

Clubware – A Collaborative Genomic Breeding Value Estimation Software Development Platform

M. Coffey and R. Mrode

SAC, EGENES, Sir Stephen Watson Building, Bush Estate, Penicuik, EH26, 0PH, UK

Introduction

Genomics has developed rapidly culminating in the availability of dense marker (single nucleotide polymorphisms -SNP) panels at relatively low cost. These are being used extensively by animal breeding companies to more accurately and more quickly identify animals for entry into progeny testing or for selecting a team of young bulls for marketing. The organisations that have taken a lead in the use of SNP's have written their own software to estimate the SNP effects and are therefore able to compute genomic breeding values (Meuwissen *et al.*, 2001). The cost of such software and of genotyping bulls in sufficient numbers could be a barrier to the exploitation of this technology in some countries with less resource. Alternatively to avoid re-inventing the wheel, some countries may prefer a more collaborative way of working as the most efficient way forward. This is the founding philosophy for Clubware.

Proposition

The acquisition of knowledge used to develop software that solves animal breeding problems arises as a result of international research, scientific reporting, conference debates and collaboration between individuals or organisations. Given that this knowledge is exchanged freely in most instances, the proposition in this paper is that the software arising from the assimilation of knowledge relating to genomic selection could also be exchanged freely.

A number of people who share that principle found each other at a number of meetings and decided to meet formally to further that ideal. From that meeting, the idea of an International Club to jointly develop a suite of programs to enable the genetic evaluation of animals using a combination of

phenotypes and genotypes emerged. The first meeting was held at London Stansted airport on October 15th 2008 in order to establish what a club would do and whether there was sufficient interest and reason to continue (see Acknowledgements). The conclusion was that there was a case to continue and the group established the Clubware project as a platform to share development of software in a club based environment (as opposed to freeware). Minutes and notes from that meeting are available on www.genomicselection.net.

The project has been established initially to solve one problem as defined in diagram 1. This enables easier organisation and definition of success. However, if the project foundations are built sensibly then the same project may be the basis of other collaborations or an enduring project for collaborative animal breeding software development.

Of particular note is that the framework allows multiple solutions to the same problem and could provide a more efficient framework to adapt software to take up new scientific results over time. Integral to the proposed solution is a testing environment against which developed software can be audited and of course, a mark of assurance applied to those softwares that pass tests. This may be an ideal role for Interbull.

Results

The following results have been achieved by the Club so far:

1. An organising committee is operational
2. A web site for collaboration (www.genomicselection.net) has been established and is hosted by SLU
3. Individuals have agreed to lead on the writing of components

- a. Pre-processor and validation of input data – Ismo Stranden and Esa Mantysaari
- b. Solvers for the estimation of SNP effects – Donagh Berry, Jan-Thijs van Kaam, Theo Meuwissen
- c. Combination of genomic and traditional breeding values – Raph Mrode
- d. Test environment – Hossein Jorjani

The web site has a software version control function that allows the uploading of source files with tracked changes highlighted. The intention is for sub-projects to be created that encapsulate each separate function of the overall project (as defined in Diagram 1). The lead author will place the source in the sub project and others may then download, improve and upload the software. Thus each separate module will evolve over time as a result of many developers' input.

The site will contain documentation giving details of the communication protocols and format adopted by each module. Thus other developers can add modules that communicate with existing modules. Users of the site may propose new modules and that will initiate a new sub-project.

A module considered to be of high importance is the testing environment. It will be populated with simulated data that produces known answers. Thus new modules can be tested against this fixed data using documented tests. An idea being considered is that software could receive a quality mark from Interbull once it passes agreed tests. Users can then note that mark in refereed publications when the software is used.

Discussion

The current state of knowledge in genomic selection suggests that it may have an impact on genetic improvement. As such, the availability of software to exploit SNP data that is potentially available to all countries provides the most efficient way to utilise this technology. This effectively means that

competitive advantage is gained not through denying technology but through more rapidly exploiting widely available technology. This premise underpins Clubware.

The use of software on the Clubware site is subject to licence. It is not Freeware. Participation in Clubware gains rights to use software under the use of that particular licence. Use of software must be acknowledged in publications. The copyright will be retained by the original author. The licences are applied at module level i.e. some modules will be totally Open source whereas others will be compiled modules thereby rendering the original source unavailable. This is a decision of the original author but competition will determine the degree of restriction – an Open source version of a module may find a more widespread user base than a protected one. Of course the protected module may be substantially better and so could end up being used. That's competition for you.

The other major barrier to exploitation of genomic selection for any particular country is the cost of genotyping sufficient number of bulls that have daughters with phenotypic records in that country. On the assumption that genotype by environment exists, then it is necessary for each country to establish a reference population (Hayes *et al.*, 2009) in order to estimate SNP effects and therefore genomic values for their population. This will allow the calculation of genomic breeding values for imported semen bulls that have been tested in another country. To achieve higher accuracy for the prediction of genomic breeding values, adequate numbers of bulls (about 4000), should be genotyped (VanRaden *et al.*, 2008). Having 1000 bulls genotyped by 4 different countries and then traded internationally would reduce costs for all 4 countries. This is another area that Interbull could take a pivotal role because it could act as an independent third party for countries wanting to collaborate or trade genotypes.

Conclusion

A framework for joint development of software for genomic selection has been

established. Its future activity and scope is down to the user base and will be influenced, organised and developed by the international community that is Interbull. Its activity need not be restricted to dairy – Interbeef may be interested to use such software for genomic selection in beef. The framework has been established. Its future success depends on contributions by the international animal breeding community that shares its ideals.

This framework may be extended to the sharing of genotypes between countries that share common bulls. As a result, all barriers to uptake of this technology are removed and dairy farmers and society will begin to reap the benefits of a technology whose delivery has been a long time coming.

Acknowledgements

The efforts of the organising committee are acknowledged for bringing the project this far;

Esa Mantysaari, Ismo Strandén, Erik Boncam-Rudloff, Raphael Mrode, Donagh Berry, Jan-Thijs Van Kaam, Robin Thompson, John Woolliams, Theo Meuwissen, Martin Norling is thanked for setting up and maintaining the web site.

References

- Hayes, B.J., Bowman, P.J., Chamberlain, A.J. & Goddard, M.E. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92, 433-443.
- Meuwissen, T.H., Hayes, B. & Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819-1829.
- VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel R.D. & Schenkel, F. 2008. Genomic data and copoperation results in faster progress. Paper presented for *Interbull open meeting*, Niagara Falls, June, 2008.

Diagram 1. The problem to be solved.



