Searching for genes of interest in sheep

Carole Moreno-Romieux Moreno, Jean Michel J. M. Elsen, Andres Legarra, Rachel Rupp, Jacques J. Bouix, Francis F. Barillet, Isabelle Palhiere Palhière, Helene H. Larroque, Daniel D. Allain, Dominique Francois, et al.

SEARCHING FOR GENES OF INTEREST IN SHEEP

Carole R Moreno[1], J.M. Elsen[1], A Legarra[1], R Rupp[1], J. Bouix[1], F. Barillet[1], I. Palhiere[1], H. Larroque[1], D. Allain[1], D. Francois[1], C. Robert[1], G. Tosser-Klopp[2], L. Bodin[1] and P. Mulsant[2]

[1]INRA, UR631, Station d'Amélioration Génétique des Animaux,
31320 Castanet-Tolosan, France
[2] UMR 444 INRA-ENVT– Génétique cellulaire, 31320 Castanet-Tolosan, France
carole.moreno@toulouse.inra.fr

SUMMARY

In the past, several studies have been performed to look for genes affecting traits of interest using microsatellite markers in sheep. Even if numerous chromosomal regions have been detected, few genes and causal mutations have been identified. The recent availability of high density SNP chip in sheep is a technological revolution for genomic researches and breeding selection. Firstly, genomic research in small ruminants will be boosted towards the identification of causative mutations underlying large genetic effects on sustainability traits, or the identification of very closely linked genetic markers that allow selecting genes in ovine breeding programmes. Secondly, such new molecular tools allow considering a new type of selection: the genomic selection. The principle of this selection is to use marker effects (without any information about the underlying genes) estimated within a part of the population which is phenotyped and genotyped (called the training population) and applying these effects to the rest of the genotyped population. However, the genomic selection might not always be a profitable strategy in sheep.

INTRODUCTION:

Classically, genetic improvement of sheep breeds has been obtained using pedigree-predicted breeding values and measured phenotypes on a large part of the population. In the last 15 years, several quantitative trait loci (QTL) studies have been carried out in sheep (http://sphinx.vet.unimelb.edu.au/QTLdb/). Numerous economically important traits, including wool, milk and meat production and/or quality, reproduction, behaviour and disease resistance traits, recorded on farm or in specialized experimental designs have been considered. Usually, whole genome scans were performed by genotyping 100 to 200 microsatellite markers on sire or grand-sire designs. However, due to the relatively low number of markers used, some QTLs were likely missed, and most of the discovered QTLs were located within large confidence intervals (20-100 cM). Such results would be technically sufficient to carry out marker-assisted selection (MAS) programmes, like, for instance, in the French dairy cattle breeds [Guillaume, et al. 2008]. However, such "1st generation" MAS programmes are unrealistic in sheep breeding, mainly for economical reasons. On the other hand, the identification of causal mutations having major impact on economically important traits (e.g. BMP15, myostatin or PrnP mutations) would be highly profitable to the sheep industry. The use of the newly developed 59,000 ovine SNP chip should allow an easier and quicker fine mapping of such genes but also allows performing a new type of selection, the so called genomic selection.

RECENT REVOLUTION IN MOLECULAR GENETIC: THE HIGH DENSITY SNP CHIP

In order to find chromosomal regions affecting traits of interest, molecular markers are used to scan animal genome. Before 2009, microsatellite markers were used but they have several drawbacks. There are less than 500 markers which are available today (http://www.ncbi.nlm.nih.gov/mapview/) but only a part of them can be used in a genotyping panel (in general 150 to 200 markers). The cost of the genotyping is proportional to the number of marker, and reach about 200€ as soon as 150 markers are used

Since January 2009, thanks to the sequencing efforts of the international sheep genomics consortium (ISGC consortium), 270,000 high quality single nucleotide polymorphisms (SNP) have been detected, distributed around the whole genome (http://www.sheephapmap.org/) have been detected. From this 270 000, 50,000 were selected to create a 50kovine SNP chip comprising equally spaced SNPs. The Illumina Ovine SNP50 BeadChip operating on the iSelect Infinium II™ technology from Illumina became recently available (http://www.illumina.com). More precisely, the BeadChip comprises 59,454 evenly spaced SNPs (average spacing 46Kb). This tool is a real revolution both in terms of cost and high number of markers. In fact, there are several hundred times more markers than in the microsatellite panel for a lower price.

Table 1: comparison of molecular tools: Microsatellite versus SNP panels

	Microsatellite panel	SNP50 Beadchip
Number of markers	150 to 200	59 454
Number of polymorphic alleles	2 to 20	2
Number of marker by cM	0.5	30
Location	Between genes	Within and between genes

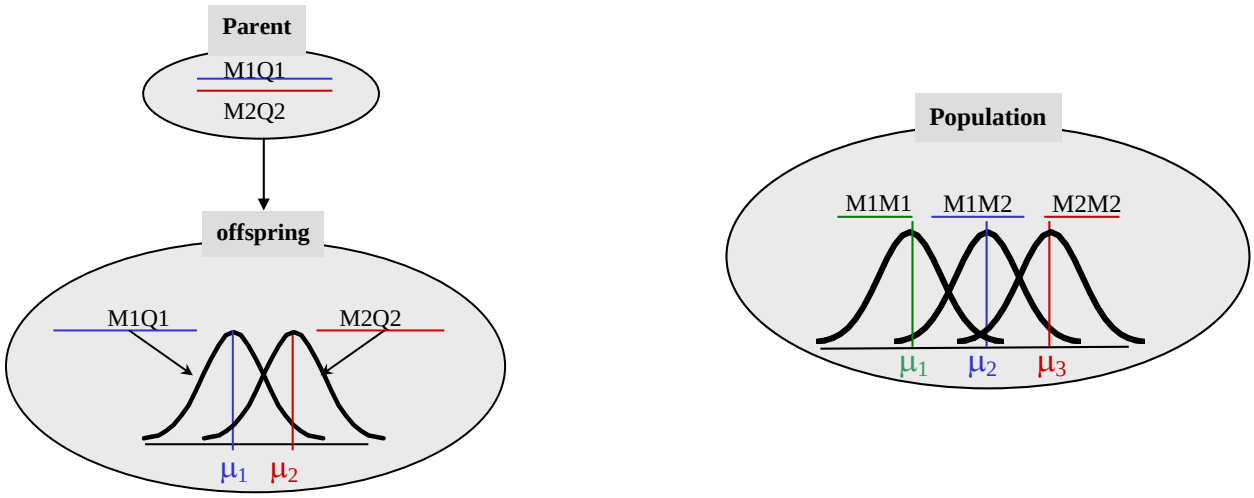
GENERAL PRINCIPLE OF QTL DETECTION

Linkage analyses. In sheep, QTL detection studies are classically performed using linkage analysis approaches (figure 1). Typically, a linkage analysis is performed using large half sib families. The general principle of linkage analysis is to observe the transmission of markers M between sire and progeny, assuming a QTL Q is located as in the vicinity of the marker M, i.e. with only a few recombinations between both loci. The higher the total number of genotyped markers is, the higher the probability of having a marker close enough to the QTL position is. Therefore, the use of SNP chip will allow the presence of a marker very close to any genome position, while microsatellite panels leave large areas of the genome without any marker. If the QTL Q exists, and assuming that the sire is heterozygous both at the QTL and at its flanking marker (with a genotype M1M2) a significant difference is observed between the offspring group which receives M1 from the sire and the group which receives M2 (figure 1). If not, both groups are similar. In this design and statistical approach, only heterozygote sires are used to detect QTLs. Consequently, a high number of families are needed for the detection of QTLs with low allelic frequencies. Moreover, a high number of progeny by family is needed to detect QTL with small effect.

Association studies. In the association studies, the picture of the population is taken without considering allele transmission from parents to progeny. It is supposed that any significant effect of a marker on the phenotype is explained by the presence of a tightly linked segregating QTL. In this approach, it is assumed that a QTL allele appeared by mutation within a specific marker haplotype in the past (often before the creation of breeds). From the

onset of the mutation to the study, there were generally several hundred of generations and thus many possible crossing-overs between the QTL Q and its flanking markers M. Consequently, a linkage between Q and M is only observed at a very short distance, and the only high density chips made the approach possible. A major disadvantage of this type of methods is the high proportion of false QTLs. In fact, a part of the observed differences between marker genotypes (M1M1, M1M2 and M2M2 in the example of figure 1) are due to random events of the population history and structure (*e.g.* family structures, bottle necks, selection, derive). Therefore, these methods allow detecting small chromosomal regions carrying QTL but a part of these QTL are mistakes.
In order to use the complementary advantages of association studies and linkages analyses, LDLA (Linkage Desequlibrium-Linkage Analysis) methods were created using simultaneously the family structure information and the history information.

Figure 1: Principles of linkage analysis and association study



M is the closest marker to the QTL Q. M1 and M2 are the observed polymorphisms at the marker M and Q1 and Q2 are the assumed polymorphisms at the QTL Q. For the linkage analysis, μ_1 and μ_2 are the means of the offspring which received M1 and M2 respectively from their parent. For the association study, μ_1, μ_2 and, μ_3 are the means of the populations with M1M1, M1M2 and M2M2 genotypes respectively.

IDENTIFICATION OF GENE AND CAUSAL MUTATION:

Despite the numerous and fruitful QTL studies performed in sheep, only a few causal mutations have been identified so far (table 2). In fact, most of the genes which were detected have a major effect and were evidenced following a candidate gene approach. Generally, the candidate gene approach was performed in chromosomic regions where QTL have been previously detected. In this approach, genes are studied when they have a known function related to the studied trait or when they were already demonstrated as controlling the trait of interest in other species, generally human and/or model species such as mice. The candidate gene is then sequenced to detect polymorphisms in a target population such as population with extreme phenotypes.
The Booroola gene has been localised with a wide genome scan undertaken on several families and finally discovered after linkage analysis of two very large families. The first mutation of the BMP15 gene was also found by a similar way. It has been studied in the

Lacaune population as a good candidate gene because its crucial role in the ovarian function had previously been demonstrated in sheep, mice and humans [Bodin, et al. 2007].
However, even after a powerful QTL study, the candidate gene approach is often a risky opportunity for several reasons. Firstly, the confidential interval of a QTL on the genome contains a large number of genes: 300 to 600 genes when microsatellite panels were used and hopefully 50 to 100 candidate genes using high throughput chip. Secondly, only a part of these genes have a known function. Consequently, there is often a "better" candidate gene waiting to be discovered. Thirdly, some causal mutations are located out of the coding sequence of the gene and plenty of time can be lost sequencing only the candidate gene while the mutation was so close. This misfortune occurred during the discovery of *GDF8* mutation. In fact, a mutation out of the gene creates a target site for microRNAs (miRNAs) that are highly expressed in skeletal muscle. This causes translational inhibition of the *GDF8* gene and hence contributes to the muscular hypertrophy of Texel sheep [Clop, et al. 2006].

Table 2: Identified causal mutations affecting production trait in sheep

Gene	Causal mutation	Affected trait(s)	references
PrP	*A136V, R154H, L141F,R171Q…*	Scrapie resistance	[Goldmann, et al. 2005; Hunter, et al. 1989; Moreno, et al. 2007]
BMPR1B	$FecB^{B}$ (booroola)	Ovulation rate	[Mulsant, et al. 2001; Souza, et al. 2001; Wilson, et al. 2001]
GDF9	$FecI^{T}$,$FecG^{H}$	Ovulation rate	[Hanrahan, et al. 2004; Nicol, et al. 2009]
BMP15	$FecX^{I}$, $FecX^{H}$, $FecX^{G}$, $FecX^{B}$,$FecX^{L}$, $FecX^{R}$	Ovulation rate	[Bodin, et al. 2007; Davis, et al. 2001; Galloway, et al. 2000; Hanrahan, et al. 2004]
GDF8	*G to A transition in the 3' UTR*	Muscle mass	[Clop, et al. 2006]
αs1-casein	A to E versus D	Casein content	[Barillet, et al. 2005]
β-lactoglobulin	A versus B	cheese-making properties; milk yield	[Barillet, et al. 2005]
ASIP	*Gene duplication*	White dominant color	[Norris and Whan 2008]
TYRP1	*G869 T transversion in exon IV*	Light fleece color	[Gratten, et al. 2007]
Hairless (hr)	*A739T, G823A and C1312T in exon III*	Hypotrichosis	[Finocchiaro, et al. 2003]

Recent advances in microarray technology, allow exploration of the expression of up to thousands of genes in the context of complex biological functions. Such an approach was used, for example to compare gene expression in sheep resistant or susceptible to gastrointestinal nematodes [Keane, et al. 2006]. The most recent tool available for sheep is a 15K oligonucleotide microarray by Agilent. Further techniques based on high throughput sequencing are also promising because they don't need any prior knowledge about the genes to study. Combined to specific genetic models, the use of such transcriptomic tools will certainly help to nominated functional candidate genes for QTLs or to better understand how the identified major genes act.

Even if today most causal mutations were found by candidate gene approaches, the evolution of sequencing and transcriptomic techniques will allow to study large DNA sequences of the genome and their effect on the genome expression in the near future. A European project '3SR' will use these invasive techniques to better understand the genetic basis of mastitis, parasitism and ovulation rate phenotypes in goat and sheep.

USE OF GENES IN SELECTION VERSUS GENOMIC SELECTION

Genes Selection. If causal mutations are known, the gene selection is the most efficient and easy way to select traits. In fact, no performance has to be collected and only tissue samples have to be taken for genotyping of causal mutation. This selection way is particularly interesting for traits difficult and expensive to measure such as behaviour traits or disease resistance (e.g. PrP gene to fight against Scrapie in sheep) or traits for which information are available late in life (longevity, fertility). However, the preliminary step of gene identification is performed for only a small group of traits (table 1). Moreover, because of the high efficiency of the gene selection, some precautions are necessary before using a causal mutation. The most important one is to estimate the effect of this mutation on the other traits of interest in order to escape an unintentional negative selection. For example, considering genes such as MHC in chicken, studies showed that the selection of resistant alleles for a disease is able to select susceptibility to other diseases. Finally, concerning disease resistance, because of the possible adaptation of the pathogen to the resistant host, the selection on several genes simultaneously or firstly on gene, and after on phenotype within "resistant" animals, is a less risky strategy than selection on a unique gene. The breakdown of resistance was previously observed during the selection of apple trees to *Venturia inaequalis* [Parisi, et al. 2004].

Genomic selection. The principle of genomic selection is to estimate simultaneously the effect of a large number of markers (54,000 markers in sheep) in a large phenotyped and genotyped population (minimum 1000 animals) so-called 'training population'. The estimation methods of SNP effects are related to association studies. During the following step, the genomic estimations are used in the rest of the breed to predict the genomic breeding values, only by marker genotyping and without phenotyping. This strategy is particularly attractive when generation interval is large or when progeny testing has to be performed (for example in milk breeds). Genomic evaluations are rapidly replacing traditional evaluation systems used for dairy cattle selection [Goddard and Hayes 2009]. Despite the widespread excitement about the potential for genomic selection to provide new approaches for the improvement of sustainability traits, in our opinion, there are a number of reasons to be cautious about this approach in small ruminant species.

Firstly, there is the challenge of many different breeds and production environments: whilst it is entirely feasible to 'train' genomic selection for Holstein dairy cattle where there are relatively few variations in production environments, very little is currently known about the 'portability' of genomic selection predictions across different breeds or environments. Early results for beef cattle indicate that denser SNP arrays than are currently available are likely to be needed [Kizilkaya, et al. 2010]. This type of investment is conceivable in beef cattle but it is more difficult in the case of small meat sheep breeds with the present chip costs.

This brings us to the second concern, viz. cost. This has two dimensions: 'training' of genomic selection requires phenotyping of significant numbers of animals and if this is required for each breed and major environmental type it will be prohibitively expensive. Furthermore, the economics of the application of the technology (even with optimistic costs will still be around 80€ per animal) is a challenge in species where (i) each individual has a relatively low economic value, (ii) the number of progeny produced by an improved animal is generally small due to small litter size and limited use of AI, (iii) current selection

programmes are relatively low cost and (iv) there are few cost savings or benefits from reduced generation intervals.
The situation is however different in French dairy sheep breeding schemes based on large open nuclei using extensively artificial insemination (AI) and progeny testing of several hundred of AI rams per year (*e.g.* in Lacaune or Manech red faced breeds) but also in Australia or New Zealand where very large wool and meat sheep populations are raised in a similar environment.
On an other hand, in meat sheep breeding schemes numerous evaluations of merit are performed on a small number of young males issued from elite parents, like growth, in-vivo body composition traits, even resistance to parasites by faecal egg count, or meat traits of young adults progeny tested. In these situations the potential benefits of genomic selection would appear of relative interest.

CONCLUSION

The availability of high density chip technology is a major revolution to study the genetic variability of traits of interest and to perform genomic selection as well. However, the molecular revolution continues with the closed availability of sequencing of the sheep genome, the decrease of the price of sequencing technology, the more and more global and less and less expensive gene expression tools …

REFERENCES

Barillet F, Arranz JJ, Carta A 2005. Mapping quantitative trait loci for milk production and genetic polymorphisms of milk proteins in dairy sheep. Genetics Selection Evolution 37 Suppl. 1, S109-123.

Bodin L, Di Pasquale E, Fabre S, Bontoux M, Monget P, Persani L, Mulsant P 2007. A novel mutation in the bone morphogenetic protein 15 gene causing defective protein secretion is associated with both increased ovulation rate and sterility in Lacaune sheep. Endocrinology 148(1), 393-400.

Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibe B, Bouix J, Caiment F, Elsen JM, Eychenne F and others. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nature Genetics 38(7), 813-818.

Davis GH, Bruce GD, Dodds KG 2001. Ovulation rate and litter size of prolific Inverdale (FecXI) and Hanna (FecXH) sheep. Proceedings Association for Advancement in Animal Breeding and Genetics, 14, 175-*178*.

Finocchiaro R, Portolano B, Damiani G, Caroli A, Budelli E, Bolla P, Pagnacco G 2003. The hairless (hr) gene is involved in the congenital hypotrichosis of Valle del Belice sheep. Genetetics Selection Evolution 35 Suppl 1, S147-156.

Galloway SM, McNatty KP, Cambridge LM, Laitinen MPE, Juengel JL, Jokiranta TS, McLaren RJ, Luiro K, Dodds KG, Montgomery GW 2000. Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. Nature Genetics 25(3), 279-283.

Goddard ME, Hayes BJ 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics 10(6), 381-391.

Goldmann W, Baylis M, Chihota C, Stevenson E, Hunter N 2005. Frequencies of PrP gene haplotypes in British sheep flocks and the implications for breeding programmes. Journal of Applied Microbiology 98(6), 1294-1302.

Gratten J, Beraldi D, Lowder BV, McRae AF, Visscher PM, Pemberton JM, Slate J 2007. Compelling evidence that a single nucleotide substitution in TYRP1 is responsible for coat-colour polymorphism in a free-living population of Soay sheep. Proceedings of the Royal Society B 274(1610), 619.

Guillaume F, Fritz S, Boichard D, Druet T 2008. Estimation by simulation of the efficiency of the French marker-assisted selection program in dairy cattle (Open Access publication). Genetics Selection Evolution 40, 91-102.

Hanrahan JP, Gregan SM, Mulsant P, Mullen M, Davis GH, Powell R, Galloway SM 2004. Mutations in the genes for oocyte-derived growth factors GDF9 and BMP15 are associated with both increased ovulation rate and sterility in Cambridge and Belclare sheep (Ovis aries). Biology of Reproduction 70(4), 900.

Hunter N, Foster JD, Dickinson AG, Hope J 1989. Linkage of the gene for the scrapie-associated fibril protein (PrP) to the Sip gene in Cheviot sheep. Veteterinary Record 124(14), 364-366.

Keane OM, Zadissa A, Wilson T, Hyndman DL, Greer GJ, Baird DB, McCulloch AF, Crawford AM, McEwan JC 2006. Gene expression profiling of Naïve sheep genetically resistant and susceptible to gastrointestinal nematodes. BMC Genomics 7(1), 42.

Kizilkaya K, Fernando RL, Garrick DJ 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. Journal of Animal Science 88(2), 544.

Moreno CR, Moazami-Goudarzi K, Laurent P, Cazeau G, Andreoletti O, Chadi S, Elsen JM, Calavas D 2007. Which PrP haplotypes in a French sheep population are the most susceptible to atypical scrapie? Archive of Virology 152(6), 1229-1232.

Mulsant P, Lecerf F, Fabre S, Schibler L, Monget P, Lanneluc I, Pisselet C, Riquet J, Monniaux D, Callebaut I 2001. Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Merino ewes. Proceedings of the National Academy of Sciences 98(9), 5104.

Nicol L, Bishop S, Pong-Wong R, Bendixen C, Holm LE, Rhind S, McNeilly A 2009. Homozygosity for a single base pair mutation in the oocyte-specific GDF9 gene results in sterility in Thoka sheep. Reproduction 138, 921. DOI: 10.1530/REP-09-0193

Norris BJ, Whan VA 2008. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. Genome Research 18(8), 1282.

Parisi L, Fouillet V, Schouten HJ, Groenwold R, Laurens FND, Didelot F, Evans K, Fischer C, Gennari F, Kemp H 2004. Variability of the pathogenicity of Venturia inaequalis in Europe.

Souza CJ, MacDougall C, Campbell BK, McNeilly AS, Baird DT. 2001. The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic receptor type 1 B (BMPR1B) gene. Journal of Endocrinology 169(2), R1.

Wilson T, Wu XY, Juengel JL, Ross IK, Lumsden JM, Lord EA, Dodds KG, Walling GA, McEwan JC, O'Connell AR 2001. Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells. Biology of Reproduction 64(4), 1225.