GSE Genetics Selection Evolution

RESEARCH Open Access

Estimates of missing heritability for complex traits in Brown Swiss cattle

Sergio-Iván Román-Ponce[1,2,3*], Antonia B Samoré[1], Marlies A Dolezal[1], Alessandro Bagnato[1] and Theo HE Meuwissen[2]

Abstract

Background: Genomic selection estimates genetic merit based on dense SNP (single nucleotide polymorphism) genotypes and phenotypes. This requires that SNPs explain a large fraction of the genetic variance. The objectives of this work were: (1) to estimate the fraction of genetic variance explained by dense genome-wide markers using 54 K SNP chip genotyping, and (2) to evaluate the effect of alternative marker-based relationship matrices and corrections for the base population on the fraction of the genetic variance explained by markers.

Methods: Two alternative marker-based relationship matrices were estimated using 35 706 SNPs on 1086 dairy bulls. Both pedigree- and marker-based relationship matrices were fitted simultaneously or separately in an animal model to estimate the fraction of variance not explained by the markers, i.e. the fraction explained by the pedigree. The phenotypes considered in the analysis were the deregressed estimated breeding values (dEBV) for milk, fat and protein yield and for somatic cell score (SCS).

Results: When dEBV were not sufficiently accurate (50 or 70%), the estimated fraction of the genetic variance explained by the markers was around 65% for yield traits and 45% for SCS. Scaling marker genotypes with locus-specific frequencies of heterozygotes slightly increased the variance explained by markers, compared with scaling with the average frequency of heterozygotes across loci. The estimated fraction of the genetic variance explained by the markers using separately both relationships matrices followed the same trends but the results were underestimated. With less accurate dEBV estimates, the fraction of the genetic variance explained by markers was underestimated, which is probably an artifact due to the dEBV being estimated by a pedigree-based animal model.

Conclusions: When using only highly accurate dEBV, the proportion of the genetic variance explained by the Illumina 54 K SNP chip was approximately 80% for Brown Swiss cattle. These results depend on the SNP chip used and the family structure of the population, i.e. more dense SNPs and closer family relationships are expected to result in a higher fraction of the variance explained by the SNPs.

Background

Genome-wide dense marker arrays that are available for livestock populations cover all chromosomes with dense single nucleotide polymorphism (SNP) markers [1]. Many dairy cattle populations are currently being genotyped using these arrays [2-4]. The main objective is to apply genomic selection (GS) [5]. GS allows prediction of the genetic merit of young animals based on marker information in the absence of own performance data. The marker effects are estimated in a reference population, which must have both genotypic and phenotypic records. In the case of dairy bulls, phenotypic data come from genetic evaluations in the form of daughter yield deviation (DYD) or deregressed estimated breeding values (dEBV) [6].

Identity by descent (IBD) alleles refer to alleles that descend from a common ancestor in the base population [7]. The coefficient of coancestry between two animals is defined as the probability that two randomly sampled alleles from the two animals are IBD [8], and

* Correspondence: romanponce@hotmail.com
[1]Dipartimento di Scienze e Tecnologie Veterinarie per la Sicurezza Alimentare, Università degli Studi di Milano, Via Celoria 10, Milano 20133, Italia
[2]Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, P.O. Box 5003, Oslo N-1432 Ås, Norway
Full list of author information is available at the end of the article

twice the coancestry is defined as their numerator relationship [8]. This approach leads to the estimation of a matrix of relationships based on the pedigree information. The latter is fundamental to estimate the genetic parameters for complex traits such as heritability (defined as the proportion of the phenotypic variance in a population that is attributed to additive genetic effects). The relationship matrix based on pedigree data dates back to a base population, for which parents are unknown and which is considered unrelated, unselected and non-inbred. The choice of the base population affects the estimate of the additive genetic variance [9].

However, the relationship matrix can also be estimated from genome-wide genetic markers such as panels of SNPs [10-12]. Methods have been developed to construct such marker-based relationship matrices [12-15]. Recently, these relationship matrices have been used to dissect the additive genetic variance of complex traits [16].

The proportion of the genetic variance not captured by markers (C_{miss}) represents the variance that cannot be used by GS and affects the maximum accuracy that can be achieved by GS [17]. The term 'missing heritability' [18] describes the fact that marker-phenotype associations identified in genome-wide association studies do not explain all the genetic variance in complex traits (e.g. height in humans). Some strategies have been proposed to reduce C_{miss}: (1) increasing the sample size in order to also detect genes with smaller effects, (2) expanding the studies to non-European samples in human genetics, (3) enlarging the collection of phenotypes to explore gene-gene interactions, (4) changing the structure of the training population, mainly in terms of the relatedness of the included individuals, and (5) moving to the genomic selection approach instead of estimating the marker effect for each SNP individually [13,19,20]. In animal breeding, some results suggest that the Illumina Bovine54K chip array (Illumina Inc., San Diego, CA) does not capture all the additive genetic variation for all dairy traits [21-23], even when using the GS approach, it estimates simultaneously all the SNP effects.

The main objective of this study was to estimate the fraction of the genetic variance not explained by the 54 K Illumina SNP chip. Two alternative marker-based relationship matrices were used for analysis.

Methods

Genotypic and phenotypic data

A total of 1092 Italian Brown Swiss bulls were genotyped with the Illumina Bovine54K chip (Illumina Inc., San Diego, CA). These bulls were born between 1963 and 2002. Figure 1 shows the distribution of the genotyped bulls over the birth years. All the SNPs on the X-chromosome were excluded from the analysis, which left 51 582 markers. The quality control process removed 1421 SNPs that had more than 5% missing genotypes and 14 455 SNPs with a minor allele frequency lower than 5%. Six sires were deleted because their genotyping rate was lower than 95%. Editing was performed with

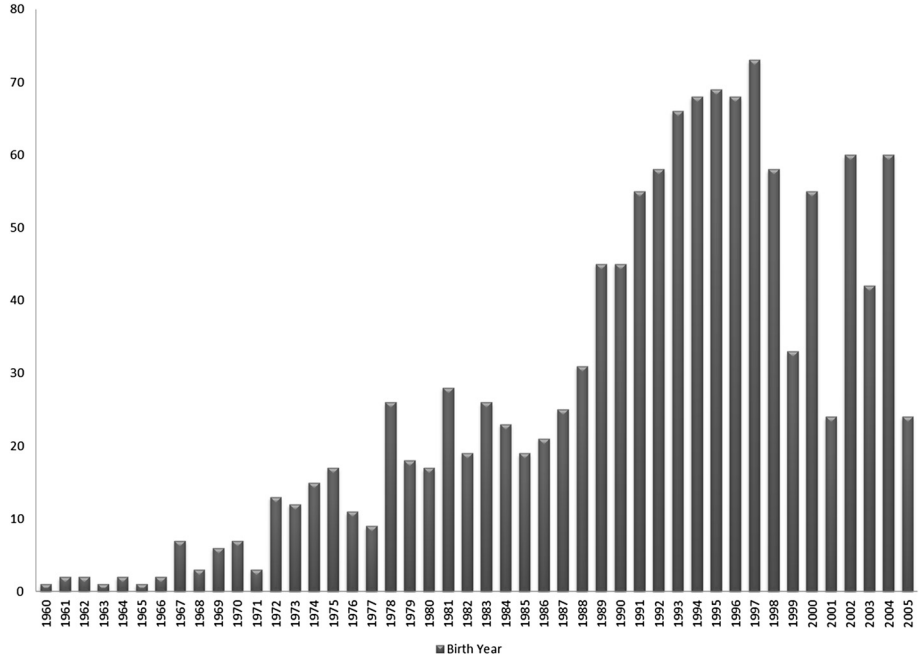

Figure 1 Distribution of birth years of the 1092 genotypes Italian Brown Swiss bulls.

two different software packages: SAS® (SAS Inst. Inc., Cary, NC) and PLINK v1.07 [24]. At the end of the quality control process, genotypes were available for 1086 sires with 35 706 SNPs and with a missing genotype rate of 0.66%.

The phenotypic data available were the EBV for fat yield (FAT), milk yield (MILK), protein yield (PROT) and somatic cell score in milk (SCS) for each bull, which were calculated by the Italian National Association of Brown Swiss (ANARB). The EBV were deregressed as proposed by Garrick [21], in order to eliminate the shrinkage contained in the EBV and to remove ancestral information. The deregressed EBV (dEBV) were used as phenotypic records for the bulls with heritability equal to the reliability of the EBV.

Three subsets were formed according to the reliability of EBV as follows: animals with a reliability of at least 50% for each trait; animals with a reliability greater than 70% for each trait; animals with a reliability of at least 90% for each trait.

Relationship matrices: A and G

A pedigree file was extracted from the Italian Brown Swiss herd book. Pedigree was traced back five generations and the pedigree file included 6826 entries. The completeness in the pedigree was 100% up to the grandparents, and decreased to ~90% thereafter. The equivalent number of known generations as calculated by the software Pedig [25] was on average 5.14 and the median was 5.23. The pedigree file was used to estimate the additive genetic relationships (**A**) with an adapted version of the procedure proposed by Meuwissen and Luo [26], as implemented in ASREML [27].

Two genomic relationship matrices (**G**) were computed for all genotyped animals. The first $\mathbf{G_V}$ was based on the method proposed by VanRaden [12]. Let **M** be the marker-genotype matrix with number of individuals (n) and number of loci (m) as dimensions. The elements in the matrix **M** were coded as −1 (homozygous for one allele) 0 (heterozygous) and 1 for (homozygous for the other allele). The nxm matrix **P** contains columns with all elements $2(p_i\text{-}0.5)$, where p_i is the frequency of the second allele at locus *i*. The matrix **P** was subtracted from **M** to give **Z = M - P**. Finally, matrix $\mathbf{G_V}$ was calculated as:

$$\mathbf{G_V} = \frac{\mathbf{ZZ}'}{2\sum_{i=1}^{m} p_i(1-p_i)}.$$

The second genomic relationship matrix ($\mathbf{G_Y}$) was computed as:

$$\mathbf{G_Y} = \frac{\mathbf{WW}'}{m},$$

where **W** is the **Z** matrix but with each element scaled based on the allele frequency of each locus as follows: $w_{ij} = \frac{z_{ij}}{\sqrt{2p_j(1-p_j)}}$ [12,14].

Correction for the base population

Both the **G** matrix and the pedigree-based relationship matrix, **A**, are expressed relative to a base population, i. e. an original population in which all animals are assumed unrelated and non-inbred, and these populations may differ between the pedigree-based and genomic relationship matrices [15]. To correct for these differences, the scale of **G** was changed to that of **A** based on Wright's F-statistic [7]. We expressed the total inbreeding of animal *i* in the **G** matrix as:

$$F_{it} = G_{ii}-1 \;\; or \;\; F_{it} = F_{st} + (1-F_{st}) \;\; F_{is},$$

where F_{st} is the average inbreeding in the population, i.e. the average of the diagonal elements of **G** minus 1, and F_{is} is the inbreeding of animal *i* relative to the population average inbreeding F_{st}, which is calculated as: $F_{is} = \frac{(F_{it}-F_{st})}{(1-F_{st})} = \frac{(G_{ii}-1-F_{st})}{(1-F_{st})}$.

The average population inbreeding of **G** was set equal to that of **A** by rescaling the diagonal element of **G** corresponding to individual *i* as:

$$G^*_{jj} = A_{st} + (1-A_{st})F_{st} + 1,$$

Where A_{st} is the average of the diagonals of **A** minus 1. The off-diagonals of **G** were rescaled similarly, using the same F_{st} and A_{st} values. Numerator relationships were transformed to kinships, ∅, i.e. by dividing the relationship by 2, and performing the base-correction on the kinship level, which is the same level as that of inbreeding, i.e.

$$\varnothing_{jis} = \frac{\left(\frac{G_{ji}}{2} - F_{st}\right)}{(1-F_{st})}, \text{and}$$

$$G^*_{ji} = 2\left[A_{st} + (1-A_{st})\varnothing_{jis}\right],$$

where $\varnothing_{jis}$ is the kinship of animal *j* and *i* relative to the base population inbreeding, F_{st}.

Estimation of variance components

To estimate the fraction of the genetic variance captured by dense markers covering the entire genome, the approach of Goddard et al. [28] was used. Both matrix **A** and **G** were fitted in the model simultaneously in order to estimate the fraction of the genetic variance captured by each of these matrices. The variance component analyses were performed by ASREML-R [29], using the following model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{u} + \mathbf{e},$$

where **y** is the vector of the dEBV; μ is the overall mean; $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are the incidence matrices for pedigree-based and genomic random animal effects, respectively; **a** is the vector of the random additive genetic animal effects using the pedigree-based relationship matrix, with $\mathbf{a} \sim N(0, \mathbf{A}\sigma^2_{\mathbf{a}})$; **u** is the vector of random additive genetic effect using the genomic relationship matrix, with $\mathbf{u} \sim N(0, \mathbf{G}\sigma^2_{\mathbf{u}})$; and finally, **e** is the vector of random residual effects. Because the number of daughters per bull was high for all bulls, the reliabilities of the dEBV were high and varied little between bulls, and a homogeneous error variance structure was assumed.

If we assume that **A** is an unbiased estimate of **G**, and write $\mathbf{G} = \mathbf{A} + \mathbf{D}$ [28], where **D** is a matrix of deviations from pedigree relationships due to the segregation of a finite number of chromosome segments in the genome, the genetic variance of the records becomes $V(\mathbf{g}) = \mathbf{G}\sigma^2_u + \mathbf{A}\sigma^2_a = \mathbf{A}(\sigma^2_u + \sigma^2_a) + \mathbf{D}\sigma^2_u$. Hence, as in a model that fits only pedigree relationships ($\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_1\mathbf{a} + \mathbf{e}$), the total genetic variance is explained by the **A** matrix and the segregation of chromosome segments that are traced by the markers is explained by σ^2_u. The fraction of genetic variance not captured by the markers on the SNP chip (C_{miss}) was thus estimated as:

$$C_{miss} = 1 - \frac{\sigma^2_u}{\sigma^2_g} = 1 - \frac{\sigma^2_u}{\left(\sigma^2_a + \sigma^2_u\right)},$$

where σ^2_g is the total genetic variance, σ^2_u is the variance due to marker-based relationships and σ^2_a is the variance due to pedigree-based relationships.

The two additive genetic variances were also estimated by fitting each separately: the additive genetic animal variance using the pedigree-based relationship matrix ($\boldsymbol{\sigma^2_{a0}}$) and the additive genetic variance using the genomic relationship matrix ($\boldsymbol{\sigma^2_{g0}}$). The estimate of $\boldsymbol{\sigma^2_{a0}}$ was used to calculate an alternative estimate for the fraction of genetic variance not addressed by the markers on the SNP chip (C_{miss2}) as follows: $C_{miss2} = 1 - \frac{\sigma^2_{u0}}{\sigma^2_{a0}}$. The estimate C_{miss2} has the advantage that σ^2_{a0} is known to yield an unbiased estimate of the genetic variance, but it has the disadvantage that σ^2_{u0} is likely to include more genetic variance than that explained by QTL that are in LD with the markers [11]. E.g. if only some of the chromosomes contain markers, these markers can explain genetic variance at the unmarked chromosomes, because the markers trace family relationships. If, in the latter case, the pedigree-based relationship matrix is fitted simultaneously with the marker-based relationship matrix, the variance due to the unmarked chromosomes is expected to be included in the polygenic variance, σ^2_a, because the pedigree-based relationship matrix more closely resembles the family relationships at the unmarked chromosomes than at the marked chromosomes, which may show relationships that (randomly) deviate from the pedigree. Thus, C_{miss2} is expected to underestimate the fraction of missing genetic variance.

Results

Descriptive statistics

Descriptive statistics for each trait and dataset are in Table 1. In the group of bulls with dEBV reliabilities of at least 50%, the dEBV average reliability was ~90% (±7%) for the production traits (FAT, PROT and MILK), and 82.6% (±10.7%) for SCS. The subset of sires with dEBV reliabilities of at least 70% had a similar average reliability of ~91% (±5%) for the production traits. The lowest average reliability in this subset was 85.7% (±7.4%) for SCS. Finally, the subset of bulls with reliabilities of at least 90% had an average reliability close to ~94% (±3%) for all traits. As expected, the differences in the average of the reliabilities between traits tended to decrease with increasing minimum reliability requirements.

Proportion of genetic variance not explained by markers

The fraction of genetic variance not explained by molecular markers based on C_{miss} was estimated for all datasets (50, 70 and 90 dEBV reliabilities) and traits (FAT, PROT, MILK and SCS). Results are in Table 2. For dFAT50, the estimate of C_{miss} was 0.373 ± 0.068 based on $\mathbf{G_V}$ and 0.363 ± 0.069 based on $\mathbf{G_Y}$. The estimates of C_{miss} were smaller for the dFAT70 subset than for the dFAT50 subset. For dFAT90, the estimate was 0.305 ± 0.074 $\mathbf{G_V}$, while the $\mathbf{G_Y}$ matrix did not result in converged variance component estimates. Algorithms other than the AI-REML algorithm might have converged (e.g. the EM-algorithm, which is known to be slow), but the convergence difficulties are probably due to the small size of the dataset, thus resulting variance component estimates would have been unreliable.

The fraction of the genetic variance not explained by molecular markers based on C_{miss2} through the additive genetic variances was estimated separately for all datasets and traits (Table 3). Results for C_{miss2} followed the same trends as for C_{miss} but the values of C_{miss2} were lower probably due to its underestimation of the fraction of the missing genetic variance.

Results for dMILK, dPROT and dSCS were similar to those described above for dFAT for both genomic relationship matrices. Estimates of C_{miss} for dMILK70 and dPROT70 hardly differed from those for dMILK50 and dPROT50, respectively. The subsets with dEBV90 resulted in estimates of C_{miss} of 0.199 (±0.101) for dMILK90 and 0.206 (±0.098) for dPROT90 when using $\mathbf{G_Y}$. These estimates were not significantly different from those obtained with the larger datasets for the same

Table 1 Descriptive statistics for de-regressed estimated breeding values (dEBV) and reliabilities (r^2) for production traits*

Trait	Subset label	Number of observations	dEBV		r^2 (%)	
			Mean	SD	Mean	SD
Fat yield	dFAT50	1034	−8.1	26.3	90.2	7.4
	dFAT70	1006	−8.7	26.1	91.0	5.8
	dFAT90	655	−12.7	25.4	94.3	2.9
Milk yield	dMILK50	1034	−205.9	666.9	90.7	7.3
	dMILK70	1014	−214.7	665.6	91.4	5.6
	dMILK90	691	−316.1	646.9	94.4	2.9
Protein yield	dPROT50	1034	−8.2	23.3	90.6	7.1
	dPROT70	1009	−8.7	23.2	91.3	5.6
	dPROT90	681	−12.1	22.9	94.4	2.9
Somatic cell score	dSCS50	978	0.246	1.206	82.6	10.7
	dSCS70	848	0.233	1.118	85.7	7.4
	dSCS90	223	0.018	0.972	95.2	2.9

*Subsets of the genotyped sire population were divided based on minimum reliabilities (50, 70, or 90); SD: standard deviation.

traits (dEBV50 or dEBV70), although they were systematically lower for all traits.

The highest estimates for C_{miss} were obtained for dSCS50, with 0.532 (±0.091) for $\mathbf{G_V}$. When using $\mathbf{G_Y}$, the corresponding C_{miss} estimate was lower (0.486 ± 0.095). The smallest C_{miss} estimate was obtained for dSCS90: 0.061 (±0.197) using $\mathbf{G_Y}$. The variance component analysis with $\mathbf{G_V}$ on the same dataset did not converge. This was the smallest dataset and, although the average reliability was the highest, estimates of C_{miss} were not significantly different from 0.

Table 2 Proportion of genetic variance not explained by markers (C_{miss}) ± standard error (SE) for dEBV for production traits*[1]

Label	G_Y	G_V
dFAT50	0.363 ± 0.069	0.373 ± 0.068
dFAT70	0.363 ± 0.072	0.369 ± 0.070
dFAT90	NC	0.305 ± 0.074
dMILK50	0.337 ± 0.076	0.357 ± 0.074
dMILK70	0.342 ± 0.077	0.358 ± 0.075
dMILK90	0.199 ± 0.101	0.245 ± 0.098
dPROT50	0.345 ± 0.077	0.363 ± 0.074
dPROT70	0.344 ± 0.078	0.357 ± 0.076
dPROT90	0.206 ± 0.098	0.235 ± 0.095
dSCS50	0.486 ± 0.095	0.532 ± 0.091
dSCS70	0.492 ± 0.101	0.530 ± 0.097
dSCS90	0.061 ± 0.197	NC

*Subsets of the genotyped sire population were divided based on minimum reliabilities (50, 70, or 90); NC: Log-likelihood was not available since the iterative procedure was not convergent; [1]$\mathbf{G_Y}$, and $\mathbf{G_V}$: genomic relationship matrices as proposed by [14] and [12], respectively, and corrected to the same base population.

In general, estimates of C_{miss2} decreased as the reliability of the dEBV increased. Estimates of C_{miss2} differed from estimates of C_{miss}, probably because C_{miss2} is expected to underestimate the fraction of the missing genetic variance.

Discussion

We estimated the fraction of the genetic variance not accounted by SNPs in the marker panel (C_{miss}) based on the Illumina 54 K SNP chip for complex traits in dairy cattle. The results showed that the estimates of C_{miss} depended on the reliability of the phenotypic traits considered, i.e. the dEBV used as response values. When the accuracy of the dEBV increases, i.e. when the correlation

Table 3 Proportion of genetic variance not explained by markers (C_{miss2}) for dEBV for production traits*[1]

Label	G_Y	G_V
dFAT50	0.116	0.097
dFAT70	0.108	0.089
dFAT90	0.026	0.024
dMILK50	0.073	0.055
dMILK70	0.075	0.057
dMILK90	0.125	0.101
dPROT50	0.054	0.035
dPROT70	0.052	0.031
dPROT90	0.031	0.008
dSCS50	0.149	0.149
dSCS70	0.152	0.152
dSCS90	−0.024	−0.024

*Subsets of the genotyped sire population were divided based on minimum reliabilities (50, 70, or 90); [1]$\mathbf{G_Y}$ and $\mathbf{G_V}$: genomic relationship matrices as proposed by [14] and [12], respectively.

between dEBV and the true breeding value increases, the proportion of the genetic variance explained by SNPs tended to increase. When the reliability of the dEBV is low, the family/pedigree information greatly contributes to the estimation of the EBV, which results in a larger fraction of the variance being explained by **A** and, in turn, in upward biases of C_{miss}. Because the estimates of the C_{miss} values, are expected to be overestimated due to the use of (family information in) dEBV, the best estimates of C_{miss} are obtained for data sets with high reliabilities, which resulted in estimates around 0.2. This implies that the maximum accuracy of GEBV is $\sqrt{(1\text{-}C_{miss})} \approx 0.9$, which agrees with the result of Daetwyler [22], who studied the increase in the accuracy of GEBV with increasing training population sizes.

For all production traits, the fraction of the genetic variance not explained by the SNPs was significantly different from 0, even when the phenotypes were very accurate (reliability > 90%), and were, therefore, very close to the true breeding values. Correction for the base population did not affect the fraction of the genetic variance explained by markers for any of the marker-based relationships here used. The differences in $C_{\mathbf{miss}}$ estimates between using $\mathbf{G_V}$ and $\mathbf{G_Y}$ were negligible for all traits and all subsets. Similarly, when using EBV instead of dEBV (results not shown), the results were virtually the same.

If original performance records of production and SCS phenotypes are used to estimates C_{miss}, instead of dEBV, the upward biases mentioned above are not expected to occur. The error variances would be higher than when using dEBV, but the value of σ_a^2 would not be inflated, because family information does not contribute to own phenotype (in contrast to dEBV phenotypes).

The sources of phenotypic information used in genomic analyses are very heterogeneous and vary from individuals with highly reliable information, i.e. progeny-tested bulls, and animals with phenotypes with low levels of accuracy, i.e. young cows. To take into account these differences in reliability in a weighted analysis, it is necessary to know the value of C_{miss} for each phenotype [22]. In addition, a polygenic effect must be included in the model to account for unmarked genetic effects. Knowledge of the fraction of the genetic variance not explained by markers is also required to predict the accuracy of the genomic predictions for each individual in the population, since it affects the maximum accuracy that can be achieved [17].

The base population correction of the genomic relationship matrix generally affected neither the proportion of genetic variance captured by markers, nor the genetic variance captured by the pedigree-based relationship matrices, which agrees with [17,30] but not with [31]. The latter authors, however, scaled the relationships in the opposite direction, i.e. when G relationships were too high, they scaled all relationships downwards, which further decreased the differences in relationships that were already small since relationships are bound by a maximum of 1 (and vice-versa when G relationships were too small). Moreover, the correction for the base population facilitates the integration of relationship matrices **A** and **G** into a single matrix (**H**), according to Legarra et al. [32], Christensen and Lund [13], and Meuwissen et al. [15].

We also estimated C_{miss2} using the pedigree-based estimate of genetic variance. The denominators of C_{miss} and C_{miss2} were significantly different from each other but both estimates revealed that the genomic relationship matrix could explain more than 95% of genetic variance if sufficiently reliable phenotypes are used (with reliabilities greater than 95%).

It should be noted that the estimates of C_{miss} and C_{miss2} depend on the SNP chip used, i.e. more dense SNP chips are expected to yield lower estimates of C_{miss} and C_{miss2} (a larger fraction of the variance is explained by the SNPs), and also on the family structure of the population [33]. Populations with more closely related individuals are expected to yield high LD between SNPs and QTL, even when they are physically quite far apart and, therefore, lower estimates of C_{miss}. The population structure of the Italian Brown Swiss population reflects that of a typical dairy breeding population, and, thus, our results probably apply also to other dairy breeding populations.

Conclusions

The fraction of genetic variance explained by genetic markers from high-density SNP panels was significantly different from 0 for the complex traits analyzed when the phenotypes are not highly accurate. The minimum fraction of the genetic variance not explained by the markers (C_{miss}) was equal to 0.2, which was estimated based on the most accurate phenotypes. This value agrees with other values reported in the literature. Correction of the genomic relationship matrix for the variance of the allele frequency of each locus ($\mathbf{G_Y}$) instead of the average frequency of heterozygotes ($\mathbf{G_V}$), hardly explained any additional genetic variance. Our estimate of C_{miss} of 0.2 implies that about 80% of the genetic variance is explained by the Illumina 54 K SNP chip. Values for C_{miss} are expected to depend on the density of the chip (a larger SNP chip is expected to explain a larger fraction of the genetic variance) and on family relationships in the population, i.e. closer family relationships are expected to reduce C_{miss}.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions
SIRP performed the study and drafted the manuscript. ABS contributed to writing the draft. SIRP, MAD and AB prepared the genotypic and phenotypic data. THEM planned and coordinated the whole study, and contributed to writing the manuscript. All the authors read and approved the final manuscript.

Acknowledgements
The helpful comments of three reviewers are gratefully acknowledged. We gratefully acknowledge the Italian Brown Cattle Breeders' Association (**ANARB**) for collecting, handling and sharing data. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 222664. ("Quantomics"). This article reflects only the author's views and the European Community is not liable for any use that may be made of the information contained herein.

Author details
[1]Dipartimento di Scienze e Tecnologie Veterinarie per la Sicurezza Alimentare, Università degli Studi di Milano, Via Celoria 10, Milano 20133, Italia. [2]Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, P.O. Box 5003, Oslo N-1432 Ås, Norway. [3]Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, C.E. Valles Centrales, CIRPAS, Melchor Ocampo 7, Etla, Oaxaca 68200, México.

Received: 24 January 2013 Accepted: 28 April 2014
Published: 4 June 2014

References

1. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TPL, Sonstegard TS, Van Tassell CP: **Development and characterization of a high density SNP genotyping assay in cattle.** *PLoS ONE* 2009, **4:**e5350.
2. Berry DP, Kearney F, Harris B: **Genomic selection in Ireland.** *Interbull Bull* 2009, **39:**29–34.
3. Schenkel FS, Sargolzaei M, Kistemaker G, Jansen GB, Sullivan P, Van Doormaal BJ, VanRaden PM, Wiggans GR: **Reliability of genomic evaluation of Holstein cattle in Canada.** *Interbull Bull* 2009, **39:**51–58.
4. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS: **Invited review: Reliability of genomic predictions for North American Holstein bulls.** *J Dairy Sci* 2009, **92:**16–24.
5. Meuwissen THE, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157:**1819–1829.
6. Calus MPL: **Genomic breeding values prediction: Methods and procedures.** *Animal* 2010, **4:**157–164.
7. Wright S: **Coefficients of inbreeding and relationship.** *Am Nat* 1922, **56:**330–338.
8. Malécot G: *Les Mathématiques de l'Hérédité.* Paris: Masson et Cie; 1948.
9. van der Werf JH, de Boer IJ: **Estimation of additive genetic variance when base populations are selected.** *J Anim Sci* 1990, **68:**3124–3132.
10. Fernando RL: **Genetic evaluation and selection using genotypic, phenotypic and pedigree information**. In *Proceedings of the 6th World Congress in Genetics Applied to Livestock Production: 11–16 January 1998; Armidale*, 26. 1998:329–336.
11. Habier D, Fernando RL, Dekkers JCM: **The impact of genetics relationship information on genome-assisted breeding values.** *Genetics* 2007, **177:**2389–2397.
12. VanRaden PM: **Efficient methods to compute genomic predictions.** *J Dairy Sci* 2008, **91:**4414–4423.
13. Christensen OF, Lund MS: **Genomic prediction when some animals are not genotyped.** *Genet Sel Evol* 2010, **42:**2.
14. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM: **Common SNPs explain a large proportion of the heritability for human height.** *Nat Genet* 2010, **42:**565–569.
15. Meuwissen THE, Luan T, Woolliams JA: **The unified approach to the use of genomic and pedigree information in genomic evaluations revisited.** *J Anim Breed Genet* 2011, **128:**429–439.
16. Lee SH, Goddard ME, Visscher PM, van der Werf JHJ: **Using the realized relationship matrix to disentangle confounding factors for the estimation of genetic variance components of complex traits.** *Genet Sel Evol* 2010, **42:**22.
17. Dekkers JC: **Prediction of response to marker-assisted and genomic selection using selection index theory.** *J Anim Breed Genet* 2007, **124:**331–341.
18. Maher B: **Personal genomes: The case of the missing heritability.** *Nature* 2008, **456:**18–21.
19. Manolio TA, Collins FS, Cox NJ, Golstein DB, Hindoff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boenhnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarrol SA, Visscher PM: **Finding the missing heritability of complex diseases.** *Nature* 2009, **461:**747–753.
20. Makowsky R, Pajewski NM, Klimentidis YC, Vazquez IA, Duarte CW, Allison DB, de los Campos G: **Beyond missing heritability: prediction of complex traits.** *PLoS Genet* 2011, **7:**e1002051.
21. Garrick DJ, Taylor JT, Fernando RL: **Deregressing estimated breeding values and weighting information for genomic regression analyses.** *Genet Sel Evol* 2009, **41:**55.
22. Daetwyler HD: **Genome-Wide Evaluation of Populations**. *PhD Thesis.* Wageningen: Wageningen University; 2009.
23. Haile-Mariam M, Nieuwhof GJ, Beard KT, Konstatinov KV, Hayes BJ: **Comparison of heritabilities of dairy traits in Australian Holstein-Friesian cattle from genomic and pedigree data and implications for genomic evaluations.** *J Anim Breed Genet* 2013, **130:**20–31.
24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC: **PLINK: a toolset for whole-genome association and population-based linkage analysis.** *Am J Hum Genet* 2007, **81:**559–575.
25. Boichard D, Maignel L, Verrier E: **The value of using probabilities of gene origin to measure genetic variability in a population.** *Genet Sel Evol* 1997, **29:**5–23.
26. Meuwissen THE, Luo Z: **Computing inbreeding coefficients in large populations.** *Genet Sel Evol* 1992, **24:**305–313.
27. Gilmour AR, Gogel BJ, Cullis BR, Thompson R: *ASREML User Guide Release 3.0.* Queensland, Australia: The Department of Primary Industries and Fisheries; 2009.
28. Goddard ME, Hayes B, Meuwissen THE: **Using the genomic relationship matrix to predict the accuracy of genomic selection.** *J Anim Breed Genet* 2011, **128:**409–421.
29. Butler D, Cullis B, Gilmour A, Gogel B: *ASReml-R Reference Manual, Version 3.* Queensland, Australia: The Department of Primary Industries and Fisheries; 2009.
30. Sorensen DA, Kennedy BW: **Estimation of genetic variances from unselected and selected populations.** *J Anim Sci* 1984, **59:**1213–1223.
31. Forni S, Aguilar I, Misztal I: **Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information.** *Genet Sel Evol* 2011, **43:**1.
32. Legarra A, Aguilar I, Misztal I: **A relationship matrix including full pedigree and genomic information.** *J Dairy Sci* 2009, **92:**4656–4663.
33. Jensen J, Su G, Madsen P: **Partitioning additive genetic variance into genomic and remaining polygenic components for complex traits in dairy cattle.** *BMC Genet* 2012, **13:**44.

doi:10.1186/1297-9686-46-36
Cite this article as: Román-Ponce *et al.*: Estimates of missing heritability for complex traits in Brown Swiss cattle. *Genetics Selection Evolution* 2014 **46**:36.