Viral metagenomics for swine health

by

Brandi Feehan

B.S., Kansas State University, 2014

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Diagnostic Medicine/Pathobiology
College of Veterinary Medicine

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Roman Pogranichniy

Copyright

Abstract

Viral metagenomic sequencing describes an unbiased technique to identify viruses present in a sample. The rise of metagenomics, through next generation sequencing (NGS) and bioinformatics, has aided disease epidemiology, diagnostics, and vaccine development for the swine industry. Metagenomic techniques originated from labor intensive functional screening which NGS overcame by significantly increase the data output. Although, with this increase of data, computational methods were needed to elucidate biological significance. Therefore, bioinformatics made NGS relevant for viral metagenomics targeting swine health. Viral metagenomics for the swine industry involves uncovering molecular characteristics to aid in the development of vaccines, diagnostics, and therapeutics.

We utilized metagenomics to identify a novel *mammalian orthorubulavirus 5* (PIV5) while attempting to isolate porcine rotavirus B (RVB). During cell passaging, the responsible technician fell ill with symptoms of a respiratory infection. On passage four, cells exhibited the cytopathic effect of syncytia formation and were sent for NGS which identified a human PIV5 strain, named Moskva. The resulting genome demonstrated an 87% nucleotide identity to a human PIV5 strain. The original cell culture and pig fecal samples were negative for Moskva by qRT-PCR. Moskva formed a distinct clade from other PIV5 strains in a whole genome based phylogenetic tree. Comparing the two major surface glycoproteins, fusion (F) and hemagglutinin-neuraminidase (HN), amongst human PIV5 strains, Moskva had 32 and 31 amino acid substitutions, respectively. The novel PIV5 strain proved that further research is needed to determine prevalence of PIV5 in different mammalian hosts and geographical locations; in addition, the effects of amino acid substitutions in the viral surface proteins F and HN of the Moskva strain, especially with relevance to human health, need to be determined.

Viral metagenomics was further applied to investigate *Betaarterivirus suid 2* (PRRSV-2), the causative agent of porcine reproductive and respiratory syndrome (PRRS). PRRSV-2 detection and genome assembly through NGS approaches indicated that current PRRSV-2 analytic methods for NGS were insufficient. Nearly half of the PRRSV-2 qRT-PCR positive samples were unsuccessfully assembled using NGS. Read classification detected a novel porcine virus, a lentivirus, which was found especially as a coinfecting agent with PRRSV-2. While PRRSV-2 glycoprotein 5 (GP5) has been used historically for phylogenetic analysis, five of the ten PRRSV-2 genes illustrated increased nucleotide diversity when compared to GP5, indicating that other PRRSV-2 genes could be suitable candidates to demonstrate PRRSV-2 diversity. Comparatively, both the whole genome and GP5 phylogenetic trees lacked clustering patterns relative to temporal or geographical distribution, which also indicated the need for the use of a different PRRSV-2 gene for phylogenetic analysis. The use of a representative gene for PRRSV-2 classification is crucial to more accurately comprehend PRRSV-2 diversity and its implications for PRRSV-2 evolution, distribution, and virulence.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to firstly thank my PI and colleagues for investing in me and my research whom all surpassed my expectations upon entering my master's degree. Dr. Douglas Marthaler instilled his passion for viral metagenomics of the swine industry from the day I set foot in the lab as an undergraduate. Doug spent the time to ensure I not only mastered techniques, but that I also understood the relevance of our work. Throughout my time in the Marthaler Lab, the lab family that formed around me inspired me further and resoundingly encouraged me in my final year as an undergraduate to commit to a master's degree. I am thankful for the opportunities to present my research, which Doug fully encouraged including my first scientific conference, American Society for Virology at the University of Minnesota. My deepest gratitude goes out to all members of the Marthaler Lab including Deepak Kumar, Michael Duff, Qinghong Ran, Matt Heffel, Suman Chaudhary, Mingpu Qi, Leeanna Burton, and Aleesha Koetting. I will forever be thankful for the experiences and memories I gained from working in my first laboratory.

I would also like to extend great appreciation to Dr. Roman Pogranichniy for providing great insights and aid in finishing my master's thesis. To my entire committee, including Dr. Susan Brown and Dr. Jürgen Richt, I appreciate the time you have devoted while providing honest feedback and continued support. You have promoted my presentation abilities, capability to extrapolate results and the power of confidence, among other skills, which I will utilize in my PhD and career.

I am also incredibly thankful for my collaborators and funding sources which allowed me to complete my master's degree. Dr. Ying Fang's laboratory offered me unique experiences and collaborations with Fangfeng Yuan and Xingyu Yan. Dr. Nora Springer was also gracious to allow me to use her lab while Joe Anderson assisted in viral isolation. For my second study of mammalian orthorubulavirus 5 (PIV5), I extend my thanks to our Russian collaborators:

Aleksey A. Penin, Alexey N. Mukhin, Anna S. Moskvina, Kizkhalum M. Khametova, Anton G. Yuzhakov, Maria I. Musienko, Alexey D. Zaberezhny, Taras I. Aliper, and Konstantin P. Alekseev. I am also thankful for the assistance the following offered during my master's degree: Kansas State Veterinary Diagnostic Laboratory (KSVDL), the Russian Academy of Sciences (Projects 0053-2019-0005), Boehringer Ingelheim, KSU professors of my courses, and swine industry partners with the Marthaler Lab.

The last group of individuals I mention are most deserving of their thanks. Throughout my childhood, my parents, Vicki and Tom Feehan, ensured I experienced many sports, traveled frequently, kept to a manageable budget among countless other experiences and lessons. They only asked for a couple of things in return: that I strived to learn through school and always gave my all in my ambitions. Thanks Mom and Dad for everything (more than even the combined pages this thesis could ever contain). Furthermore, my extended family and significant other, Phillip Haslouer, have also shown me all the support and love I needed to pursue and finish my master's degree and now a PhD.

Finally, while they say it takes a village to raise a child, I feel more inclined to say it takes a village full of amazing individuals to support a graduate student, and I am beyond thankful and proud of my village.

Chapter 1 - Introduction

1.1 The Rise of Metagenomic Sequencing

Metagenomic sequencing described an unbiased sequencing approach originating from a need to detect microbes uncapable of being cultured (Thomas et al.). Viral metagenomics aimed to elucidate the genomic sequences of all viruses within a sample including previously unknown, or novel, viruses. Metagenomics began with random shotgun sequencing and the development of Sanger sequencing, a first generation sequencing method. Sanger sequencing initiated with an amplification of the targeted region or genome through PCR using primers to a known genomic region. The process therefore amplified the target sequence. PCR amplification bypassed the need for viral cultures which fulfilled a key aspect, a more unbiased approach, of metagenomic sequencing (Tyson et al.; Venter). During Sanger sequencing, fluorescently-labelled terminating dideoxynucleotides (ddNTPs) were added, and these ddNTPs would halt sequence extension when the base was incorporated. The newly generated sequences, now with differing nucleotide lengths, were run on a gel to distinguish these varying lengths. The fluorescent ddNTP, read according to the position in the gel, indicated the nucleotide sequence of the amplified target. First generation sequencing provided a less biased approach compared to culture methods, although Sanger sequencing was still biased due to the necessity for primer design (Applications; Slatko et al.). The sequence bias demonstrated by first generation sequencing was detrimental to the discovery of novel viruses and detecting multiple pathogens, or coinfections, concurrently.

Second generation sequencing, or next generation sequencing (NGS), provided a further unbiased approach which appealed to metagenomic studies. To provide the unbiased nature of NGS, the NGS technology of Illumina utilized random hexamer primers to amplify all deoxyribonucleic acid (DNA) and complementary DNA (cDNA) during the library preparation process. While the random hexamer primers still introduced bias as a result of amplification

inefficiencies (Hansen et al.), unlike Sanger sequencing, NGS detected the sequences of novel viruses and viral coinfections without requiring uniquely designed primers for each virus (Oka et al.; Farlow et al.; X. Zhang et al.). In Illumina sequencing, DNA or cDNA from each original sequence were clonally amplified, aiding in sequence accuracy, to generate clusters on a flow cell. Every flow cell sequenced numerous clusters as one flow cell averages 1,000,000 clusters per square millimeter (*Optimal Template Loading Concentration and Cluster Densities for Illumina Instruments*). Somewhat similar to Sanger sequencing, when a fluorescence-labelled, reversible terminator deoxyribonucleotide (dNTP) was incorporated into each sequence in a cluster, the entire cluster omitted the same fluorescence signal. After the fluorescence signal had been detected indicating the nucleotide base, the terminator in the dNTP was reversed. This reversal quenched the fluorescence and allowed for another base to be incorporated into the sequence. Illumina sequencing generated paired reads of up to 300 bases per read, compared to a maximum of 1,000 bases per read with Sanger. Both first and second generation sequencing are known as sequencing by synthesis technologies as a result of base detection during nucleotide polymerization (Morozova and Marra; Pettersson et al.). NGS has become a favored technology for metagenomic studies as a result of the increasingly unbiased approach while retaining sequence accuracy.

Third generation sequencing has transitioned from high throughput with short reads to a moderate data output with significantly longer reads, typically more than 10,000 bases, but can surpass 100,000 bases. The genome of most pathogenic swine viruses is less than 20,000 bases which is readily elucidated through third generation sequencing (Philips). Additionally, the length of reads allows third generation sequencing to forego PCR amplification during library preparation (Besser et al.). In third generation sequencing technologies, including nanopore sequencing, a single strand of nucleic acid moved through a protein pore while changing the

ionic current. These ionic changes are read to decipher the nucleotide sequence. Problems arose for nanopore sequencing in nucleotide regions where the same base is repeated. Without a change in ionic current, the exact number of base repetitions were difficult to determine (Shendure et al.; Y. Wang et al.). These nucleotide inaccuracies are detrimental for genome discovery in viral metagenomics. While long reads are beneficial for viral metagenomic applications, the decrease in data output and sequence accuracy hinder the use of third generation sequencing for metagenomics.

Viral metagenomic sequencing was developed through a need to sequence nucleic acids in an unbiased approach to demonstrate comprehensive microbe presence and generate viral genomes. The first generation sequencing approaches utilized PCR to form the foundation for sequencing uncultured viruses for metagenomics. Sanger sequencing lacked the ability to detect novel viruses and viral coinfections efficiently without designing numerous primers. NGS was created where random hexamer primers amplified DNA and cDNA in a relatively unbiased technique. Additionally, Illumina sequencing proved to be high throughput with sustained read accuracy upholding its preferred use for metagenomic studies. While third generation sequencing generates long reads, its read accuracy is of high concern for elucidating quality viral genomes. Viral metagenomics studies often utilize NGS for its high throughput nature and accuracy resulting in an abundance of genetic sequences for virus detection and genome assembly.

1.2 Integrating Bioinformatics with Viral Metagenomics

Metagenomic sequencing applications have been significantly improved through bioinformatics as raw metagenomic sequencing data *per se* has minimal applications to animal/swine health. Metagenomic sequencing generates large amounts of data, ranging from 7-150Gb with Illumina NGS technologies, which requires computational methods to decode the data for biological relevance (*Sequencing Platforms | Compare NGS Platform Applications & Specifications*). Bioinformatic techniques include a wide range of applications, from initial quality and trimming of sequencing reads, to further downstream applications such as genome classification, assembling viral genomes, elucidating genetic changes, and demonstrating evolutionary relationships.

The first step after next generation sequencing is to trim reads to remove barcodes and regions of low quality. Sample barcodes, utilized in next generation sequencing, are added to reads during library preparation. These barcodes are used to sequence multiple samples during one sequencing run through a process known as multiplex sequencing. After sequencing, generated read are demultiplexed to separate reads according to the respective samples (Illumina, Inc., San Diego, CA). Without the ability to separate reads by sample through bioinformatics, multiplex sequencing would be unattainable and sequencing efficiencies would be diminished. Once reads are demultiplexed, barcodes and regions or entire reads of low quality can be removed through quality trimming. Quality scores are assigned during sequencing and trimming removes adapter sequences and regions falling below a quality threshold. A nucleotide would receive a low-quality score as a result of low confidence in the accuracy of the assigned base. For example, in Illumina sequencing, a nucleotide could be assigned a low quality score due to a weak fluorescent signal within a cluster (*Sequencing Quality Scores*). If low quality sequences are not trimmed, inaccurate genomes and genetic variants, including single nucleotide

polymorphisms (SNPs) and insertions or deletions (INDELs), could be detected and associated with virulence/attenuation of swine viruses. Trimming programs are necessary so downstream analyses and classifications only use high-quality sequences free of barcode sequences (National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi, India).

Another bioinformatic technique, *de novo* assembly, combines sequencing reads to form larger genomic sequences known as contigs or scaffolds. Unassembled, short NGS reads, are of little use for viral metagenomics applications since genes and viral genomes are larger than these short reads (*Sequencing Platforms | Compare NGS Platform Applications & Specifications*). *De novo* assembly is important to elucidate complete viral genomes and accurately determine genetic variants (Afolabi et al.). Assembler applications overlap sequencing reads and generate contigs and scaffolds using scoring algorithms that include the length of overlap, number of base mismatches, and depth of read coverage among other factors. Many *de novo* assemblers are available, including A5, IVA and Ray (Darling lab at the University of Technology Sydney, Ultimo, Australia; Pathogen Informatics, Wellcome Sanger Institute, Hinxton, Cambs, UK; Laval University, Quebec City). A5 has demonstrated increased conservation of open reading frames (ORFs), IVA assembles paired-reads of varying read coverage efficiently, while Ray consistently outputs lengthy assemblies. Although Ray also utilizes increased computer power and time (Hunt et al.; Tritt et al.; Khan et al.). Combining assemblers in a bioinformatics pipeline ensures various assembled contigs and scaffolds will be generated representing the spectrum of potential viral genomes.

Downstream genome alignments demonstrate similarity and disparities between viral sequences to elucidate biological relevance. Assembled contigs and scaffolds can be sent through a search in the Basic Local Alignment Search Tool (BLAST) to determine genomic identity to published GenBank sequences (National Center for Biotechnology Information, U.S. National

Library of Medicine, Bethesda, MD, USA). BLAST is necessary to indicate if a contig or scaffold is similar to a known viral sequence. A genome alignment, between the contig or scaffold, and published virus genomes can be generated through MAFFT (Berkeley Software Distribution, University of California, Berkeley, CA). Similar to assemblers, alignment algorithms differ in the match and mismatch scoring matrix, gap open penalty and computing power required. Genetic changes, including SNPs, INDELs, and gene truncations are discovered through alignments (Okumura et al.). Genetic changes impact phenotypic characteristics therefore influencing viral virulence characteristics. For example, specific SNPs and SNP combinations, forming a haplotype, demonstrated increased virulence of Porcine Reproductive and Respiratory Syndrome virus (PRRSV) (Walker et al.). Similarly, a truncated PA-X protein in swine influenza A viruses (IAVs) displayed increased pathogenicity and replication in cultured swine cells (Xu et al.). Metagenomic sequencing and bioinformatic techniques, including alignments, are necessary to study known and novel viruses and their genetic alterations that play a role in swine health.

Genome alignments are input into phylogenetic analyses, which demonstrates evolutionary relationships amongst viruses (Y. Zhang et al.). Numerous applications, including PhyML and MEGA are available to generate phylogenetic trees with bootstrap values, which are statistical supports of the branching points called nodes (French Institute of Bioinformatics, Èvry, France; Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA). Phylogenetic tree algorithms vary in the regression models used to determine evolutionary distance and relationship. Nelson et al. described the evolutionary transmission events of influenza A virus (IAV) between mammalian hosts through phylogenetic analyses. This study determined IAV has evolved within each host independently for at least the last century, with the most recent interhost transmission event occurring more than 100 years ago

(Nelson et al.). In an attempt to classify and evaluate previous classification cutoff values of rotavirus B (RVB) strains, phylogenetic analyses were performed on each RVB gene segment, which supported the cutoff value for four gene segments while refuted the cutoff values for the remaining seven gene segments (Shepherd et al.). Identification of accurate genotypes is important to detect and track recombination and genetic motifs. Furthermore, epidemiological approaches, including discovering spatial and temporal viral distribution, are elucidated through the combination of phylogenetic analyses with demographic information (Ayolabi et al.; W. Wang et al.). An epidemiological investigation of IAV discovered the diversity and dynamics of IAV infections within farrow-to-wean farms, increasing the knowledge on IAV evolution in this environment (Diaz et al.). Phylogenetic analyses are crucial for understanding viral evolution, especially as evolution pertains to viral virulence and represent the foundation of molecular epidemiology.

The swine health industry utilizes metagenomics alongside bioinformatics to create relevant sequencing and analysis approaches. Without bioinformatics, metagenomics would be unfeasible. Metagenomic sequencing outputs large amounts of sequencing data which requires computational analyses to determine the biological relevance of sequences in a time efficient manner. Aiding in the endeavors for biological significance, there are abundant bioinformatic techniques providing various methods and applications for viral metagenomic data targeting swine health.

1.3 Applications of Viral Metagenomic Sequencing in Swine Health

Viral metagenomics has numerous applications especially as the technique applies to swine health. Firstly, epidemiology has improved through the integration of metagenomics with bioinformatics. Similarly, diagnostic medicine of the swine industry benefits from viral metagenomics through the detection of novel pathogens, improved diagnostics for known pathogens and the development of vaccines. Collectively, metagenomic sequencing aids global swine health through innumerable applications of the technology.

Epidemiological applications for swine health have been aided by viral metagenomics. Through an epidemiological study, the spread of PRRSV-2 lineage 3 has been traced across China since 2010. Specifically, the study described two PRRSV-2 strains in question, GZgy17 and SCya18, which underwent independent recombination events with separate lineages, 8 and 1, respectively. By inoculating the two strains independently into 4 week old piglets, recombination was determined as a key component for pathogenicity, i.e. increasing mortality rates (Zhou et al.). Epidemiology allows the geographical and temporal distribution of viruses to be correlated with other characteristics, such as recombination events in this PRRSV-2 study.

Beyond tracing and studying recombinant viruses, viral metagenomics can discover coinfections and associate viral presence with clinical signs. The detection of multiple viruses within the same sample relies upon the massive parallel sequencing offered by NGS. For example, the virome of Spanish pigs exhibiting periweaning failure-to-thrive syndrome (PFTS), contained significantly less porcine parvovirus 6 (PPV6) while demonstrating an increased abundance of ungulate bocaparvovirus 2 (BoPV2), ungulate protoparvovirus 1 (PPV) and porcine circovirus 3 (PCV3) (Franzo et al.). By knowing which viruses are associated with specific clinical signs aids in developing targeting therapies and surveillance strategies for swine

herds. Determining viral and virome presence with specific illness, through epidemiological analyses, offers many applications of viral metagenomics to swine health.

Beyond epidemiology, the identification of novel viruses and viral variants through metagenomics impacts the development and validity of diagnostic assays. Diagnostic tests, such as PCRs/RT-PCRs and ELISAs, are commonplace to detect pathogens and evaluate immune responses, respectively. Although, specificity of these tests relies upon accurate pathogen information including sequence data and virus diversity. Through metagenomics, the genomic changes were elucidated that led to classical swine fever virus (CSFV) variants escaping detection by RT-PCR. The authors developed a new primer set which was able to detect the CSFV subset with the altered sequence in the E(RNS) region (Leifer et al.). ELISAs also rely upon knowledge of the detailed molecular characteristics of the pathogen studied and they need to be validated for their detection characteristics and cross-reactivity (Hause et al.). Without accurate virus genomic sequences and their respective diversity, PCRs/RT-PCRs and ELISAs for specific viruses cannot be developed. Still, viral metagenomics has further applications beyond detecting novel viruses and the development and improvement of diagnostic tests.

Results from metagenomic sequencing can also be utilized for vaccine research in the swine industry. Metagenomic techniques were applied towards the development of a live-attenuated porcine epidemic diarrhea virus (PEDV) vaccine candidate and was critical for the discovery of amino acid changes associated with attenuation of the vaccine candidate. Animal immunogenicity testing demonstrated that the vaccine virus could confer lactogenic immunity from sows to piglets (Won et al.). The development of effective and safe porcine vaccines nowadays depends on accurate metagenomic sequencing. Genetic changes, such as INDELs and SNPs, within both the viral target antigens or vaccine strains themselves, affect the efficacy of vaccines; but metagenomics, in combination with bioinformatics, aids in elucidating genetic

changes that might otherwise go undetected, and aids in the development of efficacious vaccine candidates. The integration of viral metagenomics with bioinformatics is also utilized in developing and evaluating antiviral therapies in the swine industry.

Viral metagenomics provides innumerable applications to swine health with only a few described here. Epidemiology benefits directly as viral spread and distribution can be traced through metagenomic sequencing. Furthermore, the development and testing of diagnostic assays and vaccines relies upon the accuracy that metagenomic sequencing has provided through NGS and bioinformatics. For the multitude of applications it holds, viral metagenomics offers a foundational approach to improve and protect swine health.

Chapter 2 - Accidental Discovery of a Novel *mammalian orthorubulavirus 5*

Published in *Viruses*, August 2019, Volume 11, Issue 9

2.1 Introduction

Mammalian orthorubulavirus 5 (PIV5), formerly named parainfluenza virus 5, resides within the *Paramyoxiviridae* family (Knipe and Howley; "International Committee on Taxonomy of Viruses (ICTV)"). PIV5 is globally distributed and has been associated with respiratory disease in canine, cattle, swine, and lesser panda (Oem et al.; Liu et al.; Lee and Lee; Zhai et al.). While the pathogenesis of PIV5 in humans is unknown, both PIV5 and contagious *mumps orthorubulavirus* belong to the *Orthorubulavirus* genus (Knipe and Howley). Symptoms of mumps rubulavirus in humans include fever, swollen salivary glands and possibly miscarriages or hearing loss ("Mumps - Symptoms and Causes"). We discovered a divergent Russian PIV5 strain (Moskva) in Vero cells exhibiting cytopathic effect (CPE). The cells were suspected to be accidentally contaminated by an ill laboratory technician since the swine clinical samples used for cell culture inoculation and the initial cells were negative for Moskva by RT-PCR. In addition, Moskva contained a small hydrophobic (SH) protein coding sequence (CDS) which is present in human PIV5 strains.

2.2 The Study

In November 2015, an outbreak of enteric disease occurred in three-day old piglets on a farm residing in the Republic of Buryatia, Russia (Alekseev et al.). Next Generation Sequencing (NGS) identified only porcine rotavirus B (RVB) in the fecal samples (apart from host genes or bacteria), and fecal sample supernatants were passaged on Vero cells in hopes of isolating RVB (Alekseev et al.). Concurrently, the scientist passaging the cell cultures exhibited mild respiratory symptoms including tussis and rhinorrhea. CPE, namely syncytia formation, was observed on the fourth passage of the cells. The nucleic acid from the 15^{th} passage was extracted, and NGS was performed to identify the CPE causative pathogen using an Illumina HiSeq2000 (Illumina, San Diego, CA, USA) (Alekseev et al.). NGS generated 9,280,254 paired reads while 8,949,411 reads resulted after trimming (Alekseev et al.). Kraken identified ~85% (7,602,876) paired reads as eukaryotic and ~9% (818,499) as viral. Of viral reads, ~65% (534,325) were classified as *Paramyoxiridae* and ~33% (271,890) were classified as *Retroviridae* (retrovirus) (Knutson et al.). *De novo* assembly of the viral reads generated a PIV5 genome (Moskva, MK593539), which upon BLASTn yielded an 87% nucleotide identity to human PIV5 strain DEN, isolated from the United Kingdom in 1980 (JQ743322). The highest BLASTn hit to a non-PIV5 strains was to human parainfluenza virus 2 strain (68% identity) from Vietnam (MH006623). Primers (forward, CCGGATCACGTGTCCTCAAA; and reverse, ACCAGGAACCGCACTAATGG) were designed to Moskva, and a RT-PCR was developed utilizing RevertAid H Minus First Strand cDNA Synthesis Kit, using recommended manufacturer protocols (Thermo Fisher Scientific, Waltham, MA, USA). The original Vero cell stock and fecal samples were negative for Moskva by RT-PCR. Testing of the cell culture from passage 4 demonstrated PIV5 titers of ~$10^{3.5}$ $TCID_{50}$/ml (Reed and Muench), supporting the hypothesis of human contamination of the cell culture.

Moskva is 15,218 nucleotides in length and contains a 28-nucleotide deletion in the 5' untranslated region compared to other PIV5 strains in GenBank. Moskva displays a genome organization of nucleocapsid (N), V, phosphoprotein (P), matrix (M), fusion (F), SH, hemagglutinin-neuraminidase (HN), and large (L) genes (Knipe and Howley). The M protein is the most abundant capsid protein, and the L protein is crucial for RNA-dependent RNA replication for Paramyxoviruses (Knipe and Howley). Both, the nucleocapsid and L open reading frames (ORFs) exhibited novel truncations of 9 and 11 residues, respectively. These two Moskva gene truncations exemplify previously unknown diversity of PIV5. In addition, Moskva contains a putative SH gene. The presence of the SH gene varies amongst PIV5 strains, and six of seven human strains contain the SH gene. The SH ORF is present in Moskva which again supports the hypothesis of human contamination of the Vero cell culture (Lee and Lee). The SH protein of PIV5 has been suggested to limit apoptosis via tumor necrosis factor α (TNFα) and causes a decrease in CPE *in vitro* (Knipe and Howley; Lin et al.). The identification and proposed function of the SH protein contradicts our observed CPE using Moskva in susceptible cell cultures.

Pairwise nucleotide and amino acid MAFFA alignments were generated in Geneious version 11.5.1 to calculate genomic identities (Geneious, Newark, NJ, USA). PIV5 nucleotide percent identities of the complete genome increased with the addition of Moskva from 95.5-100% to 85.4-100% (Table 2.1). Using the 80% whole genome nucleotide identity classification set by the International Committee on Taxonomy of Virus, Moskva belongs to the *Rubulavirus* genus ("International Committee on Taxonomy of Viruses (ICTV)"). Moskva showed an 84.9-88.8% nucleotide identity to the other PIV5 GenBank genomes, regardless of the gene or host species (Table 2.1), highlighting the uniqueness of Moskva within the rubulaviruses. The Moskva RNA-dependent RNA polymerase (L protein) had a 96.9% amino acid identity to the L

proteins of other PIV5 strains, confirming Moskva as a member of the PIV5 species ("International Committee on Taxonomy of Viruses (ICTV)"). The uniqueness and divergence of Moskva was reflected as a distinct lineage in a whole genome phylogenetic tree (Figure 2.1). The 500 bootstrapped Maximum Likelihood tree was created in Geneious and edited in FigTree (Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, Scotland). Host and geographic clades were absent in the tree, which was reflected in all individual gene phylogenetic trees (Supplemental Figure 5.1). The divergence in genomic identities and phylogenetic trees suggests undiscovered divergent PIV5 strains are circulating in geographically distinct regions.

The PIV5 genome encodes two glycoproteins, F and HN, which are essential for viral attachment, fusion, and entry. With Moskva exhibiting human origin, the F and HN amino acid sequences of Moskva were compared to the human PIV5 strains. The protein models illustrated the location and magnitude of substitutions using SWISS MODEL (Swiss Institute of Bioinformatics, Lausanne, Switzerland; F, 4jf7.1; HN, 4wsg.1), Phyre2 (Imperial College London, London, UK; F, c2b9bA; HN, c4jf7B) and Chimera (Resource for Biocomputing, Visualization and Informatics, University of California, San Francisco, CA, USA). Consistent with Moskva's evolutionary divergence, the F and HN proteins of Moskva had more residue substitutions (32 and 31, respectively) relative to a consensus sequence of the human PIV5 strains (Figure 2.2). Of the 32 Moskva F protein substitutions, the majority of the substitutions (n=25) occurred in unclassified regions of PIV while 5 and 2 substitutions occurred in domains I and III, respectively (Bose, Heath, et al.; Poor et al.). Domain II (including the hydrophobic loop), the fusion peptide, and heptad repeat region B (HRB) lacked residue substitutions in Moskva (Bose, Heath, et al.; Waning et al.). The hydrophobic loop plays a crucial role interacting with the HN protein for fusion activation, and the region is conserved in all the strains

of *Paramyxoviruses* (Bose, Heath, et al.; Waning et al.). In the HN protein, the stalk was more conserved than the head, with 23 residue substitutions in the head versus only two substitutions in the middle of the stalk (Figure 2.3). The stalk of HN interacts with the F protein for fusion, and residue substitutions in the stalk near the head of the protein are less detrimental than those substitutions further from the head of the protein (Bose, Welch, et al.). The described changes in the F and HN glycoproteins, further confirm the divergence of Moskva from other human PIV5 viruses while supporting regions of conservation.

2.3 Conclusion

A novel PIV5 strain (Moskva) was reported, and Moskva is hypothesized to have resulted from accidental cross-contamination of Vero cells by a human experiencing respiratory illness. While PIV5 is a global pathogen, our genetic and phylogenetic analyses indicated greater diversity of PIV5 strains than previously reported. The branching of Moskva suggests distinct PIV5 strains are likely circulating globally undetected. Moreover, Moskva had a noteworthy number of amino acid substitutions in its surface glycoproteins, as well as shorted M and L genes, indicating considerably greater PIV5 diversity than was previously known. Further research is required to understand the implications of PIV5 global presence, the role of the SH protein in human PIV5 strains, the functional relevance of the amino acid substitutions of its surface proteins, and the M and F protein truncations, especially as these topics pertain to human health.

2.4 Figures and Table

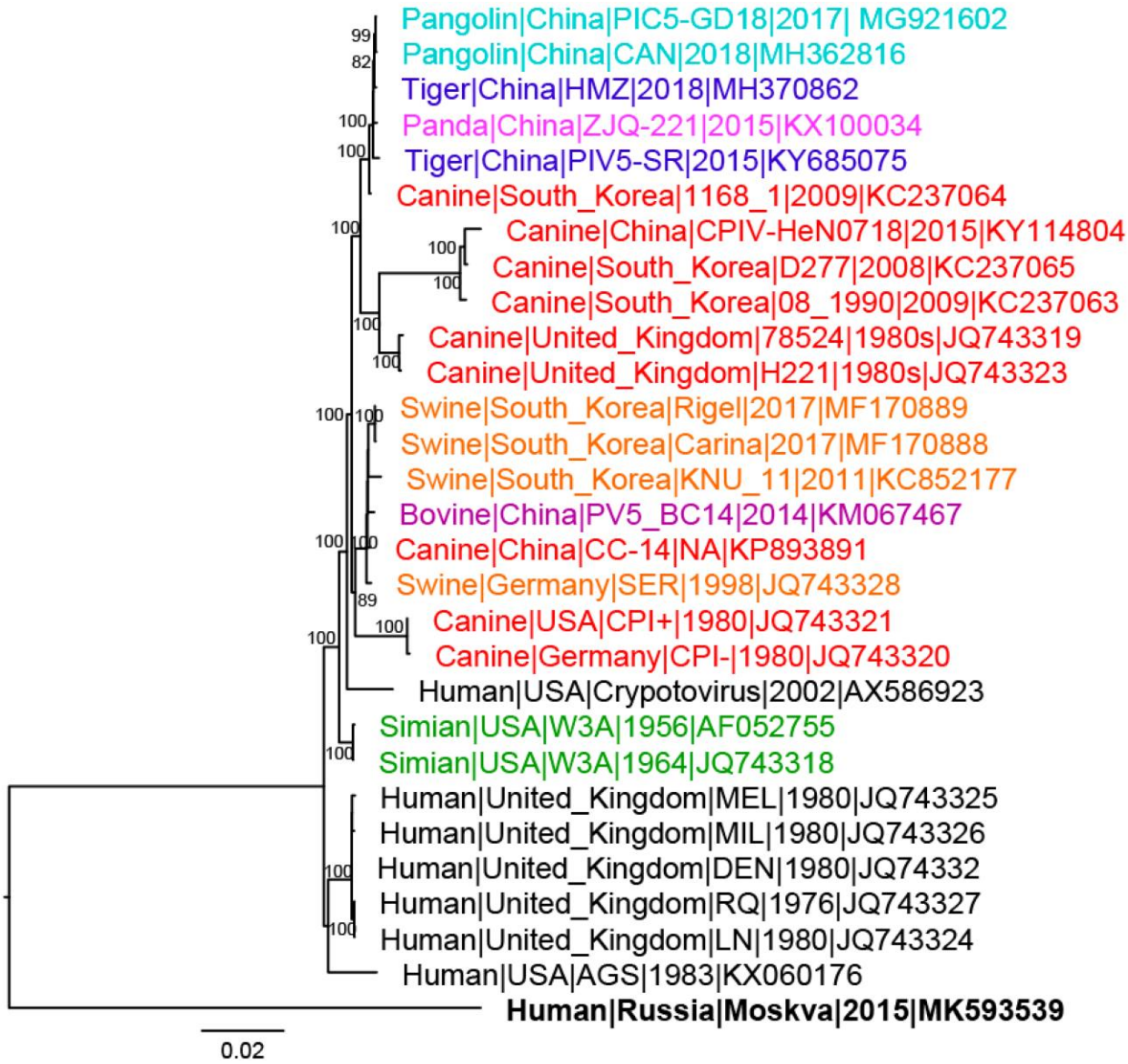


Figure 2.1 Whole genome phylogenetic tree of PIV5 strains. Bootstrap values greater than 70 are shown at major nodes. Strains are colored according to host with black representing human strains; red, canine; orange, swine; green, simian; teal, pangolin; blue, tiger; pink, panda; and purple, bovine. Moskva is in bold.

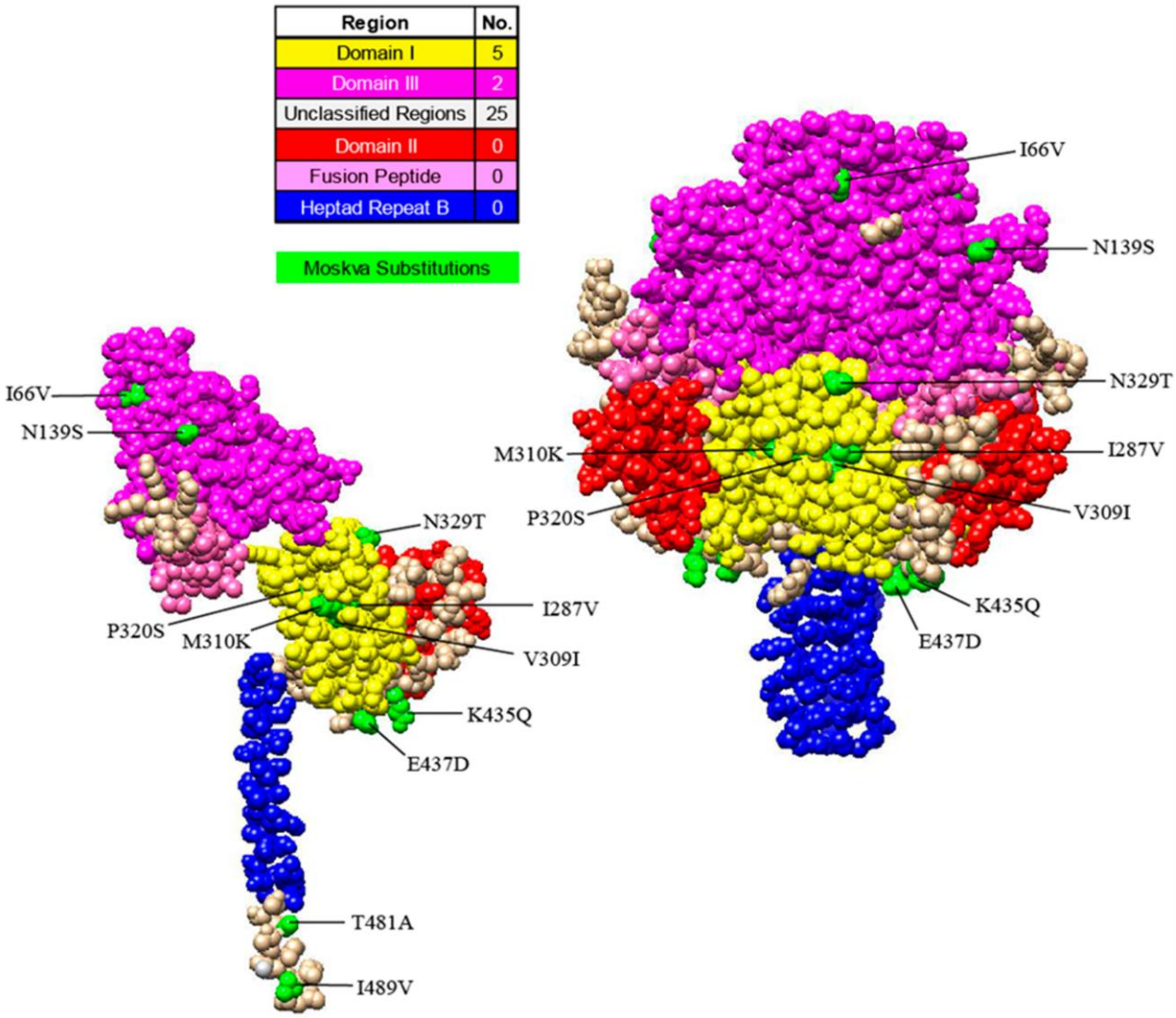


Figure 2.2 Uncleaved Fusion (F) monomer and trimer protein models illustrating the residue substitutions in Moskva. Regions and substitutions are colored according to the legend.

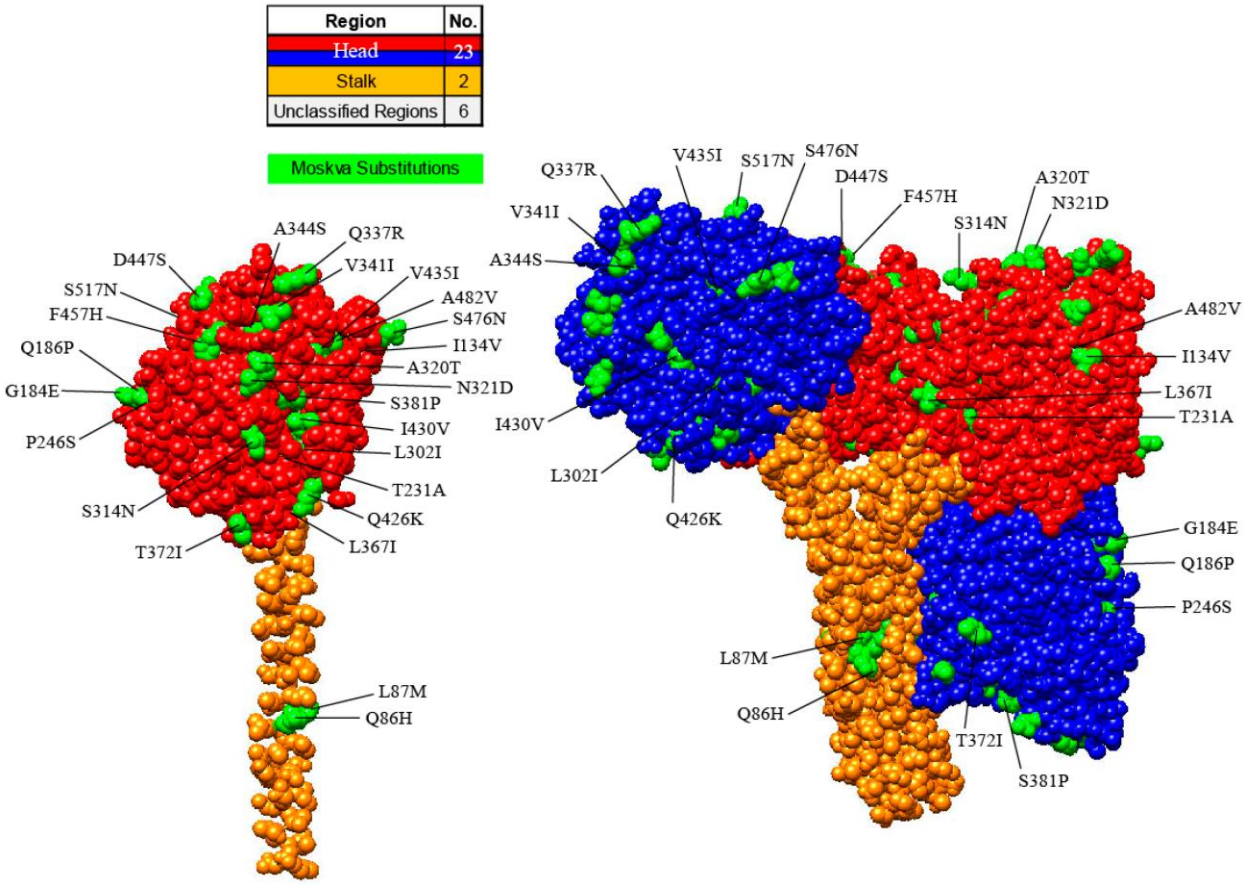


Figure 2.3 Hemagglutinin-neuraminidase (HN) monomer and tetramer protein models illustrating the residue substitutions in Moskva. Regions and substitutions are colored according to the legend.

	Length		Excluding Moskva n=26		Including Moskva n=27		Human n=7		Simian n=2		Swine n=4		Bovine, panda, pangolin, tiger n=6	
Gene/Genome	Nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa
Nucleocapsid (N)	1,530	509	95.8-100	97.7-100	86.3-100	93.5-100	86.8-87.5	**94.1-95.1**	**87.8**	94.5	87.3-87.7	93.9-94.7	87.5-87.7	93.7-94.5
V (V)	669	222	95.5-100	94.6-100	87.3-100	94.6-100	**87.7-88.8**	94.6-96.9	**88.8**	**97.3**	87.7-88.3	95.5-96.9	87.4-88.2	95.5-96.9
Phosphoprotein (P)	1,177	391	95.8-100	95.9-100	87.1-100	94.1-100	**88.1-88.6**	94.4-95.9	**88.6**	**96.2**	87.8-88.3	94.6-95.7	87.9-88.2	94.9-95.7
Matrix (M)	1,134*	377*	95.3-100	96.0-100	86.2-100	93.7-100	**87.0-87.7**	**95.0-96.6**	87.4-87.5	95.5-95.8	87.1-87.4	95.5-95.8	86.6-87.4	94.4-95.8
Fusion (F)	1,656†	551†	94.8-100	95.9-100	83.3-100	91.5-100	**84.9-85.8**	**93.3-94.6**	85.2	93.7-93.8	85.0-85.1	93.3-93.5	84.9-85.1	93.5-93.7
Hemagglutinin-neuraminidase (HN)	1,698	565	95.6-100	95.1-100	84.8-100	91.3-100	85.5-85.8	**92.8-94.5**	85.6-85.7	92.6	85.6-85.8	91.9-92.6	84.9-85.8	92.2-92.8
Large (L)	6,678‡	2,256‡	96.9-100	98.1-100	87.4-100	96.9-100	**87.9-88.2**	**97.3-97.8**	**88.2**	97.5	87.9-88.0	96.9-97.3	**87.8-88.2**	97.2-97.6
Entire genome	15,246§		95.5-100		85.4-100		**86.2-86.7**		**86.6-86.7**		86.2-86.4		86.4-86.5	

*Moskva: 1,107nt and 368aa.

†Simian strains: 1,589nt and 529aa; Human MEL strain: 1,605nt and 534aa.

‡Moskva: 6,735nt and 2,245aa.

§Moskva: 15,218nt.

¶Boldface indicates the highest percent identity to Moskva in the corresponding gene/genome.

Table 2.1 Comparison of the nucleotide (nt) and amino acid (aa) lengths of various PIV5 ORFs and homology with PIV5 strains from humans and other mammals.

Chapter 3 - Next Generation Sequencing (NGS) Targeting Coinfections and Phylogenetic Analysis of *Betaarterivirus suid* 2 (PRRSV-2) in an Integrated US Swine Producer

3.1 Introduction

Porcine reproductive and respiratory syndrome (PRRS) is an endemic disease of swine that has been estimated to cost the US swine industry $302 million annually (Holtkamp et al.). The causative agent of PPRS, discovered in 1991, is a positive stranded RNA virus, *Betaarterivirus suid* (PRRSV) ("International Committee on Taxonomy of Viruses (ICTV)"). The virus belongs to the *Arteriviridae f*amily within the Nidovirales order. *Betaarterivirus suid 1* (PRRSV-1) comprises European-type strains while *Betaarterivirus suid 2* (PPRSV-2) contains North American type strains (Meng et al.). Although the presentation of clinical disease does not greatly differ between the two genotypes, PPRSV-2 has been associated with an increased severity of respiratory disease, including pneumonia and dyspnea. Additionally, both genotypes produce comparable reproductive disease such as late term abortions, stillbirths, and decreased piglet birth weights (Ladinig et al.; Loula; Rossow et al.; Martínez-Lobo et al.). PPRSV-2 causes continual losses to the US swine industry. As a result, continued PRRSV-2 research is critical to uncover virus characteristics related to the pathogenicity, diversity, and evolution of the pathogen, to lessen the consequences of PRRSV-2 infections in US swine.

The detection of PRRSV-2 within viral coinfections is limited so far. Detection of viral coinfections in PRRSV-2 infected swine could reveal mechanisms potentially responsible for the observed severity and length of PRRSV-2 clinical signs. Furthermore, historical and current PRRSV-2 classification and comparison have mainly utilized PRRSV-2 glycoprotein (GP) 5 (GP5). Reliance solely on GP5 for classification and distinction amongst PRRSV-2 isolates

could miss genetic variation/s in other genomic regions which could be of importance to PRRSV-2 disease presentation. GP5 classification of PRRSV-2 strains is done by restriction fragment length polymorphism (RFLP) or by sequencing and subsequent phylogenetic analyses. Sequencing relies on the complete GP5 sequence while RFLP determines the genomic differences due to insertion/s or deletion/s of nucleotides at limited positions within GP5; therefore, RFLP cannot distinguish between all PRRV-2 strains (Wesley et al.). The present study aimed to (i) determine the entirePRRSV-2 genome through next generation sequencing (NGS), (ii) identify PRRSV-2 coinfections, (iii) demonstrate genomic diversity outside GP5, and (iv) classify PRRSV-2 strains phylogenetically from a multi-US state, integrated US swine producer.

3.2 The Study

In order to compare the diversity of PRRSV-2 strains circulating in a large, closed swine herd from 2012-2019, serum samples from farms in Oklahoma, Kansas and Texas were collected when miscarriage rates rose above 2% in sows beyond 70 days of gestation. Serum samples were collected and underwent processing, extraction, and qRT-PCR. Samples proceeded in our analysis pipeline if qRT-PCR indicated PRRSV-2 presence through a Ct value at or below 26 (n=91), indicative of a relatively high PRRSV-2-specific RNA presence (ThermoFischer Scientific, Waltham, MA, USA). Nuclease (DNase and RNase) inhibitor treatment, library preparation and metagenomic sequencing were conducted according to previous studies with the exception of utilizing an Illumina NextSeq sequencing machine (Bentley et al.; Tritt et al.).

Resulting sequencing reads underwent trimming and Kraken classification (Knutson et al.) which resulted in 51 of the 91 (56%) qRT-PCR PRRSV-2 positive samples being also positive the presence of PRRSV-2 genomic sequences (Table 3.1). This indicated a discrepancy in the detection of PRRSV-2 through qRT-PCR versus NGS, even with a PRRSV-2 Ct value threshold of 26. Viral coinfections, amongst the PRRSV-2 qRT-PCR positive samples, occurred most often with porcine endogenous retrovirus (n=42; 46%), lentivirus (n=34; 37%), and Torque teno sus virus (n=31; 34%). The identified porcine lentivirus is a novel virus as indicated by the absence of porcine lentivirus sequences in GenBank. As Kraken classified sequencing reads, similar to Equine Infectious Anemia Virus (EIAV), future research should assemble the lentivirus genome and determine its classification within the lentivirus genus and its similarity to EIAV. With the relatively high number of lentivirus detections by Kraken, this porcine lentivirus should be investigated as playing a role in the PRRSV-associated disease complex in future studies. Porcine endogenous retrovirus and Torque teno sus virus have been reported as ubiquitous porcine pathogens with unsubstantiated disease association, and these results support

the findings of pathogens unrelated to the investigated disease being regularly present in in swine samples (Pogranichniy et al.; Plotzki et al.; Qin et al.).

The cumulative number of distinct viruses identified as coinfections within one sample were also quantified through Kraken classification (Figure 3.1). Most samples positive for PRRSV-2 via qRT-PCR contained three viruses (n=30; 33%), followed by two viruses (n=25; 27%); ten samples (11%) contained PRSSV-2 only. Of samples positive for PRRSV-2 by NGS and Kraken, nine samples (18%) were positive for PRRSV-2 only, and most of these samples had two viruses present (n=19; 37%). Previous research has indicated PRRSV-2 being prevalent as a coinfecting agent (Saade et al.). These findings support the notion that PRRSV-2 is potentially acting in synergy with other co-infecting viruses, as 89% of all PRRSV-2 qRT-PCR positive samples, and 82% of those with PRRSV-2 genomic sequences identified with NGS and Kraken, revealed a viral coinfection. Further research is needed to elucidate the role of PRRSV-2 coinfections in PRRS pathogenesis, especially with the porcine lentivirus described, and the consequences of this pertaining to disease severity, secondary infections and PRRSV epidemiology.

Viral and unknown Kraken classified reads were input through *de novo* assembly and sent through a BLASTN search as previously described (Knutson et al.). Of the initial 91 qRT-PCR PRRSV-2 positive samples, merely 30 (33%) were identified with complete PRRSV-2 genomes consisting of ten genes (ORF1a, ORF1b, GP2, envelope (E), GP3, GP4, GP5, ORF5a, matrix (M) and nucleocapsid (N)). Therefore, even with the qRT-PCR threshold, only 30 samples had successful PRSSV-2 genome assemblies, although 51 samples contained Kraken-classified short reads. This further corroborated the need for enhanced sample screening methods to improve the rate of successfully assembling a PRRSV genome. An additional 33 complete PRRSV-2 genomes were assembled from viral isolates, originating from symptomatic swine,

which underwent the same DNA extraction, library preparation, sequencing and bioinformatic assembly. Finally, eight vaccine genomes, from vaccines utilized in the farm system, were included in downstream analysis, bringing the total to 71 PRRSV-2 genomes. Pairwise MAFFT alignments were created in Geneious version 11.1.5 which illustrated whole genomic identities ranging from 81-99% (Table 3.2; Geneious, Newark, NJ, USA). ORF5 (coding for GP5), used historically for PRRSV-2 classification, consisted of identities ranging from 82-100%. Of the ten PRRSV-2 genes, five (ORF2a, ORF5a, ORF3, ORF1ab and ORF1a) demonstrated increased diversity relative to GP5. The most divergent gene, ORF1a, consisted of 7% greater diversity (74-99%) than GP5 while the most conserved gene, ORF6 (86-100%), illustrated 4% less diversity relative to GP5. The majority of PRRSV-2 genes demonstrated increased diversity relative to ORF5. Diversity of genes outside of ORF5 seem to reflect PRRSV-2 evolution and could influence virulence; further studies should include analyses of additional PRRSV-2 genes, especially ORF1a, to determine PRRSV-2 evolution and diversity.

Phylogenetic analysis was utilized through bootstrapped, 500 replicates, maximum likelihood nucleotide trees of the 71 PRRSV-2 genomes in Geneious and edited through FigTree (Figure 3.2; Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, Scotland). The complete genome nucleotide tree identified an outgroup with three vaccine genomes (MLV, Fostera and ATP) alongside one PRRSV-2 genome originating from swine samples obtained in Colorado during 2018 (2018.01_CO_PF_3D). Genomes did not group according to year or geographical location within the US. Similarly, in the GP5 tree, the 2018.01_CO_PF_3D virus grouped with the same three vaccines (Figure 3.3). A second clade was formed representing four distinct GP5 sequences (2015.03_KS_VI_2Y, 2018.05_OK_VI_2Y_A, 2019.05_TX_S_2Y, and 2019.03_TX_S_2Y). Similar as observed for the whole genome phylogenetic tree, PRRSV-2 GP5 sequences did not group according to year or geographical location within the US. Even

though the vaccine outgroup was the same, the location of individual PRSSV-2 genomes on the two phylogenetic trees either based on the whole PRRSV-2 genome or on GP5 only was different. Similar to the genomic identities of the whole genome and GP5 varying (81-99% and 85-99%, respectively), phylogenetic analyses demonstrated a lack of distinction between the whole genome and GP5 phylogenetic trees. Neither the complete genome nor GP5 based phylogenetic trees depicted strains clustering according to year or geographical location which might indicate that the inclusion of a different PRRSV-2 gene could help to depict these demographics better for evolutionary analyses.

3.3 Conclusions

PRRSV-2 has drastic impacts, both on swine health status and economic return, for the US swine industry and the swine industry worldwide. In order to understand the use of NGS for detecting PRRSV-2 and viral coinfections, in addition to evaluating genetic diversity within and outside of GP5, samples from a large, integrated swine producer were collected, and PRSSV-2 positive samples were sequenced. Although 91 samples were positive for PRRSV-2 by qRT-PCR, only 51 (56%) were positive by Kraken read classification and only 30 (33%) of these resulted in complete PSSRV-2 assembled genomes. The efficiency of NGS to detect PRRSV-2 in swine samples and to yield complete PRSSV-2 genome did not correlate well with the previously set qRT-PCR cutoff of 26, indicating different screening methods and sequencing approaches could increase NGS efficiency. In this study, samples were qRT-PCR tested prior to nuclease inhibitor treatment. Testing the samples by qRT-PCR after nuclease inhibitor treatment is one alternative approach to screening PRRSV-2 samples prior to NGS. Decreasing the Ct cutoff value is another option, although some samples capable of assembling complete PRRSV-2 genomes could be dismissed. The three prevalent viruses in coinfection supported the ubiquitous presence of porcine endogenous retrovirus and Torque teno sus virus while identifying a novel porcine lentivirus via Kraken short read-classification. The implications of a porcine lentivirus, especially in a coinfection with PRRSV-2, are subjects for future research. PRRSV-2 genome analyses demonstrated that the majority of PRRSV-2 genes are more diverse than GP5. Furthermore, phylogenetic analysis of the whole PRRSV-2 genome and GP5 illustrated a lack of coherence between the two phylogenetic trees. Future studies are needed to investigate the diversity found in PRRSV-2 genomes outside of GP5 and uncover the implications of genomic diversity relative to temporal and geographical distribution, evolution, and virulence. Altogether, the use of NGS allowed a comprehensive understanding of the virome present in swine samples

derived from symptomatic animals, including PRRSV, while also providing the opportunity to assemble the complete genome of PRRSV and other viruses. NGS can provide insights into the PRRSV genome and was able to identify novel genetic diversity within its genome which enhances the swine industries understanding of the viral variant. PRRSV epigenetics especially benefits from NGS as scientists can work to understand disease severity and geographic distribution of PRRSV variants. With this knowledge, scientists and swine farmers will have an enhanced ability to monitor PRRSV distribution and target PSSRV variants of concern. Moreover, our results indicate that PRRSV is typically associated with other viruses. Viral co-infections may result in different disease outcomes; therefore, this aspect of PRRSV pathogenesis should be investigated in future research.

3.4 Figures and Tables

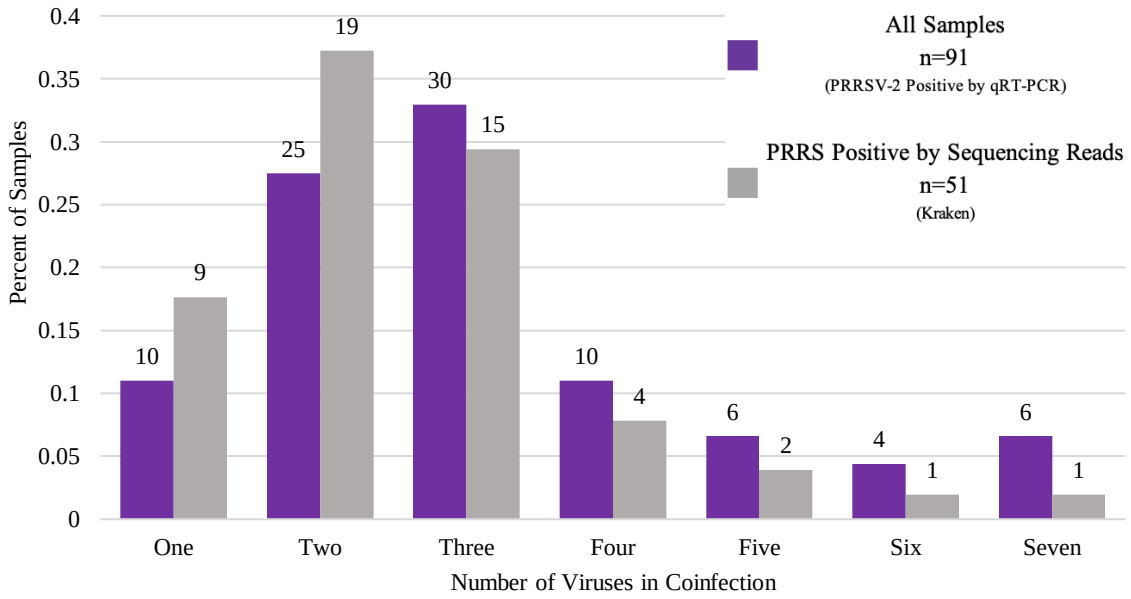
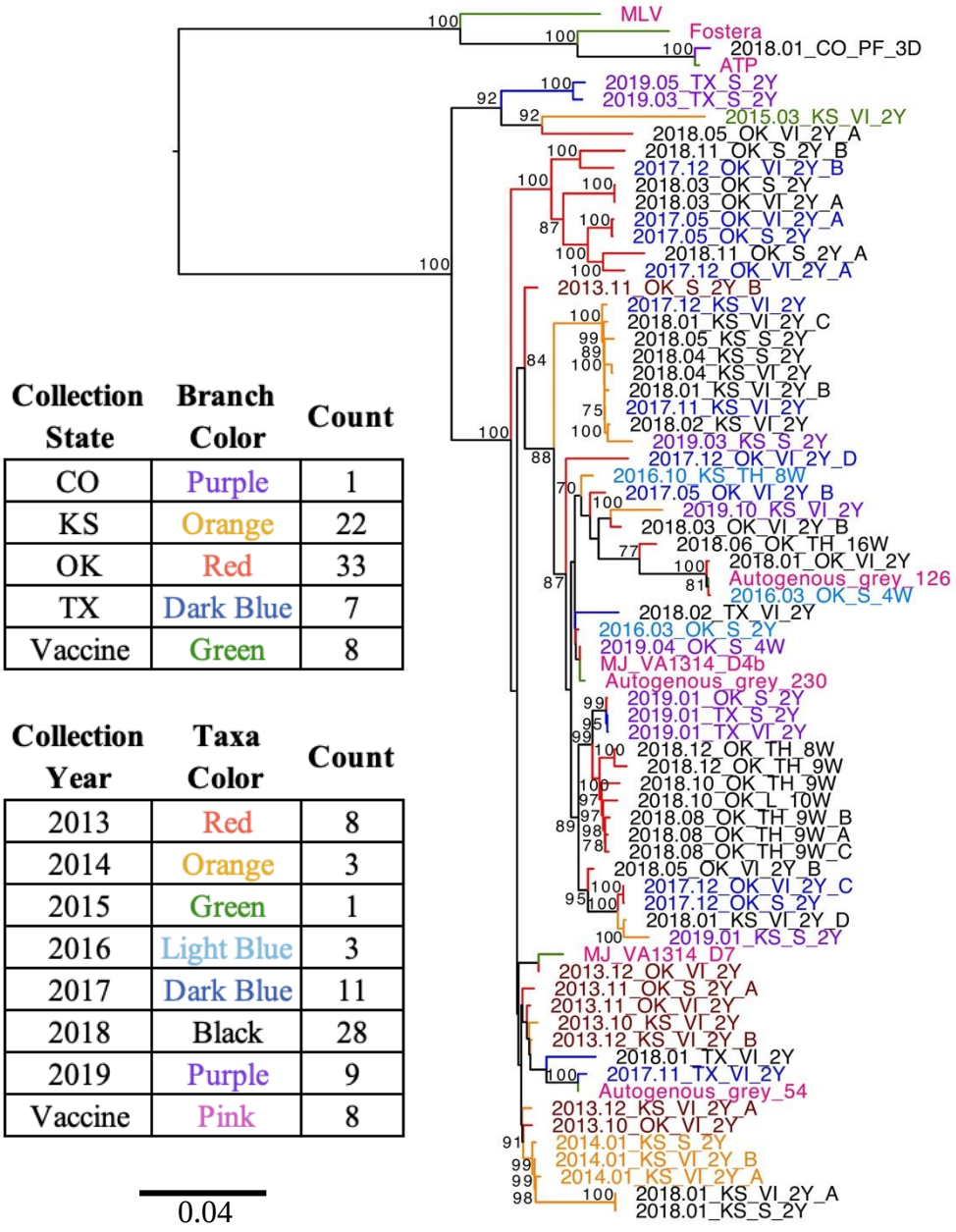


Figure 3.1 Number of Kraken identified viruses present within a sample. The number of samples (n) is indicated at the top of each bar.

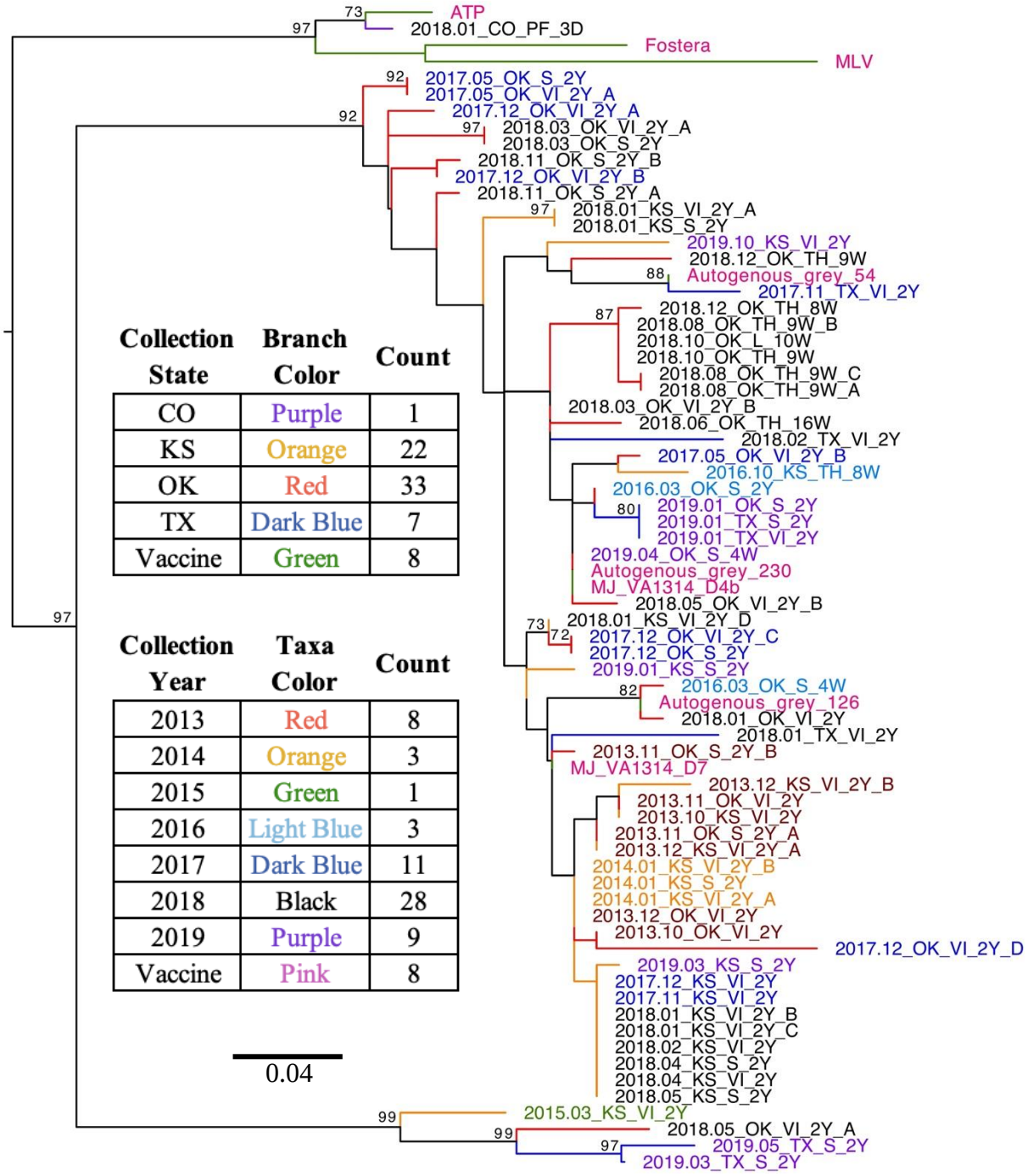


Collection State	Branch Color	Count
CO	Purple	1
KS	Orange	22
OK	Red	33
TX	Dark Blue	7
Vaccine	Green	8

Collection Year	Taxa Color	Count
2013	Red	8
2014	Orange	3
2015	Green	1
2016	Light Blue	3
2017	Dark Blue	11
2018	Black	28
2019	Purple	9
Vaccine	Pink	8

Figure 3.2 Whole genome phylogenetic tree of PRRSV-2 strains. Bootstrap values greater than 70 are shown at nodes. PRRSV-2 nomenclature is as follows:

CollectionYear.CollectionMonth_CollectionState_SampleType (L=Lung; S=Serum, PF=Processing Fluid; TH=Tissue Homogenate; VI=Viral Isolate)_AgeOfAnimal(D=Days; W=Weeks; Y=Years)_Letter (this letter distinguishes samples with the same demographic information).



Collection State	Branch Color	Count
CO	Purple	1
KS	Orange	22
OK	Red	33
TX	Dark Blue	7
Vaccine	Green	8

Collection Year	Taxa Color	Count
2013	Red	8
2014	Orange	3
2015	Green	1
2016	Light Blue	3
2017	Dark Blue	11
2018	Black	28
2019	Purple	9
Vaccine	Pink	8

Figure 3.3 Glycoprotein 5 (GP5) phylogenetic tree of PRRSV-2 strains. Bootstrap values greater than 70 are shown at nodes. PRRSV-2 nomenclature is as follows: CollectionYear. CollectionMonth_CollectionState_SampleType (L=Lung; S=Serum, PF=Processing Fluid; TH=Tissue Homogenate; VI=Viral Isolate)_AgeOfAnimal (D=Days; W=Weeks; Y=Years)_Letter (this letter distinguishes samples with the same demographic information).

Virus Identified by Sequencing Reads	Overall		PRRS Identified by Sequencing Reads (Kraken)			
			Yes		No	
	n=91		n=51		n=40	
	Count	Percent	Count	Percent	Count	Percent
PRRS	51	56%	51	100%	0	0%
Porcine Endogenous Retrovirus	42	46%	29	57%	13	33%
Lentivirus*	34	37%	14	27%	20	50%
Torque Teno Sus Virus	31	34%	18	35%	13	33%
Bacteriophage	16	18%	3	6%	13	33%
Circovirus	13	14%	2	4%	11	28%
Parvovirus	13	14%	7	14%	6	15%
Porcine Kobuvirus	11	12%	2	4%	9	23%
Adenovirus	11	12%	2	4%	9	23%
Mimivirus	8	9%	1	2%	7	18%
IAV	7	8%	1	2%	6	15%
Porcine Pegivirus	7	8%	2	4%	5	13%
Posavirus	2	2%	1	2%	1	3%
Hypovirus	2	2%	0	0%	2	5%
Coronavirus	1	1%	0	0%	1	3%

*Classified as Equine Infectious Anemia Virus

Table 3.1 Total number and relative abundance of Kraken classified viruses present in samples.

Genome/Gene Protein(s)	Nucleotide		Amino Acid	
	Length	Percent Identity	Length	Percent Identity
Entire Genome	14,947[a]	80-99		
ORF1a Nonstructural proteins	7,098[b]	74-99	2,365[b]	75-100
ORF1ab Nonstructural proteins	11,468[c]	78-99		
ORF3 GP3	765	79-100	254	75-100
ORF5a GP5a	141[d]	82-100	46[d]	76-100
ORF2a GP2a	771	82-100	256	78-100
ORF5 GP5	603	82-100	200	82-100
ORF2b 2b; E	222	83-100	73	83-100
ORF4 GP4	537	83-100	178	85-100
ORF1b Nonstructural proteins	4,377	85-100	1,458	94-100
ORF7 N	372	86-100	123	87-100
ORF6 M	525[e]	86-100	174[e]	90-100

[a]Ranging from 14,846nt to 15,412nt

[b]7,098nt and 2,365aa for 36 strains; 7,107bt and 2,368aa for 26 strains; 7,512nt and 2503aa for 1 strain and 2 vaccines; 7,494nt and 2,497aa for Fostera vaccine

[c]11,468nt for 36 strains; 11,477nt for 26 stains; 11,882 for 2018.01_CO.J_PF_3D_NA_2vx and 2 vaccine strains; 11,864 for Fostera

[d]141nt and 46aa for 46 strains and Fostera vaccine; 156nt and 51aa for 5 strains and 2 vaccine strains; 138nt and 45aa for

2019.01_KS.R_S_2Y_AT_2vx; 5 vaccine strains do not code for ORF5a (10nt short at 5' end)

[e]528nt and 175aa for 2017.12_OK.F_VI_2Y_FS_1

Table 3.2 Homology analyses of PRRSV-2 genomes assembled in this project.

Chapter 4 - References

Afolabi, Kayode Olayinka, et al. “Increasing Diversity of Swine Parvoviruses and Their Epidemiology in African Pigs.” *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, vol. 73, Apr. 2019, pp. 175–83. *PubMed*, doi:10.1016/j.meegid.2019.04.029.

Alekseev, Konstantin P., et al. “Genome Characterization of a Pathogenic Porcine Rotavirus B Strain Identified in Buryat Republic, Russia in 2015.” *Pathogens*, vol. 7, no. 2, 2018, p. 46. *www.mdpi.com*, doi:10.3390/pathogens7020046.

Applications, National Research Council (US) Committee on Metagenomics: Challenges and Functional. *Why Metagenomics?* National Academies Press (US), 2007. *www.ncbi.nlm.nih.gov*, https://www.ncbi.nlm.nih.gov/books/NBK54011/.

Ayolabi, Christianah Idowu, et al. “Detection of Dengue Viruses among Febrile Patients in Lagos, Nigeria and Phylogenetics of Circulating Dengue Serotypes in Africa.” *Infection, Genetics and Evolution*, vol. 75, Nov. 2019, p. 103947. *ScienceDirect*, doi:10.1016/j.meegid.2019.103947.

Bentley, David R., et al. “Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry.” *Nature*, vol. 456, no. 7218, Nov. 2008, pp. 53–59. *DOI.org (Crossref)*, doi:10.1038/nature07517.

Besser, John, et al. “Next-Generation Sequencing Technologies and Their Application to the Study and Control of Bacterial Infections.” *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, vol. 24, no. 4, Apr. 2018, pp. 335–41. *PubMed Central*, doi:10.1016/j.cmi.2017.10.013.

Bose, S., C. M. Heath, et al. “Mutations in the Parainfluenza Virus 5 Fusion Protein Reveal Domains Important for Fusion Triggering and Metastability.” *Journal of Virology*, vol. 87, no. 24, Dec. 2013, pp. 13520–31. *Crossref*, doi:10.1128/JVI.02123-13.

Bose, S., B. D. Welch, et al. “Structure and Mutagenesis of the Parainfluenza Virus 5 Hemagglutinin-Neuraminidase Stalk Domain Reveals a Four-Helix Bundle and the Role of the Stalk in Fusion Promotion.” *Journal of Virology*, vol. 85, no. 24, Dec. 2011, pp. 12855–66. *Crossref*, doi:10.1128/JVI.06350-11.

Diaz, Andres, et al. “Complete Genome Sequencing of Influenza A Viruses within Swine Farrow-to-Wean Farms Reveals the Emergence, Persistence, and Subsidence of Diverse Viral Genotypes.” *Journal of Virology*, vol. 91, no. 18, 15 2017. *PubMed*, doi:10.1128/JVI.00745-17.

Farlow, Jason, et al. “Intra-Epidemic Genome Variation in Highly Pathogenic African Swine Fever Virus (ASFV) from the Country of Georgia.” *Virology Journal*, vol. 15, no. 1, 14 2018, p. 190. *PubMed*, doi:10.1186/s12985-018-1099-z.

Franzo, Giovanni, et al. “Exploratory Metagenomic Analyses of Periweaning Failure-to-Thrive Syndrome-Affected Pigs.” *The Veterinary Record*, vol. 184, no. 1, Jan. 2019, p. 25. *PubMed*, doi:10.1136/vr.105125.

Hansen, Kasper D., et al. “Biases in Illumina Transcriptome Sequencing Caused by Random Hexamer Priming.” *Nucleic Acids Research*, vol. 38, no. 12, July 2010, p. e131. *PubMed*, doi:10.1093/nar/gkq224.

Hause, Ben M., et al. “Feral Swine Virome Is Dominated by Single-Stranded DNA Viruses and Contains a Novel Orthopneumovirus Which Circulates Both in Feral and Domestic Swine.” *Journal of General Virology*, vol. 97, no. 9, 2016, pp. 2090–95. *Microbiology Society Journals*, doi:10.1099/jgv.0.000554.

Holtkamp, Derald J. et al. *Economic Impact of Porcine Reproductive and Respiratory Syndrome Virus on U.s. Pork Producers*. Ames (Iowa): Iowa State University, 2012. Web.

Hunt, Martin, et al. “IVA: Accurate de Novo Assembly of RNA Virus Genomes.” *Bioinformatics (Oxford, England)*, vol. 31, no. 14, July 2015, pp. 2374–76. *PubMed*, doi:10.1093/bioinformatics/btv120.

“International Committee on Taxonomy of Viruses (ICTV).” *International Committee on Taxonomy of Viruses (ICTV)*, https://talk.ictvonline.org//taxonomy/p/taxonomy-history?taxnode_id=201856468. Accessed 30 Apr. 2019.

Khan, Abdul Rafay, et al. “A Comprehensive Study of De Novo Genome Assemblers: Current Challenges and Future Prospective.” *Evolutionary Bioinformatics Online*, vol. 14, 2018, p. 1176934318758650. *PubMed*, doi:10.1177/1176934318758650.

Knipe, David Mahan, and Peter M. Howley. *Fields’ Virology*. Lippincott Williams & Wilkins, 2007.

Knutson, Todd P., et al. “A Porcine Enterovirus G Associated with Enteric Disease Contains a Novel Papain-like Cysteine Protease.” *The Journal of General Virology*, vol. 98, no. 6, June 2017, pp. 1305–10. *PubMed Central*, doi:10.1099/jgv.0.000799.

Ladinig, Andrea, et al. “Variation in Fetal Outcome, Viral Load and ORF5 Sequence Mutations in a Large Scale Study of Phenotypic Responses to Late Gestation Exposure to Type 2 Porcine Reproductive and Respiratory Syndrome Virus.” *PLoS ONE*, edited by Xiang-Jin Meng, vol. 9, no. 4, Apr. 2014, p. e96104. *DOI.org (Crossref)*, doi:10.1371/journal.pone.0096104.

Lee, Yu Na, and Changhee Lee. “Complete Genome Sequence of a Novel Porcine Parainfluenza Virus 5 Isolate in Korea.” *Archives of Virology*, vol. 158, no. 8, Aug. 2013, pp. 1765–72. *Springer Link*, doi:10.1007/s00705-013-1770-z.

Leifer, Immanuel, et al. “Escape of Classical Swine Fever C-Strain Vaccine Virus from Detection by C-Strain Specific Real-Time RT-PCR Caused by a Point Mutation in the

Primer-Binding Site." *Journal of Virological Methods*, vol. 166, no. 1–2, June 2010, pp. 98–100. *PubMed*, doi:10.1016/j.jviromet.2010.03.004.

Lin, Y., et al. "Induction of Apoptosis by Paramyxovirus Simian Virus 5 Lacking a Small Hydrophobic Gene." *Journal of Virology*, vol. 77, no. 6, Mar. 2003, pp. 3371–83. *Crossref*, doi:10.1128/JVI.77.6.3371-3383.2003.

Liu, Ye, et al. "Parainfluenza Virus 5 as Possible Cause of Severe Respiratory Disease in Calves, China." *Emerging Infectious Diseases*, vol. 21, no. 12, Dec. 2015, pp. 2242–44. *PubMed Central*, doi:10.3201/eid2112.141111.

Loula, T. "Mystery Pig Disease." *Agri-Practice*, vol. 12, no. 1, 1991, pp. 23–34.

Martínez-Lobo, F. J., et al. "Comparative Pathogenicity of Type 1 and Type 2 Isolates of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) in a Young Pig Infection Model." *Veterinary Microbiology*, vol. 154, no. 1–2, Dec. 2011, pp. 58–68. *DOI.org (Crossref)*, doi:10.1016/j.vetmic.2011.06.025.

Meng, X. J., et al. "Sequence Comparison of Open Reading Frames 2 to 5 of Low and High Virulence United States Isolates of Porcine Reproductive and Respiratory Syndrome Virus." *Journal of General Virology*, vol. 76, no. 12, Dec. 1995, pp. 3181–88. *DOI.org (Crossref)*, doi:10.1099/0022-1317-76-12-3181.

Morozova, Olena, and Marco A. Marra. "Applications of Next-Generation Sequencing Technologies in Functional Genomics." *Genomics*, vol. 92, no. 5, Nov. 2008, pp. 255–64. *PubMed*, doi:10.1016/j.ygeno.2008.07.001.

"Mumps - Symptoms and Causes." *Mayo Clinic*, http://www.mayoclinic.org/diseases-conditions/mumps/symptoms-causes/syc-20375361. Accessed 1 Dec. 2018.

Nelson, Martha I., et al. "Human-Origin Influenza A(H3N2) Reassortant Viruses in Swine, Southeast Mexico." *Emerging Infectious Diseases*, vol. 25, no. 4, 2019, pp. 691–700. *PubMed*, doi:10.3201/eid2504.180779.

Oem, Jae-Ku, et al. "Molecular Characteristics of Canine Parainfluenza Viruses Type 5 (CPIV-5) Isolated in Korea." *Canadian Journal of Veterinary Research*, vol. 79, no. 1, Jan. 2015, pp. 64–67.

Oka, Tomoichiro, et al. "First Complete Genome Sequences of Genogroup V, Genotype 3 Porcine Sapoviruses: Common 5′-Terminal Genomic Feature of Sapoviruses." *Virus Genes*, vol. 53, no. 6, Dec. 2017, pp. 848–55. *DOI.org (Crossref)*, doi:10.1007/s11262-017-1481-8.

Okumura, N., et al. "Sequence Polymorphisms in Porcine Homologs of Murine Coat Colour-Related Genes." *Animal Genetics*, vol. 41, no. 2, Apr. 2010, pp. 113–21. *PubMed*, doi:10.1111/j.1365-2052.2009.01968.x.

Optimal Template Loading Concentration and Cluster Densities for Illumina Instruments. https://genohub.com/loading-concentrations-optimal-cluster-density/. Accessed 4 Aug. 2019.

Pettersson, Erik, et al. *Generations of Sequencing Technologies | Elsevier Enhanced Reader*. doi:10.1016/j.ygeno.2008.10.003. Accessed 4 Aug. 2019.

Philips, Ron Milo &. Ron. *» How Big Are Viruses?* http://book.bionumbers.org/how-big-are-viruses/. Accessed 4 Aug. 2019.

Plotzki, Elena, et al. "Microbiological Characterization of a Newly Established Pig Breed, Aachen Minipigs." *Xenotransplantation*, vol. 23, no. 2, Mar. 2016, pp. 159–67. *DOI.org (Crossref)*, doi:10.1111/xen.12233.

Pogranichniy, Roman M., et al. "Case-Control Study on the Association of Porcine Circovirus Type 2 and Other Swine Viral Pathogens with Postweaning Multisystemic Wasting Syndrome." *Journal of Veterinary Diagnostic Investigation: Official Publication of the American Association of Veterinary Laboratory Diagnosticians, Inc*, vol. 14, no. 6, Nov. 2002, pp. 449–56. *PubMed*, doi:10.1177/104063870201400601.

Poor, Taylor A., et al. "On the Stability of Parainfluenza Virus 5 F Proteins." *Journal of Virology*, edited by D. S. Lyles, vol. 89, no. 6, Mar. 2015, pp. 3438–41. *Crossref*, doi:10.1128/JVI.03221-14.

Qin, Sinan, et al. “Viral Communities Associated with Porcine Respiratory Disease Complex in Intensive Commercial Farms in Sichuan Province, China.” *Scientific Reports*, vol. 8, no. 1, Sept. 2018, p. 13341. *www.nature.com*, doi:10.1038/s41598-018-31554-8.

Reed, L. J., and H. Muench. “A SIMPLE METHOD OF ESTIMATING FIFTY PER CENT ENDPOINTS.” *American Journal of Epidemiology*, vol. 27, no. 3, May 1938, pp. 493–97. *academic.oup.com*, doi:10.1093/oxfordjournals.aje.a118408.

Rossow, Kurt D., et al. “Experimental Porcine Reproductive and Respiratory Syndrome Virus Infection in One-, Four-, and 10-Week-Old Pigs.” *Journal of Veterinary Diagnostic Investigation*, vol. 6, no. 1, Jan. 1994, pp. 3–12. *DOI.org (Crossref)*, doi:10.1177/104063879400600102.

Saade, Georges, et al. “Coinfections and Their Molecular Consequences in the Porcine Respiratory Tract.” *Veterinary Research*, vol. 51, no. 1, June 2020, p. 80.

Sequencing Platforms | Compare NGS Platform Applications & Specifications. https://emea.illumina.com/systems/sequencing-platforms.html?langsel=/ch/. Accessed 8 Aug. 2019.

Sequencing Quality Scores. https://emea.illumina.com/science/technology/next-generation-sequencing/plan-experiments/quality-scores.html?langsel=/ch/. Accessed 5 Aug. 2019.

Shendure, Jay, et al. “DNA Sequencing at 40: Past, Present and Future.” *Nature*, vol. 550, no. 7676, Oct. 2017, pp. 345–53. *DOI.org (Crossref)*, doi:10.1038/nature24286.

Shepherd, Frances K., et al. “Whole Genome Classification and Phylogenetic Analyses of Rotavirus B Strains from the United States.” *Pathogens (Basel, Switzerland)*, vol. 7, no. 2, Apr. 2018. *PubMed*, doi:10.3390/pathogens7020044.

Slatko, Barton E., et al. “DNA Sequencing by the Dideoxy Method.” *Current Protocols in Molecular Biology*, vol. 47, no. 1, 1999, pp. 7.4A.1-7.4A.39. *Wiley Online Library*, doi:10.1002/0471142727.mb0704as47.

Thomas, Torsten, et al. “Metagenomics - a Guide from Sampling to Data Analysis.” *Microbial Informatics and Experimentation*, vol. 2, Feb. 2012, p. 3. *PubMed Central*, doi:10.1186/2042-5783-2-3.

Tritt, Andrew, et al. “An Integrated Pipeline for de Novo Assembly of Microbial Genomes.” *PLOS ONE*, vol. 7, no. 9, Sept. 2012, p. e42304. *PLoS Journals*, doi:10.1371/journal.pone.0042304.

Tyson, Gene W., et al. “Community Structure and Metabolism through Reconstruction of Microbial Genomes from the Environment.” *Nature*, vol. 428, no. 6978, Mar. 2004, pp. 37–43. *PubMed*, doi:10.1038/nature02340.

Venter, J. C. “Environmental Genome Shotgun Sequencing of the Sargasso Sea.” *Science*, vol. 304, no. 5667, Apr. 2004, pp. 66–74. *DOI.org (Crossref)*, doi:10.1126/science.1093857.

Walker, Lianna R., et al. “Genome-Wide Association Analysis for Porcine Reproductive and Respiratory Syndrome Virus Susceptibility Traits in Two Genetic Populations of Pigs1.” *Journal of Animal Science*, vol. 97, no. 8, July 2019, pp. 3253–61. *PubMed*, doi:10.1093/jas/skz184.

Wang, Wei, et al. “Sequence and Phylogenetic Analysis of Novel Porcine Parvovirus 7 Isolates from Pigs in Guangxi, China.” *PloS One*, vol. 14, no. 7, 2019, p. e0219560. *PubMed*, doi:10.1371/journal.pone.0219560.

Wang, Yue, et al. “The Evolution of Nanopore Sequencing.” *Frontiers in Genetics*, vol. 5, Jan. 2015. *PubMed Central*, doi:10.3389/fgene.2014.00449.

Waning, D. L., et al. “Activation of a Paramyxovirus Fusion Protein Is Modulated by Inside-out Signaling from the Cytoplasmic Tail.” *Proceedings of the National Academy of Sciences*, vol. 101, no. 25, June 2004, pp. 9217–22. *Crossref*, doi:10.1073/pnas.0403339101.

Wesley, Ronald D., et al. “Differentiation of a Porcine Reproductive and Respiratory Syndrome Virus Vaccine Strain from North American Field Strains by Restriction Fragment Length Polymorphism Analysis of ORF 5.” *Journal of Veterinary Diagnostic Investigation*, vol. 10, no. 2, Apr. 1998, pp. 140–44. *DOI.org (Crossref)*, doi:10.1177/104063879801000204.

Won, Hokeun, et al. “Generation and Protective Efficacy of a Cold-Adapted Attenuated Genotype 2b Porcine Epidemic Diarrhea Virus.” *Journal of Veterinary Science*, vol. 20, no. 4, July 2019, p. e32. *PubMed*, doi:10.4142/jvs.2019.20.e32.

Xu, Guanlong, et al. “Truncation of C-Terminal 20 Amino Acids in PA-X Contributes to Adaptation of Swine Influenza Virus in Pigs.” *Scientific Reports*, vol. 6, Feb. 2016, p. 21845. *PubMed*, doi:10.1038/srep21845.

Zhai, Jun-Qiong, et al. “First Complete Genome Sequence of Parainfluenza Virus 5 Isolated from Lesser Panda.” *Archives of Virology*, vol. 162, no. 5, May 2017, pp. 1413–18. *Springer Link*, doi:10.1007/s00705-017-3245-0.

Zhang, Xiaojian, et al. “Tissue Tropisms Opt for Transmissible Reassortants during Avian and Swine Influenza A Virus Co-Infection in Swine.” *PLoS Pathogens*, vol. 14, no. 12, Dec. 2018. *PubMed Central*, doi:10.1371/journal.ppat.1007417.

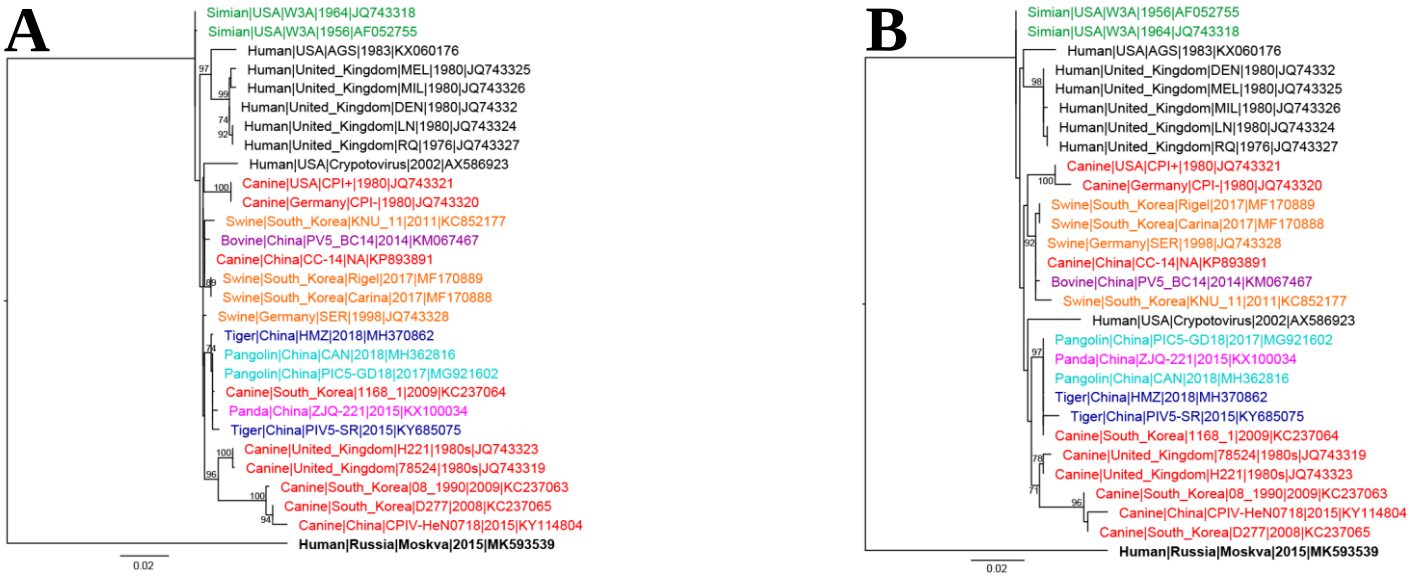
Zhang, Ye, et al. “Genetic and Biological Characteristics of Avian Influenza Virus Subtype H1N8 in Environments Related to Live Poultry Markets in China.” *BMC Infectious Diseases*, vol. 19, no. 1, May 2019, p. 458. *PubMed*, doi:10.1186/s12879-019-4079-z.

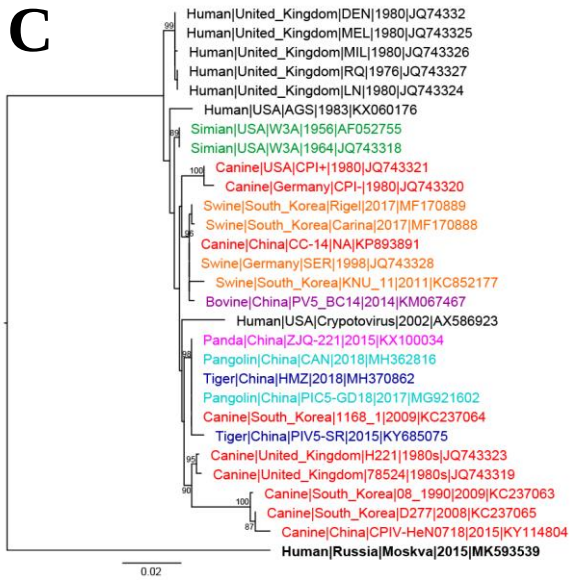
Zhou, Long, et al. “Emergence of Two Novel Recombinant Porcine Reproductive and Respiratory Syndrome Viruses 2 (Lineage 3) in Southwestern China.” *Veterinary*

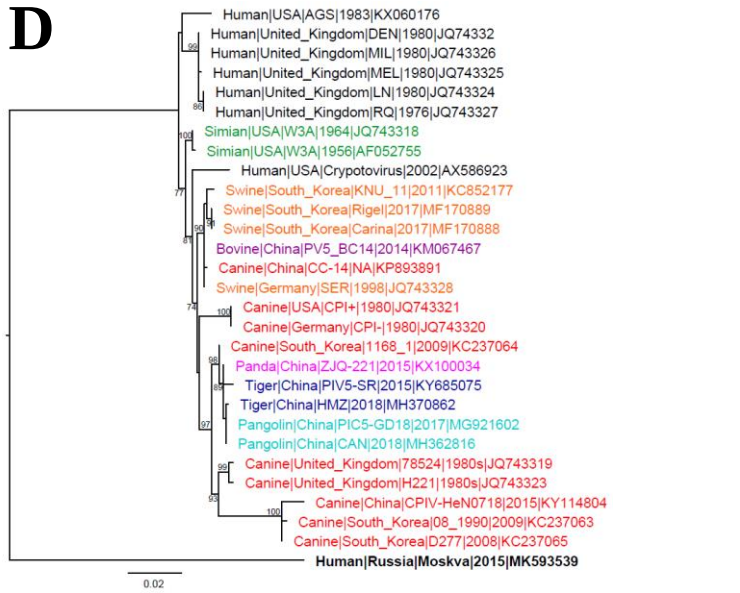
Microbiology, vol. 232, May 2019, pp. 30–41. *PubMed*,
doi:10.1016/j.vetmic.2019.01.026.

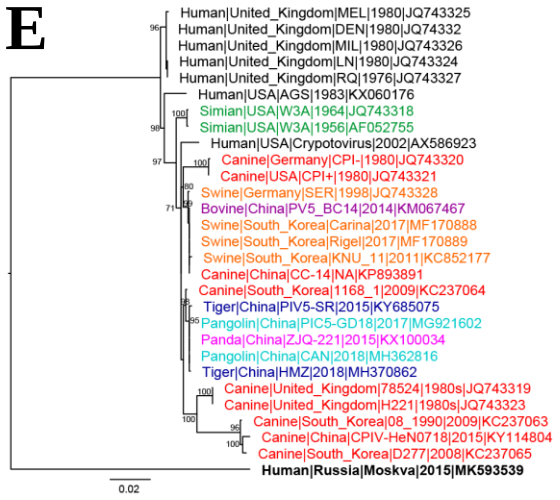
Chapter 5 - Appendix

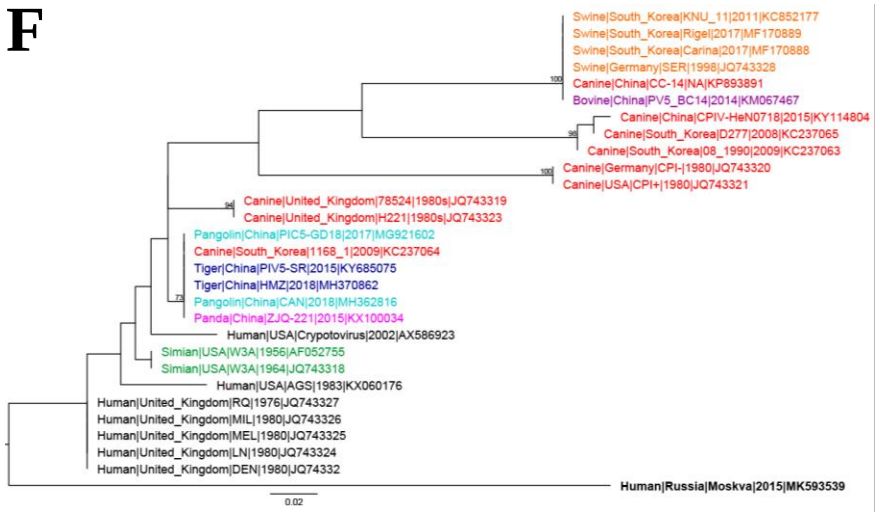
5.1 Supplemental Figure

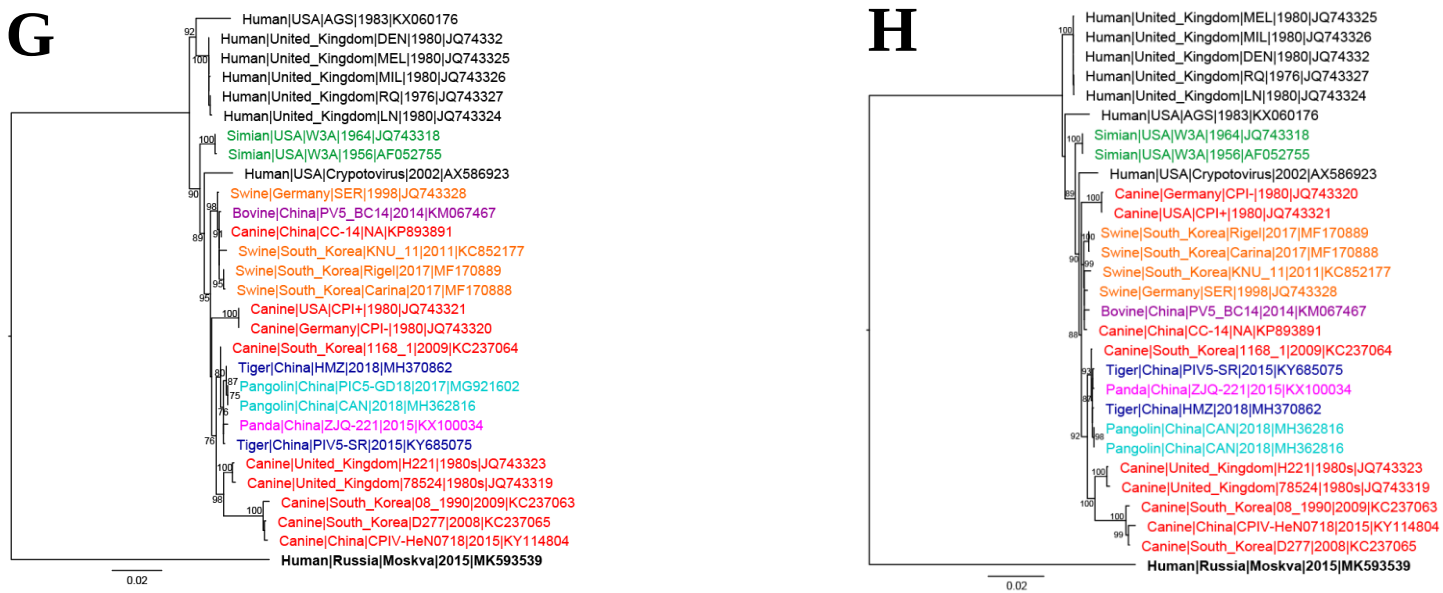



C
Human|United_Kingdom|DEN|1980|JQ74332
Human|United_Kingdom|MEL|1980|JQ743325
Human|United_Kingdom|MIL|1980|JQ743326
Human|United_Kingdom|RQ|1976|JQ743327
Human|United_Kingdom|LN|1980|JQ743324
Human|USA|AGS|1983|KX060176
Simian|USA|W3A|1956|AF052755
Simian|USA|W3A|1964|JQ743318
Canine|USA|CPI+|1980|JQ743321
Canine|Germany|CPI-|1980|JQ743320
Swine|South_Korea|Rigel|2017|MF170889
Swine|South_Korea|Carina|2017|MF170888
Canine|China|CC-14|NA|KP893891
Swine|Germany|SER|1998|JQ743328
Swine|South_Korea|KNU_11|2011|KC852177
Bovine|China|PV5_BC14|2014|KM067467
Human|USA|Crypotovirus|2002|AX586923
Panda|China|ZJQ-221|2015|KX100034
Pangolin|China|CAN|2018|MH362816
Tiger|China|HMZ|2018|MH370862
Pangolin|China|PIC5-GD18|2017|MG921602
Canine|South_Korea|1168_1|2009|KC237064
Tiger|China|PIV5-SR|2015|KY685075
Canine|United_Kingdom|H221|1980s|JQ743323
Canine|United_Kingdom|78524|1980s|JQ743319
Canine|South_Korea|08_1990|2009|KC237063
Canine|South_Korea|D277|2008|KC237065
Canine|China|CPIV-HeN0718|2015|KY114804
Human|Russia|Moskva|2015|MK593539
0.02


D
Human|USA|AGS|1983|KX060176
Human|United_Kingdom|DEN|1980|JQ74332
Human|United_Kingdom|MIL|1980|JQ743326
Human|United_Kingdom|MEL|1980|JQ743325
Human|United_Kingdom|LN|1980|JQ743324
Human|United_Kingdom|RQ|1976|JQ743327
Simian|USA|W3A|1964|JQ743318
Simian|USA|W3A|1956|AF052755
Human|USA|Crypotovirus|2002|AX586923
Swine|South_Korea|KNU_11|2011|KC852177
Swine|South_Korea|Rigel|2017|MF170889
Swine|South_Korea|Carina|2017|MF170888
Bovine|China|PV5_BC14|2014|KM067467
Canine|China|CC-14|NA|KP893891
Swine|Germany|SER|1998|JQ743328
Canine|USA|CPI+|1980|JQ743321
Canine|Germany|CPI-|1980|JQ743320
Canine|South_Korea|1168_1|2009|KC237064
Panda|China|ZJQ-221|2015|KX100034
Tiger|China|PIV5-SR|2015|KY685075
Tiger|China|HMZ|2018|MH370862
Pangolin|China|PIC5-GD18|2017|MG921602
Pangolin|China|CAN|2018|MH362816
Canine|United_Kingdom|78524|1980s|JQ743319
Canine|United_Kingdom|H221|1980s|JQ743323
Canine|China|CPIV-HeN0718|2015|KY114804
Canine|South_Korea|08_1990|2009|KC237063
Canine|South_Korea|D277|2008|KC237065
Human|Russia|Moskva|2015|MK593539
0.02


E
Human|United_Kingdom|MEL|1980|JQ743325
Human|United_Kingdom|DEN|1980|JQ74332
Human|United_Kingdom|MIL|1980|JQ743326
Human|United_Kingdom|LN|1980|JQ743324
Human|United_Kingdom|RQ|1976|JQ743327
Human|USA|AGS|1983|KX060176
Simian|USA|W3A|1964|JQ743318
Simian|USA|W3A|1956|AF052755
Human|USA|Crypotovirus|2002|AX586923
Canine|Germany|CPI-|1980|JQ743320
Canine|USA|CPI+|1980|JQ743321
Swine|Germany|SER|1998|JQ743328
Bovine|China|PV5_BC14|2014|KM067467
Swine|South_Korea|Carina|2017|MF170888
Swine|South_Korea|Rigel|2017|MF170889
Swine|South_Korea|KNU_11|2011|KC852177
Canine|China|CC-14|NA|KP893891
Canine|South_Korea|1168_1|2009|KC237064
Tiger|China|PIV5-SR|2015|KY685075
Pangolin|China|PIC5-GD18|2017|MG921602
Panda|China|ZJQ-221|2015|KX100034
Pangolin|China|CAN|2018|MH362816
Tiger|China|HMZ|2018|MH370862
Canine|United_Kingdom|78524|1980s|JQ743319
Canine|United_Kingdom|H221|1980s|JQ743323
Canine|South_Korea|08_1990|2009|KC237063
Canine|China|CPIV-HeN0718|2015|KY114804
Canine|South_Korea|D277|2008|KC237065
Human|Russia|Moskva|2015|MK593539
0.02


F
Swine|South_Korea|KNU_11|2011|KC852177
Swine|South_Korea|Rigel|2017|MF170889
Swine|South_Korea|Carina|2017|MF170888
Swine|Germany|SER|1998|JQ743328
Canine|China|CC-14|NA|KP893891
Bovine|China|PV5_BC14|2014|KM067467
Canine|China|CPIV-HeN0718|2015|KY114804
Canine|South_Korea|D277|2008|KC237065
Canine|South_Korea|08_1990|2009|KC237063
Canine|Germany|CPI-|1980|JQ743320
Canine|USA|CPI+|1980|JQ743321
Canine|United_Kingdom|78524|1980s|JQ743319
Canine|United_Kingdom|H221|1980s|JQ743323
Pangolin|China|PIC5-GD18|2017|MG921602
Canine|South_Korea|1168_1|2009|KC237064
Tiger|China|PIV5-SR|2015|KY685075
Tiger|China|HMZ|2018|MH370862
Pangolin|China|CAN|2018|MH362816
Panda|China|ZJQ-221|2015|KX100034
Human|USA|Crypotovirus|2002|AX586923
Simian|USA|W3A|1956|AF052755
Simian|USA|W3A|1964|JQ743318
Human|USA|AGS|1983|KX060176
Human|United_Kingdom|RQ|1976|JQ743327
Human|United_Kingdom|MIL|1980|JQ743326
Human|United_Kingdom|MEL|1980|JQ743325
Human|United_Kingdom|LN|1980|JQ743324
Human|United_Kingdom|DEN|1980|JQ74332
Human|Russia|Moskva|2015|MK593539
0.02



Supplemental Figure 5.1 Individual gene [A, nucleocapsid (N); B, V; C, phosphoprotein (P); D, matrix (M); E, fusion (F); F, SH; G, hemagglutinin-neuraminidase (HN); and H, large (L)] phylogenetic trees of PIV5 strains. Bootstrap values greater than 70 are located at major nodes. Strains are colored according to host with black representing human strains; red, canine; orange, swine; green, simian; teal, pangolin; blue, tiger; pink, panda; and purple, bovine. Moskva is in bold.