The Swine Genome Sequencing Project: Implications for Health and Meat Production

Lawrence B. Schook, Ph.D.
Institute for Genomic Biology
University of Illinois
1201 W. Gregory Drive, Urbana, IL 61801
(schook@uiuc.edu)

OVERVIEW. Over the past decade tremendous progress has been made mapping and characterizing the swine genome. Currently, moderate to high-resolution genetic linkage maps containing highly polymorphic loci (Type II) have been produced using independent mapping populations [Rohrer et al. 1996]. Additionally, physical mapping methods such as somatic cell hybrid analysis, *in situ* hybridization and ZOO-FISH have been employed to enrich the Type I marker map and to perform comparative analysis with map-rich species such as the human and mouse. To date, >5,000 mapped loci are cataloged for the pig genome (http://www.thearkdb.org). Recently, whole-genome radiation hybrid (WG-RH) panels have been generated for swine [Hawken et al. 1999] resulting in yet another rapid increase in the number of expressed sequences being mapped facilitating comparative mapping with other species [Rink et al. 2002]. The swine genomics community has also acquired access to resources such as bacterial artificial chromosome (BAC) libraries providing approximately 35X coverage of the swine genome. These BAC resources have facilitated the production of high-resolution physical maps in specific chromosomal regions and support the construction of sequence-ready mapping resources for the porcine genome. This includes the creation of a pig-human comparative map and the initial construction of a whole genome BAC contig. Finally, large scale sequencing of expressed sequences (ESTs) in conjunction with genomic sequencing has permitted the identification of single nucleotide polymorphisms (SNPs) that can be used to finely map traits (disease resistance). Thus, the tools and information have been developed to permit application of genomics into improving the health and performance of pigs. Clearly, low cost diagnostics based on this information will be the next wave of development. Linking new diagnostics with genomic therapies and management tools will also soon to arrive on the marketplace.

BUILDING THE GENOMIC ROADMAP

Comparative Gene Mapping: Leveraging the Human and Mouse Genomic Sequence. The pig genome is of similar size (3 x 10^9 bp), complexity and chromosomal organization (2n = 38, including meta- and acrocentric chromosomes) as the human genome. Comparative genetic maps have indicated that the porcine and human genomes are more similarly organized than either is when compared to the mouse. The mean length of conserved syntenic segments between human and pig is approximately twice as long as the average length of conserved syntenic segments between human and mouse [Ellergren et al., 1994; Rettenberger et al., 1995]. Furthermore, the organizational similarities between the human and porcine genomes reflect similarities at the nucleotide level. In more than 600 comparisons of non-coding DNAs aligned by orthologous exonic sequences on human chromosome 7, pig (and cow, cat and dog) sequences consistently grouped closer to human and non-human primate sequences than did rodent (mouse and rat) sequences [Green, 2002].

The Swine Maps: Linkage, Physical and Phenotypic. The porcine research community has a long history in quantitative genetics, and more recently in genomics research. The genetic contribution of many multi-genic traits in pigs is well documented and this knowledge has

provided the basis for the identification and mapping of a growing number of quantitative trait loci (QTL). The first linkage maps were published in 1994 and have been used to identify chromosomal regions that influence quantitative traits affecting growth, body composition, reproduction and immune response [reviewed in Bidanel and Rothschild, 2002]. Numerous cDNA libraries have been developed from many different tissues at different physiological stages and more than 100,000 ESTs have been deposited at The Institute for Genetic Research (TIGR, a public database). The TIGR cluster analysis on all EST sequences has generated > 50,000 non-redundant gene indices that are routinely used by the international research community. A subset of 1,000 ESTs has been selected for sequencing predicted introns (based upon human genomic sequence) to identify SNP for mapping on the linkage map. The same set of ESTs has been placed on the RH map [Hawken et al., 1999; Meyers et al., 2004]. The EST mapping effort is designed to improve the human-pig comparative maps. The radiation hybrid map of the porcine genome integrates microsatellite markers with ESTs selected from BLAST hits with human sequences. This map contains 5,500 STS markers and is currently being assembled with 2,000 comparative markers (http://imprh.toulouse.inra.fr). A coordinated international effort has been initiated to develop a porcine BAC map. An international consortium was developed to construct the porcine BAC map. The Sanger Institute, USDA-ARS, Roslin Institute, INRA and the University of Illinois are the current participants (www.swinegenomics.com).

High Resolution Mapping of QTL by Linkage Disequilibrium (LD) Analysis. Historically two methods have been utilized to map QTL. These have included linkage analysis and LD. Most QTL studies using families derived from divergent crosses and low-resolution maps have used linkage analysis to localize specific traits to chromosomal regions of < 10 cM. LD in contrast is highly informative but is dependent upon a high-resolution map but can incorporate both existing population and family data. The development of high-density SNP maps would thus provide a powerful approach for multi-point fine mapping and in particular traditionally difficult traits to map such as animal health. Recently Fan and Xiong [2002] have shown that a regression approach to mapping QTL by LD using population data can be achieved with a complex trait (disease in their example) with a sample size of 250 individuals with moderate LD. Thus, the utility of LD for mapping QTL is highly significant since most QTL studies are limited in the number of phenotypes scored (the number of animals that have been evaluated for a particular disease within a herd). In the analysis of complex diseases in humans, access to large families or informative populations is in most cases also rate limiting. Thus, the application of genome-wide association studies using SNPs is more effective [Nowotny et al. 2001]. LD also is referred to as “allelic association”, which is when alleles at two distinct locations in the genome are more highly associated than expected. To this end, the development of SNP-based LD maps could facilitate whole-genome association studies, leading to more efficient detection of candidate disease resistance or susceptibility genes [Nowotny et al. 2001].

Single Nucleotide Polymorphism (SNP) Discovery. The growing access to genomic human DNA sequences as well as other model organisms provides the opportunity to identify sites in which a single base-pair varies from individual to individual [Kwok and Gu 1999]. These unique single nucleotide polymorphisms (SNP) are now becoming the marker of choice to study complex genetic traits and diseases [Kwok and Gu 1999]. There are two approaches to SNP mapping that include genome-wide and candidate-gene mapping. In both of these cases, SNPs are used in an association study to establish linkage with a specific trait or disease.

Expression Profiling to Identify Animals at Risk. In addition to the genetic makeup of an animal (inherent resistance or susceptibility), the environment (housing, geographic implications, stress of mixing or transportation, and nutrition) provides significant impact into the expression of genes (level, time, tissues). Microarray technology is a powerful tool for the analysis of host-pathogen interactions as it provides an opportunity to simultaneously interrogate the transcriptional status of thousands of genes from a single animal (genetic makeup) in different enviornments. Microarrays consist of the specific and ordered placement of individual gene fragments onto a solid matrix. Labeled probes prepared from the mRNA isolated from various tissue samples (WBCs, biopsy material, bucal swabs) are then hybridized to the microarray and the probes hybridizing to individual DNA fragments are measured. The key feature of microarrays is that a single assay can measure the transcriptional response of thousands of genes to a change in their cellular state in response to changes in their environment (i.e. disease, nutrition, stress). By utilizing the power of microarrays to simultaneously quantify expression levels of thousands of genes, it is possible to comprehensively examine the regulation of hundreds of biochemical processes within the intestinal epithelium.

THE HOLY GRAIL: GENETIC SELECTION FOR DISEASE RESISTANCE

Two major issues confront the usage of genomic information to improve pig health. The first is developing content information that provides genetic information (either sequence or expression based) that can identify animals at risk for a given disease. This “content” is being developed both at the experimental level as well as on-site in swine operations. The second issue is developing high-throughput, low-cost technology platforms that permit the broad utilization of genomic information.

DESIGNER MEATS: GENETIC SELECTION FOR CONSUMER PREFERENCES

First Fruits of SNP Discovery. The challenge facing animal breeding is to capture the growing information from livestock QTL mapping studies. Most studies locate putative QTL to within a 20 cM chrosomosal interval which is clearly too large to fine map through traditional backcrossing. Thus, new approaches that incorporate low cost, robust, high-throughput genetic analysis are essential for translating experimental studies into commercial populations [Vignal et al. 2002]. Several efforts have been initiated with the use of SNPs in crop genetics. Rafalski [2002] has demonstrated that haplotype-based analysis is more informative than analysis based on individual SNPs and has more power in analyzing association with phenotypes. Goddard and Wijsman [2002] have characterized the cost-effectiveness of SNP genome scans and have concluded that to be cost-effective, SNPs should have a common allele frequency between 0.5-0.75. They also showed that no more than five loci per cluster are needed and an optimal solution may be use of maps with 2-3 SNPs per cluster. Also, relevant are the findings of Ching et al. [2002] in which they studied LD in ancestral maize populations. They concluded that there are a small number (2-8) of distinct and highly diverse haplotypes that can be distinguished, and within genes, SNP loci comprising the haplotypes are in LD with each other.

The use of SNPs has been first applied in the swine industry as the technology platform for traceability of animals from the farm to the fork. SNPs provide a robust method to genotype large numbers of animals, quickly, and is adaptable for both laboratory as well as on-site testing.

A number of studies have just been initiated to identify SNPs associated with QTLs. Recently, Van Laere et al. (2003) used SNP genotyping to map a gene associated with lean growth in

pigs. Most recently, we (www.swinegenomics.com) have initiated a SNP haplotyping project for selection of gilts with high ovulation rates.

Windows of Opportunity. Currently the U.S. research community has constructed an expression assay system containing over 13,000 pig ESTs. This microarray expression platform will be used to identify changes in gene expression as a function of: (1) tissue responses to economically important swine diseases; (2) the interplay between diseases and the animal's environment; (3) to determine whether expression profiles of WBC can be used to predict disease outbreak or the presence of sub-clinical disease; and (4) as a diagnostic tool to monitor the pig's response to therapuetics or nutrition intervention.

The most significant opportunity comes from the recent decision by the NIH to add the pig to the list of animals for complete genome sequencing (http://www.genome.gov/10002154). This scientific recognition provides the basis for creating an international consortium to secure funding to complete this initiative. When finished, this sequence will permit rapid identification of genes and targeting chromosomal regions for rapid SNP assays to create new screening tools as well as for the development of new drugs and medicines that promote animal health and performance.

FUTURE PERSPECTIVES

Clearly within the foreseeable future, real-time assessment of an individual animal's health status will be achievable. Thus, production induced stress (weaning of animals, movement of animals within the production facility, nutrition, weather, and shipment) will also be monitored through expression profiling. Inherent with this information will be nutritional and environmental intervention that affects an animal's inherent resistance to disease. Such insights will undoubtedly reduce the requirement for antibiotics as feed additives. Genomics will also provide traceability systems that will be used to address health issues and identify units that have altered animal performance, drug or antibiotic residues within a production system. Assurances of healthly, animal-based protein will be a high priority for meeting consumer demands in global markets.

ACKNOWLEDGMENTS

This work was partially supported by grants from the USDA-National Research Initiative (2002-35205-12712), the USDA Cooperative State Research Service (AG2002-34480-11828), and the USDA Agricultural Research Service (Agreement No. 58-5438-2-313).

LITERATURE CITED.

Bidanel, J. P. and M.F. Rothschild. 2002. Current status of quantitative trait loci mapping in pigs. *Pig News and Information* **23**:N39-N54.

Ching, A., K.S. Caldwell, M. Jung, M. Dolan, O.S. Smith, S. Tingey, M. Morgante and A.J. Rafalski. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbreed lines. *BMC Genetics* **3**:19-33.

Ellegren, H., B.P. Chowdhary, M. Johansson, L. Marklund, M. Fredholm, I. Gustavsson and L. Andersson. 1994. A primary linkage map of the porcine genome reveals a low rate of genetic recombination. *Genet*. **137**:1089-1100.

Fan, R. and M. Xiong. 2002. High resolution mapping of quantitative trait loci by linkage disequilibrium analysis. *Eur. J. of Human Genetics* **10**:607-615.

Goddard, K.A.B., P.J. Hopkins, J.M. Hall and J.S. Whitte. 2000. Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. *Am. J. Hum. Genet*. **66**:216-234.

Green, E. 2002. Comparative Sequencing of targeted regions in multiple vertebrates: Reconnaissance for Future Genome Explorations. Plenary Lecture at Plant, Animal & Microbial Genome meeting, San Diego, Ca.

Hawken, R.J., J. Murtaugh, G.H. Flickinger, M. Yerle, A. Robic, D. Milan, J. Gellin, C.W. Beattie, L.B. Schook and L.J. Alexander. 1999. A first generation porcine whole-genome radiation hybrid map. *Mammalian Genome* **10**:824-830.

Kwok, P.-Y. and Z. Gu. 1999. Single nucleotide polymorphism libraries: why and how are we building them? *Molecular Medicine Today* **5**:538-543.

Nezer, C. L. Moreau, D. Wagenaar and M. Georges. 2002. Results of a whole genome scan targeting QTL for growth and carcass traits in a Pietrain X Large White intercross. *Gen. Sel. Evol*. **34**:371-387.

Nowotney, P., J.M. Kwon and A.M. Goate. 2001. SNP analysis to dissect human traits. *Current Opin. in Neurobiol.* **11**:637-641.

Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology* **5**:94-100.

Rettenberger, G., C. Klett, U. Zechner, J. Kunz, W. Vogel and H. Hameister. 1995. Visualization of the conservation of synteny between humans and pigs by heterologous chromosomal painting. *Genomics* **26**:372-378.

Rink, A., E.M. Santchi, K.M. Eyer, B. Roelofs, M. Hess, M. Godfrey, E.K. Karajusuf, M. Yerle, D. Milan and C.W. Beattie. 2002. A first-generation EST RH comparative map of the porcine and human genome. *Mammalian Genome* **13**:578-587.

Rohrer, G.A., L.J. Alexander, Z. Hu, T.P.L. Smith, J.W. Keele and C.W. Beattie. 1996. A comprehensive map of the porcine genome. *Genome Res* **6**:371-391.

Van Laere, A.-S., M Nguyen, M. Braunschweig, C. Nezer, C. Collette, L Moreau, A.L. Archibald, C.S. Haley, N. Buys, M. Tally, G. Andersson, M. Georges and L. Andersson. 2003. A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. Nature **425**:832-836.

Vignal, A., D. Milan, M. San Cristobal and A. Eggen. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol*. **34**:275-305