Content

Preface

This thesis is the result of three years work on my Industrial-PhD project entitled "Genomic selection in small dairy cattle populations". The work was funded by a Ph.D. scholarship from VikingGenetics and the Ministry of Science, Technology and Innovation. The work was carried out at the Center for Quantitative Genetics and Genomics (QGG), Department of Molecular Biology and Genetics, Aarhus University and at VikingGenetics, Assentoft. The analysis for paper III was done in collaboration with Christa Egger-Danner, ZUCHTDATA, Austria and Alfons Willam, BUKU University, Vienna, during three short stays in Vienna.

Personally the last three years has been a challenging and exciting travel into the theoretical aspects of dairy cattle breeding. A special thanks to VikingGenetics and Søren Borchersen for encouraging me to spend 3 years on this project. It is my hope that the outcome of this project will contribute to a more efficient use of genomic selection.

Many people have contributed in various ways, thanks to:

My advisory group: Bernt Guldbrandtsen, Anders Christian Sørensen, Mogens S. Lund and Søren Borchersen, for inspiration and recommendations, and for asking the challenging academic questions.

Christa Egger-Danner and Alfons Willam for introducing me to ZPLAN and Vienna.

Colleques in VikingGenetics and QGG, Foulum, for creating a good working environment.

Christian (and Lasse) for taking time to discuss ADAM results in the final stage, during his paternity leave. Rasmus for assistance in "breaking the R-code", Morten and Louise for helpful corrections to the manuscript.

Family and friends for support and interest, during the project - properly still wondering how to spend three years producing four papers.

Finally, thanks to the Danish Jersey breed for providing genetics and recordings. Without you, this project had never been the same.

Jørn Rind Thomasen

March 2013

Summary

Genomic selection increases reliability of estimates of genetic merit for young breeding candidates without own performance and thereby improving genetic gain. For that reason, genomic selection has become an integral part of the basis of breeding decisions in many dairy cattle breeding schemes. However, the method has turned out to be more efficient in the numerically bigger Holstein breed compared to smaller breeds.

The overall focus of this thesis was to examine how the use of genomic selection can be optimized in terms of expected monetary genetic gain, discounted profit and rate of inbreeding, as well as the uncertainty about the actual outcomes (risk) for a small dairy cattle population. Danish Jersey was used as a model breed to exemplify a small population, but the findings are expected to be relevant for other small dairy cattle populations as well.

In **paper I**, the reliabilities of genomic predictions were evaluated in Danish Jersey for all traits included in its breeding goal. A Bayesian method was used to estimate the SNP marker effects. The reliabilities of genomic predictions were on average across all traits 0.04 higher than the reliabilities of the pedigree indices. Estimates of reliabilities depended on the validation method applied. They also varied between traits without a clear relationship to the heritability of the trait.

Danish Jersey serves as an example of an admixed breed including animals with varied breed proportions of original Danish and US Jersey. In **paper II**, it was evaluated whether the population structure known from the history of Danish Jersey is reflected in the genomic structure currently observed in the breed. Firstly, it was found that the linkage disequilibrium in the group of admixed Danish Jersey animals was lower compared to the groups of primarily either Danish or US Jersey origin. Secondly, it was found that genomic breed proportions were in agreement with the pedigree-based breed proportions. However, explicitly including genomic breed proportions in a random regression prediction model for the trait udder health did not clearly improve the reliabilities of the genomic predictions compared to a basic genomic model. This shows that current models already properly account for heterogeneous breed origin, at least as measured at the whole genome level.

In **paper III**, the optimal breeding scheme for Danish Jersey was studied using a deterministic approach. The optimal breeding scheme was characterized by a mixed use of genotyped young bulls and the older progeny tested bulls. Strong interaction effects were observed between increased reliabilities of genomic predictions and more intensive

use of young bulls. A turbo scheme using only young bulls was genetically superior if higher reliabilities of genomic predictions could be obtained. Using discounted profit as the evaluation criterion, the turbo scheme was always superior due to lower generation interval and reduction in costs of housing and feeding waiting bulls. The results from paper III demonstrated that low reliabilities of genomic predictions limit the possibilities to move towards more efficient breeding schemes with more intensive use of genotyped young bulls without a progeny test.

One way to increase reliabilities of genomic prediction is to include genotyped cows in the reference population. This aspect was studied in **paper IV,** with use of stochastic simulations. Reliabilities of genomic predictions increased remarkably when genotyped cows were added to the reference population. In addition, the highest increase in genetic gain was obtained in a turbo scheme when cows were added to the reference population. The rate of inbreeding was also lower in these schemes. The risk measured as the variance of response was highest in the turbo schemes compared to the hybrid schemes with mixed use of young and progeny tested bulls. However, to confirm these results, more replicates are needed.

In conclusion:

- The genomic structure of the Danish Jersey population reflects its population history.
- Including genomic breed proportions at whole genome level in a random regression model did not improve the reliabilities of genomic predictions. Genomic prediction models that are able to account for a more detailed population structure at individual marker level should be a future focus area.
- The Danish Jersey population is challenged by low reliabilities of genomic predictions limiting the benefit of more intense use of genotyped young bulls.
- Addition of genotyped cows to the reference population has a positive effect on the accuracy of genomic selection. Breeding schemes using young bulls more intensively in particular benefit from this.
- Ways to increase the reliabilities of genomic predictions must be explored.

It is therefore recommended that:

- Initiatives are taken to form a global Jersey reference population by exchanging of already genotyped progeny tested bulls.
- Genotyped cows are included in the reference in order to increase reliabilities of genomic breeding values and hence increase genetic gain.

Sammendrag

Genomisk selektion øger sikkerheden på avlsværdier for unge avlsdyr uden egne registreringer og forbedrer derfor avlsfremgangen. Derfor er genomisk selektion allerede blevet en del af beslutningsgrundlaget i mange avlsprogrammer hos malkekvæg. Metoden har dog vist sig at være mere effektiv i den talmæssigt store Holstein race sammenlignet med de mindre racer.

Hovedformålet for denne afhandling var at undersøge, hvorledes anvendelsen af genomisk selektion kan optimeres med hensyn til totaløkonomisk avlsfremgang, indtjening og stigning i indavl, samt usikkerheden (risikoen) på responset for en lille malkekvægpopulation. Dansk Jersey blev anvendt som model for en lille population. Det forventes, at konklusionerne fra dette studie også er relevant for andre små malkekvægpopulationer.

I **artikel I** blev sikkerhederne for de genomiske avlsværdier beregnet for alle egenskaber, der indgår i avlsmålet for Dansk Jersey. En bayesiansk metode blev anvendt til at beregne markøreffekterne. Sikkerhederne på de genomiske avlsværdier var i gennemsnit 0.04 højere end sikkerhederne på afstamningsindeksene. Sikkerhederne afhang af den anvendte valideringsmetode. Sikkerhederne varierede fra egenskab til egenskab uden en klar sammenhæng til egenskabens arvbarhed.

Dansk Jersey er et eksempel på en blandet race, der indeholder dyr med varieret racesammensætning af original dansk og amerikansk Jersey. I **artikel II** blev det analyseret, hvorvidt populationsstrukturen kendt fra Dansk Jersey kunne genfindes i den genomiske struktur, som den findes i racen i dag. For det første blev det vist, at koblingsuligevægten i gruppen af blandede Dansk Jersey dyr var mindre end i gruppen af dyr bestående hovedsagelig af oprindelige, rene danske eller amerikanske Jersey. For det andet blev det vist, at raceandele beregnet ud fra markørinformationen var i overensstemmelse med de afstamningsbaserede raceandele. En direkte anvendelse af de genomiske raceandele i en tilfældig regressionsmodel forbedrede dog ikke sikkerhederne på de genomiske avlsværdier sammenlignet med en basal genomisk model. De basale genomiske modeller tager muligvis allerede højde for denne struktur

I **artikel III** blev den optimale avlsplan for Dansk Jersey undersøgt ved hjælp af en deterministisk simulationsmodel. Den optimale avlsplan var kendetegnet ved en anvendelse af både genotypede ungtyre samt afkomsundersøgte tyre. Der blev fundet store vekselvirkninger mellem en øget sikkerhed på de genomiske avlsværdier og en mere intensiv anvendelse af ungtyre. En turbo avlsplan udelukkende med brug af ungtyre viste

sig at være genetisk overlegen, såfremt der kan opnås højere sikkerheder på de genomiske avlsværdier. Når indtjeningen blev anvendt som vurderingskriterium, var turboavlsplanen altid overlegen på grund af et kortere generationsinterval samt en reduktion af omkostningerne til opstaldning of fodring af ventetyrene. Resultaterne fra artikel III viser, at de lave sikkerheder på de genomiske avlsværdier udgør en begrænsning i forhold at indføre en mere effektiv avlsplan med en mere intensiv brug af genotypede ungtyre på bekostning af afkomsundersøgte tyre.

En måde at øge sikkerheder på de genomiske avlsværdier, er at anvende genotypede køer i reference populationen. Dette aspekt blev undersøgt i **artikel IV** ved hjælp af stokastisk simulering. Simuleringerne viser, at sikkerhederne på de genomiske avlsværdier stiger mærkbart, når genotypede køer indgår i reference populationen. Avlsfremgangen er desuden større i en turboavlsplan, sammenlignet med en hybrid avlsplan, hvor både genotypede ungtyre og afkomsundersøgte tyre anvendes. Indavlsstigningen var også lavere i turboavlsplanen. Risikoen, beregnet som variansen på responset, var størst i turboavlsplanen sammenlignet med hybrid avlsplanen. For at kunne drage sikre konklusioner kræves dog flere gentagelser af simulationerne.

Hovedkonklusionerne i dette studie er

- Den genomiske struktur i Dansk Jersey afspejler racesammensætningen baseret på afstamningen.
- Direkte hensyntagen til raceandele beregnet ud fra markørinformationen i en tilfældig regressions model forbedrede ikke sikkerhederne på de genomiske avlsværdier. Genomiske prædiktionsmodeller der tager højde en mere detaljeret populationsstruktur på markør niveau bør være et fremtidigt indsatsområde.
- Dansk Jersey er udfordret af lave sikkerheder på genomiske avlsværdier, som begrænser muligheden for at gennemføre en mere effektiv avlsplan med øget brug af genotypede ungtyre.
- Mulighederne for at øge sikkerhederne på genomiske avlsværdier bør undersøges.

Anbefalingerne er:

- Der tages initiativ til at danne en global Jersey reference population med udveksling af allerede genotypede afkomsundersøgte tyre.
- Genotypede køer inddrages i reference populationen med henblik på at øge sikkerheden på genomiske avlsværdier. Avlsplaner der anvender ungtyre mere intensivt vil i særlig grad drage nytte af dette.

List of abbreviation

DGV	Direct Genomic Value
DRP	De-Regressed Proofs
GEBV	Genomically Enhanced Breeding Values
GxE	Genotype versus Environmental Interaction
GS	Genomic Selection
LD	Linkage Disequilibrium
MAS	Marker Assisted Selection
Ne	Effective population size
QTL	Quantitative Trait Loci
SNP	Single Nucleotide Polymorphism

General introduction

Genomic selection (GS) has recently become an integrated part of the breeding decisions in many dairy cattle breeding schemes. The method seems to be the most promising new development in improving genetic merit since the introduction of artificial insemination with frozen semen.

GS has turned out to be more efficient in the numerical bigger Holstein populations compared to smaller breeds such as Jersey (Pryce and Daetwyler, 2012). This is mainly due to the availability of larger reference populations and a higher degree of homogeneity and thus more reliable GEBV in the large populations. In contrast, smaller populations suffer from relatively less reliable GEBV. This is mainly due to the limited size of the historic reference populations (Goddard and Hayes, 2009) and, for some breeds, also the lack of homogeneity (Brøndum et al., 2011). This limits the benefits of introducing GS in smaller breeds and raises the question of how to overcome this.

Accordingly, the aim of this thesis is to 1) assess the present value of GS in the smaller dairy cattle breeds exemplified by the Danish Jersey breed and 2) examine the possibilities of increasing this value.

The term "small population" covers a wide range of populations and there are multiple definitions. In this thesis a small breed is defined as "a production population with an active selective breeding program, that is genetically and economic competitive in the dairy production sector". We do not consider the very small breeds which are very restricted in population size and hence defined as a breed targeted for conservation alone. Apart from the Danish Jersey population examples of other breeds that are covered by this definition are basically other breed groups than Holstein listed in Table 1, which has remarkably smaller sire reference populations compared to Holstein: Other Jersey populations, Brown Swiss, Fleckvieh, Red Group, Montbéliarde and Normande.

Small populations as a future source of genetic variability

There are important arguments for maintaining the smaller dairy cattle populations as competitive breeds. Firstly, genetic diversity within and between populations is a basic resource for a continuous development of production efficiency in animal production and for ensuring product variability for the consumers. Secondly, future adaptations to new breeding goals, climate change, limited resources of feed and water in combination with a rapidly growing and increasingly wealthy human population (Hermansen, 2012) may also

require diverse genetic resources which are adapted to specific environments (GxE). Finally, the existence of competitive breeds with diverse genetic backgrounds is a crucial requirement for running efficient crossbreeding programs (Sørensen et al., 2008). In addition to these arguments there is a political request to ensure that genetic diversity is preserved. FAO is committed to implement a global action plan to protect animal genetic resources (FAO, 2007). Also, at the Nordic level the "Nordic Council of Ministers" has established a policy for sustainable management of farm animal genetic resources (Fimland, 2005).

From Conventional breeding schemes towards Genomic Selection

The value of a breeding scheme can be defined by the monetary genetic gain. That is genetic gain expressed in terms of net monetary units. The main contributors to a high monetary genetic gain are 1) a short generation interval, 2) reliable estimates of breeding values, 3) high selection intensity, 4) high levels of additive genetic variation in the population, and 5) low costs of the breeding scheme.

Conventional breeding schemes based on selection among bulls with information from large daughters groups have for the last four decades proved to be an effective tool to generate high monetary genetic gain in dairy cattle. However, these breeding schemes are characterized by long generation intervals and high costs related to feeding and housing of bulls as well for collection of records from daughter groups (paper III).

Some of these factors are changed, when new sources of information, such as DNA information are incorporated. Use of DNA information allows selection of breeding candidates early in life with higher reliabilities compared to conventional pedigree averages and it is therefore expected to improve genetic gain.

The idea behind the use of DNA information is that genetic markers in LD with QTL have the potential to explain part of the genetic variation. Selection decisions can then be based on these markers with higher accuracy compared to when using pedigree information alone.

MAS seemed to be a promising tool for integrating DNA information in breeding decisions and thereby select young breeding candidates with higher reliability. However, MAS requires large grandsire daughter group designs in order to capture genetic variation through powerful QTL detection. This is not feasible in small cattle populations. In addition MAS were facing two main problems. First the effects of the largest QTLs were

often overestimated and the QTL with minor effects were difficult to detect. Both problems are mostly solved using GS. Only few applications of MAS in commercial breeding program has been seen (Boichard et al., 2012).

In contrast to MAS, GS is an option to integrate marker information in breeding decisions for the smaller breeds. In 2001 Meuwissen et al. (2001) proposed the theory behind GS. The general principle behind the method is that genetic marker effects are estimated in a historic reference population where the individuals are genotyped in addition to having phenotypes or reliable estimates of genetic merit. The estimated marker effects are then used to construct a prediction model for the breeding value of candidate animals only based on marker information. Although proposed in 2001, it was not until 2008, where dense genetic marker panels covering the entire cattle genome with thousands of SNP marker (Matukumalli et al., 2009) became commercially available, that GS started to play a major role in the cattle breeding industry (Harris, 2010; Su et al., 2010; VanRaden et al., 2009).

Accuracy of genomic predictions

Several factors influence the accuracy of genomic predictions. The accuracy can be grouped in two sets of correlations (Goddard, 2009). One set evaluating the strength of LD between the markers and the QTL defining the trait, and another set defining the accuracy of estimated marker effects.

In general, how accurately marker effects are estimated depends on the heritability of the trait, the accuracy of the EBV, the density of the marker set and finally the number of genotyped animals with records, which we refer to as the historic reference population (Goddard, 2009; Hayes et al., 2009). Accordingly, the factors influencing the reliabilities of GEBV may differ between populations (Hayes et al., 2009). It is therefore important to estimate the reliabilities of GEBV in the specific population where GS is going to be applied (Paper I).

The general assumption in order for GS to work, is that all, or a high proportion of the QTLs are in sufficient LD with some of the markers. The expectation is that many QTL affects each trait. Therefore we aim for a marker set covering the whole genome. The stronger the LD between the markers in the population, the better the chance is to capture a high proportion of the genetic variation through the markers. The pattern of LD in the population is shown to be determined by the historic mating structure both recent and distant (Sved, 1971). Based on this, we expect that GS will work less efficient in breeds being subject to admixture from other populations. In Paper II we investigate the

history of the Danish Jersey breed, based on the available SNP marker information. Methods for using the population structure in the prediction of genomic breeding values are evaluated.

Also, the heritability of the trait is a parameter that influences the reliability of the DGV. With a high heritability, fewer records are needed to obtain a high reliability of the EBV used to estimate the marker effects. However, most breeding schemes are designed with big daughter group sizes that provide high reliabilities of EBV even for traits with low heritabilities. In paper I the reliabilities of DGV for a wide range of heritabilities are predicted.

A large reference population is a key factor for high reliabilities of genomic predictions. However, the numbers of animals with EBV based on daughter group information are far smaller than the number of marker effects, whose effects are to be estimated (Goddard and Hayes, 2009). An increase in the reliability of genomic predictions can be obtained by enlarging the reference population. This can be done by including more progeny tested historic bulls in the reference population within breed. Another option is to merge the reference populations for genetically closely associated breeds with similar definitions of EBV. This has shown to be a powerful way to increase reliabilities of genomic predictions in Holstein (Lund et al., 2011; Wiggans et al., 2011). Merging the largest European Holstein reference populations improved the reliabilities by 11% points compared to the Nordic Holstein reference population (Lund et al., 2011). However, if the SNP marker effects are not in LD with QTL both within and across the populations, the gain by merging the reference populations might be reduced. As an example, the combination of the Nordic Red populations only resulted in minor improvement of the reliabilities (Brøndum et al., 2011). A third option for enlarging the reference population is to merge more distantly related reference populations. However, it requires more dense markers, such that the persistence between marker and QTL across the populations are maintained. de Roos et al. (2008) found that at least 300,000 markers were required to find sufficient markers in LD with the QTL across Jersey and Holstein.

Finally the choice of prediction model may influence the reliability of the prediction (Meuwissen et al., 2001). The aim is that the model is able to capture a high proportion of the genetic variance by using marker information, the pedigree structure and the underlying QTL.

In this thesis, two of the most commonly used models for routine genomic predictions are used (Hayes et al., 2009): 1) The GBLUP approach assuming that all SNP

effects are normal distributed with same variance and 2) the Bayesian approach, which allows each marker to have its own variance of SNP effects.

In paper I, a Bayesian model (BayesB) with a common prior distribution of scaling factors was used was used to validate the reliabilities of DGV. Based on a validation study in Danish Holstein, this model was shown to be the model with the best prediction ability (Su et al., 2010). In papers II and IV, the GBLUP approach was used (Christensen and Lund, 2010; Gao et al., 2012). A detailed validation of prediction models was, however, not the focus area of this thesis.

Challenges for the smaller populations

Smaller dairy cattle populations suffer from relatively less accurate genomic predictions compared to the large populations despite of the various options for obtaining higher accuracy described above.

For smaller populations, the possibilities for generating large reference populations are few, due to the limited number of progeny tested bulls that can be generated within the population. Table 1 shows the size of reference populations for the breeds which at present operate a GS breeding program. For all breeds except Holstein, the size of the reference population is in the range of 1,000 and up to 9,000 bulls. In the Jersey breed this can, to some extent, be alleviated by merging the global Jersey sire reference populations, which is expected to increase the number of bulls in the reference population to around 8,000. This is, however, still much smaller than the Holsteins sire reference population.

Merging the reference populations across breeds is, at least in theory, another option for generating larger reference populations for the smaller populations. In practice, no significant gain has at present been obtained by merging breeds. Pryce et al. (2011) found only very little gain in reliability by combining reference populations from Jersey, Fleckvieh and Holstein using a 50k chip.

At present, the most promising option for enlarging reference populations in the smaller populations is to include genotyped cow in the reference population (Buch et al., 2011; Mc Hugh et al., 2011). This option has become more relevant with decreasing cost of genotyping and the release of new lower cost low-density marker platforms (Wiggans et al., 2013), as a high number of genotyped cows will be required to increase reliability of genomic predictions (Goddard and Hayes, 2009). Inclusion of cows in the reference

population will provide more recent information compared to the use of older historical bulls. Also, genetic variation may be lost, when the reference population only consists of bulls with a single EBV as average information content for a big daughter group. In paper IV we evaluate the value of including cows in the reference population for Danish Jersey with approximately 1000 bulls in the reference population.

Breeding schemes using genomic information

The new opportunity of using genomic information to select young breeding candidates with a higher accuracy has a potentially high impact on the optimal design of the breeding scheme (Pryce and Daetwyler, 2012).

Several studies have shown that higher genetic gain can be achieved if young males without progeny performance are used as parents for the next generation (Buch et al., 2012; de Roos et al., 2011; Schaeffer, 2006). Therefore, cattle breeding organizations move towards breeding schemes with more intensive use of young bulls as bull sires as well as for insemination of cows. This is in part due to higher reliability of genomic predictions compared to parent average reliabilities and in part due to reduced cost when keeping fewer waiting bulls. Furthermore, the cost of genotyping has decreased. This makes it even more feasible to increase the amount of genotypings followed by increased selection intensity in the young bull selection pathway.

In general, genomic information can be applied in two types of genomic breeding designs 1) a pre-selection scheme and 2) a turbo scheme. In pre-selection schemes an intensive genomic evaluation of young bulls entering progeny testing is applied. The choice is whether the number of progeny tested bull should remain constant compared to a conventional breeding scheme or reduced in order to offset for the costs of genotyping bull calves. However, the long-term value of adding progeny tested bulls to the reference population should also be taken into account, as this might have an impact on reliabilities of genomic predictions for a small breed (paper IV). In the turbo schemes, only young bulls are used as both bull sires and for insemination of cows. This reduces the generation interval. In addition, turbo schemes are cheaper to run, as costs related to feeding and housing of waiting bulls can be eliminated (König et al., 2009; Schaeffer, 2006).

Most of the simulation studies published so far has focused on the optimization of genomic breeding schemes in larger cattle populations (Pryce and

Daetwyler, 2012). In a review study, (Pryce and Daetwyler, 2012) reported an extra genetic gain from applying genomic selection in the range of +28% to +108% depending on the underlying assumptions. Since genomic information adds less to the reliabilities of genomic predictions in smaller populations, we expect that the conclusions from studies of larger populations might not apply for the smaller populations.

Two types of models are used for evaluation of the optimal breeding schemes: A deterministic approach in paper III and a stochastic approach in paper IV. As deterministic models are fast to run, a broad range of breeding schemes and interaction effects between breeding scheme parameters can be evaluated. Stochastic models simulating the build-up of SNPs and QTL and are far more computer demanding. Stochastic simulations also rates of inbreeding and variance of response, because more replicates of the same breeding scheme is performed. In addition, effects of selection (Bulmer, 1971) and inbreeding on genetic variance can easier be accounted.

Inbreeding

The conventional breeding schemes applied up to now have been successful in generating genetic gain because of intensive use of the best AI bulls and use of BLUP breeding value estimation methods. These methods use all information sources from relatives, which therefore tends to favor animals within certain families. The intensive use of few superior bull sires, has resulted in small effective population size (Sørensen et al., 2005) for Danish Jersey and Holstein.

This decrease has been very pronounced in the Jersey breeds. The effective population size is estimated to 55 in the Danish Jersey population and 53 in the US Jersey population (Stachowicz et al., 2011). A further increase is expected in the near future due to even more intensive use of the best bulls. FAO recommends a minimum effective population size of 50 or equivalently a maximum of 1% increase in inbreeding per generation (FAO, 2007), in order to maintain a sufficient amount of genetic variability within the breed.

Selection using genomic information provides more information about the Mendelian sampling term and thus more weight are put on the individual's own information for young selection candidates compared to the parent average information. This makes it possible to distinguish among full-sibs without own information. It tends to favor selection of animals from a larger number of different families. GS is therefore

expected to reduce inbreeding per generation compared to traditional BLUP selection assuming the same breeding scheme (Daetwyler et al., 2007).

However, as genomic information is expected to increase reliability of young breeding candidates, the genomic breeding schemes acts in favor of shortening the generation interval which might increase inbreeding per time unit. The net effect on the rate of inbreeding per time thus depends of the breeding scheme used (Buch et al., 2011; Lillehammer et al., 2011). In general, lower inbreeding rates are expected in genomic pre-selection schemes compared to turbo schemes (Bouquet and Juga, 2013).

In paper IV we investigate how inbreeding develops in a small dairy cattle population for different breeding schemes and for different strategies for updating the reference population.

Table 1: Overview of breeding groups with an active genomic selection breeding program

Breed group	Country/area	Number of bulls in reference	Recorded cows	Cows in reference	Reference*
Brown Swiss[1]					
	Austria	3,800	52,524	0	Egger-Danner, 2013
	France	4,000	17,430[5]	0	Fritz, 2013
	Germany	3,800	147,694	0	Egger-Danner, 2013
	USA	5,404		508	Chesnais, 2013
	Switzerland	3,729	203,000[6]	0	Schnyder, 2013
Cross breed					
	New Zealand, CRV[9]	3,460	-	2,700	Schrooten, 2013
	New Zealand, LIC[10]	510	1,876,800	1,040	Sherlock, 2013
Holstein					
	Australia	4,364	355,036	13,851	Pryce, 2013
	Canada[2]	20,822	730,054	0	Chesnais, 2013
	DFS[3]	23,779	615,000	0	Nielsen, 2013
	France[3]	24,000	1,700,044[5]	0	Fritz, 2013
	Germany[3]	25,436	1,800,000	0	Alkhoder, 2013
	Ireland[7]	4,500	-	0	Pryce and Daetwyler, 2012
	Italy[2]	19,104	1,128,626	0	Kaam, 2013
	New Zealand, CRV[9]	3,460	-	2,700	Schrooten, 2013
	New Zealand, LIC[10]	3,430	1,757,200	1,340	Sherlock, 2013
	North America[2]	20,822	2,200,000	31,342	Chesnais, 2013
	Poland[3]	2,689	656,340	0	Jedraszczyk, 2013

	Spain[3]	21,656	500,000	0	Hernández, 2013
	Switzerland[7]	2,830	70,000	0	Schnyder, 2013
	The Netherlands[3]	24,500	900,000	0	Schrooten, 2013
	USA[2]	20,822	2,200,000	31,342	Chesnais, 2013
Jersey					
	Canada[2]	2,814	33,000	0	Chesnais, 2013
	DFS	1,205	68,000	0	Nielsen, 2013
	USA[2]	2,814	268,998	8,685	Chesnais, 2013
	New Zealand, CRV[9]	3,460	-	2,700	Schrooten, 2013
	New Zealand, LIC[10]	2,080	556,600	1,100	Sherlock, 2013
	Australia	1,017	175,000	4,240	Pryce, 2013
Red Group					
	DSF[4]	7,255	365,000	0	Nielsen, 2013
	NRF	3,000	200,000	0	Heringstad, 2013
Flekvieh					
	Austria[8]	9,000	279,691	0	Egger-Danner, 2013
	Germany[8]	9,000	682,019	0	Alkhoder, 2013
Montbéliarde					
	France	2,300	405,309[5]	0	Fritz, 2013
Normande					
	France	2,100	239,666[5]	0	Fritz, 2013

DSF: Denmark, Sweden and Finland
1: International collaboration through the Brown Swiss Intergenomics project of INTERBULL
2: common reference for Holstein in USA, Canada, UK and Italy
3: exchange of reference bulls in Eurogenomics
4: bulls from Norway included in reference population
5: number of lactations
6: number of registered cows and pregnant heifers
7: exchange with Czech Republic, Poland and a few from Italy
8: exchange of reference bull between Austria and Germany
9: crossbred evaluation
*: All references are personal communications, except for Pryce and Daetwyler (2011)

Aims of the thesis

The overall objective of this thesis was to examine how genomic selection is optimized, in terms of monetary genetic gain, inbreeding and risk of incorporating genomic information in the breeding strategy for a small dairy cattle population. A small dairy cattle population was exemplified using Danish Jersey.

The objective was achieved by:

- Estimating reliabilities of direct genomic values for traits included in the breeding goal for Danish Jersey using different validation methods (Paper I).
- Quantifying the value of including marker information in the selection of young breeding candidates assessed by comparing the reliability of the genomic predictions with the pedigree index selection (Paper I).
- Investigating whether the historic pedigree population structure in Danish Jersey is reflected in its genomic structure (Paper II).
- Examining if explicitly accounting for the population structure by genome wide grouping of animals improves genomic predictions (Paper II).
- Investigating the effect of running a genomic breeding scheme compared to a conventional progeny testing program (Paper III).
- Investigating the value of increased genomic information origination either from higher reliabilities of predictions or increasing the selection intensity of young bulls by genotyping more bull calves (Paper III).
- Examining whether or not an increased value of genomic information interacts positively with more intensive use of young bulls in the breeding scheme (Paper III).
- Investigating if genotyped cows included in the reference population increase reliability of genomic predictions and if the increase in monetary genetic gain interacts with the applied breeding scheme (Paper IV).
- Evaluating the breeding schemes according to inbreeding and variance of the response (risk)

Paper I

Reliabilities of genomic estimated breeding values in Danish Jersey

J.R. Thomasen, B. Guldbrandtsen, G. Su, R. F. Brøndum, and M. S. Lund

Animal (2012), doi:10.1017/S1751731111002035

Animal (2012), **6:5**, pp 789–796 © The Animal Consortium 2011
doi:10.1017/S1751731111002035

animal

Reliabilities of genomic estimated breeding values in Danish Jersey

J. R. Thomasen[1,2†], B. Guldbrandtsen[1], G. Su[1], R. F. Brøndum[1] and M. S. Lund[1]

[1]*Department of Molecular Biology and Genetics, Faculty of Science and Technology, Aarhus University, PO Box 50, 8830 Tjele, Denmark;* [2]*VikingGenetics, Ebeltoftvej 16, 8860 Assentoft, Denmark*

(Received 8 March 2011; Accepted 25 August 2011; First published online 11 November 2011)

In order to optimize the use of genomic selection in breeding plans, it is essential to have reliable estimates of the genomic breeding values. This study investigated reliabilities of direct genomic values (DGVs) in the Jersey population estimated by three different methods. The validation methods were (i) fivefold cross-validation and (ii) validation on the most recent 3 years of bulls. The reliability of DGV was assessed using squared correlations between DGV and deregressed proofs (DRPs). In the recent 3-year validation model, estimated reliabilities were also used to assess the reliabilities of DGV. The data set consisted of 1003 Danish Jersey bulls with conventional estimated breeding values (EBVs) for 14 different traits included in the Nordic selection index. The bulls were genotyped for Single-nucleotide polymorphism (SNP) markers using the Illumina 54 K chip. A Bayesian method was used to estimate the SNP marker effects. The corrected squared correlations between DGV and DRP were on average across all traits 0.04 higher than the squared correlation between DRP and the pedigree index. This shows that there is a gain in accuracy due to incorporation of marker information compared with parent index pre-selection only. Averaged across traits, the estimates of reliability of DGVs ranged from 0.20 for validation on the most recent 3 years of bulls and up to 0.42 for expected reliabilities. Reliabilities from the cross-validation were on average 0.24. For the individual traits, the reliability varied from 0.12 (direct birth) to 0.39 (milk). Bulls whose sires were included in the reference group had an average reliability of 0.25, whereas the bulls whose sires were not included in the reference group had an average reliability that was 0.05 lower.

Keywords: cross-validation, direct genomic value, genomic selection, reliability, composite breed

Implications

Inclusion of marker information in the selection of young breeding candidates on average improves the reliability by 0.04 compared with parent index selection. To assess the reliability of genomic predictions, it is important to reduce dependency between reference and test population, which is important for estimation of genomic reliabilities. Future successful use of genomic information in Danish Jersey requires more reliable genomic breeding values. The most efficient strategy could be through a collaboration with other Jersey populations or alternatively with other cattle populations and across breed evaluations.

Introduction

In genomic selection (GS; Meuwissen *et al.*, 2001) marker effects are estimated in a genotyped reference population where individuals also have phenotypes or reliable estimates of genetic merit. The marker allele effects predicted using the reference population are then used to construct a prediction model for the breeding value of candidate animals with only marker information. These predictions are called direct genomic values (DGVs).

Reliabilities of genomic breeding values are difficult to determine. In order to optimize the use of GS in practical breeding programs, it is important to estimate the reliabilities of the DGV. Reliabilities of DGV depend on many factors such as the number of bulls in the reference population, the heritability of the trait, the genetic structure of the population and the numbers of markers used for genomic prediction (Hayes *et al.*, 2009a). All these factors may differ between populations. It is therefore important to evaluate the reliabilities of the genomic predictions in the same population from which breeding candidates are being selected. Reliabilities based on real data have been reported from Holstein populations (Hayes *et al.*, 2009a; VanRaden *et al.*, 2009; Harris and Johnson, 2010; Lund *et al.*, 2010;

† E-mail: jrt@vikinggenetics.com

Su *et al*., 2010) and Jersey populations (Hayes *et al*., 2009b; Harris and Johnson, 2010).

Three different methods are used in these studies to estimate the reliabilities of DGV. First, Su *et al*. (2010) applied a cross-validation method to real data where subsets of proven bulls in turn are used as test bulls for predictions of DGV by omitting their estimates of genetic merit. Cross-validation has previously been used for validation on simulated data (Villumsen and Janss, 2009). Second, model-estimated reliabilities have been calculated from Markov chain Monte Carlo (MCMC) samples from the Bayesian posterior distributions of the DGVs. They were used as validation criteria in the study by Su *et al*. (2010). In conventional breeding value estimation (Misztal and Wiggans, 1988; Tier and Meyer, 2004), reliabilities of the estimated breeding values (EBVs) are also calculated from approximations of prediction error variances (PEVs) from the individual animal model solutions of genetic merit. Third, validation of the reliabilities of DGV for the group of the most recent bulls with EBV was used in the studies by Harris and Johnson (2010), Lund *et al*. (2010) and VanRaden *et al*. (2009). Validation of the most recent 3 or 4 years of bulls seems to have become the most commonly used method (Lund *et al*., 2010; Van Raden *et al*., 2009). Interbull validation test for genomic evaluations is now based on validations of the most recent 4 years of bulls (Mäntysaari *et al*., 2010).

Estimates of reliabilities of DGV show substantial variation depending on the validation method used. In general, this has two consequences. First, it makes it difficult to compare the estimated reliabilities between studies using different methods. Second, the uncertainty of the true reliability of the DGV makes it difficult to predict the true value of genomic information and therefore makes it uncertain how to design the optimal breeding plan including genomic information.

The main purpose of this study is to estimate the reliabilities of DGV for traits included in the breeding goal in the Danish Jersey population using different validation methods. Three different methods for investigating the reliability of DGV are studied: fivefold cross-validation, validation for most recent 3 years of bulls and reliabilities calculated from the model. In addition, factors affecting the level of reliability are studied. Finally, the validations are compared to the predicted reliabilities (Goddard, 2009) and the reliability of inclusion of marker information in the selection of the young breeding candidates is compared with the reliability of pedigree index (PI) selection.

Material and methods

Data

The progeny-tested Jersey bulls in the analysis were born between 1985 and 2004. They belong to 81 paternal half-sib families with 2 to 71 sons in each family. Another 25 sires had only one son in the data set. All bulls were genotyped using the Illumina Bovine SNP50 BeadChip (Illumina, San Diego, CA, USA) (Matukumalli *et al*., 2009). Single-nucleotide polymorphisms (SNPs) typing was performed at the Department of Molecular Biology and Genetics at Aarhus University.

Editing of genotypic data

The genotypic data were edited both by animal and by loci. After marker data quality checking, 1003 bulls and 33 524 SNP markers were available. For animals, the requirements were a call rate above 95% except for a few old animals, which were accepted with call rates of at least 85%. Marker loci were accepted if they had a call rate of at least 95%. Loci with a minor allele frequency (MAF) less than 5% were excluded. Loci without a map position in the Btau 4.0 assembly were discarded. Animals with an average GenCall score (Illumina, 2005) of less than 0.65 were discarded. Individual marker typings with a GenCall score of less than 0.6 were discarded. On average 99.6% of the markers could be assigned to a genotype, with a range from 86.3% to 99.9% (Table 1). There was no evidence that the older test groups had lower call rates and therefore had a lower SNP marker quality. The genotyped bulls represent nearly all proven bulls born in the period from 1988 to 2004. Only 12 bulls from the years 1985 to 1987 were available.

Two sets of EBVs were used as response variables for the predictions of DGV in this study. The first was EBVs published in July 2009 as the basis for predictions used in the fivefold cross-validations. The second was EBVs published in June 2006 used for the predictions of DGV in the validation method for most recent 3 years of bulls. The traits investigated were the 14 combined traits included in the Nordic Total Merit index for Danish Jersey (Pedersen *et al*., 2010). A detailed description of the index traits and the calculations of EBV can be obtained from Team Avlsværdivurdering (2009). Averaged over all traits, the number of typed bulls with EBVs was 974.

Each combined trait consists of a varying number of traits weighted with an economical value. The range of heritabilities (Team Avlsværdivurdering, 2009) for each trait included in the

Table 1 *Call rates of marker data and structure of the whole data set and five test groups*

Test group	Number of bulls	Average call rate	Number of half-sib families	Interval of birth years	Average birth year
All	1003	0.9962	106	1985 to 2004	1996
A	185	0.9951	21	1985 to 1993	1990
B	225	0.9960	15	1991 to 1999	1993
C	207	0.9969	19	1994 to 1999	1996
D	202	0.9970	24	1997 to 2004	1999
E	184	0.9960	27	1996 to 2004	2002

Table 2 *Number, mean and s.d. of EBV, reliability of EBV for the reference bulls, range of heritabilities (h^2) for the component traits of each combined trait and predicted reliabilities of DGV*

Trait	Number of reference bulls	Mean EBV	s.d. of EBV	Reliability of EBV	Range of h^2 for component traits	Predicted reliability of DGV*
Maternal calving	998	100.46	9.25	0.53	0.01 to 0.03	0.23
Udder health	996	101.46	9.84	0.63	0.01 to 0.03	0.26
Other diseases	877	98.78	9.92	0.48	0.01 to 0.05	0.19
Direct birth	1003	99.72	8.39	0.69	0.01 to 0.11	0.27
Fertility	992	100.58	11.36	0.62	0.02 to 0.04	0.25
Temperament	992	97.71	9.21	0.32	0.05	0.15
Feet and legs	916	96.74	9.90	0.60	0.09 to 0.16	0.23
Longevity	971	103.41	9.28	0.71	0.10	0.27
Udder conformation	963	94.94	9.68	0.76	0.17 to 0.42	0.29
Milking ability	930	98.92	10.02	0.51	0.19	0.21
Protein	1000	91.46	12.32	0.93	0.22 to 0.35	0.33
Yield	1000	90.31	12.36	0.93	0.22 to 0.44	0.33
Fat	1000	91.15	11.90	0.93	0.23 to 0.38	0.33
Milk	1000	94.82	11.65	0.93	0.27 to 0.44	0.33
Average	974	99.60	10.40	0.68	–	0.26

EBV = estimated breeding value; DGV = direct genomic value.
*Calculated from the formula derived by Goddard (2009).

combined traits is shown in Table 2. The heritabilities vary from 0.01 to 0.03 for maternal calving and udder health to more than 0.20 for production traits.

Statistical model

Marker effects were estimated in a Bayesian model with all SNPs included as predictors. A detailed description of the Bayesian model is given in Su *et al.* (2010) and also in Villumsen *et al.* (2009). These procedures are used as implemented in the IBay package v1.46 (Janss, 2009). Conventional EBV was used as response variables. In the analyses, EBVs were weighted by *1/(1 – reliability of EBV).* The used reliability was the official reliability published together with the EBV. The calculations of the reliabilities are based on the expected daughter contributions.

Briefly, the following model was used to estimate marker effects:

$$\mathbf{y} = \mathbf{1}\mu + \sum_{i=1}^{m} \mathbf{X}_i \mathbf{q}_i v_i + \mathbf{e}$$

where $\mathbf{y}$ is the vector of published conventional EBV; μ is the intercept; $\mathbf{1}$ is a vector of ones; m is the number of SNP markers; $\mathbf{X}_i$ are design matrices linking animals to the allele types of marker I; $\mathbf{q}_i$ is the vector of scaled SNP effects (scaled by v_i, which is equivalent to the standard deviation) of marker i with $\mathbf{q}_i \sim N(\mathbf{0}, \mathbf{I})$, v_i ($v_i > 0$) is a scaling factor for SNP effects of marker i, and $\mathbf{e}$ is the vector of residuals with $\mathbf{e} \sim N(\mathbf{0}, \mathbf{W}^{-1}\sigma_e^2)$, where $\mathbf{W}$ is a diagonal matrix containing the weights of the EBV. The effects of SNP alleles of marker i are the products of v_i and $\mathbf{q}_i$. Scaling factors v_i were assumed to have a common prior distribution for all markers across the genome, given by

$$v_i \sim TN(0, \sigma_v^2), \quad v_i > 0$$

where *TN* is a positive half-normal distribution.

The DGV for individual k was defined as the sum of predicted effects of SNP over all markers,

$$\mathrm{DGV}_k = \hat{\mu} + \sum_{i=1}^{m} \mathbf{X}_{i(k)} \hat{\mathbf{q}}_i \hat{v}_i$$

where $\hat{\mu}$ is an estimate of the intercept and the sum $\sum_{i=1}^{m} \mathbf{X}_{i(k)} \hat{\mathbf{q}}_i \hat{v}_i$ is the deviation of the DGV from the intercept $\hat{\mu}$ for each individual k. The posterior means of the model parameters, $\hat{\mu}$ and $\hat{\mathbf{q}}_i \hat{v}_i$ are obtained from the MCMC sampler (Villumsen *et al.*, 2009).

The MCMC sampler was run as a single chain with a length of 50 000 iterations. Samples from the first 10 000 iterations were discarded as burn-in. Every fifth sample of the remaining 40 000 was used to estimate the parameters of the realized posterior distributions.

Reliability of DGV

Two different methods were used to investigate the reliabilities of the DGV: (i) fivefold cross-validation and (ii) validation for the most recent 3 years of bulls with EBV.

Deregressed EBVs (also known as deregressed proofs (DRPs) from January 2011 were used for calculation of the reliabilities of DGV. Calculations of the DRP followed the procedures described by Strandén and Mäntysaari (2010). The reliabilities of DGV were calculated as the squared correlation between DGV and DRP divided by the mean reliability of DRP for the test bulls.

Fivefold cross-validation. The reference bulls were divided into five nearly equally sized subsets (184 to 225) according to year of birth. Table 1 shows the number of bulls per test group. Half-sib families having sons born in more than one time-period were all assigned to the same subset. Cross-validations were

performed by in turn omitting EBV from one subset (test data) from the full data set, and then predicting the DGV for the test data based on the remaining data. In order to reduce dependencies between reference data and test data, bulls in the test data which had sons in the reference data were excluded from the test data set. This procedure removed 63 bulls from the calculations of the correlations.

For the validation of all bulls the reliabilities of DGV were estimated as the average of the squared correlation between DGV and DRP from each of the five test data sets. For validation of bulls with sires in the reference population or without sires in the reference population, the reliabilities of DGV were estimated as the within-year squared correlation between DGV and DRP across all the five test data sets. This procedure was used due to small number of test bulls in some of the subsets.

The most recent 3 years validation. The 860 bulls with official EBV in June 2006 were used as reference bulls. In all, 133 bulls born in 2002 to 2004 with official EBV in July 2009 were used as test bulls. The reliabilities of DGV were estimated as the squared correlation between DGV and DRP from January 2011 for these test bulls.

In order to investigate the gain in information about Mendelian sampling from the SNP marker information, the squared correlations between PI and DRP ($r^2_{PI,DRP}$) were calculated for the test bulls in the most recent 3 years validation. The gain was evaluated as the difference between ($r^2_{DGV,DRP}$) and ($r^2_{PI,DRP}$). The PI was calculated using the EBV for the sire and the maternal grandsire (MGS) of the bull as

$$PI = \tfrac{1}{2}(EBV_{SIRE} - 100) + \tfrac{1}{4}(EBV_{MGS} - 100) + 100$$

using the official EBVs from June 2006. Bulls with missing EBVs for their sire or MGS were removed from the calculations. This procedure removed between 2 and 30 animals from the calculations, depending on the trait.

In addition, model-estimated reliabilities were investigated in the most recent 3 years validation. The model-estimated reliabilities were obtained from the PEV following Su *et al.* (2010). The reliability for a candidate '$_i$' was calculated from the formula:

$$r^2_{DGV_i} = 1 - \frac{PEV_i}{\sigma^2_a}$$

where σ^2_a is additive genetic variance estimated as the sum of the variance of DGV and the mean PEV for all candidates. Model-estimated reliabilities were calculated for the test bulls in the most recent 3 years validation analysis.

Moreover, predicted reliabilities of DGV for each trait were calculated using the formula derived by Goddard (2009). The following values were used in the calculations: An effective population size of 42 in Danish Jersey estimated by Sørensen *et al.* (2005); the actual number of reference bulls for each trait (Table 2); a length of the cattle genome of 3000 cM (Bovine Hapmap database); mean reliabilities of EBV for each combined trait from Table 2. The results of the calculations of predicted reliabilities are presented in Table 2. Statistics for the EBV (evaluation in 2009) are given in Table 2. The number of genotyped bulls with EBV varies from 877 bulls for 'other diseases' up to 1003 bulls for birth index. For the reference bulls, standard deviations of EBV were calculated. For the traits fat, fertility, milk, milking ability, protein and yield index, the standard deviations were higher than 10. The bulls with EBVs included in this study are born over a long period (1985 to 2004; Table 1). In the investigated period, there is a genetic trend for all these traits, except for fertility (Danish Cattle Federation, 2010). The average reliabilities of EBV for the tested bulls range from 0.32 for temperament to 0.93 for the production traits.

Results

Reliabilities of DGV

Table 3 shows the estimated reliabilities of DGV for all 14 combined traits included in the breeding goal. The $r^2_{DGV,DRP}$ calculated from the fivefold cross-validation method for all bulls ranged from 0.12 (direct birth) to 0.39 (milk) with an average of 0.24. The group of bulls whose sires were included in the reference group had an average $r^2_{DGV,DRP}$ of 0.25 across all traits with a range from 0.11 for 'direct birth' to 0.40 for milk. For the group of bulls whose sires were not included in the reference group, the average $r^2_{DGV,DRP}$ were 0.20 across all traits. The lowest value was found for 'feet and legs' (0.06) and the highest for milk (0.35).

Reliabilities assessed by the most recent 3 years of bulls were on average 0.22, which was marginally lower (0.02) than the squared correlations calculated from all bulls in the fivefold validation (Table 3). The level of the calculated reliabilities from these two methods was on average marginally lower (0.02 to 0.04) than the predicted reliabilities for DGV (Table 2), which on average was 0.26. The ranking of the reliabilities were in general constant across the fivefold validation and the validation of the most recent 3 years of bulls. Highest reliabilities were obtained for the traits milk and 'udder conformation'. For fertility, the reliability was remarkably higher for the most recent 3 years validation. For 'udder health' the reliability in the fivefold validation was higher than in the most recent 3 years validation. However, the results for fertility and 'udder health' are in concordance with the results from the cross-validation in the youngest test group (E; Table 4).

The corrected squared correlation between the DGV and DRP ($r^2_{DGV,DRP}$) and between the DRP and PI ($r^2_{DRP,PI}$) for the most recent 3 years of bulls are shown in Table 3. Averaging over all traits, the difference in reliability between DGV and PI is 0.04. Highest gain in reliabilities (above 0.10) is obtained for the traits 'udder conformation', 'feet and leg', 'milking ability', fertility and longevity. For the milk production traits the overall improvements are less. Model-estimated reliabilities obtained from the MCMC analyses range from 0.32 to 0.62 with an average of 0.42 (Table 3), which is about twice as high as the reliability obtained from the validation.

Table 3 *Corrected squared correlation between DGV and DRP ($r^2_{DGV,DRP}$) for different groups of bulls and model-estimated reliability of DGV (calculated from prediction error variance) for bulls in the test data*

Trait	Fivefold cross-validation, all bulls $r^2_{DGV,DRP}$	Five old cross-validation, bulls with sires in reference $r^2_{DGV,DRP}$	Fivefold cross-validation, bulls without sires in reference $r^2_{DGV,DRP}$	Most recent 3 years $r^2_{DGV,DRP}$	Most recent 3 years $r^2_{DRP,PI}$	Model-estimated reliability for candidates
Maternal calving	0.17	0.15	0.08	0.11	0.12	0.51
Udder health	0.29	0.33	0.23	0.20	0.19	0.48
Other diseases	0.19	0.18	0.18	0.13	0.22	0.46
Direct birth	0.12	0.11	0.16	0.04	0.00	0.32
Fertility	0.17	0.16	0.18	0.28	0.16	0.47
Temperament	0.23	0.26	0.22	0.26	0.28	0.62
Feet and legs	0.21	0.24	0.06	0.21	0.07	0.39
Longevity	0.26	0.27	0.18	0.26	0.16	0.35
Udder conformation	0.30	0.28	0.24	0.31	0.16	0.34
Milking ability	0.26	0.24	0.22	0.30	0.17	0.34
Protein	0.26	0.27	0.21	0.25	0.29	0.39
Yield	0.20	0.22	0.13	0.16	0.20	0.35
Fat	0.25	0.25	0.16	0.16	0.14	0.37
Milk	0.39	0.40	0.35	0.40	0.37	0.49
Average	0.24	0.25	0.20	0.22	0.18	0.42

DGV = direct genomic value; DRP = deregressed proof; PI = pedigree index.
For the most recent 3 years prediction calculation of squared correlation between DRP and PI ($r^2_{DGV,PI}$) is shown. All squared correlations are adjusted for the mean reliability of DRP.

Table 4 *Reliability of EBV (REL_{EBV}) and corrected squared correlation between EBV and DGV ($r^2_{DGV,DRP}$) from cross-validation within test groups of bulls for the traits Protein, fertility and udder health*

	REL_{EBV}			$r^2_{DGV,DRP}$		
Test group	Protein	Fertility	Udder health	Protein	Fertility	Udder health
A	0.94	0.66	0.65	0.23	0.06	0.22
B	0.93	0.63	0.63	0.26	0.17	0.28
C	0.92	0.58	0.61	0.28	0.14	0.40
D	0.93	0.62	0.64	0.27	0.19	0.41
E	0.93	0.59	0.61	0.26	0.28	0.15

EBV = estimated breeding value; DGV = direct genomic value.

Table 4 presents the reliability of EBV and corrected squared correlation between DGV and DRP for each of the five test groups in the fivefold cross-validation study for the traits protein, fertility and udder health. It was observed that $r^2_{DGV,DRP}$ varied among the five subsets especially for fertility and udder health. For fertility, the reliability is remarkably higher in the last test group (E) and lowest in the oldest test group (A), whereas the reliability for 'udder health' is lowest in test group E.

Discussion

Improvements of reliabilities using marker information

Comparison of the squared correlations between DGV and DRP and the squared correlations between PI and DRP averaged across all traits shows that there is a gain in reliability of 0.04. This shows that use of genomic information adds information about Mendelian sampling, which is not obtained using PI. The information about the SNP marker effects in young selection candidates without own performance, makes DGV a strong pre-selection tool for selection of young candidates within families compared to PI information. For the Nordic Holstein population which has a three times larger reference population of 3037 reference bulls used in the validation, the gain in squared correlation was on average 0.18 for the traits protein, udder depth, somatic cell score, longevity and the fertility trait, non-return rate (Lund *et al.*, 2010). In their study, DRP was used as response variables for predictions of the genomic breeding values.

The level of the calculated reliabilities in our study for the fivefold cross-validation and predictions of most recent 3 years of bulls are on average close to predicted reliabilities (Table 2). The predictions following Goddard's formula (2009) depend strongly on the effective population size used in the predictions. Increasing the effective population size from 42 to 60 decreases the average predicted reliabilities by

0.05. Hayes *et al*. (2009c) compared predicted reliabilities with observed reliabilities of genomic breeding values for both Holstein and Jersey populations. The conclusion from their study was that predicted reliabilities agreed well with the observed reliabilities using recent estimates of effective population size.

Relationship between reliability and heritability

In this study, there was no clear relationship between the heritability used for the calculation of the response variable and the calculated reliability for the DGV. For example, the low heritability trait 'udder health' has a high reliability in the cross-validation and the high heritability production traits fat and protein have low reliabilities. A possible reason is that the response variable for the prediction of DGV is published EBV, which has a relatively high accuracy even for traits with low heritability due to large daughter groups. These results are in contrast to the findings of Luan *et al*. (2009) where a strong relationship between the accuracy of the prediction and the heritability of the trait was observed. Their study was carried out in the Norwegian Red with a smaller reference population of only 500 bulls. The phenotypic data used in their study was daughter-yield deviations (DYDs) with reliabilities between 0.33 and 0.66 for traits with low heritability. Therefore, the size of the reference population in their study may be too small to predict reliable SNP effects for traits with low heritability.

Connection between training and test data set

In a breeding plan with a short generation interval, the candidates might be selected for breeding before their sires are progeny tested and hence included in the reference population. In order to design optimal breeding plans, it is important to know the reliabilities of genomic predictions for the candidates, whose sires are not included in the reference population. In this study, it is seen that the level of reliabilities depends on whether the sire of a candidate bull is included in the reference population. The reliabilities were on average 0.05 higher for the cross-validation, when the sire of the candidate was in the reference population compared to bulls where the sire was not in the reference population. Similar results in a simulated cattle population were reported by Lund *et al*. (2009). They stated that when sires are in the training data, genomic breeding values are estimated using both information of linkage disequilibrium (LD) in the population and sire genetic information, thus having the sire in the training data set provides more information for genomic prediction of the sons and consequently higher accuracy. For animals without close relationship to the reference population it is more difficult to connect combinations of markers with performance information compared to animals with close relationship to the reference population. The most likely reason is that the number of estimated SNP marker effects was far bigger than the number of animals with records. Villumsen *et al*. (2009) investigated the decay of reliabilities over generations without phenotypic information for a Bayesian model with single marker SNP effects. They showed that the reliability was reduced from 0.71 to 0.64 in the first two generations for a simulated trait with heritability of 0.02. This decay is in concordance with the reduction of 0.05 in this study, when comparing the two groups of test bulls with and without sires in the reference population.

In this study, EBVs were used for the predictions of the genomic breeding values. Guo *et al*. (2010) have, in a simulation study, shown that using EBV as response variable for the predictions provided higher or similar reliabilities compared to the use of DYD as response variable. For a Bayesian common prior model (the model used in the present study), the EBV perform slightly better for prediction of genomic breeding values both for high and low heritability traits. An EBV has, compared to a DYD or DRP, a higher reliability, as it contains all available pedigree information. For estimation of SNP effects in small populations such as the Danish Jersey, it is important to use all available information. DRP was, in this study, however, used as proxy for the true breeding values for predictions of the reliabilities of DGV in order to reduce errors of correlations between training and test data set (Amer and Banos, 2010). The time span between the EBV used for the predictions of the DGV is maximized by use of the most recent calculated DRP from January 2011. For the most recent 3-year calculation, the time span is 5.5 years, which reduces the dependency between training and test data set. A similar design for the validation of genomic predictions is used in the study by VanRaden *et al*. (2009).

In the cross-validation study, all paternal half sibs are moved to the same subset and all sires with sons in the reference data are left out in the calculations of the reliabilities. These steps ensure that the dependencies between training and test data set are minimized, but also ensure that the design in the validation is as close as possible to the realistic selection process.

Comparison of validation methods

Averaged over traits, the reliabilities obtained from the fivefold cross-validation are slightly higher compared to those from the most recent 3-year validation. This is expected because there is a lower dependency between EBV used for the predictions of DGV and the DRP used for the test bulls in the most recent 3-year validation.

Model-estimated reliabilities for the 133 candidate bulls in the most recent 3-year study are much higher than estimates from either cross-validation or the most recent 3-year validation. Similar results are found by VanRaden *et al*. (2009), where expected reliabilities on average over 20 traits were found to be 13 percentage points higher compared to reliabilities calculated from validation of the most recent 4 years of bulls. VanRaden *et al*. (2009) listed several arguments why model-estimated reliabilities are higher. On one hand, the reliabilities obtained from the validations may be underestimated, because the test bulls have been selected based on parent average information instead of using a random sample from the population. On the other hand,

some genetic effects may not be in complete LD with the markers. This may lead to an overestimation of the model-estimated reliability. Another explanation for this overestimation could be a possible inflation of the DGV (Aguilar *et al*., 2010), which leads to an overestimated variation of the DGV and thus an overestimated reliability of DGV.

Lund *et al*. (2009) compared different methods to validate prediction models. Using simulated data, the cross-validation method turned out to be the most efficient method to validate the predictive ability of the model. The advantage of the cross-validation procedure is that this procedure makes it possible to retain a large training data set combined with a large test data set. In small cattle populations like Danish Jersey, it is particularly important that the validation procedures use the training and test data set in the most efficient way in order to reduce sampling error.

In a study on Australian Jersey, genomic EBVs for 77 candidate bulls were predicted using a reference population of 287 sires. They obtained reliabilities of 0.37, 0.14 and 0.18 for milk, fat and protein (Hayes *et al*., 2009b). The model used in their study was BayesA (Meuwissen *et al*., 2001) with deregressed breeding values as response variable. These results are marginally lower than the reliabilities we found for milk (0.40), fat (0.16) and protein (0.25) in the prediction of the most recent 3 years of bulls. However, taking the small reference population into consideration, the reliabilities found by Hayes *et al*. (2009b) are relatively high compared to the expected accuracy of genomic breeding values presented by Hayes *et al*. (2009c). The authors argue that this is due to a low effective population size and a high genomic relationship between the reference and test data set. Harris and Johnson (2010) reported an average reliability of 0.55 for the traits protein, fat, somatic cell score and fertility in the New Zealand Jersey population with 1738 reference bulls using a linear mixed model. The calculated reliabilities in our study were lower than the reliabilities reported for the Nordic Holsteins (Su *et al*., 2010) using a much bigger reference population consisting of 3330 bulls. The predictions for the fivefold cross-validation in our study were 0.16 lower than those found by Su *et al*. (2010). In the study by Su *et al*. (2010), EBV was used as response variable both for the prediction of EBV and for calculation of the reliabilities. Therefore, higher estimates of reliabilities are expected due to a higher dependency between reference and test data set. For the expected reliabilities, the level in our study was 0.13 lower than the values for Nordic Holstein (Su *et al*., 2010). The level of the reliabilities in Su *et al*. (2010) for the cross-validation is in concordance with the level predicted in Hayes *et al*. (2009c). Different models and validation methods in the reported studies are used to calculate the reliabilities of DGV, which in general makes comparisons between studies difficult.

Improving GS

The level of the reliability obtained in Danish Jersey is expected to increase as the size of the reference population increases as shown by Goddard and Hayes (2009). The reference population can be extended either by using genomic and phenotypic information from females or through collaboration between Jersey populations. The benefits of collaboration have been shown for the European Holstein populations where four reference populations have been merged (Lund *et al*., 2010). As a result of this collaboration the Nordic Holstein reference population grew from 4000 to nearly 16 000 reference bulls. On average the reliabilities increased by 11% when the reference population was quadrupled. The conclusion from that study was that merging of reference populations is a very efficient way of increasing reliabilities of genomic breeding values. The reliability is also expected to increase by blending with information from the conventional PI. Results from the study by Harris and Johnson (2010) showed a gain in reliability between 1.3% and 4.4% depending on trait from blending DGV with pedigree information for candidate bulls without own phenotypic information. Use of high-density SNP panels may allow using reference populations across breeds, which is an efficient way of expanding the reference population for a small breed such as Danish Jersey. De Roos *et al*. (2008) concluded that at least 300 000 markers are needed to obtain consistent marker effects across breeds. Procedures that combine all phenotypic information, pedigree and genomic information simultaneously for both genotyped and non-genotyped animals are expected to produce more accurate breeding values (Forni *et al*., 2011).

Conclusion

Expected gain in reliability by including genomic information in the selection decisions for new breeding candidates compared to parent index selection was on average across all traits 0.04. No clear connection between the heritability of the trait and the estimated reliability was found. Reliabilities depend on whether the sire of the candidate bull is included in the reference population. The reliability is 0.05 lower when the sire is not included in the reference population than when the sire is included. Averaged across 14 traits, the reliabilities of genomic prediction using the current reference data is in the range between 0.20 and 0.42, depending on estimation method. Estimates for the fivefold cross-validation (average 0.24) and most recent 3 years (average 0.20) provide estimates of the reliabilities that are closest to values predicted for the actual size of the reference population and the effective population size. In order to obtain reliable estimates of reliabilities it is important that the dependency between reference and test data is reduced.

Acknowledgments

This study was performed in the project 'Genomic Selection – from function to efficient utilization in cattle breeding (Grant no. 3412-08-02253)', funded by the Danish Directorate for Food, Fisheries and Agri Business, Mælkeafgiftsfonden, VikingGenetics, Nordic Genetic Evaluation and Aarhus University. We thank Danish Cattle Federation and Nordic Cattle

Genetic Evaluation for providing phenotypic data. The work is co-funded by a PhD scholarship from VikingGenetics and the Ministry of Science, Technology and Innovation.

References

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S and Lawlor TJ 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science 93, 743–752.

Amer PR and Banos G 2010. Implications of avoiding overlap between training and test datasets when evaluationg genomic predictions of genetic merit. Journal of Dairy Science 93, 3320–3330.

Danish Cattle Federation 2010. Avlsanalyser. Retrieved March 1, 2011 from http://www.landbrugsinfo.dk/kvaeg/avl/avlsanalyser/sider/startside.aspx

De Roos APW, Hayes BJ, Spelman RJ and Goddard ME 2008. Linkage disequilibrium and persistence of phase in holstein-friesian, jersey and angus cattle. Genetics 179, 1503–1512.

Forni S, Aguilar I and Misztal I 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genetics Selection Evolution 43, 1–7.

Goddard M 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136, 245–257.

Goddard ME and Hayes BJ 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Review Genetics 10, 381–391.

Guo G, Lund MS, Zhang Y and Su G 2010. Comparison between genomic predictions using daughter yield deviations and conventional estimated breeding values as response variables. Journal of Animal Breeding and Genetics 127, 423–432.

Harris BL and Johnson DL 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. Journal of Dairy Science 93, 1243–1252.

Hayes BJ, Bowman PJ, Chamberlain AJ and Goddard ME 2009a. Invited review: genomic selection in dairy cattle: progress and challenges. Journal of Dairy Science 92, 433–443.

Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K and Goddard ME 2009b. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetics Selection Evolution 41, 1–9.

Hayes BJ, Daetwyler HD, Bowman PJ, Moser G, Tier B, Crump R, Khatkar M, Raadsma HW and oddard M 2009c. Accuracy of genomic selection: comparing theory and results. Procedings from the Associations of Advancement of Animal Breeding and Genetics 18, 34–37.

Illumina Inc. 2005. Illumina gencall data analysis software – gencall software algorithms for clustering, calling, and scoring genotypes. Illumina Pub. no. 370-2004-009, 1–2.

Janss L 2009. IBAY manual version 1.46. Bayesian Solutions, Leiden, The Netherlands.

Lund M, Sahana G, de Koning DJ, Su G and Carlborg Ö 2009. Comparison of analyses of the QTLMAS XII common dataset. I: genomic selection. BMC Proceedings 3, S1.

Luan T, Woolliams JA, Lien S, Kent M, Svendsen M and Meuwissen THE 2009. The accuracy of genomic selection in norwegian red cattle assessed by cross-validation. Genetics 183, 1119–1126.

Lund MS, Roos APW, de Vries AG, Druet T, Ducrocq V, Fritz S, Guillaume F, Guldbrandtsen B, Liu Z, Reents R, Scrooten C, Seefried M and Su G 2010. Improving Genomic Prediction by EuroGenomics collaboration. In Book of abstracts of the 9th World Congress on Genetics Applied to Livestock Production (ed. G. Erhardt), p. 150. Leipzig, Germany.

Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TPL, Sonstegard TS and Van Tassell CP 2009. Development and characterization of a high density snp genotyping assay for cattle. PLoS One 4, e5350.

Mäntysaari E, Liu Z and Van Raaden P 2010. Interbull validation test for genomic evaluations. InterBull Bulletin 41. Proceedings at the 2010 Interbull workshop and industry meeting, March 4–5, Paris, France, 1–5.

Meuwissen THE, Hayes BJ and Goddard ME 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.

Misztal I and Wiggans GR 1988. Approximation of prediction error variance in large-scale animal models. Journal of Dairy Science 71 (suppl. 2), 27–32.

Pedersen J, Sørensen MK, Toivonen M and Eriksson J-Å , Aamand GP 2010. Report on economics basis for a Nordic total merit index. Retreived March 1, 2011 from http://www.nordicebv.info/NR/rdonlyres/B618C0E5-FF6F-4D31-8F86-B3CE4A140043/0/NAV TMI_report_lastversion_131108.pdf

Sørensen AC, Sørensen MK and Berg P 2005. Inbreeding in danish dairy cattle breeds. Journal of Dairy Science 88, 1865–1872.

Strandén I and Mäntysaari E 2010. A recipe for multiple trait deregression. InterBull Bulletin no. 42. Proceedings at the 2010 Interbull meeting, May 31st–June 4th. Riga, Latvia, 21–24.

Su G, Guldbrandtsen B, Gregersen VR and Lund MS 2010. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. Journal of Dairy Science 93, 1175–1183.

Team Avlsværdivurdering 2009. Årstatistik Avl – 2008/2009. Dansk Kvæg, Skejby, Denmark.

Tier B and Meyer K 2004. Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. Journal of Animal Breeding and Genetics 121, 77–89.

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF and Schenkel FS 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science 92, 16–24.

Villumsen TM and Janss L 2009. Bayesian genomic selection: the effect of haplotype length and priors. BMC Proceedings 3 (suppl. 1), S11.

Villumsen TM, Janss L and Lund MS 2009. The importance of haplotype length and heritability using genomic selection in dairy cattle. Journal of Animal Breeding and Genetics 126, 3–13.

Paper II

The admixed population structure in Danish Jersey challenge accurate genomic predictions

J.R. Thomasen, A. C. Sørensen, G. Su, P. Madsen, M. S. Lund and B. Guldbrandtsen

Submitted to Journal of Animal Science

The admixed population structure in Danish Jersey dairy cattle challenges accurate genomic predictions[1]

J.R. Thomasen[*†], A. C. Sørensen[†], G. Su[†], P. Madsen[†], M. S. Lund[†] and B. Guldbrandtsen[†]

[*]VikingGenetics, Ebeltoftvej 16, 8860 Assentoft, Denmark

[†]Center for Quantitative Genetics and Genomics, Faculty of Science and Technology, Aarhus University, DK-8830, Tjele, Denmark

Corresponding author: Jørn Rind Thomasen, e-mail: jotho@vikinggenetics.com

[1] This work was performed in the project "Genomic Selection – from function to efficient utilization in cattle breeding (grant no. 3412-08-02253)", funded by the Danish Directorate for Food, Fisheries and Agri Business, Mælkeafgiftsfonden, VikingGenetics, Nordic Genetic Evaluation and Aarhus University. We thank Danish Cattle Federation and Nordic Cattle Genetic Evaluation for providing phenotypic and pedigree data. The work is co-funded by a Ph.D. scholarship from VikingGenetics and the Ministry of Science, Technology and Innovation.

ABSTRACT: The main purpose of this study is to evaluate whether the population structure in Danish Jersey (DJ) known from the history of the breed also is reflected in its genomic structure. This is done by comparing the linkage disequilibrium and persistence of phase for subgroups of Jersey animals with high proportions of Danish (DNK) or US (USJ) origin. Furthermore, it is investigated whether a model explicitly incorporating breed origin of animals, inferred either through the known pedigree or from SNP marker data, leads to improved genomic predictions compared to a model ignoring breed origin. The study of the population structure incorporated 1,730 genotyped Jersey animals. In total 39,542 SNP markers were included in the analysis. The 1,079 genotyped bulls with de-regressed proof for udder health were used in the analysis for the predictions of the genomic breeding values. A range of random regressions models that included the breed origin were analyzed and compared to a basic genomic model that assumes a homogeneous breed structure. The main finding in this study is that the importation of germplasm from the USJ population is readily reflected in the genomes of modern DJ animals. First, linkage disequilibrium in the group of admixed DJ animals is lower compared to the groups of the original DNK and USJ animals. Second, persistence of linkage disequilibrium phase is not conserved for longer marker distances between animals with mainly Danish or US origin. Third, the STRUCTURE analysis could retrieve genomic based breed proportions in alignment to the pedigree based breed proportions. However, including this population structure in a random regression prediction model, did not clearly improve the reliabilities of the genomic predictions compared to a basic genomic model.

Key words: cattle, population structure, genomic predictions

INTRODUCTION

Genomic predictions are more accurate than traditional pedigree based evaluations (Meuwissen et al., 2001), when the evaluation of the animal does not includes progeny information. The important factors influencing reliabilities of genomic predictions (Hayes et al., 2009) are the number of animals with phenotypes in the genotyped reference population, the reliability of the phenotypes that are used to predict the SNP marker effects, and the level of linkage disequilibrium (**LD**) between SNP markers and QTL. The size of the reference population and phenotypic recording strategies can be changed through changes in the breeding strategy. In contrast, the level of LD at a given distance between loci is a function of the population history (Hayes et al., 2009).

Danish Jersey (**DJ**) is an example of an admixed breed consisting of animals with diverse origin. It includes animals with different breed proportions of original Danish Jersey (**DNK**) and US Jersey (**USJ**). The two populations share common ancestry; even though they have been separated for almost a century. However, due to the long separation, phase associations between marker and QTL alleles may differ depending on the origin of the chromosome segments, thus reducing LD across the two substructures in DJ. These effects may explain why predicted reliabilities in DJ are lower than expected given the number and quality of phenotypes in the reference population (Thomasen et al., 2012).

The main objective of this study is to investigate whether the historic pedigree population structure in DJ is reflected in its genomic structure. This is investigated by LD measures and persistency of marker phase within and between subgroups of animals with high and low proportions of DNK (vs. USJ) in their pedigree. In addition, it is investigated whether explicitly accounting for the population structure by genome wide grouping of animals improves the genomic predictions. A random regression linear model is used to model the population structure in the genomic predictions and compared to a traditional pedigree model and a genomic model that assumes a homogeneous breed structure.

MATERIALS AND METHODS

Genetic data

All animals included in this study were genotyped using the Illumina Bovine SNP50 BeadChip (Matukumalli et al., 2009). SNP typing was done in part at the Department of Molecular Biology and Genetics, Aarhus University and in part at Genoscan A/S (Tjele, Denmark). The genotypic data was edited by individual typing for each animal. For animals the requirements was a call rate above 95% except for 10 animals (0.6% of total), which were accepted with call rates of at least 85%. Marker loci were accepted if they had a call rate of at least 95%. Loci with a minor allele frequency (**MAF**) less than 1 % were excluded. Loci without a map position in the UMD 3.1 assembly (Zimin et al., 2009) were discarded. Animals with an average GenCall score of less than 0.65 were discarded. Individual marker genotypings with a GenCall score of less than 0.6 were discarded.

After marker data quality checking 39,542 informative SNP markers were available for in total 1730 Jersey animals born between 1983 and 2010. Due to missing DNA samples only 18 bulls born in the years 1983 to 1987 were included. The dataset consisted of 1480 genotyped Jersey sires and 250 genotyped cows. The fastPHASE software (Scheet and Stephens, 2006) was used to impute sporadic missing genotypes and to phase the SNP markers for the Jersey animals.

Breed Origin of Animals

For the study of the LD structures two subgroups were performed based on the breed proportions traced by the pedigree. Animals with at least 75% Danish origin in their pedigree were assigned to the DNK group consisting of 171 animals. Animals with at least 75% USJ origin in pedigree were assigned to the USJ group consisting of 131 animals. In this study 52 original USJ sires and 27 original DNK sires are included. Breed proportions of Canadian Jersey were included in USJ. Due to import of semen from a few New Zealand bulls in the 1960s many of the genotyped Jersey animals have a small part of their pedigree originated from the New Zealand Jersey population. The breed proportions with this origin were not used in this study.

The average proportion of original DNK varies from 0.25 for bulls born in 1985 up to 0.75 for the bulls born in 1988. The proportion of DNK has stabilized around 0.60 for bulls born after 2002. The standard deviation in breed proportion of DNK has been reduced over the years showing that the animals are continuously admixed.

Estimation of LD and Persistence of Phase

The r^2 was used as the measure for the LD (Lewontin, 1964) as it seems to be the most robust measure for the LD (Qanbari et al., 2010). The r^2 was calculated as the squared correlations between markers that were grouped into bins by lengths of 7 kb.

Persistence of marker phase was calculated as the Pearson correlation ρ of r across the groups DJ, USJ and DNK, where r is the correlation between markers within bin and subpopulation of a defined marker interval. The persistence of marker phase across the groups was calculated within bins of 25 kb in marker distances (de Roos et al., 2008). The larger marker distance for calculations of the r^2 was used in order to reduce stochastic sampling variance of the calculated ρ. The minimum number that was included in the calculation of ρ was 10598 in one bin for the persistence between USJ and DNK.

Historic Population Size

The present effective population size (***Ne***) is a result of its historic population evolution. Traditionally estimated Ne is based on the pedigree (Nomura, 1996). However, with the availability of dense markers both present and historic ***Ne*** can be estimated from LD information.

The historic population size ($\boldsymbol{N_T}$), T generations ago, for the DJ population was inferred from the calculated r^2 over a certain interval between loci. The expected r^2 (Sved, 1971) can for small values of marker intervals c (in Morgans) and in the absence of mutations be calculated as: $r^2 = \frac{1}{4N_e c+1}$, where N_e is the effective population size and c is the linkage map distance in Morgans. By reformulating this formula the expected value of $\boldsymbol{N_T}$ can be calculated as:

$$E(N_T) \approx \frac{\frac{1}{r^2}-1}{4c},$$

where $T = 1/2c$ (Hayes et al., 2003). Using the calculated r^2 for each bin of 7 kb marker distances $E(N_T)$ were calculated within each bin.

Measures of Breed Origin

The traditional genomic prediction model assumes that QTL effects are in perfect LD with the surrounding SNPs. However as hypothesized in this paper the LD may be reduced due to the admixture structure in DJ and hence lower the reliability of the genomic predictions if this substructure is not accounted for in the model.

We used two measures of breed origin were used in genomic prediction model to account explicit for the population structure 1) the pedigree based breed proportion and 2) a marker based breed proportion. Marker information may give a more reliable estimate of the population structure if the pedigree data is not complete (Flury et al., 2010). Therefore breed proportions were also inferred from the marker information using the software package STRUCTURE (Pritchard et al., 2000). The linkage model that uses the positions of the individual markers, were chosen to infer the breed proportions (Falush et al., 2003). The model accounts for both 1) LD due to mixture of the subpopulations which arise from variation in pedigree and leads to correlation among markers even if they are unlinked and 2) LD due to admixture that can be recognized from each of the ancestral populations. However, the linkage model ignores the common background LD that persists across the subpopulations. Therefore this model is best suited to use loosely linked markers (Falush et al., 2003). A marker spacing of around one cM was therefore used. This was achieved by choosing one out of every 100 SNP markers without any other restrictions. In total, 412 evenly spaced markers were chosen across all 29 autosomes. In the STRUCTURE analysis two subpopulations were assumed (k=2). Based on our knowledge of the history in the breed, using k=2 is a strong prior. We also tested k=3, due to the fact that a small proportion of the present DJ animals originate from the New Zealand Jersey population. However, it was not possible to infer this subpopulation as a third breed in the STRUCTURE analysis. No prior pedigree information about the breed origin was used to infer the marker based breed proportion. For burn-in the first 10,000 iterations of the Gibbs sampler were discarded. After burn-in the MCMC chain was iterated 20,000 times. This is the standard setting in the program, which was found to be sufficient for the present dataset. In a study by (Frkonja et al., 2012) including only 10,000 MCMC chains were performed.

Dependent Variable

De-regressed proofs (DRP) for udder health were used as response variable for the prediction of genomic breeding values. Official EBV from January 2011 was used for the calculations of DRP for udder health (Danish Cattle Federation, 2012). This trait was chosen because there is a relatively large difference in genetic level for this trait between the USJ and DNK groups (Danish Cattle Federation, 2012), indicating that there have been different breeding goals in the two populations. The EBVs are corrected for dominance effects due to total heterosis and recombination loss, but include differences in genetic levels between original USJ and DNK animals (Negussie et al., 2010). In total 1,079 genotyped animals had records for DRP. Calculations of the DRP followed the procedures described by Strandén and Mäntysaari (2010). The DRP for udder health varied between 60.10 and 126.7 with a variance of 118.2. Reliabilities ranged from 47% to 99%.

Genomic Predictions

The 1,079 bulls with DRP for udder health were split into a reference population consisting of 879 bulls and a test population consisting of the 200 youngest bulls. Udder health was chosen because it has low heritabililty, with a high economic value in the breeding goal. Use of marker information for increasing reliability of prediction is therefore particularly beneficial. Variance components were estimated using data from the reference population using AI-REML. Next, genomic breeding values were predicted using phenotypes and genotypes from the reference population and genotypes from the test population. The DMU software package (Madsen and Jensen, 2010) was used for the estimation of the variance components and for prediction of the genomic breeding values.

The statistical model used for the prediction of genomic breeding values explicit accounting for the population structure was:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{w}\beta + \mathbf{Z}\mathbf{a}_0 + \mathbf{W}\mathbf{a}_1 + \mathbf{e} \qquad \textbf{(1)}$$

Where $\mathbf{y}$ is the vector of DRP for udder health, $\mathbf{1}$ is a vector of ones, μ is the intercept, $\mathbf{w}$ is a vector with element w_i as the proportion of Danish origin for animal i from either pedigree or STRUCTURE, β is the fixed regression of DRP on breed proportion, the i'th element in $\mathbf{a_0}$ is the random animal intercept for animal i, the i'th

element in $\mathbf{a_1}$ is the random regression on breed proportion for animal *i*, **Z** is an incidence matrix associating $\mathbf{a}_o$ with **y**, **W** is a matrix with one nonzero element in each row presenting the proportion of Danish origin for corresponding animal, and **e** is the vector of residuals. It is assumed that $\mathbf{e} \sim N(0, \mathbf{D}\sigma_e^2)$ where **D** is a diagonal matrix having the diagonal elements $d_i = \frac{1-REL\_DRP_i}{REL\_DRP_i}$ where REL_DRP_i is the reliability of DRP for animal *i* and σ_e^2 is the residual variance. We used this procedure in order to correct for heterogeneous residual variance due to different reliabilities of DRP.

The random animal effects $\mathbf{a_0}$ **a**nd $\mathbf{a_1}$ are assumed to follow the distribution:

$$\begin{pmatrix} \mathbf{a_0} \\ \mathbf{a_1} \end{pmatrix} \sim \mathbf{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\sigma}^2_{\mathbf{a0}} & \boldsymbol{\sigma}_{\mathbf{a0a1}} \\ \boldsymbol{\sigma}_{\mathbf{a0a1}} & \boldsymbol{\sigma}^2_{\mathbf{a1}} \end{bmatrix} \otimes \mathbf{G}^*\right)$$

where $\mathbf{G}^*$ is the weighted genomic and pedigree relationship matrix (Gao et al., 2012). A weight of 0.05 on the pedigree relationship matrix is used. Different weights of the pedigree relationship matrix were tested, but were not found to influence the predictive ability.

The direct genomic breeding value (**DGV**) for animal *i* is calculated as

$$DGV_i = \beta w_i + a_{0i} + w_i a_{1i} \quad (2)$$

Ignoring the population structure equation (1) can be reduced to

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{a_0} + \mathbf{e} \quad (3)$$

Formula (3) represents the basic genomic model that assumes a homogeneous breed structure. The DGV_i for animal *i* is now reduced to:

$$DGV_i = a_{0i} \quad (4)$$

Finally, a prediction model equivalent to (3) but using the pedigree-based relationship was used to predict breeding values, in order to evaluate the advantage of genomic prediction.

Reliability of DGV

Reliability of genomic predictions has in several studies been assessed as the validation reliability, r_v^2 (Su et al, 2010; Thomasen et al., 2012; Lund et al., 2011). The validation reliability of the DGV was calculating as the squared correlation between DGV and DRP divided by the mean reliability of DRP for the bulls in the test data set, which was in line with the definition of reliability of estimated breeding

value, since $r_v^2 \approx \frac{Cov^2(DGV,DRP)}{\sigma^2_{DGV}\sigma^2_{DRP}REL_{DRP}} \approx \frac{Cov^2(DGV,a)}{\sigma^2_{DGV}\sigma^2_a}$, where a was the true breeding value (Su et al., 2012). The predictive ability for the random regression model (1) was compared to the basic genomic model (3).

RESULTS

The Level of LD

The level of r^2 in the admixed group of animals (DJ), as shown in Figure 1, is marginally lower than in the two subgroups (DNK and USJ). For a marker spacing of 100 kb the level of LD is 0.02 higher within the two subgroups compared to the admixed group including all animals. This difference in r^2 is constant up to a distance of 700 kb. The level of r^2 decreases with increasing marker distances and decrease to a level of 0.08 for all animals included in the calculations (DJ).

Persistence of Phase

The highest levels of phase persistence were observed between the DJ and DNK groups, which for the marker spacing interval 100-200 kb has a level of 0.95 (figure 2). Phase persistence between USJ and DJ was slightly lower, decreasing to a level of 0.91 at marker space of 100-200 kb. Lowest persistence is seen between the DNK and the USJ group with a value of 0.78 for the marker spacing interval 100-200 kb.

Historic Population Size

Historic effective population sizes estimated from LD data up to 20 generations in the past are shown in figure 3. The effective population size shows a minimum 4 to 6 generations ago with a level of 135. This is the time period where the import of USJ germplasm started to play a major role in the breeding program for DJ (P. G. Larson, VikingGenetics, Randers, Denmark, personal communication). As a consequence a small increase in the effective population size was seen 3 to 4 generations in the past.

From generation 20 and down to generation 6 in the past the estimated effective population size decreases approximately linearly from 190 to 135.

Population Structure

A comparison of the subpopulation structure either inferred from STRUCTURE or traced by the pedigree is shown in figure 4. Overall the comparison between the two measures shows a correlation of 0.81 with a regression of 1.01. This correlation is obtained without attaching any predefined breed origin to the animals defined as either USJ or DNK from the pedigree. For a given proportion of DNK origin expected from the pedigree the figure visualize the variation in proportion of DNK origin inferred from the markers. Animals that are estimated to be purebred DNK (proportion =1) based on the pedigree, were seen to have inferred breed proportion varying between 0.71 and 1, except for one animal. This indicates that some of purebred DNK are in fact admixed with USJ or they still carry haplotypes inherited from the shared founder population. The STRUCTURE analysis with the SNP markers was able to identify the admixture in the genome. The animals that are estimated to be purebred USJ (proportion=0) have breed proportion varying between 0.05 and 0.20. The USJ animals are on average seven years younger and therefore more distant from the common founder population.

Genomic Predictions

Table 1 shows the reliabilities of the genomic predictions for udder health using different prediction models. The highest reliability (33.7%) for the test bulls is obtained for the genomic model (model f) that use the pedigree derived breed proportion as covariate for both fixed and random regression. In this model the covariance between the random animal effects (a_0 and a_1) was restricted to 0. This model made only slightly (0.08 %) better predictions than the basic genomic model (model d) without information on population structure. In general, there are only small differences (0.9%) in predictive ability between all the genomic prediction models. The basic animal model (model h) that uses the traditional pedigree-based relationship matrix as covariance structure has a predictive ability of 24.2 %. Thus, the model using marker information (model d) increases predictive ability by 9.5%.

The three models (e, f and g) that use pedigree derived breed proportions as population structure information on average perform 0.30% better than marker-derived breed proportions. However, the correlation between the pedigree and marker derived breed proportions were relatively high (0.81), indicating that marker inferred breed proportions is consistent with the breed proportions assigned to pedigree.

DISCUSSION

Population Structure

The main finding in this study is that the importation of germplasm from the US Jersey population is readily reflected in the genomes of modern DJ animals. First, r^2 in the group of admixed DJ animals are lower compared to the groups of the original DNK and USJ animals. Second, persistence of LD phase is not conserved for longer marker distances (>40kb) between DNK and USJ animals. As we only have a relative few purebred original USJ and DNK animals included in the present study, the differences in r^2 and persistence of phase between purebred USJ and purebred DNK animals will be even higher than found in this study. Third, the STRUCTURE analysis could retrieve genomic based breed proportions in alignment to the pedigree based breed proportions.

In general the pattern of the r^2 for marker space differences up to 100kb observed in the present study follows the results obtained in other genetic structure analysis of the cattle breeds (de Roos et al., 2008; Gautier et al., 2007; McKay et al., 2007; Pryce et al., 2010). For larger SNP marker spacing (up to 1 Mb) the r^2 in the DJ population are higher than r^2 found in the Australian (Pryce et al., 2010) and New Zealand (de Roos et al., 2008) Jersey populations. These differences might reflect that the recent historic population development in the DJ has been more closed and selection more intensive, leading to a smaller effective population size.

The pattern of persistence of phase and levels of LD reflect the history of the USJ and DNK populations, as we know that both subpopulations originate from the Channel Island of Jersey; more than 20 generations in the past. DJ originates from a single import of 5200 animals from the Channel Island of Jersey around year 1900. The animals that founded the USJ population were imported to North America

around year 1860. DJ was until the 1960s developed as a closed breed without further import of animals or semen. From this period the import of USJ semen has started (P. G. Larson, VikingGenetics, Randers, Denmark, personal communication).

The STRUCTURE analysis gave inferred breed proportions in alignment with pedigree based breed proportions and a regression coefficient between the two measures close to one. The correlation of 0.81 is a bit lower than obtained by Frkonja et al., (2012) using the same number of markers (0.93). However, Frkonja et al., (2012) predefined the breed for the purebred animals, which increases the correlation. The two analyzed breeds in their study (Simmental and Red Holstein) have been separated for many more generation than to the two Jersey populations in the present study have. The analysis also supported our prior expectation that DJ consists of two major subpopulations. We used the linkage model in STRUCTURE, which accounts for LD due to the mixture of populations and LD due to admixture (Falush et al., 2003). Background LD, however, is not accounted for. Therefore, we chose a set of loosely linked markers as suggested by (Falush et al., 2003). Increasing the number of markers will increase the amount of background LD in the data, which can lead to misleading results. The relatively large variation in inferred breed proportion obtained in STRUCTURE for a given value of breed proportion from pedigree may reflect that the separation of the two subpopulations has been fairly short. Therefore, some genomic segments cannot be associated with a particular subpopulation.

Effective Population Size

The estimated current effective population size in DJ of 135 based on the markers is in alignment with the marker based estimate found in the New Zealand Jersey population (de Ross et al., 2008), which also was estimated to 135. Sørensen et al. (2005) calculated an effective population size of 53 in DJ based on the rate of inbreeding calculated from the pedigree. The estimate using pedigree information is highly dependent on mating decisions, as it measures the correlation of uniting gametes. The marker based population size is in contrast a function of associations of alleles in gametes within and between animals. The increase in effective population size 3 to 5 generations in the past indicates that the import of USJ semen has introduced new combinations of marker alleles in the DJ population. This might be

explained by a differentiation in the breeding goals between the two subpopulations and genetic drift within the populations.

Using Population Structure in Genomic Predictions

The genomic prediction model that has been validated in the DJ population (Thomasen et al., 2012) assumes a homogenous population structure in DJ. This assumption implies that LD between markers and QTL persists for all animals. However, we have shown in this study that persistence of marker phase is not unique within the DJ breed for marker distances above 10 kb, when information about the subpopulation structure is studied. We therefore hypothesized that differences in marker allele effects exits due to difference in origin of alleles (DNK versus USJ) and that use of the genetic structure either based on the information from the markers or the pedigree can improve the prediction of the genomic breeding values.

Several studies exploit merging of reference populations, as the size of the reference population has been shown to affect the reliabilities of the genomic predictions considerably (Goddard and Hayes, 2009). Merging the largest European Holstein reference populations improved the reliabilities considerably (Lund et al., 2011), whereas combination of the Nordic Red populations only resulted in minor improvement of the reliabilities (Brøndum et al., 2011). A further extension of the Nordic Red reference population with the Norwegian Red even slightly decreased the reliabilities for some fertility traits (Heringstad et al., 2011). The Nordic Red is a group of 3 cattle populations, which again includes contributions from several breeds. Persistence of phase varied from 0.80 and up to 0.94 for marker intervals up to 500 kb (Rius-Vilarrasa et al., 2011). This indication of slightly different associations might be the explanation for the small improvement, when the Nordic Red populations are combined into one reference population.

Prediction Models Including Population Structure

The overall conclusion from comparing different genomic prediction models is that including population structure in a random regression prediction model, did not clearly improve the reliabilities of the genomic predictions compared to a basic genomic model. There can be several explanations for the lack of improvement. First,

the population structure might already be modeled well by the genomic relationship matrix. In a simulation study for admixed and crossbred populations (Toosi et al., 2010) showed that the GS prediction equations performed well as long as all purebreds that contributed to the population were included in the reference set, even without origin of breed had been taken explicitly into account. Second, the population structure information was based on average marker information measure across the entire genome. This might be a too naive measure to model the differences in marker allele effects, as admixed animals consist of a mosaic of haplotypes originating from purebreds DNK and USJ animals. Therefore, more detailed random regression models should be set up, which allows the breed origin of markers within an individual to vary over the genome. For instance a specific marker allele might be linked to a positive udder health effect, when it originates from a DNK animal but linked to a negative effect, when it has a USJ origin. However, due to high overall positive correlations between SNP effects in the two breeds of origin, it was in the present study not possible to separate these effects, using average genome wide population structure information.

Genomic Predictions Across Populations

DJ is a relative small population with only 1000 bulls with phenotypic information. As reliabilities increases with the size of the reference populations, use of phenotypic information from other populations may be a strong tool to increase the reliabilities for the genomic predictions. Genomic predictions across populations were evaluated in a simulation study by (de Roos et al., 2009). The conclusion from their study was that reliabilities of genomic predictions could be improved if reference data from all populations were included in the reference set. However, it requires that LD phase persists across populations. If the persistence between populations was low, reliabilities might even be reduced (de Roos et al., 2009). Investigation of persistence of marker phases between Jersey and the other Nordic dairy cattle populations showed that the persistence of marker phases is low. Results from that study are not shown in this paper. Adding reference data from either Holstein or Nordic Red is therefore not expected to increase reliabilities. It will at least require a high density marker genotyping of all the Nordic reference

populations and a more sophisticated modeling of population structure, before the benefits from combination of the Nordic reference populations can be realized.

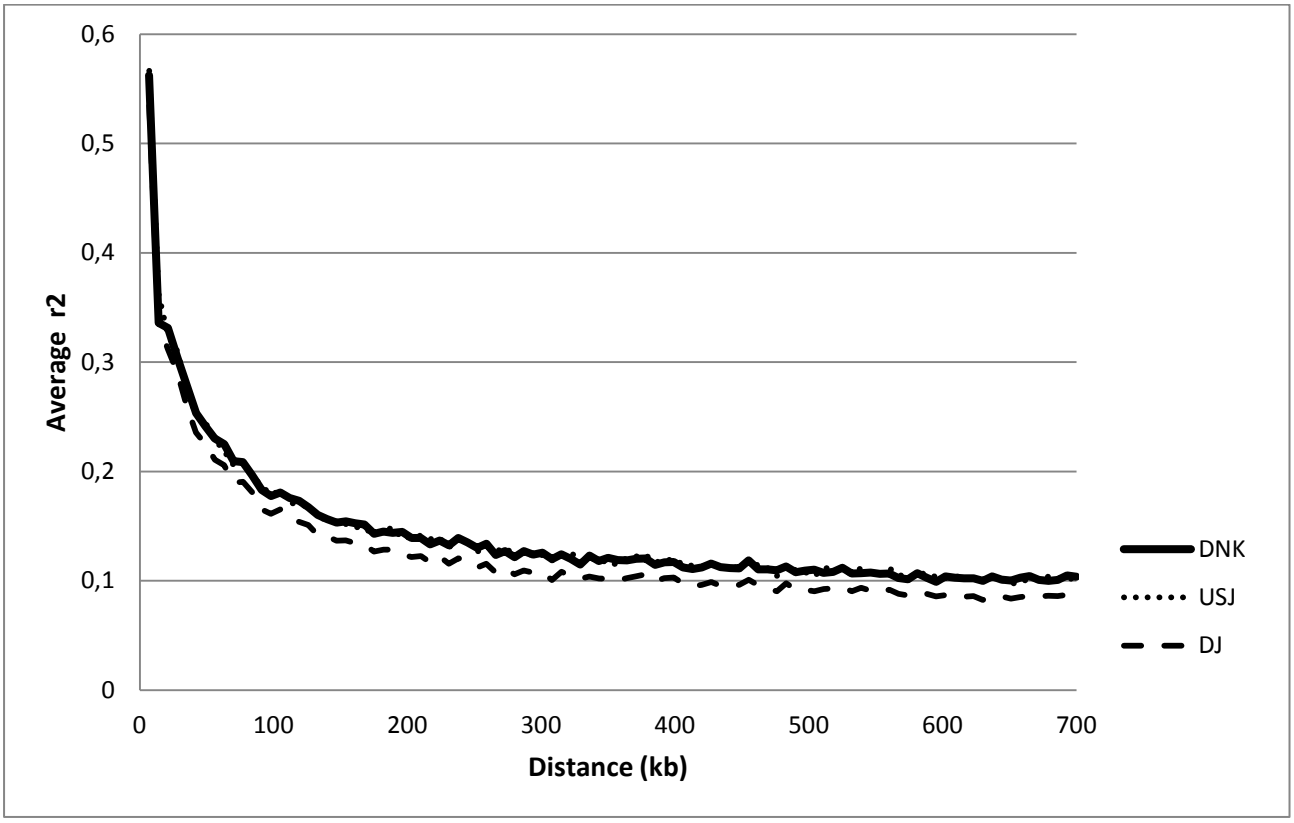

Figure 1: Linkage disequilibrium measured as r^2 between markers in Danish Jersey (DJ), original Danish Jersey (DNK) and original US Jersey (USJ). The curve for USJ is largely hidden behind the curve for DNK.

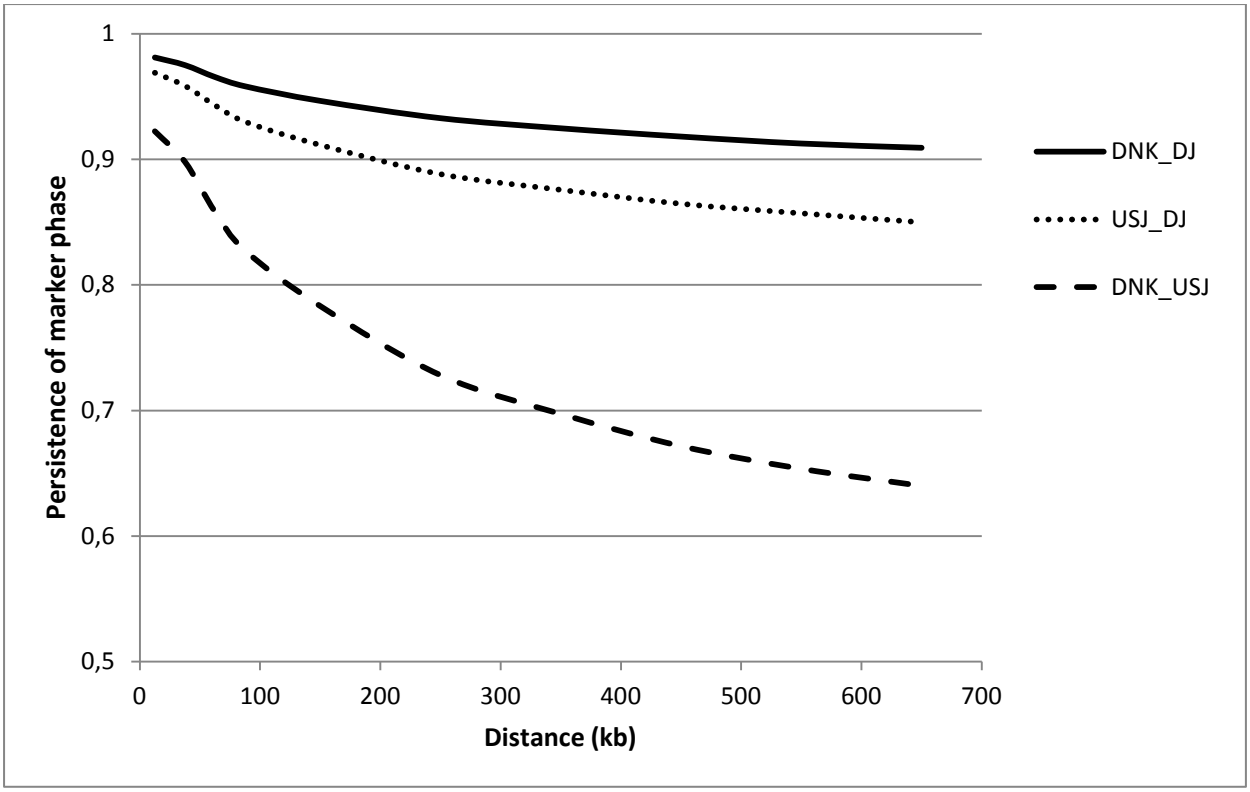

Figure 2: Persistence of marker phase as a function of increasing marker distance between the groups of Jersey animals: Danish Jersey (DJ), original Danish Jersey (DNK) and original US Jersey (USJ). Note that the y-axis starts at 0.5.

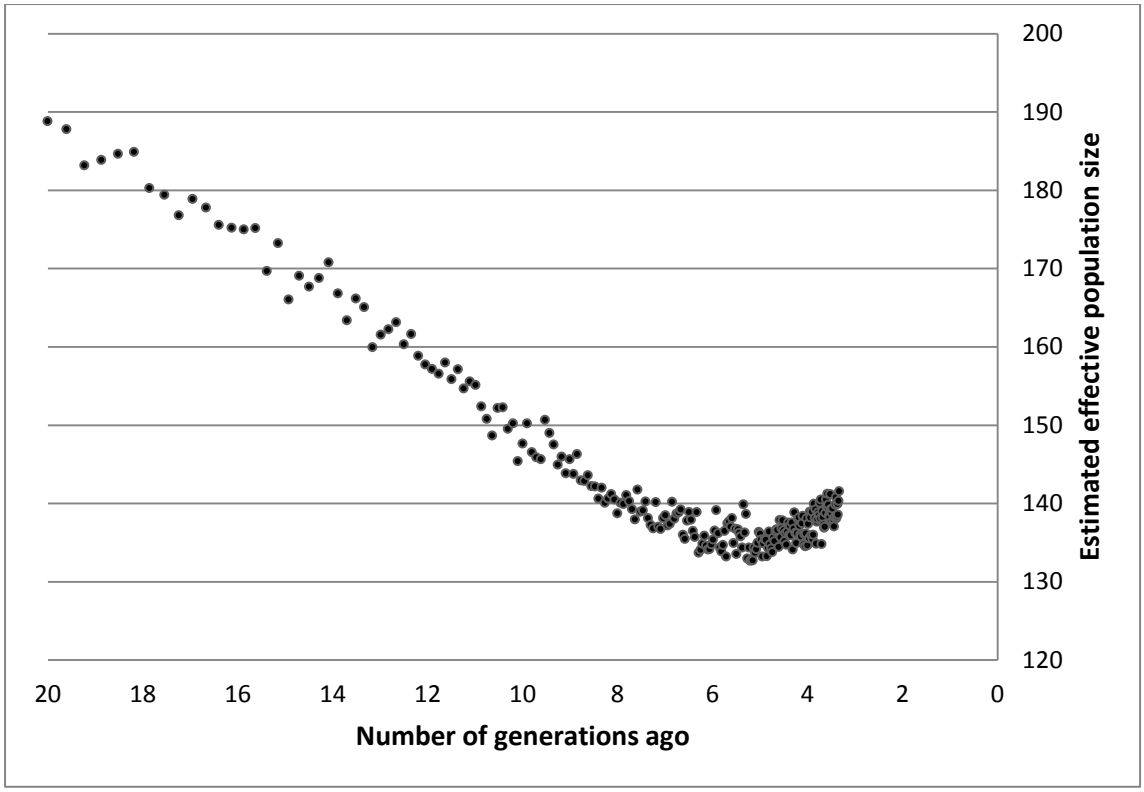

Figure 3: Estimated historic effective population size from linkage disequilibrium data. Calculations of historic effective population size are shown up to 20 generations in the past.

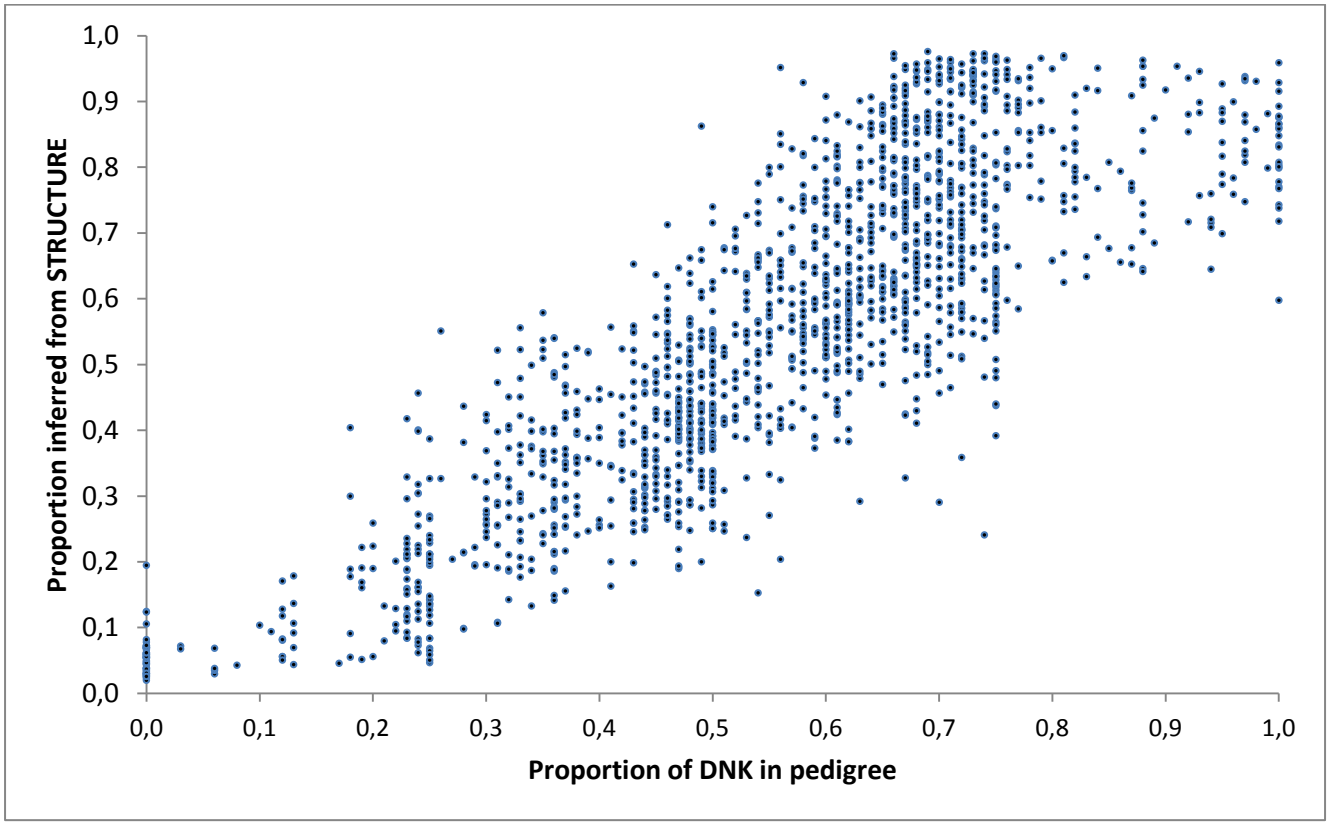

Figure 4: Proportion of original Danish Jersey (DNK) in pedigree plotted against proportion inferred from STRUCTURE for the 1730 genotyped animals.

Table 1: Prediction ability of different prediction models for udder health in Danish Jersey. Population structure information is either obtained from the markers or from the traditional pedigree. The information for the covariance structure of animals in the model is derived from either genomic or traditional pedigree information.

Model	*Population structure information*	*Covariance structure*	*Prediction ability (%) of predictions for test bulls*
(**a**) $\mathbf{y} = \mathbf{1}\mu + \mathbf{W}\beta + \mathbf{Za_0} + \mathbf{Wa_1} + \mathbf{e}$	markers	genomic	33.11
(**b**) $\mathbf{y} = \mathbf{1}\mu + \mathbf{W}\beta + \mathbf{Za_0} + \mathbf{Wa_1} + \mathbf{e}*$	markers	genomic	33.11
(**c**) $\mathbf{y} = \mathbf{1}\mu + \mathbf{W}\beta + \mathbf{Za_0} + \mathbf{e}$	markers	genomic	32.85
(**d**) $\mathbf{y} = \mathbf{1}\mu + \mathbf{Za_0} + \mathbf{e}$	none	genomic	33.64
(**e**) $\mathbf{y} = \mathbf{1}\mu + \mathbf{W}\beta + \mathbf{Za_0} + \mathbf{Wa_1} + \mathbf{e}$	pedigree	genomic	33.43
(**f**) $\mathbf{y} = \mathbf{1}\mu + \mathbf{W}\beta + \mathbf{Za_0} + \mathbf{Wa_1} + \mathbf{e}*$	pedigree	genomic	33.72
(**g**) $\mathbf{y} = \mathbf{1}\mu + \mathbf{W}\beta + \mathbf{Za_0} + \mathbf{e}$	pedigree	genomic	32.80
(**h**) $\mathbf{y} = \mathbf{1}\mu + \mathbf{Za_0} + \mathbf{e}$	none	pedigree	24.19

(*): In model (b) and (f) σ_{a0a1} was set to 0

LITERATURE CITED

Brøndum, R. F., E. Rius-Vilarrasa, I. Strandén, G. Su, B. Guldbrandtsen, W. F. Fikse, and M. S. Lund. 2011. Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. J. Dairy Sci. 94:4700-4707.

Danish Cattle Federation. 2012. Accessed December 27, 2012. In danish. http://www.landbrugsinfo.dk/kvaeg/avl/avlsanalyser/sider/startside.aspx.

de Roos, A. P. W., B. J. Hayes, and M. E. Goddard. 2009. Reliability of genomic predictions across multiple populations. Genetics 183:1545-1553.

de Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus Cattle. Genetics 179:1503-1512.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567-1587.

Flury, C., M. Tapio, T. Sonstegard, C. Drogemuller, T. Leeb, H. Simianer, O. Hanotte, and S. Rieder. 2010. Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. J. Anim. Br. Genet. 127:339-347.

Frkonja, A., B. Gredler, U. Schnyder, I. Curik, and J. Sölkner. 2012. Prediction of breed composition in an admixed cattle population. Anim. Genet. 43:696-703.

Gao, H., O. Christensen, P. Madsen, U. Nielsen, Y. Zhang, M. Lund, and G. Su. 2012. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. Genet. Sel. Evol. 44:8.

Gautier, M., T. Faraut, K. Moazami-Goudarzi, V. Navratil, M. Foglio, C. Grohs, A. Boland, J. G. Garnier, D. Boichard, G. M. Lathrop, I. G. Gut, and A. Eggen.

2007. Genetic and haplotypic structure in 14 European and African cattle breeds. Genetics 177:1059-1070.

Goddard, M. E. and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10:381-391.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92:433-443.

Hayes, B. J., P. M. Visscher, H. C. McPartlan, and M. E. Goddard. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Gen. Res. 13:635-643.

Heringstad, B., G. Su, T. R. Solberg, B. Guldbrandtsen, M. Svendsen, and M. S. Lund. 2011. Genomic predictions based on a joint reference population for Scandinavian red breeds. Book of Abstracts of the 62nd Annual Meeting of the European Federation of Animal Science, Stavanger, Norway. Wageningen Acedemic Publishers, Wageningen, the Netherlands.

Lewontin, R. C. 1964. The interaction of selection and linkage. I. General considerations; Heterotic models. Genetics 49:49-67.

Lund, M., A. de Roos, A. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Gen. Sel. Evol. 43:43.

Madsen, P. and J. Jensen. 2010. An user's guide to DMU, Version 6, release 5.0. http://dmu.agrsci.dk/dmuv6_guide.5.0.pdf . Accessed December 27, 2012.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. L. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS One 4:e5350.

McKay, S., R. Schnabel, B. Murdoch, L. Matukumalli, J. Aerts, W. Coppieters, D. Crews, E. Neto, C. Gill, C. Gao, H. MANNEN, P. Stothard, Z. Wang, C. Van Tassell, J. Williams, J. Taylor, and S. Moore. 2007. Whole genome linkage disequilibrium maps in cattle. BMC Genetics 8:74.

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

Negussie, E., M. Lindauer, E. A. Mäntysaari, I. Strandén, J. Pösö, U. S. Nielsen, J. Johansson, J. Å. Eriksson, and G. P. Aamand. 2010. Combining Test Day SCS with Clinical Mastitis and Udder Type Traits: A Random Regression Model for Joint Genetic Evaluation of Udder Health in Denmark, Finland and Sweden. Interbull Bulletin No.42. Riga, Latvia.

Nomura, T. 1996. Effective size of selected populations with overlapping generations. J. Anim. Br. and Genet. 113:1-16.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Pryce, J. E., S. Bolormaa, A. J. Chamberlain, P. J. Bowman, K. Savin, M. E. Goddard, and B. J. Hayes. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J- Dairy Sci. 93:3331-3345.

Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. R. Sharifi, and H. Simianer. 2010. The pattern of linkage disequilibrium in German Holstein cattle. Anim. Genet. 41:346-356.

Rius-Vilarrasa, E., T. Iso-Touru, I. Strandén, N. Schulman, B. Guldbrandtsen, E. Strandberg, M. S. Lund, J. Vilkki, and W. F. Fikse. 2011. Characterization of linkage disequilibrium in a Danish, Swedish and Finish Red cattle popluation. Book of Abstracts of the 62nd Annual Meeting of the European Federation of Animal Science, Stavanger, Norway. Wageningen Acedemic Publishers, Wageningen, the Netherlands.

Scheet, P. and M. Stephens. 2006. A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. American J. Human Genet. 78:629-644.

Strandén, I. and E. A. Mäntysaari. 2010. A recipe for multiple trait deregression. Interbull Bulletin No.42. Riga, Latvia.

Su, G., B. Guldbrandtsen, V. R. Gregersen, and M. S. Lund. 2010. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. J. Dairy Sci. 93:1175-1183.

Su, G., P. Madsen, U. S. Nielsen, E. A. Mäntysaari, G. P. Aamand, O. F. Christensen, and M. S. Lund. 2012. Genomic prediction for Nordic Red Cattle using one-step and selection index blending. J. Dairy Sci. 95:909-917

Sved, J. A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor. Pop. Bio. 2:125-141.

Sørensen, A. C., M. K. Sørensen, and P. Berg. 2005. Inbreeding in Danish Dairy Cattle Breeds. J. Dairy Sci. 88:1865-1872.

Thomasen, J. R., B. Guldbrandtsen, G. Su, R. F. Brøndum, and M. S. Lund. 2012. Reliabilities of genomic estimated breeding values in Danish Jersey. Animal. 6:789-796.

Toosi, A., R. L. Fernando, and J. C. M. Dekkers. 2010. Genomic selection in admixed and crossbred populations. J. Anim. Sci. 88:32-46.

Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marcais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biology 10.

Paper III

The optimal genomic selection strategy in a small dairy cattle breeding program still involves progeny testing

J.R. Thomasen, C. Egger-Danner, A. Willam, B. Guldbrandtsen, M.S. Lund and A. C. Sørensen†

Manuscript submitted to Journal of Dairy Science

INTERPRETIVE SUMMARY

The optimal genomic selection strategy in a small dairy cattle breeding program still involves progeny testing. *By Thomasen et al.*

Small dairy cattle populations are challenged because of the low reliabilities of genomic predictions. We have demonstrated that low reliabilities of genomic predictions sets limitations for moving towards more genetic efficient breeding schemes with more intensive use of young bulls without progeny testing. Strong positive interaction effects between increased reliability of genomic predictions and more intensive use of young bulls exist. From an economic perspective a juvenile scheme is always advantageous. The main future focus area for the smaller dairy cattle breeds is to join forces that increase reliabilities of genomic predictions.

GENOMIC SELECTION IN SMALL DAIRY CATTLE BREEDS

The optimal genomic selection strategy in a small dairy cattle breeding program still involves progeny testing

J.R. Thomasen,*† C. Egger-Danner,‡ A. Willam,§ B. Guldbrandtsen,† M.S. Lund,† and A. C. Sørensen†

* VikingGenetics, DK 8860, Assentoft, Denmark

† Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Faculty of Science and Technology, Aarhus University, DK-8960, Tjele, Denmark

‡ ZuchtData EDV-Dienstleistungen GmbH, 1200 Vienna, Austria

§ Division of Livestock Sciences, Department of Suistainable Agricultural Systems, University of Natural Resources and Life Sciences Vienna (BOKU), 1180 Vienna, Austria

*Corresponding author: Jørn Rind Thomasen, VikingGenetics, DK 8860, Assentoft, Denmark. Phone: (45) 87 95 94 00. Fax: (45) 87 95 94 01.

Email: jotho@vikinggenetics.com

Abstract. The main objective of this study was to evaluate the value of running a genomic breeding scheme in a small dairy cattle population that is intermediate in terms of using both young bulls (YB) and progeny tested bulls (PB). This scheme was compared to a conventional progeny testing program without use of genomic information and, as the extreme case, a juvenile scheme with genomic information where all bulls are used before progeny information is available. The population structure, cost and breeding plan parameters were chosen to reflect the Danish Jersey cattle population mimicking a small dairy cattle population. The population consists of 68,000 registered cows. 1,500 bull dams were screened to produce the 500 genotyped bull calves from which 60 YBs were selected to be progeny tested. We included two traits in the breeding goal, a production trait (h^2=0.30) and a functional trait (h^2=0.04). An increase in reliability of 5 percentage points for each trait was used in the genomic scenario. A deterministic approach was used to model the different breeding programs, where the evaluation criteria were annual monetary genetic gain and discounted profit (DP). We investigated the effect of varying the parameters 1) increase in reliability due to genomic information, 2) number of genotyped bull calves, 3) proportion of bull dam sires that are young bulls, and 4) proportion of cow sires that are young bulls. We showed that the genomic breeding scheme is both genetic and economical superior to the conventional breeding scheme even in a small dairy cattle population where genomic information give a relatively low increase in reliability of breeding values. The optimal breeding scheme was characterized by a mixed use of YB and PB both as bull sires and for inseminations in the cow population. The results from this study also supported our hypothesis that strong interaction effects exist. The strongest interaction effects were obtained between increased reliabilities of genomic estimated breeding values (GEBV) and more intensive use of YB. The juvenile scheme was genetic inferior when the increase in reliability was low (5 percentage points), but becomes genetic superior as higher reliabilities of GEBV can be obtained. Using DP as evaluation criterion the juvenile scheme was always superior due to the lower generation interval and reduction of costs for housing and feeding waiting bulls to a minimum.

Key words: genomic selection, genomic breeding plans, dairy cattle, small population

INTRODUCTION

Combining information from pedigree and SNP markers leads to increased reliabilities of genomic estimated breeding values **(GEBV)** compared to parent average estimates (Hayes et al., 2009b). With an increased reliability, young bulls **(YB)** become more competitive relative to progeny tested bulls **(PB)** in populations with considerable genetic gain. Thus, several studies have shown that higher genetic gain can be achieved if young males without progeny performance are used as parents for the next generation (Buch et al., 2012; de Roos et al., 2011). In the pre-genomic era, conventional breeding schemes were characterized by long generation intervals. Now cattle breeding organizations move towards breeding schemes with more intensive use of YB as bull sires and for inseminations of cows, partly because of higher reliability based on genomic predictions and partly because of potentially reduced cost when keeping fewer waiting bulls. Furthermore, the cost of genotyping has decreased, which has made it even more feasible to speed up the amount of genotyping followed by increased selection intensity for YB selection pathway.

The number of genotyped reference bulls with daughter proofs (Goddard and Hayes, 2009) is the most important factor for determining reliabilities of genomic predictions for future selection candidates. However, for small dairy cattle populations there are only limited numbers of potential reference bulls available. Hence, the gain from using genomic information is low (Pryce et al., 2011). A gain in reliability of 4 percentage points was estimated in the Danish Jersey breed (Thomasen et al., 2012) based on a reference population of 1,000 bulls, whereas a gain of 20 percentage points was obtained in the Nordic Holstein breed (Lund et al., 2011) based on a larger European Holstein reference population consisting of nearly 20,000 bulls. Most of the simulation studies published so far focus on optimization of genomic breeding schemes in larger cattle populations (Pryce and Daetwyler, 2012). Since genomic information adds less to reliabilities in smaller populations, we expect that conclusions from studies for larger populations might not apply for smaller populations. E.g. the annual genetic gain is positively influenced by the interaction effect between the added value of genotypic information and a more intensive use of YB (Buch et al., 2012). Hence, a key design parameter for such a small breeding scheme is which proportion of YB is optimal in order to maximize annual genetic gain, as we expect PB to remain competitive with YB.

The main objective of this paper is to evaluate the impact of running a genomic selection breeding scheme compared to a conventional progeny testing program for a small dairy cattle population and to evaluate the value of increased genomic information originating either from higher reliability of genomic predictions or increasing the selection intensity of YB by genotyping more bull calves. A second objective is to analyze the effect of changing the proportion of genomically selected YB both as bull sires and as cow sires. We hypothesize that the added value of genomic information interacts positively with more intensive use of YB in the breeding scheme. The main evaluation criteria are annual monetary genetic gain and discounted profit. Furthermore, the balance of annual monetary genetic gain for a production and functional trait is presented.

MATERIALS AND METHODS

Experimental design

To test our main hypothesis three overall breeding designs are compared: 1) progeny testing scheme without use of genomic information **(conventional)**, 2) juvenile scheme with genomic information where bulls are used before progeny information is available **(turbo)**, and 3) scheme that is intermediate in terms of using both young and progeny tested bulls **(hybrid)**.

Our second objective is investigated by varying parameters effecting annual genetic gain in the hybrid scheme: 1) increase in reliability due to genomic information, 2) number of genotyped bull calves, 3) proportion of bull dams mated with YB and 4) proportion of cows mated with YB.

The population structure, parameters and breeding scheme is chosen to mimic practically feasible options for the Danish Jersey cattle population as an example of a small dairy cattle population.

Hybrid scheme

The hybrid breeding scheme reflects the current breeding scheme with use of genomic information as it is carried out in the Danish Jersey breed. The general structure of the hybrid breeding scheme is illustrated in Figure 1. The population

consists of 68,000 cows with records. The 1,500 cows with the highest estimated breeding value **(EBV)** according to the breeding goal are screened as bull dam candidates. It is assumed that these cows are inseminated with relevant bull sires to produce the 500 bull calves that will be genotyped. 60 YB are selected for progeny testing according to their GEBV. These YB are randomly used for 50% of the inseminations in the cow population. Eventually, each YB obtains 113 daughter records for the production trait and 104 daughter records for the functional trait, which are the current daughter group sizes in the Danish Jersey population. The 15 YB with the highest GEBV are selected as bull sires and mated to 25% of the bull dams. Finally, four PB are selected both for use as bull sires, contributing with 75% of the inseminations, and for inseminations in the cow population, contributing with 50% of the inseminations. These PB are available because a waiting bull system is run until their daughter proofs are available. For this breeding scheme four key parameters are varied:

Increase in reliability of GEBV: The value of genomic information is measured by the increase of reliability of genomic predictions above the reliability of the parent average. An increase in reliability of 5 percentage points is used in the reference scenario. This is the current effect of including the genomic information in Danish Jersey (Thomasen et al., 2012). The reliability is increased in steps up to the level of the reliability of a progeny tested bull, which is obtained by adding 40 percentage points to the reliability of the parent average. In this study the EBV for the PB only includes the information from daughter records but no genomic information

Number of genotyped young bulls: The number of genotyped bull calves is varied from 500 up to 2,000 in order to evaluate the effect of increasing the selection intensity of bull calves that go into progeny testing.

A price of 100 Euro is connected to each SNP-typing (Table 1). When increasing the number of genotyped bull calves, the number of bull dams is increased in order to have a fixed ratio of three bull dams per genotyped bull calf. In the reference scenario of the hybrid scheme 500 bull calves out of 1500 bull dams are genotyped yearly (Table 1), which is the current number of bull calves genotyped in the Danish Jersey breeding program.

The costs for obtaining a higher reliability is highly dependent of the information source and, therefore, difficult to predict. No costs are attached the increase in reliability.

The proportion of bull dams mated with YB is varied from 0 to 1. In the reference scenario a proportion of 0.25 is used. The remaining bull dams are inseminated with PB.

The proportion of cows mated with YB is determined by the farmers. The proportion is varied from 0.25 to 1.

Interactions: For testing interaction effects between the mentioned parameters all two-way combinations of the four parameters were investigated.

Turbo scheme

An extreme breeding scheme was studied. This scheme maximizes the use of genomic information. Only YB's are used as bull sires and as sires of dams. Thereby, the generation interval is minimized. It also reduces the cost of the breeding plan since YBs were slaughtered as soon as enough semen had been produced.

The conventional scheme

A conventional progeny testing breeding program without use of genomic information was investigated. We used the same breeding plan parameters as used in the Danish Jersey breeding scheme before genomic selection was introduced. This breeding program has a lower proportion of cows mated with YB (Table 1) compared to the hybrid scheme and an exclusive use of PB as sires of sons. In this breeding scheme, five PB were yearly selected compared to four in the hybrid breeding scheme. The numbers of YB and bull dams are the same as in the reference hybrid scheme.

Method and evaluation criteria

A deterministic approach is used to simulate and evaluate the value of the different breeding strategies. ZPLAN software (Willam et al., 2002) was used. It allows evaluation of selection strategies mainly based on the gene flow method (Hill, 1974) combined with a selection index procedure for predicting reliabilities. With ZPLAN it is possible to evaluate both the genetic and the economic consequences of the different breeding strategies for a given investment horizon. We used two main

criteria to compare the value of the different breeding strategies: 1) annual monetary genetic gain (**AMGG**), defined as the average increase per year in monetary superiority of the progeny of the selected animals after one round of selection; and 2) discounted profit (**DP**), defined as the discounted monetary profit based on the genetic superiority and expressed as the improved profit per animal in the total population over the given investment period. In this study we used an investment period of 15 years. Returns were discounted by six % per year, while costs were discounted by four % per year.

ZPLAN requires population-, biological- as well as cost parameters, which are given in table 1. The biological parameters are obtained partly from the official milk recording statistics (Lauritsen, 2012) and partly from an analysis of bull statistics for Danish Jersey in VikingGenetics (P. G. Larson, VikingGenetics, Randers, Denmark, *pers. comm.*). The included cost parameters exclusively reflect the variable costs related to the breeding program (Table 1) while fixed costs are being ignored. All results for AMGG and DP are expressed as relative values referring to the values of the reference scenario of the hybrid scheme, which are set to 100.

Breeding goal and traits

The breeding goal is a weighted sum of two traits. The first trait represents milk production traits (h^2=0.30) and the second trait functional traits (h^2=0.04). An unfavorable genetic correlation of -0.30 between the two traits was assumed, while the residuals were assumed uncorrelated. The economic values were set to 83 Euro for the milk production trait (**PT**) and 82 Euro for the functional trait (**FT**) per additive genetic standard deviation. These economic values ensure that the correlation between milk production and breeding goal is the same as in the Nordic total merit index (Buch et al., 2012). All animals are selected for the overall breeding goal.

Reliability of estimated breeding values

ZPLAN calculates the reliability of the index for each selection group separately. The index used is constructed as a selection index including different sources of

information as own performance, maternal and paternal half sibs, half sibs of sire and dam and progeny.

The reliability of genomic predictions originates both from linkage disequilibrium (LD) between SNP-markers and QTL as well as family relationship information (Habier et al., 2007). The LD source persists over generations, whilst the reliability due to family relationship is dependent on the candidates' relationship to the reference population.

The gain in reliability of GEBV for YB compared to parent average is modeled by adding the gain of each index trait. We assume that the gain in reliability only originates from LD information and is independent on family relationship information. The same gain in reliability is added to each trait as it is assumed that the gain in reliability due to genomic information is independent of the heritability of the trait (de Roos et al., 2011; Thomasen et al., 2012).

In a breeding scheme where YB are used as bull sires the reliability of GEBV is lower compared to a breeding scheme where PB are used as bull sires. We modeled this reduction in reliability of GEBV due to family relationship by removing the paternal half sib information for the proportion of YB that are used as bull sires.

The reliabilities of the indices were 29%, 67%, and 35% for the YB, PB and cows respectively in the conventional breeding scheme. In the turbo scheme the reliabilities of the YB were lowered to 21% as an effect of the bull sires not are included in the reference population with daughter group information. In the hybrid breeding scheme the reliability of the YB were 29%, which obviously is the same as in the conventional breeding scheme. However, this reliability is a result of adding 5 percentage points to parent average due to genomic selection and reducing the parental half sib information for the proportion (0.25) of the YB having an YB as bull sire.

RESULTS

Overall comparisons of breeding schemes

Table 2 shows the results for the comparison of the three main breeding schemes. The level of AMGG and DP for the hybrid scheme is set to 100. The hybrid scheme is 6.8% superior for AMGG compared to the conventional breeding scheme. This difference is mainly due to a shorter generation interval in the hybrid scheme

compared to the conventional scheme (4.14 years versus 3.58 years). This reduction in generation interval more than compensates for the lower reliability of YB compared to PB. The hybrid scheme is also superior for AMGG compared to turbo scheme (3.9%).

In contrast, among the three main breeding schemes the turbo scheme gives the highest DP. The turbo scheme results in a 49.7% higher DP than the hybrid scheme. The hybrid scheme has 11.3% higher DP than the conventional scheme. The FT has the highest contribution to AMGG in the turbo plan (36.2%), which is 3.4 percentage points higher than in the conventional scheme.

Effect of increasing reliability of genomic prediction

The value of increasing reliability of GEBV in YB compared to parent average is shown in Table 3. Increasing the reliability of GEBV from +5 to +10 percentage points increases AMGG just by 1.1%. Increasing the reliability up to +40 percentage points, which is equivalent to assuming that the genomic information yielded the same reliability as daughter proofs, increases AMGG by 7.5% and DP by 13%. The increase in DP only expresses the discounted return since no cost is attached to increased reliability.

Effect of increasing the number of genotyped young bulls

Doubling the number of genotyped YB from 500 to 1,000 increases AMGG with 3% (Table 4). The marginal effect on AMGG is lower (2.2 %) when the number of genotypings is increased from 1,000 to 2,000 bull calves. Increasing the number of genotyping's increases the cost of the breeding scheme. In addition to genotyping cost, more bull dams have to be inspected (Table 1). Taking this variable cost into account the DP increases by 4.4% by doubling the number of bull calves 500 to 1000.

Effect of alternative strategies using young bulls

The use of YB was varied independently around the default values both for inseminations of bull dams (Table 5) and inseminations of cows (Table 6). In the hybrid breeding scheme the default proportion of bull dams inseminated with YB´s

is 25%. This value results in the highest AMGG. A higher proportion of bull dams inseminated with YB´s give, in general, a marginally lower AMGG (down to -1.8%) but a higher DP (up to 6.5%). The proportion of AMGG that originates from PT is reduced from 66.3% to 65%.

Using YB only for insemination of cows results in the highest AMGG (+3.1%), highest DP (12.7%) and largest contribution of FT to AMGG (36.5%) given default values for the other parameters in the hybrid scheme.

Interaction effects of breeding scheme parameters

The strongest interaction effects are observed between increased reliability of GEBV and a more intensive use of YB for inseminating bull dams both for AMGG (Figure 2) and DP (Figure 3). With a gain in reliability for GEBV of +5 percentage points compared to parent average, the optimal proportion of bull dams inseminated with YB is 0.25. For higher levels of gain in reliability of GEBV the optimum shifts towards only using YB as bull sires. For a gain in reliability of +40 percentage points AMGG increases relatively with 16.3% by increasing the use of YB for inseminating bull dams from 25% to 100%. For all levels of gain in reliability the highest discounted profit is obtained with exclusively YB used as bull sires. In general, the strongest interaction effects are observed for all parameters related to increased gain in reliability of GEBV (results not shown). Furthermore, higher marginal gains in AMGG and DP are obtained by genotyping more bull calves, when YB are used more intensively in the population (results not shown). Similarly, the marginal value of extra genotypings is larger with larger gains in reliability of GEBV.

Optimal use of YB in the population with regard to AMGG depends on the use of YB as bull sires and *vice versa* (Figure 4). If farmers decide to exclusively use YB for inseminating cows, the highest AMGG is obtained for a breeding scheme using only PB as bull sires. The optimal breeding scheme shifts towards using a higher percentage of bull dams inseminated with YB, when fewer cows (<70%) are inseminated with YB. In a breeding scheme only using YB as bull sires, all levels of using YB in the population nearly resulted in the same level of AMGG (96.9 to 98.3). Using DP as the evaluation criterion, there is however no interaction between the use of YB in the population and the use of YB for inseminating bull dams. Exclusive use of YB in the population is superior for all levels of use of YBs as bull sires, and

exclusively use of YB as bull sires is superior for all levels of use of YB in the population (Figure 5).

DISCUSSION

In this study we can show that a genomic breeding scheme is superior to a conventional breeding scheme even in a small dairy cattle population where genomic information gives a relatively low increase in reliability. The optimal breeding scheme is characterized by a mixed use of YB and PB both for use as bull sires and for inseminations in the cow population. The results from this study also support our hypothesis that strong interaction effects exist between increased reliabilities of GEBV and a more intensive use of YB. As a result, the turbo scheme only becomes superior when higher reliabilities of GEBV can be obtained.

Optimal breeding scheme for a small dairy cattle population

All genomic breeding schemes tested yielded higher AMGG and DP compared to the conventional scheme. However, as we expected, the genetic superiority found in the present study is lower compared to other studies (Pryce and Daetwyler, 2012), which are based on larger population sizes and hence larger reference populations. In pre-selection schemes with reliabilities of GEBV around 60% the improvement in annual genetic gain was 12% (de Roos et al., 2011) to 16% (Pryce et al., 2010). Breeding schemes modeling the Norwegian Red cattle population consisting of 120,000 cows were simulated by Lillehammer et al. (2011). In their pre-selection scenario with 750 genotyped bull calves, 60 PB and reliabilities of GEBV of 46% the annual genetic gain was increased by 11% compared to the conventional breeding scheme. Compared to our study this is a doubling of superiority in annual genetic gain, but obtained for a reference population consisting of thrice the number of bulls.

We also found that the genomic turbo scheme is genetically inferior to the hybrid scheme with mixed use of YB and PB. This is in contrast to several studies, where turbo schemes provide consistently higher annual genetic gains (König et al., 2009; Pryce and Daetwyler, 2012). In these schemes reliabilities of selection of young bull candidates were higher than in our study. The main reason is that the YB in our study are selected with a lower reliability of GEBV, which is not compensated by the shortened generation interval in the turbo scheme. In the turbo scheme the sires of

genotyped bull calves have not yet obtained a progeny test. As we assume that a part of the reliability arise from family relationship information (Habier et al., 2007; Habier et al., 2010; Wientjes et al., 2012) a reduction of the value of this information source has a high impact on the reliability of GEBV of YB, when the value of the genomic information is low. However, with increased reliabilities of GEBV we also found that turbo schemes become genetically superior compared to the hybrid scheme.

The turbo scheme, however, is superior when evaluated by DP instead of AMGG. There are two main reasons for this. Firstly, YB are slaughtered as soon as enough semen doses are produced to supply the population. Housing and feeding the bulls in the waiting period is by far the most expensive part of the breeding scheme. In the turbo scheme this cost is reduced to a minimum. Secondly, AMGG is expressed earlier in the turbo scheme due to a shorter generation interval (from 3.54 to 2.94 years). The difference in DP between the different breeding schemes depends on the discounting rates used as well as the investment period. However, a sensitivity analysis shows that the turbo scheme remains a superiority of 42 % with a reduction of the interest rate for costs and return to 4% and 2% and an investment horizon extended to 20 years.

Only two studies incorporate a total evaluation of the economic value of genomic breeding schemes (König et al., 2009; Schaeffer, 2006). Both studies find substantially higher profit for the turbo scheme compared to a conventional breeding scheme. However, these results are obtained for scenarios where reliabilities of GEBV are assumed to be considerable higher than in the present study. In these studies they did not model the reduction in reliability due to less parental family information, which may explain a part of the difference in reliabilities. König et al. (2009) used in addition a much higher selection intensity in the YB selection path.

A consequence of moving from conventional to genomic breeding schemes was a shift in the composition of genetic progress towards a relatively larger contribution from the low heritability trait, as the relative gain in reliability for this trait was higher. This effect was more pronounced at higher gains in reliabilities and with more intensive use of genomic information in selection decisions, i.e. selection of bulls before progeny information is available.

Optimal use of young bulls for inseminating cows

The proportion of cows inseminated with YB is a decision made by the farmers. In our reference scenario of the hybrid scheme 50% of the cows are inseminated with YB, which reflect the actual use in the Danish Jersey population. We have shown that exclusive use of YB can increase AMGG with 3.1%. It requires, however, that farmers accept to use YB with a lower reliability compared to PB. This fact was questioned by König et al. (2009), who also recommended a breeding scheme with 50% use of PB for inseminations of cows, even though the scheme resulted in a lower AMGG compared to exclusive use of YB. In contrast, we find that similar AMGG is obtained, no matter which strategy the farmers has for using YB, when only PB are used as bull sires in the breeding scheme.

Exclusive use of YB is always superior and without interaction effects, when the evaluation criterion is DP. The reason is that discounting favors breeding schemes, where the return is realized as quickly as possible, i.e. when the genetic superiority is disseminated into the population as quickly as possible. If increased reliability of GEBV can be obtained, it will be more attractive for the farmers to use YB.

Increased value of genomic information in hybrid scheme

In the present study extra use of genomic information is expressed either through higher reliabilities of GEBV or by additional genotypings of bull calves. The value of increased reliability of GEBV in the YB path is limited, provided that all other breeding parameters are kept constant in the hybrid scheme, i.e. when PB are used more intensively than YB. Genotyping of 1,000 bull calves provides the same improvement in AMGG as an increase of 20 percentage points in reliability. Doubling the number of genotyped bull calves will be attractive as DP also increases compared to the reference scenario of the hybrid scheme. Hence, even with the current cost of genotyping the potential return can pay for the additional genotypings.

With more than 1,000 additional genotypings of bull calves we observe a diminishing return in terms of AMGG. This finding is supported by other studies (Henryon et al., 2012; Sørensen and Sørensen, 2009). However, the rate at which the return diminishes is smaller with increasing reliabilities of GEBV, which was also

found by Henryon et al. (2012). We also observe a diminishing return in terms of DP, which is, however, smaller than observed for AMGG (results not shown).

Perspectives for effective breeding schemes in small dairy cattle populations

Small dairy cattle populations are challenged due to low reliabilities of genomic predictions. In this study, we have demonstrated that a low reliability sets limitations for moving towards more efficient breeding schemes with more intensive use of YB. This also limits the opportunities to run a more cost effective breeding scheme with lower housing and feeding costs for waiting bulls. Such savings of cost could be used for genotyping more bull calves in order to increase selection intensity in the male selection path.

Therefore, the key focus for smaller dairy cattle breeds should be to increase reliabilities of GEBV. Collaboration exchanging SNP-marker information of internationally evaluated bulls has shown to be effective for the Holstein population (Lund et al., 2011). For a small dairy cattle breed like Danish Jersey with only 1,000 bulls in the reference population, the marginal effect might be even bigger, provided strong genetic links between the Jersey subpopulations exist and reliable EBV are available for all important traits in the breeding goal. The North American Jersey population is the most promising collaborator for the Danish Jersey (Thomasen et al., 2013). Another option is to include genotyped females with own records in the reference population (Egger-Danner et al., 2012; Mc Hugh et al., 2011). Using reference populations from the bigger Holstein populations has so far been unsuccessful for increasing reliabilities of genomic predictions for small dairy cattle populations (Erbe et al., 2012; Hayes et al., 2009a).

ACKNOWLEDGMENTS

Parts of the analysis in this study were carried out during stays of J. R. Thomasen at Zuchtdata, Austria and BOKU University in Vienna, Austria in spring 2012. This work was performed in the project "Genomic Selection – from function to efficient utilization in cattle breeding (grant no. 3412-08-02253)", funded by the Danish Directorate for Food, Fisheries and Agri Business, Mælkeafgiftsfonden, VikingGenetics, Nordic Genetic Evaluation and Aarhus University. The work is co-funded by a Ph.D. scholarship from VikingGenetics and the Ministry of Science, Technology and Innovation.

Table 1. Essential input parameters used for modeling Danish Jersey breeding schemes. The conventional breeding scheme without use of genomic information, the hybrid breeding scheme with a combined use of genomic selected young bulls and progeny tested bulls and the turbo breeding scheme only with use of young bulls.

	Conventional	Hybrid	Turbo
Population parameters			
Number of cows in population	68,000	68,000	68,000
Number of young bulls mated with cows	60	60	60
Proportion of cows mated with young bulls	0.3	**0.5**	1.0
No. of proven bulls selected per year	5	4	-
No. of young bulls mated with bull dams	0	15	15
Proportion of bull dams mated with young bulls	0	**0.25**	1.0
No. of selected bull dams per genotyped bull calf	-	3	3
No. of selected bull dams per year	1,500	-	-
Inseminations per 1st lactation record	10	10	10
No of genotyped bull calves	0	**500**	500
Increased reliability of GEBV (percentage points)	0	**+5**	+5
Biological coefficients			
Average calving interval (years)	1.1	1.1	1.1
Inseminations per pregnancy	2.2	2.2	2.2
Rearing percentage for heifers and bull calves	0.9	0.9	0.9
Calving percentage	0.9	0.9	0.9
Survival rate for cows	0.8	0.8	0.8
Survival rates for waiting bulls	0.95	0.95	0.95
Use of proven bulls (years)	2.0	2.0	-
Use of young bulls (years)	0.3	0.3	0.6
Generation interval for bull dams	2.4	2.4	2.4
Generation interval for production cows	3.2	3.2	3.2
Variable cost parameters (EURO)			
Inspection of bull dams (per bull dam)	5	5	5
Inspection of bull calves (per genotyped calf)	-	40	40

Costs for SNP-typing per genotyped calf	-	100	100
Variable costs covering feeding and labor (per test bull per year)	2,500	2,500	2,500
Interest rate for return and costs (%)	6/4	6/4	6/4
Investment period (years)	15	15	15

Figures with bold are varied in the hybrid breeding scheme

Table 2. Comparison of the three main breeding schemes: conventional, hybrid and turbo. Breeding schemes are evaluated for annual monetary genetic gain (AMGG), discounted profit, generation interval and proportion of AMGG from production trait (PT) and functional trait (FT). Value of AMGG and discounted profit for the hybrid scheme is standardized to 100

	Conventional	Hybrid	Turbo
AMGG (%)	93.2	100	96.1
Discounted profit in %	88.7	100	149.7
Generation interval in years	4.14	3.58	2.48
AMGG from PT versus FT	67.2/32.8	66.3/33.7	63.8/36.2

Table 3. Effect of increasing reliability of GEBV on annual monetary genetic gain (AMGG), discounted profit, generation interval and proportion of AMGG from production trait (PT) and functional trait (FT) for the hybrid breeding scheme.

Percentage points	+5[1]	+ 10	+20	+30	+40
AMGG (%)	100	101.1	103.4	105.5	107.5
Discounted profit (%)	100	102.1	105.9	109.5	113.0
Generation interval in years	3.58	3.58	3.58	3.58	3.58
AMGG from PT vs FT	66.3/33.7	66.2/33.8	65.8/34.2	65.4/34.6	65.2/34.8

[1]Reference scenario of hybrid scheme

Table 4. Effect of increasing number of genotyped young bulls (YB) on annual monetary genetic gain (AMGG), discounted profit, generation interval and proportion of AMGG from production trait (PT) and functional trait (FT) for the hybrid breeding scheme.

No. of genotyped YB	500[1]	750	1,000	1,500	2,000
AMGG (%)	100	101.9	103.0	104.6	105.2
Discounted profit (%)	100	102.8	104.4	106.3	107.1
Generation interval in years	3.58	3.58	3.58	3.58	3.58
AMGG from PT vs FT	66.3/33.7	66.0/34.0	65.8/34.2	65.6/34.4	65.4/34.6

[1]Reference scenario of hybrid scheme

Table 5. Effect of proportion of bull dams mated with young bulls (YB) on annual monetary genetic gain (AMGG), discounted profit, generation interval and proportion of AMGG from production trait (PT) and functional trait (FT) for the hybrid breeding scheme.

Prop. of bull dams mated with YB	0	0.25[1]	0.50	0.75	1
AMGG (%)	99.5	100	99.9	99.3	98.2
Discounted profit (%)	97.9	100	102.1	104.1	106.5
Generation interval in years	3.79	3.58	3.37	3.15	2.94
AMGG from PT versus FT	66.7/33.3	66.3/33.7	65.9/34.1	65.4/34.6	65.0/35.0

[1]Reference scenario of hybrid scheme

Table 6. Effect of proportion of cows mated with young bulls (YB) on annual monetary genetic gain (AMGG), discounted profit, generation interval and proportion of AMGG from production trait (PT) and functional trait (FT) for the hybrid breeding scheme.

Proportion of cows mated withYB	0.25	0.50[1]	0.75	1
AMGG (%)	95.4	100	102.1	103.1
Discounted profit (%)	91.0	100	106.7	112.7
Generation interval in years	3.82	3.58	3.33	3.09
AMGG from PT versus FT	69.5/30.5	66.3/33.7	64.5/35.5	63.5/36.5

[1]Reference scenario of hybrid scheme

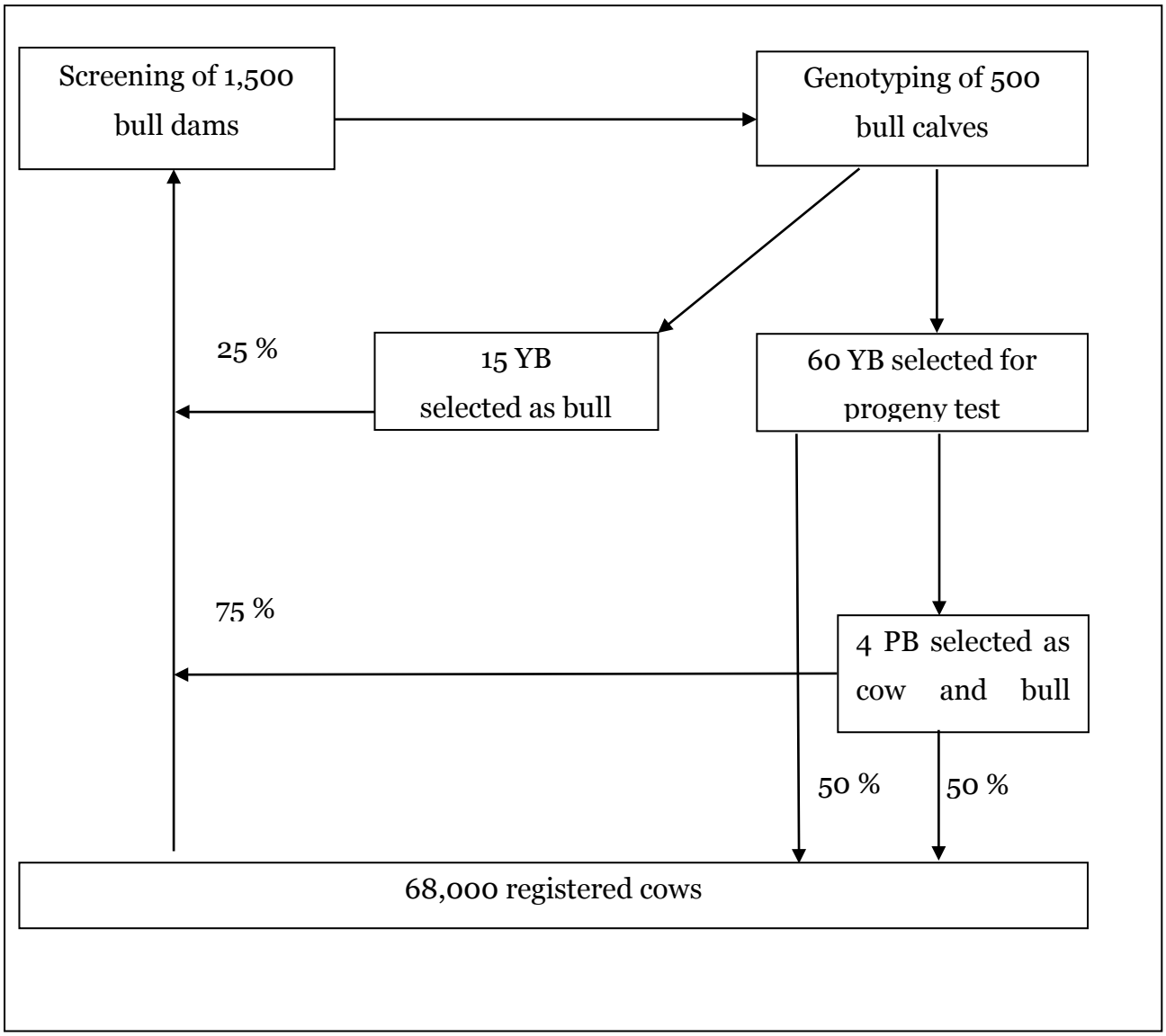

Figure 1: Illustration of selection steps in the genomic hybrid breeding scheme. Proportion of inseminations of bull dams and production cows by young bulls (YB) and proven bulls (PB) refers to the reference hybrid breeding scheme.

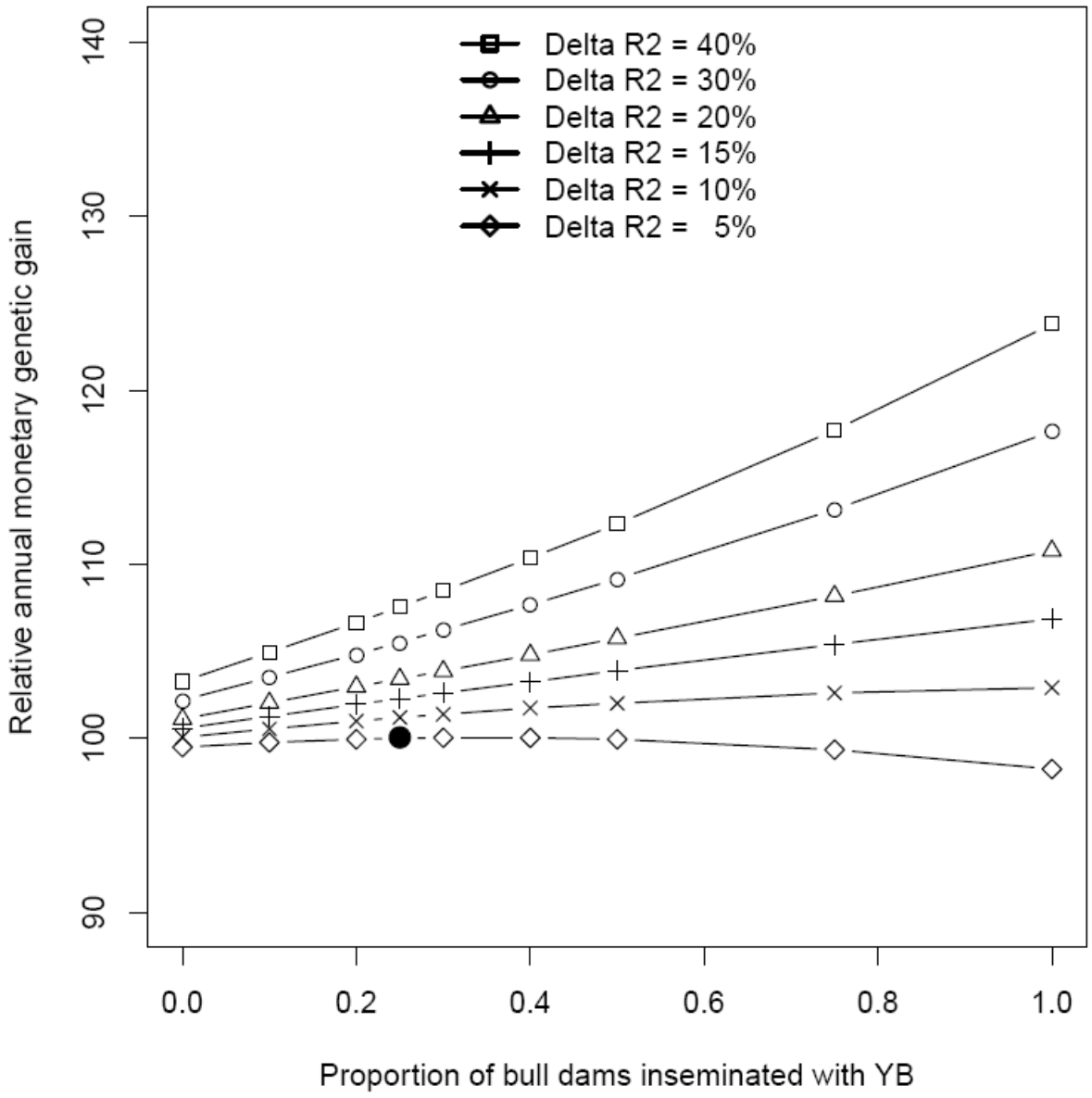

Figure 2. Annual monetary genetic gain as a function of increased proportion of bull dams inseminated with YB for different levels of increase in reliability due to added value of genomic information. The relative value of the reference scenario is set to 100 (●)

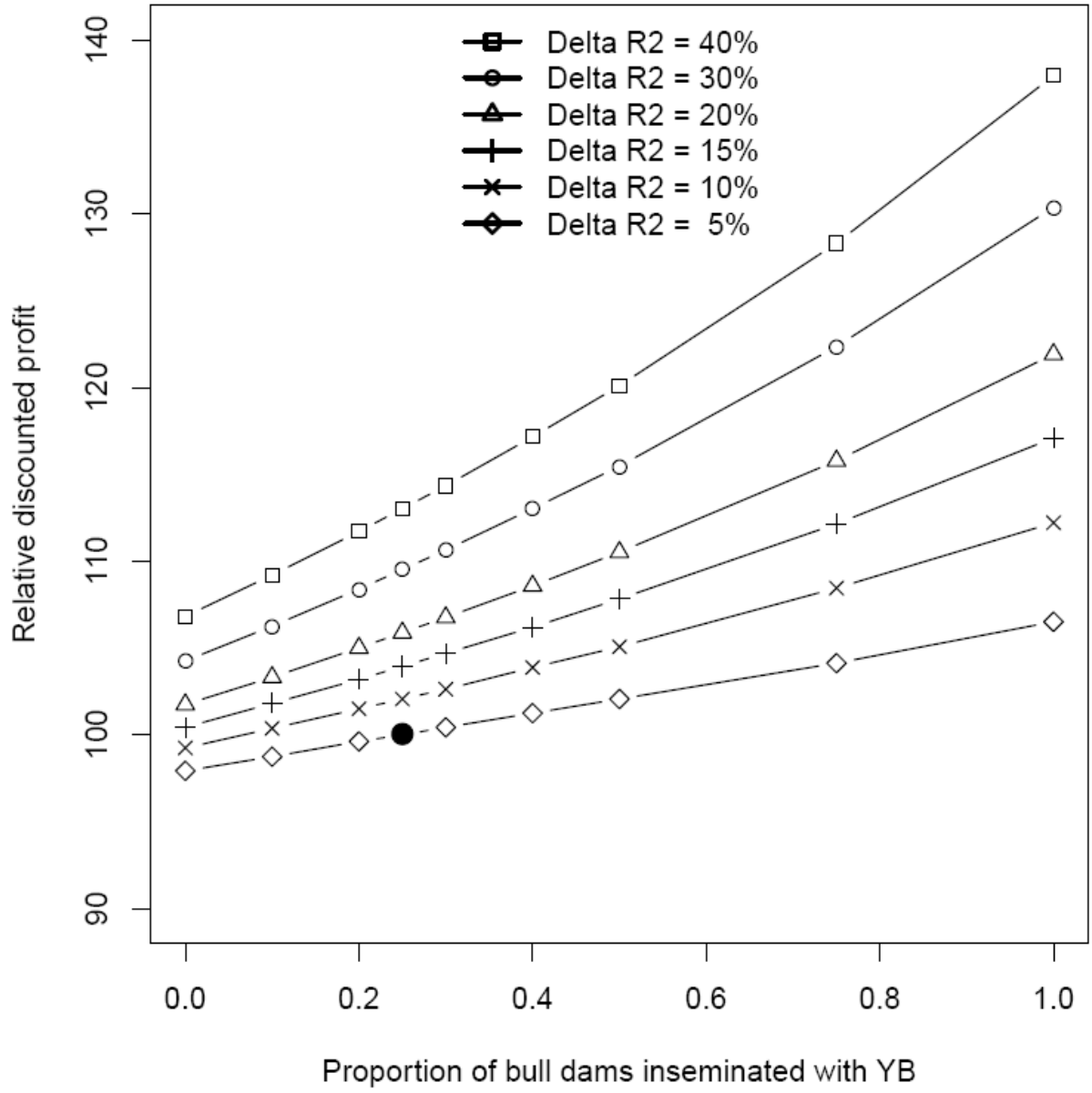

Figure 3. Discounted profit as a function of increased proportion of bull dams inseminated with YB for different levels of increase in reliability due to added value of genomic information. The relative value of the reference scenario is set to 100 (●)

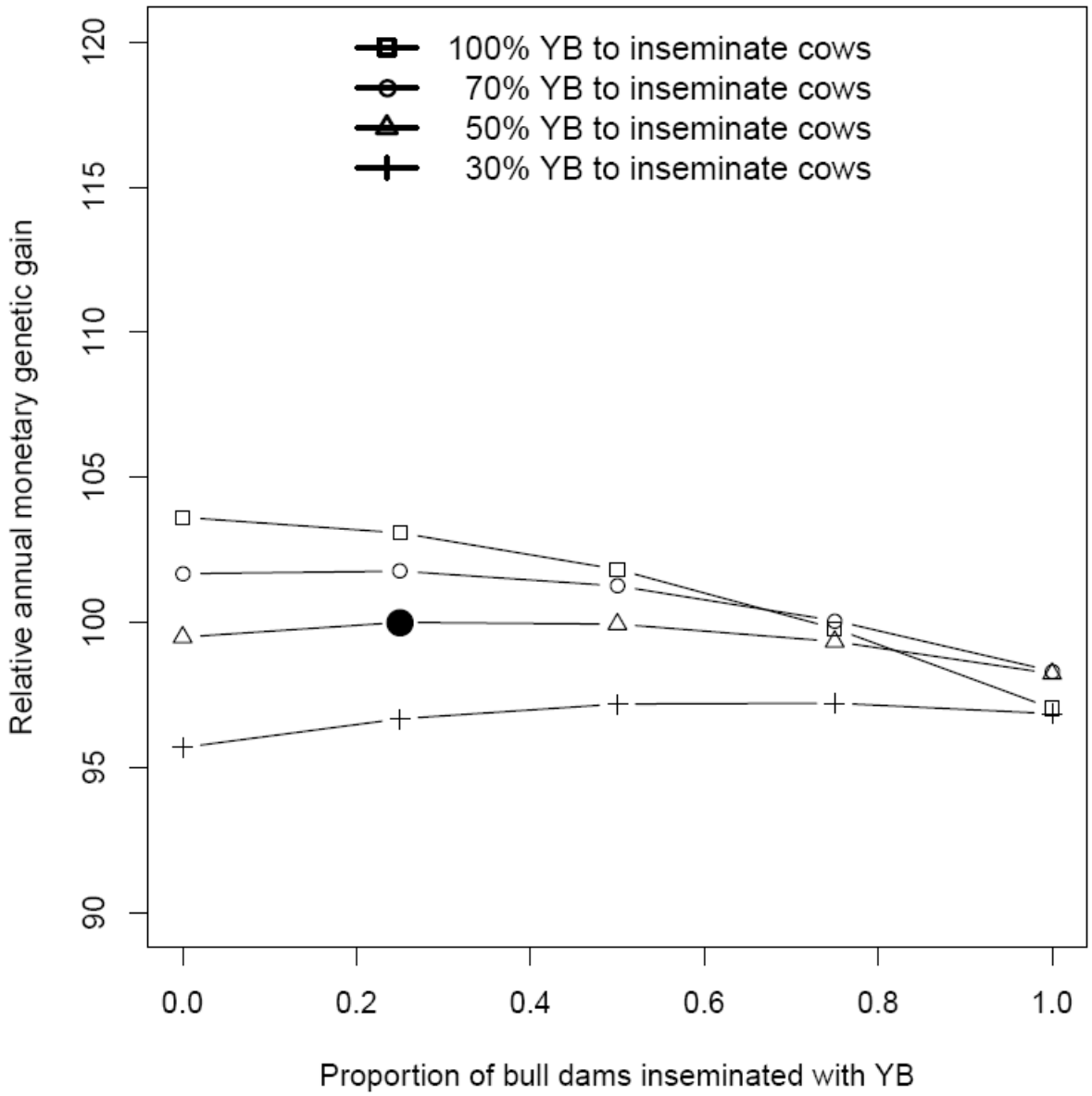

Figure 4. Annual monetary genetic gain as a function of proportion of bull dams inseminated with YB for different levels of YB use for inseminations of cows. The relative value of reference scenario is set to 100 (●)

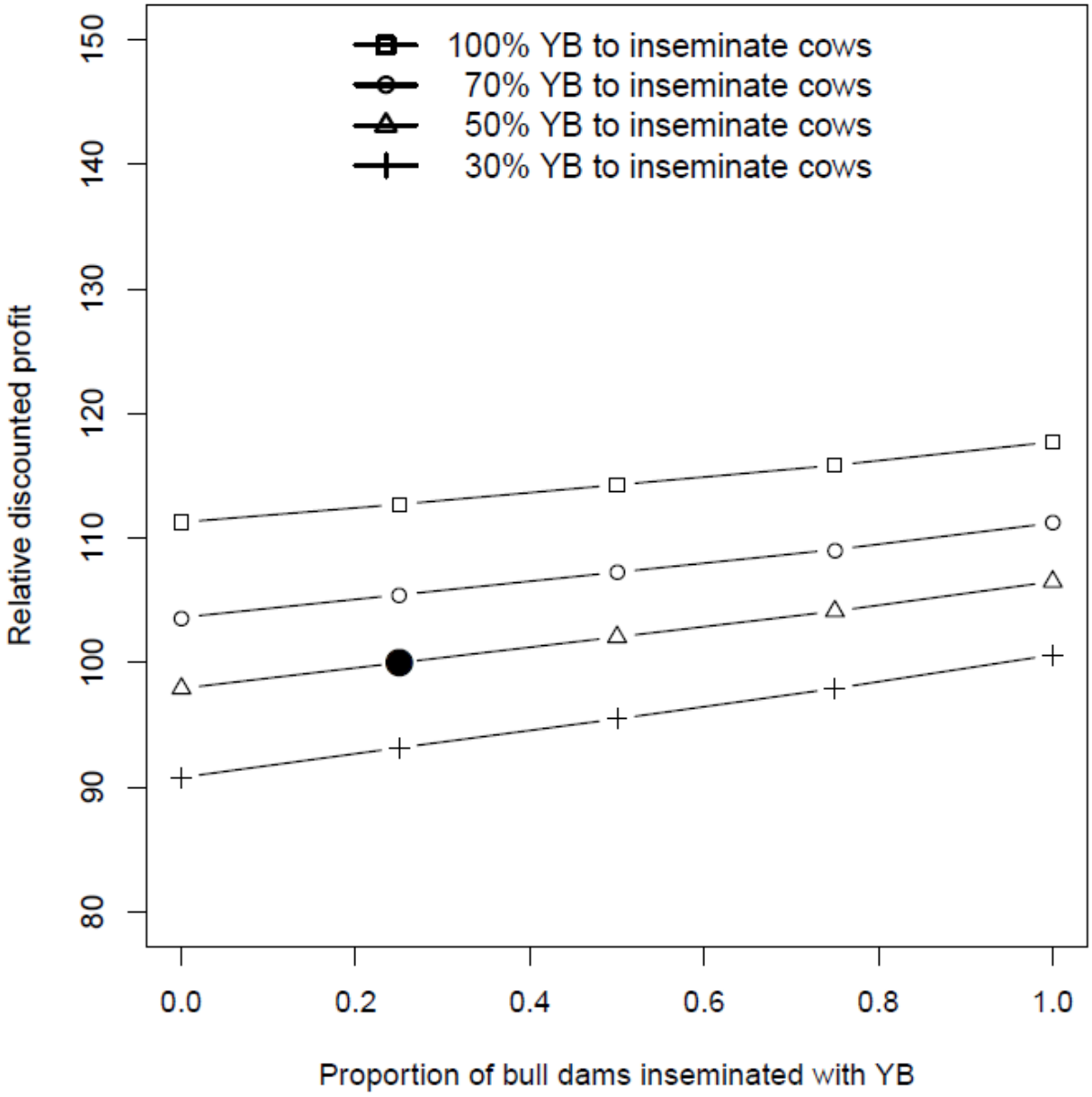

Figure 5. Discounted profit as a function of proportion of bull dams inseminated with YB for different levels of YB use for inseminations of cows. The relative value of the reference scenario is set to 100 (●)

REFERENCES

Buch, L. H., M. K. Sørensen, P. Berg, L. D. Pedersen, and A. C. Sørensen. 2012. Genomic selection strategies in dairy cattle: Strong positive interaction between use of genotypic information and intensive use of young bulls on genetic gain. Journal of Animal Breeding and Genetics 129(2):138-151.

de Roos, A. P. W., C. Schrooten, R. F. Veerkamp, and J. A. M. van Arendonk. 2011. Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. J. Dairy Sci. 94(3):1559-1567.

Egger-Danner, C., H. Schwarzenbacher, and A. Willam. 2012. Genotyping of cows to speed up the avaliability of GEBVs for direct health traits - genetic and economic aspects based on Fleckvieh (Simmental) Austria. Book of Abstracts of the 63rd Annual Meeting of the European Federation of Animal Science, Bratislava, Slovakia

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95(7):4114-4129.

Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10(6):381-391.

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. Genetics 177(4):2389-2397.

Habier, D., J. Tetens, F. R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genetics Selection Evolution 42(1):5.

Hayes, B. J., P. J. Bowman, A. C. Chamberlain, K. Verbyla, and M. E. Goddard. 2009a. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetics Selection Evolution 41.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009b. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92(2):433-443.

Henryon, M., P. Berg, T. Ostersen, B. Nielsen, and A. C. Sørensen. 2012. Most of the benefits from genomic selection can be realized by genotyping a small proportion of available selection candidates. J. Anim Sci.(accepted)

Hill, W. G. 1974. Prediction and evaluation of response to selecton with overlapping generations. Animal Production 18:117-139.

König, S., H. Simianer, and A. Willam. 2009. Economic evaluation of genomic breeding programs. J. Dairy Sci. 92(1):382-391.

Lauritsen, U. 2012. Tal om Kvæg. Online Assessed 28th of September 2012. http://www.landbrugsinfo.dk/Kvaeg/Tal-om-kvaeg/Sider/mndstatmain.aspx.

Lillehammer, M., T. H. E. Meuwissen, and A. K. Sonesson. 2011. A comparison of dairy cattle breeding designs that use genomic selection. J. Dairy Sci. 94(1):493-500.

Lund, M., A. de Roos, A. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genetics Selection Evolution 43(1):43.

Mc Hugh, N., T. H. E. Meuwissen, A. R. Cromie, and A. K. Sonesson. 2011. Use of female information in dairy cattle genomic breeding programs. J. Dairy Sc.i 94(8):4109-4118.

Pryce, J. E., and H. D. Daetwyler. 2012. Designing dairy cattle breeding schemes under genomic selection: a review of international research. Anim. Prod. Sci. 52(3):107-114.

Pryce, J. E., M. E. Goddard, H. W. Raadsma, and B. J. Hayes. 2010. Deterministic models of breeding scheme designs that incorporate genomic selection. J. Dairy Sci. 93(11):5455-5466

Pryce, J. E., B. Gredler, S. Bolormaa, P. J. Bowman, C. Egger-Danner, C. Fuerst, R. Emmerling, J. Sölkner, M. E. Goddard, and B. J. Hayes. 2011. Short communication: Genomic selection using a multi-breed, across-country reference population. J. Dairy Sci. 94(5):2625-2630.

Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics 123(4):218-223.

Sørensen, A. C., and M. K. Sørensen. 2009. Inbreeding rates in breeding programs with different strategies for using genomic selection. Interbull Bulletin No. 40.

Thomasen, J. R., B. Guldbrandtsen, G. Su, R. F. Brøndum, and M. S. Lund. 2012. Reliabilities of genomic estimated breeding values in Danish Jersey. Animal 6(5):789-796.

Thomasen, J. R., A. C., Sørensen, G. Su, P. Madsen, M. S. Lund and B. Guldbrandtsen. 2013. The admixed population structure in Danish Jersey challenges accurate genomic predictions. Submittet to J. of Anim Sci.

Wientjes, Y., R. VeerKamp, and M. Calus. 2012. Effect of linkage disequilibrium, haplotypes and family relations on reliability of genomic predictions. Book of Abstracts of the 63rd Annual Meeting of the European Federation of Animal Science, Bratislava, Slovakia

Willam, A., G. Nitter, H. Bartenschlager, K. Karras, E. Niebel, H.-U. Graser. 2008. ZPLAN - Manual for a PC-Program to Optimize Livestock Selection Schemes. Manual Version 2008 for Source Code “z10.for”.

Paper IV

Adding cows to the reference population increases reliability of genomic predictions in a small dairy population

J.R. Thomasen, A. C. Sørensen, M. S. Lund B. Guldbrandtsen,

Manuscript in preparation

GENOMIC SELECTION

Adding cows to the reference population increases reliability of genomic predictions in a small dairy population

J.R. Thomasen,* †A. C. Sørensen,† M. S. Lund, † and B. Guldbrandtsen, †

* VikingGenetics, DK 8860, Assentoft, Denmark

† Center for Quantitative Genetics and Genomics, Faculty of Science and Technology, Aarhus University, DK-8960, Tjele, Denmark

*Corresponding author: Jørn Rind Thomasen, VikingGenetics, DK 8860, Assentoft, Denmark. Phone: (45) 87 95 94 00. Fax: (45) 87 95 94 01.

Email: jotho@vikinggenetics.com

ABSTRACT

The main objective of this study was to evaluate the effect of including cows in the reference population for small dairy cattle population with a limited number of sires in the historic reference population, which therefore at present is challenged by having inaccurate genomic predictions. Three main scenarios for maintaining and updating the reference population were investigated for a period of 15 years. 1) Number of progeny tested bulls per year was tested at four levels: 15, 40 60 and 100. 2) Each year 2,000 first lactation cows were randomly selected from the cow population for genotyping (COW30000), and 3) an additional 2,000 first lactation cows were randomly selected and typed each of the first two years (COW34000). The effects were evaluated in two main breeding schemes. A turbo scheme exclusively using genotyped young bulls (turbo) and a breeding scheme with a mixed use of genotyped young bulls and progeny tested bulls (hybrid). The populations were simulated in three steps: 1) 500 generations of a historic population for build-up of linkage disequilibrium and a pool of segregating QTL. 2) 20 years of a conventional progeny testing program for build-up of a reference population consisting of 1000 progeny tested genotyped bulls and 3) followed by 15 years of the genomic selection schemes (turbo or hybrid). The breeding schemes were chosen to mimic practically feasible options for the Danish Jersey cattle population as an example of a small dairy cattle population. Two traits were included in the breeding goal, a production trait (h^2=0.30) and a functional trait (h^2=0.04). A stochastic approach was used to model the different strategies, where the evaluation criteria were annual monetary genetic gain, rate of inbreeding, reliability of genomic predictions and variance of response. Including cows in the reference population increased monetary genetic gain and decreased the rate of inbreeding compared to breeding schemes only updating the reference populations with 60 progeny tested bulls annually. The increase in genetic gain was larger for the turbo schemes characterized by exclusive use of young bulls and hence shorter generation intervals. The risk, measured as variance of response, of running the turbo schemes with genotyping of males was generally higher compared to the schemes using progeny tested bulls, due to the lower reliability of the bulls used intensively. The annual genetic gain and the reliability of genomic predictions were slightly higher in the COW34000 scenario compared to the COW30000 scenario. Inclusion of cows in the reference population is a fast way to increase reliabilities of genomic predictions in a small population.

Key words: genotyped cows, genomic selection, genomic breeding plans

INTRODUCTION

In genomic selection (**GS**) the effects of dense sets of genetic markers are estimated in a reference population of genotyped and phenotyped individuals (Meuwissen et al., 2001). The estimated effects of the markers are then used to predict genomic enhanced breeding values (**GEBV**) for selection candidates with no or only limited phenotypic information in addition to pedigree information. In dairy cattle, up to now reference populations mostly consisted of progeny tested sires with reliably predicted breeding values (**EBV**). These EBVs include information from hundred up to thousands of daughters. Adding more animals to the reference population has been shown to be the most efficient way of increasing reliabilities of GEBV's (Lund et al., 2011; Wiggans et al., 2011).

Small dairy cattle populations have small reference populations of progeny tested bulls. These populations, therefore, have low reliabilities of GEBV (Thomasen et al., 2012). This poses a challenge for their future genetic gain. Thomasen et al. (2013a) showed that low reliabilities of genomic predictions limit genetic gain in breeding schemes with more intensive use of young bulls without a progeny test. Therefore, an important objective for smaller dairy cattle breeds is to increase reliabilities of GEBV's. Increasing the amount of information within a breed can be achieved either by increasing the number of progeny tested bulls included in the reference population or by including genotyped females with own records directly in the reference population.

Genotyping cows have become more relevant with decreasing costs of genotyping in general and particularly the introduction of the less costly low density SNP marker panels. These are optimized for imputation to denser SNP panels. In the Nordic cow population a very high proportion of the cows have phenotypic recordings for all the traits in the breeding goal (Lauritsen, 2012). This makes genotyping of cows an option to increase reliabilities of genomic predictions. Buch et al., (2012) showed that a reference population consisting of all cows with a specific phenotype results in higher reliability compared to a reference population including only the proofs of the sires of these cows in the reference population.

Therefore higher genetic gain, less inbreeding, higher reliabilities of genomic predictions are to be expected in a small dairy cattle population where 1) genotyped females with own phenotypic records are added to the existing sire reference population, or 2) the annual number of progeny tested bulls included in the reference population is increased. Two strategies for genotyping cows were evaluated. One scenario modeled the effect of adding 2000 cows annually to the reference population. Another scenario modeled the effect of increasing this number of cows over a period of two years of selection.

The effects were evaluated in two main breeding schemes. A turbo scheme exclusively using genotyped young bulls (**turbo**) and a breeding scheme with a mixed use of genotyped young bulls and progeny tested bulls (**hybrid**). The turbo scheme will reduce the accuracies relatively more because their sires are not in the reference population (Thomasen et al., 2012). Thus genotyping of cows will have more information to contribute. Therefore we hypothesize that genotyped females contributes relatively with a higher genetic gain in the turbo scheme compared to the breeding scheme with a mixed use of genotyped young and progeny tested bulls.

In the present study the objective was to evaluate the merit of adding genotypes of bulls and cows in terms of monetary genetic gain, reliability of GEBV, loss of genetic variation and variability of the genetic gain in two different breeding schemes.

MATERIALS AND METHODS

Experimental design

The main objective is investigated by varying the number of progeny tested bulls and the number of cows added to the reference population over a given time horizon. A finite locus model was used to simulate the different breeding strategies.

The populations were simulated in three steps:

1) 500 generations of a historic population for build-up of linkage disequilibrium (**LD**) and a pool of segregating QTL. This step was computationally intensive and was shared between all scenarios.
2) 20 years of a conventional progeny testing program for build-up of a reference population consisting of 1000 progeny tested genotyped bulls, followed by

3) 15 years of the genomic selection schemes (turbo or hybrid)

The breeding schemes are chosen to mimic practically feasible options for the Danish Jersey cattle population as an example of a small dairy cattle population. A more detailed description of the breeding scheme parameters can be found in Thomasen et al. (2013a).

Hybrid scheme: The hybrid scheme reflects the current genomic breeding scheme in the Danish Jersey breed. 60 young bulls (**YB**) are selected annually for progeny testing according to their GEBV from a group of 500 genotyped bull calves. These YB obtain 100 daughter records for the production trait (**PT**) and 92 daughter records for the functional trait (**FT**). The 15 YBs with the highest GEBV are selected as bull sires (one to four years old) and mated to 25% of the bull dams and 50% of the cow population. Finally, four progeny tested bulls (**PBs**) are selected both for use as bull sires (five to six years old), contributing with 75% of the inseminations, and for inseminations in the cow population contributing with 50% of the inseminations. These PBs are available because the bulls are alive their daughter proofs are available.

Turbo scheme: This is a breeding scheme where only YBs are used as bull sires and for inseminations of cows. Thereby, the generation interval is minimized.

In both schemes the simulated population consists of 20,000 cows with records from 100 herds in both the turbo and hybrid scheme. The 1,500 cows with the highest EBV for the breeding goal across all herds are screened as bull dam candidates. It is assumed that these cows are inseminated with young bull sires to produce the 500 bull calves that will be genotyped.

Historic population: The goal for the formation of the historic population is twofold: first to generate a dense SNP marker set with a LD structure reflecting the Danish Jersey population (Thomasen et al., 2012), and secondly building a genetic architecture fitting the traits included in the breeding goal (Thomasen et al., 2013b). We model the genetic architecture of the traits by a finite locus model, where traits are influenced by a large but finite number of QTL each having a small but non-infinitesimal effect.

The historic population was simulated over 500 non-overlapping generations. In the base generation 200 males and 200 females were mated. This number was

reduced linearly down to 125 males and 125 females in the last generation. The genome for each animal contained 30 chromosomes of 100 cM containing a total of $3x10^8$ evenly distributed potential SNP markers. The base population was completely homozygous. A mutation rate of $1.8x10^{-6}$ converted the original allele to an alternate allele. Every 32nd base was a potential QTL. If a mutation occurred at a potential QTL, segregating QTL was generated. After the 500 generations 157,374 markers and 5,055 QTL were segregating. The LD structure was evaluated by calculating the r^2 (Lewontin, 1964) for different marker distances (Figure 1). The r^2 was calculated as the squared correlations between markers that were grouped into bins by lengths of 10 kb. All breeding schemes were started by sampling from the haplotypes of the same last generation of the historic population. We assumed the total genetic variance in the historic population for each trait be 1.0.

Sire reference population: The sire reference population was generated over a period of 20 years (time step 1 to 20) and was aimed to reflect the current size of the Danish Jersey reference population. Each year 50 bulls were progeny tested, adding up to a total of 1,000 progeny tested bulls. Each year 5 PB were selected for further inseminations. Only PBs were used as bull sires in this scheme. A detailed overview of the breeding scheme parameters is given in Table 1. This breeding scheme model at the same time the conventional progeny testing scheme (**C-B50**) before the start of genomic selection. We used the results from time step to 11 to 20 for evaluation of the conventional breeding scheme (Table 1).

Strategies for maintaining future reference population

Three scenarios for maintaining and updating the reference population was investigated for both the hybrid and turbo breeding schemes: 1) Number of progeny tested bulls was varied between 15 and 100 at 4 different levels (**B15, B40, B60 and B100**). In all scenarios the number of YB used as bull sires are 15. 2) Each year 2,000 first lactation cows were randomly selected from the cow population for genotyping adding up to 30,000 cows for the period of the 15 years (**COW30000**), and 3) an additional 2,000 first lactation cows were randomly selected and typed each of the first two years in order to evaluate the effect of increasing the reliabilities of the genomic predictions more intensively compared to scenario 2. The remaining 13 years 2,000 cows were randomly selected and genotyped for a total of 34,000 genotyped cows over the whole period (**COW34000**). The scenarios COW30000

and COW34000 were only simulated in the hybrid and turbo reference schemes with 60 progeny tested bulls annually.

The daughter group size of the YBs was the same in all scenarios. In practice, the daughter group size should be reduced, when the number of tested YBs is increased assuming a fixed number of cows available for insemination with YB. Therefore, comparisons of breeding schemes will only be made between schemes with the same number of progeny tested bulls.

Breeding goal and traits

The breeding goal includes two traits. The first trait represents milk production traits (h^2=0.30) and the second represent functional traits (h^2=0.04). The effects of the QTLs affecting the two traits were sampled from a bivariate normal distribution. An unfavorable genetic correlation of -0.30 between the two traits was assumed, while the residuals were assumed uncorrelated. The economic values were set to 83 Euro for the PT and 82 Euro for the FT per additive genetic standard deviation. These economic values were chosen such that the correlation between milk production and breeding goal is the same as it is in the Nordic total merit index (Buch et al., 2012). The total merit index reflecting the overall breeding goal was used as selection criteria for all animals. The variance of the breeding goal is 10000 Euro.

Sampling of breeding values and phenotypes

The true breeding values (**TBV**) were constructed by summing the QTL allelic effects within and across loci. The phenotypes were simulated by adding a residual term sampled from a normal distribution with a mean of 0. A residual variance of 2.33 was used for the PT and 24 for the FT.

Estimation of breeding values

Two different genetic evaluations were used: 1) a single step genomic BLUP (**GBLUP**) for prediction of GEBV and 2) a BLUP animal model using traditional pedigree relationship for prediction of EBV. The GBLUP procedure was used for prediction of GEBVs for the selection of bull calves to genotype and YBs. Conventional BLUP was used for prediction of EBV in the selection steps of bull dams, cows and PBs.

The procedures for estimation of GEBV using GBLUP were previously described (Christensen and Lund, 2010; Gao et al., 2012). As we did not expect that the SNP markers capture all the genetic variance explained by the QTL, due to incomplete LD between markers and QTL, the relationship matrix used for the genotyped animals **($\mathbf{G}_w$)** was a linear combination of the original genomic relationship matrix **(G)** and the relationship matrix **(A**), as described by (Gao et al., 2012). The $\mathbf{G}_w$ was calculated as $\boldsymbol{G_w} = (1-w)\boldsymbol{G} + w\boldsymbol{A}$ **,** where w was the weight representing the fraction of genetic variance not captured by markers. In this study a weight of 0.1 was used. Prediction of GEBV and EBV was done using DMU (Madsen and Jensen, 2008)

Method and evaluation criteria

The stochastic simulation tool ADAM (Pedersen et al., 2009) was used to simulate the different breeding strategies. For the 8 scenarios involving the variable number of progeny tested YBs, 20 replicates of each scenario were performed. Due to long computation time for GBLUP breeding value estimation in the scenarios involving genotyped cows, only 5 replicates were performed for the four scenarios involving genotyping cows. Analyses of annual monetary genetic gain (**ΔG**), variance of ΔG (**V**) and rate of inbreeding per generation (**ΔF**) were evaluated in the years from 21 to 35.

For each replicate ΔG was calculated as the regression coefficient of mean TBV on birth year of animal. In this study ΔG was presented as mean of replicates for each scenario. V was calculated as the variance of ΔG between replicates. Inbreeding rates were presented per generation and estimated from the pedigree. The inbreeding rate per generation was calculated for each replicate as one minus the exponential of the regression coefficient of the natural logarithm of mean inbreeding on the average generation equivalent for each year. The presented ΔF was then obtained by averaging over replicates. Differences in ΔG and ΔF between scenarios were compared to the least significant difference (**LSD**) using a confidence level of 95%.

The reliabilities of the estimated breeding values were calculated as the squared correlation between the estimated breeding values and TBV of the available candidates in a specific selection group. The reliabilities were calculated as averages across replicates within each selection group and birth year.

RESULTS

Updating the reference population with bulls

For all levels of progeny tested bulls the turbo scheme gave significantly ($P<0.05$) higher annual ΔG compared to the hybrid schemes (Table 2). The highest difference between the two schemes was seen for 15 YB progeny tested per year, with a difference of 18.1%. The smallest difference between the two schemes was obtained for 60 YB, where the turbo scheme was genetically superior by 8.6%. The variance of response for all turbo schemes was in the range from 2.86 to 8.09. For the hybrid schemes the interval was from 2.48 to 4.00.

The genomic schemes provided significantly ($P<0.05$) higher ΔG compared to the conventional scheme (R-B50). The H-B60 was 14% superior and the T-B60 was 33% superior.

The increase in inbreeding per generation was in general lower in the turbo schemes (1.65% to 1.86%) compared to the hybrid schemes (1.81% to 2.15%), but only significantly so ($P<0.05$) for 15 and 100 progeny tested bulls. The lowest increase in inbreeding was observed for the conventional scheme (1.36%).

In the hybrid schemes bulls were a mixture of YBs and PBs (Table 1). The generation interval in the hybrid schemes was around 3.75 years. In the turbo schemes with exclusive use of YB the generation interval was reduced to approximately 2.6 years.

Updating the reference population with genotyped cows

Updating the reference population with genotyped cows provided €5.6 higher annual ΔG (20%) in the turbo scheme compared to the hybrid scheme (Table 3). The COW34000 provided higher annual ΔG in both the hybrid (€1.8) and the turbo scheme (€1.0). However, due to few replicates only the H-COW30000 provided significant lower annual ΔG. In the cow schemes the V range from 0.34 in the T-COW34000 and up to 8.83 in the H-COW34000.

Compared to the hybrid scheme (H-B60) in Table 2, the H-COW30000 provided €1.8 (7%) higher annual ΔG, whereas the T-COW30000 scenario yielded €5.6 (20%) higher annual ΔG than T-B60 scheme. The H-COW34000 provided €3.4 (14%) higher annual ΔG than the H-B60. For the T-COW34000 the difference to the T-B60 was €6.6 (24%).

The lowest rate of inbreeding was observed for the scenarios involving genotyping of cows. The reduction is 3.0% for the hybrid scheme and 35% for the turbo scheme.

Reliabilities of GEBV

The development in reliabilities for genotyped bull calves in the three turbo schemes T-B60, T-COW30000 and T-COW34000 are shown in Figure 2. In the T-B60 scheme the reliability only increased slightly over the period of 15 years, with an average of 0.19 in the first five years reaching a level of 0.27 on average over last five years. The reliability of BLUP parent average EBV was 0.1 in year 21 (results not shown), and hence the gain in reliability due to genomic information in the sire reference population was 0.1.

The T-COW30000 and T-COW34000 schemes showed a higher increase in reliability during the first four years. After this period the reliability increases with a lower rate. Adding 4000 extra cows in the reference population year 21 and 22 (T-COWS34000) increased reliability of GEBV up to year 30. After this period there was no difference in reliabilities between the two strategies. For both schemes the reliability increased from approximately from 0.3 to just below 0.6.

Figure 3 shows the development in reliabilities for genotyped bull calves in the three hybrid schemes H-YB60, H-COW30000 and H-COW34000. The development in reliabilities followed the same pattern at a slightly higher level as for the turbo schemes.

DISCUSSION

Genotyping of cows for inclusion in the reference population increases genetic gain and decreases the rate of inbreeding compared to breeding schemes only updating reference populations with a limited number of progeny tested bulls. This increase in genetic gain is larger for the turbo schemes characterized by exclusive use of YBs and hence shorter generation intervals. The risk, measured as variance of response, of running the turbo schemes with genotyping of males is generally higher compared to the schemes using progeny tested bulls due to the lower reliability of the bulls used intensively. Inclusion of cows in the reference population is a fast way to increase reliabilities of genomic predictions.

Value of genotyping cows

Adding genotypes of cows had a major positive effect on reliabilities of GEBV. We proposed genotyping of cows to update and enlarge the reference population for a small population with a small reference population and a limited annual number of progeny tested bulls. Sixty bulls progeny tested annually only resulted in a minor increase in reliability of the genomic predictions (+0.07) in both the turbo scheme and the hybrid scheme over the 15 years. Already after four years, with genotyping of a total of 8,000 cows in COW30000 scheme the reliability of GEBV was increased by +0.20. After the fourth year with genotyping of cows, the reliability increased with a lower rate, where the added information from the genotyped cows and progeny tested bulls was sufficient to offset the decreased value (aging) of the historic sire reference population.

Depending on the breeding scheme applied, the increase in genetic gain by including cows in the reference varied between 5.5% and 14%. (Mc Hugh et al., 2011) studied the value of including cows in the reference population for a small dairy cattle population with 500 bulls in the reference population. The genetic gain was increased with 9% by adding 500 genotyped cows to the reference and increased 44% by adding 3500 genotyped cows to the reference population annually. The marginal value of including cows was found to be higher, since the reference population in the study by (Mc Hugh et al., 2011) was only half the size compared to the present study.

The genotyped cows were randomly chosen from the approximately 5,000 first lactation cows. However, strategies for selection of the genotyped females may have an impact on the reliabilities of GEBV. Pszczola et al. (2012) showed that the family structure in the reference population influences the reliability of GEBV. The highest reliability of GEBV were obtained when the relationship between animals in the reference population were lowest. So strategies which minimize the relationship between the cows chosen for genotyping are expected produce the biggest improvements in reliability, e.g. strategies that ensure a balanced family size of genotyped cows. In contrast relationship between animals in the reference and selection candidates should be increased (Pszczola et al., 2012). The bull dam candidates were not genotyped in the setup for our study. This was chosen not to do in order to evaluate the effect on monetary genetic gain of including genotyped production cows in the reference population. However, using genotype information in selection of females will result in increased reliabilities of young bull dam

candidates without own performance and hence reduce the generation interval in this selection path and increase genetic gain. Dassonneville et al. (2012) showed that inclusion of elite females in the reference population leads to over-estimated genomic predictions for production traits. Inclusion of genotyped females in the reference population will therefore probably not increase reliabilities of genomic predictions.

Inbreeding

Including cows in the reference population reduces average relationship in reference population and with the highest extend in the turbo schemes. Without genomic information, candidates without own phenotypic information or offspring information have breeding values based on their parents' information. Genomic information adds information about the Mendelian sampling term for young selection candidates. Information about the Mendelian segregation leads to more efficient selection within families. GS is therefore expected to reduce rate of inbreeding per generation compared to traditional BLUP selection assuming the same breeding scheme (Deatwyler et al., 2007). In the present study, the lowest rate of inbreeding was seen for the schemes using cows in the reference population compared to the schemes where only progeny tested bulls are used to update the reference population.

In GS breeding schemes, where the reference population only consists of bulls, the general expectation is that GS schemes provide lower rates of inbreeding compared to conventional breeding schemes. We did not observe this. It might have two explanations. First, the gain in reliability due to genomic information is low and hence information about the Mendelian sampling term is still inaccurate. Secondly, the hybrid schemes use a few PB quite intensively.

Selection of breeding candidates with a lower reliability increases the risk of the breeding schemes. In the breeding schemes only with genotyping of males, we found on average a lower variance of response in the hybrid schemes compared to the turbo schemes, where reliabilities selection decisions on average are based on GEBV with lover reliabilities as no PB bulls are used. In the schemes including genotyping of cows this conclusion could not be verified. Due to the low number of replicates in the schemes including more replicates are probably needed to confirm this finding.

Gain in reliability

The reliabilities of GEBVs obtained by GBLUP were 10% points higher than the reliabilities of traditional parent averages obtained from traditional BLUP estimates. This gain was twice the gain recently estimated from real data in the Danish Jersey population with 1,000 reference bulls (Thomasen et al., 2012). There might be several reasons for that. First we do not know the true genetic architecture of the traits. In this study, 5,055 purely additive QTL explained the additive genetic variance of the traits. We used 157,374 segregating markers many of which had small minor allele frequencies (**MAF**). In the real Danish Jersey population 38,242 markers are segregating (Thomasen et al., 2012). However, this is a selected set of SNP markers biased towards high MAF. The effect of this difference is unknown, as we do not know how SNPs at low frequencies contribute to tracking the QTL explaining the total breeding value. In addition, other genetic effects besides additive genetic effects might be present and complicate prediction.

The level of LD is also known to influence the reliability of GEBV. In case QTLs are closely linked to markers, then the LD in this simulation study is higher compared to the level of LD in the real population (Figure 1) and result in higher reliabilities of genomic predictions. But as stated earlier we do not know the true genetic architecture behind the traits.

Optimal breeding scheme

We find that turbo schemes are always superior to the hybrid schemes, irrespective of the number of progeny tested bulls. Thomasen et al. (2013a) found in a deterministic simulation study, using similar breeding parameters, the optimal breeding scheme being dependent on the reliabilities of GEBV. The study showed that a turbo scheme just become genetically superior, when the gain in reliability is 10% points. However, for a lower increase in reliabilities (+5%) the optimal breeding scheme still involves intensive use of progeny tested bulls. A further investigation to clarify the optimal number of YB when cows are included in reference population would be interesting, both from a genetic and economic perspective, as progeny testing of bulls are a costly process.

Table 1. Parameters used for modeling of breeding schemes. The conventional reference breeding scheme for buildup of the sire reference population (C-B50) and the two future breeding schemes, the hybrid breeding scheme (hybrid) with a combined use of genomic selected young bulls and progeny tested bulls and the turbo breeding scheme (turbo) only with use of young bulls.

Breeding scheme parameters	C-B50	Hybrid	Turbo
Number of cows in breeding population	20,000	20,000	20,000
Number of herds	100	100	100
Proportion of cows mated with young bulls	0.3	**0.5**	1.0
No. of proven bulls selected per year	5	4	-
Age distribution of proven bulls (years)	5 to 6	5 to 6	-
No. of young bulls mated with bull dams	0	15	15
Age distribution of bull dams (years)	2 to 3	2 to 3	2 to 3
Number of progeny tested young bulls	50	**60**	**60**
Age distribution of young bulls (years)	1	1	1
No. of young bulls as bull sires	0	15	15
Age distribution of young bulls sires(years)	-	1 to 4	1 to 4
Proportion of bull dams mated with young bulls	0	0.25	1.0
No. of selected bull dams per year	1,500	1,500	1,500
No of genotyped bull calves	0	500	500
Culling rates bull calves (year 0 to 1)	0.15	0.15	0.15
Culling rates bull calves (year 2 to 6)	0.05	0.05	0.05
Culling rates females	0.10	0.10	0.10
Max. no paternal half sibs genotyped	-	50	50
Max. no paternal half sibs as young bull sires (per year/total)	-	4/5	4/5
Time period (years)	1 to 20	21 to 35	21 to 35
No. of genotyped cows year 21 to 35 (per year)	-	**2000**	**2000**
Additional no. of genotyped cows year 21 and 22 in total	-	**4000**	**4000**

Figures with bold are varied

Table 2. Annual monetary genetic gain (ΔG), variance of response (V), generation interval (GI) and rate of inbreeding in percentage per generation averaged over years 21 to 35 and averaged over 20 replicates for the hybrid (H) and turbo (T) breeding schemes and varied number of young bulls (B) tested. The figures for the conventional breeding scheme (C-B50) are average over years 11 to 20 over 160 replicates. Standard errors are given in brackets.

Scheme	ΔG*	V	GI	ΔF* (%)
C-B50	21.8^{c} (0.063)	0.48	4.96	1.83^{c} (0.013)
H-B15	23.2^{a} (0.352)	2.48	3.74	2.15^{a} (0.134)
T-B15	27.4^{b} (0.481)	4.63	2.59	1.82^{bc} (0.090)
H-B40	24.9^{a} (0.386)	2.98	3.79	2.03^{ac} (0.090)
T-B40	27.7^{b} (0.503)	5.06	2.62	1.77^{ac} (0.105)
H-B60	25.6^{a} (0.358)	2.57	3.77	1.81^{ac} (0.087)
T-B60	27.8^{b} (0.379)	2.86	2.62	1.86^{ac} (0.071)
H-B100	26.1^{a} (0.448)	4.00	3.72	1.99^{ac} (0.086)
T-B100	29.8^{b} (0.636)	8.09	2.61	1.65^{bc} (0.082)

*: Figures with different letters are significant different (P<0.05). For hybrid (H) and turbo schemes (T) comparisons are only valid between the same numbers of progeny tested bulls.

Table 3. Annual monetary genetic gain (ΔG), variance of response (V), generation interval (GI) and rate of inbreeding in percentage per generation averaged over years 21 to 35 and average of 4 replicates for different breeding schemes and different strategies for genotyping of cows. Standard errors are given in brackets.

Scheme	ΔG*	V	GI	ΔF* (%)
H-COW30000	27.4^{a} (0.737)	2.17	3.94	1.49^{a} (0.104)
H-COW34000	29.2^{ab} (1.486)	8.83	3.74	1.98^{a} (0.506)
T-COW30000	33.4^{bc} (1.471)	8.65	2.57	1.29^{a} (0.085)
T-COW34000	34.4^{c} (0.293)	0.34	2.57	1.14^{a} (0.096)

*: Figures with different letters are significant different (P<0.05).

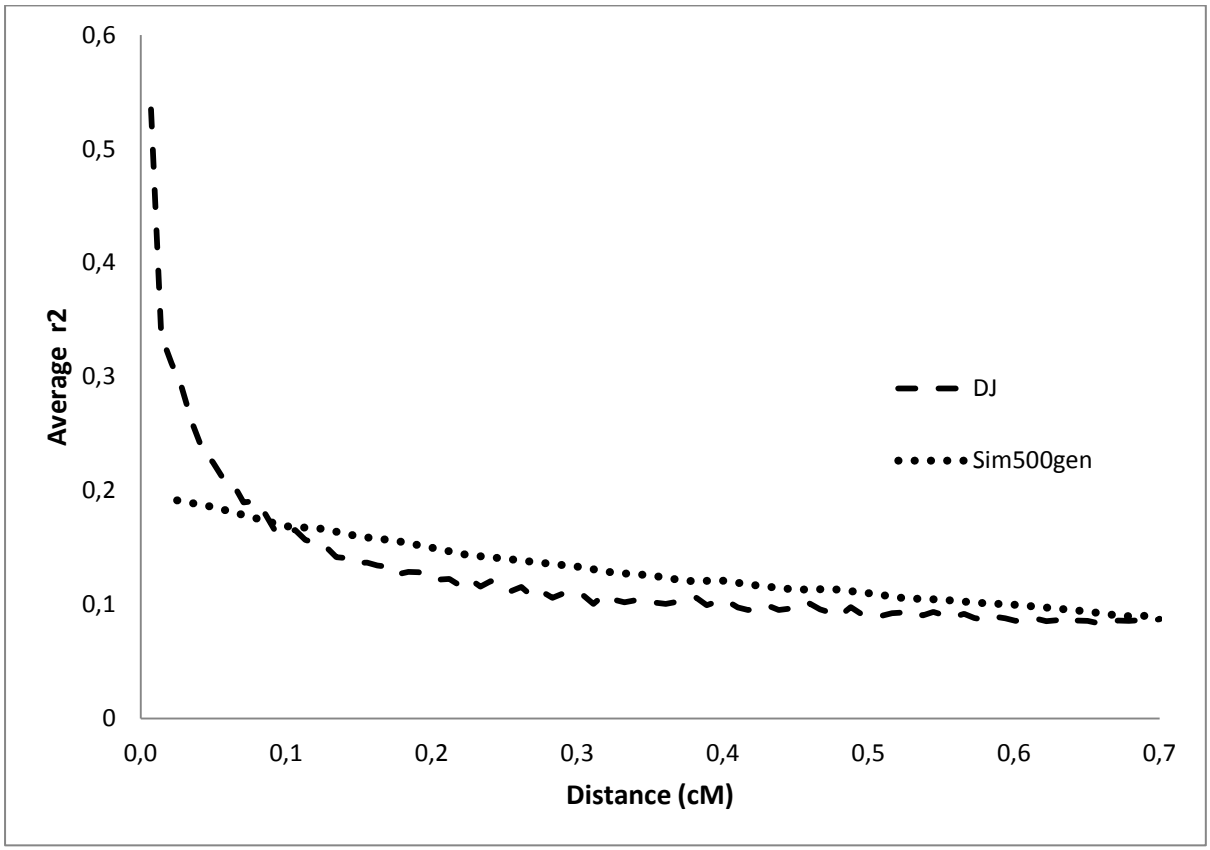

Figure 1: Linkage disequilibrium calculated as average r2 for different marker distances in cM. DJ line shows the r2 in the Danish Jersey population and the Sim500gen line shows the r2 in the last generation of the historic population.

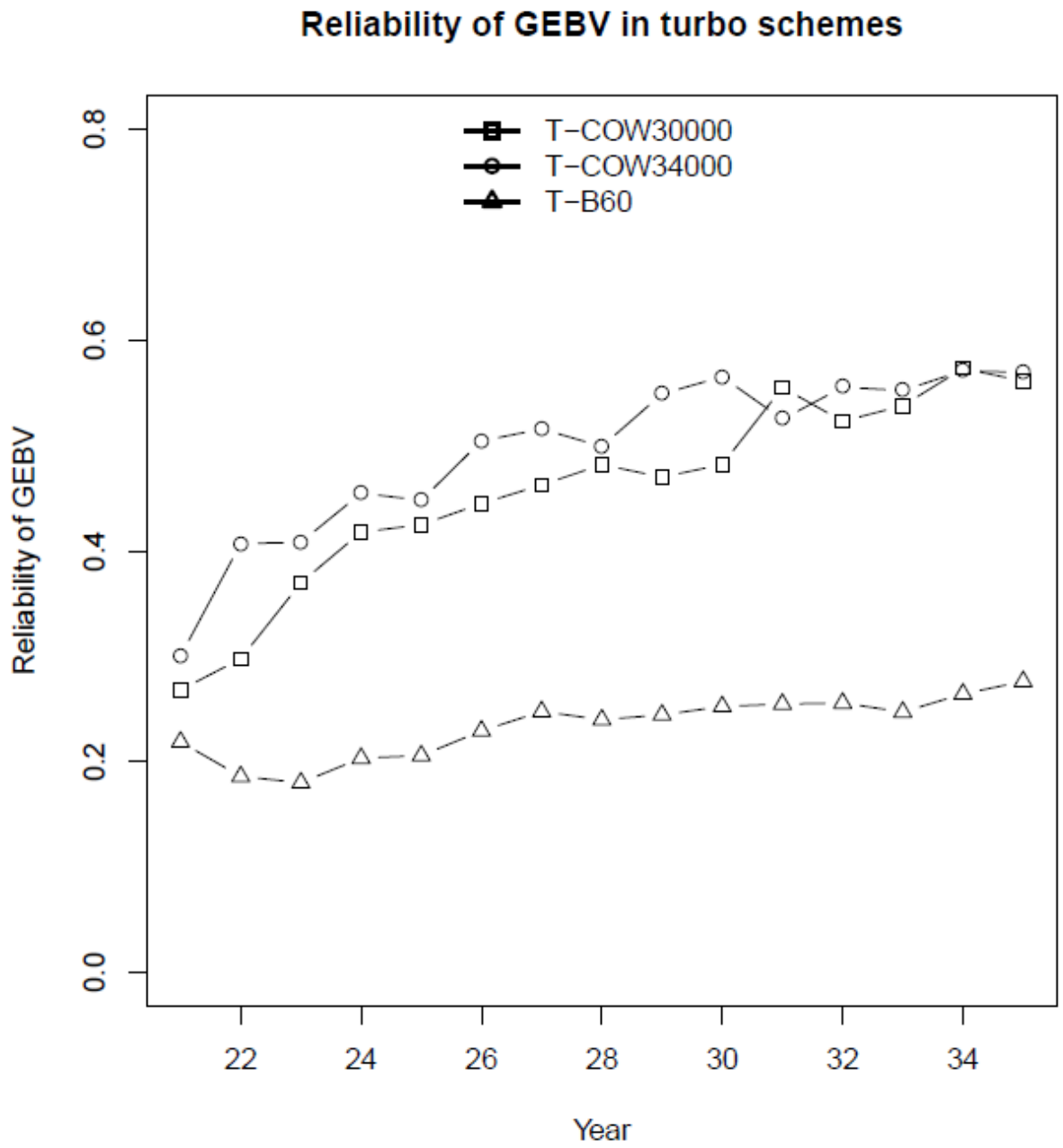

Figure 2: Development in reliability of GEBV over time for the turbo schemes adding progeny tested young bulls to the reference population (T-B60), and genotyped cows to the reference population (T-COW30000 and T-COW34000).

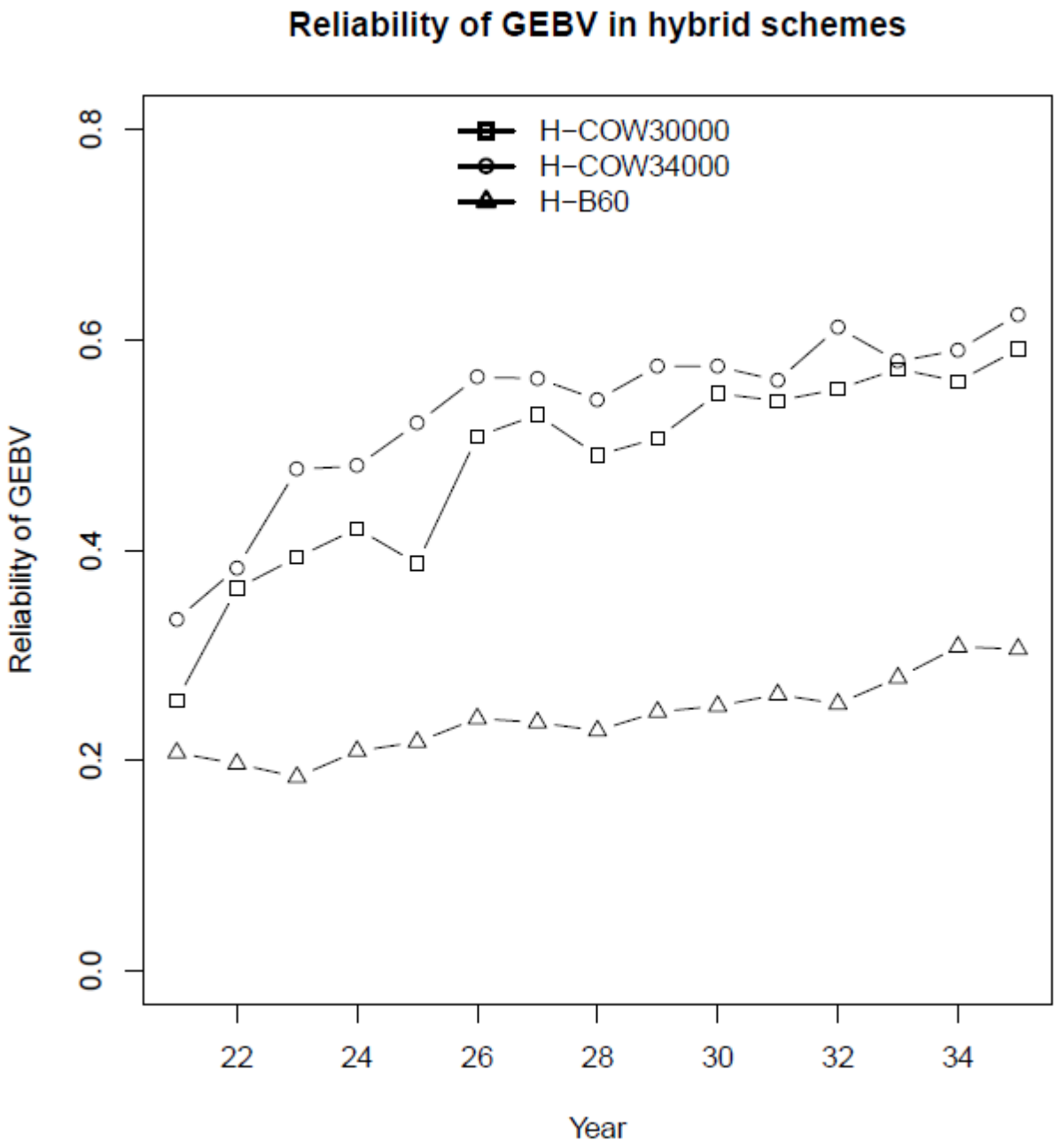

Figure 3: Development in reliability of GEBV over time for the hybrid schemes adding progeny tested young bulls to the reference population (H-B60), and genotyped cows to the reference population (COW30000 and COW34000).

ACKNOWLEDGMENTS

This work was performed in the project “Genomic Selection – from function to efficient utilization in cattle breeding (grant no. 3412-08-02253)”, funded by the Danish Directorate for Food, Fisheries and Agri Business, Mælkeafgiftsfonden, VikingGenetics, Nordic Genetic Evaluation and Aarhus University. The work is co-funded by a Ph.D. scholarship from VikingGenetics and the Ministry of Science, Technology and Innovation.

REFERENCES

Buch, L. H., M. Kargo, P. Berg, J. Lassen, and A. C. Sørensen. 2012. The value of cows in reference populations for genomic selection of new functional traits. Animal 6(6):880-886.

Buch, L. H., M. K. Sørensen, P. Berg, L. D. Pedersen, and A. C. Sørensen. 2012. Genomic selection strategies in dairy cattle: Strong positive interaction between use of genotypic information and intensive use of young bulls on genetic gain. Journal of Animal Breeding and Genetics 129(2):138-151.

Christensen, O., and M. Lund. 2010. Genomic prediction when some animals are not genotyped. Genetics Selection Evolution 42(1):2.

Dassonneville, R., A. Baur, S. Fritz, D. Boichard, and V. Ducrocq. 2012. Inclusion of cow records in genomic evaluations and impact on bias due to preferential treatment. Genetics Selection Evolution 44(1):40.

Deatwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. J Anim Breed Genet 124:369-376 .

Gao, H., O. Christensen, P. Madsen, U. Nielsen, Y. Zhang, M. Lund, and G. Su. 2012. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. Genetics Selection Evolution 44(1):8.

König, S., H. Simianer, and A. Willam. 2009. Economic evaluation of genomic breeding programs. J Dairy Sci 92(1):382-391.

Lauritsen, U. 2012. Tal om Kvæg. Online Assessed 28th of September 2012. http://www.landbrugsinfo.dk/Kvaeg/Tal-om-kvaeg/Sider/mndstatmain.aspx

Lewontin, R. C. 1964. The interaction of selection and linkage I. General considerations; Heterotic models. Genetics 49(1):49-67.

Lund, M., A. de Roos, A. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A

common reference population from four European Holstein populations increases reliability of genomic predictions. Genetics Selection Evolution 43(1):43.

Madsen, P., and J. Jensen. 2008. An user's guide to DMU, Version 6, release 4.7.

Mc Hugh, N., T. H. E. Meuwissen, A. R. Cromie, and A. K. Sonesson. 2011. Use of female information in dairy cattle genomic breeding programs. J Dairy Sci 94(8):4109-4118.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157(4):1819-1829.

Pedersen, L. D., A. C. Sørensen, M. Henryon, S. Ansari-Mahyari, and P. Berg. 2009. ADAM: A computer program to simulate selective breeding schemes for animals. Livestock Science 121(2-3):343-344.

Pryce, J. E., and H. D. Daetwyler. 2012. Designing dairy cattle breeding schemes under genomic selection: A review of international research. Anim. Prod. Sci. 52(3):107-114.

Pryce, J. E., M. E. Goddard, H. W. Raadsma, and B. J. Hayes. 2010. Deterministic models of breeding scheme designs that incorporate genomic selection. J Dairy Sci 93(11):5455-5466

Pszczola, M., T. Strabel, J. A. M. van Arendonk, and M. Calus. 2012. Reference population design affects reliability of selection of (un)genotyped animals. Book of Abstracts of the 63rd Annual Meeting of the European Federation of Animal Science, Bratislava, Slovakia

Thomasen, J. R., C. Egger-Danner, A. Willam, B. Guldbrandtsen, M. S. Lund, and A. C. Sørensen. 2013a. The optimal genomic selection breeding scheme in a small dairy cattle breeding scheme still involves progeny testing. J Dairy Sci (Submitted).

Thomasen, J. R., B. Guldbrandtsen, G. Su, R. F. Brøndum, and M. S. Lund. 2012. Reliabilities of genomic estimated breeding values in Danish Jersey. Animal 6(5):789-796.

Thomasen, J. R., A. C. Sørensen, B. Guldbrandtsen, G. Su, P. Madsen, and M. S. Lund. 2013b. The admixed population structure in Danish Jersey challange accurate genomic predictions. J. Anim Sci.:(Submitted).

Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, future. J Dairy Sci 94(6):3202-3211.

Main findings

The overall aim of this study was to examine how genomic selection is optimized, in terms of monetary genetic gain, discounted profit, rate of inbreeding and the risk of incorporating genomic information in the breeding strategy for a small dairy cattle population. Danish Jersey was used as a model population to illustrate the problems facing the smaller dairy cattle populations (Table 1). The findings in this thesis are expected to be relevant for other small dairy cattle populations as well.

In order to evaluate the efficiency of a genomic breeding scheme it is essential to have knowledge of the reliabilities of the genomic predictions for all traits included in the breeding goal in the specific population. In paper I, the reliabilities of genomic predictions were estimated in Danish Jersey. A Bayesian method was used to estimate the SNP marker effects. The reliabilities assessed as corrected squared correlations between DGV and DRP were, on average across all traits, 0.04 higher than between DRP and the pedigree index for animals without own performance. Averaged across traits, the estimates of reliability of DGVs ranged from 0.20 for validation on the three most recent years of bulls and to 0.42 for expected reliabilities from the posterior distribution of DGV. Reliabilities from the cross validation were on average 0.24. For the individual traits the cross-validation reliability varied from 0.12 (direct birth) to 0.39 (milk production). Bulls whose sires were included in the reference group had an average reliability of 0.25, whereas for bulls whose sires were not included in the reference group the average reliability was 0.05 lower. The reliabilities of DGV were slightly lower than the expected reliabilities estimated from the formula derived by (Goddard, 2009).

Danish Jersey serve as an example of an admixed breed including animals with varied breed proportions of original Danish and US Jersey. This may explain why estimated reliabilities (paper I) in Danish Jersey are lower than expected reliabilities. In paper II, it was evaluated whether the population structure known from the history of Danish Jersey is reflected in the genomic structure currently observed in the population. This was done by comparing LD and persistence of phase between subgroups of Jersey animals with either high proportion of Danish or US origin. Firstly, it was found that the LD across all groups of pure and admixed Danish Jersey animals was lower compared to within groups of either original Danish or original US Jersey animals. Secondly, it was found that the STRUCTURE analysis could retrieve genomically based breed proportions in agreement with the pedigree based estimates of breed proportions. However, including

the STRUCTURE inferred breed proportions in a random regression prediction model for the trait udder health, did not improve the reliabilities of the genomic predictions compared to a basic genomic model. The inclusion of the population structure was based on average marker information measured across the entire genome. This might be too naïve a measure to model the differences in marker allele effects and therefore the reason why reliabilities did not improve. Current models may already account for heterogeneous breed origin.

In paper III, a multitude of breeding schemes for Danish Jersey were studied using a deterministic approach. The breeding schemes using genomic information were shown to be superior with respect to monetary genetic gain and discounted profit compared to the conventional breeding scheme. The optimal breeding scheme was characterized by a mixed use of genotyped young bulls and progeny tested bulls both as bull sires and cow sires (hybrid scheme). Strong interaction effects were observed between increased reliabilities of GEBV and more intensive use of young bulls. The turbo scheme, where only young bulls were used, was inferior when the gains in reliability of GEBV were at the same level as for Danish Jersey at present. However, the turbo scheme becomes superior if higher reliabilities of GEBV can be obtained. Using discounted profit as evaluation criterion, the turbo scheme was always superior due to the lower generation interval and the reduction in costs for housing and feeding waiting bulls. The results from paper III demonstrate that low reliabilities of genomic predictions limits the move towards more efficient breeding schemes with more intensive use of young bulls and no use of progeny testing. The low reliability of DGV is, therefore, the major limitation for more efficient breeding schemes.

Based on the findings in paper III, it can be concluded that a main future focus area should be to increase reliabilities of genomic predictions. A solution to this could be to include genotyped cows in the reference population. This aspect was studied in paper IV using a finite locus model in a stochastic approach. Strategies for genotyping cows were evaluated. The effects were assessed for the same two breeding schemes as in paper III: a turbo scheme and a hybrid scheme. Genetic gain was increased by 20% in the turbo schemes and 7% in the hybrid schemes, when cows were added to the reference population. The rate of inbreeding was decreased by 3% (hybrid) and 30% (turbo). The variance of response is higher in the turbo schemes compared to the hybrid schemes. However, to confirm the results for variance of response in the analysis including cows, more replicates of the simulations are needed. Reliabilities of GEBV increased gradually up

to 0.6, as more cows were added to the reference population. In contrast, the increase in reliabilities was much lower (from 0.19 to 0.27), when 60 progeny tested bulls were added to the reference population yearly. Genotyping of cows is therefore an effective way to increase reliabilities of genomic breeding values and hence increase genetic gain in a population that is limited in terms of small reference populations of progeny tested bulls.

General discussion

Reliabilities of genomic predictions

Accurate genomic predictions are a prerequisite for running an efficient genomic breeding scheme (paper III) and increase the monetary genetic gain, especially in small dairy cattle populations, where reliabilities of genomic predictions are low. There are multiple ways to increase the reliabilities of genomic predictions. Either by improving the prediction model with use of existing information, increasing the marker density or by enlarging the reference population. In paper I, it was found that use of genomic information only increased reliabilities in the present Danish Jersey population with 4 percentage points compared to parent average reliabilities. In paper II, it was proposed, that inclusion of the population structure by use of existing information in a random regression model could increase reliabilities of genomic predictions. However, using the inferred breed proportions of old Danish versus US Jersey from the SNP markers did not improve genomic prediction reliability (paper II). There are two possible reasons for this. One is that the 54K marker panel used in this study may not be dense enough to accurately capture and account for the population structure (de Roos et al., 2008). Another reason is that inclusion of the population structure in the prediction model as one average level across the entire genome might be not efficient to model the differences in marker allele effects. Prediction models that are able to pick up a more detailed population structure at individual marker level and, hence, strengthen the LD between available markers and QTL in heterogeneous populations should be a future focus area. Simulation studies have shown that prediction models using haplotypes instead of single SNP markers improve reliabilities of genomic predictions (Villumsen et al., 2009). In a recent study using Holstein data, (Cuyabano, 2013) found that the use of a haplotype model in combination with denser marker maps (777K) instead of a model based on single SNP markers resulted in some increase in reliabilities. Increasing the marker density from 54K to 777K using single marker prediction only resulted in minor improvement of the prediction ability (Su et al., 2009). Better clustering of haplotypes might be a way to further improve this method (Edriss et al., 2013). Inclusion of the US Jersey data in the analysis (Table 1) might also increase the probability to capture the breed specific marker effects.

Enlarging the reference population can either be achieved by combining Jersey sire reference populations or creating an across breeds reference with e.g. by combining with the Holstein reference. Formation of a global Jersey sire reference population is the

most cost efficient way to increase reliabilities of genomic predictions. Collaborative efforts in other populations have already demonstrated the benefits by such efforts (Lund et al., 2011; Wiggans et al., 2011) (Table 1). Seen from the perspective of a small dairy cattle population, the main competitor is increasingly going to be the Holstein breed rather than other Jersey populations. This argues in favor of increased sharing and cooperation between the Jersey populations. A merge of the existing Jersey sire reference populations would increase the reference population to about 10,000 sires (Table 1). However, inclusion of sires and cows from different climate zones and production systems in a global reference population requires development of models that take account of the differences in population structures and environmental conditions.

The results from the simulation study in paper IV show that inclusion of cows in the reference population is an effective and quick way to increase reliabilities of genomic predictions. Other studies support that cows as reference population for genomic prediction are feasible (Buch et al., 2011; Mc Hugh et al., 2011; Zhou et al., 2013). Already after two years of selection and inclusion of 4,000 genotyped cows in the reference the reliability of GEBV increased from 0.2 to 0.4. However, in practice more genotyped cows might be needed to obtain the same increase in reliabilities. The discounted profit by genotyping cows was not evaluated in this study. Egger-Danner et al. (2012) evaluated the economic consequences for Fleckvieh of genotyping cows. They found a positive profit ranging from 2.6% to 7.1%, depending on the number of genotyped cows. They used a cost of 50€ per genotyped cow. However, the cost per genotyped SNP is decreasing, and the availability of cheaper low density chips that are optimized for imputation (Wiggans et al., 2013), increasingly makes genotyping of cows a cost efficient strategy to increase reliabilities of genomic prediction. A few cattle breeding organizations, has already started the process of including genotyped cows in the reference population (Table1). However, the question remains, if low-density genotyped cows are valuable as part of the reference population or whether higher-density information is needed without imputation.

From a Nordic perspective, an alternative to a global reference population is a common multi-breed reference population, which would facilitate the possibility of using the entire Nordic registered cow population actively for prediction in all the three Nordic dairy cattle breeds: Holstein, Nordic Red and Jersey. It is expected that GxE effects will be of minor influence in a Nordic multi-breed compared to a global Jersey reference population.

However, the differences in population structures are, in contrast to within breed predictions, expected to be larger and more advanced models will be needed before this approach becomes practically feasible. Use of multi-breeds reference populations has so far only resulted in minor improvements (Erbe et al., 2012), even with use of high density marker panels. The reason is that this approach requires prediction models that are able to capture SNP marker effects that are in LD with QTL both within and across genetically distant breeds. The optimal use of heterogeneous reference populations still needs more research attention in order to improve genomic predictions.

Optimal genomic breeding scheme for a small population

In a small dairy cattle breed, where reliabilities of genomic predictions are low, the optimal breeding scheme is characterized by a mixed use of genotyped young bulls and proven bulls (paper III). This finding is in contrast to previous studies of optimal genomic breeding schemes where turbo schemes have been shown to be genetically superior (Buch et al., 2012; de Roos et al., 2011; Lillehammer et al., 2011). In these studies reliabilities of genomic predictions were, however, higher than the reliabilities used in paper III. Accordingly, we also find that young bull schemes become genetically superior when reliabilities of genomic predictions are increased. There are strong positive interactions between increased reliability of genomic predictions and a more intensive use of young bulls. Increasing the reliability without any adjustment to the breeding plan only has a minor effect on genetic gain. Therefore, optimization of the breeding scheme is a process that needs to be re-evaluated if and when the value of genomic information changes.

The simulated breeding schemes in paper III and IV were chosen to mimic the Danish Jersey population, as an example of a small dairy cattle population. As the same breeding scheme parameters were used in the two simulation studies, it allows us to compare the increase in annual monetary genetic gain in the deterministic (paper III) and the stochastic approach (paper IV). However, one major difference between the two methods is the way the reliabilities were modeled. In paper III, the reliabilities of GEBV were modeled as a selection index with a fixed percentage added to parent average, for the whole evaluation period of 15 years. In contrast, the reliabilities in paper IV were based on the outcome of the breeding scheme tested, and increased over time as more information was added to the reference population. In paper III a fixed value of 5% gain in reliability in the standard genomic breeding scheme was used based on the finding in paper I. For the

bull schemes (paper IV) reliabilities increased from 10% and up to 20% over the 15 years of evaluation. The monetary genetic gains were 14% to 30% lower in the stochastic simulations (paper IV) depending on breeding scheme. The second main difference between the two methods is that the results from the stochastic simulations take account of the Bulmer effects, which reduces the genetic variance due to selection. Interaction effects between increased reliabilities of genomic predictions and more intensive use of young bulls were found in both studies. The results in this thesis show that a combined use of the two simulation methods provides a stronger basis for the search of the optimal breeding scheme both from a genetic and economic perspective and according to inbreeding and variance of response (risk).

Using genotyped cows in the reference population will shift the focus of the breeding scheme from progeny testing of bulls towards an optimization of the cows entering the reference population with respect to the number and the genetic relationship to the existing reference animals. The optimal number of progeny tested bulls might not be the same in a breeding scheme using genotyped cows. This would therefore be an interesting area of investigation. Also, maximizing the use of the genotyped cows is expected to increase emphasis on more widespread use of female reproduction technologies such as MOET and OPU as the bull dams can be selected with higher accuracy. In conventional breeding schemes, widespread use of MOET results in high rates of inbreeding. In a genomic selection schemes the adverse effects of MOET on rates of inbreeding are expected to be lower (Pedersen et al., 2012).

Inbreeding

Information about the Mendelian segregation through the marker information leads to more efficient selection within families. GS is therefore expected to reduce the rate of inbreeding per generation compared to traditional BLUP selection assuming the same breeding scheme (Deatwyler et al., 2007). Estimates of inbreeding were only obtainable from the stochastic simulation study in paper IV. The increase in rate of inbreeding per generation was investigated for the breeding schemes: conventional, hybrid and turbo. The inbreeding rates varied from 1.14% up to 2.15% per generation. In general, the turbo schemes provided lower rates of inbreeding per generation compared to schemes with mixed use of young bulls and progeny tested bulls. This is to some extend surprising as the average reliability of selection of candidates in schemes with use of only young bulls are

lower. This is expected to favor co-selection of relatives in turbo schemes and thereby increase inbreeding (de Roos et al., 2011). In our case, however, only a few progeny tested bulls were used intensively in both the conventional and hybrid schemes, leading to a high inbreeding rate in these schemes. The lowest rate of inbreeding was seen for the turbo schemes using cows in the reference. The main reason for this is the higher reliability in the scenarios including genotyping of cows. In general in other studies inbreeding rates are in general observed to be lower (de Roos et al., 2011; Pryce and Daetwyler, 2012). However, these results were based on larger populations and higher reliabilities of genomic predictions. More replicates of the study in paper IV will be needed in order to confirm the difference between the evaluated scenarios.

None of the tested breeding schemes resulted in inbreeding rates below 1% per generation, as recommended by FAO, in order to avoid undesirable effects of inbreeding. Lowering the inbreeding rates in a small population will require that actions are taken to use genetically more diverse group of bulls. A solution is to control inbreeding by minimizing the co-ancestry of parents using of optimum contribution selection (Meuwissen, 1997). The inbreeding rates presented in this study were estimated from the pedigree. However, regions of the genome containing QTL with large effects will be applied to a more intensive selection pressure, when selection is based on marker information. This tends to increase local inbreeding at important QTL sites, and result in hitch-hiking effects over large genome regions. Therefore, a measure of genome based inbreeding rates might be a more accurate measure of the future true inbreeding in a breeding scheme using genomic selection (Pedersen et al., 2010; Sonesson et al., 2012) Accordingly, optimum contribution selection should also be based on genomic information as proposed by (Sonesson et al., 2012).

Conclusions and perspectives

This study has shown that small dairy cattle populations, exemplified by Danish Jersey, are challenged because of the low reliabilities of genomic predictions. The results have also demonstrated that low reliabilities of genomic predictions limits the possibilities for moving towards more efficient breeding schemes with more intensive use of young bulls. An important focus area for the smaller dairy cattle breeds is therefore to seek for ways to increase the reliabilities of genomic predictions.

As described in the introduction, the reliabilities of genomic predictions depend on several factors. When the reference population is small, there is a high marginal effect on reliability by enlarging the reference population. This can be done either by adding more genotyped bulls with daughter proof or by including genotyped cows directly.

It is recommended that initiatives are taken to form a global Jersey reference population. This solution is very cost efficient, as it only requires exchange of all ready genotyped bulls. Exchange of genotyped bulls with daughter proof is already established for Holstein, Fleckvieh, and Brown Swiss (Table 1), and has been proven to increase reliabilities of genomic predictions (Lund et al., 2011).

It is also recommended to include genotyped cows in the reference population. The results in this study show a clear increase in reliabilities of genomic predictions, an increase in monetary genetic gain, and a reduction in the rate of inbreeding compared to when the reference population was updated only with bulls. Furthermore, the monetary genetic gain can be increased by running a turbo scheme provided that the gain in reliability due to genomic information is increased by at least 10%. This is fulfilled by genotyping at least 2,000 registered cows each year.

References

Boichard, D., F. Guillaume, A. Baur, P. Croiseau, M. N. Rossignol, M. Y. Boscher, T. Druet, L. Genestout, J. J. Colleau, L. Journaux, V. Ducrocq, and S. Fritz. 2012. Genomic selection in French dairy cattle. Anim. Prod. Sci. 52(3):115-120.

Bouquet, A., and J. Juga. 2013. Integrating genomic selection into dairy cattle breeding programmes: a review. Animal (7)5: 705–713.

Brøndum, R. F., E. Rius-Vilarrasa, I. Strandθn, G. Su, B. Guldbrandtsen, W. F. Fikse, and M. S. Lund. 2011. Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. J Dairy Sci 94(9):4700-4707.

Buch, L. H., M. Kargo, P. Berg, J. Lassen, and A. C. Sørensen. 2011. The value of cows in reference populations for genomic selection of new functional traits. Animal 6(6):880-886.

Buch, L. H., M. K. Sørensen, P. Berg, L. D. Pedersen, and A. C. Sørensen. 2012. Genomic selection strategies in dairy cattle: Strong positive interaction between use of genotypic information and intensive use of young bulls on genetic gain. Journal of Animal Breeding and Genetics 129(2):138-151.

Bulmer, M. G. 1971. The Effect of Selection on Genetic Variability. The American Naturalist 105(943):201-211.

Christensen, O., and M. Lund. 2010. Genomic prediction when some animals are not genotyped. Genetics Selection Evolution 42(1):2.

Cuyabano, B. C. D. 2013. Personal communication.

de Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage Disequilibrium and Persistence of Phase in Holstein-Friesian, Jersey and Angus Cattle. Genetics 179(3):1503-1512.

de Roos, A. P. W., C. Schrooten, R. F. Veerkamp, and J. A. M. van Arendonk. 2011. Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. J Dairy Sci 94(3):1559-1567.

Deatwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. in .

Edriss, V., R. Fernando, G. Su, M. Lund, and B. Guldbrandtsen. 2013. The effect of using genealogy-based haplotypes for genomic prediction. Genetics Selection Evolution 45(1):5.

Egger-Danner, C., H. Schwarzenbacher, and A. Willam. 2012. Genotyping of cows to speed up the availability of GEBVs for direct health traits - genetic and economic aspects based on Fleckvieh (Simmental) Austria. Book of Abstracts of the 63rd Annual Meeting of the European Federation of Animal Science, Bratislava, Slovakia

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J Dairy Sci 95(7):4114-4129.

FAO. 2007. Global plan of action for animal genetic resources and the interlagen decleration. in Commision on genetic resources for food and agriculture organisation of the united nations, Rome.

Fimland, E. 2005. Bærekraftig forvaltning av husdyrgenetiske ressurer. Nordisk Genbank Husdyr, Ås.

Gao, H., O. Christensen, P. Madsen, U. Nielsen, Y. Zhang, M. Lund, and G. Su. 2012. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. Genetics Selection Evolution 44(1):8.

Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet 10(6):381-391.

Goddard, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136(2):245-257.

Harris, B. L. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 92(2):433-443.

Hermansen, J. 2012. Challenges for livestock production and animal science in a low carbon economy. Book of Abstracts of the 63rd Annual Meeting of the European Federation of Animal Science, Bratislava, Slovakia,

König, S., H. Simianer, and A. Willam. 2009. Economic evaluation of genomic breeding programs. J Dairy Sci 92(1):382-391.

Lillehammer, M., T. H. E. Meuwissen, and A. K. Sonesson. 2011. A comparison of dairy cattle breeding designs that use genomic selection. J Dairy Sci 94(1):493-500.

Lund, M., A. de Roos, A. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genetics Selection Evolution 43(1):43.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. L. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS One 4(4):e5350.

Mc Hugh, N., T. H. E. Meuwissen, A. R. Cromie, and A. K. Sonesson. 2011. Use of female information in dairy cattle genomic breeding programs. J Dairy Sci 94(8):4109-4118.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157(4):1819-1829.

Pryce, J. E., and H. D. Daetwyler. 2012. Designing dairy cattle breeding schemes under genomic selection: a review of international research. Anim. Prod. Sci. 52(3):107-114.

Pryce, J. E., B. Gredler, S. Bolormaa, P. J. Bowman, C. Egger-Danner, C. Fuerst, R. Emmerling, J. S+Âlkner, M. E. Goddard, and B. J. Hayes. 2011. Short communication: Genomic selection using a multi-breed, across-country reference population. J Dairy Sci 94(5):2625-2630.

Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics 123(4):218-223.

Sørensen, A. C., M. K. Sørensen, and P. Berg. 2005. Inbreeding in Danish Dairy Cattle Breeds. J Dairy Sci 88(5):1865-1872.

Sørensen, M. K., E. Norberg, J. Pedersen, and L. G. Christensen. 2008. Invited Review: Crossbreeding in Dairy Cattle: A Danish Perspective. J Dairy Sci 91(11):4116-4128.

Stachowicz, K., M. Sargolzaei, F. Miglior, and F. S. Schenkel. 2011. Rates of inbreeding and genetic diversity in Canadian Holstein and Jersey cattle. J Dairy Sci 94(10):5160-5175.

Su, G., R. F. Brøndum, P. Ma, B. Guldbrandtsen, G. P. Aamand, and M. S. Lund. 2012. Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. J Dairy Sci 95(8):4657-4665.

Su, G., B. Guldbrandtsen, V. R. Gregersen, and M. S. Lund. 2010. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. J Dairy Sci 93(3):1175-1183.

Sved, J. A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theoretical population biology 2:125-141.

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92(1):16-24.

Villumsen, T. M., L. Janss, and M. S. Lund. 2009. The importance of haplotype length and heritability using genomic selection in dairy cattle. Journal of Animal Breeding and Genetics 126(1):3-13.

Wiggans, G. R., T. A. Cooper, C. P. Van Tassell, T. S. Sonstegard, and E. B. Simpson. 2013. Technical note: Characteristics and use of the Illumina BovineLD and GeneSeek Genomic Profiler low-density bead chips for genomic evaluation. J Dairy Sci 96(2):1258-1263.

Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, future. J Dairy Sci 94(6):3202-3211.

Publication list

Peer-reviewed publications outside the scope of the thesis

2013

Effects of *Bos taurus* autosome 9-located quantitative trait loci haplotypes on the disease phenotypes of dairy cows with experimentally induced *scherichia coli* mastitis. Khatun, M.; Sørensen, P.; Jørgensen, H. B. H.; Sahana, G.; Sørensen, L. P.; Lund, M. S.; Ingvartsen, K. L.; Buitenhuis, B.; Vilkki, J.; Bjerring, M.; **Thomasen, J. R.**; Røntved, C. M..
Journal of Dairy Science, **Vol**. 96(3), 2013, 1820-1833.

2011

QTL explaining variation in production traits and udder health in the Danish Holstein population. Thomsen, H.; **Thomasen, J. R.**; Guldbrandtsen, B.; Lund, M. S.
Archiv fuer Tierzucht, **Vol**. 54(4), 2011, s. 348-359.

2008

Quantitative Trait Loci Affecting Calving Traits in Danish Holstein Cattle. **Thomasen, J R**; Guldbrandtsen, B; Sørensen, P; Thomsen, B; Lund, M S.
Journal of Dairy Science, **Vol**. 91(5), 2008, 2098-2105.

Pathogen-Specific Effects of Quantitative Trait Loci Affecting Clinical Mastitis and Somatic Cell Count in Danish Holstein Cattle.
Sørensen, L. P.; Guldbrandtsen, B; **Thomasen, J. R.**; Lund, M S.
Journal of Dairy Science. **Vol** 91(6). 2493-2500.

2007

Detection of Quantitative Trait Loci Affecting Lameness and Leg Conformation Traits in Danish Holstein Cattle. Buitenhuis, A. J.; Lund, M. S.; **Thomasen, J. R.**; Thomsen, B.; Nielsen, V. H.; Bendixen, C.; Guldbrandtsen, B.
Journal of Dairy Science, **Vol**. 90, Nr. 1, 2007, s. 472-481.

Conference abstracts in relation to the thesis

2013

Genotyping cows for the reference increases reliability of genomic predictions in a small breed.

Thomasen, J. R.; Sørensen, A. C.; Lund, M. S. Guldbrandtsen, B..
Book of Abstracts of the 64th Annual Meeting of the European Association for Animal Production.

2011

Analysis of subpopulation structure in Danish Jersey.

Thomasen, J. R.; Sørensen, A. C.; Brøndum, R. F.; Lund, M. S.; Guldbrandtsen, Bernt.
Book of Abstract of the 62nd Annual Meeting of the European Federation of Animal Science. **Vol**. 17. Wageningen Academic Publishers, 2011. p 27.

2010

Reliabilities of Genomoc Estimated Breeding Values in Danish Jersey.

Thomasen, J R; Guldbrandtsen, B; Su, G; Brøndum, R F; Lund, M S.
Poster session presented at World Congress on Genetics Applied to Livestock Production, WCGALP, Leipzig, Germany.