of genomic relationship matrices computed based on QTL (G_Q) to predict GEBVs gave higher accuracy of prediction than G_M or G_A at 1% rate of true inbreeding (Table 2). The accuracy of prediction of male selection using Q_M was 12.7% higher than M_M at 1% rate of true inbreeding (Table 2).

7702 QTL

With 7702 QTL, the six genomic relationships matrix combinations for prediction and control of inbreeding realized almost the same rate of genetic gain at the same rate of true inbreeding (Figure 3). However, with an increase of the rate of true inbreeding >1%, the differences between the rate of genetic gain obtained by the different genomic relationship matrices became visible and the rate of genetic gain of Q_M and Q_A became slightly higher than the other four genomic relationships matrices combinations. However, all these differences in rate of genetic gain at the same rate of true inbreeding were rather marginal.

Discussion

Our findings partly supported our hypothesis that prediction with QTL and inbreeding control with markers, Q_M, realizes more genetic gain at the same rates of true inbreeding than prediction and inbreeding control with markers, M_M. We found that prediction with QTL and inbreeding control with markers realized more genetic gain when 500 and 1000 QTL controlled the trait under selection. However, when the trait was controlled by 7702 QTL, prediction and inbreeding control with markers realized just as much genetic gain as prediction with QTL and inbreeding control with markers. These findings are important because they highlight that when traits under selection are controlled by small numbers of QTL, we need to select directly for the QTL to maximize genetic gain at pre-defined rates of true inbreeding. This implies that we need to know where the QTL are located on the genome. On the other hand, we do not need to select directly for the QTL or know where the QTL are located to maximize genetic gain when traits are controlled by large numbers of QTL. In this scenario, prediction using markers that are in linkage disequilibrium with the QTL (and inbreeding control using markers in LD with IBD alleles) is sufficient. Therefore, the method used in prediction and inbreeding control when using OCS depends on the number of QTL controlling the trait under selection. We need to select for the QTL directly with small numbers of QTL to maximize genetic gain. With large numbers of QTL, we can simply use markers.

Prediction with QTL and inbreeding control with markers only realized more genetic gain when small numbers of (500 and 1000) QTL controlled the trait under selection for two reasons. First, prediction with QTL was more accurate than prediction with markers when small numbers of QTL controlled the trait. Prediction with QTL generated accurate breeding values because it had perfect knowledge of the true genetic (co)variance among individuals for the trait under selection.

Prediction with markers was not as accurate because there was insufficient LD between the markers and QTL. Many of the markers where not located near QTL. With high numbers of QTL, prediction with markers had similar accuracy of prediction as prediction with QTL. There was more LD between the markers and QTL when many QTL were distributed across the genome. The second possible reason was that inbreeding control restricted changes in QTL-allele frequencies less when there were small numbers of QTL controlling the trait under selection. Inbreeding control traced and penalised changes in marker-allele frequencies brought about by realised genetic drift and selection (Woolliams et al., 2015). It penalised changes in allele frequencies at all marker loci. Because these marker alleles were in linkage disequilibrium with QTL alleles, it restricted changes in QTL-allele frequencies. Inbreeding control is spread over the whole genome. With few QTL, much of the inbreeding control could be at regions of the genome that do not harbour QTL. With many QTL, inbreeding control is at all regions of the genome that harbour QTL – it penalised changes in allele frequencies at all loci, when we need to allow allele-frequency changes at some QTL loci. Therefore, genomic relationship matrices used to predict GEBV may differ from that used to control the inbreeding. In addition, genomic relationship matrices that consider the true genetic architecture of a trait under selection and allow differentiating the inbreeding rates at the QTL from the general rates at the genomic level could realize higher rates of genetic gain at the same rate of true inbreeding.

Prediction with QTL and inbreeding control with markers cannot be implemented directly in practical breeding schemes. However, this scheme does teach us some principles that apply to practical breeding schemes. Genomic relationship matrices based on QTL are not currently available in practice and it is unlikely to be available soon since exact number and position of QTL controlling a trait are not known. Although, it is not possible to get a genomic-relationship matrix

based on QTL, a trait specific genomic relationship matrix could be available from genome-wide association studies. Previous studies have shown that trait-specific genomic-relationship matrices realize higher prediction accuracy than marker-based genomic-relationship matrices but it realizes lower accuracy of prediction than the QTL-based genomic-relationship matrix (Nejati-Javaremi et al., 1997, Zhang et al., 2010, Fragomeni et al., 2017). Until QTL-based genomic-relationship matrices become available, trait-specific genomic-relationship matrices could be used for prediction. Moreover, the inbreeding can be controlled using markers present across a genome that have no association with the trait under selection to relax inbreeding control around the QTL regions. Such combinations of genomic relationship matrices for prediction and control of inbreeding realize more genetic gain than using markers for both prediction and control of inbreeding. This is true when the trait of interest deviates from the infinitesimal model assumptions. Therefore, when a trait under selection is controlled by small number of QTL, optimum contribution selection that incorporate information about the trait in the construction of genomic relationship matrix for prediction realizes more genetic gain and there is incentive for research work aiming to obtain more biological information about the QTL that code for the trait.

Our findings with small numbers of (500 and 1000) QTL controlling the trait under selection are supported by several studies that assessed the role of alternative genomic relationship matrices on accuracy of genomic estimated breeding values (Nejati-Javaremi et al., 1997; Zhang et al., 2010; Fragomeni et al., 2017). These studies addressed the role of alternative genomic-relationship matrices on prediction without considering inbreeding control. However, our results are generally supported by their findings as we extended the study towards optimum contribution selection and assessed both the prediction and control of inbreeding part together. Fragomeni et al. (2017) showed that, adding causative QTN (quantitative trait nucleotides) in unweighted genomic

relationship matrix improved the accuracy of prediction by 0.04. However, using weighted genomic relationship matrix with weights obtained by genome-wide association study, they reported an increase of accuracy by 0.1. Moreover, they reported also accuracies closer to 1.00 using genomic relationship matrices computed based on causative QTN. These findings agree with our findings on (500 and 1000) QTL where Q_M realized the highest accuracy of prediction and genetic gain. Moreover, prediction with all loci (both markers and QTL) without weighing in the construction of genomic relationship matrix) and inbreeding control with markers, A_M, realized higher accuracy of prediction than M_M. Our results also showed that the difference in rate of genetic gain obtained between Q_M and M_M became smaller as the number of QTL became larger, i.e., 500 and 1000 QTL. Moreover, this difference became insignificant as the number of QTL further increased in to 7702 QTL. Similarly, Nejati-Javaremi et al. (1997) reported higher accuracies and response to selection using genomic relationship matrices constructed based on QTL genotypes when a trait is controlled by a small number of loci. Fragomeni et al. (2017) also reported lower accuracy using 1000 QTL than 100 QTL. Therefore, with small numbers of QTL controlling the trait under selection, QTL based genomic relationship matrices realized the highest accuracies and genetic gain in both cases when prediction was studied separately or together with control of inbreeding in case of OCS.

However, care has to be taken not to focus the G matrix for the prediction of GEBV too strongly on too few QTL. The accuracy of selection will be reduced when part of the genetic variation, which is due to polygenes, is neglected. Moreover, the GBLUP model will overestimate the accuracies of the GEBV, since it assumes that there are few QTL underlying the trait, and predicts accuracies according to this assumption. The use of OCS seems to shift the balance more towards fewer QTL in the G matrix for prediction of GEBV, in order to increase the difference between

the G matrix for prediction, favoring allele frequency changes (in the right direction), and the G matrix for inbreeding control, which attempts to minimize allele frequency changes.

Conclusions

This study showed that as the number loci involved in the control of the trait of interest are large, genomic relationship matrices based on markers for both prediction and control of inbreeding, M_M, perform as good as genomic relationship matrix constructed based on QTL (Q_M) in genomic optimum contribution selection. Whereas, when the trait is controlled by a small number of genes, genomic relationship matrix constructed based on QTL (Q_M) realize higher rates of genetic gain than genomic relationships constructed based on markers (M_M) at the same rate of true inbreeding in genomic optimum contribution selection.

Table 1: Rate of genetic gain and rate of pedigree based inbreeding in genomic optimum contribution selection using different genomic information to predict GEBV and control inbreeding. Rates of true inbreeding were 1 and 0.5% and there were 1000 QTL.

G-matrices	$\Delta IBD_{=0.01}$		$\Delta IBD_{=0.005}$	
	$\Delta G_{(SE)}$	ΔF (SE)	ΔG (SE)	ΔF (SE)
Q_M	0.677(0.005)	0.012(0.0002)	0.586(0.007)	0.0086(0.0003)
Q_A	0.672(.005)	0.012(0.0002)	0.586(0.006)	0.0085(0.0003)
A_M	0.537(0.005)	0.012(0.0002)	0.433(0.006)	0.0092(0.0002)
A_A	0.544(.004)	0.012(0.0002)	0.401(0.005)	0.0093(0.0003)
M_M	0.522(0.005)	0.012(0.0003)	0.383(0.006)	0.0085(0.0003)
M_A	0.557(0.005)	0.012(0.0002)	0.451(0.007)	0.0086(0.0003)

Rate of genetic gain (ΔG), Rate of true inbreeding (ΔIBD), Rate of inbreeding based on pedigree (ΔF_s) and SE is standard errors based on 100 replicates. A_A (both prediction and inbreeding-control use genomic relationship matrix based on both marker and QTL), M_A (prediction of GEBV using genomic relationship matrix based on marker and control of inbreeding using G_A), Q_A (prediction of GEBV using G_Q and control of inbreeding using G_A), A_M (prediction of GEBV using G_A and control of inbreeding using G_M), M_M (both prediction and control of inbreeding using G_M), Q_M (prediction of GEBV using G_Q and control of inbreeding using G_M). G, M, Q and A represent genomic relationship matrix, markers, QTL and both markers and QTL respectively.

Table 2: Accuracy of GEBVs for alternative genomic relationship matrices used to predict GEBVs and control inbreeding in genomic optimum contribution selection in generation 11 and at a rate of true inbreeding of 0.01. The number of QTL=1000.

G-matrices	Male accuracy (SE)	Female accuracy (SE)
Q_M	0.728 (0.005)	0.738 (0.005)
Q_A	0.721 (0.005)	0.720 (0.005)
A_M	0.661 (0.005)	0.661 (0.005)
A_A	0.656 (0.005)	0.660 (0.005)
M_M	0.646 (0.006)	0.646 (0.006)
M_A	0.659 (0.006)	0.659 (0.005)

A_A (both prediction and inbreeding-control use genomic relationship matrix based on both marker and QTL), M_A (prediction of GEBV using genomic relationship matrix based on marker and control of inbreeding using G_A), Q_A (prediction of GEBV using G_Q and control of inbreeding using G_A), A_M (prediction of GEBV using G_A and control of inbreeding using G_M), M_M (both prediction and control of inbreeding using G_M), Q_M (prediction of GEBV using G_Q and control of inbreeding using G_M). G, M, Q and A represent genomic relationship matrix, markers, QTL and both markers and QTL respectively.

Author contribution statement

GG performed the study and drafted the manuscript. GG, THEM, ACS and MH designed the study. MH modified the ADAM program. THEM and ACS planned and coordinated the whole study. THEM and MH contributed to writing the manuscript. All authors read and approved the final manuscript.

Conflict of interest statement

Mark Henryon is employed at SEGES. All other authors declare no competing interests.

Acknowledgements

This study was funded by the Center for Genomic Selection in Animals and plants (GenSAP) supported by Innovation Fund Denmark.

References

Avendaño, S., Woolliams, J., and Villanueva, B. (2005). Prediction of accuracy of estimated Mendelian sampling terms. *Journal of Animal Breeding and Genetics* 122(5)**,** 302-308.

Berg, P., Nielsen, J., and Sørensen, M.K. (2006). "EVA: Realized and predicted optimal genetic contributions", in: *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006*: Instituto Prociência), 27-09.

Eding, H., and Meuwissen, T. (2001). Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *Journal of Animal Breeding and Genetics* 118(3)**,** 141-159.

Fragomeni, B.O., Lourenco, D.A., Masuda, Y., Legarra, A., and Misztal, I. (2017). Incorporation of causative quantitative trait nucleotides in single-step GBLUP. *Genetics Selection Evolution* 49(1)**,** 59.

Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136. doi: 10.1007/s10709-008-9308-0.

Gómez-Romano, F., Villanueva, B., Fernández, J., Woolliams, J.A., and Pong-Wong, R. (2016). The use of genomic coancestry matrices in the optimisation of contributions to maintain genetic diversity at specific regions of the genome. *Genetics Selection Evolution* 48(1)**,** 1.

Habier, D., Fernando, R.L., and Garrick, D.J. (2013). Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics* 194(3)**,** 597-607.

Hayes, B.J., Visscher, P.M., and Goddard, M.E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics research* 91(01)**,** 47-60.

Henryon, M., Ostersen, T., Ask, B., Sørensen, A.C., and Berg, P. (2015). Most of the long-term genetic gain from optimum-contribution selection can be realised with restrictions imposed during optimisation. *Genetics Selection Evolution* 47(1)**,** 1.

Jannink, J.-L. (2010). Dynamics of long-term genomic selection. *Genetics Selection Evolution* 42(1)**,** 1.

Madsen, P., Sørensen, P., Su, G., Damgaard, L.H., Thomsen, H., and Labouriau, R. (Year). "DMU-a package for analyzing multivariate mixed models", in: *8th World Congress on Genetics Applied to Livestock Production*).

Meuwissen, T. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *Journal of animal science* 75(4)**,** 934-940.

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

Nejati-Javaremi, A., Smith, C., and Gibson, J. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of animal science* 75(7)**,** 1738-1745.

Pedersen, L., Sørensen, A., Henryon, M., Ansari-Mahyari, S., and Berg, P. (2009). ADAM: A computer program to simulate selective breeding schemes for animals. *Livestock Science* 121(2)**,** 343-344.

Roughsedge, T., Pong-Wong, R., Woolliams, J., and Villanueva, B. (2008). Restricting coancestry and inbreeding at a specific position on the genome by using optimized selection. *Genetical Research* 90(2)**,** 199.

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

VanRaden, P. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science* 91(11)**,** 4414-4423.

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

Wang, L., Sørensen, P., Janss, L., Ostersen, T., and Edwards, D. (2013). Genome-wide and local pattern of linkage disequilibrium and persistence of phase for 3 Danish pig breeds. *BMC genetics* 14(1)**,** 115.

Woolliams, J., Berg, P., Dagnachew, B., and Meuwissen, T. (2015). Genetic contributions and their optimization. *Journal of Animal Breeding and Genetics* 132(2)**,** 89-99.

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A., Lee, S.H., et al. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature genetics* 47(10)**,** 1114.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature genetics* 42(7)**,** 565-569.

Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning, D.-J., and Zhang, Q. (2010). Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PloS one* 5(9)**,** e12648.

Figure 1. Schematic representation of the simulated breeding scheme.

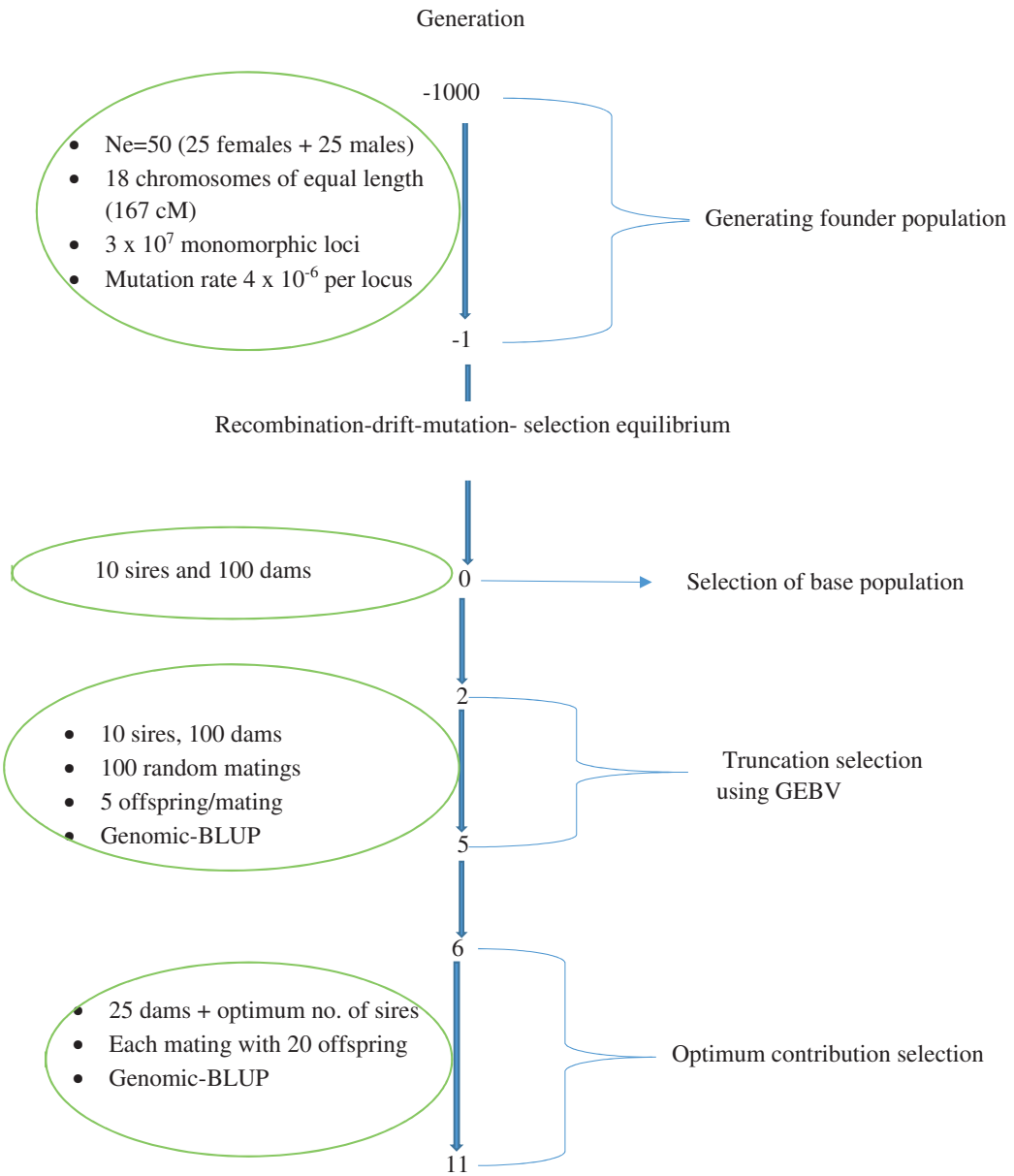
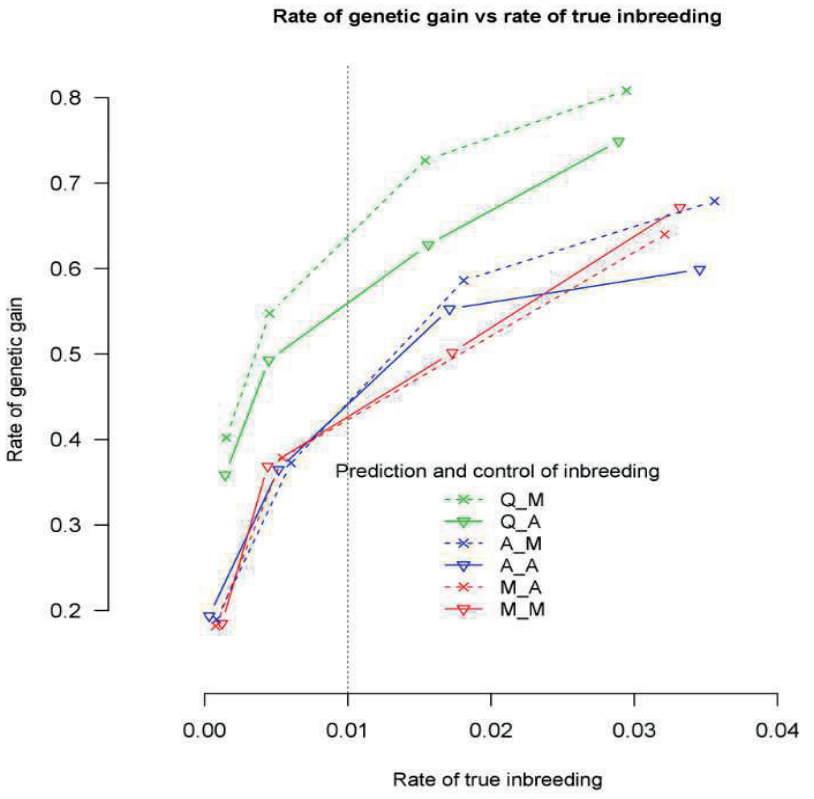
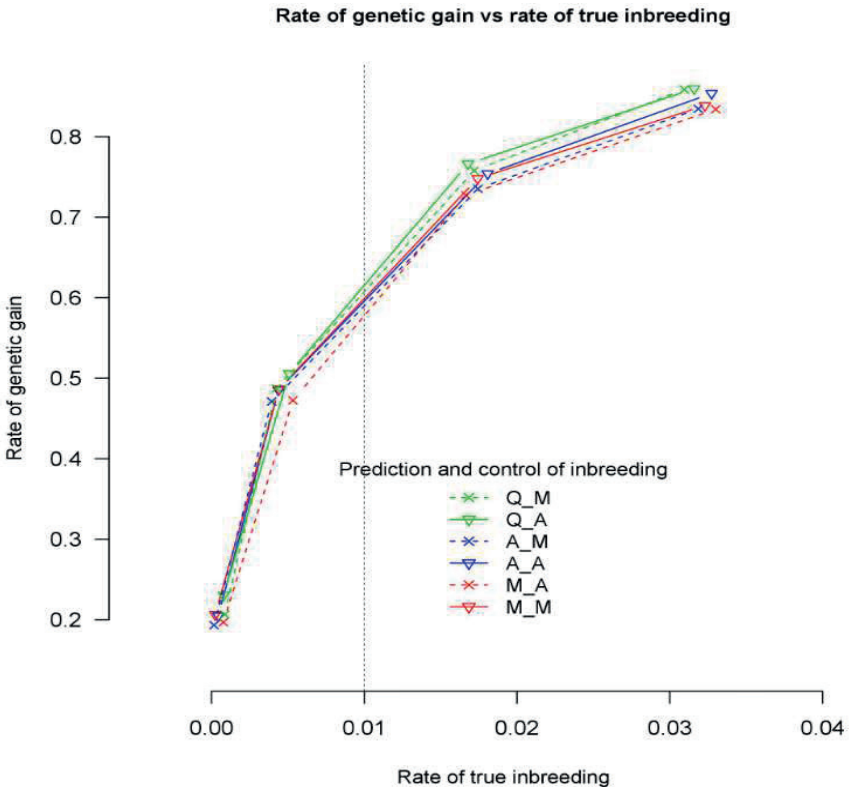


Figure 2. Rate of genetic gain and true inbreeding (IBD) using 500 QTL in genomic optimum contribution selection.



A_A (both prediction and inbreeding-control use genomic relationship matrix based on both marker and QTL), M_A (prediction of GEBV using genomic relationship matrix based on marker and control of inbreeding using G_A), Q_A (prediction of GEBV using G_Q and control of inbreeding using G_A), A_M (prediction of GEBV using G_A and control of inbreeding using G_M), M_M (both prediction and control of inbreeding using G_M), Q_M (prediction of GEBV using G_Q and control of inbreeding using G_M). G, M, Q and A represent genomic relationship matrix, markers, QTL and both markers and QTL respectively.

Figure 3: Rate of genetic gain and true inbreeding (IBD) using 7702 QTL.



A_A (both prediction and inbreeding-control use genomic relationship matrix based on both marker and QTL), M_A (prediction of GEBV using genomic relationship matrix based on marker and control of inbreeding using G_A), Q_A (prediction of GEBV using G_Q and control of inbreeding using G_A), A_M (prediction of GEBV using G_A and control of inbreeding using G_M), M_M (both prediction and control of inbreeding using G_M), Q_M (prediction of GEBV using G_Q and control of inbreeding using G_M). G, M, Q and A represent genomic relationship matrix, markers, QTL and both markers and QTL respectively.

Paper 2

Use of Bayesian genomic prediction methods in genomic optimum contribution selection.

G T Gebregiwergis, A Christian Sørensen and Theo HE Meuwissen

Submitted to Journal: Genetic Selection Evolution

Use of Bayesian genomic prediction methods in genomic optimum contribution selection.

G T Gebregiwergis[1*] A Christian Sørensen[2] Theo HE Meuwissen .[1]

[1] Department of Animal and Aquaculture Sciences, Norwegian University of life Sciences, Norway,

[2] Department of Molecular Biology and Genetics, Aarhus University, Tjele DK-8830, Denmark

*Corresponding author

Email addresses:

GG: gebreyohans.tesfaye.gebregiwergis@nmbu.no

THEM: theo.meuwissen@nmbu.no

ACS: achristian.sorensen@mbg.au.dk

ABSTRACT

Background

Variable selection methods such as Bayesian genomic prediction apply increased weights to SNPs that appear to be important for genomic prediction and yield higher accuracy than genomic best linear unbiased prediction (GBLUP). The present study aims to investigate whether variable selection methods improve genetic progress when applied in genomic optimum contribution selection (GOCS), and to investigate its effect on the distribution of the genetic drift over the genome.

Methods

We carried out a stochastic simulation study to estimate rate of genetic gain realized in GOCS using GBLUP and the Bayesian model BayesP. Three genetic architectures with 7702, 1000 and 180 QTL and 54218 markers were examined. Selection was for a single trait with additive genetic variance 1.0 and heritability 0.2. Genomic estimated breeding values predicted using GBLUP or BayesP were used as selection criteria, whereas the genomic relationship matrix was used to control the inbreeding. All selection candidates were phenotyped and genotyped before selection.

Results

We found that BayesP realized 5.7%, 1.7% and 2.7% more short-term rate of genetic gain than GBLUP with 180, 1000 and 7702 QTL respectively at 1% short-term rate of inbreeding in genomic optimum contribution selection. Moreover, accuracies of selection were higher with BayesP than GBLUP and this difference in accuracies of selection was especially large with few QTL. BayesP also caused markedly lower signatures of selection in the QTL regions than GBLUP.

Conclusions

BayesP gave higher accuracy and more genetic gain than GBLUP. Since the extra genetic gain seems to come mainly from the extra accuracy of BayesP, it seems that the choice between GBLUP and Bayesian methods depends only on the accuracy also in case of GOCS. Moreover, BayesP gave smaller selective sweeps in the QTL regions.

Background

Methods for genomic predictions are categorized into non-variable and variable selection methods [1, 2], where the variable selection methods apply increased weights to SNPs which appear to be important for genomic prediction. Many of the variable selection methods are of Bayesian nature, and specify prior distributions for SNP effects with spikes at zero (i.e. no effect of the SNP), which results in the down-weighing of large numbers of SNPs. Genomic best linear unbiased prediction (GBLUP), which is a typical non-variable selection method, weighs all SNPs equally, i.e. it regresses the apparent effect of all the SNP (measured by its least square solution) down by an equal factor, which depends on the number of SNPs and the trait heritability. This explains why GBLUP yields approximately the same accuracy for a given number of individuals in the training population and heritability regardless of numbers of QTL [1]. BayesB, a variable selection method, yielded higher accuracy than GBLUP at low numbers of QTL and its accuracy decreased and finally asymptoted as the number of QTL increased [1]. However, the advantage of BayesB over GBLUP was generally small in empirical studies [3, 4] . Although, BayesB performed as good as GBLUP and even better at low number of QTL, it is computationally demanding due to MCMC (Monte Carlo Markov Chain) sampling, and hence little used in practice. BayesP, which is also a variable selection method could be an alternative to BayesB, because it is iterative in nature and often computationally a factor 100 faster than BayesB, and at the same time yielding similar accuracies [5, 6].

In genomic selection, increasing accuracies of genomic estimated breeding values do not only result in increased rates of genetic gain but may also increase the loss of genetic diversity [7], in particular if the use of genomic selection results in a reduction of the generation interval. However, the loss of genetic diversity in livestock breeding populations can be managed while increasing

genetic gain using optimum-contribution selection (OCS) [8]. OCS achieves this by optimizing the number of genetic contribution of selection candidates to the next generation. The genetic contribution of selection candidates to the next generation are optimized based on the estimated breeding values (EBVs) and the additive-genetic relationship between selection candidates [8]. In the OCS method, the selection decisions and the evaluations of EBVs are separated and this allows the use of the most accurate prediction method [9]. Moreover, in the presence of genomic information, relationship between selection candidates can be estimated more accurately using marker information than the pedigree-based relationships (A-matrix) [10]. Furthermore, the marker information is not only used to estimate the relationship between selection candidates more accurately but also used to manage ancestral contributions [11, 12]. OCS can be further extended to genomic optimum contribution selection (GOCS) by using the genomic relationship matrix (G) instead of A to manage the increased relationships of selection candidates [9, 13]. These methods showed that more genetic gain can be achieved if there is a difference between the genomic relationships used for inbreeding control and those used for EBV estimation, which is achieved by the use of variable selection methods for EBV estimation since they weigh the SNPs differently. In addition, variable selection methods may achieve higher selection accuracy and manage genetic drift in QTL rich regions since Woolliams, Berg [9] showed that GOCS is directed at managing genetic drift at loci across the genome. The aim of this paper is to investigate whether variable selection methods improve genetic progress when applied in GOCS, and to investigate its effect on the distribution of the genetic drift over the genome.

Methods

We carried out a stochastic simulation study to estimate rate of genetic gain realized in GOCS using GBLUP and the Bayesian model BayesP [5], since its computational speed enables its application in stochastic simulations of entire breeding schemes over many generations. Three genetic architectures with 7702, 1000 and 180 QTL were examined. Genomic estimated breeding values predicted using GBLUP or BayesP were used as selection criteria for selection schemes, whereas the genomic relationship matrix was used to control the inbreeding.

GBLUP analysis:

The G-BLUP model used to predict GEBVs was:

y=1μ**+Zg+e**,

where **y** is a vector of phenotypes, μ is the overall mean, **1** is a vector of ones, **Z** is a design matrix allocating records to breeding values, **g** is a vector of breeding values for all animals with Var (**g**)= $\mathbf{G}\boldsymbol{\sigma}_g^2$ **,** where **G** is the genomic relationship matrix and $\boldsymbol{\sigma}_g^2$ is the additive genetic variance. **e** is a vector of the residuals with variance σ_e^2. We used the VanRaden method to compute **G** [14] and written as:

$$G = \frac{NN'}{\sum 2p_j(1-p_j)}$$

Where N_{ij} is the genotype of animal i for SNP j coded as $-2p_j$, $1-2p_j$ or $2-2p_j$ for homozygote, heterozygote and alternative homozygote respectively. We used the allele frequencies in the unselected base population as p_j.

BayesP analysis:

In BayesP, the prior distribution of the SNP effects was based on the Pareto principle [5, 15]: a priori x % of the SNPs with largest effect are assumed responsible for (100-x) % of the genetic variance (Vg). I.e., BayesP assumes that SNP effects come from a mixture of two normal distribution, one with large variance (σ_1^2) and one with small variance (σ_2^2), respectively. If we assume the prior for the mixing frequency to equal π=x/100 then using the Pareto principle, $\sigma_1^2=\frac{(1-\pi)Vg}{\pi M}$ and $\sigma_2^2=\frac{\pi Vg}{(1-\pi)M}$ are the variances of the large and small SNP effects respectively, where, M is the total number of SNPs. Hence, the total variance equals M ($\pi\sigma_1^2$+ (1-π) σ_2^2) = Vg. The π value used for BayesP was set to the ratio of the number of QTL simulated to the number of SNPs used [5]. The linear model used to estimate SNP effects for BayesP approach was:

$$\mathbf{y}=\boldsymbol{\mu}+\sum_{j=1}^{M}\mathbf{x}_j\mathrm{b}_j+\ \mathbf{e},$$

Where **y** is a (nx1) vector of n phenotypes; **μ** is the overall mean; M is the total number of SNPs; $\mathbf{x}_i$ is a (Mx1) vector of the M standardized SNP genotypes, i.e., $\mathbf{x}_j=\frac{-2p_j}{\sqrt{(2p_j(1-p_j))}}$, $\frac{1-2p_j}{\sqrt{(2p_j(1-p_j))}}$, or $\frac{2(1-p_j)}{\sqrt{(2p_j(1-p_j))}}$ for SNP genotype ' 0 0','0 1',or ' 1 1', respectively, and p_j is the allele frequency of SNP j; $\mathbf{b}_j$ is the effect of the j-th SNP genotype; **e** is the vector of environmental effect, with Var(**e**) =$\mathbf{I}\sigma_{\mathbf{e}}^2$; and summation is over all SNPs. BayesP used the Iterative conditional Expectation (ICE) algorithm of Meuwissen, Solberg [16] and details of the BayesP algorithm can be found Yu and Meuwissen [5]. Genomic estimated breeding values were calculated from the estimates of the SNPs effects as:

$$\hat{\mathrm{g}}_i=\sum_{j=1}^{M}X_{ij}\hat{\mathrm{b}}\mathrm{j}\ ,$$

where $\hat{g}_i$ genomic estimated breeding value of individual *i* ; X_{ij} is the standardized SNP genotype of individual i for SNP j; and $\hat{b}j$ is the estimate of the SNP effect.

Simulation of founder populations

The founder population was simulated for 1000 discrete generations with 50 animals (25 males and 25 females) in each generation using the program ADAM [17]. At 1000 generations, the population reached a mutation-selection-recombination-drift equilibrium and at this generation their chromosomes were pooled. From this chromosome pool of 100 haplotypes, we sampled a base population with a size of 110 animals (10 males and 100 females) in generation 0.

Simulation of the Genome

The genome simulation were performed with pig breeding in mind, in particular the commercial breeds of the Danish pig breeders. Thus, the genome consisted of 18 chromosomes of equal length (167 cM) with 30,000,000 loci (both markers and QTL). Linkage disequilibria between the QTL and markers alleles was established during simulation of the founder population with a Fisher-Wright inheritance model [18, 19]. The probability of mutation occurring at each locus was assumed 2×10^{-6} per meiosis. Approximately, 428 SNPs per chromosome were sampled randomly without replacement and used as QTL. As the result of this sampling, there were 7702 QTL in the genome. The number of QTL across the genome were reduced to a desired number, i.e., 1000 QTL or 180 QTL by random sampling from these 7702 QTL. QTL effects were sampled from an exponential distribution and the proportion of mutations that generate a positive effect at QTL loci was assumed 0.10. A total of 54218 SNPs over all chromosomes were sampled as genetic markers, i.e. approximately 3012 markers per chromosome from the remaining SNPs after the QTL were sampled.

IBD loci (Template markers)

The 334 (2 per cM) IBD loci were positioned evenly on each chromosome of the base animals and used to measure true inbreeding. Each base animal was provided unique alleles at each IBD locus. Hence, there are $2n_b$ unique alleles at each IBD locus among the n_b base animals. Thus, if an animal in generation, e.g., 10 carries two identical IBD marker alleles, this implies the animal is IBD at this position, since both alleles are a copy of the same unique base population allele. The IBD loci were not available for selection purposes but used only to monitor the identity by descent status in each IBD locus

Selection methods

Truncation selection

The 100 females and 10 males base population were randomly mated to give 500 offspring with equal sex ratio. These offspring were phenotyped and genotyped in the first generation. Next, truncation selection was carried out from generation 2 up to generation 5 using the genomic breeding value estimate and 100 dams and 10 sires were selected every generation. Each selected sire was mated once with a randomly sampled dam without replacement and each mating produced five offspring. Hence, these random matings produced 500 offspring with equal sex ratio.

Optimum contribution selection

EVA (Evolutionary Algorithms) [20] was used to optimize individual genetic contributions by maximizing a quadratic function, U_t ,with respect to **c**:

$$U_t(\mathbf{c}) = \boldsymbol{c}'\hat{\mathbf{g}} + \omega\mathbf{c}'\mathbf{G}\mathbf{c}$$

Where **c** is a n vector of genetic contributions of parents to the next generation, *n* is the number of animals in the population traced back from the current generation selection candidates to the base population, *t* is generations, $\hat{\mathbf{g}}$ is a n vector of genomic estimated breeding values,ω, is the penalty

applied to the average relationship of the current generation, **G** is a n x n genomic relationship matrix among all selection candidates. **G** is computed as described for prediction GEBVs in the GBLUP analysis section. With optimum contribution selection, 25 dams were selected at each generation based on GEBVs and the number of selected sires varied as this number was optimized by GOCS. The random allocation of a sire to the selected dam was proportional to the optimal contribution of the sires and each selected dam was randomly mated to one by OCS selected sire to produce 20 offspring with an equal sex ratio. The optimum contribution selection was carried out from generations t=6….15 and discrete generations were assumed.

Calculation of true breeding values and phenotypic values

All animals in each simulation were assigned phenotypes and breeding values prior to selection. The true breeding value (TBV) of an individual was calculated as:

$$TBV_i = \sum_{j=1}^{N} (x_{ij1} g_{j1} + x_{ij2} g_{j2})$$

Where x_{ijk} is the number of copies of the k^{th} allele that individual i has at the j^{th} QTL position, N is number of QTL, g_{jk} is the effect of the k^{th} allele at the j^{th} position and k=alleles 1 or 2. The QTL effects were standardised so that the total genetic variance was 1.0. The phenotypic values, P_i of individuals were simulated by:

$$P_i = TBV_i + e_i$$

Where e_i is an error term for individual i, which was sampled from the Normal distribution with mean zero and variance 4 resulting in a heritability of 0.2.

Data analyses

For each scheme, we plotted the short- and long-term rate of genetic gain against short- and long-term rate of IBD respectively at different penalties (ω). For each scheme, we searched for a penalty that gave a rate of short term IBD of 1% based on the 100 replicates and compared the

accompanying rate of short-term genetic gain. We also compared the accuracies of male and female selection.

In this study, short- and long-term refer to generations 6 - 10 and 11 - 15 respectively. Rates of genetic gain were calculated as the linear regression of G_t on t where G_t is the average true-breeding value of animals born in generation t. Rates of true inbreeding (IBD) (using the template markers) was calculated as 1-exp(β), where β is a linear regression of ln(1-F_t) on t and F_t is the average coefficient of true inbreeding for animals born at generation t (t = 6 … 10) for the short-term and (t = 11 … 15) for the long term [21]. F_t for true inbreeding was calculated using the d = 6012 IBD markers as $F_t=\frac{1}{n_t\, d}\sum_{i=1}^{n_t}\sum_{i=1}^{d}\delta ij$, where n_t is the number of animals born in generation t-and δ_{ij} is the IBD status at IBD-marker locus j (j = 1 … d) for animal i (i = 1 … n_t). δ_{ij} was equal to 1 if locus j for animal i was IBD for a unique (base) allele, and 0 otherwise.

Software

The simulations were run using the program, ADAM [17]. GEBV were estimated using DMU6 [22] or BayesP [5]. OCS was carried out by EVA[20].

Results

Rate of genetic gain with 180 QTL

With180 QTL, BayesP prediction method gave 5.7% higher rate of genetic gain than GBLUP at approximately 1% rate of IBD (Table 1). However, they had different long-term rates of IBD and the GBLUP had 21% higher long-term rate of IBD than BayesP (Table 1). The long-term rate of genetic gain was 9% higher in GBLUP than BayesP (Table 1).

Table 1: Rate of short- and long-term genetic gain and IBD in genomic optimum contribution selection using BayesP and GBLUP prediction methods (QTL=180).

Prediction	ω	ΔG_S (SE)	ΔG_L (SE)	ΔIBD_s (SE)	ΔIBD_L (SE)
BayesP	-15.5	0.369 (0.001)	0.243 (0.001)	0.010 (0.000)	0.012(0.000)
GBLUP	-15.0	0.349 (0.008)	0.265 (0.006)	0.011 (0.000)	0.014 (0.001)

Short-term rate of genetic gain (**ΔG_S**), Long-term rate of genetic gain (**ΔG_L**), Short-term rate of true inbreeding (**ΔIBD_S**), Long-term rate of true inbreeding (**ΔIBD_L**), standard error (SE) and weight applied on average relationship (ω).

Figure 1 shows the average IBD across replicates for the three chromosomes with the largest QTL. BayesP gives somewhat higher favorable QTL alleles frequencies change and markedly lower IBD of template markers around the QTL region than GBLUP (GBLUP has wider IBD peaks surrounding the QTL).

Accuracies of selection

The mean accuracies for selection candidates across 100 replicates were higher for BayesP than GBLUP (Figure 2). The genetic variance was similar across generations (Figure 3).

Rate of genetic gain with 1000 QTL

Table 2 shows at a short-term rate of IBD of 1%, the short-term rate of genetic gain of BayesP was 1.7% higher than GBLUP (Table 2). The weights applied on average relationship (ω) in GOCS in order to achieve approximately 1% short-term rate of IBD using BayesP and GBLUP were -16.2 and -15.8 respectively. Genetic variances at each generation for both prediction methods are presented in Figure 4. Both prediction methods had almost equal amounts of genetic variance at each generation (Figure 4).

Table 2: Rate of short- and long-term genetic gain and IBD in genomic optimum contribution selection using BayesP and GBLUP prediction methods (QTL=1000)

Prediction	ω	ΔG_S (SE)	ΔG_L (SE)	ΔIBD_s (SE)	ΔIBD_L (SE)
BayesP	-16.2	0.525 (0.001)	0.437 (0.001)	0.010 (0.000)	0.013(0.000)
GBLUP	-15.8	0.516 (0.008)	0.452 (0.007)	0.010 (0.000)	0.014 (0.001)

Accuracies of selection

The mean accuracies for selection candidates across 100 replicates were higher for BayesP than GBLUP in the short term. However, the difference in accuracies between the prediction methods became smaller over time (Figure 5). At generation 6 and 7 both prediction methods had equal genetic variance, however at these generations the accuracies of males and females selection candidates were higher in case of BayesP prediction compare to GBLUP (Figure 4 and 5)

Rate of genetic gain with 7702 QTL

At higher number of QTL (QTL=7702), prediction using BayesP gave 2.7 % higher rate of short-term genetic gain than the GBLUP prediction method at approximated 1% rate of short term IBD in genomic optimum contribution selection (Table 3 at ω =-16.5 &-15.5). In the long term, BayesP gave 4 % lower rate of genetic gain than GBLUP, however long- term rate of IBD in case of the GBLUP was 6% higher than BayesP.

Table 3: Rate of short- and long-term genetic gain and IBD in genomic optimum contribution selection using BayesP and GBLUP prediction methods (QTL=7702)

Prediction	ω	ΔG_S (SE)	ΔG_L (SE)	ΔIBD_s (SE)	ΔIBD_L (SE)
BayesP	-16.5	0.574 (0.001)	0.527 (0.001)	0.011 (0.000)	0.013 (0.000)
GBLUP	-15.5	0.559 (0.008)	0.549 (0.006)	0.011 (0.000)	0.014 (0.001)

Discussion

We investigated the effect of prediction methods, i.e. GBLUP and BayesP on rates of genetic gain at the same rate of true inbreeding in genomic optimum contribution selection. Our findings showed that using BayesP realized more genetic gain than GBLUP at the same rate of IBD (in the short-term) in genomic optimum contribution selection (Tables 1, 2 and 3). In the short-term, both prediction methods realized very similar rate of true inbreeding however, in the long-term, both prediction methods had somewhat different long-term rate of true inbreeding. Thus, the comparison between the prediction methods is fair only during the short-term rate of genetic gain. The higher genetic gain of the BayesP method is mainly due to the higher accuracy (Figures 2 and 4) since the relative differences in accuracy of selection are larger than those in genetic gain. This difference in accuracy of selection was especially large if there were few QTL as expected by Daetwyler, Pong-Wong [1] (Figures 2 and 4). The difference in rate of genetic gain between the prediction methods at the same rate of IBD was also larger at a low number of QTL (Table 1 and 3). However, the accuracy of selection becomes more similar in the long term in both prediction methods (Figures 2 and 4). This diminishing of accuracy differences between the prediction methods in the long term may be explained by the higher frequencies of QTL alleles with large effects, which implies that BayesP has less benefit of focusing on large QTL and accuracy differences between the methods become smaller. Hence, the better performance of BayesP to predict genomic breeding values than GBLUP also makes BayesP to perform better in genomic

optimum contribution selection than GBLUP and realizes more genetic gain at the same rate of IBD.

GBLUP yields higher IBD around the QTL (Figure 1), this is probably because GBLUP selects for a set of markers surrounding the QTL that improves the frequencies of haplotypes that carry the positive QTL allele. Large haplotypes have low initial frequencies, resulting in large selective sweeps, and IBD signals surrounding the QTL. On the other hand, BayesP tries to identify few SNPs in close LD with the QTL, which results in a smaller selective sweep and IBD signal (Figure 1). However, in the presence of a large number of QTL, these IBD signals overlap with each other and are less visible, i.e. the QTL peaks are too close to each other. It seems advantageous to have smaller selective sweeps, and the BayesP method benefits from this in the short term by changing QTL allele frequencies more, and in the long term its rate of IBD is lower than that of GBLUP.

Our finding, the relative better performance of BayesP than GBLUP in accuracies of males and females selection (Figures 2 and 4) are in agreement with studies of Yu and Meuwissen [5] and Iheshiulor, Woolliams [6]. Using a simulation study, Iheshiulor, Woolliams [6] reported that BayesP gave higher accuracies than GBLUP at a density of 45 QTL/Morgan as well as with whole-genome sequence data. However, they reported also similar accuracies for GBLUP and BayesP if the trait is lowly heritable and controlled by a large number of QTL. However in our finding, BayesP had higher accuracies than GBLUP at 7702 QTL as well as 180 QTL at the same rate of

IBD. Nevertheless, we could not compare the relative difference in accuracies between BayesP and GBLUP at large number of QTL (QTL=7702) and small number of QTL (QTL=180) since the accuracies of selection candidates depended on the weight applied on average relationship (ω).The latter is because a high ω implies the selection of many sires and thus small sire families, resulting in a lower selection accuracies.

In the case of BayesP prediction methods we assumed that the number of QTL that affect the trait is known and we set the π value to the ratio of the number of QTL relative to the number of SNPs. However, the number of QTL affecting a trait is generally not known. Nevertheless, optimal π values for the analysis could be found by cross validation [1]. It seems that BayesP improved genetic predictions compared to GBLUP and reduced the consequence of selection in the form of smaller selective sweeps.

Conclusions

BayesP gave higher accuracy than GBLUP and more genetic gain. Since the extra genetic gain seems to come from the extra accuracy of BayesP, and not from other improvements, it seems that the choice between GBLUP and Bayesian methods depends only on the accuracy in GOCS. BayesP also gave smaller selective sweeps around the QTL than GBLUP, causing extra genetic gain in the short term, and in the long term its rate of inbreeding was lower than that of GBLUP.

Competing interest

The authors declare that they have no competing interests.

Authors' contributions

GG performed the study and drafted the manuscript. GG, THEM and ACS designed the study. ACS incorporated the BayesP program in ADAM program. THEM planned and coordinated the whole study and contributed to writing the manuscript. All authors read and approved the final manuscript.

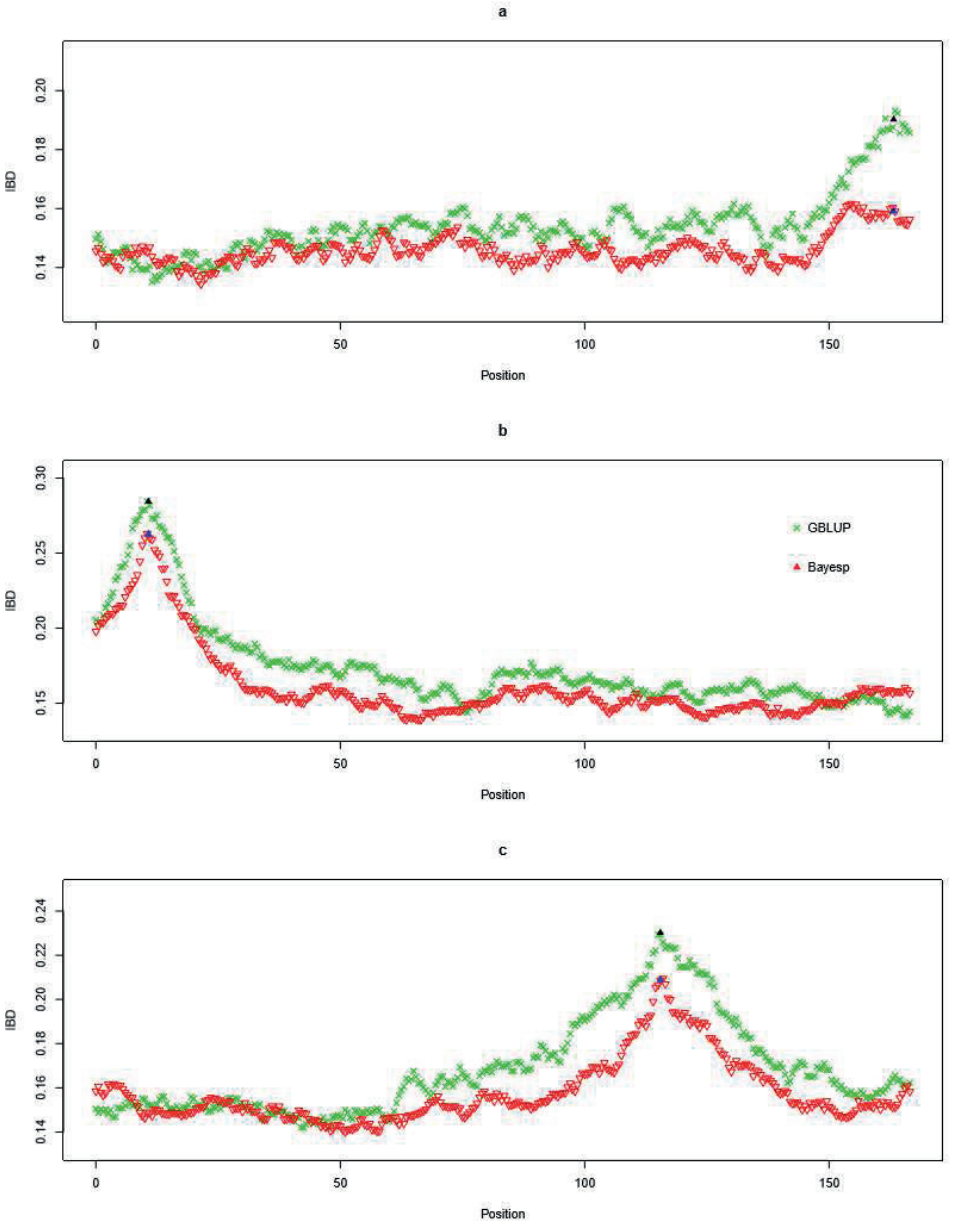
Acknowledgements

This study was funded by the Center for Genomic Selection in Animals and plants (GenSAP) supported by Innovation Fund Denmark. We also thank Mark Henryon for his help in ADAM program and allow us to use his haplotype structure.

References

1. Daetwyler, H.D., et al., *The impact of genetic architecture on genome-wide evaluation methods.* Genetics, 2010. **185**.
2. Meuwissen, T.H., B.J. Hayes, and M.E. Goddard, *Prediction of total genetic value using genome-wide dense marker maps.* Genetics, 2001. **157**(4): p. 1819-1829.
3. Hayes, B.J., et al., *Invited review: Genomic selection in dairy cattle: Progress and challenges.* Journal of dairy science, 2009. **92**(2): p. 433-443.
4. VanRaden, P., et al., *Invited review: Reliability of genomic predictions for North American Holstein bulls.* Journal of dairy science, 2009. **92**(1): p. 16-24.
5. Yu, X. and T.H. Meuwissen, *Using the Pareto principle in genome-wide breeding value estimation.* Genetics Selection Evolution, 2011. **43**(1): p. 35.
6. Iheshiulor, O.O., et al., *Within-and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels.* Genetics Selection Evolution, 2016. **48**(1): p. 15.
7. De Beukelaer, H., et al., *Moving Beyond Managing Realized Genomic Relationship in Long-Term Genomic Selection.* Genetics, 2017. **206**(2): p. 1127-1138.
8. Meuwissen, T., *Maximizing the response of selection with a predefined rate of inbreeding.* Journal of animal science, 1997. **75**(4): p. 934-940.
9. Woolliams, J., et al., *Genetic contributions and their optimization.* Journal of Animal Breeding and Genetics, 2015. **132**(2): p. 89-99.
10. Hayes, B.J., P.M. Visscher, and M.E. Goddard, *Increased accuracy of artificial selection by using the realized relationship matrix.* Genetics research, 2009. **91**(01): p. 47-60.
11. Wray, N.R. and R. Thompson, *Prediction of rates of inbreeding in selected populations.* Genetics Research, 1990. **55**(1): p. 41-54.
12. Henryon, M., P. Berg, and A.C. Sørensen, *Animal-breeding schemes using genomic information need breeding plans designed to maximise long-term genetic gains.* Livestock Science, 2014. **166**: p. 38-47.
13. Sonesson, A.K., J.A. Woolliams, and T.H. Meuwissen, *Genomic selection requires genomic control of inbreeding.* Genetics Selection Evolution, 2012. **44**(1): p. 27.
14. VanRaden, P., *Efficient methods to compute genomic predictions.* Journal of dairy science, 2008. **91**(11): p. 4414-4423.
15. Juran, J.M., *The non-Pareto principle; mea culpa*. 1975.
16. Meuwissen, T.H., et al., *A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value.* Genetics Selection Evolution, 2009. **41**(1): p. 2.
17. Pedersen, L., et al., *ADAM: A computer program to simulate selective breeding schemes for animals.* Livestock Science, 2009. **121**(2): p. 343-344.
18. Wright, S., *Evolution in Mendelian populations.* Genetics, 1931. **16**(2): p. 97-159.
19. Fisher, R., *The Genetical Theory of Natural Selection Clarendon Press Oxford Google Scholar.* 1930.
20. Berg, P., J. Nielsen, and M.K. Sørensen. *EVA: Realized and predicted optimal genetic contributions*. in *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006.* 2006. Instituto Prociência.
21. Hillestad, B., *Inbreeding determined by the amount of homozygous regions in the genome.* 2017.
22. Madsen, P., et al. *DMU-a package for analyzing multivariate mixed models*. in *8th World Congress on Genetics Applied to Livestock Production*. 2006.

Figure 1: Plot of IBD of template markers at genomic regions where the effect of favorable QTL alleles are greater than 0.191 at generation 10.



The black and blue triangular symbols in the peaks represent the IBD of the QTL in GBLUP and BayesP respectively. In Figure 1a, The effect of favorable QTL allele was 0.191 and its frequencies were 0.287 and 0.28 for BayesP and GBLUP respectively. In Figure 1b, The effect of favorable QTL allele was 0.54 and its frequencies were 0.332 and 0.333 in BayesP and GBLUP respectively. In Figure 1c, The effect of favorable QTL allele was 0.24 and its frequencies were 0.429 and 0.4 in BayesP and GBLUP respectively.

Figure 2: Accuracies of males and females selection candidates at each generation from generation 6 up to generation 15 (QTL=180).

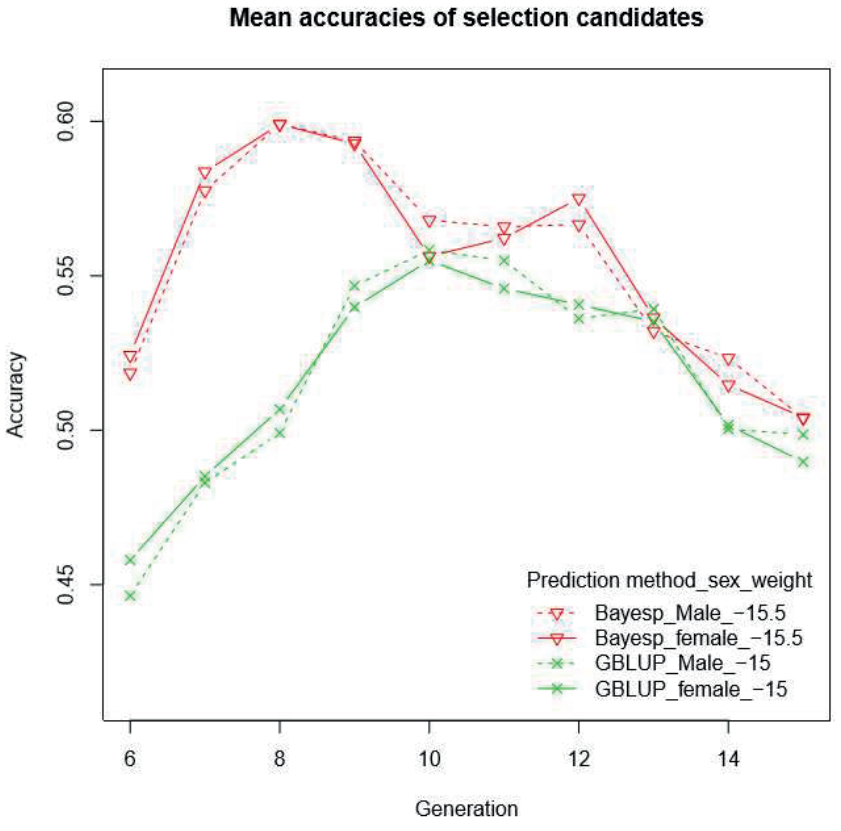


Figure 3: Genetic variance at each generation in BayesP at ω =-15.5 and GBLUP at ω =-15 in genomic optimum contribution selection (QTL=180).

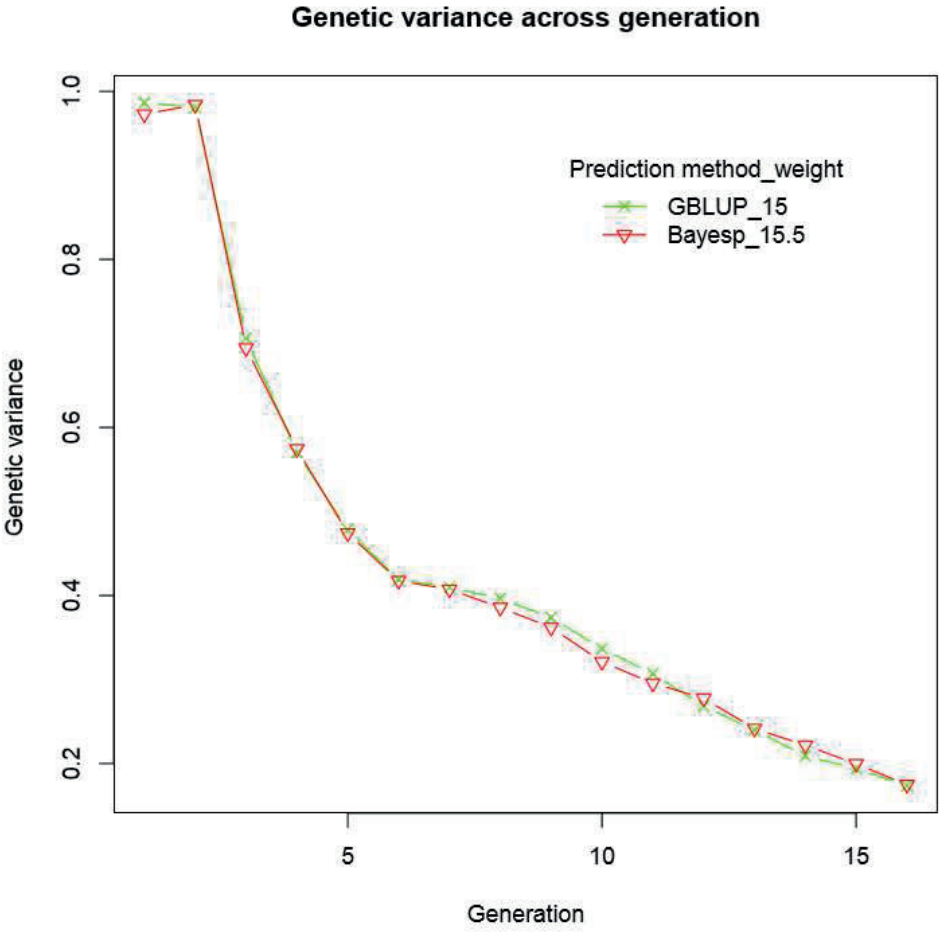


Figure 4: Genetic variance at each generation in BayesP at ω =-16.2 and GBLUP at ω =-15.8 in genomic optimum contribution selection (QTL=1000).

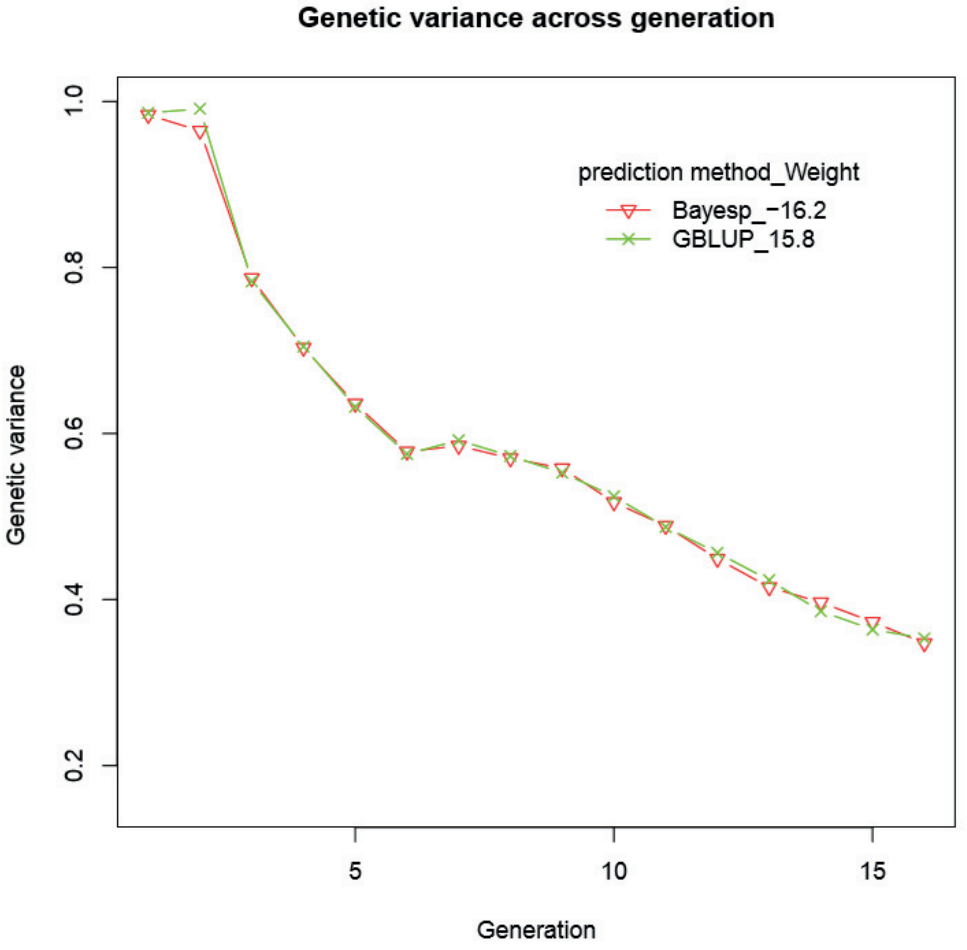
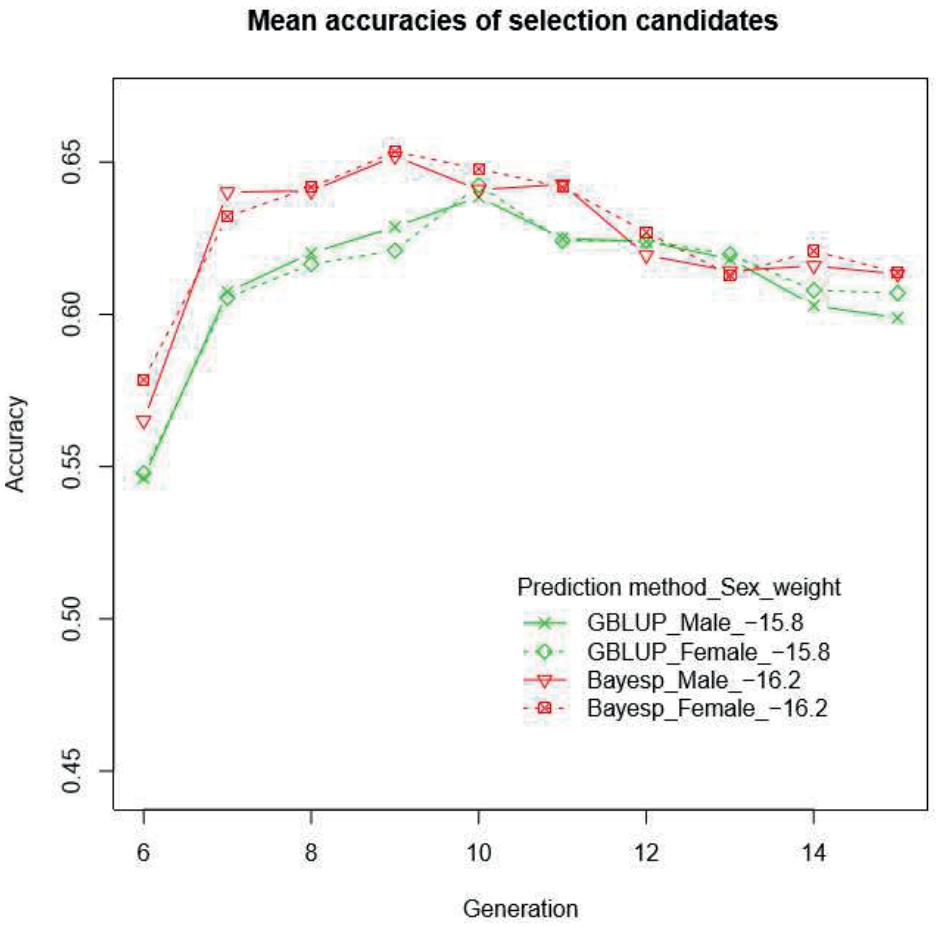


Figure 5: Accuracies of males and females selection candidates at each generation from generation 6 up to generation 15 (QTL=1000).



Paper 3

Estimation of inbreeding depression of semen quality traits using different measures of genomic inbreeding in a swine breeding population.

G.T.Gebregiwergis, Christian Maltecca, Francesco Tiezzi, Kent A.Gray, Vance D.Brown, Yijian Huang, Jeremy T.Howard, Solomon Antwi Boison and Theo HE Meuwissen

Submitted to Journal: Animal Breeding and Genetics

Estimation of inbreeding depression of semen quality traits using different measures of genomic inbreeding in a swine breeding population.

Authors: *G.T.Gebregiwergis[1], Christian Maltecca[2], Francesco Tiezzi[2], Kent A.Gray[3], Vance D.Brown[3], Yijian Huang[3], Jeremy T.Howard[3], Solomon Antwi Boison[4] and Theo HE Meuwissen[1]

[1] Department of Animal and Aquaculture Sciences, Norwegian University of life Sciences, Norway, [2]Department of animal science and genetics program, North Carolina state University, Raleigh, NC, USA, [3] Smithfield Premium Genetics, Rose Hill, NC, USA,[4] Nofima: The Norwegian Institute of Food, Fisheries and Aquaculture Research, Osloveien1, 1430 Ås, Norway

Abstract

The current availability of large numbers of single nucleotide polymorphisms (SNP) creates the opportunity to obtain more precise estimates of IBD of an individual than pedigree information. As a result, different methods have been proposed to quantify genomic inbreeding and inbreeding depression using SNP marker genotypes. However, there is no consensus on which genomic inbreeding measure to use for inbreeding depression studies. In this study, we compare alternative pedigree and genomic measures of inbreeding to estimate inbreeding depression in semen quality traits in a Large White pig population. Genomic inbreeding measures of 1701 genotyped individuals were obtained based on runs of homozygosity (F_{ROH}), average homozygosity (Fhomo), excess of homozygosity (FEx_homo), probability of uniting gametes (FU) and from the diagonal of genomic relationships matrix (FGRM). We obtained semen quality phenotypes (motility, progress motility, normal morphology, proximal droplets, and distal droplets) of 357 boars. Inbreeding depression was estimated by regressing the phenotype of the semen quality traits on alternative inbreeding coefficients. We did not detect a significant effect of inbreeding on all traits using Fped, FGRM, and Fu. However, inbreeding measured using F_{ROH}, FEx_homo, and Fhomo showed a significant effect on all semen quality traits ($P < 0.05$). An increase of 1% in ROH based genomic inbreeding (FROH>2Mb) causes a reduction of 2.8%, 2.5% and 2.7% of a phenotypic standard deviation of progressive motility, motility, and normal morphology respectively, and it causes proximal and distal droplets to increase by 3% and 2% respectively. The results highlight that inbreeding has significant effects on semen quality traits. Moreover, ROH based and excess of homozygosity measures of genomic inbreeding appear to capture inbreeding depression in semen quality data better. Hence, control of inbreeding strategies should be considered in the pig industry in order to get high-quality semen in sufficient quantity.

Keywords: genomic inbreeding, inbreeding depression, semen quality traits, sperm motility, progressive motility, runs of homozygosity, swine

Introduction

In a finite population, mating of animals with common ancestors is unavoidable and results in the accumulation of inbreeding over generations (Falconer, 1960). This accumulated inbreeding can cause homozygosity at loci with deleterious alleles to increase, and heterozygosity at loci displaying heterozygous advantage to decrease (Falconer, 1960). As a consequence, the mean of the population for a quantitative trait such as fertility, size, yield and fitness may be reduced. This reduction in performance is called inbreeding depression. It has been documented by several studies in different livestock species on a wide variety of traits (Silió et al., 2013; Leroy, 2014; Saura et al., 2015; Ferenčaković et al., 2017; Martikainen et al., 2017). Hence, measuring and managing inbreeding and quantifying inbreeding depression are active research topics.

The level of inbreeding in an individual is represented by its inbreeding coefficient and is defined as the probability that two alleles at any locus in that individual are identical by descent (IBD) (Wright, 1922; Falconer, 1960). This probability can be computed from pedigree information. However, the pedigree based inbreeding coefficient generally underestimates the true proportion of the genome that is IBD since the individual can be IBD due to more distant ancestors than those included in the pedigree (Kardos et al., 2016). Moreover, due to the effect of linkage, even in the presence of all common ancestors of parents in the pedigree, the true proportion of the genome that is IBD is not perfectly predicted from pedigree information (Kardos et al., 2015) . The current availability of a large number of single nucleotide polymorphisms (SNP) can yield more precise estimates of IBD of an individual than the pedigree alone (Kardos et al., 2015). In addition, unlike pedigree-based measures of inbreeding, genomic-based measures of inbreeding can provide an estimate of inbreeding for specific genomic regions (Curik et al., 2017), which opens new opportunities to study and quantify inbreeding and inbreeding depression.

Different methods have been proposed to quantify genomic inbreeding using SNP marker information. It can be calculated based on the percent homozygosity of all SNP markers (Kardos et al., 2015), based on excess of homozygosity as described by Keller et al. (2011) and based on the correlation between uniting gametes (Wright, 1922). In addition, genomic inbreeding coefficients are also obtained by subtracting one from the diagonal of genomic relationship matrices (VanRaden, 2007; 2008; Yang et al., 2010), although inbreeding coefficients obtained by this method are sensitive to allele frequencies in the base population. As the result of this, negative values might arise and alternative measures of genomic inbreeding can potentially be negatively correlated (Zhang et al., 2015). All the above measures of genomic inbreeding methods can be used to obtain more precise inbreeding coefficients than the pedigree method but they do not distinguish between identity by state (IBS) and IBD alleles (Howrigan et al., 2011; Kardos et al., 2015). Genomic inbreeding coefficients estimated using runs of homozygosity (ROH) instead implicitly correct for IBS probability by declaring only (arbitrarily) long runs of homozygosity as IBD (Keller et al., 2011; Kardos et al., 2015). In addition, the inbreeding coefficient obtained using ROH is more correlated with the homozygous mutation load than the above measures of genomic inbreeding (Keller et al., 2011). Finally, genomic inbreeding measured using runs of homozygosity can differentiate recent inbreeding from old inbreeding (Howrigan et al., 2011). As the result of this, ROH measures of genomic inbreeding received a lot of attention in the literature. However, in a simulation study considering a broad spectrum population parameters Kardos et al. (2015) found that the SNP-by-SNP-based estimators of inbreeding were as good as ROH to estimate true IBD and Bjelland et al. (2013) reported more inbreeding depression in dairy cattle using genomic inbreeding obtained from the diagonal of the genomic relationship matrix (F_{GRM}) than from

ROH. Similarly, Bérénos et al. (2016) reported that FGRM detected inbreeding depression more consistently across both body size and fitness traits. Hence, currently there is no full consensus on which genomic inbreeding measure to use for inbreeding depression studies.

The objective of this study was to compare alternative pedigree and genomic estimators of inbreeding to estimate inbreeding depression in semen quality traits in a Large White pig population. The estimator that shows the highest association to trait-depressions is expected to best reflect the inbreeding depression aspect of inbreeding.

Materials and methods

Phenotypic, pedigree and genotype information on a Large White nucleus population was obtained from Smithfield premium Genetics (Rose Hill, NC). All genotypic and phenotypic data came from other studies. Hence, no animal care approvals were required for the current study.

Pedigree and Phenotypic data

The pedigree data consisted of 6447 individuals (animals with phenotypes and their ancestors) with average depth ~6. The following phenotypic data for semen traits were available: sperm motility, which is the proportion of sperm cells actively moving in an ejaculate; sperm progressive motility, defined as the proportion of sperm cells moving in a straight line; the percentage of total normal morphology, which are sperm cells with no morphological abnormalities; distal droplet, defined as percentage of sperm cells with a swelling at farther down of the tail of the sperm; proximal droplet, defined as percentage of sperm cells with a swelling at the junction of the head and tail of the sperm. All traits were measured as percentages. Based on Q_Q plots of traits (results not presented) there were clear deviations from normality for the traits proximal droplets and distal droplets. Hence, the phenotypic values were log transformed for both traits. In addition, we checked each trait for outliers and descriptive statistics of the final data used for the analysis are presented in Table 1.

Genomic data

Four different SNP chips were used for genotyping: 179 animals were genotyped with the 10,241 SNP chip (Genomic Profiler 10k BeadChip; GeneSeek, Neogen Corp., Lincoln, NE) , 44 animals were genotyped using the 61,565 SNP chip (Infinium PorcineSNP60 v2 BeadChip; Illumina, Inc., San Diego, CA), 270 animals had information with the 62,163 SNP chip (Infinium PorcineSNP60 v1 BeadChip; Illumina, Inc.), and 1208 animals had information for the 68,528 SNP chip (PorcineSNP80 BeadChip; GeneSeek, Neogen Corp.). For all animals, genotypes were imputed to the 61,565 SNP chip using FImpute v2.2 (Sargolzaei et al., 2014). The genomic data was subject to Quality Control (QC) before the analysis. This QC removed SNP with minor allele frequency < 0.05 and call rates < 0.9. Only SNPs on autosomes were kept. After QC and imputation, 45,840 SNP were available for 1701 samples to this study. All measures of genomic inbreeding were calculated based on the 1701 genotyped Large White individuals.

Estimation of inbreeding coefficients

Inbreeding coefficients were estimated using five alternative measures of inbreeding as described below.

1. Pedigree-based inbreeding coefficients (Fped):- Fped was calculated using all pedigree information with the *pedigree* R package (Coster and Coster, 2010).
2. Genomic inbreeding based on excess of homozygosity (FEx_homo), which is a measure of the excess in the observed number of homozygous SNP markers within an individual relative to the expected number of homozygous SNP markers under random mating as described by Keller et al. (2011) and FEx_homo was calculated as:

$$\text{FEx\_homo}_i = \frac{O(\text{Hom}_i) - E(\text{Hom})}{m - E(\text{Hom})},$$

where $O(\mathrm{Hom}_i)$ is the observed number of homozygous loci for the *i*th individual, and $E(\mathrm{Hom})$ is expected average number of homozygous SNP markers across m loci assuming Hardy-Weinberg equilibrium. FEx_homo estimates were obtained using the –het function in PLINK. Genomic inbreeding of an individual were also calculated based on the proportion of observed number of homozygous loci (Fhomo).

3. ROH based genomic inbreeding (F_{ROH}):- ROH based genomic inbreeding coefficients for a given individual were defined as the proportion of its genome that is in ROH based on given ROH length cutoff (McQuillan et al., 2008). We used PLINK 1.9 to detect ROHs and used similar criteria as Purfield et al. (2012) to define a ROH. Default values in PLINK were used to establish the size of sliding windows (50 SNPs), the maximum gap length between two consecutive homozygous SNPs (1 Mb), the number of heterozygotes allowed in a window (1) and the minimum proportion of overlapping windows that must be homozygous (0.05). However, to avoid the effect of low SNPs density on ROH length, the minimum SNP density required to define a ROH was set to 1 SNP per 120 Kb. In addition, the minimum ROH length was set to 500 Kb in order to remove short ROHs which may occur by chance (the corresponding PLINK parameters are –homozyg-window-snp 50—homozyg-window-het 1 –homozyg-window-threshold 0.05 –homozyg-snp 5 –homozyg-kb 500 –homozyg-density 120 –homozyg-gap 1000). ROH based measures of inbreeding coefficients of an individual were calculated using the following formula: $\frac{\sum_k \mathrm{length}(\mathrm{ROH_k})}{L}$, where k=number of ROH discovered based on a given length cutoff for each animal, and L=total length of the genome. The length of ROH was measured in kilobases, with a L=2,449,138 kb (McQuillan et al., 2008).Three different measures of genomic inbreeding were estimated based on ROHs. The first measures of inbreeding were estimated based on ROHs length greater than 2 Mb (FROH> 2Mb).The length of ROH correlates to the age of inbreeding and a long ROH length is most likely due to recent inbreeding and/or it may consist of several adjacent ROHs (i.e. the entire segment is not truly autozygous; (Keller et al., 2011). To assess the relative importance of distant versus recent inbreeding, inbreeding coefficients of an individual were also estimated using short and long ROH. Inbreeding coefficient of an individual due to short ROH (FROH< 5Mb) was defined as the proportion of its genome that was in ROH of between 1 and 5 Mb, whereas, inbreeding coefficients based on long ROH (FROH>= 5Mb) were estimated using 5 Mb as minimum ROH length (Saura et al., 2015)
4. Genomic inbreeding coefficients from the genomic relationship matrix (FGRM). FGRM were calculated based on the variance of additive genetic values following VanRaden (2008). FGRM was calculated as it is presented by (Yang et al., 2011; Zhang et al., 2015) as follows:

$$\frac{[x_i - E(x_i)]^2}{h_i} - 1 = \frac{(x_i - 2\hat{q}_i)^2}{h_i} - 1,$$

Where x_i is 0, 1 and 2 for homozygote for reference allele, heterozygous for reference and homozygous for non-reference allele for the i^{th} SNP respectively, q_i is the observed fraction of the reference allele at locus *i*, $h_i = 2q_i(1-q_i)$.Finally, an individual inbreeding coefficient (FGRM) from all SNPs was obtained by averaging the estimates over all of the SNPs .

5. Genomic inbreeding coefficient based on correlation between uniting gametes (F_U) was estimated as follow (Wright, 1922):

$$\frac{x_i^2 - (1 + 2q_i)x_i + 2q_i^2}{h_i}$$

Where x_i ,q_i and h_i are the same as for F_{GRM} (Yang et al., 2011) and the individual inbreeding coefficient (F_U) obtained by averaging the estimates over all of the SNPs. The calculations for F_U and FGRM were computed using the option *–ibc* from GCTA (Genome-wide Complex Trait Analysis) software.

Pearson's correlation coefficients were calculated between alternative measures of inbreeding coefficients using the R (R Core Team, 2018).

Estimation of inbreeding depression

The alternative inbreeding coefficients were used to test their association with inbreeding depression. All measures of inbreeding were estimated with all the 1701 animals with genotypes, however for the estimate of inbreeding depression; only 357 animals out of the 1701 had phenotypes. Inbreeding depression was estimated by regressing the phenotypic values on the alternative inbreeding coefficients estimated using models of the form:

$$y_{ijkrylnm} = \mu + season_i + age_j + rest_r + year_y + stud_l + bF_{ijryln} + b1day_{ijryln} + p_n + a_n + e_{ijrylnm},$$

Where $y_{ijrylnm}$ is the sperm characteristic measured on m^{th} ejaculate of the n^{th} boar of the l^{th} *stud*, μ is the overall mean, $season_i$ is the effect of the season (Four levels) at collection, age_j is the effect of the age classes of the boar, $rest_r$ is the effect of the interval between the present and previous semen collection, $year_y$ is the effect of year class y (5 levels), $stud_l$ is the effect of the boar stud (10 levels), $b1$ is the regression coefficient of age in days and day_{ijryln} is the age of an individual in days, b is the regression coefficient on the inbreeding coefficient F_{ijlnm}, which was one of the alternative measures of inbreeding (Fped, FEx_homo, FROH>2Mb, FROH>=5Mb, FROH<5Mb, FGRM and F_U), p_n is the permanent environmental effect of the boar ,a_n is the additive genetic effect of the boar, and e_{ijlnm} is the residual effect. Monthly intervals were used for the effect of the age classes except for the traits motility and normal morphology. 5 classes (< 12 months, 13- 18 months, 19-24 months, 25- 28 months and > 28 months) were formed for the effect of age classes for the traits motility and normal morphology. For the effect of the interval between the present and previous semen collection, the classes were formed with an interval of 1 day for interval less than 12 days. For intervals greater than 12 days two classes were formed: 12-16 days, and >16 days. All fixed effects of the model were selected based on the available data and known models from the literature (Wolf and Smital, 2009; Ferenčaković et al., 2017; Marques et al., 2017). In addition, all the fixed effects were tested for their significances for each trait. As the result of these significance tests age_j was not incorporated in the model for the traits proximal droplet and distal droplet. The number of days between successive collections ($rest_r$) was incorporated in the model for the traits normal morphology and proximal droplets only. The effect of age in days as a covariate was significant for the trait motility only. All statistical analyses were performed with ASReml-4(Gilmour et al., 2015).

Results

Runs of homozygosity

A total of 146977 ROHs were detected in the 1701 genotyped animals. The number of ROHs with length less than or equal to 1Mb were only 0.1% out of 146977 total ROHs and the number of ROHs with length 1-2Mb were only 9.9%.The number of SNPs per ROH varied from 5 to 3174. Out of 146977 ROHs that were detected, only 7, 9 and 23 ROHs were with SNP number 5, 10 and 15 respectively. More than 99.98% of the ROHs were with SNP number greater than 15. The number of SNP varied from 21 to 3174 for the ROHs with length greater than 2Mb. In case of ROH> 5Mb, the number of SNPs varied from 47 to 3174.

Inbreeding coefficients

There were a total of 8 estimates of inbreeding coefficients for each animal and the average inbreeding coefficients estimated across all individuals using different approaches are presented in Table 2. The average inbreeding of all genotyped animals ranged from -0.007 to 0.664. Inbreeding coefficients based on excess of homozygosity (FEx-homo) and average homozygosity (Fhomo) gave the least and the highest average inbreeding respectively. The average inbreeding using the pedigree data is less than the ROH based average inbreeding. The average inbreeding coefficients using the ROH based inbreeding coefficients varied from 0.062 to 0.248 depending on the length of ROHs. However, the pedigree based inbreeding coefficient (Fped), the genomic relationship matrix based inbreeding coefficients (FGRM), and inbreeding coefficients based on correlation between uniting gametes (Fu) resulted in very similar average inbreeding levels.

The correlation between the alternatives measures of inbreeding is presented in Figure 1. Among the genomic measures of inbreeding, FROH>2Mb, FROH>=5Mb, Fhomo and FEx-homo had high and positive Pearson correlations with each other. However, they had low and positive correlation with the pedigree based inbreeding coefficients (Fped). Fped was almost uncorrelated with FGRM, Fu and FROH<5Mb. Measures of inbreeding based on (excess of) homozygosity (Fhomo and FEx-homo) had low correlation with both FGRM and Fu, whilst they had a correlation of 1 between themselves. However, FGRM and Fu were highly correlated (0.96) to each other.

Estimates of Inbreeding depression

The inbreeding depression as a proportion of the phenotypic standard deviation and the estimates of inbreeding depression for all traits are presented in **Table 3 and supplementary 1**. Statistically significant inbreeding depression estimates were observed for the traits motility, progressive motility, normal morphology, distal droplets and proximal droplets ($P<0.05$). However, the inbreeding depression estimates differed between the different measures of inbreeding. An increase of 1% in FROH>2Mb, FROH>=5Mb ,FEx_homo and Fhomo reduced motility by approximately 2.5%, 2.4%, 2.2% and 5.6% of the phenotypic standard deviation of the trait respectively. Similarly, a 1% increase in FROH>2Mb, FROH>=5Mb, FEx_homo and Fhomo reduced the progressive motility by 2.8%, 2.8%, 2.2% and 5.6% of a phenotypic standard deviation of the trait, respectively. A significant reduction of normal sperm morphology was also observed using FEx_homo, Fhomo and ROH based measures of genomic inbreeding coefficients. In addition, distal and proximal droplets also showed a significant increment for inbreeding coefficients measured using FROH>2Mb, FROH>=5Mb, FEx_homo and Fhomo. However, none of the traits showed significant inbreeding depression estimates using FGRM, Fu and Fped.

Discussion

In this study, we observed that inbreeding significantly reduces boar semen quality (e.g. motility, progressive motility, normal morphology, distal droplets and proximal droplets) using (excess of) homozygosity and ROH based measures of genomic inbreeding (**Table 3 and supplementary 1**). The impact of inbreeding on semen quality traits has also been reported using genomic measures of inbreeding in cattle (Ferenčaković et al., 2017). However, there is lack of information on the impact of inbreeding on semen quality traits using genomic measures of inbreeding in pigs. Although there are several studies that have used pedigree based measures of inbreeding to study inbreeding depression on semen quality traits in pigs (Van Eldik et al., 2006; Zajitschek et al., 2009; Ruiz-Lopez et al., 2010; Maximini et al., 2011). In our study, we did not detect a significant association with pedigree inbreeding coefficients (Table 3). This could be due to the presence of quite low inbreeding levels (Table 2) and low variability of F_{ped} which may lead to low statistical power for significance testing. Dorado et al. (2017) and Maximini et al. (2011) found that low levels of inbreeding did not affect sperm motility in bulls. However, when inbreeding levels were high, there was a significant reduction of sperm motility. Using ROH based measures of genomic inbreeding Terán et al. (2018) and Azcona et al. (2019) reported the presence of a potential influence of inbreeding on sperm morphometry and motility in bulls. Overall, our findings showed that sperm quality traits are affected by inbreeding depression using the genomic measures of inbreeding coefficients. Therefore, control of genomic inbreeding strategies should be considered in the livestock industry in order to get high-quality semen in sufficient quantity, in addition to that this may also alleviate inbreeding depression on traits that are not investigated here.

In our study we detected a significant effect of inbreeding on normal morphology of sperm cells using all ROH based, FEx_homo and Fhomo measures of inbreeding coefficients. A similar finding has been reported by Van Eldik et al. (2006) which showed significant reductions of normal spermatozoa morphology as inbreeding coefficients increased. However, Dorado et al. (2017) did not report a significant effect of inbreeding on sperm morphology in beef bulls, which may be due to the size of their study.

The alternative measures of inbreeding coefficients considered in this study resulted in different estimates of inbreeding coefficients (Table 2). As pedigree provides an incomplete picture of inheritance patterns, lower inbreeding levels were expected with pedigree inbreeding compared to ROH and Fhomo measures. In our study we obtained higher average inbreeding coefficients using Fhomo compared to ROH based methods (Table 2). Fhomo, unlike ROH tracks both IBS and IBD alleles and inbreeding coefficients obtained using Fhomo could be overestimated, partially explaining the differences in magnitude between the two estimates. It is worth noting that average inbreeding coefficients using F_{ROH} could also be overestimated as they may arise from false positive short ROHs (Ferenčaković et al., 2013). In our study 58% (85,726 of 146,977) of all ROHs were shorter than 5 Mb. Ferenčaković et al. (2013) suggested that a 50K panel could produce false positive findings of short ROHs thus leading to an overestimation of inbreeding coefficients obtained by this measure, however, they did not find differences between panels for the number of ROHs longer than 4 Mb. On the other hand, measures of inbreeding coefficients based on ROH greater than 5 Mb could also underestimate the average inbreeding coefficients by excluding shorter IBD regions.

The correlations between inbreeding coefficients estimates of alternative measures of inbreeding coefficients varied from very high (0.97) to close to zero. Among the genomic measures of

inbreeding coefficients, FGRM showed very low negative correlations with other estimators except with F_U. Similar findings have been reported using 50K SNP chip genotypes and whole genome sequence data by (Zhang et al., 2015). The high correlations between inbreeding coefficients measured using ROH based with homozygosity measures (FEx-homo or Fhomo) (figure 1) are also in agreement with previous findings (Saura et al., 2015; Zhang et al., 2015; Zanella et al., 2016; Martikainen et al., 2017). Similar to our finding using 50K SNP chip data, Zhang et al. (2015) reported positive and high correlations between FGRM and Fu. However, they reported negative correlation between Fu and FEx-homo. Using one locus as an example, Zhang et al. (2015) further showed that inbreeding coefficients estimated using FGRM, Fu and FEx-homo depend strongly on the estimates of allele frequencies in the base population. This dependency on allele frequency estimates might yield less accurate inbreeding coefficients depending on the distribution of alleles frequencies in the population (Zhang et al., 2015). Given this premise, inbreeding coefficients obtained using genomic relationship matrix (FGRM) and the correlation between uniting gametes (Fu) in our study could be less accurate than the other genomic measures of inbreeding coefficients and partially explain their low correlation with other measures of inbreeding presented. However, the presence of a high correlation between ROH based measures of inbreeding coefficients with Fhomo and FEx-homo are probably because they are similar (F_{homo} represents basically a ROH of 1 SNP). Hence, measures of inbreeding coefficients using (excess of) homozygosity gave more accurate inbreeding coefficients than FGRM or F_U.

Fhomo and FEx-homo showed a correlation of 1 in Figure 1, which was expected since the calculation of FEx-homo involves only a rescaling of the actual homozygosity (Fhomo). However, the regression coefficients on inbreeding are very different for Fhomo and FEx-homo, where the regression coefficient of FEx-homo is much more in line with the other regression coefficients (Table 3). This is probably because Fhomo is not corrected for IBS probabilities, and thus expresses inbreeding at a totally different level (average Fhomo = 0.66) than the other inbreeding measures (Table 2), which also reduces its standard deviation (SD (Fhomo)=0.018 vs. SD(FEx-homo)=0.055), since its values are much closer to the maximum of probability of identity of 1. Thus, although measures of inbreeding that are corrected for IBS or not may be highly correlated, their estimates of inbreeding depression (regression coefficients on F) should be interpreted very differently since they are expressed on a different scale. The measure that is corrected for IBS seems to be preferred since it is more in line with other measures of inbreeding, including pedigree based F, and less sensitive to the choice of the SNP panel (which affects the average homozygosity).

In the literature, there is no consensus on which genomic estimator of inbreeding is best for studying inbreeding depression. Keller et al. (2011) showed that among the alternative measures of inbreeding coefficients (F_{ROH}, Fped, FEx-homo and FGRM), F_{ROH} was most associated with the recessive mutation load. Moreover, F_{ROH} showed higher power in regression analyses compared to Fped because of its higher variance (Keller et al., 2011).In spite of this, there are reports which show that other genomic measures of inbreeding are as good as ROH based measures to detect inbreeding depression (Bjelland et al., 2013; Bérénos et al., 2016). In our study, FROH>2Mb and FROH>= 5Mb outperformed Fped, FEx-homo, FGRM and Fu for the detection of inbreeding depression in semen quality traits (i.e. resulted in more significant P values). Hence, we recommend FROH>= 5Mb for detecting inbreeding depression, and expect that it best describes the loss of heterozygosity that causes inbreeding depression. Possibly this is because FROH>= 5Mb describes rather recent inbreeding (~2-10 generation ago), and it is known that inbreeding

depression is more severe for recent inbreeding than ancient inbreeding due to the purging of deleterious alleles in the longer term (Fisher, 1954; Hedrick and Garcia-Dorado, 2016).

Conclusions

In our study ROH based and excess of homozygosity measures of genomic inbreeding showed stronger associations with inbreeding depression in semen quality data. Motility, progressive motility, normal morphology and proximal and distal droplets were all significantly affected by inbreeding accumulation. An increase of 1% in ROH based genomic inbreeding (FROH>2Mb) causes a reduction of 2.8%, 2.5% and 2.7% of phenotypic standard deviation of progressive motility, motility, and normal morphology respectively, but it causes proximal and distal droplets to increase by 3% and 2% respectively. Hence, strategies to control inbreeding should be considered by the pig industry in order to get high-quality semen in sufficient quantity in addition to avoiding inbreeding depression effects on other traits of interest.

Acknowledgments

This study was funded by the Center for Genomic Selection in Animals and plants (GenSAP) supported by Innovation Fund Denmark and Norwegian University of life Sciences research department covered travel and accommodation cost of GTG. We are also very grateful to SPG and NCST for providing genotypes, pedigree and semen quality data used in this study.

Conflict of interest statement

The authors declare that they have no competing interests.

References

Azcona, F., Sole, M., Dorado, J., Ross, P., Terán, E., and Demyda-Peyrás, S. (2019). 155 Whole genome association analysis suggests an influence of inbreeding on bull sperm morphometry. *Reproduction, Fertility and Development* 31(1)**,** 202-203.

Bérénos, C., Ellis, P.A., Pilkington, J.G., and Pemberton, J.M. (2016). Genomic analysis reveals depression due to both individual and maternal inbreeding in a free-living mammal population. *Molecular ecology* 25(13)**,** 3152-3168.

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

Coster, A., and Coster, M.A. (2010). Package 'pedigree'. *R package version* 1.

Curik, I., Ferenčaković, M., and Sölkner, J. (2017). Genomic dissection of inbreeding depression: a gate to new opportunities. *Revista Brasileira de Zootecnia* 46(9)**,** 773-782.

Dorado, J., Cid, R.M., Molina, A., Hidalgo, M., Ariza, J., Moreno-Millán, M., et al. (2017). Effect of inbreeding depression on bull sperm quality and field fertility. *Reproduction, Fertility and Development* 29(4)**,** 712-720.

Falconer, D.S. (1960). *Introduction to quantitative genetics.* Oliver And Boyd; Edinburgh; London.

Ferenčaković, M., Sölkner, J., and Curik, I. (2013). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution* 45(1)**,** 42.

Ferenčaković, M., Sölkner, J., Kapš, M., and Curik, I. (2017). Genome-wide mapping and estimation of inbreeding depression of semen quality traits in a cattle population. *Journal of dairy science* 100(6)**,** 4721-4730.

Fisher, R.A. (1954). A fuller theory of "junctions" in inbreeding. *Heredity* 8(2)**,** 187.

Gilmour, A., Gogel, B., Cullis, B., and Thompson, R. (2015). "ASReml user guide release 4.0. Hemel Hempstead: VSN International Ltd; 2014".).

Hedrick, P.W., and Garcia-Dorado, A. (2016). Understanding inbreeding depression, purging, and genetic rescue. *Trends in ecology & evolution* 31(12)**,** 940-952.

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

Kardos, M., Luikart, G., and Allendorf, F. (2015). Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity* 115(1)**,** 63.

Kardos, M., Taylor, H.R., Ellegren, H., Luikart, G., and Allendorf, F.W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary applications* 9(10)**,** 1205-1218.

Keller, M.C., Visscher, P.M., and Goddard, M.E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189(1)**,** 237-249.

Leroy, G. (2014). Inbreeding depression in livestock species: review and meta-analysis. *Animal genetics* 45(5)**,** 618-628.

Marques, D., Lopes, M.S., Broekhuijse, M., Guimarães, S., Knol, E., Bastiaansen, J., et al. (2017). Genetic parameters for semen quality and quantity traits in five pig lines. *Journal of animal science* 95(10)**,** 4251-4259.

Martikainen, K., Tyrisevä, A.-M., Matilainen, K., Pösö, J., and Uimari, P. (2017). Estimation of inbreeding depression on female fertility in the Finnish Ayrshire population. *Journal of Animal Breeding and Genetics* 134(5)**,** 383-392.

Maximini, L., Fuerst-Waltl, B., Gredler, B., and Baumung, R. (2011). Inbreeding depression on semen quality in Austrian dual-purpose simmental bulls. *Reproduction in domestic animals* 46(1), e102-e104.

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., et al. (2008). Runs of homozygosity in European populations. *The American Journal of Human Genetics* 83(3), 359-372.

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

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. (Vienna, Austria).

Ruiz-Lopez, M.J., Evenson, D.P., Espeso, G., Gomendio, M., and Roldan, E.R. (2010). High levels of DNA fragmentation in spermatozoa are associated with inbreeding and poor sperm quality in endangered ungulates. *Biology of reproduction* 83(3), 332-338.

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

Saura, M., Fernández, A., Varona, L., Fernández, A.I., De Cara, M.Á.R., Barragán, C., et al. (2015). Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. *Genetics Selection Evolution* 47(1), 1.

Silió, L., Rodríguez, M., Fernández, A., Barragán, C., Benítez, R., Óvilo, C., et al. (2013). Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *Journal of Animal Breeding and Genetics* 130(5), 349-360.

Terán, E., Goszczynski, D., Molina, A., Ross, P., Dorado, J., Giovambattista, G., et al. (2018). 150 Runs of Homozygosity Analysis and their Possible Influence on Sperm Motility in Highly Consanguineous Bulls. *Reproduction, Fertility and Development* 30(1), 215-215.

Van Eldik, P., Van Der Waaij, E., Ducro, B., Kooper, A., Stout, T., and Colenbrander, B. (2006). Possible negative effects of inbreeding on semen quality in Shetland pony stallions. *Theriogenology* 65(6), 1159-1170.

VanRaden, P. (2007). Genomic measures of relationship and inbreeding. *INTERBULL bulletin* (37), 33.

VanRaden, P. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science* 91(11), 4414-4423.

Wolf, J., and Smital, J. (2009). Quantification of factors affecting semen traits in artificial insemination boars from animal model analyses. *Journal of animal science* 87(5), 1620-1627.

Wright, S. (1922). Coefficients of inbreeding and relationship. *The American Naturalist* 56(645), 330-338.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature genetics* 42(7), 565.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* 88(1), 76-82.

Zajitschek, S., Lindholm, A., Evans, J., and Brooks, R. (2009). Experimental evidence that high levels of inbreeding depress sperm competitiveness. *Journal of evolutionary biology* 22(6), 1338-1345.

Zanella, R., Peixoto, J.O., Cardoso, F.F., Cardoso, L.L., Biegelmeyer, P., Cantão, M.E., et al. (2016). Genetic diversity analysis of two commercial breeds of pigs using genomic and pedigree data. *Genetics Selection Evolution* 48(1), 24.

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

Table 1: Descriptive statistics of semen quality traits. Progressive motility (PM), Motility (MO), distal droplet (DD), proximal droplet (PD) and Normal morphology (NM)

Trait	Number of observations	Number of animal	mean	SD	Min	max
Motility	1148	357	56.440	22.860	0.300	99.300
Progressive Motility	1142	355	37.580	20.930	0.200	85.300
Normal morphology	1088	342	83.900	8.510	32.500	96.700
Distal droplet (Log transformed)	1142	356	1.903	0.428	0.588	3.408
Proximal droplet (Log transformed)	1088	342	1.290	0.634	-0.916	3.934

Table 2: Descriptive statistics of inbreeding coefficients estimated using alternative measures of inbreeding.

Measure of inbreeding	Mean	SD	Min	max
FROH>2Mb	0.248	0.042	0.047	0.359
FROH>=5Mb	0.190	0.040	0.002	0.344
FROH<5Mb	0.062	0.010	0.014	0.094
FEx_homo	-0.007	0.055	-0.287	0.166
Fhomo	0.664	0.018	0.570	0.722
FGRM	-0.081	0.059	-0.232	0.287
Fu	-0.012	0.029	-0.103	0.087
F_{ped}	0.009	0.011	0.000	0.129

F_{ped}, FGRM, Fu, FEx_homo and Fhomo are inbreeding coefficients based on pedigree, genomic relationships matrix, correlation between uniting gametes, excess of homozygosity and homozygosity respectively.FROH>2Mb, FROH>=5Mb and FROH<5Mb are inbreeding coefficients based on ROH length cutoff greater than 2Mb, greater or equal to 5Mb and maximum of 5 Mb and minimum of 1 Mb respectively.

Table 3: Proportion of phenotypic standard deviation of semen quality traits [Progressive motility (PM), Motility (MO), distal droplet (DD), proximal droplet (PD) and Normal morphology (NM)] per 1% increase in alternative measures of inbreeding.

Measure of Inbreeding	PM		MO		DD		PD		NM
	Proportion[a]	LogL[b]	Proportion[a]	LogL	Proportion[a]	LogL	Proportion[a]	LogL	Proportion[a]	LogL
FROH>2Mb	**-0.028******	**-3652**	**-0.025*****	-3881	0.020 **	619	0.030****	128	-0.027**	-2480
FROH>=5Mb	**-0.028******	-3652	-0.024**	-3881	0.017	618	0.032****	128	-0.024**	-2480
FROH<5Mb	-0.033	-3655	-0.052	-3882	0.075	620	0.004	125	-0.096**	-2479
FEx_homo	**-0.022*****	-3655	-0.017**	-3882	0.015**	618	0.017**	126	-0.021**	-2480
Fhomo	**-0.056*****	-3652	-0.050**	-3881	0.044**	620	0.051**	127	-0.063**	-2479
FGRM	0.007	-3652	-0.015	-3882	-0.008	613	-0.002	140	0.003	-2483
F_U	0.004	-3657	0.001	-3884	0.002	617	0.023	126	-0.017	-2482
F_{ped}	0.003	-3655	-0.062	-3882	-0.072	620	-0.030	126	0.032	-2481

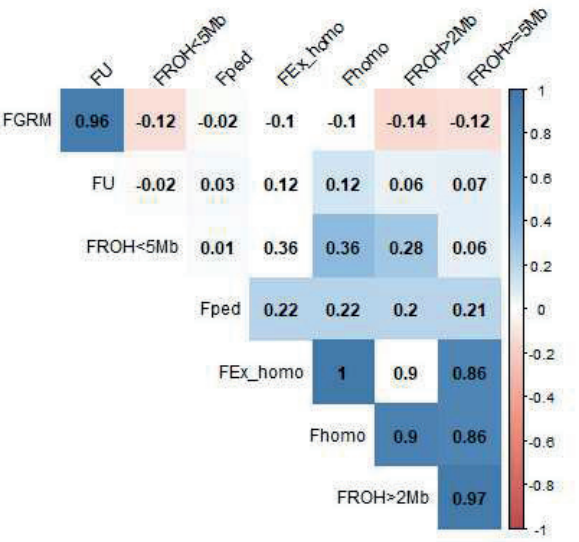
**0.01 <p-value $\leq$ 0.05, ***0.005 < p-value $\leq$ 0.01, ****p-value <0.005

[a]Estimate of inbreeding depression divided by phenotypic standard deviation

[b]LogL= Log likelihood value of the model

F_{ped}, FGRM, Fu, FEx_homo and Fhomo are inbreeding coefficients based on pedigree, genomic relationships matrix, correlation between uniting gametes, excess of homozygosity and homozygosity respectively. FROH>2Mb, FROH>=5Mb and FROH<5Mb are inbreeding coefficients based on ROH length cutoff greater than 2Mb, greater or equal to 5Mb and maximum of 5 Mb and minimum of 1 Mb respectively.

Figure 1: The correlations of the different measure of inbreeding.



F_{ped}, FGRM, Fu, FEx_homo and Fhomo are inbreeding coefficients based on pedigree, genomic relationships matrix, correlation between uniting gametes, excess of homozygosity and homozygosity respectively. FROH>2Mb, FROH>=5Mb and FROH<5Mb are inbreeding coefficients based on ROH length cutoff greater than 2Mb, greater or equal to 5Mb and maximum of 5 Mb and minimum of 1 Mb respectively.

Table supplementary 1: Effect of 1% increase in alternative measures of inbreeding on semen quality traits and standard errors of the estimates in brackets. Progressive motility (PM), Motility (MO), distal droplet (DD), proximal droplet (PD) and Normal morphology (NM).

Measure of Inbreeding	PM	MO	DD	PD	NM
FROH>2Mb	**-0.591** (0.188) ****	**-0.582** (0.207)***	0.008 (0.004) **	0.019 (0.007)***	-0.226 (0.091)**
FROH>=5Mb	**-0.585** (0.195) ****	-0.537 (0.214)**	0.007 (0.004)	0.020 (0.007) ****	-0.201 (0.095)**
FROH<5Mb	-0.694 (0.802)	-1.200 (0.879)	0.032 (0.018)	0.003 (0.029)	-0.815 (0.379)**
FEx_homo	**-0.460** (0.173) ***	-0.382 (0.161) **	0.007 (0.003 [a])**	0.011 (0.005) **	-0.180 (0.070)**
Fhomo	**-1.169** (0.339)***	-1.145 (0.482) **	0.019 (0.010)**	0.032 (0.016)**	-0.540 (0.210)**
FGRM	0.139 (0.039)	-0.342 (0.143)	-0.003 (0.003)	-0.001 (0.001)	0.027 (0.019)
Fu	0.088 (0.266)	0.024 (0.291)	0.001 (0.006)	0.014 (0.010)	-0.143 (0.128)
F_{ped}	0.069 (0.922)	-1.417 (1.006)	-0.031 (0.020)	-0.019 (0.033)	0.274 (0.430)

*0.05<p-value ≤ 0.1, **0.01 <p-value ≤ 0.05, ***0.005 < p-value ≤ 0.01, ****p-value <0.005

F_{ped}, F_{GRM}, F_U, F_{Ex_homo} and F_{homo} are inbreeding coefficients based on pedigree, genomic relationships matrix, correlation between uniting gametes, excess of homozygosity and homozygosity respectively. FROH>2Mb, FROH>=5Mb and FROH<5Mb are inbreeding coefficients based on ROH length cutoff greater than 2Mb, greater or equal to 5Mb and maximum of 5 Mb and minimum of 1 Mb respectively.

ISBN: 978-82-575-1587-4
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no

104326 / ANDVORDGRAFISK.NO