RESEARCH REPORT
SWINE HEALTH

Title: Genome sequencing of *Haemophilus parasuis* for improved swine health - **NPB #07-038**

Investigator: Gregory J. Phillips

Institution: Iowa State University

Date Submitted: July 8, 2008

Industry Summary:

Although *H. parasuis* remains a concern to producers in the United States, it remains a challenging microorganism to study and thus control. For example, since *H. parasuis* is difficult to culture and exists as multiple serotypes, effective, cross protective vaccines are not available. Identifying all of the genes in *H. parasuis* by DNA sequencing represents a powerful new strategy to better understand how the bacterium grows, its mechanisms of virulence, and to potentially identify new strategies for its detection and elimination. Fortunately, the recent advent of new sequencing technologies has dramatically reduced the cost of whole-genome sequencing. We have used one of these new sequencing systems (454, FLX system) to determine the DNA sequence of the majority (>99%) of the *H. parasuis* strain 29755 genome. The raw sequence information is continually being analyzed to understand how this microorganism grows and causes disease. Specifically, through collaborations with swine respiratory disease researchers at the National Animal Disease Center, we have identified and begun to characterize new genes encoding outer membrane proteins that may be used in future vaccine trials as a broadly protective antigen. As additional *H. parasuis* genomes are sequenced, we will be able to perform comparative studies to identify genes that are unique to highly virulent strains, as well as sequences that are shared among multiple isolates. These sequences can potentially be used in PCR-based diagnostic assays to differentiate between *H. parasuis* strains. Although the formal sequencing phase of this project is completed, analysis of the data continues, both by us and by *H. parasuis* researchers world-wide. The draft sequence information has been uploaded to the National Center for Biotechnology Information Genome Project Center, a depository of sequence information maintained by the U.S. government. This allows investigators to access to the draft sequence of the *H. parasuis* genome for their own analysis. This draft genome represents the first sequence of *H. parasuis* available to the public and should yield new insights and applications into *H. parasuis* for years to come.

Contact information: Greg Phillips, Ph.D./Department of Veterinary Microbiology and Preventive Medicine, Iowa State University, 1802 University Boulevard, Ames, IA 50011
Ph., 515-294-1525, Fax., 515-294-1401, e-mail, gregory@iastate.edu.

Scientific Abstract

This respiratory pathogen *Haemophilus parasuis* is the causative agent of porcine polyserositis, or Glässers disease, and is now considered a significant swine health problem in the United States. Despite its importance, many fundamental features of *H. parasuis* remain poorly understood, including determinants of host specificity, virulence factors, and the genetic basis for the multiple serotypes characteristic of the species. Unfortunately, *H. parasuis* is challenging to study by traditional microbiological methods since it is difficult to culture directly from its natural environment and exhibits extensive strain variation. *H. parasuis* infections are also frequently associated with other bacterial pathogens, further complicating its detection and diagnosis. Our lack of knowledge and difficulty in working *H. parasuis* makes it difficult to consistently detect, prevent, and treat respiratory diseases in pigs. Genomic DNA sequencing represents a means to improve our understanding of *H. parasuis* virulence, as well as a means to identify new vaccine targets and sequences for strain detection and differentiation.

Introduction

Segregated Early Weaning (SEW) and related systems, while lessening the severity of many infectious diseases, have ironically also fostered the emergence of respiratory pathogens, including Haemophilus parasuis. H. parasuis the causative agent of porcine polyserositis, or Glässers disease, is now considered a significant swine health problem world-wide. While H. parasuis is most devastating in the nursery unit, it can also enter the herd in the finishing unit or after introduction of breeding stock, resulting in high morbidity and mortality at all ages of production. Symptoms include high fever, respiratory distress, coughing, apathy, loss of appetite and anorexia. Surviving animals also can develop infectious arthritis and inflammation of the elbow and stifle joints.

Given the nearly 70 million market hog inventory of the U.S. (data from the National Agricultural Statistics Service), H. parasuis has been predicted to result in loss of over 150 million dollars per year to the U.S. swine industry. Since H. parasuis is difficult to culture in vitro, much of the basic characterization necessary to develop new vaccines or diagnostic tools is lacking. Further confounding these efforts is their phenotypic and underlying genetic heterogeneity. At least 15 distinct serovars have been identified that vary widely in virulence, site of isolation and disease manifestations. Multiple types, based on serology or restriction site polymorphisms, can be found in the same heard and even in the same animal. Without adequate molecular tools to distinguish among types it has not been possible to accurately identify and characterize virulent isolates.

The development of new therapeutic agents has been limited since few specific virulence factors have been identified. H. parasuis isolates possessing characteristics associated with virulent strains can also be obtained from healthy pigs, further frustrating attempts to identify the basis for virulence. A number of bacterial and viral agents are known to enhance and complicate the diagnosis and study of Glässer's disease, making it difficult to assess the contribution of H. parasuis to disease. New diagnostic tools are needed to improve disease diagnosis by distinguishing between H. parasuis and other respiratory pathogens. More specific probes to distinguish between the multiple serovars of H. parasuis and un-typable variants also need to be developed. Currently molecular tools for definitive identification of H. parasuis and subsequent strain genotyping are limited to a search for random polymorphisms.

Objectives:

The *objective* of this research project is to determine the genomic DNA sequence of the swine pathogen *H. parasuis* leading to the development of new diagnostic tools and vaccines for improved animal health.

Materials & Methods

We initially determined the genome size of *H. parasuis* strain 29755 by pulsed field gel electrophoresis (PFGE) and obtained a genome size of ~1.9 Mb, slightly larger than *H. influenzae* (1.8 Mb).

H. parasuis 29755 was cultured on blood agar plates and isolated genomic DNA using a standard protocol recommended by 454 Life Sciences. Genomic DNA (5 mg) was sent to SUNY-Buffalo Genome Sequencing Facility for sequencing using 454 technology. This facility recently acquired a 454 Sciences Corporation sequencing instrument and provided sequencing services to us at an extremely reasonable price.

Data were made available to us from the SUNY-Buffalo Sequencing Center in standard FASTA format and will include quality scores to allow us to assess the reliability of the sequencing reads. The individual sequencing reads was assembled into longer contigs using the *Newbler* program integrated into the 454 sequencing system. 454 sequencing technology also provided 20x-30x depth of coverage, superior to the ~8x coverage generated through shotgun sequencing methods.

Annotation was done using the NCBI annotation programs. To use the programs, the *H. parasuis* genome sequence was uploaded to the NCBI Genome Center by ftp. Using standard programs, putative genes were identified based on similarity with other known sequences. The data were also analyzed to identify potential mutations and duplications in the sequence. These were discarded until additional sequence analysis can be performed. Predicted gene products with a pair-wise match to a hypothetical protein from another species, were named "conserved hypothetical protein", while gene products with no database matches remained "hypothetical protein".

While automated annotation process is extremely good, our experience tells us that the final product still requires significant manual input for clean up, particularly to reconcile poorly annotated data derived from other genome projects. This was also necessary to identify and resolve additional ambiguities in the sequence that need manual attention and further sequence cleanup. Manual annotation was facilitated by the TIGR utility Manatee, a web-based gene evaluation and genome annotation interface.

The annotated *H. parasuis* genome was scanned by bioinformatics tools to identify proteins predicted to be secreted or localized to the cell surface. Proteins localized on the cell surface are more likely to be recognized by host antibodies, many of which will be protective against bacterial infection. This information will be incorporated into a database that will allow direct comparisons of potential vaccine targets across multiple *H. parasuis* genomes, as they become available. This database is currently being developed and will be made available via a dedicated web site in the future for use by investigators in both academic and industrial settings to begin vaccine development.

Results

A total of 2,224,137 unique bases of the *H. parasuis* genome were determined by the 454 sequencing system. After the sequencing phase, the 454 system assembled the randomly generated DNA sequences into contiguous sequences (contigs) by searching for overlapping sequence information. This resulted in a total of 246 contigs. The largest contig was 157,179 bases, with an average contig size of 13,514 bases. On average, 600 bases were determined per sequencing reaction. Since the amount of genomic sequence information exceeds the predicted genome size of *H. parasuis*, as obtained by pulsed-field gel electrophoresis, we predict that the sizes of the gaps are small and that we are not likely missing significant amounts of sequence information.

An account was established with the NCBI Genome Project Center and the sequence information has been formatted for GenBank submission. The sequence, along with the annotation of the predicted genes is available at:
(http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj&cmd=Retrieve&dopt=Overview&list_uids=20127).

Although additional sequencing will be required to close the remaining gaps and generate a complete genome, the majority of *H. parasuis* genes were identified from this draft sequence. We have initiated a collaboration with the swine respiratory disease group at the National Animal Disease Center (NADC) to identify new targets for vaccine development and diagnostic reagents. We have also identified all putative genes that encode membrane-bound proteins as they may represent new targets for vaccine development. The outer membrane proteins P2 and P5 are of particular interest as potential vaccine targets, as well use of their genes for strain discrimination. Previous attempts to clone the genes using PCR primers based on the *H. influenzae* genome sequence were unsuccessful, however, primers were successfully designed using the strain 29755 sequence.

Discussion

The genomic sequence of *H. parasuis* will contribute to improved pig health in multiple ways, including addressing the need for more effective cross-protective vaccines. Currently, vaccination against one or more serotypes does not guarantee protection against other serotypes. The prevalence of serovar diversity, including a high percentage of un-typeable strains, accentuates the limitations of the vaccines currently available. Comparison of genes from strain 29755 with similar genes from other isolates will likely lead to identification of new sequences shared among different *H. parasuis* strains. In particular, the genes encoding the P2 and P5 outer membrane proteins have been identified is several *H. parasuis* strains and are currently being analyzed for their potential as vaccine candidates

The analysis of the *H. parasuis* 29755 genome, and the P2 and P5 genes, is being performed through a collaboration between the College of Veterinary Medicine at ISU and the NADC. The long-term goal of our research team is to better understand the biology of *H. parasuis* and ultimately develop improved diagnostic tools and improved vaccines to prevent Glässer's disease, and holds good promise to obtain future funding. We have submitted a proposal to the NSF/USDA Microbial Genome Sequencing program to request funds to complete the closure phase of the *H. parasuis* strain 29755 genome, and to sequence additional *H. parasuis* isolates to the draft phase. Despite being consistently ranked as "highly meritorious" and among the top proposals for two years, the *H. parasuis* 29755 sequencing project was not funded. The program director has indicated that support from the NPB will be vital to demonstrate the importance of *H. parasuis* to the swine industry.

The resulting genome sequences will be used for the strategy of "reverse" vaccine development that is based on the principle that antibody targets can be identified by analysis of genome sequences. The approach has yielded new vaccines against multiple pathogens where more traditional approaches of preparing killed vaccines have failed. To this end, the *H. parasuis* genome has been scanned using computational tools to identify proteins predicted to be secreted or localized to the cell surface (eg., P-Sort, BOMP, Lipo-P). Proteins localized on the cell surface are more likely to be recognized by host antibodies, many of which will be protective against bacterial infection and/or disease. Future studies will be required to test the potential of these proteins as vaccine targets.

Additional functional genomics studies are now possible with the availability of the draft *H. parasuis* sequence. For example, high-density microarrays may be constructed to monitor changes in transcription of *H. parasuis* genes during different growth phases and during infection and to make precise comparisons of genome content among multiple isolates.

Publication in preparation:

Comparative analysis and characterization of the genes encoding *Haemophilus parasuis* outer membrane proteins P2 and P5. Michael A. Mullins, Karen B. Register, Tracy Nicholson, Crystal Loving, Darrell O. Bayles, Susan Brockmeier, Gregory J. Phillips and David W. Dyer.