



Research Needs in Numerical Genomics and Quantitative Genetics of Livestock.

Summary from the SABRE Workshop 13 June 2006

This report was compiled by:

C.S. Haley - Roslin Institute

D.J. de Koning - Roslin Institute

J.M. Elsen - INRA

R.F. Veerkamp - ASG Lelystad

SABRE is an Integrated Project supported by funding under the 6th Research Framework Programme of the European Union. European Commission, Directorate E03 – Security of food production systems.



This publication represents the views of the Authors, not the EC.
The EC is not liable for any use that may be made of the information.



Food Quality and Safety

Research Needs in Numerical Genomics and Quantitative Genetics of Livestock:

Summary from the SABRE Workshop: 13 June 2006

Summary

This workshop was organised as part of Work Package 1 on Numerical Genomics of the SABRE Integrated Project. It was a satellite meeting to the Genesis Faraday “Symposium on High-Throughput Genotyping, Numerical Genomics and Quantitative Genetics in Animal Breeding”. The objective of the workshop was to review the need for new statistical and numerical methods and software for the analysis, interpretation and exploitation of genomics data in livestock. It was attended by 30 scientists drawn from the SABRE project or other international experts working in research or industry.

Major conclusions from the discussion were:

- Both gene identification and use of genome-wide selection without gene identification were part of the future of livestock research and application and hence needed the support of appropriate numerical methods.
- Better understanding and methods were needed for analysis of high-density SNP data.
- Methods should be developed that handle complex genetic systems including interactions between genes and with environmental factors.
- Methods were needed to enhance our abilities to identify mapped genes, including methods to integrate information from multiple sources.
- The livestock community could both learn from and contribute to the wider genetics research community.
- Optimising marker-assisted and genome-wide selection over a range of situations, such as partial genotyping, requires further study.
- More research is needed on the use of genomic tools to harness non-additive genetics effects.

- Application of genomic technology is limited by the need for good software. Big companies can patch something together themselves or in collaboration with research groups, but moderate sized companies do not have the skills or resources.
- Software tools needed have to cover use of genomic information in the context of breeding programme design and optimisation, prediction of genetic progress and breeding value estimation.
- Public sector research, including that in SABRE, would feed into this software development, but long-term support and development required that the software be commercially produced.

Introduction

The workshop was organised as part of Work Package 1: Numerical Genomics of the EC SABRE Integrated Project (FOOD-CT-2006-016250). It took place as a satellite meeting following the Genesis Faraday “Symposium on High-Throughput Genotyping, Numerical Genomics and Quantitative Genetics in Animal Breeding”.

The workshop programme covered:

- 1) The need for new methods and tools to facilitate the dissection and understanding of complex traits and the identification of underlying genes
- 2) The need for new methods and tools to apply genomic information in breeding programmes.
- 3) Issues in the implementation of new tools through software.

Prior to the discussion on methods and tools, the workshop debated whether the development of new approaches such as genome-wide selection made trait dissection and gene identification redundant.

The programme of the Workshop is shown at Annex 1 and of the Symposium is shown at Annex 2. Invited attendees at the workshop were: all SABRE work package 1 participants, all other work package leaders in SABRE and SABRE management, all presenters and chairs from the Symposium of the previous day. The attendees included representative from both the research community and industry, including

representative from the European Forum for Farm Animal Breeding (EFFAB) and from the Sustainable Farm Animal Breeding and Reproduction Technology Platform (FABRE). The list of the Workshop attendees is shown at Annex 3.

The Value of Trait Dissection

Traditional selection, which treats the genetic basis of complex traits as a “black box” and does not attempt to understand it beyond estimation of heritabilities and genetic correlations, has proven very successful. With the advent of genomic tools much effort has been expended on mapping QTLs and then studying these to higher and higher resolution using a variety of tools with the aim of identifying genes and nucleotide substitutions underlying variation. There have been a number of notable successes in this endeavour, and some genes so identified are being utilised in selection. However some groups have focused on selection using closely linked markers without trying to identify the precise gene or causative mutation. The advent of very dense SNP maps (e.g. now 10,000 in livestock with the prospect of moving to 500,000 or more) has led to the development of ideas on genome-wide selection which may be very effective without the need to identify the genes being selected. This provokes the question as to whether livestock genetic researchers still need to spend some energy on gene identification (and hence to develop tools for this purpose).

This debate produced no overall consensus, with cogent views being expressed both for and against the need to continue work on trait dissection. Overall, there was a strong view that genome-wide selection has great potential to revolutionise animal breeding without the need to identify specific genes. However, genome-wide selection was likely to be restricted to the few species with in-depth genomic information (i.e. chickens, cattle, pigs), particularly sequence data, and to more highly-managed breeding schemes, for some years. There were thus good reasons for continuing gene identification the foreseeable future, both for reasons of application and to improve biological understanding.

Arguments raised *in favour* of continuing gene identification were:

- i. to understand genetic variation, inheritance and changes in variance during selection

- ii. to understand genetic correlations and allow prediction of direct and correlated effects of selection and GxE
- iii. new traits like group selection require understanding and may not be amenable to the “black-box” approach
- iv. for utilising resistance to disease it is very important to understand the functioning of the gene
- v. information from the “black-box” does not extend outside animal breeding, or even outside the specific population – little general information is gained from it
- vi. our research sponsors and the general research community expect it and will benefit from this approach
- vii. to learn from and link in to other species/human/mouse
- viii. genomic selection is not for all species for the foreseeable future (apart from cattle, pigs and poultry)
- ix. for when genomic selection does not work

Arguments raised *against* continuing gene identification were:

- i. genetic variation changes and is variable – using individual genes is less responsive to these changes than genome-wide selection
- ii. dissecting all traits is impossible (already >40 in some breeding programmes)
- iii. return on investment is much higher from black box approaches
- iv. there are many individual genes to deal with
- v. QTLs effects differ between lines, so drilling down to the individual genes is a waste of time. The background might change the function.

Methods and Tools for Trait Dissection

Molecular tools to study livestock genomes are developing rapidly. Methods to analyse and interpret the data generated are also developing, but many are aimed explicitly at data from human populations. What are the current and predicted future gaps in our analytical methods in this area and to what extent can livestock geneticists harness methods developed in other areas – for the study of human data, for example?

One general point made is that phenotypic information is often a limiting factor now, and the development of high-density SNP information may increase the demand for high-quality phenotypic information. The development of a sophisticated electronic infrastructure may aid in the capture and retrieval of data (e.g. from slaughter houses or retailers). In addition, high-density SNP information may help reconstruct pedigree information where none has been recorded. Thus bioinformatic tools for data capture, storage and retrieval are an important part of the picture, and effective ways of reconstructing pedigree relationships from SNP data may also be of value.

Utilising information from Linkage Disequilibrium

The need for tools in this area will change with developing marker information. For example, use of SNP haplotypes may be more valuable when SNP information is relatively sparse (e.g. 10-50k SNPs) compared to when it is quite dense (e.g. 500k SNPs). Thus more complex approaches that allow one to infer haplotypes and utilise this information may be more valuable when there is less SNP information. Some lessons are being learnt from analyses underway in livestock and those that have been undertaken in other species, particularly humans. However it is clear that there is still much to be learnt and optimum methods of analysis are not yet well defined.

Experience from an analysis of data from 10k SNPs in cattle (Mike Goddard) suggested that use of haplotype information increased power to detect QTL, but not the precision of their mapping. Simulation studies from Jack Dekkers and co-workers also suggest relatively limited value in the use of multiple marker information. On the other hand, studies in mice (from the group of Jonathon Flint) suggested that QTL could only be detected and located with reasonable precision and power utilising haplotype information. These discrepancies may reflect different models underlying the association between markers and a causative polymorphism (for example, whether associations were a consequence of the original mutation, or to subsequent drift or

population admixture). In some cases a single SNP may be predictive of the genotype of the causative polymorphism, whereas in others it may require a haplotype of several SNPs to predict the causative polymorphism.

Resolving this and other questions will require analysis of sufficient livestock SNP scans to be undertaken and published to allow us to make general conclusions about the genetic architecture of LD in livestock. To further facilitate this, a number of areas for research and development can be suggested:

- i. Tools for high-throughput association analysis of individual SNPs in livestock data accounting for pedigree structure
- ii. Methods for haplotype reconstruction appropriate for livestock data with large families and complex pedigrees
- iii. Further evaluation of alternative approaches to the use of multiple SNP data.

Tools to Study Genetic Complexity:

Most attempts to dissect complex traits focus on additive effects of genes. However, well defined examples of more complex genetic architecture, including imprinting and epistasis (gene x gene interaction) are known. Opinion is divided on the importance of including these more complex effects in genetic analysis and this was reflected in the workshop. Reasons for allowing for complexity were:

- it is known to exist and found experimentally in crosses
- biological considerations suggest that the proper way to go about is to assume epistasis is everywhere
- knowledge of some known genes demonstrates presence of non-additive genetic inheritance
- it will not be possible to fully understand control of complex traits without its consideration

Reasons for being more cautious about the inclusion of complexity were:

- it is difficult to estimate
- large experimental data sets are required

- it is difficult to utilise in breeding programmes.

Overall the view was that methods should be developed to incorporate these effects, but that much of the genetic variation will be captured without them.

Experimental and Bioinformatic Issues

Handling large volumes of SNP data presents significant challenges. Identifying genotyping errors and inferring missing or incorrect genotypes were issues that were raised as problems. The software Genoprob by Mark Thalman was suggested as one solution.

The workshop discussed the possibility of putting a probability on the concordance of a polymorphism with a particular QTL allele. It was suggested that it might be possible to conclude that concordance was unlikely unless the polymorphism was actually causative. There was no clear resolution, but specific information such as that coming from the cattle Hapmap project would help quantify probabilities of concordance. The view was expressed that statistical concordance on its own could never be sufficient to prove the involvement of a particular polymorphism and that functional information would always be needed.

Methods for going from high-resolution mapping to gene identification were discussed. Tools for building pathways that put genes into a physiological context and link them to traits are of great value, but require more livestock species information than is currently available. One experimental approach is using gene expression data, for example in eQTL studies. A complementary approach is to examine the results of gene manipulations (e.g. gene knockout or RNAi mediated knockdown) on gene expression pathways and relating this to the effect of individual QTL. To screen many potential candidate genes in this way would require a high throughput system, probably *in vitro* based.

Livestock Community Needs and Contributions

Many tools and approaches are being developed by researchers studying the genomes of humans and model species. Tools developed in other research areas can be utilised for livestock, but to what extent are the needs of the livestock community unique and can they contribute to the needs of other research areas. Overall the view was that there is much to be gained from methods developed for human data (for example the

livestock community should take more advantage of coalescence models), but using some of the specific software could be problematic. Data from livestock may often be more challenging to analyse than that from humans for several reasons including large population sizes, large family sizes, pedigree complexity with loops and deep pedigrees. Thus in some cases software will need to be adapted or written from scratch. Livestock geneticists have greater experience on accounting for pedigree using mixed model approaches to contribute to human genetics.

Methods and Tools for Breeding

These methods cover the pipeline from data collection, collation and management, through programme design and progress prediction to breeding value estimation. Many of the issues discussed concerned the lack of software to implement genomic approaches in an effective manner (see next section). In addition a number of implementation/technical issues were raised, such as computational issues with large data sets. The more specific research questions that were raised were:

- i. Optimal use of genomic information in combination with pedigree and prediction of genetic progress
- ii. Identification of individuals to genotype to optimise information from genotyping a fraction of the population
- iii. Breeding value prediction when only a fraction of the population is genotyped or different animals have different densities of marker information.
- iv. The ability to recreate pedigree structure from genotype information (and whether there is the need to do so when using genome-wide selection approaches)
- v. Utilisation of marker information when breeding for non-additive effects in line crosses, etc.
- vi. So far software that advises on optimal breeding strategies, ignore genomics information.

- vii. Do you develop generic tools or be specific for a given company and generalise retrospectively?
- viii. Also it was agreed that the best way to develop tools is to provide an early draft to industry to play with, rather than make them wait for the finished product which turns out is doing the wrong thing.
- ix. Capacity is limiting: tools can just do the job currently but not quickly enough due largely to limits of the current software.

User Friendly Software for Numerical Genomics

Many good methods will not see wide application unless embedded in user-friendly software. The tradition in livestock genetics and breeding has largely been for individual scientists to distribute code to those that request it. However, individual scientists are not necessarily well placed to develop user-friendly software – it is not their main area of expertise or a major objective set by their employers. Resource limitations may also mean code developed by researchers may lack documentation and support, particularly in the longer term. Even widely used code, such as PEST, may be frozen due to lack of resource for further development. Thus application of genomics for both trait dissection and for breeding may be limited by lack of suitable software to implement it.

The advent of high-throughput genomics increases the challenge due to the sheer potential volume of the data. This is exacerbated by the increasing complexity of breeding objectives and the large number of traits that have to be considered in some breeding programmes. The major breeding companies may have the resource and expertise to put together a pipeline for analysis of genomic data and incorporation into breeding programmes, but even for them it is not easy. Large cooperative breeding structures may have the expertise to develop software, but can have limited room for manoeuvre due to the size and complexity of their breeding programmes. Moderate and small organisations do not have the resource to implement themselves and require outside assistance and off-the-shelf solutions.

The complete breeding programme package would include tools for data checking, storage, manipulation, etc, tools for breeding programme design, tools to predict progress and tools for breeding value estimation. The magnitude of the potential

difference in ethos and implementation of genome-wide selection *versus* use of individual genes and loci means that different tools are needed for these two approaches to breeding value estimation, at least in the first instance. Many of the individual tools are available for these steps from the efforts of individual groups and researchers. However they need to be streamlined, documented and packaged into a coherent framework.

The overall conclusion with respect to breeding software was that the development of the coherent package of breeding programme software was very necessary. However, this should be undertaken largely outside of academic and public research organisations. The role of SABRE and similar research projects was to develop the methodology and provide modules (software or published methods) which could be incorporated into coherent integrated frameworks by others. One such initiative to develop an integrated framework was already underway as a collaboration organised from the Netherlands. Such software would need to be at least partly commercially based and supported by payment for its use.

Acknowledgements

Input, ideas and suggestions from workshop participants is gratefully acknowledged.

The workshop was organised with the help of the Genesis Faraday team, in particular Alex Clop, and was supported by SABRE funded under EU Framework 6. Additional financial support came from Genesis Faraday and the BBSRC.

This report was compiled by C.S. Haley, D.J. de Koning, J.M. Elsen and R.F. Veerkamp.

Annex 1

Tuesday, June 13 – Workshop: Numerical Genomics in SABRE

Royal College of Surgeons of Edinburgh, Edinburgh

The meeting builds on the Symposium of the previous day. It will explore the need for methods and software in numerical genomics that will release the full potential the genomics revolution in livestock. Through discussion we hope to come to a broad consensus on the key targets for study in the short and medium term and the need for software for trait dissection and to implement genomic breeding. The workshop needs the expertise and views of all participants – please come prepared to contribute your opinions.

09.00 – WORKING GROUP 1. Research needs in numerical genomics and quantitative genetics of livestock.

Numerical genomics in SABRE

Using of high-throughput genomics data to dissect traits

Mapping, functional genomics, bioinformatics, modeling

The needs of research

What are the analytical bottle necks now? 5 years? 10 years?

What will be done for us (in human genetics), what must we do ourselves?

10.30 – Coffee break

11.00 - WORKING GROUP 2. Needs for breeding tools

High-throughput genomics in animal breeding

MAS, genome wide selection, diversity, pedigree control

The industry view

The challenges between theory and application

What needs to be done now? In 5 years, 10 years?

12.30 – Lunch

13.30 - WORKING GROUP 3. Developing the tools

Putting it into practise – characteristics of the ideal software.

Practical challenges.

Software development in the public and private sectors

What needs to be done now? In 5 years, 10 years?

14.45 – Remarks and conclusions by Prof. Chris Haley, Roslin Institute

15.00 – Workshop Close (Tea)

15.30 – SABRE WP1 Business meeting

Update on progress, implementing the lessons from the symposium.

17.00 – Business meeting close

Annex 2



Symposium on High-Throughput Genotyping, Numerical Genomics and Quantitative Genetics in Animal Breeding

Monday - Tuesday, June 12 - 13 2006

Moredun Research Institute, Edinburgh

This is a one and a half day Workshop in the framework of the SABRE project (cutting-edge genomics for **Sustainable Animal Breeding**). The aim of the Workshop is to help explore the potential for high-throughput low cost genotyping technologies to contribute to improved quantitative selection programmes. With this purpose, we will:

- a) explore the current status of these molecular technologies and forecast the possible future availability of low-cost high density genomic information
- b) review the relevant numerical genomics methods and analytical tools/software and define opportunities and needs at their interface.

The first day of this meeting is a symposium open to all SABRE partners and Genesis-Faraday Partnership (GFP) members. The workshop on day two are restricted to SABRE partners, invited contributors and GFP staff.

DRAFT Programme

Monday, June 12 – OPEN SYMPOSIUM

09.00 - Registration

09.50 - Welcome message by *Chris Warkup, Director of Genesis Faraday Partnership*

MORNING SESSION High-throughput low cost genotyping technologies

10.00 – New sequencing technology – *Dr. Jeff Schloss (NHGRI)*

10.35 – SNP technologies - *Dr. Panagiotis Deloukas (Sanger Institute)*

11.10 – Coffee break

11.40 – 4 billion bases a day – *Dr. Cameron Neylon (Southampton University)*

12.15 - DISCUSSION SESSION Forecasting the future costs of using high-throughput genotyping for research and industry

12.45 – Lunch

AFTERNOON SESSION Numerical genomics and approaches to incorporate genotyping information in the analysis of phenotypes

13.45 – Trends in quantitative genetics – *Prof. Mike Goddard (University of Melbourne)*

14.20 – Marker Assisted BLUP – *Prof. Rohan Fernando (Iowa State University)*

14.55 - Coffee break

15.25 – Genomic selection – *Prof. Theo Meuwissen (Norwegian University of Life Sciences)*

16.00 – Operational genetics – *Prof. John Woolliams (Roslin Institute)*

16.35 – Group selection – *Prof. Bill Muir (Purdue University)*

17.10 – Discussion and closing remarks of the symposium

19.30 – Meeting dinner

Annex 3 - Workshop Attendees

NAME	INSTITUTION
Joel Weller	ARO
Li Jiang	DIAS
Mogens Sando Lund	DIAS
Peter Sorensen	DIAS
Pascale Le Roy	INRA
Egbert Knol	IPG
Jan Merks	IPG
Roel Veerkamp	IPG
Johanna Vilkki	MTT Agrifood Res Finland
Simone Guimaraes	UFV
Mario Callus	WUR
Dr. Henri Heuven	WUR
DJ de Koning	RI
Anne Marie Neeteson	EFFAB/FABRE
Gary Evans	GENUS
Jeff Schloss	NHGRI
Mike Goddard	U Melbourne
Rohan Fernando	Iowa State U
Theo Meuwissen	Norway
John Woolliams	RI
Bill Muir	Purdue U
Beatriz Villanueva	SAC
Graham Plastow	Independent
Curt van Tassell	USA
Jean Michel Elsen	INRA
Chris Haley	RI
Santiago Avendano	Aviagen
Alex Clop	Genesis Faraday
Chris Warkup	Genesis Faraday