



Value of the Dutch Holstein Friesian germplasm collection to increase genetic variability and improve genetic merit

H. P. Doekes,*†¹ R. F. Veerkamp,* P. Bijma,* S. J. Hiemstra,† and J. Windig*†

*Animal Breeding and Genomics, and

†Centre for Genetic Resources the Netherlands, Wageningen University and Research, PO Box 338, 6700 AH, Wageningen, the Netherlands

ABSTRACT

National gene bank collections for Holstein Friesian (HF) dairy cattle were set up in the 1990s. In this study, we assessed the value of bulls from the Dutch HF germplasm collection, also known as cryobank bulls, to increase genetic variability and improve genetic merit in the current bull population (bulls born in 2010–2015). Genetic variability was defined as 1 minus the mean genomic similarity (SIM_{SNP}) or as 1 minus the mean pedigree-based kinship (f_{PED}). Genetic merit was defined as the mean estimated breeding value for the total merit index or for 1 of 3 subindices (yield, fertility, and udder health). Using optimal contribution selection, we minimized relatedness (maximized variability) or maximized genetic merit at restricted levels of relatedness. We compared breeding schemes with only bulls from 2010 to 2015 with schemes in which cryobank bulls were also included. When we minimized relatedness, inclusion of genotyped cryobank bulls decreased mean SIM_{SNP} by 0.7% and inclusion of both genotyped and nongenotyped cryobank bulls decreased mean f_{PED} by 2.6% (in absolute terms). When we maximized merit at restricted levels of relatedness, inclusion of cryobank bulls provided additional merit at any level of mean SIM_{SNP} or mean f_{PED} except for the total merit index at high levels of mean SIM_{SNP} . Additional merit from cryobank bulls depended on (1) the relative emphasis on genetic variability and (2) the selection criterion. Additional merit was higher when more emphasis was put on genetic variability. For fertility, for example, it was 1.74 SD at a mean SIM_{SNP} restriction of 64.5% and 0.37 SD at a mean SIM_{SNP} restriction of 67.5%. Additional merit was low to nonexistent for the total merit index and higher for the subindices, especially for fertility. At a mean SIM_{SNP} of 64.5%, for example, it was 0.60 SD for the total merit index and 1.74 SD for fertility. In conclusion, Dutch HF cryobank bulls can be used to increase genetic variability and improve genetic

merit in the current population, although their value is very limited when selecting for the current total merit index. Anticipating changes in the breeding goal in the future, the germplasm collection is a valuable resource for commercial breeding populations.

Key words: gene bank collection, genetic diversity, genetic improvement, conservation, dairy cow

INTRODUCTION

The Holstein Friesian (HF) breed is the dominating dairy cattle breed worldwide. Despite its census size of millions of individuals, the breed has an effective population size of 18 to 115 (Danchin-Burge et al., 2011; Rodríguez-Ramilo et al., 2015; Doekes et al., 2018a). In the early 1990s, national HF gene bank collections were established to safeguard genetic variability (Danchin-Burge et al., 2011). Since then, ex situ in vitro conservation has been used as a complementary strategy to in situ in vivo management of genetic variability in the breed (FAO, 2015).

In vitro conservation has several advantages and potential uses. One advantage is that the stored material harbors genetic variation of the population at the time of sampling, which may include variation that has since been lost in vivo due to selection and drift. Material from cryobank individuals, therefore, could be used to restore or increase genetic variability in the current live population (Leroy et al., 2011; Eynard et al., 2018). In an extreme scenario in which the live population becomes highly endangered or extinct (e.g., due to a disease outbreak), the stored material could also be used to re-establish the population. Other potential uses of gene bank collections include the management of inbreeding in small populations (Sonesson et al., 2002; Shepherd and Woolliams, 2004), the documentation of genetic trends (Smith, 1977; Garcia and Baselga, 2002), and the introgression of specific genetic variants into live populations (e.g., introgression of the polled allele).

Recently, Doekes et al. (2018a) reported a decrease in genetic variability in the Dutch-Flemish HF breeding program from 1986 through 2015, with a particularly fast decrease since the introduction of genomic selec-

Received June 15, 2018.

Accepted August 1, 2018.

¹Corresponding author: harmen.doekes@wur.nl

tion. This recent decrease suggests that old bulls from the Dutch HF germplasm collection could be used to increase variability in the current population.

A disadvantage of old cryobank bulls is that their genetic level, measured by EBV, is generally lower than that of recently born bulls. Consequently, the use of cryobank bulls to increase genetic variability in the current population is expected to reduce genetic merit. This hypothesis, however, does not necessarily hold for all traits because not all traits currently of interest have been continuously selected for in the past. For example, HF breeding goals consisted of mainly yield and conformation traits before 2000, whereas they now also include many traits related to health, reproduction, and longevity (Miglior et al., 2005; Cole and VanRaden, 2018; Doekes et al., 2018a). When the focus of the current breeding goal would shift toward one of the latter traits, cryobank bulls might have value for the population in terms of both genetic variability and genetic merit.

The aim of this study was to assess the value of the Dutch HF germplasm collection to increase genetic variability and improve genetic merit in the current bull population. We considered 3 scenarios: (1) maximizing genetic variability, (2) maximizing genetic merit for the total merit index while maintaining variability, and (3) maximizing genetic merit for a subindex (yield, fertility, or udder health) while maintaining variability. In addition to an SNP-based assessment, we performed a pedigree-based evaluation to include bulls from the germplasm collection that had no genotype data.

MATERIALS AND METHODS

Germplasm Collection, Groups, and Data

The Dutch HF germplasm collection was set up in 1993 and is now managed by the Centre for Genetic Resources the Netherlands of Wageningen University and Research (Wageningen, the Netherlands). The collection consists of 5,457 HF bulls (fraction HF >87.5%, either red or black). The majority of these bulls are from progeny testing schemes of 2 commercial companies: the Dutch-Flemish cattle improvement cooperative (CRV; Arnhem, the Netherlands) and Alta Genetics (Feerwerd, the Netherlands).

In this study, we used 5,783 HF bulls (both cryobank and noncryobank). To assess the additional value of the germplasm collection to the current bull population, we defined 4 groups: very young bulls (VYB; $n = 212$), young bulls (YB; $n = 762$), cryobank bulls with genotype data (CBG; $n = 2,888$), and cryobank bulls with only pedigree data (CBP; $n = 1,921$). The VYB consisted of bulls born in 2014 to 2015 with EBV based

on only genomic and parental information. The YB consisted of bulls born in 2010 to 2013 with EBV based on genomic and parental information (23% of bulls) or on genomic, parental, and daughter information (77% of bulls). The mean reliability of EBV was about 60% in the VYB and about 80% in the YB. Of the VYB and YB, 68 bulls (32%) and 551 bulls (72%), respectively, were also stored in the germplasm collection. The CBG consisted of genotyped cryobank bulls born in 1985 to 2009. The CBP consisted of cryobank bulls born in 1978 to 2015 that had pedigree but no genotype data. Figure 1 shows the number of bulls by group and year of birth.

Pedigree and genotype data were provided by CRV. The total pedigree comprised 429,981 individuals. Pedigree completeness per bull was assessed with the number of ancestral generations completely known (NCG) and the complete generation equivalent (CGE). We computed the CGE for each bull as the sum of $(1/2)^n$ over all its known ancestors, with n being the generation number of a given ancestor. Bulls with an NCG lower than 2 were excluded from the analyses ($n = 29$; these bulls had no genotype data and were excluded from the abovementioned group sizes). Genotyping was performed with the Illumina (San Diego, CA) BovineSNP50 BeadChip (versions v1 and v2) or CRV custom-made 60k Illumina panel (versions v1 and v2). Genotypes were imputed to 76k from the different panels following Druet and Georges (2010). Prior to imputation, SNP with a call rate lower than 0.85, a minor allele frequency lower than 0.025, or a difference of more than 0.15 between observed and expected heterozygosity were discarded. We also discarded SNP with unknown position on the Btau4.0 genome assembly. After quality control and imputation, the data set consisted of 75,538 autosomal SNP per bull.

Measures of Genetic Variability

Genetic variability was defined as 1 minus the mean relatedness in a population. We considered 2 measures of relatedness: the genomic similarity (SIM_{SNP}) and the pedigree-based kinship (f_{PED}).

The $SIM_{SNP_{ij}}$ was defined as the probability that 2 alleles at a random SNP, 1 sampled from bull i and 1 from bull j , were identical by state (Malécot, 1948). To calculate $SIM_{SNP_{ij}}$, we first computed a genomic relationship matrix for all genotyped individuals with allele frequencies fixed to 0.5 using *calc_grm* (Calus and Van-den-plas, 2013). We then scaled the obtained relationships (G_{ij}) to genomic similarities according to

$$SIM_{SNP_{ij}} = \frac{G_{ij} + 2}{4} \text{ (for derivation, see Additional File 1)}$$

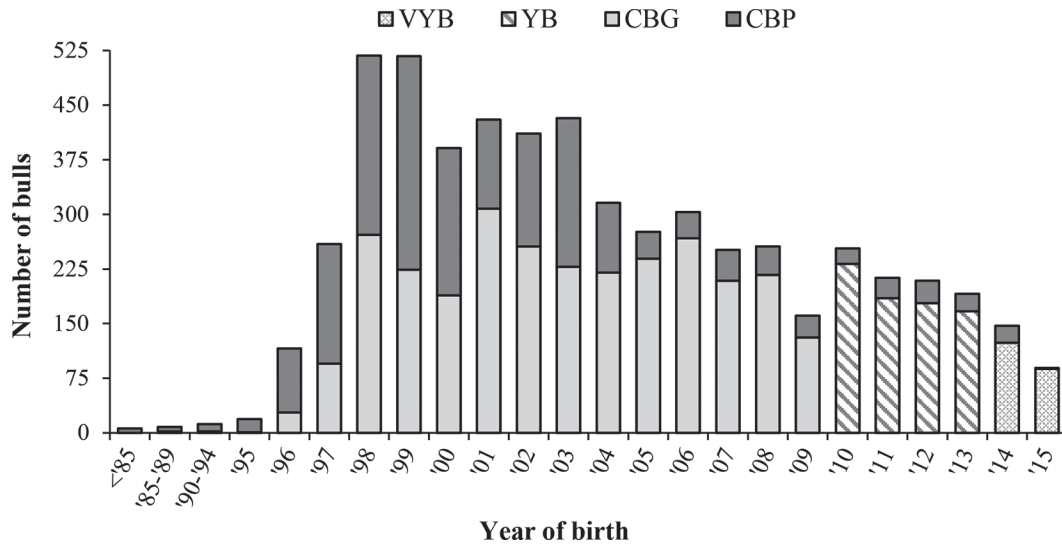


Figure 1. Number of bulls by year of birth and group. VYB = very young bulls ($n = 212$); YB = young bulls ($n = 762$); CBG = cryobank bulls with genotype data ($n = 2,888$); CBP = cryobank bulls with only pedigree data ($n = 1,921$).

of Eynard et al., 2015). Note that the mean SIM_{SNP} in a population is equal to 1 minus the expected heterozygosity.

The $f_{PED_{ij}}$ was defined as the pedigree-based probability that 2 alleles at a random (imaginary) selection-free locus, 1 sampled from bull i and 1 from bull j , were identical by descent with reference to a base population (Falconer and Mackay, 1996). Founders in the pedigree were considered as base population. We first computed an additive genetic relationship matrix for all bulls with *calc_grm* (Calus and Vandenplas, 2013) according to the algorithms of Sargolzaei et al. (2005) and Colleau (2002). We then obtained the $f_{PED_{ij}}$ as half of the additive genetic relationship.

Measures of Genetic Merit

Genetic merit was defined as the mean EBV for a selection index. We considered 4 selection indices: 1 total merit index (**NVI**) and 3 subindices—namely, yield (**INET**), daughter fertility (**FERT**), and udder health (**UH**). All EBV were obtained from the December 2017 publication of the organization for genetic evaluation of bulls in the Netherlands and Flanders (CRV, 2018a). The NVI is the Dutch-Flemish total merit index that includes INET, FERT, UH, longevity, conformation, and birth traits with relative weights (based on the sum of genetic SD) of 26, 14, 14, 11, 30, and 5%, respectively (CRV, 2018b). The breeding value for INET comprised the EBV for lactose yield, fat yield, and protein yield and was calculated as $INET = 0.3 \times \text{lactose yield} + 2.1 \times \text{fat yield} + 4.1 \times \text{protein yield}$ (CRV, 2018b). The

breeding value for FERT comprised the EBV for the interval between first and last insemination (IFL) and the calving interval and was calculated as $FERT = 0.52 \times (\text{IFL} - 100) + 0.52 \times (\text{calving interval} - 100) + 100$ (CRV, 2018b). The breeding value for UH comprised the EBV for subclinical mastitis (SCM) and clinical mastitis (CM) and was calculated as $UH = 0.477 \times (\text{SCM} - 100) + 0.641 \times (\text{CM} - 100) + 100$ (CRV, 2018b). The EBV for FERT and UH were rescaled such that the mean equaled 100 and the SD at population level was 4, whereas the NVI and INET were used on their original scales (CRV, 2018b).

Optimal Contribution Selection

Since its introduction in the late 1990s (Meuwissen, 1997; Grundy et al., 1998), optimal contribution selection (**OCS**) has become the gold standard to maximize the mean EBV in the next generation while restricting the mean relatedness to a predefined value. The restriction on relatedness is generally based on the desired rate of inbreeding (e.g., 0.5 or 1%). In addition to balancing genetic merit and variability, OCS may be used to maximize variability by minimizing relatedness irrespective of genetic gain (e.g., Eynard et al., 2018).

In this study, we used OCS to assess the value of cryobank bulls to the current bull population. We compared results of OCS schemes with only bulls from the current population with those with both bulls from the current population and cryobank bulls. More specifically, we considered the following 4 populations: (1) VYB, (2) VYB + YB, (3) VYB + YB + CBG, and (4) VYB + YB + CBG + CBP. For each population, we first

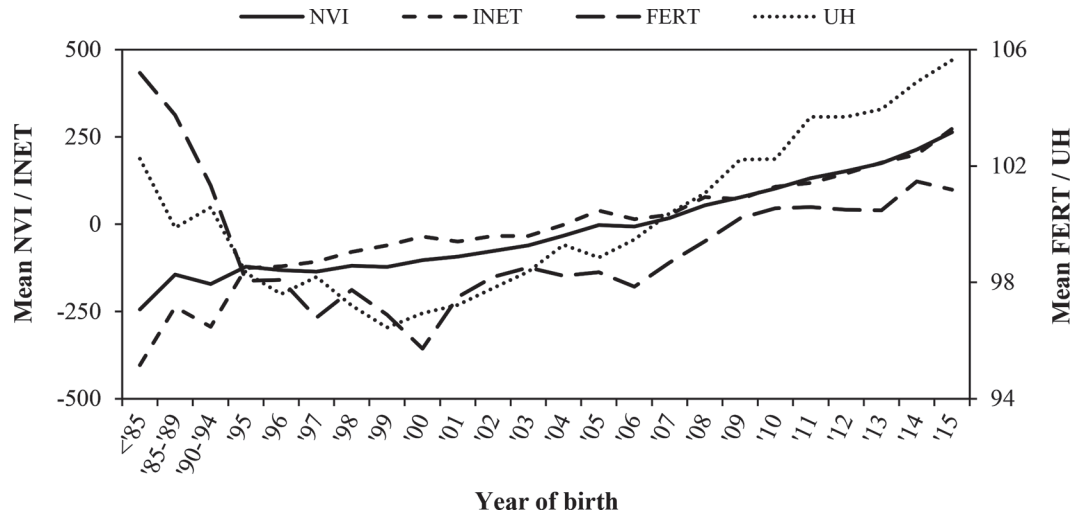


Figure 2. Mean EBV for 4 selection indices for all bulls ($n = 5,783$) by year of birth. NVI = total merit index; INET = yield index; FERT = daughter fertility index; UH = udder health index. Note that NVI and INET are shown on the primary y-axis and that FERT and UH are shown on the secondary y-axis.

ran OCS to maximize variability (i.e., minimize mean SIM_{SNP} or mean f_{PED}). We then maximized the mean EBV for NVI, INET, FERT, or UH while restricting the mean SIM_{SNP} or mean f_{PED} to predefined values. These predefined values ranged from a minimum (previously determined by minimizing SIM_{SNP} or f_{PED}) to a maximum of 68% for SIM_{SNP} and 12.5% for f_{PED} . The chosen maxima corresponded to a rate of inbreeding of about 5% when considering the VYB as the current generation. Between the minimum and maximum, we ran scenarios at intervals of 0.04% for SIM_{SNP} and of 0.08% for f_{PED} . Analyses were performed with Gencont software (Meuwissen, 2002), which uses the Lagrangian multiplier approach to solve the optimization problem (Meuwissen, 1997).

To compare results across traits, we present the differences in merit between populations not only on the original index scales but also in standard deviations (which equaled 128.3 for NVI, 136.0 for INET, 4.7 for FERT, and 5.1 for UH). To further visualize the value of cryobank bulls, we evaluated the total contribution that was assigned to each subgroup when running OCS with all genotyped bulls (VYB + YB + CBG) for SIM_{SNP} or with all bulls with pedigree (VYB + YB + CBG + CBP) for f_{PED} .

RESULTS

Genetic Trends

Mean NVI and INET have increased continuously over the last 30 yr (Figure 2). Mean FERT and mean

UH first decreased until they were included in the breeding goal around 2000. Since then, the genetic level for FERT and UH has steadily increased.

Descriptive Statistics Across Groups

Mean pedigree completeness (NCG and CGE), mean relatedness (SIM_{SNP} and f_{PED}), and mean EBV (for all traits) were greatest in the VYB followed by the YB, CBG, and CBP (Table 1). Means for the CBG and CBP were relatively similar compared with means in the VYB and YB. Cryobank bulls showed most variation in EBV (i.e., greatest SD) followed by the YB and the VYB. Cryobank bulls also showed the lowest minimum EBV followed by the YB and the VYB. The maximum of the total merit index NVI was smaller in the CBG (268) and CBP (287) than in the YB (374) and VYB (380). In fact, the maxima for NVI in the CBG and CBP were less than 1 SD above the mean EBV in the VYB. For the subindices (INET, FERT, and UH), however, the maximum was similar across groups, and the maxima for the CBG and CBP were in the upper tail of the corresponding distributions in the VYB and YB (i.e., above the mean +2 SD).

Maximizing Genetic Variability

Genetic variability was maximized by minimizing relatedness (either mean SIM_{SNP} or mean f_{PED}) with OCS. Minimization of SIM_{SNP} in the VYB decreased mean SIM_{SNP} from the current generation (65.9%) to the next generation (65.4%) by 0.5% (Table 2). This is

equivalent to a 0.5% increase in mean heterozygosity. When the YB was added to the VYB, the mean SIM_{SNP} in the next generation further decreased by 0.9% (to 64.5%). Adding the CBG resulted in a further decrease of 0.7%. In other words, inclusion of genotyped cryobank bulls increased expected heterozygosity by 0.7% when compared with a scheme in which only VYB and YB were used. Minimization of mean f_{PED} in the VYB decreased mean f_{PED} from the current generation (7.8%) to the next generation (7.0%) by 0.8%. Stepwise adding the YB, CBG, and CBP resulted in further decreases of 1.8, 1.8, and 0.8%, respectively. Thus, the inclusion of genotyped and nongenotyped cryobank bulls decreased mean f_{PED} by 2.6% when compared with the scenario in which only VYB and YB were used. Note that although in absolute terms the realized decrease in mean f_{PED} was larger than that for mean SIM_{SNP} , in relative terms (i.e., scaled by the noninbred part) they were quite similar. The difference between the VYB before OCS and the VYB + YB + CBG after OCS, for example, was 2.1% for SIM_{SNP} and 4.3% for f_{PED} in absolute terms and 6.1% for SIM_{SNP} and 4.7% for f_{PED} in relative terms.

The decrease in mean relatedness that was achieved by including additional groups of selection candidates

was accompanied by an increase in the number of candidates that was selected and by a decrease in the percentage of candidates that was selected. The increase in number of selected candidates was especially apparent when moving from VYB to VYB + YB for both SIM_{SNP} and f_{PED} and when moving from VYB + YB to VYB + YB + CBG for SIM_{SNP} (Table 2). Moving from the VYB to larger populations was also accompanied by a redistribution of contributions among selected bulls.

Maximizing genetic variability, irrespective of genetic merit, decreased mean EBV for all indices (Figure 3). For example, minimizing mean SIM_{SNP} in the VYB decreased the mean NVI from the current generation (241.1) to the next generation (227.8) by 13.3 points. Stepwise adding the YB and CBG resulted in further decreases of 107.7 points (to 120.1) and 193.1 points (to -73.0), respectively. Thus, there was a clear cost in merit when selecting only for variability.

Maximizing Genetic Merit While Maintaining Genetic Variability

The inclusion of additional groups of (old) selection candidates, such as cryobank bulls, resulted in more merit at the same level of variability (Figure 3). At a

Table 1. Descriptive statistics for pedigree completeness (2 measures), relatedness (2 measures), and EBV (4 selection indices) by group of bulls

Item ¹	Pedigree completeness ²		Relatedness ³		EBV ⁴			
	NCG	CGE	SIM_{SNP} (%)	f_{PED} (%)	NVI	INET	FERT	UH
Mean								
VYB	7.27	13.95	65.85	7.82	241.1	237.5	101.5	105.3
YB	6.61	12.67	65.25	6.91	141.8	139.5	100.7	103.4
CBG	5.92	10.86	64.48	5.20	-46.7	-7.38	98.0	98.5
CBP	5.71	10.48	NA ⁵	5.14	-78.9	-51.1	97.8	98.2
SD								
VYB	0.89	0.44	0.97	2.11	46.7	73.7	2.2	2.9
YB	1.12	0.55	1.05	2.39	64.7	100.6	3.6	3.5
CBG	0.98	0.72	0.88	2.05	99.6	117.1	5.0	4.9
CBP	1.07	0.91	NA	2.19	101.5	118.0	4.3	4.8
Minimum								
VYB	4	12.79	63.13	4.00	124	36	96	95
YB	2	10.59	62.43	2.23	-96	-181	88	89
CBG	2	7.38	61.92	0.48	-413	-410	76	80
CBP	2	5.90	NA	0.19	-438	-487	84	75
Maximum								
VYB	9	14.89	75.23	31.43	380	448	107	112
YB	9	14.06	75.78	32.38	374	415	111	114
CBG	8	12.71	75.75	33.22	268	472	113	111
CBP	9	14.06	NA	31.47	287	382	114	113

¹VYB = very young bulls (n = 212); YB = young bulls (n = 762); CBG = cryobank bulls with genotype data (n = 2,888); CBP = cryobank bulls with only pedigree data (n = 1,921).

²NCG = number of completely known generations; CGE = complete generation equivalent.

³ SIM_{SNP} = genomic similarity (excluding self-similarities); f_{PED} = pedigree-based kinship (excluding self-kinships).

⁴NVI = total merit index; INET = yield index; FERT = daughter fertility index; UH = udder health index.

⁵Not applicable.

mean SIM_{SNP} equal to the mean SIM_{SNP} of the current VYB (65.58%), for example, maximization for INET resulted in an INET of 322.8 when using VYB, 365.6 when using VYB + YB, and 389.3 when using VYB + YB + CBG. An exception was found for scenarios in which NVI was maximized at high levels of mean SIM_{SNP} . For these scenarios, the VYB + YB + CBG provided slightly less merit than the VYB + YB (see Discussion).

The benefit of including additional groups of (old) selection candidates, such as genotyped cryobank bulls, was greater when more emphasis was put on genetic variability. In other words, the difference between the curves in Figure 3 was greater at lower levels of relatedness. For example, the additional merit for FERT obtained by adding the CBG to the VYB + YB at mean SIM_{SNP} levels of 64.5, 65.5, 66.5, and 67.5% was 8.1 (1.74 SD), 2.4 (0.51 SD), 1.9 (0.4 SD), and 1.7 (0.37 SD), respectively.

The benefit of including additional groups of (old) selection candidates, such as genotyped cryobank bulls, at specific levels of relatedness differed across selection indices (Figure 4). The additional merit obtained by adding the CBG to the VYB + YB at a mean SIM_{SNP} of 64.5%, for example, was 0.60 SD (71.2 points) for NVI, 1.06 SD (144.3 points) for INET, 1.74 SD (8.1 points) for FERT, and 1.18 SD (6.0 points) for UH. For NVI, there was no additional merit of including the CBG at mean SIM_{SNP} above 65.24%. For UH, there was no additional merit (i.e., <0.01 SD) of the CBG at mean SIM_{SNP} above 65.80%. For INET, additional merit of the CBG was relatively stable (~0.16 SD) for mean SIM_{SNP} above 65%. Of the 4 indices, FERT showed the greatest additional merit for the CBG. Although the additional merit for FERT also decreased when mean SIM_{SNP} increased, there was still benefit of including CBG at high levels of mean SIM_{SNP} (e.g., 0.36 SD more merit at a mean SIM_{SNP} of 67.5%).

Results for f_{PED} were similar to those for SIM_{SNP} (Figure 3). The VYB + YB + CBG + CBP resulted in more genetic merit compared with the VYB + YB + CBG, but this difference was present only at very low levels of mean f_{PED} . For NVI, INET, and FERT, the additional merit of the CBP quickly decreased with increasing f_{PED} , and for mean f_{PED} of $\geq 5\%$ it was approximately zero. For FERT, there was still a bit of additional merit at higher levels of mean f_{PED} (~0.13 SD).

Contributions of Groups

When SIM_{SNP} was minimized with all genotyped bulls (VYB + YB + CBG), 89% of the contributions were assigned to the CBG, 10% to the YB, and 1% to the VYB (Figure 5). When f_{PED} was minimized with all bulls available (VYB + YB + CBG + CBP), 64% of the contributions were assigned to the CBP, 34% to the CBG, 2% to the YB, and 0% to the VYB.

As expected from Figure 3, the contribution of cryobank bulls (CBG and CBP) generally decreased when the restriction on relatedness became less stringent. The exact pattern differed across selection indices. For the total merit index NVI, the contribution of cryobank bulls continued to decrease with increasing levels of relatedness. At very high levels of relatedness, only bulls from the VYB (~70%) and YB (~30%) were selected. For INET, the contribution of cryobank bulls also decreased with increasing relatedness but remained stable (at ~40%) for mean SIM_{SNP} of $\geq 65.5\%$ and for mean f_{PED} of $\geq 8\%$. For FERT, about 90% of the contributions were assigned to cryobank bulls at any level of relatedness. For UH, the contribution of cryobank bulls also decreased with an increase in relatedness. This index showed a relatively high contribution of the YB. For UH, a single bull in the CBP (born in 2004) with a very high EBV (113) was assigned a contribution of about 15% in the scenarios with high mean f_{PED} .

Table 2. Mean relatedness before and after minimizing relatedness with optimal contribution selection as well as number of selected bulls (n_sel) for 4 populations and 2 relatedness measures

Population ¹	No.	Minimized relatedness measure ²					
		SIM_{SNP}			f_{PED}		
		Before (%)	After (%)	n_sel	Before (%)	After (%)	n_sel
VYB	212	65.85	65.37	91	7.82	7.02	127
VYB + YB	974	65.33	64.51	151	7.02	5.24	232
VYB + YB + CBG	3,862	64.50	63.78	225	5.26	3.45	236
VYB + YB + CBG + CBP	5,783	NA ³	NA	NA	5.16	2.63	247

¹VYB = very young bulls (n = 212); YB = young bulls (n = 762); CBG = cryobank bulls with genotype data (n = 2,888); CBP = cryobank bulls with only pedigree data (n = 1,921).

² SIM_{SNP} = genomic similarity; f_{PED} = pedigree-based kinship.

³Not applicable.

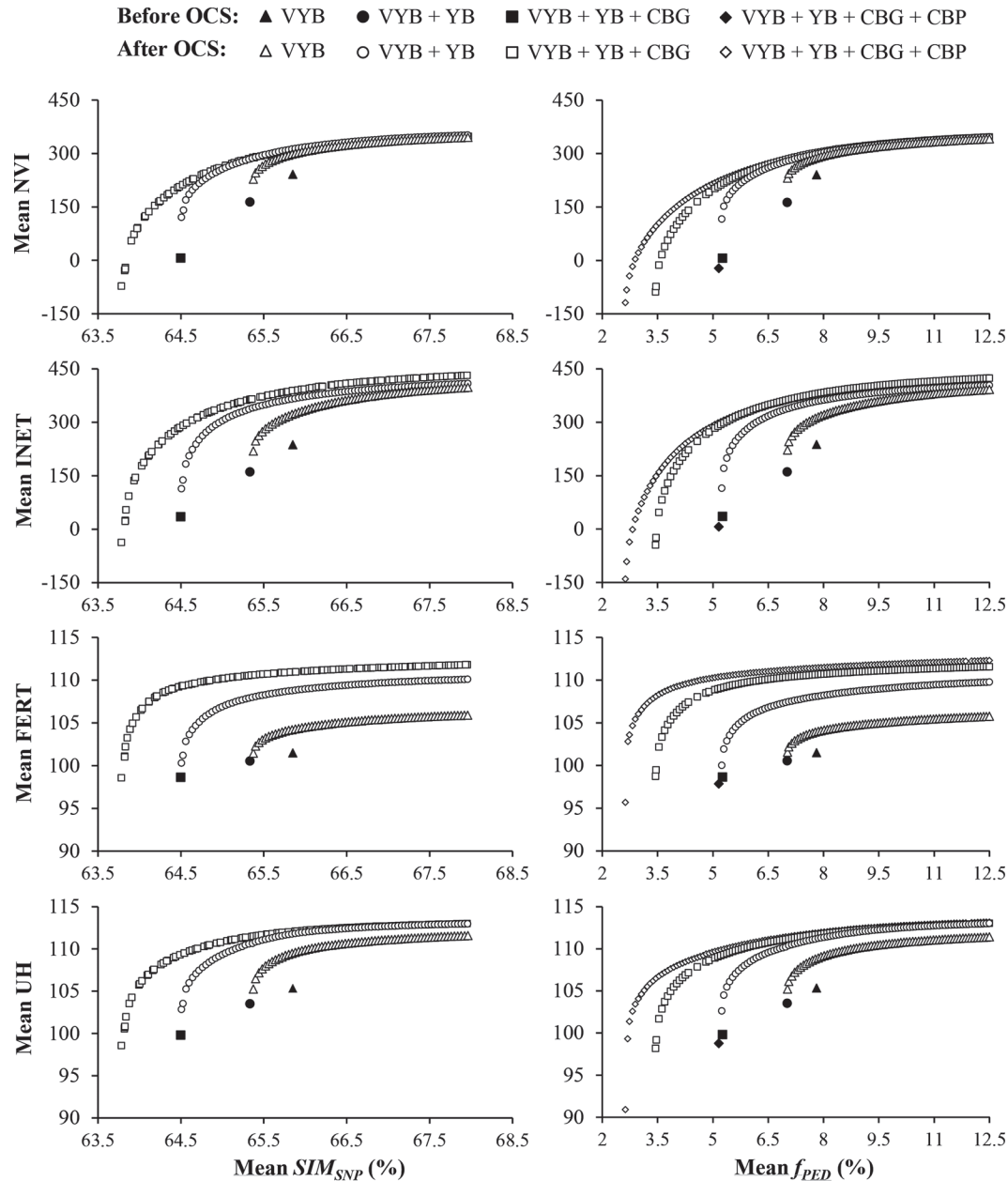


Figure 3. Mean genetic merit at restricted levels of relatedness (SIM_{SNP} left, f_{PED} right) before and after maximizing merit for 1 of 4 selection indices (NVI = total merit index; INET = yield index; FERT = fertility index; UH = udder health index) with optimal contribution selection (OCS) in 4 populations. Levels of relatedness ranged from a minimum, obtained by minimizing SIM_{SNP} or f_{PED} , to a maximum of 68 and 12.5% for SIM_{SNP} and f_{PED} , respectively. SIM_{SNP} = genomic similarity; f_{PED} = pedigree-based kinship. VYB = very young bulls (n = 212); YB = young bulls (n = 762); CBG = cryobank bulls with genotype data (n = 2,888); CBP = cryobank bulls with only pedigree data (n = 1,921).

DISCUSSION

Value of the Germplasm Collection

Our primary objective was to assess the value of the Dutch HF germplasm collection to increase genetic variability and improve genetic merit in the current

bull population. The selection of almost exclusively cryobank bulls when maximizing genetic variability (Figure 5) shows that diversity of the current bull population is well captured by the germplasm collection. This finding is in line with the results of Danchin-Burge et al. (2011). Results based on minimization of mean SIM_{SNP} and mean f_{PED} furthermore suggest that the

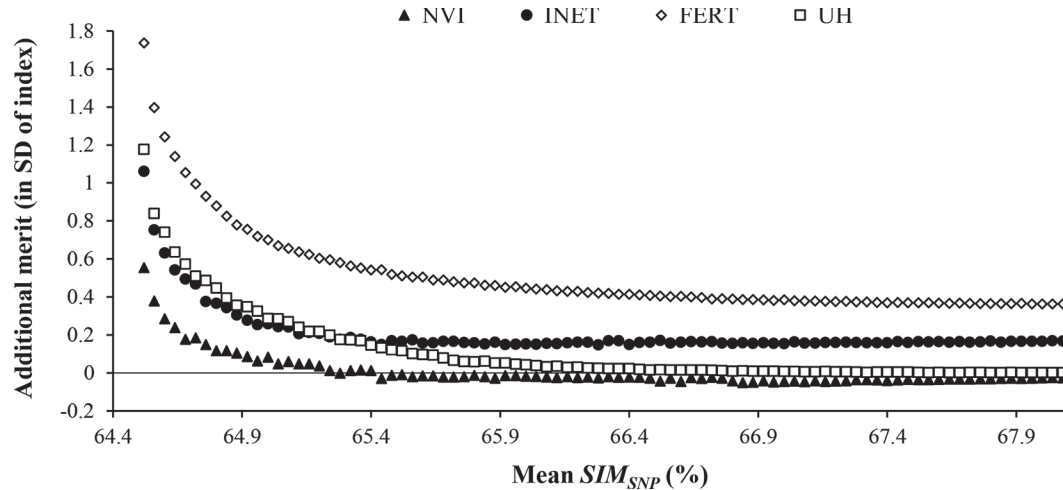


Figure 4. Additional merit (expressed in SD of selection indices) achieved with VYB + YB + CBG compared with VYB + YB at various levels of genomic similarity (SIM_{SNP}). VYB = very young bulls ($n = 212$); YB = young bulls ($n = 762$); CBG = cryobank bulls with genotype data ($n = 2,888$). NVI = total merit index; INET = yield index; FERT = fertility index; UH = udder health index.

germplasm collection can be used to increase genetic variability in the current population (Table 2). When genetic merit and genetic variability are balanced, the inclusion of cryobank bulls in addition to bulls from the current population may result in more merit at the same level of variability or, equivalently, in more variability at the same level of merit (Figure 3). We found that the benefit of using cryobank bulls depended on 2 factors: (1) the relative emphasis on variability and (2) the selection criterion (i.e., the index).

The additional merit of cryobank bulls, and the percentage of contributions assigned to cryobank bulls, was greater when more emphasis was put on genetic variability (Figures 3 and 4). This is not surprising because both relatedness and genetic merit have increased over time (Figure 2; Doekes et al., 2018a).

Additional merit of cryobank bulls was relatively low for the current total merit index NVI and greater for the subindices INET, UH, and FERT (Figures 3 and 4). For the NVI, there was almost no value in including cryobank bulls as selection candidates (except when we put a lot of emphasis on genetic variability). The limited additional merit for the NVI can be explained by the fact that the NVI has been the main index under selection in recent decades and that bulls from the current population simply have the highest NVI (Figure 2; Table 1). This finding also suggests that selection for NVI in the past has been effective, at least for the EBV. The observed difference between additional merit of cryobank bulls for the NVI on one hand and the subindices on the other reflects the principle that “single-trait selection often suffers from antagonistic correla-

tions with traits not in the selection objective,” whereas “multiple-trait selection avoids those problems at the cost of less-than-maximal progress for individual traits” (Cole and VanRaden, 2018). Past selection for NVI has resulted in less-than-maximal progress for the subindices. Some bulls that were assigned high contributions in this study when selecting for a subindex (e.g., yield), may not have been used so much in practice because they scored relatively low for other traits. The relatively high additional merit for FERT can be explained by the availability of cryobank bulls with high FERT, which were born before the intense selection for yield traits in the last decades of the 20th century and, thus, before the associated decrease in FERT (Figure 2) that was due to the well-known antagonistic correlation between fertility and yield traits.

Cryobank bulls with only pedigree data (i.e., CBP) were assigned up to 70% of the contributions when maximizing merit at low levels of mean f_{PED} (Figure 5). This finding suggests that, based on pedigree, there are bulls in the CBP that are of interest with regard to genetic variability. To determine whether these bulls are also less related at the genomic level or whether they were simply selected by OCS because they had limited pedigree completeness, they would have to be genotyped.

It is important to note that the actual breeding program is more dynamic than the schemes we evaluated. First, selection in practice occurs in both sexes, whereas we considered only bulls. We believe that the use of only bulls was sufficient for the purpose of this study because genetic variability in the population is

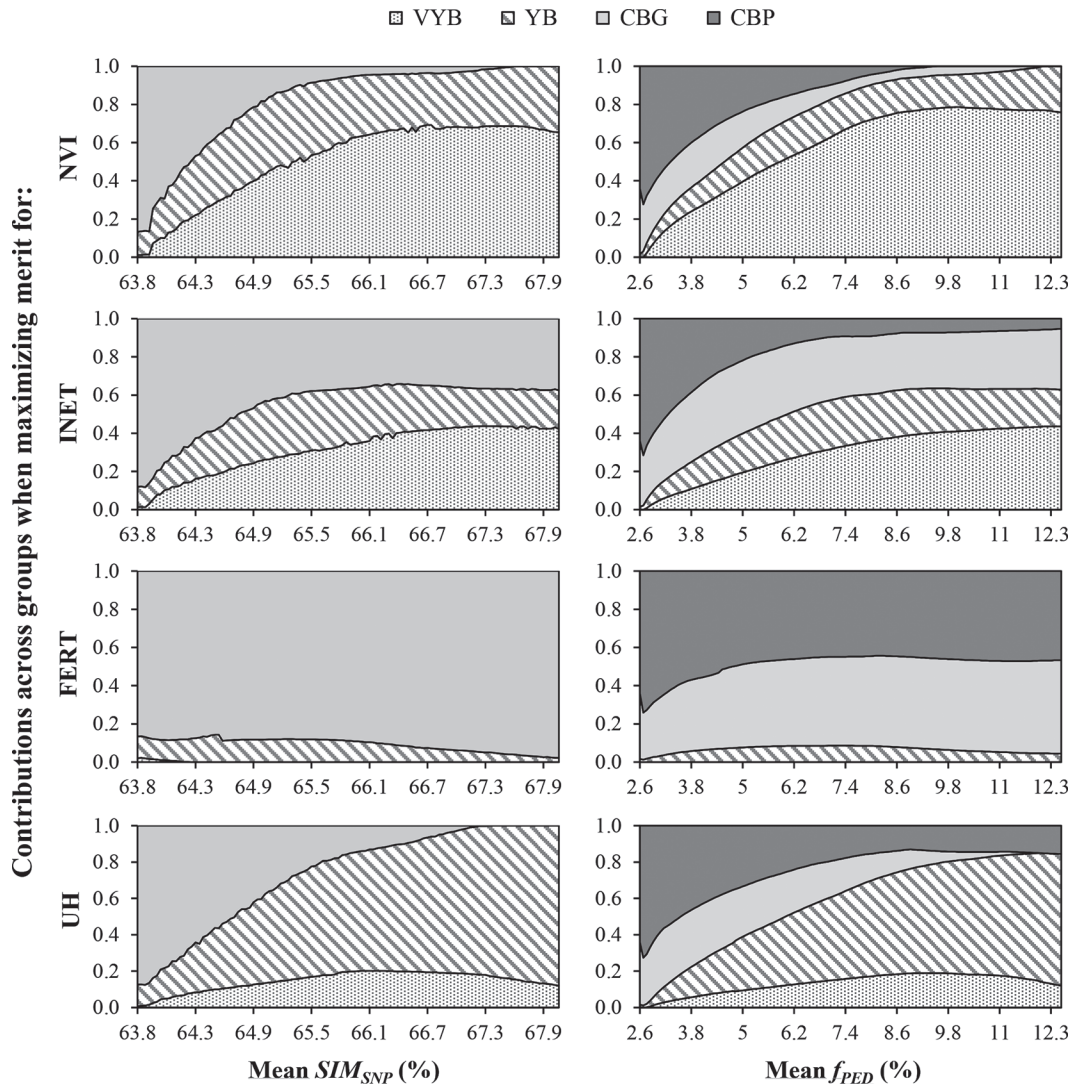


Figure 5. Contributions across groups when maximizing genetic merit for 1 of 4 selection indices (NVI = total merit index; INET = yield index; FERT = fertility index; UH = udder health index) at restricted levels of mean SIM_{SNP} in the VYB + YB + CBG population (left) or at restricted levels of mean f_{PED} in the VYB + YB + CBG + CBP population (right). SIM_{SNP} = genomic similarity; f_{PED} = pedigree-based kinship. VYB = very young bulls (n = 212); YB = young bulls (n = 762); CBG = cryobank bulls with genotype data (n = 2,888); CBP = cryobank bulls with only pedigree data (n = 1,921).

largely driven by relatedness between bulls (due to substantial use of AI). Nevertheless, selection at the cow side may offer some possibilities for management of genetic variability, especially because cows show lower relatedness levels than AI bulls (Doekes et al., 2018b). Second, the definition of the current bull population is not straightforward. We focused on bulls from a single breeding program, whereas in practice bulls from other breeding programs (from other countries) also may be used as selection candidates. Furthermore, we used 2 groups of bulls to represent the current bull population: 1 consisting of only VYB, born in 2013 to 2015, and

1 consisting of VYB and YB, born in 2010 to 2015. In practice, mostly top bulls from the VYB will be selected for breeding together with some top bulls from earlier years. To investigate the effect of focusing on top bulls, we performed OCS using the 50 top bulls from the VYB with the highest NVI. For any mean SIM_{SNP} above 66.3%, the NVI obtained with the top bulls was approximately identical to the NVI obtained with the entire VYB. When using only top bulls, a mean SIM_{SNP} below 66.3% could not be realized, whereas a mean SIM_{SNP} of 65.4% could be realized when using the entire VYB. This finding emphasizes that focusing

only on top bulls further reduces genetic variability and highlights the importance of also storing lower ranked bulls in the gene bank collection.

In this study, we considered only a single generation of OCS. As shown by Leroy et al. (2011), using cryobank bulls to increase genetic variability in a single generation will have no effect in the long term if their offspring are not selected subsequently. Genetic variability, therefore, should receive attention in subsequent generations as well. Continuous use of OCS may ensure that offspring of cryobank bulls are selected depending on the relative emphasis put on genetic variability. It would be interesting to investigate the decrease in long-term contributions of cryobank bulls considering various constraints on loss of genetic variability.

Limit of the Lagrangian Multiplier Approach

An unexpected result was found when maximizing NVI at high levels of mean SIM_{SNP} in the VYB + YB and VYB + YB + CBG populations. For these scenarios, the obtained NVI was greater for the VYB + YB than for the VYB + YB + CBG (although the difference was small; i.e., <0.05 SD; Figure 4). This result is theoretically impossible because all bulls in the former population were also part of the latter population. If the CBG would have provided no additional merit at all, the obtained NVI at a given mean SIM_{SNP} for the VYB + YB + CBG should have been at least equal to that for the VYB + YB. The observed pattern, therefore, has to be an artifact of the used algorithm. A difficulty of running OCS with the Lagrangian multiplier approach is that some contributions in the obtained optimal solution may be negative (Meuwissen, 1997). This problem is remedied in Gencont by eliminating selection candidates with negative contributions and repeating the optimization procedure until no negative contributions remain. The drawback of this approach, however, is that some of the candidates that were eliminated in early iterations may have had a positive contribution in the true optimal solution (Pong-Wong and Woolliams, 2007; Woolliams et al., 2015). When we compared high mean SIM_{SNP} scenarios between the VYB + YB and the VYB + YB + CBG, we observed that some bulls with moderate to high contributions in the former population were not selected at all in the latter population. This could be due to the elimination procedure. A possibility to get closer to the true optimal solution is to remove only a single candidate per iteration (the one with the most negative contribution) instead of eliminating all candidates with negative contributions. This remedy, however, would increase

computation time. Alternatively, one may consider a completely different approach to solve the optimization problem, such as semidefinite programming (Pong-Wong and Woolliams, 2007).

Future Perspectives for HF Germplasm Collections

The Dutch HF germplasm collection is a rather unique collection containing material from many AI bulls over a period of approximately 40 yr. In addition to the national gene bank collection, AI companies and farmers have stored germplasm over time. These companies and individual farmers, however, cannot guarantee the availability of stored material in the long term. Systematically storing genetic material in national collections, therefore, is required to safeguard the material for future use.

Various national HF collections exist across the globe. For the collections of the Netherlands, France, and the United States, Danchin-Burge et al. (2011) showed that there is substantial overlap in terms of the stored (pedigree-based) variability. An interesting question is whether national collections should be combined to reduce storage redundancy. We believe that although cooperation between gene banks is important to efficiently allocate resources and ensure that global HF variability is stored, there is substantial value in having separate national collections. The main advantage of having separate collections is that material from a national collection is more readily available for local breeding programs. In addition, having separate gene bank collections is an insurance against calamities.

An important aspect of gene bank management is to determine which and how many individuals are stored in a collection. In the case of the Dutch HF, a vast majority of AI bulls have been stored in the collection over time, with generally 25 straws per bull. These pre-selected AI bulls will not have covered all genetic variability and genetic potential for various traits present in the national population. To optimize collections, one could also store material from bulls that are not used for AI (and use OCS to determine exactly which bulls to store). The number of straws stored per individual is important with regard to the potential use, and therefore the value, of the collection. Today, fast genetic gain in HF is realized by producing many embryos through superovulation (Jaton et al., 2016), which requires a lot of semen. One may question whether cryobank bulls, with a limited number of straws, can make a significant contribution in the current system. Producing 90% of all offspring in 1 generation with cryobank bulls (such as in the hypothetical FERT scenario in this study), for

example, is not possible. Across multiple generations, however, gene bank material could be disseminated throughout the population.

In line with the simulation study of Leroy et al. (2011), our results suggest that gene bank collections are mostly valuable when the aim is to increase genetic variability or when major changes in selection objectives or practices occur (and when the use of animals from other breeds is not preferred). Breeding goals for HF have changed in the past (Miglior et al., 2005; Cole and VanRaden, 2018; Doekes et al., 2018a) and are expected to change further in the future. A complete shift from the total merit index to a subindex such as FERT is very unlikely. Instead, we expect that relative weights for specific trait groups will gradually shift over time and that novel traits will be added to the breeding goal. Various factors may influence the shift in breeding goal composition, including production economics (e.g., milk quota), societal demands (e.g., animal welfare), environment (e.g., climate change), technology (e.g., mid-infrared spectroscopy), and breeding value estimation (e.g., genomic prediction). By separating phenotype recording from the selection process, genomic prediction has removed the need for large-scale phenotyping and enabled selection for novel traits that are difficult to measure. An overview of such novel traits is provided by Boichard and Brochard (2012), Egger-Danner et al. (2015), and Cole and VanRaden (2018). Cryobank bulls may have relatively high EBV for novel traits because these traits have not been directly selected for in the past. Anticipating changes in the HF breeding goal in the future, the germplasm collection is a valuable resource in terms of both genetic variability and genetic merit.

CONCLUSIONS

Bulls from the Dutch HF germplasm collection can be used to increase genetic variability in the current breeding population. When genetic merit and genetic variability are to be balanced, the benefit of including cryobank bulls as selection candidates in addition to bulls from the current population depends on (1) the relative emphasis on genetic variability and (2) the selection criterion. Additional merit from cryobank bulls is higher when more emphasis is put on variability. Additional merit from cryobank bulls is very low for the current total merit index but higher for the subindices INET, UH, and FERT (especially high for fertility). Anticipating changes in the HF breeding goal in the future, the germplasm collection is a valuable resource for commercial breeding populations in terms of both genetic variability and genetic merit.

ACKNOWLEDGMENTS

The research leading to these results was conducted as part of the IMAGE project, which received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 677353. In addition, the Dutch Ministry of Agriculture, Nature and Food Quality (The Hague, the Netherlands) contributed financially (KB-21-004-003). The authors acknowledge CRV (Arnhem, the Netherlands) for providing pedigree and genotype data and thank Ina Hulsegge (Animal Breeding and Genomics, Wageningen University and Research, Wageningen, the Netherlands) for technical support.

REFERENCES

- Boichard, D., and M. Brochard. 2012. New phenotypes for new breeding goals in dairy cattle. *Animal* 6:544–550.
- Calus, M. P. L., and J. Vandenplas. 2013. Calc_grm—A programme to Compute Pedigree, Genomic, and Combined Relationship Matrices. Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Wageningen, the Netherlands.
- Cole, J. B., and P. VanRaden. 2018. Symposium review: Possibilities in an age of genomics: The future of selection indices. *J. Dairy Sci.* 101:3686–3701.
- Colleau, J. J. 2002. An indirect approach to the extensive calculation of relationship coefficients. *Genet. Sel. Evol.* 34:409–421.
- CRV. 2018a. Zoek stier. Accessed Jan. 29, 2018. <https://www.cooperatie-crv.nl/zoek-stier/>.
- CRV. 2018b. E-Hoofdstukken. Accessed Jan. 29, 2018. <https://www.cooperatie-crv.nl/downloads/fokwaarden/methodiek-2/e-hoofdstukken/>.
- Danchin-Burge, C., S. J. Hiemstra, and H. Blackburn. 2011. Ex situ conservation of Holstein-Friesian cattle: Comparing the Dutch, French, and US germplasm collections. *J. Dairy Sci.* 94:4100–4108.
- Doekes, H. P., R. F. Veerkamp, P. Bijma, S. J. Hiemstra, and J. J. Windig. 2018a. Trends in genome-wide and region-specific genetic diversity in the Dutch-Flemish Holstein-Friesian breeding program from 1986 to 2015. *Genet. Sel. Evol.* 50:15.
- Doekes, H. P., R. F. Veerkamp, S. J. Hiemstra, P. Bijma, S. van der Beek, and J. J. Windig. 2018b. Genomic selection and inbreeding and kinship in Dutch-Flemish Holstein Friesian cattle. Page 90 in *Proc. 11th World Congr. Genet. Appl. Livest. Prod.*, Auckland, New Zealand. Massey University, Palmerston North, New Zealand.
- Druet, T., and M. Georges. 2010. A hidden Markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. *Genetics* 184:789–798.
- Egger-Danner, C., J. Cole, J. Pryce, N. Gengler, B. Heringstad, A. Bradley, and K. F. Stock. 2015. Invited review: Overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. *Animal* 9:191–207.
- Eynard, S. E., J. J. Windig, I. Hulsegge, S. J. Hiemstra, and P. L. Calus Mario. 2018. The impact of using old germplasm on genetic merit and diversity—A cattle breed case study. *J. Anim. Breed. Genet.* <https://doi.org/10.1111/jbg.12333>.
- Eynard, S. E., J. J. Windig, G. Leroy, R. Van Binsbergen, and M. P. Calus. 2015. The effect of rare alleles on estimated genomic relationships from whole genome sequence data. *BMC Genet.* 16:24.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th ed. Longman, Harlow, UK.
- FAO (Food and Agriculture Organization of the United Nations). 2015. Part 3: The state of capacities. Section D: Conservation programmes. Pages 295–307 in *The Second Report on the State of*

- the World's Animal Genetic Resources for Food and Agriculture. B. D. S. D. Pilling, ed. FAO Commission on Genetic Resources for Food and Agriculture Assessments, Rome, Italy.
- Garcia, M., and M. Baselga. 2002. Estimation of genetic response to selection in litter size of rabbits using a cryopreserved control population. *Livest. Prod. Sci.* 74:45–53.
- Grundy, B., B. Villanueva, and J. A. Woolliams. 1998. Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genet. Res.* 72:159–168.
- Jaton, C., A. Koeck, M. Sargolzaei, F. Malchiodi, C. A. Price, F. S. Schenkel, and F. Miglior. 2016. Genetic analysis of superovulatory response of Holstein cows in Canada. *J. Dairy Sci.* 99:3612–3623.
- Leroy, G., C. Danchin-Burge, and E. Verrier. 2011. Impact of the use of cryobank samples in a selected cattle breed: A simulation study. *Genet. Sel. Evol.* 43:36.
- Malécot, G. 1948. *Mathématiques de l'Hérédité*. Masson & Cie, Paris, France.
- Meuwissen, T. H. E. 1997. Maximizing the response of selection with a predefined rate of inbreeding. *J. Anim. Sci.* 75:934–940.
- Meuwissen, T. H. E. 2002. GENCONT: An operational tool for controlling inbreeding in selection and conservation schemes. Pages 769–770 in *Proc. 7th World Congr. Genet. Appl. Livest. Prod.*, Montpellier, France. Editions Quae, Montpellier, France.
- Miglior, F., B. L. Muir, and B. J. van Doormaal. 2005. Selection indices in Holstein cattle of various countries. *J. Dairy Sci.* 88:1255–1263.
- Pong-Wong, R., and J. A. Woolliams. 2007. Optimisation of contribution of candidate parents to maximise genetic gain and restricting inbreeding using semidefinite programming. *Genet. Sel. Evol.* 39:3–25.
- Rodríguez-Ramilo, S. T., J. Fernández, M. A. Toro, D. Hernández, and B. Villanueva. 2015. Genome-wide estimates of coancestry, inbreeding and effective population size in the Spanish Holstein population. *PLoS One* 10:e0124157.
- Sargolzaei, M., H. Iwaisaki, and J. J. Colleau. 2005. A fast algorithm for computing inbreeding coefficients in large populations. *J. Anim. Breed. Genet.* 122:325–331.
- Shepherd, R. K., and J. A. Woolliams. 2004. Minimising inbreeding in small populations by rotational mating with frozen semen. *Genet. Res.* 84:87–93.
- Smith, C. 1977. Use of stored frozen semen and embryos to measure genetic trends in farm livestock. *J. Anim. Breed. Genet.* 94:119–130.
- Sonesson, A. K., M. E. Goddard, and T. H. Meuwissen. 2002. The use of frozen semen to minimize inbreeding in small populations. *Genet. Res.* 80:27–30.
- Woolliams, J. A., P. Berg, B. S. Dagnachew, and T. H. E. Meuwissen. 2015. Genetic contributions and their optimization. *J. Anim. Breed. Genet.* 132:89–99.