



Application of molecular information in sustainable animal breeding

Dirk-Jan de Koning

Division of Genetics and Genomics, Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Roslin, Midlothian, EH25 9PS UK

ABSTRACT - Livestock genomics is aimed at dissecting the genetic control of variation in economically important trait, such as disease resistance and product yield/quality. Unraveling the genetic control of such complex traits remains very challenging but farm animals are now well placed to bridge the gap between human biology and traditional model species. Livestock species share with model species the benefits of controlled breeding, while their biology is often much closer to that of humans. Livestock genetics can exploit the abundant genetic variation between divergent breeds as well as segregating variation within breeds, thus getting the best of both worlds. Large numbers of QTL have been detected for a variety of traits but for only a handful has the functional DNA mutation been discovered. The main challenge is how to exploit this information for sustainable animal breeding. The proposed applications for Marker Assisted Selection (MAS) vary from selecting on individual known mutations to using genome-wide SNP data for genome-wide selection. Molecular markers are also important tools to assess genetic variation within and between populations. Sustainable animal breeding should meet the needs of the current generation without compromising the needs of future generations. This contribution will discuss the status of molecular genetics in livestock and how this could support sustainable animal breeding.

Key Words: Animal Breeding, Functional Genomics, Marker Assisted Selection, Sustainable Agriculture

State of Play in Livestock Genomics

Many complex traits such as disease resistance, behaviour, and agricultural product quality are controlled by the actions and interactions of several or many quantitative trait loci (QTL) combined with environmental influences. Identifying the genetic and environmental factors that control variation and elucidating the biochemical and physiological pathways that connect these factors to the phenotype provides understanding of trait biology and would aid drug discovery and the development of personalised medicine. However, dissecting the genetic control of variation in complex traits and identifying underlying loci controlling such variation has proved to be very challenging. QTL detection by linkage mapping, using either existing genetic variation or creating such variation via controlled breeding, has been successful in identifying chromosomal regions associated with a wide range of complex traits in many different species. However, the relatively low accuracy of the linkage mapping means that mapped QTL may span a region of genome containing hundreds of potential positional candidate genes.

In genome-wide association or linkage disequilibrium studies, a population with thousands of individuals segregating for a trait of interest, or a diverse collection of inbred populations, is genotyped for very dense markers across the genome. In humans, this approach has been championed by the HapMap project and one recent achievement is the association between the FTO gene and obesity (Frayling et al., 2007). When collections of inbred lines are available, associations between traits and marker genotypes or haplotypes can be studied across lines as has been proposed for several species (Wade et al., 2002; Yu and Buckler, 2006; Zhao et al., 2007). The number of lines that need to be genotyped is much lower compared to the individual genotyping needed for a within-population association study. However, analyses of such experiments must account for potential confounding factors that may give spurious associations, such as genetic relationships among inbred lines.

The status of genome maps and genome sequencing in livestock species has recently been summarized (Womack, 2005). Linkage maps, physical maps and, more recently, limited QTL maps for nine livestock species and two aquaculture species can be

found at ArkDB (<http://www.thearkdb.org>) The availability of genome sequence information is crucial in pinpointing gene location and other conserved genome regions via comparative mapping. However, for complex trait dissection it is equally important to have the molecular techniques to characterise the genetic variation at the genome level for a population of interest. For the main livestock species, we now have extensive marker sets with large genome-wide SNP panels already available or rapidly coming onto the market. Also in terms of gene-expression microarrays, there is now a range of chips available for most livestock species. These tools have been successfully applied to QTL mapping studies in livestock. QTL mapping in livestock has focussed mainly on traits that are of economic importance such as efficiency of meat, milk and egg production, yield and quality of product, and more recently disease resistance. The principles of QTL mapping in livestock populations are outlined by Andersson and Georges (Andersson and Georges, 2004). To provide a comprehensive overview of livestock QTL mapping results as well as to facilitate comparative mapping of QTL between species, a database of livestock QTL was recently established (<http://www.animalgenome.org/QTLdb/>). The utility of databases like AnimalQTLdb is likely to improve further with ongoing efforts to define the “Minimal Information or QTLs and Association Studies” (MIQAS; <http://miqas.sourceforge.net/>). This initiative aims to provide the genome mapping equivalent of MIAME (currently the standard for microarray experiments) and was initiated by the livestock genome mapping community.

While the number of QTL detected in livestock species is quite impressive, the number of QTL for which the underlying causal genetic variant (QTN) has been identified is very small. In dairy cattle, a QTL on chromosome 14 affecting fat composition in milk was resolved to a mutation in *DGAT1* (Grisart et al., 2002). In pigs, a QTL with parent-of-origin effects on chromosome 2, affecting muscle depth in pigs, was resolved into a QTN in the imprinted *IGF2* locus (Van Laere et al., 2003). The causal single nucleotide polymorphism was identified by re-sequencing the target locus across pigs of known QTL genotype, and from a diverse range of genetic material (breeds). In sheep, the callipyge locus, resulting in muscular

hypertrophy in sheep, was resolved into a mutation in the novel *CLPG* gene (Freking et al., 2002). For callipyge, the mutation shows a very peculiar type of imprinting coined as ‘polar overdominance’. For this mutation, only individuals that are heterozygous for the mutation AND which have received the mutant allele through their sire show the callipyge phenotype. Livestock genomics combines genetic diversity with appropriate molecular and statistical tools to dissect QTL and identify QTN with complex gene action. An emerging lesson from livestock genomics is that regulatory mutations, as opposed to non-synonymous mutations that cause amino acid changes in proteins, are a major driving force behind variation in complex traits. While the number of identified functional mutations in livestock is small, their gene action has relevance for human genetics and biology in general.

No single species is sufficient for the dissection of complex traits that underpin most important variation in vertebrates. Combining information across several or many species is likely to be necessary to build a complete picture of the control of a particular complex trait. The natural advantages of livestock, combined with the rapid development of genomic tools and resources for these species, now make them a synergistic partner between studies of humans and those of rodents and other model species. The scene is set for livestock genomics to make ever greater contributions to our understanding of complex traits.

The Use of Molecular information in livestock Breeding I: Marker Assisted Selection

While clearly successful in academic research as a vehicle for QTL detection, the utilisation of these QTL in commercial breeding is limited in comparison to the number of QTL described. The technical aspects and potential implications of implementing MAS in livestock are discussed elsewhere (Dekkers, 2004). Adopting the terminology of Dekkers (Dekkers, 2004) there are three levels of MAS: gene assisted selection (GAS) where the functional mutation and its effects are known; linkage disequilibrium MAS (LD-MAS) where a marker (or marker haplotypes) is in population-wide disequilibrium with a QTL; and linkage equilibrium MAS (LE-MAS) where markers are in Hardy-Weinberg

equilibrium with the QTL at the population level, but linkage disequilibrium exists within families. A fourth type of MAS that was recently proposed is ‘genome-wide MAS’ (GW-MAS), where dense markers (i.e. SNPs) across the genome are used to predict the genetic merit of an individual without targeting any individual QTL or measuring (expensive) phenotypes on every generation (Meuwissen et al., 2001). Integrating current evaluations with MAS is most straightforward for GAS and LD-MAS because the QTL effect can be included in routine evaluations as a fixed effect. Linkage equilibrium-MAS, on the other hand, requires extensive genotyping and fairly complicated statistical procedures (Wang et al., 1998), while GW-MAS reduces the genome to a ‘black-box’ but does not require selection of QTL using arbitrary thresholds. Furthermore, the dense marker information required for GW-MAS may dispense with, often faulty, pedigree records because all pedigree information is encoded in the genome-wide genotypes.

The Use of Molecular information in livestock Breeding I: monitoring diversity between and within breeds.

Besides mapping QTL, molecular markers are very important for population genetics, in particular the management of genetic resources. The role of molecular markers in characterization of genetic variability is multifold (FAO, 2007): 1) Quantify genetic variation within and between populations, even in the absence of phenotypic and pedigree records. 2) In the absence of (reliable) pedigree data, genetic markers can be used as a measure of the effective population size (N_e). 3) Identify geographical locations of particular populations as well as admixture between populations. 4) Provide information on evolutionary relationships as well as areas of domestication and migration routes. 5) Identify family relationships (e.g. parentage) in the absence of pedigree data. For a more informed discussion of molecular tools for livestock conservation, the author recommends: “Utilisation and Conservation of farm animal Genetic Resources” (Oldenbroek, 2007).

Do we need molecular information for sustainable livestock production?

The author prefers a straightforward definition of sustainability: “*Meeting the needs of today’s generation*

without compromising the needs of future generations”. The demand for animal protein as a high quality food source is unlikely to decline in the face of global population growth and increasing wealth for developing countries. Coupled to the competition for land resources between ‘crops for fuel’ and ‘crops for food’, leading to increased cereal prices, one could argue that selection for increased production efficiency in livestock is firmly back on the agenda. At the same time, livestock production at industrial scales, may lead towards genetic ‘monocultures’ and erosion of farm animal genetic resources. Any form of genetic selection implies change in expected phenotypic performance as well as change in allele frequencies. It has been shown that direct application of ‘traditional’ BLUP selection favours the selection of relatives, thereby increasing inbreeding if no additional safeguards are in place. Using molecular markers, e.g. when selecting for favourable alleles of functional genes (gene-assisted selection or GAS), those genes will rapidly go to fixation, as well as a genomic region around these genes as a result of hitchhiking. Molecular markers can accelerate genetic progress but at the same time increase inbreeding and hence fail to contribute to sustainability. When, on the other hand, molecular markers are employed to monitor genetic diversity without additional selection, large investments are incurred with no short term economic benefit. Therefore, to warrant large investments in molecular technology for sustainable livestock production, this should address genetic improvement as well as genetic diversity. The emergence of high density SNP chips for many livestock species could actually facilitate such a joined-up strategy: In terms of genetic progress, genomic selection (Meuwissen et al., 2001) would not act very strongly on any individual locus but select on the total genome instead. Compared to traditional selection the level of inbreeding under genomics selection may be lower because genomic selection would act on the Mendelian sampling term, which is not exploited by BLUP selection. As a result, genomic selection can differentiate between offspring of elite parents and prevent the increased selection of relatives. Furthermore, the availability of dense genome-wide marker information provides to means to monitor levels of inbreeding on the basis of actual data rather than by pedigree-based prediction. At the population level,

genome-wide SNP panels provide insight into population history in terms of bottlenecks and admixture but also the phylogenetic relationships between populations and family relationships within a population. Therefore, genome-wide SNP chips are the potential tool of choice for both selection and conservation.

However, to date the promises of genomic selection are only delivered via *in-silico* prediction. While genomics selection is currently being implemented across the globe (in particular for cattle breeding), we may be waiting for quite a while before the actual effectiveness of this technology, especially in the longer term, is established.

Also, while the 0.05\$ genotype may be a fact, the sheer number of genotypes and animals to be genotyped means that the technology is still very expensive and only within the grasp of large companies and/or highly sponsored government initiatives.

For genome-wide selection and conservation to be implemented on a global scale several developments are required:

- 1) The accessibility and affordability of the technology needs to increase. Some platform are currently 'in-house' or only available to consortium members.
- 2) We need a global data infrastructure to share and access molecular data, in combination with performance traits. Again, this requires a re-think of how to approach intellectual property and how to protect company investments. Because all genetic relationships (both family and phylogenetic) are encoded in the molecular data, the recording structure does not require pedigree information, which is a major benefit.
- 3) "Dissemination Scenarios" must be developed on how to deliver the genetic progress to the developing world. One could imagine that genomics selection is practised in a local nucleus herd from which selected sires (and surplus dams) are sold/distributed to farmers. Given the unique local environmental challenges, it is important to maintain a local level of input into the selection programme.

- 4) For all this to happen, considerable funds need to be allocated both at national and international levels.

In summary, for molecular tools to make a positive contribution to sustainable livestock production we need a joined-up strategy addressing genetic progress as well as conservation, rather than piecemeal approaches addressing only part of the puzzle. The recent FAO publication on animal genetic resources for food and agriculture provides a good starting point for the development of such a strategy (FAO, 2007).

Acknowledgements

The author acknowledges financial support from the BBSRC.

Reference List

- ANDERSSON, L. and M. GEORGES. 2004. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat. Rev. Genet.* 5:202-212.
- DEKKERS, J. C. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.* 82 E-Suppl:E313-E328.
- FAO. 2007. The state of the world's animal genetic resources for food and agriculture. FAO, Rome.
- FRAYLING, T. M., N. J. TIMPSON, M. N. WEEDON, E. ZEGGINI, R. M. FREATHY, C. M. LINDGREN, J. R. PERRY, K. S. ELLIOTT, H. LANGO, N. W. RAYNER, B. SHIELDS, L. W. HARRIES, J. C. BARRETT, S. ELLARD, C. J. GROVES, B. KNIGHT, A. M. PATCH, A. R. NESS, S. EBRAHIM, D. A. LAWLOR, S. M. RING, Y. BEN-SHLOMO, M. R. JARVELIN, U. SOVIO, A. J. BENNETT, D. MELZER, L. FERRUCCI, R. J. LOOS, I. BARROSO, N. J. WAREHAM, F. KARPE, K. R. OWEN, L. R. CARDON, M. WALKER, G. A. HITMAN, C. N. PALMER, A. S. DONEY, A. D. MORRIS, G. VEY-SMITH, A. T. HATTERSLEY, and M. I. MCCARTHY. 2007. A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science*.
- FREKING, B. A., S. K. MURPHY, A. A. WYLIE, S. J. RHODES, J. W. KEELE, K. A. LEYMASTER, R. L. JIRTLE, and T. P. SMITH. 2002. Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res.* 12:1496-1506.
- GRISART, B., W. COPPIETERS, F. FARNIR, L. KARIM, C. FORD, P. BERZI, N. CAMBISANO, M. MNI, S. REID, P. SIMON, R. SPELMAN, M. GEORGES, and R. SNELL. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGATI gene with major effect on milk yield and composition. *Genome Res.* 12:222-231.

- MEUWISSEN, T. H., B. J. HAYES, and M. E. GODDARD. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829.
- Oldenbroek, K. 2007. *Utilisation and Conservation of farm animal Genetic Resources*. Wageningen Academic Publishers, Wageningen.
- VAN LAERE, A. S., M. NGUYEN, M. BRAUNSCHWEIG, C. NEZER, C. COLLETTE, L. MOREAU, A. L. ARCHIBALD, C. S. HALEY, N. BUYS, M. TALLY, G. ANDERSSON, M. GEORGES, and L. ANDERSSON. 2003. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* 425:832-836.
- WADE, C. M., E. J. KULBOKAS, III, A. W. KIRBY, M. C. ZODY, J. C. MULLIKIN, E. S. LANDER, K. LINDBLAD-TOH, and M. J. DALY. 2002. The mosaic structure of variation in the laboratory mouse genome. *Nature* 420:574-578.
- WANG, T., R. L. FERNANDO, and M. GROSSMAN. 1998. Genetic evaluation by best linear unbiased prediction using marker and trait information in a multibreed population. *Genetics* 148:507-515.
- WOMACK, J. E. 2005. Advances in livestock genomics: opening the barn door. *Genome Res.* 15:1699-1705.
- YU, J. and E. S. BUCKLER. 2006. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17:155-160.
- ZHAO, K., M. J. ARANZANA, S. KIM, C. LISTER, C. SHINDO, C. TANG, C. TOOMAJIAN, H. ZHENG, C. DEAN, P. MARJORAM, and M. NORDBORG. 2007. An Arabidopsis Example of Association Mapping in Structured Samples. *PLoS. Genet.* 3:e4.