

Work in Progress on Italian Holstein Genomic Evaluation

J.B.C.H.M. van Kaam^{a}, G.B. Jansen^b, E.L. Nicolazzi^a and F. Canavesi^a*

^a – ANAFI, Via Bergamo 292, 26100 Cremona (CR), Italy,

jtkaaam@anafit.it, ezequielnicolazzi@anafit.it, fabiolacanavesi@anafit.it

^b – Dekoppel consulting, Via Rovera 10, 10010 Chiaverano (TO), Italy, *jansen.gerald@gmail.com*

* – Corresponding author: Tel. +39-0372-474278, Fax: +39-0372-474203

Abstract

In Italy various projects are working, which have as one of their aims to support the development of genomic selection for the Italian cattle breeding system. Anafi, the Italian Holstein association, and several Italian AI centres participate in these three projects: (1) the SelMol project, which involves several species and breeds (2) the ProZoo project is active in Holstein cattle and pigs (3) the Elica project, which is started to further increase the reference population for Holstein. During 2010 Anafi expects to introduce Holstein genomic breeding values.

Keywords: genomic selection, SNP, Holstein cattle

Introduction

Genomic selection is based on the simultaneous selection for thousands of single nucleotide polymorphisms (SNPs), which cover the entire genome at a density sufficiently high to exploit population-wide linkage disequilibrium between these SNP markers and quantitative trait loci. Since the demonstration of the possibilities by Meuwissen *et al.* (2001), the description of the possible impact on the breeding programs by Schaeffer (2006) and the availability of a dense SNP chip (Van Tassell *et al.*, 2008) enabling large scale genotyping, intense research efforts have been launched in all major dairy countries. Key advantage of the genomic selection approach is that the genetic values of animals can be estimated more accurately at younger age, therefore enabling faster selection. The use of DNA information also provides a manner to verify the breeding values of elite cows.

Italy is fully participating in the worldwide development and application of genomic selection. The objective of this paper is to describe the current status of the research in which Anafi, the Italian Holstein association, is involved.

Reference population

Since 1986, Anafi operates a national quarantine centre for Holstein progeny test candidate bulls. Since 2001, genetic material has been routinely collected from all these bulls. Older genetic material is partly available as well.

In Italy, various projects are active in the field of genomic selection. In this paper we will present a short description of the projects in which Anafi is currently involved. Genomic information in Italy will be available to Anafi and not locked within individual breeding companies, which is important, because Anafi is responsible for Holstein breeding value estimation.

From 2007-2010, a national research project is active in Italy in the field of molecular genetics. This project named SelMol is supported by the Italian ministry of agriculture with a multi-million euro budget. The project includes many organisations, in particular about a dozen universities and the Parco Tecnologico Padano research institute and has the support from AI centres and herd book organisations. The project works on dairy

cattle (Holstein, Brown Swiss, and Simmental), meat cattle (Piemontese and Marchigiana), buffalos, pigs, and sheep. The Illumina Bovine 50k technology is used in this project for Holstein, Brown Swiss, Simmental, Piemontese and Marchigiana cattle. The SelMol project plans to genotype 1000 to 1200 Italian progeny tested Holstein bulls. Genotyping for all breeds is nearly completed.

A second project is the ProZoo project, which runs from 2009-2013. This project works on Holstein cattle and pigs. This project is coordinated by the Parco Tecnologico Padano research institute. Anafi and some AI centres are participating in this project. The ProZoo project plans to genotype another 1000 to 1200 Italian progeny tested Holstein bulls. In September about half of these genotypes will become available.

The SelMol and ProZoo projects have established an agreement to share genotypic information. The participants of the projects will be able to use the information for scientific research, whereas breed associations will use the same information for starting up genomic selection.

Anafi has established a target list of 3700 bulls, which is the aim for our initial reference population. These are 3200 Italian progeny tested bulls, born since 2000 and with breeding values in April 2010, plus 500 bulls which are their sires and maternal plus paternal grandsires. Given that the first two projects will genotype around 2200 Holstein bulls in total, a third project, named Elica, is initiated to genotype the remaining part of the target list. The Elica project is a collaboration between Anafi and several Italian AI centres.

Italy is the third country in Europe in terms of Holstein progeny testing. Although by ourselves we can establish a competitive genomics program, international cooperation is an option that remains open.

Simulation

A simulation has been undertaken to assess the level of accuracies that can be obtained with

different numbers of bulls in the reference training population. In the simulation 54 000 SNPs and 100 genes were segregating in three pedigrees based on the actual population of bulls used in the Italian projects. The pedigree included up to 20 generations. The three sizes of the reference population, which were used roughly reflect the reference population after (1) the first project (SelMol), (2) the first two projects (SelMol and ProZoo) and (3) the total of three projects (SelMol, ProZoo and Elica). Phenotypes for three traits with high, medium and low heritability levels were simulated for 80 daughters per bull. On average 44 000 SNPs were selected and their effects were estimated with a GS-BLUP approach. The direct genomic values were then calculated for 400 bulls born between 1/9/2007 and 31/8/2008 (one year of progeny testing in Italy). The effective accuracies were calculated.

Table 1 shows accuracies that can be obtained with the first, first two and all three projects on Holstein genomic selection in Italy. Accuracies are shown for high, medium and low heritability traits.

Table 1. Simulation results for bull genomic value accuracies with three sizes of the training population and three heritability levels.

Bulls in training	High heritability ($h^2=0.40$)	Medium heritability ($h^2=0.25$)	Low heritability ($h^2=0.10$)
1034	40.1 %	37.5 %	30.8 %
2109	54.4 %	51.4 %	34.5 %
3727	67.0 %	62.9 %	41.0 %

Results of the simulation are comparable to results presented by VanRaden *et al.* (2008) and Meuwissen *et al.* (2001). As shown a good size reference population is quite important, even above 2000 bulls there is still a substantial increase in accuracies for all heritability levels.

Software development

Deployment of genomic selection requires a lot of new software. Anafi has deployed a web-based development platform using the Trac system. This system allows developers to collaborate even when not always physically in the same place. Currently software is available for simulation, deregression, SNP preparation, SNP checking, haplotype reconstruction, estimation of SNP effects, and estimation of direct genomic values. Software is developed primarily in Fortran 95 and Python. All necessary files and file formats were defined initially so developers have a clear route to follow.

In addition to internal development Anafi is also participating in the Clubware initiative for common development with other countries. Also the Clubware initiative has a web-based development platform using Redmine. Software is available for estimation of SNP effects using BLUP, BayesA, BayesB and FastBayesB, and for blending. Software is developed in Fortran 95.

Furthermore ideas and programs can be exchanged with researchers participating in the above national projects.

Genotypes available

At Anafi, we chose the following coding for SNPs. For each SNP, the alphabetically lowest nucleotide present is labelled allele 1, whereas the other allele is labelled allele 2. SNP genotypes were coded as:

- 0: Missing both alleles
- 1: Homozygous allele 1
- 2: Heterozygous with unknown phase
- 3: Homozygous allele 2
- 4: Heterozygous with phase 1
- 5: Heterozygous with phase 2
- 6: Allele 1 from sire, dam allele missing
- 7: Sire allele missing, allele 1 from dam
- 8: Allele 2 from sire, dam allele missing
- 9: Sire allele missing, allele 2 from dam

Genotypes 0-3 are output from the lab. Genotypes 4 and 5 result from deriving the phase of heterozygous genotypes. Genotypes 6-9 occur when single alleles are derived from relatives. This coding system is more extensive than the (0, 1, 2, 5) coding used by others and can therefore contain more information.

Table 2 shows the current status of proven bull genotyping. In September 2009, Anafi will have enough bull genotypes available to start internal genomic evaluations.

Table 2. Current status of proven bull genotyping per project based on the Illumina Bovine 50k technology. Note that still more bulls will be genotyped by both projects.

Breeds	SelMol	ProZoo
Holstein	887	570
Brown Swiss	759	
Simmental	488	
Piemontese	375	
Marchigiana	227	

First DNA results show a call rate of 99.60%. SNPs not resulting in genotypes account for 0.12%, whereas 11.50% of SNPs were monomorphic.

Conclusions

With the joint effort of all projects, Anafi will be ready to introduce genomic breeding values for Holstein cattle in 2010 in Italy. The size of the reference population is expected to be adequate to achieve good accuracy levels from the start.

Acknowledgements

The authors wish to acknowledge the support received from all projects and all people and organizations involved in the projects.

References

- Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157, 1819–1829.
- Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123, 218–223.
- Van Tassell, C.P., Smith, T.P.L., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C. & Sonstegard, T.S. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* 5, 247-252.
- VanRaden, P.M., Van Tassell, C., Wiggans, G., Sonstegard, T., O’Connell, J., Schnabel, B., Taylor, J. & Schenkel, F. 2008. Reliability of Genomic Predictions for North American Dairy Bulls. *ADSA/ASAS annual meeting*, Indianapolis, July 2008.