RESEARCH REPORT
SWINE HEALTH

Title: Use of Mobile Nanopore Sequencing to Detect and Genotype Porcine Reproductive and Respiratory Syndrome Virus: NPB # 16-205

Investigator: Kevin Lahmers

Institution: Virginia Tech

Date: Submitted: 1/1/18, resubmitted 5/9/18

Industry Summary:

Porcine Reproductive and Respiratory Syndrome Virus continues to be a leading cause of disease and decreased production in the swine industry. Detection of virus is not sufficient to determine if this is modified live vaccine, continuing low level infection, or introduction of new strains of the virus that have entered a herd or facility. Therefore, this project looked to combine detection and genotyping using a newly available sequencing method. This project was able to successfully detect and genotype virus from experimentally and naturally infected pigs using serum. In experimental infections where the virus used is known, we were able to correctly identify the inoculated virus each time and were able to detect mixed infections when two samples were mixed. In addition, we have developed a method of data analysis that allows us to identify PRRS virus sequences from among the millions of sequences obtained and to classify the virus within 90 seconds. We are refining this method so that it is automated and better classifies viruses that do not match those sequences found in published databases. This sequencing technology has the ability to selectively sequence and reject unwanted genetic material. Attempts to use this approach were of limited success. Due to the high speed of sequencing reactions and relatively short read lengths, the enrichment was not sufficient to warrant the approach. Alternate enrichment approaches are funded and underway with the intent of making this as fast or faster than current PCR results and at a similar cost. In addition, to increase throughput and efficiency while providing value-added results, we are sequencing ORF5 amplicons at a materials cost similar to that of traditional Sanger sequencing but with the added benefit of being able to detect multiple viruses or modified live vaccine and wildtype virus in the same sample. If you would like to discuss these findings or further progress, please feel free to contact me.

Kevin Lahmers, DVM, PhD, DACVP
Virginia Tech University
klahmers@vt.edu
540-231-7632

Keywords: PRRS virus, sequencing, genotyping

Scientific Abstract:
Porcine Reproductive and Respiratory Syndrome Virus continues to be a leading cause of disease and decreased production in the swine industry. Detection of virus is not sufficient to determine if this is modified live vaccine, continuing low level infection, or introduction of new strains of the virus that have entered a herd or facility. Therefore, this project looked to combine detection and genotyping using nanopore-based shotgun sequencing. This project was able to successfully detect and genotype virus from experimentally and naturally infected pigs using serum. Using reverse transcription and nanopore sequencing, we were able to detect PRRSV and correctly genotype the virus used for inoculation by using a custom database and the Centrifuge metagenomic classifier. This method is able to detect mixed infections when two samples were mixed. We have demonstrated that this is repeatable with one read per 1600 total reads in multiple library preparations of the same sample. In addition, we have developed a method of data analysis that allows us to identify PRRS virus sequences from among the millions of sequences obtained and to classify the virus within 90 seconds. We are refining this method so that it is automated and better classifies viruses that do not match those sequences found in published databases. This sequencing technology has the ability to selectively sequence and reject unwanted genetic material. Attempts to use this approach were of limited success. Due to the high speed of sequencing reactions and relatively short read lengths, the enrichment was not sufficient to warrant the approach. Alternate enrichment approaches are funded and underway with the intent of making this as fast or faster than current PCR results and at a similar cost. In addition, to increase throughput and efficiency while providing value-added results, we are sequencing ORF5 amplicons at a materials cost similar to that of traditional Sanger sequencing but with the added benefit of being able to detect multiple viruses or modified live vaccine and wildtype virus in the same sample. PCR amplification and sequencing (amp-seq) allows for detection of PRRSV in lower titer samples (preliminary results) as well as detection of multiple isolates within a sample, which is an advantage to sequencing. The amp-seq approach is capable of distinguishing isolates with 99.33% sequence identity over the 604 bp product. The results of this funding will result in 2 manuscripts in preparation, 3 presentations at international meetings, and over $500,000 in additional grant funding to further this line of research in pig disease. We will make protocols and this testing available to clients in the next 2 months.

Introduction:
Since the initial identification of Porcine Reproductive and Respiratory Syndrome in North America in 1987, PRRS has had a major economic impact on the swine industry [1]. The low fidelity of RNA polymerase allows this positive-strand RNA virus to mutate frequently. In addition to mutation, recombination and immune selection lead to greater genetic diversity. In fact the European and North American strains (type I and type II, respectively) only share 50-70% nucleotides [2]. This genetic diversity has interfered with development of effective control of PRRS [3]. Variability in genetic makeup has also interfered with the ability to detect the PRRS virus using our most consistent method, polymerase chain reaction (PCR).

In addition to identifying the presence or absence of the PRRS virus, some realtime PCR assays allow the differentiation of type I and II viruses. However, PCR is limited in its ability to further classify the virus. Most recently, subclassification has been done using restriction fragment length polymorphism (RFLP) of ORF5 [1]. Studies have demonstrated that RFLP does not correlate well with lineage and may be of limited utility beyond the farm level [1, 2]. In addition, this typing requires manipulation of amplified DNA which increases the risk of cross contamination and also requires experienced personnel to interpret results. Genotyping has promise in more effectively subclassifying virus isolates, but this process is time consuming with added manipulation required and several days in order to obtain results. Clearly, an improved approach to genotyping of PRRS virus would be beneficial in attempts to control PRRS.

Until recently, all sequencing modalities required synthesis of a complementary strand of DNA with detection of addition of individual nucleotides by color, light or pH detection. These sequencing instruments are expensive often costing hundreds of thousands of dollars making them difficult to obtain for routine diagnostics. In addition, turn around time is at least 1-2 days and often longer. Lastly, the read length of most of these sequencing approaches is limited to 500 bp or less which does not allow for sequencing of entire PRRS genomes in a single read and requires bioinformatic approaches to reconstruct the fragments of sequence DNA. Despite these challenges, deep sequencing has been used to detect and characterize disease agents in food animal species [4-11]. It has been used in detecting and characterizing recent diseases in the swine industry such as Porcine Epidemic Diarrhea Virus and Seneca Valley Virus [12-14].

A new approach to sequencing has recently been developed by Oxford Nanopore Technologies. Their MinION does not use sequencing by synthesis but instead is based on a small protein pores embedded in a membrane. A current passes through these pores. DNA is fed through the pores and the structural features of the nucleotides alter the current passing through individual pores. This change in current can be analyzed and interpreted as DNA sequence. Recent improvements have led to 95-99% accuracy (personal experience) and increased read speeds of >450 basepairs per second for each pore. This sequence data is available for analysis immediately rather than at the end of the run as in other sequencing approaches. This allows sequencing until desired depth or genotyping is completed before switching sample.

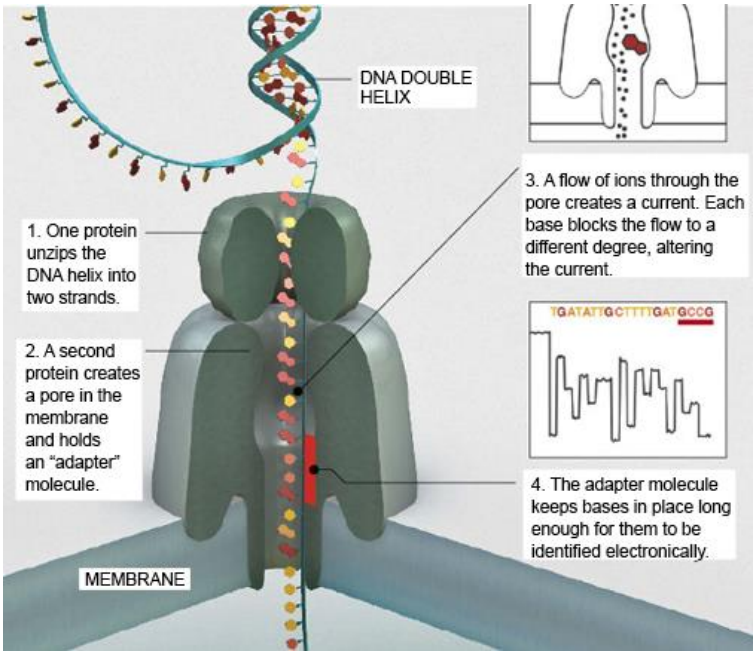

Fig. 1: Mechanism of Nanopore sequencing. As DNA is moved through a nanopore, the change in ion flow is used to identify the bases within the nanopore.
From http://www2.technologyreview.com/news/427677/nanopore-sequencing/

Objectives:
A labor efficient and cost effective method of detection and subtyping of PRRS strains with improved turn around time is needed. To accomplish this we will use a novel mobile nanopore sequencing approach combined with concurrent data analysis and positive selection of PRRS virus from serum.

Objective 1: Demonstrate efficacy and determine sensitivity of 'positive selection while sequencing' approach in identifying PRRS in biological samples. Identify depth of sequencing required to effectively identify PRRS virus spiked into serum at different concentrations. Samples will be pooled in order to determine the limit of detection in pooled samples.
Objective 2: Field validation of the 'positive selection while sequencing' approach to PRRS detection and genotyping. Field samples will be used and compared to PCR as the gold standard.

Objective 3: Streamline a dedicated sequencing pipeline to identify PRRS virus strain based on sequence of ORF5. A cloud-based tool to map sequences to reference sequences (WIMP) is available for the MinION and requires no bioinformatic training. This tool will be refined to focus on porcine pathogens, particularly PRRS virus, and to be run on the local computer.

Materials & Methods:

1.1. Samples: Porcine serum samples collected as part of routine surveillance from regional swine farms were used as the basis for this aim.

Experimental Spiking of Serum: Initial samples were PCR-negative serum or PBS spiked with 5 serial ten-fold dilutions of PRRSV (strain VR2385). The spiked samples helped determine the linearity of detection in varying virus loads. In addition, the efficacy of genotyping was compared between low- and high-titer samples. Based on this bench validation data, depth of sequence (i.e., amount of sequence) required for sensitivity equivalence equal or better than PCR was roughly determined.

1.2. Nanopore sequencing:

RNA extraction and DNA synthesis: Viral RNA was extracted from serum using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany), according to manufacturer's protocol. For quality control, extracted RNA was quantified using RNA High Sensitivity Assay kits (ThermoFisher Scientific, USA) on a Qubit Fluorometer). cDNA was synthesized using the SuperScript™ III reverse transcriptase kit (Invitrogen, Carlsbad, CA) with random oligodeoxynucleotide primers. cDNA was prepared for sequencing using the cDNA-by-ligation protocol (Oxford Nanopore) according to the manufacturer's specifications.

MinION sequencing and bioinformatics analysis: Prepared samples were loaded on the R9.4.1 and 9.4 flow cells after quality control assessment as part of the MinKnow application 1.7.4 (Oxford Nanopore). Sequences and associated metadata were locally basecalled using Albacore 2.1.3 (Oxford Nanopore). Reads were mapped locally to a custom-build index based on NCBI RefSeq viral genomic sequences using Centrifuge [15]. Metadata for PRRSV-positive were analyzed to identify the time and percent positive reads. Classified metagenomics data was visualized using Pavian.

Positive selection using "Read Until": This was developed in collaboration with the researcher who initially described this process, Dr. Matthew Loose. Briefly, the raw nanopore data are alterations in current through a pore (a.k.a., squiggle data), which are used to predict nucleotide sequence. In order to increase the speed of positive selection, instead of using base-called squiggles to target sequencing in real time, the nucleotide sequences of the targets that are to be positively selected are converted to predicted squiggle data. The predicted squiggle data is then compared, in real time, to the active squiggle data by digital time warping. Sequencing is continued for positive matches; however, negative matches are rejected by reversing the polarity of the pore. This polarity reversal not only spatially ejects the molecule from the pore, but also disrupts the sequencing chemistry for that molecule making it less likely to be resequenced. This occurs within two seconds with termination occurring within the first 500 bp.

Parameters measured: Samples were analyzed for number of sequence reads until first targeted virus detection (i.e., threshold read, Rt; see Validation Plan), read ratio (total number of viral reads/total number of reads), time until first PRRSV detection, and accuracy of the genome sequence relative to the reference genome. Limit of detection will be expressed as relative to genome copy number (as determined by RT-qPCR). Genotyping and comparison of the sequences to recent geographically isolates will also be conducted using phylogenetic analyses similar to those used in the preliminary data. These same parameters were measured while using the selective sequencing protocol "Read Until" to evaluate efficacy of enrichment.

1.3. RT-PCR: Reverse transcriptase, quantitative PCR (RT-qPCR) was performed for PRRSV using VetMax (Applied Biosystems, Foster City, CA) according to manufacturer's specifications.

Results: Report your research results by objective.

Objective 1: Demonstrate efficacy and determine sensitivity of 'positive selection while sequencing' approach in identifying PRRS in biological samples. Identify depth of sequencing required to effectively identify PRRS virus spiked into serum at different concentrations. Samples will be pooled in order to determine the limit of detection in pooled samples.

Initial experiments demonstrated the ability to detect PRRS virus sequence from samples in which virus culture was spiked into PBS and into serum. The virus was correctly identified based on the known virus spiked into the matrix. The results were similar between PBS and serum. Due to the chemistry of the reverse transcription, sequencing fragments averaged 2,000-3,000 bases. This was sufficient for accurately identifying the PRRS virus strains. However, when 'read until' was applied, it was determined that due to the speed of sequencing and the time required to identify the sequence of the cDNA fragment in the pore, at least half of the read was through the pore before it could be reversed. Therefore, the 'read until' process did not lead to sufficient enrichment to be of significant benefit. We are working to increase read lengths by alternate sample preparation as well as other forms of enrichment such as a 'bait and capture' approach on the flow cell. Depth of sequence required for detection varied from sample to sample with most samples having a positive within the first 8,000 reads or 20 minutes. However, some samples did not have positive samples early and there were not sufficient numbers of positive reads to determine with confidence that there was only one virus strain in the sample. Therefore, as enrichment procedures are being developed, we determined that 400,000 sequencing reads or 30 positive PRRS virus sequences would be the threshold for sequencing.

Objective 2: Field validation of the 'positive selection while sequencing' approach to PRRS detection and genotyping. Field samples will be used and compared to PCR as the gold standard.

We initially focused on experimentally infected pigs for this objective. This was chosen because we already had data regarding infection kinetics and the virus used to inoculate was known. We sequenced serum from pigs from three separate challenges.

Challenge	**PRRS reads**	**Unique reads**
VR2385	32	24 (to VR2385)
MN184	19	5 (to MN184)
NADC20	178	67 • SD98-163_P83 18 • 1692-98 8 • P129 7 • SDU73 4

NADC20 full length sequence is not in Genbank so was not included in our analysis pipeline. The unique reads identified were nearly identical to NADC20 and represent the best fit of the sequencing data to the database. When *de novo* sequence for NADC20 was added to the database, all 67 reads were correctly reclassified NADC20.

Additional samples were sequenced from field samples collected as part of routine diagnostic samples. In these samples, PRRS virus could be identified in relatively high titer samples, but samples with threshold cycles above 25 in realtime PCR could not consistently yield positive samples in the first 400,000 reads. Preliminary exploration of oral fluids as the test sample have yielded similar results.

Two additional items are included in this aim. In the experimental infection with MN184 we also detected high numbers of sequencing reads with porcine parvovirus 6. This sample was collected in 2013 and this virus was first described in China in 2014 [16] and subsequently in retrospective samples in the United States in 2015 [17]. This demonstrates the power of metagenomic sequencing in that it can potentially identify other infections including diseases

that have not been previously identified or known to exist in a particular species or location. Our sequencing has also identified concurrent mycoplasmal, streptococcal, Haemophilus, influenza and PEDV infections. Second, we artificially mixed samples in order to determine if we could identify co-infections with this sequencing approach. In these mixed samples, we were able to correctly identify the known strains (VR2385, MN184, and NADC20) when two or all three viruses were included. This lead us to another approach in which we amplified *orf5* as has traditionally done for Sanger sequencing and then sequenced the amplicon. We have been able to produce sequence corresponding to Sanger sequence from the amplicon at 99-100% sequence identity between Sanger and consensus sequence. We are able to correctly cluster reads with as few as 4 nucleotide differences in the 604 bp amplicon (99.3% identical). In addition, we have correctly identified mixed samples. Lastly, by multiplexing the samples using barcodes, we can perform this at roughly the cost of Sanger sequencing making it potentially an alternative with the ability to detect mixed infections that is currently not available.

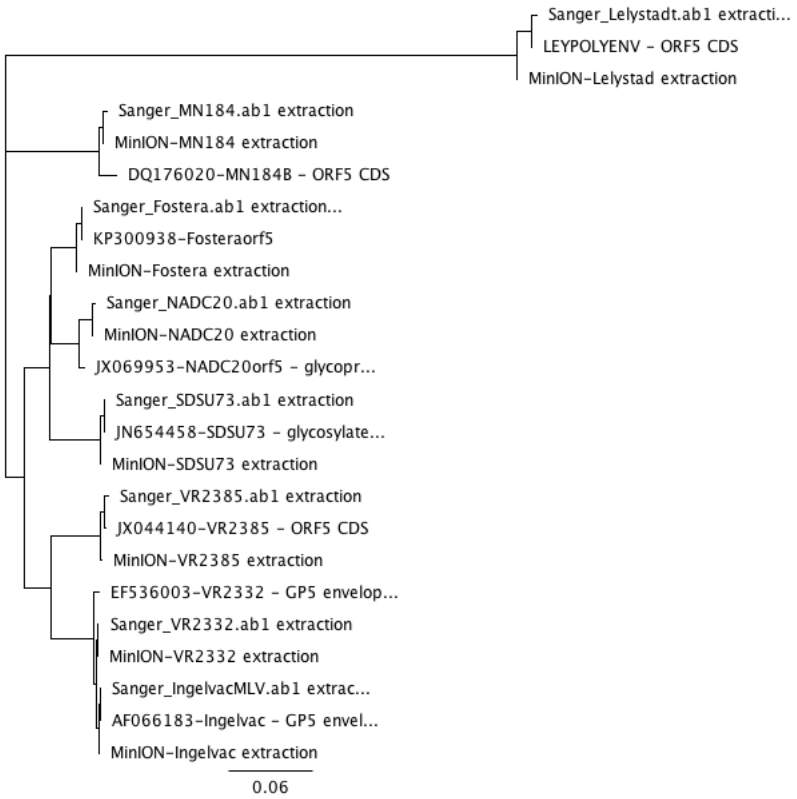

Figure 2. Comparison of multiplexed sequence of orf5 with Sanger sequence and published sequence. Sanger sequence, MinION sequence and published NCBI sequence were compared in a phylogenetic tree. All three sequences for each strain clustered appropriately.

Objective 3: Streamline a dedicated sequencing pipeline to identify PRRS virus strain based on sequence of ORF5. A cloud-based tool to map sequences to reference sequences (WIMP) is available for the MinION and requires no bioinformatic training. This tool will be refined to focus on porcine pathogens, particularly PRRS virus, and to be run on the local computer.

We have moved base-calling to local cluster computing allowing us to match basecalling process on Albacore to the speed of the sequencing as more and more rapid sequencing is available. Basecallled reads are then imported into the Centrifuge python program. This program is an open source software that allows for custom databases. We used a standard database containing representative species of all known bacteria, virus, and the human genome and created a database from all complete PRRS virus genomes in Genbank (761 genomes). We first apply the general database which gives an estimate of PRRS virus read numbers and to detect other

potential pathogens (e.g. porcine parvovirus 6). The reads are rerun through the PRRS specific database to genotype the virus reads in the sample. Each run is completed in 90 seconds or less on a basic Macbook Pro. For confirmation, reads were then BLASTed using the NCBI database. At this point, none of the Centrifuge classified reads have been categorized as misclassified by BLAST. Centrifuge classification of genotypes was highly sensitive and specific when the database included the viral sequence. When the virus is not included in the database such as wildtype virus, the closest matches are identified by the Centrifuge process. This can be improved by extracting viral reads and creating contigs using Geneious 10.1. This approach also allows for separation of two wildtype strains that do not have 100% identity in the database. We have expanded this to apply to amplicon-based sequencing of PRRS as well as other veterinary pathogens. If needed, metagenomic data can be visualized using Pavian, an open source pthyon script using R (Figure 3).

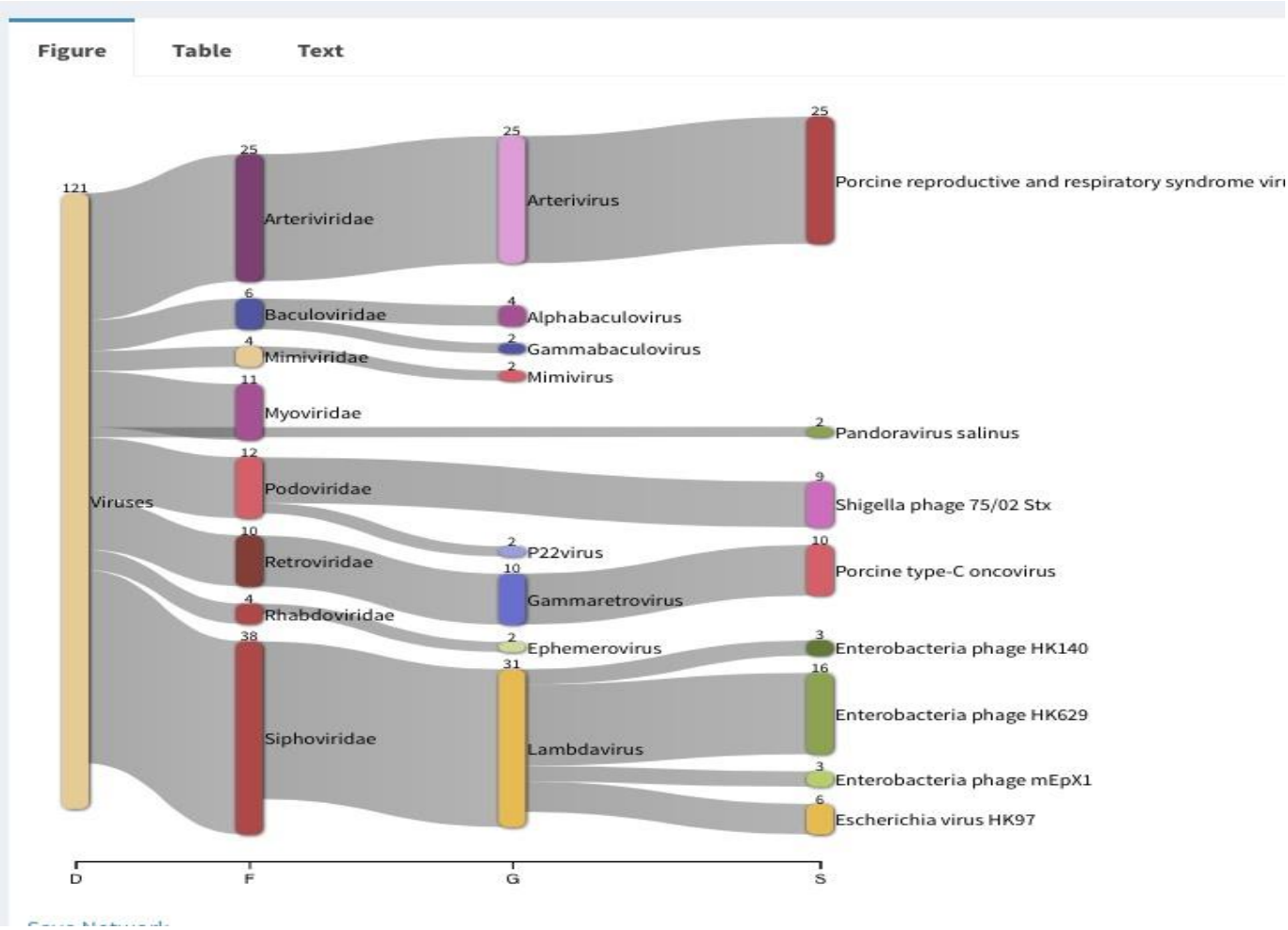

Figure 3. Visualization of shotgun metagenomic classifications by Centrifuge using Pavian, an R based data visualization software program. Numbers over the nodes represent numbers of reads.

This pipeline is functional and capable of correctly identifying PRRS virus genotypes even with mixed infections. We are improving the process to better classify wildtype strains and to ease identification of mixed infections.

Discussion: In summary, in the past year, we have demonstrated that PRRS virus can be identified from serum samples using MinION sequencing. Using Centrifuge for metagenomic analysis, we can accurately genotype the PRRS virus in the serum sample using standard computing resources in a short period of time. This approach is able to correctly genotype virus and detect mixed infection although this process is still being formalized. The 'read until' approach to sequence enrichment does not significantly improve sequence yield at this point.

Other approaches for PRRS virus sequence enrichment are being tested. Preliminary data evaluating sequencing of *orf5* amplicons has demonstrated promise in being able to correctly genotype and detect mixed PRRS infections in serum as well as aggregate samples. This is being further evaluated for a shortterm (validated within 1-3 months) approach to genotyping while other approaches are being improved. In addition, MinION sequencing has demonstrated the ability to detect other pathogens in the sample. As a result of this project, MinION sequencing for metagenomics analysis is currently available and being used for investigational purposes in swine disease investigations. Results from this project have been presented at the American Association of Veterinary Laboratory Diagnosticians annual meeting in San Diego, CA on October 15, 2017, the American College of Veterinary Pathologists annual meeting in Vancouver, BC on November 5, 2017, and at the PRRS Symposium in Chicago, IL on December 1, 2017. An additional student poster at the American Association of Swine Veterinarians is accepted for March 2018. Lastly, as a result of this proposal, an additional $250,000 in federal funding has been obtained to further this study and apply it to oral fluids using enrichment techniques (funding agency has not made a public announcement).

References

1. Brar, M.S., et al., *Porcine reproductive and respiratory syndrome virus in Ontario, Canada 1999 to 2010: genetic diversity and restriction fragment length polymorphisms.* Journal of General Virology, 2011. **92**(6): p. 1391-1397.
2. Han, J., Y. Wang, and K.S. Faaberg, *Complete genome analysis of RFLP 184 isolates of porcine reproductive and respiratory syndrome virus.* Virus Research, 2006. **122**(1-2): p. 175-182.
3. Mitra, N., et al., *Metagenomic characterization of the virome associated with bovine respiratory disease in feedlot cattle identified novel viruses and suggests an etiologic role for influenza D virus.* Journal of General Virology, 2016.
4. Borgarelli, M., et al., *Mitral valve repair in dogs using an ePTFE chordal implantation device: a pilot study.* J Vet Cardiol, 2017.
5. Palinski, R.M., et al., *Widespread detection and characterization of porcine parainfluenza virus 1 in pigs in the United States.* Journal of General Virology, 2015. **97**(2): p. 281-286.
6. Tokarz, R., et al., *Discovery of a novel nidovirus in cattle with respiratory disease.* Journal of General Virology, 2015. **96**(8): p. 2188-2193.
7. Hause, B.M., et al., *Discovery of a novel putative atypical porcine pestivirus in pigs in the USA.* Journal of General Virology, 2015. **96**(10): p. 2994-2998.
8. Hause, B.M., et al., *Isolation of a novel swine influenza virus from Oklahoma in 2011 which is distantly related to human influenza C viruses.* PLoS pathogens, 2013. **9**(2): p. e1003176.
9. Hause, B.M., R.A. Hesse, and G.A. Anderson, *Identification of a novel Picornavirales virus distantly related to posavirus in swine feces.* Virus Genes, 2015. **51**(1): p. 144-147.
10. Collin, E.A., et al., *Cocirculation of two distinct genetic and antigenic lineages of proposed influenza D virus in cattle.* Journal of Virology, 2015. **89**(2): p. 1036-1042.
11. Neill, J.D., D.O. Bayles, and J.F. Ridpath, *Simultaneous rapid sequencing of multiple RNA virus genomes.* Journal of Virological Methods, 2014. **201**: p. 68-72.
12. Jarvis, M.C., et al., *Genomic and evolutionary inferences between American and global strains of porcine epidemic diarrhea virus.* Preventive veterinary medicine, 2015. **123**: p. 175-184.
13. Bracht, A.J., et al., *Real-Time Reverse Transcription PCR Assay for Detection of Senecavirus A in Swine Vesicular Diagnostic Specimens.* PLoS One, 2016. **11**(1): p. e0146211.
14. Zhang, X., et al., *Identification and genomic characterization of the emerging Senecavirus A in southeast China, 2017.* Transbound Emerg Dis, 2017.
15. Kim, D., et al., *Centrifuge: rapid and sensitive classification of metagenomic sequences.* Genome Res, 2016. **26**(12): p. 1721-1729.
16. Ni, J., et al., *Identification and genomic characterization of a novel porcine parvovirus (PPV6) in China.* Virol J, 2014. **11**: p. 203.
17. Schirtzinger, E.E., et al., *First identification of porcine parvovirus 6 in North America by viral metagenomic sequencing of serum from pigs infected with porcine reproductive and respiratory syndrome virus.* Virol J, 2015. **12**: p. 170.