

# PIG TALES

## Newsletter of the International Swine Genome Sequencing Consortium (SGSC) Pig Genome Sequence Project

### Project Directors

**Jane Rogers**  
Wellcome Trust Sanger Institute  
**Larry Schook**  
University of Illinois-Urbana

Technical (\*) and  
Steering Committee(#)

**A. Archibald\*##**  
Roslin Institute  
**C. Beattie\***  
University of Illinois-Chicago  
**J. Beever\*##**  
University of Illinois-Urbana  
**M. Boggess#**  
National Pork Board  
**J. Cassidy#**  
North Carolina State University  
**P. Chardon\*##**  
INRA-Jouy-en-Joysas  
**K. Eversole#**  
Alliance for Animal Genome  
Research  
**M. Fredholm\***  
Danish Ag and Vet  
**R. Green#**  
USDA/ ARS  
**D. Hamernik#**  
USDA /CSREES  
**S. Humphray\*##**  
Wellcome Trust Sanger Institute  
**B. Liu\***  
Beijing Genome Center  
**D. Milan\*##**  
INRA-Toulouse  
**S-J Oh\***  
National Livestock Research  
Institute  
**F.A. Ponce de Leon\***  
University of Minnesota  
**G. Rohrer\***  
USDA/ ARS  
**M. Rothschild\*##**  
Iowa State University  
**H. Uenishi\***  
National Institute of Agrobiological  
Sciences  
**M-C Wu\***  
Taiwan Livestock Research Institute

### Successful PAG and St. Croix Workshops

The past six months have been extremely busy ones for the SGSC! During the Plant and Animal Genomics Meeting in San Diego, over 75 participants attended the Swine Genome Sequencing Workshop. During the workshop, Sean Humphray provided an update on the status of the pig genome project and Larry Schook, Jane Rogers and Alan Archibald discussed related sequencing activities. There was considerable discussion on the future development of a porcine SNP chip as well as activities associated with the annotation and assembly of the sequence. These topics provided the basis for an extended workshop that was held in St. Croix, US Virgin Islands on March 15-17, 2007. That workshop was supported in part by Monsanto Choice Genetics and included a session organized by Mike Grosz (Monsanto Choice Genetics) and Mark Boggess (NPB).

The first session at the St. Croix workshop addressed technical issues and was co-Chaired by Jane Rogers and Larry Schook. Larry provided an project overview and Sean Humphray presented a project status report. Mario Caccamo (Sanger) demonstrated how the sequence information is being used for chromosome assembly builds and annotation. Finally, Craig Beattie gave an update of RH mapping and strategies on targeting gaps. The second session "Translating DNA Sequence into Profits: the Pork Industry Enters the Post-Genome Sequence era" was co-Chaired by Mike Grosz and Mark Boggess. Mike Grosz provided an overview and mapping the profit QTL. Mark Boggess, shared the producer's perspective on how they see the pig genome sequence aiding industry needs. An European industry perspective was presented by Gerard Albers (Nutreco). Strategies for commercializing animal genomics and immediate applications was provided by Dennis Fantin, MMIGenomics and Elisabetta Giuffra, Tecnoparco, illustrated how genomic information is being used to address disease resistance.

Discussions defining the next phase of the pig genome project was Chaired by Deb Hamernik. Deb provided a CS-REES-USDA historical perspective on animal genomics and presented the agency's future vision. Lars Boland provided an update on the extended Sino-European collaboration for additional genome sequencing and Alan Archibald provided a summary of current genetic resources and databases. Jane Rogers then shared her thoughts and strategies for ensuring a high quality genome sequence and the required resources to achieve this goal. The last session addressed new initiatives and was co-Chaired by Alan Archibald and Jon Beever. Jon and Hirohide Uenishi shared their perspectives and activities on full Length cDNA library construction and sequencing, Martien Gronen outlined the SNP discovery pipeline they have established and Rosane Oliveira demonstrated how she has utilized iRNA for gene validation. Merete Fredholm shared their innovative bioinformatics approaches to identify non-coding RNA and regulation of gene expression. Emøke Bendixen updated the group with respect to her groups activities related to creating a proteomics initiative and a pig DB.



Left to Right: Gerard Albers, Christian Bendixen, Deb Hamernik, Barbara Harlizius, Alan Archibald, Emøke Bendixen, Merete Fredholm, Craig Beattie, Mario Caccamo, Jane Rogers, Sean Humphray, Jon Beever, Mark Boggess, Larry Schook, Harris Lewin, Michael Grosz, Rosane Oliveira, Bruce Schatz, Chan-Kyu Park, Martin Gronen, Dennis Fantin, Eung-Woo Park, Elisabetta Giuffra, Alan Mathew, Hirohide Uenishi, and Lars Bolund.

## Annotation and analysis of the pig genome sequence

The emerging sequence of the pig genome will be an invaluable resource. However, to understand the functioning of the genes and regulatory elements, and to design sensible molecular biological experiments to test hypotheses, the genome sequence must be related to the extant functional data for that organism. Thus, analysis and annotation of the pig genome sequence is an essential and integral part of international Pig Genome Sequencing Project. SGSC Members Alan Archibald, Sean Humphray and Jane Rogers have secured a grant of almost £1 million from the UK's Biotechnology and Biological Sciences Research Council (BBSRC) to support analysis and annotation of the pig genome sequence. The grant will fund four posts – one in each of Ewen Birney and Tim Hubbard's Ensembl team at the European Bioinformatics Institute (EBI) and the Wellcome Trust Sanger Institute (WTSI) respectively, one in Jane Rogers team at WTSI and one at Roslin Institute with Alan Archibald.

They will use the well-established Ensembl system as the main tool for storage, management and dissemination of pig genome data. Thus, the Pre-Ensembl, Ensembl and Vega systems hosted at the Hinxton Genome campus will be the primary repositories for annotations of the pig genome. The BAC map-based view is on-line in Pre-Ensembl ([http://pre.ensembl.org/Sus\\_scrofa\\_map/index.html](http://pre.ensembl.org/Sus_scrofa_map/index.html)), the emerging draft sequences of chromosomes 7 and 14 are also available in Pre-Ensembl ([http://pre.ensembl.org/Sus\\_scrofa/index.html](http://pre.ensembl.org/Sus_scrofa/index.html)) and the manually annotated and curated limited 'finished' sequence of the MHC and part of chromosome 17 are available in Vega ([http://vega.sanger.ac.uk/Sus\\_scrofa/index.html](http://vega.sanger.ac.uk/Sus_scrofa/index.html)). The BBSRC-funded effort will provide a focus for analysis and annotation of the pig genome sequence. The standard Ensembl pipeline will be used to annotate the sequence with genes and other key features and to develop a complete gene build for the pig in the medium term.

They will integrate the pig genome sequence with diverse pre-existing data sets, including SNPs, ESTs and quantitative trait loci (QTL) and integrate the sequence with maps (genetic, physical) and physical resources (clones, microarrays) providing a seamless route for interrogation and development of experimentation tools. Finally computational approaches, integrating the above resources and also leveraging the comparative genomics potential in the mammalian clade will be used to analyse and present the genome in a user friendly format.

The Distribute Annotation System (DAS) can be used to decorate the sequence with other information and Alan would be pleased to hear from groups interested in providing such DAS tracks on the Ensembl framework. They will also engage the wider community in the annotation effort, for example, through annotation jamborees. The SGSC is aiming to provide high quality consistent annotation and will organise training workshops in late 2007 or early 2008 in order to achieve the desired consistency and quality. An annotated pig genome sequence will dramatically accelerate research on the pig as an important animal for agriculture and human biology. Together with the wider pig research community our aim is to make the pig genome sequence maximally useful by delivering an annotated sequence of the highest quality in a user friendly manner.

Contact: Alan Archibald, Roslin Institute; [alan.archibald@bbsrc.ac.uk](mailto:alan.archibald@bbsrc.ac.uk)

### SGSC CALENDAR OF EVENTS

Oct. 26, 2007	SGSC Workshop Hosted by INRA, Paris, France
Jan. 13, 2008	2008 SGSC Workshop at PAG, San Diego
TBA	Annotation and Assembly Workshop Hinxton, UK

### SPONSORS

European Union SABRE Funding  
Institute for Pig Genetics (TOPIGS), Netherlands  
INRA Genescope  
Iowa Pork Producers Association  
Iowa State University  
Korean National Livestock Research Institute  
National Institute of Agrobiological Sciences, Japan  
National Pork Board, U.S.A.  
North Carolina Pork Council  
North Carolina Agricultural Research Service  
North Carolina State University  
University of Illinois  
Wellcome Trust Sanger Institute  
BBSRC, UK



Sean Humphray and Jane Rogers stop by the University of Illinois on way to PAG to meet T.J. Tabasco



The 3X BAC skim of the porcine genome is supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number #2006-35216-16668.