The porcine genome initiative: Implications for digestive physiology

L.B. Schook *

Departments of Animal Sciences, Pathobiology and Nutritional Sciences, Regenerative Biology and Tissue Engineering Theme, 1201 West Gregory Drive, 383 ERML, College of Agricultural, Consumer and Environmental Sciences, Institute for Genomic Biology, University of Illinois at Urbana-Champaign Urbana, IL 61801, United States

Abstract

Completion of the pig genome sequencing will celebrate the Chinese Year of the Pig in 2007. The International Swine Genome Sequencing Consortium has established an integrated approach to create genome sequences, chromosomal contigs, whole-genome libraries, full-length cDNA libraries and to generate single nucleotide polymorphisms (SNPs). The project is being conducted in collaboration with The Wellcome Trust Sanger Institute. Project information as well as monitoring progress of the project can be viewed at http://www.sanger.ac.uk/Projects/S_scrofa/. During the next year, numerous physical maps and functional genomics analyses will be published. These advances will have significant implications for studying digestive physiology. Examples illustrating as to how effective genomic approaches for addressing critical questions in digestive physiology are presented. This includes approaches showing how genomics can be used to monitor the process of GI colonization and the effects of various nutrients are explored. The development of comparative genome maps also permits the incorporation of experimental studies in other species (mice and humans) to identify new hypotheses for testing in the pig. This comparative genomics approach further supports the use of the pig in biomedical research studies for human clinical studies. As this new genomic information becomes available, the necessity to create interactive databases, that provide portals for using genomic information to create gene based hypotheses will become rate limiting. Thus, the digestive physiology community contributions towards building web-based interfaces are essential.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Pig genomics; Functional genomics; DNA genome sequence; Physiology; Gut

1. Introduction

Completion of the human genome sequence provides a foundation for understanding genetic complexity and how it contributes to diverse phenotypes and disease. It is clear that model organisms such as the pig will continue to play an invaluable role in the synthesis of this understanding. The pig represents an evolutionary clade distinct from primates or rodents and thus, provides considerable power in the analysis of DNA sequence and phenotypic diversity. The pig, a domesticated eutherian mammal, has co-evolved with humans and represents a taxum with diverse selected phenotypes (Bidanel and Rothschild, 2002). The pig has played a central role in the scientific and medical communities, thus providing scientific justification for understanding
the porcine genome with respect to physiological models of growth and development, health, and reproduction (Tumbleson and Schook, 1996).

2. Comparative gene mapping: leveraging the mammalian genomic sequencing projects

The pig (Sus scrofa domestica) genome is of similar size (∼2.6×10⁹ bp), complexity and chromosomal organization (2n=38) as the human genome. Comparative maps have indicated that the porcine and human genomes are more similarly organized than when either is compared to the mouse (Fronicke et al., 1996). 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 the human and mouse (Rettenberger et al., 1995). Furthermore, the organizational similarities between the human and porcine genomes are reflected in similarities at the nucleotide level. In more than 600 comparisons of non-coding DNAs aligned by orthologous exonic sequences on human chromosome 7 (HSA7), pig sequences consistently grouped closer to human and non-human primate sequences than did rodent (mouse and rat) sequences, reflecting that rodent genomes appear to be evolving at a faster rate than those of primates, carnivores and artiodactyls (Thomas et al., 2003). Recently, we compared the genome organization of eight phylogenetically distinct species from five mammalian orders in order to address fundamental questions relating to mammalian chromosomal evolution. We observed that the rates of chromosome evolution within mammalian orders were found to increase since the Cretaceous/Tertiary boundary and nearly 20% of chromosome breakpoint regions were reused during mammalian evolution; these re-use sites are also enriched for centromeres (Murphy et al., 2005).

3. The swine maps: linkage, physical, and DNA sequencing

Over the past decade, tremendous progress has been made in mapping and characterizing the swine genome. Currently, moderate to high-resolution genetic linkage maps containing highly polymorphic loci have been produced using independent mapping populations (Archibald et al., 1995). Additionally, physical mapping methods such as somatic cell hybrid analysis, in situ hybridization and ZOO-FISH (Chowdhary et al., 1998) have been employed to enrich the gene map and to perform comparative analysis with map-rich species such as the human and mouse. To date, over 5000 mapped loci have been assigned to the pig genome (http://www.toulouse.inra.fr/lgc/lgc.htm). Whole-genome radiation hybrid (WG-RH) panels (7500 and 12,500 rad) have been generated for swine (Hawken et al., 1999; Yerle et al., 2002) resulting in yet another rapid increase in the number of comparative mapped loci. More recently, the swine genomics community has acquired access to resources such as bacterial artificial chromosome (BAC) libraries (Anderson et al., 2000; Fahrenkrug et al., 2001; Rogel-Gaillard et al., 1999; http://bacpac.chori.org/porcine242.htm) providing approximately 35X coverage of the swine genome. These BAC resources have facilitated the production of high-resolution physical maps in specific chromosomal regions (Rogel-Gaillard et al., 1999; Milan et al., 2000) that support construction of sequence-ready mapping resources for the porcine genome. Finally, the first whole-genome shotgun sequencing efforts have been reported (Wernersson et al., 2005) which will provide new opportunities for SNP discovery.

Collaborations with colleagues at INRA, the Roslin Institute and The Wellcome Trust Sanger Institute have led to the development of a porcine BAC map with 20X coverage using fingerprinting and BAC end-sequencing (Meyers et al., 2005; Rogatcheva et al., 2004; Rogatcheva et al., submitted for publication; Schook et al., 2005). These resources (http://sanger.ac.uk/Projects/S_scrofa/) supported the selection of a minimum tilling path (MTP) of BAC clones to complement a whole-genome shotgun sequencing approach. This MTP has permitted the “piggy BACing” of human sequences to construct a high-resolution (approximately 1 Mb) comparative map of SSC13 that represents a fusion of HSA3 and HSA21 (Rogatcheva et al., 2004, submitted for publication).

4. First fruits of gene discovery

During the past year, a number of studies in pigs and cattle have been completed to define mutations (and their single nucleotide polymorphisms, SNPs) and SNP associations with QTL that affect fitness traits (Mackowsky et al., 2005), growth and carcass yield (Buchanan et al., 2005). These findings clearly support efforts to create a haplotype map to dissect growth and development traits. Also relevant to is developing SNP discovery in porcine expressed genes (Fahrenkrug et al., 2001). Using locus specific amplification and comparative sequencing they generated PCR products and allelic information from their resource population. They were able to discover 1650 SNPs in 403 amplicons. Finally, using MS data, McRae et al. (2002) have studied LD in domestic sheep. They were able to demonstrate high
levels of LD that extended for tens of cM in two populations and concluded that the prospects of LD mapping in livestock are encouraging but that disequilibrium (D') may be skewed when rare alleles are present (or typing errors). They recommended that the statistical significance of LD is used in conjunction with coefficients such as D' to determine the true extent of LD.

5. Genomic sequencing: creating a single, integrated physical map

The Swine Genome Sequencing Consortium (SGSC) was formed in September 2003 by academic, government and industry representatives to provide international coordination for sequencing the pig genome. The SGSC’s Mission is to advance biomedical research for animal production and health by the development of DNA-based tools and products resulting from the sequencing of the swine genome that will be available to all parties at no cost. During the past two years, the SGSC has met biannually to develop a strategic roadmap for creating the required scientific resources, to integrate existing physical maps, and to create a sequencing strategy that captured international participation and a broad funding base (Schook et al., 2005). These interactions have supported the integration of physical mapping data with the goal of creating a MTP to serve as the genome-sequencing template. In addition, whole-genome (WG)-shotgun libraries have been constructed and are currently being sequenced in various laboratories around the globe. We have constructed a physical map of the pig genome by integrating restriction fingerprints and BAC end sequences generated from 4 BAC libraries with radiation hybrid markers, genetic markers and contig alignments to the human genome. The map covers the 18 pig autosomes and the X chromosome in 176 contigs. The level of map contiguity is higher than that achieved using similar approaches with other genomes and chromosome 13 is represented by a single contig (Humphray et al., submitted for publication). The map is accessible through WebFPC (http://sanger.ac.uk/Projects/S_scrofa/) and represents an entry point for rapid electronic positional cloning of genes and fine mapping of QTLs. The map provides a resource for the selection of a minimum tilepath for clone based sequencing and for targeted SNP and functional genomics studies and has been used for our SNP Discovery Program.

6. Nutritional genomics

Genomics is the scientific study of structure, function and interrelationships of both individual genes and the genome in its entirety. Fadiel et al. (2005) provide an excellent review including relevant website resources that support genomics and its use in life science research. These genomic tools and methods (nutrigenomics) provide new approaches and challenges for the nutrition and related food industry (Van der Werf et al., 2001). Applications can identify specific biomolecular molecules controlling biochemical processes or they can be used to access the response(s) to these molecules. Thus, new bioactives (novel functional foods) can be developed, including rapid and mechanistic approaches for food safety evaluation, and the detection of microbiota, (even if not cultureable). Perhaps the largest opportunity will be developing a metabolomics approach for nutrition and digestive physiology that is hypothesis-driven to evaluate host responses to food and nutrients.

As a whole, biological research has recently shifted its focus from reductionism to holistic approaches to study complex systems, a strategy often termed “systems biology”. Nutritional research has progressed similarly, incorporating phenotype, genotype, genomics, and bioinformatics into an integrative molecular research model to study digestive physiology and the impact of nutritional status on health and disease (Kuzmuk et al., 2005; Hauck et al., 2005; Swanson and Schook, 2006). Given the importance of livestock and companion animals as large animal models for humans, many of these genomes including that of the dog, cow, cat, chicken and pig have been sequenced or are being sequenced (Womack, 2005). While anatomical and physiological similarities are noted between these species, genomic approaches permits hypotheses based on sequence and gene expression variation to predict physiological outcomes. This review aims to reiterate the importance of the pig model in gastrointestinal research (digestive physiology) and its relevance to both agricultural and biomedical research.

Clinically there are many situations where the nutritional status of the patient has a direct impact on their health. Genomics provides a new approach to develop models to define the underlying mechanisms. For example, formula-fed infants are known to have more episodes of acute diarrhea and intestinal infection than do breast-fed infants (Milo et al., 2004). Thus, recent studies have focused on the impact of various formula supplements to alter gene expression in the neonatal digestive tract. The impact of infant formula supplements (fermentable substrates) on the gene expression of proinflammatory cytokines can be determined by monitoring proinflammatory cytokine RNA abundance in the digestive tract and circulating IL-6 activity in the serum. Although no significant mRNA
differences were observed at the selected time points, insights into the relative levels of gene expression were determined. Future studies using laser capture microdissection for gene expression studies of the digestive tract can permit more targeted approaches to gene expression in specific cell types rather than intestinal regions. Extension of these studies has included the addition of fiber to infant formulas to reduce the recovery time following pathogenic infection of infants. The working hypothesis is that fermentable fibers will reduce infection-associated symptoms and enhance intestinal structure and function in neonatal piglets. Using methylcellulose, soy polysaccharides or fructooligosaccharides reduced *S. typhimurium* infection and enhanced intestinal function could be obtained (Correa-Matos et al., 2003). In humans, a functional genomics approach has supported expression profiling and defining changes in gene expression following iron-induced oxidative stress in the small intestine (Troost et al., in press). Human volunteers were used to isolate duodenal tissue sampling by gastroduodenoscopy, a perfusion catheter was inserted orogastrically to perfuse a 40 cm segment of the proximal small intestine with saline and with either 80 or 400 mg iron as ferrous gluconate. After the intestinal perfusion, a second duodenal tissue sample was obtained. Thiobarbituric acid reactive substances, an indicator of lipid peroxidation increased significantly and dose-dependently in intestinal fluids within 30 min. The expression of 89 gene reporters was significantly altered by both iron interventions. Functional mapping showed that both iron dosages mediated 6 distinct processes, three involving G protein receptor coupled pathways and the others associated with cell cycle, complement activation and calcium channels. The utility of a functional genomics approach will soon be available to porcine models as the pig genome sequencing and the annotation of pigs is completed.

The eukaryotic GI tract is colonized by a vast number of bacteria, where the commensal microbiota play an important role in defining the healthy gut (Mutch et al., 2004). The influence of the commensal microflora on regions of the host GI tract’s transcriptome is a subject of intense study. Gene expression profiles of the corpus jejenum, descending colon and rectum of conventional and germ-free mice were examined using the Affymetrix Mu74Av2 GeneChip. Differentially regulated genes were identified and Gene Ontology functions were assigned. Cluster analysis revealed that the microbiota had the greatest effect (affecting 267 genes) on the jejunum and the lowest on the rectum had the fewest (137 genes). Clustering genes by function demonstrated that, despite the large number of differentially regulated genes, the residential microbiota most significantly modified genes involved in immune and water function pathways along the length of the gut. Region-specific communication between the host and microbiota were identified in the corpus and jejunum, where kallikrein and apoptosis regulator activities were modulated, respectively. These findings identify important interactions between the microbiota and gut tissues affecting transcription and implicate a coordinated microenvironment.

Mice with a null mutation in their gastrin hormone (GH) gene have impaired gastric acid secretion (Jain et al., 2006). These investigators evaluated changes in the acid-secreting parietal cell in gastrin-deficient mice. Transcription profiling showed several transcripts encoding parietal cell proteins involved in gastric acid secretion were decreased. Comparison of gene expression in the GH-knockout and wild type mice identified 47 transcripts that differed by greater than or equal to 2 fold, suggesting that gastrin affects parietal cell gene expression in a specific manner. The differentially expressed genes included several genes in signaling pathways with a substantial number (20%) known to be target genes of *Wnt* and *Myc* pathways.

The intestinal brush border fructose transporter GLUT5 (SLC2A5) typically appears in rats after weaning is completed (Cui et al., 2004). However, the precocious consumption of dietary fructose or *in vivo* perfusion for 4 h of the small intestine with high fructose specifically stimulates *de novo* synthesis of GLUT5 mRNA and protein prior to completion of weaning. Since no intermediary signals linking the substrate (fructose) to GLUT5 transcription are known, transcriptional profiling was used to identify genes whose expression levels were altered as a result of fructose perfusion. Expression of GLUT5 and Nap12b (intestinal Na+-dependent phosphate transporter) dramatically increased and decreased respectively after 4 h of perfusion. Expression of >20 genes including key gluconeogenic enzymes was increased. These findings suggest that the gluconeogenic enzymes and their common metabolic intermediate fructose-6-phosphate may regulate fructose metabolism and GLUT5 expression in the small intestine.

Bacterial colonization modulates postpartum maturation of the gut resulting in an efficient barrier to luminal antigens and bacteria. Thus, the use of broad-spectrum antibiotics in pediatrics and commercial livestock operations should alter bacterial colonization and thus impair gut maturation and function. Schumann et al. (2005) tested this hypothesis using suckling rats that received a daily gavage of antibiotic (Clamoxyl, amoxicillin) or saline postnatal day 7 until day 17 or
day 21. Luminal microbiota composition and global gene expression profiles from the small intestine and colon of each group were determined. Antibiotic treatment resulted in an almost complete eradication of Lactobacillus in the whole intestine and drastically reduced total aerobic and anaerobic bacteria. The global gene expression analysis revealed that antibiotic treatment affected the maturation process of 249 and 149 genes in the proximal and distal small intestine, respectively, and 163 genes in the colon. Expression of genes coding for Paneth cell function (defensins, matrilysin and phospholipase A2) was significantly downregulated. The MHC class Ia and II genes associated with antigen presentation were also depressed. Conversely, mast cell proteases expression was upregulated. These results suggest that early antibiotic treatment affecting the microbiota can significantly alter gene expression and gut function and potentially compromising animal health.

7. Genomic approaches to pig digestive physiology

Genome research in livestock species is contributing to our understanding of chromosome evolution and to informing the human genome (Womack, 2005). The introduction of “omics” to the nutritional community highlights how these tools can be used to create nutritional models to monitor the interactions of genetics, nutrition, age and exercise (Swanson et al., 2003). This permits new hypotheses driven by proposed changes in gene activity to be developed and support the exploration of how diet and lifestyle can affect an animal’s health and performance.

This includes the application of DNA-based tools (monitoring both sequence variation and gene expression levels) to animal health and production. This spectrum of study ranges from DNA sequence, to understanding the relationship between sequence and expression variation contributing to phenotypes. Moreover, the tools to test hypotheses will continue to emerge. Thus, the identification of genes through linkage analysis or expression profiling can be directly tested using RNA interference and the use of cloned genetically modified animals. The understanding of host-pathogen (microbe) interactions associated with animal and human health will be the next frontier. Currently the first generation porcine oligonucleotide set, representing 13,297 cDNAs and ESTs has been designed by Qiagen-Operon for transcriptional profiling (Zhao et al., 2005). This microarray is comprised of 70-mer oligonucleotide (Qiagen-NRSP8 array) and has been recently validated and shown to be informative and specific and thus an appropriate tool for porcine functional genomics studies and for modeling the pig in physiological questions. Recently Machado et al. (2005) have used this functional genomics approach to monitor changes in gene expression in jejunal Peyer’s patches in both juvenile and adult pigs.

8. Conclusions and future perspectives

The applications of genomic tools to support digestive physiology will continue to emerge. The transformation of increasing genomic information into knowledge and the creation of new hypotheses requires new approaches to the development of bioinformatics tools and integrated data bases. The growing need for innovative computational approaches and integrative models to link genes to complex function has been a critical component of the rat physiology community (Cowley, 2005). They have proposed a broad goal of physiological genomic research to link genes to their functions using appropriate experimental and computational techniques (Kwittek and Jacob, 2005). Vast quantities of gene sequence and expression data are being generated but the interpretation of this data requires the integration of information derived from many diverse sources. The Rat Genome Database (RGD; http://rgd.mcw.edu) provides computational tools and strategies specifically supporting the goal of linking genes with their physiological functions in rats and through the use of comparative genomics to humans and mouse. These computational tools support strategies for the application of these resources in physiological genomics.

The future of the pig digestive physiology community to fully utilize genomics also will require a similar approach of developing an international integrated database. Efforts are currently underway at the University of Illinois in conjunction with the National Center for Supercomputer Applications (NCSA) to establish a pig genome database similar to that currently being used by the rat physiology community (www.swinegenomics.com). The challenge to the broader community is to develop strong communication linkages and enhance the exchange of data between global researchers. Through the leadership of the digestive physiology group significant progress will be attainable.

Acknowledgements

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). The
author also wishes to recognize the support of the College of Agricultural, Environmental and Consumer Sciences, University of Illinois.

References


Humphray, S., Scott, C., Clark, R., Mann, B., Plumb, R., et al., submitted for publication. An integrated sequence ready BAC map of the pig (Sus scrofa domestica) genome.


