Livestock Science 166 (2014) 101–110

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

Genomic evaluation of cattle in a multi-breed context ☆

Mogens Sandø Lund *, Guosheng Su, Luc Janss, Bernt Guldbrandtsen, Rasmus Froberg Brøndum

Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Faculty of Science and Technology, Aarhus University, Blichers Allé 30, Tjele 8830, Denmark

ARTICLE INFO

Keywords:
Multi-breed
Genomic Prediction
Linkage disequilibrium

ABSTRACT

In order to obtain accurate genomic breeding values a large number of reference animals with both phenotype and genotype data are needed. This poses a challenge for breeds with small reference populations. One option to overcome this obstacle is to use a multi-breed reference population. However, combining populations across breeds is not straightforward due to differences in linkage disequilibrium structure and weak relationships between breeds. This study offers a review of the available literature on the use of reference populations compiled from different cattle breeds. Results show that the effect of multi-breed reference populations on the accuracy of genomic prediction is highly affected by the genetic distance between breeds. When combining populations of the same breeds from different countries, large increases in accuracy are seen, whereas for admixed populations with some exchange of sires, substantial but smaller gains are found. Little or no benefit is found when combining distantly related breeds such as Holstein and Jersey and using the widely used genomic BLUP model. By using more sophisticated Bayesian variable selection models that put more focus on genomic markers in strong linkage disequilibrium with causative variants in combination with denser markers sets or functional subsets of markers, it is however possible to utilize information across distantly related breeds to increase the accuracy of genomic prediction. The further development of multi-breed genomic prediction models offers not only increases in the accuracy of genomic breeding values for small breeds, but will also give a stronger persistence of the accuracy over generations within larger breeds.

1. Introduction

A key factor for a successful genomic selection scheme is the ability to accurately predict genomic breeding values (GEBV). This requires a reference population from which marker effects can be estimated precisely. The accuracy of the resulting GEBV relies heavily on the number of individuals in the reference population (Goddard, 2009).

Small dairy cattle populations are often restricted by 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 relative to breeds with large reference populations. Thomasen et al. (2014) showed that low reliabilities of genomic prediction are the single most important factor limiting genetic gain in smaller populations with more

☆ This paper is part of the special issue entitled: Genomics Applied to Livestock Production, Guest Edited by Jose Bento Sterman Ferraz.

* Corresponding author. Tel.: +45 87 15 80 24.
E-mail addresses: mogens.lund@agrsci.dk (M.S. Lund), guosheng.su@agrsci.dk (G. Su), luc.janss@agrsci.dk (L. Janss), bernt.guldbrandtsen@agrsci.dk (B. Guldbrandtsen), rasmusf.brondum@agrsci.dk (R.F. Brøndum).

http://dx.doi.org/10.1016/j.livsci.2014.05.008
1871-1413/

intensive use of young bulls without a progeny test. Therefore, an important objective for smaller dairy cattle breeds in order to stay competitive should be to increase reliabilities of GEBVs. Joining two or more populations from the same or different breeds into a common reference population is an obvious way to achieve this, as it is expected to increase the accuracies of GEBV without adding extra costs for genotyping and phenotyping.

Studies such as Goddard (2009) that describe the effectiveness of genomic prediction, typically assume a panmictic population. However, many populations are structured to some degree. Population structure can appear at a number of levels. At one end of the scale national or regional populations of cattle belonging to the same breed often have close genetic ties and are effectively nearly panmictic. An example of this would be the Holstein breed of dairy cattle, where extensive global exchange of genetic material occurs. However, even within what is nominally the same breed such as the Jersey population of Denmark and North America, Thomasen et al. (2012) observed differences in linkage disequilibrium (LD) despite extensive use particularly of North American genetic material in the Danish Jersey population. Populations belonging to different breeds of cattle are an extra step away from truly panmictic. Here genetic similarities are determined by the time since separation of the breeds and the extent of subsequent exchange of genetic material between the breeds. In particular, if genetic exchange has been limited between the breeds, then similarities in LD between the populations will be due to conserved LD from the shared ancestral population. As a consequence, the formula describing the effectiveness of genomic prediction (Goddard, 2009) cannot be used directly in the multi-breed context.

When predicting GEBV, the basic assumption is that all causal variants are in LD with genomic markers. Linkage disequilibrium can arise in a number of ways: mutations are born in LD, through genetic drift (Hill and Robertson, 1968; Sved, 1971), selection can generate LD between alleles both affecting a trait (Felsenstein, 1965; Bulmer, 1971) and through admixture, migration or introgression between populations with different allele frequencies.

New mutations are effectively unique by nature. A mutation arises in one particular genome – in LD with all the alleles in that genome – regardless of linkage distance. However, LD rapidly dissipates except within short distances for which LD can persist for very long time. In an infinite randomly mated population the LD parameter shrinks by a factor (1-R) in each generation, where R is the recombination rate between the loci involved. Effectively only LD between fairly closely linked loci persists for more than a few generations.

In livestock populations genetic drift constantly generates large LD (also termed 'linkage'), and by strong selection and high usage of individuals selected for breeding this linkage may dominate the genomic structure of the populations. Consequently, the genomic markers will predict the family structure and multiple co-segregating causative variants. The LD generated will in most cases be weak for each pair of alleles, but as it is generated for all pairs of alleles regardless of whether they are linked or not, it is generated for a large number of pairs of alleles. However, as much of this LD is between unlinked or loosely linked alleles, it dissipates very quickly.

Figure 3 in Habier et al. (2010) illustrates these differences. As one moves away from the population with the phenotypic information the ability of genomic prediction declines first steeply due to loss of loose linkage (relationship). Later it declines in a nearly linear (on the number of generations) fashion even into another population reflecting slow decay of LD between more tightly linked markers. In contrast, the accuracy of conventional pedigree based BLUP declines much more quickly. Therefore, many recent findings show that the accuracy of GEBV depends heavily on family relationships between the reference and test populations (Habier et al., 2010; Daetwyler et al., 2012; Wientjes et al., 2013; De Los Campos et al., 2013).When reference and test individuals are from different populations, relationships may be relatively weak. Assuming that we know the causal variants and use them to estimate genomic relationships, the phenotypes in one population will contribute to the accuracy of GEBV in the other population (De Los Campos et al., 2013). However, because we calculate genomic relationships using markers in imperfect LD with causal variants, there is a loss of information. With close relationships the loss is small but for weak relationships the loss is considerable (De Los Campos et al., 2013). Therefore, the increase in accuracy of GEBV when combining populations into a single reference, will depend on how genetically related the populations are, what proportion of the LD information in genomic predictions originates from tightly linked markers and from relationship information. The former is to some degree shared between related breeds making composite reference populations usable. The latter is not, as information on individuals from one breed will not be useful to predict relationship information in another breed. The magnitude of LD is largely dictated by historic effective population size. For tightly linked loci, Sved (1971), Eqn. 7, using an argument akin to a coalescent argument, found that for small recombination rates, R, the expected squared correlation between pairs of markers is

$$E(r^2) = 1/(1+4N_eR),$$

where N_e is the long-term effective population size.

On the other hand the amount of information due to relationship largely depends on recent effective populations. For unlinked loci Sved's equation predicts $E(r^2)=1/(1+12N_e)$. The effective population sizes now are much smaller than historic effective population sizes (e.g. Hayes et al., 2003). This tends to increase the significance of relationship information relative to LD information.

Two conclusions arise from this argument: First of all, the use of a composite reference population will only be successful if either an appreciable fraction of the information explained by the markers arises from tight linkage between markers and causative variants or the breeds making up the composite reference have relatively recent genetic exchange between them. Second, method development (including choice of marker panels) for prediction based on composite reference populations needs to pay

special attention to exploiting information from markers tightly linked to the causative variants.

2. Statistical models

2.1. Single-trait GBLUP

The most straightforward approach for multi-breed prediction is to apply regular single-trait GBLUP (VanRaden, 2008). This approach will mainly be sensible when populations are relatively close, e.g., subpopulations of the same breed. The prerequisite for GBLUP to function well is to have many recent relationships, i.e. use of a significant amount of common sires or grandsires. A limitation of GBLUP for pooling across breed data is that relationships across breeds are typically small. When these relationships get very small the correlation between genomic relationships at causal loci and genomic relationships calculated from genome wide markers becomes very low (De Los Campos et al., 2013). Consequently, they essentially become "noise" and can cause estimation problems.

2.2. Multi-trait GBLUP

Multi-trait GBLUP models can be used when the phenotypes measured in different breeds are considered different traits (Olson et al., 2012; Zhou et al., 2013). This approach can accommodate for phenotypes not being measured in exactly the same way, for possible SNP by population (genetic background) interactions and SNP by environment interactions. The multi-trait GBLUP model is straightforwardly developed from a classical pedigree-based multi-trait model where the pedigree relationship matrix is replaced by a genomic relationship matrix. An equivalent model to the multi-trait GBLUP model is a multi-breed SNP-BLUP model which has uniform variance across the genome for SNP effects within traits and uniform covariance across the genome for SNP effects between traits. Sørensen et al. (2012) used such a multi-trait SNP-BLUP implementation on mastitis related traits in Danish Holstein cows and showed that such a model allows for the genomic correlation to locally deviate somewhat from the overall SNP correlation. In general, however, a limitation of this multi-trait GBLUP/SNP-BLUP model remains that the covariance across the genome is assumed uniform.

2.3. Bayesian approaches

Application of Bayesian variable selection methods has the potential to improve multi-breed evaluations in two ways. Firstly, as Bayesian variable selection methods allocate more genomic variance to markers that show a high association to phenotypes, they may be able to better separate the linkage and LD contributions in genomic predictions. Consequently predictors are more based on LD, which is expected to improve the sharing of information across populations or breeds. Secondly, Bayesian approaches can alleviate the strong assumptions in GBLUP approaches that SNP variances and covariances are uniform across the genome. This will be a great advantage when the LD phase between markers and causative variants are different in the combined breeds or a causative variant is only segregating in one of the breeds.

The rationale that Bayesian methods construct predictors more based on LD can be verified in situations where relationships get weaker. Habier et al. (2013) demonstrated in a simulation study that the Bayesian methods have better 'persistency' to predict several generations ahead, and this effect must be based on stronger capturing of LD instead of linkage. Also in real data this has been shown, for instance, Gao et al. (2013) obtained improved predictions from Bayesian mixture models for animals that had no parents with information in the reference data. The LD between markers and causative variants can be further increased by using haplotypes of linked markers close to the causative variant.

Some work has been done on multi-trait Bayesian genomic models. Calus and Veerkamp (2011) presented multi-trait mixture models, which are variations of BayesB and BayesCπ, and applied them to simulated single breed data; however the covariance for SNPs across the genome in their models is the same and based on a prior pedigree based multi-trait BLUP analysis of the genetic (genomic) covariance of the traits. So far, no work has been done to alleviate these assumptions of constant variance and covariance, and to apply such multi-trait models to the case of multi-breed evaluation.

3. Results on joining populations into a common reference population

3.1. Dairy cattle breeds

A number of studies have compared the predictive ability of genomic models trained in a joint reference population by combining populations of the same breed or populations of different breeds. Table 1 presents the results from genomic predictions by combining different dairy cattle populations. From the results it is clear that joining populations of the same breed increases the accuracies of GEBVs. This is particularly clear in the cases where the exchange of genetic material between populations is large as for Holstein Friesian (HF). Here large improvements are realized when combining populations in North America (Schenkel et al., 2009; Vanraden et al., 2012) and in the EuroGenomics collaboration (Lund et al., 2010). Similarly, genomic predictions for Chinese HF using a joint reference with Nordic HF increases accuracies substantially (Zhou et al., 2013). This could be regarded as more surprising since environmental and management factors are very different between Chinese and Nordic Holstein populations.

Another category of results are from joining breeds that are more distinct but admixed in the sense that bulls to some degree have been used across the breeds. This is particularly clear for the Nordic red breeds: Danish Red (DR), Swedish Red (SR), Finnish Ayrshire (FA), and Norwegian Red (NR). In these breeds principle components analysis clearly shows the consequence of a high exchange of genetic material between SR and FA as well as the use of SR bulls in DR (Kadri, 2014). As a consequence SR and FA

Table 1
Increase in accuracy/reliability when using joint dairy reference compared to a single reference population for milk-, protein and fat yield, fertility and Somatic Cell Score (SCS). All studies are performed using 54 k genotype data. Ref1 is the breed and country of origin for the single reference population, and Ref2 is the breeds and countries of origin for the joint reference. Reference sizes are given as number of bulls (+number of cows). R or R^2 in column five states whether the original paper uses the correlation or squared correlation to measure the validation accuracy. Breed codes: HF=Holstein-Friesian, JE=Jersey, BS=Brown-Swiss, DR=Danish Red, SR=Swedish Red, FA=Finnish Ayrshire, NR=Norwegian Red, VR=Danish/Swedish/Finnish Red, MB=Montbéliarde, NM, Normande. Country Codes: US=United States, IT=Italy, CA=Canada, UK=United Kingdom, CH= Czech Republic, AT=Austria, DE=Germany, NL=Netherlands, FR=France, CI=China, NO=Nordic, AS=Australia. Trait codes: NRR= Non Return Rate, CR=Calving Rate, UHI=Udder Health Index, DPR=Daughter Pregnancy Rate, IFC=Interval between Calving and First insemination, FC=Fat Content.

Ref1	Ref2	Ref1 size	Ref2 Size		Milk	Protein	Fat	Fertility	SCS	Method	Citation
HF (US)	HF (US+IT+CA+UK)	10,534+22800	18,508+22800	R^2	2.1	2.3	2.3	3.8^{DPR}	3.5	GBLUP	Vanraden et al. (2012)
BS (US)	BS (CH+DE+AT)	812+374	1682 +374	R^2	5.3	2.7	1.1	-3^{DPR}	0.8	GBLUP	Vanraden et al. (2012)
HF (CA)	HF (US)	1097	4127	R^2	9	8	12	3	10	GBLUP	Schenkel et al. (2009)
HF (NO)	HF (NO+DE+FR+NL)	3077	10,880	R^2		13		5^{NRR}	13	GBLUP	Lund et al., (2011)
HF (DE)	HF (NO+DE+FR+NL)	3676	14,479	R^2		2		10^{NRR}	15	GBLUP	Lund et al., (2011)
HF (FR)	HF (NO+DE+FR+NL)	3071	12,078	R^2		4		10^{CR}	8	QTL-BLUP	Lund et al., (2011)
HF (NL)	HF (NO+DE+FR+NL)	3472	9618	R^2		5		3^{IFC}	8	Bayesian 2-mixture	Lund et al., (2011)
HF (CI)	HF (CI+NO)	13+1572	4411+1572	R^2	29	32	25			Multitrait GBLUP	Zhou et al. (2013)
HF (CI) cows	HF (CI+NO)	80+1572	4478+1572	R^2	11	5	5			Multitrait GBLUP	(Zhou et al., 2013)
DR	VR	929	3735	R^2	2	4	1	-3^{NRR}	2^{UHI}	Bayesian	Brøndum et al. (2011)
SR	VR	1551	3735	R^2	9	18	7	9^{NRR}	6^{UHI}	Bayesian	Brøndum et al. (2011)
FA	VR	1562	3735	R^2	12	13	6	5^{NRR}	10^{UHI}	Bayesian	Brøndum et al. (2011)
VR	VR+NR	3367	5717	R	1	1	2	0^{NRR}	2^{UHI}	GBLUP	Zhou et al. (2014a)
NR	VR+NR	2076	5433	R	5	8	5	2^{NRR}		GBLUP	Zhou et al. (2014a)
VR	VR+HF (NO)	3437	6552	R	1.4	1.1	1.0	0.4^{NRR}	0.4	GBLUP	Zhou et al. (2014b)
DR	VR+HF (NO)	3437	6552	R	5	3	2	2^{NRR}	1	GBLUP	Zhou et al. (2014b)
SR	VR+HF (NO)	3437	6552	R	2	2	2	0^{NRR}	0	GBLUP	Zhou et al. (2014b)
FA	VR+HF (NO)	3437	6552	R	1	0	0	0^{NRR}	0	GBLUP	Zhou et al. (2014b)
HF(NO)	VR+HF (NO)	3115	6552	R	0.6	0	0.4	-0.4^{NRR}	0.4	GBLUP	Zhou et al. (2014b)
MB	MB+NM+HF (FR)	950	4896	R^2	2		6^{FC}	0^{CR}		GBLUP	Karoui et al. (2012)
NM	MB+NM+HF (FR)	970	4896	R^2	2		0^{FC}	0^{CR}		GBLUP	Karoui et al. (2012)
HF (FR)	MB+NM+HF (FR)	2976	4896	R^2	1		1^{FC}	0^{CR}		GBLUP	Karoui et al. (2012)
BS (US)	BS+JE+HF (US)	506	7168	R^2	4	3	4	-1^{DPR}	−1	Non-linear GBLUP	Olson et al. (2012)
JE(US)	BS+JE+HF (US)	1361	7168	R^2	−3	−2	−4	0^{DPR}	0	Non-linear GBLUP	Olson et al. (2012)
HF(US)	BS+JE+HF (US)	5331	7168	R^2	−4	−3	−3	0^{DPR}	0	Non-linear GBLUP	Olson et al. (2012)
HF (AS)	HF+JE (AS)	1897	2351	R	−1	0	0			GBLUP	Erbe et al. (2012)
JE (AS)	HF+JE (AS)	454	2351	R	−3	−2	−3			GBLUP	Erbe et al. (2012)

largely overlap in a plot of the first and second principal components, while there is a smaller overlap between DR and SR. A similar situation is present in NR, which has frequent exchange of genetic material with SR. This structure is clearly favorable for an increase in reliabilities in GEBVs when going from a single breed reference to a joint reference. Generally, the increases when combining these related breeds are substantial, but smaller than combining populations of the same breed. For FA and SR large increases in reliabilities were observed when their reference populations were combined, while the added effect on the reliability for these two breeds by including DR as well was negligible. On the contrary DR had the smallest increase in accuracy when using a multi-breed reference of DR, SR and FA (Brøndum et al., 2011). Similarly, the accuracies for GEBVs in NR increased substantially when Danish, Swedish and Finnish Red (VR) animals were added to the reference (Heringstad et al., 2011; Zhou et al., 2014a).

A third group of studies attempt to join populations of more distantly related breeds. One study combined the three French populations of Holstein, Normande, and Montbéliard (Karoui et al., 2012). This study found a slight increase in reliabilities for production traits of the breed with the smallest population size. However, no increase was found for fertility, where the genetic correlation between the trait-performances measured in different breeds was low. Zhou et al. (2014b) investigated genomic prediction across the Nordic HF and VR populations, and reported that the joint reference population slightly increased the reliability in DSF, but differences were negligible in HF. Among the three sub-populations of VR, accuracies increased more for DR than for SR and FA, because of closer genetic relationships between DR and Nordic HF.

A number of studies (Hayes et al., 2009; Pryce et al., 2011; Olson et al., 2012; Erbe et al., 2012) report on the effect of combining HF with Jersey. Here the relationships across breeds are weak although probably higher for the Australian HF and Jersey, since the Australian Jersey is upgraded to Australian HF by systematic crossing with HF (Pryce et al., 2011). Generally, no improvements are observed in the accuracies of GEBV for HF when Jersey animals are added to the reference population, and for Jersey animals results are similar or worse when using 54 k data and GBLUP methods (Hayes et al., 2009; Erbe et al., 2012). However, when using denser SNP panels, functional subsets of markers or Bayesian methods, increases in accuracy for the Jersey animals have been observed when adding HF to the reference (Erbe et al., 2012). Olson et al. (2012) studied the effect on reliabilities when combining Brown Swiss (BS), Jersey, and HF and using a single trait GBLUP model ($GBLUP_{ST}$), assuming that all data are from one uniform population or a multi-trait GBLUP ($GBLUP_{MT}$), in which SNP effects in different breeds were correlated. Using $GBLUP_{ST}$, the GEBV reliabilities on average increased slightly for BS but decreased for Jersey and HF when the reference populations were combined. When $GBLUP_{MT}$ was used for prediction of protein yield, the negative effects of combining reference populations were not observed and a small positive effect was observed for BS and Holstein while there was no benefit for Jersey.

3.2. Beef cattle breeds

In general, beef cattle have more breeds, but smaller populations than dairy cattle within a country. Therefore, a combined reference population of various breeds to increase the size of the reference population is usually used for genomic prediction. Weber et al. (2012) investigated the accuracy of genomic predictions for six growth and carcass traits for populations including many breeds. Genomic breeding values were predicted using a univariate BayesCπ (Habier et al., 2011) model, and using a dataset comprising purebred animals (2000_BULL) and a dataset comprising crossed animals (USMARC_GPE). Cross-validation was performed by taking one dataset (or a subset) as reference population and the other (or a subset) as test population. The study reported that genomic predictions using multi-breed reference populations were more accurate than those obtained using a single-breed reference population. For example, the accuracies, on average over 5 traits (birth weight, weaning weight, yearling weight, rib eye area and marbling score), were 0.30 for both crossed animals with breed proportion of Angus larger than 0.25 (AN25) and those with breed proportion of Hereford larger than 0.25 (H25) when using the whole 2000_BULL reference population, while the accuracies were 0.17 for AN25 and 0.24 for H25 when using Angus or Hereford single-breed reference population (subset of the 2000_BULL). However, in another study on genomic prediction of weaning weight and yearling weight for purebred animals in the US beef cattle population, based on the purebred data from US 2000 bull project implementing a univariate BayesC or BayesCπ model, Kachman et al. (2013) reported that, for breeds in the reference data, genomic predictions from multi-breed and single-breed reference populations had similar accuracies.

Chen et al. (2013) studied genomic predictions for residual feed intake in Canadian Angus and Charolais beef cattle populations, applying a univariate GBLUP model and a univariate BayesB model. In the first validation scenario where the data was split into reference and test datasets by birth year of the animals (relatively strong relationship between test and reference animals), the combined reference data did not lead to a higher accuracy of genomic prediction than a single-breed reference. However, using an alternative cross-validation where the data was split into reference and test datasets according to sire families (weak relationship between test and reference animals), the combined reference data increased accuracies of 1–2% points in Angus and 3–4% points in Charolais. Bolormaa et al. (2013) assessed the accuracy of genomic predictions for 19 traits including feed efficiency, growth, and carcass and meat quality traits in Australian beef cattle populations, using a GBLUP model and a BayesR model. This study showed that a combined reference population performed better than a single-breed reference population. Using the GBLUP model, the gain in accuracy by moving from the single-breed reference population to the combined reference population was 4% points, averaged over traits and breeds (the paper did not present the improvement from a combined reference population when using BayesR).

3.3. Gain in accuracy relative to own and joint reference size

Gains in accuracy when combining reference data for admixed or distantly related animals depend on the size of the within breed reference populations. When combining Australian HF and Jersey animals little or no gains are seen for HF (Hayes et al., 2009; Erbe et al., 2012) whereas with increased marker density or Bayesian methods gains are seen for the Jersey animals (Erbe et al., 2012). When combining DR, SR and FA with NR the largest gains were seen for NR which had the smallest within breed reference size (Zhou et al., 2014a), and when combining North American BS, Jersey and HF animals the largest gain was also seen for the smallest breed, i.e. BS. Hozé et al. (2014) investigated the effect of own reference size in a study on French Normande animals, which were used in a multi-breed genomic prediction scheme with varying sizes of within breed reference data. Results showed that for small numbers of Normande animals ($N=200,400$) the benefit of adding Holstein and Montbéliarde animals to the reference was larger (3% vs 0.5%) than when a large within breed reference was available ($n=1600$).

Results from Hozé et al. (2014) also showed that the gain in accuracy from the multi-breed reference was much larger for animals without their sire in the reference (12–16% vs 1–2%) showing that the benefit of multi-breed references depends on the relationship between reference and test animals, as was also observed in the study on beef cattle by Chen et al. (2013). According to the results from the above studies, the superiority of combined reference population over single-breed reference population is large if the relationship between the test animals and the within-breed reference animals is weak, and vice versa.

4. Using different genomic models and marker densities

It is clear from the results that with increased genetic distance between the populations being combined into a joint reference population, the increase in accuracy of GEBV is smaller. This is because the LD between markers and causative mutations within populations does not persist across populations. However, conditional on the same causative variants being present and segregating in the combined populations, it should be possible to estimate the effects across populations and thereby increase the accuracy. To achieve this in distantly related breeds at least two technical requirements must be fulfilled. First, the marker density used has to be sufficient to achieve consistent LD between causative variants and markers across breeds. Second, genomic prediction models must allocate more genomic variance to markers in strong LD with the causative variantion.

A few studies compared the accuracy of different statistical models for genomic prediction. Results from these are shown in Table 2. Olson et al. (2012) compared the accuracy

Table 2
Increase in accuracy/reliability when using joint dairy reference compared to a single reference population for milk-, protein and fat yield, fertility and somatic cell score (SCS). Results from different methods as well as different marker densities are shown. Ref1 is the breed and country of origin for the single reference population, and Ref2 is the breeds and countries of origin for the joint reference. R or R^2 in column three states whether the original paper uses the correlation or squared correlation to measure the validation accuracy. Breed codes: HF=Holstein-Friesian, JE=Jersey, BS=Brown-Swiss, NR=Norwegian Red, VR=Danish/Swedish/Finnish Red. Country Codes: US=United States, AS=Australia. Trait codes: NRR= Non Return Rate, UHI=Udder Health Index.

Ref1	Ref2		Milk	Protein	Fat	Fertility	SCS	Method	Citation
54 K									
HF (AS)	HF+JE (AS)	R	−1	0	0			GBLUP	Erbe et al. (2012)
HF (AS)	HF+JE (AS)	R	−1	1	1			BayesR	Erbe et al. (2012)
JER (AS)	HF+JE (AS)	R	−3	−2	−3			GBLUP	Erbe et al. (2012)
JER (AS)	HF+JE (AS)	R	−4	1	2			BayesR	Erbe et al. (2012)
HF (AS)	HF+JE (AS)	R	0	0	1			GBLUP	Hayes et al. (2009)
HF (AS)	HF+JE (AS)	R	−4	0	0			BayesA	Hayes et al. (2009)
JE (AS)	HF+JE (AS)	R	−7	0	1			GBLUP	Hayes et al. (2009)
JE (AS)	HF+JE (AS)	R	4	−1	14			BayesA	Hayes et al. (2009)
VR	NR	R	1	1	2	0^{NRR}	2^{UHI}	GBLUP	Zhou et al. (2014a)
VR	NR	R	2	2	0	1^{NRR}	-7^{UHI}	Bayesian 4-mixture	Zhou et al. (2014a)
NR	VR	R	5	8	5	2^{NRR}		GBLUP	Zhou et al. (2014a)
NR	VR	R	9	13	6	3^{NRR}		Bayesian 4-mixture	Zhou et al. (2014a)
BS (US)	BS+JE+HF (US)	R^2		3				GBLUP	Olson et al. (2012)
JE(US)	BS+JE+HF (US)	R^2		−2				GBLUP	Olson et al. (2012)
HF(US)	BS+JE+HF (US)	R^2		−3				GBLUP	Olson et al. (2012)
BS (US)	BS+JE+HF (US)	R^2		2				Multitrait GBLUP	Olson et al. (2012)
JE(US)	BS+JE+HF (US)	R^2		0				Multitrait GBLUP	Olson et al. (2012)
HF(US)	BS+JE+HF (US)	R^2		1				Multitrait GBLUP	Olson et al. (2012)
800 K									
HF (AS)	HF+JE (AS)	R	0	0	1			GBLUP	Erbe et al., (2012)
HF (AS)	HF+JE (AS)	R	−1	0	1			BayesR	Erbe et al. (2012)
JE (AS)	HF+JE (AS)	R	−1	1	1			GBLUP	Erbe et al. (2012)
JE (AS)	HF+JE (AS)	R	3	5	3			BayesR	Erbe et al. (2012)
TRANSCRIBED MARKERS									
HF (AS)	HF+JE (AS)	R	0	0	0			GBLUP	Erbe et al. (2012)
HF (AS)	HF+JE (AS)	R	−1	1	1			BayesR	Erbe et al. (2012)
JE (AS)	HF+JE (AS)	R	6	6	−3			GBLUP	Erbe et al. (2012)
JE (AS)	HF+JE (AS)	R	4	10	−2			BayesR	Erbe et al. (2012)

of a single- versus multi-trait GBLUP in a combined population of American HF, Jersey and BS. Although the observed accuracy for the Brown-Swiss animals was higher in the single trait model, looking across the three breeds, the multi-trait model performed the best. However, as this was tested only for protein yield, further studies are needed. Zhou et al. (2014a) found that a Bayesian model gave higher accuracies for NR animals than GBLUP when NR and VR animals were combined. Prediction accuracies were not increased for VR animals in general, but only for DR. In a study by Hayes et al. (2009) where Australian HF and Jerseys were combined, larger increases in accuracy for the Jersey animals were found when using a Bayesian method than when using a GBLUP method, and the largest gain was observed for fat yield, which might be explained by a better ability to estimate the effect of the DGAT1 mutation (Grisart et al., 2002) in the combined dataset. Results confirm that more advanced models, putting more emphasis on strong LD between markers and causative variants performs better for multi-breed prediction.

Another vein of models attempt to include information on breed proportions in admixed populations. In a study on the Danish Jersey which is an admixed population of Danish and American Jersey, Thomasen et al. (2013) investigated the implications of including inferred breed proportions from pedigree or genomic data in a random regression model. Although structures were observed in the population, no increase in the accuracy of genomic predictions was found when including these in the model. In the admixed VR red cattle population Makgahlela et al. (2012) showed that accounting for interactions between breed of origin and marker effects in the prediction model could improve the reliability of genomic prediction by 2–3% for protein and milk yield.

4.1. Marker density

In dairy cattle it has been estimated that 300 k markers are necessary to achieve a LD phase persistence sufficiently high for multi-breed genomic prediction (De Roos et al., 2008). This level can be achieved by using the Illumina Bovine High Density (HD) chip (Illumina, Inc., San Diego, CA) with 777 k SNP markers. Generally a subset of the reference population is genotyped with the Bovine HD chip and subsequently individuals genotyped with a lower density array are imputed to the HD level. This procedure is relatively reliable and can provide imputed HD genotypes with less than 1% error rate (Ma et al., 2013; Brøndum et al., 2012a; Berry et al., 2013; Hozé et al., 2013; Vanraden et al., 2013) with an appropriate reference population. Using a Bayesian prediction model Hozé et al. (2014) reported that when combining a small population of Normande cattle ($N < 500$) with Holstein and Montbéliarde the increase in accuracy from the multi-breed reference population was 2% higher when using the HD panel compared to using the 54 k panel. Similar results were observed in Erbe et al. (2012) where 3% higher accuracies were found for the Bayesian multi-breed prediction using an HD panel compared to using the 54 k panel.

Currently, cost effective whole genome sequencing (WGS) data are becoming available. From now on marker density will no longer be a limiting factor as all SNP (including causative variants) of sequenced individuals in principle can be imputed in all genotyped individuals. The 1000 bulls genome project (Daetwyler et al., 2014) provides a panel of sequenced bulls that can be used as a reference to impute WGS variants in all genotyped individuals. Initial results on imputation from HD to sequence data show a mean accuracy (correlation of true and imputed genotype) of 0.8 within the Holstein breed which has the largest number of sequenced animals. This is lower than previous results on imputation from 50 k to HD data where accuracies are typically larger than 0.95 (Brøndum et al., 2012a; Berry et al., 2013). It is however clear that for rare variants (which dominate the WGS data) the imputation accuracy is very low, i.e. for loci with MAF less than 0.05 the imputation accuracy is below 0.5 (Daetwyler et al., 2014). For numerically smaller breeds sequence imputation accuracies are lower than for the Holsteins, results however suggest that they can be increased by including other breeds in the reference (Daetwyler et al., 2014), and for both small and large breeds imputation accuracies are expected to increase as the number of animals in the 1000 Bull genomes database increases.

4.2. Prior information

Further prospects to improve across breed predictions could be to use biological information, such as information on genes and pathways. Although experimental data has shown that QTL replicate poorly in different genetic backgrounds, expression patterns and gene-networks associated to traits remain much more consistent (Huang et al., 2012). In human genetics it is observed that although different (sub)populations may show different mutations to affect a trait, these mutations may cluster in the same genes, and that, therefore, the gene-trait association remains relatively consistent. In the same line of thought, variances explained by genes or genomic regions may be relatively consistent between breeds or populations. Brøndum et al. (2012b) developed an approach somewhat along these lines by assuming that the number of important markers in genomic regions could be consistent across breeds. Here posterior proportions of mixture distributions estimated in chromosomal regions in one population were used as priors for the breed of interest in a mixture model with four Normal distributions. The study showed an increase in accuracy of up to 3.5% for the Jersey population when using priors derived from Australian HF, compared to a model without location specific priors. However, for most traits, the increases in accuracy were lower than those for prediction using combined reference populations. The small gain in accuracy of genomic predictions suggests that the priors from other breeds were too vague to efficiently utilize information across breeds.

As realized or imputed WGS data are becoming available on a large scale, it is becoming increasingly important to use any prior information available on the probability of particular SNP to be functional elements. In Erbe et al. (2012) a subset of markers on the Bovine HD chip in or within a distance of 1 kB to transcribed genomic regions was used to calculate GEBV. Genomic models were trained

in a reference population of Jersey bulls, HF bulls, or a combination. Results showed large increases for the prediction accuracy in the Jersey in the multi-breed scenario using both GBLUP and BayesR. This suggests that when using transcribed markers the relationship modeled with the G matrix is closer to the true functional relationship than when using 54 k or HD markers.

5. Perspectives

From results presented it is clear that combining reference populations of the same or closely related breeds is an efficient way to increase accuracies of GEBV. For distantly related breeds it is crucial that markers are in close LD with causative variants and that prediction methods attempt to focus the predictions on those markers. Although few results exist so far, indications are that Bayesian variable selection models which allocate a large part of the genomic variance to markers in strong LD with causative variants are beneficial when predicting GEBV across breeds. Methods based on haplotype sharing between breeds can potentially improve predictions further, as haplotypes of linked markers may exist that are in higher LD with the causative variants than any individual SNP. Mapping results using haplotype sharing across breeds has sometimes obtained very precise localization of causative mutations (Grisart et al., 2002).

These models need to be developed in a multi-trait framework. Although Calus and Veerkamp (2011) have shown some work on multi-trait Bayesian genomic models, the covariance for SNPs across the genome in their models is the same and based on a prior pedigree based multi-trait BLUP analysis of the genetic (genomic) covariance of the traits. Models need to be developed that alleviate the assumption of equal covariance across the traits for each SNP.

With the availability of WGS data the requirement of using markers in strong LD with causative variants can be achieved by sequencing a subset of individuals and imputing all SNP to all genotyped individuals. This generates more than 20 million SNPs to handle in genomic predictions, which is a technical challenge for Bayesian variable selection models in large populations. Another limitation in this approach is, that the accuracy of imputation for rare variants is very poor (Daetwyler et al., 2014). If the complex traits in the breeding goal are regulated by rare variants, the imputed sequence will provide limited extra information. To overcome this limitation a substantially larger number of individuals could be sequenced. If focus on rare variants is needed, sequencing many more animals may be required. This could call for alternative cost effective sequencing strategies to complement the current medium coverage sequences available. Alternative strategies could be sequencing a large numbers of individuals with low coverage (Li et al., 2011), exome sequencing, or genotyping by sequencing approaches. Another alternative is to retrieve the potentially most efficient markers for across breed predictions from mining the sequence data and add these SNP to genotyping chips that are used to screen a large number of individuals. These SNP could be selected as those SNP from sequence based genome-wide association studies (GWAS) that show the highest association to the most important traits or that explain genetic covariance across breeds. Alternatively, SNP could be selected that are most likely to be functional when assessing the annotation information. This approach has the advantage that sequence variants which are inaccurately imputed in all genotyped individuals, could be genotyped accurately in phenotyped individuals. Consequently the associations may reach their full potential to improve genomic predictions.

Using information across breeds will only be efficient for the fraction of the genetic variance caused by causative variants that segregate in both breeds. It is unclear how large the shared variance across breeds is. Generally, QTL studies are not consistent in finding the same regions across different breeds. This is likely due to a lack of power to identify QTL in smaller breeds but also indicates that part of the genetic variance is likely to be private to specific breeds. With continuously more detailed genetic information from genotyping and sequencing, it would be useful to perform a powerful analysis to assess to which extend genetic variance is private within breeds or shared among breeds. Another reason for the lack of consistency across breeds could be epistatic interactions among genes. In this case the effect of a particular QTN depends on the frequency of genes it interacts with (e.g. Carlborg et al., 2003, Huang et al., 2012)). As these could be different among breeds it results in breed specific effects. It would be useful to study if the lack of shared genetic variance among breeds is caused by such epistatic interactions. This could probably most efficiently be carried out in large datasets of genotyped and phenotyped individuals, where markers in LD with potentially interacting QTL can be observed for the same individuals as the phenotypes. Such large datasets are presently becoming available in dairy cattle. Accommodating for gene-background interactions can remain a bottleneck in across breed prediction and may not be trivial. Including marker by breed interactions in multi-breed models will essentially disconnect the marker effects between breeds, and therefore fall back to a within-breed analysis. Thus, relatively sophisticated models should be developed, for instance, to separate markers in those that are consistent between breeds and those that are not consistent between breeds or to allow marker effects to be partitioned into shared and breed specific components.

An expected added value of predicting GEBVs across breeds is an increased persistency of the predictive ability of GEBV within breeds, because prediction models focus more on SNP close to causative variants. This becomes increasingly important as more generations of genotyped individuals are available in the reference populations. In addition, as genomic predictions are becoming the primary information when ranking selection candidates, younger animals without parents or even grandparents in the reference population are selected for breeding. With this increased genetic distance to the reference population it becomes increasingly important to increase the persistency of genomic predictions. This is a factor that is largely ignored when studies evaluate the accuracies of genomic predictions, which generally evaluate the predictive ability one generation away from the reference population.

Conflict of interest statement

None declared.

References

Berry, D.P., McClure, M.C., Mullen, M.P., 2013. Within- and across-breed imputation of high-density genotypes in dairy and beef cattle from medium- and low-density genotypes. J. Anim. Breed. Genet. 131, 165–172, http://dx.doi.org/10.1111/jbg.12067.

Bolormaa, S., Pryce, J., Kemper, K., 2013. Accuracy of prediction of genomic breeding values for residual feed intake, carcass and meat quality traits in *Bos taurus*, *Bos indicus* and composite beef cattle. J. Anim. Sci. 91, 3088–3104, http://dx.doi.org/10.2527/jas2012-5827.

Brøndum, R.F., Ma, P., Lund, M.S., Su, G., 2012a. Short communication: Genotype imputation within and across Nordic cattle breeds. J. Dairy Sci. 95, 6795–6800, http://dx.doi.org/10.3168/jds.2012-5585.

Brøndum, R.F., Rius-Vilarrasa, E., Strandén, I., Su, G., Guldbrandtsen, B., Fikse, W.F., Lund, M.S., 2011. Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. J. Dairy Sci. 94, 4700–4707, http://dx.doi.org/10.3168/jds.2010-3765.

Brøndum, R.F., Su, G., Lund, M.S., Bowman, P.J., Goddard, M.E., Hayes, B.J., 2012b. Genome position specific priors for genomic prediction. BMC Genomics 13, 543, http://dx.doi.org/10.1186/1471-2164-13-543.

Bulmer, M., 1971. The effect of selection on genetic variability. Am. Nat. 105, 201–211.

Calus, M.P.L., Veerkamp, R.F., 2011. Accuracy of multi-trait genomic selection using different methods. Genet. Select. Evol. 43, 26, http://dx.doi.org/10.1186/1297-9686-43-26.

Carlborg, O., Kerje, S., Schütz, K., Jacobsson, L., Jensen, P., Andersson, L., 2003. A global search reveals epistatic interaction between QTL for early growth in the chicken. Genome Res. 13, 413–421, http://dx.doi.org/10.1101/gr.528003.

Chen, L., Schenkel, F., Vinsky, M., Crews Jr, D.H., Li, C., 2013. Accuracy of predicting genomic breeding values for residual feed intake in Angus and Charolais beef cattle. J. Anim. Sci. 91, 4669–4678, http://dx.doi.org/10.2527/jas2013-5715.

Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R.F., Liao, X., Grohs, C., Jung, S., Baur, A., Esquerré, D., Gollnick, N., Legrain, S., Laurent, P., Djari, A., Rossignol, M.-N., Klopp, C., Rocha, D., Fritz, S., Eggen, A., Bowman, P., Coote, D., Chamberlain, A., VanTassel, C.P., Hulsegge, I., Goddard, M., Guldbrandtsen, B., Lund, M.S., Veerkamp, R., Boichard, D., Fries, R., Hayes, B.J., 2014. The 1000 bull genomes project. Nat. Genet.. (in press).

Daetwyler, H.D., Kemper, K.E., van der Werf, J.H.J., Hayes, B.J., 2012. Components of the accuracy of genomic prediction in a multi-breed sheep population. J. Anim. Sci. 90, 3375–3384, http://dx.doi.org/10.2527/jas.2011-4557.

Erbe, M., Hayes, B.J., Matukumalli, L.K., Goswami, S., Bowman, P.J., Reich, C.M., Mason, B. a, Goddard, M.E., 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95, 4114–4129, http://dx.doi.org/10.3168/jds.2011-5019.

Felsenstein, J., 1965. The effect of linkage on directional selection. Genetics 52, 349–363.

Gao, H., Su, G., Janss, L., Zhang, Y., Lund, M.S., 2013. Model comparison on genomic predictions using high-density markers for different groups of bulls in the Nordic Holstein population. J. Dairy Sci. 96, 4678–4687, http://dx.doi.org/10.3168/jds.2012-6406.

Goddard, M., 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136, 245–257, http://dx.doi.org/10.1007/s10709-008-9308-0.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M., Snell, R., 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12, 222–231, http://dx.doi.org/10.1101/gr.224202.

Habier, D., Fernando, R.L., Garrick, D.J., 2013. Genomic-BLUP decoded: a look into the black box of genomic prediction. Genetics 194, 597–607, http://dx.doi.org/10.1534/genetics.113.152207.

Habier, D., Fernando, R.L., Kizilkaya, K., Garrick, D.J., 2011. Extension of the bayesian alphabet for genomic selection. BMC Bioinf. 12, 186, http://dx.doi.org/10.1186/1471-2105-12-186.

Habier, D., Tetens, J., Seefried, F., Lichtner, P., Thaller, G., 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet. Select. Evol. 42, 5, http://dx.doi.org/10.1186/1297-9686-42-5.

Hayes, B.J., Bowman, P.J., Chamberlain, A.C., Verbyla, K., Goddard, M.E., 2009. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet. Select. Evol. 41, 51, http://dx.doi.org/10.1186/1297-9686-41-51.

Hayes, B.J., Visscher, P.M., McPartlan, H.C., Goddard, M.E., 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Res. 13, 635–643, http://dx.doi.org/10.1101/gr.387103.

Heringstad, B., Su, G., Solberg, T.R., Guldbrandtsen, B., Svendsen, M., Lund, M.S., 2011. Genomic predictions based on a joint reference population for Scandinavian red breeds. In Book of Abstracts of the 62nd Annual Meeting of the European Federation of Animal Science. Waageningen Academic Publishers, Stavanger, Norway29.

Hill, W.G., Robertson, A., 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38, 226–231, http://dx.doi.org/10.1007/BF01245622.

Hozé, C., Fouilloux, M.-N., Venot, E., Guillaume, F., Dassonneville, R., Fritz, S., Ducrocq, V., Phocas, F., Boichard, D., Croiseau, P., 2013. High-density marker imputation accuracy in sixteen French cattle breeds. Genet. Select. Evol. 45, 33, http://dx.doi.org/10.1186/1297-9686-45-33.

Hozé, C., Fritz, S., Phocas, F., Boichard, D., Ducrocq, V., Croiseau, P., 2014. Efficiency of multi-breed genomic selection for dairy cattle breeds with different sizes of reference population. J. Dairy Sci, 1–12, http://dx.doi.org/10.3168/jds.2013-7761.

Huang, W., Richards, S., Carbone, M.A., Zhu, D., Anholt, R.R.H., Ayroles, J.F., Duncan, L., Jordan, K.W., Lawrence, F., Magwire, M.M., Warner, C.B., Blankenburg, K., Han, Y., Javaid, M., Jayaseelan, J., Jhangiani, S.N., Muzny, D., Ongeri, F., Perales, L., Wu, Y.-Q., Zhang, Y., Zou, X., a Stone, E., a Gibbs, R., Mackay, T.F.C., 2012. Epistasis dominates the genetic architecture of Drosophila quantitative traits. Proc. Natl. Acad. Sci. U. S. A. 109, 15553–15559, http://dx.doi.org/10.1073/pnas.1213423109.

Kachman, S.D., Spangler, M.L., Bennett, G.L., Hanford, K.J., Kuehn, L. a, Snelling, W.M., Thallman, R.M., Saatchi, M., Garrick, D.J., Schnabel, R.D., Taylor, J.F., Pollak, E.J., 2013. Comparison of molecular breeding values based on within- and across-breed training in beef cattle. Genet. Select. Evol. 45, 30, http://dx.doi.org/10.1186/1297-9686-45-30.

Kadri, N.K., 2014. Methods For Association Mapping Using Multiple Population Data, Ph.D. thesis submitted to Aarhus University ISBN: 978-87-92869-90-6.

Karoui, S., Carabaño, M.J., Díaz, C., Legarra, A., 2012. Joint genomic evaluation of French dairy cattle breeds using multiple-trait models. Genet. Select. Evol. 44, 39, http://dx.doi.org/10.1186/1297-9686-44-39.

Li, Y., Sidore, C., Kang, H.M., Boehnke, M., Abecasis, G.R., 2011. Low-coverage sequencing: implications for design of complex trait association studies. Genome Res. 21, 940–951, http://dx.doi.org/10.1101/gr.117259.110.

De Los Campos, G., Vazquez, A.I., Fernando, R., Klimentidis, Y.C., Sorensen, D., 2013. Prediction of complex human traits using the genomic best linear unbiased predictor. PLoS Genet. 9, e1003608, http://dx.doi.org/10.1371/journal.pgen.1003608.

Lund, M.S., De Roos, A.P.W., De Vries, A.G., Druet, T., Ducrocq, V., Guillaume, F., Guldbrandtsen, B., Liu, Z., Reents, R., Schrooten, C., Seefried, M., Su, G., 2010. Improving genomic prediction by EuroGenomics collaboration. In: Proceedings of the WCGALP 2010. Leipzig. pp. 7–10.

Lund, M.S., deRoos, A.P.W., deVries, A.G., Druet, T., Ducrocq, V., Fritz, S., Guillaume, F., Guldbrandtsen, B., Liu, Z.T., Reents, R., Schrooten, C., Seefried, F., Su, G.S., 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet. Select. Evol. 43.

Ma, P., Brøndum, R.F., Zhang, Q., Lund, M.S., Su, G., 2013. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. J. Dairy Sci. 96, 4666–4677, http://dx.doi.org/10.3168/jds.2012-6316.

Makgahlela, M.L., Mäntysaari, E.A., Strandén, I., Koivula, M., Nielsen, U.S., Sillanpää, M.J., Juga, J., 2012. Across breed multi-trait random regression genomic predictions in the Nordic Red dairy cattle. J. Anim. Breed. Genet. 130, 10–19, http://dx.doi.org/10.1111/j.1439-0388.2012.01017.x.

Olson, K.M., Vanraden, P.M., Tooker, M.E., 2012. Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 95, 5378–5383, http://dx.doi.org/10.3168/jds.2011-5006.

Pryce, J.E., Gredler, B., Bolormaa, S., Bowman, P.J., Egger-Danner, C., Fuerst, C., Emmerling, R., Sölkner, J., Goddard, M.E., Hayes, B.J., 2011. Short communication: genomic selection using a multi-breed, across-country

reference population. J. Dairy Sci. 94, 2625–2630, http://dx.doi.org/10.3168/jds.2010-3719.
De Roos, A.P.W., Hayes, B.J., Spelman, R.J., Goddard, M.E., 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics 179, 1503–1512, http://dx.doi.org/10.1534/genetics.107.084301.
Schenkel, F., Sargolzaei, M., Kistemaker, G., Jansen, G., Sullivan, P., Van Doormaal, B.J., Vanraden, P.M., Wiggans, G.R., 2009. Reliability of genomic evaluation of Holstein cattle in Canada. Interbull. Bull. 39, 51–58.
Sved, J., 1971. Linkage disequilibrium of chromosome segments. Theor. Popul. Biol. 2, 125–141.
Sørensen, L.P., Janss, L., Madsen, P., Mark, T., Lund, M.S., 2012. Estimation of (co)variances for genomic regions of flexible sizes: application to complex infectious udder diseases in dairy cattle. Genet. Select. Evol. 44, 18, http://dx.doi.org/10.1186/1297-9686-44-18.
Thomasen, J., Sørensen, A., Su, G., Madsen, P., Lund, M., Guldbrandtsen, B., 2013. The admixed population structure in Danish Jersey dairy cattle challenges accurate genomic predictions. J. Anim. Sci. 91, 3105–3112, http://dx.doi.org/10.2527/jas2012-5490.
Thomasen, J.R., Guldbrandtsen, B., Su, G., Brøndum, R.F., Lund, M.S., 2012. Reliabilities of genomic estimated breeding values in Danish Jersey. Animal 6, 789–796, http://dx.doi.org/10.1017/S1751731111002035.
Thomasen, J.R., Willam, A., Guldbrandtsen, B., Lund, M.S., Sørensen, A.C., 2014. Genomic selection strategies in a small dairy cattle population evaluated for genetic gain and profit. J. Dairy Sci. 97, 458–470, http://dx.doi.org/10.3168/jds.2013-6599.
VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91, 4414–4423, http://dx.doi.org/10.3168/jds.2007-0980.
Vanraden, P.M., Null, D.J., Sargolzaei, M., Wiggans, G.R., Tooker, M.E., Cole, J.B., Sonstegard, T.S., Connor, E.E., Winters, M., van Kaam, J.B.C.H.M., Valentini, G. a, Van Doormaal, B.J., Faust, M. a, Doak, G. a, 2013. Genomic imputation and evaluation using high-density Holstein genotypes.
J. Dairy Sci. 96, 668–678, http://dx.doi.org/10.3168/jds.2012-5702.
Vanraden, P.M., Olson, K.M., Null, D.J., Sargolzaei, M., Winters, M., van Kaam, J.B.C.H.M., 2012. Reliability increases from combining 50, 000- and 777, 000- marker genotypes from four countries. Interbull. Bull. 46, 75–79.
Weber, K., Thallman, R., Keele, J., Snelling, W., Bennett, G., Smith, T., McDaneld, T., Allan, M., Van Eenennaam, A., Kuehn, L. a, 2012. Accuracy of genomic breeding values in multibreed beef cattle populations derived from deregressed breeding values and phenotypes. J. Anim. Sci. 90, 4177–4190, http://dx.doi.org/10.2527/jas2011-4586.
Wientjes, Y.C.J., Veerkamp, R.F., Calus, M.P.L., 2013. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. Genetics 193, 621–631, http://dx.doi.org/10.1534/genetics.112.146290.
Zhou, L., Ding, X., Zhang, Q., Wang, Y., Lund, M.S., Su, G., 2013. Consistency of linkage disequilibrium between Chinese and Nordic Holsteins and genomic prediction for Chinese Holsteins using a joint reference population. Genet. Select. Evol. 45, 7, http://dx.doi.org/10.1186/1297-9686-45-7.
Zhou, L., Heringstad, B., Su, G., Guldbrandtsen, B., Meuwissen, T., Svendsen, M., Grove, H., Nielsen, U.S., Lund, M.S., 2014a. Genomic predictions based on a joint reference population for the Nordic Red cattle breeds. J. Dairy Sci. 10.3168/jds.2013-7580, in press.
Zhou, L., Lund, M.S., Wang, Y., Su, G., 2014b. Genomic predictions across Nordic Holstein and Nordic Red using the GBLUP model with different relationship matrices. J. Anim. Breed. Genet.