



A mating advice system in dairy cattle incorporating genomic information

T. R. Carthy,^{1*} J. McCarthy,² and D. P. Berry¹

¹Teagasc, Animal and Grassland Research and Innovation Centre, Fermoy, Co. Cork, Ireland P61 P302

²Irish Cattle Breeding Federation, Bandon, Co. Cork, Ireland P72 X050

ABSTRACT

This study investigated the effects of alternative mating programs that incorporate genomic information on expected progeny herd performance and inbreeding, as well as methods to include un-genotyped animals in such mating programs. A total of 54,535 Holstein-Friesian cattle with imputed high-density genotypes (547,650 SNP after edits) were available. First, to quantify the accuracy of imputing un-genotyped animals (often an issue in populations), a sub-population of 729 genotyped animals had their genotypes masked, and their allele dosages were imputed, using linear regression exploiting information on genotyped relatives. The reference population for imputation included all genotyped animals, excluding the 729 selected animals and their sires, dams, and grandsires, and had either (1) their sires' genotypes, (2) their dams' genotypes (3) both their sires' and their dams' genotypes, or (4) both their sires' and maternal grandsires' genotypes introduced into the reference population. The correlations between true genotypes and the imputed allele dosages ranged from 0.58 (sire only) to 0.68 (both sire and dam). A herd of 100 cows was then simulated (1,000 replicates) from the sub-population of 729 imputed animals. The top 10 bulls from the genotyped population, based on their total genetic merit index (TMI) were selected to be used as sires. Three mating allotment methods were investigated: (1) random mating, (2) sequential mating based on maximizing only the expected TMI of the progeny, and (3) linear programming to maximize a generated index constructed to maximize genetic merit and minimize expected progeny inbreeding as well as intra- and inter-progeny variability in genetic merit. Relationships among candidate parents were calculated using either the pedigree relationship matrix or the genomic relationship matrix; the latter was constructed using either the true genotypes of both parents or the true genotypes of the sire plus the imputed allele dos-

ages of the dam. Using the genomic co-ancestry estimates resulted in lower average herd expected genomic inbreeding levels compared with using the pedigree-based co-ancestry estimates. Additionally, if the dams were not genotyped, using their imputed allele dosages also resulted in lower average herd expected inbreeding levels compared with using the pedigree co-ancestry estimates. The inter-progeny coefficient of variation for selected traits, milk and fertility, estimated breeding values were reduced by 12 to 65% using the linear programming method compared with sequential mating.

Key words: mating program, genomic relationship, imputation

INTRODUCTION

Mating programs provide a support tool for breeders and producers, facilitating identification of the best combination of parental matings to maximize long-term genetic gain in their herds within the constraints of biology (i.e., number of progeny per dam) and the desires of the breeder or producer. Mating programs tend to allocate the matings based on the expected genetic merit of the prospective progeny, avoiding the mating of (closely) related individuals to prevent excessive inbreeding (Weigel and Lin, 2000). Mating programs can control inbreeding at both the population and the individual animal level (Pryce et al., 2012). At the population level, optimal contribution methods are designed to maximize long-term genetic gain while also constraining the rate of inbreeding (Meuwissen, 1997; Sonesson and Meuwissen, 2000). At the individual animal level, mating allotment can be applied to avoid excessive inbreeding among progeny (Kinghorn, 1998). Many producers, especially those with large herds, also seek as homogeneous a herd as possible and, in doing so, strive to have all animals conform to the on-farm infrastructure (e.g., cubicle size); thus, mating programs can also aid in providing uniform management decisions across the entire herd.

One of the most important aspects of any mating program is the control of inbreeding. Inbreeding occurs when 2 individuals transmit to their progeny the same

Received January 11, 2019.

Accepted April 24, 2019.

*Corresponding author: Tara.CCarthy@teagasc.ie

part of the genome inherited from a common ancestor. Inbreeding depression, the unfavorable repercussion of inbreeding, has been associated with compromised fitness in many species, including pigs (Fernández et al., 2002), chickens (Flock et al., 1991), sheep (Barczak et al., 2009), beef cattle (Barczak et al., 2009), and dairy cattle (McParland et al., 2007; Miglior et al., 2008). Inbreeding can occur in mating programs if precautions are not taken to minimize the likelihood of such consanguineous mating. The expected inbreeding of the progeny from a given mating has traditionally been calculated based on the pedigree relationship between the parents (Weigel, 2001); however, the depth of the pedigree available, coupled with parentage errors known to exist (Purfield et al., 2016), affect the efficacy of such an approach. Furthermore, pedigree methods fail to differentiate between individuals with the same pedigree that inherit different parts of the genome and thus have different realized relationships. Genomic information can, however, provide a more accurate estimate of the relationship between parents (Sun et al., 2013) and, in doing so, provide a more accurate estimate of the expected inbreeding of the progeny.

The objective of the present study was to investigate alternative producer-level mating programs, taking into consideration the demands of producers, as well as the incorporation of genomic information into mating decision support tools. Because not all females are generally genotyped yet, the ability to impute an ungenotyped population was quantified, and the effects of incorporating such imputed genotypes into the mating advice tool were evaluated.

MATERIALS AND METHODS

Genotype Data

Genotypes from a total of 54,535 Holstein-Friesian cattle (35,030 males and 19,505 females) were available. All animals were genotyped on one of the following genotype panels: 168 on the Illumina BovineHD BeadChip (Illumina, San Diego, CA; 777,962 SNP); 8,223 on the Illumina BovineSNP50 BeadChip (54,001 SNP); 9,422 on the International Dairy and Beef (IDB) Version 1 (11,103 SNP); 13,147 on the IDB version 2 (40,196 SNP); 11,947 on the IDB version 3 (41,484 SNP); 10,098 on the Illumina BovineLD Genotyping BeadChip (6,909 SNP); and 1,630 on the Illumina Bovine 3k Genotyping BeadChip (2,900 SNP). All panels were imputed to the high density (HD; i.e., 777,962 SNP) using FIMPUTE Version 2.2 (Sargolzaei et al., 2014), with a multi-breed reference population consisting of 5,195 HD genotyped dairy and beef animals. Mendelian inconsistencies were used to validate animal

parentage as well as to discard autosomal SNP that did not adhere to expected Mendelian inheritance patterns. A total of 89,565 SNP with a minor allele frequency <1%, as well as 94,254 SNP that deviated from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-9}$), were discarded. Only autosomal SNP with a known genomic position were retained. A total of 547,650 SNP were available for analysis.

Prediction of Genotypes in Ungenotyped Animals: Predicted Allele Dosage

Allele dosage is the number of copies of a particular allele that an animal has at a given locus (Gengler et al., 2007). Gengler et al. (2007) proposed a gene-dropping method to predict the allele dosage of an individual where no genotype existed for that locus; the allele dosage produced is a linear transformation of the 3 genotype probabilities to a single value per locus per individual, ranging from 0 to 2. The proposed approach exploits the known allele dosage of relatives to predict the unknown allele dosage at that locus, as follows:

$$q_x = \left(\mathbf{1} \quad \mathbf{A}_{xy} \mathbf{A}_y^{-1} \right) \begin{pmatrix} \mu \\ \mathbf{q}_y - \mathbf{1}\mu \end{pmatrix},$$

where q_x is the predicted allele dosage of the ungenotyped individual x , \mathbf{q}_y is a vector of the known allele dosage of genotyped individual y , $\mathbf{1}$ is a vector of ones, \mathbf{A}_{xy} is a matrix representing the additive genetic relationship between the individual x with unknown genotypes and the genotyped relatives y , \mathbf{A}_y^{-1} represents the inverse of the additive genetic relationship matrix among all genotyped individuals, and μ is the mean allele frequency of the total genotyped population at the given locus.

Accuracy of Allele Dosage Prediction

A sub-population of 729 animals (204 males and 525 females) with both parents and maternal grandsire genotyped were extracted from the 54,535 genotyped animals; the sires of the 729 selected animals did not appear as a maternal grandsire for any of the other 729 animals. The entire genotypes of the 729 individuals in the sub-population were masked and their allele dosage per locus imputed using the above algorithm implemented in the SNP1101 software (Sargolzaei, 2014).

The reference population for imputation initially included all genotyped animals, excluding the 729 selected animals as well as their sires, dams, and grandsires. Different combinations of relatives were then introduced into the reference population, and the

imputation accuracy was computed for the different scenarios investigated. Four different scenarios were investigated: (1) only the sire genotypes was introduced into the reference population; (2) only the dam genotypes were introduced into the reference population; (3) both sire and dam genotypes were introduced into the reference population; and (4) both sire and maternal grandsire genotypes were introduced into the reference population. The accuracy of imputation per animal was measured as the correlation between the true genotypes (0, 1, or 2) minus the mean allele dosage for the respective SNP with the imputed allele dosage (on a scale of 0 to 2) minus the mean allele dosage for the respective SNP (Mulder et al., 2012) across all SNP in the animal. The imputation accuracy per SNP was measured as the correlation between the true genotypes (0, 1, or 2) and the imputed allele dosage (on a scale of 0 to 2) per SNP across all individuals.

Pedigree and Genomic Relationships

Pedigree Relationships. The pedigrees of all 54,535 genotyped animals were traced back at least 5 generation, where possible; the pedigree consisted of 231,798 animals. The coefficients of pedigree relationships among animals were estimated using a modified algorithm described by Meuwissen and Luo (1992) in the RelaX2 software (Strandén and Vuori, 2006).

Genomic Relationships. The correlation between expected inbreeding in progeny and observed inbreeding is stronger when base population allele frequencies are used compared with the use of the allele frequencies of the current population (VanRaden et al., 2011). The base population in the present study was defined as the oldest animals in the pedigree with unknown parents, and the associated allele frequencies were calculated using the gene-dropping method proposed by Gengler et al. (2007) implemented in the SNP1101 software (Sargolzaei, 2014); this method can be adapted, if required, to account for different genetic groups, as may exist in a multi-breed population.

The genomic relationship (a) between 2 animals, i and j , was extracted from a genomic relationship matrix calculated using the approach outlined by Yang et al. (2010):

$$a_{ij} = \frac{1}{N} \sum_m \frac{(x_{mi} - 2p_m)(x_{mj} - 2p_m)}{2p_m(1 - p_m)},$$

where N is the number of SNP, x_{mi} is the allele dosage of the reference allele for the m th SNP in the i th individual, x_{mj} is the allele dosage of the reference allele

for the m th SNP in the j th individual, and p_m is the base frequency of the reference allele for the m th SNP.

The genomic relationship (a) of an animal with itself (i.e., the genomic inbreeding) was calculated using the approach outlined by Yang et al. (2010), which provides an unbiased estimate of inbreeding:

$$a_{ii} = \frac{1}{N} \sum_m \frac{x_{mi}^2 - (1 + 2p_m)x_{mi} + 2p_m^2}{2p_m(1 - p_m)} + 1,$$

where N is the number of SNP, x_{mi} is the allele dosage of the reference allele for the m th SNP in the i th individual, and p_m is the base frequency of the reference allele for the m th SNP.

Genomic relationships were derived using the entire imputed HD genotype data, with 547,650 autosomal SNP per individual. However, to quantify the effect of using the more common 50k genotype panels, only every 10th SNP was retained from the imputed HD genotype, and genomic relationships were calculated using just this reduced SNP set.

Adjustment of Genome-Based Estimated Relationships. To combine genomic and pedigree-based relationships, the genome-based estimates of relationships were rescaled to reflect pedigree-based estimates. To achieve this, the genome-based and pedigree-based relationship coefficients (including the relationship of an animal with itself) of 529 Holstein-Friesian AI bulls, with at least 5 full generations of pedigree, were extracted. The genomic relationship matrix was regressed on the pedigree-based relationship matrix by solving the following equations:

$$G_{ij}^* = \alpha + \beta G_{ij},$$

where G_{ij} is the genomic relationship between animals i and j , G_{ij}^* is the adjusted genomic relationship between animals i and j , and α and β were obtained from

$$\begin{bmatrix} 1 & \overline{G_{ij}} \\ 1 & \overline{G_{ii}} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \overline{A_{ii}} \\ \overline{A_{ij}} \end{bmatrix},$$

where $\overline{G_{ij}}$ is the mean of the off-diagonals of the genomic relationship matrix, $\overline{G_{ii}}$ is the mean of the diagonals of the genomic relationship matrix, $\overline{A_{ij}}$ is the mean of the off-diagonals of the pedigree relationship matrix, and $\overline{A_{ii}}$ is the mean of the off-diagonals of the pedigree relationship matrix. Using this approach, the mean of the diagonal and the off-diagonal of the genomic relationship matrix was regressed toward the mean of the

respective pedigree-based estimates, with an equal emphasis on both the diagonal and the off-diagonal elements (P. Van Raden, USDA Animal Genomics Improvement Laboratory, Beltsville, MD; personal communication).

Mating Algorithm

Generated Index. An index was generated that was optimized using linear programming techniques, to (1) maximize the herd's mean EBV for a total merit index (**TMI**) while minimizing the variability among progeny in the estimated TMI, (2) penalize matings with a greater expected inbreeding in the resulting progeny, (3) maximize the mean expected EBV of 2 chosen traits of interest while minimizing the expected inter-progeny variability in these EBV (this can easily be expanded to more than 2 traits), and (4) minimize the differential in the expected progeny EBV of both traits (i.e., achieve a progeny balanced for each trait). The index value per mating between parents i and j was defined as follows:

$$I_{ij} = \frac{TMI_{ij} - \gamma f_{ij}}{TMI_{SD_{pop}}} + \sum_{n=1}^2 \left(\frac{Trait_{n_{ij}}}{Trait_{n_{SD_{pop}}}} - \frac{|Trait_{n_{avg}} - Trait_{n_{ij}}|}{Trait_{n_{SD_{herd}}}} \right) - \frac{|Trait_{1_{ij}} - Trait_{2_{ij}}|}{var_{SD_{pop}}}$$

where I_{ij} is the expected index value of progeny from mating parents i and j , TMI_{ij} is the estimated TMI of the potential progeny (i.e., simply the average parental TMI), f_{ij} is the expected progeny inbreeding from that mating, γ is the effect of inbreeding on TMI, $Trait_{n_{ij}}$ is the EBV of the potential progeny for the trait n for that mating, $Trait_{n_{avg}}$ is the average EBV of all potential progeny in the herd for the trait n , $Trait_{1_{ij}}$ and $Trait_{2_{ij}}$ are the EBV of any 2 traits that require the variability between them to be minimized, $TMI_{SD_{pop}}$ is the standard deviation of the TMI in the total population of AI sires, $Trait_{n_{SD_{pop}}}$ is the standard deviation of the EBV for trait n in the total population of AI sires, $Trait_{n_{SD_{herd}}}$ is the standard deviation of the EBV for trait n of all potential progeny in the herd, and $var_{SD_{pop}}$ is the standard deviation of the variance between the 2 traits of interest in the total population of AI sires.

For the purposes of this study, the TMI used was the Irish national breeding index, the economic breeding index (**EBI**). The EBI is composed of 7 sub-indices,

2 of which are milk and fertility and which constitute the majority of the emphasis in the index (Roche et al., 2017). The milk and fertility indices were used to represent the 2 traits of interest in the current study. The effect of inbreeding (γ) was a monetary penalty calculated based on the inbreeding depression effects reported by McParland et al. (2007) for Irish dairy cows and the current economic values associated with the respective traits used in the EBI.

Herd Simulation. A herd of 100 dams was randomly selected without replacement from the 525 females that were part of the sub-population used for imputation; this step was repeated 1,000 times to generate 1,000 different herds. All dams had a genotype available, as well as imputed allele dosages derived from the 4 different imputation scenarios previously described. The top 10 AI sires, ranked by total genetic merit index (**TMI**), that had a genotype available were selected as the team of bulls to be used in each replicate. The EBV for the TMI, milk sub-index, and fertility sub-index for all candidate sires and dams were obtained from the national genetic evaluations.

Mating Allotment Methods. Three methods of allocating a candidate sire to a candidate dam within a herd (i.e., replicate) were investigated. All 3 mating strategies had similar constraints applied. In all cases, all dams could only be mated once. It was assumed that the potential sires were already preselected, and there was a limit to the number of services per sire for first service. Equal usage per sire was assumed; thus, each sire was allowed to be mated to a maximum of 10 dams. A maximum level of expected progeny inbreeding of 6.25% was enforced; any mating expected to exceed this limit was not considered further. Expected progeny inbreeding was calculated as either (1) the rescaled genomic estimates of co-ancestry between both parents, calculated using the true genotypes from both parents, (2) the rescaled genomic estimates of co-ancestry between both parents, calculated using the true genotypes from the sire but using imputed allele dosages from the dam, or (3) the pedigree estimates of co-ancestry between both parents.

The first method of mating allotment was based on random mating in the herd. The second method was based on sequentially selecting the candidate sire and dam that produced the progeny with the greatest expected TMI within the pre-defined constraints (i.e., a dam can only be mated once, and a sire can be mated a maximum of 10 times). The third method used linear programming, implemented using the R package LP-solve. The linear programming method maximizes the sum of the generated index value described previously for the whole herd:

Table 1. Mean within-animal and within-SNP imputation accuracy¹ of imputed allele dosage on 729 animals when both parents (sire and dam), sire and maternal grandsire only, sire only, or dam only were included in the reference population

Item	Sire and dam	Sire and maternal grandsire	Sire only	Dam only
Within-animal accuracy	0.68	0.58	0.58	0.67
Within-SNP accuracy	0.67	0.57	0.57	0.66

¹Imputation accuracy was calculated as the correlation between true genotypes and imputed allele dosages after correction for the mean allele dosage.

$$f_{\max}(I) = \sum_{i=1}^a \sum_{j=1}^b I_{ij} x_{ij},$$

where I_{ij} is the index value of the progeny of dam i and sire j , a is the number of dams in the herd, b is the number of sires to be used in the herd, and x_{ij} is equal to 1 or 0 based on the following constraints:

$$\sum_{j=1}^b x_{ij} \leq 1; \sum_{i=1}^a x_{ij} \leq 10.$$

The results averaged across the 1,000 replicates are presented. The herd standard deviations, minimum and maximum EBV for the TMI, and the 2 traits (milk and fertility indices) were used to compare the performance of each mating allotment method. The average expected progeny genomic inbreeding coefficients in the herd, regardless of which of the 3 methods was used to estimate expected progeny inbreeding (genomic, pedigree, or imputed genomic), was used to evaluate the effects of the mating allotment methods.

RESULTS

Prediction of Allele Dosage

Allele Dosage Prediction Accuracy per Animal. The mean within-animal accuracy of imputation for the 729 un-genotyped animals was a function of which ancestors had true genotypes introduced into

the imputation reference population (Table 1). The mean within-animal imputation accuracy ranged from 0.58 when only the sire's genotype was in the reference population to 0.68 when both parental genotypes were in the reference population.

Allele Dosage Prediction Accuracy per SNP. The accuracy of imputation in the un-genotyped animals was poorer for SNP with low minor allele frequency (MAF; <0.1) than for higher-MAF SNP (>0.4; see Table 2); more variability in imputation accuracy was also evident for low-MAF SNP compared with higher-MAF SNP. The mean accuracy of imputation per SNP increased with the proportion of ancestors genotyped (Table 1).

Measures of Inbreeding and Co-Ancestry

Panel Density. The correlations between the genomic inbreeding and co-ancestry coefficients calculated using the full HD genotypes and the subset of 50k genotypes were 0.997 and 0.996, respectively. The linear regression coefficient of the genomic inbreeding estimates calculated with the HD genotypes on the genomic inbreeding estimates calculated using the 50k genotypes was 0.996; similarly, the linear regression coefficient of the co-ancestry estimates calculated with the HD genotypes on those calculated with 50k genotypes was also 0.997.

Inbreeding Coefficients. The correlation between the pedigree inbreeding coefficients and genomic inbreeding coefficients calculated on the 54,535 genotyped animals was 0.23 (Figure 1). The regression coefficient

Table 2. Mean imputation accuracy¹ per SNP categorized by minor allele frequency (MAF) group when both parents (sire and dam), sire and maternal grandsire only, sire only, or dam only were included in the reference population

MAF	Sire and dam	Sire and maternal grandsire	Sire only	Dam only
0.0–0.1	0.65	0.55	0.55	0.64
0.1–0.2	0.67	0.57	0.57	0.66
0.2–0.3	0.67	0.58	0.58	0.67
0.3–0.4	0.68	0.58	0.58	0.67
0.4–0.5	0.68	0.58	0.58	0.67

¹Imputation accuracy was calculated as the correlation between true genotypes and imputed allele dosages.

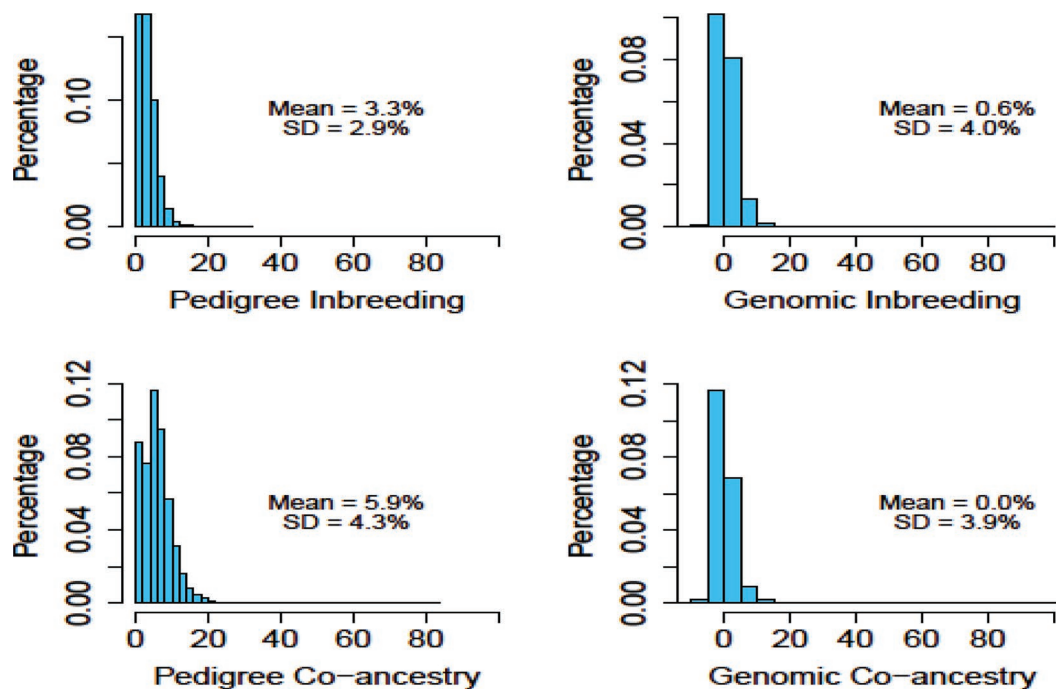


Figure 1. Distribution of both inbreeding (top) and co-ancestry (bottom) estimates, using either pedigree or genomic information on 54,535 genotyped animals. SD = standard deviation.

of the genomic inbreeding coefficients on pedigree inbreeding coefficients was 0.37, with a standard error (SE) of 0.01, and the intercept was 0.63 (SE = 0.01). The correlation between the genomic inbreeding coefficients calculated when true genotypes were available and those calculated from imputed allele dosage (assuming the animal was un-genotyped) on the 729 evaluated animals ranged from 0.19 when only the dam’s genotype was included in the imputation reference population to 0.51 when both the sire’s and the dam’s genotypes were included in the imputation reference population (Table 3). The linear regression coefficient of individual animal inbreeding calculated from the true genotypes on those calculated from imputed allele dosage ranged from 0.10

(SE = 0.02) with an intercept of 0.95 (SE = 0.01) when only the dam’s genotype was available to 0.37 (SE = 0.02) with an intercept of 0.81 (SE = 0.01) when both the sire and the dam had a genotype available.

Co-Ancestry Coefficients. The correlation between the pedigree and the genomic co-ancestry estimates for the 54,535 genotyped animals was 0.57. The linear regression coefficient of the genomic co-ancestry estimates on the pedigree co-ancestry estimates was 0.71 (SE = 0.00002) with an intercept of 0.05 (SE = 0.000001). Using the imputed allele dosage from the 729 animals in the sub-population and the true genotype of the remaining population, the correlation of the 729 imputed animals and the remaining population

Table 3. Correlation and regression coefficient (standard error in parentheses) of pedigree-based measures and genome-based measures using imputed allele dosages of inbreeding (729 animals) and relationships between animals (729 imputed animals with 53,625 genotyped animals) on genomic-based measures using true genotypes when both parents (sire and dam), sire and maternal grandsire, sire only, or dam only were included in the reference population

Item	Pedigree	Imputed allele dosage relationships			
		Sire and dam	Sire and maternal grandsire	Sire only	Dam only
Inbreeding					
Correlation	0.39	0.51	0.35	0.33	0.19
Regression coefficient	0.45 (0.04)	0.37 (0.02)	0.20 (0.02)	0.18 (0.02)	0.10 (0.01)
Co-ancestry					
Correlation	0.64	0.94	0.92	0.92	0.93
Regression coefficient	0.49 (0.0001)	1.01 (0.0001)	1.02 (0.0001)	1.02 (0.0001)	1.00 (0.0001)

genomic co-ancestry coefficients ranged from 0.92 when only the sire's genotype was in the population to 0.94 when both parental genotypes were in the population (Table 3). The linear regression coefficients of the imputed allele dosage co-ancestry coefficients on genomic co-ancestry coefficients were all close to 1 (1.00 to 1.01; SE = 0.0001) with the corresponding intercepts near zero (Table 3).

Rescaling Genomic Measures to Pedigree Measures. Using the small cohort of 529 AI bulls that had a deep pedigree, the intercept (α) and the regression coefficient (β) obtained from regressing the genomic relationships on pedigree using the described method were 0.07 and 0.96, respectively. After rescaling the genomic relationship coefficients for all 54,535 genotyped animals, the mean genomic inbreeding (3.6%) and mean genomic co-ancestry (7.0%) coefficients were more similar to the pedigree inbreeding (3.3%) and co-ancestry (5.9%) estimates. The regression coefficient of the pedigree relationship coefficient on the rescaled genomic relationship coefficient remained similar to the regression coefficient of the pedigree relationship coefficient on the unadjusted genomic relationship coefficient.

Mating Advice Systems

Mating Algorithms Performance. Similar trends in expected average progeny performance between each of the mating allotment methods were observed regardless of which approach was used to estimate co-ancestry between candidate parents. Therefore, only the results from each mating allotment method that applied genomic co-ancestry between parents are discussed in detail (Table 4). The mean expected EBI of the progeny was the same regardless of which mating allotment method was used (Table 4). Mating selection based on the linear programming approach resulted in a smaller within-replicate standard deviation for TMI, milk, and fertility EBV compared with both random selecting and sequential mating (Table 4). The inter-progeny coefficient of variation in the herd TMI was reduced by 12% and 27% when using linear programming compared with random and sequential mating, respectively. Greater reductions were achieved in the inter-progeny coefficient of variation for both the milk EBV (61% and 65% compared with random and sequential mating, respectively) and the fertility EBV (41% and 57% compared with random and sequential mating, respectively) when using linear programming. The reduction in the coefficient of variation was primarily driven by the minimum expected progeny EBV in the herd for both the milk and the fertility sub-indices increasing to a greater degree than the maximum expected progeny

EBV reduced (Table 4). The average within-replicate herd genomic inbreeding level was reduced by 18% using the linear programming approach compared with the other 2 approaches investigated.

Herd Inbreeding. Based on the linear programming mating allotment approach, substituting the genomic co-ancestry estimates of the parents with the traditionally used pedigree co-ancestry estimates resulted in a higher average herd genomic inbreeding level (Table 5). Using imputed allele dosages for the candidate dam along with the true genotypes of the sire to calculate genomic co-ancestry also resulted in a lower average herd genomic inbreeding level compared with using pedigree-based co-ancestry (Table 5). Applying the maximum threshold of 6.25% expected progeny inbreeding successfully limited the maximum herd inbreeding level, regardless of how expected progeny inbreeding was estimated. Similar results for herd inbreeding levels were observed when mating decisions were based on the random mating and sequential mating approaches (results not shown).

DISCUSSION

Controlling the expected inbreeding in progeny is one of the main rationales for implementing a mating advice program. Genetic evaluations and downstream mating programs tend to favor genetically superior sires that originate from the same or similar family lines and, through artificial insemination, these sires are then widely used in the commercial population (Weigel, 2001). The use of a small pool of sires can increase the risk of inbreeding; however, the use of mating programs that take into account the relationships between the candidate sires and dams in a herd has been shown to reduce progeny inbreeding (Weigel, 2001).

Inbreeding arises due to the inheritance of identical parts of the genome from a common ancestor (Wright, 1922). Both the pedigree-based and the genome-based methods of estimating relationships presented in this study attempt to estimate the proportion of the genome that is identical by descent (**IBD**). Pedigree-based methods trace expected co-ancestry via common ancestors and therefore are limited by the depth of the pedigree (i.e., an animal with no pedigree is assumed to be non-inbred and unrelated to the rest of the population). Genome-based methods, on the other hand, estimate the realized proportion of the genome that is identical by examining the genotypes of all animals. However, genome-based methods cannot distinguish between the parts of the genome that are identical due to inheritance from a common ancestor (i.e., **IDB**) and those that are identical in state but not due to inheritance from a common ancestor. Since the true

Table 4. The mean, standard deviation (SD), minimum, and maximum total merit index (TMI)₁, milk EBV, fertility EBV, expected genomic inbreeding, and expected pedigree inbreeding of progeny for different breeding system simulations¹

Metric	Breeding system ²	Mean	SD	Minimum	Maximum
TMI	Random	214.26 (2.79)	35.18 (2.42)	100.15 (19.75)	278.72 (5.51)
	EBV	214.26 (2.79)	39.97 (2.30)	95.59 (18.60)	289.94 (4.06)
	Index	214.26 (2.79)	31.45 (2.40)	107.91 (18.97)	272.95 (3.92)
Milk	Random	66.35 (1.24)	17.20 (1.01)	20.90 (5.55)	104.01 (4.671)
	EBV	66.35 (1.24)	17.70 (1.09)	17.67 (3.52)	101.84 (4.03)
	Index	66.35 (1.24)	10.68 (0.99)	34.68 (4.35)	96.08 (5.62)
Fertility	Random	99.06 (2.06)	27.98 (1.82)	13.135 (15.66)	155.36 (5.88)
	EBV	99.06 (2.06)	31.03 (1.81)	5.06 (16.66)	156.92 (5.547)
	Index	99.06 (2.06)	19.78 (1.89)	32.74 (14.62)	144.54 (4.88)
Genomic inbreeding	Random	3.45 (0.09)	1.07 (0.08)	1.15 (0.28)	6.05 (0.18)
	EBV	3.45 (0.10)	1.14 (0.06)	1.13 (0.13)	6.12 (0.13)
	Index	2.82 (0.08)	0.89 (0.05)	0.98 (0.17)	5.28 (0.41)
Pedigree inbreeding	Random	3.39 (0.11)	1.46 (0.09)	0.15 (0.05)	6.92 (0.44)
	EBV	3.36 (0.11)	1.52 (0.09)	0.12 (0.04)	7.18 (0.36)
	Index	3.14 (0.11)	1.33 (0.07)	0.14 (0.05)	6.13 (0.20)

¹Data averaged across 1,000 replicates of a 100-cow herd. Standard deviation of the replicates given in parentheses.

²Breeding system used: Random = random mating; EBV = sequential mating based on EBV for total genetic merit; Index = linear program based on maximizing the generated index.

proportion of the genome that is IBD is unknown in most populations, no one measure of inbreeding and co-ancestry is completely accurate. Using simulated data for which the true proportion of IBD was known, Kardos et al. (2015) reported that genome-based measures of inbreeding were a better estimate of the true IBD than pedigree-based measures were.

Due to the very strong correlations (>0.99) and the regression coefficients of near unity between the genomic inbreeding and co-ancestry estimated using either the full HD genotypes or the reduced SNP panel, very similar results would be expected if the reduced SNP panel was used. Therefore, if imputation is not feasible, using a lower-density panel would have little

Table 5. Mean, standard deviation (SD), minimum, and maximum expected progeny genomic inbreeding (%) using the linear program-based mating program, with expected progeny inbreeding calculated using either genomic, pedigree, or imputed genomic information with both the sire and dam, sire and maternal grandsire, sire only, or dam only included in the reference population¹

Item	Genomic	Pedigree	Imputed			
			Sire and dam	Sire and maternal grandsire	Sire only	Dam only
Mean	2.82 (0.08)	3.14 (0.09)	2.93 (0.08)	2.97 (0.14)	2.98 (0.08)	2.95 (0.08)
SD	0.89 (0.05)	1.02 (0.06)	0.96 (0.06)	0.68 (0.09)	0.98 (0.06)	0.97 (0.06)
Minimum	0.98 (0.17)	1.08 (0.22)	0.98 (0.16)	1.12 (0.27)	0.99 (0.18)	1.00 (0.17)
Maximum	5.28 (0.41)	5.84 (0.33)	5.57 (0.44)	5.53 (0.68)	5.83 (0.59)	5.68 (0.51)

¹Data averaged across 1,000 replicates of a 100-cow herd. Standard deviation of the replicates given in parentheses.

effect on the overall results presented herein. Keller et al. (2011), using simulated data, reported weak correlations between genomic inbreeding coefficients and pedigree inbreeding coefficients, ranging from 0.16 to 0.22, similar to those found in the present study. Weak correlations between the 2 measures of inbreeding have also been reported in other dairy cattle populations (0.29 by Pryce et al., 2012; 0.13 to 0.44 by Marras et al., 2015). Reports of correlations between the pedigree and genomic co-ancestry coefficients are scarce, but stronger correlations (0.67 to 0.88) than those presented here have been reported in dairy cattle (VanRaden et al., 2011; Pryce et al., 2012). The genomic relationship matrix was used in the current study as a measure of both the inbreeding and the co-ancestry estimates. However, multiple studies have reported stronger correlations between runs of homozygosity (ROH) and pedigree-based inbreeding coefficient (Pryce et al., 2012; Ferenčaković et al., 2013; Zhang et al., 2015) than both those previously reported and those presented in this study using the genomic relationship matrix. Additionally, a stronger correlation between a ROH method of co-ancestry and pedigree-based co-ancestry estimates (Pryce et al., 2012) has also been reported using the genomic relationship matrix.

One limitation to adopting genomic measures of relationships in mating programs is that the females in a population are generally not all genotyped. The imputation strategy in the present study attempted to overcome this shortcoming by imputing the entire genotypes of individuals with no genomic information through exploitation of the information of genotyped relatives. Based on selection index theory, the expected imputation accuracy of allele dosage of un-genotyped animals is 0.71 when both parents have a genotype, 0.61 when either the sire and maternal grandsire or the dam and paternal grandsire have a genotype, and 0.50 when only the sire or the dam have a genotype (Bouwman et al., 2014). The accuracy of imputation reported in the present study was comparable with the theoretical imputation accuracy, although the accuracy of imputation reported when only one parent had a genotype was slightly greater (0.58 to 0.69) than expected (0.50), which is consistent with what has been reported previously (Bouwman et al., 2014). The imputed allele dosages in the current study explained between 33 and 46% of the variation in the true genotypes. Considering that pedigree-based co-ancestry coefficients account for only 33% of the variation in the genomic co-ancestry coefficients, the variation in the true genotypes explained by imputed allele dosages is somewhat expected. Interestingly, the accuracy of imputation was, on average, 9 percentage units greater

when only the dam's genotype (as opposed to the sire's genotype) was in the reference population compared with when only the sire's genotype was in the reference population. This could be partly attributable to the greater number of direct progeny (and other relatives) of the sire compared with the dam in the imputation reference population. Sires, through use of AI, are more likely to have a greater number of progeny than are dams; in the current study, the average number of genotyped progeny in the reference population for the sires was 48, whereas the average number of genotyped progeny per dam was 1.7. Therefore, a greater number of paternal half-siblings were available to improve the imputation accuracy when the sire was not genotyped compared with the number of maternal half-siblings when the dam was not genotyped.

Using the imputed allele dosages of the 729 animals to estimate the genomic relationships with the remaining genotyped population provides a good representation of the true genomic relationships between these animals. The correlations and the regression coefficients of the genomic relationship estimated using true genotypes of both animals and the genomic relationship estimated using imputed allele dosage from one of the animals were better than the corresponding pedigree-based estimates. This suggests that, where possible, using the imputed allele dosages may be a better indicator of the actual relationships among animals than simply using pedigree information, especially if the dam's genotype is in the imputation reference population.

Mating Advice Systems

When selection is based on genomic breeding values, it is more effective to control inbreeding using genome-based measures rather than pedigree-based measures (Sonesson et al., 2010). Because the upper threshold on the level of expected inbreeding in the progeny was imposed on all mating allotment methods, the average herd inbreeding level remained relatively low, regardless of how the co-ancestry among parents was estimated. However, use of the rescaled genomic relationship matrix to estimate co-ancestry resulted in lower average herd genomic inbreeding levels compared with using pedigree relationships. The addition of a negative value associated with expected progeny inbreeding in the linear programming index resulted in an 18% reduction in the average level of inbreeding in the herd compared with both the random and the sequential mating methods. Of the 2 strategies investigated to estimate genomic relationships when females in the herd had no genotype (i.e., pedigree relationships or genomic relationships calculated with the imputed allele dosages of

the dam), using the imputed allele dosages was a better approach to reduce expected progeny genomic inbreeding than using pedigree co-ancestry estimates.

We assumed in the present study that the sires had been preselected by the producer and that cows in the herd could only be mated once. Thus, as expected, regardless of which mating allotment method was applied, the average EBV of the progeny did not differ. Nevertheless, despite no differences in the average EBV of the progeny, substantial differences in inter-progeny variability were evident depending on the method used. The 2 selected traits in the current study, milk and fertility, account for the majority of the Irish TMI because of their substantial effect on herd profitability. Producers often want to limit the range in EBV for a given trait between the genetically elite and inferior progeny (i.e., produce a more homogeneous herd), while still producing the best progeny possible. The proposed index used for the linear programming method attempted to maximize the generated index value for all matings, based not only on expected TMI of the progeny (after adjusting for the unfavorable repercussions of inbreeding) but also on the expected genetic merit of the progeny for the selected traits, after adjusting for how far that progeny's EBV deviated from the mean EBV of all the potential progeny in the herd. By accounting for both the expected genetic merit of the selected trait and the deviation from the mean of the selected trait in the herd, matings that produce progeny that differed from the mean of the herd were penalized and the expected genetic merit of the progeny was still taken into account. Consequently, matings that produce progeny close to or better than the mean EBV of the herd are ranked higher on the generated index, whereas matings that produce progeny worse than mean are ranked lower on the generated index. Based on the herd standard deviation of the EBV reported in the present study, using the linear programming approach resulted in more homogeneous progeny than did either the random or the sequential mating methods. However, the reduction in inter-progeny variability was mainly achieved by an improvement of the EBV of the genetically inferior progeny rather than by a reduction in genetic merit of the genetically elite progeny. Thus, the linear programming strategy based on the index generated in the present study addresses producer demands to simultaneously produce high genetic merit progeny in the herd and reduce within-herd inter-animal variability.

One concern that often arises when producing a more homogeneous herd is the potential reduction of long-term genetic gain in the herd. However, due to the low intensity of selection generally applied to the dams-to-dams selection pathway, only 4% of genetic gain is attributed to the dam of the cow (Van Tassell and Van

Vleck, 1991; Schaeffer, 2006). Therefore, regardless of how the matings are allotted within a herd, long-term genetic gain is unlikely to be severely affected (independent of any effect of inbreeding on long-term genetic diversity). Furthermore, it has been reported that the best strategy to advance genetic gain in the herd is to exclude the bottom 20% of the herd (based on their EBV) from breeding replacement heifers, rather than to preferentially mate the top EBV animals (Johnson et al., 2018). Such a luxury of selection, however, depends on excellent reproductive performance.

CONCLUSIONS

The use of genomic relationships between candidate sires and dams to estimate expected progeny inbreeding is a superior approach to controlling the level of inbreeding in a herd than the traditionally used pedigree-based measures. An approach to overcome lack of genotype data among dams was presented that exploits the information of genotyped relatives to impute the missing genotypes. The resulting imputed allele dosages can be used in the calculation of genomic relationships, providing a better estimate than the traditionally used pedigree-based measures. Using both the genomic relationships and a linear programming method of mating allotment within the framework of the proposed index not only reduced the level of expected genomic inbreeding in the herd but also reduced inter- and intra-progeny variability in genetic merit for selected traits.

ACKNOWLEDGMENTS

This work was funded in part by a research grant from Science Foundation Ireland (award number 14/IA/2576) and by a research grant from Science Foundation Ireland and the Department of Agriculture, Food and Marine, on behalf of the Government of Ireland, under the grant 16/RC/3835 (VistaMilk; Dublin, Ireland).

REFERENCES

- Barczak, E., A. Wolc, J. Wójtowski, P. Ślósarz, and T. Szwaczkowski. 2009. Inbreeding and inbreeding depression on body weight in sheep. *J. Anim. Feed Sci.* 18:42–50. <https://doi.org/10.22358/jafs/66366/2009>.
- Bouwman, A. C., J. M. Hickey, M. P. Calus, and R. F. Veerkamp. 2014. Imputation of non-genotyped individuals based on genotyped relatives: Assessing the imputation accuracy of a real case scenario in dairy cattle. *Genet. Sel. Evol.* 46:6. <https://doi.org/10.1186/1297-9686-46-6>.
- Ferenčaković, M., J. Sölkner, and I. Curik. 2013. Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. *Genet. Sel. Evol.* 45:42. <https://doi.org/10.1186/1297-9686-45-42>.

- Fernández, A., J. Rodríguez, M. Toro, M. Rodríguez, and L. Silió. 2002. Inbreeding effects on the parameters of the growth function in three strains of Iberian pigs. *J. Anim. Sci.* 80:2267–2275. <https://doi.org/10.1093/ansci/80.9.2267>.
- Flock, D. K., H. Ameli, and P. Glodek. 1991. Inbreeding and heterosis effects on quantitative traits in a White Leghorn population under long-term reciprocal recurrent selection. *Br. Poult. Sci.* 32:451–462. <https://doi.org/10.1080/00071669108417370>.
- Gengler, N., P. Mayeres, and M. Szydlowski. 2007. A simple method to approximate gene content in large pedigree populations: Application to the myostatin gene in dual-purpose Belgian Blue cattle. *Animal* 1:21–28. <https://doi.org/10.1017/S1751731107392628>.
- Johnson, T., K. Eketone, L. McNaughton, K. Tiplady, J. Voogt, R. Sherlock, G. Anderson, M. Keehan, S. Davis, and R. Spelman. 2018. Mating strategies to maximize genetic merit in dairy cattle herds. *J. Dairy Sci.* 101:4650–4659. <https://doi.org/10.3168/jds.2017-13538>.
- Kardos, M., G. Luikart, and F. W. Allendorf. 2015. Measuring individual inbreeding in the age of genomics: Marker-based measures are better than pedigrees. *Heredity* 115:63–72. <https://doi.org/10.1038/hdy.2015.17>.
- Keller, M. C., P. M. Visscher, and M. E. Goddard. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense SNP data. *Genetics* 189:237–249. <https://doi.org/10.1534/genetics.111.130922>.
- Kinghorn, B. P. 1998. Mate selection by groups. *J. Dairy Sci.* 81:55–63. [https://doi.org/10.3168/jds.S0022-0302\(98\)70154-7](https://doi.org/10.3168/jds.S0022-0302(98)70154-7).
- Marras, G., G. Gaspa, S. Sorbolini, C. Dimauro, P. Ajmone-Marsan, A. Valentini, J. L. Williams, and N. P. Macciotta. 2015. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim. Genet.* 46:110–121.
- McParland, S., J. F. Kearney, M. Rath, and D. P. Berry. 2007. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. *J. Dairy Sci.* 90:4411–4419. <https://doi.org/10.3168/jds.2007-0227>.
- Meuwissen, T., and Z. Luo. 1992. Computing inbreeding coefficients in large populations. *Genet. Sel. Evol.* 24:305–313. <https://doi.org/10.1186/1297-9686-24-4-305>.
- Meuwissen, T. H. E. 1997. Maximizing the response of selection with a predefined rate of inbreeding. *J. Anim. Sci.* 75:934–940. <https://doi.org/10.2527/1997.754934x>.
- Miglior, F., B. Van Doormaal, and G. Kistemaker. 2008. Phenotypic analysis of inbreeding depression for traits measured in Canadian dairy cattle breeds. Accessed Oct. 2, 2018. <https://www.cdn.ca/Articles/GEBMAY2008/Filippo%20-%20Inbreeding%20Depression%20-%20GEB%2004-2008.pdf>.
- Mulder, H. A., M. P. L. Calus, T. Druet, and C. Schrooten. 2012. Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. *J. Dairy Sci.* 95:876–889. <https://doi.org/10.3168/jds.2011-4490>.
- Pryce, J. E., B. J. Hayes, and M. E. Goddard. 2012. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. *J. Dairy Sci.* 95:377–388. <https://doi.org/10.3168/jds.2011-4254>.
- Purfield, D. C., M. McClure, and D. P. Berry. 2016. Justification for setting the individual animal genotype call rate threshold at eighty-five percent. *J. Anim. Sci.* 94:4558–4569. <https://doi.org/10.2527/jas.2016-0802>.
- Roche, J. R., D. P. Berry, A. M. Bryant, C. R. Burke, S. T. Butler, P. G. Dillon, D. J. Donaghy, B. Horan, K. A. Macdonald, and K. L. Macmillan. 2017. A 100-year review: A century of change in temperate grazing dairy systems. *J. Dairy Sci.* 100:10189–10233. <https://doi.org/10.3168/jds.2017-13182>.
- Sargolzaei, M. 2014. SNP1101 User's Guide. Version 1.0.
- Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15:478.
- Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123:218–223.
- Sonesson, A., J. Woolliams, and T. Meuwissen. 2010. Maximising genetic gain whilst controlling rates of genomic inbreeding using genomic optimum contribution selection. Abstract no. 892 in Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Germany.
- Sonesson, A. K., and T. H. Meuwissen. 2000. Mating schemes for optimum contribution selection with constrained rates of inbreeding. *Genet. Sel. Evol.* 32:231. <https://doi.org/10.1051/gse:2000116>.
- Strandén, I., and K. Vuori. 2006. RelaX2: Pedigree analysis programme. Pages 27–30. in Proceedings of the 8th World Congress on Genetics Applied to Livestock Production. Instituto Prociência.
- Sun, C., P. VanRaden, J. O'Connell, K. Weigel, and D. Gianola. 2013. Mating programs including genomic relationships and dominance effects. *J. Dairy Sci.* 96:8014–8023. <https://doi.org/10.3168/jds.2013-6969>.
- Van Tassell, C., and L. D. Van Vleck. 1991. Estimates of genetic selection differentials and generation intervals for four paths of selection. *J. Dairy Sci.* 74:1078–1086. [https://doi.org/10.3168/jds.S0022-0302\(91\)78258-1](https://doi.org/10.3168/jds.S0022-0302(91)78258-1).
- VanRaden, P. M., K. M. Olson, G. R. Wiggans, J. B. Cole, and M. E. Tooker. 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *J. Dairy Sci.* 94:5673–5682. <https://doi.org/10.3168/jds.2011-4500>.
- Weigel, K. A. 2001. Controlling inbreeding in modern breeding programs. *J. Dairy Sci.* 84:E177–E184. [https://doi.org/10.3168/jds.S0022-0302\(01\)70213-5](https://doi.org/10.3168/jds.S0022-0302(01)70213-5).
- Weigel, K. A., and S. W. Lin. 2000. Use of computerized mate selection programs to control inbreeding of Holstein and Jersey cattle in the next generation. *J. Dairy Sci.* 83:822–828. [https://doi.org/10.3168/jds.S0022-0302\(00\)74945-9](https://doi.org/10.3168/jds.S0022-0302(00)74945-9).
- Wright, S. 1922. Coefficients of inbreeding and relationship. *Am. Nat.* 56:330–338.
- Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, and G. W. Montgomery. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–569.
- Zhang, Q., M. P. Calus, B. Guldbbrandtsen, M. S. Lund, and G. Sahana. 2015. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genet.* 16:88. <https://doi.org/10.1186/s12863-015-0227-7>.