Relatedness between numerically small Dutch Red dairy cattle populations and possibilities for multibreed genomic prediction

J. Marjanovic,[1] **B. Hulsegge,**[1,2] **and M. P. L. Calus**[1]*
[1]Animal Breeding and Genomics, Wageningen University & Research, Droevendaalsesteeg 1, 6700AH Wageningen, the Netherlands
[2]Centre for Genetic Resources, Wageningen University & Research, Droevendaalsesteeg 1, 6700AH Wageningen, the Netherlands

ABSTRACT

Red dairy breeds are a valuable cultural and historical asset, and often a source of unique genetic diversity. However, they have difficulties competing with other, more productive, dairy breeds. Improving competitiveness of Red dairy breeds, by accelerating their genetic improvement using genomic selection, may be a promising strategy to secure their long-term future. For many Red dairy breeds, establishing a sufficiently large breed-specific reference population for genomic prediction is often not possible, but may be overcome by adding individuals from another breed. Relatedness between breeds strongly decides the benefit of adding another breed to the reference population. To prioritize among available breeds, the effective number of chromosome segments (M_e) can be used as an indicator of relatedness between individuals from different breeds. The M_e is also an important parameter in determining the accuracy of genomic prediction. The M_e can be estimated both within a population and between 2 populations or breeds, as the reciprocal of the variance of genomic relationships. We investigated relatedness between 6 Dutch Red cattle breeds, Groningen White Headed (GWH), Dutch Friesian (DF), Meuse-Rhine-Yssel (MRY), Dutch Belted (DB), Deep Red (DR), and Improved Red (IR), focusing primarily on the M_e, to predict which of those breeds may benefit from including reference animals of the other breeds. All of these breeds, except MRY, are under high risk of extinction. Our results indicated high variability of M_e, especially between M_e ranging from ~3,500 to ~17,400, indicating different levels of relatedness between the breeds. Two clusters are especially important, one formed by MRY, DR, and IR, and the other comprising DF and DB. Although relatedness between breeds within each of these 2 clusters is high, across-breed genomic prediction is still limited by the current number of genotyped individuals, which for many breeds is low. However, adding MRY individuals would increase the reference population of DR substantially. We estimated that between 11 and 133 individuals from other breeds are needed to achieve accuracy of genomic prediction equivalent to using one additional individual from the same breed. Given the variation in size of the breeds in this study, the benefit of a multibreed reference population is expected to be lower for larger breeds than for the smaller ones.

Key words: Red dairy cattle breed, effective number of chromosome segments, genomic selection

INTRODUCTION

Genomic selection of candidates has become the method of choice in many breeding programs. In dairy cattle, the main appeal of selection based on genomic breeding values (**GEBV**) instead of traditional BLUP, is often in the possibility to increase genetic gain by reducing the generation interval. The process of estimating GEBV is known as genomic prediction, and it requires a reference population where SNP effects can be estimated accurately. Hence, success of genomic selection strongly relies on the accuracy of genomic prediction. Two main factors that determine the accuracy of genomic prediction are the size of the reference population (Meuwissen et al., 2001) and the relationship between reference individuals and selection candidates (Habier et al., 2007; Pszczola et al., 2012). The latter is especially important when the reference population is small (Wientjes et al., 2013). Limited-size reference populations are usually encountered in numerically small breeds. Many of the European Red dairy breeds fall into that category.

Red dairy breeds are a valuable cultural and historical asset, and often a source of unique genetic diversity (van Breukelen et al., 2019). However, they have difficulties to compete with other, highly productive, dairy breeds (Hiemstra et al., 2010; Addo et al., 2017). In the Netherlands, up to the beginning of the 1980s, the native dual-purpose breeds [i.e., breeds used for both milk

and meat production, such as Groningen White Headed (**GWH**), Dutch Friesian (**DF**), and Meuse-Rhine-Yssel (**MRY**)] were most popular for milk production (Van Eijndhoven, 2014). Currently, the high-yielding Holstein Friesian (**HF**) is the dominant breed in the Dutch dairy cattle population, and as a result of replacement of Red dairy cattle by HF, Red dairy breeds suffered a significant decrease in numbers (Hiemstra et al., 2010; Van Eijndhoven, 2014). Due to their functional characteristics that tend to be better compared with mainstream breeds, for example, fertility, udder health, calving ease, and because of their adaptation to local environments, interest in Red dairy breeds is still substantial, especially for niche production and crossbreeding (Schäler et al., 2019).

Maintaining, and preferably improving, competitiveness of Red dairy breeds by accelerating genetic improvement of production traits using genomic selection may be a promising strategy to secure their long-term future. For example, genomic evaluations for the MRY breed showed 15.9% gain in reliability over parent average EBV for production traits, and 9.6% for conformation traits, indicating clear benefits of using GEBV (Stoop et al., 2017). However, for many Red dairy breeds, establishing a sufficiently large breed-specific reference population is often not possible. This issue may be overcome by relying on reference animals from other breeds (i.e., by forming a multibreed reference population; Hozé et al., 2014). The reference animals from the target breed are much more informative than those from other breeds; hence, many animals from other breeds are needed to achieve the same accuracy as using reference animals from the target breed (Meuwissen et al., 2013). In addition, not all breeds are equally relevant. Relatedness between breeds strongly decides the benefit of adding another breed to the reference population, where a higher increase in accuracy is expected when closely related breeds are combined in the reference population, whereas no or only low increases in accuracy are expected when those breeds are more distant (Habier et al., 2007; Brøndum et al., 2011; Hozé et al., 2014). Different methods can be used to estimate relatedness between breeds (Eding and Meuwissen, 2001). In this domain the effective number of chromosome segments (M_e) is especially interesting, as it not only indicates relatedness between individuals from different breeds, but it can also be used to predict expected accuracy of genomic prediction that relies on use of information from other breeds (Wientjes et al., 2016).

The M_e can be estimated both within a population and between 2 populations or breeds. The M_e within the population describes the number of chromosome segments that are segregating independently in the population. The value of within M_e is directly related to the effective population size (N_e; Brard and Ricard, 2015; Lee et al., 2017). Low N_e is associated with higher relatedness among individuals and consequently to higher extent of linkage disequilibrium (**LD**; Sved, 1971; Falconer and Mackay, 1996) and lower number of segregating chromosome segments. The M_e between populations gives insight into the consistency of LD between the 2 populations (Wientjes et al., 2016). Conceptually, we can consider the M_e between populations as the effective number of chromosome segments that are segregating in the joint reference population consisting of the 2 populations of interest (Wientjes et al., 2016). Low M_e between populations indicates high relatedness between 2 populations, whereas between populations that were split more generations ago usually a higher value of M_e is observed (see general discussion in Wientjes, 2016).

In this study we investigated relatedness between 6 native Dutch Red cattle breeds, focusing primarily on the effective number of chromosome segments, to predict which of those breeds may benefit from including reference animals of the other breeds. To achieve our aim, we used genotype data on 5 Dutch Red dairy breeds, GWH, DF, MRY, Dutch Belted (**DB**), Deep Red (**DR**), and one Dutch Red beef breed, Improved Red (**IR**). Out of these breeds only MRY has an active breeding program, with a recently implemented genomic evaluation system (Stoop et al., 2017). In addition, GWH, DF, DB, DR, and IR are under high risk of extinction (CGN, 2019).

MATERIALS AND METHODS

Data

Genotype data on 6 breeds were generated by the Centre for Genetic Resources, the Netherlands, and in addition genotypes of 59 MRY bulls, belonging to CRV, a cattle breeding organization in the Netherlands, were added to the data set (Table 1). The majority of individuals originated from the Netherlands; however, a small number also came from Germany, France, Belgium, and the United States. Data included both bulls and cows in a 3:1 ratio. Genotyping was performed using BovineSNP50 BeadChip version 2 and 3 (Illumina Inc., San Diego, CA). After combining the different genotyping batches, a set of 41,764 markers remained. This initial genotype data set was put through quality control where only individuals with a call rate >90%, markers with call rate >95%, and markers with an allele present at least 3 times in the data set (equivalent

Table 1. Numbers and source of genotyped animals per breed, and current status of each of the breeds

Breed	Number of individuals	Number of individuals per sex			Country of origin[1]					Total number of cows per breed[2]	Effective population size[3]	Level[3,4] of inbreeding as average $FR_{OH} > 2$ Mb
		Bulls	Cows	Unknown	NL	BE	DE	FR	US			
Meuse-Rhine-Yssel	678	546	132	0	674	0	4	0	0	11,006	90	0.080
Groningen White Headed	128	68	60	0	126	0	2	0	0	1,991	62	0.150
Deep Red	44	23	21	0	0	1	43	0	0	1,216	86	0.065
Dutch Belted	39	39	0	0	38	0	0	0	1	2,500	47	0.110
Dutch Friesian	333	232	100	1	322	0	0	11	0	2,402	99	0.085
Improved Red	24	24	0	0	24	1	0	0	0	726	86	0.040
Total	1,246	939	314	2	1,184	1	49	11	1			

[1] NL = the Netherlands; BE = Belgium; DE = Germany; FR = France; US = United States.

[2] Data from Centre for Genetic Resources (CGN; 2019).

[3] Unpublished data (C. Schmidtmann, Christian-Albrechts-Universitat, Kiel, Germany).

[4] $FR_{OH} > 2$ Mb = the breeding coefficient computed from runs of homozygosity of at least 2 Mbp length.

to using a minor allele frequency threshold of 0.001) were retained. The final genotype data included 39,336 markers for 1,246 individuals.

Principal component analysis (**PCA**) was used to visualize the relatedness of the different breeds based on the genotype data. Principal components were computed using calc_grm software (Calus and Vandenplas, 2016) as a singular value decomposition of a genomic relationship matrix computed following the first method proposed by VanRaden (2008), using allele frequencies computed across the entire data set, where principal components are computed using eigenvectors and eigenvalues as proposed by Abraham and Inouye (2014). In addition, genetic differentiation among the breeds was calculated using pairwise F_{ST} (Weir and Cockerham, 1984) in PLINK v1.9 (Purcell and Chang; www.cog-genomics.org/plink/1.9/).

Estimation of Effective Number of Chromosome Segments (M_e)

The within population M_e was estimated as variation in genomic relationships between the individuals (Hayes et al., 2009; Goddard, 2009; Goddard et al., 2011). The following equation was used:

$$M_e = \frac{1}{Var\left(G_{ij} - A_{ij}\right)}, \qquad [1]$$

where G_{ij} is the genomic and A_{ij} is the pedigree relationship between individual i and j, and the variance is taken over all pairs ij in the population. However, for all breeds except MRY, pedigrees were not available; hence, for those breeds within M_e was estimated using only variance of genomic relationships (see Discussion). So, effectively, the equation used reduced to $M_e = \frac{1}{Var\left(G_{ij}\right)}$. Using an adapted version of equation 1, M_e between populations can be estimated as follows (Wientjes et al., 2013):

$$M_e = \frac{1}{Var\left(G_{pop1_i pop2_j} - A_{pop1_i pop2_j}\right)}, \qquad [2]$$

where $G_{pop1_i pop2_j}$ is the genomic relationship between individual i from population 1 and individual j from population 2, and $A_{pop1_i pop2_j}$ is the corresponding pedigree relationship, with the variance taken across all pairs of individuals from population 1 and 2. Note that pedigree relationships across breeds to be used in this

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix} =$$

$$\begin{bmatrix} \dfrac{\mathbf{Z}_1 \mathbf{Z}_1^{'}}{\sum 2p_{1k}\left(1-p_{1k}\right)} & \dfrac{\mathbf{Z}_1 \mathbf{Z}_2^{'}}{\sqrt{\sum 2p_{1k}\left(1-p_{1k}\right)}\sqrt{\sum 2p_{2k}\left(1-p_{2k}\right)}} \\ \dfrac{\mathbf{Z}_2 \mathbf{Z}_1^{'}}{\sqrt{\sum 2p_{1k}\left(1-p_{1k}\right)}\sqrt{\sum 2p_{2k}\left(1-p_{2k}\right)}} & \dfrac{\mathbf{Z}_2 \mathbf{Z}_2^{'}}{\sum 2p_{2k}\left(1-p_{2k}\right)} \end{bmatrix},$$

formula can be constructed relative to an across-breed base (VanRaden, 1992; Legarra et al., 2015). Whether or not such an across-breed base or any other arbitrary base is used in $A_{pop1_i\,pop2_j}$ likely will hardly affect the results, as changing this base effectively results in adding a constant value to all relationships. However, due to an absence of pedigrees, for estimation of between M_e only genomic relationships were used, that is

$$M_e = \frac{1}{Var\left(G_{pop1_i\,pop2_j}\right)} \quad \text{(see Discussion).}$$

The M_e was estimated using calc_grm software (Calus and Vandenplas, 2016) and an exponential function was fitted to adjust **G-A** values to be on average 0 across the range of pedigree relationships (Wientjes et al., 2016). The matrix **G** was calculated using the equation shown above.

In the equation, $\mathbf{G}_{11}$ is a matrix with genomic relationships in population 1, $\mathbf{G}_{22}$ is a matrix with genomic relationships in population 2, whereas $\mathbf{G}_{12}$ and $\mathbf{G}_{21}$ are matrices with genomic relationships between population 1 and 2 (Wientjes et al., 2016). $\mathbf{Z}_1(\mathbf{Z}_2)$ matrix contains genotypes for all individuals from population 1 (population 2) at all loci, centered by subtracting twice the allele frequency per locus, and $p_{1k}(p_{2k})$ is the allele frequency of marker k in population 1 (population 2). $\mathbf{Z}_1\mathbf{Z}_2^{'}$ and $\mathbf{Z}_2\mathbf{Z}_1^{'}$ are matrices of genetic covariance between the genetic values of 2 populations, divided by the SDs of the genotypes in each population $\sqrt{\sum 2p_{1k}\left(1-p_{1k}\right)}$ and $\sqrt{\sum 2p_{2k}\left(1-p_{2k}\right)}$. These genomic relationship matrices were built for all pairwise combinations of breeds.

To predict how valuable individuals from another breed are compared with individuals from the selection breed, we used the equation of Wientjes (2016), $n_{P,2} = \dfrac{h_1^2}{h_2^2}\dfrac{M_{e_{1,2}}}{M_{e_1}}n_{P,1}$. In this equation, M_{e_1} refers to within M_e of the selection breed, and $M_{e_{1,2}}$ refers to between population M_e between selection breed and any other (reference) breed; $n_{P,2}$ is the number of indi-

viduals from another population/breed that can obtain the same accuracy as $n_{P,1}$. The genetic correlations between breeds were assumed to be 1, and in our calculations we consider heritabilities from both breeds $\left(h_1^2 \text{ and } h_2^2\right)$ to be equal, such that these drop out of the equation.

RESULTS

We performed PCA as a first step to assess the relationships between the breeds. The first 10 principal components explained 16.3% of the total variation. The first principal component explained 7.29% of total variance (Figure 1) and indicated separation of DB and DF from MRY and DR. In addition, we could observe high overlap between MRY, DR, and IR. The second principal component explained 3.85% of the total variance and showed clear separation of GWH from other breeds. The next principal components did not provide further separation of the breeds (results not shown). Genetic differentiation among breeds, measured using pairwise F_{ST} (Table 2), confirmed closest relationships between DR and IR (0.028), MRY and DR (0.035),
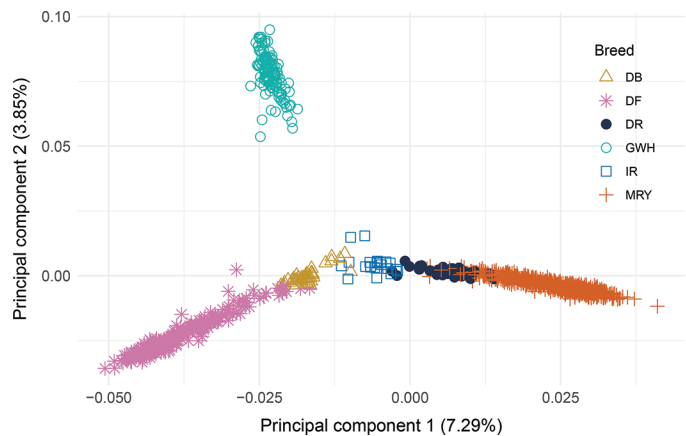
Figure 1. Principal component analysis of 6 Dutch cattle breeds: Groningen White Headed (GWH), Dutch Friesian (DF), Meuse-Rhine-Yssel (MRY), Dutch Belted (DB), Deep Red (DR), Improved Red (IR).

and MRY and IR (0.050). Pairwise F_{ST} ranged between 0.028 and 0.110, with high values especially observed between GWH and other breeds.

Table 3 presents estimates of within and between population M_e. Within all breeds M_e was low, ranging from 125 to 336 segments. The low values are not surprising, given the numerically small size of these breeds, and reported N_e values and inbreeding levels (Table 1). The highest within M_e value of 336 has been found for the largest of the breeds, MRY. However IR, even though being the smallest of all breeds, had within M_e of 208, in line with having the lowest inbreeding level and relatively high N_e compared with the other breeds (Table 1). Estimates of between M_e were significantly larger than within M_e, and ranged from ~3,500 to ~17,400. Based on these values, the closest relationships were between MRY, DR, and IR. These breeds have common (recent) history, as both DR and IR have been derived from MRY, by selecting individuals with a deeper red color and in IR selection was also oriented toward meat production rather than dual purpose. Dutch Friesian was most closely related to DB, and vice versa. Groningen White Headed was genetically most distant from other breeds. These relationships were also confirmed by PCA plot and pairwise F_{ST}.

Table 4 gives more tangible insight in consistency of LD between individuals from different breeds. It shows the predicted number of individuals from the reference population that are needed to achieve the same improvement in accuracy as when using one individual from the same breed as selection population. For DR, at least 24 and 45 individuals are needed from its closest related breeds MRY and IR, respectively, whereas many more are needed from a distantly related breed. For smallest breed of all, IR, number of individuals was between 32 and 75; for largest breed, MRY, between 11 and 52 are needed to achieve the same accuracy as using one MRY individual in the reference population.

DISCUSSION

We analyzed relatedness between 6 native Dutch Red cattle breeds using effective number of chromo-

Table 2. Estimated pairwise genome-wide fixation indices (F_{ST})[1]

Breed	DF	DR	GWH	IR	MRY
DB	0.063	0.060	0.110	0.050	0.081
DF		0.070	0.095	0.060	0.077
DR			0.100	0.028	0.035
GWH				0.100	0.100
IR					0.050

[1]Deep Red (DR), Dutch Belted (DB), Dutch Friesian (DF), Groningen White Headed (GWH), Meuse-Rhine-Yssel (MRY), Improved Red (IR).

Table 3. Estimates of the effective number of chromosome segments (M_e) within and between breeds[1,2]

Breed	DR	DB	DF	GWH	MRY	IR
DR	149					
DB	16,648	125				
DF	14,466	7,956	211			
GWH	17,088	15,972	14,128	153		
MRY	3,541	14,528	15,975	17,387	336	
IR	6,695	15,525	10,932	13,680	8,543	208

[1]Deep Red (DR), Dutch Belted (DB), Dutch Friesian (DF), Groningen White Headed (GWH), Meuse-Rhine-Yssel (MRY), Improved Red (IR).
[2]Within population M_e values are on the diagonal; between population M_e values are off the diagonal.

some segments. Our results show high variability of M_e, especially between M_e, indicating different level of relatedness between the breeds, which can be used to tailor multibreed reference populations for the considered breeds. The multibreed reference population should be much larger than the available single-breed reference population for a predicted breed, since, depending on the breed, between 11 and 133 individuals from other breeds are needed to achieve accuracy of genomic prediction equivalent to using one additional individual from the same breed. Therefore, given the variation in size of the breeds in the study, the benefit of a multibreed reference population is expected to be lower for larger breeds than for smaller ones.

Within Population M_e

For 6 breeds in this study estimated within M_e ranged between 125 and 336 segments (Table 3). These values are lower than those found for numerically much larger HF populations where within M_e was 1,000 to 2,000 (Erbe et al., 2013; Wientjes et al., 2016; van den Berg

Table 4. Number of individuals from the reference population needed to obtain the same increase in accuracy as one individual from the selection population[1,2]

Selection population	Reference population					
	DR	DB	DF	GWH	MRY	IR
DR	1	112	97	115	24	45
DB	133	1	64	128	116	124
DF	69	38	1	67	78	52
GWH	112	104	92	1	114	89
MRY	11	43	48	52	1	25
IR	32	75	53	66	41	1

[1]Deep Red (DR), Dutch Belted (DB), Dutch Friesian (DF), Groningen White Headed (GWH), Meuse-Rhine-Yssel (MRY), Improved Red (IR).
[2]For example, to achieve the same increase in accuracy as adding one individual from DR, 112 DB individuals are needed, 97 DF, 115 GWH, and so on.

et al., 2019), but similar to those of numerically smaller Brown Swiss population, 150 to 400 segments (Erbe et al., 2013), and Jersey, 470 segments (van den Berg et al., 2019). Because M_e is linked to the effective population size, it is expected that breeds with lower genetic diversity have smaller M_e. Based on the calculated LD, measured as the squared correlation between the alleles at 2 different loci (r^2), a recent study estimated N_e for several Red cattle breeds, including 6 breeds used here, and found lowest values for DB (47) and GWH (62), and highest for MRY (90) and DF (99) (Table 1), which corresponds to the M_e values we found. For all breeds except MRY, pedigrees were not available, hence M_e was estimated using only variation in genomic relationships between the individuals. When pedigree was used, within M_e for MRY increased from 307 to 677, indicating the effect of pedigree relationships. Note that out of 678 MRY individuals 171 did not have pedigree, hence these were excluded from the computation. Therefore, with and without pedigree comparison was based on the remaining 507 individuals. High relationships, such as those between full- and half-sibs, parent-offspring, and parent-grand offspring, can considerably add to the variance across all relationships in the population. Subtracting the pedigree from the genomic relationships before computing the variance of the genomic relationships (Equation 1) will lead to a lower variance and, therefore, an increase in the estimated M_e. The increase in within M_e due to considering pedigree relationships can be substantial, 2-fold as shown for MRY in our study, but even 3- to 4-fold, according to a recent simulation study (Marjanovic and Calus, 2020). When pedigrees are incomplete or not available, it is possible to derive parent-offspring relationships from genotypic information, which is done routinely in many national databases (Hayes, 2011; McClure et al., 2018), or to derive even lower degree relationships, and use this to build the pedigree relationship matrix. Although absence of pedigree may be an issue for single-breed genomic prediction, in small breeds, where emphasis is on multibreed genomic prediction, between M_e is required, which is not influenced by pedigree information unless recent introgression occurred. The computation of $n_{P,2}$, on the other hand, requires within M_e, and therefore numbers in the Table 3 may be overestimated. However, assuming that the effect of using pedigree is similar for all breeds, such that all within M_e values are equally affected, suggests that the pattern observed in terms of which breeds are most closely related stays the same.

The within M_e can be estimated using different formulas, such as those that rely on N_e and size of the genome (Hayes et al., 2009), or the number of eigenvalues that explain, for example, 98% (**EIG98**) of the variance of the genomic relationship matrix (Misztal,

2016; Pocrnic et al., 2016), in addition to the approach used in our study. The EIG98 approach proved to be effective when applied to large data sets (Pocrnic et al., 2016), but in our case with only 24 to 678 genotypes being available per breed, would not give meaningful results as the maximum value of EIG98 is by definition restricted to be lower than the number of genotypes used. Depending on the formula, estimates of M_e can vary significantly (Wientjes et al., 2013; Brard and Ricard, 2015), especially when N_e is used, because N_e can be estimated from either pedigree or genomic data and different computational methods (Leroy et al., 2013; Wang et al., 2016). Computing M_e as $2N_eL/\ln(4N_eL)$ (Goddard, 2009), assuming a genome size (L) of 31.6 Morgan (Ihara et al., 2004) and using N_e values reported in Table 1, gave values that were 1.6 to 3.8 times higher than the M_e values reported in Table 3 (results not shown). For MRY, the value based on this formula was 609, which was reasonably close the M_e value of 677 computed based on the variance of genomic minus pedigree relationships, in line with observations of Wientjes et al. (2013). Computing M_e based on the variance of genomic relationships as in our study allows us to consider specific characteristics of a population, such as population structure, and in addition this approach can be extended to 2 breeds, allowing for computation of between M_e, necessary to predict the accuracy for across-breed genomic prediction (Wientjes et al., 2013). A recent study by van den Berg et al. (2019) found that prediction accuracy using within M_e computed as done in our study resulted in overestimation of genomic prediction accuracy. In the study by van den Berg et al. (2019) the true within M_e may have been underestimated due to close relationships among some animals in the reference population, which could also be expected in numerically small breeds. Thus, although being sensitive to nonrandom sampling of genotype data from a population, estimating M_e from the variance of genomic relationships is attractive because it is straightforward to compute, and it is reasonably robust against using small sample sizes as long as these are a proper representation of the population (Marjanovic and Calus, 2020).

Between Population M_e

Between population M_e values ranged from ~3,500 to ~17,400, with lowest values observed between MRY and 2 breeds recently derived from MRY, being DR and IR (Table 3). Using similar data as in our study, van Breukelen et al. (2019) analyzed genetic distances between breeds with a neighbor-joining tree, which revealed that IR and DR bulls had a high mean similarity with MRY bulls, with IR bulls forming a separate

cluster in the neighbor-joining tree. Indeed, between population M_e from our study showed higher relatedness of MRY with DR (~3,500) than IR (~8,500), also confirmed by the PCA plot (Figure 1) and pairwise F_{ST} values, which could be explained by IR being selected for meat production by using Belgian Blue bulls to some extent (van Breukelen et al., 2019). For DB the lowest between M_e was with DF. The high relatedness between these breeds has also been observed in our PCA plot and in other studies using different measures of genetic relatedness (Eding and Meuwissen, 2001; van Breukelen et al., 2019), which also suggested that DF bulls have been used in the past to upgrade the DB breed. The most distant breed was GWH, where between M_e ranged from ~13,700 to ~17,400. In the PCA plot GWH also formed a distinct cluster away from the other breeds. Similarly, pairwise F_{ST} values were highest between GWH and all other breeds. Previous analysis of GWH showed the lowest level of admixture compared with other Dutch Red cattle breeds (van Breukelen et al., 2019). In general, between M_e followed the pattern of genetic relatedness estimated with alternative methods, which is a further confirmation that between M_e can be used as an indicator to select breeds suitable for a multibreed reference population.

The between M_e value was between 10× and 100× larger than within M_e, which is not surprising as LD structure, upon which M_e is dependent, is at least partly different between different breeds (de Roos et al., 2008; Larmer et al., 2014; Wientjes et al., 2015). This also illustrates challenges of across-breed genomic prediction, since to predict the GEBV of individuals from a certain population, effects of all independent segments need to be estimated (Meuwissen et al., 2013; Wientjes et al., 2016). The accuracy of genomic prediction increases as the number of segments decreases (Daetwyler et al., 2008). In a study on GWH, HF, and MRY, M_e between HF and MRY was 24,000 and between GWH and HF 18,000, approximately 10× higher than within M_e (Wientjes et al., 2015). Another study found M_e of ~15,400 between HF and Jersey, and ~18,400 between HF and Austrian Red, which is 15 to 32× higher than within M_e of these breeds (van den Berg et al., 2019). It has been speculated previously that M_e of about 20,000 could be used to predict the accuracy of across-breed genomic prediction for closely related breeds (Wientjes et al., 2015), which is higher than values found in our study. Hence, potentially all breeds in this study could be included in the reference population for each of the breeds. When there has been high exchange of genetic material in the past, the increase in reliabilities of GEBV can be substantial as shown on the example of Nordic Red cattle breeds (Brøndum et al., 2011; Heringstad et al., 2011; Zhou et al., 2014). For

breeds with weaker relationships, such as the frequently studied HF and Jersey, usually little improvement has been observed in accuracies of GEBV when breeds were combined together, except for small breeds when either higher SNP density are analyzed with Bayesian methods (Erbe et al., 2012; Calus et al., 2018) or a model component is added to model SNPs selected from sequence data (Raymond et al., 2018). It is clear from our results that 2 clusters can be identified among the Dutch Red breeds, one formed by MRY, DR, and IR, and another comprising DF and DB. Given the number of genotyped individuals (Table 1), MRY would benefit very little from inclusion of IR and DR in the reference population, but the opposite is the case for DR where adding MRY would substantially increase the reference population (Table 3). Similarly, for the smaller DB breed, adding DF individuals could be beneficial, assuming more individuals are genotyped in the future. Based on the relatedness between breeds within those 2 clusters there appears to be potential for across-breed genomic prediction, however this is still limited by the current number of genotyped individuals. Future efforts should be focused on increasing number of genotyped individuals of breeds within the clusters, where combining their reference populations may be beneficial when using a standard 50k SNP panel, whereas for combining breeds outside the cluster a higher SNP density may be needed.

In this study we did not consider larger mainstream breeds, although they may also be added to the reference population (e.g., HF for selection of DF, as these 2 breeds evolved from a common population). Increasing the size of the reference population with a larger breed will lead to a reference population being dominated by one breed and consequently reliability to predict the other breed may be reduced (van den Berg et al., 2020). To overcome this, a subset of the breed or modeling the same trait in different breeds as correlated traits in a multitrait model could be used (van den Berg et al., 2020). Nevertheless, the implication may be that the small breed is selected in the direction of the mainstream breed, making them more alike, instead of preserving the uniqueness of the small breed.

CONCLUSIONS

Within and between M_e of 6 analyzed Dutch Red cattle breeds showed high variability, indicating different levels of relatedness between the breeds, which can be used to tailor multibreed reference populations for each of those breeds. Based on our results, 2 clusters are especially important, one formed by MRY, DR, and IR, and the other comprising DF and DB. Although relatedness between breeds within each of these 2 clusters

is high, across-breed genomic prediction is still limited by the current number of genotyped individuals, which for many breeds is low. In particular, adding MRY individuals would increase the reference population of DR substantially. Future efforts should be focused on increasing number of genotyped individuals of breeds within the clusters.

REFERENCES

Abraham, G., and M. Inouye. 2014. Fast principal component analysis of large-scale genome-wide data. PLoS One 9:e93766. https://doi.org/10.1371/journal.pone.0093766.

Addo, S., J. Schäler, D. Hinrichs, and G. Thaller. 2017. Genetic diversity and ancestral history of the German Angler and the Red-and-White dual-purpose cattle breeds assessed through pedigree analysis. Agric. Sci. 8:1033–1047. https://doi.org/10.4236/as.2017.89075.

Brard, S., and A. Ricard. 2015. Is the use of formulae a reliable way to predict the accuracy of genomic selection? J. Anim. Breed. Genet. 132:207–217. https://doi.org/10.1111/jbg.12123.

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. https://doi.org/10.3168/jds.2010-3765.

Calus, M., and J. Vandenplas. 2016. Calc_grm—A program to compute pedigree, genomic, and combined relationship matrices. ABGC, Wageningen UR Livestock Research, Wageningen, the Netherlands.

Calus, M. P. L., M. E. Goddard, Y. C. J. Wientjes, P. J. Bowman, and B. J. Hayes. 2018. Multibreed genomic prediction using multitrait genomic residual maximum likelihood and multitask Bayesian variable selection. J. Dairy Sci. 101:4279–4294. https://doi.org/10.3168/jds.2017-13366.

CGN. 2019. Rassenlijst Nederlandse landbouwhuisdierrassen en hun risicostatus op basis van aantal volwassen vrouwelijke dieren in Nederland. Wageningen Univ. Res. https://edepot.wur.nl/418082.

Daetwyler, H. D., B. Villanueva, and J. A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One 3:e3395. https://doi.org/10.1371/journal.pone.0003395.

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. https://doi.org/10.1534/genetics.107.084301.

Eding, H., and T. H. E. Meuwissen. 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. J. Anim. Breed. Genet. 118:141–159. https://doi.org/10.1046/j.1439-0388.2001.00290.x.

Erbe, M., B. Gredler, F. R. Seefried, B. Bapst, and H. Simianer. 2013. A function accounting for training set size and marker density to model the average accuracy of genomic prediction. PLoS One 8:e81046. https://doi.org/10.1371/journal.pone.0081046.

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. 95:4114–4129. https://doi.org/10.3168/jds.2011-5019.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to Quantitative Genetics. 4th ed. Longman Group, Essex, UK.

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

Goddard, M. E., B. J. Hayes, and T. H. E. Meuwissen. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. J. Anim. Breed. Genet. 128:409–421. https://doi.org/10.1111/j.1439-0388.2011.00964.x.

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397. https://doi.org/10.1534/genetics.107.081190.

Hayes, B. J. 2011. Technical note: Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. J. Dairy Sci. 94:2114–2117. https://doi.org/10.3168/jds.2010-3896.

Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. (Camb.) 91:47–60. https://doi.org/10.1017/S0016672308009981.

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. Page 29 in Proc EAAP, Norway. Wageningen Academic Publishers, Wageningen, the Netherlands.

Hiemstra, S. J., Y. de Haas, A. Mäki-Tanila, and G. Gandini. 2010. Local cattle breeds in Europe: development of policies and strategies for self-sustaining breeds. 1st ed. Wageningen Academic Publishers, Wageningen, the Netherlands.

Hozé, C., S. Fritz, F. Phocas, D. Boichard, V. Ducrocq, and P. Croiseau. 2014. Efficiency of multi-breed genomic selection for dairy cattle breeds with different sizes of reference population. J. Dairy Sci. 97:3918–3929. https://doi.org/10.3168/jds.2013-7761.

Ihara, N., A. Takasuga, K. Mizoshita, H. Takeda, M. Sugimoto, Y. Mizoguchi, T. Hirano, T. Itoh, T. Watanabe, K. M. Reed, W. M. Snelling, S. M. Kappes, C. W. Beattie, G. L. Bennett, and Y. Sugimoto. 2004. A comprehensive genetic map of the cattle genome based on 3802 microsatellites. Genome Res. 14(10a):1987–1998. https://doi.org/10.1101/gr.2741704.

Larmer, S. G., M. Sargolzaei, and F. S. Schenkel. 2014. Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. J. Dairy Sci. 97:3128–3141. https://doi.org/10.3168/jds.2013-6826.

Lee, S. H., W. M. S. P. Weerasinghe, N. R. Wray, M. E. Goddard, and J. H. J. van der Werf. 2017. Using information of relatives in genomic prediction to apply effective stratified medicine. Sci. Rep. 7:42091. https://doi.org/10.1038/srep42091.

Legarra, A., O. F. Christensen, Z. G. Vitezica, I. Aguilar, and I. Misztal. 2015. Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. Genetics 200:455–468. https://doi.org/10.1534/genetics.115.177014.

Leroy, G., T. Mary-Huard, E. Verrier, S. Danvy, E. Charvolin, and C. Danchin-Burge. 2013. Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. Genet. Sel. Evol. 45:1. https://doi.org/10.1186/1297-9686-45-1.

Marjanovic, J., and M. P. L. Calus. 2020. Factors affecting accuracy of estimated effective number of chromosome segments for numerically small breeds. J. Anim. Breed. Genet. jbg.12512. https://doi.org/10.1111/jbg.12512.

Van Eijndhoven, M. 2014. Genetic variation of milk fatty acid composition between and within dairy cattle breeds. PhD thesis, Wageningen University, Wageningen, the Netherlands.

McClure, M. C., J. McCarthy, P. Flynn, J. C. McClure, E. Dair, D. K. O'Connell, and J. F. Kearney. 2018. SNP data quality control in a national beef and dairy cattle system and highly accurate SNP based parentage verification and identification. Front. Genet. 9:84. https://doi.org/10.3389/fgene.2018.00084.

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.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2013. Accelerating improvement of livestock with genomic selection. Annu. Rev. Anim. Biosci. 1:221–237. https://doi.org/10.1146/annurev-animal-031412-103705.

Misztal, I. 2016. Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. Genetics 202:401–409. https://doi.org/10.1534/genetics.115.182089.

Pocrnic, I., D. A. L. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2016. The dimensionality of genomic information and its effect on genomic prediction. Genetics 203:573–581. https://doi.org/10.1534/genetics.116.187013.

Pszczola, M., T. Strabel, H. A. Mulder, and M. P. L. Calus. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. J. Dairy Sci. 95:389–400. https://doi.org/10.3168/jds.2011-4338.

Raymond, B., A. C. Bouwman, Y. C. J. Wientjes, C. Schrooten, J. Houwing-Duistermaat, and R. F. Veerkamp. 2018. Genomic prediction for numerically small breeds, using models with pre-selected and differentially weighted markers. Genet. Sel. Evol. 50:49. https://doi.org/10.1186/s12711-018-0419-5.

Schäler, J., D. Hinrichs, and G. Thaller. 2019. The benefit of native uniqueness in a local red cattle breed from Northern Germany. J. Anim. Breed. Genet. 136:518–525. https://doi.org/10.1111/jbg.12421.

Stoop, M., G. Veninga, G. de Jong, and F. Reinhardt. 2017. Genomics in small populations: The MRY breed. Interbull Bull. 51:26–28.

Sved, J. A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor. Popul. Biol. 2:125–141. https://doi.org/10.1016/0040-5809(71)90011-6.

van Breukelen, A. E., H. P. Doekes, J. J. Windig, and K. Oldenbroek. 2019. Characterization of genetic diversity conserved in the gene bank for Dutch cattle breeds. Diversity (Basel) 11:229. https://doi.org/10.3390/d11120229.

van den Berg, I., I. M. MacLeod, C. M. Reich, E. J. Breen, and J. E. Pryce. 2020. Optimizing genomic prediction for Australian Red dairy cattle. J. Dairy Sci. 103:6276–6298. https://doi.org/10.3168/jds.2019-17914.

van den Berg, I., T. H. E. Meuwissen, I. M. MacLeod, and M. E. Goddard. 2019. Predicting the effect of reference population on the accuracy of within, across, and multibreed genomic prediction. J. Dairy Sci. 102:3155–3174. https://doi.org/10.3168/jds.2018-15231.

VanRaden, P. M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. J. Dairy Sci. 75:3136–3144. https://doi.org/10.3168/jds.S0022-0302(92)78077-1.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980.

Wang, J., E. Santiago, and A. Caballero. 2016. Prediction and estimation of effective population size. Heredity 117:193–206. https://doi.org/10.1038/hdy.2016.43.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Wientjes, Y. C. J. 2016. Multi-population genomic prediction. PhD thesis. Department of Animal Breeding and Genetics, Wageningen University, Wageningen, the Netherlands.

Wientjes, Y. C. J., P. Bijma, R. F. Veerkamp, and M. P. L. Calus. 2016. An equation to predict the accuracy of genomic values by combining data from multiple traits, populations, or environments. Genetics 202:799–823. https://doi.org/10.1534/genetics.115.183269.

Wientjes, Y. C. J., R. F. Veerkamp, P. Bijma, H. Bovenhuis, C. Schrooten, and M. P. L. Calus. 2015. Empirical and deterministic accuracies of across-population genomic prediction. Genet. Sel. Evol. 47:5. https://doi.org/10.1186/s12711-014-0086-0.

Wientjes, Y. C. J., R. F. Veerkamp, and M. P. L. Calus. 2013. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. Genetics 193:621–631. https://doi.org/10.1534/genetics.112.146290.

Zhou, L., B. Heringstad, G. Su, B. Guldbrandtsen, T. H. E. Meuwissen, M. Svendsen, H. Grove, U. S. Nielsen, and M. S. Lund. 2014. Genomic predictions based on a joint reference population for the Nordic Red cattle breeds. J. Dairy Sci. 97:4485–4496. https://doi.org/10.3168/jds.2013-7580.

ORCIDS

J. Marjanovic ⊙ https://orcid.org/0000-0001-7395-2529
B. Hulsegge ⊙ https://orcid.org/0000-0002-7015-6540
M. P. L. Calus ⊙ https://orcid.org/0000-0002-3213-704X