The Pig Genome Project Has Plenty to Squeal about

B. Fan^{a, b} D.M. Gorbach^a M.F. Rothschild^a

^aDepartment of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames, Iowa, USA; ^bKey Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Education and College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, PR China

Key Words

Genetic resources • Genome analysis tools • Mapping • Pig genomics • Porcine

Abstract

Significant progress on pig genetics and genomics research has been witnessed in recent years due to the integration of advanced molecular biology techniques, bioinformatics and computational biology, and the collaborative efforts of researchers in the swine genomics community. Progress on expanding the linkage map has slowed down, but the efforts have created a higher-resolution physical map integrating the clone map and BAC end sequence. The number of QTL mapped is still growing and most of the updated QTL mapping results are available through PigQTLdb. Additionally, expression studies using high-throughput microarrays and other gene expression techniques have made significant advancements. The number of identified non-coding RNAs is rapidly increasing and their exact regulatory functions are being explored. A publishable draft (build 10) of the swine genome sequence was available for the pig genomics community by the end of December 2010. Build 9 of the porcine genome is currently available with Ensembl annotation; manual annotation is ongoing. These drafts provide useful tools for such endeavors as comparative genomics and SNP

scans for fine QTL mapping. A recent community-wide effort to create a 60K porcine SNP chip has greatly facilitated whole-genome association analyses, haplotype block construction and linkage disequilibrium mapping, which can contribute to whole-genome selection. The future 'systems biology' that integrates and optimizes the information from all research levels can enhance the pig community's understanding of the full complexity of the porcine genome. These recent technological advances and where they may lead are reviewed.

Copyright © 2011 S. Karger AG, Basel

Archaeological evidence from figurines and bones revealed that the domestication of the pig (*Sus scrofa*) occurred as early as 9,000 years ago in the Middle East, and some evidence indicated the domestication of the pig could date back even earlier in China [Department of Animal Science Oklahoma State University, 1995; Clutton-Brock, 1999]. After transformation of pigs from hunted wild animals to fully managed domestic ones, relationships between humans and pigs became closer, yet more diverse and complex [Albarella et al., 2007]. Humans have developed a variety of pig breeds for meat production and a source of fat during the co-evolution of both species. To date, there are 541 reported pig breeds world-

wide [Rischkowsky and Pilling, 2007]; however, only 6 of them including Large White (Yorkshire), Landrace, Duroc, Hampshire, Berkshire and Piétrain are dominant in the commercial pork industry [Rothschild and Ruvinsky, 2010].

Enormous progress has been made for swine breeding with the utilization of classic genetics and quantitative genetics approaches. Many studies identified candidate genes and quantitative trait loci (QTL) significantly associated with economically important traits and some of them have been utilized in the commercial pig industry [Rothschild et al., 2007]. It is predicted that even more progress can be achieved after pig genomics information is better understood and then integrated into marker-assisted selection. Coordinated efforts to better understand the pig genome were initiated in the early 1990s with gene identification and mapping efforts. After entering the new millennium, a historical 'White Paper' outlining the roles pigs play in agriculture and as biological models for humans was announced, with the objective to sequence the whole swine genome [Rohrer et al., 2002]. The annotated build 9 of the porcine genome was made available in September 2009. The publishable build 10 was completed by the end of December 2010 with Ensembl annotation to come a few months later. At the same time, significant progress has been made on gene expression using high-throughput arrays and functional studies of the pig genome [Tuggle et al., 2007]. Additionally, the 60K porcine SNP chip [Ramos et al., 2009] released in January of 2009 has facilitated whole-genome association analyses and research into whole-genome selection. Swine genome science has come of age and the advances in swine genomics promise to greatly benefit the pig industry and consumers in the future.

Linkage Mapping and Gene Mapping Progress

The first 2 pig linkage maps were generated by the PiGMaP and USDA-MARC genome mapping projects around 15 years ago [Archibald et al., 1995; Rohrer et al., 1996]. The USDA-MARC linkage map was comprised of ~1,200 microsatellite markers, and since its first publication additional genetic markers such as microsatellites, amplified fragment length polymorphisms (AFLPs), single nucleotide polymorphisms (SNPs) and expressed sequence tags (ESTs) have been being identified and merged (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9823; <http://www.marc.usda.gov/genome/swine/swine.html>). The ArkDB database combines infor-

mation from various linkage maps, including the USDA-MARC map, and has information on nearly 1,600 genes and 3,300 markers (<http://www.thearkdb.org/arkdb/do/getChromosomeDetails?accession=ARKSPC000000001>).

Two recently developed SNP linkage maps include a primary map composed of ~1,800 in silico identified variants [Li et al., 2008] and a map consisting of 66 SNPs derived from bone-related genes based on Berkshire × Yorkshire pigs [Onteru et al., 2008]. The steps to expand the linkage map have slowed down in recent years, partially due to the advances in the genome sequencing project and high-density porcine SNP chip development. Instead, efforts have moved to developing the pig-human comparative map and integrating the linkage, physical and cytogenetic maps.

Driven by the pig genome sequencing project, a new human-pig comparative map mainly consisting of ESTs and BAC-end sequences was constructed as a first step towards sequencing the pig genome [Meyers et al., 2005]. A large-scale porcine BAC physical map was further developed and used for the selection of BACs to construct a minimal tiling path for genome sequencing and targeted gap filling [Rogatcheva et al., 2008]. Eventually, an updated physical map integrating the clone map and BAC end sequence data was reported [Humphray et al., 2007], which not only greatly contributed to the assembly of the pig genome but also facilitated electronic positional cloning of genes and human-pig comparative genomics studies.

As one of the principal and most accurate mapping techniques, FISH is still being used for porcine gene mapping in a few studies. A 7,000-rad radiation hybrid (RH) panel (ImpRH) [Yerle et al., 1998], which contains nearly 6,000 markers including microsatellites and over 2,000 ESTs, and a somatic cell hybrid panel [Yerle et al., 1996] have been broadly utilized for gene mapping in recent years. An advanced 12,000-rad RH map was also developed, and it allows for even greater precision for mapping within and across species [Yerle et al., 2002]. The utilization of these valuable tools has led to the rapid increase in the number of genes mapped to pig chromosomes during the last decade.

QTL and Candidate Genes

For pork production to be cost-effective, it requires an efficient growth rate, a low feed-to-gain ratio, high levels of reproductive success and survivability and increased pig health. Consumers are also looking for pork products

with good carcass merit and high meat quality. Using pig resource families founded by commercial and exotic pig breeds, researchers have identified quantitative trait loci (QTL) affecting these traits. A large number of QTL have been reported on nearly all chromosomes for growth, carcass and meat quality traits and several chromosomes for reproduction and feet and leg structure. The QTL affecting immune response and disease resistance traits are far less numerous, but this is an area where gene expression studies are particularly valuable. An extensive summary of pig QTL is available at the PigQTLdb (<http://www.animalgenome.org/QTLdb/pig.html>) [Hu et al., 2005], which combines all the published QTL information, integrates linkage and physical maps, and is an excellent repository for updated QTL mapping results. Currently, there are 6,344 QTL reported in the PigQTLdb, which represent 593 different traits.

Some valuable lessons learned from recent QTL mapping studies include: (1) Many traits result from a complex interaction of several different factors. A QTL study that focuses on a single phenotype measurement may result in an incomplete understanding of the multitude of factors that combine to produce the trait of interest. Instead, a systematic study of factors that may contribute to the analyzed trait can help illustrate the underlying causes of a specific phenotype with better accuracy. For example, ovulation rate, follicle-stimulating hormone levels, uterine length, feed intake and environmental factors may likely add value to QTL affecting litter size. (2) Joint and QTL analyses which combine existing results from different studies may better utilize resources. Due to funding limitations, sample sizes are frequently smaller than ideal. By combining the data from multiple QTL studies, possible experimental limitations can be corrected while increasing the power for detecting real QTL [Kim et al., 2005; Muñoz et al., 2009]. Development of new approaches will also improve multi-factorial and complex trait analysis for QTL [Rothschild et al., 2007].

With the identification of a large number of pig SNPs and the application of high-throughput genotyping platforms, fine mapping within specific QTL areas is being carried out. On the end of SSC17 where QTL for meat quality traits were mapped, around 50 additional SNPs were mined. Combining the linkage disequilibrium association and QTL analyses, several genes such as *CTSZ* and *AURKA* were found to be significantly associated with meat color and lactate-related traits [Ramos, 2006]. On SSC2, fine mapping and haplotype analyses confirmed that the *CAST* gene was one of the important can-

didate genes influencing meat quality [Meyers et al., 2007]. Fine mapping was also performed on SSC8 to identify 2 candidate genes responsible for ovulation rate [Campbell et al., 2003]. Due to the availability of the newly developed high-density 60K porcine SNP chip, large-scale association studies are discovering many new QTL.

A few studies in swine have started to focus on non-traditional QTL effects, such as eQTL and imprinted QTL. The study of pig eQTL, which measures and treats each observation of gene expression from a microarray as a new trait, shows promise for traits with a strong environmental influence on gene expression. A simulation study on selective transcriptional profiling and data analysis strategies for eQTL mapping in outbred pig populations was conducted, which suggested a strategy for future eQTL mapping analyses [Cardoso et al., 2008]. Another non-traditional type of QTL study involves imprinted QTL, of which many have been found. For example, imprinted QTL were found on SSC2 [Van Laere et al., 2003] on which the *IGF2* gene is located, and on SSC6 and SSC14 [de Koning et al., 2000; Holl et al., 2004]. The discovery of imprinted QTL provides an opportunity to further understand the genetic control of traits of interest in pigs.

Linkage disequilibrium (LD) plays a vital role in QTL mapping, so genotyping of high-density SNPs in genes or chromosomal regions of interest is an essential prerequisite to discover the LD structure within a region. This information allows causative mutations, LD blocks and selective sweeps in a population to be discovered. For example, as evidence that LD levels can be low even within a single gene, an absence of LD was observed between 2 mutations (Arg236His and Asp298Asn) in the *MC4R* gene, and both mutations were significantly associated with fatness and growth traits [Fan et al., 2009a]. Hence, high-density SNPs may give further clues about causative mutations in certain genes. Similarly, large-scale SNP analyses need to be validated in different populations and breeds due to the variability in some gene sequences and LD levels among different breeds. For instance, studies of the sequence variation of 3 different genes showed no evidence of reduction in genetic variability among different pig breeds for *FABP4* and *IGF2* genes, but a significant reduction in genetic variability for *FABP5* in different breeds [Ojeda et al., 2008a, b]. Chromosome-wise LD analyses showed a difference in LD among different commercial pig breeds, as well as in European and Chinese indigenous pig breeds [Du et al., 2007; Amaral et al., 2008]. The LD blocks were extended up to 400 kb for European breeds and only 10 kb for Chinese breeds. The

average extent of LD in the pig was a bit higher than in humans, which ranged from 10–30 kb. Simulations and case studies have demonstrated that the extent of LD is correlated with the distance between markers, and it depends on the genetic structures of breeds/lines and mating systems [Nsengimana et al., 2004]. More information on LD levels should be obtained in the near future by the swine HapMap project's utilization of the new 60K SNP chip.

Candidate gene analyses have been employed to investigate a variety of economic traits. To date, significant association has been demonstrated for candidate genes with litter size (*ESR1*, *PRLR*, *RBP4*, *FSHB*), growth (*MC4R*), muscle mass (*IGF2*, *MSTN*), meat quality (*HAL*, *RN*, *PRKAG3*, *CAST*), disease resistance (*K88*, *FUT1*, *SLA*, *NRAMP*) and coat color (*KIT*, *MC1R*). The commercial pig industry is actively using many of these gene markers in combination with traditional performance information to improve production by marker-assisted selection (MAS).

While the traits listed above have received most of the attention, many traits of economic importance to the swine industry have been greatly overlooked. As QTL studies have become more prevalent, there is more interest in trying to determine QTL for non-traditional traits. For example, recent studies using a high-throughput genotyping technique revealed that the genes *CCR7* and *CPT1A* were associated with sow longevity traits [Mote et al., 2009], the genes *CALCR* and *COL1A2* were significantly associated with leg action [Fan et al., 2009b], and the genes *ELF5*, *KIF18A*, *COL23A1* and *NPTX1* were associated with scrotal hernias in pigs [Du et al., 2009].

Databases and Online Resources

With the rapid advances of the porcine genome sequence and other related projects, a tremendous amount of data is being generated every day. Thus, large databases to store this information and make it publicly available are becoming more essential, as are online resources that enable analysis of this new data. Significant pig bioinformatics efforts have been undertaken by the Roslin Institute, Sanger Institute, TGAC, NCBI, NAGRP Bioinformatics and Pig Genome Coordination Programs, and other research institutes around the world. Information is shown in table 1 about the online database resources beneficial for the pig genomics community.

Functional Genomics

In recent years, a number of studies have been carried out to profile expression of the pig genome. The number of porcine expressed sequence tags (ESTs) has expanded dramatically, with over 1,600,000 ESTs deposited in GenBank (<http://www.ncbi.nlm.nih.gov/sites/entrez>). There are over 180 porcine cDNA libraries which have each contributed more than 1,000 sequences to UniGene. These libraries came from almost 50 categories of different tissues, such as muscle, brain, intestine, ovary, blood and embryonic tissue. In UniGene Build 40, the mRNAs (8,377), high-throughput cDNAs (19,562) and ESTs (1,250,988) have been clustered into 51,270 UniGene sets (<http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=9823>). Overall, data on pig gene expression is being produced at a rapid rate.

The combination of expression data and the completed genome sequence will certainly contribute to the better understanding of the physiological complexity of the pig transcriptome and uncover genes and gene pathways that control traits of economic importance. Many studies have shown that the differences that exist in the porcine transcriptome are dependent upon breed and developmental stage, among other factors, so there is much left to learn through expression studies. Recently developed techniques such as microarray, quantitative real-time PCR, serial analysis of gene expression (SAGE) and ChIP-chip (chromatin immunoprecipitation on a microarray chip) are being utilized for expression studies in livestock, and these tools can greatly enhance the swine community's understanding of the porcine genome.

Microarray studies in particular have been used extensively to determine the biological basis for such traits as reproduction, muscle development, feed intake and host response to infection in pigs [Tuggle et al., 2007]. Cooperative efforts led by the US Pig Genome Coordinator and a committee of interested scientists have produced a second-generation 70-mer cDNA or oligo spotted array (<http://www.pigoligoarray.org/>). The array is a 20,400-element oligonucleotide microarray with probes designed based on comparison of pig ESTs to known vertebrate proteins. Affymetrix also sells a 23,937-probe porcine microarray.

Another technique for expression profiling is SAGE, which is a sequencing-based method. The sequenced SAGE tags are a digital representation of cellular gene expression and have been used to identify genes differentially expressed at various stages of development. Successful examples using SAGE profiling for 2 critical em-

Table 1. The online resources and databases useful for the pig genomics community

Category	Online resource/database	Description	URL
Genome	US pig genome mapping site	Comprehensive	http://www.animalgenome.org/pigs/
Genome	Pig genome resource at NCBI	Comprehensive	http://www.ncbi.nlm.nih.gov/genome/guide/pig/
Genome	Swine Genome Sequencing Consortium (SGSC)	Comprehensive	http://www.piggenome.org/
Sequencing	Porcine genome sequencing project at Sanger Institute	Physical mapping and genome sequencing	http://www.sanger.ac.uk/Projects/S_scrofa/
Sequencing	Sino-Danish pig genome sequencing project	genome sequences and EST sequences	http://piggenome.dk/
Sequencing	Ensembl (<i>Sus scrofa</i>)	Assemblies for chromosomes 1–18 and X	http://www.ensembl.org/Sus_scrofa/Info/Index
Sequencing	VEGA (<i>Sus scrofa</i>)	Preliminary annotation of the sequenced region on SSC6, 7 and 17	http://vega.sanger.ac.uk/Sus_scrofa/index.html
SNP	dbSNP (<i>Sus scrofa</i>)	Database of SNPs and other genetic variation	http://www.ncbi.nlm.nih.gov/SNP/snp_batchSearch.cgi?org=9823&type=SNP
Linkage map	ArkDB	linkage mapping, RH mapping, physical mapping	http://www.thearkdb.org/arkdb/do/getChromosomeDetails?accession=ARKSPC00000001
Linkage map	USDA/MARC Linkage Map	Genetic linkage map	http://www.marc.usda.gov/genome/swine/swine.html
Physical map	Cytogenetic Map	pig-human comparative map	https://www-lgc.toulouse.inra.fr/pig/cyto/cyto.htm
Physical map	IMpRH maps	IMpRH mapping	https://www-lgc.toulouse.inra.fr/pig/RH/Menuchr.htm
Expression	DFCI Pig Gene Index	ESTs annotation and cluster	http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=pig
Expression	Agenae Project	Pig ESTs	http://public-contigbrowser.sigenae.org:9090/Sus_scrofa/index.html
Expression	Pig array	Porcine microarray	http://www.pigoligoarray.org/
Expression	UniGene (<i>Sus scrofa</i>)	transcribed sequences and gene-based clusters	http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=9823
Disease	Online Mendelian Inheritance in Animals	Porcine genetic disorders	http://omia.angis.org.au/adv_search_results.shtml?field1=sci_name&query1=Sus+scrofa
Disease	International PRRS Genomics Consortium	PRRS genomics	http://www.prrs.org/
Disease	MHC database	Swine SLA sequences	http://www.ebi.ac.uk/ipd/mhc/sla/index.html
Disease	IMGT®	Immunogenetics	http://imgt.cines.fr/textes/IMGTveterinary/#Pig
QTL	PigQTLdb	QTL repository and integration of linkage and physical maps	http://www.animalgenome.org/QTLdb/pig.html
QTL	GridQTL	QTL mapping	http://www.gridqtl.org.uk/
Non-coding RNA	miRBase	published miRNA sequences and annotation	http://microrna.sanger.ac.uk/cgi-bin/targets/v5/genome.pl
Non-coding RNA	Rfam database	Non-coding RNAs annotation	http://rfam.sanger.ac.uk/
Bioinformatics	NAGRP Bioinformatics Coordination Program	Bioinformatics tools	http://www.animalgenome.org/bioinfo/
Bioinformatics	Geocities	Bioinformatics tool list	http://www.geocities.com/bioinformaticsweb/tools.html
Bioinformatics	BLAST (<i>Sus scrofa</i>)	BLAST Pig Sequences	http://animalgenome.org/blast/blast.php?bdb=pig10

bryo development stages and 3 skeletal development stages in 2 pig breeds have been reported [Blomberg and Zuelke, 2004; Tang et al., 2007].

ChIP-chip technology received its name because it determines the results from chromatin immunoprecipitation through use of a microarray. ChIP-chip is used to determine which DNA regions are bound to different proteins and can be used to identify promoter regions, enhancers, repressors, and silencers. Though this technology is not currently utilized in pigs, it may come soon after completion of the genome sequence.

Recently, numerous studies demonstrated that non-protein-coding RNAs, including small nucleolar RNA (snoRNA), microRNA (miRNA), short interfering RNA (siRNA), piwi-interacting RNA (piRNA), and natural antisense RNA (NAT), function in gene expression regula-

tion at many levels, including chromatin architecture, RNA editing, and RNA stability and translation, for example. Porcine miRNAs were reported from the Sino-Danish 0.66× coverage pig genome sequence, but were not extensively examined [Wernersson et al., 2005]. More recent studies have identified and begun characterization of miRNAs by means of direct cloning and sequencing, in silico prediction and northern hybridization validation [Huang et al., 2008; Reddy et al., 2009]. The present studies suggest some miRNAs are involved in reproduction, muscle growth and development. Further work on miRNA target site identification, functional mechanisms and biological significance is being explored.

RNA interference technology is a method for selectively silencing genes by creating double-stranded RNA. The pilot work using an RNAi technique in pigs was car-

ried out for pig transgenesis studies [Clark et al., 2007]. The successful knock-down of target genes such as porcine endogenous retroviruses (PERVs) by siRNA transformation into cells was observed [Miyagawa et al., 2005; Palmer et al., 2006]. Cong et al. [2010] also showed the possibility of using *Salmonella choleraesuis* as a vector for siRNA administration to swine, but a standard, efficient platform for RNAi experiments in pigs has yet to be developed.

Sequencing Progress

In addition to the recently sequenced livestock animals of biological significance, such as cow, dog, chicken and horse, the pig will be the next one with its entire genome published. With the endeavors of the Swine Genome Sequencing Consortium (SGSC) and many other research institutes, tremendous progress has been made in sequencing the pig genome [Archibald et al., 2010] and developing the related scientific projects such as the porcine HapMap project and a high-density SNP chip. The first swine genome sequence draft with an overall depth of 4× coverage and 6× in some locations (build 9) became available to the pig genomics community in April of 2009 (http://www.ensembl.org/Sus_scrofa/Info/Index/). An updated sequence alignment (build 10) incorporating 24× coverage with Illumina whole genome shotgun (WGS) reads from BGI, 1× coverage of WGS capillary reads from National Livestock Research Institute-Korea, and additional short-read sequences from Sanger Institute was available by the end of December 2010 [International Swine Genome Sequencing Consortium 2010].

The primary strategy for sequencing the pig genome was to combine the WGS approach with a minimal tiling path (MTP) of bacterial artificial chromosomes (BACs). The MTP was constructed from previous landmarks, BAC end sequences, and restriction fingerprinting of over 260,000 BACs derived from 4 BAC libraries that have been used for positional cloning and QTL mapping studies [Schook et al., 2005]. It was estimated that the length of the entire pig genome was about 2.6 Gb, and the original plan was to sequence 16,000–17,000 BAC clones and generate a draft 6× coverage of the genome from a 3× coverage of BACs comprising the MTP and a 3× coverage of WGS sequencing. Sequencing work and assembly are ongoing at the Wellcome Trust Sanger Institute and the Genome Analysis Centre, respectively, through this international collaboration composed of many different laboratories.

Different from the methods used for human and mouse genome sequencing, a high-resolution physical map integrating the fingerprint map and BAC end sequence data was constructed before the initiation of whole-genome sequencing. This map comprised 172 high continuity clone contigs distributed over all 18 autosomes and the 2 sex chromosomes, X and Y, representing ~2.6 Gb [Humphray et al., 2007]. A total of 98.3% of this integrated physical map has been selected in 16,273 BAC clones for sequencing. To date (January 24th, 2011), a total of 16,193 of these clones have been sequenced, which contributes 3,013 Mb (180.7 Mb finished quality) (http://www.sanger.ac.uk/Projects/S_scrofa/) [International Swine Genome Sequencing Consortium 2010]. In addition, efforts are being made to close the remaining gaps by incorporating next-generation sequence reads from several sources. Information on the current status of the genome sequencing project, as well as on each chromosome, is available from the Sanger Institute website (http://www.sanger.ac.uk/Projects/S_scrofa/) and NAGRP Pig Genome Coordination Program website (<http://www.animalgenome.org/pigs/genomesequence>).

Particular regions on a few chromosomes were selected for deeper sequencing (6–8× coverage) due to the potential biological effects of these regions and the availability of funding. The Institute for Pig Genetics in the Netherlands funded the additional sequencing on all of chromosome 4, while the EU SABRE project funded additional sequencing on chromosomes 7 and 14. Chromosome 17 was also made a high priority for sequencing. Further work on the sex chromosomes has been funded and the Wellcome Trust Sanger Institute plans to further annotate those chromosomes. The sequences from chromosomes 1–18 and X have been released into the ENSEMBL website (http://www.ensembl.org/Sus_scrofa/Info/Index/). The pig X chromosome sequencing project is also making excellent progress through the sequencing pipeline at the Sanger Institute, with sequence data now available for 879 clones.

Manual annotation of 3 interesting regions from the pig genome was completed, which included 2.4 Mb of the major histocompatibility complex (MHC) region on SSC7p1.1–q1.1, ~8 Mb of chromosome 17, and part of a ~300-kb region on chromosome 6 orthologous to the human leukocyte receptor complex (LRC). The annotated data are publicly available via the Vertebrate Genome Annotation (VEGA) website (http://vega.sanger.ac.uk/Sus_scrofa/index.html). The work on assembly and annotation of the remaining sequenced regions is ongoing. Even after build 10 of the genome is published, more targeted

sequencing will still be needed to fill the remaining gaps for annotation to be complete.

Using the new genome sequence, the genetic effects observed in pigs when they are studied as models of human diseases can be more easily translated into predictions for humans. Though we currently do not know the level of similarity between orthologous regions in pigs and humans, the previous draft 0.66× coverage of the genome completed by the Sino-Danish Pig Genome Project [Wernersson et al., 2005] and the annotated region on SSC17 [Hart et al., 2007] demonstrated that the pig was much closer to human than mouse, as shown from the types of syntenic data. Pigs also seem to have an isochore structure most similar to that in humans.

One discovery from human genome variation studies that is ongoing in pigs with the new genome sequence is the identification of copy number variants (CNVs) [Fadista et al., 2010]. CNVs represent those DNA segments longer than 1 kb present at variable copy number between different individuals, and they might be responsible for phenotype variances of disease associated traits. Few CNVs had been uncovered in pigs until recently when a custom tiling oligonucleotide array was used for screening unrelated individuals on the regions of SSC4, 7, 14 and 17 already sequenced and assembled. This screen identified 37 CNV regions [Fadista et al., 2008]. A feasible approach for CNV genotyping similar to the human CNV array should be developed in the future.

SNP Chip

The recent developments on the pig genome sequence and the next-generation sequencing technologies have driven the production of a high-density porcine SNP chip. A 60K chip has been developed and distributed by the International Swine SNP Consortium and Illumina for the pig community, and it will serve as an extremely useful tool for a variety of research fields including animal genetic evaluation, whole-genome association, LD mapping, population genetics and evolution genomics studies.

The porcine SNP chip was designed as an Illumina Infinium iSelect™ SNP BeadChip with a selection of 64,232 SNPs, which were chosen from a total of over 549,000 SNPs [Ramos et al., 2009]. The SNPs genotyped by this SNP chip include already validated SNPs from the public dbSNP database and those mined from the deposited porcine shotgun sequences [Kerstens et al., 2009], as well as SNPs identified de novo using high-throughput

sequencing techniques from reduced representation libraries (RRL) [Wiedmann et al., 2008; Ramos et al., 2009]. Decisions about which SNPs to include on the SNP chip were primarily based on each SNP's position in the genome to ensure even coverage across chromosomes. Other criteria considered include the estimated minor allele frequency (MAF) and whether the SNP passed other quality control measures and standards for Illumina Infinium iSelect SNP BeadChip preparation. Based on build 9 of the porcine genome, the median distance between SNPs is 28 kb, though stretches as long as 1.7 Mb exist. The long gaps could be a result of current issues with the genome assembly, and their lengths will likely change in the upcoming build. Most of the SNPs reside on the autosomes and X chromosome.

A pilot study using the 60K SNP chip was carried out at Wageningen University using a panel of 158 individuals representing the original discovery panel. From this study, 62,121 SNPs (97.5%) were successfully scored. Minor allele frequencies were fairly high with 58,994 SNPs being polymorphic of which 57,109 had MAF >0.05. The average MAF for all scorable SNPs was 0.274 [Ramos et al., 2009]. When the panel was tested on other *Sus* species, 76–99% of the SNPs were monomorphic [Groenen et al., 2009].

A Possible Future for Swine Genomics

Large-scale whole-genome association analyses are already underway with the new 60K SNP chip in pigs, and they will change the conventional association model with one-trait-one-marker and accelerate progress in the search for genetic variation underlying the inheritance of important production traits and complex genetic diseases [Georges, 2007; Rothschild and Plastow, 2008; Andersson, 2009]. The vast amounts of data produced will allow for haplotype block construction and linkage disequilibrium mapping analyses to determine the important chromosomal regions and facilitate positional cloning. In addition, selective sweeps and positive selection events on breed formation history can be identified through high-density SNP genotyping. These new chips will make it much easier to combine information from multiple studies to complete meta-analyses, since the markers used in each study will be consistent.

Another method which requires high-density markers is genomic selection, which integrates genotypic, phenotypic and pedigree data for breeding value prediction and marker-assisted selection on a genome wide scale. Differ-

ent methods for genomic selection have been developed over the past decade, and simulations show it can increase the accuracy of estimates of genetic merit while reducing the generation interval [Meuwissen, 2007]. These genomic breeding values (GEBV) have already been combined into the selection model for the USDA's genetic evaluation of dairy cattle in the US, and similar utilization of genomic information to evaluate pigs will be feasible in the near future.

The rapid development of structural genomics and functional genomics is making 'systems biology' feasible in pigs. The integration of datasets from various levels of the genome, transcriptome, proteome, epigenome and others will contribute to better understanding of individual genes, genetic pathways and regulatory networks to identify significant variation in genes controlling important traits of interest in pigs [Brazma et al., 2006].

Interest in the pig as a biological model for human biology has continued for several years. Because of the size and physiological similarity of pigs to humans, the pig has become a model of choice for a vast array of disciplines such as nutrition, digestive physiology, cardiovascular disease, diabetes, skin formation and healing, maternal mortality, melanoma and infectious disease [Lunney, 2007; Quilter et al., 2008]. One ongoing project utilizing pigs as a biological model for human disease involves measuring the phenotypic responses of thousands of pigs to Porcine Reproductive and Respiratory Syndrome Virus (PRRSV). The International PRRS Host Genetics Consortium is working to map QTL for response to PRRSV, in hopes of discovering loci that can make pigs more resistant to PRRSV and gaining knowledge that could be applied to infectious disease response in humans. In particular, these results could provide information about viral persistence and transmission pathways [Lunney, 2007].

Conclusions

The rapid advances in structural and functional genomics of the pig provide a good opportunity to better understand the genetic variation underlying economically important and complex phenotypes and may also be of value for understanding human health. The high synteny between the pig and human genome sequences, as revealed by several newly developed high-resolution comparative physical maps, suggests that the pig may be the best biomedical model to study many human diseases. Use of the new high-density porcine SNP chip in as-

sociation analyses will identify many interesting gene effects, but collaboration and joint analyses between different research groups will become more critical. The development of novel statistics methods for linkage disequilibrium mapping, fine mapping and whole-genome selection are urgent, so as to speed the utilization of the existing and forthcoming pig genomics information. The discovery and characterization of genome sequence variants such as SNPs, CNVs and indels, as well as non-coding RNA sequences like miRNA, piRNA and NATs, are necessary. The organization, integration, and optimization of results from all research levels using a 'systems biology' approach will provide a bountiful harvest for the pig genomics community. Overall the pig genomics community has made great strides in recent years and will reap the considerable rewards in terms of new knowledge applicable to both agriculture and human health.

Acknowledgements

The authors acknowledge the partial support provided by the National Pork Board, Iowa Pork Producers Association, U.S. Pig Genome Coordination Program, the College of Agriculture and Life Sciences at Iowa State University, State of Iowa and Hatch funding and the activities of the International Swine Sequencing Consortium. Support for D.M. Gorbach is provided by USDA National Needs Graduate Fellowship Competitive Grant No. 2007-38420-17767 from the National Institute of Food and Agriculture.

References

- Albarella U, Dobney K, Ervynck A, Rowley-Conwy PE: Pigs and Humans: 10,000 Years of Interaction (Oxford University Press, Oxford 2007).
- Amaral AJ, Megens HJ, Crooijmans RP, Heuven HC, Groenen MA: Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* 179:569–579 (2008).
- Andersson L: Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. *Genetica* 136:341–349 (2009).
- Archibald AL, Haley CS, Brown JF, Couperwhite S, McQueen HA, et al: The PiGMaP consortium linkage map of the pig (*Sus scrofa*). *Mamm Genome* 6:157–175 (1995).
- Archibald AL, Boland L, Churcher C, Fredholm M, Groenen MAM, et al: Pig genome sequence – analysis and publication strategy. *BMC Genomics* 11:438 (2010).
- Blomberg LA, Zuelke KA: Serial analysis of gene expression (SAGE) during porcine embryo development. *Reprod Fert Develop* 16:87–92 (2004).

- Brazma A, Krestyaninova M, Sarkans U: Standards for systems biology. *Nat Rev Genet* 7: 593–605 (2006).
- Campbell EM, Nonneman D, Rohrer GA: Fine mapping a quantitative trait locus affecting ovulation rate in swine on chromosome 8. *J Anim Sci* 81:1706–1714 (2003).
- Cardoso FF, Rosa GJ, Steibel JP, Ernst CW, Bates RO, Tempelman RJ: Selective transcriptional profiling and data analysis strategies for expression quantitative trait loci mapping in outbred F2 populations. *Genetics* 180:1679–1690 (2008).
- Clark KJ, Carlson DF, Fahrenkrug SC: Pigs taking wing with transposons and recombinases. *Genome Biol* 8 Suppl 1:S13 (2007).
- Clutton-Brock J: *A Natural History of Domesticated Mammals*, ed 2 (Cambridge University Press, Cambridge; Natural History Museum, New York 1999).
- Cong W, Jin H, Jiang C, Yan W, Liu M, et al: Attenuated *Salmonella choleraesuis*-mediated RNAi targeted to conserved regions against foot-and-mouth disease virus in guinea pigs and swine. *Vet Res* 41:30 (2010).
- de Koning DJ, Rattink AP, Harlizius B, van Arendonk JA, Brascamp EW, Groenen MA: Genome-wide scan for body composition in pigs reveals important role of imprinting. *Proc Natl Acad Sci USA* 97:7947–7950 (2000).
- Department of Animal Science Oklahoma State University: Breeds of Livestock: Swine <http://www.ansi.okstate.edu/breeds/swine> (1995).
- Du FX, Clutter AC, Lohuis MM: Characterizing linkage disequilibrium in pig populations. *Int J Biol Sci* 3:166–178 (2007).
- Du ZQ, Zhao X, Vukasinovic N, Rodriguez F, Clutter AC, Rothschild MF: Association and haplotype analyses of positional candidate genes in five genomic regions linked to scrotal hernia in commercial pig lines. *Plos One* 4:e4837 (2009).
- Fadista J, Nygaard M, Holm LE, Thomsen B, Bendixen C: A snapshot of CNVs in the pig genome. *Plos One* 3:e3916 (2008).
- Fadista J, Zhan B, Hedegaard J, Bendixen C: Structural Variation in the Pig Genome Using Next-Generation Sequencing (32nd Conference of the International Society for Animal Genetics, Edinburgh, Scotland 2010).
- Fan B, Onteru SK, Plastow GS, Rothschild MF: Detailed characterization of the porcine *MC4R* gene in relation to fatness and growth. *Anim Genet* 40:401–409 (2009a).
- Fan B, Onteru SK, Mote BE, Serenius T, Stalder KJ, Rothschild MF: Large-scale association study for structural soundness and leg locomotion traits in the pig. *Genet Sel Evol* 41:14 (2009b).
- Georges M: Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. *Annu Rev Genomics Hum Genet* 8:131–162 (2007).
- Groenen MAM, Crooijmans RPMA, Ramos AM, Amaral AJ, Kerstens H, et al: Design of the Illumina Porcine 50K+ SNP iSelect Beadchip and Characterization of the Porcine HapMap Population (Plant and Animal Genome Conference XVII, San Diego, CA, USA 2009).
- Hart EA, Caccamo M, Harrow JL, Humphray SJ, Gilbert JG, et al: Lessons learned from the initial sequencing of the pig genome: comparative analysis of an 8 Mb region of pig chromosome 17. *Genome Biol* 8:R168 (2007).
- Holl JW, Cassady JP, Pomp D, Johnson RK: A genome scan for quantitative trait loci and imprinted regions affecting reproduction in pigs. *J Anim Sci* 82:3421–3429 (2004).
- Hu ZL, Dracheva S, Jang W, Maglott D, Bastiaansen J, et al: A QTL resource and comparison tool for pigs: PigQTLDB. *Mamm Genome* 16:792–800 (2005).
- Huang TH, Zhu MJ, Li XY, Zhao SH: Discovery of porcine microRNAs and profiling from skeletal muscle tissues during development. *Plos One* 3:e3225 (2008).
- Humphray SJ, Scott CE, Clark R, Marron B, Bender C, et al: A high utility integrated map of the pig genome. *Genome Biol* 8:R139 (2007).
- International Swine Genome Sequencing Consortium: Improving the High-Quality Draft Swine Genome Reference: The Biology of Genomes (Cold Spring Harbor, NY, USA 2010).
- Kerstens HH, Kollers S, Kommadath A, Del Rosario M, Dibbits B, et al: Mining for single nucleotide polymorphisms in pig genome sequence data. *BMC Genomics* 10:4 (2009).
- Kim JJ, Rothschild MF, Beever J, Rodriguez-Zas S, Dekkers JC: Joint analysis of two breed cross populations in pigs to improve detection and characterization of quantitative trait loci. *J Anim Sci* 83:1229–1240 (2005).
- Li XP, Hu ZL, Moon SJ, Do KT, Ha YK, et al: Development of an in silico coding gene SNP map in pigs. *Anim Genet* 39:446–450 (2008).
- Lunney JK: Advances in swine biomedical model genomics. *Int J Biol Sci* 3:179–184 (2007).
- Meuwissen T: Genomic selection: marker assisted selection on a genome wide scale. *J Anim Breed Genet* 124:321–322 (2007).
- Meyers SN, Rogatcheva MB, Larkin DM, Yerle M, Milan D, et al: Piggy-BACing the human genome. II. A high-resolution, physically anchored, comparative map of the porcine autosomes. *Genomics* 86:739–752 (2005).
- Meyers SN, Rodriguez-Zas SL, Beever JE: Fine-mapping of a QTL influencing pork tenderness on porcine chromosome 2. *BMC Genet* 8:69 (2007).
- Miyagawa S, Nakatsu S, Nakagawa T, Kondo A, Matsunami K, et al: Prevention of PERV infections in pig to human xenotransplantation by the RNA interference silences gene. *J Biochem* 137:503–508 (2005).
- Mote BE, Koehler KJ, Mabry JW, Stalder KJ, Rothschild MF: Identification of genetic markers for productive life in commercial sows. *J Anim Sci* 87:2187–2195 (2009).
- Muñoz G, Ovilo C, Silió L, Tomás A, Noguera JL, Rodríguez MC: Single- and joint-population analyses of two experimental pig crosses to confirm quantitative trait loci on *Sus scrofa* chromosome 6 and leptin receptor effects on fatness and growth traits. *J Anim Sci* 87:459–468 (2009).
- Nsengimana J, Baret P, Haley CS, Visscher PM: Linkage disequilibrium in the domesticated pig. *Genetics* 166:1395–1404 (2004).
- Ojeda A, Estelle J, Folch JM, Perez-Enciso M: Nucleotide variability and linkage disequilibrium patterns at the porcine *FABP5* gene. *Anim Genet* 39:468–473 (2008a).
- Ojeda A, Huang LS, Ren J, Angiolillo A, Cho IC, et al: Selection in the making: a worldwide survey of haplotypic diversity around a causative mutation in porcine *IGF2*. *Genetics* 178: 1639–1652 (2008b).
- Onteru SK, Fan B, Rothschild MF: SNP detection and comparative linkage mapping of 66 bone-related genes in the pig. *Cytogenet Genome Res* 122:122–131 (2008).
- Palmer ML, Lee SY, Carlson D, Fahrenkrug S, O'Grady SM: Stable knockdown of CFTR establishes a role for the channel in P2Y receptor-stimulated anion secretion. *J Cell Physiol* 206:759–770 (2006).
- Quilter CR, Gilbert CL, Oliver GL, Jafer O, Furlong RA, et al: Gene expression profiling in porcine maternal infanticide: a model for periparturient psychosis. *Am J Med Genet B Neuropsychiatr Genet* 147B:1126–1137 (2008).
- Ramos AM: Molecular Dissection of Genomic Regions Influencing Important Economic Traits in the Pig: Animal Science (Iowa State University, Ames, IA, USA 2006).
- Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, et al: Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *Plos One* 4:e6524 (2009).
- Reddy AM, Zheng Y, Jagadeeswaran G, Macmill SL, Graham WB, et al: Cloning, characterization and expression analysis of porcine microRNAs. *BMC Genomics* 10:65 (2009).
- Rischkowsky B, Pilling D: The State of the World's Animal Genetic Resources for Food and Agriculture (FAO, Rome 2007).
- Rogatcheva MB, Chen K, Larkin DM, Meyers SN, Marron BM, et al: Piggy-BACing the human genome. I. Constructing a porcine BAC physical map through comparative genomics. *Anim Biotechnol* 19:28–42 (2008).
- Rohrer GA, Alexander LJ, Hu ZL, Smith TPL, Keele JW, Beattie CW: A comprehensive map of the porcine genome. *Genome Res* 6:371–391 (1996).

- Rohrer G, Beever JE, Rothschild MF, Schook L, Gibbs R, Weinstock G: Porcine genomic sequencing initiative. <http://www.animalgenome.org/pigs/community/WhitePaper/2002.html> (2002).
- Rothschild MF, Plastow GS: Impact of genomics on animal agriculture and opportunities for animal health. *Trends Biotechnol* 26:21–25 (2008).
- Rothschild MF, Ruvinsky A: *The Genetics of the Pig*, ed 2. (CABI, Wallingford, Oxfordshire, UK 2010).
- Rothschild MF, Hu ZL, Jiang Z: Advances in QTL mapping in pigs. *Int J Biol Sci* 3:192–197 (2007).
- Schook LB, Beever JE, Rogers J, Humphray S, Archibald A, et al: Swine Genome Sequencing Consortium (SGSC): a strategic roadmap for sequencing the pig genome. *Comp Funct Genomics* 6:251–255 (2005).
- Tang ZL, Li Y, Wan P, Li XP, Zhao SH, et al: LongSAGE analysis of skeletal muscle at three prenatal stages in Tongcheng and Landrace pigs. *Genome Biology* 8:R115 (2007).
- Tuggle CK, Wang YF, Couture O: Advances in swine transcriptomics. *Int J Biol Sci* 3:132–152 (2007).
- Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, et al: A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. *Nature* 425:832–836 (2003).
- Wernersson R, Schierup MH, Jorgensen FG, Gorodkin J, Panitz F, et al: Pigs in sequence space: A 0.66X coverage pig genome survey based on shotgun sequencing. *BMC Genomics* 6:70 (2005).
- Wiedmann RT, Smith TP, Nonneman DJ: SNP discovery in swine by reduced representation and high throughput pyrosequencing. *BMC Genet* 9:81 (2008).
- Yerle M, Echard G, Robic A, Mairal A, Dubut-Fontana C, et al: A somatic cell hybrid panel for pig regional gene mapping characterized by molecular cytogenetics. *Cytogenet Cell Genet* 73:194–202 (1996).
- Yerle M, Pinton P, Robic A, Alfonso A, Palvadeau Y, et al: Construction of a whole-genome radiation hybrid panel for high-resolution gene mapping in pigs. *Cytogenet Cell Genet* 82:182–188 (1998).
- Yerle M, Pinton P, Deltros C, Arnal N, Milan D, Robic A: Generation and characterization of a 12,000-rad radiation hybrid panel for fine mapping in pig. *Cytogenet Genome Res* 97:219–228 (2002).