Mixed Strain Identification of Porcine Reproductive and Respiratory Syndrome Virus in Multiplexed Samples using Nanopore Sequencing

Maria Paz Buman Ruiz Diaz

Thesis presented to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science

In

Biomedical and Veterinary Sciences

Kevin K. Lahmers, Chair

Tanya LeRoith

Daniel Slade

December 5, 2023

Blacksburg, Virginia

Keywords: Nanopore, sequencing, porcine reproductive and respiratory syndrome virus

Mixed Strain Identification of Porcine Reproductive and Respiratory Syndrome Virus in Multiplexed Samples using Nanopore Sequencing

Maria Paz Buman Ruiz Diaz

ABSTRACT

For over thirty years, Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) has been a major contributor to morbidity and mortality in the commercial swine industry across the globe. This highly mutagenic RNA virus causes significant economic losses wherever it is prevalent, leading to $664 million in annual losses in the United States. Unfortunately, the current prevention and diagnostic techniques available have proven to be insufficient in controlling the spread of this disease. We describe an alternative diagnostic method exploiting the rapid turnaround time and long-read capacity of Oxford Nanopore Technology's MinION next-generation sequencer. We have developed a novel primer set designed to span Open Reading Frames 3 through 7 of the PRRSV genome, which has allowed for multiplexing of samples, thus reducing individual cost of testing, while yielding significantly more information than previously available. This novel primer pair and sequencing technique have distinguished mixed infections within individual animals and may be used to determine vaccination status. This new approach will help producers and veterinarians make better-informed decisions about co-mingling of animals and vaccination strategies, thus reducing the emergence of new, pathogenic strains of PRRSV.

Mixed Strain Identification of Porcine Reproductive and Respiratory Syndrome Virus in Multiplexed Samples using Nanopore Sequencing

Maria Paz Buman Ruiz Diaz

GENERAL AUDIENCE ABSTRACT

Porcine reproductive and respiratory syndrome virus (PRRSV) is a common, economically important pathogen in commercial swine production. The virus was first identified in the late 1980's during outbreaks in the United States and Europe. In female pigs, the disease is characterized by abortion storms, and the delivery of mummified fetuses or very weak, ill piglets. Neonates often display signs of pneumonia, respiratory distress, and many die from hypoxia. Surviving piglets are highly susceptible to other diseases and are poor growers compared to other, unaffected piglets. Boars may show signs of respiratory disease and can also have decreased libido and reproductive success for months at a time. The virus is prone to mutating once a pig is infected, preventing herds from mounting sufficient immunity to protect against new, mutant strains. Identifying infected pigs early and accurately is crucial to managing PRRSV outbreaks. Currently available diagnostic tests for PRRSV have many limitations, thus we have developed a new diagnostic test using next-generation sequencing technology. Oxford Nanopore Technology provides a commercially available nanopore sequencer, the MinION, that can read long DNA strands in real-time. With this technology we have expanded the area of the PRRSV genome that can be sequenced, which allows us to better identify and distinguish strains of PRRSV in infected, and vaccinated pigs. This new testing method will allow veterinarians and

practitioners across the country to better identify and predict outbreaks in their herds, helping them develop better management strategies against PRRSV.

ACKNOWLEDGEMENTS

Thank you to my committee members, Dr. Kevin Lahmers, Dr. Tanya LeRoith, and Dr. Daniel Slade for their guidance and insight into this project. Special thanks to Dr. Kevin Lahmers for lending a hand during a difficult time in my life, and always being supportive of my future goals. In that same vein, thank you to the excellent team of pathologists in the Department of Biomedical Sciences and Pathobiology for all their support and encouragement in this, and future endeavors.

Thank you to the brilliant women in science who provided a safe space for me to ask all my questions. To Michelle Todd for her indispensable laboratory knowledge, to Catie Burgess for her experience, friendship, and comedy, to Annmarie Budde for her colorful technical support, and to Dr. Anna Hassebroek for patiently lending a listening ear.

Lastly, thank you to Tim for his unrelenting care and kindness.

TABLE OF CONTENTS

LIST OF FIGURES

Thesis Figures

Manuscript Figures

LIST OF TABLES

Thesis Tables

Manuscript Tables

PREFACE

The manuscript included in chapter II is a collaborative project as detailed by multiple authorship. However, most of the benchtop labor, data analysis, and all the writing of the manuscript was done by me, the primary author. Michelle Todd performed some of the early sequencing experiments from 2016 to 2018, which are outlined in the *Shotgun sequencing* portion of the experiment. Gillian Rodgers also performed some unpublished experiments for this project and is credited to outlining some of the sequencing pipeline. Samples were provided by collaborative researchers who also participated in the editing of the manuscript. Dr. Kevin Lahmers is the principal investigator and corresponding author in this research endeavor and is credited for writing the grant which provided funding for this project.

CHAPTER I: LITERATURE REVIEW

Part One: The Virus

Porcine reproductive and respiratory syndrome virus (PRRSV) is a 15,000-base pair (bp), positive-sense, single-stranded, RNA virus with remarkable mutational ability.[1,2] It was first recognized in the late 1980's by near-simultaneous epidemics in Europe (The Netherlands, 1990) and North America (USA, 1987), and is now the most economically devastating swine disease in the United States.[2-4] PRRSV is classified into two subtypes: Type 1 is the European subtype (PRRSV1), and Type 2 is the North American subtype (PRRSV2), though Type 2 is actually most prevalent in Canada, the U.S., and Asia. This project is focused on the North American subtype (PRRSV2).

History and phylogeny

Porcine reproductive and respiratory syndrome virus is an *Arterivirus* (order *Nidovirales,* family *Arteriviridae*), along with Equine Arteritis Virus, Lactate dehydrogenase-elevating virus, and Simian hemorrhagic fever virus.[5] The North American and European subtypes only share about sixty-percent similarity in nucleotide sequence, with up to a twenty-percent variation in nucleotides between strains of each subtype.[6] There has been much debate over the origin of PRRSV, but extensive phylogenetic analysis suggests the two subtypes originated from a common ancestor, but evolved separately and coincidentally emerged in a similar time frame.[7,8]

The North American subtype is further classified into nine lineages (Table 1), with some having a higher prevalence in North America (Lineages 1, 5, and 8) or Asia (Lineages 3 and 4).[8]

Lineage 1 (L1) carries the viral strains MN184 a, b, and c, which led to severe outbreaks in the U.S. in the late 1990s and early 2000's.[8] Lineage 5 (L5) holds the first isolated strain of PRRSV2, VR-2332, and Lineage 8 (L8) holds strains JA-142, NADC20, and SDSU73.[9,10] The prevalence, importance, and pathogenicity of strains vary by geography and prevention strategies.

Table 1 – PRRSV2 Lineages and Brief Descriptions

Lineage	Important Strains	Comments[11]
1	• MN184 a, b, c • NADC30	The most prevalent lineage globally in 2023. There are multiple sub-lineages in L1, with the L1C variant increasing in prevalence since 2015.
2		Occasionally detected in Canada.
3		Primarily found across Asia.
4		Primarily found in Japan.
5	• VR2332; Ingelvac PRRS MLV vaccine • VR2385	VR2332 is recognized as the original strain detected in North America in the late 1980s.
6		Rarely isolated in the U.S.
7	• Neb-1; PrimePac vaccine	The PrimePac vaccine was discontinued in 2005 after which the prevalence of new, recombinant strains in this lineage dissipated considerably.
8	• JA-142; Ingelvac PRRS ATP vaccine • Fostera (P129) vaccine • NADC20 • SDSU73	Lineage 8 has led to multiple field outbreaks in the last 30 years and is still considered a prevalent lineage of PRRSV.
9		Prevalence has dropped significantly since 2010.

Disease and transmission

As the name suggests, PRRSV causes an array of reproductive and respiratory signs in pigs of varying ages, and these are the most prominent modes of transmission as well. Horizontal transmission occurs by three routes: oronasal aerosolization, transport, and artificial insemination. PRRSV is readily aerosolized in the respiratory tract and can spread far distances by sneezing and coughing (approximately 30 meters), thus making direct contact and distance-dependent transmission relatively simple.[12] Fomites, including humans, are also a major method of transmission that can overcome even larger distances, further stressing the importance of stringent biosecurity practices in animal facilities.[12] A phylogeographical analysis of PRRSV2 performed by Shi et al., also demonstrates the major role of interstate transportation and artificial insemination in disease transmission within the United States.[8,13] Even with strict quarantine measures for pigs and shower-in/shower-out protocols for human caretakers, disease still often spreads amongst newly mingled animals via aerosolization and nose-to-nose contact. Truck washes are also an imperfect method of sanitation as we cannot ensure these vessels are perfectly sanitized after every transport, allowing PRRSV to be transported in wheel wells, fenders, and other vehicle crevices. In the realm of reproduction, improved handling of semen, and stringent selection of PRRSV-negative boars, has led to ever-decreasing levels of disease transmission through artificial insemination.[8] Vertical transmission occurs from the sow to the piglet, and sows infected with PRRSV2 in late gestation can give birth to viremic piglets.[14]

In sows and gilts, the time of infection during gestation often dictates the impact the virus will have on piglets, while clinically affected dams may exhibit signs of pneumonia, lethargy, and general malaise.[15-18] When infection occurs during early to mid-gestation, piglets are usually born without clinical signs. However, infection during late-gestation often leads to late-term abortions, stillbirths, or mummified fetuses.[19] Newborn piglets often present with varying degrees of respiratory disease such as pneumonia, respiratory distress, and cyanosis, which can lead to death.[2,18-20] Piglets who survive this period often show poor growth, increased disease susceptibility and are generally unthrifty when compared to others.[19] Boars infected with PRRSV often have a decreased libido and breeding rate for months at a time.[20]

Once infected, PRRSV replicates within the cytoplasm of porcine alveolar macrophages. From here, the virus produces a generalized viremia in the pig characterized by interstitial pneumonia, multi-organ edema, lymph node hypertrophy, hyperplasia and degeneration, and varying degrees of myometritis and endometritis. There may also be scattered lesions across the heart, large vessels, brain, kidneys, nasal mucosa, and stomach.[18,19] Aborted, stillborn, and mummified piglets infrequently show gross lesions, but the sows do.[19]

Infection with PRRSV also leads to a peculiar immune response in pigs. Within twelve hours of infection there is an initial spike in circulating leukocyte concentration. However, this response is dampened shortly afterward, leading to a decreased circulating leukocyte count.[17] This dampened immune response allows the virus to remain in lymph nodes and lung tissue, where it continues to mutate for up to 260 days.[21] The virus in these persistently infected animals eventually accumulates enough mutations (single nucleotide polymorphisms or recombinations) that a novel viral strain is produced. This novel strain can now infect the entire herd again, leading to cyclical outbreaks of PRRSV in large herds.[16,17,21]

In summary, the virus spreads throughout the body and greatly diminishes the pigs' immune capacity, thus making the animal significantly more susceptible to other diseases.[17] PRRSV2 is readily spread by direct contact with nasal discharge, blood, and semen. However, long-distance transportation is a hallmark of the swine industry in the United States and must be recognized as a major catalyst for super-spreading of PRRSV2, allowing for novel, recombinant strains to emerge.

Economic impact and importance

Unsurprisingly, respiratory, and reproductive diseases have a significant economic impact on swine production, an industry which is based on the constant replenishment of animals and their daily growth rates. PRRSV is considered endemic across U.S. swine herds and leads to an estimated $664 million in losses annually in the U.S., and up to £698,000 of losses in the United Kingdom.[3,22] Thousands of animals are lost every year to PRRSV, with an approximate 11% decreased farrowing rate and 6-11% increased mortality in the grower and finisher stages of production.[6] The severity of the outbreak also plays a major role in the economic impact PRRSV may have on a farm. One study from China found a range of ¥1049.78–¥1609.70 ($143 – $221*) of increased costs per sow from the breeding to grower stage (*Currency exchange as of October 2023).[23] At the time of this study, China is reported to have the highest population of swine in the world, with over 450 million animals, making this disease a major factor of profitability in the Chinese, and global swine market.[24]

Virology

PRRSV2 is approximately 15-kb long, capped with a poly-adenylated tail at the 3' end and is divided into at least nine Open Reading Frames (ORF), depending on the source.[1,7,25] On the 5' end is a significant untranslated region (5'-UTR) which serves as an anchor for ribosomal transcription, and another 3'-UTR past the polyadenylated tail. Of the translated genome, ORF1a and ORF1b cover the majority and encode for non-structural proteins, which are responsible for the replication and transcription complex of the virus.[20,25] The structural proteins, glycoproteins (GP) 2 – 5, N, M, and E, are encoded by ORFs 2 – 7.[26] The mature viral particle is composed of a capsule created by GP2 – GP5 tetramers, M protein, and sialic acid.[25] The RNA genome is enshrouded within the virion particle by nucleocapsid (N) protein.[25]

The virus first contacts host alveolar macrophages via its membrane proteins, GP2 – GP5. Glycoproteins 2, 3, and 4 are shielded by glycan molecules to avoid triggering an immune response in the host.[27] GP5 interacts with multiple host cell mediators including CD163 and CD169, which, along with M protein, allow for swift entry of the virus into alveolar macrophages.[25,28] Viral RNA and N protein can now traverse through the cell cytoplasm to a budding site, or begin replication.[25] CD163 and CD169 have been shown to be major mediators in disease development with CD163 knockout pigs being resistant to PRRSV infection.[29]

As mentioned, PRRSV has a remarkable mutational rate with 4.71×10^2 to 9.8×10^2 synonymous sites/year, equivalent to one mutation (point mutation or recombination) per replication cycle.[21,30] This high mutation rate is attributed to PRRSV RNA-dependent RNA polymerase (RdRp) inability to perform proofreading checks during transcription, allowing mutations to accumulate at an alarming rate. PRRSV is thought to be the most mutagenic ssRNA virus and is only comparable to the mutation rate of Human Influenza Virus, a retrovirus which also lacks RdRp proofreading abilities.[1,7] ORFs 1a, 1b, 3 and 5 are notoriously variable amongst PRRSV strains, and ORF5 is responsible for encoding GP5, a major component of the viral capsule and the infectivity of the virus, often attributed to mutations resulting in gain of fitness.[30,31] It is this genetic variation and role in immunity that has led to ORF5 being used as a diagnostic marker for PRRSV since its discovery.

Part Two: Diagnostics and Prevention

Over the years, multiple preventive strategies have been established for PRRSV, including vaccines and diagnostic tests. Currently, there are three major vaccines available in the United States, all with variable success. The virus has also been tested for via multiple classes of diagnostic tests including Enzyme-Linked Immunosorbent Assays (ELISA), Restriction Fragment Length Polymorphism (RFLP) analysis, and real-time PCR (rt-PCR) (Table 2). Other ancillary tests have also been developed and left behind over the years, such as virus isolation, serum virus neutralization, and indirect fluorescent antibody tests on serum. *In situ* hybridization and immunohistochemistry can also be performed on histology samples from necropsy cases.[32] However, we will not be covering these ancillary tests in this review as they are not pertinent to the following research.

Vaccine Strategies

In veterinary medicine, there are two major groups of vaccines available: killed vaccines and modified live vaccines (MLV). Killed vaccines provide limited immunity against PRRSV2 or require multiple boosters to reduce clinical disease and increase viral clearance, and were discontinued in the U.S. in 2005.[33] Thus far, only MLVs have shown to stimulate a sufficient immune response in pigs in a commercially viable way.[26] Boehringer-Ingelheim has developed two widely used MLVs for PRRSV. The Ingelvac PRRS MLV and PRRS ATP vaccines are used widely throughout North America and provide adequate immunity against their own, respective lineages, with mild to moderate immunity against heterologous strains. The Ingelvac PRRS MLV vaccine is derived from the L5 strain, VR-2332, and the PRRS ATP vaccine is derived from the L8 strain, JA-142. Zoetis, another competing pharmaceutical company, has also developed an

MLV vaccine, Fostera (a.k.a. strain P129), which also belongs to Lineage 8.[11,34] There are other, minor vaccines available for specific strains used in other countries, or being researched, though these are not relevant to the present study.[26,33]

Modified live vaccines are not without consequences. These vaccines are derived from live viruses that have been altered in a way to reduce their ability to cause disease (a.k.a. attenuated live vaccines). This means that when vaccinated, the animal is inoculated with some version of a virus that, in the case of PRRSV, may be capable of reverting to its pathogenic state. A notable phenomenon with PRRS MLV vaccines is that they have also been shown to recombine with wildtype strains of virus, leading to the development of novel field strains.[15,21,35,36] The limited protection and recombinant ability of PRRS MLVs make them ill-suited to combat such a dynamic opponent.

Diagnostic Testing

When efficacious, diagnostic tests can allow producers and veterinarians to detect infected animals early, and quarantine them from the rest of the herd, thus reducing the spread of disease. We will discuss three major diagnostic methods of importance (Table 2).

Enzyme-linked immunosorbent assay testing

ELISA are relatively easy to develop and inexpensive to distribute. PRRSV2 is tested by an indirect antibody ELISA that provides rapid results for individual animals, or "composite samples of oral fluid."[37] IDEXX's PRRS X3 Ab Test has a reported specificity of 98.8% and sensitivity of 99.9%, making it an attractive screening test for PRRSV2, and is widely used in the United States.[38,39] However, these ELISAs are only capable of detecting antibodies to PRRSV, which are produced after exposure to the virus. To detect antibodies, we must allow the immune system to develop them, a process called seroconversion. Replication of the virus occurs rapidly in macrophages (~12 hours), but peak viremia typically occurs within 7-10 days of infection, which is when ELISAs are most helpful.[14,18,20] The IDEXX PRRS X3 Ab Test specifically tests for nucleocapsid protein, and has been reported to provide results as early as 5 days after infection.[40-42] Regardless, this puts swine veterinarians 5-10 days behind the start of an outbreak, allowing many animals to go undiagnosed during that time. Also, antibody testing only provides information about an animal's exposure to PRRSV and cannot distinguish between Type 1 or Type 2 infections. They also cannot differentiate between MLV vaccination and natural infection, thus providing limited information on an animal's overall infection status.

Restriction-fragment length polymorphism analysis

Restriction fragment length polymorphism analysis is a technique in which restriction enzymes digest a nucleic acid substrate (i.e., PCR product), and are analyzed via southern blotting.[43] For PRRSV2, ORF5 is amplified via PCR and then digested by enzymes MIuII, HincII, and SacII, providing a "digestion pattern" characterized by a 3-digit code.[44] In contrast to ELISA tests, RFLP analysis detects actual virus in a sample (serum, semen, oral fluids, etc.), which can, theoretically, allow for earlier detection starting at 12 hours post-infection. Over time, new RFLP patterns were detected for PRRSV2, ultimately leading to the question of whether RFLP was an accurate method of characterizing new PRRSV2 strains.

Recent studies performed by Cha et al., and Trevisan et al., recognize the limitations of RFLP analysis. Cha et al. performed a multiple passage study of pigs inoculated with VR2332 and found that RFLP patterns are "unpredictable and inconsistent" during *in vivo* replication of PRRSV2. This study also noted multiple genetic changes in ORF5, which were not identified by RFLP analysis, suggesting the RFLP patterns are not accurately representing genetic diversity of PRRSV2.[44] Trevisan et al. also performed a genetic diversity study comparing ORF5 sequences via Sanger and RFLP analysis. Ultimately, they identified multiple clusters within certain RFLP patterns, but recognized a disparity of genetic diversity between their RFLP reports and Sanger sequencing analyses.[44,45] These reports highlight the significant flaws in RFLP analysis, and the need for more detailed PRRSV2 testing.

Table 2 – Summary of currently available diagnostic tests for PRRSV

Testing Method	Advantages	Disadvantages	Time to testing
ELISA	• Fast • Inexpensive • Pooled testing available	• Cannot differentiate vaccination from infection • Cannot differentiate PRRSV1 vs PRRSV2 infection	5 to 10 days post-infection
RFLP analysis	• Differentiates PRRSV1 from PRRSV2 • Some potential for identification of vaccine strains	• Slow • Provides limited information of the viral genome • Often incapable of identifying novel strains of virus, or recombinant strains	>12 hours post-infection
Sanger sequencing of ORF5	• Differentiates PRRSV1 from PRRSV2 • Gold-standard sequencing method	• Slow (5+ days for results) • Expensive (~$200/sample)[46] • Provides very limited information past ORF5 • Often incapable of identifying novel strains of virus, or recombinant strains	> 12 hours post-infection

Sanger sequencing of Open Reading Frame 5

A more descriptive diagnostic method involves the use of PCR followed by Sanger sequencing, one of the first DNA sequencing methods developed.[47] Despite Sanger sequencing becoming so readily available, it is limited in that it can only provide accurate sequencing analysis for approximately 800 bp at a time. The ORF5 gene measures approximately 600 bp, which makes it suitable for Sanger sequencing, but still leaves over 90% of the translated genome undefined. By only analyzing ORF5, we often miss recombinant strains of PRRSV, where the ORF5 sequence may belong to an L1 strain, but other ORFs may belong to another lineage altogether, making it difficult to accurately identify and classify novel PRRSV2 strains.[21,48]

As mentioned, there are growing reports of new, recombinant strains of PRRSV2 across the world. Recombination often occurs between wildtype strains, between wildtype and vaccine strains, or sometimes, between two vaccine strains.[21,36,48] Mosaic strains are also reported, though in a smaller proportion.[36,48] Sanger sequencing and RFLP analysis only provide a keyhole view of the PRRSV2 genome via ORF5 analysis. By expanding genome sequencing analysis, we can gather more information about new PRRSV2 strains, thus improving our clinical decision-making to reduce, or eliminate, the development of pathogenic, recombinant strains. An ideal diagnostic test would expand past ORF5, and accurately classify the infective strain(s) at minimum by lineage, and at best by individual strain. The test would also discern the animal's vaccine status, all at a low cost with rapid turnaround time. Doing so would allow producers to safely mingle animals that have been infected or vaccinated with the same lineages of PRRSV2, or genetically similar strains, thus reducing the impact of recombination and point mutations leading to new, pathogenic strains of virus. Detection of multiple lineages within individual pigs would also allow for identification of the causative agent of disease in post-vaccination outbreaks, which could lead to safer, and more judicious use of modified live vaccines.

Major strides in next-generation sequencing (NGS) have allowed for rapid, high-throughput genetic analyses, which are capable of exactly that.

Part Three: Next-generation sequencing with Oxford Nanopore

In 1977, Frederick Sanger and his colleagues developed a ground-breaking technique for DNA sequencing, rightfully earning him his second Nobel Prize in Chemistry in 1980. Sanger sequencing allowed researchers to sequence the entire human genome for the first time and became pivotal in the development of DNA sequencing as a research field and for clinical applications.[49] Over 40 years later, we now have access to a new realm of next-generation sequencing techniques.

Next-generation sequencing

Although Sanger sequencing became adopted as the gold-standard for DNA sequencing, “next-generation” techniques were developed for faster, high-throughput sequencing. In 2015, Oxford Nanopore Technologies (ONT) released the first commercially available nanopore sequencing device called the MinION, currently capable of sequencing individual nucleotides at ~400 bp/second, with ultra-long reads up to 4 million bases.[50-52] Next-generation sequencing typically occurs in three steps: library preparation of DNA or RNA samples, sequencing, and data analysis (bioinformatics). Much of the information detailed here is from personal experience working with nanopore sequencing technology.

Library preparation

Library preparation (library prep) involves the priming of DNA strands, a.k.a. the library, via enzyme-assisted chemical reactions that enable sequencing on the selected sequencing device. In the case of nanopore sequencing, DNA strands must be “end-prepped” and “adapter-ligated” so they may be processed by the nanopore membrane (Figure 1). The end-prep procedure is relatively universal across nanopore library prep protocols and allows the 5’ and 3’ ends of the DNA strand to accept the nanopore adapter to be ligated. Adapter ligation involves the addition of a specialized adapter to the 5' end of the DNA strand which serves as an anchor to the nanopore itself. The MinION is capable of sequencing entire genomes of mammalian DNA, down to the plasmids and bacteriophages of unicellular organisms. This wide array of DNA templates requires specified library prep protocols which can be found on the ONT website.[52]

Another important aspect to library prep and nanopore sequencing is the ability to multiplex samples during sequencing. Sample multiplexing, multiplex sequencing, or simply multiplexing involves the pooling of multiple samples (the collective libraries) so they may be sequenced at the same time on the same device. To do so, samples are tagged via barcodes: unique nucleotide sequences developed by ONT for multiplexing. These barcodes are attached after end-prepping and are read and sorted by the sequencer by a process called "bin-sorting" or demultiplexing (Figure 1). A single MinION sequencer is currently capable of identifying 96 barcodes, allowing for multiplexing of up to 96 samples at a time.[50,52,53]

The MinION is also capable of sequencing RNA molecules, though due to the fragility of RNA, direct RNA sequencing is typically performed in smaller genome fragments. Another approach for sequencing RNA genomes is to synthesize cDNA via reverse transcription, which is a method we employed in this project.

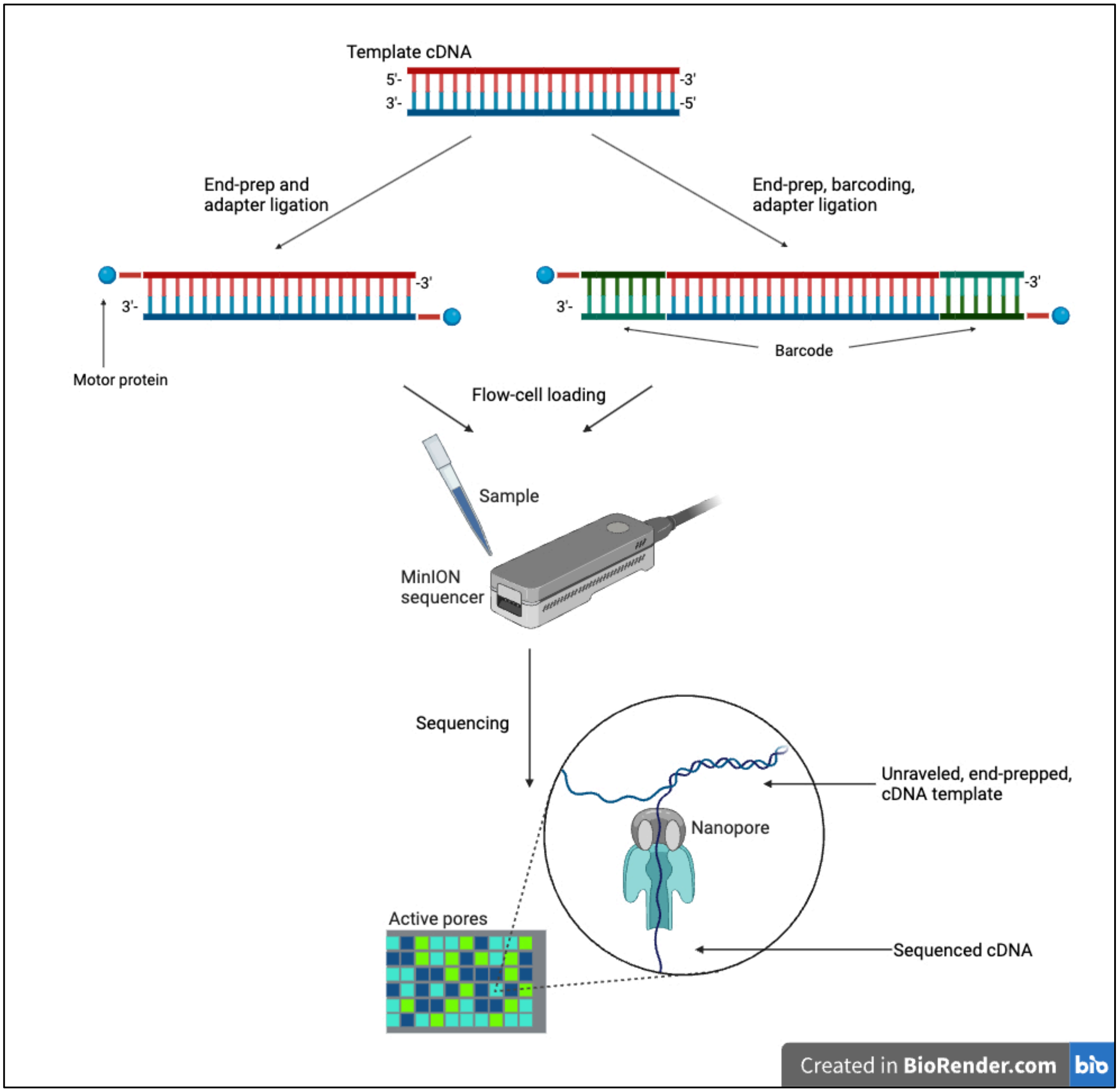


Figure 1 – Schematic view of alternative options for library prep, followed by nanopore sequencing.

Nanopore Sequencing

After library prep, the pooled library is loaded onto the MinION's flow cell to begin sequencing. Nanopore sequencing works by running the end-prepped DNA strands through individual nanopores, powered by an ionic current.[52,54] Each flow cell contains one nanopore membrane, outfitted with 512 "nanowells", each with a single nanopore inside. Flongles are smaller, cheaper, disposable flow cells with up to 126 nanopores, and can be used in the MinION and Mk1C devices. As the sequencing run begins, an ionic current charges the nanopore membrane, thus attracting the adapter-ligated DNA strands. Each nanopore contains a tether strand which grasps onto the adapter, ratcheting the DNA strand three nucleotides at a time.

As the nucleotides pass through the charged nanopore, they produce an ionic current that is identified and processed into "squiggle data." This squiggle data is the raw data produced by the MinION, which can then be processed by ONT's MinKNOW software and later basecalled by ONT's proprietary basecalling software.[52] Basecalling is the process by which a sequencing software assigns nucleotides based on the raw data received.[52] In the case of the MinION, the raw squiggle data is saved as a fast5 file and, after basecalling, is converted into fastA files. These fastA files can be imported into bioinformatics programs for data analysis. However, newer versions of MinION basecalling software (i.e., *Guppy*, *Bonito, Dorado, etc.*) can further process the fast5 files and assign quality scores to the nucleotide readings as well. This information is saved as fastQ files and includes the nucleotide sequence along with a basecalling quality score, which becomes helpful during data analysis to establish the trustworthiness of reads.[50,52,53]

Nanopore sequencing power can be tailored to the sequencing capacity required. The MinION Mk1B is the smallest, most portable device provided by ONT. It weighs less than 90 grams, is entirely powered by a USB 3.0 cable connected to a laptop computer and starts at $1,000 per device, making it accessible to laboratories worldwide.[52] The sequencing output from the Mk1B provides only raw data in the form of fast5 files that have not been basecalled. The MinION Mk1C however, provides more powerful sequencing runs as it contains an integrated graphical processing unit, capable of sequencing, basecalling, and bin-sorting in real-time. The Mk1C provides unprecedented access to sequencing data allowing for immediate data analysis and decision-making, while sequencing is still occurring.[52] There are also larger, more powerful devices available through ONT. The PromethION 48 is the largest sequencer available, capable of holding 48 PromethION flow cells at a time, with up to 2,675 nanopores per flow cell, and a theoretical maximum output of 280 Gigabytes.[52] Sequencing can be stopped at any time, and due to its real-time basecalling abilities, results can be available whenever the researcher deems sufficient reads have been attained, sometimes within minutes.

MinION sequencing devices enable unprecedented portability. The smaller devices, Mk1B and Mk1C, are capable of processing small, whole genomes to targeted amplicons of larger, complex genomes.[52] This makes them excellent for rapid diagnosis in the field. In fact, the MinION has already been deployed for field work during the Zika and Ebola outbreaks, in the latest SARS-CoV-2 epidemic, and has even traveled to space.[51,55-59]

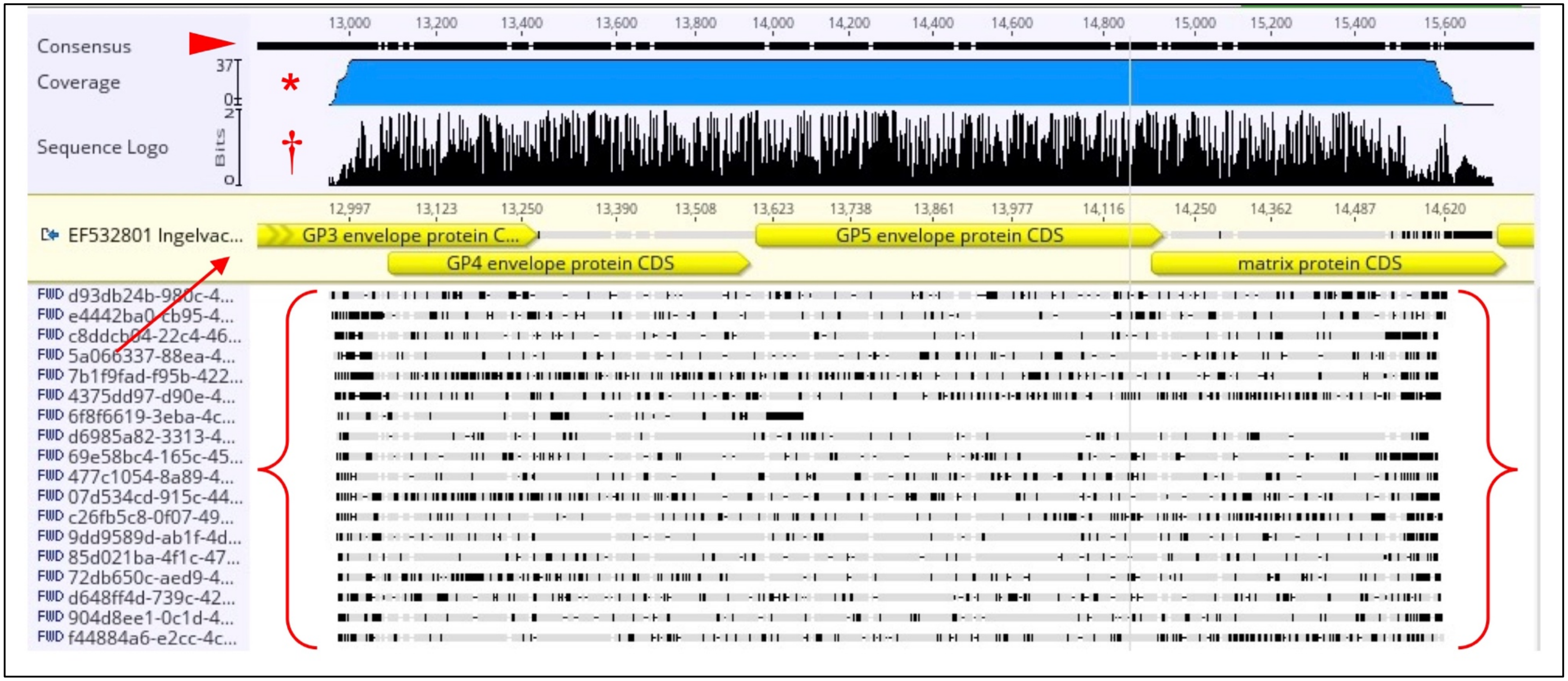


Figure 2 – Screenshot of Geneious Prime describing Mapping to Reference Genome results. Sequences (in red brackets) are aligned to the reference genome (red arrow). The depth of coverage is determined by the number of reads aligned to the sequence, and the relative coverage can be visualized above the reference genome (red asterisk). The sequence logo shows the degree of alignment between all the aligned sequences and the reference genome (red cross). When magnified, the sequence logo can show multiple nucleotides in the same locus. The size of the nucleotide signifies the ratio of sequences with that nucleotide compared to others. Larger nucleotides in a locus indicate most sequences have that same nucleotide in that position. When there are multiple nucleotides in a position of the sequence logo, this can denote a single nucleotide polymorphism, and may warrant further analysis. The consensus sequence (red arrowhead) is the average nucleotide reading from the sequence logo. The consensus sequence can be extracted and used for BLAST analysis, phylogeny analysis, pairwise alignment, or many other commands.

Data analysis and bioinformatics

Once the sequencing run has ended, the actual nucleotide data can be analyzed by any number of bioinformatics software programs. Each individual strand that was sequenced and processed is termed a "read," and in the case of nanopore sequencing, the "depth of reads" or "coverage" corresponds to the number of reads of the same portion of the genome (Figure 2). Greater coverage often indicates more accurate sequencing, reduced basecalling error, and more reliable input data.

Typically, the sequencing data is analyzed by mapping to a known reference genome, or via *de novo* assembly. When sequences are mapped to a reference genome, they are individually aligned to the template genome wherever they align best, allowing us to compare the reference genome to the sequence reads and easily identify areas of mutation. Reads can also be mapped to multiple genomes, and the degree of alignment to different genomes may signify that multiple variants are present in a sample.[53] However, mapping to reference genomes forces the reads to map wherever the software determines is best, even if the alignment is not of very high quality. This approach may impede identification of novel strains of a virus, for example. After mapping, a "consensus sequence" is generated, delineating the average nucleotide sequence between the reference genome and input reads. The consensus sequence will show areas of shallow coverage, or of multiple nucleotides represented. When this occurs, this may represent a single nucleotide polymorphism (SNP) or a recombination event in the event of larger gaps in the consensus. This may provide insight into a novel strain of virus, but requires further investigation, such as with *de novo* assembly.

De novo assembly allows reads to be pieced together to produce larger DNA strands from shorter, fragmented reads.[60] This is especially helpful when only short portions of DNA can be sequenced, or the quality of DNA is poor due to shearing, transport, or handling problems.[60] This patchwork approach is also beneficial with tiling experiments, when a whole genome has too much variability to be sequenced at once, so multiple, overlapping amplicons can be sequenced and assembled together.[61] *De novo* assembly is also used to identify novel variants of viruses or new species. The quality of this approach, however, is dictated by that of the source material. If there is a low depth of reads, or the reads are of poor quality, there can be many errors in assembly, leading to inappropriate assemblies or even incorrect identification of results altogether. *De novo* assembly is also a notoriously time-consuming process, with some assemblies taking multiple days to finish, or requiring super-computer capabilities.[60]

Once the consensus sequence is generated, it can now be interpreted, and results gathered. In the case of multiplexing, all samples should go through the same bioinformatic analysis, known as the "pipeline." This reduces the chances of human manipulation in individual samples, yielding more accurate conclusions. Though, at times, manipulation of data is necessary to diminish the effect of low-quality reads, inappropriate assemblies, or establishing minimum read coverage depending on the experiment conducted.

While processing the sequencing data from our study, we found that demultiplexing can also pose challenges to nanopore sequencing. Increasing the stringency of demultiplexing and barcode detection requirements enabled us to gather higher quality input reads, thus providing more accurate analysis further downstream. Ultimately, the bioinformatics approach decided upon depends on the goal of the research being conducted, the computing power available, and the quality of source data gathered.

As of 2023, nanopore sequencing is still a relatively new field of molecular biology, but with highly accomplished strengths. It is with these capabilities that we provide an avenue towards a new diagnostic method for porcine reproductive and respiratory syndrome virus.

Hypothesis

Long-amplicon nanopore sequencing can distinguish wildtype and vaccine strains in PRRSV2 positive pigs in a cost-effective and timely manner.

Objectives:

1) To determine if shotgun sequencing is a viable method of detecting PRRSV in inoculated pig serum.
2) To determine if long-amplicon sequencing with a novel primer pair is a viable method of detecting PRRSV in viral stock cultures, dilutions of viral stocks, and inoculated pig sera.
3) To determine if long-amplicon sequencing with a novel primer pair is a viable method of differentiating between PRRSV wildtype strains and vaccine strains in infected pig sera.
4) To determine if long-amplicon nanopore sequencing is a potential, cost-effective and rapid diagnostic test for PRRSV2.

CHAPTER II: RAPID IDENTIFICATION AND DISTINCTION OF PRRSV2 WILDTYPE AND VACCINE STRAINS IN MULTIPLEXED SAMPLES USING NANOPORE SEQUENCING

Maria P. Buman Ruiz Diaz[1], Michael Roof[2], S. Michelle Todd[1], Gillian Rodgers[1], Pablo Piñeyro[3], Kevin K. Lahmers[1]

1. Virginia-Maryland College of Veterinary Medicine, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24060

2. Boehringer Ingelheim

3. College of Veterinary Medicine, Iowa State University

Corresponding Author: Kevin K. Lahmers, Virginia-Maryland College of Veterinary Medicine, Virginia Polytechnic Institute and State University, 205 Duck Pond Drive, Blacksburg, Virginia 24061, USA, klahmers@vt.edu

Running head: PRRSV Mixed Infection Rapid Identification using Nanopore Sequencing

ABSTRACT

For over thirty years, Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) has been a major contributor to morbidity and mortality in the commercial swine industry across the globe. This highly mutagenic RNA virus causes significant economic loss wherever it is prevalent and leads to hundreds of thousands of animals lost annually. Unfortunately, the current prevention and diagnostic techniques available have proven to be insufficient in controlling the spread of this disease. We describe a new diagnostic method exploiting the rapid turnaround time and long-read capacity of Oxford Nanopore Technology's MinION next-generation sequencer. We have developed a novel PCR primer set designed to amplify open reading frames 3 through 6 of the PRRSV genome, yielding significantly more information than previously available. This PCR approach allows for multiplexing of samples, thus reducing individual cost of testing, This novel primer pair and sequencing technique have distinguished mixed infections within individual animals and may be used to determine vaccine persistence, or reversion to a pathogenic form. This new approach will help producers and veterinarians make more rapid and better-informed decisions about co-mingling of animals and vaccination strategies, thus reducing the emergence of new, pathogenic strains of PRRSV.

INTRODUCTION

Porcine reproductive and respiratory syndrome virus (PRRSV) is a 15,000-base pair, enveloped, positive-sense RNA virus composed of at least nine open reading frames (ORFs). [1,2] This virus has caused substantial global economic loss since its discovery in the late 1980s, estimated at $664 million in annual losses in the United States[3]. Consequently, PRRSV is the country's most economically devastating swine disease and is considered endemic in all commercial swine herds. [3] The virus is classified into two types: European Type 1 (PRRSV1) and North American Type 2 (PRRSV2), with approximately 60% similarity in nucleotide sequence between the two types and about 20% dissimilarity amongst the numerous strains within each subtype. [4-7] PRRSV is traditionally diagnosed using either enzyme-linked immunosorbent assays (ELISA), restriction fragment length polymorphism (RFLP) analysis, or PCR and Sanger sequencing, with variable success and accuracy. [8-10] Type 2 infections are classified into one of nine lineages via Sanger sequencing of ORF5 and, historically, restriction fragment length polymorphism (RFLP) analysis. [7,11] These methods have been efficacious over the last four decades but are proving to be less useful in distinguishing diverse PRRSV strains due to the constant mutation, recombination, and global evolution of the virus. [12]

We have seen increasing accounts of newly mutated, pathogenic strains of PRRSV in recent years. [13,14] Most alarming, however, are the ever-growing reports of recombinant strains of PRRSV. Researchers have proven many of these strains arise within closed, certified PRRSV-free, or vaccinated herds. [13-17] One report even describes a pathogenic, recombinant strain of PRRSV that arose from a vaccinated, PRRSV-free herd. [17] Unfortunately, by only testing for PRRSV via ORF5 analysis, many of these recombinant strains are misdiagnosed, or missed altogether, because the current diagnostics only identify a small part of the viral genome[18]. It is this gap in diagnostic ability that has led us to explore alternative options.

Breakthroughs in next-generation sequencing technology have allowed for more accessible and practical diagnostic workflows. Oxford Nanopore Technology's (ONT) MinION device has been repeatedly proven to be reliable and timely in its sequencing abilities. [19-30] Advantages to this technology include real-time basecalling and demultiplexing, unrestricted read length capacity, and maximized affordability at an estimated \$1-\$6 of library prep per sample. [31,32] The MinION is also capable of sequencing up to ninety-six multiplexed samples at a time, with the potential for identification of mixed infections, and perhaps discerning vaccine status. [26,27,32]

The decreasing relevance of standard PRRSV diagnostic techniques and the refinement of nanopore sequencing technology has led us to develop a novel long-amplicon diagnostic test for PRRSV. This test capitalizes on the MinION's long-read sequencing ability to identify and distinguish wildtype and vaccine strains of PRRSV2 in infected pigs in a cost-effective and timely manner.

MATERIALS AND METHODS

We first evaluated the use of nanopore sequencing by testing pig sera for PRRSV via shotgun sequencing, and then by long-amplicon sequencing of viral cultures and serum samples using a novel primer pair, designed in our laboratory.

Shotgun sequencing of inoculated pig serum

We initially performed shotgun sequencing of a serum sample (102.7) from a specific-pathogen-free (SPF) pig inoculated with PRRSV2 strain MN184 (SPF-MN184), collected 7 days post-inoculation (dpi) (X. J. Meng, Virginia Tech). We extracted RNA using a QIAmp Viral RNA Mini Kit, per the manufacturer instructions (Qiagen). We then performed cDNA synthesis using the SuperScript First Strand Synthesis System for RT-PCR, per manufacturer instructions (ThermoFisher Scientific), followed by Second Strand Synthesis using the NEBNext UltraII Non-Directional RNA Second Strand Synthesis Module (New England Biolabs). Second Strand Synthesis was carried out in an 80 uL reaction with 15 uL First Strand cDNA, per manufacturer instructions, including incubation at 16°C for two and a half hours. We then proceeded to dsDNA purification using AmPure XP Beads, per manufacturer instructions (Beckman Coulter Life Sciences). We then performed library preparation using the SQK-LSK108 kit (Legacy store, Oxford Nanopore Technologies), flow cell 9.4.1, and sequenced on a MinION for 24 hours.

We basecalled these reads using *Guppy* v. 6.5.7 (Oxford Nanopore Technologies) and classified them using Centrifuge v. 1.0.4 using the compressed human, virus, bacteria, and archaea database.[33] We also analyzed these reads using Geneious Prime version 2022.2.1 (Dotmatics) and mapped them to five PRRSV2 reference genomes: MN184, VR2332, VR2385, NADC20, and SDSU73 (GenBank Accession Numbers: EF484031, EF536003, JX044140, MK837936, JN654458, respectively), using the Geneious Assembler (default settings).

Long-amplicon sequencing with novel primer pair on viral cultures and pig sera

For all the following experiments we first extracted RNA genomes from culture media or serum samples (Viral RNA Mini Kit, Qiagen) and then synthesized cDNA (SuperScript IV First-Strand Synthesis System, ThermoFisher Scientific) per manufacturer instructions. We performed conventional PCR throughout these experiments using the Platinum PCR SuperMix High Fidelity kit (ThermoFisher Scientific), per manufacturer instructions. We designed a novel primer set using Geneious Prime, spanning from the end of ORF3 through ORF7 (Forward: 5'-TAGAATGGCTGCGTCCCTTC-3'; Reverse: 5'- CAATCGGATGAAAGCCTGCG-3'), with an expected amplicon of 1,537 bp (prrs2-amp primers). We amplified the samples in a 50 uL reaction consisting of 45 uL Platinum High Fidelity SuperMix, 200 nM each Forward primer and Reverse primer, and 3 uL cDNA template or 3 uL nuclease-free water as a negative template control (NTC). We performed conventional PCR using a Biometra TProfessional Thermocycler with cycle times of 94°C$^{2:00}$, (94°C$^{0:30}$, 55°C$^{0:30}$, 68°C$^{2:00}$) x34, 68°C$^{10:00}$, 7°C$^{0:00}$, with the lid pre-heated to 94°C. After PCR, we electrophoresed the samples in a 1% TAE-agarose gel dissolved in 1x TAE buffer at 100V for 60 minutes, and measured bands using HyperLadder 1 kb (Meridian Bioscience). We visualized and captured electrophoretograms using an iBright CL1500 Imaging System (ThermoFisher Scientific). We then purified PCR products with the QIAquick PCR purification Kit, per manufacturer instructions (Qiagen). All amplicon samples were sequenced with either MinION Mk1b or Mk1c devices using R9.4.1 flow cells or flongles (Oxford Nanopore Technologies).

We performed all sequencing and basecalling using ONT's proprietary MinKNOW and *Guppy* software v. 6.5.7. We then performed all bioinformatic analysis using Geneious Prime version 2022.2.1 and ISU *PRRSView.com* (Iowa State University, https://prrsv.vdl.iastate.edu/)

Library preparation and sequencing

We first tested the prrs2-amp primers on viral stock cultures of PRRSV2 strains MN184, VR2332, VR2385, SDSU73, and NADC20 grown on MARC-145 cells (Dr. Pablo Piñeyro, Iowa State University), with the methods described above. We then tested a dilution series containing a mix of two of these viral stock cultures (NADC20 and VR2332 strains). We made this dilution series by extracting RNA from the viral stock cultures and quantifying them using the Qubit RNA HS Assay Kit (Invitrogen by ThermoFisher Scientific) on a Qubit 3 Fluorometer, per manufacturer instructions (ThermoFisher Scientific). We pooled the NADC20 and VR2332 strains at concentration ratios of 1:100, 1:10, 1:1, 10:1, and 100:1, respectively (Table 1). We resumed our experiments as previously described for cDNA synthesis, PCR, and gel electrophoresis.

Table 1 – Absolute quantities of RNA used to prepare the viral stock dilution series.

NADC20:VR2332 Ratio	NADC20 (ng)	VR2332 (ng)
1:100	5	455
1:10	12	108
1:1	60	60
10:1	108	12
100:1	455	5

Next, we tested the primers on serum samples from two SPF pigs inoculated with PRRSV2 strains, NADC20 and VR2385, sampled at 7 dpi (X.J. Meng, Virginia Tech). We also evaluated these primers on twenty serum samples from vaccinated pigs with mixed virus challenges, at 31 days post-vaccination (mixed infection samples, Boehringer-Ingelheim). The experimental groups consisted of seven pigs vaccinated with Ingelvac PRRS MLV vaccine

(Boehringer-Ingelheim), seven pigs vaccinated with Ingelvac PRRS ATP vaccine (Boehringer-Ingelheim), and six unvaccinated pigs. The MLV vaccine is protective against L5 strains, and the ATP vaccine is protective against L8 strains.[34] All pigs were then challenged with a cocktail of PRRSV2 lineages 1, 5, and 8, with a single strain from each lineage. Upon receiving these serum samples, we were also provided with the genome sequences for these challenge lineages and vaccines. The challenge lineages were harvested from wildtype infections, and after BLAST analysis, we determined these lineages corresponded to strains 9982R-S5-L001 (L1), PRR715665-S8-L001 (L5), lung249621-S5-L001 (L8) (GenBank Accession Numbers: MN073092, MN073134, MN0733159). Lastly, we performed a sequencing run to determine demultiplexing and mapping error. We utilized bovine serum as a negative template control (NTC) and a randomly selected positive template control (PTC) (mixed infection sample 132), and performed our experiment as described.

For the long-amplicon sequencing portion of our study, we performed all library preparation using ONT's SQK-LSK109 kit and barcoded samples using EXP-NBD104/114 (Legacy store, Oxford Nanopore Technologies). Samples were sequenced on an Mk1C device using R9.4.1 flow cells for 12-48 hours, depending on the number of samples sequenced.

Bioinformatics analysis

For long-amplicon sequencing, we performed all our basecalling in real-time on the Mk1C using *Guppy*. We altered the demultiplexing command line to be more stringent, only allowing barcode sorting of sequences with the same barcode on both the 5' and 3' ends of cDNA, and not allowing an internal barcode in the basecalled sequence. This reduced our barcode bleeding error rate and improved the overall bioinformatics pipeline (*unpublished data)*.

After demultiplexing, we serially sampled up to 20,000 sequence reads and imported the sequence FASTQ files to Geneious Prime for bioinformatics analysis.

For the viral stock cultures, dilution series, and SPF pig samples, we mapped the imported sequences to the five aforementioned PRRSV2 reference genomes (see Shotgun sequencing), using the Geneious Assembler (default settings). For the mixed infection samples, we mapped sequences to the five reference genomes provided by Boehringer-Ingelheim. We also performed another mapping analysis by mapping the sequences to nine reference genomes, with three representative genomes from lineages 1, 5, and 8, the most prevalent lineages in the U.S (Table 2). [7] We then repeated this analysis using *minimap2* on samples 103, 117, and 121 with default settings.

Table 2 – Summary of lineages used for mapping

Genome	Lineage	Accession Number
pMN184	L1	EF484031
9982R-S5-L001	L1	MN073092
NADC30	L1	MG500776
VR2332	L5	EF536003
PRR715665-S8-L001	L5	MN073134
RespPRRS MLV	L5	AF066183
NADC20	L8	MK837936
lung249621-S5-L001	L8	MN073159
Ingelvac ATP	L8	EF532801

For *de novo* assembly, we sampled the first 4,000 sequences from each mixed infection sample with sensitivity set to low. We analyzed the contigs that measured between 1,530 bp – 1,670 bp, with a 95% High Quality read percentage, and used the PRRSV ORF5 BLAST Tool on Iowa State University's *PRRSView* website to classify the contigs by their lineages (https://prrsv.vdl.iastate.edu/).

RESULTS

Shotgun sequencing of inoculated pig serum

After reviewing results from Centrifuge, we found the SPF-MN184 sample was positive for both PRRSV2 strain MN184, and Porcine parvovirus-6 (PPV6). *Centrifuge results from KL will be added later.* Upon mapping the sequences to the five reference genomes in Geneious Prime we found 1/3,384 reads mapped to MN184 and 18/3,384 reads mapped to VR2385 (Figure 1a, 1b). However, the reads mapped to VR2385 were all aligned at the 3' poly-A tail of the viral genome, indicating inappropriate mapping. The single read mapped to MN184 covered a 706 bp portion of ORF1, with 90.72% identity after pairwise alignment.

The Centrifuge results hitting to PPV6 were unexpected, so we decided to try mapping the reads to randomly selected, representative porcine parvovirus genomes (GenBank Accession Numbers: OR046036, KP765690, MZ577032, NC_014665, OR046032, MW853956). In doing so, we found that 63/3,384 reads mapped to the PPV6 strain with dispersed coverage throughout the genome (Figure 1c). BLAST analysis of the consensus sequence found a 99.21% identity to PPV isolate PPV6-KF4. This demonstrates the ability of shotgun metagenomic sequencing to identify nontarget pathogens, but with such low efficiency and specificity, this is currently not a viable approach.

Long-amplicon sequencing of viral cultures and pig sera

Gel electrophoresis

To enrich for target sequence, long amplicons were generated using the prrs2-amp primers. All the agarose gels for the amplifications performed with our prrs2-amp primers revealed consistent bands at approximately 1,500 base pairs, as expected (Figure 2). Mapping with Geneious Prime showed the amplicons aligned appropriately within ORF3 to ORF7 (Figure 3).

Demultiplexing command

We tested the stringency of our altered demultiplexing command by sequencing an NTC and PTC sample together. In this analysis we found the PTC mapped 99.84% of reads to their appropriate genome (L1), while the NTC did not produce any mapped results to PRRSV2 reference genomes. This implies that any sequences pulled from the NTC were likely the ONT barcode sequence used encompassing some non-specific binding of the NTC sample, and not PRRSV2 sequences. This confirms that our demultiplexing command prevented sorting of improperly barcoded samples.

Viral stock samples

After mapping the imported sequences of the viral stock cultures to the five PRRSV2 reference genomes, we found at least 99.87% of the imported reads mapped to their corresponding genomes across all samples (Table 3). For the MN184 culture, 99.98% of reads mapped to MN184, with only 0.02% of reads mapping to VR2385. Viral stock strain NADC20 was the only sample to map to all five reference genomes.

Dilution series of mixed viral stock samples

Upon analysis of the sequencing reads, we noticed a potentially direct relationship of mapped reads to concentration of each strain (Table 3). The higher the concentration of NADC20 to VR2332, the more reads mapped to NADC20, and the fewer to VR2332, and vice versa (Figure 4). The 1:1 dilution was created with 60 nanograms of extracted RNA from each strain, and we expected to have an equal number of reads mapped between the two strains. However, we found that 41.59% of reads mapped to strain NADC20, and 58.23% mapped to strain

VR2332. The 10:1 dilution mapped sequences to all reference genomes, but with 99.92% of reads mapping to either VR2332 or NADC20. Thus far, we have shown our primers are effective at amplifying the PRRSV2 genome in various concentrations of viral stocks, down to 5 ng of starting RNA, and in viral stock mixes.

Specific pathogen-free serum samples

After analysis with Geneious Prime, 99.76% of reads from the SPF-NADC20 sample mapped to NADC20, and 99.91% of reads from SPF-VR2385 mapped to VR2385 (Table 3). We also began seeing reads mapping with low coverage ($n < 13$), to other genomes. In these low-coverage cases, we extracted the consensus sequence from these results and compared to the best mapped to genome, and found 33.46% identity, at best. We saw this pattern repeat itself in future bioinformatic analyses of serum samples when there was poor coverage to a reference genome, insinuating the need to be judicious about interpreting results with poor coverage and/or low-quality reads. We established a cut-off of 60x coverage based on studies published by Bull et al., and Whitford et al., and the limitations described above, to consider a result positive. [21,26]

Mixed infection serum samples

Table 4 provides a detailed description of the sequencing results and mapped reads for all twenty mixed infection samples. In summary, all samples mapped with great coverage to the L1 sequence (27.72% at minimum, 100% at maximum). Samples 103, 117, 123, and 127, all of which were unvaccinated, also mapped to L5 and L8. Samples 111, 119, 121, 131, and 133, which were either unvaccinated or vaccinated with the MLV vaccine, also mapped to L8. Sample 131 is also the only sample to have mapped best to L8 (72.18%) rather than L1 (27.72%). Samples 110, 121, and 129, which were vaccinated with the MLV vaccine, mapped 1.51 – 39.27% of reads to the MLV genome. With this first analysis, we have shown the potential to identify mixed infections in serum samples. However, this approach is not indicative of a real-world scenario, so we decided to expand our mapping approach.

To make this study more clinically relevant, we mapped these mixed infection samples to nine reference genomes representative of the three most relevant PRRSV2 lineages. We expected the unvaccinated animals in this experimental group to all be positive for the L1, L5, and L8 inoculum strains. All these unvaccinated animals were positive for L1 (9982R-S5-L001) and L8 (249621-S5-L001). Of the unvaccinated animals, 103, 117, and 127 were positive for at least one of the pathogenic L5 strains. However, when evaluated as a collective, animals 123 and 127 are also positive for L5. The VR2332 and PRR715665-S8-L001 genomes have a 99.62% identity, which we suspect impaired with the mapping ability of our software.

We tested this by mapping the same sequences to the VR2385 genome rather than PRR715665-S8-L001 and found they mapped considerably better to VR2332 in this scenario (*unpublished data)*. We found the same pattern in MLV-vaccinated samples 110, 121, and 129, except in this scenario reads mapped between the VR2332 and MLV genomes. The MLV vaccine shares a 99. 98% identity with its parent strain, VR2332. We do not suspect these animals were infected with the L5 inoculum, but the software algorithm could not appropriately map reads to the pathogenic and vaccine strains because of their high percent identity. With these results, we recommend researchers interpret their results with prudence and understand the limitations of the equipment being utilized.

Regardless of this limitation, we identified animals 119, 121, and 133 were infected with the L8 inoculum, as they were not protected by the MLV vaccine. None of the animals vaccinated with the ATP vaccine were positive for either the L5 or L8 inoculum, though they were all positive for L1. Only animal 130 had a more than six reads mapped to the ATP genome.

However, we considered this result inconclusive as we noted all these amplicons mapped accurately within the ORF3-ORF7 region, and the extracted consensus sequence had a 99.47% identity to the ATP genome.

Ultimately, we were able to identify mixed infections in nine of the twenty inoculated pigs and discerned the vaccine history in three MLV-vaccinated pigs. There are limitations to this bioinformatical approach, and they ought to be considered. Nevertheless, this method provided accurate, timely results which suggests that mapping to reference genomes is a practical approach to mixed infection identification of PRRSV2.

Mapping with *minimap2* provided similar results to the above, but with some important differences. Samples 103, 117, and 121 were positive for the same inoculum strains, L1 and L8, with similar depths of coverage as in the previous analysis. However, based on the *minimap2* results, 103 and 117 were also erroneously positive for NADC20, which was not utilized in this infection experiment. The lineage 5 results also had similar numbers of reads mapped to all three L5 strains. Sample 103 had 310, 308, and 308 reads mapped to VR2332, PRR715665-S8-L001, and the MLV genome, respectively. This same pattern repeated itself in samples 117 and 121. With these numbers in mind, we found the total number of reads reported ($n \geq 21{,}791$) exceeded the number of reads used for the mapping command ($n \leq 19{,}826$). We suspect the *minimap2* tool may have difficulty determining what genome to align a read to when genomes have a high percent identity, and so maps the same read multiple times. This method overinflates results which can lead to inappropriate conclusions, and we discourage the use of the *minimap2* aligner in these methods.

We also performed *de novo* assembly for analysis of the mixed infection samples and uploaded the selected contigs ($n \geq 14$) to *PRRSview*, where we found similar results to the mapping analyses. L1A was identified in all samples. L5 was identified in samples 103 and 121, and L8 was found in samples 103, 111, 123, 127, and 131. However, this method failed to identify at least one lineage in nine of twenty samples (Table 2), when compared to the mapping approach. We believe this may be due to the stringent contig selection parameters established. Regardless, we believe *de novo* assembly is not a feasible method when used alone for identifying multiple strains of PRRSV2.

This assembly method requires significant computational power to be performed in a timely manner. Assembling 4,000 reads from ten mixed infection samples by *de novo* required ~72 hours, with our computer solely running Geneious Prime (iMac, OS Catalina, 3.1 GHz Quad-Core Intel Core i7, 8 GB RAM) ultimately yielding less information than we retrieved by mapping. The long assembly time could be overcome by using a computer with higher processing ability, such as the newly announced iMac with M3 chip, or a supercomputer.[35] However, this defeats the purpose of ONTs overarching goal to "sequence anything, anywhere, anytime."[32] Because of this, we believe that by mapping to reference genomes we can gain accurate results, so long as the researcher understands the limitations of this approach.

Table 3 – Summary of mapping results to five reference genomes from viral stock cultures, viral stock dilution series, and SPF pig serum samples.												
Experiment	Sample	Total mapped reads	MN184	MN184 %*	NADC20	NADC20 %*	SDSU73	SDSU73 %*	VR2332	VR2332 %*	VR2385	VR2385%*
Viral Stock Cultures	MN184	19,597	19,594	99.98%	0	0.00%	0	0.00%	0	0.00%	3	0.02%
	NADC20	19,643	1	0.01%	19,617	99.87%	18	0.09%	3	0.02%	4	0.02%
	SDSU73	19,776	1	0.01%	7	0.04%	19,765	99.94%	0	0.00%	3	0.02%
	VR2332	19,826	1	0.01%	2	0.01%	0	0.00%	19,818	99.96%	5	0.03%
	VR2385	19,223	0	0.00%	7	0.04%	0	0.00%	15	0.08%	19,201	99.89%
Dilution series of viral stock cultures (NADC20: VR2332)	1:100	8,876	0	0.00%	32	0.36%	0	0.00%	8,841	99.61%	3	0.03%
	1:10	13,040	0	0.00%	666	5.11%	1	0.01%	12,369	94.85%	4	0.03%
	1:1	7,494	0	0.00%	3,117	41.59%	10	0.13%	4,364	58.23%	3	0.04%
	10:1	19,584	1	0.01%	17,818	90.98%	13	0.07%	1,750	8.94%	2	0.01%
	100:1	14,478	0	0.00%	14,307	98.82%	8	0.06%	161	1.11%	2	0.01%
SPF serum samples	NADC20†	7,537	1	0.01%	7,519	99.76%	13	0.17%	2	0.03%	2	0.03%
	VR2385†	16,255	0	0.00%	4	0.02%	1	0.01%	9	0.06%	16,241	99.91%

* Percentages describe the number of mapped reads to each genome to the total number of mapped reads obtained.
† The strain of PRRSV2 the pathogen-free pig was inoculated with.

Table 4 – Results of mixed infection samples mapped to the reference genomes provided by Boehringer Ingelheim

Sample	Vaccinated with*	Total mapped reads	Lineage 1	Lineage 5	Lineage 8	MLV	ATP	Lineages identified by *PRRSView*
102	ATP	19,865	19,864	0	1	0	0	L1A
103	None	19,826	15,992	338	3,494	1	1	L1A, L5, L8
104	ATP	19,820	19,820	0	0	0	0	L1A
106	MLV	19,543	19,542	0	0	0	1	L1A
110	MLV	7,792	7,674	0	0	118	0	L1A
111	None	19,652	18,597	19	1,035	1	0	L1A, L8
114	ATP	19,847	19,845	1	0	0	1	L1A
117	None	19,407	15,314	231	3,862	0	0	L1A, L8
119	MLV	19,863	19,767	2	74	20	0	L1A
121	MLV	19,654	11,773	2	158	7,720	1	L1A, L5
123	None	19,540	17,952	67	1,519	1	1	L1A, L8
124	ATP	19,902	19,894	0	2	6	0	L1A
125	ATP	19,783	19,775	8	0	0	0	L1A
127	None	19,544	11,457	193	7,893	1	0	L1A, L8
129	MLV	19,448	19,164	0	0	284	0	L1A
130	ATP	19,858	19,821	0	0	0	37	L1A
131	None	14,432	4,000	13	10,417	0	2	L1A, L8
132	ATP	16,562	16,558	4	0	0	0	L1A
133	MLV	19,360	19,226	0	134	0	0	L1A
136	MLV	16,501	16,500	0	0	1	0	L1A
NTC + PTC	NTC	0	0	0	0	0	0	-
	PTC	16,105	16,079	23	1	2	0	-

* ATP: Ingelvac PRRS ATP vaccine; MLV: Ingelvac PRRS MLV vaccine

Table 5 – Results of mixed infection samples mapped to nine representative genomes from lineages 1, 5, and 8.

				Lineage 1			Lineage 5			Lineage 8		
Sample	Vaccine	Challenge	Total mapped reads	MN184	9982R-S5-L001	NADC30	VR2332	PRR715665-S8-L001	RespPRRS MLV	NADC20	249621-S5-L001	Ingelvac ATP
102	ATP	1, 5, 8	19,861	0	19,860	0	0	0	0	1	0	0
103	-	1, 5, 8	19,826	0	16,003	1	137	90	81	17	3,497	0
104	ATP	1, 5, 8	19,819	0	19,817	2	0	0	0	0	0	0
106	MLV	1, 5, 8	19,545	1	19,542	1	1	0	0	0	0	0
110	MLV	1, 5, 8	7,791†	0	7,672	0	8	58*	53*	0	0	0
111	-	1, 5, 8	19,652	0	18,596	1	10	4	5	2	1,034	0
114	ATP	1, 5, 8	19,848	1	19,846	1	0	0	0	0	0	0
117	-	1, 5, 8	19,402	0	15,311	0	93	67	67	4	3,859	1
119	MLV	1, 5, 8	19,861	0	19,766	2	3	10	6	0	74	0
121	MLV	1, 5, 8	19,660	2	11,772	1	1,577	3,086	3,061	4	156	1
123	-	1, 5, 8	19,538	0	17,949	4	28	21	20	2	1,514	0
124	ATP	1, 5, 8	19,905	1	19,895	1	0	0	0	0	2	6
125	ATP	1, 5, 8	19,777	0	19,771	0	4	1	1	0	0	0
127	-	1, 5, 8	19,531	1	11,475	1	82	41*	54*	9	7,877	0
129	MLV	1, 5, 8	19,448	1	19,160	2	65	118	102	0	0	0
130	ATP	1, 5, 8	19,857	1	19,818	1	0	0	0	0	0	37
131	-	1, 5, 8	14,432	0	4,019	0	0	0	0	6	10,405	2
132	ATP	1, 5, 8	16,562	0	16,555	3	1	1	2	0	0	0
133	MLV	1, 5, 8	19,349	0	19,215	0	0	0	0	1	133	0
136	MLV	1, 5, 8	16,498	0	16,495	2	0	1	0	0	0	0

*Due to the similarity of these genomes and likely mapping error, we consider these samples positive.
†Despite equimolar flow cell loading, this sample had 8,876 passed reads, of which 7,791 were used for mapping.

DISCUSSION

Shotgun sequencing with nanopore

In reviewing our results from shotgun sequencing, we do not believe this is a feasible approach for nanopore sequencing of PRRSV2. Neither of the PRRSV2 results are reliable enough to be conclusive for diagnosis. However, the analysis with Centrifuge proved interesting in that it was able to identify a mixed infection of PRRSV2 and PPV6, further confirmed by mapping with Geneious. Perhaps there is potential for collateral sequencing of pathogens using the MinION, and this may be a future avenue of research.

Primer Design and Applicability

We have effectively designed a primer set that has consistently proven to amplify approximately 1,500 bp, from the middle of ORF 3 to 7, across all the samples tested. We recommend using these primers as a universal set for future experiments involving PRRSV2. Diel et al. recently published a similar targeted amplicon sequencing method[27], though we did not utilize rtPCR for detection of PRRSV2 in samples prior to sequencing, and we found our prrs2-amp primers with nanopore sequencing to be useful in mixed infections as well.

Challenges with demultiplexing

One of the features we tried to highlight in this study is the MinION's capacity for sample multiplexing. However, demultiplexing can sometimes be a challenge due to inappropriate bin-sorting or errors in the native barcoding process. Knowing the potential issues with demultiplexing, we decided to increase the stringency of our demultiplexing command. We only allowed bin sorting of samples that possessed the same native barcode (NBD) attached at the 5' and 3' ends of the sequence and did not contain an NBD in the middle of the sequence – a known error during barcoding. This greatly reduced the total number of reads sorted but increased the overall stringency of sorting. After doing so, we yielded the results shown, with very few (0-2) instances of "barcode bleeding." We recommend this, or similar, approaches for future multiplexing studies.

Sequencing of viral cultures

Upon sequencing the viral stock cultures and viral stock dilution series, bioinformatic analyses showed a minimum 99.87% mapping accuracy of sequenced amplicons to their respective reference genome (Table 2), signifying high quality reads and a reliable mapping result.

In the dilution series, we found a potential direct relationship pattern between RNA concentration to the number of reads mapping to the respective reference genome. In the samples with low concentrations of NADC20 or VR2385, we saw low coverage to the NADC20 or VR2385 genomes (Figure 4), and vice versa. These samples proved our ability to identify mixed infections in viral stock, even at low starting RNA concentrations. A limit of detection analysis would also be beneficial in this scenario to better determine the minimum viral load required to accurately identify PRRSV2 strains.

Sequencing of serum samples

The experimental method elucidated here, along with other similar studies, [27,36] have the potential of becoming a novel diagnostic test for PRRSV2. In utilizing this approach, we were able to ultimately identify mixed infections in the serum of twenty pigs inoculated with a cocktail of PRRSV2 L1, L5, and L8 strains, and accurately identified the vaccine status of three pigs, 31 days after their vaccination.

Regardless, discretion is required from the researcher or technician performing this analysis. By mapping to reference genomes, we allow software to decide what reads best align to pre-determined genomes, which may be inaccurate when too few reference genomes are used. To overcome this limitation, we encourage the establishment of a universally available PRRSV2 genome database. By utilizing a comprehensive, universally available genome bank of clinically relevant strains of PRRSV2, a researcher or diagnostic lab could readily map their nanopore sequencing results to thousands of curated PRRSV2 genomes, thus easily identifying mixed infections, vaccine strains, and mutant, recombinant strains of the virus.

One of the hallmarks of the swine industry in the U.S. is the transportation and mingling of animals across state lines into nurseries, grower facilities, or finishing farms. By utilizing our prrs2-amp primers and nanopore sequencing to diagnose mixed infections in pigs, veterinarians can better inform producers of at least the lineages, or, at best, the individual strains present in their pigs. By only mingling pigs with similar PRRSV2 infection histories, the effects of recombination and the likelihood of highly pathogenic outbreaks decrease. This approach can also be helpful in identifying the cause of outbreaks that occur shortly after vaccination with MLVs. Veterinarians can determine whether the outbreak is caused by the vaccine itself reverting to a wildtype form, or by another viral strain present in the herd, thereby elucidating the safety and efficacy of the utilized vaccine.

ACKNOWLEDGEMENTS

Thank you to all those who contributed to this work. We thank Dr. X.J. Meng for providing samples without which our work could not have been done. We also thank Ms. Annmarie Budde for her technical assistance, and Dr. Nadia Saklou, and Catie Burgess for their guidance during this endeavor.

DECLARATION OF CONFLICTING INTERESTS

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

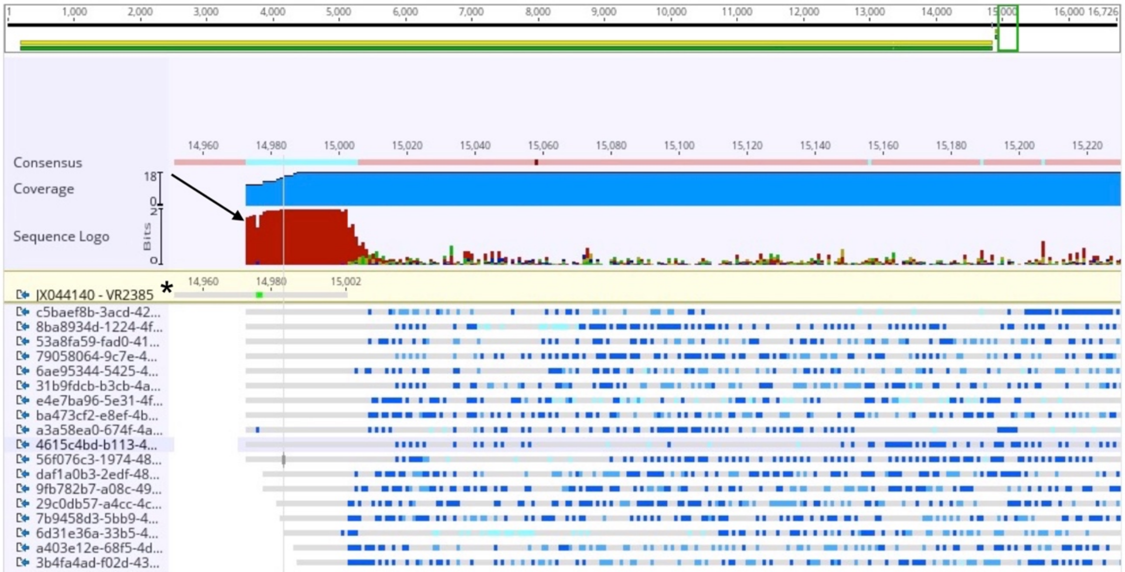
FUNDING

United States Department of Agriculture – National Institute of Food and Agriculture

National Pork Board

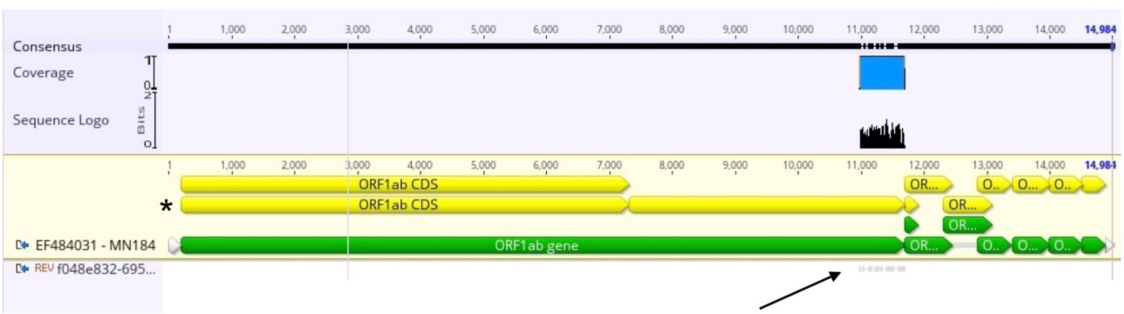
NC229

FIGURES



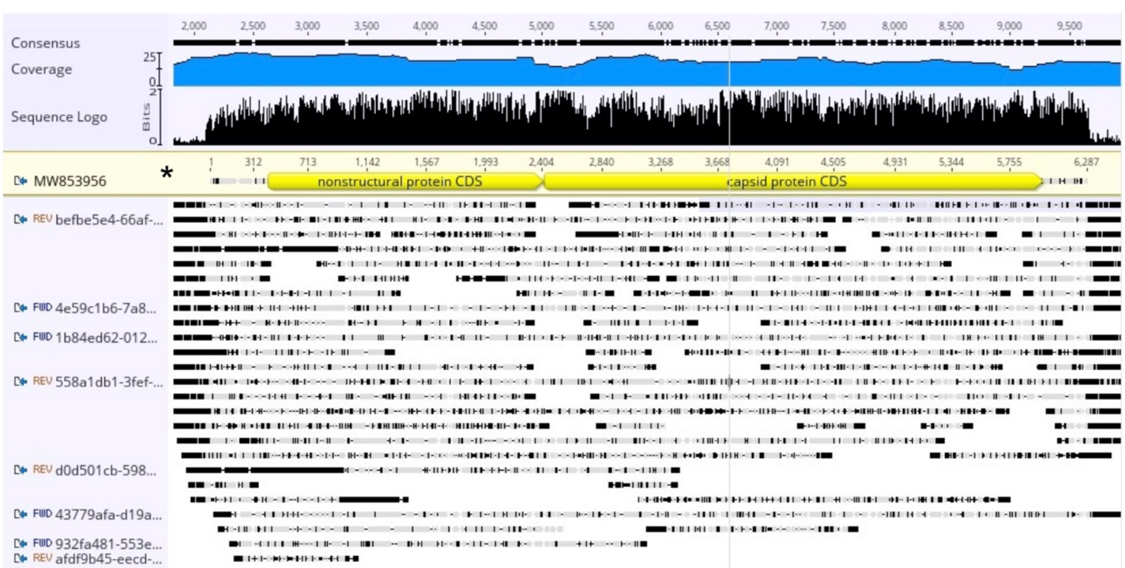
Consensus
Coverage
Sequence Logo
JX044140 - VR2385

A



Consensus
Coverage
Sequence Logo
ORF1ab CDS
ORF1ab CDS
EF484031 - MN184
ORF1ab gene
REV f048e832-695...

B



Consensus
Coverage
Sequence Logo
MW853956
nonstructural protein CDS
capsid protein CDS

C

Figure 1 – Screenshots of Geneious Prime showing shotgun mapping results. A) The reference genome, VR2385, is shown in the yellow area denoted by an asterisk, with the poly-A tail shown for context (arrow). The dashed lines below the reference genome are the sequences from the MinION which have mapped to the poly-A tail of VR2385. B) Screenshot of Geneious Prime. The MN184 genome is shown in the yellow area, denoted by an asterisk, with annotations of ORF1 – ORF7. A single sequence from SPF-MN184 is shown mapping at the 5' end of the ORF1ab gene (arrow), spanning 706 bp. C) The PPV6 reference genome, AH-PPV620178-3 (asterisk) is shown in its entirety. Sequences from the SPF-MN184 sample are shown mapping throughout the PPV6 reference genome, with a coverage of 63/3,384 reads. The consensus sequence was extracted and analyzed by BLAST, which produced a 99.21% identity to PPV6-KF4.

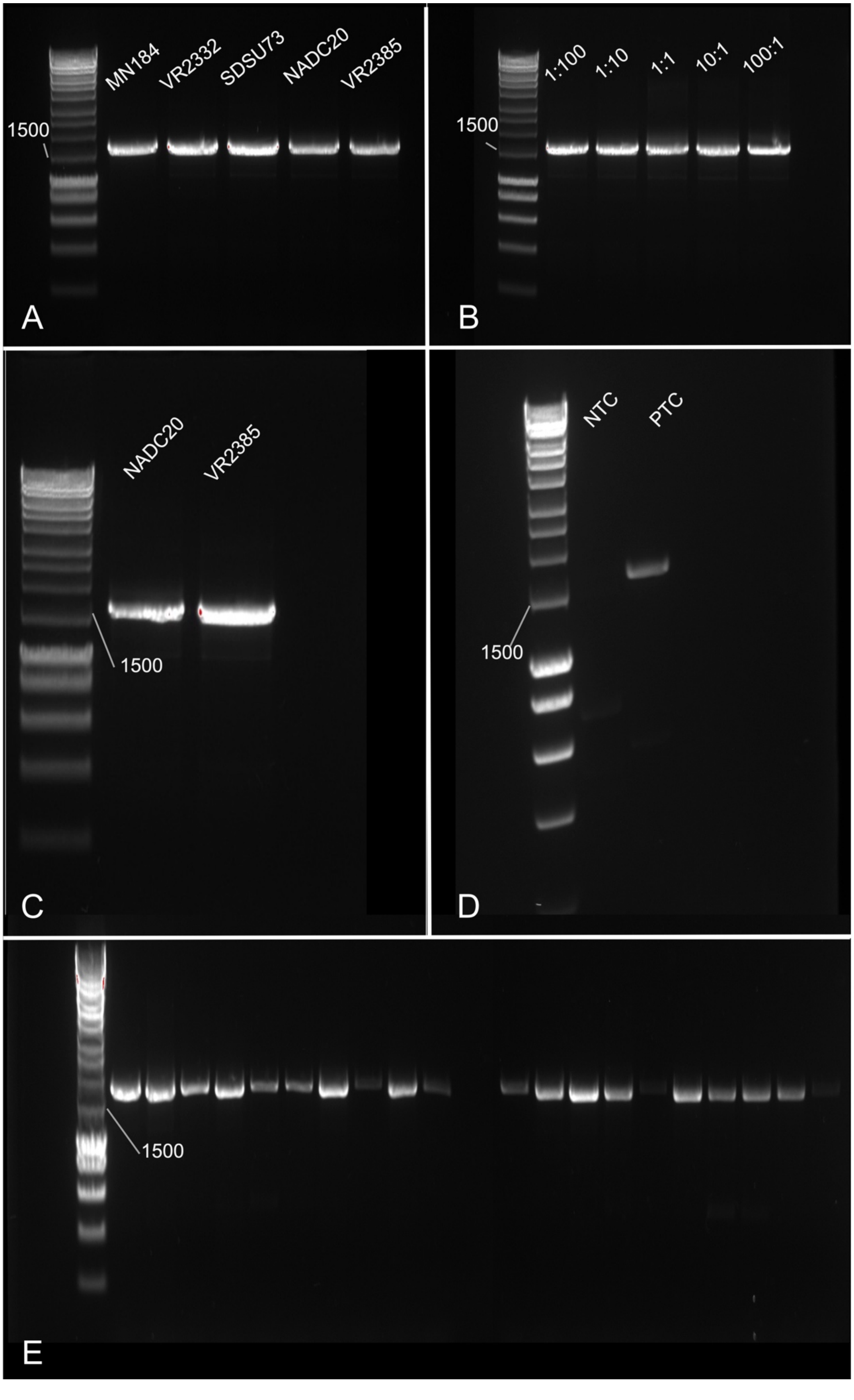

MN184
VR2332
SDSU73
NADC20
VR2385
1500
A
1:100
1:10
1:1
10:1
100:1
1500
B
NADC20
VR2385
1500
C
NTC
PTC
1500
D
1500
E

Figure 2 – Agarose gel electrophoresis (1% TAE-agarose gel) images of samples tested with prrs2-amp primers. Lane 1 contains HyperLadder 1 kb marker (Meridian Bioscience) in all gels. A) Viral stock cultures. Lanes 2-6 are of the labeled viral stocks, lane 7 is the NTC. B) Dilution series of viral stock cultures. Lanes 2-6 include the dilution series. Lane 7 is the NTC. C) Serum from SPF pigs inoculated with NADC20 (Lane 2) and VR2385 (Lane 3). Lane 4 is of the NTC. D) Negative (Lane 2) and positive (Lane 3) template control series. E) Mixed infection/vaccination samples. Lanes 1-10 and 12-21 are of the mixed infection samples, lane 11 is of the NTC.

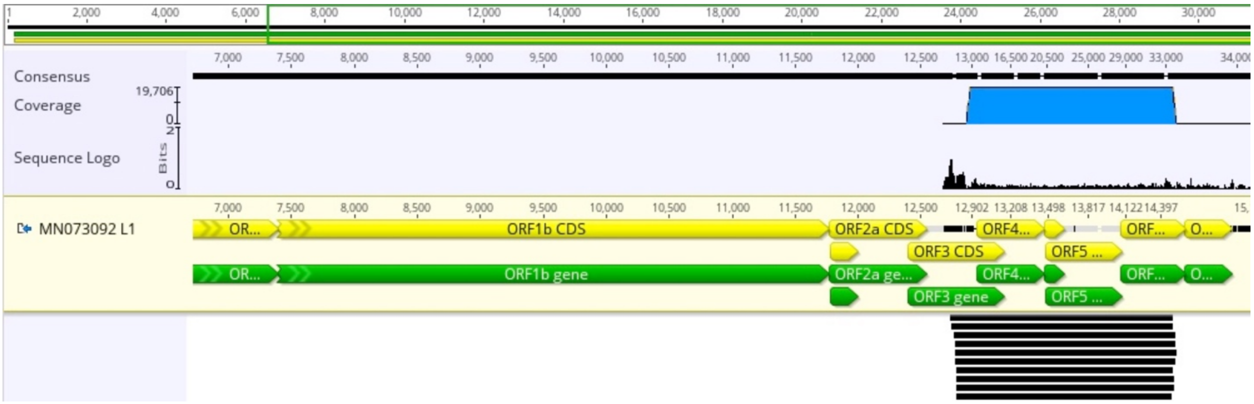


Figure 3 – Screenshot of Geneious Prime showing amplicon aligning to a PRRSV2 genome (L1 inoculum). The amplicons, shown as black bars below the reference genome, consistently align from the middle of ORF3 up to ORF7, for approximately 1,537 bp.

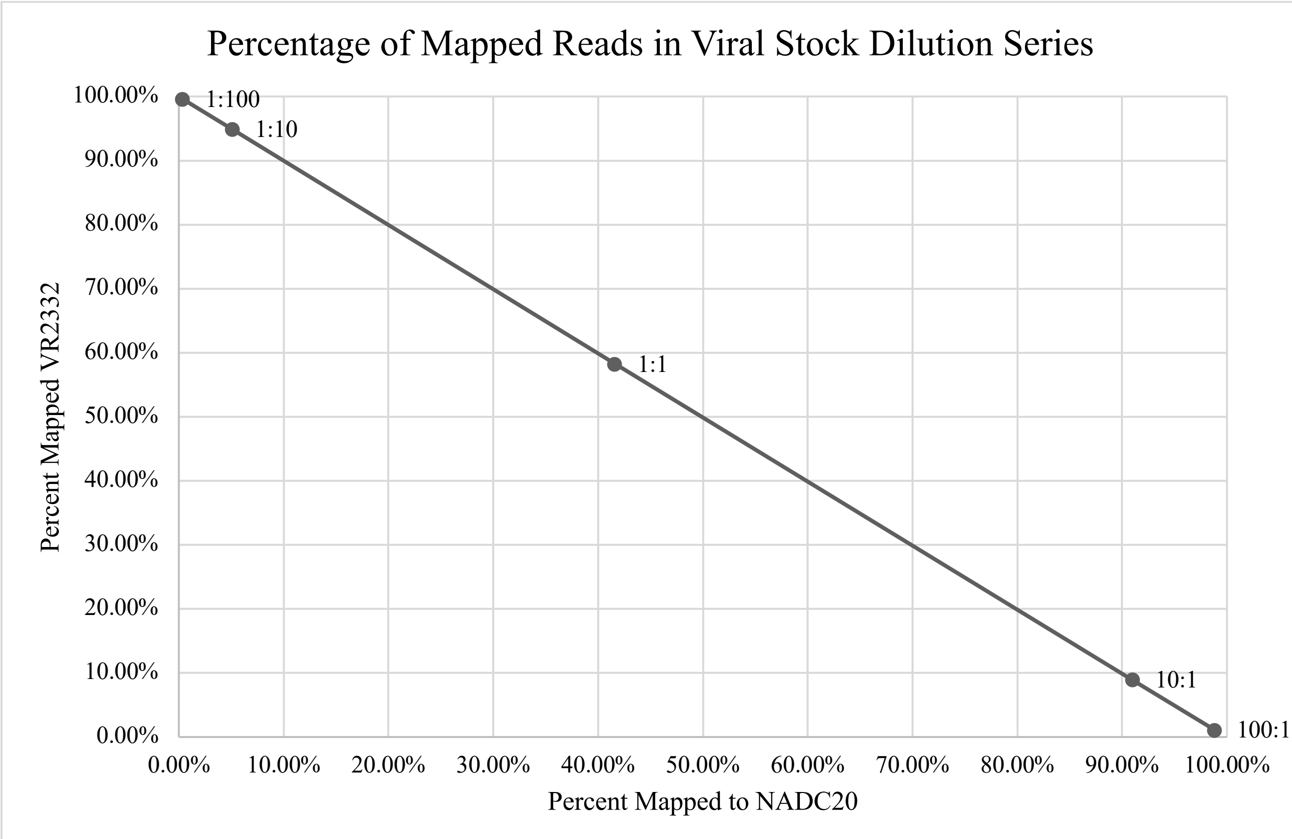


Figure 4 – Graph demonstrating the direct relationship between the starting RNA load and percentage of reads mapped to the corresponding PRRSV2 genome in SPF serum samples. As the starting RNA load increases of the SPF sample, so do the number of reads mapped to the corresponding genome, and vice versa.

REFERENCES

1. Montaner-Tarbes S, Del Portillo HA, Montoya M, Fraile L. Key Gaps in the Knowledge of the Porcine Respiratory Reproductive Syndrome Virus (PRRSV). *Front Vet Sci* 2019;6:38.

2. Rossow KD. Review Article Porcine Reproductive and Respiratory Syndrome Virus. *Journal of Veterinary Pathology* 1998;35:1-20.

3. Neumann EJ, Kliebenstein, J. B., Johnson, C. D., Mabry, J. W., Bush, E. J., Seitzinger, A. H., Green, A. L., Zimmermann, J. J. Assessment of the economic impact of porcine reproductive and respiratory syndrome on swine production in the United States. *Journal of the American Veterinary Medical Association* 2005;227:385-392.

4. Melmer DJ, Friendship, R., O'Sullivan, T. L., Greer, A. L., Novosel, D., Ojkic, D., Poljak, Z. Classification of porcine reproductive and respiratory syndrome virus in Ontario using Bayesian phylogenetics and assessment of temporal trends. *The Canadian Journal of Veterinary Research* 2020;85:83-92.

5. Shi M, Lam TT, Hon CC, et al. Molecular epidemiology of PRRSV: a phylogenetic perspective. *Virus Res* 2010;154:7-17.

6. Hanada K, Suzuki Y, Nakane T, et al. The origin and evolution of porcine reproductive and respiratory syndrome viruses. *Mol Biol Evol* 2005;22:1024-1031.

7. Shi M, Lam TT, Hon CC, et al. Phylogeny-based evolutionary, demographical, and geographical dissection of North American type 2 porcine reproductive and respiratory syndrome viruses. *J Virol* 2010;84:8700-8711.

8. Gerbera PF, Giménez-Lirola, L. G., Halbura, P. G., Zhou, L., Meng, X.J., Opriessnig, T. Comparison of commercial enzyme-linked immunosorbent assays and fluorescent microbead

immunoassays for detection of antibodies against porcine reproductive and respiratory syndrome virus in boars. *Journal of Virological Methods* 2013;197:63-66.

9. Sattler T, Wodak, E., Revilla-Fernandez, S., Schmoll, F. Comparison of different commercial ELISAs for detection of antibodies against porcine respiratory and reproductive syndrome virus in serum. *BMC Veterinary Research* 2014;10:2-6.

10. Prickett J, Simer, R., Christopher-Hennings, J., Yoon, K., Evans, R.B., Zimmerman, J.J. Detection of Porcine reproductive and respiratory syndrome virus infection in porcine oral fluid samples- a longitudinal study under experimental conditions. *Journal of Veterinary Diagnostic Investigation* 2008;20:156-163.

11. Dee SA, Torremorell, M., Rossow, K., Mahlum, C., Otake, S., Faaberg, K. Identification of genetically diverse sequences (ORF 5) of porcine reproductive and respiratory syndrome virus in a swine herd. *The Canadian Journal of Veterinary Research* 2001;65:254-260.

12. Martin-Valls GE, Cortey, M., Allepuz, A., Illas, F., Mateu, E. Description of a New Clade within Subtype 1 of Betaarterivirus suid 1 Causing Severe Outbreaks in Spain. *American Society for Microbiology Journals* 2022;11:1-3.

13. Guzman M, Melendez R, Jimenez C, et al. Analysis of ORF5 sequences of Porcine Reproductive and Respiratory Syndrome virus (PRRSV) circulating within swine farms in Costa Rica. *BMC Vet Res* 2021;17:217.

14. Kang H, Yu JE, Shin JE, et al. Geographic distribution and molecular analysis of porcine reproductive and respiratory syndrome viruses circulating in swine farms in the Republic of Korea between 2013 and 2016. *BMC Vet Res* 2018;14:160.

15. Martin-Valls GE, Kvisgaard LK, Tello M, et al. Analysis of ORF5 and full-length genome sequences of porcine reproductive and respiratory syndrome virus isolates of genotypes 1 and 2

retrieved worldwide provides evidence that recombination is a common phenomenon and may produce mosaic isolates. *J Virol* 2014;88:3170-3181.

16. Andreyev VG, Wesley R. D., Mengeling, W. L., Vorwald, A. C., Lager, K. M. Genetic variation and phylogenetic relationships of 22 porcine reproductive and respiratory syndrome virus (PRRSV) field strains based on sequence analysis of open reading frame 5. *Archives of Virology* 1996;142:993-1001.

17. Cui XY, Xia DS, Huang XY, et al. Recombinant characteristics, pathogenicity, and viral shedding of a novel PRRSV variant derived from twice inter-lineage recombination. *Vet Microbiol* 2022;271:109476.

18. Cha SH, Chang CC, Yoon KJ. Instability of the restriction fragment length polymorphism pattern of open reading frame 5 of porcine reproductive and respiratory syndrome virus during sequential pig-to-pig passages. *J Clin Microbiol* 2004;42:4462-4467.

19. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* 2016;14:265-279.

20. Butt SL, Erwood, E. C., Zhang, J., Sellers, H. S., Young, K., Lahmers, K. K., Stanton, J. B. Real-time, MinION-based, amplicon sequencing for lineage typing of infectious bronchitis virus from upper respiratory samples. *Journal of Veterinary Diagnostic Investigation* 2021;33:179-190.

21. Bull RA, Adikari TN, Ferguson JM, et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat Commun* 2020;11:6272.

22. Kilianski A, Haas JL, Corriveau EJ, et al. Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. *Gigascience* 2015;4:12.

23. Castro-Wallace SL, Chiu CY, John KK, et al. Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Sci Rep* 2017;7:18022.

24. Hoenen T, Groseth A, Rosenke K, et al. Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. *Emerg Infect Dis* 2016;22:331-334.

25. Kono N, Arakawa K. Nanopore sequencing: Review of potential applications in functional genomics. *Dev Growth Differ* 2019;61:316-326.

26. Whitford W, Hawkins V, Moodley KS, et al. Proof of concept for multiplex amplicon sequencing for mutation identification using the MinION nanopore sequencer. *Sci Rep* 2022;12:8572.

27. Caserta LC, Zhang J, Pineyro P, Diel DG. Rapid genotyping of porcine reproductive and respiratory syndrome virus (PRRSV) using MinION nanopore sequencing. *PLoS One* 2023;18:e0282767.

28. King J, Harder T, Beer M, Pohlmann A. Rapid multiplex MinION nanopore sequencing workflow for Influenza A viruses. *BMC Infect Dis* 2020;20:648.

29. Stahl-Rommel S, Jain M, Nguyen HN, et al. Real-Time Culture-Independent Microbial Profiling Onboard the International Space Station Using Nanopore Sequencing. *Genes (Basel)* 2021;12.

30. Hansen S, Faye O, Sanabani SS, et al. Zika Virus Amplification Using Strand Displacement Isothermal Method and Sequencing Using Nanopore Technology. *Methods Mol Biol* 2020;2142:123-136.

31. Laver T, Harrison J, O'Neill PA, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol Detect Quantif* 2015;3:1-8.

32. Oxford Nanopore Technologies. 2023; https://nanoporetech.com/. Accessed October 28, 2023.

33. Kim D, Song, L., Breitwieser, F.P., Salzberg S.L. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Research* 2016;26.

34. Zeller MA SA, Sharma A, Trevisan G, Zhang J, Harmon K, Gauger, PC. ISU PRRSView, 2023.

35. Apple. Apple Event 2023 Scary Fast, 2023.

36. Gagnon CA, Lalonde, C., Provost, C. Porcine reproductive and respiratory syndrome virus whole-genome sequencing efficacy with field clinical samples using a poly(A)-tail viral genome purification method. *Journal of Veterinary Diagnostic Investigation* 2021;33:216-226.

CHAPTER III: CONCLUSION AND FUTURE DIRECTIONS

In this study we have identified mixed infections of PRRSV2 in viral stocks, and the serum of pigs challenged with individual strains, and a cocktail of strains from lineages 1, 5, and 8, using Nanopore sequencing. We were also able to identify the vaccine status of three animals vaccinated with the PRRS MLV vaccine, 31 days post-vaccination. With the materials and methods described, we estimate a total cost of $34 in consumables, per sample for 24 samples tested, and 3-5 days of turnaround time from when samples arrive to the laboratory until results are submitted. When including the cost of technician time, this makes our approach competitive with currently available diagnostic tests for PRRSV2.

When we originally began the primer selection process in this project, we attempted to amplify ORFs 1-7 of the genome, but were unsuccessful. We encourage researchers to continue searching for primers that can span larger portions of the PRRSV2 genome for more accurate strain identification. We also encourage experimentation with oral fluid samples, in addition to serum samples. In pigs, blood is often drawn from an ear vein or, when unsuccessful, from the cranial vena cava, with potentially lethal consequences. This is a stressful procedure, making oral fluids a much more attractive sampling method for producers, veterinarians, and pigs. Rope sampling is an inexpensive and effective technique for pooling saliva from multiple pigs in a pen, and ought to be considered as a future sampling method of study.

The study presented here provides a pathway to future avenues of research into Nanopore sequencing for PRRSV2 diagnosis. A limit of detection study is warranted to determine how soon after infection the disease can be diagnosed. We also encourage the development of a universally accessible database of clinically relevant strains of PRRSV2, and their geographic distribution. These studies would improve surveillance efforts for PRRSV2, which can help in the prevention of future outbreaks, and the prediction of novel, pathogenic strains of virus.

One of the limitations we found in our research is the importance of selecting the appropriate bioinformatics software and pipeline for the research being conducted. By understanding the advantages and disadvantages of bioinformatical techniques, such as mapping to reference genomes and *de novo* assembly, the researcher can more accurately interpret the results of their study and determine when further analysis is required.

This project provides a promising starting point for the development of a novel diagnostic test for PRRSV2. ELISA is a rapid, inexpensive testing method that only provides information about the exposure of an animal to PRRS, and cannot distinguish between infections, vaccinations, or Types 1 and 2. RFLP analysis is slow, outdated, provides little information about the PRRSV2 genome, and can also be an inaccurate test when trying to identify novel strains of virus. Sanger sequencing is more specific than the other methods mentioned, can differentiate Type 1 and Type 2 infections, but only elucidates a small portion of the PRRSV2 genome (600/15,000 bp). On the other hand, Nanopore sequencing of long amplicons, such as in the technique described in this project, can be used to multiplex up to 96 samples at a time, while yielding more information in a faster turnaround time, at a competitive price point. All these attributes make Nanopore sequencing a promising diagnostic test for PRRSV2, deserving of further research.

REFERENCES

1. Kappes MA, Faaberg KS. PRRSV structure, replication and recombination: Origin of phenotype and genotype diversity. *Virology* 2015;479-480:475-486.

2. Rossow KD. Review Article Porcine Reproductive and Respiratory Syndrome Virus. *Journal of Veterinary Pathology* 1998;35:1-20.

3. Neumann EJ, Kliebenstein, J. B., Johnson, C. D., Mabry, J. W., Bush, E. J., Seitzinger, A. H., Green, A. L., Zimmermann, J. J. Assessment of the economic impact of porcine reproductive and respiratory syndrome on swine production in the United States. *Journal of the American Veterinary Medical Association* 2005;227:385-392.

4. Terpstra C, Wensvoort G, Pol JM. Experimental reproduction of porcine epidemic abortion and respiratory syndrome (mystery swine disease) by infection with Lelystad virus: Koch's postulates fulfilled. *Vet Q* 1991;13:131-136.

5. Brinton MA, Snijder, E. J. *Encyclopedia of Virology (Third Edition)*. 3 ed: Elsevier, 2008.

6. Melmer DJ, Friendship, R., O'Sullivan, T. L., Greer, A. L., Novosel, D., Ojkic, D., Poljak, Z. Classification of porcine reproductive and respiratory syndrome virus in Ontario using Bayesian phylogenetics and assessment of temporal trends. *The Canadian Journal of Veterinary Research* 2020;85:83-92.

7. Shi M, Lam TT, Hon CC, et al. Molecular epidemiology of PRRSV: a phylogenetic perspective. *Virus Res* 2010;154:7-17.

8. Shi M, Lam TT, Hon CC, et al. Phylogeny-based evolutionary, demographical, and geographical dissection of North American type 2 porcine reproductive and respiratory syndrome viruses. *J Virol* 2010;84:8700-8711.

9. Paploski IAD PN, Makau DN, Rovira A, Corzo CA, Schroeder DC, Cheeran MC-J, Doeschl-Wilson A, Kao RR, Lycett S, VanderWaal K. Phylogenetic Structure and Sequential Dominance of Sub-Lineages of PRRSV Type-2 Lineage 1 in the United States. *Vaccines* 2021;9:608.

10. Paploski IAD CC, Rovira A, Murtaugh MP, Sanhueza JM, Vilalta C, Schroeder DC, VanderWaal K. Temporal Dynamics of Co-circulating Lineages of Porcine Reproductive and Respiratory Syndrome Virus. *Frontiers in Microbiology* 2019;10.

11. Zeller MA SA, Sharma A, Trevisan G, Zhang J, Harmon K, Gauger, PC. ISU PRRSView, 2023.

12. Otake S, Dee, S. A., Rossow, K. D., Deen, J., Soo Joo, H., Molitor, T. W., Pijoan, C. Transmission of porcine reproductive and respiratory syndrome virus by fomites (boots and coveralls. *Journal of Swine Health and Production* 2002;10:59-65.

13. Christopher-Hennings J, Nelson, E. A., Hines, R. J., Nelson, J. K., Swenson, S. L., Zimmerman, J. J., Chase, C. C. L., Yaeger, M. J., Benfield, D. A. Persistence of porcine reproductive and respiratory syndrome virus in serum and semen of adult boars. *Journal of Veterinary Diagnostic Investigation* 1995;7:456-464.

14. Pileri E, Mateu E. Review on the transmission porcine reproductive and respiratory syndrome virus between pigs and farms and impact on vaccination. *Vet Res* 2016;47:108.

15. Li X, Galliher-Beckley A, Wang L, et al. Comparison of Immune Responses in Pigs Infected with Chinese Highly Pathogenic PRRS Virus Strain HV and North American Strain NADC-20. *Open Virol J* 2017;11:73-82.

16. Rahe MC, Murtaugh MP. Effector mechanisms of humoral immunity to porcine reproductive and respiratory syndrome virus. *Vet Immunol Immunopathol* 2017;186:15-18.

17. Lunney JK, Fang Y, Ladinig A, et al. Porcine Reproductive and Respiratory Syndrome Virus (PRRSV): Pathogenesis and Interaction with the Immune System. *Annu Rev Anim Biosci* 2016;4:129-154.
18. Rossow KD, Collins J. E., Goyal, S. M., Nelson, E. A., Christopher-Hennings, J., Benfield, D. A. Pathogenesis of Porcine Reproductive and Respiratory Syndrome Virus Infection in Gnotobiotic Pigs. *Journal of Veterinary Pathology* 1995;32:361-373.
19. Lager KM, Halbur, P. G. Gross and microscopic lesions in porcine fetuses infected with porcine reproductive and respiratory syndrome virus. *Journal of Veterinary Diagnostic Investigation* 1996;8:275-282.
20. Balasuriya UB, Carossino M. Reproductive effects of arteriviruses: equine arteritis virus and porcine reproductive and respiratory syndrome virus infections. *Curr Opin Virol* 2017;27:57-70.
21. Risser J, Ackerman M, Evelsizer R, et al. Porcine reproductive and respiratory syndrome virus genetic variability a management and diagnostic dilemma. *Virol J* 2021;18:206.
22. Nathues H, Alarcon P, Rushton J, et al. Cost of porcine reproductive and respiratory syndrome virus at individual farm level - An economic disease model. *Preventive Veterinary Medicine* 2017;142:16-29.
23. Zhang Z, Li Z, Li H, et al. The economic impact of porcine reproductive and respiratory syndrome outbreak in four Chinese farms: Based on cost and revenue analysis. *Front Vet Sci* 2022;9:1024720.
24. Service USDoAFA. Livestock and Poultry: World Markets and Trade, 2023;1-18.
25. Montaner-Tarbes S, Del Portillo HA, Montoya M, Fraile L. Key Gaps in the Knowledge of the Porcine Respiratory Reproductive Syndrome Virus (PRRSV). *Front Vet Sci* 2019;6:38.

26. Charerntantanakul W. Porcine reproductive and respiratory syndrome virus vaccines: Immunogenicity, efficacy and safety aspects. *World J Virol* 2012;1:23-30.

27. Chongxu Shi YL, Yaozhong Ding, Yongguang Zhang, Jie Zhang. PRRSV receptors and their roles in virus infection. *Arch Microbiol* 2015;197:503-512.

28. Veit M, Matczuk AK, Sinhadri BC, et al. Membrane proteins of arterivirus particles: structure, topology, processing and function. *Virus Res* 2014;194:16-36.

29. Whitworth KM, Rowland RR, Ewen CL, et al. Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus. *Nat Biotechnol* 2016;34:20-22.

30. Hanada K, Suzuki Y, Nakane T, et al. The origin and evolution of porcine reproductive and respiratory syndrome viruses. *Mol Biol Evol* 2005;22:1024-1031.

31. Chen N, Trible BR, Kerrigan MA, et al. ORF5 of porcine reproductive and respiratory syndrome virus (PRRSV) is a target of diversifying selection as infection progresses from acute infection to virus rebound. *Infect Genet Evol* 2016;40:167-175.

32. Christopher-Hennings J, Faaberg, K. S., Murtaugh, M. P., Nelson, E. A., Roof, M. B., Vaughn, E. M., Yoon, K., Zimmerman, J. J. Porcine reproductive and respiratory syndrome (PRRS) diagnostics: Interpretation and limitations. *Journal of Swine Health and Production* 2002;10:213-218.

33. Renukaradhya GJ, Meng XJ, Calvert JG, et al. Inactivated and subunit vaccines against porcine reproductive and respiratory syndrome: Current status and future direction. *Vaccine* 2015;33:3065-3072.

34. Chae C. Commercial PRRS Modified-Live Virus Vaccines. *Vaccines (Basel)* 2021;9.

35. Kang H, Yu JE, Shin JE, et al. Geographic distribution and molecular analysis of porcine reproductive and respiratory syndrome viruses circulating in swine farms in the Republic of Korea between 2013 and 2016. *BMC Vet Res* 2018;14:160.

36. Cui XY, Xia DS, Huang XY, et al. Recombinant characteristics, pathogenicity, and viral shedding of a novel PRRSV variant derived from twice inter-lineage recombination. *Vet Microbiol* 2022;271:109476.

37. IDEXX. IDEXX PRRS Oral Fluids Ab Test. 2023; https://www.idexx.com/en/livestock/livestock-tests/swine-tests/idexx-prrs-oral-fluids-ab-test-na/#:~:text=The%20IDEXX%20PRRS%20Oral%20Fluids,from%20a%20pen%20of%20pigs. Accessed October 25, 2023.

38. Cho HJ, Deregt, D., Joo, H. S. An ELISA for Porcine Reproductive and Respiratory Syndrome- Production of Antigen of High Quality. *Canadian Journal of Veterinary Research* 1996;60:89-93.

39. Sattler T, Wodak, E., Revilla-Fernandez, S., Schmoll, F. Comparison of different commercial ELISAs for detection of antibodies against porcine respiratory and reproductive syndrome virus in serum. *BMC Veterinary Research* 2014;10:2-6.

40. Gerbera PF, Giménez-Lirola, L. G., Halbura, P. G., Zhou, L., Meng, X.J., Opriessnig, T. Comparison of commercial enzyme-linked immunosorbent assays and fluorescent microbead immunoassays for detection of antibodies against porcine reproductive and respiratory syndrome virus in boars. *Journal of Virological Methods* 2013;197:63-66.

41. Cong Y, Huang Z, Sun Y, et al. Development and application of a blocking enzyme-linked immunosorbent assay (ELISA) to differentiate antibodies against live and inactivated porcine reproductive and respiratory syndrome virus. *Virology* 2013;444:310-316.

42. Biernacka K, Podgorska K, Tyszka A, Stadejek T. Comparison of six commercial ELISAs for the detection of antibodies against porcine reproductive and respiratory syndrome virus (PRRSV) in field serum samples. *Res Vet Sci* 2018;121:40-45.

43. Jarcho J. Restriction fragment length polymorphism analysis. *Curr Protoc Hum Genet* 2001;Chapter 2:Unit 2 7.

44. Cha SH, Chang CC, Yoon KJ. Instability of the restriction fragment length polymorphism pattern of open reading frame 5 of porcine reproductive and respiratory syndrome virus during sequential pig-to-pig passages. *J Clin Microbiol* 2004;42:4462-4467.

45. Trevisan G, Sharma, A., Gauger, P., Harmon, K. M., Zhang, J., Main, R., Zeller, M., Linhares, L. C. M., Linhares, D. C. PRRSV2 genetic diversity defined by RFLP patterns in the United States from 2007 to 2019. *Journal of Veterinary Diagnostic Investigation* 2021;33:920-931.

46. University IS. Porcine reproductive and respiratory syndrome virus test fee. 2023; https://vetmed.iastate.edu/vdl/diagnostic-tests/14. Accessed November 4, 2023.

47. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977;74:5463-5467.

48. Martin-Valls GE, Kvisgaard LK, Tello M, et al. Analysis of ORF5 and full-length genome sequences of porcine reproductive and respiratory syndrome virus isolates of genotypes 1 and 2 retrieved worldwide provides evidence that recombination is a common phenomenon and may produce mosaic isolates. *J Virol* 2014;88:3170-3181.

49. Walker J. Frederick Sanger. *Nature*, 2014.

50. Pugh J. The Current State of Nanopore Sequencing. *Methods Mol Biol* 2023;2632:3-14.

51. Hoenen T, Groseth A, Rosenke K, et al. Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. *Emerg Infect Dis* 2016;22:331-334.

52. Oxford Nanopore Technologies. 2023; https://nanoporetech.com/. Accessed October 28, 2023.

53. Metzker ML. Sequencing technologies — the next generation. *Nature Reviews Genetics* 2010;11:31-46.

54. Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. *Nat Biotechnol* 2016;34:518-524.

55. Bull RA, Adikari TN, Ferguson JM, et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat Commun* 2020;11:6272.

56. Castro-Wallace SL, Chiu CY, John KK, et al. Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Sci Rep* 2017;7:18022.

57. King J, Harder T, Beer M, Pohlmann A. Rapid multiplex MinION nanopore sequencing workflow for Influenza A viruses. *BMC Infect Dis* 2020;20:648.

58. Stahl-Rommel S, Jain M, Nguyen HN, et al. Real-Time Culture-Independent Microbial Profiling Onboard the International Space Station Using Nanopore Sequencing. *Genes (Basel)* 2021;12.

59. Hansen S, Faye O, Sanabani SS, et al. Zika Virus Amplification Using Strand Displacement Isothermal Method and Sequencing Using Nanopore Technology. *Methods Mol Biol* 2020;2142:123-136.

60. Baker M. De novo genome assembly: what every biologist should know. *Nature Methods* 2012;9:333-337.

61. Caserta LC, Zhang J, Pineyro P, Diel DG. Rapid genotyping of porcine reproductive and respiratory syndrome virus (PRRSV) using MinION nanopore sequencing. *PLoS One* 2023;18:e0282767.