Use of Sequencing Technologies to Improve Swine Infectious Disease Management

A DISSERTATION

SUMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Shaoyuan Tan

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

ADVISED BY DR. MICHAEL P. MURTAUGH & DR. CHERYL M.T. DVORAK

FEBURARY 2020

Acknowledgements

I would like to take this opportunity to express my gratitude to all those who helped me to make my thesis work a success.

First and foremost, I would like to thank my mentors and advisors, Dr. Michael Murtaugh and Dr. Cheryl Dvorak. Mike gave me the opportunity to work in his lab. I am very thankful for Mike's education of scientific thinking and his devotion to the lab. He continues to be an inspiration to me even after his passing. Cheryl has been the most important person to the completion of my PhD journey. She gave me very helpful guidance and tremendous support regarding my research and manuscript writing. I can't imagine what things would have been like without her. I feel lucky to have the two excellent advisors.

Many thanks to my committee members, Dr. Mark Rutherford, Dr. Andres Perez, Dr. Chad Myers, and Dr. Douglas Marthaler, for their inspiring input to the research. I greatly enjoyed talking to them about research and career. I appreciated their time and efforts.

I would like to acknowledge my friends and colleagues for their friendship and support. I am grateful to the members of the Murtaugh lab, Kyra Martins, Jordan Young, Frances Shepherd, and Suzanne Stone, for the enjoyable working environment. I also want to thank the Veterinary Medicine graduate program (VMED) and the College of Veterinary Medicine for the opportunity to pursue my graduate study. I would especially like to thank Lisa Hubinger for her efforts in coordinating the graduate program.

My research was supported by grant funding from Boehringer Ingelheim and the National Pork Board. I would like to thank the Leman China Conference and MnDRIVE

Global Food Ventures for fellowships to pursue my research. I am also grateful for several travel grants provided by College of Veterinary Medicine.

I would like to thank my partner, Patrick Mathwig, for the endless encouragement, love and patience. Last but not least, I cannot express depth of my gratitude to my parents and my sister. Their unconditional love and support have driven me through ups and downs of life. My family's values of having an open mind and positive attitude will continue to guide me in the future.

Dedication

This dissertation is dedicated to my beloved parents and sister,

who have been with me every step of the way.

Abstract

The swine industry is an indispensable part of the food production and agricultural system. However, infectious disease poses great threats to the industry. The current situation is that “old” infectious diseases are not totally under control and “new” pathogens keep emerging. In addition, via food chain and direct or indirect contact, some swine pathogens can infect humans and antimicrobial resistance arising from swine pathogens can also adversely impact public health. The threat of infectious disease to the swine population has been further magnified by globalization which increases the rate and breadth of disease transmission, thus having a more devastating impact.

Fortunately, scientific advances have greatly bolstered our ability to develop effective disease control strategies. Sequencing technology has emerged as a powerful solution to deal with the challenging infectious disease situations we are facing today. The advantages of sequencing as a diagnostic tool are numerous. From our research, we confirmed that sequencing has improved the disease diagnostic resolution to the strain level, providing more precise and effective control strategies; it is able to generate additional genomic information for functional prediction of infectious agents, such as antimicrobial resistance profiling; and sequencing can facilitate a prompt response against emerging eventualities due to its ability to rapidly detect pathogens in a sample, including the novel or unexpected ones.

Going forward, the more frequent use of sequencing in swine medicine can enhance our ability to predict and control the emergence and transmission of infectious disease within swine populations.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AMR	antimicrobial resistance
ASF	African swine fever
BLAST	basic local alignment search tool
DRS	direct RNA sequencing
E. coli	*Escherichia coli*
EID	emerging infectious disease
ELISA	enzyme-linked immunosorbent assay
FMD	foot-and-mouth disease
HIV	human immunodeficiency virus
MLST	multilocus sequence typing
PCR	polymerase chain reaction
PCS	cDNA-PCR sequencing
PCV2	porcine circovirus type 2
PEDV	porcine epidemic diarrhea virus
PRDC	porcine respiratory disease complex
PRRS	porcine reproductive and respiratory syndrome
PRRSV	porcine reproductive and respiratory syndrome virus
RCS	RNA Calibration Strand
RT-qPCR	real-time quantitative reverse transcription polymerase chain reaction
S. suis	*Streptococcus suis*
ST	sequence type

SVA	senecavirus A
WGS	whole genome sequencing

Chapter I - A review of challenges existing in current swine industry and how sequencing technologies can improve swine infectious disease management

1. Introduction

The swine industry is an important component of agricultural and food production (FAO 2019). Due to the increasing human population and the improved economic status of people, there is a greater demand for animal-based proteins (Vallat 2014). The intensification of the swine industry is an inevitable trend (Davies 2012). Expanding swine farm size and increase in frequency of global trading activities has created an opportunity for emergence and spreading of infectious disease.

Infectious disease is now among the primary constraints to swine production and are expected to be an ongoing challenge (VanderWaal and Deen 2018). The direct adverse impact of infectious disease on the swine industry is decreased production and economic loss. It is estimated that porcine reproductive and respiratory syndrome (PRRS) outbreaks reduce swine production by approximately 7.4% on average annually (Valdes-Donoso et al. 2018). In addition to the direct impacts on the swine industry, swine pathogens can infect human populations (Smith 2015; Davies 2012) such as swine influenza virus, a swine zoonotic pathogen, which caused the "2009 H1N1 Pandemic" in humans, affecting over 200 countries and resulting in more than 12 thousands deaths in the U.S. alone (Sullivan et al. 2010). There is also a growing occurrence of antimicrobial resistance (AMR) in swine infectious agents that can transfer to the human population and have a great potential to

threaten public health (Yang et al. 2019). Therefore, effective swine infectious disease management is greatly needed to maintain the stability and productivity of swine production, as well as to ensure public health.

Pigs can be infected by a complex matrix of infectious agents, including bacteria, viruses, parasites and protozoans (VanderWaal and Deen 2018). Laboratory diagnostics are needed to investigate the cause of infectious diseases. The current widely used molecular diagnostic methods, such as polymerase chain reaction (PCR), and serology methods, such as enzyme-linked immunosorbent assay (ELISA), can detect target pathogens at species level with a prior hypothesis of possible pathogens under investigation. While this method is effective in most situations for routine disease surveillance and outbreaks investigation, the huge and continuously increasing diversity of pathogen species/variants can cause misdiagnosis using the previously described method due to its inability to detect new or unexpected pathogens/variants (Morens and Fauci 2013). In addition, it is common that swine diseases, especially the respiratory diseases and enteric diseases (Theuns et al. 2018; Qin et al. 2018), are caused by multiple pathogens. Previously, it has been almost impossible to correctly predict the presence of all pathogens in a sample since pathogen detection is performed mainly using hypothesis-driven testing, such as PCR. Unbiased, hypothesis-free diagnostic methods are needed to tackle the challenging swine infectious disease situation and to assist in better disease management.

Sequencing is one of the new technologies that has been utilized as an effective tool for disease investigation in the field of veterinary medicine (Van Borm et al. 2015). Sequencing has been used in swine for precise detection and characterization of bacterial and viral pathogens, such as novel pathogen/variant detection, pathogen lineage

identification, sequence typing, genotyping, etc. (Han et al. 2019; Keller et al. 2018). Sequencing has also been used for unbiased microbial population diversity characterization, thus providing a new approach to investigate coinfections and complex diseases (Hersh and Vachani 2017; Qin et al. 2018).

The facts are, infectious diseases have no country boundary and animal pathogens can spread to human populations. Thus, combined global efforts from multiple disciplines are needed to win the battle against infectious diseases. This review summarizes the current situations of swine infectious diseases and examines the impact of recent sequencing technologies on infectious disease management.

2. Situation of current swine infectious diseases

2.1. Emerging infectious diseases, especially RNA viruses, pose a big threat to the swine industry

Emerging infectious diseases, that include emergence of new species or strains, have always represented a major threat to the swine industry throughout history and are expected to remain a considerable challenge in the foreseeable future (Barrett et al. 1998; Morens and Fauci 2013). There are two major categories of emerging infections, newly emerging (those that have not previously been recognized) and reemerging infectious diseases (those that existed in the past but are now rapidly increasing either in incidence or in geographical or host range) (Morens, Folkers, and Fauci 2004).

Currently, the swine industry is facing a variety of infectious pathogens, and new pathogens keep emerging. Among these pathogens, viruses and bacteria account for more than 75% of the top pathogens under investigation according to a recent analysis of 57,471 publications from 1996 to 2016 (VanderWaal and Deen 2018). Unlike bacterial genetic

material, which is always DNA, viral genomes may be DNA or RNA (Bethesda 2007). RNA viruses are predicted to be the biggest emerging threat to the swine industry due to their higher mutation rates and adaptability compared to pathogens with DNA genomes (bacterial or DNA viruses) (Marston et al. 2014). This prediction has been proven by the observed fact that most of emerging swine infectious diseases in recent years are caused by RNA viruses, including the most devastating ones like PRRS virus (PRRSV) and porcine epidemic diarrhea virus (PEDV). Diseases caused by RNA viruses are difficult to prevent and control, because they evolve rapidly and produce new variants continuously (Drew 2011), which can cause missed diagnosis and vaccination failure. In addition, there is a special concern of swine RNA viruses causing zoonotic disease (such as Nipah virus) due to their intrinsic nature of rapid adaptive rates to new host species (Carrasco-Hernandez et al. 2017) .

PRRS which is caused by PRRSV, was indicated to be one of the most devastating swine pathogens between 2006–2016 for North America and Eastern Asia (VanderWaal and Deen 2018; Murtaugh et al. 2010). PRRSV infection has been a severe threat to the swine industry globally since it was first observed in the 1980s. Due to the frequency of global trade, pathogenic PRRSV strains in one place can transmit to other places, and new virulent strains keep emerging due to mutation and recombination. For example, the NADC30 strain first broke out in America and then later in China, where it recombined with the high pathogenic PRRSV and other local strains, generating new variants (Han et al. 2019; Zhou et al. 2018). It is evident that globalization facilitates the pathogen spread at a larger geographical scale and magnifies the economic loss. This has also been demonstrated by transmission of PEDV, which was first identified in the 1980s in China

(Wang, Fang, and Xiao 2016), spread to the U.S. in 2013 and had impacted about half of U.S. swine herds, resulting in the deaths of more than seven million piglets (Goede and Morrison 2016). Today, the swine industry in China is a hot zone of African Swine Fever (ASF) (Wang, Sun, and Qiu 2018). The disease spread throughout the continent of Africa, then Europe, and finally to China, transmitting from north to south and causing an estimated reduction of more than half the pig population. It is currently spreading beyond China to its neighboring countries with a huge potential to spread to more even further, including to the U.S. (Sur 2019).

Emerging infectious diseases are unpredictable, but inevitable and very likely to be devastating. The ever-increasing connection and interdependence of the world creates opportunities for infectious agents' transmission globally and further magnifies the adverse impact of diseases. When an emerging infectious agent enter into a population, most likely, there is no immunity against it, no effective vaccine or treatments for it. It is of crucial importance to react promptly and implement effective intervention strategies when dealing with emerging infectious diseases, which relied on the development and implementation of reliable, accurate, and rapid diagnostic methods.

2.2. Co-infections and disease complexes complicated disease diagnosis and treatments

Co-infection is the simultaneous infection of a host caused by multiple pathogens, which happens frequently in humans and animals (Brealey et al. 2015; Ouyang et al. 2019). Co-infections have been shown to reduce disease severity in very few cases (Gonzalez et al. 2018). In most cases, they will augment disease severity (Ouyang et al. 2019). One extreme example was human immunodeficiency virus (HIV) co-infections in human,

which always have adverse, even deadly, consequences (Naidoo, Mahomed, and Moodley 2017; Rajbhandari et al. 2016).

Co-infection is generally involved in pig infectious diseases (Choi, Goyal, and Joo 2003). For example, PRRSV is a significant swine pathogen, and it can damage a hosts defenses, resulting in increased susceptibility to infections by secondary pathogens (Niederwerder et al. 2016). So far, facts from PRRSV co-infection studies included PRRS enhances susceptibility to *Streptococcus suis (S. suis)* challenge (Thanawongnuwech et al. 2000); co-infection of PRRSV and swine influenza virus created a very severe and lasting respiratory disease (Van Reeth, Nauwynck, and Pensaert 2001); PRRSV infection enhances porcine circovirus type 2 (PCV2) replication (Rovira et al. 2002); and pigs infected with PRRSV have a marked decrease in their ability to kill bacteria (Solano et al. 1998). In addition to PRRSV co-infections with other pathogens, co-infections of multiple PRRSV strains also happen frequently in swine farms (Han et al. 2019; Liu et al. 2018), possibly causing recombination and generating novel pathogenic variants. It is obvious that the co-infections and interaction of PRRSV with other infectious agents made disease investigation and control more complicated. PCV2 is endemic on most farms and most animals are infected with PCV2. However, PCV2 alone doesn't usually cause clinical disease unless there is a co-infection. PCV2 tends to co-infect with other pathogens, such as PRRSV, porcine parvovirus, *Mycoplasma hyopneumoniae*, and *Salmonella spp.*, as well as different PCV2 strains (Ouyang et al. 2019). Co-infections lead to complex diseases. Several diseases syndromes relate to PCV2, including porcine dermatitis and nephropathy syndrome (PDNS), postweaning multisystemic wasting syndrome (PMWS), and porcine respiratory disease complex (PRDC) (Ouyang et al. 2019). Such complex diseases posed a

challenge for diagnostics and disease management. Take PRDC as an example, it is caused by a combination of bacteria, viruses, management, environment and swine genetic conditions (Opriessnig, Giménez-Lirola, and Halbur 2011). There was a long time that no single-reaction diagnostic test existed for the simultaneous detection of major pathogens causing PRDC, which led to limited understanding and poor control of this disease. Investigation of pathogens using specific PCR for each pathogen are too time consuming. Multiplex PCR assay can detect multiple pathogens at one time (Liu et al. 2013; El Kholy et al. 2016). However, it is still not an unbiased and robust method, because prior hypothesis of existence and prior knowledge of genetic information for primer design are needed. As a result, diagnostic laboratories usually fail to detect all disease relevant infectious agents because of the high specificity of current diagnostic assays and conflicting results and conclusions were drawn from the study of the same disease complex (McArdle, Turkova, and Cunnington 2018). In order to address existing co-infections and complex diseases challenges in the swine industry, an unbiased, broad-range detection method is urgently needed.

2.3. Zoonotic swine diseases indicated the importance of transdisciplinary collaboration

The ultimate purpose of rearing domestic swine is to supply nutritious protein to humans, which is meant to be a positive impact on human health. However, swine zoonotic diseases, caused by bacteria, viruses, parasites and other infectious agents, adversely impact human health. It is estimated that zoonotic agents comprise 60% of the known pathogens to humans (Jones et al. 2008; Tomley and Shirley 2009). This indicates an urgent need to understand the dynamics of zoonotic diseases in order to mitigate the threats to animal and public health (Cleaveland, Laurenson, and Taylor 2001). Pathogens infecting

swine have a higher probability of spreading to humans compared to pathogens from other animals due to the intensity of swine farming worldwide and the close contact between pigs and humans (Uddin Khan et al. 2013).

There are a number of known pathogens infecting swine that can transmit to humans, such as *S. suis* and swine influenza virus (Smith 2015; Smith et al. 2011). Since the first human *S. suis* cases were diagnosed in Denmark (Perch, Kristjansen, and Skadhauge 1968), human infections of *S. suis* infection have been reported in more than 24 countries worldwide so far (Gottschalk, Segura, and Xu 2007; Hughes et al. 2009). Especially in Asia, *S. suis* is classified as an emerging zoonotic pathogen and has caused outbreaks of severe human infections. The swine influenza virus situation is more complicated since it can infect multiple hosts. Pigs' susceptibility to both human and avian influenza viruses permit them to be infected with viruses originating from both mammals and birds. This in return leads to reassortment of genetic material between multiple subtypes and generates new influenza viruses. In April 2009, the novel swine-origin H1N1 virus was detected in humans. The virus quickly spread to over 200 countries, causing significant harm (Sullivan et al. 2010). The Centers for Disease Control and Prevention (CDC) estimated that between 43 and 89 million individuals were infected in the United States within one year (https://www.cdc.gov/flu/pandemic-resources/2009-h1n1-pandemic.html). Until now, influenza infections in humans continue to occur, posing a significant risk to the public (Bowman et al. 2017; Bailey et al. 2018).

While it is more concerning that infectious diseases of animals can infect human beings, transmission of some diseases from humans to domestic animals is also common (Pearce-Duvet 2006). In fact, reverse zoonosis of influenza passed from humans to swine

has been shown far more frequent than swine to human (Nelson and Vincent 2015). The frequent transmission of pathogens among humans and other animals strongly indicates the importance and necessity for a “One Health” approach and collaborations from different fields to deal with infectious diseases.

2.4. The rising AMR calls for high resolution diagnosis and precision treatments

Right after Fleming discovered penicillin in 1946, he gave a warning that “the thoughtless person playing with penicillin treatment is morally responsible for the death of the man who finally succumbs to infection with the penicillin-resistant organism” (Fleming 1945). Unfortunately, Fleming's warning was unheeded, bacteria that are resistant to penicillin emerged, followed by emergence of bacteria that are resistant to other antimicrobials. AMR is a natural product of use of antimicrobials, which can be accelerated by overuse and improper use (Chantziaras et al. 2014). The continuously increasing AMR in bacteria is now one of the major concerns in human and veterinary medicine. There are a number of drivers responsible for this situation and the widespread and uncontrolled use of antibiotics in animal production is one of them (Landers et al. 2012).

Antibiotics have been used routinely in farm animal production since the 1950s to keep animals healthy and thus increase productivity (Lekagul 2019). It is estimated that 80% of antibiotics sold in the U.S. are used in animals, which include therapeutic use to treat bacterial infections and subtherapeutic purposes to promote growth (Ventola 2015b). The emergence of AMR allows pathogens to evade the action of antimicrobial drugs and cause reduction of pig production due to poor disease control outcomes. In addition, it can adversely impact human health through the transfer of resistant organisms and associated genes via the food chain (Ventola 2015b; 'The antibiotic alarm' 2013). To make things

worse, antibiotic development is no longer an economically wise investment, generating fewer available new antibiotics (Bartlett, Gilbert, and Spellberg 2013; Ventola 2015b; Piddock 2012).

How to recover the miracle of antibiotics has become a real question that needs to be addressed immediately (Cars 2014). Completely curtailing the use of antibiotics in livestock is unrealistic considering animal health, welfare and economics. In order to regulate antibiotic use, the FDA released a new policy in 2018 for the use of medically important antimicrobials in livestock that requires a rigorous evaluation before therapeutic use of these drugs (U.S. Food and Drug Administration, https://www.fda.gov/AnimalVeterinary/SafetyHealth/AntimicrobialResistance, 7/1/2019). It is important to recognize particular needs which can be determined using molecular diagnostic methods and treat animals with specifically targeted antimicrobials, instead of empiric order, which will effectively control disease and at the same time avoid overuse (Davies 2012; 'The antibiotic alarm' 2013; Ventola 2015a; Lekagul 2019). Ideal diagnostic methods should be able to detect the genes related to AMR at the point of care, enabling rapid AMR identification and effective disease treatments. Hopefully, understanding the prevalence of pathogens in pigs, levels of resistance and susceptibility to different antibiotics, and prudent use of antibiotics can slow down the emergence of AMR in animal production and avoid the risk of humanity losing control of the ability to manage bacterial infections with antibiotics (Cars 2014; Bartlett, Gilbert, and Spellberg 2013).

3. Current strategies to control swine infectious diseases

Infectious diseases will continue to emerge and re-emerge (Morens and Fauci 2013). When it comes to combating infectious diseases, prevention is the first step followed by

prompt reaction and effective control of disease spread. Current prevention solutions to swine diseases primarily focuses on developing vaccines to reduce the susceptibility of pigs, along with implementing biosecurity strategies at farms and between farms to reduce the risk of spread. Routine disease surveillance and prompt outbreak response are key elements in controlling the spread infectious diseases, which depends on rapid clinical diagnosis and infectious agent identification. The awareness of biosecurity, the development of vaccines, the improvements on diagnostic methods, and the regulated use of antibiotics are some key elements currently standard practice for infectious disease control in the swine industry.

3.1. Vaccination and biosecurity

Vaccines are one of the most amazing advances to improve animal health (Rose and Andraud 2017). The basic principle of vaccines is to confer protection to the vaccinated individual. Beyond the individual protection, the indirect protection of non-immune individuals in contact with vaccinated ones also contributes to controlling pathogen spread at the population scale; a phenomenon known as herd immunity (Rose and Andraud 2017). So far, in the swine industry, vaccines have been successfully developed and used to control many diseases caused by viral and bacterial pathogens, including PRRS, swine influenza, *E. coli* diarrhea, etc. (Nan et al. 2017; Romagosa et al. 2011; Riising, Murmans, and Witvliet 2005). The main drawbacks of vaccination include the lack of robust protection across all variants of a species and the challenge introduced by vaccine to develop diagnostics allowing for discrimination of infected as opposed to vaccinated animals.

Biosecurity is a combination of management practices designed to prevent the introduction and transmission of infectious agents locally, nationally and internationally (Zimmerman et al. 2019). Good biosecurity can prevent the introduction of disease to the herd and can help prevent that disease from spreading to other herds if a disease is already present in the herd. In addition, farm animals are also a source of human food-borne pathogens, as a result, effective biosecurity strategies can prevent zoonotic disease transmission and have a broader impact on the environment and biodiversity. However, there is still biosecurity breaches. It has been reported that current biosecurity measures appear to have been generally more successful for constraining swine bacterial diseases (*Mycoplasma hyopneumoniae* being an important exception) but less effective for viral diseases in the swine industry. Thus, the most problematic issue in swine health management today is the capacity for local spread of viruses among farms (Davies 2012), especially the airborne viruses such as PRRS and influenza (VanderWaal and Deen 2018). In such cases, vaccination plays a major role for disease control, and development of effective vaccines are priority research goals.

3.2. Infectious disease surveillance, diagnostics and treatments

Diagnostic testing is widely used for disease outbreak investigation and for the routine surveillance of infectious diseases. Expanded surveillance is imperative to identify novel pathogens/strains of potential importance and to understand disease transmission within species as well as between species to prevent events like the 2009 pandemic A/H1N1 (Berger et al. 2018). Once outbreaks happen, rapid, accurate identification of infectious agents is necessary for expediting implementation of appropriate control measures. In addition, diagnostic testing can aid in investigation of vaccination failure and

biosecurity breaches for more effective infectious disease management. Previously, infectious disease reporting was often spotty and delayed, due to the lack of diagnostic facilities and the long turnaround time of diagnostic methods. In recent years, there have been major improvements both in diagnostic technology and in digital surveillance abilities, greatly enhancing pathogen identification, information sharing and timely disease control (Morse 2012).

We are in the midst of a transformation where microbiological diagnostics are based on a combination of old and new technologies, that include traditional pathogen culture methods, more recent well-developed molecular diagnostic methods such as PCR, and under-developed technologies such as sequencing. Culture of organisms is a 150-year-old standard for contemporary clinical microbiology. The development of molecular diagnostic methods in recent years, that can detect unique nucleic acid or biochemical composition, improved timeliness and resolution of infectious disease investigation (Chan et al. 2010). Today, PCR is the gold standard for sensitive and specific detection of viral and bacterial pathogens in clinical samples. Nucleic acid sequencing is an emerging powerful tool that adds discriminatory power to veterinary molecular diagnostics. The features of sequencing complements detection by "gold standard" PCR by further confirming the pathogen of interest, providing strain information and functional genotyping. Sequencing has the detection resolution at the strain-level, which can be used to differentiate vaccine from wild strains. Sequencing has also been used to type bacterial and viral isolates, such as detecting virulence, antimicrobial resistance, sequence types, or strains (Han et al. 2019; Keller et al. 2018; Estrada et al. 2019).

In the swine industry, the use of antibiotics is common for disease prevention and treatment, as well as for improvement of feed efficiency and promotion of animal growth (Lekagul 2019; Cromwell 2002). The benefits of using antibiotics to improve production is undeniable. However, the emergence of AMR is a growing concern to public health. It is estimated that by 2050, 10 million human lives a year may be lost to AMR, and the cumulative cost in terms of lost global production would be an enormous 100 trillion USD if no actions are taken (Resistance 2016; Sugden, Kelly, and Davies 2016). To address this issue, it is important to enhance alternative methods for disease control, such as good biosecurity and effective vaccination. If antibiotics are actually needed, definitive diagnosis and comprehensive assessment should be performed to guide prudent and targeted antimicrobial use (Sugden, Kelly, and Davies 2016).

4. Application of sequencing technologies to combat swine infectious diseases

4.1. Review of sequencing technologies

So far, there are numerous sequencing platforms, which have been well summarized (Maljkovic et al. 2019). The first generation sequencing method, which appeared in the mid-1970s represented by Sanger sequencing, was a real revolution, since it brought molecular diagnostics to the nucleic acid level (Sanger, Nicklen, and Coulson 1977). From the 1970s to the early 2000s, Sanger sequencing remained the predominant DNA sequencing method. Until now, Sanger sequencing is still being widely used in the swine industry for pathogen investigation due to its low cost, ease of use, and solid validation. The major disadvantage of the first generation sequencing is it often requires prior knowledge of the target genome for specific primer design and template amplification, thus it can only be used when the pathogen causing disease is known. Since one sequencing

reaction covers less than 1000 bp of the genome, several specific primers are often needed if the whole genome is desired.

The second generation of sequencing methods were developed in the 2000s. Many sequencing platforms were developed during this time, including Roche 454 sequencing, Hiseq and Miseq from Illumina, and Ion Torrent platforms from Life Technologies Co., as well as others (Goodwin, McPherson, and McCombie 2016). The main advantages of the second generation sequencing methods include high throughput and unbiased detection, which makes detection of multiple agents and discovery of novel microorganisms in a sample possible (Goodwin, McPherson, and McCombie 2016). The third generation Oxford Nanopore MinION sequencing method recently emerged as a good candidate for infectious disease diagnosis. The unique features of MinION sequencing include low infrastructure cost, high accessibility and portability, easy operation, real-time data streaming, direct DNA/RNA strand detection, and long sequence read generation. While the third generation sequencing technologies still suffer from high error rates, it has been used successfully for outbreak investigations (Hoenen et al. 2016; Faria et al. 2016) and infectious agent genome characterizations (Ashton et al. 2015). Research has also been done to benchmark MinION sequencing as a routine diagnostic tool against standard methods (Tan, Dvorak, and Murtaugh 2019). Of note, MinION sequencing can not only sequence DNA, it can sequence RNA in its native form (Garalde et al. 2018). This makes it a good fit for RNA viral infectious diseases investigation (Tan, Dvorak, and Murtaugh 2019; Keller et al. 2018). This is significant as RNA viruses have been shown to cause the most emerging disease outbreaks due to their high mutation rates (Carrasco-Hernandez et al. 2017).

All generations of sequencing have been used for swine infectious diseases and have demonstrated its great power to impact our understanding and more effectively control swine infectious diseases. Below we demonstrate some important fields where sequencing has been used for disease diagnosis.

4.2. Sequencing technologies aid in different areas of infectious disease diagnosis

4.2.1. Sequencing improved diagnostic precision and accuracy

Antonie van Leeuwenhoek, the "father of the microscope," changed our perspective on infectious diseases when he enabled the visualization of the microbial world. Since then, the goal of infectious disease diagnostics has been to detect the infectious agents responsible for an infection and enable an appropriate response (Caliendo, Gilbert, Ginocchio, Hanson, May, Quinn, Tenover, Alland, Blaschke, and Bonomo 2013). Over the years, the resolution of diagnostics and our understanding of infectious agents has been improving gradually. The microscope and culture-based methods have been used for centuries. However, there is a risk of missing the organism in direct observation, especially for viral diseases because of the size of virus particles (Mongan, Tuda, and Runtuwene 2019), not to mention that there are many viruses and other microorganisms that are unculturable. In recent years, molecular techniques have brought great advances in infectious disease diagnostics (Jannes and De Vos 2006). Today PCR is considered as a milestone in diagnostic history and a gold standard in diagnostic labs due to its short turnaround time, high detection accuracy and sensitivity (Caliendo, Gilbert, Ginocchio, Hanson, May, Quinn, Tenover, Alland, Blaschke, and Bonomo 2013).

Sequencing is a process of determining the order of nucleotides in DNA/RNA. Nearly all infectious agents contain DNA or RNA genomes, making sequencing an

attractive approach for pathogen detection. Sequencing has improved the resolution of diagnostics from detecting the species down to detection of the strain. The strain-level detections allow swine veterinarians to troubleshoot biosecurity breaches, to determine whether vaccination or treatment was successful, and to identify the emergence of new strains. Take PRRSV as an example, the PRRSV vaccine is widely used for PRRS control. However, the PRRSV vaccine is not completely effective at preventing and controlling infection due to the high genetic diversity of the virus, thus outbreaks still occur in vaccinated herds (Murtaugh and Genzow 2011; Opriessnig et al. 2002; Wang, Marthaler, et al. 2015; Zhou et al. 2017). A diagnostic method which can provide genetic information about the strain causing an infection would potentially allow for the determination of vaccination failures (Kim et al. 2015; Storgaard, Oleksiewicz, and Botner 1999). The reoccurrence of PRRS outbreaks in a swine farm is common, which could be caused by new strains or resident strains. Strain-level detection can also differentiate these two scenarios and thus guide control strategies. Using PRRSV as an example, we demonstrated that sequencing can provide higher detection resolution and thus can allow for more clear understanding of outbreaks and guide more effective interventive strategies. In addition to strain-level detection, sequencing can go further down to new variant detection (Gwinn, MacCannell, and Khabbaz 2017b) and viral quasispecies studies (Domingo, Sheldon, and Perales 2012), which gives an unprecedent ultra-precise diagnosis. Sequencing has indeed brought precision medicine into infectious disease diagnosis by its power to reveal the genetic code of microorganisms.

Sequencing can support other diagnostic methods. It can provide more information to confirm the diagnostic results from other tests, making diagnostic results more accurate.

In addition to the extra confirmation provided by sequencing, other molecular diagnostics (such as PCR) can be modified in light of pathogen sequence information generated during an outbreak. For example, during the 2013–2016 Ebola epidemic, rapidly generated viral genome sequences from a portable sequencer were used to update PCR-based diagnostics, allowing for accurate pathogen detection (Sozhamannan et al. 2015).

4.2.2. Sequencing is an all-inclusive diagnostic test that can reveal a lot of functional information in a single assay

Identification of causative agents of infectious disease is the primary goal of clinical microbiology laboratories. Beyond that, additional information about a pathogen, including sequencing types, AMR profiles, virulence, etc., can be desirable to provide insights on pathogenic mechanisms and guide effective control strategies (Roe et al. 2019; Thomas et al. 2017; Collineau et al. 2019). Conventionally, typing of pathogens mainly relies on well-established separate methods (Ribot et al. 2006). While informative, these methods only capture a small subset of the total genomic information during a test. If more information is desired, multiple tests are needed to be run which makes the procedure laborious, time-consuming, and expensive. Sequencing can provide complete genetic information to reveal multiple functional indications in a single assay (Joensen et al. 2014). Sequencing-based genotyping, therefore, could simplify workflow and eliminate the need for multiple individual assays.

For example, using MinION sequencing the MLST and AMR of *S. suis* can be identified in a single assay (Tan et al. 2019). After sequencing, MLST's were determined bioinformatically, and AMR genes were identified using an RGI online tool, both of which were able to be determined in a few hours after obtaining sequence (Tan et al. 2019). By

comparison, MLST determination conventionally involves PCR amplification of targeted regions followed by DNA sequencing, which usually takes days to get a result. AMR can be identified through antimicrobial susceptibility testing, which includes microorganism cultures and then antimicrobial susceptibility pattern identification by exposing them to different type of antimicrobials, a long and laborious process (Mongan, Tuda, and Runtuwene 2019). Of note, long-read sequencing, such as MinION, provides new solutions to assist in sequencing repetitive regions, such as horizontally acquired virulence genes or antibiotic resistance genes that are difficult to detect using short read sequencing due to difficulty assembling repetitive or duplicated regions using short reads (Ashton et al. 2015).

4.2.3. The feature of hypothesis-free detection allows sequencing to identify co-infections and discover new pathogens/variants

Hypothesis-driven molecular testing such as PCR can involve numerous individual tests for targeted organisms but may still miss a novel pathogen or strain, which decreases the breadth and sensitivity of a detection. A hypothesis-free diagnostic approach, that has the potential to detect all pathogens present in a sample, is helpful to be open about all possible causes rather than just focused on the common ones. Metagenomic sequencing allows screening of most if not all organisms and genetic composition in parallel, providing insight into microbial community biodiversity and function (Wang et al. 2013). The metagenomic sequencing approach is a hypothesis-free approach, because it doesn't target particular pathogens but equally applies to the detection of all pathogens in a sample. Metagenomic sequencing has been successfully used in the investigation of microbiome communities or co-infections, which not only included multiple species, but also included

different genotypes (Machado et al. 2019) and strain detection (Salipante et al. 2015) due to its high differentiation resolution.

Respiratory and enteric diseases in swine are often caused by multiple pathogens and metagenomic sequencing is a useful tool to deal with such challenging situations where hypothesis-driven methods fail. For example, porcine respiratory disease complex (PRDC) is a common piglet disease and causes substantive economic losses in pig farming. One feature of PRDC is its association with co-infections containing multiple pathogens. The metagenomic sequencing study of PRDC indicated that sequencing can not only identify a greater number of species related to PRDC but also improve the detection resolution by differentiating different genotypes of one pathogen (Qin et al. 2018). The robustness of the sequencing approach is important since disease situations differ significantly among farms, regions and countries. Thus, sequencing as a diagnostic method for PRDC will provide the most appropriate treatment and control regimens to individual farms (Qin et al. 2018).

In addition to simultaneous multiple pathogen detection, metagenomic sequencing offers a solution to detect unexpected pathogens (Gu, Miller, and Chiu 2019; Al Rwahnih et al. 2018). For example, in a study of porcine viral enteric disease complexes, porcine kobuvirus was identified for the first time as an important enteric virus (Theuns et al. 2018). Emerging disease, that is caused by newly discovered pathogens or existing pathogens that extend into a new geographical region, always come unexpectedly. When dealing with emerging diseases, the delayed response can lead to disease spread, which highlights the gaps of current infectious disease surveillance systems and the need for prompt identification of infectious agents (Gardy and Loman 2018). It is important to improve the ability for rapid detection of emerging pathogens, since increasing globalization,

agricultural intensification, and climatic changes have caused more frequent occurrence of emerging infectious diseases. Metagenomic sequencing can fill existing gaps and serve as a valuable diagnostic method for the swine industry.

4.2.4. Pathogen genome sequencing provides and understanding of pathogen evolution and transmission

Variability is not actually caused by man; he only unintentionally exposes organic beings to new conditions of life, and then nature acts on the organization and causes it to vary.

— Charles Darwin

As pathogens replicate and spread, their genomes accumulate mutations. Microorganisms have relatively higher mutation rates than their host. RNA viruses especially accumulate changes in their genomes with a typical rate of 0.0005 to 0.005 changes per site per year (Duffy, Shackelton, and Holmes 2008). Sequencing can be used to study pathogen evolution dynamics. In the past, due to technology limitations, gene targets were selected to study disease transmission and evolution, such as the ORF5 (GP5) gene for PRRSV, HA gene for influenza A virus, ORF2 (capsid) gene for porcine circovirus 2, and spike protein gene for porcine epidemic diarrhea virus (PEDV). These changes can now be detected via inexpensive and rapid whole-genome sequencing (WGS) on an unprecedented scale. Sequencing has been used successfully to understand the dynamics of the evolution process which includes genetic changes in terms of virulence, pathogenicity, and drug resistance (Bryant, Chewapreecha, and Bentley 2012).

An infectious disease is caused by the invasion of an infectious agent to a host causing damage to the host, followed by transmission of the infectious agent to other hosts

(Bethesda 2007). Disease transmission includes any mechanism by which an infectious agent is spread from an infected host to a susceptible host. Preventing or blocking transmission of infectious disease is one of the cornerstones of veterinary medicine. The ability of microbial genome sequencing to identify traceable differences in the pathogen genome within a short turnaround time and the application of transmission or phylogenetic tree analysis improves the epidemiological understanding of the temporal and spatial spread of infectious diseases (Neher and Bedford 2018; Kao et al. 2014). In fact, sequencing data plus epidemiological data is a rapidly evolving field and is referred to as "genomic epidemiology" (Gardy and Loman 2018; Ladner et al. 2019). With the development of sequencing technologies and their decreasing cost, whole genome sequences of pathogens can be easily obtained in near real time, which can provide timely information about pathogen transmission and prevent disease outbreaks at early stages of infection (Gardy and Loman 2018; Kan et al. 2018). By better understanding pathogen transmission and evolution, we are potentially able to predict the future, and effective prevention and control methods can be developed to combat infectious diseases.

4.2.5. Sequencing and other bioinformatics prediction tools provide support for vaccine development

Vaccination is one of the most effective disease prevention strategies. Especially, given the inevitable emergence and increasing prevalence of antibiotic resistance and ineffective biosecurity for all pathogen prevention, vaccines are a valuable strategy to prevent infectious diseases (Jansen, Knirsch, and Anderson 2018). So far, great progress has been made in the area of vaccines to prevent human diseases (such as influenza, hepatitis B, and measles) and animal disease (such as PRRS, PCV2). However, vaccination

is not always successful, the failure of which is mostly due to the variability in pathogen genomes and the poor understanding of immune escape (Luciani, Bull, and Lloyd 2012). Sequencing has been helpful for effective vaccination in many ways that include sequencing pathogen genomes, identifying host immune cell responses, determining pathogen-host interactions, etc. (Luciani, Bull, and Lloyd 2012; Dhiman, Smith, and Poland 2009).

Conventionally in veterinary medicine, vaccines are mostly modified live vaccines and inactivated vaccines. The former has the drawback of potential residual virulence and reversion to pathogenic wild types and the latter has the drawback of less protection power (Meeusen et al. 2007). More recent vaccine types include subunit, genetically engineered, live viral vector vaccine, etc. Emerging and rapidly mutating viral pathogens are a particular challenge for effective vaccine development since their genomes can change rapidly and vaccines can lack cross-protection. Sequencing provides a promising solution for pathogens that are highly diverse (such as PRRSV, influenza virus) and newly emerged (such as PEDV). When a pathogen has high genetic diversity, cross-protection is important. Genomics can be a useful tool for guiding vaccine development, since it allows us to monitor genetic changes over time and the genetic diversity in pathogen populations ('Pathogen genomics leading to vaccines' 2019). For example, given the degree of genetic diversity observed among the PRRSV strains worldwide, the current vaccines, which are all based on one selected virus strain for each commercial vaccine, will unlikely provide a sufficient level of protection against the heterologous field strains (Kim et al. 2015). Sequencing can examine the genomic variations in pathogens and identify new potential candidate genes and antigens to design a universal vaccine that can provide broader

protection against multiple strains of the pathogen (Erbelding et al. 2018; Brockmeier et al. 2018). For pathogens that have no efficacious vaccines yet, sequencing can provide insight on potential protein subunits as vaccine candidates and assist future vaccine development (Brockmeier et al. 2018; Tuanthap et al. 2019). Global genomic approaches to screen for potential vaccine candidates can be a powerful way to increase the chance of success to protect swine populations. In addition, sequencing can be used for vaccine contaminant detection (Onions et al. 2011), and can be used to investigate vaccine failure.

4.2.6. A portable sequencer is a possible solution for the uneven resource distribution issue

The uneven distribution of health resources within countries and communities has been noticed long ago (Akhtar and Izhar 1986; Nishiura et al. 2004). Research showed that while emerging disease hotspots are more concentrated in lower-latitude developing countries, there is more surveillance and investigation resources in richer, developed countries, the phenomenon of which is known as "the inverse care law" (Jones et al. 2008).

In the swine industry, it has been estimated that the most vulnerable scenario may be the rapid expansion and intensification of swine industries in developing countries where there is a lack of health resources (Davies 2012). For example, it has been identified that regions of Southeast Asia, Eastern Europe, Central America, and sub-Saharan Africa have high potential for influenza virus spillover, however, it is found that 78.43% of high-risk areas lacked evidence of influenza virus surveillance (Berger et al. 2018). It is also estimated that the next emerging swine infectious disease, possibly zoonotic disease, will most likely happened in developing countries, for there is rapid expansion and intensification of swine industries but a lack of basic biosecurity measures and diagnostic facilities (Davies 2012). Because of globalization and frequent trade of animals and animal

products, infectious disease originating from one place can transmit rapidly throughout the world. This indicates that for better control of global infectious disease more scientific resources are needed in places where emerging infectious diseases are more likely to originate. The dilemma is that infectious surveillance depends on diagnostic facilities, which are lacking in resource-limited areas where emerging infectious diseases most often occur. While the construction of medical infrastructure is not a simple task, it is promising that a number of molecular technologies are adaptable to resource limited settings for disease surveillance and can facilitate the decentralization of technology (Lipkin 2010).

Portable diagnostic devices are a promising tool to solve this dilemma, since they can diagnose infectious diseases in places that have a limited laboratory capacity. The MinION, from Oxford Nanopore Technologies, is a pocket-sized sequencer (Jain et al. 2016). The portable sequencer can be shipped and operated under rough field conditions to provide rapid, unbiased pathogen detection. It has been used for on-site outbreak investigations in remote areas such as during the Ebola epidemic in West Africa (Quick et al. 2016). In addition to single pathogen detection, it has been used for biodiversity investigation in a remote tropical forest (Pomerantz et al. 2018). Especially, as the development of additional equipment such as Voltrax (an add-on for the flowcell which will automate library preparation) or MinIT (a GPU computer) may eliminate the need for an extraction kit or laptop computer, and even make it more suitable for field work (Mongan, Tuda, and Runtuwene 2019).

4.3. Limitations of sequencing

Sequencing provides opportunities for infectious disease control, but limitations do exist.

- **Cost.** Although the cost of sequencing has dropped, it is still relatively costly compared to other conventional methods. It is also still a technology that requires structured laboratories and equipment.
- **Resources.** The use of sequencing for disease investigation is making bioinformatics an integral part of clinical diagnostics. There is a great need for experts with the right kind of skills and experience, as well as computational and IT infrastructure to support the analyses of large amounts of data produced by high throughput sequencing.
- **Technical obstacles.** When using metagenomic sequencing for unbiased pathogen detection from clinical samples, most DNA/RNA would be host derived and the microbial nucleic acids of interest are a minority of the sample (Yamagishi, Runtuwene, Hayashida, Mongan, Thi, Thuy, Nhat, Limkittikul, Sirivichayakul, and Sathirapongsasuti 2017). This poses a significant needle-in-a-haystack challenge for detecting pathogens from metagenomic data. Methods for host depletion and target enrichment have been developed (Thoendel et al. 2016; Charalampous et al. 2018), however, more efforts are needed to improve the efficiency.
- **Interpretation and validation.** Whole genome sequencing can provide an understanding of a diseases' functions. Since the genotype and phenotype are still not well understood and databases are incomplete, interpretation of sequencing results can be biased (Gu, Miller, and Chiu 2019). Validation and standardization of sequencing technology is greatly needed. There is no standard method for interpreting sequencing results.

5. Future directions to improve global swine health

Currently, infectious diseases are still not under control in the swine industry and are continuous threats to swine health. Undoubtedly, infectious diseases are becoming better controlled nowadays due to the improvement of technologies and control strategies ('Pathogen genomics leading to vaccines' 2019). However, infectious pathogens are also evolving, and new pathogens keep emerging. The battle against emerging infectious diseases is an ongoing process (Morens and Fauci 2013; Morens, Folkers, and Fauci 2004).

Concentration of pig production makes control or eradication of some diseases more difficult. At the same time, the transportation and movement of pigs across regions and countries has provided opportunities for global disease transmission. Thus, collaboration between regions and countries is needed to improve animal health. Animal health is also closely related to human health, since pathogens as well as AMR can jump among animals and humans. A "One Health" approach is a trend to deal with infectious diseases. New technologies, such as sequencing, can be harnessed as a powerful support for infectious disease control. The sequencing data, however, creates more challenges in terms of sequencing analysis, sequencing platforms, data sharing and concerns over data safety (Moscoso et al. 2019) that need to be addressed.

5.1. Big data brings opportunities and challenges

Advances in genomics are transforming our approach to understanding swine disease diagnosis (Pak and Kasarskis 2015). As high throughput sequencing methods are applied more on veterinary medicine, massive amount of data ("big data") was generated. Big data in the field of veterinary medicine has been mostly focused on spatial and temporal epidemiological analysis (VanderWaal et al. 2017; Kao et al. 2014). Recent reports have

called for the integration of genomic data with epidemiology streams (Flahault et al. 2017; Gardy and Loman 2018). The epidemiological potential of this pathogen genomic surveillance system is profound. Sequencing can provide detailed information about pathogens and disease surveillance can help us understand global distribution and transmission of the vast majority of infectious diseases (Hay et al. 2013).

The development of technologies and the "big data" era provide new opportunities to respond rapidly to infectious diseases and improve animal health. With rapidly increasing volumes of sequence data, cloud-based platforms are needed for data storage and sharing, along with experts in database management. It is a challenge to define and extract the appropriate information from the large amounts of sequence data that is generated. Thus, to facilitate the use of whole genome sequencing data and benefit the most from big data, experts with genetic knowledge and bioinformatics analysis skills are greatly needed.

5.2. The "One Health" approach is needed to improve public health

Rising zoonotic diseases (Ang, Lim, and Wang 2018; Al-Muharrmi 2010) and the AMR concern (Hernando-Amado et al. 2019) requires integrative and cross-disciplinary efforts, to achieve optimum health for people and animals. "One Health" recognizes that the health of animals is connected to the health of people, and that infectious disease can cross boundaries. The "One Health" concept aims to raise awareness of interdisciplinary, and inter-institutional collaborative efforts to tackle current infectious diseases (Dixon, Dar, and Heymann 2014; Uddin Khan et al. 2013). It is estimated that 6 out of every 10 infectious diseases in humans are spread from animals (Jones et al. 2008). Because the microbiomes of humans, animals, plants, water and soils are interconnected and most

ecosystems contribute to the emergence, acquisition and spread of AMR. AMR can no longer be addressed by simply studying the problem within a single system (Hernando-Amado et al. 2019). In general, a closer cooperation between human and veterinary medicine is urgently needed (Houe et al. 2019). Actions have already been made under the guidance of "One Health", such as One Health surveillance systems (Stärk et al. 2015). Responses to Hendra virus outbreaks served as a successful story of using a One Health practice to effectively fight against zoonotic diseases (Wang and Crameri 2014).

The world is increasingly connected, which allows for the express movement of goods, people and infectious diseases across the globe. In the swine industry, there are currently significant international-scale infectious disease outbreaks, such as PRRSV, PEDV, and ASF. This not only requires advanced diagnostic methods, such as sequencing, for effective disease management, but also needs increased communication and collaboration among regions and countries across a variety of sectors including human, agriculture and the environment (Kluberg et al. 2016; Sugden, Kelly, and Davies 2016) (Chan et al. 2010).

Chapter II - MinION sequencing of *Streptococcus suis* allows for functional characterization of bacteria by multilocus sequence typing and antimicrobial resistance profiling

A manuscript published in *Journal of Microbiological Methods*

Tan, S., C. M. T. Dvorak, A. Estrada, C. Gebhart, D. G. Marthaler, and M. P. Murtaugh. 2019. 'MinION sequencing of *Streptococcus suis* allows for functional characterization of bacteria by multilocus sequence typing and antimicrobial resistance profiling', *J Microbiol Methods*: 105817.

Author Contributions: MPM conceived the idea. MPM and CMTD supervised and administered the project. ST conducted all the research, and was responsible for experiments, analysis, visualization and preparing the original draft. CMTD revised the manuscript. AE provided *S. suis* isolates and Illumina raw sequencing data. DGM and CG provided important suggestions for *S. suis* MLST and AMR analysis. Funding acquisition: MPM, CMTD and ST. All authors have edited and commented the manuscript. All authors have read and approved the final version of the manuscript.

In recent years, high-throughput sequencing has revolutionized disease diagnosis by its powerful ability to provide high resolution genomic information. The Oxford Nanopore MinION sequencer has unparalleled potential as a rapid disease diagnostic tool due to its high mobility, accessibility, and short turnaround time. However, there is lack of a rigorous quality assessment and control processes standardizing the testing on the MinION, which is necessary for incorporation into a diagnostic workflow. Thus, our study examined the use of the MinION sequencer for bacterial whole genome generation and characterization. Using *Streptococcus suis* as a model, we optimized DNA isolation and treatments to be used for MinION sequencing and standardized de novo assembly to quickly generate a full-length consensus sequence achieving a 99.4% average accuracy. The consensus genomes from MinION sequencing were able to accurately predict the multilocus sequence type in 8 out of 10 samples and identified antimicrobial resistance profiles for 100% of the samples, despite the concern of a high error rate. The inability to unequivocally predict sequence types was due to difficulty in differentiating high identity alleles, which was overcome by applying additional error correction methods to increase consensus accuracy. This manuscript provides methods for the use of MinION sequencing for identification of S. suis genome sequence, sequence type, and antibiotic resistance profile that can be used as a framework for identification and classification of other pathogens.

1. Introduction

Precise identification of specific pathogenic organisms is one of the crucial goals for public health. Ideal diagnostics would be able to provide a rapid diagnosis while

creating a stream of genomic data revealing critical functional information to aid in precise treatments. Although the resolution of diagnostic technologies has been improving from the cell level down to the molecular level, there are still unmet needs.

For bacterial pathogens, disease control relies heavily on the use of antibiotic drugs, but the rise of antimicrobial resistance (AMR) presents a global threat to public health due to the high use of antibiotics (Centner 2016). To effectively prevent the potential rise of antibiotic-resistant bacteria, the FDA released a new policy in 2018 for the use of medically important antimicrobials in livestock that requires a rigorous evaluation before therapeutic use of these drugs (U.S. Food and Drug Administration, https://www.fda.gov/AnimalVeterinary/SafetyHealth/AntimicrobialResistance, 7/1/2019). Therefore, there is now a pressing need for identification of AMR to guide in effective antimicrobial usage. Sequencing technologies have emerged as a potential candidate to fill this unmet need by providing genetic information without any prior knowledge of possible pathogen presence, allowing for accelerated pathogen detection as well as providing valuable information on strain discrimination, and genome functions including antibiotic resistance profiles (Lecuit and Eloit 2014).

Currently, short-read next generation sequencing, such as the Illumina platform (Illumina, Inc.), is the dominant method for obtaining genomic sequences (Deurenberg et al. 2017). However, the fragmented short-read sequences can present serious limitations for genome assemblies and downstream analyses (Cao et al. 2017; Ricker, Qian, and Fulthorpe 2012). Long-read single-molecule sequencers, represented by Oxford Nanopore MinION (Oxford Nanopore Technologies Ltd.) and PacBio Sequel (Pacific Biosciences of California, Inc.), are able to overcome this obstacle by providing longer assemblies (Ardui

et al. 2018; Jain et al. 2018). In addition, MinION has other unique advantages such as potential on-site detection due to its small size, real-time data stream, and its universal accessibility given its relatively affordable price. This technology has been tested for genome sequencing research of bacteria, viruses, and yeast, and it has shown a great potential for infectious disease diagnosis (Giordano et al. 2017; Loman, Quick, and Simpson 2015; Wang, Moore, et al. 2015; Mitsuhashi et al. 2017). However, the shift from research to routine diagnostic applications requires rigorous quality control processes and standardization of testing. Thus, the purpose of this study was to optimize methods and provide benchmarking for bacterial detection and genome characterization using MinION sequencing.

Streptococcus suis (S. suis) is a model pathogen for optimization of MinION sequencing for bacterial detection and characterization, as *S. suis* is an important pathogen for pigs and can be transmitted to humans. Additionally, it has received growing attention for its increasing and complex antibiotic resistance (Devi et al. 2017; Yongkiettrakul et al. 2019). *S. suis* strains have been reported to encode antibiotic resistance genes against antimicrobials including tetracyclines, macrolides, β-lactams, aminoglycosides, trimethoprim–sulfamethoxazole, chloramphenicol, and fluoroquinolones (Palmieri, Varaldo, and Facinelli 2011). It is crucial to understand and determine the AMR profile to develop a better understanding of *S. suis* AMR mechanisms and importance as a reservoir of antibiotic resistance determinants. *S. suis* is also highly diverse with more than 1000 sequence types (STs) registered in the *S. suis* multilocus sequence type (MLST) database, which makes control and surveillance of this pathogen more challenging (Gauthier-Levesque et al. 2016; King et al. 2002; Segura et al. 2017). The genome size of *S. suis*

varies greatly from 1.6 to 2.1 million bases, reflecting again its high diversity (Feng et al. 2014). Establishment of a MinION workflow for *S. suis* whole genome sequencing will allow for disease surveillance, sequence typing, AMR profiling, and will lead towards the identification of resistance mechanisms. Our research has provided an optimized operational *S. suis* sequencing workflow using Oxford Nanopore MinION technology and has demonstrated its potential as a widely applicable technology for diagnostics and characterization of pathogens including MLST classification and antibiotic resistance profiling. This method can be used as a reference to develop a framework for MinION sequencing and characterization of other pathogens.

2. Materials and Methods

2.1. Bacterial strains

T7 Express Competent *Escherichia coli* (*E. coli*) (New England Biolabs, Inc) was used for evaluation of the DNA purification method. *E. coli* was grown in Luria Broth for 12h and then used for DNA extraction. Stock cultures of 10 *S. suis* strains were kindly provided by Dr. Connie Gebhart at the University of Minnesota Veterinary Diagnostic Laboratory, along with Illumina HiSeq sequencing raw reads for each sample (BioSample accession numbers: SAMN11854274, SAMN11854276, SAMN11854295, SAMN11854277, SAMN11854278, SAMN11854305, SAMN11854328, SAMN11854241, SAMN11854253, SAMN11854163).

2.2. Bacterial DNA isolation

E. coli cultures were used to compare four different genomic DNA extraction methods starting with the same amount of culture for each method. An organic method for DNA extraction using phenol, chloroform, and isoamyl alcohol was performed based on

Wright, et al with a final volume of 100 μL of DNA in nuclease-free water (Wright, Adelskov, and Greene 2017). A silica-based magnetic bead kit (Qiagen MagAttract HMW DNA kit), a silica-based membrane spin column (Qiagen DNeasy Blood & Tissue kit), and a differential precipitation kit (Qiagen Gentra Purgene Yeast/Bact. kit) were all used following the manufacturer's instructions with a final elution of DNA in 100 μL of nuclease-free water.

After the DNA isolation comparison, genomic DNA was isolated for the remainder of the experiments using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's protocol for gram-positive bacteria, replacing the enzymatic lysis buffer with 160 μl Buffer P1 (Qiagen) containing 5 μl mutanolysin (25 KU/μl, MilliporeSigma). This was followed by a 2h incubation at 37°C and a final elution in 100μl of nuclease-free water. DNA concentrations were measured using a Qubit 3.0 fluorometer (ThermoFisher Scientific) following manufacturer's protocols. DNA purity was measured using a Nanodrop ND-1000 spectrophotometer (ThermoFisher Scientific) to examine the OD 260/280 and OD 260/230 ratios. DNA integrity was determined using an Agilent 4200 TapeStation (Agilent Technologies).

To improve samples with poor DNA purity, drop dialysis of DNA samples was performed by floating a 0.025 μm, mixed cellulose ester, MF-Millipore membrane filter (MilliporeSigma) on the surface of sterile water in a petri dish, adding the DNA sample onto the membrane, and incubating for 1 hour at room temperature. The DNA sample was then recovered by pipetting it off the membrane and was measured using a NanoDrop ND-1000 spectrophotometer (ThermoFisher Scientific) to examine the DNA concentration, as well as the purity of the DNA by OD 260/280 and OD 260/230 ratios.

2.3. DNA shearing and size selection

Four combinations of DNA size selection and shearing methods were evaluated using three *S. suis* samples and sequence results from the four treatments were compared. The four combination groups were gDNA (genomic DNA with no size selection or shearing), gtube (sheared genomic DNA using g-TUBE), BlueP (size selected genomic DNA using BluePippin), and gtube/BlueP (sheared, size selected genomic DNA using g-TUBE followed by BluePippin).

DNA was sheared using a g-TUBE (Covaris, Inc.) according to manufacturer's protocols targeting 18kb fragments as determined by mass of DNA and centrifuge speed. A total of 12ug of genomic DNA was centrifuged at 5800 rpm for 1 minute to obtain fragments of approximately 18kb. DNA was size selected using BluePippin (Sage Science) following manufacturer's protocols targeting fragments >6kb. Briefly, a 6000 to 50,000 bp size range for DNA was selected in the program and 2 lanes of 5ug of each sample were loaded into a 0.75% agarose cassette. DNA fragment size following treatments was examined using an Agilent 4200 TapeStation (Agilent Technologies) following manufacturer's instructions.

2.4. MinION library preparation and sequencing

A MinION Mk I sequencer and flow cell versions R9.4.1, R9.5 and R9.5.1 were used for this study (Oxford Nanopore Technologies Ltd). An R9.4.1 flow cell, the Ligation sequencing kit and the rapid barcoding kit were used for sequencing to compare DNA treatments (Oxford Nanopore Technologies Ltd). An R9.5.1 flow cell and the $1D^2$ sequencing kit were used for optimizing de novo assembly following the manufacturer's instructions (Oxford Nanopore Technologies Ltd). Sequencing results from all versions of

flow cells and sequencing kits were used for comparisons with Illumina sequencing for de novo assembly and genome characterization.

The input DNA quantity for library preparation was calculated using the average fragment size and a targeted molarity of 0.3 pmol. Library preparation was performed using the 1D or $1D^2$ Ligation sequencing kits (Oxford Nanopore Technologies Ltd) following manufacturer's protocols. Briefly, DNA were end-repaired using the NEBNext® End Repair Module (New England Biolabs), and subsequently cleaned up by AMPure XP beads (Beckman Coulter). Then, Blunt/TA Ligase Master Mix (NEB) was added to ligate adapters and purified using AMPure XP beads. Prepared DNA libraries were then added to the flow cells in the MinION (Oxford Nanopore Technologies Ltd), the MinION was connected to the MinKNOW software (Oxford Nanopore Technologies Ltd), and sequencing was initiated and run for the desired amount of time.

2.5. MinION sequence analysis

MinION raw sequencing reads were basecalled using Albacore (version 2.3.1, Oxford Nanopore Technologies Ltd) with a default quality filter of 7 to determine pass reads. The total raw read quality was examined using MinIONQC (Lanfear et al. 2019) to determine yields, mean read length, and mean quality. MinION raw reads, for all 10 isolates have been deposited in the NCBI Sequence Read Archive (SRA): SRR10240057 to SRR10240066.

De novo assembly of MinION reads was performed using Canu (version1.6) with default parameters, which has 3 main steps; correct, trim, and assemble (Koren et al. 2017). Quast (version 4.5) was used to evaluate assembly quality, including the number of contigs and N50 (Gurevich et al. 2013). Alignment of the MinION raw fastq file from basecall to

the draft genome from Illumina sequencing was performed using Graphmap (version 0.5.2) (Sovic et al. 2016). AlignQC (version 2.0.5) was used to evaluate the alignment and to determine the single read error rates (Weirather et al. 2017). Qualimap (version2.2.1) was used to characterize the coverage distribution of the alignment across the whole genome (Garcia-Alcalde et al. 2012).

All bioinformatics methods and source code used in this manuscript are also provided in the public GitHub repository at github.com/ShaoyuanTan/ssuisproject

2.6. De novo assembly optimization

Optimization of de novo assembly was performed by examining different fold coverage levels, adjusting read quality filters, and applying different error correction methods. Different fold coverage subsets (5X, 10X, 20X, 40X, 60X, 80X, 100X, 200X, and 400X) were obtained by randomly selecting reads from a single sequencing dataset using Seqtk (github.com/lh3/seqtk) and fastq-tools (github.com/dcjones/fastq-tools). Different quality filter cut-off subsets from the same dataset were generated using R (version 3.4.0), followed by generation of 25X, 50X, and 100X fold coverage sets for each quality filter subset (Team 2016). De novo assembly was then performed on the subsets using Canu. Consensus accuracy and assembly quality of each were evaluated using Mummer (version 4.0.0beta2) (Marcais et al. 2018) and Quast to determine the optimal cutoffs.

MinION consensus sequences generated using Canu have been deposited at NCBI under BioProject PRJNA545328. This Whole Genome project has been deposited at GenBank under the accession VIEM00000000 to VIEV00000000. The version described in this paper is version VIEM01000000 to VIEV01000000. For AMR gene analysis, Canu

parameters were modified by adding corOutCoverage=1000 to allow for generation of plasmids before examination of antimicrobial resistance (AMR) genes. MinION consensus sequences generated using the modified Canu parameters that gave different AMR profiles from original consensus sequences have been deposited at NCBI under BioProject PRJNA547797. This Whole Genome project has been deposited at GenBank under the accession VIFG00000000 and VIFH00000000. The version described in this paper is version VIFG01000000 and VIFH01000000.

Improvement of the assembly consensus accuracy from 5X to 100X coverage was performed using 3 different open source software programs (racon, nanopolish and Pilon). Racon (version 1.3.1) is a consensus module that was used to correct draft genome assemblies using MinION raw reads (Vaser et al. 2017). Nanopolish (version 0.10.1) exploits a single-level algorithm that was used to calculate an improved consensus sequence for the draft genome assembly using MinION raw reads (github.com/jts/nanopolish). MinION consensus genomes corrected using nanopolish have been deposited at NCBI under BioProject PRJNA547794. This Whole Genome project has been deposited at GenBank under the accession VIFI00000000 and VIFJ00000000. The version described in this paper is version VIFI01000000 and VIFJ01000000.

A hybrid genome was assembled to provide a long and accurate consensus sequence by combining long reads from MinION sequencing and accurate reads from Illumina sequencing for error correction using Pilon (version 1.22) (Walker et al. 2014). Hybrid consensus genomes have been deposited at NCBI under BioProject PRJNA547793. This Whole Genome project has been deposited at GenBank under the accession

VIEW00000000 to VIFF00000000. The version described in this paper is version VIEW01000000 to VIFF01000000.

2.7. Illumina sequence analysis

The S. suis Illumina HiSeq sequencing was performed by the University of Minnesota BioMedical Genomics Center and raw data was provided to us by Dr. Connie Gebhart at the University of Minnesota Veterinary Diagnostic Laboratory (Estrada et al. 2019). Raw reads from Illumina HiSeq sequencing were evaluated by FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) and then trimmed using Trimmomatic (version 0.36) (Bolger, Lohse, and Usadel 2014). *De novo* assembly was performed using SPAdes (version3.11.1) (Bankevich et al. 2012). Assembly quality was assessed using Quast. Illumina consensus genomes have been deposited at NCBI under BioProject PRJNA547781. This Whole Genome project has been deposited at GenBank under the accession VIEC00000000 to VIEL00000000. The version described in this paper is version VIEC01000000 to VIEL01000000.

In addition to the Genbank, all consensus genomes are also provided in the following figshare project: https://figshare.com/articles/Consensus_sequences_for_S_suis_project/9955211

2.8. Characterization of the S. suis genomes

The consensus genomes assembled using Oxford Nanopore MinION sequence (MinION), Illumina Hiseq sequence (Illumina), and the hybrid assembly (hybrid) were compared using the hybrid consensus as the reference genome.

Quast was used to assess all assemblies, and the identity of Illumina contigs and MinION contig(s) to the hybrid consensus were generated using Mummer (version

4.0.0beta2). The ability of each sequencing and assembly method to generate a *S. suis* consensus, predict the MLST, and identify antibiotic resistance genes were compared. MLST profiles of the *S. suis* genomes were predicted using PubMLST (pubmlst.org/). A total of 7 housekeeping genes (*aroA*, *cpn60*, *dpr*, *gki*, *mutS*, *recA*, *thrA*) for *S. suis* MLST analysis were downloaded from the PubMLST website (6/21/2018 update). Alignment of the MinION, Illumina, and hybrid assemblies to the dataset of 7 housekeeping genes was performed using BLASTn. Knowing that the consensus genome from MinION was less than 100% accurate, the highest bit score from BLASTn for each gene was recorded and used even if the identity was less than 100% and both the BLAST score and percent identity were recorded (Supplemental Table S1). The alleles with the highest BLAST score for each of the 7 genes were then used to predict the sequence type by filling out the allelic profile on PubMLST using the "exact match only" search option (Jolley, Bray, and Maiden 2018). Antimicrobial resistance (AMR) genes were characterized by the Resistance Gene Identifier (RGI) online analysis tool from the Comprehensive Antibiotic Resistance Database (CARD), which uses CARD's curated AMR detection models to predict the complete resistome from genome sequences (https://card.mcmaster.ca/analyze/rgi). AMR genes from Illumina and hybrid assemblies were detected under the strict algorithm and MinION assemblies used the loose algorithm, all using a custom cutoff of "% Identity of Matching Region" > 90. The genes detected by RGI were then confirmed using BLASTn to compare the AMR gene sequence to the MinION genome using a cutoff of alignment length/AMR gene length > 0.8 (coverage > 80%).

2.9. Sanger sequencing to confirm the new gki allele

Primers were designed for PCR amplification of the *gki* gene region using Primer3 (forward primer 5’ TTGGGTGTTGGTATGGGGT 3’, reverse primer 5’ TAGCCGTTACATCTTGCCCA 3’) (Untergasser et al. 2012). DNA was amplified using AccuStart PCR SuperMix (Quantabio). PCR was performed under standard conditions of 94°C for 3 min, 35 cycles of 94°C for 20 sec, 55°C for 20sec, and 72°C for 1.5 min, followed by a final elongation of 72°C for 5 min (Dvorak et al. 2013). PCR products were then cleaned up using ExoSAP-IT (Thermo Fisher Scientific). Sanger sequencing was performed on the PCR products at the University of Minnesota Genomics Center. Results were visualized by Unipro UGENE (Okonechnikov, Golosova, and Fursov 2012). The new allele (gki_360) and new sequence type (ST1214) were submitted to and are available at PubMLST (https://pubmlst.org/).

2.10. Statistical methods

Two-way ANOVA and Tukey's multiple comparisons test were performed using Graphpad Prism (v. 8.0.0, GraphPad Software).

3. Results

3.1. Optimization of DNA isolation

Extraction of high quality and quantity DNA is an essential preliminary step for microbial sequencing. Oxford Nanopore MinION sequencing requires high quality input DNA with a recommended OD 260/280 ratio of 1.8 and OD 260/230 ratio of 2.0–2.2. Four different DNA extraction methods were examined for isolation of *E. coli* DNA; organic, silica-based magnetic beads, silica-based membrane spin columns, and differential precipitation. Following extraction, DNA recovery and the presence of contaminants such as protein and salts were examined.

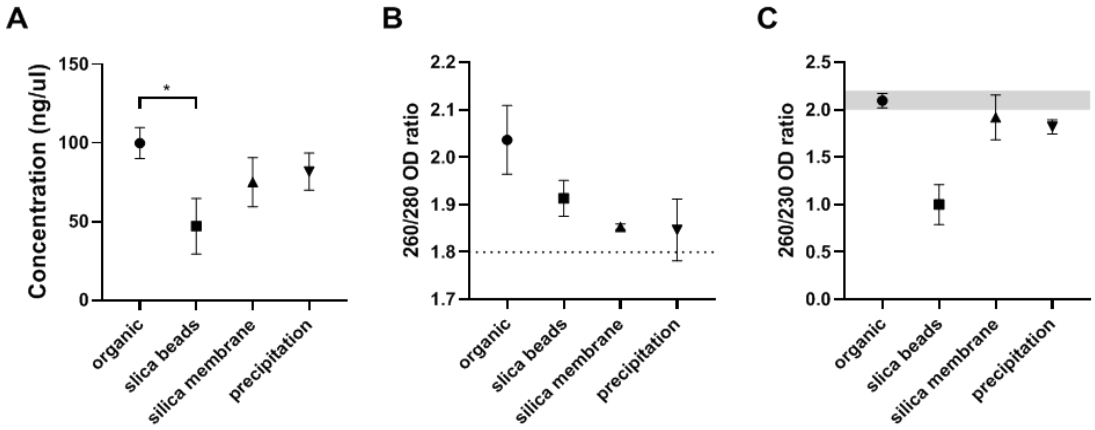

Figure 1. A comparison of DNA quantity and quality using 4 different DNA extraction methods.

The identical amount of a single *E. coli* culture was used for DNA extraction using 4 different methods (organic, silica-based magnetic beads, silica-based membrane spin column, and differential precipitation). The final output volume was identical for each method to allow for direct comparison. A) The DNA concentration was measured using Qubit (*$p<0.05$) (n=3). B) A260/280 and C) A260/230 absorbance ratios were examined using a nanodrop spectrophotometer with the ideal ratio shown as a dashed line (B) or grey bar (C) (n=3).

The highest amount of DNA was obtained using the organic extraction method, which was significantly higher than that obtained using silica beads ($p<0.05$), but not significantly different from those using the silica membrane or precipitation methods (Figure 1A). The purity of the DNA samples was examined by analyzing the 260/280

absorbance ratio, looking for an optimal ratio of 1.8, with either lower or higher ratios suggesting poor quality DNA (Figure 1B). Both the silica-membrane and precipitation methods gave 260/280 ratios around 1.8 while the silica bead method was a little higher at 1.9 and the organic method gave a ratio over 2.0, suggesting RNA contamination (Figure 1B). Another DNA purity measure is examination of the 260/230 absorbance ratio, which can be used to measure the salt contamination in the sample, with an optimal ratio being 2.0-2.2. The silica-membrane, precipitation and organic methods gave the best ratios at 1.8-2.1, and the silica-bead method had a much lower ratio of 1.0 suggesting a high salt contamination (Figure 1C). However, the DNA quality of samples with a lower OD 260/230 ratio was able to be improved using drop dialysis to remove salts (data not shown). Overall, the two optimal methods of DNA extraction were using the silica-membrane or differential precipitation methods. The silica-based membrane spin column extraction method was used for DNA extraction of *S. suis* samples for the remainder of the study.

3.2. Optimization of DNA shearing and size selection

Although the MinION sequencer has the ability to generate long reads, high molecular weight (HMW) genomic DNA is recommended to be sheared and size selected prior to MinION sequencing to theoretically optimize the sequence adapter binding efficiency and thus sequencing output. Four methodologies were examined to determine which technique resulted in optimal yield, read size, and read quality. The 4 methodologies were as follows; 1) intact HMW genomic DNA (gDNA), 2) sheared DNA using a g-TUBE to target 18kb fragments (gtube), 3) size selected DNA for fragments >6kb using BluePippin (BlueP), and 4) g-TUBE sheared DNA that was then size selected using BluePippin (gtube/BlueP). Following BluePippin treatment, DNA was treated by drop

dialysis to improve the 260/230 ratio. The DNA recovery after BluePippin treatment was only about 40%, thus a much higher amount of starting material was needed to obtain similar output DNA amounts as compared to the other treatments. The average DNA fragment size following treatment was examined (Figure 2A). Genomic DNA showed an average size of 23 kb, g-TUBE sheared DNA gave an average size of 20kb, BluePippin size selection gave an average size of 18kb, and g-TUBE sheared DNA followed by BluePippin size selection gave an average size of 17kb (Figure 2A).

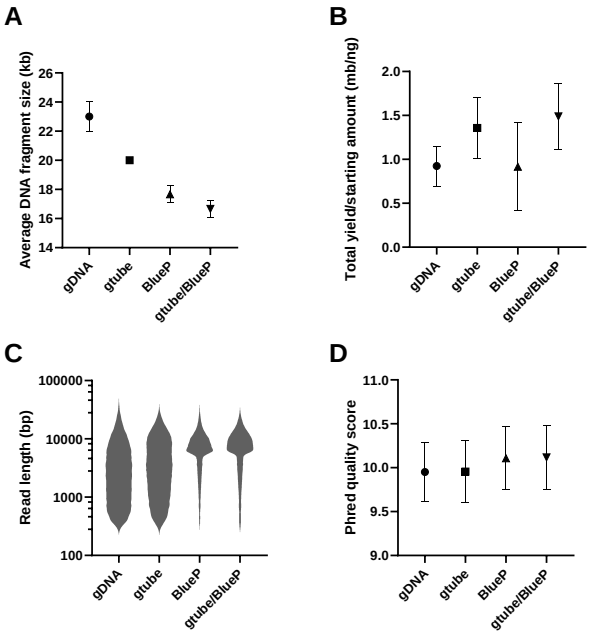

Figure 2. The effect of DNA size selection and shearing on sequencing read quality.

Four DNA size selection and shearing treatments were performed on a *S. suis* stock DNA sample. A) Following each treatment, the average DNA fragment size was determined. A total of 0.3pmol of DNA from each treatment was used for MinION library preparation and sequencing (n=3). Sequencing output was examined for B) the ratio of total sequence yield per total amount of DNA (ng) used for library preparation, C) the sequencing read size distribution, and D) the Phred quality score of the sequencing reads.

DNA from each of the different treatments (0.3 pmol each) was barcoded for MinION library preparation and sequencing. Using the same DNA molarity for all samples instead of the same quantity allowed for a reduction of the confounding variable of binding efficiency to the adapters during library preparation. Sequence outputs were then compared for yield, size, and quality of the output reads (Figure 2B-2D). The ratio of total yield (mb, million bases) over the total amount used for library preparation (ng) was highest in both of the g-tube treatment groups, but no

significant difference among groups was observed (Figure 2B). An examination of the read length distribution from the four groups showed that while BluePippin did remove many small fragments, long reads were present in all groups (Figure 2C). No significant difference in the read quality between treatments was observed and all treatments showed an average Phred quality score around 10 (Figure 2D).

The influence of DNA shearing or size selection on de novo assembly from raw reads was examined by comparing the number of contigs, N50, and consensus accuracy at 25X, 50X, and 100X coverage. The N50 is a measure of the mean contig length that gives more weight to longer contigs. De novo assembly was not significantly impacted by any of the shearing or size selection treatments (Figure 3A-3C). However, consensus accuracy was significantly different between 25X and 100X coverage in all groups, with 100X

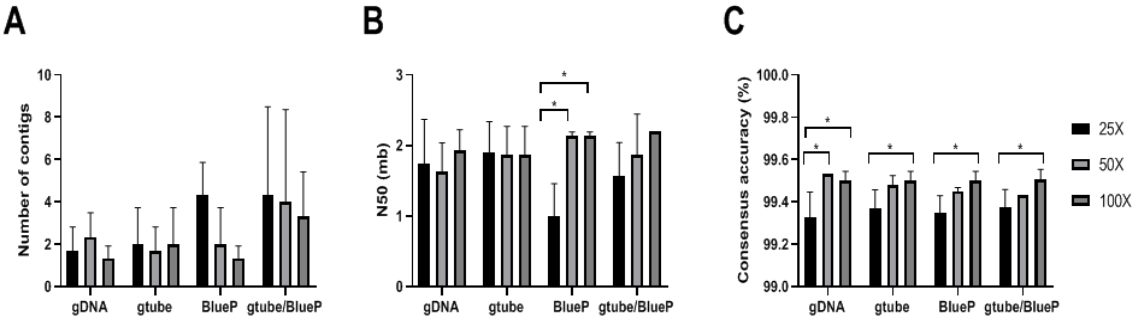

Figure 3. The effect of size selection and shearing on de novo sequence assembly.

Four DNA size selection and shearing treatments were performed on a *S. suis* stock DNA sample followed by library preparation using 0.3 pmol input DNA and MinION sequencing. De novo assembly of raw reads was performed and bioinformatic analysis of the de novo assembly from each treatment at 25X, 50X, and 100X coverage was examined (n=3). A) The number of contigs created, B) the N50 (mean contig length), and C) the consensus accuracy were compared between treatments (*$p<0.05$).

coverage giving a significantly higher consensus accuracy (Figure 3C). Overall, sequencing of genomic DNA without any DNA shearing or size selection already gave good results, but the average fragment sizes were big and required more input DNA to

obtain the same molar amount for library production. Following BluePippin treatment, there was only about a 40% recovery of the DNA, and the de novo assembly results were similar among groups, indicating this was an unnecessary step. The g-TUBE treatment had almost a 100% recovery following treatment, and since there were more DNA ends available for adapter ligation after shearing, less DNA was needed for library preparation. More reproducible sequencing results from lower amounts of starting material, regardless of the genome size of the pathogen, would be generated using g-TUBE treatment prior to library preparation. Thus, further sequencing experiments in this manuscript were performed using g-TUBE treatment after DNA extraction prior to library preparation and sequencing.

3.3. Optimization of de novo assembly

Optimization of de novo assembly was performed by examining assembly results using different fold coverage levels, different read quality filters, and by applying different error correction methods. The minimal amount of coverage necessary for optimal de novo assembly was examined by comparing consensus accuracy and contig assembly at sequence coverages from 5X-400X. Between 5X to 20X coverage, consensus accuracy and N50 increased and the number of contigs dropped considerably creating as few contigs as possible (Figure 4A-4C). From 20-100X coverage relatively small variations were observed in consensus accuracy, N50, and the number of contigs and they were near

optimal by 100X coverage (Figure 4A-4C). Thus, 20-40X coverage is adequate, but 100X or higher is ideal for de novo assembly.

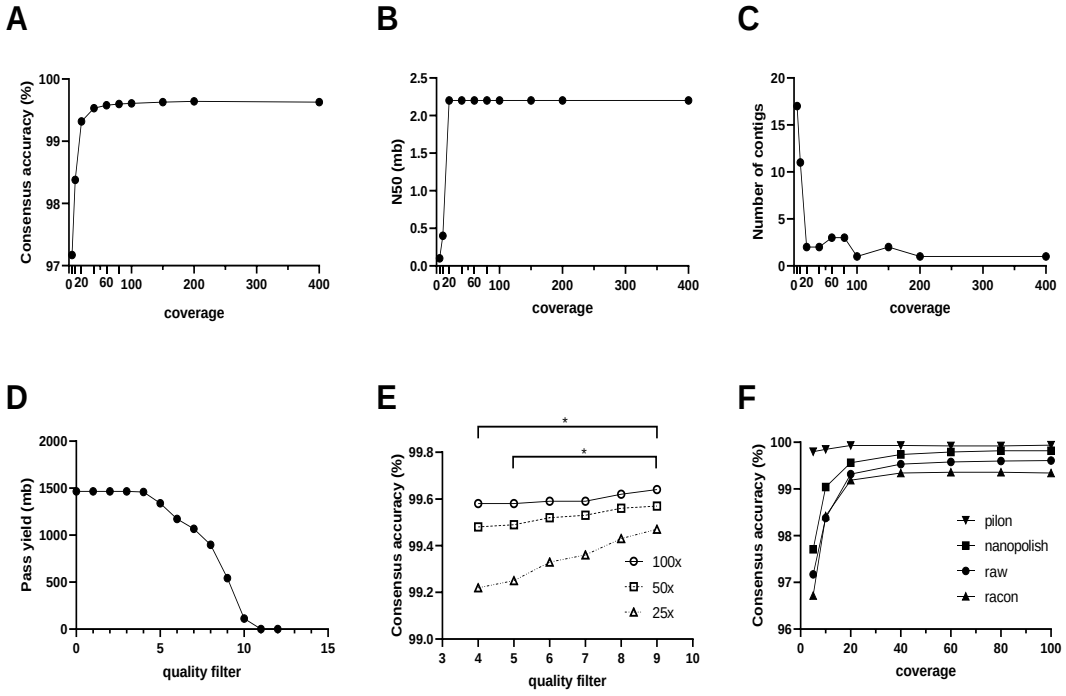

Figure 4. Analysis and optimization of consensus accuracy following de novo assembly.

A consensus sequence was assembled from MinION sequencing reads using de novo assembly. A) The consensus accuracy, B) the N50, and C) the number of contigs were examined at coverage levels from 5X-400X. D) The total number of bases that passed quality filter cut-offs ranging from 0-12 were examined. E) At 25X, 50X, and 100X coverage the consensus accuracy following de novo assembly using different quality filter cut-offs from 4-9 was determined. F) The effect of error correction methods on the consensus accuracy of de novo assembly using nanopolish, racon, or pilon was examined at coverage levels from 4X to 100X. (*$p<0.05$)

Use of a strict quality filter can optimize the quality of input reads for de novo assembly. However, the higher the quality filter the fewer reads will pass the quality filter (pass reads), which leads to lower coverage during de novo assembly. Examination of the number of bases available after different quality filters shows a dramatic drop between a quality filter of 4 to 9 from 1.5 $x10^9$ to 5.4 $x10^8$ total bases available, which means at a quality filter of 9 only 40% of the data generated is used (Figure 4D). Next, an examination of different

quality filter cut-offs from 4 to 9 at three coverage levels (25X, 50X, and 100X, representing low, medium and high coverage) were performed to determine optimal or acceptable quality filter cutoffs (Figure 4E). As quality filter cut-offs increased, consensus accuracy increased, however, the consensus accuracy of lower coverage samples (25X and 50X) was never able to reach the accuracy of higher coverage samples even using lower quality filter cut-offs (100X, quality 4 cut-off) (Figure 4E). This suggests that coverage is more important than read quality for consensus accuracy and if the total sequence yield gives less than 100X coverage, no quality filter should be used. If the coverage is greater than 100X, then use of the quality filter can significantly improve consensus accuracy (Figure 4E). No significant difference was observed in the N50 or the number of contigs generated as quality filter cut-offs increased at each coverage level (data not shown). For the remainder of this study, a quality filter of 7 was applied, which is the default filter when using Albacore basecalling software.

Besides coverage and read quality optimization, consensus accuracy can also be improved by using an error correction program. Three different error correction methods were compared to identify which gave optimal improvement in consensus accuracy (Figure 4F). The Pilon method is a hybrid error correction method incorporating Illumina sequencing reads to correct the MinION consensus and gives a high consensus accuracy regardless of the coverage level but requires Illumina as well as MinION sequencing data (Figure 4F). For the raw read consensus, and the racon and nanopolish error correction methods, the consensus accuracy improved greatly from 5X to 40X coverage and then maintained a similar accuracy out to 100X coverage (Figure 4F). At 100X coverage, the raw consensus accuracy with no error correction was 99.6%, racon gave poor results

(99.3%), self-correction using nanopolish increased the accuracy to 99.8%, and Pilon generated the highest accuracy at 99.9% (Figure 4F). Knowing that nanopolish and Pilon could improve the consensus accuracy, the time and computing power needed to run these error correction methods needs to be weighed against the accuracy that is needed. Using the University of Minnesota Supercomputing Institute High Performance Computing (HPC) system with 252gb supercomputing memory, acquisition of the raw consensus using Canu with no error correction takes about 1 hour. Performing a self-correction on this raw consensus using nanopolish requires approximately 12 hours of computing time while a hybrid-correction using Pilon requires less than 10 minutes of computing time, but requires Illumina data.

3.4. Determination of sequence type and antibiotic resistance

A total of 10 different *S. suis* strains were sequenced following optimal sample preparation as determined above by isolating DNA using the silica-membrane spin column extraction and g-TUBE shearing followed by MinION (long-read) library preparation and sequencing. Raw reads from each sample were assembled into contigs using Canu. Illumina raw reads were assembled using SPAdes. A hybrid consensus genome incorporating the long contigs from MinION and the high accuracy reads from Illumina was generated for each sample using Pilon. This hybrid genome was used as the reference genome for comparing genomes generated by the other 2 methods (Table 1). The average Phred quality score was vastly different between the 2 methods with a Phred score of 8.5 for MinION and >30 for Illumina, thus highlighting the differences in sequence accuracy. A comparison between the MinION and Illumina assembled genomes indicated that although Illumina sequencing generated more sequence data and created contigs with

higher accuracy than MinION contigs, it was unable to assemble into long contigs (Table 1), thus confirming that MinION sequencing is superior to Illumina for generation of large, continuous contigs.

Table 1. De novo assembly results for MinION and Illumina sequencing.

BioSample accession	Sample ID	Total input for assembly (mb)		Number of contigs		N50 (mb)		% Identity*	
		MinION	Illumina	MinION	Illumina	MinION	Illumina	MinION	Illumina
SAMN11854274	1	196	859	2	41	2.1	0.8	98.82	99.87
SAMN11854276	2	434	785	1	129	2.2	0.1	99.64	99.95
SAMN11854295	3	200	856	1	59	2.1	0.2	99.55	99.96
SAMN11854277	4	632	893	1	39	2.2	1.1	99.15	99.91
SAMN11854278	5	870	978	1	102	2.1	0.1	99.54	99.94
SAMN11854305	6	586	678	1	48	2.1	0.2	99.49	99.98
SAMN11854328	7	458	825	1	62	2.2	0.2	99.47	99.95
SAMN11854241	8	597	640	1	63	2.1	0.1	99.54	99.99
SAMN11854253	9	286	666	16	54	2.1	0.2	99.22	99.59
SAMN11854163	10	280	632	3	112	2.1	0.2	99.48	99.83
	Average	454	781	2.8	71	2.1	0.3	99.39	99.90

*% identity as compared to the genome generated using the hybrid method.

Multilocus sequence typing (MLST) was performed using the 10 *S. suis* consensus genomes generated by each MinION, Illumina, and hybrid method. For the Illumina and hybrid consensus genomes, MLST was determined by 100% identity matching to alleles. Because MinION does not generate highly accurate sequence, MLST cannot be performed using the normal parameters of matching an allele using 100% identity. However, the lack of accuracy is due to sequencing error across the genome as opposed to nucleotide changes in a particular region which is observed between alleles, so we assessed the possibility that sequence types (ST) could be accurately predicted by finding the best match allele, even though it would have less than a 100% identity due to random base changes throughout the gene. To test if MinION data could be used to predict the sequence type and to determine

the accuracy of this prediction, the allele with the highest BLAST score was identified. These alleles were then used to determine the ST for each of the 10 *S. suis* samples by PubMLST online tools (pubmlst.org) (Table 2). For sample 4, the *gki* gene didn't match any existing *gki* alleles, indicating the discovery of a new allele and thus a new sequence type. The gene was identified in all 3 assemblies and was PCR amplified and sequenced using Sanger sequencing. The new allele was confirmed, submitted to PubMLST and the allele is now identified as gki_360, and the sequence type is assigned to be ST1214.

Table 2. MLST identification using MinION and Illumina sequencing data.

S. suis sample ID	MLST			
	MinION	MinION + nanopolish*	Illumina	Hybrid
1	975	-	975	975
2	1001	-	1001	1001
3	1 or 134	1	1	1
4†	1214	-	1214	1214
5	971	-	971	971
6	1	-	1	1
7	988	-	988	988
8	1 or 127	1	1	1
9	995	-	995	995
10	1	-	1	1

*Nanopolish was only performed on the MinION sequences for samples 3 and 8 to determine if this could resolve the ambiguity observed in the MinION MLST.
†newly identified ST from this study

The MinION assemblies unequivocally identified the same ST/allelic profile as the Illumina and hybrid genomes for 8 of the 10 samples (Table 2). The MinION genome identified two possible STs for samples 3 and 8 due to 2 alleles giving the same BLAST score (cpn60_1 and cpn60_61 for sample 3; thrA_1 and thrA_18 for sample 8). An examination of the indistinguishable alleles using the clustalW alignment tool identified a 99.7% identity between the alleles for both genes, suggesting that the consensus accuracy of the MinION sequence cannot distinguish between highly similar alleles. After increasing

the accuracy of the MinION genome by performing nanopolish error correction on these two samples, the ST was re-examined. A single allele was now identified using the highest BLAST score for each of the MinION+nanopolish genomes, identifying the same ST/allelic profile as the Illumina and hybrid and genomes with no ambiguities (Table 2). Thus, the original MinION genome, with an average accuracy of 99.6%, was able to predict the majority of STs but had difficulty differentiating high identity alleles (>99.7% identity) while the MinION genome that was error corrected using nanopolish, which increased the average accuracy to 99.8%, was able to correctly predict all STs. Thus, MinION data can be used to predict the MLST of an isolate with 99.7% identity and with higher accuracy if sequencing errors in the genome are corrected using nanopolish.

Antimicrobial resistance (AMR) profiles are an important characterization of bacteria and are necessary for identification of effective treatment of infection. Commonly used antibiotics to treat pigs infected by *S. suis* included ampicillin, chlortetracycline, clindamycin, florfenicol, gentamicin, neomycin, oxytetracycline, penicillin, spectinomycin, tiamulin, trimethoprim, and tylosin. To determine if MinION sequencing could detect AMR genes, especially for those antibiotics involved in pig production listed above, the MinION, Illumina, and hybrid consensus sequences were analyzed using the RGI analysis tool (Jia et al. 2017), followed by confirmation using BLASTn analysis. The RGI analysis tool matches translated protein sequences to the protein sequence of the known AMR genes. Due to insertions and deletions in MinION data, only short sections of the gene are usually able to be translated correctly to determine a >90% identity between the consensus genome and the AMR gene. To further confirm that the AMR gene was present in the genome, BLASTn was used to compare the nucleotide sequence of the identified AMR genes to the

consensus genomes using a cutoff of coverage > 80% to confirm the presence of the entire gene. A total of 6 AMR genes, including tetO, ermB, ANT(6)-la, lnuB, lnuA and tetL, were identified in at least one sample from at least one genome sequence using the RGI tool. Among the 6 AMR genes identified, only tetO, conferring resistance to chlortetracycline and oxytetracycline, and ermB, conferring resistance to tylosin and clindamycin, were antibiotics relevant to pig production (Table 3). The AMR detection results from RGI were consistent among three methods for all samples except sample1, whose results were different depending upon the consensus genome. BLASTn confirmation of gene coverage confirmed RGI results for all the samples except sample 1. BLASTn identified that the RGI results for the sample1 MinION results only covered a small region of the tetO gene, thus the gene was not actually present (Table 3). Another 4 AMR genes, not relevant to pig production, were found to be present in sample 1 (ANT(6)-la, lnuB) and in sample 9 (lnuA, tetL) by RGI, but only 2 of these were confirmed with BLASTn analysis, sample 1 (lnuB; lincosamide) and sample 9 (lnuA; lincosamide).

Table 3. AMR genes detected in the 10 S. suis samples.

AMR genes	tetO		ermB	
Method	RGI detection	BLASTn confirmation	RGI detection	BLASTn confirmation
1*	+/-/+†	-	+	+
2*	+	+	-	-
3	+	+	+	+
4	+	+	+	+
5	+	+	+	+
6	+	+	-	-
7	+	+	+	+
8	+	+	+	+
9	+	+	+	+
10	+	+	+	+

+ denotes gene presence and – denotes gene absence

*AMR detected using modified Canu parameters and contigs

†detection varies based on the sequence method and are shown as MinION/Illumina/Hybrid results

4. Discussion

Rapid diagnostics are important for effective treatment and control of infectious diseases. Identification and treatment of bacterial infections have become routine, but with the emergence of antibiotic resistant bacteria, new approaches are necessary to guide appropriate antibiotic use (Spellberg, Bartlett, and Gilbert 2013; Schurch and van Schaik 2017). Whole genome sequencing (WGS) offers a potential solution for the control of antibiotic resistance and bacterial characterization by providing all relevant genomic information about a pathogen in a single inexpensive, quickly performed assay (Reuter et al. 2013; Ashton et al. 2015; Koser, Ellington, and Peacock 2014). In this study, optimization of MinION sequencing for bacterial detection and characterization was performed to determine if this technology could be effectively used for disease diagnostics and identification of antibiotic resistance.

Since it became commercially available in December 2014, Oxford Nanopore sequencing has been predicted to be a good fit for rapid infectious disease detection (Hoenen et al. 2016). Complete bacterial genome sequences have been generated using the MinION sequencer, either by following a MinION-Illumina hybrid approach or with MinION-only data (Risse et al. 2015; Goldstein et al. 2019). Here we optimized the Oxford Nanopore MinION sequencing workflow for bacterial characterization using *S. suis* as a model organism, since it causes disease in both livestock (pigs) and humans. The methods we developed here can be used as a starting point to optimize sequencing of other bacterial pathogens.

Major disadvantages of MinION sequencing include its need for high quality input DNA and its high sequencing error rates. Analysis of different methods for obtaining high

quality input DNA, including an examination of shearing or size selecting DNA, was performed to determine the most effective method for production of high quality genomic sequence. Our results indicated that use of a silica-based membrane spin column DNA extraction method combined with shearing of this DNA using a g-TUBE gave optimal performance. To overcome the high sequencing error rates associated with MinION sequencing, de novo assembly using Canu was exploited to generate a more accurate consensus sequence. It was determined that 20X-40X coverage was adequate for assembly, but 100X coverage was ideal for optimal consensus accuracy. Use of a quality filter was found to improve consensus accuracy if >100X coverage was obtained, but with lower coverage levels (<100X) no quality filter should be used since coverage seems to be more important than quality. Bioinformatic error correction using nanopolish for self-correction was able to generate sequence with a consensus accuracy of 99.8%, although if Illumina data was available, a hybrid error correction using Pilon generated the highest consensus accuracy (99.9%) in much fewer contigs than that of Illumina alone. However, nanopolish takes a large amount of computing power and a relatively long time to run, so it is not recommended unless high accuracy is necessary, and Pilon requires Illumina sequencing data from the same sample, which is unlikely to be an option. However, if Illumina sequence has already been produced and long contigs are further desired, a quick sequencing run with MinION can be performed and the data combined with the Illumina sequence to obtain high accuracy long read sequence.

Previous research has indicated that long-read sequencing technology (such as MinION) may be more suitable for de novo assembly than short-read sequencing (such as Illumina) (Goodwin et al. 2015; Bainomugisa et al. 2018; Tyson et al. 2018). Our research

confirms this observation, as Illumina generated higher accuracy sequence, but MiniON data was superior for generation of long continuous contigs. Use of a hybrid method to generate long contigs using MinION data as a reference genome and then Illumina data to correct for accuracy gave optimal sequence results and is perfect if both contig length and accuracy are desired. However, in most situations there is no need for a highly accurate consensus sequence and sequencing using 2 different methods is not usually performed.

MinION sequence from de novo assembly (coverage >100X and quality >7) was successfully used for MLST and AMR detection. MLST unequivocally identified 8 of the 10 strains examined and the other 2 strains were typed to 2 possible alleles which differed by less than 0.3%. If more accurate typing was necessary, error correction of the sequence using nanopolish could be performed to correctly identify ambiguous STs. Our study accurately predicted the STs for clinical *S. suis* samples using MinION sequencing, in spite of low accuracy sequence. This indicated that MinION sequencing errors, which are observed across the genome, only give a noisy background of error, instead of a bias for single nucleotide polymorphisms (SNPs) predictive of differing alleles. Of note, MinION sequencing was able to detect the real SNPs and discover a new ST just as well as was performed with Illumina sequencing. Identification of AMR genes using the MinION consensus genome was performed successfully for all genes present in the bacterial genome. However, modification of default assembler parameters was needed to allow for the generation of plasmids, which may contain AMR genes. Also, since consensus sequences from MinION had relatively high errors, the loose algorithm from RGI was used instead of the strict algorithm, which was used for Illumina and hybrid consensus

sequences. In addition, BLASTn was exploited to erase possible false positive results due to identification of partial protein sequence and to confirm detection of AMR genes.

In this study, we optimized the MinION whole genome sequencing workflow for *S. suis* bacterial characterization and demonstrated its use for sequence typing and antibiotic resistance profiling. As sequencing technologies mature, rapid, comprehensive disease diagnosis in a single assay will gradually become a reality. Although the advantage of WGS for disease diagnosis is clear, we have only begun to identify the first steps in standardization of the WGS workflow to decrease costs and shorten the turnaround time for routine application of the technology. Notably, since MinION is a portable sequencer, on-site detection of adverse health events using real-time streaming of information will need to become available to aid in rapid precision treatment solutions (Strong et al. 2014).

Acknowledgments: The authors would like to thank the University of Minnesota Veterinary Diagnostic Laboratory for providing samples for testing. The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported within this paper. URL: http://www.msi.umn.edu.

Funding Information: National Pork Board Grant number 18-040, Boehringer Ingelheim Vetmedica, Inc. Grant, MnDRIVE Global Food Ventures Fellowship awarded to ST, and University of Minnesota Veterinary Population Medicine Department Leman China student support awarded to ST.

Conflicts of Interest: The authors declare that there are no conflicts of interest.

Chapter III - Rapid, unbiased PRRSV strain detection using MinION direct RNA sequencing and bioinformatics tools

A manuscript published in *Viruses*

Tan, S., C. M. T. Dvorak, and M. P. Murtaugh. 2019. 'Rapid, Unbiased PRRSV Strain Detection Using MinION Direct RNA Sequencing and Bioinformatics Tools', *Viruses*, 11.

Author Contributions: conceptualization, MPM; methodology, software, validation, formal analysis, investigation, data curation, visualization, and writing-original draft preparation, ST; writing-review and editing, ST and CMTD; supervision, CMTD; project administration, CMTD; funding acquisition, MPM, CMTD and ST.

Prompt detection and effective control of porcine reproductive and respiratory syndrome virus (PRRSV) during outbreaks is important given its immense adverse impact on the swine industry. However, the diagnostic process can be challenging due to the high genetic diversity and high mutation rate of PRRSV. A diagnostic method that can provide more detailed genetic information about pathogens is urgently needed. In this study, we evaluated the ability of Oxford Nanopore MinION direct RNA sequencing to generate a PRRSV whole genome sequence and detect and discriminate virus at the strain-level. A nearly full length PRRSV genome was successfully generated from raw sequence reads, achieving an accuracy of 96% after consensus genome generation. Direct RNA sequencing reliably detected the PRRSV strain present with an accuracy of 99.9% using as few as 5 raw sequencing reads and successfully differentiated multiple co-infecting strains present in a sample. In addition, PRRSV strain information was obtained from clinical samples containing 10^4 to 10^6 viral copies or more within 6 hours of sequencing. Overall, direct viral RNA sequencing followed by bioinformatic analysis proves to be a promising approach for identification of the viral strain or strains involved in clinical infections, allowing for more precise prevention and control strategies during PRRSV outbreaks.

1. Introduction

The swine industry plays an important role in feeding the world, as pork is one of the highest consumed animal proteins in the world (FAO 2019; Lager and Buckley 2019). Emerging and re-emerging viral infectious diseases have been posing great challenges to the swine industry, among which porcine reproductive and respiratory syndrome (PRRS) is one of the most devastating diseases (Nathues et al. 2017; Neumann et al. 2005). PRRS

virus (PRRSV) is the causative agent of PRRS which contains a positive-sense, single-stranded, 15 kb RNA genome (Snijder and Meulenberg 1998). PRRSV is categorized into two genotypes, type 1 (European type) and type 2 (North American type), which differ by approximately 40% at the genomic level between the two genotypes (Forsberg 2005; Murtaugh et al. 1998; Nelsen, Murtaugh, and Faaberg 1999), and strains within each genotype also vary considerably with genomic differences as high as 20% (Meng 2000).

Globally, PRRS remains a threat to the swine industry despite many years of combined efforts to combat and control infection and disease (Han et al. 2019). One of the challenges for PRRSV control is the frequent recurrence of PRRS outbreaks in swine farms (Tousignant et al. 2015), with a prediction that of farms reporting an outbreak today, 71% will have a recurrence of PRRSV infection within the following two years (Carles Vilalta 2019). The PRRSV recurrence is either caused by introduction of a new strain or a re-break with the resident virus strain. The knowledge of which type of break is crucial to determine the necessary control methods. Controlling against a new introduction from outside the farm indicates a need for enhanced biosecurity, while a re-break of a resident strain suggests better strategies for an elimination or vaccination program are needed. Another big challenge for PRRS control is that PRRSV vaccine is not completely effective at preventing and controlling infection due to the high genetic diversity of the virus, thus outbreaks still occur in vaccinated herds (Murtaugh and Genzow 2011; Opriessnig et al. 2002; Wang, Marthaler, et al. 2015; Zhou et al. 2017). A diagnostic method which can provide genetic information about the strain causing infection would allow for identification of potential reasons for vaccination failure, such as limited cross-protection due to high genetic divergence from vaccine (Kim et al. 2015), or in the case of genetic

similarity to vaccine, perhaps a reversion to virulence (escape mutation) of the vaccine itself (Storgaard, Oleksiewicz, and Botner 1999). In addition to the clinical challenges mentioned above, PRRSV has widely divergent genetic lineages and is a rapidly evolving pathogen with novel variants which seem to be more divergent and virulent than those in the past (Murtaugh et al. 2010; Hanada et al. 2005; Kappes and Faaberg 2015; Han et al. 2019). The continuous emergence of new virulent strains causes unexpected devastating outbreaks, such as the severe outbreaks of HP-PRRSV in China and MN184 and NADC30 outbreaks in the United States (Zhou and Yang 2010; Han, Wang, and Faaberg 2006; Brockmeier et al. 2012; Han et al. 2017). The increasing incidence of co-infections of multiple strains further complicates PRRS diagnosis and control (Yang et al. 2017). Hence, diagnostic tools that can provide more genetic information are extremely important for investigation, prevention, and control strategies for PRRSV outbreaks.

Prompt detection of pathogens during an outbreak is essential for efficient disease control. Real-time quantitative reverse transcription polymerase chain reaction (RT-qPCR) testing, which has the advantages of short turnaround time and high sensitivity (Schmitt and Henderson 2005; Prachayangprecha et al. 2014), is currently the primary molecular diagnostic method for PRRSV detection and is performed routinely in diagnostic laboratories. The current PCR methods can quickly detect the presence of PRRSV in general (Lurchachaiwong et al. 2008), but cannot identify specific strains unless targeted primers are designed which requires prior knowledge about the strains of interest (Yang et al. 2017; Wernike et al. 2012). Nucleic acid sequencing technologies have been incorporated as a new diagnostic tool in recent years to provide necessary support, such as strain information, for PCR in clinical sample diagnostics (Van Borm et al. 2015; Gwinn,

MacCannell, and Armstrong 2019). Sequencing approaches have been applied successfully to various fields in virology so far, such as the discovery of novel viruses/strains, whole viral genome sequencing, quasispecies detection, and epidemiological investigations (Capobianchi, Giombini, and Rozera 2013; Houldcroft, Beale, and Breuer 2017). The fact that sequencing is robust and doesn't need prior knowledge of the pathogens/strains under detection is particularly important for rapid responses to highly variable pathogens, such as PRRSV (Liu et al. 2018). The routine use of viral genome sequencing and genomic surveillance will not only serve as a powerful tool for PRRSV detection (Parker and Chen 2017), but also provides researchers with a better understanding of PRRSV epidemiology and how the virus transmits, spreads, and evolves, thus facilitating effective prevention and control measures (Zhao et al. 2012).

Conventional sequencing methods for RNA viruses usually includes reverse transcription and PCR amplification during library preparation which is then followed by amplicon sequencing. These extra steps not only introduce bias, but also increase the sequencing time, making rapid diagnosis difficult (Lahens et al. 2014; Hoenen et al. 2016; Garalde et al. 2018). The Oxford Nanopore MinION sequencer allows for sequencing of the RNA molecule directly, in its native format. This feature, together with other characteristics such as low start-up costs, portability, and real-time data streaming, makes the Oxford Nanopore MinION sequencer a good candidate for rapid RNA virus detection, even in resource limited or remote areas (Gallagher et al. 2018; Faria et al. 2016; Hoenen et al. 2016). Influenza virus was the first pathogen to be successfully sequenced in its native RNA format by direct RNA sequencing (DRS) using Oxford Nanopore MinION technology (Keller et al. 2018). Since then, studies have been performed for other viruses,

confirming the potential of MinION technology to aid in the detection of infectious viral agents (Viehweger et al. 2018; Depledge et al. 2019). PRRSV whole genome sequencing (WGS) has been carried out previously using traditional Sanger sequencing (Han et al. 2019) and next-generation short-read sequencing platforms (Kvisgaard et al. 2013). Incorporating a bioinformatic approach, we systematically evaluated and standardized third-generation long-read MinION DRS for PRRSV whole genome generation and determined its ability to detect the viral strain present, its analytical sensitivity for strain-level diagnosis of clinical samples, and its feasibility for differentiation of co-existing multiple strains in a single sample. To our knowledge, our study is the first time PRRSV has been sequenced in its native RNA format without amplification.

2. Materials and Methods

2.1. Viral strains and samples

A PRRSV type 2 isolate, VR2332 (GenBank: EF536003.1), was used as the main reference strain. A PRRSV type 1 isolate PRRSV1/USA/Lab6 (SDEU, GenBank: MN175678) and PRRSV type 2 isolate PRRSV2/USA/Lab3 (1-7-4, GenBank: MN175677) were used for experiments examining the detection of multiple viral isolates in a single sample. All viral isolates were propagated on MARC 145 cells as previously described (Robinson et al. 2018). Clinical samples and the corresponding ORF5 sequences from Sanger sequencing, were obtained from Dr. Sunil Kumar Mor at the University of Minnesota Veterinary Diagnostic Laboratory (St. Paul, MN).

2.2. RNA extraction and viral copy number determination

PRRSV RNA was extracted from cell culture supernatants, virus-negative pig serum spiked with PRRSV, and clinical PRRSV-positive serum samples using the QIAamp

Viral RNA mini kit (Qiagen, Germantown, MD) following manufacturer's instructions without the addition of carrier RNA and with a final elution in 50 ul nuclease-free water. A high concentration PRRSV stock (supernatants from virus grown in MARC-145 cells) was extracted to generate a large amount of high concentration RNA for whole genome sequencing. Generation of known concentrations of virus in serum samples (spike-in samples) was performed by adding the PRRSV stock to virus negative pig serum, half of which was used for sequencing and the other half for determining the number of viral copies present. For clinical samples, RNA was extracted from 300 ul of serum, two thirds of which was used for sequencing and the remaining third was used to determine the number of viral copies present. Viral copies were determined using an RT-qPCR assay as described previously using a standard curve to determine the number of viral copies and then calculating the total number of copies sequenced (Robinson et al. 2018).

Since MinION RNA sequencing requires a high amount of input RNA for library preparation (>500ng), lower viral RNA concentration samples were supplemented with exogenous cellular RNA for sequencing library preparation. This exogenous cellular RNA was obtained by extracting total RNA from MARC-145 cells using the Qiagen RNeasy mini kit (Qiagen, Germantown, MD) according to the manufacturer's protocol with the addition of on-column DNAse digestion. When needed, concentration of RNA was performed using a SpeedVac lab concentrator (Savant, NY, USA). A Qubit 3.0 fluorometer (Life technologies, Carlsbad, CA) and a Nanodrop1000 spectrophotometer (Thermo Scientific, Waltham, MA) were used for quantitative and qualitative assessments.

2.3. MinION direct RNA sequencing

Sequencing libraries were generated from 600ng of extracted viral RNA or a combination of viral RNA and exogenous cellular RNA using the direct RNA sequencing kit (Oxford nanopore Technologies Ltd, Oxford, United Kingdom) according to the manufacturer's protocol (Garalde et al. 2018). The sequencing library was then loaded onto a R9.4.1 SpotON flow cell and sequenced using a MinION Mk I sequencer (Oxford nanopore Technologies Ltd, Oxford, United Kingdom) which was connected to a computer and remotely controlled by the MinKNOW software (Oxford nanopore Technologies Ltd, Oxford, United Kingdom). The estimated yield was monitored in real-time, samples were sequenced for approximately 6 hours and adjusted for more or less time if needed.

For evaluation of whole viral genome generation from MinION direct RNA sequencing, two duplicate runs were performed starting with 600 ng PRRSV VR2332 genomic RNA. Sequencing of mixed-strain samples combined 300 ng of VR2332 RNA and 300ng of strain 1-7-4 or SDEU RNA, or 600 ng VR2332 RNA total as a control. Other samples that contained less than 600 ng of PRRSV RNA, such as clinical samples, were supplemented with exogenous cellular RNA to obtain a total of 600 ng RNA for use in library preparation.

2.4. Evaluation of sequencing reads and consensus sequences

Basecalling of raw reads was performed using Albacore (Oxford nanopore Technologies Ltd, Oxford, United Kingdom) to generate FASTQ files. Total yield, total reads, read quality, and read length from whole genome sequencing were analyzed using NanoPlot (Van Broeckhoven et al. 2018). To obtain raw error rates and error patterns, sequencing reads were mapped to the VR2332 reference sequence using minimap2 (Li

2016), processed with SAMtools (Li et al. 2009) to generate BAM files, and then evaluated by AlignQC (Weirather et al. 2017).

A consensus genome was generated using the longest PRRSV read from the sequencing data as a scaffold. The longest PRRSV read was extracted from the FASTQ file using an awk command, all other raw reads were then mapped to this sequence using minimap2 (Li 2016), and then the map file was processed using Racon (Vaser et al. 2017). A comparison of this consensus genome to the reference genome was analyzed by pairwise alignment using Geneious software (version 8.0.5) (Kearse et al. 2012). Depth of coverage across the consensus genome was analyzed using Qualimap (Okonechnikov, Conesa, and Garcia-Alcalde 2016). The average coverage and accuracy across the genome were then evaluated using a window size of 1000 bp and visualized using GraphPad Prism 8 (GraphPad Software, San Diego, CA).

2.5. Evaluation of analytical sensitivity

The analytical sensitivity of MinION direct RNA sequencing was analyzed by examining the sequencing yield needed for viral strain detection, as well as the number of viral copies needed to generate detectable viral sequence. The sequencing yield needed for viral strain detection was examined by generating datasets with targeted yields ranging from 3,000 to 30,000,000 bases from the two whole genome sequencing runs. Specifically, the text summary of the sequencing file from basecalling was analyzed using R (version 3.4.0) (Team 2016) and groups with the desired yields were generated by setting a cutoff at the sequencing time in which the desired yield was reached. Examination of the number of viral copies needed in a sample in order to detect the virus was performed by sequencing viral RNA extracted from cell supernatant samples, spike-in samples, and clinical samples

containing different amounts of virus. Because samples with a relatively low number of viral copies yielded low amounts of viral RNA, exogenous cellular RNA was added to achieve efficient library production. Following sequencing of the libraries containing both viral RNA and cellular RNA, the PRRSV sequences needed to be extracted for further analysis. First, a custom PRRSV sequence database containing 951 PRRSV whole genome sequences was generated by downloading all PRRSV whole genome sequences available in GenBank (949 sequences including our VR2332 strain, download date: Nov 2018) with the addition of sequences from our SDEU and 1-7-4 lab strains. Then, the PRRSV reads were able to be identified and obtained by mapping the raw sequencing reads to this custom PRRSV database using minimap2 (Li 2016) and extracting the mapped reads using SAMtools (Li et al. 2009).

Identification of the viral strain present in the sample was examined using Basic Local Alignment Search Tool (BLAST) with a significance filter of expect value (E) < 10^{-50} to examine the PRRSV sequence reads. The PRRSV raw reads were compared to the custom PRRSV database using nucleotide BLAST (BLASTn) and the top match, based on bit score, was regarded as the strain detected in the sample. This detected sequence was then aligned to the known reference genome using Geneious software version R8.0.5 (Kearse et al. 2012) and the percent identity was recorded to show the accuracy of detection. For supernatant and spike-in samples, both the VR2332 whole genome and the ORF5 sequence were known and designated as the reference sequence to compared to the MinION generated sequences. For clinical samples, only the ORF5 sequence was known and was used as the reference sequence for comparison. A consensus genome was

generated, if possible, for each dataset or sample using the longest PRRSV read as a scaffold followed by analysis of consensus length and accuracy as described above.

Linear regression analysis was performed to compare PRRSV sequencing reads to viral RNA copies using GraphPad Prism 8 (GraphPad Software, La Jolla, CA). In order to normalize among different sequencing runs with varying total reads, the ratio of PRRSV reads to total reads was used to allow for comparison. The viral RNA copies were determined by RT-qPCR and reported as total viral copies per sequencing run.

2.6. Differentiation of multiple viral isolates in a single sample

Samples containing a mixture of two viral isolates, or VR2332 alone as a control, were sequenced as above. In order to identify the yields needed for accurate strain detection and differentiation, datasets with yields from 30,000 to 30,000,000 bases were generated randomly from total reads using fastq-tools (https://homes.cs.washington.edu/~dcjones/fastq-tools/). PRRSV reads were extracted by mapping all reads to the PRRSV database using minimap2 (Li 2016). In order to detect PRRSV strains, PRRSV reads were first BLASTn analyzed to identify the top BLAST hit as determined by bit score (BLAST filter of $E < 10^{-50}$ plus alignment identity > 80% and length > 900 bp). Then, all PRRSV reads were mapped to the this top BLAST hit using minimap2 with "map-ont" preset option (Li 2016) and mapped reads were extracted using SAMtools (Li et al. 2009). The unmapped reads were also extracted and were analyzed against the PRRSV database a second time to detect any other strain existing in the same sample. The top BLAST hit was recorded and the mapped and unmapped reads to the second top match were again separated. This was repeated until no PRRSV strain was detected in the extracted unmapped reads. The read length and accuracy were based on the

results of the analytical sensitivity experiment, where the detection limit was approximately 900 bp and 80% identity. The top BLAST hits were compared to the targeted known strain (1-7-4, SDEU, or VR2332) and the percent identity was recorded. The percentages of reads matching the detected isolates to total PRRSV reads were also recorded.

The investigation of previous-run contamination was conducted by extracting all reads from the suspected sequencing results that mapped to the reference sequence of the contaminating strain. The "read_id" of the contaminating reads were extracted using SAMtools. As an indication of when during the sequencing run the contaminating read was observed, the "start_time" that matched the "read_id" of the contaminating reads was extracted using R (version 3.4.0) (Team 2016). The number of total contaminating reads over the time course of the sequencing run was analyzed using GraphPad Prism 8 (GraphPad Software, La Jolla, CA).

2.7. Computer codes and sequencing data

The main bioinformatic methods and codes used in this study can be found here: https://github.com/ShaoyuanTan/PRRSVproject

The sequencing data has been deposited to NCBI Sequence Read Archive (SRA) under accession numbers: SRR10292736 to SRR10292741.

3. Results

3.1. Evaluation of MinION RNA sequencing for generation of viral genomes

A high concentration cell culture grown PRRSV VR2332 stock was used for RNA isolation and evaluation of MinION direct RNA whole genome sequencing. PRRSV RNA was extracted using the QIAamp Viral RNA mini kit, which has shown consistently good

performance in several studies (Cornelissen et al. 2017; Conceicao-Neto et al. 2015). A total of 600 ng RNA was used for library preparation and sequencing, which was performed in duplicate. Since the whole genome sequencing was under ideal conditions using 600 ng RNA starting material, one-hour of sequencing was sufficient to generate more than enough reads for sequence analysis (Table 4).

Table 4. Assessment of raw reads from direct RNA sequencing

	Run #1	Run #2
Available pores (group 1)	474	495
Sequencing time	1 hour	1 hour
Total pass bases	20,351,741	27,167,775
Total pass reads	14,963	23,547
Mean read length (bp)	1,360	1,154
Mean read quality	8.2	8.5
Mappable reads / percentage	13,284 / 88.8%	19,549 / 83.0%
Longest read (bp) / accuracy	15,026 / 86.3%	15,060 / 86.7%
Consensus length (bp) / accuracy	15,140 / 95.5%	15,055 / 95.3%

Raw reads from the first hour of sequencing were extracted and evaluated for yield, read quality, read length, raw error rates and consensus generation (Table 4). Both sequencing runs generated more than 20 megabases (mb) total yield within one-hour of sequencing with the longest raw read over 15,000 bp in length, very close to the full length VR2332 reference sequence (15,182 bp) (Table 4). Comparing the longest raw read to the VR2332 reference sequence gave an identity of approximately 86.5%, and the sequence accuracy improved to 95.4% after generating a consensus using the longest raw read as a scaffold (Table 4). Further examination of the error rates between the raw reads and the

reference sequence identified total error rates at 13.9%, including 6.3% deletion (45% of total error), 4.1% mismatch (30% of total error), and 3.5% insertion (25% of total error) error types (Figure 5a). Of note, error patterns showed that insertion and deletion of T nucleotides, and C/T mismatches were the most frequently observed error patterns (Figure 5b).

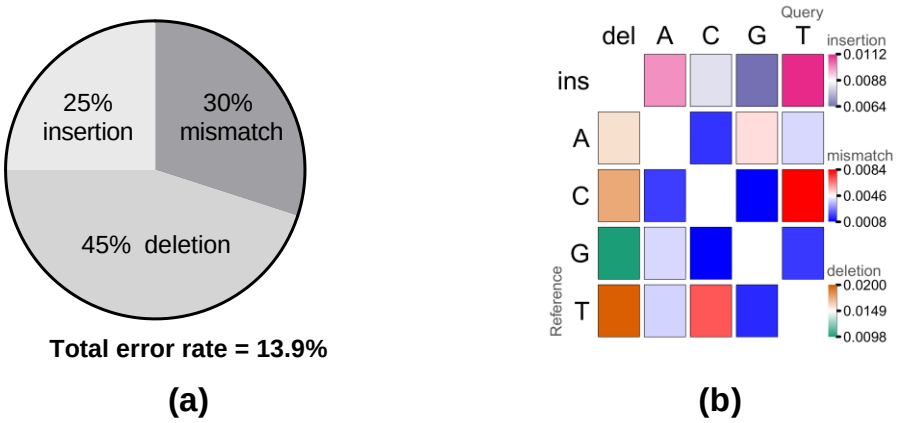

Figure 5. Analysis of direct RNA sequencing errors.

To obtain raw error rates and error patterns, raw reads were mapped to the VR2332 reference sequence, followed by evaluation of the mapping. (a) The percent of each error type is shown as well as the total error rate. (b) The error patterns of insertions (first row with darker pink indicating higher errors), deletions (first column with darker orange indicating higher errors) and mismatches (center matrix with darker red indicating higher error). The U bases in the query sequence were adjusted to T automatically by the minimap program in order to map to the reference sequence which was DNA.

The depth of coverage across the PRRSV genome was observed to be extremely uneven with higher coverage on the 3' end of the genome and gradually decreasing towards the 5' end, which agrees with what has been observed previously (Figure 6) (Viehweger et al. 2018; Keller et al. 2018). This is not surprising since the sequence adaptor was ligated to the poly(A) tail on the 3' end and this is where sequencing began. If the RNA was partially degraded or RNA second structure hampered the movement of the RNA through the nanopores, then only the 3' end would be sequenced, thus resulting in uneven coverage distribution. Despite the uneven coverage, the accuracy across the genome was similar,

around 95%, with the middle region of the genome having a higher accuracy (97%) and the 3’ end having the lowest accuracy (93%) (Figure 6). This was surprising since higher coverage can correct random sequencing errors and usually results in higher accuracy, which would suggest the 3’ end would have a much higher accuracy instead of a lower accuracy. Such conflicts imply the existence of technological bias resulting in sequencing errors that cannot be corrected by depth of coverage.

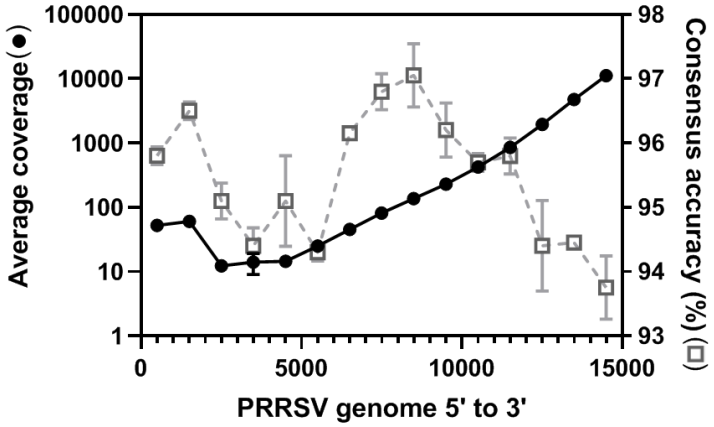

Figure 6. Depth of coverage and consensus accuracy across the PRRSV whole genome.

Raw reads were mapped to the longest raw read which served as a scaffold to generate a consensus genome. The coverage distribution (left y-axis, black closed circles) was evaluated by Qualimap. The consensus accuracy (right y-axis, grey open square, dashed line) was generated using a Geneious pair-wise alignment. Both the average coverage and accuracy were evaluated using a window size of 1000bp and visualized using GraphPad prism software.

3.2. Analytical sensitivity of MinION direct RNA sequencing

3.2.1 Examination of sequencing yield needed for strain detection

Analytical sensitivity of MinION direct RNA sequencing was first evaluated by examining sequencing results over a range of sequence yields to determine the lowest sequencing yield at which the PRRSV virus could be identified and at which a consensus genome could be generated. A range of sequence yields from 3 kilobases (kb) to 30,000 kb were obtained from the two whole genome sequencing runs above. Total reads were

analyzed against a custom PRRSV database using BLASTn and the top match for each sequence yield, even those with only a few reads, was GenBank ID KC469618.1 (15,458 bp). A 99.9% identity was observed between the known sequence of the VR2332 strain used in this experiment (GenBank ID EF536003.1, 15,182 bp) and the top BLAST match, KC469618.1, with an alignment length of 15,183 bp and only 15 base changes, suggesting they are basically the same isolate, especially since PRRSV has a high mutation rate estimated at (4.71–9.8) × 10^{-2}/ site/year (Hanada et al. 2005).

The length and accuracy of the longest reads, and generation of consensus sequences were further examined at the different sequence yields (Table 5). As sequencing yield increased, the length of the longest reads obtained increased, as did the length of the consensus sequence, reaching a maximal level at a yield of 15,000 kb (Table 5). The accuracy of the longest read at the different yields didn't change, confirming that read quality was similar over time and that read quality was not related to read length. However, the accuracy of the consensus sequence increased from about 92% to 95% from 15 kb to 7,500 kb input yield, due to the increased depth of coverage (Table 5). Consensus accuracy generated from yields more than 7,500 kb were consistently above 95% (Table 5). A nearly full length, 15,101 bp in length (breadth of coverage 99.5%), PRRSV consensus genome sequence with a sequence accuracy of 95.2%, was generated from a sequence yield of 15 mb (Table 5). The minimal sequencing yield required for accurate PRRSV strain detection was found to be 3 kb (~5 reads) (Table 5). A total sequencing yield of 15 mb (~6×10^4 reads) allowed for accurate whole PRRSV genome generation (Table 5).

Table 5. Detection power of MinION direct RNA sequencing at different sequence yields[a]

Sequencing yield (kb)	Number of total reads	Longest read		Consensus sequence		Coverage	
		Length (bp)	Accuracy	Length (bp)	Accuracy	Breadth[b]	Depth[c]
3	5	1,606	84.2%	-	-	-	-
15	109	1,899	88.1%	1,861	92.10%	12.3%	1
75	443	4,036	84.1%	4,081	91.95%	26.9%	5
150	790	4,496	83.4%	4,548	92.45%	30.0%	10
750	3,500	7,533	84.9%	7,609	92.55%	50.1%	49
1500	6,857	8,382	85.8%	8,435	94.20%	55.6%	99
7500	32,571	12,990	86.6%	12,988	95.20%	85.5%	494
15000	64,860	15,043	86.7%	15,101	95.20%	99.5%	988
30000	127,411	15,081	86.0%	15,171	95.50%	99.9%	1,976

[a]average of two sequence runs

[b]consensus genome length (bp)/reference genome length (bp)

[c]sequencing yield (bp)/reference genome size (bp)

3.2.2 Determination of minimal viral copy level needed for sequencing

The high amounts of viral RNA used for evaluation of MinION sequencing yields above are unrealistic and do not represent amounts of virus that can be found in field samples. Thus, analytical sensitivity was next examined using samples with a more realistic amount of viral copies present. A total of 5 lower concentration cell culture samples, 3 serum samples with known amounts of virus spiked-in, and 6 clinical samples containing varying amounts of virus were sequenced. The total number of viral copies that were used for each MinION sequencing reaction was determined using RT-qPCR, with a range of 3.2 x 10^4 to 5.9 x 10^9 viral copies per sequencing reaction in these samples (Table 6). The PRRSV strain was determined by analyzing total raw reads from sequencing against the custom PRRSV database and the top BLAST match was used to identify the viral strain present in the sample (Table 6). MinION sequencing was able to detect PRRSV in spike-in samples containing as low as 3.4 x 10^4 viral copies and in clinical samples at 3.8 x 10^6

viral copies (Table 6). The analytical sensitivity difference related to sample type was unexpected, but, in fact, reasonable. One possible reason for this sensitivity difference could be related to viral RNA quality. Viral RNA extracted from cell culture supernatants are produced cleanly in a lab and are quickly stored properly to minimize viral and RNA degradation, thus giving higher quality samples. Clinical samples, on the other hand, are usually obtained on farm and the subsequent handling, shipping, and storage of clinical samples will inevitably increase viral and RNA degradation and decrease sample quality, resulting in lower sequencing yields, while RT-qPCR, which is less sensitive to these conditions, can still detect the presence of the virus (Sanchez-Romero et al. 2019).

The detection accuracy of the raw PRRSV reads was determined by comparing the top BLAST hit to the known ORF5 sequence and/or whole genome sequence (Table 6). For cell supernatant and spike-in samples, the detection accuracy remains almost the same even as the viral copy number increased from an order of 10^4 to 10^9, and the top hits all showed more than a 99% identity to the reference whole genome sequence (Table 6). For clinical samples, more than $2.3x10^5$ viral copies were needed in order to detect viral sequence (Table 6). At 3.8×10^6 viral copies the detection accuracy, comparing the top BLAST hit to the known ORF5 sequence, was 94%, increasing to 97% as the number of viral copies increased (Table 6).

Table 6. Analytical sensitivity of direct PRRSV RNA sequencing.

Sample type	Viral copies/ reaction	# of total reads	Top BLAST match	Identity to ORF5/whole genome %	# of PRRSV reads	Consensus/longest read	
						Length (bp)	Accuracy %
Cell supernatant	5.9×10^9	19,198	KC469618.1	100.0/99.9	1,247	8,282	94.5
	2.0×10^9	23,068	KC469618.1	100.0/99.9	949	7,167	93.8
	1.7×10^9	83,192	KC469618.1	100.0/99.9	831	9,187	93.2
	6.8×10^8	116,698	KC469618.1	100.0/99.9	699	5,975	93.1
	3.7×10^8	118,879	KC469618.1	100.0/99.9	422	6,028	93.7
Spike-in	1.5×10^9	322,778	KC469618.1	100.0/99.9	1,589	15,021	93.0
	9.4×10^6	300,143	KC469618.1	100.0/99.9	45	3,743	90.5
	3.4×10^4	161,569	CS484777.1	99.0/99.4	3	905*	82.1
Clinical	1.4×10^8	77,468	MF327000.1	96.8/-	42	1,984	90.5
	2.4×10^7	266,120	KX192112.1	97.0/-	16	2,431	88.4
	3.8×10^6	286,680	KT581982.1	94.4/-	6	940*	83.7
	2.3×10^5	201,887	ND	-	0	-	-
	6.5×10^4	240,944	ND	-	0	-	-
	3.2×10^4	307,822	ND	-	0	-	-

ND: not detected

*Longest raw read was used

PRRSV consensus sequences were obtained from each of the samples, if possible, in order to evaluate the ability of DRS to generate accurate consensus sequence from low viral copy samples (Table 6). MinION sequencing produced a large number of total raw reads, most of which were from the added exogenous cellular RNA necessary for successful library preparation. The desired PRRSV reads were obtained through mapping raw reads against the custom PRRSV database and using those that matched to generate a consensus sequence. A consensus sequence was not able to be obtained for 2 of the samples (spike-in 3.4×10^4 viral copies and clinical 3.8×10^6 viral copies) because of the low number of PRRSV reads present, so the longest PRRSV read was used for accuracy analysis instead of a consensus sequence (Table 6). The accuracy of the consensus sequence (or longest PRRSV read) was determined by comparing it to the known whole genome and/or ORF5 sequence (Table 6). Not surprisingly, there was a general trend that

longer and more accurate consensus sequences were generated when more viral copies were sequenced, with slight fluctuation due to variations in sequencer performance (Table 6). Notably, a basically full-length genome with a consensus accuracy of 93.0% was observed in the spike-in sample containing 1.5×10^9 viral copies (Table 6). The other three samples in which more than 10^9 viral copies were used as the input sample were also able to generate a consensus genome with an accuracy higher than 93%, but were not full-length genomes, perhaps due to the low number of PRRSV reads (and total reads) even though the percent of PRRSV reads per total reads was higher in these samples. Thus, more than 10^9 viral copies with perhaps 1,500 PRRSV reads are recommended if the goal is to obtain a full-length genome sequence, but if identification of the viral strain involved in infection is all that is needed, then clinical serum samples need only have 10^6-10^7 viral copies to be successful (Table 6).

3.2.3 Determination of sequencing as a quantitative method

A comparison between the number of viral copies and the number of viral reads from sequencing was performed to determine if there was a quantitative relationship between input PRRSV RNA amounts and output PRRSV sequencing reads. Of note, the total raw reads varied greatly even though the same amount of total RNA was used for library preparation (Table 6), which was mainly due to the variation of flow cell performance, such as available pores. In order to normalize the comparison, the ratio of PRRSV reads to total reads was calculated and compared to the input viral copies and a strong positive correlation (r^2=0.88) was observed. Thus, the knowledge of the number of viral copies in a sample can predict the approximate number of raw reads that will be obtained after sequencing allowing for more successful sequencing results, and the number

of reads obtained from sequencing can be used to estimate the number of viral copies present in a sample.

3.3. Detection of multiple viral isolates present in a single sample

In swine farms PRRS outbreaks can occur even in herds that are vaccinated, therefore it is necessary to be able to differentiate the presence of infectious field strains from vaccine to aid in outbreak investigation (Murtaugh and Genzow 2011; Opriessnig et al. 2002; Zhou et al. 2017). To address this issue, we explored the use of MinION DRS for detection of multiple PRRSV strains in the same sample using a stepwise BLAST approach. Samples were created that contained the VR2332 strain (parental strain to the type 2 PRRSV MLV vaccine) to represent vaccine, and either a type 1 PRRSV strain (SDEU, 61.4% similarity with VR2332) or another type 2 PRRSV strain (1-7-4, 82.4% similarity with VR2332). After sequencing, PRRSV reads were extracted from total reads. PRRSV reads were BLAST analyzed against the custom PRRSV database to identify the top match strain and all PRRSV reads that were able to map to this strain were obtained. The unmapped reads were then BLAST analyzed a second time against the custom PRRSV database to identify the top match of these remaining sequences and they were then mapped to this second top match. If unmapped sequences remained, this pipeline was repeated to identify more than 2 PRRSV strains present in the sample. Results showed that even at a total sequence yield of 30 kb (20-26 PRRSV reads), MinION sequencing was able to identify a PRRSV strain with >99.9% identity to the input VR2332 strain (Table 7). The control samples did not identify a second PRRSV sequence present (at any sequence yield) which was promising, since VR2332 was the only virus present. In the mixed virus samples, the second viral strain was not detectable at a total yield of 30 kb. However, at 300 kb or

higher yields (245 or more PRRSV reads), the second strain could be identified with an accuracy > 99.8% (Table 7). Thus, if enough virus is present from both strains, they could be successfully detected in a single sample. Interestingly, in the VR2332 + 1-7-4 sample, SDEU sequences were also detected, which was not expected since that strain was not present in the sample. Previously, others have observed between-run carryover contamination on the same MinION flowcell (Greninger et al. 2015; Ma, Stachler, and Bibby 2017). Our observation also indicates the carryover contamination from our previous VR2332 + SDEU sample sequencing. This reiterates the need for effective washing of flow cells, as well as good records of what is run on each flow cell previously, especially if flow cells are used for diagnostics. Further investigation into the SDEU carry over contamination showed that SDEU reads were consistently generated during the entire sequencing run, thus contaminating reads could not be minimized by removing the first few minutes of sequencing, they contaminated the entire sequencing run. Although this experiment was designed to differentiate field strains from vaccine strain, it can be applied to the investigation of multiple co-infection strains. Since the identification of the strains present is based on the top BLAST match, any strain with a known genome or similar genome to one in the database could be identified. The strains examined here had at least an 82.4% identity and it is not known if strains with higher identity to each other would both be able to be distinguished, but with an adjustment of the minimap parameters used to map reads to the top BLAST hit, they should be able to be observed.

Table 7. Mapping status of direct RNA sequencing on samples containing multiple viral strains.

Groups	Total yield (kb)	# of total reads	# of PRRSV reads	PRRSV reads / total reads	First match			Second match			Third match		
					Top BLAST match (identity %)	# of matching reads	% of PRRSV reads	Top BLAST match (identity %)	# of matching reads	% of PRRSV reads	Top BLAST match (identity %)	# of matching reads	% of PRRSV reads
Control	30	25	20	80%	KC469618.1 (99.9)	20	100%	ND					
	300	245	210	86%	KC469618.1 (99.9)	208	99%	ND					
	3,000	2,451	2,079	85%	KC469618.1 (99.9)	2,044	98%	ND					
	30,000	24,512	20,819	85%	KC469618.1 (99.9)	20,472	98%	ND					
VR2332 + SDEU mixed sample	30	38	23	61%	KC469618.1 (99.9)	19	83%	ND					
	300	375	234	62%	KC469618.1 (99.9)	194	83%	CS421743.1 (99.8)	35	15%	ND		
	3,000	3,748	2,281	61%	KC469618.1 (99.9)	1,722	75%	SDEU (100.0)	514	23%	ND		
	30,000	37,478	23,004	61%	KC469618.1 (99.9)	17,610	77%	SDEU (100.0)	4,879	21%	ND		
VR2332 + 1-7-4 mixed sample	30	34	26	76%	JA894280.1 (100.0)	18	69%	ND					
	300	335	272	81%	KC469618.1 (99.9)	224	82%	1-7-4 (100.0)	38	14%	ND		
	3,000	3,351	2,699	81%	KC469618.1 (99.9)	2,287	85%	SDEU (100.0)	14	1%	1-7-4 (100.0)	348	13%
	30,000	33,512	26,917	80%	KC469618.1 (99.9)	22,668	84%	SDEU (100.0)	136	1%	1-7-4 (100.0)	3,633	13%

ND: not detected

From this study we also noticed that the percentage of PRRSV reads that mapped to the first BLAST hit could be used as an indicator for the presence of other PRRSV strains (Table 7). The samples that only contained VR2332 had > 98% of PRRSV reads mapping to VR2332, while in the mixed strain samples less than 85% of the PRRSV reads mapped to the first BLAST match, VR2332 (Table 7). Thus, less than 98% of reads mapping to the first BLAST hit would suggest the presence of a second strain (Table 7).

4. Discussion

PRRSV has been a severe threat to the swine industry worldwide ever since it was first described in the late 1980s (Albina 1997). Control of PRRSV is difficult, but important for animal welfare and swine production, where the development and implementation of reliable, accurate and rapid diagnostic methods play a key role. Several methods have been developed and applied to PRRSV diagnosis, which are well described by Ko, et al (Ko et al. 2015). Currently, PRRSV diagnostics mainly includes anti-PRRSV antibody detection by serological testing and nucleic acid detection using PCR based assays. Sequencing of PRRSV began in the mid-1990s, to discriminate between strains, which mainly focused on open reading frame 5 (ORF5) or other short regions of interest, but rarely encompassed the complete genome due to technological and monetary limitations (Murtaugh 2012; Kapur et al. 1996). PRRSV ORF5 shows extensive genetic diversity and has been used for providing insight into PRRSV epidemiology, however it is only 5% of the whole genome, thus 95% of the genomic information remains for prediction of genetic variation. Whole genome sequencing is greatly needed to provide a more complete picture of the virus (Slatko, Gardner, and Ausubel 2018; van Dijk et al. 2018), which is now gradually becoming more feasible with the rapid development and innovation of new sequencing

technologies (Lu, Archibald, and Ait-Ali 2014). Oxford Nanopore direct RNA sequencing (DRS) is revolutionary for sequencing RNA viral genomes, since it can sequence the RNA directly, allowing for detection of methylation sites and decreasing bias inherent in reverse transcription and PCR amplification of samples prior to sequencing, and it can generate long reads, allowing for the elucidation of recombination events (Viehweger et al. 2019).

This study was planned and performed to assess the feasibility of Oxford Nanopore MinION DRS in clinical PRRSV diagnostics to identify the viral strains involved in infection. The key interests addressed in this study included whether sequencing can detect PRRSV strains to identify an outbreak as occurring due to the introduction of a new strain or recirculation of a previous outbreak, whether sequencing can generate whole genome information to aid in further understanding of PRRSV epidemiology, and whether sequencing can detect and differentiate multiple strains in a single sample to investigate outbreaks that occur in vaccinated herds or co-infection of multiple field-strains. Previously, PRRSV whole genomes have been generated using Sanger and Illumina sequencing technologies (Han et al. 2019; Zhang et al. 2017; Kvisgaard et al. 2013). While both sequencing technologies can generate whole PRRSV genome with more than 99.9% accuracy, the raw reads produced are usually less than 1500bp. As a result, in order to generate a PRRSV whole genome, multiple primer sets and multiple individual sequencing reactions are needed for Sanger sequencing which is labor and time consuming; or for Illumina, computing resource intensive genome assembly is needed which requires time and knowledge to perform effectively. Oxford Nanopore MinION sequencing, on the other hand, can generate ultra-long raw reads which are in theory only limited by input fragment length (Jain et al. 2018). This feature is beneficial, since it saves time and effort when

generating a whole genome sequence. In this study, we successfully generated PRRSV raw reads up to the length of the entire genome (15 kb) with an approximate 86% identity to the known input genome sequence. A bioinformatics approach was developed that used the longest raw read as a scaffold to effectively generate a consensus sequence, improving the accuracy to 96% identity of the input genome.

Sequencing can be incorporated as a supportive tool for PCR to aid in diagnostic strain level PRRSV detection. It has been reported that both Sanger and Illumina sequencing can accurately detect PRRSV strains present in a sample, but both require transcription of RNA into cDNA followed by PCR amplification prior to sequencing (Han et al. 2019; Zhang et al. 2017). Differing from this, MinION technology directly sequences RNA strands for detection of PRRSV strains. Most importantly, the MinION sequencer is cost-effective and easily accessible, without the investment of expensive sequencing and bioinformatics infrastructure. Despite the low raw read accuracy of direct RNA sequencing (~86%), which is the main concern with this technology, PRRSV strains were identified with 99.9% accuracy using as few as 5 raw reads (3kb total yield). This accurate strain-level detection, even though the sequence accuracy is low, allows for guidance on determining effective control methods due to the precise detection of the circulating strains on a farm.

Now knowing the potential of DRS for strain level detection of pathogens as determined through this study as well as others (Kilianski et al. 2015), we next investigated the analytical sensitivity of PRRSV detection to determine its usefulness for obtaining reliable sequencing results. Previous research examining analytical sensitivity of next-generation sequencing has reported sensitivities that are similar or less sensitive than RT-

qPCR (Prachayangprecha et al. 2014; Huang et al. 2019a), and the third-generation nanopore DRS has previously shown a sensitivity of 1.89×10^7 viral copies in an influenza virus study (Keller et al. 2018). Our results indicated that samples with a minimum of 10^4 to 10^6 viral copies, depending on the sample type, can be successfully sequenced to accurately identify strains after about 6 hours of sequencing. Although DRS is not as sensitive as PCR for use as a diagnostic tool identifying viral presence (Zhao et al. 2019; Hu et al. 2015), it can be used for further investigation of the strain causing an outbreak, either directly from high viral load samples or following amplification of virus in cell culture. Additionally, a very strong correlation was observed between the number of viral reads generated through sequencing and the starting number of viral copies, indicating sequencing reads can be predicted by viral copies in a sample and vice versa, which has been confirmed by other studies as well (Prachayangprecha et al. 2014). Interestingly, the observation that the sensitivity of sequencing was higher from cell culture virus spiked into serum as opposed to clinical serum samples suggests that sample handling or perhaps the quality of the sample was an important factor for sequencing sensitivity (Relova et al. 2018), thus emphasizing the importance of careful handling, transporting and storing of clinical samples to protect the viral RNA from degradation (Houseley and Tollervey 2009; Gonzalez-Perez et al. 2010). This also suggests that on-site sequencing of samples as opposed to a centralized diagnostic system may allow for higher sensitivity of detection due to the ability to immediately process samples after sampling.

In addition to a single strain infection, clinical situations have been shown to be more complicated, sometimes involving infection with multiple strains simultaneously, such as co-infection of multiple field strains or co-existence of field strain(s) with vaccine

strain (Wang, Marthaler, et al. 2015; Zhao et al. 2015). This not only poses challenges to disease diagnosis but also increases the chance of PRRSV recombination, which is considered to be one of the most important mechanisms in PRRSV evolution (Han et al. 2019; Wang et al. 2018). In order to address this issue, Oxford Nanopore DRS was evaluated to determine if it could be used to discriminate co-infection by two PRRSV strains from different genotypes (61.4% similarity) as well as from the same genotype (82.4% similarity) in a single sample. In fact, the strains were easily differentiated, and the same method could be used to identify more than 2 strains present in a single sample.

This study begins the process of developing rapid and high-resolution PRRSV diagnostics for use in clinical situations where genomic data is urgently needed. This includes situations of potential infection, outbreak investigation, vaccine design guidance, and producer desires for more specific information. The PRRSV RNA genetic material presents the same technical demands for extraction, processing and sequencing as do influenza virus, coronaviruses, picornaviruses, rotaviruses, and many foreign animal disease viruses for which rapid pathogen identification and discrimination can be critically important. Knowledge gained from PRRSV in this study can be immediately translatable to aid in rapid diagnostic detection and strain-specific identification of an entire class of important swine pathogens. In fact, MinION sequencing technology might end up being a useful and affordable diagnostic tool for swine veterinary medicine in general. This technology can provide a complete readout of RNA viruses and RNA from other pathogens present in a sample without the need for pre-existing knowledge of what might be present (Theuns et al. 2018).

The current evaluation of this sequencing technology indicates that it can be used successfully along with qPCR for diagnosis of a pathogen, whole genome generation, strain-level pathogen detection and differentiation. As the DRS technology continues to develop and RNA isolations are optimized for use outside of a research laboratory, the future goal is to realize on-site infectious diseases investigation using the Oxford Nanopore MinION portable sequencer to allow for quicker diagnosis and facilitation of more rapid decision-making, an important consideration in an industry in which delays in moving animals due to unknown health status can disrupt flow patterns and schedules, or cause disease outbreaks with great economic losses.

Funding: This research was funded by National Pork Board Grant, grant number #18-040.

Acknowledgments: The authors would like to thank Dr. Sunil Kumar Mor and Tracy Otterson at the University of Minnesota Veterinary Diagnostic Laboratory (St.Paul, MN) for providing PRRSV clinical samples and ORF5 sequences. The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported within this paper. URL: http://www.msi.umn.edu.

Conflicts of Interest: The authors declare no conflict of interest.

Chapter IV - Characterization of emerging swine viral disease through Oxford Nanopore sequencing using Senecavirus A as a model

A manuscript in preparation

Emerging viral infectious diseases present a major threat to the global swine industry. Since 2015, Senecavirus A (SVA) has been observed as a cause of vesicular disease in different countries throughout the world and is considered an emerging disease. Despite the growing concern about this virus, there is a lack of information about SVA for development of prevention and diagnostic strategies, which is a problem for all newly emerging infectious diseases. Using SVA, we demonstrated that Oxford Nanopore MinION sequencing could be used as a robust tool for investigation of emerging infectious diseases caused by RNA viruses. Our results identified the presence of a pathogen from a clinical sample, allowing for identification at the species and strain level. SVA whole genome sequences were generated using both direct RNA sequencing and cDNA-PCR sequencing, with a consensus accuracy of 94% and 99% respectively. The advantages of direct RNA sequencing lie in its simplicity of library preparation and direct RNA strand information which can indicate potential nucleic acid modifications, while cDNA-PCR sequencing excelled at creating highly accurate sequences. This study developed whole genome sequencing methods to facilitate the understanding of SVA and provide a reference for investigations of other emerging infectious diseases.

1. Introduction

Emerging and re-emerging viral diseases have caused a significant adverse impact on swine production and will be an ongoing challenge. Emerging infections can be caused by previously unknown/undetected agents or known pathogens spreading to a new geographic locations or populations. The appearance of emerging diseases has been usually characterized by sudden unpredictable outbreaks, which then spread across regions and countries (Meng 2012).

The onset feature of emerging infectious diseases drives the need for effective disease management via rapid novel pathogen detection and efficient epidemiological surveillance (Devaux 2012; Burrell, Howard, and Murphy 2017; Jones et al. 2008). In the last 40 years, important emerging swine diseases include, but are not limited to, porcine reproductive and respiratory syndrome virus (PRRSV) first described in the late 1980s (Lunney, Benfield, and Rowland 2010), porcine circovirus type 2 (PCV2) discovered in the late 1990s (Segales, Kekarainen, and Cortey 2013) and more recent porcine epidemic diarrhea virus (PEDV) appearing in the early 2010s (Goede and Morrison 2016). The current introduction of African Swine Fever (ASF) into China again confirmed the devastating impact of viral diseases to the swine industry (Wang, Sun, and Qiu 2018). It has been estimated that by the end of 2019 half of China's pork supply, roughly a quarter of world's pork supply, will have been culled due to ASF disease (Loeb 2019). So far, ASF has spread from China to the neighboring countries and it is very likely to eventually enter other disease-free regions, such as the United States, despite all attempts to keep it out (Jurado et al. 2019). It is evident that globalization, increased travel and trade activities have been facilitating the transmission of pathogens and magnifying the adverse impact of

infectious diseases. The United States is now free of FMDV, ASF, and CSFV, but they are very likely to eventually arrive to the US in the closely connected world.

Senecavirus A (SVA), is a single-stranded, positive-sense RNA virus that is classified into the Senecavirus genus of the Picornaviridae family (Hales et al. 2008). The whole genome of SVA is approximately 7.3 kilobases (kb), which contains a 5′ untranslated region (UTR), a single open reading frame and a 3′ UTR with a poly(A) tail (Venkataraman et al. 2008). Previously, the study of SVA mainly focused on its oncolytic virotherapy (Burke 2016). More recently SVA has been detected in association with outbreaks of vesicular disease and neonatal mortality in pigs of many places worldwide including the United States, Brazil, Thailand and China (Leme et al. 2015; Wu et al. 2016; Canning et al. 2016; Saeng-Chuto et al. 2018). Due to its rapid spread SVA is now considered an emerging infectious disease (Segalés et al. 2017). So far, the risk of SVA causing an epidemic is still possible and the spread of SVA needs to be better controlled.

Effective surveillance and rapid response to SVA infections are imperative to reduce its adverse impact, which depend on the development of diagnostic tools. A variety of diagnostic methods have been developed, including virus isolation, direct detection in tissues by immunohistochemistry (Oliveira et al. 2017), serology assays such as enzyme-linked immunosorbent assay (Dvorak et al. 2017) and virus neutralization (Goolia et al. 2017), and nucleic acid detection using PCR (Leme et al. 2015; Fowler et al. 2017; Feronato et al. 2018; Zhang et al. 2019). The detection methods mentioned above play important roles in SVA detection and control. However, the limitation is that all methodologies are hypothesis-driven and need prior information about the pathogen detected, which will inevitably slow down the rapid identification of an unexpected

emerging virus or existing mutated virus where little or no information is present. PCR is considered to be a sensitive and accurate method, yet it suffers from inconsistent results as indicated by other studies, which suggests the necessity of supporting diagnostic tools for confirmation (Gimenez-Lirola et al. 2016). An additional complication is SVA-induced vesicular lesions look exactly like the ones caused by foot-and-mouth disease (FMD). FMD is an OIE-listed disease (http://www.oie.int) and a priority foreign animal disease for many countries, including the United States. Currently, the USDA actively investigates farms that show vesicular lesions to rule out the presence of FMDV, even when SVA is suspected (APHIS 2014). That makes the rapid and accurate detection of infectious agents responsible for vesicular lesions even more crucial (Segalés et al. 2017). A diagnostic method is needed for quick identification and confirmation of the presence or absence of FMDV to minimize the economic impact of infection on swine farm production.

High-throughput sequencing technologies have created a new possibility for surveillance and diagnostics of emerging viruses (Huang et al. 2019a). The robust, unbiased feature of sequencing allows for large-scale virus detection, as well as the discovery of unexpected even novel pathogens. Additionally, the subsequent whole genome generation from sequencing can enlarge the genome database for further investigation (Qian et al. 2016; Zhao et al. 2017). This is beneficial in the case of SVA, since very few SVA whole genomes are available in NCBI GenBank (i.e. 203 complete genomes as of Jan 2020). As more SVA genomes are sequenced and analyzed the more our understanding of SVA will improve, including SVA genotyping, molecular epidemiology, infection dynamics, and vaccine development.

Oxford Nanopore Technologies (ONT) MinION (ONT, Oxford, UK) is a single-molecule, long-read sequencer that determines bases by detecting ionic current changes as a DNA/RNA strand passes through a nanoscale protein pore (Loman and Watson 2015). The MinION sequencer is highly accessible to any research or diagnostic laboratory due to its low capital investment. As opposed to all other existing sequencing technologies, MinION is smaller than a smart phone, thus is potentially able to fulfill a point-of-care test in field environments (Jain et al. 2016), which has already been done and has been proven to work by several studies including Ebola, Zika and influenza surveillance and outbreak investigations (Quick et al. 2016; Rambo-Martin et al. 2019; Quick et al. 2017).

Viral infectious diseases include RNA viruses and DNA viruses. RNA viruses are a particularly high risk for causing the next emerging events, due to their high mutation rate and adaptability to new environment and hosts (Woolhouse and Brierley 2018). In this study we evaluated Oxford Nanopore sequencing for SVA investigation, aiming to provide insight on investigating emerging RNA viral diseases through sequencing. Oxford Nanopore provided two methodologies for RNA viral genome sequencing: traditional amplicon sequencing (cDNA-PCR sequencing, PCS), which has lower error rates and higher throughput, but involves reverse transcription and PCR amplification that is time consuming and loses some RNA genome structure information through the process; and direct RNA sequencing (DRS), which is an innovative technique under development that can sequence RNA strands directly, thus eliminating the length limitations possibly coming from reverse transcription and allowing for detection of nucleic acid base modifications. Of note, both methods used in this study are independent of previous sequence information such as is needed for primer design, however a poly(A) tail, which is present in the SVA

genome, is needed for adapter ligation. The approach lends itself readily to the sequencing of RNA viruses with the 3′ poly(A). Many of the swine RNA viruses have a 3′poly(A) tail, such as coronaviruses, picornaviruses, PRRSV/porcine arterivirus, and rhabdoviruses. However, sequencing of RNA pathogens (such as rotaviruses) which do not contain a poly(A) tail is possible through enzymatic addition of a 3′ poly(A) tail.

This study evaluated Oxford Nanopore DRS and PCS for detection and genome generation of RNA viruses using SVA as a model. SVA was successfully detected at the species and strain level from clinical samples. SVA whole genome sequence was able to be generated using cell culture grown viruses as well as clinical samples. The methods established here will accelerate more precise SVA detection in the field and increase the number of whole genome sequences available in GenBank to improve our understanding of the virus and thus better control of this disease. Additionally, this work will provide insights into the potential of the portable MinION sequencer as a diagnostic tool to aid in the investigation of emerging viral infectious diseases.

2. Materials and Methods

2.1. SVA samples

A Senecavirus A (SVA) lab isolate (GenBank: MN164664) and clinical samples of SVA positive vesicular fluid were provided by Dr. Fabio A. Vannucci at the University of Minnesota Veterinary Diagnostic Lab. The SVA lab isolate was propagated in cell culture in NCI-H1299 non-small cell lung carcinoma cell lines (ATCC CRL-5803) as previously described (Joshi et al. 2016).

2.2. RNA extraction

SVA RNA was extracted from cell culture SVA supernatants (cell culture samples), virus-free pig serum spiked with the SVA lab isolate (spike-in samples), and clinical vesicular fluids (clinical samples) using the QIAamp Viral RNA mini kit (Qiagen, Germantown, MD) following manufacturer's instructions without the addition of carrier RNA and with a final elution in 50ul nuclease-free water. Concentration of viral RNA was performed using a SpeedVac lab concentrator (Savant, NY, USA). A Qubit3.0 fluorometer (Life technologies, Carlsbad, CA) and a Nanodrop1000 spectrophotometer (Thermo Scientific, Waltham, MA) were used for RNA quantity and quality assessments.

2.3. Oxford Nanopore sequencing

Viral RNA was sequenced using 2 different kits, the direct RNA sequencing kit or the cDNA-PCR sequencing kit (ONT, Oxford, UK). The input RNA for the direct RNA sequencing (DRS) library preparation was extracted SVA RNA with the addition of the RNA calibration strand (RCS, 1314bp) which is provided in the sequencing kit. This RCS is Enolase II mRNA (YHR174W, NCBI Reference Sequence: NC_001140.6) provided at a concentration of approximately 50ng/ul. Library preparation was performed according to the direct RNA sequencing online protocol (ONT, Oxford, UK) (Garalde et al. 2018). The input RNA for the cDNA-PCR sequencing was the extracted SVA RNA only. CDNA-PCR sequencing (PCS) libraries were generated according to the cDNA-PCR sequencing online protocol (ONT, Oxford, UK). Briefly, 50ng or less poly(A)+ RNA was ligated to a poly(T) primer at 65 ℃ for 5min, followed by reverse transcription using a strand-switching primer and Maxima H Minus Reverse Transcriptase (ThermoFisher, Waltham, MA). Then, PCR was performed using a set of primers annealing to the end of Poly(T) primer and strand-

switching primer to amply the full-length transcripts. Exonuclease I (New England Biolabs, Ipswich, MA) was added to cleave nucleotides from the end of the PCR products, which were then purified by adding a 0.8-fold of Agencourt AMPure XP beads (Beckman Coulter, Brea, CA) and eluted in 12µl elution buffer. Following this, sequencing adapters were ligated to the amplified cDNA library. Immediately before sequencing, 12µl of the RNA library was mixed with 25.5µl of loading beads and 37.5µl of running buffer, yielding 75µl of final library. After library preparation, the DRS and PCS libraries were loaded onto a R9.4.1 SpotON flow cell and sequenced using a MinION Mk I sequencer (ONT, Oxford, UK) which was connected to a computer and remotely controlled by the MinKNOW software (ONT, Oxford, UK).

For genome sequencing of the cell culture SVA lab isolate, two sequencing replicates were performed for both DRS and PCS, with DRS starting with 60ng SVA RNA plus 300ng RCS and for PCS 60ng SVA RNA. The RCS was added to the DRS sequencing because more input RNA (~500ng total RNA) is suggested for optimized results. For sequencing runs of clinical samples and negative pig serum spiked with the SVA lab isolate, the same amount of SVA RNA was used for DRS and PCS library preparation, with the addition of 300ng RCS added to each of the DRS samples to increase the amount of total RNA present. The estimated sequence yield was monitored in real-time and samples were sequenced for approximately 6 hours with more/less time adjusted if needed to obtain the desired yield.

2.4. Bioinformatics analysis

Basecalling was carried out using guppy (ONT, Oxford, UK). Only raw reads that passed the quality filter of Phred quality score ≥ 7 (pass reads) were used for downstream

analysis. For DRS, the RNA Control Strand (RCS, 1314bp), which was used to enhance library generation and sequencing performance, was filtered out of the desired raw reads by turning on the corresponding guppy parameter. Total yield, pass reads yield, read quality, and read length of raw reads from whole genome sequencing were analyzed using MinIONQC (Lanfear et al. 2019), a script written in R to provide quality control for Oxford Nanopore data.

For sequencing of cell culture samples, raw pass reads were mapped to the SVA reference genome (GenBank: MN164664) using minimap2 (Li 2016), then analyzed using Qualimap (Okonechnikov, Conesa, and Garcia-Alcalde 2016), generating raw error rates and coverage information which was visualized using GraphPad prism software (GraphPad Software, San Diego, CA). The reads that mapped to the SVA reference genome (SVA reads) were extracted using SAMtools (Li et al. 2009), and the SVA yield, average reads length and quality were determined using NanoPlot (Van Broeckhoven et al. 2018).

For the clinical and spike-in samples, taxonomic analysis at the species level was performed to identify pathogens existing in the sample using what's in my pot (WIMP), which is provided by ONT's subsidiary Metrichor (Juul et al. 2015). An SVA custom database was created to analyze SVA sequencing reads by downloading all of the SVA whole genomes from GenBank (132 complete SVA genomes as of March 2019). To detect SVA at the strain level, pass reads were analyzed against this SVA database using Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990) to find the strain with the best match.

2.5. SVA consensus generation

Comparison was first performed among different assemblers, Canu (Koren et al. 2017), Miniasm (Li 2016), Racon (Vaser et al. 2017), and wtdbg2 (Ruan and Li 2019), using reads from direct RNA sequencing and cDNA-PCR sequencing. The best assembler for each sequencing method was selected to generate consensus. For DRS, a consensus sequence was generated, without need for reference or assembly, by extracting the longest read among all sequencing reads as a scaffold and mapping all pass reads to this longest read sequence using minimap2 (Li 2016) followed by consensus generation using Racon (Vaser et al. 2017). For PCS, de novo assembly of pass reads was performed using the Canu assembler (Koren et al. 2017). Evaluation and optimization of consensus generation for both sequencing methods was performed in terms of input, total sequence yield, and pre-treatment of raw reads using the cell culture virus in which the whole genome sequence was already known. Groups containing different yields, ranging from 70 kilobases (kb) to 7 megabases (mb), were generated by random selection using fastq-tools (https://homes.cs.washington.edu/~dcjones/fastq-tools/) from the total pass reads dataset. In the same yield group, three subgroups were formed using different raw reads filters; 1) original pass reads without further filters, 2) pass reads with read length > 1314bp to get rid of short reads, and RCS in the case of this study, and 3) pass reads that can be mapped to the SVA database. The consensus length and accuracy were the two main parameters evaluated for comparison. Consensus accuracy was determined by a comparison of the consensus genome to the reference genome and was analyzed by the ClustalW pairwise alignment in Geneious 8.0.5 software (https://www.geneious.com) (Kearse et al. 2012).

2.6. Sanger sequencing and analysis

Primers were designed for PCR amplification of the 3' end (3' UTR, 3D and partial 3C genomic regions) of SVA similar to previous studies (Wu et al. 2017) (forward primer_1 5' GGGTGACGACTTACAAGGGA 3', reverse primer_1 5' GAGCCAGTGCCGTGTGAAGAGT 3'; forward primer_2 5' CGCCAAGTTTCAATCCCATC 3', reverse primer_2 5' TCCCTTTTCTGTTCCGACTG 3'). SVA RNA was PCR-amplified using AccuStart PCR SuperMix (Quantabio, Beverly, MA) (Dvorak et al. 2013). PCR products were then cleaned up using ExoSAP-IT (Thermo Fisher Scientific, Waltham, MA). Sanger sequencing was performed on the PCR products at the University of Minnesota Genomics Center. Results were visualized and reads quality were checked using CLC Genomics Workbench 11.0 (https://www.qiagenbioinformatics.com/). The consensus sequence for each clinical sample was generated using Geneious (Kearse et al. 2012).

2.7. Analytical sensitivity determination

Analytical sensitivity was determined for both sequencing methods using spike-in and clinical samples containing a range of virus amounts. An SVA viral stock was 10-fold serially diluted from 1X to 10000X to generate decreasing amounts of viruses which were then added into SVA-free pig serum. The Ct value and viral copies for all samples were determined by RT-qPCR at the University of Minnesota Veterinary Diagnostic Lab. Eight clinical samples ranging from 10^2 to 10^6 viral copies/ml (Ct value range 24 to 13). Viral RNA was extracted from 1ml of sample with half of the sample used for direct RNA sequencing (DRS) and half for cDNA-PCR sequencing (PCS).

Taxonomic analysis was performed using WIMP, which allowed for pathogen detection, including SVA, at the species level. To further detect SVA at the strain level, BLAST analysis was carried out as described above. For spike-in samples, the sequence of the detected strain was determined using BLAST and then compared to the known reference genome using the ClustalW pairwise alignment in the Geneious software (Kearse et al. 2012) to identify the accuracy of the strain level detection. For clinical samples, the consensus sequence determined using Sanger sequencing was used as the reference. Detection accuracy was determined by comparing the strain identified as the best BLAST match to the reference sequence obtained from Sanger sequencing using ClustalW pairwise alignment (Larkin et al. 2007).

A consensus sequence was able to be generated for most samples. In the spike-in and clinical samples, SVA reads first needed to be separated from background sequence for successful SVA consensus generation. SVA reads were extracted by mapping all of the reads against the custom SVA database (generated above), followed by consensus generation using Racon for DRS and Canu for PCS as described above. The consensus length and accuracy were calculated to indicate the performance of consensus generation using varying amounts of viral copies.

To test if DRS and PCS were quantitative diagnostic methods, correlation analysis was performed. The total number of reads varied for each sequencing reaction, thus reads were normalized by calculating the ratio of SVA reads / total reads in order to compare between samples. Linear regression analysis was then performed to determine if there was any correlation using the reads/total reads ratio and the amount of input viral copies using GraphPad prism software (GraphPad Software, San Diego, CA).

3. Results

3.1. Raw reads assessment from direct RNA sequencing and cDNA-PCR sequencing

In order to evaluate and compare the general performance of DRS and PCS, two whole genome sequencing runs were carried out for each method. The available pores for each sequencing run were recorded to indicate the condition of the flow cell. All runs started with 60 ng of SVA RNA for library preparation and sequenced for 6 hours. SVA reads were extracted and analyzed. PCS had a much better performance than DRS in terms of SVA yield (66.1mb vs 4.5mb), average read length (1721bp vs 1267bp), and lower raw error rates (11.23% vs 15.14%) (Table 8). These differences could be explained by the intrinsic feature of Oxford Nanopore DNA sequencing and RNA sequencing, where the latter is a novel technology still under development. The main reason for higher SVA yield of PCS was that the two methods started with same amount of SVA RNA, but PCS involved a PCR amplification step increasing the number of SVA DNA strands.

Table 8. A comparison between sequencing statistics of the direct RNA and cDNA-PCR methods after 6h sequencing.

Sequencing statistics	direct RNA sequencing			cDNA-PCR sequencing		
	#1	#2	average	#1	#2	average
No. of available pores	474	332	403	434	408	421
Total pass yield mb	68	68	68	109	75	92
SVA yield mb	4.9	4.1	4.5	77.5	54.6	66.1
No. of SVA reads	3812	3306	3559	45964	31124	38544
Mean SVA read length	1295	1239	1267	1687	1754	1721
SVA reads error rates	14.91%	15.36%	15.14%	11.06%	11.39%	11.23%

Coverage analysis showed that PCS is able to generate more even coverage distribution than DRS (Figure 7). The uneven distribution of DRS has been observed previously (Tan, Dvorak, and Murtaugh 2019; Keller et al. 2018). Possible reasons for

uneven DRS coverage include partially degraded RNA and RNA second structures hampering the movement of the RNA through the nanopores.

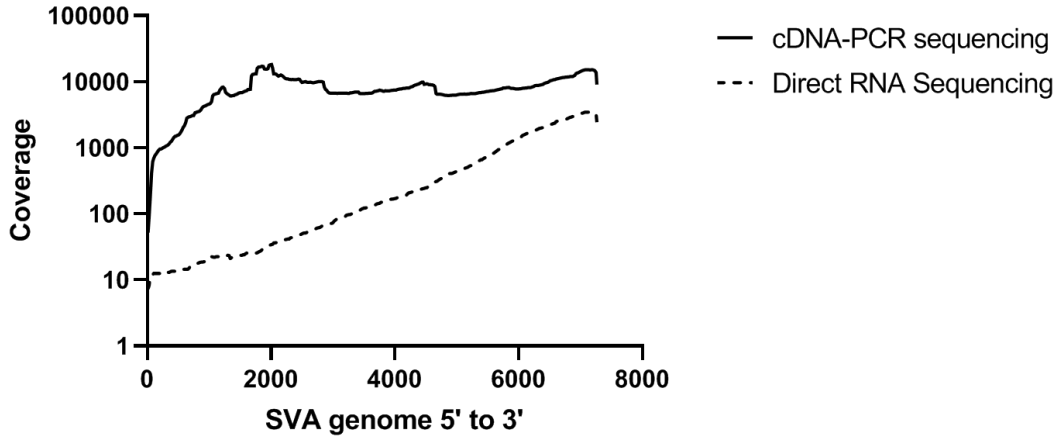

Figure 7. Coverage distribution of direct RNA sequencing and cDNA-PCR sequencing.

SVA reads were mapped to reference genome by minimap2 and analyzed by Qualimap to generate coverage information file which was then visualized by GraphPad prism software. The solid black line represents PCR-cDNA sequencing reads distribution over SVA genome, and the dash black like represents direct RNA sequencing reads distribution.

3.2. Consensus generation and comparison for direct RNA sequencing and cDNA-PCR sequencing

From the comparison of different assemblers in terms of consensus length and accuracy, PCS datasets worked best with Canu (Koren et al. 2017), and DRS datasets worked best with Racon (Vaser et al. 2017). Following this, we determined the sequence yield and pre-assembly reads filter needed for optimized consensus generation. After basecalling, different yield datasets (0.7, 7, or 70mb) were generated by randomly selecting reads from the total pass reads dataset. For each yield dataset, three groups were generated based on different filters, with group_1 containing all the pass reads (quality > 7), group_2 consisting of pass reads with a length filter >1314 bp, and group_3 with pass reads that mapped to the SVA database. The rationale behind the length filter was to test if a dataset with longer reads on average would help consensus generation. Also, the RCS which was

added during DRS library preparation was 1314bp, thus this cutoff not only removed short reads but also all RCS sequences. The rationale for use of the mappable filter was the assumption that a "less noisy" dataset would be beneficial for SVA assembly and consensus generation, especially in a clinical sample setting where large portions of sequencing reads may be host mRNA. The dataset statistics were first evaluated. For DRS, the yield after the length filter was 11% of the total input pass reads and after the mappable filter was 7%. The low recovery was due to the removal of RCS reads. For PCS, the recovery rate after the length filter and mappable filter was 75% and 73%, respectively (Table 9). The average read length and Phred quality score of PCS is greater than DRS. The average read length of DRS group_1 is especially low due to the presence of short RCS reads (Table 9).

Table 9. Sequencing statistics of datasets using different reads filters.

Sequencing statistics		direct RNA sequencing			cDNA-PCR sequencing		
		#1	#2	average	#1	#2	average
Group_1 Read quality > 7	Yield mb	68	68	68	70	70	92
	Average length bp	619	622	621	1519	1492	1506
	Average quality	8.7	8.2	8.5	8.6	8.5	8.6
Group_2 Read quality > 7 & length > 1314 bp	Yield mb	7.5	6.9	7.2	52	53.1	52.55
	Average length bp	2338	2289	2314	2960	2994	2977
	Average quality	8.6	8.6	8.6	8.9	8.9	8.9
Group_3 Read quality > 7 & mapped to SVA database	Yield mb	4.9	4.1	4.5	49.8	51	66.1
	Average length bp	1295	1239	1267	1691	1760	1726
	Average quality	8.3	8.2	8.3	8.6	8.5	8.6

For the DRS groups, as the starting yield increased, the consensus length and accuracy increased (Table 10). The highest consensus length and accuracy were observed at the 70mb yield (Table 10). At the same yield level, the performance of consensus generation with different filters were similar, indicating that the sequencing yield is the

leading factor for consensus generation and the raw reads filters didn't have much influence on results. A similar observation was observed for PCS sequencing, as within a sequencing yield, the different filters showed similar consensus generation results. However, for PCS an increase in yield didn't always result in better consensus performance, as 70 mb pass reads generated lower accuracy and a shorter consensus than the 7 mb read group did (Table 10). The best consensus for PCS was generated using a total sequencing yield of 7 mb (~5mb SVA reads). For both DRS and PCS, since there were no obvious advantages of applying extra filters to our original pass reads datasets, pass reads were used for further consensus evaluation and comparison.

Using the optimized consensus sequences generated from DRS in the 70mb total yield group (~5mb SVA reads) and PCS in the 7mb total yield group (~5mb SVA reads), a comparison was conducted in terms of consensus length and accuracy. While both DRS and PCS can generate nearly full length SVA, the consensus from PCS can achieve 99% accuracy, much higher than that of DRS in which the highest consensus accuracy was 94% (Table 10). Of note, in our research there is no obvious difference among the three different filter groups in the same yield group, however, it doesn't mean that reads filters were unnecessary for all situations. In some clinical samples, consensus generation may by improved by only using mappable SVA reads, especially if a large amount of host mRNA is present in the samples. We used the SVA mappable reads (Group_3 filter set) for consensus generation from following spike-in and clinical samples.

Table 10. Performance of consensus generation using different raw reads filters and at different starting yield.

Groups		*Group_1*						Group_2						Group_3					
Reads filters		*Read quality > 7*						Read quality > 7 & length > 1314 bp						Read quality > 7 & mapped to SVA database					
Runs		#1		#2		Average		#1		#2		Average		#1		#2		Average	
	Yield mb	Length bp	Accuracy	Length bp	Accuracy	Length bp	Accuracy	Length bp	Accuracy	Length bp	Accuracy	Length bp	Accuracy	Length bp	Accuracy	Length bp	Accuracy	Length bp	Accuracy
DRS	0.7	5654	87.6%	4542	91.0%	5098	89.3%	6103	88.5%	3659	91.1%	4881	89.8%	5248	88.6%	6865	83.7%	6057	86.2%
	7	7152	91.8%	7157	89.7%	7155	90.8%	6391	93.3%	4653	89.0%	5522	91.2%	7059	89.9%	7123	91.7%	7091	90.8%
	70	7150	94.4%	7175	94.1%	7163	94.3%	7075	94.6%	7117	94.1%	7096	94.4%	7095	94.7%	7124	94.0%	7110	94.4%
PCS	0.7	7038	98.9%	6438	99.0%	6738	99.0%	6698	99.1%	5933	95.7%	6316	97.4%	6246	97.3%	6937	98.7%	6592	98.0%
	7	7173	98.6%	7360	99.2%	7267	98.9%	7251	99.1%	7224	98.9%	7238	99.0%	6965	99.0%	7192	99.0%	7079	99.0%
	70	3077	82.3%	3028	98.6%	3053	90.5%	5250	87.8%	7941	95.4%	6596	91.6%	6057	86.5%	7464	89.3%	6761	87.9%

3.3. Analytical sensitivity of direct RNA sequencing and cDNA-PCR sequencing

Selection of certain diagnostic methods depends on the features of a diagnostic method and the purpose of an investigation. Analytical sensitivity of a diagnostic method is necessary to know in order to guide diagnostic use depending upon what is needed in different situations. The analytical sensitivity of nanopore DRS and PCS was evaluated using spike-in and clinical samples over a range of virus concentrations from 4.7×10^2 to 1.0×10^7 total viral copies.

After sequencing, the number of total reads from each run was determined (Table 11). In order to detect SVA at species level in an unbiased, hypothesis-free way. Taxonomic analysis was performed using WIMP. The number of reads classified as SVA was recorded (Table 11). In general, results showed that SVA can be detected using both method in spike-in and clinical samples with viral copies of 10^4 or more. Using DRS to investigate spike-in samples, we could detect SVA reads using as low as 4.7×10^2 SVA viral copies, while for PCS, we could detect SVA reads with viral copies of 1.2×10^4. In clinical samples, the detection limit was 9.2×10^2 viral copies for DRS and 2.2×10^3 for PCS. The number of total reads indicated the overall performance of sequencing, while the SVA reads suggested the presence and abundance of SVA in a sample (Table 11).

We noticed that the total reads generated from DRS varied a lot from 10^5 to 10^6 using similar sequencing time. PCS total reads varied from 10^4 to 10^6 (Table 11). Of note, SVA was not detected using PCS from a clinical sample with viral copies of 1.2×10^4 but can be detected in a spike-in sample with viral copies of 1.2×10^4 and clinical samples with less viral copies (7.6×10^3, 2.3×10^3, and 2.2×10^3) (Table 11). This could be explained by the few total reads generated. Even though 1.2×10^4 is 5 time of 2.2×10^3, the total

reads generated of the former sample is 16 times less than the latter, which makes it reasonable that the sample with more viral copies didn't detect SVA while the sample with less did detect SVA. Our observation of varying total and SVA reads on the other hand confirmed the inconsistency condition of the current flow cells, that would lead to inconsistent sequencing output for each run. Our experience agreed with observation by other researchers (Lu, Giordano, and Ning 2016). In order to minimize the inconsistency between runs, our suggestion is sequencing yield should be used as a target instead of sequencing time. This is easy to perform because we can monitor the sequencing process and approximate yield from MinKNOW interface in real-time.

In order to investigate whether SVA can be determined at strain level by DRS and PCS, total reads were BLASTn analyzed against SVA database, the best match was detected and was considered as the SVA strain existing in the sample (Table 11). The similarity between the detected strain sequence and the reference was identified to show the accuracy of the strain level detection. For spike-in samples, a lab isolate of known sequence was used and was compared to the consensus sequence or top BLAST hit to evaluate the detection accuracy. For each of the clinical samples, a partial genome sequence was obtained using Sanger sequencing which was then used to determine detection accuracy. The detection for SVA strain was accurate using both DRS and PCS methods, since the top hit from BLAST analysis was the reference strain for all spike-in samples. The average strain-level detection accuracy of clinical samples, which was determined by comparing sequence of BLAST top hit and reference sequence from Sanger sequencing, was 97.6% for both methods.

SVA reads were extracted by mapping total reads to the SVA custom database and the number mappable SVA reads was determined (Table 11). Consensus sequence was generated for all samples using the extracted SVA reads. Consensus sequences were generated using Racon for DRS and Canu for PCS as described above. The results showed that from a spike-in sample containing 1.2×10^6 viral copies, a nearly whole SVA (breadth of coverage > 95%) consensus was generated from both sequencing methods with an accuracy of 91% for DRS and 99% for PCS. Similar results were observed for a clinical sample containing 1.1×10^6 viral copies, which generated SVA consensus sequences (breadth of coverage > 97%) that had an accuracy of 99.6% or 97.0% (Table 11). For samples containing less than 10^6 viral copies, lower consensus accuracy was observed. A similar limit of detection and consensus generation for the DRS and PCS sequencing was observed, with DRS being slightly more sensitive and PCS generating slightly more accurate consensus sequences.

Table 11. Analytical sensitivity comparison of DRS and PCS

					Species detection		Strain detection				Consensus generation					
	Ct value	Viral copies	No. of total reads		SVA reads from WIMP		blast_DRS		blast_PCS		SVA reads from minimap		DRS_consensus		PCS_consensus	
			DRS	PCS	DRS	PCS	top blast	Identity%	top blast	Identity%	DRS	PCS	Length bp	Accuracy %	Length bp	Accuracy %
Spike-in	10	1.0×10^7	94572	16515	5722	2913	MN164664	100.0	MN164664	100.0	42936	4658	7189	94.8	7395	99.2
	13	1.2×10^6	40555	5775	236	1182	MN164664	100.0	MN164664	100.0	1961	1611	7177	91.0	6873	99.0
	18	6.5×10^4	185259	5071	17	296	MN164664	100.0	MN164664	100.0	43	420	2551	90.6	6770	99.4
	20	1.2×10^4	219789	3889	3	3	MN164664	100.0	MN164664	100.0	10	3	1306	90.3	916*	80.8
	25	4.7×10^2	110641	7880	1	0	MN164664	100.0	NA	NA	1	0	456*	85.1	NA	NA
Clinical	13	1.1×10^6	45478	78075	299	436	KX019804.1	97.1	KX019804.1	97.1	377	630	7157	97.0	7013	99.6
	16	1.3×10^5	148513	5256	83	7	KX019804.1	98.2	KU058182.1	97.8	124	9	3421	96.4	880*	89.0
	18	5.0×10^4	47411	174551	19	24	KY618836.1	97.4	KY618836.1	97.4	20	27	6534	97.1	2285*	95.3
	20	1.2×10^4	54964	5943	3	0	KU051394.1	97.9	NA		4	0	1957	82.4	NA	NA
	21	7.6×10^3	39465	242239	3	4	KY618835.1	97.7	MK256736.1	97.4	3	5	1158	94.3	988*	87.0
	22	2.3×10^3	53359	96661	1	4	KY618836.1	97.4	KT827250.1	97.7	1	7	511*	88.8	7745*	85.0
	23	2.2×10^3	58929	633632	4	1	KX019804.1	98.2	MH634514.1	98.2	4	1	3206	92.1	1171*	94.5
	24	9.2×10^2	41645	135552	1	0	MH490944.1	97.0	NA	NA	1	0	300*	80.0	NA	NA

*Longest read was used because no consensus was generated.

We also investigated if there was a quantitative relationship between the output SVA reads and the input SVA viral copies. In order to minimize the inter-sequencing variations, SVA reads were normalized for each sequencing run based on the total reads generated. Thus, a correlation analysis between the ratio of SVA reads/total reads and SVA viral copies was performed. The results showed that DRS has a stronger linear regression with an r^2=0.99 while PCS has a weak linear regression with an r^2=0.54, indicating DRS is a quantitative method while PCS is not (Figure 8).

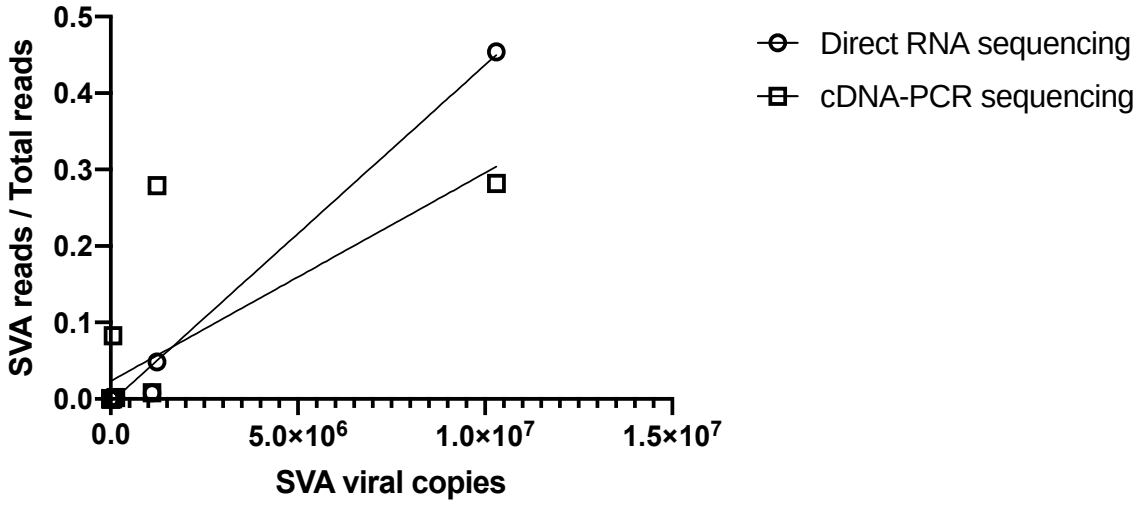

Figure 8. Correlation of SVA input viral copies and sequencing output.

Sequencing results from Table 10 were used to determine if there was a correlation between input viral copies per reaction with the ratio of SVA reads to total reads. Both direct RNA sequencing (DRS) and PCR-cDNA sequencing (PCS) were evaluated. For DRS $r^2 = 0.99$, $p < 0.0001$, and for PCS $r^2 = 0.51$, $p = 0.03$.

4. Discussion

Early and reliable detection of infectious agents soon after the onset of infection is essential for disease control as delays and misdiagnosis inevitably lead to increased spread of disease and escalation of adverse impacts. When a new pathogen emerges in a

population, there is usually no existing immunity, as well as no vaccine or specific treatment against it. While prompt diagnosis and treatment are especially crucial for emerging and re-emerging diseases, these infections are more likely to be missed during routine disease surveillance or adverse event investigation since most diagnostic methods are based on previously known agents and are not prepared for unexpected pathogens. Diagnostic methods that are rapid, available at the point-of-care, able to detect new pathogens, and robustly applicable across a wide range of pathogens are greatly needed to effectively fight against emerging eventualities (Caliendo, Gilbert, Ginocchio, Hanson, May, Quinn, Tenover, Alland, Blaschke, Bonomo, et al. 2013; Blaschke et al. 2015). Sequencing has been proven to be an accurate, precise, and hypothesis-free method for infectious disease investigation (Lefterova et al. 2015; Gwinn, MacCannell, and Khabbaz 2017a). It has been successfully used to investigate disease outbreaks (Houlihan et al. 2018), reveal genetic diversity among microbial populations (Zukurov et al. 2016) and discover emerging infectious agents (Kruppa et al. 2018). Among all pathogens, RNA viruses have the highest mutation rates, and are anticipated to have the highest possibility to cause the next emerging event (Woolhouse and Brierley 2018). Also, RNA viruses are of special concern regarding zoonotic transmission due to their high adaptability to new hosts (Carrasco-Hernandez et al. 2017).

In this study we used Oxford Nanopore sequencing for SVA detection at both the species and strain level in a research setting to develop a robust, on-site sequencing tool for investigation of SVA and to serve as a reference for investigation of other emerging RNA viral diseases. Our workflow involved a robust and unbiased species-level detection of pathogens via WIMP where we can get a general picture of all pathogens involved in a

sample. Combined with clinical symptoms and professional knowledge, suspected pathogen(s) can be targeted. Strain-level detection of the targeted pathogen(s) was then performed to provide more information about the investigation. We also optimized whole genome generation of SVA to provide support to expand the SVA database for further research of pathogenic mechanism or disease epidemiology.

Table 12. Summary of direct RNA sequencing and cDNA-PCR sequencing.

	cDNA-PCR sequencing	Direct RNA sequencing
Laboratory time (sample prep)	5 h	3 h
Laboratory time (generation of sequences)	6 h	6 h
Amount of RNA required for input	2 ng	500 ng*
Pass yield per hour mb	11	15
Sensitivity (derivation of sequences from low viral copies samples)	10^3 to 10^4	10^2 to 10^3
Recommended consensus generation program	Canu	Racon
Raw reads accuracy	89%	85%
Consensus accuracy	99%	94%
Consensus genome coverage	100%	100%
Depth of sequencing distribution	Even distribution	Coverage biases
Key attributes	Accessible, portable,	Rapid, accessible, portable potential RNA structure detection
Key concerns	need for amplification, time consuming	High requirement for input RNA amount, high error rate

*RCS or addition of exogenous RNA can be used to make up the majority of the RNA needed.

In addition, we provided thorough comparisons for PCS and DRS, which are summarized in Table 12, aiming to provide guidance on selection of a sequencing method when in different clinical situations and for different purposes. PCS takes more time but can generate a more accurate consensus, the advantage of which was especially obvious with higher viral copy number samples. DRS is less accurate, just as sensitive, but has unique and promising features such as detection of nucleic acid modifications as confirmed by other studies (Ovcharenko and Rentmeister 2018; Viehweger et al. 2018). Despite the differences between read accuracy, both methods can detect SVA at strain level correctly

using raw reads without the need for any further bioinformatics analysis. This allows anyone to run the assay easily since one of the limitations of sequencing as a diagnostic tool is its high requirement for data management and analysis.

When it comes to analytical sensitivity of sequencing diagnostics, it is important to define sequencing time, since longer time means more reads thus higher sensitivity. In our study, we sequenced samples for 6 hours in order to set a shorter turnaround time frame for generation of same day results. During disease outbreaks, time from sample collection to a clinical report should occur as quickly as possible and same day results would be ideal. Surprisingly, the analytical sensitivity of DRS and PCS was shown to be similar, approximately 10^2 to 10^4 viral copies, with DRS being slightly more sensitive than PCS. Other sequencing studies have discovered a detection limit of 10^2-10^3 genome copies/ml for influenza virus detection when sequencing samples for 48h instead of 6h (Lewandowski et al. 2019). It can't be emphasized enough the importance of short time frame on clinical management decisions. Because diseases can spread rapidly without infectious agents identified and proper control, especially when there is usually no effective vaccine or treatment available in the case of emerging pathogens. Of note, we observed that direct RNA sequencing is a quantitative assay while cDNA-PCR sequencing (amplicon sequencing) is not. Similar conclusions were drawn from other research groups. In one influenza virus research the authors concluded that there was a strong relationship between the viral titer and the proportion of influenza reads using Oxford Nanopore direct RNA sequencing technology (Lewandowski et al. 2019). However, in a hepatitis B virus (HBV) study using Oxford Nanopore amplicon sequencing, authors concluded that the amplicon sequencing is not quantitative where they observed considerable variability in total yields

and the proportion of mapped HBV reads between sequencing runs (McNaughton et al. 2019). The reason could be that cDNA-PCR sequencing included more steps during library preparation, including amplification and PCR products selection which could possibly introduce bias, while the process of direct RNA library preparation is simple and straightforward.

This study evaluated the ability of MinION sequencing as a diagnostic tool using SVA as a model. The method established in this study provided a framework for investigation of other emerging viral diseases and can be applied to aid in real-world problems directly. There are a few limitations to this sequencing method. Our method of species and strain detection largely depend on the genome database, GenBank. While investigating emerging viral diseases that are caused by known viruses expanding to new hosts or geographical regions, GenBank will have the viral genome information. Some previously unknown, newly discovered pathogens won't be detected directly from the databases due to lack of sequence information. However, the information is present, and the emerging pathogen can still be determined by carefully examining the unclassified sequencing reads and determining if they are at all similar to previously identified pathogens.

So far, the analytical sensitivity of MinION sequencing is lower than that of diagnostic PCR assays, which are limited by availability of assays for detection of known pathogens. However, it has been indicated that with further development, short-read second-generation sequencing will have a sensitivity to detect viruses that is comparable to PCR assays (Huang et al. 2019b). We believe as MinION sequencing continues to develop, its ability for pathogen detection will increase. Even at the current state, MinION

sequencing can be used as a support for PCR detection to provide strain information for more effective disease control. At a high enough viral presence, sequencing can be used to generate whole genome sequence and expand the genome database.

The portability of the MinION sequencer makes it suitable for fieldwork. There is a lack of veterinary infrastructure, expertise, diagnostic laboratories and surveillance in developing countries or remote places where emerging diseases are more likely to occur (Jones et al. 2008). However, emerging viruses know no country. In fact, the next emerging swine pathogen is mostly likely to happen in those resource-limited regions due to the lack of biosecurity measures and diagnostic facilities (Davies 2012). With globalization, they can transmit easily across borders. Thus, in order to effectively prevent emerging diseases globally, high priority should be given to provide surveillance and diagnostics in areas that are far away from established diagnostic laboratories. While most sequencing is generally restricted to large university laboratories and reference laboratories, due to its low accessibility, high instrument cost, extensive pre-sequencing library preparation and complex after-sequencing bioinformatics analysis, the portable sequencer, MinION, appears to be a perfect solution to address the issue of poorly allocated resources, since it can be taken to places with limited access to the critical infrastructure. In fact, several field studies have been conducted to confirm such advantages of a portable sequencer. During the Zika Virus (ZIKV) outbreak of 2015–2016 in Brazil, the incidence of ZIKV infection and study ZIKV vector dynamics were tracked in real-time by researchers using MinION (Faria et al. 2016). During the Ebola outbreak in 2015, MinION sequencing device were taken to the site for disease detection, which demonstrated the possibility of real-time genomic surveillance in resource-limited settings (Quick et al. 2016). Research for Dengue

virus serotyping using LAMP's simplicity and MinION's portability has been determined for Dengue virus investigation in the field in developing countries (Yamagishi, Runtuwene, Hayashida, Mongan, Thi, Thuy, Nhat, Limkittikul, Sirivichayakul, Sathirapongsasuti, et al. 2017). While it is surely a promising revolution, challenges to the full realization of this potential still need to be addressed, including automation of wet lab experiments in the field.

Infectious diseases will continue to emerge, getting prepared for the next one is the first step (Morens and Fauci 2013). In this study, we demonstrated that the portability, easy-operation, and low-maintenance of the MinION platform is a promising tool for disease surveillance and emerging infectious agent detection (Russell et al. 2018).

Chapter V - General Discussion

Infectious disease is a major constraint for the swine industry. There is a complex matrix of swine infectious agents, including bacteria, viruses, parasites and protozoans. Among these pathogens, viruses and bacteria account for more than 75% of the top important pathogens (VanderWaal and Deen 2018). RNA viruses are especially tricky to deal with, since they can mutate rapidly and adapt to new environments quickly. The most devastating infectious diseases in the swine industry, such as PRRS, PED, FMD, etc., are caused by RNA viruses. At the same time, bacteria have been gaining more attention in recent years due to the emergence of antimicrobial resistance (AMR), which can be transferred to human beings. Details about the current challenges to combat infectious disease in the swine industry have been described in **Chapter I, Section 2**.

The bright side is that we have made continuous improvements in infectious disease control. The detection resolution of diagnosis has been improved continuously, from observation of clinical signs and necropsy, to the use of culture and microscopy for diagnosis (Smith and Wolfe 1980), then to the detection of the pathogens nucleic acid and proteins. Sequencing is a rapidly evolving technology that has been used for infectious disease surveillance and outbreak investigation. The advantages and application of sequencing have been thoroughly reviewed in **Chapter I , Section 4**. Throughout our research, we evaluated a portable sequencer, MinION (Oxford Nanopore Technologies), to aid in swine infectious disease management. Other sequencing platforms usually need a core facility mainly due to expensive investment for sequencing equipment and operation difficulties. In comparison, MinION has the features of low cost, portability, ease of access

and operation. As a result, MinION sequencing is very much a possibility for becoming a first-line strategy for pathogen detection and characterization in clinical and public health settings. Lack of thorough evaluation and validation is a major limitation of MinION sequencing; thus, we evaluated and provided benchmark for this technology.

To test the performance of MinION sequencing on bacterial genome characterization, AMR detection, and provide benchmarks for future clinical use, we performed comprehensive analysis and optimization of MinION sequencing using *S. suis* in **Chapter II**. The results showed that full-length *S. suis* consensus sequences with a 99.4% average accuracy can be generated easily following our pipeline of optimized DNA extraction, DNA treatments, whole genome sequencing, and bioinformatic analysis. The MLST can be predicted and AMR can be accurately detected using the consensus genomes. This study illustrated the advantages of MinION sequencing as an "all-in-one" diagnostic assay that can provide high diagnostic resolution and reveal quite a bit of functional information about a pathogen. The method is a revolution in disease diagnosis because of the depth of information it is able to reveal.

Control of RNA viral disease is of high priority in the swine industry. RNA viruses cause enormous adverse impact on pigs, among which PRRS is one of the most devastating diseases (Nathues et al. 2017; Neumann et al. 2005). In order to address some issues we are facing today, such as vaccine failure, co-infections by multiple strains, and continuously emerging virulent strains, we applied MinION sequencing to PRRSV investigation in **Chapter III**. We successfully detected single PRRSV strain and differentiated multiple strains in a single sample using MinION direct RNA sequencing. The strain-level detection makes sequencing a great support for PCR, which is the gold

standard for diagnosing infections currently in diagnostic labs, to investigate challenging clinical cases. In addition, we were also able to generate the PRRSV whole genome (15kb), which opens the door to additional types of analysis, such as pathogenic mechanisms, evolution, and disease epidemiology, thus facilitating effective disease control. The method can be immediately applied to rapid diagnostic detection and strain-specific identification of an entire class of important swine pathogens. In fact, MinION sequencing might end up being a useful diagnostic tool for swine veterinary medicine in general.

Of note, in this study we not only developed a rapid PRRSV strain-level detection and whole genome generation workflow. We also demonstrated the unbiased detection feature of sequencing, that can detect all pathogens and host genomes present in a sample. We used what's in my pot (WIMP) to reveal all pathogens in our clinical samples. This can be a useful first step for disease investigation in clinical samples to generate a general picture about the pathogen species composition. Ideally, a workflow for disease investigation should be robust first to detect all pathogens and then be specific to investigate pathogens of interest. As the WIMP program develops, strain-level detection and other genomic analyses are highly possible.

Emerging viral infectious diseases present a major threat to the global swine industry. SVA is considered an emerging swine pathogen. Like other emerging infectious diseases, there is a lack of understanding of SVA, and extensive research is needed to guide effective disease control. In **Chapter IV**, using SVA as a model, we provided a sequencing approach to detect swine emerging viral pathogens at the species and strain levels. Additionally, in this study two different sequencing methods, direct RNA sequencing and cDNA-PCR sequencing, were compared in terms of yields, accuracy, de novo assembly,

and analytical sensitivity. cDNA-PCR sequencing is the traditional approach to sequencing RNA that includes transcription and PCR amplification. Direct RNA sequencing is a new technology that can sequence the RNA strand directly, without the need for PCR amplification. Our main results showed that SVA whole genomes can be generated from both direct RNA sequencing and cDNA-PCR sequencing, with a consensus accuracy of 94% and 99%, respectively. Despite of the difference in sequencing accuracy, both methods can be used for species and strain level pathogen detection directly from clinical samples using just raw reads. The advantages of direct RNA sequencing lay in its short turn-around time, more sensitive pathogen detection, and direct RNA strand information present indicating detection of potential nucleic acid modifications. In comparison, cDNA-PCR sequencing excelled at its higher accuracy, especially at higher viral copies when it has better performance for whole genome generation than direct RNA sequencing. The comprehensive comparison of the two methods for RNA virus investigation highlighted their advantages and limitations, which can serve as a guidance for decisions about method selection for different situations and purposes. Basically, our results suggest that for low viral copy clinical samples, direct RNA sequencing should be performed; while for high viral copy clinical or cell culture viral samples, cDNA-PCR sequencing should be performed, since it can produce more accurate raw reads and generate consensus genomes with higher accuracy.

From our study, we demonstrated that sequencing combined with bioinformatics is indeed a revolutionary method for infectious diseases diagnosis. It can be incorporated into the current infectious disease diagnostic system for more accurate detection, characterization of pathogen and further epidemiology analysis. Oxford Nanopore

sequencing is rarely used for swine infectious disease diagnosis so far, but we demonstrated that it is a promising diagnostic tool to improve swine infectious disease management. The low device cost makes it highly accessible to diagnostic labs or individual farms. During sequencing, data can be streamed and analyzed in real-time. This can save time in during outbreak investigations. In addition to the feature of real-time data generation and analysis, MinION is a portable device that can be taken to the field, allowing for on-site diagnosis. The development of additional devices, such as VolTRAX, an automatic library preparation device to get your biological sample ready for analysis, Flongle, an adapter of MinION suitable for field work, and MinIT, a preconfigured, portable replacement for the sequencing laptop, making this technology progressive as compared to other sequencing technologies in its ability to be used for field applications. The limitations of Oxford Nanopore sequencing are acknowledged (**Chapter I, Section 4.3**), which includes high cost, lack of interdisciplinary experts, pathogen genome separation from host background, and data interpretation.

Applications of genome sequencing in infectious disease surveillance and outbreak detection will considerably benefit public health, in addition to swine health. Because pathogens of veterinary interest are not only related to animal health but have significant impact on public health. It has been generally accepted that most infectious diseases affecting human beings originated from pathogens of other mammalian species. There is a high chance of swine pathogens spread to humans and human pathogens spread to swine due to the intensity of swine farming worldwide and the close contact between pigs and humans (Uddin Khan et al. 2013). Similarly, AMR can be transmitted back and forth between hosts since microbials of humans and animals, which carry AMR genes, are

interconnected. AMR determinants can transfer from animals or the environment to human beings via direct/indirect contact or through the food chain (Ventola 2015b; 'The antibiotic alarm' 2013). The frequent transmission of pathogens and AMR among humans and other animals strongly indicates the importance and necessity for a "One Health" approach. In addition, the world is increasingly connected, which makes it convenient for disease transmission. The frequent international-scale infectious disease outbreaks in the swine industry called for global efforts for effective disease control. While technological development is the foundation for disease control, more important is the global collaboration across a variety of sectors (human, agricultural and environmental).

Undoubtedly, we are making great progress in controlling and preventing infectious diseases worldwide. Infectious diseases were even considered soon to be eradicated in the 1960s, due to the development of vaccination and use of antibiotics (Brachman 2003). However, adaptation of pathogens to environmental pressure encouraged the emergence and reemergence of infectious disease and created new challenges. Fortunately, new technologies and strategies are developed to control infectious disease effectively. Sequencing as a diagnostic tool can provide great support to the current diagnostic systems and has huge potential. In the future, a portable sequencer can be used as an on-site diagnostic tool for routine disease surveillance and outbreak investigation, to provide rapid insights for disease control. Challenges exist to realize this future goal. First, while sequencing as a complement to PCR can be used to investigate challenging clinical cases, more efforts are needed to realize the routine and independent use of sequencing as a diagnostic tool. The current analytical sensitivity of sequencing is still relatively low, compared with PCR. It can only detect clinical samples with medium to high viral titers.

So, sequencing can be used for situations such as disease outbreaks or cell cultured virus, where the number of viral copies is usually high. But the application of sequencing for routine surveillance is limited, since the viral titer is not always high enough in the sample. Secondly, sequencing can reveal almost everything present in a sample, which is an advantage for co-infection investigation and microbial community characterization. However, this can also generate cofounders and make it difficult to highlight the pathogens of interest. Especially for sequencing technologies like Oxford Nanopore sequencing that has low accuracy, the sequencing errors can be confused with real mutations. Various methods have been developed to enrich the pathogens under investigation and minimize the host material background. However, its efficiency is still questionable and needs to be improved. Next, for emerging pathogen detection, the challenges described above are an issue. In addition, if the pathogen genome is not present in GenBank, then it won't be identified, as most classification software is based on the GenBank database. The classification accuracy is also something to take into consideration. How similar/different two sequences are will allow for them to be classified into the same/different species or strains? Again, the accuracy of MinION sequencing is far from 100%, and this further complicates the classification process. Also, a "portable diagnostic lab" is attractive, since it can be used in regions that lack diagnostic capacities and it can allow for rapid disease investigation. To realize on-site infectious disease investigation, on-site extraction, library preparation, and sequencing are needed. In addition, electricity and Wi-Fi are needed for device operation and necessary data analysis. Finally, sequencing as a diagnostic tool needs further validation. In addition to analytical sensitivity which we evaluated in our study, diagnostic sensitivity, specificity, and repeatability need to be evaluated.

We are in a continuous battle with infectious disease. The fact that pathogens (especially RNA viruses) evolve faster than their host (us and our animals) puts us on the inferior side of the battle against infectious diseases. Biological evolution is like Pandora's Box, and novel agents will continue to emerge and surprise us. We need the ability to deal with the current diseases and prepare for upcoming ones. Especially due to the climate change, population expanding, frequent travel and integration between humans and animals, it is estimated that more and more emerging eventualities are going to happen. So, infectious disease is and will be a continuous threat to animal and human health. New technologies, such as sequencing, are needed to help better control infectious disease. The adverse impact of infectious disease nowadays can be profound, since the world is highly connected, allowing for quick movement of pathogens across the globe. Collaboration among regions and information sharing are essential, especially in the case of emerging diseases, to limit the disease spread. Swine infectious disease can not only influence swine health and pork production, it can also adversely impact human health, such as zoonotic disease and antibiotic resistance. A coordinated, "One Health" response is urgently needed for infectious disease control in general.

Bibliography

Akhtar, R., and N. Izhar. 1986. 'The spatial distribution of health resources within countries and communities: examples from India and Zambia', *Soc Sci Med*, 22: 1115-29.

Al-Muharrmi, Z. 2010. 'Understanding the Influenza A H1N1 2009 Pandemic', *Sultan Qaboos Univ Med J*, 10: 187-95.

Al Rwahnih, Maher, Adib Rowhani, Nathaniel Westrick, Kristian Stevens, Alfredo Diaz-Lara, Florent P Trouillas, John Preece, Craig Kallsen, Kristen Farrar, and Deborah Golino. 2018. 'Discovery of viruses and virus-like pathogens in pistachio using high-throughput sequencing', *Plant disease*, 102: 1419-25.

Albina, E. 1997. 'Epidemiology of porcine reproductive and respiratory syndrome (PRRS): an overview', *Vet Microbiol*, 55: 309-16.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. 'Basic local alignment search tool', *J Mol Biol*, 215: 403-10.

Ang, B. S. P., T. C. C. Lim, and L. Wang. 2018. 'Nipah Virus Infection', *J Clin Microbiol*, 56.

'The antibiotic alarm'. 2013. *Nature*, 495: 141-41.

APHIS, USDA. 2014. 'Foot-and-mouth disease response plan', *The Red Book*.

Ardui, S., A. Ameur, J. R. Vermeesch, and M. S. Hestand. 2018. 'Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics', *Nucleic Acids Res*, 46: 2159-68.

Ashton, P. M., S. Nair, T. Dallman, S. Rubino, W. Rabsch, S. Mwaigwisya, J. Wain, and J. O'Grady. 2015. 'MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island', *Nat Biotechnol*, 33: 296-300.

Bailey, E. S., J. Y. Choi, J. K. Fieldhouse, L. K. Borkenhagen, J. Zemke, D. Zhang, and G. C. Gray. 2018. 'The continual threat of influenza virus infections at the human-animal interface: What is new from a one health perspective?', *Evol Med Public Health*, 2018: 192-98.

Bainomugisa, A., T. Duarte, E. Lavu, S. Pandey, C. Coulter, B. J. Marais, and L. M. Coin. 2018. 'A complete high-quality MinION nanopore assembly of an extensively drug-resistant Mycobacterium tuberculosis Beijing lineage strain identifies novel variation in repetitive PE/PPE gene regions', *Microb Genom*.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2012. 'SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing', *J Comput Biol*, 19: 455-77.

Barrett, Ronald, Christopher W Kuzawa, Thomas McDade, and George J Armelagos. 1998. 'Emerging and re-emerging infectious diseases: the third epidemiologic transition', *Annual review of anthropology*, 27: 247-71.

Bartlett, John G, David N Gilbert, and Brad Spellberg. 2013. 'Seven ways to preserve the miracle of antibiotics', *Clinical Infectious Diseases*, 56: 1445-50.

Berger, Kathryn A, David M Pigott, Francesca Tomlinson, David Godding, Sebastian Maurer-Stroh, Biruhalem Taye, Fernanda L Sirota, Alvin Han, Raphael TC Lee, and Vithiagaran Gunalan. 2018. "The geographic variation of surveillance and zoonotic spillover potential of influenza viruses in domestic poultry and swine." In *Open forum infectious diseases*, ofy318. Oxford University Press US.

Bethesda, National Institutes of Health ;. 2007. 'Understanding Emerging and Re-emerging Infectious Diseases.', *Biological Sciences Curriculum Study*, NIH Curriculum Supplement Series [Internet]: https://www.ncbi.nlm.nih.gov/books/NBK20370/.

Blaschke, A. J., A. L. Hersh, S. E. Beekmann, D. Ince, P. M. Polgreen, and K. E. Hanson. 2015. 'Unmet diagnostic needs in infectious disease', *Diagn Microbiol Infect Dis*, 81: 57-9.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30: 2114-20.

Bowman, Andrew S, Rasna R Walia, Jacqueline M Nolting, Amy L Vincent, Mary Lea Killian, Michele M Zentkovich, Joshua N Lorbach, Sarah E Lauterbach, Tavis K Anderson, and C Todd Davis. 2017. 'Influenza A (H3N2) virus in swine at agricultural fairs and transmission to humans, Michigan and Ohio, USA, 2016', *Emerg Infect Dis*, 23: 1551.

Brachman, Philip S. 2003. "Infectious diseases—past, present, and future." In.: Oxford University Press.

Brealey, J. C., P. D. Sly, P. R. Young, and K. J. Chappell. 2015. 'Viral bacterial co-infection of the respiratory tract during early childhood', *FEMS Microbiol Lett*, 362.

Brockmeier, S. L., C. L. Loving, A. C. Vorwald, M. E. Kehrli, Jr., R. B. Baker, T. L. Nicholson, K. M. Lager, L. C. Miller, and K. S. Faaberg. 2012. 'Genomic sequence and virulence comparison of four Type 2 porcine reproductive and respiratory syndrome virus strains', *Virus Res*, 169: 212-21.

Brockmeier, Susan L, Crystal L Loving, Tracy L Nicholson, Jinhong Wang, Sarah E Peters, Lucy Weinert, Roy Chaudhuri, David J Seilly, Paul R Langford, and Andrew Rycroft. 2018. 'Use of proteins identified through a functional genomic screen to develop a protein subunit vaccine that provides significant protection against virulent Streptococcus suis in pigs', *Infection and immunity*, 86: e00559-17.

Bryant, Josephine, Claire Chewapreecha, and Stephen D Bentley. 2012. 'Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences', *Future microbiology*, 7: 1283-96.

Burke, M. J. 2016. 'Oncolytic Seneca Valley Virus: past perspectives and future directions', *Oncolytic Virother*, 5: 81-9.

Burrell, Christopher J., Colin R. Howard, and Frederick A. Murphy. 2017. 'Chapter 15 - Emerging Virus Diseases.' in Christopher J. Burrell, Colin R. Howard and Frederick A. Murphy (eds.), *Fenner and White's Medical Virology (Fifth Edition)* (Academic Press: London).

Caliendo, A. M., D. N. Gilbert, C. C. Ginocchio, K. E. Hanson, L. May, T. C. Quinn, F. C. Tenover, D. Alland, A. J. Blaschke, R. A. Bonomo, K. C. Carroll, M. J. Ferraro, L. R. Hirschhorn, W. P. Joseph, T. Karchmer, A. T. MacIntyre, L. B. Reller, and A. F. Jackson. 2013. 'Better tests, better care: improved diagnostics for infectious diseases', *Clin Infect Dis*, 57 Suppl 3: S139-70.

Caliendo, Angela M, David N Gilbert, Christine C Ginocchio, Kimberly E Hanson, Larissa May, Thomas C Quinn, Fred C Tenover, David Alland, Anne J Blaschke, and Robert A Bonomo. 2013. 'Better tests, better care: improved diagnostics for infectious diseases', *Clinical Infectious Diseases*, 57: S139-S70.

Canning, P., A. Canon, J. L. Bates, K. Gerardy, D. C. Linhares, P. E. Pineyro, K. J. Schwartz, K. J. Yoon, C. J. Rademacher, D. Holtkamp, and L. Karriker. 2016. 'Neonatal Mortality, Vesicular Lesions and Lameness Associated with Senecavirus A in a U.S. Sow Farm', *Transbound Emerg Dis*, 63: 373-8.

Cao, M. D., S. H. Nguyen, D. Ganesamoorthy, A. G. Elliott, M. A. Cooper, and L. J. Coin. 2017. 'Scaffolding and completing genome assemblies in real-time with nanopore sequencing', *Nat Commun*, 8: 14515.

Capobianchi, M. R., E. Giombini, and G. Rozera. 2013. 'Next-generation sequencing technology in clinical virology', *Clin Microbiol Infect*, 19: 15-22.

Carles Vilalta, Juan Sanhueza, Mariana Kikuti, Cesar Corzo. 2019. 'Porcine Reproductive and Respiratory Syndrome break recurrence', *https://www.pig333.com*.

Carrasco-Hernandez, R, Rodrigo Jácome, Yolanda López Vidal, and Samuel Ponce de León. 2017. 'Are RNA viruses candidate agents for the next global pandemic? A review', *ILAR journal*, 58: 343-58.
Cars, Otto. 2014. 'Securing access to effective antibiotics for current and future generations. Whose responsibility?', *Upsala journal of medical sciences*, 119: 209-14.
Centner, T. J. 2016. 'Recent government regulations in the United States seek to ensure the effectiveness of antibiotics by limiting their agricultural use', *Environ Int*, 94: 1-7.
Chan, Emily H, Timothy F Brewer, Lawrence C Madoff, Marjorie P Pollack, Amy L Sonricker, Mikaela Keller, Clark C Freifeld, Michael Blench, Abla Mawudeku, and John S Brownstein. 2010. 'Global capacity for emerging infectious disease detection', *Proceedings of the National Academy of Sciences*, 107: 21701-06.
Chantziaras, I., F. Boyen, B. Callens, and J. Dewulf. 2014. 'Correlation between veterinary antimicrobial use and antimicrobial resistance in food-producing animals: a report on seven countries', *J Antimicrob Chemother*, 69: 827-34.
Charalampous, Themoula, Hollian Richardson, Gemma L. Kay, Rossella Baldan, Christopher Jeanes, Duncan Rae, Sara Grundy, Daniel J. Turner, John Wain, Richard M. Leggett, David M. Livermore, and Justin O'Grady. 2018. 'Rapid Diagnosis of Lower Respiratory Infection using Nanopore-based Clinical Metagenomics', *bioRxiv*: 387548.
Choi, Young Ki, Sagar M Goyal, and Han Soo Joo. 2003. 'Retrospective analysis of etiologic agents associated with respiratory diseases in pigs', *The Canadian veterinary journal*, 44: 735.
Cleaveland, S, MK Laurenson, and LH Taylor. 2001. 'Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356: 991-99.
Collineau, L., P. Boerlin, C. A. Carson, B. Chapman, A. Fazil, B. Hetman, S. A. McEwen, E. J. Parmley, R. J. Reid-Smith, E. N. Taboada, and B. A. Smith. 2019. 'Integrating Whole-Genome Sequencing Data Into Quantitative Risk Assessment of Foodborne Antimicrobial Resistance: A Review of Opportunities and Challenges', *Front Microbiol*, 10: 1107.
Conceicao-Neto, N., M. Zeller, H. Lefrere, P. De Bruyn, L. Beller, W. Deboutte, C. K. Yinda, R. Lavigne, P. Maes, M. Van Ranst, E. Heylen, and J. Matthijnssens. 2015. 'Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis', *Sci Rep*, 5: 16532.
Cornelissen, M., A. Gall, M. Vink, F. Zorgdrager, S. Binter, S. Edwards, S. Jurriaans, M. Bakker, S. H. Ong, L. Gras, A. van Sighem, D. Bezemer, F. de Wolf, P. Reiss, P. Kellam, B. Berkhout, C. Fraser, and A. C. van der Kuyl. 2017. 'From clinical sample to complete genome: Comparing methods for the extraction of HIV-1 RNA for high-throughput deep sequencing', *Virus Res*, 239: 10-16.
Cromwell, G. L. 2002. 'Why and how antibiotics are used in swine production', *Anim Biotechnol*, 13: 7-27.
Davies, Peter R. 2012. 'One world, one health: the threat of emerging swine diseases. A North American perspective', *Transbound Emerg Dis*, 59: 18-26.
Depledge, D. P., K. P. Srinivas, T. Sadaoka, D. Bready, Y. Mori, D. G. Placantonakis, I. Mohr, and A. C. Wilson. 2019. 'Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen', *Nat Commun*, 10: 754.
Deurenberg, R. H., E. Bathoorn, M. A. Chlebowicz, N. Couto, M. Ferdous, S. Garcia-Cobos, A. M. Kooistra-Smid, E. C. Raangs, S. Rosema, A. C. Veloo, K. Zhou, A. W. Friedrich, and J. W. Rossen. 2017. 'Application of next generation sequencing in clinical microbiology and infection prevention', *J Biotechnol*, 243: 16-24.
Devaux, C. A. 2012. 'Emerging and re-emerging viruses: A global challenge illustrated by Chikungunya virus outbreaks', *World J Virol*, 1: 11-22.

Devi, Mrinalee, Jyoti B Dutta, Swaraj Rajkhowa, Dhireswar Kalita, Girindra Kumar Saikia, Bipin Chandra Das, Razibuddin Ahmed Hazarika, and Gauranga Mahato. 2017. 'Prevalence of multiple drug resistant Streptococcus suis in and around Guwahati, India', *Veterinary world*, 10: 556.

Dhiman, Neelam, David I Smith, and Gregory A Poland. 2009. 'Next-generation sequencing: a transformative tool for vaccinology', *Expert Rev Vaccines*, 8: 963-67.

Dixon, M. A., O. A. Dar, and D. L. Heymann. 2014. 'Emerging infectious diseases: opportunities at the human-animal-environment interface', *Vet Rec*, 174: 546-51.

Domingo, Esteban, Julie Sheldon, and Celia Perales. 2012. 'Viral Quasispecies Evolution', *Microbiology and Molecular Biology Reviews*, 76: 159-216.

Drew, T. W. 2011. 'The emergence and evolution of swine viral diseases: to what extent have husbandry systems and global trade contributed to their distribution and diversity?', *Rev Sci Tech*, 30: 95-106.

Duffy, Siobain, Laura A Shackelton, and Edward C Holmes. 2008. 'Rates of evolutionary change in viruses: patterns and determinants', *Nature Reviews Genetics*, 9: 267.

Dvorak, C. M., Z. Akkutay-Yoldar, S. R. Stone, S. J. Tousignant, F. A. Vannucci, and M. P. Murtaugh. 2017. 'An indirect enzyme-linked immunosorbent assay for the identification of antibodies to Senecavirus A in swine', *BMC Vet Res*, 13: 50.

Dvorak, C. M., M. P. Lilla, S. R. Baker, and M. P. Murtaugh. 2013. 'Multiple routes of porcine circovirus type 2 transmission to piglets in the presence of maternal immunity', *Vet Microbiol*, 166: 365-74.

El Kholy, AA, NA Mostafa, AA Ali, MMS Soliman, SA El-Sherbini, RI Ismail, N El Basha, RI Magdy, N El Rifai, and DH Hamed. 2016. 'The use of multiplex PCR for the diagnosis of viral severe acute respiratory infection in children: a high rate of co-detection during the winter season', *European Journal of Clinical Microbiology & Infectious Diseases*, 35: 1607-13.

Erbelding, Emily J, Diane J Post, Erik J Stemmy, Paul C Roberts, Alison Deckhut Augustine, Stacy Ferguson, Catharine I Paules, Barney S Graham, and Anthony S Fauci. 2018. 'A universal influenza vaccine: the strategic plan for the National Institute of Allergy and Infectious Diseases', *The Journal of infectious diseases*, 218: 347-54.

Estrada, A. A., M. Gottschalk, S. Rossow, A. Rendahl, C. Gebhart, and D. G. Marthaler. 2019. 'Serotype and Genotype (Multilocus Sequence Type) of Streptococcus suis Isolates from the United States Serve as Predictors of Pathotype', *J Clin Microbiol*.

FAO. 2019. 'Food Outlook: Biannual Reports on global Food Markets (Food Agric Organ United Nations, Rome)'.

Faria, N. R., E. C. Sabino, M. R. Nunes, L. C. Alcantara, N. J. Loman, and O. G. Pybus. 2016. 'Mobile real-time surveillance of Zika virus in Brazil', *Genome Med*, 8: 97.

Feng, Y., H. Zhang, Z. Wu, S. Wang, M. Cao, D. Hu, and C. Wang. 2014. 'Streptococcus suis infection: an emerging/reemerging challenge of bacterial infectious diseases?', *Virulence*, 5: 477-97.

Feronato, C., R. A. Leme, J. A. Diniz, A. M. D. Agnol, A. F. Alfieri, and A. A. Alfieri. 2018. 'Development and evaluation of a nested-PCR assay for Senecavirus A diagnosis', *Trop Anim Health Prod*, 50: 337-44.

Flahault, A., A. Geissbuhler, I. Guessous, P. J. Guerin, I. Bolon, M. Salathe, and G. Escher. 2017. 'Precision global health in the digital age', *Swiss Med Wkly*, 147: w14423.

Fleming, A. 1945. 'Penicillin's finder assays its future', *New York Times*, 26: 21.

Forsberg, R. 2005. 'Divergence time of porcine reproductive and respiratory syndrome virus subtypes', *Mol Biol Evol*, 22: 2131-4.

Fowler, V. L., R. H. Ransburgh, E. G. Poulsen, J. Wadsworth, D. P. King, V. Mioulet, N. J. Knowles, S. Williamson, X. Liu, G. A. Anderson, Y. Fang, and J. Bai. 2017.

'Development of a novel real-time RT-PCR assay to detect Seneca Valley virus-1 associated with emerging cases of vesicular disease in pigs', *J Virol Methods*, 239: 34-37.
Gallagher, M. D., I. Matejusova, L. Nguyen, N. M. Ruane, K. Falk, and D. J. Macqueen. 2018. 'Nanopore sequencing for rapid diagnostics of salmonid RNA viruses', *Sci Rep*, 8: 16307.
Garalde, D. R., E. A. Snell, D. Jachimowicz, B. Sipos, J. H. Lloyd, M. Bruce, N. Pantic, T. Admassu, P. James, A. Warland, M. Jordan, J. Ciccone, S. Serra, J. Keenan, S. Martin, L. McNeill, E. J. Wallace, L. Jayasinghe, C. Wright, J. Blasco, S. Young, D. Brocklebank, S. Juul, J. Clarke, A. J. Heron, and D. J. Turner. 2018. 'Highly parallel direct RNA sequencing on an array of nanopores', *Nat Methods*, 15: 201-06.
Garcia-Alcalde, F., K. Okonechnikov, J. Carbonell, L. M. Cruz, S. Gotz, S. Tarazona, J. Dopazo, T. F. Meyer, and A. Conesa. 2012. 'Qualimap: evaluating next-generation sequencing alignment data', *Bioinformatics*, 28: 2678-9.
Gardy, Jennifer L, and Nicholas J Loman. 2018. 'Towards a genomics-informed, real-time, global pathogen surveillance system', *Nature Reviews Genetics*, 19: 9.
Gauthier-Levesque, L., L. Bonifait, N. Turgeon, M. Veillette, P. Perrott, D. Grenier, and C. Duchaine. 2016. 'Impact of serotype and sequence type on the preferential aerosolization of Streptococcus suis', *BMC Res Notes*, 9: 273.
Gimenez-Lirola, L. G., C. Rademacher, D. Linhares, K. Harmon, M. Rotolo, Y. Sun, D. H. Baum, J. Zimmerman, and P. Pineyro. 2016. 'Serological and Molecular Detection of Senecavirus A Associated with an Outbreak of Swine Idiopathic Vesicular Disease and Neonatal Mortality', *J Clin Microbiol*, 54: 2082-9.
Giordano, F., L. Aigrain, M. A. Quail, P. Coupland, J. K. Bonfield, R. M. Davies, G. Tischler, D. K. Jackson, T. M. Keane, J. Li, J. X. Yue, G. Liti, R. Durbin, and Z. Ning. 2017. 'De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms', *Sci Rep*, 7: 3935.
Goede, D., and R. B. Morrison. 2016. 'Production impact & time to stability in sow herds infected with porcine epidemic diarrhea virus (PEDV)', *Prev Vet Med*, 123: 202-07.
Goldstein, S., L. Beka, J. Graf, and J. L. Klassen. 2019. 'Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing', *BMC Genomics*, 20: 23.
Gonzalez-Perez, I., A. A. Cayarga, Y. P. Hernandez, I. G. de la Rosa, Y. J. Gonzalez, C. S. Leon, and R. R. Alvarez. 2010. 'Long-term conservation of HCV RNA at 4 degrees C using a new RNA stabilizing solution', *J Virol Methods*, 168: 207-11.
Gonzalez, A. J., E. C. Ijezie, O. B. Balemba, and T. A. Miura. 2018. 'Attenuation of Influenza A Virus Disease Severity by Viral Coinfection in a Mouse Model', *J Virol*, 92.
Goodwin, S., J. Gurtowski, S. Ethe-Sayers, P. Deshpande, M. C. Schatz, and W. R. McCombie. 2015. 'Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome', *Genome Res*, 25: 1750-6.
Goodwin, Sara, John D McPherson, and W Richard McCombie. 2016. 'Coming of age: ten years of next-generation sequencing technologies', *Nature Reviews Genetics*, 17: 333.
Goolia, M., F. Vannucci, M. Yang, D. Patnayak, S. Babiuk, and C. K. Nfon. 2017. 'Validation of a competitive ELISA and a virus neutralization test for the detection and confirmation of antibodies to Senecavirus A in swine sera', *J Vet Diagn Invest*, 29: 250-53.
Gottschalk, Marcelo, Mariela Segura, and Jiangu Xu. 2007. 'Streptococcus suis infections in humans: the Chinese experience and the situation in North America', *Anim Health Res Rev*, 8: 29-45.
Greninger, A. L., S. N. Naccache, S. Federman, G. Yu, P. Mbala, V. Bres, D. Stryke, J. Bouquet, S. Somasekar, J. M. Linnen, R. Dodd, P. Mulembakani, B. S. Schneider, J. J. Muyembe-Tamfum, S. L. Stramer, and C. Y. Chiu. 2015. 'Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis', *Genome Med*, 7: 99.

Gu, Wei, Steve Miller, and Charles Y Chiu. 2019. 'Clinical metagenomic next-generation sequencing for pathogen detection', *Annual Review of Pathology: Mechanisms of Disease*, 14: 319-38.
Gurevich, Alexey., Vladislav. Saveliev, Nikolay. Vyahhi, and Glenn. Tesler. 2013. 'QUAST: quality assessment tool for genome assemblies', *Bioinformatics*, 29: 1072-75.
Gwinn, M., D. R. MacCannell, and R. F. Khabbaz. 2017a. 'Integrating Advanced Molecular Technologies into Public Health', *J Clin Microbiol*, 55: 703-14.
Gwinn, Marta, Duncan MacCannell, and Gregory L. Armstrong. 2019. 'Next-Generation Sequencing of Infectious PathogensNext-Generation Sequencing of Infectious PathogensNext-Generation Sequencing of Infectious Pathogens', *Jama*, 321: 893-94.
Gwinn, Marta, Duncan R MacCannell, and Rima F Khabbaz. 2017b. 'Integrating advanced molecular technologies into public health', *Journal of Clinical Microbiology*, 55: 703-14.
Hales, L. M., N. J. Knowles, P. S. Reddy, L. Xu, C. Hay, and P. L. Hallenbeck. 2008. 'Complete genome sequence analysis of Seneca Valley virus-001, a novel oncolytic picornavirus', *J Gen Virol*, 89: 1265-75.
Han, G., H. Xu, K. Wang, and F. He. 2019. 'Emergence of Two different recombinant PRRSV strains with low neutralizing antibody susceptibility in China', *Sci Rep*, 9: 2490.
Han, J., Y. Wang, and K. S. Faaberg. 2006. 'Complete genome analysis of RFLP 184 isolates of porcine reproductive and respiratory syndrome virus', *Virus Res*, 122: 175-82.
Han, J., L. Zhou, X. Ge, X. Guo, and H. Yang. 2017. 'Pathogenesis and control of the Chinese highly pathogenic porcine reproductive and respiratory syndrome virus', *Vet Microbiol*, 209: 30-47.
Hanada, K., Y. Suzuki, T. Nakane, O. Hirose, and T. Gojobori. 2005. 'The origin and evolution of porcine reproductive and respiratory syndrome viruses', *Mol Biol Evol*, 22: 1024-31.
Hay, Simon I, Dylan B George, Catherine L Moyes, and John S Brownstein. 2013. 'Big data opportunities for global infectious disease surveillance', *PLoS Med*, 10: e1001413.
Hernando-Amado, Sara, Teresa M Coque, Fernando Baquero, and José L Martínez. 2019. 'Defining and combating antibiotic resistance from One Health and Global Health perspectives', *Nature microbiology*, 4: 1432-42.
Hersh, C. P., and A. Vachani. 2017. 'Whole-Genome Sequencing in Common Respiratory Diseases. Ready, Set, Go!', *Am J Respir Crit Care Med*, 196: 121-22.
Hoenen, T., A. Groseth, K. Rosenke, R. J. Fischer, A. Hoenen, S. D. Judson, C. Martellaro, D. Falzarano, A. Marzi, R. B. Squires, K. R. Wollenberg, E. de Wit, J. Prescott, D. Safronetz, N. van Doremalen, T. Bushmaker, F. Feldmann, K. McNally, F. K. Bolay, B. Fields, T. Sealy, M. Rayfield, S. T. Nichol, K. C. Zoon, M. Massaquoi, V. J. Munster, and H. Feldmann. 2016. 'Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool', *Emerg Infect Dis*, 22: 331-4.
Houe, H., S. S. Nielsen, L. R. Nielsen, S. Ethelberg, and K. Molbak. 2019. 'Opportunities for Improved Disease Surveillance and Control by Use of Integrated Data on Animal and Human Health', *Front Vet Sci*, 6: 301.
Houldcroft, C. J., M. A. Beale, and J. Breuer. 2017. 'Clinical and biological insights from viral genome sequencing', *Nat Rev Microbiol*, 15: 183-92.
Houlihan, C. F., D. Frampton, R. B. Ferns, J. Raffle, P. Grant, M. Reidy, L. Hail, K. Thomson, F. Mattes, Z. Kozlakidis, D. Pillay, A. Hayward, and E. Nastouli. 2018. 'Use of Whole-Genome Sequencing in the Investigation of a Nosocomial Influenza Virus Outbreak', *J Infect Dis*, 218: 1485-89.
Houseley, J., and D. Tollervey. 2009. 'The Many Pathways of RNA Degradation', *Cell*, 136: 763-76.
Hu, L., X. Y. Lin, Z. X. Yang, X. P. Yao, G. L. Li, S. Z. Peng, and Y. Wang. 2015. 'A multiplex PCR for simultaneous detection of classical swine fever virus, African swine fever virus, highly pathogenic porcine reproductive and respiratory syndrome virus, porcine

reproductive and respiratory syndrome virus and pseudorabies in swines', *Pol J Vet Sci*, 18: 715-23.

Huang, B., A. Jennsion, D. Whiley, J. McMahon, G. Hewitson, R. Graham, A. De Jong, and D. Warrilow. 2019a. 'Illumina sequencing of clinical samples for virus detection in a public health laboratory', *Sci Rep*, 9: 5409.

Huang, Bixing, Amy Jennsion, David Whiley, Jamie McMahon, Glen Hewitson, Rikki Graham, Amanda De Jong, and David Warrilow. 2019b. 'Illumina sequencing of clinical samples for virus detection in a public health laboratory', *Sci Rep*, 9: 5409.

Hughes, James M, Mary E Wilson, Heiman FL Wertheim, Ho Dang Trung Nghia, Walter Taylor, and Constance Schultsz. 2009. 'Streptococcus suis: an emerging human pathogen', *Clinical Infectious Diseases*, 48: 617-25.

Jain, M., S. Koren, K. H. Miga, J. Quick, A. C. Rand, T. A. Sasani, J. R. Tyson, A. D. Beggs, A. T. Dilthey, I. T. Fiddes, S. Malla, H. Marriott, T. Nieto, J. O'Grady, H. E. Olsen, B. S. Pedersen, A. Rhie, H. Richardson, A. R. Quinlan, T. P. Snutch, L. Tee, B. Paten, A. M. Phillippy, J. T. Simpson, N. J. Loman, and M. Loose. 2018. 'Nanopore sequencing and assembly of a human genome with ultra-long reads', *Nat Biotechnol*, 36: 338-45.

Jain, M., H. E. Olsen, B. Paten, and M. Akeson. 2016. 'The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community', *Genome Biol*, 17: 239.

Jannes, Geert, and Daniel De Vos. 2006. 'A review of current and future molecular diagnostic tests for use in the microbiology laboratory.' in, *Diagnostic Bacteriology Protocols* (Springer).

Jansen, Kathrin U, Charles Knirsch, and Annaliesa S Anderson. 2018. 'The role of vaccines in preventing bacterial antimicrobial resistance', *Nature medicine*, 24: 10.

Jia, B., A. R. Raphenya, B. Alcock, N. Waglechner, P. Guo, K. K. Tsang, B. A. Lago, B. M. Dave, S. Pereira, A. N. Sharma, S. Doshi, M. Courtot, R. Lo, L. E. Williams, J. G. Frye, T. Elsayegh, D. Sardar, E. L. Westman, A. C. Pawlowski, T. A. Johnson, F. S. Brinkman, G. D. Wright, and A. G. McArthur. 2017. 'CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database', *Nucleic Acids Res*, 45: D566-d73.

Joensen, K. G., F. Scheutz, O. Lund, H. Hasman, R. S. Kaas, E. M. Nielsen, and F. M. Aarestrup. 2014. 'Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli', *J Clin Microbiol*, 52: 1501-10.

Jolley, K. A., J. E. Bray, and M. C. J. Maiden. 2018. 'Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications', *Wellcome Open Res*, 3: 124.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. 2008. 'Global trends in emerging infectious diseases', *Nature*, 451: 990-3.

Joshi, Lok R, Kristin A Mohr, Travis Clement, Kyle S Hain, Bryan Myers, Joseph Yaros, Eric A Nelson, Jane Christopher-Hennings, Danielle Gava, and Rejane Schaefer. 2016. 'Detection of the emerging picornavirus Senecavirus A in pigs, mice, and houseflies', *Journal of Clinical Microbiology*, 54: 1536-45.

Jurado, C., L. Mur, M. S. Perez Aguirreburualde, E. Cadenas-Fernandez, B. Martinez-Lopez, J. M. Sanchez-Vizcaino, and A. Perez. 2019. 'Risk of African swine fever virus introduction into the United States through smuggling of pork in air passenger luggage', *Sci Rep*, 9: 14423.

Juul, Sissel, Fernando Izquierdo, Adam Hurst, Xiaoguang Dai, Amber Wright, Eugene Kulesha, Roger Pettett, and Daniel J. Turner. 2015. 'What's in my pot? Real-time species identification on the MinION™', *bioRxiv*: 030742.

Kan, B., H. Zhou, P. Du, W. Zhang, X. Lu, T. Qin, and J. Xu. 2018. 'Transforming bacterial disease surveillance and investigation using whole-genome sequence to probe the trace', *Front Med*, 12: 23-33.

Kao, Rowland R, Daniel T Haydon, Samantha J Lycett, and Pablo R Murcia. 2014. 'Supersize me: how whole-genome sequencing and big data are transforming epidemiology', *Trends Microbiol*, 22: 282-91.
Kappes, M. A., and K. S. Faaberg. 2015. 'PRRSV structure, replication and recombination: Origin of phenotype and genotype diversity', *Virology*, 479-480: 475-86.
Kapur, V., M. R. Elam, T. M. Pawlovich, and M. P. Murtaugh. 1996. 'Genetic variation in porcine reproductive and respiratory syndrome virus isolates in the midwestern United States', *J Gen Virol*, 77 (Pt 6): 1271-6.
Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. 'Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*, 28: 1647-9.
Keller, M. W., B. L. Rambo-Martin, M. M. Wilson, C. A. Ridenour, S. S. Shepard, T. J. Stark, E. B. Neuhaus, V. G. Dugan, D. E. Wentworth, and J. R. Barnes. 2018. 'Direct RNA Sequencing of the Coding Complete Influenza A Virus Genome', *Sci Rep*, 8: 14408.
Kilianski, A., J. L. Haas, E. J. Corriveau, A. T. Liem, K. L. Willis, D. R. Kadavy, C. N. Rosenzweig, and S. S. Minot. 2015. 'Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer', *Gigascience*, 4: 12.
Kim, T., C. Park, K. Choi, J. Jeong, I. Kang, S. J. Park, and C. Chae. 2015. 'Comparison of Two Commercial Type 1 Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) Modified Live Vaccines against Heterologous Type 1 and Type 2 PRRSV Challenge in Growing Pigs', *Clin Vaccine Immunol*, 22: 631-40.
King, S. J., J. A. Leigh, P. J. Heath, I. Luque, C. Tarradas, C. G. Dowson, and A. M. Whatmore. 2002. 'Development of a multilocus sequence typing scheme for the pig pathogen Streptococcus suis: identification of virulent clones and potential capsular serotype exchange', *J Clin Microbiol*, 40: 3671-80.
Kluberg, S. A., S. R. Mekaru, D. J. McIver, L. C. Madoff, A. W. Crawley, M. S. Smolinski, and J. S. Brownstein. 2016. 'Global Capacity for Emerging Infectious Disease Detection, 1996-2014', *Emerg Infect Dis*, 22: E1-6.
Ko, Jung Ho, Phat-Loc Nguyen, Ji-Young Ahn, Hobaek Yoon, Jiho Min, Lyon Lee, Sung-Jin Cho, Simranjeet Singh Sekhon, and Yang-Hoon Kim. 2015. 'The global research trend of porcine reproductive and respiratory syndrome virus (PRRSV): A mini review', *Toxicology and Environmental Health Sciences*, 7: 241-50.
Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, and A. M. Phillippy. 2017. 'Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation', *Genome Res*, 27: 722-36.
Koser, C. U., M. J. Ellington, and S. J. Peacock. 2014. 'Whole-genome sequencing to control antimicrobial resistance', *Trends Genet*, 30: 401-7.
Kruppa, J., W. K. Jo, E. van der Vries, M. Ludlow, A. Osterhaus, W. Baumgaertner, and K. Jung. 2018. 'Virus detection in high-throughput sequencing data without a reference genome of the host', *Infect Genet Evol*, 66: 180-87.
Kvisgaard, L. K., C. K. Hjulsager, U. Fahnoe, S. O. Breum, T. Ait-Ali, and L. E. Larsen. 2013. 'A fast and robust method for full genome sequencing of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) Type 1 and Type 2', *J Virol Methods*, 193: 697-705.
Ladner, Jason T, Nathan D Grubaugh, Oliver G Pybus, and Kristian G Andersen. 2019. 'Precision epidemiology for infectious disease control', *Nature medicine*, 25: 206-11.
Lager, K. M., and A. C. Buckley. 2019. 'Porcine Anti-viral Immunity: How Important Is It?', *Front Immunol*, 10: 2258.

Lahens, N. F., I. H. Kavakli, R. Zhang, K. Hayer, M. B. Black, H. Dueck, A. Pizarro, J. Kim, R. Irizarry, R. S. Thomas, G. R. Grant, and J. B. Hogenesch. 2014. 'IVT-seq reveals extreme bias in RNA sequencing', *Genome Biol*, 15: R86.

Landers, Timothy F, Bevin Cohen, Thomas E Wittum, and Elaine L Larson. 2012. 'A review of antibiotic use in food animals: perspective, policy, and potential', *Public health reports*, 127: 4-22.

Lanfear, R., M. Schalamun, D. Kainer, W. Wang, and B. Schwessinger. 2019. 'MinIONQC: fast and simple quality control for MinION sequencing data', *Bioinformatics*, 35: 523-25.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. 'Clustal W and Clustal X version 2.0', *Bioinformatics*, 23: 2947-8.

Lecuit, M., and M. Eloit. 2014. 'The diagnosis of infectious diseases by whole genome next generation sequencing: a new era is opening', *Front Cell Infect Microbiol*, 4: 25.

Lefterova, M. I., C. J. Suarez, N. Banaei, and B. A. Pinsky. 2015. 'Next-Generation Sequencing for Infectious Disease Diagnosis and Management: A Report of the Association for Molecular Pathology', *J Mol Diagn*, 17: 623-34.

Lekagul, Angkana. 2019. 'Patterns of antibiotic use in global pig production: a systematic review', *Veterinary and Animal Science*: 100058.

Leme, R. A., E. Zotti, B. K. Alcantara, M. V. Oliveira, L. A. Freitas, A. F. Alfieri, and A. A. Alfieri. 2015. 'Senecavirus A: An Emerging Vesicular Infection in Brazilian Pig Herds', *Transbound Emerg Dis*, 62: 603-11.

Lewandowski, Kuiama, Yifei Xu, Steven T. Pullan, Sheila F. Lumley, Dona Foster, Nicholas Sanderson, Alison Vaughan, Marcus Morgan, Nicole Bright, James Kavanagh, Richard Vipond, Miles Carroll, Anthony C. Marriott, Karen E Gooch, Monique Andersson, Katie Jeffery, Timothy EA Peto, Derrick W. Crook, A Sarah Walker, and Philippa C. Matthews. 2019. 'Metagenomic Nanopore sequencing of influenza virus direct from clinical respiratory samples', *bioRxiv*: 676155.

Li, H. 2016. 'Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences', *Bioinformatics*, 32: 2103-10.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25: 2078-9.

Lipkin, W Ian. 2010. 'Microbe hunting', *Microbiol. Mol. Biol. Rev.*, 74: 363-77.

Liu, Jian-Kui, Chun-Hua Wei, Xiao-Yan Yang, Ai-Ling Dai, and Xiao-Hua Li. 2013. 'Multiplex PCR for the simultaneous detection of porcine reproductive and respiratory syndrome virus, classical swine fever virus, and porcine circovirus in pigs', *Molecular and cellular probes*, 27: 149-52.

Liu, Y., J. Li, J. Yang, H. Zeng, L. Guo, S. Ren, W. Sun, Z. Chen, X. Cong, J. Shi, L. Chen, Y. Du, J. Li, J. Wang, J. Wu, and J. Yu. 2018. 'Emergence of Different Recombinant Porcine Reproductive and Respiratory Syndrome Viruses, China', *Sci Rep*, 8: 4118.

Loeb, Josh. 2019. "ASF to affect a quarter of the world's pigs." In.: BMJ PUBLISHING GROUP BRITISH MED ASSOC HOUSE, TAVISTOCK SQUARE, LONDON WC1H ….

Loman, N. J., J. Quick, and J. T. Simpson. 2015. 'A complete bacterial genome assembled de novo using only nanopore sequencing data', *Nat Methods*, 12: 733-5.

Loman, N. J., and M. Watson. 2015. 'Successful test launch for nanopore sequencing', *Nat Methods*, 12: 303-4.

Lu, H., F. Giordano, and Z. Ning. 2016. 'Oxford Nanopore MinION Sequencing and Genome Assembly', *Genomics Proteomics Bioinformatics*, 14: 265-79.

Lu, Z. H., A. L. Archibald, and T. Ait-Ali. 2014. 'Beyond the whole genome consensus: unravelling of PRRSV phylogenomics using next generation sequencing technologies', *Virus Res*, 194: 167-74.

Luciani, Fabio, Rowena A Bull, and Andrew R Lloyd. 2012. 'Next generation deep sequencing and vaccine design: today and tomorrow', *Trends in biotechnology*, 30: 443-52.
Lunney, J. K., D. A. Benfield, and R. R. Rowland. 2010. 'Porcine reproductive and respiratory syndrome virus: an update on an emerging and re-emerging viral disease of swine', *Virus Res*, 154: 1-6.
Lurchachaiwong, W., S. Payungporn, U. Srisatidnarakul, C. Mungkundar, A. Theamboonlers, and Y. Poovorawan. 2008. 'Rapid detection and strain identification of porcine reproductive and respiratory syndrome virus (PRRSV) by real-time RT-PCR', *Lett Appl Microbiol*, 46: 55-60.
Ma, Xiao, Elyse Stachler, and Kyle Bibby. 2017. 'Evaluation of Oxford Nanopore MinION Sequencing for 16S rRNA Microbiome Characterization', *bioRxiv*: 099960.
Machado, Lais Ceschini, Mariana Carolina de Morais-Sobral, Tulio de Lima Campos, Mylena Ribeiro Pereira, Clément Gilbert, Rafael Freitas Oliveira Franca, and Gabriel Luz Wallau. 2019. 'Genome sequencing reveals coinfection by multiple chikungunya virus genotypes in a recent outbreak in Brazil', *PLoS neglected tropical diseases*, 13: e0007332.
Maljkovic, I Berry, MC Melendrez, KA Bishop-Lilly, W Rutvisuttinunt, S Pollett, E Talundzic, L Morton, and RG Jarman. 2019. 'Next Generation Sequencing and Bioinformatics Methodologies for Infectious Disease Research and Public Health: Approaches, Applications, and Considerations for Development of Laboratory Capacity', *The Journal of infectious diseases*.
Marcais, G., A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg, and A. Zimin. 2018. 'MUMmer4: A fast and versatile genome alignment system', *PLoS Comput Biol*, 14: e1005944.
Marston, Hilary D, Gregory K Folkers, David M Morens, and Anthony S Fauci. 2014. 'Emerging viral diseases: confronting threats with new technologies', *Science Translational Medicine*, 6: 253ps10-53ps10.
McArdle, Andrew J, Anna Turkova, and Aubrey J Cunnington. 2018. 'When do co-infections matter?', *Current opinion in infectious diseases*, 31: 209.
McNaughton, A. L., H. E. Roberts, D. Bonsall, M. de Cesare, J. Mokaya, S. F. Lumley, T. Golubchik, P. Piazza, J. B. Martin, C. de Lara, A. Brown, M. A. Ansari, R. Bowden, E. Barnes, and P. C. Matthews. 2019. 'Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV)', *Sci Rep*, 9: 7081.
Meeusen, Els NT, John Walker, Andrew Peters, Paul-Pierre Pastoret, and Gregers Jungersen. 2007. 'Current status of veterinary vaccines', *Clin Microbiol Rev*, 20: 489-510.
Meng, X. J. 2000. 'Heterogeneity of porcine reproductive and respiratory syndrome virus: implications for current vaccine efficacy and future vaccine development', *Vet Microbiol*, 74: 309-29.
———. 2012. 'Emerging and re-emerging swine viruses', *Transbound Emerg Dis*, 59 Suppl 1: 85-102.
Mitsuhashi, S., K. Kryukov, S. Nakagawa, J. S. Takeuchi, Y. Shiraishi, K. Asano, and T. Imanishi. 2017. 'A portable system for rapid bacterial composition analysis using a nanopore-based sequencer and laptop computer', *Sci Rep*, 7: 5657.
Mongan, Arthur Elia, Josef Sem Berth Tuda, and Lucky Ronald Runtuwene. 2019. 'Portable sequencer in the fight against infectious disease', *Journal of human genetics*: 1-6.
Morens, D. M., and A. S. Fauci. 2013. 'Emerging infectious diseases: threats to human health and global stability', *PLoS Pathog*, 9: e1003467.
Morens, D. M., G. K. Folkers, and A. S. Fauci. 2004. 'The challenge of emerging and re-emerging infectious diseases', *Nature*, 430: 242-9.
Morse, Stephen S. 2012. 'Public health surveillance and infectious disease detection', *Biosecurity and bioterrorism: biodefense strategy, practice, and science*, 10: 6-16.

Moscoso, Carlos G, Kelly R Potz, Shaoyuan Tan, Pamala A Jacobson, Kavita M Berger, and Clifford J Steer. 2019. 'Precision medicine, agriculture, and genome editing: science and ethics', *Ann N Y Acad Sci*.

Murtaugh, M. P., K. S. Faaberg, J. Laber, M. Elam, and V. Kapur. 1998. 'Genetic variation in the PRRS virus', *Adv Exp Med Biol*, 440: 787-94.

Murtaugh, M. P., and M. Genzow. 2011. 'Immunological solutions for treatment and prevention of porcine reproductive and respiratory syndrome (PRRS)', *Vaccine*, 29: 8192-204.

Murtaugh, M. P., T. Stadejek, J. E. Abrahante, T. T. Lam, and F. C. Leung. 2010. 'The ever-expanding diversity of porcine reproductive and respiratory syndrome virus', *Virus Res*, 154: 18-30.

Murtaugh, Michael. 2012. 'Use and interpretation of sequencing in PRRSV control programs'.

Naidoo, J., N. Mahomed, and H. Moodley. 2017. 'A systemic review of tuberculosis with HIV coinfection in children', *Pediatr Radiol*, 47: 1269-76.

Nan, Y., C. Wu, G. Gu, W. Sun, Y. J. Zhang, and E. M. Zhou. 2017. 'Improved Vaccine against PRRSV: Current Progress and Future Perspective', *Front Microbiol*, 8: 1635.

Nathues, H., P. Alarcon, J. Rushton, R. Jolie, K. Fiebig, M. Jimenez, V. Geurts, and C. Nathues. 2017. 'Cost of porcine reproductive and respiratory syndrome virus at individual farm level - An economic disease model', *Prev Vet Med*, 142: 16-29.

Neher, Richard A, and Trevor Bedford. 2018. 'Real-time analysis and visualization of pathogen sequence data', *Journal of Clinical Microbiology*, 56: e00480-18.

Nelsen, C. J., M. P. Murtaugh, and K. S. Faaberg. 1999. 'Porcine reproductive and respiratory syndrome virus comparison: divergent evolution on two continents', *J Virol*, 73: 270-80.

Nelson, Martha I, and Amy L Vincent. 2015. 'Reverse zoonosis of influenza to swine: new perspectives on the human–animal interface', *Trends Microbiol*, 23: 142-53.

Neumann, E. J., J. B. Kliebenstein, C. D. Johnson, J. W. Mabry, E. J. Bush, A. H. Seitzinger, A. L. Green, and J. J. Zimmerman. 2005. 'Assessment of the economic impact of porcine reproductive and respiratory syndrome on swine production in the United States', *J Am Vet Med Assoc*, 227: 385-92.

Niederwerder, Megan C, Crystal J Jaing, James B Thissen, Ada Giselle Cino-Ozuna, Kevin S McLoughlin, and Raymond RR Rowland. 2016. 'Microbiome associations in pigs with the best and worst clinical outcomes following co-infection with porcine reproductive and respiratory syndrome virus (PRRSV) and porcine circovirus type 2 (PCV2)', *Vet Microbiol*, 188: 1-11.

Nishiura, H., S. Barua, S. Lawpoolsri, C. Kittitrakul, M. M. Leman, M. S. Maha, and S. Muangnoicharoen. 2004. 'Health inequalities in Thailand: geographic distribution of medical supplies in the provinces', *Southeast Asian J Trop Med Public Health*, 35: 735-40.

Okonechnikov, K., A. Conesa, and F. Garcia-Alcalde. 2016. 'Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data', *Bioinformatics*, 32: 292-4.

Okonechnikov, K., O. Golosova, and M. Fursov. 2012. 'Unipro UGENE: a unified bioinformatics toolkit', *Bioinformatics*, 28: 1166-7.

Oliveira, T. E. S., M. M. Z. Michelazzo, T. Fernandes, A. G. de Oliveira, R. A. Leme, A. F. Alfieri, A. A. Alfieri, and S. A. Headley. 2017. 'Histopathological, immunohistochemical, and ultrastructural evidence of spontaneous Senecavirus A-induced lesions at the choroid plexus of newborn piglets', *Sci Rep*, 7: 16555.

Onions, D, C Cote, B Love, B Toms, S Koduri, A Armstrong, A Chang, and J Kolman. 2011. 'Ensuring the safety of vaccine cell substrates by massively parallel sequencing of the transcriptome', *Vaccine*, 29: 7117-21.

Opriessnig, T, LG Giménez-Lirola, and PG Halbur. 2011. 'Polymicrobial respiratory disease in pigs', *Anim Health Res Rev*, 12: 133-48.

Opriessnig, T., P. G. Halbur, K. J. Yoon, R. M. Pogranichniy, K. M. Harmon, R. Evans, K. F. Key, F. J. Pallares, P. Thomas, and X. J. Meng. 2002. 'Comparison of molecular and biological characteristics of a modified live porcine reproductive and respiratory syndrome virus (PRRSV) vaccine (ingelvac PRRS MLV), the parent strain of the vaccine (ATCC VR2332), ATCC VR2385, and two recent field isolates of PRRSV', *J Virol*, 76: 11837-44.

Ouyang, T., X. Zhang, X. Liu, and L. Ren. 2019. 'Co-Infection of Swine with Porcine Circovirus Type 2 and Other Swine Viruses', *Viruses*, 11.

Ovcharenko, A., and A. Rentmeister. 2018. 'Emerging approaches for detection of methylation sites in RNA', *Open Biol*, 8.

Pak, Theodore R, and Andrew Kasarskis. 2015. 'How next-generation sequencing and multiscale data analysis will transform infectious disease management', *Clinical Infectious Diseases*, 61: 1695-702.

Palmieri, C., P. E. Varaldo, and B. Facinelli. 2011. 'Streptococcus suis, an Emerging Drug-Resistant Animal and Human Pathogen', *Front Microbiol*, 2: 235.

Parker, J., and J. Chen. 2017. 'Application of next generation sequencing for the detection of human viral pathogens in clinical specimens', *J Clin Virol*, 86: 20-26.

'Pathogen genomics leading to vaccines'. 2019. *Nat Genet*, 51: 923.

Pearce-Duvet, Jessica MC. 2006. 'The origin of human pathogens: evaluating the role of agriculture and domestic animals in the evolution of human disease', *Biological Reviews*, 81: 369-82.

Perch, Beate, P Kristjansen, and KN Skadhauge. 1968. 'Group R streptococci pathogenic for man: two cases of meningitis and one fatal case of sepsis', *Acta Pathologica Microbiologica Scandinavica*, 74: 69-76.

Piddock, L. J. 2012. 'The crisis of no new antibiotics--what is the way forward?', *Lancet Infect Dis*, 12: 249-53.

Pomerantz, Aaron, Nicolás Peñafiel, Alejandro Arteaga, Lucas Bustamante, Frank Pichardo, Luis A Coloma, César L Barrio-Amorós, David Salazar-Valenzuela, and Stefan Prost. 2018. 'Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building', *Gigascience*, 7: giy033.

Prachayangprecha, S., C. M. Schapendonk, M. P. Koopmans, A. D. Osterhaus, A. C. Schurch, S. D. Pas, A. A. van der Eijk, Y. Poovorawan, B. L. Haagmans, and S. L. Smits. 2014. 'Exploring the potential of next-generation sequencing in detection of respiratory viruses', *J Clin Microbiol*, 52: 3722-30.

Qian, S., W. Fan, P. Qian, H. Chen, and X. Li. 2016. 'Isolation and full-genome sequencing of Seneca Valley virus in piglets from China, 2016', *Virol J*, 13: 173.

Qin, S., W. Ruan, H. Yue, C. Tang, K. Zhou, and B. Zhang. 2018. 'Viral communities associated with porcine respiratory disease complex in intensive commercial farms in Sichuan province, China', *Sci Rep*, 8: 13341.

Quick, J., N. D. Grubaugh, S. T. Pullan, I. M. Claro, A. D. Smith, K. Gangavarapu, G. Oliveira, R. Robles-Sikisaka, T. F. Rogers, N. A. Beutler, D. R. Burton, L. L. Lewis-Ximenez, J. G. de Jesus, M. Giovanetti, S. C. Hill, A. Black, T. Bedford, M. W. Carroll, M. Nunes, L. C. Alcantara, Jr., E. C. Sabino, S. A. Baylis, N. R. Faria, M. Loose, J. T. Simpson, O. G. Pybus, K. G. Andersen, and N. J. Loman. 2017. 'Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples', *Nat Protoc*, 12: 1261-76.

Quick, J., N. J. Loman, S. Duraffour, J. T. Simpson, E. Severi, L. Cowley, J. A. Bore, R. Koundouno, G. Dudas, A. Mikhail, N. Ouedraogo, B. Afrough, A. Bah, J. H. Baum, B. Becker-Ziaja, J. P. Boettcher, M. Cabeza-Cabrerizo, A. Camino-Sanchez, L. L. Carter, J. Doerrbecker, T. Enkirch, I. G. G. Dorival, N. Hetzelt, J. Hinzmann, T. Holm, L. E. Kafetzopoulou, M. Koropogui, A. Kosgey, E. Kuisma, C. H. Logue, A. Mazzarelli, S.

Meisel, M. Mertens, J. Michel, D. Ngabo, K. Nitzsche, E. Pallash, L. V. Patrono, J. Portmann, J. G. Repits, N. Y. Rickett, A. Sachse, K. Singethan, I. Vitoriano, R. L. Yemanaberhan, E. G. Zekeng, R. Trina, A. Bello, A. A. Sall, O. Faye, O. Faye, N. Magassouba, C. V. Williams, V. Amburgey, L. Winona, E. Davis, J. Gerlach, F. Washington, V. Monteil, M. Jourdain, M. Bererd, A. Camara, H. Somlare, A. Camara, M. Gerard, G. Bado, B. Baillet, D. Delaune, K. Y. Nebie, A. Diarra, Y. Savane, R. B. Pallawo, G. J. Gutierrez, N. Milhano, I. Roger, C. J. Williams, F. Yattara, K. Lewandowski, J. Taylor, P. Rachwal, D. Turner, G. Pollakis, J. A. Hiscox, D. A. Matthews, M. K. O'Shea, A. M. Johnston, D. Wilson, E. Hutley, E. Smit, A. Di Caro, R. Woelfel, K. Stoecker, E. Fleischmann, M. Gabriel, S. A. Weller, L. Koivogui, B. Diallo, S. Keita, A. Rambaut, P. Formenty, S. Gunther, and M. W. Carroll. 2016. 'Real-time, portable genome sequencing for Ebola surveillance', *Nature*, 530: 228-32.

Rajbhandari, Ruma, Tomi Jun, Hamed Khalili, Raymond T Chung, and Ashwin N Ananthakrishnan. 2016. 'HBV/HIV coinfection is associated with poorer outcomes in hospitalized patients with HBV or HIV', *Journal of viral hepatitis*, 23: 820-29.

Rambo-Martin, Benjamin L., Matthew W. Keller, Malania M. Wilson, Jacqueline M. Nolting, Tavis K. Anderson, Amy L. Vincent, Ujwal Bagal, Yunho Jang, Elizabeth B. Neuhaus, C. Todd Davis, Andrew S. Bowman, David E. Wentworth, and John R. Barnes. 2019. 'Mitigating Pandemic Risk with Influenza A Virus Field Surveillance at a Swine-Human Interface', *bioRxiv*: 585588.

Relova, D., L. Rios, A. M. Acevedo, L. Coronado, C. L. Perera, and L. J. Perez. 2018. 'Impact of RNA Degradation on Viral Diagnosis: An Understated but Essential Step for the Successful Establishment of a Diagnosis Network', *Vet Sci*, 5.

Resistance, Review on Antimicrobial. 2016. *Tackling drug-resistant infections globally: final report and recommendations* (Review on antimicrobial resistance).

Reuter, S., M. J. Ellington, E. J. Cartwright, C. U. Koser, M. E. Torok, T. Gouliouris, S. R. Harris, N. M. Brown, M. T. Holden, M. Quail, J. Parkhill, G. P. Smith, S. D. Bentley, and S. J. Peacock. 2013. 'Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology', *JAMA Intern Med*, 173: 1397-404.

Ribot, E. M., M. A. Fair, R. Gautom, D. N. Cameron, S. B. Hunter, B. Swaminathan, and T. J. Barrett. 2006. 'Standardization of pulsed-field gel electrophoresis protocols for the subtyping of Escherichia coli O157:H7, Salmonella, and Shigella for PulseNet', *Foodborne Pathog Dis*, 3: 59-67.

Ricker, N., H. Qian, and R. R. Fulthorpe. 2012. 'The limitations of draft assemblies for understanding prokaryotic adaptation and evolution', *Genomics*, 100: 167-75.

Riising, H. J., M. Murmans, and M. Witvliet. 2005. 'Protection against neonatal Escherichia coli diarrhoea in pigs by vaccination of sows with a new vaccine that contains purified enterotoxic E. coli virulence factors F4ac, F4ab, F5 and F6 fimbrial antigens and heat-labile E. coli enterotoxin (LT) toxoid', *J Vet Med B Infect Dis Vet Public Health*, 52: 296-300.

Risse, J., M. Thomson, S. Patrick, G. Blakely, G. Koutsovoulos, M. Blaxter, and M. Watson. 2015. 'A single chromosome assembly of Bacteroides fragilis strain BE1 from Illumina and MinION nanopore sequencing data', *Gigascience*, 4: 60.

Robinson, S. R., M. C. Rahe, D. K. Gray, K. V. Martins, and M. P. Murtaugh. 2018. 'Porcine reproductive and respiratory syndrome virus neutralizing antibodies provide in vivo cross-protection to PRRSV1 and PRRSV2 viral challenge', *Virus Res*, 248: 13-23.

Roe, Chandler C, Adam Vazquez, Eliana Pia Esposito, Raffaele Zarrilli, and Jason William Sahl. 2019. 'Diversity, virulence, and antimicrobial resistance in isolates from the newly emerging Klebsiella pneumoniae ST101 lineage', *Front Microbiol*, 10: 542.

Romagosa, Anna, Matt Allerson, Marie Gramer, Han Soo Joo, John Deen, Susan Detmer, and Montserrat Torremorell. 2011. 'Vaccination of influenza a virus decreases transmission rates in pigs', *Vet Res*, 42: 120.

Rose, Nicolas, and Mathieu Andraud. 2017. 'The use of vaccines to control pathogen spread in pig populations', *Porcine health management*, 3: 8.

Rovira, Albert, M Balasch, J Segales, L Garcia, J Plana-Duran, C Rosell, H Ellerbrok, A Mankertz, and M Domingo. 2002. 'Experimental inoculation of conventional pigs with porcine reproductive and respiratory syndrome virus and porcine circovirus 2', *J Virol*, 76: 3232-39.

Ruan, J., and H. Li. 2019. 'Fast and accurate long-read assembly with wtdbg2', *Nat Methods*.

Russell, J. A., B. Campos, J. Stone, E. M. Blosser, N. Burkett-Cadena, and J. L. Jacobs. 2018. 'Unbiased Strain-Typing of Arbovirus Directly from Mosquitoes Using Nanopore Sequencing: A Field-forward Biosurveillance Protocol', *Sci Rep*, 8: 5417.

Saeng-Chuto, K., P. Rodtian, G. Temeeyasen, M. Wegner, and D. Nilubol. 2018. 'The first detection of Senecavirus A in pigs in Thailand, 2016', *Transbound Emerg Dis*, 65: 285-88.

Salipante, Stephen J, Dhruba J SenGupta, Lisa A Cummings, Tyler A Land, Daniel R Hoogestraat, and Brad T Cookson. 2015. 'Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology', *Journal of Clinical Microbiology*, 53: 1072-79.

Sanchez-Romero, M. I., J. M. Garcia-Lechuz Moya, J. J. Gonzalez Lopez, and N. Orta Mira. 2019. 'Collection, transport and general processing of clinical specimens in Microbiology laboratory', *Enferm Infecc Microbiol Clin*, 37: 127-34.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. 'DNA SEQUENCING WITH CHAIN-TERMINATING INHIBITORS', *Proceedings of the National Academy of Sciences of the United States of America*, 74: 5463-67.

Schmitt, B., and L. Henderson. 2005. 'Diagnostic tools for animal diseases', *Rev Sci Tech*, 24: 243-50.

Schurch, A. C., and W. van Schaik. 2017. 'Challenges and opportunities for whole-genome sequencing-based surveillance of antibiotic resistance', *Ann N Y Acad Sci*, 1388: 108-20.

Segalés, J., D. Barcellos, A. Alfieri, E. Burrough, and D. Marthaler. 2017. 'Senecavirus A:An Emerging Pathogen Causing Vesicular Disease and Mortality in Pigs?', *Veterinary Pathology*, 54: 11-21.

Segales, J., T. Kekarainen, and M. Cortey. 2013. 'The natural history of porcine circovirus type 2: from an inoffensive virus to a devastating swine disease?', *Vet Microbiol*, 165: 13-20.

Segura, M., N. Fittipaldi, C. Calzas, and M. Gottschalk. 2017. 'Critical Streptococcus suis Virulence Factors: Are They All Really Critical?', *Trends Microbiol*, 25: 585-99.

Slatko, B. E., A. F. Gardner, and F. M. Ausubel. 2018. 'Overview of Next-Generation Sequencing Technologies', *Curr Protoc Mol Biol*, 122: e59.

Smith, Dipendra ThapaliyaBlake M. HansonAshley KatesCassandra A. KlostermannRajeshwari NairShylo E. WardynTara C. 2015. 'Zoonotic Diseases of Swine: Food-borne and Occupational Aspects of Infection.', *In: Sing A. (eds) Zoonoses*, Infections Affecting Humans and Animals. : pp 23-68.

Smith, JW, and MS Wolfe. 1980. 'Giardiasis', *Annual review of medicine*, 31: 373-83.

Smith, Tara C, Abby L Harper, Rajeshwari Nair, Shylo E Wardyn, Blake M Hanson, Dwight D Ferguson, and Anne E Dressler. 2011. 'Emerging swine zoonoses', *Vector-Borne and Zoonotic Diseases*, 11: 1225-34.

Snijder, E. J., and J. J. Meulenberg. 1998. 'The molecular biology of arteriviruses', *J Gen Virol*, 79 (Pt 5): 961-79.

Solano, Gloria I, Elida Bautista, Thomas W Molitor, Joaquim Segales, and Carlos Pijoan. 1998. 'Effect of porcine reproductive and respiratory syndrome virus infection on the clearance

of Haemophilus parasuis by porcine alveolar macrophages', *Canadian journal of veterinary research*, 62: 251.
Sovic, I., M. Sikic, A. Wilm, S. N. Fenlon, S. Chen, and N. Nagarajan. 2016. 'Fast and sensitive mapping of nanopore sequencing reads with GraphMap', *Nat Commun*, 7: 11307.
Sozhamannan, Shanmuga, Mitchell Holland, Adrienne Hall, Daniel Negrón, Mychal Ivancich, Jeffrey Koehler, Timothy Minogue, Catherine Campbell, Walter Berger, and George Christopher. 2015. 'Evaluation of signature erosion in Ebola virus due to genomic drift and its impact on the performance of diagnostic assays', *Viruses*, 7: 3130-54.
Spellberg, B., J. G. Bartlett, and D. N. Gilbert. 2013. 'The future of antibiotics and resistance', *N Engl J Med*, 368: 299-302.
Stärk, Katharina DC, Montserrat Arroyo Kuribreña, Gwenaelle Dauphin, Sandra Vokaty, Michael P Ward, Barbara Wieland, and Ann Lindberg. 2015. 'One Health surveillance–more than a buzz word?', *Prev Vet Med*, 120: 124-30.
Storgaard, T., M. Oleksiewicz, and A. Botner. 1999. 'Examination of the selective pressures on a live PRRS vaccine virus', *Arch Virol*, 144: 2389-401.
Strong, M. J., G. Xu, L. Morici, S. Splinter Bon-Durant, M. Baddoo, Z. Lin, C. Fewell, C. M. Taylor, and E. K. Flemington. 2014. 'Microbial contamination in next generation sequencing: implications for sequence-based analysis of clinical samples', *PLoS Pathog*, 10: e1004437.
Sugden, Rebecca, Ruth Kelly, and Sally Davies. 2016. 'Combatting antimicrobial resistance globally', *Nature microbiology*, 1: 16187.
Sullivan, Seth J., Robert M. Jacobson, Walter R. Dowdle, and Gregory A. Poland. 2010. '2009 H1N1 influenza', *Mayo Clinic proceedings*, 85: 64-76.
Sur, J. H. 2019. 'How far can African swine fever spread?', *J Vet Sci*, 20: e41.
Tan, S., C. M. T. Dvorak, A. Estrada, C. Gebhart, D. G. Marthaler, and M. P. Murtaugh. 2019. 'MinION sequencing of Streptococcus suis allows for functional characterization of bacteria by multilocus sequence typing and antimicrobial resistance profiling', *J Microbiol Methods*: 105817.
Tan, S., C. M. T. Dvorak, and M. P. Murtaugh. 2019. 'Rapid, Unbiased PRRSV Strain Detection Using MinION Direct RNA Sequencing and Bioinformatics Tools', *Viruses*, 11.
Team, R Core. 2016. 'R: A language and environment for statistical computing. R Foundation for Statistical Computing , Vienna, Austria. URL https://www.R-project.org/.'.
Thanawongnuwech, R., G. B. Brown, P. G. Halbur, J. A. Roth, R. L. Royer, and B. J. Thacker. 2000. 'Pathogenesis of porcine reproductive and respiratory syndrome virus-induced increase in susceptibility to Streptococcus suis infection', *Vet Pathol*, 37: 143-52.
Theuns, S., B. Vanmechelen, Q. Bernaert, W. Deboutte, M. Vandenhole, L. Beller, J. Matthijnssens, P. Maes, and H. J. Nauwynck. 2018. 'Nanopore sequencing as a revolutionary diagnostic tool for porcine viral enteric disease complexes identifies porcine kobuvirus as an important enteric virus', *Sci Rep*, 8: 9830.
Thoendel, Matthew, Patricio R Jeraldo, Kerryl E Greenwood-Quaintance, Janet Z Yao, Nicholas Chia, Arlen D Hanssen, Matthew P Abdel, and Robin Patel. 2016. 'Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing', *Journal of Microbiological Methods*, 127: 141-45.
Thomas, Milton, Gavin John Fenske, Linto Antony, Sudeep Ghimire, Ronald Welsh, Akhilesh Ramachandran, and Joy Scaria. 2017. 'Whole genome sequencing-based detection of antimicrobial resistance and virulence in non-typhoidal Salmonella enterica isolated from wildlife', *Gut pathogens*, 9: 66.
Tomley, Fiona M, and Martin W Shirley. 2009. "Livestock infectious diseases and zoonoses." In.: The Royal Society.

Tousignant, S. J., A. M. Perez, J. F. Lowe, P. E. Yeske, and R. B. Morrison. 2015. 'Temporal and spatial dynamics of porcine reproductive and respiratory syndrome virus infection in the United States', *Am J Vet Res*, 76: 70-6.

Tuanthap, Supansa, Sompong Vongpunsawad, Cherdpong Phupolphan, Ausanee Duang-in, Suphot Wattanaphansak, Pornchalit Assavacheep, Apiradee Theamboonlers, Supol Luengyosluechakul, Alongkorn Amonsin, and Yong Poovorawan. 2019. 'Analysis of the spike, ORF3, and nucleocapsid genes of porcine epidemic diarrhea virus circulating on Thai swine farms, 2011–2016', *PeerJ*, 7: e6843.

Tyson, J. R., N. J. O'Neil, M. Jain, H. E. Olsen, P. Hieter, and T. P. Snutch. 2018. 'MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome', *Genome Res*, 28: 266-74.

Uddin Khan, Salah, Kalina R Atanasova, Whitney S Krueger, Alejandro Ramirez, and Gregory C Gray. 2013. 'Epidemiology, geographical distribution, and economic consequences of swine zoonoses: a narrative review', *Emerg Microbes Infect*, 2: 1-11.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. 2012. 'Primer3--new capabilities and interfaces', *Nucleic Acids Res*, 40: e115.

Valdes-Donoso, Pablo, Julio Alvarez, Lovell S Jarvis, Robert B Morrison, and Andres M Perez. 2018. 'Production losses from an endemic animal disease: Porcine reproductive and respiratory syndrome (PRRS) in selected Midwest US sow farms', *Front Vet Sci*, 5: 102.

Vallat, Bernard. 2014. 'Feeding the world better by controlling animal diseases', *OIE Website, Editorials. Accessed May 21st*.

Van Borm, S., S. Belak, G. Freimanis, A. Fusaro, F. Granberg, D. Hoper, D. P. King, I. Monne, R. Orton, and T. Rosseel. 2015. 'Next-generation sequencing in veterinary medicine: how can the massive amount of information arising from high-throughput technologies improve diagnosis, control, and management of infectious diseases?', *Methods Mol Biol*, 1247: 415-36.

Van Broeckhoven, Christine, Marc Cruts, Wouter De Coster, Svenn D'Hert, and Darrin T Schultz. 2018. 'NanoPack: visualizing and processing long-read sequencing data', *Bioinformatics*, 34: 2666-69.

van Dijk, E. L., Y. Jaszczyszyn, D. Naquin, and C. Thermes. 2018. 'The Third Revolution in Sequencing Technology', *Trends Genet*, 34: 666-81.

Van Reeth, Kristien, Hans Nauwynck, and Maurice Pensaert. 2001. 'Clinical effects of experimental dual infections with porcine reproductive and respiratory syndrome virus followed by swine influenza virus in conventional and colostrum-deprived pigs', *Journal of Veterinary Medicine, Series B*, 48: 283-92.

VanderWaal, K., and J. Deen. 2018. 'Global trends in infectious diseases of swine', *Proc Natl Acad Sci U S A*, 115: 11495-500.

VanderWaal, Kimberly, Robert B Morrison, Claudia Neuhauser, Carles Vilalta, and Andres M Perez. 2017. 'Translating big data into smart data for veterinary epidemiology', *Front Vet Sci*, 4: 110.

Vaser, R., I. Sovic, N. Nagarajan, and M. Sikic. 2017. 'Fast and accurate de novo genome assembly from long uncorrected reads', *Genome Res*, 27: 737-46.

Venkataraman, S., S. P. Reddy, J. Loo, N. Idamakanti, P. L. Hallenbeck, and V. S. Reddy. 2008. 'Structure of Seneca Valley Virus-001: an oncolytic picornavirus representing a new genus', *Structure*, 16: 1555-61.

Ventola, C Lee. 2015a. 'The antibiotic resistance crisis: part 2: management strategies and new agents', *Pharmacy and Therapeutics*, 40: 344.

Ventola, C. L. 2015b. 'The antibiotic resistance crisis: part 1: causes and threats', *P t*, 40: 277-83.

Viehweger, A., S. Krautwurst, K. Lamkiewicz, R. Madhugiri, J. Ziebuhr, M. Holzer, and M. Marz. 2019. 'Direct RNA nanopore sequencing of full-length coronavirus genomes

provides novel insights into structural variants and enables modification analysis', *Genome Res*, 29: 1545-54.
Viehweger, Adrian, Sebastian Krautwurst, Kevin Lamkiewicz, Ramakanth Madhugiri, John Ziebuhr, Martin Hölzer, and Manja Marz. 2018. 'Nanopore direct RNA sequencing reveals modification in full-length coronavirus genomes', *bioRxiv*: 483693.
Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, and A. M. Earl. 2014. 'Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement', *PLoS One*, 9: e112963.
Wang, D., L. Fang, and S. Xiao. 2016. 'Porcine epidemic diarrhea in China', *Virus Res*, 226: 7-13.
Wang, H. M., Y. G. Liu, Y. D. Tang, T. X. Liu, L. L. Zheng, T. Y. Wang, S. G. Liu, G. Wang, and X. H. Cai. 2018. 'A natural recombinant PRRSV between HP-PRRSV JXA1-like and NADC30-like strains', *Transbound Emerg Dis*, 65: 1078-86.
Wang, J., N. E. Moore, Y. M. Deng, D. A. Eccles, and R. J. Hall. 2015. 'MinION nanopore sequencing of an influenza genome', *Front Microbiol*, 6: 766.
Wang, J., J. Qi, H. Zhao, S. He, Y. Zhang, S. Wei, and F. Zhao. 2013. 'Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease', *Sci Rep*, 3: 1843.
Wang, L. F., and G. Crameri. 2014. 'Emerging zoonotic viral diseases', *Rev Sci Tech*, 33: 569-81.
Wang, T., Y. Sun, and H. J. Qiu. 2018. 'African swine fever: an unprecedented disaster and challenge to China', *Infect Dis Poverty*, 7: 111.
Wang, X., D. Marthaler, A. Rovira, S. Rossow, and M. P. Murtaugh. 2015. 'Emergence of a virulent porcine reproductive and respiratory syndrome virus in vaccinated herds in the United States', *Virus Res*, 210: 34-41.
Weirather, J. L., M. de Cesare, Y. Wang, P. Piazza, V. Sebastiano, X. J. Wang, D. Buck, and K. F. Au. 2017. 'Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis', *F1000Res*, 6: 100.
Wernike, K., B. Hoffmann, M. Dauber, E. Lange, H. Schirrmeier, and M. Beer. 2012. 'Detection and typing of highly pathogenic porcine reproductive and respiratory syndrome virus by multiplex real-time rt-PCR', *PLoS One*, 7: e38251.
Woolhouse, Mark EJ, and Liam Brierley. 2018. 'Epidemiological characteristics of human-infective RNA viruses', *Scientific data*, 5: 180017.
Wright, M. H., J. Adelskov, and A. C. Greene. 2017. 'Bacterial DNA Extraction Using Individual Enzymes and Phenol/Chloroform Separation', *J Microbiol Biol Educ*, 18.
Wu, Q., X. Zhao, Y. Bai, B. Sun, Q. Xie, and J. Ma. 2017. 'The First Identification and Complete Genome of Senecavirus A Affecting Pig with Idiopathic Vesicular Disease in China', *Transbound Emerg Dis*, 64: 1633-40.
Wu, Q., X. Zhao, Y. Chen, X. He, G. Zhang, and J. Ma. 2016. 'Complete Genome Sequence of Seneca Valley Virus CH-01-2015 Identified in China', *Genome Announc*, 4.
Yamagishi, J., L. R. Runtuwene, K. Hayashida, A. E. Mongan, L. A. N. Thi, L. N. Thuy, C. N. Nhat, K. Limkittikul, C. Sirivichayakul, N. Sathirapongsasuti, M. Frith, W. Makalowski, Y. Eshita, S. Sugano, and Y. Suzuki. 2017. 'Serotyping dengue virus with isothermal amplification and a portable sequencer', *Sci Rep*, 7: 3510.
Yamagishi, Junya, Lucky R Runtuwene, Kyoko Hayashida, Arthur E Mongan, Lan Anh Nguyen Thi, Linh Nguyen Thuy, Cam Nguyen Nhat, Kriengsak Limkittikul, Chukiat Sirivichayakul, and Nuankanya Sathirapongsasuti. 2017. 'Serotyping dengue virus with isothermal amplification and a portable sequencer', *Sci Rep*, 7: 3510.
Yang, H., L. Paruch, X. Chen, A. van Eerde, H. Skomedal, Y. Wang, D. Liu, and J. Liu Clarke. 2019. 'Antibiotic Application and Resistance in Swine Production in China: Current Situation and Future Perspectives', *Front Vet Sci*, 6: 136.

Yang, K., Y. Tian, D. Zhou, Z. Duan, R. Guo, Z. Liu, F. Yuan, and W. Liu. 2017. 'A Multiplex RT-PCR Assay to Detect and Discriminate Porcine Reproductive and Respiratory Syndrome Viruses in Clinical Specimens', *Viruses*, 9.

Yongkiettrakul, Suganya, Krissana Maneerat, Buppa Arechanajan, Yuwares Malila, Potjanee Srimanote, Marcello Gottschalk, and Wonnop Visessanguan. 2019. 'Antimicrobial susceptibility of Streptococcus suis isolated from diseased pigs, asymptomatic pigs, and human patients in Thailand', *BMC Vet Res*, 15: 5.

Zhang, J., Y. Zheng, X. Q. Xia, Q. Chen, S. A. Bade, K. J. Yoon, K. M. Harmon, P. C. Gauger, R. G. Main, and G. Li. 2017. 'High-throughput whole genome sequencing of Porcine reproductive and respiratory syndrome virus from cell culture materials and clinical specimens using next-generation sequencing technology', *J Vet Diagn Invest*, 29: 41-50.

Zhang, Z., Y. Zhang, X. Lin, Z. Chen, and S. Wu. 2019. 'Development of a novel reverse transcription droplet digital PCR assay for the sensitive detection of Senecavirus A', *Transbound Emerg Dis*, 66: 517-25.

Zhao, K., C. Ye, X. B. Chang, C. G. Jiang, S. J. Wang, X. H. Cai, G. Z. Tong, Z. J. Tian, M. Shi, and T. Q. An. 2015. 'Importation and Recombination Are Responsible for the Latest Emergence of Highly Pathogenic Porcine Reproductive and Respiratory Syndrome Virus in China', *J Virol*, 89: 10712-6.

Zhao, X., Q. Wu, Y. Bai, G. Chen, L. Zhou, Z. Wu, Y. Li, W. Zhou, H. Yang, and J. Ma. 2017. 'Phylogenetic and genome analysis of seven senecavirus A isolates in China', *Transbound Emerg Dis*, 64: 2075-82.

Zhao, Y., F. Liu, Q. Li, M. Wu, L. Lei, and Z. Pan. 2019. 'A multiplex RT-PCR assay for rapid and simultaneous detection of four RNA viruses in swine', *J Virol Methods*, 269: 38-42.

Zhao, Z., Y. Qin, Z. Lai, L. Peng, X. Cai, L. Wang, X. Guo, and H. Yang. 2012. 'Microbial ecology of swine farms and PRRS vaccine vaccination strategies', *Vet Microbiol*, 155: 247-56.

Zhou, L., R. Kang, B. Xie, Y. Tian, X. Yang, J. Yu, and H. Wang. 2018. 'Complete Genome Sequence of a Recombinant NADC30-Like Strain, SCnj16, of Porcine Reproductive and Respiratory Syndrome Virus in Southwestern China', *Genome Announc*, 6.

Zhou, L., B. Yang, L. Xu, H. Jin, X. Ge, X. Guo, J. Han, and H. Yang. 2017. 'Efficacy evaluation of three modified-live virus vaccines against a strain of porcine reproductive and respiratory syndrome virus NADC30-like', *Vet Microbiol*, 207: 108-16.

Zhou, L., and H. Yang. 2010. 'Porcine reproductive and respiratory syndrome in China', *Virus Res*, 154: 31-7.

Zimmerman, Jeffrey J, Locke A Karriker, A Ramirez, Kent J Schwartz, Gregory W Stevenson, and JQ Zhang. 2019. 'Diseases of swine', *Diseases of swine, 11th Edition.*: 123 - 57.

Zukurov, Jean P, Sieberth do Nascimento-Brito, Angela C Volpini, Guilherme C Oliveira, Luiz Mario R Janini, and Fernando Antoneli. 2016. 'Estimation of genetic diversity in viral populations from next generation sequencing data with extremely deep coverage', *Algorithms for Molecular Biology*, 11: 2.