Variance of gametic diversity and its application in selection programs

D. J. A. Santos,[1,2]* J. B. Cole,[3] T. J. Lawlor Jr.,[4] P. M. VanRaden,[3] H. Tonhati,[2] and L. Ma[1]*
[1]Department of Animal and Avian Sciences, University of Maryland, College Park 20742
[2]Departamento de Zootecnia, Universidade Estadual Paulista, Jaboticabal, 14884-900, Brazil
[3]Henry A. Wallace Beltsville Agricultural Research Center, Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350
[4]Holstein Association USA, Brattleboro, VT 05302-0808

ABSTRACT

The variance of gametic diversity $\left(\sigma^2_{\text{gamete}}\right)$ can be used to find individuals that more likely produce progeny with extreme breeding values. The aim of this study was to obtain this variance for individuals from routine genomic evaluations, and to apply gametic variance in a selection criterion in conjunction with breeding values to improve genetic progress. An analytical approach was developed to estimate σ^2_{gamete} by the sum of binomial variances of all individual quantitative trait loci across the genome. Simulation was used to verify the predictability of this variance in a range of scenarios. The accuracy of prediction ranged from 0.49 to 0.85, depending on the scenario and model used. Compared with sequence data, SNP data are sufficient for estimating σ^2_{gamete}. Results also suggested that markers with low minor allele frequency and the covariance between markers should be included in the estimation. To incorporate σ^2_{gamete} into selective breeding programs, we proposed a new index, relative predicted transmitting ability, which better utilizes the genetic potential of individuals than traditional predicted transmitting ability. Simulation with a small genome showed an additional genetic gain of up to 16% in 10 generations, depending on the number of quantitative trait loci and selection intensity. Finally, we applied σ^2_{gamete} to the US genomic evaluations for Holstein and Jersey cattle. As expected, the *DGAT1* gene had a strong effect on the estimation of σ^2_{gamete} for several production traits. However, inbreeding had a small impact on gametic variability, with greater effect for more polygenic traits. In conclusion, gametic variance, a potentially important parameter for selection programs, can be easily computed and is useful for improving genetic progress and controlling genetic diversity.

Key words: Mendelian sampling, gamete, heterozygosity, selective breeding, dairy cattle

INTRODUCTION

Since the introduction of marker-assisted selection and genomic selection, technological improvements have resulted in widespread incorporation of molecular information into genetic evaluations (Nejati-Javaremi et al., 1997; Meuwissen et al., 2001; Schaeffer, 2006). Increased prediction accuracy, along with reduced generation intervals, has made genomic selection an important tool for achieving fast progress in dairy selection programs (García-Ruiz et al., 2016). Despite concerns about inbreeding in selection and mating designs, most selection programs only consider breeding values when making selection decisions. Even with genomic selection models, genomic breeding value or PTA and evaluation of future progeny are mostly based on expected breeding values without consideration of the variability of those values due to Mendelian sampling.

In addition to breeding value or PTA, other selection strategies have been proposed to increase the rate of genetic progress. One idea was to select animals that will provide greater genetic gains in the future rather than choosing the best animals in the current population. Goiffon et al. (2017) showed improved genetic gains when selecting for the best gametes from a subset of individuals in a population. Segelke et al. (2014) discussed the potential use of the variation within groups of offspring, which allows the assignment of probabilities to obtain progeny with a breeding value over a given threshold, as well as the number of matings required. In a follow-up study, Bonk et al. (2016) showed how exact within-family genetic variation can be calculated using data from phased genotypes. Recently, Müller et al. (2018) proposed a new selection criterion based on the expected maximum haploid breeding value. Collectively, these studies suggest that the incorporation of variation of future gametic values into mating decisions

can improve genetic progress on top of the selection on breeding values.

However, a few questions need to be answered before the application of gametic variance to breeding programs, such as how to assess the variation of future gametic values of an individual, how large is the gametic variance, how to use this information for selection, and how to estimate the variance of gametic diversity and use it in existing genomic evaluations. In this study, we aimed to address these questions from a statistical point of view, demonstrating the equivalence between gametic variance and Mendelian sampling variance in the classical BLUP (pedigree) model. We also sought to explore how this variance can be used as a selection criterion in conjunction with breeding values, with the goal of maximizing future genetic gains. We propose an approach for estimating this variance from routine genomic evaluations, verifying the adequacy of the estimates for individuals with and without progeny, and estimating the variance of breeding values of future progeny for a given mating. Finally, we evaluate the application of gametic variance to improve the selection of dairy traits in the US Holstein and Jersey populations.

MATERIALS AND METHODS

Estimation of the Variance of Gametic Diversity

We refer to the variance of gametic diversity as σ^2_{gamete}, which is equivalent to half of the Mendelian sampling variance (Appendix A1). σ^2_{gamete} measures the deviation of progeny breeding values from parent average and can be calculated using the probabilities of transmission of alleles at all QTL from an individual to its gametes. Gametic variance represents the variability of all possible gametic values generated by the permutation and recombination of each parental chromosome. In fact, only the heterozygous loci of an individual contribute to σ^2_{gamete}, so we only consider heterozygous loci in the following text.

Let's first consider one locus. For a biallelic locus j of an individual i with allele substitution effect α_j, σ^2_{gamete} can be calculated from a binomial variance of $\sigma^2_{[j]} = np(1-p)\alpha_j^2$, where the probability of transmission of a reference allele to a gamete $p = 0.5$ and the number of alleles transmitted to a gamete $n = 1$. When 2 loci, j and k, are considered for an individual i, the resulting variance can be obtained as

$$\sigma^2_{[j+k]} = \sigma^2_{[j]} + \sigma^2_{[k]} + 2\sigma_{jk}$$

and $\sigma_{jk} = \left(p_{jk} - p_j p_k\right)\alpha_j \alpha_k,$ [1]

where $p_j = p_k = 0.5$, and p_{jk} is the probability that the 2 reference alleles of the 2 loci are transmitted together; p_{jk} can be obtained from the linkage phase and recombination rate between the 2 loci. For example, $p_{jk} = 0.25$ and $\sigma_{jk} = 0$ when the loci are in linkage equilibrium; $p_{jk} = 0.5$ and $\sigma_{jk} = 0.25\alpha_j\alpha_k$ when the 2 reference alleles are on the same chromosome and the loci are in complete linkage.

Extending this calculation from 2 loci to all QTL on the genome, the σ^2_{gamete} of individual i can be obtained as the sum across all N heterozygous QTL:

$$\sigma^2_{\text{gamete}} = \sum_{j=1}^{N} \sigma^2_{[j]} + 2\sum_{j=1}^{N} \sum_{k=j+1}^{N} \sigma_{jk}.$$

This can be represented in matrix format as follows:

$$\sigma^2_{\text{gamete}} = \left[\alpha_1 \ldots \alpha_N\right]\mathbf{M}\left[\alpha_1 \ldots \alpha_N\right]',$$ [2]

where $\alpha_j \left(j = 1, \ldots, N\right)$ are the allele substitution effects, and $\mathbf{M}$ is the (co)variance matrix of the Mendelian transmission probabilities for the N heterozygous loci:

$$\mathbf{M} = \begin{bmatrix} 0.25 & \cdots & al_{1,N}\left(-\dfrac{cM_{1,N}}{200} + 0.25\right) \\ \vdots & \ddots & \vdots \\ al_{N,1}\left(-\dfrac{cM_{N,1}}{200} + 0.25\right) & \cdots & 0.25 \end{bmatrix},$$

where al_{jk} is a phase indicator for loci j and k, with value 1 when both loci have the reference allele on the same chromosome and -1 otherwise; cM_{jk} is the genetic distance between the 2 loci (in centimorgans). Any 2 loci with genetic distance >50 cM on the same chromosome, or on different chromosomes, are assumed to be independent and thus have zero values for the corresponding elements of $\mathbf{M}$. When all the loci are independent,

$$\mathbf{M} = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 0.25 \end{bmatrix}.$$

Instead of using genetic distances, $\mathbf{M}$ can be set up when direct recombination rates are available.

To estimate gametic variance in real data where genomic evaluation is available, we proposed to use the estimated SNP effects to replace true QTL effects in Equation [2]. This approximation of QTL with SNP marker effects is similar to that described by Bonk et al. (2016). Note that using estimated SNP effects in [2] may bias the estimation due to the covariance between estimated effects of SNP in linkage disequilibrium (LD) and potential biases from shrunken estimators of SNP effects, which warrants further investigation.

Application of Gametic Variance in Selection Programs

A new selection strategy using σ^2_{gamete} can be proposed, focusing on the future genetic progress (Bijma et al., 2018). When a small proportion of animals are selected for breeding, σ^2_{gamete} can help identify those that are most likely to produce progeny with extreme breeding values. Assuming selection intensity is maintained across generations, the average genetic value of the animals selected in the future will be related to the variance of gametes of the selected animals in the current generation. The average breeding value transmitted to future progeny can be calculated by summing the expected value and i times the standard deviation of gametic diversity $\left(i\sigma_{\text{gamete}}\right)$. The selection intensity (i) represents the number of standard deviations between the population average and the average of selected individuals. The same intensity can be applied when using PTA as the expected value and σ_{gamete} as standard deviation, to obtain the mean breeding value transmitted to the selected individuals in the next generation. Similar approaches have been proposed by Lehermeier et al. (2017) via a usefulness criterion (**UC**) with genomic EBV (**GEBV**) and the standard deviation of a given mating.

Here, we propose a new selection criterion relative to the intensity of selection applied in the next generation (i_f) for an individual i (unknowing mating),

$$RPTA_i = PTA_i + \sigma_{\text{gamete}\_i} \times i_f, \qquad [3]$$

where $RPTA_i$ (relative PTA) is the average of the genetic values relative to the group of progeny that will be selected in the future (see Appendix A2). In addition, we introduce a new concept of coefficient of relative variation (**CRV**) as a measure of the variability of the additive genetic values (u) transmitted from an individual to its progeny (Appendix A3). The CRV of an individual i is defined as follows (where E indicates expected value):

$$CRV_i = \frac{\sigma_{\text{gamete}}}{0.5\sqrt{E\left(u_i^2\right)}}. \qquad [4]$$

Simulation

To verify the estimation of σ^2_{gamete} by genomic models and the use of this new parameter to aid selection, we simulated different scenarios with various QTL, genotype, and phenotype data using the QMSim version 1.10 software (Sargolzaei and Schenkel, 2009). In brief, we simulated a historical population, a 10-generation recent population, and a 10-generation future population (Table 1).

To mimic real populations, a historical population was simulated with the same proportion of males and females that were mated randomly. This population was generated in 3 phases: the first phase consisted of 500 generations with a constant population size of 1,000 individuals; the second phase had 500 generations with a constant reduction of population size from 1,000 to 200 to generate LD and establish drift-mutation balance; and the third phase included 10 generations of expansion, where the population size increased from 200 to 3,000. From the last generation of this historical population, 200 males and 800 females were randomly selected as founders to generate the study population, which consisted of 10 generations with 5 progeny per dam and a ratio of 50% males in the offspring. We simulated a selection for breeding values estimated by the classical BLUP (Henderson, 1975). The replacement ratio was set at 20% for dams and 60% for sires (Brito et al., 2011), and mating was random among selected individuals. The replacement ratio is the proportion of animals to be culled and replaced in each generation.

From the study population (last 10 generations of the simulation), genotype and QTL data were obtained for the 9th generation (treated as a reference population) and the 10th generation (the validation population). The marker effects were first estimated in the reference generation. The σ^2_{gamete} values for all individuals were estimated for both the reference and validation populations using the marker effects estimated in the reference generation. For comparison, true gametic variance was also calculated using the QTL effects and their genotype data from the simulation.

To reduce computational load, a small genome, with 4 autosomal chromosomes of 50 cM each, was simulated. The mutation rate was fixed at 2.5×10^{-5} in the historical population. The number of crossovers was sampled from a Poisson distribution. A total of 200,000 markers and different sets of QTL were simulated to be randomly distributed along the genome. After the ge-

nome was simulated, a panel with 10% of the polymorphic markers was sampled every 0.5 cM and another panel with 20% of the markers was sampled every 0.5 cM. The first panel was chosen to mimic a high-density SNP panel and the second for sequence data. A detailed description of the parameters is reported in Table 1.

Six traits were simulated with heritabilities of 0.1, 0.3, and 0.5 and 20 QTL (i.e., 0.1 QTL per cM) or 200 QTL (i.e., 1 QTL per cM), respectively. We used 2 QTL densities similar to those used by Meuwissen et al.

(2001). The QTL effects were generated based on a gamma distribution with parameter $\beta = 0.4$ (Hayes and Goddard, 2001). The phenotypic variance was assumed to be 1 for all traits. Four replicates were used for each trait. In addition, 10 future generations were simulated where the individuals were selected either by the true breeding value (T_PTA) or by true RPTA (T_RPTA) to verify and compare the genetic gains obtained using these criteria. To assess the effect of these indices on selection in the future generations, the replacement ra-

Table 1. Summary of simulation parameters

Parameter	Value
Genome parameter	
Genome size	200 cM
Number of chromosomes	4
Number of QTL	20 and 200
Number of markers	10,000 (high-density panel) and 20,000+ QTL (sequence data)
Mutation rate, QTL	2.5×10^{-5}
Mutation rate, marker	2.5×10^{-3}
Marker positions in genome	Evenly spaced
QTL position in genome	Random (uniform distribution)
QTL allele effect	Gamma distribution ($\beta = 0.4$)
Trait parameters	
Number of traits	6
Heritability	0.10, 0.30, 0.50
Phenotypic variance	1
Sex-limited trait	No
Population structure parameters	
Historical generation	
Phase 1	
Number of generations	500
Number of animals	Constant (500 males and 500 females)
Mating	Random
Phase 2	
Number of generations	500
Initial number of animals	1,000
Final number of animals	200 (100 males and 100 females)
Mating	Random
Phase 3	
Number of generations	10
Initial number	200 (100 males and 100 females)
Final number	3,000 (1,500 males and 1,500 females)
Mating	Random
Recent generation	
Number of generations	10
Reference population	9th
Validation population	9th and 10th
Number of offspring per dam	5
Founders	1,000 (200 males 800 females)
Mating	Random
Selection	BLUP
Cutting	BLUP
Replacement	20% females and 60% males
Overlapping generation	Yes
Generation 9–10 (predictability)	Correlation $\left(\sigma^2_{\text{gamete}},\ \sigma^2_{\text{gamete\_estimated}}\right)$
Future generation	
Number of generations	10
Criterion of selection[1]	T_PTA = TRUE/2 or T_RPTA (TRUE/2) + σ^2_{gamete}
Number of offspring per dam	5 or 10
Replacement	100% females and 100% males
Better criterion	Genetic gain per generation

[1]T_PTA = true PTA; T_RPTA = true relative PTA; σ^2_{gamete} = variance of gametic diversity.

tio was maintained at 100% and the number of offspring per dam was 5 (corresponding to a selection intensity of 0.996 for females and 1.76 for males) or 10 (corresponding selection intensities of 1.4 for females and 2.06 for males). As the predicted σ^2_{gamete} is a latent variance, its realized value depends on the number of progeny of an individual. Any inference using this variance should be regarded as a bet (probability of an event considering the number of attempts). Therefore, the selection intensity applied to RPTA (i_f) may need to be adjusted accordingly, and 3 values of i_f (0.5, 0.8, and 1) were tested in this study.

Genomic Analysis

Because σ^2_{gamete} depends on the marker effects in genomic models, we used a model that assumed homogeneity of variance of marker effects, **GBLUP** (SNP-BLUP), and another model that allowed heterogeneity of marker effects with differential shrinkage through the improved Bayesian LASSO (**BLASSO**; Legarra et al., 2011). The analyses were performed using the GS3 v.3 software (Legarra et al., 2015). The model included the population mean, marker effects, and residual. Only markers with minor allele frequency (**MAF**) >0.05 were considered. For estimation of additive and residual variances, the simulated true values were used as initial values to reduce computational complexity, followed by 20,000 iterations with the burn-in of 2,000 initial chains.

Application of Gametic Variance to Real Data

The data used were part of the 2017 US genomic evaluations from the Council on Dairy Cattle Breeding (CDCB, Bowie, MD), consisting of 1,364,278 Holstein and 164,278 Jersey cattle from the national dairy cattle database. Five dairy traits based on up to 5 lactations were analyzed: milk (**MY**), fat (**FY**) and protein (**PY**) yields, and fat (**F%**) and protein (**P%**) percentages. The genotype data were generated from different SNP arrays with the number of SNP ranging from 7K to 50K. All individuals were imputed to a common panel of 60,671 SNP and their linkage phase were determined by FindHap version 3 (VanRaden et al., 2011). The σ^2_{gamete} was calculated using Equation [2] with estimated SNP effects $(\hat{\alpha}_1)$. The marker effects were derived from the PTA obtained from the genomic evaluation. Sex-specific recombination rates between SNP markers in Holstein and Jersey cattle were directly used in this study (Ma et al., 2015; Shen et al., 2018). Thus, a modification to the off-diagonal elements of the **M** matrix in Equation [2] was applied to incorporate recombination rate

$$\mathbf{M}_{jk} = al_{jk}\left(-\frac{rate_{jk}}{2} + 0.25\right),$$

when the recombination rate is <0.5; and $\mathbf{M}_{jk} = 0$ when the rate ≥0.5.

RESULTS AND DISCUSSION

Estimation of Gametic Variance with Genomic Models

The variance of progeny breeding values has been investigated in previous studies (Cole and VanRaden, 2011; Segelke et al., 2014; Bonk et al., 2016). Here, we sought to use simulation to evaluate the predictability of gametic variance as a parameter for selection. To evaluate the predictability, a comparison with classical simulation studies with genomic prediction was adopted. The variance of gametic diversity $\left(\sigma^2_{\text{gamete}}\right)$ was calculated considering both dependence and independence between loci, using all QTL and QTL with MAF ≥5%, and utilizing high-density SNP and sequence data with marker effects obtained from genomic models. The Pearson correlation between the true and estimated σ^2_{gamete} ranged from medium to high (Table 2), similar to those studies on breeding values (Meuwissen et al., 2001; Daetwyler et al., 2010; Clark et al., 2011). In general, the correlation increased when the heritability (h^2) of traits increased, whereas the same relation was not apparent when the number of QTL was large. Differently, for the GEBV prediction, the increase in accuracy has been reported with increased h^2 and for scenarios with a small number of QTL, particularly when these were estimated by differential shrinkage models (Daetwyler et al., 2010; Clark et al., 2011).

We observed higher correlations between the true and predicted σ^2_{gamete} using BLASSO compared with GBLUP in all scenarios (Table 2). These results were partly due to the small genome and large QTL effects simulated. Although GBLUP can have a similar or slightly better performance for prediction of GEBV than differential shrinkage models for scenarios with a large number of QTL (Daetwyler et al., 2010), the accuracy of the estimated marker effects, mainly for QTL regions, is greater from differential shrinkage models (Meuwissen et al., 2001; Shepherd et al., 2010; Legarra et al., 2011). For estimating σ^2_{gamete}, the marker effect has a greater impact than for GEBV prediction because

σ^2_{gamete} uses the squared marker effects as well as the dependency of the chromosome segments. Therefore, this observation can also be attributed to the greater accuracy of the marker effects estimated by BLASSO and to the high dependency of the chromosome segments simulated.

The effect on prediction was inferred by a linear regression between true and estimated σ^2_{gamete}. For the intercept of regression (a), GBLUP had a lower scale effect (close to zero) than BLASSO but the difference was not large (Table 3). A low scale effect is important for σ^2_{gamete} prediction because it affects the precision of

the limit values of the confidence interval for future progeny PTA. The scale effect may be affected by the prediction models and by factors inherent to the trait. However, GBLUP had a larger prediction bias, worse values of mean squared error, and regression coefficients (b) more different from 1 (Table 3). For genomic prediction, lower bias has also been reported for differential shrinkage models (Meuwissen et al., 2001). Our result can be attributed to the accuracy of the estimated marker effects and to the small number of independent chromosome segments simulated.

For a trait with $h^2 = 0.10$ and 20 QTL (Table 2), the correlations between σ^2_{gamete} obtained with all QTL and

Table 2. Pearson correlations between variance of gametic diversity for all QTL $\left(\sigma^2_g\right)$, for QTL with minor allele frequency (MAF) ≥ 0.05 $\left(\sigma^2_{gm}\right)$ and disregarding the covariances for all QTL $\left(\sigma^2_d\right)$ and QTL with MAF ≥ 0.05 $\left(\sigma^2_{dm}\right)$, and their estimations using a high-density marker panel and sequence data by genomic BLUP (bp) and Bayesian LASSO (ls), considering $\left(\sigma^2_{gbp} \text{ and } \sigma^2_{gls}\right)$ and disregarding $\left(\sigma^2_{dbp} \text{ and } \sigma^2_{dls}\right)$ the dependency between the markers[1]

Trait		Gametic variance	High-density SNP				Sequence data				QTL data			
h^2	QTL (no.)		σ^2_{gbp}	σ^2_{gls}	σ^2_{dbp}	σ^2_{dls}	σ^2_{gbp}	σ^2_{gls}	σ^2_{dbp}	σ^2_{dls}	σ^2_g	σ^2_{gm}	σ^2_d	σ^2_{dm}
0.1	20	σ^2_g	0.49	**0.56**	0.17	0.39	0.46	**0.57**	0.20	0.40	—	0.75	0.96	0.69
		σ^2_{gm}	0.53	**0.74**	0.21	0.54	0.48	**0.75**	0.25	0.55	0.75	—	0.66	0.93
		σ^2_d	0.45	**0.53**	0.15	0.43	0.43	**0.53**	0.19	0.43	0.96	0.66	—	0.71
		σ^2_{dm}	0.50	**0.74**	0.18	0.61	0.45	**0.73**	0.24	0.61	0.69	0.93	0.71	—
	200	σ^2_g	0.50	**0.60**	0.29	0.37	0.46	**0.61**	0.29	0.40	—	0.96	0.50	0.48
		σ^2_{gm}	0.48	**0.61**	0.29	0.39	0.45	**0.63**	0.30	0.41	0.96	—	0.46	0.49
		σ^2_d	0.29	0.28	**0.51**	0.30	0.28	0.27	**0.48**	0.31	0.50	0.46	—	0.97
		σ^2_{dm}	0.27	0.29	**0.52**	0.32	0.26	0.29	**0.49**	0.33	0.48	0.49	0.97	—
0.3	20	σ^2_g	0.64	**0.83**	0.28	0.66	0.59	**0.83**	0.07	0.65	—	0.94	0.95	0.90
		σ^2_{gm}	0.65	**0.87**	0.28	0.68	0.59	**0.87**	0.07	0.68	0.94	—	0.90	0.95
		σ^2_d	0.60	**0.81**	0.30	0.69	0.54	**0.81**	0.07	0.68	0.95	0.90	—	0.95
		σ^2_{dm}	0.60	**0.85**	0.30	0.71	0.55	**0.85**	0.07	0.70	0.90	0.95	0.95	—
	200	σ^2_g	0.63	**0.77**	0.25	0.49	0.59	**0.77**	0.29	0.48	—	0.95	0.55	0.52
		σ^2_{gm}	0.62	**0.78**	0.25	0.51	0.57	**0.78**	0.29	0.49	0.95	—	0.53	0.53
		σ^2_d	0.42	0.48	0.52	**0.63**	0.40	0.49	0.54	**0.62**	0.55	0.53	—	0.99
		σ^2_{dm}	0.41	0.48	0.52	**0.63**	0.39	0.48	0.54	**0.63**	0.52	0.53	0.99	—
0.5	20	σ^2_g	0.54	**0.67**	0.28	0.50	0.48	**0.66**	0.18	0.49	—	0.86	0.94	0.81
		σ^2_{gm}	0.51	**0.67**	0.26	0.47	0.44	**0.65**	0.16	0.46	0.86	—	0.79	0.93
		σ^2_d	0.52	**0.64**	0.30	0.53	0.46	**0.63**	0.19	0.51	0.94	0.79	—	0.85
		σ^2_{dm}	0.49	**0.64**	0.28	0.51	0.43	**0.63**	0.18	0.49	0.81	0.93	0.85	—
	200	σ^2_g	0.79	**0.85**	0.37	0.51	0.76	**0.84**	0.29	0.51	—	0.95	0.65	0.62
		σ^2_{gm}	0.77	**0.86**	0.37	0.55	0.74	**0.86**	0.30	0.55	0.95	—	0.64	0.65
		σ^2_d	0.53	0.61	0.49	**0.83**	0.52	0.61	0.38	**0.83**	0.65	0.64	—	0.98
		σ^2_{dm}	0.51	0.61	0.49	**0.85**	0.50	0.61	0.37	**0.85**	0.62	0.65	0.98	—

[1]Values in bold represent the best estimates.

Table 3. Mean squared prediction (MSE), intercept (a), and coefficient (b) of the linear regression between the variance of gametic diversity for QTL and its estimates using a high-density SNP panel and sequence data by genomic models[1]

Trait			High-density SNP			Sequence data		
h^2	QTL (no.)	Model[2]	MSE	a	b	MSE	a	b
0.1	20	GBLUP	0.0014	**−0.0010**	0.27	0.0022	**−0.00033**	0.20
		BLASSO	**8e-05**	0.0027	**1.20**	**8e-05**	0.00185	**1.26**
	200	GBLUP	0.0010	**0.0058**	0.23	0.0016	**0.00637**	0.18
		BLASSO	**0.0001**	0.0074	**1.01**	**0.0001**	0.00681	**1.03**
0.3	20	GBLUP	0.0017	**−0.00697**	0.43	0.0028	**−0.00625**	0.35
		BLASSO	**0.0002**	0.00282	**1.46**	**0.0002**	0.00247	**1.41**
	200	GBLUP	0.0021	0.00979	0.40	0.0035	0.01123	0.33
		BLASSO	**0.0004**	**0.00945**	**1.14**	**0.0004**	**0.00950**	**1.13**
0.5	20	GBLUP	0.0019	**−0.00294**	0.26	0.0030	**−0.002039**	0.19
		BLASSO	**0.0001**	0.00188	**1.41**	**0.0001**	0.001866	**1.37**
	200	GBLUP	0.0022	**0.00560**	0.62	0.0033	**0.006547**	0.56
		BLASSO	**0.0008**	0.00851	**1.10**	**0.0007**	0.008799	**1.09**

[1]Values in bold represent the least-biased estimates.
[2]GBLUP = genomic BLUP; BLASSO = Bayesian LASSO.

with QTL of MAF ≥5% were of moderate to high magnitude, lower than that of other traits (high magnitude), resulting in lower correlations with the σ^2_{gamete} estimated by genomic models. Although this result may be due to allele frequency fluctuations in historical population, it also implies that QTL with low MAF are important for obtaining accurate estimates of σ^2_{gamete}. This variance does not depend directly on population allele frequencies but on the individual's heterozygote status. Although MAF filtering (≥5%) can be used to improve the prediction of GEBV (Uemoto et al., 2015), markers with low MAF may have greater linkage disequilibrium with QTL with low MAF, providing better predictions of gametic variance.

To facilitate rapid calculation of σ^2_{gamete}, we tested some scenarios without considering the covariance (dependence) between markers. However, the correlation between true and estimated σ^2_{gamete} was always lower compared with the full model, with the difference ranging from moderate to high when the estimates were obtained from QTL, and from low to high when obtained from the marker effects (Table 2). However, the high correlation observed for one of the scenarios ($h^2 = 0.30$ and QTL = 20) can be attributed to the random distribution of QTL in the genome. Therefore, covariance between markers should always be included for calculating σ^2_{gamete}, and thus, be preferred over the traditional Mendelian sampling variance (Appendix A1). This result is consistent with Bonk et al. (2016), who recommended the use of haplotype and direct recombination data (Cole and VanRaden, 2011).

No difference in correlation between true and estimated σ^2_{gamete} from BLASSO was observed between the high-density SNP and sequence data scenarios (Table 2), indicating that SNP panels with moderate densities are sufficient for estimating σ^2_{gamete}. However, a decrease in correlation was observed for estimates obtained with GBLUP when the sequence data panel was used, regardless of the number of simulated QTL. For GEBV prediction, Clark et al. (2011) observed a small difference in performance using differential shrinkage with sequence data compared with medium-density SNP panels. Pérez-Enciso et al. (2017) also reported a modest increase in accuracy using differential shrinkage model on sequence data. Therefore, sequence data are unlikely to offset SNP panels for predicting GEBV when the number of loci is large and the prior given to each SNP is uniform. Although no improvement in accuracy for σ^2_{gamete} was observed with an increased number of markers, the difference in performance between the 2 types of methods was in line with the literature on GEBV studies. This fact, together with the increase in overestimation due to an increased number of markers (Table 3), confirms the preference of shrinkage models for estimating σ^2_{gamete} in our simulation of small genome and relative large QTL effects.

The correlation between true and predicted CRV was lower than that of σ^2_{gamete} (Supplemental Table S1; https://doi.org/10.3168/jds.2018-15971). There was no unanimous model, but GBLUP showed better prediction performance for many scenarios, whereas BLASSO had better results when ignoring the covariance between markers in scenarios with moderate heritability and a small number of QTL. Generally, the prediction with high-density markers showed a higher accuracy than that with sequence data. The CRV is a relative parameter that indicates how variable the GEBV of an individual is when transmitted to its gametes. The magnitude of the correlation showed that this parame-

ter can be predicted, although the decreased accuracy with an increased number of markers indicated some difficulty for prediction in these cases.

These results may also be explained by a partition of CRV (Appendix A3). Similar results were observed for σ^2_{gamete} and CRV in the 10th generation using the marker effects estimated from the 9th generation. It means that predictions for these parameters can follow the same design in genomic selection programs to calculate GEBV, and σ^2_{gamete} can be estimated using the training data from previous generations (Habier et al., 2007).

Application of Gametic Variance in Selection Programs

The percentage of additional genetic gain ($\mathbf{\Delta G}$) per generation in selection by using RPTA compared with PTA ($\Delta G_{\text{RPTA-PTA}}/\Delta G_{\text{PTA}}$), as well as the accumulated gain for a period of 10 generations, was used to assess the suitability of the new selection index (Figure 1 and Supplemental Figure S1; https://doi.org/10.3168/jds.2018-15971). The accumulated genetic gains obtained with RPTA were higher than those obtained with PTA when the number of QTL increased. No significant increase was observed for a small number of QTL (20). However, in scenarios with more QTL, the genetic gain was close to expected (Appendix A2), with ΔG ranging from 5 to 16% in 10 generations, indicating an advantage of RPTA for traits with large numbers of QTL. These results were in agreement with those reported by Daetwyler et al. (2015) using a genomic optimal haploid value (**OHV**) for selection. In addition, we noted that RPTA tended to increase the frequency of the best alleles in the population more quickly than PTA, which resulted in a permanent effect over generations (Supplemental Figure S1; https://doi.org/10.3168/jds.2018-15971). This trend was also observed by Daetwyler et al. (2015) using OHV, which verified the increase of genetic gain and a smaller reduction of genetic diversity compared with genomic selection. According to those authors, the selection of the individuals with the highest breeding values can lead to the loss of rare favorable alleles in the population, but individuals that carry these alleles may be more favorable in the long term, even though their GEBV can be below the truncation point. The importance of the selection for favorable minor alleles was also reported by Sun and VanRaden (2014) using a weighted genomic selection (**WGS**) by the favorable MAF.

The percentages of additional genetic gain per generation using RPTA were generally positive across scenarios (Figure 1 and Supplemental Figure S1; https://doi.org/10.3168/jds.2018-15971). Some were negative in scenarios with a few QTL but they were all positive and relatively large for scenarios with more QTL, between 3 and 8%. These results highlight the advantage of using RPTA compared with conventional PTA. Although for all scenarios we observed that the first generation under the RPTA selection obtained less genetic gain than PTA, a large increase was obtained in subsequent generations. Daetwyler et al. (2015) reported similar trends in the first few generations of selection using OHV. Thus, the selection by RPTA initially provided an increase of variability, with a subsequent reduction by selection of the best alleles in the following generations.

The effect of h^2 was not evident across the scenarios, and in general, the best RPTA performance was observed for scenarios with moderate heritability. Lehermeier et al. (2017) compared UC and OHV in 2 scenarios with low and high h^2 and observed greater gain in the latter scenario. However, as the true values (PTA and RPTA) were used for selection in this study, heritability would not affect the estimation but the magnitude of genetic variances.

When the number of progeny per dam increased, the genetic gain also increased for both PTA and RPTA, but the RPTA achieved a faster, and therefore greater, genetic gain (Figure 1 and Supplemental Figure S1; https://doi.org/10.3168/jds.2018-15971). These results agreed with those reported by Daetwyler et al. (2015), who observed a rapid increase in genetic gain using OHV when the number of progeny increased from 10 to 1,000. Lehermeier et al. (2017) also observed greater gains for higher selection intensities for UC and OHV. For the scenarios with a few QTL, there was no large gain by using RPTA when increasing selection intensity (Figure 1 and Supplemental Figure S1). For scenarios with many QTL, we observed a greater gain using RPTA in early generations than in scenarios with lower selection intensity, as predicted by the expected gain (Appendix A2). For traits with many QTL, higher values of i_f had better performance than those with lower intensities, although the trend was not observed for scenarios with a few QTL. It suggests that for scenarios with many QTL, the increase in selection intensity allows the use of higher values for i_f.

One of the advantages of using RPTA is the choice of i_f for weighting σ^2_{gamete} in the index (Figure 1 and Supplemental Figure S1). Initially, integral values (without adjustment) for i_f were tested but gains were much smaller than those with traditional selection. The component $i_f \times \sigma^2_{\text{gamete}}$ from RPTA is stochastic, and high values for i_f can be risky when the number of progeny per individual is small. Lehermeier et al. (2017) used integral value of i_f to obtain UC; however, they simu-
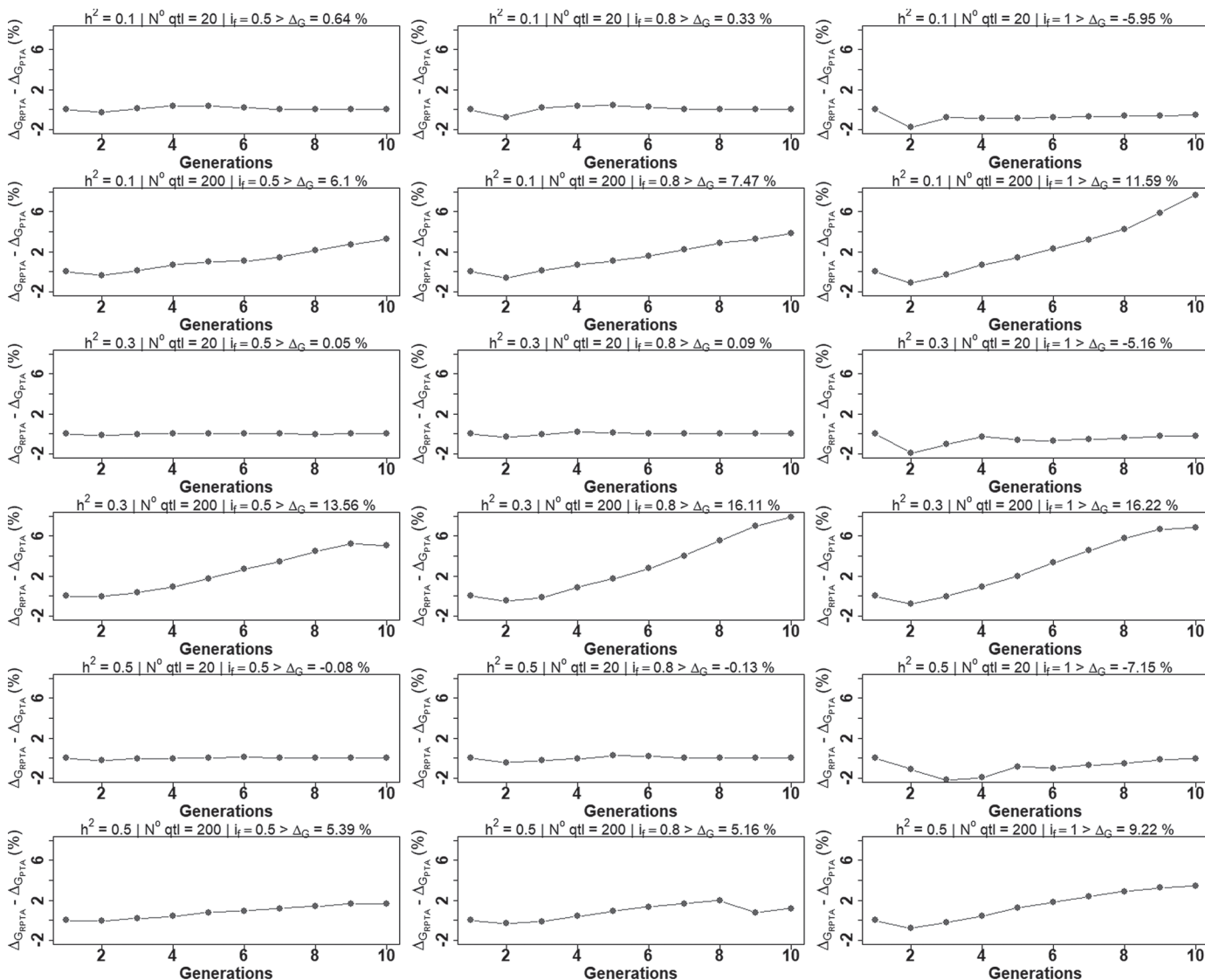
Figure 1. Difference in percentage of genetic gain per generation between true PTA (ΔG_{PTA}) and relative PTA (ΔG_{RPTA}) with different adjusted future selection intensity values ($i_f = 0.5$, 0.8, and 1) and heritabilities ($h^2 = 0.1$, 0.3, and 0.5) for 10 generations and 20 QTL with real section intensity around 0.996 for females and 1.755 for males (5 offspring per dam).

lated plant crosses with large numbers of progeny (100) to realize the predicted variance in the offspring. The increase in i_f proved to be unsuitable for scenarios with few QTL, where lowest values should be prioritized for this type of trait. However, for scenarios with many QTL, large i_f values are desired, especially when selection intensity is high. Thus, the risk related to the number of progeny should be considered and standardized equally for all individuals, preferably by increasing the minimum number of offspring. Lehermeier et al. (2017) obtained the selection intensity value in UC from the proportion of individuals selected within plant homogeneous crosses, which is equivalent to using dam

selection intensity in animal breeding. Empirically, to find the optimal value for i_f, we can adjust the real value of the future selection intensity of dams (have the least number of offspring) by $\left(1 - \overline{PV}\right)$, so the adjusted intensity value is $i_f^* = \left(1 - \overline{PV}\right) \times i_f$, where the individual percentage of variation (PV) is obtained from equation [A3.1] in Appendix A3. Given the number of progeny per dam (n) and the average CRV of the population, we have $\overline{PV} = \dfrac{\left(Z_{1-\alpha/2}\right) \times \overline{CRV}}{\sqrt{n}}$, where Z is the corresponding percentile value from a standard normal distribution.

In this section, we verified the feasibility of using σ^2_{gamete} in selection programs to explore the whole additive genetic potential of individuals. The proposed index (RPTA) is easy to obtain and apply. As the variance of GEBV of future progeny is $\sigma^2_{\text{gameteSire}} + \sigma^2_{\text{gameteDam}}$, the UC for a given mating can be easily obtained as $\text{RPTA}_{\text{sire}} + \text{RPTA}_{\text{dam}}$. With greater genetic gains and better preservation of genetic diversity (Supplemental Figures S2 and S3; https://doi.org/10.3168/jds.2018-15971), the RPTA can accelerate genetic selection compared with other indices that also preserve diversi-

ty, such as the OHV, WGS, optimal population value, and genotype building (Goiffon et al., 2017). This statement is supported by the literature, because first Goiffon et al. (2017) showed a better performance of OHV than cited indices, and then Lehermeier et al. (2017) introduced the UC (similar to RPTA, but used for mating) and demonstrated its superiority to OHP within the crosses when different selection intensities are applied. Thus, RPTA can be a better option to maximize genetic gain per generation and the profitability of a breeding program. In contrast, although
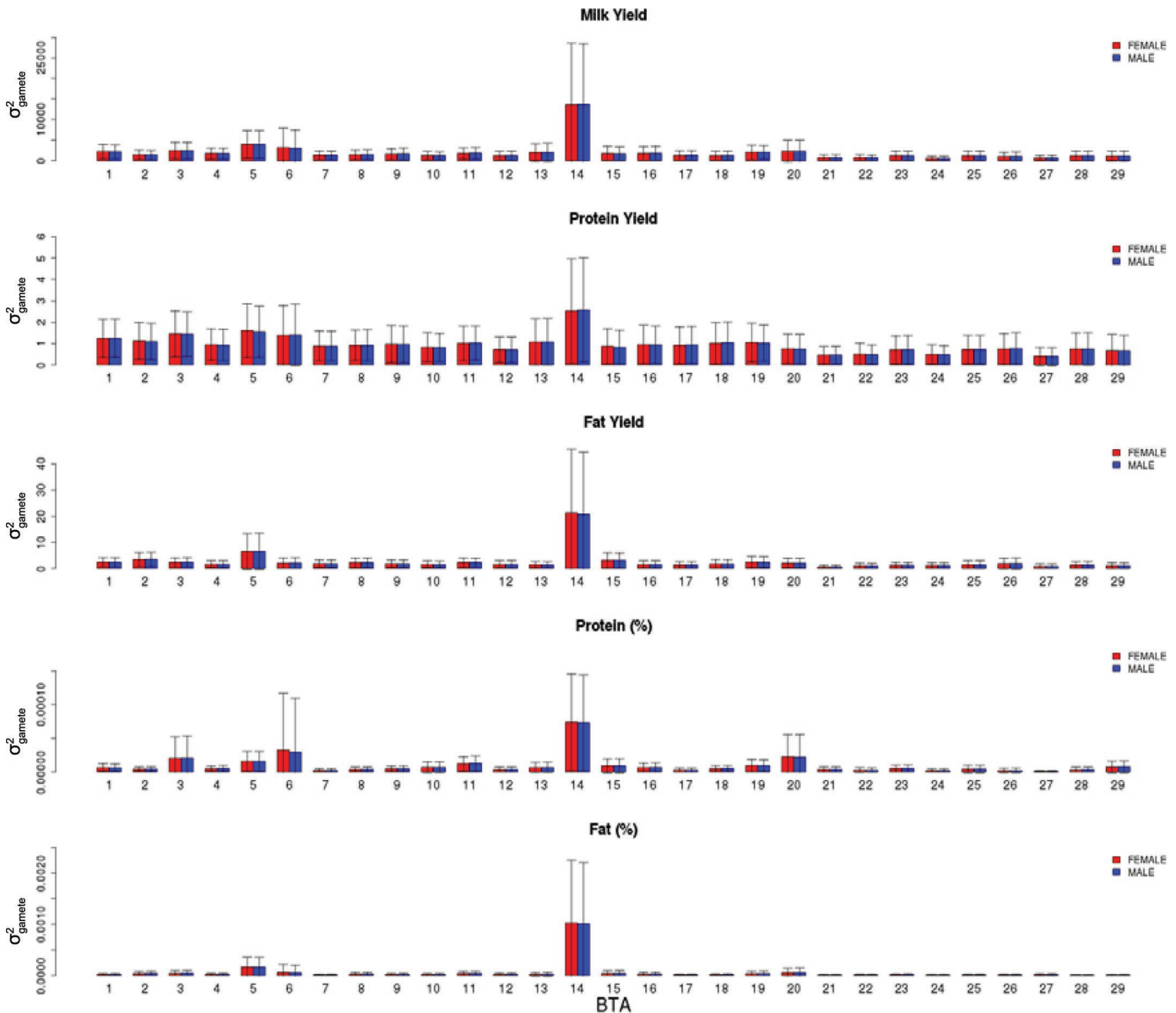

Figure 2. Distribution of variance of gametic diversity $\left(\sigma^2_{\text{gamete}}\right)$ for milk, fat, and protein yields and fat and protein percentages by chromosome and sex in US Holstein cattle. Bars indicate averages and whiskers represent standard deviations.

RPTA represents an optimal equilibrium between the expected value (PTA) and the variability, the weighted value for σ^2_{gamete} $\left(i^*_f\right)$ can be modified to preserve diversity and accelerate genetic progress. Another point to consider is that although RPTA and WGS have a similar purpose, the expected component of RPTA (PTA) can still be adjusted for a greater emphasis in the selection of favorable minor alleles if desired, as suggested by Sun and VanRaden (2014). Besides, although the genetic gain obtained using RPTA with random mating has been impressive for traits with many QTL, greater gains can be obtained with this criterion using a sophisticated mating design (Allaire, 1980; Sonesson and Meuwissen, 2000; Lehermeier et al., 2017).

Estimation and Application of Gametic Variance in Real Data

The suitability of applying σ^2_{gamete} in livestock breeding programs was evaluated using 2 dairy populations, Holstein and Jersey, for 5 milk production traits. Because chromosomes are independent genome segments, σ^2_{gamete} was calculated separately for each chromosome (Figure 2 and Supplemental Figure S4; https://doi.org/10.3168/jds.2018-15971). The average, standard deviation, and amplitude of the estimated σ^2_{gamete} were largest on BTA14 for all production traits. This was expected because BTA14 contains the largest QTL for milk production, *DGAT1* (Grisart et al., 2002; Bouwman et al., 2011, 2018). However, the 5 traits had different distributions of σ^2_{gamete} among chromosomes. The σ^2_{gamete} for PY was more evenly distributed, but for other traits, especially for F%, σ^2_{gamete} showed a skewed distribution with major mass concentrated on BTA14. This is due to the greater effect of *DGAT1* on milk fat than on milk protein (Thaller et al., 2003). Similar results were observed in both Holstein and Jersey cattle. Although the recombination rate is different between males and females in cattle (Shen et al., 2018), little difference was observed for σ^2_{gamete} between the 2 sexes (Figure 2 and Supplemental Figure S4).

The distribution of σ^2_{gamete} varied in the 2 cattle populations (Figure 3 and Supplemental Figure S5; https://doi.org/10.3168/jds.2018-15971). The results showed a distribution close to the typical Gaussian curve for PY, but non-Gaussian curves for other traits. For F%, the distribution had 2 peaks. Similar results were observed for Holstein and Jersey, but the effect of BTA14 was more pronounced in Jersey, possibly because this breed has a higher milk fat percentage, as well as different composition of fatty acid content in milk (White et al., 2001). Segelke et al. (2014), studying genetic variation
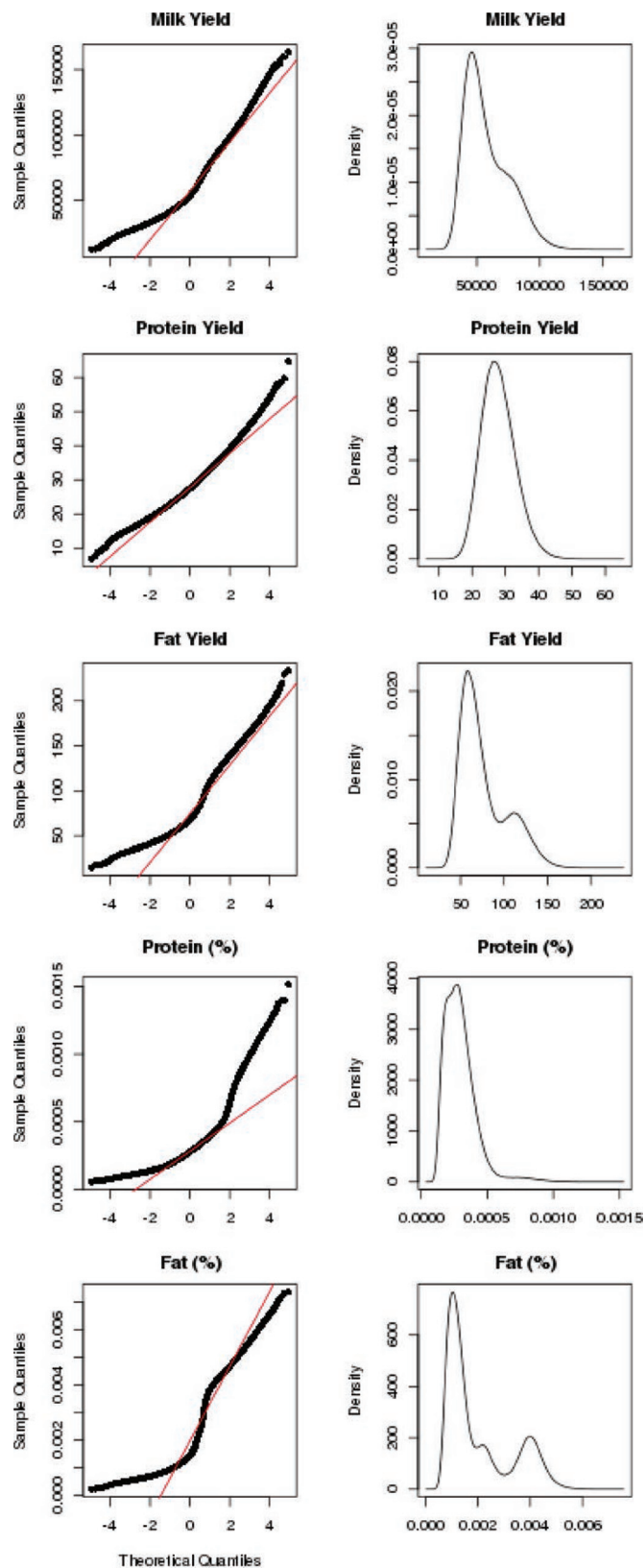

Figure 3. Normal quantile-quantile (Q-Q) and kernel density plots of variance of gametic diversity distribution for milk, fat, and protein yields as well as fat and protein percentages in US Holstein cattle.

in Holstein, observed a similar distribution for PY and FY.

The predictability of the offspring variance of dairy traits for bulls was assessed using a Pearson correlation between the variance of progeny breeding values and the σ^2_{gamete} of bulls estimated by genomic models (Table 4). This correlation increased with an increased number of offspring. The moderate to high correlation indicated the feasibility of predictions. These results were consistent with Segelke et al. (2014), who reported the same trend with slightly higher values for the correlation with standard deviations of gamete breeding value (**SDGBV**) in Holstein. Although the variance of progeny GEBV also contains a dam effect, our results using only sires validated the σ^2_{gamete} estimated by genomic models as a predictor of the variance of GEBV for future offspring. In general, traits with larger coefficient of variation, such as PY, MY, and FY, had a lower correlation, whereas traits with lower coefficient of variation, F%, and P%, exhibited larger correlations. The difference in variability also explains the second trend, where traits with a biased distribution of σ^2_{gamete} per chromosome had a better prediction than those showing even distributions. Segelke et al. (2014) also reported a larger correlation for FY than PY using SDGBV.

The correlations of σ^2_{gamete} between traits were all positive, ranging from moderate to high magnitude, although most of the estimated correlations between production and content percentage traits were negative (Table 5). In general, many large correlations of σ^2_{gamete} were observed with MY, indicating that selection using gametic variation of MY can result in the preservation of variability in other production traits as well. The magnitude of correlation for FY and PY was similar to those studies using other measures of progeny variation by Segelke et al. (2014) and by Bonk et al. (2016).

Correlations of σ^2_{gamete} with inbreeding coefficients, given as the diagonal of the Wright (1931) and genomic (VanRaden, 2008) relationship matrices, were negative but close to zero across dairy traits, with the highest correlation observed for PY and the lowest for F% (Table 5). The results observed for PY and FY were similar to those reported by Segelke et al. (2014) using SDGBV. The negative correlation was expected because the greater the inbreeding, the lower the heterozygosity of the loci and consequently the lower the σ^2_{gamete}. However, because σ^2_{gamete} depends on the heterozygosity of those genomic regions with QTL effects, the low correlations were not unexpected, because inbreeding coefficients are global indicators of whole-genome heterozygosity.

CONCLUSIONS

This study verified the feasibility of estimating and applying variance of gametic diversity in livestock breeding programs. The σ^2_{gamete} can be accurately ob-

Table 4. Pearson correlations (r) of σ^2_{gamete} for milk, protein, and fat yields (*MY*, *PY*, and *FY*, respectively) and protein and fat percentages (*P%* and *F%*, respectively) with variances of progeny breeding values for different minimum numbers of offspring per sire[1]

Breed and minimum no. of offspring	Sires (no.)	r_{MY}	r_{PY}	r_{FY}	$r_{P\%}$	$r_{F\%}$
Jersey						
10	1,109	0.24	0.16	0.20	0.30	0.58
50	451	0.40	0.33	0.46	0.50	0.75
100	311	0.53	0.34	0.47	0.60	0.85
200	183	0.64	0.31	0.49	0.77	0.95
300	128	0.68	0.40	0.55	0.86	0.96
400	97	0.66	0.43	0.61	0.90	0.97
500	77	0.66	0.51	0.62	0.90	0.97
600	66	0.69	0.54	0.66	0.92	0.97
Holstein						
10	6,797	0.29	0.16	0.32	0.39	0.57
50	2,753	0.55	0.24	0.60	0.69	0.85
100	1,887	0.66	0.27	0.67	0.78	0.91
200	1,241	0.71	0.23	0.70	0.83	0.93
300	903	0.75	0.26	0.74	0.85	0.94
400	706	0.77	0.29	0.77	0.87	0.95
500	569	0.78	0.30	0.78	0.89	0.96
600	478	0.78	0.30	0.77	0.89	0.95

[1]σ^2_{gamete} = variance of gametic diversity.

Table 5. Pearson correlations between σ^2_{gamete} for different traits (above diagonal) and breeding values (below diagonal) and inbreeding coefficients (right side columns)[1]

Breed and trait	MY	FY	PY	F%	P%	F_G	F_P
Jersey						
MY	—	0.77	0.61	0.92	0.84	−0.07	−0.03
FY	0.47	—	0.55	0.83	0.73	−0.08	−0.02
PY	0.86	0.74	—	0.41	0.37	−0.15	−0.08
F%	−0.67	0.34	−0.30	—	0.87	−0.03	−0.00
P%	−0.59	0.24	−0.10	0.83	—	−0.07	−0.01
F_G	0.04	−0.03	0.03	−0.07	−0.03	—	0.48
F_P	0.19	0.13	0.20	−0.10	−0.06	0.48	—
Holstein						
MY	—	0.73	0.62	0.86	0.67	−0.12	−0.03
FY	0.44	—	0.43	0.90	0.46	−0.11	−0.04
PY	0.83	0.68	—	0.42	0.28	−0.19	−0.07
F%	−0.50	0.56	−0.12	—	0.59	−0.06	−0.01
P%	−0.40	0.33	0.18	0.69	—	−0.10	−0.03
F_G	0.24	0.30	0.31	0.07	0.10	—	0.51
F_P	0.24	0.31	0.33	0.08	0.13	0.51	—

[1]MY = milk yield; PY = protein yield; FY = fat yield; F% = fat percentage; P% = protein percentage; F_G and F_P = genomic and pedigree inbreeding coefficients, respectively; σ^2_{gamete} = variance of gametic diversity.

tained from genomic models. To improve the estimation of σ^2_{gamete}, covariance between markers needs to be considered. σ^2_{gamete} can be especially useful for traits with many QTL using a newly developed selection index, RPTA. Additionally, the confidence level of this index can be adjusted by the number of future progeny, making it suitable for dairy cattle breeding. For Holstein and Jersey cattle, *DGAT1* had a large effect on the prediction of σ^2_{gamete} across all production traits. Inbreeding coefficients had a small impact on gametic variability of dairy traits, with a greater effect on traits less affected by *DGAT1*. Collectively, σ^2_{gamete} can be easily obtained and applied to existing genomic selection programs to improve genetic progress and control genetic diversity.

REFERENCES

Allaire, F. R. 1980. Mate selection by selection index theory. Theor. Appl. Genet. 57:267–272.

Bijma, P., Y. C. J. Wientjes, and M. P. L. Calus 2018. Increasing genetic gain by selecting for higher Mendelian sampling variance. Page 11.47 in Proc. World Congr. Genet. Appl. Livest. Prod., Auckland, New Zealand. http://www.wcgalp.org/proceedings/2018/increasing-genetic-gain-selecting-higher-mendelian-sampling-variance.

Bonk, S., M. Reichelt, F. Teuscher, D. Segelke, and N. Reinsch. 2016. Mendelian sampling covariability of marker effects and genetic values. Genet. Sel. Evol. 48:36.

Bouwman, A. C., H. Bovenhuis, M. H. P. W. Visker, and J. A. M. Van Arendonk. 2011. Genome-wide association of milk fatty acids in Dutch dairy cattle. BMC Genet. 12:43. https://doi.org/10.1186/1471-2156-12-43.

Bouwman, A. C., H. D. Daetwyler, A. J. Chamberlain, C. H. Ponce, M. Sargolzaei, F. S. Schenkel, G. Sahana, A. Govignon-Gion, S. Boitard, M. Dolezal, H. Pausch, R. F. Brøndum, P. J. Bowman, B. Thomsen, B. Guldbrandtsen, M. S. Lund, B. Servin, D. J. Garrick, J. Reecy, J. Vilkki, A. Bagnato, M. Wang, J. L. Hoff, R. D. Schnabel, J. F. Taylor, A. A. E. Vinkhuyzen, F. Panitz, C. Bendixen, L. E. Holm, B. Gredler, C. Hozé, M. Boussaha, M. P. Sanchez, D. Rocha, A. Capitan, T. Tribout, A. Barbat, P. Croiseau, C. Drögemüller, V. Jagannathan, C. V. Jagt, J. J. Crowley, A. Bieber, D. C. Purfield, D. P. Berry, R. Emmerling, K. U. Götz, M. Frischknecht, I. Russ, J. Sölkner, C. P. Van Tassell, R. Fries, P. Stothard, R. F. Veerkamp, D. Boichard, M. E. Goddard, and B. J. Hayes. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat. Genet. 50:362–367.

Brito, F. V., J. Braccini Neto, M. Sargolzaei, J. A. Cobuci, and F. S. Schenkel. 2011. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. BMC Genet. 12:80.

Clark, S. A., J. M. Hickey, and J. H. J. Van der Werf. 2011. Different models of genetic variation and their effect on genomic evaluation. Genet. Sel. Evol. 43:18. https://doi.org/10.1186/1297-9686-43-18.

Cole, J. B., and P. M. VanRaden. 2011. Use of haplotypes to estimate Mendelian sampling effects and selection limits. J. Anim. Breed. Genet. 128:446–455.

Daetwyler, H. D., M. Hayden, G. Spangenberg, and B. Hayes. 2015. Selection on optimal haploid value increases genetic gain and preserves more genetic diversity relative to genomic selection. Genetics 200:1341–1348.

Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams. 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics 185:1021–1031.

Dempfle, L. 1990. Problems in the use of the relationship matrix in animal breeding. Pages 454–473 in Advances in Statistical Methods for Genetic Improvement of Livestock. Vol 1. D. Gianola and K. Hammond, ed. Springer, New York, NY.

García-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggans, F. J. Ruiz-López, and C. P. Van Tassell. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc. Natl. Acad. Sci. USA 113:E3995–E4004. https://doi.org/10.1073/pnas.1519061113.

Goiffon, M., A. Kusmec, L. Wang, G. Hu, and P. S. Schnable. 2017. Improving response in genomic selection with a population-based selection strategy: Optimal population value selection. Genetics 206:1675–1682.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 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.

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.

Hayes, B., and M. E. Goddard. 2001. The distribution of effects of genes affecting quantitative traits in livestock. Genet. Sel. Evol. 33:209–229.

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447.

Legarra, A., A. Ricard, and O. Filangio. 2015. GS3 Genomic Selection — Gibbs Sampling — Gauss Seidel (and BayesCπ). Accessed Jan. 20, 2015. http://snp.toulouse.inra.fr/~alegarra/manualgs3_last.pdf.

Legarra, A., C. Robert-Granié, P. Croiseau, F. Guillaume, and S. Fritz. 2011. Improved Lasso for genomic selection. Genet. Res. (Camb.) 93:77–87.

Lehermeier, C., S. Teyssèdre, and C. C. Schön. 2017. Genetic gain increases by applying the usefulness criterion with improved variance prediction in selection of crosses. Genetics 207:1651–1661.

Ma, L., J. R. O'Connell, P. M. VanRaden, B. Shen, A. Padhi, C. Sun, D. M. Bickhart, J. B. Cole, D. J. Null, G. E. Liu, Y. Da, and G. R. Wiggans. 2015. Cattle sex-specific recombination and genetic control from a large pedigree analysis. PLoS Genet. 11:e1005387. https://doi.org/10.1371/journal.pgen.1005387.

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.

Mrode, R. A. 2005. Linear Models for the Prediction of Animal Breeding Values. 2nd ed. CAB International, Wallingford, UK.

Müller, D., P. Schopp, and A. E. Melchinger. 2018. Selection on expected maximum haploid breeding values can increase genetic gain in recurrent genomic selection. G3 (Bethesda) 8:1173–1181. https://doi.org/10.1534/g3.118.200091.

Nejati-Javaremi, A., C. Smith, and J. P. Gibson. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. J. Anim. Sci. 75:1738–1745.

Pérez-Enciso, M., N. Forneris, G. de Los Campos, and A. Legarra. 2017. Evaluating sequence-based genomic prediction with an efficient new simulator. Genetics 205:939–953.

Sargolzaei, M., and F. Schenkel. 2009. QMSim: A large-scale genome simulator for livestock. Bioinformatics 25:680–681.

Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123:218–223.

Segelke, D., F. Reinhardt, Z. Liu, and G. Thaller. 2014. Prediction of expected genetic variation within groups of offspring for innovative mating schemes. Genet. Sel. Evol. 46:42. https://doi.org/10.1186/1297-9686-46-42.

Shen, B., J. Jiang, E. Seroussi, G. E. Liu, and L. Ma. 2018. Characterization of recombination features and the genetic basis in multiple cattle breeds. BMC Genomics 19:304. https://doi.org/10.1186/s12864-018-4705-y.

Shepherd, R. K., T. H. Meuwissen, and J. A. Woolliams. 2010. Genomic selection and complex trait prediction using a fast EM algorithm applied to genome-wide markers. BMC Bioinformatics 11:529 https://doi.org/10.1186/1471-2105-11-529.

Sonesson, A. K., and T. H. E. Meuwissen. 2000. Mating schemes for optimum contribution selection with constrained rate of inbreeding. Genet. Sel. Evol. 32:231–248.

Sun, C., and P. M. VanRaden. 2014. Increasing long-term response by selecting for favorable minor alleles. PLoS One 9:e88510.

Thaller, G., W. Krämer, A. Winter, B. Kaupe, G. Erhardt, and R. Fries. 2003. Effects of DGAT1 variants on milk production traits in German cattle breeds. J. Anim. Sci. 81:1911–1918.

Thomasen, J. R., A. C. Sørensen, G. Su, P. Madsen, M. S. Lund, and B. Guldbrandtsen., 2013. The admixed population structure in Danish Jersey dairy cattle challenges accurate genomic predictions. J. Anim. Sci. 91:3105–3112.

Uemoto, Y., S. Sasaki, T. Kojima, Y. Sugimoto, and T. Watanabe. 2015. Impact of QTL minor allele frequency on genomic evaluation using real genotype data and simulated phenotypes in Japanese Black cattle. BMC Genet. 16:134. https://doi.org/10.1186/s12863-015-0287-8.

Van Belle, G., and D. C. Martin. 1993. Sample size as a function of coefficient of variation and ratio of means. Biometrics 58:612–620.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.

VanRaden, P. M., J. R. O'Connell, G. R. Wiggans, and K. A. Weigel. 2011. Genomic evaluations with many more genotypes. Genet. Sel. Evol. 43:10. https://doi.org/10.1186/1297-9686-43-10.

White, S. L., J. A. Bertrand, M. R. Wade, S. P. Washburn, J. T. Green Jr., and T. C. Jenkins. 2001. Comparison of fatty acid content of milk from Jersey and Holstein cows consuming pasture or a total mixed ration. J. Dairy Sci. 84:2295–2301.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

APPENDIX A1

Variance of Gametic Diversity in Traditional BLUP Models

The additive genetic value of an individual (u) can be expressed as the sum of half the value of sire (S), half of the dam (D), and a Mendelian error (m). The components transmitted by the sire (T_S) and dam (T_D) can be represented as

$$T_S = \left(\frac{u_S}{2}\right) + m_S$$

$$\text{and } T_D = \left(\frac{u_D}{2}\right) + m_D, \qquad [A1.1]$$

where the variance of m, the Mendelian sample variance, $var(m)$, is obtained as in Mrode (2005):

$$var(m) = \frac{(1 - F)\sigma_u^2}{4}. \qquad [A1.2]$$

Here, the $var(m)$ of an EBV is obtained from the parental inbreeding coefficients (F) (Dempfle, 1990). From the genomics point of view, $var(m)$ can be obtained considering locus by locus of the parent. In the traditional BLUP model, the diagonal elements of the relationship matrix are obtained by doubling the expected variance if this diploid individual mated with itself. This variance can be divided into 2 parts for the progeny having the same alleles $\left(a_i = a_j\right)$ for a given locus or different alleles $\left(a_i \neq a_j\right)$. Considering that a homozygous locus has probability $\Pr\left(a_i = a_j\right) = 1$ and a heterozygous locus $\Pr\left(a_i = a_j\right) = 0.5$ and $\Pr\left(a_i \neq a_j\right) = 0.5$, the total variance of a hypothetical mating will be a function of the number of homozygous loci ($NHom$) and the number of heterozygous loci ($NHet$):

$$2 \times var\left(T_S + T_D\right) = \frac{2 \times \left(NHom + NHet\right)}{\sum_i^N 2p_i q_i}\sigma_u^2. \quad [A1.3]$$

Separating by components related to equality $\left(a_i = a_j\right)$ and to the difference $\left(a_i \neq a_j\right)$ among the loci of the future gametes, we have

$$2 \times var\left(T_S + T_D\right) = \frac{2 \times \left[NHom + NHet\,\Pr\left(a_i = a_j\right)\right]}{\sum_i^N 2p_i q_i}\sigma_u^2$$

$$+ \frac{2 \times NHet\,\Pr\left(a_i \neq a_j\right)}{\sum_i^N 2p_i q_i}\sigma_u^2$$

$$2 \times var\left(T_S + T_D\right) = (1 + F)\sigma_u^2 + \left(1 - F\right)\sigma_u^2.$$

Knowing that $\sigma_u^2 = \sum_i^N 2p_i q_i \sigma_a^2$, where σ_a^2 is the variance of the effect of allelic substitution considered homogeneous for all loci in BLUP, $var(m)$ is

$$var(m) = \frac{1 - F}{4}\sigma_u^2 = 0.25\frac{2 \times NHet\,\Pr\left(a_i \neq a_j\right)}{\sum_i^N 2p_i q_i}\sigma_u^2$$

$$var(m) = 0.25 \times NHet\sigma_a^2.$$

When homogeneous variance is assumed for the allelic substitution effects across all loci that are inherited independently:

$$var(m) = \sigma_{\text{gamete}}^2. \qquad [A1.4]$$

APPENDIX A2

Expected Genetic Gain Using the Relative PTA

The $RPTA_i$ (relative PTA) refers to the average of the genetic value relative to the group of gametes that will be selected in the future. Thus, from the key equation of genetic change, the future genetic gain considering the selection of animals can be estimated. We will differentiate between the 2 selection intensities, i_r (recent) and i_f (future), although they are identical in most cases. The relative selection differential (SR) for the next generation is given by $\text{SR} = \text{E}\left[RBV_{i>\text{selected}}\right]$, where the $RBV_i = 2 \times \left(PTA_i + \sigma_i \times i_f\right)$. The variance of $RBV = 2 \times \left(PTA + \sigma_{gamete} \times i_f\right)$ can be obtained as

$$E\left\{\left[RBV - E\left(RBV\right)\right]^2\right\} =$$
$$\sigma_u^2 + 4 \times i_f^2 \times \left[E\left(\sigma_{gamete}^2\right) - E\left(\sigma_{gamete}\right)^2\right]. \quad [A2.1]$$

The relative genetic gain (ΔGR) can then be obtained as

$$GR = r \times i_r \times \sqrt{\sigma_u^2 + 4 \times var\left(\sigma_{gamete}^2\right) \times i_f^2 \times r_g^2},$$

[A2.2]

where σ_u^2 is the additive genetic variance, $var\left(\sigma_{gamete}^2\right)$ is the variance of gametic diversity, and r and r_g^2 are, respectively, accuracies of the genetic evaluation and the prediction of the σ_{gamete}^2.

The increase in rate of genetic gain from using the relative criterion in place of the traditional criterion, disregarding the accuracy of σ_{gamete}^2, will then be

$$\Delta GR - \Delta G = \sqrt{\sigma_u^2 + 4i_f^2 var\left(\sigma_{gamete}^2\right)} - \sqrt{\sigma_u^2}. \text{ [A2.3]}$$

The expected genetic gain can be stratified using different i_f for sex, where the term $4i_f^2 var\left(\sigma_{gamete}^2\right)$ is expanded to explicitly express male (S) and female (D) contributions:

$$i_{fS}^2 var\left(\sigma_{gameteS}^2\right) + i_{fD}^2 var\left(\sigma_{gameteD}^2\right)$$
$$+ \; 2 \times i_{fS} i_{fD} cov\left(\sigma_{gameteS} \sigma_{gameteD}\right).$$

[A2.4]

APPENDIX A3

Sample Size and Coefficient of Relative Variation

Given σ_{gamete}^2, it is useful to find the optimal number of progeny for an individual to realize the expected variability in its offspring. This is similar to the optimal number of daughters needed for a reliable conventional progeny test. It is difficult to obtain such an estimate using a general model, but estimating a percentage variability in real data is possible. Van Belle and Martin (1993) proposed an approach to obtain the number of samples from the coefficient of variation considering the margin of error as a percentage of variation. Be-

cause the gametic variation is a random component proportional to the additive genetic variance, a modification of Van Belle and Martin (1993) by substituting the coefficient of variation for the relative variation (CRV) of the value transmitted to the progeny can be used to estimate the required sample size (n); CRV (equation [4]) considers the value related to the average transmission of additive variance of an individual i to its progeny $\left[\sqrt{E\left[(0.5u_i)^2\right]}\right]$ as $0.5\sqrt{\sum\limits_i^{NHom} 2\alpha_i^2 + \sigma_{gamete}^2}$.

Thus, we can use the CRV to obtain the sample size according to levels of percentage variation admitted between an estimated value from the sample and the expected value (Van Belle and Martin 1993). The percentage of variation in a sample relative to the expected value (PV) in $0.5\sqrt{E\left[u_i^2\right]}$ can be represented as $PV_{0.5u} = \dfrac{0.5T_1 - 0.5T}{0.5\sqrt{E\left[u_i^2\right]}}$, where T_1 represents the change (alternative) in the mean of PTA that may differ from the expected transmitted value (T). The smaller the change desired in PTA, the greater the number of progeny will be required and, consequently, greater variability will be observed among these progenies. The number of progeny (n) to be used from the percentage of variation $\left(PV_{0.5u_i}\right)$ as a margin of error can be represented as

$$n = \frac{\left(Z_{1-\alpha/2}\right)^2 \times \left(CRV_i\right)^2}{\left(PV_{0.5u_i}\right)^2},$$

[A3.1]

where $Z_{1-\alpha/2}$ is the critical value associated with the degree of confidence. For example, at significance level of 95% if we accept a percentage change in the mean of only 10% $\left(PV_{0.5u_i} = 0.1\right)$ of the PTA, we have $n = 400\left(CRV_i\right)^2$, where the homozygous animals will have a lower CRV than the heterozygous animals, requiring a smaller number of offspring for progeny testing.