final report

Project code: **BSC.0083**

Prepared by: **Hutton Oddy**

University of New England

Date published: **December 2011**

ISBN: **9781741917550**

PUBLISHED BY
Meat & Livestock Australia Limited
Locked Bag 991
NORTH SYDNEY NSW 2059

Sheep Genome research tool development and a strategic framework for capturing benefits

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

Abstract

Australia and the international community invested in the development of sheep genomics resources to provide the knowledge, technical capability and human resources to enable effective R & D in genomic solutions for the sheep industry. The task of producing genomic resources for sheep was too big for any one country to do alone. The International Sheep Genomics Consortium (ISGC) was formed to co-ordinate international efforts to obtain genomics resources for sheep.

The value of such co-operation was exemplified by the development and release of a 50k ovine SNP chip in January 2009, of which more then 35000 chips have been used around the world. In March 2011 the ISGC made available the first sheep genome reference sequence and has since completed a haplotype map of more than 3000 sheep from 74 breeds. These resources underpin future R & D into sheep biology and genetics and ultimately delivery of genome tools to the sheep industry. Australia is the major user of the resources developed by the ISGC.

We have formed a Ruminant Genome Biology Consortium to share data and experience about how to build and use genomic resources for different domesticated ruminant species.

The primary driver for demand for new genomic information will be on-going R & D into the biology of sheep traits. Testing of new ideas which use genomics tools such as genomic Breeding Values for difficult to measure traits, SNP derived pedigree markers and markers for some single locus traits will inform the value of genomics tools to industry.

The important role played by MLA (and AWI) in fostering global collaboration in sheep genome resource development should continue. It is not yet clear the extent to which there is a need to invest in development of a new high density sheep SNP chip. It may be that technological improvements will mean that the best way to capture the information in the sheep genome sequence for industry will be through re-sequencing. Only time and on-going R & D will reveal the best strategy.

Executive Summary

Over the past decade, MLA, AWI, the Commonwealth Government and a number of Australian research agencies (CSIRO, Universities) have invested, together with international partners (USA, China, UK, NZ and EU) in the development of sheep genomics resources. The purpose of the Australian investments was to provide the knowledge, technical capability and human resources to enable effective R & D in genomic solutions for the sheep industry.

Developing the resources required for such research is clearly an undertaking of international as well as national significance. The International Sheep Genomics Consortium (ISGC) was formed in 2003 to co-ordinate international efforts to obtain genomics resources for sheep. It was recognised that the task of producing genomic resources for sheep was too big for any one country to do alone and that there were considerable cost savings to each country to be made by combining resources within the global sheep research community. The value of such co-operation was exemplified by the development and release of the 50k ovine SNP chip in January 2009. Since that time more than 35000 ovine 50k SNP chips have been used around the world, with heavy use in Australia and New Zealand.

In the second half of 2009 the ISGC commenced the development of a reference genome sequence. The expectation was that the reference sequence would deliver a high quality and invaluable resource to underpin future sheep research and benchmark other resources and outputs including the delivery of additional high fidelity genomic tools for industry use.

The first public assembly of the sheep reference genome was made available in March 2011. This was a joint work between the ISGC and Chinese scientists working with the Beijing Genomics Institute (BGI, Shenzhen) and the Kunming Institute of Biology. Scientists, and funding agencies, from Australia, UK, France, Germany, Denmark (EU),USA, New Zealand and China contributed to this work. It is available at :- http://www.livestockgenomics.csiro.au/sheep/oar2.0.php

Following this, details of the genetic structure of the global sheep population, including information about the mixture of sheep during domestication and more recent events such as breed formation has been uncovered. The Haplotype Mapping paper (which describes genome structure of more than 3000 sheep of 74 breeds) has been accepted for publication in PLoS Biology and will appear in early 2012.

There is growing interest in comparative biology of ruminants, particularly around the basis of their adaptations to markedly different environments, size and ability to be domesticated. We have formed a Ruminant Genome Biology Consortium with the object of sharing learnings about how to build and use genomic resources for different domesticated ruminant species (currently :- Sheep, Cow, Goat, Alpaca, several species of Deer, Buffalo and Bison) and some wild ruminants with interesting adaptations (Tibetan Antelope, Moose). The diversity of adaptations in the ruminant clade (adapted to pre-digestion of forage organic matter) may well lead to a better understanding of how to more effectively manipulate traits in domestic animals.

The important role played by MLA and AWI in fostering an environment in which successful global collaboration in sheep genome resource development is well recognised. The extent to which this needs to continue is not yet clear. The availability of a sheep reference genome (which will continue to be improved) and knowledge of breed diversity in sheep will assist understanding of current activities to genotype sheep with the 50k SNP chip and other products of genome research (SNP

tests for pedigree, single marker tests for agouti, horn / poll, muscling – myostatin etc). Although the re-sequencing of more than 100 sheep will generate the knowledge, it is not clear that this will lead to a high density ovine SNP chip. In part this is because it is not clear that the technology to re-sequence genomes will become so cheap as to make construction of a high density chip unnecessary. The other aspect is that the global need for a genotyping resource that drove development of the 50k ovine SNP chip has not yet been established for a SNP chip with ~ 1 million markers.

What is clear is that ISGC has done a good job at providing the co-ordination to cost effectively build sheep genomics resources. It would be remiss of organisations like MLA (and AWI) to permit the ISGC to fold without a clear understanding of the value of applied genomics to the Australian sheep industry. At present there are a number of trials underway to establish the value of genomics tools for estimation of breeding values of hard to measure traits. There is also a need to understand the biology which lead to the different traits, and it is here the research tools developed by the ISGC will have their major short term application. Without ongoing research into biology of sheep, genomic tool development will probably wither. However, both ongoing R & D into basic biology and continued attempts to use that knowledge in practical fields such as animal breeding will signal the need for further resources.

As for a strategy for MLA (and AWI) in this space, perhaps the wisest course is to watch and wait. It is unlikely in the short term that new resources to support development of gEBVs will need to be developed, but almost certainly they will be in the longer term. By watching this space, regularly updating knowledge and awareness of managers and the industry about developments and possibilities in genomics generally, and acting to ensure that some elements of international coordination continue, options will remain open and the good work to date will not be lost.

Contents

1) Background

Over the past decade, MLA, AWI, the Commonwealth Government and a number of research agencies (CSIRO, Universities) have invested, together with international partners (USA, UK, NZ the EU, and China) in the development of sheep genomics resources. The resources include “libraries” of sheep DNA sequence, maps of marker positions on the physical genome, a “virtual” and draft sheep genome sequence, genotyping (SNP) chips and the human capability to exploit genomic information in the Australian sheep industry. The purpose of investments in sheep genomics resources was to provide the knowledge, technical capability and human resources to enable effective R & D in genomic solutions for the sheep industry. This is clearly an undertaking of international as well as national significance. The International Sheep Genomics Consortium (ISGC) was formed in 2003 to co-ordinate international efforts to obtain genomics resources for sheep.

In the second half of 2009 the ISGC commenced the development of a reference genome of the sheep. The expectation was that the reference sequence would deliver a high quality and invaluable resource to underpin future sheep research and benchmark other resources and outputs including the delivery of additional high fidelity genomic tools for industry use. In February 2010 the BGI (Shenzhen) (formerly the Beijing Genome Institute) and Kunming Institute of Zoology announced they were close to completing a next-generation sequencing of a sheep, a goat and an antelope genome. Although their method would yield a lower quality genome sequence than the then current ISGC international project, their was a risk that their sequence would reach press well ahead of the ISGC sequence, thus demoting the subsequent ISGC publications to much lower ranking journals and impact. The University of New England funded Dr Hutton Oddy to attend a workshop called by the Chinese project team and he was able to negotiate with Prof Wen Wang of Kunming Institute of Zoology and Prof Jun Wang, Executive Director of BGI that BGI, Kunming and ISGC join forces to develop a combined comprehensive sheep genome paper with ISGC as lead authors, to be published in the highest level international scientific journal(s).

At that time it was anticipated that a high quality sheep genome sequence paper would be partnered by a paper on the sheep haplotype map. The haplotype map paper would provide new information on impact of domestication and past selection on the genome structure of a wide number of sheep breeds around the world. These publications provide unprecedented open access to resources for research and development into different aspects of sheep biology around the world. Potential applications arising from development of the sheep reference genome sequence range from improved tools for genetic selection to identification of the biology contributing to important traits and subsequent non-genetic applications (which range from diagnostic tools to therapeutic applications).

It is important that this work is seen as the product of far sighted, diligent and well managed investments by, among others, the Australian sheep industry. The publication of this work provides an opportunity for MLA to advertise their contributions to these internationally important research outcomes. Continued dialog between the scientific community responsible for publishing the papers, past and current investors and the sheep industry is important to maximise the impact of this work on the wider community.

The technology base is moving so quickly, at ever reducing cost, and the animal/phenotype resources for potential exploitation accumulating at an ever increasing rate that the opportunities for finding new ways to use genomic information are both great and not immediately obvious. In such

an environment, it is essential to have a clear strategy to guide our response to future challenges and shape the future in a way that brings positive outcomes to the Australian Sheep industry.

To guide that strategy requires a mapping of the science and technology landscape (including the people required to use and understand such technology) and an overview of industry opportunities that this technology may present in the foreseeable future. Ideally, the current sheep industry framework for investment in genetic improvement can be integrated with this mapping to take account of the opportunities and risks presented by the rapid change in the technology landscape. From an industry perspective, any technology mapping exercise must address the implications for industry application. It is also important that during the mapping process and subsequent development of a strategy that we retain engagement of all potential players in the science, translation and practice of the technology and its outcomes.

The work described here goes some way to addressing the needs outlined above. It describes the successful completion by the ISGC of the first reference genome of a sheep, briefly touches on the work carried out within the Haplotype Mapping component of the ISGC and presents information / views that can inform a high level strategy to about Australia's ongoing contributions in this space. The author is aware of a similar MLA supported strategy mapping process underway by Dr Rob Woolaston.

2) Project Objectives

1. Coordinate the interaction between ISGC and Chinese collaborators to ensure submission of a high value sheep genome publication by third quarter 2010.

2. Represent Australian industry needs within ISGC discussions and planning with aim to maximise value of ISGC outputs to Australia (examples include discussions about design of a reduced SNP assay and opportunities to divert resources within existing USDA and other grants to new activities).

3. Maintain Australian profile as a lead country within ISGC (notwithstanding small funding commitments).

4. Maximise opportunity for MLA to obtain high-profile PR from release of sheep genome and related publications.

5. Develop a broad assessment of future opportunities for sheep genome research and applications of relevance to Australia.

MILESTONES:

1. Assume the role of Strategy Leader and Coordinator of ISGC and provide leadership and oversight to the ISGC collaboration, promote the activities of the ISGC to ensure that Australian interests are properly addressed and recognised. Guide the redirection of investments in the USDA grant and the funds held at Baylor College of Medicine to obtain

maximum impact for Australia from our work with the KIZ/BGI group, providing guidance and recommendations on development of a reduced SNP assay for sheep. [15 June 2010]

Ongoing.

2. Plan and facilitate a workshop to establish an international consortium for Sheep Genome Applications at the ISAG meeting in Edinburgh in July.[31 July 2010]

   Completed July 2010

3. Facilitate co-ordinated publication of the sheep genome sequence and sheep haplotype map papers. Provide guidance and specific advice (including draft material for press releases) in a timely fashion to ensure that past investors (MLA and AWI) receive appropriate recognition on release of the sheep genome and HapMap publications. [30 Sept 2010]

   Sheep Genome reference sequence publically available March 2011. HapMap paper submitted for publication May 2011. Accepted December 2011.

4. Upon publication of sheep genome embark on the activities necessary to maximise the downstream benefits to the international sheep community. Work with the Australian and international sheep community to map the future framework for sheep genomics activities and applications and establish a rational basis for future investments in this space within Australia.

3) Results and Discussion

a) Completion and publication of the first sheep reference genome sequence.

In March 2011 the first reference genome sequence of any mammalian species to be constructed primarily from next generation (i.e. short read) sequencing technologies was made available on the web. http://www.livestockgenomics.csiro.au/sheep/oar2.0.php

Oar v2.0 was built using next generation sequence derived from one female and one male Texel. The primary *de novo* assembly was performed using 75 fold Illumina GA sequence from the female Texel. Mate pair characteristics of the paired end reads were used to produce scaffolds spanning 2.71 Gb or approximately 91% of the sheep genome (scaffold build v1.2). A further 45 fold coverage of the male Texel was used for gap filling, before scaffolds were anchored onto the 27 sheep chromosomes. Scaffolds that were clearly chimeric were identified by comparison with the bovine UMD3 assembly and manually split in the gap between adjacent contigs mapped to two different bovine chromosomes. Superscaffolds were built from the set of scaffolds and split scaffolds >2 kb in length using the end sequences derived from the male Texel BAC library, CHOR-243 and the predicted locations on OARv1 of SNPs included on the Illumina Ovine SNP50 BeadChip. This was undertaken as a single integrated process and non-congruent BACs and out of position SNPs were minimised. Several rounds of manual checking and final error correction were carried out using the end sequences of the BACs in the bovine CHORI-240 library and 454 mate pair sequence data

derived from 8kb and 20kb insert libraries of the male Texel. Ambiguous positions were resolved using the predicted location of the SNPs based on OARv1.0 and conserved synteny with the UMD3 bovine genome assembly. Superscaffolds were initially ordered and oriented into chromosomes using the locations of the SNP in OARv1.0. The positions of the SNPs in the sheep linkage map and the sheep RH map were used to identify remaining errors and to refine the assembly.

The female sheep on which the reference genome is based is a Texel ewe from a Danish producer. Sequence from the Male Texel used to generate the CHORI-243 high copy number BAC library was also incorporated into the reference genome sequence.

The reference genome sequence resulted from close collaboration between the International Sheep Genomics Consortium and scientists at Beijing Genomics Institute (BGI, Shenzhen) and the Kunming Institute of Zoology. Sequencing of the female Texel was done at BGI (Shenzhen) using Illumina GA2 methodology. Illumina GA 2 sequence of the male Texel was performed at the Roslin Research Institute Edinburgh. Initial assembly of contigs was performed using SOAPdenovo (Short Oligonucleotide Analysis Package http://soap.genomics.org.cn/soapdenovo.html), then refined as described above. Data for construction of the Radiation Hybrid (RH) Panel order was obtained from genotyping the INRA RH Panel with the Illumina Ovine SNP50 chip. A RH Map was constructed by the team lead by Dr Thomas Faraut at INRA Toulouse, France. Dr Jillian Maddox (University of Melbourne) provided a refined linkage map using data from genotyping the International Mapping Flock and the FMFS / sheepGENOMICS flock with the Illumina Ovine SNP50 chip. These maps were a direct output of the ISGC, were essential tools to cross check assembled sequence fragment (contig) order during initial assembly and eventually were subsumed into the final reference genome sequence.

At present (December 2011) there is no published paper describing the procedures used to generate the sheep reference genome (although the procedures used are described in http://www.livestockgenomics.csiro.au/sheep/oar2.0.php).

b) Completion and Publication of the Sheep Haplotype Map.

The ISGC co-ordinated the collection of DNA samples from more than 3000 animals of 74 diverse sheep breeds (details of sample structure http://www.sheephapmap.org/hapmap.php) and arranged for genotyping using the Illumina Ovine SNP50 beadchip. The resulting genotype information has been used to construct a Haplotype Map of sheep. An earlier publication using this population, using a prototype SNP chip with 2470 markers, provided useful information on breed diversity (Kijas et al, 2009). The map developed using the output from the 50kSNP chip is beginning to provide insights into the historic mixture of domesticated sheep and clues as to the contributions of different ancestors to modern sheep breeds. The data reveal that the majority of sheep populations have a higher effective population size than most cattle or dog breeds. An initial scan for selection sweep signals revealed at least 30 regions containing genes for coat pigmentation, skeletal morphology, body size, growth, reproduction and absence of horns.

The HapMap data generated by the ISGC is available for download for research purposes from

http://www.sheephapmap.org/download.php

and the initial analysis has been accepted for publication in PLoS Biology (Kijas *et al*, 2012 see Bibliography).

c) Development of a Ruminant Genome Biology Consortium.

At the meeting called by BGI (Shenzhen) and Kunming Institute of Zoology in March 2010, two major topics were discussed.

1. Completion of the sheep reference genome

2. Development of an international consortium of scientists to further the work on sequencing genomes of ruminants with a particular emphasis on developing a process for studying the comparative genomics of ruminants.

I undertook the task of forming a Ruminant Genome Biology Consortium to address the co-ordination of ruminant genome projects and especially the co-operative analysis of the data generated. In August 2010 I attended ISAG and held an out of session meeting in which more the 60 scientists indicated their willingness to participate in a Ruminant Genome Biology Consortium (RGBC). On return from ISAG I developed a framework for a website for the RGBC and negotiated its hosting in the animal genome group at Ohio State University. The prototype site is available to view http://www.animalgenome.org/ruminants/ and is improved from time to time when new material becomes available. The nature of the consortium is outlined on that site.

I prepared a draft paper and circulated to co-authors to announce the RGBC to the scientific community, but as yet that has not progressed to submission.

d) Looking Forward – planned activities with ISGC and implications for future applications

i) Communications.

The ISGC presented 3 scientific papers at the Plant Animal Genome Conference (San Diego, USA) in January. These were presented by a) John McEwan on behalf of Brian Dalrymple “Progress towards assembly of the sheep reference genome” (W137) b) Noelle Cockett “Physical mapping and the genome assembly of sheep” (W133) c) Yu Jiang “Sequencing and assembly of the sheep genome reference sequence (W138) and d) Chunhua Wu “An update on the sheep RH map” P552.

MLA and AWI were informed of an intended press release to coincide with the announcement of the assembly of the sheep reference genome at PAG, but this did not proceed because at the last minute Brian Dalrymple was unable to attend due to the floods in Brisbane, and there were uncertainties about the positioning of some contigs. The floods subsided and the contig positions were subsequently resolved and the ISGC put the reference genome up on the web in early March 2011. There has been no formal press release, but there is an announcement on the ISGC website (http://www.sheephapmap.org). The sequence has been submitted to NCBI and will be available on their website later this year.

ii) Next steps for the ISGC.

Following the release of the sheep reference genome in March, the ISGC focussed on completion of the HapMap paper (accepted by PloS Biology) and moved onto the next phase of activity (see below).

Next major ISGC Project.
The availability of a reference genome assembly for sheep (OARv2.0) has paved the way for a whole genome re-sequencing project, with the following objectives:

(i) identification of 30 million SNP using a genetically diverse sample of domestic sheep genomes

(ii) design of a high density SNP chip

iii) characterisation and annotation of structural variation including copy number variation

iv) provision of whole genome sequences to improve the accuracy of genomic selection

The conduct of this study was discussed by the ISGC at the Plant Animal Genome meeting in January 2011. It received in-principle agreement to proceed. At that meeting, John McEwan indicated that although 5 million SNPS were identified across the two Texel sequences, these SNPs would probably not be useful by themselves to create a high density SNP chip or equivalent. Therefore, he proposed that a selected group of the HapMap animals be sequenced at 1X for 50 animals and 10X for 10 animals who represented major ”clades”. It is anticipated that the remaining funds (approx $200K) at the Baylor College of Medicine Human Genome Resource Center could be used for this.

The design of the re-sequencing study has been completed and samples sent to the Human Genome Sequencing Centre. It is anticipated that sequencing will commence early 2012, with data available by mid 2012.

A rebuild of Oar2.0 is underway and it is anticipated this will be publically released as Oar3.0 at the Plant Animal Genome Meeting in January 2012. Plans to submit a paper outlining the steps required to build, and the quality of the build, of a sheep reference sequence have been deferred until the release of Oar3.0. This is in part to meet the needs of the European Union 3SR project co-funding the post-Doc (Dr Yu Jiang) working on the project.

iii) Implications of access to sheep reference genome sequence and outputs from ongoing re-sequencing projects for end users and potential future investment

The public release of a sheep reference genome sequence, has resulted in corrections to the positions of SNP on the existing 50k sheep SNP chip. This has improved our capacity to construct haplotypes from genotyped populations (incorrect SNP order generates a discontinuity in the haplotype map). An immediate outcome of this is to improve haplotype reconstruction and to increase the accuracy of any inferences from genomic selection. It facilitates accurate imputation of genome sequence in related individuals.

Availability of the reference genome facilitates re-sequencing of many more animals because it permits rapid assembly of their genomes. This can enable genome wide SNP discovery on a much larger scale than previously possible. It also makes it practical to re-sequence key individuals from any chosen population. Coupled with lower density genotyping (50k or less) these methodologies permit accurate imputation of genotype and reconstruction of population structure. There is increasing evidence that variation in a considerable number of traits is associated with variation in copy number (CNV) of genomic regions. The ability to rapidly re-sequence whole genomes increases the likelihood of identifying regions where CNV is associated with traits. It may also be possible to design chips that inform of CNVs as well as SNP genotypes. By having access to a large number of SNPs (derived from the re-sequencing of individuals of many breeds) it is possible also to generate much larger SNP chips than presently in use.

However, the best strategy to utilise this additional information about genome structure is not yet clear. For example:

The 50K SNP Chip was made possible because

a) a need to genotype 1000's of phenotyped animals from various studies and from sheep industries around the world

b) more than 500,000 SNPs were available from which to select a suitable panel

c) willingness for members of the ISGC to work together to make it feasible for a commercial provider to enter into production of a chip, and

d) the ISGC co-ordinated purchase orders to assist the manufacturers to assess demand, and facilitated communication during the design and testing phase.

This process resulted in the manufacture and use of more than 35,000 ovine 50kSNP chips, which have been used over the past 3 years.

With the advent of a reference genome sequence, some corrections to SNP order have been made. As noted above this has facilitated more accurate haplotype construction. In theory this should make calculation of gEBVs more robust.

The ISGC is currently re-sequencing (to a depth of 10-12X) more than 100 sheep that reflect the diversity of the global population (to enable design of a High Density ~1000000 SNP Chip), to provide specific information on single locus disorders and to provide sequence for a number of breeds in commercial use internationally.

It is difficult to foresee if demand for a high density sheep SNP chip will ever be sufficient to encourage commercial manufacture. With the cost of sequencing continuing to fall, it is possible that re-sequencing of key individuals will become the norm, followed by genotyping of their relatives with the current 50K SNP chip (or a tool with a smaller number of SNP if this is economical).

The ISGC have confronted these issues, and recognise that they will be predominantly market and technology driven. Accordingly, current ISGC activities to provide high quality sequence and assembly in the public domain is recognised as enabling / knowledge building. Implementation

awaits specific applications, which are subject to competitive forces with sheep industries around the world, or arise from opportunities which emerge in parallel industries / applications.

Implications for future investments by MLA in sheep genomic resources.

Background.

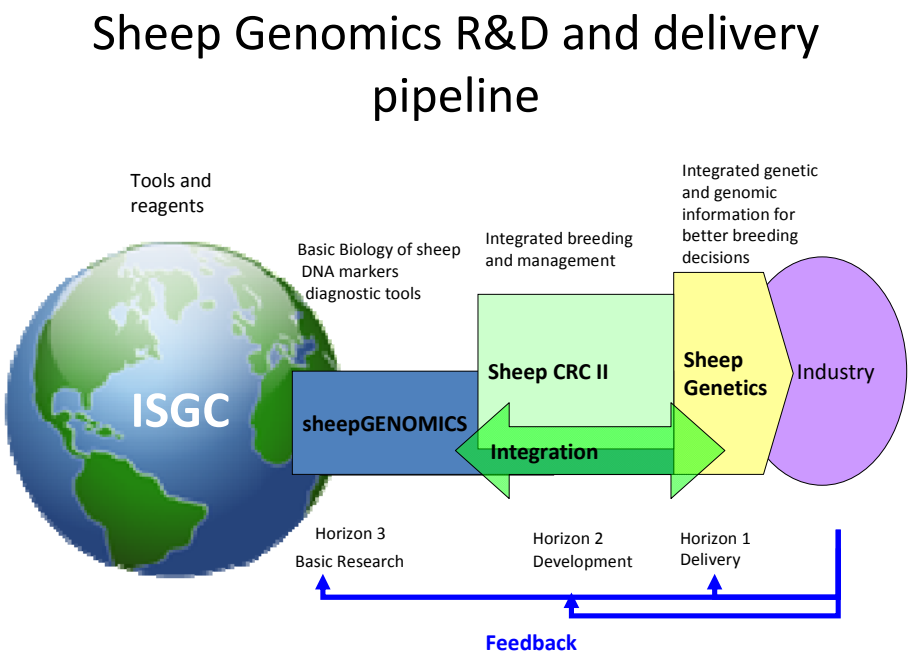

Figure 1. The role of the different participants in sheep genomics research, development and delivery to industry (from Oddy *et al* 2007). The key feature of this landscape is the key role of the ISGC in provision of genomics resources (tools and reagents – from the underpinning knowledge of genome structure to SNP chips and their implications). The practical activities arising from access to sheep genomics tools and reagents range from basic research into sheep biology (once conducted by sheepGENOMICS), translational research into near to market products such as genomic EBVs and pedigree markers (currently conducted within the Sheep CRC) and delivery to industry through Sheep Genetics.

An overview of the integration of the global effort to generate tools and reagents for sheep genomics and the co-ordinated Australian efforts to apply these tools is shown in Figure 1. The primary Australian users of sheep genomics tools (derived from the ISGC) are the Sheep CRC (and associated projects including the Reducing Emissions from Livestock program), Sheep Genetics (through various pilot projects conducted through the Sheep CRC). In addition some specific studies

on aspects of biology using sheep as a model organism continue. For example, A/Prof David Groth (UWA and Edith Cowan University) has been sequencing the major histocompatability complex (MHC) region of the sheep genome. Some who were involved with the sheepGENOMICS program are continuing to use the resources for follow up research on muscle development (e.g. Ross Tellam, CSIRO, LI) and as part of the ISGC HapMap Project (James Kijas CSIRO and Ben Hayes DPIVic). These latter activities are to my knowledge not currently funded by the Australian sheep industry.

The core activities of sheepGENOMICS, which relied on access to genomics tools and knowledge (understanding host resistance to parasites, wool biology and regulation of muscle growth) are not now directly supported by the Australian sheep industry. Instead, near to practice applications of sheep genome resources are focused on estimation of genomic breeding values for hard to measure traits currently used within industry, and for new traits measured within the Information Nucleus flock of the Sheep CRC.

Traits measured by the Sheep CRC are, with the exception of a few nutrient content traits of meat, similar to those measured in the FMFS Mapping Flock established by MLA and AWI in the sheepGENOMICS program. The data from both have been combined into parallel databases and are currently being managed through the Sheep CRC, but will return to Sheep Genetics at the end of the Sheep CRC. Pilot trials to test genomic breeding values (gEBVs) determined in the sheepGENOMICS and Sheep CRC research flocks have been conducted. These have relied exclusively on genotypes derived from the Illumina ovine SNP50 chip developed by the ISGC. Although there has been some activity to generate a reduced (cheaper, 3 - 7k SNP) chip to facilitate imputation of genotypes. To my knowledge this has not yet found widespread use. Current R & D is directed to development of computational tools to derive breeding values from genomic information, further collection of phenotypes and genotypes and development of small scale genotyping tools for specific purposes (e.g. parentage tests).

In Australia, for a period, the community working with practical use of the ovine SNP50 chip were largely unconnected from the community that developed the 50k SNP chip, the HapMap, tools for determining pedigree from SNPs and the reference sequence. This was not optimal, and is only now being addressed, through projects run through the Sheep CRC using industry animals. These now include testing of new SNP pedigree tools and genotyping of industry animals as precursors to release of gEBVs for some traits.

International.

Use of genomics tools for selective breeding of industry animals is being explored in both commercial and research settings in New Zealand. The elements of such programs are being assembled throughout Europe. In China, there is also interest in use of transgenics.

Many of the traits being targeted are associated with resistance to disease, and increased reproductive performance together with reduced environmental footprint of the sheep industry. Improving production traits per se seems to be less of an immediate goal. It is worth noting that the global sheep population has been falling since the early 1990s, and that the 2 largest exporters of sheep meat (Australia and New Zealand) have maintained production from a considerably reduced flock. At the same time, the world production of wool has fallen. This suggests that future uses of genomics tools for the sheep industry will increasingly be for sustainable production of food rather

than fibre. Reproductive efficiency is now more important than it has been previously. The use of genomics tools is well suited to improvement of sex linked, and difficult to measure traits, Moreover, there is already good evidence for a number of single gene effects on reproductive performance of sheep. We anticipate that with access to a reference genome, discovery of causative associations with single locus traits will be easier than in the past.

Other issues.

Maintaining capability remains an issue. The information about sheep reference genome sequence is now publically available, there are only a small number of Australian scientists who understand how to use such information. There is ongoing effort in supporting the development of genomic assisted breeding values (gEBVs), but this is just one subset of what is possible.

With reduced funding for use of genomics derived tools (gene expression, proteomics) in sheep it is less likely that new scientists will be attracted into the area to conduct industry specific R & D. The model where the sheepGENOMICS program carried out basic R & D alongside a Sheep CRC focussed on development of new sheep genetic parameters and joint translational research into genomic tools (Figure 1) that was present at the inception of the Sheep CRC has long since vanished. This is unfortunate in that it has resulted in lost opportunity and reduced capacity of the sheep industry to harness diverse points of view to deliver the value of genomic and genetic R & D.

Importing practical outcomes from other countries is not likely to be a fruitful exercise. This is because the populations of sheep in each of the major sheep producing countries has a different genetic base (at least in recent times). It would be anticipated that associations between traits and genome regions will be to some extent unique to the populations (and more so in the case of gEBVs derived from imputation over a relatively small number of generations). An illustration of this is the high frequency of the g<6723G<A myostatin rearrangement in Australian Texel sheep compared to NZ Texels.

This means that there is no substitute for translational R & D in Australia before application to industry. It suggests that maintaining a base of scientists and skills in the field of sheep genomics is an appropriate response to retain capacity to respond to new scientific advances and new industry challenges.

Summary of factors which impact on formulation of a strategic response.

The section above has outlined many of the factors which will affect the immediate future for R & D and application of genomic tools. It was intended to provide a background for thinking about a strategic response. Below is a more explicit listing of current and short term future activities that complete the background on which a strategic response can be framed.

Points 1-5 are from an Australian perspective alone, points 6-9 are made knowing what the ISGC and others are planning to do around the world.

Australia.

1. Work will continue to collect phenotypes on the progeny of the Information Nucleus flock both within the Sheep CRC and beyond. At least some of these progeny will be genotyped using the 50k SNP chip.

2. In the short term (for at least the next 2 years) it is unlikely that a higher density sheep SNP chip will be manufactured because:

    a. It will be necessary to complete the complete the evaluation of data collected (in Australia within the Sheep CRC, in NZ within their industry and R & D program) before deciding what the limitations are with current tools. Although it is true that a chip with more markers would be particularly useful for Merinos because LD is shorter than in terminal and maternal sire breeds, it remains unclear that with only relatively recent pedigree information that this is a serious constraint to inferring genotype and associations with hard ot measure phenotypes.

    b. It is unlikely that there will be sufficient global interest (in the next 2 years) to form a consortium to build a high density chip to drive down costs.

3. A smaller (3 - 7,000 SNP) chip may well be used to assist with inference of genotype within closely related animals (although the now small difference in cost between genotyping using the 50k SNP chip and the smaller chips may make this less of an option).

4. Some key sires identified by industry or by Sheep CRC phenotyping will be sequenced to a limited depth (say 10x) to provide a resource for imputation of genotypes (using either the 50k or 3-7k SNP chip).

5. The (Australian) scientists who lead the assembly of the sheep reference genome sequence are currently not funded from Australian sources to work on sheep. There is limited funding from international consortia to provide Post Doctoral support for assembly of Oar3.0 (the next version of the sheep reference genome) and assembly of sequence from the re-sequencing project. This will become a serious constraint to ISGC activities when sheep re-sequence data becomes available in 2012 (see below),

International.

6. The ISGC will conduct a re-sequencing study of selected individuals representing diverse breeds to generate a comprehensive suite of SNPs and provide explicit knowledge of the extent of variation in copy number (CNV) of genomic regions. It is anticipated that the information from this will be available mid to late 2012.

7. Following the re-sequencing the ISGC will work to position the SNPs and CNVs onto the sheep genome reference sequence. It is anticipated that the availability of additional sequence will lead to repositioning of some parts of the sheep reference genome and development of a comprehensive map of variation in the sheep genome. This will enable rational design of a new high density sheep SNP chip (if required) and also provide some insight into evolution of sheep genome sequence compared to other ruminants.

8. The ISGC will remain the group of choice to co-ordinate international investment in development of new genotyping and gene expression tools for sheep. If there is sufficient international interest, the ISGC is the right group to co-ordinate the construction of a high density SNP Chip.

9. The work required to re-assemble many sheep genomes and match them to the reference sequence (point 7 above) is currently under funded. A small amount of funding from a USDA and European Union 3SR grant is available. Some support from within Australia (either directly from CSIRO or jointly with RDCs or other Commonwealth grants) may need to be secured to guarantee that the work is conducted to the standards used to generate the reference genome. This work is of national and international significance and rightfully fits within basic science support funding (e.g that provided previously through the Commonwealth Governments International Science Linkage Grant Scheme).

The above background is intended to make it relatively simple to see the strategic responses (below).

Potential Strategic Response (by MLA and AWI on behalf of the Australian sheep industry).

Given the background discussion, the assumptions above, and the desire of the Australian sheep industry to remain at the forefront of sheep genomics R & D and delivery, the following strategy is proposed to ensure that genomics resources continue to be developed in a timely manner and not constrain potential applications.

1. MLA (and AWI) continue to support the goals of the ISGC (i.e. to develop sheep genomics resources) but at arms length and recognise that resources generated by the ISGC will continue to be delivered into the public domain.

This suggests that the key strategic response is to watch and be aware of the forces driving development of genomic resources and their potential use in industry.

2. Operate at arms length in the development of sheep genomics resources.

Specifically :-

a) Foster co-operation with other national and international agencies to modify the funding environment in a timely manner to ensure that resources are available when required. A successful example of this mode of influence at the international level has been through discussions at the Tri-Nations Lamb Working Group (Australia, New Zealand USA). At the national level it has been through past support of applications through Commonwealth funding opportunities such as the International Science Linkage grants. Where specific opportunities arose, it was through direct negotiation with AgResearch and Meat and Wool New Zealand (now Beef and Lamb NZ).

b) Only step in (i.e. provide funds) where it is apparent that no other player can support the development of resources required by the Australian industry, and the process is at risk of failure.

The processes used to step in (if required) should be nimble i.e. quick when needed and not subject to detailed evaluation at the time of action (I would strongly suggest a process of ongoing evaluation of possibilities to enable this to happen).

Implement a policy to maintain a close watch on the activities of the ISGC with respect to the implications for current activities within Australia. The purpose of the watching brief policy it for MLA and AWI to quickly intervene to prevent catastrophic failure (or grasp an otherwise unseen opportunity) as and when required.

3. Act to increase awareness among diverse users of sheep genomics resources in Australia of International activities to secure / develop genomics resources. This will make implementation of 2 above easier. The simplest way to do this is to encourage members of the ISGC to publish in a timely manner and to support some investment in R & D into underpinning biology of sheep. The latter will provide new insights into possibilities that can inform practical applications and identify missing resources in a timely fashion. It is my view that without an active R & D program in the field of sheep biology / genomics it is unlikely that the impetus to continue development of sheep genomics resources can be sustained at an international level. It is also unlikely given the cost of generating such resources that any one country can do it alone (although costs are falling).

4) Success in Achieving Objectives

All project objectives were achieved. Specifically :

4.1 Coordinate the interaction between ISGC and Chinese collaborators to ensure submission of a high value sheep genome publication by third quarter 2010.

Completed in March 2011. The current version of the reference sheep genome is available here

http://www.livestockgenomics.csiro.au/sheep/oar2.0.php.

Work is underway to improve the sequence (fill gaps, re-order small segments) with anticipated release of oar3.0 in January 2012. Publically available annotation (placement of gene names) will be done for oar3.0.

4.2 Represent Australian industry needs within ISGC discussions and planning with aim to maximise value of ISGC outputs to Australia (examples include discussions about design of a reduced SNP assay and opportunities to divert resources within existing USDA and other grants to new activities).

Achieved. Assisted with rearrangement of contract with Baylor College of Medicine to ensure that work required for development of next generation SNP chip will be possible. Continue to discuss ways to maximise advantage from co-funding from USDA and European funds.

4.3 Maintain Australian profile as a lead country within ISGC (notwithstanding small funding commitments).

Achieved. Australia still retains a lead role in the ISGC. Australia played a significant role in initiating the Ruminant Genome Biology Consortium, and will continue to do so as that consortium develops.

4.4 Maximise opportunity for MLA to obtain high-profile PR from release of sheep genome and related publications.

We have ensured that MLA and AWI obtained due recognition for their investments in the ISGC. Prepared press release and provided to MLA and AWI, although we are still awaiting CSIROs preferred method for publically releasing the story behind development of the sheep reference genome.

4.5 Develop a broad assessment of future opportunities for sheep genome research and applications of relevance to Australia.

See this report.

5) Impact on Meat and Livestock Industry – now & in five years time

5.1 Impact on Meat and Livestock Industry – now & in five years time

The impact of access to worlds best resources in sheep genomics is dependent on downstream use within the Australian R & D community generally and the sheep industry. Access to the 50k SNP Chip and flow on products such as the pedigree SNP test is enabling the Sheep CRC and Sheep Genetics to explore use of genomic breeding values (gEBVs). If these tests are able to deliver more than current methodology (from phenotype and pedigree information alone) then industry will obtain value in terms if improved livestock for breeding. It is too early to speculate on outcomes from improved knowledge of the sheep genome sequence.

6) Conclusions and Recommendations

6.1 Conclusions and Recommendations

6.1.1 **Conclusions**

The ISGC has done a remarkable job at securing and developing the resources required to conduct world class genomics R & D for the sheep industry, including the ground work that enabled the production of a 50kSNP Chip. The ISGC have recently completed a reference genome sequence of the sheep and made that publically available, and published a Haplotype Map of more than 3000 sheep from 74 breeds around the world. This provides valuable information about the origin of current sheep breeds and of selection sweeps for particular traits. The ISGC have embarked on re-sequencing more than 100 animals to provide the knowledge to generate a high density SNP chip if required by the international sheep community, and to provide deeper knowledge of genome structure.

It is unlikely that the full value of these resources will be extracted without ongoing R & D into understanding the biological basis of sheep phenotypes. Nonetheless, some of the resources

developed by the ISGC are providing significant new information to the Australian sheep industry in the form of gEBVS and SNP based parentage tests.

The work of the ISGC has been an example of how an essential resource can be cost – effectively developed through international collaboration. MLA and AWI are to be commended for timely provision of support to enable the ISGC to function, and the wisdom to watch and provide oversight, and when required, resources to facilitate action, rather than become actively involved in the work of the ISGC.

6.1.2 **Recommendations**

The ISGC will continue while there is a need to develop new sheep genomics resources or improve existing ones. The primary driver for demand for new genomic information will be on-going R & D into the biology of sheep traits. Testing of new ideas which use genomics tools such as gEBVs for difficult to measure traits, SNP derived pedigree markers and markers for some single locus traits will inform the value of genomics tools to industry.

MLA (and AWI) should continue to support the development of genomics tools for the sheep. It is not yet clear the extent to which there is a need to invest in development of a new high density sheep SNP chip. It may be that technological improvements will mean that the best way to capture the information in the sheep genome sequence for industry will be through re-sequencing. Only time and on-going R & D will reveal the best strategy.

Nonetheless, MLA (and AWI) should watch activities in the ISGC closely and act to ensure it continues while there is a need for sheep genome resources. It is desirable that this action be through influence (e.g. through the Tri-Nations lamb group) as well as targeted investment.

7) Bibliography

Recent publications from the ISGC.

Archibald AL, Cockett NE, Dalrymple BP, Faraut T, Kijas JW, Maddox JF, McEwan JC, Hutton Oddy V, Raadsma HW, Wade C, Wang J, Wang W, Xun X. (2010) The sheep genome reference sequence: a work in progress. International Sheep Genomics Consortium,
Anim Genet. 41(5):449-53.

Becker D, Tetens J, Brunner A, Bürstel D, Ganter M, Kijas J; International Sheep Genomics Consortium, Drögemüller C (2010). Microphthalmia in Texel sheep is associated with a missense mutation in the paired-like homeodomain 3 (PITX3) gene. PLoS One. 5(1):e8689

Dominik, S., Henshall J.M. Hayes B.J. (2011) Identification of a SNP on sheep chromosome 10 which is highly predictive of Poll in Australian Merino. Animal Genetics (In Press)
García-Gámez E, Reverter A, Whan V, McWilliam SM, Arranz JJ; International Sheep Genomics Consortium, Kijas J. (2011) Using regulatory and epistatic networks to extend the findings of a genome scan: identifying the gene drivers of pigmentation in merino sheep. PLoS One. 6(6):e21158

Goldammer T, Di Meo GP, Lühken G, Drögemüller C, Wu CH, Kijas J, Dalrymple BP, Nicholas FW, Maddox JF, Iannuzzi L, Cockett NE. (2009) Molecular cytogenetics and gene mapping in sheep (Ovis aries, 2n = 54) Cytogenet Genome Res. 126(1-2):63-76.

Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, Tang D, McEwan J, Cockett N, Oddy VH, Nicholas FW, Raadsma H; International Sheep Genomics Consortium. (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One. 4(3):e4668.

Kijas, JW, Lenstra, JA, Hayes, B, Boitard, S, Porto Neto, LR., San Cristobal, M., Savin, B., McCulloch, R, Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., Dalrymple, B, and other members of the ISGC (2012) Genome wide analysis of the world's sheep breeds reveal high levels of historic mixture and strong recent selection. PLoS Biology (in press).

Magee DA, Park SD, Scraggs E, Murphy AM, Doherty ML, Kijas JW; International Sheep Genomics Consortium, MacHugh DE. (2010)Technical note: High fidelity of whole-genome amplified sheep (Ovis aries) deoxyribonucleic acid using a high-density single nucleotide polymorphism array-based genotyping platform.
J Anim Sci. 88(10):3183-6.

Miller JM, Poissant J, Kijas JW, Coltman DW; International Sheep Genomics Consortium (2011). A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. Mol Ecol Resour. 11(2):314-22.

Mömke S, Kerkmann A, Wöhlke A, Ostmeier M, Hewicker-Trautwein M, Ganter M, Kijas J; International Sheep Consortium, (2011) A frameshift mutation within LAMC2 is responsible for Herlitz type junctional epidermolysis bullosa (HJEB) in black headed mutton sheep. PLoS One. 6(5):e18943.

Oddy H, Dalrymple B, Mcewan J, Kijas J, Hayes B, van der Werf J, Emery D, Hynd P, Longhurst T, Fischer T, Ferguson D, Forage R, Cockett, Nicholas F (2007) Sheepgenomics and the international sheep genomics consortium. *Proc Assoc. Advmt. Anim. Breed. Genet.* **17:** 411-417