Applications of Viral Metagenomics in the Veterinary Field

Looking for the Unknown

Anne-Lie Blomström

Faculty of Veterinary Medicine and Animal Sciences
Department of Biomedical Sciences and Veterinary Public Health
Uppsala

Applications of Viral Metagenomics in the Veterinary Field
Looking for the Unknown

Abstract

Viral metagenomics provide a powerful technology to investigate the viral flora of healthy and sick animals. Using these methodologies, we gain a better understanding in the etiology of diseases, as well as deepen our knowledge into the viruses circulating in nature and the complex interaction between virus and host.

The aim of this thesis was to utilize viral metagenomics into different areas of interest for veterinary science. One multifactorial disease complex was studied, as were a disease of unknown etiology, where traditional methods had failed to identify a causative agent. The final study was on viruses in an arthropod vector. Postweaning multisystemic wasting syndrome (PMWS) was the multifactorial disease chosen here. Though, porcine circovirus type 2 (PCV-2) has been found the causative agent other factors, such as viruses, are believed to influence the disease. Using multiple displacement amplification followed by large-scale sequencing we discovered, apart from PCV-2 and torque teno virus, a novel porcine bocavirus in the background of PCV-2 in lymph nodes collected from pigs suffering from PMWS. This co-infection was seen in a high percentage (71%) of pigs suffering from PMWS compared to only 33% in pigs without PMWS. Shaking mink syndrome (SMS) was used to investigate a disease of unknown etiology. By random amplification and large-scale sequencing we found an astrovirus in the brain of minks experimentally infected with brain homogenate from diseased animals. Astrovirus was also detected in the brain of naturally infected minks. Vectors are important transmitters of disease, and in the final study, soft ticks (*Ornithodoros*) collected from a warthog burrow in Uganda were investigated for the presence of viruses. Among others, we discovered a possible novel RNA virus that showed a distant relationship to hepatitis E virus.

In conclusion, viral metagenomics have successfully been applied to investigate three important areas for veterinary science and through these studies three novel viruses were discovered and genetically characterized.

Keywords: viral metagenomics, Postweaning multisystemic wasting syndrome, Shaking mink syndrome, soft ticks (*Ornithodoros*), bocavirus, astrovirus

Author's address: Anne-Lie Blomström, SLU, Department of Biomedical Sciences an
Veterinary Public Health, P.O. Box 7028, 750 07 Uppsala, Sweden
E-mail: Anne-Lie.Blomstrom@bvf.slu.se

Contents

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Blomström, A-L, Belák, S., Fossum, C., McKillen, J., Allan, G., Wallgren, P. & Berg, M. (2009). Detection of a novel porcine boca-like virus in the background of porcine circovirus type 2 induced postweaning multisystemic wasting syndrome. *Virus Research* 146(1-2), 125-129.

II Blomström, A-L, Belák, S., Fossum, C., Fuxler, L., Wallgren, P. & Berg, M. (2010). Studies of porcine circovirus type 2, porcine boca-like virus and torque teno virus indicate the presence of multiple viral infections in postweaning multisystemic wasting syndrome pigs. *Virus Research* 152(1-2), 59-64.

III Blomström, A-L, Widén, F., Hammer, A-S., Belák, S. & Berg, M. (2010). Detection of a novel astrovirus in brain tissue of mink suffering from shaking mink syndrome using viral metagenomics. (Manuscript accepted to be published in Journal of clinical microbiology; doi:10.1128/JCM.01940-10).

IV Blomström, A-L, Ståhl, K., Masembe, C., Okurut, A. R., Atmned, P., Kempe, S., Bishop, R., Belák, S. & Berg, M. (2010). Viral metagenomic analysis of soft ticks (*Ornithodoros*) in Uganda reveals a novel RNA virus with distant relationship to hepatitis E virus. (Manuscript)

Papers I–III are reproduced with the permission of the publishers.

Abbreviations

BLAST   Basic local blast alignment search tool
cDNA   Complementary DNA
DNA   Deoxyribonucleic acid
DNAse   Deoxyribonuclease
EID   Emerging infectious disease
EM   Electron microscopy
HEV   Hepatitis E virus
MDA   Multiple displacement amplification
PCR   Polymerase chain reaction
PCV-2   Porcine circovirus type 2
PMWS   Postweaning multisystemic wasting syndrome
RCA   Rolling circle amplification
RDA   Representational difference display
RNA   Ribonucleic acid
RNase   Ribonuclease
rPCR   Random PCR
SISPA   Sequence independent, single-primer amplification
SMS   Shaking mink syndrome
TTV   Torque teno virus
VI   Virus isolation

1 Introduction

Throughout recorded history, viruses have caused diseases in animals and humans, with descriptions of possible viral infections appearing long before viruses were first discovered. For example, a hieroglyph from 1400 B.C. shows a man displaying the typical symptoms of a poliovirus infection, and this picture is only one of many reports of historical occurrences of viral diseases (Flint *et al.*, 2009). The concept of a virus (meaning poison or toxic) was introduced in 1898 by Beijerinck, who found that the agent causing tobacco mosaic disease in plants could not, as bacteria, be filtered out and could not grow by itself but required living and growing cells in which to replicate (Beijerinck, 1898). After Beijerinck's discovery, many viruses were found. The first animal virus (foot-and-mouth-disease virus) was discovered the same year by Loeffler and Frosch (Mahy, 2005).

Today, we know that viruses are all around us. The ocean is estimated to contain 10^6 to 10^9 virus particles per milliliter of seawater (Suttle, 2005); and we encounter viruses daily via the air, and contact with other people and animals. In most cases, these encounters do not harm us because the immune system clears the infection or the virus is not able to establish an infection due to physical barriers or a lack of permissive and susceptible cells. However, viruses have many different ways to evade the host immune system, and when they are able to cross cellular barriers and complete their infectious cycle disease can occur, though some infections are latent despite successful infection (Flint *et al.*, 2009).

While we currently have extensive knowledge of the agents of many diseases, there are still many disorders for which the etiology is unknown, though evidence indicates that the cause is likely infectious. This introduction will focus on modern molecular techniques that can be used to study viral populations and discover "unknown" viruses in samples from

both healthy and sick individuals but classical methods for virus discovery will also be described.

1.1 Emerging infectious diseases

Emerging infectious diseases (EIDs) occur due to the introduction of a pathogen into a new and susceptible population, and during the past 30 years, over 50 new human pathogens have been reported by the World Health Organization (www.who.int), some of which are shown in Table 1. A vast number of these diseases are zoonoses that spread through transmission of the virus from animals to humans. Many interactions between the host and the virus are in a steady state, meaning that both survive and can multiply while the infection is in (many cases) asymptomatic for the host (Feldmann *et al.*, 2002). However, a number of factors can disturb this interaction and an EID can arise. These factors include viral evolution, ecological factors, international traveling, and climatic changes. Many EIDs are the result of a change in the host and/or viral ecology. As humans encroach on wildlife habitats through the expansion of urban areas, humans and their domestic animals come into contact with wildlife carrying viruses that they have not previously encountered. In some cases, transmission to new hosts leads to disease (Daszak *et al.*, 2000; Morse, 1995). For example, in 1998 in Malaysia, a number of people died from a neurological disease caused by an agent that was, at that time unrecognized. The disease in humans was correlated with a mild respiratory impairment and occasional neurological disease in pigs. The causative agent was found to be a novel paramyxovirus: Nipah virus. Nipah virus had a stable virus-host interaction with fruit bats, but due to urbanization, the domestic pigs became infected, possibly by eating fruit contaminated with bat saliva or urine. The disease was then spread from the pigs to humans through aerosols. Apart from 105 human deaths, more than one million pigs were killed (Daniels *et al.*, 2007; Chua *et al.*, 2000). Another EID from the last 30 years is the Hantavirus Pulmonary Syndrome, caused by Sin Nombre virus. In this instance, climate changes were a factor leading to the spread of the virus from the deer mouse to humans in 1993 (Klein & Calisher, 2007). In 2003, the first case of severe acute respiratory syndrome (SARS) was reported. SARS coronavirus, the causative agent of SARS, was spread to humans from bats through civets in the animal market. The disease spread to > 8000 people in 29 countries, highlighting the impact of international travel on viral spread (Berger *et al.*, 2004). The same is true for animal

trading, which can spread disease globally. Apart from being a threat to animal and human health these outbreaks can also have severe economic consequences for both individual farmers and affected nations, due to animal loss and trade restriction. Even if the disease does not lead to animal loss it can still be a considerable problem because it can cause reproduction disorders and decreased growth. Therefore, it is important to have tools that can rapidly identify the causative agents behind new EIDs.

Table 1. Examples of emerging and re-emerging diseases and some of the contributing factors behind their emergence

Disease	Virus	Contributing factors	Hosts
SARS	Coronavirus	Urbanization, international travel	Fruit bat, civet, human
Hantavirus pulmonary syndrome	Sin Nombre virus	Climate change	Deer mouse, human
Hendra virus disease	Hendravirus	Encroachment	Fruit bat, horse, human
West Nile Fever	West Nile virus	Climate, migrating birds	Mosquito, bird, human, horse
Dengue fever/dengue shock syndrome	Dengue virus	Urbanization, international travel, no effective mosquito control	Mosquito, human, lower primates

When faced with a disease, an investigation is initiated (Fig. 1) to identify the etiological agent(s), prevent the spread of disease, and if possible, treat the affected animals. Clinical symptoms, as well as necropsy findings and histopathological changes can often provide an insight to the etiology. However, to further confirm the etiology, additional laboratory diagnostic procedures are required. There are a many different diagnostic assays that can be used to detect a specific agent in a clinical specimen. For virus detection, virus isolation (VI), electron microscopy (EM), polymerase chain reaction (PCR), immunoflourescence (IF), and antigen ELISA (AgELISA), are termed "direct diagnostic methods" because these assays directly detect the virus particle or its components (e.g., viral nucleic acids or viral proteins). Other techniques, for example serum-neutralization (SN), hemagglutination-inhibition (HI) tests, or ELISA, are used to detect the antibody response of the host to the virus. Because these assays do not

directly detect the virus particles but estimate the immune response of the host, they are termed "indirect diagnostic methods". The various direct and indirect methods all have their specific strengths and weaknesses. For example, an advantage of the VI and the EM methods is that they have the capacity to detect a wide range of viruses in clinical samples without prior knowledge in to the virus (Quinn *et al.*, 2002; Dinter, 1989). The majority of the molecular diagnostic methods, such as various PCR assays, do not have this broad detection capacity because they are constructed for the specific detection and identification of targeted viruses or virus families. Molecular methods, such as microarrays and various sequencing methods, are/have been developed to, as VI and EM, detect a very wide range of known and unknown viruses (Tang & Chiu, 2010; Ambrose & Clewley, 2006; Muerhoff *et al.*, 1997).
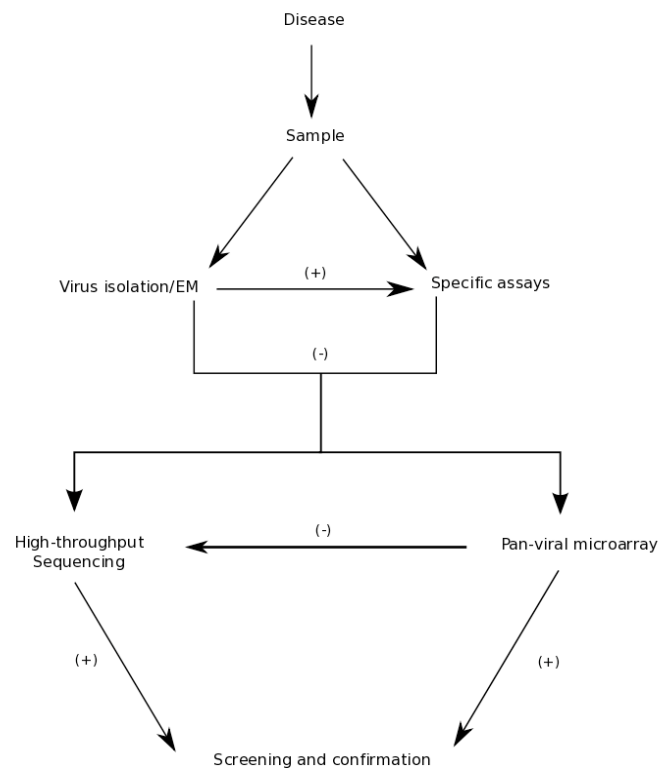

Figure 1. Flow chart of possible actions taken during a disease outbreak. (+) and (–) indicate whether a test is positive or negative, respectively.

12

1.2 Classical virus discovery methods

One important contribution to virus research was the establishment of systems with the ability to isolate and propagate viruses. Unlike bacteria, viruses require living cells to survive and replicate. In the early history of virology, viruses were propagated through passages in animals. However, because this method was both time-consuming and expensive, researches turned to the use of embryonated eggs and, later, cell-culture systems for the propagation and isolation of viruses (Leland & Ginocchio, 2007). Vaccinia virus (Steinhardt *et al*., 1913) is an example of a virus that was cultivated by cell culture in the early 1900s, but it was not until the mid–1900s that cell culture had a true impact on virology. At this time, antibiotics were introduced in cell media, reducing the problem of bacterial contamination. Another important advance was the development of continuous cell lines. The first immortal cell line, still widely used today, is that derived from tumor tissue taken from Henrietta Lacks in 1951, from whom the cell line derives its name: HeLa (Flint *et al*., 2009). At present, there are a number of continuous cell lines in use. These two advances made virus propagation more reliable and reproducible.

There are numerous classical methods of virus diagnosis, but in the following two sections, only VI and EM will be discussed because unlike many other techniques (e.g., IF and VN), they do not test for a specific virus in a sample but have a broad applicability.

1.2.1 Virus isolation

Isolating an unknown virus (or even known virus) is not an easy task. Because each virus, in general, has the capacity to replicate on certain types of cells, virological laboratories prepare and maintain a wide range of cell cultures. After inoculation, the cells are studied for the cytopathic effect (CPE) over a period of time, which can vary in length (days to weeks). If virus replication results in the development of CPE, the changes in the cells facilitate diagnosis. However, a number of viruses do not replicate in cell cultures, and sometimes only weak/non existent CPE is seen. In the absence of apparent CPEs other techniques such as EM (discussed in the next section) and hemadsorption (HAD) can be used to find the virus. HAD is used to identify viruses with the ability to agglutinate erythrocytes, a property which is exhibited by a number of viruses (e.g., influenza virus, Newcastle disease virus and adenovirus) (Leland & Ginocchio, 2007; Dinter, 1989). Virus isolation has an advantage over many other methods of analysis in that it shows that there is viable virus in the sample. Another important

consideration is the necessity to obtain virus isolates and strains for many purposes, such as detailed studies of virus biology, experimental infections, as well as vaccine development and production, among other applications.

1.2.2 Electron microscopy

One way to detect viruses and visualize them is through EM, which can be used for both viruses that are able to replicate in cell cultures and for those that do not. Apart from infected cells, various fluids (e.g., urine and serum) can be examined as well as feces and tissues. By using EM, the virus particles are visualized, and with knowledge of the viral structure and the location of the virus particles, one can identify the type of virus (i.e., whether it is a DNA or RNA virus or which family it belongs to). Negative staining is often used and is a rapid procedure, though the sensitivity of this method is somewhat low (10^6 particles/ml) (Goldsmith & Miller, 2009; Wild, 2008; Hazelton & Gelderblom, 2003). One of the first viruses to be visualized through EM was the tobacco mosaic virus in 1939 and since that time a vast number of viruses have been studied by this technique. Norovirus was discovered and associated with gastroenteritis by EM visualization (Kapikian *et al.*, 1972) and the Nipah virus (discussed under EIDs) was also discovered via EM (Chua *et al.*, 2000). With this technique, one does not only see the virus and its structure but also various viral processes, such as viral budding, can be studied.

1.3 Viral Metagenomics

As previously mentioned, not all microorganisms can be cultivated, and therefore, some may go unnoticed and unstudied. Various molecular methods such as PCR have been used to detect specific viruses or viral families. These methodologies are routinely used and have both high sensitivity and specificity. PCR can be used to detect a few copies of a particular nucleic acid, and the viral load in a sample can be quantified with real-time PCR (Mackay *et al.*, 2002). Although many of the PCR assays investigate the presence of a specific virus, assays have been developed to detect several viruses simultaneously. These multiplex PCR assays have been developed for various viruses connected to different syndromes such as encephalitic and respiratory syndromes (Elnifro *et al.*, 2000). For example, multiplex PCR was successfully used in 2010 to detect between two and four different respiratory viruses, such as influenza A and B, adenovirus and

coronavirus, in each assay (Lassauniere *et al.*, 2010). The multiplexing capacity of such assays is hindered by the availability and number of fluorescent dyes, primer-dimer formation, and other considerations. A MassTag PCR approach combining PCR with mass spectrometry, has been applied to successfully detect 22 different respiratory pathogens such as influenza, coronavirus and rhinovirus (Briese *et al.*, 2005). However, it is not always feasible to utilize these specific assays because they are based on the use of specific sequences to detect their targets. Metagenomics is a more general approach to study the genetic composition of an uncultured sample (Handelsman, 2004). Thus, viral metagenomics investigates the complete genetic viral population in the sample studied (Delwart, 2007). During this decade metagenomics has been used to study viral populations in a number of environments. In 2002, shotgun sequencing was applied to DNA extracted from 200 L of seawater. More than 65% of the sequences could not be identified, but a large portion of the identified sequences was from phages, covering most of the major dsDNA phage families (Breitbart *et al.*, 2002). Since then, other studies have been performed, including the viral RNA world (Culley *et al.*, 2006). Similar techniques have also been used to study the viral population in various environments, such as feces (Breitbart *et al.*, 2003), blood (Breitbart & Rohwer, 2005) and potential viral reservoirs (e.g., bats) (Li *et al.*, 2010). Because these techniques investigate the complete viral community within a sample, it is possible to study both the natural viral flora and emerging pathogens in disease complexes. The following sections briefly describe the major techniques used to study viral metagenomics, as well as discuss the problem of sample preparation. Figure 2 shows a possible workflow for a viral metagenomic study, using sequencing as a final output.
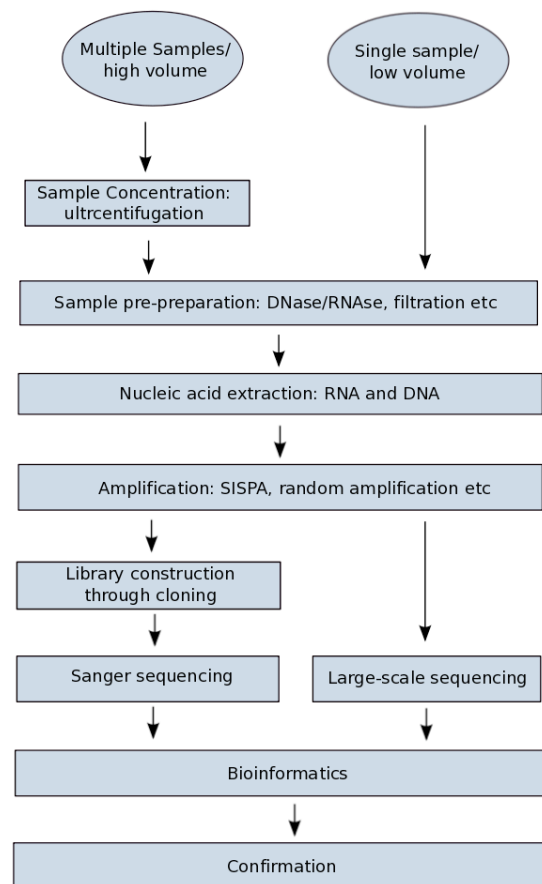
```
              ┌────────────────────┐       ┌────────────────┐
              │  Multiple Samples/ │       │ Single sample/ │
              │    high volume     │       │   low volume   │
              └─────────┬──────────┘       └───────┬────────┘
                        │                          │
                        ▼                          │
              ┌────────────────────┐               │
              │ Sample Concentration:│             │
              │   ultrcentifugation │               │
              └─────────┬──────────┘               │
                        │                          │
                        ▼                          ▼
              ┌─────────────────────────────────────────┐
              │ Sample pre-preparation: DNase/RNAse,     │
              │            filtration etc                │
              └───────────────────┬─────────────────────┘
                                  ▼
              ┌─────────────────────────────────────────┐
              │   Nucleic acid extraction: RNA and DNA   │
              └───────────────────┬─────────────────────┘
                                  ▼
              ┌─────────────────────────────────────────┐
              │  Amplification: SISPA, random amplification etc │
              └─────────┬───────────────────┬───────────┘
                        │                   │
                        ▼                   │
              ┌────────────────────┐        │
              │ Library construction│       │
              │  through cloning   │        │
              └─────────┬──────────┘        │
                        │                   │
                        ▼                   ▼
              ┌────────────────────┐  ┌────────────────────┐
              │  Sanger sequencing │  │ Large-scale sequencing │
              └─────────┬──────────┘  └─────────┬──────────┘
                        │                       │
                        ▼                       ▼
              ┌─────────────────────────────────────────┐
              │              Bioinformatics              │
              └───────────────────┬─────────────────────┘
                                  ▼
              ┌─────────────────────────────────────────┐
              │               Confirmation               │
              └─────────────────────────────────────────┘
```

Figure 2. Scheme for a possible workflow using sequence–independent amplification and sequencing to study viral metagenomics.

1.3.1 Sample preparation

For all of the methods mentioned below sample selection and preparation is crucial. One of the major obstacles in attempting to detect novel viruses using viral metagenomic approaches in clinical samples is a lack of sensitivity and the "contamination" of host nucleic acid. The viral nucleic acid often only constitutes a small proportion of the total nucleic acids from an extracted sample. Therefore, it is crucial to concentrate the virus particles, separate them from the cells and/or degrade the nucleic acids of the host. The most common methods are centrifugation, filtration and nuclease

treatment (Thurber *et al.*, 2009; Delwart, 2007). One possible way to increase the likelihood of identifying viruses is by pooling a large number of samples suspected to contain the agent. Pooling can be done with ease, especially when working with fluid samples, such as serum, blood, and suspended swabs. Ultracentrifugation can subsequently be used to concentrate the virus.

Bacteria and larger cell complexes can be removed by filtration. Filters (0.22-μm pore) are frequently used to remove bacteria when searching for viruses because it was historically believed that all viruses could pass through filters of that size (Delwart, 2007). However, in recent years, viruses have been discovered that are almost as large as bacteria. For example, mimivirus is a DNA virus with a genome of an approximately 1.2 Mb and a capsid with a diameter of approximately 500 nm (Van Etten *et al.*, 2010), which would therefore likely not be detected in an assay using 0.22-μm pore filters. Viral particles can also be separated from other particles based on density using sucrose cushions and cesium-chloride gradient separation (Thurber *et al.*, 2009).

Nucleases are frequently used to reduce the host nucleic-acid background. The principle behind this technique is that the viral capsid protects the viral nucleic acid from the nuclease, while the host nucleic acid is degraded (Allander *et al.*, 2001). This notion is true to a certain extent, and currently, DNase treatment is included in most viral metagenomic approaches, especially in those approaches using sequencing as the output. One should, however, keep in mind that different viruses have different stabilities and may have varying sensitivities to nuclease treatment. To remove host RNA, RNase is sometimes used (Tang & Chiu, 2010; Delwart, 2007). The major problem with RNA contamination is that it is difficult to degrade ribosomal RNA (rRNA). As a consequence, a high proportion of rRNA remain, even after RNase treatment. Thus, achieving the total degradation/removal of host nucleic acid is difficult, but using nuclease treatment, it may be possible to equilibrate the host/viral nucleic acid ratio. It is important to be aware of this balance when trying to degrade the host nucleic acid while preserving that of the virus and remembering that the virus may be, for example, in an episomal form and, therefore, not well-protected.

A non-nuclease approach to eliminate rRNA was suggested by Endoh *et al.* (2005). They designed a set of 96 primers that was used during cDNA synthesis instead of random primers. The primers were designed so that they would not prime rRNA but would still target all known mammalian viruses in the database (Endoh *et al.*, 2005). It is also possible to extract ribosomal sequences from a sample. An example of this is RiboMinus technology

(Invitrogen) that contains a number of highly conserved ribosomal sequences coupled to streptavidin. The total nucleic acid is incubated with the coupled ribosomal sequences, allowing hybridization between the synthetic sequences and those of the sample, and the hybridized sequences can then be removed from the sample.

1.3.2 Microarray

Microarray analysis was initially employed to study gene expression but has been increasingly utilized in the detection of various pathogens, such as viruses. Microarray analysis has a multiplexing capacity that, by far, outnumbers PCR, and it can be used to detect up to several thousand different viruses. The microarray is based on designing viral probe sequences that will hybridize to their corresponding viral sequence in the sample (if available). To increase sensitivity, the nucleic acid can be amplified, either specifically or randomly (Miller & Tang, 2009). In some cases, the sample can also be subjected to DNase treatment, as discussed in the previous section. Different microarrays have been designed to probe for viruses known to cause specific disorders. Microarray analysis has also been used to subtype and genotype various viruses, such as influenza (Metzgar *et al.*, 2010; Gall *et al.*, 2009). In recent years, panviral microarrays have been designed in an attempt to detect all known viruses. The two major panviral microarrays are the Virochip, detecting an estimated 1000 viruses with 10,000 probes (Wang *et al.*, 2003; Wang *et al.*, 2002), and the GreenChip, detecting 1710 viruses with 9477 viral probes (Palacios *et al.*, 2007). They differ in a number of aspects, one being the probe design. Whereas Greenchip is based on protein family database (Pfam), Virochip is, on the other hand, based on viral nucleotide sequences from GenBank. Both arrays, however, utilize random amplification before hybridization and have relatively long probes, approximately 70-mers. Both arrays use these longer probes, compared to, for example, 25-mers, to detect more divergent viruses because longer probes allow for more mismatches to occur. It is also possible to recover a hybridized sequence and determine its exact nucleotide order by physical scraping or elution, followed by sequencing of the nucleic acid. The Virochip was utilized to discover a novel coronavirus in beluga whales (Mihindukulasuriya *et al.*, 2008) and an avian bornavirus in birds suffering from proventricular dilatation disease (Kistler *et al.*, 2008). Thus, these techniques can be used to detect novel and unexpected viruses in samples, though the success of the assay is dependent on the probe sequences designed and placed on the array.

1.3.3 Subtraction hybridization assays

Representational difference display (RDA), developed in 1993, is based on subtraction hybridization and is used to determine differences between two complex DNA populations (Fig. 3) (Lisitsyn & Wigler, 1993). Thus, two samples are required: one (the tester) containing the unknown target, for example, viral nucleic acid; and one (the driver) without the target. It is important that these samples, apart from the target, be as similar as possible in nucleic acid composition. Ideally, one would have samples from the same individual before and after infection; in natural infections, these samples are, of course, usually not available. Indeed, the major limitation of RDA is the problem of finding an appropriate driver sample. Both samples are digested with restriction enzymes to reduce genome complexity before ligation of adaptors to the cut tester sample. The tester and the driver are then mixed and allowed to hybridize, and the ends are filled in. Three different possible hybridization products are available: 1) hybridization of unique tester sequences; 2) tester sequence hybridized to a driver; and 3) excess driver. From these three variants, it is only the first that can be amplified efficiently by PCR because only unique tester sequences contain the adaptor sequence on both ends. Thus, by PCR, one can enrich for the unique sequences, which can then be cloned and sequenced. To further enrich for these sequences, the PCR product can be used as a template in additional tester-driver hybridizations to discard common sequences that remain. Since the first description of RDA, several modified protocols have been developed, such as Selectively Primed Adaptive Driver – RDA (SPAD-RDA) (Birkenmeyer *et al.*, 2003) and the Primer Extension Enrichment Reaction (PEER) (Ganova-Raeva *et al.*, 2006), with various improvements to increase the sensitivity of the assay. Only one year after the first publication of RDA this technique was used to discover human herpesvirus 8, which is associated to Kaposi's sarcoma (Chang *et al.*, 1994). Subsequently, several other viruses have been discovered using RDA such as torque teno virus (TTV) (Nishizawa *et al.*, 1997) and GBV-A and B (Simons *et al.*, 1995).
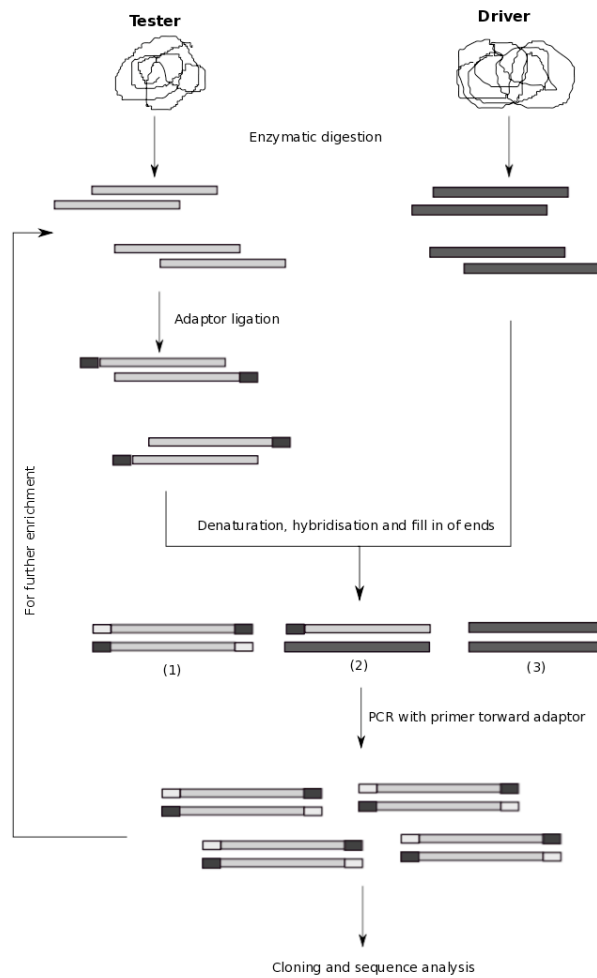
Figure 3. The principle of classical RDA. Tester and driver DNA are fragmented, and adaptor sequences are then ligated to the tester sequences. The two samples are mixed and allowed to hybridize. After an end fill-in reaction, PCR is performed using primers to the adaptor sequence. Only in the case where tester DNA hybridizes to another tester DNA sequence (1) can amplification efficiently occur, and the unique sequence can be amplified.

1.3.4 Sequence-independent amplification assays

By using assays that do not require specific primers or probes, it is possible to identify highly divergent and novel viruses that may otherwise go unrecognized. It is also possible to study coinfections by several viruses (Ambrose & Clewley, 2006). Shotgun DNA libraries can be created with DNA fragmented by sonication, partial DNase treatment or restriction enzyme digestion, followed by ligation into a vector and screening for the

20

genetic composition by sequencing or immunoscreening without the need for amplification (Deininger, 1983; Anderson, 1981). One problem with this approach is that a large amount of DNA is needed. Therefore, various strategies for amplification have been used to randomly amplify all of the nucleic acid in the sample. In 1989, Akowitz and Meanuelidis developed a protocol based on the ligation of adaptors to cDNA, followed by PCR using primers against the adaptors to generate cDNA libraries from small amounts of RNA (Akowitz & Manuelidis, 1989). A very similar approach, sequence-independent, single-primer amplification (SISPA), was published in 1991. Using SISPA, it was possible to create a library of 10^6 recombinants from 1 pg of RNA (Reyes & Kim, 1991). This technique was, in the same year, used in combination with immunoscreening to identify a Norwalk virus sequence from 1.5 g feces (Matsui *et al.*, 1991). Different versions of SISPA have been developed, and many SISPA protocols include digestion of the DNA and cDNA with restriction enzymes before ligation of the adaptor. Due to the high host background, filtration and DNase steps were introduced (as discussed under sample preparation) to aid the search for novel/previously unrecognized viruses. For instance, using the DNase-SISPA protocol, two bovine parvoviruses were discovered in bovine sera (Allander *et al.*, 2001).

To avoid the need for restriction enzyme digestion and ligation, random amplification protocols have been developed. This technique was developed in 1992 for RNA using a random primer with an attached adaptor sequence at the 5' end. Through a ds cDNA synthesis, the cDNA is labeled at both ends, and a PCR can be performed using a primer against the adaptor (Froussard, 1992). Similar protocols were developed for DNA (Fig. 4). By applying random PCR (rPCR) following filtration and DNase treatment to respiratory tract samples, a library of approximately 860 clones was produced. By sequencing and data analysis, 20% of these were found to be of viral origin, including a novel human bocavirus, and the sensitivity was estimated to 10^5 when sequencing 96 clones (Allander *et al.*, 2005). Several modified rPCR protocols have also been developed for SISPA. With the new sequencing technologies discussed in the next section, rPCR has become a useful tool for virus discovery.
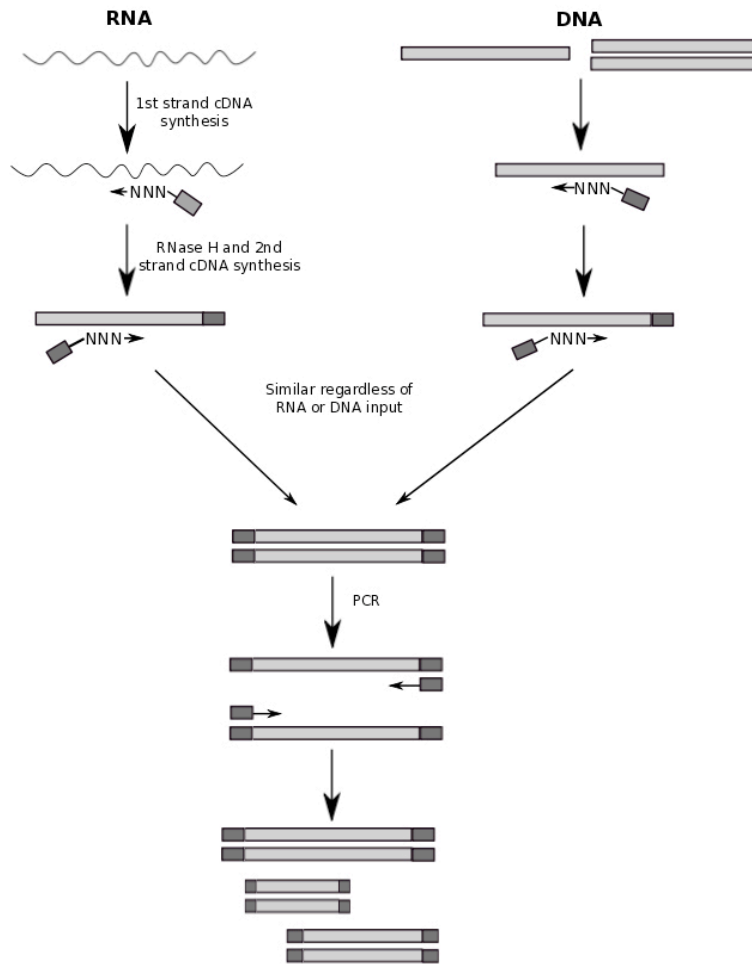
Figure 4. The principle of random amplification of RNA and DNA using a random primer with an arbitrary sequence attached.

Nucleic acid can also be efficiently amplified by phi29 in an isothermal reaction. Starting with only a few nanograms of DNA, several milligrams can be produced. Phi29 is a displacement polymerase that is very processive and can incorporate over 70,000 bases before detaching from its template. Apart from polymerization activity, the enzyme also has a 3' − 5' DNA exonuclease activity with an error rate of 1 in $10^6 - 10^7$ (Johne *et al.*, 2009). Random primers were, in combination with phi29, used to amplify circular DNA targets, increasing it 10,000-fold through rolling-circle amplification

(RCA) (Fig. 5a) (Dean *et al.*, 2001). Because a number of viruses have circular genomes, they can be amplified in a similar manner. In 2004, random primer and RCA techniques were used to identify a papillomavirus from tissue without prior sequence knowledge (Rector *et al.*, 2004). Although circular targets are ideal for this type of amplification, linear targets can also be used by multiple displacement amplification (MDA) (Fig. 5b). In MDA, the efficiency of the reaction is, to some extent, based on the length of the target, which may be unfavorable for short linear DNA viruses and short cDNA. However, ligation can be used to overcome this problem, and several RNA viruses were successfully amplified from tissues using phi29 (Berthet *et al.*, 2008).
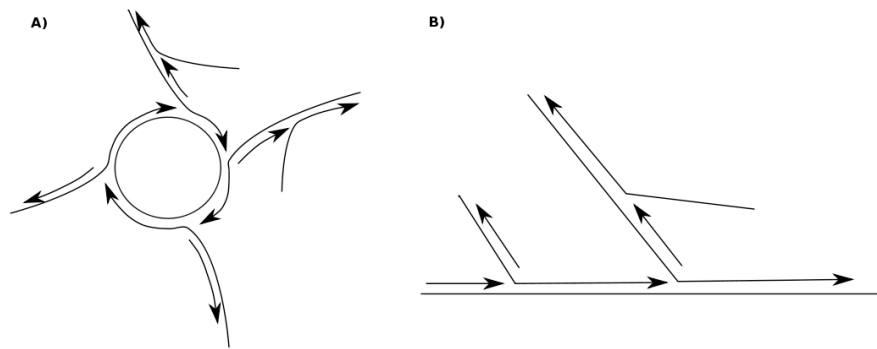

Figure 5. Representation of the principle of rolling-circle amplification (A) on a circular target and multiple displacement amplification (B) of a linear target.

1.3.5 Large-scale sequencing

Until recently, the products of the above-mentioned amplification technologies have usually been cloned into bacteria to create libraries that are characterized by Sanger sequencing to identify any potential viruses in viral metagenomic studies. However, this process is quite laborious, and due to a combination of the high background of contaminating host nucleic acid and the occasionally low levels of virus, a vast number of clones may have to be sequenced before a viral sequence is identified.

Around 2004, three independent platforms were introduced for sequencing that both obviate the need for traditional cloning after amplification and yield several hundred thousand sequence reads in one run. These technologies are often referred to as next-generation sequencing or high-throughput sequencing (Kircher & Kelso, 2010; Shendure & Ji, 2008). The

first platform to be introduced was the 454 technology, shortly followed by Illumina and the SOLiD platforms. These techniques have different sequencing principles but all have high throughput (relative to Sanger sequencing) in common. None of them require a bacterial cloning step prior to sequencing.

The 454 platform uses emulsion PCR to clonally amplify each DNA fragment on an individual bead. The beads are then added together with all of the various sequencing enzymes in wells on a picotiter plate containing approximately two million wells. The sequencing occurs via pyrosequencing, and the emitted light is captured with a high-resolution, charge-coupled device. Approximately 750 Mb sequence data can be obtained from one run, with reads of a sequence length between $300 - 500$ nt (Kircher & Kelso, 2010; Margulies *et al.*, 2005).

The SOLiD technology also uses emulsion PCR to clonally amplify the DNA sequences but is based on a different sequencing strategy. The sequencing reaction is not dependent on the action of a polymerase incorporating nucleotides. Instead, it is based on the ligation of a fluorescence-labeled probe to the target sequence. Compared to the 454 platform, more data are obtained, but the sequence reads are shorter, usually ranging between 25 and 75 nt.

Illumina does not clonally amplify the DNA using emulsion PCR but instead uses a bridge PCR. Additionally, the sequencing technology more closely resembles Sanger sequencing than the other technologies. It is based on the incorporation of reversible terminators, which are fluorescence-labeled, and after each cycle, images for four channels are obtained. Like SOLiD, this technique yields more data than the 454 platform, but the read length is also shorter, approximately 100 nt (Kircher & Kelso, 2010; Shendure & Ji, 2008).

All of these sequencing technologies have their own strengths and weaknesses that should be considered before choosing which platform to use. For metagenomic studies, 454 sequencing is often used, mainly due to the longer read length, which makes de novo assembly easier. Due to the vast amount of data produced by the different sequencing technologies, knowledge of bioinformatics and access to bioinformatic tools are necessary to extract the information of interest. Thus, a number of tools (both web applications and computer software) have been developed to aid with assembly and annotation of the sequences. (Nowrousian, 2010). For example, in the MEGAN software, BLAST outputs from a metagenomic study are used as inputs, and these data are used to produce a taxonomical analysis in which each read is assigned to a node in the NCBI taxonomy

(Huson *et al.*, 2009; Huson *et al.*, 2007). An example of a web-based tool to utilize BLAST is the personal BLAST navigator (PLAN). In this platform, the sequences are uploaded, and automated BLAST searches are performed. The data can then be searched, filtered and downloaded into spreadsheets (He *et al.*, 2007). A problem with these tools is, however, that they are not always maintained and updated for a longer period after relase/publication.

Using these sequencing technologies in combination with sequence-independent amplification and bioinformatics, several novel viruses have been discovered. A new arenavirus was found in three patients who died of a febrile illness after they received organs from the same donor (Palacios *et al.*, 2008). Independently of the detection of the avian bornavirus in birds with proventricular dilatation disease using Virochip (Kistler *et al.*, 2008) a study using random amplification and 454 sequencing also found an avian bornavirus in birds with the same disease (Honkavuori *et al.*, 2008). The aforementioned sequencing technologies are, at present, the most commonly used large-scale sequencing platform. However, as the interest in this field is considerable, new techniques and platforms are currently being developed.

1.4 Causation

Viral metagenomic approaches have the potential to detect nucleic acids from RNA and DNA viruses in samples from both healthy and diseased individuals (Tang & Chiu, 2010; Delwart, 2007). However, it is important to remember that the presence of nucleic acid from a specific virus does not necessarily mean that this virus is the cause of the disease. Several viruses are, for example, considered ubiquitous in the population without a known link to a disease manifestation. As an increasing number of microorganisms was discovered in the 19th century and the beginning of the 20th century, the need emerged for distinguishing pathogenic microorganisms from normal, nonpathogenic flora and to link specific agents to specific diseases. Near the end of the 19th century, Robert Koch presented several criteria, usually referred to as Koch's postulates, which must be fulfilled for a microorganism to be linked to a specific disease. According to those postulates, the microorganism should: i) always be found in the diseased tissue of the sick host; ii) it must be isolated from a diseased host and grown in culture; and iii) the isolated organism must be able to reproduce the disease in experimentally infected healthy hosts. At the time when these criteria were proposed, the concept of viruses and how they affected animals were not

known; studies of microorganisms investigated bacteria and parasites. Because many viruses cannot be grown in culture, they do not fulfill the second postulate. Also, while many viruses do not cause disease in all infected hosts, some may do so when the clinical signs no longer appear. Therefore, if an agent fulfills the postulates, then there is a strong indication that it is responsible of the disease; however, this agent could still be the causative agent even if it does not fulfill these postulates (Fredericks & Relman, 1996).

2 Aims of the Thesis

The main aim of this study was to employ techniques of viral metagenomics in the veterinary field to highlight important areas where these techniques can be used.

- The first study aimed to examine a complex disease with possible viral coinfections. Postweaning multisystemic wasting syndrome (PMWS) was chosen as a target disease. Although porcine circovirus type 2 (PCV-2) is connected with the disease, other viral agents have been suggested to be involved in aiding to trigger the complete clinical picture of PMWS (papers I and II).

- In the second study, we investigated a disease of completely unknown etiology. Shaking mink syndrome (SMS) was chosen because the causative agent of this disease was unknown, despite intensive efforts to elucidate an etiological connection (paper III).

- In the last study, we investigated the usefulness of the technology for studying vectors that have the potential to spread disease. Soft ticks collected from a warthog burrow in Uganda were examined for the presence of viruses (paper IV).

3 Material and Methods

The following sections provide a brief description of the material and techniques used in the various studies. More details can be found in the individual papers. The first section does, however, deal with unpublished data from trials prior to or during the work presented in the various articles included in this thesis.

3.1 Optimization of the technology

Prior to the usage of viral metagenomic approaches on real cases, various steps were optimized using infected cells, tissue spiked with different viruses and clinical material.

3.1.1 Amplification

Virus-infected cells were used to investigate the amplification efficiency of rPCR. We used Borna disease virus (BDV) as an RNA virus model and herpesvirus and PCV-2 as dsDNA and ssDNA virus models, respectively. For all three of the viruses, the random amplification protocol was performed as described by Allander *et al.* (2005). Real-time PCR was then used to investigate the amplification of the viral nucleic acid. For each virus, a sample was taken for a real-time PCR reaction at two different steps: after the extraction of nucleic acid and after the rPCR. Our results clearly demonstrated that the random amplification protocol was an efficient technique to amplify viral nucleic acid (Table 2).

Table 2. Amplification of three viruses: BDV (RNA virus), herpesvirus (dsDNA virus) and circovirus (ssDNA virus) using rPCR. Ct values are shown.

	BDV	Herpesvirus	Circovirus
RNA/DNA	21.4	27	20
PCR	7.1	11	8.5

Using clinical material with a known virus infection, we also observed an increase of the viral nucleic acid after random amplification. SISPA was also tested, but because the amplification was greater using rPCR as well as faster and more straightforward, we did not continue using SISPA for clinical samples.

3.1.2 Ribosomal RNA

Because contamination with host rRNA appeared to be one of the major obstacles, the effect of RNase treatment was tested. First, various concentrations of RNase were tested on brain samples. An in-house 18S real-time PCR assay was performed on the various samples to determine the effect of the treatment. We found that the RNase treatment reduced the 18S contamination but did not completely remove it. A reduction of approximately 5 Ct compared to the sample without RNase was observed, but there were no differences due to the various RNase concentrations tested. The principle is based upon the notion that the viral nucleic acid is protected within its capsid, but we also investigated the effect of RNase on the virus. Therefore, brain samples were spiked with either BDV or bovine viral diarrhea virus (BVDV). Comparing these samples, we found that the two viruses displayed different sensitivities to the treatment. For BDV, a gradient-like decrease of the virus was observed as the concentration of RNase was increased. However, BVDV was unaffected by the RNase treatment, indicating that there is a difference to how sensitive various viruses are.

RiboMinus (Invitrogen) was also tested as a means of removing rRNA. This was performed both on C6 cells infected with BDV and on a liver sample from a cat suffering from feline infectious peritonitis. Via Bioanalyser analysis of RNA and 18S real-time PCR of the two samples, we observed a small reduction of 18S rRNA.

3.2 Papers I and II – Postweaning multisystemic wasting syndrome

In the first study (I), lymph node samples for 454-sequencing analysis were obtain from two Swedish pigs confirmed to have PMWS. In the follow-up study (II), 58 lymph node samples were collected from pigs with and without PMWS.

3.2.1 Sample preparation and 454-sequencing (paper I)

The two lymph nodes were individually homogenized in 1x DNase buffer (Roche) before centrifugation to remove cell debris. The supernatant was filtered through a 0.45-μm syringe filter (Millipore), and the filtrate was split into 200-μl aliquots before a 2h DNase/RNase treatment at 37°C was performed. DNA was extracted and then amplified with random primers in an MDA reaction using Genomiphi v2 DNA amplification kit (GE Healthcare). The purified MDA product from each of the samples was pooled and sequenced at the Royal Institute of Technology, Stockholm, Sweden using the 454 technology.

3.2.2 454-data analysis (paper I)

The 454 sequences were assembled using SeqMan v.8 (Lasergene), and the obtained contigs and singletons were blasted against GenBank using blastn and blastx.

3.2.3 Verification of 454-sequencing results (paper I)

To confirm the hits from the BLAST searches, PCR assays were performed for each of the detected viruses (PCV-2, TTV and a bocavirus). For PCV-2 detection, a previously published real-time PCR protocol was used (Balint *et al.*, 2009). A classical PCR for TTV and the bocavirus were performed using primers based on the sequences obtained by the 454 sequencing. The products were sequenced at Macrogen (Korea).

3.2.4 Prevalence (paper II)

To determine the prevalence of these viruses, lymph nodes from 58 pigs with and without PMWS were investigated. For PCV-2, SYBR Green real-time PCRs were used, which discriminate between the three genogroups identified in Sweden: SG1, SG2 and SG3 (Timmusk *et al.*, 2008). To detect

TTV and discriminate between TTV-1 and 2, a classical PCR was performed using the primers from a previously published study (Segales *et al.*, 2009). The detection of the bocavirus was achieved using primers designed from the 454-data.

3.2.5 Sequence analysis and phylogenetic studies (paper I and II)

A number of the TTV-1 and 2-positive PCR products were sequenced, as were a number of the products from the porcine bocavirus PCR. The sequences were edited using SeqMan v7.2 (Lasergene). TTV-1 and 2 sequences were retrieved from GenBank, as was a region of the VP1/VP2 from a number of parvoviruses. For a number of bocaviruses from different species, the NP-1 protein sequence was also retrieved. A phylogenetic tree of TTV-1 and TTV-2 was constructed using MEGA4 (Tamura *et al.*, 2007). For the parvoviruses, a tree was constructed with parvoviruses from different species based on an element of the VP1/VP2. The sequence identity of the NP-1 gene was compared between different bocaviruses and the detected porcine bocavirus.

3.3   Paper III – Shaking mink syndrome

Brain tissue from three experimentally infected minks was used for random amplification and large-scale sequencing. Brain tissue from three mink kits with a natural occurrence of SMS was used to confirm the findings. In addition, the brain tissue from six healthy minks was included.

3.3.1 Sample preparation and 454-sequencing

Brain tissue samples from the experimentally infected mink kits were mechanically homogenized in 1x DNase buffer (Roche). The homogenate was centrifuged at 4000 rpm for 10 min before the supernatant was collected and filtered through a 0.45-μm syringe filter (Millipore). In aliquots of 200 μl, the filtrate was treated with DNase and RNase for 2 h (37 °C) to reduce the nucleic acid contamination from the host. Both RNA and DNA were extracted from the samples. To efficiently and randomly amplify the viral nucleic acid, it was labeled at both ends. For the RNA, this labeling was performed during the cDNA synthesis step using a random primer consisting of random nucleotides at the 3' end and a tag sequence at the 5' end (Allander *et al.*, 2005). The cDNA synthesis was performed with

the previously mentioned primer, together with Superscript III (Invitrogen) in the first strand reaction. For the second strand synthesis, the Klenow fragment (3' − 5' exo⁻) (NEB) was used. The same primer was used to label the extracted DNA in a two-step Klenow fragment (3' − 5' exo⁻) (NEB) reaction. An rPCR, using AmpliTaq Gold (Applied Biosystems), was performed on the labeled DNA and cDNA with a primer identical to the tag sequence. The amplified products were purified and then sent for 454 sequencing at the Royal Institute of Technology in Stockholm, Sweden.

3.3.2 454-data analysis

The 454-sequencing data was assembled using SeqMan v7.2 (Lasergene). Blastn searches using PLAN webserver (He *et al.*, 2007) were performed on the created contigs and on the singleton sequences. An additional BLAST search were performed using blastx on the sequences with an e-value greater than 10^{-3}.

3.3.3 Verification of 454-sequencing results

Primers were designed based on the obtained astrovirus sequences. A classical PCR was performed with AmpliTaq Gold DNA polymerase (Applied Biosystems) using these primers. This PCR used cDNA from the three experimentally infected mink kits, from the brain tissue of the three SMS minks and from the six healthy mink kits.

3.3.4 Full-length sequencing

Because the eight astrovirus sequences obtained from the 454-sequencing were dispersed over the genome, primers were designed to fill in the gaps between them. Based on the genome of a mink astrovirus associated with preweaning mink diarrhea (MiAstV) available in GenBank, a primer was designed at the extreme 5' end. FirstChoice RLM-RACE kit (Invitrogen) was used to determine the 3' end of the genome. The various PCR products were purified and then sequenced at Macrogen (Korea), either directly or after a cloning step.

3.3.5 Sequence analysis and phylogenetic studies

The sequences were edited and then assembled into the complete genome sequence. Astrovirus genome sequences from other species, as well as other

mink astrovirus sequences, were collected from GenBank. Phylogenetic trees were constructed using the MEGA4 software (Tamura *et al.*, 2007). One tree contained the full-length sequence of the astrovirus detected in this study (SMS-AstV), together with the full-length sequences of astroviruses from various species. The other tree was based on a partial region of ORF1b from SMS-AstV and the corresponding region of mink astrovirus isolates from Denmark and Sweden. Using Bioedit, a sequence identity matrix comparing SMS-AstV to MiAstV was performed on both the nucleotide and the protein sequence of the different ORFs.

3.4 Paper IV – Arthropod vector

In this study, 24 soft ticks collected from a warthog burrow in Lake Mburo National park, Uganda were investigated for the presence of viruses.

3.4.1 Sample preparation and 454 sequencing

The ticks were homogenized in 1x DNase buffer (Roche), either individually (n = 1) or in a pool (n = 20). After centrifugation of the homogenate, the supernatant was filtered through a 0.45-μm syringe filter (Millipore). DNase and RNase treatment were performed at 37°C for 2h before extraction of nucleic acid. The nucleic acid was labeled at both ends using a tag sequence. For the RNA, double stranded cDNA synthesis using a random primer with the tag sequence at the 5' end (Allander *et al.*, 2005) was performed. The first-strand synthesis was performed with Superscript III (Invitrogen), and the second strand was synthesized using Klenow fragment (3' – 5' exo$^-$) (NEB). The DNA was also labeled using the same primer and Klenow fragment (3' – 5' exo$^-$) (NEB). A PCR assay was performed with a primer identical to the tag sequence using AmpliTaq Gold DNA polymerase (Applied Biosystems). The product was purified and sent for 454 sequencing at Inqaba Biotech (South Africa).

3.4.2 454-data analysis

CLC Genomic workbench v.7 was used to assemble the sequences. The contigs and singleton sequences were blasted through the PLAN web server (He *et al.*, 2007). Sequences giving no hit or an e-value greater than 10^{-3} using blastn were processed further with blastx searches.

3.4.3 Verification of 454-sequencing results

PCR was performed on nucleic acid isolated from the ticks to detect each of the three viruses (PCV-2, TTV and a hepatitis E-related virus) discovered using large-scale sequencing. For TTV, primers for both genotypes 1 and 2 were used (Segales *et al.*, 2009), and real-time PCR was performed for PCV-2 (Balint *et al.*, 2009). The primers for the HEV-related virus were designed based on the obtained 454 sequences.

3.4.4 Sequencing and phylogenetic studies of the HEV-related virus

The PCR product from the amplification of the HEV-related virus was purified and sequenced using BigDye Terminator v3.1 chemistry (Applied Biosystems). The sequence was edited using SeqMan v8 (Lasergene). The sequence was translated using transeq (EMBOSS) and the corresponding sequences of the different genogroups were collected from GenBank. The sequences were aligned and a phylogenetic tree was constructed using MEGA4 (Tamura *et al.*, 2007). Using the same sequences and region, a sequence identity matrix was performed in Bioedit.

4 Results and Discussion

4.1 Papers I and II – Postweaning multisystemic wasting syndrome

Although PCV-2 is considered the causative agent of PMWS, infection by this virus is not sufficient to develop the disease. In fact, PCV-2 is considered a ubiquitous virus in the pig population around the world (Segales *et al.*, 2005). Therefore, various factors have been suggested to play a role in the development of the disease. One of these factors is coinfection with other viruses. Through experimental studies, complete PMWS development was reconstituted by coinfection with PCV-2 and porcine parvovirus and porcine reproductive and respiratory syndrome virus (PRRSV) (Allan *et al.*, 2004). Therefore, we decided to use samples from pigs with PMWS to study any possible viral coinfections using viral metagenomics.

In this case, MDA technology was used to amplify the nucleic acid prior to 454 sequencing. This efficiently amplified the DNA, yielding a product of high molecular weight (> 20 kb).

Approximately 9000 reads were obtained after the large-scale sequencing, and from these, 99.5% were determined via BLAST searches to belong to PCV-2. In fact, a minimum number of the sequences were from hosts, indicating both the enormous amount of PCV-2 in these samples and the success of the DNase treatment. Apart from PCV-2, a few TTV sequences were also obtained. Although the presence of TTV had, at that time, not been investigated in pigs in Sweden, it was shown in studies in Europe and other parts of the world to be highly prevalent in both domestic and wild pig populations (Okamoto, 2009; McKeown *et al.*, 2004). A third virus was also present in the investigated lymph nodes, exhibiting close similarity to

bocavirus. At that time, no porcine bocavirus was known, and therefore, this discovery was further investigated through additional sequencing. By designing primers based on the reads from the 454 sequencing, a 2000-nt-long region was obtained. Due to the presence of the NP1 gene and phylogenetic studies, we classified this newly discovered virus as a porcine bocavirus in the parvovirus family (Fig. 6).
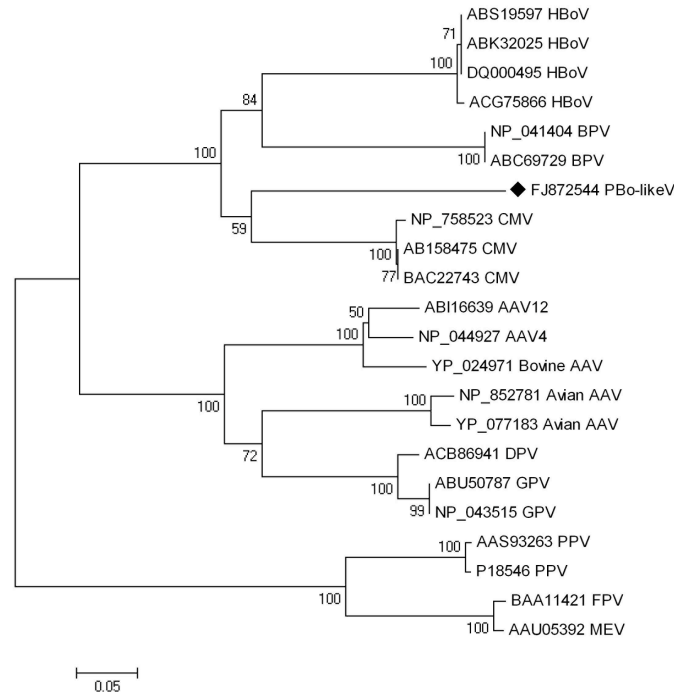

Figure 6. Neighbor-joining tree with a bootstrap value of 2000, exhibiting the phylogenetic relationship between the porcine bocavirus discovered in this study (PBo-likeV), shown with a ♦, and parvoviruses from different species.

To further study the prevalence of both TTV and the novel porcine bocavirus in Sweden, DNA from lymph nodes from 58 pigs with and without PMWS were analyzed by PCR. TTV, which is comprised of two genogroups (TTV-1 and TTV-2), was found at a high prevalence in the investigated pigs, regardless of whether they had PMWS. Although the porcine bocavirus could be found in both pigs with and without PMWS, the positive percentage was higher in the pigs suffering from PMWS (88% compared to 46%) (Fig. 7a). Studying the coinfection showed that in 71% of the PMWS-positive pigs, apart from PCV-2, all three viruses (TTV-1, TTV-2 and porcine bocavirus) were present, compared to only 33% in the

non–PMWS pigs (Fig. 7b). The largest group of pigs without PMWS was that showing a coinfection of TTV-1 and 2 together with PCV-2.
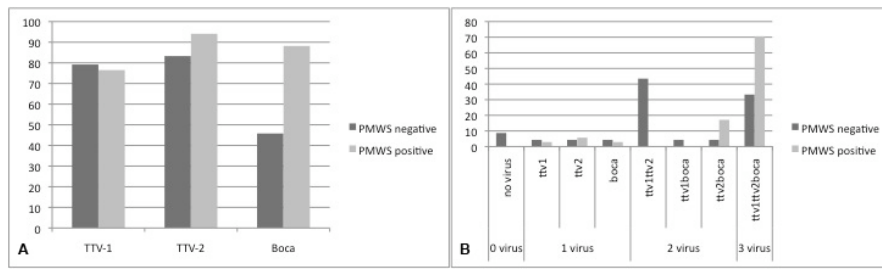

Figure 7. A: Percentage of each of the three viruses in pigs with and without PMWS. B: Percentage of the different possible virus combinations in positive and negative PMWS pigs.

Together, the results of this study indicated that the chosen viral metagenomic approach worked well, and apart from the expected discovery of PCV-2 and TTV, we also detected a novel porcine bocavirus. The consequences of this coinfection is not known, and further studies must be performed to determine if these viruses interact or are affected by each other's presence.

Unlike the other viral metagenomic studies presented in this thesis, random MDA was used instead of rPCR. The choice of amplification method may have influenced the large-scale sequencing outcome, regarding the dominant presence of PCV-2 compared with host nucleic acid. Also, the presence of RNA viruses was not investigated in this study. However, as previously discussed, it is possible to use similar techniques to amplify nucleic acid from RNA viruses. MDA proved to amplify the DNA efficiently, but it would be of interest to perform tests on various DNA viruses (especially short linear viruses) to determine how well their genome can be amplified using this approach or if modifications to the protocol must be made to not discriminate against these viruses.

4.2 Paper III - Shaking mink syndrome case

Shaking mink syndrome (SMS) is a disease that has gone undiagnosed for a decade. It first appeared in Denmark in 2000 and then spread to other Nordic countries. The affected minks display neurological symptoms, such as shaking and a staggering gait. Despite efforts to discover the causative agent using detection assays for pathogens known to cause neurological disorders (and more general approaches such as virus isolation), none was

found. Experimental infection trials have also been conducted, demonstrating that the disease can be transmitted by injection of brain homogenate from a sick individual to a healthy one (Gavier-Widen *et al.*, 2004). Because all of these studies were performed with a negative result, SMS was considered an interesting case for a viral metagenomic study.

The amplification approach used in this study was rPCR, which worked well for both the RNA and DNA, producing a smear product. The large-scale sequencing yielded approximately 49,000 reads that, after assembly, consisted of 1403 contigs and 10,756 singleton sequences. After analysis of the data set, it was clear that the majority of the sequences were host sequences. However, five sequences were of viral origin, all of them showing high similarity to a mink astrovirus (MiAstV) associated with preweaning diarrhea. Through PCR, the astrovirus was confirmed to be present in the brain of both the experimentally infected minks and in minks with a natural occurrence of SMS. Because the obtained reads were evenly distributed over the genome, the complete genome of the viral SMS astrovirus (SMS–AstV) could be sequenced from one of the SMS minks using primers based on the 454 reads. The ends were elucidated with 3' RACE and with a primer based on the 5' end of the MiAstV sequence from GenBank. The complete genome of SMS–AstV was found to contain 6,614 nt and was 80% identical to the MiAstV genome. Phylogenetic studies (Fig. 8) indicated that SMS–AstV grouped with MiAstV, rather than with a human astrovirus associated with encephalitis (HAstV-PS) (Quan *et al.*, 2010).
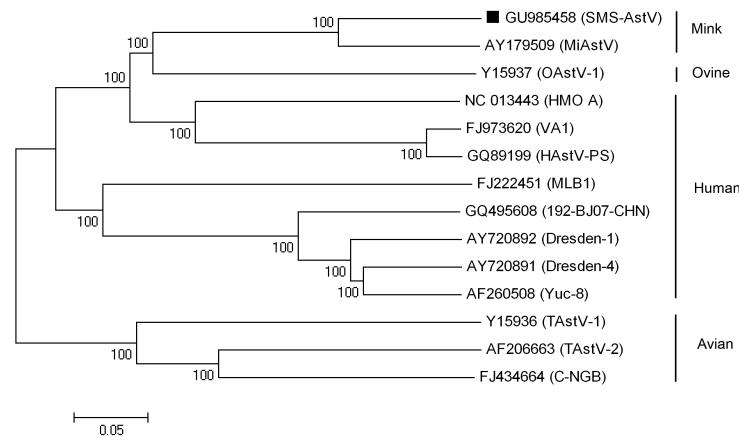

Figure 8. Neighbor-joining tree exhibiting the phylogenetic relationship of SMS–AstV (■) and astroviruses from various species.

40

It was a clear difference in how similar the different ORFs were between SMS-AstV and MiAstV. ORF1a and b, which encode the non-structural protein, were more similar between the two viruses than was ORF2, which codes for the capsid protein. While ORF1a and b of SMS-AstV showed a protein identity of greater than 90% to MiAstV, the same identity for ORF2 was only 59%.

Thus, in this study, viral metagenomics was used to discover a possible causative agent of SMS in minks. However, further studies are required to confirm this notion. This finding also has the potential to help solve other neurological disorders. Astroviruses have previously only been connected to enteric disorders such as diarrhea, but as this paper was under review, another article was published that described an astrovirus found in a boy with encephalitis (Quan *et al.*, 2010). These two findings demonstrate that there may be astroviruses that have roles in completely different disorders from what have traditionally been thought.

4.3   Paper IV - Arthropod vector

Many of the recent EIDs have occurred due to transmission of a virus that has a wildlife reservoir but that, for some reason, has come into contact with and infected humans and/or domestic animals, causing disease (Daszak *et al.*, 2000). Therefore, it is important to have knowledge of the viruses circulating in nature. Arthropods, such as mosquitoes and ticks, are known to be important vectors for spreading viruses and other pathogens from wildlife to domestic animals and humans (Weaver & Reisen, 2009). In this study, we investigated the presence of virus in soft ticks (*Ornithodoros*) collected from a warthog burrow in Uganda. This tick species is known to be the vector for African swine fever virus (Kleiboeker & Scoles, 2001), but it may also carry other viruses and is therefore worth testing.

Twenty-four ticks were used in this study; four were handled individually, while the rest were pooled. The rPCRs worked well, giving the expected smear, and after the 454 sequencing of the products, approximately 143,000 reads were obtained. After assembly, 1329 contigs and 52,682 singletons were obtained. From this data set, three different viruses were found. A few of the sequences were PCV-2 and TTV, but these findings could not be confirmed by PCR, possibly due to mismatches in the primer sequence or low amounts of virus. Ten sequences showing low similarity at the protein level to hepatitis E virus (HEV) were also discovered. This finding was followed by sequencing a longer region (620 nt) between the two reads

obtained by the 454 sequencing that were farthest away from each other. In the phylogenetic study, the discovered virus (HEV-related virus) did not group with HEV (Fig. 9).
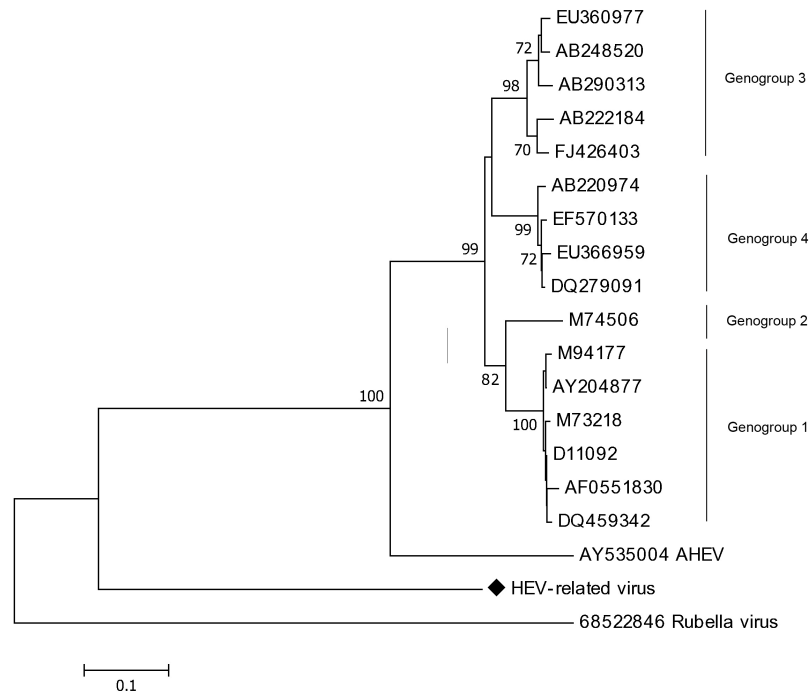

Figure 9. Phylogenetic study of the HEV-related tick virus (◆) and the different genogroups of HEV, as well as AHEV and rubella virus.

Sequence analysis suggested that there is a considerable difference between the HEV-related virus and all known genogroups of HEV, as there is to the more divergent avian HEV. Indeed, only approximately 35% identity to these other viruses was found at the protein level. The sequence analyses suggested that this may not be a novel genogroup of HEV but, rather, a novel genus within *Hepaviridae* or even a novel virus family. Thus, so far, only a small portion of the viral genome has been sequenced, and further studies are required to ultimately classify this virus.

In, conclusion, using random amplification and large-scale sequencing, soft ticks was investigated and a novel RNA virus was found. This result indicates that there are many viruses present in nature, which we are not aware of and do not know the role they have or if they have the potential to spread to other species and cause EIDs.

42

5 Concluding Remarks

Random amplification and large-scale sequencing techniques was successfully utilized to investigate several areas where viral metagenomic studies could be of interest for veterinary science. Through these studies, several novel viruses were also discovered.

- By studying PMWS, it was shown that in the case of an excess amount of one virus, in this case PCV-2, the metagenomic techniques were able to detect even additional viruses in the background. When studying a disease, it is important to remember that even though one agent may be considered the causative agent, the effect of other viral infections can influence the outcome of the disease. In the case of PMWS, coinfection by PCV-2, TTV-1 and 2, and a novel porcine bocavirus was discovered. The effect of this coinfection is not known, but one can state that a high percentage of the diseased animals harbored a coinfection of all of these viruses.

- In the SMS study, the opportunity to work with a disease of unknown etiology, for which classical methods had failed to uncover the causative agent was possible. Using viral metagenomics, an astrovirus was identified as a possible agent behind this disease. Although further studies must be performed to confirm or reject this hypothesis, this study shows the strength of the technique for detecting novel or unexpected viruses in connection with new diseases.

- By studying vectors and the pathogens that they carry, we will be better prepared for future EIDs. In this study of soft ticks in Uganda, it was shown that it is possible to discover novel viruses in this type of vector

organism. The impact of the newly detected viruses in vectors is currently being further investigated.

6 Implications for Future Research

Through viral metagenomics, genetic populations can be studied without the need for cultivation and prior sequence knowledge. This approach can provide new information, both on the natural viral flora of a host and the host's environment, as well as on known disease complexes and new EIDs. Random PCR and large-scale sequencing have been shown, both in the studies presented in this thesis and through other studies, to be powerful tools to investigate these issues. The work presented in this thesis opens new avenues in the veterinary field and contributes to a better understanding of the complex etiology of infectious diseases in animal hosts. Disease outbreaks in animals have the potential to cause devastating economic effects at both the country level and for animal owners. Many infectious agents also have zoonotic capacities, in which the infections initiate in animals and are then transmitted to humans. According to the Center for Disease Control and Prevention, approximately 75% of recently emerging infectious diseases affecting humans are originating from animals. Accordingly, it is evident that future research will put more focus on the wide-range detection of pathogens in animals and humans hosts, in order to investigate known and emerging new diseases, to better understand the infection biology of complex maladies and to support the development of effective controls measures. Metagenomic studies also provide the opportunity to study virus/host interactions in a non-diseased host to better understand viral ecology.

References:

Akowitz, A. & Manuelidis, L. (1989). A novel cDNA/PCR strategy for efficient cloning of small amounts of undefined RNA. *Gene* 81(2), 295-306.

Allan, G.M., McNeilly, F., Ellis, J., Krakowka, S., Botner, A., McCullough, K., Nauwynck, H., Kennedy, S., Meehan, B. & Charreyre, C. (2004). PMWS: experimental model and co-infections. *Vet Microbiol* 98(2), 165-8.

Allander, T., Emerson, S.U., Engle, R.E., Purcell, R.H. & Bukh, J. (2001). A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci U S A* 98(20), 11609-14.

Allander, T., Tammi, M.T., Eriksson, M., Bjerkner, A., Tiveljung-Lindell, A. & Andersson, B. (2005). Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proc Natl Acad Sci U S A* 102(36), 12891-6.

Ambrose, H.E. & Clewley, J.P. (2006). Virus discovery by sequence-independent genome amplification. *Rev Med Virol* 16(6), 365-83.

Anderson, S. (1981). Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res* 9(13), 3015-27.

Balint, A., Tenk, M., Deim, Z., Rasmussen, T.B., Uttenthal, A., Csagola, A., Tuboly, T., Farsang, A., Fossum, C., Timmusk, S., Berg, M. & Belak, S. (2009). Development of Primer-Probe Energy Transfer real-time PCR for the detection and quantification of porcine circovirus type 2. *Acta Vet Hung* 57(3), 441-52.

Beijerinck, M.W. (1898). *Concerning a contagium vivum fluidum as cause of the spot disease of tobacco leaves*. American Phytopathological Society. (Phytopatological Classics

Berger, A., Drosten, C., Doerr, H.W., Sturmer, M. & Preiser, W. (2004). Severe acute respiratory syndrome (SARS)--paradigm of an emerging viral infection. *J Clin Virol* 29(1), 13-22.

Berthet, N., Reinhardt, A.K., Leclercq, I., van Ooyen, S., Batejat, C., Dickinson, P., Stamboliyska, R., Old, I.G., Kong, K.A., Dacheux, L., Bourhy, H., Kennedy, G.C., Korfhage, C., Cole, S.T. & Manuguerra, J.C. (2008). Phi29 polymerase based random amplification of viral RNA as an alternative to random RT-PCR. *BMC Mol Biol* 9, 77.

Birkenmeyer, L.G., Leary, T.P., Muerhoff, A.S., Dawson, G.J., Mushahwar, I.K. & Desai, S.M. (2003). Selectively primed adaptive driver RDA (SPAD-RDA): an improved method for subtractive hybridization. *J Med Virol* 71(1), 150-9.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P. & Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* 185(20), 6220-3.

Breitbart, M. & Rohwer, F. (2005). Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *Biotechniques* 39(5), 729-36.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F. & Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A* 99(22), 14250-5.

Briese, T., Palacios, G., Kokoris, M., Jabado, O., Liu, Z., Renwick, N., Kapoor, V., Casas, I., Pozo, F., Limberger, R., Perez-Brena, P., Ju, J. & Lipkin, W.I. (2005). Diagnostic system for rapid and sensitive differential detection of pathogens. *Emerg Infect Dis* 11(2), 310-3.

Chang, Y., Cesarman, E., Pessin, M.S., Lee, F., Culpepper, J., Knowles, D.M. & Moore, P.S. (1994). Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* 266(5192), 1865-9.

Chua, K.B., Bellini, W.J., Rota, P.A., Harcourt, B.H., Tamin, A., Lam, S.K., Ksiazek, T.G., Rollin, P.E., Zaki, S.R., Shieh, W., Goldsmith, C.S., Gubler, D.J., Roehrig, J.T., Eaton, B., Gould, A.R., Olson, J., Field, H., Daniels, P., Ling, A.E., Peters, C.J., Anderson, L.J. & Mahy, B.W. (2000). Nipah virus: a recently emergent deadly paramyxovirus. *Science* 288(5470), 1432-5.

Culley, A.I., Lang, A.S. & Suttle, C.A. (2006). Metagenomic analysis of coastal RNA virus communities. *Science* 312(5781), 1795-8.

Daniels, P.W., Halpin, K., Hyatt, A. & Middleton, D. (2007). Infection and disease in reservoir and spillover hosts: determinants of pathogen emergence. *Curr Top Microbiol Immunol* 315, 113-31.

Daszak, P., Cunningham, A.A. & Hyatt, A.D. (2000). Emerging infectious diseases of wildlife--threats to biodiversity and human health. *Science* 287(5452), 443-9.

Dean, F.B., Nelson, J.R., Giesler, T.L. & Lasken, R.S. (2001). Rapid amplification of plasmid and phage DNA using Phi 29 DNA

polymerase and multiply-primed rolling circle amplification. *Genome Res* 11(6), 1095-9.

Deininger, P.L. (1983). Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal Biochem* 129(1), 216-23.

Delwart, E.L. (2007). Viral metagenomics. *Rev Med Virol* 17(2), 115-31.

Dinter, Z. (1989). *Diagnostic virology*.

Elnifro, E.M., Ashshi, A.M., Cooper, R.J. & Klapper, P.E. (2000). Multiplex PCR: optimization and application in diagnostic virology. *Clin Microbiol Rev* 13(4), 559-70.

Endoh, D., Mizutani, T., Kirisawa, R., Maki, Y., Saito, H., Kon, Y., Morikawa, S. & Hayashi, M. (2005). Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription. *Nucleic Acids Res* 33(6), e65.

Feldmann, H., Czub, M., Jones, S., Dick, D., Garbutt, M., Grolla, A. & Artsob, H. (2002). Emerging and re-emerging infectious diseases. *Med Microbiol Immunol* 191(2), 63-74.

Flint, S.J., Enquist, L.W., Racaniello, V.R. & Skalka, A.M. (2009). *Principles of virology*. 3rd edition. American Society for Microbiology.1).

Fredericks, D.N. & Relman, D.A. (1996). Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev* 9(1), 18-33.

Froussard, P. (1992). A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. *Nucleic Acids Res* 20(11), 2900.

Gall, A., Hoffmann, B., Harder, T., Grund, C., Ehricht, R. & Beer, M. (2009). Rapid haemagglutinin subtyping and pathotyping of avian influenza viruses by a DNA microarray. *J Virol Methods* 160(1-2), 200-5.

Ganova-Raeva, L., Zhang, X., Cao, F., Fields, H. & Khudyakov, Y. (2006). Primer Extension Enrichment Reaction (PEER): a new subtraction method for identification of genetic differences between biological specimens. *Nucleic Acids Res* 34(11), e76.

Gavier-Widen, D., Brojer, C., Dietz, H.H., Englund, L., Hammer, A.S., Hedlund, K.O., Hard af Segerstad, C., Nilsson, K., Nowotny, N., Puurula, V., Thoren, P., Uhlhorn, H., Weissenbock, H., Agren, E. & Klingeborn, B. (2004). Investigations into shaking mink syndrome: an encephalomyelitis of unknown cause in farmed mink (Mustela vison) kits in Scandinavia. *J Vet Diagn Invest* 16(4), 305-12.

Goldsmith, C.S. & Miller, S.E. (2009). Modern uses of electron microscopy for detection of viruses. *Clin Microbiol Rev* 22(4), 552-63.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68(4), 669-85.

Hazelton, P.R. & Gelderblom, H.R. (2003). Electron microscopy for rapid diagnosis of infectious agents in emergent situations. *Emerg Infect Dis* 9(3), 294-303.

He, J., Dai, X. & Zhao, X. (2007). PLAN: a web platform for automating high-throughput BLAST searches and for managing and mining results. *BMC Bioinformatics* 8, 53.

Honkavuori, K.S., Shivaprasad, H.L., Williams, B.L., Quan, P.L., Hornig, M., Street, C., Palacios, G., Hutchison, S.K., Franca, M., Egholm, M., Briese, T. & Lipkin, W.I. (2008). Novel borna virus in psittacine birds with proventricular dilatation disease. *Emerg Infect Dis* 14(12), 1883-6.

Huson, D.H., Auch, A.F., Qi, J. & Schuster, S.C. (2007). MEGAN analysis of metagenomic data. *Genome Res* 17(3), 377-86.

Huson, D.H., Richter, D.C., Mitra, S., Auch, A.F. & Schuster, S.C. (2009). Methods for comparative metagenomics. *BMC Bioinformatics* 10 Suppl 1, S12.

Johne, R., Muller, H., Rector, A., van Ranst, M. & Stevens, H. (2009). Rolling-circle amplification of viral DNA genomes using phi29 polymerase. *Trends Microbiol* 17(5), 205-11.

Kapikian, A.Z., Wyatt, R.G., Dolin, R., Thornhill, T.S., Kalica, A.R. & Chanock, R.M. (1972). Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. *J Virol* 10(5), 1075-81.

Kircher, M. & Kelso, J. (2010). High-throughput DNA sequencing-- concepts and limitations. *Bioessays* 32(6), 524-36.

Kistler, A.L., Gancz, A., Clubb, S., Skewes-Cox, P., Fischer, K., Sorber, K., Chiu, C.Y., Lublin, A., Mechani, S., Farnoushi, Y., Greninger, A., Wen, C.C., Karlene, S.B., Ganem, D. & DeRisi, J.L. (2008). Recovery of divergent avian bornaviruses from cases of proventricular dilatation disease: identification of a candidate etiologic agent. *Virol J* 5, 88.

Kleiboeker, S.B. & Scoles, G.A. (2001). Pathogenesis of African swine fever virus in Ornithodoros ticks. *Anim Health Res Rev* 2(2), 121-8.

Klein, S.L. & Calisher, C.H. (2007). Emergence and persistence of hantaviruses. *Curr Top Microbiol Immunol* 315, 217-52.

Lassauniere, R., Kresfelder, T. & Venter, M. (2010). A novel multiplex real-time RT-PCR assay with FRET hybridization probes for the detection and quantitation of 13 respiratory viruses. *J Virol Methods* 165(2), 254-60.

Leland, D.S. & Ginocchio, C.C. (2007). Role of cell culture for virus detection in the age of technology. *Clin Microbiol Rev* 20(1), 49-78.

Li, L., Victoria, J.G., Wang, C., Jones, M., Fellers, G.M., Kunz, T.H. & Delwart, E. (2010). Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. *J Virol* 84(14), 6955-65.

Lisitsyn, N. & Wigler, M. (1993). Cloning the differences between two complex genomes. *Science* 259(5097), 946–51.

Mackay, I.M., Arden, K.E. & Nitsche, A. (2002). Real-time PCR in virology. *Nucleic Acids Res* 30(6), 1292–305.

Mahy, B.W. (2005). Introduction and history of foot-and-mouth disease virus. *Curr Top Microbiol Immunol* 288, 1–8.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F. & Rothberg, J.M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057), 376-80.

Matsui, S.M., Kim, J.P., Greenberg, H.B., Su, W., Sun, Q., Johnson, P.C., DuPont, H.L., Oshiro, L.S. & Reyes, G.R. (1991). The isolation and characterization of a Norwalk virus-specific cDNA. *J Clin Invest* 87(4), 1456–61.

McKeown, N.E., Fenaux, M., Halbur, P.G. & Meng, X.J. (2004). Molecular characterization of porcine TT virus, an orphan virus, in pigs from six different countries. *Vet Microbiol* 104(1-2), 113-7.

Metzgar, D., Myers, C.A., Russell, K.L., Faix, D., Blair, P.J., Brown, J., Vo, S., Swayne, D.E., Thomas, C., Stenger, D.A., Lin, B., Malanoski, A.P., Wang, Z., Blaney, K.M., Long, N.C., Schnur, J.M., Saad, M.D., Borsuk, L.A., Lichanska, A.M., Lorence, M.C., Weslowski, B., Schafer, K.O. & Tibbetts, C. (2010). Single assay for simultaneous detection and differential identification of human and avian influenza virus types, subtypes, and emergent variants. *PLoS One* 5(2), e8995.

Mihindukulasuriya, K.A., Wu, G., St Leger, J., Nordhausen, R.W. & Wang, D. (2008). Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J Virol* 82(10), 5084-8.

Miller, M.B. & Tang, Y.W. (2009). Basic concepts of microarrays and potential applications in clinical microbiology. *Clin Microbiol Rev* 22(4), 611–33.

Morse, S.S. (1995). Factors in the emergence of infectious diseases. *Emerg Infect Dis* 1(1), 7–15.

Muerhoff, A.S., Leary, T.P., Desai, S.M. & Mushahwar, I.K. (1997). Amplification and subtraction methods and their application to the discovery of novel human viruses. *J Med Virol* 53(1), 96-103.

Nishizawa, T., Okamoto, H., Konishi, K., Yoshizawa, H., Miyakawa, Y. & Mayumi, M. (1997). A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochem Biophys Res Commun* 241(1), 92-7.

Nowrousian, M. (2010). Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. *Eukaryot Cell* 9(9), 1300-10.

Okamoto, H. (2009). TT viruses in animals. *Curr Top Microbiol Immunol* 331, 35-52.

Palacios, G., Druce, J., Du, L., Tran, T., Birch, C., Briese, T., Conlan, S., Quan, P.L., Hui, J., Marshall, J., Simons, J.F., Egholm, M., Paddock, C.D., Shieh, W.J., Goldsmith, C.S., Zaki, S.R., Catton, M. & Lipkin, W.I. (2008). A new arenavirus in a cluster of fatal transplant-associated diseases. *N Engl J Med* 358(10), 991-8.

Palacios, G., Quan, P.L., Jabado, O.J., Conlan, S., Hirschberg, D.L., Liu, Y., Zhai, J., Renwick, N., Hui, J., Hegyi, H., Grolla, A., Strong, J.E., Towner, J.S., Geisbert, T.W., Jahrling, P.B., Buchen-Osmond, C., Ellerbrok, H., Sanchez-Seco, M.P., Lussier, Y., Formenty, P., Nichol, M.S., Feldmann, H., Briese, T. & Lipkin, W.I. (2007). Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerg Infect Dis* 13(1), 73-81.

Quan, P.L., Wagner, T.A., Briese, T., Torgerson, T.R., Hornig, M., Tashmukhamedova, A., Firth, C., Palacios, G., Baisre-De-Leon, A., Paddock, C.D., Hutchison, S.K., Egholm, M., Zaki, S.R., Goldman, J.E., Ochs, H.D. & Lipkin, W.I. (2010). Astrovirus encephalitis in boy with X-linked agammaglobulinemia. *Emerg Infect Dis* 16(6), 918-25.

Quinn, P.L., Makey, B.K., Carter, M.E., Donnelly, W.J. & Leonard, F.C. (2002). *Veterinary Microbiology and Microbial Disease*.

Rector, A., Tachezy, R. & Van Ranst, M. (2004). A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. *J Virol* 78(10), 4993-8.

Reyes, G.R. & Kim, J.P. (1991). Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Mol Cell Probes* 5(6), 473-81.

Segales, J., Allan, G.M. & Domingo, M. (2005). Porcine circovirus diseases. *Anim Health Res Rev* 6(2), 119-42.

Segales, J., Martinez-Guino, L., Cortey, M., Navarro, N., Huerta, E., Sibila, M., Pujols, J. & Kekarainen, T. (2009). Retrospective study on swine Torque teno virus genogroups 1 and 2 infection from 1985 to 2005 in Spain. *Vet Microbiol* 134(3-4), 199-207.

Shendure, J. & Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotechnol* 26(10), 1135-45.

Simons, J.N., Pilot-Matias, T.J., Leary, T.P., Dawson, G.J., Desai, S.M., Schlauder, G.G., Muerhoff, A.S., Erker, J.C., Buijk, S.L., Chalmers, M.L. & et al. (1995). Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proc Natl Acad Sci U S A* 92(8), 3401-5.

Steinhardt, E., Israeli, C. & Lambert, R.A. (1913). Studies on the Cultivation of the Virus of Vaccinia. *The Journal of Infectious Diseases* 13(2), 294-300.

Suttle, C.A. (2005). Viruses in the sea. *Nature* 437(7057), 356-61.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24(8), 1596-9.

Tang, P. & Chiu, C. (2010). Metagenomics for the discovery of novel human viruses. *Future Microbiol* 5(2), 177-89.

Thurber, R.V., Haynes, M., Breitbart, M., Wegley, L. & Rohwer, F. (2009). Laboratory procedures to generate viral metagenomes. *Nat Protoc* 4(4), 470-83.

Timmusk, S., Wallgren, P., Brunborg, I.M., Wikstrom, F.H., Allan, G., Meehan, B., McMenamy, M., McNeilly, F., Fuxler, L., Belak, K., Podersoo, D., Saar, T., Berg, M. & Fossum, C. (2008). Phylogenetic analysis of porcine circovirus type 2 (PCV2) pre- and post-epizootic postweaning multisystemic wasting syndrome (PMWS). *Virus Genes* 36(3), 509-20.

Van Etten, J.L., Lane, L.C. & Dunigan, D.D. (2010). DNA viruses: the really big ones (giruses). *Annu Rev Microbiol* 64, 83-99.

Wang, D., Coscoy, L., Zylberberg, M., Avila, P.C., Boushey, H.A., Ganem, D. & DeRisi, J.L. (2002). Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci U S A* 99(24), 15687-92.

Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R., Hickenbotham, M., Magrini, V., Eldred, J., Latreille, J.P., Wilson, R.K., Ganem, D. & DeRisi, J.L. (2003). Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* 1(2), E2.

Weaver, S.C. & Reisen, W.K. (2009). Present and future arboviral threats. *Antiviral Res* 85(2), 328-45.

Wild, P. (2008). Electron microscopy of viruses and virus-cell interactions. *Methods Cell Biol* 88, 497-524.

Acknowledgements

The work was performed at the Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences and was financed by the Swedish International Development Cooperation Agency

I would like to express my gratitude for the support and contribution to my work:

Mikael Berg, my main supervisor, for your continuous support, encouragement and guidance. For having an open door and to whom I can always vent my thoughts.

Sandor Belák, co-supervisor, for providing me the opportunity to start and carry out my phD in virology. For your support and helping me become a better scientific writer.

Frederik Widén, co-supervisor, for you positive support and knowledge in the veterinary field.

Mikhayil Hakhverdyan, co-supervisor, for always being ready to help – may it be a real-time analysis or a broken machine. And for you nice company in Pirbright during the LAMP study.

All the people that have been at the SLU virology and SVA Vir/FOU during my 5 years in the group: *Adam, Akos, Alia, Anna, Anna-Malin, Attila, Babu, Balaje, Claudia, Femi, Fredrik G, Giorgi, Hongyan, Istvan, Jay, Jonas, Kalle, Karin U, Karin W, Lihong, Linda, Mehdi, Mikael L, Munir, Neil, Peter, Sandra, Shahid, Shaman, Siamak and Sophie* (hope I did not miss any of you,

in that case I blame the late hour….). Thanks for all of your company both in the lab as well as during barbeques, "knytkalas", "fikas" etc.

With special thanks to:

Peter Gyarmati, my very good friend, who without my first three years would have been much duller. Your company during fishing, making barbeque pizza, O´Conners etc etc was invaluable. And for always being there when needed and listening.

Anna-Malin Linde, for your friendship and sharing your thoughts both on science and life.

Giorgi Metreveli, for being a very good friend and being such a good company on all our spontaneous AWs.

Jonas Wensman, whom I started my "scientific career" together with 6 years ago at BMC. For you friendship and sharing the phD struggles.

Shaman Muradrasoli, although only a short stay at our virology group but a highly appreciated member, for our chats and nice coffee breaks.

Siamak Zohari, our lab course guru. For your great company and friendship (and for providing me with energy boosters in the form of "snusklubbor" and coffee powder ☺)

Muhammed Munir my former office mate. For sharing your work and thoughts on life.

Mehdi Bidokhti, for arranging all those nice social events such as bowling, movie nights, canoeing etc.

Lihong Liu and *Hongyan Xia* for your friendship and positive spirits.

Anna Rothman for always helping with all kinds of practical matters.

Berka, for always being ready to help but mostly for being you and our nice chats throughout the years.

To the immunology group, *Caroline, Bettan, Lotta, Tommy, Parvin, Tanja, Frida, Berndt*, where I did my degree project but have continued to collaborate with. Especially to *Caroline Fossum*, thanks for all your help and support in my PMWS related work. To *Lisbeth "Bettan" Fuxler* for your kindness and help. *Lotta Wik* for all our nice chats and your company during the lab course and having someone trying to avoid HA/HI with…..

Per Wallgren, for all you help and support in the PMWS studies.

To some past and present phD students: *Eva M, Sofia, Sara, Marlene and Eva E* thanks for your company during courses and AWs. And to *Malin Hagberg* for your nice company, both during (long) coffee breaks and in front of the real-time computer, during your short time as a virologist.

To *Åsa* and *Maria* for both of your friendships. For the great company during our years at SLU – from waiting for a frog in the botanical garden to all our wine and salmon evenings in Gottsunda.

To *Jessica*, my beloved sister, and to your family with Dan and your three wonderful children, Anton, Pontus and Sara, for your love – wish you all would live closer!!

To my *mum*, for your continuous support and love. For always being there for me (and for all our laughter's trying to construct yet another IKEA furniture on each visit… ☺)