Genomic Predictions for Dry Matter Intake Using the International Reference Population of gDMI

Yvette de Haas[1], Jennie E. Pryce[2], Mario P.L. Calus[1], Ina Hulsegge[1], Diane M. Spurlock[3], Donagh Berry[4], Eileen Wall[5], Peter Løvendahl[6], Kent A. Weigel[7], Kevin Macdonald[8], Filippo Miglior[9], Nina Krattenmacher[10], Roel F. Veerkamp[1].

[1]Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, 6700 AH Wageningen, The Netherlands; [2]Dept. of Environment and Primary Industries & Dairy Futures CRC, Agribio, 5 Ring Road, La Trobe University, Bundoora, 3083, Australia; [3]Department of Animal Science, Iowa State University Ames, IA 50011, USA; [4]Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Co. Cork, Ireland; [5]Animal and Veterinary Sciences, SRUC, Easter Bush Campus, EH25 9RG, Scotland, United Kingdom; [6]Dept. Molecular Biology and Genetics, Aarhus University, AU-Foulum, 8330 Tjele, Denmark; [7]Department of Dairy Science, University of Wisconsin, Madison 53706, USA; [8]DairyNZ, Private Bag 3221, Hamilton 3240, New Zealand; [9]Canadian Dairy Network & University of Guelph, Guelph, ON, Canada; [10]Institute of Animal Breeding and Husbandry, Christian-Albrechts-University, Hermann-Rodewald-Straße 6, 24118 Kiel, Germany.

Abstract

In this study, we have demonstrated that using dry matter intake (DMI) phenotypes from multiple countries increases the accuracy of genomic breeding values for this important trait, provided a multi-trait approach is used. Data from Australia, Canada, Denmark, Germany, Ireland, the Netherlands, New Zealand, United Kingdom and two institutions in the United States were combined to estimate the accuracy of genomic prediction for DMI multi-trait models. The average accuracies was 0.44, and ranged from 0.37 (Denmark) to 0.54 (the Netherlands). Enlarging the combined dataset with unique phenotypes does increase the accuracy of the genomic prediction for DMI. This stimulates further international collaboration.

1. Introduction

Feed cost is the single-largest expense of dairy production (Vallimont et al., 2011) and has increased substantially over the last few years (Garcia, 2009). Although it is a crucial factor in the profitability of the dairy industry, little attention has been paid to improve feed efficiency through direct selection (Linn, 2006; Zamani et al., 2008). This is mainly due to the difficulties and costs associated with individual feed intake measurements (Kelly *et al*., 2010).

Therefore, for dry matter intake (DMI) and other difficult to measure traits, there has been recent interest in combining data from international research populations for genetic analysis (Banos *et al*., 2012; de Haas *et al*., 2012; Veerkamp *et al*., 2012; Berry *et al*., 2014). Reasons why genotypes and phenotypes from different research organisations are combined include adding statistical power to genome wide association studies and/or trying to improve the accuracy of genomic prediction.

Challenges when combining phenotypes from several countries include genotype by environment interactions and differences in trait definitions. A multi-trait model can handle traits that are measured in different environments as separate traits, and therefore treat both the genotype by environment interaction and differences in trait definitions properly. Genomic predictions for multiple traits are straightforward if a genomic BLUP (G-BLUP) methodology is used, as also demonstrated by De Haas *et al*. (2012). The objective of this study was to estimate the accuracy of genomic prediction for DMI, when analysed in a multiple-trait analysis, using the largest existing international database, with individual DMI records from Europe, North America and Australasia.

2. Material and Methods

Available data. Data on individual daily feed intake of Holstein-Friesian cows and heifers were available from nine countries in Europe,

North America and Australasia, with some countries providing data from more than one population of animals. Only data from parity one to five were retained for inclusion in the analysis; feed intake data from growing heifers (<2 years of age) in Australia and New Zealand were also available and retained for the analysis. Data on feed intake was transformed into DMI by multiplying wet feed intake by the respective dry matter content of that particular diet for the purposes of analysis. A more detailed description of the merging of the data sources and variance components across the different herds is given by Berry *et al*. (2014).

With this collaborative effort between research organizations in Australia (AUS), Canada (CAN), Denmark (DNK), Germany (GER), Ireland (IRL), the Netherlands (NLD), New Zealand (NZL), United Kingdom (UK) and Iowa and Wisconsin in United States of America (IOWA and WISC, respectively), resources were available on 233,189 feed intake records from 12,425 parities on 8,737 cows and heifers, of which 1,784 of them were nulliparous (Berry *et al*., 2014).

Phenotype. Fitted values for DMI per animal on day 70 in parity 2, predicted from estimated quadratic DMI curve for each animal by 5 parities was used as phenotype. Dry matter intake at 70 days in milk was chosen as the phenotype because this was the period when the largest number of actual DMI observations existed within the dataset and it is also close to the critical period of early lactation (Berry *et al*., 2014).

Validation sets. The accuracies of genomic predictions of DMI in 10 groups of animals (validation populations) were estimated, by excluding each of those groups one at a time from the reference population. Validation populations were subsets of the dataset based on progeny groups of sires in the different countries. Each validation population had animals from all countries. With this approach it can be shown if the accuracy of a bull's genomic estimated breeding value (GEBV) can be increased by using a multi-country reference population.

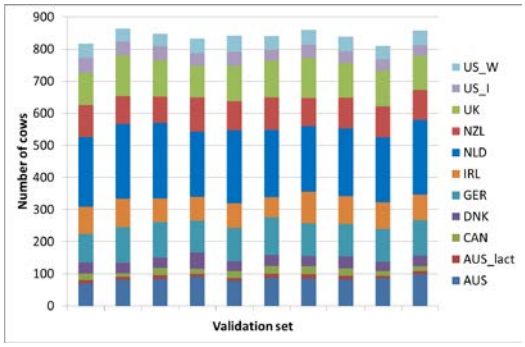

Figure 1. Number of cows of each country per validation set.

Generation of relationship matrix. Single nucleotide polymorphism genotypes were available on 5,999 animals, of which 5,429 had phenotypic information in this study. A total of 1,888 animals had Illumina high density genotype information and 4,111 had genotype information from the Illumina Bovine50 Beadchip. Imputation of Illumina high density genotypes for 5,999 animals to 591,213 SNP is described in detail by Pryce *et al*. (2014). Monomorphic SNPs as well as SNPs deviating from Hardy-Weinberg equilibrium were discarded and only autosomal SNPs were retained. Following editing, 583,375 SNPs remained for the calculation of the genomic relationship matrix (VanRaden, 2008; Yang *et al*., 2010). Pedigree information of all animals was traced back to the founder population; aliases in the pedigree were removed through the use of the INTERBULL identification cross reference tables and manual correction of the pedigree. The computation of the combined pedigree and genomic relationship matrix (H^{-1}) followed Aguilar *et al*. (2010) and Christensen and Lund (2010), and is for this data described in more detail by Berry *et al*. (2014). The H^{-1} matrix consisted of 51,486 identities.

Variance components. Fifty-five bivariate analyses were run to estimate all genetic and residual (co)variances for weighted average DMI at 70 DIM in parity 2 using animal linear mixed models in ASReml software (Gilmour *et al*., 2009), using the H^{-1} matrix. The model used was:

$$Yi Yj = \mu + \text{animal} + \text{residual}$$

Bending. The variance-covariance matrices for the genetic and residual effects were bent to make one positive definite matrix out of the 55 bivariate analyses. This was done with an iterative bending process where, in three steps the poorly estimated correlations (high s.e.) were bent until the smallest eigenvalue remained constant, conditional on the more accurate estimated correlations.

Accuracies of genomic predictions. The main objective of this study was to determine how well the genomic breeding values predict the true breeding values of individual animals. If the true breeding value of individuals were known, the accuracy of the genomic breeding values would be the correlation between the genomic breeding values and the true breeding values. In practice, the true breeding values are unknown, and the only data available are phenotypes, which are made up of the effect of the true breeding value and the environmental effect. Given this, the accuracy of the genomic breeding values has been derived as follows.

For each of the 10 validation sets in the dataset, a single validation set was removed from the dataset one by one. The SNP effects for DMI were calculated in the remaining dataset using the G-BLUP analysis in MiXBLUP (Mulder *et al.*, 2010). Genomic estimated breeding values were obtained for all animals from the same analysis for the animals in the validation set. For each validation set that was removed from the dataset, the GEBV were then correlated with a vector of phenotypes (phen) of DMI, corrected for the fixed effects as described above. This gave r(GEBV,phen). To adjust that phenotypes were used and not the true breeding values, r(GEBV,phen) was divided by h, where h was the square root of the estimated heritability of DMI in that country.

Correlations were calculated for each of the different validation populations between the breeding values estimated for all individuals in that validation population with an 11-trait analysis in MiXBLUP (Mulder *et al.*, 2010). These correlations were then averaged across validation sets within each country.

3. Results and Discussion

Genetic parameters. The estimated within country heritabilities for the weighted average DMI on 70 DIM in parity 2 range from 0.12 to 0.53 (Table 1). Many of the within country heritability estimates were in close proximity to each other and are consistent with previously published heritabilities for these populations (Coffey *et al.*, 2001; Berry *et al.*, 2007; de Haas *et al.*, 2012; Spurlock *et al.*, 2012) and elsewhere (Sondergaard *et al.*, 2002; Vallimont *et al.*, 2011).

Table 1. Estimated within country heritabilities for weighted average of DMI on 70d in parity 2.

	Heritability
Canada	0.21
Denmark	0.46
Australian heifers	0.32
New Zealand heifers	0.24
Germany	0.17
Iowa, US	0.53
Ireland	0.26
Netherlands	0.38
United Kingdom	0.26
Wisconsin, US	0.12

Genetic correlations were high between some countries (>0.8); e.g., Denmark and Germany, and Denmark and the Netherlands (Table 2). Whereas other countries show low, and even negative correlations amongst each other. This was the case for Ireland and New Zealand.

Some countries had comparatively low numbers of DMI records and consequently s.e. estimates of many genetic correlations were high. The genetic correlations with large s.e. may have affected the bending procedure. Based on the matrix of bended genetic correlations, countries can be grouped together (Figure 2). The three groups are: (1) the Australian heifer data and New Zealand plus

the Irish cows; (2) the Australian lactating data, plus the Dutch, Canadian and UK data; and (3) the data collected at both universities in US plus Germany and Denmark.

Accuracies of genomic predictions. The average accuracy of the genomic prediction (r(GEBV,TBV) for the fitted value of DMI on 70d in parity 2 is 0.44 (0.08). The accuracies range from 0.37 to 0.54 (Table 3). This accuracy is higher than the accuracy De Haas *et al*. (2012) estimated for a combined dataset of Australian, Dutch and UK data (0.35). Therefore, enlarging the reference set does improve the accuracy of genomic prediction of DMI.

4. Conclusions

This international consortium has created the world's biggest collection of data for feed intake on genotyped dairy cattle. In this study, we demonstrated that, provided a multi-trait approach is used, combining similar phenotypes across countries can increase the accuracy of genomic breeding values for important traits, such as DMI. Enlarging the combined dataset with unique phenotypes does increase the accuracy of the genomic prediction. This stimulates further international collaboration.

5. Acknowledgements

This work is performed by the gDMI consortium in which 15 parties from 10 countries have combined their knowledge, data and expertise. Financial support for gDMI from CRV (Arnhem, the Netherlands), ICBF (Cork, Ireland), CONAFE (Madrid, Spain), and DairyCo (Warwickshire, UK), The Natural Science and Engineering Research Council of Canada and DairyGen Council of Canadian Dairy Network (Guelph, ON, Canada) is gratefully appreciated, as well as the EU FP7 IRSES SEQSEL (Grant no. 317697). The Australian Research was funded by the Gardiner Foundation. The New Zealand research was funded by MBIE, DairyNZ, NZ Trade & Enterprise and LIC. The US research is funded by the National Institute of Food and Agriculture (NIFA) Grant no. 0224899, from the United States Department of Agriculture (USDA). The RobustMilk project was financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-21170.

6. References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S. & Lawlor, T.J. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci. 93:2,*743-752.

Banos, G., Coffey, M.P., Veerkamp, R.F., Berry, D.P. & Wall, E. 2012. Merging and characterising phenotypic data on conventional and rare traits from dairy cattle experimental resources in three countries. *Animal 6:7,* 1040-1048.

Berry, D.P., Coffey, M.P., Pryce, J.E., de Haas, Y., Løvendahl, P., Krattenmacher, N., Crowley, J.J., Wang, Z., Spurlock, D., Weigel, K., Macdonald, K. & Veerkamp, R.F. 2014. International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. *J. Dairy Sci. 97:6,* 3894-3905.

Berry, D.P., Horan, B., O'Donovan, M., Buckley, F., Kennedy, E., McEvoy, M. & Dillon, P. 2007. Genetics of grass dry matter intake, energy balance, and digestibility in grazing Irish dairy cows. *J. Dairy Sci. 90:10,* 4835-4845.

Christensen, O.F. & Lund, M.S. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol. 42:2*.

Coffey, M.P., Emmans, G.C. & Brotherstone, S. 2001. Genetic evaluation of dairy bulls for energy balance traits using random regression. *Anim. Sci. 73,* 29-40.

de Haas, Y., Calus, M.P.L., Veerkamp, R.F., Wall, E., Coffey, M.P., Daetwyler, H.D., Hayes, B.J. & Pryce, J.E. 2012. Improved accuracy of genomic prediction for dry matter intake of dairy cattle from combined European and Australian data sets. *J. Dairy Sci. 95:10,* 6103-6112.

Garcia, A. 2009. *Dairy profitability 1010: Milk quality and feed efficiency.* Page http://pubstorage.sdstate.edu/AgBio_Publications/articles/ExEx4042.pdf in Extension Extra. South Dakota State University, USDA.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. 2009. *ASREML. Program user manual.* NSW Agriculture, Orange Agricultural Institute, Forest Road, Orange, NSW, 2800, Australia.

Kelly, A.K., McGee, M., Crews, D.H., Fahey, A.G., Wylie, A.R. & Kenny, D.A. 2010. Effect of divergence in residual feed intake on feeding behavior, blood metabolic variables, and body composition traits in growing beef heifers. *J. Anim. Sci. 88:1,* 109-123.

Linn, J. 2006. Feed efficiency: Its economic impact in lactating cows. *WCDS Adv. Dairy Technol. 18,* 19-28.

Mulder, H.A., Lidauer, M., Stranden, I., Mäntysaari, E., Pool, M.H. & Veerkamp, R.F. 2010. *MiXBLUP manual.* Wageningen, the Netherlands.

Pryce, J.E., Johnston, B., Hayes, J., Sahana, G., Weigel, K.A., McParland, S., Spurlock, D., Krattenmacher, N., Spelman, R.J., Wall, E. & Calus, M.P.L. 2014. Imputation of genotypes from low density (50,000 markers) to high density (700,000 markers) of cows from research herds in Europe, North America, and Australasia using 2 reference populations. *J. Dairy Sci. 97:3,* 1799-1811.

Søndergaard, E., Sørensen, M.K., Mao, I.L. & Jensen, J. 2002. Genetic parameters of production, feed intake, body weight, body composition, and udder health in lactating dairy cows. *Livest. Prod. Sci. 77:1,* 23-34.

Spurlock, D.M., Dekkers, J.C.M., Fernando, R., Koltes, D.A. & Wolc, A. 2012. Genetic parameters for energy balance, feed efficiency, and related traits in Holstein cattle. *J. Dairy Sci. 95:9,* 5393-5402.

Vallimont, J.E., Dechow, C.D., Daubert, J.M., Dekleva, M.W., Blum, J.W., Barlieb, C.M., Liu, W., Varga, G.A., Heinrichs, A.J. & Baumrucker, C.R. 2011. Short communication: Heritability of gross feed efficiency and associations with yield, intake, residual intake, body weight, and body condition score in 11 commercial Pennsylvania tie stalls. *J. Dairy Sci. 94:4,* 2108-2113.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci. 91:11,* 4414-4423.

Veerkamp, R.F., Coffey, M.P., Berry, D.P., de Haas, Y., Strandberg, E., Bovenhuis, H., Calus, M.P.L. & Wall, E. 2012. Genome-wide associations for feed utilisation complex in primiparous Holstein-Friesian dairy cows from experimental research herds in four European countries. *Animal 6:11,* 1738-1749.

Yang, J.A., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E. & Visscher, P.M. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genet. 42:7,* 565-571.

Zamani, P., Miraei-Ashtiani, S.R. & Mohammadi, H. 2008. Genetic parameters of residual energy intake and its correlations with other traits in holstein dairy cattle. *Turk. J. Vet. Anim. Sci. 32:4,* 255-261.

Table 2. Matrix of bended genetic correlations between 11 countries (Australia lactating (AUS), Canada (CAN), Denmark (DNK), Australian heifers (AU_h), New Zealand heifers (NZ_h), Germany (GER), Iowa-USA (US_I), Ireland (IRL), the Netherlands (NLD), United Kingdom (UK) and Wisconsin-USA (US_W).

	AUS	CAN	DNK	AU_h	NZ_h	GER	US_I	IRL	NLD	UK
CAN	0.66									
DNK	0.56	0.32								
AU_h	0.27	0.30	0.05							
NZ_h	-0.26	0.09	-0.19	0.22						
GER	0.32	0.13	0.85	0.17	0.17					
US_I	0.36	0.14	0.79	-0.14	-0.06	0.68				
IRL	0.00	0.04	0.16	0.39	0.56	0.45	-0.15			
NLD	0.83	0.77	0.82	0.20	-0.14	0.62	0.63	0.02		
UK	0.57	0.80	0.37	0.61	0.40	0.46	0.08	0.50	0.68	
US_W	0.53	0.52	0.75	0.29	0.15	0.86	0.50	0.35	0.80	0.76

Table 3. The average of the approximated accuracy (acc) of genomic prediction estimated in a multivariate run between all countries in the gDMI dataset. In all analyses, a multi-country reference set was taken consisting of all data except the validation set. The corresponding standard errors (se) are shown separately.

	AUS	CAN	DNK	AU_h	NZ_h	GER	US_I	IRL	NLD	UK	US_W
acc	0.48	0.40	0.37	0.39	0.44	0.45	0.46	0.48	0.54	0.49	0.38
se	0.08	0.19	0.08	0.04	0.05	0.07	0.06	0.06	0.04	0.08	0.15

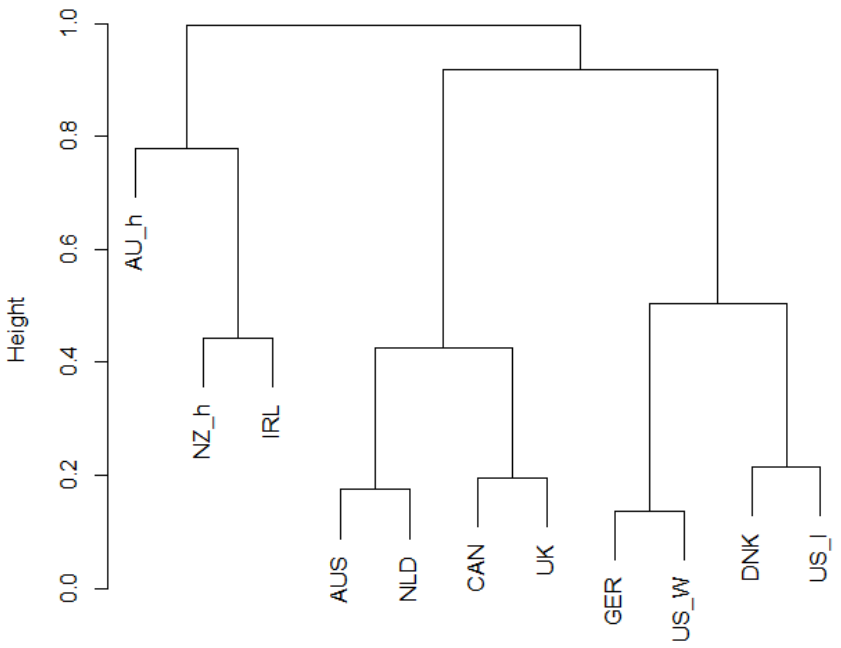

Figure 2. Dendogram of the 11 countries, showing which countries group together, based on the matrix of bended genetic correlations.